

CARACTÉRISATION DU REPLIEMENT EN TEMPS RÉEL DU RIBORÉGULATEUR ADÉNINE

par

Patrick St-Pierre

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

FACULTÉ DES SCIENCES

UNIVERSITÉ DE SHERBROOKE

Sherbrooke, Québec, Canada, mai 2016

Le 19 mai 2016

*le jury a accepté la thèse de Monsieur Patrick St-Pierre
dans sa version finale.*

Membres du jury

Professeur Daniel Lafontaine
Directeur de recherche
Département de biologie

Professeure Pascale Legault
Évaluateur externe
Université de Montréal
Département de biochimie et médecine moléculaire

Professeur Martin Bisailon
Évaluateur interne
Département de biochimie

Professeur Vincent Burrus
Président rapporteur
Département de biologie

SOMMAIRE

Vers la fin des années 1990 et au début des années 2000, l'idée que l'ARN puisse interagir directement avec de petits métabolites pour contrôler l'expression de certains gènes devient de plus en plus acceptée. Des recherches menées à cette époque ont permis la découverte de plusieurs structures d'ARN hautement conservées nommées riborégulateurs. La structure de ces ARN leur permet de reconnaître spécifiquement un ligand. La reconnaissance du ligand entraîne ensuite un changement de conformation dans l'ARN responsable du contrôle de l'expression génétique. Le but de cette thèse est d'étudier la structure et les changements de conformation du riborégulateur associé au gène *add* liant l'adénine chez *Vibrio vulnificus*. Ce riborégulateur étant relativement simple, les informations recueillies lors de cette étude pourront servir à comprendre le fonctionnement de riborégulateurs plus complexes.

Dans l'introduction, la découverte des riborégulateurs sera décrite en plus des caractéristiques particulières et de l'importance de ces ARN. Par la suite, quelques exemples démontrant l'importance des structures d'ARN seront abordés. Ensuite, les techniques de fluorescence utilisées pour étudier les structures d'ARN au cours de cette thèse seront présentées. Enfin, les recherches effectuées sur les riborégulateurs adénine seront détaillées afin d'aider le lecteur à bien comprendre le type de riborégulateur au centre de cette thèse.

Le chapitre 1 traite du repliement de l'aptamère suite à la liaison avec l'adénine. Dans ce chapitre, il est démontré que l'aptamère peut adopter trois conformations. Une modification de la séquence de l'aptamère de type sauvage a permis d'isoler ces trois conformations. Il a ensuite été possible d'identifier les caractéristiques propres à chacun des états.

Le chapitre 2 s'intéresse à une région précise du riborégulateur adénine. Dans ce chapitre, la conformation du cœur de l'aptamère est étudiée plus en profondeur. Il y est possible de constater que le repliement du cœur de l'aptamère influence l'interaction boucle-boucle en présence de magnésium et de ligand. De plus, la présence de ligand, en concentration suffisante, permet le repliement du cœur et favorise le rapprochement des tiges P2 et P3 dans un aptamère muté pour empêcher la formation de l'interaction boucle-boucle. Il semble donc que le repliement du cœur de l'aptamère influence la structure globale de l'aptamère.

Finalement, les travaux présentés dans les chapitres 1 et 2 seront mis en contexte avec la littérature scientifique disponible. Cette discussion tentera de réconcilier certaines observations contradictoires. Il sera ensuite question de l'impact que les travaux présentés dans cette thèse peuvent avoir dans le domaine de l'ARN. Enfin, quelques études à réaliser en continuité avec ces travaux seront proposées.

REMERCIEMENTS

Obtenir un diplôme de doctorat en biologie moléculaire représente bien plus que les nombreuses heures passées à lire des articles scientifiques, à répéter des expériences qui devraient fonctionner et à préparer des présentations scientifiques. Bien que tout cela soit nécessaire, la réussite d'un doctorat passe avant tout par l'acquisition d'une méthode de réflexion, efficace, rigoureuse et honnête. L'utilisation constante de ce mode de pensée lors des études graduées a nécessairement un effet sur la vie de tous les jours d'un doctorant. Je suis persuadé que la façon de réfléchir acquise lors des études graduées favorise le fonctionnement en société et nous aide à devenir de meilleures personnes. Il s'agit d'un très grand avantage qu'il faut utiliser à bon escient. D'ailleurs, le monde entier gagnerait à avoir plus de dirigeants ayant une formation en sciences ou faisant confiance aux scientifiques. Mais je m'égare, car je veux ici simplement remercier ceux qui m'ont permis d'apprendre à réfléchir.

Il y a tout d'abord mes parents, Louise Fiset et Réjean St-Pierre, qui m'ont toujours donné les meilleurs outils dont ils disposaient pour permettre ma réussite scolaire. Je voudrais aussi remercier monsieur Benoît Fournier, un enseignant en biologie au secondaire, qui a une grande influence sur mon choix de domaine d'étude. Au niveau universitaire, j'aimerais d'abord remercier mon directeur de recherche à la maîtrise, le docteur Jean-Pierre Perreault, puisque c'est sous sa tutelle que j'ai réalisé l'importance de la méthode scientifique. Je dois également mentionner mon directeur de recherche au doctorat, le docteur Daniel Lafontaine avec qui j'ai pu tester et perfectionner ma façon de penser. Nous avons eu d'excellentes discussions scientifiques et philosophiques et j'espère qu'il y en aura beaucoup d'autres. La majorité des résultats présentés dans cette thèse n'auraient pu être obtenus sans l'aide du docteur Juan Carlos Penedo-Esteiro de l'University of St-Andrews. Évidemment, je me dois de remercier les membres de mon jury de thèse les docteurs Martin Bisailon, Vincent Burrus

et Pascale Legault pour tout le temps qu'ils m'ont accordé afin de lire cette thèse et de l'évaluer.

Réussir un doctorat peut prendre beaucoup de temps. Par contre, le temps passe très vite lorsque l'on est en bonne compagnie. Je ne crois pas qu'il y ait meilleurs collègues de laboratoire que ceux que j'ai connus au cours de mon doctorat. Bien sûr, il y a eu quelques accrochages, mais on a toujours trouvé un moyen de s'entendre au final. De toute façon, les beaux moments, les discussions mémorables, les soirées arrosées, les parties de volley-ball et les sorties de laboratoire surpassent amplement tous les différends que nous avons pu avoir. Je ne vous nommerai pas de peur d'en oublier, mais sachez que j'ai adoré travailler avec vous chers membres présents et passés du laboratoire Lafontaine.

Finalement, je dois mentionner deux personnes dont l'importance peut être sous-estimée lors de l'obtention d'un doctorat. Il y a tout d'abord celle avec qui je partage ma vie depuis plus de 11 ans. Sans sa patience et son soutien, je ne déposerais peut-être pas cette thèse aujourd'hui. Je veux aussi parler de ma fille qui aura deux ans au moment où j'aurai officiellement le titre de docteur. Je souhaite à tous ceux qui auront des enfants pendant leurs études graduées d'avoir un bébé comme elle. Avec une famille aussi extraordinaire, les écueils sur la route menant au doctorat disparaissent d'eux même.

Je vous aime Véronique et Léonie.

TABLES DES MATIÈRES

SOMMAIRE.....	IV
REMERCIEMENTS	VI
TABLES DES MATIÈRES	VIII
LISTE DES ABBRÉVIATIONS.....	X
LISTE DES TABLEAUX	XIII
LISTE DES FIGURES	XIV
INTRODUCTION	1
1. L'ARN.....	1
1.1 Les bases de la biologie moléculaire	1
1.2 Le rôle central de l'ARN en biologie moléculaire	4
1.3 Niveaux structuraux des ARN.....	10
1.4 L'importance de la structure des ARN.....	14
2. Les riborégulateurs	16
2.1 Définition des riborégulateurs	16
2.2 Importance des riborégulateurs	17
2.3 Identification des riborégulateurs.....	19
2.4 Classes de riborégulateurs	20
2.3 Plateforme d'expression et mécanismes de régulation.....	29
3. Riborégulateurs reconnaissant les purines.....	39
3.1 Les purines.....	39
3.2 Types de riborégulateurs liant les purines	40

3.3	Structure des riborégulateurs liant les purines.....	43
3.4	Études des riborégulateurs purines par fluorescence.....	48
4.	Objectifs du projet de recherche.....	53
CHAPITRE 1	56
	Patron de repliement du riborégulateur <i>add</i> observé par FRET en mode molécule unique..	56
CHAPITRE 2	88
	Repliement du cœur du riborégulateur adénine observé au niveau de la molécule unique...	88
CHAPITRE 3	114
	Discussion.....	114
	Mise en contexte dans la littérature scientifique	114
	Impact potentiel dans le domaine de l'ARN	119
	Perspectives	120
	Conclusion.....	124
ANNEXE	126
	Fluorescence tools to investigate riboswitch structural dynamics.	126
BIBLIOGRAPHIE	177

LISTE DES ABBRÉVIATIONS

2AP : 2-aminopurine

A : Adénine

ADN : Acide désoxyribonucléique

AdoCbl : 5'-deoxy-5'-adenosylcobalamine

ARN : Acide ribonucléique

ARNase : Enzyme dégradant l'ARN

ARNt : ARN de transfert

ARNt^{leu} : ARN de transfert spécifique à la leucine

ARNm : ARN messenger

ARNr : ARN ribosomal

ATP : Adénosine triphosphate

B. subtilis : *Bacillus subtilis*

C : Cytosine

Cas : CRISPR *associated*

c-di-AMP : Adénosine monophosphate cyclique

c-di-GMP : Guanosine monophosphate cyclique

C. reinhardtii : *Chlamydomonas reinhardtii*

CRISPR : *Clustered regularly interspaced short palindromic repeats*

DAP : 2-6-diaminopurine

E. coli : *Escherichia coli*

F : Conformation repliée (*Folded*)

FMN : Flavine mononucléotide

FRET : *Förster Resonance Energy Transfer*

G: Guanine

GlcN6P : Glucosamine-6-phosphate

GTP : Guanosine triphosphate

HDV : Virus de l'hépatite D

I : Conformation intermédiaire (*Intermediate*)

J : Jonction

L : Boucle (*Loop*)

Mg²⁺ : magnésium

Moco : Cofacteur du molybdène

NADH : Nicotinamide adénine dinucléotide

P : Tige (*Paired region*)

pH : potentiel Hydrogène

RBS : Site de liaison du ribosome (*Ribosome binding site*)

RMN : Résonance magnétique nucléaire

PréQ₁ : Pré-queuosine (7-aminométhyl-7-deazaguanine)

SAH : S-adénosyl-homocystéine

SAM : S-adénosyl-méthionine

SELEX : *Systematic evolution of ligand by exponential enrichment*

SHAPE : *Selective 2'Hydroxyl acylation Analyzed by Primer Extension*

siARN : Petit ARN interférent (*small interfering RNA*)

sm-FRET : FRET en mode molécule unique (*single molecule FRET*)

T : Thymine

THF : Tétrahydrofolate

TPP : Thiamine pyrophosphate

Tuco : Cofacteur du tungstène

U : Uracile

U : Conformation non repliée (*Unfolded*)

UTR : Région non-traduite (*Untranslated region*)

V. vulnificus : *Vibrio vulnificus*

ZTP : 5-aminoimidazole-4-carboxamide riboside 5'-triphosphate

LISTE DES TABLEAUX

Introduction

Tableau 1.1	Classes de riborégulateurs, leurs aptamères et leurs ligands	20
--------------------	--	----

LISTE DES FIGURES

Introduction

Figure 1	Hypothèse centrale de la biologie moléculaire	2
Figure 2	Mécanisme de régulation transcriptionnel rho-indépendant	31
Figure 3	Mécanisme de régulation transcriptionnel dépendant de la protéine Rho	32
Figure 4	Mécanisme de régulation traductionnel	33
Figure 5	Mécanisme de régulation contrôlant la dégradation	35
Figure 6	Mécanisme de régulation de l'épissage alternatif	36
Figure 7	Structure secondaire des riborégulateurs liant les purines	43
Figure 8	Représentation schématique de la structure cristalline du complexe adénine-aptamère adénine	45
Figure 9	Diagramme de Jablonski expliquant la fluorescence et la phosphorescence	48
Figure 10	Efficacité de FRET en fonction de la distance des fluorophores	51
Figure 11	FRET en fonction du temps pour un aptamère liant l'adénine	52

Chapitre 1

Figure 1	Structure secondaire du riborégulateur adénine de <i>Vibrio vulnificus</i>	62
Figure 2	Analyse sm-FRET du vecteur P2-P3 de l'aptamère adénine	68
Figure 3	Analyse sm-FRET du vecteur P1-P3 de l'aptamère adénine	70
Figure 4	Analyse sm-FRET du vecteur P1-P2 de l'aptamère adénine	72
Figure 5	Analyse sm-FRET du vecteur P2-P3 de l'aptamère adénine en fonction de la concentration de magnésium	74
Figure 6	Analyse SHAPE de l'aptamère adénine	76
Figure 7	Patron de repliement de l'aptamère adénine <i>add</i>	79
Figure supplémentaire 1	Analyse sm-FRET du vecteur P2-P3 en fonction de la concentration de magnésium	81
Figure supplémentaire 2	Analyse sm-FRET du vecteur P2-P3 muté dans la L3	82
Figure supplémentaire 3	Analyse sm-FRET du vecteur P1-P2	83
Figure supplémentaire 4	Analyse SHAPE complète de l'aptamère adénine	84

Chapitre 2

Figure 1	Structure secondaire du riborégulateur adénine de <i>Vibrio vulnificus</i>	93
Figure 2	Efficacité de FRET en fonction de la concentration de magnésium	98
Figure 3	Histogramme de l'efficacité de FRET pour le vecteur P1-J2-3 à différentes concentrations de magnésium et d'adénine.	101
Figure 4	Histogramme de l'efficacité de FRET pour le vecteur P1-J2-3 muté dans la boucle L2 à différentes concentrations de magnésium et d'adénine	105
Figure 5	Représentation schématique des interactions intramoléculaires de l'aptamère adénine	108

Chapitre 3

Figure 1	Patron de repliement de l'aptamère <i>add</i> déterminé par spectroscopie de force atomique	118
-----------------	---	-----

Annexe

Figure 1	Structures chimiques et résumé des interactions de type paire de bases impliquant les analogues de nucleotides 2AP et Pyrrolo-dC	131
Figure 2	Observation des changements de conformations locaux chez les riborégulateurs liant les purines	136
Figure 3	Observation des changements de conformations longue-distance induits par l'empilement hélical et la formation du pseudonoeud chez le riborégulateur liant la S-adenosylmethionine	141
Figure 4	Utilisation de la 2AP, comme ligand intrinsèquement fluorescent mimant l'adénine, pour l'investigation du processus de reconnaissance du ligand chez les riborégulateurs liant les purines	147
Figure 5	Patrons de repliement proposés pour divers aptamères	157
Figure 6	Méthode de dénaturation chimique en mode molécules uniques afin d'étudier les interactions entre le ligand et l'aptamère liant l'adénine	163

INTRODUCTION

1. L'ARN

L'objet de cette thèse est l'étude des changements de structure d'un acide ribonucléique (ARN). Une telle étude peut sembler d'un intérêt limité de prime abord. Pourtant, la compréhension des interactions régissant la conformation des ARN est devenue incontournable dans le domaine de la biologie moléculaire puisque plusieurs mécanismes cellulaires sont directement régulés par la structure de l'ARN. Afin de bien apprécier l'importance des études structurales de l'ARN, il est nécessaire de faire un survol de certaines découvertes faites depuis les 65 dernières années. Ce survol permettra de démontrer le rôle central qu'occupe l'ARN dans l'expression génétique de même que l'importance de la structure de certains ARN.

1.1 Les bases de la biologie moléculaire

Le terme biologie moléculaire a été utilisé pour la première fois en 1938 par Warren Weaver (Howe, Wang et al. 2015) pour désigner une discipline scientifique qui se trouve au croisement de la génétique, de la biochimie et de la physique. Cette discipline a connu ses premiers moments de gloire dans les années 50. En effet, en 1952 Alfred Hershey confirme hors de tout doute que l'information génétique est transmise par l'acide désoxyribonucléique (ADN), grâce à ses études sur les bactériophages (Hershey et Chase 1952). L'année suivante, James Watson et Francis Crick déterminent la structure physique de l'ADN en interprétant les données expérimentales de diffraction des rayons X sur des molécules d'ADN (Watson et Crick 1953). La démonstration de la conformation en hélice de deux brins d'ADN leur vaudra le prix Nobel de physiologie et médecine en 1962 en compagnie de Maurice Wilkins. À partir

de ce moment, l'ADN était reconnu comme étant le support de l'information génétique. Les scientifiques ont ensuite tenté d'expliquer comment l'information génétique encodée dans l'ADN pouvait être exprimée dans les cellules vivantes.

Francis Crick publia un article en 1970 qui résume les informations disponibles et les hypothèses émises par la communauté scientifique à ce sujet (Crick 1970). Cet article établit, avec les connaissances disponibles à l'époque, le «dogme» de la biologie moléculaire. À ce sujet, il est intéressant de noter que Francis Crick a lui-même admis, dans son autobiographie (Roy, Hohng et al. 2008), avoir mal utilisé le terme dogme. En effet, Crick ne concevait pas un dogme comme quelque chose qui ne peut pas être mis en doute. Il utilisait plutôt le terme comme phrase-choc pour parler d'une «Hypothèse centrale» de la biologie moléculaire qui n'avait pas beaucoup de fondement expérimental à ce moment. La figure 1 est tirée de l'article de 1970 et résume bien cette hypothèse centrale où l'information génétique est surtout transmise de l'ADN à l'ARN et ensuite aux protéines.

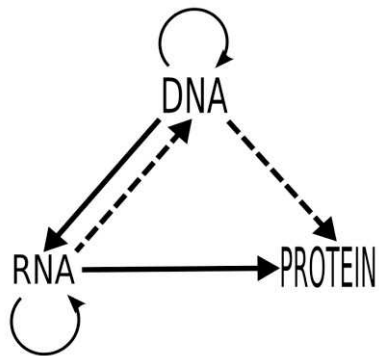


Figure 1 Tirée de (Crick 1970). Hypothèse centrale de la biologie moléculaire selon les connaissances de 1970

Les flèches représentent la direction dans laquelle l'information est transmise. Les flèches pointillées sont des hypothèses émises par Francis Crick.

À ce moment, les scientifiques s'accordaient pour dire que l'ADN était transcrit en acides ribonucléiques par des ARN polymérases telles que découvertes chez *Escherichia coli*

(Stevens 1960, Hurwitz et Bresler 1961) et dans les pois (Huang, Maheshwari et al. 1960) et que ces ARN étaient ensuite traduits en protéines par les ribosomes (Palade 1955). Selon ce qui était connu à l'époque, l'ARN ne servait que d'intermédiaire entre l'information contenue dans l'ADN et la fonctionnalité incarnée par les protéines. Cette idée que l'ARN ne servait que de messenger vient sans doute de sa grande ressemblance avec l'ADN et de sa plus faible stabilité par rapport à ce dernier. Du point de vue de la ressemblance, il faut savoir que tous les deux sont des acides nucléiques formés par un enchaînement de nucléotides. Les nucléotides les plus abondants dans la nature sont composés d'un groupement phosphate, d'un sucre à 5 carbones (le désoxyribose pour l'ADN, le ribose pour l'ARN) et d'une base azotée. Il existe cinq types de nucléotides naturels différents et comme le groupement phosphate et le sucre à 5 carbones sont toujours les mêmes, c'est la base azotée qui détermine la nature d'un nucléotide. Les bases azotées retrouvées dans l'ADN sont l'adénine (A), la thymine (T), la guanine (G) et la cytosine (C) alors que dans l'ARN la thymine est remplacée par l'uracile (U) (Voet et Voet 2004). Pour ce qui est de la stabilité, il a été constaté que le ribose de l'ARN est plus réactif que le désoxyribose de l'ADN causant plus facilement des cassures dans les brins d'ARN (Voet et Voet 2004). De plus, il avait été démontré que l'ADN se retrouvait sous une forme double brin (Crick 1970) ce qui favorise sa stabilité et sa réparation alors que l'ARN était majoritairement retrouvé sous forme simple brin, donc plus fragile et difficile à réparer (Voet et Voet 2004). Tout cela portait à croire que l'ADN était l'unique support de l'information génétique alors que l'ARN ne servait que de copie temporaire pouvant être jetée une fois l'information transmise aux protéines.

Il était aussi accepté que seules les protéines puissent remplir diverses fonctions biologiques en raison de la diversité des acides aminés qui les composent. En effet, contrairement à l'ADN et à l'ARN qui ne possèdent que 4 molécules de base (les nucléotides), il existe plus de 20 acides aminés naturels qui peuvent s'enchaîner pour créer une protéine. Ce qui veut dire que si l'on considère uniquement l'ordre dans lequel s'enchaînent 10 groupements de base, les protéines ont 10^{16} possibilités supplémentaires pour s'assembler par rapport aux acides nucléiques. Cette observation suggère que les protéines ont plus d'opportunité que les acides

nucléiques pour remplir des fonctions particulières. Il avait aussi été démontré que les protéines pouvaient former des enzymes catalysant des réactions biochimiques importantes, qu'elles pouvaient jouer un rôle structural dans la cellule et qu'elles formaient des anticorps et des hormones (Voet et Voet 2004).

Des travaux réalisés par le groupe de Nirenberg ont permis de déterminer comment l'information passe des ARN aux protéines. Ils ont démontré qu'un codon formé de 3 nucléotides dans l'ARN détermine l'identité de l'acide aminé retrouvé dans une protéine (Nirenberg et Matthaei 1961). A la suite de ces travaux, il a été possible de déchiffrer le code génétique en trouvant à quel acide aminé chaque codon correspondait. L'ordre dans lequel les acides aminés s'enchaînent pour former une protéine est donc déterminé par l'enchaînement des codons de l'ARN et, par conséquent, par la séquence d'ADN.

L'ensemble des travaux présentés précédemment, permet de mettre en place ce qu'il convient d'appeler les bases de la biologie moléculaire. C'est-à-dire que l'information génétique transite de l'ADN vers les protéines en passant par un intermédiaire d'ARN. Déjà en 1970, Francis Crick supposait que ce modèle de base était trop simple et proposait que l'ADN puisse transmettre directement l'information aux protéines et que l'ARN pouvait modifier l'information encodée dans l'ADN (Figure 1). Même si ces hypothèses furent avant-gardistes à l'époque, il était loin de se douter à quels points les interactions ADN-ARN-protéines étaient plus complexes qu'il ne l'imaginait et le rôle central qu'occupe l'ARN dans ce système.

1.2 Le rôle central de l'ARN en biologie moléculaire

Au début des années 1970, le rôle central de l'ARN dans l'expression génétique n'est pas considéré. Pourtant, cette molécule est habituellement placée entre l'ADN et les protéines dans

les schémas de l'époque. Cette perception était sur le point de changer puisque les scientifiques se sont rapidement rendu compte que l'ARN peut aussi contenir de l'information génétique et la transmettre à l'ADN. En effet, des enzymes capables de produire de l'ADN à partir d'une matrice d'ARN ont été identifiées dans des extraits viraux en 1970 (Baltimore 1970, Temin et Mizutani 1970). De plus, des virus dont le génome est composé exclusivement d'ARN avaient déjà été identifiés (Voet et Voet 2004). Il a aussi été démontré que le complexe responsable de la formation des protéines, le ribosome, est composé de protéine et d'ARN (Voet et Voet 2004) sans oublier que les ARN de transfert (ARNt) sont nécessaires au déchiffrement du code génétique (Nirenberg et Matthaei 1961). Dès lors, l'importance de l'ARN dans l'expression génétique ne faisait plus aucun doute et l'intérêt pour ce type de molécules n'a fait que s'accroître.

Cet intérêt a mené à la démonstration de l'activité catalytique de l'ARN par deux groupes indépendants au début des années 1980. L'équipe de Thomas Cech a tout d'abord mis en évidence la capacité catalytique de l'ARN permettant l'excision des introns du groupe I (Cech, Zaug et al. 1981). L'année suivante, le laboratoire de Sidney Altman démontra qu'un ARN est responsable de l'activité catalytique de la RNase P, une enzyme nécessaire à la maturation des ARNt (Guerrier-Takada, Gardiner et al. 1983). Ces découvertes leur ont valu conjointement le prix Nobel de chimie en 1989 et ont ouvert la voie à l'identification d'une panoplie d'ARN catalytiques naturels et artificiels (Talini, Branciamore et al. 2011, Martin, Unrau et al. 2015). Le nom ribozyme a été donné à ce type d'ARN et provient de la contraction des mots «*ribonucleic acid*» et «enzyme». Ce nom évoque clairement la capacité de l'ARN à remplir des fonctions catalytiques jusqu'alors associées uniquement aux protéines. De nos jours, l'importance des ARN catalytiques n'est plus à faire. En effet, en plus des deux exemples mentionnés précédemment, des ribozymes sont retrouvés dans le génome du virus de l'hépatite D (HDV) (Sharmeen, Kuo et al. 1988), dans les viroïdes (Cote et Perreault 1997, Navarro et Flores 1997), et dans le complexe d'épissage eucaryote (Papasaïkas et Valcarcel 2015). Enfin, il est désormais acquis que le centre catalytique du ribosome est composé presque exclusivement d'ARN (revue dans (Ramakrishnan 2014)). Comme il a déjà été démontré que

les ribosomes traités avec des protéases (Noller, Hoffarth et al. 1992) conservent une certaine capacité à produire des liens peptidiques, il est fort probable que l'ARN soit responsable de la production de toutes les protéines retrouvées dans les organismes vivants. Même avant ces découvertes, l'importance de l'ARN devenait de plus en plus évidente. Cependant, peu de gens s'attendaient à ce que l'ARN puisse contrôler l'expression génétique. Cette fonction a longtemps été attribuée uniquement à l'activité de certaines protéines.

Le contrôle de l'expression génétique est essentiel à la vie. Par exemple, une bactérie doit pouvoir rapidement s'adapter à son environnement au risque de mourir. Si leur génome était exprimé de façon constante, cela impliquerait que les bactéries devraient modifier leur génome pour acquérir de nouveaux gènes ou éliminer des gènes inutiles afin de s'adapter aux changements de conditions environnementales. De telles modifications sont possibles, mais ne se font pas instantanément (Voet et Voet 2004). Les bactéries seraient mortes avant d'avoir le temps de s'adapter. Les organismes eucaryotes font face au même problème à la différence que les mécanismes d'acquisition de nouveaux gènes prennent plus de temps à se mettre en place (Voet et Voet 2004). De plus, toutes les cellules d'un organisme eucaryote possèdent le même génome. Pourtant, les cellules formant le cœur ont une physiologie et des fonctions extrêmement différentes de celles du cerveau. Il est alors évident que toute l'information pour créer un organisme et lui permettre de conserver son homéostasie se retrouve encodée dans son génome, mais que tous les gènes ne sont pas exprimés de façon constante et dans toutes les cellules. Le mode de contrôle le plus répandu est la régulation de l'initiation de la transcription par des protéines au niveau du promoteur des gènes (Voet et Voet 2004). Ce mode de régulation est tellement répandu et efficace que, jusqu'à la fin des années 1990, il semblait impossible que l'ARN puisse participer à la régulation de l'expression génétique de façon significative.

Cette perception a été brisée en 1998 alors que les chercheurs Graig Mello et Andrew Fire ont découvert le système d'interférence par l'ARN (Fire, Xu et al. 1998). Le fonctionnement de ce

système implique des petites séquences d'ARN de 21 à 24 nucléotides, appelés petits ARN interférents (siRNA de l'anglais *small interfering RNA*), qui sont utilisés par des complexes protéiques pour cibler des ARN messagers (ARNm) et ainsi initier leur dégradation ou empêcher leur traduction (Svoboda 2014). Ces phénomènes affectent directement le niveau d'expression des gènes retrouvés sur les ARNm ciblés. Le mécanisme d'ARN interférence tel que décrit par A. Fire et C. Mello semble être unique aux eucaryotes. Cependant, un mécanisme semblable a été découvert chez les procaryotes environ une décennie plus tard (Bolotin, Quinquis et al. 2005). Le mécanisme procaryote fonctionne à l'aide d'ARN appelés CRISPR (*Clustered Regularly Interspaced Short Palindromic Repeats*) et des protéines Cas (CRISPR associated) et sert de véritable système immunitaire adaptatif chez les bactéries les protégeant des bactériophages et d'autres acides nucléiques envahisseurs (revue dans (Marraffini 2015)). Dans les mêmes années, plusieurs petits ARN régulateurs non codant ont été étudiés chez *E. coli*. Il a été démontré que ces petits ARN contrôlent directement l'expression génétique de plusieurs mécanismes essentiels aux bactéries comme le transport du fer, l'utilisation des sucres, la production de toxine et bien d'autres (revue dans (Gottesman 2004, Repoila et Darfeuille 2009)). Tout comme pour les siARN et les ARN CRISPR, la plupart de ces ARN nécessitent l'intervention d'une protéine, notamment Hfq, pour remplir leur fonction. Tous ces exemples, impliquent la participation de molécules d'ARN et de protéines pour assurer un contrôle génétique. Cependant, une classe d'ARN découverte également au début des années 2000, suggère que certains ARN peuvent réguler l'expression génétique sans l'aide des protéines.

Une branche importante de la biologie moléculaire consiste à isoler et caractériser les protéines responsables de la régulation de l'expression des gènes. Malgré les nombreux succès obtenus au fil des années, il a été impossible dans certains cas de découvrir quelle protéine ou quel groupe de protéines étaient responsables de la régulation génétique. Au début des années 2000, après l'accumulation de plusieurs observations, il est suggéré que l'ARNm peut détecter directement la concentration de métabolites présents dans une cellule (Gelfand, Mironov et al. 1999, Stormo et Ji 2001). La base de cette hypothèse était la démonstration

qu'il est possible de développer des molécules d'ARN artificiels, appelées aptamères, capables de lier de petits métabolites (Jenison, Gill et al. 1994, Gold, Brown et al. 1997). Le développement des aptamères artificiels se fait par la technique appelée la SELEX (*S*ystematic *E*volution of *L*igand by *E*xponential *E*nrichement). La SELEX permet de sélectionner, à partir de millions de séquences possibles, les molécules d'ARN qui possèdent une caractéristique voulue, que ce soit la liaison d'un ligand ou une activité enzymatique. Devant la facilité avec laquelle il a été possible de sélectionner des ARN capables de lier des métabolites variés comme l'adénosine triphosphate (ATP) (Sassanfar et Szostak 1993), la théophylline (Jenison, Gill et al. 1994), ou le colorant Hoechst (Nguyen, DeFina et al. 2002), certains chercheurs ont soupçonné que des aptamères peuvent se retrouver dans des ARNm de façon naturelle.

Ces soupçons ont été renforcés lorsqu'une recherche bio-informatique a permis d'identifier une séquence hautement conservée dans la partie 5' non traduite (UTR de l'anglais *untranslated region*) des gènes de synthèse et de transport de la thiamine chez les bactéries Gram positives et Gram négatives (Miranda-Rios, Navarro et al. 2001). De plus, lorsque des variations étaient observées dans cette séquence, des variations complémentaires étaient présentes pour conserver la structure prédite de l'ARN. Dans cette étude, il a aussi été démontré que cette région, nommée *thibox*, était nécessaire à la régulation de l'expression des gènes de synthèse et de transport de la thiamine. Le modèle de régulation proposé par l'équipe de recherche implique que le repliement de l'ARN empêche la reconnaissance du site d'initiation de la traduction et que la présence de thiamine libère le site de liaison du ribosome. Les auteurs de l'article disaient avoir été incapables d'identifier des facteurs protéiques responsables de cette régulation. Si bien que dans leur conclusion, ils suggèrent que l'ARN interagit directement avec la thiamine pour réguler l'expression génétique. Enfin, les auteurs ont remarqué qu'un mécanisme semblable à celui qu'ils proposent pourrait expliquer la régulation déjà observée pour des gènes impliqués dans le transport de la cobalamine (Ravnum et Andersson 2001) et dans la synthèse de la riboflavine (Gelfand, Mironov et al. 1999).

L'année suivante, le groupe de Ronald Breaker démontre hors de tout doute l'existence de ce mécanisme de régulation. En effet, ils prouvent qu'un opéron de gènes responsables de la synthèse de la vitamine B₁ chez *Bacillus subtilis* est régulé au niveau de l'initiation de la traduction par la liaison directe de la 5'-deoxy-5'-adenosylcobalamine (AdoCbl) à la région 5'-UTR de son ARNm (Nahvi, Sudarsan et al. 2002). Ce laboratoire a aussi fait la démonstration que la régulation de l'expression des gènes *thiC* et *thiM* d'*E. coli* obéit au même principe (Winkler, Nahvi et al. 2002).

Vers la fin de l'année 2002, les groupes de Ronald Breaker et d'Evgeny Nudler démontrent simultanément que l'interaction directe entre un ARNm et un métabolite permet le contrôle de la transcription chez la bactérie *B. subtilis*. Le groupe de Nudler a d'abord montré que la thiamine pyrophosphate (TPP) et la flavine mononucléotide (FMN) contrôlent respectivement l'expression de l'opéron *tenA* et du gène *ribC* en interagissant directement avec la région 5' UTR des ARNm pour forcer l'arrêt de la transcription (Mironov, Gusarov et al. 2002). Quelques jours plus tard, le groupe de Breaker publie un article montrant le même mécanisme impliquant lui aussi la FMN, mais cette fois pour la régulation de l'opéron *ribDEAHT* (Winkler, Cohen-Chalamish et al. 2002). C'est suite à la publication de ces articles que le nom riborégulateur (*riboswitch* en anglais) a été accepté pour désigner des séquences d'ARN non codant capables de contrôler l'expression génétique en interagissant directement avec un métabolite. L'identification des riborégulateurs a permis de constater que l'ARN peut contrôler la transcription et la traduction, les deux grandes étapes de l'expression génétique, sans l'aide des protéines solidifiant ainsi son rôle central au sein de la biologie moléculaire.

En effet, à la fin de 2002, il était connu que l'ARN était nécessaire à la production de protéine (ARNm, ARNt et ARN ribosomiaux (ARNr)), pouvait être le support de l'information génétique (virus à ARN, viroïdes), avoir des fonctions catalytiques (ribozymes), interagir

directement avec des métabolites (aptamères, riborégulateurs) et contrôler l'expression génétique (siARN, petits ARN non codant, riborégulateurs). Depuis ce temps, chacune de ces fonctions remplies par l'ARN ont été confirmées et la compréhension du mode de fonctionnement n'a cessé de s'améliorer (revue dans (Gottesman 2004, Repoila et Darfeuille 2009, Ramakrishnan 2014, Mellin et Cossart 2015)). Le nombre et la diversité des fonctions remplies par l'ARN peuvent paraître étonnants. Après tout, les molécules d'ARN sont presque exclusivement composées des quatre mêmes nucléotides, ce qui limite les possibilités d'organisations différentes. Cependant, la séquence d'un ARN n'est pas l'unique déterminant de l'information ou de la fonction qu'il détient. L'autre déterminant capital est la structure qu'un ARN peut adopter sous différentes conditions.

1.3 Niveaux structuraux des ARN

La structure des ARN peut être divisée en 3 niveaux : primaire, secondaire et tertiaire. La structure primaire représente la séquence d'enchaînement des nucléotides d'une molécule d'ARN. C'est l'information de base contenue dans la molécule. Immédiatement après la transcription, la séquence de l'ARN correspond à celle retrouvée dans le génome. Cependant, plusieurs phénomènes post-transcriptionnels comme l'épissage peuvent modifier la séquence de l'ARN. Par exemple, jusqu'à 15% de la séquence d'un ARN de transfert peut être modifiée par rapport à sa séquence génomique. Ces modifications vont de la simple méthylation au changement de l'identité du nucléotide (adénosine vers inosine) (Su, Chan et al. 2014). Les techniques de séquençage et de chromatographie permettent de déterminer la structure primaire des ARN.

La structure secondaire d'un ARN est une représentation en 2 dimensions des interactions entre les nucléotides. En effet, les acides ribonucléiques peuvent interagir entre eux de différentes manières (Leontis, Stombaugh et al. 2002). Les plus abondantes sont les

interactions appelées Watson-Crick où l'adénine s'associe avec l'uracile et la guanine avec la cytosine. Un autre type d'interaction importante est appelé Wobble et se produit lorsque la guanine interagit avec l'uracile. Lorsque suffisamment de ces deux types d'interactions surviennent à proximité l'une de l'autre, la molécule d'ARN se replie sur elle-même pour former une structure double-brin d'ARN, aussi appelée tige. La structure secondaire d'un ARN est donc une représentation des régions double brin et simple brin. Sur cette représentation, les régions double brin sont nommées P, de l'anglais *paired region*, alors que les régions simple-brin peuvent être nommées L ou J. La nomenclature L vient de l'anglais *Loop* (boucle en français) et désigne une région simple brin dont les deux extrémités proviennent de la même tige (région P). Cet arrangement est l'une des structures d'ARN les plus fréquentes et est aussi appelé structure en épingle à cheveux (*hairpin*) en raison de sa ressemblance avec l'article de coiffure. Quant à la nomenclature J elle est associée aux régions simple brin qui relient deux tiges ensemble. En plus des lettres utilisées pour distinguer le type de structure se retrouvant sur un ARN, un numéro est associé à chacune des régions. Ce numéro est accordé en fonction de l'ordre dans laquelle chaque région est observée en partant de l'extrémité 5'-phosphate de l'ARN.

Le premier véritable exemple de détermination de la structure secondaire d'un ARN a été fait sur un ARNt. À cette époque, ce type d'ARN était plus facile à observer en raison de sa grande stabilité et de son abondance dans les cellules. Ainsi, en 1965, Robert W. Holley suggère trois structures secondaires possibles à l'ARNt de levure suite à des essais de dégradation avec des ARNases pancréatiques (Holley, Apgar et al. 1965). De ces trois structures, celle adoptant une structure de type «cruciforme» (*Cloverleaf*), s'avèrera très près de la réalité.

De nos jours, la détermination de la structure secondaire d'un ARN de plus en plus complexe nécessite l'emploi de plusieurs techniques basées sur le concept de la réactivité différentielle des structures d'ARN. Par exemple, une enzyme dégradant l'ARN (ARNase) peut avoir plus

de facilité à cliver les ARN simple brin que les ARN double-brin. En couplant un essai de dégradation par une ARNase avec un séquençage, il est possible d'identifier les régions protégées contre l'activité de l'enzyme. Dans le laboratoire du docteur Daniel Lafontaine, les techniques les plus couramment utilisées pour déterminer la structure secondaire des ARN sont la cartographie enzymatique (T1, RNase A) (Lemay, Penedo et al. 2006, Blouin et Lafontaine 2007), la cartographie par coupure en ligne (qui utilise le magnésium pour dégrader l'ARN) (Mulhbachter et Lafontaine 2007, Mulhbachter, Brouillette et al. 2010) et le SHAPE (*Selective 2'Hydroxyl acylation Analyzed by Primer Extension*) (Blouin, Chinnappan et al. 2011, Lemay, Desnoyers et al. 2011, Tremblay, Lemay et al. 2011, Smith-Peter, Lamontagne et al. 2015). Contrairement à la cartographie en ligne et par les ARNases, le SHAPE utilise la différence de réactivité du groupement 2'-OH des ribonucléotides pour sonder la structure des ARN. Lors de ces essais, un groupement chimique est ajouté au 2'OH des ribonucléotides les plus flexibles, donc les plus réactifs. Les conditions expérimentales sont ajustées afin que chaque ARN ne soit modifié en moyenne qu'une seule fois. Ensuite, une transcription inverse est faite en utilisant une ADN polymérase dépendante de l'ARN. La polymérase s'arrêtera aux modifications, donnant des ADN dont la longueur dépendra des positions modifiées sur l'ARN. Enfin, une migration sur gel de polyacrylamide sépare les ADN de différentes longueurs. Sur ce gel, les régions les plus accessibles au réactif utilisé pour le SHAPE, habituellement les régions simple brin, donneront un signal plus intense que les régions double-brin. Le SHAPE devient de plus en plus populaire pour déterminer les structures d'ARN. En effet, cette technique comporte moins de limites que les cartographies par coupure en ligne et par les ARNases. Par exemple, la cartographie en ligne utilise des ions divalents pour enclencher la dégradation de l'ARN. Il est alors difficile d'étudier la structure des ARN dans différentes conditions impliquant des ions divalents. Les ARNases quant à elles, sont spécifiques à certains nucléotides. Si ces nucléotides sont absents ou se retrouvent en très faible quantité dans la structure à étudier, les ARNases ne seront pas très utiles. Ces limites n'affectent pas le SHAPE puisque les modifications se font sur tous les nucléotides, peu importe les concentrations d'ions présents.

Connaître la structure secondaire d'un ARN permet d'en apprendre beaucoup sur ce dernier et d'émettre des hypothèses sur son mode de fonctionnement. Cependant, cette représentation en 2 dimensions ne reflète pas totalement la réalité. En effet, des régions qui peuvent sembler très éloignées dans la structure secondaire peuvent en fait interagir l'une avec l'autre dans l'ARN fonctionnel. C'est pourquoi la détermination du repliement de l'ARN en trois dimensions, la structure tertiaire, est essentielle pour bien comprendre le fonctionnement d'un ARN. La cristallographie s'est révélée une méthode très efficace pour obtenir ces informations. L'idée est d'obtenir un cristal de l'ARN et d'étudier le patron de diffraction des rayons X frappant ce cristal. Ce patron de diffraction est unique et dépend de la position de chacun des atomes dans la structure cristalline. Le concept est simple, mais son application demande temps, efforts et patience. Cette technique est utilisée depuis les années 1930 et a fait ses preuves plusieurs fois en révélant la structure tertiaire de nombreuses protéines. Depuis, cette technique a par la suite été utilisée avec succès pour obtenir les structures de nombreux ARN et complexe ARN-protéine notamment celle du ribosome (Mitra, Schaffitzel et al. 2005, Schuwirth, Borovinskaya et al. 2005). À cet égard, la structure cristalline du ribosome fournit l'indice le plus concluant indiquant que le ribosome puisse être un ribozyme (Mitra, Schaffitzel et al. 2005, Schuwirth, Borovinskaya et al. 2005).

L'ARNt a également été le premier type d'ARN dont la structure tertiaire a été déterminée. Les pionniers dans ce domaine ont été les groupes d'Alexander Rich et d'Aaron Klug qui ont publié presque simultanément leur structure du cristal de l'ARNt de la phénylalanine (Kim, Suddath et al. 1974, Robertus, Ladner et al. 1974). Leurs travaux ont démontré que l'ARNt adopte une forme ressemblant à un L en trois dimensions. Dans cette conformation, l'anticodon et le site de liaison de l'acide aminé se retrouvent aux extrémités de la molécule. Cette observation suggérait que le site de reconnaissance du codon est éloigné du site catalysant la formation du lien peptidique dans le ribosome. Les efforts mis en place pour déterminer les structures d'ARN ont permis de comprendre comment fonctionnent certains processus biologiques comme la traduction.

1.4 L'importance de la structure des ARN

La structure d'un ARN est intimement reliée à sa fonction. Encore une fois, l'étude des ARNt fournit de nombreux exemples de ce phénomène. L'un de ces exemples est l'ARNt mitochondrial liant la leucine (ARNt^{leu}). Des mutations dans la séquence de cet ARNt sont associées à plusieurs maladies humaines comme le diabète, la cardiomyopathie et l'encéphalopathie. L'une de ces mutations correspond au nucléotide U à la position 3271 qui est changé par un C (U3271C). Il a été démontré que cette mutation n'affecte pas la reconnaissance de l'ARNt^{leu} par les enzymes de modification ou de chargement des ARNt identifiés à ce jour. Le seul effet de cette mutation est de déstabiliser la tige de l'anticodon (Wittenhagen et Kelley 2003). La structure de l'ARNt^{leu} étant modifiée, ce dernier entre plus difficilement dans le ribosome et reconnaît moins efficacement le codon qui lui est associé. Les protéines produites grâce à un tel ARNt ont plus de chance de ne pas être assemblées correctement, ce qui entraîne de graves problèmes dans une cellule. Les personnes possédant la mutation U3271C souffrent généralement de myopathie et de diabète (Wittenhagen et Kelley 2003). Cet exemple médical permet de rendre plus concrète l'importance de la structure de l'ARN. Pour un biologiste moléculaire par contre, l'élucidation de la structure tertiaire du ribosome reste la démonstration ultime de l'importance de connaître la conformation des ARN.

L'élucidation de la structure tertiaire du ribosome s'est échelonnée sur plusieurs années. Rappporter cette histoire de façon complète et précise pourrait être l'objet d'une thèse en soi. C'est pourquoi il convient mieux, ici, de référer le lecteur à un excellent article paru dans la revue *Cell* en 2014 (Ramakrishnan 2014) qui relate de façon accessible, les grandes étapes dans l'étude des ribosomes. L'une de ces étapes s'est avérée être un tournant dans la compréhension de ce mécanisme si important pour la vie. Il s'agit de la cristallisation de la structure de la grande sous-unité du ribosome (Ban, Nissen et al. 2000). Avant cette

découverte, il était connu que le ribosome était composé d'ARN et de protéines. Jusqu'alors, il y avait un débat pour savoir quelle molécule remplissait la fonction catalytique dans le ribosome. La structure proposée par le laboratoire de Thomas A. Steitz a montré que le cœur catalytique du ribosome était composé uniquement d'ARN. Cette structure a donc fortement suggéré que le ribosome est un ribozyme nécessitant de nombreuses protéines pour soutenir sa structure. L'ARN ribosomal est ainsi devenu la molécule fonctionnelle d'une des macromolécules les plus importants à la vie. Sans la détermination de la structure du ribosome, le débat sur le rôle de l'ARN ribosomal ferait sans doute encore rage.

En plus de clore un débat houleux, les observations faites sur la structure tertiaire du ribosome ont révélé plusieurs motifs structuraux récurrents. Les motifs les plus abondants sont les tétraboucles, la boucle E, les jonction trois voies, les pseudonoeuds et les interactions triples. Certains de ces motifs sont des interactions longue distance permettant de compacter l'ARN et de stabiliser des groupements fonctionnels. Par exemple, les pseudonoeuds sont formés par des interactions de type Watson-Crick entre deux régions éloignées qui semblent simple brin sur la structure secondaire. D'autres motifs servent en réalité de médiateur entre l'ARN et d'autres molécules comme les protéines et les ions (Sweeney, Roy et al. 2015). La plupart des motifs identifiés chez le ribosome ont ensuite été observés chez d'autres ARN comme les ribozymes, les ARN du mécanisme d'épissage, les sARNnc et les riborégulateurs (Sweeney, Roy et al. 2015).

De nos jours, il est accepté que les ARN ne soient pas uniquement de longs filaments linéaires flottant dans la cellule, mais plutôt des molécules compactes et structurées. Ceci implique que les ARN messagers ont eux aussi des structures qui leur sont propres. Comme il l'a déjà été mentionné, la structure d'un ARN est intimement liée à sa fonction. Ceci implique donc que les ARN messagers peuvent remplir des fonctions supplémentaires à celle de transmettre l'information à traduire par le ribosome. En fait, dans la section 1.2 il a été question de ce phénomène. Il y était question des riborégulateurs qui sont des régions d'ARN messagers

structurées capables de réguler l'expression génétique. Ce type de molécules a été étudié avec un grand intérêt en raison de leur abondance chez les bactéries et de leur capacité à réguler l'expression génétique sans l'intervention des protéines. Ainsi, de nombreux types de riborégulateurs ont été identifiés au cours des dernières années. Ceux-ci interagissent avec une grande variété de métabolites et utilisant plusieurs mécanismes différents pour réguler l'expression génétique. La prochaine section permettra de bien définir ce que sont les riborégulateurs et d'apprécier leur diversité et leur importance dans l'expression génétique.

2. Les riborégulateurs

Dans la section 1, il a été mis en évidence que les ARN sont essentiels à la vie. Il a également été démontré que la structure des ARN est étroitement liée à leurs fonctions. Les études sur les riborégulateurs ont donc une importance particulière à la compréhension de la vie. En effet, ces ARN affectent la régulation génétique grâce à leurs structures et leurs changements de conformations. Comme cette régulation se fait sans l'intervention de protéines, ce mécanisme limite les intermédiaires entre la réception de l'information et la réponse à donner. Ceci diminue directement l'énergie requise pour maintenir l'homéostasie. Ce mécanisme est tellement simple et élégant que, pour certains scientifiques, il pourrait également s'agir des vestiges d'un des plus vieux modes de régulation génétique à être apparu avec la vie (Kloc, Foreman et al. 2011). Avant d'amener cette réflexion plus loin, il est nécessaire de bien définir ce que sont les riborégulateurs et la façon dont ils ont été découverts.

2.1 Définition des riborégulateurs

Les riborégulateurs sont des éléments de contrôle génétique généralement retrouvés dans la région 5'-UTR des ARNm bactériens. La structure adoptée par un riborégulateur lui permet de

lier un ligand particulier et de réguler l'expression du gène, ou de l'opéron de gènes, qui est encodé sur la même molécule d'ARN. L'action des riborégulateurs est habituellement celle d'un rétro-inhibiteur, puisque les gènes régulés sont souvent associés à la voie de synthèse ou au transport du métabolite que lie le riborégulateur.

Un riborégulateur peut être séparé en deux parties, l'aptamère et la plateforme d'expression. Comme mentionné précédemment, un aptamère est une région d'ARN structurée qui permet la reconnaissance spécifique d'un métabolite cellulaire. Pour une classe de riborégulateurs donnée, l'aptamère est relativement bien conservé. La plateforme d'expression, quant à elle, est responsable de la modulation de l'expression génétique et sa structure peut varier considérablement entre deux espèces bactériennes. En général, la liaison d'un ligand à l'aptamère entraîne un changement dans la structure de la plateforme d'expression et c'est ce changement de structure qui affecte l'expression du ou des gènes associés au riborégulateur.

2.2 Importance des riborégulateurs

La régulation par les riborégulateurs est très importante chez les bactéries. Par exemple, près de 4% des gènes de la bactérie *B. subtilis* sont régulés de cette façon (Winkler et Breaker 2005). Il existe au moins 25 classes de riborégulateurs qui peuvent tous lier un ligand particulier. La majorité d'entre eux ont été retrouvés chez les bactéries, mais quelques-uns, comme celui liant la thiamine pyrophosphate, ont aussi été identifiés chez les plantes et les champignons (Kim et Breaker 2008). Rien n'exclut la possibilité de retrouver d'autres riborégulateurs chez les eucaryotes étant donné la grande quantité d'ARN non codant retrouvée chez ces organismes. Tous ces ARN non codants ont la possibilité d'interagir avec des métabolites et d'initier un mécanisme de régulation. Malgré plusieurs efforts (Drory Retwitzer, Kifer et al. 2015), très peu de riborégulateurs ont été confirmés chez les eucaryotes. L'explication de ce phénomène est peut-être attribuée à la présence du noyau. En effet, chez

les procaryotes, la transcription se fait directement dans le cytoplasme et les ARN subissent peu de modifications. Les ARNm eucaryotes quant à eux sont produits dans le noyau avant d'être modifiés post-transcriptionnellement et exportés au cytoplasme. Les échanges entre le noyau et le cytoplasme sont limités et peu de métabolites peuvent traverser cette barrière (Voet et Voet 2004). Beaucoup de riborégulateurs identifiés jusqu'à présent semblent lier leur ligand pendant la transcription. Il est donc possible que la présence du noyau ait mis une pression sélective défavorable à l'existence des riborégulateurs chez les eucaryotes.

Il semble que la faible représentation des riborégulateurs chez les eucaryotes en comparaison aux procaryotes puisse être mise à profit dans le domaine de la santé. En effet, il est évident que les riborégulateurs sont impliqués dans la régulation de gènes essentiels à la survie des bactéries. Donc, affecter le fonctionnement normal d'un riborégulateur pourrait entraîner la mort de la bactérie. Comme il y a peu de riborégulateurs chez les eucaryotes, il y a de fortes chances que l'hôte de la bactérie ne soit pas affecté en même temps. Quelques groupes de recherche ont tenté de développer une nouvelle classe d'antibiotiques ciblant spécifiquement des riborégulateurs. Bien que certains résultats semblent prometteurs (Blount et Breaker 2006, Kim, Blount et al. 2009, Lee, Blount et al. 2009, Long, Ji et al. 2010, Mulhbacher, Brouillette et al. 2010, Deigan et Ferre-D'Amare 2011, Lunse, Schuller et al. 2014, Howe, Wang et al. 2015), il reste beaucoup de travail à faire avant de voir la première utilisation d'un tel antibiotique pour traiter une maladie animale ou humaine. Tel que suggéré dans une revue de littérature à laquelle a participé l'auteur de cette thèse (Mulhbacher, Brouillette et al. 2010), il faudra, entre autres, améliorer l'efficacité de molécules identifiées jusqu'à maintenant et démontrer leur innocuité envers l'organisme malade. Jusqu'à maintenant, seules quelques classes de riborégulateurs semblent se prêter à cette démarche. L'identification de nouveaux riborégulateurs ouvrira peut-être de nouvelles voies pour le développement de ces antibiotiques.

2.3 Identification des riborégulateurs

L'identification des riborégulateurs se fait habituellement en recherchant des séquences ou des structures conservées dans la région 5'-UTR des ARNm encodant des gènes qui répondent de façon concertée à un signal connu ou prédit (Henkin 2008). L'arrivée de la bio-informatique a facilité l'identification des riborégulateurs potentiels et a permis d'augmenter la complexité des motifs recherchés (Weinberg, Barrick et al. 2007). Ainsi, la liste des riborégulateurs potentiels s'est allongée et une plus grande variabilité dans le patron des riborégulateurs déjà identifiés a été démontrée. Les analyses bio-informatiques permettent d'identifier de potentiels riborégulateurs en recherchant des séquences et des structures d'ARN conservées en avant d'un groupe de gènes aux fonctions similaires chez plusieurs espèces. Afin de confirmer l'identité d'un riborégulateur, il est nécessaire de démontrer qu'un motif d'ARN peut lier un ligand sans l'aide d'une protéine ou d'autres facteurs cellulaires et que cette liaison affecte l'expression d'un gène. Des expériences dans des lignées cellulaires comme la cartographie chimique ou enzymatique et la fusion du potentiel riborégulateur avec un gène rapporteur *in vivo* sont donc nécessaires pour confirmer l'existence d'un riborégulateur. Dans la majorité des études bio-informatiques effectuées pour identifier des riborégulateurs chez les bactéries, c'est la région 5'-UTR des ARNm qui était ciblée. Cependant, la structure d'ARN liant la thiamine-pyrophosphate chez les plantes se retrouve dans la région 3'-UTR et dans des introns (Cheah, Wachter et al. 2007). Ceci met en évidence la nécessité d'élargir les recherches pour incorporer toutes les régions des ARN susceptibles de lier un métabolite, particulièrement pour les ARNm des eucaryotes. De plus, jusqu'à maintenant il y a presque toujours une association physique entre le riborégulateur et les gènes régulés ce qui facilite l'identification des riborégulateurs. Par contre, les riborégulateurs agissants en *trans* sur d'autres molécules d'ARN que celle sur laquelle ils se retrouvent sont plus difficiles à identifier et à caractériser (Henkin 2008) par les méthodes conventionnelles.

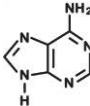
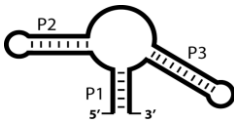
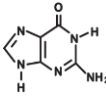
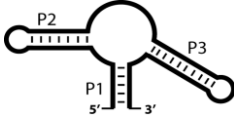
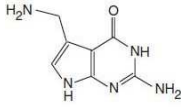


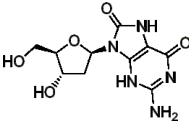
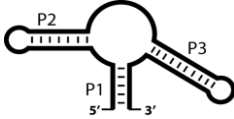
2.4 Classes de riborégulateurs

Depuis la découverte du premier riborégulateur en 2002, près de 25 classes de riborégulateurs liant des ligands particuliers ont été confirmées et il existe au moins autant de motifs d'ARN orphelins qui ont le potentiel d'être des riborégulateurs (Batey 2012). Les ligands et les aptamères des riborégulateurs confirmés sont présentés dans le tableau 1.1.

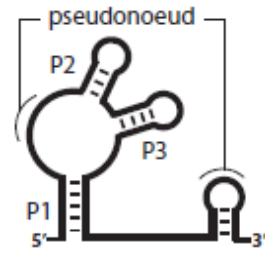
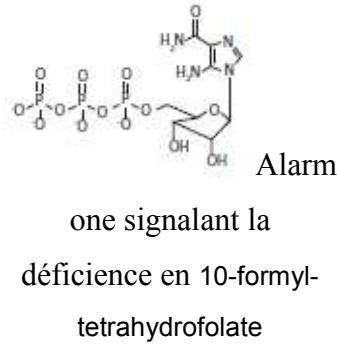
En regardant ce tableau, il est évident que la nature des ligands est très variée de même que la structure des aptamères permettant leur reconnaissance. Dans les lignes qui vont suivre, quelques particularités intéressantes sur différentes classes de riborégulateurs seront présentées. Dans cette section, une attention particulière sera apportée aux aptamères. Les plateformes d'expression seront présentées dans la section suivante. Les riborégulateurs liant les purines seront présentés en détail dans leur propre section étant donné qu'il s'agit du groupe auquel appartient l'ARN étudié dans cette thèse.

Les riborégulateurs présentés dans le tableau 1.1 sont ceux dont le rôle a été confirmé, mais il faut savoir que plusieurs autres riborégulateurs potentiels existent et attendent d'être caractérisés (Weinberg, Wang et al. , Batey 2012). De toutes ces classes de riborégulateurs, seul celui liant la thiamine pyrophosphate (TPP) est retrouvé dans des cellules eucaryotes (Sudarsan, Barrick et al. 2003, Galagan, Calvo et al. 2005, Croft, Moulin et al. 2007). Il est difficile d'expliquer pourquoi un seul type de riborégulateur a été identifié chez les cellules eucaryotes. Une première hypothèse suggère que la façon d'identifier les riborégulateurs ne soit pas optimale pour les eucaryotes. Comme il l'a été mentionné précédemment, il faut comparer les séquences génomique de plusieurs organismes afin d'identifier des motifs d'ARN conservés. Le nombre de génomes eucaryotes connu est plus restreint que celui des procaryotes, ce qui diminue la quantité de comparaisons possibles. Pour cette raison, les chercheurs ont surtout tenté de retrouver les motifs bactériens dans les cellules eucaryotes.

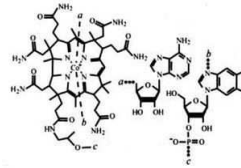
Tableau 1.1 Classes de riborégulateurs, leurs aptamères et leurs ligands

Riborégulateur / ligand	Ligand / Fonctions biologiques	Structure secondaire de l'aptamère
Adénine	 <p>Purine : composante de l'ADN et de l'ARN.</p>	
Guanine	 <p>Purine : composante de l'ADN et de l'ARN.</p>	
Pré-queuosine₁ (PréQ ₁)	 <p>Intermédiaire dans la biosynthèse de la queuosine.</p>	<p>PréQ₁-I</p>  <p>PréQ₁-II</p> 
2'-déoxy-guanosine	 <p>Composante de l'ADN.</p>	

ZTP
 (5-aminoimidazole -4-
 carboxamide riboside 5'-
 triphosphate)



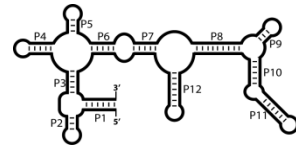
Adénosyl-cobalamine
 (AdoCbl)



ou

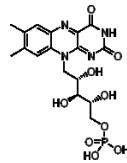
Coenzyme B₁₂
 (CoB₁₂)

Réactions de
 transméthylation et
 d'isomérisation.

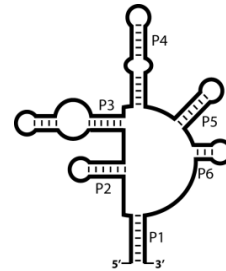


Méthylation de certains
 métaux.

Flavine mononucléotide
 (FMN)

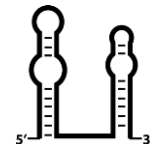


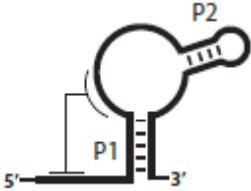
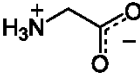
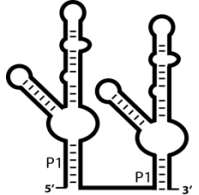
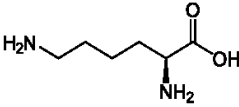
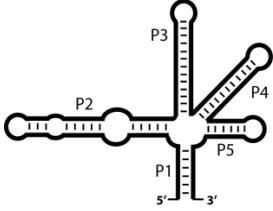
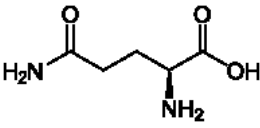
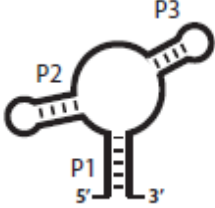
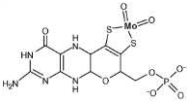
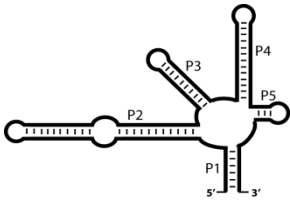
Réactions
 d'oxydoréduction.

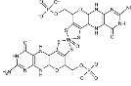
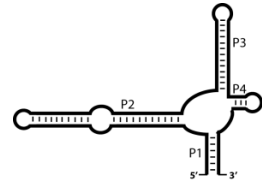
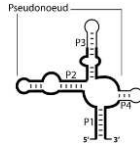
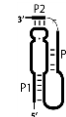
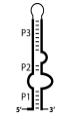
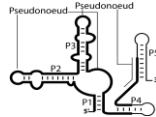

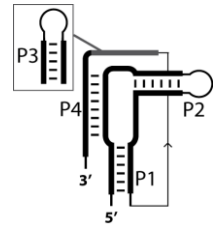
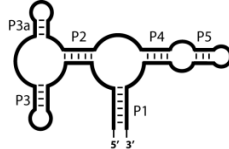


Ion magnésium
 (Mg²⁺)

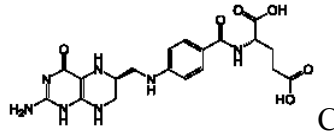
Ion divalent essentiel à
 diverses réactions
 biochimiques.



<p>Ion fluor (F⁻)</p>	<p>Ion monovalent.</p>	
<p>Glycine</p>	<p>Composante des protéines.</p>	 
<p>Lysine</p>	<p>Composante des protéines.</p>	 
<p>Glutamine</p>	<p>Composante des protéines.</p>	 
<p>Cofacteur du molybdène (Moco)</p>	<p>Biosynthèse d'enzymes et de protéines.</p>	 

<p>Cofacteur du tungstène (Tuco)</p>	 <p>Réactions d'oxydoréduction</p>	
<p>S-adenosyl-méthionine (SAM)</p>	<p>Dérivé de la méthionine. Réactions de méthylation (ADN, ARN, protéines).</p>	<p>SAM-I</p>  <p>SAM-II</p>  <hr/> <p>SAM-III</p>  <p>SAM-IV</p>  <hr/> <p>SAM-V</p> 
<p>S-adenosyl-homocystéine (SAH)</p>	<p>Intermédiaire de synthèse de cystéine.</p>	
<p>Thiamine pyrophosphate (TPP)</p>	<p>Participe aux réactions de décarboxylation.</p>	

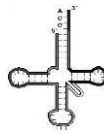
Tétrahydrofolate
(THF)



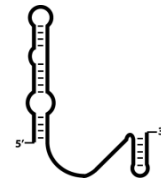
cofacteur impliqué dans le transfert d'atome de carbone



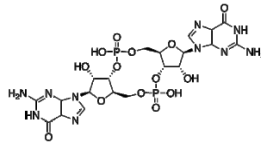
ARN de transfert libre
(ARNt)



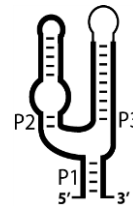
Synthèse des protéines.



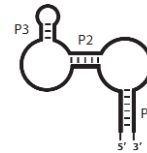
Guanosine monophosphate cyclique
(c-di-GMP)



Adaptation bactérienne (mobilité, biofilm et virulence).

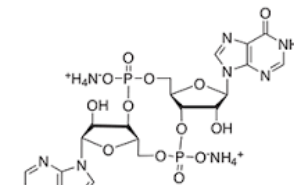


Classe I

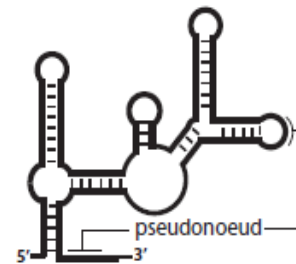


Classe II

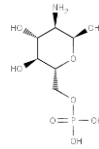
Adénosine monophosphate cyclique
(c-di-AMP)



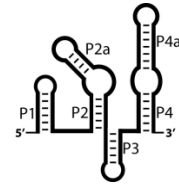
Messenger secondaire impliqué dans la réponse osmotique et la sporulation



**Glucosamine -6-
phosphate
(GlcN6P)**

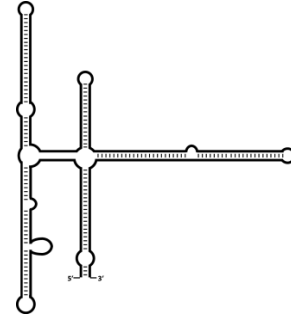


Glucide



pH

pH



Il est probable que la structure des riborégulateurs bactériens soit très différente de celle des eucaryotes; limitant les chances d'identification de cette manière. Une deuxième hypothèse suppose qu'il y a très peu de riborégulateurs chez les eucaryotes. Par exemple, l'apparition du noyau pourrait avoir limitée les interactions entre les ARN en cours de transcription et les métabolites dans le cytoplasme. L'utilisation des riborégulateurs serait ainsi devenue moins favorable chez les eucaryotes et leurs fonctions auraient été de plus en plus remplies par des protéines. Selon cette hypothèse, les riborégulateurs retrouvés chez les eucaryotes seraient des vestiges de l'évolution. Enfin, une troisième hypothèse propose que la présence du riborégulateur TPP dans les cellules eucaryote s'explique par un très improbable transfert horizontal de gènes (Miranda-Rios, Navarro et al. 2001). Cette hypothèse est supportée par l'identification de séquences correspondant aux riborégulateurs liant le bis-(3'-5')-cyclic dimeric guanosine mono-phosphate (c-di-GMP) chez des bactériophages (Sudarsan, Lee et al. 2008). En effet, si les riborégulateurs peuvent se retrouver dans des phages, il est possible qu'ils passent d'un type bactérien à un autre. Le transfert horizontal des riborégulateurs est donc possible. Bien qu'il y ait sans doute eu un moment au cours de l'évolution où le transfert de gènes entre les cellules eucaryotes et procaryotes a été possible, de tel transfert sont très improbable de nos jours.

Le tableau 1.1 montre les aptamères indépendants. Cependant, il est possible de retrouver des aptamères en tandem. C'est notamment le cas pour les riborégulateurs répondant à la glycine. Tous les riborégulateurs glycine identifiés jusqu'à maintenant possèdent deux aptamères liant l'acide aminé de façon coopérative (Mandal, Lee et al. 2004). Il semble que cette liaison coopérative serve à activer plus efficacement l'expression des gènes de l'opéron *gcvT* qui sont impliqués dans l'utilisation de la glycine comme source d'énergie par la cellule. Bien qu'il soit généralement retrouvé seul, l'aptamère liant la TPP a également été observé en tandem (Welz et Breaker 2007) tout comme l'aptamère reconnaissant la cobalamine (vitamine B₁₂) (Sudarsan, Hammond et al. 2006). Par contre, l'aptamère liant la cobalamine est uniquement retrouvé en tandem avec d'autres aptamères. Cet arrangement permet de contrôler l'expression des gènes à l'aide de deux signaux distincts ce qui permet plusieurs niveaux de contrôle de l'expression des gènes en fonction des conditions cellulaires (Sudarsan, Hammond et al. 2006).

Le tableau 1.1 permet aussi de constater une relation entre la taille du ligand et celle de l'aptamère qui reconnaît cette molécule. Généralement plus le ligand est volumineux et complexe en terme de groupement chimique, plus l'aptamère sera complexe lui aussi. Une exception notable à cette règle est l'aptamère liant la lysine qui est de taille imposante en comparaison à la simplicité de la lysine (Serganov et Patel 2009). En plus de la relation entre la taille du ligand et de l'aptamère, il est aussi possible de constater que chaque ligand n'est associé qu'à un type d'aptamère. Cette observation est surprenante étant donné que des aptamères, comme ceux reconnaissant la TPP, sont très semblables entre eux et se retrouvent dans des types cellulaires très différents (Kim et Breaker 2008). Les aptamères liant la S-adenosylmethionine (SAM) et le c-di-GMP font ici figure d'exceptions. En effet, 5 familles d'aptamères liant la SAM et deux familles liant le c-di-GMP sont connues de nos jours. Leurs structures secondaires et tertiaires ont pour la plupart, été caractérisées (Tausch, Xu et al. 2014). La variété de structures capables de lier le ligand SAM est impressionnante. Par exemple, la famille SAM-I adopte une architecture à 4 voies alors que la famille SAM-II est organisée autour d'un pseudonoeud et que la famille SAM-III possède un aptamère à 3 tiges-

boucles. Cette diversité de structure reconnaissant le même ligand suggère que ces trois riborégulateurs sont apparus indépendamment au cours de l'évolution (Peselis et Serganov 2014). Il est possible que tous ces aptamères différents soient apparus afin de pouvoir répondre à différentes concentrations de ligand ou en raison de l'importance ubiquitaire de la molécule SAM dans la survie cellulaire.

Au premier coup d'œil, les aptamères présentés dans le tableau semblent très différents. Cependant, un regard plus approfondi permet de remarquer plusieurs caractéristiques communes. Par exemple, dans la plupart des cas, les régions d'ARN double brin sont reliées entre elles par des jonctions simple brin qui forme le site de liaison de l'aptamère. Le site de liaison est souvent appelé le cœur de l'aptamère et est généralement représenté comme étant la partie centrale de la molécule d'ARN. Il est aussi possible de constater que la tige P1 est toujours formée par l'interaction des régions 5' et 3' de l'aptamère. Ceci signifie que la tige P1 fait partie de la plateforme d'expression et qu'elle est toujours impliquée dans le mécanisme de régulation. Le riborégulateur liant la glucosamine-6-phosphate (GlcN6P) est une exception notable à ce principe étant donné qu'il a été démontré que ce riborégulateur ne change pas significativement de conformation lorsqu'il lie son ligand. En effet, ce riborégulateur est également un ribozyme qui utilise le GlcN6P comme cofacteur pour enclencher la coupure de l'ARN et entraîner sa dégradation (Barrick, Corbino et al. 2004, Brooks et Hampel 2009).

L'observation révèle aussi des motifs structuraux récurrents. L'un d'entre eux sur lequel il convient de porter une attention particulière, est la jonction 3 voies. Ce motif survient lorsque trois tiges d'ARN double brin sont organisées autour de jonctions simple brin. Dans le cas des riborégulateurs liant les purines, la glycine, le glutamate, le c-di-GMP et la SAM ce motif correspond à l'ensemble de l'aptamère. De plus, la jonction trois voies se retrouve également dans des aptamères plus complexes comme ceux liant l'AdoCbl, le GlcN6P et la TPP, ce qui signifie que ce motif est très commun chez les riborégulateurs. Il est donc fort probable que les

informations qui seront obtenues lors des études des aptamères liant les purines pourront être extrapolées pour comprendre le fonctionnement de ces autres ARN.

2.3 Plateforme d'expression et mécanismes de régulation

La reconnaissance d'un ligand à elle seule n'explique pas comment un riborégulateur peut moduler l'expression des gènes encodés sur un ARNm. Il faut se rappeler qu'un riborégulateur est composé de deux parties; un aptamère et une plateforme d'expression. La liaison du ligand à l'aptamère entraîne des changements dans sa structure et affecte la conformation de la plateforme d'expression. En fait, un riborégulateur possède généralement deux conformations mutuellement exclusives, l'une liée au ligand et l'autre sans ligand. En fonction de la nature de la plateforme d'expression et des changements de conformation qui l'affectent, l'expression des gènes peut être modulée au niveau de la transcription, de la traduction, de l'épissage et de la stabilité de l'ARNm. Les plateformes d'expression ne sont pas associées exclusivement à certaines classes de riborégulateurs. Ainsi, il a été démontré que les riborégulateurs naturels reconnaissant la SAM peuvent contrôler la transcription ou la traduction des gènes en fonction de la plateforme d'expression qui est associée à leur aptamère (Wang et Breaker 2008). De plus, l'auteur de cette thèse a participé à un article montrant que les riborégulateurs adénine *pbuE* et *add* régulent respectivement la transcription et la traduction des gènes auxquels ils sont associés (Lemay, Desnoyers et al. 2011). Théoriquement, il serait possible de retrouver tous les mécanismes de régulation chez toutes les classes de riborégulateurs. À titre d'exemple, des groupes de recherche ont conçu des riborégulateurs artificiels en associant des aptamères avec diverses plateformes d'expression (Berens, Groher et al. 2015). Pour les riborégulateurs naturels identifiés à ce jour, les mécanismes de régulation transcriptionnelle et traductionnelle sont les plus répandus.

Régulation de la transcription

L'une des façons dont les riborégulateurs peuvent moduler l'expression des gènes est en arrêtant la transcription de l'ARN messager. Il existe deux mécanismes pour terminer la transcription de l'ARNm. Le premier mécanisme implique la formation d'une structure tige-boucle suivie d'une séquence riche en U dans l'ARNm. Cette structure particulière est appelée terminateur de transcription. En favorisant ou non la formation d'un terminateur, un riborégulateur peut ainsi contrôler l'expression d'un gène ou d'un opéron. L'exemple du riborégulateur FMN permet de bien comprendre ce mécanisme. La séquence de ce riborégulateur peut former deux structures mutuellement exclusives; un terminateur et un anti-terminateur. À faible concentration de FMN, l'ARN naissant, comportant le riborégulateur, adoptera préférentiellement la forme anti-terminatrice (à gauche dans la figure 2). L'ARN sera alors entièrement transcrit et les gènes de biosynthèse de la flavine seront exprimés. À forte concentration de FMN, l'aptamère forme un complexe avec la FMN ce qui empêche la formation de l'anti-terminateur (à droite dans la figure 2). Le terminateur pourra alors se former pour empêcher la transcription des gènes responsables de la production de la flavine (Wickiser, Winkler et al. 2005). Un mécanisme semblable peut être utilisé pour activer l'expression d'un gène. Dans ce cas, l'anti-terminateur est formé lorsque la concentration de ligand est élevée. Pour une raison encore inconnue, cette régulation au niveau de la transcription est retrouvée majoritairement chez les bactéries Gram positives. Le deuxième mécanisme d'arrêt de la transcription est dépendant de la protéine Rho. Ce mécanisme sert surtout à arrêter la transcription à la fin de la région codante des gènes. Il y a quelques années, il a été découvert que ce mécanisme est aussi exploité par les riborégulateurs associés aux gènes *mgtA* et *ribB* répondant respectivement aux concentrations de magnésium et de vitamine FMN (Hollands, Proshkin et al. 2012). La figure 3 suggère un mécanisme de fonctionnement pour ce type de régulation chez le gène *mgtA*. Il y est proposé qu'un site de liaison de la protéine Rho est accessible en forte concentration de magnésium et que la diminution de la concentration de magnésium change la structure du riborégulateur et séquestre ce site de liaison.

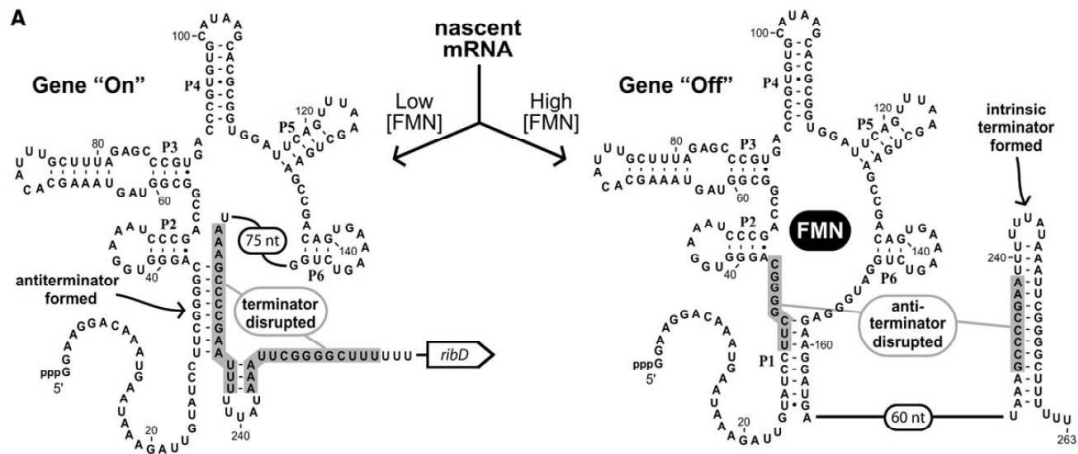


Figure 2 Tirée de (Wickiser, Winkler et al. 2005). Mécanisme de régulation transcriptionnelle rho-indépendant. Exemple du riborégulateur FMN de l'opéron ribDEATH de *B. subtilis*.

Le riborégulateur FMN emploie un mécanisme de contrôle des gènes qui fonctionne grâce à la présence mutuellement exclusive des structures appelées terminateur et anti-terminateur. À faible concentration de FMN (à gauche), le transcrit d'ARN forme l'anti-terminateur et la transcription de l'ARN sera complète. À haute concentration de FMN (à droite), le FMN forme un complexe avec l'aptamère ce qui empêche la formation de l'anti-terminateur, le terminateur est alors formé arrêtant la transcription. La structure du riborégulateur ne varie probablement pas entre ces deux états sur un laps de temps significatif au point de vue biologique. Chaque ARNm s'engage sans doute dans l'une des deux voies de repliement en fonction de la concentration de FMN présente au moment de leur transcription et conservera la conformation adoptée à ce moment.

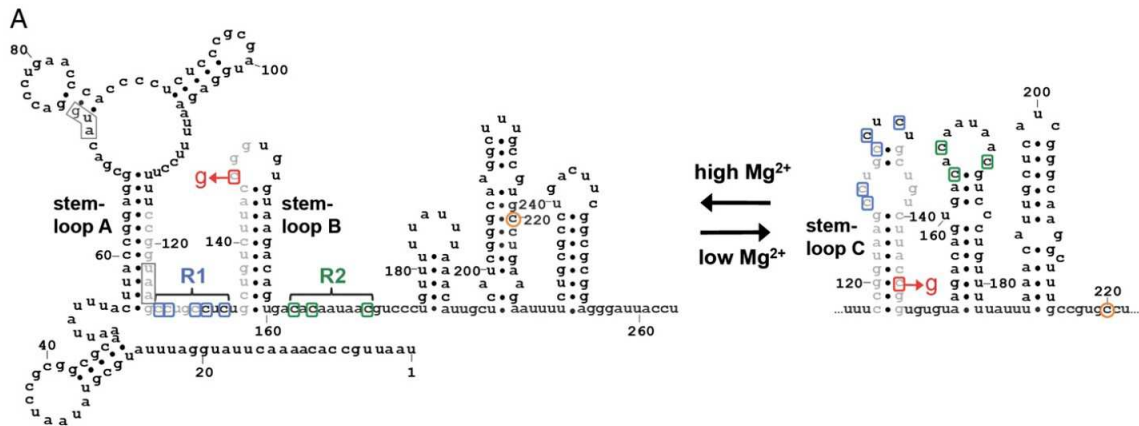


Figure 3 Tirée de (Hollands, Proshkin et al. 2012) Mécanisme de régulation transcriptionnelle dépendant de la protéine Rho. Exemple du riborégulateur magnésium *mgtA* de *Salmonella enterica*.

Un terminateur dépendant de la protéine Rho dans la séquence leader du gène *mgtA* prévient la transcription de la séquence codante lorsque la concentration de magnésium est élevée. Le schéma montre la structure secondaire de l'ARNm leader *mgtA*. À forte concentration de magnésium (à gauche) les tige-boucles A et B sont formées. La tige-boucle C est favorisée en faible concentration de magnésium. La position approximative du site de terminaison dépendant de Rho est encerclée en orange.

Régulation de la traduction

Le deuxième mode de régulation très répandu chez les riborégulateurs est la modulation de la traduction des gènes. Contrairement à la régulation transcriptionnelle, la régulation au niveau de la traduction est un phénomène beaucoup plus présent chez les bactéries Gram négatives. Ce mécanisme implique aussi la formation de deux structures mutuellement exclusives. Par contre, dans ce cas-ci, l'ARNm est toujours transcrit sur toute sa longueur et c'est la formation d'une tige qui séquestre le site de liaison du ribosome (RBS) qui est responsable de la régulation de l'expression des gènes (Figure 4). Dans le cas du riborégulateur liant la lysine qui se retrouve devant le gène *lysC* d'*E. coli*, le site de liaison du ribosome est libre d'accès en absence de lysine. Lorsqu'il y a suffisamment de lysine pour lier l'aptamère, il y a un changement de conformation qui permet la formation d'une structure tige-boucle séquestrant

le site de liaison du ribosome. Alors, le ribosome ne reconnaît plus cet ARN et le gène *lysC* n'est pas traduit. Ce mécanisme d'inhibition du gène *lysC* est nécessaire, car il encode pour une aspartokinase requise à la production d'acides aminés comme la thréonine, la méthionine et la lysine. Une surproduction d'acides aminés, ou d'intermédiaire de réaction de leur production pourrait être toxique pour les bactéries (Serganov et Patel 2009). L'inhibition de l'expression du gène *lysC* semble être primordiale pour la survie des bactéries. En effet, en plus de l'inhibition de la traduction, ce riborégulateur contrôle aussi l'arrêt de transcription de façon Rho dépendante (D. Lafontaine, communication personnelle) et possède également un mécanisme de régulation impliquant la dégradation de l'ARN (Caron, Bastet et al. 2012).

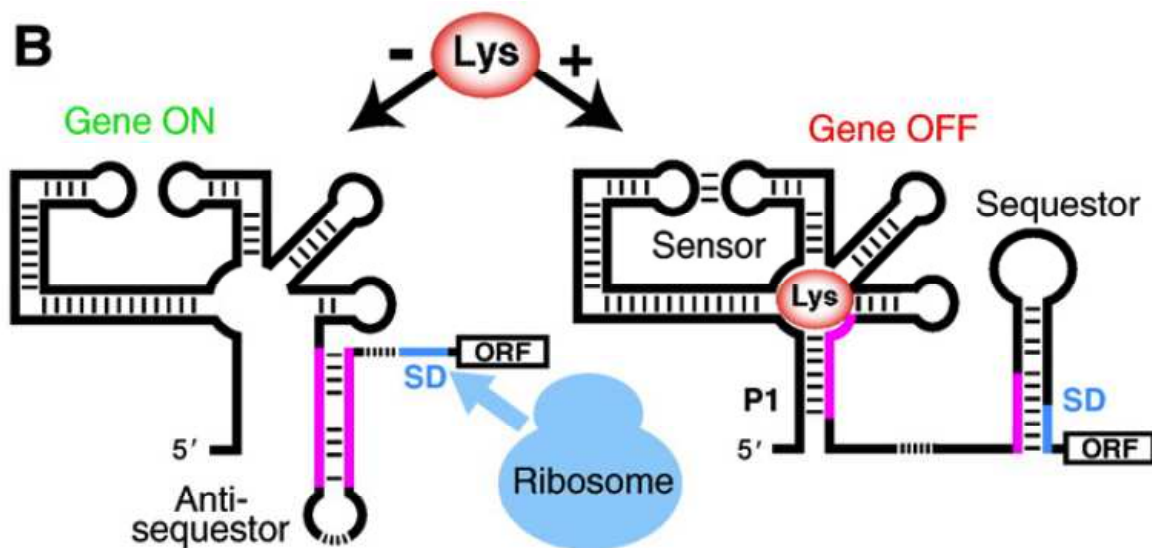


Figure 4 Tirée de (Serganov and Patel 2009) **Mécanisme de régulation traductionnelle. Exemple du riborégulateur lysine du gène *lysC* d'*E. coli***

Répression de la traduction du gène *lysC* par le riborégulateur liant la lysine. En absence de lysine (à gauche), le site de liaison du ribosome (SD) est libre et le gène *lysC* est traduit. En présence de lysine (à droite), la lysine forme un complexe avec l'aptamère ce qui permet la formation d'une tige d'ARN double-brin dans la plateforme d'expression qui séquestre le site de liaison du ribosome. Le gène *lysC* ne peut alors pas être traduit.

Régulation de la dégradation

La dégradation de l'ARN est un processus nécessaire à la survie cellulaire. Ce phénomène permet d'éliminer les ARN inutiles et de réutiliser les nucléotides afin de produire de nouveaux ARN nécessaires à la cellule. Dans certains cas, la dégradation est un mécanisme de protection contre les ARN envahisseurs comme les virus. Il existe plusieurs enzymes et complexes protéiques responsables de la dégradation. Le plus important est le complexe appelé dégradosome qui est responsable du renouvellement de la majorité des ARNm. Ce complexe interagit avec des séquences spécifiques d'ARN pour débiter la dégradation. Un tel site de reconnaissance existe dans la séquence 5' non codante du gène *lysC* d'*E. coli*. En 2012, il a été démontré que l'accessibilité de ce site de liaison du dégradosome est contrôlée par l'activité du riborégulateur présent dans cette région (Caron, Bastet et al. 2012). La figure 5 montre comment les mécanismes de régulation traductionnel et de dégradation cohabitent afin de réguler l'expression du gène *lysC*. Un mécanisme semblable a été aussi proposé pour le riborégulateur associé au gène *mgtA* (Hollands, Sevostiyanova et al. 2014). Pour d'autres riborégulateurs, il semble que la diminution du nombre de ribosomes présent sur l'ARN, suite à l'arrêt de traduction, entraîne la dégradation de l'ARN sans que ceux-ci contrôlent directement l'accessibilité d'un site de reconnaissance d'une ARNase (Caron, Bastet et al. 2012).

L'initiation de la dégradation de l'ARN ne nécessite pas nécessairement l'action du dégradosome. Par exemple, le riborégulateur liant la glucosamine -6-phosphate, aussi appelé riborégulateur *glmS*, utilise un mécanisme unique pour enclencher la destruction de l'ARN sur lequel il se trouve. Ce riborégulateur est retrouvé devant le gène *glmS* de *B. subtilis* et de 17 autres bactéries Gram positives. Ce gène encode pour une glutamine-fructose-6-phosphate amidotransférase, une enzyme qui utilise le fructose -6-phosphate et la glutamine pour générer du GlcN6P. Cette réaction est la première étape d'une voie de synthèse de la paroi cellulaire (Winkler, Nahvi et al. 2004). Évidemment, cette enzyme est inutile si la cellule possède

suffisamment de GlcN6P, c'est pourquoi les bactéries utilisent un riborégulateur spécifique au GlcN6P pour contrôler l'expression de ce gène. Le riborégulateur *glmS* est en fait un ribozyme dont l'activité d'autocoupeure est dépendante de sa liaison avec le GlcN6P. À faible concentration de GlcN6P, l'ARNm n'est pas affecté et le gène *glmS* est exprimé. En présence d'une concentration suffisante de GlcN6P, le ribozyme catalyse la coupure de l'ARNm ce qui enclenche sa dégradation et empêche la production de l'enzyme (Winkler, Nahvi et al. 2004). Il s'agit ici du seul riborégulateur-ribozyme identifié à ce jour.

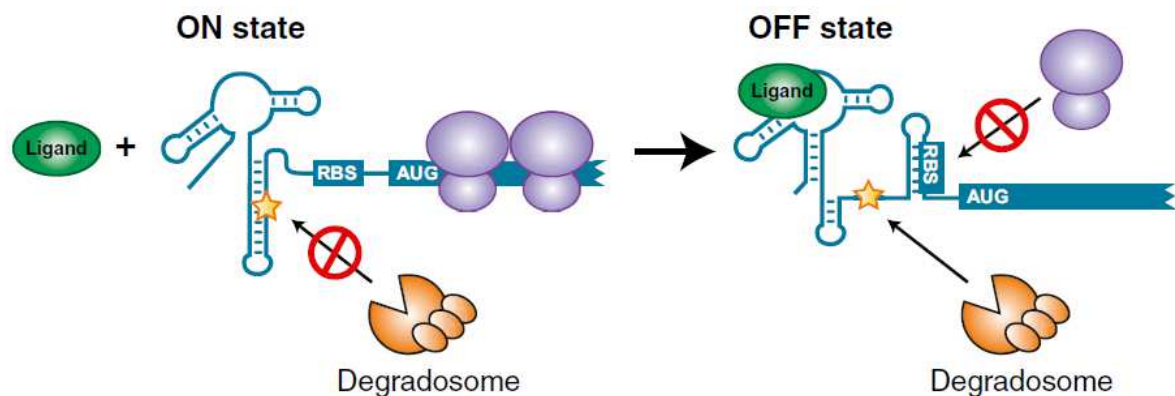


Figure 5 Tiré de (Caron, Bastet et al. 2012). **Mécanisme de régulation contrôlant la dégradation. Exemple du riborégulateur lysine du gène *lysC* de *E. coli***

Mécanismes de régulation du riborégulateur *lysC* contrôlant l'initiation de la traduction et la dégradation de l'ARNm. En absence de ligand (à gauche), le riborégulateur se replie dans la conformation ON. Ceci séquestre le site de coupure de l'ARNase E (indiqué par une étoile) et libère le site de liaison du ribosome (RBS) permettant la traduction. La liaison du ligand favorise la formation de la conformation OFF (à droite). Ceci expose le site de coupure par la ARNase E et séquestre le RBS, assurant la répression de l'expression du gène *lysC*.

Régulation de l'épissage

En plus d'être identifiés chez les bactéries, les riborégulateurs ont aussi été retrouvés chez les eucaryotes. Étant données les différences importantes entre ces deux types de cellules, il n'est pas surprenant de constater que les riborégulateurs eucaryotes emploient un mode de régulation particulier. Lors de la découverte de ce riborégulateur chez *Aspergillus oryzae*, il a

été démontré que la régulation s'effectuait au niveau de l'épissage alternatif du gène *thiA* (Kubodera, Watanabe et al. 2003). Par la suite, un mécanisme semblable a été retrouvé chez le champignon *Neurospora crassa* (Cheah, Wachter et al. 2007), les algues vertes *Chlamydomonas reinhardtii* (Croft, Moulin et al. 2007) et dans les plantes (Wachter, Tunc-Ozdemir et al. 2007). Le mécanisme de régulation est simple, la liaison de la TPP à l'aptamère altère la disponibilité d'un site d'épissage alternatif, modifiant ainsi la composition des introns conservés dans l'ARNm mature. Par exemple, le sixième intron contenu dans le gène *thiC* chez *Chlamydomonas reinhardtii* contient un tel riborégulateur (Figure 6). Ce gène encode

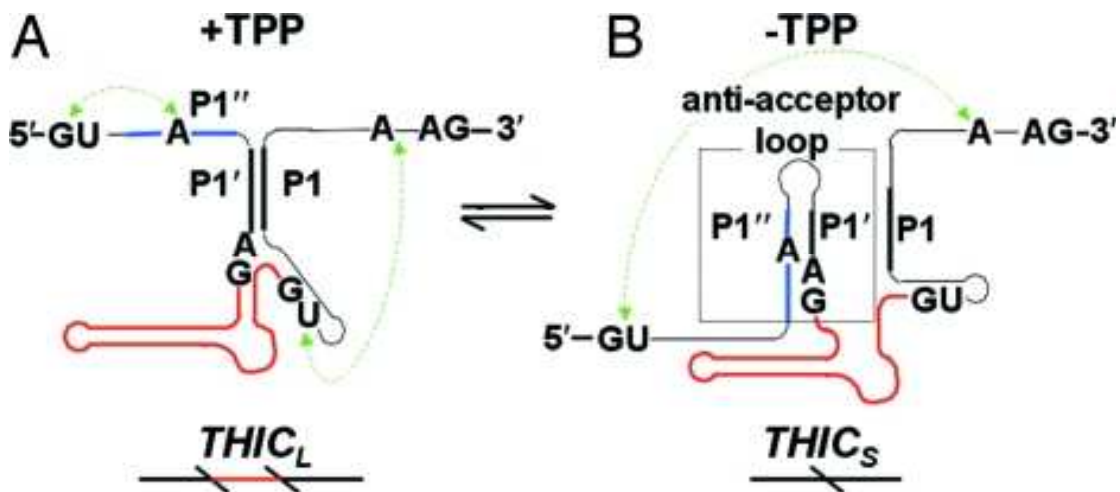


Figure 6 Tirée de (Croft, Moulin et al. 2007). **Mécanisme de régulation de l'épissage alternatif.** Exemple du riborégulateur liant le TPP dans le transcrit *THIC* de *C. reinhardtii*

(A) Diagramme du sixième intron du gène *THIC* contenant le riborégulateur. Les frontières de l'intron sont identifiées (GU/AG) de même que le site accepteur de l'épissage (A) et l'exon alternativement épissé en rouge. La ligne verte montre quelles régions seraient jointes pour former le lariat. Les lignes plus épaisses montrent les régions formant une tige d'ARN double-brin. À haute concentration de TPP, la TPP va lier le riborégulateur ce qui permettra l'interaction entre P1 et P1'. Dans l'intron, il y a deux sites accepteur de disponibles. Il y aura donc deux événements d'épissage de part et d'autre du riborégulateur. Ceci permet la production du transcrit *THIC_L* qui encode une protéine tronquée. (B) Lorsqu'il n'y a pas de TPP liée au riborégulateur, il y a formation d'une tige d'ARN double-brin alternative entre la région P1' et la région P1''. La région P1'' (en bleu) contient un site accepteur qui ne sera plus disponible pour l'épissage. Seul le site d'épissage en 3' de l'intron sera accessible et l'intron sera entièrement enlevé et le transcrit *THIC_S* encodant la protéine fonctionnelle sera produit.

une enzyme nécessaire dans les premières étapes de la formation de la thiamine. *C. reinhardtii* peut acquérir la thiamine de l'environnement sous sa forme phosphorylée (TPP). Lorsqu'il n'y a pas de TPP disponible, la bactérie est capable de produire elle-même cette molécule via une voie de biosynthèse impliquant de nombreuses enzymes. Lorsque la concentration de TPP est suffisante, il y a deux sites d'épissage qui sont disponibles de part et d'autre du riborégulateur et le sixième intron est incorporé dans l'ARNm mature. L'ARNm produit est appelé THIC_L et encode une protéine tronquée et inactive. À faible concentration de TPP, la région acceptrice du site d'épissage en 5' du riborégulateur est séquestrée dans une structure tige-boucle. L'ARNm produit (THIC_S) ne comportera donc pas le sixième intron et encodera la protéine fonctionnelle, ce qui permettra la formation de la thiamine (Croft, Moulin et al. 2007). Une particularité intéressante de ce type de régulation est que le riborégulateur ne se retrouve pas dans la région 5'-UTR de l'ARNm. Ce qui montre que la recherche de nouveaux riborégulateurs, particulièrement chez les eucaryotes, ne doit pas se limiter uniquement à la région 5'-UTR des ARNm.

Régulation en *trans*

Les riborégulateurs agissent habituellement en *cis*, c'est-à-dire sur les gènes situés en aval sur la molécule d'ARN. Il existe aussi des riborégulateurs qui peuvent moduler l'expression de gènes encodés sur un autre ARNm, comme le font les petits ARN régulateurs (en *trans*). Cette découverte a été faite dans la bactérie *Listeria monocytogenes*. Des riborégulateurs potentiels liant la SAM avaient auparavant été identifiés chez cette bactérie par bio-informatique. Ces potentiels riborégulateurs ont été nommés SreA-G et ils sont retrouvés devant des gènes impliqués dans le métabolisme et le transport de la cystéine et de la méthionine (Toledo-Arana, Dussurget et al. 2009). Deux de ces riborégulateurs, SreA et SreB, ont été confirmés et il a été démontré qu'ils agissent au niveau de la transcription, ce qui produit des petits ARN terminés prématurément lorsque la concentration de SAM est suffisante (Loh, Dussurget et al. 2009). Dans cet article, ils ont aussi démontré que les petits ARN formés par l'arrêt de

transcription de SreA et SreB peuvent réguler l'expression du gène de virulence *prfA* en se liant à la région 5'-UTR de son ARNm par complémentarité de séquence. Cette interaction se produit indépendamment de la liaison de la SAM à l'aptamère. Ces observations ouvrent la voie à un nouveau mécanisme de régulation impliquant les riborégulateurs. Il est possible que les riborégulateurs modulant la transcription de leurs gènes puissent aussi agir en *trans*. En effet, lorsque la transcription est arrêtée de façon prématurée, il y a toujours formation de petits ARN. La séquence de ces petits ARN correspond généralement à l'aptamère des riborégulateurs. En utilisant un mécanisme semblable à celui du riborégulateur SAM de *Listeria monocytogenes*, on peut facilement imaginer que les riborégulateurs modulant la transcription ne contrôlent pas seulement l'expression des gènes associés à la biosynthèse et au transport d'un métabolite précis, mais qu'il puisse aussi avoir un impact sur une ou des voies métaboliques plus ou moins rapprochées.

Les exemples présentés dans cette section montrent la grande variété et la grande complexité des riborégulateurs. Cette complexité est visible autant au niveau des structures adoptées par ces ARN qu'au niveau des mécanismes de régulation qu'ils emploient. La découverte et la caractérisation de nouveaux riborégulateurs, particulièrement chez les eucaryotes, demeurent des voies de recherches primordiales à la compréhension des mécanismes nécessaires à la vie. Cependant, il convient aussi d'améliorer les connaissances disponibles sur les riborégulateurs identifiés à ce jour. Car, même si les mécanismes généraux sont connus, certains détails de fonctionnement sont toujours nébuleux. L'une des approches à adopter pour obtenir des informations plus détaillées sur un système est d'en étudier un modèle simplifié. En extrapolant à partir des données acquises de cette façon, il est possible de mieux comprendre les parties les plus complexes du système. Dans le cas des riborégulateurs, ceux reconnaissant les purines méritent une attention particulière puisque leur structure est relativement plus simple que pour les autres classes de riborégulateurs.

3. Riborégulateurs reconnaissant les purines

Les riborégulateurs liant les purines sont d'excellents modèles d'étude en raison de leurs structures simples, de leur stabilité, de la variété de mécanismes des régulations et de la constance qu'ils affichent lors de leur repliement. De plus, le motif d'ARN adopté par l'aptamère, est retrouvé dans la structure de plusieurs autres riborégulateurs plus complexes. Mais, avant d'entrer dans les détails à propos de cette classe de riborégulateur, il est important de connaître la nature de leur ligand.

3.1 Les purines

Les purines sont des molécules azotées hétérocycliques formées d'un cycle pyrimidine fusionné à un cycle imidazole. Dans une cellule, l'adénine et la guanine sont les deux purines les plus abondantes et les plus importantes. En effet, ces deux molécules, ainsi que les pyrimidines thymine, cytosine et uracile, font partie des nucléotides formant l'ADN et l'ARN. En plus de leur participation dans la formation des nucléotides, les purines sont des composantes importantes de plusieurs molécules essentielles à la survie cellulaire comme l'ATP, le GTP, l'AMP cyclique, le NADH et la coenzyme A. La voie de biosynthèse des purines est complexe et elle implique la participation de beaucoup d'enzymes de même que la formation de plusieurs intermédiaires de réactions (Voet et Voet 2004). Il faut beaucoup d'énergie à la cellule pour produire toutes les enzymes nécessaires à la formation des purines. Pour éviter de gaspiller de l'énergie, les cellules doivent contrôler précisément l'expression de des gènes encodant ces enzymes. Il y a plusieurs mécanismes de régulation employés dans cette voie métabolique, dont, évidemment, l'utilisation de riborégulateurs.

3.2 Types de riborégulateurs liant les purines

Les purines reconnues par les riborégulateurs de façon naturelle sont la guanine (Mandal, Boese et al. 2003), l'adénine (Mandal et Breaker 2004), le 2-désoxyguanosine (Kim, Roth et al. 2007), preQ1 (Roth, Winkler et al. 2007) et le ZTP (Kim, Nelson et al. 2015). Plusieurs aptamères liants des purines ou des dérivés des purines comme le GTP, l'ATP, l'AMPc et la xanthine ont été obtenus de façon artificielle par SELEX (Kiga, Futamura et al. 1998, Koizumi et Breaker 2000, Davis et Szostak 2002). Les structures des aptamères artificiels sont différentes de celles des aptamères retrouvées dans la nature. Ceci suggère qu'il existe peut-être d'autres types de riborégulateurs qui peuvent lier les purines ou leurs dérivés et qu'ils restent encore à être identifiés. En effet, l'identification de nouveaux riborégulateurs se fait par bio-informatique en recherchant des structures d'ARN conservées chez plusieurs espèces. Il faut donc avoir une idée de la structure attendue pour mener une recherche efficace. Il est alors possible que des aptamères naturels n'aient pas été remarqués parce que leur structure n'a pas été recherchée. Une autre hypothèse pouvant expliquer ces différences entre aptamères naturels et artificiels est que la conformation retrouvée dans la nature est extrêmement favorisée par la sélection naturelle. Dans ce cas, il semble que les pressions sélectives naturelles n'ont pas été reproduites en laboratoire.

Le premier indice de l'existence de riborégulateurs liant les purines a été obtenu en 1997. Christiansen et al. ont remarqué que l'expression de l'opéron *xpt-pbuX*, qui encode des gènes de transport et de métabolisme des purines, de la bactérie *B. subtilis* était dépendante de la concentration de guanine. Cependant, aucun facteur protéique pouvant déterminer la concentration de guanine n'a été identifié à ce moment (Christiansen, Schou et al. 1997). En comparant la séquence du 5'UTR de l'opéron avec celles d'autres gènes impliqués dans le métabolisme des purines, une région hautement conservée a été observée. Cette région a vite été proposée comme étant un aptamère appartenant à un riborégulateur (Mandal, Boese et al. 2003). Cette séquence consensus a aussi été retrouvée devant des gènes impliqués dans le

transport et le métabolisme des purines de plusieurs autres organismes. La structure secondaire et le fait que la liaison de la guanine à ce riborégulateur favorise la formation d'un terminateur de transcription ont été démontrés grâce à des techniques de cartographie enzymatique et chimique (Soukup et Breaker 1999). Ce riborégulateur reconnaît spécifiquement la guanine et non l'adénine et plusieurs dérivés des purines. La constante de dissociation (K_d) de l'aptamère pour la guanine est aux alentours de 5 nM (Gilbert, Stoddard et al. 2006). La séquence consensus déterminée pour le riborégulateur *xpt-pbuX* a aussi été retrouvée dans le 5'UTR de gènes associés au transport et au métabolisme de l'adénine (Mandal, Boese et al. 2003, Mandal et Breaker 2004), notamment le gène *pbuE* chez *B. subtilis*. Contrairement à la majorité des riborégulateurs connus, celui-ci active l'expression du gène suite à la liaison du ligand (Lemay, Desnoyers et al. 2011). À ce jour, il est l'un des rares exemples de riborégulateur activateur. Les autres riborégulateurs reconnaissant l'adénine, identifiés par la suite (Porter, Marcano-Velazquez et al. 2014), répriment tous l'expression des gènes auxquels ils sont associés.

Récemment, deux autres classes de riborégulateurs se sont jointes au groupe reconnaissant les purines. Il s'agit des riborégulateurs dont l'aptamère est capable de lier la 7-aminométhyl-7-deazaguanine (preQ₁) et le 5-aminoimidazole-4-carboxamide riboside 5'-triphosphate (ZTP). Ces deux aptamères sont beaucoup moins répandus que ceux liant l'adénine et la guanine, et en raison de leur identification relativement récente, moins d'informations sont disponibles à leur sujet. Il convient cependant de les présenter brièvement ici en raison de leur appartenance au groupe des riborégulateurs liant les purines.

PreQ₁ est un intermédiaire dans la voie de synthèse de la queuosine (Q), un nucléotide extrêmement modifié qui se retrouve dans les ARNt. L'aptamère reconnaissant cette molécule est très simple et consiste en une séquence d'ARN de 34 nucléotides capable de former une structure en tige-boucle (Roth, Winkler et al. 2007). La structure cristalline de cet aptamère lié à preQ₁ a été obtenu en 2009 et a permis de constater la présence d'un pseudonœud qui n'avait

pas été suggéré auparavant (Klein, Edwards et al. 2009). Une deuxième classe de riborégulateurs capables de lier preQ₁ (preQ₁-II) a plus tard été identifiée et la structure de son aptamère est un peu plus complexe que celui de la première classe (Meyer, Roth et al. 2008). Celui-ci contient quatre tiges, dont la tige P3 qui est en même temps un pseudonoeud. Ces deux aptamères étant très simples, leur structure a été étudiée en profondeur en utilisant de nombreuses techniques comme la résonance magnétique nucléaire (RMN), la cristallographie au rayon X et d'autres méthodes biochimiques. Une excellente revue de littérature résumant les études de la structure de ces riborégulateurs a été publiée récemment (Eichhorn, Kang et al. 2014).

La ZTP est une purine modifiée qui sert de messenger cellulaire en cas de déficience en 10-formyl-tetrahydrofolate. Près de 30 ans après son identification, il a été démontré qu'il transmet son message à l'aide d'un riborégulateur (Kim, Nelson et al. 2015). La structure secondaire de cet aptamère se sépare en deux parties. La première partie est composée des tiges P1, P2 et P3 qui s'organisent en jonction à trois voies. La deuxième partie contient uniquement la tige P4 qui est séparée de la jonction trois voies par une région de 15 nucléotides non structurée. De plus, la boucle L4 forme un pseudonoeud avec la jonction J1-2 (Kim, Nelson et al. 2015). La même année de son identification, cet ARN a été cristallisé et sa structure confirme la présence du pseudonoeud et révèle les nucléotides responsables de la reconnaissance du ligand (Jones et Ferre-D'Amare 2015).

Avant l'identification des aptamères liant PréQ₁ et la ZTP, le groupe d'aptamères reconnaissant les purines était composé exclusivement d'aptamères liant l'adénine et la guanine. Étant donné la grande ressemblance de structure entre ces deux aptamères, l'expression «riborégulateurs liant les purines» est devenue acceptée pour désigner ces riborégulateurs et la structure particulière qu'ils adoptent. Cette expression sera utilisée dans ce sens pour le reste de cette thèse. Cependant, il est évident qu'avec la découverte de

boucles L2 et L3 en fonction de la tige sur laquelle elles se retrouvent (dans la figure 7, elles sont nommées L1 et L2 en raison d'une ancienne nomenclature). Les segments d'ARN simple brin entre les tiges sont nommées en fonction des tiges qu'ils joignent (J1-2, J2-3 et J3-1). Le patron de conservation observé entre les nucléotides des boucles suggère la formation d'une interaction boucle-boucle qui a été confirmée par l'étude des structures cristallines des aptamères.

Les interactions nécessaires aux riborégulateurs liant les purines pour la reconnaissance spécifique de leur ligand ont été déterminées par des analyses de cristallographie et de RMN. Les cristaux du riborégulateur associé à l'opéron *xpt-pbuX* chez *Bacillus subtilis* et celui du riborégulateur adénine associé au gène *add* de *Vibrio vulnificus* liés à leur ligand respectif ont été obtenus (Batey, Gilbert et al. 2004, Serganov, Yuan et al. 2004). La structure tridimensionnelle des deux aptamères est presque identique et elle permet d'englober complètement le ligand (Figure 8). La structure globale de l'aptamère permet d'expliquer certaines caractéristiques de la séquence consensus de ces riborégulateurs. Par exemple, les boucles L2 et L3 forment une interaction boucle-boucle et la longueur des tiges P2 et P3 est toujours d'au moins 7 paires de base en raison des limites structurales imposées par leur interaction. De plus, les nucléotides qui forment les régions simple brin entre les différentes tiges sont très conservés parce qu'ils sont nécessaires à la reconnaissance du ligand (Figures 7 et 8). Par exemple, la structure cristalline de l'aptamère *pbuX* a permis de confirmer que le ligand forme une interaction de type Watson-Crick avec le nucléotide C74 de la J3-1 qui confère la spécificité de l'aptamère. Cette interaction est stabilisée par une interaction triple avec le nucléotide U51 de la J2-3 et par un pont hydrogène avec le nucléotide U22 de la J1-2 (Serganov, Yuan et al. 2004). Les mêmes interactions ont été identifiées dans la structure cristalline de l'aptamère *add* (Batey, Gilbert et al. 2004). Ainsi, un seul nucléotide présent dans le cœur de l'aptamère est responsable de sa spécificité en formant une paire de base de type Watson-Crick avec la purine qu'il doit reconnaître. En effet, le seul changement du résidu C en U à la position 74 de l'aptamère *pbuX* permet de modifier sa spécificité pour la guanine vers l'adénine (Mandal et Breaker 2004). Les structures cristallines

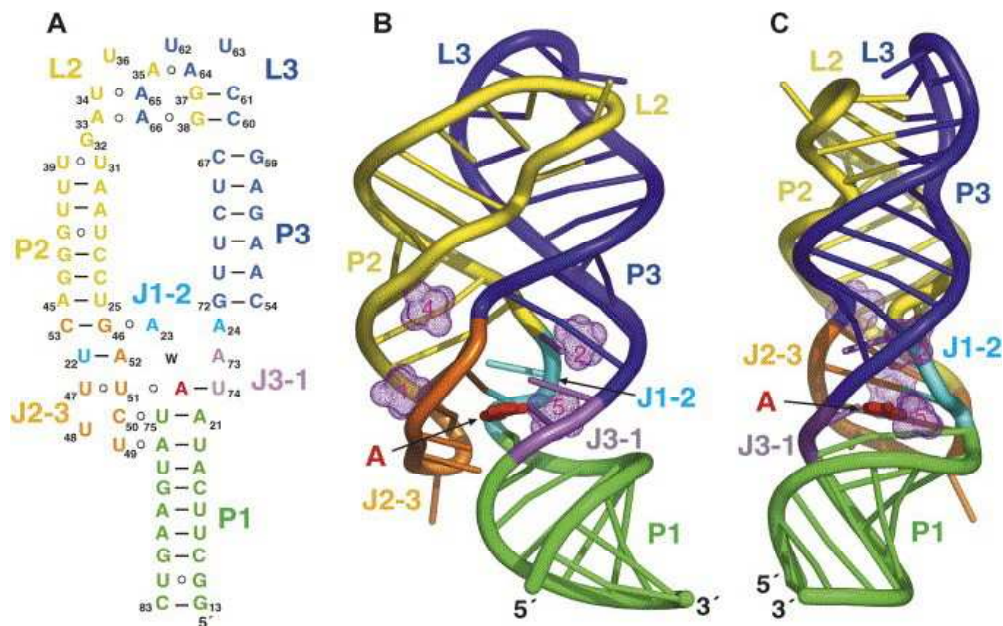


Figure 8 Tirée de (Serganov, Yuan et al. 2004). Représentation schématique de la structure cristalline du complexe adénine-aptamère adénine

(A) Représentation schématique des interactions tertiaires dans la structure repliée du complexe adénine-aptamère adénine. Le code de couleur est le suivant : Les tiges P1, P2 et P3 sont verte, jaune et bleue respectivement. La boucle L2 est jaune et la boucle L3 est bleue. Les segments de jonction entre les tiges J1-2, J2-3 et J3-1 sont cyan, or et violet respectivement. L'adénine liée au complexe est en rouge. Les interactions tertiaires impliquant des paires de bases Watson-Crick et des paires de bases non-canoniques sont représentées par des lignes pleines et pointillées respectivement. (B et C) Représentation en ruban du complexe adénine-aptamère adénine (rotation selon un angle de 90° par rapport à l'axe verticale). Le code de couleur est le même qu'en A. Quatre ions magnésium sont représentés par des surfaces mauves pointillées.

ont aussi révélé que le ligand est pratiquement inaccessible au solvant. Ces observations ont aussi permis de proposer un mécanisme d'action du riborégulateur par lequel la liaison du ligand induit un changement de conformation permettant la régulation des gènes. Selon cette hypothèse, les nucléotides 49 et 50, à proximité du site de liaison, forment une interaction triple avec la P1 stabilisant sa formation. Ainsi, l'extrémité 3' de la P1 ne peut pas former l'anti-terminateur et l'expression du gène est diminuée (Mandal, Boese et al. 2003). Cette particularité a été observée dans plusieurs autres riborégulateurs et il est généralement accepté que la P1 participe toujours aux mécanismes de régulation (Haller, Souliere et al. 2011).

Les structures cristallines des aptamères liés à leur ligand ont fourni plusieurs informations capitales à la compréhension des riborégulateurs liant les purines. Cependant, les analyses faites à partir de la structure cristalline ont leur limite. En effet, étant donné la nature même du cristal, il ne montre pas la structure de l'ARN en solution. De plus, les conditions pour obtenir la structure cristalline ne sont pas nécessairement représentatives de ce qui est retrouvé *in vivo*. Aussi, il a été possible d'obtenir seulement la structure cristalline des aptamères liés à leur ligand. Aucune structure cristalline de la forme non liée n'est disponible. En fait, la structure d'un aptamère purine non lié à un ligand est effectivement disponible. Par contre, cet aptamère comporte un nucléotide muté en guanine à la position 39 qui interagit spécifiquement avec le C nécessaire à la reconnaissance de la guanine. Cette interaction mime la liaison du ligand à l'aptamère et la structure cristalline correspond presque parfaitement à ceux obtenus en présence de ligand mentionnés précédemment (Delfosse, Bouchard et al. 2010). Enfin, une structure cristalline ne montre que la structure la plus stable de l'ARN. Ainsi, si l'ARN peut adopter plusieurs conformations, ce ne sera pas visible à partir d'une structure cristalline. Dans certains cas, la cristallographie a permis l'identification de changements de conformations. Il s'agit cependant de cas particuliers, comme le ribosome (Zhou, Lancaster et al. 2013), pour lesquels les cristaux de plusieurs conformations ont été obtenus. Pour les aptamères, les structures cristallines ont permis d'observer les interactions nécessaires à la reconnaissance du ligand sont connues, mais pas de savoir quelles sont les étapes qui mènent à sa liaison. En effet, si le ligand n'est pas accessible au solvant, il est évident qu'il n'a pas pénétré directement dans le site de liaison dans sa conformation observable en cristallographie. Afin Pour véritablement comprendre le fonctionnement d'un riborégulateur, déterminer l'ordre dans lequel se replie chacune des parties de l'aptamère devient donc aussi important que d'identifier les nucléotides responsables de la reconnaissance du ligand.

Afin d'avoir une idée de l'ordre de repliement du riborégulateur, des études de spectroscopie en mode molécule unique ont été réalisées. Ces études ont permis de déterminer un patron de repliement hiérarchique selon lequel la P2 et la P3 se forment en premier suivies par l'interaction entre la boucle L2 et la boucle L3. À ce moment, l'aptamère possède un site de

liaison du ligand compétent, ce qui inclut probablement quelques interactions particulières à la jonction trois voies. La tige P1 est la dernière structure à se former. En absence de ligand, l'aptamère compétent dans lequel la P1 n'est pas formée est la forme la plus abondante alors qu'en présence de ligand, la conformation complètement repliée est majoritaire (Greenleaf, Frieda et al. 2008).

Plusieurs groupes de recherche ont opté pour l'utilisation de techniques exploitant la fluorescence afin d'améliorer les connaissances sur les riborégulateurs. Une revue de littérature publiée par St-Pierre et al (St-Pierre, McCluskey et al. 2014) récapitule les principales techniques de fluorescences utilisées au cours des dernières années. Cet article permet de constater l'importance de ces techniques dans les études sur la dynamique de repliement des ARN, plus particulièrement des riborégulateurs. La revue de littérature dont il est question se retrouve, pour consultation, en annexe de cette thèse sous le titre *Fluorescence tools to investigate riboswitch structural dynamics*. Cet article est très complet, c'est pourquoi seulement les techniques utilisées au cours des travaux menés pour cette thèse seront présentées dans les prochaines sections.

3.4 Études des riborégulateurs purines par fluorescence

La fluorescence est un phénomène physique permettant la production de lumière. Le diagramme de Jablonski présenté à la figure 9 permet de bien comprendre ce phénomène. Au départ, une source d'énergie fait monter un électron d'une molécule fluorescente à un niveau

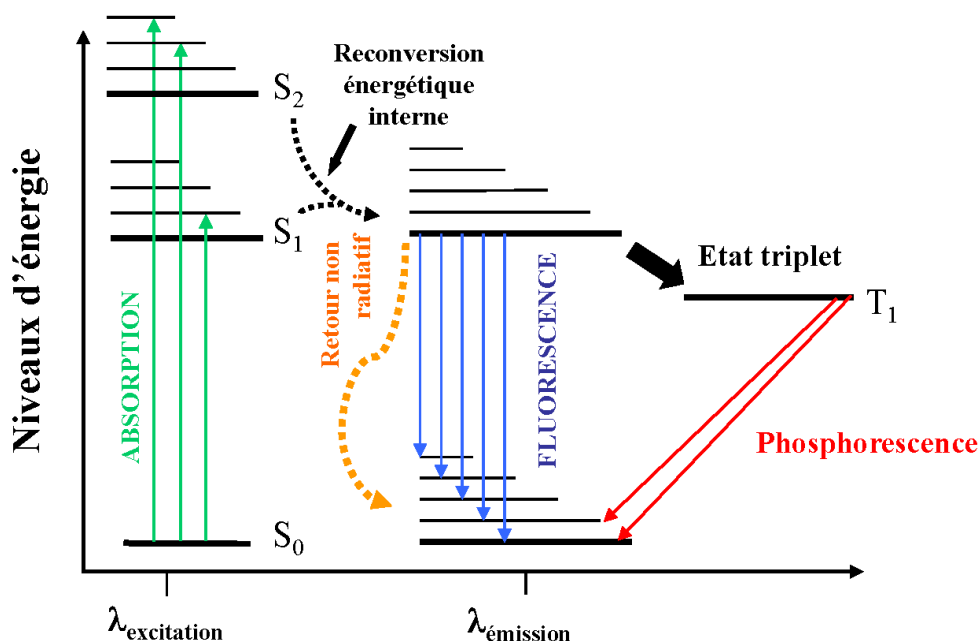


Figure 9 Diagramme de Jablonski expliquant la fluorescence et la phosphorescence.

Une fois excité, un électron d'une molécule fluorescente passe à un niveau d'énergie supérieur (S_1 ou S_2). Chaque niveau d'énergie possède plusieurs états vibrationnels qui peuvent être occupés par l'électron. Les états vibrationnels sont représentés par les lignes plus fines. La reconversion énergétique interne est un phénomène au cours duquel l'électron se retrouve au niveau d'énergie supérieur le plus stable suite à une perte d'énergie, généralement sous forme de chaleur. Un électron à un niveau d'énergie supérieur doit spontanément retourner à son état le plus stable (S_0). La perte d'énergie nécessaire au retour à l'état S_0 peut se faire par perte de chaleur (retour non radiatif) ou par émission d'un photon (fluorescence). Si le spin de l'électron n'est pas favorable au retour à l'état S_0 , l'électron se retrouve alors dans un état triplet (T_1). L'électron quittera l'état T_1 lorsque son spin redeviendra favorable émettant par le fait même un photon (phosphorescence).

d'énergie supérieur (S_1). En fonction de la quantité d'énergie fournie à l'électron, celui-ci peut atteindre différents niveaux d'énergie plus élevés (S_{1V1} , S_{1V2} , S_2 , etc.). Cependant, des événements d'interconnexions, qui se manifestent sous forme de perte de chaleur, ramènent habituellement l'électron au niveau S_1 . Spontanément, l'électron doit passer du niveau S_1 à son niveau d'énergie original (S_0). L'énergie perdue par l'électron lors de son retour à l'état le plus stable est émise sous forme de photon. Ce phénomène se produit en moins de 10^{-7} secondes après l'excitation. Principalement en raison des pertes sous forme de chaleur, le photon émis par fluorescence est toujours de plus faible énergie par rapport à la source permettant l'excitation initiale. Concrètement, cela signifie que si l'excitation se fait avec une source lumineuse, la longueur d'onde d'excitation sera toujours plus petite que la longueur d'onde d'émission.

La 2-aminopurine

L'existence d'un analogue fluorescent des purines, la 2-aminopurine (2AP), facilite l'étude des riborégulateurs liant les purines. Cette molécule fluorescente possède une longueur d'onde d'excitation optimale de 315 nm et émet à une longueur d'onde d'émission de 370 nm. Cette base peut s'apparier avec l'uracile et avec la cytosine ce qui en fait à la fois un analogue de l'adénine et de la guanine dans les structures d'ARN. De plus, cette molécule perd sa capacité d'émettre de la fluorescence lorsqu'elle interagit avec d'autres nucléotides. Par exemple, la 2AP peut être reconnue par l'aptamère reconnaissant les purines et cette liaison empêche l'émission de lumière par la 2AP. Il a été démontré que la constante de dissociation de la 2AP avec l'aptamère est semblable à celle de l'adénine (St-Pierre, McCluskey et al. 2014). En mesurant la variation de fluorescence, il est possible de déterminer l'efficacité de liaison de l'aptamère. Cette méthode a été utilisée pour déterminer l'impact de mutations sur la fonctionnalité de l'aptamère et ainsi identifier les nucléotides importants à son fonctionnement (St-Pierre, McCluskey et al. 2014). En plus d'être utilisée comme ligand, la 2AP a été incorporée chimiquement dans la séquence de plusieurs riborégulateurs afin de suivre les

changements de conformation des ARN (Heppell, Mulhbachter et al. 2009, St-Pierre, McCluskey et al. 2014). Ces techniques permettent d'obtenir de l'information sur l'organisation locale de la structure, mais fournissent peu de renseignements sur les interactions longues distances que possède un ARN. Heureusement, d'autres méthodes complémentaire permettent d'acquérir des précisions sur l'organisation globale de l'ARN.

Le transfert d'énergie par résonance de type Förster.

Le transfert d'énergie par résonance de type Förster, aussi appelé FRET de l'acronyme anglais pour *Förster Resonance Energy Transfer*, est utilisé depuis plusieurs années pour l'étude des macromolécules. La technique FRET exploite la capacité d'une molécule fluorescente à transmettre son énergie à une autre molécule fluorescente compatible. Ce transfert d'énergie est intimement relié à la distance entre les deux fluorophores selon la formule mathématique suivante :

$$E_{\text{FRET}} = (1 + [R_0/R]^6)^{-1}$$

Dans cette formule, E_{FRET} représente l'efficacité de transfert, R indique la distance entre le centre des deux fluorophores, et R_0 correspond à la constante de Förster pour le couple de fluorophores utilisé. La constante de Förster est égale à la distance entre le centre des deux fluorophores lorsque E_{FRET} est égale à 0,5. La constante de Förster pour la couple cyanine 3 (Cy3) et cyanine 5 (Cy5) est de $R_0 = 56 \text{ \AA}$ (Roy, Hohng et al. 2008). Avec cette formule, il est donc possible d'estimer la distance entre deux molécules fluorescentes. La figure 10 est une représentation graphique de cette formule et montre comment le E_{FRET} varie en fonction de la distance entre les deux fluorophores. Enfin, le phénomène de FRET est observable sur des distances allant de 0 à 120 \AA , ce qui est idéal pour l'étude d'interaction entre les macromolécules (Roy, Hohng et al. 2008). Les fluorophores peuvent être facilement attachés

aux extrémités 5' et 3' de l'ARN aussi bien qu'à un nucléotide à l'intérieur de la molécule. La conjugaison se fait en utilisant un nucléotide possédant un groupement chimique amine (-NH₂) à la position 5 de la base thymidine. Ce nucléotide est placé à la position désirée pour l'incorporation du fluorophore lors de la synthèse chimique de l'ARN. Le fluorophore quant à lui contient un groupement NHS-ester qui lui permettra de se lier spontanément et spécifiquement au groupement amine. De cette façon, les fluorophores sont insérés précisément aux endroits choisis dans la molécule d'ARN. La technique FRET a été utilisée à maintes reprises et avec succès pour étudier la structure des riborégulateurs (St-Pierre, McCluskey et al. 2014). Par contre, quelques informations capitales sur le repliement de l'ARN peuvent rester invisibles au FRET.

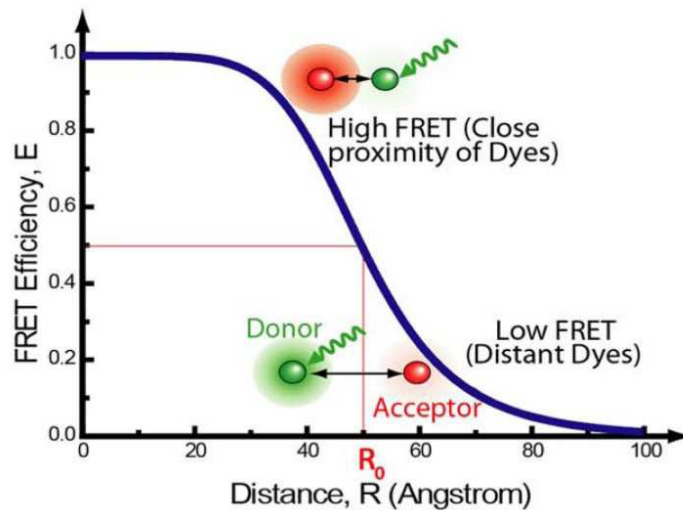


Figure 10 Tirée de (Roy, Hohng et al. 2008). Efficacité de FRET en fonction de la distance des fluorophores.

Efficacité de FRET (E) en fonction de la distance entre les fluorophores mesurée en Angstrom (R). Le fluorophore donneur est directement excité par un laser et émet de la fluorescence ou transmet son énergie au fluorophore accepteur en fonction de la proximité entre les deux. Lorsque $R = R_0$, $E = 0,5$. À plus faible distance $E > 0,5$ et *vice versa* selon la fonction montrée en bleu.

En effet, en FRET, comme pour beaucoup d'autres techniques biochimiques, le signal de toutes les molécules est comptabilisé en même temps, ce qui entraîne donc l'observation d'une valeur moyenne. Ceci implique qu'il est difficile de déterminer si les molécules ont toutes la même valeur de FRET ou s'il existe différentes populations de molécules. Pour la même raison, il est impossible de savoir si les molécules changent de conformation lors de la prise de donnée. Comme il est difficile de synchroniser le repliement des molécules d'ARN, la technique de FRET en mode molécules uniques (sm-FRET de l'anglais *Single Molecule FRET*) a été mise au point (Ha, Enderle et al. 1996). Cette technique mise sur l'observation individuelle des molécules. Ainsi, les molécules n'ont pas besoin d'être synchronisées et si plusieurs états peuvent exister en même temps sur différentes molécules, il est possible de les observer. Étant donnée l'utilisation abondante faite de cette technique lors des travaux présentés dans cette thèse, une description détaillée du fonctionnement du sm-FRET serait pertinente. Cependant, il n'est pas nécessaire de comprendre tous les détails associés à cette technique pour apprécier les informations que l'on peut en tirer. De plus, cette description serait très longue et n'est pas directement reliée au sujet de cette thèse. C'est pourquoi les lecteurs intéressés à en savoir plus à ce sujet sont invités à consulter une excellente revue de littérature intitulée *A practical guide to single-molecule FRET* (Roy, Hohng et al. 2008)

Des expériences de sm-FRET ont permis de montrer que lors de la liaison du ligand au riborégulateur purine, les tiges P2 et P3 se rapprochent l'une de l'autre. Ce rapprochement dépend de la présence d'ions magnésium. Il semble donc que l'aptamère soit préreplié et prêt à recevoir la purine qui stabilisera la structure de l'aptamère. Il semble aussi exister un intermédiaire de repliement pour l'aptamère du riborégulateur associé au gène *add* de *V. vulnificus* (Lemay, Penedo et al. 2006). La figure 11 présente un résultat obtenu lors d'une expérience de sm-FRET qui démontre l'existence de l'intermédiaire de repliement. Il est intéressant de noter que la majorité des molécules observées lors de cette expérience transitaient par la forme intermédiaire. L'existence de cette conformation ne fait alors pas de doute. Cependant, l'existence de cette structure est transitoire, ceci limite les expériences pouvant être réalisées afin de caractériser la conformation intermédiaire. Il a quand même été

proposé que la structure de l'intermédiaire corresponde à l'empilement des tiges P1 et P3 juste avant que la boucle L2 de la tige P2 n'interagisse avec la boucle L3 de la tige P3.

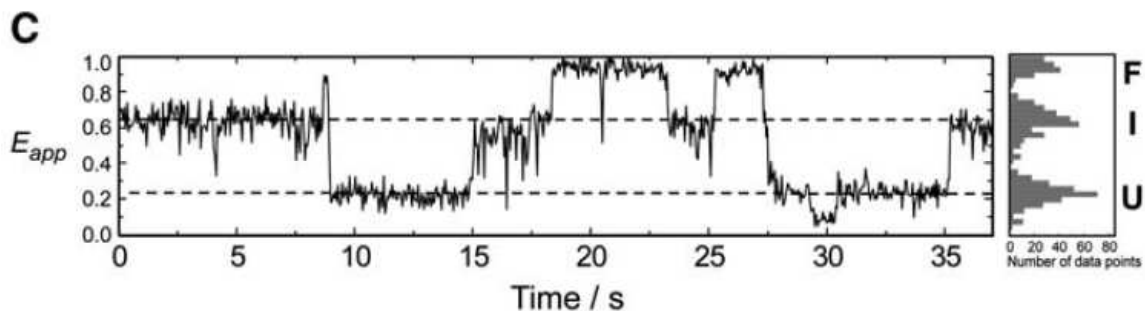


Figure 11 Tirée de (Lemay, Penedo et al. 2006). eFRET en fonction du temps pour un aptamère liant l'adénine

Le vecteur utilisé pour obtenir ce résultat est un aptamère liant l'adénine possédant les fluorophores sur les boucles L2 et L3. Tracé du signal FRET (E_{app}) en fonction du temps pour une seule molécule d'aptamère liant l'adénine en présence de 0,02 mM Mg^{2+} . Plusieurs interconversions entre les formes non repliée (U, $E_{app} = 0,25$), intermédiaire (I, $E_{app} = 0,5$) et repliée (F, $E_{app} = 0,9$) sont visibles.

Il est important de mentionner qu'un tel intermédiaire n'a pas été observé à la suite d'expériences similaires sur l'aptamère du riborégulateur *xpt-pbuX* (Brenner, Scanlan et al.). Ceci laisse supposer que le mécanisme de repliement n'est pas exactement le même pour tous les riborégulateurs liant les purines. L'observation de l'intermédiaire de repliement est à l'origine des travaux réalisés dans cette thèse.

4. Objectifs du projet de recherche

Le but des travaux présentés dans cette thèse est de préciser le patron de repliement de l'aptamère liant l'adénine associé au gène *add* chez *V. vulnificus*. Ce but peut paraître surprenant étant donné que la classe des riborégulateurs purines est la plus étudiée depuis la

découverte de ce type de mécanisme de régulation. En effet, la structure de l'aptamère et les mécanismes de régulation associés à ces riborégulateurs sont bien connus et plusieurs études portant sur le repliement de cet ARN ont été réalisées (Porter, Marcano-Velazquez et al. 2014). Cependant, ces études ont été réalisées uniquement sur l'aptamère lié à son ligand. Peu d'informations sont disponibles sur la conformation non liée et sur les conformations intermédiaires de nature transitoire que peuvent adopter cet ARN. Par exemple, il est évident que l'aptamère adopte une conformation prête à accueillir le ligand en présence de magnésium et que cette conformation est différente de celle observée lors d'expériences de cristallographie. Pourtant, le niveau de préorganisation de cette conformation reste indéterminé. De plus, seul l'ordre de formation des structures secondaires a été déterminé (Porter, Marcano-Velazquez et al. 2014). Pour l'instant, il est impossible de savoir si le cœur de l'aptamère se replie avant la formation de l'interaction boucle-boucle par exemple. Le niveau de préorganisation nécessaire à la reconnaissance du ligand reste lui aussi inconnu. L'identification en sm-FRET d'une conformation intermédiaire de l'aptamère du riborégulateur associé au gène *add* chez *V. vulnificus* a ouvert la porte à la détermination de cette conformation. En effet, s'il est possible d'isoler et de caractériser cet intermédiaire, il sera possible de déterminer sa structure. Il sera aussi possible de démontrer si la conformation intermédiaire peut lier le ligand, ou si la liaison du ligand entraîne la formation de l'état intermédiaire suivit de l'état replié. Cette caractérisation fournira des informations sur un niveau de préorganisation précédant la formation de l'état replié de l'aptamère.

Le principal objectif du chapitre 1 a été d'isoler la conformation intermédiaire de l'aptamère *add* de *V. vulnificus*. Cette isolation a été faite en utilisant la technique de sm-FRET puisqu'il s'agit de la technique ayant permise l'identification de l'intermédiaire en premier lieu. De plus, c'est l'une des rares techniques disponibles permettant de suivre le repliement de molécules d'ARN individuelles dans le temps. Plusieurs avenues ont été envisagées pour permettre l'isolation de l'intermédiaire. Par exemple, l'auteur de cette thèse a participé à une étude dans laquelle l'urée a été utilisée comme agent dénaturant afin d'isoler la conformation intermédiaire de l'aptamère *add*. Cette étude a été faite en collaboration avec le

laboratoire du docteur Juan Carlos Penedo de l'université de St-Andrews et a été publiée dans la revue scientifique *Nucleic Acid Research* (Shaw, St-Pierre et al. 2014). Par contre, la voie retenue ici pour isoler l'intermédiaire a été de muter les nucléotides de la boucle L2 pour empêcher l'interaction avec la L3. De cette façon, les molécules ne devraient pas pouvoir adopter la conformation repliée et l'ajout de magnésium devrait permettre d'isoler la conformation intermédiaire. Cette hypothèse de travail est basée sur les travaux réalisés sur l'aptamère *add* qui suggèrent que la structure de l'intermédiaire de repliement est caractérisée par un empilement des tiges P1 et P3 sans que les boucles L2 et L3 n'interagissent entre elles (Lemay, Penedo et al. 2006). Une fois le premier objectif atteint, il est devenu intéressant d'étudier l'organisation globale des trois conformations identifiées pour cet aptamère (Non-replié, Intermédiaire et Replié). Pour ce faire, des fluorophores ont été positionnés sur les différentes tiges de l'aptamère afin de déterminer comment elles interagissent l'une par rapport à l'autre. Ces informations ont été obtenues pour chaque conformation isolée. Le premier chapitre se conclut avec des expériences de SHAPE donnant un aperçu des interactions secondaires pour chaque conformation de l'aptamère.

Les données présentées au chapitre 1 ont permis d'observer l'organisation globale des différentes conformations de l'aptamère *add*. Les résultats obtenus lors de cette étude permettent de supposer que les nucléotides du cœur de l'aptamère, particulièrement ceux se retrouvant dans la région J2-3, dirigent le repliement de l'aptamère. Selon la structure cristalline de l'aptamère il serait possible de placer un fluorophore sur un nucléotide de la région J2-3 sans affecter de façon significative la structure de l'aptamère. Le principal objectif du chapitre 2 est donc d'observer le repliement du cœur de l'aptamère en temps réel. Pour ce faire, des expériences de sm-FRET ont été réalisées sur des aptamères dont les fluorophores se retrouvent dans la J2-3 et sur la P1.

CHAPITRE 1

Patron de repliement du riborégulateur *add* observé par FRET en mode molécule unique

Une première étude du repliement de l'aptamère *add* par sm-FRET a démontré que la structure de cet aptamère varie entre trois conformations nommées non repliée (U), intermédiaire (I) et repliée (F). La structure de l'état replié a facilement été associée à la conformation observée en cristallographie. Cependant, les structures des états U et I n'ont pas été déterminées bien qu'il a été proposé que la conformation U soit caractérisée par l'empilement coaxial des tiges P1 et P3 en absence de l'interaction boucle-boucle (Gilbert, Stoddard et al. 2006). Afin de tester cette hypothèse, l'interaction boucle-boucle a été empêchée en modifiant les nucléotides de la boucle L2 pour éviter la formation de paire de base de type Watson-Crick avec les nucléotides de la boucle L3. L'intermédiaire de repliement a ainsi été isolé et la première caractérisation de la structure de la conformation I été réalisée par sm-FRET et SHAPE.

Lors de cette caractérisation, il a été remarqué que le 2-6 diaminopurine (DAP), un analogue de l'adénine, pouvait interagir avec l'aptamère modifié et promouvoir le rapprochement des tiges P2 et P3. Ce résultat révèle que le ligand se lie à l'aptamère avant la formation de l'interaction boucle-boucle et qu'il peut même induire la formation de cette interaction. Ce résultat était inattendu étant donné les données disponibles dans la littérature qui suggèrent que l'interaction boucle-boucle est nécessaire à la formation d'un aptamère compétent à la liaison du ligand (Greenleaf, Frieda et al. 2008). Les résultats de ces travaux seront publiés dans la revue scientifique *Nucleic Acid Research*.

Pour cet article, Samuel Jacques a obtenu les résultats préliminaires en sm-FRET pour l'aptamère muté dans la boucle L2, Patrick St-Pierre a obtenu la majorité des résultats de sm-FRET pour l'aptamère muté et tous les résultats pour l'aptamère de type sauvage et Frédéric Picard-Jean a fait les expériences de SHAPE. L'analyse des résultats de SHAPE a été faite par Frédéric Picard-Jean et Patrick St-Pierre. La rédaction du manuscrit a été faite par Patrick St-Pierre et Daniel Lafontaine. Juan Carlos Penedo a permis la mise au point du sm-FRET.

Folding pathway of the *add* adenine riboswitch monitored using single-molecule FRET

Patrick St-Pierre¹, Samuel Jacques¹, Frédéric Picard-Jean¹, J. Carlos Penedo*^{2,3} and Daniel A. Lafontaine*¹

¹Department of Biology, Faculty of Science, RNA Group, Université de Sherbrooke, Sherbrooke, Québec J1K 2R1, Canada

²SUPA, School of Physics and Astronomy, University of St Andrews, St Andrews, Fife, KY16 9SS, UK

³Biomedical Sciences Research Complex, University of St Andrews, St Andrews, Fife, KY16 9SS, UK

*To whom correspondence should be addressed. Email: daniel.lafontaine@usherbrooke.ca and jcp10@st-andrews.ac.uk.

ABSTRACT

Riboswitches are regulatory elements that perform specific metabolite sensing. The *Bacillus subtilis* adenine riboswitch positively controls translation initiation upon adenine binding and is part of the purine-binding riboswitch class. Although a lot of work has been devoted to the characterization of riboswitch ligand recognition mechanisms, little is known regarding how the folding process of riboswitches is involved in ligand sensing. Here, by employing a single-molecule approach, we have characterized the folding pathway of the adenine aptamer by monitoring the global structure of the RNA as a function of magnesium and ligand cofactors. By introducing fluorophores at various positions within the aptamer, we were able to characterize in detail the folding pathway. Furthermore, we also find that the presence of a long-range tertiary structure, the loop-loop interaction, is not required for ligand recognition and for the adoption of a global structure similar to the one adopted by the wild-type molecule. Lastly, biochemical probing of an intermediate conformer of the folding pathway reveals how the formation of the loop-loop interaction influences the structure of the core domain. Our study suggests that transient intermediate conformers can be part of riboswitch folding pathways to participate in the formation of ligand binding site.

INTRODUCTION

Gene expression requires the assistance of numerous cofactors to provide the specificity and timing of regulation, all of which aiming to ensure cellular homeostasis. Accumulating evidence indicates that RNA folding is a central player in gene regulation, particularly by providing specific protein binding sites and by allowing enzymatic reactions to be performed. A striking example about the importance of RNA folding in gene expression was recently obtained with the discovery of riboswitches, which are regulatory elements mostly found in the 5' untranslated region (5' UTR) of bacterial messenger RNA (mRNA) (1). Such mRNA elements are involved in the control of gene expression by modulating their structure upon the binding of specific cellular metabolites that are often related to the controlled gene(s). The riboswitch-metabolite interaction is mediated by the aptamer domain, which exhibits a high degree of sequence conservation and structural complexity (2). Ligand binding also modulates the structure of the riboswitch expression platform that is directly implicated in the regulation of gene expression, either by modulating premature transcription termination or by sequestering the ribosome binding site (RBS) (1). In addition, riboswitches have been shown to modulate splicing (3, 4) and RNA degradation (5), and also to regulate gene expression in a trans-acting manner (6).

Riboswitches recognize a large variety of metabolites such as amino acids, nucleobases and cofactors. Among these, the purine-sensing riboswitches are particularly interesting due to their small size and relatively simple structure (7). Three different natural classes of purine riboswitches have been identified so far where each one exhibits high affinity toward adenine (8), guanine (9) or deoxyguanosine (10). The *Vibrio vulnificus* adenine-sensing *add* riboswitch positively modulates the translation of an adenine deaminase (Figure 1A), making it one of the rare riboswitches to increase gene expression upon ligand binding (11–13). In this regulation mechanism, the presence of adenine allows the Shine-Dalgarno (SD) and AUG start codon sequences to be accessible for ribosome binding, thereby promoting translation initiation

(Figure 1A). However, the absence of adenine results in the formation of a sequestrator stem-loop making inaccessible both SD and AUG sequences (Figure 1B), resulting in translation inhibition. As other purine riboswitches, the *add* aptamer is organized around a three-way helical junction in which three helices (P1, P2 and P3) are connected through a single-stranded core region (Figure 1C). Crystal structures of purine aptamers (11, 14) have revealed that the fold of the RNA is highly compact and that several interactions involving the bound ligand are performed to ensure high binding affinity and specificity (15). As often observed in three-way junctions, helices P1 and P3 are involved in the formation of a coaxial stacking interaction (Figure 1C). Additionally, the aptamer exhibits a long-range loop-loop tertiary interaction occurring between L2 and L3 loops (Figures 1B and 1C), which was observed to be important for ligand binding. While numerous X-ray crystallography studies have been done on the ligand-bound form of purine riboswitches, no ligand-free aptamer crystal is available.

Furthermore, although several biochemical techniques have been used to characterize the structure of purine aptamers, there is relatively little information about the folding pathway of purine aptamers and its role in ligand recognition. For instance, it is not clear to which extent the core of the aptamer needs to be preorganized to allow ligand binding (15).

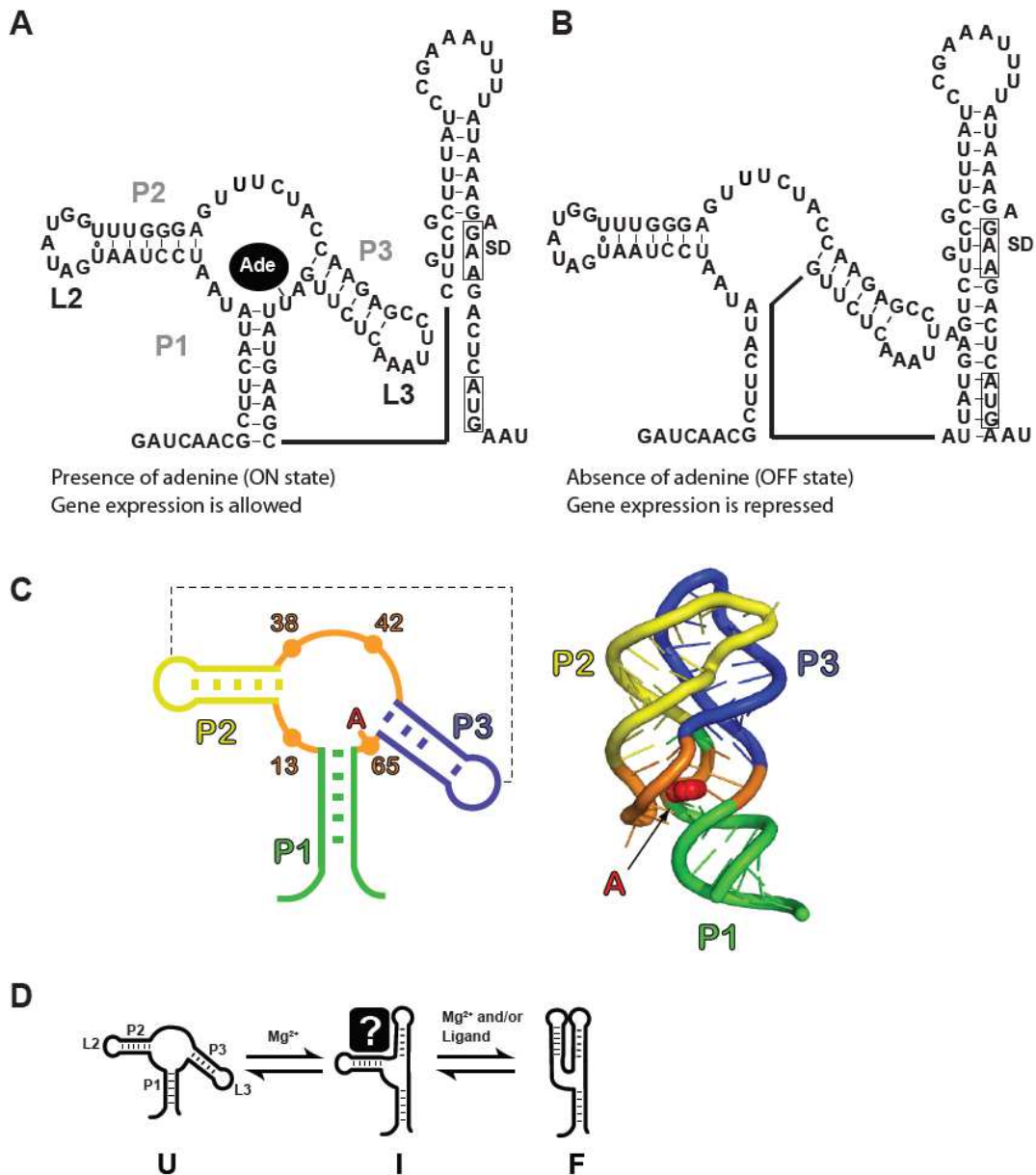


Figure 1 Secondary structures of the *Vibrio vulnificus* adenine riboswitch. (A) Secondary structure model of the ON conformer in which the Shine-Dalgarno (SD) and the AUG initiation codon (boxed) are accessible for translation initiation. (B) Secondary structure model of the adenine riboswitch in which the absence of adenine results in the repression of translation initiation. (C) Cartoon of the adenine aptamer showing the P1, P2 and P3 stems in both a secondary and tertiary structure representations. The four residues within the core region (13, 38, 42 and 65) are directly interacting with adenine in the crystal structure (11). The dotted line represent the L2-L3 loop-loop interaction. (D) Proposed folding pathway of the adenine aptamer according to a previous study (16). The folding pathway shows that the adenine aptamer adopts three different structural configurations corresponding to an unfolded (U), intermediate (I) and folded (F) structures. The actual configuration of the U and I structures are not known, but speculation (25) have been made regarding the I state based on our previous work (16).

Single-molecule experiments provide vital information by directly allowing the identification and correlation of folding intermediates. We have previously employed single-molecule Förster resonance energy transfer (sm-FRET) to study the folding pathway of the *add* riboswitch aptamer by measuring FRET occurring between fluorophores introduced in the L2 and L3 loops (16), effectively allowing to follow the formation of the loop-loop interaction. By monitoring individual molecules, we showed that the *add* aptamer exhibits great heterogeneity in folding and unfolding rates, both of which were significantly reduced upon adenine binding. Real-time FRET traces revealed that the aptamer fluctuates between unfolded (U), intermediate (I) and folded (F) states, all of which were observed within single aptamer molecules. The folded conformation was confidently associated to the crystallized structure bound to the ligand. Strikingly, the intermediate state was promoted by ligand binding suggesting that the intermediate is likely an obligate state along the folding pathway. Based on our study and available crystal structures, it was proposed by Gilbert and Batey that the intermediate state might be a partially folded structure in which the P1-P3 coaxial stack is formed but in which the loop-loop interaction does not occur (Figure 1D). However, despite the significant number of biochemical studies performed on purine aptamers, the nature of the intermediate still remains elusive.

Here, we have studied the folding pathway of the *add* aptamer by monitoring the aptamer conformational changes using sm-FRET and SHAPE (Selective 2'-hydroxyl acylation analyzed by primer extension) analysis. By measuring multiple FRET efficiencies reporting on key conformational changes, we have established the global folding pathway of the *add* aptamer as a function of magnesium ions and ligand cofactors. In addition to the natural sequence, we have used an aptamer variant in which the L2-L3 loop-loop interaction is disabled as a way to stabilize the intermediate conformation. In contrast to the wild-type, we found that the folding of the aptamer variant to be limited to the formation of the intermediate conformation. However, ligand binding to the L2 variant unexpectedly resulted in a conformation very similar to the ligand-bound wild-type aptamer, suggesting that ligand binding induced the docking of loops L2 and L3 even when corresponding sequences are not

compatible for intramolecular interactions. Using SHAPE analysis, the core region of the intermediate conformer was shown to differ from the unfolded and folded states, where several interactions were already present suggesting that a pre-folded architecture is required for ligand recognition.

MATERIAL and METHODS

Preparation of dual-labelled aptamer molecules

Vibrio vulnificus adenine aptamer molecules were assembled as previously reported (16). Briefly, fluorescent molecules were reconstituted by ligating two RNA molecules each carrying either a Cy3 or Cy5 fluorophore. The RNA strand corresponding to the 5' portion of the *add* aptamer carried a biotin at the 5' end to allow surface attachment and an amino linker nucleotide in the L2 loop for labeling. The RNA strand corresponding to the 3' end carried a phosphate group at the 5' end to allow ligation using T4 RNA ligase and an amino linker nucleotide in the L3 loop for labeling. RNA strands were labeled using NHS ester fluorophores (Fisher Scientific). The labeling and purification of RNA strands were performed as previously described (16). The oligonucleotides used to reconstitute the aptamer are the following (positions used for labeling are identified by underlined nucleotides and mutations in the L2 variant are shown in bold):

Strands used for vectors P2-P3:

WT 5': Biotin-GCGCGAGCGUUGUAUAAUCCUAAUGAUAUGGUUUGGGAGU

WT 3': PO₄-UUCUACCAAGAGCCUUAAACUCUUGAUUACAACGCUCGCGC

L2 5': Biotin-GCGCGAGCGUUGUAUAAUCCUAAUGAAUUCCUUUGGGAGU

L3 3': PO₄-UUCUACCAAGAGGGUAUAACUCUUGAUUACAACGCUCGCGC

Strands used for vector P1-P3:

WT 5': Biotin-GCGCGAGCGUUGUAUAAUCCUAAUGAUUAGGUUUGGGAGU

WT 3': PO₄-UUCUACCAAGAGCCUUAAACUCUUGAUUACAACGCUCGCGC

L2 5': Biotin-GCGCGAGCGUUGUAUAAUCCUAAUGAAUCCUUUGGGAGU

Strands used for vector P1-P2:

WT 5': Biotin-GCGCGAGCGUUGUAUAAUCCUAAUGAUUUGGUUUGGGAGU

WT 3': PO₄-UUCUACCAAGAGCCUAAACUCUUGAUUACAAUGCUCGCGC

L2 5': Biotin-GCGCGAGCGUUGUAUAAUCCUAAUGAAUUCCUUUGGGAGU

Single-molecule FRET analysis

Sm-FRET analysis was essentially performed as previously reported (16). Briefly, molecules were isolated on quartz microscope slides using successive treatment with 1 mg/mL biotinylated BSA (Sigma) and 0.2 mg/mL streptavidin (Invitrogen). After treatment with streptavidin, 50-250 pM of fluorophore-labeled aptamers were added to the slide. Fluorescence data at donor and acceptor wavelengths were acquired from single molecules using a total internal fluorescence (TIR) microscope coupled with a back-illuminated EMCCD camera (iXon, Andor Technology). Unless indicated, the data were acquired with an integration time of 50 ms. The buffer used for all experiments is derived from Aitken and *al.* (17) and contains 10 mM Tris-HCl pH 8.0, 10 mM NaCl, 1 μM protocatechuic acid (Fisher

Scientific), 0.25 μ M protocatechuic dioxygenase (Sigma) and 0.5 mM 6-Hydroxy-2,5,7,8-tetramethyl-chroman-2-carboxylic acid (TROLOX) (Sigma). The specified magnesium and 2,6-diaminopurine (DAP, Sigma) concentrations were directly added to this buffer. Measurements were obtained at 22°C. Single-molecule FRET efficiency (E_{app}) after background correction was determined by: $E_{app} = I_A/(I_A+I_D)$ where I_A and I_D are the fluorescence intensities of Cyanine 5 (Cy5) and Cyanine 3 (Cy3), respectively. The value of E_{app} determined this way closely matches the true efficiency of transfer since the quantum yield of Cy3 and Cy5 are very close. Raw data transformation and analysis was made using laboratory-written analysis routine developed in IDL 6.0 and MATLAB 7. For each tested condition, population histograms were obtained by averaging the E_{app} of the first 10 frames of 1000 molecules randomly selected after manual scanning against photobleaching and blinking. Gaussian models were fitted on the histograms to identify the aptamer populations. Single molecule traces show E_{app} values measured over time as previously described (16).

SHAPE chemical probing of adenine aptamers

Transcripts corresponding to the wild-type and L2 mutant aptamers were produced as previously described (18, 19). Briefly, SHAPE reactions were prepared by mixing 1 pmol of purified aptamers resuspended in a 2:1 mixture of 0.5X TE buffer: 3.3X folding buffer (333 mM K-HEPES, pH 8.0, 333 mM NaCl) and the required concentration of MgCl₂ and DAP. Samples were reacted with *N*-methylisatoic anhydride (NMIA, Invitrogen) (18). Reverse transcription reactions were performed on NMIA-reacted RNA and products were separated on 5% denaturing polyacrylamide gels.

RESULTS

Monitoring the loop-loop interaction of the *add* aptamer using sm-FRET

To study the conformational changes occurring in the *add* aptamer, we analyzed the dynamics of the aptamer using sm-FRET. Dual-labeled aptamers were attached to a quartz slide via biotin and were visualized using a wide-field total internal reflection (TIR) microscope. We first analyzed the wild-type molecule in the presence of 10 mM NaCl in which only the secondary structure is predicted to form (16). As expected, single-molecule histograms showed the presence of a major population (84%) centered at an apparent FRET efficiency (E_{app}) of ~ 0.35 (Figure 2A, top panel), corresponding to the unfolded (U) state in which the loop-loop interaction is not taking place. However, a small fraction (16%) of the aptamer population was also found to adopt the folded state (F state, $E_{app} \sim 0.92$), consistent with the formation of the loop-loop interaction (16). Upon the addition of magnesium ions, the majority of molecules folded into the F state (Figure 2A, middle panel), in agreement with the loop-loop interaction being stabilized upon Mg^{2+} binding (16). To characterize the ligand binding activity of the aptamer, we used 2,6-diaminopurine (DAP) that was previously shown to exhibit ~ 30 -fold more affinity toward the adenine-sensing aptamer (8). The addition of DAP did not show any significant effect on the aptamer population (Figure 2A, lower panel), consistent with the loop-loop interaction not being altered upon ligand binding.

The transient nature of the intermediate conformation (Figure 1D)(16) makes difficult to characterize its folding and structural properties. Given that the E_{app} value (~ 0.60) of the intermediate conformer strongly suggests that the loop-loop interaction is not formed (16), we reasoned that the intermediate structure should be favored by disabling the loop-loop interaction. We thus introduced mutations within the L2 loop (U25A, A26U, G28C and G29C) to prevent the formation of the loop-loop interaction. Sm-FRET analysis of the L2 mutant

revealed that the unfolded state is very similar to the wild-type in presence of NaCl only (Figure 2B, top panel). The absence of the F state in this experimental condition is consistent

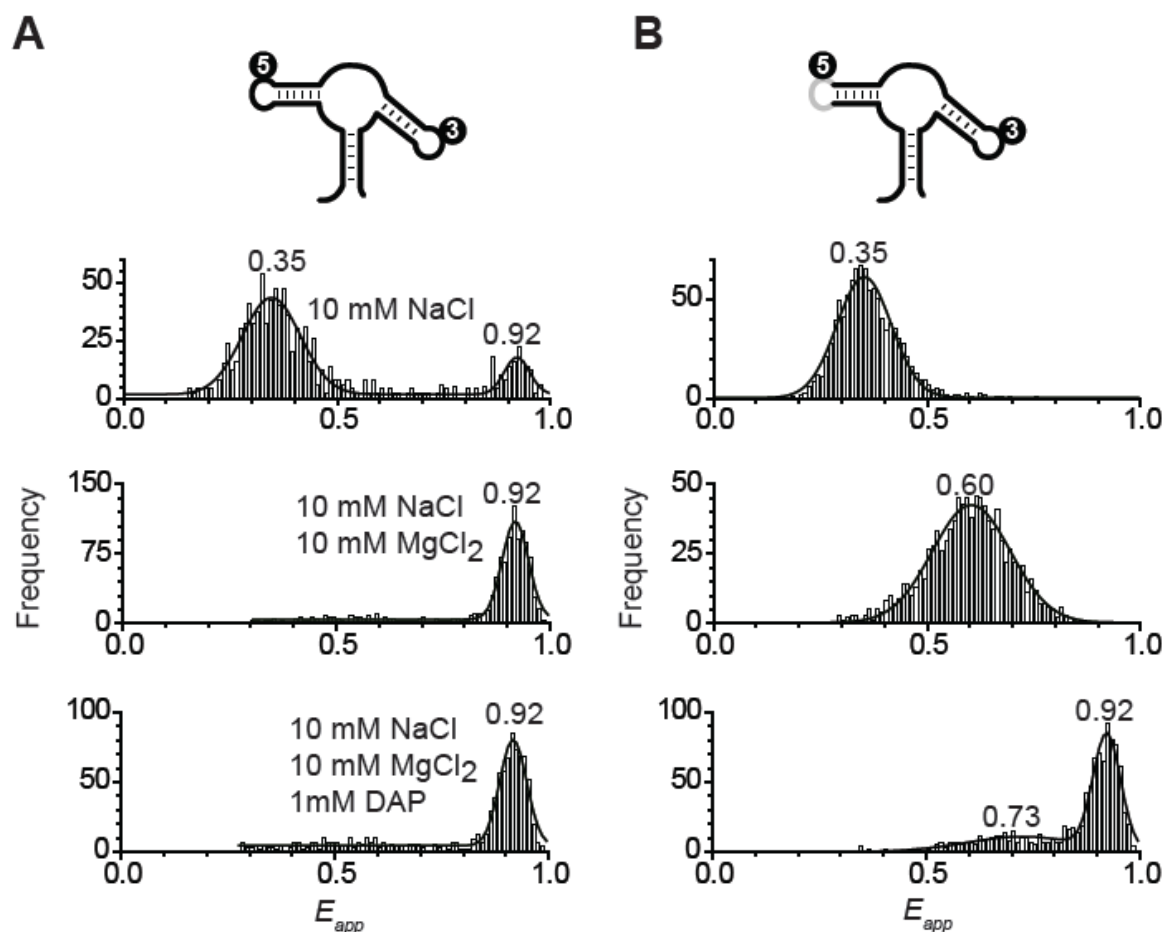


Figure 2 Sm-FRET analysis for the vector P2-P3 of the adenine aptamer. (A) Sm-FRET experiments performed for the wild-type aptamer. The cartoon represents the dual-labeled aptamer used for these studies. Cy3 (3) and Cy5 (5) fluorophores are indicated. Population histograms for the wild-type aptamer in presence of 10 mM NaCl (top panel), 10 mM NaCl and 10 mM MgCl₂ (middle panel) and 10 mM NaCl, 10 mM MgCl₂ and 1 mM DAP (lower panel). Individual time traces were manually filtered to remove the contribution from fluorophore blinking and photobleaching events. E_{app} values determined by fitting analysis for each population are indicated. (B) Sm-FRET experiments performed for the L2 mutant aptamer. The cartoon represents the dual-labeled aptamer used for these studies. Cy3 (3) and Cy5 (5) fluorophores are indicated. Population histograms for the mutant aptamer were obtained in the same conditions as in A. Individual time traces were manually filtered to remove the contribution from fluorophore blinking and photobleaching events. E_{app} values determined by fitting analysis for each population are indicated.

with the L2 mutant not supporting the formation of the loop-loop interaction. However, in contrast to the wild-type aptamer adopting the F state in the presence of 10 mM Mg^{2+} ions (Figure 2A, middle panel), the L2 mutant showed a single population exhibiting an E_{app} of ~ 0.60 in the same condition suggesting that the binding of Mg^{2+} ions to the L2 mutant stabilizes the presence of the intermediate conformer. A magnesium ions titration showed that the adoption of the intermediate conformer was gradually obtained (Supplementary Figure 1) from the shifting of the U population, indicating that a fast interconversion exists between both U and I states. Strikingly, the addition of 1 mM DAP induced a major proportion of aptamer molecules to fold into the F state (Figure 2B, lower panel), indicating that DAP can bind to the intermediate conformer to promote the folded state conformation. Since the high FRET state is observed with the L2 mutant, it suggests that the loop-loop interaction is not required for ligand binding. Similar results were obtained when L3 loop nucleotides were mutated (Supplementary Figure 2), consistent with the Mg^{2+} -dependent stabilization of the intermediate being directly related to the disruption of the loop-loop interaction.

Monitoring the formation of the P1-P3 helical stacking unit

The *add* aptamer crystal structure shows that the ligand binding site is located adjacently to a helical stacking unit formed by stems P1 and P3 (11). To determine the role of the helical stack in the *add* folding pathway, fluorophores were incorporated in stems P1 and P3 to directly monitor the formation of the P1-P3 helical stacking unit. In the presence of NaCl, the wild-type aptamer adopted a single population exhibiting an E_{app} of ~ 0.64 (Figure 3A, top panel). Addition of Mg^{2+} ions reproducibly shifted the aptamer population to an E_{app} of ~ 0.68 (Figure 3A, middle panel). However, the FRET value was not significantly changed by adding 1 mM DAP (Figure 3A, lower panel). Given that FRET efficiencies are relatively similar in the presence of NaCl and Mg^{2+} ions, it makes relatively difficult to determine whether the P1-P3 helical stack is formed upon Mg^{2+} binding or whether it is already stacked in the presence of NaCl. Such very similar FRET efficiencies did not allow both populations to be

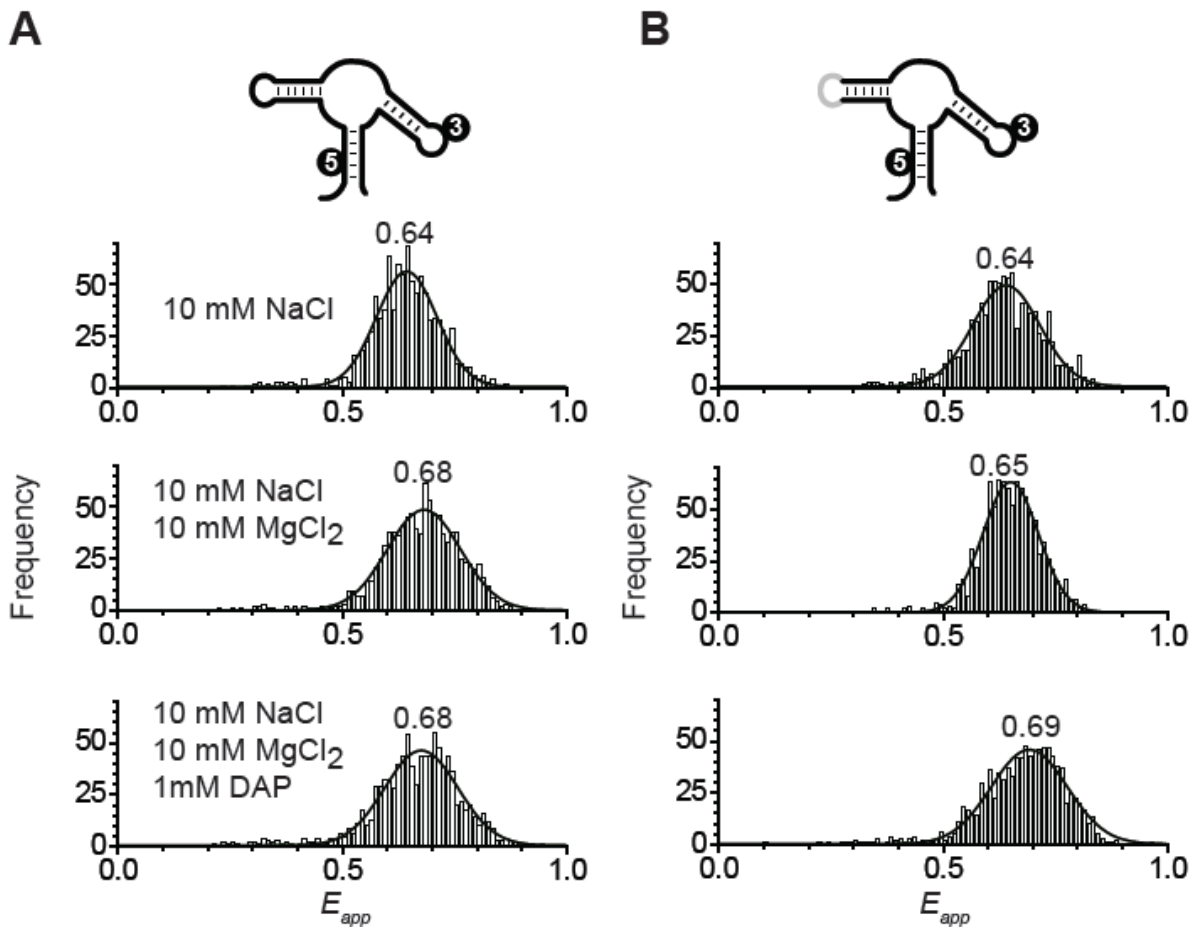


Figure 3 Sm-FRET analysis for the vector P1-P3 of the adenine aptamer. (A) Sm-FRET experiments performed for the wild-type aptamer. Population histograms for the wild-type aptamer in presence of 10 mM NaCl (top panel), 10 mM NaCl and 10 mM MgCl₂ (middle panel) and 10 mM NaCl, 10 mM MgCl₂ and 1 mM DAP (lower panel). Individual time traces were manually filtered to remove the contribution from fluorophore blinking and photobleaching events. E_{app} values determined by fitting analysis for each population are indicated. (B) Sm-FRET experiments performed for the L2 mutant aptamer. Population histograms for the mutant aptamer were obtained in the same conditions as in A. Individual time traces were manually filtered to remove the contribution from fluorophore blinking and photobleaching events. E_{app} values determined by fitting analysis for each population are indicated.

deconvoluted from time traces. The formation of the P1-P3 helical stack was next monitored in the context of the L2 mutant. Very similar results to the wild-type molecule were obtained in all tested conditions (Figure 3B), suggesting that the P1-P3 helical stack is not significantly affected by the formation of the loop-loop interaction. Considering the small changes in E_{app} caused by the addition of DAP, it most probably indicates that stems P1 and P3 are stacked in all conformations of the *add* aptamer as observed in the crystal structure (11). The E_{app} change from ~ 0.65 to ~ 0.69 (Figure 3B) is probably caused by a compaction of the core domain of the aptamer that occurs upon ligand binding, as suggested by in-line probing data (8). The absence of a large FRET change when monitoring the P1-P3 helical stack suggests that the conformational positioning of the P2 stem is the major cause of FRET changes observed when monitoring the loop-loop interaction (Figure 2A).

The P1-P2 conformational changes are highly dependent on cofactor binding

To obtain a complete view of the *add* aptamer folding pathway, we next introduced fluorophores in the P1 stem and in the P2 loop, thereby allowing to monitor any conformational changes involving helices P1 and P2. In contrast to all fluorescent constructs used in this study, the P1-P2 vector is the only one for which structural transitions were detected in both in the presence of Mg^{2+} ions and ligand. When monitoring the FRET efficiency in a background of 10 mM NaCl, a unique aptamer population exhibiting an E_{app} of ~ 0.57 was observed (Figure 4A, top panel). A significant decrease in efficiency ($E_{app} \sim 0.37$) was observed in the presence of Mg^{2+} ions, a condition allowing the loop-loop interaction to form (Figure 4A, middle panel). The addition of 1 mM DAP further reduced the FRET value to ~ 0.31 (Figure 4A, lower panel), indicating that ligand binding results in a conformational change altering the positioning of the P1 stem relative to the loop-loop interaction. A ligand titration revealed that the aptamer population was shifted from the high FRET state to the low FRET state at ligand saturation (Supplementary Figure 3), consistent with fast interconversion between populations.

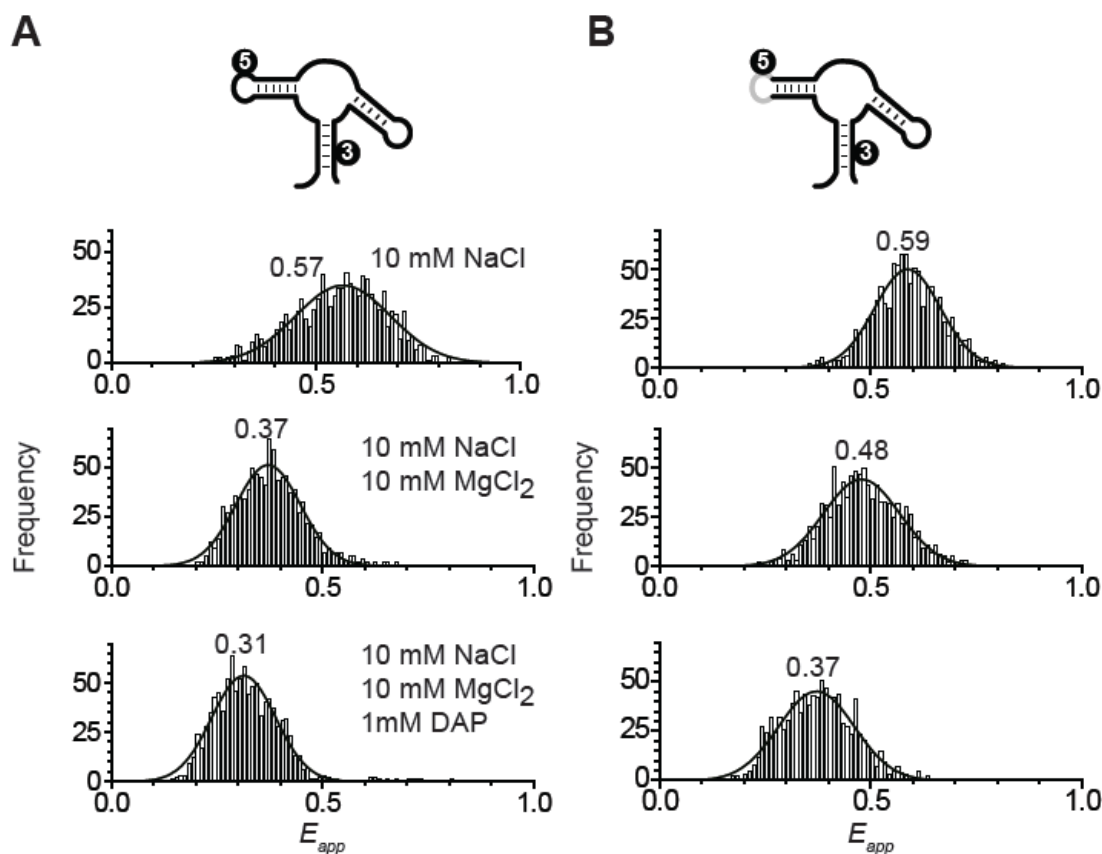


Figure 4 Sm-FRET analysis for the vector P1-P2 of the adenine aptamer. (A) Sm-FRET experiments performed for the wild-type aptamer. Population histograms for the wild-type aptamer in presence of 10 mM NaCl (top panel), 10 mM NaCl and 10 mM MgCl₂ (middle panel) and 10 mM NaCl, 10 mM MgCl₂ and 1 mM DAP (lower panel). Individual time traces were manually filtered to remove the contribution from fluorophore blinking and photobleaching events. E_{app} values determined by fitting analysis for each population are indicated. (B) Sm-FRET experiments performed for the L2 mutant aptamer. Population histograms for the mutant aptamer were obtained in the same conditions as in A. Individual time traces were manually filtered to remove the contribution from fluorophore blinking and photobleaching events. E_{app} values determined by fitting analysis for each population are indicated.

We next monitored the P1-P2 conformational change in the context of the L2 mutant. In the presence of NaCl only, one population was detected with an E_{app} of ~ 0.59 (Figure 4B, top panel). The addition of Mg^{2+} ions decreased the FRET efficiency to a value of ~ 0.48 (Figure 4B, middle panel). The presence of 1 mM DAP further reduced the E_{app} value to ~ 0.37 (Figure 4B, lower panel), suggesting that ligand binding caused the L2 mutant to adopt a different conformation, similarly to the wild-type aptamer (Figure 4A, lower panel).

Ligand binding does not allow aptamer folding in absence of Mg^{2+} ions

Since our results show that the Mg^{2+} -dependent loop-loop interaction is not required for ligand binding (Figure 2B, lower panel), we next determined whether the presence of Mg^{2+} ions were necessary for ligand binding to occur. For these experiments, we employed the L2 mutant aptamer for which a clear ligand binding effect can be observed based on the E_{app} value (Figure 2B, lower panel). When incubating the L2 mutant in a buffer containing 10 mM NaCl in the absence or presence of 1 mM DAP, no significant change was detected for both E_{app} values (Figure 5, compare the top panels of A and B). This is suggesting that the L2 aptamer mutant is not able to bind DAP in the absence of Mg^{2+} ions. However, when these experiments were repeated in similar conditions but which contained 250 μM Mg^{2+} ions, a significant proportion of aptamers were found to adopt the F state ($E_{app} \sim 0.9$) only in the presence of DAP (Figure 5, compare the middle panels of A and B). The U and I aptamer populations were also detected in the absence (E_{app} U state ~ 0.37 and I state ~ 0.55) and presence of DAP (E_{app} U state ~ 0.28 and I state ~ 0.49), as expected from our previous study (16). Lastly, increasing the Mg^{2+} ions concentration at 1 mM yielded the L2 mutant aptamer to adopt a higher proportion the F state in presence of DAP, consistent with the importance of Mg^{2+} ions for ligand binding. Together, these observations suggest that the preorganization of the core driven by Mg^{2+} ion is required to allow binding of the ligand to the aptamer. This is supported by observations made in the crystal structure of the adenine aptamer showing five magnesium ions present within the aptamer core domain (11). The possibility that a direct interaction

between the aptamer, the ligand and at least one magnesium ion is required for the binding cannot be excluded. However, this seems unlikely considering that, in the crystal structure (11), only one Mg^{2+} ion is close enough to the ligand to directly interact with it, and this ion is already involved in an interaction with the RNA core domain.

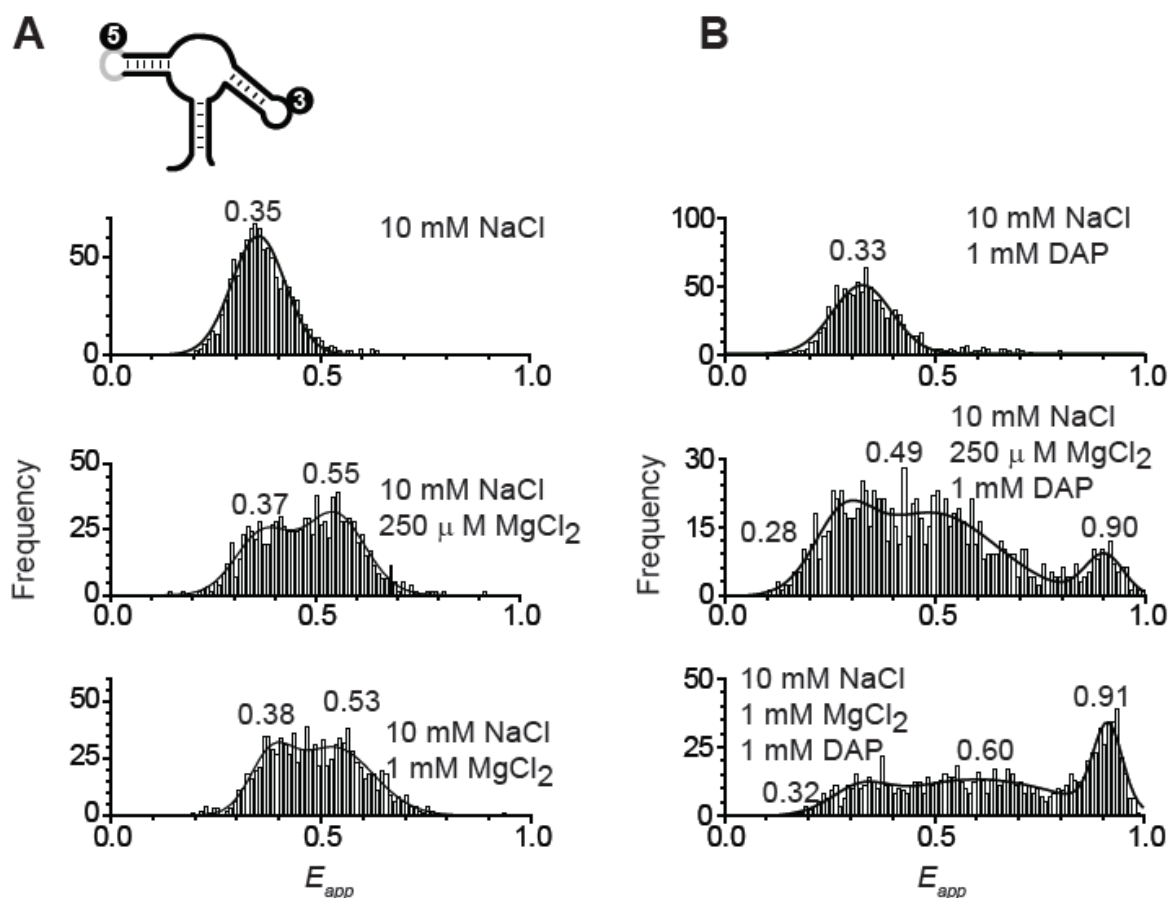


Figure 5 Sm-FRET analysis for the vector P2-P3 of the adenine L2 mutant aptamer as a function of ion and DAP concentrations. (A) Sm-FRET experiments performed for the L2 mutant aptamer in absence of DAP. Population histograms were collected in presence of 10 mM NaCl (top panel), 10 mM NaCl and 250 μ M $MgCl_2$ (middle panel) and 10 mM NaCl and 1 mM $MgCl_2$ (lower panel). Individual time traces were manually filtered to remove the contribution from fluorophore blinking and photobleaching events. E_{app} values determined by fitting analysis for each population are indicated. (B) Sm-FRET experiments performed for the L2 mutant aptamer in the presence of 1 mM DAP. Population histograms for the mutant aptamer were obtained in the same conditions as in A. Individual time traces were manually filtered to remove the contribution from fluorophore blinking and photobleaching events. E_{app} values determined by fitting analysis for each population are indicated.

Organization of the single stranded regions observed by SHAPE analysis

To explore the conformation of the *add* aptamer core domain, we employed SHAPE analysis to characterize the conformational states of the wild-type and L2 mutant aptamers. This technique allows detection of changes in accessibility of each nucleotide of the aptamer in any experimental conditions. According to our sm-FRET analysis, the wild-type and L2 mutant aptamers adopt the F (folded) and I (intermediate) conformations, respectively, when incubated in a buffer containing 10 mM NaCl and 10 mM Mg²⁺ ions (Figure 2, compare middle panels of A and B). Thus, by using these conditions, it is possible to probe the F and I conformations using SHAPE analysis.

SHAPE analysis was first performed on the wild-type molecule using a range of Mg²⁺ ion concentrations (Supplementary Figure 4). In the absence of Mg²⁺ ions, SHAPE products were detected for most single stranded regions as expected from previous studies (20, 21). However, the addition of magnesium ions resulted in the protection of regions consistent with the reorganization of the aptamer upon Mg²⁺ ion binding. For example, residues of the L3 region (residues 56-57) became more protected in the presence of 10 mM Mg²⁺ ions (Supplementary Figure 4), consistent with the formation of the loop-loop interaction. The addition of DAP resulted in the protection of several regions within the aptamer domain. However, the residue U39 was found to be highly reactive (Figure 6A), as expected from its high solvent exposition as revealed by the crystal structure (11). When these experiments were done with the L2 mutant, a globally similar folding pattern was obtained (Supplementary Figure 4). However, no protection of residues 56-57 was observed upon increasing the Mg²⁺ ions concentration, in agreement with our sm-FRET data (Figure 2B, middle panel). Interestingly, compared to the wild-type, we observed a higher reactivity for residue A43 (Figure 6A, see arrow) in the presence of 10 mM Mg²⁺ ions. Given that A43 is involved in a Watson-Crick base pair interaction with U13 in the ligand-bound crystal structure (Figure 6B, see dashed lines), it suggests that the U13-A43 base pair is not formed in the L2 mutant

aptamer in 10 mM Mg²⁺ ions. However, a strong protection is observed for A43 in presence of DAP (Figure 6B) suggesting that the base U13-A43 base pair is formed upon ligand binding. No increase of reactivity has been observed for AC44, which is involved in a Watson-Crick base pair with G37 in 10 mM Mg²⁺ (Figure 6A), indicating that this base pair is not perturbed in the context of the L2 mutant.

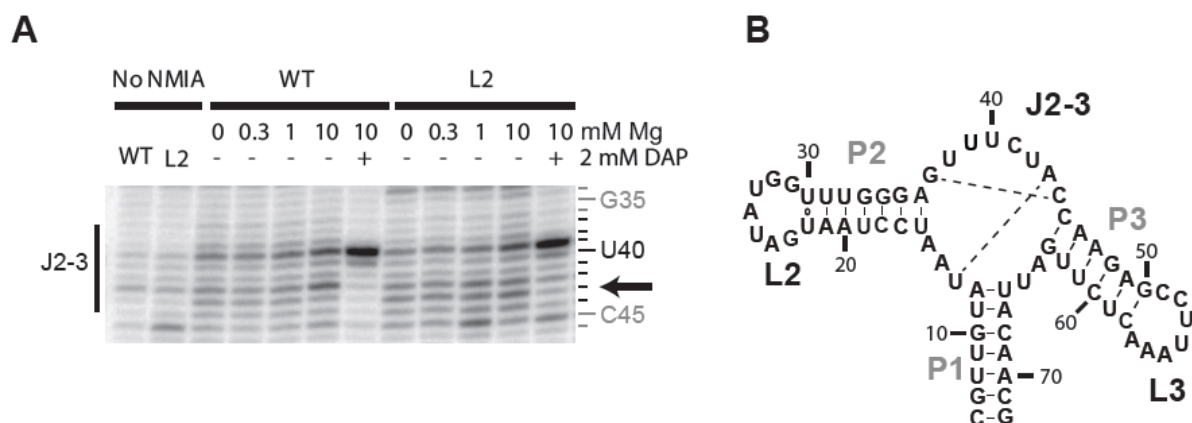


Figure 6 SHAPE analysis of the adenine aptamer. (A) SHAPE experiments performed for the wild-type (WT) and the L2 mutant (L2) aptamers. Reactions were done using various concentrations of magnesium ions (0, 0.3, 1 and 10 mM) in the absence or presence of 2 mM DAP. Control experiments were performed for both the wild-type and the L2 mutant in which the NMIA reagent was replaced with DMSO. Only the region J2-3 of the aptamer is shown. The reacting nucleotides are shown on the right of the gel and the arrow shows position A43. (B) Secondary structure and sequence of the *add* aptamer showing the presence of Watson-Crick base pairs occurring within the core domain of the aptamer (dotted lines). The complete gel is shown in the Supplementary Figure 4.

DISCUSSION

The intrinsic structural flexibility of RNA molecules makes them capable of participating in a remarkable wide range of biological systems. Such structural prowess requires RNA to follow a folding pathway that is particularly important to restrain the number of possible structural conformers for which RNA must sample before adopting biologically relevant structures (1, 7, 22). In the last decades, the folding of large RNA molecules—such as group I/II introns and RNase P—was shown to contain multiple possible folding pathways involving several long-lived intermediate states. In the case of the group I intron, the intermediate corresponds to a misfolded state that slows down the kinetics of RNA folding. Alternatively, the group II intron was shown to contain an obligate intermediate along the folding pathway, whose formation being rate limiting for ribozyme folding. Clearly, the diversity of RNA folding pathways and the presence of discrete intermediate conformers appear to be an integral part of RNA-based biological activity and regulation.

Using a combination of single-molecule FRET and biochemical approaches, we have characterized the folding pathway of the *add* adenine-sensing riboswitch. Although the architecture of the ligand-bound aptamer is well-known from various crystal structures obtained with several purine analogs (11), much less is known about the structure of the unbound aptamer. An interesting feature observable from crystal structures is that the bound ligand is totally surrounded by the RNA, suggesting that ligand binding is achieved through a partially folded aptamer exhibiting a less compact conformation. The organization level of such a partially folded conformer is not known in high detail, as no purine aptamer has been crystallized without ligand. We previously observed a transient intermediate conformer along the *add* aptamer folding pathway that was promoted upon ligand binding (16). By disabling the loop-loop interaction, we found here that the intermediate conformer was adopted in the absence of ligand, thus enabling sm-FRET experiments to deduce the global structure of all *add* structural conformers. Our model predicts that stems P1 and P3 are most probably

coaxially stacked in all aptamer conformations (Figure 7), as demonstrated by similar E_{app} values for all tested conditions (Figure 3A). The stacking of P1 and P3 stems is also supported by SHAPE data showing no significant change in the J3-1 region upon adding Mg^{2+} ions or ligand. Furthermore, SHAPE results suggest that J1-2 nucleotides are more flexible in the intermediate conformer than in the free and ligand-bound folded states indicating a higher flexibility of the P2 helix. Interestingly, these results suggest that the U13-A43 base pair located in the aptamer core domain is stabilized upon the formation of the loop-loop interaction. It should be mentioned that the RNA used in this study encompasses only the aptamer domain and not the expression platform. Because of this, it is not possible to know the role of the expression platform in the folding pathway of the aptamer. However, it is clear that the interactions observed in the unfolded and intermediate conformations can respectively be the first to happen and do not need to be altered to bind DAP (or adenine) since the intermediate is able to perform ligand binding (16). As a result, it is possible to classify structural transitions and suggest the order in which each group folds. Based on force microscopy analysis (23), it is most likely that the secondary structure of the aptamer forms in the order of stems P2, P3 and P1. Once the P1 stem is formed, the following tertiary interactions are formed in an indeterminate order: U40•A12-U66; C41•U11-A67 and G37-C44 inducing the stacking of the stems P1-P3. Together, these interactions are inducing the formation of the intermediate conformer. From there, the aptamer can transit via a folding pathway comprising either the $F_{unbound}$ or F_{bound} conformation (24) depending whether the ligand is available or not. In the absence of ligand, the loop-loop interaction is formed as well as U13-A43 base pair to produce the $F_{unbound}$ state. Upon ligand binding, the F_{bound} state is obtained by the formation of the loop-loop interaction as well as A14•G37-C44, U13-A43 and U38-U42-U65-ligand. The folding pathway proposed here is shown in Figure 7.

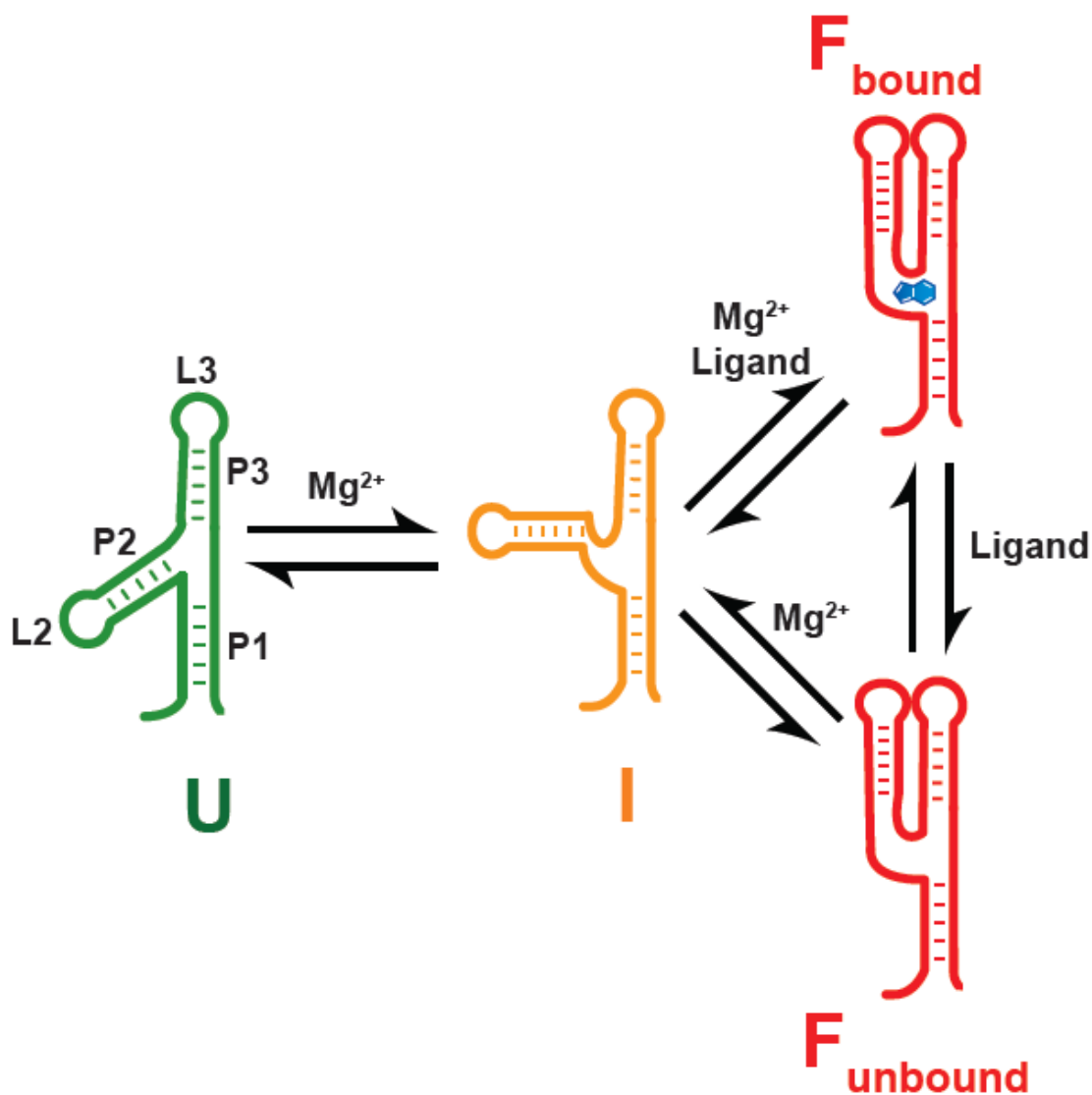


Figure 7 Folding pathway of the *add* adenine aptamer. The folding pathway is constituted by three different conformers which are the unfolded (U), the intermediate (I) and the folded (F) state. Based on our results, the P1 and the P3 stems are already stacked in the U state. The binding of magnesium ions results in the aptamer to adopt the I state, which is characterized by the global reorganization of the P2 stem relatively to the P1-P3 stacking unit. It is expected that the core region of the I state comprises the formation of intramolecular interaction such as the Watson-Crick base pair G37-C44. However, our results obtained by SHAPE analysis indicate that the U13-A43 base pair is not formed at this step. The subsequent folding of the aptamer leads to the formation of the F state, which can be obtained either in the absence ($F_{unbound}$) or presence (F_{bound}) of ligand. An interconversion between these two states is also possible (24).

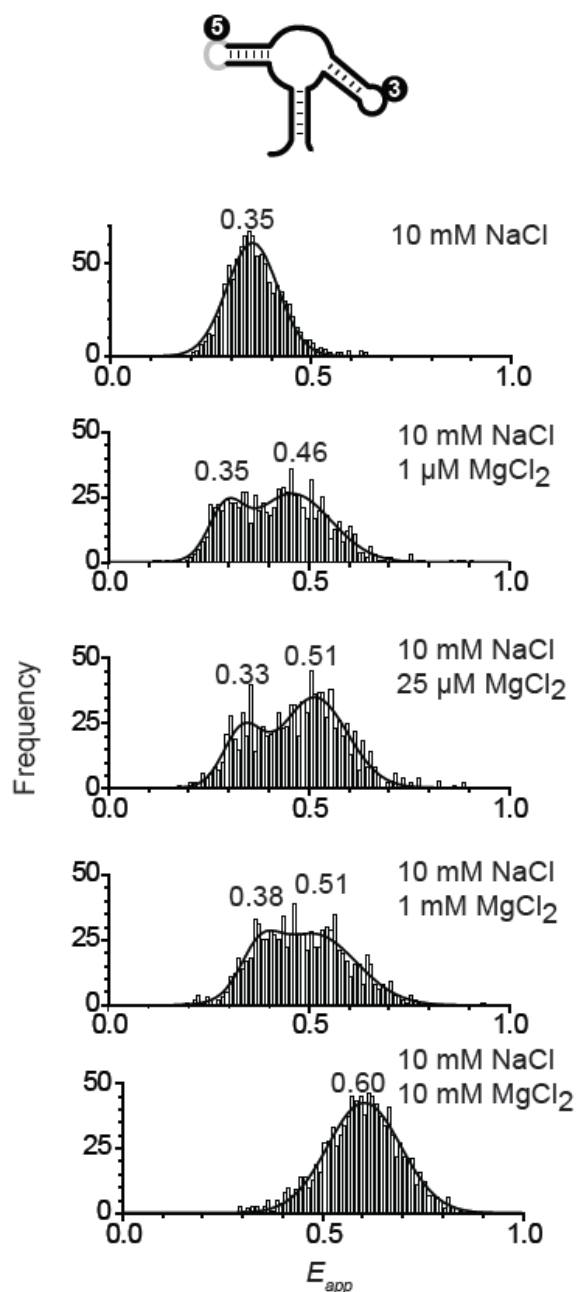
Lastly, it should also be considered that our experiments are performed post-transcriptionally. It is likely that the folding of the RNA is affected by the transcription process. However, force spectroscopy showed that, on a fully transcribed aptamer, the P1 stem is the last structure to be formed (23). Since transcription is producing the 3' part of the P1 stem at the end of the aptamer transcription process, the folding pathway presented here remains valid. Thus, this work proposes a folding pathway for the adenine-sensing aptamer and suggests a hierarchical process involving the role of an intermediate conformer along the folding pathway. The results presented here could serve in the study of more complex riboswitches or RNA-based regulating mechanisms.

ACKNOWLEDGEMENT

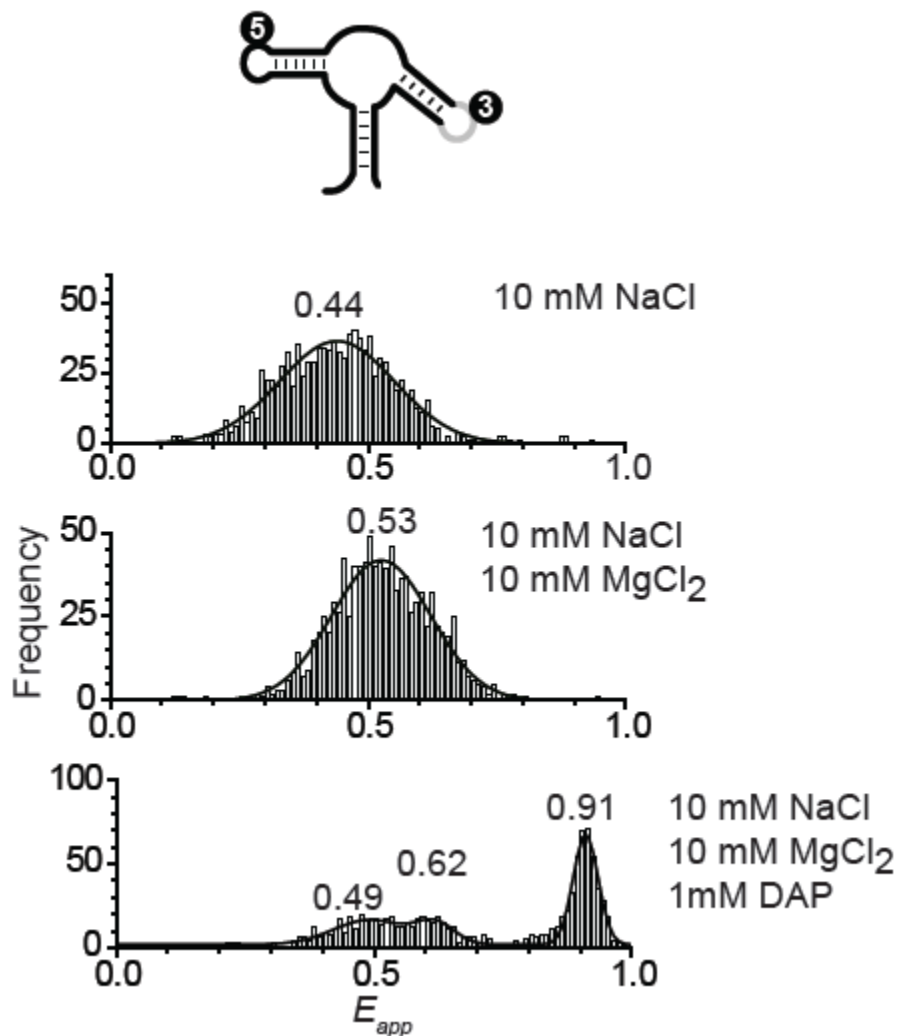
We thank members of the Lafontaine laboratory for discussion and Dr Alain Lavigueur for critical reading of the manuscript.

FUNDING

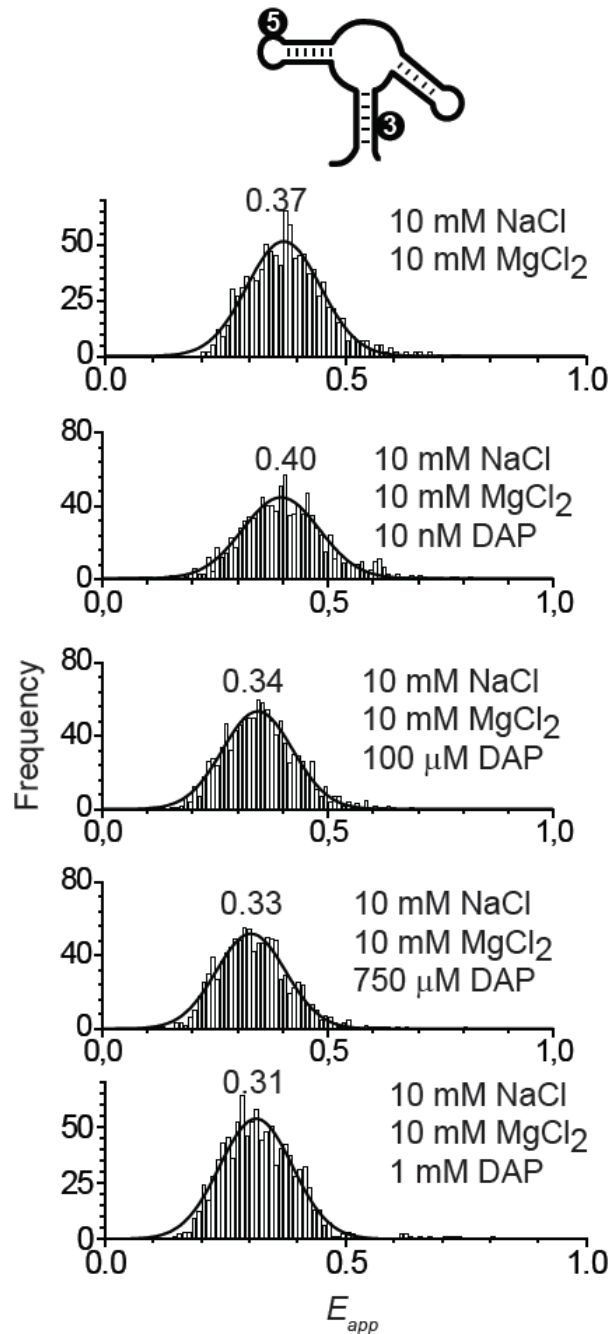
National Sciences and Engineering Research Council of Canada (NSERC); Doctoral fellowship from NSERC (to P.S.P); Canadian Institute for Health Research (CIHR) New Investigator scholar (to D.A.L.).



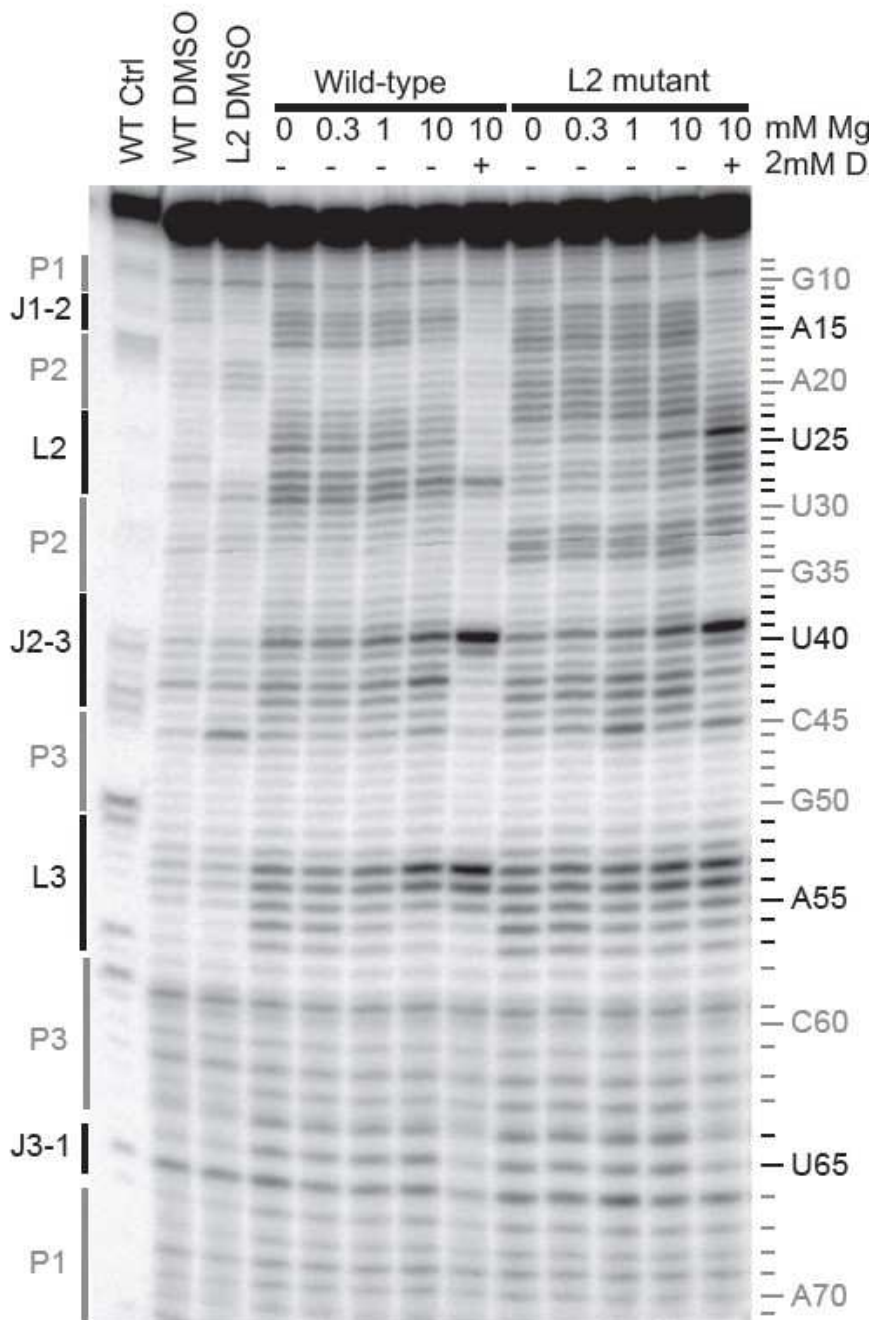
Supplementary Figure 1. Sm-FRET analysis for the vector P2-P3 as a function of magnesium ion concentrations. Sm-FRET experiments performed for the L2 mutant aptamer. Population histograms were collected in presence of 10 mM NaCl, 1 μ M MgCl₂, 25 μ M MgCl₂, 1 mM MgCl₂, and 10 mM MgCl₂. Individual time traces were manually filtered to remove the contribution from fluorophore blinking and photobleaching events. E_{app} values determined by fitting analysis for each population are indicated.



Supplementary Figure 2. Sm-FRET analysis for the vector P2-P3 of the adenine aptamer. Sm-FRET experiments performed for the L3 mutant aptamer. The cartoon represents the dual-labeled aptamer used for these studies. Population histograms were obtained in presence of 10 mM NaCl (top panel), 10 mM NaCl and 10 mM MgCl₂ (middle panel) and 10 mM NaCl, 10 mM MgCl₂ and 1 mM DAP (lower panel). Individual time traces were manually filtered to remove the contribution from fluorophore blinking and photobleaching events. E_{app} values determined by fitting analysis for each population are indicated.



Supplementary Figure 3. Sm-FRET analysis for the vector P1-P2 of the adenine aptamer. Sm-FRET experiments performed for the wild-type mutant aptamer as a function of DAP concentration. Population histograms were obtained in the buffer conditions described in the figure. Individual time traces were manually filtered to remove the contribution from fluorophore blinking and photobleaching events. E_{app} values determined by fitting analysis for each population are indicated.



Supplementary Figure 4. SHAPE analysis of the adenine aptamer. SHAPE experiments performed for the wild-type (WT) and the L2 mutant (L2) aptamers. Reactions were done using various concentrations of magnesium ions (0, 0.3, 1 and 10 mM) in the absence or presence of 2 mM DAP. Control experiments were performed for both the wild-type and the L2 mutant in which the NMIA reagent was replaced with DMSO.

REFERENCES

1. Roth,A. and Breaker,R.R. (2009) The structural and functional diversity of metabolite-binding riboswitches. *Annu Rev Biochem*, **78**, 305–334.
2. Serganov,A. and Patel,D.J. (2007) Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat Rev Genet*, **8**, 776–790.
3. Cheah,M.T., Wachter,A., Sudarsan,N. and Breaker,R.R. (2007) Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature*.
4. Bocobza,S., Adato,A., Mandel,T., Shapira,M., Nudler,E. and Aharoni,A. (2007) Riboswitch-dependent gene regulation and its evolution in the plant kingdom. *Genes Dev*, **21**, 2874–2879.
5. Caron,M.P., Bastet,L., Lussier,A., Simoneau-Roy,M., Masse,E. and Lafontaine,D.A. (2012) Dual-acting riboswitch control of translation initiation and mRNA decay. *Proc Natl Acad Sci U S A*, **109**, E3444–53.
6. Loh,E., Dussurget,O., Gripenland,J., Vaitkevicius,K., Tiensuu,T., Mandin,P., Repoila,F., Buchrieser,C., Cossart,P. and Johansson,J. (2009) A trans-acting riboswitch controls expression of the virulence regulator PrfA in *Listeria monocytogenes*. *Cell*, **139**, 770–779.
7. Serganov,A. and Nudler,E. (2013) A decade of riboswitches. *Cell*, **152**, 17–24.
8. Mandal,M. and Breaker,R.R. (2004) Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat Struct Mol Biol*, **11**, 29–35.
9. Mandal,M., Boese,B., Barrick,J.E., Winkler,W.C. and Breaker,R.R. (2003) Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria. *Cell*, **113**, 577–586.

10. Kim, J.N., Roth, A. and Breaker, R.R. (2007) Guanine riboswitch variants from *Mesoplasma florum* selectively recognize 2'-deoxyguanosine. *Proc Natl Acad Sci U S A*, **104**, 16092–16097.
11. Serganov, A., Yuan, Y.R., Pikovskaya, O., Polonskaia, A., Malinina, L., Phan, A.T., Hobartner, C., Micura, R., Breaker, R.R. and Patel, D.J. (2004) Structural Basis for Discriminative Regulation of Gene Expression by Adenine- and Guanine-Sensing mRNAs. *Chem Biol*, **11**, 1729–1741.
12. Lemay, J.F., Desnoyers, G., Blouin, S., Heppell, B., Bastet, L., St-Pierre, P., Masse, E. and Lafontaine, D.A. (2011) Comparative Study between Transcriptionally- and Translationally-Acting Adenine Riboswitches Reveals Key Differences in Riboswitch Regulatory Mechanisms. *PLoS Genet*, **7**, e1001278.
13. Rieder, R., Lang, K., Graber, D. and Micura, R. (2007) Ligand-Induced Folding of the Adenosine Deaminase A-Riboswitch and Implications on Riboswitch Translational Control. *Chembiochem*, **8**, 896–902.
14. Batey, R.T., Gilbert, S.D. and Montange, R.K. (2004) Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature*, **432**, 411–415.
15. Porter, E.B., Marcano-Velázquez, J.G. and Batey, R.T. (2014) The purine riboswitch as a model system for exploring RNA biology and chemistry. *Biochim. Biophys. Acta - Gene Regul. Mech.*, **1839**, 919–930.
16. 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.
17. Aitken, C.E., Marshall, R.A. and Puglisi, J.D. (2008) An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophys. J.*, **94**, 1826–35.
18. Blouin, S., Chinnappan, R. and Lafontaine, D.A. (2010) Folding of the lysine riboswitch:

- importance of peripheral elements for transcriptional regulation. *Nucleic Acids Res*, **39**, 3373–3387.
19. Heppell,B., Blouin,S., Dussault,A.M., Mulhbach,J., Ennifar,E., Penedo,J.C. and Lafontaine,D.A. (2011) Molecular insights into the ligand-controlled organization of the SAM-I riboswitch. *Nat. Chem Biol*, **7**, 384–392.
 20. Tremblay,R., Lemay,J.F., Blouin,S., Mulhbach,J., Bonneau,E., Legault,P., Dupont,P., Penedo,J.C. and Lafontaine,D.A. (2011) Constitutive regulatory activity of an evolutionary-excluded riboswitch variant. *J Biol Chem*, **286**, 27406–27415.
 21. Stoddard,C.D., Gilbert,S.D. and Batey,R.T. (2008) Ligand-dependent folding of the three-way junction in the purine riboswitch. *RNA*, **14**, 675–684.
 22. Garst,A.D., Edwards,A.L. and Batey,R.T. (2011) Riboswitches: structures and mechanisms. *Cold Spring Harb Perspect Biol*, **3**.
 23. 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 (80-.)*, **319**, 630–633.
 24. Dalgarno,P.A., Bordello,J., Morris,R., St-Pierre,P., Dube,A., Samuel,I.D., Lafontaine,D.A. and Penedo,J.C. (2013) Single-molecule chemical denaturation of riboswitches. *Nucleic Acids Res*, **41**, 4253–4265.
 25. Gilbert,S.D. and Batey,R.T. (2006) Riboswitches: fold and function. *Chem Biol*, **13**, 805–807.

CHAPITRE 2

Repliement du cœur du riborégulateur adénine observé au niveau de la molécule unique

Les structures cristallines et les techniques biochimiques comme la cartographie par coupure en ligne, la cartographie aux ARNases et le SHAPE fournissent beaucoup d'informations sur le domaine de liaison des aptamères. Cependant, il est toujours difficile d'observer la dynamique de repliement du domaine de liaison d'un riborégulateur. En effet, les nucléotides composants le cœur des aptamères sont presque toujours impliqués dans l'organisation de la pochette de liaison. Il est donc difficile de modifier les nucléotides de cette région sans affecter le fonctionnement de l'ARN. Or, les techniques permettant d'observer le repliement des molécules d'ARN requièrent presque toujours la modification d'au moins un nucléotide. Par exemple, il peut s'agir de l'incorporation d'une molécule fluorescente de 2AP ou de l'ajout de fluorophores.

Il s'avère que l'aptamère *add* possède un nucléotide dans la J2-3 qui peut être modifié. Dans l'article présenté dans ce chapitre, cette possibilité a été exploitée afin d'étudier la dynamique de repliement de la pochette de liaison. Un fluorophore a été attaché à ce nucléotide afin d'observer le repliement de l'aptamère en sm-FRET. Les expériences de sm-FRET ont été réalisées sur des aptamères de type sauvage et des aptamères, empêchant la formation de l'interaction boucle-boucle (L2). Les résultats obtenus lors de cette étude démontrent que le repliement du cœur induit par le magnésium et par le ligand influence la structure globale de l'aptamère.

Dans cette étude, Patrick St-Pierre a réalisé les expériences de FRET et de sm-FRET pour l'aptamère de type sauvage, Jonathan Grondin a réalisé les expériences de sm-FRET pour l'aptamère mutant et Daniel Lafontaine a écrit le manuscrit qui sera soumis à la revue scientifique *Nucleic Acid Research*.

Folding of the adenine riboswitch core domain monitored at the single-molecule level

Patrick St-Pierre, Jonathan Grondin and Daniel A. Lafontaine*

Department of Biology, Faculty of Science, RNA Group, Université de Sherbrooke, Sherbrooke, Québec J1K 2R1, Canada

*To whom correspondence should be addressed. Tel: 891-821-8000; Email: daniel.lafontaine@usherbrooke.ca

ABSTRACT

The *Bacillus subtilis* adenine riboswitch is one of the rare examples of a riboswitch positively modulating gene expression upon metabolite binding. The folding pathway of this riboswitch was previously characterized using an array of biochemical and biophysical techniques. From these studies, it was found that a loop-loop interaction is central to adenine recognition at physiological magnesium concentrations. Here, using a single-molecule approach, we further characterize the folding of the core riboswitch domain that is directly involved in adenine sensing. We find that the folding of the core domain occurs at a magnesium concentration similar to that at which the loop-loop is adopted, suggesting that both structural rearrangements are interrelated. In addition, we observe that the core domain undergoes structural rearrangement upon adenine binding. By preventing the formation of the loop-loop interaction, we determine that the magnesium-dependent folding of the core domain is perturbed. However, addition of ligand restores the folding of the core domain in the context of the mutant riboswitch. This work directly shows that the loop-loop interaction and the core domain are both involved in the pre-organization of the riboswitch structure and ligand recognition.

INTRODUCTION

Riboswitches are genetic regulatory elements that are located in the 5' untranslated region (UTR) of messenger RNA (1, 2). These control elements perform the sensing of specific metabolites and, as a result, alter their structure to control gene expression (3). Riboswitches are composed of two domains: the aptamer and the expression platform. While the aptamer domain is involved in the specific recognition of the cellular metabolite, the expression platform is involved in the control of gene expression, which is usually operated at the levels of transcription, translation or mRNA decay (1). In all cases, the structure of the expression platform is directly controlled by the ligand-bound state of the aptamer domain. To date, more than 25 different classes of riboswitches have been discovered, each specifically binding a different metabolite such as amino acids, carbohydrates, coenzymes, and nucleobases. Other RNA switches responding to temperature, pH and ions have also been discovered (1), suggesting that RNA-based molecular mechanisms are highly versatile in regulating cellular processes.

The purine-specific class of riboswitches is one of the most studied class and responds to adenine, guanine, ZTP and preQ1 metabolites (3). The adenine-responsive riboswitch is one of the rare type of riboswitches that positively regulates gene expression upon ligand binding. Specifically, in the case of the *Vibrio vulnificus* adenine-sensing riboswitch, ligand binding allows the expression platform to adopt a structure in which the Shine-Dalgarno (SD) sequence is exposed (Figure 1A), therefore allowing translation initiation to be achieved. However, in the absence of adenine, a sequestrator stem-loop is formed in the expression platform that prevents translation initiation (Figure 1B), and consequently, gene expression is not allowed. The *V. vulnificus* adenine riboswitch controls the expression of an adenosine deaminase (*add*), which is involved in the metabolism of purine-related cellular metabolites. An adenine-responsive riboswitch is also found in *Bacillus subtilis* and has been shown to positively control transcription elongation (4, 5). Although both riboswitches are highly

similar, they exhibit very different ligand binding mechanisms (6). This suggests that the regulation level, that is translation or transcription, may influence how riboswitches use metabolite recognition to regulate gene expression. In the last decade, purine-responsive riboswitches have been characterized using a combination of biochemical and biophysical approaches (7). From these studies, it has emerged that the folding pathway of the aptamer mainly consists in the adoption of a loop-loop interaction involving loop 2 (L2) and loop 3 (L3) (Figure 1A) (4, 8, 9). Furthermore, biochemical assays such as in-line probing and SHAPE (Selective 2'-hydroxyl acylation analyzed by primer extension) have revealed that ligand binding to the aptamer results in the reorganization of the core domain, which consists in the single stranded regions connecting stems P1, P2 and P3 (Figure 1A). This is consistent with crystal structures (10, 11) showing that the ligand binding site is confined within the core domain, the three-way junction, and that a particular uracil residue (Figure 1A) is involved with the bound adenine in a Watson-Crick base pair interaction. Here, we characterize in more detail the folding of the core domain of the adenine riboswitch by using a single-molecule approach. By introducing a fluorophore within the core domain of the adenine riboswitch aptamer, we monitor the folding as a function of magnesium ions and adenine. We find that the folding of the core domain occurs at low magnesium ion concentrations, similarly to the loop-loop interaction. Furthermore, we find that the formation of the loop-loop interaction is involved in the folding of the core domain, suggesting that both folding processes are intimately linked. These results directly show how the formation of the loop-loop interaction is involved in ligand sensing by allowing the pre-folding of the riboswitch core domain.

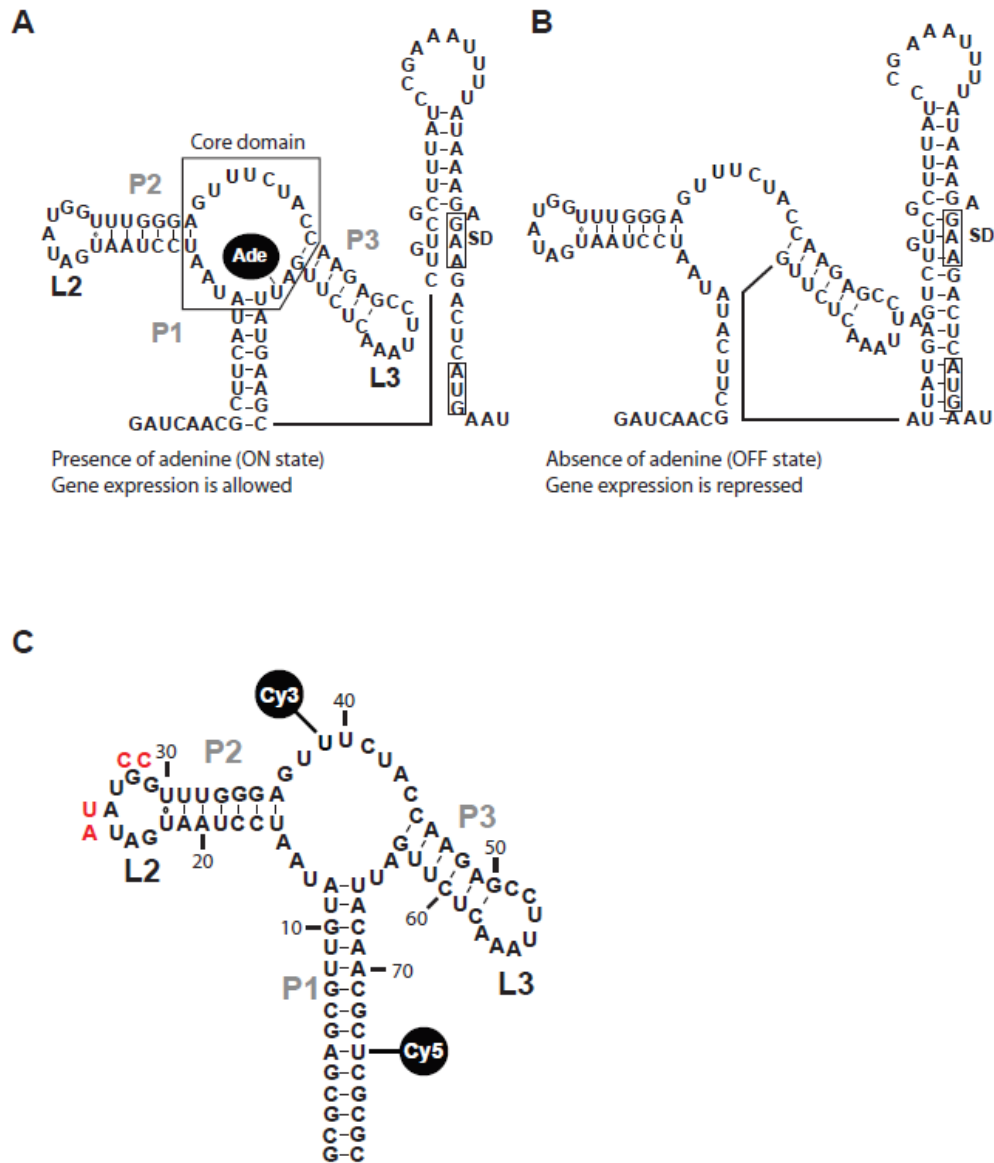


Figure 1 Secondary structures of the *Vibrio vulnificus* adenine riboswitch. (A) Secondary structure associated with the ON state in which the Shine-Dalgarno (SD) and the AUG initiation codon (boxed) are accessible for translation initiation. (B) Secondary structure of the adenine riboswitch in absence of adenine where translation initiation is not allowed. Note the presence of a translation stem-loop sequestering preventing ribosome translation initiation. (C) Representation of the adenine aptamer showing the donor (fluorescein) and acceptor (Cy3) fluorophores attached at their respective locations by conjugation to 5-amino-allyl uridine nucleotides.

MATERIAL and METHODS

Preparation of dual-labelled aptamer molecules

Vibrio vulnificus adenine aptamer molecules were constructed from two oligonucleotides:

WT 5' strand: Biotin-

GCGCGAGCGUUGUAUAAUCCUAAUGAUAUGGUUUGGGAGUUUCUACCAAGAGC
CUU

WT 3' strand: PO₄-AAACUCUUGAUUACAACGCUCGCGC

The L2 mutant was prepared using a different 5' strand (mutations are shown in bold):

L2 5' strand: Biotin-

GCGCGAGCGUUGUAUAAUCCUAAUGAAUUCCUUGGGAGUUUCUACCAAGAGC
CUU

RNA molecules were obtained from IDT Canada with 5'-amino-allyl thymidine at defined positions (underlined nucleotides). Fluorophores were introduced at defined positions using succinimidyl ester derivatives according to the manufacturer's protocol (12). Labeled RNA strands were purified by denaturing gel electrophoresis as previously described (12). Purified fluorescent RNA molecules were annealed together (molar ratio 1:1) by heating a mixture to 75°C in 10 mM Tris (pH 7.5), 50 mM NaCl and slowly cooling it to room temperature.

Annealed complexes were then subjected to ligation by using the T4 RNA ligase procedure (12). Next, ligated RNA molecules were further purified by denaturing gel electrophoresis, precipitated with ethanol and redissolved in water.

FRET measurement in the steady state

Fluorescence spectroscopy was performed on a Quanta Master fluorometer. Fluorescence experiments were collected at 25°C in a buffer containing 50 mM Tris-HCl, pH 8.3. For each experiments, the concentrations of NaCl, magnesium ions and ligand are indicated in the corresponding figures and figure legends. Fluorescence spectra were corrected for background and polarization artefacts were avoided by setting excitation and emission polarizers crossed at 54.74°. Values of eFRET were determined using the acceptor normalization method (13, 14). Steady state FRET assays were performed with the 5' and the 3' strands labelled with Cy3 and fluorescein, respectively.

The folding of a macromolecule as a function of magnesium ions can be modelled using a simple two-state model assuming an all-or-none conformational transition. This approach was previously used to characterize the folding of the loop-loop interaction of the adenine aptamer (12). We have used a similar approach to model the folding of the core domain of the adenine aptamer. In all cases, the reported errors are the standard error in the fitting, which is assumed to be approximated by the standard deviation of the points from the fitted curve (12).

Single-molecule FRET analysis

Fluorescence data were obtained from single molecules by using a total internal reflection fluorescence microscope with a 532 nm laser excitation and an EMCCD camera (iXON, Andor Technology). Microscope slides were successively treated with biotinylated bovine serum albumin (BSA) and streptavidin. Dual-labelled aptamer molecules were then added to the slide and data were acquired with a 50 ms integration time. The reacting buffer used for sm-FRET analyses was made as previously described (12). Measurements were performed at room temperature. Data analysis was performed by using analysis routine developed in MATLAB 7. Single-molecule histograms were built by averaging the first 10 frames of each FRET trace for every individual molecule after manual filtering.

RESULTS and DISCUSSION

Design of dual-labelled aptamers to monitor the folding of the riboswitch core domain

Previous studies have established that the adenine-sensing aptamer follows a folding pathway in which a loop-loop interaction is formed between the regions L2 and L3 (Figure 1A) (4, 10, 11, 15–17). Although this interaction was not found to be essential for ligand binding, it was showed to be highly important to allow ligand binding at physiological concentrations of magnesium ions (18). Single-molecule Förster Resonance Energy Transfer (sm-FRET) approaches have previously been used to directly monitor the folding kinetics of the loop-loop interaction and have revealed that an intermediate conformer is part of the folding process (12). Interestingly, an sm-FRET study performed on the highly similar *Bacillus subtilis* guanine-sensing aptamer has not revealed the presence of such an intermediate in the folding

pathway (19), suggesting that both riboswitch aptamers may employ different strategies to perform efficient ligand recognition.

Several investigations have revealed that the core domain of the riboswitch is reorganized upon the binding of magnesium ions (5, 20, 21). Given that the core domain is the ligand binding site for the adenine aptamer (10, 11), it suggests a model where the core region of the adenine aptamer adopts a pre-folded state that is required prior to adenine binding. To investigate this process in more detail, we have designed a dual-labelled fluorescent construct amenable to sm-FRET analysis in which a Cy3 dye is incorporated within the core region of the aptamer at position 39 (Figure 1C). This position was chosen to introduce the Cy3 dye since (i) it does not show any sequence conservation (5, 17, 21), (ii) it tolerates the presence of a 2-aminopurine fluorophore without perturbing the ligand binding activity (20, 21), and (iii) it does not interact with the rest of the aptamer molecule according to crystal structures (10, 11). Therefore, by introducing a Cy5 dye within the P1 helical domain (Figure 1C), it allows to monitor the FRET efficiency occurring between both the Cy3 and Cy5 fluorophores. The introduction of a Cy5 within the P1 stem does not perturb ligand binding (St-Pierre and al., manuscript in preparation). As a result, this fluorescent aptamer construct allows to monitor any structural changes arising within the aptamer core domain—or the P1 stem—that affects the distance separating both Cy3 and Cy5 dyes.

Folding of the riboswitch core domain characterized by FRET assay

Before analyzing the folding of the adenine aptamer at the single-molecule level, we first monitored the aptamer folding under bulk in steady state conditions, where obtained FRET signals represent the average conformation(s) adopted by aptamers in solution. These assays are important as they enable to calculate the affinity of magnesium ions during the folding transition. We first monitored the FRET efficiency as a function of magnesium ion

concentrations (Figure 2). The addition of magnesium ions induced an increase of FRET in the micromolar range, suggesting that the core domain is reorganized upon magnesium binding (Figure 2). We have fitted our data to a simple two-state binding model assuming an all-or-none conformational transition induced by the noncooperative binding of magnesium ions (see Material and Methods). From the model, we can determine the concentration of magnesium ions ($[Mg^{2+}]_{1/2}$) at which the folding transition is half complete. In the absence of ligand, our fitting analysis yielded $[Mg^{2+}]_{1/2} = 38 \pm 3 \mu M$. However, when the experiments were repeated

in the presence of a saturating concentration of $10 \mu M$ adenine, a similar folding transition was observed but for which a $[Mg^{2+}]_{1/2} = 23 \pm 1 \mu M$ was calculated. The decrease in the $[Mg^{2+}]_{1/2}$ value in the presence of adenine indicates that the folding of the core domain is

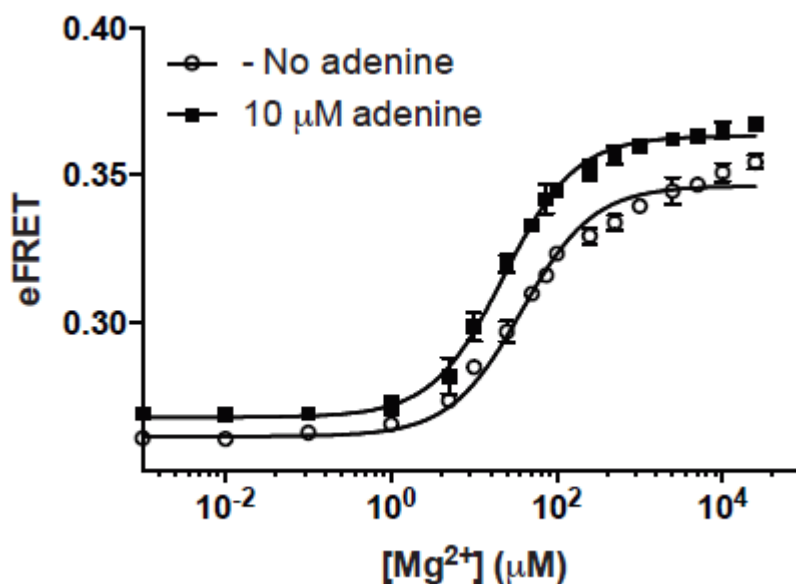


Figure 2. Plot of the efficiency of FRET as a function of magnesium ion concentration in the absence (open circles) or the presence (black squares) of $10 \mu M$ adenine. FRET efficiency (eFRET) increased with the addition of magnesium ions, corresponding to an ion-induced folding of the core domain region. The experimental data were fitted (lines) by regression to a simple two-state model in which the binding of metal ions to RNA induces a structural change.

influenced by adenine binding. These results are supported by in-line probing analysis showing that the core region is reorganized upon adenine binding (5). Together, these results indicate that the adenine riboswitch core domain is reorganized upon the binding of magnesium ions and that adenine is involved in the folding process.

We previously reported using bulk FRET assays that the $[Mg^{2+}]_{1/2}$ value for the formation of the L2-L3 interaction occurs in the low micromolar range (12), and is also decreased by a factor of ~2-fold in the presence of adenine. Thus, these results suggest that both the riboswitch core domain and the loop-loop interaction exhibit very similar folding properties. Given that it was previously observed that the presence of the loop-loop interaction is important for the efficiency of ligand binding (11, 18), it indicates that the formation of the loop-loop structure is somehow related to the folding of the riboswitch core domain and to the recognition of adenine. Similarly, a loop-loop interaction has previously been demonstrated to be important for the folding of another three-way junction RNA molecule, the hammerhead ribozyme, which was shown to be important for the catalytic activity to be attained at physiological magnesium ions concentrations (22). Given that the magnesium-dependent autocatalytic cleavage of the hammerhead ribozyme occurs within the core region of the RNA molecule, it suggests that the formation of the hammerhead loop-loop interaction modulates the folding of the core domain, thus exhibiting a folding property similar to the adenine riboswitch. Finally, a peripheral structural element within the *S*-adenosylmethionine (SAM) riboswitch was also found to influence the folding of the riboswitch core region (23). In this case, the peripheral element corresponds to a single stem-loop structure (the P4 stem) and the stability of the latter was shown to be proportional to the riboswitch ligand binding affinity. Since ligand binding is performed within the riboswitch core domain, it strongly suggests that the stability of the P4 stem is directly influencing the folding of the core domain, and consequently, the ability of the SAM riboswitch to achieve ligand recognition.

Single-molecule FRET assays reveals that the riboswitch core domain adopts two distinct structural conformers

We next analyzed the folding of the adenine riboswitch aptamer using sm-FRET analysis. These assays are very powerful to determine the folding pathway of any biomolecules by providing information about the folding dynamics of individual molecules (24, 25), thereby avoiding the difficulties obtained by using ensemble measurements. Dual-labelled adenine aptamers were attached to a quartz slide via streptavidin-biotin linkage (see Material and Methods) and were monitored using a wide-field total internal reflection (TIR) microscope (12, 23, 26). Single-molecule time trajectories have been acquired at different magnesium ion concentrations and are represented using apparent FRET efficiency (E_{app}) histograms (Figure 3A). In absence of magnesium ions (10 mM NaCl), low ($E_{app} = \sim 0.33$) and high ($E_{app} = \sim 0.56$) FRET efficiencies were detected (Figure 3A, top panel) indicating that adenine aptamers are adopting at least two structural conformers. When these experiments were repeated as a function of magnesium ions, the efficiency of the low and high FRET population increased and reached values of ~ 0.42 and ~ 0.74 , respectively, at 10 mM Mg^{2+} ions (Figure 3A, lower panel). Given the small change occurring for the low eFRET population between the absence and presence of 10 mM magnesium ions (~ 0.33 vs ~ 0.42), it makes difficult to assess whether a structural change is really occurring between both experimental conditions. In contrast, the addition of 10 mM magnesium ions caused a clear structural change for the high E_{app} population where a value of ~ 0.74 was obtained (Figure 3A, lower panel). Thus, compared to the absence of magnesium ions (Figure 3A, top panel, $E_{app} = \sim 0.33$), it suggests that the core domain undergoes structural change upon the binding of magnesium ions. These results are very similar to those obtained in bulk conditions (Figure 2). It should be noted that no direct comparison between the FRET values can be made for bulk and single-molecule conditions since different sets of dyes were used for each set of experiments.

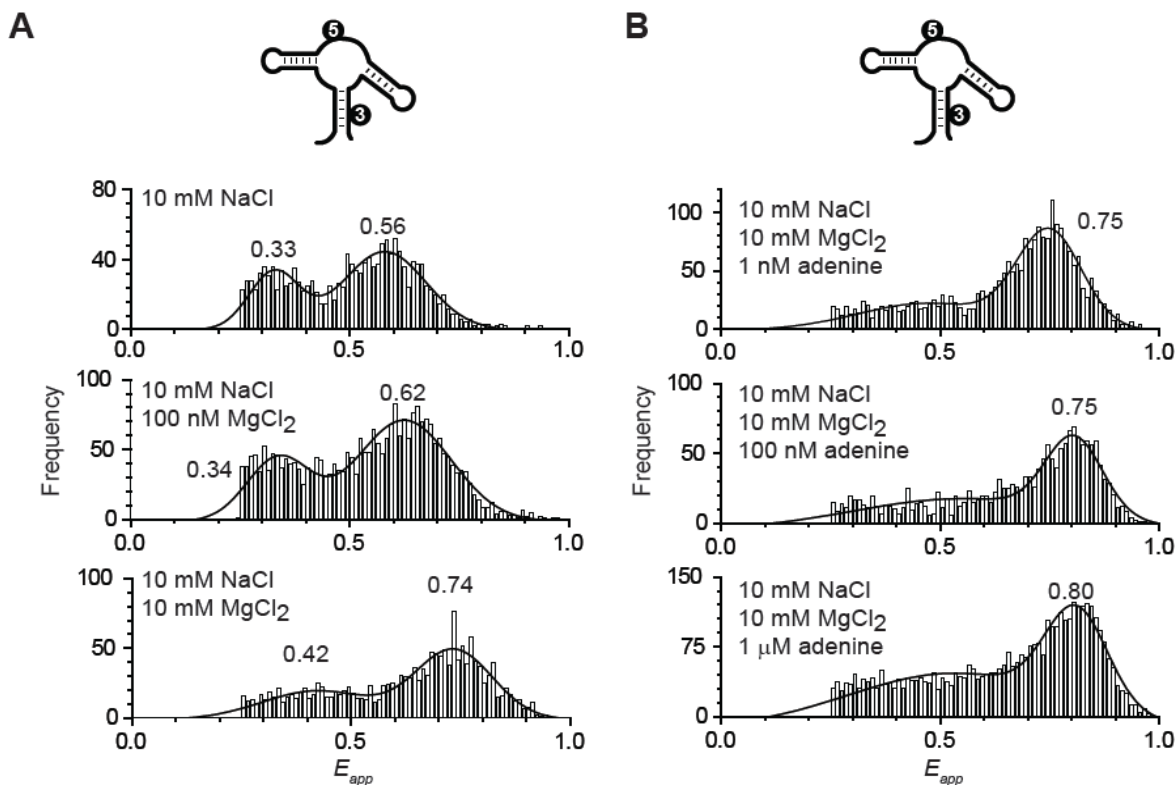


Figure 3 Histograms of FRET efficiency summed over multiple single molecules at different magnesium ion concentrations. (A) Sm-FRET experiments were performed in the absence of adenine or DAP. The cartoon represents the dual-labeled aptamer used for these studies. Cy3 (3) and Cy5 (5) fluorophores are indicated. Individual time traces were manually filtered to remove the contribution from fluorophore blinking and photobleaching events. E_{app} values determined by fitting analysis for each population are indicated. (B) Sm-FRET experiments were performed as a function of the adenine concentration. eFRET values determined by fitting analysis for each population are indicated.

We next monitored the influence of adenine on the folding of the riboswitch core domain. As it can be observed in Figure 3B (top panel), the addition of 1 nM adenine resulted in an E_{app} value of ~ 0.75 , which is very similar to the value obtained in absence of adenine (Figure 3A, lower panel, $E_{app} = \sim 0.74$). No value was determined for the low E_{app} population due to the absence of a defined peak representing this aptamer population. When we repeated sm-FRET assays using higher adenine concentrations, we observed that the E_{app} was increased to a value of ~ 0.80 when using 1 μM adenine (Figure 3B, lower panel), suggesting that the core domain is slightly rearranged at higher adenine concentrations. This suggests that the binding of adenine to the riboswitch reorganizes the core domain, consistent with previously reported biochemical data showing the reorganization of the core upon ligand binding (5, 6, 12, 21).

Using in-line probing assays, it was previously observed that the core domain of the *B. subtilis* adenine riboswitch adopts a more constrained structure upon ligand binding (5), suggesting that the structure of the riboswitch aptamer is reorganized by adenine recognition. Additional insights about the aptamer core reorganization was subsequently obtained by introducing a 2-aminopurine fluorophore within the aptamer core domain (20, 21), showing that the 2-aminopurine fluorescence is highly increased upon ligand binding. These results gave rise to a folding model in which the structure of the adenine riboswitch aptamer is extensively reorganized upon ligand binding (20, 21). Furthermore, 2-aminopurine fluorescence indicated that the core domain of the adenine aptamer is modulated upon magnesium ions binding (21), consistent with our sm-FRET analysis (Figure 3A). However, it is difficult to ascertain to which degree the core domain is pre-folded upon the binding of magnesium ions. Although all previous studies were analyzing the effect of ligand binding on the core domain by monitoring the local structure, our sm-FRET analysis reveals that the global structure of the core domain is also reorganized both in the presence of Mg^{2+} ions and ligand. It seems highly probable that the magnesium-dependent step is important for adenine recognition by pre-folding the aptamer core domain into a ligand-binding competent state.

The importance of magnesium ions in the pre-folding of riboswitches has also previously been reported for the SAM, lysine and thiamine pyrophosphate (TPP) riboswitches (23, 27–29). In each case, the presence of magnesium ions was found to induce the riboswitch to adopt a structure different from the ligand bound state. One of the most characterized case is the SAM riboswitch for which magnesium binding results in a structure highly similar to the folded state (23), but in which key folding processes have not yet been attained due to the absence of the SAM ligand. Similarly to the adenine riboswitch, the magnesium-induced structural state of the SAM riboswitch relies on the formation of a long-range tertiary interaction, a pseudoknot structure, to allow further progress in the folding pathway and ultimately SAM binding. Indeed, the introduction of mutations destabilizing the pseudoknot structure severely inhibits the folding pathway as observed using comparative gel analysis (23). Clearly, as determined here for the adenine aptamer, magnesium ions are highly involved in the pre-folding of riboswitch aptamer core domains, suggesting that they have an important role in ligand recognition.

The formation of the loop-loop interaction is highly important for the pre-folding of the riboswitch core domain

To establish the molecular requirements of the folding process of the core domain, we next investigated the importance of the loop-loop interaction using sm-FRET experiments. Based on crystal structures (10, 11) and previous biophysical studies (12), we introduced mutations (U25A, A26U, G28C and G29C) within the L2 region (L2 mutant) to disable the formation of the loop-loop interaction (see Material and Methods for mutations). When performing sm-FRET experiments using the L2 mutant in absence of magnesium ions and ligand, we observed that adenine aptamers folded into two different populations exhibiting E_{app} values of ~ 0.37 and ~ 0.61 (Figure 4A, top panel), similarly to the wild-type aptamer (Figure 3A, top panel). These results indicate that the presence of two populations observed for the wild-type molecule in the absence of Mg^{2+} ions (Figure 3A, top panel) does not depend on the presence

of a functional loop-loop interaction. Unexpectedly, the addition of magnesium ions resulted in the adoption of a single aptamer population exhibiting eFRET values of ~ 0.44 and ~ 0.49 at $10\ \mu\text{M}$ and $100\ \mu\text{M}$ magnesium ions, respectively (Figure 4A, second and third panels). The further increase of the Mg^{2+} ions concentration at $10\ \text{mM}$ yielded an E_{app} value of ~ 0.59 (Figure 4A, lower panel). These results indicate that, in contrast to the wild-type aptamer, the L2 mutant folds homogeneously into a single E_{app} population at a saturating Mg^{2+} ions concentration. Given that the $E_{app} \sim 0.59$ population does not correspond to the one observed for the wild-type molecule (Figure 3A, lower panel, $E_{app} = \sim 0.74$), it suggests that the L2 mutant is not able to fold into a structure similar to the wild-type aptamer. Most probably, the inability of the L2 mutant to perform the loop-loop interaction results in the formation of the intermediate aptamer conformation exhibiting partial aspects of the native aptamer structure (12)(St-Pierre and *al.*, in preparation).

Given that the introduction of mutations preventing the loop-loop formation severely decreases ligand binding affinity (18), we next investigated whether the ligand-dependent folding of the aptamer core domain could still occur within the L2 mutant. In these experiments, DAP was employed instead of adenine due to its ~ 30 -fold higher binding affinity toward the adenine aptamer (5). As it can be observed in the Figure 4B (top panel), a single population exhibiting an E_{app} value of ~ 0.67 was detected in the presence of $1\ \text{mM}$ DAP. However, an additional E_{app} value of ~ 0.85 was detected when using $200\ \mu\text{M}$ DAP (Figure 4B, middle panel). This E_{app} value is highly similar to what obtained for the wild-type aptamer in the presence of $1\ \text{mM}$ adenine (Figure 3B, lower panel), indicating that the core domain in the context of the L2 mutant folds similarly to the wild-type aptamer. Furthermore, when these experiments were repeated in the presence of $1\ \text{mM}$ DAP (Figure 4B, lower panel), a single

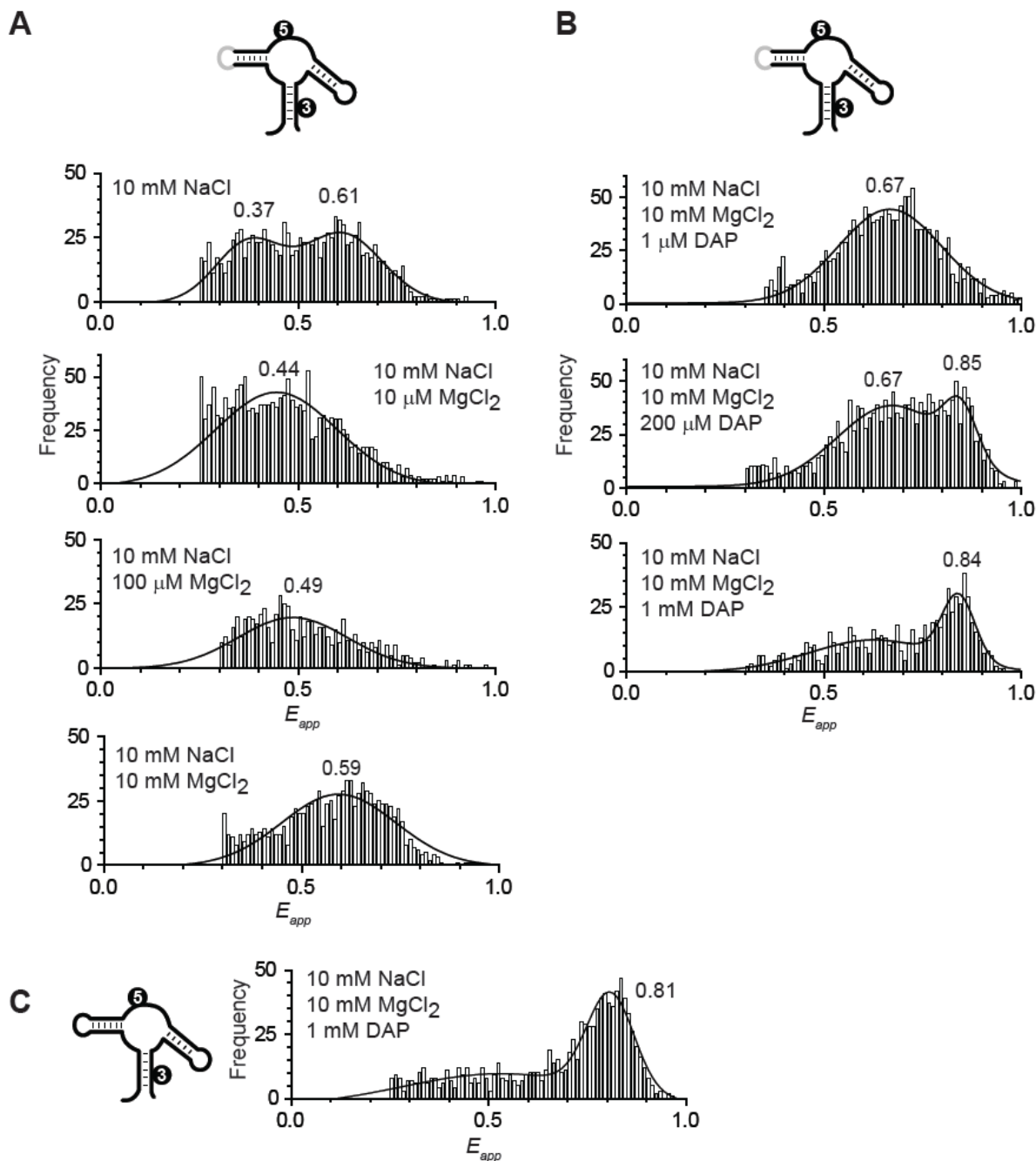


Figure 4 Histograms of FRET efficiency performed for the L2 mutant. (A) Sm-FRET experiments were performed in the absence of ligand. E_{app} values determined by fitting analysis for each population are indicated. (B) Sm-FRET experiments were performed as a function of the adenine concentration. E_{app} values determined by fitting analysis for each population are indicated. (C) Sm-FRET experiments were performed in the presence of 10 mM 2,6-diaminopurine (DAP). The E_{app} value determined by fitting analysis is indicated for the aptamer population.

E_{app} population of ~ 0.84 was obtained. As a control, sm-FRET experiments were performed with DAP in the context of the wild-type aptamer (Figure 4C). A very similar E_{app} value of ~ 0.81 was obtained when using 1 mM DAP compared to 1 mM adenine (Figure 3B, lower panel), indicating that DAP binding to the aptamer results in a global structure very similar than the one obtained through adenine binding.

CONCLUSION

In this work, we have monitored the folding of the core domain of the adenine riboswitch using bulk and sm-FRET experiments. In contrast to previously published sm-FRET studies characterizing the folding of riboswitches, this work describes the folding process of the adenine aptamer by introducing one fluorophore in the core domain, which is directly involved in the recognition of adenine (10, 11). By using such a fluorescent construct, it has allowed us to directly monitor the folding of the aptamer at the level of the junction core domain as a function of Mg^{2+} ions and ligand cofactors. The introduction of a fluorophore in the adenine aptamer core domain is achievable at position 39 (Figure 1C) given that this residue is completely exposed to the solvent and not interacting with the remainder of the aptamer domain according to available crystal structures (10, 11). Since previous fluorescence studies reported that the introduction of a 2AP fluorophore at position 39 responds to the addition of Mg^{2+} ions and ligand (20, 21), clearly indicating that the local structure surrounding U39 experiences structural changes upon cofactors binding. The presence of such residues lying close to ligand binding sites and allowing the introduction of fluorophores is not common among riboswitches (1, 8, 9), since most of nucleotides located in ligand binding sites are involved in interactions with the bound ligand.

Based on our FRET data, the core domain of the adenine aptamer responds to the presence of magnesium ions by undergoing a structural transition (Figures 2 and 3A). According to our previous analysis, the addition of Mg^{2+} ions at saturating concentrations (i.e., 10 mM) results

in the formation of the loop-loop interaction (12). Thus, it is expected that the population having an E_{app} value of ~ 0.74 (Figure 3A, lower panel) corresponds to aptamers exhibiting a loop-loop interaction. As expected, aptamers having a disabled loop-loop interaction (L2 mutant) do not show the same FRET efficiency at 10 mM Mg^{2+} ions concentration (Figure 4A, lower panel, $E_{app} \sim 0.59$). We recently determined that L2 mutant aptamers rather adopts an intermediate structure in such conditions (St-Pierre and *al.*, in preparation), therefore suggesting that the core domain is also adopting an intermediate fold in the context of the mutant.

Our previous sm-FRET results suggested that the folding of the core domain is directly influencing the formation of the loop-loop interaction (12). Our present work further extends this notion by showing that the formation of the loop-loop interaction, in absence or presence of ligand, is directly modulating the structure of the core domain. Thus, these results suggest a model in which a mutual interaction exists between both the loop-loop interaction and the core domain, where the structure of each motif is influencing the state of the other motif (Figure 5). This process could be described as a "scissor-like" movement in which both folding motifs are modulated during the same global folding process. Given that the core domain is formed by residues involved in intramolecular interactions (Figure 5) as well as ligand interactions (10, 11), it indicates that a folding process is required to attain such a complex degree of interactions. Our sm-FRET results show that ligand binding is also directly modulating the

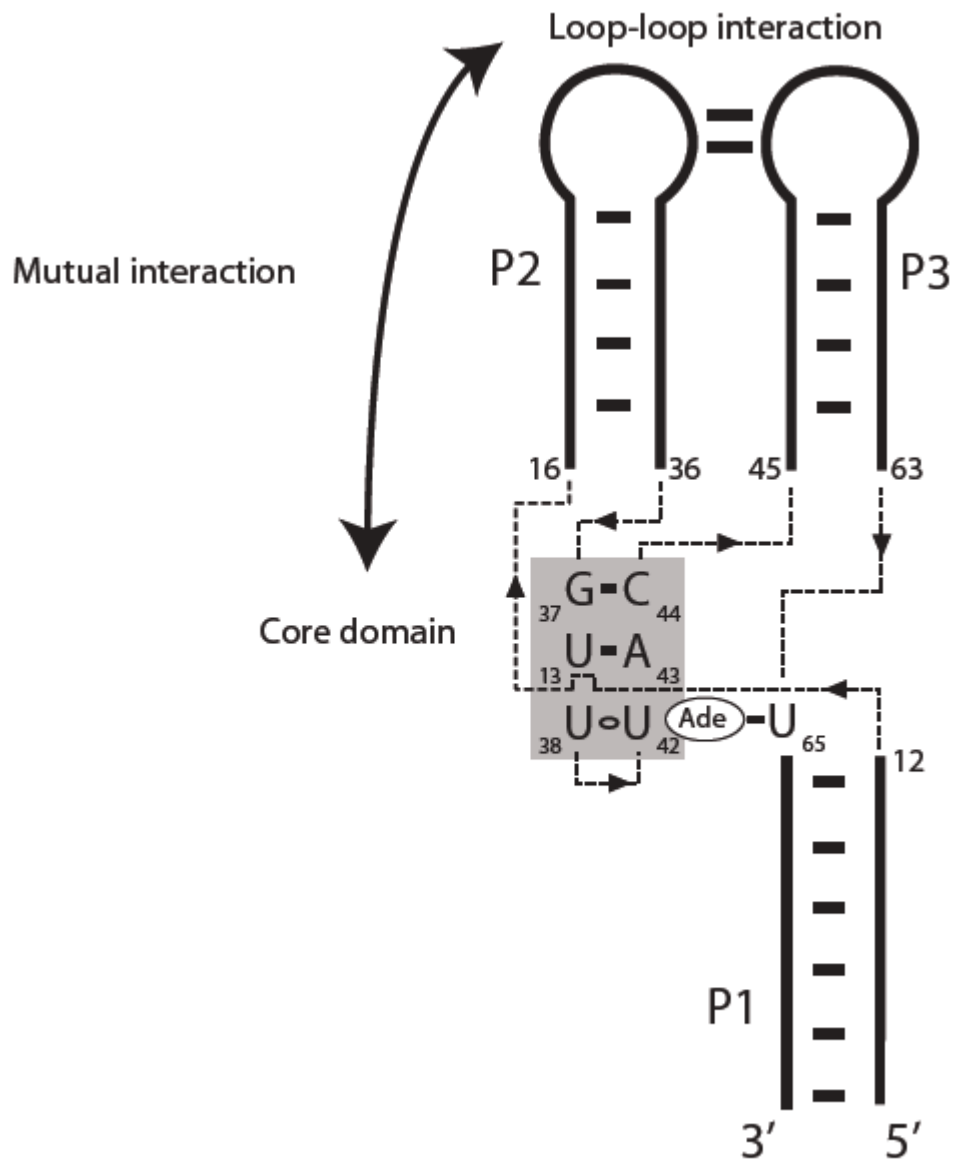


Figure 5 Schematic representation of the intramolecular interactions occurring between the aptamer and the bound adenine. Only selected residues are shown. Solid lines represent the molecule backbone and dotted lines indicate the polarity of the nucleic acids. A curve arrow represents a mutual interactions occurring between the formation of the loop-loop interaction and the folding of the core domain (highlighted in gray).

global structure of the core domain (Figure 3B), although to a lesser extent than the binding of Mg^{2+} ions (Figure 3A). Although it is still unknown whether ligand binding can be performed after the formation of the loop-loop interaction, the presence of a ligand-dependent folding process suggests a mechanism in which the structure of the aptamer domain could be reorganized locally at the level of the core region, in agreement with the change observed in sm-FRET assays (Figure 3B).

Our work provides the first observation at the single-molecule level of the reorganization of the core region upon ligand binding to the riboswitch. This should provides a useful method to characterize the folding process of more complex riboswitches or regulatory RNA molecules.

ACKNOWLEDGEMENT

The authors thank members of the Lafontaine laboratory and Dr Carlos Penedo for discussion.

FUNDING

National Sciences and Engineering Research Council of Canada (NSERC); Doctoral fellowship from NSERC (to P.S.P); Canadian Institute for Health Research (CIHR) New Investigator scholar (to D.A.L.).

REFERENCES

1. Serganov,A. and Nudler,E. (2013) A decade of riboswitches. *Cell*, **152**, 17–24.
2. Breaker,R.R. (2012) Riboswitches and the RNA World. *Cold Spring Harb Perspect Biol*, **4**, 1–15.
3. Garst,A.D., Edwards,A.L. and Batey,R.T. (2011) Riboswitches: structures and mechanisms. *Cold Spring Harb Perspect Biol*, **3**.
4. Mandal,M., Boese,B., Barrick,J.E., Winkler,W.C. and Breaker,R.R. (2003) Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria. *Cell*, **113**, 577–586.
5. Mandal,M. and Breaker,R.R. (2004) Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat Struct Mol Biol*, **11**, 29–35.
6. Lemay,J.F., Desnoyers,G., Blouin,S., Heppell,B., Bastet,L., St-Pierre,P., Masse,E. and Lafontaine,D.A. (2011) Comparative Study between Transcriptionally- and Translationally-Acting Adenine Riboswitches Reveals Key Differences in Riboswitch Regulatory Mechanisms. *PLoS Genet*, **7**, e1001278.
7. St-Pierre,P., McCluskey,K., Shaw,E., Penedo,J.C. and Lafontaine,D.A. (2014) Fluorescence tools to investigate riboswitch structural dynamics. *Biochim Biophys Acta*, **1839**, 1005–1019.
8. Montange,R.K. and Batey,R.T. (2008) Riboswitches: emerging themes in RNA structure and function. *Annu Rev Biophys*, **37**, 117–133.
9. Peselis,A. and Serganov,A. (2014) Themes and variations in riboswitch structure and function. *Biochim Biophys Acta*, **1839**, 908–918.
10. Serganov,A., Yuan,Y.R., Pikovskaya,O., Polonskaia,A., Malinina,L., Phan,A.T.,

- Hobartner,C., Micura,R., Breaker,R.R. and Patel,D.J. (2004) Structural Basis for Discriminative Regulation of Gene Expression by Adenine- and Guanine-Sensing mRNAs. *Chem Biol*, **11**, 1729–1741.
11. Batey,R.T., Gilbert,S.D. and Montange,R.K. (2004) Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature*, **432**, 411–415.
 12. 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.
 13. 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.
 14. 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.
 15. Lemay,J.F. and Lafontaine,D.A. (2006) [The adenine riboswitch: a new gene regulation mechanism.]. *Med Sci*, **22**, 1053–1060.
 16. Neupane,K., Yu,H., Foster,D.A., Wang,F. and Woodside,M.T. (2011) Single-molecule force spectroscopy of the add adenine riboswitch relates folding to regulatory mechanism. *Nucleic Acids Res*, **39**, 7677–7687.
 17. Lemay,J.F. and Lafontaine,D.A. (2007) Core requirements of the adenine riboswitch aptamer for ligand binding. *RNA*, **13**, 339–350.
 18. Stoddard,C.D., Gilbert,S.D. and Batey,R.T. (2008) Ligand-dependent folding of the three-way junction in the purine riboswitch. *RNA*, **14**, 675–684.
 19. Brenner,M.D., Scanlan,M.S., Nahas,M.K., Ha,T. and Silverman,S.K. (2010) Multivector fluorescence analysis of the xpt guanine riboswitch aptamer domain and the

- conformational role of guanine. *Biochemistry*, **49**, 1596–1605.
20. Gilbert,S.D., Stoddard,C.D., Wise,S.J. and Batey,R.T. (2006) Thermodynamic and Kinetic Characterization of Ligand Binding to the Purine Riboswitch Aptamer Domain. *J Mol Biol*, **359**, 754–768.
 21. Tremblay,R., Lemay,J.F., Blouin,S., Mulhbachter,J., Bonneau,E., Legault,P., Dupont,P., Penedo,J.C. and Lafontaine,D.A. (2011) Constitutive regulatory activity of an evolutionary-excluded riboswitch variant. *J Biol Chem*, **286**, 27406–27415.
 22. 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.
 23. Heppell,B., Blouin,S., Dussault,A.M., Mulhbachter,J., Ennifar,E., Penedo,J.C. and Lafontaine,D.A. (2011) Molecular insights into the ligand-controlled organization of the SAM-I riboswitch. *Nat. Chem Biol*, **7**, 384–392.
 24. Walter,N.G., Harris,D.A., Pereira,M.J. and Rueda,D. (2001) In the fluorescent spotlight: global and local conformational changes of small catalytic RNAs. *Biopolymers*, **61**, 224–242.
 25. Bokinsky,G. and Zhuang,X. (2005) Single-molecule RNA folding. *Acc Chem Res*, **38**, 566–573.
 26. Dalgarno,P.A., Bordello,J., Morris,R., St-Pierre,P., Dube,A., Samuel,I.D., Lafontaine,D.A. and Penedo,J.C. (2013) Single-molecule chemical denaturation of riboswitches. *Nucleic Acids Res*, **41**, 4253–4265.
 27. Haller,A., Rieder,U., Aigner,M., Blanchard,S.C. and Micura,R. (2011) Conformational capture of the SAM-II riboswitch. *Nat Chem Biol*, **7**, 393–400.
 28. Fiegand,L.R., Garst,A.D., Batey,R.T. and Nesbitt,D.J. (2012) Single-molecule studies of the lysine riboswitch reveal effector-dependent conformational dynamics of the aptamer

domain. *Biochemistry*, **51**, 9223–9233.

29. Haller,A., Altman,R.B., Souliere,M.F., Blanchard,S.C. and Micura,R. (2013) Folding and ligand recognition of the TPP riboswitch aptamer at single-molecule resolution. *Proc Natl Acad Sci U S A*, **110**, 4188–4193.

CHAPITRE 3

Discussion

Les travaux présentés dans cette thèse visent, au départ, à préciser le patron de repliement de l'aptamère liant l'adénine. L'impact des observations faites lors de ces travaux sur la compréhension du fonctionnement des riborégulateurs associés aux aptamères liant l'adénine a été discuté dans les chapitres 1 et 2. Il s'avère donc inutile de répéter cette discussion ici. Cependant, certains résultats obtenus lors de ces travaux entrent en conflit avec des informations disponibles dans la littérature scientifique. Les prochaines lignes serviront à placer les travaux présentés dans cette thèse dans le contexte de la littérature scientifique disponible. Ensuite, il sera question de l'importance que ces travaux peuvent avoir dans le domaine scientifique touchant l'ARN. Enfin, des expériences permettant de poursuivre les travaux entamés lors de cette thèse seront proposées.

Mise en contexte dans la littérature scientifique

Des efforts pour isoler l'intermédiaire de repliement de l'aptamère *add* ont été entrepris il y a plusieurs années. Un article paru en 2013, dans la revue scientifique *Nucleic Acids Research* (Dalgarno, Bordello et al. 2013), décrit des expériences de sm-FRET utilisant l'urée comme agent dénaturant pour observer les étapes limitantes dans le repliement de l'ARN. Les résultats présentés dans cet article supportent l'existence de l'état intermédiaire. Ils ont aussi permis de dire que le ligand induit un changement de conformation à l'aptamère. Ce phénomène est nommé *induced-fit* en anglais et pourrait se traduire par : mécanisme d'induction conformationnelle. Ce mécanisme d'action est opposé à celui appelé capture conformationnelle (*conformational capture*) au cours duquel le ligand stabilise un état que l'ARN peut adopter

par lui-même sans la présence de ce dernier. Les résultats présentés au chapitre 1 montrant que la DAP, mais pas le magnésium, entraîne un changement de conformation de l'aptamère L2 muté supportent l'hypothèse du mécanisme d'induction conformationnelle. Dans l'article de 2013, il a aussi été suggéré que l'intermédiaire pourrait être isolé en utilisant les bonnes concentrations d'ion divalent et d'urée. Malheureusement, les tentatives faites pour isoler l'intermédiaire en utilisant l'urée comme agent dénaturant se sont avérées infructueuses. De plus, il a été observé que l'urée interférait avec plusieurs techniques de détermination de la structure secondaire des ARN. Il aurait donc été impossible de caractériser précisément la structure secondaire de l'intermédiaire. Malgré ces inconvénients, la dénaturation des ARN observés en sm-FRET reste une avenue intéressante pour comprendre les dynamiques de repliement des riborégulateurs. Comme il l'a été remarqué dans un chapitre de livre auquel j'ai participé (Shaw, St-Pierre et al. 2014), l'utilisation d'agent dénaturant afin d'étudier la dynamique de repliement des ARN est une pratique très peu utilisée.

En sm-FRET, la position des fluorophores sur la molécule d'ARN définit ce qui est appelé un vecteur. Ainsi, dans l'étude présentée au chapitre 1, trois vecteurs ont été utilisés pour observer le repliement des aptamères *add* de type sauvage et L2 mutée. Une étude semblable avait déjà été réalisée sur l'aptamère de type sauvage *xpt* liant la guanine (Brenner, Scanlan et al. 2010). Comme il s'agit de deux aptamères liant les purines et que leur structure secondaire est très semblable, il était attendu que les résultats pour les molécules de type sauvage correspondent. Ce qui n'est pas tout à fait le cas. Premièrement, contrairement à ce qui a été observé pour l'aptamère *add* (Lemay, Penedo et al. 2006), aucun intermédiaire de repliement n'a été détecté pour l'aptamère *xpt*. Deuxièmement, les résultats obtenus pour l'aptamère *add* propose que les tiges P1 et P2 s'éloignent l'une de l'autre après l'ajout de magnésium alors que ceux réalisés avec l'aptamère *xpt* suggèrent le contraire. Troisièmement, le même type de différence est observable, dans une moins grande amplitude, pour l'interaction entre les tiges P1 et P3, mais cette fois-ci, les tiges se rapprocheraient pour l'aptamère *add* et s'éloigneraient pour l'aptamère *xpt*. Quelques hypothèses peuvent expliquer ces différences.

Une première hypothèse porte sur la localisation des fluorophores. En effet, même si elles permettent de suivre les mêmes interactions, les positions des fluorophores ne sont pas les mêmes dans les deux articles. Par exemple, les fluorophores permettant de suivre le repliement entre les tiges P2 et P3 se retrouve dans les boucles L2 et L3 pour l'aptamère *add*, alors qu'ils ont été positionnés directement sur les tiges dans le cas de l'aptamère *xpt*. Pour tester cette hypothèse, il faudrait positionner les fluorophores au même endroit pour les deux aptamères. Cependant, des limitations techniques empêchent de faire ce test simplement. En effet, avec les techniques utilisées au cours de cette thèse, l'incorporation des fluorophores se fait uniquement sur des résidus thiamine modifiés. Comme la thiamine forme les mêmes interactions de type Watson-Crick que l'uridine, seuls les U d'une séquence d'ARN peuvent être marqués avec un fluorophore. Afin de mettre les fluorophores aux mêmes positions pour les deux aptamères, il faut nécessairement modifier la séquence d'un aptamère. Le plus simple serait de modifier les tiges P2 et P3 de l'aptamère *add* afin de pouvoir y incorporer les fluorophores aux mêmes positions que celles utilisées pour l'aptamère *xpt*. Par contre, il faudrait s'assurer par SHAPE que la structure et l'affinité pour l'adénine de l'aptamère ne sont pas trop affectées par ces changements.

Une deuxième possibilité expliquant les différences observées entre les aptamères *add* et *xpt* concerne la nature du ligand. Le principe de fonctionnement des riborégulateurs indique que le ligand affecte le repliement de l'aptamère. Il est donc possible que la nature du ligand permette ou non de voir un intermédiaire de repliement. Il est possible que le pont hydrogène supplémentaire présent dans l'interaction G-U entre la guanine et l'aptamère *xpt* affecte le repliement de l'aptamère. Ceci pourrait empêcher la formation de l'intermédiaire ou favoriser le repliement complet plus rapidement empêchant l'observation de l'intermédiaire. Cette hypothèse est plus facile à tester que la première. En effet, il est connu que la spécificité de l'aptamère *add* peut être changée en modifiant un seul nucléotide de la jonction J3-1 (Lemay et Lafontaine 2007). La modification du U à la position 65 pour un C n'affecte pas la structure

de l'aptamère, mais permet de le faire passer d'un aptamère liant l'adénine à un spécifique à la guanine. Des expériences de sm-FRET avec des aptamères possédant cette mutation révéleraient si la nature du ligand affecte la présence de l'intermédiaire de repliement.

Dans le chapitre 1, il est montré que la mutation de la L2 permet d'isoler la conformation intermédiaire en ajoutant uniquement du magnésium. Il a ensuite été observé que l'ajout du ligand DAP entraîne la formation de l'état repliée. Ce qui suggère que le ligand peut se lier à l'état intermédiaire qui ne possède pas d'interaction boucle-boucle, bien que cette interaction soit plus difficile. En effet, ces expériences ont été réalisées avec la DAP qui lie l'aptamère avec une affinité 30 fois supérieure à celle de l'adénine (Mandal et Breaker 2004). Cependant, une étude de spectroscopie de force atomique en mode molécule unique suggère que la formation de l'interaction boucle-boucle est une étape préalable à la liaison du ligand (Neupane, Yu et al. 2011).

La figure 1 est tirée de cet article où la spectroscopie de force atomique est utilisée afin d'observer le repliement de l'aptamère liant les purines. Il y est démontré que la présence d'adénine n'introduit pas de différence dans le patron de repliement de l'aptamère avant la formation de la conformation appelée P1_U. Les résultats de cet article suggèrent fortement que la conformation P1_U correspond à un état où seule la tige P1 n'est pas formée. Selon ces résultats, il semble donc que l'interaction L2-L3 soit préalable à la liaison du ligand. En combinant l'ensemble de ces résultats, il semble plus juste de dire que l'interaction boucle-boucle aide à positionner le cœur de l'aptamère dans une conformation ayant une meilleure affinité pour le ligand, mais que cette interaction n'est pas absolument nécessaire à la liaison.

Les résultats de spectroscopie de force atomique ne montrent pas l'existence d'un intermédiaire de repliement correspondant à l'empilement des tiges P1 et P3. Afin de comprendre pourquoi les résultats obtenus en sm-FRET et en spectroscopie de force atomique

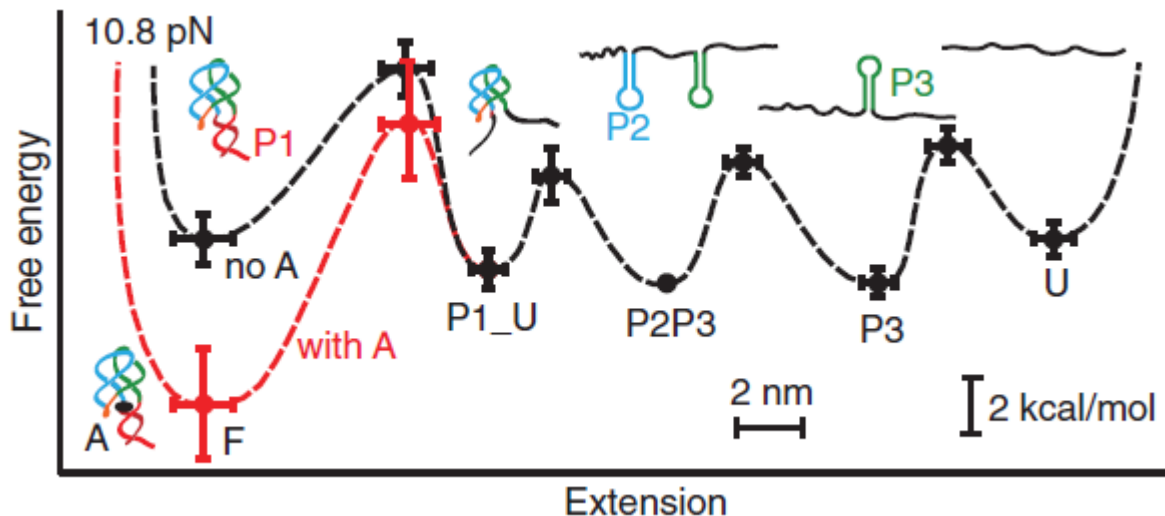


Figure 12 Tiré de (Neupane, Yu et al. 2011) Patron de repliement de l'aptamère *add* déterminé par spectroscopie de force atomique. Les étapes clés du patron de repliement natif en 5 étapes ont été reconstruites à partir d'analyse de transition entre deux états pour chaque transition. L'énergie et l'extension sont déterminées en référence à l'état P2P3. Les barres d'erreur correspondent à l'écart type. Les lignes pointillées indiquent le patron théorique en présence (noir) et en absence (rouge) d'adénine. Les dessins montrent les caractéristiques structurales associées à chaque état.

ne mènent pas à l'identification des mêmes structures, il est important de réaliser que, lors de ces deux expériences, la formation de la tige P1 est grandement affectée. En effet, lors des expériences réalisées par Neupane (Neupane, Yu et al. 2011), la tige P1 ne peut être que la dernière structure secondaire à se former en raison de la force appliquée sur la molécule. De l'autre côté, la tige P1 est toujours formée lors des expériences de sm-FRET en raison de la grande quantité de paires de base qui la maintiennent formée en absence de force appliquée à ces extrémités. Conséquemment, ces deux techniques ne peuvent pas donner exactement la même information sur le repliement de l'aptamère *add*. Il est alors possible que la conformation I, observée en sm-FRET, n'existe que lorsque la P1 est formée et que la conformation P1_U, observée en spectroscopie, n'existe que lorsque la P1 n'est pas formée.

Pour l'instant, il est impossible d'affirmer que cette explication est la bonne. Par contre, elle permet d'expliquer pourquoi deux expériences mènent à des conclusions contradictoires. Si cette hypothèse est acceptée, il faut alors déterminer quelle situation représente le plus ce qui se produit lors du repliement naturel de l'aptamère *add*. Le plus simple serait de dire que l'étude montrant que le P1 est la dernière structure secondaire formée (Neupane, Yu et al. 2011) représente fidèlement la réalité. En effet, lors de la transcription de l'aptamère, la partie 3' de la tige P1 est la dernière partie produite. Donc la P1 devrait logiquement être la dernière structure à se former. Cependant, cette étude montre aussi que la P3 se forme avant la P2 alors que cette dernière devrait être transcrite en premier et donc se former la première. Il apparaît donc que l'étude de spectroscopie de force atomique ne reflète pas fidèlement le repliement natif de l'ARN. De plus, il a été démontré que le riborégulateur *add* fonctionne de façon thermodynamique (Lemay, Desnoyers et al. 2011). C'est-à-dire que le riborégulateur peut être transcrit en entier et quand même lier le ligand. Ceci implique que le cœur de l'aptamère reste accessible au ligand même si la tige P1 peut se former. En forçant la P1 à se défaire en premier, les expériences de spectroscopie de force atomique ne permettent pas de voir ce phénomène. Cette critique ne veut pas dire pour autant que les expériences de sm-FRET représentent fidèlement la réalité. En effet, il est évident que la tige P1 n'est pas toujours formée lorsque les autres structures secondaires et tertiaires apparaissent. Pour l'instant, il semble plus sage de dire que le repliement naturel de l'aptamère *add* se fait selon un patron hybride entre ceux proposés jusqu'à maintenant.

Impact potentiel dans le domaine de l'ARN

La structure secondaire de l'aptamère *add* est un motif d'ARN appelé jonction à trois voies. Comme décrit dans l'introduction, ce motif est composé de trois tiges d'ARN double-brin organisées entre elles par des régions d'ARN simple brin. Ce motif d'ARN est récurrent dans plusieurs types d'ARN et est associé à plusieurs fonctions. Par exemple, il est retrouvé chez d'autres riborégulateurs tels que ceux liant la glycine (Serganov et Patel 2009), la SAM (Wang

et Breaker 2008), le GlcN6P (Brooks et Hampel 2009), la TPP (Serganov, Polonskaia et al. 2006), la vitamine B12 (Nahvi, Barrick et al. 2004) et le c-di-GMP (Kulshina, Baird et al. 2009). Ce motif permet aussi de décrire la structure du ribozyme en tête de marteau (Hammann, Luptak et al. 2012) et se retrouve dans les ARN ribosomiaux (Ban, Nissen et al. 2000) et bien plus. Un article publié en 2006 dresse la liste de plus d'une trentaine de jonctions trois voies dont le cristal est disponible à ce jour et tente de les classer afin de faire ressortir leurs caractéristiques particulières (Lescoute et Westhof 2006)

Les travaux présentés dans cette thèse pourraient aider à comprendre le repliement et le fonctionnement de tous ces ARN. En effet, il est difficile d'étudier les repliements de très longs ARN. Il convient souvent mieux d'étudier de petits ARN et d'extrapoler les résultats obtenus sur des ARN plus complexes. Le riborégulateur *add* remplit parfaitement ce rôle en raison de sa petite taille et de la régularité de son repliement. Dans l'article mentionné précédemment (Lescoute et Westhof 2006), l'aptamère liant les purines est un représentant d'une famille de jonction à 3 voies comprenant 17 membres. Il est donc envisageable que le patron de repliement de l'aptamère *add* permette de comprendre le repliement des 16 ARN plus complexes de la même famille. En plus d'être un excellent modèle structural pour les jonctions à 3 voie, l'aptamère *add* serait aussi un modèle d'étude pour le mécanisme de reconnaissance du ligand par induction conformationnelle.

Perspectives

Les études présentées dans cette thèse se sont attardées au repliement de l'aptamère du riborégulateur *add*. Ces études et toutes celles réalisées sur cet aptamère ont grandement aidé à la compréhension de la structure et du mécanisme de fonctionnement du riborégulateur liant l'adénine. Cependant, les riborégulateurs sont composés d'un aptamère et d'une plateforme d'expression et cette dernière est rarement prise en compte lors d'étude structurale des

ribo-régulateurs. Les chercheurs intéressés par la structure des ARN ne négligent pas d'étudier la plateforme d'expression par choix, mais plutôt en raison de difficulté technique. Par exemple, il n'a pas été possible d'obtenir de cristal de ribo-régulateur complet avec la plateforme d'expression. Cette difficulté technique vient probablement du fait que l'introduction de la plateforme d'expression permet l'existence de plusieurs conformations de l'ARN empêchant une cristallisation efficace. Aussi, la taille de la majorité des ribo-régulateurs excède celle habituellement utilisée lors d'expériences de Résonance Magnétique Nucléaire (RMN) (Peselis et Serganov 2014). La même limitation s'applique aux expériences de sm-FRET. En effet, la façon la plus efficace de positionner précisément les deux fluorophores requis est de synthétiser les ARN chimiquement. La synthèse chimique n'étant pas efficace pour de longs ARN, il n'est souvent pas possible d'étudier des ribo-régulateurs complets avec cette technique. Le ribo-régulateur SAM de classe II est le seul ribo-régulateur complet à avoir été étudié de cette façon en raison de sa très petite taille (52 nt) (Haller, Rieder et al. 2011).

En considérant sa plateforme d'expression, le ribo-régulateur *add* a une taille de 120 nucléotides. Ce qui en fait un des plus petits ribo-régulateurs connus. Au début des travaux entrepris pour cette thèse, il n'était pas possible de créer un ARN de cette taille pour des études en sm-FRET. Cependant, l'efficacité de la synthèse chimique des ARN s'est accrue depuis ce temps il serait maintenant possible de produire un tel ARN. Ces développements rendent accessible l'étude du repliement du ribo-régulateur *add* en entier par sm-FRET. Une telle étude fait directement suite à celles réalisées au cours de cette thèse. La quantité de données disponibles sur la structure et le repliement de l'aptamère de ce ribo-régulateur faciliteraient la prise de décision quant aux expériences à réaliser de même que l'interprétation des résultats. Une telle étude permettrait, entre autres, de suivre la formation de la tige P1 dans un environnement plus près de la réalité que ne l'ont permis les expériences réalisées à ce jour. En effet, il serait relativement simple de positionner les fluorophores aux extrémités de la tige P1 pour observer sa formation. Il serait aussi possible de déterminer l'effet du ligand sur la formation de la tige P1. Réaliser cette étude ne demanderait l'acquisition d'aucun matériel

(outre les ARN synthétiques) ni de connaissances techniques supplémentaires à ceux utilisées au cours de cette thèse. Ceci rend cette perspective très accessible et réaliste à court terme.

Les connaissances acquises au cours des travaux présentés ici permettent de poursuivre un objectif beaucoup plus ambitieux. Grâce à l'expertise développée sur la technique de sm-FRET, il serait possible de développer une méthode permettant l'étude du repliement des ARN en cours de transcription. Une telle méthode répondrait à un besoin pressant dans le domaine de recherche de la structure de l'ARN. En effet, depuis toujours la structure de l'ARN est étudiée après sa transcription complète dans sa conformation la plus stable thermodynamiquement. Il est cependant évident que la structure de l'ARN change au cours de sa transcription, qu'une même séquence d'ARN peut adopter différentes conformations et que la conformation thermodynamique n'est pas nécessairement celle la plus présente naturellement (Woodson 2010). Depuis quelques années, quelques techniques ont été développées afin d'étudier la structure de l'ARN en cours de transcription avec plus ou moins de succès (Woodson 2010). La découverte d'une technique efficace permettant l'étude de la structure des ARN en cours de transcription serait une véritable révolution dans le domaine. Ceci rendrait possible l'étude des structures d'ARN se rapprochant le plus de celles retrouvées dans les cellules. Il serait alors possible de parler de structures natives des ARN. Également, nombre de conformation que peut adopté un ARN naturellement pourrait être déterminé de même que leur distributio.

Le sm-FRET se prête bien l'étude de l'ARN en cours de transcription puisque cette technique permet d'observer différentes conformations d'ARN en même temps, ne requiert pas la synchronisation du repliement et permet de suivre différents paramètres de repliement. Sans entrer dans les détails, le concept de la technique proposée est d'observer un complexe transcriptionnel en sm-FRET. Des fluorophores seraient positionnés à des endroits précis sur l'ARN grâce au principe de transcription par étape (*stepwise transcription*) (Sohn, Shen et al. 2003). L'ADN des complexes de transcription serait utilisé comme ancre pour isoler les

molécules lors de l'observation. Le développement de cette approche est un projet ambitieux puisqu'il faut combiner deux techniques difficiles à maîtriser, le sm-FRET et la transcription par étape. Développer une telle approche demande énormément d'effort. Par exemple, cette thèse devait à l'origine porter sur le développement de cette approche. Cependant, de nombreux problèmes techniques ont retardé l'avancement de ces travaux et ont forcé une révision du projet de recherche. Par contre, cette idée n'a pas été abandonnée et, en plus de l'auteur de cette thèse, trois étudiants du laboratoire du docteur Daniel Lafontaine travaillent à faire avancer ce projet. Des développements récents permettent de croire que cette technique sera fonctionnelle prochainement. Il sera alors pertinent de comparer le patron de repliement natif de l'aptamère *add* avec celui déterminé thermodynamiquement notamment grâce aux résultats obtenus au cours de cette thèse.

Conclusion

La compréhension du fonctionnement d'un système complexe passe souvent par l'étude de modèle simplifié de ce système. En raison de sa petite taille, de sa structure simple, de sa stabilité, et de la constance qu'il affiche lors de son repliement, le riborégulateur *add* s'est avéré être un excellent modèle d'étude. Les travaux réalisés avec ce riborégulateur, dont ceux présentés dans cette thèse, permettent de mieux comprendre le mécanisme d'action des riborégulateurs, le comportement des structures d'ARN dans différentes conditions et l'importance du motif d'ARN en jonction à 3 voies. Il est même possible que des expériences faites avec le riborégulateur *add* permettent, un jour, de mieux comprendre l'origine de la Vie.

En effet, les ARN semblent être des molécules essentielles à l'existence de la vie telle qu'elle est présente sur la Terre de nos jours. Cette observation, de même que la découverte de l'activité catalytique de l'ARN, notamment celle du ribosome, et au fait que de nombreux cofacteurs sont composés en partie de nucléotides ont permis de proposer l'hypothèse d'un Monde à ARN. Cette hypothèse est l'une des plus populaires pour expliquer l'apparition de la Vie (Martin, Unrau et al. 2015). Elle suggère que la Vie a émergé d'un environnement où l'ARN serait le précurseur de toutes les macromolécules biologiques. Certains scientifiques sont septiques à l'idée que l'ARN, telle qu'elle est observée aujourd'hui, soit le point de départ de l'existence de la Vie sur Terre. Par contre, ils sont généralement d'accord pour dire qu'à un moment dans l'Histoire, l'ARN devait remplir simultanément le rôle de support de l'information génétique et d'effecteur de cette information (Cech 2012).

Il semble évident que l'ARN a dû interagir avec d'autres molécules pour permettre l'évolution vers la Vie telle qu'on la connaît aujourd'hui. C'est exactement ce que font les riborégulateurs pour permettre le contrôle de l'information génétique. Ainsi, l'étude des riborégulateurs est peut-être une fenêtre qui permet de voir comment l'ARN primordial fonctionnait. De plus, la

structure des ARN étant centrale à leur fonctionnement, il est évident que toutes études s'intéressant à la structure des ARN peuvent éventuellement fournir des informations sur certaines caractéristiques du Monde à ARN. Si l'étude de la structure de l'ARN et l'étude des riborégulateurs peuvent permettre d'étoffer ou de confirmer l'hypothèse du Monde à ARN, il est fort probable que les études structurales menées sur un riborégulateur modèle jouent un rôle important dans cette démonstration. Après tout, il est toujours permis de rêver.

Annexe

Fluorescence tools to investigate riboswitch structural dynamics.

Fluorescence tools to investigate riboswitch structural dynamics

Patrick St-Pierre¹, Kaley McCluskey², Euan Shaw², Penedo, J.C^{2,3*} and Lafontaine, D. A^{1*}

¹RNA Group, Department of biology, Faculty of Science, Université de Sherbrooke, Sherbrooke, QC, Canada, J1K 2R1; ²SUPA, School of Physics and Astronomy University of St Andrews, St Andrews, Fife, United Kingdom, KY16 9SS; ³Biomedical Sciences Research Complex, University of St Andrews, St Andrews, Fife, United Kingdom, KY16 9SS;

* To whom correspondence should be addressed.

daniel.lafontaine@usherbrooke.ca
jcp10@st-andrews.ac.uk

ABSTRACT

Riboswitches are novel regulatory elements that respond to cellular metabolites to control gene expression. They are constituted of highly conserved domains that have evolved to recognize specific metabolites. Such domains, so-called aptamers, are folded into intricate structures to enable metabolite recognition. Over the years, the development of ensemble and single-molecule fluorescence techniques has allowed to probe most of the mechanistic aspects of aptamer folding and ligand binding. In this review, we summarize the current fluorescence toolkit available to study riboswitch structural dynamics. We first describe those methods based on fluorescent nucleotide analogues, mostly 2-aminopurine (2AP), to investigate short-range conformational changes, including some key steady-state and time-resolved examples that exemplify the versatility of fluorescent analogues as structural probes. The study of long-range structural changes by Förster resonance energy transfer (FRET) is mostly discussed in the context of single-molecule studies, including some recent developments based on the combination of single-molecule FRET techniques with controlled chemical denaturation methods.

1. Introduction

The last decade has witnessed the discovery of an astonishing variety of non-protein coding RNA sequences that perform a diverse range of functional roles in living organisms. An emerging theme arising from a combination of complementary biochemical, biophysical and crystallographic studies suggests that this complex range of RNA functionalities stems, at least in part, from their ability not only to fold into a specific 3D structure but also to alter its conformation in response to subtle changes in specific cellular signals such as the presence of ions, proteins, small molecules and temperature variations (Breaker 2012, Serganov and Nudler 2013). RNA's ability to sense and adapt to the cellular environment as an intrinsic part of its functional role is exemplified by the ever-growing family of metabolite-sensing regulatory mRNAs (Breaker 2012, Serganov and Nudler 2013). These RNA elements, so-called riboswitches, are mostly located within the 5' untranslated region of prokaryotic messenger RNA (mRNA) and usually regulate gene expression in a *cis*-fashion by directly binding to specific cellular cues without protein assistance (Montange and Batey 2008, Breaker 2012, Serganov and Nudler 2013). To understand the regulatory mechanism of riboswitches, it is therefore of primordial importance to study the secondary and tertiary structures of these RNAs, as well as their folding pathway. Crystallography of riboswitch aptamers provided the main source of information for the structure of most riboswitches (Serganov and Nudler 2013). However, the RNA needs to be in a highly stable conformation to be crystallized. Because of this limitation, crystallography mainly provided information about riboswitch aptamers bound to their cognate ligand. The examination of aptamer crystals reveals that RNA molecules make extensive contacts with the bound metabolite, mostly resulting in the latter being completely buried within the aptamer. This suggests that the riboswitch needs to adopt an open conformation to allow efficient ligand recognition and that a folding mechanism must occur to allow the formation of a riboswitch-ligand complex. This review will describe different fluorescent techniques that have allowed researchers to examine, at least in part, the folding pathways of various aptamers as a response to ligand binding.

These techniques also allowed the observation of the bound ('closed') conformation of aptamer domains and provided a way to identify what conformational changes are involved in ligand binding. Finally, these tools have allowed to understand the dynamics of these RNAs by providing information about folding and unfolding rates as well as direct measurements of the binding constant of aptamers.

2. Fluorescent analogs of adenine and cytosine: spectroscopic properties of 2-aminopurine and pyrrolo-dC

Fluorescent molecules that have the ability to form hydrogen bonds and preserve the self-recognition properties of natural nucleobases have played a pivotal role in our understanding of DNA and RNA structures and how these are altered by environmental conditions including ions, denaturant agents, temperature and small molecules (Rist and Marino 2002, Wilhelmsson 2010). Among these fluorescent nucleobase analogs, 2-aminopurine (2AP) (Fig. 1a) has been extensively used to investigate interactions between DNA and RNA regions and to follow their association to other biomolecules including proteins and large ribonucleoprotein complexes (Hershey and Chase 1952, Guest, Hochstrasser and al. 1991, Stivers 1998, Rachofsky, Osman and al. 2001, Rist and Marino 2002, Wilhelmsson 2010). 2AP is an analogue of adenine (Fig. 1a) that, like adenine (Fig. 1b), forms stable base pairs with uracil (RNA) and thymine (DNA) (Fig. 1c and Fig. 1d) but also moderately stable base pairs with cytosine in a wobble configuration (DNA or RNA) (Fig. 1e). Although 2AP exhibits a low extinction coefficient ($\Sigma \sim 6000 \text{ M}^{-1} \text{ cm}^{-1}$), the use of 2AP in fluorescence studies of nucleic acid structure is very attractive because its photophysical properties, in particular its fluorescence quantum yield, lifetime and anisotropy, are extremely sensitive to local changes in the surrounding environment (Hershey and Chase 1952, Guest, Hochstrasser and al. 1991, Holmen, Norden and al. 1997, Stivers 1998, Jean and Hall 2001, Rachofsky, Osman and al. 2001, Jean and Hall 2002). Moreover, the position of the lowest energy absorption band (305 nm) lies outside the absorption band of the nucleic acid structure, thus making it possible to

selectively excite 2AP without affecting natural nucleobases. This is an important advantage that enables 2AP to act as a reporter of localized structural changes when compared to other techniques such as circular dichroism and absorption spectroscopy that monitor the entire RNA present in solution (Rist and Marino 2002, Wilhelmsson 2010).

The high quantum yield of 2AP free in solution (~0.68) is drastically quenched when incorporated into nucleic acids (Hershey and Chase 1952, Jean and Hall 2001, Wilhelmsson 2010). Several spectroscopic and computational studies have investigated the molecular basis of this quenching mechanism as a function of position within the nucleic acid structure,

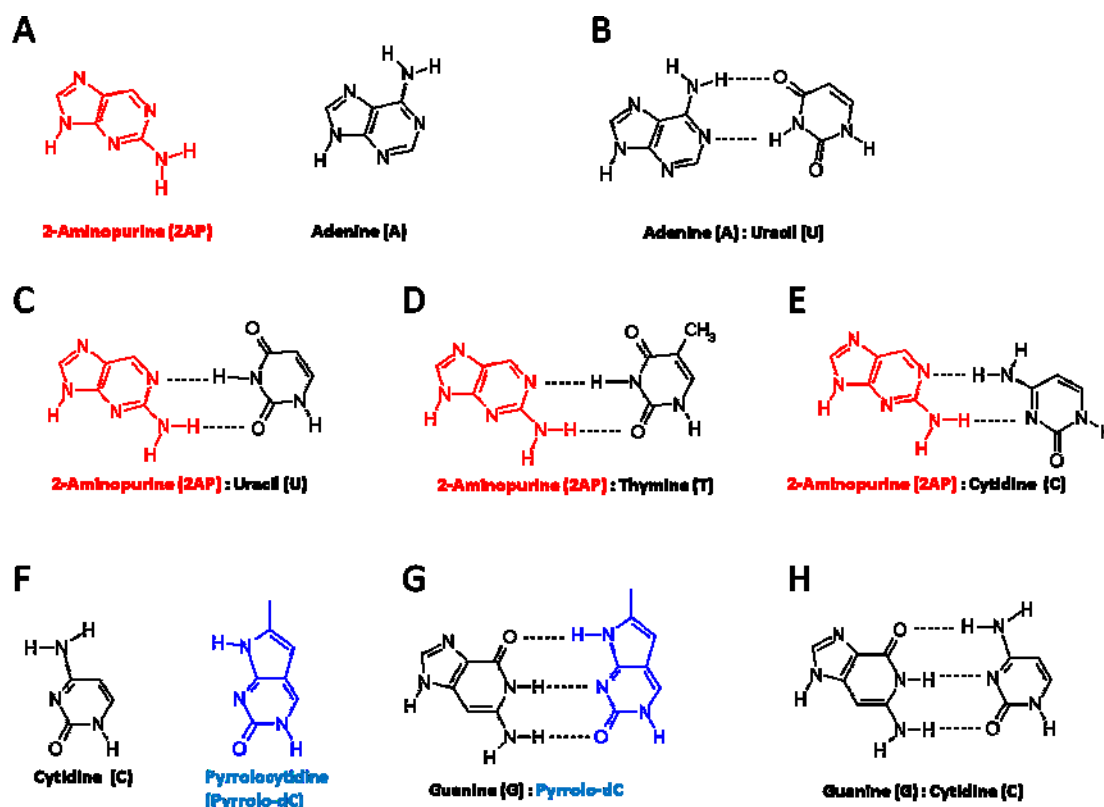


Fig. 1. Chemical structures and summary of base pairing interactions involving nucleotide analogs 2AP and Pyrrolo-dC.

sequence, nature of the adjacent bases and stacking geometries (Hershey and Chase 1952, Holmen, Norden and al. 1997, Stivers 1998, Jean and Hall 2001, Rachofsky, Osman and al. 2001, Jean and Hall 2002, Rai, Cole and al. 2003, Hardman and Thompson 2006). In general, the conformational motion of the stacked bases surrounding the fluorophore is known to be crucial for the quenching mechanism (Hershey and Chase 1952, Holmen, Norden and al. 1997, Stivers 1998, Jean and Hall 2001, Rachofsky, Osman and al. 2001, Jean and Hall 2002, Rai, Cole and al. 2003, Hardman and Thompson 2006). In fact, it has been shown that the emission intensity of 2AP in duplex DNA increases upon cooling and becomes close to that of free 2AP at low temperatures (O'Neill and Barton 2004). In contrast, hydrogen bonding and base pairing do not usually affect 2AP fluorescence emission significantly (Jean and Hall 2001, Rai, Cole and al. 2003). This sensitivity to the stacked micro-environment has made 2AP an ideal probe to investigate DNA and RNA structural changes involving nucleotide base flipping (Holz, Klimasauskas and al. 1998, Neely, Daujotyte and al. 2005), nucleic acid interactions (Edwards, Klein and al. 2007), local melting (Crick 1970) and denaturation (Ballin, Bharill and al. 2007). 2AP fluorescence has also been extensively applied to study a variety of DNA and RNA processing enzymes (Bloom, Otto and al. 1993, Dunlap and Tsai 2002, Hariharan and Reha-Krantz 2005, Neely, Daujotyte and al. 2005, Finger, Patel and al. 2013) and to elucidate the relationship between folding and catalysis in small nucleolytic RNA enzymes (ribozymes) (Menger, Tuschl and al. 1996, Walter, Chan and al. 2001, Walter, Harris and al. 2001, Harris, Rueda and al. 2002). However, as discussed in the next sections, it is in the study of riboswitch folding and regulatory function where, in recent years, the potential of 2AP substitutions at specific regions of the mRNA sequence has been fully exploited (Borovok, Gorovitz and al. 2006, Lang, Rieder and al. 2007, Rieder, Lang and al. 2007, Blouin, Chinnappan and al. 2010, Haller, Souliere and al. 2011, Dambach, Sandoval and al. 2015).

A fluorescent derivative of cytidine, pyrrolocytosine (pyrrolo-dC) (Fig. 1f), was introduced in the late 90s following earlier studies on a fluorescent nucleobase named furano-dT (Liu and Martin 2001, Berry, Jung and al. 2004). It was found that furano-dT incorporated into an

oligonucleotide was converted into pyrrolo-dC after ammonia treatment at the final stage of the solid-phase synthesis (Berry, Jung and al. 2004). It has been shown that DNA duplexes containing the pyrrolo-dC:G base pair (Fig. 1g) were as stable as unmodified ones (Fig. 1h) (Berry, Jung and al. 2004). Pyrrolo-dC is commercially available and exhibits its absorption maximum at 350 nm with an extinction coefficient very similar to that of 2AP ($\Sigma \sim 5900 \text{ M}^{-1} \text{ cm}^{-1}$) and a fluorescence maximum around 473 nm when incorporated into a DNA duplex (Liu and Martin 2001). The fluorescence quantum yield of the pyrrolo-dC monomer has been reported to be 0.2 and decreases to values of 0.03 in dsDNA and 0.05 in dsRNA (Liu and Martin 2001). In contrast to 2AP, very little is known about the influence of adjacent bases on the emission properties of pyrrolo-dC. Nevertheless, the use of pyrrolo-dC to investigate DNA and RNA structure continues to grow (Hardman and Thompson 2006, Zhang and Wadkins 2009). For example, pyrrolo-dC has been used to investigate the HIV-1 polypurine tract (PPT) using DNA/RNA hybrids (Dash, Rausch and al. 2004) and to characterize the mechanism of DNA repair by human alkyltransferase (Zang, Fang and al. 2005). Furthermore, the combination of 2AP and pyrrolo-dC in a molecular beacon-like fashion to investigate nucleic acid hybridization has already been reported (Marti, Li and al. 2006). In the context of metabolite-sensing mRNA sequences, the use of pyrrolo-dC substitutions to report folding and/or ligand binding has been very limited, and to date, only a single study has been reported applying pyrrolo-dC to investigate the conformational heterogeneity of the SAM-I riboswitch ON state (Huang, Kim and al. 2012). It is clear that to exploit the full potential of pyrrolo-dC as a probe for RNA structure, more efforts are needed; particularly time-resolved and computational data to determine the exact nature of the quenching mechanism and the influence of the surrounding sequence. In the next sections, we will focus on discussing specific examples on the application of the more widely employed 2AP analog to unravel the structural dynamics, ligand-binding properties and gene regulation mechanisms of metabolite-sensing mRNAs.

3. Strategies to identify structure-responsive 2-aminopurine substitutions within regulatory mRNA sequences

Monitoring mRNA folding by following the changes in 2AP steady-state fluorescence emission caused by variations in the stacked positioning of the neighboring bases is the most common application of 2AP labeled mRNAs (Rist and Marino 2002, Wilhelmsson 2010, Finger, Patel and al. 2013). This is mostly due to the simplicity of the steady-state assay that requires only a conventional fluorimeter, commonly present in most laboratories, and to the fact that mRNA sequences labeled with 2AP at specific positions are commercially available at moderate cost. With the exception of the adenine riboswitch for which it has been shown that the 2AP analogue can mimic the natural adenine ligand ($K_{Dapp} \sim 251$ nM) (Borovok, Gorovitz and al. 2006, Dambach, Sandoval and al. 2015), most studies of riboswitch structural changes relied on 2AP replacement of specific adenine bases or non-conserved nucleotides along the mRNA sequence (Gelfand, Mironov and al. 1999, Lang, Rieder and al. 2007, Lemay and Lafontaine 2007, Heppell and Lafontaine 2008, Heppell, Mulhbachter and al. 2009, Blouin, Chinnappan and al. 2010, Haller, Rieder and al. 2011, Haller, Souliere and al. 2011, Heppell, Blouin and al. 2011, Tremblay, Lemay and al. 2011, Garst, Porter and al. 2012). Selection of appropriate 2AP substitutions to monitor secondary and tertiary RNA folding while still preserving the functional properties of the riboswitch can be carried out by detailed inspection of the crystal structure or sequence non-conservation (Borovok, Gorovitz and al. 2006, Heppell, Blouin and al. 2011, Dambach, Sandoval and al. 2015). However, in the absence of crystallographic data, the choice of 2AP labeling positions is a tedious and expensive task involving a quasi-random sampling of non-conserved regions of mRNA sequences to find folding- or binding-sensitive locations. Remarkably, a strategy recently developed by Micura and coworkers (Souliere, Haller and al. 2011, Souliere and Micura 2014) provides a means to quantitatively estimate which 2AP substitutions may act as sensitive reporters of local structural changes in a given riboswitch. The method explores a direct correlation found between the 2'-OH flexibility of nucleotides reported by 2'-hydroxyl acylation analyzed by primer extension (SHAPE) and the fluorescence response observed

following 2AP substitution at highly SHAPE-responsive nucleotides. SHAPE is a widely employed method to explore secondary and tertiary interactions between nucleotides within an mRNA sequence that takes advantage of the chemical coupling between the ribose 2'-OH nucleophile and electrophilic groups such as N-methylisatoic anhydride (NMIA) or benzoyl cyanide (BzCN) to form 2'-O-adducts (Wilkinson, Merino and al. 2006, Mortimer and Weeks 2009). The extent of adduct formation depends on the local flexibility of the nucleotide position and thus SHAPE provides a means to differentiate less flexible and reactive nucleotides engaged in stable base-pairing from those in unconstrained locations. Using the BzCN SHAPE reagent in three model riboswitches (adenine, SAM-II and preQ1-class II) it was found that variations in SHAPE-reactivity profiles, with respect to a control experiment in the absence of BzCN, correlate strongly with an increase or decrease in 2AP fluorescence emission, respectively (Souliere, Haller and al. 2011, Souliere and Micura 2014). Moreover, the quantitative variation in SHAPE reactivity and 2AP emission showed a positive correlation in all three RNAs investigated, suggesting that nucleotides that are more flexible usually correspond with positions that are looped out or partially unstacked (Souliere, Haller and al. 2011, Souliere and Micura 2014). Thus, the SHAPE analysis might be used to predict not only suitable 2AP substitutions to report folding but also whether we should expect an increase or a decrease in 2AP fluorescence output.

4. Monitoring local conformational changes in purine-sensing riboswitches using 2-aminopurine substitutions

The ability of 2AP variants to provide real-time information about local conformational changes is exemplified by early studies on the aptamer domain of the adenosine deaminase *add* riboswitch from *Vibrio vulnificus* that controls gene expression at the translational level (Rieder, Lang and al. 2007). The crystal structure of this RNA-ligand complex revealed an intricate network of interactions that result in complete encapsulation of the ligand (Fig. 2a) (Serganov, Yuan and al. 2004). For the ligand-free state in the presence of Mg^{2+} ions, NMR and biochemical data suggested that the aptamer domain is already pre-organized with the long-range loop-loop interaction involving L2 and L3 being mostly developed, but with the

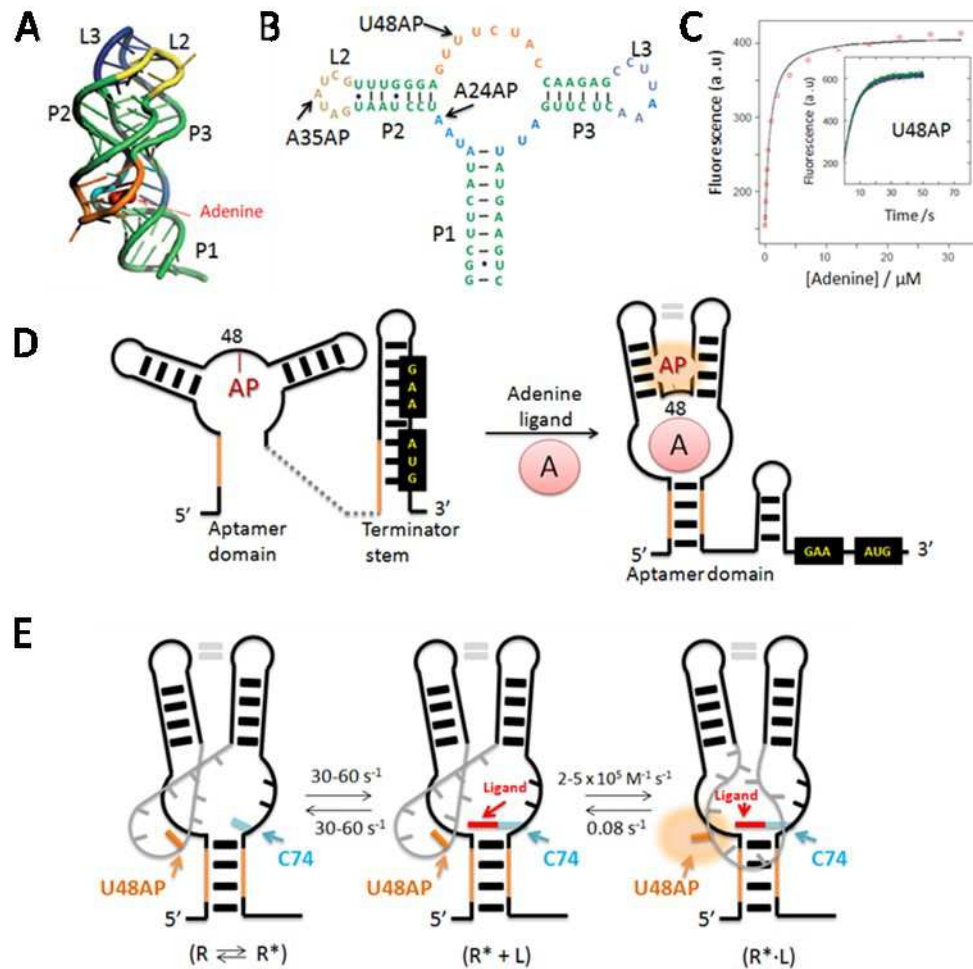


Fig. 2. Monitoring local conformational changes in purine-sensing riboswitches using 2AP substitutions. (A) Crystal structure of the add adenine riboswitch from *Vibrio vulnificus* in complex with adenine (Serganov, Yuan and al. 2004). (B) Cartoon representation of the secondary structure of the add adenine riboswitch indicating the location of the 2AP substitutions discussed in the text. (C) Example of thermodynamically determined dissociation constant from the variation in 2AP fluorescence versus adenine concentration. Inset: Representative example of stopped-flow fluorescence spectroscopy to monitor the kinetics of adenine binding to the add U48AP variant (71 nt). The concentration of aptamer and adenine ligand were $0.3 \mu\text{M}$ and $3 \mu\text{M}$, respectively (D) Schematics of 2AP-substitution experiments on full-length add adenine riboswitches. Ligand-binding to the aptamer domain triggers extrusion of the 2AP into the solvent which is monitored as an increase in fluorescence emission. (E) Proposed mechanism for ligand recognition by the guanine riboswitch. The riboswitch exists in an equilibrium of states with a disordered region (grey). Ligand binding selectively targets those conformers with an exposed C74 nucleotide (R^*), triggering the expulsion of U48AP into the solvent which is monitored as an increase in fluorescence emission and the disorder-to-order transition of the three-way junction.

ligand-binding pocket remaining largely disordered (Gelfand, Mironov and al. 1999, Serganov, Yuan and al. 2004, Borovok, Gorovitz and al. 2006, Lemay and Lafontaine 2007, Noeske, Schwalbe and al. 2007, Lemay, Penedo and al. 2009, Sharma, Bulusu and al. 2009, Delfosse, Bouchard and al. 2010, Leipply and Draper 2011, Tremblay, Lemay and al. 2011). It was also suggested that upon ligand binding the aptamer core undergoes a transition from this disordered "open" state to a locally organized conformation (Kubodera, Watanabe and al. 2003, Buck, Furtig and al. 2007, Stoddard, Gilbert and al. 2008, Prychyna, Dahabieh and al. 2009, Jain, Zhao and al. 2010). This "induced-fit" folding model was further investigated using 2AP substitutions in close vicinity (U48AP and A24AP) but also distant (A35AP) from the binding site (Fig. 2b) (Rieder, Lang and al. 2007). Upon addition of adenine to the U48AP variant a significant increase in fluorescence was observed indicating that the 2AP at this position becomes unstacked and protrudes into solution when the ligand binds (Fig. 2c), in agreement with the X-ray structure (Serganov, Yuan and al. 2004). Whilst the U48AP substitution showed very little variation in 2AP fluorescence upon addition of Mg^{2+} ions, two-thirds of the total quenching observed for the variant A24AP were caused by the addition of Mg^{2+} ions alone (Rieder, Lang and al. 2007), suggesting that this region of the binding pocket is mostly pre-organized in the presence of Mg^{2+} ions. Such decrease in 2AP fluorescence suggests that A24 becomes less flexible upon folding and agrees with the X-ray structure showing A24 intercalated between nucleotides G72 and A73.

The Mg^{2+} -dependence of the intercalation is further supported by the presence of a defined Mg^{2+} -binding site next to A24 that may stabilize the intercalated configuration (Serganov, Yuan and al. 2004). Additional evidence for an "induced-fit" model in which ligand binding modifies the aptamer architecture at regions distant from the binding site were obtained by replacing position A35, which is located at the loop-loop interface, by 2AP (Rieder, Lang and al. 2007). The A35AP variant showed a marked fluorescence quenching upon addition of Mg^{2+} and ligand. Moreover, the fact that more than one-third of the total observed quenching was caused by Mg^{2+} -induced folding provided evidence for the parallel orientation of stems P2 and P3 and a significant pre-organization of the loop-loop interaction in the absence of

ligand, as previously observed by other techniques (Kubodera, Watanabe and al. 2003, Buck, Furtig and al. 2007, Lemay and Lafontaine 2007, Stoddard, Gilbert and al. 2008, Prychyna, Dahabieh and al. 2009, Jain, Zhao and al. 2010, Tremblay, Lemay and al. 2011). A stopped-flow fluorescence approach revealed comparable folding rates ($k \sim 3.1 \pm 0.05 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$) for all the 2AP variants investigated (U48AP, A24AP and A35AP), suggesting that the organization of the binding pocket and the formation of the loop-loop interaction take place in a concerted manner (Rieder, Lang and al. 2007).

A similar U48AP substitution was employed in the context of the *xpt-pbuX* guanine riboswitch to investigate the ligand binding dynamics of the guanine analog 7-deazaguanine ($K_D \sim 0.0495 \mu\text{M}$) (Gelfand, Mironov and al. 1999). The biphasic behavior obtained for the association rates was proposed as evidence for an 'induced fit' binding mechanism involving at least two discrete steps as previously described for the theophylline aptamer (Fig. 2e) (Jucker, Phillips and al. 2003). The kinetic data were consistent with a model in which the RNA displays a manifold of ligand-free states slowly interconverting ($k_1 = 30\text{-}60 \text{ s}^{-1}$) between binding-incompetent and binding-competent configurations. In these "open conformations" the RNA structure has been mostly formed but the core remains highly dynamic and different from the ligand-bound state. Recognition of the ligand by these "open" structures may proceed by establishing a critical Watson-Crick pairing with nucleotide at position 74 (Fig. 2e) followed by the stretch of nucleotides connecting stems P2 and P3 (J2/3) acting as a lid to completely seal the ligand and form a network of hydrogen bonds that spread through the aptamer core.

Interestingly, significant differences in ligand-binding response were found when similar experiments using the U48AP variant were repeated in the context of the full length *add* adenine riboswitch (Fig. 2c) and compared to the equivalent G48AP-labelled *pbuE* riboswitch from *Bacillus subtilis* that controls regulation at the transcriptional level (Fig. 2d) (Rieder, Lang and al. 2007). A pronounced fluorescence increase in the U48AP-labelled *add* riboswitch in the presence of adenine ligand strongly suggested that the translational repressor

stem did not inhibit the formation of the RNA-ligand complex. In contrast, no change in fluorescence was detected for the G48AP-labelled *pbuE* riboswitch suggesting that ligand binding was significantly disrupted. The different response to ligand binding was rationalized by examining the interface between the aptamer and the expression platform in both riboswitches (Fig. 2d). For the *pbuE* riboswitch, the interface contains a more stable terminator stem that shares long stretches of nucleotides with the P3 and P1 stems of the aptamer domain. Once this terminator stem is formed, even high concentrations of adenine ligand cannot revert the structure towards the ligand-bound aptamer domain (Gelfand, Mironov and al. 1999, Rieder, Lang and al. 2007, Tremblay, Lemay and al. 2011, Dambach, Sandoval and al. 2015). On the contrary, the interface of the *add* riboswitch includes a much shorter stretch of shared nucleotides, thus still allowing partial assembly of the aptamer domain (P2, P3 stems and the L2/L3 loop-loop interaction) even when the repressor stem is formed (Fig. 2d). In the context of the gene regulation mechanism, the inability of the transcriptional-regulative *pbuE* riboswitch to bind the ligand after the terminator stem has been transcribed indicates that the *pbuE* must operate under kinetic control (Borovok, Gorovitz and al. 2006, Haller, Souliere and al. 2011, Lemay, Desnoyers and al. 2011). Transcriptional pausing at U-rich regions has been also discussed as an additional mechanism to provide the aptamer region with more time to sense the ligand (Borovok, Gorovitz and al. 2006, Haller, Souliere and al. 2011, Lemay, Desnoyers and al. 2011). In contrast, the thermodynamically-controlled *add* riboswitch exists in an equilibrium between the repressor stem and a ligand-binding competent structure and might be transcribed without pausing. These differences in ligand-binding response constitute an excellent example of how riboswitches within the same class have evolved to efficiently accommodate different regulatory requirements.

5. Dissecting global conformational rearrangements using 2AP substitutions: helical stacking and pseudoknot formation during riboswitch folding

A similar 2AP substitution strategy as that used to investigate the *add* and *pbuE* adenine riboswitches was applied to investigate the local conformational changes induced by metal ions and ligand binding in the aptamer domains of the SAM-I (Heppell and Lafontaine 2008, Heppell, Blouin and al. 2011), SAM-II (Haller, Rieder and al. 2011), TPP (Lang, Rieder and al. 2007), lysine (Blouin, Chinnappan and al. 2010, Garst, Porter and al. 2012) and preQ1 riboswitches (Rieder, Kreutz and al. 2010, Souliere, Haller and al. 2011). Among these, the folding pathway of SAM-I and SAM-II riboswitches has been extensively explored and provides an excellent example to discuss how 2AP substitutions can be tailored to monitor the folding of complex regulatory elements. Moreover, whereas the adenine riboswitch represents a paradigm of ligand-recognition models based on "induced-fit", SAM-I and SAM-II riboswitches, have become models of ligand-recognition following a "conformational capture" mechanism where the ligand binding shifts the equilibrium towards one of the multiple conformers present in solution (Heppell, Blouin and al. 2011). Despite recognizing the same ligand, both SAM-sensing riboswitches exhibit remarkably different structures. The SAM-I riboswitch is organized around a four-way junction composed of stems P1 to P4 connected through unpaired nucleotides (Fig. 3a) (Hale, Zhao and al. 2009, Ren, Rajashankar and al. 2012). The crystal structure of SAM-I showed the four helical stems (P1 to P4) being organized as two sets of coaxially stacked helices oriented with an angle of 70° relative to each other (Ren, Rajashankar and al. 2012). One of these coaxial stacks involves the P1 and P4 helical stems whilst the other one is formed between the P2 and P3 helices (Fig. 3a and Fig. 3b). In a recent study, the formation of the P2-P3 and P1-P4 helical stacks as a function of Mg²⁺ and SAM ligand was monitored by replacing nucleotides A71 and A138 at the interface of each stacking unit by 2AP, respectively (Fig. 3c) (Heppell, Blouin and al. 2011). For the A71AP variant, a relatively high 2AP emission in the presence of Na⁺ ions was progressively quenched by addition of Mg²⁺ ions, whereas no change was detected by addition of SAM ligand (Fig. 3c, upper panel). In marked contrast, the A138AP variant showed a significant

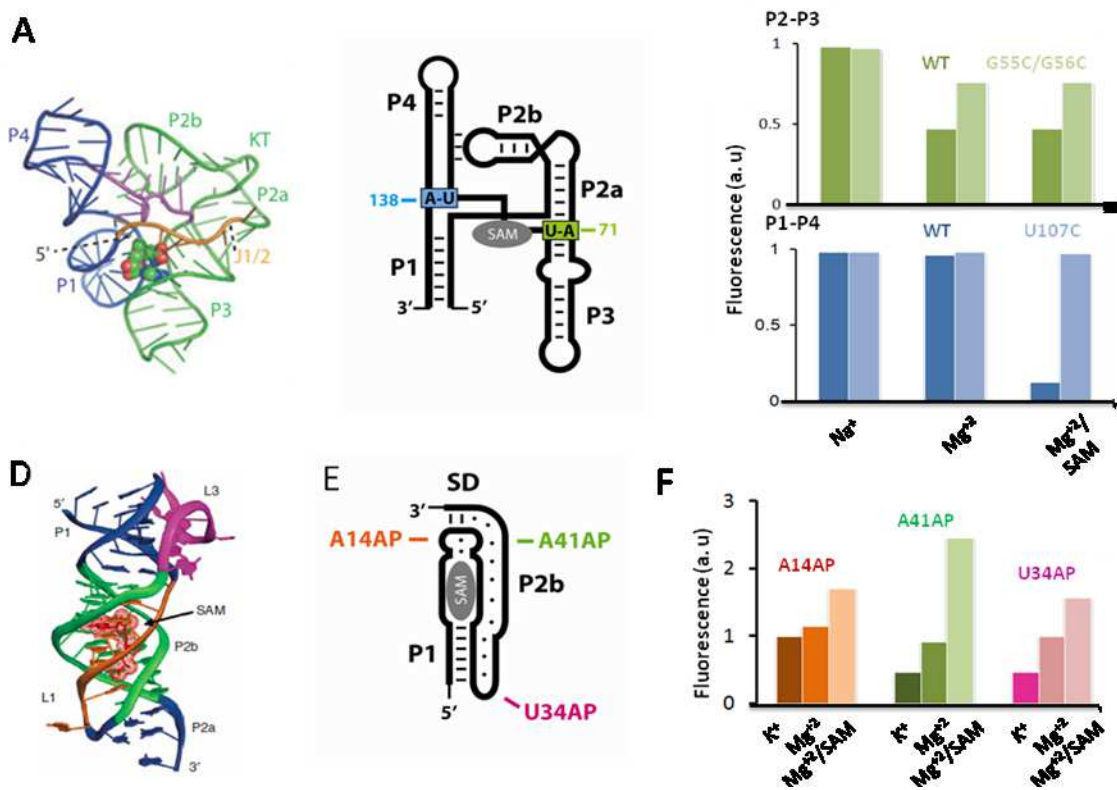


Fig. 3. Monitoring long-range conformational changes induced by helical stacking and pseudoknot formation in S-adenosylmethionine (SAM) riboswitches using 2AP substitutions. (A) Crystal structure of the SAM-I riboswitch in complex with the SAM ligand (Ren, Rajashankar and al. 2012). (B) Secondary structure and tertiary arrangement as deduced from the crystal structure (Ren, Rajashankar and al. 2012). The ligand position and the location of 2AP substitutions at positions 138 and 71 to monitor helical stacking are also indicated. (C) Upper panel: normalized 2AP fluorescence intensity monitoring the P2-P3 stacking. The 2AP fluorescence emission at position 71 was monitored in the context of the wild type and of a G55C/G56C pseudoknot-deficient mutant at the indicated experimental conditions. Bottom panel: Normalized 2AP fluorescence intensity monitoring of the P1-P4 stacking. The 2AP fluorescence emission at position 138 was monitored in the context of the wild type and a U107C ligand binding-deficient mutant at the indicated experimental conditions. [Adapted from reference 39]. (D) Crystal structure of the SAM-II riboswitch in complex with SAM ligand (Gilbert, Rambo and al. 2008). (E) Secondary structure and tertiary arrangement as deduced from the crystal structure (Gilbert, Rambo and al. 2008). The ligand position and the relative location of 2AP substitutions discussed in the text are shown. (F) Relative variation in 2AP fluorescence intensity obtained for the indicated 2AP substitutions normalized with respect to the fluorescence value obtained for the A14AP value in the absence of magnesium and ligand [Adapted from reference 43]

decrease in 2AP fluorescence only upon addition of SAM ligand (Fig. 3c, bottom panel), suggesting that while Mg^{2+} ions are mostly responsible for stacking of the P2-P3 helical stems, the equivalent coaxial alignment of the P1 and P4 stems is driven almost exclusively by ligand-binding (Heppell, Blouin and al. 2011). The importance of ligand-

binding in P1-P4 stacking was further confirmed by the lack of 2AP quenching using a U107C mutant of the A138AP construct (Fig. 3c, bottom panel) that does not support ligand binding. The influence of the pseudoknot motif on the folding of the aptamer core was also examined using a similar 2AP fluorescence assay and mutations (G55C and G56C) that abolish pseudoknot formation (P2 mutant) (Fig. 3c, upper panel) (Heppell and Lafontaine 2008, Heppell, Blouin and al. 2011). It was found that disruption of the pseudoknot tertiary interaction markedly decreased the fluorescence quenching observed for the A71AP construct carrying the P2 mutant in the presence of Mg^{2+} and Mg^{2+} /SAM when compared to the A71AP wild-type structure (Fig. 3c, upper panel). Interestingly, the fluorescence quenching was efficiently restored by using a compensatory mutant, suggesting that pseudoknot formation can occur as long as the base-pairing potential is maintained (Heppell and Lafontaine 2008, Heppell, Blouin and al. 2011). Similar Mg^{2+} -induced organization of the SAM-I aptamer was recently observed by Sanbonmatsu and coworkers [94]. In their study, they found that while Mg^{2+} could produce a partially closed state, ligand was still required to fully form the closed state. Moreover, they also showed that ligand binding could pre-organize the closed state, suggesting that Mg^{2+} and ligand cofactors can both be involved in the pre-organization of the aptamer. In addition, they also found that the formation of the closed state relies on key tertiary contacts that modulate the structure of the expression platform [94, 95].

The structure of the SAM-II riboswitch is also organized as a classical H-type pseudoknot comprising two Watson-Crick-paired helices (P1 and P2a) and two loops (L1 and L3) (Fig. 3d and Fig. 3e) (Gilbert, Rambo and al. 2008). Notably, formation of the pseudoknot induced by Mg^{2+} ions and stabilization of the ligand-binding competent state sequesters the Shine-Dalgarno sequence by forming the P2a stem via two Watson-Crick base pairs with nucleotides of loop L1 (Gilbert, Rambo and al. 2008, Haller, Rieder and al. 2011) (Fig. 3e). Pseudoknot formation was investigated using three fluorescent variants incorporating single 2AP

substitutions at specific locations. The first 2AP substitution was located 'above' the ligand binding site (U34AP) to report on P1/L3 folding (Fig. 3e). A second variant was engineered close to where the SAM-ligand nucleobase intercalates (A41AP), and a third substitution was placed 'below' the binding site to report the formation of the P2a stem and the sequestration of the Shine-Dalgarno sequence (A14AP) (Fig 3e). In agreement with the X-ray structure (Gilbert, Rambo and al. 2008), the three variants showed an increase in 2AP fluorescence upon addition of Mg^{2+} ions and a further increase after addition of SAM ligand (Fig. 3f), suggesting that they adopt a common unstacked configuration in the ligand-bound state (Haller, Rieder and al. 2011). Interestingly, the rate constant for ligand binding monitored using the A14AP variant was slightly faster (~ 1.8) than those observed for the U34AP and A41AP substitutions, suggesting that conformational changes around nucleotides U34 and A14 are rate-limiting. Taking these data together, a sequential folding model was proposed in which pre-organization of the L1 loop initiates pseudoknot formation and allows SAM binding followed by compaction of loop L3 and formation of the P2a helix [46].

Available data about SAM-I and SAM-II riboswitches are consistent with a model in which only a subset of structural conformations are competent for ligand binding [42, 46]. Because both riboswitches may sample all conformations along the folding pathway and that ligand binding only stabilizes a subset of them, the SAM-I and SAM-II riboswitches employ a “conformational capture” mechanism [42, 46]. This process is distinct from that of induced-fit where ligand-bound structures are only accessed in presence of ligand, which probably involves more local RNA rearrangement. Sm-FRET assays have shown that the presence of ligand shifts the equilibrium towards the closed state in the case of SAM-I and SAM-II riboswitches [42, 46]. This is in contrast to the purine riboswitches for which folding of the core domain is only achieved in presence of ligand [30, 31, 45]. In principle, it is entirely possible for a riboswitch to employ both folding mechanisms to achieve the ligand-bound native state. For example, one could envisage that the SAM-I riboswitch also employs an induced-fit mechanism to reorganize the ligand binding site upon SAM binding. Therefore,

riboswitches could employ various folding pathways to recognize cellular metabolites and to regulate gene expression.

2AP substitutions have also been used to investigate the gene regulation mechanism of the lysine riboswitch (Blouin, Chinnappan and al. 2010, Garst, Porter and al. 2012) and the molecular basis of ligand specificity (Garst, Porter and al. 2012). A 2AP substitution at position G193 was initially used by Lafontaine and coworkers (Heppell, Blouin and al. 2011) to monitor the folding and ligand-binding process. As expected from chemical footprint and X-ray data, a strong increase in 2AP fluorescence was observed upon the addition of lysine, suggesting a transition from a stacked conformation in the unbound aptamer to a solvent exposed conformation (Blouin, Chinnappan and al. 2010). A two-piece five-way lysine aptamer carrying the G193AP substitution was subsequently engineered by Batey and coworkers to determine whether the *B. subtilis lysC* aptamer could productively bind different lysine analogs and other amino acids (Garst, Porter and al. 2012). The aptamer exhibited >1000 fold specificity for lysine ($K_D \sim 53 \mu\text{M}$) over other amino acids and only serine showed some affinity, although at much higher concentrations ($\sim 20 \text{ mM}$) than those present in the cell ($\sim 70 \mu\text{M}$) (Garst, Porter and al. 2012). Using the G193AP variant, it was found that the aptamer specificity is highly sensitive to the precise positioning of the Σ -amino group to form electrostatic contacts with the backbone of G110, whereas alterations to the main-chain functional groups are better tolerated. Notably, lysine analogs lacking a formal charge on the side Σ -amino chain displayed a very moderate reduction in affinity (4-5 fold). This indicates that the positive charge of the amino group is not essential for ligand binding if the network of hydrogen bonds is maintained.

The binding kinetics of the *lysC* riboswitch was also analyzed by stopped-flow fluorescence spectroscopy using the G139AP substitution. The extracted association and dissociation rates were used to model theoretical values of half-maximal regulatory activity (T_{50}) as a function

of transcription speed and compared these with T_{50} values obtained using *E. coli* RNAP acting under single round *in vitro* transcription conditions (Garst, Porter and al. 2012). As predicted by the model and consistent with T_{50} values obtained using *B. subtilis* RNAP (Wilson-Mitchell, Grundy and al. 2012), at low NTP concentrations (50 μM), a good agreement was observed between the aptamer affinity ($K_D \sim 53 \pm 5 \mu\text{M}$) and the half-maximal regulatory activity ($T_{50} \sim 61 \pm 9 \mu\text{M}$), suggesting that the lysine riboswitch operates under thermodynamic control. At high NTP concentrations, similar to those found in rapidly dividing cells ($>400 \mu\text{M}$), and in excellent agreement with the values modeled using the 2AP-substitution assay ($T_{50} \sim 260 \mu\text{M}$), the experimental T_{50} values increased by a factor of ~ 5 -fold ($290 \pm 9 \mu\text{M}$), a hallmark of kinetic control of gene regulation. Together, these data indicate a drastic shift from a thermodynamic to a kinetic control of the regulation process as a function of transcription rate. Because the concentration of NTPs is known to change by a factor of 2- to 5-fold during different bacterial growth phases (Buckstein, He and al. 2008), the regulatory efficiency of the *lysC* riboswitch could be finely tuned by the metabolic state of the cell.

6. Understanding ligand-recognition mechanisms using intrinsically fluorescent ligands

In the previous sections we described the use of specific 2AP substitutions along the mRNA sequence to investigate the folding, ligand-binding dynamics and molecular basis of ligand specificity in a variety of riboswitches. However, in addition to using 2AP as an indirect reporter to characterize the RNA state, its ability to efficiently replace the natural ligand in adenine-sensing riboswitches has also been explored to directly monitor the dynamics of ligand binding (Borovok, Gorovitz and al. 2006, Eskandari, Prychyna and al. 2007, Lemay and Lafontaine 2007, Jain, Zhao and al. 2010, Tremblay, Lemay and al. 2011, Dambach, Sandoval and al. 2015).

The 2AP binding affinity for a set of wild-type adenine-sensing aptamers was tested using a fluorescence assay. As exemplified in Fig. 4a, as the concentration of RNA aptamer increases, the 2AP fluorescence becomes progressively quenched due to the formation of the bound complex until reaching a plateau at saturating ligand concentrations (Lemay and Lafontaine 2007, Dambach, Sandoval and al. 2015). The fluorescence data fitted to a two-state model yielded similar dissociation constants with values ranging from 116 ± 14 for the *Vibrio vulnificus* (VV) *add* riboswitch to 552 ± 17 for *Clostridium perfringens* (CP) *add* riboswitch, thus confirming that 2AP can efficiently replace the natural ligand in all variants investigated (Lemay and Lafontaine 2007). Notably, VV aptamers carrying a P1 stem containing only 2 bp showed a 54-fold reduction in affinity, suggesting that the length of the P1 stem is critical to stabilize the RNA-ligand complex (Lemay and Lafontaine 2007). 2AP binding to the adenine aptamer was also used to experimentally test the consensus structure proposed based on in-line probing and X-ray data. The different aptamer variants tested and their corresponding affinity values for 2AP are shown in Fig. 4b and Fig. 4c, respectively. Based on these data, a new consensus structure was proposed to account for variations in the helical regions and also in the core of the aptamer.

Among all variants, it is worth to discuss in detail the effect of position 39 in ligand binding and its functional correlation with position 65 (positions 48 and 74 using guanine riboswitch nomenclature, Fig. 2), which dictates ligand-binding specificity (Lemay and Lafontaine 2007). In the context of the adenine riboswitch (A-box), the consensus sequence suggests that position 39 can be any base except an adenine (Fig. 4c). Similarly, the consensus sequence for the guanine riboswitch (G-box) indicates that position 39 can be any base except of guanine.

Using 2AP as ligand, significant reductions in binding affinity were observed for the A-box variant G39A ($K_D \sim 1977 \pm 139$ nM) and the G-box variant U39G ($K_D \sim 1977 \pm 139$ nM), whereas other mutations induced only a very moderate reduction in affinity (Fig. 4c). A crystal structure of the *B. subtilis pbuE* aptamer carrying the G39-C65 provided the rationale for the

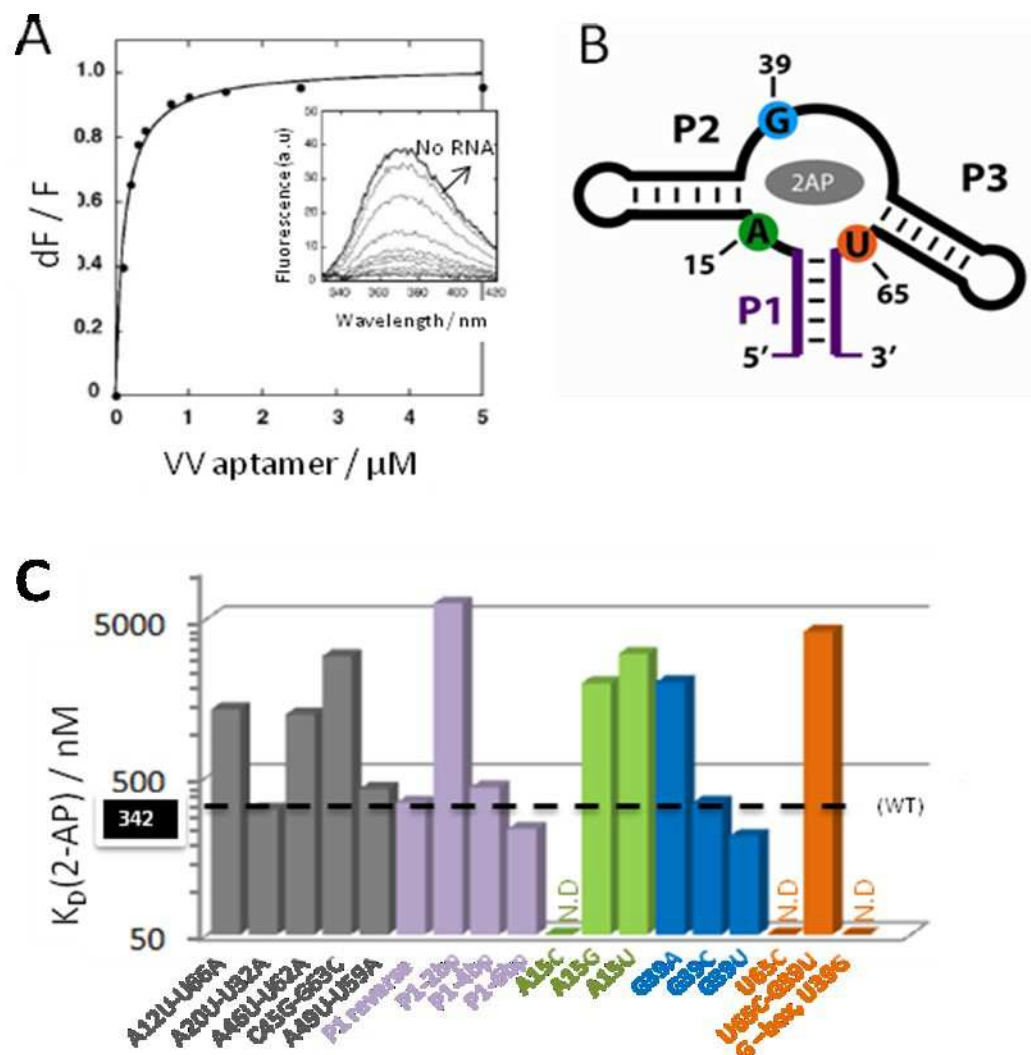


Fig. 4. 2AP as an intrinsically fluorescent ligand mimic of adenine to investigate ligand recognition processes in purine-sensing riboswitches. (A) Normalized 2AP fluorescence intensity plotted as a function of wild type adenine aptamer. Changes in fluorescence (dF) were normalized to the maximum fluorescence measured in the absence of RNA (F). Inset: fluorescence emission spectra from 330 to 420 nm for each RNA concentration. [Adapted from reference 45]. (B) Schematic showing the position of 2AP substitutions discussed in the text. The reduction in 2AP binding affinity observed for positions 15, 39, and 65 is also indicated. 2AP binding affinities of the A15C and U65C variants could not be determined due to their very low affinity [Adapted from reference 45]. (C) Bar plot summarizing the variation in affinity values for 2AP binding to the adenine aptamer for the indicated aptamer variants [Adapted from reference 45]. The dashed line indicates the affinity value obtained for the wild type adenine aptamer.

sequence conservation observed at position 39 (Delfosse, Bouchard and al. 2010). The X-ray structure revealed a compact ligand-free conformation in which the presence of the base pair induces a network of interactions that mimics those present in the wild type aptamer (Delfosse, Bouchard and al. 2010). In a later study, it was found that stable ligand-free combinations such as A39-U65 or G39-C65 also showed a perturbed gene regulation efficiency (Delfosse, Bouchard and al. 2010, Tremblay, Lemay and al. 2011). These findings are consistent with the idea that purine riboswitches have evolved to exclude non-productive ligand-free conformations (Tremblay, Lemay and al. 2011).

The variation in fluorescence quantum yield between free and bound 2AP was further exploited to assess whether the *pbuE* riboswitch operates under thermodynamic or kinetic control (Borovok, Gorovitz and al. 2006). The association rate constant for 2AP binding using a stopped-flow approach and pseudo-first order conditions of RNA in excess of ligand were similar to those obtained using the U48AP variant previously described ($\sim 10^4 \text{ M}^{-1} \text{ s}^{-1}$) (Rieder, Lang and al. 2007). Competition experiments between adenine (or 2,6-diaminopurine (DAP)) and a pre-incubated RNA:2AP complex were performed using a dilution-relaxation strategy in which the complex was diluted into fluorescently inert ligand (adenine, DAP). The time-dependent recovery of 2AP fluorescence after dilution was used to extract the 2AP dissociation rate (k_{off}) and the lifetime of the RNA:2AP complex ($1/k_{\text{off}}$) yielding values ranging from 94 s (20°C) to 18 s (30°C). The experimental values for the association and dissociation constants obtained from these experiments for the three ligands investigated (adenine, 2AP and DAP) were used in combination with kinetic simulations to describe the response of the system to changes in the binding-competent time window (\otimes_{RNAP}), which cause the regulatory mechanism to switch between kinetic and equilibrium regimes. For the *pbuE* riboswitch, the formation of a binding competent aptamer requires the incorporation of at least 80 nucleotides. The binding-competent window (\otimes_{RNAP}) spans from this stage to the point at which the terminator will form and involves the incorporation of approximately 30 additional nucleotides (80-110) (Borovok, Gorovitz and al. 2006). As an example, if we

assume that the RNA polymerase (RNAP) incorporates 15 nt/s, the value of \otimes_{RNAP} will be approximately 2 s. Taking the association rate between the *pbuE* riboswitch and the adenine ligand obtained at 25°C using the 2AP competition assay ($2.6 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$), a 2 μM concentration of free adenine ligand is required to significantly bind the aptamer within this 2 s temporal window (Borovok, Gorovitz and al. 2006). While more research is required to completely understand the details of the gene regulation mechanism, the potential of the *pbuE*, and the lysine riboswitch as previously discussed, to function as a kinetically or thermodynamically controlled switch depending on transcription parameter (i.e., transcription rate, NTPs concentration) emphasizes the ability of metabolite-sensing RNAs to quickly adapt to changes in the cellular environment and perform sophisticated tasks long before proteins had emerged during evolution.

The flavin mononucleotide-sensing (FMN) riboswitch is the only known riboswitch displaying a comparable potential to the adenine riboswitch-2AP system to directly monitor ligand binding using a fluorescence assay due to the intrinsic emission of the FMN ligand. In fact, similarly to 2AP, the fluorescence of the free FMN ligand near 520 nm becomes also quenched upon forming a complex with the aptamer. As observed for other metabolite-sensing RNAs (Henkin 2008, Horvath and Barrangou 2010), the full-length FMN riboswitch from the *ribD* operon of *B. subtilis* (263 nt) showed a substantial difference between the K_D and the T_{50} values, and such disparity has been proposed as a signature of kinetically-controlled riboswitches (Horvath and Barrangou 2010). In fact, two putative transcriptional pausing sites, PA and PB, have been identified near positions U215 (PA) and U245 (PB) of the *ribD* sequence that could play an important role in the function of the FMN riboswitch (Horvath and Barrangou 2010). The lifetime of these pause sites was determined to be ~ 10 s and ~ 1 min, respectively. The affinity of the FMN ligand for transcriptional intermediates mimicking the PA (200 *ribD*) and PB (230 *ribD*) pausing sites was assessed by monitoring the decrease in fluorescence of the FMN ligand upon binding (Horvath and Barrangou 2010). The value obtained ($K_D \sim 11$ nM) for the aptamer domain (165 *ribD*) was in good agreement with that reported using in-line probing (Makarova, Grishin and al. 2006, Horvath and Barrangou

2010). The dissociation constant values obtained for the 200 *ribD* (PA mimic) were only moderately increased by 2-fold compared to the aptamer domain, whereas the value for the 230 *ribD* (PB mimic) displayed an 8-fold increase (~ 89 nM at 20°C), which is still an order of magnitude better than the T_{50} value observed using *B. subtilis* RNAP. Additional dilution-relaxation experiments using stopped-flow fluorescence techniques were employed to gain insight into the lifetime of the RNA-FMN complex. In these experiments, the pre-equilibrated RNA-FMN complex was diluted in buffer to induce the dissociation of the complex and the subsequent recovery of the FMN emission. From the very poor k_{off} values obtained ($\sim 10^{-3}$ s $^{-1}$ for 165 *ribD*), it was concluded that once the riboswitch is bound to FMN, it is unlikely to release the ligand and achieve equilibrium before the RNAP reaches the site of intrinsic termination and the riboswitch completes its regulatory function.

7. Dynamics and heterogeneity of ligand-bound states using ultrafast time-resolved 2AP fluorescence

High resolution crystal structures of ligand-bound states exist for many metabolite-sensing mRNAs providing an atomic-level description of the architecture of the RNA-ligand complex (Serganov, Yuan and al. 2004, Gilbert, Rambo and al. 2008, Delfosse, Bouchard and al. 2010, Ren, Rajashankar and al. 2012). However, these structures represent only an average static image of the complete set of RNA-ligand conformers that may dynamically coexist in equilibrium in solution. Although ligand-bound states are certainly less flexible than ligand-free states, the presence of the ligand might not be sufficient to completely lock down the complex into a rigid structure, and in fact, a slightly flexible bound state might be more efficient, particularly for riboswitches operating under kinetic control. Therefore, exploring the dynamic range of these interactions between the ligand and the bases that constitute the binding pocket is important and may help to establish a more detailed picture of riboswitch function. Similarly to the variation in quantum yield observed under steady-state illumination conditions, the fluorescence lifetime of the 2AP excited-state is known to be strongly

influenced by the local environment (Ballin, Bharill and al. 2007). The fluorescence decay profile of 2AP free in solution following excitation with an ultra-short laser pulse (usually picoseconds or shorter) exhibits a mono-exponential behavior with a characteristic lifetime of ~11-12 ns (Ballin, Bharill and al. 2007, Neely, Magennis and al. 2007). The presence of shorter lifetime components in the fluorescence decay is normally associated to the existence of 2AP fluorophores engaged in stacking interactions with neighboring bases (Neely, Daujotyte and al. 2005, Ballin, Bharill and al. 2007, Neely, Magennis and al. 2007). Thus, multiple 2AP-RNA environments that interconvert in a time scale that is slow compared to the lifetime of the excited-state (~11-12 ns) will be resolved as complex multi-exponential fluorescence decays. In such ensemble measurements, the fluorescence lifetimes values (τ_i) and their corresponding fractional contributions (f_i) provide insights into the number of bound conformers present in solution and their relative populations, respectively (Neely, Daujotyte and al. 2005, Neely, Magennis and al. 2007). Remarkably, the ability to extract the relative population of each 2AP-RNA species present in solution enables the quantification of the binding affinity to specific conformers of the RNA-2AP complex. This is a feature not accessible using steady-state 2AP assays, where only an average value from all the RNA-2AP conformers present in solution can be obtained.

Recent studies on the adenine riboswitch in complex with 2AP exemplified how the conformational heterogeneity of ligand-bound aptamers can be captured using these time-resolved fluorescence techniques (Eskandari, Prychyna and al. 2007, Prychyna, Dahabieh and al. 2009). The fluorescence decay profiles of 2AP in the presence of the adenine riboswitch were well described using a four-exponential model. The longest lifetime matched that of unbound 2AP, whereas the contribution to the fluorescence decay with a lifetime of < 0.3 ns (in the limit of experimental resolution) was assigned to a closed conformer (C_3) resembling the X-ray structure of the bound state in which 2AP is tightly bound and buried within the RNA structure. The remaining two contributions to the experimental fluorescence decay with lifetimes of ~1.9 ns (conformer C_1) and 0.5 ns (conformer C_2) were interpreted as arising from

less compact intermediate conformations of the RNA-2AP complex (Eskandari, Prychyna and al. 2007). This assignment was further supported by looking at the variation in the relative contribution of each lifetime component to the total fluorescence decay as a function of ligand concentration, Mg^{2+} concentration and temperature. The weaker ligand-binding ($K_D(2AP) \sim 635 \pm 25$ nM), Mg^{2+} -binding ($K_D(Mg^{2+}) \sim 1.5-2$ mM) affinities and thermal stabilities ($T_m \sim 41-46^\circ C$) observed for the C_1 and C_2 open conformers compared to that obtained for the C_3 closed state ($K_D(2AP) \sim 60 \pm 50$ nM, $K_D(Mg^{2+}) \sim 90 \pm 50$ nM, $T_m \sim 35^\circ C$) confirmed that the C_1 and C_2 species are structurally different to the closed C_3 state and represent bound states in which 2AP remains poorly stacked (Eskandari, Prychyna and al. 2007, Prychyna, Dahabieh and al. 2009). Further mutational analysis compromising the stability of the P1 stem (G13C) and the formation of the loop-loop interaction (G37C) were consistent with a model in which ligand binding directs riboswitch folding both at the binding pocket and at the remote loop-loop interaction (Prychyna, Dahabieh and al. 2009).

In the previous studies the lifetime corresponding to the closed RNA-2AP complex (C_3 conformer) was near the time resolution of the experimental setup (Eskandari, Prychyna and al. 2007, Prychyna, Dahabieh and al. 2009). A more recent 2AP-based strategy using femto-second time-resolved fluorescence up-conversion techniques was applied to expand the dynamic range to time scales down to picoseconds (Jain, Zhao and al. 2010). The combination of this technique with time-resolved anisotropy measurements confirmed dynamic motions of the ligand inside the binding pocket of the *add*, *pbuE* and *xpt-pbuX* riboswitches occurring on fast (picosecond) and slow (1-2 ns) time scales. Such motions of the ligand are most likely coupled to base fluctuations within the binding pocket and possibly even to the entire aptamer domain. In agreement with molecular dynamics simulations on the adenine (Priyakumar and MacKerell 2010) and guanine (Villa, Wohner and al. 2009) riboswitches, it was hypothesized that these motions may represent fluctuations of 2AP between two populations in which the ligand stacks with either the adenine or the uracil bases of the A21-U75 and A52-U22 base pairs that form the floor and ceiling of the binding pocket. Although, this sliding motion of the 2AP ligand needs to be confirmed, there is no doubt that ultrafast time-resolved fluorescence

technique have the potential to resolve complex population distributions within an RNA ensemble and can complement single-molecule and NMR studies to increase our understanding of RNA-ligand interactions.

8. Single-molecule dynamics of long-range structural rearrangements within the aptamer sensing domain

The use of fluorescent nucleotide analogs provides a wealth of information on RNA folding mechanism in riboswitches. However, X-ray crystal structures of riboswitches show many long-range interactions that are beyond the local probing range of 2AP and pyrrolo-dC analogs. To understand in more detail the formation of these long-range interactions, Förster Resonance Energy Transfer (FRET) can be used to monitor global conformational changes (Selvin 2000). FRET requires two compatible fluorescent dyes, a donor and an acceptor, that can sustain resonance energy transfer upon the specific excitation of the donor. The transfer of energy is highly dependent on the distance between the two dyes and thus variations in FRET efficiency are highly sensitive indicators of RNA global structural changes (Selvin 2000).

Commonly used FRET pairs consist in fluorescein and Cyanine 3 (Cy3) and Cyanine 3-Cyanine 5 (Cy5). Dyes can be incorporated at a specific location of the RNA using the chemical coupling between an N-hydroxysuccinimide ester (NHS) group of the dye and the amino linker group on the RNA. The amino linker group can be added during solid-phase RNA synthesis and it is possible to add this modification at the 5' or 3' end as well as internally. However, internal labelling is only available on the C5 position of a deoxy-thymidine, meaning that the internal labels should always be positioned at a uridine in the natural RNA to preserve the Watson-Crick base pairing ability of the nucleotide. It is important to make sure that the dyes do not disrupt the structure and function of the studied RNA. To do so, crystal structures are particularly useful given that nucleotides fully exposed in a crystal are excellent candidates for fluorescent labelling. In riboswitch studies, to ensure

that the labelling does not disrupt the structure it is important to do a ligand binding experiment and compare the result with an unlabelled RNA.

As it will be discussed below, FRET is a very powerful technique to understand the global folding and ligand-recognition mechanisms of riboswitches. Furthermore, by performing FRET measurements at the single-molecule level (sm-FRET), it is possible to circumvent the problems associated to the averaging of many molecules characteristic of conventional ensemble experiments. Thus, single-molecule FRET allows to extract complex folding trajectories and clearly identify transient states along the folding pathway that are inaccessible by ensemble approaches. Sm-FRET relies on the spatial isolation of fluorescently-labelled RNA molecules on a surface to allow the analysis of individual folding trajectories and the associated conformational changes in solution. A common way to isolate molecules for sm-FRET experiments is to exploit the interaction between a biotin-modified RNA and surface-coated streptavidin as a way to attach molecules on a quartz microscope slide and observe them for long periods of time. Readers interested to learn more about sm-FRET techniques are encouraged to consult an excellent review recently written by Taekjip Ha (Roy, Hohng and al. 2008). Once molecules are isolated on a surface, it is possible to independently monitor the FRET process of every molecule, enabling the detection of rare folding events that are usually hidden in bulk studies. Thus, sm-FRET bypasses the need to synchronize molecular folding processes because the folding pathways of monitored molecules can be observed independently allowing for easy observation of unsynchronized and fast folding events. A combination of FRET and sm-FRET experiments were used to study with success the folding pathway of a variety of riboswitches, such as purine (Brenner, Scanlan and al. 2010, Dalgarno, Bordello and al. 2013, Dambach, Sandoval and al. 2015), SAM (Haller, Rieder and al. 2011, Heppell, Blouin and al. 2011), Lysine (Fiegland, Garst and al. 2012), TPP (Haller, Altman and al. 2013) and c-di-GMP riboswitches (Wood, Ferre-D'Amare and al. 2012).

The purine binding aptamers are well-conserved RNA structures that are organized around a three-way junction connecting three helical domains (P1, P2 and P3). The crystal structure of

the *xpt* guanine aptamer shows a tertiary interaction occurring between the loops of P2 and P3 stems, which was originally proposed based on sequence alignment (Ren, Rajashankar and al. 2012). FRET experiments were carried out to investigate the molecular requirements of this long-range interaction (Dambach, Sandoval and al. 2015). Dyes were incorporated at positions U27 (P2 loop) and U53 (P3 loop) of the *Vibrio vulnificus add* aptamer because of their high accessibility as shown in the crystal (Serganov, Yuan and al. 2004). FRET was measured at different MgCl_2 concentrations and showed two main states in which the aptamer was observed, an unfolded (U) and folded (F) state (Fig. 5a). By performing a Mg^{2+} titration, it is possible to calculate the concentration of magnesium ions ($[\text{Mg}^{2+}]_{1/2}$) at which the transition from the unfolded to the folded state is 50% complete. When the titration was performed in absence of adenine, the calculated $[\text{Mg}^{2+}]_{1/2}$ corresponded to 22 μM . However, in presence of adenine, the obtained value was 12 μM suggesting that the affinity for magnesium is increased in presence of adenine. This can be explained by a model in which the loop-loop interaction is assisted by the presence of adenine.

To further understand the role of ligand in the formation of the loop-loop interaction, sm-FRET experiments were carried out in as a function of Mg^{2+} and ligand. In addition to the U and F states observed in bulk experiments, sm-FRET demonstrated that *add* aptamer can adopt a transient intermediate conformation (I). Analysis of 846 single traces at low magnesium concentration (500 μM) suggested that the intermediate is an obligatory step in the folding of the *add* aptamer as only 6% of the transitions monitored showed direct transition from the unfolded to the folded state, and vice versa (Dambach, Sandoval and al. 2015). These results allowed to propose the folding pathway for the *add* adenine aptamer as shown in Fig. 5a. The intermediate was detectable with and without adenine suggesting that the existence of this conformation is not dependent on the binding of the ligand. However, experiments made at low magnesium concentration (2 nM), with and without 50 μM adenine, showed an increase in the amount of intermediate in presence of adenine. This indicate that the ligand promotes the formation of the intermediate state (Dambach, Sandoval and al. 2015). Sm-FRET also allowed measuring the dwell time in folded and unfolded conformations. Fitting the dwell

times using single-exponential functions allowed to calculate the rate constants for folding ($k_{\text{fold}} = 0.76 \pm 0.05 \text{ s}^{-1}$) and unfolding ($k_{\text{unfold}} = 1.3 \pm 0.1 \text{ s}^{-1}$) in the absence of adenine as well as in presence of 50 μM adenine ($k_{\text{fold}} = 1.5 \pm 0.1 \text{ s}^{-1}$) and ($k_{\text{unfold}} = 0.84 \pm 0.05 \text{ s}^{-1}$). These results showed that the stabilization of the folded state by the ligand results from changes to both folding and unfolding rates (Dambach, Sandoval and al. 2015).

A recent sm-FRET study addressed the folding pathway of another purine binding riboswitch, the *B. subtilis xpt* guanine riboswitch (Brenner, Scanlan and al. 2010). While the *add* aptamer binds adenine, the *xpt* aptamer preferentially binds guanine, xanthine and hypoxanthine. This difference is explained by the presence of a cytosine at position 74 of the *xpt* aptamer which does a Watson-Crick interaction with guanine, while a uracil is present at this position in the *add* aptamer. Otherwise, both aptamer are very similar and are composed of three RNA stems in which the P2 and P3 stem-loops are known to interact with each other, as seen in crystals (Serganov, Yuan and al. 2004, Ren, Rajashankar and al. 2012). Sm-FRET experiments conducted on the *xpt* aptamer to monitor folding transitions occurring between P1 and P2 stems revealed that two states exist without magnesium (E_{FRET} values of ~ 0.3 and ~ 0.65). Upon addition of MgCl_2 , both states collapsed into a single population exhibiting a value of ~ 0.4 , which increased to 0.5 when magnesium concentrations were increased over 10 mM. When similar experiments were done with 500 nM guanine, the authors found that 3 states exist without MgCl_2 with E_{FRET} values of ~ 0.3 , ~ 0.45 and ~ 0.65 . Examination of individual traces showed that all three states could be visited by a single molecule. Increasing concentrations of magnesium lead to a collapse of all those states to a population having a E_{FRET} value of ~ 0.5 , which is slightly higher than what was observed without guanine. This observation suggests that guanine lead to a higher level of compaction of the aptamer in presence of magnesium ions. This effect is dependent of guanine as the addition of 50 μM adenine did not produce similar compaction. In contrast, the folding of the P1-P3 vector showed a relatively simpler folding transition where only one population exhibiting a E_{FRET} of

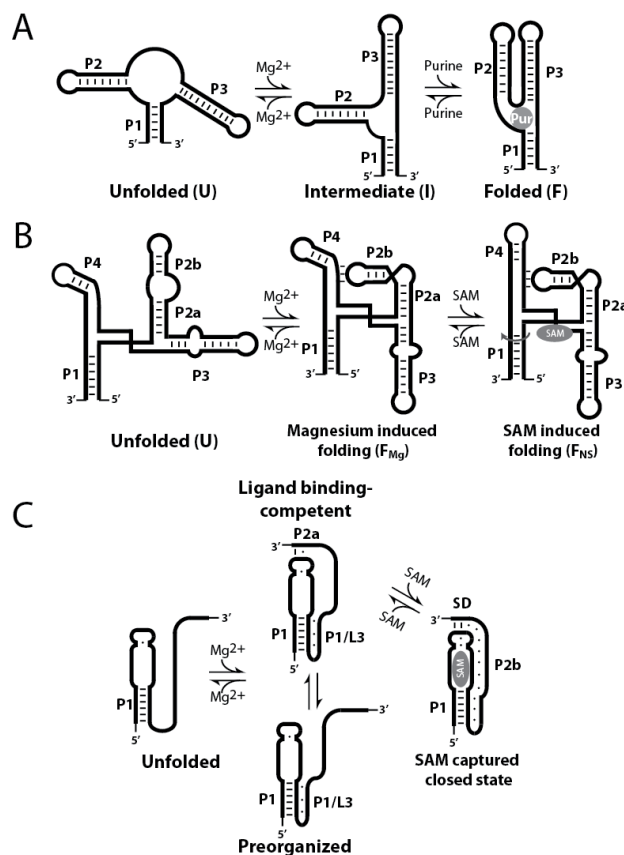


Fig. 5. Folding pathway proposed for purine (A), SAM-I (B) and SAM-II (C) aptamer. For the purine aptamer (A), the formation of the folded state determined to be by the interaction between the loops of stems P2 and P3 is well characterized. It can be formed in presence of magnesium only and is stabilized by the presence of the ligand. Sm-FRET experiments demonstrated that the *add* aptamer goes from the unfolded state to the folded state by passing through an intermediate state. The intermediate is hypothesized to be a stacking between the stems P1 and P3. No intermediate was detected for the *xpt* aptamer suggesting that the *add* and *xpt* aptamers do not follow the same folding pathway. The folding pathway of the SAM-I aptamer as determined by 2AP, CGE and sm-FRET experiment is shown in B. Magnesium ions promote the formation of the pseudoknot between the loop 2 and J3/4 as well as the stacking of stems P2 and P3 to give the F_{Mg} state. The binding of the ligand SAM induces the stacking of stems P4 and P1 and an helical twist of the P1 stem, indicated by the arrow, to produce the FNS state. The folding pathway of the SAM-II riboswitch is shown in C. Here, magnesium ions allow the formation of the pre-organized state and the binding-competent state. The pre-organized state is characterized by the formation of a three strands RNA interaction giving the P1/L3 structure. The binding-competent state contains the same structure in addition to a two base pairs interaction partially hiding the Shine-Dalgarno sequence. Only the binding-competent state allows SAM binding to produce the SAM captured closed state where non-canonical base pair interactions lock the ligand-riboswitch complex.

0.4 was detected, which decreased to a value of 0.35 upon the addition of magnesium ions. This behavior is independent of the presence of guanine.

Lastly, the P2-P3 vector followed a folding transition similar to the one observed for *add* aptamer (Dambach, Sandoval and al. 2015). In this case, two populations were visible without magnesium (E_{FRET} values of 0.4 and 0.9). Those two populations collapsed into a single one exhibiting an E_{FRET} value 0.8 upon addition of magnesium. The folding transition was also found to be independent of guanine. This is in contrast with what was observed for the *add* aptamer where three states were identified in presence of ligand at low concentrations of magnesium ions (Fig. 5a). This difference may be attributed to different folding pathways that adenine and guanine aptamers employ to achieve their native state. Another difference found between both aptamers concerns the rate of folding. While both folding and unfolding rates of *add* aptamer were shown to be less heterogeneous upon addition of adenine (Dambach, Sandoval and al. 2015), the *xpt* aptamer was found to only be slightly affected by the presence of guanine. Therefore, in agreement with the sm-FRET data available, it is probable that even with their considerable sequence and structural homology the *xpt* and *add* aptamer have different folding pathways. More data will be needed to understand how two similar aptamers may employ different folding pathways to attain their respective native states.

SAM riboswitches can be categorized into six classes according to their predicted secondary structure. The aptamer domain of the SAM-I riboswitch is arranged around a four-way junction and is able to specifically recognize the *S*-adenosyl methionine (SAM) metabolite. To determine the folding pathway of the aptamer domain, several FRET vectors were engineered to monitor the formation of specific regions of the aptamer (Heppell, Blouin and al. 2011). For example, given that the crystal structure shows that P1 and P3 stems are arranged into close juxtaposition to form a part of the ligand binding site, a FRET vector was designed to monitor the folding transition occurring between stems P1 and P3. While bulk FRET experiments showed an increase of E_{FRET} as a function of magnesium ions, no change was observed when 5 μ M SAM was added to the solution. These results suggest that the binding of magnesium ions

is sufficient to promote the P1-P3 close juxtaposition, independently of SAM (Fig. 5b). On the contrary, when analyzing the vector P1-P4, changes in FRET were only observed upon the addition of SAM. In this case, a SAM-dependent decrease in FRET was monitored, suggesting that SAM probably induces the coaxial stacking of helices P1 and P4 (Fig. 5b), as observed in the crystal. Lastly, the vector P3-P4 was found to report both magnesium- and SAM-folding transitions, reflecting data obtained for both vectors 1-3 and 1-4. Together with results obtained from comparative gel electrophoresis and 2-aminopurine substitutions discussed in previous sections, these results were interpreted in relation to the crystal structure of the SAM-I riboswitch (Ren, Rajashankar and al. 2012) and allowed to establish the folding pathway of the aptamer as a function of both magnesium ions and SAM cofactors. While the addition of magnesium ions induces the formation of the pseudoknot structure, P2-P3 helical stacking unit and the P1 and P3 close juxtaposition, the binding of SAM to the aptamer promotes the stacking of P1 and P4 and the reorganization of the aptamer core domain (Fig. 5b).

To further characterize the conformational changes taking place in the SAM aptamer, sm-FRET experiments were performed using the vector P1-P3. In absence of magnesium ions, sm-FRET histograms showed a single Gaussian peak with an apparent FRET efficiency of ~ 0.28 . Increasing the concentration of magnesium ions (1 mM) resulted in the formation of a major peak progressively shifting to a higher FRET value of ~ 0.4 . These peaks were respectively associated to the unfolded (U) and Mg^{2+} -induced folded state (F_{Mg}), respectively. Sm-FRET traces of molecules at low concentrations of magnesium ions concentration revealed that the shifting peak corresponded to a mixture of U and F_{Mg} states and that the fast dynamic of the RNA accounted for the shifting of the major peak. At higher concentration of magnesium ions (3 mM), a peak, corresponding to 32% of the molecules, was also observed with a FRET efficiency of 0.62. The proportion of molecules exhibiting this FRET value did not increase any further when the concentration of $MgCl_2$ was increased to 10 mM. However, this peak became highly prominent when 100 μM SAM ligand was added to the sample, suggesting that SAM binding induce an additional reorganization of the P1-P3 vector. This bound state was designated as the native state (F_{NS}). Analysis of sm-FRET traces for single molecules

demonstrated that an interconversion exists between U, F_{Mg} and F_{NS} states at intermediate concentrations of Mg^{2+} ions in absence of SAM, and that increasing concentrations of Mg^{2+} mostly promoted transition between the F_{Mg} and F_{NS} state. It also showed that SAM slows the interconversion dynamics by stabilizing the F_{NS} state. Further investigation of the F_{NS} state allowed the identification of the novel folding mechanism in which ligand binding induces the rotation of the P1 stem around its helical axis (Heppell, Blouin and al. 2011). The folding pathway identified in this study is reported here in Fig. 5b. Additional experiments need to be performed to fully understand the mechanism and role of P1 twisting in riboswitch genetic regulation.

SAM-II riboswitches are particularly interesting because of their small size, which corresponds to 52 nt for the whole riboswitch. The crystal structure of the metX SAM-II riboswitch bound to SAM shows that it folds into a classical H-type pseudoknot (Haller, Rieder and al. 2011). Because of its small size, it was the first complete riboswitch, with aptamer and expression platform, to be studied using sm-FRET. Two conformations in solution were identified for this riboswitch, one with a low FRET (0.30), where the RNA is relatively open, and one with a high FRET (0.85) where the pseudoknot is formed. Both states were observed in the absence of Mg^{2+} ions and it was shown that the addition of Mg^{2+} ions stabilized the high FRET state by increasing the dwell time of the RNA in the high-FRET conformation. In the presence of Mg^{2+} ions and SAM ligand the riboswitch showed a dominant high-FRET population suggesting a ligand-induced stabilization of a compacted pseudoknot conformation. This was confirmed by seeing only rare low-FRET events in individual traces. These data allowed to understand the ligand-free structure of the full length SAM-II riboswitch and also helped to demonstrate that the SAM-II riboswitch follows a conformational capture binding mechanism as it can be seen in Fig. 5c (Haller, Rieder and al. 2011).

Three other riboswitches have been studied using sm-FRET, namely the *lysC* lysine riboswitch from *Bacillus subtilis* (Fiegland, Garst and al. 2012), the *thiM* TPP riboswitch from

Escherichia coli (Haller, Altman and al. 2013) and the *tfoX* c-di-GMP riboswitch from *Vibrio vulnificus* (Wood, Ferre-D'Amare and al. 2012). In the case of the lysine riboswitch (Fiegland, Garst and al. 2012), sm-FRET experiments aimed at understanding the kinetic between the free aptamer and the aptamer bound to its ligand. It was shown that the rate of unfolding (k_{open}) value did not change much in function of the concentration of lysine while the rate of folding (k_{close}) had a significant dependence on the concentration of lysine as exemplified by its 20-fold increase from 25 μM to 2000 μM lysine. In this study, the average dwell time for the aptamer was estimated to be 0.7s and 5 s for the open and closed conformation, respectively. If these values are indicative of *in vivo* conditions, it means that while the aptamer is able to bind lysine on a very short timescale window (0.7s), it releases lysine only after a significant slower timescale (5 s). Thus, these results suggest that once the aptamer region of the riboswitch is transcribed, it should immediately bind available lysine. However, since lysine should not be released before 5 s, it should allow enough time to transcribe the expression platform and take the decision to stop gene expression (Fiegland, Garst and al. 2012). Moreover, this study suggests that the lysine aptamer can adopt three different conformations. Conversely to what was observed for the purine aptamer (Dambach, Sandoval and al. 2015), only two FRET states were observed for the lysine aptamer. The hypothesis of an additional third conformation arises from the analysis of the dynamic of the lysine aptamer observed using sm-FRET in freely diffusing molecules as well as tethered ones. No conformation for this third state is predicted yet (Fiegland, Garst and al. 2012). This is interesting considering that all the riboswitches studied by sm-FRET so far show three major conformations. In addition to the unfolded and folded conformations, the third conformation is usually identified as a pre-folded apo-complex ready to bind the ligand.

9. Single-molecule chemical denaturation methods

Undoubtedly, the application of single-molecule FRET-based methods to investigate riboswitch dynamics has already provided a wealth of information about the regulatory

mechanism difficult to obtain by any other method as discussed in the previous section. However, it is worth noting that although sharing a common objective, the application of these smFRET techniques in the protein and RNA folding fields has followed different approaches. Studies in the single-molecule protein folding field have predominantly used denaturant agents (i.e. urea, guanidinium chloride) to elucidate the unfolding mechanism and the rate-limiting step (Schuler and Eaton 2008, Nienhaus 2009). In contrast, single-molecule RNA science has focused almost exclusively on understanding the effect of metal ions as folding promoters and stabilizers of the RNA structure (Bokinsky and Zhuang 2005). Interestingly, a recent study has highlighted the advantages of merging both approaches to investigate the dynamics and stability of RNA-ligand complexes (Dalgarno, Bordello and al. 2013). In this study, smFRET was combined with chemical denaturation to demonstrate that the competing interplay between folding (Mg^{2+} , ligand) and unfolding agents such as urea enables the extraction of information that otherwise will remain hidden using conventional smFRET approaches (Dalgarno, Bordello and al. 2013). The *add* riboswitch aptamer from *V. vulnificus* was labeled as in previous studies (Dambach, Sandoval and al. 2015) with the Cy3-Cy5 FRET pair at the P2 and P3 stems loops and used as a model compound to explore the potential of merging smFRET and urea-induced denaturation. Aptamers incubated at 4 mM Mg^{2+} to induce the formation of the loop-loop interaction (docked state, $E_{\text{app}} \sim 0.9$) were subjected to increasing concentrations of urea to promote the formation of the undocked state (UD, $E_{\text{app}} \sim 0.3$). It is worth to noting that the E_{app} values obtained in the presence of urea for both states, docked and undocked, were in excellent agreement with those previously reported when examining the folding pathway exclusively in the presence of Mg^{2+} ions (Dambach, Sandoval and al. 2015), suggesting that urea-induced unfolding takes place through identical conformational states as the Mg^{2+} -induced folding process. Single-molecule histograms obtained under these conditions for the ligand-free docked state (D_{LF}) (Fig. 6a) and the ligand-bound docked state (D_{LB}) (Fig. 6b) revealed distinct effects induced by ligand binding. As expected, in the presence of adenine ligand, higher concentrations of urea were required to disrupt the loop-loop interaction to levels similar to those observed for the ligand-free state. More importantly, a significantly higher dynamic heterogeneity for the undocked step was observed as the concentration of ligand increased. In fact, the single-molecule dwell-time histograms of

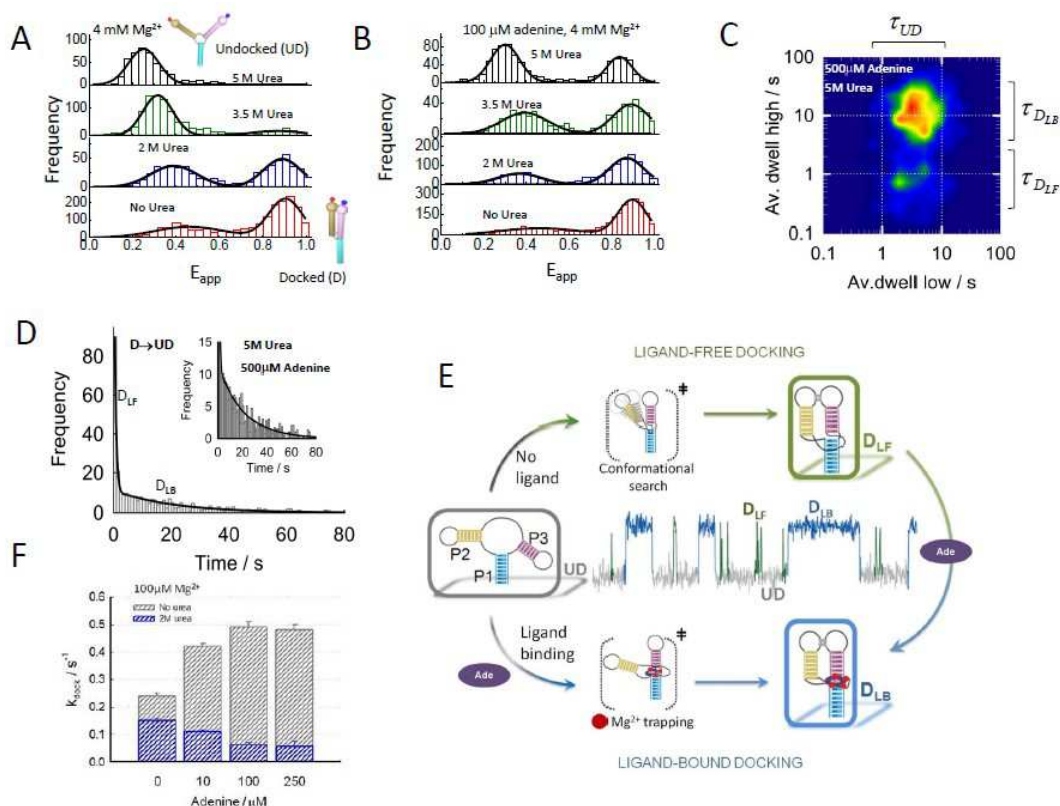


Fig. 6. Single-molecule chemical denaturation methods to investigate RNA-ligand interactions in the context of the adenine aptamer [Adapted from reference 82]. (A) Urea-induced undocking of ligand-free adenine aptamers. Single-molecule histograms of FRET efficiency are shown as a function of urea concentration. All histograms were obtained in a background of 4 mM Mg^{2+} and 50 mM Na^+ ions. (B) Urea-induced undocking of ligand-bound adenine aptamers. Single-molecule FRET histograms were obtained at the indicated concentrations of urea in a background of 4 mM Mg^{2+} ions and 100 μM adenine ligand. (C) Two-dimensional contour plot showing the distribution of average docking and undocking dwell times at 4mM Mg^{2+} , in the presence of 5 M urea and 500 μM adenine ligand. (D) Single-molecule dwell time histogram for the undocking process at the conditions indicated in (C). Solid line represents the non-linear squares fitting to a bi-exponential decay function. From the fitting we obtained values of rate constants $2.1 \pm 0.1 \text{ s}^{-1}$ (minor component) and $0.045 \pm 0.003 \text{ s}^{-1}$ (major component). (E) Model proposed for ligand-induced switching of the rate-limiting step for adenine aptamer docking. In the absence of ligand, the rate-limiting step for stabilization of the ligand-free docked state (D_{LF}) involves the formation of key tertiary contacts between the undocked (UD) P2 and P3 stem loops (*conformational search model*). Ligand binding to partially folded or unfolded aptamer domains induces a switch to a rate-limiting step involving the consolidation of the ligand-aptamer complex (D_{LB}) around specifically captured Mg^{2+} ions (*ion trapping model*). Inset: representative single-molecule FRET trajectory showing the coexistence of ligand-free (green), ligand-bound (blue) and undocked states (grey) that can be distinguish by the different stability against urea-induced denaturation [Adapted from reference 80].

undocking rates (Fig. 6c) obtained at 4mM Mg^{2+} , 5 M urea and 500 μ M adenine ligand showed a clear bi-exponential decay (Fig. 6d). Based on the variation of the relative contribution of each rate with respect to the amount of ligand present, it was concluded that the fast component, with a value similar to that obtained in the absence of ligand, corresponded to urea-induced undocking events taking place from ligand-free states ($\sim 2 \pm 0.01$ s⁻¹). The slow component ($\sim 0.045 \pm 0.003$ s⁻¹), whose contribution increased with added ligand, was assigned to undocking events from ligand-bound states. The relative magnitude of both rates was used to estimate a 50-fold ligand-induced stabilization of the docked state due to ligand binding. It is interesting to realize that the D_{LB} and D_{LF} states have exactly the same E_{app} value (~ 0.9) and therefore conventional smFRET techniques cannot distinguish between them. Remarkably, the combined approach resolved this ambiguity even for D_{LB} and D_{LF} states coexisting within a single FRET trajectory (Fig. 6e, inset trace), purely using their different stability against urea-induced denaturation. Although still in its infancy, chemical denaturation methods hold significant potential in smFRET studies of RNA-ligand complexes for which ligand binding does not induce significant conformational changes in the RNA platform.

In the same study, the combined approach was used to extract information regarding the nature of the rate-limiting step for the formation of the loop-loop docked state and how this might be influenced by the ligand (Dalgarno, Bordello and al. 2013). The application of chemical denaturation methods to investigate rate-limiting steps in RNA folding is not new in ensemble studies (Pan and Sosnick 1997, Bartley, Zhuang and al. 2003, Sosnick and Pan 2003) but no systematic study was previously reported at the single-molecule level. Comparative analysis of the docking rates of the *add* adenine aptamer with and without added ligand as a function urea concentration revealed previously unknown details, in particular regarding the influence of ligand binding at early stages of the folding process. For the ligand-free state, the docking rate progressively decreased in the presence of increasing concentrations of urea. Moreover, the relative change between docking rates in the presence and absence of denaturant (~ 2 -3 fold) was practically constant across a 40-fold variation in the

concentration of Mg^{2+} ions. Urea-induced acceleration of the folding rate has been observed for many RNAs and it is accepted to be a hallmark of the presence of kinetic traps and/or misfolded states along the folding pathway (Pan and Sosnick 1997, Sosnick and Pan 2003). In contrast, the observation of a decrease in the loop-loop docking rate in the presence of urea and its independence of the concentration of divalent metal ions were taken as evidence for a trap-free conformational search mechanism for the folding of the ligand-free *add* adenine aptamer (Fig. 6e). Urea-induced deceleration of the docking kinetics has also been reported for the two-way hairpin ribozyme, which also carries two stem loops interacting with each other in the docked state (Bokinsky, Rueda and al. 2003). Based on the analogies between both RNA elements, a rate-limiting step involving productive encounters between the peripheral loops to trigger the formation of specific tertiary contacts was hypothesized as a feature that may be common to other functional RNAs.

In the presence of ligand, urea also decreased the docking rate but this effect was now strongly dependent on the concentration of Mg^{2+} ions. The docking rate decreased by 2-fold at saturating Mg^{2+} (4mM) and by ~8-fold at Mg^{2+} concentrations where the docked state should only be transiently formed (~100 μ M). This dependence was taken as evidence for a different rate-limiting step for the ligand-bound state in which divalent metal ions are essential to achieve the native state. Importantly, at sub-saturating concentrations of Mg^{2+} ions, the relative decrease in k_{dock} was a function of the ligand concentration (Fig. 6f). Based on these findings, it was proposed that when loop-loop docking is compromised, the formation of partially folded ligand-bound states requires the assistance of Mg^{2+} ions specifically positioned along the aptamer structure (Fig. 6e). Indeed, such a Mg^{2+} -trapping rate-limiting step could provide an additional mechanistic role for the five Mg^{2+} ions present in the crystal structure of the aptamer domain (Serganov, Yuan and al. 2004), two of them located near the ligand-binding pocket. The proposed overall mechanism is consistent with an 'induced-fit' model for ligand binding as suggested by other biochemical and NMR studies (Gelfand, Mironov and al. 1999, Kubodera, Watanabe and al. 2003, Noeske, Schwalbe and al. 2007, Sharma, Bulusu and al. 2009, Haller, Rieder and al. 2011). However, the extend of the ligand-

induced conformational changes might be just restricted to the aptamer core, at saturating Mg^{2+} concentrations, or might also involved the formation of long-range tertiary contacts when Mg^{2+} ions are scarce. In the latter scenario, specifically trapped Mg^{2+} ions could be critical to funnel an ensemble of partially docked/folded states towards the tightly packed native state.

10. Future directions

Riboswitches heavily rely on the balance between folding dynamics, binding kinetics and transcription rate to sense the metabolite concentration and to achieve genetic regulation. To date, most studies on riboswitch function have focused on elucidating the conformational changes taking place during the transition from a ligand-free to a ligand-bound state and the influence of ligand binding along this process. However, a deeper understanding of the switching mechanism and how metabolite-recognition is coupled to the expression machinery requires to move from aptamer-only studies to full-length riboswitches. In this context, since most riboswitches have relatively high molecular weights, novel methods combining chemical and enzymatic reactions are needed to produce in higher yields fluorescent mRNAs. Moreover, although several techniques have been used to characterize the mechanism of ligand-binding, they all relied on indirect methods and almost no information is available and no method has been developed to directly report ligand-binding kinetics in real time. This is due to the fact that very few fluorescent natural ligands are available, which restrains the advancement in the field. The development of fluorescent ligands, or ligands carrying a fluorescent moiety, showing unperturbed metabolite-recognition properties is necessary to gather essential knowledge during the sensing phase of the regulatory process. Finally, it is known that for certain riboswitches mostly operating under kinetic control, the delicate balance between the rate of ligand binding and the transcription rate dictates the regulatory outcome. Thus, new fluorescence-based methods at single-molecule level are eagerly awaited to address the importance of co-transcriptional folding in riboswitch function.

References

- [1] A. Serganov, E. Nudler, A decade of riboswitches, *Cell*, 152 (2013) 17-24.
- [2] R.R. Breaker, Riboswitches and the RNA World, *Cold Spring Harb Perspect Biol*, 4 (2012) 1-15.
- [3] R.K. Montange, R.T. Batey, Riboswitches: emerging themes in RNA structure and function, *Annu Rev Biophys*, 37 (2008) 117-133.
- [4] L.M. Wilhelmsson, Fluorescent nucleic acid base analogues, *Quarterly reviews of biophysics*, 43 (2010) 159-183.
- [5] M.J. Rist, J.P. Marino, Fluorescent nucleotide base analogs as probes of nucleic acid structure, dynamics and interactions, *Curr. Org. Chem.*, 6 (2002) 775-793.
- [6] C.R. Guest, R.A. Hochstrasser, L.C. Sowers, D.P. Millar, Dynamics of mismatched base pairs in DNA, *Biochemistry*, 30 (1991) 3271-3279.
- [7] E.L. Rachofsky, R. Osman, J.B. Ross, Probing structure and dynamics of DNA with 2-aminopurine: effects of local environment on fluorescence, *Biochemistry*, 40 (2001) 946-956.
- [8] J.T. Stivers, 2-Aminopurine fluorescence studies of base stacking interactions at abasic sites in DNA: metal-ion and base sequence effects, *Nucleic Acids Res*, 26 (1998) 3837-3844.
- [9] D.C. Ward, E. Reich, L. Stryer, Fluorescence studies of nucleotides and polynucleotides. I. Formycin, 2-aminopurine riboside, 2,6-diaminopurine riboside, and their derivatives, *J Biol Chem*, 244 (1969) 1228-1237.
- [10] A. Holmen, B. Norden, B. Albinsson, Electronic Transition Moments of 2-Aminopurine, *J. Am. Chem. Soc.*, 119 (1997) 3114-3121.
- [11] J.M. Jean, K.B. Hall, 2-Aminopurine fluorescence quenching and lifetimes: role of base stacking, *Proc Natl Acad Sci U S A*, 98 (2001) 37-41.

- [12] J.M. Jean, K.B. Hall, 2-Aminopurine electronic structure and fluorescence properties in DNA, *Biochemistry*, 41 (2002) 13152-13161.
- [13] S.J. Hardman, K.C. Thompson, Influence of base stacking and hydrogen bonding on the fluorescence of 2-aminopurine and pyrrolocytosine in nucleic acids, *Biochemistry*, 45 (2006) 9145-9155.
- [14] P. Rai, T.D. Cole, E. Thompson, D.P. Millar, S. Linn, Steady-state and time-resolved fluorescence studies indicate an unusual conformation of 2-aminopurine within ATAT and TATA duplex DNA sequences, *Nucleic Acids Res*, 31 (2003) 2323-2332.
- [15] M.A. O'Neill, J.K. Barton, DNA-mediated charge transport requires conformational motion of the DNA bases: elimination of charge transport in rigid glasses at 77 K, *J Am Chem Soc*, 126 (2004) 13234-13235.
- [16] B. Holz, S. Klimasauskas, S. Serva, E. Weinhold, 2-Aminopurine as a fluorescent probe for DNA base flipping by methyltransferases, *Nucleic Acids Res*, 26 (1998) 1076-1083.
- [17] R.K. Neely, D. Daujotyte, S. Grazulis, S.W. Magennis, D.T. Dryden, S. Klimasauskas, A.C. Jones, Time-resolved fluorescence of 2-aminopurine as a probe of base flipping in M.HhaI-DNA complexes, *Nucleic Acids Res*, 33 (2005) 6953-6960.
- [18] M. Rist, J. Marino, Association of an RNA kissing complex analyzed using 2-aminopurine fluorescence, *Nucleic Acids Res*, 29 (2001) 2401-2408.
- [19] S.M. Law, R. Eritja, M.F. Goodman, K.J. Breslauer, Spectroscopic and calorimetric characterizations of DNA duplexes containing 2-aminopurine, *Biochemistry*, 35 (1996) 12329-12337.
- [20] J.D. Ballin, S. Bharill, E.J. Fialcowitz-White, I. Gryczynski, Z. Gryczynski, G.M. Wilson, Site-specific variations in RNA folding thermodynamics visualized by 2-aminopurine fluorescence, *Biochemistry*, 46 (2007) 13948-13960.

- [21] L.B. Bloom, M.R. Otto, J.M. Beechem, M.F. Goodman, Influence of 5'-nearest neighbors on the insertion kinetics of the fluorescent nucleotide analog 2-aminopurine by Klenow fragment, *Biochemistry*, 32 (1993) 11247-11258.
- [22] C. Hariharan, L.J. Reha-Krantz, Using 2-aminopurine fluorescence to detect bacteriophage T4 DNA polymerase-DNA complexes that are important for primer extension and proofreading reactions, *Biochemistry*, 44 (2005) 15674-15684.
- [23] L.D. Finger, N. Patel, A. Beddows, L. Ma, J.C. Exell, E. Jardine, A.C. Jones, J.A. Grasby, Observation of unpaired substrate DNA in the flap endonuclease-1 active site, *Nucleic Acids Res*, 41 (2013) 9839-9847.
- [24] C.A. Dunlap, M.D. Tsai, Use of 2-aminopurine and tryptophan fluorescence as probes in kinetic analyses of DNA polymerase beta, *Biochemistry*, 41 (2002) 11226-11235.
- [25] N.G. Walter, D.A. Harris, M.J. Pereira, D. Rueda, In the fluorescent spotlight: global and local conformational changes of small catalytic RNAs, *Biopolymers*, 61 (2001) 224-242.
- [26] M. Menger, T. Tuschl, F. Eckstein, D. Porschke, Mg(2+)-dependent conformational changes in the hammerhead ribozyme, *Biochemistry*, 35 (1996) 14710-14716.
- [27] D.A. Harris, D. Rueda, N.G. Walter, Local conformational changes in the catalytic core of the trans-acting hepatitis delta virus ribozyme accompany catalysis, *Biochemistry*, 41 (2002) 12051-12061.
- [28] N.G. Walter, P.A. Chan, K.J. Hampel, D.P. Millar, J.M. Burke, A base change in the catalytic core of the hairpin ribozyme perturbs function but not domain docking, *Biochemistry*, 40 (2001) 2580-2587.
- [29] J.K. Wickiser, M.T. Cheah, R.R. Breaker, D.M. Crothers, The kinetics of ligand binding by an adenine-sensing riboswitch, *Biochemistry*, 44 (2005) 13404-13414.

- [30] R. Rieder, K. Lang, D. Graber, R. Micura, Ligand-Induced Folding of the Adenosine Deaminase A-Riboswitch and Implications on Riboswitch Translational Control, *Chembiochem*, 8 (2007) 896-902.
- [31] J.F. Lemay, J.C. Penedo, R. Tremblay, D.M. Lilley, D.A. Lafontaine, Folding of the adenine riboswitch, *Chem Biol*, 13 (2006) 857-868.
- [32] A. Haller, M.F. Souliere, R. Micura, The Dynamic Nature of RNA as Key to Understanding Riboswitch Mechanisms, *Acc Chem Res*, (2011).
- [33] K. Lang, R. Rieder, R. Micura, Ligand-induced folding of the thiM TPP riboswitch investigated by a structure-based fluorescence spectroscopic approach, *Nucleic Acids Res*, 35 (2007) 5370-5378.
- [34] S. Blouin, R. Chinnappan, D.A. Lafontaine, Folding of the lysine riboswitch: importance of peripheral elements for transcriptional regulation, *Nucleic Acids Res*, 39 (2010) 3373-3387.
- [35] D.A. Berry, K.Y. Jung, D.S. Wise, A.D. Sercel, W.H. Person, H. Mackie, J.B. Randolph, R.L. Somers, Pyrrolo-dC and pyrrolo-C: fluorescent analogs of cytidine and 2'-deoxycytidine for the study of oligonucleotides, *Tetrahedron Lett*, 45 (2004) 2457-2461.
- [36] C. Liu, C.T. Martin, Fluorescence characterization of the transcription bubble in elongation complexes of T7 RNA polymerase, *J Mol Biol*, 308 (2001) 465-475.
- [37] X. Zhang, R.M. Wadkins, DNA hairpins containing the cytidine analog pyrrolo-dC: structural, thermodynamic, and spectroscopic studies, *Biophys J*, 96 (2009) 1884-1891.
- [38] C. Dash, J.W. Rausch, S.F. Le Grice, Using pyrrolo-deoxycytosine to probe RNA/DNA hybrids containing the human immunodeficiency virus type-1 3' polypurine tract, *Nucleic Acids Res*, 32 (2004) 1539-1547.
- [39] H. Zang, Q. Fang, A.E. Pegg, F.P. Guengerich, Kinetic analysis of steps in the repair of damaged DNA by human O6-alkylguanine-DNA alkyltransferase, *J Biol Chem*, 280 (2005) 30873-30881.

- [40] A.A. Marti, X. Li, S. Jockusch, Z. Li, B. Raveendra, S. Kalachikov, J.J. Russo, I. Morozova, S.V. Puthanveetil, J. Ju, N.J. Turro, Pyrene binary probes for unambiguous detection of mRNA using time-resolved fluorescence spectroscopy, *Nucleic Acids Res*, 34 (2006) 3161-3168.
- [41] W. Huang, J. Kim, S. Jha, F. Aboul-Ela, Conformational heterogeneity of the SAM-I riboswitch transcriptional ON state: a chaperone-like role for S-adenosyl methionine, *J Mol Biol*, 418 (2012) 331-349.
- [42] B. Heppell, S. Blouin, A.M. Dussault, J. Mulhbacher, E. Ennifar, J.C. Penedo, D.A. Lafontaine, Molecular insights into the ligand-controlled organization of the SAM-I riboswitch, *Nature Chem Biol*, 7 (2011) 384-392.
- [43] A.D. Garst, E.B. Porter, R.T. Batey, Insights into the regulatory landscape of the lysine riboswitch, *J Mol Biol*, 423 (2012) 17-33.
- [44] B. Heppell, D.A. Lafontaine, Folding of the SAM aptamer is determined by the formation of a K-turndependent pseudoknot, *Biochemistry*, 47 (2008) 1490-1499.
- [45] S.D. Gilbert, C.D. Stoddard, S.J. Wise, R.T. Batey, Thermodynamic and Kinetic Characterization of Ligand Binding to the Purine Riboswitch Aptamer Domain, *J Mol Biol*, 359 (2006) 754-768.
- [46] A. Haller, U. Rieder, M. Aigner, S.C. Blanchard, R. Micura, Conformational capture of the SAM-II riboswitch, *Nat Chem Biol*, 7 (2011) 393-400.
- [47] B. Heppell, J. Mulhbacher, J.C. Penedo, D.A. Lafontaine, Application of fluorescent measurements for characterization of riboswitch-ligand interactions, *Methods Mol Biol*, 540 (2009) 25-37.
- [48] J.F. Lemay, D.A. Lafontaine, Core requirements of the adenine riboswitch aptamer for ligand binding, *RNA*, 13 (2007) 339-350.

- [49] R. Tremblay, J.F. Lemay, S. Blouin, J. Mulhbacher, E. Bonneau, P. Legault, P. Dupont, J.C. Penedo, D.A. Lafontaine, Constitutive regulatory activity of an evolutionary-excluded riboswitch variant, *J Biol Chem*, 286 (2011) 27406-27415.
- [50] M.F. Souliere, A. Haller, R. Rieder, R. Micura, A Powerful Approach for the Selection of 2-Aminopurine Substitution Sites to Investigate RNA Folding, *J Am Chem Soc*, (2011).
- [51] M.F. Souliere, R. Micura, Use of SHAPE to select 2AP substitution sites for RNA-ligand interactions and dynamics studies, *Methods Mol Biol*, 1103 (2014) 227-239.
- [52] K.A. Wilkinson, E.J. Merino, K.M. Weeks, Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution, *Nat Protoc*, 1 (2006) 1610-1616.
- [53] S.A. Mortimer, K.M. Weeks, Time-resolved RNA SHAPE chemistry: quantitative RNA structure analysis in one-second snapshots and at single-nucleotide resolution, *Nat Protoc*, 4 (2009) 1413-1421.
- [54] A. Serganov, Y.R. Yuan, O. Pikovskaya, A. Polonskaia, L. Malinina, A.T. Phan, C. Hobartner, R. Micura, R.R. Breaker, D.J. Patel, Structural Basis for Discriminative Regulation of Gene Expression by Adenine- and Guanine-Sensing mRNAs, *Chem Biol*, 11 (2004) 1729-1741.
- [55] J.F. Lemay, J.C. Penedo, J. Mulhbacher, D.A. Lafontaine, Molecular basis of RNA-mediated gene regulation on the adenine riboswitch by single-molecule approaches, *Methods Mol Biol*, 540 (2009) 65-76.
- [56] J. Noeske, H. Schwalbe, J. Wohnert, Metal-ion binding and metal-ion induced folding of the adenine-sensing riboswitch aptamer domain, *Nucleic Acids Res*, 35 (2007) 5262-5273.
- [57] M. Sharma, G. Bulusu, A. Mitra, MD simulations of ligand-bound and ligand-free aptamer: molecular level insights into the binding and switching mechanism of the add A-riboswitch, *RNA*, 15 (2009) 1673-1692.

- [58] V. Delfosse, P. Bouchard, E. Bonneau, P. Dagenais, J.F. Lemay, D.A. Lafontaine, P. Legault, Riboswitch structure: an internal residue mimicking the purine ligand, *Nucleic Acids Res*, 38 (2010) 2057-2068.
- [59] D. Leipply, D.E. Draper, Effects of Mg²⁺ on the free energy landscape for folding a purine riboswitch RNA, *Biochemistry*, 50 (2011) 2790-2799.
- [60] J. Noeske, C. Richter, M.A. Grundl, H.R. Nasiri, H. Schwalbe, J. Wohnert, An intermolecular base triple as the basis of ligand specificity and affinity in the guanine- and adenine-sensing riboswitch RNAs, *Proc Natl Acad Sci U S A*, 102 (2005) 1372-1377.
- [61] C.D. Stoddard, S.D. Gilbert, R.T. Batey, Ligand-dependent folding of the three-way junction in the purine riboswitch, *Rna*, 14 (2008) 675-684.
- [62] O. Prychyna, M.S. Dahabieh, J. Chao, M.A. O'Neill, Sequence-dependent folding and unfolding of ligand-bound purine riboswitches, *Biopolymers*, 91 (2009) 953-965.
- [63] N. Jain, L. Zhao, J.D. Liu, T. Xia, Heterogeneity and dynamics of the ligand recognition mode in purine-sensing riboswitches, *Biochemistry*, 49 (2010) 3703-3714.
- [64] J. Buck, B. Furtig, J. Noeske, J. Wohnert, H. Schwalbe, Time-resolved NMR methods resolving ligand-induced RNA folding at atomic resolution, *Proc Natl Acad Sci U S A*, 104 (2007) 15699-15704.
- [65] F.M. Jucker, R.M. Phillips, S.A. McCallum, A. Pardi, Role of a heterogeneous free state in the formation of a specific RNA-theophylline complex, *Biochemistry*, 42 (2003) 2560-2567.
- [66] J.F. Lemay, G. Desnoyers, S. Blouin, B. Heppell, L. Bastet, P. St-Pierre, E. Masse, D.A. Lafontaine, Comparative Study between Transcriptionally- and Translationally-Acting Adenine Riboswitches Reveals Key Differences in Riboswitch Regulatory Mechanisms, *PLoS Genet*, 7 (2011) e1001278.

- [67] U. Rieder, C. Kreutz, R. Micura, Folding of a transcriptionally acting preQ1 riboswitch, *Proc Natl Acad Sci U S A*, 107 (2010) 10804-10809.
- [68] W.C. Winkler, A. Nahvi, N. Sudarsan, J.E. Barrick, R.R. Breaker, An mRNA structure that controls gene expression by binding S-adenosylmethionine, *Nature structural biology*, 10 (2003) 701-707.
- [69] R.K. Montange, R.T. Batey, Structure of the S-adenosylmethionine riboswitch regulatory mRNA element, *Nature*, 441 (2006) 1172-1175.
- [70] S.D. Gilbert, R.P. Rambo, D. Van Tyne, R.T. Batey, Structure of the SAM-II riboswitch bound to S-adenosylmethionine, *Nat Struct Mol Biol*, 15 (2008) 177-182.
- [71] S.N. Wilson-Mitchell, F.J. Grundy, T.M. Henkin, Analysis of lysine recognition and specificity of the *Bacillus subtilis* L box riboswitch, *Nucleic Acids Res*, 40 (2012) 5706-5717.
- [72] M.H. Buckstein, J. He, H. Rubin, Characterization of nucleotide pools as a function of physiological state in *Escherichia coli*, *J Bacteriol*, 190 (2008) 718-726.
- [73] S. Eskandari, O. Prychyna, J. Leung, D. Avdic, A. O'Neill M, Ligand-Directed Dynamics of Adenine Riboswitch Conformers, *J Am Chem Soc*, (2007).
- [74] N. Sudarsan, J.K. Wickiser, S. Nakamura, M.S. Ebert, R.R. Breaker, An mRNA structure in bacteria that controls gene expression by binding lysine, *Genes Dev*, 17 (2003) 2688-2697.
- [75] W.C. Winkler, R.R. Breaker, Genetic control by metabolite-binding riboswitches, *ChemBiochem*, 4 (2003) 1024-1032.
- [76] W.C. Winkler, S. Cohen-Chalamish, R.R. Breaker, An mRNA structure that controls gene expression by binding FMN, *Proc Natl Acad Sci U S A*, 99 (2002) 15908-15913.
- [77] R.K. Neely, S.W. Magennis, S. Parsons, A.C. Jones, Photophysics and X-ray structure of crystalline 2-aminopurine, *Chemphyschem : a European journal of chemical physics and physical chemistry*, 8 (2007) 1095-1102.

- [78] U.D. Priyakumar, A.D. MacKerell, Jr., Role of the adenine ligand on the stabilization of the secondary and tertiary interactions in the adenine riboswitch, *J Mol Biol*, 396 (2010) 1422-1438.
- [79] A. Villa, J. Wohnert, G. Stock, Molecular dynamics simulation study of the binding of purine bases to the aptamer domain of the guanine sensing riboswitch, *Nucleic Acids Res*, 37 (2009) 4774-4786.
- [80] P.R. Selvin, The renaissance of fluorescence resonance energy transfer, *Nature structural biology*, 7 (2000) 730-734.
- [81] R. Roy, S. Hohng, T. Ha, A practical guide to single-molecule FRET, *Nat Methods*, 5 (2008) 507-516.
- [82] P.A. Dalgarno, J. Bordello, R. Morris, P. St-Pierre, A. Dube, I.D. Samuel, D.A. Lafontaine, J.C. Penedo, Single-molecule chemical denaturation of riboswitches, *Nucleic Acids Res*, 41 (2013) 4253-4265.
- [83] M.D. Brenner, M.S. Scanlan, M.K. Nahas, T. Ha, S.K. Silverman, Multivector fluorescence analysis of the xpt guanine riboswitch aptamer domain and the conformational role of guanine, *Biochemistry*, 49 (2010) 1596-1605.
- [84] L.R. Fiegand, A.D. Garst, R.T. Batey, D.J. Nesbitt, Single-molecule studies of the lysine riboswitch reveal effector-dependent conformational dynamics of the aptamer domain, *Biochemistry*, 51 (2012) 9223-9233.
- [85] A. Haller, R.B. Altman, M.F. Souliere, S.C. Blanchard, R. Micura, Folding and ligand recognition of the TPP riboswitch aptamer at single-molecule resolution, *Proc Natl Acad Sci U S A*, 110 (2013) 4188-4193.
- [86] S. Wood, A.R. Ferre-D'Amare, D. Rueda, Allosteric tertiary interactions preorganize the c-di-GMP riboswitch and accelerate ligand binding, *ACS Chem Biol*, 7 (2012) 920-927.

- [87] G.U. Nienhaus, Single-molecule fluorescence studies of protein folding, *Methods Mol Biol*, 490 (2009) 311-337.
- [88] B. Schuler, W.A. Eaton, Protein folding studied by single-molecule FRET, *Curr Opin Struct Biol*, 18 (2008) 16-26.
- [89] G. Bokinsky, X. Zhuang, Single-molecule RNA folding, *Acc Chem Res*, 38 (2005) 566-573.
- [90] T.R. Sosnick, T. Pan, RNA folding: models and perspectives, *Curr Opin Struct Biol*, 13 (2003) 309-316.
- [91] L.E. Bartley, X. Zhuang, R. Das, S. Chu, D. Herschlag, Exploration of the transition state for tertiary structure formation between an RNA helix and a large structured RNA, *J Mol Biol*, 328 (2003) 1011-1026.
- [92] T. Pan, T.R. Sosnick, Intermediates and kinetic traps in the folding of a large ribozyme revealed by circular dichroism and UV absorbance spectroscopies and catalytic activity, *Nature structural biology*, 4 (1997) 931-938.
- [93] G. Bokinsky, D. Rueda, V.K. Misra, M.M. Rhodes, A. Gordus, H.P. Babcock, N.G. Walter, X. Zhuang, Single-molecule transition-state analysis of RNA folding, *Proc Natl Acad Sci U S A*, 100 (2003) 9302-9307.
- [94] S.P. Hennesly, I.V. Novikova, K.Y. Sanbonmatsu, The expression platform and the aptamer: cooperativity between Mg²⁺ and ligand in the SAM-I riboswitch, *Nucleic Acids Res*, 41 (2013) 1922-1935.
- [95] S.P. Hennesly, K.Y. Sanbonmatsu, Tertiary contacts control switching of the SAM-I riboswitch, *Nucleic Acids Res*, 39 (2011) 2416-2431.

BIBLIOGRAPHIE

Baltimore, D. (1970). "RNA-dependent DNA polymerase in virions of RNA tumour viruses." *Nature* 226(5252): 1209-1211.

Ban, N., P. Nissen, J. Hansen, P. B. Moore and T. A. Steitz (2000). "The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution." *Science* 289(5481): 905-920.

Barrick, J. E., K. A. Corbino, W. C. Winkler, A. Nahvi, M. Mandal, J. Collins, M. Lee, A. Roth, N. Sudarsan, I. Jona, J. K. Wickiser and R. R. Breaker (2004). "New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control." *Proc Natl Acad Sci U S A* 101(17): 6421-6426.

Batey, R. T. (2012). "Structure and mechanism of purine-binding riboswitches." *Q Rev Biophys* 45(3): 345-381.

Batey, R. T., S. D. Gilbert and R. K. Montange (2004). "Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine." *Nature* 432(7015): 411-415.

Berens, C., F. Groher and B. Suess (2015). "RNA aptamers as genetic control devices: the potential of riboswitches as synthetic elements for regulating gene expression." *Biotechnol J* 10(2): 246-257.

Blouin, S., R. Chinnappan and D. A. Lafontaine (2011). "Folding of the lysine riboswitch: importance of peripheral elements for transcriptional regulation." *Nucleic Acids Res* 39(8): 3373-3387.

Blouin, S. and D. A. Lafontaine (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(8): 1256-1267.

Blount, K. F. and R. R. Breaker (2006). "Riboswitches as antibacterial drug targets." *Nat Biotechnol* 24(12): 1558-1564.

Bolotin, A., B. Quinquis, A. Sorokin and S. D. Ehrlich (2005). "Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin." *Microbiology* 151(Pt 8): 2551-2561.

Brenner, M. D., M. S. Scanlan, M. K. Nahas, T. Ha and S. K. Silverman (2010). "Multivector fluorescence analysis of the xpt guanine riboswitch aptamer domain and the conformational role of guanine." *Biochemistry* 49(8): 1596-1605.

Brooks, K. M. and K. J. Hampel (2009). "A rate-limiting conformational step in the catalytic pathway of the glmS ribozyme." *Biochemistry* 48(24): 5669-5678.

Caron, M. P., L. Bastet, A. Lussier, M. Simoneau-Roy, E. Masse and D. A. Lafontaine (2012). "Dual-acting riboswitch control of translation initiation and mRNA decay." *Proc Natl Acad Sci U S A* 109(50): E3444-3453.

Cech, T. R. (2012). "The RNA worlds in context." *Cold Spring Harb Perspect Biol* 4(7): a006742.

Cech, T. R., A. J. Zaugg and P. J. Grabowski (1981). "In vitro splicing of the ribosomal RNA precursor of Tetrahymena: involvement of a guanosine nucleotide in the excision of the intervening sequence." *Cell* 27(3 Pt 2): 487-496.

Cheah, M. T., A. Wachter, N. Sudarsan and R. R. Breaker (2007). "Control of alternative RNA splicing and gene expression by eukaryotic riboswitches." *Nature* 447(7143): 497-500.

Christiansen, L. C., S. Schou, P. Nygaard and H. H. Saxild (1997). "Xanthine metabolism in *Bacillus subtilis*: characterization of the xpt-pbuX operon and evidence for purine- and nitrogen-controlled expression of genes involved in xanthine salvage and catabolism." *J Bacteriol* 179(8): 2540-2550.

Cote, F. and J. P. Perreault (1997). "Peach latent mosaic viroid is locked by a 2',5'-phosphodiester bond produced by in vitro self-ligation." *J Mol Biol* 273(3): 533-543.

Crick, F. (1970). "Central dogma of molecular biology." *Nature* 227(5258): 561-563.

Croft, M. T., M. Moulin, M. E. Webb and A. G. Smith (2007). "Thiamine biosynthesis in algae is regulated by riboswitches." *Proc Natl Acad Sci U S A* 104(52): 20770-20775.

Dalgarno, P. A., J. Bordello, R. Morris, P. St-Pierre, A. Dube, I. D. Samuel, D. A. Lafontaine and J. C. Penedo (2013). "Single-molecule chemical denaturation of riboswitches." *Nucleic Acids Res* 41(7): 4253-4265.

Davis, J. H. and J. W. Szostak (2002). "Isolation of high-affinity GTP aptamers from partially structured RNA libraries." *Proc Natl Acad Sci U S A* 99(18): 11616-11621.

Deigan, K. E. and A. R. Ferre-D'Amare (2011). "Riboswitches: discovery of drugs that target bacterial gene-regulatory RNAs." *Acc Chem Res* 44(12): 1329-1338.

Delfosse, V., P. Bouchard, E. Bonneau, P. Dagenais, J. F. Lemay, D. A. Lafontaine and P. Legault (2010). "Riboswitch structure: an internal residue mimicking the purine ligand." *Nucleic Acids Res* 38(6): 2057-2068.

Drory Retwitzer, M., I. Kifer, S. Sengupta, Z. Yakhini and D. Barash (2015). "An Efficient Minimum Free Energy Structure-Based Search Method for Riboswitch Identification Based on Inverse RNA Folding." *PLoS One* 10(7): e0134262.

Eichhorn, C. D., M. Kang and J. Feigon (2014). "Structure and function of preQ riboswitches." *Biochim Biophys Acta* 1839(10): 939-950.

Fire, A., S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver and C. C. Mello (1998). "Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*." *Nature* 391(6669): 806-811.

Galagan, J. E., S. E. Calvo, C. Cuomo, L. J. Ma, J. R. Wortman, S. Batzoglou, S. I. Lee, M. Basturkmen, C. C. Spevak, J. Clutterbuck, V. Kapitonov, J. Jurka, C. Scazzocchio, M. Farman, J. Butler, S. Purcell, S. Harris, G. H. Braus, O. Draht, S. Busch, C. D'Enfert, C. Bouchier, G. H. Goldman, D. Bell-Pedersen, S. Griffiths-Jones, J. H. Doonan, J. Yu, K. Vienken, A. Pain, M. Freitag, E. U. Selker, D. B. Archer, M. A. Penalva, B. R. Oakley, M. Momany, T. Tanaka, T. Kumagai, K. Asai, M. Machida, W. C. Nierman, D. W. Denning, M. Caddick, M. Hynes, M. Paoletti, R. Fischer, B. Miller, P. Dyer, M. S. Sachs, S. A. Osmani and B. W. Birren (2005). "Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*." *Nature* 438(7071): 1105-1115.

Gelfand, M. S., A. A. Mironov, J. Jomantas, Y. I. Kozlov and D. A. Perumov (1999). "A conserved RNA structure element involved in the regulation of bacterial riboflavin synthesis genes." *Trends Genet* 15(11): 439-442.

Gilbert, S. D., C. D. Stoddard, S. J. Wise and R. T. Batey (2006). "Thermodynamic and kinetic characterization of ligand binding to the purine riboswitch aptamer domain." *J Mol Biol* 359(3): 754-768.

Gold, L., D. Brown, Y. He, T. Shtatland, B. S. Singer and Y. Wu (1997). "From oligonucleotide shapes to genomic SELEX: novel biological regulatory loops." *Proc Natl Acad Sci U S A* 94(1): 59-64.

Gottesman, S. (2004). "The small RNA regulators of *Escherichia coli*: roles and mechanisms." *Annu Rev Microbiol* 58: 303-328.

Greenleaf, W. J., K. L. Frieda, D. A. Foster, M. T. Woodside and S. M. Block (2008). "Direct observation of hierarchical folding in single riboswitch aptamers." *Science* 319(5863): 630-633.

Guerrier-Takada, C., K. Gardiner, T. Marsh, N. Pace and S. Altman (1983). "The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme." *Cell* 35(3 Pt 2): 849-857.

Ha, T., T. Enderle, D. F. Ogletree, D. S. Chemla, P. R. Selvin and S. Weiss (1996). "Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor." *Proc Natl Acad Sci U S A* 93(13): 6264-6268.

Haller, A., U. Rieder, M. Aigner, S. C. Blanchard and R. Micura (2011). "Conformational capture of the SAM-II riboswitch." *Nat Chem Biol* 7(6): 393-400.

Haller, A., M. F. Souliere and R. Micura (2011). "The dynamic nature of RNA as key to understanding riboswitch mechanisms." *Acc Chem Res* 44(12): 1339-1348.

Hammann, C., A. Luptak, J. Perreault and M. de la Pena (2012). "The ubiquitous hammerhead ribozyme." *RNA* 18(5): 871-885.

Henkin, T. M. (2008). "Riboswitch RNAs: using RNA to sense cellular metabolism." *Genes Dev* 22(24): 3383-3390.

Heppell, B., J. Mulhbachter, J. C. Penedo and D. A. Lafontaine (2009). "Application of fluorescent measurements for characterization of riboswitch-ligand interactions." *Methods Mol Biol* 540: 25-37.

Hershey, A. D. and M. Chase (1952). "Independent functions of viral protein and nucleic acid in growth of bacteriophage." *J Gen Physiol* 36(1): 39-56.

Hollands, K., S. Proshkin, S. Sklyarova, V. Epshtein, A. Mironov, E. Nudler and E. A. Groisman (2012). "Riboswitch control of Rho-dependent transcription termination." *Proc Natl Acad Sci U S A* 109(14): 5376-5381.

Hollands, K., A. Sevostyanova and E. A. Groisman (2014). "Unusually long-lived pause required for regulation of a Rho-dependent transcription terminator." *Proc Natl Acad Sci U S A* 111(19): E1999-2007.

Holley, R. W., J. Apgar, G. A. Everett, J. T. Madison, M. Marquisee, S. H. Merrill, J. R. Penswick and A. Zamir (1965). "STRUCTURE OF A RIBONUCLEIC ACID." *Science* 147(3664): 1462-1465.

Howe, J. A., H. Wang, T. O. Fischmann, C. J. Balibar, L. Xiao, A. M. Galgoci, J. C. Malinverni, T. Mayhood, A. Villafania, A. Nahvi, N. Murgolo, C. M. Barbieri, P. A. Mann, D. Carr, E. Xia, P. Zuck, D. Riley, R. E. Painter, S. S. Walker, B. Sherborne, R. de Jesus, W. Pan, M. A. Plotkin, J. Wu, D. Rindgen, J. Cummings, C. G. Garlisi, R. Zhang, P. R. Sheth, C. J. Gill, H. Tang and T. Roemer (2015). "Selective small-molecule inhibition of an RNA structural element." *Nature* 526(7575): 672-677.

Huang, R. C., N. Maheshwari and J. Bonner (1960). "Enzymatic synthesis of RNA." *Biochem Biophys Res Commun* 3: 689-694.

Hurwitz, J. and A. E. Bresler (1961). "The incorporation of ribonucleotides into ribonucleic acid." *J Biol Chem* 236: 542-548.

Jenison, R. D., S. C. Gill, A. Pardi and B. Polisky (1994). "High-resolution molecular discrimination by RNA." *Science* 263(5152): 1425-1429.

Jones, C. P. and A. R. Ferre-D'Amare (2015). "Recognition of the bacterial alarmone ZMP through long-distance association of two RNA subdomains." *Nat Struct Mol Biol* 22(9): 679-685.

Kiga, D., Y. Futamura, K. Sakamoto and S. Yokoyama (1998). "An RNA aptamer to the xanthine/guanine base with a distinctive mode of purine recognition." *Nucleic Acids Res* 26(7): 1755-1760.

Kim, J. N., K. F. Blount, I. Puskarz, J. Lim, K. H. Link and R. R. Breaker (2009). "Design and antimicrobial action of purine analogues that bind Guanine riboswitches." *ACS Chem Biol* 4(11): 915-927.

Kim, J. N. and R. R. Breaker (2008). "Purine sensing by riboswitches." *Biol Cell* 100(1): 1-11.

Kim, J. N., A. Roth and R. R. Breaker (2007). "Guanine riboswitch variants from *Mesoplasma florum* selectively recognize 2'-deoxyguanosine." *Proc Natl Acad Sci U S A* 104(41): 16092-16097.

Kim, P. B., J. W. Nelson and R. R. Breaker (2015). "An ancient riboswitch class in bacteria regulates purine biosynthesis and one-carbon metabolism." *Mol Cell* 57(2): 317-328.

Kim, S. H., F. L. Suddath, G. J. Quigley, A. McPherson, J. L. Sussman, A. H. Wang, N. C. Seeman and A. Rich (1974). "Three-dimensional tertiary structure of yeast phenylalanine transfer RNA." *Science* 185(4149): 435-440.

Klein, D. J., T. E. Edwards and A. R. Ferre-D'Amare (2009). "Cocrystal structure of a class I preQ1 riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase." *Nat Struct Mol Biol* 16(3): 343-344.

Kloc, M., V. Foreman and S. A. Reddy (2011). "Binary function of mRNA." *Biochimie* 93(11): 1955-1961.

Koizumi, M. and R. R. Breaker (2000). "Molecular recognition of cAMP by an RNA aptamer." *Biochemistry* 39(30): 8983-8992.

Kubodera, T., M. Watanabe, K. Yoshiuchi, N. Yamashita, A. Nishimura, S. Nakai, K. Gomi and H. Hanamoto (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(3): 516-520.

Kulshina, N., N. J. Baird and A. R. Ferre-D'Amare (2009). "Recognition of the bacterial second messenger cyclic diguanylate by its cognate riboswitch." *Nat Struct Mol Biol* 16(12): 1212-1217.

Lee, E. R., K. F. Blount and R. R. Breaker (2009). "Roseoflavin is a natural antibacterial compound that binds to FMN riboswitches and regulates gene expression." *RNA Biol* 6(2): 187-194.

Lemay, J. F., G. Desnoyers, S. Blouin, B. Heppell, L. Bastet, P. St-Pierre, E. Masse and D. A. Lafontaine (2011). "Comparative study between transcriptionally- and translationally-acting adenine riboswitches reveals key differences in riboswitch regulatory mechanisms." *PLoS Genet* 7(1): e1001278.

Lemay, J. F. and D. A. Lafontaine (2007). "Core requirements of the adenine riboswitch aptamer for ligand binding." *RNA* 13(3): 339-350.

Lemay, J. F., J. C. Penedo, R. Tremblay, D. M. Lilley and D. A. Lafontaine (2006). "Folding of the adenine riboswitch." *Chem Biol* 13(8): 857-868.

Leontis, N. B., J. Stombaugh and E. Westhof (2002). "The non-Watson-Crick base pairs and their associated isostericity matrices." *Nucleic Acids Res* 30(16): 3497-3531.

Lescoute, A. and E. Westhof (2006). "Topology of three-way junctions in folded RNAs." *RNA* 12(1): 83-93.

Loh, E., O. Dussurget, J. Gripenland, K. Vaitkevicius, T. Tiensuu, P. Mandin, F. Repoila, C. Buchrieser, P. Cossart and J. Johansson (2009). "A trans-acting riboswitch controls expression of the virulence regulator PrfA in *Listeria monocytogenes*." *Cell* 139(4): 770-779.

Long, Q., L. Ji, H. Wang and J. Xie (2010). "Riboflavin biosynthetic and regulatory factors as potential novel anti-infective drug targets." *Chem Biol Drug Des* 75(4): 339-347.

Lunse, C. E., A. Schuller and G. Mayer (2014). "The promise of riboswitches as potential antibacterial drug targets." *Int J Med Microbiol* 304(1): 79-92.

Mandal, M., B. Boese, J. E. Barrick, W. C. Winkler and R. R. Breaker (2003). "Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria." *Cell* 113(5): 577-586.

Mandal, M. and R. R. Breaker (2004). "Adenine riboswitches and gene activation by disruption of a transcription terminator." *Nat Struct Mol Biol* 11(1): 29-35.

Mandal, M., M. Lee, J. E. Barrick, Z. Weinberg, G. M. Emilsson, W. L. Ruzzo and R. R. Breaker (2004). "A glycine-dependent riboswitch that uses cooperative binding to control gene expression." *Science* 306(5694): 275-279.

Marraffini, L. A. (2015). "CRISPR-Cas immunity in prokaryotes." *Nature* 526(7571): 55-61.

Martin, L. L., P. J. Unrau and U. F. Muller (2015). "RNA synthesis by in vitro selected ribozymes for recreating an RNA world." *Life (Basel)* 5(1): 247-268.

Mellin, J. R. and P. Cossart (2015). "Unexpected versatility in bacterial riboswitches." *Trends Genet* 31(3): 150-156.

Meyer, M. M., A. Roth, S. M. Chervin, G. A. Garcia and R. R. Breaker (2008). "Confirmation of a second natural preQ1 aptamer class in Streptococcaceae bacteria." *RNA* 14(4): 685-695.

Miranda-Rios, J., M. Navarro and M. Soberon (2001). "A conserved RNA structure (thi box) is involved in regulation of thiamin biosynthetic gene expression in bacteria." *Proc Natl Acad Sci U S A* 98(17): 9736-9741.

Mironov, A. S., I. Gusarov, R. Rafikov, L. E. Lopez, K. Shatalin, R. A. Kreneva, D. A. Perumov and E. Nudler (2002). "Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria." *Cell* 111(5): 747-756.

Mitra, K., C. Schaffitzel, T. Shaikh, F. Tama, S. Jenni, C. L. Brooks, 3rd, N. Ban and J. Frank (2005). "Structure of the E. coli protein-conducting channel bound to a translating ribosome." *Nature* 438(7066): 318-324.

Mulhbacher, J., E. Brouillette, M. Allard, L.-C. Fortier, F. Malouin and D. A. Lafontaine (2010). "Novel Riboswitch Ligand Analogs as Selective Inhibitors of Guanine-Related Metabolic Pathways." *PLoS Pathog.* 6(4):e1000865

Mulhbacher, J. and D. A. Lafontaine (2007). "Ligand recognition determinants of guanine riboswitches." *Nucleic Acids Res* 35(16): 5568-5580.

Nahvi, A., J. E. Barrick and R. R. Breaker (2004). "Coenzyme B12 riboswitches are widespread genetic control elements in prokaryotes." *Nucleic Acids Res* 32(1): 143-150.

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

Navarro, B. and R. Flores (1997). "Chrysanthemum chlorotic mottle viroid: unusual structural properties of a subgroup of self-cleaving viroids with hammerhead ribozymes." *Proc Natl Acad Sci U S A* 94(21): 11262-11267.

Neupane, K., H. Yu, D. A. Foster, F. Wang and M. T. Woodside (2011). "Single-molecule force spectroscopy of the add adenine riboswitch relates folding to regulatory mechanism." *Nucleic Acids Res* 39(17): 7677-7687.

Nguyen, D. H., S. C. DeFina, W. H. Fink and T. Dieckmann (2002). "Binding to an RNA aptamer changes the charge distribution and conformation of malachite green." *J Am Chem Soc* 124(50): 15081-15084.

Nirenberg, M. W. and J. H. Matthaei (1961). "The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides." *Proc Natl Acad Sci U S A* 47: 1588-1602.

- Noller, H. F., V. Hoffarth and L. Zimniak (1992). "Unusual resistance of peptidyl transferase to protein extraction procedures." *Science* 256(5062): 1416-1419.
- Palade, G. E. (1955). "Studies on the endoplasmic reticulum. II. Simple dispositions in cells in situ." *J Biophys Biochem Cytol* 1(6): 567-582.
- Papasaikas, P. and J. Valcarcel (2015). "The Spliceosome: The Ultimate RNA Chaperone and Sculptor." *Trends Biochem Sci.* 41(1):33-45
- Peselis, A. and A. Serganov (2014). "Themes and variations in riboswitch structure and function." *Biochim Biophys Acta* 1839(10): 908-918.
- Porter, E. B., J. G. Marcano-Velazquez and R. T. Batey (2014). "The purine riboswitch as a model system for exploring RNA biology and chemistry." *Biochim Biophys Acta* 1839(10): 919-930.
- Ramakrishnan, V. (2014). "The ribosome emerges from a black box." *Cell* 159(5): 979-984.
- Ravnum, S. and D. I. Andersson (2001). "An adenosyl-cobalamin (coenzyme-B12)-repressed translational enhancer in the cob mRNA of *Salmonella typhimurium*." *Mol Microbiol* 39(6): 1585-1594.
- Repoila, F. and F. Darfeuille (2009). "Small regulatory non-coding RNAs in bacteria: physiology and mechanistic aspects." *Biol Cell* 101(2): 117-131.
- Robertus, J. D., J. E. Ladner, J. T. Finch, D. Rhodes, R. S. Brown, B. F. Clark and A. Klug (1974). "Structure of yeast phenylalanine tRNA at 3 Å resolution." *Nature* 250(467): 546-551.
- Roth, A., 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 and R. R. Breaker (2007). "A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain." *Nat Struct Mol Biol* 14(4): 308-317.
- Roy, R., S. Hohng and T. Ha (2008). "A practical guide to single-molecule FRET." *Nat Methods* 5(6): 507-516.
- Sassanfar, M. and J. W. Szostak (1993). "An RNA motif that binds ATP." *Nature* 364(6437): 550-553.

Schuwirth, B. S., M. A. Borovinskaya, C. W. Hau, W. Zhang, A. Vila-Sanjurjo, J. M. Holton and J. H. Cate (2005). "Structures of the bacterial ribosome at 3.5 Å resolution." *Science* 310(5749): 827-834.

Serganov, A. and D. J. Patel (2009). "Amino acid recognition and gene regulation by riboswitches." *Biochim Biophys Acta* 1789(9-10): 592-611.

Serganov, A., A. Polonskaia, A. T. Phan, R. R. Breaker and D. J. Patel (2006). "Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch." *Nature* 441(7097): 1167-1171.

Serganov, A., Y. R. Yuan, O. Pikovskaya, A. Polonskaia, L. Malinina, A. T. Phan, C. Hobartner, R. Micura, R. R. Breaker and D. J. Patel (2004). "Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs." *Chem Biol* 11(12): 1729-1741.

Sharmeen, L., M. Y. Kuo, G. Dinter-Gottlieb and J. Taylor (1988). "Antigenomic RNA of human hepatitis delta virus can undergo self-cleavage." *J Virol* 62(8): 2674-2679.

Shaw, E., P. St-Pierre, K. McCluskey, D. A. Lafontaine and J. C. Penedo (2014). "Using sm-FRET and denaturants to reveal folding landscapes." *Methods Enzymol* 549: 313-341.

Smith-Peter, E., A. M. Lamontagne and D. A. Lafontaine (2015). "Role of lysine binding residues in the global folding of the lysC riboswitch." *RNA Biol.* 12(12):1372-82

Sohn, Y., H. Shen and C. Kang (2003). "Stepwise walking and cross-linking of RNA with elongating T7 RNA polymerase." *Methods Enzymol* 371: 170-179.

Soukup, G. A. and R. R. Breaker (1999). "Relationship between internucleotide linkage geometry and the stability of RNA." *RNA* 5(10): 1308-1325.

St-Pierre, P., K. McCluskey, E. Shaw, J. C. Penedo and D. A. Lafontaine (2014). "Fluorescence tools to investigate riboswitch structural dynamics." *Biochim Biophys Acta* 1839(10): 1005-1019.

Stevens, A. (1960). "Incorporation of the adenine ribonucleotide into RNA by cell fractions from *E. coli* B." *Biochemical and Biophysical Research Communications* 3(1): 92-96.

Stormo, G. D. and Y. Ji (2001). "Do mRNAs act as direct sensors of small molecules to control their expression?" *Proc Natl Acad Sci U S A* 98(17): 9465-9467.

Su, D., C. T. Chan, C. Gu, K. S. Lim, Y. H. Chionh, M. E. McBee, B. S. Russell, I. R. Babu, T. J. Begley and P. C. Dedon (2014). "Quantitative analysis of ribonucleoside modifications in tRNA by HPLC-coupled mass spectrometry." *Nat Protoc* 9(4): 828-841.

Sudarsan, N., J. E. Barrick and R. R. Breaker (2003). "Metabolite-binding RNA domains are present in the genes of eukaryotes." *RNA* 9(6): 644-647.

Sudarsan, N., M. C. Hammond, K. F. Block, R. Welz, J. E. Barrick, A. Roth and R. R. Breaker (2006). "Tandem riboswitch architectures exhibit complex gene control functions." *Science* 314(5797): 300-304.

Sudarsan, N., E. R. Lee, Z. Weinberg, R. H. Moy, J. N. Kim, K. H. Link and R. R. Breaker (2008). "Riboswitches in eubacteria sense the second messenger cyclic di-GMP." *Science* 321(5887): 411-413.

Svoboda, P. (2014). "Renaissance of mammalian endogenous RNAi." *FEBS Lett* 588(15): 2550-2556.

Sweeney, B. A., P. Roy and N. B. Leontis (2015). "An introduction to recurrent nucleotide interactions in RNA." *Wiley Interdiscip Rev RNA* 6(1): 17-45.

Talini, G., S. Branciamore and E. Gallori (2011). "Ribozymes: Flexible molecular devices at work." *Biochimie* 93(11): 1998-2005.

Temin, H. M. and S. Mizutani (1970). "RNA-dependent DNA polymerase in virions of Rous sarcoma virus." *Nature* 226(5252): 1211-1213.

Toledo-Arana, A., O. Dussurget, G. Nikitas, N. Sesto, H. Guet-Revillet, D. Balestrino, E. Loh, J. Gripenland, T. Tiensuu, K. Vaitkevicius, M. Barthelemy, M. Vergassola, M. A. Nahori, G. Soubigou, B. Regnault, J. Y. Coppee, M. Lecuit, J. Johansson and P. Cossart (2009). "The *Listeria* transcriptional landscape from saprophytism to virulence." *Nature* 459(7249): 950-956.

Trausch, J. J., Z. Xu, A. L. Edwards, F. E. Reyes, P. E. Ross, R. Knight and R. T. Batey (2014). "Structural basis for diversity in the SAM clan of riboswitches." *Proc Natl Acad Sci U S A* 111(18): 6624-6629.

Tremblay, R., J. F. Lemay, S. Blouin, J. Mulhbach, E. Bonneau, P. Legault, P. Dupont, J. C. Penedo and D. A. Lafontaine (2011). "Constitutive regulatory activity of an evolutionarily excluded riboswitch variant." *J Biol Chem* 286(31): 27406-27415.

Voet, D. and J. G. Voet. Biochemistry 4th edition, Wiley, Kendallville. (2004). pp. 1248

Wachter, A., M. Tunc-Ozdemir, B. C. Grove, P. J. Green, D. K. Shintani and R. R. Breaker (2007). "Riboswitch control of gene expression in plants by splicing and alternative 3' end processing of mRNAs." *Plant Cell* 19(11): 3437-3450.

Wang, J. X. and R. R. Breaker (2008). "Riboswitches that sense S-adenosylmethionine and S-adenosylhomocysteine." *Biochem Cell Biol* 86(2): 157-168.

Watson, J. D. and F. H. Crick (1953). "Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid." *Nature* 171(4356): 737-738.

Weinberg, Z., 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 and R. R. Breaker (2007). "Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline." *Nucleic Acids Res* 35(14): 4809-4819.

Weinberg, Z., J. X. Wang, J. Bogue, J. Yang, K. Corbino, R. H. Moy and R. R. Breaker (2010). "Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes." *Genome Biol* 11(3): R31.

Welz, R. and R. R. Breaker (2007). "Ligand binding and gene control characteristics of tandem riboswitches in *Bacillus anthracis*." *RNA* 13(4): 573-582.

Wickiser, J. K., W. C. Winkler, R. R. Breaker and D. M. Crothers (2005). "The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch." *Mol Cell* 18(1): 49-60.

Winkler, W., A. Nahvi and R. R. Breaker (2002). "Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression." *Nature* 419(6910): 952-956.

Winkler, W. C. and R. R. Breaker (2005). "Regulation of bacterial gene expression by riboswitches." *Annu Rev Microbiol* 59: 487-517.

Winkler, W. C., S. Cohen-Chalamish and R. R. Breaker (2002). "An mRNA structure that controls gene expression by binding FMN." *Proc Natl Acad Sci U S A* 99(25): 15908-15913.

Winkler, W. C., A. Nahvi, A. Roth, J. A. Collins and R. R. Breaker (2004). "Control of gene expression by a natural metabolite-responsive ribozyme." *Nature* 428(6980): 281-286.

Wittenhagen, L. M. and S. O. Kelley (2003). "Impact of disease-related mitochondrial mutations on tRNA structure and function." *Trends Biochem Sci* 28(11): 605-611.

Woodson, S. A. (2010). "Compact intermediates in RNA folding." *Annu Rev Biophys* 39: 61-77.

Zhou, J., L. Lancaster, J.P. Donohue and H.F. Noller (2013). "Crystal structures of EF-G-ribosome complexes trapped in intermediate states of translocation." *Science* 340(6140): 1236086

