

Centre Armand-Frappier Santé Biotechnologie

IDENTIFICATION DE POTENTIELS RIBO-INTERRUPTEURS DE LA GUANIDINE

Par
Quetia Joseph

Mémoire présenté pour l'obtention du grade de
Maître ès Sciences (M.Sc.)
en microbiologie appliquée

Jury d'évaluation

Président du jury et
examineur interne

Dr Charles Dozois
INRS-Centre Armand-Frappier Santé
Biotechnologie

Examineur externe

Dr Benoît Cousineau
Microbiologie et immunologie
Université McGill

Directeur de recherche

Dr Jonathan Perreault
INRS-Centre Armand-Frappier Santé
Biotechnologie

REMERCIEMENTS

Je tiens à remercier toutes les personnes qui ont contribué de près ou de loin à la réalisation de ce projet de recherche ainsi qu'à toutes celles qui m'ont soutenue au cours de ma maîtrise. Je remercie de tout coeur le Dr Jonathan Perreault pour cette opportunité et pour sa confiance. Je remercie également les membres du laboratoire avec qui j'ai vécu des années enrichissantes. Un merci particulier à Radia Benmiloud pour sa contribution dans la recherche bio-informatique.

Je tiens également à remercier sincèrement mes parents. Je n'aurais pas été capable de me rendre jusqu'ici sans eux à mes côtés. Je remercie également tous mes frères et sœurs qui m'ont toujours encouragé à persévérer malgré les difficultés.

Je remercie également les docteurs Dozois et Cousineau pour le temps consacré à l'examen de ce mémoire.

Finalement, je remercie aussi la Fondation Armand-Frappier qui m'a octroyé une bourse.

RÉSUMÉ

Les ARN non codants (ARNnc) de type ribo-interrupteurs (*riboswitches*) sont principalement retrouvés dans la région 5'UTR (*untranslated region*) et permettent d'influencer l'expression de nombreux gènes en intervenant soit au niveau de la transcription ou encore au niveau de la traduction. C'est le cas du ribo-interrupteur *mini-ykkC* également appelé guanidine-II qui lie la guanidine, un composé toxique pour la bactérie. Associé à certaines pompes à efflux comme SugE, guanidine-II permettrait donc de réguler l'expulsion de la guanidine hors de la cellule. Or, ce ribo-interrupteur possède un motif simple composé d'une pochette de liaison de seulement six nucléotides. En nous basant sur la simplicité de son motif, nous avons émis l'hypothèse que cet ARN pourrait être largement répandu au sein de divers génomes. Au cours de cette recherche, nous avons pu identifier, par des méthodes bio-informatiques, environ 280 000 candidats correspondants au motif guanidine-II disséminés au travers de différents génomes tels que chez les bactéries, chez les archaea et même au sein d'un génome où aucun ribo-interrupteur n'a été trouvé jusqu'alors, chez l'être humain. De plus, les résultats de cartographie *in vitro* par *in-line* (*in-line probing*) ont permis d'identifier un possible ribo-interrupteur dans un ARN bactérien déjà annoté, soit la ribonucléase P.

Mots-clés : *ykkC*; *mini-ykkC*; *ykkC-III*; *sugE*; *small multidrug resistance*; ribo-interrupteur guanidine; cartographie *in vitro* par *in-line*; RNase P; *Salinibacter ruber*

ABSTRACT

Riboswitches are non-coding RNAs (ncRNA) primarily found in 5' UTR and can influence gene expressions via intervention during transcription or translation processes. This is the case of riboswitches mini-*ykkC* also called guanidine-II which binds guanidine, a toxic element for bacteria. Associated with some efflux pumps such as SugE, guanidine-II would be able to regulate guanidine expulsion out of the bacterial cell. However, this riboswitch has a simple motif containing a binding pocket of only six nucleotides. Based on the simplicity of its motif, we hypothesized that this RNA could be widely spread throughout different genomes. During this research, we were able to identify, through bioinformatics tools, approximately 280 000 hits corresponding to the guanidine-II motif scattered through various genomes such as bacteria, archaea and even within a genome where no riboswitches have been found to date, namely in humans. Furthermore, in-line probing results have identified a possible riboswitch in a bacterial RNA already annotated, that is ribonuclease P.

Keywords : *ykkC*; mini-*ykkC*; *ykkC*-III; *sugE*; small multidrug resistance; guanidine riboswitches; in-line probing; RNase P; *Salinibacter ruber*

TABLE DES MATIÈRES

REMERCIEMENTS	III
RÉSUMÉ	V
ABSTRACT	VII
TABLE DES MATIÈRES	IX
LISTE DES FIGURES.....	XI
LISTE DES TABLEAUX	XIII
LISTE DES ÉQUATIONS	XIV
LISTE DES ABRÉVIATIONS.....	XV
1 INTRODUCTION.....	1
1.1 LES ACIDES RIBONUCLEIQUES (ARN).....	1
1.1.1 <i>Les ARN non codants</i>	2
1.2 LES RIBO-INTERRUPTEURS	2
1.2.1 <i>Découverte</i>	3
1.2.2 <i>Structure et fonctionnement</i>	4
1.3 FAMILLES RIBO-INTERRUPTEURS GUANIDINE.....	7
1.3.1 <i>Guanidine-I</i>	7
Motif et gènes associés	7
Structures secondaire et tertiaire	10
1.3.2 <i>Guanidine-II</i>	12
Motif et gènes associés	12
Structures secondaire et tertiaire	14
1.3.3 <i>Guanidine-III</i>	16
Motif et gènes associés	16
Structures secondaire et tertiaire	17
1.3.4 <i>Guanidine IV</i>	20
Motif et gènes associés	20
Structures secondaire et tertiaire	21
Récapitulatif	22
2 PROBLÉMATIQUE.....	23
3 HYPOTHÈSE ET OBJECTIFS.....	25
4 MATÉRIEL ET MÉTHODES	27

4.1	RECHERCHE DES MOTIFS DE GUANIDINE-II.....	27
4.1.1	<i>Banque de données RNAs_{tem}</i>	27
4.1.2	<i>RNArobo</i>	28
4.1.3	<i>Sélection des candidats</i>	29
4.2	TRANSCRIPTION.....	30
4.2.1	<i>Préparation des matrices</i>	30
4.2.2	<i>Transcription in vitro</i>	33
4.3	MARQUAGE EN 5' A L'AT ³² P.....	34
4.4	CARTOGRAPHIE <i>IN VITRO</i> PAR <i>IN-LINE</i>	35
5	RÉSULTATS.....	39
5.1	RECHERCHE BIO-INFORMATIQUE.....	39
5.1.1	<i>RNAs_{tem}</i>	39
5.1.2	<i>RNArobo</i>	41
5.1.3	<i>Gènes associés aux candidats RNArobo</i>	43
5.2	CARTOGRAPHIQUE <i>IN VITRO</i> PAR <i>IN-LINE</i>	53
5.2.1	<i>Ribo-interrupteur glutamine</i>	53
5.2.2	<i>Ribonucléase P</i>	56
5.3	ÉVALUATION <i>IN VIVO</i> DE LA FONCTION INTERRUPTRICE DU CANDIDAT RNASE P.....	59
6	DISCUSSION.....	61
6.1	GENES ASSOCIES AUX CANDIDATS BACTERIENS.....	61
6.2	ÉVALUATION DU CANDIDAT <i>GLNA</i>	61
6.3	ÉVALUATION DU CANDIDAT RNASE P.....	62
6.4	IMPORTANCE DU MOTIF ET POTENTIEL DES MOTIFS A DEUX TIGES.....	65
7	CONCLUSION ET PERSPECTIVES.....	67
8	BIBLIOGRAPHIE.....	69
9	ANNEXE.....	75

LISTE DES FIGURES

FIGURE 1. DIVERSITE DE LIGANDS DES RIBO-INTERRUPTEURS.	3
FIGURE 2. STRUCTURE ET MECANISME D'UN RIBO-INTERRUPTEUR TYPIQUE.....	5
FIGURE 3. MECANISMES REGULATEURS COURAMMENT UTILISES PAR LES RIBO-INTERRUPTEURS.	6
FIGURE 4. LE MOTIF YKKC ET LES PROTEINES ASSOCIEES.	8
FIGURE 5. STRUCTURES CHIMIQUES DE METABOLITES.....	9
FIGURE 6. CONSENSUS MIS A JOUR DU RIBO-INTERRUPTEUR GUANIDINE-I.	11
FIGURE 7. STRUCTURE DU RIBO-INTERRUPTEUR GUANIDINE-II ET LES PROTEINES ASSOCIEES.	13
FIGURE 8. STRUCTURES SECONDAIRE ET TERTIAIRE DU RIBO-INTERRUPTEUR GUANIDINE-II.....	15
FIGURE 9. STRUCTURE DU RIBO-INTERRUPTEUR GUANIDINE-III ET LES PROTÉINES ASSOCIÉES.	17
FIGURE 10. STRUCTURES SECONDAIRE ET TERTIAIRE DU RIBO-INTERRUPTEUR GUANIDINE-III.....	19
FIGURE 11. STRUCTURE SECONDAIRE DU RIBO-INTERRUPTEUR GUANIDINE IV.	21
FIGURE 12. CAPTURE D'ECRAN DE LA REQUETE RNASTEM.....	27
FIGURE 13. DESCRIPTEURS UTILISES POUR LA RECHERCHE RNAROBO.....	28
FIGURE 14. CRITERES DE SELECTION DES CANDIDATS RNASTEM ET RNAROBO.	29
FIGURE 15. SCHEMA DE LA PCR D'ASSEMBLAGE DES CANDIDATS <i>GLNA</i> ET <i>PFL</i>	31
FIGURE 16. SCHEMA DU CLIVAGE SPONTANE DES LIENS PHOSPHODIESTERS DE L'ARN.	36
FIGURE 17. SCHEMA DU GEL D' <i>IN-LINE</i>	37
FIGURE 18. COMPARAISON DES ECHANTILLONS STATISTIQUE.	52
FIGURE 19. STRUCTURES DU RIBO-INTERRUPTEUR <i>GLNA</i> ET DE SON LIGAND.	54
FIGURE 20. LIAISON <i>IN VITRO</i> DE LA GUANIDINE PAR LE RIBO-INTERRUPTEUR <i>GLNA</i>	55
FIGURE 21. LIAISON <i>IN VITRO</i> DE LA GUANIDINE PAR LE CANDIDAT RNASE P.	57
FIGURE 22. STRUCTURES SECONDAIRE ET PRIMAIRE DU CANDIDAT RNASE P.....	58

LISTE DES TABLEAUX

TABLE 1. CANDIDATS RNASTEM SELECTIONNES POUR LA VALIDATION EXPERIMENTALE	40
TABLE 2. NOMBRE DE CANDIDATS TOTAUX TROUVES DANS DIFFERENTS GENOMES ET IGR	41
TABLE 3. CANDIDATS RNAROBO TESTES EXPERIMENTALEMENT	42
TABLE 4. ANALYSE DES PROTEINES ASSOCIEES AUX CANDIDATS RNAROBO	45
TABLE 5. CONTROLE GENES ALEATOIRES	47
TABLE 6. CONTROLE GENES ASSOCIES AU RIBO-INTERRUPTEUR GLYCINE.....	48
TABLE 7. CONTROLE GENES ASSOCIES AU RIBO-INTERRUPTEUR MOCO	50
TABLE 8. CONTROLE GENES ASSOCIES A L'ARN <i>RYBB</i>	51

LISTE DES ÉQUATIONS

Score arbitraire pour la sélection des candidats.....	29
---	----

LISTE DES ABRÉVIATIONS

3D : trois dimensions
A : adenine
ADN : acide désoxyribonucléique
APS : persulfate d'ammonium
ARN : acide ribonucléique
ARNnc : ARN non codants
ARNP : ARN polymérase
ARNr : ARN ribosomal
ARNt : ARN de transfert
AT. : antiterminateur
C : cytosine
EDTA : Éthylènediaminetétraacétique
G : guanine
GGAM-1 : Guanidine-Gene-Associated Motifs 1
IGR : région intergénique
lncRNA : long ARN non codant
O2' : oxygène en 2'
O5' : oxygène en 5'
P : phosphore
PAGE : (électrophorèse sur) gel de polyacrylamide
pré-ARNt : ARN précurseur
ribo-interrupteur *glnA* : ribo-interrupteur glutamine
Rnase P : ribonucléase P
RT : transcriptase inverse
SD : Shine-Dalgarno
SMR : small multidrug resistance
T : thymine
T7 ARNp : T7 ARN polymérase
TAE : Tris Acétate EDTA
TBE : Tris Borate EDTA
TEMED : tétraméthyléthylènediamine
Tris : Tris(hydroxyméthyl)aminométhane
U : uracile
UV : ultraviolet
ZMP : 5-aminoimidazole-4-carboxamide ribonucleotide
ZTP : 5-amino-4-imidazole carboxamide riboside 5'-triphosphate

1 INTRODUCTION

Le dogme central de Crick originellement discuté à la fin des années 1950 a permis d'établir les balises pour la compréhension du transfert de l'information génétique entre l'acide désoxyribonucléique (ADN), l'acide ribonucléique (ARN) et les protéines (Cobb, 2017). Dans un document publié en 1970, Crick dégage quatre grandes directions dans le transfert d'information génétique et qui sont soutenues aujourd'hui: 1. ADN vers ADN, 2. ADN vers ARN, 3. ARN vers protéines et 4. ARN vers ARN (Crick, 1970). Alors qu'il est question de réplication dans la première et la quatrième directions, les deuxième et troisième placent clairement l'ARN comme étape intermédiaire dans la synthèse des protéines. Puis, la découverte d'une transcriptase inverse (RT) vient soutenir une 5e direction dans le transfert d'information génétique, soit ARN vers ADN (Nature, 1970).

1.1 Les acides ribonucléiques (ARN)

L'ARN est une macromolécule présente chez tous les organismes vivants. Peu importe l'organisme où il est trouvé, son squelette reste le même : une succession de sucres ribose auxquels sont attachés différentes bases azotées, soient adénine (A), guanine (G), cytosine (C) et uracile (U), ainsi que des groupements phosphates formant les liens phosphodiesters qui relient les carbones 5' et 3' des sucres entre eux. L'ARN est dit simple brin, mais sa longueur et sa structure peuvent varier. Ainsi, il est possible pour les bases azotées d'être liées entre elles par des liens hydrogènes qui se formeront, généralement, entre l'adénine et l'uracile ainsi qu'entre la cytosine et la guanine ((Wang & Farhana, 2020). La différence entre cette dernière molécule et son homologue, l'ADN, se situe également dans le squelette. En effet, l'ADN contient le sucre désoxyribose au lieu du ribose et utilise plutôt la base thymine (T) au lieu de l'uracile. Finalement, l'ADN, quant à lui, adoptera plutôt une conformation en double hélice dite double brin. Les bases azotées se lieront donc entre elles selon les mêmes règles, mais avec la thymine prenant la place de l'uracile (Ghannam *et al.*, 2020). Alors que l'ADN est l'unité de base de l'information génétique chez tous les organismes vivants où il est trouvé, les fonctions de l'ARN sont un peu plus diversifiées. En fait, ce dernier sera classé dans différentes catégories selon sa fonction, mais fera essentiellement partie de l'une des deux classes suivantes à savoir les ARN codants ou les ARN non codants (ARNnc) (Wang & Farhana, 2020).

1.1.1 Les ARN non codants

Si chez les bactéries, la majorité des transcrits sont codants et permettent donc la synthèse de protéines (Cases *et al.*, 2003), inversement, chez les organismes dits supérieurs, cette proportion est à moins de 3% (Li & Liu, 2019). Les ARN non codants sont donc les transcrits qui n'encodent pas pour une protéine. Ceux-ci restent toutefois fonctionnels, et sont impliqués dans plusieurs processus biologiques tels que la transcription ou la traduction. C'est le cas notamment des petits ARN nucléaires impliqués dans l'épissage des ARNm chez les eucaryotes ou encore des ARN ribosomiaux (ARNr) et de transfert (ARNt) impliqués dans la synthèse des protéines (Noller, 1991 ; Yarus, 1972). Ces derniers exemples sont des ARN domestiques, soient des ARN dont le gène est toujours transcrit (Romano *et al.*, 2017). D'autres ARNnc seront plutôt impliqués dans des processus de régulations.

La fonction des ARNnc régulateurs est souvent liée à leur séquence et à leur repliement (Hubé & Francastel, 2018). Ces derniers sont classifiés selon leur taille, soit les longs ARNnc (lncRNA) composés de 200 à plusieurs milliers de nucléotides ainsi que les petits ARNnc (moins de 200 nucléotides) (Ma *et al.*, 2013 ; Mattick & Makunin, 2006). Les ARNnc régulateurs sont également classifiés selon leur mode d'action, à savoir en *trans* ou en *cis*. Le mode d'action en *trans* implique que les ARNnc régulateurs peuvent réguler l'expression génique en intervenant auprès d'une autre molécule, le plus souvent en formant un complexe plus ou moins parfait avec celle-ci. On y retrouve notamment, les petits ARN interférents et les micro ARN (eucaryotes) ainsi que les petits ARN (procaryotes). Les lncRNA sont, quant à eux, retrouvés à la fois chez les eucaryotes et les procaryotes (Carthew & Sontheimer, 2009 ; Jørgensen *et al.*, 2020). L'autre mode d'action est dit en *cis*. Dans ce cas-ci, il s'agit souvent d'ARNnc se trouvant dans les régions intergéniques, les régions non traduites (UTR) en 3' et 5' et les introns dans le cas des eucaryotes. On y retrouve notamment certains lncRNA capables d'agir en *cis*, les amplificateurs ou *enhancers* (eucaryotes) ainsi que les ribo-interrupteurs également appelés *riboswitches* en anglais (Gil & Ulitsky, 2020 ; Guil & Esteller, 2012 ; Pavlova *et al.*, 2019).

1.2 Les ribo-interrupteurs

Les ribo-interrupteurs sont des ARNnc trouvés majoritairement dans la région 5'UTR des ARNm. Ils répondent à une variété de ligands (Figure 1) et peuvent contrôler l'expression du gène se trouvant en aval par divers mécanismes (Serganov & Nudler, 2013). Le terme « interrupteur » (ou *switch* de son nom en anglais) fait donc référence au fait que, tout comme un interrupteur, ces ARN sont capables, chez la bactérie, d'activer ou d'inhiber un gène selon l'abondance

intracellulaire du ligand auquel le ribo-interrupteur répond (Breaker, 2011; Nahvi *et al.*, 2002). En effet, les ribo-interrupteurs sont largement répandus chez les procaryotes, bien qu'un exemple ait également été trouvé chez certains eucaryotes (Bocobza *et al.*, 2007; Galagan *et al.*, 2005).

1.2.1 Découverte

Avant la découverte des ribo-interrupteurs, l'un des mécanismes de régulation de l'expression des gènes le plus connus chez la bactérie était la rétro-inhibition impliquant l'intervention de protéines appelées répresseurs. Lorsque l'inhibition des vitamines B₁, B₂ et B₁₂ par la thiamine, la riboflavine et la cobalamine respectivement ont été montrés, il n'a pas été étonnant de voir les efforts converger vers la découverte des répresseurs possiblement impliqués dans ce mécanisme (Miranda-Ríos *et al.*, 2001 ; Nou & Kadner, 1998). Rien n'a été trouvé en ce sens, mais il a plutôt été trouvé que les mécanismes de régulation en cause sont dus à des séquences d'ARNm régulateurs et que ceux-ci sont capables de répondre à un changement de concentrations de certains métabolites, notamment des dérivés de vitamines (Gelfand *et al.*, 1999 ; McDaniel *et al.*, 2003 ; Miranda-Ríos *et al.*, 2001 ; Nou & Kadner, 2000 ; Stormo & Ji, 2001).

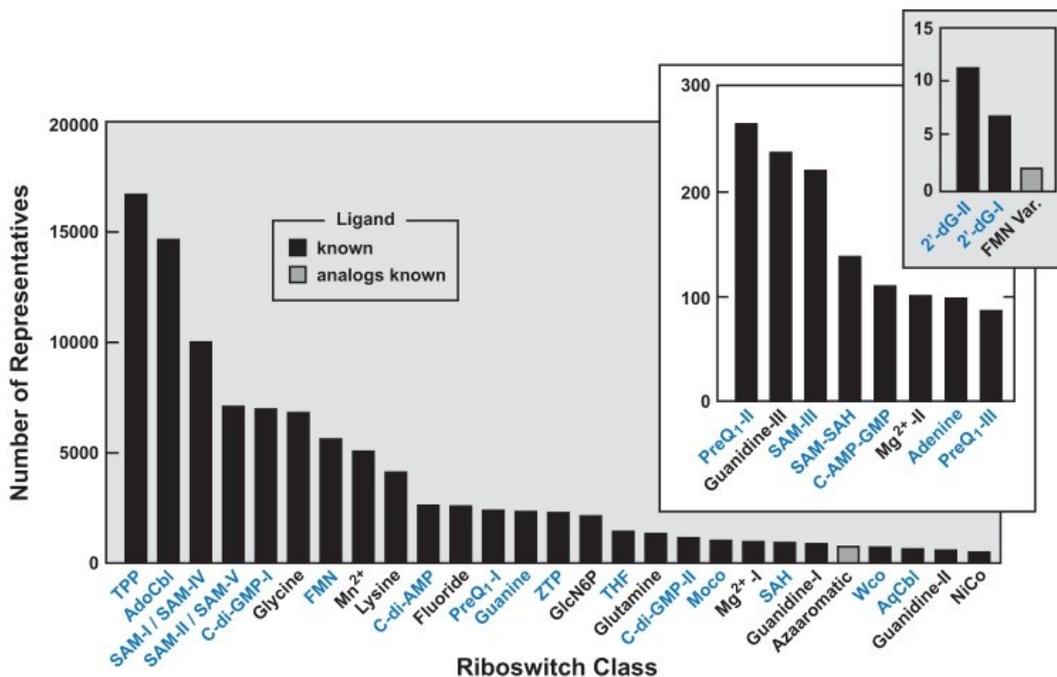


Figure 1. Diversité de ligands des ribo-interrupteurs.

Les classes de ribo-interrupteurs sont nommées conformément à leur ligand. Dans le cas de plusieurs classes possédant le même ligand, ces derniers sont identifiés par des chiffres romains. Les classes sont présentées par rang selon leur abondance dans la base de données génomique. Les barres grises et noires représentent les ribo-interrupteurs dont les ligands naturels sont non connus (*analogs known*) et connus (*known*) respectivement. Les textes en bleu correspondent aux ribo-interrupteurs pour lesquels les ligands sont des dérivés de nucléotides d'ARN ou de leurs précurseurs. Tiré de (McCown *et al.*, 2017).

1.2.2 Structure et fonctionnement

Les ribo-interrupteurs comportent deux domaines distincts, à savoir l'aptamère et la plateforme d'expression (Figure 2 A). L'aptamère est le domaine hautement conservé qui permet au ribo-interrupteur de lier son ligand avec affinité et spécificité. En effet, ce domaine, adopte une structure en trois dimensions (3D) formant ainsi la pochette de liaison du ligand qui sera capable de discriminer le ligand spécifique de molécules possédant des structures similaires à ce dernier (Winkler & Breaker, 2003). La plateforme d'expression, quant à elle, est le domaine permettant de réguler l'expression des gènes (Lotz & Suess, 2018). En fait, ces deux domaines se chevauchent pour former la séquence interruptrice. À la liaison du ligand au domaine aptamère, la séquence interruptrice dirige le repliement de l'ARN au niveau de la plateforme d'expression afin qu'elle se conforme dans l'un de ces deux états : *on* (en marche) ou *off* (éteint) (Figure 2 B). Ces deux états, chacun traduit par une structure distincte de la plateforme d'expression, sont mutuellement exclusifs et ne peuvent donc coexister (Garst *et al.*, 2011). La plateforme d'expression, contrairement à l'aptamère, est moins conservée. En effet, il est montré que des aptamères homologues présents dans différents organismes sont souvent accompagnés de plateformes d'expression différentes. Ceci suggère que bien que la séquence de l'aptamère reste conservée, celle de la plateforme d'expression, quant à elle, peut varier au cours de l'évolution afin de s'ajuster aux besoins de régulation (Jones & Ferré-D'Amaré, 2017).

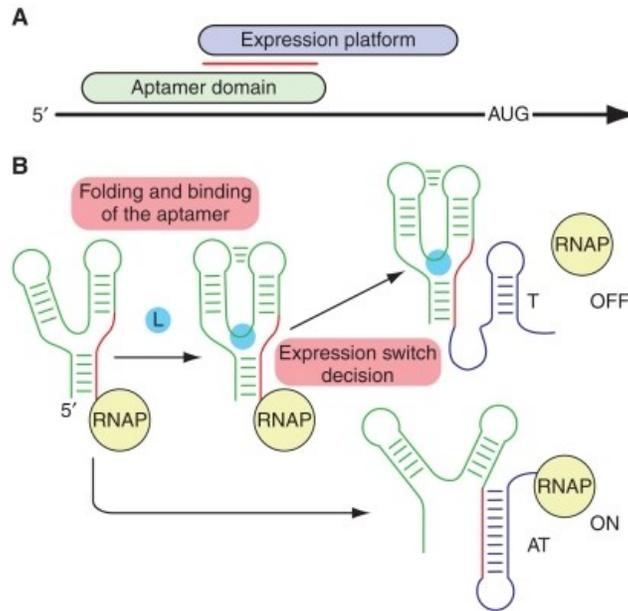


Figure 2. Structure et mécanisme d'un ribo-interrupteur typique.

A) Présentation des éléments structuraux d'un ribo-interrupteur et de leur réarrangement tel que présent dans la séquence primaire de l'ARNm : domaine aptamère (vert), plateforme d'expression (violet), séquence interruptrice (ligne rouge). La séquence interruptrice est formée par les deux domaines se chevauchant. B) Durant la transcription du ribo-interrupteur par l'ARN polymérase (RNAP), plusieurs facteurs interviennent afin de permettre la bonne régulation de l'expression des gènes. L'un des facteurs arrivant tôt durant la transcription est le repliement de l'aptamère et une potentielle liaison du ligand (L, cyan). Dépendamment de la liaison du ligand, le ribo-interrupteur adoptera une conformation au niveau de la plateforme d'expression (T : terminateur rho-indépendant, AT : antiterminateur) qui déterminera le type de réponse régulatrice (ON ou OFF). Tiré de (Garst et al., 2011). Droits d'auteur de *Cold Spring Harbor Laboratory Press*.

Pour réguler l'expression de gènes, bien que certains ribo-interrupteurs puissent agir en *trans* (De Lay & Garsin, 2016; Loh *et al.*, 2009), la grande majorité utilise une diversité de mécanismes dits en *cis* en régulant l'expression des gènes se trouvant en aval. Ces mécanismes comprennent notamment l'atténuation de la transcription, l'initiation de la traduction, la dégradation de l'ARNm (*mRNA decay*) et l'épissage de l'ARNm (Bédard *et al.*, 2020).

Pour agir au niveau de la transcription, les ribo-interrupteurs font le plus souvent intervenir la formation d'un terminateur Rho-indépendant ou d'un antiterminateur (AT) situé dans la plateforme d'expression. Ainsi, à la liaison du ligand au domaine aptamère, la plateforme d'expression pourrait permettre la formation du terminateur bloquant ainsi l'accès à l'ARN polymérase (ARNP) ce qui mettra fin à la transcription (OFF). En l'absence du ligand, c'est plutôt l'AT qui est formé et la transcription, par défaut, continue (ON) (Figure 2 B et Figure 3) (Kim & Breaker, 2008). À l'inverse, la liaison du ligand pourrait activer la transcription en permettant la formation de l'AT (Figure 3) (Serganov & Nudler, 2013). Pour ce qui est de la traduction, les ribo-interrupteurs

interviennent le plus souvent en séquestrant ou en inhibant la séquence de Shine-Dalgarno (SD) à la liaison du ligand. Ceci correspondra à l'inhibition (OFF) ou à l'activation de la traduction (ON) respectivement (Figure 3) (Kim & Breaker, 2008 ; Serganov & Nudler, 2013).

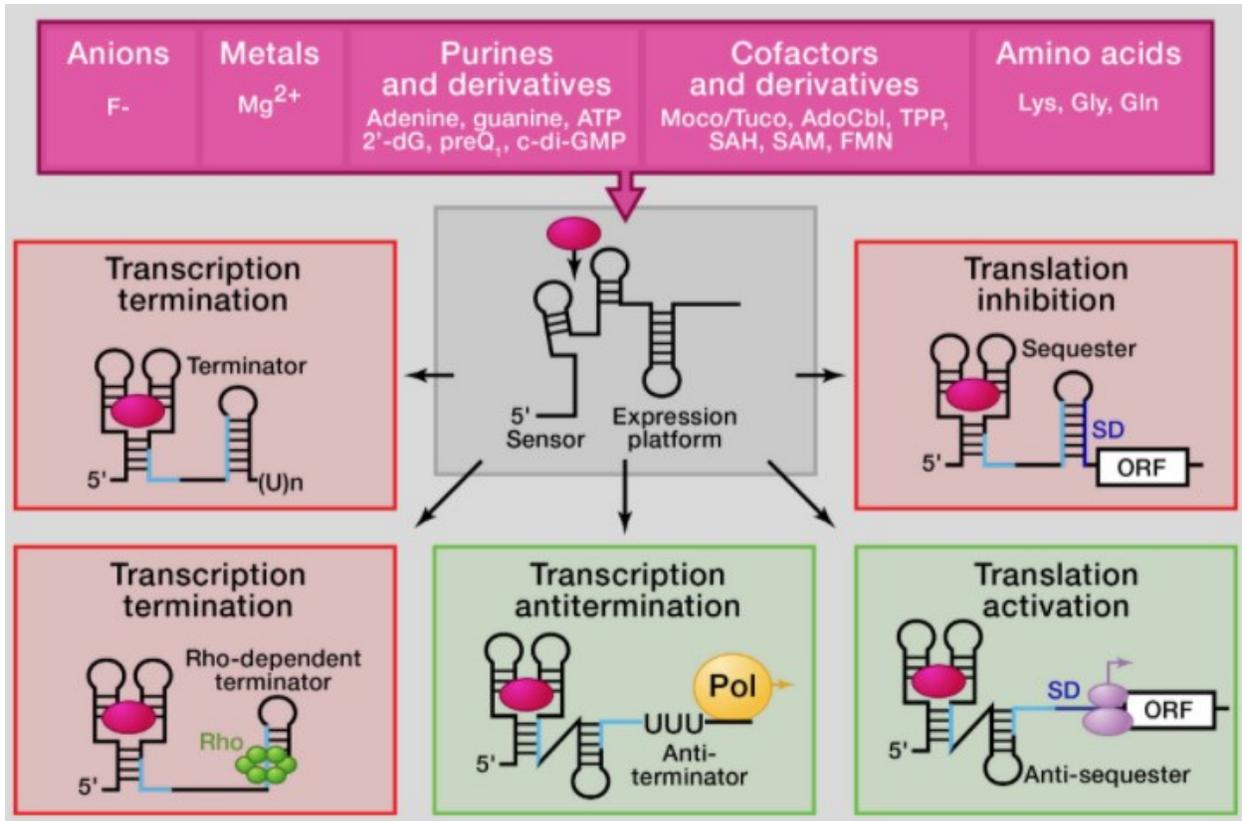


Figure 3. Mécanismes régulateurs couramment utilisés par les ribo-interrupteurs.

Exemples de ligands (encadré rose, haut de la figure) pouvant être liés par des ribo-interrupteurs (encadré gris au centre) et leur classification, soient (de gauche à droite) des anions, des métaux, des purines et leurs dérivés, des cofacteurs et leurs dérivés, des acides aminés. Sont également montrés au centre et au bas de la figure plusieurs mécanismes d'activation (encadrés verts) et d'inhibition (encadrés roses) de la transcription et de la traduction à la liaison d'un ligand (cercle rose). Les mécanismes intervenant au niveau de la transcription peuvent impliquer des terminateurs Rho-indépendant (centre gauche) ou dépendant (coin inférieur gauche) menant à la fin prématurée de la transcription ou encore des antiterminateurs qui libéreront l'accès à l'ARN polymérase (Pol) (en bas au centre) afin de permettre la transcription. Les mécanismes intervenant au niveau de la traduction impliquent plutôt la séquestration de la séquence de Shine-Dalgarno (SD) situé en amont du cadre ouvert de lecture (ORF) inhibant ainsi la traduction (centre droit) ou encore la libération de la séquence de SD permettant ainsi à la traduction d'avoir lieu (coin inférieur droit). Tiré de (Serganov & Nudler, 2013).

1.3 Familles ribo-interrupteurs guanidine

Lorsqu'on a affaire à un ribo-interrupteur dont le ligand est encore inconnu, celui-ci est qualifié d'orphelin. Pour trouver son ligand, la méthode est généralement simple : regarder la fonction du gène en aval. En effet, la fonction du gène peut donner une indication sur le ligand du ribo-interrupteur. Par exemple, si la fonction du gène associé au ribo-interrupteur est le transport d'un composé XYZ, il est fort probable que le composé XYZ soit le ligand recherché du ribo-interrupteur (Breaker *et al.*, 2017; Meyer *et al.*, 2011). Pour les membres des familles des ribo-interrupteurs guanidine, cela s'est toutefois avéré être plus difficile depuis la découverte du premier motif appelé motif *ykkC/yxkD*.

1.3.1 Guanidine-I

Motif et gènes associés

Le motif *ykkC/yxkD* a été découvert lors d'une analyse de séquences comparatives qui avait pour but de trouver de nouveaux ribo-interrupteurs (Barrick *et al.*, 2004). Le contexte génomique de ce candidat était, toutefois, vaste et diversifié ce qui a rendu fastidieuse la découverte de son ligand (Meyer *et al.*, 2011). En effet, ce motif possède deux sous-types qui répondent à des ligands différents. Le ligand spécifique du premier sous-type (Figure 4 A) est la guanidine et correspond donc à la famille de ribo-interrupteur guanidine-I, également appelé *ykkC*. Pour ce qui est du deuxième sous-type celui-ci est décliné en quatre familles différentes, soient 2a, 2b, 2c et 2d (non représentés). Tout comme le sous-type 1, ces derniers sont également associés à une variété de gènes, mais différents de ceux retrouvés sous le contrôle du sous-type 1 (Sherlock & Breaker, 2020; Sherlock *et al.*, 2019). Les sous-types 2a et 2b répondent respectivement aux ligands ppGpp (guanosine tétraphosphate, soit un second messager important dans la réponse stringente) et PRPP (phosphoribosylpyrophosphate impliqué, entre autres, dans le métabolisme des purines) (Sherlock & Breaker, 2020; Sherlock *et al.*, 2019). Quant au sous-type 2c, des essais par cartographie *in vitro* par *in-line* montrent qu'il répond, entre autres, à l'adénosine et la cytidine diphosphates, mais l'identité précise de son ligand naturel reste encore à être déterminée (Sherlock *et al.*, 2019). En effet, plusieurs représentants de ce sous-type ont montré une capacité à pouvoir lier un nucléoside diphosphate d'ARN ou d'ADN avec une affinité similaire (Sherlock & Breaker, 2020). Il est donc possible que le sous-type 2c réponde à une plus grande diversité de nucléosides diphosphates qu'initialement présentée (Sherlock & Breaker, 2020 ; Sherlock *et al.*, 2019). En ce qui concerne le dernier sous-type 2d, celui-ci est encore à ce jour orphelin (Sherlock

& Breaker, 2020). Dans le cas de la présente recherche, seul le motif *ykkC* sous-type 1 sera discuté.

Plusieurs gènes auxquels l'ARN *ykkC* est associé encodent, entre autres, des arginases, des carboxylases ou des allophanates hydrolases, mais également plusieurs transporteurs comme des transporteurs de bicarbonate de type ABC ou associés à la résistance aux médicaments comme SugE et YkkC (Figure 4 B) (Barrick *et al.*, 2004 ; Meyer *et al.*, 2011). C'est d'ailleurs de cette dernière protéine que vient le nom de cette première famille de ribo-interrupteur guanidine. Devant cette diversité de fonctions, mais surtout de transporteurs auxquels il est associé, Nelson et collaborateurs (Nelson *et al.*, 2017) émettent l'hypothèse que ce régulateur reconnaît possiblement un composé toxique pour la bactérie. Après environ 2000 conditions testées, le groupe de chercheurs démontre, en effet, que ce ribo-interrupteur répond à l'ajout de guanidine (Nelson *et al.*, 2017). Avec cette découverte, plusieurs des gènes retrouvés en aval du ribo-interrupteur tels que l'urée carboxylase ont dû être rebaptisés au nom du nouveau ligand, à savoir guanidine carboxylase. Il a en effet été démontré que cette enzyme avait 40 fois plus d'affinité pour la guanidine que pour l'urée (Nelson *et al.*, 2017).

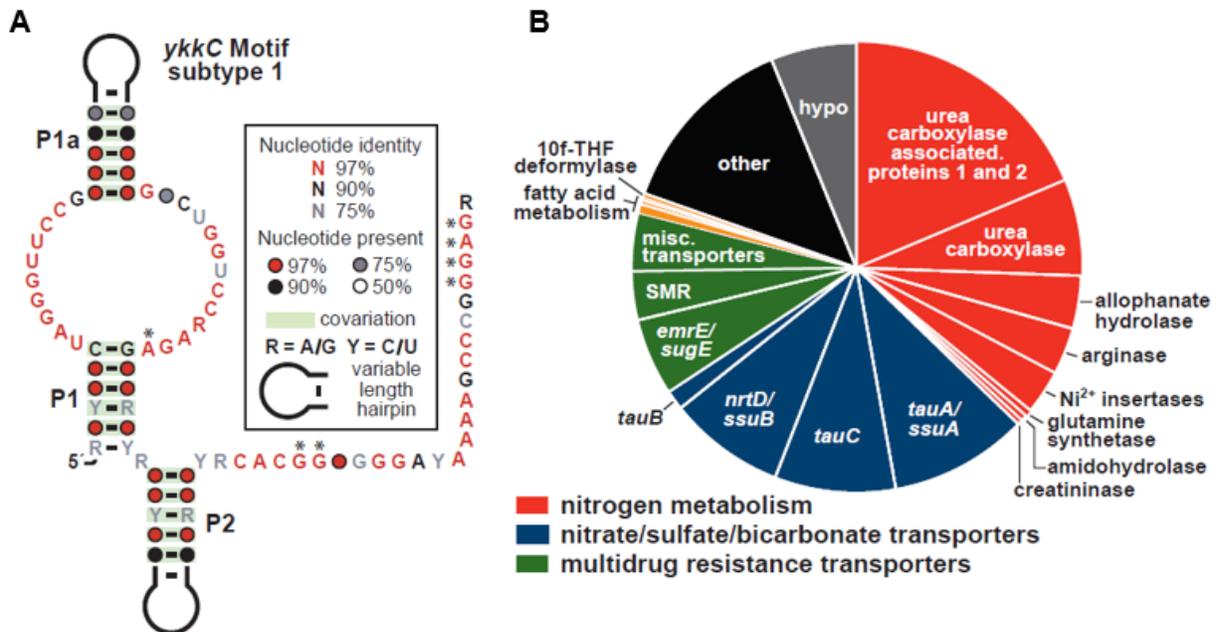


Figure 4. Le motif *ykkC* et les protéines associées.

A) Séquence consensus et structure secondaire du motif *ykkC*. Le modèle a été obtenu après analyse d'environ 1500 motifs *ykkC* sous-type 1. Les astérisques mettent en évidence les nucléotides qui diffèrent dans le motif sous-type 2 (non représenté). B) Prédiction de protéines contrôlées par le motif sous-type 1. Tiré de (Nelson *et al.*, 2017).

Peu de choses sont connues sur la biologie de la guanidine incluant son rôle physiologique. Jusqu'alors, ce composé était surtout connu pour sa capacité dénaturante à haute concentration (Greene & Pace, 1974). En observant sa structure chimique, il est toutefois possible d'observer quelques similitudes avec d'autres métabolites connus tels que l'arginine, l'agmatine et l'urée (Figure 5) (Battaglia *et al.*, 2017). Au cours de leur recherche pour identifier le ligand de ce ribo-interrupteur, Nelson et collaborateurs (Nelson *et al.*, 2017) ont démontré que ce métabolite est produit naturellement par les bactéries. L'hypothèse a donc été émise que ce métabolite serait produit lorsque ces microorganismes n'auraient pas accès à leur source préférentielle d'énergie telle que le glucose et que les familles de ribo-interrupteur jusqu'alors découvertes serviraient donc uniquement à activer des gènes servant à la détoxification.

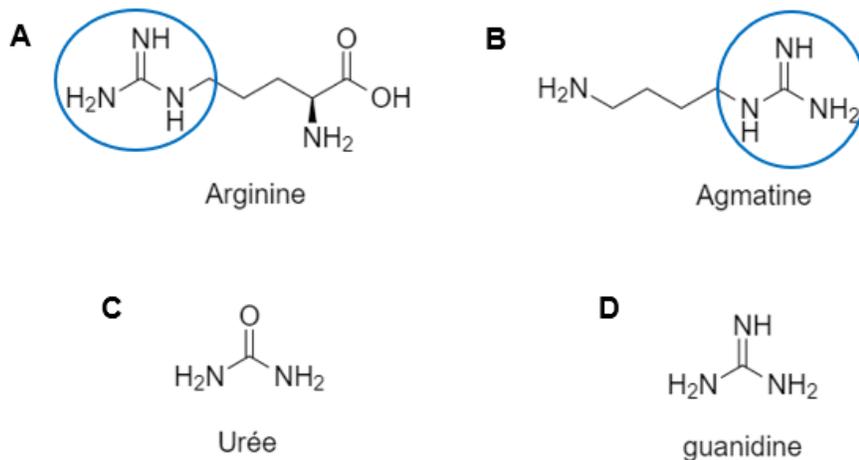


Figure 5. Structures chimiques de métabolites.

A) Structure chimique de l'acide aminé arginine, B) Structure chimique de l'agmatine C) Structure chimique de l'urée et D) Structure chimique de la guanidine. Les groupements chimiques guanidine présents chez l'arginine et l'agmatine sont mis en évidence par les cercles bleus.

Structures secondaire et tertiaire

Le ribo-interrupteur guanidine-I lie spécifiquement la guanidine avec beaucoup d'affinité, tout en excluant des molécules similaires par leurs structures ou qui possèdent le même groupement chimique telles que l'arginine, l'agmatine ou l'urée (

Figure 5), bien que ces dernières soient aussi présentes en concentrations appréciables chez la bactérie (Caldara *et al.*, 2008; Hamana, 1996; Nelson *et al.*, 2017). Cette spécificité est possible grâce à la pochette de liaison du ribo-interrupteur. Afin de mieux comprendre les mécanismes de liaison de la guanidine au ribo-interrupteur guanidine-I, des chercheurs ont résolu la structure du régulateur lié au métabolite par cristallographie aux rayons X (Battaglia *et al.*, 2017; Reiss *et al.*, 2017).

La structure 3D a permis de révéler des éléments structuraux additionnels (Figure 6 A et B) (Battaglia *et al.*, 2017; Reiss *et al.*, 2017) qui n'avaient initialement pas été trouvés par les recherches bio-informatiques (Nelson *et al.*, 2017) (Figure 4). Avec ces nouveaux résultats, la structure secondaire consensus du ribo-interrupteur guanidine-I a donc été mise à jour. Celle-ci révèle que la grande boucle interne initialement révélée par prédiction bio-informatique (Nelson *et al.*, 2017) forme une hélice nommée P1a et située entre les hélices P1 et P1b. De plus, une hélice supplémentaire, P3, est également révélée en 3' (Reiss *et al.*, 2017) dans ce qui avait initialement été prédit comme une longue séquence simple brin par Nelson et collaborateurs (Nelson *et al.*, 2017).

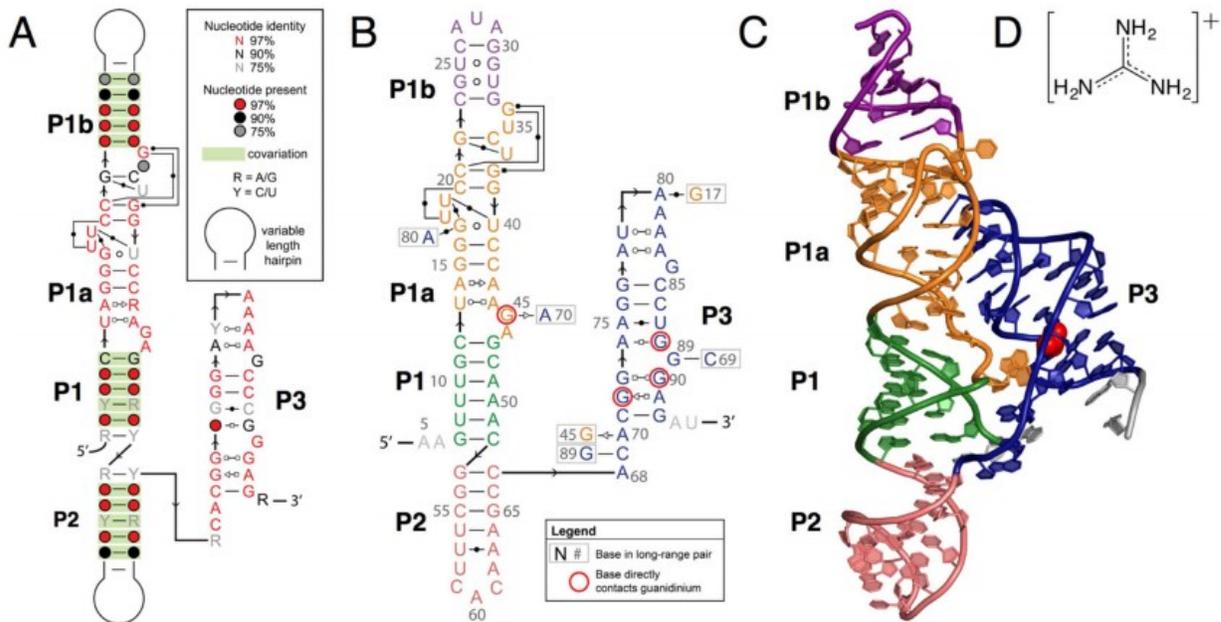


Figure 6. Consensus mis à jour du ribo-interrupteur guanidine-I.

A) Consensus mis à jour du motif ribo-interrupteur guanidine-I reflétant les observations de la structure du cristal. Les éléments pairs (hélices) sont notés par l'abréviation P. Le motif comporte donc, de 5' en 3', cinq hélices notées P1a, P1b, P1, P2 et P3 respectivement. **B)** Structure secondaire du cristal provenant de *Sulfolobus acidophilus*. Les hélices P1, P1a, P1b, P2 et P3 sont montrés en vert, orange, violet, saumon et bleu respectivement. Les nucléotides directement en contact avec le ligand sont encerclés en rouge. **C)** Diagramme en ruban de la structure tertiaire du ribo-interrupteur guanidine-I de *S. acidophilus*. Les couleurs sont les mêmes qu'en B). Le ligand guanidinium est montré par des sphères rouges. **D)** Structure chimique du ligand guanidine sous sa forme ionique (guanidinium). Tiré de (Reiss *et al.*, 2017).

La structure montre également que la pochette de liaison est formée entre les hélices P3 et P1a. Plus précisément, le ribo-interrupteur lie le ligand sous sa forme ionique via différents ponts hydrogènes (Figure 6 B) (Reiss *et al.*, 2017). Toujours selon cette recherche, la proximité du ligand chargé positivement avec le phosphate chargé négativement en G73 procurerait un effet stabilisateur ionique pour la liaison de la guanidine. Un modèle de surface également effectué par le même groupe montre que 99,5% du cation est engouffré par le ribo-interrupteur. En fait, la pochette de liaison, telle qu'elle se forme, ne laisserait pas de place à une molécule plus grosse que la guanidine pour s'y lier sauf pour un atome de plus via une petite pochette supplémentaire présente près d'un des groupes amino de la guanidine. Ceci explique pourquoi une précédente étude (Nelson *et al.*, 2017) a montré que des analogues de la guanidine tels que le méthylguanidine ou l'aminoguanidine peuvent se lier au ribo-interrupteur, alors qu'une molécule aussi large que l'arginine en est incapable bien que cette dernière possède un groupement

guanidinium. Cette même recherche montrait également que l'affinité du ribo-interrupteur se situant à l'échelle du μM (Nelson *et al.*, 2017) était beaucoup plus grande pour la guanidine que ses analogues confirmant ainsi la spécificité du ligand identifié. Cette même étude montre que l'aptamère de ce ribo-interrupteur est souvent associé à un terminateur de transcription. Des essais *in vitro* ont, en fait, montré que le ribo-interrupteur, une fois lié à son ligand, permettait d'obtenir des transcrits entiers suggérant ainsi un mécanisme d'activation de la transcription pour ce ribo-interrupteur.

1.3.2 Guanidine-II

Motif et gènes associés

Le motif mini-*ykkC* a d'abord été découvert suite à une recherche en génomique comparative et a été nommé basé sur sa petite taille et la similarité qu'il y avait entre les gènes auxquels il était associé et ceux du motif ARN *ykkC* (Weinberg *et al.*, 2007). Cet ARN ne faisait, toutefois, pas partie des candidats potentiels pour un ribo-interrupteur. En effet, la structure simpliste de mini-*ykkC* ainsi que son nombre peu élevé de nucléotides conservés ne laissait pas présager qu'il pourrait agir comme régulateur (Weinberg *et al.*, 2007). Néanmoins, suite à l'identification du ligand pour le ribo-interrupteur guanidine-I, Sherlock et collaborateurs (Sherlock *et al.*, 2017) émettent l'hypothèse que ce motif pourrait également lier la guanidine en présence d'un cofacteur, la présence d'une répétition de structures en tige-boucle dans le motif suggérant un domaine de liaison aux protéines (Figure 7 A). Or, cela s'est avéré ne pas être le cas. En effet, il a été montré que le motif mini-*ykkC* était capable de lier directement le cation guanidinium et d'activer l'expression des gènes auxquels il était associé. Le motif mini-*ykkC* a ainsi été renommé ribo-interrupteur guanidine-II (Sherlock *et al.*, 2017).

Plus précisément, le ribo-interrupteur guanidine-II est composé de deux tiges riches en GC nommées P1 et P2. Chacune des tiges se ferme sur une boucle ACGR, R représentant une purine A ou G. Ces deux tiges-boucles, ou épingles à cheveux, sont ainsi reliées par une séquence variant de 7 à 40 nucléotides. La recherche bio-informatique a révélé que le motif se trouvait souvent à proximité d'une séquence de Shine-Dalgarno, le site de fixation des ribosomes, et d'un codon de départ. Il est également montré que le motif est associé aux mêmes gènes que le ribo-interrupteur guanidine-I. En effet, dans ce cas-ci, on retrouve notamment des gènes encodant pour des carboxylases, des allophanates hydrolases, de même que des gènes appartenant à la famille des *small multidrug resistance* (SMR) encodant pour des transporteurs (Figure 7 B). Ceci, en combinaison avec la toxicité de la guanidine à haute concentration, suggère que le ligand, une

fois lié au domaine aptamère du ribo-interrupteur, libère la séquence de Shine-Dalgarno, autrement séquestrée, et permet la traduction de l'ARNm en protéines. Ceci dans le but de réduire la concentration intracellulaire de guanidine (Sherlock *et al.*, 2017).

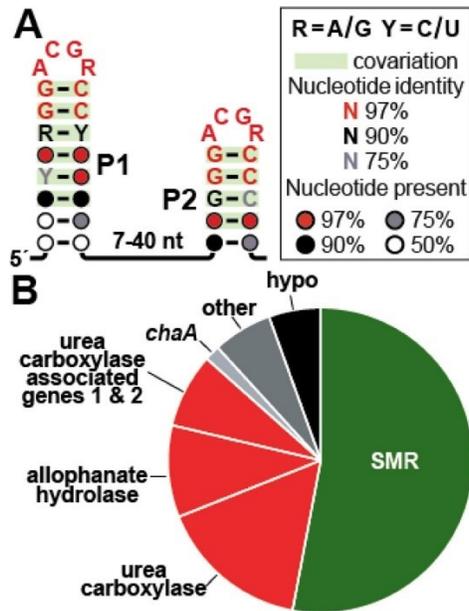


Figure 7. Structure du ribo-interrupteur guanidine-II et les protéines associées.

A) Séquence consensus et modèle structurel du ribo-interrupteur guanidine-II. B) Protéines associées au ribo-interrupteur guanidine-II incluant des transporteurs faisant partie de la famille des *Small Multidrug Resistance* (SMR) (vert), des urées carboxylases et leurs protéines associées (rouge), d'autres protéines (gris) et des protéines hypothétiques (hypo, noir). Reproduit avec la permission de (Sherlock *et al.*, 2017). Droits d'auteur (2021) *American Chemical Society*.

Structures secondaire et tertiaire

Afin de mieux comprendre la liaison de la guanidine pour ce ribo-interrupteur, deux groupes de chercheurs ont résolu la structure du ribo-interrupteur par cristallographie aux rayons X (Huang *et al.*, 2017a ; Reiss & Strobel, 2017). De prime abord, la petite structure de ribo-interrupteur guanidine-II laissait présager que cette famille aurait une moins bonne sélectivité pour son ligand comparativement à guanidine-I (Huang *et al.*, 2017a ; Sherlock *et al.*, 2017). En effet, la structure 3D de guanidine-II révèle que comparativement à guanidine-I qui engouffre complètement le ligand lors de sa liaison, guanidine-II a moins d'interactions directes avec le ligand (Figure 8 B et C). Ce dernier possède également une petite ouverture sur le côté de la pochette pouvant accommoder une petite chaîne. C'est ce qui contribue à la moins grande sélectivité observée pour ce ribo-interrupteur comparativement à la 1^{re} famille (Huang *et al.*, 2017a; Reiss & Strobel, 2017). Ainsi, tout comme pour guanidine-I, la pochette de liaison de guanidine-II laisse suffisamment de place pour des analogues à la guanidine tels que le méthylguanidine et discrimine contre des molécules plus grosses tels que l'arginine et l'agmatine en empêchant stériquement les deux dernières de se lier simultanément (Huang *et al.*, 2017a; Nelson *et al.*, 2017; Sherlock *et al.*, 2017). L'affinité du ribo-interrupteur observée pour les analogues est toutefois moindre (c'est-à-dire que le K_D est plus grand) faisant de la guanidine le ligand spécifique de guanidine-II (Sherlock *et al.*, 2017).

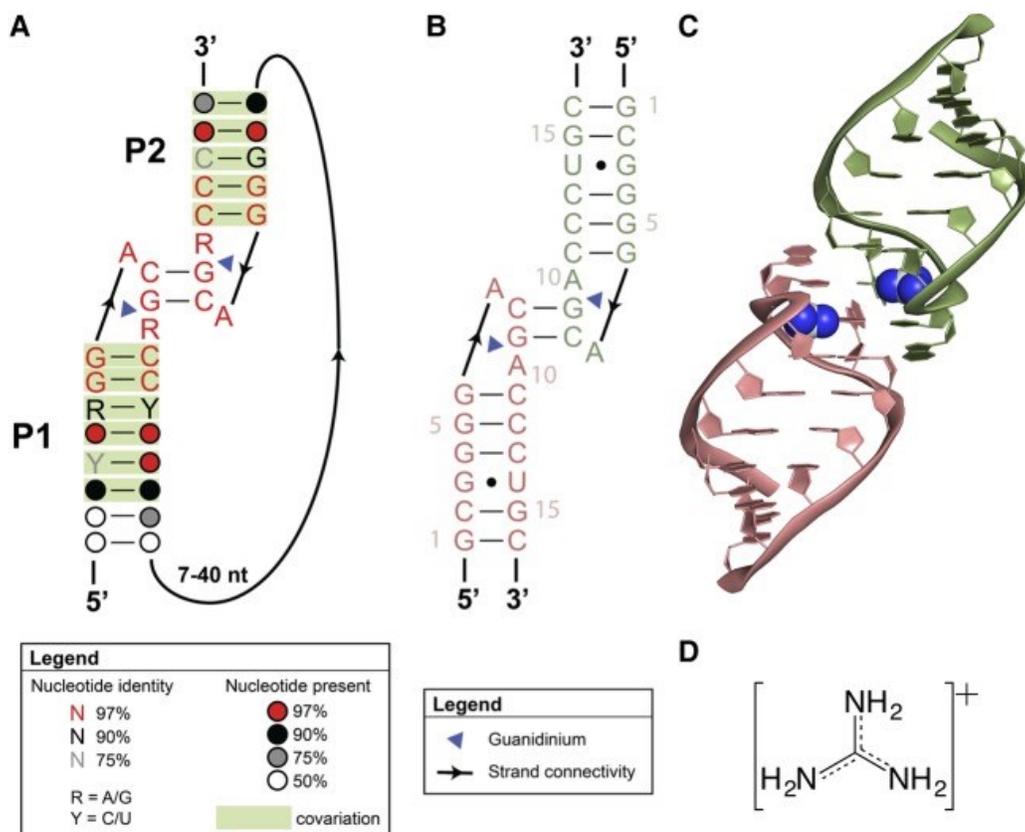


Figure 8. Structures secondaire et tertiaire du ribo-interrupteur guanidine-II.

A) Séquence consensus mise à jour afin de refléter la structure du cristal. B) Structure secondaire du crystal de ribo-interrupteur guanidine-II provenant de *Pseudomonas aeruginosa*. Les deux structures en épingle à cheveux du ribo-interrupteur sont distinguées par les couleurs rose et verte. Les triangles bleus représentent le cation guanidinium. C) Diagramme en ruban de la structure 3D du ribo-interrupteur guanidine-II (*P. aeruginosa*). Les couleurs sont les mêmes qu'en B). Les cations guanidines sont représentés par des sphères bleues. D) Structure chimique du cation guanidinium. Tiré de (Reiss & Strobel, 2017).

Plus précisément, le crystal du ribo-interrupteur guanidine-II obtenu à partir de *Pseudomonas aeruginosa* montre que les deux structures en épingle à cheveux sont séparées par une rotation de 180° ce qui n'était pas observé dans la première structure secondaire consensus du ribo-interrupteur (Figure 7 et Figure 8). Cette dimérisation forme les deux pochettes de liaison, alors que la liaison de la guanidine se fait strictement au niveau de la boucle ACGR et la première paire de bases GC se trouvant au-dessous (Reiss & Strobel, 2017). Ceci explique également pourquoi chacune des tiges est également capable de lier le cation seul (Huang *et al.*, 2017a ; Reiss & Strobel, 2017; Sherlock *et al.*, 2017). Cette dimérisation est, en fait, au centre de la fonction du ribo-interrupteur qui lie son ligand de manière coopérative (Huang *et al.*, 2017a ;

Wuebben *et al.*, 2020). Grâce à cette coopérativité, l'affinité du ribo-interrupteur pour le ligand en est augmentée et se fait dans des concentrations dans l'ordre du μM (Sherlock *et al.*, 2017). Ceci nous donne une idée du niveau de toxicité de la guanidine pour les bactéries. Une concentration intracellulaire au-delà du μM serait donc néfaste pour ces microorganismes puisque c'est à ce moment que les gènes permettant la détoxification sont activés par le ribo-interrupteur (Huang *et al.*, 2017a; Sherlock *et al.*, 2017).

1.3.3 Guanidine-III

Motif et gènes associés

Le motif *ykkC*-III a été découvert à la suite d'une analyse génomique comparative effectuée au sein de bactéries et d'archaea (Weinberg *et al.*, 2010). L'hypothèse émise était que ce potentiel ribo-interrupteur pourrait répondre au même ligand que les deux premières familles de ribo-interrupteurs (Weinberg *et al.*, 2010). En effet, bien que *ykkC*-III ne possédât aucune similarité au niveau de sa structure primaire ou secondaire avec *ykkC*, il était toutefois associé aux mêmes gènes que ce dernier notamment ceux encodant pour des transporteurs SMR (Figure 9) (Sherlock & Breaker, 2017). De plus, le motif *ykkC*-III possède également des similarités au niveau de sa séquence conservée avec celle de guanidine-II. Ce dernier comporte deux boucles ACGR conservées et cette séquence est également retrouvée chez *ykkC*-III bien qu'elle ne se présente pas sous la forme d'une boucle. (Sherlock & Breaker, 2017; Weinberg *et al.*, 2010). Tout comme, guanidine-II, le motif *ykkC*-III se présente également souvent à proximité d'une séquence de SD. Tout ceci suggère que cette famille de ribo-interrupteurs agirait également comme un activateur de la traduction qui permettrait de renverser la toxicité de la guanidine lorsque sa concentration intracellulaire est trop élevée (Sherlock & Breaker, 2017). Cette hypothèse a été confirmée via cartographie *in vitro* par *in-line* montrant que ce motif constitue, en effet, une 3^e famille de ribo-interrupteur guanidine avec une affinité similaire à celle de guanidine-I et II.

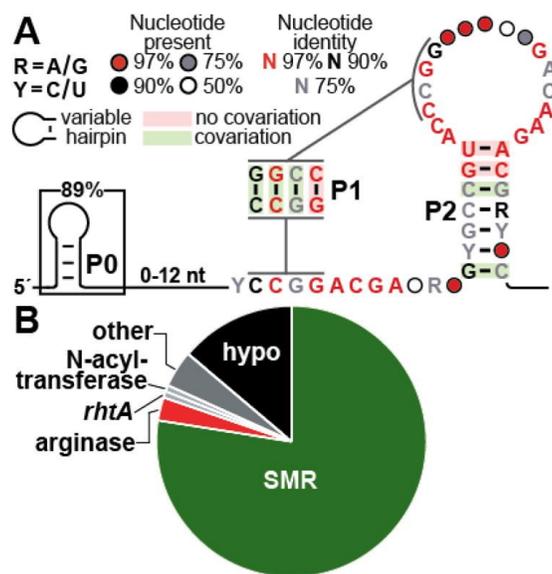


Figure 9. Structure du ribo-interrupteur guanidine-III et les protéines associées.

A) Séquence consensus et modèle structural de guanidine-III. B) Protéines dont les gènes qui les encodent sont sous le contrôle du ribo-interrupteur. On y retrouve les transporteurs de la famille des *small multidrug resistance* (SMR) (vert), les arginases (rouge), les autres protéines (gris) et des protéines hypothétiques (hypo, noir). Reproduit avec la permission de (Sherlock & Breaker, 2017). Droits d'auteur (2021) *American Chemical Society*.

Structures secondaire et tertiaire

La structure secondaire du ribo-interrupteur guanidine-III présente un pseudonoeud, une structure tertiaire de l'ARN, comprenant une partie des éléments notés P1 et P2. On observe également la présence de l'élément noté P0 dans la séquence consensus de Sherlock et collaborateurs (Sherlock & Breaker, 2017) qui ne contient pas de nucléotides hautement conservés (Figure 9). Ceci expliquerait également l'absence de ce dernier dans certains des exemples du motif trouvés. L'étude a d'ailleurs montré par cartographie *in vitro* par *in-line* que la délétion de cet élément n'affectait pas de manière significative l'affinité du ribo-interrupteur pour son ligand (Sherlock & Breaker, 2017). Il est par contre possible qu'*in vivo*, cette hélice ait un autre rôle biologique ou biochimique que la liaison du ligand ((Sherlock & Breaker, 2017). Afin de mieux comprendre l'affinité et la spécificité de ce ribo-interrupteur pour le ligand guanidine, la structure de cette famille de ribo-interrupteurs a également été résolue par cristallographie aux rayons X (Figure 10 E à G) (Huang *et al.*, 2017b). Le repliement de guanidine-III se fait de manière complètement différente de guanidine-I et II. En effet, le pseudonoeud présent dans cette structure contribue à la formation d'une triple hélice en P2 et qui forme ainsi la pochette de liaison. Dans le cas de guanidine-I, on

observait plutôt la présence de deux tiges-boucles, l'une d'entre elles empêchant la formation d'un terminateur de transcription lorsque le ligand était présent (Battaglia *et al.*, 2017 ; Reiss *et al.*, 2017). Dans le cas de guanidine-II, on observait également la présence de deux tiges-boucles, mais celle-ci interagissaient entre elles afin de former les pochettes de liaison (Huang *et al.*, 2017a ; Reiss & Strobel, 2017). Bien que le repliement ne se fasse pas de la même manière pour guanidine-III, on observe que, dans les trois cas, le ligand interagit principalement avec des bases guanine de l'ARN formant des ponts hydrogènes avec ceux-ci pour s'y lier (Battaglia *et al.*, 2017 ; Huang *et al.*, 2017a ; Huang *et al.*, 2017b ; Reiss & Strobel, 2017 ; Reiss *et al.*, 2017). De plus, tout comme guanidine-II, guanidine-III possède une petite ouverture sur le côté de sa pochette de liaison ce qui le rend moins sélectif que guanidine-I. En effet, il est montré que cette ouverture est capable d'accueillir une petite chaîne latérale ce qui explique pourquoi ce ribo-interrupteur est également capable de lier des analogues de la guanidine tels que le méthylguanidine ainsi qu'une molécule plus grosse telle que l'agmatine. Toutefois, contrairement à guanidine-II, il n'a pas été possible de mesurer l'affinité de guanidine-III pour l'aa arginine (Sherlock & Breaker, 2017). Huang et collaborateurs (Huang *et al.*, 2017b) postulent que ceci pourrait être dû au groupement carboxylique présent dans cet aa qui ajouterait une déstabilisation empêchant de mesurer l'affinité de l'ARN pour ce ligand.

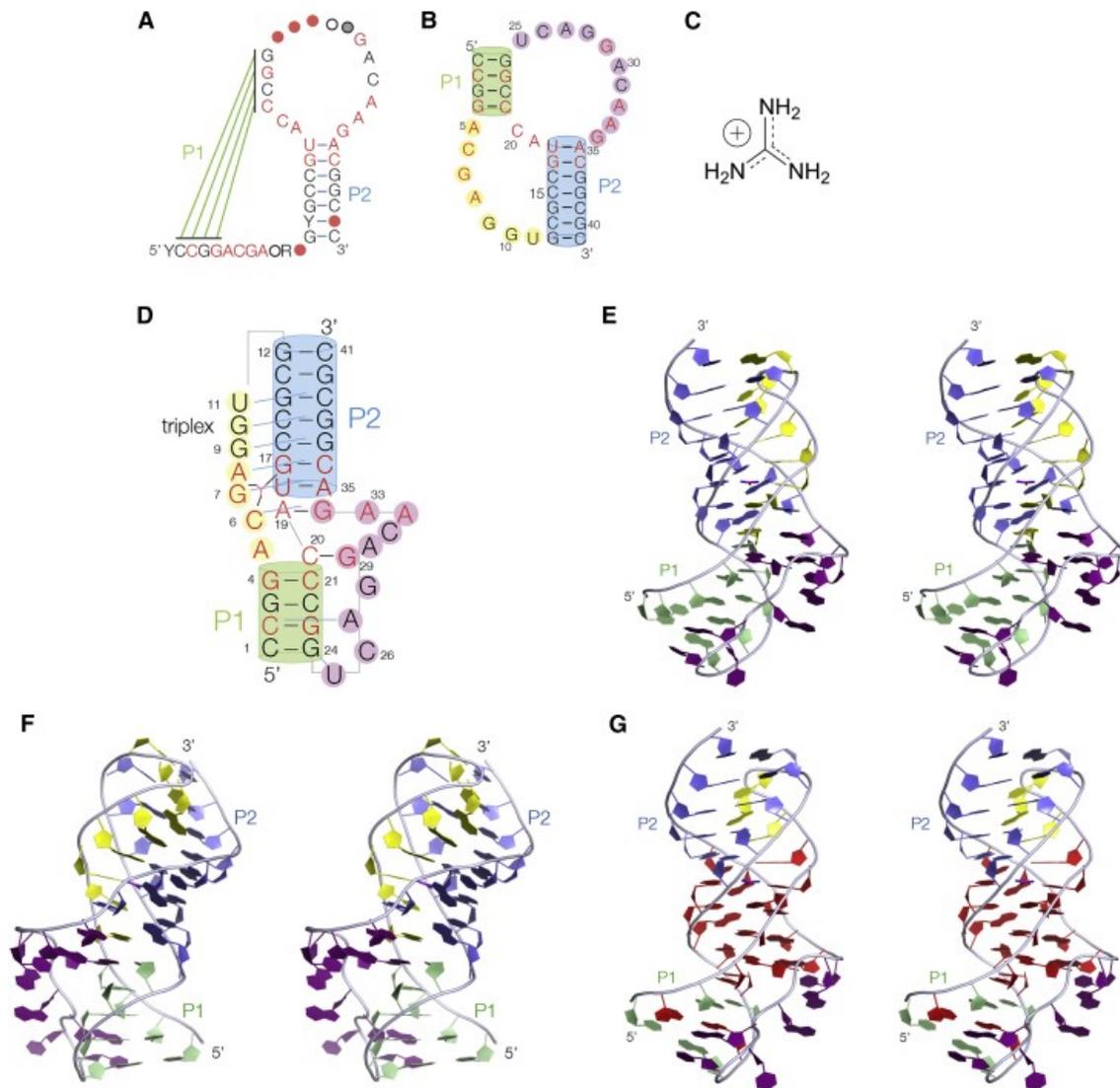


Figure 10. Structures secondaire et tertiaire du ribo-interrupteur guanidine-III.

A) Séquence et structure secondaire du ribo-interrupteur guanidine-III. Adapté de (Sherlock & Breaker, 2017) par l'auteur original (Huang *et al.*, 2017b). B) Séquence du ribo-interrupteur guanidine-III de *Thermobifida fusca* repliée dans sa structure secondaire comprenant le pseudonoeud ainsi que les tiges-boucles P1 et P2. C) Structure du cation guanidinium. D) Illustration schématique de la structure. Les nucléotides conservés à 97% sont en rouge. Les liaisons hydrogène sont indiquées par les lignes cyan. E) et F) Deux vues de la structure du cristal provenant de deux côtés opposés. G) Vue de la structure du cristal provenant de la même perspective qu'en E). Les nucléotides conservés à 97% sont représentés en rouge ce qui correspond également à l'entièreté de la section centrale. Reproduit avec la permission de (Huang *et al.*, 2017b). Droits d'auteur (2021) *American Chemical Society*.

1.3.4 Guanidine IV

Motif et gènes associés

Le motif *mepA* a été découvert à la suite d'une recherche bio-informatique qui avait pour but de découvrir de nouveaux ribo-interrupteurs et ARNnc (Salvail *et al.*, 2020). Il porte, en fait, le nom d'un des gènes auquel il est souvent associé. Le motif comporte deux hélices annotées P1 et P2 ainsi qu'un terminateur de transcription intrinsèque. L'hélice P1 et le terminateur sont liés par une séquence variant de 2 à 11 nucléotides. Les nucléotides conservés sont principalement trouvés près de la boucle de P1 et celle du terminateur de transcription (Figure 11 A). Les gènes associés à ce motif se trouvent également être associés aux précédentes familles de ribo-interrupteur incluant *mepA* qui fait partie des familles des SMR (Figure 11 B). Par ailleurs, en effectuant une recherche des régions non traduites de ces gènes, un autre groupe de recherche, Lenkeit et collaborateurs (Lenkeit *et al.*, 2020), est parvenu à trouver le même motif qu'ils ont appelé *Guanidine-Gene-Associated Motifs 1* (GGAM-1). Comme son nom l'indique, le but de cette recherche était de pouvoir spécifiquement trouver une autre classe de ribo-interrupteur guanidine. Bien que les deux groupes de chercheurs n'aient pas utilisé la même démarche bio-informatique, ils sont tous deux parvenus à identifier la 4^e famille de ribo-interrupteurs guanidine. En effet, des essais *in vitro* et *in vivo* ont montré que ce ribo-interrupteur répondait à la guanidine avec une affinité similaire à celle des précédentes familles et qu'il agissait comme un activateur de la transcription (Lenkeit *et al.*, 2020; Salvail *et al.*, 2020).

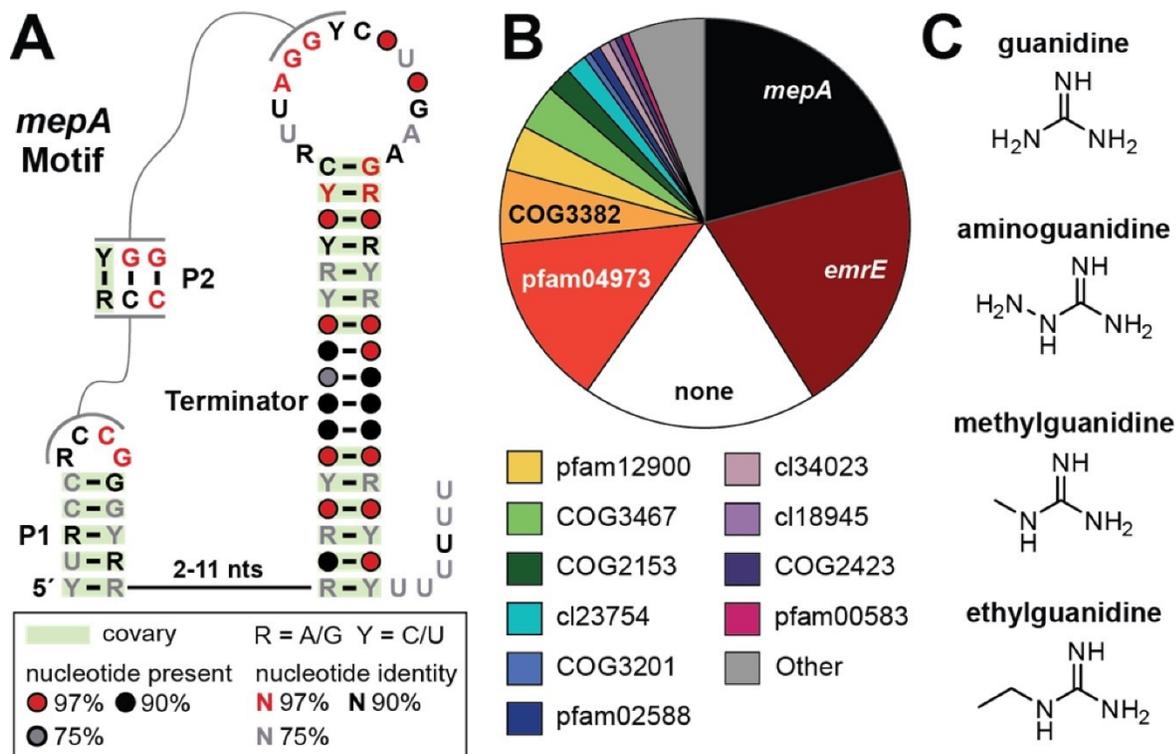


Figure 11. Structure secondaire du ribo-interrupteur guanidine IV.

A) Séquence consensus et structure secondaire du motif *mepA* (ribo-interrupteur guanidine IV) basées sur 1275 séquences homologues, mais non identiques. P1 et P2 sont des tiges prédites et pour laquelle P2 est incompatible avec la formation du terminateur de transcription intrinsèque prédit. B) Contexte génétique du motif *mepA*. Les protéines pour lesquelles les gènes les encodant sont immédiatement en aval sont rapportées (1250 au total). C) Structures chimiques de la guanidine et de ses analogues qui ont été capables d'être liés à l'aptamère de l'ARN *mepA*. Reproduit avec la permission de (Salvail et al., 2020). Droits d'auteur (2021) American Chemical Society.

Structures secondaire et tertiaire

À ce jour, aucune structure 3D du ribo-interrupteur guanidine-IV n'est disponible. Cependant, des essais *in vitro* ainsi que la prédiction de structure nous permettent de comprendre la formation du domaine aptamère de ce régulateur et ce qui constitue donc sa pochette de liaison. En effet, selon les analyses bio-informatiques et la structure secondaire prédite, la séquence liant la tige-boucle P1 avec celle du terminateur semble être trop courte pour permettre la formation des deux boucles de manière simultanée (Figure 11 A) (Salvail et al., 2020). Leur formation semble donc être mutuellement exclusive ce qui correspondrait, en effet, au mécanisme d'un ribo-interrupteur activateur de transcription. En d'autres mots, la liaison du ligand au domaine aptamère stabiliserait la formation de l'hélice P2 ce qui, à son tour, empêcherait la formation du terminateur de transcription (Salvail et al., 2020). C'est ce que Salvail et collaborateurs (Salvail et al., 2020) ont pu confirmer via cartographie *in vitro* par *in-line*. Par ces mêmes essais, ils ont également pu

confirmer que l'élément P2 se formait suite à la liaison de la guanidine et est donc important pour la fonction du domaine aptamère. D'un autre côté, la séquence liant P1 et le terminateur, quant à elle, semble relativement non structurée en l'absence de ligand. Ceci a donc pu confirmer l'hypothèse selon laquelle la liaison du ligand au domaine aptamère empêcherait la formation du terminateur de transcription (Salvail *et al.*, 2020).

Il a également été montré que ce ribo-interrupteur possède une grande sélectivité pour la guanidine, alors qu'il n'arrive pas à lier des molécules aux structures similaires telles que l'urée ou l'arginine (Lenkeit *et al.*, 2020; Salvail *et al.*, 2020). De plus, tout comme guanidine-II et III, il arrive également à lier, mais avec moins d'affinité que pour la guanidine, des analogues de ce dernier tels que le méthylguanidine, l'aminoguanidine et l'éthylguanidine (Figure 11 C) (Salvail *et al.*, 2020). Bien qu'aucune structure 3D ne soit disponible à ce jour, en se fiant aux précédentes familles du ribo-interrupteur, on peut laisser présager que la guanidine doit se lier à guanidine-IV via des ponts hydrogènes et qu'il interagirait principalement avec des bases guanine. De plus, on peut également émettre l'hypothèse que la pochette de liaison de ce ribo-interrupteur ne doit pas complètement engouffrer le ligand comme dans le cas de guanidine-I (Battaglia *et al.*, 2017; Reiss *et al.*, 2017), mais plutôt qu'elle doit posséder une petite ouverture permettant d'accueillir une petite chaîne latérale lui permettant de lier des analogues de la guanidine comme le font les ribo-interrupteurs guanidine-II et III (Huang *et al.*, 2017a ; Huang *et al.*, 2017b; Reiss & Strobel, 2017).

Récapitulatif

En somme, il existe à ce jour quatre familles de ribo-interrupteurs guanidine étant capables de répondre à ce ligand. Deux d'entre elles, guanidine-I et IV interviennent au niveau de la transcription, alors que guanidine II et III interviennent plutôt au niveau de la traduction. Guanidine-I est le ribo-interrupteur le plus sélectif des quatre familles, alors que les guanidine-II à IV arrivent également à lier des analogues de ce dernier, mais avec une affinité moindre. Toutes ces familles, mise à part guanidine-II, possèdent une structure et pochette de liaison complexes ce qu'on peut également imaginer pour guanidine-IV bien que sa structure 3D n'ait pas encore été résolue. De son côté, guanidine-II est le plus petit des quatre familles avec deux pochettes de liaison très simples ne comportant que six nucléotides chacune. Malgré sa petite taille et sa simplicité, il est également le seul des quatre familles à posséder plus qu'une pochette de liaison.

2 PROBLÉMATIQUE

La présente recherche porte sur la 2^e famille de ribo-interrupteur guanidine, soit mini-*ykkC* aussi appelé ribo-interrupteur guanidine-II. Étant donné la simplicité du motif guanidine-II, il a d'abord été supposé que ce ribo-interrupteur ne pouvait fonctionner seul, mais pourrait assurer sa fonction seulement à l'aide d'une protéine. Or, cela n'est pas le cas (Sherlock *et al.*, 2017). D'un point de vue évolutif, cela laisse donc supposer qu'un tel motif apparu au hasard et devant un gène pour lequel sa fonction serait pertinente aurait pu maintenir sa fonction et même la raffiner au cours du temps. Le motif guanidine-II, contrairement aux autres familles d'ARN plus complexes, pourraient donc avoir plusieurs origines distinctes, mais qui auraient évoluées de manière convergente. Ainsi, le ribo-interrupteur guanidine-II offre un avantage pour la recherche d'un point de vue évolutif via la simplicité de son motif, mais rend difficile la recherche *de novo* pour la découverte de nouveaux ribo-interrupteurs. En effet, les familles d'ARN tels que les ribo-interrupteurs possèdent, en général, des structures et une pochette de liaison complexes ce qui facilite la recherche *de novo* via leur structure secondaire parmi différents génomes. En fait, cela est dû à la complexité de leur structure qui laisse une empreinte distincte permettant de mieux les identifier. Toutefois, avec une pochette de liaison comportant seulement six nucléotides, et présentée sous la forme d'une simple tige-boucle, ce motif passerait inaperçu lors d'une recherche bio-informatique, car il peut assez facilement être trouvé par hasard dans des séquences suffisamment longues. Qui plus est, le motif guanidine-II, avec sa petite taille, pourrait facilement se présenter sous différentes configurations fonctionnelles, mais qui ne seraient pas repérées au cours d'une même recherche pour les mêmes raisons (soit parce que la structure serait trop simple, soit parce qu'il y aurait trop de configurations différentes, avec seul point commun la pochette de liaison de six nucléotides).

Or, la présente étude propose d'éviter cette problématique par l'utilisation d'outils bio-informatiques qui exploiteront plutôt cette simplicité du motif guanidine-II pour la découverte de nouveaux aptamères et possiblement ribo-interrupteurs guanidine. Plus précisément, deux outils bio-informatiques seront utilisés. Premièrement, la base de données RNASTEM permettant d'extraire des informations sur la structure secondaire de différents ARN connus. Deuxièmement, le logiciel RNArobo qui permettra de rechercher la pochette de liaison du motif guanidine-II dans une variété de génomes et parmi des séquences aléatoires également. Ensuite, des approches expérimentales permettront de valider les résultats bio-informatiques obtenus. De plus, comme le ribo-interrupteur possède deux pochettes de liaison correspondant à chacune des tiges-boucles,

aussi bien le motif à 1 tige qu'à 2 tiges seront utilisés au cours de cette recherche. Le présent projet offre donc également la possibilité de trouver des ribo-interrupteurs guanidine avec des configurations différentes de celle de guanidine-II.

3 HYPOTHÈSE ET OBJECTIFS

Étant donné sa simplicité, nous croyons qu'il est possible que le motif guanidine-II pourrait avoir une origine polyphylétique, mais ayant évolué de manière convergente de sorte à garder et à raffiner sa fonction au cours du temps. Nous émettons l'hypothèse que le motif guanidine-II pourrait être retrouvé parmi plusieurs séquences d'ARN connues et distribué parmi divers génomes. Nous croyons également que ces motifs nouvellement trouvés conserveraient leur capacité de lier la guanidine au niveau de la boucle, mais dans diverses configurations avec une seule ou plusieurs tiges. À cet effet, cette recherche comporte deux principaux objectifs :

1. Identifier les motifs guanidine-II via divers outils bio-informatiques, soient
 - a) RNAsstem : recherche du motif au sein d'ARN connus
 - b) RNArobo : recherche du motif au sein de séquences d'ARN provenant de différents génomes (champignons, bactéries, archaea, virus, humain) et de régions intergénomiques.
2. Évaluer la capacité de ces motifs nouvellement trouvés à lier la guanidine via des tests *in vitro*

4 MATÉRIEL ET MÉTHODES

4.1 Recherche des motifs de guanidine-II

Afin de découvrir de nouveaux ribo-interrupteurs, le motif guanidine-II à 1 tige et à 2 tiges a été recherché à l'aide d'outils bio-informatiques. Chaque outil comprend ses spécificités et avantages décrits ci-bas.

4.1.1 Banque de données RNASTem

La banque de données RNASTem a été utilisée exclusivement pour les recherches du motif à 1 tige comme décrit en 1.3.3.2. Cette banque de données est basée sur Rfam qui répertorie les familles de séquences d'ARN structurés incluant les ARN non codants et les éléments cis régulateurs. RNASTem se base donc sur cette banque de données pour permettre la recherche d'informations à partir de la structure secondaire de ces ARN. Pour le motif guanidine-II, il a donc été demandé à RNASTem de sortir tous les ARN comportant une boucle de 4 nucléotides, soient ACGR se fermant par une paire de bases GC (Figure 12).

The screenshot shows the RNASTem search interface. At the top, there is a 'Limits' section with an 'Add' button. Below this, a checkbox labeled 'Restrict result to:' is checked. Underneath, there are two radio buttons: 'Simple stem-loop' (selected) and 'Multi-stem junction stem-loop'. The search criteria are defined in three numbered steps:

- Step 1: 'ST: Loop/Junction size' is set to 'EQUAL' and the value '4'.
- Step 2: 'ST: Loop/Junction sequence' is set to 'MATCH' and the sequence 'ACG[AG]'. A 'Del' button is visible next to the dropdown.
- Step 3: 'ST: Loop/Junction closing pair' is set to 'MATCH' and the sequence 'GC'. A 'Del' button is visible next to the dropdown.

Each step is separated by a dotted line and includes an 'And' dropdown menu to indicate logical conjunction between criteria.

Figure 12. Capture d'écran de la requête RNASTem.

La requête est une boucle à 4 nucléotides (1) dont la séquence est ACGA ou bien ACGG (2) et se fermant sur une paire de bases GC (3).

4.1.2 RNArobo

Le logiciel RNArobo permet la recherche de motif d'ARN au sein de longues séquences génomiques (Rampášek *et al.*, 2016). Pour cela, il utilise une combinaison de séquences et de contraintes de structure qui lui sont dictées par le descripteur fourni par l'utilisateur. La recherche avec RNArobo a donc été utilisée afin d'étendre la recherche au-delà d'ARN déjà caractérisés c'est-à-dire en incluant des séquences aléatoires au sein de divers génomes où on retrouve la guanidine telle quelle (humain, bactéries) ou possiblement sous forme de groupement chimique (champignons, virus et bactéries). Pour cela, des descripteurs (Figure 13) détaillant les motifs à une et à deux tiges mentionnés en 1.3.2 ont été soumis au logiciel pour la recherche.

<p>A) # motif map s1 h1 s2 h1' s3</p> <p>#loop s1 0 <u>NNNNNNNNNNNNNNNNNNNNNNNN[100]</u> h1 0:0 *****<u>NNNNNNG:CNNNNNN</u>***** s2 0 ACGR s3 0 <u>NNNNNNNNNNNNNNNNNNNNNNNN[100]</u></p>	<p>B) # motif map s1 h1 s2 h1' s3 h2 s4 h2' s5</p> <p>#loop s1 0 <u>NNNNNNNNNNNNNNNNNNNNNNNN[100]</u> h1 0:0 *****<u>NNNNNNG:CNNNNNN</u>***** s2 0 ACGR s3 0 <u>NNNNNNNN[40]</u> h2 0:0 ***<u>NNNG:CNNN</u>*** s4 0 ACGR s5 0 <u>NNNNNNNNNNNNNNNNNNNNNNNN[100]</u></p>
--	--

Figure 13. Descripteurs utilisés pour la recherche RNArobo.

La recherche a été effectuée avec la version 2.1.0. Le descripteur contient une carte du motif (motif map) listant de manière individuelle chaque élément structural de la séquence de 5' à 3'. Les éléments simples brins sont identifiés par la lettre « s », alors que les éléments faisant partie d'une hélice sont identifiés par les lettres « h » et « h' » pour les brins positif et négatif respectivement. Par la suite, chaque élément structural est détaillé de manière spécifique. Le premier chiffre indique le nombre de disparités (mismatch) que l'utilisateur permet. Les éléments dits en hélice peuvent comporter un deuxième chiffre précédé d'un deux-points (:) indiquant, quant à lui, le nombre de mésappariements (mispair) permis par l'utilisateur. Puis, la séquence primaire est indiquée à l'aide des symboles de l'IUPAC. Les éléments suivis du symbole # ne sont pas lus par le logiciel et sont des commentaires fournis par l'utilisateur pour s'aider à se repérer dans le descripteur. A) Descripteur du motif guanidine-II à 1 tige. Les éléments s1 et s3 correspondent aux séquences en amont et en aval du motif respectivement allant chercher de 20 jusqu'à 100 bases avant et après celui-ci. L'hélice h1 comporte une paire de base GC qui vient fermer la boucle et minimum 6 nucléotides supplémentaires suivis de 1 à 5 nucléotides dont la présence est facultative et qui est notée par des astérisques (*). L'élément s2 correspond à la boucle ACGR. B) Descripteur du motif guanidine-II à 2 tiges. Les éléments s1, h1 et s2 sont tels que décrits en A). L'élément s3 correspond au linker de 7 à 40 nucléotides séparant les 2 tiges P1 et P2 tels que décrit 1.3.3.2. Les éléments s4 et h2 correspondent respectivement à la boucle et à la tige de P2. Finalement, l'élément s5 va chercher de 20 à 100 nucléotides en aval du motif complet.

4.1.3 Sélection des candidats

La sélection des candidats de RNAstem s'est basée sur un score établi arbitrairement et basé sur deux critères principaux à savoir la longueur des tiges (l) et le nombre de *wobbles* (w), c'est-à-dire des paires G-U ou U-G. Ainsi, le score a été calculé de la manière décrite ci-dessous :

$$score = l - (0,5 * w) \quad (1)$$

Les candidats présentant le score le plus élevé ont été sélectionnés pour les validations expérimentales. Pour ce qui est des candidats de RNArobo, la sélection s'est faite de la même manière, mais à cela s'est ajouté des critères supplémentaires, soient la conservation des séquences, la prédiction de structures, la présence de plusieurs tiges ou non et la fonction des gènes se trouvant en aval des motifs trouvés (Figure 14). Ces critères supplémentaires ont été ajoutés afin de raffiner les résultats, car le nombre de candidats obtenus était élevé.

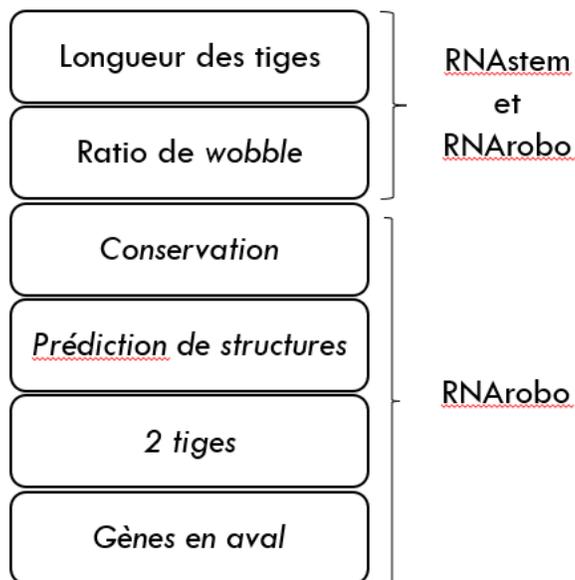


Figure 14. Critères de sélection des candidats RNAstem et RNArobo.

4.2 Transcription

4.2.1 Préparation des matrices

Séquence complète de *glnA* et *pfl*

Des oligonucléotides synthétiques d'ADN correspondant aux séquences complètes des ribo-interrupteurs *glnA* et *pfl* (ZMP/ZTP) (Table 1) ont été commandés chez Sigma-Aldrich pour amplification par PCR d'assemblage. La séquence *glnA* a été amplifiée à l'aide de deux oligonucléotides matrices (sens et anti-sens) se chevauchant avec l'oligonucléotide sens précédé du promoteur T7 ARN polymérase (T7 ARNP) suivi de trois G afin de faciliter la transcription *in vitro* à l'étape subséquente (Figure 15 A). Pour la séquence *pfl*, celui-ci a été amplifié à l'aide de quatre oligonucléotides synthétiques se chevauchant également. Le premier oligonucléotide sens toujours précédé d'un promoteur T7 ARNP suivi de trois G (Figure 15 B).

Les réactions de PCR ont été assemblées dans un volume final de 50 μ L à l'aide de 1,25 U de l'enzyme *HotStarTaq DNA polymerase* (Qiagen), de son tampon 1X (Tris-Cl ; KCl ; $(\text{NH}_4)_2\text{SO}_4$; 15 mM MgCl_2 à pH 8,7 ; 20°C) ; 200 μ M de dNTP et le volume complété à l'eau milli-Q. Pour ce qui est des matrices décrites plus haut (Table 1) elles ont été ajoutées à une concentration de 10 μ M chacune à la réaction PCR pour l'amplification de *glnA*. Pour la réaction *pfl*, les deux matrices extérieures ont été ajoutées à une concentration de 100 μ M (pour faire office d'amorces), alors que les matrices intérieures étaient à une concentration de 10 μ M chacune.

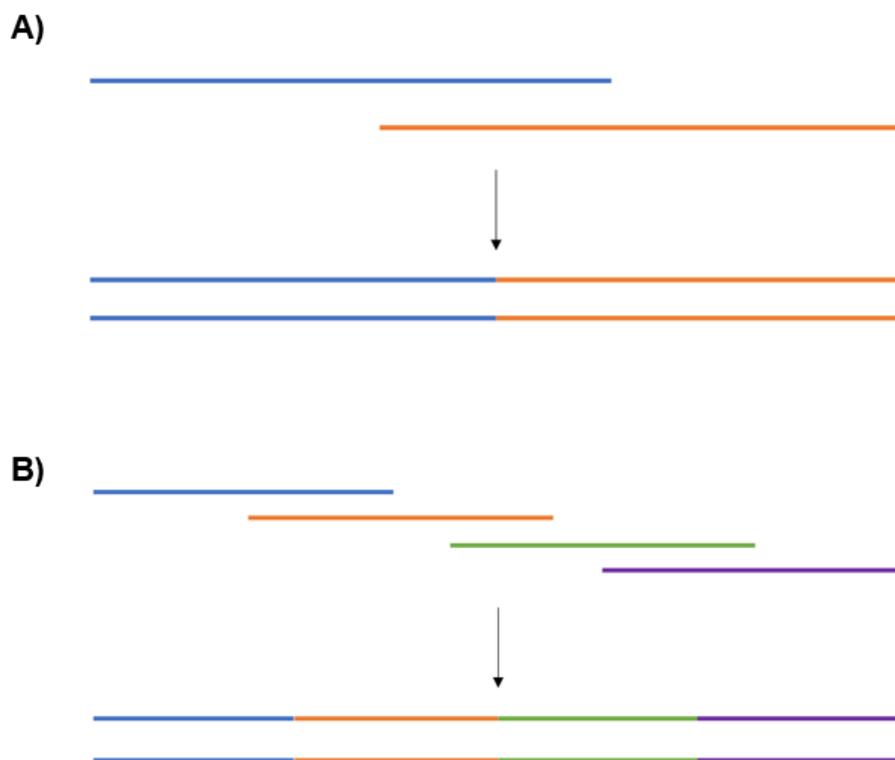


Figure 15. Schéma de la PCR d'assemblage des candidats *glnA* et *pfl*.

A) Le candidat *glnA* a été amplifié à l'aide de deux oligonucléotides synthétiques sens (bleu) et anti-sens (orange) se chevauchant. **B)** Le candidat *pfl* a été amplifié à l'aide de quatre nucléotides se chevauchant. Les oligonucléotides externes (bleu et violet) ont servi d'amorces, alors que les oligonucléotides internes (orange et vert) ont servi de matrices.

Chaque PCR a ensuite suivi le protocole suivant pendant 25 cycles (étapes 3 à 5) dans le thermocycleur C1000 touch-thermal cyler (Bio-Rad) :

1. Activation de l'enzyme : 95°C, 15 minutes
2. Dénaturation initiale : 95°C, 30 secondes
3. Dénaturation : 95°C, 30 secondes
4. Hybridation : 59°C, 30 secondes
5. Élongation : 72°C, 30 secondes
6. Élongation finale : 72°C, 5 minutes

Électrophorèse sur gel d'agarose

L'obtention des produits de PCR de la bonne taille a été vérifiée en comparant la taille des échantillons à une échelle moléculaire d'ADN de 100 paires de bases (Froggabo) sur gel d'agarose 2%. Le gel d'agarose a été préparé en dissolvant 0,8 g d'agarose (Biobasic Canada) dans 40 mL de Tris(hydroxyméthyl)aminométhane Acétate Éthylènediaminetétraacétique 1X (TAE) (soient 40 mM Tris(hydroxyméthyl)aminométhane (Tris) ; 20 mM acétate ; 1mM Éthylènediaminetétraacétique (EDTA) ; pH 8,6) et contenant 0,5 µL de GelStain (TransGen Biotech Co.). Les échantillons pour la migration sont, quant à eux, préparés de la manière suivante : 5 µL de produit PCR sont mélangés à 1 µL de bleu 6X (40% saccharose ; 0,05% xylène de cyanol et 0,05% une pincée de bleu de bromophénol). Une fois le gel polymérisé, les échantillons y sont déposés pour la migration qui se fait dans du tampon TAE 1X à 100 V pendant environ 1h ou jusqu'à ce que le bleu de bromophénol ait migré au $\frac{3}{4}$ du gel. Finalement, les échantillons sont visualisés au GelDoc et le logiciel Image Lab (Bio-Rad) grâce au GelStain qui a servi d'intercalant entre les bases d'ADN et qui est capable d'émettre une fluorescence lorsqu'exposé à la lumière ultraviolette (UV).

Les autres matrices comprenant le motif guanidine-II ont également été commandées de chez Sigma-Aldrich. Elles ont été synthétisées avec une séquence complémentaire au promoteur T7 ARNP en 5' suivi de trois G afin de faciliter la transcription, puis de la matrice codant pour l'ARN d'intérêt et ont été transcrites directement *in vitro* sans amplification telle que décrite ci-dessous.

4.2.2 Transcription *in vitro*

Toutes les matrices ont été transcrites directement *in vitro* exceptées pour les matrices *glnA* (séquence complète) et *pfl* qui ont d'abord été amplifiées par PCR comme mentionné ci-dessus. Les réactions de transcription ont été assemblées dans un volume réactionnel de 100 μ L comprenant 40 pmol de la matrice, 80 pmol de promoteur T7 ARNP (Sigma-Aldrich) pour former une matrice double brin partiel, 3 mM rNTP, du tampon 1X fait maison (80 mM HEPES-KOH pH 7,5 ; 24 mM MgCl₂ ; 40 mM DTT et 2 mM spermidine), 4U d'ARN polymérase T7 (Université de Sherbrooke), 1U d'inhibiteur de RNase (ribolock, ThermoFisher), 1U de pyrophosphatase et de l'eau milli-Q pour compléter le volume. Les réactions ont, par la suite, été incubées pendant 3h à 37°C. Les matrices d'ADN ont ensuite été dégradées en ajoutant 2U de DNase I (NEB) et en laissant incuber 10 minutes à la même température. Pour les réactions de transcription de *glnA* et de *pfl*, le même protocole a été suivi sans l'ajout de 80 pmol de promoteur T7 ARNP, les matrices étant déjà sous forme de double brin. De l'eau milli-Q a donc été ajoutée à ces réactions pour garder un volume de 100 μ L.

Extraction des ARN

Les ARN ont ensuite été purifiés par la méthode d'extraction phénol/chloroforme qui consiste à ajouter 1X le volume de phénol/chloroforme (1 :1) au mélange réactionnel de transcription. Ensuite, après un vortex de 30 secondes, les tubes sont centrifugés 14 800 rpm pendant 5 minutes. La phase aqueuse, contenant l'ARN, est ensuite récupérée dans un autre tube pour une deuxième extraction au chloroforme seulement (1X le volume) pour éliminer toutes traces de phénol. Les ARN sont ensuite précipités par l'ajout de glycogène à une concentration finale de 1 μ g/ μ L; 0,1 volume d'acétate de sodium et 2,5X volume d'éthanol 100%. Les échantillons sont ensuite incubés à -20°C pendant minimum 2h. Par la suite, les tubes sont centrifugés à 4°C pendant 30 minutes. Le culot est ensuite lavé avec 150 μ L d'éthanol 70% et centrifugé à la même température pendant 10 minutes. Le culot est, par après, récupéré et est mis à sécher. Finalement, ce dernier est réhydraté dans 20 μ L d'eau milli-Q et est gardé sur glace pour la prochaine étape de purification.

Purification sur gel de polyacrylamide

Les échantillons sont ensuite déposés dans un gel dénaturant de polyacrylamide de 20% (ratio d'acrylamide/bis-acrylamide de 19 : 1 ; 8 M urée ; 89 mM Tris base ; 89 mM acide borique, 2 mM EDTA ; pH 8) afin de compléter l'étape de purification. Du tétraméthyléthylènediamine (TEMED) et du persulfate d'ammonium (APS) sont ajoutés dans un ratio de 20 : 1 respectivement afin de catalyser la réaction de polymérisation du gel . Une fois polymérisé, les échantillons qui ont préalablement été mélangés à 1X le volume de bleu 2X dénaturant (10 mM EDTA diluée dans le formamide; 0,01% bleu de bromophénol et 0,01% xylène de cyanol), sont déposés dans le gel pour la migration dans du Tris Borate EDTA (TBE) 1X à 15 watts pendant environ 1h ou jusqu'à ce que le bleu de bromophénol ait atteint le bas du gel (en laissant 1-2 cm). Une fois la migration terminée, les ARN sont visualisés par la technique de l'ombrage ultraviolet (UV) (*UV shadowing*) et la bande de gel correspondante est découpée pour l'élution.

Élution des ARN

La bande de gel, une fois découpée, est déposée dans un tube contenant 400 µL de tampon d'élution (0,3M NaCl) pour l'élution à température pièce pendant 2h. Une fois élué, l'ARN est précipité à l'éthanol tel que décrit plus haut avant d'être quantifié au NanoDrop™ 2000c (ThermoFisher).

4.3 Marquage en 5' à l'AT³²P

Déphosphorylation des ARN

Afin de marquer en 5' les différents ARN, il faut tout d'abord les déphosphoryler. Ceci se fait à l'aide d'un mélange réactionnel de 20 µL contenant 0,25 U d'*Antartic Phosphatase* (New England BioLabs Inc.), son tampon commercial à 1X (50 mM Bis-Tris-Propane-HCl ; 1mM MgCl₂ ; 0,1 mM ZnCl₂ ; pH 6 à 25°C), 200 pmol d'ARN, et le tout complété à l'eau milli-Q. Le mélange est ensuite incubé pendant 20 minutes à 37°C avant l'inactivation de la phosphatase à 80°C pendant 2 minutes.

Marquage radioactif

Par la suite, afin de pouvoir visualiser les ARN lors des réactions de cartographie *in vitro* par *in-line* décrites ci-bas, les ARN sont marqués à la radioactivité. Pour cette étape, 10 pmol d'ARN déphosphorylé sont utilisés dans un mélange réactionnel de 20 μ L contenant également 1,25 U de T4 polynucléotide kinase (New England BioLabs Inc.) ; son tampon commercial 1X (70 mM Tris-HCl, 10 mM MgCl₂ ; 5 mM DTT ; pH 7,6 à 25°C) ; 1 μ L (10 μ Ci) de [γ -³²P] ATP (PerkinElmer) et le tout complété avec de l'eau milli-Q. Le mélange est ensuite incubé pendant 1h à 37°C. Finalement, les échantillons sont ensuite purifiés sur un gel de polyacrylamide dénaturant dans les mêmes conditions que décrites plus haut. À la fin de la migration, le gel est exposé pendant 5 minutes à une plaque phosphore *Storage Phosphor Screen* (Molecular Dynamics) (préalablement protégé d'une pellicule plastique. Une fois l'exposition terminée, la cassette est scannée par le Typhoon™ FLA 9500 (GE healthcare Life Sciences) et analysé à l'aide du logiciel ImageQuant TL v8.1. Finalement, les ARN sont élués et précipités tels que décrits plus haut avant d'être réhydratés dans 60 μ L à 90 μ L d'eau milli-Q dépendamment du nombre de réactions de cartographie effectuées (4.4).

4.4 Cartographie *in vitro* par *in-line*

Principe

La cartographie *in vitro* par *in-line* (*in-line probing*) est une technique permettant d'élucider la structure secondaire d'un ARN, mais également d'évaluer la capacité de liaison d'un ligand par le domaine aptamère d'un ribo-interrupteur. Elle repose sur le clivage spontané des ARN qui vise les liens phosphodiesters dans les régions simple brin de son squelette. Plus spécifiquement, lorsque les oxygènes en 2' (O2') et 5' (O5') sont alignés au phosphore (P) du centre, l'O2' agit comme groupe nucléophile et attaque le P du centre résultant au clivage de l'ARN (Figure 16). La fréquence des clivages dépend des structures locales de l'ARN. Ainsi, un ARN simple brin dépourvu de structures secondaires aura plus de probabilité d'adopter une conformation *in-line* permettant le clivage des liens phosphodiesters. À l'inverse, si les liens phosphodiesters se trouvent dans une structure hautement repliée de l'ARN lorsque le ligand vient se lier à celui-ci, le clivage à cet endroit ne pourra se faire. Ainsi, pour évaluer la capacité de liaison du domaine aptamère, les différents fragments obtenus suite au clivage sont séparés sur un gel dénaturant de polyacrylamide. Le patron de dégradation obtenu pour l'ARN incubé avec son ligand potentiel est ensuite comparé à celui obtenu pour l'ARN incubé sans ligand (Figure 17).

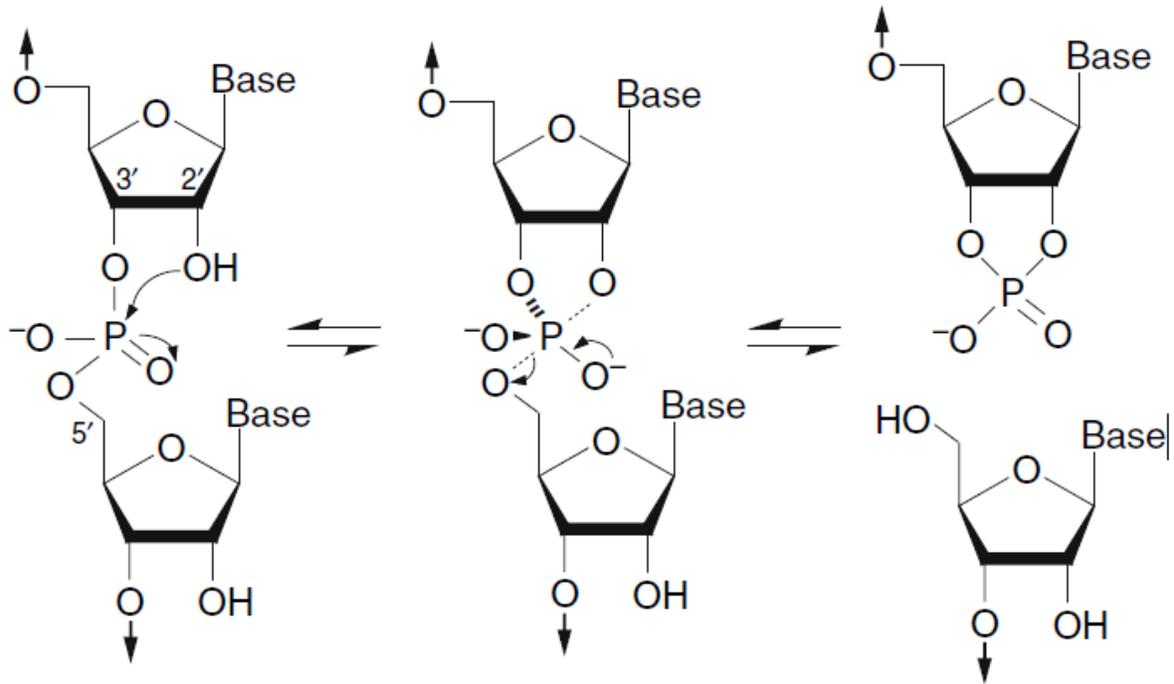


Figure 16. Schéma du clivage spontané des liens phosphodiesters de l'ARN.

Les flèches incurvées décrivent le mouvement des électrons partant de l'oxygène 2' (groupe nucléophile). Ceci permet une attaque nucléophile sur le phosphore (P) du centre impliqué dans le lien phosphodiester et libère ainsi le groupe attaché à l'oxygène 5' (lignes pointillées). Tiré de (Regulski & Breaker, 2008).

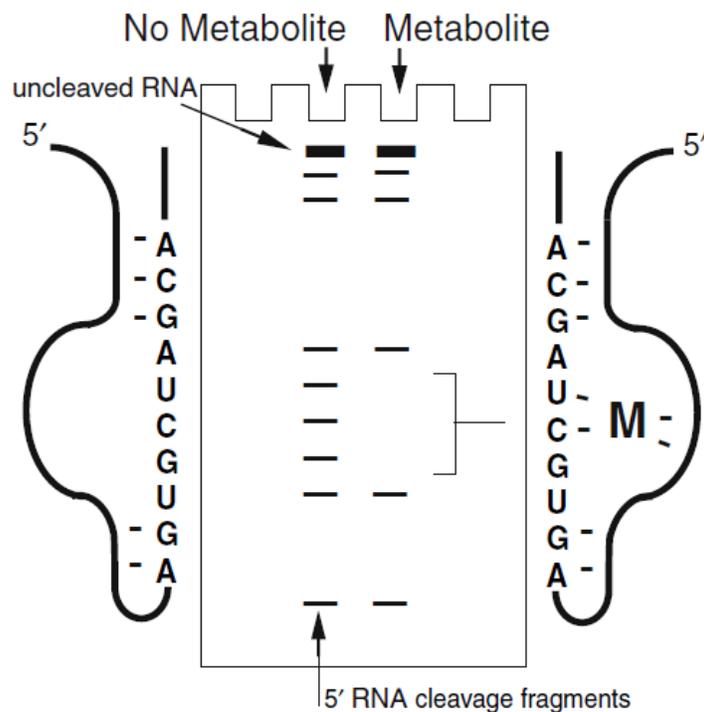


Figure 17. Schéma du gel d'*in-line*.

Les réactions incubées sans métabolite (*no metabolite*) et avec métabolite (*metabolite*) sont déposées dans un gel de polyacrylamide pour migration. Les différents fragments d'ARN 5' (*5' RNA cleavage fragments*) sont visibles grâce au marquage au ^{32}P . La différence entre les deux patrons de dégradations (mise en évidence par le crochet sur le schéma) résulte du fait que le métabolite (indiqué par la lettre M) a pu se lier à l'ARN. Ceci a contribué à changer sa structure secondaire et a réduit le nombre de clivages spontanés, l'alignement nécessaire n'ayant pas pu se faire. L'ARN non clivé (*uncleaved RNA*) est également visible sur le haut du gel. Tiré de (Regulski & Breaker, 2008).

Méthode

Pour les réactions dites d'*in-line*, un mélange réactionnel de 20 μL est préparé dans un tube en combinant 3 μL d'ARN radiomarqué ou chaud ; 1X du tampon *in-line* (50 mM Tris-HCl (pH 8,3 à 20°C); 20 mM MgCl_2 ; 200 mM KCl); la solution du métabolite à la concentration finale désirée (variant de 10 μM à 100 mM) et de l'eau milli-Q pour compléter le volume. Dans la méthode originale, les mélanges réactionnels sont, ensuite, incubés à température pièce pendant 40h à 48h. Cependant, la structure de la tige-boucle guanidine-II étant vraiment stable et n'étant naturellement pas prompt à adopter une conformation *in-line*, les mélanges réactionnels de la présente expérience sont plutôt incubés pour le même temps, mais à 37°C pour favoriser le clivage spontané des ARN.

Une fois le temps d'incubation presque complété, les trois réactions de contrôles suivantes doivent être préparées : non réagit (NR), digestion partielle à l'aide d'une RNase (T1) clivant après chaque G et la digestion alcaline (OH) clivant après chaque nucléotide. Ces réactions sont préparées dans un volume réactionnel de 20 μL . Pour commencer, la réaction nommée NR est un contrôle négatif comprenant le précurseur de l'ARN non digéré et aucun ligand. Cette réaction permet de s'assurer de l'intégrité de l'ARN et donc de vérifier qu'il n'a pas été dégradé par des ribonucléases. Pour ce faire, un mélange réactionnel comportant 3 μL d'ARN marqué au ^{32}P et complété à l'eau milli-Q est préparé. Ensuite, la réaction T1, quant à elle, comprend 6 μL d'ARN marqué, 25 mM de citrate de sodium (pH 5,0 à 23°C), 0,25 mU RNase T1 et 3 μL de formamide (Sigma-Aldrich) et de l'eau milli-Q pour compléter la réaction. Cette dernière réaction vise à digérer partiellement l'ARN via la RNase T1 qui ne clive l'ARN qu'après chaque nucléotide G. Finalement, la réaction nommée OH, comprend également 6 μL d'ARN marqué, ajouté de 10% final de Na_2CO_3 et de l'eau milli-Q pour compléter la réaction. Cette réaction alcaline, quant à elle, sera capable d'aller cliver l'ARN après chaque nucléotide. L'analyse de ces deux digestions (une fois migrée sur gel) permet d'établir la structure secondaire de l'ARN et de repérer à quel endroit le ligand a pu se lier.

Toutes ces réactions, soient les réactions d'*in-line* et de contrôle, sont ensuite déposées sur un gel de polyacrylamide de 20% à 60 W pendant 3h30. Le gel est ensuite démoulé et entouré de pellicule plastique. Une cassette d'autoradiographie (FisherBiotech) y est ensuite apposée pour une exposition de 12h à 4°C dans une boîte à Plexiglas. Une fois le temps d'exposition terminée, la cassette est scannée par le Typhoon™ FLA 9500 (GE healthcare Life Sciences) et analysé à l'aide des logiciels ImageQuant TL v8.1. Ici, ce logiciel permet de quantifier l'intensité des différentes bandes modulée par la variation en concentration du ligand. Finalement, le K_D est déterminé à l'aide du logiciel GraphPad Prism 8.0.2 en mettant en évidence les fractions liées selon la concentration du ligand (en échelle logarithmique).

5 RÉSULTATS

5.1 Recherche bio-informatique

5.1.1 RNAstem

La première partie de la recherche bio-informatique consistait en la recherche du motif à 1 tige de guanidine-II au sein d'une banque d'ARN déjà annotée, à savoir RNAstem. Cette recherche a permis de trouver plus de 819 représentations du motif en tige épingle parmi lesquelles 768 ribointerrupteurs guanidine-II potentiellement actifs ont été trouvés confirmant ainsi la validité des critères de recherche sélectionnés. Parmi ces candidats, 14 (Table 1) ont été sélectionnés pour la validation expérimentale selon les critères mentionnés en 4.1.3. L'un des candidats, nommé *glnA*, a été testé avec sa séquence complète et sa séquence tronquée, c'est-à-dire ne contenant que la boucle du motif et la tige complète de deux paires de bases. La liste brute des candidats trouvés par RNAstem se trouve en annexe électronique.

Table 1. Candidats RNAsstem sélectionnés pour la validation expérimentale

R	Nom Rfam	Boucle	Séquence candidat
R F	5S_rRNA	ACGA	GCCGAUGGUACUGGGCG GACGAC CGCUGGGAGAGUAG GU
R	RNaseP_bact_a	ACGA	GCGUUCUG GACGAC CGGAGCGGC
R	RRE	ACGA	CAGCAG GACGAC CUGCUA
R	7SK	ACGG	UCGAAGAG GACGGC CUCAG
R F	SSU_rRNA_bacteria	ACGA	GACAUCUGGCUUCAG GACGAC UGACUUCGGUUCAGUG GC
R	SSU_rRNA_bacteria	ACGA	GCCU GACGAC CGACCGAACGAAUAGUGGC
R F	SSU_rRNA_bacteria	ACGA	GUCUUUAGUUUACUAUAG GACGAC CGGUUAGUGGC
R F	Trp_leader	ACGA	GUGGCGCACCUCUG GACGAC GGGCGGCGUGAUCGC
R F	mir-592	ACGG	GAUGAUGUGUUGUG GACGGC ACAGCGUCAUCACGUGG UGACGCAACAUC AUGACGUAA
R F	<i>glnA</i> _complet	ACGA	AUCGUUCAUCCUAAACAGGACGGAAGUAAG GACGAC UC GGAACGAAGC
R	<i>glnA</i> _séquence_tronq	ACGA	AGACGACU
R F	<i>Pfl</i> (ZMP/ZTP)	ACGA	GAUUCGCGCAACUGGCAAGGUAAGAUG GACGAC CAU CGAGGAGCGUGGGAAA
R	SSU_rRNA_eukarya	ACGA	UUACCCAAUCCC GACGAC GGGGAGGUA
R	SSU_rRNA_eukarya	ACGA	CGACGAC GGCG

Le Rfam ID et le nom Rfam correspondent respectivement à l'identifiant Rfam et le nom de l'ARN tel que trouvé dans cette base de données. Les noms identifiés en rouge sont les candidats qui ont été testés par *in-line*. Les séquences de la boucle du motif ainsi que la séquence ARN du candidat y sont également indiquées. Les éléments de la séquence en gras correspondent au motif guanidine-II.

5.1.2 RNArobo

Cette deuxième partie de la recherche bio-informatique a permis d'étendre les résultats de la recherche aux séquences aléatoires également. De plus, la recherche avec RNArobo (v. 2.1.0) s'est faite au sein de plusieurs génomes, soient chez l'être humain, la bactérie, l'archaea, le virus et également dans les régions intergéniques (IGR) pour les motifs à 1 tige et à 2 tiges. Ainsi, avec cet outil informatique, il a été possible de trouver exactement 283 909 candidats correspondant au motif guanidine-II. Le nombre de candidats trouvés pour chaque génome ainsi que les IGR est indiqué dans la Table 2. À partir de ces candidats, environ une dizaine de candidats par génome et pour les IGR (Table S1, annexe) ont été sélectionnés pour passer l'étape de la validation expérimentale et trois d'entre eux ont été testés (Table 3). Une liste brute de tous les candidats RNArobo est disponible dans l'annexe électronique.

Table 2. Nombre de candidats totaux trouvés dans différents génomes et IGR

	1 tige	2 tiges
IGR procaryotiques	4 829	150
Humain	972	11
Bactérie	162 040	7 877
Archaea	30 457	2 565
Virus	1 307	33
Champignons	72 298	1 370

Les résultats correspondent au nombre de candidats trouvés pour les motifs à 1 tige ou à 2 tiges de génomes et régions intergéniques (IGR) procaryotiques.

Table 3. Candidats RNArобо testés expérimentalement

Nom	Boucle	Score
<i>Bacillus subtilis strain TLO3</i>	ACGG	24
<i>Streptomyces pactum strain ACT12</i>	ACGA	24
<i>Lactobacillus delbrueckii subsp. indicus strain JCM 15610</i>	ACGG	24

Candidats RNArобо qui ont été testés expérimentalement via la cartographie *in vitro* par *in-line*. Le score a été calculé tel que décrit dans la section 4.1.3

5.1.3 Gènes associés aux candidats RNArobo

Pour aller plus loin dans l'analyse des candidats obtenus, une recherche de gènes associés, soient les gènes se trouvant en aval des motifs associés et sur le même brin, a été effectuée pour les candidats obtenus dans le génome bactérien avec la recherche du motif à une tige. Plus précisément, cette analyse permet de vérifier s'il est possible de retrouver les mêmes gènes en aval des candidats trouvés par RNArobo que ceux qui sont connus dans la littérature pour être associés à la famille de ribo-interrupteur guanidine-II. Une liste de protéines encodées par les gènes associés à guanidine-II a donc été établie à l'aide de la base de données Ribogap 2 (Naghdi *et al.*, 2017), puis leur occurrence a été dénombrée. Ensuite, le nombre d'occurrences de ces mêmes protéines a été dénombré à chaque fois qu'elles étaient associées à un candidat trouvé par RNArobo. Le ratio en pourcentage des occurrences de RNArobo par rapport aux occurrences trouvées par Ribogap 2 (Naghdi *et al.*, 2017) a été calculé. Plus le ratio est élevé plus la recherche RNArobo a été capable de trouver un candidat dont le gène en aval est connu pour être associé au ribo-interrupteur guanidine-II. Les résultats ainsi obtenus sont présentés dans la Table 4. En moyenne, ce ratio équivaut à 3,57%.

Afin de formuler une conclusion sur ce ratio, une comparaison avec quatre différents contrôles négatifs choisis arbitrairement a été effectuée. Le premier contrôle négatif s'est fait à partir d'une liste de gènes générée aléatoirement. Ces gènes n'ont donc de prime abord aucun lien avec le ribo-interrupteur guanidine-II. Un ratio plus bas est donc attendu dans le cas de ce contrôle. Encore une fois, le nombre d'occurrences totales de chaque gène a été dénombré en utilisant la base de données Ribogap 2 (Naghdi *et al.*, 2017). Puis, le nombre d'occurrences de ces gènes pour les candidats de RNArobo a également été dénombré afin de pouvoir établir un ratio. En moyenne, le ratio est de 0,13% (Table 5).

Pour les trois autres contrôles négatifs, ceux-ci ont été choisis de manière arbitraire, à savoir les ribo-interrupteurs glycine, moco et l'ARN *rybB* ont été choisis. Pour chacun des ribo-interrupteurs et des ARN, une liste de gènes associés (ou des protéines pour lesquels ils encodent) a été déterminée à l'aide de la même base de données, puis leur nombre d'occurrences a été dénombré. Par la suite, le nombre d'occurrences de chacun de ces gènes ou protéines a été dénombré pour les candidats de RNArobo. Dans ces derniers cas également, ces gènes n'ont pas d'associations avec le ribo-interrupteur guanidine-II. Un faible ratio est donc attendu. En moyenne, les ratios obtenus étaient de 0,12% ; 0,63% et 0,42% pour les contrôles effectués avec les ribo-interrupteurs glycine, moco et l'ARN *rybB* respectivement (Table 6, Table 7 et Table 8).

Afin de déterminer si les ratios obtenus pour les contrôles étaient significativement différents de celui obtenu pour le ribo-interrupteur guanidine-II, le test statistique de Wilcoxon-Mann-Whitney (Mann & Whitney, 1947; Wilcoxin, 1947) a été effectué. Les résultats montrent qu'il y a une différence significative entre l'échantillon statistique guanidine-II et les contrôles gènes aléatoires, glycine et moco, mais pas avec le contrôle *rybB* (Figure 18). Ce dernier contrôle avait un ratio moyen de 0,42%, soit bien en deçà du 3,57% obtenu pour guanidine-II. Le fait que le test statistique n'ait pas révélé de différence significative peut être dû au fait que ce dernier contrôle ne comptait que 13 observations ce qui a pu affecter les résultats étant donné la petite taille de cet échantillon. Néanmoins, la tendance reste la même.

Toutefois, à la vue du ratio obtenu pour les candidats de RNArobo (3,57%), il semblerait que ces derniers soient de potentiels ribo-interrupteurs de la guanidine si on considère qu'ils soient associés aux mêmes gènes que ceux présentés dans la littérature pour le ribo-interrupteur guanidine-II. Les ratios faibles obtenus pour les contrôles négatifs, c'est-à-dire pour les gènes aléatoires ainsi que ceux associés aux ribo-interrupteurs glycine et moco viennent également confirmer ces résultats, alors que ces gènes et protéines ne sont pas connus pour avoir une association avec guanidine-II. Finalement, bien que le test statistique n'ait pas relevé de différence significative, probablement dû au nombre peu élevé d'observations, le ratio de 0,42% obtenu pour l'ARN *rybB* reste inférieur à celui de 3,57% obtenu pour guanidine-II. Ceci viendrait aussi confirmer la validité des candidats obtenus par RNArobo pour le génome bactérien avec la recherche du motif à une tige.

Table 4. Analyse des protéines associées aux candidats RNArobo

Liste exhaustive des produits de gènes associés	RNArobo	Ribogap	Ratio %
urea carboxylase	2	434	0,46
urea amidolyase	4	45	8,89
allophanate hydrolase	4	8	50,00
sugE protein	15	201	7,46
quaternary ammonium compound-resistance protein sugE	40	234	17,09
multidrug resistance protein	20	1075	1,86
SMR family multidrug efflux pump	6	118	5,08
multidrug transporter, SMR superfamily	9	1638	0,55
molecular chaperone	31	6697	0,46
molecular chaperone SugE	0	6	0,00
putative chaperone	1	420	0,24
Chaperone-like protein transmembrane protein	0	1	0,00
nucleotidyltransferase substrate-binding protein	0	41	0,00
two component LuxR family transcriptional regulator	1	1568	0,06
suppressor of groEL	4	9	44,44
calcium/proton antiporter	1	159	0,63
2-hydroxyhepta-2,4-diene-1,7-dioate isomerase	0	341	0,00
malonyl CoA-acyl carrier protein transacylase	4	1410	0,28
transporter	731	420491	0,17
phytase	1	346	0,29
short-chain dehydrogenase/reductase SDR	21	11944	0,18
biotin carboxylase	8	2795	0,29
DMT family permease	1	335	0,30
putative cytoplasmic protein	4	2018	0,20
cytochrome bd ubiquinol oxidase	3	796	0,38
oxidoreductase	279	96005	0,29
sulfatase	7	5732	0,12
aspartate-semialdehyde dehydrogenase	0	2734	0,00
peptide chain release factor 3	4	1494	0,27

potassium-transporting ATPase A	1	173	0,58
putative transposase	21	6332	0,33
beta-lactamase domain-containing protein	4	3205	0,12
F0F1 ATP synthase subunit A	1	1964	0,05
ribonucleoside-diphosphate reductase, adenosylcobalamin-dependent	0	129	0,00
succinate dehydrogenase flavoprotein	6	2153	0,28
fatty acid desaturase	3	1974	0,15
transcriptional regulatory protein EntR (Entericidin R)	0	6	0,00
transcriptional regulator	1356	298554	0,45
Response regulator containing a CheY-like receiver domain and an HTH DNA-binding domain	1	735	0,14
Membrane transporters of cations and cationic drugs	0	20	0,00
LuxR family transcriptional regulator	33	7216	0,46
urea short-chain amide or branched-chain amino acid	1	9	11,11
IS1 repressor protein InsA	0	254	0,00
Total	2628	881819	-
		moy	3,57%

Le nombre d'occurrences pour chaque produit de gènes (protéines) associés a été dénombré pour la construction du tableau de la manière décrite ci-contre. 1^{re} colonne : liste exhaustive des candidats RNArobo en amont de produits de gènes associés au ribo-interrupteur guanidine-II. 2^e colonne : nombre de candidats de RNArobo pour chaque produit de gènes indiqué (nombre RNArobo). 3^e colonne : nombre d'occurrences trouvées pour ces mêmes produits de gènes par la base de données Ribogap 2 (nombre Ribogap). 4^e colonne : ratio qui équivaut à : (nombre RNArobo/nombre Ribogap)x100. Le nombre total d'occurrences trouvées pour les candidats RNArobo ou Ribogap 2 sont indiquées au bas (total) tout comme la moyenne des ratios de la dernière colonne (moyenne).

Table 5. Contrôle gènes aléatoires

Gènes	RNArobo	Ribogap	Ratio (%)
infA	13	1752	0,74
InfA	13	1752	0,74
marC	1	160	0,63
sfsA	2	331	0,60
eutA	1	190	0,53
recQ	2	800	0,25
moaC	2	1082	0,18
cheY	1	888	0,11
flhA	1	956	0,10
recF	1	1330	0,08
hsrA	0	50	0,00
fesA	0	6	0,00
atoS	0	125	0,00
icd2	0	53	0,00
deoD	0	891	0,00
yqaB	0	101	0,00
cheY1	0	56	0,00
yfeC	0	108	0,00
rppH	0	80	0,00
yoaF	0	62	0,00
cho2	0	2	0,00
uxaC	0	248	0,00
ybhB1	0	2	0,00
nfrB	0	89	0,00
ybhB	0	1	0,00
hoIA	0	1066	0,00
rnb1	0	2	0,00
tusD	0	69	0,00
ydiU	0	58	0,00
yihY	0	76	0,00

Total	37	12386	-
		moy	0,13

Un contrôle a été effectué avec une liste de gènes générée aléatoirement et présentée dans la première colonne. Les annotations sont les mêmes qu'en Table 3, mais avec des gènes plutôt que le produit de ces derniers.

Table 6. Contrôle gènes associés au ribo-interrupteur glycine

Gènes	Rnarobo	Ribogap	Ratio (%)
proV	11	328	3,35
alaS	3	1918	0,16
lgt	3	1261	0,24
aroP	2	265	0,75
nhaA	1	631	0,16
xthA	1	513	0,19
narK	1	419	0,24
fbp	1	1248	0,08
odc	2	1070	0,19
dadA	0	439	0,00
glyA	0	1955	0,00
gcvT	0	1164	0,00
dagA	0	62	0,00
sdhB	0	1255	0,00
sdhB1	0	13	0,00
M1-554	0	2	0,00
Irp1	0	22	0,00
gcvP	0	789	0,00
attE	0	10	0,00
glyP	0	7	0,00
pcrA	0	551	0,00
gcvP1	0	64	0,00
rplM	0	1779	0,00
hemC	0	1355	0,00
alsT	0	123	0,00
dagA1	0	3	0,00

grdX	0	16	0,00
comGA	0	108	0,00
trxB1	0	130	0,00
bcr	0	503	0,00
raxST	0	7	0,00
lon2	0	39	0,00
dadA3	0	24	0,00
gcvB	0	24	0,00
yqhH	0	70	0,00
cycA	0	302	0,00
tyrS2	0	19	0,00
ydgF	0	55	0,00
gcvH	0	798	0,00
napF	0	171	0,00
ampC	0	341	0,00
dadA4	0	8	0,00
gcvH_1	0	6	0,00
Total	25	19867	-
		moy	0,12

Un contrôle a été effectué avec les gènes associés au ribo-interrupteur glycine et présentée dans la première colonne. Les annotations sont les mêmes qu'en Table 3, mais avec des gènes plutôt que le produit de ces derniers. Les gènes associés aux ribo-interrupteurs glycine ont été déterminés à l'aide de la base de données Ribogap 2.

Table 7. Contrôle gènes associés au ribo-interrupteur moco

Gènes	Rnarobo	Ribogap	Ratio (%)
ybhK	7	75	9,33
narQ	2	107	1,87
argF	2	834	0,24
AtoC	1	154	0,65
moeA	2	958	0,21
modA	2	723	0,28
moaA	1	1158	0,09
aorA	0	10	0,00
TupB	0	25	0,00
THI80	0	13	0,00
torZ	0	55	0,00
ytfE	0	81	0,00
torY	0	53	0,00
fdhA	0	252	0,00
purR	0	524	0,00
HyaA	0	98	0,00
SpIB	0	114	0,00
dinG	0	594	0,00
moaE	0	741	0,00
etfQ	0	2	0,00
Total	17	6571	-
		moy	0,63

Un contrôle a été effectué avec les gènes associés au ribo-interrupteur moco et présentés dans la 1^{ère} colonne. Les annotations sont les mêmes qu'en Table 3, mais avec des gènes plutôt que le produit de ces derniers. Les gènes associés aux ribo-interrupteurs moco ont été déterminés à l'aide de la base de données Ribogap 2.

Table 8. Contrôle gènes associés à l'ARN *rybB*

Protéines associées	Rnarobo	Ribogap	Ratio (%)
transcriptional regulator, ArsR	24	722	3,32
permease YbjG	0	1	0,00
DeoR family transcriptional	38	3468	1,10
phosphoesterase PA-	0	563	0,00
saccharopine dehydrogenase	4	1015	0,39
undecaprenyl pyrophosphate	2	997	0,20
membrane protein	318	122407	0,26
phosphatase	116	68045	0,17
inner membrane-associated	0	1	0,00
trans-2-enoyl-CoA reductase	0	293	0,00
serine transport protein	0	2	0,00
deoxyribose-phosphate	0	2004	0,00
DNA-binding transcriptional	0	2143	0,00
Total	502	201661	-
		moy	0,42

Un contrôle a été effectué avec les produits de gènes (protéines) associés à l'ARN *rybB* et présentés dans la 1^{ère} colonne. Les annotations sont les mêmes qu'en Table 3. Les produits de gènes associés à l'ARN *rybB* ont été déterminés à l'aide de la base de données Ribogap 2.

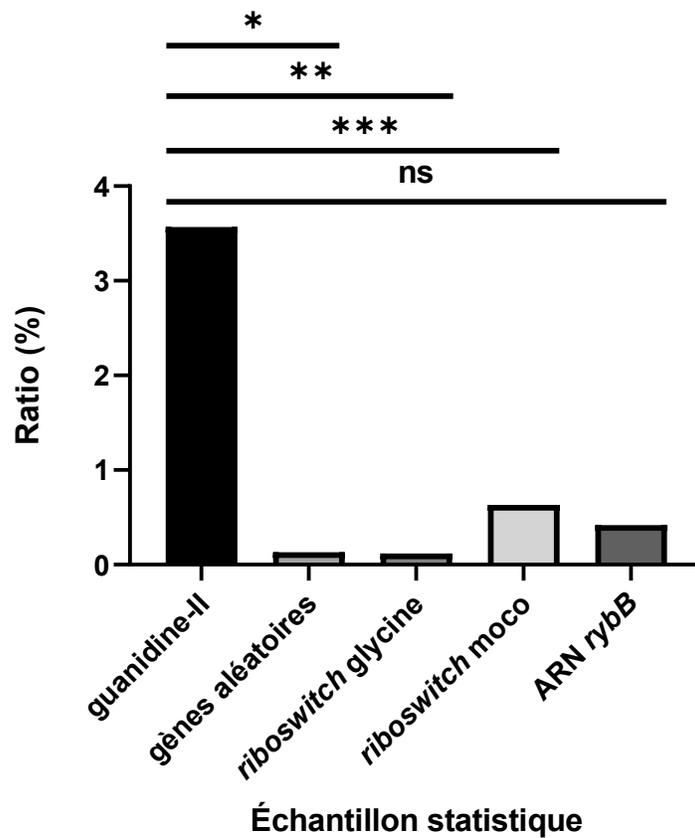


Figure 18. Comparaison des échantillons statistique.

Comparaison des échantillons statistiques entre guanidine-II et les quatre contrôles (gènes aléatoires, ribo-interrupteurs glycine et moco ainsi que l'ARN rybB) a été effectuée à l'aide du test statistique Wilcoxon-Mann-Whitney. Les différents p-values obtenus étaient de 0,0002582 ; 5,651e-08 ; 0,003669 et 0,05806 pour les comparaisons de guanidine-II avec les gènes aléatoires, le ribo-interrupteur glycine, le ribo-interrupteur moco et l'ARN rybB respectivement. Ces p-values ont permis d'établir qu'il y avait une différence significative avec les trois premiers contrôles (montrée par des *) (p-value < 0,05), mais une différence non significative (ns) pour le dernier contrôle rybB (p-value > 0,05).

5.2 Cartographique *in vitro* par *in-line*

Parmi les candidats sélectionnés de RNAs_{stem} et RNAr_{obo}, plus d'une quinzaine ont pu être évalués via la cartographie *in vitro* par *in-line* (identifiés dans l'annexe électronique).

Parmi eux, deux ont donné des résultats indiquant une interaction avec la guanidine, soient le ribo-interrupteur complet glutamine (*glnA*) ainsi que la ribonucléase P (Rnase P). Leurs résultats sont présentés plus en profondeur dans les sections 5.2.1 et 5.2.2 ci-dessous.

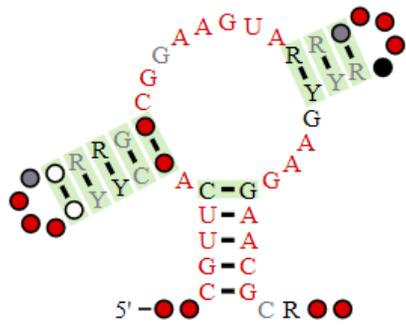
5.2.1 Ribo-interrupteur glutamine

Le ribo-interrupteur glutamine (*glnA*) (Figure 19 A) est un des candidats qui a été identifiée par la banque de données RNAs_{stem}. En effet, bien que le ligand de ce ribo-interrupteur ait été identifié comme étant la glutamine (Figure 19 B), selon notre hypothèse, le motif guanidine-II qu'il comporte laisse croire qu'il serait aussi capable de lier la guanidine et pourrait potentiellement agir comme ribo-interrupteur guanidine. La cartographie a donc été utilisée afin d'évaluer, en premier lieu, si cet ARN serait capable de lier la guanidine également, alors que ces deux molécules sont impliquées dans le métabolisme de l'azote.

Le gel d'*in-line* effectué en absence de glutamine (Figure 20 A) semble montrer quelques modulations de l'ARN clivé au niveau de la boucle ACGR (encadré) et sous la boucle (flèche) suggérant ainsi la liaison de la guanidine avec le ribo-interrupteur *glnA*. En effet, on constate une augmentation croissante de l'intensité des bandes avec une augmentation croissante du ligand. En d'autres mots, la liaison du ligand à l'ARN aurait ainsi perturbé la structure de l'ARN ce qui permet d'observer cet effet. Après quantification et analyse du gel, un K_D a pu être calculé pour ce candidat qui montre une affinité d'environ 5 mM en concentrations traces et une affinité légèrement plus élevée avec un K_D de 4 mM à une concentration de 2 μ M (Figure 20 C).

Le ribo-interrupteur *glnA* possédant déjà un ligand spécifique, à savoir la glutamine, il était donc également important de déterminer si la guanidine pourrait avoir un effet sur la liaison du ribo-interrupteur à son ligand et donc, indirectement, un effet sur la fonction de ce ribo-interrupteur. Pour ce faire, un autre gel *in-line* a été effectué, mais cette fois-ci en présence de glutamine à une concentration de 20 mM (Figure 20 B). Cette dernière concentration correspond, en fait, à quatre fois le K_D du ribo-interrupteur ce qui permet de s'assurer que le ligand glutamine ne soit pas un facteur limitant. Après quantification de ce dernier gel d'*in-line* (Figure 20 D), on constate que le ribo-interrupteur *glnA* perd de l'affinité pour son ligand naturel. À la vue de ces résultats, il semblerait donc que la guanidine rendrait le ribo-interrupteur *glnA* non fonctionnel.

A)



B)

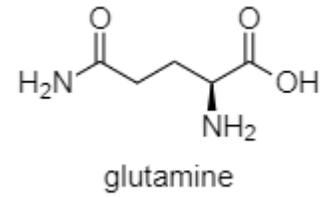


Figure 19. Structures du ribo-interrupteur *glnA* et de son ligand.

A) Structure secondaire du ribo-interrupteur *glnA* telle que trouvée par le logiciel R-scape (Rivas *et al.*, 2017) (tiré de Rfam 14.4). B) Structure de l'acide aminé glutamine.

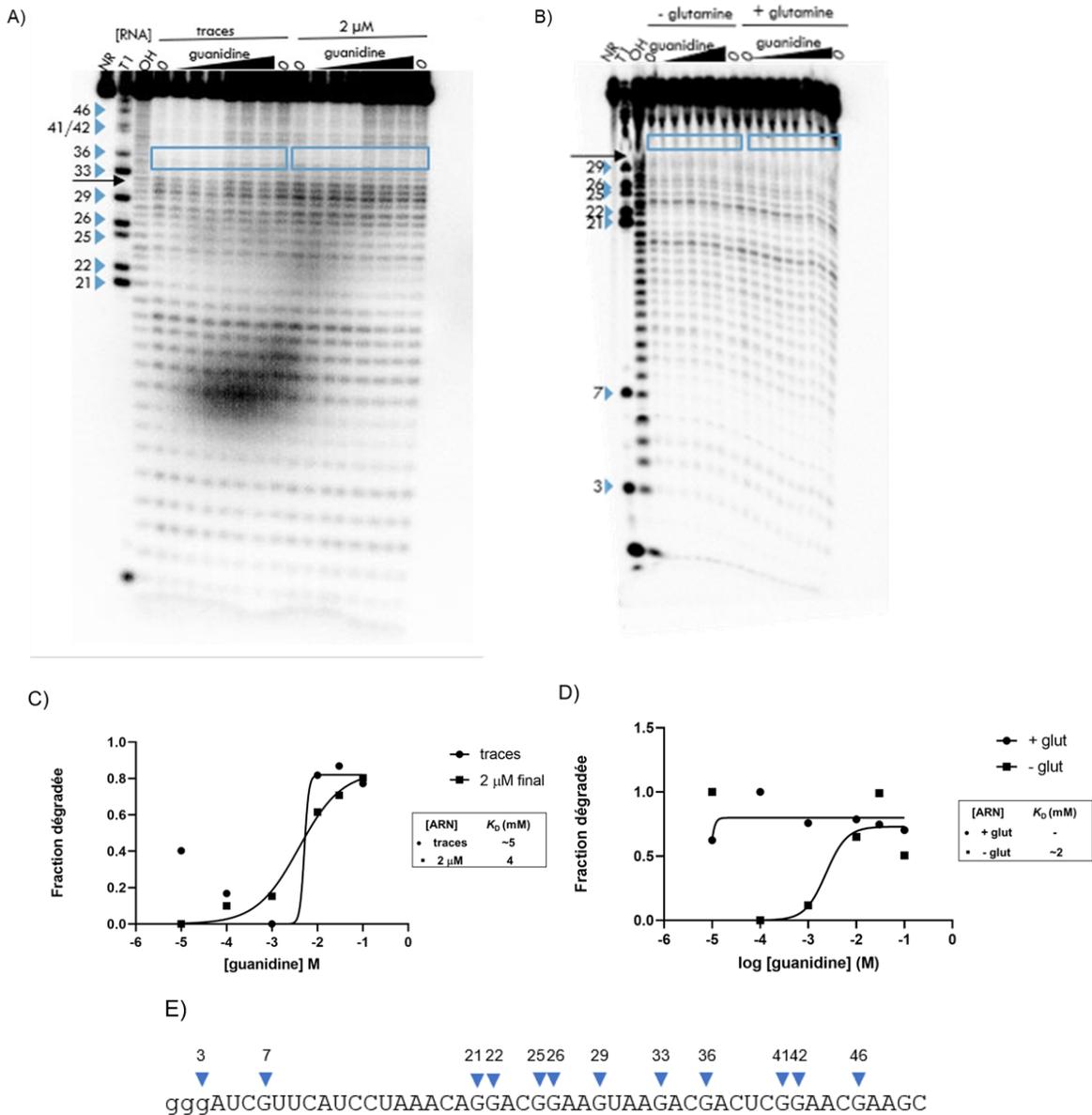


Figure 20. Liaison *in vitro* de la guanidine par le ribo-interrupteur *glnA*.

A) Analyse du gel de polyacrylamide résultant de la réaction de cartographie *in vitro* par *in-line* du candidat ribo-interrupteur *glnA*. Ce dernier a été testé en concentrations traces ([RNA]) et à 2 μM final en absence (0) et en présence de guanidine d'une concentration variant de 10 μM à 100 mM. NR, T1 et OH correspondent respectivement à non réagit, digestion partielle à la ribonucléase T1 (clive après chaque G) et digestion alcaline (clive après chaque nucléotide). Les positions des résidus G sont indiquées à l'aide des flèches bleues. L'encadré correspond à la boucle ACGA du ribo-interrupteur *glnA* et la flèche noire indique la bande qui a été quantifiée. **B)** Analyse du gel de polyacrylamide résultant de la réaction de cartographie *in vitro* par *in-line* du candidat ribo-interrupteur *glnA* en présence de son ligand spécifique glutamine. Le candidat ribo-interrupteur *glnA* a été testé en concentrations traces dans les mêmes conditions décrites en A) en présence de glutamine (+glut, 20 mM) et en absence (- glut). Les annotations sont les mêmes qu'en A). **C)** et **D)** Courbes de la fraction d'ARN lié selon le logarithme de la concentration de guanidine (M). Les courbes ont été tracées selon les données quantifiées des gels obtenus en A) et B) respectivement. **E)** Séquence primaire du ribo-interrupteur *glnA*. Les flèches en bleus indiquent les positions des G clivés en B). Les 3 « g » en 5' de la séquence correspondent aux nucléotides G ajoutés à la fin du promoteur T7 dans le but de faciliter la transcription *in vitro* comme mentionné dans la section 4.2.

5.2.2 Ribonucléase P

La ribonucléase P (RNase P) de l'organisme *Salinibacter ruber*, une bactérie extrême halophile, est un des candidats qui a été identifié par la banque de données RNASTEM. Ce ribozyme, retrouvé dans les 3 domaines de la vie, est essentiel pour la maturation des ARNt. Plus précisément, la RNase P est composée d'une sous-unité d'ARN qui lui confère son pouvoir catalytique ainsi que d'une ou plusieurs sous-unités protéiques (Ellis & Brown, 2009). C'est donc la partie ARN qui est responsable du clivage des ARN précurseurs (pré-ARNt) du côté 5' libérant ainsi l'extension 5' aussi appelé 5' *leader*. Les sous-unités protéiques, quant à elles, sont des facteurs essentiels pour l'activité *in vivo* de cette endonucléase, mais pas *in vitro* (Randau *et al.*, 2008). En effet, l'activité catalytique de l'ARN est capable de se maintenir *in vitro* sans l'aide de cofacteurs (Guerrier-Takada *et al.*, 1983).

La fonction de ce candidat est donc déjà bien caractérisée. Toutefois, selon notre hypothèse, le motif guanidine-II présent dans cet ARN suggère qu'il serait capable de lier la guanidine et pourrait potentiellement agir comme ribo-interrupteur également. La même méthode citée ci-haut a donc été utilisée afin de valider ou d'infirmer notre hypothèse.

Les gels d'*in-line* (Figure 21 A et B) montrent quelques modulations de l'ARN clivé au niveau de la boucle ACGR (encadré). En effet, dans ce cas-ci également, on constate une diminution croissante de l'intensité des bandes avec une augmentation croissante de la guanidine. Ceci suggère donc la liaison de la guanidine avec le candidat RNase P selon un K_D d'environ ~ 33 mM pour les concentrations traces et < 10 μ M pour la concentration à 2 μ M finaux (Figure 21 C et D respectivement).

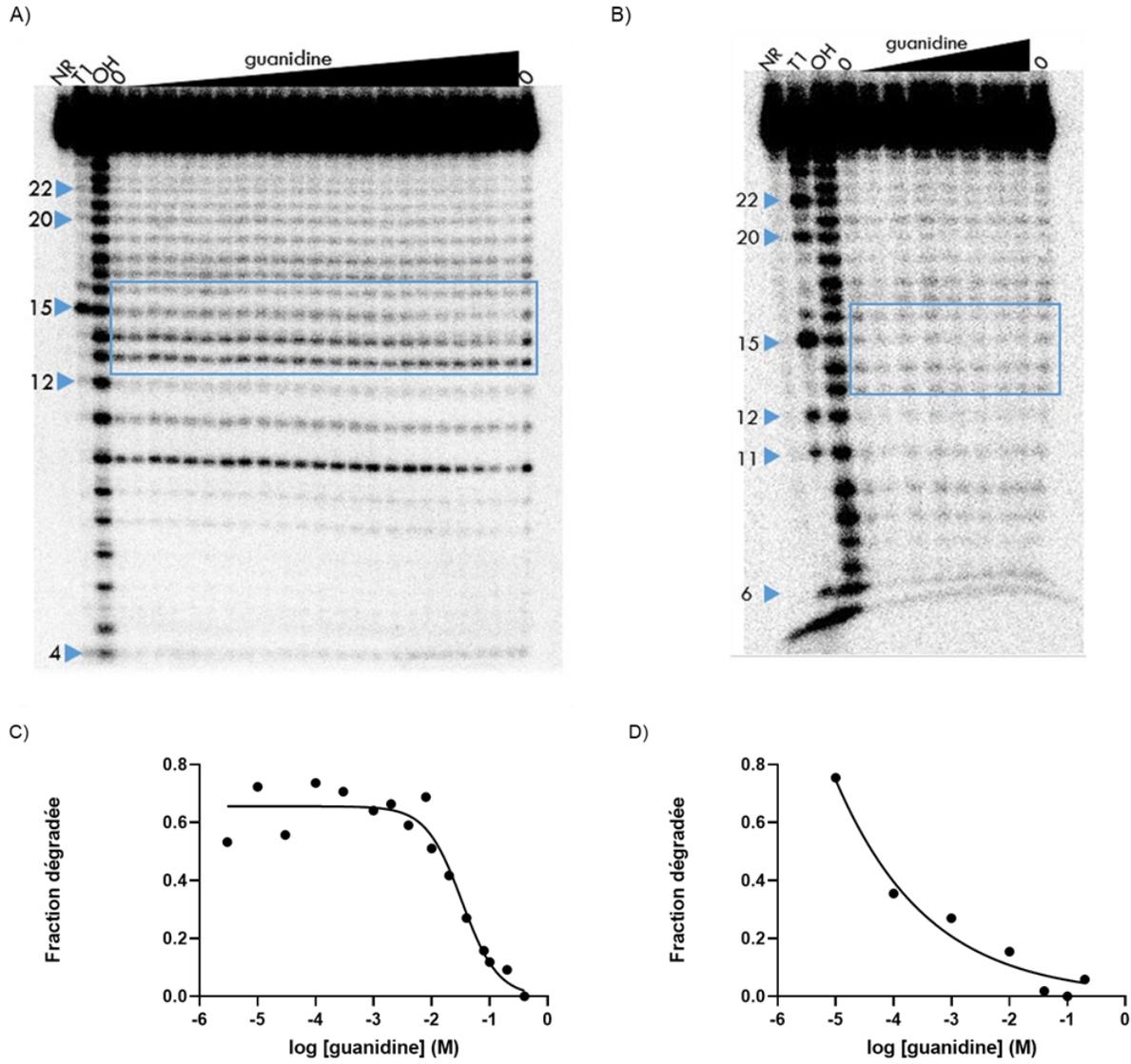


Figure 21. Liaison in vitro de la guanidine par le candidat RNase P.

A) Gel de polyacrylamide (PAGE) résultant de la réaction de cartographie in vitro par in-line du candidat Rnase P. Ce dernier ([RNA]) a été testé en concentrations traces en absence (0) et en présence de guanidine d'une concentration variant de 10 μ M à 400 mM. NR, T1 et OH correspondent respectivement à non réagit, digestion partielle à la ribonucléase T1 (clive après chaque G) et digestion alcaline (clive après chaque nucléotide). Les positions des résidus G sont indiquées à l'aide des flèches bleues. L'encadré correspond à la boucle ACGA de la RNase P. B) Gel PAGE résultant de la réaction de cartographie in vitro par in-line du candidat Rnase P. Ce dernier ([RNA]) a été testé à une concentration finale de 2 μ M final en absence (0) et en présence de guanidine d'une concentration variant de 10 μ M à 200 mM. Les annotations sont les mêmes qu'en A). C) et D) Courbes de la fraction d'ARN lié selon le log de la concentration de guanidine (M) selon les données quantifiées des gels obtenus en A) et B) respectivement.

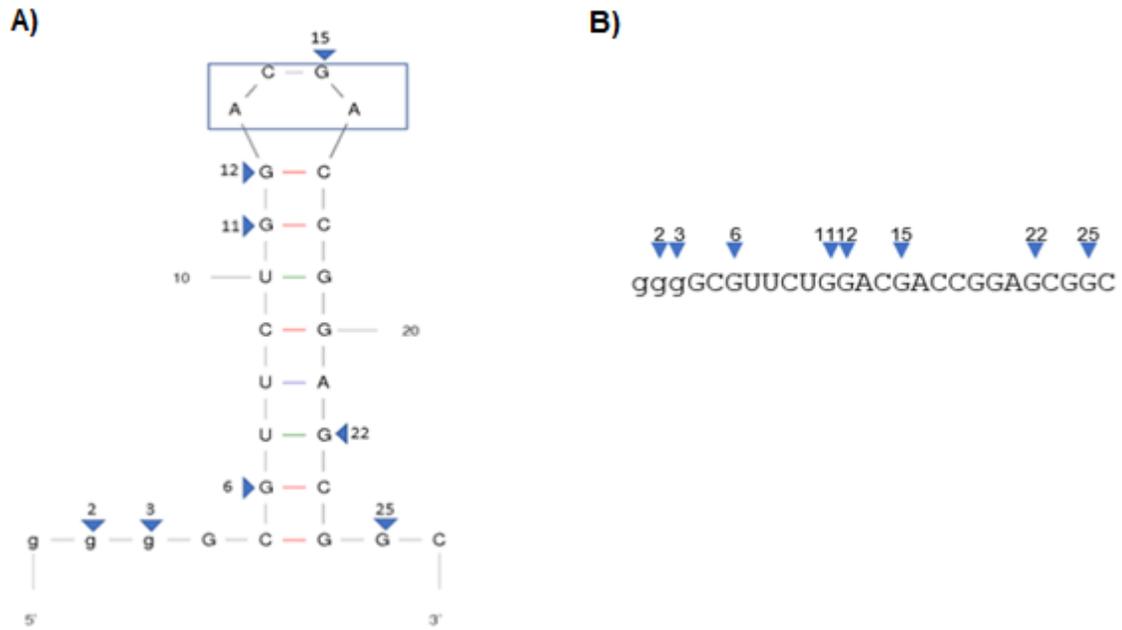


Figure 22. Structures secondaire et primaire du candidat RNase P.

A) Structure secondaire du candidat RNase P tel que prédit par Mfold (Zuker, 2003). Les flèches en bleu mettent en évidence quelques positions des nucléotides G pour une lecture plus facile. **B)** Séquence du candidat RNase P tel qu'identifié par RNAstem. Les flèches en bleu sont les mêmes qu'en A).

5.3 Évaluation *in vivo* de la fonction interruptrice du candidat RNase P

Afin de valider les résultats *in vitro* obtenus, des essais *in vivo* sont prévus chez la bactérie *Salinibacter ruber* où le candidat a été trouvé. À ce jour, les résultats ne sont pas encore disponibles et ne seront donc pas présentés dans la présente recherche.

6 DISCUSSION

6.1 Gènes associés aux candidats bactériens

Plusieurs des gènes associés aux candidats trouvés par RNArobo correspondent à ceux identifiés dans la littérature pour les ribo-interrupteurs guanidine. On retrouve notamment des gènes qui encodent pour des allophanates hydrolases, des urée (guanidine) carboxylases et pour plusieurs transporteurs faisant partie de la famille des SMR telle que SugE. En calculant la moyenne des ratios obtenus pour ces gènes trouvés, on retrouve également un ratio largement supérieur à ceux obtenus pour les différents contrôles négatifs (Table 4 à Table 8). Comme les contrôles négatifs sont des gènes ou des protéines pour lesquels aucune association n'est connue avec les ribo-interrupteurs guanidine, il était attendu que ces ratios soient plus faibles. De prime abord, les résultats sont donc intéressants et semblent indiquer que les candidats bactériens trouvés soient de possibles ribo-interrupteurs guanidine. Toutefois, cette recherche effectuée par l'entremise de ribogap 2 comporte un biais important qu'il n'est possible d'enlever selon les paramètres actuels de la base de données. En effet, il n'est pas possible d'extraire des résultats obtenus, les ribo-interrupteurs guanidine-II qui ont déjà été identifiés. Ce faisant, le ratio supérieur aux contrôles négatifs observé dans la Table 4 a pu largement être influencé par ce nombre. À la lumière de ces observations, il est donc impossible d'arriver à une conclusion ferme vis-à-vis des chiffres obtenus. Toutefois, bien que nous ignorions la véritable proportion que ces gènes associés occupent, les candidats bactériens gardent leur potentiel d'être des ribo-interrupteurs guanidine si on considère le fait qu'ils aient le même motif que guanidine-II et qu'un certain nombre d'entre eux semblent, effectivement, être associés aux mêmes gènes que ce dernier.

6.2 Évaluation du candidat *glnA*

La banque de données RNASTem a permis d'identifier le motif guanidine-II présent dans le ribo-interrupteur *glnA* de *Synechococcus phage S-PM2*. Ce candidat a donc été soumis à une cartographie *in vitro* par *in-line* afin de, tout d'abord, déterminer s'il était capable de lier la guanidine. Le premier gel effectué (Figure 20 A) semble confirmer notre hypothèse selon laquelle ce candidat serait bel et bien capable de lier la guanidine via le motif guanidine-II qu'il contient. Nous avançons cette conclusion avec précaution, car ce gel semble contenir des artéfacts marqués par les différentes intensités des puits 0 ne contenant aucun ligand. Néanmoins, il est à noter que le 2^e gel d'*in-line* (Figure 20 B) vient appuyer notre analyse puisqu'il est possible d'observer une affinité pour la guanidine en absence de glutamine dans ce cas-ci également.

Dans les expériences futures, ces gels seraient, toutefois, à refaire pour s'assurer de la validité de notre analyse. À ce stade de nos expériences, il serait, toutefois, possible de croire que le mode de liaison de la guanidine au ribo-interrupteur *glnA* se ferait de manière similaire à celui de guanidine-II, soit via des ponts hydrogènes. Par ailleurs, la charge positive de ce ligand pourrait facilement interagir avec la charge négative du squelette de *glnA* résultant en une stabilisation ionique tout comme cela a été observé pour guanidine-II. La dernière moitié du gel d'*in-line* présente à la figure 20 B, quant à elle, semble montrer une perte d'affinité à la guanidine en présence de glutamine.

Considérant ceci, pour les expériences futures, en plus de refaire ces expériences et afin de pousser plus loin les analyses, il serait intéressant, d'une part, de continuer les expériences *in vitro* avec la cartographie sous différentes conditions afin d'évaluer les différentes possibilités du ribo-interrupteur. Premièrement, un ribo-interrupteur qui serait inactivé par la guanidine, et donc rendu incapable de détecter la glutamine. Deuxièmement, la possibilité inverse, soit le ribo-interrupteur qui serait inactivé par la glutamine, et donc rendu incapable de détecter la guanidine. Troisièmement, le ribo-interrupteur serait plutôt fonctionnel en présence de guanidine ce qui lui permettrait de se replier correctement et donc apte à lier la glutamine. Dernièrement, le cas inverse, soit la glutamine permettrait au ribo-interrupteur de se replier correctement et, dans ce cas, le rendrait apte à lier la guanidine. D'une autre part, il serait également intéressant de poursuivre les études *in vivo* afin de déterminer si la guanidine affecterait le phage, plus précisément ses mécanismes d'infection et de réplication. Pour cela, il serait intéressant, de faire croître le phage en présence de guanidine afin d'observer si sa croissance en serait affectée ou d'observer quelques phénotypes et, en parallèle, de mesurer l'activité du ribo-interrupteur via un gène promoteur de type *lacZ*, par exemple.

6.3 Évaluation du candidat RNase P

Un autre candidat identifié par la banque RNAs_{tem} a également été testé. Il s'agissait cette fois-ci d'une ribonucléase bactérienne de chez *Salinibacter ruber*. En effet, selon la recherche bio-informatique effectuée, la RNase P comporte également un motif guanidine-II qui serait, selon notre théorie, un potentiel ribo-interrupteur de la guanidine, ou en ferait plutôt une version allostérique de la RNase P. Pour valider notre hypothèse, nous avons donc à nouveau effectué une cartographie *in vitro* par *in-line*.

Dans ce cas-ci, les gels d'*in-line* montrent également la liaison de la guanidine par la RNase P avec un K_D de ~33 mM pour les concentrations d'ARN traces et inférieur à 10 μ M pour une

concentration d'ARN à 2 μM finaux. Cette dernière concentration a été testée afin de simuler la présence de deux boucles, soit le motif guanidine-II complet. Pour cette dernière concentration, il était donc attendu que l'affinité à la guanidine soit meilleure et c'est ce qui est observé. En effet, l'affinité observée est 1 000 000 fois plus élevée dans le dernier cas. Cet exercice a également été fait dans la littérature avec un motif à 1 tige trouvée chez *Gloeobacter violaceus* (Sherlock *et al.*, 2017). Les résultats ont montré un K_D de 1 mM à des concentrations traces d'ARN et de 0,2 mM à une concentration finale d'ARN de 2 μM . L'affinité calculée dans la présente expérience avec le candidat RNase P est donc moins bonne à ce qui est retrouvé dans la littérature pour des concentrations traces (33 mM > 1 mM), mais amélioré pour une concentration finale de 2 μM (10 μM < 0,2 mM). La modulation du gel effectué à une concentration d'ARN à 2 μM est moins observable que celle présente pour le gel à concentrations traces. Ce gel à 2 μM effectué avec moins de variations de concentration de la guanidine comporte donc quelques incertitudes, alors qu'un K_D n'a pu être calculé. Néanmoins, les fractions dégradées calculées pour les concentrations zéro se situant proches de la valeur une (non présentées) montrent qu'en élargissant la gamme inférieure de concentrations de guanidine testés, il aurait été possible d'observer un plateau et ainsi avoir une courbe typique de liaison à un ligand.

Pour les expériences futures, nous suggérons de faire à nouveau ce gel afin de confirmer les analyses. Par ailleurs, il serait également intéressant de tester d'autres composés tels que des analogues de la guanidine afin de caractériser la spécificité de la RNase P pour ce ligand. Il est, en effet, possible que tout comme guanidine-II, la RNase P soit moins spécifique pour la guanidine comparativement aux autres familles et qu'en plus de lier des groupements aminés, elles soient capables de lier des analogues également.

Par ailleurs, il serait également intéressant de continuer l'analyse et d'évaluer si la guanidine serait capable d'affecter positivement ou négativement la fonction de la RNase P de manière allostérique. En effet, étant donné la fonction de maturation des pré-ARNt de cette endonucléase, il est possible de croire que ce ribozyme pourrait fonctionner de manière allostérique via la liaison à la guanidine et que, par exemple, la liaison à ce ligand favoriserait la maturation des ARNt en 5'. De plus, il est également possible de croire que le motif guanidine-II présent dans la RNase P pourrait interagir avec une autre boucle interne présente chez cette dernière favorisant ainsi la liaison de la guanidine tel que cela est observé chez le ribo-interrupteur guanidine-II qui possède deux pochettes de liaison. Afin de tester ces hypothèses, des expériences complémentaires de cartographie *in vitro* par *in-line* comprenant, cette fois-ci, la séquence complète de la RNase P de *Salinibacter ruber* seraient nécessaires. Il serait également intéressant de faire des tests *in vivo*

en faisant croître la bactérie en présence de guanidine dans la possibilité d'y observer des phénotypes différents, par exemple. Il serait également pertinent d'augmenter les concentrations de sels lors des expériences *in vitro* afin que celles-ci reflètent les conditions *in vivo* du microorganisme qui peuvent s'élever jusqu'à 15% {Antón, 2002 #112}.

Peu de choses sont connues sur la biologie de la guanidine si ce n'est qu'elle est produite de manière naturelle par les bactéries, probablement en situation de carence nutritionnelle, et qu'elle semble être toxique pour ces dernières au-delà du μM (Nelson *et al.*, 2017). Jusqu'à présent, les familles de ribo-interrupteurs guanidine semblaient participer à la détoxification de la cellule bactérienne en activant les gènes responsables tels que des transporteurs SMR. Une récente étude vient toutefois remettre en question la détoxification comme rôle unique de ces familles de ribo-interrupteurs. En effet, les résultats de cette recherche suggèrent que des entérobactéries seraient capables d'importer la guanidine via des transporteurs de type ABC dans le but de l'utiliser comme source d'azote (Sinn *et al.*, 2021). Cette récente découverte jette une nouvelle lumière sur les voies métaboliques entourant la production et l'importation de guanidine qui restent, toutefois, encore bien méconnues. Les résultats à 2 μM finaux pour le candidat RNase P candidat s'avère à posséder une fonction interruptrice déclenchée par la RNase P, il semble donc peu probable qu'à cette concentration, cette fonction participe à la détoxification de la cellule même en accélérant la maturation des ARNt. D'un autre côté, cela pourrait être une indication que des halophiles tels que *S. ruber* soient plus sensibles à ce ligand et ce, peut-être dans le but d'avoir un avantage par rapport aux autres microorganismes de son environnement. En effet, il est montré que *S.ruber* compose 5-25% de la communauté procaryotique où il est trouvé, alors que c'est une tendance plutôt observée chez les archaea (Antón *et al.*, 2002).

6.4 Importance du motif et potentiel des motifs à deux tiges

La simplicité de la pochette de liaison de guanidine-II permet d'avoir un aperçu sur la manière dont l'évolution permettrait l'émergence de telles séquences fonctionnelles. En effet, parmi les 256 combinaisons possibles de tétraboucles (4^4), le motif ACGR constitue 2 combinaisons (0,78% des possibilités). Or, sur les 6,67 millions de tétraboucles répertoriées dans RNAs_{stem}, seules 819 boucles ACGR sont trouvées ce qui inclut également la famille guanidine-II. Ceci constitue une proportion de 0,01% bien en deçà du 0,78% théorique. Cette sous-représentation montrerait donc l'importance de l'évolution dans le raffinement de cette boucle, et plus précisément sur sa fonction. En effet, il est possible de croire que le motif ACGR aurait pu apparaître de manière aléatoire au sein de diverses séquences et n'aurait été conservé que devant des gènes pour lequel sa fonction aurait un intérêt. Ce genre de motif est intéressant pour une telle étude. En effet, il serait possible que les nombreuses boucles ACGR trouvées dans le cadre de ce projet, mais pour lesquelles aucune capacité de liaison à la guanidine n'a été montrée, n'auraient tout simplement pas encore atteint leur fonction où seront amenées à disparaître au cours du temps si leur présence à cet endroit n'est pas pertinente. Cela est d'autant plus possible, en considérant que cette tetraboucle semble généralement défavorisée dans les structures d'ARN dont la fonction n'est pas reliée à la guanidine.

De plus, nous savons également que bien que chaque boucle ACGR est théoriquement capable de lier la guanidine, l'affinité du ribo-interrupteur guanidine-II est augmentée par la présence de deux de ces boucles. Ainsi, malgré les nombreuses tiges simples qui n'ont pas montré de capacité de liaison au cours de ce projet, les candidats à deux tiges, pour leur part, ont un plus grand potentiel de confirmer notre hypothèse de départ. En effet, ils possèdent le motif complet de guanidine-II et seraient potentiellement rendus plus loin dans le processus évolutif. Il serait donc intéressant de poursuivre la validation expérimentale de ces candidats.

7 CONCLUSION ET PERSPECTIVES

L'objectif principal de cette recherche était de découvrir de nouveaux ribo-interrupteurs guanidine au travers de séquences ARN connus et inconnus et possédant le motif guanidine-II. Les résultats bio-informatiques ont permis d'identifier un peu plus de 270 000 candidats correspondant aux motifs recherchés (1 tige et 2 tiges) disséminés au sein de plusieurs génomes (bactéries, archaea, virus, champignons, humains) et régions intergéniques. De ces candidats, environ 80 ont été sélectionnés pour évaluation via la cartographie *in vitro* par *in-line* et plus d'une quinzaine ont pu être testés au cours de cette recherche dont le ribo-interrupteur *glnA* de *Synechococcus phage S-PM2* et la RNase P de la bactérie *S. ruber*.

Pour ce qui est du ribo-interrupteur *glnA*, les tests d'*in-line* ont montré que la guanidine est capable de se lier à cet ARN en absence de son ligand naturel, soit la glutamine. Une fois cette dernière ajoutée, en présence de guanidine, la quantification du gel montre cette fois-ci que ce ribo-interrupteur perdrait son affinité pour la glutamine. En d'autres mots, il semblerait que la guanidine arrive à rendre non fonctionnel ce ribo-interrupteur glutamine. Il serait intéressant de poursuivre des tests *in vivo* afin de pouvoir observer si la même tendance se poursuit dans ces conditions. Lors de l'infection de la bactérie, *Synechococcus* prend contrôle de la machinerie de l'hôte pour s'approvisionner en azote extracellulaire (Waldbauer *et al.*, 2019). En effet, une étude d'infection virale effectuée sur *Synechococcus* WH8102 montre que la cyanobactérie continue à synthétiser des protéines et ribozymes malgré que sa machinerie soit sous le contrôle du phage (Waldbauer *et al.*, 2019). Ainsi, durant l'infection par le cyanophage S-SM1, il semblerait que la bactérie produise des éléments qui seraient utiles à sa défense tels que le facteur de terminaison Rho, la ribonucléase J, mais également de nombreuses protéines qui restent à caractériser. À ce jour, alors que les différents facteurs entourant le métabolisme de la guanidine restent encore inconnus, il serait intéressant de poursuivre la recherche en ce sens afin de déterminer si ces protéines ne joueraient pas un rôle de défense via la production indirecte de guanidine. En fait, il serait possible de croire que l'inactivation du ribo-interrupteur *glnA* serait, en fait, un moyen pour *Synechococcus phage S-PM2* de s'adapter à ce système de défense de surproduction de la guanidine par l'hôte.

Pour ce qui est de la RNase P, cet ARN dont la fonction a déjà été caractérisée, présente un motif guanidine-II qui montre *in vitro* une capacité à lier la guanidine. Son affinité (K_D 33 mM) est cependant plus faible à ce qui est trouvé dans la littérature (K_D 1 mM) pour une concentration traces d'ARN testé. De plus, à une concentration d'ARN plus élevée permettant de simuler la

présence de deux boucles, l'affinité de cet ARN pour la guanidine ($K_D < 10 \mu\text{M}$) s'en est retrouvée augmentée selon ce qui était attendu et avec une affinité meilleure à ce qui est retrouvé dans la littérature ($K_D 0,2\text{mM}$).

Afin de déterminer si le candidat RNase P est également un ribo-interrupteur, il faudrait tester si l'ajout de guanidine a une influence sur la fonction de cette endonucléase. Or, il est, pour l'instant, impossible d'établir une conclusion à ce niveau-là, ce test n'ayant encore été effectué. Avec les résultats obtenus au cours de la présente recherche, il est toutefois suggéré que la RNase P pourrait être au minimum un aptamère étant donné sa capacité de liaison.

Jusqu'à présent, quatre familles de ribo-interrupteur guanidine ont été découvertes et il est possible de croire qu'il reste encore d'autres familles à découvrir. En effet, la présence de nombreuses familles de ce régulateur pourrait être expliquée par le fait que le cation guanidinium pourrait facilement interagir avec la charge négative des ARN. D'un autre côté, les ribo-interrupteurs guanidine pourraient être impliqués dans plus d'aspects biologiques qu'il ne semble a priori. Les récentes études publiées montrent, en effet, que certains ribo-interrupteurs guanidine participeraient à l'importation de ce métabolite plutôt que son exportation (Sinn *et al.*, 2021). Par ailleurs, il est intéressant de constater que plusieurs de ces ribo-interrupteurs sont également associés à de nombreux transporteurs SMR qui sont impliqués dans la résistance aux antibiotiques tels que SugE. En fait, de nombreux pathogènes résistants tels qu'*Escherichia coli* et *Klebsiella pneumoniae* qui sont connus pour causer des infections urinaires et des pneumonies possèdent des ribo-interrupteurs guanidine. Devant ces nombreuses associations, une recherche bio-informatique et expérimentale plus approfondies permettraient vraisemblablement de découvrir un nouveau mécanisme de régulation via les ribo-interrupteurs guanidine chez ces pathogènes, mais également de découvrir de nouveaux ribo-interrupteurs guanidine en amont de gènes de résistances.

8 BIBLIOGRAPHIE

- Barrick JE, Corbino KA, Winkler WC, Nahvi A, Mandal M, Collins J, Lee M, Roth A, Sudarsan N, Jona I, Wickiser JK & Breaker RR (2004) New RNA motifs suggest an expanded scope for ribo-interrupteurs in bacterial genetic control. *Proc Natl Acad Sci U S A* 101(17):6421-6426.
- Battaglia RA, Price IR & Ke A (2017) Structural basis for guanidine sensing by the *ykkC* family of ribo-interrupteurs. *RNA* 23(4):578-585.
- Bédard AV, Hien EDM & Lafontaine DA (2020) Ribo-interrupteur regulation mechanisms: RNA, metabolites and regulatory proteins. *Biochim Biophys Acta Gene Regul Mech* 1863(3):194501.
- Bocobza S, Adato A, Mandel T, Shapira M, Nudler E & Aharoni A (2007) Ribo-interrupteur-dependent gene regulation and its evolution in the plant kingdom. *Genes Dev* 21(22):2874-2879.
- Breaker RR (2011) Prospects for ribo-interrupteur discovery and analysis. *Mol Cell* 43(6):867-879.
- Breaker RR, Atilho RM, Malkowski SN, Nelson JW & Sherlock ME (2017) The Biology of Free Guanidine As Revealed by Ribo-interrupteurs. *Biochemistry* 56(2):345-347.
- Caldara M, Dupont G, Leroy F, Goldbeter A, De Vuyst L & Cunin R (2008) Arginine biosynthesis in *Escherichia coli*: experimental perturbation and mathematical modeling. *J Biol Chem* 283(10):6347-6358.
- Carthew RW & Sontheimer EJ (2009) Origins and Mechanisms of miRNAs and siRNAs. *Cell* 136(4):642-655.
- Cases I, de Lorenzo V & Ouzounis CA (2003) Transcription regulation and environmental adaptation in bacteria. *Trends Microbiol* 11(6):248-253.
- Cobb M (2017) 60 years ago, Francis Crick changed the logic of biology. *PLoS Biol* 15(9):e2003243.
- Crick F (1970) Central dogma of molecular biology. *Nature* 227(5258):561-563.
- De Lay NR & Garsin DA (2016) The unmasking of 'junk' RNA reveals novel sRNAs: from processed RNA fragments to marooned ribo-interrupteurs. *Curr Opin Microbiol* 30:16-21.
- Ellis JC & Brown JW (2009) The RNase P family. *RNA Biol* 6(4):362-369.
- Galagan JE, Calvo SE, Cuomo C, Ma LJ, Wortman JR, Batzoglou S, Lee SI, Baştürkmen M, Spevak CC, Clutterbuck J, Kapitonov V, Jurka J, Scazzocchio C, Farman M, Butler J, Purcell S, Harris S, Braus GH, Draht O, Busch S, D'Enfert C, Bouchier C, Goldman GH, Bell-Pedersen D, Griffiths-Jones S, Doonan JH, Yu J, Vienken K, Pain A, Freitag M, Selker EU, Archer DB, Peñalva MA, Oakley BR, Momany M, Tanaka T, Kumagai T, Asai K, Machida M, Nierman WC, Denning DW, Caddick M, Hynes M, Paoletti M, Fischer R, Miller B, Dyer P, Sachs MS, Osmani SA & Birren

- BW (2005) Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature* 438(7071):1105-1115.
- Garst AD, Edwards AL & Batey RT (2011) Ribo-interrupteurs: structures and mechanisms. *Cold Spring Harb Perspect Biol* 3(6).
- 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.
- Ghannam JY, Wang J & Jan A (2020) Biochemistry, DNA Structure. *StatPearls*, © 2020, StatPearls Publishing LLC., Treasure Island FL.
- Gil N & Ulitsky I (2020) Regulation of gene expression by cis-acting long non-coding RNAs. *Nat Rev Genet* 21(2):102-117.
- Greene RF, Jr. & Pace CN (1974) Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, alpha-chymotrypsin, and beta-lactoglobulin. *J Biol Chem* 249(17):5388-5393.
- Guerrier-Takada C, Gardiner K, Marsh T, Pace N & Altman S (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* 35(3 Pt 2):849-857.
- Guil S & Esteller M (2012) Cis-acting noncoding RNAs: friends and foes. *Nat Struct Mol Biol* 19(11):1068-1075.
- Hamana K (1996) Distribution of diaminopropane and acetylspermidine in Enterobacteriaceae. *Can J Microbiol* 42(2):107-114.
- Huang L, Wang J & Lilley DMJ (2017a) The Structure of the Guanidine-II Ribo-interrupteur. *Cell Chem Biol* 24(6):695-702 e692.
- Huang L, Wang J, Wilson TJ & Lilley DMJ (2017b) Structure of the Guanidine III Ribo-interrupteur. *Cell Chem Biol* 24(11):1407-1415 e1402.
- Hubé F & Francastel C (2018) Coding and Non-coding RNAs, the Frontier Has Never Been So Blurred. *Front Genet* 9:140.
- Jones CP & Ferré-D'Amaré AR (2017) Long-Range Interactions in Ribo-interrupteur Control of Gene Expression. *Annu Rev Biophys* 46:455-481.
- Jørgensen MG, Pettersen JS & Kallipolitis BH (2020) sRNA-mediated control in bacteria: An increasing diversity of regulatory mechanisms. *Biochim Biophys Acta Gene Regul Mech* 1863(5):194504.
- Kim JN & Breaker RR (2008) Purine sensing by ribo-interrupteurs. *Biol Cell* 100(1):1-11.
- Lenkeit F, Eckert I, Hartig JS & Weinberg Z (2020) Discovery and characterization of a fourth class of guanidine ribo-interrupteurs. *Nucleic Acids Res* 48(22):12889-12899.
- Li J & Liu C (2019) Coding or Noncoding, the Converging Concepts of RNAs. *Front Genet* 10:496.
- Loh E, Dussurget O, Gripenland J, Vaitkevicius K, Tiensuu T, Mandin P, Repoila F, Buchrieser C, Cossart P & Johansson J (2009) A trans-acting ribo-interrupteur

- controls expression of the virulence regulator PrfA in *Listeria monocytogenes*. *Cell* 139(4):770-779.
- Lotz TS & Suess B (2018) Small-Molecule-Binding Ribo-interrupteurs. *Microbiol Spectr* 6(4).
- Ma L, Bajic VB & Zhang Z (2013) On the classification of long non-coding RNAs. *RNA Biol* 10(6):925-933.
- Mann HB & Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics* 18:50-60.
- Mattick JS & Makunin IV (2006) Non-coding RNA. *Hum Mol Genet* 15 Spec No 1:R17-29.
- McCown PJ, Corbino KA, Stav S, Sherlock ME & Breaker RR (2017) Ribo-interrupteur diversity and distribution. *RNA* 23(7):995-1011.
- McDaniel BA, Grundy FJ, Artsimovitch I & Henkin TM (2003) Transcription termination control of the S box system: direct measurement of S-adenosylmethionine by the leader RNA. *Proc Natl Acad Sci U S A* 100(6):3083-3088.
- Meyer MM, Hammond MC, Salinas Y, Roth A, Sudarsan N & Breaker RR (2011) Challenges of ligand identification for ribo-interrupteur candidates. *RNA Biol* 8(1):5-10.
- 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. *Proc Natl Acad Sci U S A* 98(17):9736-9741.
- 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, Sudarsan N, Ebert MS, Zou X, Brown KL & Breaker RR (2002) Genetic control by a metabolite binding mRNA. *Chem Biol* 9(9):1043.
- Nature (1970) Central dogma reversed. *Nature* 226(5252):1198-1199.
- Nelson JW, Atilho RM, Sherlock ME, Stockbridge RB & Breaker RR (2017) Metabolism of Free Guanidine in Bacteria Is Regulated by a Widespread Ribo-interrupteur Class. *Mol Cell* 65(2):220-230.
- Noller HF (1991) Ribosomal RNA and translation. *Annu Rev Biochem* 60:191-227.
- Nou X & Kadner RJ (1998) Coupled changes in translation and transcription during cobalamin-dependent regulation of *btuB* expression in *Escherichia coli*. *J Bacteriol* 180(24):6719-6728.
- Nou X & Kadner RJ (2000) Adenosylcobalamin inhibits ribosome binding to *btuB* RNA. *Proc Natl Acad Sci U S A* 97(13):7190-7195.
- Pavlova N, Kaloudas D & Penchovsky R (2019) Ribo-interrupteur distribution, structure, and function in bacteria. *Gene* 708:38-48.
- Rampášek L, Jimenez RM, Lupták A, Vinař T & Brejová B (2016) RNA motif search with data-driven element ordering. *BMC Bioinformatics* 17(1):216.

- Randau L, Schröder I & Söll D (2008) Life without RNase P. *Nature* 453(7191):120-123.
- Regulski EE & Breaker RR (2008) In-line probing analysis of ribo-interrupteurs. *Methods Mol Biol* 419:53-67.
- Reiss CW & Strobel SA (2017) Structural basis for ligand binding to the guanidine-II ribo-interrupteur. *RNA* 23(9):1338-1343.
- Reiss CW, Xiong Y & Strobel SA (2017) Structural Basis for Ligand Binding to the Guanidine-I Ribo-interrupteur. *Structure* 25(1):195-202.
- Rivas E, Clements J & Eddy SR (2017) A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat Methods* 14(1):45-48.
- Romano G, Veneziano D, Acunzo M & Croce CM (2017) Small non-coding RNA and cancer. *Carcinogenesis* 38(5):485-491.
- Salvail H, Balaji A, Yu D, Roth A & Breaker RR (2020) Biochemical Validation of a Fourth Guanidine Ribo-interrupteur Class in Bacteria. *Biochemistry* 59(49):4654-4662.
- Serganov A & Nudler E (2013) A decade of ribo-interrupteurs. *Cell* 152(1-2):17-24.
- Sherlock ME & Breaker RR (2017) Biochemical Validation of a Third Guanidine Ribo-interrupteur Class in Bacteria. *Biochemistry* 56(2):359-363.
- Sherlock ME & Breaker RR (2020) Former orphan ribo-interrupteurs 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 Ribo-interrupteur Class in Bacteria. *Biochemistry* 56(2):352-358.
- Sherlock ME, Sadeeshkumar H & Breaker RR (2019) Variant Bacterial Ribo-interrupteurs Associated with Nucleotide Hydrolase Genes Sense Nucleoside Diphosphates. *Biochemistry* 58(5):401-410.
- Sinn M, Hauth F, Lenkeit F, Weinberg Z & Hartig JS (2021) Widespread bacterial utilization of guanidine as nitrogen source. *Mol Microbiol* 10.1111/mmi.14702.
- Stormo GD & Ji Y (2001) Do mRNAs act as direct sensors of small molecules to control their expression? *Proc Natl Acad Sci U S A* 98(17):9465-9467.
- Waldbauer JR, Coleman ML, Rizzo AI, Campbell KL, Lotus J & Zhang L (2019) Nitrogen sourcing during viral infection of marine cyanobacteria. *Proc Natl Acad Sci U S A* 116(31):15590-15595.
- Wang D & Farhana A (2020) Biochemistry, RNA Structure. *StatPearls*, © 2020, StatPearls Publishing LLC., Treasure Island FL.
- Weinberg Z, Barrick JE, Yao Z, Roth A, Kim JN, Gore J, Wang JX, Lee ER, Block KF, Sudarsan N, Neph S, Tompa M, Ruzzo WL & Breaker RR (2007) Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Res* 35(14):4809-4819.

- 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 Biol* 11(3):R31.
- Wilcoxon F (1947) Probability tables for individual comparisons by ranking methods. *Biometrics* 3(3):119-122.
- Winkler WC & Breaker RR (2003) Genetic control by metabolite-binding ribo-interrupteurs. *ChemBiochem* 4(10):1024-1032.
- Wuebben C, Vicino MF, Mueller M & Schiemann O (2020) Do the P1 and P2 hairpins of the Guanidine-II ribo-interrupteur interact? *Nucleic Acids Res* 48(18):10518-10526.
- Yarus M (1972) Intrinsic precision of aminoacyl-tRNA synthesis enhanced through parallel systems of ligands. *Nat New Biol* 239(91):106-108.
- Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31(13):3406-3415.

9 ANNEXE

Table S1. Liste des candidats RNArabo sélectionnés pour validation

Génome	# accession	Nom	Score
archaea 1 tige	NZ_KE386573.1	<i>Haloplanus natans</i> DSM 17983	24
	NZ_AOJK01000072.1	<i>Halorubrum californiensis</i> DSM 19288	24
	NZ_AEMG01000002.1	<i>Haladaptatus paucihalophilus</i> DX253	24
	NZ_ANPG01000079.1	<i>Haloferax</i> sp. BAB2207	24
	NZ_AOLO01000015.1	<i>Haloferax mediterranei</i> ATCC 33500	24
	NZ_AOLZ01000073.1	<i>Halobiforma lacisalsi</i> AJ5	24
	NZ_JROF01000001.1	<i>Halopiger salifodinae</i> strain KCY07-B2	24
	NZ_JROF01000025.1	<i>Halopiger salifodinae</i> strain KCY07-B2	24
	NZ_AOJD01000065.1	<i>Halorubrum tebenquichense</i> DSM 14210	24
	NZ_AOIV01000024.1	<i>Halogeometricum pallidum</i> JCM 14848	24
archaea 2 tiges	NZ_CP019327.1	<i>Haloterrigena daqingensis</i> strain JX313	24,5
	NZ_CP007055.1	<i>Halostagnicola larsenii</i> XH-48	25
	NZ_AOMC01000167.1	<i>Halococcus morrhuae</i> DSM 1307	26
	NZ_FTNP01000002.1	<i>Haloterrigena daqingensis</i> strain CGMCC 1.8909	24,5
	NZ_AOLG01000047.1	<i>Haloferax prahovense</i> DSM 18310	28
	NZ_LN831302.1	<i>Halobacterium hubeiense</i> strain JI20-1	39

bactérie 1 tige	CP021169.1	<i>Bacillus subtilis strain TLO3</i>	24
	NZ_CP020896.1	<i>Rhizobium etli strain Bra5</i>	24
	NZ_CP020947.1	<i>Rhizobium etli CIAT 894</i>	24
	NZ_CP020722.1	<i>Bacillus subtilis strain Bs-115</i>	24
	NZ_LT629712.1	<i>Bifidobacterium longum strain Su859</i>	24
	NZ_CP019724.1	<i>Streptomyces pactum strain ACT12</i>	24
	NZ_CP019724.1	<i>Streptomyces pactum strain ACT12</i>	24
	NZ_CP019724.1	<i>Streptomyces pactum strain ACT12</i>	24
	NZ_CP018627.1	<i>Streptomyces hygroscopicus strain XM201</i>	24
	NZ_CP018614.1	<i>Lactobacillus delbrueckii subsp. indicus strain JCM 15610</i>	24
bactérie 2 tiges	NZ_CP019724.1	<i>Streptomyces pactum strain ACT12</i>	48
	CP019724.1	<i>Streptomyces pactum strain ACT12</i>	48
	NZ_CP010313.1	<i>Bradyrhizobium japonicum strain E109</i>	25,5
	NC_017249.1	<i>Bradyrhizobium japonicum USDA 6 DNA</i>	25,5
	NC_015312.1	<i>Pseudonocardia dioxanivorans CB1190</i>	30
	NC_013929.1	<i>Streptomyces scabiei 87.22</i>	23
	NC_008278.1	<i>Frankia alni str. ACN14A chromosome</i>	27
	NC_003888.3	<i>Streptomyces coelicolor A3(2) chromosome</i>	34
fungi 1 tige	NW_006271969.1	<i>Cordyceps militaris CM01</i>	24

	NW_006383032.1	<i>Chaetomium thermophilum var. thermophilum</i> DSM 1495	24
	NC_021436.1	<i>Rhizoctonia solani strain AG3 Rhs1AP</i> mitochondrion	24
	NW_006921540.1	<i>Pseudocercospora fijiensis CIRAD86</i>	24
	NW_007361095.1	<i>Colletotrichum graminicola M1.001</i>	24
	NW_012234701.1	<i>Pseudogymnoascus destructans</i>	24
	NW_006767446.1	<i>Punctularia strigosozonata HHB-11173 SS5</i>	24
	NW_006912911.1	<i>Capronia epimyces CBS 606.96 aczIQ-</i> <i>supercont1.8</i>	24
	NW_013550599.1	<i>Rhinocladiella mackenziei CBS 650.93</i>	24
fungi 2 tiges	NW_006271974.1	<i>Cordyceps militaris CM01</i>	32,5
	NW_006271975.1	<i>Cordyceps militaris CM01</i>	31
	NW_006711154.1	<i>Trichoderma reesei QM6a</i>	33,5
	NW_006913794.1	<i>Eutypa lata UCREL1</i>	32
	NW_006920871.1	<i>Pseudozyma flocculosa PF-1</i>	30,5
	NW_008804547.1	<i>Gaeumannomyces graminis var. tritici R3-</i> <i>111a-1</i>	32
	NW_012133795.1	<i>Pseudozyma hubeiensis SY62 DNA, scaffold:</i> <i>PHS13</i>	32,5
	NW_007361089.1	<i>Colletotrichum graminicola M1.001</i>	38
	NW_006711190.1	<i>Trichoderma reesei QM6a</i>	30,5

	NW_002988901.1	<i>Postia placenta Mad-698-R POSPL</i>	30
humain 1 tige	NC_000001.11	<i>Homo sapiens chromosome 1, GRCh38.p7</i>	23,5
	NC_000002.12	<i>Homo sapiens chromosome 2, GRCh38.p7 Primary Assembly</i>	22,5
	NC_000002.12	<i>Homo sapiens chromosome 2, GRCh38.p7 Primary Assembly</i>	23,5
	NC_000003.12	<i>Homo sapiens chromosome 3, GRCh38.p7 Primary Assembly</i>	22
	NC_000006.12	<i>Homo sapiens chromosome 6, GRCh38.p7 Primary Assembly</i>	23
	NC_000006.12	<i>Homo sapiens chromosome 6, GRCh38.p7 Primary Assembly</i>	24
	NC_000007.14	<i>Homo sapiens chromosome 7, GRCh38.p7 Primary Assembly</i>	24
	NC_000008.11	<i>Homo sapiens chromosome 8, GRCh38.p7 Primary Assembly</i>	22
	NC_000008.11	<i>Homo sapiens chromosome 8, GRCh38.p7 Primary Assembly</i>	23
	NC_000012.12	<i>Homo sapiens chromosome 12, GRCh38.p7 Primary Assembly</i>	22,5
	NC_000014.9	<i>Homo sapiens chromosome 14, GRCh38.p7 Primary Assembly</i>	22
	NC_000017.11	<i>Homo sapiens chromosome 17, GRCh38.p7 Primary Assembly</i>	21,5

humain 2 tiges	NC_000009.12	<i>Homo sapiens chromosome 9, GRCh38.p7 Primary Assembly</i>	30
	NC_000009.12	<i>Homo sapiens chromosome 9, GRCh38.p7 Primary Assembly</i>	30
	NC_000009.12	<i>Homo sapiens chromosome 9, GRCh38.p7 Primary Assembly</i>	29,5
	NC_000018.10	<i>Homo sapiens chromosome 18, GRCh38.p7 Primary Assembly</i>	40,5
	NC_000019.10	<i>Homo sapiens chromosome 19, GRCh38.p7 Primary Assembly</i>	30,5
IGR 1 tige	NC_012528.1	seq115709	24
	NC_016109.1	seq134747	24
	NC_008709.1	seq196152	24
	NC_004557.1	seq341437	24
	NC_005085.1	seq512336	24
	NC_021252.1	seq638907	24
	NC_002677.1	seq744501	24
	NC_010804.1	seq868243	24
	NC_002977.6	seq948601	24
	NC_010162.1	seq1081966	24
IGR 2 tiges	NC_010805.1	seq89863	15
	NC_023064.1	seq169108	17

	NC_023064.1	seq169517	17
	NC_008709.1	seq196152	16,5
	NC_014729.1	seq894729	16,5
	NC_014721.1	seq1403564	15,5
	NC_014151.1	seq1475142	18,5
	NC_006351.1	seq1585646	15
	NC_006351.1	seq1585647	15
	NC_010338.1	seq2235326	16
	NC_014391.1	seq2393342	18,5
	NC_002946.2	seq2536553	15
	NC_003888.3	seq2748033	16
	NC_010086.1	seq2867956	15
virus 1 tige	NC_021858.1	<i>Pandoravirus dulcis</i>	22,5
	NC_022098.1	<i>Pandoravirus salinus</i>	23
	NC_023677.1	<i>Chimpanzee alpha-1 herpesvirus strain 105640</i>	24
	NC_023721.1	<i>Mycobacterium phage Drago</i>	24
	NC_028379.1	<i>Elephant endotheliotropic herpesvirus 4 isolate North American NAP69</i>	24
	NC_031936.1	<i>Gordonia phage GAL1</i>	22,5
	NC_030230.1	<i>Tokyovirus A1 DNA</i>	22,5
	NC_033464.1	<i>Spheniscid herpesvirus 2 isolate lib01004</i>	23,5

	NC_008196.1	<i>Mycobacterium phage Llij</i>	24
	NC_009233.1	<i>Heliothis virescens ascovirus 3e</i>	24
virus 2 tiges	NC_020231.1	<i>Caviid herpesvirus 2 strain 21222</i>	30
	NC_021307.1	<i>Mycobacterium phage Severus</i>	31
	NC_022063.1	<i>Myobacteriophage Phelemich,</i>	30,5
	NC_022098.1	<i>Pandoravirus salinus</i>	31
	NC_023602.1	<i>Mycobacterium phage 32HC</i>	30,5
	NC_026591.1	<i>Mycobacterium phage Trike</i>	31
	NC_008520.2	<i>Anticarsia gemmatalis nucleopolyhedrovirus</i>	30,5
	NC_002169.1	<i>Spodoptera exigua nucleopolyhedrovirus</i>	35
	NC_026440.1	<i>Pandoravirus inopinatum isolate KlaHel</i>	30
	NC_026440.1	<i>Pandoravirus inopinatum isolate KlaHel</i>	36,5
	NC_028818.1	<i>Streptomyces phage TP1604</i>	36
	NC_029065.1	<i>Pseudomonas phage VCM</i>	30,5
	NC_030912.1	<i>Gordonia phage McGonagall</i>	35
	NC_031761.1	<i>Anticarsia gemmatalis multicapsid nucleopolyhedrovirus isolate AgMNPV-37</i>	30,5
	NC_032001.1	<i>Only Syngen Nebraska virus 5</i>	31
NC_005137.2	<i>Choristoneura fumiferana DEF MNPV</i>	30	

Environ une dizaine de candidat par génome a été sélectionné pour la validation expérimentale.