

Centre Armand – Frappier Santé Biotechnologie

ARN non codants régulateurs en *cis* et *riboswitchs* : développement de méthodes et caractérisation exploratoire.

Par
Aurélie Devinck

Thèse présentée pour l'obtention du grade
de Philosophiae Doctor (Ph.D.) en Biologie

Jury d'évaluation

Président du jury et
examineur interne

Pr Charles Calmettes
INRS – Centre Armand Frappier Santé
Biotechnologie

Examineur externe

Pr Benoit Cousineau
Mc Gill University – School of
Biomedical Sciences in Microbiology
and Immunology

Examineur externe

Pre Pascale Romby
CNRS – Institut de biologie moléculaire
et cellulaire. Strasbourg (France)

Directeur de recherche

Pr Jonathan Perreault
INRS – Centre Armand Frappier Santé
Biotechnologie

REMERCIEMENTS

La réalisation et la réussite de ce doctorat n'aurait pas été possible sans l'aide et le soutien de plusieurs personnes que je tenais à remercier dans cette partie.

Mes premiers merci vont à mon directeur de recherche le Pr Jonathan Perreault. Merci tout d'abord d'avoir eu confiance en moi en 2016 et de m'avoir confié ce beau projet que j'ai tant aimé et on peut le dire tant détester parfois. Et oui, c'est ça aussi d'optimiser des méthodes. Merci pour tous tes conseils, ton soutien et ton écoute durant ces six années de maîtrise et doctorat.

J'aimerais également remercier les membres de mon jury. Tout d'abord, le Pr Charles Calmettes qui m'a suivi dans tous mes examens de doctorat jusqu'à cette dernière étape de mon parcours. Ainsi que la Pre Pascale Romby et le Pr Benoit Cousineau d'avoir accepté d'évaluer mon travail effectué au cours de ces dernières années. Je tiens aussi à remercier le Pr Sébastien Faucher pour avoir été présent pour l'évaluation de mes autres examens de doctorat.

Je tenais également à remercier Marie-Christine Groleau et Charles Morin, du laboratoire du Pr. Éric Déziel pour leur aide avec mes questions sur *Pseudomonas* et aussi de m'avoir fourni les souches nécessaires à ce travail.

D'énormes merci également à ma collègue de toujours, Emilie. Nos parcours se sont suivis tout du long de nos maîtrises et doctorats respectifs. À deux, nous avons vécu beaucoup de choses comme des heures d'optimisation pour le SR-PAGE, d'écoute, de brainstorm et d'interrogation sans oublier nos différents congrès en Californie ou encore à Washington DC. Ton soutien a aussi été un grand pilier pour moi pour la durée de ce doctorat.

Je ne pouvais pas faire ces remerciements sans penser à Audrey-Anne, de gros gros merci Dédé pour tous ces mots de soutien et toute ton aide toujours avec un timing parfait dans les cellules de crises au laboratoire. On ne lâche pas !!!

Sans oublier les collègues et ami.e.s des autres laboratoires : Nicolas, Ève, Martin. Merci pour tous ces bons moments passés ensemble aussi bien à l'INRS qu'à l'extérieur. Merci Nico pour toutes ces sessions de brainstorm quand mes expériences ne fonctionnaient pas, et non je ne fais pas de sorcellerie. Merci également aux dernières arrivées Apolline, Marie, Eléonore, Mathilde et Adeline qui ne m'ont pas vu sous mon meilleur jour mais vos sourires, vos petits mots m'ont réchauffé le cœur.

J'ai également une grosse pensée pour quelqu'un qui j'espère est fier de moi aujourd'hui de là-haut : mon papa. Tu nous as quitté il y a plusieurs années mais je sentais ta force pour m'aider à travers ce projet de doctorat. Merci papa !

Je ne serai jamais arrivé si loin aussi sans le soutien indéfectible de ma maman et de joli papa, vous avez toujours cru en moi merci !! Merci également à mes beaux-parents, Danièle et Philippe, pour votre soutien.

Et enfin je ne peux pas finir ces remerciements sans te dire d'énormes merci à toi mon mari. Merci pour ton soutien de tous les jours, des bons résultats au mauvais tu as toujours été là pour m'aider à me relever dans les coups durs. Merci pour ton aide aussi pour la réalisation des scripts qui ont rendus certaines de mes tâches beaucoup plus rapides.

RÉSUMÉ

Au cours des dernières décennies de nouvelles classes d'ARN ont été découvertes. Celles-ci ne contiennent pas d'informations permettant la synthèse de protéines mais ont des fonctions régulatrices. De nombreuses études ont permis de montrer leur importance dans le maintien de l'homéostasie de la cellule, mais aussi pour leur adaptation à l'environnement. Ces sont les ARN non codants (ARNnc). Ils sont classés en différentes catégories à savoir les ARN en *trans*, qui sont transcrits de manière indépendante par rapport à leur cible, ou les ARN en *cis*, qui eux sont situés à proximité de leur cible voire sur le même ARN messager (ARNm). Ce sont ces derniers qui nous intéresseront dans le cadre de cette thèse de doctorat.

Certaines structures d'ARN complexes sont situées dans les parties 5' non traduites des ARNm et permettent de réguler l'expression du gène situé directement en aval. On peut en distinguer plusieurs catégories en fonction du stimulus permettant sa fonction, comme les thermorégulateurs, les T-box, les *r-leader* ou encore les *riboswitchs*. La plupart des ARNnc en *cis* ont été découverts grâce à des méthodes de bioinformatique basées sur des comparaisons de structures. Cependant, ces méthodes présentent des limites. Elles permettent de prédire de nombreuses structures qui ont la nécessité d'être validées expérimentalement. Ceci crée un déséquilibre dans la caractérisation de ces ARN entre les bactéries disposant de nombreux outils génétiques, comme *Escherichia coli* et *Bacillus subtilis* et celles en ayant moins comme *Pseudomonas aeruginosa* et *Burkholderia thailandensis*. Nous avons ainsi développé au laboratoire, un plasmide nommé pVK-f2.2-lux permettant de faire des fusions traductionnelles entre des ARNnc prédits et un opéron permettant de produire de la luminescence. Ceci nous a permis de caractériser l'importance de la structure de six ARNnc prédits chez *Pseudomonas* pour permettre la régulation du gène situé en aval de ceux-ci. Afin de pallier à une autre limite des outils de bioinformatique, nous avons développé une méthode expérimentale pour la découverte de nouveaux *riboswitchs*. Cette méthode est basée sur la caractéristique principale des *riboswitchs* à changer de conformation en présence de leur ligand, provoquant un changement de migration dans un gel de polyacrylamide natif. Cette méthode a été validée grâce à des *riboswitchs* connus et nous a permis de sélectionner un *riboswitch* avec une affinité modifiée pour son ligand grâce à un SELEX (Systematic Evolution of Ligand by EXponential enrichment).

Mots-clés : *rmf*, *sucA-II*, *Pseudomon-Rho*, *rne-II*, *fixA*, *Pseudomon-GGDEF*, ppGpp, di-GMPc, électrophorèse, TPP.

ABSTRACT

During the last decades new classes of RNA have been discovered. These do not contain information for protein synthesis but have regulatory functions. Numerous studies have shown their importance in the maintenance of cell homeostasis, but also for their adaptation to the environment. These are the non-coding RNAs (ncRNAs). They are classified in different categories, namely RNAs in trans, which are transcribed independently from their target, or RNAs in cis, which are located near their target or on the same messenger RNA (mRNA). These are the ones that will interest us in this PhD thesis.

Some complex RNA structures are located in the 5' untranslated parts of the mRNAs and allow to regulate the expression of the gene located directly downstream. Several categories can be distinguished depending on the stimulus allowing its function, such as thermoregulators, T-boxes, r-leaders or riboswitches. Most of the cis-regulatory RNAs discovered have been discovered using bioinformatics tools based on structure comparisons. However, these methods have some limitations. They predict many structures that need to be validated experimentally. This creates an imbalance in the characterization of these RNAs between bacteria with many genetic tools, such as *Escherichia coli* and *Bacillus subtilis*, and those with fewer such as *Pseudomonas aeruginosa* and *Burkholderia thailandensis*. We have thus developed in the laboratory a plasmid named pVK-f2.2-lux allowing translational fusions between predicted ncRNAs and an operon that allows luminescence production. This allowed us to characterize the importance of the structure of six predicted ncRNAs in *Pseudomonas* to allow regulation of the gene located downstream of them. In order to overcome another limitation of bioinformatics tools, we have developed a new experimental method for the discovery of new riboswitches. This method is based on the main characteristic of riboswitches to change their conformation in the presence of their ligand causing a migration change in a native polyacrylamide gel. This method was validated with known riboswitches and we were able to select a riboswitch with a modified affinity for its ligand using SR-PAGE with a SELEX (Systematic Evolution of Ligand by EXponential enrichment).

Keywords : *rmf*, *sucA-II*, *Pseudomon-Rho*, *rne-II*, *fixA*, *Pseudomon-GGDEF*, ppGpp, di-GMPc, TPP, electrophoresis.

TABLE DES MATIÈRES

REMERCIEMENTS.....	III
RÉSUMÉ.....	V
ABSTRACT.....	VII
TABLE DES MATIÈRES	IX
LISTE DES FIGURES.....	XIII
LISTE DES TABLEAUX	XVII
LISTE DES ABRÉVIATIONS	XIX
1 INTRODUCTION	1
1.1 LA RÉGULATION CHEZ LES PROCARYOTES.....	1
1.1.1 <i>Les facteurs de transcription</i>	1
1.1.2 <i>Les systèmes à deux composants</i>	2
1.1.3 <i>Les ARN non codants régulateurs</i>	3
1.2 LES ARN NON CODANTS RÉGULANT EN CIS.....	5
1.2.1 <i>Les T-box</i>	5
1.2.2 <i>Les ARN thermorégulateurs</i>	9
1.2.3 <i>Ribosomal RNA Leader (r-leader)</i>	10
1.2.4 <i>Les riboswitchs</i>	12
1.3 LES RIBOSWITCHS CHEZ LES PROCARYOTES	12
1.3.1 <i>Découverte des riboswitchs</i>	14
1.3.2 <i>Régulation de la transcription</i>	15
1.3.3 <i>Régulation de la traduction</i>	18
1.3.4 <i>Régulation au niveau de la stabilité de l'ARNm</i>	19
1.3.5 <i>Différents types de ligands</i>	20
1.4 MÉTHODES DE DÉCOUVERTE DE NOUVEAUX ARN NON CODANTS AGISSANT EN CIS	21
1.4.1 <i>Méthodes bioinformatique</i>	21
1.4.2 <i>Méthode basée sur la structure des ARN</i>	23
1.4.3 <i>Term-seq</i>	24
1.4.4 <i>La méthode PARCEL</i>	25
1.4.5 <i>SHAPE-MaP</i>	26
1.5 DISTRIBUTION DES ARN NON CODANTS RÉGULANT EN CIS.....	28
2 PROBLÉMATIQUE, HYPOTHÈSE ET OBJECTIFS.....	31
3 CARACTÉRISATION DE NOUVEAUX ARN NON CODANTS EN CIS CHEZ LES BACTÉRIES DU GENRE PSEUDOMONAS ET BURKHOLDERIA.....	33
3.1 MISE EN CONTEXTE	33
3.1.1 <i>Organismes d'études</i>	33

3.1.2	<i>ARN non codants chez nos organismes d'études</i>	36
3.1.3	<i>Problématique</i>	42
3.2	MATÉRIEL ET MÉTHODES	45
3.2.1	<i>Design des ARNnc candidats</i>	45
3.2.2	<i>Plasmides</i>	48
3.2.3	<i>Clonage</i>	49
3.2.4	<i>Conjugaison</i>	50
3.2.5	<i>Milieux de culture utilisés</i>	51
3.2.6	<i>Essai luminescence</i>	53
3.2.7	<i>Cartographie in vitro par in-line</i>	54
3.3	PRÉSENTATION DU RAPPORTEUR D'EXPRESSION	59
3.3.1	<i>Plasmide utilisé pVK-f2-lux</i>	59
3.3.2	<i>Preuve de concept avec le riboswitch fluor de B. thailandensis</i>	60
3.3.3	<i>Design des ARN non codants candidats</i>	62
3.4	OPTIMISATION DU VECTEUR D'EXPRESSION	64
3.5	RÉSULTATS ET DISCUSSION	66
3.5.1	<i>Motif rmf</i>	66
3.5.2	<i>Motif Pseudomon-GGDEF</i>	82
3.5.3	<i>Candidats avec un rôle de régulation confirmé</i>	91
3.5.4	<i>Candidats dont le rôle de régulation reste à confirmer</i>	97
3.5.5	<i>Conclusion</i>	105
4	DÉVELOPPEMENT D'UNE NOUVELLE MÉTHODE EXPÉRIMENTALE POUR LA DÉCOUVERTE DE RIBOSWITCHS	108
4.1	MISE EN CONTEXTE ET PROBLÉMATIQUE	108
4.2	MATÉRIEL ET MÉTHODES	111
4.2.1	<i>Constructions des riboswitchs</i>	111
4.2.2	<i>Transcription in vitro</i>	115
4.2.3	<i>Préparation du gel de SR-PAGE</i>	115
4.2.4	<i>Préparation des échantillons</i>	116
4.2.5	<i>Migration inverse</i>	117
4.2.6	<i>Sélection grâce au SELEX de riboswitchs TPP avec une affinité modifiée</i>	119
4.2.7	<i>Cartographie in vitro par in-line et constante de dissociation</i>	119
4.2.8	<i>Calcul des énergies libres et structure secondaire des ARN</i>	121
4.2.9	<i>Tests adaptateurs dans le cadre d'un SELEX</i>	121
4.3	RÉSULTATS	124
4.3.1	<i>Validation de la méthode</i>	124
4.3.2	<i>Un outil d'étude pour la structure et la plateforme d'expression</i>	127

4.3.3	<i>Sélection de riboswitchs avec une affinité modifiée</i>	130
4.3.4	<i>Influence des adaptateurs sur le décalage des ARN</i>	133
4.4	DISCUSSION	135
5	DISCUSSION GÉNÉRALE ET PERSPECTIVES	138
5.1	CARACTÉRISATION D'ARN NON CODANTS CHEZ <i>PSEUDOMONAS</i> ET <i>BURKHOLDERIA</i>	138
5.1.1	<i>Le motif rmf</i>	141
5.1.2	<i>Le motif Pseudomon-GGDEF</i>	146
5.1.3	<i>Le motif Pseudomon-Rho</i>	148
5.1.4	<i>Le motif sucA-II</i>	149
5.1.5	<i>Le motif fixA</i>	150
5.1.6	<i>Le motif rne-II</i>	151
5.1.7	<i>Perspectives</i>	153
5.2	DÉVELOPPEMENT D'UNE NOUVELLE MÉTHODE EXPÉRIMENTALE POUR LA DÉCOUVERTE DE <i>RIBOSWITCHS : LE SHIFTED REVERSE-PAGE</i>	157
5.2.1	<i>Discussion SR-PAGE</i>	158
5.2.2	<i>Perspectives du SR-PAGE</i>	160
5.3	CONCLUSION.....	168
6	BIBLIOGRAPHIE	169
7	ANNEXES	187
7.1	ANNEXE 1 : SÉQUENCES DES ARNnc CANDIDATS	187
7.2	ANNEXE 2 : MOTIF <i>RMF</i>	189
7.3	ANNEXE 3 : CARTOGRAPHIE <i>IN VITRO</i> PAR <i>IN-LINE</i> DE <i>RMF</i>	190
7.4	ANNEXE 4 : COURBE DE K_D POUR LES DIFFÉRENTS ESSAIS DE LA CONSTRUCTION <i>RMF 4</i>	191
7.5	ANNEXE 5 : <i>PSEUDOMON-GGDEF</i>	192
7.6	ANNEXE 6 : CARTOGRAPHIE <i>IN VITRO</i> PAR <i>IN-LINE</i> DU MOTIF <i>PSEUDOMON-GGDEF</i>	193
7.7	ANNEXE 7 : CARTOGRAPHIE <i>IN VITRO</i> PAR <i>IN-LINE</i> <i>PSEUDOMON-GGDEF</i> ET <i>RMF</i>	194
7.8	ANNEXE 8 : <i>PSEUDOMON-RHO</i>	195
7.9	ANNEXE 9 : <i>SUCA-II</i>	196
7.10	ANNEXE 10 : <i>FIXA</i>	197
7.11	ANNEXE 11 : <i>RNE-II</i>	198
7.12	ANNEXE 12 : ÉNERGIE LIBRE <i>RIBOSWITCH</i> FLUOR.....	199
7.13	ANNEXE 13 : ÉNERGIE LIBRE <i>RIBOSWITCH</i> FMN.....	200
7.14	ANNEXE 14 : ÉNERGIE LIBRE <i>RIBOSWITCH</i> DI-GMPC	201
7.15	ANNEXE 15 : ÉNERGIE LIBRE <i>RIBOSWITCH</i> NICKEL/COBALT.....	202
7.16	ANNEXE 16 : CLONES DU <i>RIBOSWITCH</i> TPP	203
7.17	ANNEXE 17 : STRUCTURE LIBRAIRIE 3 DU <i>RIBOSWITCH</i> TPP.....	208
7.18	ANNEXE 18 : PRÉDICTION DE STRUCTURE LIBRAIRIE 3	209

7.19	ANNEXE 19 : ARTEFACT DUE À LA CONTAMINATION LUMINEUSE.....	214
7.20	ANNEXE 20 : ARTICLE DU SR-PAGE PUBLIÉ DANS BIORXIV	215

LISTE DES FIGURES

FIGURE 1.1	SCHÉMA DU MODE DE FONCTIONNEMENT DES SYSTÈMES À DEUX COMPOSANTS.....	2
FIGURE 1.2	REPRÉSENTATION SCHÉMATIQUE D'UN ARNM.....	5
FIGURE 1.3	STRUCTURE CONSENSUS DES T-BOX.....	7
FIGURE 1.4	REPRÉSENTATION SCHÉMATIQUE DU FONCTIONNEMENT DES T-BOX.....	8
FIGURE 1.5	FONCTIONNEMENT DES ARN THERMORÉGULATEURS.....	9
FIGURE 1.6	REPRÉSENTATION SCHÉMATIQUE DU MODE GÉNÉRAL DE RÉGULATION PAR LES <i>RIBOSWITCHS</i>	13
FIGURE 1.7	REPRÉSENTATION SCHÉMATIQUE D'UN RIBOSWITCH AGISSANT SUR LA TRANSCRIPTION.....	14
FIGURE 1.8	EXEMPLE DE RÉGULATION DE LA TRANSCRIPTION PAR LE <i>RIBOSWITCH YDHL</i> CHEZ <i>B.SUBTILIS</i>	16
FIGURE 1.9	MÉCANISME DE RÉGULATION DE LA TRANSCRIPTION DÉPENDANT DE LA PROTÉINE RHO.....	17
FIGURE 1.10	MÉCANISME DE RÉGULATION DE LA TRADUCTION VIA UN EXEMPLE D'INHIBITION.....	19
FIGURE 1.11	MÉCANISME DE RÉGULATION DU <i>RIBOSWITCH LYSC</i> CHEZ <i>E. COLI</i>	20
FIGURE 1.12	PIPELINE BIOINFORMATIQUE POUR LA RECHERCHE DE NOUVEAUX ARN NON CODANTS.....	22
FIGURE 1.13	SCHÉMA DU PRINCIPE DE LA MÉTHODE PARCEL.....	26
FIGURE 1.14	SCHÉMA DU PRINCIPE DU SHAPE-MAP.....	27
FIGURE 1.15	DISTRIBUTION DES <i>R-LEADER</i>	29
FIGURE 3.1	SCHÉMA DES MÉTHODES DE CARACTÉRISATION D'UN NOUVEAU MOTIF D'ARN NON CODANT.....	44
FIGURE 3.2	SCHÉMA DE LA PCR D'ASSEMBLAGE.....	46
FIGURE 3.3	LE PLASMIDE PVK-F2-LUX.....	60
FIGURE 3.4	AUGMENTATION DE LA LUMINESCENCE PRODUITE VIA PVK-F2-LUX AVEC DES CONCENTRATIONS CROISSANTES DE FLUOR.....	61
FIGURE 3.5	SCHÉMA DU DESIGN UTILISÉ POUR LA CRÉATION DES PLASMIDES.....	63
FIGURE 3.6	REPRÉSENTATION SCHÉMATIQUE DES CLONAGES DANS LES PLASMIDES PVK.....	65
FIGURE 3.7	STRUCTURE DU MOTIF <i>RMFWT</i> CHEZ <i>P. AERUGINOSA</i> PA14.....	67
FIGURE 3.8	CONTEXTE TRANSCRIPTOMIQUE DU GÈNE <i>RMF</i> DANS <i>P. AERUGINOSA</i> PA14.....	68
FIGURE 3.9	LE NIVEAU DE LUMINESCENCE DU MOTIF <i>RMFWT</i> EST PLUS GRAND QUE POUR LE MOTIF AVEC LES MUTATIONS DISRUPTIVES.....	69

FIGURE 3.10	LA CONSTRUCTION COMPENSATOIRE RÉTABLIT LE NIVEAU DE LUMINESCENCE À UN NIVEAU SIMILAIRE À LA CONSTRUCTION WT.	70
FIGURE 3.11	LE PROFIL DU NIVEAU DE LUMINESCENCE EST MODIFIÉ DANS LA SOUCHE PA01 Δ REL Δ SPOT. 72	
FIGURE 3.12	L'EXPRESSION DE RMF DIMINUE EN PHASE STATIONNAIRE.....	74
FIGURE 3.13	PRINCIPE DE LA MÉTHODE DE CARTOGRAPHIE <i>IN VITRO</i> PAR <i>IN-LINE</i>	76
FIGURE 3.14	CONFIRMATION DE LA STRUCTURE DU MOTIF <i>RMF</i>	77
FIGURE 3.15	LES MUTATIONS DISRUPTIVES ENTRAINENT UN CHANGEMENT DE CONFORMATION DE L'ARN <i>RMF</i> . 79	
FIGURE 3.16	L'ARN <i>RMF</i> SEMBLE LIER LE PPGPP ET LE PPPGPP.	81
FIGURE 3.17	STRUCTURE DU MOTIF <i>PSEUDOMON</i> -GGDEF WT DE <i>PSEUDOMONAS SYRINGAE</i> PV. <i>TOMATO DC3000</i> . 83	
FIGURE 3.18	DES DIFFÉRENCES DU NIVEAU DE LUMINESCENCE ENTRE LES CONSTRUCTIONS WT ET DISR SONT OBSERVABLES POUR CERTAINS MILIEUX TESTÉS POUR LE MOTIF <i>PSEUDOMON</i> -GGDEF.	84
FIGURE 3.19	LE NIVEAU DE LUMINESCENCE EST FORTEMENT DIMINUÉ DANS LA SOUCHE DE <i>P. AERUGINOSA</i> EXPRIMANT UN NIVEAU FAIBLE DE DI-GMPc.	86
FIGURE 3.20	PLUSIEURS STRUCTURES DE <i>PSEUDOMON</i> -GGDEF SEMBLENT COEXISTER.....	88
FIGURE 3.21	L'INFLUENCE DU PPGG SUR L'ARN <i>PSEUDOMON</i> -GGDEF N'EST VISIBLE QU'À DE HAUTES CONCENTRATIONS.	90
FIGURE 3.22	STRUCTURE DU MOTIF <i>PSEUDOMON</i> -RHO WT CHEZ <i>P. AERUGINOSA PA14</i>	91
FIGURE 3.23	CONTEXTE DE LA TRANSCRIPTION DU GÈNE <i>RHO</i> DANS <i>PSEUDOMONAS AERUGINOSA PA14</i>	92
FIGURE 3.24	LA CONSTRUCTION <i>PSEUDOMON</i> -RHO WT MONTRE UN NIVEAU DE LUMINESCENCE PLUS FORT QUE LA CONSTRUCTION AVEC LES MUTATIONS DISR POUR CERTAINS MILIEUX.	93
FIGURE 3.25	STRUCTURE DU MOTIF <i>SUCA-II</i> WT DE <i>PSEUDOMONAS AERUGINOSA PA14</i>	95
FIGURE 3.26	PROFIL DE TRANSCRIPTOMIQUE DE L'OPÉRON <i>SDHCDAB</i> ET DU GÈNE <i>SUCA</i>	96
FIGURE 3.27	LE PROFIL DU NIVEAU DE LUMINESCENCE POUR LE MOTIF <i>SUCA-II</i> EST SEMBLABLE QUEL QUE SOIT LE MILIEU TESTÉ.	96
FIGURE 3.28	STRUCTURE DU MOTIF <i>FIXA</i> WT DE <i>PSEUDOMONAS SYRINGAE</i> PV. <i>TOMATO DC3000</i>	98
FIGURE 3.29	DES DIFFÉRENCES DANS LE NIVEAU DE LUMINESCENCE NE SONT DISTINGUABLES QUE POUR CERTAINS MILIEUX POUR LE MOTIF <i>FIXA</i>	99

FIGURE 3.30	AUCUNE DIFFÉRENCE DANS LE NIVEAU DE LUMINESCENCE N'EST OBSERVABLE ENTRE LES DEUX CONSTRUCTIONS DU MOTIF FIXA EN CONDITION ANAÉROBIE.	101
FIGURE 3.31	STRUCTURE DU MOTIF <i>RNE-II</i> WT DE <i>PSEUDOMONAS AERUGINOSA</i> PA14.	102
FIGURE 3.32	CONTEXTE DE LA TRANSCRIPTION DU GÈNE <i>RNE-II</i> CHEZ <i>P. AERUGINOSA</i> PA14.	103
FIGURE 3.33	DIFFÉRENTS PROFILS DE NIVEAU DE LUMINESCENCE SONT VISIBLES SELON LE MILIEU TESTÉ POUR LE MOTIF <i>RNE-II</i>	104
FIGURE 4.1	ASSEMBLAGE DU SYSTÈME DE SR-PAGE.	116
FIGURE 4.2	REPRÉSENTATION SCHÉMATIQUE DES COUPURES RÉALISÉES LORS DU SELEX AVEC LE SR-PAGE.	118
FIGURE 4.3	PRINCIPE DE LA MÉTHODE DU SR-PAGE.	125
FIGURE 4.4	VALIDATION DE LA MÉTHODE DU SR-PAGE.	126
FIGURE 4.5	LES CONSTRUCTIONS FMN_3 ET FMN_5 PRÉSENTENT UNE ÉNERGIE LIBRE SIMILAIRE ENTRE LEURS STRUCTURES LIÉES ET NON LIÉES.	129
FIGURE 4.6	SELECTION D'UN APTAMERE THIAMINE A PARTIR DU RIBOSWITCH TPP PAR LE SR-PAGE.	132
FIGURE 4.7	LES OLIGONUCLÉOTIDES COMPLÉMENTAIRES AUX ADAPTATEURS PERMETTENT DE RÉTABLIR LE DÉCALAGE DES <i>RIBOSWITCHS</i> LORS DU SR-PAGE.	134
FIGURE 5.1	MODELE HYPOTHETIQUE DE REGULATION DU MOTIF <i>RMF</i> VIA L'INTERVENTION DE DEUX ACTIVATEURS.	144
FIGURE 5.2	COMPARAISON DU SR-PAGE AVEC LES MÉTHODES CONNUES POUR LA DÉCOUVERTE DE NOUVELLES MOLÉCULES.	160
FIGURE 5.3	MÉTHODE UTILISÉE POUR LA RÉALISATION DES DIFFÉRENTES LIBRAIRIES D'ADN GÉNOMIQUE..	162
FIGURE 5.4	REPRÉSENTATION SCHÉMATIQUE DE LA LIBRAIRIE IGR D'OLIGONUCLÉOTIDES.	165
FIGURE 5.5	DIMÉRISATION D'AMORCES LORS DU SÉQUENÇAGE DE LA LIBRAIRIE RW D'OLIGONUCLÉOTIDES.	166
FIGURE 7.1	LE MÊME PROFIL D'EXPRESSION DE LA LUMINESCENCE EST OBSERVABLE POUR TOUS LES MILIEUX TESTÉS POUR LE MOTIF <i>RMF</i>	189
FIGURE 7.2	UNE STRUCTURE IDENTIQUE SE REMARQUE POUR TOUTES LES CONSTRUCTIONS DU MOTIF <i>RMF</i>	190
FIGURE 7.3	PLUSIEURS ESSAIS DE CARTOGRAPHIE <i>IN VITRO</i> PAR <i>IN-LINE</i> PERMETTENT DE DÉTERMINER DES K_D POUR L'ARN <i>RMF</i>	191

FIGURE 7.4	POUR LA MAJORITÉ DES MILIEUX TESTÉS IL N'Y A PAS DE DIFFÉRENCE DANS L'EXPRESSION DE LA LUMINESCENCE POUR LE MOTIF <i>PSEUDOMON-GGDEF</i>	192
FIGURE 7.5	LA STRUCTURE DE <i>PSEUDOMON-GGDEF</i> EST IDENTIQUE QUEL QUE SOIT LES CONSTRUCTIONS TESTÉES.	193
FIGURE 7.6	LES MODULATIONS OBSERVÉES POUR LE PGP EST NON-SPÉCIFIQUE.....	194
FIGURE 7.7	DES VARIATIONS D'EXPRESSION DE LA LUMINESCENCE SONT VISIBLES EN FONCTION DU MILIEU DE CULTURE UTILISÉ POUR LE MOTIF <i>PSEUDOMON-RHO</i>	195
FIGURE 7.8	LE MÊME PROFIL D'EXPRESSION DE LA LUMINESCENCE EST OBSERVABLE POUR TOUS LES MILIEUX TESTÉS POUR LE MOTIF <i>SUCA-II</i>	196
FIGURE 7.9	L'EXPRESSION DE LA LUMINESCENCE VARIE EN FONCTION DES MILIEUX TESTÉS POUR LE MOTIF <i>FIXA</i> .	197
FIGURE 7.10	DES VARIATIONS D'EXPRESSION DE LA LUMINESCENCE SONT VISIBLES DÉPENDAMMENT DU MILIEU TESTÉ POUR LE MOTIF <i>RNE-II</i>	198
FIGURE 7.11	ÉNERGIE LIBRE DES DIFFÉRENTES CONSTRUCTIONS DU RIBOSWITCH FLUOR UTILISÉ LORS DES ESSAIS DE VALIDATION DU SR-PAGE DANS LEUR CONFORMATION LIÉE ET NON LIÉE.	199
FIGURE 7.12	ÉNERGIE LIBRE DES DIFFÉRENTES CONSTRUCTIONS DU RIBOSWITCH FMN UTILISÉ LORS DES ESSAIS DE VALIDATION DU SR-PAGE DANS LEUR CONFORMATION LIÉE ET NON LIÉE.	200
FIGURE 7.13	ÉNERGIE LIBRE DES DIFFÉRENTES CONSTRUCTIONS DU RIBOSWITCH DI-GMPc UTILISÉ LORS DES ESSAIS DE VALIDATION DU SR-PAGE DANS LEUR CONFORMATION LIÉE ET NON LIÉE.	201
FIGURE 7.14	ÉNERGIE LIBRE DES DIFFÉRENTES CONSTRUCTIONS DU RIBOSWITCH NICKEL-COBALT UTILISÉ LORS DES ESSAIS DE VALIDATION DU SR-PAGE DANS LEUR CONFORMATION LIÉE ET NON LIÉE.	202
FIGURE 7.15	LA LIBRAIRIE 3 DU RIBOSWITCH TPP UTILISÉE LORS DU SELEX A PERMIS DE SÉLECTIONNER DES ARN PRÉSENTANT UNE STRUCTURE TIGE-BOUCLE.	208
FIGURE 7.16	LA PROXIMITÉ D'ÉCHANTILLONS EXPRIMANT UN NIVEAU DE LUMINESCENCE PLUS FORT ENTRAÎNE UN ARTEFACT DANS LES LECTURES DE LUMINESCENCE.	214

LISTE DES TABLEAUX

TABLEAU 1.1	LISTE DES LIGANDS CONNUS POUR LES <i>RIBOSWITCHS</i>	21
TABLEAU 3.1	SEULEMENT QUELQUES ARNRÉG ONT ÉTÉ CARACTÉRISÉS CHEZ <i>P. AERUGINOSA</i> DANS DES PROCESSUS CELLULAIRES DIFFÉRENTS.....	37
TABLEAU 3.2	LES ARNnc RÉGULANT EN <i>CIS</i> SONT PLUS CARACTÉRISÉS CHEZ <i>E. COLI</i> QUE CHEZ NOS ORGANISMES D'ÉTUDES : <i>P. AERUGINOSA</i> ET <i>B. THAILANDENSIS</i>	39
TABLEAU 3.3	LISTE DES AMORCES UTILISÉES POUR LA CONSTRUCTION DES PLASMIDES.	47
TABLEAU 3.4	LISTE DES MILIEUX DE CULTURE COMMUNS À TOUS LES MOTIFS TESTÉS.....	51
TABLEAU 3.5	LISTE DES AMORCES UTILISÉES POUR LES DIFFÉRENTES CONSTRUCTIONS DES MOTIFS POUR LE <i>IN-LINE</i>	55
TABLEAU 3.6	RÉSUMÉ DE LA DIFFÉRENCE DU NIVEAU DE LUMINESCENCE ENTRE LES CONSTRUCTIONS WT ET DISR POUR LES SIX ARNnc CANDIDATS DE L'ÉTUDE ET TOUTES LES CONDITIONS TESTÉES.	106
TABLEAU 4.1	LISTE DES AMORCES UTILISÉES POUR LES DIFFÉRENTS CONTRÔLES DES <i>RIBOSWITCHS</i>	112
TABLEAU 4.2	LISTE DES AMORCES UTILISÉES POUR LA RÉALISATION DES LIBRAIRIES DU <i>RIBOSWITCH</i> TPP... ..	114
TABLEAU 4.3	LISTE DES AMORCES UTILISÉES POUR LE SELEX DU <i>RIBOSWITCH</i> TPP.....	120
TABLEAU 4.4	CONTRAINTES POUR LA RÉALISATION DES STRUCTURES DES <i>RIBOSWITCHS</i>	121
TABLEAU 4.5	LISTE DES AMORCES UTILISÉES POUR LE TEST DES ADAPTATEURS.	122
TABLEAU 5.1	LISTE DE TOUS LES CANDIDATS INITIALEMENT PRÉVUS POUR LA CARACTÉRISATION DES ARNnc CHEZ <i>PSEUDOMONAS</i> ET <i>BURKHOLDERIA</i>	139
TABLEAU 5.2	RÉCAPITULATIF DES VALEURS DE K_D OBTENUES POUR LES DIFFÉRENTS ESSAIS AVEC <i>RMF</i>	142
TABLEAU 5.3	AMÉLIORATION DES CONNAISSANCES SUR LES ARNnc EN <i>CIS</i> CHEZ <i>PSEUDOMONAS</i>	156
TABLEAU 7.1	SÉQUENCES DES CONSTRUCTIONS TESTÉES POUR CHAQUE MOTIF.....	187
TABLEAU 7.2	RÉCAPITULATIF DES CLONES SÉLECTIONNÉS LORS DU SELEX AVEC LE <i>RIBOSWITCH</i> TPP.	203
TABLEAU 7.3	PRÉDICTION DE STRUCTURE TIGE-BOUCLE DANS LA LIBRAIRIE 3 ET DANS UNE LIBRAIRIE ALÉATOIRE.	209

LISTE DES ABRÉVIATIONS

2'-dG	2' -deoxyguanosine
ADN	Acide DésoxyriboNucléique
ADNc	ADN complémentaire
Amorce F	Amorce forward
Amorce R	Amorce reverse
AMP	Adénosine MonoPhosphate
AMP-GMPc	AMP - GMP cyclique
Anti-SD	Anti Shine-Dalgarno
ARN	Acide RiboNucléique
ARNm	ARN messenger
ARNnc	ARN non codants régulateurs
ARNr	ARN ribosomaux
ARNrég	Petits ARN régulateurs
ARNt	ARN de transfert
<i>B. cereus</i>	<i>Bacillus cereus</i>
<i>B. subtilis</i>	<i>Bacillus subtilis</i>
<i>B. mallei</i>	<i>Burkholderia mallei</i>
<i>B. pseudomallei</i>	<i>Burkholderia pseudomallei</i>
<i>B. thailandensis</i>	<i>Burkholderia thailandensis</i>
BCAA	Acide Aminé à Chaîne ramifiée Branché (Branched Chain Amino Acid)
CAA	CasAmino Acid
CDD	Conserved Domain Database
CDS	Coding sequence (Séquence codante)
Comp	Compensatoire

CRP	cAMP receptor protein
di-AMPC	di-AMP cyclique
di-GMPc	di-GMP cyclique
Disr	Disruptive
DO	Densité optique
<i>E. coli</i>	<i>Escherichia coli</i>
FMN	Flavine MonoNucléotide
GMP	Guanosine MonoPhosphate
HK	Histidine Kinase
HMP-PP	4-amino-2-methyl-5-hydroxymethylpyrimidine pyrophosphate
IGR	InterGenic Regions (region intergénique)
kb	Kilo bases
<i>L. monocytogenes</i>	<i>Listeria monocytogenes</i>
Mb	Mégabase (millions de bases)
Nts	Nucléotides
<i>P. aeruginosa</i>	<i>Pseudomonas aeruginosa</i>
PCR	Polymerase Chain Reaction (réaction de polymérisation en chaîne)
pGpG	5'-phosphoguananylyl-(3',5')-guanosine
ppGpp	Guanosine tétraphosphate
pppGpp	Guanosine pentaphosphate
Pré-Q1	Pré-queuosine 1
PRPP	Phosphoribosylpyrophosphate
PSSM	Peptone Salt Succinate Medium
Pv	Pathovar
RBS	Ribosome Binding Site (site de fixation du ribosome)
RR	Régulateur de réponse

rut	rho utilization
SAM	S-AdénosylMéthionine
SAH	S-AdénosylHomocystéine
SCFM	Synthetic Cystic Fibrosis sputum Medium
SD	Shine-Dalgarno
SELEX	<i>Systematic Evolution of Ligand by Exponential Enrichment</i>
SHAPE-MaP	Selective 2'-hydroxyl Acylation analyzed by Primer Extension – Mutational profiling
SR-PAGE	Shifted Reverse PolyAcrylamide Gel Electrophoresis
Tm	Température d'hybridation des amorces (melting temperature)
TPP	Thiamine PyroPhosphate
RBS	Ribosome Binding Site
UTR	Untranslated region (région non traduite)
<i>V. cholerae</i>	<i>Vibrio cholerae</i>
WT	Wild Type / Sauvage
ZTP	5-aminoimidazole-4-carboxamide ribonucleoside-5' -triphosphate

1 INTRODUCTION

1.1 La régulation chez les procaryotes

Au cours de leur croissance les bactéries sont constamment soumises à des variations de conditions environnementales, qui peuvent être des changements de température, de pH, des carences nutritionnelles... Afin de survivre à ces changements, elles ont la nécessité de s'adapter et la régulation de l'expression de ces gènes est la solution la plus efficace. Les bactéries ont également la nécessité de devoir maintenir un niveau stable de ces composés intracellulaire, ce que l'on appelle l'homéostasie. Pour cela elles utilisent plusieurs moyens à différents niveaux qui sont en autres, les facteurs de transcription, les systèmes à deux composants et les ARN non codants.

1.1.1 Les facteurs de transcription

Un des premiers moyens pour réguler l'expression des gènes est l'utilisation de facteur de transcription. Ce sont des protéines de régulation se fixant à l'ADN permettant ainsi de réguler l'initiation de la transcription (Balleza *et al.*, 2008). Ils sont composés de deux domaines, un domaine de fixation à l'ADN et un domaine permettant la liaison d'un métabolite (Martinez-Antonio & Collado-Vides, 2003). Ils peuvent agir en activant la transcription en se fixant en amont du promoteur favorisant ainsi la liaison de l'ARN polymérase, mais également en la réprimant en se fixant directement sur le promoteur et empêchant ainsi le recrutement de la polymérase (Balleza *et al.*, 2008).

Il existe plus de 250 facteurs de transcription chez *Escherichia coli* (*E. coli*) (Shimada *et al.*, 2008). La plupart d'entre eux régulent un opéron spécifiquement ainsi qu'un nombre de gènes limités (Shimada *et al.*, 2008), comme MelR situé en amont des gènes *melAB* chez *E. coli* codant pour des protéines impliquées dans le métabolisme et le transport du mélibiose (Grainger *et al.*, 2004). Cependant, certains facteurs de transcription plus globaux régulent une multitude de gènes, comme CRP, FNR, Lrp (Martinez-Antonio & Collado-Vides, 2003; Shimada *et al.*, 2008).

L'un des facteurs de transcription le plus important et le plus étudié est CRP (cAMP receptor protein) (Visweswariah & Busby, 2015). L'étude de Shimada et collaborateurs publiée en 2011 a permis de mettre en évidence, grâce à une méthode de SELEX génomique, entre 379 et 500 cibles de CRP dans les bactéries (Shimada *et al.*, 2011). Ce système permet de ressentir les signaux présents au sein de la bactérie, mais il existe d'autres systèmes permettant de sentir et

transmettre des messages extérieurs comme les systèmes à deux composants (Dale & Park, 2010).

1.1.2 Les systèmes à deux composants

Chez les procaryotes plus de 300 systèmes à deux composants sont connus (Dale & Park, 2010). Ce système de transduction de signal est composé de deux éléments qui sont une protéine membranaire appelée histidine kinase (HK) et une protéine cytoplasmique nommée régulateur de réponse (RR) (Dale & Park, 2010). L'HK ainsi que le RR comportent deux domaines. L'HK contient un domaine détecteur situé en dehors de la bactérie qui permet de ressentir le message extérieur et le domaine transmetteur qui lui est sur la face cytoplasmique permettant d'interagir avec le régulateur de réponse (Dale & Park, 2010). L'HK peut ressentir de nombreux signaux comme des nutriments, des cations, des variations dans l'osmolarité (Goodman *et al.*, 2004). Le RR possède un domaine receveur contenant un groupement aspartate pour la réception du groupement phosphate de l'HK et un domaine réponse permettant la liaison à l'ADN (Dale & Park, 2010).

Lors de la détection d'un stimulus, il y a un changement de conformation de l'HK qui s'autophosphoryle via un groupement histidine conservé. Ce groupement phosphate sera alors transféré au RR qui pourra ainsi interagir avec l'ADN pour réguler la transcription du gène (Figure 1.1) (Dale & Park, 2010). Une étape également importante dans cette transduction du signal est la déphosphorylation de l'HK, afin d'arrêter le message, qui peut se faire directement par le RR mais aussi via une phosphatase spécifique (Dale & Park, 2010).

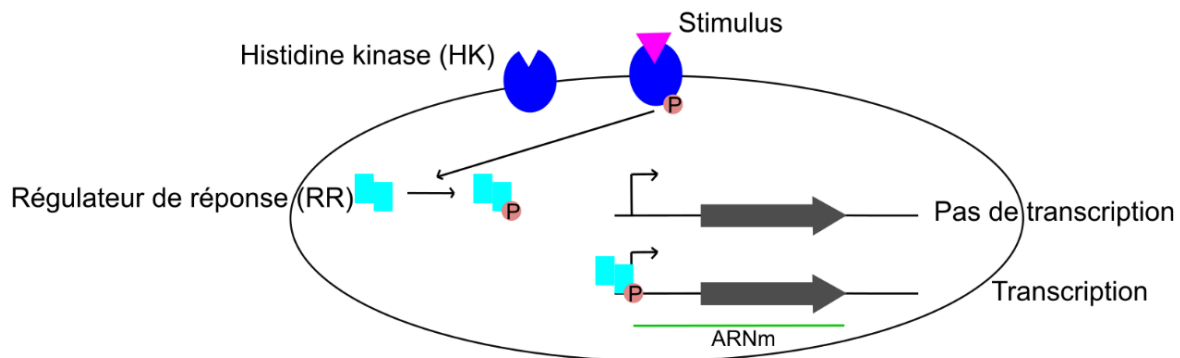


Figure 1.1 Schéma du mode de fonctionnement des systèmes à deux composants.

Le récepteur histidine kinase (HK) est représenté en bleu, et son stimulus en rose. Suite à l'autophosphorylation de ce dernier, le régulateur de réponse (RR) représenté en cyan est phosphorylé permettant la transcription du gène. Inspiré de (Dale & Park, 2010).

Les systèmes à deux composants peuvent sentir diverses molécules dans des métabolismes différents, comme par exemple, la présence de magnésium (Mg^{2+}) via le système PhoP-PhoQ chez *Salmonella* (Groisman, 2001) ou la réponse au stress oxydatif via le système ArcB-ArcA chez *E. coli* (Loui *et al.*, 2009).

Chez *Pseudomonas aeruginosa* (*P. aeruginosa*), l'un de nos organismes d'étude, plus de 50% des systèmes à deux composants présents sont impliqués dans la virulence (Francis *et al.*, 2017). Il a été montré que la virulence chez cet organisme est régulée via un réseau de multikinase c'est-à-dire plusieurs HK travaillant en même temps afin de créer un réseau complexe capable de sentir de multiples signaux simultanément (Francis *et al.*, 2017). Les systèmes à deux composants sont complexes et permettent dans certains cas, comme dans le système GacS/GacA, de réguler l'expression d'ARN non codants régulateurs (Francis *et al.*, 2017).

1.1.3 Les ARN non codants régulateurs

Au cours des dernières décennies, de nouvelles classes d'ARN ont été découvertes ne comportant pas d'informations destinées à la synthèse de protéines, mais ayant une fonction régulatrice. Ceux-ci ont un rôle important dans l'adaptation des organismes à leur environnement. En effet, ils permettent une économie d'énergie par rapport aux protéines de régulation du fait de l'absence de traduction. Il s'agit des ARN non codants régulateurs (ARNnc). Les ARNnc jouent un rôle important dans la régulation des gènes dans les trois domaines de la vie (eucaryotes, procaryotes et archées) (Hess & Marchfelder, 2012).

Les premiers ARNnc ont été découverts chez les procaryotes en 1967 (Hess & Marchfelder, 2012). On peut distinguer deux classes d'ARNnc régulateurs en fonction de leur localisation par rapport à leur cible. Tout d'abord, les ARNnc en *cis*, ils sont situés dans la partie 5' ou 3' UTR des ARNm, comme les *riboswitchs* et les T-box (Desgranges *et al.*, 2019). Certains petits ARN régulateurs agissent également en *cis*, on les appelle les ARN anti-sens et sont codés sur le brin opposé de leur cible et lui sont parfaitement complémentaires (Thomason & Storz, 2010). Il existe également les ARNnc en *trans*. Ceux-ci sont transcrits de manière indépendante et sont situés à distance de leurs cibles (Desgranges *et al.*, 2019). La classe la plus caractérisée d'ARNnc en *trans* est celle des petits ARN régulateurs (Hess & Marchfelder, 2012).

1.1.3.1 Les petits ARN régulateurs

En 1984, la fonction du premier petit ARN régulateur (ARNrég) a été caractérisée chez *E. coli*, il s'agit de l'ARN MicF (Mizuno *et al.*, 1984). Il a été montré que cet ARNrég inhibe la traduction de l'ARNm codant OmpF, une porine majeure de la membrane externe (Mizuno *et al.*, 1984). L'ARN MicF interagit au niveau de la partie 5' UTR ainsi que dans le début de la séquence codante bloquant la traduction de l'ARNm *ompF* (Mizuno *et al.*, 1984). Depuis, les ARNrég ont été largement étudiés et plusieurs ARNrég ont été identifiés et leurs fonctions déterminées. En effet, selon Rfam 549 familles d'ARNrég sont maintenant annotées dans les bactéries (Boutet *et al.*, 2022).

Les ARNrég sont généralement non codants. Leur taille est typiquement comprise entre 50 et 300 nucléotides (Lalaouna *et al.*, 2013). Ils sont structurés par des tiges-boucles et agissent sur des cibles exprimées de manière indépendante (Storz *et al.*, 2011). Ils peuvent agir à différents niveaux dans la cellule en provoquant un arrêt précoce de la transcription (Bossi *et al.*, 2012), en inhibant ou en activant la traduction, ou en diminuant la stabilité des ARNm via la dégradation par des ARNases (Lalaouna *et al.*, 2013). Les ARNrég ne présentent qu'une complémentarité limitée leur permettant de réguler plusieurs cibles situées à différents endroits dans le génome (Desgranges *et al.*, 2019). Certains ARNrég présentent un régulon très grand comme GcvB présent chez *E. coli* et *Salmonella* (Lalaouna *et al.*, 2019b). En effet, GcvB a plus de 30 cibles connues interagissant avec des parties différentes de l'ARNrég (Lalaouna *et al.*, 2019b). Afin d'aider l'interaction entre l'ARNrég et l'ARNm une chaperonne peut être nécessaire comme la protéine Hfq (Dale & Park, 2010) ou encore ProQ (Smirnov *et al.*, 2017).

1.1.3.2 Les ARN antisens

Les ARN antisens sont des ARNnc agissant en *cis*, ils sont généralement associés à des éléments mobiles comme les plasmides, phages ou transposons, mais certains exemples sont codés également au niveau du chromosome (Hess & Marchfelder, 2012). Ces transcrits antisens sont codés sur le brin d'ADN opposé à un autre gène avec lequel ils peuvent potentiellement formés des appariements de bases (Thomason & Storz, 2010).

Les ARN antisens peuvent agir au niveau de la transcription soit par un mécanisme d'interférence soit par un mécanisme d'atténuation, soit au niveau de la stabilité des ARNm par le recrutement de ribonucléases mais également au niveau de la traduction de manière directe ou indirecte (Thomason & Storz, 2010). Du fait de la complémentarité parfaite entre l'ARN antisens et son ARNm cible, son action ne nécessite pas l'aide d'une chaperonne comme Hfq (Hess &

Marchfelder, 2012). L'existence d'ARN antisens a été rapportée pour une grande variété de gènes avec des fonctions variées, comme les systèmes toxine-antitoxine, les processus de conjugaison ou de réplication (Hess & Marchfelder, 2012).

Tous les ARNnc en *cis* ne sont pas des ARN antisens. En effet, certains peuvent en présence d'un métabolite ou d'une variation de condition expérimentale (température, pH...) changer de conformation entraînant ainsi une modification de l'expression du gène situé directement en aval (Narberhaus *et al.*, 2006). Ce sont les ARNnc régulant en *cis* qui sont trouvés dans la même molécule que l'ARNm concerné, ceux-ci seront décrits plus en détail dans la section suivante.

1.2 Les ARN non codants régulant en *cis*

Le terme ARNnc en *cis* est souvent retrouvé pour désigner les ARN antisens se situant en dehors de l'ARNm de la cible (Thomason & Storz, 2010). Cependant, de nombreuses classes d'ARNnc en *cis* se retrouvent sur le même ARNm que le gène à réguler au niveau de leur région 5' ou 3' UTR (Figure 1.2). Comme par exemple les T-box (Kreuzer & Henkin, 2018), les thermorégulateurs (Hess & Marchfelder, 2012) ou les *riboswitchs* (Kim & Breaker, 2008).

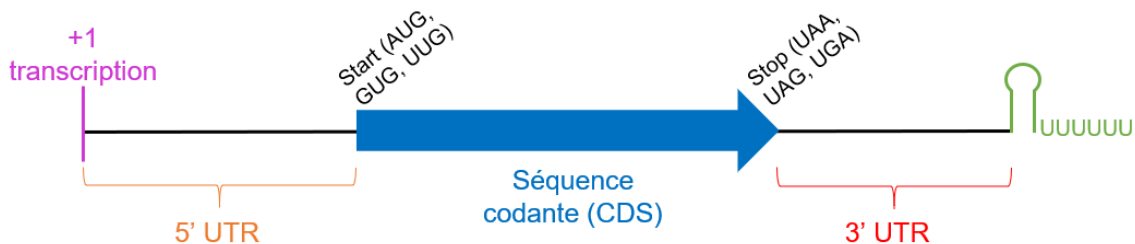


Figure 1.2 Représentation schématique d'un ARNm.

Un ARNm se compose de différents éléments. Il commence au niveau du +1 de transcription, représenté ici en rose, constituant le premier nucléotide de l'ARNm. Les nucléotides situés entre ce +1 de transcription et le codon initiateur de la traduction de la séquence codante est la région 5' non traduite (5' UTR, représenté en orange). La séquence codante (CDS) représentée par la flèche bleue est la partie qui sera traduite jusqu'au codon stop. Enfin la partie de l'ARNm présente entre le codon stop et le terminateur de transcription, ici terminateur rho indépendant (en vert) est la partie 3' UTR (rouge).

1.2.1 Les T-box

Les T-box ont été les premiers exemples de système où un ARNnc en *cis* répond directement à un signal de la cellule par un réarrangement de conformation afin de contrôler l'expression des gènes (Kreuzer & Henkin, 2018). Ils sont présents dans les bactéries à Gram positif comme les

Firmicutes, ou encore, les Actinobactéries (Kreuzer & Henkin, 2018; Vitreschak *et al.*, 2008). Le premier T-box découvert est celui présent en amont du gène *tyrS* chez *Bacillus subtilis* (*B. subtilis*) (Henkin *et al.*, 1992). Les premiers T-box ont été trouvés via la présence d'une séquence de 14 nucléotides se retrouvant en amont de plusieurs gènes codant pour des aminoacyl-tRNA synthétases (Grundy & Henkin, 1993; Kreuzer & Henkin, 2018). Il a été ensuite montré que la plupart des gènes d' aminoacyl-tRNA synthétases sont régulés par un mécanisme d'antiterminaison de la transcription (Grundy *et al.*, 1994). Une grande conservation de la séquence primaire et de la structure secondaire est présente dans les parties 5' non traduites en amont de gènes impliqués dans le métabolisme des ARN de transfert (ARNt) comme les aminoacyl-tRNA synthétases (Grundy *et al.*, 1994). La régulation de ces gènes est essentielle à la survie de la bactérie car une mauvaise régulation peut conduire à des miscacylations des ARNt ou une induction de la réponse stringente due à l'accumulation de ARNt non chargés (Sherwood *et al.*, 2015).

Les T-box sont composés de trois tiges boucles (dénommées *stem I*, *stem II* et *stem III*) et d'un pseudonœud (*stem IIA/B*) suivis par une structure mutuellement exclusive de terminateur et antitermineur de transcription (Figure 1.3) (Grundy *et al.*, 1994; Kreuzer & Henkin, 2018). La partie la plus importante est présente dans la tige boucle I. Il s'agit d'une séquence de trois nucléotides correspondant à un codon de l'acide aminé correspondant au gène situé en aval (Sherwood *et al.*, 2015). Cette séquence est appelée séquence *specifier*. Celle-ci se situe dans la seconde boucle de la tige I nommée *specifier loop* (Kreuzer & Henkin, 2018). Concernant les tiges II et IIA/B leur fonction n'est pas encore définie, car certains T-box de *B. subtilis* n'en possède pas comme *glyQS* (Yousef *et al.*, 2003) et d'autres ne sont pas fonctionnels en absence de ces tiges comme pour *tyrS* (Rollins *et al.*, 1997).

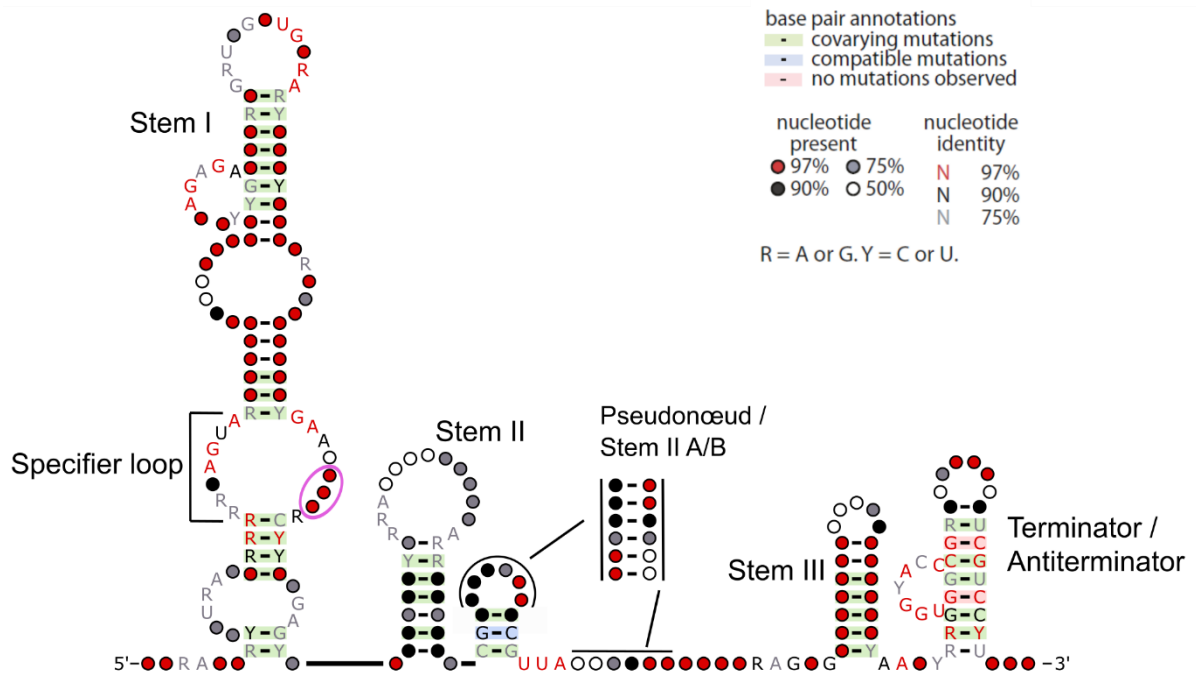


Figure 1.3 Structure consensus des T-box.

Structure réalisée à partir des séquences des T-box présents dans la base de données Rfam (Griffiths-Jones *et al.*, 2003), puis dessinée en utilisant les logiciels R2R (Weinberg & Breaker, 2011) et Inkscape.. Les codons permettant la reconnaissance de l'acide aminé sont représentés par le cercle rose présent dans la *specifier loop*. Chaque nucléotide A, U, G, C, R ou Y peuvent être annotés en couleur en fonction de leur identité dans les différentes séquences présentes en rouge 97%, en noir 90% et en gris 75%. Les nucléotides R ou Y correspondent à la nomenclature IUPAC, où R correspond à A ou G et Y correspond à C ou U. Les ronds indiquent la présence d'un nucléotide et les couleurs la probabilité que ce nucléotide soit présent avec : en rouge 97%, en noir 90%, en gris 75% et en blanc 50%. Enfin les bandes de couleurs entre les nucléotides correspondent aux variations des paires de bases entre les différentes séquences, en vert cela correspond à une covariation, en bleu à des mutations compatibles et les paires de bases conservées sont en rouge.

Les ligands des T-box sont les ARNt qui sont eux-mêmes des ARN structurés (Kreuzer & Henkin, 2018). Chaque T-box interagit spécifiquement avec son ARNt correspondant via deux éléments. Tout d'abord grâce à une interaction codon-anticodon au niveau de la séquence *specifier* dans la tige boucle I (Grundy *et al.*, 1994). Ainsi qu'avec une interaction entre la structure antiterminatrice (5' - UGGN -3') et l'ARNt (5' - NCCA -3') (Grundy *et al.*, 1994). En présence d'un ARNt non chargé, celui-ci pourra se lier aux trois nucléotides de la séquence *specifier* et à la tige antiterminatrice ce qui stabilisera cette structure et permettra ainsi la transcription du gène (Figure 1.4). Cependant en présence d'un ARNt chargé avec un acide aminé, il n'y aura pas de reconnaissance au niveau de la séquence T-box impliquant donc la formation du terminateur de transcription (Figure 1.4) (Kreuzer & Henkin, 2018).

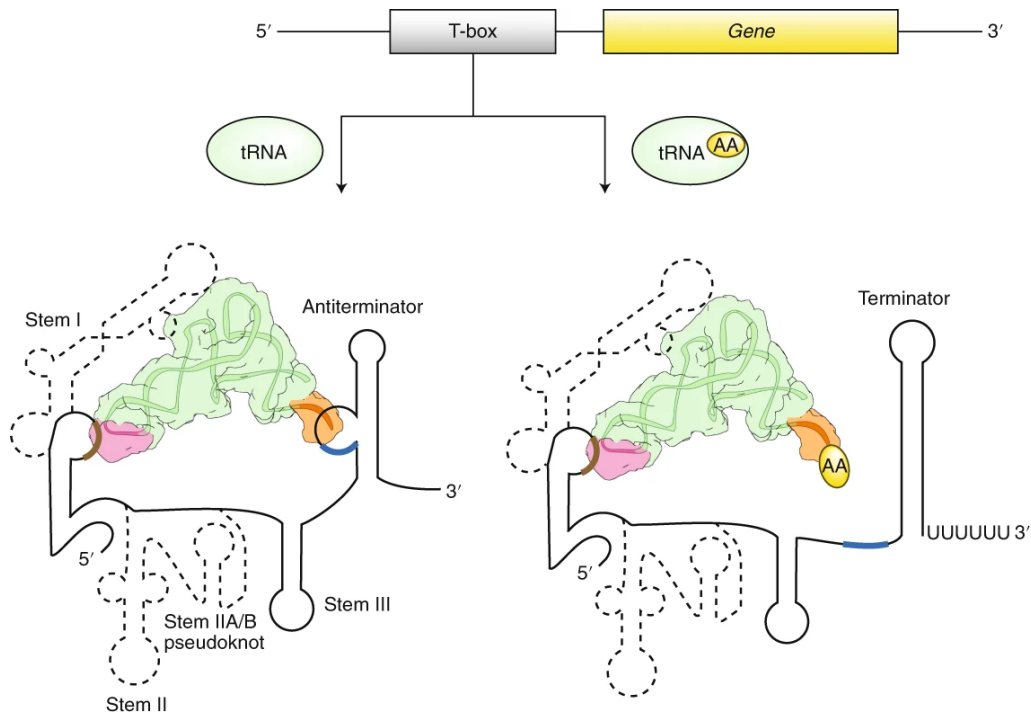


Figure 1.4 Représentation schématique du fonctionnement des T-box.

Les T-box sont situés dans la partie 5' UTR du gène qu'il régule. En présence d'ARNt non chargé, présenté dans la partie gauche de la figure, celui-ci interagit avec le site *specifier* dans la tige I (en brun) ainsi qu'avec la séquence T-box représentée en bleu. Ceci entraîne la formation d'un antiterminateur permettant donc la transcription du gène. Cependant en présence d'ARNt chargé de son acide aminé correspondant, celui-ci ne pourra interagir qu'avec le site *specifier* dans la tige I, mais pas avec la séquence T-box impliquant la présence d'un terminateur de transcription. Les parties en traits pleins sont les parties constantes des T-box retrouvées quel que soit l'espèce étudiée. Les parties en pointillés représentent les parties variables des T-box qui peuvent être observées en fonction de l'espèce dans laquelle se retrouve le T-box. Tiré de (Weaver & Serganov, 2019).

La plupart des T-box fonctionnent par un mécanisme de régulation de la transcription. Cependant, devant le gène *ileS* des Actinobactéries, il n'y a pas de terminateur de transcription apparent (Sherwood *et al.*, 2015). En effet avec l'aide d'étude de prédiction de structure il semble que la régulation se fasse au niveau de la traduction via un mécanisme de séquestration du Shine-Dalgarno (SD) (Sherwood *et al.*, 2015; Vitreschak *et al.*, 2008). Le SD est la séquence de reconnaissance du ribosome située dans le RBS (Volkenborn *et al.*, 2020). L'étude de Sherwood *et al* (2015) a également permis de mettre en évidence d'autres structures de T-box avec des tiges I plus courtes. Cependant, aucun T-box ne présente des modifications dans la tige I, II et le pseudonœud IIA/B en même temps suggérant que le système peut supporter des changements, mais pas dans tous les domaines.

1.2.2 Les ARN thermorégulateurs

La température est un facteur environnemental qui affecte le métabolisme global des bactéries, elles ont donc besoin de finement surveiller ces variations (Hess & Marchfelder, 2012). Cette surveillance se fait via divers procédés comme la présence de protéines de régulation comme des facteurs de transcription, des chaperonnes, la fluidité de la membrane, le changement dans la topologie de l'ADN ou alors la structure de l'ARN (Hess & Marchfelder, 2012; Loh *et al.*, 2018). Les thermorégulateurs sont des structures d'ARN situées dans la partie 5' UTR des ARN messagers (ARNm). Ils changent de conformation suite à une variation de température dans la cellule (Loh *et al.*, 2018). Les thermorégulateurs agissent au niveau de la traduction en séquestrant le site de fixation du ribosome (RBS) à basse température, ainsi une augmentation de la température entraînera la libération de ce dernier (Figure 1.5) (Narberhaus *et al.*, 2006). Le RBS est la région comprenant le SD et les nucléotides jusqu'au codon d'initiation de la traduction (Volkenborn *et al.*, 2020). Le fait que le changement de conformation soit un processus graduel permet que la réponse soit aussi directement corrélée avec la variation de la température (Hess & Marchfelder, 2012).

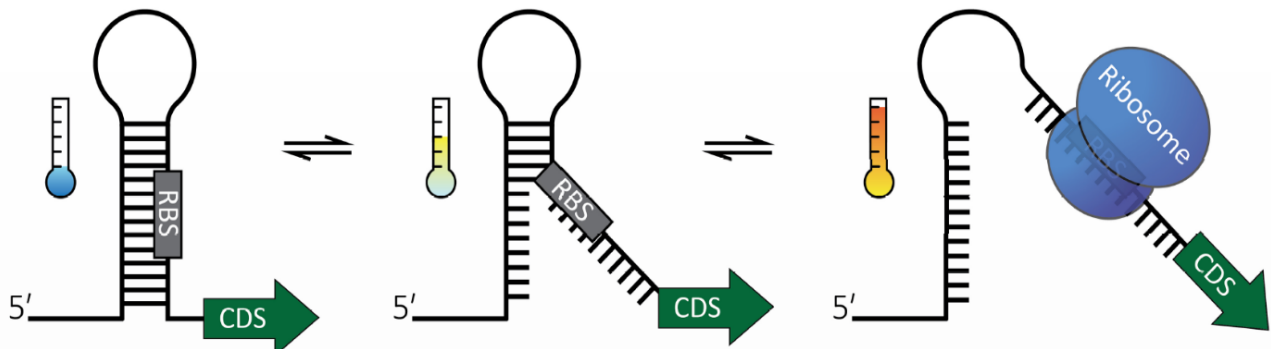


Figure 1.5 **Fonctionnement des ARN thermorégulateurs.**

Les ARN thermorégulateurs changent de conformation en réponse à des variations de température. À basse température (schéma tout à gauche) le RBS est séquestré dans une structure double brin empêchant le recrutement du ribosome, il n'y aura donc pas de traduction du gène (représenté par la flèche verte CDS). Cependant quand la température augmente la structure tige boucle s'ouvre jusqu'à ce que le RBS soit accessible pour le ribosome (représenté ici en bleu) permettant ainsi la traduction du gène en aval. Tiré de (Loh *et al.*, 2018).

Un fait intéressant sur les thermorégulateurs est que leurs séquences et leurs structures secondaires sont très diverses et peu conservées (Loh *et al.*, 2018). Une quinzaine de famille

différentes de thermorégulateurs sont répertoriées dans la base de données Rfam (Griffiths-Jones *et al.*, 2003; Kalvari *et al.*, 2021), comme ROSE (*Repression of heat Shock gene Expression*) qui est retrouvé dans plusieurs α et γ -protéobactéries (Hess & Marchfelder, 2012). Il est composé de deux, trois ou quatre tige-boucles et est d'une taille de 60 à 100 nucléotides de long (Grosso-Becerra *et al.*, 2014). Il est caractérisé par un petit motif consensus prédit pour agir comme anti-SD empêchant ainsi l'accès de la sous-unité 30S du ribosome (Hess & Marchfelder, 2012). Un autre motif pouvant être retrouvé est le FourU, où quatre uridines interagissent avec la séquence du SD (Loh *et al.*, 2018). Le plus étudié est celui présent chez *Salmonella* devant le gène *agsA* codant pour une petite protéine de réponse au stress chaud (Waldminghaus *et al.*, 2007). La plupart des gènes régulés par les thermorégulateurs sont des protéines de réponse au stress chaud ou des gènes de virulence (Narberhaus *et al.*, 2006).

Le premier thermorégulateur trouvé fut celui en amont du gène *cIII* chez le phage λ (Altuvia *et al.*, 1989). Ce thermorégulateur permet la transition entre le cycle lysogène ou lytique. En effet, à 37°C la structure favorise la libération du RBS entraînant la traduction de la protéine CIII qui à haute concentration place le phage en cycle lysogène (Narberhaus *et al.*, 2006).

Les thermorégulateurs ont également un rôle important chez les bactéries pathogènes. L'un des premiers thermorégulateurs chez les pathogènes a été trouvé devant le gène *prfA* chez *Listeria monocytogenes* (Johansson *et al.*, 2002). Chez *Pseudomonas*, la présence de plusieurs thermorégulateurs a été montrée. Notamment un motif ROSE-like présent en amont du gène *ibpA* codant pour une petite protéine de stress chaud (Krajewski *et al.*, 2013). Elle présente également deux autres thermorégulateurs présents devant les gènes *rhlA* et *lasI*, qui ne sont pas spécifiquement impliqués dans les stress de température mais permettent de réguler l'expression de nombreux facteurs de virulence via la détection du *quorum sensing* (Grosso-Becerra *et al.*, 2014). Ceci permettant de détecter l'entrée dans les cellules de l'hôte et ainsi activer la cascade de virulence.

1.2.3 Ribosomal RNA Leader (*r-leader*)

La synthèse des ribosomes est un procédé complexe comprenant plusieurs étapes de maturation (Eckert & Weinberg, 2020). Ils ont une grande importance dans le métabolisme des bactéries et consomment ainsi une grande quantité de l'énergie de la cellule (Eckert & Weinberg, 2020). Chez les bactéries les ribosomes sont composés de deux sous unités, 50S et 30S formant le ribosome 70S. Chacune de ces sous-unités sont composées d'ARN ribosomiaux (ARNr) ainsi que de protéines correspondant au final à environ une soixantaine de composants (Dale & Park, 2010).

Le maintien de la stœchiométrie de tous les composants du ribosome est essentiel pour son fonctionnement, la présence de structure dans l'ARNm est un de ces moyens de régulation (Meyer, 2018). En effet il existe un processus de rétroaction impliquant l'ARNm et les protéines ribosomales appelé *r-leader* (Eckert & Weinberg, 2020). Lors de leur découverte, il avait été posé comme hypothèse que la structure du *r-leader* mime la structure de l'ARNr correspondant (Nomura *et al.*, 1980). Cependant une étude plus récente montre que cette similarité serait plutôt inhabituelle, ou alors cantonnée aux motifs qui sont les plus répandus (Meyer, 2018).

Les *r-leader* sont situés dans la partie 5' UTR des ARNm codant pour des protéines ribosomales. À la suite de la liaison de la protéine il y aura un changement de conformation de l'ARNm entraînant sa fonction (Eckert & Weinberg, 2020). Ces changements pourront avoir plusieurs conséquences comme une inhibition de la traduction via une séquestration du SD ou un emprisonnement du ribosome, mais aussi un changement dans le taux de dégradation de l'ARNm ou encore une terminaison précoce de la transcription (Fu *et al.*, 2013; Meyer, 2018). Chez *E. coli* plus du quart des opérons codant pour des protéines ribosomales sont régulés par leur propre moyen via notamment ce mécanisme de rétroaction des protéines ribosomales sur leur région leader (Meyer, 2018). Jusqu'à présent, il existe 55 *r-leader* prédits présents aussi bien chez les bactéries que chez les archées (Eckert & Weinberg, 2020). Les *r-leader* sont généralement peu répandus dans les différents ordres. Une étude a permis de montrer que sur dix motifs présents chez *E. coli* seulement trois étaient répandus dans le phylum des eubactéries, mais les sept autres étaient quant à eux limités aux γ -protéobactéries (Fu *et al.*, 2013).

Une analyse sur 298 bactéries a permis de montrer que le gène *rpsL* fait partie des gènes les plus régulés par un mécanisme de terminaison prématurée de la transcription (Naville & Gautheret, 2010). Ce mécanisme de transcription prématurée se fait grâce à la formation d'un terminateur intrinsèque de transcription suite à la liaison de la protéine ribosomale S7 (Naville & Gautheret, 2010). Il est présent dans un opéron comprenant les gènes *rpsL*, *rpsG* codant pour les protéines ribosomales S12 et S7, ainsi que les gènes *fusA* et *tuf* codant pour des facteurs d'élongations (Naville & Gautheret, 2010). Une analyse de structure de la partie 5' UTR en amont de cet opéron a permis de mettre en évidence une séquence consensus contenant un terminateur de transcription Rho indépendant dans plusieurs espèces bactériennes (Naville & Gautheret, 2010). Il a également été montré chez *E. coli* que les gènes *rpsL* et *rpsG* étaient régulés via une boucle de rétrocontrôle de la protéine S7, produit du gène *rpsL* (Naville & Gautheret, 2010).

1.2.4 Les *riboswitchs*

Les *riboswitchs* sont des structures d'ARN situées dans la partie 5' UTR des ARNm. Ces ARNnc en *cis* étant une partie importante de ce projet de doctorat ils seront présentés plus en détails dans la section suivante (1.3).

1.3 Les *riboswitchs* chez les procaryotes

Le terme *riboswitch* a été présenté pour la première fois en 2002 par Nahvi et ses collaborateurs pour désigner un ARN ayant un rôle régulateur en présence d'un métabolite donné, mais sans l'aide de facteurs protéiques tel que démontré pour la structure d'ARN présente en amont du gène *btuB* répondant à l'adénosylcobalamine (Nahvi *et al.*, 2002).

Les *riboswitchs* sont des structures d'ARN le plus souvent retrouvées dans les parties 5' UTR des ARNm bactériens (Winkler & Breaker, 2003). Ils sont quasiment toujours situés dans la même orientation et sur le même brin que le gène qu'ils régulent (Sherlock & Breaker, 2020). Ces *riboswitchs* lient spécifiquement un ligand permettant ainsi de réguler l'expression d'un ou plusieurs gènes situés directement en aval de celui-ci (Kim & Breaker, 2008). Dans certains cas, plusieurs classes de *riboswitchs* peuvent lier le même ligand, comme par exemple les *riboswitchs* S-Adénosyl Méthionine (SAM), di-GMP cyclique (di-GMPc), préQ1 et magnésium. Il arrive souvent que la régulation orchestrée par les *riboswitchs* soit une boucle de rétrocontrôle, c'est-à-dire que la molécule synthétisée par le gène peut venir se lier au *riboswitch* afin de moduler sa propre expression (Kim & Breaker, 2008) (Figure 1.6).

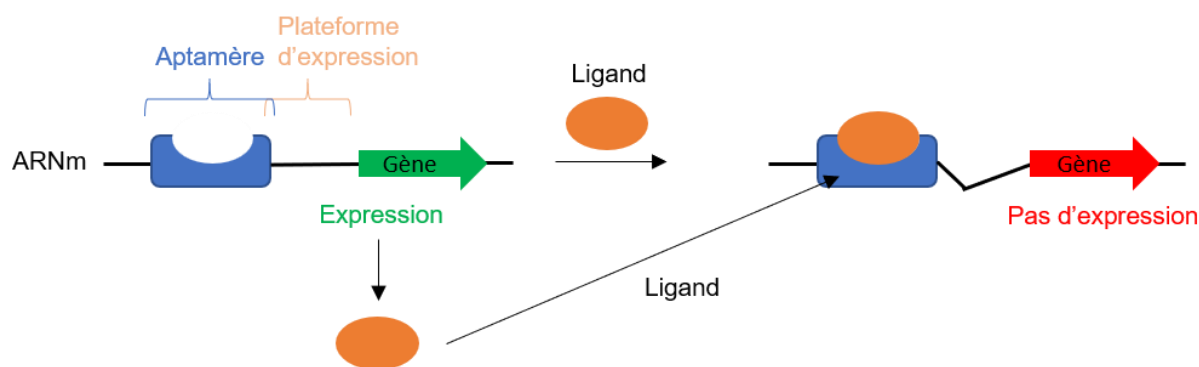


Figure 1.6 Représentation schématique du mode général de régulation par les *riboswitchs*.

La partie en bleu représente le domaine aptamère du *riboswitch*, la partie en noir représente la plateforme d'expression du *riboswitch*. Le rond orange représente le ligand qui peut soit venir du milieu, soit être le produit du gène régulé impliquant donc une boucle de rétrocontrôle. Les flèches représentent le gène régulé par le *riboswitch* dont la couleur verte ou rouge indique l'expression ou non du gène respectivement. Inspiré de (Winkler & Breaker, 2003).

Ces molécules d'ARN sont généralement constituées de deux domaines fonctionnels qui sont l'aptamère et la plateforme d'expression (Figure 1.7). Le domaine aptamère agit de façon similaire à un récepteur protéique en servant d'élément de reconnaissance (Mandal & Breaker, 2004b), il est très conservé et permet la liaison du ligand (Kim & Breaker, 2008; Winkler & Breaker, 2003). La seconde partie, nommée plateforme d'expression, est située directement en aval du domaine aptamère et c'est cette partie qui est responsable de la fonction du *riboswitch* en interagissant avec des éléments impliqués dans l'expression des gènes, comme le site de fixation du ribosome ou les tiges de terminateurs de transcription (Mandal & Breaker, 2004b; Winkler & Breaker, 2003). La plateforme d'expression est moins conservée entre les organismes permettant ainsi une vaste diversité de régulation par les *riboswitchs* (Mandal & Breaker, 2004b). L'action des *riboswitchs* est basée sur la présence de deux structures mutuellement exclusives, thermodynamiquement stables empêchant donc le changement de conformation spontané de l'ARN (Machtel *et al.*, 2016). La taille des *riboswitchs* est variable allant de 25 nucléotides pour le *riboswitch* pré-queuosine₁-I jusqu'à approximativement 200 nucléotides pour le *riboswitch* lysine (Lotz & Sues, 2018; Machtel *et al.*, 2016). Pour le moment, 55 classes de *riboswitchs* ont été validées expérimentalement dans la plupart des bactéries séquencées (Greenlee *et al.*, 2018; Kavita & Breaker, 2022; Salvail & Breaker, 2023).

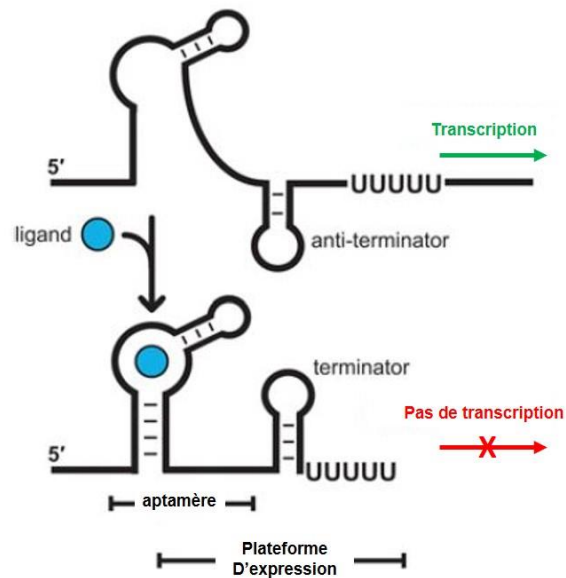


Figure 1.7 Représentation schématique d'un riboswitch agissant sur la transcription.

Les *riboswitchs* sont situés dans la région 5' non traduites des ARNm, ils sont composés de deux fonctionnels nommés aptamère, qui permet la liaison du ligand et plateforme d'expression, responsable de l'action du riboswitch via son changement de conformation. L'exemple ici montre une inhibition de la transcription, où en l'absence du ligand, représenté en bleu, il y a une structure anti-terminatrice permettant la transcription du gène en aval. Cependant après la liaison du ligand, il y a un changement de conformation de la plateforme d'expression qui se traduit par la formation d'un terminateur de transcription. Tiré de (Kim & Breaker, 2008).

Les *riboswitchs* chez les procaryotes peuvent agir à différents niveaux dans la cellule, ainsi il a été montré qu'ils pouvaient activer ou inhiber l'expression des gènes en agissant aussi bien au niveau de la transcription que de la traduction, mais également au niveau de la stabilité de l'ARNm (Machtel *et al.*, 2016).

1.3.1 Découverte des *riboswitchs*

La première notion de la présence de *riboswitch* provient d'une étude publiée en 2002 qui décrit un mécanisme d'atténuation de la transcription contrôlant la synthèse de la riboflavine (vitamine B2) et de la thiamine (vitamine B1) dans *B. subtilis* (Mironov *et al.*, 2002). Dans une étude précédente, la présence d'une structure en feuille de trèfle en amont de l'opéron permettant la biosynthèse de la riboflavine avait été mise en évidence et semblait avoir un effet sur la biosynthèse de la riboflavine (Kil *et al.*, 1992). Cette structure particulière est retrouvée très conservée dans plusieurs classes de bactéries (Gelfand *et al.*, 1999). Ainsi, plusieurs critères indiquent que cette structure semble être un élément ARN régulateur, comme le fait qu'il soit très conservé aussi bien au niveau de la structure que de la séquence, le fait qu'il soit retrouvé devant différents gènes (tous en rapport avec la riboflavine) dans différents génomes et le fait qu'il soit

sujet au transfert horizontal (Gelfand *et al.*, 1999). Cette structure sera par la suite appelée *rfn*-box (Mironov *et al.*, 2002). Il a été confirmé très peu de temps après que la *rfn*-box située en amont de l'opéron *ribDEAHT* chez *B. subtilis* permet la régulation de l'expression de cet opéron grâce à une interaction directe avec la molécule de Flavine MonoNucléotide (FMN) (Winkler *et al.*, 2002b).

Concernant la régulation de la thiamine, comme retrouvé pour la riboflavine, la présence d'une *thi*-box située en amont de différents opérons responsables de la biosynthèse de la thiamine a été observée (Miranda-Ríos *et al.*, 2001). Chez *E. coli*, la présence de ces *thi*-box en amont de trois opérons est impliquée dans la biosynthèse de la thiamine (*thiCEFSGH*, *thiMD* et *thiBPQ*). Ces mêmes opérons sont régulés de manière négative par la TPP (Miranda-Ríos *et al.*, 2001). Ainsi des mutations au niveau de ces structures ont montré une perte de la régulation en présence de TPP, suggérant un rôle important dans la régulation de ces gènes via la molécule de TPP (Miranda-Ríos *et al.*, 2001).

Ainsi au regard des connaissances déjà présentes, l'étude de 2002 (Mironov *et al.*, 2002) a montré que la régulation de ces gènes intervient au niveau de la transcription. Un terminateur de transcription est créé par l'interaction avec les molécules de FMN ou de TPP selon leur *riboswitch* respectif (Mironov *et al.*, 2002; Winkler *et al.*, 2002b). En effet, la liaison de ces molécules entraîne un changement de structure de l'ARN et sa fonction (Mironov *et al.*, 2002). Une autre étude publiée en 2002 a montré la présence de *thi*-box proche de la région de fixation du ribosome suggérant qu'un autre mode de régulation pouvait être possible par ces structures d'ARN à savoir au niveau de la traduction (Winkler *et al.*, 2002a). Depuis, de nombreux autres *riboswitchs* ont été découverts dans la plupart des génomes bactériens séquencés par des méthodes bioinformatiques qui seront décrites plus en détails dans la section 1.4.1.

1.3.2 Régulation de la transcription

1.3.2.1 Mécanisme Rho indépendant

Ce mode de régulation est très répandu (Machtel *et al.*, 2016). Dans ce cas, la liaison du ligand à l'aptamère entraînera un changement de conformation de la plateforme d'expression en la stabilisant en un terminateur de transcription ou une structure anti-terminatrice selon qu'il s'agisse d'un mécanisme d'inhibition ou d'activation (Machtel *et al.*, 2016). En effet, dans le cadre d'une inhibition de la transcription, en absence du ligand la plateforme d'expression forme une structure anti-terminatrice permettant le passage de l'ARN polymérase. Cependant, en présence du ligand,

il y aura un changement de conformation entraînant la formation d'un terminateur intrinsèque de transcription ce qui entrainera un arrêt précoce de cette dernière et comme la séquence codante ne sera pas transcrite, la protéine ne sera pas exprimée (Bédard *et al.*, 2020) (Figure 1.7).

Un exemple de *riboswitch* agissant par ce type de mécanisme, mais qui active l'expression plutôt que de la réprimer, est le *riboswitch* adénine présent devant le gène *ydhL* chez *B. subtilis* (Mandal & Breaker, 2004a). Il a été montré que le gène *ydhL*, aussi appelé *pbuE*, code pour une pompe à efflux de purines (Johansen *et al.*, 2003). Ainsi, en présence de fortes concentrations d'adénine le *riboswitch* permet une expression de cette pompe à efflux et donc une diminution de la concentration intracellulaire d'adénine. La structure en absence d'adénine présente une longue structure tige-boucle suivie d'une queue poly-U correspondant à un terminateur de transcription empêchant ainsi l'expression des gènes situés en aval (Mandal & Breaker, 2004a) (Figure 1.8). Cependant, la fixation de l'adénine provoque un changement de conformation qui déstabilise le terminateur formant ainsi une structure anti-terminatrice constituée par l'aptamère lui-même. Cette structure permettra l'expression du gène situé en aval (Mandal & Breaker, 2004a) (Figure 1.8).

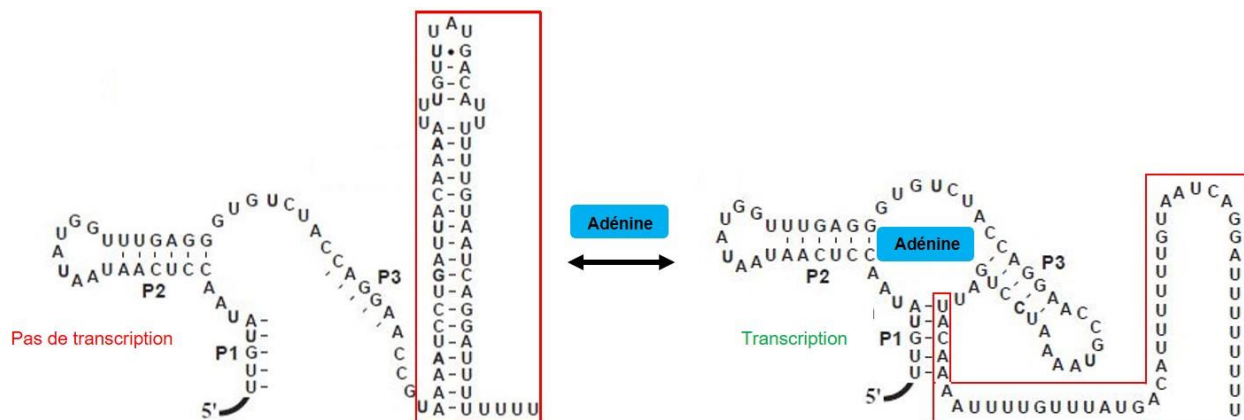


Figure 1.8 Exemple de régulation de la transcription par le *riboswitch ydhL* chez *B.subtilis*.

Les rectangles rouges représentent la plateforme d'expression du *riboswitch* et permet de visualiser le changement de conformation. Dans la conformation de gauche, il y a la présence d'un terminateur de transcription empêchant la suite de la transcription mais en présence d'adénine (conformation de droite) un des bras de cette tige-boucle sert à la liaison de la molécule défaisant cette structure permettant alors la transcription de l'ARNm. Adapté de (Mandal & Breaker, 2004a).

1.3.2.2 Mécanisme Rho dépendant.

Un autre mécanisme jouant sur la transcription est via une terminaison de transcription dépendante de la protéine Rho. Pour son action, Rho forme une structure en anneau comprenant six protéines qui se fixera à des sites nommés site *rut* (rho utilization) (Henkin & Peters, 2020). L'hexamère de Rho ne peut se fixer à son site *rut* seulement si celui-ci est accessible, par exemple il ne peut être occupé par des ribosomes (Henkin & Peters, 2020). Il est ainsi possible d'avoir deux types de régulation : directe ou indirecte (Bastet *et al.*, 2017). Dans le cas d'une régulation indirecte, c'est la séquestration du SD qui empêche la présence de ribosomes sur l'ARNm et donc le rend sensible à Rho en présence de son ligand (Bastet *et al.*, 2017) (Figure 1.9 B).

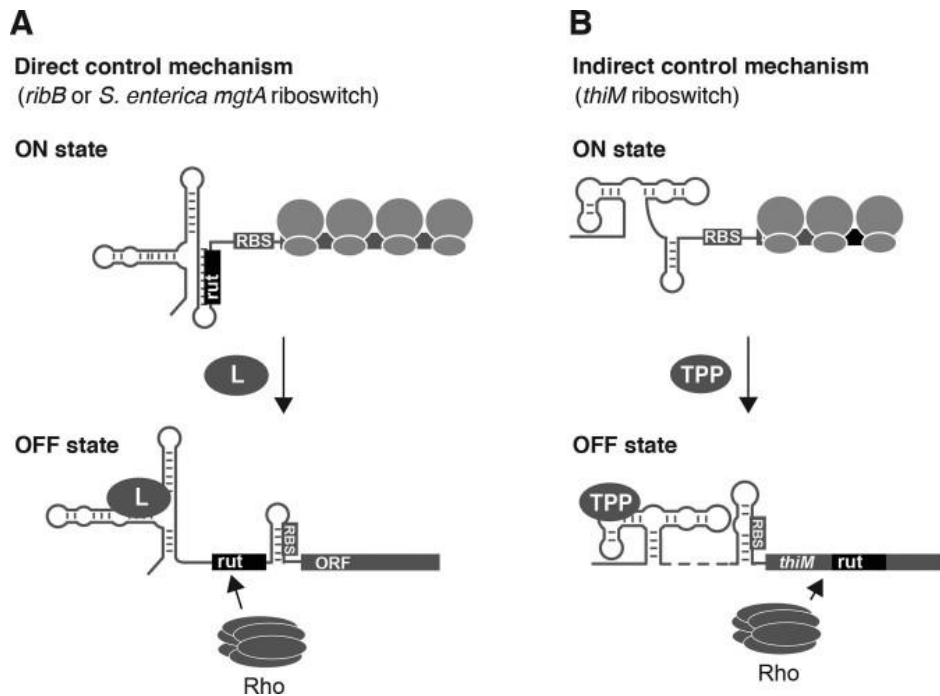


Figure 1.9 Mécanisme de régulation de la transcription dépendant de la protéine Rho.

La régulation de la transcription Rho dépendante peut se faire via un mécanisme direct ou indirect. **A** : Dans le cas du mécanisme direct, la présence du ligand, représenté par un rond gris avec un L, entraîne la libération du site *rut* permettant la liaison de l'hexamère de protéine Rho provoquant ainsi un arrêt de la transcription. **B** : Dans le cas du mécanisme indirect, ici avec l'exemple du *riboswitch thiM*, la liaison de la molécule de TPP entraîne un changement de conformation de la plateforme d'expression du *riboswitch* séquestrant alors le RBS. Les ribosomes ne pourront ainsi plus interagir avec ce dernier, donc ne protégeront plus le site de clivage *rut* présent dans le gène *thiM*. Cela entraîne donc un arrêt de la transcription de ce gène. Tiré de (Bastet *et al.*, 2017).

Un exemple de régulation directe par ce biais est le *riboswitch* magnésium en amont du gène *mgtA* chez *Salmonella enterica* serovar Typhimurium (Figure 1.9 A) (Hollands *et al.*, 2012). En effet, des études de la structure ont mis en évidence qu'il n'y avait pas de terminateur de transcription dans la partie 5' UTR de ce gène. Des essais de fusion avec *lacZ* réalisés avec l'antibiotique bicyclomycine, qui est un inhibiteur spécifique de Rho (Kohn & Widger, 2005), ont permis de montrer qu'une grande concentration en magnésium entraîne un arrêt de la transcription (Hollands *et al.*, 2012). En effet, la liaison du magnésium entraîne un changement de conformation de la plateforme d'expression libérant le site *rut*, le rendant ainsi accessible à Rho (Bastet *et al.*, 2017). Cependant, en présence de bicyclomycine, la régulation en présence de magnésium ne se fait plus suggérant donc un mécanisme de terminaison de la transcription Rho dépendant (Hollands *et al.*, 2012).

1.3.3 Régulation de la traduction

La régulation de l'expression des gènes médiée par les *riboswitchs* peut également se faire au niveau de la traduction (Machtel *et al.*, 2016). Le mécanisme observé dans ce cas ressemble beaucoup à ce que l'on peut trouver au niveau de la transcription à l'exception que le changement de conformation se produit cette fois au niveau de l'accessibilité du site de fixation du ribosome permettant la traduction de l'ARNm (Machtel *et al.*, 2016).

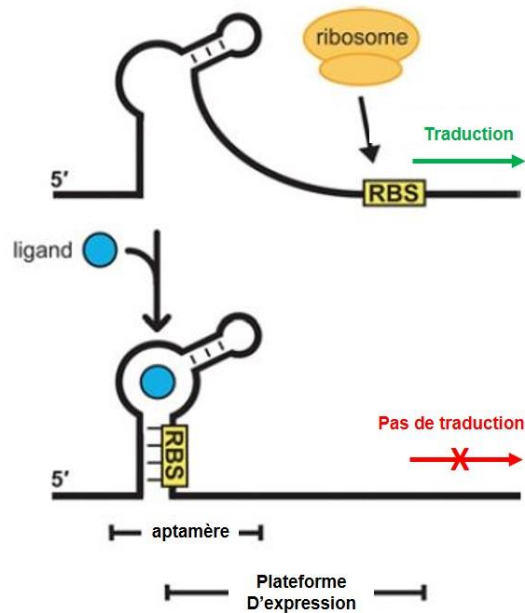


Figure 1.10 Mécanisme de régulation de la traduction via un exemple d'inhibition.

Le ligand est représenté par le rond bleu. En absence du ligand le site de fixation du ribosome (RBS) est disponible pour que la traduction puisse avoir lieu. En présence du ligand le changement de conformation de la plateforme d'expression entraîne la séquestration du site de fixation du ribosome (RBS) dans la structure double-brin empêchant la liaison de ce dernier afin de réaliser la traduction. Tiré de (Kim & Breaker, 2008).

En effet dans le cas d'une inhibition de la traduction, le SD sera en absence du ligand accessible pour le ribosome. Cependant, une fois le ligand fixé au domaine aptamère cela entraîne un changement de conformation de la plateforme d'expression séquestrant ainsi le SD qui le rend inaccessible pour le ribosome (Figure 1.10). C'est ce que l'on peut observer dans le cas du *riboswitch* SAM-II (Haller *et al.*, 2011).

1.3.4 Régulation au niveau de la stabilité de l'ARNm

Peu d'exemples sont connus pour ce type de régulation, mais c'est le cas du *riboswitch* lysine présent devant le gène *lysC* chez *E. coli* (Caron *et al.*, 2012). Dans des études précédentes, il avait été suggéré que la régulation de ce *riboswitch* se faisait via une modulation de la traduction (Rodionov *et al.*, 2003). Des essais dans des souches d'*E. coli* mutées au niveau de l'ARNase E ont été réalisés. Les souches *rne-131* conservant l'activité ribonucléolytique de l'ARNase E, mais ne pouvant pas former le dégradosome et *rne-3071* présentant une ARNase thermosensible à 44°C ont permis de montrer que le dégradosome ainsi que l'ARNase E sont impliqués dans la régulation par ce *riboswitch* (Caron *et al.*, 2012).

Ainsi en absence de lysine, la conformation de la plateforme d'expression a deux fonctions à la fois : elle permet la traduction du gène car le RBS est libre, mais elle bloque aussi un site de clivage de l'ARNase E (Figure 1.11) (Caron *et al.*, 2012). Cependant, après la liaison du ligand, le RBS est séquestré dans une structure double brin empêchant ainsi la traduction du gène *lysC*. De plus, le site de clivage de l'ARNase E se retrouve également accessible impliquant une dégradation rapide de l'ARNm (Caron *et al.*, 2012).

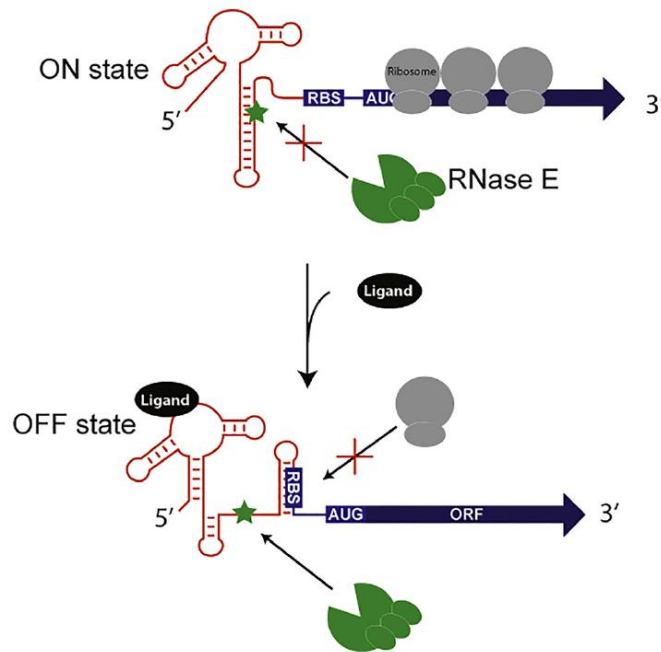


Figure 1.11 Mécanisme de régulation du *riboswitch lysC* chez *E. coli*.

En absence de ligand, le site de clivage de l'ARNase E (représenté par une étoile verte) est séquestré dans une structure double brin le rendant inaccessible pour l'ARNase E et le dégradosome (représenté en vert). Le RBS est également disponible pour permettre la fixation du ribosome et donc la traduction du gène. Cependant, en présence du ligand, le site de clivage se retrouve accessible pour l'ARNase E et le dégradosome. De plus, le RBS se retrouve bloqué dans une structure double brin empêchant ainsi la traduction du gène *lysC*. Tiré de (Caron *et al.*, 2012).

1.3.5 Différents types de ligands

Les *riboswitchs* peuvent répondre à plusieurs classes de ligands de natures différentes (McCown *et al.*, 2017). Ces ligands peuvent être des cofacteurs d'enzymes comme l'adénosylcobalamine (Nahvi *et al.*, 2002), la thiamine pyrophosphate (TPP) (Winkler *et al.*, 2002a), des dérivés de nucléotides comme la guanine (Mandal *et al.*, 2003), l'adénine (Mandal & Breaker, 2004a), des

molécules de signalisation comme le di-GMP-cyclique (di-GMPc) (Sudarsan *et al.*, 2008), des acides aminés comme la glycine (Mandal *et al.*, 2004) ou encore des ions tels que le nickel/cobalt (Furukawa *et al.*, 2015). Tous les ligands des *riboswitchs* connus sont cités dans le tableau 1.1.

Tableau 1.1 Liste des ligands connus pour les *riboswitchs*.

Cofacteurs d'enzymes	Dérivés de nucléotides	Molécules signal	Acides Aminés	Ions	Autre classe
Adénosylcobalamine	Guanine	c-di-GMP	Glycine	Magnésium	Glucosamine 6 phosphate
Aquacobalamine	Adénine	c-di-ADMP	Glutamine	Manganèse	Azaaromatique
Thiamine PyroPhosphate	Pré-Q1	c-AMP-GMP	Lysine	Fluor	Guanidine
HMP-PP	2'-dG	ZTP		Nickel/Cobalt	
Flavine MonoNucléotide	PRPP	ppGpp		Fer	
S-adénosylméthionine	(d)ADP/(d)CDP			Sodium	
S-adénosylhomocystéine	Xanthine / acide urique			Lithium	
Tétrahydrofolate					
Cofacteur Molybdène					
Cofacteur Tungstène					
NAD +					

Tableau récapitulatif de tous les ligands de riboswitchs découverts à ce jour (McCown *et al.*, 2017; Sherlock & Breaker, 2020; Sherlock *et al.*, 2018b; White *et al.*, 2022a; White *et al.*, 2022b; Xu & Cotruvo, 2022; Yu & Breaker, 2020).

1.4 Méthodes de découverte de nouveaux ARN non codants agissant en *cis*

1.4.1 Méthodes bioinformatique

La plupart des *riboswitchs* connus de nos jours ont été découverts grâce à des méthodes bioinformatiques basées sur des comparaisons des régions intergéniques à travers différentes classes de bactéries (Dar *et al.*, 2016). Dans les dernières décennies, plusieurs générations de méthodes ont été utilisées afin de rechercher des motifs d'ARNnc. Cependant, elles sont toutes

basées sur le même principe c'est-à-dire regarder les structures secondaires conservées dans des alignements de séquences (Yao *et al.*, 2007).

L'une d'entre elles est basée sur des recherches avec BLAST afin d'identifier les séquences intergéniques puis comparer les conservations de séquences en utilisant le package fasta (Barrick *et al.*, 2004). Enfin, les modèles intéressants étaient ensuite raffinés et recherchés à plus grande échelle en utilisant le programme Sequencesniffer (Barrick *et al.*, 2004). Ce programme pouvait causer des artefacts d'alignement en essayant de sur-optimiser le nombre de positions qui pourraient covariées entre les séquences représentant ainsi une limite de cet outil (Yao *et al.*, 2007).

Une autre méthode a ainsi été développée qui cette fois-ci pouvait chercher dans des séquences non alignées et utilise un nouvel outil nommé CMfinder (Yao *et al.*, 2006). CMfinder est un outil permettant de trouver des motifs d'ARN dans un ensemble de séquences non alignées avec une faible identité de séquence (Yao *et al.*, 2006). Le pipeline de cette méthode est présenté dans la figure 1.12 (Yao *et al.*, 2007).

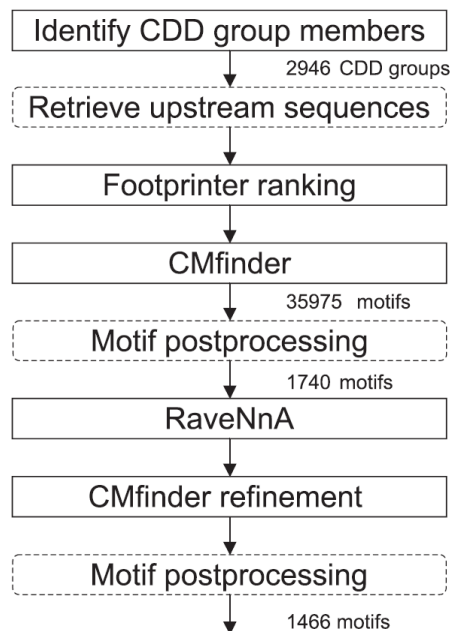


Figure 1.12 Pipeline bioinformatique pour la recherche de nouveaux ARN non codants.

Pipeline présenté pour la recherche de nouveaux ARNnc chez les procaryotes. Les rectangles pleins représentent des outils qui ont besoin de ressources computationnelles. Les étapes intermédiaires et/ou manuelles sont représentées en rectangles pointillés. Tiré de (Yao *et al.*, 2007).

Pour cette méthode, la première étape consiste à extraire, en utilisant la base de données Conserved Domain Database (CDD) de NCBI, un groupe de gènes homologues dans leurs fonctions puis d'en extraire les séquences situées dans les parties 5' non traduites (Yao *et al.*, 2007). La seconde étape consiste à utiliser l'outil Footprinter, qui est un outil de phylogénie d'ADN permettant de sélectionner dans les groupes de gènes ceux qui ont le plus de chance d'avoir des ARNnc. En effet, il a été montré que les *riboswitchs* présentent une conservation de séquence globale faible mais certaines parties sont hautement conservées (Yao *et al.*, 2007). Par la suite, les ensembles de séquences sont appliqués dans CMfinder qui permettra de rechercher des motifs d'ARN, un tri sera effectué après cette étape afin de déterminer les meilleurs candidats (Yao *et al.*, 2007). La quatrième étape consiste à regarder dans une base de données de génomes procaryotes la présence de ce motif en utilisant l'outil RaveNnA, le motif pourra ensuite être affiné à la suite des nouvelles instances en réutilisant CMfinder (Yao *et al.*, 2007). Pour terminer, des analyses manuelles au niveau du contexte génétique ainsi que ces potentielles fonctions sont effectuées (Yao *et al.*, 2007).

Comme vu précédemment, l'approche bioinformatique repose principalement sur la comparaison de séquences suffisamment semblables pour permettre de les regrouper en alignements de séquences conservées, mais également assez différentes pour permettre de déceler de la covariation. La covariation est la variation entre les deux nucléotides de l'alignement en même temps de façon à conserver une paire de base Watson-Crick (ou la paire de base compatible G-U) (Eckert & Weinberg, 2020). Lorsqu'un grand nombre de paires de base prédites covarient, le niveau de confiance accordé à la structure découverte est très élevé. En effet, si plusieurs paires de base covarient cela indique une certaine distance évolutive, et en plus si plusieurs positions dans les régions simples brin sont conservées, cela représente un bon candidat potentiel pour un ARNnc (Eckert & Weinberg, 2020). En effet, ces derniers requièrent un positionnement précis de groupements chimiques donnés pour former une pochette de liaison sélective pour leur ligand, ce qui implique la conservation de plusieurs bases à proximité de cette pochette de liaison.

1.4.2 Méthode basée sur la structure des ARN

Une autre méthode pouvant être utilisée afin de découvrir de nouveaux *riboswitchs* est également basée sur la conservation de structure des ARN, mais en prenant comme point de départ un ARN déterminé expérimentalement comme ayant des propriétés d'aptamère (Riccitelli & Lupták, 2010). Ainsi, une librairie de séquences est mise en contact avec un ligand spécifique. Les molécules interagissant avec ce ligand pourront être sélectionnées par une méthode de

SELEX (*Systematic Evolution of Ligand by Exponential Enrichment*) pour leur capacité à lier la molécule (Riccitelli & Lupták, 2010). Le SELEX consiste en la réalisation de plusieurs cycles (sélection, lavage, amplification) à partir d'une librairie complexe d'ADN ou d'ARN, afin de sélectionner les molécules ayant le plus d'affinité pour la molécule à tester (Tuerk & Gold, 1990). Quand une partie de la librairie montre l'activité escomptée celle-ci est séquencée afin de voir si un motif particulier domine dans la population (Riccitelli & Lupták, 2010). L'utilisation d'un descripteur, nommé RNABOB, permet de détailler la structure trouvée. Pour cela, la séquence est décrite en utilisant les codes de RNABOB, par exemple, pour une région simple brin celle-ci sera annotée « s » dans le descripteur (Riccitelli & Lupták, 2010). Une fois le modèle formé, il est possible de scanner les génomes complets afin de vérifier si la structure est présente (Riccitelli & Lupták, 2010).

L'une des limites de cette méthode est que même si, via le SELEX, une structure liant efficacement la molécule est trouvée, peut-être que cette structure ne se retrouve pas dans la nature. Une seconde limite concerne la sélection lors du SELEX, car le ligand est fixé sur une matrice solide limitant l'accessibilité de certains groupements pouvant interagir avec la librairie d'ARN. Cette méthode a été utilisée dans le but de trouver de nouveaux *riboswitchs* à partir d'un aptamère liant l'adénosine, décrit dans une étude précédente, afin de chercher des homologues de celui-ci chez différents organismes (Vu *et al.*, 2012). Des homologues ont ainsi été retrouvés notamment chez une bactérie (*Bacillus cytotoxicus*), chez des vertébrés (souris, ouistiti ...) et chez l'humain (Vu *et al.*, 2012). Des essais de cartographie *in vitro* par *in-line* ont permis de montrer une liaison entre l'ARN et des molécules d'ATP et d'AMP mais la fonction biologique reste encore inconnue (Vu *et al.*, 2012).

1.4.3 Term-seq

L'utilisation des méthodes de bioinformatique comporte certaines limites pour la découverte des nouveaux *riboswitchs*. Ainsi, une méthode expérimentale qui permettrait de découvrir des *riboswitchs* directement via le séquençage complet du transcriptome des bactéries ouvrirait la voie à de nouveaux motifs et c'est ce que le Term-seq propose (Dar *et al.*, 2016). Le Term-seq est une méthode expérimentale pour la découverte de nouveaux *riboswitchs* ou régulateurs utilisant un mécanisme de terminaison de la transcription dans des génomes bactériens (Dar *et al.*, 2016). Les bactéries sont, tout d'abord, cultivées dans une condition donnée. Puis les ARN sont séquencés grâce à un protocole de RNA-seq développé afin de séquencer directement les extrémités 3' des ARN dans les bactéries (Dar *et al.*, 2016). Cette méthode peut donc être

appliquée afin de découvrir des régulations impliquant un arrêt précoce de la transcription au niveau de régions 5' non traduites, comme le font plusieurs *riboswitchs* par exemple (Dar *et al.*, 2016).

Dans un premier temps, cette méthode a été appliquée sur *B. subtilis* cultivé en milieu riche (Dar *et al.*, 2016). Ceci a permis d'identifier 1443 extrémités d'ARN correspondant aux différents gènes de *B. subtilis*. Des analyses des lectures obtenues ont permis de mettre en évidence des *riboswitchs* déjà connus, comme TPP ou SAM (Dar *et al.*, 2016). Ainsi, 92% des *riboswitchs* agissant au niveau de la transcription connus chez *B. subtilis* ont été retrouvés via cette méthode, et 18 nouveaux candidats montrant un arrêt précoce de la transcription ont été découverts (Dar *et al.*, 2016).

Cependant cette méthode ne révèle pas le ligand responsable de cet arrêt de transcription précoce dans l'ARN, qui est une partie importante pour la caractérisation fonctionnelle des *riboswitchs*. Elle a donc été adaptée en comparant les résultats du Term-seq de la même bactérie cultivée en présence ou en absence du métabolite d'intérêt (Dar *et al.*, 2016). En effet, la présence du ligand change la conformation du régulateur et change ainsi le profil des ARN subissant un arrêt précoce de la transcription en présence du ligand. Cette approche a une nouvelle fois été testée avec *B. subtilis* qui a été cultivée avec ou sans lysine (Dar *et al.*, 2016). Ainsi seuls les *riboswitchs* lysine connus ont montré un changement d'expression dans les niveaux d'ARN et ceci est spécifique au ligand car aucun changement sur ces *riboswitchs* n'est observé en présence de méthionine (Dar *et al.*, 2016).

Une limite de cette méthode peut être qu'elle ne permet pas de trouver des ARN régulateurs présentant uniquement un mécanisme de régulation de la transcription, manquant alors les autres mécanismes.

1.4.4 La méthode PARCEL

PARCEL est l'acronyme pour *Parallel Analysis of RNA Conformations Exposed to Ligand binding*. Il s'agit d'une méthode expérimentale permettant de détecter les changements de conformation d'ARN en réponse à la présence d'un ligand via une analyse de séquençage à haut débit (Tapsin *et al.*, 2018). Pour cela, les ARN totaux de l'organisme d'intérêt sont extraits en enlevant par la suite les ARNr, qui sont présents en grand nombre dans la cellule. Les différentes structures du transcriptome sont, par la suite, analysées en présence ou en absence de différents ligands en utilisant la nucléase V1 clivant les structures double-brin (Tapsin *et al.*, 2018). Les fragments une fois obtenus sont clonés dans des bibliothèques d'ADN complémentaire aux ARN pour une analyse

par Illumina. Ainsi les différences de lectures obtenues entre les deux conditions montrent le changement de conformation qui se produit quand le ligand est présent (Figure 1.13) (Tapsin *et al.*, 2018).

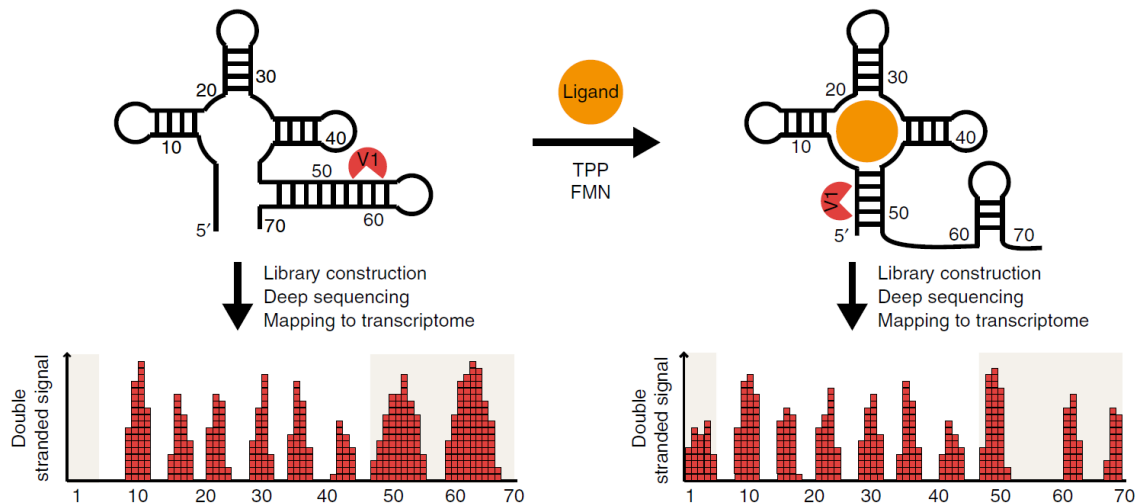


Figure 1.13 Schéma du principe de la méthode PARCEL.

Les ARN totaux des cellules sont extraits, puis mis en présence de ligand (ici TPP et FMN, représenté en orange). Ils sont ensuite traités avec la nucléase V1 (représenté en rouge) coupant au niveau des régions double brin de l'ARN. Les échantillons sont ensuite séquencés et les lectures alignées sur le transcriptome de l'organisme étudié. Les deux conditions, avec ou sans ligand, sont ensuite comparées. Ainsi si une différence est visible en présence versus en absence du ligand cela indique un changement de conformation dû à sa présence donc la présence d'un aptamère. Tiré de (Tapsin *et al.*, 2018).

Grâce à cette méthode, de nouveaux aptamères d'ARN ont été identifiés (Tapsin *et al.*, 2018). Notamment un ARN présent chez *B. subtilis* qui semble, à la suite de la liaison de FMN, réguler la maturation de l'ARN_{tm}. En effet, les changements de structure ne sont présents uniquement que dans les formes précurseurs, mais pas dans la forme mature de l'ARN_{tm} (Tapsin *et al.*, 2018).

1.4.5 SHAPE-MaP

Enfin, une autre méthode expérimentale pouvant être utilisée pour la découverte de nouveaux *riboswitchs* est le SHAPE-MaP (Siegfried *et al.*, 2014). Cette méthode est basée sur le SHAPE (Selective 2'-hydroxyl Acylation analyzed by Primer Extension), qui permet d'étudier les structures d'ARN grâce à des composés tel que le N-méthyl isatoïc anhydride (NMIA) faisant réagir les positions 2' hydroxyle des nucléotides flexibles de l'ARN (Weeks & Mauger, 2011). Cette méthode peut s'apparenter à la méthode PARCEL, décrite précédemment. En effet, ces deux méthodes

utilisent des agents chimiques / enzymatique afin d'étudier la structure de l'ARN en présence ou en absence de ligand via du séquençage à haut débit (Siegfried *et al.*, 2014; Tapsin *et al.*, 2018). L'acronyme MaP correspond à Mutational Profiling. Les nucléotides flexibles sont ceux présents dans des structures simples brins telles que les boucles (Weeks & Mauger, 2011). Les ARN modifiés par le SHAPE sont ensuite retro-transcrits dans des conditions qui permettent à la transcriptase inverse d'insérer des mutations en face de la modification induite par le SHAPE (Figure 1.14) (Siegfried *et al.*, 2014), en vue de la préparation au séquençage. Lors de l'analyse du séquençage, les « mutations » sont analysées et un profil des différentes mutations est réalisé en fonction des lectures obtenues sur la séquence de référence (Siegfried *et al.*, 2014). Finalement, le profil peut être utilisé pour faire un modèle de structure secondaire qui pourra être utilisé par la suite (Siegfried *et al.*, 2014).

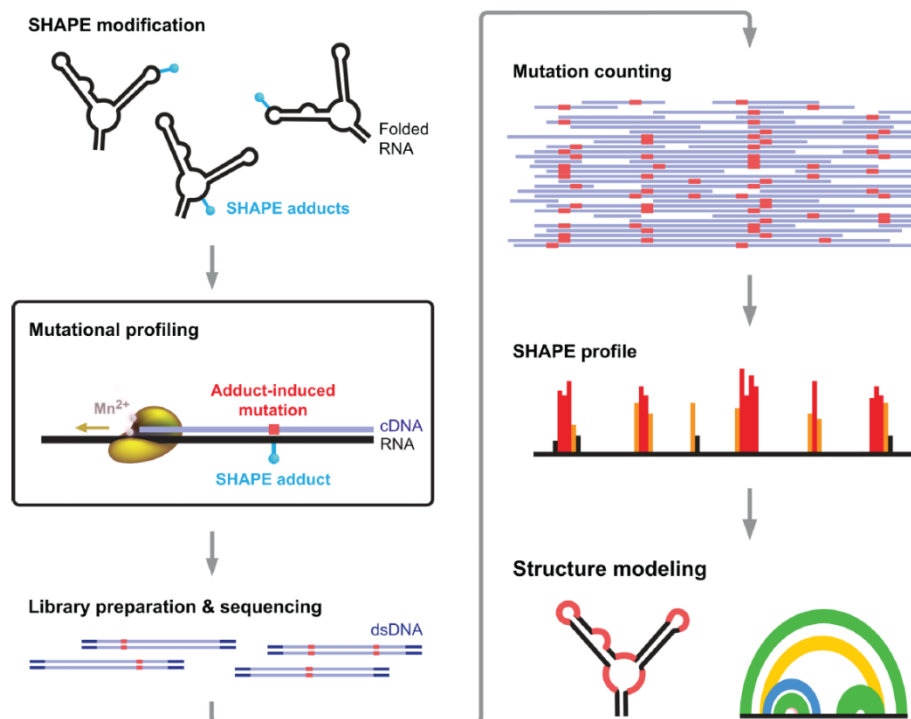


Figure 1.14 Schéma du principe du SHAPE-MaP.

Les ARN d'intérêt sont traités via un réactif de SHAPE, permettant d'ajouter des modifications sur les nucléotides libres. Ils sont ensuite retro-transcrits dans des conditions favorisant l'insertion d'un nucléotide différent à l'endroit où se trouve la modification induite par le SHAPE. Une librairie d'ADNc est ensuite synthétisée afin d'être séquençée. Les mutations obtenues seront ensuite comptées et alignées sur la séquence de référence permettant ainsi de créer la structure de l'ARN. Tiré de(Siegfried *et al.*, 2014).

Le SHAPE-MaP pourrait en principe être utilisé à grande échelle avec plusieurs milliers d'ARN en même temps (Zeller *et al.*, 2022). Pour cela une étape essentielle est la présence de code barre permettant d'identifier les différentes séquences testées lors du séquençage (Zeller *et al.*, 2022). Des conditions avec et sans le ligand d'intérêt peuvent ainsi être réalisées en comparant les profils de SHAPE entre les différentes conditions (Zeller *et al.*, 2022). En effet, des essais dans des plaques 96 puits peuvent permettre soit de tester différents ligands pour un même ARN ou alors de tester pour un même ligand des ARN différents. Des différences dans les profils des mutations indiquent des différences de structure entre les deux conditions, donc un changement de conformation qui est une caractéristique des *riboswitchs* (Zeller *et al.*, 2022). Cependant, les études réalisées avec le SHAPE-MaP n'ont pour le moment pas permis la découverte de nouveaux *riboswitchs*, celles-ci étaient davantage basées sur la découverte de nouveaux ligands synthétiques pour des *riboswitchs* connus (Zeller *et al.*, 2022).

1.5 Distribution des ARN non codants régulant en *cis*

Les T-box sont une classe d'ARNnc régulant en *cis* répandus chez les bactéries à Gram positif, notamment chez les Firmicutes qui possèdent plus de 1% de ces unités de transcription régulées par des T-box (Kreuzer & Henkin, 2018). Plus de 1000 classes de T-box annotées agissent au niveau de la transcription (Zhang, 2020). Une plus petite classe retrouvée chez les Actinobactéries permettent aussi une régulation au niveau de la traduction (Zhang, 2020).

Les *r-leader* sont présents dans différents phylum du règne des bactéries (Figure 1.15) (Fu *et al.*, 2013). Des comparaisons de structures des différents *r-leader* présents chez *E. coli* a permis de montrer que les *r-leader* les plus répandus sont ceux répondant aux protéines ribosomales L1, L10 et S2 (Figure 1.15) (Fu *et al.*, 2013). De plus, dans cette étude on peut également remarquer que tous les phylum bactériens comportent au moins un *r-leader* sauf les Acidobactéries (Eckert & Weinberg, 2020; Fu *et al.*, 2013).

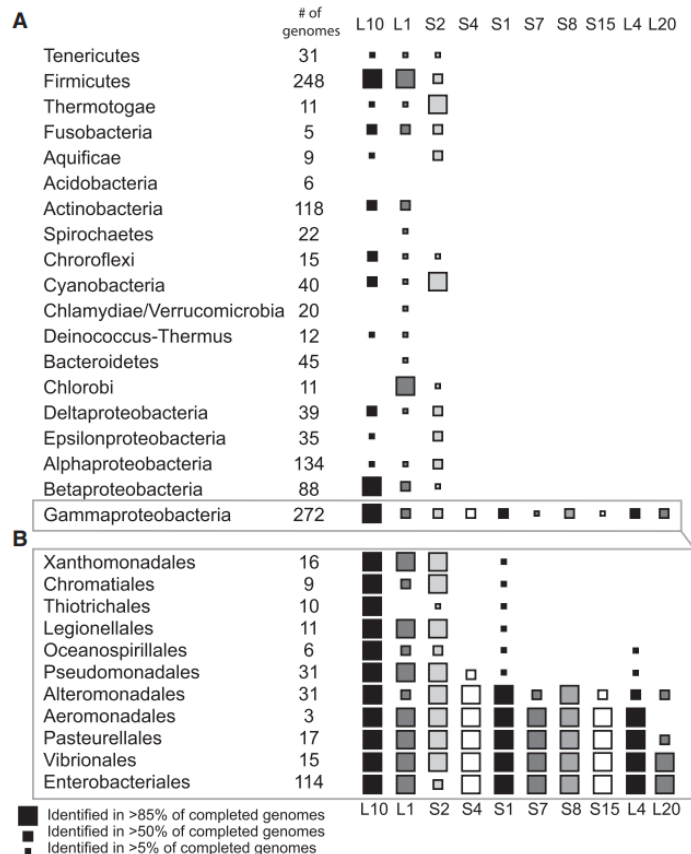


Figure 1.15 Distribution des *r-leader*.

A : Distribution des *r-leader* répondant aux protéines ribosomales L10, L1, S2, S4, S1, S7, S8, S15, L4 et L10 dans les différents phylums des bactéries. **B** : Distribution des *r-leader* dans les différents ordres des γ -protéobactéries. La taille des carrés représente le nombre de génomes du phylum ou de la famille dans lequel ce *r-leader* est retrouvé. Tiré de (Fu *et al.*, 2013).

Enfin, la présence d'une grande variété de *riboswitchs* a été montrée dans la plupart des génomes bactériens séquencés, incluant les bactéries bien étudiées comme *B. subtilis* ou *E. coli* (Breaker, 2011). Un grand nombre de *riboswitchs* a aussi été annoté chez les archées, notamment le *riboswitch* fluor (Speed *et al.*, 2018). Les *riboswitchs* sont retrouvés dans les trois domaines de la vie à savoir chez les procaryotes, les eucaryotes et les archées (Mironov *et al.*, 2002; Sudarsan *et al.*, 2003a). Cependant, le seul exemple connu jusqu'à présent de *riboswitch* chez les eucaryotes est le *riboswitch* TPP. Sa présence a été démontrée chez les plantes, comme *Arabidopsis thaliana*, ou chez les champignons comme *Neurospora crassa* (Sudarsan *et al.*, 2003a). Certains *riboswitchs* sont plus répandus que d'autres (McCown *et al.*, 2017). Les *riboswitchs* TPP et adénosylcobalamine sont les plus répandus, on les retrouve dans 6201 et 5174 espèces respectivement (Pavlova *et al.*, 2019). On peut aussi distinguer des différences de

représentation des *riboswitchs* en fonction du phylum, en effet les firmicutes et les protéobactéries présentent beaucoup plus de *riboswitchs* prédits que d'autres phylum comme les acidobactéries par exemple (McCown *et al.*, 2017).

2 PROBLÉMATIQUE, HYPOTHÈSE ET OBJECTIFS

Depuis de nombreuses années, des études ont permis de montrer l'importance des ARNnc (ARN rég, *riboswitchs*, thermorégulateurs, ribosomal leader ...) dans le métabolisme des bactéries aussi bien pour leur survie (Lalaouna *et al.*, 2019a; Massé *et al.*, 2005; Pavlova *et al.*, 2019) que pour leur pathogénicité (McKee *et al.*, 2018; Renzoni *et al.*, 1999).

La plupart des ARNnc agissant en *cis* découverts jusqu'à présent l'ont été grâce à des méthodes de bioinformatique (décrites dans la section 1.4.1). Certes, ces outils ont été très puissants pour découvrir de multiples candidats d'ARNnc (Barrick *et al.*, 2004; Weinberg *et al.*, 2007; Weinberg *et al.*, 2017a; Weinberg *et al.*, 2010) menant à la découverte de nombreuses classes de *riboswitchs* comme le *riboswitch* glycine (*gcvT*) (Mandal *et al.*, 2004), le *riboswitch* fluor (*crcB*) (Baker *et al.*, 2012) ou encore les *riboswitch* guanidine I à IV (Nelson *et al.*, 2017; Salvail *et al.*, 2020; Sherlock & Breaker, 2017; Sherlock *et al.*, 2017). Cependant, ces techniques présentent certaines limites pour la découverte de nouveaux motifs, comme les problèmes d'annotation de gènes ou la nécessité d'avoir plusieurs séquences à comparer pour la covariation. Afin de sélectionner les candidats à tester expérimentalement une analyse humaine est nécessaire mais celle-ci a tendance à mettre de côté les motifs les plus simples ou les moins répandus présentant une autre limite des outils de bioinformatique. Plusieurs études de Weinberg et collaborateurs ont permis de mettre en évidence plus de 300 motifs qui pour la plupart ne sont pas caractérisés (Weinberg *et al.*, 2007; Weinberg *et al.*, 2017a; Weinberg *et al.*, 2010). De nombreux ARNnc agissant en *cis* ont été découverts et caractérisés dans des bactéries disposant de nombreux outils génétiques, comme *E. coli* ou *B. subtilis* (Nahvi *et al.*, 2004; Winkler *et al.*, 2002b). Cependant, de nombreux ARNnc agissant en *cis* ont été prédits dans des bactéries présentant moins d'outils génétiques, comme *Pseudomonas* et *Burkholderia*. Ainsi, peu de ces « ARNnc putatifs » ont été caractérisés, notamment par le fait d'une disponibilité plus faible en outils génétiques dans ces bactéries (Munyati-Othman *et al.*, 2021; Reiss & Strobel, 2017).

Pour pallier également aux limites des outils de bioinformatique, des méthodes expérimentales ont été développées par plusieurs groupes comme le Term-seq (Dar *et al.*, 2016) et PARCEL (Tapsin *et al.*, 2018), mais ces méthodes présentent aussi des limites déjà citées. Nous avons donc développé au sein du laboratoire deux outils dans le but de découvrir et caractériser de nouveaux ARNnc en *cis*. Tout d'abord, un plasmide compatible avec les bactéries des genres *Pseudomonas* et *Burkholderia* nommé pVK-f2-lux (Korniakova *et al.*, 2021). Ce dernier permet de cloner en fusion traductionnelle, avec l'opéron *luxCDABE*, des ARNnc candidats présumés

régulateurs en *cis* afin de pouvoir étudier les variations d'expression de la luminescence dans plusieurs conditions (Korniakova *et al.*, 2021). Le second outil est également une nouvelle méthode expérimentale, appelée le Shifted-Reverse PAGE, basée sur la capacité des *riboswitchs* à changer de conformation en présence de leur ligand, ce qui provoque un changement de migration dans un gel de polyacrylamide natif (Devinck *et al.*, 2022).

Le but de ce projet de doctorat est d'optimiser et d'utiliser ces outils de manière à découvrir et caractériser de nouveaux ARNnc de type *riboswitch* ou autre type d'ARNnc agissant en *cis* dans les bactéries. Pour cela, nous avons émis deux hypothèses. La première hypothèse est que nous pourrions utiliser le plasmide pVK-f2-lux afin de combler le manque dans la caractérisation d'ARNnc dans des genres bactériens disposant de peu d'outils génétiques comme *Pseudomonas* et *Burkholderia*. Les ordres des Pseudomonadales et Burkholderiales présentent un grand nombre d'espèces et de souches séquencées (environ 1000). Ainsi, dans le cas des méthodes de découverte de nouvelles structures d'ARN conservées, de nombreux motifs candidats peuvent être trouvés. Cependant, ceux-ci demeurent non caractérisés dû au peu d'outils génétiques disponibles. Notre seconde hypothèse est que nous pourrions utiliser le SR-PAGE afin d'étudier la présence de structures d'ARN dans une grande variété d'organismes (bactéries, champignons, ADN humain ou encore des métagénomes) présentant différents niveaux de découverte de ces ARNnc en *cis*. Cette approche est également intéressante pour les organismes ayant de nombreux *riboswitchs* prédits et caractérisés. En effet, des études basées sur la prévalence des classes de *riboswitchs* suggèrent que leur découverte suit une distribution en loi de puissance (McCown *et al.*, 2017). Dans ce cas, il serait possible qu'il reste encore plus de 1000 classes de *riboswitchs* à découvrir (McCown *et al.*, 2017).

Ainsi, les objectifs de ce projet de recherche seront détaillés en deux chapitres. Un premier objectif consistera en la caractérisation de la fonction d'une vingtaine de candidats d'ARNnc régulant en *cis* présents chez les bactéries du genre *Pseudomonas* et *Burkholderia*. Des essais de luminescence seront réalisés dans de nombreuses conditions afin de caractériser ces candidats. Un second objectif consistera au développement de la méthode de SR-PAGE qui passera par des étapes d'optimisation et de validation de la méthode ainsi que de son application dans le but de découvrir de nouveaux *riboswitchs*.

3 CARACTÉRISATION DE NOUVEAUX ARN NON CODANTS EN CIS CHEZ LES BACTÉRIES DU GENRE *PSEUDOMONAS* ET *BURKHOLDERIA*

Contribution :

Le plasmide pVK-f2-lux utilisé dans ce chapitre a été conçu par Vesta Korniakova, étudiante à la maîtrise au sein du laboratoire. J'ai effectué des expériences qui ont servi pour les révisions de l'article de Vesta publié dans *Frontiers in Genetics* (Korniakova V, [Devinck A](#), Groleau MC, Déziel E & Perreault J (2021) Fluoride-Controlled Riboswitch-Based Dampening of Gene Expression for Cloning Potent Promoters. *Front Genet* 12:591543).

Concernant ce chapitre, où le plasmide en question a été utilisé pour caractériser des ARN qui n'ont rien à voir avec les *riboswitchs* dans la publication ci-haut mentionnée, j'ai fait toutes les étapes de design, de clonage des structures d'ARN, de design expérimental et d'optimisation ainsi que les essais de luminescence et de validation du rôle « cis-régulateur » des ARN en question. Pareillement, j'ai réalisé la cartographie des structures d'ARN *rnf* et *Pseudomon-GGDEF*.

Une partie des résultats de ce chapitre a été obtenue avec l'aide de Flavie Charbonneau et Hyerin Kim, des stagiaires de premier cycle lors de l'été 2022.

3.1 Mise en contexte

Plus de 300 ARNnc potentiels ont été découverts via des méthodes de bioinformatique (Weinberg *et al.*, 2007; Weinberg *et al.*, 2017a; Weinberg *et al.*, 2010) présents dans une grande diversité d'organismes. Cependant, afin d'être considérés comme un ARNnc fonctionnel ces candidats doivent être caractérisés. Nous avons développé au sein du laboratoire un rapporteur d'expression spécifique pour les bactéries des genres *Pseudomonas* et *Burkholderia* dans le but de caractériser certains des ARNnc candidats présents dans ces études.

3.1.1 Organismes d'études

3.1.1.1 *Pseudomonas*

Les *Pseudomonas* sont des bactéries à Gram négatif retrouvées communément dans l'environnement, notamment dans le sol, les marais et sur les plantes (Lyczak *et al.*, 2000; Stover

et al., 2000). Elles peuvent agir également en tant que pathogènes opportunistes dans certaines conditions (Lyczak *et al.*, 2000).

3.1.1.2 *Pseudomonas aeruginosa*

Pseudomonas aeruginosa est l'un des pathogènes de ce genre bactérien. La première infection causée par *P. aeruginosa* a été reportée en 1862, par l'observation de bacilles dans du pus d'une couleur bleu-grise dû à la présence de pyocyanine (Lyczak *et al.*, 2000). Elle est particulièrement retrouvée dans des infections du sang chez les grands brûlés, du tractus urinaire chez les patients avec des cathéters mais également dans des infections pulmonaires chez des patients sous respirateurs artificiels (Stover *et al.*, 2000). Elle est particulièrement dangereuse chez les patients atteints de fibrose kystique, représentant une des causes de mortalité de cette maladie (Stover *et al.*, 2000).

C'est une bactérie présentant un génome de 6,3 millions de bases (Mb) avec 5 570 cadres de lecture prédits (Stover *et al.*, 2000). De manière intéressante, il a été montré que plus de 9% des cadres de lectures prédits étaient des régulateurs de transcription et des systèmes à deux composants rendant cette bactérie complexe au niveau de sa régulation et capable de répondre à de large gamme de conditions environnementales (Stover *et al.*, 2000). Elle code également pour de nombreux facteurs de virulence comme des lipopolysaccharides, des flagelles, des pili de type IV, des systèmes de sécrétion de type III (Rocha *et al.*, 2019). Lors d'infection, tous ces facteurs de virulence agissent de différentes manières (Rocha *et al.*, 2019). En effet, certains de ces facteurs seront responsables davantage d'une infection aiguë alors que d'autres entraîneront une infection chronique (Jurado-Martín *et al.*, 2021). L'une des caractéristiques importante chez *P. aeruginosa* est également la présence de plusieurs systèmes de *quorum sensing* permettant la communication entre les cellules suite à des changements dans les conditions environnementales (Jurado-Martín *et al.*, 2021). En effet, elle possède quatre systèmes qui sont interconnectés: Las, Rhl, Pqs et Iqs (Jurado-Martín *et al.*, 2021). Ces systèmes sont particulièrement importants pour sa virulence et sa pathogénicité (Jurado-Martín *et al.*, 2021). Il a été montré que l'absence d'un ou plusieurs de ces facteurs entraîne une diminution significative de la virulence (Smith & Iglewski, 2003).

P. aeruginosa présente également une résistance naturelle aux antibiotiques la rendant difficile à traiter (Stover *et al.*, 2000). Celle-ci est notamment due à une faible perméabilité de la membrane limitant ainsi l'entrée de l'antibiotique dans la bactérie (Lambert, 2002). La présence de pompes à efflux particulièrement efficaces est aussi un mécanisme lui permettant de se

protéger (Lambert, 2002). De plus, l'usage des antibiotiques entraîne une pression favorisant d'autant plus la résistance par le développement de nouveaux mécanismes (Lambert, 2002). En effet, des mutations dans la sous-unité GyrA de l'ADN gyrase a permis une résistance aux fluoroquinolones (Mérens *et al.*, 2011). Ainsi, la découverte de nouveaux éléments de régulation comme des *riboswitchs* fournit de nouvelles cibles potentielles pour la recherche de nouveaux antibiotiques. Certains *riboswitchs*, comme FMN ou TPP, ont déjà été ciblés par de nouvelles classes de potentiels antibiotiques (Chakravarty & Massé, 2019).

3.1.1.3 *Pseudomonas syringae* pathovar (pv) *tomato* DC3000

Pseudomonas syringae pv. *tomato* DC3000 est un phytopathogène causant la gale bactérienne de la tomate et elle est également pathogène de l'organisme modèle des plantes *Arabidopsis thaliana* (Buell *et al.*, 2003)

Son génome d'une taille de 6,5 Mb contient un chromosome circulaire ainsi que deux plasmides représentant 5 763 cadres de lecture (Buell *et al.*, 2003). Son caractère pathogène dépend de la présence d'un système de sécrétion de type III pour permettre l'injection de facteur de virulence au sein de l'hôte (Buell *et al.*, 2003). Elle présente également certains éléments permettant de s'adapter à son environnement comme le sidérophore Pvd (pyoverdine), permettant de répondre à une limitation en fer (Jones & Wildermuth, 2011).

3.1.1.4 *Burkholderia*

Les *Burkholderia* sont des bactéries à Gram négatif retrouvées dans plusieurs niches écologiques (Eberl, 2006). Le genre comporte plus de 30 espèces dont la majorité sont des pathogènes humains (Eberl, 2006). Certaines de ces espèces sont des pathogènes opportunistes notamment pour les patients immunodéprimés ou souffrant de fibrose kystique (Eberl, 2006).

Ce genre comprend également des pathogènes dits primaires, c'est-à-dire capable d'infecter l'hôte quelles que soient ces défenses immunitaires, comme *Burkholderia pseudomallei* (*B. pseudomallei*) et *Burkholderia mallei* (*B. mallei*) (Baudart & Paniel, 2014; Eberl, 2006). Cette première bactérie est responsable de la mélioïdose, qui est une infection souvent fatale et endémique dans les pays de l'Asie du Sud Est ou dans le Nord de l'Australie (Brett *et al.*, 1998). La seconde est responsable de la morve, une maladie touchant initialement les animaux tels que les chevaux, mais qui peut contaminer également les humains (Eberl, 2006). Ces deux bactéries ont été placées par le Centre de Contrôle des Maladies (CDC) sur la liste des agents de menace biologiques présentant un risque pour être utilisé comme outils de bioterrorisme (Eberl, 2006).

Des études phylogénétiques sur différentes espèces de *Burkholderia* ont permis de montrer que ces deux bactéries (*B. pseudomallei* et *B. mallei*) étaient très proches de l'espèce *B. thailandensis* qui n'est pas pathogène chez l'humain (Brett *et al.*, 1998). *B. thailandensis* comporte deux chromosomes de 3,8 Mb et 2,9 Mb respectivement, avec 5 645 cadres de lecture annotés (Kim *et al.*, 2005). De manière intéressante, le système de sécrétion de type III nommé Bsa-TTSS, un facteur de virulence important chez *B. pseudomallei*, est hautement conservé chez *B. thailandensis* (Haraga *et al.*, 2008). Cette ressemblance rend *B. thailandensis* l'un des modèles de choix pour l'étude de ces pathogènes, qui nécessite normalement pour leur culture un laboratoire de niveau de confinement 3 (Haraga *et al.*, 2008). Ainsi des études sur cette bactérie permettra d'en découvrir davantage sur ces deux organismes hautement pathogènes et dangereux.

3.1.2 ARN non codants chez nos organismes d'études

3.1.2.1 *Pseudomonas*

Les gènes de virulence de *P. aeruginosa* sont contrôlés par des mécanismes extrêmement complexes impliquant des circuits de régulation et de multiples cascades de signalisation (Pita *et al.*, 2018). Des études ont permis de mettre en évidence lors de l'infection de l'hôte l'expression de 233 ARN rég dans notre souche d'intérêt, à savoir *P. aeruginosa* PA14 (Pita *et al.*, 2018; Pusic *et al.*, 2021). Cependant, seulement quelques exemples ont été caractérisés et sont impliqués dans divers processus cellulaires comme la virulence, le métabolisme du fer, le métabolisme de l'azote, le quorum sensing et la sensibilité aux antibiotiques (Tableau 3.1) (Pusic *et al.*, 2021).

Tableau 3.1 Seulement quelques ARNrég ont été caractérisés chez *P. aeruginosa* dans des processus cellulaires différents.

Virulence / biofilm	Métabolisme du fer	Métabolisme de l'azote	Quorum Sensing	Sensibilité aux antibiotiques
ErsA	PrrF 1/2	Pail	PhrS	ErsA
Sr0161	PrrH	NrsZ	PrrF 1/2	Sr0161
Sr006			ReaL	Sr006
PrrF 1/2				
PrrH				
ReaL				
Pail				
NrsZ				

Tableau récapitulatif des ARNrég caractérisés chez *P. aeruginosa* impliqués dans différents processus cellulaires (Pusic *et al.*, 2021).

Il est possible de remarquer que la plupart de ces ARNrég sont présents dans différents processus cellulaires en fonction de la cible de l'ARNrég. En effet, des exemples que l'on peut mentionner sont les ARN PrrF1/2 qui sont impliqués dans la virulence de *P. aeruginosa* ainsi que dans le métabolisme du fer ou encore le quorum sensing (Pusic *et al.*, 2021). Ces ARN ont une fonction similaire à l'ARNrég RyhB chez *E. coli* permettant de maintenir l'homéostasie du fer en condition de limitation en fer en réprimant la synthèse de protéines non essentielles contenant du fer (Pusic *et al.*, 2021). Ils sont également impliqués dans la répression du catabolisme de l'antranilate entraînant l'expression du Pseudomonas quinolone system (PQS), une molécule du quorum sensing (Pusic *et al.*, 2021). Ainsi, il est possible en voyant les différentes cibles de certains ARNrég de déterminer que ces processus cellulaires sont interconnectés. Un autre exemple que l'on peut mentionner est ReaL, qui est présent chez *P. aeruginosa* PA01 et PA14, il a été montré comme ayant une action pléiotropique sur la virulence notamment par la régulation du quorum sensing (Carloni *et al.*, 2017; Pusic *et al.*, 2021).

Des ARNnc agissant en *cis* sont également retrouvés dans cet organisme. Plusieurs classes ont été découvertes comme des *riboswitchs*, des thermorégulateurs ou des *r-leader* (Kalvari *et al.*, 2021). Des études ont permis de montrer un lien entre la virulence et la température chez *P. aeruginosa* (Grosso-Becerra *et al.*, 2014). La présence d'un thermorégulateur ROSE en amont du gène *rhIA* permet d'augmenter la virulence à 37°C grâce à plusieurs facteurs comme une

production plus importante de rhamnolipides ou en augmentant la transcription du gène rhIR impliqué dans le quorum sensing (Grosso-Becerra *et al.*, 2014). Plusieurs *riboswitchs* sont également présents chez *Pseudomonas* comme les *riboswitchs* fluor, manganèse ou guanidine-II (Chakravarty & Massé, 2019; Kalvari *et al.*, 2021; Reiss & Strobel, 2017). Une étude a été réalisée afin de rechercher des *riboswitchs* dépendant du quorum sensing par la méthode de Term-seq (méthode présentée dans la section 1.4.3) (Thomason *et al.*, 2019). Néanmoins, aucun *riboswitch* n'a pu être trouvé via cette méthode (Thomason *et al.*, 2019). Cependant, le Term-seq ne permet d'identifier que les mécanismes d'arrêt précoce de la transcription limitant ainsi l'étude (Dar *et al.*, 2016).

3.1.2.2 *Burkholderia*

Les bactéries du genre *Burkholderia* possèdent également plusieurs classes d'ARNnc comme des ARNrég, des *riboswitchs* et des thermorégulateurs (Kalvari *et al.*, 2021). Concernant les ARNrég, une étude de puce à ADN a permis de mettre en évidence 60 ARNrég différemment exprimés dans divers milieux (Stubben *et al.*, 2014). Ceux-ci agissent aussi bien en *cis* qu'en *trans* dans différents métabolismes (Stubben *et al.*, 2014).

Concernant les ARNnc régulant en *cis*, l'implication du *riboswitch* glycine a été démontrée chez *B. thailandensis* dans la détoxification de la glycine (Munyati-Othman *et al.*, 2021). En effet, les résultats de cette étude montrent que le *riboswitch* permet de ressentir de fortes concentrations de glycine activant ainsi l'expression du système de détoxification (Munyati-Othman *et al.*, 2021). La présence de *riboswitch* di-GMPc a également été montrée comme ayant une importance pour sa mobilité et sa virulence (Wang *et al.*, 2022).

Chez les organismes très étudiés comme *E. coli* ou *B. subtilis*, de nombreux ARNnc ont été découverts et caractérisés dans différentes voies du métabolisme (Boutet *et al.*, 2022). Nous avons ainsi voulu comparer le niveau de caractérisation des ARNnc agissant en *cis* entre *E. coli* et nos organismes d'études *P. aeruginosa* et *B. thailandensis* (Tableau 3.2). Pour cela, tous les ARNnc en *cis* annotés dans ces trois organismes ont été extraits grâce à la base de données Ribogap (Naghdhi *et al.*, 2017). Puis une étude bibliographique a été réalisée afin vérifier la caractérisation de chacun de ces ARN.

Tableau 3.2 Les ARNnc régulant en *cis* sont plus caractérisés chez *E. coli* que chez nos organismes d'études : *P. aeruginosa* et *B. thailandensis*.

	<i>E. coli</i> K12 MG1655	<i>P. aeruginosa</i> PA14	<i>B. thailandensis</i> E264	Références
<i>Riboswitchs</i>			Glycine RF00504	(Mandal <i>et al.</i> , 2004)
			ZMP/ZTP RF01750	(Kim <i>et al.</i> , 2015)
		SAH RF01057	SAH RF01057	(Wang <i>et al.</i> , 2008)
		Fluor RF01734	Fluor RF01734	(Baker <i>et al.</i> , 2012)
	FMN RF00050	FMN RF00050	FMN RF00050	(Vitreschak <i>et al.</i> , 2002)
	TPP RF00059	TPP RF00059	TPP RF00059	(Winkler <i>et al.</i> , 2002a)
	yybP-ykoY RF00080	yybP-ykoY RF00080	yybP-ykoY RF00080	(Price <i>et al.</i> , 2015)
	Cobalamine RF00174	Cobalamine RF00174	Cobalamine RF00174	(Nahvi <i>et al.</i> , 2002)
	Guanidine-II RF01068	Guanidine-II RF01068	Guanidine-II RF01068	(Sherlock <i>et al.</i> , 2017)
	Lysine RF00168			(Sudarsan <i>et al.</i> , 2003b)
	Molybdenum (MoCo) RF01055			(Regulski <i>et al.</i> , 2008)
Magnésium RF01056			(Groisman <i>et al.</i> , 2006)	
Thermorégulateurs		PA5194 RF02813		(Delvillani <i>et al.</i> , 2014)
		ROSE like RF02758, RF02759		(Grosso-Becerra <i>et al.</i> , 2014)
	ROSE_2 RF01832	ROSE_3 RF02523		(Nocker <i>et al.</i> , 2001)
	cspA RF01766	cspA RF01766	cspA RF01766	(Giuliodori <i>et al.</i> , 2010)
	trxA RF02773			(Righetti <i>et al.</i> , 2016)
	cpoB/ybgF RF02782			(Righetti <i>et al.</i> , 2016)

	htrA RF02700			(Klinkert <i>et al.</i> , 2012)
<i>r-leader</i>		rpsL leader RF01773		(Naville & Gautheret, 2010)
	S15 leader RF00114	S15 leader RF00114		(Benard <i>et al.</i> , 1996)
	alpha_RBS RF00140	alpha_RBS RF00140	alpha_RBS RF00140	(Schlax <i>et al.</i> , 2001)
			L19 leader RF00556	(Eckert & Weinberg, 2020)
Acide aminés <i>leader</i>	Tryptophane RF00513	Tryptophane RF00513		(Babitzke & Gollnick, 2001)
	Thréonine RF00506	Thréonine RF00506	Thréonine RF00506	(Lynn <i>et al.</i> , 1987)
	Leucine RF00512	Leucine RF00512	Leucine RF00512	(Grandoni <i>et al.</i> , 1993)
	Histidine RF00514			(Chan & Landick, 1989)
	Phénylalanine RF01859			(Gavini & Pulakat, 1991)
<i>Leader</i>	rnk_leader RF01771	rnk_pseudo RF01772	rnk_pseudo RF01772	(Naville & Gautheret, 2010)
	greA RF01769			(Potrykus <i>et al.</i> , 2010)
	rimP RF01770			(Yoshikazu <i>et al.</i> , 1985)
	pyrC_leader RF02372			(Wilson <i>et al.</i> , 1992)
	pyrD_leader RF02373			(Jensen, 1989)
<i>Regulatory element</i>	rne5 RF00040			(Diwa <i>et al.</i> , 2000)
	rncO RF00552			(Matsunaga <i>et al.</i> , 1996)
	icsRS RF01517			(Desnoyers <i>et al.</i> , 2009)
	SECIS_2 RF01988			(Mix <i>et al.</i> , 2007)

<i>Frameshift element</i>	DnaX RF00382			(Larsen <i>et al.</i> , 1997)	
ARN non caractérisés	RtT RF00391			(Bösl & Kersten, 1991)	
	ClpQY_prom RF02401			(Lien <i>et al.</i> , 2009)	
	Rev39_5p_UTR RF02744			(Kwenda <i>et al.</i> , 2016)	
	JUMPstart RF01707 nuoG RF01748			(Weinberg <i>et al.</i> , 2010)	
		PhotoRC-I RF01716 flg-Rhizobiales RF01736 Pseudomon-Rho RF01720 Pseudomon-groES RF01721 gyrA RF01740 rmf RF01755 rneII RF01756 sucA-II RF01758 sucC RF01759 Pseudomon-GGDEF (no Rfam name, RNA possible) ° fixA (no Rfam name, RNA possible) °			(Weinberg <i>et al.</i> , 2010)
		Rev39_5p_UTR RF02744			(Kwenda <i>et al.</i> , 2016)
		narK RF03032			(Weinberg <i>et al.</i> , 2017a)
		gabT RF01738 radC RF01754 traJ-II RF01760	gabT RF01738, radC RF01754 * traJ-II RF01760		(Weinberg <i>et al.</i> , 2010)
			sucA RF01070, MAEB (no Rfam name, RNA possible)		(Weinberg <i>et al.</i> , 2007)

			anti hemB RF02837	
			C4 antisense RNA RF01695 * sbcD RF01757** Burkholderiales -1 (no Rfam name, RNA possible)	(Weinberg <i>et al.</i> , 2010)
			int-<i>alpA</i> RF02996	(Weinberg <i>et al.</i> , 2017a)
Total ARN annotés	35	35	25	
Total ARN non caractérisés	5	16	11	
Pourcentage ARN à caractériser	14,3 %	45,7 %	44 %	

Tableau récapitulatif de tous les ARNnc agissant en cis dans les bactéries d'études à savoir *E. coli* K12 MG1655, *P. aeruginosa* UCBPP-PA14 et *B. thailandensis* E264. Les ARN annotés en vert seront l'objet de ce projet de recherche de doctorat. Les symboles °, * et ** indiquent que ces ARN ne sont pas présents dans les bactéries d'études mais dans des bactéries du même genre. °: *P. syringae* pv. *tomato* DC3000, *: *Paraburkholderia xenovorans* LB400 et **: *Burkholderia ambifaria* MC40-6. Ces ARN ont été inclus dans les calculs des ARN à caractériser.

Ainsi, nous pouvons voir que dans *Pseudomonas* et *Burkholderia* environ 45 % des ARNnc en cis annotés ne sont pas caractérisés contrairement à *E. coli* ne présentant que 14% d'ARNnc non caractérisés (Tableau 3.2). La différence est encore plus flagrante si on considère que la majorité des ARNnc caractérisés de *Pseudomonas* et *Burkholderia* l'ont été chez *E. coli* ou d'autres espèces parce que ces ARNnc sont très largement distribués. Il reste ainsi un grand travail à effectuer sur ces organismes. Une partie de ce travail de caractérisation manquant a été réalisée lors de ce projet de recherche de doctorat. En effet, l'étude de douze ARNnc candidats chez différentes souches de *Pseudomonas* ainsi que sept ARNnc chez *Burkholderia* a été planifiée. La caractérisation de ces candidats augmenterait ainsi grandement les connaissances sur ces organismes et leur métabolisme.

3.1.3 Problématique

La plupart des *riboswitchs* connus de nos jours ont été découverts via des méthodes de bioinformatique. En effet, des études de génomique comparative ont permis d'identifier des centaines d'ARNnc candidats dans de nombreux genres de bactéries (Barrick *et al.*, 2004; Weinberg *et al.*, 2007; Weinberg *et al.*, 2017a; Weinberg *et al.*, 2010). Ces ARNnc potentiels sont

appelés des *riboswitchs* orphelins car leur ligand est inconnu (Greenlee *et al.*, 2018). De ces études, plusieurs *riboswitchs* ont été validés expérimentalement comme le motif GEMM présent dans l'étude de 2007 (Weinberg *et al.*, 2007) qui a été le premier *riboswitch* retrouvé devant de nombreux gènes ayant des fonctions éloignées (Sherlock & Breaker, 2020). Il a été montré comme répondant au di-GMPc (Sudarsan *et al.*, 2008).

Afin de valider qu'un candidat est un nouvel ARNnc, il est nécessaire de le caractériser expérimentalement. Pour la caractérisation des *riboswitchs*, il existe de nombreuses méthodes (Figure 3.1) comme l'utilisation de cartographie *in vitro* par *in-line* avec différents ligands potentiels choisis en fonction du contexte génétique où se retrouve le motif. Cela est plus évident quand les motifs se retrouvent devant des gènes de même nature, comme le *riboswitch* glycine qui se retrouve en amont du gène *gcvT* impliqué dans le métabolisme de la glycine (Mandal *et al.*, 2004). Il est également possible de faire la validation expérimentale de ces candidats via des méthodes *in vivo*, comme avec des fusions du motif avec un gène rapporteur tel que la luciférase (opéron *luxCDABE*) ou encore la β -galactosidase (gène *lacZ*). Cette dernière méthode peut être faite avec des ligands précis comme pour les cartographies *in vitro* par *in-line*. Dans le cas de notre étude d'ARNnc qui pourraient, ou non, être des *riboswitchs*, des tests avec plusieurs milieux de culture de compositions différentes ont été effectués. Cela ne donne pas nécessairement le ligand correspondant, mais peut confirmer que nous sommes en présence d'un ARNnc *in cis* et nous informer sur les conditions dans lequel il est exprimé (Meyer *et al.*, 2011). Cette deuxième alternative permet à la fois la validation des *riboswitchs*, mais également d'ARNnc *in cis* d'autres catégories, comme les T-box, les *r-leaders* ou les thermorégulateurs.

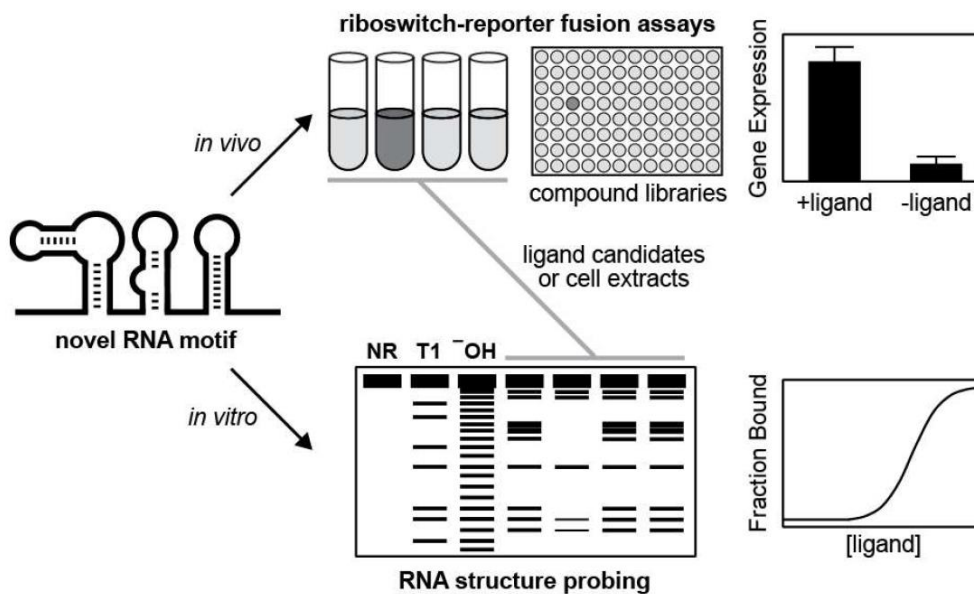


Figure 3.1 Schéma des méthodes de caractérisation d'un nouveau motif d'ARN non codant.

Quand un motif de potentiel ARNnc est découvert, il doit par la suite être validé expérimentalement. Pour cela, il peut y avoir les méthodes *in vivo* où une fusion avec un gène rapporteur est réalisée puis l'expression est comparée entre la condition avec ou sans ligand. Cette recherche de métabolite peut également se faire via l'essai de plusieurs milieux de culture. Il est également possible d'utiliser des méthodes *in vitro* notamment avec la méthode de cartographie *in vitro* par *in-line*, qui permet de vérifier le changement de conformation du *riboswitch* en présence de ligand. Elle permet par la suite de calculer la constante de dissociation associée à ce dernier (Sherlock & Breaker, 2020).

L'une des limites qui empêchait la caractérisation de nouveaux ARNnc régulant *in cis* chez des bactéries des genres *Pseudomonas* ou *Burkholderia* est la disponibilité des outils génétiques. En effet, il existe moins d'outils génétiques dans ces organismes rendant plus complexe leur travail. Cependant, nous avons développé au laboratoire un plasmide spécifiquement pour ces bactéries, qui est basé sur le plasmide mini-CTX-lux possédant l'opéron *luxCDABE* de *Xenorhabdus luminescens* (Becher & Schweizer, 2000). Ce plasmide a la capacité de s'intégrer directement dans le génome de ces bactéries le rendant ainsi plus stable (Becher & Schweizer, 2000). Ce plasmide a été modifié comme décrit dans la section 3.3 afin de faire des fusions traductionnelles entre les différents motifs à tester et l'opéron luciférase, notamment le gène *luxC*.

3.2 Matériel et méthodes

3.2.1 Design des ARNnc candidats

Les séquences des ARNnc choisis ont été extraites en utilisant l'outil BLAST dans les différents organismes étudiés (Altschul *et al.*, 1990). Les structures ont ensuite été dessinées en utilisant l'outil R2R (Weinberg & Breaker, 2011). Pour cela, les fichiers Stockholm des différents candidats ont été extraits via la plateforme Rfam (Griffiths-Jones *et al.*, 2003) ou à travers les figures supplémentaires des articles (Weinberg *et al.*, 2007; Weinberg *et al.*, 2010). Ces structures ont ensuite été analysées pour faire les différents mutants contenant des mutations disruptives et compensatoires. Les séquences des différentes constructions sont répertoriées dans l'Annexe 1.

Les amorces ont été créées en prévision du clonage par la méthode de Gibson. Pour cette méthode, il est nécessaire d'ajouter des régions dans les amorces qui sont complémentaires au plasmide. Pour nos essais, la séquence adaptatrice en 5' (5' - ggattcgaccggccatggcctgca - 3') et la séquence adaptatrice en 3' (5' - aataatgaatgaaatttttagtgaattcgctaggcaggtg - 3') ont été utilisées comme adaptateur de Gibson. Pour les constructions sauvage (WT) et disruptive (disr), les PCR ont été réalisées avec l'enzyme Q5. L'ADN de départ a été amplifié avec 1 µM final d'amorce sens et antisens, 0,2 mM final de dNTPs, 1X final de tampon Q5 et 1 U d'ADN polymérase Q5® High-Fidelity (New England Biolabs). Pour les mutants compensatoires, des PCR d'assemblage ont été réalisées également avec l'enzyme Q5 dans les mêmes conditions que précédemment (Figure 3.2).

Séquence à amplifier :

acgclttgtagaactccagttaatcgcccaccgcaacgggtg**gc**ggcgagggtcgtcacgccgcaacgaccccttcggcgaaagcttcgccaggcctccctgggggccaacgggacataacagtc

Amorce F1

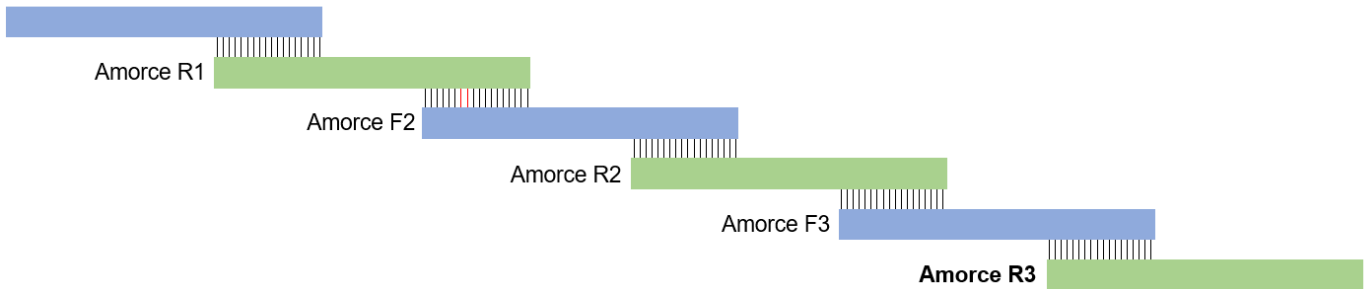


Figure 3.2 Schéma de la PCR d'assemblage.

Représentation schématique d'une réaction de PCR d'assemblage. Pour cela, différentes amorces forward (représentées en bleu) et reverse (représentées en vert) sont utilisées présentant des parties capables de s'hybrider (représentées par les traits noir). Les mutations compensatoires sont représentées en rouge. Les amorces des extrémités sont annotées en gras.

Cependant, dans ce cas, il n'y a pas d'ADN de départ. Les amorces des extrémités (F1 et R3) ont été ajoutées à une concentration finale de 1 μ M et les amorces centrales (R1, F2, R2 et F3) ont été mises à une concentration finale de 0,1 μ M (Figure 3.2). Les amorces utilisées ont été représentées dans le Tableau 3.3.

Tableau 3.3 Liste des amorces utilisées pour la construction des plasmides.

Constructions motif <i>rmf</i>
<p><i>rmf</i> WT : amplifié à partir de <i>P. aeruginosa</i> PA14</p> <ul style="list-style-type: none"> - ggattcgaccgggcatggcctgcaACGCTTTGGTTAGAACTCCAGTTAATCC - aataatgaatgaaatTTTTtagtgaattcgctaggcaggtgACGCTTAAGTCTTCTCATAGGGTGTTC
<p><i>rmf</i> disr : amplifié à partir de <i>P. aeruginosa</i> PA14</p> <ul style="list-style-type: none"> - ggattcgaccgggcatggcctgcaggACGCTTTGGTTAGAACTCCAGTTAATCGCCCACCGCAACGGTGT - aataatgaatgaaatTTTTtagtgaattcgctaggcaggtgACGCTTAAGTCTTCTCATAGGGTGTTC
<p><i>rmf</i> comp : PCR d'assemblage</p> <ul style="list-style-type: none"> - ACGCTTTGGTTAGAACTCCAGTTAATCGCCCACCGCAACGGTGTG - CGCCGAAAGGGGTCGTTGCCGGCGTGACGACCCTCGCCGCACACCGTTGCCGTGGGCGAT - CCCCTTTCCGCGAAAGCTTCGCCAGGCCTCCCCTGGGGGCCAACGGGACATAACAGTCAA - ACGCTTAAGTCTTCTCATAGGGTGTGCCCTCACTTGTTGACTGTTATGTCCCG
Constructions motif <i>Pseudomon</i> -Rho
<p><i>rho</i> WT : amplifié à partir de <i>P. aeruginosa</i> PA14</p> <ul style="list-style-type: none"> - ggattcgaccgggcatggcctgcaggTTAAATTCGGGTTCCGCGACGTT - aataatgaatgaaatTTTTtagtgaattcgctaggcaggtgGAGTTCGGTCAGATTCATAGGA
<p><i>rho</i> disr : amplifié à partir de <i>P. aeruginosa</i> PA14</p> <ul style="list-style-type: none"> - attcgaccgggcatggcctgcaggTTAAATTCGGGTTCCGCGTGGTTTTTTCCGTCGCC - aataatgaatgaaatTTTTtagtgaattcgctaggcaggtgGAGTTCGGTCAGATTCATAGGA
Constructions motif <i>rne-II</i>
<p><i>rne-II</i> WT : amplifié à partir de <i>P. aeruginosa</i> PA14</p> <ul style="list-style-type: none"> - ggattcgaccgggcatggcctgcaggAAGGAACATGTGTAGGGTGG - aataatgaatgaaatTTTTtagtgaattcgctaggcaggtgGATCAGCATTCTTTTCATGTAGTACC
<p><i>rne-II</i> disr : amplifié à partir de <i>P. aeruginosa</i> PA14</p> <ul style="list-style-type: none"> - ttcgaccgggcatggcctgcaggAAGGAACATGTGTAGGGTGGCTGATGTACGGCTGTT - aataatgaatgaaatTTTTtagtgaattcgctaggcaggtgGATCAGCATTCTTTTCATGTAGTACC
Constructions motif <i>sucA-II</i>
<p><i>sucA-II</i> WT : amplifié à partir de <i>P. aeruginosa</i> PA14</p> <ul style="list-style-type: none"> - ggattcgaccgggcatggcctgcaggAACACCTGTGGCGCCGGAG

- aataatgaatgaaatTTTTtagtgaattcgctaggcaggtgCATTACGCTTTTCGTGCATGCTAGGTC
sucA-II discr : amplifié à partir de <i>P. aeruginosa</i> PA14
- ggattcgaccggccatggcctgcaggAACACCTGTGGGCCCGGAGTTCTTCCGG - aataatgaatgaaatTTTTtagtgaattcgctaggcaggtgCATTACGCTTTTCGTGCATGCTAGGTC
Constructions motif <i>fixA</i>
<i>fixA</i> WT : amplifié à partir de <i>P. syringae</i> pv. <i>tomato</i> DC3000
- ggattcgaccggccatggcctgcaggGTCAAGATACGGGCGGTTTT - aataatgaatgaaatTTTTtagtgaattcgctaggcaggtgAGCTACAAGAACCCTTCATGG
<i>fixA</i> discr : amplifié à partir de <i>P. syringae</i> pv. <i>tomato</i> DC3000
- tcgaccggccatggcctgcaggGTCAAGATACGGGCGGTTTTGCCATTGCCGTAGGCT - aataatgaatgaaatTTTTtagtgaattcgctaggcaggtgAGCTACAAGAACCCTTCATGG
Constructions motif <i>Pseudomon</i>-GGDEF
<i>Pseudomon</i>-GGDEF WT : amplifié à partir de <i>P. syringae</i> pv. <i>tomato</i> DC3000
- ggattcgaccggccatggcctgcaggTCTCGGTTTTACCGTAAAGAAT - aataatgaatgaaatTTTTtagtgaattcgctaggcaggtgCGCTAGTTTTTTCAGCATCG
<i>Pseudomon</i>-GGDEF discr : amplifié à partir de <i>P. syringae</i> pv. <i>tomato</i> DC3000
- ggattcgaccggccatggcctgcaggTCTCGGTTTTACCGTAAAGAAT - aataatgaatgaaatTTTTtagtgaattcgctaggcaggtgCGCTAGTTTTTTCAGCATCGCAAGGGATCTCTG TTACCACAG

Tableau représentant les amorces utilisées pour amplifier les différents ARNc candidats. La bactérie qui a servi à l'amplification est indiquée. WT : correspond au motif naturel présent dans la bactérie. Discr : correspond à la structure défective du motif contenant les mutations disruptives. Comp : correspond à la construction permettant de rétablir la structure. Pour chaque amorce, la région en minuscule correspond aux régions adaptatrices pour la Gibson et celle en majuscule correspond à la séquence dans le génome bactérien.

3.2.2 Plasmides

Deux plasmides ont été utilisés. Tout d'abord le plasmide mini-CTX-lux contenant une copie unique du *riboswitch* fluor de *B. thailandensis*. Puis le plasmide mini-CTX-lux contenant une double construction de *riboswitch* fluor, un provenant de *Bacillus cereus* (*B. cereus*) et un provenant de *B. thailandensis*. Ils ont été extraits d'*E. coli* DH5 α en utilisant le kit Presto™ Midi plasmid de Geneaid conformément aux instructions du fabricant.

Pour le plasmide ne contenant qu'une copie de *riboswitch*, 1 μ g de plasmide a été digéré avec 7 U de l'enzyme de restriction Scal (ThermoFisher; ER0431) et 1,4 U de l'enzyme AarI

(ThermoFisher; ER1581). Pour compléter la réaction, 2 μL de buffer Tango 10X ont été ajoutés ainsi que 0,4 μL d'oligonucléotides 50X fournis avec l'enzyme AarI. La réaction de digestion a été incubée pendant 16h à 37°C puis inactivée pendant 20 min à 80°C. La qualité de la réaction a été ensuite vérifiée sur un gel agarose 0,8%.

Pour le plasmide contenant deux copies de *riboswitchs*, le même protocole a été utilisé mais l'enzyme Scal a été remplacée par 7 U de SdaI (ThermoFisher; ER1191).

Les plasmides ont ensuite été directement utilisés pour le clonage.

3.2.3 Clonage

Les ARNnc candidats ont été clonés via la méthode de Gibson Assembly (New England Biolabs). Pour cela, 0,02 pmol de plasmide ont été mis en contact avec 0,06 pmol d'insert ainsi que le Master mix de Gibson (HiFi DNA assembly, New England Biolabs). Les réactions ont été incubées pendant 1h à 50°C. Puis la moitié de chaque réaction a été ajoutée à 100 μL de *E. coli* DH5 α compétentes et incubées pendant 30 min sur glace. Un choc thermique de 45 sec à 42°C a été réalisé avant de replacer les cellules 5 min dans la glace. Un volume de 1 mL de milieu LB a été ajouté puis une incubation de 1h30 à 37°C a été faite. Pendant ce temps, des boîtes de Petri de LB agar ont été coulées en prenant soin d'ajouter une concentration de 15 $\mu\text{g}/\text{mL}$ de tétracycline. Un volume de 100 μL de bactéries a été étalé sur les boîtes de Petri et elles ont été incubées toute la nuit à 37°C.

Les colonies ont ensuite été relancées en culture dans du milieu LB liquide avec 15 $\mu\text{g}/\text{mL}$ de tétracycline et incubées tout la nuit en agitation à 37°C. Les plasmides des clones ayant repris en culture ont été extraits en utilisant le kit Presto™ Mini plasmid de Geneaid conformément aux instructions du fabricant. Des PCR de vérification ont été réalisées avec l'enzyme Q5 en utilisant les amorces AD123 (5' - CCATTAAATGGATGGCAAATACCAGGT - 3') et AD131 (5' - GGATTGCACTAAATCATCACTTTTCGGG - 3') à un T_m de 68°C. Un gel d'agarose 2% a été réalisé afin de visualiser le résultat. Les clones positifs ont été stockés dans 15% glycérol à -80°C puis les PCR ont été envoyées pour un séquençage Sanger au Centre d'expertise et de service Génome Québec. Les plasmides extraits ont par la suite été transformés par choc thermique, comme décrit précédemment, dans une souche d'*E. coli* sm10 λpir permettant de réaliser les conjugaisons.

3.2.4 Conjugaison

Les souches d'*E. coli* sm10 λ pir contenant les différents plasmides et de *P. aeruginosa* (PA14, PA01, PA01 $\Delta relA \Delta spoT$, PA14 *wspR*-, PA14 *wspF*-), ont été cultivées toute la nuit à 37°C en agitation, respectivement dans 3 mL de TSB avec 15 μ g/mL de tétracycline et dans 3 mL de TSB. Des volumes de 300 μ L de la souche donneuse (*E. coli* sm10 λ pir) et 150 μ L de souche receveuse (*P. aeruginosa*) ont été mélangés, lavés deux fois par des traitements de centrifugation (1 min à 8000 x g) et remis de nouveau dans 450 μ L de milieu TSB. Lors du dernier lavage, le culot a été resuspendu dans 60 μ L de TSB avant d'être placé au centre d'une boîte de Petri contenant 20 mL de TSB agar. La réaction a par la suite été incubée pendant 5 à 7h à 37°C, couvercle vers le haut. Pendant ce temps, des boîtes de Petri de milieu TSB agar ont été préparées en ajoutant 125 μ g/mL de tétracycline et 25 μ g/mL de triclosan (l'antibiotique servant de contre sélection pour *E. coli*). Les bactéries conjuguées ont été récupérées à l'aide de TSB, puis une dilution au dixième a été réalisée. Finalement, 200 μ L de chaque dilution ont été étalées sur les boîtes de Petri TSB supplémentées en tétracycline et en triclosan et ont été incubées toute la nuit à 37°C. Les colonies ont ensuite été relancées en culture dans 3 mL de TSB liquide avec 75 μ g/mL de tétracycline et incubées tout la nuit en agitation à 37°C. Un volume de 3 μ L de culture a été utilisé pour faire les PCR de vérification avec l'enzyme ADN polymérase Q5® High-Fidelity (New England Biolabs), en utilisant les amorces AD123 et AD131 (séquence disponible dans la section 3.2.3) à un T_m de 68°C. Un gel d'agarose 2% a ensuite été réalisé. Les clones positifs ont été stockés dans 15% glycérol à -80°C.

3.2.5 Milieux de culture utilisés

Tableau 3.4 Liste des milieux de culture communs à tous les motifs testés.

Milieux riche	Milieux chimiquement définis			
LB	M9 glucose	M9 Alanine	M9 Lysine	M9 Casamino Acid
LB pH 5	M9 dextrose	M9 Asparagine	M9 Méthionine	M9 0,1 M NaCl
LB pH 6	M9 fructose	M9 Cystéine	M9 Phénylalanine	M9 0,5 M NaCl
LB pH 7	M9 maltose	M9 Glutamine	M9 Proline	M9 1 M NaCl
LB pH 8	M9 glucose pH 5	M9 Glycine	M9 Sérine	M9 low Phosphate
TSB + EDTA	M9 glucose pH 6	M9 Histidine	M9 Thréonine	M63
TSB + EGTA	M9 glucose pH 7	M9 Isoleucine	M9 Tryptophane	SCFM
	M9 glucose pH 8	M9 Leucine	M9 Valine	

Tableau représentant tous les milieux utilisés pour les différents tests des plaques de luminescence pour chacun des candidats. Le milieu SCFM correspond au Synthetic Cystic Fibrosis sputum Medium (Palmer *et al.*, 2007).

LB et LB agar

Le milieu LB a été réalisé en mélangeant 25 g/L de milieu Luria Broth (Alpha Biosciences) dans de l'eau Milli-Q puis autoclavé.

Pour les essais avec les variations de pH du milieu LB (pH initial de 6,85), il a été ajusté en ajoutant du chlorure d'hydrogène 37% (HCl) ou de l'hydroxyde de sodium 10N (NaOH).

Pour le milieu agar, 1,5% d'agar a été ajouté lors de la préparation du milieu liquide et autoclavé.

TSB et TSB agar

Le milieu TSB a été fait en mélangeant 30 g/L de milieu Trypto Soy Broth (Alpha Biosciences) dans de l'eau Milli-Q puis autoclavé.

Pour les essais avec l'EDTA et l'EGTA, une concentration finale de 1 mM a été ajoutée.

Pour le milieu agar, 1,5% d'agar a été ajouté lors de la préparation du milieu liquide et autoclavé.

Pour les essais en anaérobie, une concentration finale de 1% KNO₃ a été ajoutée au milieu TSB agar.

M9 et M9 agar

Le milieu M9 de base a été réalisé en mélangeant 200 mL de M9 salt 5X autoclavé (64 g/L de Na₂HPO₄-7H₂O, 15 g/L KH₂PO₄, 2,5 g/L NaCl, 5 g/L NH₄Cl) avec 2 mL de MgSO₄ 1 M stérile, 100 µL de CaCl₂ 1 M stérile et 20 mL de glucose à 20% stérile. Le volume a été ajusté à 1 L avec de l'eau Milli-Q stérile (Elbing & Brent, 2002).

Pour les essais avec des variations dans les sources de carbone (glucose, maltose, dextrose et fructose, succinate, citrate, acétate et glycérol), une même concentration de 20% a été utilisée pour les différents composés testés.

Pour les essais avec les variations de pH du milieu M9 (pH initial de 7,16), il a été ajusté en ajoutant du chlorure d'hydrogène 37% (HCl) ou de l'hydroxyde de sodium 10 N (NaOH).

Des milieux M9 supplémentés en acides aminés ont également été utilisés. Pour cela, les mêmes conditions que le milieu M9 de base ont été faites avec comme source de carbone du glucose en ajoutant 16 mL d'une solution à 5 mg/mL des différents acides aminés. Plusieurs acides aminés ont été testés de façon individuelle (Tableau 3.3).

Pour la condition avec une faible concentration de phosphate, le M9 salt 5X a été modifié avec 0,01 g/L de KH₂PO₄ au lieu de 15 g/L qui a été compensé avec 14,9 g/L de KCl.

Milieu M63 (Elbing & Brent, 2002)

Une solution de sel 5X a été réalisée avec 10 g/L (NH₄)₂SO₄, 68 g/L KH₂PO₄, 2,5 mg FeSO₄-7H₂O. Le pH a été ajusté à 7 avec du KOH et autoclavé. Une fois stérilisé, la solution de sel a été diluée à 1X avec de l'eau Milli-Q stérile. Puis, 1 mL de MgSO₄ 1M et 10 mL de glucose 20% stériles ont été ajoutés.

SCFM (Synthetic Cystic Fibrosis sputum Medium) (Palmer *et al.*, 2007)

Les solutions stocks d'acides aminés ont été réalisées à une concentration de 100 mM dans de l'eau Milli-Q et stockés à 4°C dans le noir, hormis la tyrosine, l'aspartate et le tryptophane qui ont été resuspendus dans du NaOH à 1 M, 0,5 M et 0,2 M respectivement. Les acides aminés ont été ajoutés à un milieu tamponné (6,5 mL NaH₂PO₄ 0,2 M, 6,25 mL Na₂HPO₄ 0,2 M, 0,348 mL KNO₃ 1M, 1,084 mL K₂SO₄ 0.25M, 0,122g NH₄Cl, 1,114g KCl, 3,03g NaCl et 10 mM MOPS) à différents volumes (Palmer *et al.*, 2007). Le pH du milieu a été ajusté à 6,8 puis stérilisé grâce à

un filtre 0,22 µm. Après la stérilisation, 1,754 mL de CaCl₂ 1M, 0,606 mL de MgCl₂ 1, 1 mL de FeSO₄-7H₂O 3,6 mM et 3 mL de D-glucose 1 M stériles ont été ajoutés.

Peptone-Succinate Salt Medium (PSSM)

Le milieu Peptone-Succinate Salt Medium (PSSM) a été réalisé en mélangeant 10 g/L de Peptone (Bacto, BD), 1 g/L d'acide succinique, 1 g/L de (NH₄)₂SO₄, 1 g/L de MgSO₄ – 7 H₂O, 2 mg/L de FeCl₃ – 6 H₂O et 2 mg/L de MnSO₄ – H₂O. Le milieu a ensuite été stérilisé en utilisant un filtre 0,22 µm.

3.2.6 Essai luminescence

Les bactéries à tester ont été cultivées toute la nuit dans du milieu LB ou TSB supplémenté avec 75 µg/mL de tétracycline à 30°C ou 37°C en agitation, dépendamment si le motif se trouve initialement dans *P. syringae* pv. *tomato* ou *P. aeruginosa* respectivement. Trois lavages avec chaque milieu à tester ont été réalisés en centrifugeant 0,5 mL de culture pendant 2 min à 9000 x g. Le surnageant de culture a été jeté puis le même volume du milieu correspondant a été ajouté pour resuspendre le culot avant de refaire une étape de centrifugation. La densité optique (DO) a été mesurée à une longueur d'onde de 600 nm, puis les bactéries ont été diluées dans leur milieu correspondant pour obtenir une DO initiale de 0,05. Chaque préculture a été testée sous la forme de triplicata technique dans trois puits distincts pour nos essais de luminescence. Les essais ont été effectués dans des plaques blanches à fond plat transparent (655073; Greiner Bio One). Les lectures ont été réalisées avec le Cytation 3 pour la réalisation de cinétique avec des mesures de luminescence et de DO toutes les 20 minutes ou avec la lectrice à plaque infinie M1000 (TECAN®). Pour la lectrice à plaque TECAN®, les mesures ont été prises à différents temps à savoir : dès la fin de la réalisation de la plaque (T0), après 4 à 5h d'incubation (T5) et après 24h d'incubation (T24). Pour certains essais des prises de mesures à d'autres temps ont également été effectuées. Pour l'incubation, les plaques ont été placées à 30°C ou 37°C en agitation. Pour chacun des motifs, la DO 600 nm ainsi que la luminescence ont été mesurées à chacun de ces temps. Le ratio luminescence sur DO a été calculé en faisant la moyenne des trois répliques techniques. Un *t-test* multiple non pairé a été réalisé afin de comparer les différences d'expression de la luminescence entre les différentes constructions WT et *disr* (Manna *et al.*, 2021).

Pour le motif *fixA*, des essais de luminescence ont été réalisés en condition anaérobie. Pour cela, les différentes souches à tester ont été cultivées en agitation dans du TSB supplémenté avec 75 µg/mL de tétracycline toute la nuit à 30°C. La DO 600 nm a été mesurée afin de s'assurer que

les cultures avaient une croissance semblable pendant la nuit. Un volume de 20 µL de culture a été prélevé et utilisé pour strier sur les boîtes de Pétri contenant du milieu TSB agar ou TSB agar supplémenté avec 1% de KNO₃. Deux boîtes de Pétri pour chaque milieu ont été réalisées, du parafilm a été utilisé afin de sceller les Pétri pour ne pas que l'agar sèche lors de l'incubation. La culture a été initiée pendant 3h à 30°C en aérobie pour toutes les boîtes, puis la moitié a été placée dans une jarre anaérobie à 30°C pendant trois jours. Les Pétri en aérobie et anaérobie de chaque milieu ont été visualisées en même temps en utilisant le ImageQuant™ LAS 500 avec le programme de chimiluminescence (GE Healthcare).

3.2.7 Cartographie *in vitro* par *in-line*

3.2.7.1 Amplification par PCR

Les différentes constructions des motifs *rmf* et *Pseudomon*-GGDEF ont été amplifiées par PCR grâce à l'ADN polymérase Q5, comme décrit précédemment (section 3.2.1). Les amorces utilisées ont été répertoriées dans le Tableau 3.5.

Tableau 3.5 Liste des amorces utilisées pour les différentes constructions des motifs pour le *in-line*

Constructions motif Rmf
<p>Amorce forward commune pour toutes les constructions de <i>rmf</i></p> <ul style="list-style-type: none"> - TAATACGACTCACTATAGGTTGGTTAGAACTCCAGTTAATCCGCC
<p><i>rmf</i> WT 1 : amplifié à partir de <i>P. aeruginosa</i> PA14</p> <ul style="list-style-type: none"> - CATAGGGTGTGCCCTCACTT <p><u>Séquence ARN complète :</u></p> <p>uugguuagaacuccaguuauuccgccaccgcaacggugucggggcgagggucgucacgccggcaacgacccuuucggcgaaagc uucgccaggccucccugggggccaacgggacauaacagucaacaagugagggcaacaccuauug</p>
<p><i>rmf</i> WT 2 : amplifié à partir de <i>P. aeruginosa</i> PA14</p> <ul style="list-style-type: none"> - CCTCACTTGTTGACTGTTATGTCCC <p><u>Séquence ARN complète :</u></p> <p>uugguuagaacuccaguuauuccgccaccgcaacggugucggggcgagggucgucacgccggcaacgacccuuucggcgaaagc uucgccaggccucccugggggccaacgggacauaacagucaacaagugaggg</p>
<p><i>rmf</i> WT 3 : amplifié à partir de <i>P. aeruginosa</i> PA14</p> <ul style="list-style-type: none"> - GTTGACTGTTATGTCCCGTT <p><u>Séquence ARN complète :</u></p> <p>uugguuagaacuccaguuauuccgccaccgcaacggugucggggcgagggucgucacgccggcaacgacccuuucggcgaaagc uucgccaggccucccugggggccaacgggacauaacagucaac</p>
<p><i>rmf</i> WT 4 : amplifié à partir de <i>P. aeruginosa</i> PA14</p> <ul style="list-style-type: none"> - TTATGTCCCGTTGGCCCCCA <p><u>Séquence ARN complète :</u></p> <p>uugguuagaacuccaguuauuccgccaccgcaacggugucggggcgagggucgucacgccggcaacgacccuuucggcgaaagc uucgccaggccucccugggggccaacgggacauaa</p>
<p><i>rmf</i> disr 4 : amplifié à partir de <i>P. aeruginosa</i> PA14</p> <ul style="list-style-type: none"> - TTATGTCCCGTTGGCCCCCA <p><u>Séquence ARN complète :</u></p> <p>uugguuagaacuccaguuauuccgccaccgcaacggugucggggcgagggucgucacgccggcaacgacccuuucggcgaaagc uucgccaggccucccugggggccaacgggacauaa</p>
Constructions motif <i>Pseudomon</i> -GGDEF
<p>Amorce forward commune pour toutes les constructions de <i>Pseudomon</i>-GGDEF</p> <ul style="list-style-type: none"> - TAATACGACTCACTATAGGTTTTACCGTAAAGAATTAATCACATCC
<p><i>Pseudomon</i>-GGDEF 1 : amplifié à partir de <i>P. syringae</i> pv. <i>tomato</i> DC3000</p>

<p>- TTTCAGCATCGCAAGGCTTC</p> <p><u>Séquence ARN complète :</u></p> <p>guuuuaccguaaagaauuaaucacauccgccgagagccugacaagcgagcugauuaauaucgggcuugcagaaugagcaguaa gccacgugugcuuugacugugguaacagagaagccuugcgaugcugaa</p>
<p><i>Pseudomon</i>-GGDEF 2 : amplifié à partir de <i>P. syringae</i> pv. <i>tomato</i> DC3000</p> <p>- CGCAAGGCTTCTCTGTTACC</p> <p><u>Séquence ARN complète :</u></p> <p>guuuuaccguaaagaauuaaucacauccgccgagagccugacaagcgagcugauuaauaucgggcuugcagaaugagcaguaa gccacgugugcuuugacugugguaacagagaagccuugcg</p>
<p><i>Pseudomon</i>-GGDEF 3 : amplifié à partir de <i>P. syringae</i> pv. <i>tomato</i> DC3000</p> <p>- CCACAGTCAAAGCACACGTG</p> <p><u>Séquence ARN complète :</u></p> <p>guuuuaccguaaagaauuaaucacauccgccgagagccugacaagcgagcugauuaauaucgggcuugcagaaugagcaguaa gccacgugugcuuugacugugg</p>
<p><i>Pseudomon</i>-GGDEF 4 : amplifié à partir de <i>P. syringae</i> pv. <i>tomato</i> DC3000</p> <p>- CATCGCAAGGCTTCTCTGTT</p> <p><u>Séquence ARN complète :</u></p> <p>guuuuaccguaaagaauuaaucacauccgccgagagccugacaagcgagcugauuaauaucgggcuugcagaaugagcaguaa gccacgugugcuuugacugugguaacagagaagccuugcgaug</p>
<p><i>Pseudomon</i>-GGDEF 5 : amplifié à partir de <i>P. syringae</i> pv. <i>tomato</i> DC3000</p> <p>- TTCTCTGTTACCACAGTCAAAG</p> <p><u>Séquence ARN complète :</u></p> <p>guuuuaccguaaagaauuaaucacauccgccgagagccugacaagcgagcugauuaauaucgggcuugcagaaugagcaguaa gccacgugugcuuugacugugguaacagagaa</p>

Tableau représentant les amorces utilisées pour amplifier les différentes constructions des motifs rmf et Pseudomon-GGDEF pour réaliser les cartographies in vitro par in-line. WT : correspond au motif naturel, Disr : correspond à la structure défaire du motif contenant les mutations disruptives. Les séquences en bleu correspondent au promoteur T7 permettant la transcription en ARN. Les nucléotides en rouge correspondent aux mutations disruptives.

3.2.7.2 Transcription *in vitro*

Les réactions de PCR ont été incubées pendant 3h à 37°C avec de l'ARN polymérase T7 (20 µL de PCR [~400 ng], 20 µL 5X tampon de transcription [400 mM HEPES-KOH, pH 7.5, 120 mM MgCl₂, 10 mM spermidine, 200 mM DTT], 0,001 U pyrophosphatase (Roche), 0,5 µL d'inhibiteur d'ARNase Ribolock (Thermo Fischer/EO0382) [40 U/ µL], ARN polymérase T7 [100 U], rNTP 5 mM (GTP, ATP, CTP et UTP) et 5 µCi [α-³²P] UTP). Le volume de réaction a été par la suite ajusté à 100 µL avec de l'eau Milli-Q. Après une incubation de 30 min avec 2 U d'ADNase I, les

ARN ont été purifiés avec du chloroforme afin d'éliminer les traces d'ADN. Après une centrifugation de 10 min à 15 000 rpm, la phase aqueuse a été récupérée et les ARN ont été précipités à -80°C avec 0,1 volume d'acétate de sodium 3 M (pH 5,2) et 2 volumes d'éthanol froid 100% pendant au moins 2 heures. Les ARN précipités ont été resuspendus dans 20 µL d'eau Milli-Q et purifiés sur un gel de polyacrylamide dénaturant (8 M urée) à 6%. Ils ont été déposés sur gel grâce à un tampon de charge dénaturant 2X (contenant 0,05% de bleu de bromophénol, 0,05% de bleu de xylène cyanol, 10 mM EDTA [pH 8], 95% formamide). Les ARN ont été visualisés grâce à la méthode d'ombrage par UV (*UV shadowing*) sur une plaque de silice. Ils ont ensuite été élués avec 500 µL de tampon d'éluion (0,3 M NaCl) puis ont été précipités comme expliqué précédemment.

3.2.7.3 Déphosphorylation et marquage à la radioactivité

Les ARN précipités ont ensuite été déphosphorylés en utilisant 5 U d'antartic phosphatase (New England Biolabs; M0289S) accompagnée de son buffer 1X ainsi que 0,5 µL d'inhibiteur d'ARNase Ribolock (Thermo Fisher; EO0382) dans un volume final de 20 µL. La réaction a été incubée pendant 20 minutes à 37°C puis 5 minutes à 65°C pour désactiver l'enzyme. Environ 10 pmol d'ARN déphosphorylés ont été utilisés pour le radiomarquage avec 0,5 µL de l'enzyme polynucléotide kinase T4 (T4 PNK) (New England Biolabs; M0201S), 2 µL de T4 PNK buffer (10X) et 2 µL de [γ -³²P] ATP (Perkin Elmer) dans un volume final de 20 µL. La réaction a été incubée pendant 1h à 37°C, puis les ARN radiomarqués ont été purifiés sur un gel de polyacrylamide dénaturant 6% comme décrit précédemment. Ils ont été visualisés grâce à une exposition sur cassette de phosphore, imagés avec le Typhoon FLA9500 (GE Healthcare Life Sciences). Les ARN ont ensuite été élués et précipités comme décrit précédemment. Les ARN ont par la suite été resuspendus dans 400 µL d'eau Milli-Q stérile.

3.2.7.4 Cartographie *in vitro* par *in-line*

Un volume de 4 µL d'ARN radioactif a été mélangé avec 10 µL de tampon de *in-line* (20 mM MgCl₂, 50 mM Tris-HCl pH 8,3 (à 25°C) et 100 mM KCl) et 2 µL de ligands à différentes concentrations. Le volume a ensuite été ajusté à 20 µL avec de l'eau Milli-Q stérile. Pour *rmf*, les concentrations de ppGpp (TriLink Biosciences) et pppGpp (Jena Biosciences) testées sont de 10 nM à 100 µM. Pour *Pseudomon*-GGDEF, des concentrations allant de 10 nM à 100 µM ont été utilisées pour le di-GMPc (Sigma) et le pGpG (Jena Biosciences). Les réactions ont été

incubées pendant environ 40h à température de la pièce, puis ont été arrêtées en utilisant du tampon de charge dénaturant 2X.

Des échelles de contrôles ont été réalisées (NR : non réagi; T1 : qui est un traitement avec l'ARNase T1 permettant de couper après chaque guanine de l'ARN; OH : qui est une échelle alcaline coupant après chaque nucléotide). Celles-ci permettent d'aligner la structure de l'ARN sur la séquence. Pour l'échelle NR, 5 μ L d'ARN radioactif ont été dilués dans de l'eau Milli-Q dans un volume final de 20 μ L. Pour l'échelle T1, l'ARN radioactif a été incubé avec 50 mM de citrate de sodium et 3 μ L de formamide 100% pendant 2 min à 56°C. Un volume de 1 μ L de ARNase T1 (Thermo Scientific) a été ajouté à la réaction puis celle-ci a été incubée de nouveau pendant 5 min à 56°C. Pour l'échelle OH, l'ARN radioactif a été incubé pendant 1 min 20 à 90°C avec 50 mM de carbonate de sodium. Toutes les réactions ont été arrêtées avec 20 μ L de tampon de charge dénaturant 2X.

Un gel de polyacrylamide dénaturant 8% a été réalisé où les échelles et les échantillons ont été migrés pendant 3h à 50W. Le gel a ensuite été séché (gel dryer model 585 Biorad avec une pompe à vide CVC3000 Vaccumbrand), exposé sur cassette de phosphore et visualisé à l'aide du Typhoon FLA 9500 (GE Healthcare Life Sciences).

Pour le calcul de la constante de dissociation (K_D), les intensités des puits et des bandes qui modulent ont été quantifiées en utilisant le logiciel ImageQuant TL (GE Healthcare Life Sciences). Les intensités des bandes ont ensuite été normalisées en fonction de l'intensité de chaque puits, mais aussi entre elles, à l'aide du calcul des valeurs minimales et maximales, pour donner des valeurs comprises entre 0 et 1. La représentation des courbes de K_D ainsi que le calcul de sa valeur ont été réalisés en utilisant le logiciel GraphPad Prism 8 (Sherlock *et al.*, 2017). Pour le calcul et la représentation du K_D , la condition sans ligand (0) a été représentée par une concentration 10^{-10} car sur une échelle logarithmique il n'est pas possible d'avoir une valeur de 0. Ceci nous permet également de vérifier le plateau dans la représentation graphique. Le K_D moyen a été calculé en faisant la moyenne des valeurs de K_D obtenues pour chacune des bandes. L'écart-type a été déterminé grâce à Excel à partir des valeurs de K_D obtenues pour chacune des bandes.

3.3 Présentation du rapporteur d'expression

3.3.1 Plasmide utilisé pVK-f2-lux

Le plasmide pVK-f2-lux a été créé à partir du plasmide mini-CTX-lux, contenant l'opéron *luxCDABE* de *Xenorhabdus luminescens* (Becher & Schweizer, 2000). Ce plasmide a été modifié de manière à pouvoir réaliser des fusions traductionnelles propres ne contenant pas de sites de restrictions entre l'insert et le gène *luxC*, ce qui n'est pas toujours présent dans les plasmides avec gène rapporteur mais essentiel dans le cas de test d'ARNnc régulant en *cis* (Korniakova *et al.*, 2021). Ce plasmide a comme propriété d'être présent à un grand nombre de copies chez *E. coli* la bactérie qui nous permet de créer nos plasmides. Cependant, chez *P. aeruginosa* ou *B. thailandensis*, nos organismes d'études, il est présent en une copie unique qui sera intégrée au chromosome (Korniakova *et al.*, 2021). Cette caractéristique le rend ainsi plus stable dans la bactérie (Becher & Schweizer, 2000).

Le plasmide pVK-f2-lux contient le promoteur P1 et le *riboswitch* fluor de *B. thailandensis* (Figure 3.3 A). Le promoteur P1 est un promoteur constitutif originellement retrouvé dans le plasmide R388, un plasmide impliqué dans la résistance à la triméthopime (Zolg & Hänggi, 1981). Plusieurs sites multiples de clonage ont été ajoutés en amont (MSC1) et en aval (MSC2) du promoteur permettant à la fois de pouvoir modifier le promoteur mais aussi pouvoir enlever le *riboswitch* fluor grâce à l'enzyme de restriction AarI retrouvée dans la partie 3' du design qui a été ajoutée au plasmide (Figure 3.3 B) (Korniakova *et al.*, 2021). AarI est l'enzyme permettant de faire les fusions propres dans le plasmide, c'est-à-dire sans reste de sites de restrictions. En effet il s'agit d'une enzyme de restriction de type IIS permettant de couper l'ADN à une distance définie en aval de la séquence de reconnaissance quel que soit les nucléotides présents (Pingoud & Jeltsch, 2001). La coupure se fait à quatre nucléotides sur le brin sens et à huit nucléotides sur le brin anti-sens après le site de restriction de l'enzyme. Ce site est situé au niveau du deuxième nucléotide du deuxième codon du gène *luxC* permettant lors de la coupure d'enlever le site d'initiation de la traduction de ce gène (Korniakova *et al.*, 2021).

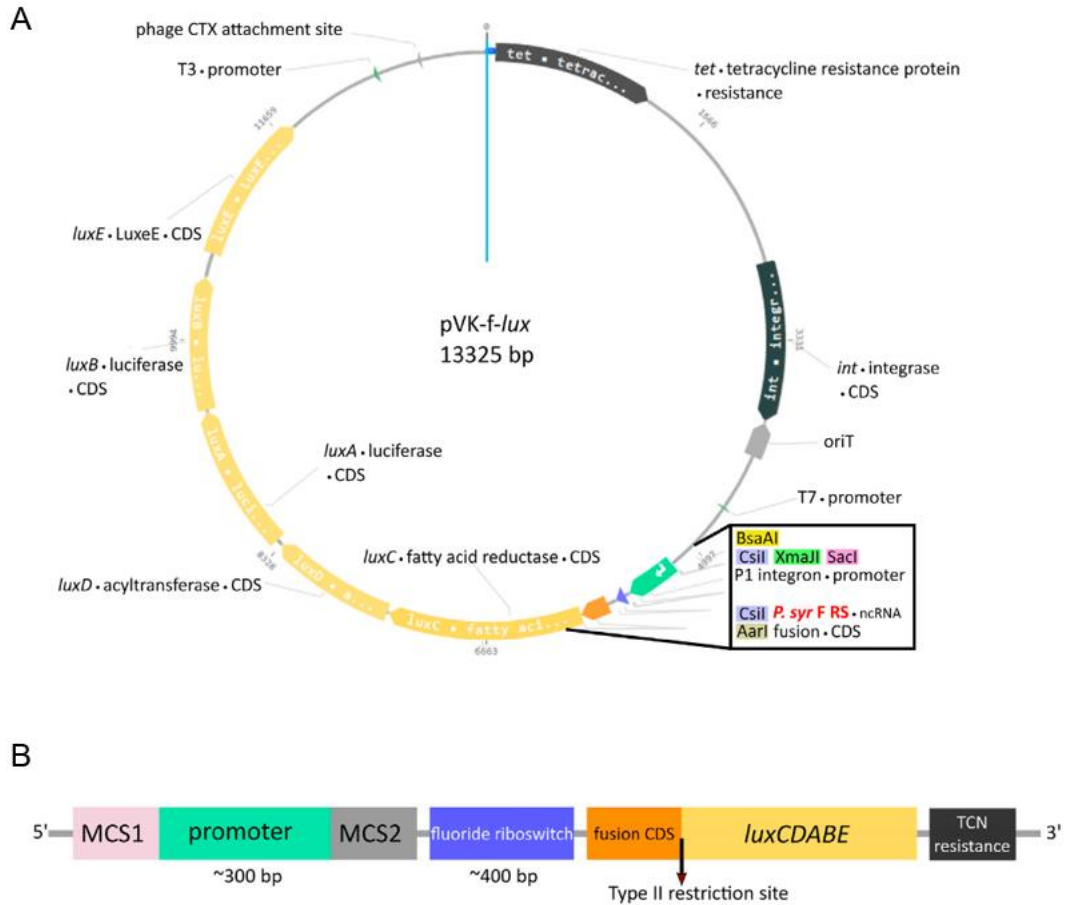


Figure 3.3 Le plasmide pVK-f2-lux.

A : Carte du plasmide pVK-f2-lux. **B :** Représentation schématique des éléments importants pour le clonage dans le plasmide pVK-f2-lux, le promoteur P1 (représenté par un rectangle vert) est entouré par les sites multiples de clonage 1 et 2 contenant les enzymes Csil, XmaJI et Sacl pour MSC1 et Sacl, NcoI, Sdal et Scal pour MSC2. Le *riboswitch* fluor de *B. thailandensis* est ensuite présent (rectangle bleu) suivi par le début du gène *crcB* en fusion grâce à AarI, une enzyme de restriction de type IIS, avec le gène *luxC*. Le gène *luxC* est le premier gène de l'opéron *luxCDABE* permettant l'expression de la luciférase. Le plasmide pVK-f2-lux présente aussi un gène de résistance à la tétracycline (TCN résistance, rectangle noir) (Korniakova *et al.*, 2021).

3.3.2 Preuve de concept avec le *riboswitch* fluor de *B. thailandensis*

Afin de valider la fonctionnalité du plasmide pVK-f2-lux contenant une copie du *riboswitch* fluor de *B. thailandensis*. Des essais de luminescence ont été réalisés avec différentes concentrations de fluor (de 0 mM à 62 mM) dans *B. thailandensis* pendant 48h puis le ratio luminescence / DO 600 nM a été calculé grâce au Cytation 3 (Figure 3.4).

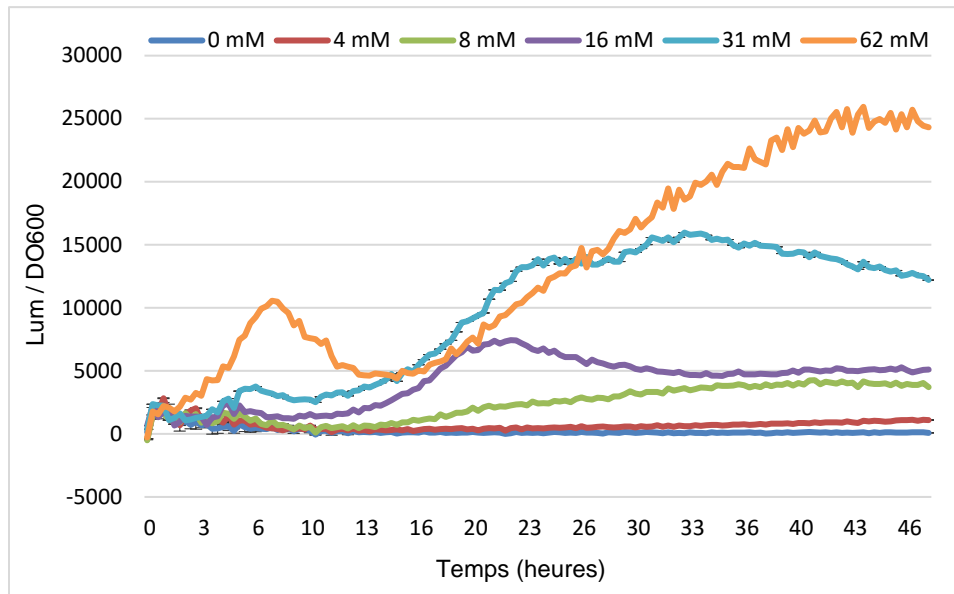


Figure 3.4 Augmentation de la luminescence produite via pVK-f2-lux avec des concentrations croissantes de fluor.

La souche de *B. thailandensis* E264 contenant le plasmide pVK-f2-lux + *riboswitch* fluor a été cultivée pendant 48h au sein du Cytation 3 à 37°C avec plusieurs concentrations de fluor (0 mM, 4 mM, 8 mM, 16 mM, 31 mM et 62 mM). Des lectures de la DO 600 nm et de la luminescence ont été effectuées toutes les 20 min et la courbe du ratio luminescence / OD600 a été réalisée.

Nous pouvons observer que selon la concentration de fluor présente dans le milieu, le niveau de luminescence est différent. En effet, en absence de fluor (0 mM) on peut voir qu'il n'y a aucune expression de l'opéron *lux* au cours du temps (Figure 3.4). Cependant, on peut remarquer que dès une concentration de 4 mM de fluor, il y a une augmentation de l'expression de l'opéron *lux* se traduisant par une augmentation du niveau de luminescence dans le temps, qui se confirme de façon dose dépendante avec les concentrations suivantes (Figure 3.4). Le fluor étant le ligand naturel de ce *riboswitch*, cela nous permet de confirmer qu'en présence du ligand l'expression de la luminescence est modifiée. Concernant les pics observés dans la Figure 3.4, il est difficile d'expliquer exactement ces variations. Cependant, nous savons que le *riboswitch* présent dans ce plasmide est *crcB* liant le fluor (Baker *et al.*, 2012). Ce *riboswitch* est présent en amont du gène *crcB* connu pour limiter la toxicité du fluor dans la bactérie via l'expression d'une pompe à efflux (Baker *et al.*, 2012). De plus, l'expression du gène rapporteur (*lux*) est le reflet de ce qui se passe pour le gène endogène dans *B. thailandensis*. Ainsi, il est possible qu'à ces différents moments la bactérie ait besoin d'exprimer la pompe à efflux pour sortir le fluor, qui n'est pas nécessaire pour sa croissance et est uniquement toxique. Le niveau de luminescence redescend par la suite car la concentration intracellulaire est redescendu il n'est donc plus nécessaire

d'exprimer cette pompe à efflux. Ces essais permettent également de montrer que le maintien de l'homéostasie dans la cellule n'est pas quelque chose de constant. Les bactéries ont besoin de se maintenir dans une zone où tous les éléments sont contrôlés.

Lors de cette première étude, nous avons également montré que grâce à ce plasmide, il était possible de cloner en amont de l'opéron luciférase des promoteurs très forts, comme le promoteur P_{S7} de *Paraburkholderia xenovorans* LB400 (Korniakova *et al.*, 2021), alors que la toxicité induite par ce niveau d'expression est normalement délétère et empêche le clonage. Cependant, le design du vecteur permet également de digérer le plasmide afin d'enlever le ou les *riboswitchs* fluor pour y mettre des ARNnc d'intérêt (Korniakova *et al.*, 2021).

3.3.3 Design des ARN non codants candidats

Pour chacun des ARNnc candidats testés, le même concept de design a été réalisé. En effet, le motif a été pris en incluant cinq nucléotides de la séquence intergénique en amont du motif naturellement présent dans la bactérie ainsi que toute la séquence intergénique jusqu'à l'AUG du gène en aval (Figure 3.5). Enfin, les six premiers codons du gène en aval ont été ajoutés au design pour être placés en fusion traductionnelle avec le gène *luxC* de l'opéron *luxCDABE* (Figure 3.5).

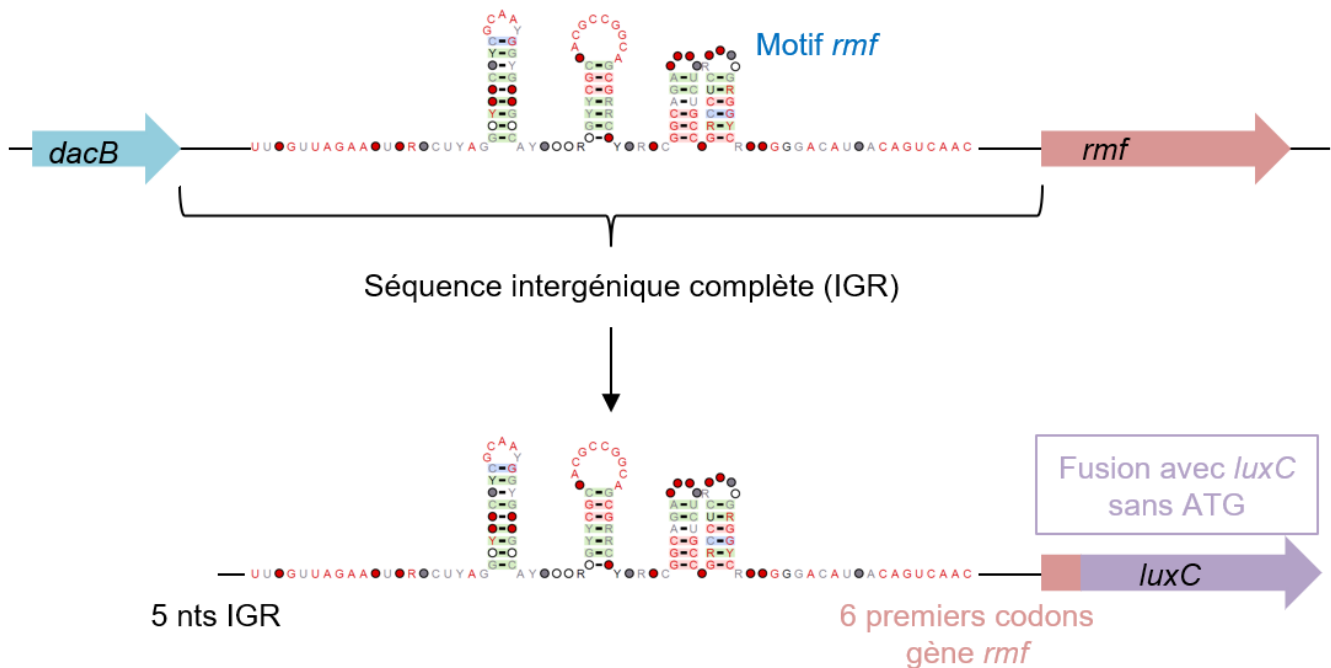


Figure 3.5 Schéma du design utilisé pour la création des plasmides.

Exemple réalisé pour le motif *rmf*, la séquence consensus a été illustrée grâce à R2R (Weinberg & Breaker, 2011). Le motif *rmf* chez *P. aeruginosa* PA14 est situé entre les gènes *dacB* (représenté par la flèche bleue) et *rmf* (représenté par la flèche rouge), la séquence entre les deux gènes constitue la région intergénique complète. Pour le design du plasmide, cinq nucléotides (nts) en amont du motif ainsi que tous les nts en aval du motif jusqu'à l'AUG du gène *rmf* ont été pris. Enfin, les six premiers codons du gène *rmf* ont également été insérés dans le design pour les mettre en fusion avec le gène *luxC* (représenté par la flèche mauve) auquel on a enlevé son propre codon start pour le remplacer par celui de *rmf* (pour cet exemple).

Ce design correspond à la construction appelée WT, qui est le motif présent naturellement dans la bactérie. Une construction disruptive (disr) a également été réalisée où une partie de la structure a été défective. Ces mutations ont été réalisées en se basant sur les structures conservées des différents motifs obtenues à partir de l'alignement des séquences disponibles pour chacun des motifs. Par exemple, une mutation dans une tige a été réalisée qui empêcherait ainsi son repliement naturel. Pour cela, une paire de base a été modifiée afin de perdre les paires de base avec la deuxième partie de la tige boucle. Ceci nous donnera un indice si la structure de l'ARN est importante pour son fonctionnement. Et enfin une construction compensatoire (comp) a été faite, celle-ci a pour but de restaurer la structure suite à la mutation disruptive. Ces différentes constructions ont été comparées entre-elles pour chaque ARNnc candidat au niveau de l'expression de luminescence dans différents milieux.

Plusieurs études récentes ont permis de mettre en évidence plus de 300 ARN non codants potentiels découverts par des méthodes de génomique comparative (Weinberg *et al.*, 2007; Weinberg *et al.*, 2017a; Weinberg *et al.*, 2010). Parmi cette liste, nous avons choisi des motifs spécifiquement présents chez *Pseudomonas* et *Burkholderia* afin de les caractériser en utilisant le vecteur d'expression pVK-f2-lux (Korniakova *et al.*, 2021). Pour *Pseudomonas*, les motifs choisis sont les suivants : *aceE*, *fixA*, *gyrA*, *icd*, *Pseudomon-GGDEF*, *Pseudomon-Rho*, *rmf*, *rnell*, *sucA-II* et *sucC* (Weinberg *et al.*, 2010). Nous avons également sélectionné le motif *Pseudomon-I* mais celui-ci a été caractérisé comme étant un petit ARN régulateur nommé ErsA, qui est impliqué dans plusieurs métabolismes comme le stress de l'enveloppe (Falcone *et al.*, 2018; Ferrara *et al.*, 2015; Ferrara *et al.*, 2021). Pour *Burkholderia*, les motifs choisis sont : MAEB, *sucA*, *gabT*, *traj-II*, *sbcD*, *radC* et *Burkholderiales-1* (Weinberg *et al.*, 2007; Weinberg *et al.*, 2010).

3.4 Optimisation du vecteur d'expression

Les premiers essais de clonage de nos ARNnc candidats ont été réalisés avec le plasmide pVK-f2-lux présenté dans la section 3.3. Cependant, ceux-ci se sont révélés infructueux. En effet, nous ne sommes pas parvenus à digérer le plasmide avec les enzymes de restriction prévues à savoir *ScaI* et *AarI* placées de manière à pouvoir enlever le *riboswitch* fluor de *B. thailandensis* et y insérer nos ARN d'intérêt (Figure 3.6 A), malgré la présence des sites de restrictions correspondants.

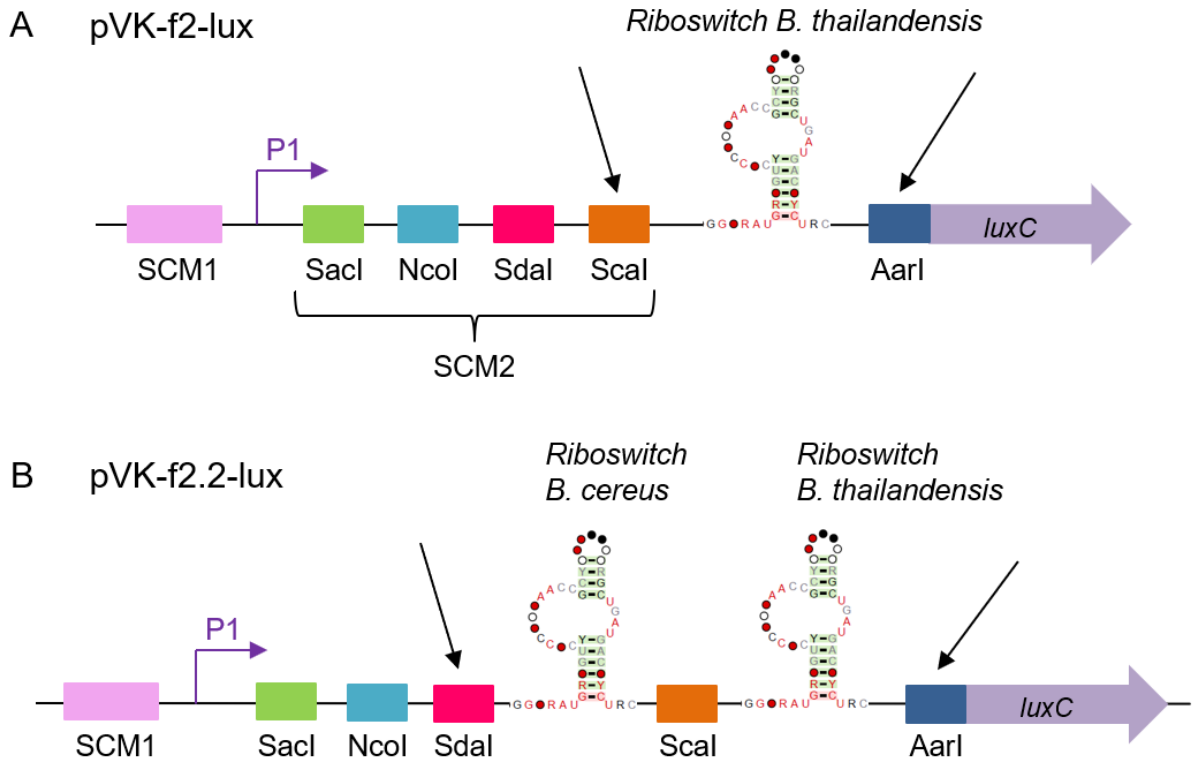


Figure 3.6 Représentation schématique des clonages dans les plasmides pVK.

A : Le plasmide pVK-f2-lux, avec les différentes enzymes de restrictions présentes dans les sites multiples de clonage (SCM 1 et 2) ainsi que l'emplacement du *riboswitch* fluor de *B. thailandensis*. Les flèches indiquent les sites de restrictions utilisés pour enlever le *riboswitch* fluor. **B** : Le nouveau plasmide utilisé avec ces différents composants, nommé pVK-f2.2-lux. Afin de pouvoir enlever les deux copies de *riboswitchs* fluor la digestion doit être faite avec les enzymes SdaI et AarI, indiquées par les flèches.

Nous avons alors opté pour une version du vecteur avec deux *riboswitchs*, celui de *B. cereus*, agissant sur la transcription du gène *crcB* (Baker *et al.*, 2012) et celui de *B. thailandensis*, agissant au niveau de la traduction (Korniakova *et al.*, 2021). Ce plasmide a été testé dans cette étude afin de pouvoir cloner le promoteur fort P_{S7} (Korniakova *et al.*, 2021). Celui-ci a été utilisé dans ce chapitre car les digestions et le clonage ont mieux fonctionné que dans pVK-f2-lux initialement prévu. Dans ce plasmide, le *riboswitch* de *B. thailandensis* se situe au même endroit que dans le plasmide pVK-f2-lux mais le *riboswitch* de *B. cereus* a été inséré au sein du SCM 2 entre les enzymes SdaI et Scal (Figure 3.6 B). Ce plasmide sera nommé pVK-f2.2-lux. Ainsi pour le clonage avec ce nouveau plasmide, il faudra le digérer avec les enzymes SdaI et AarI afin de pouvoir enlever les *riboswitchs* fluor (Figure 3.6 B). Ce plasmide a ensuite permis de faire tous les clonages de nos ARNnc candidats par la méthode de Gibson Assembly.

3.5 Résultats et discussion

3.5.1 Motif *rmf*

Le motif *rmf* se retrouve en amont du gène du même nom chez plusieurs bactéries de la famille des Pseudomonadaceae (Weinberg *et al.*, 2010). Le gène *rmf* code pour le *ribosomal modulation factor*, qui est impliqué dans la dimérisation du ribosome 70S formant ainsi le ribosome 100S (Yamagishi *et al.*, 1993). Ce ribosome 100S est un complexe de deux ribosomes entiers le rendant ainsi inactif et ne permettant plus la traduction de protéines (Polikanov *et al.*, 2012; Yamagishi *et al.*, 1993). Il a été montré que la protéine Rmf liait le ribosome 70S proche du tunnel de sortie de l'ARNm de la sous-unité 30S en interagissant avec l'ARNr 16S (Polikanov *et al.*, 2012). Cette liaison permet un changement de conformation du ribosome où les deux petites sous-unités du ribosome interagissent formant ainsi le dimère (Polikanov *et al.*, 2012). L'action de Rmf est réalisée en collaboration avec l'*hibernation promoting factor* (Hpf) (Polikanov *et al.*, 2012; Ueta *et al.*, 2005). En effet, des études ont montré que Rmf permet la dimérisation de la sous-unité 70S du ribosome formant ainsi le ribosome 90S puis que Hpf permet de stabiliser cette structure en ribosome 100S (Polikanov *et al.*, 2012; Ueta *et al.*, 2005). Le ribosome 100S se forme en phase stationnaire de croissance suite à la limitation en nutriment mais aussi en condition de stress (El-Sharoud & Niven, 2005; El-Sharoud & Niven, 2007; Izutsu *et al.*, 2001). La présence des ribosomes 100S met la bactérie dans un état pouvant ressembler à une hibernation (Yoshida & Wada, 2014). Ainsi, cette structure de ribosome permet de conserver les ribosomes sous une forme inactive, mais ils peuvent être récupérés dès que les conditions s'y prêtent (Yoshida *et al.*, 2002). Le motif *rmf* est annoté dans la base de données Rfam sous la référence RF01755 (Kalvari *et al.*, 2021). La structure de ce motif est assez simple, elle est constituée de quatre tiges boucles (Figure 3.7).

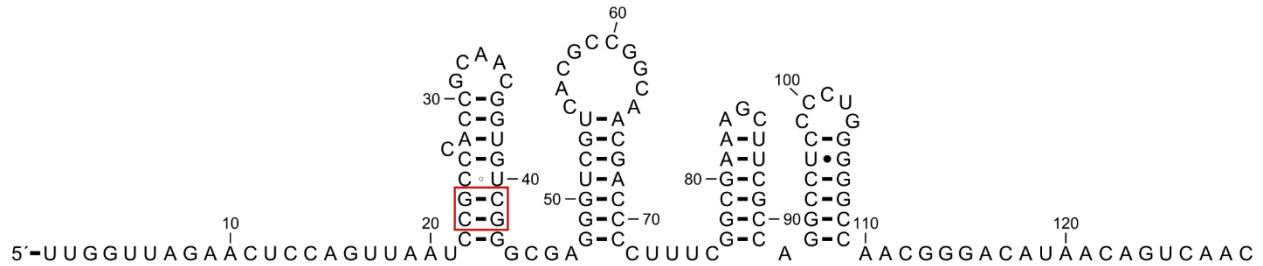


Figure 3.7 Structure du motif *rmf* WT chez *P. aeruginosa* PA14.

Structure réalisée grâce à l'outil R2R en extrayant les séquences du motif *rmf* contenues dans Rfam. Le rectangle rouge correspond à l'endroit où se situe les mutations disruptives (C23G et G24C) et compensatoires (C41G et G42C).

Une étude transcriptomique a été réalisée, par une autre équipe, afin de comparer les différences d'expression de tous les gènes de *P. aeruginosa* PA14 à 28°C vs 37°C (Wurtzel *et al.*, 2012). Ces données ont été obtenues dans du milieu LB en début de phase stationnaire à une DO 600 nm de 1,8 (Wurtzel *et al.*, 2012). Cette étude a permis d'établir une carte précise du transcriptome de cet organisme. Nous pouvons voir la présence de lecture dans l'analyse du séquençage en amont du site de début de transcription de *rmf* qui peut correspondre à la présence du motif en amont du gène activant ainsi la transcription de celui-ci (Figure 3.8). Une différence d'un facteur 2 a été montrée dans la transcription de *rmf* entre les conditions 28°C et 37°C (Wurtzel *et al.*, 2012). Or, il a été montré chez *E. coli* que la protéine Rmf pouvait être impliquée dans plusieurs stress comme les chocs osmotiques ou la résistance au stress chaud (Niven, 2004).

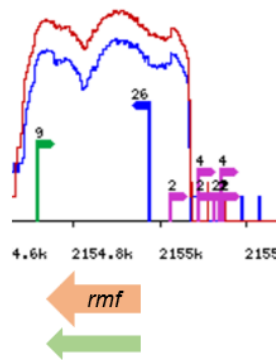


Figure 3.8 Contexte transcriptomique du gène *rmf* dans *P. aeruginosa* PA14.

Représentation schématique des résultats de données de transcriptomique du gène *rmf* (flèche orange). La flèche en bleu en amont du gène *rmf* correspond au début de transcription. La flèche verte correspond au transcrit. Les courbes en bleu et en rouge correspondent à la couverture des lectures de RNA-seq à 28°C en bleu et à 37°C en rouge. Adapté de (Wurtzel *et al.*, 2012).

Le motif *rmf* jusqu'au six premiers codons du gène en aval ont été clonés en fusion traductionnelle avec le gène *luxC* dans le plasmide pVK-f2.2-lux où les deux *riboswitchs* fluor ont été retirés, puis conjugués dans *P. aeruginosa* PA14. Les constructions WT, *dir* et *comp* ont été réalisées. Des essais de luminescence ont été effectués dans du milieu riche (LB) et du milieu chimiquement défini (M9) afin de comparer l'expression de la luminescence dans les différentes conditions. Cet essai a été réalisé avec le Cytation, puis les ratios luminescence / DO600 nm ont été calculés (Figure 3.9).

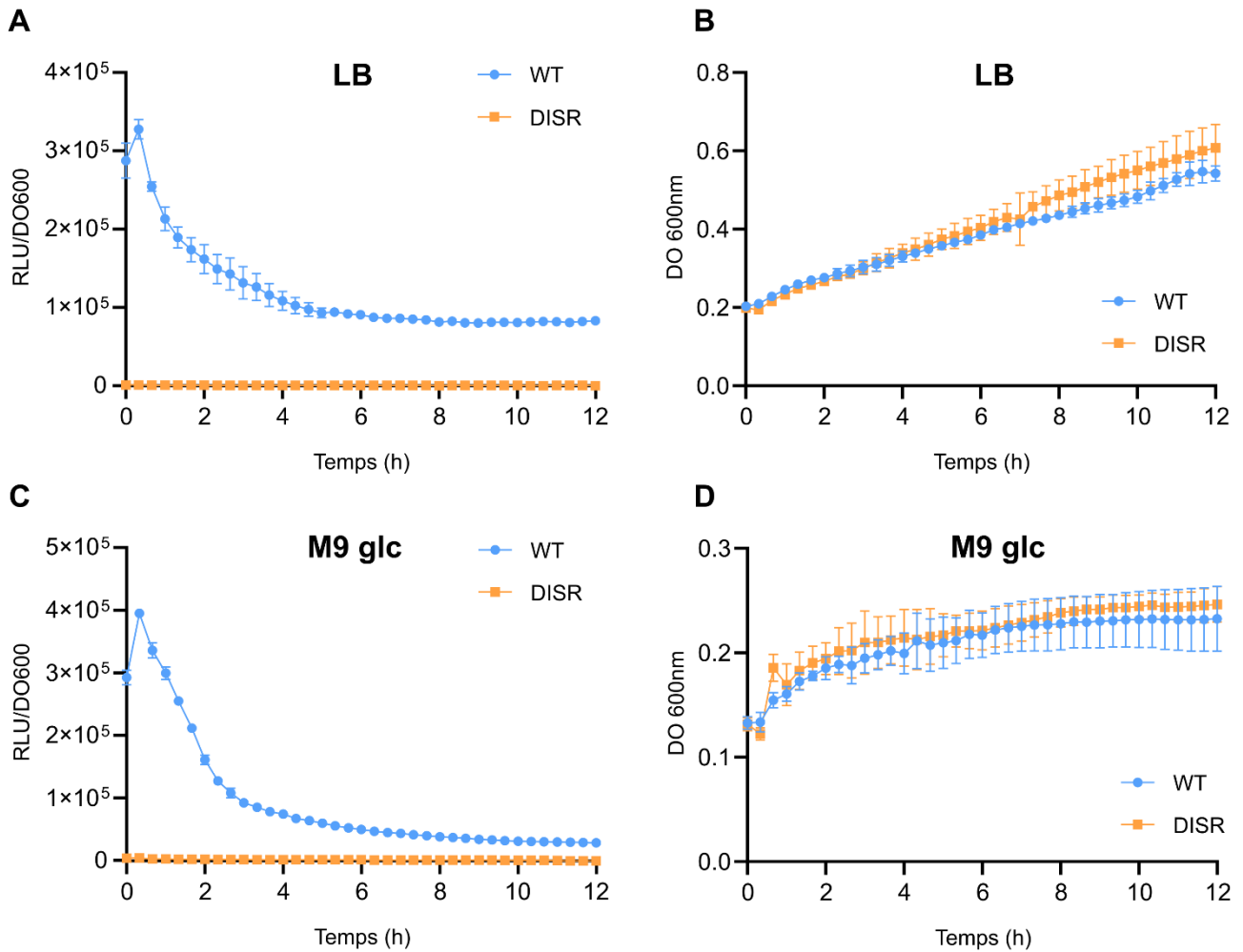


Figure 3.9 Le niveau de luminescence du motif *rmf* WT est plus grand que pour le motif avec les mutations disruptives.

Analyse du niveau de luminescence des constructions *rmf* WT (bleu) et disruptive (orange) dans *P. aeruginosa* PA14 cultivée dans du milieu LB (A) ou dans du milieu M9 (C). Courbe de croissance des constructions *rmf* WT (bleu) et disruptive (orange) dans du milieu LB (B) ou dans du milieu M9 (D). L'intensité de la luminescence et la DO 600nm ont été mesurées en utilisant le Cytation 3 pendant 12h, avec des lectures toutes les 20 min. Le ratio luminescence / DO 600nm a ensuite été calculé.

Pour les deux milieux, nous pouvons remarquer une grande différence du niveau de luminescence entre le motif WT et le motif défaut contenant les mutations *disr* (Figure 3.9). Ces différences observées sont indépendantes de la DO de la culture car les résultats présentés sont le ratio entre la luminescence et la DO 600 nm. En effet, sans cette normalisation, une expression constante (par cellule) mène à l'observation d'une augmentation de la luminescence si la densité de la culture augmente. Ce résultat nous suggère que la structure de l'ARN est importante pour

l'expression du gène en aval. En effet, la seule présence de mutations disruptives entraîne une perte de l'expression de l'opéron *lux*. Afin de confirmer que la structure est essentielle au fonctionnement de ce motif et donc à l'expression du gène en aval, des mutations compensatoires ont été réalisées. Ces dernières permettent de rétablir la paire de bases qui a été affectée lors de la construction défective. Des essais de luminescence ont été effectués grâce à la lectrice à plaque M1000 infinie (TECAN®) dans du milieu chimiquement défini (M9) et du milieu riche (LB) (Figure 3.10).

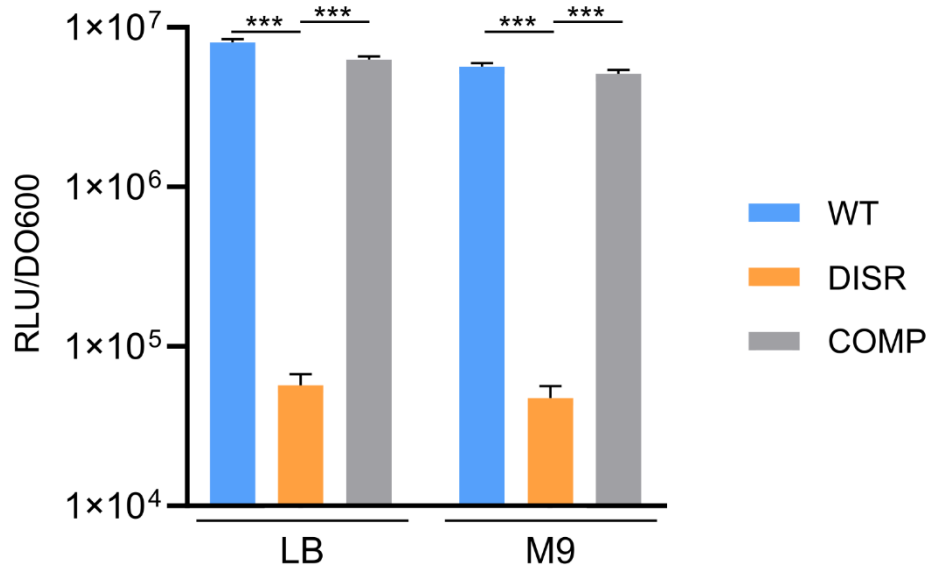


Figure 3.10 La construction compensatoire rétablit le niveau de luminescence à un niveau similaire à la construction WT.

Histogramme du ratio luminescence / DO 600nm pour les constructions du motif *rmf* WT (bleu), disruptive (orange) et compensatoire (gris) cultivée pendant 5h à 37°C dans du milieu riche (LB) ou dans du milieu chimiquement défini (M9). Pour plus de clarté dans les résultats, les valeurs ont été représentés sur une échelle logarithmique base 10. Un t-test multiple non pairé a été réalisé (***) : $p < 0,001$.

Nous pouvons remarquer que le mutant compensatoire (comp) permet de restaurer le niveau de luminescence à un niveau comparable au motif WT dans les deux milieux testés (Figure 3.10). Ces résultats supportent notre hypothèse précédente que la structure est importante pour l'expression du gène en aval. Ces trois constructions ont été testées dans tous les milieux répertoriés dans la section 3.2.5 (Tableau 3.4). La protéine Rmf étant impliquée dans la phase stationnaire et la réponse stringente (Henkin & Peters, 2020; Yoshida & Wada, 2014), des essais avec de nombreux acides aminés ont été réalisés (Tableau 3.4). Ces différents milieux ont par la suite été testés pour tous les autres motifs. Nous avons ainsi pu remarquer que dans tous les

milieux le même profil se dégageait, c'est-à-dire un niveau élevé de luminescence pour les constructions WT et comp, et un niveau faible pour la construction défectueuse. Les résultats des ratios luminescence / DO 600 nm pour tous les milieux sont présentés dans l'Annexe 2. Ces résultats pourraient indiquer que le signal permettant la transduction de signal pour l'activation de *rmf* dans la bactérie est présent quel que soit le milieu.

Chez *E. coli*, il a été montré que la transcription de la protéine Rmf est induite par ppGpp en phase stationnaire (Izutsu *et al.*, 2001). La molécule de ppGpp est l'alarmone connue pour être le messenger principal lors de la réponse stringente. En effet, la présence d'ARNt non chargés, témoins d'une carence en acides aminés, entraîne la synthèse et l'accumulation de ce nucléotide grâce aux enzymes RelA et SpoT (Henkin & Peters, 2020). Il a été montré qu'une souche d'*E. coli* délétée des gènes *relA* et *spoT* ($\Delta relA \Delta spoT$) donc incapable de synthétiser les molécules de ppGpp n'exprime pas la protéine Rmf et est incapable de faire des ribosomes 100S en phase stationnaire (Izutsu *et al.*, 2001). Or, une étude précédente a montré l'existence d'un *riboswitch* liant le ppGpp (Sherlock *et al.*, 2018a). Ce *riboswitch* est présent devant une multitude de gènes impliqués dans le métabolisme des acides aminés, comme la biosynthèse des acides aminés branchés (BCAA), ou des glutamate synthases (Sherlock *et al.*, 2018a).

Nous avons ainsi émis l'hypothèse que le motif *rmf* pourrait être une nouvelle classe de *riboswitch* ppGpp dans les bactéries du genre *Pseudomonas*. Afin de vérifier cette hypothèse, nous avons conjugué nos plasmides dans une souche de *P. aeruginosa* PA01 $\Delta relA \Delta spoT$ ne produisant pas de ppGpp (Nguyen *et al.*, 2011). Nous n'avons accès qu'à ce mutant dans la souche de *P. aeruginosa* PA01, une autre souche de référence de *Pseudomonas aeruginosa*. Ainsi, afin de s'assurer de la comparaison entre les souches nous avons également conjugué nos plasmides dans une souche de *P. aeruginosa* PA01 WT. Cette souche possède naturellement également le motif *rmf*. Des essais de luminescence ont été réalisés une nouvelle fois dans du milieu riche (LB) et du milieu chimiquement défini (M9) (Figure 3.11).

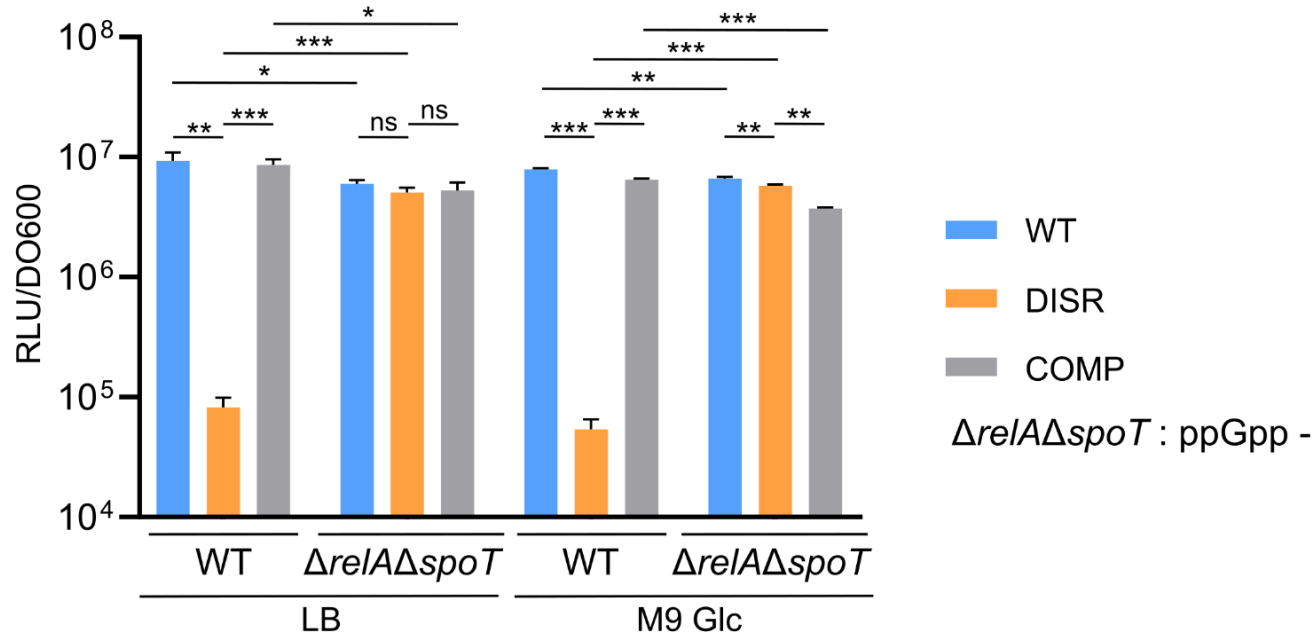


Figure 3.11 Le profil du niveau de luminescence est modifié dans la souche PA01 $\Delta relA \Delta spoT$.

Histogramme du ratio luminescence / DO 600nm pour les constructions *rmf* WT (bleu), disruptive (orange) et compensatoire (gris) dans les souches de *P. aeruginosa* PA01 WT et PA01 $\Delta relA \Delta spoT$ cultivées dans un milieu riche (LB) et dans un milieu chimiquement défini (M9) pendant 5h à 37°C en agitation. La luminescence et la DO 600nm ont été mesurées grâce à la lectrice à plaque M1000 infinie (TECAN®). Pour plus de clarté dans les résultats, les valeurs ont été représentés sur une échelle logarithmique base 10. Un t-test multiple non pairé a été réalisé (ns : non significatif ; * : $p < 0,05$; ** : $p < 0,01$; *** : $p < 0,001$).

Concernant la souche WT de *P. aeruginosa* PA01, nous pouvons observer le même profil que dans le cas de la souche PA14, c'est-à-dire que dans le cas du mutant défaut présentant les mutations disruptives il y a un niveau faible de luminescence comparé au WT et comp, ainsi qu'une restauration du niveau de luminescence proche du WT pour la construction compensatoire. Il est à noter que dans le cas du milieu M9 la restauration du compensatoire n'est pas complète. Concernant la souche de *P. aeruginosa* PA01 $\Delta relA \Delta spoT$, nous pouvons remarquer que le niveau de luminescence est statistiquement différent entre les différentes constructions testées (WT, discr et comp) comparé à la souche PA01 WT (Figure 3.11). Cependant, le profil du niveau de luminescence des constructions WT et comp est très semblable par rapport à la souche WT (Figure 3.11). De manière intéressante, nous pouvons voir que dans cette souche, quel que soit le milieu testé, (LB ou M9) le mutant défaut présente un niveau de luminescence comparable aux constructions WT et comp (Figure 3.11).

Comme indiqué précédemment, notre hypothèse est que le motif *rmf* serait un *riboswitch* ppGpp, or cette alarmone de la réponse stringente est exprimée au niveau de la phase stationnaire (Henkin & Peters, 2020). Une cinétique a ainsi été réalisée en milieu LB et en milieu M9 afin de vérifier les différences de niveau de luminescence et donc de Rmf au cours de la croissance de la bactérie (Figure 3.12). Pour cela, les souches de *P. aeruginosa* PAO1 WT et $\Delta relA \Delta spoT$ ont été cultivées dans ces deux milieux pendant 80 h à 37°C et différentes lectures de la luminescence et de la DO 600 nm ont été réalisées à 5h, 24h, 30h, 48h, 54h, 72h et 80h de croissance. Pour les essais avec le milieu LB dans la souche PAO1 WT, le même profil était observable, à savoir une expression forte de l'opéron *lux* pour les constructions WT et comp ainsi qu'une expression beaucoup plus faible pour le mutant *disr* (Figure 3.12 A, WT). Pour la souche PAO1 $\Delta relA \Delta spoT$, les trois constructions montrent un niveau de luminescence similaire (Figure 3.12 A, $\Delta relA \Delta spoT$). Cependant, on peut voir que les bactéries dans le LB n'ont pas encore tout à fait atteint leur phase stationnaire (Figure 3.12 C), ceci pourrait ainsi expliquer l'absence de différence au cours du temps.

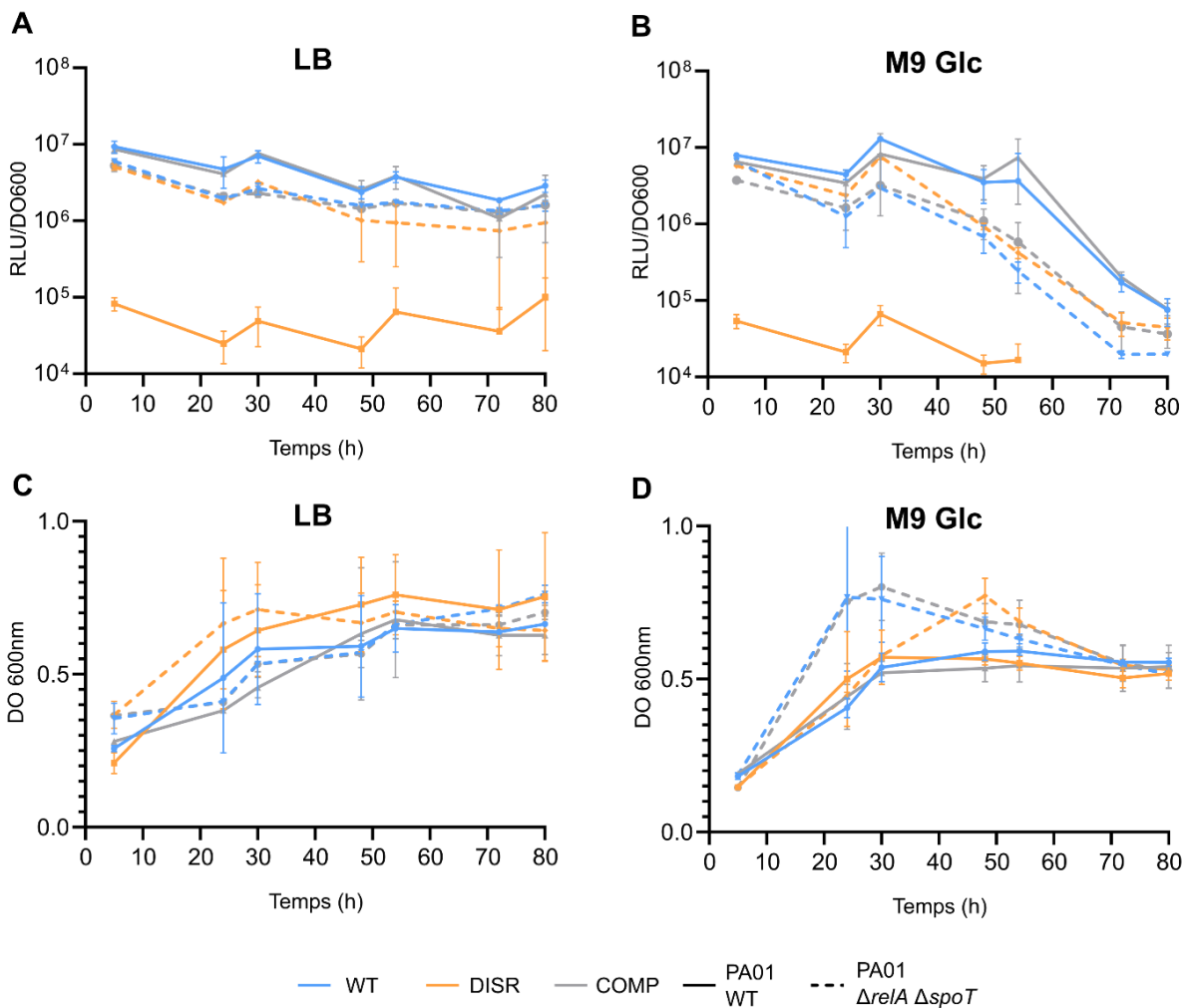


Figure 3.12 L'expression de Rmf diminue en phase stationnaire.

A et B : Cinétique de l'expression de la luminescence (ratio luminescence/DO600 nm) pour les constructions *rmf* WT (bleu), disruptive (orange) et compensatoire (gris) dans les souches de *P. aeruginosa* PA01 (traits pleins) et *P. aeruginosa* PA01 $\Delta relA \Delta spoT$ (traits pointillés) cultivés dans **A** : du milieu riche (LB) ou **B** : du milieu chimiquement défini (M9 glucose). Pour plus de clarté dans les résultats, les valeurs ont été représentées sur une échelle logarithmique base 10. **C et D :** Courbe de croissance à DO600 nm des différentes souches testées dans **C** : du milieu riche (LB) ou **D** : du milieu chimiquement défini (M9 glucose).

Par contre, pour le milieu M9, nous pouvons voir pour les constructions WT et comp dans PA01 WT un niveau élevé de luminescence jusqu'à 54h, suivi d'une diminution très marquée de celui-ci à 72h et 80h, ainsi qu'une absence d'expression de l'opéron *lux* pour le mutant *disr* tel qu'observé précédemment (Figure 3.12 B). Le même profil est observable avec le milieu M9 qu'avec le LB pour la souche $\Delta relA \Delta spoT$. En concordance avec les courbes de croissance de ces souches (Figure 3.12 D), ces résultats laissent à penser que l'expression de Rmf est inhibée

en phase stationnaire aussi bien en absence qu'en présence de ppGpp. Cependant, il est possible que cette forte diminution à partir de 72h ne soit pas uniquement due à une diminution de l'expression, mais aussi au fait d'une croissance très longue des bactéries pour lesquelles les ressources nutritionnelles viendraient à manquer.

Ces résultats ne coïncident pas avec notre hypothèse initiale, à savoir que *rmf* serait un *riboswitch* ppGpp. En effet, dans la souche PAO1 WT le niveau de luminescence est haut. Cela indique que dans ces conditions le potentiel *riboswitch* permet une expression du gène et que le mutant disruptif entraîne une inactivation de ce dernier, du fait du faible niveau de luminescence observé. Ainsi, compte tenu de notre hypothèse nous nous serions attendus à ne pas avoir de différence dans le niveau de luminescence entre la souche WT et la souche $\Delta relA \Delta spoT$ pour la construction défaite contenant les mutations disruptives car celle-ci supposément ne permettrait pas la liaison de ppGpp. De plus, du fait de l'absence de son ligand potentiel nous nous serions également attendus à des changements dans le niveau de luminescence par rapport à la souche PAO1 WT aussi bien pour la construction WT que la construction compensatoire.

Compte tenu de notre hypothèse concernant la possibilité que le motif *rmf* soit un *riboswitch* ppGpp, des essais de cartographie *in vitro* par *in-line* ont été réalisés pour le motif *rmf*. La méthode de cartographie *in vitro* par *in-line* est une méthode d'étude de la structure des ARN basée sur la dégradation naturelle des ARN en fonction de leur structure (Regulski & Breaker, 2008). Cette dégradation se fait grâce à une attaque nucléophile de l'oxygène 2' sur le centre de phosphore quand l'oxygène 2', le phosphore et l'oxygène 5' sont en conformation en ligne (*in-line*) (Regulski & Breaker, 2008). Les structures double brin compte tenu de leur manque de flexibilité seront ainsi protégées du clivage alors que les structures simple brin seront clivées. Cette méthode est couramment utilisée pour la caractérisation des *riboswitchs*. Compte tenu de la propriété des *riboswitchs* à changer de conformation suite à la liaison de leur ligand, cette liaison provoquera un changement dans le profil de dégradation de l'ARN en comparant la condition sans et avec le ligand (Figure 3.13).

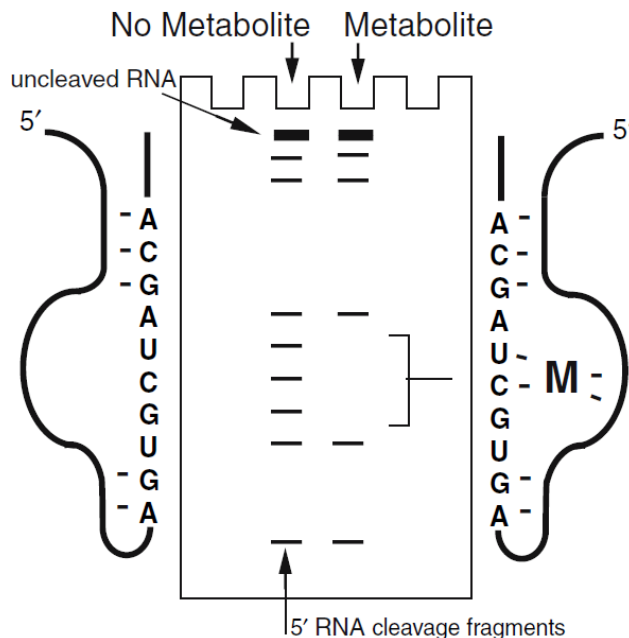


Figure 3.13 Principe de la méthode de cartographie *in vitro* par *in-line*.

Représentation schématique de la méthode de cartographie *in vitro* par *in-line*. Les ARN marqués à la radioactivité à leur extrémité 5' sont incubés avec ou sans leur ligand pendant 40h puis séparés sur un gel de polyacrylamide dénaturant. Les ARN dans cette méthode sont dégradés en fonction de leur structure, où les parties simple brin seront soumises au clivage comparé aux structures double brin qui sont moins flexibles. L'interaction entre le ligand (métabolite représenté par la lettre M) et le *riboswitch* provoque un changement de conformation de ce dernier entraînant ainsi un profil de dégradation différent avec ou sans le ligand. Tiré de (Regulski & Breaker, 2008).

Quatre constructions du motif *rmf* ont été étudiées par *in-line* avec différentes concentrations de ppGpp et pppGpp. Les séquences testées ont été répertoriées dans le Tableau 3.5. Tout d'abord, les essais de cartographie *in vitro* par *in-line* nous ont permis de réaliser des cartographies de la structure de l'ARN *rmf* (Figure 3.14).

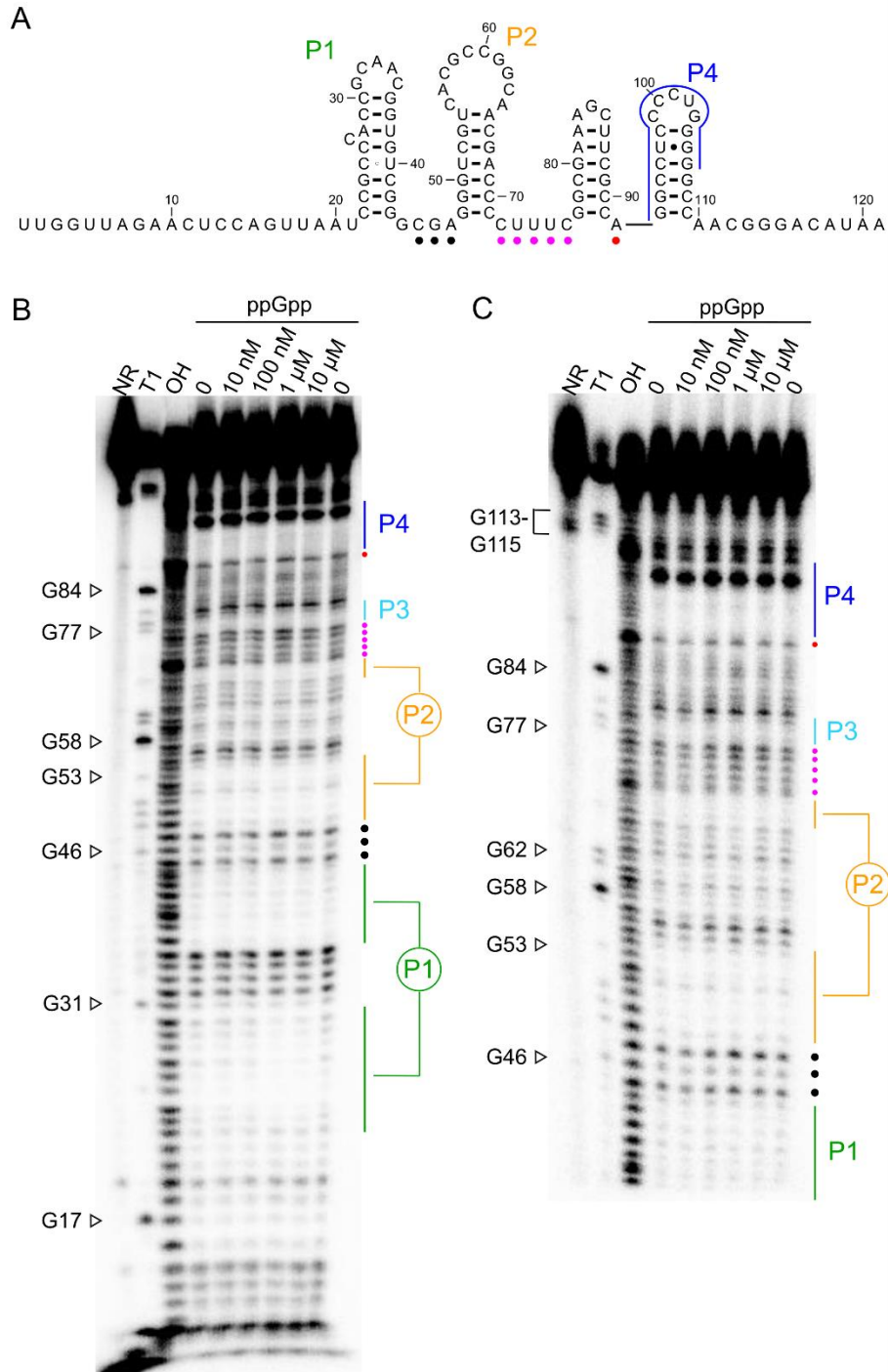


Figure 3.14 Confirmation de la structure du motif *rmf*.

A : Structure annotée de la construction 4 du motif *rmf* basée sur l'alignement de plusieurs séquences. **B** et **C** : Gels de *in-line* pour la construction 4 du motif *rmf* avec des concentrations de ppGpp de 10 nM à 10 μ M. La même construction a été testée pour ces deux gels, cependant pour le gel en C la migration a été prolongée de 2h afin d'avoir une meilleure résolution au niveau du haut du gel. NR, T1 et OH représentent les échelles d'ARN non réagi, d'ARN traité à l'ARNase T1 (clive après les G) et d'ARN soumis à un traitement alcalin (clive après chaque nucléotides) respectivement. Les régions correspondant à la structure prédite de *rmf* sont représentées par des barres ou des points suivant l'annotation en A.

Le profil de dégradation de *rmf* obtenu lors de ces essais a été comparé à la structure prédite de *rmf* issu de l'alignement de plusieurs séquences (Figure 3.14 A). Nous pouvons remarquer que la tige boucle 1 (P1) et les structures simple brin entourant les tiges boucles 1 et 2 sont bien présentes dans notre cartographie (Figure 3.14 B et C). Concernant la tige boucle 2 (P2), nous pouvons voir que la partie en 5' ainsi qu'une partie de la boucle sont bien visibles (Figure 3.14 B et C). Cependant, la boucle et la partie 3' de la tige sont plus difficilement identifiables (Figure 3.14 B et C). Concernant la tige boucle 3 et 4, il est aussi plus difficile de les identifier clairement sur notre cartographie (Figure 3.14 B et C). Notre hypothèse étant que *rmf* soit un *riboswitch*, il est possible que la structure que nous voyons sur nos cartographies soit une structure alternative de celle prédite pouvant ainsi expliquer pourquoi elles ne correspondent pas parfaitement.

Une construction du motif *rmf* *disr*, contenant les mutations disruptives, a été testée afin de valider que la présence des mutations entraînait bien un changement dans la structure de *rmf* (Figure 3.15).

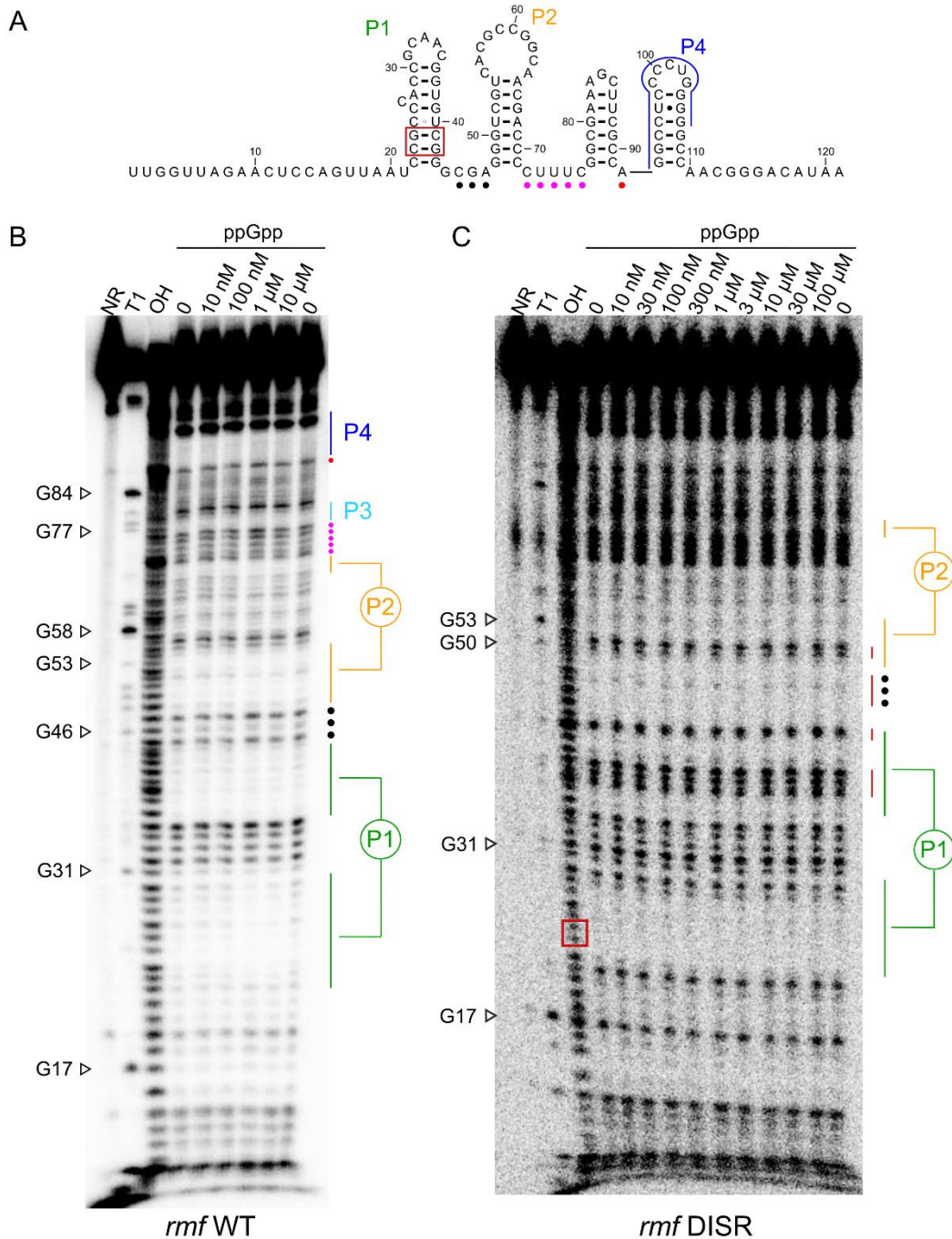


Figure 3.15 Les mutations disruptives entraînent un changement de conformation de l'ARN *rmf*.

A : Structure de la construction 4 du motif *rmf* basée sur l'alignement de plusieurs séquences réalisée avec R2R. Le rectangle rouge correspond à la position des mutations disruptives (C23G et G24C). **B** : Gel de cartographie *in vitro* par *in-line* de la construction 4 du motif *rmf* WT. Les régions correspondantes à la structure prédite de *rmf* sont représentées par des barres ou des points suivant l'annotation en A. **C** : Gels de *in-line* pour la construction 4 du motif *rmf* *disr*, contenant les mutations disruptives représentées par le rectangle rouge dans l'échelle OH. NR, T1 et OH représentent les échelles d'ARN non réagi, d'ARN traité à l'ARNase T1 (clive après les G) et d'ARN soumis à un traitement alcalin (clive après chaque nucléotides) respectivement. Les régions correspondant à la structure prédite de *rmf* sont représentées par des barres ou des points suivant l'annotation en A. Les régions ne correspondant pas à la structure de *rmf* WT sont représentées par les barres rouges.

Nous pouvons remarquer que plusieurs parties de la structure de la construction disruptive sont modifiées par rapport à la construction du motif *rmf* WT (Figure 3.15 B et C). Ce résultat nous permet de valider que la stratégie utilisée pour la création des mutants défauts contenant les mutations disruptives est efficace.

Néanmoins, le but principal de la cartographie *in vitro* par *in-line* est de visualiser si l'ARN interagit avec son ligand entraînant un changement de conformation de celui-ci, se traduisant par un profil de dégradation différent. Plusieurs essais ont été effectués pour la construction 4 du motif *rmf* avec le ppGpp et le pppGpp (Figure 3.14 et 3.16).

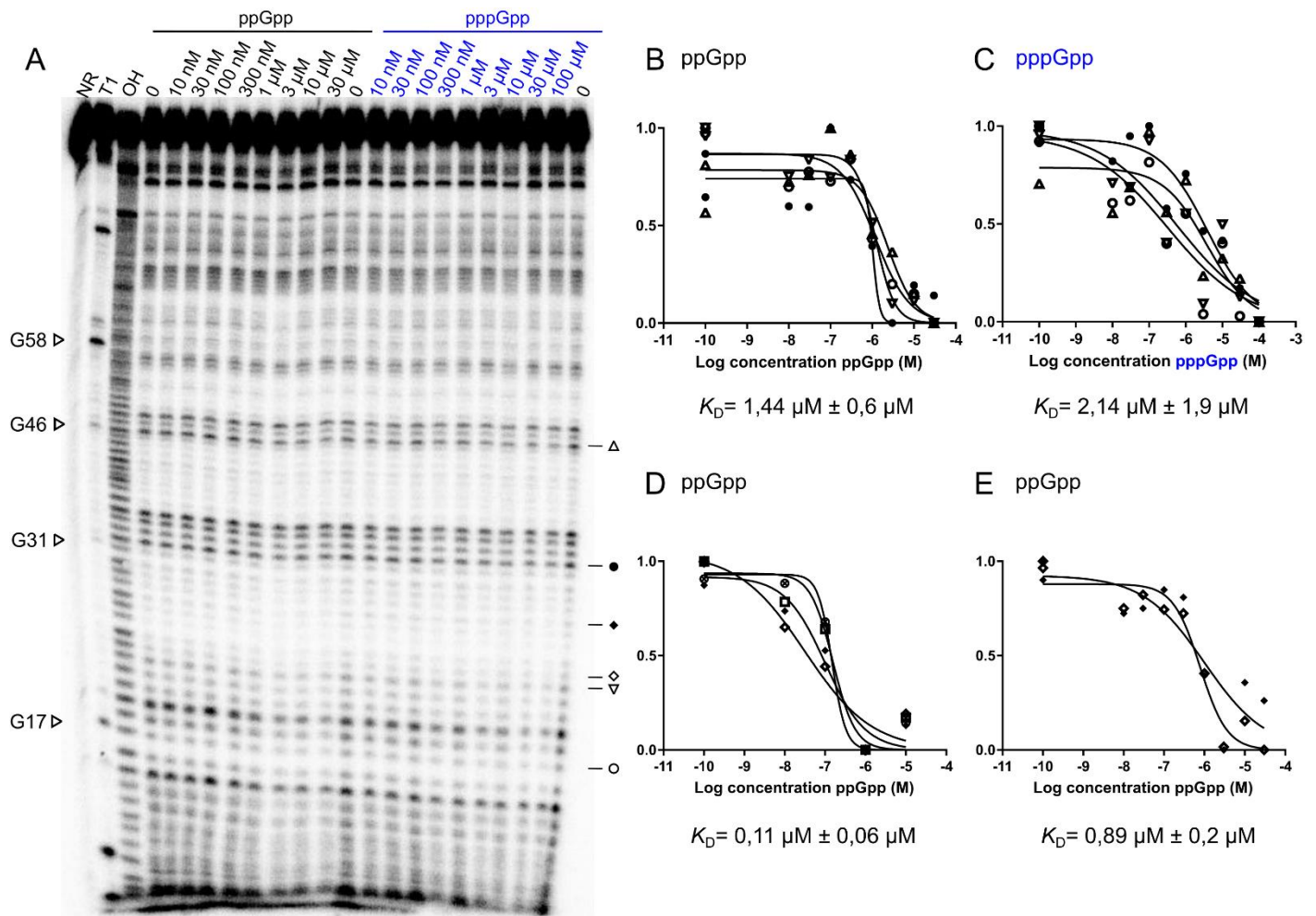


Figure 3.16 L'ARN *rmf* semble lier le ppGpp et le pppGpp.

A : Gel de cartographie *in vitro* par *in-line* de la construction *rmf* 4 WT avec des concentrations de 10 nM à 30 μ M de ppGpp et de 10 nM à 100 μ M de pppGpp. NR, T1 et OH représentent les échelles d'ARN non réagi, d'ARN traité à l'ARNase T1 (clive après les G) et d'ARN soumis à un traitement alcalin (clive après chaque nucléotides) respectivement. **B** : Courbe de K_D associée au gel présenté en A pour le ppGpp. **C** : Courbe de K_D associée au gel présenté en A pour le pppGpp. **D** : Courbe de K_D associée au gel présent dans la Figure 3.14 B. **E** : Courbe de K_D des bandes communes entre le gel présenté en A et celui de la Figure 3.14 B. Les symboles correspondant aux bandes utilisées pour déterminer la constante d'affinité de l'ARN pour le ligand testé sont représentés sur le côté du gel. La concentration 10^{-10} correspond à la condition sans ligand car il n'est pas possible de mettre la valeur 0 dans une échelle logarithmique.

Nous pouvons observer pour cette construction une modulation très subtile de l'intensité de certaines bandes en lien avec l'augmentation de la concentration de ppGpp ou pppGpp (Figure 3.16 A). Ces bandes ont été utilisées afin de déterminer les constantes de dissociation (K_D) pour les deux ligands testés (Figure 3.16 B et C). Un K_D moyen de $1,44 \mu\text{M} \pm 0,6 \mu\text{M}$ pour le ppGpp (Figure 3.16 B) et un K_D moyen de $2,14 \mu\text{M} \pm 1,9 \mu\text{M}$ pour le pppGpp (Figure 3.16 C) ont

été quantifiés. Afin de confirmer ces résultats, des répliques de la même construction ont été faits. Un des gels réalisé est présenté en Figure 3.14 B où l'on peut observer une modulation encore une fois très subtile de l'intensité de certaines bandes. Pour ce gel, nous avons été en mesure de déterminer un K_D moyen de $0,11 \mu\text{M} \pm 0,06 \mu\text{M}$ pour le ppGpp (Figure 3.16 D). Les courbes de K_D pour les autres gels sont présentes en Annexe 4. Nous sommes ainsi parvenus à déterminer des valeurs de K_D lors de quatre expériences indépendantes pour cette construction de l'ARN rnf (*rnf 4*) et une autre pour la construction *rnf 1*. Cependant, ces modulations sont à chaque fois très subtiles et nous pouvons également remarquer que ce ne sont pas toujours les mêmes bandes qui modulent dépendamment des gels. Afin de comparer les différentes modulations obtenues, les mêmes bandes que celles montrant une modulation dans la Figure 3.14 B ont été utilisées afin de calculer un K_D pour l'essai présenté dans la Figure 3.16. Deux de ces bandes ont permis de déterminer un K_D moyen de $0,89 \mu\text{M} \pm 0,2 \mu\text{M}$ (Figure 3.16 E). Ainsi, pour tous nos essais les valeurs de K_D obtenues sont comprises aux alentours de $1 \mu\text{M}$ avec un minimum de $0,11 \mu\text{M}$ et un maximum de $1,44 \mu\text{M}$ (Annexe 4).

Ces résultats semblent indiquer une interaction entre l'ARN rnf et le (p)ppGpp. Néanmoins, son mécanisme d'action précis reste encore incertain.

3.5.2 Motif *Pseudomon*-GGDEF

Le motif *Pseudomon*-GGDEF est retrouvé exclusivement chez des bactéries du genre *Pseudomonas* (Greenlee *et al.*, 2018; Weinberg *et al.*, 2010). La bactérie *P. syringae* pv. *tomato* DC3000 présente ce motif deux fois, un présent dans son chromosome et un présent sur un plasmide nommé pDC3000A (Altschul *et al.*, 1990). Ces deux motifs sont situés en amont de gènes présentant des domaines semblables (Greenlee *et al.*, 2018). Il existe 65 représentants de ce motif qui sont toujours situés en amont d'un gène codant pour une famille de protéines appelée COG5001 (Greenlee *et al.*, 2018). Les membres de cette famille contiennent généralement des domaines GGDEF et EAL qui fonctionnent comme diguanylate cyclase ou phosphodiesterase, respectivement, pour synthétiser ou dégrader les molécules de di-GMPc (Greenlee *et al.*, 2018). Le motif que nous avons choisi chez *P. syringae* pv. *tomato* DC3000 est présent dans son chromosome en amont du locus PSPTO_2907. Celui-ci se retrouve exprimé en opéron avec un répresseur transcriptionnel VanR de la famille des GntR impliqué dans le catabolisme du vanillate (Tropel & Van Der Meer, 2004; Weinberg *et al.*, 2010). La structure du motif *Pseudomon*-GGDEF a été réalisée grâce à l'alignement proposé dans l'étude de la découverte de ce motif avec l'outil

R2R (Weinberg & Breaker, 2011; Weinberg *et al.*, 2010). C'est un motif relativement simple composé de deux tiges boucles (Figure 3.17).

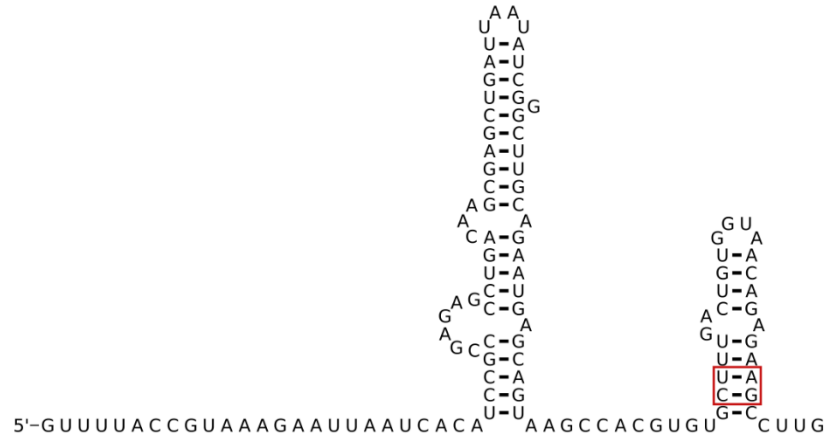


Figure 3.17 Structure du motif *Pseudomon-GGDEF WT* de *Pseudomonas syringae* pv. *tomato* DC3000.

Structure réalisée grâce à l'outil R2R à partir de la séquence présente dans l'article (Weinberg *et al.*, 2010) et le consensus de la structure. Le rectangle rouge correspond à l'endroit des mutations disruptives (C116U et U117C).

Les constructions WT et disr ont été amplifiées à partir du génome de *P. syringae* pv. *tomato*, clonées dans le plasmide pVK-f2.2-lux et conjuguées dans *P. aeruginosa* PA14. Des essais de conjugaison ont été réalisés avec la bactérie contenant naturellement ce motif, à savoir *P. syringae* pv. *tomato* DC3000. Pour cette conjugaison, le triclosan qui permet la contre sélection avec *E. coli* dans le cas de *P. aeruginosa* PA14 a été remplacé par de la rifampicine à 50 µg/mL (Scalschi *et al.*, 2014) et l'incubation a été effectuée à 30°C au lieu de 37°C, correspondant à l'intervalle de température permettant la croissance de *P. syringae* (Young *et al.*, 1977). Cependant, ces essais se sont révélés infructueux. Étant deux bactéries du même genre, nous avons décidé de conjuguer ces plasmides au sein de *P. aeruginosa* PA14 qui a été cultivée à 30°C au lieu de 37°C pour rester dans la température de croissance de *P. syringae* pv. *tomato* DC3000. Les essais de luminescence ont été réalisés dans les conditions listées dans le Tableau 3.4 (Figure 3.18).

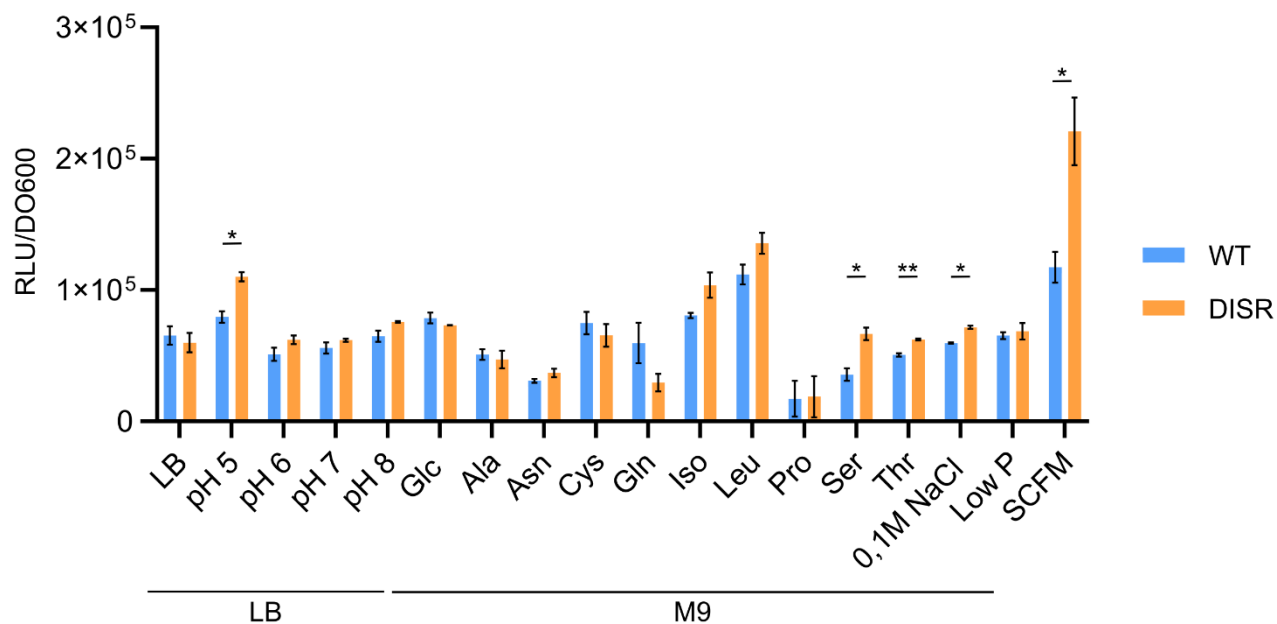


Figure 3.18 Des différences du niveau de luminescence entre les constructions WT et disr sont observables pour certains milieux testés pour le motif *Pseudomon*-GGDEF.

Histogramme des ratios luminescence / DO 600 nm pour les constructions du motif *Pseudomon*-GGDEF WT (bleu) et disruptive (orange) cultivées pendant 5h à 37°C en agitation. Les milieux de culture testés sont en milieu riche : LB, pH5 (LB pH 5), pH 6 (LB pH 6), pH 7 (LB pH 7), pH 8 (LB pH 8), et en milieux chimiquement définis : Glc (M9 glucose), Ala (M9 alanine), Asn (M9 asparagine), Cys (M9 cystéine), Gln (M9 glutamine), Iso (M9 isoleucine), Leu (M9 leucine), Pro (M9 proline), 0,1 M NaCl (M9 + 0,1 M NaCl), low P (M9 avec une faible concentration de phosphate) et SCFM. Un t-test multiple non pairé a été réalisé (* : $p < 0,05$; ** : $p < 0,01$).

On peut observer dans le cas de *Pseudomon*-GGDEF qu'une grande partie des milieux testés ne montrent pas de différence dans le niveau de luminescence entre les constructions WT et défaite (Figure 3.18). Les résultats des essais de luminescence pour les milieux non représentés dans ce graphique sont présents en Annexe 5. Pour certains milieux comme le LB pH 5, le M9 supplémenté en sérine ou thréonine ainsi que le SCFM, nous pouvons voir une différence statistiquement significative du niveau de luminescence entre les deux constructions (WT et disr). En effet, la construction défaite contenant les mutations disruptives montre un niveau de luminescence plus fort que la construction WT (Figure 3.18). Ces différences nous indiquent que la structure de l'ARN semble être importante pour la régulation de l'expression du gène en aval. Nous pouvons également remarquer dans le milieu M9 supplémenté en glutamine que la construction WT montre un niveau de luminescence plus élevé que la construction défaite contrairement aux autres conditions montrant des différences (Figure 3.18). Cependant, cette disparité étant un cas unique dans les milieux testés et ne présentant pas de différence statistiquement significative, il est difficile d'en tirer des conclusions à cette étape.

Le motif *Pseudomon*-GGDEF étant situé en amont d'un gène contenant des domaines GGDEF et EAL impliqués dans le métabolisme du di-GMPc (Greenlee *et al.*, 2018). Nous nous sommes demandé si cette molécule ne pourrait pas avoir un effet sur le motif ARNnc candidat, surtout en sachant que des *riboswitchs* existent pour le di-GMPc (Lee *et al.*, 2010; Sudarsan *et al.*, 2008). Pour cela, nous avons conjugué le plasmide contenant les constructions WT et *disr* dans des souches de *P. aeruginosa* PA14 exprimant différents niveaux de di-GMPc, à savoir *wspF*- et *wspR*-. Ces souches sont issues d'une étude réalisée par une autre équipe via une banque de transposons (Liberati *et al.*, 2006). Les gènes *wsp* sont impliqués dans la cascade de signalisation permettant la synthèse de di-GMPc (Hickman *et al.*, 2005). L'étude a montré qu'une souche Δ *wspF* entraîne une augmentation du niveau de di-GMPc chez *P. aeruginosa* (Hickman *et al.*, 2005). Le gène *wspF* code pour une méthylestérase agissant comme répresseur de WspR, en le déphosphorylant, ainsi en absence de répresseur l'expression de di-GMPc est augmentée (Hickman *et al.*, 2005). Le gène *wspR* catalyse la formation de di-GMPc une fois phosphorylée (Güvener & Harwood, 2007; Hickman *et al.*, 2005). Ainsi, en son absence le niveau de di-GMPc est diminué. Des essais de luminescence ont été réalisés dans les trois souches et les ratios luminescence / DO 600 nm ont été calculés (Figure 3.19). Pour une meilleure visibilité les résultats ont été présentés sur une échelle logarithmique base 10.

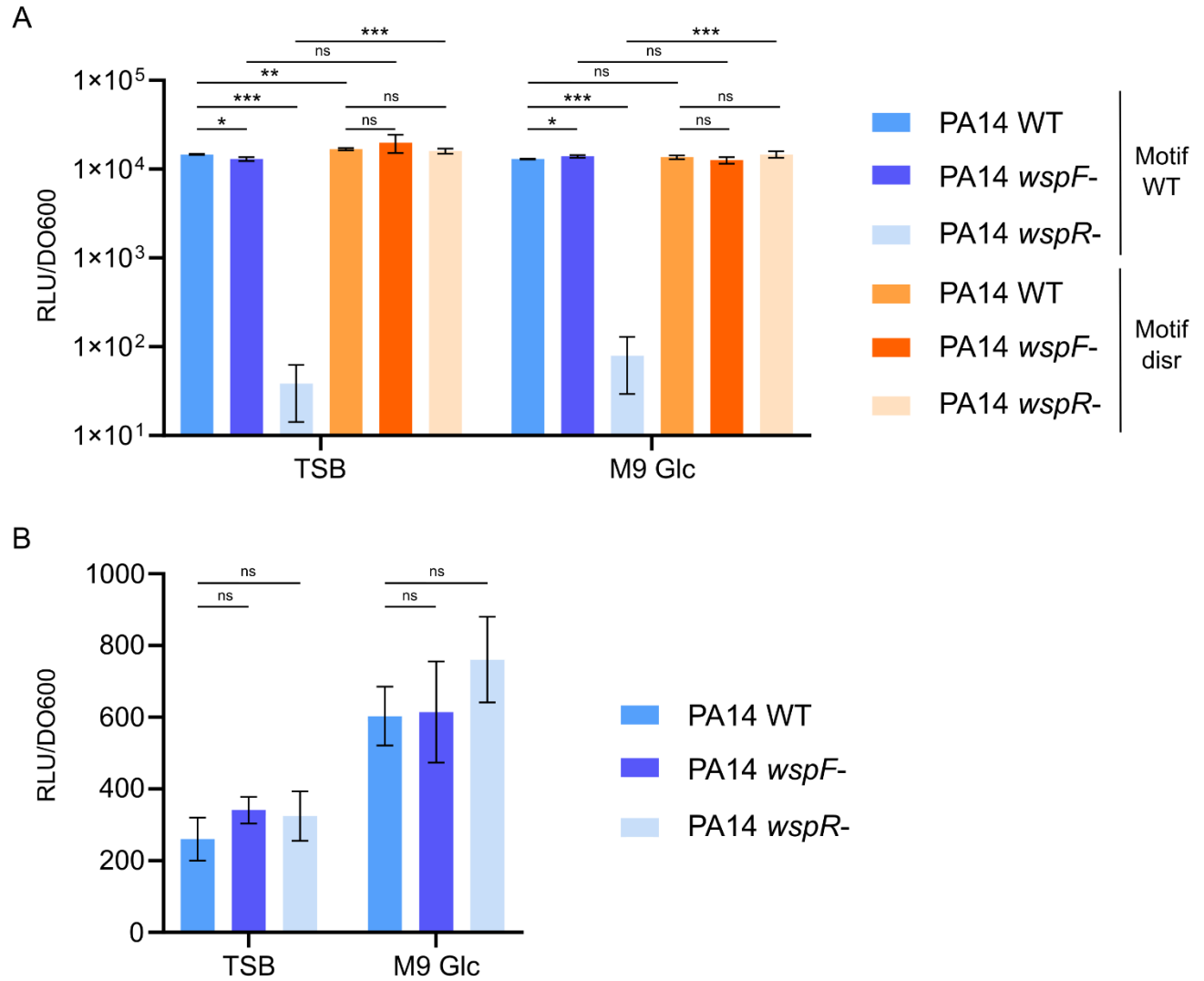


Figure 3.19 Le niveau de luminescence est fortement diminué dans la souche de *P. aeruginosa* exprimant un niveau faible de di-GMPc.

A : Histogramme des ratios luminescence / DO 600 nm pour les constructions du motif *Pseudomon*-GGDEF WT (bleu) et disruptif (orange) dans la souche PA14 WT (bleu ou orange), la souche PA14 *wspF*- produit plus de di-GMPc (bleu ou orange foncé) et la souche PA14 *wspR*- produit moins de di-GMPc (bleu ou orange clair) cultivées pendant 5h à 37°C en agitation **B** : Histogramme des ratios luminescence / DO 600 nm pour la construction *mne-II*/WT dans la souche PA14 WT (bleu), la souche PA14 *wspF*- produit plus de di-GMPc (bleu foncé) et la souche PA14 *wspR*- produit moins de di-GMPc (bleu clair). Ces tests ont été effectués en milieu riche (TSB) et en milieu chimiquement défini (M9 glucose). Un t-test multiple non pairé a été réalisé (ns : non significatif ; * : $p < 0,05$; ** : $p < 0,01$; *** : $p < 0,001$).

Nous pouvons tout d'abord remarquer que le profil du niveau de luminescence est similaire entre les deux milieux testés à savoir le milieu riche TSB et le milieu chimiquement défini M9 glucose (Figure 3.19). Nous pouvons également voir pour la construction WT que le niveau de luminescence est semblable entre les souches PA14 WT et PA14 *wspF*- produisant plus de di-

GMPc (Figure 3.19 A). Cependant, nous pouvons observer dans la souche *wspR*-, produisant moins de di-GMPc, que le niveau de luminescence est diminué de manière significative par rapport aux deux autres souches (Figure 3.19 A). Concernant la construction défaite contenant les mutations disruptives, nous pouvons voir que quel que soit la souche testée (WT, *wspF*- et *wspR*-) le niveau de luminescence est semblable (Figure 3.19 A). Ces résultats semblent montrer que la molécule de di-GMPc pourrait influencer sur le motif *Pseudomon*-GGDEF ainsi que sur l'expression du gène en aval. En effet, la diminution du niveau de luminescence pour la construction WT indique que l'expression du gène en aval est diminuée quand la concentration de di-GMPc est réduite. De plus, l'absence de différence entre les trois souches pour la construction défaite (disr) montre que si la structure de l'ARN est défaite la régulation de l'expression du gène est altérée. Les molécules de di-GMPc ont un rôle pléiotropique dans les bactéries notamment chez *Pseudomonas* (Valentini & Filloux, 2016; Wang *et al.*, 2019). Un contrôle a été effectué avec la construction WT du motif *rne-II* qui ne semble pas montrer de lien direct avec le di-GMPc afin de vérifier que cette observation soit bien dépendante du motif *Pseudomon*-GGDEF (Figure 3.19 B). Nous pouvons ainsi remarquer dans les deux milieux qu'il n'y a pas de différence dans le niveau de luminescence entre les trois souches testées (PA14 WT, PA14 *wspF*- et *wspR*-) produisant des niveaux de di-GMPc différents (Figure 3.19 B). Ce résultat permet de conforter les résultats obtenus pour le motif *Pseudomon*-GGDEF concernant son lien avec les molécules de di-GMPc, car les variations sont spécifiques à ce motif.

À la vue des résultats obtenus, nous avons émis l'hypothèse que *Pseudomon*-GGDEF pourrait être une nouvelle classe de *riboswitch* répondant au di-GMPc (Lee *et al.*, 2010; Sudarsan *et al.*, 2008). Pour vérifier notre hypothèse, des essais de cartographie *in vitro* par *in-line* ont été effectués avec cinq constructions de *Pseudomon*-GGDEF en présence de différentes concentrations de di-GMPc et de pGpG (Figure 3.20 et Annexe 6). Le motif *Pseudomon*-GGDEF étant situé en amont d'un gène possédant le domaine EAL permettant la dégradation du di-GMPc en pGpG, nous avons voulu également tester cette molécule (Römling *et al.*, 2005). Les séquences testées ont été répertoriées dans le Tableau 3.5.

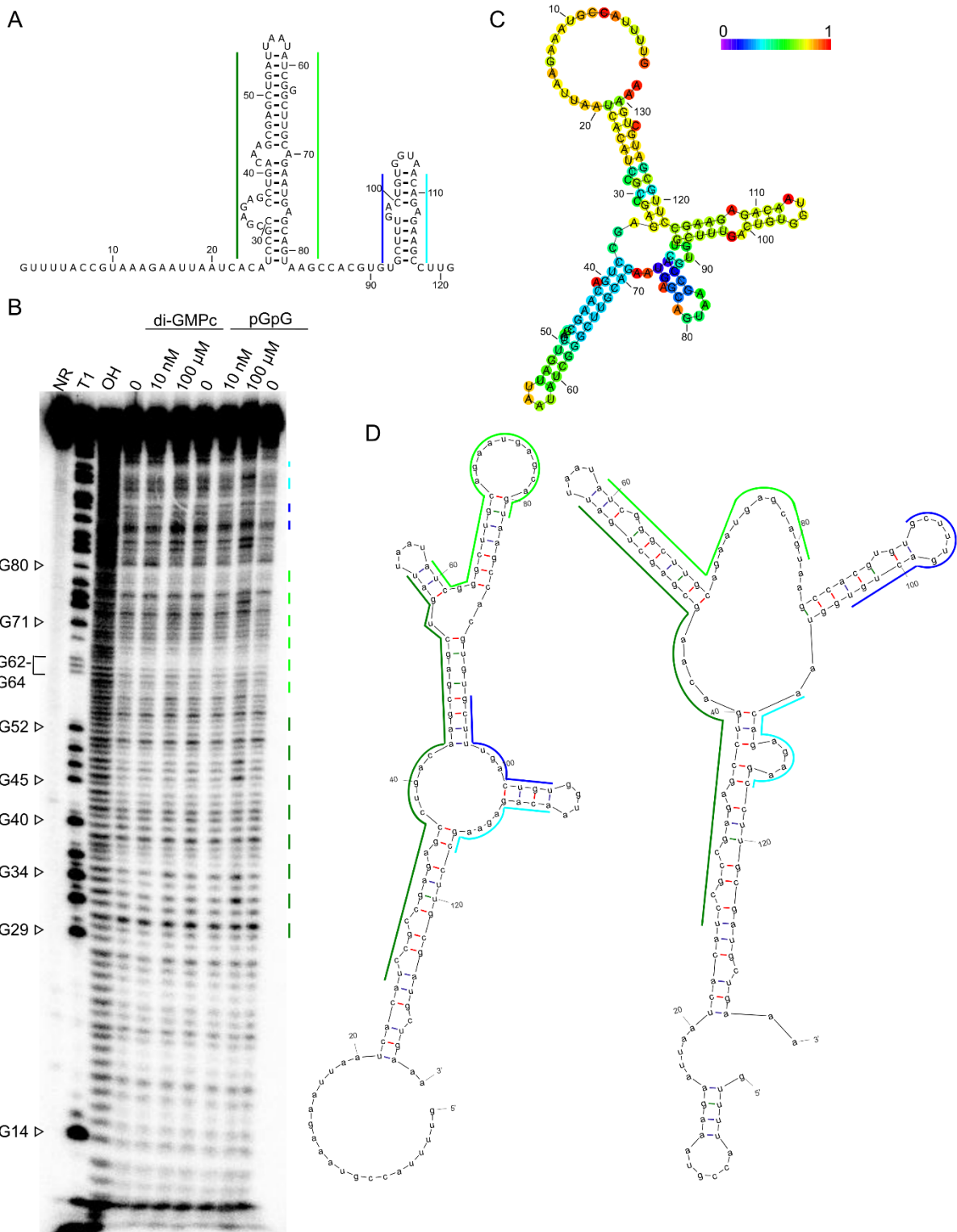


Figure 3.20 Plusieurs structures de *Pseudomon*-GGDEF semblent coexister.

A : Structure du motif *Pseudomon*-GGDEF basée sur l'alignement de plusieurs séquences réalisée avec R2R. **B** : Gel de cartographie *in vitro* par *in-line* pour la construction 1 de *Pseudomon*-GGDEF avec des concentrations de 10 nM à 100 μM de di-GMPc et de pGpG. NR, T1 et OH représentent les échelles d'ARN non réagi, d'ARN traité à l'ARNase

T1 (clive après les G) et d'ARN soumis à un traitement alcalin (clive après chaque nucléotides) respectivement. **C** : Structure de la construction 1 de *Pseudomon*-GGDEF obtenu avec RNAfold (Lorenz *et al.*, 2011). Les couleurs des ronds dans la structure représentent la probabilité que ce nucléotide soit impliqué dans une paire de base. **D** : Structures de la construction 1 de *Pseudomon*-GGDEF obtenu avec Mfold (Zuker, 2003). Les traits vert foncé et vert clair permettent de visualiser la région correspondant à la tige 1 de la structure présentée en A dans les différentes structures obtenues. Les traits bleu foncé et bleu clair permettent de visualiser la région correspondant à la tige 2 de la structure présentée en A dans les différentes structures obtenues.

Au niveau de la structure de *Pseudomon*-GGDEF, on peut observer que la structure n'est composée que de deux tiges boucles (Figure 3.20 A). Les résultats obtenus par cartographie *in vitro* par *in-line* indiquent également que cet ARN est peu structuré comme on peut le remarquer dans la Figure 3.20 B. En effet, peu de zones semblent faire partie de structures double brin stables qui peuvent être visualisées sur gel comme des zones sans dégradation (Figure 3.20 B). Des prédictions de structures ont été réalisées avec différents logiciels disponibles à savoir RNAfold (Figure 3.20 C) et Mfold (Figure 3.20 D). Ces structures obtenues ont été comparées avec la structure prédite suite à l'alignement des séquences disponibles. Les régions des tiges boucles du modèle dérivé de l'alignement ont été recherchées dans les structures alternatives et annotées (Figure 3.20 A, C et D). Nous pouvons ainsi remarquer que certaines structures sont différentes de celles présentes initialement, comme nous pouvons le voir notamment avec la deuxième partie de tige 1 (représentée en vert clair) où sur les structures alternatives de la Figure 3.20 D une grande partie semble être une boucle. Ainsi, plusieurs structures semblent coexister dans cette cartographie.

Concernant les ligands testés à savoir le di-GMPc et le pGpG. Nous pouvons tout d'abord remarquer qu'il ne semble pas y avoir de modulation pour les essais avec le di-GMPc (Figure 3.20 B). Cependant, de manière intéressante nous pouvons voir dans les conditions avec le pGpG qu'il semble y avoir une modulation à la concentration 100 μ M (Figure 3.20 B). Comme seules deux concentrations ont été testées dans la Figure 3.20 B, des essais ont également été réalisés avec une gamme de concentration plus étendue (Figure 3.21).

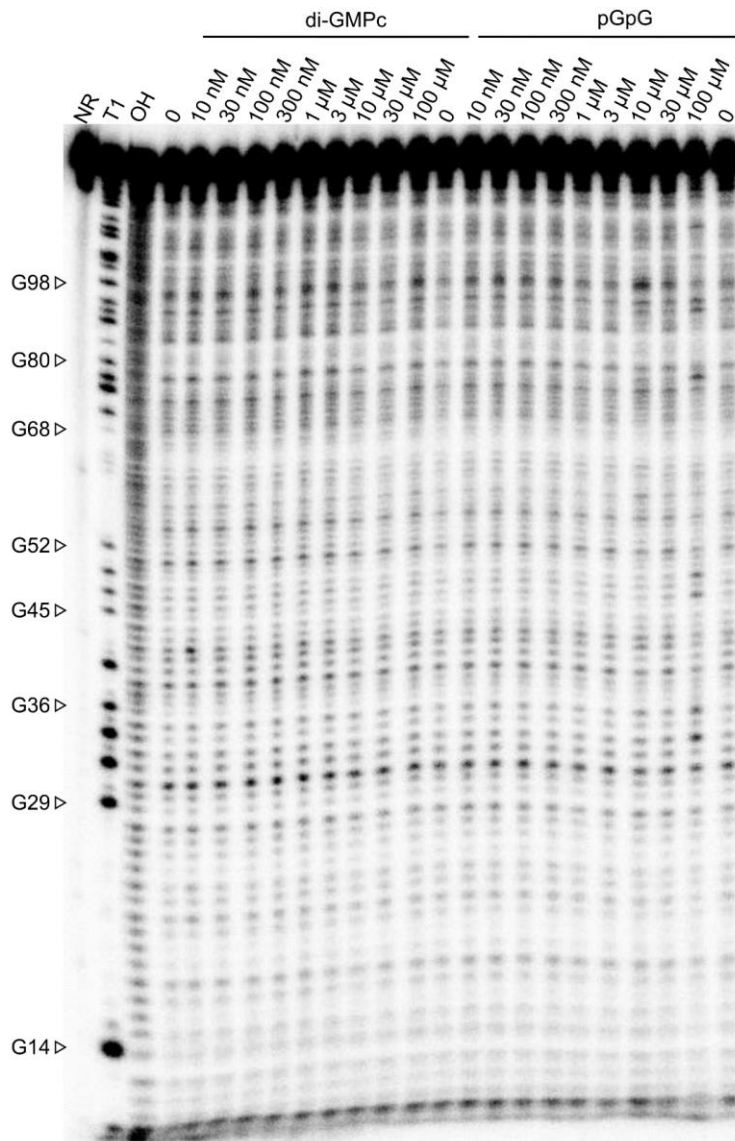


Figure 3.21 L'influence du pGpG sur l'ARN *Pseudomon-GGDEF* n'est visible qu'à de hautes concentrations.

Gel de cartographie *in vitro* par *in-line* pour la construction de *Pseudomon-GGDEF* 1 avec des concentrations de 10 nM à 100 µM de di-GMPc et pGpG, sa forme dégradée par la phosphodiesterase EAL. NR, T1 et OH représentent les échelles d'ARN non réagi, d'ARN traité à l'ARNase T1 (clive après les G) et d'ARN soumis à un traitement alcalin (clive après chaque nucléotides) respectivement.

Nous pouvons ainsi remarquer que la modulation induite par le pGpG ne se passe qu'à 100 µM, il s'agit d'une concentration assez élevée pour un *riboswitch* (Figure 3.21). De plus, il est possible à de hautes concentrations d'avoir des modulations non spécifiques à cause de contaminants présents dans la solution de la molécule testée comme il a pu être observé dans une étude

publiée en 2012 (Baker *et al.*, 2012). Ainsi, il semble peu probable que l'ARN *Pseudomon-GGDEF* soit un *riboswitch* liant le pGpG.

À la vue des résultats obtenus dans cette partie, il semble que *Pseudomon-GGDEF* soit un ARNnc en *cis* régulant l'expression du gène en aval ayant un lien avec le métabolisme du di-GMPc. Cependant, d'autres travaux seront nécessaires afin de trouver son mécanisme exact de fonctionnement.

3.5.3 Candidats avec un rôle de régulation confirmé

3.5.3.1 Motif *Pseudomon-Rho*

Le motif *Pseudomon-Rho* est retrouvé exclusivement dans les bactéries de la famille des Pseudomonadaceae (Weinberg *et al.*, 2010). Ce motif est présent devant le gène *rho* codant pour la protéine Rho impliquée dans un mécanisme de terminaison de la transcription chez les bactéries (Henkin & Peters, 2020). *Pseudomon-Rho* est présent dans la base de données Rfam sous la référence RF01720 (Kalvari *et al.*, 2021). Sa structure est présentée dans la Figure 3.22. Le motif est situé à 56 nts du codon d'initiation du gène *rho* laissant ainsi la place pour une plateforme d'expression. Cependant, aucune structure de terminateur de transcription Rho indépendant n'est visible.

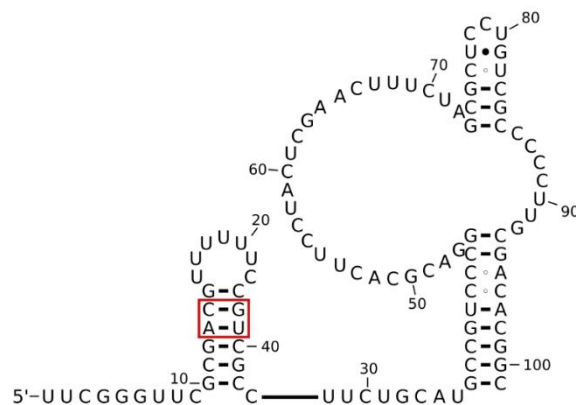


Figure 3.22 Structure du motif *Pseudomon-Rho* WT chez *P. aeruginosa* PA14.

Structure réalisée grâce à l'outil R2R en extrayant les séquences de *Pseudomon-Rho* contenues dans Rfam. Le rectangle rouge correspond à l'endroit de la mutation disruptive (A13U et C14G).

Compte tenu que ce motif se retrouve en amont du gène codant pour Rho, il est possible de se demander si cette protéine elle-même ne pourrait pas réguler son expression via une boucle de rétrocontrôle négatif entraînant un arrêt de transcription Rho dépendant. Afin de permettre un tel mécanisme, l'hexamère de Rho doit reconnaître des sites *rut* qui sont des séquences riches en cytosine (C) mais pauvres en guanine (G) dépourvues de structure secondaire contrairement aux terminateurs Rho indépendants (Ciampi, 2006; Mitra *et al.*, 2017). Or, l'analyse de la séquence de *Pseudomonas*-Rho montre que le motif contient 60 C pour 28 G pouvant concorder avec le ratio nécessaire pour la liaison de Rho sur des sites *rut*. Cependant, ce motif présente des structures tiges-boucles mais on peut penser que Rho puisse se lier au niveau de la boucle de la tige II qui est une partie riche en C. En effet, il a été montré qu'une séquence artificielle riche en C qui n'est pas un site *rut* peut permettre un recrutement de Rho (Hart & Roberts, 1991).

L'étude transcriptomique de Wurtzel et collaborateurs a permis de montrer que le transcrit du gène *rho* est un ARNm monocistronique (Figure 3.23) (Wurtzel *et al.*, 2012).

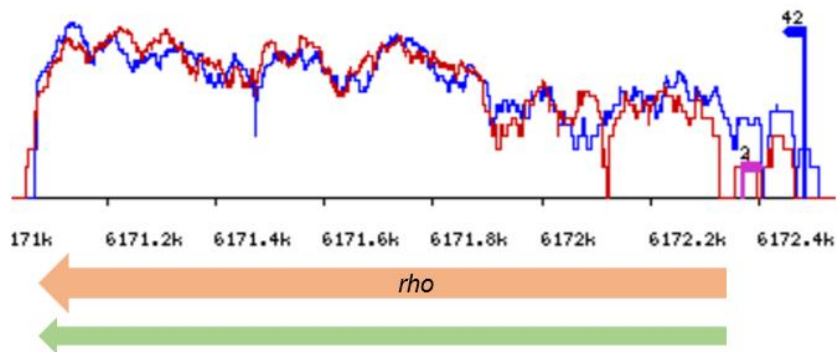


Figure 3.23 Contexte de la transcription du gène *rho* dans *Pseudomonas aeruginosa* PA14.

Représentation schématique des résultats de données de transcriptomique pour le gène *rho* (flèche orange), le transcrit est représenté par la flèche verte. La flèche bleue correspond au début de transcription du gène. Adapté de (Wurtzel *et al.*, 2012).

Les constructions WT et *disr* ont été clonées dans le plasmide pVK-f2.2-lux et conjuguées dans *P. aeruginosa* PA14. Des essais de luminescence ont été réalisés dans les conditions listées dans le Tableau 3.4. Les ratios luminescence / DO ont été calculés (Figure 3.24).

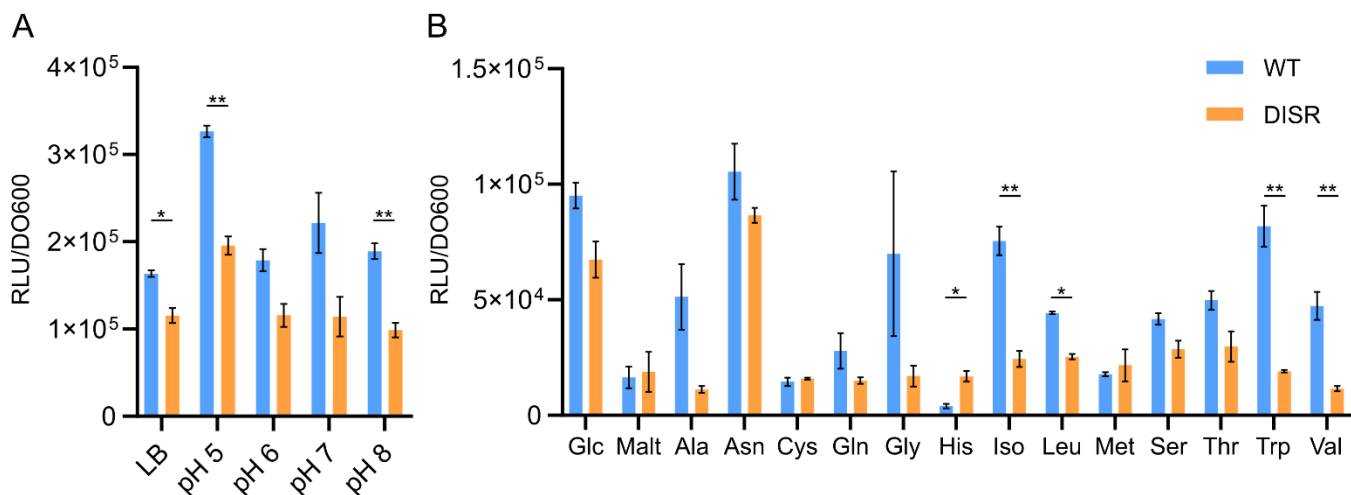


Figure 3.24 La construction *Pseudomon-Rho* WT montre un niveau de luminescence plus fort que la construction avec les mutations *disr* pour certains milieux.

Histogramme du ratio luminescence / DO 600nm pour les constructions du motif *Pseudomon-Rho* WT (bleu), disruptif (orange) cultivées pendant 5h à 37°C en agitation. **A** : Les milieux riches testés sont LB, pH 5 (LB pH 5), pH 6 (LB pH 6), pH 7 (LB pH 7), pH 8 (LB pH 8), **B** : Les milieux chimiquement définis testés sont : Glc (M9 glucose), Malt (M9 maltose), Ala (M9 alanine), Asn (M9 asparagine), Cys (M9 cystéine), Gln (M9 glutamine), Gly (M9 glycine), His (M9 histidine), Iso (M9 isoleucine), Leu (M9 leucine), Met (M9 méthionine), Ser (M9 sérine), Thr (M9 thréonine), Trp (M9 tryptophane) et Val (M9 valine). Un t-test multiple non pairé a été réalisé (* : $p < 0,05$; ** : $p < 0,01$).

Nous pouvons voir plusieurs profils de niveau de luminescence pour le motif *Pseudomon-Rho* (Figure 3.24). En effet, nous pouvons remarquer que pour certains milieux comme le M9 maltose ou encore le M9 supplémenté en cystéine ou méthionine, il n'y a pas de différence du niveau de luminescence entre la construction WT et la construction défectueuse comprenant les mutations *disr* (Figure 3.24 B). Les résultats des essais de luminescence des milieux qui ne sont pas représentés dans cette figure sont présents en Annexe 8. Nous pouvons également remarquer que le niveau de luminescence est plus fort pour la construction WT que la construction défectueuse pour la plupart des milieux montrant une différence. Des différences significatives sont notamment observables dans le milieu LB, notamment à pH 5 et pH 8 (Figure 3.24 A), ainsi que pour certaines conditions en milieu M9 comme ceux supplémentés en histidine, isoleucine, leucine, valine et tryptophane (Figure 3.24 B). La leucine, l'isoleucine et la valine sont des acides aminés appartenant à la même classe à savoir les acides aminés branchés (BCAA) pouvant ainsi expliquer que ces trois composés donnent une réponse semblable. De plus, ils sont tous les trois hydrophobes, tout comme le tryptophane. Enfin, nous pouvons également noter que le ratio WT / *disr* est inversé dans le cas du milieu M9 supplémenté en histidine (Figure 3.24 B). Il s'agit du seul milieu testé montrant ce profil ainsi il est difficile de tirer des conclusions à cette étape sur ce résultat.

À la vue des résultats obtenus, il semble que le motif *Pseudomon*-Rho puisse être un ARNnc régulant l'expression de Rho car la présence de mutations *disr* entraîne des différences d'expression du rapporteur. Cependant, le signal permettant sa régulation reste encore à déterminer car plusieurs conditions montrent des différences d'expression entre les deux constructions. De plus, ne sachant pas l'effet de la présence des mutations *disr*, qui pourraient bloquer l'ARNnc dans sa forme non liée, il est également possible que les conditions ne montrant pas de différence soient également des conditions d'intérêt. De plus, à la vue de la composition en C et G de sa structure, il est également possible de se demander si Rho ne pourrait pas interagir avec cet ARNnc dans certaines conditions afin de réguler sa propre expression.

3.5.3.2 Motif *sucA-II*

Le motif *sucA-II* se retrouve en amont du gène *sucA* exclusivement dans les bactéries de la famille des Pseudomonadaceae (Greenlee *et al.*, 2018; Weinberg *et al.*, 2010). Le gène *sucA* code pour le composant E1 du complexe 2-oxoglutarate déshydrogénase, constituant la première étape de la conversion du 2-oxoglutarate en succinyl-CoA et CO₂ dans le cycle du TCA. De manière intéressante, il existe également un motif devant le gène *sucA* dans les bactéries du genre *Burkholderia* (Weinberg *et al.*, 2007). Ces deux candidats n'ont pas la même structure mais il serait possible que l'éventuel ligand / signal soit identique et qu'il s'agisse de deux classes d'un même *riboswitch* ou ARNnc régulant en *cis* (Greenlee *et al.*, 2018).

Le motif *sucA-II* est présent dans la base de données Rfam sous la référence RF01758 (Kalvari *et al.*, 2021). Sa structure est plus complexe que d'autres motifs présentés comme *rmf* ou *Pseudomon*-GGDEF (Figure 3.25) et elle ne comporte pas de structure pouvant s'apparenter à un terminateur Rho indépendant.

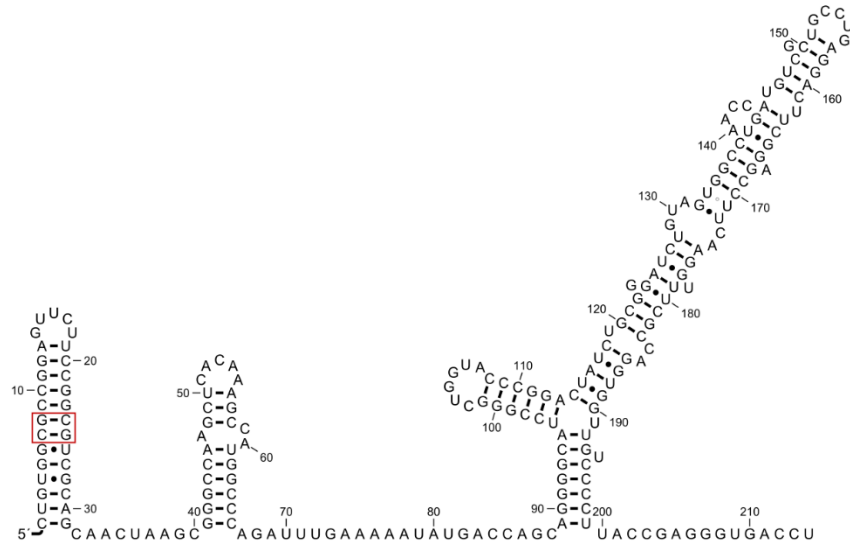


Figure 3.25 Structure du motif *sucA-II* WT de *Pseudomonas aeruginosa* PA14.

Structure réalisée grâce à l'outil R2R en extrayant les séquences du motif *sucA-II* contenues dans Rfam. Le rectangle rouge correspond à l'endroit des mutations disruptives (C7G et G8C).

Une étude transcriptomique sur *P. aeruginosa* PA14 a permis de réaliser une carte de l'expression des différents gènes dans cet organisme à 28°C et 37°C (Wurtzel *et al.*, 2012). Le gène *sucA* est transcrit avec l'opéron qui le précède à savoir les gènes *sdhC*, *sdhD*, *sdhA* et *sdhB* (Figure 3.26). Cet opéron code pour les différentes sous-unités de la succinate déshydrogénase impliquée également dans le cycle du TCA, mais à une étape plus lointaine, à savoir la conversion du succinate en fumarate. Au vu du profil de transcriptomique (Figure 3.25), nous pouvons émettre l'hypothèse que le motif *sucA*, s'il s'agit d'un ARN non codant, agisse au niveau de la traduction plutôt que sur la transcription du gène *sucA* dans les conditions testées (LB en début de phase stationnaire de croissance). En effet, on peut observer qu'il n'y a pas de différence dans le profil de transcription au niveau du motif en amont de *sucA* (Wurtzel *et al.*, 2012). Enfin, la présence du motif *sucA-II* entre l'opéron *sdhCDAB* et le gène *sucA* pourrait indiquer qu'il existe une condition pour laquelle l'expression de *sucA* doit être différente de l'expression de l'opéron présent en amont de ce gène. Une étude chez *P. aeruginosa* a montré qu'en condition de faible concentration de fer l'expression de la succinate déshydrogénase (*sdhCDAB*) est diminuée, mais, de manière intéressante, l'expression de *sucA* ne l'est pas (Nelson *et al.*, 2019). Ceci semble pointer dans la même direction que notre hypothèse à savoir que le motif doit permettre une expression différentielle entre *sdhCDAB* et *sucA*.

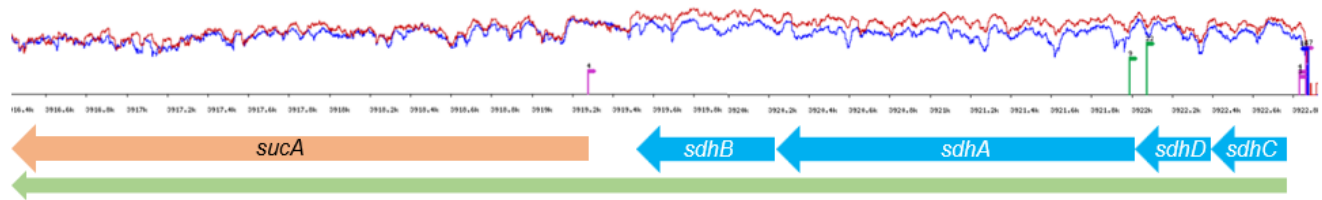


Figure 3.26 Profil de transcriptomique de l'opéron *sdhCDAB* et du gène *sucA*.

Représentation schématique des résultats de données de transcriptomique pour l'opéron *sdhCDAB* (flèches bleu clair) et le gène *sucA* (flèche orange). La flèche en bleu en amont du gène *sdhC* correspond au promoteur de ces 5 gènes. La flèche verte en-dessous correspond au transcrit poly-cistronique. Les courbes en bleu et en rouge correspondent à la couverture des lectures de RNAseq à 28°C et à 37°C, respectivement. Adapté de (Wurtzel *et al.*, 2012).

Les constructions du motif *sucA-II* WT et disruptif ont été clonées dans le plasmide pVK-f2.2-lux, puis ont été conjuguées dans la souche de *P. aeruginosa* PA14. Des essais de luminescence ont été effectués dans toutes les conditions listées dans le Tableau 3.4 (Figure 3.27 A et Annexe 9).

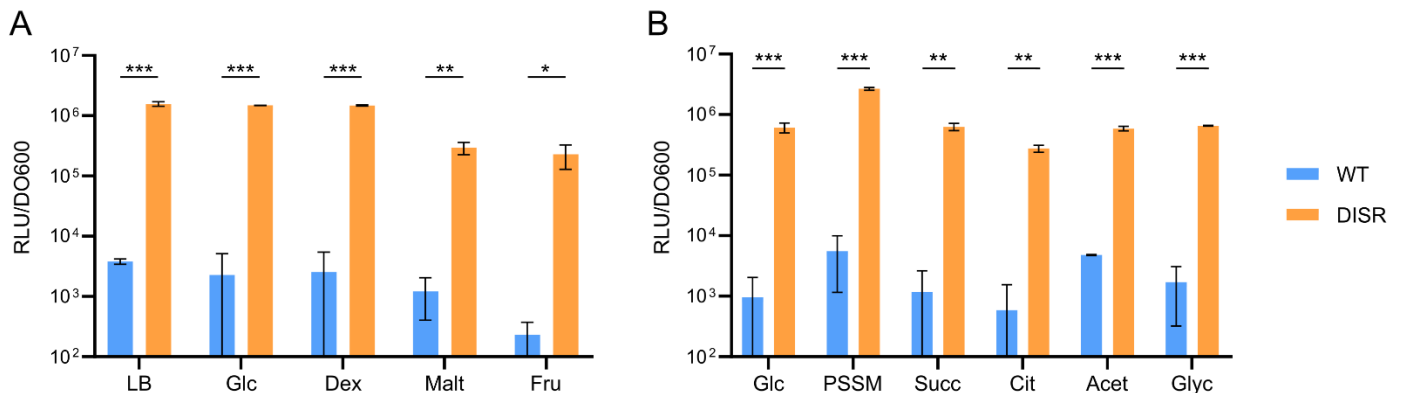


Figure 3.27 Le profil du niveau de luminescence pour le motif *sucA-II* est semblable quel que soit le milieu testé.

Histogramme du ratio luminescence / DO 600 nm pour les constructions du motif *sucA* WT (bleu), disruptive (orange) cultivées pendant 5h à 24h dépendamment du milieu. Les mesures ont ensuite été prises grâce à la lectrice à plaque M1000 infinie (TECAN®). Pour plus de clarté dans les résultats, les valeurs ont été représentés sur une échelle logarithmique base 10. Les milieux testés sont en **A** : LB, Glc (M9 glucose), Dex (M9 dextrose), Mal (M9 maltose), Fru (M9 fructose). **B** : Glc (M9 glucose), PSSM (peptone succinate salt medium), Succ (M9 succinate), Cit (M9 citrate), Acet (M9 acétate) et Glyc (M9 glycérol). Un t-test multiple non païré a été réalisé (* : $p < 0,05$; ** : $p < 0,01$; *** : $p < 0,001$).

Pour ce motif, nous pouvons voir que quel que soit le milieu testé, il y a une grande différence dans le niveau de luminescence entre les deux constructions (Figure 3.27 A). Les valeurs du ratio luminescence / DO 600nm ont été représentées sur une échelle logarithmique afin de pouvoir

visualiser convenablement les différences entre les milieux / conditions. Pour toutes les conditions testées, on peut remarquer que l'expression de la luminescence de la construction WT est beaucoup plus faible que celle de la construction avec les mutations disruptives (Figure 3.27 A). Des analyses statistiques ont été réalisées et ont permis de montrer que pour la quasi-totalité des milieux la différence d'expression entre la construction WT et la construction contenant les mutations disruptives sont statistiquement significatives (Tableau 3.6). Tous les milieux testés dans la Figure 3.27 A ont comme source de carbone du glucose ou un sucre utilisant les mêmes voies de synthèse et de dégradation (Kanehisa *et al.*, 2020; Kanehisa & Goto, 2000). Le motif *sucA-II* se trouvant en amont d'un gène impliqué dans le cycle du TCA, nous avons aussi testé d'autres sources de carbone favorisant ce cycle par rapport à la glycolyse. Des essais de luminescence ont été réalisés avec cinq différents milieux contenant comme source de carbone du succinate (M9 et peptone-succinate salt medium), du citrate, de l'acétate ou du glycérol (Figure 3.27 B). Une nouvelle fois, nous pouvons voir une grande différence de niveau de luminescence entre la construction WT et la construction défective contenant les mutations disruptives (Figure 3.27 B). Cette différence entre les deux constructions est également significative pour tous les milieux suite aux analyses statistiques (Figure 3.27 B).

Les résultats obtenus semblent indiquer que le motif *sucA-II* semble être un régulateur de l'expression du gène *sucA*. En effet, nous avons pu voir pour la construction contenant les mutations disr que l'expression de la luminescence est fortement modifiée par rapport à la construction WT. Ceci suggère que la structure de l'ARN est importante pour la régulation de l'expression du gène car quand celle-ci est défective l'expression du rapporteur est changée. Lors de ces essais, nous cherchons à identifier des conditions montrant des différences par rapport aux autres conditions testées afin de trouver le signal permettant la régulation par le motif. Cependant, ici nous ne sommes pas parvenus à trouver de telles conditions, cela pourrait signifier que le signal est présent ou absent dans toutes nos conditions. En effet, la présence des mutations disr peuvent bloquer le motif dans sa conformation active ou inactive. Plus de conditions seront à tester pour ce motif afin de déterminer son mécanisme de régulation.

3.5.4 Candidats dont le rôle de régulation reste à confirmer

3.5.4.1 Motif fixA

Le motif *fixA* est présent exclusivement dans les bactéries du genre *Pseudomonas*, plus spécifiquement dans les espèces *Pseudomonas entomophila*, *Pseudomonas fluorescens*, *Pseudomonas putida* et *Pseudomonas syringae* (Weinberg *et al.*, 2010). Ces bactéries sont soit

des bactéries pathogènes des humains, des animaux ou des plantes. Le motif est présent devant le gène *fixA* aussi nommé *effB-2*, codant pour la sous-unité β d'une flavoprotéine. Ce gène est présent dans un opéron avec le gène *effA-2* (ou *fixB*) codant pour la sous-unité α de cette flavoprotéine (Keseler *et al.*, 2021; Tsai & Saier Jr, 1995; Weinberg *et al.*, 2010). Les protéines FixA et FixB sont impliquées dans la chaîne respiratoire de la bactérie. Contrairement aux autres motifs de cette thèse, ce motif n'est pas répertorié dans la base de données Rfam. En effet, l'étude classe le motif *fixA* comme étant ambigu pour être un ARN, basé sur l'évidence de covariation dans le modèle ainsi que des variations de taille du motif dans les différentes bactéries où il est retrouvé (Weinberg *et al.*, 2007; Weinberg *et al.*, 2010). Cependant, sa présence devant des gènes ayant une fonction similaire lui confère une probabilité d'être un élément *cis*-régulateur (Weinberg *et al.*, 2007; Weinberg *et al.*, 2010). La structure 2D du motif a été réalisée à partir de l'alignement et la séquence mentionné dans l'article (Weinberg *et al.*, 2010). Sa structure est très simple, il est constitué d'une tige-boucle contenant deux boucles internes (Figure 3.28). Ces boucles pourraient être celles responsables de la liaison d'un ligand dans le cas d'un *riboswitch* hypothétique. Ce motif est situé à 56 nts du codon initiateur du gène *fixA* et ne présente pas de structure évidente de terminateur de transcription Rho-indépendant.

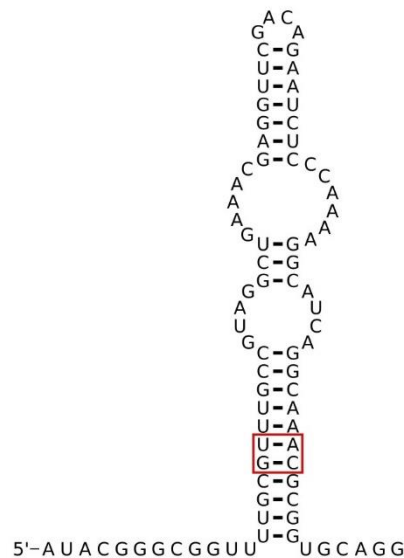


Figure 3.28 Structure du motif *fixA* WT de *Pseudomonas syringae* pv. *tomato* DC3000.

Structure réalisée grâce à l'outil R2R à partir de la séquence présente dans l'article (Weinberg *et al.*, 2010) et le consensus de la structure. Le rectangle rouge correspond à l'endroit des mutations disruptives (G17C et U18A).

Les différentes constructions ont été clonées dans le plasmide pVK-f2.2-lux. Comme pour le motif *Pseudomon-GGDEF*, les essais auraient idéalement dû être faits dans le même organisme que le motif, mais la conjugaison n'a pas été possible au sein de *P. syringae* pv. *tomato* DC3000 malgré les adaptations du protocole pour cette bactérie. Des essais de luminescence ont été réalisés dans toutes les conditions listées dans le Tableau 3.4 et les ratios luminescence / DO 600 nm ont été calculés (Figure 3.29).

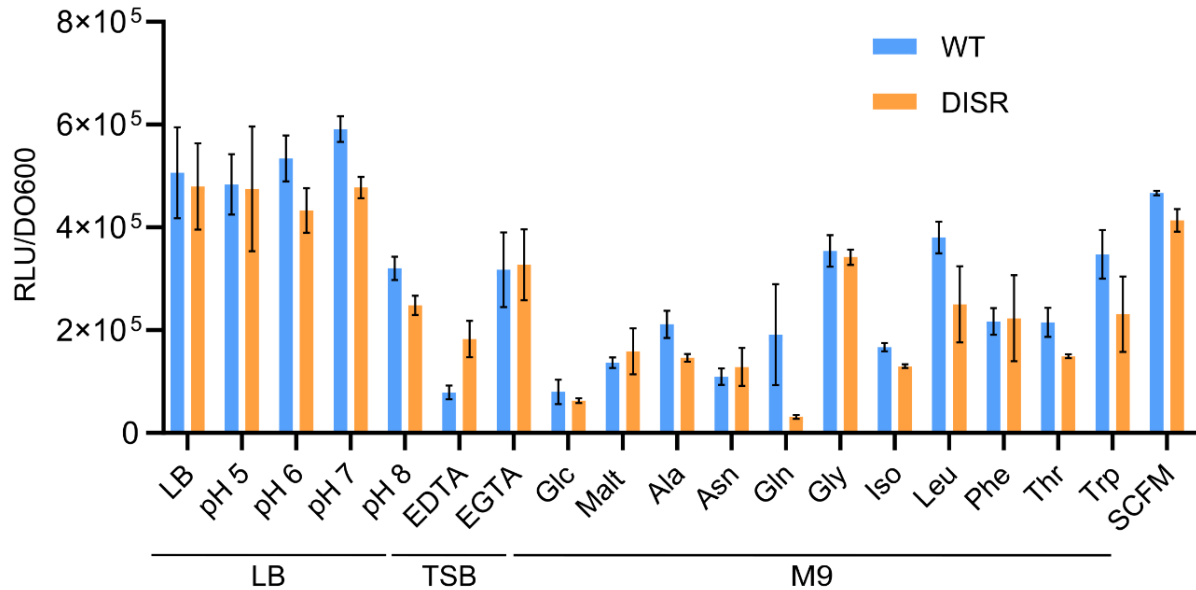


Figure 3.29 Des différences dans le niveau de luminescence ne sont distinguables que pour certains milieux pour le motif *fixA*.

Histogramme du ratio luminescence / DO 600nm pour les constructions du motif *fixA* WT (bleu), disruptive (orange) cultivées pendant 5h à 37°C en agitation. Les milieux testés sont : LB, pH 5 (LB pH 5), pH 6 (LB pH 6), pH 7 (LB pH 7), pH 8 (LB pH 8), EDTA (TSB + 1 mM EDTA), EGTA (TSB + 1 mM EGTA), Glc (M9 glucose), Malt (M9 maltose), Ala (M9 alanine), Asn (M9 asparagine), Gln (M9 glutamine), Gly (M9 glycine), Iso (M9 isoleucine), Leu (M9 leucine), Phe (M9 phénylalanine), Thr (M9 thréonine), Trp (M9 tryptophane) et SCFM.

Nous pouvons remarquer que pour la majorité des milieux testés différents profils de niveau de luminescence entre les constructions WT et *disr* sont visibles (Figure 3.29). Les résultats des essais de luminescence pour les milieux testés, mais non représentés dans la Figure 3.29, sont présentés en Annexe 10. Suite aux analyses statistiques, nous avons remarqué qu'aucun milieu testé ne montre de différence significative du niveau de luminescence entre les constructions WT et *disr* contenant les mutations *disr*. Nous pouvons cependant observer des différences apparentes pour plusieurs milieux comme le TSB supplémenté avec 1 mM EDTA (Figure 3.29). L'EDTA étant un agent chélateur d'ions divalents, nous pouvons émettre l'hypothèse que l'absence d'un des ions divalents du milieu entraîne une diminution de l'expression du gène *fixA*

grâce au motif (Finnegan & Percival, 2015). Une expression identique entre le WT et *disr* avec la condition TSB supplémenté avec 1 mM EGTA nous permet de dire que le calcium n'est pas impliqué pour la régulation pour ce motif, car l'EGTA est un agent chélateur d'ions divalents présentant une plus grande affinité pour les ions calcium (Stockbridge, 1987). Des différences apparentes sont également observables pour d'autres milieux comme le milieu M9 supplémenté en alanine, isoleucine, leucine et valine avec une expression de la luminescence plus forte pour la construction WT que la construction *disr* (Figure 3.29). Néanmoins, nous pouvons aussi remarquer des différences inverses, c'est-à-dire la construction *disr* montrant un niveau de luminescence plus élevé que la construction WT, pour les milieux M9 supplémentés en sérine ou encore le SCFM (Figure 3.29).

Le gène *fixA* étant impliqué dans la chaîne respiratoire des essais ont été réalisés en condition anaérobie (Tsai & Saier Jr, 1995). Ces essais ont été faits en boîte de Pétri sur un milieu TSB et TSB supplémenté avec du KNO_3 permettant l'apport d'une source de nitrate pour permettre la respiration de *P. aeruginosa* en anaérobie (Wu *et al.*, 2005). Des conditions aérobie et anaérobie ont été réalisés en parallèle afin de comparer le niveau de luminescence (Figure 3.30).

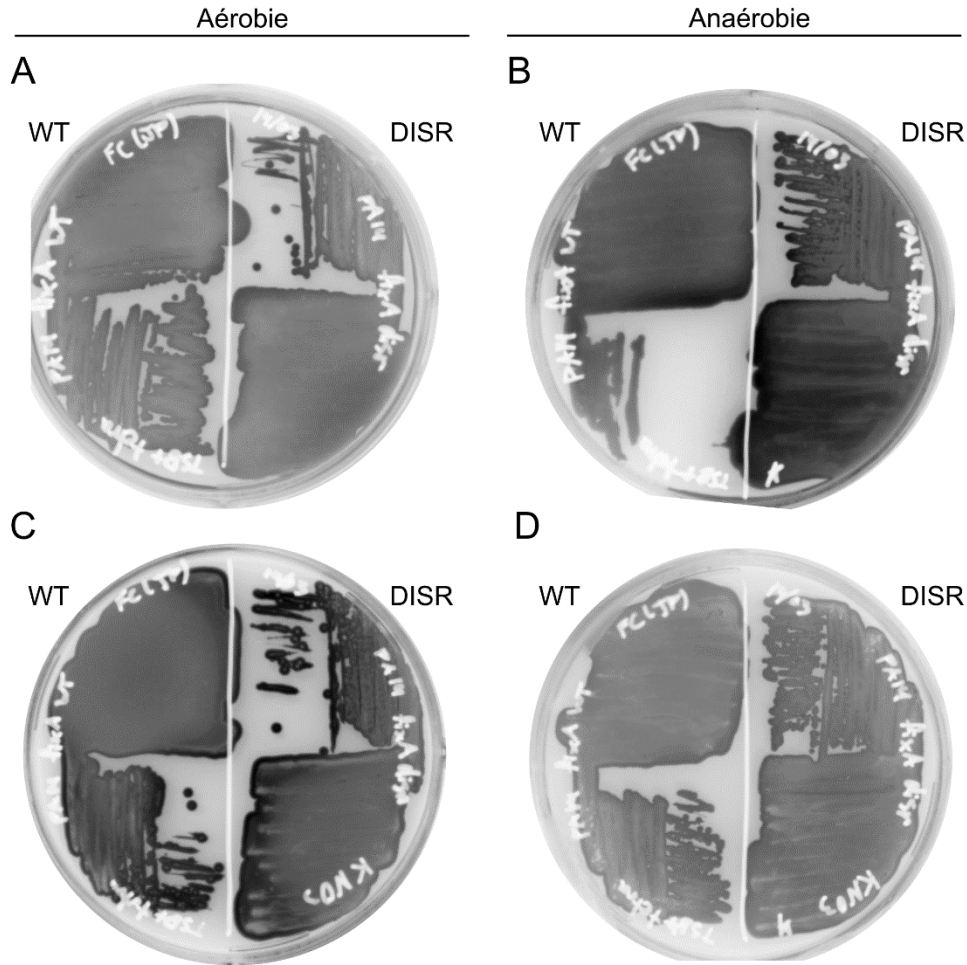


Figure 3.30 Aucune différence dans le niveau de luminescence n'est observable entre les deux constructions du motif *fixA* en condition anaérobie.

Culture sur boîte de Pétri des constructions WT et *disr* du motif *fixA* en milieu TSB agar (A et B) ou en milieu TSB + 1% KNO₃ agar (C et D) en condition aérobie (A et C) ou anaérobie (B et D). La luminescence est visualisée grâce au ImageQuant™ LAS 500 (GE Healthcare).

Nous pouvons observer que quel que soit la condition testée en anaérobie ou anaérobie le niveau de luminescence est semblable entre les constructions WT et *disr* (Figure 3.30). Pour le motif *fixA*, aucune condition ne montre de différence significative du niveau de luminescence entre les deux constructions testées, il est ainsi difficile de confirmer son rôle de régulation. D'autres travaux seront nécessaires afin de confirmer ce rôle et trouver son mécanisme de fonctionnement.

3.5.4.2 Motif *rne-II*

Notre dernier ARNnc candidat est le motif *rne-II*. Il est présent devant le gène *rne* exclusivement dans les bactéries de la famille des Pseudomonadaceae (Weinberg *et al.*, 2010). Le gène *rne* code pour l'enzyme ribonucléase E, impliquée dans le processus de dégradation de l'ARN comme partie intégrante du dégradosome (Carpousis, 2007).

Le motif *rne-II* est présent dans la base de données Rfam sous la référence RF01756. La structure du motif *rne-II* est représentée dans la figure 3.31.

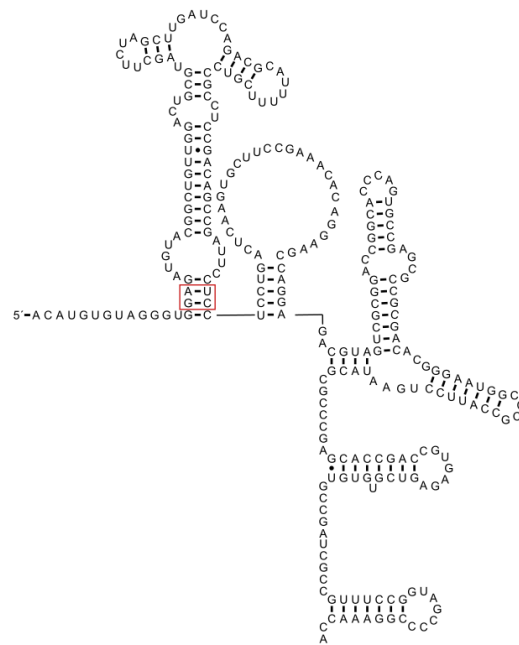


Figure 3.31 Structure du motif *rne-II* WT de *Pseudomonas aeruginosa* PA14.

Structure réalisée grâce à l'outil R2R en extrayant les séquences du motif *rne-II* contenues dans Rfam. Le rectangle rouge correspond à l'endroit des mutations disruptives (G15C et A16U).

Grâce à l'étude de Wurtzel et collaborateurs publiée en 2012, nous pouvons voir que le début de la transcription du gène *rne* coïncide avec la présence d'un ARN non codant potentiel annoté ici : PA14sr_060 (Figure 3.32) (Wurtzel *et al.*, 2012). En effet, bien qu'il ne soit pas visible le site de début de transcription est situé à la même position que le site de début de transcription de l'ARNnc potentiel à savoir à la position 2236725. En étudiant les positions de cet ARNnc potentiel et notre motif *rne-II*, il est possible de voir que ces deux ARN sont présents en amont du gène *rne* et se suivent directement (Figure 3.32) (Weinberg *et al.*, 2010; Wurtzel *et al.*, 2012). En effet, les

positions du dernier nucléotide de PA14sr_060 semble être le début de *rne-II*. Le fait que ce nucléotide soit partagé entre les deux ARN indique qu'il pourrait potentiellement s'agir d'un site de clivage par une ARNase, probablement l'ARNase E en l'occurrence. En effet, à la vue du rôle central et important de l'ARNase E dans le métabolisme bactérien, il est possible de se demander si ce motif ne pourrait pas être le siège d'une boucle de rétrocontrôle négatif où l'ARNase E viendrait se lier dans la partie 5' UTR de son ARNm et le cliver.

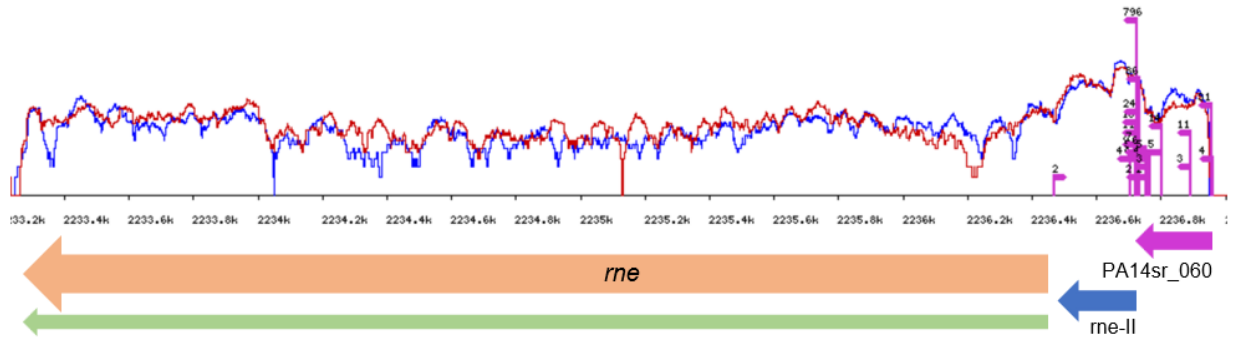


Figure 3.32 Contexte de la transcription du gène *rne-II* chez *P. aeruginosa* PA14.

Représentation schématique des résultats de données de transcriptomique pour le gène *rne* (flèche orange), l'ARN non codant potentiel PA14sr_060 (flèche rose) et notre ARN candidat *rne-II* (flèche bleue). La flèche en vert correspond au à la séquence codante du gène *rne*. Les courbes en bleu et en rouge correspondent à la couverture des lectures de RNA-seq à 28°C et à 37°C, respectivement. Adapté de (Wurtzel *et al.*, 2012).

Les constructions *rne-II* WT et *disr* ont été clonées dans le plasmide pVK-f2.2-lux, puis les plasmides ont été conjugués dans la souche de *P. aeruginosa* PA14. Des essais de luminescence ont été effectués dans toutes les conditions listées dans le Tableau 3.4 afin de comparer l'expression entre les deux constructions (Figure 3.33).

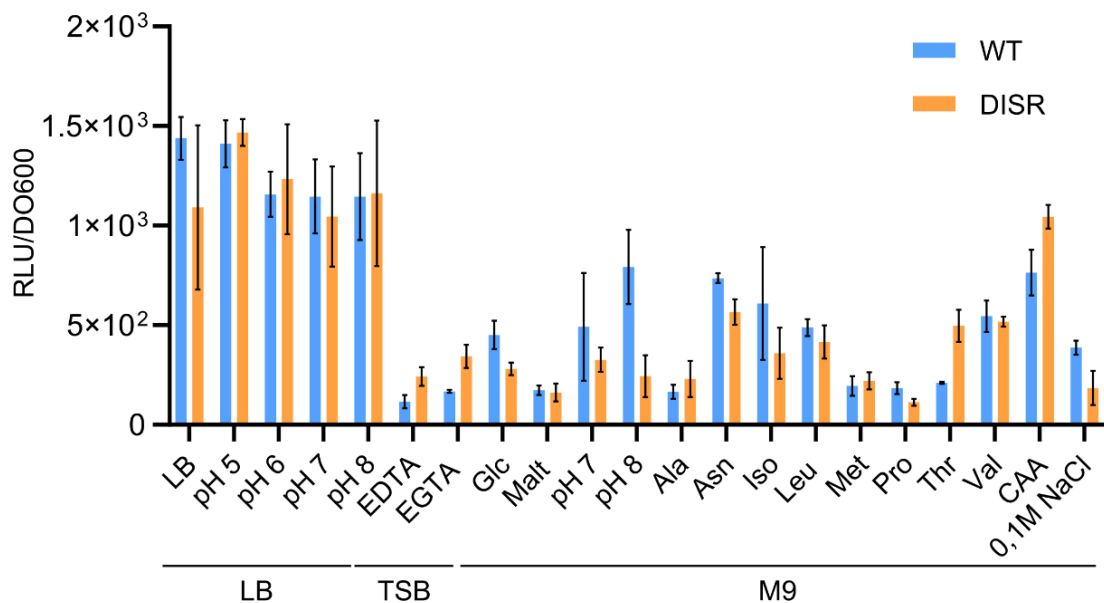


Figure 3.33 Différents profils de niveau de luminescence sont visibles selon le milieu testé pour le motif *rne-II*.

Histogramme du ratio luminescence / DO 600nm pour les constructions du motif *rne-II* WT (bleu) et disruptif (orange) cultivé pendant 5h à 24h dépendamment du milieu. Milieux riches testés : LB, pH 5 (LB pH 5), pH 6 (LB pH 6), pH 7 (LB pH 7), pH 8 (LB pH 8), EDTA (TSB + 1 mM EDTA), EGTA (TSB + 1 mM EGTA), Milieux chimiquement définis testés : Glc (M9 glucose), Malt (M9 maltose), Ala (M9 alanine), Asn (M9 asparagine), Iso (M9 isoleucine), Leu (M9 leucine), Met (M9 méthionine), Pro (M9 proline), Thr (M9 thréonine), Val (M9 valine), CAA (M9 CAA) et 0,1 M NaCl (M9 + 0,1M NaCl). Les mesures ont ensuite été prises grâce à la lectrice à plaque M1000 infinie (TECAN®).

Pour ce motif, on peut voir plusieurs profils de niveau de luminescence en fonction du milieu testé (Figure 3.33). Les résultats des ratios luminescence / DO 600 nm pour les milieux qui ne sont pas représentés dans ce graphique sont présents en Annexe 11. En effet, on peut observer que pour la plupart des milieux, incluant toutes les conditions avec LB, on ne voit pas de différence entre la construction WT et la construction avec les mutations *disr* (Figure 3.33). Ceci peut nous indiquer deux choses, premièrement il est possible que dans ces milieux le métabolite permettant la régulation par ce motif soit absent ne permettant ainsi pas son action. Ou alors que celui-ci est présent, mais que la construction avec les mutations *disr* est bloquée dans une structure ne permettant pas de voir de différences entre les deux conditions.

Pour d'autres milieux, on peut remarquer des différences dans le niveau de luminescence entre les deux constructions comme pour les milieux M9 glucose, M9 supplémenté en asparagine ou en thréonine (Figure 3.33). Il est à noter également que dans plusieurs milieux, les profils de niveau de luminescence entre le WT et le *disr* sont inversés. En effet, on peut remarquer que dans le milieu M9 supplémenté en thréonine la construction disruptive montre un niveau de

luminescence fort comparé à la construction WT, alors que dans le milieu M9 supplémenté en asparagine on peut observer le phénomène inverse (Figure 3.33). Ce résultat était inattendu car en général la régulation via un ARNnc se fait dans le même sens, à savoir une activation ou une inhibition en présence ou absence de son métabolite. Cependant, aucun des milieux testés ne montre de différence significative, entre la version mutante et WT du motif *rne-II*, suite aux analyses statistiques.

Du fait de cette disparité d'expression dans les différentes conditions testées, il est difficile de statuer sur la condition qui est importante pour la régulation par ce motif. Néanmoins, à la vue de son contexte génétique, il est également possible que cet ARN ne lie pas un métabolite et soit plutôt le siège d'une boucle de rétroaction négative via une dégradation par l'ARNase E par exemple.

3.5.5 Conclusion

En conclusion de ce chapitre, le système rapporteur pVK-f2.2-lux nous a permis d'étudier des candidats ARNnc en *cis* découverts depuis plus de 10 ans et soupçonnés de réguler l'expression des gènes situés en aval, mais qui n'avaient pas encore été caractérisés et pour lesquels nous n'avions donc aucune information confirmant un rôle dans la régulation génétique. Le rapporteur luciférase en fusion traductionnelle permet d'étudier *in vivo* des variations d'expression médiées par le motif d'ARN dans différentes conditions dans le but d'explorer le ligand d'un potentiel *riboswitch*, ou la nature de l'ARN agissant en *cis* (T-box, r-leader ...). Lors de cette étude, nous avons pu mettre en évidence l'importance des structures des ARNnc candidats pour la régulation des gènes situés en aval de ceux-ci par l'utilisation de mutations disruptives supposément non-fonctionnelles. Les résultats des essais de luminescence obtenus dans ce chapitre 1 sont résumés dans le Tableau 3.6.

Tableau 3.6 Résumé de la différence du niveau de luminescence entre les constructions WT et DISR pour les six ARNnc candidats de l'étude et toutes les conditions testées.

		Candidats avec activité régulatrice confirmée								Candidats avec activité rég. non confirmée		
		<i>rmf</i>		<i>Pseudomon-GGDEF</i>	<i>Pseudomon-Rho</i>		<i>sucA-II</i>		<i>fixA</i>	<i>rne-II</i>		
		WT vs DISR	WT vs COMP	WT vs DISR	WT vs DISR	WT vs DISR	WT vs DISR	WT vs DISR	WT vs DISR			
LB	LB	>	**	=	=	>	*	<	***	=	=	
	pH 5	>	***	=	<	*	>	**	<	***	=	=
	pH 6	>	***	=	=	=	>	<	**	=	=	
	pH 7	>	***	=	=	=	>	<	**	>	<	
	pH 8	>	***	=	=	=	>	**	<	***	>	<
TSB	EDTA	>	***	=	=	=	>	<	**	<	>	
	EGTA	>	***	=	=	=	=	<	**	=	=	
M9	Glc	>	***	=	=	=	>	<	***	=	=	
	Dex	>	***	=	=	=	>	<	***	=	=	
	Fru	>	**	=	=	=	=	<	*	=	=	
	Malt	>	***	=	=	=	=	<	**	=	=	
	pH 7	>	***	=	=	=	>	<	***	=	=	
	pH 8	>	***	=	=	=	>	<	*	=	>	
	Ala	>	***	=	=	=	>	<	**	>	=	
	Asn	>	***	=	=	=	>	<	***	=	>	
	Cys	>	***	=	=	=	=	<	**	=	=	
	Gln	>	***	=	>	=	>	<	*	>	=	
	Gly	>	**	=	=	=	=	<	**	=	=	
	His	>	***	=	=	=	<	*	<	***	=	=
	Iso	>	***	=	<	=	>	**	<	**	>	=
	Leu	>	***	=	<	=	>	***	<	***	>	=
	Lys	>	***	=	=	=	=	=	<	**	=	=
	Met	>	***	=	=	=	=	=	<	***	=	=
	Phe	>	***	=	<	=	=	=	<	**	=	=
	Pro	>	***	=	=	=	=	=	<	=	=	>
	Ser	>	***	=	<	*	=	>	<	**	>	=
Thr	>	***	=	<	**	=	>	<	**	>	<	
Trp	>	**	=	=	=	>	**	<	*	=	=	

	Val	>	***	=	=	>	*	<	**	>	=
	CAA	>	***	=	=	=		<	***	=	<
	0,1 M NaCl	>	***	=	<	**	=	<	*	=	>
	Low P	>	***	=	=	=	=	<	***	=	=
SCFM		>	***	=	<	**	>	<	**	>	=
$\Delta relA$ $\Delta spoT$	LB	=	=								
	M9	=	=								
<i>wspF</i> -	TSB				=						
	M9				=						
<i>wspR</i> -	TSB				<	***					
	M9				<	***					
M9	Succ							<	**		
	Cit							<	**		
	Acet							<	***		
	Glyc							<	***		
PSSM								<	***		
TSB agar	Aéro.										=
	Anaér.										=
TSB + 1% KNO ₃ agar	Aéro.										=
	Anaér.										=

Tableau récapitulatif de tous les résultats obtenus lors du chapitre 1 de cette thèse de doctorat. Les conditions représentées sont tous les milieux testés (LB, TSB et M9) ainsi que les essais dans différentes souches ($\Delta relA$ $\Delta spoT$ ainsi que *wspF*- et *wspR*-). Aéro signifie aérobie et Anaér signifie anaérobie. Les différences d'expression de luminescence sont représentées par des symboles « < », « > » ou « = », la référence étant la construction WT. La couleur utilisée fait référence aux précédents graphiques où la construction WT est représentée en bleue et la construction *disr* est représentée en orange. Les résultats des tests statistiques ont été représentés en fonction de la valeur de la pvalue (* pvalue < 0,05, ** pvalue < 0,01, *** pvalue < 0,001). En l'absence de symboles, cela signifie qu'il y a une différence apparente dans les graphiques mais que celle-ci n'est pas significative.

Six candidats ARNnc en *cis* ont été testés, pour la majorité d'entre eux nous avons pu établir que ces structures d'ARN sont importantes pour réguler le niveau d'expression du gène situé en aval. Cependant, pour les motifs *rne-II* et *fixA*, seul des différences apparentes dans le niveau de luminescence entre le WT et *disr* ont été observées. D'autres travaux seront nécessaires pour déchiffrer le mécanisme exact de régulation de ces différents ARN.

4 DÉVELOPPEMENT D'UNE NOUVELLE MÉTHODE EXPÉRIMENTALE POUR LA DÉCOUVERTE DE *RIBOSWITCHS*.

Contribution :

Ce chapitre a été réalisé en collaboration avec Emilie Boutet, autre étudiante au doctorat au sein du laboratoire. L'article de la méthode « Shifted Reverse PAGE : a novel approach based on structure switching for the discovery of riboswitches and aptamers » dont je suis co-première auteure avec Emilie Boutet est disponible sur bioRxiv (Devinck *et al.*, 2022).

4.1 Mise en contexte et problématique

Comme nous avons pu le voir dans le chapitre précédent, il est difficile pour certains ARNnc en *cis* candidats de déterminer le signal permettant son action du fait qu'ils soient présents devant des gènes impliqués dans des fonctions plus globales chez les bactéries. De plus, de nombreuses classes d'ARNnc régulant en *cis* existent (*riboswitch*, thermorégulateurs, r-leader, T-box) pouvant répondre à différents signaux qu'il peut être difficile de moduler via des essais de variation de milieu de culture comme nous avons pu le faire dans le chapitre précédent. Il a été également montré que certains *riboswitchs* tel que *ykkC* sont présents devant une multitude de gènes avec des fonctions différentes rendant difficile l'identification du ligand (Nelson *et al.*, 2017). Or, les méthodes de bioinformatique ne cessent de fournir des candidats ARNnc potentiels (Barrick *et al.*, 2004; Stav *et al.*, 2019; Weinberg *et al.*, 2007; Weinberg *et al.*, 2017a; Weinberg *et al.*, 2010), ceux-ci ne sont pas toujours facilement associés à leur ligand et ne font qu'agrandir la liste des *riboswitchs* ou ARNnc en *cis* dits orphelins (Greenlee *et al.*, 2018; Sherlock & Breaker, 2020). Ainsi, l'approche actuelle pour la découverte des nouveaux ARNnc en *cis* et des *riboswitchs* basée sur la comparaison de séquences et structures ne permettrait de découvrir qu'une fraction de toutes les classes de *riboswitchs* qui existent (Breaker, 2011). Les *riboswitchs* restants seraient plus rares, ce qui les rendraient beaucoup plus difficiles à découvrir avec les outils de bioinformatique actuels. De plus, les motifs présents dans des classes de bactéries présentant peu de séquences tels que les Acidobacterales ou encore les Désulfurococcales comportant une dizaine de bactéries séquencées, comparés aux Entérobactériales qui présentent plus de 3000 séquences, ne ressortent que peu lors des recherches via les outils de bioinformatique actuels. Ainsi, une méthode ne reposant pas sur les séquences disponibles dans les bases de données publiques pourrait être le meilleur moyen pour contourner les limites de ces méthodes. Nous avons ainsi développé une méthode expérimentale suite à cette

problématique nommée le Shifted-Reverse PAGE, celle-ci est créée de manière à pouvoir découvrir de nouveaux *riboswitchs* en palliant aux limites précitées des outils de bioinformatique.

L'expression des gènes chez les bactéries a besoin d'être finement régulée afin de pouvoir s'adapter à des changements constants de conditions environnementales mais aussi pouvoir maintenir l'homéostasie au sein de la cellule (Hennecke, 1990). Les cellules sont capables d'importer des molécules à partir du milieu extracellulaire et leurs concentrations peut être cruciale pour moduler des cascades enzymatiques (Hennecke, 1990). Un élément présent sur l'ARNm dans la partie 5' UTR appelé *riboswitch* peut lier des métabolites aussi appelés ligands, qui mènera à un changement de sa structure secondaire. Ce changement peut affecter la transcription ou la traduction du gène situé directement en aval de ce dernier. Le gène affecté par la présence du *riboswitch* code souvent pour la synthèse ou le transport du métabolite (Serganov & Nudler, 2013). Les *riboswitchs* sont composés de deux domaines fonctionnels : un domaine aptamère et une plateforme d'expression dont la structure dépend de la liaison du ligand associé à cet ARN. Jusqu'à maintenant, plus de 40 classes de *riboswitchs* ont été découvertes chez les bactéries (Barrick *et al.*, 2004; Mandal *et al.*, 2003; Mandal *et al.*, 2004; McCown *et al.*, 2017; Sherlock & Breaker, 2020; Sherlock *et al.*, 2018a; White *et al.*, 2022b; Winkler & Breaker, 2003; Xu & Cotruvo, 2022; Yu & Breaker, 2020). De façon intéressante, le *riboswitch* TPP a été observé dans les trois domaines de la vie, à savoir les archées, les eucaryotes et les procaryotes (McCown *et al.*, 2017).

La découverte de la plupart des *riboswitchs* connus à nos jours a été possible grâce à la disponibilité des génomes de bactéries séquencées combinée avec des approches bioinformatiques puissantes (Stav *et al.*, 2019; Weinberg *et al.*, 2007; Weinberg *et al.*, 2017a; Weinberg *et al.*, 2017b; Weinberg *et al.*, 2010). En effet, les potentiels *riboswitchs* ont été identifiés en fonction de la présence de covariation et de conservation dans la séquence car le domaine aptamère doit être hautement conservé pour reconnaître le ligand d'intérêt. La covariation résulte de la variation de deux positions correspondant à la même paire de base au sein de la tige d'une séquence de l'ARN qui ne cause pas de changement dans la structure secondaire de ce dernier, renforçant l'importance de la structure pour leur fonctionnement (Weinberg *et al.*, 2010). Cependant, ces approches bioinformatiques, bien que puissantes, présentent des limitations comme la disponibilité des séquences et leurs annotations dans les bases de données. Aussi, pour l'étude de la covariation, en théorie un alignement de trois séquences est suffisant pour une prédiction de structure secondaire et avancer vers une étude expérimentale d'une *riboswitch* potentiel. L'identification du ligand reconnu par le domaine

aptamère est souvent facilité par le gène situé en aval. Cependant, l'annotation des gènes est parfois incomplète ou mauvaise et ne permet pas toujours de donner l'information pour identifier le métabolite reconnu par le ligand.

Au cours des dernières années, des méthodes expérimentales ont été développées par d'autres groupes pour découvrir de nouveaux *riboswitchs*. Parmi celles-ci, le Term-seq comme mentionné précédemment, cette méthode permet de trouver toutes les terminaisons de transcription dans un organisme choisi en utilisant une méthode de séquençage à haut débit (Dar *et al.*, 2016). Ceci représente la limite principale de la méthode car elle permet d'identifier que les arrêts précoces de transcription qui peuvent également être provoqués par autre chose qu'un *riboswitch*. Une autre méthode expérimentale est Parallel Analysis of RNA Conformations Exposed to Ligand binding (PARCEL) (Tapsin *et al.*, 2018). Il s'agit d'une méthode *in vitro* permettant de détecter des changements structuraux des ARN en présence d'un ligand en comparant les sites de clivages par l'ARNase V1. Cependant, cette technique est limitée aux régions qui sont accessibles par l'ARNase et seulement un transcriptome peut être analysé à la fois. Une autre technique dérivée du SHAPE-MaP (Siegfried *et al.*, 2014) a été développée pour analyser l'interaction entre les molécules d'ARN et leur ligand. Elle pourrait être utilisée pour cribler des *riboswitchs* comme le propose la méthode PARCEL, mais avec des limitations similaires.

Des aptamères pour de nombreux ligands ont été identifiés grâce à des méthodes de SELEX (Systematic Evolution of Ligands by EXponential Enrichment) (Irvine *et al.*, 1991), les premiers à avoir été découverts sont un aptamère dirigé pour la T4 DNA polymérase (Tuerk & Gold, 1990) et des aptamères pour divers colorants organiques (Ellington & Szostak, 1990). L'idée générale de ces méthodes est de commencer avec une librairie extrêmement diverse de séquences ($\sim 10^{16}$ séquences) et une molécule d'intérêt. Les séquences ne liant pas cette molécule sont séparées des autres par une méthode de sélection. Les séquences sélectionnées (liant la molécule) sont amplifiées pour le nouveau cycle de sélection. Pour tous les SELEX l'idée générale reste la même, cependant diverses méthodes de sélection ont été développées, comme l'utilisation d'électrophorèse par capillaire (CE-SELEX) (Zhu *et al.*, 2019), le capture-SELEX (Boussebayle *et al.*, 2019) et l'utilisation de billes magnétiques (Espelund *et al.*, 1990) pour n'en nommer que quelques-unes (revue disponible (Bayat *et al.*, 2018; Darmostuk *et al.*, 2015)). Nous avons émis l'hypothèse qu'une méthode basée sur le SELEX pourrait être appliquée pour découvrir de nouveaux *riboswitchs*. Cependant, il y a certaines limites dans ces méthodes de SELEX classique pour trouver des *riboswitchs* naturels. En effet, ces molécules de régulation englobent souvent complètement leur ligand. Ainsi, immobiliser le ligand sur une colonne pourrait causer un

encombrement stérique qui ne permettrait pas de sélectionner de motifs. Même dans le cas où les *riboswitchs* n'enveloppent pas leur ligand, les groupements chimiques choisis pour l'immobilisation ont une probabilité élevée de faire partie des groupes nécessaires pour la liaison à l'ARN. En ce qui concerne les méthodes permettant d'immobiliser la librairie au lieu du ligand, comme le capture-SELEX (Boussebayle *et al.*, 2019), il serait difficile de l'utiliser pour trouver de nouveaux *riboswitchs* naturels car il est nécessaire d'avoir la séquence de capture au sein de l'ARN.

Nous avons ainsi développé au sein du laboratoire une nouvelle méthode expérimentale capable d'outre-passer les limites de la bioinformatique, mais aussi celles des autres méthodes expérimentales développées jusqu'à ce jour. Cette méthode est basée sur les deux caractéristiques principales des *riboswitchs* à savoir, leur capacité à lier un ligand et le changement de conformation de la plateforme d'expression en réponse à cette liaison. Ces caractéristiques sont évaluées grâce à une approche d'électrophorèse ressemblant aux gels à deux dimensions (2D), mais au lieu de migrer le gel dans une seconde dimension, les électrodes sont inversées. Cette méthode permet la détection des changements de conformation des ARN suite à la liaison du ligand comparé aux ARN qui n'ont pas interagit avec ce dernier. Nous avons appelé cette technique le Shifted-Reverse PolyAcrylamide Gel Electrophoresis (SR-PAGE).

4.2 Matériel et méthodes

4.2.1 Constructions des *riboswitchs*

Les différents *riboswitchs* testés ont été amplifiés par PCR à partir des génomes d'*E. coli* et de *B. thailandensis* pour les *riboswitchs* FMN et fluor, respectivement en prenant soin d'ajouter le promoteur T7 en amont pour réaliser la transcription *in vitro*. Les *riboswitchs* glycine (*Vibrio cholerae*), TPP (*E. coli*), c-di-GMP classe I (*Vibrio cholerae*) et Nickel/cobalt (*Listeria monocytogenes*) ont été amplifiés via une méthode de PCR d'assemblage, les amorces ont été construites à l'aide de l'outil Primerize (Tian & Das, 2017). Les réactions de PCR ont été réalisées en utilisant 2,5 U de Hot Start Taq Polymérase (Qiagen/203203) avec son tampon à une concentration finale de 1X, 0,2 mM de dNTP et 1 μ M d'amorces (forward et reverse). Pour les PCR d'assemblage, seuls les amorces des extrémités sont à une concentration de 1 μ M, les amorces centrales elles sont présentes à une concentration de 0,1 μ M. Les amorces utilisées sont listées dans le Tableau 4.1.

Tableau 4.1 Liste des amorces utilisées pour les différents contrôles des *riboswitchs*.

<i>Riboswitch</i> Fluor : amplifié à partir de <i>B. thailandensis</i>.
<ul style="list-style-type: none"> - Fluor_F : TAATACGACTCACTATAGGgcggtaccggagatgg <p>Même amorce forward pour toutes les constructions de ce <i>riboswitch</i>.</p>
<ul style="list-style-type: none"> - Fluor_const_1_R : atggcctacgacctcctg
<ul style="list-style-type: none"> - Fluor_const_2_R : tacgacctcctgcaccc
<ul style="list-style-type: none"> - Fluor_const_3_R : cctcctgcacccagg
<ul style="list-style-type: none"> - Fluor_const_4_R : tgcacccaggaacgc
<ul style="list-style-type: none"> - Fluor_const_5_R : aggaacgcgtaggcatca
<i>Riboswitch</i> FMN : amplifié à partir d'<i>E. coli</i>.
<ul style="list-style-type: none"> - FMN_F : TAATACGACTCACTATAGGgcttattctcagggcgg <p>Même amorce forward pour toutes les constructions de ce <i>riboswitch</i>.</p>
<ul style="list-style-type: none"> - FMN_const_1_R : attatggttaccagaatcagggca
<ul style="list-style-type: none"> - FMN_const_2_R : cagggcagtcttaggagt
<ul style="list-style-type: none"> - FMN_const_3_R : gtggcggcatatagccaa
<ul style="list-style-type: none"> - FMN_const_4_R : aacgtgagcgggtccat
<ul style="list-style-type: none"> - FMN_const_5_R : tcgttactctctcccatcc
<i>Riboswitch</i> di-GMPc classe I : PCR d'assemblage pour le <i>riboswitch</i> de <i>V. cholerae</i>.
<ul style="list-style-type: none"> - di-GMPc_const_1_F1 : TAATACGACTCACTATAGGgaaaaatgtcacgcacagggc - di-GMPc_const_1_R1 : ggtttaggccggaggcttgcgtcccactcttcgaatggttgcctgtgctg - di-GMPc_const_1_F2 : ctccggcctaaaccagaagacatggttaggtagcggggtaccgatggcaaaatgcataca - di-GMPc_const_1_R2 : aaagcatgcattcatagtgtcaatgatgagtcaacaaagtgtatgcattttgccatcg <p>Cette construction a été utilisée contre ADN de départ pour les autres constructions de ce <i>riboswitch</i>. L'amorce di-GMPc_const_1_F1 est l'amorce forward pour toutes les constructions.</p>

- di-GMPc_const_2_R : ggtcaatgatgagcaacaaagt
- di-GMPc_const_3_R: tgatgagcaacaaagtgtatgc
- di-GMPc_const_4_R: aaagtgtatgcattttgccatc
- di-GMPc_const_5_R: gtatgcattttgccatcgta
Riboswitch Glycine : PCR d'assemblage pour le riboswitch de <i>V. cholerae</i>.
- Gly_F1: TAATACGACTATAGGgttgaagactgcaggagagtggttgaaccagatttaacatctgagccaataac ccg
- Gly_R1 : tcgcttattcgttgccaatatatggctaagaataatgcacctgaaagattactcttcggcggggtatttggtcagatg
- Gly_F2 : ttggcaacgaataagcgaggactgtagtggaggaacctctggagagaaccgttaacggctgccgaaggagc aag
- Gly_R2 : tcctctgtcctttgcctgagagtttactctgcatatgctgcagagcttctccttcggcgaccgat
Riboswitch TPP : PCR d'assemblage pour le riboswitch d'<i>E. coli</i>.
- TPP_F : TAATACGACTCACTATAgactcggggtgcccttctgctggaaggctgagaaataccggtatcacc
- TPP_R : gtggccggtgaactccctacggctggcattatccagatcaggtgatacgggtatttctc
Riboswitch Nickel/Cobalt : PCR d'assemblage pour le riboswitch de <i>L. monocytogenes</i>.
- NiCo_const_1_F : TTCTAATACGACTCACTATAGGgcggttgagctaatacattgtgatctgaacaggcg
- NiCo_const_1_R : aaaaaaggcatacccgtagaacagaattaac
ADN de départ : construction NiCo_const_4.
- NiCo_const_2_F : TTCTAATACGACTCACTATAGGgcggttgagctaatacattgtgatctgaacaggcg
- NiCo_const_2_R : taccgtagaacagaattaactgtcccacagatgac
ADN de départ : construction NiCo_const_4.
- NiCo_const_3_F : TTCTAATACGACTCACTATAGGgcggttgagctaatacattgtgatctgaacaggcg
- NiCo_const_3_R : cccgtagaacagaattaactgtcccacagatgac
ADN de départ : construction NiCo_const_4.
- NiCo_const_4_F : TTCTAATACGACTCACTATAGGgcggttgagctaatacattgtgatctgaacaggcg
- NiCo_const_4_R : aactgtcccacagatgacctt

ADN de départ : construction NiCo_const_5.
<ul style="list-style-type: none"> - NiCo_const_5_F1 : TTCTAATACGACTCACTATAGGtcattgtgatctgaacaggcggtgaacgtaa - NiCo_const_5_R1 : tgcagcccagctgcatgaacctcgtgttacggtcaccgcctg - NiCo_const_5_F2 : cagctgggctgcaattatttgcggcagcagactatgtattctaagggcatactgtggga - NiCo_const_5_R2 : aactgtcccacagatatgccctt

Tableau représentant les amorces utilisées pour tous les *riboswitchs* contrôlés. Les lettres « F » et « R » correspondent aux amorces forward et reverse. « Const » est pour construction. Les nucléotides en majuscules représentent le promoteur T7 qui sert pour la transcription *in vitro* des ARN.

Concernant les bibliothèques pour la modification de l'affinité du *riboswitch thiM* liant le TPP, des amorces dégénérées ont été utilisées contenant 6, 10 et 18 nucléotides aléatoires (N). Les mêmes conditions de PCR d'assemblage que précédemment ont été utilisées, mais uniquement avec une amorce forward et reverse. Ces amorces ont été listées dans le Tableau 4.2.

Tableau 4.2 Liste des amorces utilisées pour la réalisation des bibliothèques du *riboswitch* TPP.

Librairie TPP 1 _ 6 nucléotides aléatoires
<ul style="list-style-type: none"> - Lib 1_F : TAATACGACTCACTATAggactcgggggtgcccttctgctggaaggctgagaaatacccgatcacc - Lib 1_R : gtgtccgtgaactccctnnnctggcattatccagnnaggatgacgggtatttctc
Librairie TPP 2 _ 10 nucléotides aléatoires
<ul style="list-style-type: none"> - Lib 2_F : TAATACGACTCACTATAggactcgggggtgcccttctgctggaaggctgagaaatacccgatcacc - Lib 2_R : gtgtccgtgaactcccnnnnntggcattatccannnnnggtgatacgggtatttctc
Librairie TPP 3 _ 18 nucléotides aléatoires
<ul style="list-style-type: none"> - Lib 3_F : TAATACGACTCACTATAggactcgggggtgcccttctgctggaaggctgagaaatacccgatcacc - Lib 3_R : gtgtccgtgaactcccnnnnnnnnnnnnnnnnnggtgatacgggtatttctc

Tableau représentant les amorces utilisées pour les PCR d'assemblage pour construire les bibliothèques initiales pour le SELEX impliquant le *riboswitch* TPP afin de modifier son affinité. Les lettres « F » et « R » correspondent aux amorces forward et reverse. Les nucléotides en majuscules représentent le promoteur T7 qui sert pour la transcription *in vitro* des ARN. Les nucléotides « n » en rouge, sont les nucléotides dégénérés.

4.2.2 Transcription *in vitro*

Pour le SR-PAGE (essais pour les *riboswitchs* connus, clones de TPP), tous les produits de PCR ont été marqués à la radioactivité au cours de la transcription avec de l' $[\alpha\text{-}^{32}\text{P}]$ UTP (Perkin Elmer). Les conditions sont les mêmes que présentées dans la section 3.2.7 mis à part la concentration de rNTP (5 mM pour ATP, GTP et CTP et 1mM pour UTP). Pour cette transcription radioactive, les ARN ont été visualisés, élués et précipités comme expliqué précédemment. Enfin, les ARN radioactifs ont été resuspendus dans 250 μL d'eau Milli-Q.

Pour les essais de cartographie *in vitro* par *in-line*, les ARN ont été marqués au niveau de leur extrémité 5' avec des $[\gamma\text{-}^{32}\text{P}]$ ATP (Perkin Elmer). La transcription *in vitro*, la déphosphorylation, et le marquage ont été réalisées dans les mêmes conditions qu'expliquées dans la section 3.2.7.

4.2.3 Préparation du gel de SR-PAGE.

Un système typique des gels de séquençage a été utilisé. Pour faire le gel, des plaques de verre de dimensions (38 cm x 45 cm) pour la grande plaque et (38 x 43 cm) pour la petite plaque ont été utilisées. La petite plaque de verre a été traitée avec une solution de Rain-X®, permettant de rendre la plaque hydrophobe. Dans un même temps, la grande plaque de verre a été traitée avec une solution d'hydroxyde de potassium (5 g KOH) et méthanol (100 mL) rendant la plaque hydrophile. Ces deux traitements sont très importants et permettent de faciliter le démoulage du gel pour des étapes prochaines. Les plaques de verre ont été ensuite assemblées en prenant soin de laisser un écart de 2 cm en bas entre la petite et la grande plaque (Figure 4.1 A). Des espaceurs (environ 0,8 mm) ont été placés entre les deux plaques de verre pour créer l'épaisseur du gel.

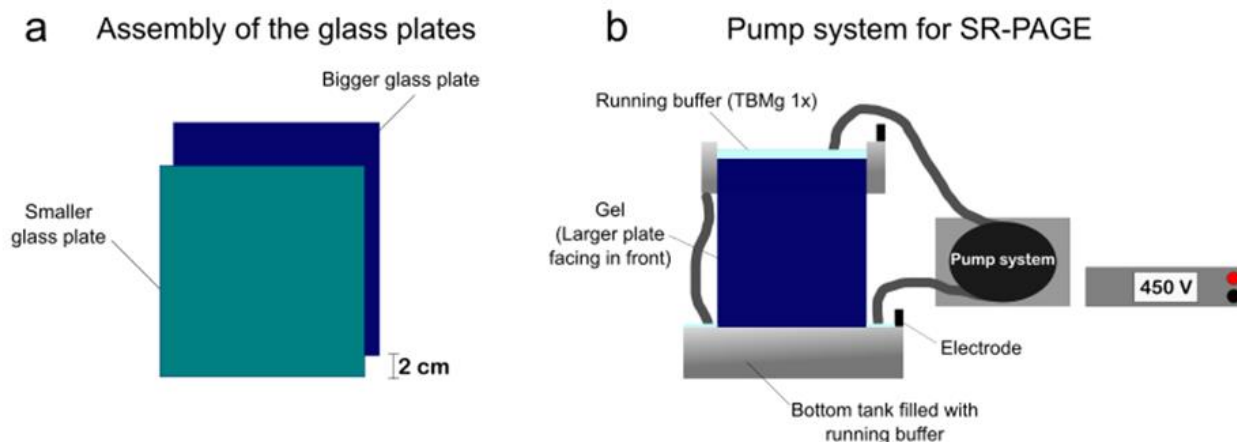


Figure 4.1 Assemblage du système de SR-PAGE.

A : La petite et la grande plaque de verre traitées avec du Rain-X® et KOH + méthanol respectivement sont assemblées en laissant un espace de 2 cm en bas de la petite plaque. **B** : Le système du SR-PAGE au cours de la migration où le gel est installé avec la grande plaque de verre sur le devant. Le système de pompe péristaltique (pump system) permet de faire remonter le tampon de migration (TBMg 1X) dans la cuve supérieure. Tiré de (Devinck *et al.*, 2022).

Une solution de 250 mL de polyacrylamide natif (29 :1) a été préparée avec une concentration finale de TBMg à 1X (0,09 M Tris Base, 0,09 M acide borique and 5 mM $Mg(CH_3COO)_2$ pH 8 à température pièce). Le volume de la solution a été ajusté avec de l'eau Milli-Q stérilisée. Un volume de 50 mL de cet acrylamide a été gardé pour être utilisé ultérieurement de manière à garder l'exacte même concentration d'acrylamide. 2 mL d'ammonium persulfate 10% (APS, Biorad) et 70 μ L de tetraméthyléthylènediamine (TEMED, Bioshop) ont été utilisés pour la polymérisation du gel. Il est à noter que plusieurs marques d'APS ont été testées et que le SR-PAGE ne fonctionnait qu'avec l'APS Biorad. Le gel a été coulé entre les deux plaques de verre et laissé à polymériser pour au moins 30 minutes. Le système a ensuite été monté comme montré dans la Figure 4.1 B. Les migrations du SR-PAGE sont longues ainsi pour permettre une bonne circulation du tampon de migration et des ions qui le composent pendant toute la durée de l'expérience, une pompe péristaltique a été installée permettant de transporter le tampon entre les deux compartiments du système (inférieur et supérieur) (Figure 4.1 B). Une pré-migration a été réalisée dans du tampon de migration (TBMg 1X) toute la nuit à 450V dans la salle à 4°C.

4.2.4 Préparation des échantillons

Les échantillons d'ARN ont été préparés avec du TBMg 1X, 3,3 μ L de bleu natif 6X (40% sucrose, 0,05% de bleu de bromophénol, 0,05% de bleu de xylène cyanol). Plusieurs contrôles ont été

utilisés pour vérifier la bonne migration, tout d'abord des ARN de tailles différentes (court et long). En effet, avec notre méthode quel que soit la taille de l'ARN s'il n'interagit pas avec le ligand vaporisé il reviendra au niveau des puits. Un autre contrôle a été l'utilisation d'une amorce fluorescente (Cy3), celle-ci permet de visualiser la migration des acides nucléiques à l'œil nu (5' - Cy3-tccccttgcatacatgtcgcggttatgcttcgaccctctagccgcacccttgg - 3' ou encore 5' - Cy3-ccatcggcactgacggcctcactgcagcagaacgtggcagcttgctgaatcggtctgcaagagtgagcccgtaacataatgg - 3') et permet d'aider l'arrêt de la seconde migration au bon moment. Enfin, un *riboswitch* connu pour être décalé (*shifté*) avec notre méthode a été utilisé comme contrôle positif. Les contrôles ont été préparés dans les mêmes conditions que nos échantillons.

Les puits ont été nettoyés avec une seringue, pour enlever les bulles et l'acrylamide non polymérisé pouvant se trouver dans les puits, avant la pré-migration mais aussi avant de déposer les échantillons. Les ARN ont été séparés dans le gel en fonction de leur taille et de leur structure lors d'une migration de 24 h à 450 V.

4.2.5 Migration inverse

Après une première migration de 24 heures, les plaques de verre ont été séparées. Une solution de ligand d'un volume de 50 mL a été préparée avec des concentrations finales équivalentes à 100 fois le K_D connu pour le *riboswitch* concerné. En effet, seulement une portion de cette solution entrera dans le gel, il faut donc mettre une concentration en excès. Les concentrations vaporisées pour chaque ligand sont les suivantes : 6 mM FH_4KO_2 (Baker *et al.*, 2012) , 5 μM FMN sodium hydraté (Howe *et al.*, 2016) (Cayman chemical company), 3 μM di-GMPc sodium (Smith *et al.*, 2009) (Sigma-Aldrich), 10 mM glycine (Huang *et al.*, 2010) (Bioshop) , 3 μM chlorure de TPP (Rentmeister *et al.*, 2007) (Sigma-Aldrich), 3 mM NiCl_2 (Furukawa *et al.*, 2015) (Fisher) et/ou 3 mM CoCl_2 (Furukawa *et al.*, 2015) (Fisher), le tout dans du TBMg 1X. Lorsqu'un ion métallique tel que le Ni^{2+} a été vaporisé, 10 mM de L-glutathion réduit ($\text{C}_{10}\text{H}_{17}\text{N}_3\text{O}_6\text{S}$) (Sigma-Aldrich) a été ajouté à la solution de ligand pour éviter l'oxydation et la précipitation. Le volume a par la suite été ajusté à 50 mL avec de l'eau Milli-Q. Cette solution a été vaporisée directement et uniformément sur le gel et laissée en contact pendant 10 minutes. Après cette incubation, l'excédent de solution a été enlevé en inclinant la plaque de verre. Les puits ont été coupés grâce à une lame et l'espace créé a été méticuleusement séché avec des Kimwipes® (Kimtech). Un volume de 70 μL de TEMED a été ajouté à la base des puits coupés pour aider à la jonction entre le gel déjà présent et celui qui sera ajouté. La petite plaque a été remise en prenant soin cette fois-ci d'aligner parfaitement le bas des deux plaques. La solution d'acrylamide qui avait été mise

de côté a été utilisée pour remplir l'espace laissé par les puits et le réalignement des plaques de verre. Les ligands vaporisés ont également été ajoutés à cette solution, mais à une concentration dix fois plus faible pour tenir compte de la proportion réellement entrée dans le gel. Enfin, pour permettre la polymérisation du gel, 700 μL d'APS 10% et 70 μL de TEMED ont été ajoutés. Le gel a été coulé puis laissé polymériser pour au moins 30 min. La migration a ensuite été relancée à 450 V en inversant les polarités du gel (négatif en bas et positif en haut).

La migration a été arrêtée quand l'amorce fluorescente arrive au niveau des puits. L'augmentation de la résistance du gel avec l'ajout d'acrylamide entraîne une migration plus rapide lors de cette étape. En effet, il a été remarqué que la migration inverse pouvait prendre entre 19 h et 23 h, au lieu des 24 h dans le sens initial. Lors de certaines expériences, il a été nécessaire de purifier les ARN ayant montré un décalage de migration, comme pour les expériences de SELEX du *riboswitch* TPP. Dans ce cas des bandes de gel au-dessus des puits ont été découpées, la bande « 0 » correspond à la ligne de base, la bande « 1 » correspond à une bande située entre 2 mm au-dessus du puits jusqu'à 1 cm au-dessus, puis enfin la bande « 2 » correspond à une bande située entre 1 cm au-dessus du puits avec tout le reste du haut du gel (Figure 4.2).

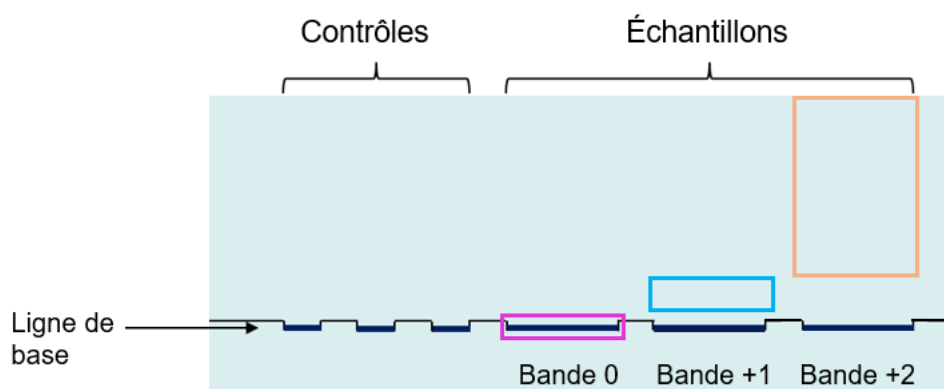


Figure 4.2 Représentation schématique des coupures réalisées lors du SELEX avec le SR-PAGE.

Après les deux migrations de SR-PAGE, les bandes de gel contenant les ARN ayant été décalés lors de la migration ont été découpées en prenant comme référence les anciens puits qui est la ligne de base. La bande « 0 » est représentée en rose, la bande « +1 » est représentée en bleue et la bande « +2 » est représentée en orange.

Les ARN de ces bandes de gel ont été élués et précipités comme décrit précédemment. Quand cette étape n'était pas nécessaire, le haut du gel (environ 15 cm) a été coupé, séché (gel dryer model 585 Biorad avec une pompe à vide CVC3000 Vaccumbrand). Le gel ou le reste du gel issu

de la découpe des bandes ont été exposés toute la nuit sur un écran de phosphore et visualisé avec le Typhoon FLA9500 (GE Healthcare Life Sciences).

4.2.6 Sélection grâce au SELEX de *riboswitchs* TPP avec une affinité modifiée

Le SR-PAGE peut être une méthode utilisée au sein d'un SELEX afin de sélectionner les ARN ayant une meilleure affinité pour un ligand. Pour la sélection des *riboswitchs* TPP ayant une affinité modifiée pour leur ligand, les différentes bibliothèques dégénérées (1, 2 et 3) décrites dans la section 4.2.1 ont été soumises à des cycles de SELEX. Les ARN ayant été décalés ont été élués du gel via un tampon d'éluion (0,3 M NaCl) et incubés en rotation toute la nuit et précipités comme décrit précédemment. Une rétrotranscription (RT) a été effectuée en utilisant la transcriptase inverse M-MuLV (New England Biolabs) conformément aux instructions du fabricant. La RT a été effectuée avec une amorce courte permettant d'éviter l'hybridation à des ARN qui auraient perdu des nucléotides du côté 3' lors de la migration. Ces ARN auraient ainsi migré plus rapidement dans la migration inverse. La moitié de la réaction de RT a été utilisée comme ADN de départ pour les réactions de PCR, qui ont été effectuées dans les mêmes conditions que pour créer la bibliothèque de départ mais avec les amorces JO-001 (5' - TAATACGACTCACTATAggactcgggggtgccctt - 3') et JP026 (5' - gtggtccggaactccc - 3'). Ces réactions de PCR ont été ensuite transcrites *in vitro* pour recommencer un cycle de sélection comme indiqué précédemment.

Pour les migrations, 10 μ M de TPP a été ajouté au tampon de migration afin que le TPP soit présent dans les deux dimensions. Pour la sélection, une solution de thiamine a été vaporisée. La concentration de thiamine a été diminuée à différents cycles afin d'augmenter la stringence de la sélection. Tout d'abord entre les générations 1 et 4, 2 mM de thiamine a été vaporisé. Pour les générations 5 et 6, 20 μ M a été vaporisé. Enfin pour les générations 7 à 10, 200 nM de thiamine a été ajouté.

4.2.7 Cartographie *in vitro* par *in-line* et constante de dissociation

Une partie des réactions PCR des bibliothèques initiales et des cycles de sélections aux générations 4, 5, 7 et 10 ont été clonées en utilisant le kit pGEM®-T easy vector (Promega) conformément aux recommandations du fabricant pour évaluer des séquences individuelles. Les mêmes échantillons ont également été séquencés par une technologie de séquençage à haut débit Illumina (Centre d'expertise et de services Génome Québec). Après analyse des résultats de séquençage Illumina, plusieurs séquences se sont retrouvées enrichies lors des cycles. Les

séquences individuelles ont été amplifiées par PCR, soit directement sur le plasmide, soit par une PCR d'assemblage dans les mêmes conditions que précédemment. Les amorces sont listées dans le Tableau 4.3. Les ARN ont ensuite été transcrits, purifiés sur un gel de polyacrylamide dénaturant à 6% élués et radiomarqués comme décrits dans des sections antérieures.

Tableau 4.3 Liste des amorces utilisées pour le SELEX du *riboswitch* TPP.

Amplification des séquences individuelles (plasmide)
<ul style="list-style-type: none"> - TPP_plasm_F: TAATACGACTCACTATAggactcggggtgcctt - TPP_plasm_RS: gtggtccgtgaactccc
Amplification des séquences enrichies (par PCR d'assemblage)
<ul style="list-style-type: none"> - TPP_enrich_F: TAATACGACTCACTATAggactcggggtgccttctgctggaaggctgagaaataccggtatcacc <p>Même amorce sens pour toutes les séquences.</p>
<ul style="list-style-type: none"> - TPP1_R: gtggtccgtgaactccctaggctggcattatccaggcgaggtgatacgggtatttctc - TPP2_R: gtggtccgtgaactccctgtctggcattaccaggtcaggtgatacgggtatttctc - TPP3_R: gtggtccgtgaactcccagcatggcattatccaaacaagggtgatacgggtatttctc - TPP4_R: gtggtccgtgaactcccaagcatggcattatccaaacaagggtgatacgggtatttctc - TPP5_R: gtggtccgtgaactcccagcattggcattatccaaaaccgggtgatacgggtatttctc - TPP6_R: gtggtccgtgaactcccaaccatggcattatccaaagcaggtgatacgggtatttctc - TPP7_R: gtggtccgtgaactccctcgctggcattatccagacgaggtgatacgggtatttctc - TPP8_R: gtggtccgtgaactccctgctctggcattatccaaactggggtgatacgggtatttctc - TPP45_R: gtggtccgtgaactcccaacaacaagcaggtgttgggtgatacgggtatttctc - TPP46_R: gtggtccgtgaactcccaacacatctccattgttgggtgatacgggtatttctc - TPP47_R: gtggtccgtgaactccctaccaacactgttgctagggtgatacgggtatttctc - TPP48_R: gtggtccgtgaactcccagcgtccgatctgctcgggtgatacgggtatttctc

Tableau représentant les amorces utilisées pour le SELEX de TPP. Tout d'abord, les amorces utilisées pour amplifier les séquences contenues dans les plasmides suite à l'utilisation du kit pGEM®-T easy vector et le séquençage Sanger. Deuxièmement, les amorces utilisées pour amplifier les séquences retrouvées enrichies lors de l'analyse du séquençage Illumina. Le nom de chaque amorce reverse correspond au numéro du clone. Pour toutes les amorces les lettres « F » et « R » correspondent à forward et reverse respectivement. Les nucléotides en majuscules correspondent au promoteur T7 pour la transcription des ARN.

Les réactions de cartographie *in vitro* par *in-line* (Regulski & Breaker, 2008) ont été effectuées comme décrit dans la section 3.2.7 en utilisant différentes concentrations de ligands allant de 30 nM à 1 mM final de thiamine et de TPP. Sauf pour l'essai avec le *riboswitch* TPP WT pour lequel des concentrations de 1 nM à 3 µM final ont été utilisées pour la molécule de TPP.

4.2.8 Calcul des énergies libres et structure secondaire des ARN

L'énergie libre de toutes les séquences d'ARN a été évaluée en utilisant le logiciel Mfold accessible à l'adresse www.unafold.org (Zuker, 2003). Les ARN ont été forcés à prendre la structure de leur aptamère en utilisant des contraintes basées sur des études de résolution atomique des différents aptamères. Les contraintes utilisées sont représentées dans le Tableau 4.4. Les structures secondaires d'ARN ont quant à elle été réalisées en utilisant les fichiers Stockholm générés par Mfold (Zuker, 2003) avec R2R (Weinberg & Breaker, 2011).

Tableau 4.4 Contraintes pour la réalisation des structures des *riboswitchs*.

<i>Riboswitch</i>	Contraintes
Fluorxx((((.....xxx((((.....))))))xxxxx))))xx.....
FMN	..(((((((..(((((.....))))..))....((((.....))))....))))..(((((((.....))))....))....((((.....)))) ..(((((((.....))))..))....))))....
di-GMPc I((((xxxxxx((...((((.....))))))....))xxx(((.....((((.....))))....))))xxx)..
Nickel-cobalt(((.....((.....((((.....))))....))....))....(((.....))....((((.....))))....))..
Stem P2-P3, Library 3 TPP((((((((((((.....))))....))))....))))....

Tableau représentant les contraintes pour la réalisation des structures d'ARN pour le calcul de leur énergie libre. Les points représentent les nucléotides n'ayant pas de contraintes. Les « x » montrent les nucléotides qui ont été forcés à rester non liés. Enfin les parenthèses représentent les interactions entre les nucléotides.

4.2.9 Tests adaptateurs dans le cadre d'un SELEX

Les constructions des *riboswitchs* (FMN_3, Glycine, TPP, di-GMPc_4 et di-GMPc_5) ont été obtenues via une amplification par PCR, comme décrit précédemment. Les constructions correspondent aux constructions précédentes (Tableau 4.1). Les amorces utilisées sont présentées dans le Tableau 4.5. Les ARN ont été transcrits avec de l'UTP radioactif et les échantillons ont été préparés comme dans la section 4.2.4 en ajoutant 0,5 µM d'oligonucléotides complémentaires (Tableau 4.5). Le mélange ARN et oligonucléotides a été incubé pendant 30 min à température pièce avant d'être déposé sur le gel. La migration a été effectuée comme décrit précédemment, en vaporisant les ligands correspondants aux échantillons présents.

Tableau 4.5 Liste des amorces utilisées pour le test des adaptateurs.

Construction FMN_3
<p>Amorces pour faire le template pour les autres PCR (pour ne pas avoir deux promoteurs T7)</p> <ul style="list-style-type: none"> - FMN_3_temp_F : gcttattctcagggcgg - FMN_3_temp_R : gtggcggcatatagccaa
<ul style="list-style-type: none"> - FMN_3_Adpt_F : TTAATACGACTCACTATAGGGA<u>gggaggacgatgcgggcttattctcagggcgg</u> - FMN_3_Adpt_R : <u>tcccgctcgtcgtctggtggcggcatatagccaa</u>
<ul style="list-style-type: none"> - FMN_3_Adpt_5'_F : TTAATACGACTCACTATAGGGA<u>gggaggacgatgcgggcttattctcagggcgg</u> - FMN_3_Adpt_5'_R : gtggcggcatatagccaa
Construction Glycine
<p>Amorces pour faire le template pour les autres PCR (pour ne pas avoir deux promoteurs T7)</p> <ul style="list-style-type: none"> - Gly_temp_F: gttgaagactgcaggagag - Gly_temp_R: tcctctgcctttgcctga
<ul style="list-style-type: none"> - Gly_Adpt_F: TTAATACGACTCACTATAGGGA<u>gggaggacgatgcgggtgaagactgcaggagag</u> - Gly_Adpt_R: <u>tcccgctcgtcgtctgcctctgcctttgcctga</u>
<ul style="list-style-type: none"> - Gly_Adpt_5'_F: TTAATACGACTCACTATAGGGA<u>gggaggacgatgcgggtgaagactgcaggagag</u> - Gly_Adpt_5'_R: tcctctgcctttgcctga
Construction TPP
<p>Amorces pour faire le template pour les autres PCR (pour ne pas avoir deux promoteurs T7)</p> <ul style="list-style-type: none"> - TPP_temp_F: ggactcggggtgccctt - TPP_temp_R: gtggtccgtcaagtcct
<ul style="list-style-type: none"> - TPP_Adpt_F: TTAATACGACTCACTATAGGGA<u>gggaggacgatgcggggactcggggtgccctt</u> - TPP_Adpt_R: <u>tcccgctcgtcgtctggtggtccgtcaagtcct</u>
<ul style="list-style-type: none"> - TPP_Adpt_5'_F: TTAATACGACTCACTATAGGGA<u>gggaggacgatgcggggactcggggtgccctt</u> - TPP_Adpt_5'_R: gtggtccgtcaagtcct
Construction di-GMPc_4 et _5
<p>Amorces pour faire le template pour les autres PCR (pour ne pas avoir deux promoteurs T7)</p> <ul style="list-style-type: none"> - di-GMPc_temp_F: ggaaaaatgtcacgcacaggg - di-GMPc_4_temp_R: caaagtgtatgcattttgccatc - di-GMPc_5_temp_R: gtatgcattttgccatcggta
<ul style="list-style-type: none"> - di-GMPc_4_Adpt_F: TTAATACGACTCACTATAGGGA<u>gggaggacgatgcggggaaaaatgtcacgcacaggg</u>

- di-GMPc_4_Adpt_R: <u>tcccgctcgtcgtctg</u> caaagtgtatgcattttgccatc
- di-GMPc_4_Adpt_5'_F : TTAATACGACTCACTATAGGGA <u>gggaggacgatgcgggg</u> aaaaatgacgcacaggg - di-GMPc_4_Adpt_5'_R: caaagtgtatgcattttgccatc
- di-GMPc_5_Adpt_F:TTAATACGACTCACTATAGGGA <u>gggaggacgatgcgggg</u> aaaaatgacgcacgcacaggg - di-GMPc_5_Adpt_R: <u>tcccgctcgtcgtctg</u> gtatgcattttgccatcgga
- di-GMPc_5_Adpt_5'_F : TTAATACGACTCACTATAGGGA <u>gggaggacgatgcgggg</u> aaaaatgacgcacaggg - di-GMPc_5_Adpt_5'_R: gtatgcattttgccatcgga
Oligonucléotides complémentaires aux adaptateurs pour tous les <i>riboswitchs</i>
- Oligo_1 : ccgcatcgtcctccc - Oligo_2 : gcatcgtcctccc
Oligonucléotides complémentaires aux adaptateurs pour la construction FMN_3
- Oligo_FMN_3 : gcccgcatcgtcctccc - Oligo_FMN_4 : aagcccgcatcgtcctccc
Oligonucléotides complémentaires aux adaptateurs pour le <i>riboswitch</i> Glycine
- Oligo_Gly_3 : cccgcatcgtcctccc - Oligo_Gly_4 : gtccccgcatcgtcctccc
Oligonucléotides complémentaires aux adaptateurs pour le <i>riboswitch</i> TPP
- Oligo_TPP_3 : gtccccgcatcgtcctccc - Oligo_TPP_4 : cccgcatcgtcctccc
Oligonucléotides complémentaires aux adaptateurs pour les constructions di-GMPc_4 et _5
- Oligo_di-GMPc_4+5_3: ttccccgcatcgtcctccc - Oligo_di-GMPc_4_5_4 : ccccgcatcgtcctccc

Tableau représentant les amorces utilisées pour le test des adaptateurs. « F » correspond aux amorces forward, « R » correspond aux amorces reverse. Les parties soulignées correspondent aux adaptateurs. Les parties en majuscules correspondent au promoteur T7.

4.3 Résultats

4.3.1 Validation de la méthode

Le SR-PAGE est une méthode basée sur la capacité des *riboswitchs* à changer de conformation en réponse à la liaison de son ligand qui causera un changement de migration dans un gel de polyacrylamide natif. La méthode est composée de deux migrations dans un gel natif, lors de la première migration les ARN sont séparés en fonction de leur taille et leur structure secondaire et tertiaire (Figure 4.3 A). Après cette première migration de 24h, les deux plaques de verre sont séparées et une solution de ligand(s) est vaporisée directement sur le gel pour permettre l'interaction avec les ARN contenus dans ce dernier (Figure 4.3 B). Après avoir coupé les puits et recoulé de l'acrylamide natif dans l'espace laissé par ceux-ci, une deuxième migration est relancée en inversant les polarités du gel. Les ARN rebrousseront ainsi chemin jusqu'au point de départ (Figure 4.3 C). Une fois la migration terminée, deux cas peuvent se présenter : les ARN n'ayant pas interagi avec l'un des ligands testés auront gardé leur structure initiale donc retourneront exactement au niveau des anciens puits. Cependant, les ARN ayant interagi avec l'un des ligands testés peuvent changer de conformation, ce qui leur conférera une structure plus compacte ou plus relâchée, et donc un changement de vitesse de migration. Cela se traduira par un décalage au-dessus ou en-dessous de la ligne des puits (Figure 4.3 C). Ainsi, le SR-PAGE est capable de reconnaître les ARN interagissant avec un ligand grâce au changement de vitesse de migration après son addition.

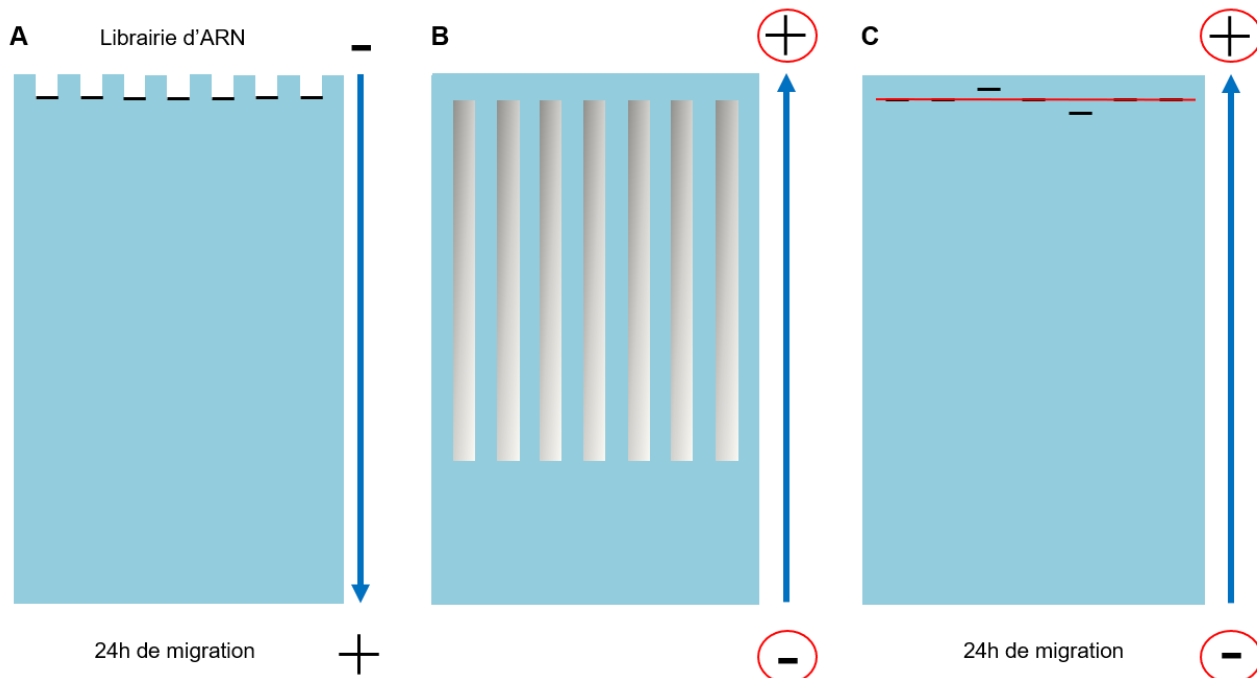


Figure 4.3 Principe de la méthode du SR-PAGE.

A : Une librairie d'ARN est déposée sur un gel de polyacrylamide natif et séparée dans le gel en absence de ligand. Les ARN sont ainsi séparés en fonction de leurs tailles et de leurs structures. **B :** Après 24h de migration, les plaques de verre sont séparées et le (ou les) ligand est vaporisé directement sur le gel. Ceci entraîne un changement de conformation des *riboswitchs* correspondant au ligand présent. La petite plaque est ensuite remplacée. Les puits sont coupés et de l'acrylamide de l'exact même concentration que le gel est coulé au niveau des anciens puits. Puis la migration est relancée en inversant la polarité. **C :** La migration du gel se fait jusqu'à ce que les ARN soient revenus au niveau des puits, appelée ligne de base (représentée en rouge). Les ARN qui ont interagi avec leur ligand auront une migration différente à cause de leurs changements de conformation, ce qui causera un décalage au-dessus ou en-dessous de la ligne de base. Les ARN n'ayant pas interagi avec un ligand reviendront au niveau de cette ligne de base.

Une preuve de concept a été réalisée pour valider que l'on pouvait avoir un changement de migration pour des *riboswitchs* connus. Cinq constructions pour chacun des motifs ont été réalisées. Différents ligands ont été testés tel que le FMN (Pedrolli *et al.*, 2015), le fluor (Breaker, 2012), le di-GMPc (classe I) (Sudarsan *et al.*, 2008), les ions nickel-cobalt (NiCo) (Furukawa *et al.*, 2015), le TPP (Rentmeister *et al.*, 2007) et la glycine (Kwon & Strobel, 2008). Pour les *riboswitchs* glycine et TPP seulement la construction cinq correspondant à l'aptamère a été réalisée basée sur des résultats préliminaires. La séquence pour le *riboswitch* glycine a été basée sur les résultats obtenus par Kwon *et al* montrant un décalage sur un gel de polyacrylamide natif (Kwon & Strobel, 2008). Pour les *riboswitchs* avec les cinq constructions, la construction minimale

ne comporte que le domaine aptamère, puis les différentes constructions sont différentes combinaisons dans la plateforme d'expression (Figure 4.4 A, B, C).

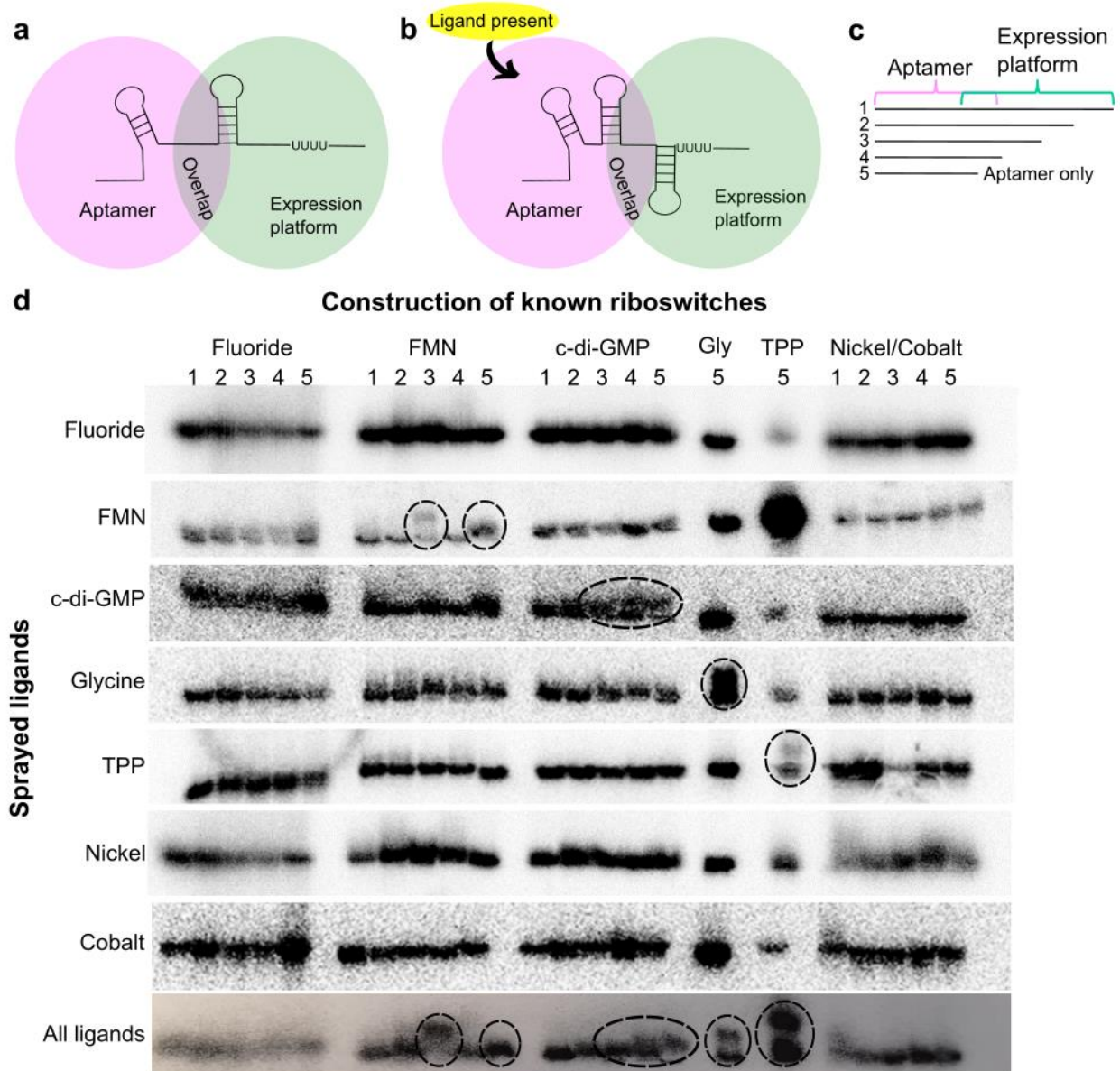


Figure 4.4 Validation de la méthode du SR-PAGE.

A : Représentation schématique d'un *riboswitch* en absence de son ligand, avec le domaine aptamère (dans le cercle rose) et la plateforme d'expression (dans le cercle vert). Les deux structures sont mutuellement exclusives grâce au chevauchement entre les deux parties. **B** : Représentation schématique d'un *riboswitch* en présence de son ligand causant un changement de conformation par rapport à A. **C** : Représentation des différentes constructions (1 à 5) pour chacun des *riboswitches* testés. **D** : Résultat de tous les gels de SR-PAGE pour la validation de la méthode. Chaque bande horizontale correspond à un gel où un ou l'ensemble des ligands a été testé. Les différentes constructions de *riboswitches* ont été marquées à la radioactivité. Les décalages observables ont été entourés. Tiré de (Devinck *et al.*, 2022).

Plusieurs essais de SR-PAGE ont été réalisés, où seulement un ligand a été vaporisé à la fois. Les ligands ont été vaporisés à des concentrations équivalentes à 100 fois leur constante de dissociation (K_D) connue (Figure 4.4 D). Il a été observé que les décalages correspondaient uniquement aux constructions correspondant au ligand vaporisé (entouré dans la figure 4.4 D). Les ARN qui n'ont pas interagi avec le ligand retournent au point de départ que l'on appelle la ligne de base.

Quand le FMN a été vaporisé, des changements ont été observés pour la construction FMN_3 et pour la construction FMN_5 avec des amplitudes de décalages différentes (Figure 4.4 D). Concernant le di-GMPc, des petits changements de migration ont été remarqués pour les constructions di-GMPc_3, di_GMPc_4 et di_GMPc_5 lorsque le ligand a été vaporisé (Figure 4.4 D). Pour les aptamères glycine et TPP, des changements de migration ont également été observés en présence de leur ligand respectif (Figure 4.4 D). Cependant, on peut remarquer qu'aucun décalage n'a pu être observé pour les différentes constructions des *riboswitchs* fluor et nickel / cobalt (Figure 4.4 D). Afin de valider notre méthode dans le contexte qui nous intéresse, à savoir un SELEX pour des multiples ligands et séquences en même temps, une solution contenant tous les ligands a été vaporisée sur le gel. Lors de cet essai, il a été remarqué que les changements de migration étaient identiques à ceux obtenus pour les ligands seuls (Figure 4.4 D). Cette dernière étape a permis de valider que le SR-PAGE pourrait être utilisé pour la recherche de multiples *riboswitchs* en même temps.

4.3.2 Un outil d'étude pour la structure et la plateforme d'expression

Nous avons également émis l'hypothèse que le SR-PAGE pourrait être une méthode permettant d'étudier la partie plateforme d'expression des *riboswitchs*. En effet, il peut être utilisé pour sélectionner une construction montrant un plus grand changement de migration dans le gel, indiquant ainsi un plus grand changement de conformation de cet ARN. Les constructions FMN_3 et FMN_5 ont été analysées plus profondément pour comprendre les différences de décalages qui peuvent être observées pour un même *riboswitch*. Théoriquement, toutes les constructions devraient avoir la même capacité à lier leur ligand. Le service Mfold (Zuker, 2003) a été utilisé pour analyser les énergies libres de toutes les constructions dans leur structure liée et non liée. La structure liée a été réalisée en forçant l'aptamère à adopter sa structure en utilisant des contraintes (Tableau 4.4), en accord avec ce qui a été publié précédemment (Serganov *et al.*, 2009). La constante d'équilibre (K_{eq}) a été mesurée en calculant la différence d'énergie libre entre les deux conformations.

Pour la construction FMN_3, l'ARN est légèrement plus stable dans sa structure non liée avec une énergie libre de -68.9 kcal/mol comparativement à sa forme liée (avec des contraintes) qui a une énergie libre de -68 kcal/mol (Figure 4.5 A). Notre interprétation est que durant la première migration la construction FMN_3 est présente dans sa conformation la plus stable, non liée, (Figure 4.5 A, partie gauche) pouvant potentiellement se convertir en structure contenant le domaine aptamère (Figure 4.5 A, partie droite). La constante d'équilibre ($K_{eq} = 0,19$) nous indique que la forme non liée serait présente à un ratio de 5 :1 comparée à la forme liée. Cependant, quand le ligand (FMN) est vaporisé sur le gel, il n'y a pas seulement un changement au niveau de l'équilibre, mais la conformation aptamère (capable d'interagir avec le ligand) est aussi capable de changer de structure tridimensionnelle. Le changement de structure entre ces deux conformations se traduit par un changement de migration qui peut se voir sur le gel de SR-PAGE (Figure 4.5 A). Comme l'énergie libre des deux conformations est proche, la conformation non liée devrait être favorisée, mais la présence du ligand est suffisante pour permettre le changement vers la structure liée. La même situation semble s'appliquer pour la construction FMN_5, avec une énergie libre de -50.70 kcal/mol et -48.70 kcal/mol pour les conformations liée et non liée respectivement (Figure 4.5 B). Cependant, les deux structures semblent vraiment similaires (Figure 4.5 B), ce qui peut expliquer un décalage beaucoup moins important. Des études précédentes ont également montré que l'aptamère du *riboswitch* FMN est préformé en absence de son ligand avec des concentrations physiologiques de magnésium (Vicens *et al.*, 2011). Dans le cas de FMN_5, malgré un K_{eq} d'une valeur de 0,026 suggérant un ratio d'environ 40:1 pour la conformation non liée comparée à la conformation liée, la proximité des deux structures peut probablement permettre un changement plus facile à l'équilibre, mais aussi un décalage plus complet comme suggéré par la quasi absence de bande d'ARN non décalé pour cette construction.

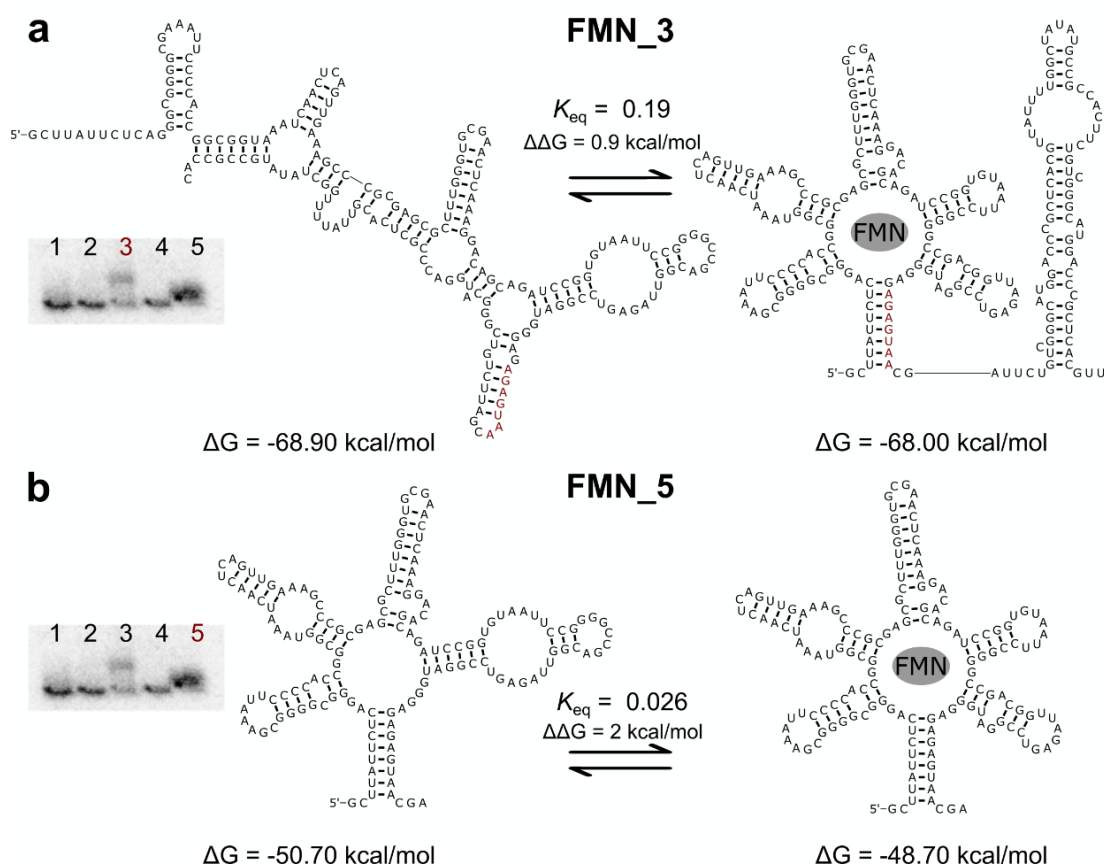


Figure 4.5 Les constructions FMN_3 et FMN_5 présentent une énergie libre similaire entre leurs structures liées et non liées.

Représentation des structures prédites, des énergies libres (ΔG) et de la constante d'équilibre (K_{eq}) pour l'ARN non lié et l'ARN lié pour les constructions FMN_3 (A) et FMN_5 (B). Les nucléotides représentés en rouge peuvent se lier afin de former la tige basale de l'aptamère FMN, permettant de former une structure mutuellement exclusive. Les décalages obtenus par le SR-PAGE sont représentés sur la gauche, avec le numéro en rouge correspondant à la structure observée. Tiré de (Devinck *et al.*, 2022).

Les constructions FMN_3 et FMN_5 sont plus stables dans leurs structures non liées et la constante d'équilibre favorise dans tous les cas cette conformation en absence de ligand. En effet, le K_{eq} présente une valeur inférieure à 1. Une analyse similaire des énergies libres et des constantes d'équilibre a été réalisée pour toutes les constructions des *riboswitchs* FMN, fluor, di-GMPc et nickel-cobalt (Annexe 12 à 15). La plus grande différence dans les énergies libre est observée pour toutes les constructions fluor et nickel-cobalt ce qui pourrait expliquer qu'aucune des constructions testées ne montrent de décalage dans la migration (Annexe 12 et 15 respectivement). De plus, la constante d'équilibre pour la plupart de ces cas présente des valeurs très faibles, suggérant que la structure liée est probablement presque absente. Pour les

constructions di-GMP_c_3-4-5, et FMN_5, la plupart des énergies libres sont très similaires avec un K_{eq} proche de 1, ce qui explique que toutes les constructions peuvent être décalées, mais les décalages sont plus subtils compte tenu des similarités entre les états liés et non liés prédits.

4.3.3 Sélection de *riboswitchs* avec une affinité modifiée

Le SR-PAGE est une méthode qui a été conçue pour être utilisée comme moyen de sélection au sein d'un SELEX. Une preuve de concept a été réalisée avec le *riboswitch* TPP où une partie de celui-ci a été modifié aléatoirement. Pour cela, trois librairies ont été créées contenant 6, 10 et 18 nucléotides aléatoires, chacune avec 4 096, 1 048 576 et 68 719 476 736 séquences possibles, respectivement (Figure 4.6 A). Les librairies d'ARN ont été séparées sur un gel de polyacrylamide natif via les deux migrations de SR-PAGE. Des milliers de séquences sont alors testées en même temps avec des structures différentes dans un même puits, une seconde migration en sens inverse est nécessaire afin de pouvoir comparer les changements de structure des ARN. En effet, avec une seule migration avec l'ajout du ligand, comme on peut le voir dans des essais de retard sur gel, il serait difficile de discerner les différences de migration d'une séquence par rapport aux autres. Avec le SR-PAGE, seuls les ARN ayant interagi avec un ligand changeront de conformation et la deuxième migration permet de les discerner des autres qui reviennent au point de départ. Pour sélectionner les ARN liant mieux la thiamine que le TPP, celui-ci est présent lors des deux migrations. Contrairement à la thiamine qui a été vaporisée sur le gel entre les deux migrations. Tout d'abord vaporisée à une concentration de 2 mM pour les générations 1 à 4, puis pour les cycles suivants la concentration de thiamine a été réduite à 20 μ M puis 200 nM afin d'augmenter la stringence de la sélection. Les ARN ayant montrés une différence de migration ont été élués du gel et amplifiés afin de réaliser un nouveau cycle de sélection. Après quatre, cinq, sept et dix cycles de sélection les séquences sélectionnées ont été clonées, mais aussi analysées via une méthode de séquençage Illumina (Bentley *et al.*, 2008) afin de déterminer les séquences se trouvant enrichies au fur et à mesure des cycles. Avec le SR-PAGE, les ARN ont été sélectionnés selon leur capacité de décalage traduisant un changement de conformation (Figure 4.6 B).

Les ARN ont été caractérisés en utilisant une méthode de cartographie *in vitro* par *in-line* afin de déterminer leur affinité avec chacun des ligands (TPP et thiamine) et leur constante de dissociation (K_D) a été calculée. Ainsi, un K_D plus petit correspond à une affinité plus grande pour son ligand. Le *riboswitch* TPP WT présente un K_D pour les molécules de TPP de moins de 10 nM et pour la thiamine de $4 \mu\text{M} \pm 1,9 \mu\text{M}$ dans nos conditions de laboratoire (Figure 4.6 F). L'un des

meilleurs clones qui a été obtenu, le clone 13, montre une légère amélioration de l'affinité pour la thiamine avec un K_D de $3 \mu\text{M} \pm 1,3 \mu\text{M}$, avec une forte diminution de l'affinité pour le TPP avec un K_D moyen de $7,8 \mu\text{M} \pm 3,5 \mu\text{M}$ (Figure 4.6 C, 4.6 D et 4.6 E). Ces valeurs de K_D représentent une moyenne des K_D obtenus suite à la réalisation de plusieurs essais de cartographie *in vitro* par *in-line* pour ce clone. D'autres clones (1, 2, 9, 13, 23, 24, 25, 41, 42, 43, 44, 45, 46, et 47) montrent également une meilleure affinité pour la thiamine que pour le TPP (Figure 4.6 F). Seulement les clones pour lesquels des valeurs de K_D ont été obtenues sont représentés. Toutes les séquences des clones sélectionnés, ainsi que leurs K_D respectifs pour chacun des ligands sont listés dans l'Annexe 16.

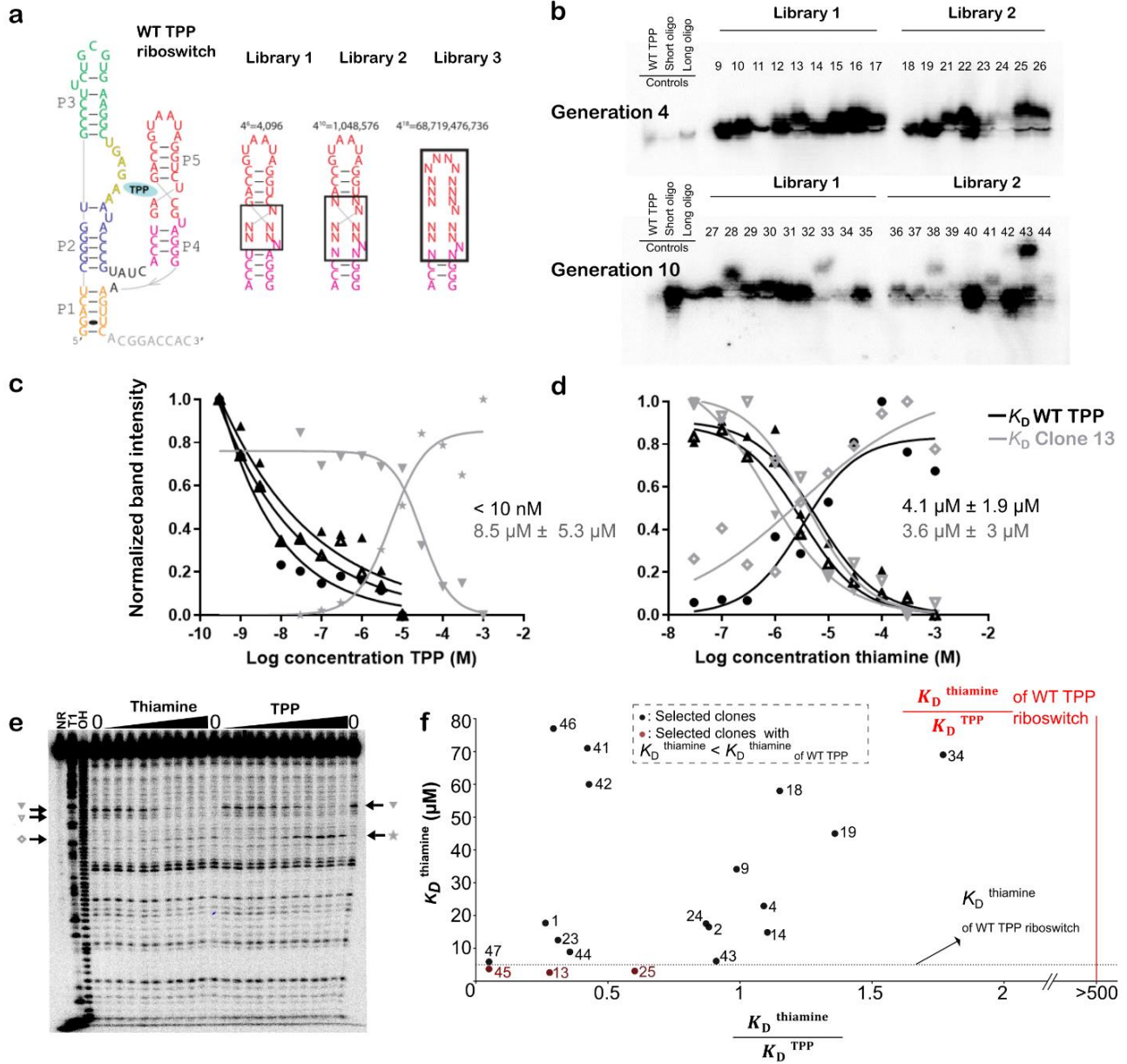


Figure 4.6 Sélection d'un aptamère thiamine à partir du riboswitch TPP par le SR-PAGE.

A : Des bibliothèques dégénérées du *riboswitch* TPP ont été réalisées. Les bibliothèques 1, 2 et 3 présentent respectivement 6, 10 et 18 nucléotides aléatoires dans la tige P5. **B** : Capacité de décalage des différents clones sélectionnés lors du SR-PAGE pour les bibliothèques 1 et 2 pendant les générations 4 et 10. Le TPP était présent dans les deux migrations à une concentration de $10 \mu\text{M}$ et la thiamine était présente lors de la seconde migration uniquement à une concentration comprise entre 2 mM et 200 nM en fonction du cycle. Les ARN ont été marqués radioactivement via une méthode de transcription chaude. Comme contrôle de migration, des oligonucléotides court et long, ainsi que le *riboswitch* WT TPP ont également été déposés sur gel. **C** : Courbe de K_D d'un essai de cartographie *in vitro* par *in-line* pour la molécule de TPP pour le *riboswitch* WT TPP (en noir) et le clone 13 (en gris). Le clone 13 est issu de la bibliothèque 1, sélectionné lors de la génération 4 de SELEX. **D** : Courbe de K_D d'un des essais de cartographie *in vitro* par *in-line* pour la molécule de thiamine pour le *riboswitch* WT TPP (en noir) et le clone 13 (en gris). **E** : Cartographie *in vitro* par *in-line* pour le clone 13. Les flèches représentent les bandes qui ont été utilisées pour le calcul du K_D représenté en C et D. Des concentrations de TPP et de thiamine de 30 nM à 1 mM ont été utilisées. **F** : Représentation de tous les clones sélectionnés montrant une affinité pour le TPP et/ou la thiamine. Chaque point représente un clone, les clones en rouge

représentent ceux qui ont une affinité pour la thiamine meilleure que le *riboswitch* WT. La ligne horizontale en pointillée représente le K_D de la thiamine pour le riboswitch WT. La ligne verticale rouge le ratio des K_D de la thiamine sur le K_D du TPP pour le *riboswitch* WT. Tiré de (Devinck *et al.*, 2022).

Dans la librairie 3, la tige P5 du *riboswitch* WT a été remplacée par 18 nucléotides aléatoires. Par conséquent, seulement une petite partie de la librairie devrait former une tige similaire au *riboswitch* WT et aucune ne devrait contenir une boucle interne car nous ne sélectionnons pas pour la liaison du TPP (Figure 4.6 A). Après les générations quatre et sept, les librairies ont été analysées par séquençage Illumina. Parmi les séquences sélectionnées suite au séquençage, les clones 45 et 47 n'ont pas montrés de liaison avec la molécule de TPP jusqu'à une concentration de 1 mM. De plus, malgré le fait que nous avons entièrement dégénéré la tige P4-P5, nous avons vraisemblablement sélectionné des ARN qui forment une tige à ce niveau-là (Annexe 17). Une prédiction des structures secondaires pour les séquences enrichies, comportant plus de quatre lectures, a été réalisée et nous a montré que nous avons sélectionné la formation d'une tige en P5 (Annexe 18). La liaison de la thiamine par les différents clones n'a pas été grandement améliorée, ainsi cette tige a pu être sélectionnée car elle a facilité le passage entre les états lié et non lié suite à la présence de la thiamine et/ou a amplifié la différence entre les versions liées et non liées de la conformation de l'ARN. En effet, il est connu que les angles entre les tiges ont un impact sur la façon dont l'ARN migre dans un gel de polyacrylamide natif (Lafontaine *et al.*, 2001).

4.3.4 Influence des adaptateurs sur le décalage des ARN

Le SR-PAGE est une méthode développée afin de pouvoir sélectionner des *riboswitches* dans une librairie d'ARN à travers un SELEX. Or, lors de ce type de méthode il est nécessaire d'ajouter de part et d'autre des ARN une séquence adaptatrice permettant d'amplifier les séquences par PCR. Nous nous sommes ainsi demandés si les adaptateurs pourraient être un frein pour notre méthode en limitant le décalage obtenu. Des contrôles ont été effectués où des séquences adaptatrices ont été ajoutées aux constructions de *riboswitches* ayant montré un décalage de migration précédemment (Figure 4.4) à savoir : FMN_3, Glycine, TPP, di-GMPc_4 et _5. Pour cela, différents oligonucléotides complémentaires aux séquences adaptatrices 5' ont été créés (Figure 4.7 A et Tableau 4.5) et mis en contact pendant 30 min avec les ARN d'intérêt avant d'être migrés sur le gel.

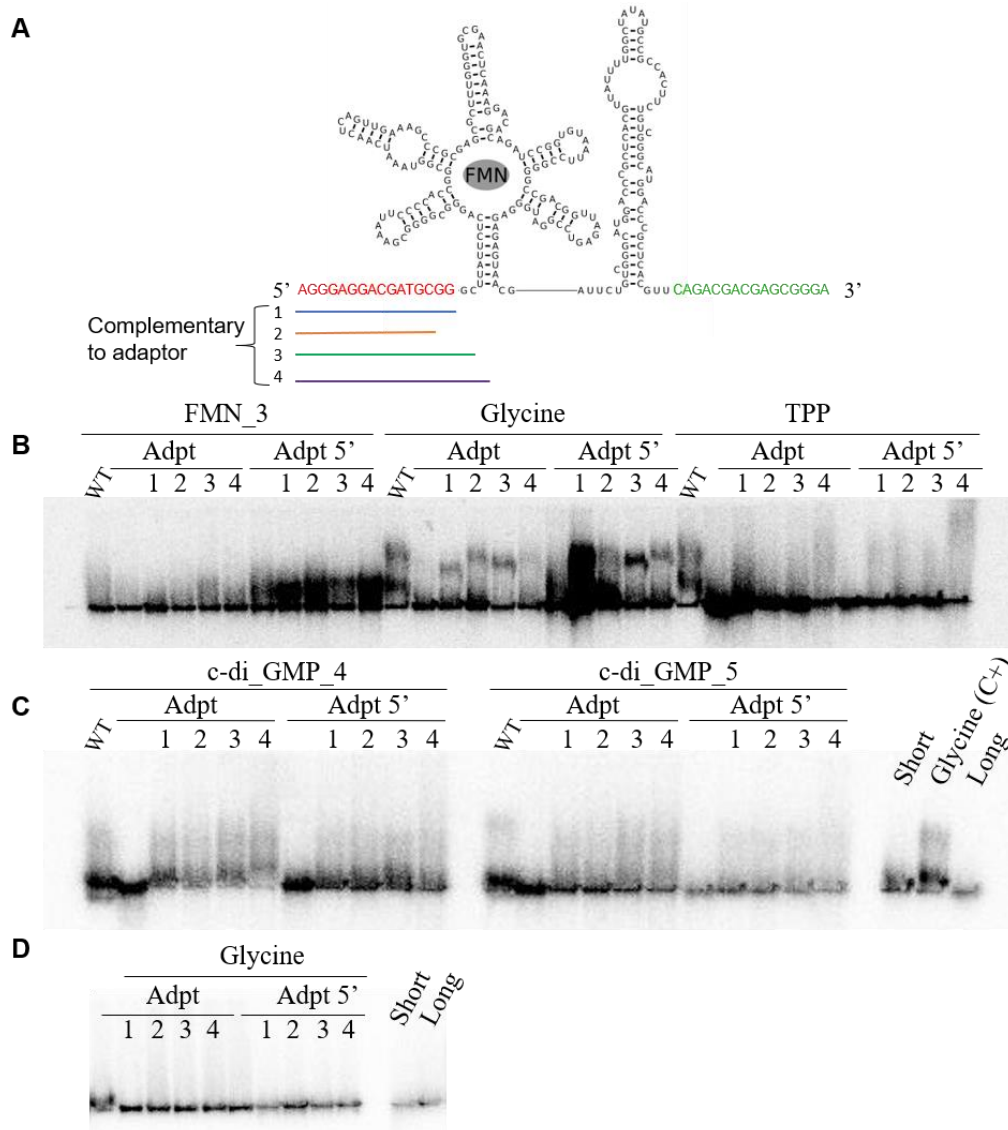


Figure 4.7 Les oligonucléotides complémentaires aux adaptateurs permettent de rétablir le décalage des *riboswitchs* lors du SR-PAGE.

A : Représentation schématique du design des oligonucléotides complémentaires avec l'exemple du *riboswitch* FMN. Les séquences adaptatrices utilisées pour l'amplification PCR sont représentées en rouge et en vert et sont identiques pour tous les *riboswitchs* testés. Les traits colorés représentent les différents oligonucléotides testés : le « 1 » est entièrement complémentaire à l'adaptateur, le « 2 » est complémentaire à l'adaptateur mais avec deux nucléotides en moins en 3', le « 3 » est complémentaire à l'adaptateur mais avec deux nucléotides supplémentaires spécifique à la séquence du *riboswitch*, enfin le « 4 » est entièrement complémentaire à l'adaptateur avec quatre nucléotides supplémentaires dans la séquence du *riboswitch*. **B** : Résultats des migrations du SR-PAGE pour les constructions FMN_3, Glycine, TPP en vaporisant sur le gel le ligand correspondant. « Adpt » correspond à la présence sur l'ARN des séquences adaptatrice en 5' et en 3'. « Adpt 5' » correspond à la présence sur l'ARN de la séquence adaptatrice uniquement en 5'. « WT » est la construction sans les séquences adaptatrices. Les puits sans numéro correspondent à l'absence d'oligonucléotides complémentaires. Les puits 1, 2, 3 et 4 correspondent à l'ARN avec l'oligonucléotides complémentaire correspondant. **C** : Résultats des migrations du SR-PAGE pour les constructions c-di-GMP_4 et c-di-GMP_5 en vaporisant sur le gel le ligand correspondant. « Adpt » correspond à la présence sur l'ARN des séquences adaptatrice en 5' et en 3'. « Adpt 5' » correspond à la présence sur l'ARN de la séquence adaptatrice uniquement en

5'. « WT » est la construction sans les séquences adaptatrices. Les puits sans numéro correspondent à l'absence d'oligonucléotides complémentaires. Les puits 1, 2, 3 et 4 correspondent à l'ARN avec l'oligonucléotides complémentaire correspondant. « Short », « Long » représentent les deux contrôles de migration, qui sont des oligonucléotides radioactifs court et long servant de contrôle de migration. Le *riboswitch* glycine est utilisé comme contrôle positif de migration. **D** : Résultats du SR-PAGE avec le *riboswitch* glycine et les oligonucléotides complémentaires aux adaptateurs sans vaporiser de ligand. « Short », « Long » représentent les deux contrôles de migration, qui sont des oligonucléotides radioactifs court et long servant de contrôle de migration. Tiré de (Devinck *et al.*, 2022).

Un décalage peut ainsi être observé pour toutes les constructions sans adaptateur (Figure 4.7 B et C). On peut également remarquer qu'aucun décalage n'est observé pour toutes les constructions avec les adaptateurs sans les oligonucléotides complémentaires, ceci représente ainsi une limite de notre méthode car même pour les *riboswitchs* connus, le décalage est perdu en présence d'adaptateurs de SELEX (Figure 4.7 B et C). Cependant, on peut distinguer qu'en présence des quatre oligonucléotides complémentaires le décalage a été restauré (Figure 4.7 B et C). Ainsi, lors d'un SELEX avec le SR-PAGE il est important de mettre des oligonucléotides complémentaires aux adaptateurs afin de s'assurer que ceux-ci n'empêchent pas le bon repliement de l'ARN.

Un contrôle a été effectué afin de s'assurer que la restauration du décalage n'était pas uniquement dû à la séparation de l'oligonucléotides de l'ARN au cours des 48h de migration. Pour cela un contrôle a été effectué avec le *riboswitch* glycine avec et sans les oligonucléotides complémentaires et sans vaporisation de glycine lors de la seconde migration. On peut alors remarquer qu'aucun décalage n'est observé (Figure 4.7 D) indiquant donc que le décalage observé est bien dû à un changement de conformation de l'ARN et pas uniquement au détachement des oligonucléotides complémentaires.

4.4 Discussion

La méthode de Shifted-Reverse PAGE (SR-PAGE) est une nouvelle méthode pour étudier des *riboswitchs* et des aptamères en outrepassant les limites des outils de bioinformatique, ainsi que celles des autres méthodes expérimentales. Cette technique a été validée avec différents *riboswitchs* connus avec des ligands de nature différentes, di-GMPc classe 1, FMN, TPP, glycine. Ces essais ont été réalisés avec des ligands individuels ou un mélange de ligands. Il a aussi été démontré que le SR-PAGE peut être utilisé au sein d'un SELEX comme méthode de sélection. En effet, il a été possible de sélectionner des *riboswitchs* avec une affinité modifiée pour son ligand en modifiant de façon aléatoire la tige P5 du *riboswitch* TPP, pour diminuer son affinité naturelle pour le TPP et augmenter son affinité pour la thiamine.

Un décalage dans la migration n'a pas été détecté pour tous les *riboswitchs* connus, comme observé pour les *riboswitchs* fluor et nickel-cobalt (Figure 4.4 D). Ceci peut être expliqué par les constructions testées. En effet, comme démontré avec les constructions FMN_3 et FMN_5 (Figure 4.5), seulement quelques nucléotides peuvent faire la différence entre un ARN montrant un décalage ou non. Ce phénomène est probablement dû à l'équilibre subtil des structures entre les conditions avec ou sans ligands. Ainsi, une légère différence dans la séquence peut résulter dans un grand changement en terme de structure secondaire. Même si toutes les constructions contiennent techniquement le domaine aptamère, cela ne signifie pas forcément que la structure de l'aptamère se formera. Nous avons aussi émis l'hypothèse qu'un décalage n'était pas observé pour tous les *riboswitchs* connus car les séquences d'ARN résultant en un plus grand changement de structure secondaire suite à la liaison du ligand n'ont pas été testées, ou parce que la différence d'énergie libre entre la structure liée et non liée était trop importante pour être surmontée même en présence du ligand. Pour les constructions montrant un décalage (FMN_3-5, c-di-GMP_3-4-5; Figure 4.4 D), la différence d'énergie libre prédite a été d'une moyenne de 1,78 kcal/mol, alors qu'une moyenne de 12,1 kcal/mol a été observée pour celles qui n'en n'ont montré (fluor_1-2-3-4-5, FMN_1-2-4, c-di-GMP_1-2, NiCo_1-2-3-4-5; Annexe 9 à 12). Dans les derniers cas, la différence d'énergie libre entre les structures liée et non liée est trop importante pour être surmontée par le ligand. Les valeurs de constante d'équilibre quand aucun décalage n'est observé est aussi petite (généralement inférieur à 10^{-3}), suggérant que la structure secondaire ne pourrait pas adopter la structure liée et/ou que l'énergie de la liaison du ligand est insuffisante pour outrepasser les différences de ΔG . Un décalage a été observé quand la constante d'équilibre était proche de 1, suggérant que la structure est capable de passer au-dessus de la différence d'énergie libre ($\Delta\Delta G$) pour être capable de lier le ligand, probablement lorsque l'interaction avec le ligand a un $\Delta\Delta G$ plus important que le $\Delta\Delta G$ des formes liées vs non liées. Il s'agit du modèle d'ajustement induit (Heppell *et al.*, 2011), quand l'ajout du ligand favorise un changement de structure secondaire d'une molécule d'ARN donnée. D'un autre côté, l'addition du ligand pourrait favoriser les ARN qui sont déjà dans la bonne structure secondaire, qui est le modèle de sélection conformationnelle (Haller *et al.*, 2011). Dans ce cas, la dynamique de changement de structure dans le gel pourrait être bloquée dans la conformation aptamère lors de l'ajout du ligand ce qui expliquerait le changement de migration. C'est probablement la combinaison de ces deux modèles qui se passe dans plusieurs cas, y compris la présence de changements plus subtils au niveau de la structure tertiaire qui pourraient conduire à des décalages détectables sur le gel.

L'amplitude de décalage semble également corrélée avec le niveau de changement de conformation observé entre les structures liées et non liées. En effet, quand les deux structures secondaires sont similaires il en résulte un plus petit changement de migration. Ce fait est observable notamment au niveau de la construction FMN_5, où le décalage est moins prononcé que celui observé pour la construction FMN_3. En effet, on a pu remarquer que la construction FMN_5 se trouvait déjà dans une structure secondaire similaire à l'aptamère dans sa conformation non liée (Figure 4.5), ceci est aussi observable pour les constructions du *riboswitch* di-GMPc (Annexe 14). Par conséquent, une des limitations de notre méthode est qu'il est possible de manquer des *riboswitchs* qui sont capables de lier un ligand, mais que le changement de conformation ne soit pas assez important pour être capable de détecter un décalage dans la migration. Cependant, cette limitation peut être utilisée à notre avantage, puisque le SR-PAGE pourrait être utilisé pour identifier les constructions produisant les plus grands décalages et ainsi gagner des informations sur la plateforme d'expression. Ceci permettrait d'augmenter nos connaissances sur les *riboswitchs* et les aptamères connus, comme suggéré avec l'analyse obtenue avec le *riboswitch* FMN (Figure 4.5).

Nous avons aussi utilisé le SR-PAGE au sein d'une méthode de SELEX afin de sélectionner des *riboswitchs* ayant une affinité modifiée pour leur ligand. Ceci peut s'apparenter à une étude effectuée par une autre équipe ayant utilisé le Capture SELEX afin de sélectionner des *riboswitchs* avec une affinité modifiée (Mohsen *et al.*, 2023). Le SR-PAGE a permis de sélectionner des molécules d'ARN en se basant sur deux propriétés : premièrement, leur capacité à lier préférentiellement la thiamine que le TPP, et deuxièmement le changement de conformation de l'ARN en présence de cette molécule produisant un décalage lors de la migration. Plus de 40 séquences isolées à partir de ce SELEX ont été étudiées séparément par une technique de cartographie *in vitro* par *in-line*. De toutes ces séquences, 14 montrent une meilleure affinité pour les molécules de thiamine que le TPP avec un ratio K_D thiamine / K_D TPP inférieur à 1, comparé à un ratio supérieur à 500 pour le *riboswitch* WT (Figure 4.6 F). De plus, trois clones (13, 25 et 45) montrent une affinité pour la thiamine potentiellement meilleure que celle du *riboswitch* WT (Figure 4.6 F). Cette expérience a permis de démontrer que le SR-PAGE pourrait être utilisé comme moyen de sélection au sein d'un SELEX pour sélectionner des aptamères avec une affinité modifiée et/ou améliorée pour un ligand, montrant par le fait même le potentiel de sélectionner des *riboswitchs* encore inconnus à partir de bibliothèques issues de génomes.

5 DISCUSSION GÉNÉRALE ET PERSPECTIVES

5.1 Caractérisation d'ARN non codants chez *Pseudomonas* et *Burkholderia*

Nous avons développé au sein du laboratoire un vecteur permettant d'étudier des promoteurs suite à la présence d'un *riboswitch* fluor entre celui-ci et l'opéron *luxCDABE*. Il s'agit du plasmide mini-CTX-lux qui a été modifié pour permettre l'insertion de promoteur (Becher & Schweizer, 2000; Korniakova *et al.*, 2021). Ce plasmide a la caractéristique d'être présent à un grand nombre de copies chez *E. coli*, mais il se retrouve intégré dans le chromosome chez les bactéries du genre *Burkholderia* et *Pseudomonas* (Becher & Schweizer, 2000), ce qui permet une évaluation plus proche des conditions physiologiques. La présence d'un *riboswitch* entre le promoteur et l'opéron permet de contrôler l'expression du promoteur et ainsi limiter la toxicité engendrée par la luminescence (Hollis *et al.*, 2001). Grâce à cet outil, nous avons été capables de cloner en amont de deux *riboswitchs* fluor, le promoteur constitutif du gène de la protéine ribosomale S7 qui est un promoteur très fort (Korniakova *et al.*, 2021).

La composition et la disposition de ce plasmide permet de l'utiliser pour un autre objectif. En effet, il est possible d'enlever le *riboswitch* fluor pour y insérer un ARNnc candidat qui sera ainsi placé en fusion traductionnelle avec le gène *luxC* du plasmide permettant d'étudier l'expression de la luminescence dans différentes conditions. Une preuve de concept a été réalisée avec le *riboswitch* fluor présent initialement dans le plasmide, qui est présent chez *B. thailandensis*. La bactérie a été cultivée avec différentes concentrations de fluor (0 mM à 62 mM) et l'expression de la luminescence a été évaluée au cours du temps. Ceci nous a permis de valider que l'expression de la luminescence dans la bactérie est bien dépendante du ligand ainsi que de la concentration de ce dernier (Figure 3.3).

Des études précédentes ont permis de mettre en évidence plus de 300 nouveaux ARNnc candidats (Weinberg *et al.*, 2007; Weinberg *et al.*, 2017a; Weinberg *et al.*, 2010). Cependant, parmi ceux-ci seulement quelques-uns ont été caractérisés comme étant des *riboswitchs* (Nelson *et al.*, 2017; Sudarsan *et al.*, 2008). Une vingtaine de candidats ont été sélectionnés pour notre étude présents chez les bactéries du genre *Pseudomonas* et *Burkholderia* (Tableau 5.1), du fait que notre plasmide soit fonctionnel spécifiquement pour ces bactéries, mais également car ces bactéries présentent de nombreux ARNnc seulement annotés, sans aucune validation expérimentale, car les outils génétiques disponibles chez ces bactéries sont plus rares que pour *E. coli* et *B. subtilis* par exemple. En effet, nous avons montré qu'environ 45% des ARNnc annotés chez *Pseudomonas* et *Burkholderia* restent non caractérisés (Tableau 3.2). Le *riboswitch* glycine

a été l'un des premiers *riboswitchs* caractérisés suite à l'étude réalisée par Barrick et ses collaborateurs (Barrick *et al.*, 2004). Celui-ci est présent en amont du gène *gcvT* impliqué dans le système de clivage de la glycine. Ainsi, le test de la glycine comme ligand potentiel était relativement évident et s'est avéré exact (Mandal *et al.*, 2004). À l'inverse, l'une des difficultés pour la caractérisation de certains nouveaux ARNnc candidats vient du fait que ces motifs sont présents en amont de gènes n'ayant pas de fonction dans une voie métabolique précise. Comme le motif *fixA* situé en amont des gènes *fixA* et *fixB* impliqués dans la respiration (Tsai & Saier Jr, 1995). Ce motif est ainsi relié à tout le métabolisme énergétique de la bactérie il est donc difficile de cibler une ou quelques molécules pouvant agir au niveau de sa régulation. Mais également le motif *rmf* qui est connecté à la phase stationnaire où l'expression de nombreux gènes est affectée rendant également difficile l'identification du signal permettant sa régulation (Waite *et al.*, 2005). Enfin, pour les motifs *Pseudomon-Rho* et *rne-II*, leur contexte génétique peut amener à penser que ce ne soit pas des *riboswitchs* liant un ligand précis mais plutôt des boucles de rétrocontrôle de la protéine en elle-même.

Tableau 5.1 Liste de tous les candidats initialement prévus pour la caractérisation des ARNnc chez *Pseudomonas* et *Burkholderia*.

<i>Burkholderia</i>		<i>Pseudomonas</i>	
Burkholderiales-I	<i>traJ-II</i>	<i>aceE</i>	<i>Pseudomon-GroES</i>
<i>gabT</i>		<i>fixA</i>	<i>Pseudomon-Rho</i>
MAEB		<i>gyrA</i>	<i>rmf</i>
<i>sbcD</i>		<i>icd</i>	<i>rnell</i>
<i>sucA</i>		Pseudomon-I	<i>sucC</i>
<i>radC</i>		<i>Pseudomon-GGDEF</i>	<i>sucA-II</i>

Tableau répertoriant tous les ARNnc candidats (issus des études (Weinberg *et al.*, 2007; Weinberg *et al.*, 2010)) prévus dans ce projet de doctorat pour être testés avec le plasmide pVK-f2-lux (Korniakova *et al.*, 2021). Les candidats en rouge sont ceux que nous ne sommes pas parvenus à cloner, le clone en orange correspond à un candidat que nous avons réussi à cloner mais qui a été caractérisé comme un petit ARN régulateur, ErsA (Ferrara *et al.*, 2021). Les candidats en vert sont ceux que nous avons réussi à cloner et caractériser.

Initialement, les ARNnc candidats devaient être clonés dans le plasmide pVK-f2-lux contenant une copie du *riboswitch* fluor de *B. thailandensis* (Korniakova et al., 2021). Cependant, ces clonages se sont révélés infructueux malgré de nombreux essais. En effet, malgré la présence des sites de restrictions correspondants et la modification de plusieurs paramètres dans la réaction (utilisation de kit de midiprep, utilisation de midiprep/miniprep nouvellement réalisées, utilisation d'enzymes de restrictions neuves), il était impossible de digérer le plasmide. Ce n'est qu'en réalisant le clonage dans un plasmide qui avait deux *riboswitchs* fluor, pVK-f2.2-lux, que les clonages furent fructueux. Quelques changements furent alors apportés au clonage, notamment pour retirer les deux *riboswitchs* fluor avant d'y insérer l'élément régulateur à l'étude (Figure 3.5). Malgré ce nouveau plasmide, nous avons de nouveau éprouvé des difficultés au niveau du clonage. En effet, nous avons pu identifier que les différentes étapes nécessaires pour la digestion du plasmide pouvaient influencer sur la réussite du clonage (qualité de la midiprep, purification sur gel). Une autre limite à ces clonages pouvait également venir de la taille du plasmide (13 kb). Il a été montré que des plasmides de grande taille avaient un rendement de transformation plus faible (Chan et al., 2002). De plus, afin de construire notre plasmide, il est nécessaire de le transformer en premier lieu dans *E. coli*. Dans cet organisme le plasmide mini-CTX-lux est répliquatif donc présent en grand nombre de copies. Il est possible que la luminescence soit donc exprimée fortement dépendamment du motif et ainsi soit toxique pour la bactérie. Or, il a été démontré qu'une forte expression de *luxCDABE* pouvait être toxique pour la cellule (Hollis et al., 2001; Korniakova et al., 2021). Compte tenu de ces difficultés de clonage rencontrées, il n'a pas été possible d'obtenir tous les clones pour les ARNnc candidats d'intérêts (Tableau 5.1). Nous sommes parvenus à obtenir sept des dix ARNnc candidats présents chez *Pseudomonas*, mais nous ne sommes pas parvenus à cloner les motifs candidats chez *Burkholderia*. Ceci constitue l'une des perspectives à ce projet.

Plus de 40 conditions ont été testées pour chacun des candidats ARNnc de notre étude (Tableau 3.4). Cependant, nous avons remarqué que certains milieux ne permettaient pas la croissance de la bactérie. En effet, on a pu voir pour tous les motifs que dans les milieux M9 pH 5, M9 pH 6, M9 supplémenté avec 0,5 mM et 1 M NaCl et M63 nous ne sommes pas parvenus à faire croître *P. aeruginosa*. Pour le milieu M9 cela peut s'expliquer par le fait que normalement le pH de ce milieu est de 7,16. Or ce milieu plus sensible au changement de pH, ce qui aurait donc pu affecter la croissance de notre organisme d'étude. Des essais ont été réalisés notamment avec le motif *sucA-II* avec des temps d'incubation plus longs (jusqu'à 30h) et la croissance était encore très faible (DO 0,03). Concernant le milieu supplémenté en NaCl, on a pu observer une bonne croissance dans le milieu avec 0,1 M NaCl, cette concentration avait été testée dans une

autre étude comme étant un milieu avec une hyper osmolarité (Scharfman *et al.*, 1996). Or pour la concentration 0,5 M on a remarqué une croissance beaucoup plus faible et une absence de croissance dans le milieu 1M NaCl. La concentration 0,1 M étant déjà un milieu avec une hyper osmolarité (Scharfman *et al.*, 1996), ces concentrations doivent être trop élevées pour permettre la croissance de *P. aeruginosa* PA14. Cependant, une étude a montré une croissance de *P. aeruginosa* dans un milieu contenant 0,5 M NaCl, mais la différence avec notre expérience est que le NaCl est ajouté une fois que la bactérie a atteint la phase exponentielle de croissance (DO 0,3) (Bazire *et al.*, 2005). D'autres essais pourraient être effectués de cette manière afin de voir l'influence d'un stress osmotique sur *P. aeruginosa* et nos ARNnc candidats. Enfin, nous avons également pu observer une absence de croissance dans le milieu M63, ce milieu chimiquement défini présente une osmolarité faible. Or, il a été montré dans plusieurs études que *P. aeruginosa* est capable de pousser dans ce type de milieu, mais elle montre une croissance plus lente (Brandenburg *et al.*, 2013). Ainsi, il serait possible de réitérer ces essais en faisant des lectures à des temps plus longs (48h, 72h). Les conditions testées n'ajoutent ou n'enlèvent pas seulement un composé, la présence ou l'absence de ce dernier peut causer de grands changements dans le métabolisme de la bactérie qui peut ainsi changer de voie de signalisation.

5.1.1 Le motif *rmf*

Le motif *rmf* présent en amont du gène codant pour le *ribosome modulation factor* est l'un de nos candidats les plus prometteurs. En effet, nous avons pu montrer une grande différence du niveau de luminescence entre les constructions WT et disruptive aussi bien dans des milieux riches que dans des milieux chimiquement définis (Figure 3.8). Le mutant compensatoire de ce motif permet une restauration du niveau de luminescence dans ces milieux (Figure 3.9 et Annexe 2).

Une première hypothèse était qu'il s'agisse d'un *riboswitch* ppGpp car des études ont montré que la transcription de *Rmf* était activée en présence de ppGpp chez *E. coli* (Izutsu *et al.*, 2001). Néanmoins, nos résultats ne semblent pas concorder avec une régulation directe de ppGpp sur cet ARN. Effectivement, les trois constructions (WT, *disr* et *comp*) ont été conjuguées dans une souche $\Delta relA \Delta spoT$ ne produisant donc pas de ppGpp (Nguyen *et al.*, 2011) et des essais de luminescence ont été réalisés. Nous avons pu voir que les trois constructions ont un niveau de luminescence semblable (Figure 3.10). Ce ne sont pas les résultats attendus. En effet, compte tenu des données obtenues lors de nos premiers essais, il semblerait que la présence du ligand permette l'activation de l'expression de *Rmf* comme le souligne le niveau de luminescence dans les différents milieux du motif WT (Figure 3.8 - 3.11). Ainsi, s'il s'agissait d'un *riboswitch* ppGpp,

en absence de ligand (dans la souche $\Delta relA \Delta spoT$) on se serait attendu à ne plus avoir d'expression des constructions WT et comp. De plus, dans ce *knock-out* la luminescence pour la construction défaite contenant les mutations disruptives est aussi intense que pour le rapporteur avec l'ARN *rmf* WT, ce qui est également inattendu. Malgré le fait que ces données *in vivo* ne coïncident pas tout à fait avec notre hypothèse initiale, nous avons réalisé des cartographies *in vitro* par *in-line* nous permettant tout d'abord de confirmer en partie la structure prédite pour *rmf* (Figure 3.14). Nous avons également pu valider que le mutant *disr* entraîne une altération de la structure de l'ARN (Figure 3.15). De plus, il est intéressant de noter que ces modifications de structures ne se produisent pas forcément à l'endroit des mutations, mais que la présence de celles-ci provoque un repliement alternatif de l'ARN à différents endroits (Figure 3.15). Ces essais de cartographie par *in-line* nous ont également permis de déterminer des valeurs de K_D pour différentes constructions de l'ARN *rmf* en présence de ppGpp et de pppGpp (Figure 3.16 et Annexe 4). Malgré que ces résultats soient reproductibles nous avons pu remarquer que ce n'était pas toujours les mêmes bandes qui montraient une modulation dans chacun de ces gels. De plus, nous avons pu remarquer que même si les valeurs de K_D sont dans des ordres de grandeur similaires celles-ci ne sont pas tout-à-fait équivalentes (Tableau 5.2).

Tableau 5.2 Récapitulatif des valeurs de K_D obtenues pour les différents essais avec *rmf*

Construction testée	K_D moyen	K_D minimum	K_D maximum
<i>rmf</i> 1	1,1 $\mu\text{M} \pm 1 \mu\text{M}$	0,1 μM	2,1 μM
<i>rmf</i> 4 (Fig 3.14)	0,11 $\mu\text{M} \pm 0,06 \mu\text{M}$	0,05 μM	0,17 μM
<i>rmf</i> 4	1,4 $\mu\text{M} \pm 2,3 \mu\text{M}$	0 μM	3,7 μM
<i>rmf</i> 4 (Fig 3.16)	1,44 $\mu\text{M} \pm 0,6 \mu\text{M}$	0,84 μM	2,04 μM
<i>rmf</i> 4 (Fig 3.16 bandes Figure 3.14)	0,89 $\mu\text{M} \pm 0,2 \mu\text{M}$	0,69 μM	1,09 μM

Tableau récapitulatif des valeurs de K_D obtenus lors des essais de cartographie *in vitro* par *in-line* avec les différentes constructions de l'ARN *rmf* WT et calculées à partir de différents gels.

À la vue de tous les résultats obtenus ici, il est difficile de conclure sur le fait que *rmf* soit effectivement un *riboswitch* ppGpp en sachant également que le *riboswitch* connu pour le ppGpp

a un K_D de 10 nm (Sherlock *et al.*, 2018a). Cependant il n'est pas exclu que *rmf* puisse être un *riboswitch* ppGpp avec une affinité moins grande que le premier *riboswitch* caractérisé car les valeurs de K_D sont dans la gamme de concentration physiologique du ppGpp dans les cellules. Une étude a évalué que la quantité de ppGpp chez *E. coli* variait de 10 à 40 μ M en phase exponentielle jusqu'à 800 μ M en phase stationnaire de croissance (Steinchen *et al.*, 2020). Ainsi, suite au rôle de Rmf il est possible qu'une concentration plus forte de ppGpp soit nécessaire afin de réguler son expression. Néanmoins, afin de poursuivre de futurs essais concernant le ppGpp et *rmf*, les constructions à réaliser devraient idéalement prendre comme base la construction *rmf* 4 ayant montré des données reproductibles (Figure 3.16 et Annexe 4).

Compte tenu du fait que certains résultats ne concordent pas parfaitement avec l'hypothèse du *riboswitch*, d'autres hypothèses ont été proposées. Une seconde hypothèse concernant *rmf* serait que celui-ci pourrait être régulé par plusieurs facteurs. En effet, Rmf est une protéine importante dans le métabolisme de la bactérie permettant l'inactivation du ribosome en formant le 100S (Yamagishi *et al.*, 1993). L'expression de cette protéine a ainsi la nécessité d'être finement régulée. Ce type de régulation à plusieurs facteurs est courant chez les bactéries. En effet, certains éléments comme la concentration de fer dans la cellule ont besoin d'être finement régulés car c'est un métal essentiel, mais qui est aussi toxique à haute concentration (Andrews *et al.*, 2003). L'expression des gènes d'acquisition et de transport du fer sont régulés notamment par Fur et par l'ARN rég RyhB (Massé *et al.*, 2005). Ainsi, en présence de fer, le facteur de transcription Fur sera actif inhibant les gènes d'acquisition du fer tout en inhibant l'ARN rég RyhB ce qui permettra l'expression des gènes de stockage du fer (Massé & Gottesman, 2002). En absence de fer, le mécanisme inverse se passe avec une activation des systèmes d'acquisition du fer et une inhibition des gènes de stockage (Massé & Gottesman, 2002). Ainsi, il serait possible que le motif *rmf* soit la cible de plusieurs molécules. Une première molécule qui permettrait l'activation de l'expression de Rmf, expliquant ainsi la forte expression qui peut être observée sous le contrôle du motif *rmf* WT (Figure 5.1 A). Cette molécule pourrait interagir au niveau de la tige boucle 1 ainsi dans le cas du mutant *disr* l'interaction ne serait pas possible dû à l'altération de la structure (Figure 5.1 A). L'alarmone ppGpp est connue pour être un second messager pléiotropique modifiant l'expression de nombreux facteurs au sein de la bactérie (Hauryliuk *et al.*, 2015). Au sein de la souche Δ *relA* Δ *spoT*, ne produisant pas ppGpp, certains facteurs pourraient être présents dû à l'absence de répression par ppGpp. Cette deuxième molécule permettrait donc l'activation de l'expression de Rmf aussi bien pour le WT, que pour le *disr* et le *comp*. Ce facteur pourrait interagir à un autre endroit sur le motif *rmf* ainsi la présence de la mutation *disr* n'empêcherait pas la liaison contrairement à la première molécule (Figure 5.1 B). À la vue des

résultats obtenus en cartographie *in vitro* par *in-line*, il serait possible de penser que l'une de ces molécules est le ppGpp lui-même compte tenu de son interaction avec l'ARN *rmf* (Figure 3.16 et Annexe 4). Ceci pourrait ainsi représenter une nouvelle classe de plateforme d'expression de *riboswitch* répondant à deux molécules.

A *P. aeruginosa* WT



B *P. aeruginosa* $\Delta relA \Delta spoT = ppGpp -$



Figure 5.1 Modèle hypothétique de régulation du motif *rmf* via l'intervention de deux activateurs.

A : Dans la souche *P. aeruginosa* WT (PA01 ou PA14) le motif *rmf* WT permettrait la liaison au niveau de la tige-boucle 1 d'un activateur 1 (représenté par le rond bleu) permettant l'expression du gène en aval (représenté par la flèche verte). Cependant, la présence des mutations *disr* (représentée en rouge) empêcheraient la liaison de cet activateur 1 en défaisant la structure de la tige-boucle 1 ne permettant ainsi pas l'expression du gène *rmf*. **B** : Dans la souche *P. aeruginosa* $\Delta relA \Delta spoT$ ne produisant pas de ppGpp. Un activateur 2 (représenté en mauve) dépendant de l'absence de ppGpp pourrait être présent et se lier à un autre endroit sur *rmf* permettant ainsi l'expression que le motif soit défait (avec des mutations *disr*) ou non.

Enfin, une autre hypothèse pourrait être que le motif *rmf* est équivalent à un *ribosomal RNA leader* (*r-leader*). Les *r-leader* sont situés en amont des gènes codant pour les protéines ribosomales dans les bactéries et permettent de réguler l'expression de ces dernières par leur interaction avec la partie 5' UTR de l'ARNm (Eckert & Weinberg, 2020). Rmf étant impliqué dans le métabolisme des ribosomes, nous avons émis l'hypothèse que cette protéine pourrait faire partie d'une boucle de rétrocontrôle où la protéine Rmf elle-même viendrait se fixer sur le motif *rmf* en amont du gène bloquant ainsi son expression, tels les *r-leader*. En effet, comme on a pu le voir lors de la cinétique effectuée en milieu M9 à partir de 54h on observe une diminution de l'expression de la luminescence donc de Rmf (Figure 3.11 A). Ceci peut suggérer qu'à un certain stade dans la croissance, quand tous les ribosomes sont bloqués, il est nécessaire d'arrêter la transcription / traduction de la protéine via la fixation au niveau du motif *rmf* de la protéine Rmf. Par contre, la limitation des ressources en phase stationnaire pourrait aussi expliquer cette diminution de l'expression.

Comme mentionné précédemment, la protéine Rmf (*ribosomal modulation factor*) permet la dimérisation du ribosome 70S entraînant la formation du ribosome 100S (Yamagishi *et al.*, 1993). Ce ribosome est inactif et ne permet pas la traduction de protéines (Yamagishi *et al.*, 1993). Une autre hypothèse concernant ce motif pourrait ainsi être que l'expression de Rmf est protégée de l'action de Rmf et du ribosome 100S contrairement aux autres gènes de la bactérie. Cette protection pourrait être due à la présence du motif *rmf*. En effet, la présence des mutations disruptives pourrait entraîner une perte de la protection du motif d'où la diminution d'expression de la luminescence observée pour cette construction. De plus, il a été montré chez *E. coli* que la transcription de *rmf* est activée par l'alarmone de la réponse stringente : ppGpp (Izutsu *et al.*, 2001). Ainsi, dans des souches $\Delta relA \Delta spoT$, produisant moins de ppGpp il a été montré une absence de la protéine Rmf ainsi que de ribosomes 100S (Izutsu *et al.*, 2001). Dans la souche PA01 $\Delta relA \Delta spoT$, du fait de l'absence de Rmf l'inhibition de l'expression ne se produirait pas, d'où la forte expression de la luminescence pour toutes les constructions du motif (WT, *disr* et *comp*). Cette hypothèse se distingue de celle décrite précédemment nécessitant l'implication de deux facteurs (Figure 5.1). Cependant, ce dernier modèle présente des lacunes. En effet, cela implique que Rmf soit toujours présent dans nos essais ce qui ne coïncide pas avec ce qui a été publié.

Rmf est également une protéine importante dans la réponse à différents stress aussi bien chez *E. coli* que chez *P. aeruginosa* (Aspedon *et al.*, 2006; Yoshida & Wada, 2014). En effet, il a été montré que l'expression de Rmf était augmentée d'environ huit fois en condition de stress

osmotique chez *P. aeruginosa* (Aspedon *et al.*, 2006). De manière intéressante, il a été montré que des souches Δrmf présentent une viabilité cellulaire diminuée en phase stationnaire par rapport à des souches WT dû notamment à l'absence de ribosomes 100S (Yamagishi *et al.*, 1993). Ainsi, l'identification du mécanisme permettant la régulation de l'expression de Rmf chez *P. aeruginosa* pourrait constituer une cible pour un potentiel nouvel antibiotique en limitant la viabilité de cet organisme.

Le motif *rmf* est un de nos candidats les plus prometteurs montrant un profil d'expression intéressant. Cependant, des travaux complémentaires seront nécessaires afin d'identifier son mécanisme d'action. Il semble y avoir un lien entre *rmf* et l'alarmone ppGpp ainsi d'autres essais de cartographie *in vitro* par *in-line* pourrait permettre d'identifier si ce lien est direct ou indirect. De plus, il est possible de penser que la protéine Rmf elle-même puisse agir sur le motif soit via un système à deux activateurs (Figure 5.1) ou via une boucle de rétrocontrôle négatif tel un *r-leader*. Des essais de retards permettraient ainsi de déterminer s'il existe une interaction entre le motif et la protéine Rmf (Hellman & Fried, 2007). Enfin, des essais de luminescence avec des souches Δrmf pourraient également être réalisés.

5.1.2 Le motif *Pseudomon*-GGDEF

Le motif *Pseudomon*-GGDEF que nous avons étudié se retrouve dans le chromosome de *P. syringae* pv. *tomato* DC3000 devant le locus PSPTO_2907 dont le produit du gène contient des domaines GGDEF et EAL. Les essais de luminescence ont été réalisés dans tous les milieux répertoriés dans le Tableau 3.4. Des analyses statistiques ont permis de mettre en évidence certains milieux présentant des différences statistiquement significatives entre les constructions WT et *disr*, notamment pour les milieux M9 supplémentés avec de la sérine ou de la thréonine. Ces acides aminés sont présents dans la même voie de synthèse pouvant expliquer leur réponse similaire (Kanehisa *et al.*, 2020; Kanehisa & Goto, 2000). Des différences significatives sont également observables pour le milieu M9 supplémenté en NaCl ainsi que le SCFM. De plus, beaucoup de milieux testés montrent une expression de la luminescence semblable entre les constructions WT et *disr*. À ce stade, il est difficile de pouvoir statuer sur le métabolite ou le signal permettant la régulation via ce motif.

Cependant, *Pseudomon*-GGDEF est situé en amont d'un gène présentant des domaines GGDEF et EAL impliqués dans la synthèse et la dégradation de di-GMPc (Valentini & Filloux, 2016). Des essais de luminescence ont ainsi été effectués avec les constructions WT et *disr* pour le motif *Pseudomon*-GGDEF dans différentes souches de *P. aeruginosa* PA14 produisant des niveaux

différents de di-GMPc (Liberati *et al.*, 2006). Nous avons observé une diminution significative du niveau de luminescence dans la souche produisant un niveau faible de di-GMPc pour la construction avec l'ARN *Pseudomon*-GGDEF WT comparée aux souches WT et *wspF*- (Figure 3.19), ainsi qu'une absence de différence dans le niveau de luminescence pour les essais avec la construction *disr* dans les trois souches testées (Figure 3.25). Nous avons ainsi émis l'hypothèse que ce motif pourrait être un *riboswitch* di-GMPc. Des essais de cartographie *in vitro* par *in-line* ont été réalisés pour cinq constructions avec différentes concentrations de di-GMPc mais n'ont pas montré de modulation (Figure 3.19, Figure 3.20 et Annexe 5). À la vue de ces résultats, nous avons cherché à étudier une autre molécule impliquée dans le métabolisme du di-GMPc à savoir sa forme catalysée : le pGpG. Cette molécule a été testée car le gène en aval du motif *Pseudomon*-GGDEF contient un domaine phosphodiesterase (EAL) responsable de la dégradation du di-GMPc en pGpG (Römling *et al.*, 2005). Pour cette molécule nous sommes parvenus à obtenir une modulation apparente pour chacune des constructions testées (Figure 3.19, Figure 3.20 et Annexe 5) mais seulement à une concentration de 100 μ M qui est une concentration assez élevée. Un essai de cartographie *in vitro* par *in-line* a été réalisé avec l'ARN *rmf* 4 avec deux concentrations de pGpG à savoir 10 nM et 100 μ M et nous pouvons voir également une modulation apparente pour cet ARN (Annexe 6). Il ne semble pas y avoir de lien entre le pGpG et *rmf* ainsi nous ne nous attendons pas à avoir une modulation due à l'interaction entre ce ligand et l'ARN. Notre hypothèse est qu'un contaminant serait présent dans la solution de pGpG entraînant une réaction non-spécifique de dégradation à haute concentration qui cause une dégradation quel que soit l'ARN testé. Ainsi, la modulation observée pour *Pseudomon*-GGDEF n'est pas la preuve que cet ARN est un *riboswitch* pGpG. Cinq constructions ont été réalisées pour ce motif avec des ligands impliqués dans le métabolisme du di-GMPc mais aucune modulation n'a pu être observée ainsi il semble peu probable que cet ARN soit un *riboswitch* di-GMPc ou pGpG, ou alors des conditions différentes seraient requises.

Une étude de RNA-seq réalisée dans *P. syringae* pv. *syringae* B728a possédant également le motif *Pseudomon*-GGDEF a permis d'étudier les différences d'expression des différents gènes de cette bactérie dans des souches produisant différents niveaux de di-GMPc (Wang *et al.*, 2019). Cependant, cette étude n'a pas permis de mettre en évidence une différence d'expression du gène *Psyr_2711* situé en aval du motif *Pseudomon*-GGDEF chez cette souche en comparant les souches produisant plus et moins de di-GMPc (Wang *et al.*, 2019). Cette absence de différence au niveau de résultats de RNA-seq ne nous amène pas à rejeter notre hypothèse car il est possible que le potentiel *riboswitch* agisse au niveau de la traduction de l'ARN et pas sur sa transcription.

Cependant, bien que l'hypothèse du *riboswitch* liant le di-GMPc ou une molécule impliquée dans son métabolisme semble intéressante, il est possible qu'un autre mécanisme intervienne. En effet, le di-GMPc est un second messager dont la synthèse et la dégradation sont régulés par de nombreux facteurs aussi bien intracellulaires qu'extracellulaires (Mills *et al.*, 2011). Ces différents facteurs permettent au di-GMPc de réguler l'expression de ces différents gènes cible à différents niveaux (Mills *et al.*, 2011). Il a déjà été montré que certaines molécules tels que la glycine ou le tryptophane soient régulées par différents moyens. Pour la glycine, il existe notamment une classe de *riboswitch* répondant à ce ligand (Mandal *et al.*, 2004), mais aussi un T-box (Apostolidi *et al.*, 2015). Ces ARNnc en *cis* n'interviennent pas au même niveau. Effectivement, le *riboswitch* interviendra au niveau de la détoxification de la glycine alors que le T-box interviendra au niveau de sa synthèse (Apostolidi *et al.*, 2015; Mandal *et al.*, 2004). Ainsi, il est possible qu'une autre classe d'ARNnc en *cis*, en plus des classes connues de *riboswitchs* de di-GMPc, puisse être présente pour réguler l'expression de gène contenant des domaines GGDEF et EAL impliqués dans la synthèse et la dégradation du di-GMPc. De plus, Il a été montré que les molécules de di-GMPc jouent un rôle très important chez *P. syringae* (Wang *et al.*, 2019). En effet, 818 gènes ont été montrés comme différentiellement exprimés dans cette bactérie en comparant des souches produisant plus ou moins de di-GMPc (Wang *et al.*, 2019). Ainsi, il est également possible que la différence d'expression de la luminescence observée ne soit pas due directement à la molécule de di-GMPc mais à l'un des gènes différentiellement exprimés ou encore qu'il ne s'agisse que d'une corrélation. Dans ce cas, la bactérie tirerait un avantage équivalent de réguler le gène en aval du motif *Pseudomon*-GGDEF par une des conséquences découlant du rôle régulateur du di-GMPc sur l'une de ces nombreuses cibles dans le génome de *Pseudomonas*.

5.1.3 Le motif *Pseudomon*-Rho

Le motif *Pseudomon*-Rho est situé en amont du gène *rho* chez *P. aeruginosa* PA14. La protéine Rho est impliquée dans le métabolisme central de la bactérie. En effet, cette protéine hexamérique permet de dissocier l'ARN polymérase de l'ADN afin d'arrêter la transcription (Banerjee *et al.*, 2006). Cette protéine est essentielle pour la plupart des organismes procaryotes (Banerjee *et al.*, 2006). La régulation de la protéine Rho n'est pas encore connue, ainsi la présence d'un motif d'ARN en amont de ce gène pourrait indiquer la présence d'ARN non codant agissant en *cis* intervenant dans sa régulation.

Les essais de luminescence pour les constructions WT et *disr* ont été réalisées dans tous les milieux listés dans le Tableau 3.4. Les analyses statistiques réalisées sur ces données ont permis

de mettre en évidence des différences significatives entre les constructions WT et disr pour quelques milieux (Tableau 3.6 et Annexe 8). Certains de ces milieux sont des milieux M9 supplémentés avec divers acides aminés. Du fait de ces différences de niveau de luminescence observées, nous avons voulu étudier les voies de signalisations dans lesquelles étaient impliqués ces acides aminés afin de voir le lien que ceux-ci pourraient avoir avec Rho. L'isoleucine, la leucine et la valine sont des acides aminés branchés (BCAA) suivant la même voie de synthèse et de dégradation menant au cycle du TCA (Kanehisa, 2019; Kanehisa *et al.*, 2020; Kanehisa & Goto, 2000). De plus, il est connu que les BCAA jouent un rôle important dans divers aspects du métabolisme bactérien notamment dans la réponse stringente, avec ppGpp mais aussi via d'autres régulateurs globaux de la transcription, comme Lrp (Kaiser & Heinrichs, 2018). Cependant, en plus des BCAA d'autres conditions comme toutes les conditions en milieu LB, le milieu SCFM ou encore d'autres milieux M9 supplémentés en acides aminés (alanine, asparagine, sérine...) montrent des différences d'expression de la luminescence entre les deux constructions. Au vu de la disparité des milieux montrant des différences il est difficile de statuer sur un métabolite permettant la régulation de Rho. De plus, du fait du rôle central de Rho dans la bactérie, nous pourrions émettre l'hypothèse que le messenger permettant l'activation de l'expression de Rho soit un messenger plus central voire même Rho lui-même.

5.1.4 Le motif *sucA-II*

Le motif *sucA-II* est présent en amont du gène *sucA* dans *P. aeruginosa* PA14 codant pour la sous unité E1 du complexe 2-oxoglutarate déshydrogénase. Le gène *sucA* est présent dans le même transcrit que l'opéron *sdhCDAB* codant pour la succinate déshydrogénase impliquée dans le cycle du TCA. Le motif *sucA-II* est présent entre cet opéron et le gène *sucA* ce qui pourrait indiquer que le motif permette une expression différentielle dans certaines conditions entre *sdhCDAB* et *sucA*. Un exemple a déjà été montré en condition de carence en fer chez *P. aeruginosa*. En effet, dans ces conditions l'opéron *sdhCDAB* se retrouve sous-exprimé par rapport à la condition WT alors que l'expression du gène *sucA* ne varie pas (Nelson *et al.*, 2019). Une autre étude a également montré que la transcription de *sucA* était diminuée de façon significative en présence d'inuline et de sa forme hydrolysée (fructooligosaccharide) chez *P. aeruginosa* PA01 (qui possède également le motif *sucA-II* (Rubio-Gómez *et al.*, 2020; Weinberg *et al.*, 2010)), alors que l'expression de *sdhCDAB* n'était pas impactée par la présence de ces composés (Rubio-Gómez *et al.*, 2020). Ces deux observations peuvent conforter notre hypothèse que la présence du motif permet une différenciation de l'expression entre l'opéron situé en amont et le gène *sucA*.

Des essais de luminescence ont été réalisés dans toutes les conditions présentées dans le Tableau 3.4 et nous avons observé que pour tous ces milieux le motif WT montre un niveau de luminescence faible, alors que le mutant *disr* pour lequel la structure a été altérée montre un niveau de luminescence beaucoup plus élevé (Figure 3.27 A). Le motif *sucA-II* étant situé dans le cycle du TCA, des tests de luminescence ont également été effectués avec d'autres sources de carbone que des sucres (succinate, acétate, glycérol, citrate). Cependant, une nouvelle fois nous avons remarqué la même différence importante entre les deux constructions (Figure 3.27 B). Le profil d'expression étant similaire entre toutes les conditions testées il est difficile de statuer sur un potentiel signal / métabolite permettant la régulation de *sucA* via *sucA-II*.

Toutefois, grâce à nos essais nous avons pu montrer que la structure de *sucA-II* semble être importante pour la régulation du gène en aval. En effet, la présence des mutations disruptives dans la première tige-boule du motif modifiant la structure de celui-ci cause un changement dans l'expression du gène *sucA* visible via les variations de l'expression de la luminescence.

5.1.5 Le motif *fixA*

Le motif *fixA* est présent en amont du gène *etfB-2* (*fixA*) qui se situe dans un opéron avec *etfA-2* (*fixB*) chez *P. syringae* pv. *tomato* (Winsor *et al.*, 2016). Ces gènes codent pour les sous-unités α et β d'une flavoprotéine (Winsor *et al.*, 2016). Il a été montré que la sous-unité β contient un groupement FAD au niveau N-terminal de la protéine (Tsai & Saier Jr, 1995). Des essais de luminescence ont été réalisés avec toutes les conditions listées dans le Tableau 3.4. Les analyses statistiques faites sur ces données n'ont pas permis de montrer des conditions pour lesquelles il y avait des différences significatives dans le niveau de luminescence entre les deux constructions. Cependant, nous avons pu voir des différences dans certains milieux notamment dans le milieu TSB supplémenté avec 1 mM EDTA, où l'expression de la luminescence de la construction WT est plus faible que la construction *disr* (Figure 3.29). Nous nous sommes ainsi intéressés à la composition en ions métalliques de ce milieu. Une étude a analysé le milieu TSB et a montré qu'il contenait 0,05 ppm de cuivre, 0,6 ppm de zinc et 0,35 ppm de fer (Akhidime *et al.*, 2019). Cette composition est toutefois seulement indicative car elle peut varier entre différents lots et compagnie. L'EDTA est un chélateur d'ions divalents qui pourraient agir au niveau de différents ions divalents notamment le fer. Il a été démontré que le complexe enzymatique formé par les produits des deux gènes *fixA* et *fixB* est capable de fournir des électrons à la ferrédoxine qui à son tour permettra la réduction de nitrogénase (Tsai & Saier Jr, 1995). De plus, la ferrédoxine est composée d'un cœur fer-souffre (Adman *et al.*, 1973). Donc en absence de fer il ne serait pas

possible de former la ferrédoxine. Ainsi, avoir un contrôle en amont de la chaîne respiratoire permettrait une économie d'énergie pour la bactérie car le transfert d'électron ne pourrait pas se faire. À la vue des résultats obtenus et de la possibilité de l'importance du fer dans ce métabolisme nous pouvons émettre l'hypothèse que la concentration en fer pourrait être importante pour la régulation de l'expression des gènes *fixA* et *fixB* via le motif *fixA*. Des essais de luminescence ont été réalisés avec l'ajout de fer, mais sans résultats probants, d'autres essais pourraient être réalisés avec un agent chélateur spécifique du fer afin de vérifier cette hypothèse.

Nous pouvons cependant remarquer que d'autres milieux montrent également des différences dans le niveau de luminescence pour ce motif (Figure 3.29). En effet, nous avons pu voir pour certains milieux M9 supplémentés en acides aminés comme l'alanine, la glutamine, la sérine ou encore la thréonine ainsi que pour le milieu SCFM que la construction WT permet un niveau de luminescence plus fort comparé à la construction contenant les mutations *disr*. Ce profil est inversé par rapport à celui observé pour le milieu TSB supplémenté avec de l'EDTA présenté plus tôt.

Compte tenu du contexte génétique de *fixA*, des essais de luminescence ont été réalisés en condition anaérobie mais aucune différence dans le niveau de luminescence n'a pu être observé (Figure 3.30).

Les bactéries *P. syringae* et *P. aeruginosa* font partie du même genre bactérien cependant ce sont deux espèces différentes ainsi que deux spectres d'hôtes différents. En effet, *P. syringae* est un pathogène des plantes (Buell *et al.*, 2003) alors que *P. aeruginosa* est un pathogène opportuniste des humains (Lyczak *et al.*, 2000) ainsi il est envisageable que les systèmes de régulation ne soient pas les mêmes dans ces différents organismes. En perspective, aussi bien pour le motif *Pseudomon-GGDEF* que *fixA*, il faudrait idéalement parvenir à conjuguer les différents plasmides dans *P. syringae* pv. *tomato* DC3000.

5.1.6 Le motif *rne-II*

Le motif *rne-II* est présent en amont du gène *rne* codant pour la ribonucléase E. Les essais de luminescence ont été réalisés dans tous les milieux présents dans le Tableau 3.4 et plusieurs profils de niveau de luminescence ont été reportés. En effet, certaines conditions montrent un profil où le niveau de luminescence médié par le motif WT est plus grand que celui médié par le motif défectueux (contenant les mutations *disr*) et dans d'autres conditions le profil inverse est visible. Ceci est un résultat inattendu avec la fonction d'un ARNnc agissant en *cis*, car ceux-ci ont généralement un mode d'action dans un seul sens par l'activation ou l'inhibition du gène en aval.

Cependant, à la vue de ces résultats nous pouvons émettre l'hypothèse que l'expression de *rne* soit influencée par plus d'un facteur permettant l'activation ou la répression de l'expression via une liaison à différents endroits sur l'ARN. En effet, le motif *rne-II* fait 268 nts avec plusieurs structures tiges boucles distinctes qui pourraient permettre la liaison de différentes molécules (Figure 3.31).

Chez *E. coli*, un ARNnc nommé *rne5* et présent dans la partie 5' non traduite en amont du gène codant pour la ribonucléase E a été montré comme agissant en *cis* sur ce gène (Diwa et al., 2000). Cet ARN a été montré comme étant le siège d'une boucle de rétrocontrôle négatif où l'ARNase E vient cliver l'ARN au niveau de cette structure (Diwa et al., 2000). La structure de *rne5* est composée de trois tiges boucles et trois segments simple brin, la fonction de cet ARN dépend de deux de ces tiges-boucles nommées hp2 et hp3 (Diwa et al., 2000). L'ARNase E est capable d'interagir avec une partie de la tige boucle hp2 permettant ainsi la dégradation de l'ARNm en aval quand les conditions cellulaires sont rencontrées (Schuck et al., 2009). Cependant, les conditions pour lesquelles ce processus se déclenche restent peu claires (Schuck et al., 2009). La structure et la longueur de notre ARNnc candidat *rne-II* ressemblent en quelques points à *rne5*, il est constitué de cinq tiges boucles et trois segments simple brin (Figure 3.31). Ainsi nous pouvons émettre l'hypothèse que chez *Pseudomonas* un processus similaire pourrait avoir lieu grâce au motif *rneII*.

Toutefois, d'après les résultats obtenus il est difficile de conclure à savoir si la structure de *rne-II* semble importante pour la régulation de l'expression de *rne*. En effet, on peut voir des différences apparentes d'expression de la luminescence entre les constructions WT et défaites (contenant les mutations *disr*) dans plusieurs conditions mais ces différences ne sont pas statistiquement significatives. Ainsi, des essais complémentaires sont à réaliser afin de confirmer ces différences. Cependant, compte tenu de notre hypothèse concernant une possible boucle de rétrocontrôle de l'ARNase E sur son propre ARNm on ne s'attend pas nécessairement à avoir de grandes variations d'expression en fonction des milieux testés. En effet, il pourrait s'agir d'un moyen de contrôle du taux d'ARNase E dans la cellule via un rétrocontrôle présent dans l'ARNm. Afin de valider cette hypothèse, des essais *in vitro* pourraient également être réalisés afin de vérifier la capacité de clivage de l'ARNase E sur l'ARN *rne-II* WT ou contenant différentes mutations. L'action de différentes ARNase pourrait aussi être évaluée afin de vérifier la spécificité d'action ainsi que l'influence de la protéine RraA connue pour être un inhibiteur de l'ARNase E (Tang et al., 2010). Cette protéine entraîne une diminution de l'interaction entre l'ARNase E et le dégradosome (Tang et al., 2010).

5.1.7 Perspectives

Le vecteur d'expression utilisé lors de ce chapitre nous a permis d'étudier six ARNnc potentiels découverts par des méthodes de bioinformatique depuis plusieurs années, mais qui n'avait pas été encore caractérisés (Weinberg *et al.*, 2007; Weinberg *et al.*, 2010). Ce plasmide nous a permis d'obtenir beaucoup de résultats permettant de démontrer dans un premier temps l'importance de la présence des différentes structures d'ARNnc pour l'expression du gène en aval. Toutefois, comme nous l'avons montré précédemment pour des motifs comme *Pseudomon*-Rho ou encore *sucA-II*, il est difficile suite aux 40 tests effectués pour chaque ARN dans cette étude de déterminer quel est le stimulus permettant l'expression du gène en aval. En effet, lors de nos tests nous faisons varier un ou plusieurs composés, par exemple une supplémentation avec un acide aminé, mais cela cause des changements beaucoup plus importants au sein de la cellule que juste la présence de cet acide aminé supplémentaire. De plus, nous ne savons pas l'effet détaillé des mutations disruptives sur la structure de notre ARN, ainsi il est possible que l'ARNnc soit bloqué dans une conformation inactive et que les conditions intéressantes soient finalement les conditions dans lesquelles on ne voit pas de différences entre les constructions. Lors ce projet de doctorat, nous avons ainsi pu déterminer un rôle régulateur de quatre candidats ARNnc testés. Cependant, le mécanisme par lequel cette régulation est faite reste à déterminer.

Pour cela, différentes pistes sont envisageables. Nous pourrions tout d'abord penser à faire de nouvelles constructions de mutants disruptifs pour chacun des motifs testés afin de comparer les différences d'expression de la luminescence pour ces différentes constructions. Ces mutants pourraient être situés dans d'autres parties de la structure, par exemple une autre tige-boucle. De plus, la réalisation des constructions compensatoires permettrait également de valider les structures présentes.

Les essais réalisés avec les différents motifs ont permis de mettre en avant une limite à nos essais de luminescence, à savoir la contamination lumineuse entre les puits des plaques. En effet, nous avons utilisé pour tous nos essais des plaques 96 puits blanches à fond transparent recommandées pour les essais de luminescence (Auld *et al.*, 2020). Ces plaques permettent une diminution de la contamination lumineuse entre les puits et une réflexion de la lumière accrue dans chaque puit pour de meilleurs signaux, ainsi qu'une autoluminescence très faible (Auld *et al.*, 2020). Cependant, certains motifs comme *rmf* expriment fortement la luminescence avec des valeurs brutes de luminescence allant jusqu'aux millions pour certains échantillons. Or, certains motifs comme *rne-II* engendrent une expression beaucoup plus faible (environ mille fois moins forte). Pour les premiers essais de luminescence, plusieurs motifs étaient testés avec les mêmes

milieux dans une plaque, mais nous avons pu voir des contaminations de luminescence de puits adjacents entre les différents motifs dû aux différences d'expressions. Une alternative a donc été de faire des plaques avec les motifs individuellement. Cependant, malgré ce changement nous avons pu remarquer pour certains motifs comme *sucA-II* ayant une grande différence du niveau de luminescence entre les constructions WT et dir que la contamination luminescence pouvait entraîner des artefacts (Annexe 19). Une alternative dans le futur pour ces essais pourrait également être l'utilisation de plaque 96 puits grises (avec fond transparent) afin de réduire les contaminations entre les puits (Auld *et al.*, 2020). Ces plaques ont une réflexion un peu plus faible de la lumière comparées aux plaques blanches, mais permettent quand même de maintenir un signal haut (Auld *et al.*, 2020). Ainsi, ces plaques sont faites pour avoir un bruit de fond bas tout en maintenant un signal de luminescence important.

Lors de cette étude, plus de 40 conditions ont été testées basées sur les produits disponibles pour la réalisation des milieux. La compagnie Biolog commercialise des plaques permettant d'analyser plus de 2000 conditions en même temps sur une plaque. Ces milliers de conditions sont personnalisables en fonction des essais qui veulent être réalisés. Ceci nous permettrait ainsi de tester facilement des milliers de conditions en même temps. Ce type de plaque a déjà été utilisé chez *P. aeruginosa* afin de déterminer les gènes impliqués dans le transport à travers la membrane de différents composés (Johnson *et al.*, 2008). De plus, ce type de plaque Biolog a déjà été utilisé pour la caractérisation du motif *yjdF* initialement retrouvé dans l'étude de Weinberg et collaborateurs (Weinberg *et al.*, 2010). En effet, celui-ci a été montré comme étant un *riboswitch* liant des composés azaaromatiques comme le chélérythrine ou le déqualium (Li *et al.*, 2016).

Concernant les motifs *rmf*, *Pseudomon-Rho* et *rne-II* du fait de l'implication directe avec l'ARN des gènes en aval, en tant que protéines liant, dégradant ou causant la terminaison de l'ARN, nous pouvons émettre l'hypothèse que ces motifs pourraient fonctionner comme des boucles de rétrocontrôle où le produit du gène à savoir Rmf, Rho ou l'ARNase E respectivement puissent venir se lier au niveau de la partie 5' UTR de son propre ARNm pour moduler son expression. De plus, comme discuté précédemment, un motif nommé *rne5* présent en amont du gène codant pour la ribonucléase E chez *E. coli* permet de réguler l'expression du gène par ce moyen (Diwa *et al.*, 2000). Afin de valider ces hypothèses, des essais de retard sur gel pourraient être réalisés. Il serait également envisageable d'utiliser des souches de banque de transposon, ou des *knock-out*, ciblant les gènes concernés afin de vérifier l'expression de la luminescence en absence de ces protéines. Dans la banque de transposon réalisée par Liberati et collaborateurs, des souches de *P. aeruginosa* PA14 comportant des transposons dans les gènes *rne*, *rmf* et *rho* sont

disponibles (Liberati *et al.*, 2006). Des essais de luminescence pourraient ainsi être réalisés dans ces souches avec les constructions WT et *disr* des trois motifs afin de voir l'influence de l'absence de ces protéines sur l'expression de notre rapporteur.

Enfin, pour la suite des travaux sur cet objectif, nous sommes en train de commencer une collaboration avec le laboratoire du Pr Côté, à l'Université de Sherbrooke. Ce laboratoire possède une grande librairie de mutants de *P. aeruginosa* PA14 ainsi qu'un automate permettant de mesurer la luminescence dans toutes ces souches mutantes en même temps (French *et al.*, 2016). Le but serait ici de faire des conjugaisons dans chacun des mutants présents et de faire des essais de luminescence avec nos constructions WT et *disruptive*.

Comme montré précédemment dans la section 3.1.2 (Tableau 3.2), il existe un grand déséquilibre de caractérisation des ARNnc en *cis* chez les bactéries disposant de peu d'outils génétiques tels que *Pseudomonas* ou *Burkholderia* par rapport à *E. coli*. Ce projet de doctorat a permis de caractériser le potentiel de régulation de quatre ARNnc en *cis* prédits chez *Pseudomonas* permettant ainsi d'augmenter les connaissances sur la régulation de cette bactérie (Tableau 5.3). Cependant, les données obtenues pour les ARNnc candidats *fixA* et *rne-III* ne permettent pas encore de confirmer leur activité régulatrice car aucune condition testée ne montre de différence significative du niveau de luminescence entre la construction WT et *disr* (Tableau 3.6).

Tableau 5.3 Amélioration des connaissances sur les ARNnc en cis chez *Pseudomonas*.

	<i>E. coli</i>	<i>P. aeruginosa</i> (avant notre étude)	<i>P. aeruginosa</i> (après notre étude)
ARNnc en cis annotés	35	35	35
ARNnc en cis avec activité régulatrice confirmée	30	19	23*
Pourcentage d'ARNnc en cis avec activité régulatrice confirmée	85,7%	54,2%	65,7%
Distribution des ARNnc étudiés			
	Étude (Weinberg <i>et al.</i> , 2010)	Rfam (Kalvari <i>et al.</i> , 2021)	Étude (Greenlee <i>et al.</i> , 2018)
<i>rmf</i>	22	88	363
<i>Pseudomon-GGDEF</i>	7	-	65
<i>Pseudomon-Rho</i>	22	90	-
<i>sucA-II</i>	22	90	294

Tableau récapitulatif de la proportion d'ARNnc en cis annotés et/ou caractérisés chez *E. coli* et *P. aeruginosa* suite à notre étude. * les motifs *fixA* et *rne-II* ne sont pas intégrés à ces données car leur rôle de régulation reste encore à confirmer.

De plus, les motifs testés sont présents de nombreuses fois dans les génomes des Pseudomonadaceae et de différents autres organismes (Tableau 5.3). La recherche du nombre d'instances de chaque ARN a été effectuée via plusieurs biais afin de voir l'augmentation des connaissances de ces candidats au cours du temps (Tableau 5.3). Nous avons tout d'abord référencé le nombre d'instances dans l'étude dans laquelle ces motifs ont été prédits (Weinberg *et al.*, 2010), puis Rfam une base de données regroupant les ARN connus (Kalvari *et al.*, 2021) et enfin une étude plus récente discutant du potentiel de certains candidats ARNnc encore non caractérisés (Greenlee *et al.*, 2018). On peut voir une nette augmentation du nombre de ces instances au cours du temps et des études réalisées notamment à cause de l'augmentation des

souches séquencées. Il est ainsi possible que les ARN que nous avons étudiés dans ce projet de recherche soit encore plus présents que ce que nous pensons actuellement.

Les résultats obtenus ici pourraient ne pas uniquement se cantonner aux Pseudomonadaceae. Nous pourrions prendre l'exemple du motif *sucA-II* qui est retrouvé 294 fois toujours en amont du gène *sucA* (Greenlee *et al.*, 2018). L'identification du mécanisme de régulation de cet ARN en *cis* pourrait être d'autant plus intéressant car la présence d'un motif en amont du gène *sucA* chez les bactéries du genre Burkholderia a été également rapporté (Weinberg *et al.*, 2007). Ce motif est retrouvé également 490 fois dans des génomes bactériens séquencés (Greenlee *et al.*, 2018). Or, il est possible que le mécanisme de régulation de ces deux ARN soient similaires ouvrant ainsi la porte à une meilleure compréhension du métabolisme de divers organismes. Ceci fut notamment le cas pour les motifs d'ARN longtemps appelés *ykkC*, *mini-ykkC* et *ykkC-III* (Barrick *et al.*, 2004; Weinberg *et al.*, 2007; Weinberg *et al.*, 2010). En effet, ces ARN étaient des *riboswitchs* orphelins jusqu'à ce que *ykkC* soit montré comme étant un *riboswitch* guanidine (Nelson *et al.*, 2017). Peu de temps après, les fonctions de *mini-ykkC* et *ykkC-III* ont été confirmés comme étant également des *riboswitchs* guanidine (Sherlock & Breaker, 2017; Sherlock *et al.*, 2017).

Bien que pVK-f2.2-lux permette une amélioration des connaissances de la régulation dans des bactéries peu étudiées au niveau des ARNnc en *cis*, une limite réside sur le fait que chacun des motifs doit être testé individuellement. Il ne s'agit ainsi pas d'une méthode à haut débit. C'est pourquoi nous avons développé une nouvelle méthode expérimentale à haut débit conçue de manière à découvrir de nouveaux *riboswitchs* à partir d'une librairie d'ARN, appelée le Shifted-Reverse PAGE.

5.2 Développement d'une nouvelle méthode expérimentale pour la découverte de *riboswitchs* : le Shifted Reverse-PAGE

Le Shifted-Reverse PolyAcrylamide Gel Electrophoresis (SR-PAGE) est une méthode expérimentale développée afin de permettre la découverte de nouveaux *riboswitchs* dans différents organismes. Elle est basée sur la caractéristique principale de ces ARN à changer de conformation suite à la liaison de leur ligand. Pour cela, les ARN sont séparés dans un gel de polyacrylamide natif lors d'une première migration puis une solution de ligand est vaporisée sur le gel permettant aux ARN correspondant d'interagir avec les différents ligands présents dans cette solution. Une seconde migration est ensuite réalisée. À la fin de celle-ci il est possible

d'observer un changement de vitesse de migration, qui se traduit par un décalage de migration, dans le cas des ARN ayant interagi avec l'une des molécules de notre solution.

5.2.1 Discussion SR-PAGE

Ce projet de doctorat a permis d'optimiser et de valider cette méthode. L'optimisation a été réalisée sur différents aspects. En effet, la partie la plus critique dans la réalisation du SR-PAGE et ayant demandé le plus d'optimisation était au niveau de l'étape de séparation des plaques de verre et de vaporisation du ligand. Comme décrit précédemment, afin de pouvoir vaporiser les ligands sur le gel les plaques de verre ont la nécessité d'être séparées, pour cela des étapes de traitement des vitres sont essentielles pour que le gel puisse adhérer à la grande plaque de verre et se décoller simplement de la petite. Une autre étape d'optimisation à cette étape a été le volume de ligand à vaporiser ainsi que les concentrations. En effet, des tests ont été effectués en vaporisant des colorants afin de vérifier la quantité de colorant entrant dans le gel de manière à appliquer ces ratios aux ligands. Le volume nécessaire de ligand était également important à évaluer car il faut suffisamment de volume afin de pouvoir recouvrir la totalité du gel, mais un trop gros volume pourrait également être délétère au moment de devoir refermer le système. Finalement, un volume de 50 mL de ligand a été choisi pour la réalisation de cette expérience. Enfin, lors des premiers essais de SR-PAGE nous avons pu remarquer que beaucoup des ARN testés semblaient bloqués dans les puits. Des tests avec des APS de différentes compagnies ont permis de résoudre ce problème, ainsi pour la réalisation du SR-PAGE il est conseillé d'utiliser de l'APS Biorad.

Suite à ces différentes étapes d'optimisation, nous sommes parvenus à valider le SR-PAGE en obtenant des décalages de migration pour plusieurs constructions de *riboswitchs* connus répondant à des ligands de natures différentes comme des seconds messagers (di-GMPc), des cofacteurs d'enzymes (FMN et TPP) ou des acides aminés (glycine). Cette partie du projet a permis de mettre en évidence une capacité du SR-PAGE qui permettrait d'étudier plus en détail les plateformes d'expression des *riboswitchs*. En effet, il serait possible d'analyser les plateformes d'expression expérimentalement par la présence d'un décalage ainsi que son amplitude. Ceci permettant de mieux comprendre les mécanismes de fonctionnement des *riboswitchs*.

Cette méthode a été pensée afin de pouvoir découvrir de nouveaux *riboswitchs* dans des bibliothèques d'ARN à la manière d'un SELEX. Nous avons aussi validé cette application prévue du SR-PAGE. En effet, nous avons été capables de sélectionner des *riboswitchs* avec une affinité modifiée pour leur ligand. La tige P4-P5 du *riboswitch* TPP permettant la liaison de la partie pyrophosphate de

la molécule de TPP a été modifiée aléatoirement via la création de trois librairies (Edwards & Ferré-D'Amaré, 2006). Ces librairies ont par la suite été soumises à des cycles de SELEX utilisant le SR-PAGE comme méthode de sélection afin de sélectionner les ARN liant préférentiellement la thiamine plutôt que le TPP. Parmi tous les clones, nous sommes parvenus à sélectionner dans une librairie contenant des milliards de séquence (librairie trois) deux clones liant la thiamine avec une meilleure affinité que le TPP et aussi une meilleure affinité par rapport au *riboswitch* WT (Figure 4.6).

Bien que le SR-PAGE semble uniquement destiné à la recherche de nouveaux *riboswitchs* cette méthode pourrait être appliquée à toutes sortes d'ADN ou d'ARN présentant un changement de structure entre deux conditions. Cette méthode semble être la seule méthode développée jusqu'à date pouvant sélectionner des séquences dans des librairies naturelles contenant plus de 10^{15} séquences (Figure 5.3). Les autres méthodes le pouvant sont des SELEX utilisant pour la plupart des librairies synthétiques.

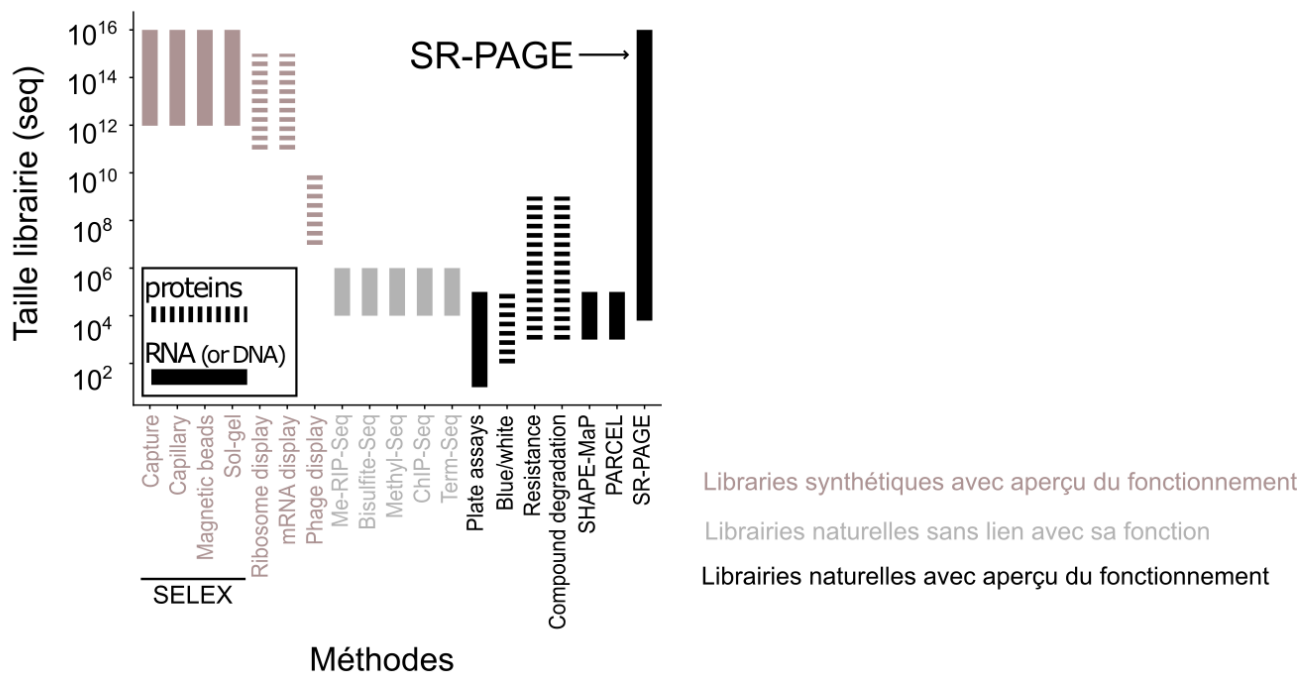


Figure 5.2 Comparaison du SR-PAGE avec les méthodes connues pour la découverte de nouvelles molécules.

Différentes méthodes utilisées pour la découverte de nouvelles molécules ont été représentées en fonction du nombre de séquences testées possibles. En brun, les méthodes sont réalisées à partir de bibliothèques synthétiques et celles-ci peuvent donner un aperçu du fonctionnement de ces molécules (Bazan *et al.*, 2012; Bruno & Kiel, 2002; He & Taussig, 2002; Jing & Bowser, 2011; Lyu *et al.*, 2021; Mendonsa & Bowser, 2004; Wilson *et al.*, 2001). En gris, les méthodes utilisent des bibliothèques naturelles mais le lien avec la fonction de l'ADN ou l'ARN n'est pas défini (Dar *et al.*, 2016; Kidder *et al.*, 2011; Lee *et al.*, 2013; Meyer *et al.*, 2012; Sun *et al.*, 2015). Ces méthodes sont limitées par les kits de préparation de librairie ainsi que par les capacités de séquençage. En noir, les méthodes utilisent des bibliothèques naturelles pouvant en plus donner un aperçu sur la fonction de la séquence sélectionnée (Devinck *et al.*, 2022; Siegfried *et al.*, 2014; Tapsin *et al.*, 2018).

Les SELEX génomiques sont les exemples les plus proches par rapport à notre méthode, utilisant des bibliothèques naturelles. Cette technique a permis de découvrir un ribozyme au sein du génome humain dans une librairie (Salehi-Ashtiani *et al.*, 2006). Pour cela, des bibliothèques d'ADN circularisé ont été utilisées mais cette technique n'est applicable que dans le cas de ribozyme étant capable d'auto-clivage (Salehi-Ashtiani *et al.*, 2006).

5.2.2 Perspectives du SR-PAGE

Le but de notre projet de recherche a été également la découverte de nouveaux *riboswitchs* via le SR-PAGE. Pour cela plusieurs méthodes de préparation de bibliothèques ont été testées. Nous avons tout d'abord préparé différentes bibliothèques à partir d'ADN génomique (Singer *et al.*, 1997). Cinq bibliothèques ont été réalisées, une comportant l'ADN génomique d'une vingtaine de bactéries,

une librairie à partir de quatre espèces de champignons, une librairie à partir du microbiote intestinal et enfin une librairie d'ADN humain. Le but pour l'étude du microbiote intestinal est que toutes les bactéries ne sont pas séquencées et sont ainsi exclues des méthodes de recherche par bioinformatique. Concernant l'ADN humain, la présence de *riboswitchs* a été montrée dans les trois domaines de la vie. Cependant, chez les eucaryotes seul le riboswitch TPP a été montré chez les champignons et les plantes (Sudarsan *et al.*, 2003a). Nous nous sommes ainsi demandés si des *riboswitchs* ne pourraient pas se trouver dans l'ADN humain. Le principe de la réalisation de ces librairies était d'obtenir une très grande couverture c'est-à-dire nucléotides par nucléotides de l'ADN complet de tous les organismes testés (Singer *et al.*, 1997). Ceci dans le but d'obtenir toutes les combinaisons possibles de l'ADN car comme nous avons pu le voir lors de nos essais de validation de la méthode, parfois quelques nucléotides de plus ou de moins peuvent permettre d'obtenir un décalage. Celui-ci pouvant également être plus ou moins important en fonction de la séquence. Le principe de la méthode pour ces librairies repose sur l'utilisation d'amorces dégénérées comportant neuf nucléotides aléatoires et une partie constante permettant l'amplification par PCR des fragments de librairie formés (Figure 5.4). Les deux brins sont synthétisés l'un à la suite de l'autre grâce à l'ADN polymérase Klenow (Figure 5.4). Les fragments de librairie contenant les parties constantes sont amplifiés par PCR permettant l'ajout du promoteur T7 pour la transcription de cette librairie en ARN (Figure 5.4).

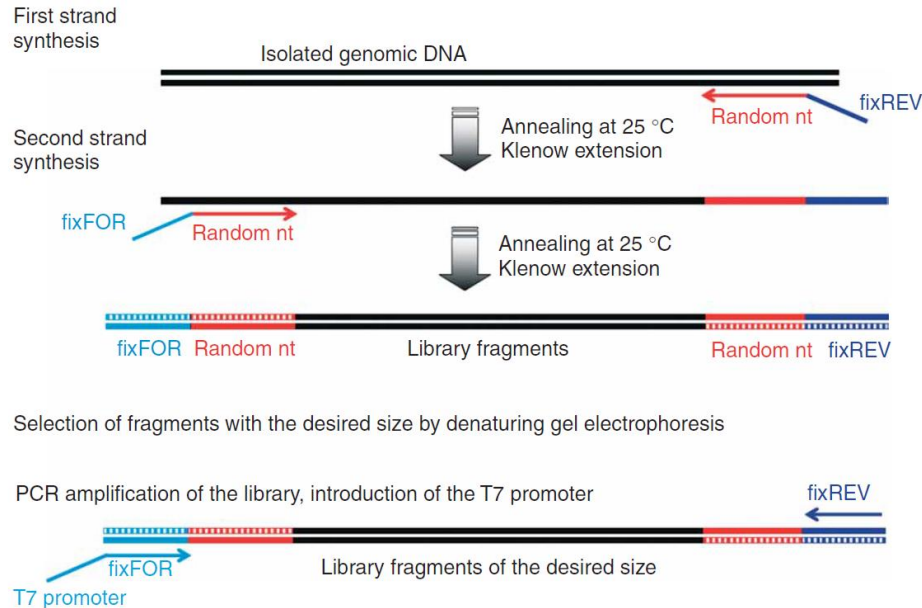


Figure 5.3 Méthode utilisée pour la réalisation des différentes librairies d'ADN génomique.

Une fois l'ADN génomique extrait, la librairie est réalisée avec des amorces dégénérées contenant chacune une séquence de neuf nucléotides aléatoires (représentés en rouge, random nt) ainsi que deux séquences connues différentes pour chacune des amorces (représentées en bleu foncé et bleu clair). Après avoir synthétisé chacun des brins contenant ainsi de part et d'autre les séquences connues, les fragments de librairie sont choisis en fonction de leur taille et amplifiés par une réaction PCR ciblant les séquences connues des amorces et permettant l'ajout du promoteur T7 pour la transcription en ARN. Tirée de (Lorenz *et al.*, 2006).

De nombreuses étapes d'optimisation ont été testées pour la réalisation de ces librairies. L'une des étapes cruciales est la purification des brins formés lors de la première synthèse. En effet, il est important d'enlever le maximum d'amorces pour la suite de la synthèse, mais aussi de perdre le moins de séquences possibles pour la couverture de notre librairie. Une autre étape importante a été la sélection des tailles des fragments de librairie et l'éluion de ces bandes d'ADN. Des fragments de 350 à 500 nucléotides étaient recherchés. Pour cela, plusieurs méthodes ont été utilisées comme l'utilisation de kit d'extraction sur gel, la dialyse ou encore l'électroélueur (Zarzosa-Álvarez *et al.*, 2010) qui a été la solution montrant le meilleur rendement. L'avantage de ce type de librairie aurait été que l'on puisse chercher des *riboswitchs* dans toutes les parties du génome et pas uniquement dans les séquences intergéniques comme le font la plupart des méthodes de bioinformatiques.

Des cycles de SELEX avec comme méthode de sélection de SR-PAGE ont été réalisés sur ces cinq librairies, puis ont été séquencés par la technologie Illumina. Les résultats obtenus ont mis en évidence la présence de biais de séquence dans la réalisation de la librairie. Nous avons

remarqué que la séquence connue des amorces pouvait se lier directement à l'ADN et avec une meilleure affinité que la partie dégénérée. En effet, toutes les amorces contiennent cette partie, ainsi ce fragment se retrouvera grandement amplifié lors de la synthèse de la librairie. Afin de résoudre ce problème, la synthèse des librairies a été refaite en utilisant cette fois une enzyme Klenow exo -. Cette polymérase est privée de son activité exonucléase 5' → 3'. Des cycles de SELEX ont été réalisés. Malgré tout, après séquençage des biais de séquences ont de nouveau été observés.

Un des problèmes rencontrés pourrait venir du fait que nous avons utilisé les amorces décrites dans l'article de la méthode (Singer *et al.*, 1997). Or, ces amorces ont été utilisées spécifiquement pour réaliser des librairies sur l'ADN humain, *E. coli* et des levures montrant de bons résultats avec une couverture nucléotides par nucléotides (Singer *et al.*, 1997). Ainsi, il est possible que du fait de la diversité des organismes que nous utilisons dans nos différentes librairies ces amorces ne soient pas adaptées. En effet, une étude plus récente utilisant le même principe de librairie indique que cette étape est cruciale et la séquence est dépendante des organismes utilisés (Lorenz *et al.*, 2006). Il est nécessaire que la partie constante ne soit pas retrouvée dans chacun des génomes ciblés (Lorenz *et al.*, 2006). Il aurait été bénéfique de créer des amorces de librairies spécifiques à chacun des organismes testés ou tout du moins des petits groupes pour lesquels les séquences connues des amorces étaient compatibles. Ces différentes librairies pourraient par la suite être assemblées pour la transcription limitant ainsi la probabilité que les séquences connues des amorces puissent s'hybrider avec l'ADN des bactéries ou champignons.

Suite aux différents essais infructueux rencontrés pour la réalisation des librairies génomiques, un autre type de librairie a été envisagé. En effet, nous avons commandé une librairie synthétisée grâce à une plateforme de synthèse électrochimique d'oligonucléotides à base de semi-conducteurs (CMOS) (Caraballo, 2018). Cette technologie mise en place par Custom Array™, permet de synthétiser sur des puces des milliers d'oligonucléotides grâce à trois étapes successives : détritulation, couplage et lavage (Caraballo, 2018). Chacun de ces cycles permet l'ajout d'un nucléotide à l'oligonucléotide. Différentes librairies ont été réalisées à savoir une librairie contenant toutes les séquences intergéniques des bactéries comprenant un terminateur de transcription Rho indépendant, une librairie contenant tous les ARN non codants annotés et une librairie contenant des *riboswitchs* annotés. La création de la librairie contenant toutes les séquences intergéniques a été réalisée via un script de bioinformatique ainsi que différents outils (Figure 5.5). Pour cela, la base de données Ribogap a été utilisée afin d'extraire toutes les séquences intergéniques contenant un terminateur de transcription Rho indépendant

(Naghdi *et al.*, 2017). Les séquences ont été triées en utilisant l'outil cdhit pour éliminer les séquences trop proches (Fu *et al.*, 2012; Li & Godzik, 2006) (Figure 5.5). Ces séquences ont été par la suite analysées par rapport à la position du terminateur. En effet, la distance entre le début du terminateur et la fin du gène en amont ainsi que celle entre la fin du terminateur et le début du gène en aval ont été mesurées. Les séquences présentant une distance entre la fin du terminateur et le début du gène en aval plus petites que la mesure en amont ont été sélectionnées. La technologie de synthèse d'oligonucléotides proposée par GenScript permet de produire des fragments inférieurs à 170 nts, les fragments synthétisés sont ainsi créés de manière à répondre à ce critère avec une séquence d'ADN de 130 nts entourés par deux séquences adaptatrices de 20 nts. Pour chaque séquence sélectionnée, trois constructions ont été réalisées en partant du milieu de la séquence du terminateur jusqu'à 130 nts puis deux décalages de 10 nts ont été réalisés de manière à avoir plusieurs combinaisons de chaque séquence intergénique (Figure 5.5).

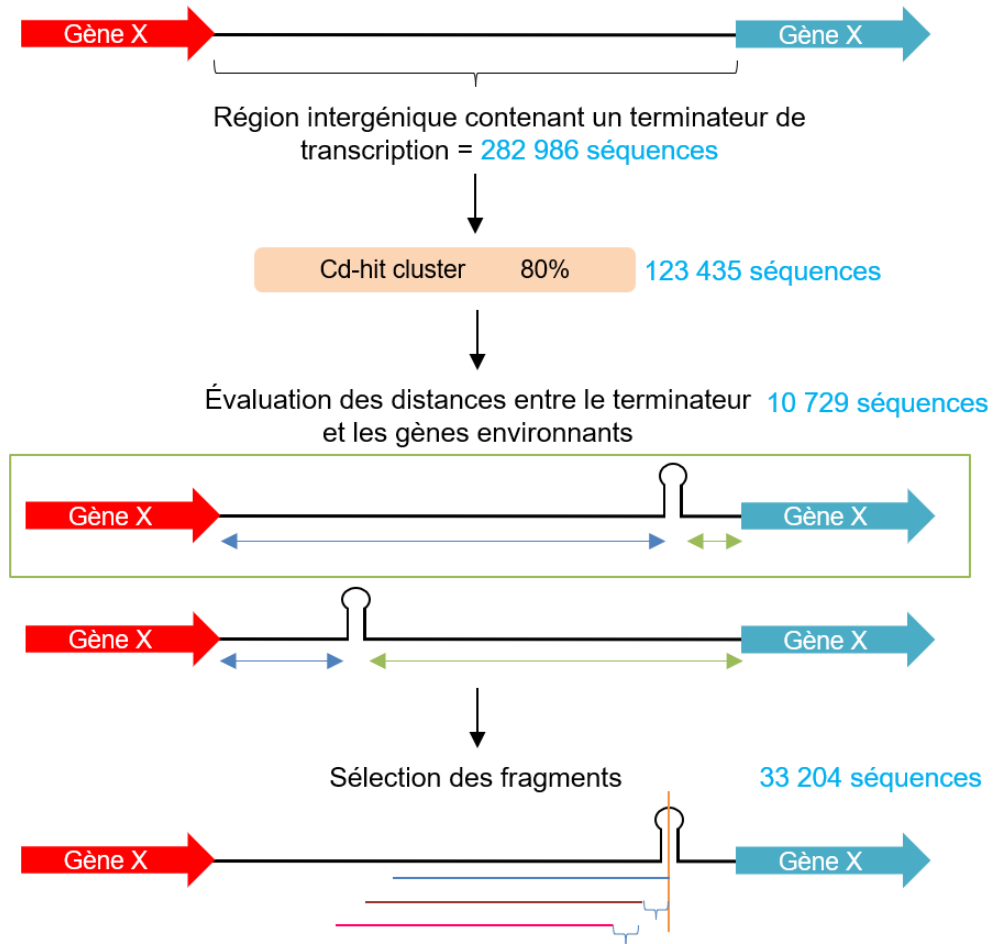


Figure 5.4 Représentation schématique de la librairie IGR d'oligonucléotides.

Les régions intergéniques sont présentes entre deux gènes (X et Y) dans les bactéries, elles ont été extraites et traitées avec cdhit avec un seuil de 80%. La position des terminateurs de transcription (représenté par une tige boucle) dans les séquences triées a été analysée. Les séquences pour lesquelles la distance entre le terminateur et le gène Y (flèche verte) est inférieure à celle entre le terminateur et le gène X (flèche bleue) ont été sélectionnées. Les fragments de librairie (représentés en bleu, marron et rose) ont été créés à partir du milieu du terminateur (représenté en orange) pour une longueur de 130 nts. Deux décalages de 10 nts ont été réalisés pour réaliser ces trois séquences.

Pour les librairies contenant tous les ARN non codants annotés et les *riboswitchs*, les séquences ont été extraites de nouveau grâce à la base de données Ribogap puis les séquences proches ont été enlevées en utilisant l'outil cdhit comme décrit précédemment. Enfin, uniquement les séquences d'ARN inférieures ou égales à 130 nts ont été sélectionnées. Des cycles de SELEX ont été réalisés avec chacune de ces librairies et séquencés par la technologie Illumina. Le séquençage a été réalisé en faisant des lectures de 75 nts à partir de chacune des amorces de nos séquences (Figure 5.5 A). Du fait de la longueur des lectures de 75 nts, le résultat attendu

aurait été d'avoir notre amorce suivie de 55 nts de notre séquence d'intérêt que l'on aurait pu retrouver grâce à notre fichier de référence et aucune séquence ne comprenant les amorces F et R (Figure 5.6 A). Cependant, l'analyse du séquençage a montré que nous avons aussi pour ces bibliothèques des artefacts dans nos séquences, mais d'une autre sorte. En effet, des dimérisations d'oligonucléotides ont pu être observées. Au lieu d'avoir comme expliqué précédemment l'amorce F ou R suivi de 55 nts de séquence, nous avons pu observer la présence des deux amorces dans les lectures de 75 nts pour environ 80% de nos lectures dans les bibliothèques initiales ayant servi aux différents cycles de SELEX. Une analyse des distances entre ces deux amorces a été réalisée. Nous pouvons voir qu'environ 2000 lectures sur près de 4000 ont moins de 15 nts entre ces deux amorces, correspondant ainsi à une dimérisation des amorces ou une favorisation des séquences de très petites tailles (Figure 5.6 B).

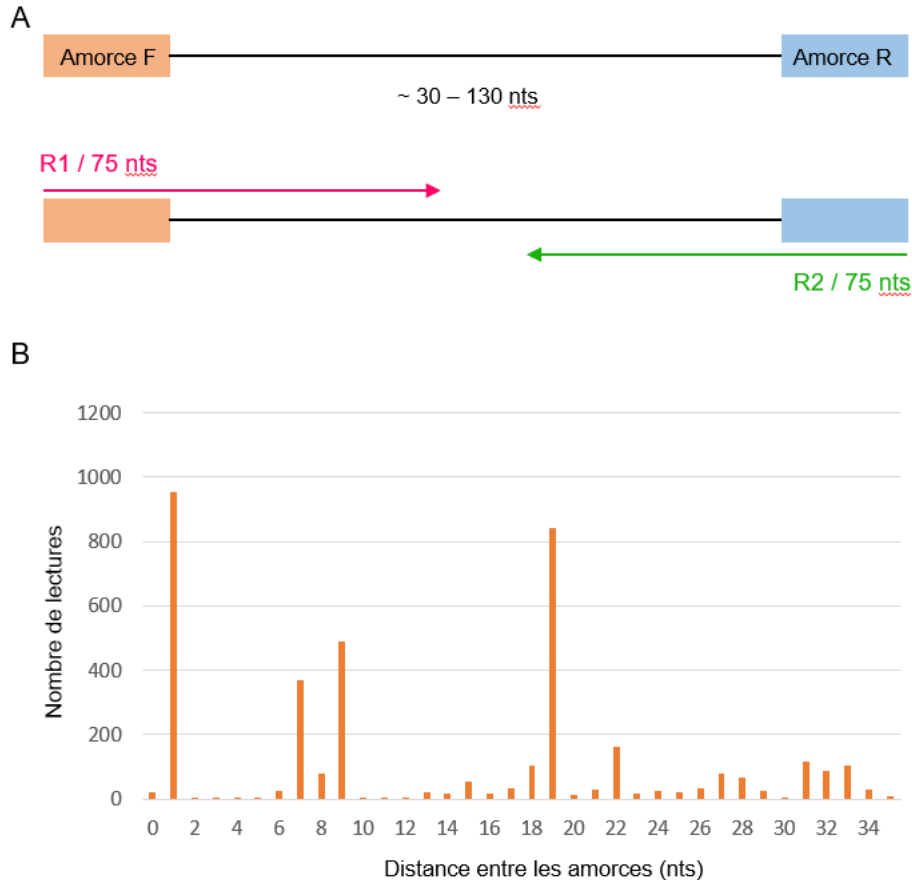


Figure 5.5 Dimérisation d'amorces lors du séquençage de la bibliothèque Rw d'oligonucléotides.

A : Les fragments de bibliothèques contiennent une amorce forward (rectangle orange) et une amorce reverse (rectangle bleu) séparées par des séquences de 30 à 130 nucléotides. Pour le séquençage Illumina, les lectures R1 et R2 d'une longueur de 75 nts ont été effectuées à partir des amorces F et R respectivement. **B** : Distribution de la distance entre les amorces F et R lors du séquençage de la bibliothèque initiale de *riboswitchs*. Une majorité de séquences >100 nts étaient attendue.

De plus, l'analyse des séquences restantes a mis en évidence une favorisation des plus petites séquences. Après discussion avec la compagnie GenScript, il pourrait s'agir d'un problème au niveau de la synthèse des oligonucléotides. La longueur de 170 nts n'est pas la longueur optimale et peut mener à avoir des séquences tronquées donc favorisant les plus petites séquences. De plus, ayant des séquences de différentes tailles à amplifier lors de la PCR il est possible que cette étape également favorise les petites séquences.

Une nouvelle librairie de *riboswitchs* a été réalisée d'une longueur de 150 nts comprenant la séquence d'ADN d'intérêt de 110 nts entourée par deux séquences adaptatrices de 20 nts. Cette librairie a comme particularité de n'avoir qu'une seule taille de séquences contrairement aux librairies réalisées auparavant. En effet, seulement les *riboswitchs* compris entre 108 et 110 nts ont été extraits. Précédemment nous avons montré que la présence des séquences adaptatrices pouvait empêcher le décalage observé pour des *riboswitchs* connus (section 4.3.4, Figure 4.6). Or nous n'avons pas évalué l'importance de la présence de ces oligonucléotides complémentaires dans la sélection au sein d'une librairie. Ainsi, pour cette nouvelle sélection avec la librairie de *riboswitch*, une comparaison entre une condition avec ou sans oligonucléotides complémentaires sera effectuée. Un cycle de sélection a été effectué avec cette librairie et deux couples de ligands. Malheureusement, les données obtenues ne montrent pas le résultat escompté. Il est difficile de visualiser dans ces résultats une sélection des *riboswitchs* correspondant aux ligands vaporisés. Ainsi, plusieurs cycles pourraient être nécessaires afin de parvenir à sélectionner spécifiquement les *riboswitchs* d'intérêts. Néanmoins, l'analyse des données a permis de montrer une diminution du nombre de dimères d'amorces comparés aux essais précédents avec les librairies contenant des fragments de tailles différentes.

En plus des différents biais observés avec les différentes méthodes de production de librairies, il est possible qu'un problème soit également la technologie utilisée pour le séquençage qui ne soit pas adaptée au type d'échantillons que nous avons. En effet, que cela soit pour les librairies d'ADN génomique ou encore les librairies d'oligonucléotides les tailles des séquences ne sont pas constantes. Or, il a été montré que les fragments de petite taille sont également favorisés lors du séquençage par rapport aux grands (Illumina, 2020).

Enfin, une autre alternative à ces librairies pourrait être l'utilisation de kit commerciaux de préparations de librairies. Plusieurs sont disponibles et présentent différentes caractéristiques (Sato *et al.*, 2019).

5.3 Conclusion

Les outils de bioinformatique utilisant les comparaisons de structure et la recherche de covariation ont permis de découvrir la plupart des ARNnc en *cis* connus de nos jours (Yao *et al.*, 2006). Ces méthodes puissantes permettent de fournir de nombreux résultats, ceci peut être vu comme un avantage mais c'est également une limite. En effet, ces nombreux candidats ont la nécessité d'être validés expérimentalement notamment par des méthodes *in vivo* avec des rapporteurs d'expression (Sherlock & Breaker, 2020). Cependant, une différence entre les bactéries présentant de nombreux outils génétiques et ceux en possédant moins se crée alors pour la caractérisation de ces ARNnc. Nous avons développé un rapporteur luminescent permettant de rééquilibrer ce niveau de caractérisation notamment pour *Pseudomonas* et *Burkholderia*. Lors de cette étude, quatre structures d'ARNnc candidats ont été caractérisées chez *Pseudomonas* réduisant ainsi le déséquilibre entre ces organismes. Les limites des outils de bioinformatique résident également dans l'annotation de gènes et la nécessité d'avoir plusieurs séquences afin de faire un alignement pour analyser cette covariation. Ainsi, des motifs plus rares ne disposant que de quelques instances ou des motifs présents dans des bactéries ne présentant que peu de souches séquencées rendra difficile la découverte de nouveaux motifs. C'est pourquoi nous avons développé la méthode de SR-PAGE pour pallier à ces limites.

Les deux méthodes développées et optimisées lors de ce projet de doctorat sont interconnectées. En effet, les potentiels *riboswitchs* trouvés via la méthode du SR-PAGE ont la nécessité d'être validés par une autre méthode, ceux-ci pourraient l'être via l'utilisation du plasmide rapporteur pVK-f2-lux, notamment dans le cas de *riboswitchs* retrouvés chez *Pseudomonas* ou *Burkholderia*. De plus, la capacité de changement de conformation des *riboswitchs* potentiels retrouvés grâce aux essais de luminescence pourrait être vérifiée par le SR-PAGE. Des essais ont d'ailleurs été réalisés pour *Pseudomon-GGDEF* et *rmf* avec le di-GMPc ou le ppGpp respectivement, mais n'ont pas donné de décalage apparent.

Enfin, ces deux méthodes pourraient permettre de découvrir et de caractériser de multitudes de nouveaux ARNnc en *cis* dans différents organismes en outre-passant les limites des outils de bioinformatique ou complétant les lacunes de ces méthodes. Ces découvertes permettraient ainsi d'augmenter nos connaissances sur la régulation et le métabolisme de divers organismes moins étudiés qu'*E. coli*. Ces méthodes n'ont pas pour but d'éliminer les outils de bioinformatique mais plutôt les compléter.

6 BIBLIOGRAPHIE

- Adman ET, Sieker LC & Jensen LH (1973) The structure of a bacterial ferredoxin. *Journal of Biological Chemistry* 248(11):3987-3996.
- Akhidime ID, Saubade F, Benson PS, Butler JA, Olivier S, Kelly P, Verran J & Whitehead KA (2019) The antimicrobial effect of metal substrates on food pathogens. *Food and bioprocesses processing* 113:68-76.
- Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403-410.
- Altuvia S, Kornitzer D, Teff D & Oppenheim AB (1989) Alternative mRNA structures of the cIII gene of bacteriophage λ determine the rate of its translation initiation. *Journal of molecular biology* 210(2):265-280.
- Andrews SC, Robinson AK & Rodríguez-Quñones F (2003) Bacterial iron homeostasis. *FEMS Microbiology Reviews* 27(2-3):215-237.
- Apostolidi M, Saad NY, Drainas D, Pournaras S, Becker HD & Stathopoulos C (2015) A glyS T-box riboswitch with species-specific structural features responding to both proteinogenic and nonproteinogenic tRNAGly isoacceptors. *RNA* 21(10):1790-1806.
- Aspedon A, Palmer K & Whiteley M (2006) Microarray analysis of the osmotic stress response in *Pseudomonas aeruginosa*. *Journal of Bacteriology* 188(7):2721-2725.
- Auld DS, Coassin PA, Coussens NP, Hensley P, Klumpp-Thomas C, Michael S, Sittampalam GS, Trask OJ, Wagner BK & Weidner JR (2020) Microplate selection and recommended practices in high-throughput screening and quantitative biology. *Assay Guidance Manual [Internet]*.
- Babitzke P & Gollnick P (2001) Posttranscription initiation control of tryptophan metabolism in *Bacillus subtilis* by the trp RNA-binding attenuation protein (TRAP), anti-TRAP, and RNA structure. *Journal of bacteriology* 183(20):5795-5802.
- Baker JL, Sudarsan N, Weinberg Z, Roth A, Stockbridge RB & Breaker RR (2012) Widespread genetic switches and toxicity resistance proteins for fluoride. *Science* 335(6065):233-235.
- Balleza E, Lopez-Bojorquez LN, Martínez-Antonio A, Resendis-Antonio O, Lozada-Chávez I, Balderas-Martínez YI, Encarnación S & Collado-Vides J (2008) Regulation by transcription factors in bacteria: beyond description. *FEMS microbiology reviews* 33(1):133-151.
- Banerjee S, Chalissery J, Bandey I & Sen R (2006) Rho-dependent transcription termination: more questions than answers. *Journal of microbiology (Seoul, Korea)* 44(1):11.
- Barrick JE, Corbino KA, Winkler WC, Nahvi A, Mandal M, Collins J, Lee M, Roth A, Sudarsan N & Jona I (2004) New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proceedings of the National Academy of Sciences* 101(17):6421-6426.
- Bastet L, Chauvier A, Singh N, Lussier A, Lamontagne AM, Prévost K, Massé E, Wade JT & Lafontaine DA (2017) Translational control and Rho-dependent transcription termination are intimately linked in riboswitch regulation. *Nucleic Acids Res* 45(12):7474-7486.
- Baudart J & Paniel N (2014) Sources et devenir des micro-organismes pathogènes dans les environnements aquatiques. *Revue Francophone des Laboratoires* 2014(459):29-39.

- Bayat P, Nosrati R, Alibolandi M, Rafatpanah H, Abnous K, Khedri M & Ramezani M (2018) SELEX methods on the road to protein targeting with nucleic acid aptamers. *Biochimie* 154:132-155.
- Bazan J, Calkosiński I & Gamian A (2012) Phage display--a powerful technique for immunotherapy: 1. Introduction and potential of therapeutic applications. *Hum Vaccin Immunother* 8(12):1817-1828.
- Bazire A, Dheilly A, Diab F, Morin D, Jebbar M, Haras D & Dufour A (2005) Osmotic stress and phosphate limitation alter production of cell-to-cell signal molecules and rhamnolipid biosurfactant by *Pseudomonas aeruginosa*. *FEMS Microbiol. Lett.* 253(1):125-131.
- Becher A & Schweizer HP (2000) Integration-proficient *Pseudomonas aeruginosa* vectors for isolation of single-copy chromosomal lacZ and lux gene fusions. *Biotechniques* 29(5):948-952.
- Bédard AV, Hien EDM & Lafontaine DA (2020) Riboswitch regulation mechanisms: RNA, metabolites and regulatory proteins. *Biochim Biophys Acta Gene Regul Mech* 1863(3):194501.
- Benard L, Philippe C, Ehresmann B, Ehresmann C & Portier C (1996) Pseudoknot and translational control in the expression of the S15 ribosomal protein. *Biochimie* 78(7):568-576.
- Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL & Bignell HR (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *nature* 456(7218):53-59.
- Bösl M & Kersten H (1991) A novel RNA product of the tyrT operon of *Escherichia coli*. *Nucleic Acids Res* 19(21):5863-5870.
- Bossi L, Schwartz A, Guillemardet B, Boudvillain M & Figueroa-Bossi N (2012) A role for Rho-dependent polarity in gene regulation by a noncoding small RNA. *Genes Dev* 26(16):1864-1873.
- Boussebayle A, Groher F & Suess B (2019) RNA-based capture-SELEX for the selection of small molecule-binding aptamers. *Methods* 161:10-15.
- Boutet E, Djerroud S & Perreault J (2022) Small RNAs beyond Model Organisms: Have We Only Scratched the Surface? *Int. J. Mol. Sci.* 23(8):4448.
- Brandenburg KS, Rodriguez KJ, McAnulty JF, Murphy CJ, Abbott NL, Schurr MJ & Czuprynski CJ (2013) Tryptophan Inhibits Biofilm Formation by *Pseudomonas aeruginosa*. *Antimicrob. Agents Chemother.* 57(4):1921-1925.
- Breaker R (2012) New insight on the response of bacteria to fluoride. *Caries research* 46(1):78-81.
- Breaker RR (2011) Prospects for riboswitch discovery and analysis. *Molecular cell* 43(6):867-879.
- Brett PJ, DeShazer D & Woods DE (1998) Note: *Burkholderia thailandensis* sp. nov., a *Burkholderia pseudomallei*-like species. *International Journal of Systematic and Evolutionary Microbiology* 48(1):317-320.
- Bruno JG & Kiel JL (2002) Use of magnetic beads in selection and detection of biotoxin aptamers by electrochemiluminescence and enzymatic methods. *Biotechniques* 32(1):178-183.
- Buell CR, Joardar V, Lindeberg M, Selengut J, Paulsen IT, Gwinn ML, Dodson RJ, Deboy RT, Durkin AS & Kolonay JF (2003) The complete genome sequence of the *Arabidopsis* and

- tomato pathogen *Pseudomonas syringae* pv. tomato DC3000. *Proceedings of the National Academy of Sciences* 100(18):10181-10186.
- Caraballo M (2018) *Oligos pool : Design, Synthesis, and Research Applications*. <https://www.genscript.com/gsfiles/techfiles/Oligo-Pools-design-synthesis-and-research-applications-slides.pdf> (Consulté le 24 novembre 2022)
- Carlioni S, Macchi R, Sattin S, Ferrara S & Bertoni G (2017) The small RNA ReaL: a novel regulatory element embedded in the *Pseudomonas aeruginosa* quorum sensing networks. *Environ Microbiol* 19(10):4220-4237.
- Caron MP, Bastet L, Lussier A, Simoneau-Roy M, Massé E & Lafontaine DA (2012) Dual-acting riboswitch control of translation initiation and mRNA decay. *Proc Natl Acad Sci U S A* 109(50):E3444-3453.
- Carpousis AJ (2007) The RNA degradosome of *Escherichia coli*: an mRNA-degrading machine assembled on RNase E. *Annu Rev Microbiol* 61:71-87.
- Chakravarty S & Massé E (2019) RNA-dependent regulation of virulence in pathogenic bacteria. *Frontiers in Cellular and Infection Microbiology* 9:337.
- Chan C & Landick R (1989) The *Salmonella typhimurium* his operon leader region contains an RNA hairpin-dependent transcription pause site: mechanistic implications of the effect on pausing of altered RNA hairpins. *Journal of Biological Chemistry* 264(34):20796-20804.
- Chan V, Dreolini LF, Flintoff KA, Lloyd SJ & Mattenley AA (2002) The effect of increasing plasmid size on transformation efficiency in *Escherichia coli*. *Journal of Experimental Microbiology and Immunology* 2:207-223.
- Ciampi MS (2006) Rho-dependent terminators and transcription termination. *Microbiology* 152(9):2515-2528.
- Dale J & Park S (2010) *Molecular genetics of bacteria*. Wiley-Blackwell, Chichester, West Sussex, England, 5th ed.
- Dar D, Shamir M, Mellin J, Koutero M, Stern-Ginossar N, Cossart P & Sorek R (2016) Term-seq reveals abundant ribo-regulation of antibiotics resistance in bacteria. *Science* 352(6282):aad9822.
- Darmostuk M, Rimpelova S, Gbelcova H & Ruml T (2015) Current approaches in SELEX: An update to aptamer selection technology. *Biotechnology Advances* 33(6, Part 2):1141-1161.
- Delvillani F, Sciandrone B, Peano C, Petiti L, Berens C, Georgi C, Ferrara S, Bertoni G, Pasini ME & Dehò G (2014) Tet-Trap, a genetic approach to the identification of bacterial RNA thermometers: application to *Pseudomonas aeruginosa*. *RNA* 20(12):1963-1976.
- Desgranges E, Marzi S, Moreau K, Romby P & Caldelari I (2019) Noncoding RNA. *Microbiology Spectrum* 7(2).
- Desnoyers G, Morissette A, Prévost K & Massé E (2009) Small RNA-induced differential degradation of the polycistronic mRNA iscRSUA. *The EMBO journal* 28(11):1551-1561.
- Devinck A, Boutet E, Ouellet J, Rouag R, Sellamuthu B & Perreault J (2022) Shifted Reverse PAGE: a novel approach based on structure switching for the discovery of riboswitches and aptamers. *bioRxiv* 10.1101/2022.07.26.501614:2022.2007.2026.501614.

- Diwa A, Bricker AL, Jain C & Belasco JG (2000) An evolutionarily conserved RNA stem-loop functions as a sensor that directs feedback regulation of RNase E gene expression. *Genes & development* 14(10):1249-1260.
- Eberl L (2006) Quorum sensing in the genus Burkholderia. *International Journal of Medical Microbiology* 296(2):103-110.
- Eckert I & Weinberg Z (2020) Discovery of 20 novel ribosomal leader candidates in bacteria and archaea. *BMC microbiology* 20(1):1-12.
- Edwards TE & Ferré-D'Amaré AR (2006) Crystal Structures of the Thi-Box Riboswitch Bound to Thiamine Pyrophosphate Analogs Reveal Adaptive RNA-Small Molecule Recognition. *Structure* 14(9):1459-1468.
- El-Sharoud WM & Niven GW (2005) The activity of ribosome modulation factor during growth of Escherichia coli under acidic conditions. *Archives of Microbiology* 184(1):18-24.
- El-Sharoud WM & Niven GW (2007) The influence of ribosome modulation factor on the survival of stationary-phase Escherichia coli during acid stress. *Microbiology (Reading)* 153(Pt 1):247-253.
- Elbing K & Brent R (2002) Media preparation and bacteriological tools. *Current Protocols in Molecular Biology* 59(1):1.1. 1-1.1. 7.
- Ellington AD & Szostak JW (1990) In vitro selection of RNA molecules that bind specific ligands. *nature* 346(6287):818-822.
- Espelund M, Stacy R & Jakobsen K (1990) A simple method for generating single-stranded DNA probes labeled to high activities. *Nucleic acids research* 18(20):6157.
- Falcone M, Ferrara S, Rossi E, Johansen HK, Molin S & Bertoni G (2018) The small RNA ErsA of Pseudomonas aeruginosa contributes to biofilm development and motility through post-transcriptional modulation of AmrZ. *Frontiers in microbiology* 9:238.
- Ferrara S, Carloni S, Fulco R, Falcone M, Macchi R & Bertoni G (2015) Post-transcriptional regulation of the virulence-associated enzyme A Ig C by the σ^{22} -dependent small RNA ErsA of Pseudomonas aeruginosa. *Environmental Microbiology* 17(1):199-214.
- Ferrara S, Carrubba R, Santoro S & Bertoni G (2021) The small RNA ErsA impacts the anaerobic metabolism of Pseudomonas aeruginosa through post-transcriptional modulation of the master regulator Anr. *Frontiers in microbiology* :2365.
- Finnegan S & Percival SL (2015) EDTA: an antimicrobial and antibiofilm agent for use in wound care. *Advances in wound care* 4(7):415-421.
- Francis VI, Stevenson EC & Porter SL (2017) Two-component systems required for virulence in Pseudomonas aeruginosa. *FEMS Microbiol. Lett.* 364(11).
- French S, Mangat C, Bharat A, Côté J-P, Mori H & Brown ED (2016) A robust platform for chemical genomics in bacterial systems. *Molecular biology of the cell* 27(6):1015-1025.
- Fu L, Niu B, Zhu Z, Wu S & Li W (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28(23):3150-3152.
- Fu Y, Deiorio-Haggart K, Anthony J & Meyer MM (2013) Most RNAs regulating ribosomal protein biosynthesis in Escherichia coli are narrowly distributed to Gammaproteobacteria. *Nucleic acids research* 41(6):3491-3503.

- Furukawa K, Ramesh A, Zhou Z, Weinberg Z, Vallery T, Winkler WC & Breaker RR (2015) Bacterial riboswitches cooperatively bind Ni²⁺ or Co²⁺ ions and control expression of heavy metal transporters. *Molecular cell* 57(6):1088-1098.
- Gavini N & Pulakat L (1991) Role of translation of the pheA leader peptide coding region in attenuation regulation of the Escherichia coli pheA gene. *Journal of bacteriology* 173(15):4904-4907.
- Gelfand MS, Mironov AA, Jomantas J, Kozlov YI & Perumov DA (1999) A conserved RNA structure element involved in the regulation of bacterial riboflavin synthesis genes. *Trends Genet* 15(11):439-442.
- Giuliodori AM, Di Pietro F, Marzi S, Masquida B, Wagner R, Romby P, Gualerzi CO & Pon CL (2010) The cspA mRNA is a thermosensor that modulates translation of the cold-shock protein CspA. *Molecular cell* 37(1):21-33.
- Goodman AL, Kulasekara B, Rietsch A, Boyd D, Smith RS & Lory S (2004) A signaling network reciprocally regulates genes associated with acute infection and chronic persistence in Pseudomonas aeruginosa. *Developmental cell* 7(5):745-754.
- Grainger DC, Overton TW, Reppas N, Wade JT, Tamai E, Hobman JL, Constantinidou C, Struhl K, Church G & Busby SJ (2004) Genomic studies with Escherichia coli MelR protein: applications of chromatin immunoprecipitation and microarrays. *Journal of bacteriology* 186(20):6938-6943.
- Grandoni J, Fulmer S, Brizzio V, Zahler S & Calvo J (1993) Regions of the Bacillus subtilis ilv-leu operon involved in regulation by leucine. *Journal of bacteriology* 175(23):7581-7593.
- Greenlee EB, Stav S, Atilho RM, Brewer KI, Harris KA, Malkowski SN, Mirihana Arachchilage G, Perkins KR, Sherlock ME & Breaker RR (2018) Challenges of ligand identification for the second wave of orphan riboswitch candidates. *RNA Biol* 15(3):377-390.
- Griffiths-Jones S, Bateman A, Marshall M, Khanna A & Eddy SR (2003) Rfam: an RNA family database. *Nucleic Acids Research* 31(1):439-441.
- Groisman E, Cromie M, Shi Y & Latifi T (2006) A Mg²⁺-responding RNA that controls the expression of a Mg²⁺ transporter. *Cold Spring Harbor symposia on quantitative biology*. Cold Spring Harbor Laboratory Press, p 251-258.
- Groisman EA (2001) The pleiotropic two-component regulatory system PhoP-PhoQ. *Journal of bacteriology* 183(6):1835-1842.
- Grosso-Becerra MV, Croda-García G, Merino E, Servín-González L, Mojica-Espinosa R & Soberón-Chávez G (2014) Regulation of Pseudomonas aeruginosa virulence factors by two novel RNA thermometers. *Proceedings of the National Academy of Sciences* 111(43):15562-15567.
- Grundy FJ & Henkin TM (1993) tRNA as a positive regulator of transcription antitermination in B. subtilis. *Cell* 74(3):475-482.
- Grundy FJ, Rollins SM & Henkin TM (1994) Interaction between the acceptor end of tRNA and the T box stimulates antitermination in the Bacillus subtilis tyrS gene: a new role for the discriminator base. *Journal of bacteriology* 176(15):4518-4526.
- Güvener ZT & Harwood CS (2007) Subcellular location characteristics of the Pseudomonas aeruginosa GGDEF protein, WspR, indicate that it produces cyclic-di-GMP in response to growth on surfaces. *Mol. Microbiol.* 66(6):1459-1473.

- Haller A, Rieder U, Aigner M, Blanchard SC & Micura R (2011) Conformational capture of the SAM-II riboswitch. *Nat. Chem. Biol.* 7(6):393-400.
- Haraga A, West TE, Brittnacher MJ, Skerrett SJ & Miller SI (2008) Burkholderia thailandensis as a model system for the study of the virulence-associated type III secretion system of Burkholderia pseudomallei. *Infection and immunity* 76(11):5402-5411.
- Hart CM & Roberts J (1991) Rho-dependent transcription termination. Characterization of the requirement for cytidine in the nascent transcript. *Journal of Biological Chemistry* 266(35):24140-24148.
- Haurlyuk V, Atkinson GC, Murakami KS, Tenson T & Gerdes K (2015) Recent functional insights into the role of (p) ppGpp in bacterial physiology. *Nature Reviews Microbiology* 13(5):298-309.
- He M & Taussig MJ (2002) Ribosome display: cell-free protein display technology. *Brief Funct Genomic Proteomic* 1(2):204-212.
- Hellman LM & Fried MG (2007) Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. *Nature protocols* 2(8):1849-1861.
- Henkin TM, Glass BL & Grundy FJ (1992) Analysis of the Bacillus subtilis tyrS gene: conservation of a regulatory sequence in multiple tRNA synthetase genes. *J Bacteriol* 174(4):1299-1306.
- Henkin TM & Peters JE (2020) *Snyder & Champness molecular genetics of bacteria*. ASM Press; Wiley, Washington, DC; Hoboken, NJ, Fifth edition.
- Hennecke H (1990) Regulation of bacterial gene expression by metal–protein complexes. *Mol. Microbiol.* 4(10):1621-1628.
- Heppell B, Blouin S, Dussault A-M, Mulhbachter J, Ennifar E, Penedo JC & Lafontaine DA (2011) Molecular insights into the ligand-controlled organization of the SAM-I riboswitch. *Nature chemical biology* 7(6):384-392.
- Hess WR & Marchfelder A (2012) *Regulatory RNAs in prokaryotes*. Springer, Vienna ;(Consulté e Date Accessed)
- Hickman JW, Tifrea DF & Harwood CS (2005) A chemosensory system that regulates biofilm formation through modulation of cyclic diguanylate levels. *Proceedings of the National Academy of Sciences* 102(40):14422-14427.
- Hollands K, Proshkin S, Sklyarova S, Epshtein V, Mironov A, Nudler E & Groisman EA (2012) Riboswitch control of Rho-dependent transcription termination. *Proceedings of the National Academy of Sciences* 109(14):5376-5381.
- Hollis R, Lagido C, Pettitt J, Porter AJR, Killham K, Paton GI & Glover LA (2001) Toxicity of the bacterial luciferase substrate, n-decyl aldehyde, to Saccharomyces cerevisiae and Caenorhabditis elegans. *Febs Letters* 506(2):140-142.
- Howe JA, Xiao L, Fischmann TO, Wang H, Tang H, Villafania A, Zhang R, Barbieri CM & Roemer T (2016) Atomic resolution mechanistic studies of ribocil: A highly selective unnatural ligand mimic of the E. coli FMN riboswitch. *RNA biology* 13(10):946-954.
- Huang L, Serganov A & Patel DJ (2010) Structural insights into ligand recognition by a sensing domain of the cooperative glycine riboswitch. *Molecular cell* 40(5):774-786.

- Illumina (2020) *How short inserts affect sequencing performance?*, <https://support.illumina.com/bulletins/2020/12/how-short-inserts-affect-sequencing-performance.html> (Consulté le 10 novembre 2022)
- Irvine D, Tuerk C & Gold L (1991) Selection: Systematic evolution of ligands by exponential enrichment with integrated optimization by non-linear analysis. *Journal of molecular biology* 222(3):739-761.
- Izutsu K, Wada A & Wada C (2001) Expression of ribosome modulation factor (RMF) in *Escherichia coli* requires ppGpp. *Genes Cells* 6(8):665-676.
- Jensen KF (1989) Regulation of *Salmonella typhimurium* pyr gene expression: effect of changing both purine and pyrimidine nucleotide pools. *Microbiology* 135(4):805-815.
- Jing M & Bowser MT (2011) Isolation of DNA aptamers using micro free flow electrophoresis. *Lab Chip* 11(21):3703-3709.
- Johansen LE, Nygaard P, Lassen C, Agersø Y & Saxild HH (2003) Definition of a second *Bacillus subtilis* pur regulon comprising the pur and xpt-pbuX operons plus pbuG, nupG (yxjA), and pbuE (ydhL). *Journal of bacteriology* 185(17):5200-5209.
- Johansson J, Mandin P, Renzoni A, Chiaruttini C, Springer M & Cossart P (2002) An RNA thermosensor controls expression of virulence genes in *Listeria monocytogenes*. *Cell* 110(5):551-561.
- Johnson DA, Tetu SG, Phillippy K, Chen J, Ren Q & Paulsen IT (2008) High-throughput phenotypic characterization of *Pseudomonas aeruginosa* membrane transport genes. *PLoS genetics* 4(10):e1000211.
- Jones AM & Wildermuth MC (2011) The Phytopathogen *Pseudomonas syringae* pv. tomato DC3000 Has Three High-Affinity Iron-Scavenging Systems Functional under Iron Limitation Conditions but Dispensable for Pathogenesis. *Journal of Bacteriology* 193(11):2767-2775.
- Jurado-Martín I, Sainz-Mejías M & McClean S (2021) *Pseudomonas aeruginosa*: An audacious pathogen with an adaptable arsenal of virulence factors. *Int. J. Mol. Sci.* 22(6):3128.
- Kaiser JC & Heinrichs DE (2018) Branching Out: Alterations in Bacterial Physiology and Virulence Due to Branched-Chain Amino Acid Deprivation. *MBio* 9(5):e01188-01118.
- Kalvari I, Nawrocki EP, Ontiveros-Palacios N, Argasinska J, Lamkiewicz K, Marz M, Griffiths-Jones S, Toffano-Nioche C, Gautheret D & Weinberg Z (2021) Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Research* 49(D1):D192-D200.
- Kanehisa M (2019) Toward understanding the origin and evolution of cellular organisms. *Protein Science* 28(11):1947-1951.
- Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M & Tanabe M (2020) KEGG: integrating viruses and cellular organisms. *Nucleic Acids Research* 49(D1):D545-D551.
- Kanehisa M & Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* 28(1):27-30.
- Kavita K & Breaker RR (2022) Discovering riboswitches: the past and the future. *Trends in Biochemical Sciences*.
- Keseler IM, Gama-Castro S, Mackie A, Billington R, Bonavides-Martínez C, Caspi R, Kothari A, Krummenacker M, Midford PE, Muñiz-Rascado L, Ong WK, Paley S, Santos-Zavaleta A,

- Subhraveti P, Tierrafría VH, Wolfe AJ, Collado-Vides J, Paulsen IT & Karp PD (2021) The EcoCyc Database in 2021. *Front Microbiol* 12:711077.
- Kidder BL, Hu G & Zhao K (2011) ChIP-Seq: technical considerations for obtaining high-quality data. *Nat Immunol* 12(10):918-922.
- Kil YV, Mironovi VN, Gorishin IY, Kreneva RA & Perumov DA (1992) Riboflavin operon of *Bacillus subtilis*: unusual symmetric arrangement of the regulatory region. *Molecular and General Genetics MGG* 233(3):483-486.
- Kim HS, Schell MA, Yu Y, Ulrich RL, Sarria SH, Nierman WC & DeShazer D (2005) Bacterial genome adaptation to niches: Divergence of the potential virulence genes in three *Burkholderia* species of different survival strategies. *BMC Genomics* 6(1):174.
- Kim JN & Breaker RR (2008) Purine sensing by riboswitches. *Biol Cell* 100(1):1-11.
- Kim PB, Nelson JW & Breaker RR (2015) An ancient riboswitch class in bacteria regulates purine biosynthesis and one-carbon metabolism. *Molecular cell* 57(2):317-328.
- Klinkert B, Cimdins A, Gaubig LC, Roßmanith J, Aschke-Sonnenborn U & Narberhaus F (2012) Thermogenetic tools to monitor temperature-dependent gene expression in bacteria. *Journal of biotechnology* 160(1-2):55-63.
- Kohn H & Widger W (2005) The molecular basis for the mode of action of bicyclomycin. *Current Drug Targets-Infectious Disorders* 5(3):273-295.
- Korniakova V, Devinck A, Groleau MC, Déziel E & Perreault J (2021) Fluoride-Controlled Riboswitch-Based Dampening of Gene Expression for Cloning Potent Promoters. *Front Genet* 12:591543.
- Krajewski SS, Nagel M & Narberhaus F (2013) Short ROSE-like RNA thermometers control IbpA synthesis in *Pseudomonas* species. *PLoS ONE* 8(5):e65168.
- Kreuzer KD & Henkin TM (2018) The T-box riboswitch: tRNA as an effector to modulate gene regulation. *Microbiology spectrum* 6(4):6.4. 22.
- Kwenda S, Gorshkov V, Ramesh AM, Naidoo S, Rubagotti E, Birch PRJ & Moleleki LN (2016) Discovery and profiling of small RNAs responsive to stress conditions in the plant pathogen *Pectobacterium atrosepticum*. *BMC Genomics* 17(1):47.
- Kwon M & Strobel SA (2008) Chemical basis of glycine riboswitch cooperativity. *Rna* 14(1):25-34.
- Lafontaine DA, Norman DG & Lilley DM (2001) Structure, folding and activity of the VS ribozyme: importance of the 2-3-6 helical junction. *The EMBO Journal* 20(6):1415-1424.
- Lalaouna D, Baude J, Wu Z, Tomasini A, Chicher J, Marzi S, Vandenesch F, Romby P, Caldelari I & Moreau K (2019a) RsaC sRNA modulates the oxidative stress response of *Staphylococcus aureus* during manganese starvation. *Nucleic Acids Research* 47(18):9871-9887.
- Lalaouna D, Eyraud A, Devinck A, Prévost K & Massé E (2019b) GcvB small RNA uses two distinct seed regions to regulate an extensive targetome. *Mol Microbiol* 111(2):473-486.
- Lalaouna D, Simoneau-Roy M, Lafontaine D & Massé E (2013) Regulatory RNAs and target mRNA decay in prokaryotes. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1829(6):742-747.
- Lambert P (2002) Mechanisms of antibiotic resistance in *Pseudomonas aeruginosa*. *Journal of the royal society of medicine* 95(Suppl 41):22.

- Larsen B, Gesteland RF & Atkins JF (1997) Structural probing and mutagenic analysis of the stem-loop required for Escherichia coli dnaX ribosomal frameshifting: programmed efficiency of 50%. *Journal of molecular biology* 271(1):47-60.
- Lee E-J, Luo J, Wilson JM & Shi H (2013) Analyzing the cancer methylome through targeted bisulfite sequencing. *Cancer Letters* 340(2):171-178.
- Lee ER, Baker JL, Weinberg Z, Sudarsan N & Breaker RR (2010) An allosteric self-splicing ribozyme triggered by a bacterial second messenger. *science* 329(5993):845-848.
- Li S, Hwang XY, Stav S & Breaker RR (2016) The yjdF riboswitch candidate regulates gene expression by binding diverse azaaromatic compounds. *Rna* 22(4):530-541.
- Li W & Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22(13):1658-1659.
- Liberati NT, Urbach JM, Miyata S, Lee DG, Drenkard E, Wu G, Villanueva J, Wei T & Ausubel FM (2006) An ordered, nonredundant library of Pseudomonas aeruginosa strain PA14 transposon insertion mutants. *Proceedings of the National Academy of Sciences* 103(8):2833-2838.
- Lien H-Y, Yu C-H, Liou C-M & Wu WF (2009) Regulation of clpQ+ Y+ (hsIV+ U+) gene expression in Escherichia coli. *The Open Microbiology Journal* 3:29.
- Loh E, Righetti F, Eichner H, Twittenhoff C & Narberhaus F (2018) RNA thermometers in bacterial pathogens. *Microbiology spectrum* 6(2):6.2. 13.
- Lorenz C, von Pelchrzim F & Schroeder R (2006) Genomic systematic evolution of ligands by exponential enrichment (Genomic SELEX) for the identification of protein-binding RNAs independent of their expression levels. *Nature Protocols* 1(5):2204-2212.
- Lorenz R, Bernhart SH, Höner zu Siederdisen C, Tafer H, Flamm C, Stadler PF & Hofacker IL (2011) ViennaRNA Package 2.0. *Algorithms for molecular biology* 6:1-14.
- Lotz TS & Suess B (2018) Small-molecule-binding riboswitches. *Microbiology Spectrum* 6(4):6.4. 26.
- Loui C, Chang AC & Lu S (2009) Role of the ArcAB two-component system in the resistance of Escherichia coli to reactive oxygen stress. *BMC microbiology* 9(1):1-14.
- Lyczak JB, Cannon CL & Pier GB (2000) Establishment of Pseudomonas aeruginosa infection: lessons from a versatile opportunist. *Microbes and infection* 2(9):1051-1060.
- Lynn SP, Burton WS, Donohue TJ, Gould RM, Gumpert RI & Gardner JF (1987) Specificity of the attenuation response of the threonine operon of Escherichia coli is determined by the threonine and isoleucine codons in the leader transcript. *Journal of molecular biology* 194(1):59-69.
- Lyu C, Khan IM & Wang Z (2021) Capture-SELEX for aptamer selection: A short review. *Talanta* 229:122274.
- Machtel P, Bałowska-Żywicka K & Żywicki M (2016) Emerging applications of riboswitches—from antibacterial targets to molecular tools. *Journal of applied genetics* 57(4):531-541.
- Mandal M, Boese B, Barrick JE, Winkler WC & Breaker RR (2003) Riboswitches Control Fundamental Biochemical Pathways in Bacillus subtilis and Other Bacteria. *Cell* 113(5):577-586.
- Mandal M & Breaker RR (2004a) Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nature structural & molecular biology* 11(1):29-35.

- Mandal M & Breaker RR (2004b) Gene regulation by riboswitches. *Nature Reviews Molecular Cell Biology* 5(6):451-463.
- Mandal M, Lee M, Barrick JE, Weinberg Z, Emilsson GM, Ruzzo WL & Breaker RR (2004) A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* 306(5694):275-279.
- Manna S, Truong J & Hammond MC (2021) Guanidine Biosensors Enable Comparison of Cellular Turn-on Kinetics of Riboswitch-Based Biosensor and Reporter. *ACS Synthetic Biology* 10(3):566-578.
- Martinez-Antonio A & Collado-Vides J (2003) Identifying global regulators in transcriptional regulatory networks in bacteria. *Current opinion in microbiology* 6(5):482-489.
- Massé E & Gottesman S (2002) A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc Natl Acad Sci U S A* 99(7):4620-4625.
- Massé E, Vanderpool CK & Gottesman S (2005) Effect of RyhB small RNA on global iron use in *Escherichia coli*. *Journal of bacteriology* 187(20):6962-6971.
- Matsunaga J, Simons EL & Simons RW (1996) RNase III autoregulation: structure and function of rncO, the posttranscriptional "operator". *RNA* 2(12):1228-1240.
- McCown PJ, Corbino KA, Stav S, Sherlock ME & Breaker RR (2017) Riboswitch diversity and distribution. *Rna* 23(7):995-1011.
- McKee RW, Harvest CK & Tamayo R (2018) Cyclic Diguanylate Regulates Virulence Factor Genes via Multiple Riboswitches in *Clostridium difficile*. *mSphere* 3(5):e00423-00418.
- Mendonça SD & Bowser MT (2004) In vitro evolution of functional DNA using capillary electrophoresis. *J Am Chem Soc* 126(1):20-21.
- Mérens A, Delacour H, Plésiat P, Cavallo J-D & Jeannot K (2011) *Pseudomonas aeruginosa* et résistance aux antibiotiques. *Revue Francophone des Laboratoires* 2011(435):49-62.
- Meyer Kate D, Saletore Y, Zumbo P, Elemento O, Mason Christopher E & Jaffrey Samie R (2012) Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3' UTRs and near Stop Codons. *Cell* 149(7):1635-1646.
- Meyer MM (2018) rRNA mimicry in RNA regulation of gene expression. *Microbiology Spectrum* 6(2):6.2. 03.
- Meyer MM, Hammond MC, Salinas Y, Roth A, Sudarsan N & Breaker RR (2011) Challenges of ligand identification for riboswitch candidates. *RNA biology* 8(1):5-10.
- Mills E, Pultz IS, Kulasekara HD & Miller SI (2011) The bacterial second messenger c-di-GMP: mechanisms of signalling. *Cellular microbiology* 13(8):1122-1129.
- Miranda-Ríos J, Navarro M & Soberón M (2001) A conserved RNA structure (thi box) is involved in regulation of thiamin biosynthetic gene expression in bacteria. *Proceedings of the National Academy of Sciences* 98(17):9736-9741.
- Mironov AS, Gusarov I, Rafikov R, Lopez LE, Shatalin K, Kreneva RA, Perumov DA & Nudler E (2002) Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell* 111(5):747-756.
- Mitra P, Ghosh G, Hafeezunnisa M & Sen R (2017) Rho protein: roles and mechanisms. *Annual review of microbiology* 71:687-709.

- Mix H, Lobanov AV & Gladyshev VN (2007) SECIS elements in the coding regions of selenoprotein transcripts are functional in higher eukaryotes. *Nucleic acids research* 35(2):414-423.
- Mizuno T, Chou M-Y & Inouye M (1984) A unique mechanism regulating gene expression: translational inhibition by a complementary RNA transcript (micRNA). *Proceedings of the National Academy of Sciences* 81(7):1966-1970.
- Mohsen MG, Midy MK, Balaji A & Breaker Ronald R (2023) Exploiting natural riboswitches for aptamer engineering and validation. *Nucleic Acids Research* 51(2):966-981.
- Munyati-Othman N, Appasamy SD, Damiri N, Emrizal R, Alipiah NM, Ramlan EI & Firdaus-Raih M (2021) Regulation of Glycine Cleavage and Detoxification by a Highly Conserved Glycine Riboswitch in Burkholderia spp. *Current Microbiology* 78(8):2943-2955.
- Naghdi MR, Smail K, Wang JX, Wade F, Breaker RR & Perreault J (2017) Search for 5'-leader regulatory RNA structures based on gene annotation aided by the RiboGap database. *Methods* 117:3-13.
- Nahvi A, Barrick JE & Breaker RR (2004) Coenzyme B12 riboswitches are widespread genetic control elements in prokaryotes. *Nucleic acids research* 32(1):143-150.
- Nahvi A, Sudarsan N, Ebert MS, Zou X, Brown KL & Breaker RR (2002) Genetic Control by a Metabolite Binding mRNA. *Chemistry & Biology* 9(9):1043-1049.
- Narberhaus F, Waldminghaus T & Chowdhury S (2006) RNA thermometers. *FEMS microbiology reviews* 30(1):3-16.
- Naville M & Gautheret D (2010) Premature terminator analysis sheds light on a hidden world of bacterial transcriptional attenuation. *Genome Biology* 11(9):R97.
- Nelson CE, Huang W, Brewer LK, Nguyen AT, Kane MA, Wilks A & Oglesby-Sherrouse AG (2019) Proteomic analysis of the Pseudomonas aeruginosa iron starvation response reveals PrrF small regulatory RNA-dependent iron regulation of twitching motility, amino acid metabolism, and zinc homeostasis proteins. *Journal of bacteriology* 201(12):e00754-00718.
- Nelson JW, Atilho RM, Sherlock ME, Stockbridge RB & Breaker RR (2017) Metabolism of Free Guanidine in Bacteria Is Regulated by a Widespread Riboswitch Class. *Mol Cell* 65(2):220-230.
- Nguyen D, Joshi-Datar A, Lepine F, Bauerle E, Olakanmi O, Beer K, McKay G, Siehnel R, Schafhauser J, Wang Y, Britigan BE & Singh PK (2011) Active starvation responses mediate antibiotic tolerance in biofilms and nutrient-limited bacteria. *Science* 334(6058):982-986.
- Niven GW (2004) Ribosome modulation factor protects Escherichia coli during heat stress, but this may not be dependent on ribosome dimerisation. *Archives of microbiology* 182(1):60-66.
- Nocker A, Hausherr T, Balsiger S, Krstulovic N-P, Hennecke H & Narberhaus F (2001) A mRNA-based thermosensor controls expression of rhizobial heat shock genes. *Nucleic acids research* 29(23):4800-4807.
- Nomura M, Yates JL, Dean D & Post LE (1980) Feedback regulation of ribosomal protein gene expression in Escherichia coli: structural homology of ribosomal RNA and ribosomal protein mRNA. *Proceedings of the National Academy of Sciences* 77(12):7084-7088.

- Palmer KL, Aye LM & Whiteley M (2007) Nutritional cues control *Pseudomonas aeruginosa* multicellular behavior in cystic fibrosis sputum. *J Bacteriol* 189(22):8079-8087.
- Pavlova N, Kaloudas D & Penchovsky R (2019) Riboswitch distribution, structure, and function in bacteria. *Gene* 708:38-48.
- Pedrolli D, Langer S, Hobl B, Schwarz J, Hashimoto M & Mack M (2015) The ribB FMN riboswitch from *Escherichia coli* operates at the transcriptional and translational level and regulates riboflavin biosynthesis. *The FEBS journal* 282(16):3230-3242.
- Pingoud A & Jeltsch A (2001) Structure and function of type II restriction endonucleases. *Nucleic Acids Res* 29(18):3705-3727.
- Pita T, Feliciano JR & Leitão JH (2018) Small Noncoding Regulatory RNAs from *Pseudomonas aeruginosa* and *Burkholderia cepacia* Complex. *Int. J. Mol. Sci.* 19(12):3759.
- Polikanov YS, Blaha GM & Steitz TA (2012) How Hibernation Factors RMF, HPF, and YfiA Turn Off Protein Synthesis. *Science* 336(6083):915-918.
- Potrykus K, Murphy H, Chen X, Epstein JA & Cashel M (2010) Imprecise transcription termination within *Escherichia coli* greA leader gives rise to an array of short transcripts, GraL. *Nucleic acids research* 38(5):1636-1651.
- Price IR, Gaballa A, Ding F, Helmann JD & Ke A (2015) Mn²⁺-sensing mechanisms of yybP-ykoY orphan riboswitches. *Molecular cell* 57(6):1110-1123.
- Pusic P, Sonnleitner E & Bläsi U (2021) Specific and global RNA regulators in *Pseudomonas aeruginosa*. *Int. J. Mol. Sci.* 22(16):8632.
- Ramesh A & Winkler WC (2010) Magnesium-sensing riboswitches in bacteria. *RNA biology* 7(1):77-83.
- Regulski EE & Breaker RR (2008) In-line probing analysis of riboswitches. *post-transcriptional gene regulation*, Springer. p 53-67.
- Regulski EE, Moy RH, Weinberg Z, Barrick JE, Yao Z, Ruzzo WL & Breaker RR (2008) A widespread riboswitch candidate that controls bacterial genes involved in molybdenum cofactor and tungsten cofactor metabolism. *Mol. Microbiol.* 68(4):918-932.
- Reiss CW & Strobel SA (2017) Structural basis for ligand binding to the guanidine-II riboswitch. *RNA* 23(9):1338-1343.
- Rentmeister A, Mayer G, Kuhn N & Famulok M (2007) Conformational changes in the expression domain of the *Escherichia coli* thiM riboswitch. *Nucleic acids research* 35(11):3713-3722.
- Renzoni A, Cossart P & Dramsi S (1999) PrfA, the transcriptional activator of virulence genes, is upregulated during interaction of *Listeria monocytogenes* with mammalian cells and in eukaryotic cell extracts. *Mol Microbiol* 34(3):552-561.
- Riccitelli NJ & Lupták A (2010) Computational discovery of folded RNA domains in genomes and in vitro selected libraries. *Methods* 52(2):133-140.
- Righetti F, Nuss AM, Twittenhoff C, Beele S, Urban K, Will S, Bernhart SH, Stadler PF, Dersch P & Narberhaus F (2016) Temperature-responsive in vitro RNA structurome of *Yersinia pseudotuberculosis*. *Proceedings of the National Academy of Sciences* 113(26):7237-7242.
- Rocha AJ, Barsottini MRdO, Rocha RR, Laurindo MV, Moraes FLLd & Rocha SLd (2019) *Pseudomonas aeruginosa*: virulence factors and antibiotic resistance genes. *Brazilian Archives of Biology and Technology* 62.

- Rodionov DA, Vitreschak AG, Mironov AA & Gelfand MS (2003) Regulation of lysine biosynthesis and transport genes in bacteria: yet another RNA riboswitch? *Nucleic acids research* 31(23):6748-6757.
- Rollins SM, Grundy FJ & Henkin TM (1997) Analysis of cis-acting sequence and structural elements required for antitermination of the *Bacillus subtilis* tyrS gene. *Mol. Microbiol.* 25(2):411-421.
- Römling U, Gomelsky M & Galperin MY (2005) C-di-GMP: the dawning of a novel bacterial signalling system. *Mol. Microbiol.* 57(3):629-639.
- Rubio-Gómez JM, Santiago CM, Udaondo Z, Garitaonandia MT, Krell T, Ramos J-L & Daddaoua A (2020) Full transcriptomic response of *Pseudomonas aeruginosa* to an inulin-derived fructooligosaccharide. *Frontiers in microbiology* 11:202.
- Salehi-Ashtiani K, Lupták A, Litovchick A & Szostak JW (2006) A genomewide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene. *Science* 313(5794):1788-1792.
- Salvail H, Balaji A, Yu D, Roth A & Breaker RR (2020) Biochemical validation of a fourth guanidine riboswitch class in bacteria. *Biochemistry* 59(49):4654-4662.
- Salvail H & Breaker RR (2023) Riboswitches. *Current Biology* 33(9):R343-R348.
- Sato MP, Ogura Y, Nakamura K, Nishida R, Gotoh Y, Hayashi M, Hisatsune J, Sugai M, Takehiko I & Hayashi T (2019) Comparison of the sequencing bias of currently available library preparation kits for Illumina sequencing of bacterial genomes and metagenomes. *DNA Res* 26(5):391-398.
- Scalschi L, Camañes G, Llorens E, Fernández-Crespo E, López MM, García-Agustín P & Vicedo B (2014) Resistance inducers modulate *Pseudomonas syringae* pv. tomato strain DC3000 response in tomato plants. *PLoS ONE* 9(9):e106429.
- Scharfman A, Kroczyński H, Carnoy C, Van Brussel E, Lamblin G, Ramphal R & Roussel P (1996) Adhesion of *Pseudomonas aeruginosa* to respiratory mucins and expression of mucin-binding proteins are increased by limiting iron during growth. *Infect Immun* 64(12):5417-5420.
- Schlax PJ, Xavier KA, Gluick TC & Draper DE (2001) Translational repression of the *Escherichia coli* α operon mRNA: importance of an mRNA conformational switch and a ternary entrapment complex. *Journal of Biological Chemistry* 276(42):38494-38501.
- Schuck A, Diwa A & Belasco JG (2009) RNase E autoregulates its synthesis in *Escherichia coli* by binding directly to a stem-loop in the rne 5' untranslated region. *Mol. Microbiol.* 72(2):470-478.
- Serganov A, Huang L & Patel DJ (2009) Coenzyme recognition and gene regulation by a flavin mononucleotide riboswitch. *nature* 458(7235):233-237.
- Serganov A & Nudler E (2013) A decade of riboswitches. *Cell* 152(1-2):17-24.
- Sherlock ME & Breaker RR (2017) Biochemical validation of a third guanidine riboswitch class in bacteria. *Biochemistry* 56(2):359-363.
- Sherlock ME & Breaker RR (2020) Former orphan riboswitches reveal unexplored areas of bacterial metabolism, signaling, and gene control processes. *RNA* 26(6):675-693.
- Sherlock ME, Malkowski SN & Breaker RR (2017) Biochemical validation of a second guanidine riboswitch class in bacteria. *Biochemistry* 56(2):352-358.

- Sherlock ME, Sudarsan N & Breaker RR (2018a) Riboswitches for the alarmone ppGpp expand the collection of RNA-based signaling systems. *Proceedings of the National Academy of Sciences* 115(23):6052-6057.
- Sherlock ME, Sudarsan N, Stav S & Breaker RR (2018b) Tandem riboswitches form a natural Boolean logic gate to control purine metabolism in bacteria. *eLife* 7.
- Sherwood AV, Grundy FJ & Henkin TM (2015) T box riboswitches in Actinobacteria: translational regulation via novel tRNA interactions. *Proceedings of the National Academy of Sciences* 112(4):1113-1118.
- Shimada T, Fujita N, Yamamoto K & Ishihama A (2011) Novel roles of cAMP receptor protein (CRP) in regulation of transport and metabolism of carbon sources. *PLoS ONE* 6(6):e20081.
- Shimada T, Ishihama A, Busby SJ & Grainger DC (2008) The Escherichia coli RutR transcription factor binds at targets within genes as well as intergenic regions. *Nucleic acids research* 36(12):3950-3955.
- Siegfried NA, Busan S, Rice GM, Nelson JA & Weeks KM (2014) RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat Methods* 11(9):959-965.
- Singer BS, Shtatland T, Brown D & Gold L (1997) Libraries for genomic SELEX. *Nucleic Acids Res* 25(4):781-786.
- Smirnov A, Wang C, Drewry LL & Vogel J (2017) Molecular mechanism of mRNA repression in trans by a ProQ-dependent small RNA. *The EMBO journal* 36(8):1029-1045.
- Smith KD, Lipchock SV, Ames TD, Wang J, Breaker RR & Strobel SA (2009) Structural basis of ligand binding by a c-di-GMP riboswitch. *Nature structural & molecular biology* 16(12):1218.
- Smith RS & Iglewski BH (2003) P. aeruginosa quorum-sensing systems and virulence. *Curr Opin Microbiol* 6(1):56-60.
- Speed MC, Burkhart BW, Picking JW & Santangelo TJ (2018) An Archaeal Fluoride-Responsive Riboswitch Provides an Inducible Expression System for Hyperthermophiles. *Appl Environ Microbiol* 84(7).
- Stav S, Atilho RM, Mirihana Arachchilage G, Nguyen G, Higgs G & Breaker RR (2019) Genome-wide discovery of structured noncoding RNAs in bacteria. *BMC microbiology* 19(1):1-18.
- Steinchen W, Zegarra V & Bange G (2020) (p)ppGpp: Magic Modulators of Bacterial Physiology and Metabolism. *Front Microbiol* 11:2072.
- Stockbridge N (1987) EGTA. *Computers in Biology and Medicine* 17(5):299-304.
- Storz G, Vogel J & Wassarman KM (2011) Regulation by small RNAs in bacteria: expanding frontiers. *Molecular cell* 43(6):880-891.
- Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P, Hickey MJ, Brinkman FSL, Hufnagle WO, Kowalik DJ, Lagrou M, Garber RL, Goltry L, Tolentino E, Westbrock-Wadman S, Yuan Y, Brody LL, Coulter SN, Folger KR, Kas A, Larbig K, Lim R, Smith K, Spencer D, Wong GKS, Wu Z, Paulsen IT, Reizer J, Saier MH, Hancock REW, Lory S & Olson MV (2000) Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen. *Nature* 406(6799):959-964.

- Stubben CJ, Micheva-Viteva SN, Shou Y, Buddenborg SK, Dunbar JM & Hong-Geller E (2014) Differential expression of small RNAs from *Burkholderia thailandensis* in response to varying environmental and stress conditions. *BMC Genomics* 15(1):385.
- Sudarsan N, Barrick JE & Breaker RR (2003a) Metabolite-binding RNA domains are present in the genes of eukaryotes. *RNA* 9(6):644-647.
- Sudarsan N, Lee E, Weinberg Z, Moy R, Kim J, Link K & Breaker R (2008) Riboswitches in eubacteria sense the second messenger cyclic di-GMP. *Science* 321(5887):411-413.
- Sudarsan N, Wickiser JK, Nakamura S, Ebert MS & Breaker RR (2003b) An mRNA structure in bacteria that controls gene expression by binding lysine. *Genes & development* 17(21):2688-2697.
- Sun Z, Cunningham J, Slager S & Kocher J-P (2015) Base resolution methylome profiling: considerations in platform selection, data preprocessing and analysis. *Epigenomics* 7(5):813-828.
- Tang J, Luo M, Niu S, Zhou H, Cai X, Zhang W, Hu Y, Yin Y, Huang A & Wang D (2010) The Crystal Structure of Hexamer RraA from *Pseudomonas Aeruginosa* Reveals Six Conserved Protein-Protein Interaction Sites. *The Protein Journal* 29(8):583-590.
- Tapsin S, Sun M, Shen Y, Zhang H, Lim XN, Susanto TT, Yang SL, Zeng GS, Lee J & Lezhava A (2018) Genome-wide identification of natural RNA aptamers in prokaryotes and eukaryotes. *Nature communications* 9(1):1-10.
- Thomason MK & Storz G (2010) Bacterial antisense RNAs: how many are there, and what are they doing? *Annu Rev Genet* 44:167-188.
- Thomason MK, Voichek M, Dar D, Addis V, Fitzgerald D, Gottesman S, Sorek R & Greenberg EP (2019) A rhII 5' UTR-derived sRNA regulates RhIR-dependent quorum sensing in *Pseudomonas aeruginosa*. *MBio* 10(5):e02253-02219.
- Tian S & Das R (2017) Primerize-2D: automated primer design for RNA multidimensional chemical mapping. *Bioinformatics* 33(9):1405-1406.
- Tropel D & Van Der Meer JR (2004) Bacterial transcriptional regulators for degradation pathways of aromatic compounds. *Microbiology and molecular biology reviews* 68(3):474-500.
- Tsai M & Saier Jr M (1995) Phylogenetic characterization of the ubiquitous electron transfer flavoprotein families ETF- α and ETF- β . *Research in microbiology* 146(5):397-404.
- Tuerk C & Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249(4968):505-510.
- Ueta M, Yoshida H, Wada C, Baba T, Mori H & Wada A (2005) Ribosome binding proteins YhbH and YfiA have opposite functions during 100S formation in the stationary phase of *Escherichia coli*. *Genes Cells* 10(12):1103-1112.
- Valentini M & Filloux A (2016) Biofilms and Cyclic di-GMP (c-di-GMP) Signaling: Lessons from *Pseudomonas aeruginosa* and Other Bacteria*. *Journal of Biological Chemistry* 291(24):12547-12555.
- Vicens Q, Mondragón E & Batey RT (2011) Molecular sensing by the aptamer domain of the FMN riboswitch: a general model for ligand binding by conformational selection. *Nucleic acids research* 39(19):8586-8598.

- Visweswariah SS & Busby SJ (2015) Evolution of bacterial transcription factors: how proteins take on new tasks, but do not always stop doing the old ones. *Trends in microbiology* 23(8):463-467.
- Vitreschak AG, Mironov AA, Lyubetsky VA & Gelfand MS (2008) Comparative genomic analysis of T-box regulatory systems in bacteria. *Rna* 14(4):717-735.
- Vitreschak AG, Rodionov DA, Mironov AA & Gelfand MS (2002) Regulation of riboflavin biosynthesis and transport genes in bacteria by transcriptional and translational attenuation. *Nucleic acids research* 30(14):3141-3151.
- Volkenborn K, Kuschmierz L, Benz N, Lenz P, Knapp A & Jaeger K-E (2020) The length of ribosomal binding site spacer sequence controls the production yield for intracellular and secreted proteins by *Bacillus subtilis*. *Microbial Cell Factories* 19(1):154.
- Vu MM, Jameson NE, Masuda SJ, Lin D, Larralde-Ridaura R & Lupták A (2012) Convergent evolution of adenosine aptamers spanning bacterial, human, and random sequences revealed by structure-based bioinformatics and genomic SELEX. *Chem Biol* 19(10):1247-1254.
- Waite RD, Papakonstantinou A, Littler E & Curtis MA (2005) Transcriptome analysis of *Pseudomonas aeruginosa* growth: comparison of gene expression in planktonic cultures and developing and mature biofilms. *Journal of bacteriology* 187(18):6571-6576.
- Waldminghaus T, Heidrich N, Brantl S & Narberhaus F (2007) FourU: a novel type of RNA thermometer in *Salmonella*. *Mol. Microbiol.* 65(2):413-424.
- Wang JX, Lee ER, Morales DR, Lim J & Breaker RR (2008) Riboswitches that sense S-adenosylhomocysteine and activate genes involved in coenzyme recycling. *Molecular cell* 29(6):691-702.
- Wang T, Cai Z, Shao X, Zhang W, Xie Y, Zhang Y, Hua C, Schuster SC, Yang L & Deng X (2019) Pleiotropic Effects of c-di-GMP Content in *Pseudomonas syringae*. *Appl Environ Microbiol* 85(10).
- Wang Z, Xie X, Shang D, Xie L, Hua Y, Song L, Yang Y, Wang Y, Shen X & Zhang L (2022) A c-di-GMP Signaling Cascade Controls Motility, Biofilm Formation, and Virulence in *Burkholderia thailandensis*. *Applied and Environmental Microbiology* 88(7):e02529-02521.
- Weaver JW & Serganov A (2019) T-box RNA gets boxed. *Nature Structural & Molecular Biology* 26(12):1081-1083.
- Weeks KM & Mauger DM (2011) Exploring RNA Structural Codes with SHAPE Chemistry. *Accounts of Chemical Research* 44(12):1280-1291.
- Weinberg Z, Barrick JE, Yao Z, Roth A, Kim JN, Gore J, Wang JX, Lee ER, Block KF & Sudarsan N (2007) Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic acids research* 35(14):4809-4819.
- Weinberg Z & Breaker RR (2011) R2R-software to speed the depiction of aesthetic consensus RNA secondary structures. *BMC bioinformatics* 12(1):1-9.
- Weinberg Z, Lünse CE, Corbino KA, Ames TD, Nelson JW, Roth A, Perkins KR, Sherlock ME & Breaker RR (2017a) Detection of 224 candidate structured RNAs by comparative analysis of specific subsets of intergenic regions. *Nucleic acids research* 45(18):10811-10823.

- Weinberg Z, Nelson JW, Lünse CE, Sherlock ME & Breaker RR (2017b) Bioinformatic analysis of riboswitch structures uncovers variant classes with altered ligand specificity. *Proceedings of the National Academy of Sciences* 114(11):E2077-E2085.
- Weinberg Z, Wang JX, Bogue J, Yang J, Corbino K, Moy RH & Breaker RR (2010) Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome biology* 11(3):1-17.
- White N, Sadeeshkumar H, Sun A, Sudarsan N & Breaker RR (2022a) Lithium-sensing riboswitch classes regulate expression of bacterial cation transporter genes. *Sci. Rep.* 12(1):19145.
- White N, Sadeeshkumar H, Sun A, Sudarsan N & Breaker RR (2022b) Na⁺ riboswitches regulate genes for diverse physiological processes in bacteria. *Nat. Chem. Biol.* 18(8):878-885.
- Wilson DS, Keefe AD & Szostak JW (2001) The use of mRNA display to select high-affinity protein-binding peptides. *Proceedings of the National Academy of Sciences* 98(7):3750-3755.
- Wilson HR, Archer CD, Liu J & Turnbough Jr CL (1992) Translational control of pyrC expression mediated by nucleotide-sensitive selection of transcriptional start sites in Escherichia coli. *Journal of bacteriology* 174(2):514-524.
- Winkler W, Nahvi A & Breaker RR (2002a) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* 419(6910):952-956.
- Winkler WC & Breaker RR (2003) Genetic control by metabolite-binding riboswitches. *Chembiochem* 4(10):1024-1032.
- Winkler WC, Cohen-Chalamish S & Breaker RR (2002b) An mRNA structure that controls gene expression by binding FMN. *Proc Natl Acad Sci U S A* 99(25):15908-15913.
- Winkler WC, Nahvi A, Sudarsan N, Barrick JE & Breaker RR (2003) An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nature Structural & Molecular Biology* 10(9):701-707.
- Winsor GL, Griffiths EJ, Lo R, Dhillon BK, Shay JA & Brinkman FS (2016) Enhanced annotations and features for comparing thousands of Pseudomonas genomes in the Pseudomonas genome database. *Nucleic Acids Res* 44(D1):D646-653.
- Wu M, Guina T, Brittnacher M, Nguyen H, Eng J & Miller SI (2005) The Pseudomonas aeruginosa proteome during anaerobic growth. *Journal of bacteriology* 187(23):8185-8190.
- Wurtzel O, Yoder-Himes DR, Han K, Dandekar AA, Edelheit S, Greenberg EP, Sorek R & Lory S (2012) The single-nucleotide resolution transcriptome of Pseudomonas aeruginosa grown in body temperature.
- Xu J & Cotruvo JA (2022) Reconsidering the czcD (NiCo) Riboswitch as an Iron Riboswitch. *ACS Bio & Med Chem Au* 10.1021/acsbiochem.1c00069.
- Yamagishi M, Matsushima H, Wada A, Sakagami M, Fujita N & Ishihama A (1993) Regulation of the Escherichia coli rnf gene encoding the ribosome modulation factor: growth phase- and growth rate-dependent control. *EMBO J* 12(2):625-630.
- Yao Z, Barrick J, Weinberg Z, Neph S, Breaker R, Tompa M & Ruzzo WL (2007) A computational pipeline for high-throughput discovery of cis-regulatory noncoding RNA in prokaryotes. *PLoS computational biology* 3(7):e126.
- Yao Z, Weinberg Z & Ruzzo WL (2006) CMfinder—a covariance model based RNA motif finding algorithm. *Bioinformatics* 22(4):445-452.

- Yoshida H, Maki Y, Kato H, Fujisawa H, Izutsu K, Wada C & Wada A (2002) The Ribosome Modulation Factor (RMF) Binding Site on the 100S Ribosome of *Escherichia coli*. *The Journal of Biochemistry* 132(6):983-989.
- Yoshida H & Wada A (2014) The 100S ribosome: ribosomal hibernation induced by stress. *Wiley Interdiscip. Rev. RNA* 5(5):723-732.
- Yoshikazu N, Jacqueline P, Jacques D & Marianne G-M (1985) Evidence for autoregulation of the nusA-infB operon of *Escherichia coli*. *Gene* 36(1-2):189-193.
- Young J, Luketina R & Marshall A (1977) The effects on temperature on growth in vitro of *Pseudomonas syringae* and *Xanthomonas pruni*. *Journal of Applied Bacteriology* 42(3):345-354.
- Yousef MR, Grundy FJ & Henkin TM (2003) tRNA requirements for glyQS antitermination: a new twist on tRNA. *RNA* 9(9):1148-1156.
- Yu D & Breaker RR (2020) A bacterial riboswitch class senses xanthine and uric acid to regulate genes associated with purine oxidation. *RNA* 26(8):960-968.
- Zarzosa-Álvarez AL, Sandoval-Cabrera A, Torres-Huerta AL & Bermudez-Cruz RM (2010) Electroeluting DNA fragments. *JoVE (Journal of Visualized Experiments)* (43):e2136.
- Zeller MJ, Favorov O, Li K, Nuthanakanti A, Hussein D, Michaud A, Lafontaine DA, Busan S, Serganov A, Aubé J & Weeks KM (2022) SHAPE-enabled fragment-based ligand discovery for RNA. *Proceedings of the National Academy of Sciences* 119(20):e2122660119.
- Zhang J (2020) Unboxing the T-box riboswitches—A glimpse into multivalent and multimodal RNA–RNA interactions. *Wiley Interdiscip. Rev. RNA* 11(6):e1600.
- Zhu C, Yang G, Ghulam M, Li L & Qu F (2019) Evolution of multi-functional capillary electrophoresis for high-efficiency selection of aptamers. *Biotechnology advances* 37(8):107432.
- Zolg JW & Hänggi UJ (1981) Characterization of a R plasmid-associated, trimethoprim-resistant dihydrofolate reductase and determination of the nucleotide sequence of the reductase gene. *Nucleic Acids Research* 9(3):697-710.
- Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research* 31(13):3406-3415.

7 ANNEXES

7.1 Annexe 1 : séquences des ARNnc candidats

Tableau 7.1 Séquences des constructions testées pour chaque motif

Constructions <i>rmf</i>
WT : acgcttgggtagaactccagttaatccgccaccgcaacgggtgctggggcagggtcgtcacgccggcaacgaccccttcggcgaaagcttcgc caggcctcccctggggccaacgggacataacagtaaac aagtgagggaacaccctatgagaagacttaagcgt
Disr : acgcttgggtagaactccagttaatgcccaccgcaacgggtgctggggcagggtcgtcacgccggcaacgaccccttcggcgaaagcttcgc caggcctcccctggggccaacgggacataacagtaaac aagtgagggaacaccctatgagaagacttaagcgt
Comp : acgcttgggtagaactccagttaatgcccaccgcaacgggtgctggggcagggtcgtcacgccggcaacgaccccttcggcgaaagcttcgc ccaggcctcccctggggccaacgggacataacagtaaac aagtgagggaacaccctatgagaagacttaagcgt
Constructions <i>Pseudomon</i> -GGDEF
WT : tctcgttttaccgtaaagaattaatcacatccgccgagagcctgacaagcgagctgattaatatcgggcttcgagaatgagcagtaagccacgt gtgcttgacttggtaacagagaagccttgcatgctgaaaaaactagcg
Disr : tctcgttttaccgtaaagaattaatcacatccgccgagagcctgacaagcgagctgattaatatcgggcttcgagaatgagcagtaagccacgt gtgcttgacttggtaacagagatcccttgcatgctgaaaaaactagcg
Constructions <i>Pseudomon</i> -Rho
WT : ttaaatcgggttcgacgctttttccgtcgccttctgcatgccgtcccggacgcacttctactcgaactttctagcgtcctgtcgtcccccttcgac acggccttccaagctcgcgcttttccaattcctttctgcgtacctatctttctatgaactgaccgaactc
Disr : ttaaatcgggttcgacgctggttttccgtcgccttctgcatgccgtcccggacgcacttctactcgaactttctagcgtcctgtcgtcccccttcgac acggccttccaagctcgcgcttttccaattcctttctgcgtacctatctttctatgaactgaccgaactc
Constructions <i>sucA-II</i>
WT : aacacctgtggcgccggagttctccggcgtcgcagcaactaagcgggccaagctcacaagccatggcccagattgaaaaatagaccag caggggatccgggctgtaccggactatctcgggatctgtagtgccataccgatgctcgtcctgaggacttcggagccttcaaggttcg ccagggtgtgtcccctaccgaggtgacctagcatgcacgaaagcgtaatg

<p>Disr :</p> <p>aacacctgtgggcccggagttctccggcgtgcagcaactaagcgggccaagctcacaagccatggcccagattgaaaaatagaccagcaggggcatccgggctggtaccggactatctgcgggatctgtagtggccataccgatgctgctgctgaggacttcggagcctcaaggttcg ccaggtggtgtcccctaccgaggtgacctagcatgcacgaaagcgtaatg</p>
<p>Constructions <i>fixA</i></p>
<p>WT :</p> <p>gtcaagatacggggcggtttgcgttgccgtaggctgaaacgaggttcgacagaatctccaaaaggcatcaggcaaacgcggtgcaggcaa caggtaacgtcgcacggcgcattttacacaccctgccgctgcttgacgggtgacgcccggcattataaggccttttaaagccctgccgctggcc tgcacacattgcgattgctgtgaggacactatcgatcgctggtattgccgggcgacatctttcaccggagagtgaggaatccatgaaggttct tgtagct</p>
<p>Disr :</p> <p>gtcaagatacggggcggtttgccattgccgtaggctgaaacgaggttcgacagaatctccaaaaggcatcaggcaaacgcggtgcaggcaa caggtaacgtcgcacggcgcattttacacaccctgccgctgcttgacgggtgacgcccggcattataaggccttttaaagccctgccgctggcc tgcacacattgcgattgctgtgaggacactatcgatcgctggtattgccgggcgacatctttcaccggagagtgaggaatccatgaaggttct tgtagct</p>
<p>Constructions <i>rnell</i></p>
<p>WT :</p> <p>aaggaacatgtgtagggtggagatgtacggctgtggactgctgtagcttctagcttgatccagacgcattttcgtccgctccgacagccgattcct cctcctgactcaagtgcttctgaaacacaggaagcaggagacgtagtcgcgaccggcaccagtgccgagcggcctggacacgggaat ggcccgcattcctgaatacgcgcccagcaccgaccgtgagagtcgtgtgcccgatcgccgttccggtagccccgaaaccattggtacta catgaaaagaatgctgac</p>
<p>Disr :</p> <p>aaggaacatgtgtagggtgctgatgtacggctgtggactgctgtagcttctagcttgatccagacgcattttcgtccgctccgacagccgattcct cctcctgactcaagtgcttctgaaacacaggaagcaggagacgtagtcgcgaccggcaccagtgccgagcggcctggacacgggaat ggcccgcattcctgaatacgcgcccagcaccgaccgtgagagtcgtgtgcccgatcgccgttccggtagccccgaaaccattggtacta catgaaaagaatgctgac</p>

Tableau répertoriant toutes les séquences d'ADN pour les constructions des différents motifs. « WT » correspond au motif naturellement présent dans la bactérie. « Disr » correspond à la mutation disruptive du motif, en gras sont représentés les nucléotides de la mutation. La séquence du motif est représentée en rouge. Les nucléotides en vert correspondent au début de la séquence codante du gène en aval.

7.2 Annexe 2 : motif *rmf*

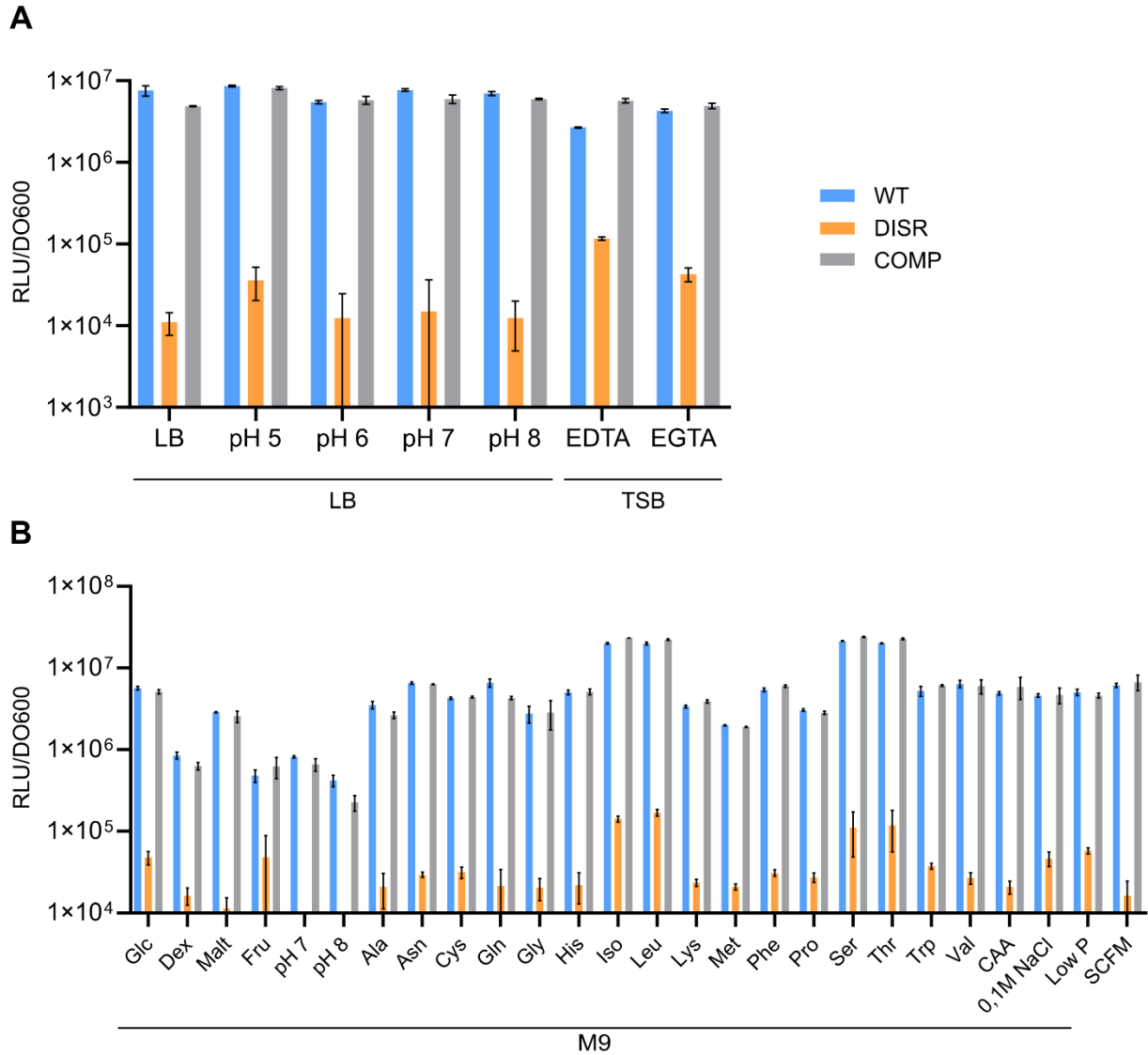


Figure 7.1 Le même profil d'expression de la luminescence est observable pour tous les milieux testés pour le motif *rmf*.

Histogramme des ratios luminescence / DO 600 nm obtenus pour les constructions WT (bleu), disruptive (orange) et compensatoire (gris) du motif *rmf*. **A** : Les milieux testés sont le LB, LB pH 5 (pH 5), LB pH 6 (pH 6), LB pH 7 (pH 7), LB pH 8 (pH 8), TSB + 1mM EDTA (EDTA), TSB + 1 mM EGTA (EGTA). **B** : Les milieux testés sont le M9 glucose (Glc), M9 dextrose (Dex), M9 maltose (Malt), M9 fructose (Fru), M9 pH 7 (pH 7), M9 pH 8 (pH 8), M9 alanine (Ala), M9 asparagine (Asn), M9 cystéine (Cys), M9 glutamine (Gln), M9 glycine (Gly), M9 histidine (His), M9 isoleucine (Iso), M9 leucine (Leu), M9 lysine (Lys), M9 méthionine (Met), M9 phénylalanine (Phe), M9 proline (Pro), M9 sérine (Ser), M9 thréonine (Thr), M9 tryptophane (Trp), M9 valine (Val), M9 casamino acid (CAA), M9 + 0,1 M NaCl (0,1 M NaCl), M9 low phosphate (Low P) et SCFM.

7.3 Annexe 3 : cartographie *in vitro* par *in-line* de *rmf*

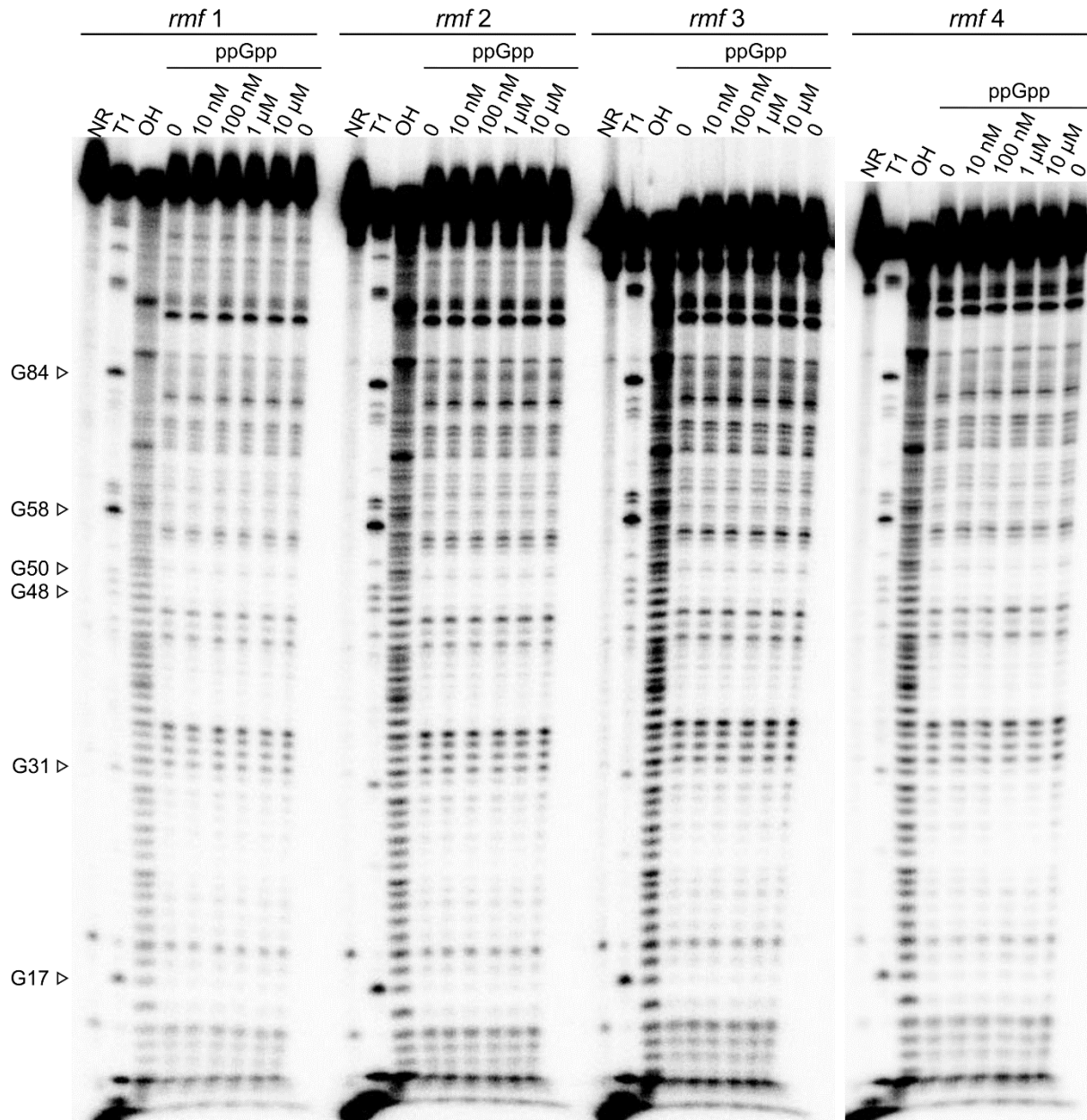


Figure 7.2 Une structure identique se remarque pour toutes les constructions du motif *rmf*.

Gels de cartographie *in vitro* par *in-line* pour les constructions de l'ARN *rmf* WT avec des concentrations de 10 nM à 10 μM de ppGpp. NR, T1 et OH représentent les échelles d'ARN non réagi, d'ARN traité à l'ARNase T1 (clive après les G) et d'ARN soumis à un traitement alcalin (clive après chaque nucléotides) respectivement. Certains résidus guanine ont été annotés basés sur l'échelle T1 et OH.

7.4 Annexe 4 : Courbe de K_D pour les différents essais de la construction *rmf 4*

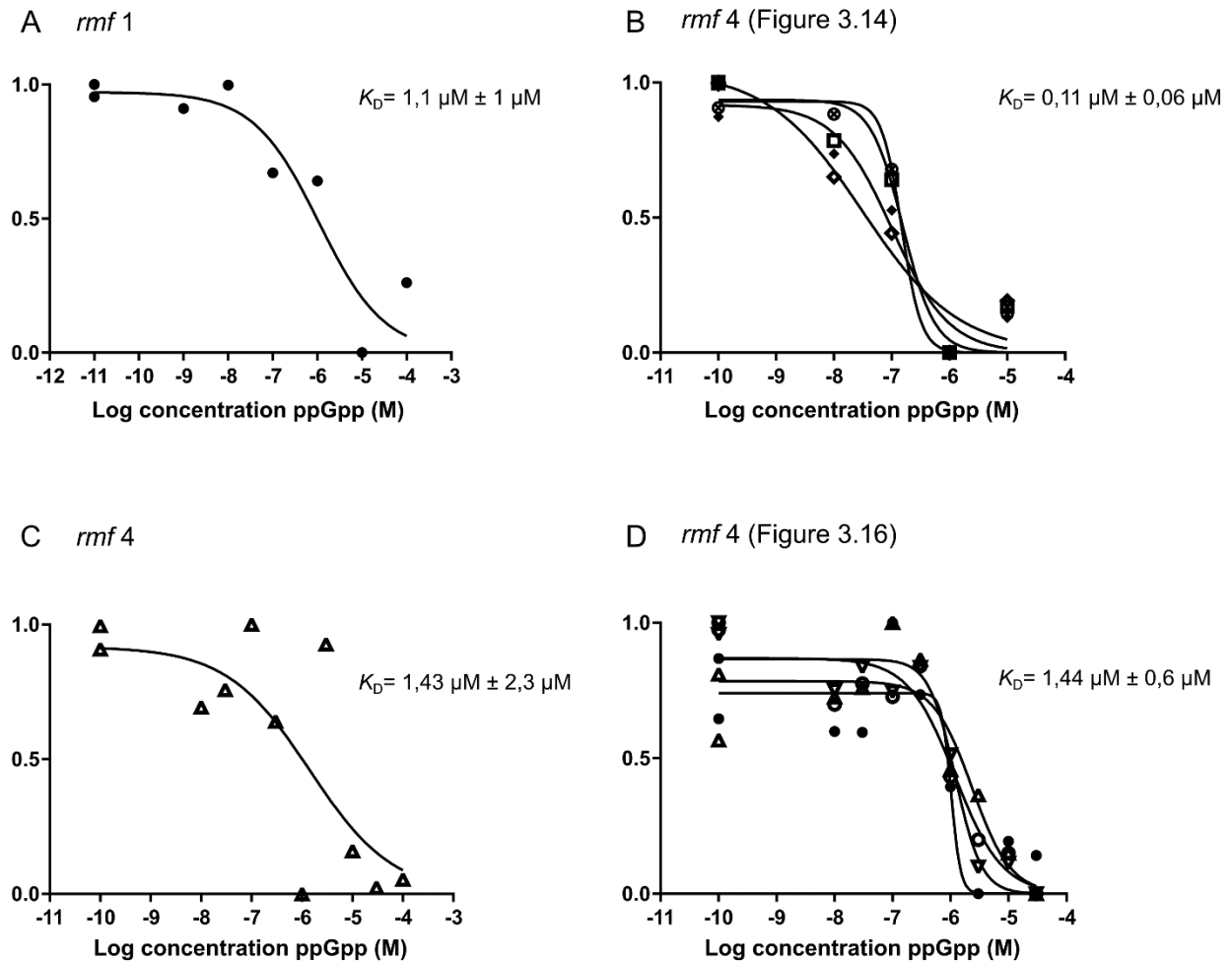


Figure 7.3 Plusieurs essais de cartographie *in vitro* par *in-line* permettent de déterminer des K_D pour l'ARN *rmf*.

Courbes de K_D issus de différents essais de cartographie *in vitro* par *in-line* pour les constructions **A** : *rmf 1* et **B**, **C** et **D** : *rmf 4*. La valeur de K_D représentée sur le graphique correspond à la moyenne des K_D des différentes courbes et l'écart type représente la déviation entre les valeurs de ces différentes courbes.

7.5 Annexe 5 : *Pseudomon*-GGDEF

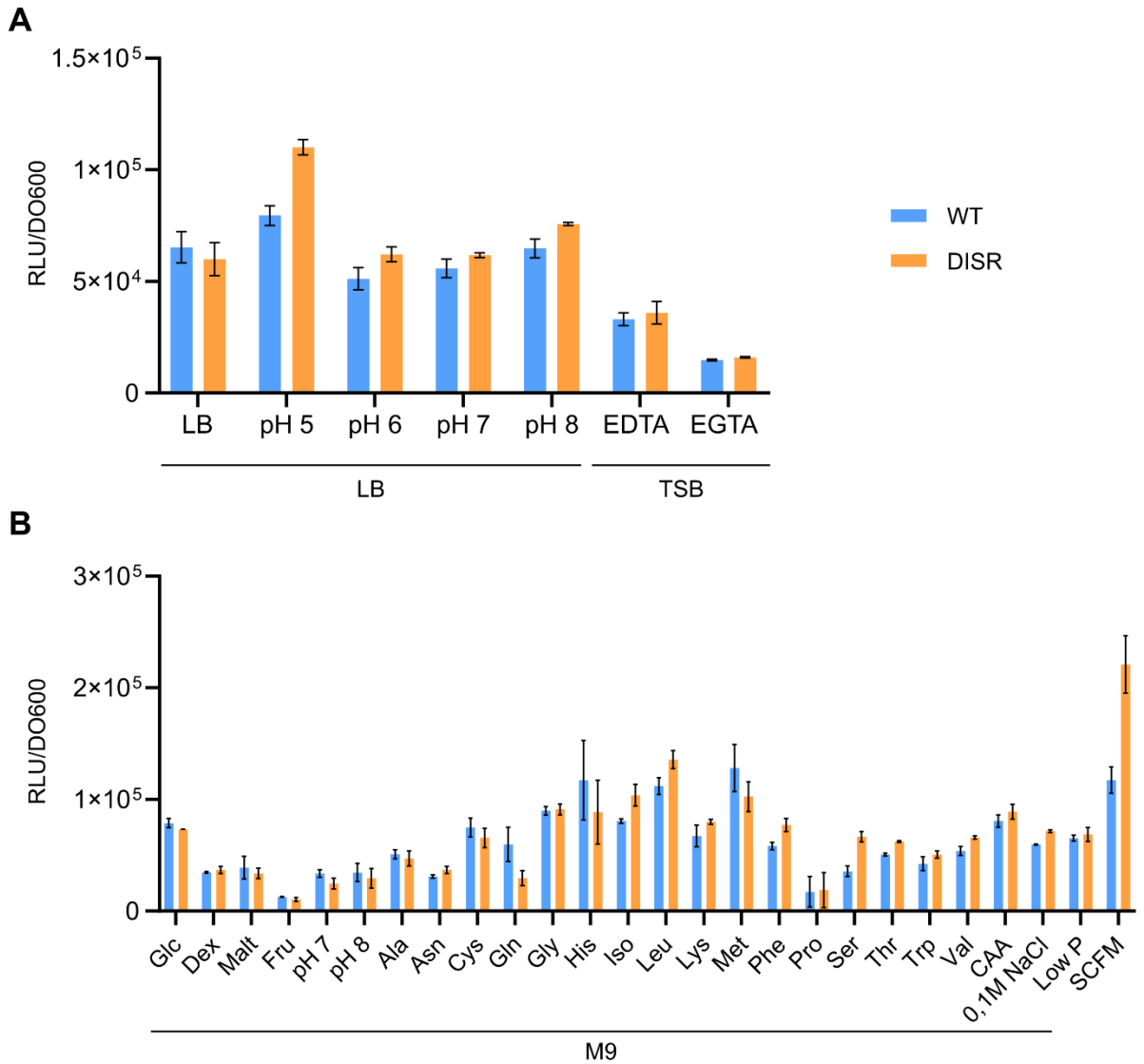


Figure 7.4 Pour la majorité des milieux testés il n’y a pas de différence dans l’expression de la luminescence pour le motif *Pseudomon*-GGDEF.

Histogramme des ratios luminescence / DO 600 nm obtenus pour les constructions WT (bleu), disruptive (orange) du motif *Pseudomon*-GGDEF. **A** : Les milieux testés sont le LB, LB pH 5 (pH 5), LB pH 6 (pH 6), LB pH 7 (pH 7), LB pH 8 (pH 8), TSB + 1mM EDTA (EDTA) et TSB + 1 mM EGTA (EGTA). **B** : Les milieux testés sont le M9 glucose (Glc), M9 dextrose (Dex), M9 maltose (Malt), M9 fructose (Fru), M9 pH 7 (pH 7), M9 pH 8 (pH 8), M9 alanine (Ala), M9 asparagine (Asn), M9 cystéine (Cys), M9 glutamine (Gln), M9 glycine (Gly), M9 histidine (His), M9 isoleucine (Iso), M9 leucine (leu), M9 lysine (Lys), M9 méthionine (Met), M9 phénylalanine (Phe), M9 proline (Pro), M9 sérine (Ser), M9 thréonine (Thr), M9 tryptophane (Trp), M9 valine (Val),) M9 casamino acid (CAA), M9 + 0,1 M NaCl (0,1 M NaCl), M9 low phosphate (Low P) et SCFM.

7.6 Annexe 6 : Cartographie *in vitro* par *in-line* du motif *Pseudomon*-GGDEF

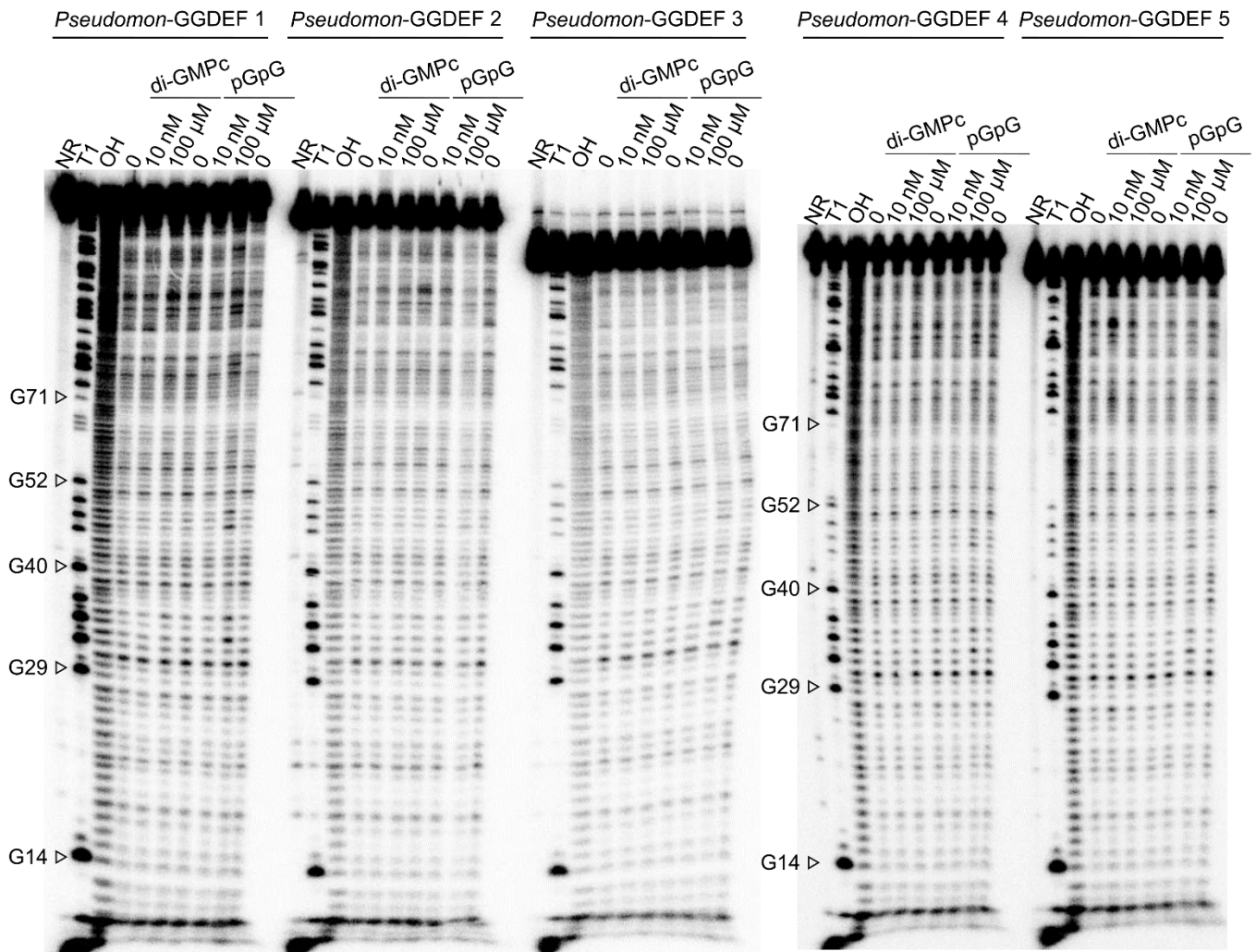


Figure 7.5 La structure de *Pseudomon*-GGDEF est identique quel que soit les constructions testées.

Gel de cartographie *in vitro* par *in-line* pour les différentes constructions de *Pseudomon*-GGDEF avec des concentrations de 10 nM à 100 μM de di-GMPc ou de pGpG. NR, T1 et OH représentent les échelles d'ARN non réagi, d'ARN traité à l'ARNase T1 (clive après les G) et d'ARN soumis à un traitement alcalin (clive après chaque nucléotides) respectivement. Certains résidus guanine ont été annotés basés sur l'échelle T1 et OH.

7.7 Annexe 7 : Cartographie in vitro par in-line *Pseudomon*-GGDEF et *rnf*

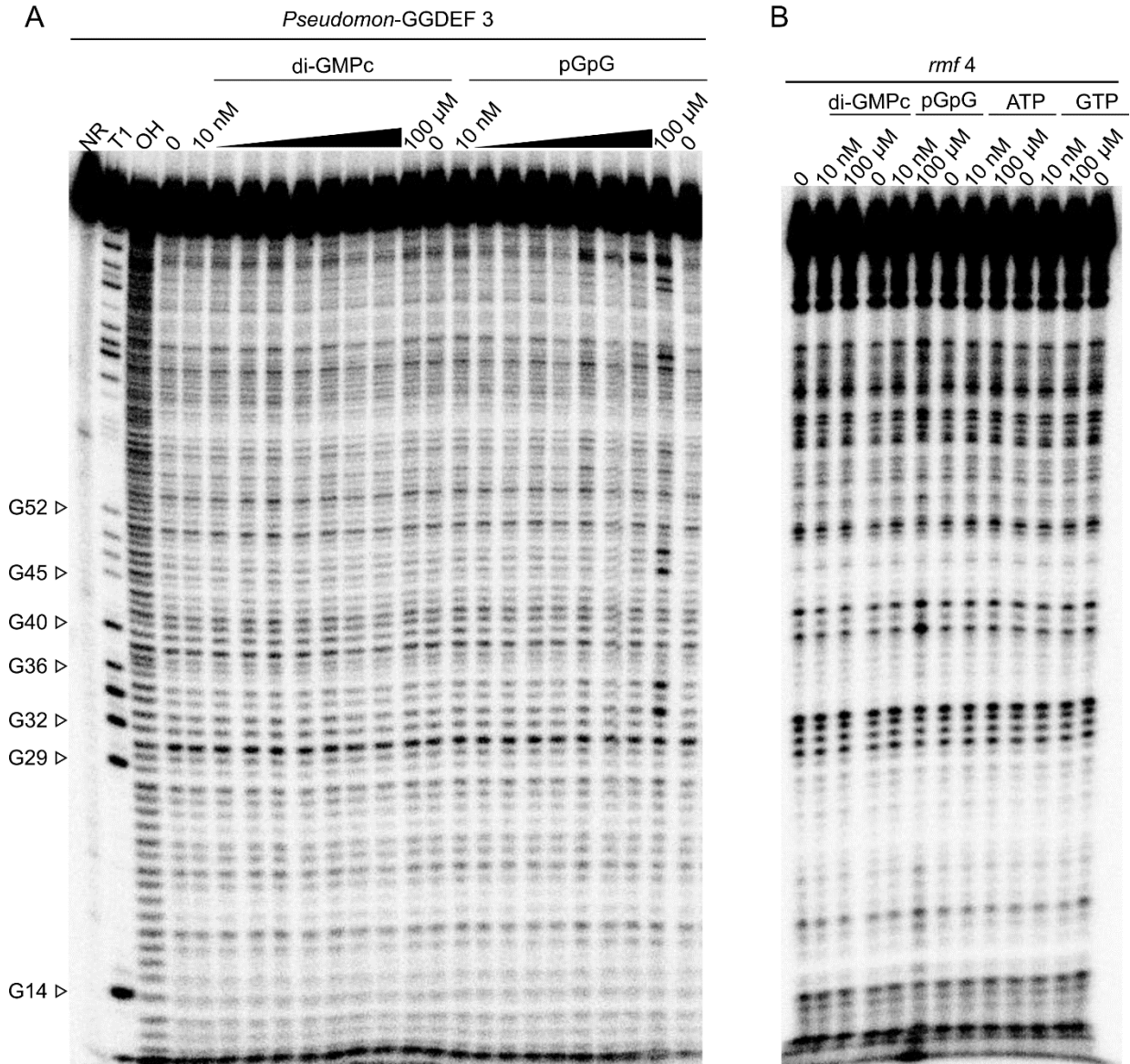


Figure 7.6 Les modulations observées pour le pGpG est non-spécifique.

A : Gel de cartographie *in vitro* par *in-line* pour la construction *Pseudomon*-GGDEF 3 avec des concentrations de 10 nM à 100 μM de di-GMPc ou de pGpG. NR, T1 et OH représentent les échelles d'ARN non réagi, d'ARN traité à l'ARNase T1 (clive après les G) et d'ARN soumis à un traitement alcalin (clive après chaque nucléotides) respectivement. Certains résidus guanine ont été annotés basés sur l'échelle T1 et OH. **B** : Gel de cartographie *in vitro* par *in-line* pour la construction *rnf* 4 WT avec des concentrations de 10 nM et 100 μM de di-GMPc, de pGpG, d'ATP et de GTP.

7.8 Annexe 8 : *Pseudomon*-Rho

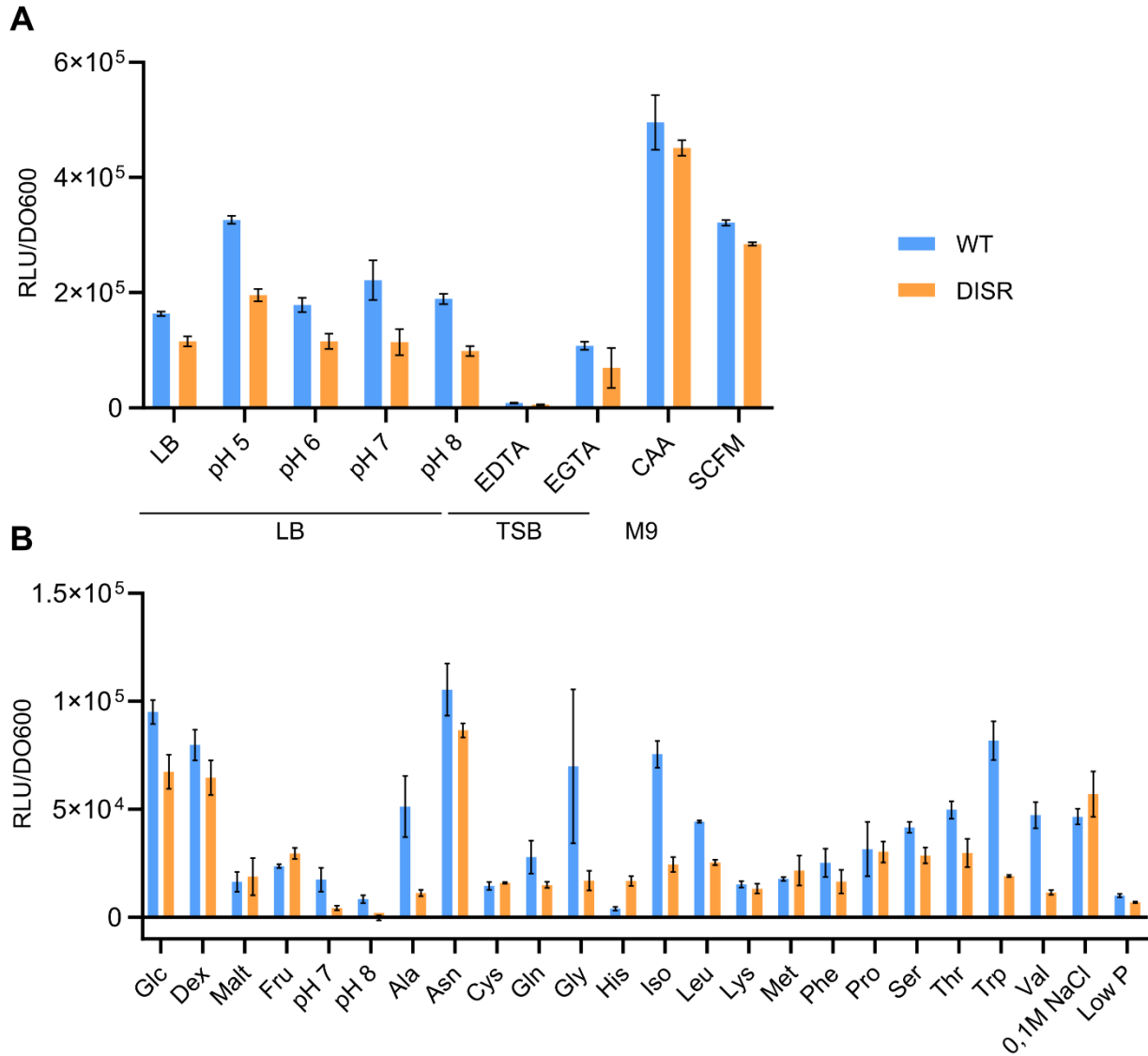


Figure 7.7 Des variations d'expression de la luminescence sont visibles en fonction du milieu de culture utilisé pour le motif *Pseudomon*-Rho.

Histogramme des ratios luminescence / DO 600 nm obtenus pour les constructions WT (bleu), disruptive (orange) du motif *Pseudomon*-Rho. **A** : Les milieux testés sont le LB, LB pH 5 (pH 5), LB pH 6 (pH 6), LB pH 7 (pH 7), LB pH 8 (pH 8), TSB + 1mM EDTA (EDTA), TSB + 1 mM EGTA (EGTA), M9 casamino acid (CAA) et SCFM. **B** : Les milieux testés sont le M9 glucose (Glc), M9 dextrose (Dex), M9 maltose (Malt), M9 fructose (Fru), M9 pH 7 (pH 7), M9 pH 8 (pH 8), M9 alanine (Ala), M9 asparagine (Asn), M9 cystéine (Cys), M9 glutamine (Gln), M9 glycine (Gly), M9 histidine (His), M9 isoleucine (Iso), M9 leucine (Ieu), M9 lysine (Lys), M9 méthionine (Met), M9 phénylalanine (Phe), M9 proline (Pro), M9 sérine (Ser), M9 thréonine (Thr), M9 tryptophane (Trp), M9 valine (Val), M9 + 0,1 M NaCl (0,1 M NaCl) et M9 low phosphate (Low P).

7.9 Annexe 9 : *sucA-II*

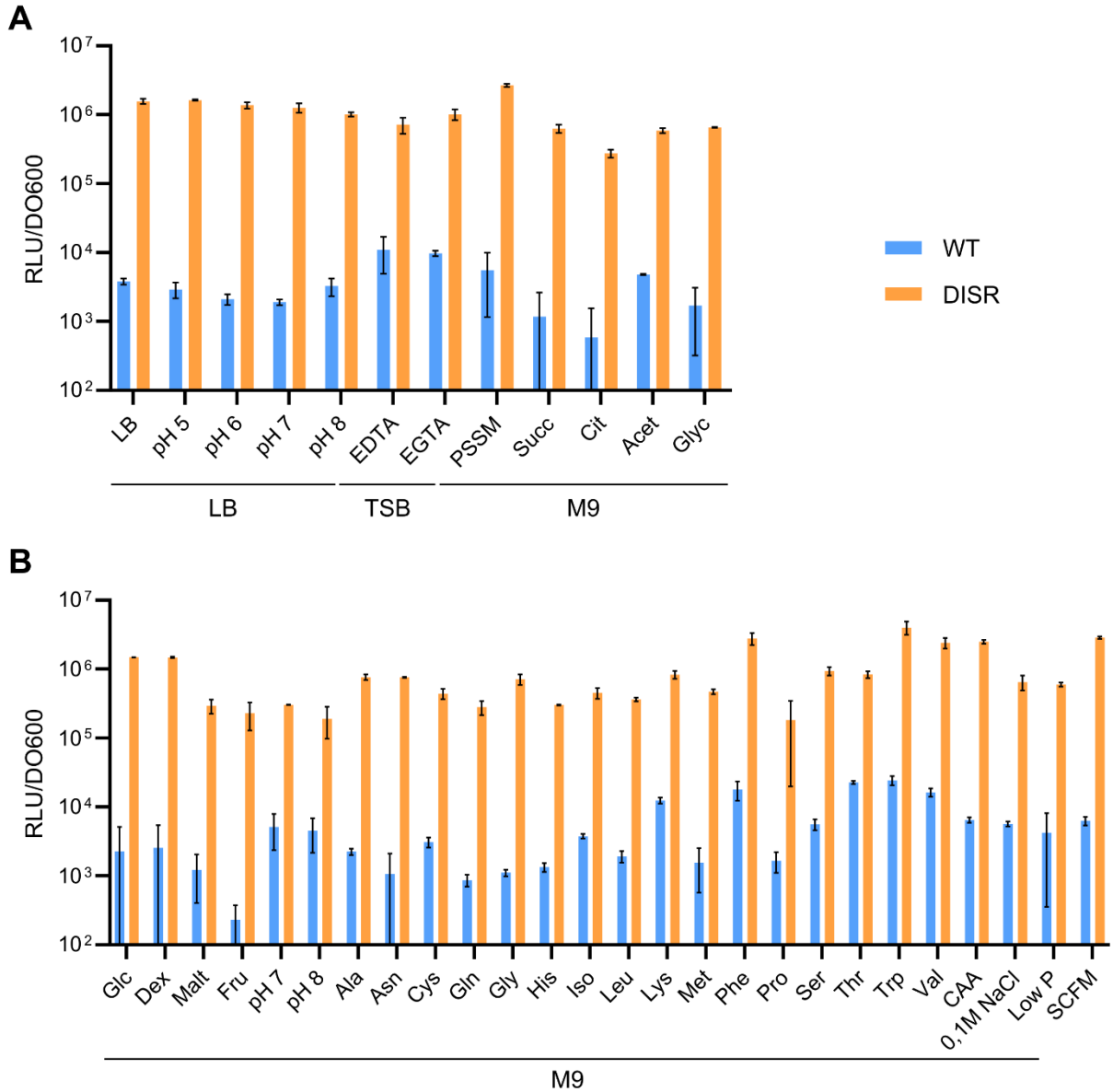
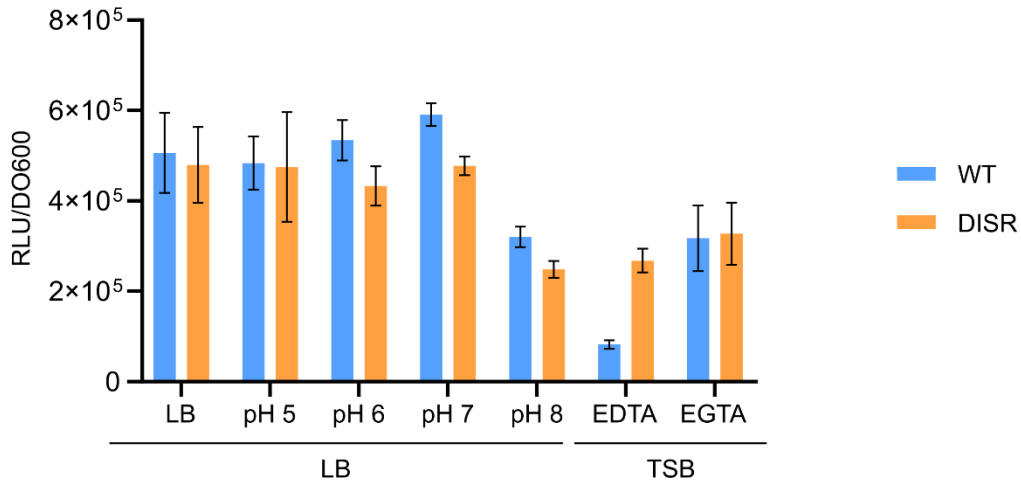


Figure 7.8 Le même profil d'expression de la luminescence est observable pour tous les milieux testés pour le motif *sucA-II*.

Histogramme des ratios luminescence / DO 600 nm obtenus pour les constructions WT (bleu), disruptive (orange) du motif *sucA-II*. **A** : Les milieux testés sont le LB, LB pH 5 (pH 5), LB pH 6 (pH 6), LB pH 7 (pH 7), LB pH 8 (pH 8), TSB + 1mM EDTA (EDTA) et TSB + 1 mM EGTA (EGTA). **B** : Les milieux testés sont le M9 glucose (Glc), M9 dextrose (Dex), M9 maltose (Malt), M9 fructose (Fru), M9 pH 7 (pH 7), M9 pH 8 (pH 8), M9 alanine (Ala), M9 asparagine (Asn), M9 cystéine (Cys), M9 glutamine (Gln), M9 glycine (Gly), M9 histidine (His), M9 isoleucine (Iso), M9 leucine (Leu), M9 lysine (Lys), M9 méthionine (Met), M9 phénylalanine (Phe), M9 proline (Pro), M9 sérine (Ser), M9 thréonine (Thr), M9 tryptophane (Trp), M9 valine (Val), M9 casamino acid (CAA), M9 + 0,1 M NaCl (0,1 M NaCl), M9 low phosphate (Low P) et SCFM.

7.10 Annexe 10 : *fixA*

A



B

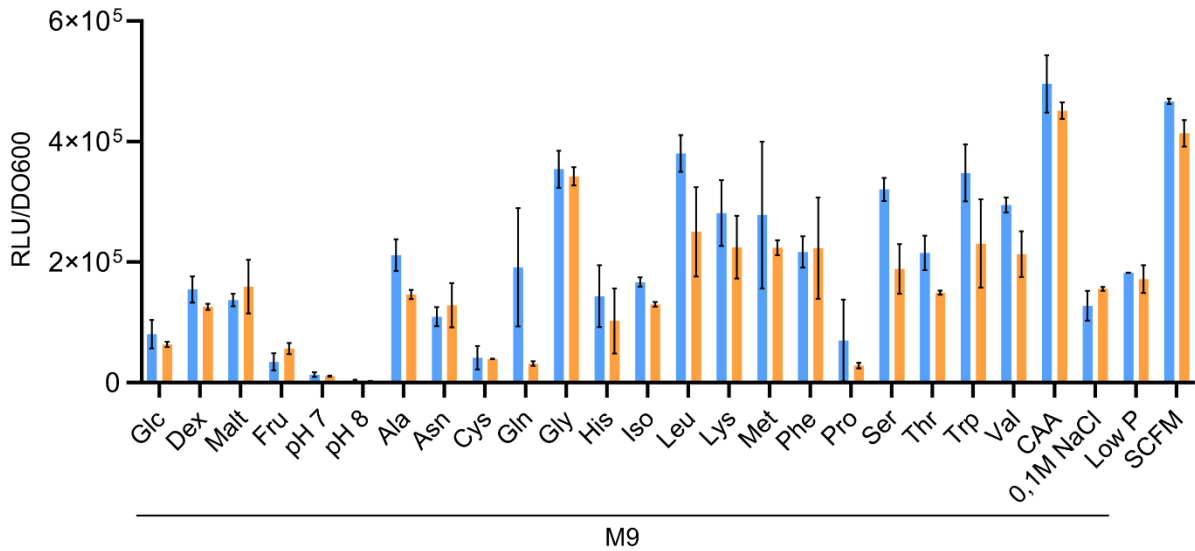


Figure 7.9 L'expression de la luminescence varie en fonction des milieux testés pour le motif *fixA*.

Histogramme des ratios luminescence / DO 600 nm obtenus pour les constructions WT (bleu), disruptive (orange) du motif *fixA*. A : Les milieux testés sont le LB, LB pH 5 (pH 5), LB pH 6 (pH 6), LB pH 7 (pH 7), LB pH 8 (pH 8), TSB + 1mM EDTA (EDTA) et TSB + 1 mM EGTA (EGTA). B : Les milieux testés sont le M9 glucose (Glc), M9 dextrose (Dex), M9 maltose (Malt), M9 fructose (Fru), M9 pH 7 (pH 7), M9 pH 8 (pH 8), M9 alanine (Ala), M9 asparagine (Asn), M9 cystéine (Cys), M9 glutamine (Gln), M9 glycine (Gly), M9 histidine (His), M9 isoleucine (Iso), M9 leucine (Ieu), M9 lysine (Lys), M9 méthionine (Met), M9 phénylalanine (Phe), M9 proline (Pro), M9 sérine (Ser), M9 thréonine (Thr), M9 tryptophane (Trp), M9 valine (Val), M9 casamino acid (CAA), M9 + 0,1 M NaCl (0,1 M NaCl), M9 low phosphate (Low P) et SCFM.

7.11 Annexe 11 : *rne-II*

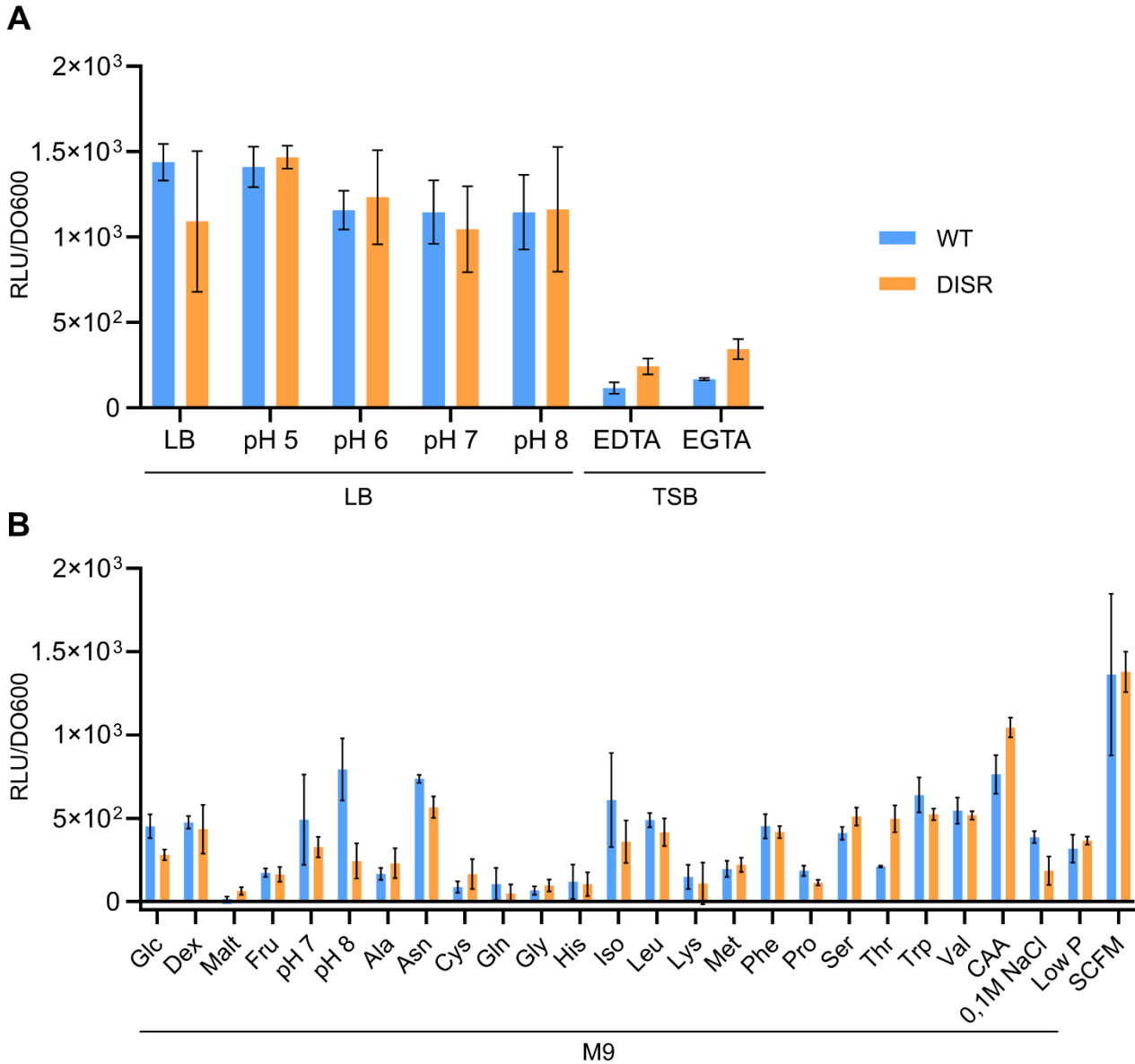


Figure 7.10 Des variations d'expression de la luminescence sont visibles dépendamment du milieu testé pour le motif *rne-II*.

Histogramme des ratios luminescence / DO 600 nm obtenus pour les constructions WT (bleu), disruptive (orange) du motif *rne-II*. **A** : Les milieux testés sont le LB, LB pH 5 (pH 5), LB pH 6 (pH 6), LB pH 7 (pH 7), LB pH 8 (pH 8), TSB + 1mM EDTA (EDTA) et TSB + 1 mM EGTA (EGTA). **B** : Les milieux testés sont le M9 glucose (Glc), M9 dextrose (Dex), M9 maltose (Malt), M9 fructose (Fru), M9 pH 7 (pH 7), M9 pH 8 (pH 8), M9 alanine (Ala), M9 asparagine (Asn), M9 cystéine (Cys), M9 glutamine (Gln), M9 glycine (Gly), M9 histidine (His), M9 isoleucine (Iso), M9 leucine (Leu), M9 lysine (Lys), M9 méthionine (Met), M9 phénylalanine (Phe), M9 proline (Pro), M9 sérine (Ser), M9 thréonine (Thr), M9 tryptophane (Trp), M9 valine (Val), M9 casamino acid (CAA), M9 + 0,1 M NaCl (0,1 M NaCl), M9 low phosphate (Low P) et SCFM.

7.12 Annexe 12 : Energie libre *riboswitch* fluor

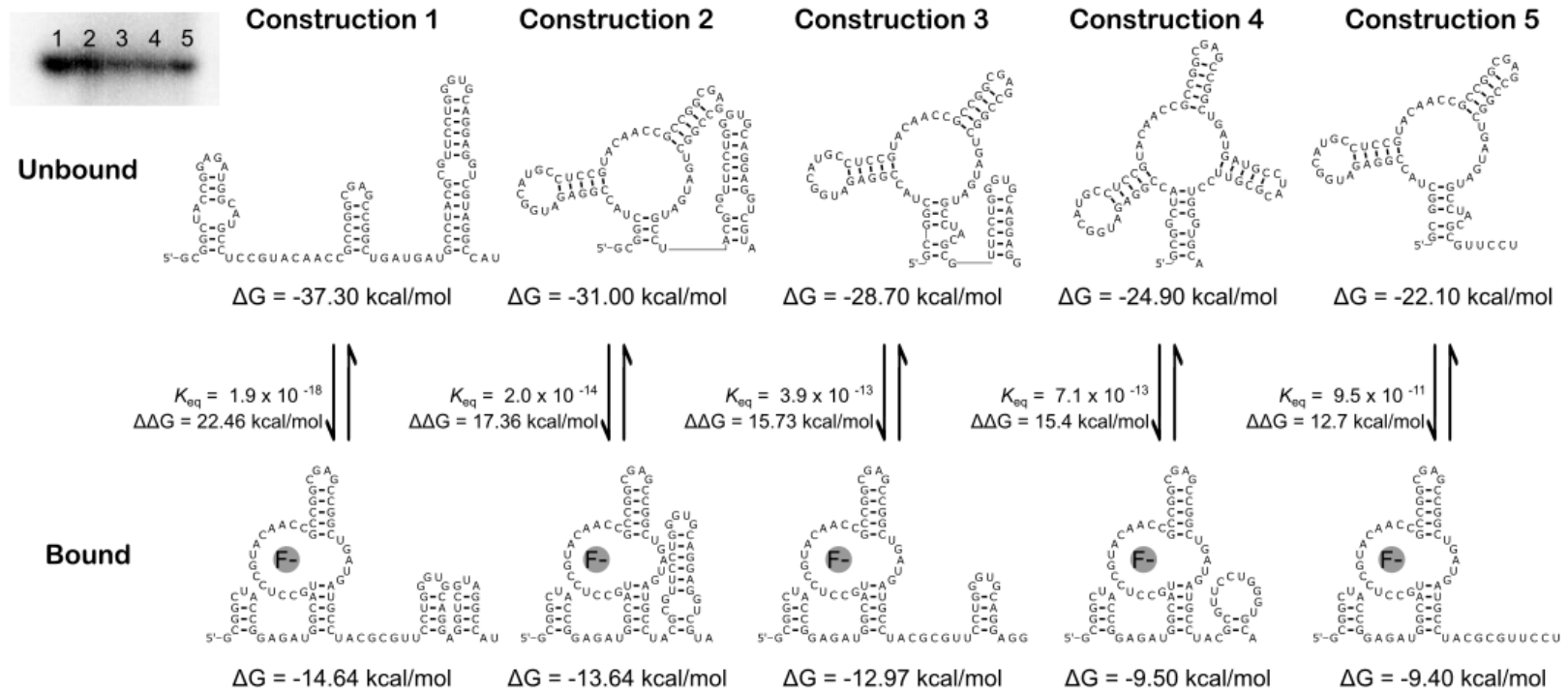


Figure 7.11 Énergie libre des différentes constructions du riboswitch fluor utilisé lors des essais de validation du SR-PAGE dans leur conformation liée et non liée.

Les structures liées (bond) représentées ont été obtenues en utilisant la plateforme Mfold en utilisant les contraintes décrites dans le Tableau 4.4. Ces contraintes ont été réalisées en suivant la structure connue de la forme liée de l'aptamère basée sur des études de résolution atomique disponibles pour chaque aptamère, cela avec les limites de Mfold (exclusion des paires de bases non canoniques). Concernant les structures non liées (unbound), elles correspondent à la structure présentée sans contrainte dans Mfold et présentant l'énergie libre minimale (MFE). Les structures ont par la suite été réalisées avec l'outil R2R.

7.13 Annexe 13 : Énergie libre *riboswitch* FMN

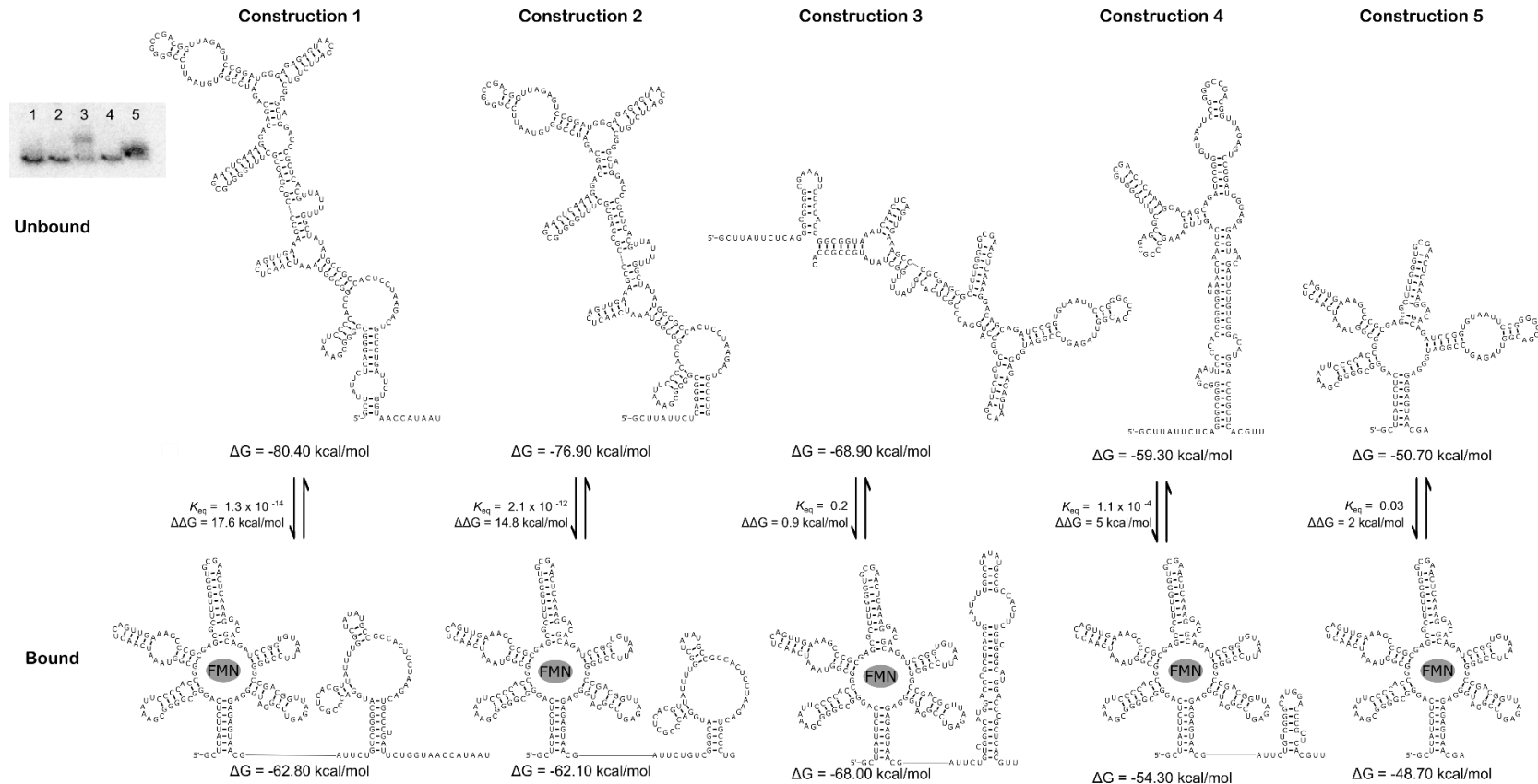


Figure 7.12 Énergie libre des différentes constructions du *riboswitch* FMN utilisé lors des essais de validation du SR-PAGE dans leur conformation liée et non liée.

Les structures liées (bond) représentées ont été obtenues en utilisant la plateforme Mfold en utilisant les contraintes décrites dans le Tableau 4.4. Ces contraintes ont été réalisées en suivant la structure connue de la forme liée de l'aptamère basée sur des études de résolution atomique disponibles pour chaque aptamère, cela avec les limites de Mfold (exclusion des paires de bases non canoniques). Concernant les structures non liées (unbound), elles correspondent à la structure présentée sans contrainte dans Mfold et présentant l'énergie libre minimale (MFE). Les structures ont par la suite été réalisées avec l'outil R2R.

7.14 Annexe 14 : Énergie libre *riboswitch* di-GMPc

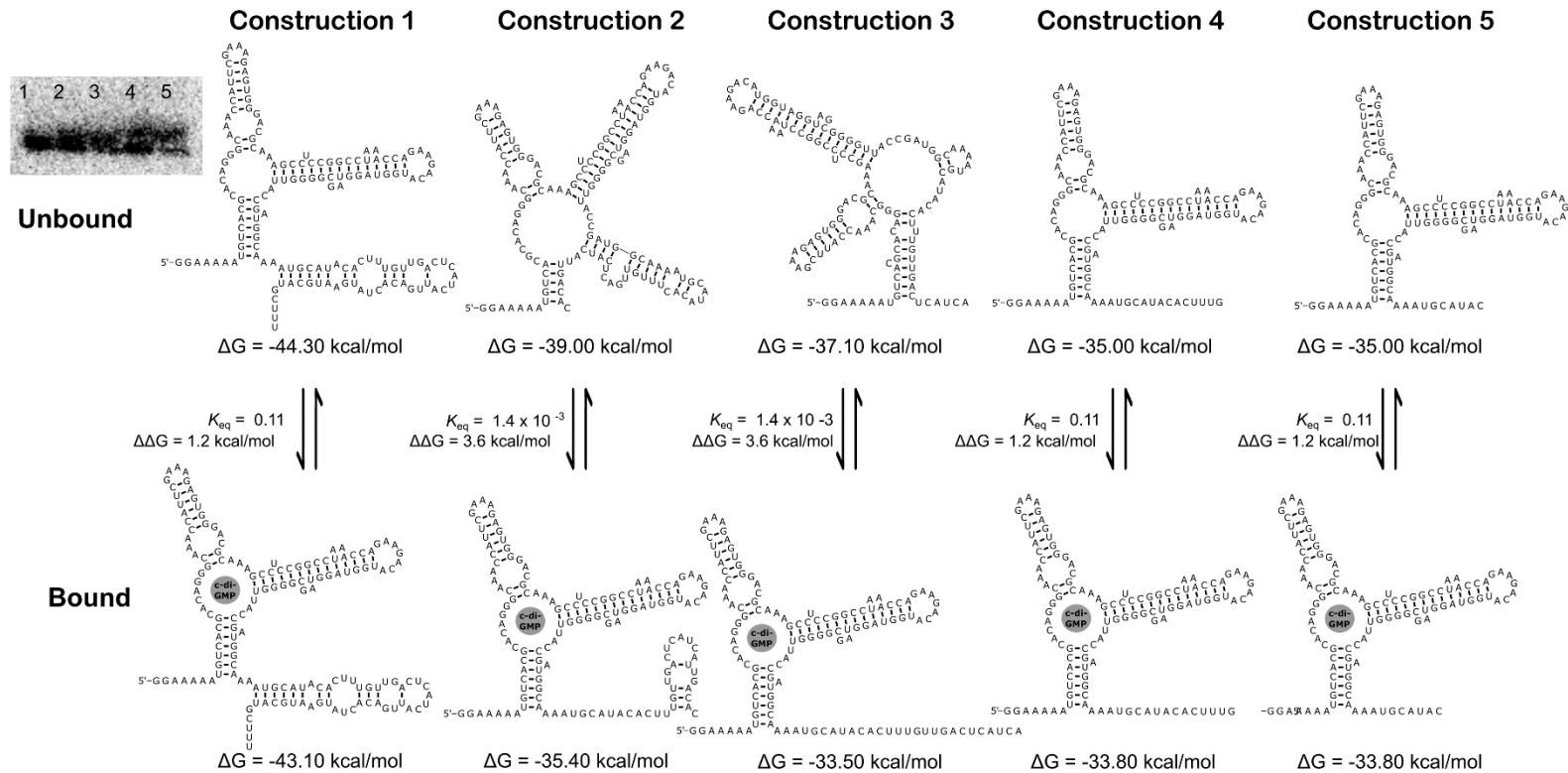


Figure 7.13 Énergie libre des différentes constructions du *riboswitch* di-GMPc utilisé lors des essais de validation du SR-PAGE dans leur conformation liée et non liée.

Les structures liées (bond) représentées ont été obtenues en utilisant la plateforme Mfold en utilisant les contraintes décrites dans le Tableau 4.4. Ces contraintes ont été réalisées en suivant la structure connue de la forme liée de l'aptamère basée sur des études de résolution atomique disponibles pour chaque aptamère, cela avec les limites de Mfold (exclusion des paires de bases non canoniques). Concernant les structures non liées (unbound), elles correspondent à la structure présentée sans contrainte dans Mfold et présentant l'énergie libre minimale (MFE). Les structures ont par la suite été réalisées avec l'outil R2R.

7.15 Annexe 15 : Énergie libre *riboswitch* Nickel/Cobalt

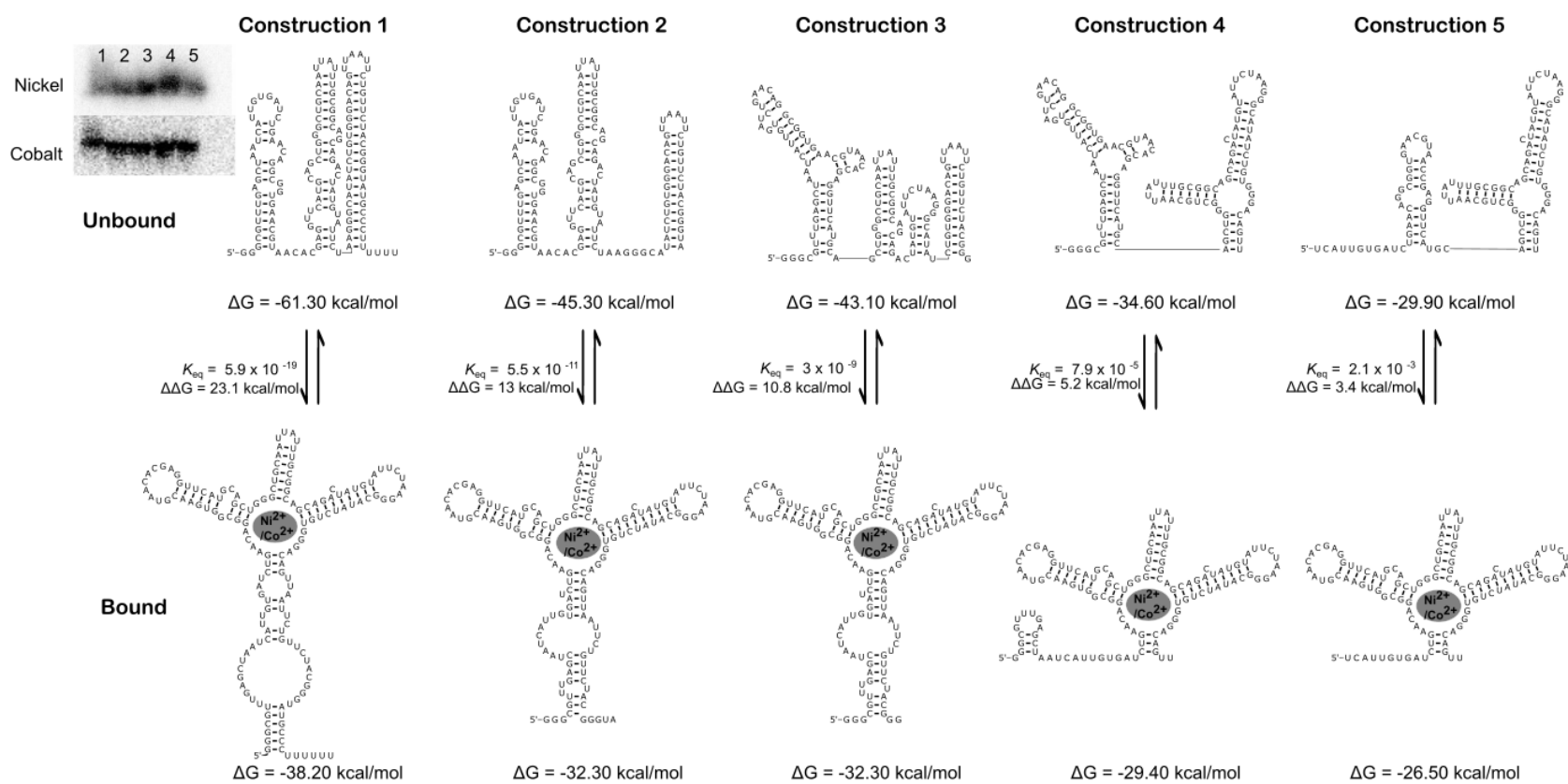


Figure 7.14 Énergie libre des différentes constructions du *riboswitch* Nickel-cobalt utilisé lors des essais de validation du SR-PAGE dans leur conformation liée et non liée.

Les structures liées (bond) représentées ont été obtenues en utilisant la plateforme Mfold en utilisant les contraintes décrites dans le Tableau 4.4. Ces contraintes ont été réalisées en suivant la structure connue de la forme liée de l'aptamère basée sur des études de résolution atomique disponibles pour chaque aptamère, cela avec les limites de Mfold (exclusion des paires de bases non canoniques). Concernant les structures non liées (unbound), elles correspondent à la structure présentée sans contrainte dans Mfold et présentant l'énergie libre minimale (MFE). Les structures ont par la suite été réalisées avec l'outil R2R.

7.16 Annexe 16 : Clones du *riboswitch* TPP

Tableau 7.2 Récapitulatif des clones sélectionnés lors du SELEX avec le *riboswitch* TPP.

Clone issu de	Numéro du clone	Lib. (Fig. 4.3A)	Sélectionné dans la bande	K_D Thiamine (μ M)	K_D TPP (μ M)	Ratio K_D^T/K_D^{TPP}	
Illumina	1	1	2	17,7	67,1	0,26	
	Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUAUCACCUCGCCUGGAUAA UGCCAGCCUAGGGAAGUUCACGGACCAC						
	2	1	2	16,5	18,5	0,88	
	Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUAUCACCUGACCUGGAUAA UGCCAGACAAGGGAAGUUCACGGACCAC						
	3	2	2	/	/	/	
	Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUAUCACCUUGUUUGGAUAA UGCCAUGCUCGGGAAGUUCACGGACCAC						
	4	2	2	22,9	21	1,09	
	Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUAUCACCUUGUUUGGAUAA UGCCAUGCUCGGGAAGUUCACGGACCAC						
	5	2	2	/	/	/	
	Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUAUCACCGUUUUGGAUAA UGCCAUGCUCGGGAAGUUCACGGACCAC						
	6	2	2	/	/	/	
	Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUAUCACCUGCUUUGGAUAA UGCCAUGGUUGGGAAGUUCACGGACCAC						
	7	2	1	/	/	/	
	Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUAUCACCUCGUCUGGAUAA UGCCAGCGAAGGGAAGUUCACGGACCAC						
8	2	1	/	/	/		
Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUAUCACCC GAGUUUGAUAAUGCCAGAGCAGGGAAGUUCACGGACCAC							
	9	1	0	34,12	34,55	0,99	
	Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCUGAAGGCUGAGAAAUACCCGUAUCACCUGUACUGGAUAAUG CCAGAAUAGGGAAGUUCACGGACCAC						
	10	1	0	/	/	/	

Gen. 4	Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUAUCACCUACACUGGAAAU GCCAGACAAGGGAAGUUCACGGACCAC					
	11	1	0	/	/	/
	Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUAUCACCUUUCUGGUAUA UGCAGUAUAGGGAAGUUCACGGACCAC					
	12	1	1	/	/	/
	Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUAUCACCUAGUCUGGAAAU GCCAGAUGAGGGAAGUUCACGGACCAC					
	13	1	1	3	7,8	0,38
	Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUAUCACCUUACUGGUAUA UGCCAGCGAGGGAAGUUCACGGACCAC					
	14	1	1	14,86	13,46	1,1
	Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUAUCACCUUCUGCGGUAUA UGCCAGGAUAGGGAAGUUCACGGACCAC					
	15	1	2	/	/	/
	Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUAUACCUUGUCUGGUAUAUG CCAGUUCAGGGAAGUUCACGGACCAC					
	16	1	2	/	/	/
	Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUAUCACCUAUUCUGGUAUA UGCCAGUAAGGGAAGUUCACGGACCAC					
	17	1	2	/	/	/
	Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUAUCACCUAGACUGGUAUAUG CCAGUAUAGGGAAGUUCACGGACCAC					
	18	2	0	58	50,4	1,15
	Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUAUCACCUUCGCUUGGAUAU GCCACAUGAGGGAAGUUCACGGACCAC					
	19	2	0	45	33,1	1,36
	Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUAUCACCUUCGCUUGGAUAU GCCACAUGAGGGAAGUUCACGGACCAC					
20	2	0	/	/	/	
Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUAUCACCUUGGAUUGGAUA UGCCAAGUUGGGAAGUUCACGGACCAC						
21	2	1	/	/	/	
Séquence d'ARN:						

GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUAUCACCUGUGCUGGAUAA UGCCAGAGUAGGGAAGUUCACGGACCAC					
22	2	1	/	/	/
Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGCUGAGAAAUACCCGUAUCACCCAUGAUGGAUAAU GCCAAUAAGGGGAAGUUCACGGACCAC					
23	2	1	12,5	40,2	0,31
Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUAUCACCUAAAUUGGAUAA UGCCACUAGAGGGAAGUUCACGGACCAC					
24	2	2	17,53	20,11	0,87
Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUAUCACCUAUUUGGAUAA UGCCACCUAAGGGAAGUUCACGGACCAC					
25	2	2	3,07	5,1	0,6
Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUAUCACCAAGGAUGGAUAA UGCCACUAAGGGGAAGUUCACGGACCAC					
26	2	2	/	/	/
Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUAUCACCUAAACUGGAUAA UGCCAUAAAGGGGAAGUUCACGGACCAC					
27	1	0	/	/	/
Séquence d'ARN: GGACUCGGGGUGCCCUUCUGUGAAAGAAGAGAAAUACCCGUAUCACCUUUGCUGGAAAUGU CAGCUUAGGGAAGUUCACGGACCAC					
28	1	0	/	/	/
Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCCGAAGGCUGAGAAAUACCCGUAUCACCUAGACUGGAUAAU CCAGCCCAGGGAAGUUCACGGACCAC					
29	1	0	/	/	/
Séquence d'ARN: GGACUCGGGGUGCCCUUCUGUGCUGAGAAAGGAAUACCCGUAUACCUAUGCUGGAAAUGCC AGUCAGGGAAGUUCACGGAC					
30	1	1	/	/	/
Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCUGAAGGCUGAGAAAUACCCGUAUCACCUUGUCUGUAAUGC AGUCGAGGGAAGUUCACGGACCAC					
31	1	1	/	/	/
Séquence d'ARN: GGCUCGGGGUGCCCUUCUGUGAAGCUGAGAAAUACCCGUAUCACCUAAACUGGUAAUGCC AGUCUAGGGAAGUUCACGGACCAC					
32	1	1	/	/	/
Séquence d'ARN: GGACUCGGGGUGCCCUUCUGGUGAAGGUGAGAAAUCCUGUAUCACCUGAUCUGGAUAAUGC CGCCAGGGAAGUUCACGGACCAC					

Gen.10	33	1	2	/	/	/
	Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUACACCUACACUGGAUAAU GCCAGCUAAGGGAAGUUCACGGACCAC					
	34	1	2	69	39	1,77
	Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUUACACCU GAAGAUGUUUCUGCUCGGGGAAGUUCACGGACCAC					
	35	1	2	/	/	/
	Séquence d'ARN: GGACUCGGGGUGCCCUUCUGGUGAAGGUGAGAAAUCCUGUAUCACCUGAUCUGGAUAAUGC CGCCAGGGAAGUUCACGGACCAC					
	36	2	0	/	/	/
	Séquence d'ARN: GGACUCGGGGUGUCUUCUUCUGUGAGGCUGAGAAUAUCCGAUACCAGACUCC					
	37	2	0	/	/	/
	Séquence d'ARN: GGACUCGGGGUGGCGNCCUGCGUGAAGCUGAGAAUCUCGUUCUCUAAAGCCCN					
	38	2	0	/	/	/
	Séquence d'ARN: GGACUCGGGGUGUCCUUCUGCGUGAGGCUGAGAAUCNUCGUUCACUCCCGACCCNN					
	39	2	1	/	/	/
	Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUUACACCUUGGAUAA UGCCACUUGGGGAAGUUCACGGACCAC					
	40	2	1	/	/	/
	Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGCAUCACCAGCACUGGUAAU GCCAAUACUGGGAAGUUCACGGACCAC					
	41	2	1	71	168	0,42
	Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUUUAUACCGCGUGUGGAUAA UCCAGCGUCGGGAAGUUCAC					
	42	2	2	60	140	0,43
	Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUUACACCAACGAACGCUAG AACGUUGGGAAGUUCACGGACCAC					
	43	2	2	6,1	6,7	0,91
	Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUUACACCAAGGAUAA UGCCAGGCUGGGAAGUUCACGGACCAC					
	44	2	2	8,9	25	0,36
	Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUUACACCC AAACUGGAUAAUGCCAUCCGGGGGAAGUUCACGGACCACAAUCAC					

Illumina	45	3	2	3,7	1000*	0,004
	Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUAUCACCAA ACACUGCUUGUUGUUUGGGAAGUUCACGGACCAC					
	46	3	2	77	262	0,29
	Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUAUCACCAA CAAUGGAAGAUGUGUUGGGAAGUUCACGGACCAC					
	47	3	2	5,89	1000*	0,005
	Séquence d'ARN: GGACUCGGGGUGCCCUUCUGCGUGAAGGCUGAGAAAUACCCGUAUCACCUA GCAACAAGUGUUGGUAGGGAAGUUCACGGACCAC					

Tableau répertoriant toutes les séquences sélectionnées lors du SELEX avec le *riboswitch* TPP modifié testées par la méthode de cartographie *in vitro* par *in-line*. Pour certaines séquences le symbole « / » est présent dans la colonne correspondant au K_D , pour ces clones le calcul du K_D n'a pas été possible. Pour les clones 45 et 47 (représenté par un astérisque), aucune modulation n'était visible pour le TPP. Nous avons indiqué une concentration de 1mM afin de calculer le ratio car il s'agit de la concentration la plus haute testée.

7.17 Annexe 17 : structure librairie 3 du *riboswitch* TPP

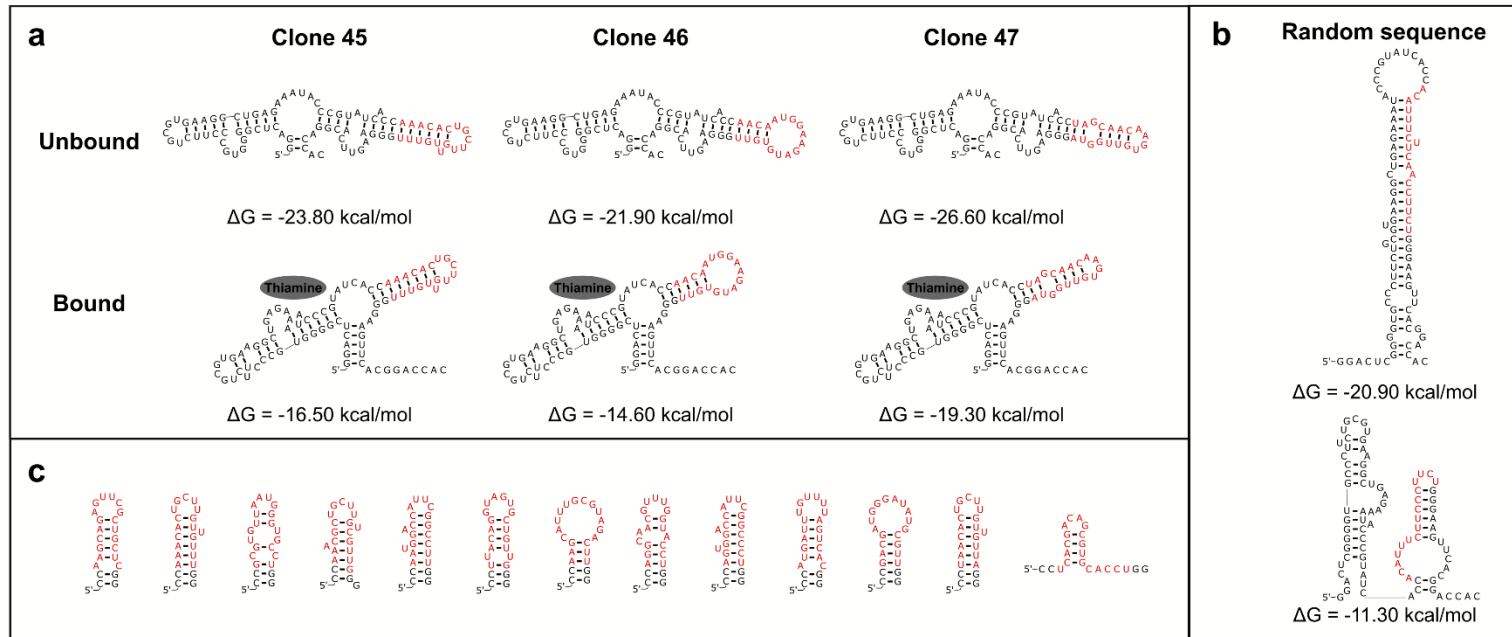


Figure 7.15 La librairie 3 du *riboswitch* TPP utilisée lors du SELEX a permis de sélectionner des ARN présentant une structure tige-boucle.

A : Les structures secondaires des clones ont été réalisées comme pour les annexes 9 à 12 en contraignant uniquement la tige P2-P3 dans laquelle se trouve la poche de liaison de la thiamine. Les trois clones issus de la librairie 3 montrent une structure globale liée / non liée semblable en formant une tige au niveau des régions aléatoires représentées en rouge. **B** : Représentation d'une structure aléatoire que l'on aurait pu retrouver dans la librairie de départ avec les mêmes contraintes qu'en A et C. **C** : La représentativité des structures a été évaluée pour quelques séquences retrouvées sélectionnées dans la librairie trois présentant plus de quatre lectures, la majorité des séquences peuvent former des tiges. A l'inverse, seule une minorité de séquences aléatoires peuvent former des tiges (Annexe 15).

7.18 Annexe 18 : prédiction de structure librairie 3

Tableau 7.3 Prédiction de structure tige-boucle dans la librairie 3 et dans une librairie aléatoire.

Seq ID	Sequence (5'-3')	ΔG value	GC% (within the 18 bases)	Final GC pairing occurs or not (position)
>22_1	CCAAGCAGAGTTCGCTGCTCGG	-7.80	56%	Yes (1,22)
>6_2	CCAAACACTGCTTGTTGTTTGG	-6.30	33%	Yes (1,22)
>6_3	CCGCGTGTTAATGGGTGCCTGG	-2.60	56%	Yes (1,22)
>5_5	CCAAACGCTGCTTGTCGTTGGG	-4.40	50%	Yes (1,21)
>5_6	CCAATGGACCATTTCGGCCTTGG	-6.70	50%	Yes (1,22)
>5_4	CCTTACAGGTAGTGCTGTTGGG	-6.10	44%	Yes (1,22)
>4_11	CCAAGCATTTCGCTAGACTTGG	-2.80	39%	Yes (1,22)
>4_9	CCAGGCAACGTTTGGTACCTGG	-6.10	50%	Yes (1,22)
>4_7	CCAGTGGACCATTTCGGCCCTGG	-9.10	61%	Yes (1,22)
>4_8	CCATGATTTGTTTTAGTCACGG	-4.10	28%	Yes (1,22)
>4_10	CCGACGATGGGATATGCGTTGG	-5.40	50%	Yes (1,22)
>4_13	CCTAACACTGCTTGTTGTTAGG	-6.70	33%	Yes (1,22)
>4_12	CCTCACGACAGCGTGCACCTGG	-2.90	61%	No
>3_17	CCTATGATTGCTCAGTCGTAGG	-9.50	39%	Yes (1,22)
>3_22	CCTATGCATCCGAGACACAAGG	0.30	44%	Yes (1,22)
>3_14	CCTCCATCAGACGGGATGGCGG	-8.20	61%	Yes (1,22)
>3_23	CCTCTGATGAGTGTGAGAAGGG	-8.40	44%	No
>3_28	CCTGATAGACTAGTGCTAAGGG	-1.70	39%	No
>3_18	CCTGATTAGCACTGCGAAATGG	0.30	39%	Yes (1,22)
>3_33	CCTTATGTTGGAGTAGCTAGGG	-3.50	39%	Yes (1,22)
>3_29	CCTTTTGTTGCGCAGTTTTAGG	0.00	33%	Yes (1,22)
>3_69	CCAACAATGGAAGATGTGTTGG	-4.40	33%	Yes (1,22)
>3_71	CCAACACGTGTCTGATTGTTGG	-4.30	39%	Yes (1,22)
>3_73	CCAACAGTGTCTCGCTGCTTGG	-7.10	50%	Yes (1,22)
>3_75	CCAACGAATTGATGTCGGTTGG	-3.90	39%	Yes (1,22)
>3_77	CCAAGATTCGCACCGATCTTGG	-7.90	44%	Yes (1,22)
>3_79	CCAATCATGGAAGATGGGTTGG	-4.10	39%	Yes (1,22)
>3_81	CCAATGCTCTTGACGGTGTGG	-6.70	44%	Yes (1,22)
>3_83	CCACCAGTTTACGGCTGGCGGG	-10.00	61%	No
>3_85	CCACCTCATGGGTATGAGGGGG	-9.70	56%	Yes (1,22)

>3_87	CCACGAAGAACAGTGTGCGAGGG	-2.90	50%	No
>3_89	CCACGCACTGTTGGGGTGCTGG	-8.20	61%	Yes (1,22)
>3_91	CCACGGAATGGCTAACCGCAGG	-1.60	56%	Yes (1,22)
>3_93	CCACTCATGGCTCATGTGTTGG	-3.70	44%	Yes (1,22)
>3_95	CCACTGATTCTTGTTGAAAGGG	-1.10	53%	No
>3_97	CCAGACAGGATGCTGTCATGGG	-9.60	61%	No
>3_99	CCAGACTGCTGTTGCCGCCTGG	-5.30	70%	Yes (1,22)
>3_101	CCAGATCTACCGTTGGATTTGG	-8.50	57%	Yes (1,22)
>3_103	CCAGCCCTACTTTGTCATATGG	0.50	56%	Yes (1,22)
>3_105	CCAGCGGCTGCTGCATGCCTGG	-6.60	74%	Yes (1,22)
>3_107	CCAGGCCACTTCTGTTACCTGG	-7.50	50%	Yes (1,22)
>3_109	CCAGGCTGGAGCCCTTGCTTGG	-8.40	61%	Yes (1,22)
>3_111	CCAGGTGCGGTCTAAAACCAGG	-2.60	50%	Yes (1,22)
>3_113	CCCAAATGCCAATTGGCCAGGG	-6.50	50%	Yes (1,22)
>3_115	CCCCATGCGTTACGTTTGGAGG	-5.70	50%	Yes (1,22)
>3_117	CCCCTACAAGAGACTGCTGGGG	-6.80	56%	Yes (1,22)
>3_119	CCCTACCATCTGCTTGCGAGGG	-4.70	56%	Yes (1,22)
>3_121	CCCTACCTGAGCTAGTTAAGGG	-2.60	44%	Yes (1,22)
>3_123	CCCTTGCCGCTGGTATGTCGGG	-3.70	61%	Yes (1,22)
>3_125	CCCTTGACACTTCCGGGGGGG	-12.20	61%	No
>3_127	CCGAAGGCTTGTGCTTAGTAGG	-3.20	44%	Yes (1,22)
>3_129	CCGCTTGATCCGTATATGCCGG	-0.90	50%	No
>3_131	CCGGACAGCACTCGCTACCCGG	-8.60	67%	Yes (1,22)
>3_133	CCGGAGCATTCTGTGCCCCGG	-10.80	67%	Yes (1,22)
>3_135	CCGGCTGTTCTGTACCGCCCGG	-5.60	67%	Yes (1,22)
>3_137	CCGGGATCGCTCGACGCCCCGG	-8.30	78%	Yes (1,22)
>3_139	CCGTCAGCCACGTGCTTAAGG	-2.70	56%	Yes (1,22)
>3_141	CCGTTGATGCTTTCGCAACCGG	-3.30	50%	Yes (1,22)
>3_143	CCTAAGCCCATCCAACCTTAGG	-3.90	44%	Yes (1,22)
>3_145	CCTAAGCGTTGTCGCGCCTGGG	-8.10	61%	Yes (1,22)
>3_147	CCTAGATGCTCTTAGCATGGGG	-5.80	44%	Yes (1,22)
>3_149	CCTAGCGGTTTGCAAAGTCGGG	-2.60	50%	Yes (1,22)
>3_151	CCTAGTTCTGGATCGCACTGGG	-4.30	50%	Yes (1,22)
>3_153	CCTATCGCTGAGTGCGCCTGGG	-5.60	61%	Yes (1,22)
>3_155	CCTCCACGCTCGGATTACCCGG	-3.00	61%	No

>3_157	CCTCTTATTTACCGAAGGAGG	-4.50	39%	Yes (1,22)
>3_159	CCTGAACCAGACGCGAATTTGG	-2.50	44%	No
>3_161	CCTGAAGGTAGGAATACTAGGG	-3.60	39%	No
>3_163	CCTGAATGCACGTAGCTCATGG	-2.10	44%	Yes (1,22)
>3_165	CCTGACCATCTGGGTTGTTGGG	-5.40	50%	Yes (1,22)
>3_167	CCTGACTTGCGTGTACCTTGGG	-1.50	50%	Yes (1,22)
>3_169	CCTGAGATTTGCCGTCTCGAGG	-6.50	50%	Yes (1,22)
>3_171	CCTGCGAATGCGCGTTCATGGG	-4.90	56%	Yes (1,22)
>3_173	CCTGCTCGTTGGTAGTGCCAGG	-3.10	56%	No
>3_175	CCTGGACGTTTTACACGCTGGG	-6.20	50%	Yes (1,22)
>3_177	CCTGGTGTTTGGCTATGCTGGG	-6.50	50%	Yes (1,22)
>3_179	CCTGTGCGATGTTGTCGATGGG	-10.70	50%	Yes (1,22)
>3_181	CCTGTTAGATGTGCTCTACAGG	-6.40	39%	Yes (1,22)
>3_183	CCTGTTACGGGCGTGCGTAGG	-6.60	61%	Yes (1,22)
>3_185	CCTTCACTATTGCTGGGATGGG	-3.20	44%	No
>3_187	CCTTCCACACTGCCGCACCTGG	0.50	61%	No
>3_189	CCTTCCGGAGAGGAGCCATCGG	-3.00	61%	No
>3_191	CCTTGCACAATGCTGTGCGGGG	-12.00	56%	Yes (1,22)
>3_193	CCTTGTCATCTGCTGTGCTGGG	-1.80	50%	Yes (1,22)
>3_195	CCTTGTTGCAATTGGCGCAGGG	-7.40	50%	Yes (1,22)
>3_197	CCTTTGCGCGTTGGGTGCCTGG	-5.00	61%	Yes (1,22)
Random Seq ID	Random sequence (5'-3')	ΔG value	GC% (within the 18 bp)	Final GC pairing occurs or not (position)
1.	ccTTCCCTTGCATATATGTTgg	-0.90	33%	No
2.	ccACATTTCTTCAACCTTCTgg	-0.20	33%	No
3.	ccAATTGCACCCTTAGGACGgg	-2.60	50%	No
4.	ccAAGACAGATATGTTCTTAgg	-3.00	28%	Yes (1, 22)
5.	ccCCTATATTTATCATTGGgg	-5.00	33%	Yes (1,22)
6.	ccCAACGGGATCGCATGTCCgg	-5.90	61%	No
7.	ccCACGTAAAACATTGTAAgg	1.00	28%	No
8.	ccACCCTCAGTTTTTGAGCgg	-4.90	50%	No
9.	ccGACAAAACTTTAAAAAGgg	0.30	22%	No
10.	ccAAATTCGCGCTCATAACTgg	1.10	39%	Yes (1,22)
11.	ccGTTAGGCCACGATTGCGTgg	-5.90	56%	No
12.	ccGAGTTTCGGCCCTGTGCTgg	-3.50	61%	No

13.	ccGCGCTGTATAGCCGATTCgg	-3.40	56%	No
14.	ccTCATTCGGGCCTTATATCgg	-0.60	44%	No
15.	ccTGGAAACCCCAACCTATTgg	-1.70	44%	No
16.	ccTAGACAGCATCATTGGCCgg	0.10	50%	No
17.	ccGAAGTTATTGGGCATATTgg	-2.00	33%	No
18.	ccCACCGTAAAGTCCTCCTCgg	-1.00	56%	No
19.	ccGGGCGTCCCTCCTTTAAAgg	-3.30	56%	No
20.	ccAGATGATAAGCTCCGGCAgg	-1.50	50%	No
21.	ccAAGGATCGGTGATATTAagg	0.50	33%	No
22.	ccCAAAGATTCGGCACATTAgg	0.10	39%	No
23.	ccCTCTTGTTGGTGTGGTATgg	-0.30	44%	No
24.	ccCGCTTAACTGCGTGGCGGgg	-10.20	67%	No
25.	ccAGCCTTATGGCAAATCGgg	-5.00	44%	No
26.	ccTTCGGGAATGATTCTGGTgg	-5.80	44%	Yes (1,22)
27.	ccAACGCTAAAGGTCCATAGgg	0.40	44%	No
28.	ccCACATACATCGCAACCTGgg	-2.00	50%	Yes (1,22)
29.	ccGCATGCGTTCAATTTGACgg	-2.00	44%	No
30.	ccGATCGCTTGGCGCTAAGAgg	-1.10	56%	No
31.	ccTTAAAGCGGCTGCACTGCgg	-4.30	56%	No
32.	ccTGTAAGGACGATTACGGAagg	-5.00	44%	No
33.	ccGTGGGCGGCCTGGGGGGAagg	-5.00	83%	No
34.	ccGCACTACCCCATCGACCTgg	-0.60	61%	No
35.	ccGTACAGGAACACTCTATAgg	-1.00	39%	No
36.	ccTTGCTCTCAGACGAACAagg	-5.70	44%	Yes (1,22)
37.	ccATTACTAGAGTGCCGCTTgg	-0.50	44%	No
38.	ccTCAGCCCCCTGTCGTGCGgg	-3.00	72%	No
39.	ccCGTTGTTGTGATTGACTgg	-3.70	44%	Yes (1,22)
40.	ccCTATTGAGGCATCAACTGgg	-6.60	44%	Yes (1,22)
41.	ccAATGAATCGGCCTATGTCgg	-3.00	44%	No
42.	ccCCCGATGTCGTTAGTGAagg	0.20	50%	No
43.	ccGGTTCCGACGCATACCTCgg	-2.80	61%	No
44.	ccCTTCGTTGAGAACCCACAagg	0.00	50%	No
45.	ccATCATACAACCTGGGGACAagg	-5.80	44%	No
46.	ccTAATCCCTACGCCATCAagg	0.40	50%	No
47.	ccTCTACACGCGTCTCTGTGgg	-3.40	56%	No

48.	ccGCTCCAGTTCATGTGCTGgg	-6.80	56%	No
49.	ccGGAGAGCACCCCTCCACAAgg	-5.90	61%	No
50.	ccGGTCTAGTGGTATGGTGGgg	-3.10	56%	No
51.	ccTGATACACGCGGCAGGGGgg	-3.60	67%	No
52.	ccTAGGACCATCGGTAGTAGgg	-6.80	50%	No
53.	ccCTGACCACTGCCTATAGGgg	-4.20	56%	No
54.	ccAGAGTGTGTCAGCCAGTGTAgg	0.40	50%	No
55.	ccACCCACGAGGATCCGAGgg	-5.00	67%	No
56.	ccAAGGCGAACCGGGCCAGAgg	-4.30	67%	No
57.	ccCTCAAAGCCGCGGCGCGAgg	-7.60	72%	No
58.	ccAGTAGCCCCGGGGTGAACgg	-4.70	67%	Yes (1,22)
59.	ccACCTATGGGGCTGGATAAgg	-4.40	50%	No
60.	ccAACTGCCCTGGTGAAGCGgg	-4.10	67%	No
61.	ccTCTGCTGCTCGAGGCCGTgg	-4.00	67%	No
62.	ccTCGCCGATGCTTGCTGCGgg	-3.90	67%	No
63.	ccTCCCCAGCCGCTACATCTgg	-1.90	61%	No
64.	ccGTCTCTTTGCCGACTAATgg	-1.90	44%	No
65.	ccGCGAACAACCACACCATAgg	0.30	50%	No
66.	ccGCGATTTCGTCGGGGCGCCgg	-5.20	78%	No
67.	ccTCGGAATACGGTATGGGCgg	-1.60	56%	No
68.	ccTCGCGGACGCCAGGCATCgg	-4.50	72%	No
69.	ccGTGCAGGTAGCGGAGGCCgg	-4.20	72%	No
70.	ccCGCACGCGAGACGAACTGgg	-2.90	67%	Yes (1,22)

Tableau récapitulatif du calcul des ΔG moyen des structures sélectionnées dans la librairie trois montrant plus de quatre lectures lors de l'analyse du séquençage haut débit et de séquences aléatoires (telles qu'on les trouverait dans la librairie initiale avant la sélection par le SR-PAGE). Les séquences aléatoires d'une taille de 18 nts présentant deux paires de bases complémentaires CG (représentées en minuscule) ont été créés en utilisant le site Random DNA sequence Generator (<http://www.faculty.ucr.edu/~mmaduro/random.htm>) avec sept valeurs de GC% (0,35, 0,40, 0,45, 0,50, 0,55, 0,60 et 0,65). Les séquences ont été soumises à Mfold afin d'évaluer leur capacité à former des structures.

7.19 Annexe 19 : Artefact due à la contamination lumineuse

A

	1	2	3	4	5	6	7	8	9	10	11	12
A	blanc M9 - glucose			sucAlli WT M9 - glucose			sucAlli WT M9 - fructose			sucAlli WT M9 - maltose		
B	sucAlli WT M9 - sérine			sucAlli WT M9 - glutamine			sucAlli WT M9 - histidine			sucAlli WT M9 - glycine		
C	sucAlli WT M9 - methionine			sucAlli WT M9 - proline			sucAlli WT M9 - cysteine			sucAlli WT M9 - alanine		
D	sucAlli WT M9 - leucine			sucAlli WT M9 - threonine			sucAlli WT M9 - isoleucine			sucAlli WT M9 - lysine		
E	blanc M9 - glucose			sucAlli disr M9 - glucose			sucAlli disr M9 - fructose			sucAlli disr M9 - maltose		
F	sucAlli disr M9 - sérine			sucAlli disr M9 - glutamine			sucAlli disr M9 - histidine			sucAlli disr M9 - glycine		
G	sucAlli disr M9 - methionine			sucAlli disr M9 - proline			sucAlli disr M9 - cysteine			sucAlli disr M9 - alanine		
H	sucAlli disr M9 - leucine			sucAlli disr M9 - threonine			sucAlli disr M9 - isoleucine			sucAlli disr M9 - lysine		

B

<>	1	2	3	4	5	6	7	8	9	10	11	12
A	126	134	126	153	176	175	164	143	153	166	326	244
B	207	206	229	202	227	233	273	315	258	249	239	233
C	227	291	310	356	346	425	370	421	430	469	431	363
D	430	591	625	2260	2341	2396	675	781	852	2071	1533	1513
E	776	1013	1127	30559	29493	28178	30748	16043	34656	38839	36907	44472
F	77283	81729	84662	25981	43982	25034	33763	34068	40191	92424	100550	82745
G	29588	29421	29709	28055	15478	13528	34747	31698	31806	83981	93980	102120
H	71610	70502	70850	103620	91179	103462	66920	67631	62821	92640	99244	97450

Figure 7.16 La proximité d'échantillons exprimant un niveau de luminescence plus fort entraîne un artefact dans les lectures de luminescence.

A : Carte de plaque pour les constructions WT et disr du motif *sucA-II* présent chez *P. aeruginosa*. **B** : Lecture de la luminescence de la plaque après 5h d'incubation à 37°C en utilisant le M1000 infinite (TECAN®). Le rectangle cyan permet de voir que les puits à proximité de la construction disr ont un niveau de luminescence beaucoup plus élevé que les autres puits ceci uniquement dû à la contamination lumineuse avec les puits du dessous. Les puits D1 à D3 sont des blancs identiques aux puits C1 à C3 et on peut observer une luminescence beaucoup plus importante.

7.20 Annexe 20 : Article du SR-PAGE publié dans BiorXiv

Shifted Reverse PAGE: a novel approach based on structure switching for the discovery of riboswitches and aptamers.

Aurélie Devinck^{1*}, Emilie Boutet^{1*}, Jonathan Ouellet^{1,2}, Rihab Rouag¹, Balasubramanian Sellamuthu¹, Jonathan Perreault¹

*These authors have an equivalent contribution

¹ INRS – Centre Armand-Frappier Santé Biotechnologie, Laval, Qc, H7V 1B7, Canada,
² Now at Monmouth University, NJ, 07764, USA.

Corresponding author:

Dr. Jonathan Perreault

Tel: (450) 687-5010 ext: 4411

Fax: (450) 687-5010

E-mail: jonathan.perreault@inrs.ca

Keywords: native PAGE, gel shift, electrophoretic mobility shift assay (EMSA), biochemical screening, RNA structure, Rho-independent transcription terminator, thiamine pyrophosphate (TPP), thiamine, SELEX, riboswitch, aptamer

Abstract

Riboswitches are regulatory sequences composed of an aptamer domain capable of binding a ligand and an expression platform that allows the control of the downstream gene expression based on a conformational change. Current bioinformatic methods for their discovery have various limitations. To circumvent this, we developed an experimental technique to discover new riboswitches called SR-PAGE (Shifted Reverse Polyacrylamide Gel Electrophoresis). A ligand-based regulatory molecule is recognized by exploiting the conformational change of the sequence following binding with the ligand within a native polyacrylamide gel. Known riboswitches were tested with their corresponding ligands to validate our method. SR-PAGE was imbricated within an SELEX to enrich switching RNAs from a TPP riboswitch-based degenerate library to change its binding preference from TPP to thiamine. The SR-PAGE technique allows performing a large screening for riboswitches, search in several organisms and test more than one ligand simultaneously.

Introduction

Microbial gene expression has to be tightly regulated to adapt to constant changing environments (Hennecke, 1990). Cells are able to import molecules from the extracellular environment and their concentration can be crucial to modulate enzymatic pathways (Hennecke, 1990). A specific element of the mRNA mostly found in the 5' untranslated region (UTR) called riboswitch can bind these molecules (ligand) which leads to a change in their secondary structure. This change will affect transcription or translation of the downstream gene. The affected genes often encode for the synthesis or the transportation of the bound molecule (Serganov & Nudler, 2013). These riboswitches thus comprise an aptamer domain and an expression platform, the structure of which depends on the binding of the aptamer's cognate ligand. To date, more than 40 riboswitches specifically binding with different metabolites and ligands have been reported (Barrick *et al.*, 2004; Mandal *et al.*, 2003; Mandal *et al.*, 2004; McCown *et al.*, 2017; Ramesh & Winkler, 2010; Sherlock *et al.*, 2018a; Winkler *et al.*, 2003) in bacteria. Interestingly, the thiamine pyrophosphate (TPP) riboswitches were also observed in archaea, plants and fungi (McCown *et al.*, 2017).

The discovery of most riboswitches was made possible by the availability of sequenced genomes for numerous bacteria combined with powerful bioinformatic approaches (Stav *et al.*, 2019; Weinberg *et al.*, 2007; Weinberg *et al.*, 2017a; Weinberg *et al.*, 2017b; Weinberg *et al.*, 2010). Potential riboswitches are identified based on the presence of sequence covariation and conservation since the aptamer domain of these regulatory molecules is highly conserved. Covariation occurs when changes in the nucleotide sequence does not affect the RNA secondary structure, reinforcing the importance of its formation (Weinberg *et al.*, 2010). However, bioinformatic tools are limited by sequence availability and their annotations within databases. Also, an alignment of three sequences is theoretically enough to provide covariation that will support a secondary structure prediction and sufficient to warrant experimental evaluation of a putative riboswitch. However, in practice, alignments leading to discovering new riboswitches typically had dozens to hundreds of sequences. The identification of the ligand recognized by the aptamer domain is often facilitated by the nature of the downstream gene. Problematically, gene annotation is sometimes missing or misleading and does not always provide the necessary

information to identify the metabolite recognized by the riboswitch. A list of orphan riboswitches for which the associated ligands are currently unknown resulted from bioinformatics-based discovery of numerous putative regulatory RNA structures (Greenlee *et al.*, 2018; Stav *et al.*, 2019; Weinberg *et al.*, 2007; Weinberg *et al.*, 2017a; Weinberg *et al.*, 2017b; Weinberg *et al.*, 2010). From the current approach for riboswitch discoveries, it is expected that a very small number of all riboswitch classes have been unveiled (Breaker, 2011). The remaining riboswitches would unfortunately be scarcer, making them much harder to be discovered by bioinformatic tools. Thus, a method which does not rely on sequences available in public databases might be the best way to circumvent this problem.

Some experimental methods have been developed by other groups in recent years to discover new riboswitches. Term-seq enables to screen all the transcription termination within chosen organisms using a high-throughput sequencing approach (Dar *et al.*, 2016). This method can be used to identify regulations involving a premature stop of transcription as can be found with some riboswitches. However, this method omits all riboswitches that have an impact on translation. Also, Term-seq cannot differentiate transcription termination caused by a riboswitch from transcription termination caused by a protein, which could be derived from indirect regulation. Another experimental method that can be used to discover new riboswitches is Parallel Analysis of RNA Confirmations Exposed to Ligand Binding (PARCEL) (Tapsin *et al.*, 2018). It is an *in vitro* method to detect structural changes of natural RNA in presence of a ligand by comparing the targeted sites of the RNase V1, an enzyme that cleaves base-paired regions. A change in the degradation pattern of the RNase V1 between tested conditions is a good indicator that the ligand induces a change of conformation in the RNA, which is characteristic of riboswitch (Tapsin *et al.*, 2018). However, this technique is limited to regions that are accessible by RNase and only one genome can be analyzed at the time. Another technique derived from SHAPE-MaP (Siegfried *et al.*, 2014) was developed to investigate RNA molecules that interact with ligands. It was recently used to screen hundreds of ligands for their ability to bind RNA molecules (Zeller *et al.*, 2022). Nevertheless, it could also be used to screen for riboswitches as with PARCEL, but with similar limitations, except perhaps that changes in RNA secondary structure upon binding is resolved to the nucleotide level.

Aptamers for various ligands have successfully been identified based on SELEX (Selective Evolution of Ligands by Exponential Enrichment) (Irvine *et al.*, 1991), the first ones being RNA aptamers selected against T4 DNA polymerase (Tuerk & Gold, 1990) and various organic dyes (Ellington & Szostak, 1990). The general idea is to start with a mix of an extremely diverse library of sequences ($\sim 10^{16}$ sequences) and a target of interest. Unbound sequences are separated from those showing an affinity for the target by a selection method. Selected sequences are amplified for the next round of enrichment. Although the general idea remains the same, many selection methods have been used within a SELEX to isolate aptamers, including capillary electrophoresis (CE-SELEX) (Zhu *et al.*, 2019), capture-SELEX (Boussebayle *et al.*, 2019) and magnetic bead SELEX (Espelund *et al.*, 1990) to name a few (reviewed in (Bayat *et al.*, 2018; Darmostuk *et al.*, 2015)). We hypothesized that a SELEX-based approach could also be applied to discover riboswitches. Nonetheless, there are certain restrictions preventing the use of typical SELEX experiments to find natural riboswitches. These regulatory sequences are often engulfing their ligand, guaranteeing that any attempt to immobilize the ligand on a column would cause

steric hindrance. Even when riboswitches do not envelop their ligand, the chemical groups chosen for immobilization are likely to be among the groups bound by the potential riboswitch. As for methods that immobilize the library instead of the ligand (e.g., capture-SELEX (Boussebayle *et al.*, 2019)), it would be difficult to use them to find riboswitches because users need to define a capture sequence within the random region, which is not possible for a library containing unknown natural riboswitches.

We have developed a novel method capable of getting over these obstacles, addressing at the same time the disadvantages of other experimental techniques. This method relies only on the two main characteristics of riboswitches: their ability to bind a ligand and their conformational change upon such binding. These characteristics are evaluated by electrophoresis approach analogous to 2D gels, but instead of running through a second dimension, electrodes are reversed. During the second migration, the RNA runs backwards in presence of a ligand. This allows the detection of a conformational change of the RNA upon binding with a ligand compared to RNA that does not interact with one. We named our technique Shifted Reverse PolyAcrylamide Gel Electrophoresis (SR-PAGE). Our method was validated with known riboswitches by verifying the shifting ability of several constructions. The power of this method was also demonstrated by selecting for a thiamine aptamer from a degenerate library of TPP riboswitches with the SR-PAGE imbricated within a SELEX as a selection method.

Results

Validating SR-PAGE with known riboswitches

SR-PAGE can identify riboswitches by taking advantage of the change in the conformation of the RNA sequence following binding with its cognate ligand within a native polyacrylamide gel. Briefly, the method is composed of two native gel migrations: during the first one, the potential regulatory RNA sequences migrate from the wells towards the positive electrode getting separated based on their size and their structure (Fig. 1a). After 24 hours of migration, the gel is unmolded, and the solution of ligand(s) is sprayed on the gel to allow interaction with the RNA molecules (Fig. 1b). The second migration is then started with the same condition as the first migration, but the polarity of the gel is reversed, forcing all RNA sequences to migrate back to the starting point (Fig. 1c). Upon binding with the ligand, the RNA is subjected to a conformational change, either assuming a more compact form that would allow it to migrate pass the well line or creating a more relaxed conformation that would slow down its migration in the gel (Fig. 1c). SR-PAGE can recognize the ligand-based regulatory molecule by selecting those that had a change in migration after the addition of the ligand.

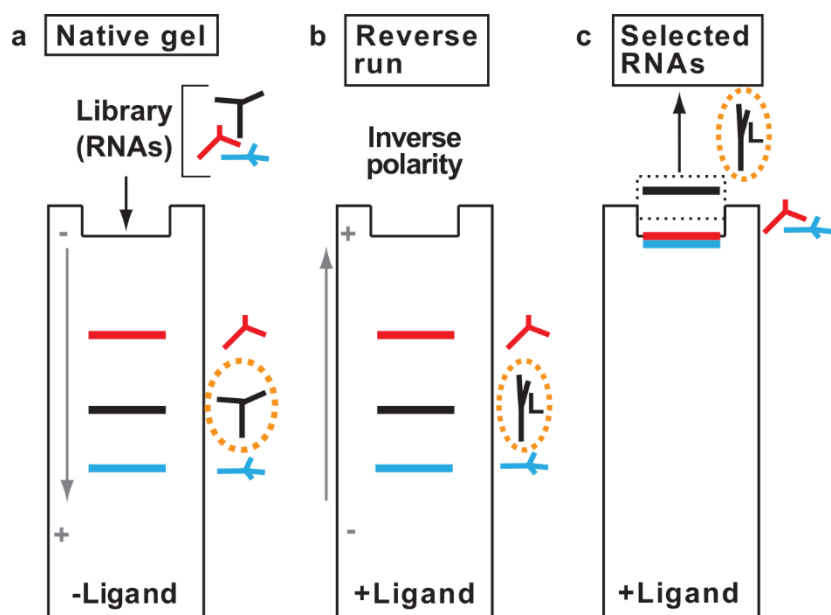


Fig. 17 | Overview of SR-PAGE. **a**, An RNA library is loaded on a native gel and migrated in absence of ligand. RNAs are separated based on structure and size. **b**, The top glass plate is taken off and the ligand is sprayed on the gel, which only changes the conformation of corresponding riboswitches (within the orange dotted circle) upon binding. The plate is put back on and migration is started with inverted polarity. **c**, The gel is run until the RNA comes back to its starting point. RNAs that interact with the ligand will have a change of migration due to a change in their secondary structure upon ligand binding. A more detailed schematic of the SR-PAGE method is available in the supplementary material (Supplementary Fig. 1).

As a proof of concept, we first wanted to validate that a change in migration upon binding of the ligand could be observed using SR-PAGE with known riboswitches. Up to five different constructions of six known riboswitches were designed, namely the known riboswitches for the following metabolites: flavin mononucleotide (FMN) (Pedrolli *et al.*, 2015), fluoride (Breaker, 2012), *c*-di-GMP (type I) (Sudarsan *et al.*, 2008), nickel-cobalt (NiCo) (Furukawa *et al.*, 2015), thiamine pyrophosphate (TPP) (Rentmeister *et al.*, 2007) and glycine (Kwon & Strobel, 2008). The sequence for the glycine riboswitch was based on results by Kwon *et al.* where a shift was observed in a native gel during migration with glycine (Kwon & Strobel, 2008). Their constructions varied in length, ranging from only the aptamer to incorporating different portions of the expression platform (Fig. 2 a,b,c; Supplementary Table 1).

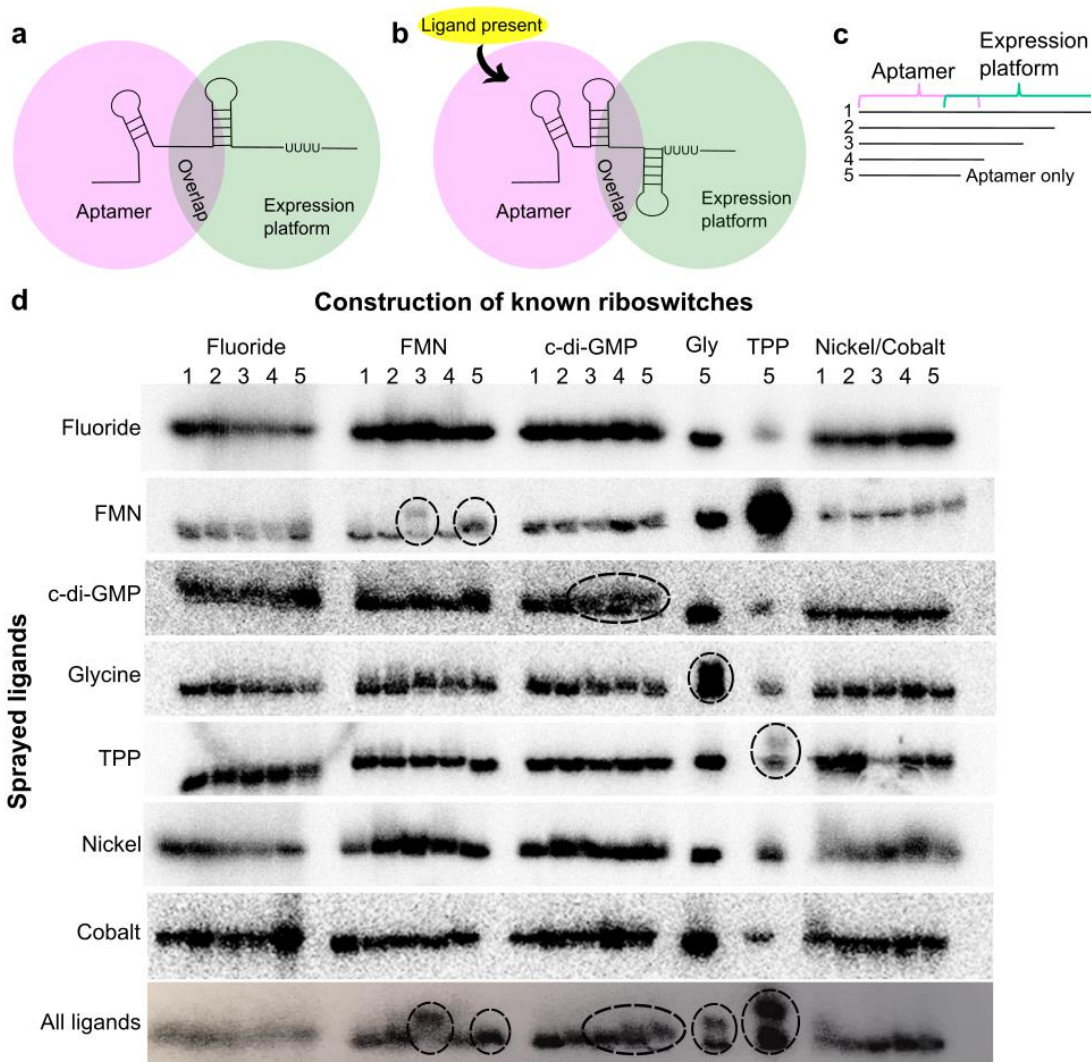


Fig. 2 | Validation of SR-PAGE with known riboswitches. **a**, Schematic representation of a riboswitch in its unbound conformation, with the aptamer domain and expression platform emphasized in pink and green respectively. The overlap between those two regions is shown. **b**, Schematic of a riboswitch in its bound conformation, where a ligand is bound to the aptamer domain, leading to a change in the secondary and tertiary structure. As an example, the change caused the formation of a Rho-independent transcription terminator in this case. **c**, Representation of how the different constructions (1 to 5) of the known riboswitches were created. **d**, Results from all SR-PAGE. Each group of horizontal bands corresponds to a different gel, where distinct ligands were sprayed into the gel before the second migration (only fluoride, FMN, c-di-GMP, glycine, TPP, nickel, cobalt or all ligands). Five different radiolabeled constructions for each known riboswitch were tested, except for the TPP and glycine riboswitch, where only the aptamer construction was assayed. Observable shifts in migration are circled.

Multiple SR-PAGE experiments were performed, where only one ligand at a time was sprayed at a concentration 100-fold higher than their respective known dissociation constant (K_D) (Fig. 2d).

Only the riboswitch constructions corresponding to the sprayed cognate ligands interacted with the metabolites, resulting in a shift of migration (circled in Fig. 2d). RNA molecules that did not interact with the sprayed ligands returned to the starting point in a straight line.

When FMN was sprayed, a significant change in migration for FMN_3 and a slight shift for FMN_5 constructs were observed (Fig. 2d). Small changes in migration for c-di-GMP_3, c-di-GMP_4 and c-di-GMP_5 constructs were also detected when the corresponding ligand was sprayed (Fig. 2d). A change in aptamer migration was also observed for the glycine and TPP aptamers (Fig. 2d) when the corresponding ligands were sprayed. Interestingly, the identical shift in migrations was observed when all ligands were sprayed at the same time on the gel, indicating that SR-PAGE can be used to select for multiple riboswitches with different ligand affinity at the same time (Fig. 2d).

SR-PAGE to investigate structure changes and expression platforms

We also hypothesized that SR-PAGE could be a means to investigate the expression platform of a riboswitch. It could be used to select for the construction that results in a bigger change in migration, implying it has a bigger conformational change. FMN_3 and FMN_5 were further analyzed to understand why different shift levels were observed between constructions of the same riboswitch: they technically both have similar ability to interact with the metabolite, since they are constituted of the same aptamer domain (Fig. 3). The Mfold web service (Zuker, 2003) was employed to analyze the free energy of both constructions in their bound and unbound forms. To mimic the bound conformation, sequences were forced to adopt the FMN aptamer domain structure (Serganov *et al.*, 2009) using constraints (constraints used to force the formation of the aptamer can be found in Supplementary material, Table 2) . The equilibrium constant was then measured from the difference in free energies between the two conformations.

For FMN_3, the RNA is only slightly more stable in its unbound conformation (unconstrained) compared to when constraints are used to force the aptamer conformation, with free energies of -68.9 kcal/mol and -68 kcal/mol respectively (Fig. 3a). Our interpretation is that during the first migration, the FMN_3 assumes its more stable unbound conformation (Fig. 3a, left) which can potentially interconvert with the aptamer-containing structure (Fig. 3a, right), with a preference of 5:1 for the former compared to the latter, according to the calculated equilibrium constant. However, when the FMN ligand is sprayed onto the gel, not only does the equilibrium change, but the “aptamer state” of the RNA (capable of interacting with the ligand) also likely has changes in its tertiary structure. The change in structure from the more stable one (on the left) to the aptamer state results in the shift of migration observed with SR-PAGE (Fig. 3a). Even if the so-called unbound state should be favored, since the free energy of both structures is similar, the presence of the ligand suffices to permit the switch. The same situation seems to apply for FMN_5 (Fig. 3b), with free energies of -50.70 kcal/mol and -48.70 kcal/mol for the unbound and bound conformations, respectively. However, the two structures appear closely related, which may explain the less prominent shift. Previous work also demonstrated that the aptamer of the FMN riboswitch is pre-formed in the absence of the metabolite at physiological concentrations of magnesium (Vicens *et al.*, 2011). In this case, in spite of a K_{eq} suggesting a ratio of ~40:1 of the most stable vs. the aptamer state, the closeness of both structures may permit an easier switch

in equilibrium, but also a more complete one, as suggested by the quasi-absence of a non-shifted band for FMN_5.

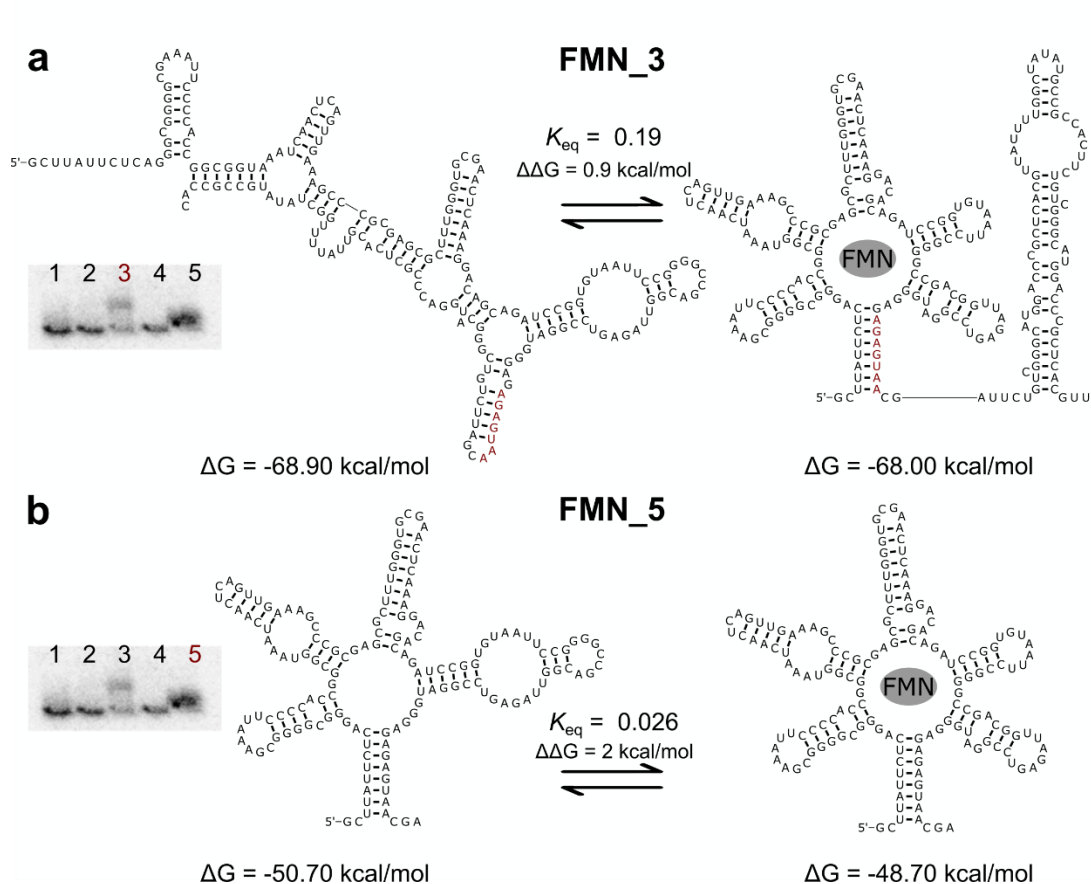


Fig. 3 | Shifting constructions 3 and 5 of the FMN riboswitch have similar free energies for their bound and unbound states. a,b Predicted structures, free energies (ΔG) and equilibrium constant (K_{eq}) of the unbound RNA and ligand-bound state of FMN_3 and FMN_5, respectively. Nucleotides emphasized in red can basepair to form the basal stem of the aptamer, leading to a mutually exclusive conformation. Shifts observed by SR-PAGE are shown, with corresponding construction number highlighted in red for a and b.

Both FMN_3 and FMN_5 structures are more stable in their unbound state, so the equilibrium constant favors in each case this conformation (in absence of ligand), represented by a K_{eq} value smaller than 1. Some of the RNA molecules were still able to adopt the ligand-binding conformation, since the K_{eq} value is close to 1. A similar analysis of the free energies and equilibrium constants of ligand-bound state compared to free RNA for all constructions of riboswitches FMN, fluoride, c-di-GMP and nickel-cobalt is available in the supplementary material (Supplementary Figs. 2–5). The larger differences in free energies observed for all the constructions of fluoride and nickel-cobalt riboswitches may explain that none of the tested constructs shifted (Supplementary Figs. 2 and 5). Moreover, the equilibrium constant for most of

these cases had very small values, meaning that the “aptamer state” is likely almost absent. As for c-di-GMP_3-4-5, FMN_3 and FMN_5, most free energies are very similar with K_{eq} closer to 1, which explains that the constructs all seem to shift, but the shifts are more subtle given the similarity between the predicted bound and unbound states.

Selection of thiamine aptamer from degenerate libraries of TPP riboswitch

As a proof of concept that the SR-PAGE can be used as a method of selection in a SELEX, the aptamer sequence from the TPP riboswitch was partially randomized to create three libraries possessing each 4,096, 1,048,576 and 68,719,476,736 possible sequence combinations (Fig. a). Libraries 1, 2 and 3 contained 6, 10 and 18 degenerated nucleotides respectively. Briefly, RNA libraries were run on native polyacrylamide gel electrophoresis during SR-PAGE. Thiamine was added to the second run (reversed polarity) of SR-PAGE, first at 2 mM (generations 1 to 4) and in later rounds of selection, at a lower concentration of 20 μ M (generation 5) and 200 nM (generations 7 to 10) to increase the stringency. Shifted RNAs were purified and amplified, then loaded on gel together with TPP (as a negative selection ligand) for the next round of selection, where thiamine is again added only for the reverse migration. We therefore selected for RNA molecules that preferentially bind thiamine rather than its native TPP. After four, five, seven and ten rounds of selection, selected sequences were cloned and also analyzed by Illumina sequencing (Bentley *et al.*, 2008) to identify the enriched sequences. RNAs were selected based on their shifting abilities during the SR-PAGE (Fig. 4b).

A smaller dissociation constant (K_D) represents a higher affinity of a molecule to its ligand. While the wildtype aptamer has a K_D smaller than 10 nM for TPP and of 4 μ M \pm 1.9 μ M for thiamine in our tested conditions (Fig. c), selected sequences using SR-PAGE as a selection method kept their binding affinity for thiamine and lost affinity for TPP through rounds of SELEX (Fig. 4f). K_D of candidates were determined by in-line probing (Regulski & Breaker, 2008). One of the best clones (clone 13) showed a slightly improved binding for thiamine molecules, with a K_D of 3 μ M \pm 1.3 μ M for thiamine and 7.8 μ M \pm 3.5 μ M for TPP (Fig. 4c, d, e). Other clones (1, 2, 9, 13, 23, 24, 25, 41, 42, 43, 44, 45, 46 and 47) also showed that the aptamer recognize thiamine with a better affinity than TPP molecules (Fig. 4f and Supplementary Table 3). Only clones where a K_D could be estimated from in-line probing are shown in the figure. Sequences of all selected clones as well as their K_D with thiamine and TPP when measurable can be found in Supplementary Table 3 (nomenclature is explained in Supplementary Fig. 6).

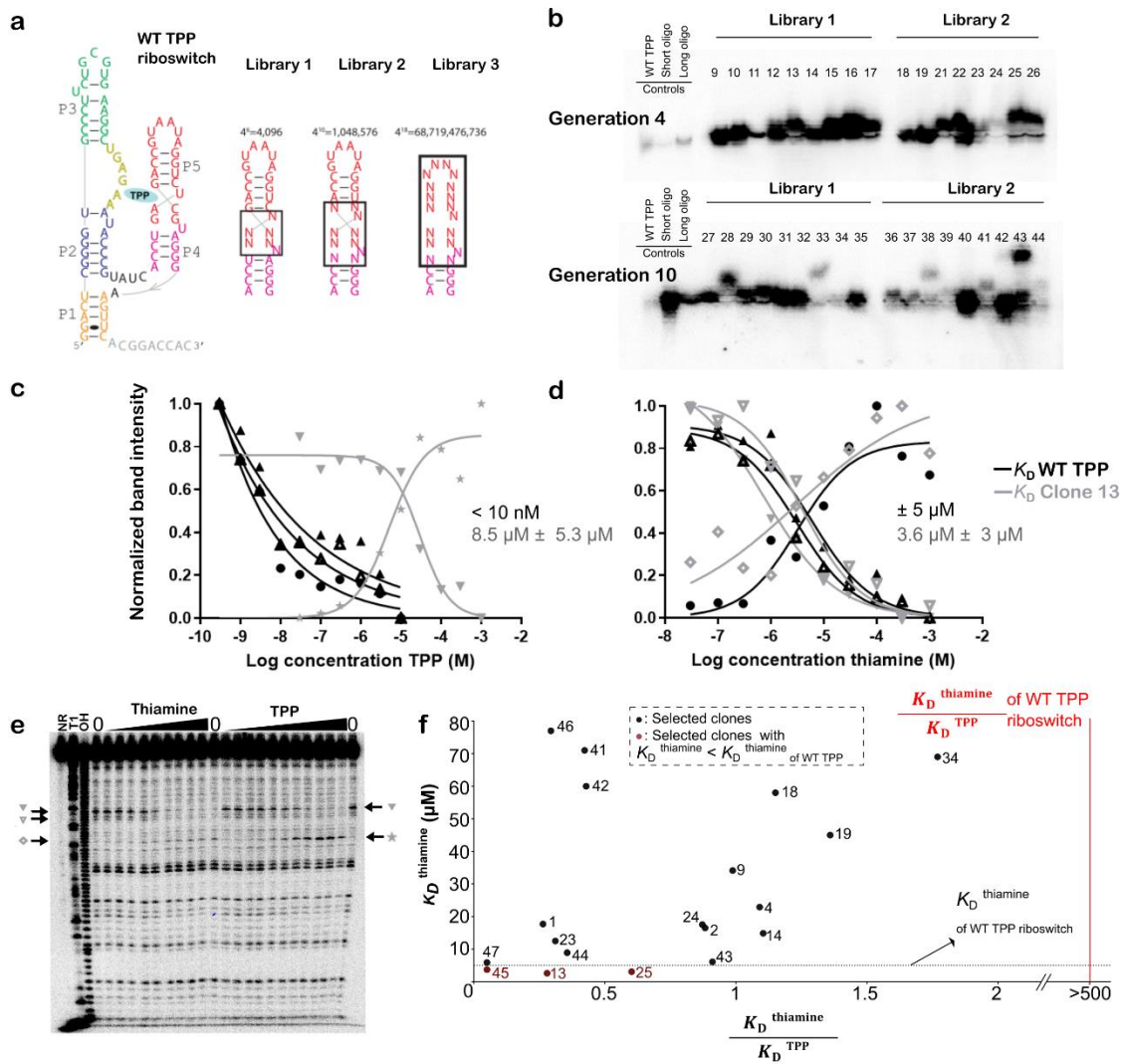


Fig. 4| Selection of a thiamine aptamer from degenerated libraries of the TPP riboswitch using the SR-PAGE as a selection tool within a SELEX. a, Degenerate libraries of the TPP riboswitch were created. Libraries 1, 2 and 3 had 6, 10 and 18 degenerated nucleotides respectively within stem P5. **b,** Shifting abilities as a result of SR-PAGE from libraries 1 and 2 with clones selected during generations 4 and 10 of the SELEX. TPP was present during the first migration within the running buffer (10 μM), and thiamine (20 μM) was sprayed on the gel before the second migration. RNA molecules can be observed since they are radioactively labeled. As controls, a short and long oligos as well as the WT TPP riboswitch were migrated on the same gel. **c,** K_D curve for WT TPP riboswitch (in black) and selected sequence 13 from library 1 after four generations of SELEX (in gray) with TPP. **d,** K_D curve for WT TPP riboswitch (in black) and selected sequence 13 from library 1 after four generations of SELEX (in gray) with thiamine. **e,** Inline probing gel of clone 13. The arrows represent the bands used for K_D calculation (graphed in c). Tested concentrations for thiamine and TPP ranged from 30 nM to 1 mM. **f,** Affinity of all selected sequences with thiamine and TPP. All dots represent individual selected sequences, where those emphasized in red are clones with a better affinity for thiamine than the WT TPP riboswitch. The dotted horizontal line depicts the K_D of the WT TPP riboswitch for thiamine. The vertical red line represents the ratio of K_D for thiamine over K_D for TPP of the WT TPP riboswitch.

In library 3, the P5 stem of the WT TPP riboswitch was initially replaced by random nucleotides, therefore only a small proportion of sequences are expected to form a similar stem, and none are expected to have the WT internal loop, since we do not select for TPP binding (Fig. 4a). After four and seven rounds of SELEX, selected libraries were analyzed by Illumina sequencing. Two of the clones (45 and 47) that we tested from this initial library show no binding to the TPP molecule up to a concentration of 1 mM (Fig. 4f). In spite of the fact that we completely degenerated stem P4-P5 (except for two bp), we apparently still selected for a stem (Supplementary Fig. 7a). Overall, secondary structure prediction for enriched sequences (i.e., more than three reads/sequence) showed that we selected for the formation of a stem in P5, as opposed to random sequences that do not typically form a stem (Supplementary Fig. 7; and Supplementary Table 4). Because the thiamine binding is not greatly improved (as suggested by K_D s similar to that of WT), this rather suggests that the stem was selected because it improved the shift between unbound states and thiamine-bound. Indeed, angles between stems are known to critically impact how RNA runs in a native PAGE (Lafontaine *et al.*, 2001).

One concern of using the SR-PAGE within a SELEX was the impact of the adapters necessary for the PCR amplification of the selected sequences in a SELEX round. Our results show that the adapters can indeed prevent riboswitch shifting using our positive controls (Supplementary Fig. 8). The addition of oligonucleotides complementary to the adapters helps to overcome this problem (Supplementary Fig. 8). All sequences for the aptamers and oligonucleotides can be found in the supplementary materials (Supplementary Table 1).

Discussion

In this paper, we presented the Shifted Reverse PAGE (SR-PAGE), a novel method to study riboswitches and aptamers going beyond the limits of bioinformatics and other existing experimental methods. This technique was validated with several known riboswitches, namely c-di-GMP-I, FMN, TPP and glycine, including when multiple ligands were used at the same time. We also demonstrated that the SR-PAGE could be imbricated within an SELEX as a method of selection, allowing us to identify degenerated TPP riboswitch sequences that lost their affinity for TPP and improved their affinity for thiamine, as confirmed by in-line probing.

A shift in migration for all known riboswitches was not always detectable, like for the fluoride and nickel-cobalt riboswitches for example (Fig. 2d). This can be explained by the tested constructions. Indeed, as demonstrated with the FMN_3 and FMN_5 constructions (Fig. 3), a few nucleotides make the difference between an RNA that shifts and an RNA that does not shift, likely due to a subtle equilibrium of structures between the conditions with and without ligands. Thus, a slight difference in sequence can result in a big change in terms of secondary structure. Even if all tested constructions technically contained the aptamer domain, it does not necessarily mean that the aptamer was forming. We hypothesized that a shift was not observed for all known riboswitches simply because the RNA sequence resulting in a bigger change of secondary structure upon ligand binding was not tested, or likely because the difference in free energy between the unbound and bound structure was too large to be overcome in other cases. The difference in predicted free energies was in average of 1.78 kcal/mol in constructions where a

shift in migration was observed (FMN_3, FMN_5, c-di-GMP_3-4-5; Fig. 2d), whereas it was on average 12.1 kcal/mol for those where no change in migration was seen (FMN_1-2-4, c-di-GMP_1-2, NiCo_1-2-3-4-5, fluoride_1-2-3-4-5; Fig. 2d). In these latter cases, the free energy difference between the unbound state and the ligand-bound state is too large to be overcome by the ligand. The equilibrium constant values in cases where no shift was observed were also small (typically $<10^{-3}$), suggesting that the secondary structures probably could not adopt the ligand-bound conformation and/or that the energy of ligand-binding was insufficient to overcome the gap between ΔG s. A shift was observed in cases where the equilibrium constant was closer to 1, suggesting the RNA structure could overcome the difference in free energy to be able to bind the ligand. This is the induced-fit model (Heppell *et al.*, 2011), where the addition of the ligand favors a change in secondary structure of a given RNA molecule. On another end, the addition of the ligand could favor RNA molecules that are already in the proper secondary structure, which is the conformational selection model (Haller *et al.*, 2011). In this case, the dynamic interchange of structures within the gel would be locked into the “aptamer state” upon addition of ligand to explain the change in migration. It is likely that combinations of both models occur in many cases, including also more subtle changes of the tertiary structure which could nevertheless lead to detectable shifts in the gel.

The amplitude of the shift (the height of the shift) appears to correlate with the level of structural difference between the unbound and the bound structure, where similar secondary structure result in a smaller change in migration. This was observed for example in FMN_5, where the shift was less pronounced than that of FMN_3, since it was already in a secondary structure closely related to that of the aptamer (Fig. 3), as well as c-di-GMP constructions, which were similar for the most part. Therefore, one of the limitations of SR-PAGE is that we might miss riboswitches that are able to bind a ligand, but the conformational change is not drastic enough to result in a detectable change in migration. However, this limitation can be used to our advantage, since the SR-PAGE can be used to identify constructions that will produce larger shifts and gain insight on the expression platform, further improving our knowledge of existent riboswitches and aptamers, as suggested with the analysis of the FMN riboswitch construction (Fig. 3).

We have also used SR-PAGE within a SELEX strategy to select riboswitches with a modified affinity for their ligands. The SR-PAGE selected RNA molecules based on two properties: first, for their ability to preferentially bind thiamine rather than TPP molecules, and for a change in the RNA conformation involving a shift by SR-PAGE upon that binding. More than forty sequences identified from that SELEX were tested separately by in-line probing. Out of all these sequences, 14 showed a better affinity for thiamine compared to that of the TPP molecule (K_D thiamine / K_D TPP <1 ; compared to >500 for WT). Moreover, three clones (13, 25 and 45) had a better affinity for thiamine than the WT TPP riboswitch (5 μ M). This experiment successfully demonstrated that SR-PAGE could be used as a selection step within a SELEX to select for affinity modified and/or enhanced aptamers.

Now that the SR-PAGE method has been validated, it could be employed to identify potential novel riboswitches starting with genomic libraries of intergenic sequences of bacteria, for example. The validation of the SR-PAGE technique presents a new approach to discover riboswitches, which has advantages over the normally used bioinformatics methods. Only

intergenic sequences are studied with bioinformatics analysis, whereas SR-PAGE allows researchers to look at the entire genomic sequences. Less complex RNA secondary structure can be overlooked when using bioinformatics, which SR-PAGE would not disregard. For example, mini-ykkC (aka guanidine-II) was deemed too simple to be a riboswitch until it was assayed with guanidine (Sherlock *et al.*, 2017). Moreover, novel regulatory structures will be directly linked to their ligand, contrary to bioinformatics, which often leads to orphan riboswitches, where the specificity of the ligand is not known (Greenlee *et al.*, 2018; Meyer *et al.*, 2011; Weinberg *et al.*, 2007; Weinberg *et al.*, 2017a; Weinberg *et al.*, 2010). The SR-PAGE selects for sequences that change conformation upon binding of a ligand. Therefore, potential riboswitches still need to be validated *in-vivo* using for example riboswitch-reporter fusion assay to confirm it has an impact on the downstream gene expression. Nevertheless, the use of SR-PAGE is not intended to overshadow bioinformatics techniques, but rather to complement them, since both have their advantages. This technique can be applied to several types of ligands such as coenzymes, amino acids and metal ions. Each new discovery will provide new insights into the biochemical capacity of RNA and open opportunities for technological advances.

Methods

PCR construction of riboswitches

Forward and reverse primers were designed to amplify the different constructions of the FMN and the fluoride riboswitches from the genomes of *Escherichia coli* (*E. coli*) and *Burkholderia thailandensis* (*B. thailandensis*) respectively, with the addition of the T7 promoter. For the glycine, TPP, nickel-cobalt and c-di-GMP class I riboswitches, PCR assembly was used to create the template using Primerize (Tian & Das, 2017). The different constructions of the c-di-GMP riboswitch were then created in the same manner as for the FMN and the fluoride riboswitches. All oligonucleotides used are listed in Supplementary Table 1. The PCR reactions were done with 2.5 U Hot Start Taq Polymerase (Qiagen/203203) with Hot Start buffer (1X), dNTPs (200 μ M), forward and reverse primers (1 μ M) completed with Milli-Q water. For assembly PCR, the end primers have a concentration of 1 μ M and the internal oligonucleotides have a concentration of 0.1 μ M.

Three libraries derived from *E. coli thiM* TPP riboswitch were designed to obtain an aptamer that binds thiamine much better than TPP. To do so, the pyrophosphate binding pocket within P5 was randomized to different degrees with degenerated primers (Fig. 4a) to create the libraries 1, 2 and 3. These oligonucleotides are also listed in the Supplementary Table 1.

***In vitro* RNA transcription**

For SR-PAGE, all PCR products of the known riboswitches as well as those from the clones derived from the SELEX with degenerated TPP riboswitches were radiolabeled with [α - 32 P] UTP (Perkin Elmer) during *in vitro* transcription using T7 RNA polymerase for 3 hours at 37 °C (20 μ L PCR products [~400 ng], 20 μ L 5X transcription buffer [400 mM HEPES-KOH, pH 7.5, 120 mM MgCl₂, 10 mM spermidine, 200 mM DTT], 0.001 U pyrophosphatase (Roche), 0.5 μ L RNase

inhibitor Ribolock (Thermo Fischer/EO0382) [40 U/ μ L], T7 polymerase [100 U], 5 mM ATP, 5 mM GTP, 5 mM CTP, 1 mM UTP and 5 μ Ci [α - 32 P] UTP). The volume was completed to 100 μ L with Milli-Q water. The transcribed RNA was precipitated at -80°C with 0.1 volume of sodium acetate 3 M (pH 5.2) and 2 volumes of 100% chilled ethanol for at least 2 hours. The precipitated RNA was resuspended in RNase-free Milli-Q water and purified on a 6% denaturing polyacrylamide gel electrophoresis (8 M urea PAGE). Samples were loaded on the gel with 2X denaturing loading buffer (0.05% bromophenol blue, 0.05% xylene cyanol, 10 mM EDTA [pH 8], 95% formamide). The expected size bands corresponding to our RNA were eluted in the elution buffer (0.3 M NaCl) and the eluate was precipitated as described before. Radiolabeled RNAs were resuspended in 250 μ L of Milli-Q sterilized water.

For in-line probing, RNA molecules were radiolabeled at their 5' end. *In vitro* transcription was performed as before, but with equal concentrations of all rNTPs, since radioactive [α - 32 P] UTP were not used. After an *in vitro* transcription, 5' ends were dephosphorylated using 5 U of Antarctic phosphatase (New England BioLabs/M0289S). A mix containing the RNA sample (~10 pmol), Antarctic phosphatase buffer 1X and 0.5 μ L RNase inhibitor Ribolock (Thermo Fischer/EO0382). The volume was adjusted to 20 μ L with sterilized Milli-Q water. The reaction was incubated at 37 °C for 20 minutes and deactivated at 65°C for 5 minutes. RNA molecules were radiolabeled at their 5' end in a reaction with ~10 pmol dephosphorylated RNA samples, 0.2 μ L T4 Polynucleotide Kinase (PNK) enzyme (New England BioLabs/M0201S), 1 μ L 10X T4 PNK buffer and 2 μ L [γ - 32 P] ATP (Perkin Elmer). The volume was adjusted to 10 μ L with water and incubated at 37°C for 1 hour. Labeled RNA samples were purified on a 6% denaturing PAGE, imaged and purified as described before.

SR-PAGE gel preparation

A slab gel system, typical of sequencing gels, was used. To mold the gel, we used glass plates of dimensions (38 cm x 45 cm) for the larger one and of (38 x 43 cm) for the smaller one. The smaller glass plate was treated with a solution of Rain-X®, a hydrophobic solution, whereas the larger glass plate was treated with a solution of potassium chloride (KOH) and methanol (5 g of KOH in 100 mL of methanol), a hydrophilic solution. These steps become important later when it is time to disassemble the gel to ensure that the gel sticks to the larger hydrophilic plate rather than on the smaller hydrophobic one. The plates are then assembled leaving a 2 cm gap at the bottom of the gel between the small and large plate (Supplementary Fig. 1a). Spacers are placed between the two plates, representing the thickness of the gel (approximately 0.8 mm). A 250 mL solution of native polyacrylamide gel (29:1) 10% was prepared with a final TBMg concentration of 1X (0.09 M Tris base, 0.09 M boric acid and 5 mM Mg(CH₃COO)₂ pH 8.0 at ambient temperature). The volume was completed with sterilized Milli-Q water. A volume of 50 mL of this solution was kept at 4°C for later use to maintain the exact same percentage of acrylamide for the solution used to cover the void left by the well removal. To polymerize the gel, 2 mL of 10% ammonium persulfate (APS, Bio-Rad) and 70 μ L of tetramethylethylenediamine (TEMED, Bioshop) was added to the remaining 200 mL preparation of the native polyacrylamide gel. Once the gel was poured, the gel was left to polymerize for at least 30 minutes. The assembly of the SR-PAGE was made as shown in Supplementary Fig.1b. To allow a good circulation of the buffer throughout the

gel, a peristaltic pump system was installed to carry the buffer from the bottom tank to top tanks. The buffer was also allowed to flow from the top tank to the bottom one via a tube (Supplementary Fig. 1b). Pre-migration was carried out overnight at 450 V with TBMg 1X as running buffer. From this point on, all the steps were carried in a 4°C room.

RNA samples were prepared in TBMg 1X, 3,3 μ L native blue 6X loading dye (40% sucrose, 0.05 % bromophenol blue, 0.05% xylene cyanol). Different controls were used to monitor proper migration. We used radioactive RNAs of different sizes (short and long) to show that regardless of the size, the RNA comes back to the starting point if it did not interact with the ligand. Another control is a Cy3 fluorescent oligonucleotide to visualize the migration within the gel with the naked eye (Supplementary Table 1). Finally, depending on the SR-PAGE assay being run, a riboswitch construct known to shift in native gels was often used as a positive control (validated by our preliminary results). These controls were prepared in the same way as the RNA samples. The wells were cleaned with a syringe to remove all potential bubbles and unpolymerized acrylamide in the wells before loading the samples and controls. RNA molecules were separated according to their size and structure for 24 hours in the 10% native polyacrylamide gel at 450 V.

SR-PAGE reverse migration

After the first migration of 24 hours, the glass plates were carefully separated. A 50 mL solution containing the ligands of interest was prepared. The final vaporized concentration of each known riboswitch's ligand is equivalent to hundred times their respective known K_D . The sprayed concentrations for each ligand were as follows: 6 mM FH_4KO_2 (Baker *et al.*, 2012) , 5 μ M FMN sodium salt, hydrate (Howe *et al.*, 2016) (Cayman chemical company), 3 μ M c-di-GMP sodium salt (Smith *et al.*, 2009) (Sigma-Aldrich), 10 mM glycine (Huang *et al.*, 2010) (Bioshop) , 3 μ M TPP chloride (Rentmeister *et al.*, 2007) (Sigma-Aldrich), 3 mM NiCl_2 (Furukawa *et al.*, 2015) (Fischer), 3 mM CoCl_2 (Furukawa *et al.*, 2015) (Fisher) in TBMg 1X. When we worked with metal ions, 10 mM L-glutathione ($\text{C}_{10}\text{H}_{17}\text{N}_3\text{O}_6\text{S}$) (Sigma-Aldrich) was also added to the sprayed solution as a reducing agent. The volume was adjusted to 50 mL with sterilized Milli-Q water. This solution was then sprayed directly onto the gel and left in contact for ten minutes (Supplementary Fig.1c). It is important to spray the solution across all the surface of the gel as uniformly as possible. The wells were cut out and the space left by the removal of these wells was carefully dried with Kimwipes® paper (Kimtech). TEMED solution (70 μ L) was added to the base of the wells to help seal the junction between the old and newly added acrylamide solution. The glass plates were reassembled, this time perfectly aligning the bottom of the two plates. The previously prepared native polyacrylamide solution was degassed and used to fill the space left by the wells, taking care to add the ligands sprayed into the solution at a final concentration 10 times lower than that present in the sprayed solution, since we do not rely on passive diffusion in the gel in this case. To allow the gel to polymerize, 700 μ L of 10% APS and 70 μ L of TEMED were added to the 50 mL native polyacrylamide gel preparation. Note that these steps were carried at 4°C, so higher concentrations of APS and TEMED are required for the gel to polymerize. The gel was left to polymerize for at least 30 minutes. The SR-PAGE was then reassembled as before, but the polarity of the gel was reversed, with the negative electrode now at the bottom and the positive at the top. The migration was restarted at 450 V (Supplementary Fig.1c). Using a marker, the

previous well line was delineated on the glass plates to use as a reference point to know when to stop the second migration.

Migration was stopped once the fluorescent oligonucleotide reached the wells. Due to a decreased resistance, the fluorescent oligonucleotide reached the baseline in a little less time than the first migration. This time could range anywhere between 19h to 23h, so it was important to monitor closely the SR-PAGE when we reached the end of the second migration. When it was not necessary to purify RNA from the gel, as for radioactively labeled riboswitches from Figs. 2d and 4b, the top of the gel was cut off and then dried using a gel dryer (model 583 Biorad, with a vacuum pump CVC3000 Vacuubrand). The gel was then exposed overnight on a phosphor screen and visualized with a Typhoon FLA9500 (GE Healthcare Life Sciences). For SELEX assays, gel strips were cut. Band 0 corresponds to the well baseline, band 1 corresponds to ~2 mm above the well up to 1 cm above the wells and band 2 corresponds to all the rest of the top of the gel; then the RNAs were eluted as described before.

SELEX of thiamine-binding TPP-derived riboswitches

When SR-PAGE was imbricated in an SELEX, like for the selection of a modified aptamer starting from degenerate libraries of TPP riboswitches, gel regions corresponding to “shifted-RNA” were cut out. RNAs were eluted in elution buffer overnight and ethanol-precipitated like described before. Reverse transcription (RT) was performed with M-MuLV reverse transcriptase (NEB) according to manufacturer instructions with a short oligonucleotide to prevent annealing of the oligonucleotides on RNAs that lost a few nucleotides at the 3' end during the migration (and would thus run faster in the reverse run, co-migrating with “shifted” RNA). Half of the RT was used as a PCR template for a 100 μ L reaction with conditions like library generation, but with 30 cycles for the first generation, and then progressively less for subsequent generations. RNAs were transcribed *in vitro* to start a new round of selection with an SR-PAGE like described before.

During selection, 10 μ M of TPP was added to the running buffer of the SR-PAGE. For the second migration, thiamine was sprayed on the gel, with a decreasing concentration as the SELEX generation increased to improve stringency. Therefore, 2 mM of thiamine was sprayed on the gel for generation 1 through 4. Then, 20 μ M of thiamine was added for generation 5. Finally, 200 nM of thiamine was sprayed for the last cycles (generations 7 to 10).

In-line probing and dissociation constant determination.

Sequences from the initial library or from generation 4, 5, 7 and 10 were either cloned with the pGEM®-T kit (Promega) to assay individual sequences and/or sequenced by Illumina (Centre d'expertises et de services Genome Québec). After analysis of Illumina sequencing results, sequences that were enriched during the cycles were selected. Individual sequences were amplified by PCR, transcribed to RNA by *in vitro* transcription, purified on 6% 8 M urea PAGE, eluted and precipitated as described before.

In-line probing reactions were performed with different concentrations of ligands ranging from 30 nM to 1 mM in a final volume of 20 μ L containing 20 mM $MgCl_2$, 50 mM Tris pH 8.3 (at 25°C) and 100 mM KCl at room temperature for approximately 40 hours. For the WT TPP riboswitch, the concentrations of ligands were ranging from 30 nM to 1 mM for thiamine and from 1 nM to 30 μ M for TPP. The volume of the reaction was adjusted with Milli-Q water. The resulting reaction was quenched with 2X loading buffer. Three ladders (NR: No Reaction, i.e., RNA in water; T1: RNase T1 which cleaves at guanine bases; and OH: alkaline reaction which cleaves all positions) were made to map the RNA sequence. For the T1 ladder, 10 μ L of the radioactively labeled RNA was incubated in 50 mM of sodium citrate, 3 μ L of formamide 100% and 4 μ L of Milli-Q water. The solution was incubated at 56°C for 2 minutes before 1.5 μ L of T1 RNase 1 U/ μ L was added to the T1 solution. It was left to react at 56°C for 5 minutes. For the alkaline digestion (OH), 10 μ L of the radioactively labeled RNA was incubated with 8 μ L of Milli-Q water and 50 mM sodium carbonate at 90°C for 90 seconds. The non-reacted ladder contained only 10 μ L of the radioactively labeled RNA with Milli-Q water to a final volume of 20 μ L. All reactions for the creation of the ladders were stopped with 20 μ L of 2X loading buffer. Samples and ladders were loaded on an 8% urea-PAGE for approximately 3 hours at 50 W (Regulski & Breaker, 2008). The gel was then dried and exposed on a phosphor screen overnight before being visualized on a Typhoon FLA9500 (GE Healthcare Life Sciences).

For dissociation constant determination, the intensity of the lanes and modulating bands were quantified using the ImageQuant TL software (GE Healthcare Life Sciences). These bands were normalized to give values between 0 and 1 according to minimum and maximum intensity of the bands. The K_D representation and determination were done with logarithmic curves on GraphPad 7 software (Sherlock *et al.*, 2017).

Free energy calculation and RNA secondary structure

Free energy of all RNA sequences were assessed using the website service of Mfold accessible at www.unafold.org (Zuker, 2003). RNAs were forced in the aptamer state using constraint based on consensus of secondary structure in Rfam (Kalvari *et al.*, 2021). All constraints applied to Mfold are available in supplementary material (Supplementary Table 2). RNA secondary structures were designed using the Stockholm file generated by Mfold (Zuker, 2003) as input for R2R (Weinberg & Breaker, 2011).

Equilibrium constant calculation

The equilibrium constants between the bound and unbound conformation of the tested RNA molecules were calculated using the following formula:

$$K_{eq} = e^{\frac{-\Delta\Delta G}{RT}}$$

The $\Delta\Delta G$ (in joules) is the difference in free energy (ΔG) of the bound and unbound conformation. The gas constant (R) is 8.314 J/mol·K. The temperature is 277 K, since all SR-PAGE experiments have been carried out at 4 ° C.

Acknowledgements

Authors wish to thank Gaël Montagne for technical help. Early work on SR-PAGE was started within the Breaker lab, JP wishes to thank Ronald Breaker for this, as well as members of the Breaker lab, including Rüdiger Welz who had performed gel shifts with the glycine riboswitch, providing the first positive control for SR-PAGE.

Author Contributions

A.D. and E.B. contributed equally to this work. J.P. conceived the SR-PAGE method. A.D., E.B., J.O., R.R., B.S. and J.P. optimized the method. J.O. realized the SELEX of thiamine-binding TPP-derived riboswitches. A.D. and E.B. performed all other experiments presented in the article. Figures were created by J.P., J.O., A.D. and E.B. Manuscript was written by J.P., A.D. and E.B. Project was supervised by J.P. Manuscript was revised by all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper.

Correspondence and request for materials should be addressed to J.P.