Multiples mécanismes de régulation post-transcriptionnelle chez les bactéries : des structures d'ARN messager aux ARN régulateurs

Jonathan Jagodnik

▶ **To cite this version:**

Jonathan Jagodnik. Multiples mécanismes de régulation post-transcriptionnelle chez les bactéries : des structures d'ARN messager aux ARN régulateurs. Microbiologie et Parasitologie. Université Sorbonne Paris Cité, 2017. Français. NNT : 2017USPCC111 . tel-02000219

HAL Id: tel-02000219

https://theses.hal.science/tel-02000219

Submitted on 1 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thèse de doctorat

de l'Université Sorbonne Paris Cité

Préparée à l'Université Paris Diderot

Ecole doctorale BioSPC – ED562

CNRS UMR8261 « Expression Génétique Microbienne » / Equipe Eliane Hajnsdorf

« Régulation de l'expression génétique par l'ARN »

Multiples mécanismes de régulation post-transcriptionnelle chez les bactéries : des structures de l'ARN messager aux ARN régulateurs

Par Jonathan Jagodnik

Thèse de doctorat de Microbiologie

Dirigée par Maude Guillier & Claude Chiaruttini

Présentée et soutenue publiquement à Paris le 15 septembre 2017

Président du jury :	Nadal, Marc / Professeur / Université Paris Diderot
Rapporteurs :	Romby, Pascale / DR1 / CNRS, Université de Strasbourg
	Bossi, Lionello / DR Emérite / CNRS, Université Paris Saclay
Examinateurs :	Massé, Eric / Professeur / Université de Sherbrooke
	Delepelaire, Philippe / DR2 / CNRS, Université Paris Diderot
Directrice de thèse :	Guillier, Maude / CR1 / CNRS
Co-directeur de thèse :	Chiaruttini, Claude / CR1 / CNRS

Résumé

Titre: ***Multiples mécanismes de régulation post-transcriptionnelle chez les bactéries : des structures de l'ARN messager aux ARN régulateurs***

Résumé: Chez les bactéries, la régulation de l'expression génétique est fondamentale pour permettre une adaptation optimale à l'environnement. De nombreux contrôles existent, notamment au niveau post-transcriptionnel par de nombreux ARN régulateurs (sRNA pour « small RNAs »). Ceux-ci ciblent des ARN messager (ARNm), permettant une régulation rapide de la synthèse de protéines. Le plus souvent, ces sRNAs interagissent avec leur(s) cible(s) au niveau du site de fixation du ribosome (RBS pour « ribosome binding site »), entrant dès lors en compétition avec le ribosome pour la fixation à l'ARNm et entraînant une régulation négative de l'expression des gènes cibles. Pour autant, il existe de nombreux mécanismes alternatifs de régulation par les sRNA.

Nous avons ainsi pu démontrer que les deux sRNAs OmrA et OmrB, conservés au sein des entérobactéries, répriment la synthèse du récepteur FepA aux complexes fer-entérobactine en ciblant une structure de l'ARNm *fepA*. Cette structure en tige-boucle est située en aval du RBS de *fepA*, et de façon surprenante, elle contrôle positivement la synthèse de FepA *via* une activation de la fixation de la sous-unité 30S du ribosome à l'ARNm. Des structures similaires ont pu être prédites dans d'autres ARNm, à l'image de *bamA*, codant la sous-unité essentielle du complexe Bam d'adressage des protéines de membrane externe en tonneaux β. Comme pour *fepA*, la tige-boucle de l'ARNm *bamA* favorise la fixation du ribosome, suggérant que ce mécanisme de régulation pourrait être bien plus général du fait de la grande conservation de *bamA* au sein des bactéries à Gram négatif. De surcroît, ces résultats constituent la première illustration que les structures d'ARNm peuvent avoir un effet positif sur la traduction.

Par ailleurs, deux autres sRNAs répriment également et indépendamment l'expression de *fepA*, à savoir SdsR et RseX. A chaque fois, le mécanisme de régulation impliqué est différent. Ainsi, SdsR s'apparie vraisemblablement à deux régions différentes de l'ARNm *fepA*, impliquant notamment une compétition classique avec la fixation du ribosome. La répression par RseX nécessite quant à elle la présence d'autres séquences du 5'UTR de *fepA*, à plus d'une centaine de nucléotides (nts) en amont du RBS. Enfin, chacun de ces sRNAs semble répondre à des stimuli différents, ce qui enrichit considérablement notre connaissance des signaux contribuant à la régulation de *fepA*, dont jusqu'ici seule la carence en fer était connue comme un signal de dérépression par le facteur de transcription Fur.

Ce travail est une nouvelle illustration de l'immense diversité des mécanismes de régulation impliquant des ARN, dont la grande flexibilité de structure et de séquence constitue une importante source de diversité à la disposition de l'évolution.

Mots clefs: ARNm ; sRNA ; Hfq ; régulation post-transcriptionnelle ; « Starting Block » ; toeprint ; réponse au stress ; fer ; mécanismes

Abstract

Title: ***Multiple post-transcriptional regulatory mechanisms in bacteria: from mRNA structures to regulatory RNAs***

Abstract: In order to perfectly adapt to their environment, bacteria require a tight gene expression regulation. This can occur through post-transcriptional control by numerous regulatory RNAs (or small RNAs, sRNAs). These sRNAs can target mRNAs, leading to a fast regulation of protein synthesis. Most often, sRNAs base-pair with their target mRNAs at the ribosome binding site (RBS), therefore competing with the ribosome for the binding with the mRNA and repressing gene expression. However, many other regulatory mechanisms involve sRNAs.

We have demonstrated that the two sRNAs OmrA and OmrB, conserved among enterobacteria, repress the synthesis of the FepA receptor for iron-enterobactin complexes through base-pairing with a secondary structure within *fepA* mRNA. This stem-loop structure is located downstream of *fepA* RBS, and most surprisingly, promotes 30S ribosomal subunit binding to *fepA* mRNA, therefore activating FepA synthesis. Similar stem-loop structures have been predicted in other mRNAs, such as the *bamA* mRNA encoding the essential subunit of the Bam outer membrane protein assembly complex. As for *fepA* mRNA, the stem-loop found in *bamA* mRNA also promotes ribosome binding, showing that this regulatory mechanism could be widespread considering the strong conservation of *bamA* among Gram negative bacteria. Moreover, these results challenge the commonly admitted view of mRNA secondary structures being repressors of gene expression.

Two other sRNAs also repress *fepA* expression in an OmrA/B-independent fashion, namely SdsR and RseX. For each of these sRNAs, the regulatory mechanism involved is different. Indeed, SdsR most likely acts through two distinct binding sites, one of which leading to a classical competition with the ribosome binding. Meanwhile, RseX repression requires most of *fepA* 5'UTR, including sequences at about 100nt upstream of the start codon. Finally, each of these sRNAs is expressed upon diverse stimuli, considerably extending our knowledge of the signals leading to *fepA* regulation, for which only the Fur-dependent derepression upon iron starvation was known.

This work highlights the great versatility of regulatory mechanisms involving RNAs. This illustrates how RNAs structural flexibility and sequence diversity is a key source of diversity for evolution.

Keywords : mRNA ; sRNA ; Hfq ; post-transcriptional regulation ; starting block ; toeprint ; stress response ; iron ; mechanisms

À mes quatre grands-parents, et à leur courage

Remerciements

Maude, si je ne te remercie pas la première, qui me resterait-il à remercier ? J'ai été touché par ta confiance. Tu m'as soutenu les yeux fermés à d'innombrables reprises, emmené à un formidable congrès au Mexique et envoyé à un cours de génétique captivant à Cold Spring Harbor Laboratories. Nos discussions scientifiques ont été de grands moments de bonheur, et me laisseront un souvenir impérissable – de même que ton inénarrable sens de l'humour ! Merci infiniment pour tout cela, et merci encore pour tout ce que tu m'as enseigné, depuis les astuces de la génétique jusqu'au raisonnement scientifique le plus irréprochable (en passant par le temps de cuisson de la semoule). Qu'aurais-je pu attendre de plus ? J'espère que cette thèse n'est que la première (quoiqu'épaisse) page d'une longue collaboration scientifique : l'avenir nous le dira.

Merci également à toi Claude, de m'avoir appris l'art ancestral du « toeprint » et des expériences *in vitro* – nous pouvons aujourd'hui affirmer avoir navigué à travers les eaux troubles de l'acrylamide, de la fluorescence et du 1-cyclohexyl-2-morpholino-trunchtruc ! Merci également pour tes nombreux conseils et pour le temps accordé à la réflexion sur les structures secondaires des ARN. Mathias, merci de m'avoir pris sous ton aile dès mon arrivée. Ton soutien et tes judicieuses remarques scientifiques m'ont été précieux.

Je remercie mes rapporteurs Pascale Romby et Lionello Bossi d'avoir accepté de juger mon travail de thèse, mais aussi pour nos passionnantes discussions scientifiques à l'occasion des divers cours et congrès. Je remercie également mes examinateurs Éric Massé (qui a courageusement accepté de traverser l'Atlantique !) et Philippe Delepelaire (qui a non moins courageusement accepté de traverser l'Institut !). Merci également à Philippe d'avoir été disponible pour discuter du fer. Merci enfin à Marc Nadal d'avoir accepté de présider mon jury de thèse avec enthousiasme.

Eliane et Fanette, merci de m'avoir permis de réaliser cette thèse, et ce dans les meilleures conditions. Eliane, ta bienveillance est légendaire. Maxence, merci d'avoir été un super étudiant de M1, et de faire aujourd'hui le bonheur du laboratoire en thèse. Je remercie

à l'évidence l'intégralité de l'équipe (involontairement mais judicieusement orthographiée « épique » il y a un instant), à savoir également Joël, Anaïs, Nguyen, Mark et Rudi pour leur bonne humeur de tous les jours, de même que les nombreux membres passés, et notamment Audrey pour m'avoir consacré du temps avant de quitter le laboratoire. Travailler à vos côtés a été une grande source de satisfaction !

Je remercie également l'intégralité de l'unité « Expression Génétique Microbienne » pour son ambiance familiale, ainsi que ceux de l'institut avec qui j'ai passé du temps, nommément : Maxence pour sa gentillesse (et sa fourberie), Hilal et Nathalie pour leur joie de vivre et leur folie des chats (longue vie aux chiens !), Justin et Nisha parce qu'ils sont mignons, Wojtek parce qu'il est bien trop gentil avec Nathalie, et tous les autres auprès desquels je m'excuse platement parce que cette liste devient longue !

Un remerciement spécial à ma famille, éparpillée aux quatre coins du monde, qui m'a supporté à bout de bras durant tout mon parcours et qui a cru en moi ! J'en profite pour remercier ma seconde famille, mes trois frères d'adoption : Arthur, Aviv et Elliott (dans l'ordre alphabétique, pour ne pas faire de jaloux !), qui ont été là à chaque instant. Merci Emilie d'avoir été à mes côtés. Merci à tous mes amis, y compris ceux que je n'ai pas suffisamment vus au cours de cette longue thèse : Sami, Wahid, Stan, Rémi, Pierre, Iffat, Elise, Bruno, Samuel, Arnold, Saz, Thomas, les comparses du master ARN et Evolution, et tous les autres. Je pense aussi aux nouvelles et formidables rencontres du cours à Cold Spring Harbor ! Sans oublier Pauline, pour avoir partagé avec moi le plus clair de cette thèse et avoir courageusement accepté de relire toute mon introduction pour la purger de ses fautes d'orthographe, merci ! Je te souhaite bonne chance pour la suite.

Un remerciement personnel à la musique et notamment à Nemost pour m'avoir appris la persévérance, mais aussi à Porcupine Tree, Chopin, Opeth, Pain of Salvation, Jimi Hendrix, Bring me the Horizon, Sia, Katatonia, et tous les autres artistes qui m'ont accompagné pendant la thèse et tout particulièrement pendant l'écriture de ce manuscrit !

Enfin, je remercie chaleureusement le Labex DYNAMO d'avoir rendu cette thèse possible du début à la fin.

Abréviations

aa	acide aminé
ADN	acide désoxyribonucléique
AGE	« agarose gel electrophoresis »
anti-SD	région anti-Shine Dalgarno de l'ARN ribosomique 16S
ARN	acide ribonucléique
ARNm	ARN messager
ARNr	ARN ribosomique
ARNt	ARN de transfert
asRNA	ARN antisens
CMCT	1-cyclohéxyl-(2-morpholinoéthyl)carbodiimide métho-p-toluène sulfonate
DHB-Ser	dihydroxybenzoyl-serine
dip	2,2'-dipyridyl
DMS	diméthylsulfate
E. coli	*Escherichia coli*
facteur σ	facteur sigma
Fe-S	centre fer-soufre
FT	facteur de transcription
LPS	Lipopolysaccharides
NB	« Northern blot »
nt	nucléotide
OmrA/B	OmrA et OmrB
OMP	« outer membrane protein », protéine de membrane externe
ORF	« open reading frame », phase ouverte de lecture, cistron
PAGE	« polyacrylamide gel electrophoresis »
pb	paire de bases
RBS	« ribosome binding site », site de fixation du ribosome
RNase	ribonucléase
RNAseq	séquençage d'ARN à haut débit
RT	« retrotranscription », transcription inverse
SB	« starting block »
SD	séquence Shine-Dalgarno de l'ARNm
SDC	système à deux composants
sRNA	« small RNA », par extension : ARN régulateur
TIR	« translation initiation region », région de démarrage de la traduction
TSS	« transcription start site », site de démarrage de la transcription
UTR	« untranslated transcribed region », régions non traduites des ARNm
WB	« Western blot »
wt	« wild type », sauvage

Table des matières

Résumé 2
Abstract 3
Remerciements 5
Abréviations 7
Table des matières 8
Index des Figures et Tableaux 11
Introduction 13
Chapitre I : Place de l'ARN dans l'évolution et Traduction bactérienne 14

A. L'ARN au fil de l'évolution 14
1. Propriétés chimiques de l'ARN 14
2. Ribozymes et origines de la vie 16
a. Les premières molécules aux origines de la vie 16
b. L'introduction du concept de ribozyme 17
c. Le monde ARN 19
3. Évolution vers la traduction, puis vers l'ADN 22
a. Conservation de la traduction et dernier ancêtre commun 22
b. Utilisation de l'ADN comme support de l'information génétique 24

B. Rouages de la traduction bactérienne 26
1. Généralités et principaux constituants ARN 26
a. Définition de la traduction 26
b. ARN messagers 26
c. ARN de transfert et code génétique 28
d. Ribosomes et ARN ribosomiques 30
2. Démarrage de la traduction 33
a. Sous-unité ribosomique 30S et complexe de pré-démarrage 33
b. Autres facteurs nécessaires au démarrage de la traduction 35
c. Sous-unité ribosomique 50S et formation du complexe de démarrage de la traduction 37
3. Allongement et terminaison de la traduction 39
a. Allongement de la traduction : synthèse des polypeptides 39
b. Terminaison de la traduction et recyclage des ribosomes 41
Chapitre II : Expression génétique et sa régulation par l'ARN chez les entérobactéries 43

A. Transcription et régulation transcriptionnelle 43
1. Régulateurs protéiques de la transcription 43
a. Facteurs interagissant avec l'ARN polymérase 43
b. Facteurs de transcription 44
2. Régulateurs ARN de la transcription 45
a. ARN 6S : un ARN mimant un ADN 45
b. Riboswitches et terminaison transcriptionnelle 46
c. Régulateurs transcriptionnels agissant par appariement de séquence : les ARN antisens 47

B. Régulation post-transcriptionnelle par les protéines 49
1. CsrA, un régulateur post-transcriptionnel protéique 49
2. RNases : la stabilité et la dégradation des ARNm 50

C. Régulation post-transcriptionnelle par les ARN 50
1. Séquences, RBS atypiques et rôles de la protéine ribosomique S1 51
a. Séquences « enhancer » et site de fixation de la protéine S1 51
b. RBS atypiques 53
c. Fréquence et rôles des codons 54
2. ARNm et structures régulatrices 55
a. Riboswitches et effets post-transcriptionnels 55

b. Thermosenseurs ... 57
c. Autres structures ayant un impact sur l'expression génétique ... 58
3. ARN régulateurs agissant par complémentarité de séquence ... 61
a. asRNA agissant au niveau post-transcriptionnel ... 62
b. ARN agissant par appariement imparfait, ARN Hfq-dépendants ... 63
c. **Revue n°1**: "*Mechanistic study of base-pairing small regulatory RNAs in bacteria*" ... 66
d. ARN « éponges » ... 77
Chapitre III : Adaptation des entérobactéries à l'environnement, réponse au stress et ARN régulateurs ... 78

A. L'enveloppe cellulaire des entérobactéries ... 78
1. Composition de l'enveloppe chez les bactéries à Gram négatif ... 78
a. Phospholipides et lipopolysaccharides ... 79
b. Paroi peptidoglycane ... 80
c. Les protéines membranaires et périplasmiques ... 81
2. OMP, adressage et rôles divers ... 82
a. OMP et physiologie de la cellule ... 82
b. Adressage des OMP à la membrane externe ... 84
3. Motilité et adhésion ... 87
a. Motilité : les flagelles ... 87
b. Adhésion, biofilms et protéines Curli ... 89

B. Réponse au stress, ARN régulateurs et circuits de régulation ... 90
1. Le système à deux composants EnvZ-OmpR et les ARN régulateurs OmrA et OmrB ... 90
a. EnvZ-OmpR et la réponse au stress membranaire ... 91
b. OmrA et OmrB, deux sRNA Hfq-dépendants similaires ... 94
2. Facteurs sigma et ARN régulateurs ... 99
a. Réponse au stress membranaire : le facteur σE et le sRNA RseX ... 99
b. Phase stationnaire de croissance : le facteur σS et le sRNA SdsR ... 102

C. Cas de figure : la réponse à la carence en fer ... 105
1. Physiologie du fer chez les entérobactéries ... 106
a. Protéines à centre fer-soufre ... 106
b. Fer libre et cytotoxicité ... 106
2. Import de fer : sidérophores et transporteurs ... 107
a. Sidérophores ... 107
b. Transporteurs de fer ... 108
3. Réponse à la carence en fer ... 110
a. Fur et RyhB : la régulation de l'homéostasie du fer ... 110
b. Régulation de *fepA* et de l'utilisation de l'entérobactine ... 112
Résultats ... 115
Chapitre I : Activation de la traduction par des structures secondaires de l'ARN messager ... 115

A. Avant-propos ... 115

B. Article de recherche: « Stem-loop structures within mRNA coding sequences activate translation initiation and mediate control by small regulatory RNAs » ... 116

C. Structures secondaires et activation de la traduction ... 154
1. L'hypothèse du « switch » ... 154
2. Implication des éléments en *cis* de l'ARNm ... 158
3. L'hypothèse du thermosenseur ... 162
4. L'existence d'autres structures secondaires dans l'ARNm *fepA* ... 164

D. Le rôle des RNases et de Hfq dans la régulation par OmrA/B ... 167
1. L'implication de la RNase E ... 167
2. L'implication de la RNase III ... 168
3. L'implication de Hfq *in vivo* ... 169
Chapitre II : À la croisée des mécanismes et des signaux amenant à la régulation de l'expression de *fepA* ... 171

A. Quatre sRNA Hfq-dépendants régulent la synthèse de FepA ... 171
1. La régulation de l'expression de *fepA* par SdsR et RseX ... 171

a. Démonstration de la régulation 171
b. Le rôle du 5'UTR de l'ARNm *fepA* et l'effet d'une délétion des sRNA 174
2. Les mécanismes de régulation de *fepA* par SdsR et RseX 177
a. Eléments mécanistiques communs 177
b. Le mécanisme de régulation de l'expression de *fepA* par SdsR 179
c. Le mécanisme de régulation de l'expression de *fepA* par RseX 183

B. Stimuli à l'origine de la régulation de l'expression de *fepA* 185
1. La réponse d'OmrA/B au fer 186
a. L'effet d'une carence en fer 186
b. La recherche d'un régulateur et l'induction par les ions Fe^{2+} 189
2. La recherche de régulateurs de RseX 194
a. La région promotrice de *rseX* 195
b. La recherche d'un régulateur transcriptionnel de RseX 199

Discussion 203

A. Eléments de réflexion sur la traduction et les ARNm 203
1. Mécanismes de régulation de la traduction reposant sur une structure secondaire activatrice de la traduction 203
a. Le modèle du starting block 203
b. Le modèle du recrutement 205
2. Structures SB : où les trouver ? 208
3. La définition du RBS 209
4. L'ARNm et le code de régulation génétique 209

B. De multiples régulations par les ARN 211
1. L'accumulation de signaux pour la régulation de l'expression de *fepA* 211
2. L'évolution et la régulation par les ARN 213

Matériels et Méthodes 218

A. Souches et milieux 218
a. Construction des souches 220
b. Milieux utilisés 221

B. Oligonucléotides, sondes et anticorps 222

C. Protocoles expérimentaux 222
a. Allongement d'amorce 222
b. RNAseq (Annexe 2) 223

Annexe 1 : **Revue n°2** « Small noncoding RNAs and stress responses » 224

Annexe 2 : Les autres cibles d'OmrA/B 238

A. L'interaction entre OmrA/B et l'ARNm *fecA* 238

B. Recherche de nouvelles cibles d'OmrA/B par RNAseq 240

Annexe 3 : Phénotypes liés aux sRNA 243

A. Au moins une autre cible commune à OmrA/B et SdsR? 243

B. OmrA/B et la taille des bactéries 245

Bibliographie 247

Index des Figures et Tableaux

Figure 1 – ARN, nucléotides et bases azotées 14
Figure 2 – Exemples de structures secondaires formées par l'ARN 16
Figure 3 – Exemple de mécanisme d'excision d'un « self-splicing intron » 18
Figure 4 – Conservation des RNases P à travers l'évolution 19
Figure 5 – Modèles d'émergence de systèmes auto-réplicatifs 20
Figure 6 – Modèle de cellule primitive dans le monde ARN 21
Figure 7 – Formation d'une liaison peptidique entre deux acides aminés 26
Figure 8 – Les différentes régions de l'ARNm 27
Figure 9 – Structure d'un ARNt 28
Figure 10 – Structure de la sous-unité 30S du ribosome bactérien 31
Figure 11 – structure de la sous-unité 50S du ribosome bactérien 32
Figure 12 – Complexe de pré-démarrage de la traduction 36
Figure 13 – Démarrage de la traduction bactérienne 38
Figure 14 – Allongement de la traduction bactérienne 40
Figure 15 – Promoteur et transcription 44
Figure 16 – Terminateur transcriptionnel Rho-indépendant 46
Figure 17 – Régulation transcriptionnelle par des asRNA 48
Figure 18 – Les diverses régions d'un RBS bactérien 52
Figure 19 – Diversité des mécanismes de régulation post-transcriptionnelle par des riboswitches 56
Figure 20 – Une structure activatrice de la traduction en amont du RBS de fis 59
Figure 21 – Effet hypothétique d'une tige-boucle au début d'une séquence codante 60
Figure 22 – Structure tridimensionnelle de la protéine Hfq 65
Figure 23 – Formation de duplexes sRNA – mRNA assistée par Hfq 65
Figure 24 – Composition de l'enveloppe des bactéries à Gram négatif 79
Figure 25 – Diverses OMP à tonneau β chez les bactéries à Gram négatif 82
Figure 26 – Adressage des OMP à tonneau β via les systèmes Sec et BAM 85
Figure 27 – Modèle de mécanisme d'intégration des OMP à la membrane externe par le complexe BAM 86
Figure 28 – Le flagelle bactérien 88
Figure 29 – Composition du complexe Curli chez les entérobactéries 89
Figure 30 – Le système à deux composants EnvZ-OmpR 92
Figure 31 – Régulon du système à deux composants EnvZ-OmpR 93
Figure 32 – Locus et conservation des deux sRNA OmrA et OmrB 94
Figure 33 – Régulon des sRNA OmrA et OmrB 96
Figure 34 – Sites d'appariement d'OmrA/B sur leurs ARNm cibles 97
Figure 35 – Prédictions d'appariement in silico entre les OmrA/B et les ARNm fepA, fecA et gntP 98
Figure 36 – Le facteur σE et la réponse au stress membranaire 100
Figure 37 – Le sRNA Hfq-dépendant RseX 101
Figure 38 – Rôle central du facteur σS dans la réponse au stress 103
Figure 39 – Le sRNA Hfq-dépendant SdsR 104
Figure 40 – Le régulon du sRNA SdsR 105
Figure 41 – La molécule d'entérobactine 107
Figure 42 – FepA et l'import de fer TonB-dépendant 109
Figure 43 – Régulons simplifiés de Fur et du sRNA RyhB 111
Figure 44 – Régulation de l'expression de fepA et de l'utilisation de l'entérobactine 114
Figure 45 – Un switch de structure pour expliquer l'effet régulateur d'OmrA/B sur fepA ? 155
Figure 46 – Le mutant Switch1 de la fusion fepA-lacZ est toujours régulé par OmrA/B 156
Figure 47 – Le mutant Switch2 de la fusion fepA-lacZ est toujours régulé par OmrA/B 157
Figure 48 – Le mutant Switch2-Loop de la fusion fepA-lacZ est toujours régulé par OmrA/B 158
Figure 49 – Les éléments en cis de l'ARNm sont importants pour l'effet de la structure SB de fepA 160
Figure 50 – La structure SB de fepA ne peut remplacer celle de bamA 162

Figure 51 – La structure SB de fepA répond à la température 163
Figure 52 – Structure secondaire de l'ARNm fepA 165
Figure 53 – Des oligonucléotides antisens ciblant la structure SL2 de fepA ont un effet sur le toeprint 166
Figure 54 – La RNase III n'est pas nécessaire à la régulation de la synthèse de FepA par OmrA/B 168
Figure 55 – Effets de Hfq et de la RNase E sur la régulation de la synthèse de FepA et sur OmrA/B 170
Figure 56 – Plusieurs sRNA Hfq-dépendants ont un effet sur la synthèse de FepA 172
Figure 57 – La régulation par RseX et SdsR dépend du 5'UTR de l'ARNm fepA 173
Figure 58 – Les deux formes du 5'UTR de fepA et l'effet des sRNA sur la synthèse de FepA 175
Figure 59 – La régulation de fepA par RseX ou SdsR est indépendante de l'effet d'autres sRNA 177
Figure 60 – RseX et SdsR ont un effet très rapide sur l'ARNm fepA 178
Figure 61 – SdsR s'apparie au RBS de fepA, et probablement également plus en amont au sein de son 5'UTR 180
Figure 62 – L'interaction de SdsR avec le RBS de l'ARNm fepA inhibe la fixation du ribosome 182
Figure 63 – L'appariement entre RseX et l'ARNm fepA reste à démontrer 183
Figure 64 – Une carence en fer active l'expression d'OmrA et OmrB au niveau transcriptionnel 187
Figure 65 – Le FeSO4 compense l'effet du dip à basse concentration et semble activer la transcription d'OmrB à haute concentration 191
Figure 66 – Seuls les ions Fe^{2+} induisent l'expression d'OmrB 192
Figure 67 – Le fer et le dip affectent l'expression d'ompC et ompF au niveau post-transcriptionnel 193
Figure 68 – Une forme plus courte de RseX pourrait exister in vivo 195
Figure 69 – Expression génétique à partir des promoteurs P1 et P2 de RseX 196
Figure 70 – Autres promoteurs hypothétiques de rseX d'après le programme « Neural Network Promoteur Prediction » 197
Figure 71 – L'expérience de crible génétique 200
Figure 72 – Le modèle du « Starting Block » 204
Figure 73 – Modèle de recrutement de la sous-unité ribosomique 30S 206
Figure 74 – La régulation de l'expression de fepA 212
Figure 75 – La séquence consensus du site de fixation d'OmrA/B 216
Figure 76 – OmrA/B répriment l'expression de fecA via un appariement 239
Figure 77 – Gènes dont les niveaux d'ARNm sont affectés par OmrA/B d'après les résultats de RNAseq 240
Figure 78 – Phénotype de sédimentation lié à omrA/B et sdsR 243
Figure 79 – Phénotype de motilité lié à OmrA/B, SdsR et RseX 244
Figure 80 – OmrA/B ont un effet sur la taille d'Escherichia coli 246

Tableau 1 – Gènes candidats à un rôle dans la régulation de RseX 201
Tableau 2 – Souches et plasmides 220
Tableau 3 – Sondes, oligonucléotides et anticorps 222

Introduction

Par leur nature plastique, pouvant servir de relais à l'information génétique et à même de former des structures complexes, les acides ribonucléiques (ARN) sont le support d'une immense variété de fonctions, dont beaucoup sont indispensables à l'ensemble des organismes vivants. C'est notamment par leur biais que s'effectue un grand nombre de régulations post-transcriptionnelles de l'expression génétique, permettant un contrôle de la synthèse des protéines d'une extrême finesse. Chez les bactéries, ces régulations de l'expression génétique apportent une capacité unique à s'adapter à de nombreuses perturbations de l'environnement.

Mon travail de thèse a débuté par l'étude chez *Escherichia coli* (*E. coli*) du mécanisme de régulation de l'ARN messager (ARNm) *fepA* par deux petits ARN régulateurs (sRNA pour « small RNAs ») OmrA et OmrB, pour aboutir à un mécanisme plus général de régulation de la traduction, impliquant des structures d'ARNm. Ce travail a également permis la découverte de la régulation de *fepA* par deux autres ARN régulateurs, SdsR et RseX, à chaque fois par le biais de mécanismes de régulation différents.

C'est la raison pour laquelle j'ai choisi d'orienter mon introduction vers les caractéristiques des ARN, la traduction bactérienne, mais aussi la régulation post-transcriptionnelle de l'expression génétique chez les entérobactéries qui implique souvent les ARN. Par ailleurs, l'ensemble de ces régulations permettant d'intégrer de nombreux signaux et d'aboutir à l'adaptation des bactéries à leur environnement, ce thème sera également abordé.

Chapitre I : Place de l'ARN dans l'évolution et Traduction bactérienne

A. L'ARN au fil de l'évolution

1. Propriétés chimiques de l'ARN

Les ARN sont des macromolécules qui entrent dans la composition des cellules dans les trois domaines du vivant. Ce sont des polymères de nucléotides (nt) (Fig. 1.A et 1.B), chaque nt étant constitué d'un pentose : le ribose, d'une base azotée et d'un groupement phosphate. Il existe quatre bases azotées différentes utilisées dans l'ARN (Fig. 1.C), à savoir deux purines : l'adénine et la guanine, ainsi que deux pyrimidines : la cytidine et l'uracile. On nomme respectivement les nucléosides formés avec ces quatre bases l'adénosine (A), la guanosine (G), la cytosine (C) et l'uridine (U).

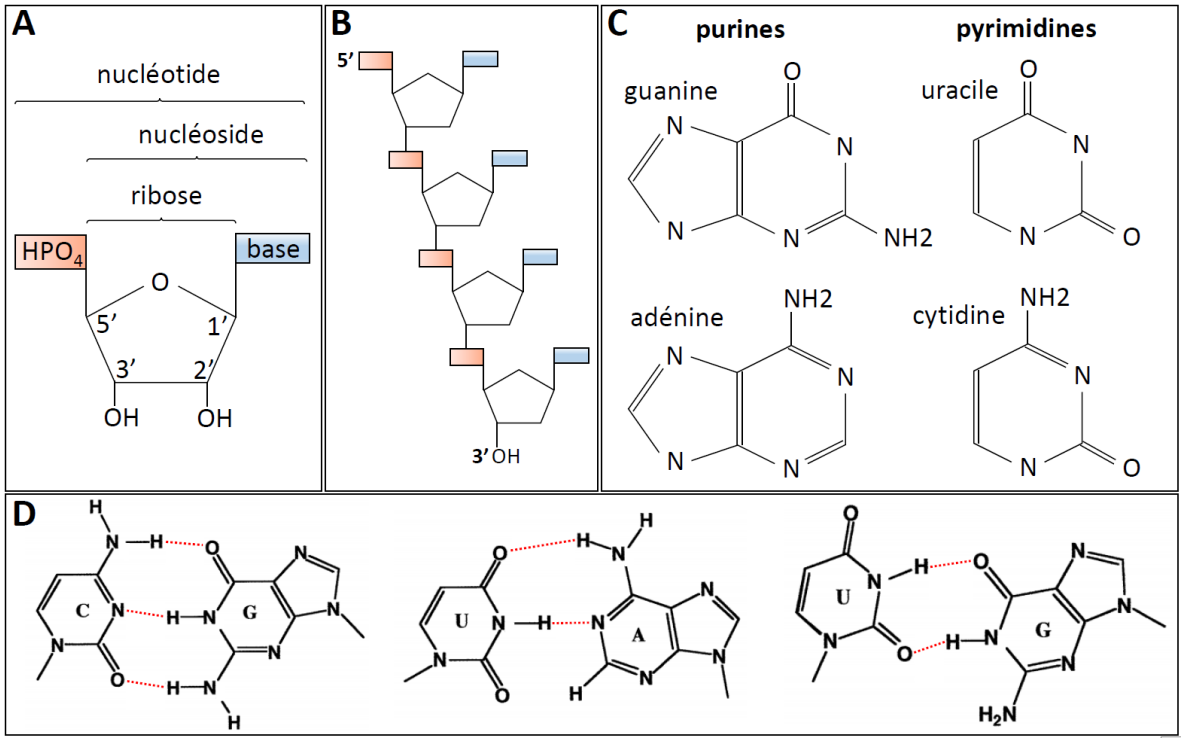


Figure 1 – ARN, nucléotides et bases azotées
A. Composition d'un nt d'ARN. Chaque trait noir représente une liaison covalente. B. Schéma d'une séquence de quatre nt ; les couleurs correspondent à la légende en A. avec les groupements phosphate en rouge et les bases azotées en bleu ; les extrémités 5'-phosphate et 3'-OH sont indiquées. C. Les quatre bases azotées utilisées dans les molécules d'ARN. D. Appariements entre base azotées au sein d'un duplex formé par deux brins d'ARN complémentaires. Sont représentés les deux appariements

Watson-Crick C-G et U-A, ainsi qu'un appariement Wobble U-G. Les liaisons d'hydrogène entre deux bases azotées sont représentées par des pointillés rouges. D'après (Ramos et al., 2004)

Les molécules d'ARN sont synthétisées *via* le processus de transcription, par une enzyme nommée ARN polymérase (ARNpol). Une molécule d'ARN débute par une extrémité 5'-triphosphate sauf quand elle est le résultat d'un clivage : dans ce cas, la nouvelle extrémité 5' générée est généralement monophosphate. La molécule d'ARN s'achève par une extrémité 3'-hydroxyle (3'), à l'image des molécules d'acide désoxyribonucléique (ADN). L'ARN et l'ADN sont également comparables par leur capacité à constituer des doubles hélices entre deux molécules de séquences complémentaires – d'ailleurs, une molécule d'ADN et une molécule d'ARN complémentaires peuvent s'apparier pour donner une double hélice hybride.

L'ADN et l'ARN diffèrent en revanche par la présence d'un groupement OH à la position 2' du ribose de l'ARN, qui rend ce dernier plus sensible à l'hydrolyse en pH alcalin. En plus des appariements de type Watson-Crick trouvés dans les double-hélices d'ADN, l'ARN est capable de former des paires de bases moins classiques telles que les appariements Wobble entre les nt U et G (Fig. 1.D). Cette dernière caractéristique, ainsi que leur capacité à établir des appariements imparfaits en laissant des segments d'ARN sous forme simple brin - les « bulges » - leur confère une grande flexibilité. Cela leur permet notamment d'adopter de nombreuses structures différentes (Fig. 2), d'où peut découler une grande diversité de fonction. C'est aussi, en plus du 2'-OH, une autre raison pour laquelle les ARN sont généralement moins stables que les molécules d'ADN, notamment parce qu'une double-hélice parfaite résiste mieux à l'augmentation de la température ou aux attaques de nucléases, des enzymes dégradant les acides nucléiques. Ainsi, la demi-vie des ARNm chez les bactéries dépasse rarement quelques minutes, ce qui se répercute sur l'expression des gènes. La présence de deux extrémités OH (2' et 3') sur les molécules d'ARN permet néanmoins de nombreuses réactions chimiques et explique en partie leur capacité à interagir avec de nombreuses molécules biologiques. Enfin, les molécules d'ARN contiennent des bases uraciles à comparer aux bases thymines trouvées dans l'ADN, ces bases ne différant que par un groupement méthyl supplémentaire dans la thymine.

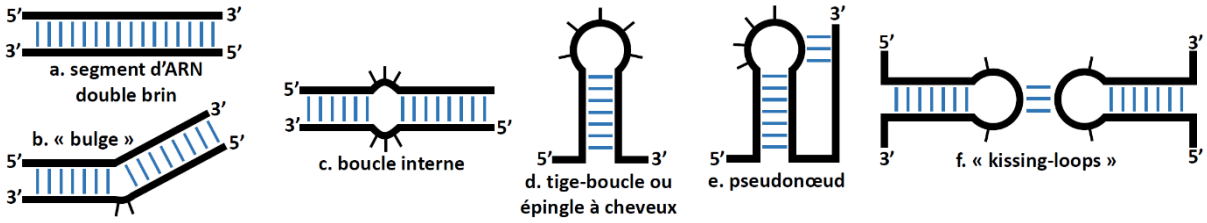


Figure 2 – Exemples de structures secondaires formées par l'ARN
Représentations schématiques de structures secondaires adoptées par les ARN, avec en noir épais les brins d'ARN, en traits noirs fins les nt non appariés et en traits bleu les paires de bases. a. Une double hélice peut se former entre deux régions complémentaires d'un ARN. b. Les « bulges » sont des portions d'ARN exclues d'une double hélice d'ARN. c. Des boucles internes prennent forme quand des nt restent non appariés de chaque côté d'un ARN double brin. Elles peuvent être symétriques ou asymétriques si le nombre de nt non appariés de chaque côté est respectivement identique ou différent. d. Une tige-boucle se forme lorsque les deux régions appariées sont proches l'une de l'autre, avec généralement 3 à 8 nt dans la boucle simple brin. e. Un pseudonœud correspond à l'appariement de nt de la boucle d'une structure d'ARN avec une région simple brin éloignée de l'ARN. f. Quand les boucles de deux structures en épingle à cheveux ont des séquences complémentaires, elles peuvent former des appariements appelés « kissing-loops ». Toutes ces structures peuvent être intramoléculaires, mais les structures a, b, c et f peuvent également être intermoléculaires. D'après (Tian et al., 2004).

2. Ribozymes et origines de la vie

L'omniprésence des ARN dans les trois domaines du vivant et leur forte conservation au cœur de mécanismes essentiels à la vie en font des molécules indispensables à tous les organismes connus à ce jour. L'évolution des ARN depuis les origines de la vie offre un point de vue original sur la position centrale que ces molécules occupent aujourd'hui ; c'est pourquoi j'ai considéré que cela pouvait apporter une lumière et une introduction intéressantes à l'ensemble de cette thèse sur les ARN.

a. Les premières molécules aux origines de la vie

Les origines de la vie constituent un vaste domaine de recherche. Elles ont été intensément étudiées au cours du siècle dernier et regorgent encore aujourd'hui d'énigmes qu'il reste à résoudre. Une des grandes questions qui a fait débat durant des décennies peut se formuler ainsi : quelles ont été les premières molécules à avoir supporté des fonctions biologiques essentielles aux formes de vie les plus primitives ? En effet, pendant longtemps l'hypothèse de l'ADN comme molécule primordiale – un excellent support de l'information génétique du fait de sa stabilité – a coexisté avec l'hypothèse privilégiant les protéines. Ces dernières peuvent effectivement former des machineries de synthèse perfectionnées et les acides aminés (aa) qui constituent ces protéines sont les produits de réactions chimiques

remarquablement simples, observables dans des expériences mimant des conditions pré-biotiques (Miller 1953) ou volcaniques (Johnson et al., 2008) par exemple. Ces hypothèses reposent principalement sur l'observation des organismes vivants modernes, dont la reproduction est axée autour de la réplication de l'ADN porteur de l'information génétique par des protéines. Mais le problème devient circulaire dès lors que l'on sait qu'il faut une information génétique, donc de l'ADN, pour parvenir à la sélection évolutive et la synthèse correcte d'une protéine complexe, et qu'à son tour la duplication de l'ADN nécessite des protéines sophistiquées. Une troisième hypothèse toujours soutenue à l'heure actuelle défend le rôle primordial des lipides, permettant la compartimentation en liposomes et donc l'augmentation locale de la concentration en molécules, favorisant les réactions chimiques (Segré et al., 2001). Cette dernière hypothèse ne saurait cependant expliquer l'accumulation d'information, aussi n'est-elle pas séparable de l'évolution en parallèles de molécules ayant des propriétés catalytiques, permettant la réplication d'un matériel génétique et donc l'évolution.

La solution à cette énigme correspond très vraisemblablement à une quatrième hypothèse qui, si elle n'exclut pas un rôle des autres molécules, positionne l'ARN à l'origine de la vie dans ce que l'on appelle « le monde ARN ». Cette hypothèse, formulée pour la première fois en 1986 et depuis reprise par d'autres (Gilbert, 1986; Sankaran, 2016), propose que les ARN soient à l'origine de l'évolution. C'est la théorie la plus consensuelle à l'heure actuelle, malgré quelques fervents défenseurs de l'hypothèse protéique notamment (Kurland, 2010). La théorie du « monde ARN » a ainsi vu le jour peu après la mise en évidence des premiers ribozymes, des ARN dotés de propriétés catalytiques à l'image d'une enzyme protéique. Il peut être intéressant de comprendre en quoi ces ribozymes ont changé la vision de la place occupée par l'ARN afin d'appréhender au mieux la théorie du monde ARN.

b. L'introduction du concept de ribozyme

Le premier exemple de ribozyme, découvert en 1982, est un intron de l'ARN ribosomique (ARNr) 26S catalysant sa propre excision en l'absence de toute protéine chez l'eucaryote monocellulaire cilié *Tetrahymena thermophila* (Kruger et al., 1982). Doté d'une capacité à réaliser son propre clivage *in vitro* pour produire un ARNr mature pour la traduction, ce type d'ARN intronique est retrouvé sous diverses formes dans tous les domaines du vivant.

C'est par exemple le cas de l'ARNm de la thymidilate synthase du bactériophage T4 d'*E. coli*, où un intron de plus d'un kilo-base est capable de générer son propre épissage. On appelle communément ces ARN des introns auto-excisants ou « self-splicing introns » (Fig. 3).

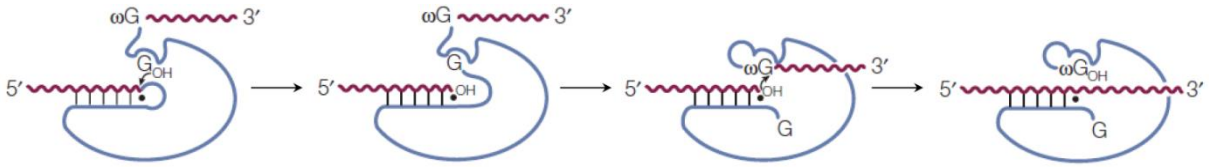


Figure 3 – Exemple de mécanisme d'excision d'un « self-splicing intron »
Le ribozyme intronique est représenté en bleu, les deux exons reliés après épissage sont indiqués en violet et les paires de base en noir. Dans un premier temps, le ribozyme recrute une guanosine qui sert de cofacteur au clivage l'ARN en attaquant, par son groupement 3'-OH, le groupement phosphate à la jonction de l'exon 5' et de l'intron ; cette guanosine se liera ensuite à l'extrémité 5' de l'intron. Puis, le nt G en 5' de l'exon 3' (ωG) se positionne au site actif du ribozyme, et va être attaqué par l'extrémité 3'-OH de l'exon 5'. Cette réaction va aboutir à la liaison des deux exons et à la séparation complète de l'intron. D'après (Fedor and Williamson, 2005).

Un second exemple de ribozyme a été publié dans un article de 1983, portant sur la ribonucléase (RNase) P, une ribonucléoprotéine essentielle composée de sous-unités protéique et ARN et conservée chez les bactéries, archées et eucaryotes (Fig. 4). Cet article démontre que l'activité endoribonucléasique de la RNase P, c'est-à-dire sa capacité à cliver une molécule d'ARN, dépend de sa sous-unité ARN (Guerrier-Takada et al., 1983) aussi bien chez la bactérie à Gram négatif *E. coli* que chez la bactérie à Gram positif *Bacillus subtilis* (*B. subtilis*). Chez ces deux bactéries très éloignées d'un point de vue évolutif, la sous-unité ARN de la RNase P est même capable de cliver les substrats naturels de cette enzyme, à savoir les extrémités 5' des ARN de transfert (ARNt), en l'absence de la sous-unité protéique. Il est à noter toutefois que si cette activité catalytique de la sous-unité ARN est conservée dans tous les domaines du vivant, il existe chez les eucaryotes de nombreux exemples où ce sont des protéines (nommées PRORP pour « protein only RNase P ») qui portent l'activité catalytique, et ce en l'absence de toute sous-unité ARN (Fig. 4).

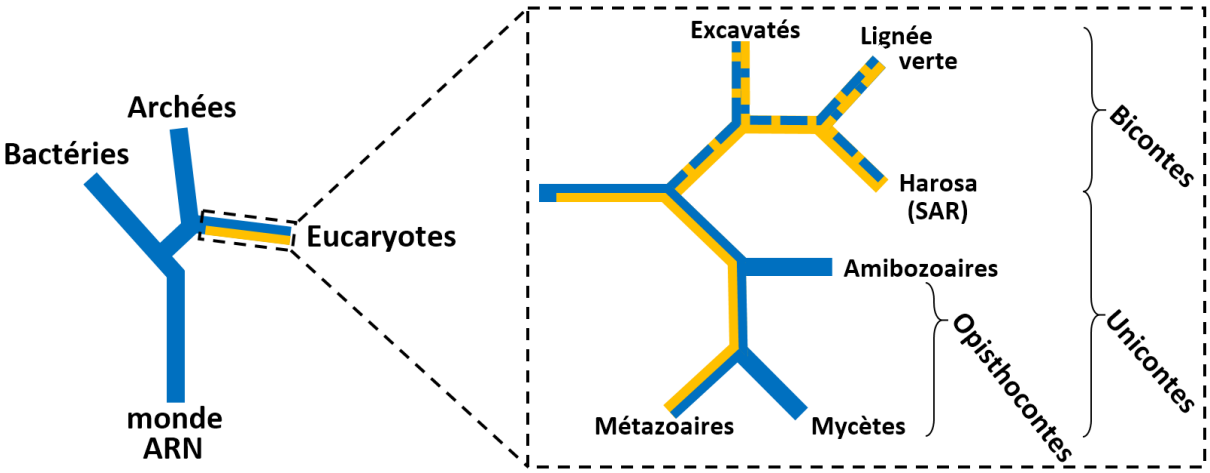


Figure 4 – Conservation des RNases P à travers l'évolution
La présence de RNase P fonctionnant avec un ribozyme (bleu) ou de PRORP fonctionnant avec une protéine catalytique (jaune) est indiquée. Chez les eucaryotes, on retrouve souvent les deux, à l'exception des mycètes et des amibozoaires où l'on ne trouve que le ribozyme. Les pointillés indiquent que certaines espèces dans l'embranchement ont perdu l'une ou l'autre forme de l'enzyme. Remarque : les groupes phylogénétiques sont en débat constant et peuvent être amenés à changer. D'après (Klemm et al., 2016).

De nombreux autres ribozymes sont connus aujourd'hui, et il a été en particulier démontré que c'est le cas des domaines catalytiques des ribosomes, à savoir les machineries effectuant la synthèse des protéines *via* le processus de traduction à travers l'ensemble du vivant (voir **Introduction I.B**).

c. Le monde ARN

De nombreux ARN recèlent donc des activités enzymatiques, et la grande conservation évolutive des deux types de ribozymes sus-cités évoque une origine très lointaine dans l'évolution, précédant probablement la différentiation du vivant en trois domaines. D'où l'hypothèse du monde ARN, molécules porteuses d'activités enzymatiques et d'information génétique, aux origines de la vie. Cette hypothèse est soutenue par une littérature très étoffée, faisant la démonstration étape par étape des éléments nécessaires à un monde ARN en évolution. Ainsi, des réactions chimiques simples sont à même de produire les constituants de base de l'ARN, à savoir les nt, dans diverses conditions compatibles avec l'atmosphère et l'océan pré-biotique (Powner et al., 2009, 2010). Des surfaces abiotiques dérivées de silices ont pu servir de matrices à la polymérisation aléatoire des premiers oligonucléotides d'ARN (Cassidy et al., 2011). La polymérisation spontanée de brins complémentaires à partir de matrices ARN a également été observée (Deck et al., 2011; Lehman and Hayden, 2011). Ce

dernier processus a pu contribuer à introduire des mutations dans les copies d'ARN nouvellement synthétisées, en constituant une première forme d'évolution darwinienne. Enfin, différents modèles ont été imaginés pour expliquer l'émergence des premiers systèmes subvenant à leur propre réplication (Meyer et al., 2012), une étape clé pour générer une évolution darwinienne plus rapide, qui a pu amener l'apparition d'ARN de plus en plus complexes (Fig. 5).

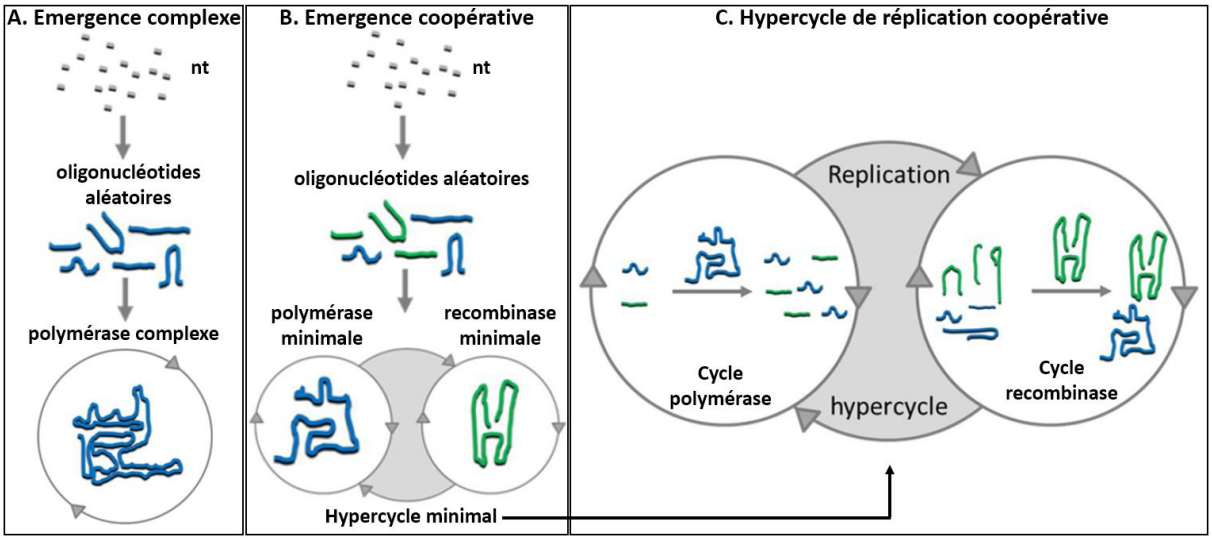


Figure 5 – Modèles d'émergence de systèmes auto-réplicatifs
Exemples de modèles expliquant l'émergence de systèmes capables de subvenir à leur propre réplication et soumis à l'évolution darwinienne. Sont représentés en bleu les ribozymes à activité polymérase, et en vert les ribozymes à activité recombinase. A. Scénario d'émergence complexe, proposant que des ribozymes à faible activité polymérase soient apparus, puis par sélection naturelle ont fini par évoluer en polymérase complexe capable de se répliquer. B. Scénario d'émergence coopérative, où l'apparition d'un ribozyme à faible activité polymérase a pu coexister et coopérer avec un ribozyme à faible activité recombinase pour former un hypercycle de réplication. C. Hypercycle de réplication en détails, avec un cycle de polymérisation permettant la synthèse de fragments de polymérase et de recombinase, et un cycle de recombinaison permettant la ligature de ces fragments pour produire des ribozymes fonctionnels. D'après (Meyer et al., 2012).

Si aujourd'hui la recherche n'offre encore que des projections sur les différents scénarios plausibles aux origines de la vie, imaginer les premières étapes de l'évolution peut constituer un exercice intéressant. En partant de l'hypothèse du monde ARN, la première fonction qui a pu être conservée par un ARN est une activité permettant sa réplication ou du moins la favorisant. À partir de là, l'introduction d'erreurs, mais peut-être également d'autres ARN parasites dans le processus de réplication, a pu générer de nombreux ARN ayant des fonctions enzymatiques différentes. Ceux-ci ont pu s'assembler pour constituer un chromosome ARN primitif, répliqué par un ribozyme dédié (Fig. 6). D'autres ARN intégrés à ce chromosome ont pu avoir une activité ribonucléasique, à l'instar des « self-splicing introns »

ou de la RNase P évoqués précédemment, et cette activité a pu servir notamment à séparer chacun des composants du chromosome ARN afin qu'ils puissent réaliser leurs activités enzymatiques. Cette séparation en de multiples ARN différents évoque une expression génétique, si l'on considère les ARN associés dans le chromosome comme des gènes, et les ARN séparés comme les molécules effectrices ayant des fonctions diverses. Une autre fonction des ribozymes à activité ribonucléasique a pu être la digestion d'ARN étrangers au chromosome auxquels ils sont intégrés, afin de générer les nt nécessaires à la réplication, constituant une forme primitive de prédation. On peut alors imaginer que des spécificités de séquence ont permis à ces nucléases de différencier les molécules d'ARN étrangères. Cette activité ribonucléasique a également pu s'appliquer dans des proportions différentes à tous les ARNs séparés du chromosome ARN, permettant un contrôle de la quantité de chaque espèce d'ARN et évoquant cette fois-ci une régulation de l'expression génétique.

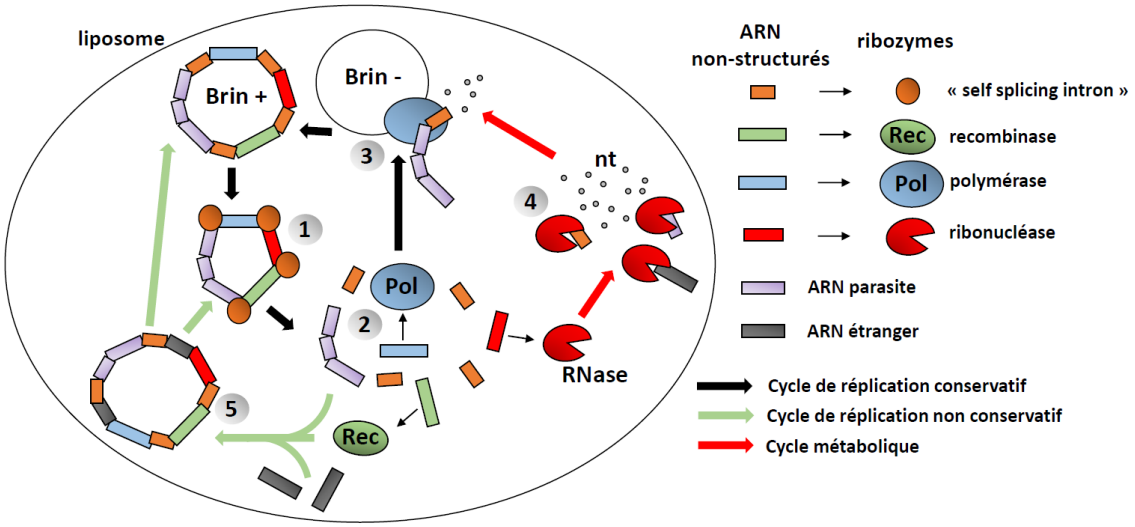


Figure 6 – Modèle de cellule primitive dans le monde ARN

D'après ce scénario, un chromosome constitué d'ARN peut être séparé en plusieurs morceaux après l'excision de « self-splicing introns » (1). Ces morceaux de chromosome peuvent alors adopter leurs structures fonctionnelles (2). Ainsi un ribozyme à activité polymérase peut répliquer un chromosome servant de matrice ou brin- afin de générer un nouveau chromosome fonctionnel ou brin+ (3). Parmi les autres unités fonctionnelles, on peut imaginer un ribozyme à activité RNase (4) qui va pouvoir dégrader des ARN parasites ou étrangers, des ARN dont la cellule primitive n'a plus besoin afin de générer des nt pour la réplication, ou encore d'autres ARN fonctionnels dont il faudrait limiter la quantité. On peut également imaginer qu'un ribozyme à activité recombinase puisse réassembler les blocs d'ARN fonctionnels ensembles selon un nouvel arrangement et en intégrant des ARN étrangers (5), ce qui va générer de la diversité dans les chromosomes disponibles pour la réplication et dans les unités fonctionnelles générées. On notera que le chromosome matrice peut lui-même contenir d'autres unités d'ARN fonctionnelles, et que la polymérase peut également utiliser comme matrices le brin+, les chromosomes réassemblés par la recombinase ou encore les ARN séparés du chromosome. Par ailleurs, ce scénario nécessite la séparation des différents constituants de l'environnement extérieur afin de favoriser les réactions chimiques : on peut ainsi imaginer que les divers ARN soient ici inclus dans un liposome, dont la division asymétrique dépendrait de sources extérieures de lipides et de la pression osmotique régnant à l'intérieur du liposome.

Quel qu'ait pu être le véritable scénario à l'origine de la vie, l'hypothèse ARN semble à plus d'un titre cohérente et raisonnable. De façon importante pour la suite de mon propos, cela met en relief la variété des fonctions attribuées à l'ARN, dont certaines trouvent leur origine dans les prémices de l'évolution sur Terre. De plus, cette position originelle de l'ARN permet de concevoir la place centrale occupée par l'ARN aujourd'hui dans les processus du vivant comme un héritage logique de l'évolution, que ce soit pour l'expression des gènes ou pour la régulation de l'expression génétique par exemple. L'évolution a ainsi pu construire autour de ce qui existait déjà, diversifiant les fonctions au fur et à mesure, et mettant à profit l'usage de nouvelles molécules telles que les protéines dotées d'activités catalytiques plus sophistiquées, ou l'ADN pour un stockage plus stable de l'information génétique.

3. Évolution vers la traduction, puis vers l'ADN

Parce que la traduction est apparue au début de l'histoire évolutive de la vie telle qu'on la connaît aujourd'hui, elle constitue la suite logique de ce propos. En effet, l'immense diversité combinatoire des 20^n polypeptides potentiels à n aa laisse imaginer le nombre de protéines qu'il est possible de produire et les fonctions nouvellement accessibles grâce à leur synthèse. Ce nombre est à mettre en relief avec les 4^n molécules d'ARN différentes à n nt possibles du fait de l'usage de seulement quatre nt différents. On comprend dès lors que l'apparition des premiers complexes capables de synthétiser des protéines, à fortiori à partir d'une séquence nucléotidique soumise à l'évolution *via* des cycles de réplication et mutation, a constitué une véritable révolution qui est aujourd'hui conservée dans l'ensemble du vivant. La traduction sera ici abordée du point de vue de l'évolution et du monde ARN, mais ses mécanismes généraux seront décrits plus en détails par la suite (voir **Introduction I.B**).

a. Conservation de la traduction et dernier ancêtre commun

L'une des caractéristiques frappantes de la traduction est l'usage de machineries traductionnelles très semblables dans les trois domaines du vivant, à savoir les ribosomes, constitués d'ARNr et de protéines. C'est d'ailleurs grâce à la conservation d'un des constituants des ribosomes, l'ARNr 16S, qu'a été édifiée la classification moderne du vivant en trois domaines : Bactéries, Archées et Eucaryotes (Woese et al., 1975) (Fig. 4). Or, les ARNr sont des éléments cruciaux de ces ribosomes, qui sont au cœur de la structure des ribosomes,

mais aussi et surtout les porteurs de l'activité catalytique des ribosomes dans la synthèse des protéines (voir **Introduction I.B.1.d**). Ce sont donc des ribozymes extrêmement conservés, d'où naît l'hypothèse que les premiers ribosomes ont pu être constitués uniquement d'ARN.

De façon tout aussi saisissante, on retrouve à travers l'ensemble du vivant une conservation quasiment parfaite du code génétique. Ainsi on dénombre à l'heure actuelle à peine plus d'une trentaine d'exceptions à ce code génétique, la plupart retrouvées dans les mitochondries d'eucaryotes et correspondant souvent à des permutations entre codons sens et codons « stop » (Bezerra et al., 2015) (voir **Introduction I.B.1.a et c**). Parmi les autres éléments qui varient entre différents domaines du vivant, on trouve l'utilisation d'un aa formyl-méthionine spécifié par le codon de démarrage de la traduction chez les bactéries, les mitochondries et les chloroplastes, à la place de la méthionine (Caskey et al., 1968; Sherman et al., 1985). Toutes proportions gardées, ces quelques exceptions restent toutefois mineures comparées à la conservation des 64 codons au cours de l'évolution. Le code génétique fait donc partie des caractéristiques qui unissent l'ensemble du vivant et trouve par conséquent son origine avant le dernier ancêtre commun à l'ensemble des clades existant aujourd'hui. Ce sont à nouveau des ARN qui jouent un rôle primordial dans la conservation et la lecture du code génétique, à savoir ARNt (voir **Introduction I.B.1.c**). Le fait que la structure de ces ARNt soit si conservée, et que la lecture du code génétique repose sur leur capacité à s'apparier avec les ARNm ainsi qu'à interagir avec les aa laisse imaginer qu'ils sont à l'origine de l'apparition d'une traduction fidèle à un code génétique unique.

On retrouve aujourd'hui de nombreuses enzymes non ribosomiques capables de réaliser des liaisons peptidiques sans matrice ARNm (Giessen and Marahiel, 2012), et la grande sous-unité du ribosome semble également posséder cette faculté en présence de certaines protéines (Shen et al., 2015). Il est ainsi possible de proposer un scénario où de premiers ribozymes – ou des proto-ribosomes - ayant la capacité de catalyser la synthèse de courtes chaînes d'aa sans suivre les instructions d'une matrice ont vu le jour. Puis, simplement parce que certains aa ont une affinité accrue pour les codons ARN qui les spécifient (Knight and Landweber, 1998), il est permis d'imaginer que certaines molécules d'ARN - ou des proto-ARNm - ont pu conduire les proto-ribosomes à synthétiser des peptides en suivant une séquence donnée et elle-même soumise à l'évolution. Parce que ces peptides apportaient un

avantage sélectif aux cellules primitives, des « proto-ARNt » interagissant à la fois avec les proto-ARNm et avec les aa ont pu co-évoluer avec les proto-ribosomes et favoriser la synthèse de peptides plus fidèles à la séquence matrice. Enfin la synthèse des protéines, soumise à une forte pression de sélection pour être la plus performante possible, s'est accélérée et est devenue de plus en plus fidèle, notamment grâce à l'intégration de protéines de plus en plus sophistiquées au sein même des ribosomes.

En résumé, comme évoqué précédemment, l'évolution a permis au vivant d'accéder à de nouvelles fonctions en construisant et étoffant autour de ce qui existait déjà. En l'occurrence, le mécanisme de la synthèse protéique a évolué autour de ribozymes et autres ARN de plus en plus complexes, jusqu'à ce que les protéines elles-mêmes deviennent suffisamment efficaces et élaborées pour intégrer ce processus. C'est probablement la raison pour laquelle on retrouve jusqu'à ce jour les ARN au cœur-même du mécanisme de la traduction, et qu'au sein des ribosomes les protéines se situent plutôt à la périphérie des petite et grande sous-unités des ribosomes (voir **Introduction I.B.1.d**).

b. Utilisation de l'ADN comme support de l'information génétique

Il est intéressant de considérer la chronologie de l'apparition de l'ADN comme support de l'information génétique, puisqu'il s'agit également d'un héritage commun à l'ensemble du vivant. S'il est possible de se demander si l'ADN a été intégré par les cellules primitives avant ou après l'usage de polypeptides, quelques indices dans la conservation et la fonction de certaines protéines actuelles laissent penser que ces dernières ont évolué avant que l'ADN ne serve à conserver l'information génétique.

Le premier indice est l'existence d'ARN polymérases ARN-dépendantes (RPRD pour « RNA polymerase RNA-dependent »). Celles-ci sont conservées au sein d'une variété de virus dits « à ARN » car leur génome est entièrement constitué d'ARN, et répliqué par ces enzymes (Bruenn, 2003). D'une manière générale, ces enzymes sont donc capables de répliquer efficacement des chromosomes ARN (Deng et al., 2006) et l'on conçoit l'utilité d'une telle enzyme dans la réplication au sein de cellules primitives basées sur l'ARN. Or, ces virus infectant des organismes aussi variés que des bactéries et des eucaryotes, il est envisageable qu'ils aient co-évolué avec leurs hôtes depuis le dernier ancêtre commun à l'ensemble du

vivant, et conservé ces enzymes qui sont aujourd'hui superflues dans la plupart des organismes utilisant l'ADN comme support de l'hérédité. Une des caractéristiques importantes de ces enzymes est donc leur capacité à synthétiser de l'ARN non seulement pour copier un ARN génomique, mais également pour produire des ARNm qui peuvent ensuite être traduits en protéines.

Un second indice découle du fait que ces RPRD sont conservées avec d'autres enzymes nommées transcriptases inverses (Makeyev and Grimes, 2004), des ADN polymérases ARN-dépendantes responsables de la transcription inverse (« retrotranscription », RT) par opposition aux ARNpol qui effectuent la transcription d'ADN en ARN. A l'instar des RPRD, on trouve ces transcriptases inverses dans des virus dénommés rétrovirus, dont l'ARN génomique subit une réaction de RT afin d'être répliqué par la machinerie de l'hôte infecté. Tout comme les virus portant une RPRD, les rétrovirus infectent des organismes extrêmement éloignés sur l'arbre du vivant et trouvent vraisemblablement leur origine avant cette diversification du vivant. De plus, cette enzyme protéique a pu rendre très facile la transition entre génome ARN et génome ADN, avant de devenir obsolète une fois que la réplication d'ADN à partir d'ADN est devenue possible grâce aux ADN polymérases ADN-dépendantes (ADNpol).

Les homologies structurales et mécanistiques existant entre toutes ces polymérases (Steitz, 1998) suggèrent ainsi qu'elles ont pu évoluer à partir de la même enzyme. Cela a pu permettre aux cellules primitives, développant toujours plus de protéines différentes grâce à la traduction et donc toujours plus d'ARNm différents, d'effectuer une transition vers l'ADN. Ce dernier est en effet plus stable et a donc probablement permis un stockage plus efficace de cette quantité croissante d'information génétique. Ce sont dès lors aux ARNpol qu'a échu le rôle de synthétiser les ARN et aux ADNpol celui de répliquer l'information génétique, rendant désuète l'utilisation des RDRP et transcriptases inverses – sauf chez certains virus qui ont préservé l'ARN comme support de l'information génétique, potentiellement parce que la taille très réduite de leur génome le permet.

B. Rouages de la traduction bactérienne

1. Généralités et principaux constituants ARN

a. Définition de la traduction

La traduction est le processus de polymérisation d'aa par des ribosomes, suivant les instructions fournies par une séquence nucléotidique composée de codons de trois nt. C'est le code génétique qui détermine l'aa spécifié par chaque codon. La séquence de codons est donc lue et traduite par les ribosomes en séquence d'aa. Les polymères d'aa sont appelés polypeptides car la liaison entre deux aa est une liaison de type peptidique (Fig. 7).

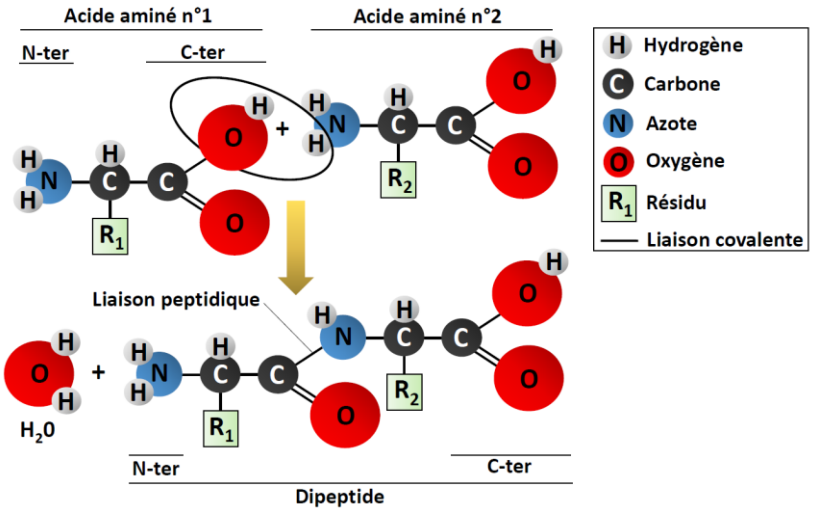


Figure 7 – Formation d'une liaison peptidique entre deux acides aminés
C'est la réaction entre l'extrémité C-terminale (C-ter) d'un aa avec l'extrémité N-terminale (N-ter) de l'aa suivant qui va former une liaison peptidique, en générant un dipeptide de séquence R1-R2 et une molécule d'eau. Le résidu est la seule partie qui diffère d'un aa à l'autre.

Chaque aa diffère des autres par son résidu, chaque résidu conférant des propriétés différentes aux protéines finales. Ainsi, l'usage de 20 aa distincts dans l'assemblage des polypeptides permet l'association de 20^n polypeptides différents, où n est égal à la longueur du polypeptide. Au cours de sa synthèse ou de façon post-traductionnelle, le polypeptide va pouvoir acquérir des conformations tridimensionnelles dictées en grande partie par la séquence d'aa, pour former une protéine fonctionnelle.

b. ARN messagers

Le processus de traduction requiert une matrice de lecture sur laquelle se baser pour fabriquer un polypeptide : cette matrice est l'ARNm. Lui-même transcrit à partir d'une matrice

ADN par l'ARNpol, l'ARNm constitue l'intermédiaire entre le gène codant et la protéine encodée fonctionnelle. En effet si l'on s'en tient au dogme central de la biologie moléculaire, chaque gène codant est transcrit en ARNm. L'ARNm est donc, comme l'ADN, porteur de la séquence codante nécessaire à la synthèse du polypeptide spécifié par le gène.

Plus particulièrement, l'ARNm contient deux régions non traduites : le 5'UTR et le 3'UTR (pour « untranslated transcribed region »), situées de part et d'autre de la séquence codante. Cette dernière, également nommée phase ouverte de lecture (ORF pour « open reading frame ») ou cistron (Fig. 8), consiste en une suite de triplets de nt formant des codons. Le début de l'OFR (et le cadre de lecture) est défini par un codon de démarrage de la traduction et la fin par un codon « stop » ou codon non-sens (voir **Introduction I.B.3.b**). Le codon de démarrage de la traduction correspond le plus souvent au codon AUG, traduit en aa formyl-méthionine chez les bactéries. Il s'agit là d'une exception pour les codons AUG, puisque les codons AUG internes sont traduits en méthionine sans groupement formyl, et que chez les eucaryotes le codon de démarrage AUG est également traduit en simple méthionine. Par ailleurs, les bactéries disposent de plusieurs autres codons de démarrage de la traduction moins fréquemment utilisés (voir **Introduction I.B.1.b et I.B.2.a**)

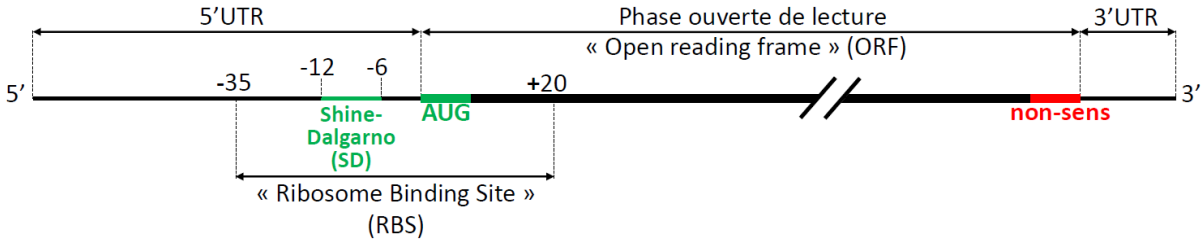


Figure 8 – Les différentes régions de l'ARNm
Un ARNm se décompose le plus souvent en deux régions non-traduites 5' et 3' (5'UTR et 3'UTR), et une région traduite appelée phase ouverte de lecture (ORF), correspondant à la partie épaissie, et s'achevant par un codon non-sens représenté en rouge. Le début de cette phase ouverte de lecture est déterminée par le codon de démarrage de la traduction AUG en vert ici, ainsi que par la présence d'une séquence SD quelques nt en amont, également en vert. La région de l'ARNm des nt -35 à +20 du codon AUG de démarrage constitue le site de fixation du ribosome ou « ribosome binding site » (RBS).

Un autre élément permettant de distinguer le premier codon d'une séquence codante chez les bactéries est la séquence dite Shine-Dalgarno (SD) entre 5 et 10 nt en amont du codon de démarrage (Shine and Dalgarno, 1974). Cette séquence SD est reconnue par l'une des sous-unités du ribosome comme un site de fixation (voir **Introduction I.B.2.a**)

c. ARN de transfert et code génétique

La traduction de la séquence de codons en aa suit les règles définies par le code génétique, décrypté dans les années 1960 (Caskey et al., 1968; Crick et al., 1961; Nirenberg et al., 1965). La lecture du code génétique repose beaucoup sur les ARNt, des ARN non-codants dont la fonction est de reconnaître les codons de l'ARNm et de leur attribuer l'aa correspondant (Giegé, 2008). La structure des ARNt (Fig. 9), strictement conservée à travers l'évolution, est au cœur du mécanisme de reconnaissance des codons. En effet elle place l'anticodon, une séquence de trois nt complémentaires au codon à traduire, dans la boucle d'une structure en épingle à cheveux, le rendant accessible et favorisant les interactions d'appariement avec les ARNm.

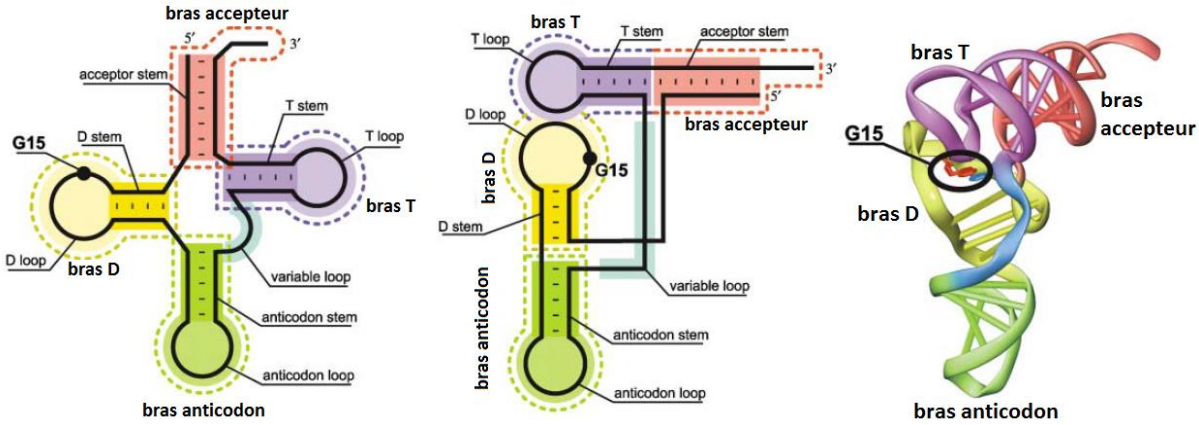


Figure 9 – Structure d'un ARNt
De gauche à droite : schéma en deux dimensions de la structure secondaire classique d'un ARNt avec trois bras formés par des tige-boucles (« stem-loop »); schéma en deux dimensions du repliement tertiaire classique d'un ARNt permis par l'interaction entre les boucles des bras T et D ; structure tridimensionnelle d'un ARNt. Le bras anticodon est toujours représenté libre d'interagir avec son codon spécifique, tandis que le bras accepteur pourra être chargé d'un aa à son extrémité 3'-CCA. D'après (Ishitani et al., 2003)

À l'autre extrémité de la structure tridimensionnelle des ARNt, qui se trouve également correspondre à l'extrémité 3' de la structure primaire de ces ARN, on retrouve les nt CCA auxquels un aa correspondant au codon reconnu par l'ARNt sera chargé. Puisque 20 aa sont majoritairement utilisés pour la synthèse de protéines, on considère qu'il existe 20 familles d'ARNt alloaccepteurs, c'est-à-dire chargés avec un aa différent ; au sein de chaque famille les ARNt sont dits isoaccepteurs puisqu'ils sont chargés par le même aa. On peut en outre signaler l'exception de l'ARNt formyl-méthionine ($ARNt^{fMet}$), reconnaissant spécifiquement des codons de démarrage de la traduction (voir **Introduction I.B.2.a**), bien qu'on puisse l'assimiler à la

famille des ARNt méthionine ($ARNt^{Met}$). Une autre exception existe aussi pour l'ARNt seleno-cystéine ($ARNt^{Sec}$), reconnaissant certains codons UGA atypiques. Habituellement non-sens, ces codons se distinguent par la présence immédiatement en 3' sur l'ARNm de séquences spécifiques et d'une structure en tige-boucle, qui sont autant de signaux pour leur traduction en seleno-cystéine (Berg et al., 1991; Heider et al., 1992). Ce dernier cas élargit d'ailleurs d'une manière intéressante la notion de code génétique, à l'image du fait que certains codons puissent intervenir dans la reconnaissance des codons qui les flanquent (Bossi and Roth, 1980).

Il est intéressant de remarquer ici que s'il existe (essentiellement) 20 aa, l'ensemble des combinaisons de codons possibles est égale à 4^3 (codon de trois nt avec un choix de quatre nt différents, A U G et C, pour chaque position), soit 64 en tout, dont 61 sont des codons sens. Cela s'explique par le fait que le code génétique est redondant : plusieurs codons différents peuvent être traduits par le même aa (on dit alors qu'ils sont synonymes). On compte par exemple six codons serine, deux codons isoleucine ou encore un seul codon tryptophane. Le nombre de gènes d'ARNt est encore plus élevé, avec chez *E. coli* 86 ARNt différents. En effet, plusieurs ARNt différents peuvent reconnaître le même codon, tandis qu'un même ARNt peut parfois reconnaître plusieurs codons synonymes. Cela est notamment dû à une propriété de l'anticodon des ARNt, portant des nt modifiés permettant une reconnaissance ambigüe de nt situés en troisième position des codons, appelée aussi position Wobble. La découverte d'au moins quatre modifications possible pour la position Wobble complémentaire sur l'anticodon (Inosine, m^5C, queuosine et mcm^3s^2U) permet ainsi d'affiner à l'extrême la spécificité de reconnaissance de codons par les ARNt (Tuorto and Lyko, 2016). L'existence de cette redondance du code génétique et d'un si grand nombre de gènes d'ARNt peut constituer un moyen de réguler l'expression génétique, comme cela sera abordé par la suite (voir **Introduction II.C.1.c**).

Enfin, les aminoacyl-ARNt synthétases (aa-RS pour « aminoacyl-tRNA synthase ») sont d'une grande importance dans la maintenance du code génétique. En effet, ces enzymes reconnaissent un ARNt donné et établissent une liaison ester entre sa séquence 3'-CCA et l'aa spécifié par la séquence anticodon. L'aa restera ainsi fixé à son ARNt sous forme d'aminoacyl-ARNt (aa-ARNt) jusqu'à ce qu'il soit inclus dans un polypeptide lors du processus de

traduction. L'ARNt peut alors être recyclé et reconnu à nouveau par une aa-RS pour qu'un nouvel aa y soit fixé. Il est donc primordial que l'aa-RS soit d'une grande spécificité pour fixer le bon aa à l'ARNt correspondant, afin d'éviter des erreurs lors de la traduction. Cette spécificité des aa-RS repose en partie et logiquement sur la séquence anticodon de l'ARNt, ainsi que sur leur capacité de correction (Giegé et al., 1998; Guo and Schimmel, 2012; XiaoLong and EnDuo, 2013).

d. Ribosomes et ARN ribosomiques

La traduction repose sur l'un des plus importants complexes ribonucléoprotéiques connus, présent dans les cellules de tous les organismes vivants : le ribosome. Ce dernier est le moteur de la traduction ; il réunit et interagit avec l'ensemble des principaux acteurs et est responsable de la réaction de liaison peptidique décrite précédemment (Fig. 7). Il est constitué de nombreuses protéines et des ARNr, qui sont non-codants à l'image des ARNt, et qui possèdent des propriétés structurales et catalytiques indispensables à l'activité du ribosome. Le ribosome se décompose en petite et grande sous-unité, nommées respectivement 30S et 50S chez les bactéries du fait de leur coefficient de sédimentation sur gradient de saccharose. Une fois assemblées, ces sous-unités forment le ribosome entier ou 70S.

La structure tridimensionnelle du ribosome bactérien et de ses sous-unités a été entièrement résolue au début des années 2000, permettant des avancées considérables dans la compréhension des mécanismes de la traduction. La structure de la sous-unité 30S (Fig. 10) a notamment été résolue grâce à la diffraction des rayons X sur des cristaux du ribosome de *Thermus thermophilus* (*T. thermophilus*) (Clemons et al., 1999; Schluenzen et al., 2000; Wimberly et al., 2000). En effet, les ARN et les protéines de cette bactérie thermophile sont très stables, car adaptés évolutivement pour la résistance aux hautes températures ; il est ainsi plus facile d'obtenir leur cristallisation. La sous-unité 30S du ribosome inclue l'ARNr 16S, un ARN fortement structuré (Fig. 10.A et 10B) qui forme le principal échafaudage de cette sous-unité et auquel se fixent les protéines de cette petite sous-unité (Fig. 10.C et 10.D), nommées S1 à S22 (S pour «Small subunit»).

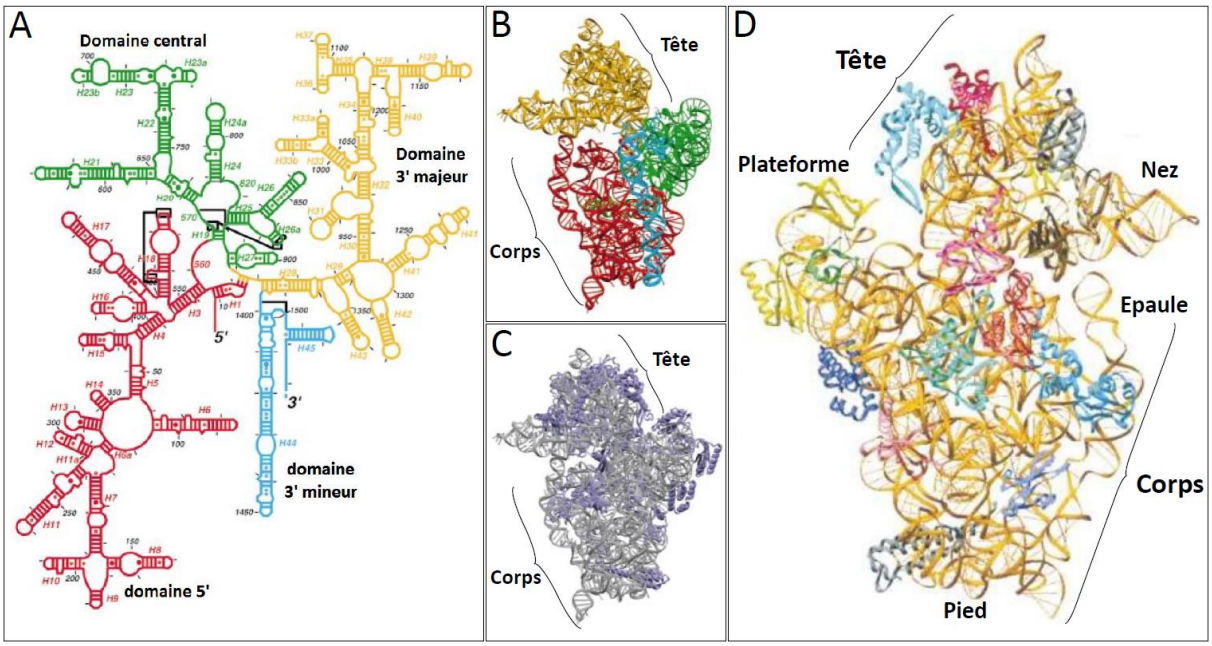


Figure 10 – Structure de la sous-unité 30S du ribosome bactérien
A. Structure secondaire de l'ARNr 16S, séparé en quatre domaines de couleurs différentes. B. Structure tridimensionnelle de l'ARNr 16S, respectant le même code couleur qu'en A. C. Structure tridimensionnelle de face de la sous-unité 30S du ribosome. L'ARNr 16S est représenté en gris, et les protéines ribosomiques qui y sont associées, en violet. D. Même structure de dos, avec l'ARNr 16S en doré et les protéines ribosomiques chacune d'une couleur différente. D'après (Clemons et al., 1999; Wimberly et al., 2000).

La sous-unité 50S du ribosome correspond à la grande sous-unité du ribosome bactérien et comprend les ARNr 23S et 5S, auxquels s'ajoutent plus de 33 protéines ribosomiques numérotées L1 à L33 (L pour « Large subunit »). Si la petite sous-unité interagit directement avec l'ARNm, c'est la grande sous-unité qui établit la liaison peptidique entre deux aa successifs tout en les dissociant de leur ARNt. La structure détaillée de la sous-unité 50S du ribosome bactérien a également été résolue dans les années 2000, en faisant cette fois-ci soit usage de la machinerie traductionnelle de *T. thermophilus* (Yusupov et al., 2001), soit de celle de la bactérie mésophile *Deinococcus radiodurans* (*D. radiodurans*) (Harms et al., 2001) (Fig. 11).

Enfin, la structure du ribosome entier en association avec un ARNm et un ou plusieurs ARNt a également été élucidée. Cette structure met en évidence de nombreuses interactions entre les deux sous-unités du ribosome et avec l'ARNm, ainsi que le tunnel emprunté par l'ARNm au sein du ribosome et les sites A, P et E occupés par les ARNt (Myasnikov et al., 2005; Yusupov et al., 2001; Yusupova et al., 2001) (voir Fig. 12).

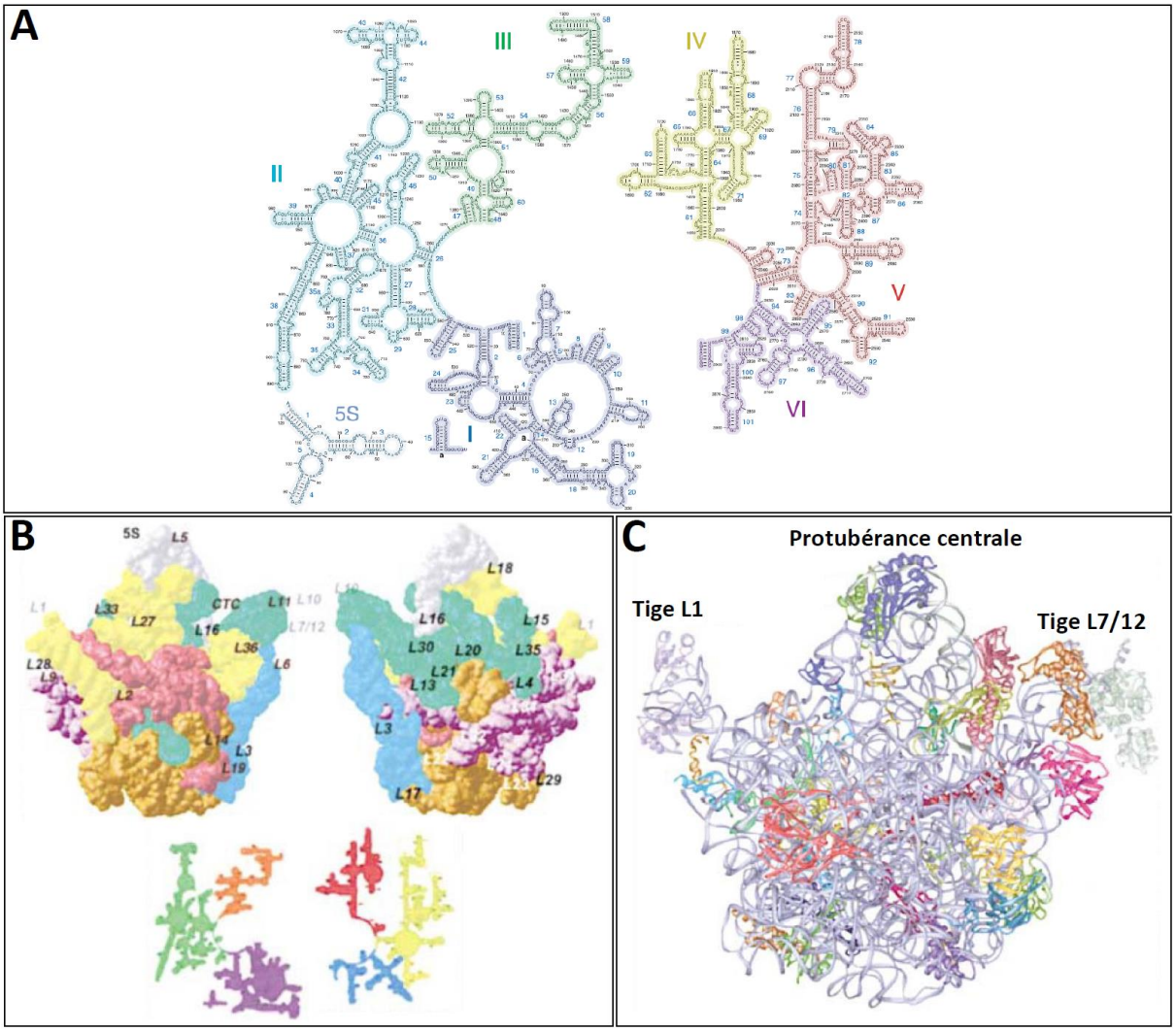


Figure 11 – structure de la sous-unité 50S du ribosome bactérien
A. Structures secondaires des ARNr 5S et 23S de T. thermophilus. La structure secondaire de l'ARNr 23S est composée de six domaines. B. Structure tridimensionnelle de l'ARNr 23S de la grande sous-unité du ribosome de D. radiodurans avec les positions des différentes protéines de la grande sous-unité du ribosome. En dessous, le schéma de la structure secondaire de l'ARNr 23S indique le code-couleur utilisé pour la représentation des six domaines. C. Structure tridimensionnelle détaillée de la sous-unité 50S de D. radiodurans avec les ARNr 23S et 5S en gris et les protéines ribosomiques en couleurs. D'après (Harms et al., 2001; Yusupov et al., 2001).

2. Démarrage de la traduction

Cette partie s'articulera autour des constituants ribosomiques et non ribosomiques nécessaires au démarrage de la traduction, une étape limitante primordiale pour la synthèse des protéines et dont les mécanismes ont été étudiés avec une extrême minutie dans la littérature.

a. Sous-unité ribosomique 30S et complexe de pré-démarrage

Le démarrage de la traduction nécessite la formation d'un complexe de pré-démarrage de la traduction, réunissant un certain nombre de facteurs protéiques et d'ARN qui seront détaillés ici. Chez les bactéries, cette étape de la traduction repose notamment sur la sous-unité 30S du ribosome.

L'une des clés du démarrage de la traduction se trouve dans la séquence de l'ARNr 16s de la petite sous-unité du ribosome. On trouve ainsi à l'extrémité 3' de cet ARN une séquence d'une grande importance pour la formation du complexe de pré-démarrage de la traduction, à savoir la séquence anti-Shine Dalgarno (anti-SD). Celle-ci fait immédiatement référence à la séquence SD citée précédemment et faisant partie intégrante des ARNm (Fig. 8). L'anti-SD permet une reconnaissance spécifique du site de démarrage de la traduction par la sous-unité 30S, par simple appariement avec la séquence SD de l'ARNm (Jacob et al., 1987; Shine and Dalgarno, 1974). Le degré de complémentarité entre les séquences anti-SD et SD, variable d'un ARNm à un autre, est un déterminant crucial de l'affinité de la sous-unité 30S du ribosome pour un ARNm donné. En découlent la stabilité du complexe de pré-démarrage de la traduction et la fréquence à laquelle cet ARNm va être traduit en protéine. Les séquences SD et anti-SD sont donc d'une grande importance pour l'expression génétique chez les bactéries, et représentent une partie de ce que l'on appelle le site de fixation du ribosome ou « ribosome binding site » (RBS, voir **Introduction II.C.1** et Fig. 18).

Hormis le degré de complémentarité entre SD et anti-SD, l'accessibilité de la séquence SD – c'est-à-dire l'absence de structure secondaire de l'ARNm au niveau du RBS, régit également la capacité du ribosome à reconnaître son site de fixation (de Smit and van Duin, 1994) (voir **Introduction II.C.2.c**). Par ailleurs, le codon de démarrage lui-même varie d'un

messager à un autre, et si les codons de démarrage AUG sont majoritaires, la traduction peut démarrer à des codons GUG, UUG, voire même AUC (Chalut and Egly, 1995), AUA (Romero and García, 1991) ou encore AUU pour le gène *infC* codant le facteur de démarrage IF3 (Butler et al., 1986; Tedin et al., 1999). D'une manière générale, un codon de démarrage AUG est bien plus favorable à une forte expression génétique que les autres (Ringquist et al., 1992; Romero and García, 1991). Un point de vue intéressant sur la présence de codons de démarrage alternatifs pour certains gènes est que cette caractéristique a pu être favorisée évolutivement par un avantage sélectif lié à la régulation gènes en question (voir **Introduction II.C.1.c**). Dans d'autres cas, une dérive génétique a pu avoir lieu, et des codons alternatifs ont pu apparaître simplement parce qu'une expression faible des gènes considérés était suffisante à la survie normale des bactéries.

Un autre déterminant d'une grande importance pour la formation du complexe de pré-démarrage de la traduction est la distance entre le SD et le codon de démarrage de la traduction (Chen et al., 1994; Gualerzi and Pon, 2015). Cela découle du fait que le codon de démarrage de la traduction est un élément de l'ARNm reconnu par le complexe formé par la sous-unité 30S avec l'ARNt^{fMet} de démarrage sous sa forme aminoacylée (fMet-ARNt^{fMet}). En effet, ce dernier a ceci de particulier qu'il n'interagit pas avec le ribosome sous sa forme entière 70S contrairement aux autres ARNt. Au lieu de cela, il interagit spécifiquement avec la sous-unité 30S, au niveau d'une région appelée site P (pour « peptidyl-tRNA ») – alors qu'un ARNt aminoacylé classique interagit avec le site A (pour « aminoacyl-tRNA »). On note que cette interaction entre site P et fMet-ARNt^{fMet} semble pouvoir avoir lieu avant ou après la reconnaissance d'un SD par la petite sous-unité du ribosome (Gold, 1988; Gualerzi and Pon, 2015). Cette interaction positionne la boucle anticodon de l'ARNt^{fMet} à faible distance de l'anti-SD de l'ARNr 16S, d'où l'existence d'une distance optimale entre SD et codon de démarrage de la traduction sur l'ARNm, d'environ 6 à 10 nt (Ringquist et al., 1992).

En résumé, le complexe de pré-démarrage de la traduction est composé de la sous-unité 30S du ribosome en association avec l'ARNm au niveau du RBS, et de l'ARNt^{fMet} aminoacylé en interaction avec le site P du ribosome et avec le codon de démarrage de la traduction de l'ARNm. La formation de ce complexe s'observe d'ailleurs assez aisément *in*

vitro, en l'absence d'autres facteurs de démarrage de la traduction, par de simples expériences de RT que l'on nomme « toeprints » (Hartz et al., 1988) (voir **Revue n°1**).

b. Autres facteurs nécessaires au démarrage de la traduction

Plusieurs protéines ribosomiques et facteurs de démarrage (IF pour « Initiation factor ») sont nécessaires à la séquence d'évènements aboutissant à la formation du complexe de démarrage de la traduction ; cette partie du propos leur sera consacrée.

Chez les bactéries, il existe trois facteurs de démarrage connus à ce jour : IF1, IF2 et IF3 (Simonetti et al., 2009), dont la structure en interaction avec la sous-unité 30S du ribosome a été résolue (Hussain et al., 2016; Simonetti et al., 2008) (Fig. 12). L'une des fonctions importantes de ces facteurs concerne l'incorporation de l'ARNt de démarrage au complexe de démarrage de la traduction. En effet, cet ARNt de démarrage nécessite une étape de maturation après son aminoacylation. Il est aminoacylé par un résidu méthionine classique par la méthionyl ARNt synthase, et ce résidu doit donc subir l'ajout d'un groupement formyl avant de pouvoir être intégré au complexe de pré-démarrage de la traduction ; c'est le rôle de la méthionyl-ARNt formyltransférase (Kozak, 1983; RajBhandary, 1994; Ramesh et al., 1997). En l'absence de cette modification, l'ARNt de démarrage ne peut être reconnu par le facteur IF2, ce qui empêche son incorporation au site P de la sous-unité 30S. La littérature semble d'ailleurs s'accorder sur le fait qu'IF2 interagit de prime abord avec la sous-unité 30S ou avec le fMet-ARNt^{fMet}, tandis qu'un cas de figure où l'ARNt de démarrage interagirait d'abord avec la sous-unité ribosomique serait plus rare (Milon et al., 2010; Tsai et al., 2012).

Par la suite, IF3 permet d'assurer que le bon ARNt de démarrage est parvenu au site P et participe à la reconnaissance du codon de démarrage par l'anticodon, puis aux changements de conformation nécessaires à la poursuite du processus de démarrage de la traduction. Le rôle particulier d'IF3 dans la discrimination des codons de démarrage de la traduction a été relevé de nombreuses fois dans la littérature (Meinnel et al., 1999; Sussman et al., 1996; Tedin et al., 1999) tandis qu'IF1 occupe le site A de la sous-unité 30S et fournit des sites de d'ancrage à IF2 et IF3. Ces facteurs de démarrage entrent donc chacun en interaction avec la sous-unité 30S pour le bon déroulement du processus de démarrage (Gualerzi and Pon, 2015; Hussain et al., 2016; Simonetti et al., 2009) (Fig. 12). Il convient de

spécifier que l'importance du groupement formyl de l'ARNt de démarrage et du point de contrôle effectué par les trois facteurs IF tient notamment au fait qu'ils empêchent l'allongement précoce de la traduction.

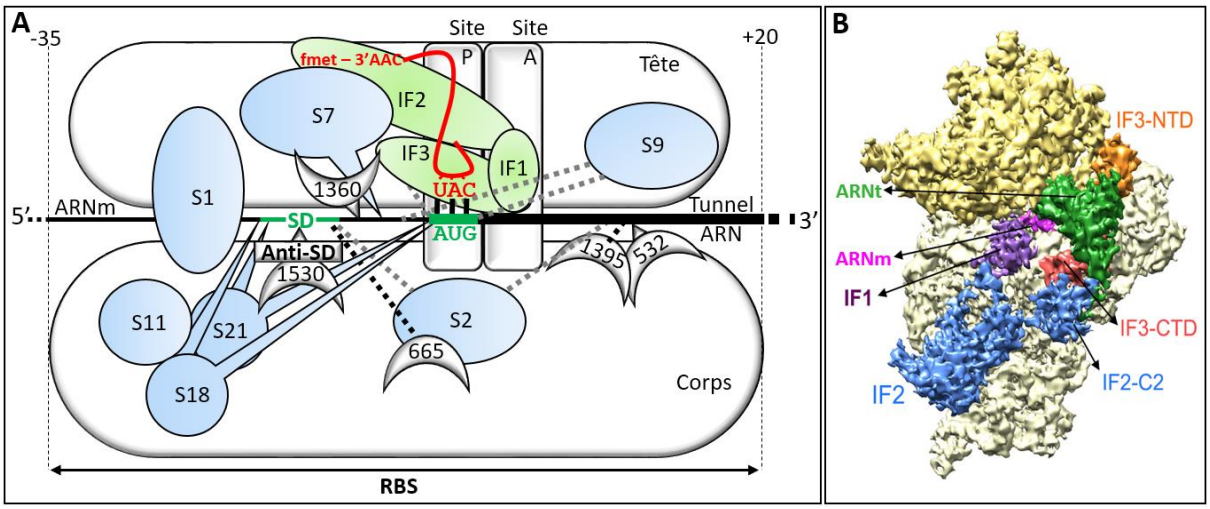


Figure 12 – Complexe de pré-démarrage de la traduction
A. Schéma des interactions entre divers composants du complexe de pré-démarrage de la traduction avec l'ARNm. Le corps et la tête de la sous-unité 30S du ribosome sont représentés avec certaines protéines ribosomiques qui y sont associées. L'ARNm est situé au niveau du tunnel ARN, avec les séquences représentatives du RBS indiquées en vert. Le fmet-ARNt^{fMet} de démarrage est représenté en rouge, au niveau du site P de la sous-unité 30S, avec sa boucle anticodon appariée au codon AUG de démarrage de la traduction. Les demi-lunes correspondent à des régions de l'ARNr 16S (voir Fig. 10.A) en interaction avec l'ARNm d'après des réactions de pontage (La Teana et al., 1995), où un triangle correspond à une interaction forte et les tirets noirs à une interaction plus faible. Il en va de même pour les protéines ribosomiques représentées en bleu, et dont les interactions fortes avec diverses régions du RBS sont indiquées par des triangles bleus, tandis que les interactions plus faibles sont représentées par des tirets gris. Pour la protéine S1, son interaction avec l'ARNm en amont du SD et sa position est déduite de (Sengupta et al., 2001). Enfin, les facteurs de démarrage de la traduction sont représentés en vert : IF1 au niveau du site A de la sous-unité 30S, IF2 en interaction avec IF1 et le groupement fMet de l'ARNt de démarrage, et IF3 à proximité du codon de démarrage de la traduction et en interaction avec les deux autres facteurs. Remarque : ce schéma correspond à une projection en deux dimensions, et les échelles ne sont pas respectées. D'après (Gualerzi and Pon, 2015; La Teana et al., 1995). B. Structure tridimensionnelle du complexe de prédémarrage de la traduction associé aux IF. Les différents domaines des facteurs IF sont ici mis en évidence. NTD : « N-terminal domain » ; CTD : « C-terminal domain » ; C2 : domaine d'interaction d'IF2 avec le groupement fmet de l'ARNt de démarrage. D'après (Hussain et al., 2016).

Certaines protéines ribosomiques ont également une fonction importante dans le démarrage de la traduction ; c'est notamment le cas de la protéine S1. Il s'agit de la plus grande protéine ribosomique, mais aussi de celle dont le lien avec le démarrage de la traduction est le plus évident, puisqu'elle stabilise l'interaction entre l'ARNm et la sous-unité 30S en interagissant avec les deux (Boni et al., 1991). Effectivement, S1 est en mesure d'interagir avec une séquence de 11 nt en amont de la séquence SD (Sengupta et al., 2001).

Plus récemment, son rôle dans la dénaturation de structures d'ARN à proximité du SD a été mis en évidence, permettant une meilleure reconnaissance par la sous-unité 30S (Duval et al., 2013). Des expériences de pontage (ou « crosslinking ») (La Teana et al., 1995), ainsi que les nombreuses études structurales précédemment évoquées, ont dévoilé de multiples interactions entre l'ARNm et d'autres protéines ribosomiques (Fig. 12.A).

c. Sous-unité ribosomique 50S et formation du complexe de démarrage de la traduction

L'assemblage entre les sous-unités 30S et 50S est une étape nécessaire pour former un ribosome complet. Ce n'est qu'après la formation du complexe de pré-démarrage de la traduction que le recrutement de la sous-unité 50S devient possible du fait de la dissociation d'IF3 puis d'IF1 du reste du complexe (Fabbretti et al., 2007) (Fig. 13). La relocalisation de l'ARNt de démarrage dans une position plus adéquate à l'interaction avec la sous-unité 50S est importante, et requiert l'activité GTPase d'IF2 (Antoun et al., 2003; Grigoriadou et al., 2007), qui consiste en l'hydrolyse d'une molécule de guanosine triphosphate (GTP). S'ensuit le relargage du facteur IF2, et c'est un facteur d'allongement de la traduction, nommément EF-Tu, qui va entrer en interaction avec le ribosome et amener l'aa-ARNt^{aa} spécifié par le deuxième codon de l'ARNm au site A (Voorhees and Ramakrishnan, 2013; Yamamoto et al., 2013), déclenchant le processus d'allongement de la traduction. L'ensemble des étapes résumant la formation des complexes de pré-démarrage et de démarrage de la traduction sont représentées dans la figure 13.

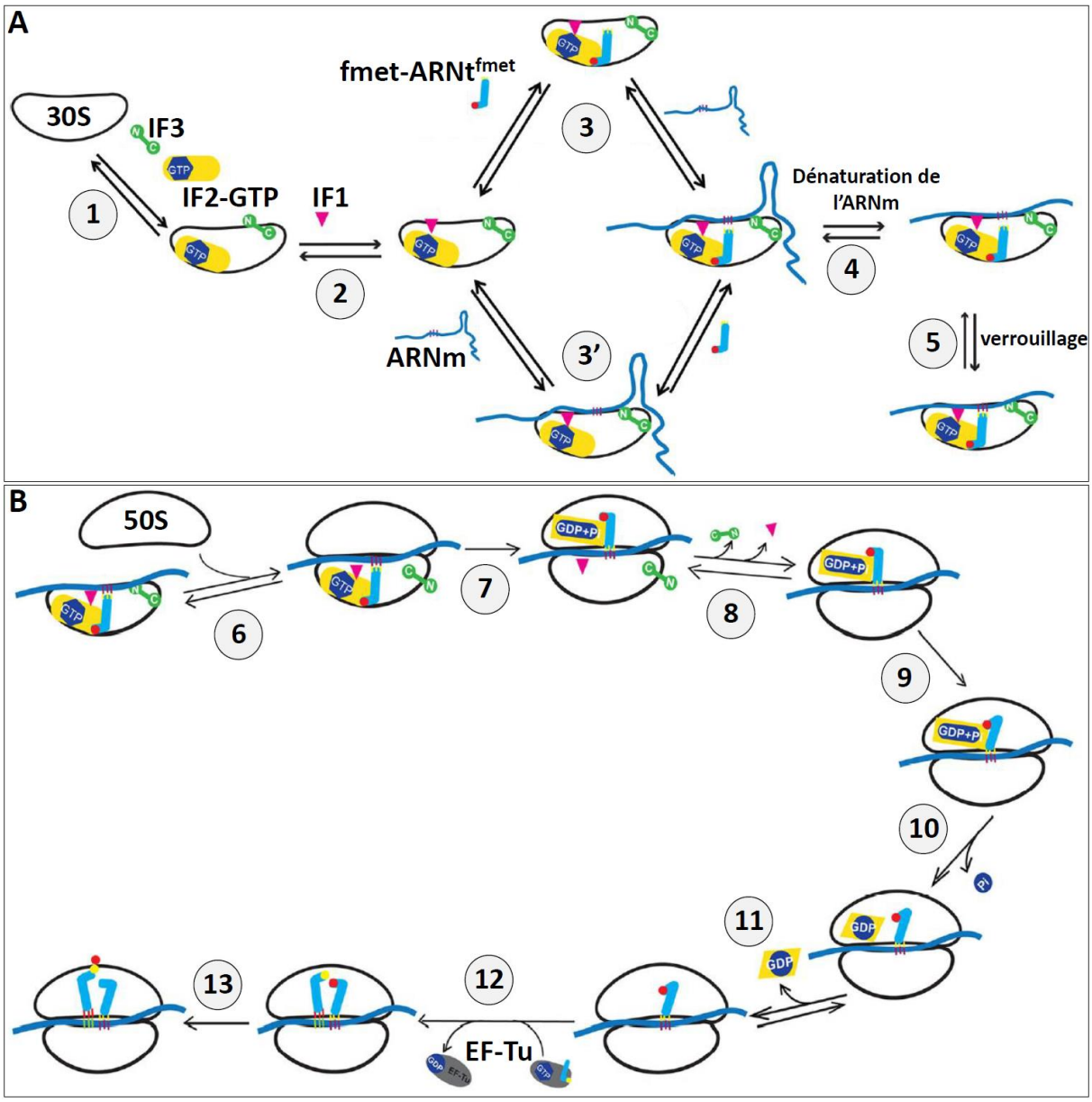


Figure 13 – Démarrage de la traduction bactérienne

Principales étapes de la formation des complexes de pré-démarrage (A) et de démarrage (B) de la traduction. Brièvement, les facteurs IF2 et IF3 (1) puis IF1 (2) se fixent à la sous-unité 30S ; l'ARNt de démarrage est recruté par IF2 et l'ARNm entre en interaction avec la sous-unité 30S sans ordre préférentiel (3, 3') ; les structures de l'ARNm sont défaites par l'activité hélicase du ribosome (4) et le complexe de pré-démarrage est verrouillé dans cette configuration quand IF3 ne décèle pas de problème de correspondance et de conformation entre les différents composants du complexe de pré-démarrage (5). La sous-unité 50S peut ensuite être recrutée (6), coïncidant avec le relargage d'IF3 (7) puis d'IF1 (8) ; IF2 positionne correctement l'ARNt de démarrage via l'hydrolyse d'une molécule de GTP + phosphate en GDP + pyrophosphate (10), puis IF2 est détaché du complexe (11) ; l'allongement de la traduction peut alors commencer avec l'intervention de EF-Tu qui participe au recrutement au site A de l'aa-ARNt^{aa} décodant le deuxième codon de l'ARNm (12), suivie par la formation de la liaison peptidique entre le fmet de l'ARNt de démarrage et l'aa du second ARNt, une réaction catalysée par le ribosome (13). Adapté de (Gualerzi and Pon, 2015).

3. Allongement et terminaison de la traduction

a. Allongement de la traduction : synthèse des polypeptides

Le ribosome 70S, l'ARNm, les ARNt, ainsi que les deux facteurs d'allongement EF-Tu et EF-G, constituent la machinerie nécessaire à l'allongement de la traduction bactérienne. Au moment où un cycle d'allongement commence, un aa-ARNt^{aa} vient d'être amené au site A par EF-Tu (Gualerzi and Pon, 2015; Kavaliauskas et al., 2012), tandis qu'un peptidyl-ARNt est situé au site P. Le ribosome est alors en conformation pré-déplacement (PRE) (ou « pre-translocation ») (Yamamoto et al., 2013), et va catalyser la formation d'une liaison peptidique en transférant la chaîne polypeptidique liée au peptidyl-ARNt du site P vers l'extrémité N-terminale de l'aa-ARNt^{aa} du site A. Ce dernier devient ainsi le nouveau porteur de l'ensemble de la chaine polypeptidique néoformée. C'est le facteur EF-G qui permet ensuite le passage de la conformation PRE en conformation post-déplacement (POST) (ou « post-translocation »), en utilisant l'énergie fournie par l'hydrolyse d'une molécule de GTP (Ratje et al., 2010; XiaoLong and EnDuo, 2013).

Le déplacement (ou « translocation ») en question correspond au décalage des ARNt des site A et P respectivement vers les sites P et E (E pour « exit ») du ribosome (Ratje et al., 2010), ce dernier étant situé près des protéines S7 et L1 (Yusupov et al., 2001). Ce mouvement s'accompagne également du déplacement de l'ARNm d'un codon vers le site E en suivant le mouvement des ARNt, c'est-à-dire dans la direction A-P-E ; un nouveau codon peut alors être décodé au niveau du site A. L'ARNt qui a été désacylé avant le changement de conformation PRE-POST est donc maintenant situé au site E et va être exclu du complexe, laissant le ribosome prêt pour l'ajout d'un nouvel aa-ARNt^{aa} au site A par EF-Tu et pour un nouveau cycle d'allongement. En résumé, chaque cycle d'allongement aboutit à la lecture d'un nouveau codon de l'ARNm, à l'addition d'un aa à l'extrémité C-terminale de la chaîne polypeptidique et au relargage d'un ARNt désacylé (Fig. 14).

D'autres facteurs d'allongement tels que EF-4 et BipA sont trouvés chez les bactéries. S'ils ne semblent pas indispensables à la traduction en conditions optimales de croissance, ils semblent avoir un rôle plus important à jouer lors de conditions de stress telles qu'un pH acide, de basses températures ou une carence en nutriments (Ero et al., 2016). EF-4 a ainsi été impliqué dans un mécanisme de correction des erreurs de traduction et permet de

débloquer des ribosomes dont le cycle d'allongement ne s'est pas déroulé correctement (Qin et al., 2006) (Fig. 14), un phénomène plus courant en conditions de stress.

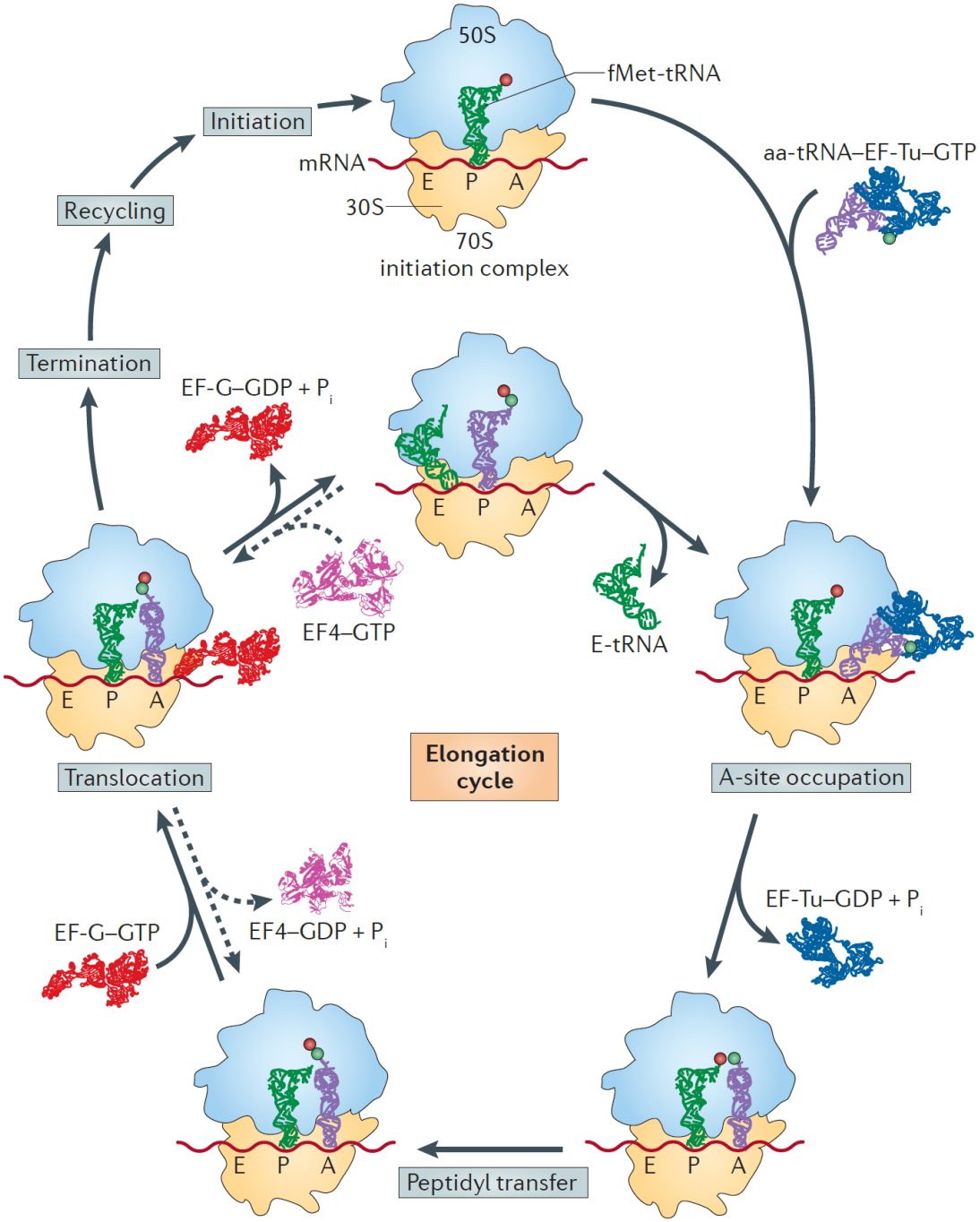


Figure 14 – Allongement de la traduction bactérienne
Représentation schématique d'un cycle d'allongement de la traduction inclus dans le processus complet de traduction. Chacun des principaux intervenants est représenté par sa structure tridimensionnelle hormis le ribosome, et les aa sont représentés par des billes de couleurs différentes. Issu de (Yamamoto et al., 2013).

On peut noter plusieurs propriétés intéressantes de l'allongement de la traduction. Ainsi, le même ARNm est le plus souvent traduit par une succession de ribosomes, formant ce que l'on appelle des polysomes, et permettant la synthèse de plusieurs copies d'une protéine à partir d'une molécule d'ARNm. Il est aussi intéressant de remarquer que les ribosomes peuvent être amenés à effectuer des pauses lors de la traduction, notamment lorsqu'ils rencontrent des séquences semblables à des SD (ou séquences « SD-like ») (Li et al., 2012). Les structures secondaires au sein de la séquence codante des ARNm sont d'ailleurs une autre source de pauses lors de l'allongement de la traduction (Wen et al., 2008), et permettraient parfois un adressage correct de certaines protéines vers la membrane pendant la traduction (Fluman et al., 2014).

Une autre propriété, retrouvée uniquement chez les bactéries, est la traduction cotranscriptionnelle. Il faut en effet savoir que le démarrage de la traduction bactérienne ne requiert pas que l'ARNm soit entièrement transcrit, et que transcription et traduction ont lieu toutes les deux dans le cytoplasme. Un complexe de pré-démarrage de la traduction peut ainsi se former dès lors que le RBS de l'ARNm a été transcrit et émerge du tunnel de sortie de l'ARNpol. Il est d'ailleurs aujourd'hui admis que le premier ribosome à traduire un ARNm détermine la vitesse de transcription par l'ARNpol, et empêche celle-ci de faire marche arrière (« backtracking ») à l'occasion de pauses dans la transcription (Proshkin et al., 2010).

b. Terminaison de la traduction et recyclage des ribosomes

La dernière étape du processus de traduction est la terminaison, aboutissant à la libération du polypeptide néo-synthétisé de son interaction avec le ribosome 70S, puis à la dissociation les deux sous-unités du ribosome afin qu'elles puissent être réutilisées pour la traduction d'autres ARNm. La fin de la séquence codante de l'ARNm doit également être libérée pour que les ribosomes suivants du polysome puissent y accéder. Chez les bactéries on connait à ce jour trois facteurs de terminaison de la traduction (RF pour « release factor ») et un facteur de recyclage (RRF pour « ribosome recycling factor ») dont les rôles seront décrits ici.

Lorsqu'un codon non-sens se trouve au site A du ribosome, il n'est pas reconnu par un ARNt mais par un facteur de terminaison. Ainsi, le codon UAA (ocre) sera reconnu par RF1 ou

RF2, tandis que le codon UAG (ambre) sera reconnu uniquement par RF1 et le codon UGA (opale) uniquement par RF2. L'un ou l'autre facteur va alors rompre la liaison covalente entre le polypeptide complet et l'extrémité 3'-CCA du peptidyl-ARNt situé au site P (Caskey et al., 1977; Kisselev and Buckingham, 2000; Pallesen et al., 2013). C'est ensuite RF3 qui va interagir avec la protéine ribosomique L12, puis libérer RF1 ou RF2 du site A du ribosome avant de s'en séparer lui-même, *via* l'hydrolyse d'une molécule de GTP (Freistroffer et al., 1997; Pallesen et al., 2013).

Il faut alors l'intervention de RRF couplée à la fixation d'EF-G à proximité des sites A et P, ainsi que l'hydrolyse d'une nouvelle molécule de GTP, pour provoquer la dissociation des sous-unités 30S et 50S (Karimi et al., 1999; Peske et al., 2005; Sternberg et al., 2009; Zavialov et al., 2005). La sous-unité 30S reste toutefois associée à l'ARNt du site P, encore apparié à l'avant-dernier codon de la phase ouverte de lecture de l'ARNm. C'est IF3 qui va permettre le recyclage de cet ARNt et de la sous-unité 30S (Karimi et al., 1999). Il s'agit là d'une étape importante car IF3 empêche la réassociation entre les deux sous-unités, qui pourrait bloquer indéfiniment les ribosomes en fin de traduction. La petite sous-unité du ribosome est alors généralement libre de se dissocier de l'ARNm.

Toutefois, une particularité intéressante des ARNm chez les bactéries est qu'ils peuvent être polycistroniques, c'est-à-dire porteurs de plusieurs ORF se suivant sur la molécule d'ARNm. Dans ce cas de figure, on considère que lorsque la traduction d'une ORF s'achève, la sous-unité 30S peut scanner l'ARNm pour trouver la région de démarrage du cistron suivant au lieu de s'en dissocier. Il a d'ailleurs été démontré que la dissociation du ribosome 70S elle-même n'est pas forcément requise, au moins lorsque le codon de démarrage du cistron suivant à traduire se trouve à proximité directe du codon de terminaison du cistron qui vient d'être traduit (Yamamoto et al., 2016). Cela peut d'ailleurs être une façon d'exprimer et réguler uniformément des gènes faisant partie d'un même complexe ou d'une même voie de biosynthèse, qui se retrouvent souvent sur un même opéron transcrit en un ARNm polycistronique.

Chapitre II : Expression génétique et sa régulation par l'ARN chez les entérobactéries

Cette partie s'articulera autour de l'expression génétique et de sa régulation, qui est fondamentale pour que les gènes adéquats soient exprimés dans des conditions où leurs produits sont nécessaires à la cellule, et inversement pour que soient réprimés les gènes dont les produits sont inutiles voir délétères dans ces conditions. Une attention particulière sera accordée à la variété des mécanismes de régulation reposant sur des ARN, surtout au niveau post-transcriptionnel. En effet, si des facteurs non-ARN sont également impliqués et seront évoqués ici, je m'étendrais davantage sur les ARN et leurs fonctions régulatrices, qui sont plus directement liés au travail effectué durant cette thèse. Seront notamment décrits les riboswitches et les structures secondaires de l'ARNm, de même que les ARN régulateurs agissant par appariement de séquence. Je ne m'attarderai pas ici sur les répercutions physiologiques de ces régulations, qui seront abordées plus en détails par la suite (voir **Introduction III**).

A. Transcription et régulation transcriptionnelle

La transcription est la première étape de l'expression génétique et correspond à la synthèse d'ARN à partir d'une matrice ADN par une ARNpol. Cette dernière démarre la transcription au niveau d'un promoteur et l'achève au niveau d'un terminateur. Ici seront détaillés les éléments régulant cette transcription.

1. Régulateurs protéiques de la transcription

a. Facteurs interagissant avec l'ARN polymérase

La plupart des gènes possèdent au moins un promoteur défini, qui consiste généralement en une séquence de nt reconnue par l'ARNpol (Fig. 15). Plus précisément, chez les bactéries des facteurs sigma (σ) protéiques s'associent à l'ARNpol, et reconnaissent chacun des séquences consensus différentes (Österberg et al., 2011). Ces séquences consensus sont le plus souvent composées de deux « boîtes » de 6 nt situées respectivement à 35 nt et 10 nt en amont du site de démarrage de la transcription (TSS pour « Transcription Start Site »). On

retrouve les deux séquences de 6 nt idéalement espacées de 17 nt pour une fixation optimale de l'ARNpol à l'ADN. En dehors du facteur σ70 (aussi appelé facteur σ de ménage), exprimé de façon constitutive en phase exponentielle de croissance, ces facteurs σ sont le plus souvent exprimés ou libérés d'une interaction inhibitrice avec une protéine anti-σ en réponse à des stress (voir **Introduction II.A.2.a et III.B.2**).

Ainsi, puisque différents promoteurs possèdent des affinités différentes pour les facteurs σ, chaque promoteur peut être reconnu par un ou plusieurs facteurs σ. Certains gènes possèdent d'ailleurs plusieurs promoteurs, dont chacun peut dépendre d'un facteur σ spécifique. Ces facteurs sont des activateurs transcriptionnels de l'expression génétique, dans la mesure où l'ARNpol va transcrire spécifiquement les gènes dont elle reconnaît les promoteurs. En fonction des conditions environnementales, les facteurs σ associés à l'ARNpol vont donc varier, entraînant une modification de l'ensemble de gènes transcrits – ou transcriptome. Il est à noter que d'autres protéines sont capables d'entrer en interaction avec l'ARNpol indépendamment des facteurs σ et peuvent également drastiquement changer le transcriptome de la cellule, à l'image de la protéine DksA (Ross et al., 2016).

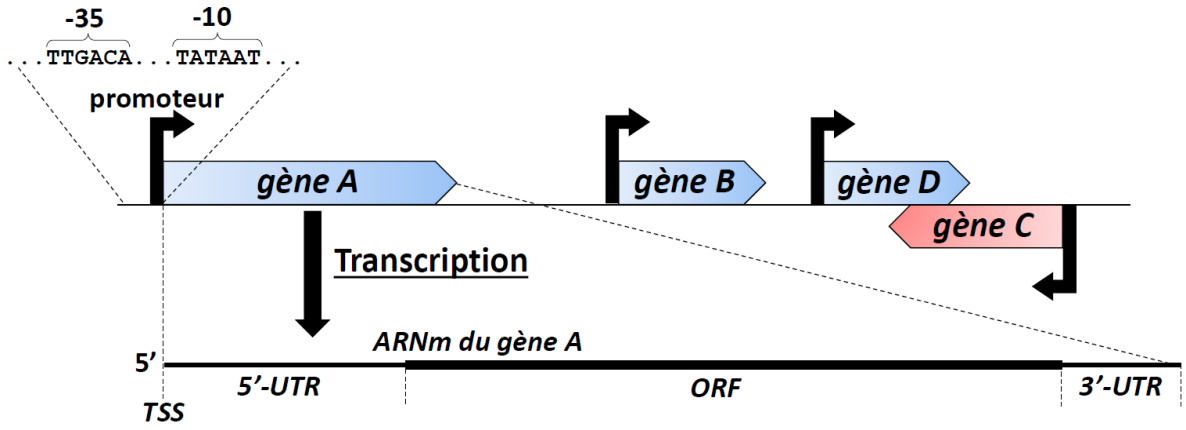


Figure 15 – Promoteur et transcription
Représentation schématique d'un locus chromosomique portant quatre gènes. Chaque gène est représenté dans la direction dans laquelle il est transcrit depuis la matrice ADN. Les flèches noires articulées correspondent aux promoteurs de ces gènes ; le détail du promoteur du gène A est représenté ici, avec les séquences -35 et -10 consensus reconnues par σ70. La numérotation -35 et -10 correspond à la distance par rapport au TSS.

b. Facteurs de transcription

Certains régulateurs transcriptionnels interagissent principalement avec l'ADN, le plus souvent à proximité des promoteurs des gènes dont ils régulent l'expression. On parle ici de

facteurs de transcription (FT) qui reconnaissent souvent des séquences consensus à l'instar des facteurs σ. Cependant, leur mécanisme d'action peut aussi bien amener à une régulation positive que négative de l'expression génétique (Browning and Busby, 2016). En effet, dans certains cas il s'agit de répresseurs transcriptionnels qui peuvent occlure le site de fixation de l'ARNpol aux boîtes -35 et -10, ou le rendre inaccessible en imposant une torsion à la molécule d'ADN. Dans d'autres cas ils peuvent au contraire rendre les promoteurs plus accessibles, ou encore participer directement au recrutement de l'ARNpol. Leur rôle ainsi que celui des facteurs sigma dans la réponse aux stimuli environnementaux sera plus longuement abordée dans les parties **Introduction III.B et III.C**.

2. Régulateurs ARN de la transcription

Si les ARN régulateurs connus à ce jour agissent dans leur grande majorité au niveau post-transcriptionnel, il est intéressant de s'attarder sur ceux opérant au niveau transcriptionnel, afin de percevoir au mieux l'étendue de la diversité des mécanismes à travers lesquels les ARN peuvent réguler l'expression génétique.

a. ARN 6S : un ARN mimant un ADN

L'ARN 6S a été le premier petit ARN non-codant identifié (Banerjee et al., 1969), mais sa fonction dans la régulation transcriptionnelle n'a été élucidée que trente ans plus tard (Wassarman and Storz, 2000), alors que la science vivait un grand bouleversement lié à la découverte du rôle de nombreux petits ARN régulateurs dans le contrôle de l'expression génétique, et ce dans les trois domaines du vivant. Conservé chez les bactéries, cet ARN 6S est aujourd'hui l'un des mieux connus. On sait ainsi qu'il séquestre des ARNpol associées à σ70 et permet l'expression de gènes dépendants d'autres facteurs σ. On sait ainsi qu'en phase stationnaire, où il est le plus abondant, l'ARN 6S séquestre les ARNpol associées à σ70 et permet ainsi l'expression de gènes dépendants d'autres facteurs σ comme, par exemple, σS (Cavanagh et al., 2008; Trotochaud and Wassarman, 2005). On remarque ici la subtilité de ce mécanisme, puisque la structure et la séquence du 6S ont évolué pour imiter celles d'un promoteur σ70-dépendant, dès lors reconnues par les ARNpol associées à σ70. Cette séquestration est levée dès lors que les cellules rentrent à nouveau en phase exponentielle de croissance, avec la dégradation de l'ARN 6S. (Pour plus de détails : voir **Revue n°2**).

b. Riboswitches et terminaison transcriptionnelle

Il existe deux mécanismes principaux *via* lesquels la transcription d'un ARN peut s'achever, impliquant des terminateurs de transcription. D'une part, 20 à 30% des gènes bactériens ont des terminateurs Rho-dépendants et dépendent de l'intervention du facteur protéique Rho, capable de se fixer à l'ARN et de déloger l'ARNpol lors d'une pause transcriptionnelle (Boudvillain et al., 2013). D'autre part, on trouve dans la majorité des gènes chez les bactéries des terminateurs Rho-indépendants (Fig. 16). Ces régions situées à l'extrémité 3' des ARN forment une structure secondaire en tige-boucle et fonctionnent de concert avec une séquence poly-U en aval de la tige-boucle, qui forme un hybride instable avec la séquence poly-A du brin ADN utilisé comme matrice par l'ARNpol, permettant le détachement de l'ARNpol (Greive and von Hippel, 2005).

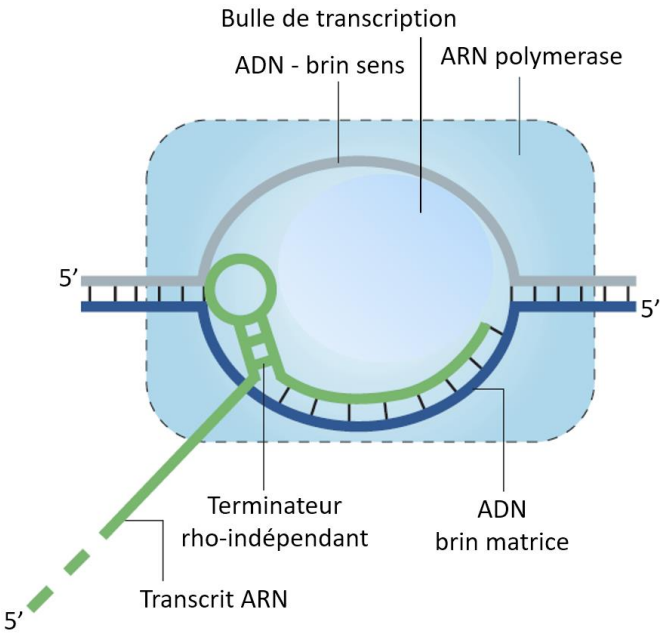


Figure 16 – Terminateur transcriptionnel Rho-indépendant
L'ARN est représenté en vert et apparié avec le brin matrice de l'ADN en bleu ici. Cet appariement implique un hybride entre une séquence ARN poly-U et une séquence ADN poly-A. La bulle de transcription correspond ici à la portion d'ADN qui ne forme pas une double hélice entre le brin sens et le brin matrice, car ce dernier est occupé par l'ARNpol et l'ARN néo-synthétisé. D'après (Greive and von Hippel, 2005).

Dans un grand nombre de cas la terminaison de la transcription est un phénomène non régulé. Cependant, certains gènes ont la capacité de moduler leur propre expression en tirant avantage de l'aptitude de certaines structures ARN à mettre prématurément fin à la transcription – on appelle ce mécanisme l'atténuation transcriptionnelle (Naville and

Gautheret, 2010). Cela peut se faire *via* l'action de riboswitches, des senseurs ARN détectant la présence de molécules spécifiques, telles que des métabolites ou des ions, et s'y liant via une région aptamère. Cette interaction va entraîner le changement de structure d'une seconde région, la plateforme d'expression. Ce changement de structure (ou « switch ») peut notamment faire apparaître un terminateur précoce de la transcription ou au contraire d'un anti-terminateur perturbant la structure d'un terminateur préexistant, permettant ainsi la régulation génétique (Serganov and Nudler, 2013).

L'un des premiers riboswitches régulateur de la transcription à avoir été décrit est le riboswitch présent dans le 5'UTR de l'ARNm polycistronique *ribGBAHT* de *B. subtilis* (Mironov et al., 2002). Cet opéron code les enzymes de la voie de biosynthèse de la riboflavine (ou vitamine B2), entrant dans la composition du dinucléotide flavine-adénine (FAD). Ainsi le riboswitch inclus dans le 5'UTR de l'ARNm change de structure en réponse au FAD pour former un terminateur précoce de transcription, prévenant la synthèse de flavine quand le produit final FAD utilisé par la cellule est présent en quantité suffisante, et formant une boucle de rétrocontrôle négatif. Cette régulation permettrait donc à la cellule d'économiser à la fois la synthèse des enzymes de la voie de biosynthèse de la flavine et la synthèse de la flavine elle-même, et d'utiliser plus efficacement les ressources dont elle dispose.

Il existe par ailleurs chez les entérobactéries des exemples de riboswitches provoquant l'apparition d'un terminateur rho-dépendant, c'est-à-dire en l'occurrence une structure d'ARN reconnue par Rho (Hollands et al., 2012). D'autres riboswitches peuvent néanmoins avoir des effets régulateurs post-transcriptionnels (voir **Introduction II.C.2.a**), tandis que l'atténuation ne se limite pas aux effets de riboswitches, et peut notamment impliquer des changements de structures d'ARNm répondant à la fixation de facteurs protéiques ou ARN (Naville and Gautheret, 2010).

c. Régulateurs transcriptionnels agissant par appariement de séquence : les ARN antisens

Certains ARN non codants peuvent avoir un effet régulateur transcriptionnel du fait de leur localisation par rapport au gène régulé. C'est le cas des ARN antisens (asRNA) transcrits à partir des brins complémentaires aux brins ADN matrices d'ARNm (Thomason and Storz,

2010). Ces asRNA sont très répandus du fait de la transcription cryptique ou « envahissante » (« pervasive ») à partir de nombreux locus semblables à des promoteurs. Si leurs mécanismes d'action restent généralement mal compris, plusieurs mécanismes ont été proposés et étudiés. La configuration des promoteurs convergeant peut par exemple provoquer un phénomène d'interférence transcriptionnelle où les ARNpol convergeant l'une vers l'autre lors de la synthèse de l'ARNm ou de l'asRNA entrent en collision (Fig. 17.A). C'est ce qui a été observé avec les promoteurs convergents *pL* et *pR*, déterminant respectivement les phases lysogénique ou lytique du bactériophage 186 infectant *E. coli* (coliphage) (Callen et al., 2004). En effet, c'est l'orientation des promoteurs qui provoque la régulation du promoteur le plus faible (*pL*) par le plus fort (*pR*), tandis qu'une expression des deux gènes à partir de loci différents ne permet plus d'observer cette régulation.

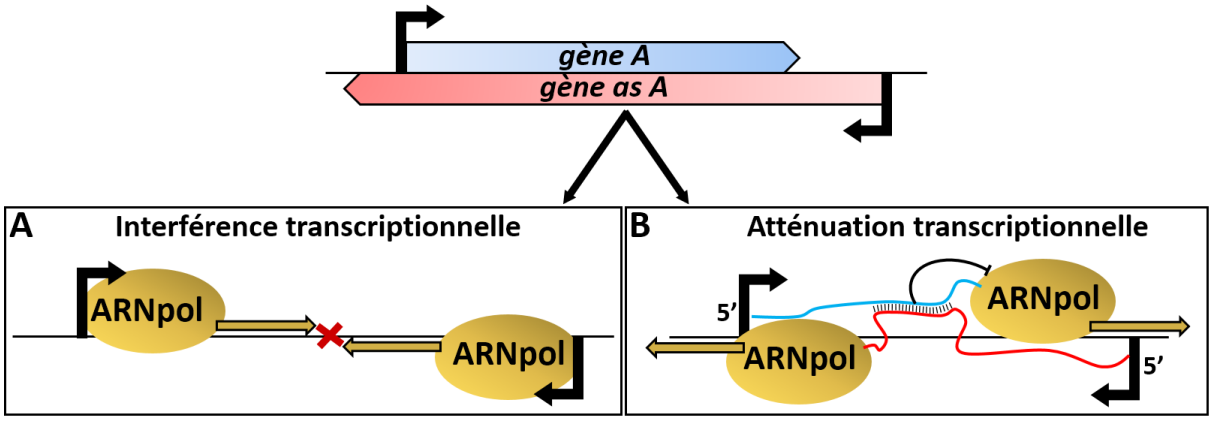


Figure 17 – Régulation transcriptionnelle par des asRNA
Le gène régulé est représenté en bleu, avec un gène antisens (as) en rouge exprimé depuis le brin ADN complémentaire. Cette configuration génétique peut entraîner une interférence transcriptionnelle (A), avec une collision entre les deux ARNpol convergeant l'une vers l'autre. Cela peut aussi aboutir au phénomène d'atténuation transcriptionnelle (B), avec l'asRNA (en rouge) s'appariant à l'ARNm néo-synthétisé (en bleu), modifiant ainsi la structure de l'ARNm pour former un terminateur arrêtant la transcription par l'ARNpol. Les flèches jaunes indiquent le sens de la transcription.

Mais ce n'est pas le seul mécanisme de régulation par les asRNA, qui peuvent également interagir directement avec l'ARNm du fait de leur complémentarité parfaite. Cela peut notamment entrainer un effet d'atténuation transcriptionnelle semblable à celui de certains riboswitches, en modifiant la structure secondaire de l'ARNm (Fig. 17.B), à l'image de l'asRNA RNAIII avec l'ARNm plasmidique *repR* chez *B. subtilis* (Brantl et al., 1993). Du fait de cette complémentarité parfaite, les asRNA peuvent également avoir des effets post-transcriptionnels qui seront abordés par la suite (voir **Introduction II.C.3.a**).

Ce regard sur les divers modes de régulation transcriptionnelle de l'expression génétique laisse déjà entrevoir les nombreux mécanismes de régulation impliquant les ARN. Une considération intéressante vis-à-vis des régulations transcriptionnelles par l'ARN est qu'elles sont nécessairement apparues après la transcription elle-même au cours de l'histoire évolutive. Cette catégorie d'ARN régulateurs a donc vraisemblablement évolué au fur et à mesure que la transcription devenait un mécanisme de plus en plus important au sein des organismes. Des ARN préexistants ont ainsi pu acquérir de nouvelles fonctions régulatrices, tandis que d'autres ont pu voir le jour à partir de promoteurs plus ou moins cryptiques à proximité de gènes codants, à l'instar de certains asRNA. Les ARN actuels ne sont donc pas cantonnés aux fonctions héritées des origines de la vie et sont également évolutivement adaptés à de nouvelles fonctions, notamment régulatrices, au sein des cellules.

B. Régulation post-transcriptionnelle par les protéines

Si beaucoup de mécanismes de régulation post-transcriptionnelle dépendent de régulateurs ARN (décrits dans la partie **Introduction II.C**), les premiers régulateurs post-transcriptionnels découverts ont été des protéines. J'évoquerai ici quelques exemples de protéines impliquées dans la régulation post-transcriptionnelle des gènes, à savoir CsrA, mais aussi les RNases qui jouent un rôle primordial dans la dégradation et la stabilité des ARN.

1. CsrA, un régulateur post-transcriptionnel protéique

Le gène *csrA* a été décrit suite à un crible génétique réalisé afin de rechercher des facteurs impliqués dans la régulation de la biosynthèse du glycogène (Romeo et al., 1993). Les effets pléiotropiques de mutants du gène *csrA*, conservé chez la plupart des bactéries, a conduit à de nombreuses études sur le fonctionnement de la protéine CsrA. Ainsi CsrA se fixe aux ARNm, souvent (mais pas uniquement) au niveau du RBS du fait de l'affinité de CsrA pour des séquences proches du SD. C'est par exemple le cas avec les ARNm *glgCAP* et *cstA* chez *E. coli*, pour lesquels la fixation de CsrA au RBS empêche la fixation du ribosome et entraîne une répression post-transcriptionnelle (Baker et al., 2002; Dubey et al., 2003). Notons que les ARN non-codants CsrB, CsrC et leurs divers homologues trouvés chez de nombreuses bactéries exprimant CsrA régulent l'activité de CsrA. Plus précisément, ces ARN régulateurs fonctionnent comme des « éponges » en séquestrant de nombreuses protéines CsrA, via de

multiples sites de fixations (Liu et al., 1997; Timmermans and Van Melderen, 2010). Ce concept d'ARN « éponges » fonctionnant par la séquestration de régulateurs sera abordé plus en détails dans la partie **Introduction II.C.3.d**. D'autres ARN régulateurs ciblent également la protéine CsrA pour en réguler l'activité, à l'image de McaS, qui a la particularité de pouvoir également s'apparier à des ARNm (Jorgensen et al., 2013). (Pour plus de détails concernant CsrA, voir la **Revue n°2**)

2. RNases : la stabilité et la dégradation des ARNm

Comme évoqué auparavant, la nature chimique des ARN les rend particulièrement vulnérables au clivage et à la dégradation par des RNases, et ce dans les trois domaines du vivant. Si dans un certain nombre de cas, ces RNases permettent la maturation et la synthèse des formes fonctionnelles d'ARN non-codants tels que les ARNr et les ARNt, elles sont également impliquées dans la régulation de l'expression génétique. En effet, elles sont notamment responsables de la demi-vie relativement courte des ARNm chez les bactéries.

Ainsi la RNase E, essentielle et conservée chez la plupart des bactéries, catalyse le clivage de nombreux ARNm à des régions simple-brin et plutôt riches en nt A-U (Mackie, 2012). Ces ARNm sont dégradés par la RNase E notamment lorsqu'ils sont moins traduits et moins protégés par les ribosomes, mais aussi dans de nombreux cas suite à un appariement avec un ARN régulateur (voir **Revue n°1**). De même, la RNase III (conservée dans les trois domaines du vivant) est impliquée dans de nombreuses réactions de clivages d'ARNm, cette fois-ci dans des régions d'ARN double-brin (Court et al., 2013). Le clivage catalysé par la RNase III peut également dépendre de l'appariement d'un ARN régulateur à l'ARNm (voir **Revue n°1**). À condition que ces clivages ne soit pas constitutifs mais plutôt dépendants de la présence d'un sRNA ou d'un facteur protéique, ils peuvent entraîner une régulation négative ou positive selon les cas.

C. Régulation post-transcriptionnelle par les ARN

Pendant et après la transcription d'un ARNm de nombreuses modifications structurales de l'ARNm peuvent altérer l'accessibilité de son RBS, ou encore sa stabilité, constituant autant de régulations post-transcriptionnelles de l'expression génétique. Une

particularité notable des effets sur l'accessibilité du RBS est qu'en l'absence de ribosomes traduisant l'ARNm, celui-ci peut rapidement devenir la cible de RNases et être dégradé (Radhakrishnan and Green, 2016), d'où un effet régulateur d'abord sur la traduction, puis sur la stabilité des ARNm. Si le rôle de certaines protéines dans la régulation post-transcriptionnelle sera abordé, cette partie sera surtout consacrée aux nombreuses régulations médiées par des ARN. En effet, dès sa transcription l'ARNm devient une cible privilégiée pour des appariements aussi bien intramoléculaires qu'intermoléculaires avec des ARN régulateurs.

1. Séquences, RBS atypiques et rôles de la protéine ribosomique S1

Comme évoqué précédemment, la capacité du ribosome à démarrer correctement la traduction dépend en grande partie du RBS de l'ARNm, et notamment de l'appariement entre séquences SD et anti-SD. Toutefois, plusieurs autres séquences et régions de l'ARNm ont été caractérisées pour leur importance pour l'expression génétique, notamment parce qu'elles ont un effet positif sur le démarrage de la traduction en comparaison à des ARNm dépourvus de ces séquences. S'il ne s'agit pas de régulation à proprement parler, il est tout à fait envisageable (et parfois même observé) que ces séquences et RBS atypiques soient l'objet de régulations spécifiques, c'est pourquoi leurs particularités seront abordées ici.

a. Séquences « enhancer » et site de fixation de la protéine S1

La séquence epsilon a été identifiée pour la première fois au sein du gène 10 du coliphage T7. On y trouve, ainsi que dans de nombreux autres gènes de ce phage, une séquence riche en nt A-U directement en amont du SD, activant d'un facteur 100 la synthèse protéique quand elle est positionnée devant un gène rapporteur *lacZ* (Olins and Rangwala, 1989). A l'origine les auteurs supposaient que cette séquence epsilon permettait une interaction ARNm-ARNr similaire à celle observée avec la séquence SD du fait d'une complémentarité de séquence avec une partie de la région 437-497 de l'ARNr 16S (Fig. 12.A tige-boucle « H17 »). Il a cependant été démontré depuis que l'activation de la traduction était indépendante de l'appariement prédit (O'Connor and Dahlberg, 2001).

Parallèlement, des séquences poly-U en amont du SD de gènes du coliphage Qβ, assimilables aux séquences epsilon, ont été caractérisées en tant que sites de fixation de la protéine ribosomique S1 (Boni et al., 1991). Par la suite, cette interaction entre la protéine S1 et une région de 11 nt en amont du SD, souvent riche en nt A-U, s'est révélée être un phénomène communément retrouvé pour les ARNm bactériens (Hook-Barnard et al., 2007; Nafissi et al., 2012; Sengupta et al., 2001) (Fig. 18). La disponibilité de la protéine S1 permet donc la régulation epsilon-dépendante de multiples ARNm chez les bactéries.

Une considération importante concerne l'hypothèse que plus la séquence SD de l'ARNm est proche de la séquence consensus complémentaire à l'anti-SD, plus efficace serait la synthèse protéique. Ce serait ignorer qu'après la fixation du ribosome, celui-ci doit être en mesure de se défaire de cet appariement afin de procéder à l'allongement de la traduction. Il semblerait ainsi que le site de fixation de la protéine S1 joue ici un rôle d' « enhancer » – ou accélérateur – de la traduction. En effet, l'interaction entre S1 et cette séquence epsilon déstabiliserait l'interaction SD/anti-SD pour permettre un allongement de la traduction plus efficace (Komarova et al., 2002; Takahashi et al., 2013).

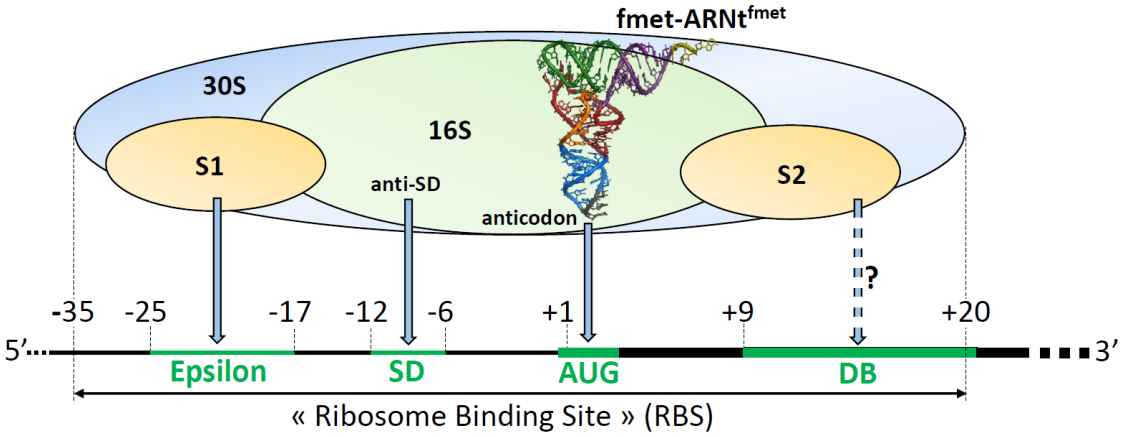


Figure 18 – Les diverses régions d'un RBS bactérien
Représentation schématique d'un ARNm en noir, et des différentes régions du RBS permettant le démarrage de la traduction en vert. Les différentes protéines et ARN interagissant avec ces régions du RBS sont représentées en complexe avec la sous-unité 30S du ribosome (en bleu), avec les protéines ribosomiques (en jaune), l'ARNr 16S (en vert pâle) et l'ARNt de démarrage (multicolore). Ces interactions sont représentées par des flèches bleues, pointillées quand elles sont hypothétiques. Les longueurs et positions des séquences epsilon et DB sont ici inspirées de (Sprengart et al., 1996).

Une autre séquence au rôle d'« enhancer » similaire à celui des séquences epsilon, dite séquence « downstream box » (DB) et située en aval du codon de démarrage de la traduction (Fig. 18), a été identifiée dans le gène *0.3* du coliphage T7 (Sprengart et al., 1990).

Ce gène contient l'un des RBS les plus forts trouvés chez *E. coli* et ses phages, permettant de parasiter efficacement la machinerie traductionnelle d'*E. coli*. À l'instar de ce qui a été proposé pour la séquence epsilon, les auteurs ont suggéré que cette séquence située entre 15 et 26 nt en aval du codon de démarrage pouvait s'apparier aux nt 1469 à 1483 de l'ARNr 16S, avant que cette hypothèse ne soit infirmée par une autre équipe (O'Connor et al., 1999). A ce jour le mécanisme impliqué n'est pas connu, et une interaction avec la protéine S1 est certainement exclue du fait de la position de la séquence DB en aval du codon de démarrage. On peut cependant supposer qu'une autre protéine ribosomique connue pour son interaction avec l'ARNm puisse être impliquée ici. Cela pourrait être le cas de la protéine S2 dont un mutant a déjà été associée à l'activation traductionnelle par la séquence DB du gène *cI* chez le prophage lambda (Shean and Gottesman, 1992), et qui pourrait établir un contact avec cette région de l'ARNm (La Teana et al., 1995) (Fig. 12.A, Fig. 18).

b. RBS atypiques

La sous-unité 30S du ribosome doit être à même de reconnaître le RBS d'un ARNm pour s'y fixer. Cependant il existe des ARNm portant un RBS atypique, où il semble manquer certains éléments – or si la synthèse protéique à partir de ces ARNm peut être affectée, elle reste généralement possible et parfois même efficace. Ces cas particuliers seront abordés ici.

Il existe ainsi de nombreux exemples d'ARNm curieusement dépourvus de séquence SD canonique : c'est le cas de plus de 15% des gènes chez les gamma-protéobactéries dont font partie les entérobactéries (Omotajo et al., 2015). Etonnamment, la seule condition qui semble nécessaire à l'expression normale de ces gènes semble être l'absence de structure au niveau du codon de démarrage de la traduction (Scharff et al., 2011). Dans d'autres cas, ce sont les séquences epsilon (Nafissi et al., 2012) et DB (Sprengart et al., 1996) qui peuvent pallier l'absence de séquence SD. Il a aussi été démontré qu'un RBS modulaire pouvait parfois être formé par une combinaison de structures de l'ARNm, réunissant des morceaux de séquences SD éparses dans la région 5'UTR de l'ARNm avec le codon de démarrage de la traduction. C'est ce qui est observé pour le gène *rpsA*, codant la protéine ribosomique S1 (Boni et al., 2001; Skorski et al., 2006). L'une des particularités intéressantes de ce système est que la protéine S1 elle-même peut se fixer à son propre ARNm quand elle est en excès dans la cellule, provoquant un changement de structure de l'ARNm qui prévient la formation de ce

RBS modulaire. De cette manière, la protéine S1 exerce un rétrocontrôle négatif sur sa propre synthèse.

Il existe un cas encore plus extrême de RBS atypique, avec les ARNm que l'on appelle « leaderless » ou sans 5'UTR. Ces derniers peuvent notamment être produits par la toxine MazF, une endoribonucléase clivant les ARNm en 3' de séquences ACA (Vesper et al., 2011) et souvent à proximité du codon de démarrage, les séparant de ce fait de leurs séquences SD tout comme d'éventuelles séquences epsilon. De manière surprenante, les auteurs de cette étude ont démontré que MazF catalyse également un clivage à l'extrémité 3' des ARNr 16S les privant de séquence anti-SD, et concluent que la population de sous-unité 30S ainsi générée est capable de démarrer spécifiquement la traduction des ARNm sans 5'UTR. Cela permettrait à la cellule de largement et rapidement modifier son profil de synthèse protéique – ou protéome – dans les conditions où MazF est exprimée (Sauert et al., 2016).

c. Fréquence et rôles des codons

Comme expliqué dans la partie **Introduction I.B.1.c**, il existe une forte redondance du code génétique ainsi qu'un grand nombre d'ARNt, plusieurs pouvant décoder le même codon. Or il existe une corrélation forte entre l'abondance des différents ARNt et la fréquence des codons qu'ils reconnaissent (Ikemura, 1981). Certains codons sont ainsi appelés codons rares car ils sont peu utilisés à travers le génome d'une espèce donnée. Ils peuvent avoir un impact important sur l'expression des gènes qui en contiennent dans la mesure où peu d'ARNt sont disponibles pour les décoder (Konigsberg and Godson, 1983).

Plus généralement, il est aujourd'hui admis qu'un biais dans l'usage de codons, qu'il soit dans l'usage de codons rares ou par exemple dans la répétition du même codon plusieurs fois d'affilée, peut contribuer à moduler l'expression génétique en imposant au ribosome des pauses (Quax et al., 2015) voire des glissements du cadre de lecture (Kim and Tinoco, 2016). Il a ainsi été établi récemment que la nature des codons d'une ORF a une influence plus forte sur le niveau d'expression des protéines que la formation de structures secondaires, à l'exception notable des 16 premiers codons, c'est-à-dire au niveau du démarrage de la traduction (Boël et al., 2016). Néanmoins, diverses structures secondaires d'ARNm peuvent

avoir une influence au niveau post-transcriptionnel. Certaines sont justement abordées dans le sous-chapitre suivant.

2. ARNm et structures régulatrices

a. Riboswitches et effets post-transcriptionnels

Si certains riboswitches régulent la transcription, beaucoup agissent au niveau post-transcriptionnel, et contrôlent fréquemment le démarrage de la traduction. En effet, l'accessibilité du RBS est déterminante pour initier correctement la traduction, et de nombreux riboswitches permettent des réarrangements de structure d'ARNm à même d'occulter ou au contraire de libérer le RBS afin que la sous-unité 30S du ribosome puisse le reconnaître et s'y fixer (Fig. 19.A). On trouve ainsi chez *E. coli,* au sein du 5'UTR des deux ARNm *thiM et thiC* codant des enzymes impliquées dans la synthèse de la thiamine (ou vitamine B_1), deux riboswitches similaires capables de lier la thiamine (Winkler et al., 2002). En réponse à une forte concentration en thiamine, ces riboswitches changent de conformation pour séquestrer le SD dans une structure secondaire. Cela entraîne une inhibition de la fixation du ribosome, et une régulation négative de l'expression de ces deux gènes.

Cependant, il existe une grande diversité de mécanismes de régulation post-transcriptionnelle par des riboswitches. Par exemple, certains riboswitches sont également des ribozymes, et sont capables de cliver l'ARNm dont ils font partie en réponse à une interaction avec leur ligand glucosamine-6-phosphate, induisant la dégradation de l'ARNm *via* l'action d'une RNase (Winkler et al., 2004) (Fig. 19.B). D'autres riboswitches se rapprochent des « self-splicing introns » évoqués précédemment, et permettent un épissage de l'ARNm dont ils font partie en réponse au ligand c-di-GMP, libérant ainsi un RBS pour la fixation du ribosome (Lee et al., 2010) (Fig. 19.C).

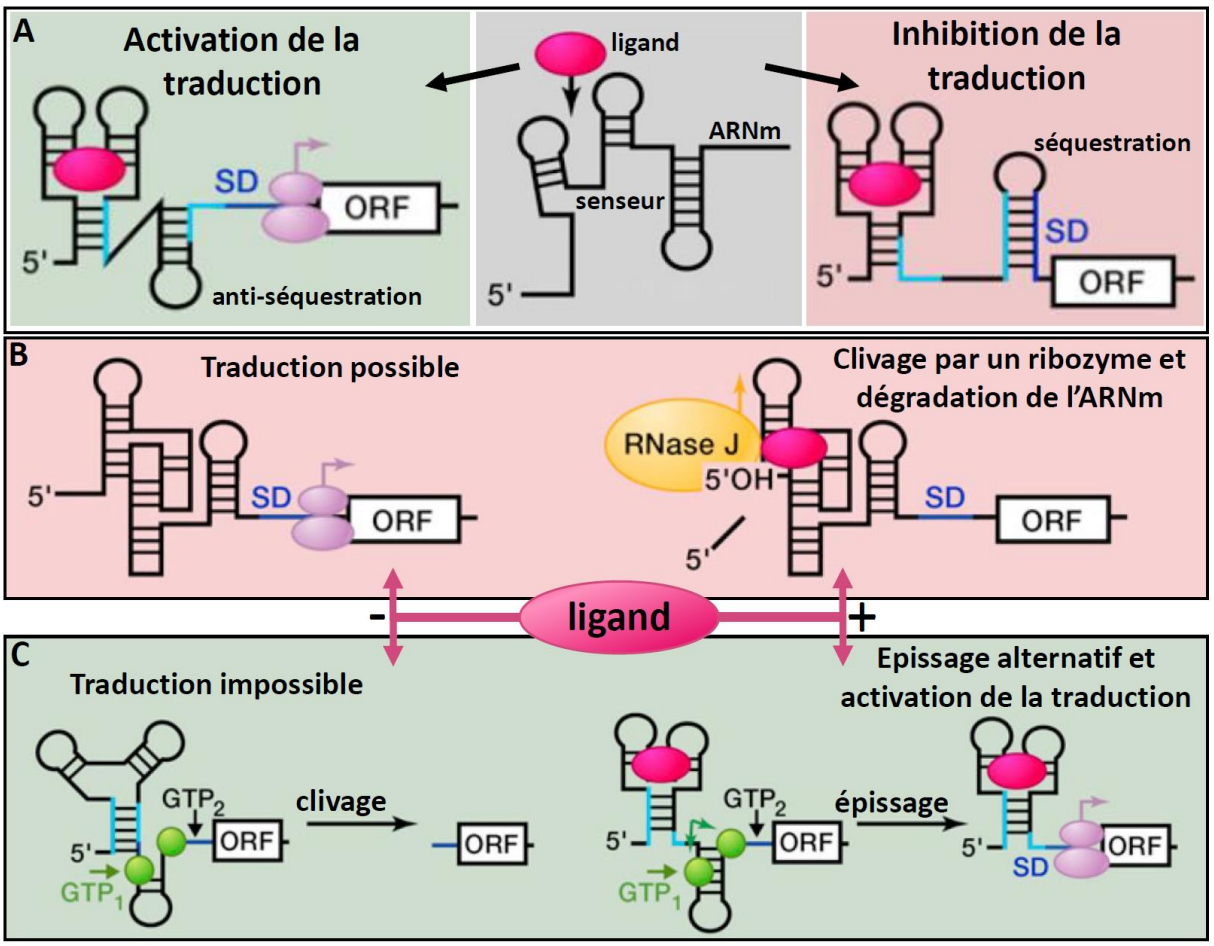


Figure 19 – Diversité des mécanismes de régulation post-transcriptionnelle par des riboswitches
Exemples de mécanismes de régulation post-transcriptionnelle impliquant des riboswitches. A. En interagissant avec le ligand, le riboswitch peut changer de structure pour séquestrer le SD ou le libérer d'une structure séquestrant le SD, entraînant respectivement une inhibition ou une activation de la traduction. B. Le ligand peut enclencher une activité ribozyme du riboswitch pour cliver l'ARNm, libérant une extrémité 5'-OH. L'ARNm peut alors être dégradé par une exonucléase telle que la RNase J chez les bactéries à Gram positif. C. En réponse à la fixation du ligand, le riboswitch peut modifier une activité catalytique préexistante pour générer un épissage alternatif, permettant un rapprochement entre une séquence SD et le codon de démarrage de la traduction et activant ainsi la traduction. D'après (Serganov and Nudler, 2013).

D'une façon intéressante, ces riboswitches sont souvent impliqués dans des boucles de rétrocontrôle négatif, à l'image de ce qui a été évoqué pour la régulation transcriptionnelle de l'opéron *rib* par un riboswitch répondant au FAD chez *B. subtilis*. Cela suggère que ces riboswitches ont été conservés car ils sont efficaces notamment pour réguler des voies de biosynthèse impliquant plusieurs enzymes, comme dans le cas de la synthèse de la thiamine ou du FAD, des vitamines ou dérivés de vitamines d'une grande importance pour les bactéries mais dont la production peut être coûteuse en ressources. Par ailleurs, leur conservation découle vraisemblablement également du fait qu'ils sont situés en *cis* des ARNm dont ils

régulent la transcription ou la traduction, ce qui augmente la probabilité qu'ils soient conservés de concert avec le restant du gène régulé.

On peut également souligner le fait que ces riboswitchs ont souvent une très grande affinité pour les ligands auxquels ils répondent. Cette caractéristique a certainement été d'une importance majeure au cours de l'évolution, car elle a pu permettre à des ARN catalytiques d'interagir très spécifiquement avec divers métabolites et ions nécessaires aux cellules, et ce avant même l'apparition des protéines. A bien des égards, il est possible de considérer ces riboswitch régulateurs de l'expression génétique comme une adaptation moderne d'ARN catalytiques auparavant impliqués plus directement dans des voies de biosynthèse. Néanmoins, une autre façon d'appréhender l'apparition de certains riboswitches est de considérer que les ARN peuvent très rapidement évoluer pour développer des affinités très spécifiques du moment qu'une pression de sélection appropriée leur est imposée. C'est d'ailleurs cette caractéristique qui est souvent exploitée pour la production d'aptamères – le plus souvent des ARN, ayant une grande affinité pour un ligand – *via* la technique de SELEX (pour « systematic evolution of ligands by exponential enrichment) (Gold et al., 1995).

b. Thermosenseurs

Les thermosenseurs sont parfois assimilés à des riboswitches dans la mesure où ce sont des éléments en *cis* d'ARNm répondant à un stimulus en changeant de structure, conduisant à une régulation de l'expression génétique. Ils en diffèrent néanmoins, du fait qu'ils n'interagissent avec aucun ligand, et sont au lieu de cela sensibles à la température (Nocker et al., 2001). Il faut porter ici une attention particulière au fait que toute structure d'ARN est sensible à la température et peut être dénaturée par la chaleur. Les thermosenseurs ont cela de particulier qu'ils sont sensibles à des températures très spécifiques, ayant une répercussion physiologique pour la bactérie. Par exemple, on comprendra qu'une augmentation de la température de 30°C à 37°C peut correspondre à l'entrée en contact avec un hôte humain pour une bactérie commensale ou pathogène. Bien entendu, d'autres températures peuvent avoir un sens physiologique pour des bactéries adaptées à d'autres milieux et soumises à des variations de température différentes.

Si certains thermosenseurs peuvent avoir un effet d'atténuation transcriptionnelle, d'autres agissent au niveau post-transcriptionnel. C'est par exemple le cas d'un élément en *cis* de l'ARNm *prfA* codant un facteur de virulence chez la bactérie à Gram positif *Listeria monocytogenes*, une bactérie pathogène pour l'être humain (Johansson et al., 2002). Les auteurs démontrent ainsi qu'une structure de cet ARNm entravant l'accès au RBS à 30°C est dénaturée à 37°C, permettant la traduction à cette température. Il pourrait s'agir d'un moyen de contrôler l'expression de ce facteur de virulence pour ne l'exprimer qu'à des températures signalant une potentielle interaction avec un hôte.

c. Autres structures ayant un impact sur l'expression génétique

D'autres éléments en *cis* de l'ARNm peuvent former des structures secondaires affectant l'expression génétique au niveau post-transcriptionnel. Si ces structures n'ont pas d'effet régulateur à elles-seules, certaines pourraient être ciblées par divers régulateurs. Ainsi, l'altération de ces structures peut entraîner des effets importants sur l'expression génétique.

On peut ainsi souligner le cas d'une tige-boucle trouvée chez *E. coli* en amont de la séquence epsilon du gène *fis*, qui est impliqué dans la formation du nucléoïde, c'est-à-dire la région du cytoplasme bactérien contenant l'ADN chromosomique. Ce gène est exprimé à partir de l'opéron *dusB-fis*, et la tige-boucle en question ainsi que la séquence epsilon sont toutes deux incluses à la fin du gène *dusB* (Fig. 20), respectivement 50 nt et 31 nt en amont du codon de démarrage du gène *fis*, qui semble par ailleurs dépourvu de séquence SD.

De façon surprenante, les auteurs démontrent qu'une majorité de mutations provoquant l'ouverture de la tige-boucle entraîne une diminution importante de l'expression de *fis*, tandis que des mutations compensatoires restaurant la structure restaurent systématiquement l'expression génétique. Les auteurs se sont également assurés que l'effet était strictement post-transcriptionnel à l'aide de fusions transcriptionnelles et traductionnelles de *fis* avec le rapporteur *lacZ*. Si le mécanisme de régulation n'est pas élucidé, les auteurs proposent que cette structure participe au démarrage de la traduction de concert avec la séquence epsilon, peut-être *via* une interaction avec la protéine S1, et ce malgré la position lointaine de la tige-boucle vis-à-vis du codon de démarrage de *fis*. Il serait par ailleurs

particulièrement intéressant de savoir si des ARN régulateurs ou d'autres facteurs en *cis* ou en *trans* peuvent cibler cette structure afin de réguler la synthèse de la protéine Fis.

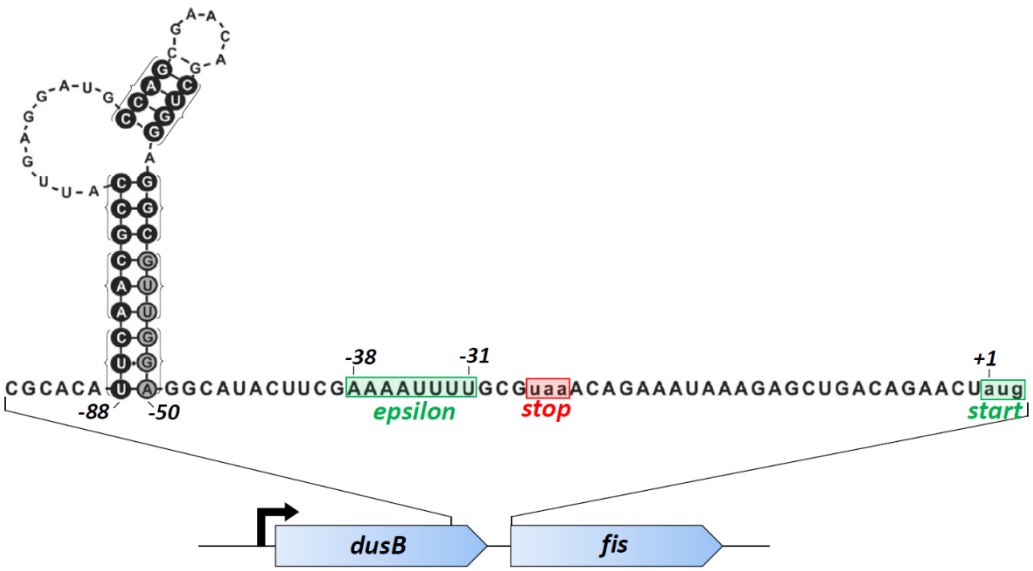


Figure 20 – Une structure activatrice de la traduction en amont du RBS de fis
L'opéron dusB-fis est exprimé depuis un promoteur en amont de dusB ; les flèches bleues correspondent aux ORF des deux gènes. Le détail des structures primaire et secondaire de la fin du cistron dusB et de la région intergénique est représenté ici jusqu'au le codon de démarrage du gène fis (codon « start »). Les éléments de séquence impliqués dans l'expression de fis sont encadrés en vert, tandis que le codon « stop » de dusB est encadré en rouge. La position des nt est indiquée par rapport à la position du codon de démarrage de fis. Sont respectivement encerclés en noir ou en gris les nt de la tige boucle activatrice affectant fortement ou faiblement l'expression de fis lorsqu'ils sont mutés. Les accolades correspondent aux groupes de mutations introduites par les auteurs. D'après (Nafissi et al., 2012).

On retrouve un autre exemple de structure secondaire ayant une influence positive sur l'expression génétique au niveau post-transcriptionnel dans une étude caractérisant l'effet de l'insertion de structures à proximité ou dans le RBS de divers gènes (Paulus et al., 2004). L'équipe en question cherchait ici à favoriser la synthèse de protéines hétérologues, c'est-à-dire provenant d'autres espèces, chez *E. coli* ou par la machinerie traductionnelle purifiée d'*E. coli*. Les auteurs ont donc modifié le 5'UTR de ces ARNm hétérologues afin qu'ils contiennent les éléments d'un RBS classique chez *E. coli*, et inséré des structures en tige-boucle à des positions différentes en aval du codon de démarrage pour en étudier les effets. (Fig. 21). Les auteurs ont pu démontrer ainsi qu'en fonction de la position d'insertion de la structure secondaire, mais aussi de sa stabilité, elle peut avoir des effets bien différents.

En résumé, à proximité directe du codon de démarrage de la traduction une tige-boucle devient inhibitrice quand son énergie libre atteint -7,8kcal.mol^{-1}, soit l'équivalent d'une

tige-boucle parfaite de 6 paires de bases (pb). L'insertion d'une tige-boucle parfaite de 8pb réprime quant à elle l'expression génétique pour autant qu'elle soit située à moins 13 nt en aval du codon de démarrage de la traduction (+13). Inversement, cette insertion semble devenir légèrement activatrice à partir de la position +16 et au-delà. Il est difficile de juger de l'effet exact de ces tiges-boucles du fait de l'absence de mutants démontrant que l'effet provient uniquement de la structure et non des codons insérés. Les auteurs ont cependant démontré que l'insertion de séquences riches en nt A-U – généralement simple-brin – entre le codon AUG et la tige-boucle favorise la traduction indépendamment de l'usage des codons. D'après cette étude, une expression optimale est obtenue grâce à l'insertion d'une structure en aval du 6ème codon, soit au nt +19, précédée d'une séquence riche en A-U et simple-brin. Cette configuration offrirait un contexte optimal pour le démarrage de la traduction, en limitant la formation de structures inhibitrices de la traduction entre le RBS et des régions en aval de l'ARNm (Fig. 21), bien que cette hypothèse ne soit pas explorée dans le détail.

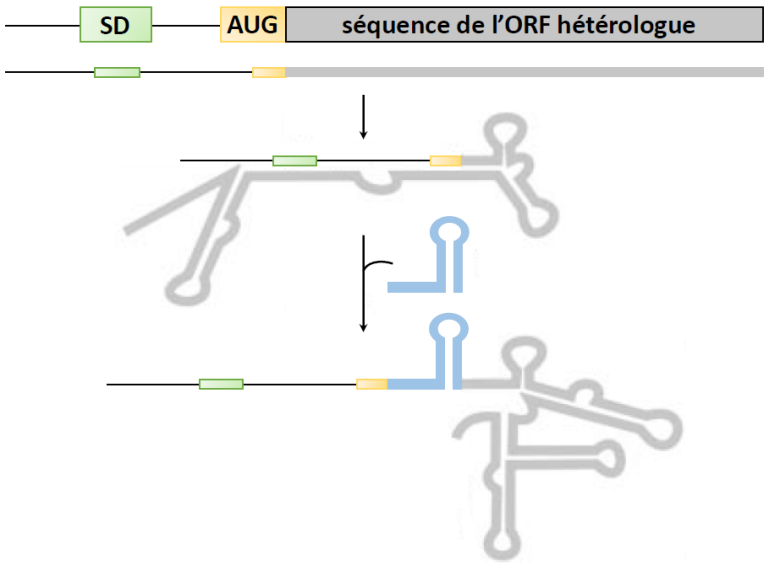


Figure 21 – Effet hypothétique d'une tige-boucle au début d'une séquence codante
ARNm chimérique proposé par (Paulus et al., 2004). L'ajout d'une séquence riche en nt A-U simple brin suivie d'une tige boucle (en bleu) permettrait d'éviter la formation de structures secondaires de l'ARNm séquestrant les séquences SD et AUG et réprimant la traduction. Adapté de (Paulus et al., 2004).

Il existe également quelques exemples de structures détectées dans les ORF qui semblent provoquer un changement fréquent de cadre de lecture (Kim and Tinoco, 2016), permettant l'expression de deux protéines différentes à partir d'un même ARNm ou provoquant l'apparition prématurée d'un codon de terminaison de la traduction. Par ailleurs et comme indiqué précédemment, quand des tiges boucles sont incluses dans la séquence

codante, elles peuvent entraîner des pauses dans la traduction (Fluman et al., 2014; Wen et al., 2008). Il semble toutefois s'agir de cas exceptionnels, et la forte activité hélicase du ribosome (Takyar et al., 2005) ainsi que l'analyse de l'ensemble des structures d'ARNm adoptées *in vivo* – ou structuromes – chez *E. coli* (Del Campo et al., 2015) et *Yersinia pseudotuberculosis* (Righetti et al., 2016) suggèrent que les structures capables de générer des pauses dans l'allongement de la traduction sont peu fréquentes. Enfin, une étude récente s'intéressant aux structures des ARNm polycistroniques d'*E. coli* révèle que ces ARNm sont structurés par ORF (Burkhardt et al., 2017). Ainsi, des structures secondaires se forment au sein de chaque ORF beaucoup plus fréquemment qu'entre différentes ORF du même ARNm, y compris pour les structures secondaires entre des régions distantes au niveau de la structure primaire. Cette structuration ORF-dépendante pourrait conférer un taux de traduction différent à chaque cistron de certains ARNm polycistroniques, ce qui permettrait par exemple un ajustement stœchiométrique de la synthèse des protéines issues des ARNm.

Pour conclure, il convient de rappeler que dans la grande majorité des cas recensés, des tiges-boucles à proximité du RBS ont un effet inhibiteur de la traduction tout simplement parce qu'elles empêchent la fixation du ribosome en limitant l'accès au SD, comme cela a été décrit précédemment (Johansson et al., 2002; de Smit and van Duin, 1994; Winkler et al., 2002). Cela permet de supposer qu'il existe une forte pression de sélection pour l'absence de structures dans les RBS, au moins dans le cas des gènes relativement bien exprimés, et cette absence de structures est une caractéristique bel et bien observée chez les bactéries (Ignatova and Narberhaus, 2017; Scharff et al., 2011).

3. ARN régulateurs agissant par complémentarité de séquence

Déjà mentionnés dans la partie consacrée aux asRNA agissant au niveau transcriptionnel, une multitude d'ARN régulateurs agissent par complémentarité de séquence avec des cibles ARNm pour exercer un contrôle de l'expression génétique. Je m'attarderai ici sur la régulation post-transcriptionnelle par ces ARN qui ont été d'une importance primordiale pour la compréhension de nombreux mécanismes de régulation chez les bactéries. Après un nouveau détour par les asRNA, seront abordés les petits ARN exprimés *en trans*, c'est à dire à partir d'un autre locus que celui de leurs cibles ARNm, et agissant par appariement imparfait de séquence. Ce sont d'ailleurs les ARN auxquels on prête le plus souvent l'appellation de

sRNA (pour « small RNA »). S'ensuivra une revue explicitant les techniques utilisées pour étudier ces sRNA ainsi que les divers mécanismes de régulation dans lesquels ils peuvent être impliqués. Je terminerai cette partie par les ARN « éponges », ces ARN qui peuvent séquestrer des sRNA, permettant ainsi la dérégulation de l'expression des cibles de ces sRNA. Par ailleurs du fait de l'immense diversité des ARN régulateurs, j'ai choisi de ne pas m'attarder sur les ARN CRISPR (« clustered regularly interspaced short palindromic repeats ») s'appariant aux ADN et impliqués dans un système immunitaire des bactéries contre tout matériel génétique étranger.

a. asRNA agissant au niveau post-transcriptionnel

Depuis le début des années 1980, les asRNA sont connus pour leur implication dans la régulation génétique. Ils ont d'abord été observés sur des plasmides ou au sein de phages et prophages. Depuis, de nombreux asRNA chromosomiques ont été découverts et semblent repandus dans les trois domaines du vivant (Wagner and Romby, 2015; Wagner et al., 2002).

L'une des caractéristiques principales des asRNA est leur complémentarité parfaite avec les ARNm dont ils régulent l'expression, puisqu'ils sont transcrits depuis le brin d'ADN opposé. Ils peuvent ainsi former des duplex d'ARN double-brin de plusieurs dizaines voire centaines de pb. Si leurs mécanismes d'action sont rarement bien compris, on note que certains cas suggèrent un effet sur la fixation du ribosome du fait de la complémentarité étendue de l'asRNA avec le RBS de la cible ARNm. C'est par exemple le cas de l'asRNA *symR* qui régule la synthèse de la toxine SymE chez *E. coli* (Kawano et al., 2007). Cependant, les études effectuées manquent dans la majorité des cas d'une démonstration de l'effet direct sur la fixation du ribosome (Thomason and Storz, 2010), qui pourrait être fournie par des expériences de toeprints (voir **Revue n°1**). Aussi, un effet de ces asRNA sur la traduction est rarement dissocié d'un effet sur la stabilité de l'ARNm, sans distinction de cause et de conséquence entre ces deux effets.

Un effet plus évident de certains asRNA est directement lié à la stabilité de leur cible ARNm. Cela s'explique par la formation des duplex d'ARN qui peuvent par exemple être ciblés par la RNase III clivant spécifiquement les ARN double-brin, ce qui est observé dans le cas de l'antisens plasmidique Sok du gène *hok* encodant une toxine (Gerdes et al., 1992). Dans

d'autres cas, ces interactions peuvent libérer des régions simple-brin situées ailleurs sur l'ARNm, qui peuvent alors être clivées spécifiquement par la RNase E. C'est le cas de l'asRNA AmgR qui provoque la dégradation RNase E-dépendante de l'ARNm polycistronique *mgtCBR* chez l'entérobactérie *Salmonella enterica* (*S. enterica*) (Lee and Groisman, 2010). Il convient cependant de noter que dans ces divers exemples les dégradations démontrées n'excluent pas un effet traductionnel des asRNA, qui pourraient notamment empêcher la fixation du ribosome aux ARNm. Un mécanisme opposé est par ailleurs possible, où l'asRNA protège l'ARNm ciblé contre une dégradation par une RNase (Thomason and Storz, 2010).

Curieusement, dans de multiples exemples l'ARNm régulé code une protéine toxique, et l'asRNA est l'antitoxine empêchant la synthèse de cette toxine. Il serait intéressant de savoir si la production d'un antisens constitue un moyen évolutif rapide pour développer une antitoxine. En effet si on imagine qu'un gène de toxine a été transmis à une bactérie dépourvue d'antitoxine, par conjugaison avec une autre bactérie ou par le biais d'un plasmide ou d'un phage, l'apparition d'un promoteur convergeant vers ce gène a pu rapidement être sélectionnée et constituer un moyen de protection contre cette toxine. Rechercher s'il existe des antitoxines protéiques et asRNA réprimant l'activité ou la synthèse d'une toxine homologue chez des bactéries différentes pourrait notamment apporter de la consistance à cette hypothèse.

b. ARN agissant par appariement imparfait, ARN Hfq-dépendants

De nombreux sRNA bactériens agissent par appariement imparfait de séquence pour réguler des cibles ARNm exprimées en *trans*. Beaucoup d'entre eux ont été identifiés au début des années 2000 (Argaman et al., 2001; Vogel, 2003; Wassarman, 2001). D'une longueur moyenne d'environ 100 nts, on les trouve chez toutes les bactéries et ils sont impliqués dans la plupart des réseaux de régulation et de réponse au stress, où leurs rôles et leurs effets en tant que régulateurs ARN pourraient être complémentaires à ceux des régulateurs protéiques (voir **Revue n°2**). La recherche de sRNA s'est longtemps focalisée sur les régions intergéniques de chromosomes et de plasmides, néanmoins la découverte récente de sRNA issus de l'extrémité 3' d'ARNm (Chao et al., 2012; Guo et al., 2014; Peng et al., 2016) ou d'ARNt (Lalaouna et al., 2015) indique qu'il reste certainement encore beaucoup de sRNA inconnus à ce jour, même au sein des espèces les plus étudiées.

L'une des particularités intéressantes des sRNA découle de leur capacité à former des appariements de séquence imparfaits, leur permettant souvent de cibler plusieurs ARNm (voir **Introduction III.B.1.b** pour un exemple avec les sRNA OmrA et OmrB). L'appariement peut ainsi s'effectuer dans des régions et avec des séquences qui varient d'une cible ARNm à l'autre. On remarque en revanche que c'est souvent le même site d'un sRNA qui est impliqué dans l'appariement avec ses cibles (Storz et al., 2011), hormis quelques exceptions notables. Ainsi le sRNA GcvB possède au moins deux sites distincts impliqués dans des appariements avec des cibles différentes (Coornaert et al., 2013). Ce cas de figure apporte d'ailleurs une perspective intéressante sur la compétition qui peut exister entre les différentes cibles d'un sRNA pour la régulation de leur expression génétique. On peut par exemple se demander si la même molécule de sRNA n'interagit qu'avec une seule molécule cible ARNm avant d'être dégradée, ou si elle peut au contraire réguler séquentiellement l'expression de plusieurs cibles. Dans le cas de GcvB une autre interrogation repose sur la faculté de ce sRNA à lier deux cibles ARNm en même temps. Il semblerait toutefois qu'au moins dans certains cas de régulation négative, le sRNA soit dégradé en même temps que sa cible ARNm, à l'image du sRNA RyhB et sa cible *sodB* (Massé et al., 2003; Prevost et al., 2011). Cela indiquerait que la molécule de sRNA peut être consommée et dégradée lors de la régulation d'une seule et unique molécule d'ARNm, permettant à la cellule de contrôler très finement l'expression génétique.

Parmi ces sRNA, beaucoup sont stabilisés par des protéines chaperonne à ARN telles que Hfq (Wagner and Romby, 2015), ou encore la plus récemment décrite ProQ (Smirnov et al., 2017). A ce jour, les sRNA les plus étudiés et les mieux compris sont très certainement les sRNA Hfq-dépendants. Cette classe d'ARN régulateurs trouvés chez les bactéries à Gram négatif a été tout particulièrement étudiée chez les entérobactéries, occasionnant la découverte de nombreux mécanismes de régulation post-transcriptionnelle (voir **Revue n°1**). Ils ont ainsi largement contribué au bouleversement de la communauté scientifique lié à la notion d'ARN régulateur. La protéine Hfq a d'abord été identifiée comme un facteur nécessaire à la réplication du coliphage Qβ (d'où son nom « host factor for phage Qβ ») – vraisemblablement parce qu'elle stabilise un ou plusieurs sRNA du phage et/ou de la bactérie importants pour l'infection par ce phage. Sa structure homohexamérique, élucidée peu après

la découverte de son implication dans la régulation de l'expression génétique par des ARN (Sauter, 2003) (Fig. 22), se décompose en face proximale, face distale et bordure.

Figure 22 – Structure tridimensionnelle de la protéine Hfq
Hfq adopte une structure en « doughnut » (selon l'expression des auteurs), avec à gauche la vue de face proximale et à droite la vue latérale (bordure). Un monomère est représenté avec son extrémité N-terminale en hélice-α en rouge, suivie de cinq feuillets β en bleu. D'après (Sauter, 2003).

Les sRNA Hfq-dépendants peuvent être divisés en deux classes, I et II, selon les régions d'Hfq avec lesquelles ils interagissent : si les sRNA des deux classes interagissent avec la face proximale *via* une séquence poly-AAN, ceux de la classe I interagissent également avec la bordure tandis que ceux de la classe II se lient à la face distale *via* une séquence poly-U (Schu et al., 2015; Zhang et al., 2013) (Fig. 23).

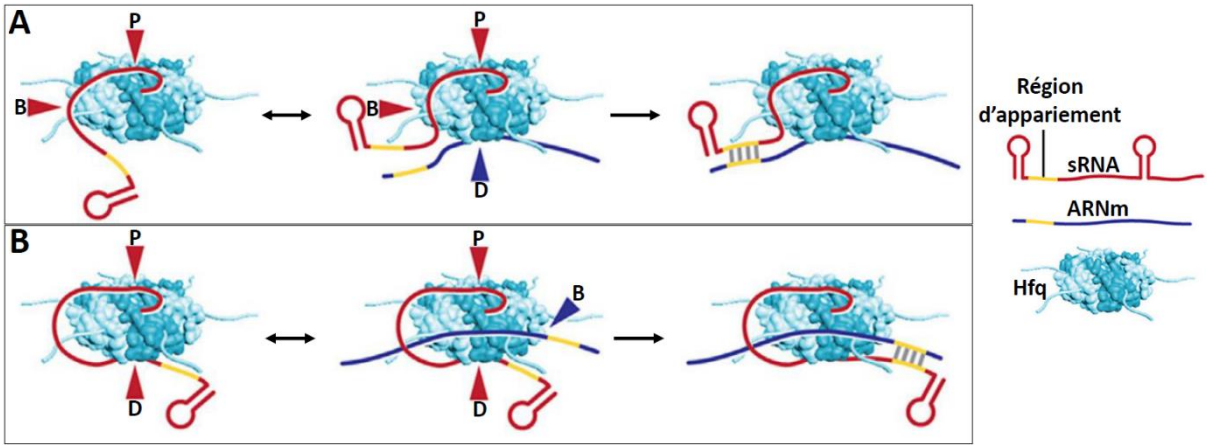


Figure 23 – Formation de duplexes sRNA – mRNA assistée par Hfq
Les deux classes de sRNA Hfq-dépendants nécessitent l'intervention de Hfq pour s'apparier à leurs cibles ARNm. Sont représentés en rouge les sRNA et en bleu les ARNm ; les flèches indiquent des interactions ARN – protéine impliquant diverses régions de Hfq : P = face proximale, B = bordure, D = face distale. Les sRNA de classe I (A) s'associent à la face proximale et à la bordure de Hfq pour s'apparier à des ARNm se liant à la face distale. Les sRNA de classe II (B) s'associent aux faces proximale et distale pour s'apparier à des ARNm liant la bordure de Hfq. Adapté de (Updegrove et al., 2016).

De manière intéressante, pour chaque classe de sRNA la région de Hfq se fixant au sRNA est différente de celle qui se fixe à l'ARNm, régions de Hfq reconnaissant des séquences consensus et des structures secondaires spécifiques. De plus, un sRNA et un mRNA ne pourront donner lieu à une interaction productive *in vivo* que si leurs sites de fixation à Hfq sont compatibles, c'est à dire qu'ils interagissent avec différentes régions d'Hfq. (N. Figueroa-Bossi & L. Bossi, communication personnelle).

La grande flexibilité dans les duplex formés entre sRNA et ARNm rend notamment difficile la prédiction *in silico* de cibles des différents sRNA connus, bien que plusieurs équipes s'y soient essayées (Busch et al., 2008; Tjaden et al., 2006; Wright et al., 2013). Afin de découvrir les cibles de ces différents sRNA, des approches globales ont donc été utilisées, à l'image de techniques de transcriptomique telles que les puces à ADN (Guillier and Gottesman, 2006; Massé et al., 2005), ou plus récemment de séquençages d'ARN haut débit (RNAseq) et de « ribosome profiling » (Guo et al., 2014; Wang et al., 2015). Il existe en outre une technique basée sur la purification des cibles ARNm appariées à des sRNA fusionnés à un « tag » MS2 issu du coliphage homonyme, suivie d'un RNAseq (Lalaouna et al., 2016). Cette dernière technique permet ainsi d'aborder le problème d'une façon différente dans la mesure où ce sont uniquement les cibles directes du sRNA qui sont considérées. Plus spécifiquement pour les sRNA Hfq-dépendants, l'immunoprécipitation de Hfq combinée à la ligature entre cibles ARNm et sRNA, suivie d'un RNAseq, a aussi décuplé le nombre de couples de sRNA-cible ARNm potentiels (Melamed et al., 2016) tout en fournissant des indications précieuses sur les éventuels sites d'appariement concernés. Toutefois lors de la recherche de nouvelles cibles d'un sRNA, l'idéal reviendrait à combiner diverses approches globales et bioinformatiques. Enfin, seule une étude au cas par cas permet la confirmation des interactions et/ou régulations ainsi mises en évidence. La compréhension des mécanismes de régulation impliqués constitue une seconde étape de l'étude de ces sRNA ; la grande diversité de ces mécanismes est détaillée dans la **Revue n°1**.

c. **Revue n°1**: "*Mechanistic study of base-pairing small regulatory RNAs in bacteria*"

Methods 117 (2017) 67–76

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Mechanistic study of base-pairing small regulatory RNAs in bacteria

Jonathan Jagodnik, Anaïs Brosse, Thao Nguyen Le Lam, Claude Chiaruttini, Maude Guillier*

CNRS UMR8261, Associated with University Paris Diderot, Sorbonne Paris Cité, Institut de Biologie Physico-Chimique, 75005 Paris, France

ARTICLE INFO

Article history:
Received 3 August 2016
Accepted 22 September 2016
Available online 28 September 2016

Keywords:
Gene expression regulation
Post-transcriptional control
RNA stability
RNA processing
SRNA-mRNA base pairing
RNA structure probing
SHAPE
Toeprint
Hfq

ABSTRACT

In all three kingdoms of life, RNA is not only involved in the expression of genetic information, but also carries out extremely diverse cellular functions. This versatility is essentially due to the fact that RNA molecules can exploit the power of base pairing to allow them to fold into a wide variety of structures through which they can perform diverse roles, but also to selectively target and bind to other nucleic acids. This is true in particular for bacterial small regulatory RNAs that act by imperfect base-pairing with target mRNAs, and thereby control their expression through different mechanisms. Here we outline an overview of *in vivo* and *in vitro* approaches that are currently used to gain mechanistic insights into how these sRNAs control gene expression in bacteria.

© 2016 Elsevier Inc. All rights reserved.

Contents

1. Introduction . . . 67
2. *In vivo* approaches . . . 69
   2.1. Studying the effect of an sRNA on a target using reporter gene fusions . . . 69
   2.2. Effect of the sRNA on mRNA levels and stability . . . 70
   2.3. A role for other actors in addition to the sRNA-mRNA duplex? . . . 71
3. *In vitro* approaches . . . 71
   3.1. Synthesis, labeling and detection of RNA molecules for *in vitro* studies . . . 71
   3.2. *In vitro* characterization of the sRNA-mRNA interaction . . . 71
   3.3. *In vitro* study of translation initiation: the toeprinting assay . . . 73
   3.4. *In vitro* translation experiments . . . 74
   3.5. Studying the degradation or processing of sRNA-mRNA duplexes *in vitro* . . . 74
4. Conclusion . . . 74
   Funding . . . 75
   Acknowledgements . . . 75
   References . . . 75

1. Introduction

It is now recognized that the transcribed fraction of genomes is much higher than previously anticipated and that RNA molecules play versatile roles in processes other than the translation of genetic expression into proteins [1]. The characterization of several of these RNAs highlighted their role in a wide variety of different biological functions, which include enzymatic activity,

Abbreviations: CMCT, 1-cyclohexyl-32-(morpholinoethyl)carbodiimide metho-*p*-toluene sulfonate; DMS, dimethyl sulfate; FACS, fluorescence-activated cell sorting; kethoxal, -ethoxy-α-ketobutyraldehyde; 1M7, 1-methyl-7-nitroisatoic anhydride; NMIA, N-methylisatoic anhydride; nt, nucleotide; ORF, open reading frame; PAGE, polyacrylamide gel electrophoresis; RBS, ribosome binding site; RT, reverse-transcription; SHAPE, selective 2′-hydroxyl acylation analyzed by primer extension; UTR, untranslated region.

* Corresponding author.
E-mail address: maude.guillier@ibpc.fr (M. Guillier).

http://dx.doi.org/10.1016/j.ymeth.2016.09.012
1046-2023/© 2016 Elsevier Inc. All rights reserved.

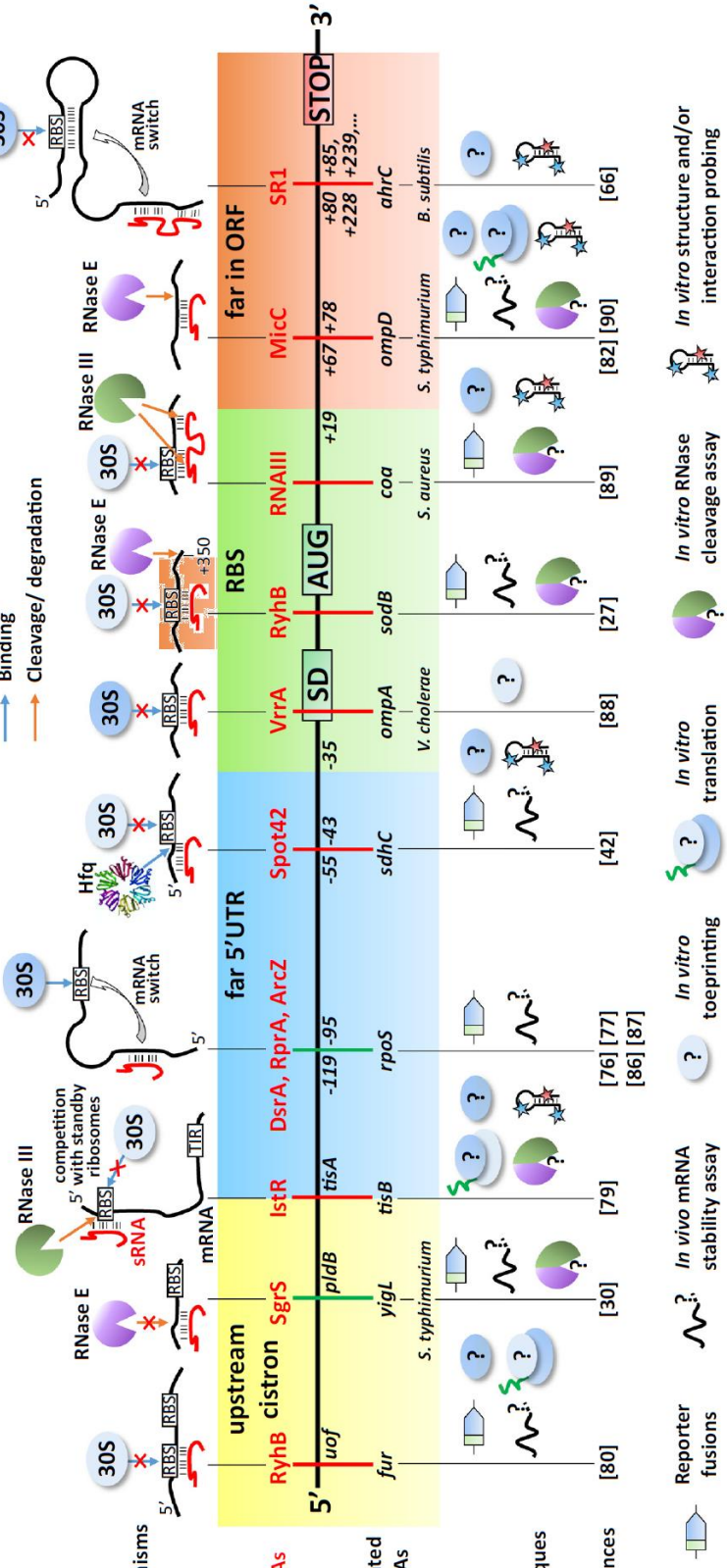


Fig. 1. Diversity of sRNA-based regulatory mechanisms and the techniques used to dissect them. Selected examples of sRNAs that activate (green vertical bars) or repress (red vertical bars) expression of target mRNAs by pairing at various positions of the transcript. The regulatory mechanism and the experimental approaches that were used are schematically depicted above and below, respectively. Four distinct regions for sRNA-binding are shown on the mRNA, with the RBS corresponding to the nts −35 to +19 as determined in [85]. Positions of the mRNA (or the cistron when relevant) where the sRNA binds are shown when pairing occurs outside of the RBS. When different from *E. coli*, the organism in which the study was done is indicated. See text for details. TIR: translation initiation region. (See above-mentioned references for further information.)

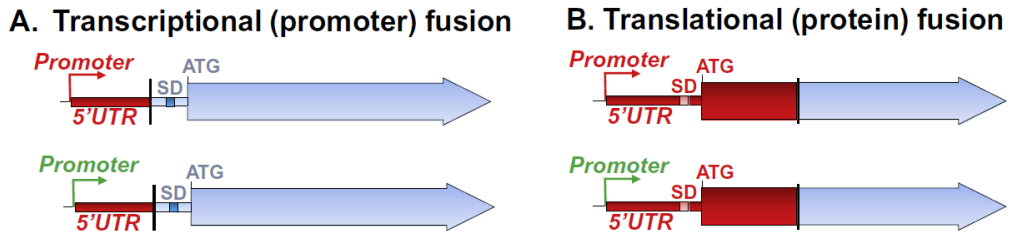


Fig. 2. Schematics of the transcriptional and translational fusions. Sequences of the gene of interest and of the reporter gene are shown in red and blue, respectively. Black vertical bars indicate the fusion points. Green arrows represent heterologous promoters, either constitutive or inducible, that replace the promoter of the gene of interest. Note that transcriptional fusions can carry some or all of the coding sequence of the target gene as well.

maintenance of genome integrity by protection against foreign DNA, processing of precursor RNAs, methylation or pseudouridylation of rRNA, and modification of protein activity [2].

RNA molecules have also been found to play important regulatory roles at each stage of gene expression *via* extremely diverse mechanisms. This is particularly true in bacteria where, in a few model organisms at least, detailed understanding of their modes of action has elucidated novel modes of RNA-based control [3].

Bacterial regulatory RNAs are extremely diverse from different standpoints. If we consider their synthesis for instance, they can be *bona fide* primary transcripts made from free-standing genes, but also originate from promoters internal to coding sequences (e.g. MicL [4]), or 3′ untranslated regions (e.g. DapZ [5]). Regulatory RNAs can also be generated by the processing of longer mRNAs (e.g. CpxQ [6]), tRNAs precursors [7], or even from longer isoforms of regulatory RNAs (e.g. SdsN [8]). Lastly, regulatory RNAs can constitute an integral part of an mRNA: this is the case for riboswitches for instance [9]. Another source of diversity within regulatory RNAs comes from their various modes of action. While riboswitches affect expression of *cis* encoded genes in response to ligand binding and conformational rearrangement, other regulatory RNAs modify the activity of a protein or more often act *via* base-pairing with mRNA. Base-pairing RNAs fall into two main classes: (i) the antisense RNAs (asRNAs), transcribed from the DNA strand opposite to their target gene, and that consequently have perfect and extended complementarity to the mRNA and (ii) the trans-encoded regulatory RNAs that target mRNA(s) encoded by a different *locus via* imperfect complementarity and therefore can regulate multiple genes in most cases [10].

The aim of this review is to provide an overview of the different approaches that can be used to study mechanisms by which bacterial regulatory RNAs control gene expression. We have chosen to focus more specifically on RNAs acting by imperfect base-pairing because this represents a major class of regulatory RNAs. Hereafter we will refer to these as small RNAs (sRNAs), as their typical size is in the 70–200 nts range. Furthermore, we will not cover the techniques aiming at the identification of direct mRNA targets of sRNAs (reviewed in [3], see also [11] for a more recent technique); we will thus focus on sRNAs and targets that were previously shown to interact directly. Most approaches mentioned below have nonetheless much broader applications and can be used in particular to study the action of other regulators, including RNAs, that do not act *via* imperfect base-pairing.

The best way to get insights into the detailed mode of action of an sRNA is to combine *in vivo* and *in vitro* approaches, that will bring complementary information. The techniques to be used will depend on the sRNA/target pair being studied and the different questions that it raises. The most common techniques that have previously successfully been used in the study of sRNAs are described below, starting with the *in vivo* approaches. The various mechanisms and techniques that are discussed in this review are summarized in Fig. 1.

2. *In vivo* approaches

Imperfectly base-pairing sRNAs typically affect the translation and/or the stability of target mRNAs either positively or negatively [11] and can also directly impact transcription of a target gene [12,13]. It is thus useful, once an sRNA has been shown to directly affect the levels of a given target mRNA or protein to study which step of gene expression is affected. This is generally performed *in vivo* by assessing whether the expression of the sRNA affects that of its target gene, as measured with transcriptional or translational fusions to a reporter gene, and by looking at the mRNA levels and/or stability.

2.1. Studying the effect of an sRNA on a target using reporter gene fusions

The general idea underlying the use of fusions is to monitor the expression of a given target gene with a read-out that is easily and precisely quantifiable. Two types of fusions exist depending on the region of the gene of interest attached to the reporter gene (Fig. 2). In the case of transcriptional fusions (or promoter fusions), the transcription signals are those of the target gene while the translation signals are from the reporter gene. Changes in the expression of a promoter fusion thus indicate changes in the transcription levels of the target gene or in the stability of the reporter mRNA. The situation is different with translational fusions (or protein fusions) where transcription and translation are driven by sequences from the target gene resulting in a chimeric protein (Fig. 2B). It is also common to use translational fusions expressed from a heterologous promoter to focus on regulatory events occurring downstream of transcription initiation. Depending on the goal, this promoter can be constitutive or inducible. The same promoter swapping approach can be performed with promoter fusions to discriminate effects at the promoter level from effects on transcription elongation or stability of the fusion mRNA.

Translational fusions have been extensively used to look at the effect of mutations in the regulatory regions located on both the trans-encoded sRNA and its target mRNA or to validate base-pairing interactions between sRNA and target mRNA *via* compensatory changes, but fusions can provide much more information regarding the effect of sRNAs. They can for instance allow deciphering the level of control by comparing the action of sRNAs on transcriptional and translational fusions. They can also be useful to narrow down the regions of a given gene that are required for sRNA control by constructing fusions carrying various regions of the target. In this regard, it is possible to change the promoter as already mentioned above, or to introduce deletions in the 5′UTR

or in the ORF of the target gene. Another powerful application of gene fusions is that they can be exploited to identify different factors involved in sRNA control of a given gene through using genetic screens or selections. A genetic selection aiming at identifying regulators of the ChiP chitoporin in *Salmonella* revealed that ChiP synthesis is regulated by Hfq and by the ChiX sRNA for instance, while a genetic screen led to further characterization of the system and ultimately demonstrated that levels of the ChiX sRNA are controlled by an inducible "bait target" of this sRNA [14]. In the same study, mutants were isolated in *chiP* or ChiX *in cis*, highlighting that genetics and gene fusions can also identify bases important for the control in the sRNA or the target.

Different reporter genes have been successfully used, with the most common ones being *lacZ*, the gene encoding β-galactosidase, and genes coding for fluorescent proteins such as *gfp*. These reporters have different advantages: *lacZ* provides a highly sensitive assay for gene expression and, because it gives rise to easily discernable phenotypes on plates, it has been widely used in genetic screens or selections [see [15] for another example], while *gfp* (or other fluorescent reporters) can be used to detect heterogeneity of gene expression in a clonal population, by FACS or microscopy, for instance [16]. Fusions have been engineered on the chromosome of several bacterial species at different *loci*, but they have also been expressed from plasmids in several cases [17,18].

Fusion of a reporter gene to a target gene results in the synthesis of a chimeric mRNA that might no longer be controlled by its cognate sRNA. This could be due to specific features in the structure of the chimeric mRNA which might prevent sRNA binding, or affect its translation and/or its stability. In cases where a gene fusion does not accurately report control by an sRNA, it is possible to use smaller tags instead (e.g. Flag or 6xHis tag) with the same idea that it is easy to follow gene expression [19]. Even these smaller tags can interfere with "natural" gene regulation and in this case, it is still possible to study the sRNA control by directly following the target protein or mRNA levels in response to changes in the sRNA expression, using western- or northern-blotting, respectively. In some cases, it is also possible to easily follow the levels of a protein based on its function, if it gives rise to a discernable phenotype or a quantifiable measurement for instance.

One last point that is worth mentioning here is that fusions are not restricted to the target mRNAs. Fusions to sRNA promoters have been widely used to study their transcriptional control. More importantly for the scope of this review, chimeras between different domains of sRNAs have been used to show that a limited region of an sRNA is necessary and sufficient for control [20,21].

2.2. Effect of the sRNA on mRNA levels and stability

If the expression of a gene is regulated by an sRNA, it might be interesting to determine whether the levels of mRNA transcribed from this gene are correlated with the expression of the regulatory sRNA, but also whether the stability of this mRNA is affected. The most straightforward way to assess the stability of an mRNA is to measure its decay rate after blocking its synthesis, which is commonly achieved *via* the addition of rifampicin to cultures. To test whether an sRNA controls the stability of a target mRNA, isogenic strains expressing or lacking this sRNA are thus typically grown up to a specific point in the growth-phase before addition of rifampicin. Total RNA is then extracted at several time-points just before and following rifampicin addition and the amount of target mRNA remaining is determined for each of these time-points, by northern-blotting for instance. The half-life of the mRNA, which is defined as the time required to decrease the amount of the mRNA 2-fold, is calculated using linear fits to a semilog plot of the hybridization data *versus* time.

If, in such an experiment, an sRNA affects the stability of its target, this raises several questions, and in particular (i) whether this is a primary effect of the sRNA-mRNA interaction, (ii) whether this change in stability is required for the control, and (iii) what are the RNases involved and what is their role?

Regarding the first point, it is well established that translating ribosomes protect the mRNA from ribonucleases [22–24] and decreasing the translation of an mRNA with an sRNA is thus expected to decrease its stability. Nonetheless, sRNAs can also directly target mRNA for degradation, independently of translational control. This can be assessed by determining whether degradation of the target mRNA is still sRNA dependent in an *in vitro* degradation assay, i.e. in the absence of the translation machinery (see Section 3.5). It is also possible to study *in vivo* whether the sRNA affects the stability of the target mRNA in the absence of translating ribosomes. It has been shown for instance that the SgrS sRNA induced a strong decrease in the *ptsG* mRNA levels, when the translation efficiency was reduced by a mutation in the Shine-Dalgarno sequence, or by the addition of kasugamycin, a translation initiation inhibitor [25,26]. Similarly, RyhB sRNA was found to repress *sodB* mRNA levels even when *sodB* translation was blocked by the introduction of a premature stop codon [27].

Recently, a ribosome profiling approach following the induction of the RyhB sRNA has proven useful to discriminate between targets that are regulated at the level of translation and/or at the level of stability [28]. While this does not address the question of whether changes in stability of targets are a consequence of changes in translation (for targets for which both stability and translation are affected), this study identified novel RyhB targets, some of which are regulated only at the level of translation, and this approach is likely to be applied to other sRNAs in the near future.

Importantly, even though an sRNA might change the stability and thus the levels of a target mRNA, this does not necessarily mean that these changes are required for the control of the protein levels. If we consider again the regulation of *ptsG* by SgrS for instance, the synthesis of the *ptsG*-encoded $IICB^{Glc}$ protein is still controlled in the absence of *ptsG* degradation [25]. In contrast, the control of the surface protein A by the regulatory RNAIII in *Staphylococcus aureus* requires that the RNase III endonuclease cleaves the resulting RNAIII-mRNA duplex [29]. In other words, it is important to determine whether degradation of the mRNA is required for sRNA control of the protein level on a "case-by-case" basis.

Note that sRNAs can also increase target mRNAs stability: in several cases, sRNAs have been reported to induce changes in the processing of their targets, leading to the production of a more stable form of the transcript. Such processing events can be followed *in vivo* by analyzing the 5′ or 3′ ends of the target, by primer extension (for the 5′ end) or RACE experiments [30,31].

Another important question to address the mechanism of control is to determine which RNase(s) are responsible for the observed change in stability or processing of the target mRNA. This can be done through a candidate approach by testing whether these changes are still observed in strains mutated for known RNases. Obvious candidates that can be tested are RNase E (in Gram-negative bacteria), RNases Y and J1 (in Gram-positive bacteria) and RNase III. These enzymes have major impact in cellular RNA turnover and have already been shown to play important roles in control mediated by sRNAs [32,33]. Note however that even if these RNases are involved in the control *in vivo*, this can be indirect, and again an *in vitro* assay is required to discriminate direct and indirect effects (see Section 3.5).

If no single RNase is identified in such an assay, one can combine mutations for various RNases as this issue could be due to (partial) functional redundancy. Alternatively, one can look for other factors involved in the control, e.g. by genetic methods,

which could reveal previously uncharacterized RNase(s) as well as other genes whose function in sRNA-mediated regulation will have to be deciphered in further detail. The role of PNPase in mediating the stability of several sRNAs was discovered by a combined genetic screen and selection for example [15].

2.3. A role for other actors in addition to the sRNA-mRNA duplex?

Other factors can be required for regulation of gene expression by sRNAs acting by base-pairing with mRNAs. For instance, in Gram-negative bacteria, the Hfq RNA chaperone has been shown to be involved in both the stabilization of the sRNAs and sRNA-mRNA duplex formation [34]. More generally, the precise understanding of the regulatory mechanism requires that these factors be identified and their action further characterized. A good example comes from the genetic study showing that the product of the *rapZ* gene was involved in the control of *glmS* by the GlmZ sRNA [35], while further work later showed that RapZ specifically directs the RNase E endonuclease to this sRNA or to the GlmY "mimic", thus protecting GlmZ from cleavage in this latter case [36]. If we want to identify factors required for the action of an sRNA on a given mRNA, it is *a priori* possible to take advantage of genetic screens or selections to look for factors that affect control when mutated (or overproduced). A more in-depth characterization of these factors and of their mode of action is then required, possibly through *in vitro* experiments. Because sRNAs may only have a subtle effect on expression of some targets, getting such genetic screens to work can be difficult and all factors may not be identified by this approach. One can thus envision to search for partners interacting with a given sRNA by a biochemical approach in parallel [37,38]. Immunoprecipitation of RNA-binding proteins and subsequent analysis of associated RNAs has also been used to identify protein partners for sRNAs (and mRNAs) [e.g. 39–41]. How these factors intervene in the regulation then requires further analysis, and *in vitro* approaches might be particularly helpful for this purpose.

3. *In vitro* approaches

While *in vivo* results are essential and highly informative in regard to the key aspects of sRNA regulation of gene expression, the use of *in vitro* techniques is necessary to further elucidate the molecular mechanisms and to define the minimal number of actors required for control and to decipher how they act. For instance, the Hfq chaperone is required *in vivo* for the action of many sRNAs, but it is often unclear whether it is because it stabilizes them, facilitates duplex formation and/or has a more direct role in the control. This question can be addressed *in vitro*: it has been shown for instance that, in the control of *sdhC* mRNA by Spot42 sRNA, binding of the sRNA to the target mRNA recruits Hfq to the vicinity of the RBS and that Hfq is the factor that competes with ribosome binding [42]. In contrast, in many other cases, the sRNA is sufficient to compete with ribosome binding *in vitro* without addition of Hfq [e.g. 43].

It is important to emphasize that *in vitro* experiments are subject to artefacts, as the experimental conditions necessarily vary from the cellular context. Even if concentrations and ratios of the different molecules present in the *in vitro* assay are carefully chosen to reflect those encountered *in vivo* as much as possible, this change in experimental conditions, including the absence of all other cellular components, can lead to non-specific effects. *In vitro* approaches are thus usually used to follow up an *in vivo* study and must necessarily include specificity controls; in the case of sRNAs, such a control can be a mutant derivative that no longer regulates target gene expression *in vivo*. *In vitro* approaches involve working with purified and/or *in vitro* synthesized components. Proteins are most often purified from strains in which they are overexpressed, and general considerations for the preparation of RNA components are described below.

3.1. Synthesis, labeling and detection of RNA molecules for in vitro studies

In most cases, RNA molecules for *in vitro* studies are obtained by *in vitro* transcription using RNA polymerase from the T7 bacteriophage [44–46]. Depending on the experiment, it might be necessary to label the RNA, which is most often performed either during transcription, by including a labeled NTP in the transcription mixture, or post-transcriptionally, e.g. at the 5′ end using T4 polynucleotide kinase and [32P-γ]ATP, or at the 3′ end using T4 RNA Ligase and [^{32}P]pCp. This labeling provides an easy detection of the full-length transcripts and cleavage products in RNA structure probing experiments or processing/cleavage assays (see below). The different RNA fragments are separated by denaturing PAGE and their sizes (as well as cleavage sites) inferred from ladders generated by partial digestion of RNA by T1 RNase, usually coupled with partial alkaline digestion. Note however that it is essential that the RNA molecules that are to be labeled are homogeneous in size, which requires a PAGE purification; for this reason direct labeling can only be used for relatively short RNAs (<200 nucleotides). Alternatively, tags have been developed that allow purification of longer transcripts with homogeneous ends [47].

Another approach that is widely used to analyze cleavages of *in vitro* (or *in vivo*) transcripts is their detection as stops to cDNA extension from a DNA primer labeled at its 5′ end either with ^{32}P or with a fluorescent dye. The resulting cDNA fragments are subsequently separated by denaturing PAGE alongside with DNA fragments generated by sequencing reactions with the same primer. Sequencing reactions can be performed either on the *in vitro* transcripts or on the DNA template used for transcription. The sites of cleavage are inferred from the length of the resulting cDNA fragments compared to that of the DNA fragments generated by sequencing. Even if it involves an extra step prior to analysis, there are several advantages to this method. The first is that it is not restricted to the analysis of cleavage products but it also applies to the analysis of chemical modification of RNA or formation of translation initiation complex (see below). The second advantage is that RNA of any size can be probed because labeled primers can be annealed anywhere along the RNA molecule, and multiple primers can be used in separate extension reactions to detect modifications along the entire molecule. The third advantage is that there is no need to gel-purify the RNA or to label it in advance. This method will not work however if the region of interest is located very close to the 3′ end of the RNA under study since cDNA extension requires annealing of a primer downstream of the concerned region.

Before starting *in vitro* experiments, which are often time-consuming and labor-intensive, it must be realized that the probability of misfolding increases dramatically for larger RNAs where several potential isoforms can exist. It is thus especially important to restrict, if possible, the region of the mRNA that will be used in the *in vitro* experiments, for instance when there are *in vivo* indications that only a given region is sufficient for control (see above). Regarding the sRNAs, the probability of misfolding is not considered to be a major issue, given their short sizes in general. Nonetheless, sRNAs have sometimes been replaced by short DNA or RNA oligonucleotides mimicking the seed-pairing region in several *in vitro* approaches, which turned out to be both easier from an experimental point of view and highly informative [21,48].

3.2. In vitro characterization of the sRNA-mRNA interaction

Even when one sRNA-mRNA interaction site has been validated *in vivo*, usually *via* introduction of compensatory changes that re-establish sRNA-mRNA base pairing, it might be informative to

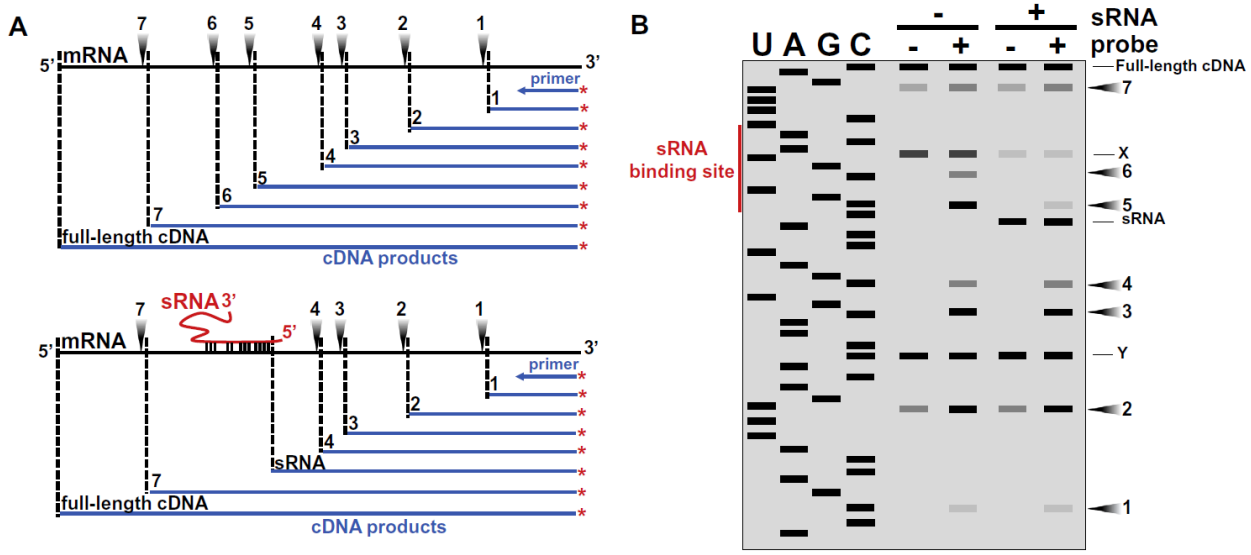


Fig. 3. Principle of the RNA structural probing experiments and analysis by RT. (A) Nucleotides reacting with the probe are indicated by black arrowheads and blue lines represent the cDNA species produced by RT of the labeled DNA primer. RT is blocked one nt downstream of a reactive residue and/or RNA structure. (B) Scheme of the fractionation of the RT products after sequencing PAGE. Nucleotides reactive to the probe present different reactivities (here, moderate for positions 1 and 7; strong for positions 2 and 6). Reactivities of positions 5 and 6 to the probe are highly dependent on binding of the sRNA, and other RT stops – potentially caused by mRNA secondary structures or cleavages – are depicted as either being sensitive to sRNA binding (X) or unchanged upon sRNA binding (Y).

determine (i) whether the interaction site is restricted to the base pairs that have been validated or is longer, and (ii) whether additional interaction sites exist. While *in vivo* data can provide partial answers to these questions, for instance by narrowing down the regions required for the control, it is also possible to precisely analyze the sRNA-mRNA duplex *in vitro*. Possibly the simplest experiment for this purpose is the analysis of reverse-transcription (RT) stops or pauses induced by the binding of the sRNA to its cognate mRNA. This analysis may provide an idea of the position of the 3′ end of the duplex on the mRNA side and may also indicate whether several pairing-regions exist. However, not all sRNA-mRNA duplexes are sufficiently stable to give rise to such signals and the absence of RT-stops upon sRNA binding is thus not sufficient to rule out a potential pairing. Alternatively, a detailed view of the *in vitro* secondary structure of the sRNA and mRNA, alone or in complex, can be provided by RNA secondary structure probing, whose principle is shown in Fig. 3.

In most cases, intermolecular RNA-RNA interactions rely on nucleotides that are located in single-stranded regions, such as loops or unstructured regions. Determination of the secondary structure of either the sRNA or the mRNA is thus expected to identify the nucleotides which are potentially available for base pairing, but also those directly involved in the interaction, if probing is performed after duplex formation. RNA secondary structure probing in solution consists of monitoring the reactivity of RNA nucleotides towards different probes, such as chemical reagents or RNases that display different specificities. RNases such as T1 (cleaves after unpaired guanines), T2 (cleaves after unpaired nts, with a preference for adenines), or V1 and RNase III (both of which cleave in double-stranded regions) have been extensively used for probing structure of sRNA, mRNA or sRNA-mRNA complex [49–51]. However, nucleotides ≪buried≫ inside the RNA molecule are difficult for RNases to access (because of their large size) and thus interpretation of probing data with RNases has to be restricted to the positive detection of cleavages rather than the absence of cleavages. Chemical probes can also be used for RNA structure probing. Because they are much smaller than RNases, they can reach almost every part of the RNA molecule and they are thus ideally suited for the study of RNA structure. A first group of chemical reagents is sequence-dependent and forms stable covalent adducts at accessible and reactive Watson-Crick positions with one or two of the four RNA bases. The most popular base-specific probes are DMS, which reacts with adenosine and cytosine residues [52], kethoxal, which modifies guanosine residues [53–55] and CMCT, which probes guanosine and uridine residues [56]. A second group of reagents, including lead(II) [57] or SHAPE reagents such as NMIA or 1M7 [58], targets the ribose-phosphate backbone with the advantage of being thus sequence-independent. They ultimately measure local nucleotide flexibility or dynamics with the assumption that more "flexible" regions (such as single-stranded regions) are modified more efficiently than more "rigid" regions (such as double-stranded regions). However it should be noted that in contrast to SHAPE reagents which form adducts at the 2′ hydroxyl group of the ribose moiety, lead (II) cleavage of the RNA backbone has strong biases towards sites of metal coordination that may make the interpretation of secondary structure measurements ambiguous [59,60]. Finally, in-line probing [61] uses the structure-dependent autocleavage capability of RNA and single-stranded regions are expected to degrade over time as they are more flexible and hence unstable.

Probing methods that interrogate all nucleotides are particularly concise and highly informative because they present two advantages compared to the use of base-specific probes: (i) the reactivities of the RNA nucleotides can be measured for almost every residue in a single experiment while base-specific probes are typically limited in their range of selectivity and should be used in parallel and (ii) the reactivities can be quantitatively compared. For more detailed information about the chemistries of the various modifying reactions, the reader should refer to some excellent reviews or articles describing the use of chemical RNA structure probing techniques [58,61–65].

In addition to precisely mapping the interaction regions for a given RNA-RNA duplex, *in vitro* RNA probing can provide key insights about possible structural rearrangement of the mRNA following sRNA binding. In some cases, the existence of such alternative structures is essential for the action of the sRNA, possibly

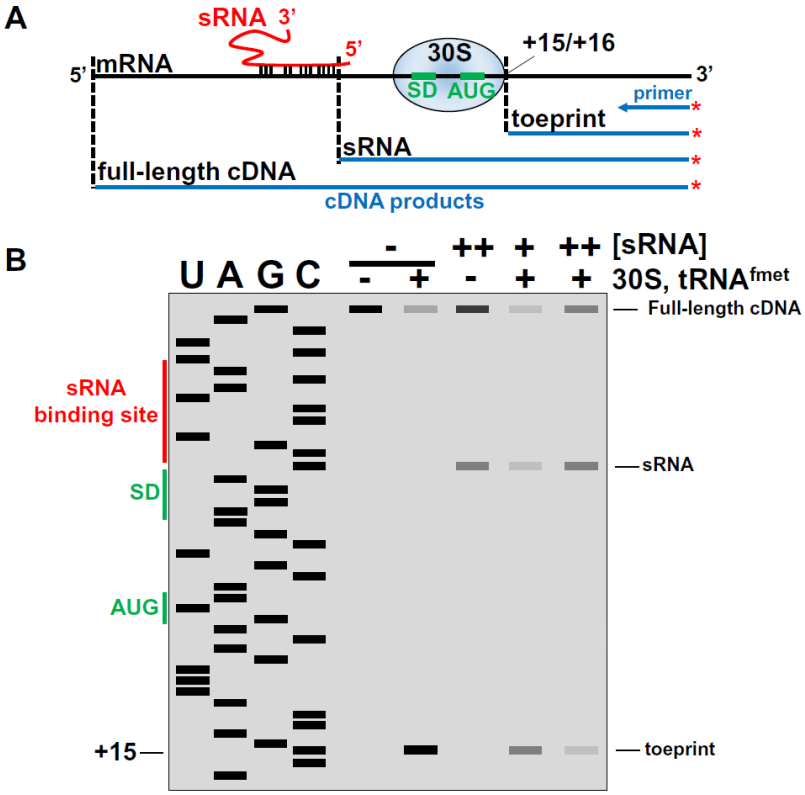


Fig. 4. Principle of the toeprinting experiment. (A) Translation signals are indicated in green, and positions of the toeprint signal are given relative to the translation start. Different cDNA produced by RT from a labeled DNA primer are shown as blue lines. (B) Scheme of the fractionation of the RT products after sequencing PAGE. In this example, sRNA binding to the mRNA inhibits the toeprint signal in a concentration-dependent manner.

because they will ultimately affect the accessibility of the RBS [66]. Similar studies have also been performed upon addition of Hfq to determine its role in mRNA or sRNA structure [67,68].

Interestingly, some of the chemical probes described above, such as DMS, lead (II) or SHAPE reagents, can enter living cells and thus be used for characterization of RNA structures *in vivo* [54,65,69,70]. RNA structure characterization has undergone critical advances recently by combining chemical probing with high-throughput sequencing, providing comprehensive insights into the RNA structure of most transcripts of a given organism in a single experiment [69,71]. The "RNA structurome" of some bacteria was for instance determined using RNase probing [72,73] and this is likely to be repeated soon in more bacterial species and/or using differents probing reagents, possibly *in vivo*. It is thus advisable to take advantage of these recent data sets (if applicable) before starting RNA structure probing experiments, which can be tedious and time-consuming.

3.3. In vitro study of translation initiation: the toeprinting assay

Many sRNAs act at the level of translation initiation by affecting binding of the 30S ribosomal subunit to the target mRNA. Therefore, one of the first mechanisms that should be investigated, when studying sRNAs that modify translation of targets, is inhibition of ribosome binding to the mRNA. For this purpose, a straightforward experiment to perform is toeprinting. Basically, toeprinting is a primer extension inhibition assay which is used to examine the formation of the translation initiation ternary complex formed between the mRNA, the 30S ribosomal subunit and the initiator $tRNA^{f\text{-}Met}$ [74]. This assay relies on the fact that reverse transcriptase is "stopped" (usually at nts 15/16 relative to the first nt of the ORF) when it encounters a ribosome sitting at the RBS, thereby allowing to determine the precise position of the ribosome on mRNA (Fig. 4). When a toeprint assay is performed in presence of sRNAs, those have to be incubated with the mRNA prior to the addition of the 30S ribosomal subunit and the initiator tRNA, as they are generally unable to displace the ternary initiation complex. They are thus usually added after denaturation and refolding of the mRNA [75], but can also be denatured and renatured together with the mRNA, which should favor sRNA-mRNA duplex formation. Note also that other factors, e.g. Hfq, might be required for toeprinting inhibition by sRNAs. This is either because it is directly involved in the competition with ribosome binding [42], or because it can increase the formation or the stability of the sRNA-mRNA duplex to more efficiently compete with ternary complex (our unpublished results).

Some essential controls must be included in the toeprinting assay: (i) a control performed in the absence of both the ribosome and the sRNA will determine the premature primer extension arrests that are due to stable mRNA secondary structure or cleavages; (ii) a control lacking ribosomes but containing the sRNA at the highest concentration used in the assay will determine the primer extension arrests that are due to base pairing of the sRNA with mRNA whenever detectable; (iii) a control performed using a mutant version of the sRNA, which has lost its regulatory effect *in vivo* and is expected not to affect toeprinting *in vitro*, will ensure that the observed effect is specific.

In most instances, sRNAs decrease the toeprint signal. In such cases, if the sRNA interacts with the mRNA at or close to the RBS, signal reduction is a strong indication that it acts *via* competition with the ribosome to bind the mRNA. If it interacts instead with a region located outside of the RBS, a possible scenario is that the sRNA-mRNA interaction triggers a rearrangement of the mRNA secondary structure, which will ultimately impede access of the ribosome to the RBS [66]. In this regard, probing the structure of

the mRNA following sRNA binding (see above) can be highly informative regarding the occurrence of such a structural switch. If, in contrast, the outcome of sRNA binding is an increase in the toeprint signal, it is likely that the sRNA base pairs with an mRNA region located outside of the RBS and a plausible mechanism is disruption of an inhibitory secondary structure that sequesters the RBS of the target mRNA by the sRNA. This is a classical and well-documented mechanism that is used by several sRNAs to activate gene expression [76–78]. In this scenario also, probing mRNA structure in the context of an sRNA-mRNA duplex should be informative. Finally, if the sRNA has an effect on the toeprint signal without binding at or in the vicinity of the RBS and without indication of a structural switch, alternative mechanisms have to be envisioned and further elucidated. For instance, the sRNA can instead compete with ribosomes for binding to a "ribosome standby" site located far upstream of the RBS of the target mRNA [79], or control expression of the target mRNA by translational coupling after acting on the expression of an upstream ORF [80].

3.4. In vitro translation experiments

Cell-free protein synthesis using bacterial extracts can be used to recapitulate control by an sRNA, as well as to study the effect of specific factors and to determine the level at which control takes place [50]. This kind of approach has become more popular with the development of an *in vitro* protein synthesis assay including T7 RNA polymerase, ribosomes and the other purified components of the *E. coli* translation machinery [81]. Although this assay is best suited to the study of the sRNA-mRNA pairs existing in *E. coli*, it might also be adaptable to the study of sRNA-mRNA pairs from other bacteria as long as the mRNA under study is translated using the *E. coli* machinery.

Using this assay, one can follow the effect of a regulatory sRNA on the expression of a given gene by analyzing the synthesis of the encoded protein either by protein gel staining, western-blotting, or ^{35}S labeling, depending on translation efficiency. As previously, an essential control that must be performed is looking at translation in the presence of a mutated sRNA that is unable to regulate the translation of the target mRNA *in vivo*.

This *in vitro* assay allows mRNA transcription and protein synthesis from a template DNA carrying a T7 promoter. If an sRNA effect on protein synthesis is detected with such a template, the use of a heterologous promoter strongly suggests that the sRNA does not act at the promoter level. It is also possible to infer whether the effect is transcriptional or translational from *in vivo* data provided by the use of fusions with reporter genes (see above). It must be realized that transcription by the T7 RNA polymerase is much faster than transcription by the *E. coli* enzyme [22], which will result in an uncoupling of transcription and translation. This might impair regulatory mechanisms that require transcription and translation to be coupled, and lead to formation of mRNA structures that differ from the native ones, possibly preventing the action of sRNAs. Alternatively, an mRNA template can be used instead. Even if it is also synthetised by the T7 RNA polymerase, it can be denatured and refolded prior to be added to this assay, which will make more unlikely the possibility of unwanted RNA structures. Another advantage of using an mRNA template is that it can help to precisely monitor mRNA levels and the sRNA-mRNA ratios used in the experiment.

If an sRNA is shown to affect the translation of a given gene in this assay, it could be due to an effect either on translation initiation or at a later step. To discriminate between these possibilities, data obtained from toeprinting experiments could be highly informative in supporting or dismissing translation initiation as the key-step of control. In the event where it seems that the sRNA controls translation elongation or termination (even though no such case has been reported so far), it would be interesting to look for a truncated polypeptide resulting from premature translation termination or elongation pausing triggered by sRNA.

Because no ribonucleases are present in the commercialized *in vitro* translation assay, another important question this test can help address is the relative contribution of translation and mRNA degradation in the control of gene expression by an sRNA, which, as mentioned above, is extremely difficult to determine by *in vivo* analyses. Alternatively, this can also be deciphered by the reverse approach of studying how an sRNA affects the target mRNA degradation or processing *in vitro*, in the absence of translation.

3.5. Studying the degradation or processing of sRNA-mRNA duplexes in vitro

Whenever an sRNA is suspected to affect the stability and/or the processing of a target mRNA, it is worth undertaking an *in vitro* study of the mRNA fate, as well as that of the sRNA, in order (i) to determine whether this change in stability is a consequence of a change in translation or, in contrast, occurs in the absence of translation as well, (ii) to identify the nuclease(s) involved and possibly (iii) to uncover other mechanistic details. One can for instance follow the cleavage of a labeled target mRNA by a purified RNase, in the presence or absence of the regulatory sRNA. In this way, it was shown that the MicC sRNA induced a specific cleavage of *ompD* mRNA by the catalytic domain of RNase E after pairing within the *ompD* ORF [82]. Furthermore, the addition of MicC variants with different 5′ ends in these experiments showed that the 5′ monophosphorylated MicC was more efficient at inducing cleavage of *ompD* than the triphosphorylated form, suggesting that, in addition to driving the RNase E cleavage close to the site of interaction, MicC also activated the enzyme by providing a 5′ monophosphorylated end. This *in vitro* assay thus confirmed that degradation of an sRNA-targeted mRNA was independent of translation and provided mechanistic insights into how the degradation takes place.

In vitro cleavage assays have also proven useful in the study of positively acting sRNAs: for instance, the activation of the sugar phosphatase encoding *yigL* gene by the SgrS sRNA is also an RNase E-dependent process [30]. In this case, SgrS binding in the upstream ORF of the *pldB-yigL* bicistronic operon was shown to stabilize the *yigL* mRNA by preventing cleavage by RNase E at a critical site. *In vitro* cleavage reactions aiming at understanding the role of RNA processing in the action of sRNAs have also been performed with purified RNA degradosomes [27], with the RNase III endonuclease [79], or with the PNPase 3′–5′ exonuclease [83], and are likely to be developed for other exo- or endonucleases. As for all *in vitro* approaches, these analyses have to be interpreted with caution and results must be supported by *in vivo* data. An *in vitro* cleavage of an sRNA-mRNA duplex by RNase III could for example simply be due to the fact that this endonuclease recognizes and cleaves double-stranded structures (which is the property that is used for RNA structural probing). It is thus essential in this case to show that a deletion of the *rnc* gene also affects control, or at least the levels of the target mRNA *in vivo* [see e.g. 79,84].

4. Conclusion

The different approaches discussed in this review can be generally applied to the study of the regulation of gene expression by bacterial base pairing sRNAs. So far, they have provided an in-depth understanding of the molecular mechanism underlying the action of sRNAs for a few sRNA-mRNA pairs in only a small number of bacteria. In the future, there is no doubt that their application, certainly with some adjustments, to other sRNAs from a larger

set of bacteria, will lead to uncover yet other mechanisms. It is also likely that the use of these techniques in conjunction with high-throughput sequencing, aiming for instance to the determination of RNA structuromes, will become more and more widespread. This will be highly interesting as the availability of more experimental data will give us a better sense of how reliable these approaches are.

The techniques described here can also be useful to study how other regulatory RNAs, including the antisense RNAs, that are widely present in all genomes, regulate gene expression. Because the perfect complementarity between an antisense RNA and its target might render the analysis more difficult, the mode of action of antisense RNAs is still extremely poorly characterized in most cases. Overall, mechanistic studies are likely to provide exciting results in the years to come.

Funding

Work in our lab is supported by the CNRS, ANR grant (ANR-14-CE10-0004-01) to MG and by the "Initiative d'Excellence" program from the French State (Grant "DYNAMO", ANR-11-LABX-0011). JJ is the recipient of a "Labex Dynamo" Ph.D. fellowship.

Acknowledgements

We are grateful to Ciaran Condon, Sylvain Durand and Mathias Springer for critical reading of the manuscript.

References

[1] J.T. Wade, D.C. Grainger, Pervasive transcription: illuminating the dark matter of bacterial transcriptomes, Nat. Rev. Microbiol. 12 (2014) 647–653.
[2] T.R. Cech, J.A. Steitz, The noncoding RNA revolution-trashing old rules to forge new ones, Cell 157 (2014) 77–94.
[3] E.G. Wagner, P. Romby, Small RNAs in bacteria and archaea: who they are, what they do, and how they do it, Adv. Genet. 90 (2015) 133–208.
[4] M.S. Guo, T.B. Updegrove, E.B. Gogol, S.A. Shabalina, C.A. Gross, et al., MicL, a new sigmaE-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein, Genes Dev. 28 (2014) 1620–1634.
[5] Y. Chao, K. Papenfort, R. Reinhardt, C.M. Sharma, J. Vogel, An atlas of Hfq-bound transcripts reveals 3′ UTRs as a genomic reservoir of regulatory small RNAs, EMBO J. 31 (2012) 4005–4019.
[6] Y. Chao, J. Vogel, A 3′ UTR-derived small RNA provides the regulatory noncoding arm of the inner membrane stress response, Mol. Cell 61 (2016) 352–363.
[7] D. Lalaouna, M.C. Carrier, S. Semsey, J.S. Brouard, J. Wang, et al., A 3′ external transcribed spacer in a tRNA transcript acts as a sponge for small RNAs to prevent transcriptional noise, Mol. Cell 58 (2015) 393–405.
[8] Y. Hao, T.B. Updegrove, N.N. Livingston, G. Storz, Protection against deleterious nitrogen compounds: role of sigmaS-dependent small RNAs encoded adjacent to *sdiA*, Nucleic Acids Res. 44 (2016) 6935–6948.
[9] B.J. Tucker, R.R. Breaker, Riboswitches as versatile gene control elements, Curr. Opin. Struct. Biol. 15 (2005) 342–348.
[10] L.S. Waters, G. Storz, Regulatory RNAs in bacteria, Cell 136 (2009) 615–628.
[11] S. Melamed, A. Peer, R. Faigenbaum-Romm, Y.E. Gatt, N. Reiss, et al., Global Mapping of Small RNA-Target Interactions in Bacteria, Mol. Cell 63 (2016) 884–897.
[12] X. Wang, S.C. Ji, H.J. Jeon, Y. Lee, H.M. Lim, Two-level inhibition of *galK* expression by Spot42: degradation of mRNA mK2 and enhanced transcription termination before the *galK* gene, Proc. Natl. Acad. Sci. U.S.A. 112 (2015) 7581–7586.
[13] N. Sedlyarova, I. Shamovsky, B.K. Bharati, V. Epshtein, J. Chen, sRNA-Mediated Control of Transcription Termination in *E. coli*, Cell 167 (2016) 111–121 e13.
[14] N. Figueroa-Bossi, M. Valentini, L. Malleret, F. Fiorini, L. Bossi, Caught at its own game: regulatory small RNA inactivated by an inducible transcript mimicking its target, Genes Dev. 23 (2009) 2004–2015.
[15] N. De Lay, S. Gottesman, Role of polynucleotide phosphorylase in sRNA function in *Escherichia coli*, RNA 17 (2011) 1172–1189.
[16] N. Grantcharova, V. Peters, C. Monteiro, K. Zakikhany, U. Romling, Bistable expression of CsgD in biofilm development of *Salmonella enterica* serovar typhimurium, J. Bacteriol. 192 (2010) 456–466.
[17] J.H. Urban, J. Vogel, Translational control and target recognition by *Escherichia coli* small RNAs in vivo, Nucleic Acids Res. 35 (2007) 1018–1037.
[18] C.P. Corcoran, D. Podkaminski, K. Papenfort, J.H. Urban, J.C. Hinton, et al., Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA, Mol. Microbiol. 84 (2012) 428–445.
[19] M. Guillier, S. Gottesman, The 5′ end of two redundant sRNAs is involved in the regulation of multiple targets, including their own regulator, Nucleic Acids Res. 36 (2008) 6781–6794.
[20] K. Papenfort, M. Bouvier, F. Mika, C.M. Sharma, J. Vogel, Evidence for an autonomous 5′ target recognition domain in an Hfq-associated small RNA, Proc. Natl. Acad. Sci. U.S.A. 107 (2010) 20435–20440.
[21] M. Bouvier, C.M. Sharma, F. Mika, K.H. Nierhaus, J. Vogel, Small RNA binding to 5′ mRNA coding region inhibits translational initiation, Mol. Cell 32 (2008) 827–837.
[22] O.V. Makarova, E.M. Makarov, R. Sousa, M. Dreyfus, Transcribing of *Escherichia coli* genes with mutant T7 RNA polymerases: stability of *lacZ* mRNA inversely correlates with polymerase speed, Proc. Natl. Acad. Sci. U.S.A. 92 (1995) 12250–12254.
[23] A. Deana, J.G. Belasco, Lost in translation: the influence of ribosomes on bacterial mRNA decay, Genes Dev. 19 (2005) 2526–2533.
[24] M. Dreyfus, Killer and protective ribosomes, Prog. Mol. Biol. Transl. Sci. 85 (2009) 423–466.
[25] T. Morita, Y. Mochizuki, H. Aiba, Translational repression is sufficient for gene silencing by bacterial small noncoding RNAs in the absence of mRNA destruction, Proc. Natl. Acad. Sci. U.S.A. 103 (2006) 4858–4863.
[26] H. Kawamoto, T. Morita, A. Shimizu, T. Inada, H. Aiba, Implications of membrane localization of target mRNA in the action of a small RNA: mechanism of post-transcriptional regulation of glucose transporter in *Escherichia coli*, Genes Dev. 19 (2005) 328–338.
[27] K. Prevost, G. Desnoyers, J.F. Jacques, F. Lavoie, E. Masse, Small RNA-induced mRNA degradation achieved through both translation block and activated cleavage, Genes Dev. 25 (2011) 385–396.
[28] J. Wang, W. Rennie, C. Liu, C.S. Carmack, K. Prevost, et al., Identification of bacterial sRNA regulatory targets using ribosome profiling, Nucleic Acids Res. 43 (2015) 10308–10320.
[29] E. Huntzinger, S. Boisset, C. Saveanu, Y. Benito, T. Geissmann, et al., *Staphylococcus aureus* RNAIII and the endoribonuclease III coordinately regulate *spa* gene expression, EMBO J. 24 (2005) 824–835.
[30] K. Papenfort, Y. Sun, M. Miyakoshi, C.K. Vanderpool, J. Vogel, Small RNA-mediated activation of sugar phosphatase mRNA regulates glucose homeostasis, Cell 153 (2013) 426–437.
[31] N. Obana, Y. Shirahama, K. Abe, K. Nakamura, Stabilization of *Clostridium perfringens* collagenase mRNA by VR-RNA-dependent cleavage in 5′ leader sequence, Mol. Microbiol. 77 (2010) 1416–1428.
[32] D. Lalaouna, M. Simoneau-Roy, D. Lafontaine, E. Masse, Regulatory RNAs and target mRNA decay in prokaryotes, Biochim. Biophys. Acta 2013 (1829) 742–747.
[33] S. Durand, A. Tomasini, F. Braun, C. Condon, P. Romby, SRNA and mRNA turnover in Gram-positive bacteria, FEMS Microbiol. Rev. 39 (2015) 316–330.
[34] J. Vogel, B.F. Luisi, Hfq and its constellation of RNA, Nat. Rev. Microbiol. 9 (2011) 578–589.
[35] F. Kalamorz, B. Reichenbach, W. Marz, B. Rak, B. Gorke, Feedback control of glucosamine-6-phosphate synthase GlmS expression depends on the small RNA GlmZ and involves the novel protein YhbJ in *Escherichia coli*, Mol. Microbiol. 65 (2007) 1518–1533.
[36] Y. Gopel, K. Papenfort, B. Reichenbach, J. Vogel, B. Gorke, Targeted decay of a regulatory small RNA by an adaptor protein for RNase E and counteraction by an anti-adaptor RNA, Genes Dev. 27 (2013) 552–564.
[37] N. Said, R. Rieder, R. Hurwitz, J. Deckert, H. Urlaub, et al., In vivo expression and purification of aptamer-tagged small RNA regulators, Nucleic Acids Res. 37 (2009) e133.
[38] A. Smirnov, K.U. Forstner, E. Holmqvist, A. Otto, R. Günster, et al., Grad-seq guides the discovery of ProQ as a major small RNA binding protein, Proc. Natl. Acad. Sci. U.S.A. (2016), in press. http://dx.doi.org/10.1073/pnas.1609981113.
[39] A. Zhang, D.J. Schu, B.C. Tjaden, G. Storz, S. Gottesman, Mutations in interaction surfaces differentially impact *E. coli* Hfq association with small RNAs and their mRNA targets, J. Mol. Biol. 425 (2013) 3678–3697.
[40] E. Holmqvist, P.R. Wright, L. Li, T. Bischler, L. Barquist, et al., Global RNA recognition patterns of post-transcriptional regulators Hfq and CsrA revealed by UV crosslinking *in vivo*, EMBO J. 35 (2016) 991–1011.
[41] L. Attaiech, A. Boughammoura, C. Brochier-Armanet, O. Allatif, F. Peillard-Fiorente, et al., Silencing of natural transformation by an RNA chaperone and a multitarget small RNA, Proc. Natl. Acad. Sci. U.S.A. 113 (2016) 8813–8818.
[42] G. Desnoyers, E. Masse, Noncanonical repression of translation initiation through small RNA recruitment of the RNA chaperone Hfq, Genes Dev. 26 (2012) 726–739.
[43] A. Coornaert, C. Chiaruttini, M. Springer, M. Guillier, Post-transcriptional control of the *Escherichia coli* PhoQ-PhoP two-component system by multiple sRNAs involves a novel pairing region of GcvB, PLoS Genet. 9 (2013) e1003156.
[44] P. Davanloo, A.H. Rosenberg, J.J. Dunn, F.W. Studier, Cloning and expression of the gene for bacteriophage T7 RNA polymerase, Proc. Natl. Acad. Sci. U.S.A. 81 (1984) 2035–2039.
[45] S. Tabor, C.C. Richardson, A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes, Proc. Natl. Acad. Sci. U.S.A. 82 (1985) 1074–1078.
[46] C. Cazenave, O.C. Uhlenbeck, RNA template-directed RNA synthesis by T7 RNA polymerase, Proc. Natl. Acad. Sci. U.S.A. 91 (1994) 6972–6976.
[47] G. Di Tomasso, A. Salvail-Lacoste, J. Bouvette, J.G. Omichinski, P. Legault, Affinity purification of *in vitro* transcribed RNA with homogeneous ends using a 3′-ARiBo tag, Methods Enzymol. 549 (2014) 49–84.
[48] S. Chen, A. Zhang, L.B. Blyn, G. Storz, MicC, a second small-RNA regulator of Omp protein expression in *Escherichia coli*, J. Bacteriol. 186 (2004) 6689–6697.

[49] C. Romilly, C. Lays, A. Tomasini, I. Caldelari, Y. Benito, et al., A non-coding RNA promotes bacterial persistence and decreases virulence by regulating a regulator in *Staphylococcus aureus*, PLoS Pathog. 10 (2014) e1003979.
[50] E. Holmqvist, J. Reimegard, M. Sterk, N. Grantcharova, U. Romling, et al., Two antisense RNAs target the transcriptional regulator CsgD to inhibit curli synthesis, EMBO J. 29 (2010) 1840–1850.
[51] A. Eyraud, P. Tattevin, S. Chabelskaya, B. Felden, A small RNA controls a protein regulator involved in antibiotic resistance in *Staphylococcus aureus*, Nucleic Acids Res. 42 (2014) 4892–4905.
[52] D.A. Peattie, W. Gilbert, Chemical probes for higher-order structure in RNA, Proc. Natl. Acad. Sci. U.S.A. 77 (1980) 4679–4682.
[53] R. Shapiro, J. Hachmann, The reaction of guanine derivatives with 1,2-dicarbonyl compounds, Biochemistry 5 (1966) 2799–2807.
[54] R. Shapiro, B.I. Cohen, S.J. Shiuey, H. Maurer, On the reaction of guanine with glyoxal, pyruvaldehyde, and kethoxal, and the structure of the acylguanines. A new synthesis of N2-alkylguanines, Biochemistry 8 (1969) 238–245.
[55] H.F. Noller, J.B. Chaires, Functional modification of 16S ribosomal RNA by kethoxal, Proc. Natl. Acad. Sci. U.S.A. 69 (1972) 3115–3118.
[56] N.W. Ho, P.T. Gilham, The reversible chemical modification of uracil, thymine, and guanine nucleotides and the modification of the action of ribonuclease on ribonucleic acid, Biochemistry 6 (1967) 3632–3639.
[57] M. Lindell, P. Romby, E.G. Wagner, Lead(II) as a probe for investigating RNA structure *in vivo*, RNA 8 (2002) 534–541.
[58] E.J. Merino, K.A. Wilkinson, J.L. Coughlan, K.M. Weeks, RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE), J. Am. Chem. Soc. 127 (2005) 4223–4231.
[59] M. Forconi, D. Herschlag, Metal ion-based RNA cleavage as a structural probe, Methods Enzymol. 468 (2009) 91–106.
[60] R.C. Spitale, P. Crisalli, R.A. Flynn, E.A. Torre, E.T. Kool, et al., RNA SHAPE analysis in living cells, Nat. Chem. Biol. 9 (2013) 18–20.
[61] E.E. Regulski, R.R. Breaker, In-line probing analysis of riboswitches, Methods Mol. Biol. 419 (2008) 53–67.
[62] C. Brunel, P. Romby, Probing RNA structure and RNA-ligand complexes with chemical probes, Methods Enzymol. 318 (2000) 3–21.
[63] C. Chiaruttini, F. Allemand, M. Springer, Structural probing of RNA thermosensors, Methods Mol. Biol. 540 (2009) 233–245.
[64] C. Ehresmann, F. Baudin, M. Mougel, P. Romby, J.-P. Ebel, et al., Probing the structure of RNAs in solution, Nucleic Acids Res. 15 (1987) 9109–9128.
[65] K.M. Weeks, Advances in RNA structure analysis by chemical probing, Curr. Opin. Struct. Biol. 20 (2010) 295–304.
[66] N. Heidrich, I. Moll, S. Brantl, *In vitro* analysis of the interaction between the small RNA SR1 and its primary target *ahrC* mRNA, Nucleic Acids Res. 35 (2007) 4331–4346.
[67] Y. Peng, J.E. Curtis, X. Fang, S.A. Woodson, Structural model of an mRNA in complex with the bacterial chaperone Hfq, Proc. Natl. Acad. Sci. U.S.A. 111 (2014) 17134–17139.
[68] V. Bordeau, B. Felden, Curli synthesis and biofilm formation in enteric bacteria are controlled by a dynamic small RNA module made up of a pseudoknot assisted by an RNA chaperone, Nucleic Acids Res. 42 (2014) 4682–4696.
[69] P. Ge, S. Zhang, Computational analysis of RNA structures with chemical probing data, Methods 79–80 (2015) 60–66.
[70] K.M. Weeks, D.M. Mauger, Exploring RNA structural codes with SHAPE chemistry, Acc. Chem. Res. 44 (2011) 1280–1291.
[71] K.E. Watters, A.M. Yu, E.J. Strobel, A.H. Settle, J.B. Lucks, Characterizing RNA structures in vitro and in vivo with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq), Methods 103 (2016) 34–48.
[72] C. Del Campo, A. Bartholomaus, I. Fedyunin, Z. Ignatova, Secondary structure across the bacterial transcriptome reveals versatile roles in mRNA regulation and function, PLoS Genet. 11 (2015) e1005613.
[73] F. Righetti, A.M. Nuss, C. Twittenhoff, S. Beele, K. Urban, et al., Temperature-responsive *in vitro* RNA structurome of Yersinia pseudotuberculosis, Proc. Natl. Acad. Sci. U.S.A. 113 (2016) 7237–7242.
[74] D. Hartz, D.S. McPheeters, R. Traut, L. Gold, Extension inhibition analysis of translation initiation complexes, Methods Enzymol. 164 (1988) 419–425.
[75] P. Fechter, C. Chevalier, G. Yusupova, M. Yusupov, P. Romby, et al., Ribosomal initiation complexes probed by toeprinting and effect of trans-acting translational regulators in bacteria, Methods Mol. Biol. 540 (2009) 247–263.
[76] N. Majdalani, C. Cunning, D. Sledjeski, T. Elliott, S. Gottesman, DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription, Proc. Natl. Acad. Sci. U.S.A. 95 (1998) 12462–12467.
[77] N. Majdalani, D. Hernandez, S. Gottesman, Regulation and mode of action of the second small RNA activator of RpoS translation, RprA, Mol. Microbiol. 46 (2002) 813–826.
[78] L. Archambault, J. Linscott, N. Swerdlow, K. Boyland, E. Riley, et al., Translational efficiency of *rpoS* mRNA from *Borrelia burgdorferi*: effects of the length and sequence of the mRNA leader region, Biochem. Biophys. Res. Commun. 433 (2013) 73–78.
[79] F. Darfeuille, C. Unoson, J. Vogel, E.G. Wagner, An antisense RNA inhibits translation by competing with standby ribosomes, Mol. Cell 26 (2007) 381–392.
[80] B. Vecerek, I. Moll, U. Blasi, Control of Fur synthesis by the non-coding RNA RyhB and iron-responsive decoding, EMBO J. 26 (2007) 965–975.
[81] Y. Shimizu, A. Inoue, Y. Tomari, T. Suzuki, T. Yokogawa, et al., Cell-free translation reconstituted with purified components, Nat. Biotechnol. 19 (2001) 751–755.
[82] K.J. Bandyra, N. Said, V. Pfeiffer, M.W. Gorna, J. Vogel, et al., The seed region of a small RNA drives the controlled destruction of the target mRNA by the endoribonuclease RNase E, Mol. Cell 47 (2012) 943–953.
[83] K.J. Bandyra, D. Sinha, J. Syrjanen, B.F. Luisi, N.R. De Lay, The ribonuclease polynucleotide phosphorylase can interact with small regulatory RNAs in both protective and degradative modes, RNA 22 (2016) 360–372.
[84] J. Vogel, L. Argaman, E.G. Wagner, S. Altuvia, The small RNA IstR inhibits synthesis of an SOS-induced toxic peptide, Curr. Biol. 14 (2004) 2271–2276.
[85] A. Huttenhofer, H.F. Noller, Footprinting mRNA-ribosome complexes with chemical probes, EMBO J. 13 (1994) 3892–3901.
[86] P. Mandin, S. Gottesman, Integrating anaerobic/aerobic sensing and the general stress response through the ArcZ small RNA, EMBO J. 29 (2010) 3094–3107.
[87] C.A. McCullen, J.N. Benhammou, N. Majdalani, S. Gottesman, Mechanism of positive regulation by DsrA and RprA small noncoding RNAs: pairing increases translation and protects rpoS mRNA from degradation, J Bacteriol. 192 (2010) 5559–5571.
[88] T. Song, F. Mika, B. Lindmark, Z. Liu, S. Schild, et al., A new Vibrio cholerae sRNA modulates colonization and affects release of outer membrane vesicles, Mol. Microbiol. 70 (2008) 100–111.
[89] C. Chevalier, S. Boisset, C. Romilly, B. Masquida, P. Fechter, et al., Staphylococcus aureus RNAIII binds to two distant regions of coa mRNA to arrest translation and promote mRNA degradation, PLoS Pathog. 6 (2010) e1000809.
[90] V. Pfeiffer, K. Papenfort, S. Lucchini, J.C. Hinton, J. Vogel, Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation, Nat. Struct. Mol. Biol. 16 (2009) 840–846.

d. ARN « éponges »

Parmi les multiples mécanismes de régulation impliquant des ARN, un mécanisme tout à fait original de régulation implique des ARN « éponges », qui interagissent avec des régulateurs. Ce faisant, ils empêchent l'interaction classique du régulateur avec ses cibles, entraînant une dérégulation de ces dernières. Hormis les « éponges » à protéine tels que CsrB et CsrC évoqués précédemment, il existe également des « éponges » à ARN, interagissant avec des ARN par complémentarité de séquence (Bossi and Figueroa-Bossi, 2016). C'est notamment le cas de l'ARNm *chbBCARFG*, qui interagit avec le sRNA Hfq-dépendant ChiX et le déstabilise, permettant ainsi la dérégulation de l' ARNm *chiP* habituellement ciblé par ChiX (Figueroa-Bossi et al., 2009). Plus récemment, il a été démontré que l'ARN Hfq-dépendant 3'ETS^{leuZ} pouvait séquestrer les sRNA Hfq-dépendants RybB et RyhB, afin de les maintenir inactifs dans des conditions où leur action ne serait pas requise (Lalaouna et al., 2015). Cet ARN 3'ETS^{leuZ} provient du clivage par la RNase E de l'extrémité 3' de l'ARN *glyW-cysT-leuZ*, combinant trois ARNt non maturés à la suite. Cela montre que ces ARN de séquestration peuvent avoir des origines extrêmement variées, illustrant le fait que l'évolution peut façonner des ARN régulateurs à partir d'ARN ayant des fonctions originelles différentes.

Chapitre III : Adaptation des entérobactéries à l'environnement, réponse au stress et ARN régulateurs

L'évolution, depuis plus de trois milliards d'années, a conféré aux organismes vivants une capacité d'adaptation à leur environnement d'une grande sophistication. C'est en particulier vrai pour les entérobactéries, qui sont confrontées à des milieux d'une extrême variété entre les divers organes de leurs multiples hôtes et l'environnement extérieur. De nombreux mécanismes permettent ainsi à ces bactéries de s'accommoder à des bouleversements environnementaux soudains et de survivre à une multitude de stress différents avec une étonnante résilience. Seront abordés dans cette partie les différents éléments de la cellule permettant de « sentir » ces stress et d'y répondre. J'introduirai plus en détail les divers acteurs ARN et protéiques auxquels j'ai pu m'intéresser au cours de ce travail de thèse, qui sont tous impliqués d'une façon ou d'une autre dans la réponse au stress chez *E. coli*.

A. L'enveloppe cellulaire des entérobactéries

Parce qu'elle est en contact direct avec l'environnement, l'enveloppe cellulaire des entérobactéries est logiquement au centre des échanges entre la cellule et le milieu extérieur. C'est également le plus souvent par son biais que les bactéries ressentent les stimuli extérieurs, avant que ceux-ci n'aboutissent à la régulation de l'expression de certains gènes et à une réponse au stress appropriée. Les principaux éléments entrant dans la composition de l'enveloppe cellulaire des entérobactéries, et plus généralement des bactéries à Gram négatif, seront décrits ici.

1. Composition de l'enveloppe chez les bactéries à Gram négatif

Les bactéries possédant une enveloppe cellulaire composée d'une membrane interne, d'une paroi périplasmique et d'une membrane externe (Fig. 24) sont classifiées en bactéries à Gram négatif. Elles sont nommées ainsi car elles ne retiennent pas la coloration au violet de gentiane après un lavage à l'alcool, un procédé nommé coloration de Gram d'après son auteur. Cette classification se fait par opposition aux bactéries à Gram positif qui ne possèdent

qu'une membrane interne, mais aussi une paroi externe beaucoup plus épaisse et imperméable à l'alcool, leur permettant de retenir la coloration de Gram. Je m'attarderai uniquement sur les bactéries à Gram négatif dont font partie les entérobactéries.

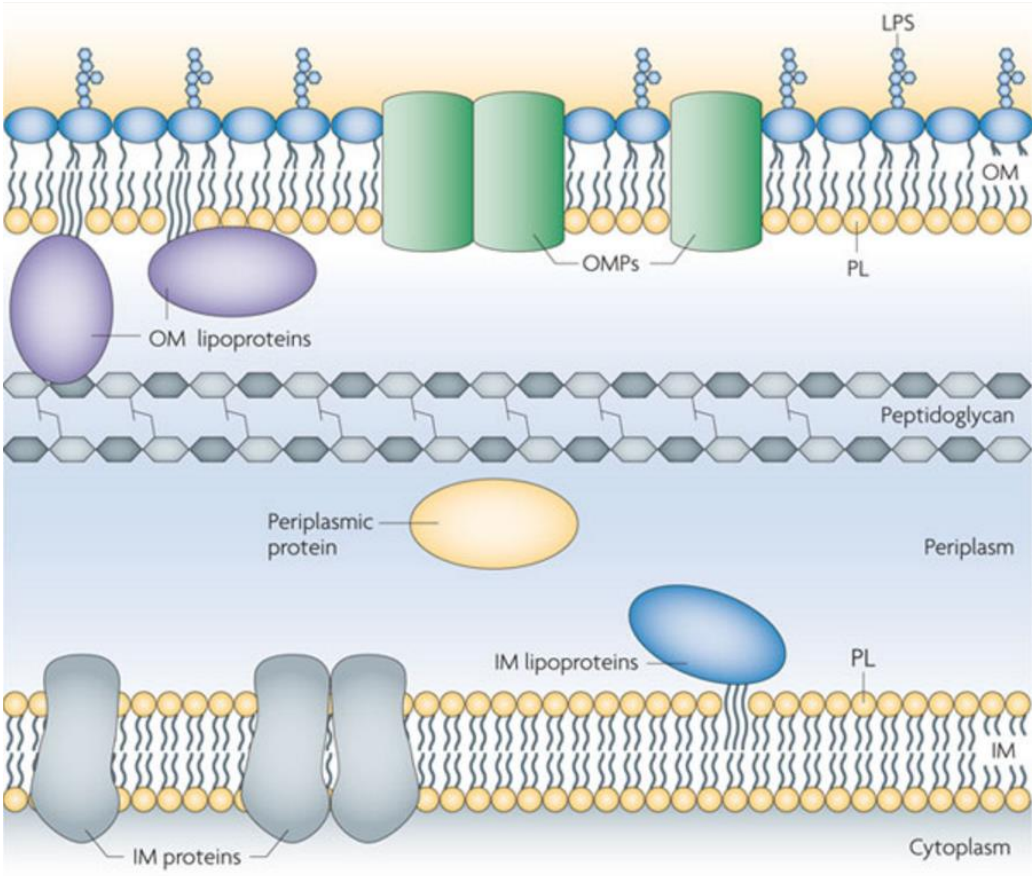


Figure 24 – Composition de l'enveloppe des bactéries à Gram négatif
Sont représentées la membrane interne (IM – « inner membrane ») avec les protéines de membrane interne en gris ; la membrane externe (OM – « outer membrane ») avec les protéines de membrane externe (OMP – « outer membrane protein ») en vert. Les phospholipides (PL) sont les principaux composants de ces membranes, de même que les lipopolysaccharides (LPS) pour le feuillet externe de la membrane externe. On trouve dans le périplasme la paroi peptidoglycane, des protéines périplasmiques ainsi que des lipoprotéines liées à l'une des membranes. D'après (Ruiz et al., 2009).

a. Phospholipides et lipopolysaccharides

L'enveloppe des bactéries à Gram négatif est composée de deux membranes, chacune formée d'une double couche de phospholipides (PL). Il s'agit de molécules amphiphiles caractérisées par une tête hydrophile portant un groupement glycérol-phosphate (Fig. 24, en beige), et une queue hydrophobe lipidique. Ces propriétés sont primordiales pour la formation des membranes, puisque la bicouche de PL est le plus souvent en contact avec l'eau et doit elle-même être imperméable à l'eau et aux ions. Cette dernière fonction est remplie par sa

partie interne hydrophobe formée des chaînes carbonées des PL, tandis que les têtes hydrophiles font face soit au cytoplasme et au périplasme pour la membrane interne, soit au périplasme et au milieu extérieur pour la membrane externe (Fig. 24). Ces PL sont donc essentiels à la compartimentation des cellules et à leur isolation de l'environnement, mais aussi à leur réplication par exemple, puisque seule une synthèse suffisante en PL peut leur permettre de croître en taille jusqu'à être en mesure de se diviser en deux cellules filles.

Les lipopolysaccharides (LPS) portent quant à eux une chaîne de sucres variés ou chaîne polysaccharide en lieu et place du groupement glycérol-phosphate des PL. Ils sont exclusivement localisés au feuillet externe de la membrane externe, où ils jouent un rôle important dans la capacité des bactéries à former des biofilms (voir **Introduction III.A.3.b** pour une définition). Plus généralement, les LPS sont impliqués dans l'interaction entre bactéries et hôtes puisque les divers polysaccharides des LPS peuvent beaucoup varier d'une espèce à l'autre, et être reconnus par le système immunitaire de l'hôte en tant qu'antigènes (Maldonado et al., 2016). Ils permettent parfois même de distinguer certaines souches pathogènes d'espèces autrement commensales, comme c'est le cas chez *E. coli*. Par ailleurs, les LPS sont des constituants essentiels chez la plupart des bactéries à Gram négatif. Leur export à la membrane externe requiert l'intervention d'une des deux seules protéines de membrane externe (OMP pour « outer membrane protein ») essentielles chez les bactéries à Gram négatif, à savoir LptD. Une fois la biosynthèse des LPS achevée, LptD les exporte depuis le périplasme vers l'extérieur de la membrane externe (Chng et al., 2010).

b. Paroi peptidoglycane

Un autre élément important de l'enveloppe des bactéries à Gram négatif est leur paroi peptidoglycane. Celle-ci est composée de longs polymères réguliers de sucres, les glycanes ; en l'occurrence il s'agit ici d'une répétition de disaccharides N-acétyl-glucosamine/N-acétyl-acide muramique. Ces glycanes sont interconnectés par des chaînes latérales peptidiques, formant ainsi un réseau autour de la bactérie, constituant un cytosquelette qui détermine la forme des bactéries (Huang et al., 2008) et supporte les contraintes physiques auxquelles ces bactéries sont confrontées. On note que cette paroi est dans la plupart des cas essentielle et permet par exemple aux bactéries de résister aux chocs osmotiques (Silhavy et al., 2010). La paroi et sa biosynthèse sont de ce fait des cibles privilégiées de nombreux antibiotiques, à

l'image des β-lactames qui inhibent l'action des transpeptidases catalysant la connexion entre les chaînes peptidiques latérales de la paroi peptidoglycane (Schneider and Sahl, 2010).

c. Les protéines membranaires et périplasmiques

Si les membranes confèrent aux bactéries la capacité de s'isoler de l'environnement, il leur faut toutefois un moyen d'interagir avec cet environnement pour en extraire les nutriments nécessaires à leur développement, mais aussi pour en détecter les éventuels changements. Ces fonctions reposent au moins en partie sur de multiples protéines localisées à l'enveloppe cellulaire. Leurs rôles divers sont résumés ici, mais certains cas de figure seront abordés de manière plus détaillée dans la suite de ce chapitre.

De nombreuses protéines membranaires assurent ainsi l'essentiel des échanges avec le milieu extérieur, avec des porines et des récepteurs permettant l'import d'une grande variété d'ions et de nutriments. Les protéines membranaires sont notamment importantes dans la réponse au stress, dans la mesure où réponse efficace peut parfois impliquer une modification conséquente du contenu des membranes en protéines (voir **Introduction III.B et III.C**). Cela peut notamment s'expliquer par la nécessité des bactéries de s'adapter aux divers sucres et ions à sa disposition, tout en limitant l'import de ceux déjà présents en excès dans la cellule.

D'autres protéines peuvent former des appendices externes à l'image des flagelles et des protéines Curli, respectivement impliqués dans la motilité et l'adhésion des cellules (voir **Introduction III.A.3**). Ces complexes protéiques sont d'une grande importance notamment dans les processus de formation de biofilm et pour la virulence de certains pathogènes. Les pili sont également des appendices protéiques, permettant pour leur part l'échange de matériel génétique entre bactéries *via* la conjugaison.

Enfin, de nombreuses protéines périplasmiques sont importantes pour le bon assemblage et l'adressage correct des différents constituants de l'enveloppe cellulaire, ou encore pour la réponse et la résistance au stress. C'est notamment le cas de certaines protéines chaperonnes qui préservent les protéines membranaires sous une forme dénaturée

dans le périplasme, afin qu'elles puissent être correctement adressées à la membrane externe (voir **Introduction III.A.2.b**).

2. OMP, adressage et rôles divers

De nombreuses protéines membranaires sont d'une importance majeure pour la physiologie cellulaire. De même, leur adressage correct est essentiel pour qu'elles puissent assurer leurs fonctions. Il existe divers mécanismes d'adressage des OMP et des protéines de membrane interne. Plutôt que d'aborder la plupart de ces protéines et voies d'adressage d'une façon succincte, j'ai choisi de m'attarder sur les OMP car elles ont un lien plus évident avec le travail effectué au cours de cette thèse.

a. OMP et physiologie de la cellule

La plupart des OMP ont des régions transmembranaires constituées de feuillets β dont l'organisation forme des tonneaux β (Horne and Radford, 2016). Les résidus hydrophobes de ces feuillets β font ainsi face à la partie interne hydrophobe de la membrane, tandis que les résidus hydrophiles sont situés à l'intérieur des tonneaux β. De cette manière, l'intérieur de ces tonneaux forme un conduit perméable à l'eau et à divers nutriments ou molécules spécifiques de chaque protéine membranaire. C'est par leur intermédiaire qu'est assurée l'immense majorité des échanges de molécules entre la bactérie et l'environnement.

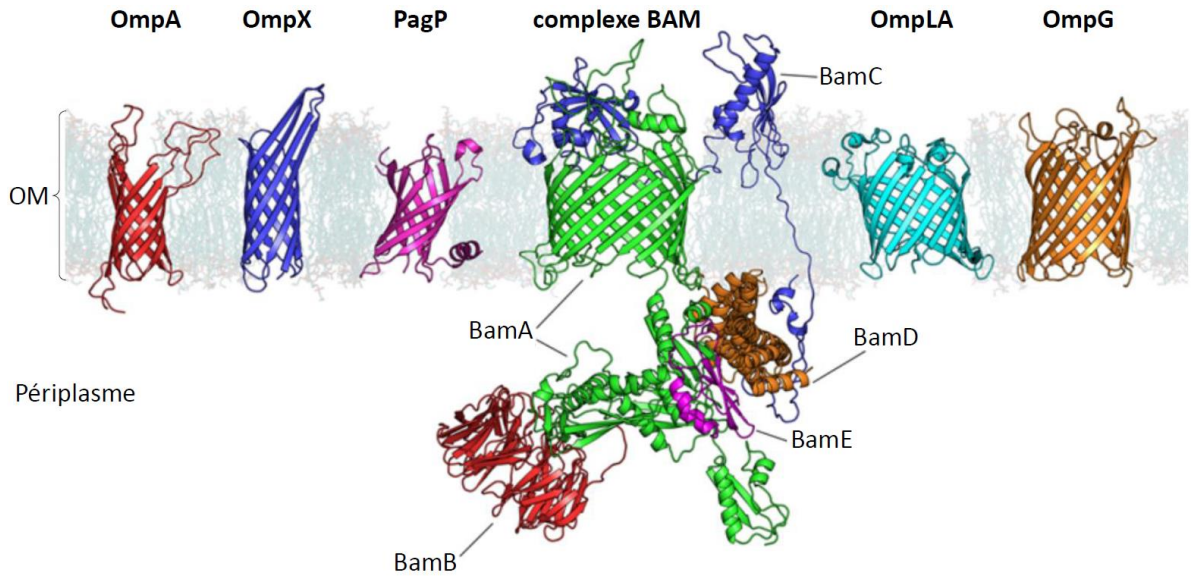


Figure 25 – Diverses OMP à tonneau β chez les bactéries à Gram négatif

Représentation tridimensionnelle de diverses OMP présentes chez les bactéries à Gram négatif, dans leur conformation fonctionnelle au sein de la membrane externe (OM – « outer membrane »). Sont également représentés les différents partenaires protéiques du complexe BAM, en interaction avec l'OMP BamA (en vert). Les flèches colorées au sein des structures protéiques représentent des feuillets β, dont l'assemblage forme des tonneaux β traversant la membrane. D'après (Horne and Radford, 2016).

Chez les entérobactéries, on compte parmi les OMP majeures les porines OmpC, OmpF et OmpA, qui font certainement partie des protéines les plus abondantes de la cellule. Si les porines OmpC et OmpF ont des spécificités relativement larges – c'est-à-dire qu'elles laissent transiter un nombre important de solutés différents – on note cependant que leur spécificité se dirige plus vers les cations que vers les anions (Nikaido, 2003). De plus, OmpF peut laisser transiter des solutés d'une taille plus importante qu'OmpC, ce qui pourrait expliquer les régulations de ces deux protéines dans diverses conditions par le système EnvZ-OmpR (Hall and Silhavy, 1981; Nikaido, 2003) (voir **Introduction III.B.1**). En effet, en condition de forte pression osmotique, une trop grande perméabilité risque d'aboutir à la mort de la cellule qui requiert des concentrations en ions bien précises. Si le rôle d'OmpA en tant que porine est plus discuté, il s'agit d'une protéine exprimée constitutivement chez les entérobactéries (Nikaido, 2003). Cette protéine pourrait toutefois être impliquée dans le processus de formation de biofilms, lors duquel elle est particulièrement abondante (Orme et al., 2006). Ces porines sont par ailleurs suffisamment importantes dans la physiologie des bactéries pour que la délétion de leur gène cause une réduction significative de la virulence, notamment chez des souches pathogènes d'*E. coli* (Hejair et al., 2017).

Néanmoins, la composition de la membrane externe en protéines ne se limite pas à des porines. Il existe par exemple de multiples transporteurs et récepteurs qui ont une spécificité beaucoup plus grande et permettent généralement l'import d'une seule molécule ou complexe de molécules. Ainsi BtuB est un transporteur ayant une grande affinité pour la vitamine B12, tandis que FepA est un récepteur très spécifique des complexes fer-entérobactine (Nikaido, 2003) (voir **Introduction III.C.2.b et III.C.3.b**). Ces diverses OMP sont d'ailleurs souvent également des récepteurs utilisés par les phages pour infecter les bactéries, et la résistance d'une bactérie à un phage donné passe fréquemment par la mutation de l'OMP ciblée par le phage. Enfin, certaines OMP ont d'autres fonctions que le transport de molécules, à l'image de la protéase OmpT ancrée à la membrane externe. OmpT est notamment impliquée dans la réponse au stress *via* la dégradation de protéines dénaturées dans le

périplasme (Kukkonen and Korhonen, 2004) et dans la résistance à des toxines transitant par la membrane externe (Duche et al., 2008).

b. Adressage des OMP à la membrane externe

L'adressage des OMP à la membrane est un processus conservé non seulement chez les bactéries à Gram négatif, mais également chez certaines organelles eucaryotes telles que les mitochondries et les chloroplastes. Cet adressage consiste en une succession d'évènements détaillés ici, assurés par une série de complexes protéiques pour la plupart essentiels (Fig. 26).

Les OMP portent à leur extrémité N-terminale un peptide-signal d'environ 25 aa qui interagit avec une chaperonne nommée « trigger factor » (« facteur de déclenchement », FD). Ce peptide-signal est ensuite être reconnu et pris en charge par les chaperonnes SecB puis SecA, qui permettent l'adressage des OMP au système de translocation (ou translocon) SecYEG ancré à la membrane interne (Paetzel, 2014). S'il semblerait que les OMP soient maintenues dénaturées par les chaperonnes SecA, SecB ou encore FD jusqu'à la fin de leur synthèse dans le cytoplasme avant l'adressage, il se pourrait toutefois que dans certains cas les OMP puissent être adressées au translocon au cours de leur synthèse. Le mécanisme d'adressage à travers SecY est en revanche mieux compris et requiert l'activité ATPase de SecA (Allen et al., 2016). Après le clivage du peptide-signal par la protéase DegP dans le périplasme, les chaperonnes Skp et SurA prennent le relai pour maintenir les OMP dénaturées (Rollauer et al., 2015). Dans le cytoplasme et le périplasme – des milieux aqueux – les domaines transmembranaires hydrophobes des OMP ont tendance à interagir ensemble, c'est pourquoi tant de chaperonnes sont nécessaires afin d'éviter leur agrégation.

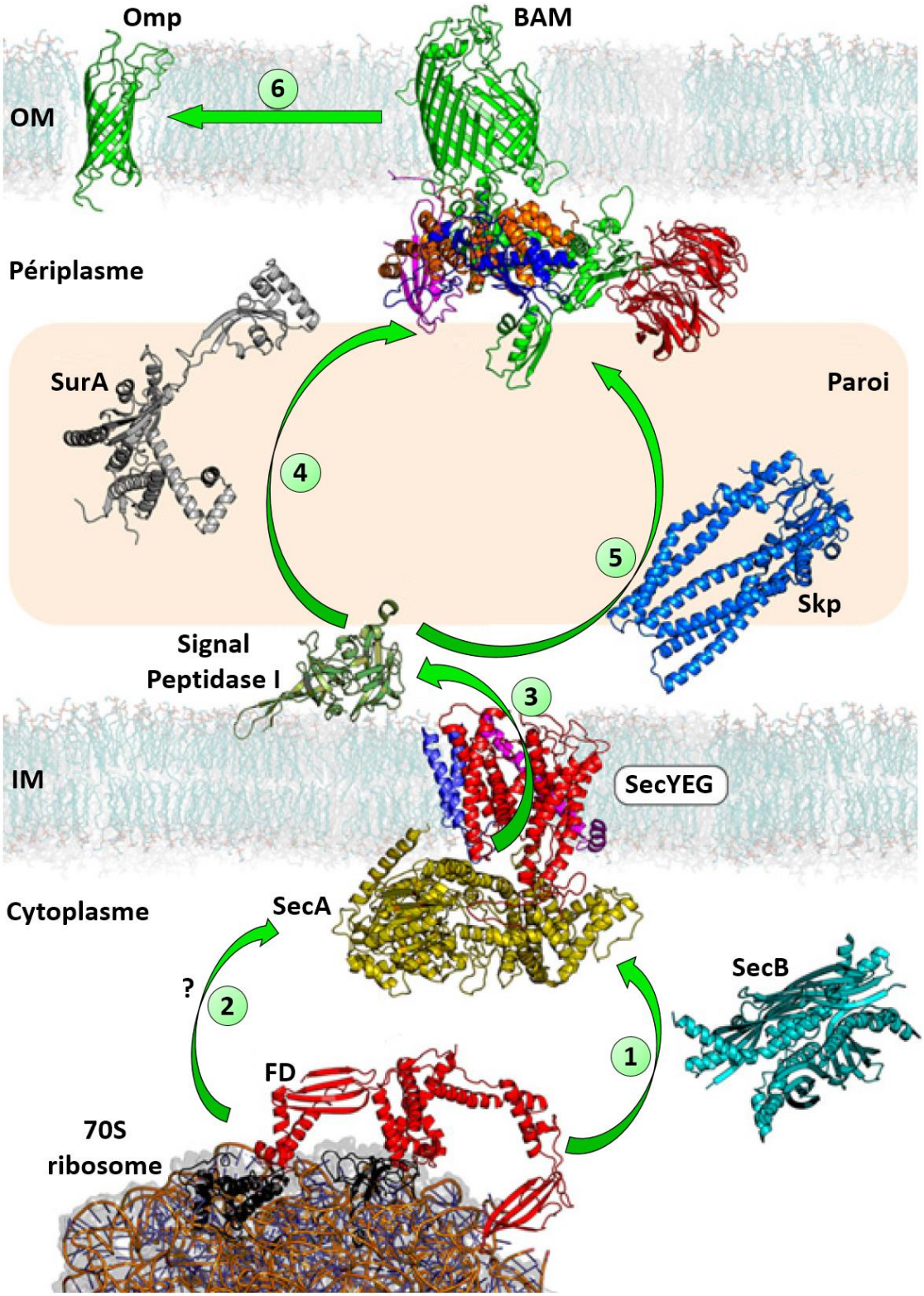


Figure 26 – Adressage des OMP à tonneau β via les systèmes Sec et BAM

Lors de la synthèse d'une OMP, le peptide-signal émergeant du tunnel de sortie du ribosome entre en interaction avec la chaperonne FD. Le peptide-signal est reconnu par la chaperonne SecB (1) ou directement par la chaperonne SecA (2), avant d'être adressé au translocon SecYEG. Celui-ci va procéder à l'export de l'OMP à travers la membrane interne (IM, « inner membrane) vers le périplasme (3), où la signal-peptidase DegP va cliver le peptide-signal. L'OMP est de nouveau prise en charge par une chaperonne périplasmique, à savoir SurA (4) ou Skp (5). S'ensuivra alors la reconnaissance de

l'extrémité C-terminale de l'OMP par le complexe BAM, qui va intégrer cette OMP à la membrane externe (OM, « outer membrane ») via BamA (6). D'après (Horne and Radford, 2016).

Une fois en interaction avec Skp ou SurA, les OMP doivent encore être intégrées correctement à la membrane externe. Pour la plupart des OMP à tonneau β, cette fonction a été attribuée au complexe BAM il y a une dizaine d'années (Kim et al., 2007; Wu et al., 2005) Ce complexe inclut BamA, la seule autre OMP essentielle chez les bactéries à Gram négatif avec LptD mentionnée précédemment, mais aussi les protéines BamB, BamC, BamD et BamE qui semblent toutes importantes pour un adressage optimal des OMP. La structure du complexe BAM a été entièrement résolue plus récemment (Bakelar et al., 2016) (Fig. 25 et 26) et a permis aux auteurs d'imaginer un mécanisme d'intégration des OMP à la membrane externe d'une grande sophistication (Fig. 27).

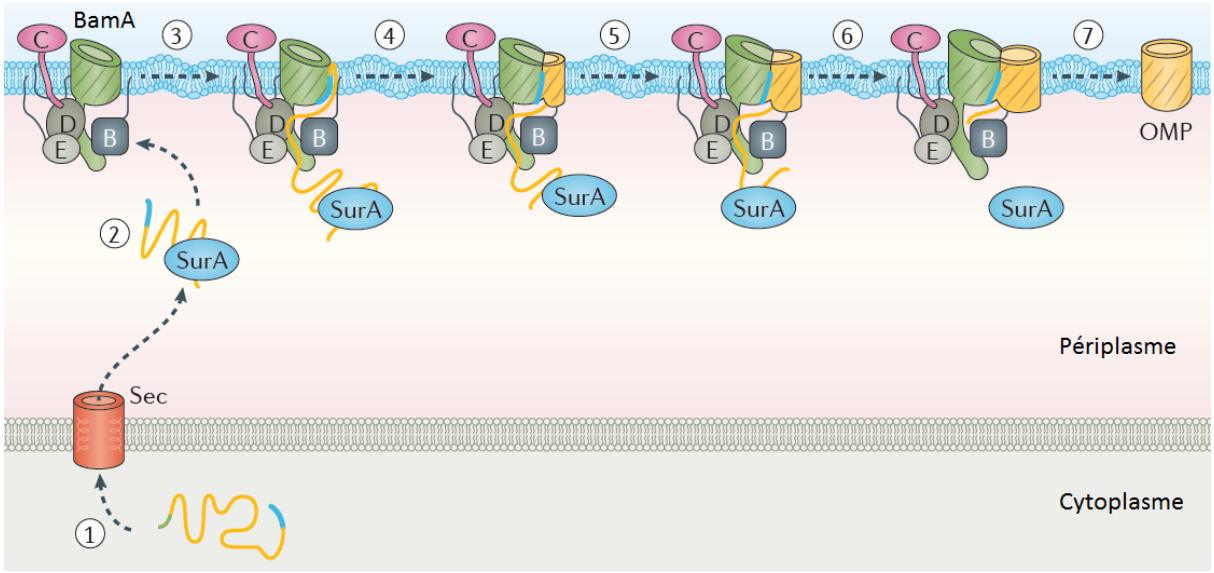


Figure 27 – Modèle de mécanisme d'intégration des OMP à la membrane externe par le complexe BAM

Après l'adressage de l'OMP (en jaune) vers le périplasme par le biais du translocon Sec (1), celle-ci est prise en charge par une chaperonne telle que SurA (2) ou Skp. Elle sera alors adressée au complexe BAM via son extrémité C-terminale qui porte un signal β (en bleu) impliqué dans l'interaction avec BAM (3). L'OMP va ensuite être intégrée à la membrane externe par additions successives de ses feuillets β au tonneau β de BamA (4, 5), et ce grâce aux changements de conformation de BamA imposés par l'ensemble du complexe BAM. Enfin, le complexe BAM va joindre les premier et dernier feuillets β de l'OMP (6), permettant à la nouvelle OMP de se défaire de son interaction avec BamA (7). D'après (Noinaj et al., 2017).

Ainsi, des changements de conformations successifs de BamA permettraient l'intégration des OMP feuillet β par feuillet β au sein du tonneau β de BamA. Puis, lors d'un ultime changement de conformation, les premier et dernier feuillets β de l'OMP à avoir été intégrés seraient rapprochés l'un de l'autre. L'ensemble du tonneau β de l'OMP nouvellement

formée serait ensuite éjecté vers la membrane externe tandis que BamA retrouverait sa conformation initiale (Noinaj et al., 2017) (Fig. 27). Un second modèle a également été proposé par les mêmes auteurs, où le complexe BAM pourrait imposer une torsion ou un rétrécissement local à la bicouche phospholipidique, permettant l'intégration spontanée de certaines OMP à la membrane externe. BamA est donc une protéine essentielle à l'adressage des OMP à la membrane externe chez les bactéries à Gram négatif. Si sa régulation transcriptionnelle par le facteur σE est connue (Dartigalongue et al., 2001), peu d'informations existent sur sa régulation post-transcriptionnelle, qui sera abordée dans ce mémoire.

3. Motilité et adhésion

Parmi les protéines localisées à l'enveloppe, certaines ne sont pas responsables des échanges d'ions et de molécules, ni de l'assemblage des protéines membranaires. Il existe notamment des protéines formant des appendices de surface, souvent impliqués soit dans la motilité des cellules, soit dans leur capacité à adhérer à divers substrats. Certains de ces complexes protéiques seront abordés ici.

a. Motilité : les flagelles

De nombreuses espèces bactériennes sont capables de se déplacer dans un milieu liquide ou sur une surface souple grâce à un ou plusieurs flagelles. Malgré la grande diversité de protéines requises dans l'assemblage de cet imposant complexe selon les espèces, on trouve chez les entérobactéries et même au-delà une certaine conservation de ses divers composants (Pallen and Matzke, 2006). Le flagelle se décompose en un filament et un crochet formant l'appendice extracellulaire, une partie basale ancrée à l'enveloppe cellulaire, un système de sécrétion de protéines et un moteur moléculaire (Fig. 28).

On note parmi les nombreuses subtilités du flagelle bactérien que le filament est un polymère de flagellines, des protéines capables de s'auto-assembler. Ainsi, chez les entérobactéries le système de sécrétion de type III ancré à la membrane interne permet d'injecter la flagelline FliC vers l'intérieur creux du filament, où FliC va pouvoir s'assembler avec d'autres protéines FliC pour former le filament (Chevance and Hughes, 2008). Le moteur moléculaire à la base du flagelle utilise la force protomotrice, générée par la différence de

concentration en protons entre cytoplasme et périplasme, pour faire tourner le flagelle. Cette rotation du flagelle autour d'un axe central ancré à la membrane peut se comparer au mouvement d'une hélice et propulse les bactéries.

Par ailleurs de nombreuses études ont été menées sur les conséquences biologiques de la motilité des bactéries. Ainsi par un procédé que l'on nomme chimiotaxie, les bactéries ont la capacité de sentir la présence d'une molécule bénéfique pour la cellule et de nager dans sa direction (Micali and Endres, 2016). La motilité est également d'une grande importance dans la virulence des bactéries pathogènes (Micali and Endres, 2016).

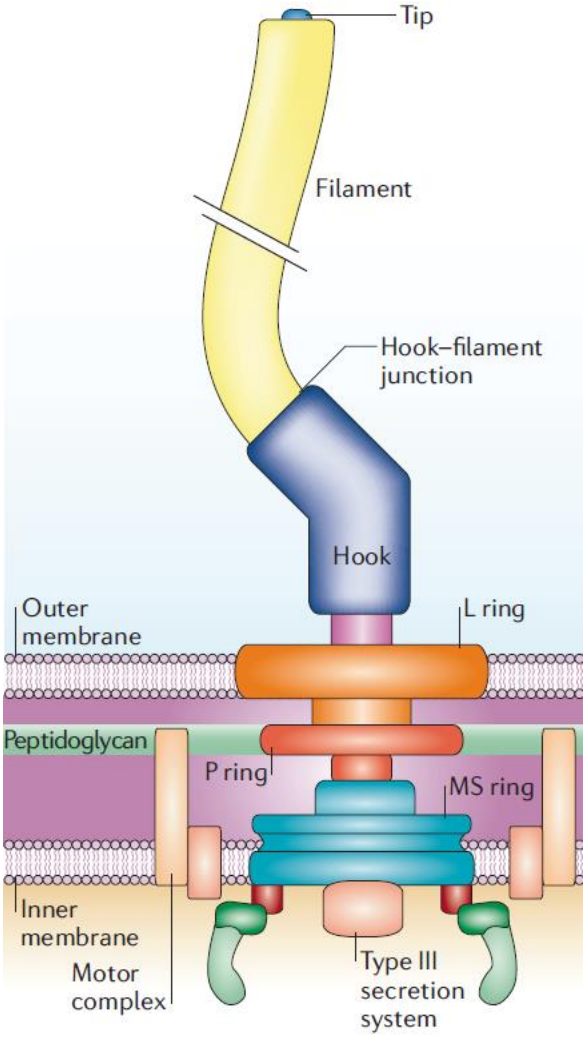


Figure 28 – Le flagelle bactérien
Schéma de la composition d'un flagelle bactérien, à partir du modèle présent chez S. enterica. Le flagelle est composé d'une partie basale ancrée à la membrane interne, à la membrane externe et à la paroi peptidoglycane. Le filament est le principal composant du flagelle et consiste en un polymère de flagelline, qu'un crochet (« hook ») rattache à la partie ancrée à la membrane. Un complexe moteur fournit l'énergie nécessaire au flagelle pour effectuer sa rotation autour de l'axe d'ancrage à la membrane, ce qui permet au filament de propulser la bactérie à la manière d'une hélice. Enfin, un

système de sécrétion élaboré permet l'export de la flagelline afin de prolonger le filament. D'après (Pallen and Matzke, 2006).

b. Adhésion, biofilms et protéines Curli

La capacité des bactéries à adhérer à divers substrats est particulièrement étudiée du fait de son importance dans la formation des biofilms, ces communautés multicellulaires de bactéries engluées au sein d'une matrice extracellulaire. L'adhésion et la formation de biofilms participent notamment à la virulence de bactéries pathogènes. En effet, les biofilms confèrent aux bactéries une résilience importante lors de divers traitements aux antibiotiques et pourraient permettre le développement de souches résistantes.

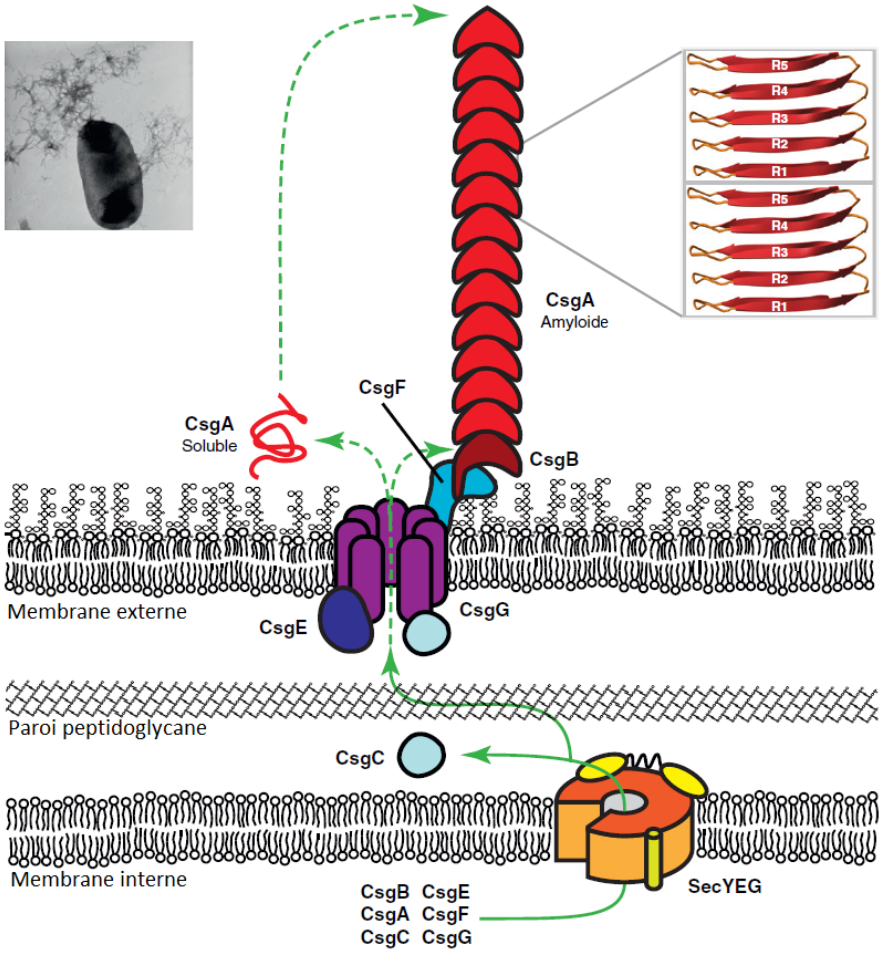


Figure 29 – Composition du complexe Curli chez les entérobactéries
Schéma d'un complexe Curli. Après leur export vers le périplasme via le translocon SecYEG, les divers composants du complexe Curli sont assemblés à la membrane externe, où la protéine CsgG forme une OMP à tonneau β. Celle-ci va permettre la sécrétion de la principale composante des complexes Curli, la protéine CsgA. Dans le périplasme CsgA est maintenue sous forme soluble par la chaperonne CsgC qui a une activité anti-amyloïde. En revanche, une fois sécrétée CsgA forme des plaques amyloïdes du fait de l'agrégation de ses feuillets β (encadré de droite), formant de longs appendices extracellulaires visibles en microscopie électronique à transmission (encadré de gauche). D'après (Blanco et al., 2012).

L'un des complexes protéiques jouant un rôle prédominant dans l'adhésion est le complexe Curli. Ce dernier est l'une des composantes extracellulaires majeures des biofilms (Blanco et al., 2012) (Fig. 29), avec les exopolysaccharides sécrétés par les bactéries. La polymérisation de CsgA, la principale sous-unité des Curli, forme des plaques amyloïdes conférant notamment une grande résistance mécanique aux biofilms bactériens (Mika and Hengge, 2014). Cela étant dit, divers éléments de l'enveloppe cellulaire participent également à la capacité d'adhésion des bactéries, à savoir les pili, les LPS, mais aussi les flagelles. En effet si leur première fonction est la motilité des cellules, les flagelles peuvent par exemple entrer en interaction avec la surface des cellules épithéliales eucaryotes, y entraînant l'adhésion des bactéries (Haiko and Westerlund-Wikström, 2013).

B. Réponse au stress, ARN régulateurs et circuits de régulation

Les bactéries sont exposées à de nombreux stress et changements de conditions environnementales, qui requièrent une adaptation rapide. La clé de cette adaptation réside dans la régulation de l'expression génétique, avec la répression des gènes dont les produits ne sont pas nécessaires voire néfastes dans les nouvelles conditions et l'activation de ceux permettant une réponse adéquate. La sélection naturelle a donc favorisé l'apparition de nombreux régulateurs de l'expression génétique, mais aussi de mécanismes permettant leur induction et leur action au moment propice. Ces régulateurs sont par ailleurs souvent impliqués dans des réseaux de régulation d'une grande complexité et peuvent contrôler la synthèse de nombreuses protéines membranaires, permettant une adaptation rapide à l'environnement. Divers mécanismes de réponse au stress et réseaux de régulation seront présentés dans cette partie, avec un accent sur les facteurs de transcription et les ARN régulateurs auxquels je me suis intéressé au cours de ce travail de thèse. Ces derniers partagent tous le point commun d'être fortement impliqués dans la régulation de l'expression de protéines membranaires.

1. Le système à deux composants EnvZ-OmpR et les ARN régulateurs OmrA et OmrB

Les facteurs de transcription (FT) répondent à des stimuli environnementaux d'une grande diversité, et beaucoup de ces FT sont inclus dans des systèmes à deux composants (SDC). Ces derniers fonctionnent *via* l'activation d'une protéine kinase senseur ancrée à la membrane qui va s'autophosphoryler en réponse à un stimulus donné. S'ensuivra la phosphorylation spécifique d'un FT que l'on nomme dans ce cas de figure « response regulator » (et que j'appellerai ici régulateur). Le plus souvent, le régulateur n'est actif qu'une fois phosphorylé, et va pouvoir contrôler l'expression de divers gènes – l'ensemble de ces gènes forment le régulon du SDC. Etant donné que la kinase et le régulateur sont présents dans la cellule avant le stimulus, ces SDC permettent une réponse particulièrement rapide. Cette réponse est encore accrue du fait que les SDC régulent souvent positivement leur propre synthèse en réponse à ce stimulus. Les régulons des SDC permettent ensuite, du moins dans un certain nombre de cas, de répondre et de s'adapter au stimulus qui a conduit à l'activation de la kinase senseur.

Je m'intéresserai ici à l'exemple du SDC EnvZ-OmpR, d'une grande importance dans la régulation de nombreuses protéines membranaires, ainsi qu'aux sRNA OmrA et OmrB. Ces deux sRNA ont été au centre de mon travail de thèse, et comme leur nom l'indique (« OmpR-regulated sRNAs »), ils font partie du régulon d'EnvZ-OmpR et l'élargissent en régulant eux-mêmes la synthèse de nombreuses protéines membranaires.

a. EnvZ-OmpR et la réponse au stress membranaire

EnvZ-OmpR est l'un des SDC les plus étudiés et ce depuis plus d'une trentaine d'années. Les deux protéines sont encodées à partir d'un même opéron *ompR-envZ*, dont le locus chromosomique a été initialement identifié pour son effet régulateur sur la transcription des gènes *ompC* et *ompF* (Hall and Silhavy, 1981). Il s'est depuis avéré qu'EnvZ est une protéine formant un homodimère à la membrane interne, avec un domaine senseur périplasmique permettant notamment de détecter de fortes osmolarités (Yaku and Mizuno, 1997) (Fig. 30). Ce signal est ensuite transmis par le domaine kinase cytoplasmique d'EnvZ au régulateur OmpR par le biais d'une cascade de phosphorylations (Igo et al., 1989).

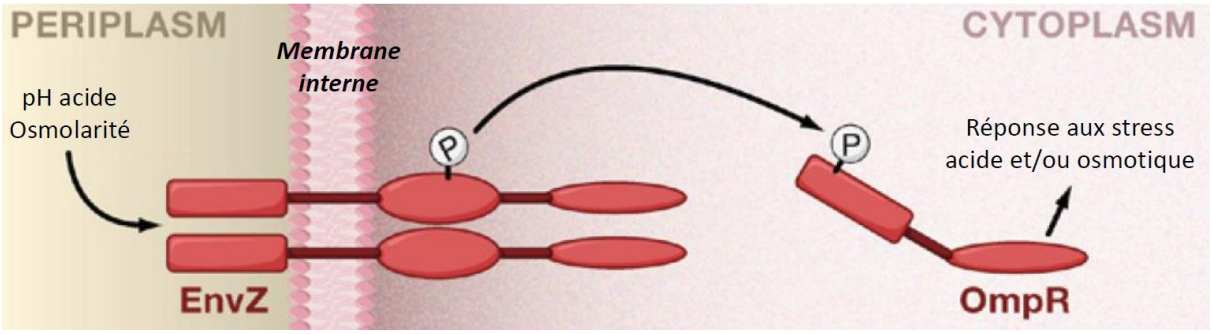


Figure 30 – Le système à deux composants EnvZ-OmpR
Représentation schématique du SDC EnvZ-OmpR. La kinase senseur EnvZ forme un homodimère à la membrane interne. Son domaine périplasmique senseur peut détecter divers stimuli, qui sont intégrés via son domaine kinase cytoplasmique. Ce dernier s'autophosphoryle avant de phosphoryler le régulateur OmpR. Sous forme phosphorylée, OmpR va pouvoir activer la transcription de divers gènes permettant d'apporter une réponse au stimulus initial. Adapté de (Kohanski and Collins, 2008)

Le stress osmotique détecté par EnvZ reste à ce jour mal compris, bien qu'il ait été démontré plus récemment qu'EnvZ est également en mesure de détecter un pH acide chez les entérobactéries *E. coli* et *S. enterica* (Quinn et al., 2014). S'il existe, le lien entre stress osmotique et stress acide n'est pas tout à fait évident. On peut toutefois avancer qu'un pH acide correspond à une forte concentration en ions oxonium H_3O^+, exerçant ainsi une pression osmotique sur les cellules à l'instar des fortes concentrations en NaCl ou en saccharose caractérisées comme des stress osmotiques activant EnvZ. Une hypothèse pourrait être que les conditions de pH acide et de forte osmolarité provoquent un stress membranaire similaire, qui serait le signal perçu par EnvZ. Le fait que l'activité kinase d'EnvZ soit également fortement activée par la procaïne, un composé chimique causant une dépolarisation de la membrane et donc un stress membranaire, conforte cette hypothèse (Rampersaud and Inouye, 1991). Il est d'ailleurs tout à fait imaginable que d'autres stimuli puissent induire une réponse de ce SDC, dont l'étude se poursuit encore aujourd'hui.

Sous sa forme phosphorylée OmpR-P, ce régulateur active la transcription de nombreux gènes ayant souvent un rapport direct avec l'enveloppe cellulaire (Fig. 31). Ainsi chez *E. coli* le régulon du SDC EnvZ-OmpR inclut les gènes *ompC* et *ompF* codant les porines majeures OmpC et OmpF évoquées précédemment, mais aussi le gène *dtpA* (Goh et al., 2004) codant un transporteur de dipeptides et tripeptides. OmpR-P contrôle également l'expression des gènes *flhDC* (Shin and Park, 1995) et *csgD* (Ogasawara et al., 2010), codant respectivement des régulateurs majeurs de la synthèse du flagelle et de la synthèse des complexes Curli. Enfin, la régulation de l'expression de *bolA* par OmpR-P (Yamamoto et al., 2000) pourrait avoir une

influence sur la morphologie cellulaire. La régulation de l'expression de l'ensemble de ces gènes permet aux bactéries de répondre au stress de manière adéquate, en modifiant notamment la perméabilité de la membrane *via* sa composition en OMP, mais aussi par exemple en activant la formation des biofilms par le biais de *csgD*.

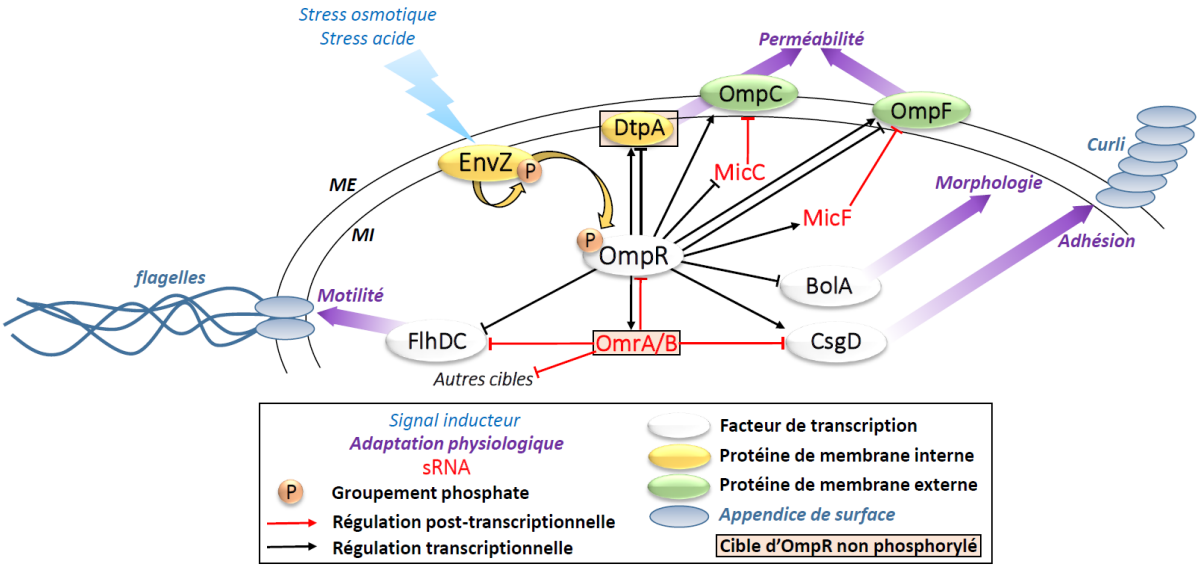


Figure 31 – Régulon du système à deux composants EnvZ-OmpR
En réponse à divers stress, la kinase senseur EnvZ va s'autophosphoryler puis phosphoryler OmpR. Ce régulateur va alors activer ou réprimer l'expression d'un certain nombre de gènes impliqués dans la physiologie de la membrane, permettant d'apporter une réponse au signal inducteur. Le régulon d'EnvZ-OmpR comprend des protéines membranaires et des régulateurs protéiques et ARN de l'expression génétique. MI : membrane interne ; ME : membrane externe.

Cela étant dit, OmpR-P n'agit pas uniquement sur l'expression de gènes codants. En effet, ce régulateur active la transcription d'au moins trois sRNA Hfq-dépendants, à savoir MicF, OmrA et OmrB, et réprime celle de MicC (Fig. 31). L'ensemble de ces sRNA ayant leurs propres cibles, cela étend d'autant plus le nombre de gènes et de fonctions régulés par le SDC EnvZ-OmpR. Cela imprime également une dynamique sur la régulation de certains gènes, dans la mesure où leur expression peut répondre au contrôle par OmpR-P dans un premier temps, puis par le sRNA dans un second temps. De plus, certaines régulations vont dans des sens apparemment contradictoires (par exemple, OmpR-P active l'expression de *csgD* tandis que les OmrA/B la répriment), et méritent une réflexion plus poussée (voir **Revue n°2**).

Par ailleurs, notre laboratoire a récemment pu mettre en évidence pour la première fois une activité de la forme non phosphorylée d'OmpR sur l'expression de certaines cibles, à savoir OmrA, OmrB et *dtpA* (Brosse et al., 2016). Cette découverte est particulièrement

intrigante dans le cas des deux sRNA OmrA et OmrB, qui sont sensibles aux variations de niveaux de la protéine OmpR alors que les autres cibles d'OmpR ne sont sensibles qu'à la forme phosphorylée d'OmpR. De ce fait, OmpR agirait sur l'ensemble du régulon d'OmrA et OmrB même dans des conditions où le reste du régulon d'OmpR reste peu affecté. Il se pourrait par exemple que certains stimuli activent la phosphorylation par EnvZ tandis que d'autres n'affecteraient que les niveaux d'OmpR non phosphorylé. Dans le premier cas de figure, l'effet des sRNA OmrA et OmrB se joindrait à celui d'OmpR-P sur l'ensemble du régulon d'EnvZ-OmpR, tandis que dans le second cas ces sRNA seraient les intermédiaires privilégiés de la réponse au stimulus en question. De plus, cette caractéristique de la régulation par OmpR non-phosphorylé place l'expression d'OmrA et de OmrB dans une voie d'autorégulation intéressante dans mesure où ces deux sRNA, en affectant les niveaux d'OmpR, répriment leur propre synthèse sans affecter les niveaux d'OmpR-P et donc le reste du régulon du SDC EnvZ-OmpR. Quoi qu'il en soit, cela donne un aperçu de la complexité des réseaux de régulation impliquant des ARN régulateurs (voir **Revue n°2**).

b. OmrA et OmrB, deux sRNA Hfq-dépendants similaires

Les deux sRNA OmrA et OmrB (OmrA/B) font partie des sRNA Hfq-dépendants les mieux caractérisés à ce jour et représentent l'un des centres d'intérêt de notre laboratoire, et l'un des miens en particulier. Je ferai ici état des connaissances à mon arrivée au laboratoire, afin de les distinguer des résultats que j'ai obtenus et qui seront présentés par la suite.

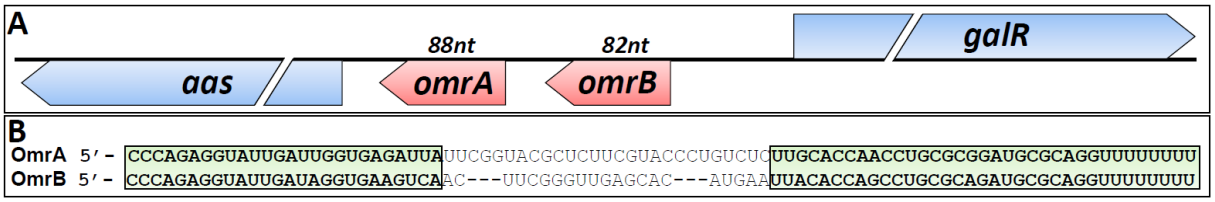


Figure 32 – Locus et conservation des deux sRNA OmrA et OmrB

A. Locus chromosomique des gènes omrA et omrB (en rouge) chez les espèces d'entérobactéries où l'on trouve les deux gènes. Ceux-ci sont situés dans la région du chromosome qui sépare les gènes codants galR et aas. Ces derniers encodent respectivement un répresseur transcriptionnel de la voie de catabolisme du D-galactose et une acyl-ACP synthétase impliquée dans la ligature d'acides gras. On note que leurs rôles sont à priori dissociés de ceux d'OmrA/B. B. Conservation de séquence entre les sRNA OmrA et OmrB. Les régions 5' et 3' encadrées en vert sont à la fois conservées entre les deux sRNA et dans les versions d'OmrA et/ou d'OmrB retrouvées chez l'ensemble des entérobactéries.

On trouve souvent les deux sRNA OmrA/B, et au moins l'un des deux, chez la plupart des entérobactéries. Ils sont également conservés entre eux, dans la mesure où leurs

extrémités 5’ sont quasiment identiques, de même que leurs extrémités 3’ (Fig. 32). Les deux gènes qui les encodent sont adjacents sur le chromosome et résultent donc vraisemblablement d’une duplication génétique, qui a pu survenir chez l’ancêtre commun d’*E. coli* et des entérobactéries les plus proches, à l’image des salmonelles et des klebsielles.

S’ils ont été identifiés comme des sRNA Hfq-dépendants chez *E. coli* dès les premières recherches de sRNA à l’échelle du génome (Argaman et al., 2001; Wassarman, 2001), leur véritable caractérisation a démarré avec la mise en évidence de leur régulation par OmpR (Guillier and Gottesman, 2006). Cette même étude marque une première étape dans la découverte des cibles d’OmrA/B, qui régulent au niveau post-transcriptionnel l’expression de nombreux gènes dont la fonction est liée à l’enveloppe cellulaire (Fig. 33).

Ainsi les gènes d’OMP *cirA*, *fecA* et *fepA*, impliqués dans l’import de fer (voir **Introduction III.C.2.b**), font partie des cibles de ces deux sRNA, de même que le gène *ompT* évoqué précédemment. Les ARNm *cirA* et *ompT* sont d’ailleurs des cibles directes d’OmrA/B, (Guillier and Gottesman, 2008), de même que l’ARNm *ompR-envZ*, codant leur propre régulateur. Comme évoqué précédemment, cette dernière régulation introduit une boucle de rétrocontrôle négatif des deux sRNA sur leur propre synthèse. Le gène codant GntP, un transporteur de glucuronate et fructuronate, fait également partie des gènes régulés par OmrA et OmrB (Tjaden et al., 2006), bien que l’interaction directe entre ces sRNA et l’ARNm n’ait pas été démontrée, à l’instar des cibles *fepA* et *fecA* avant mon travail de thèse. On note toutefois que des prédictions d’appariement entre les OmrA/B et ces trois cibles existent. Enfin, certaines cibles sont régulées à la fois par les OmrA/B et OmpR, à savoir *csgD* (Holmqvist et al., 2010) et *flhDC* (De Lay and Gottesman, 2012).

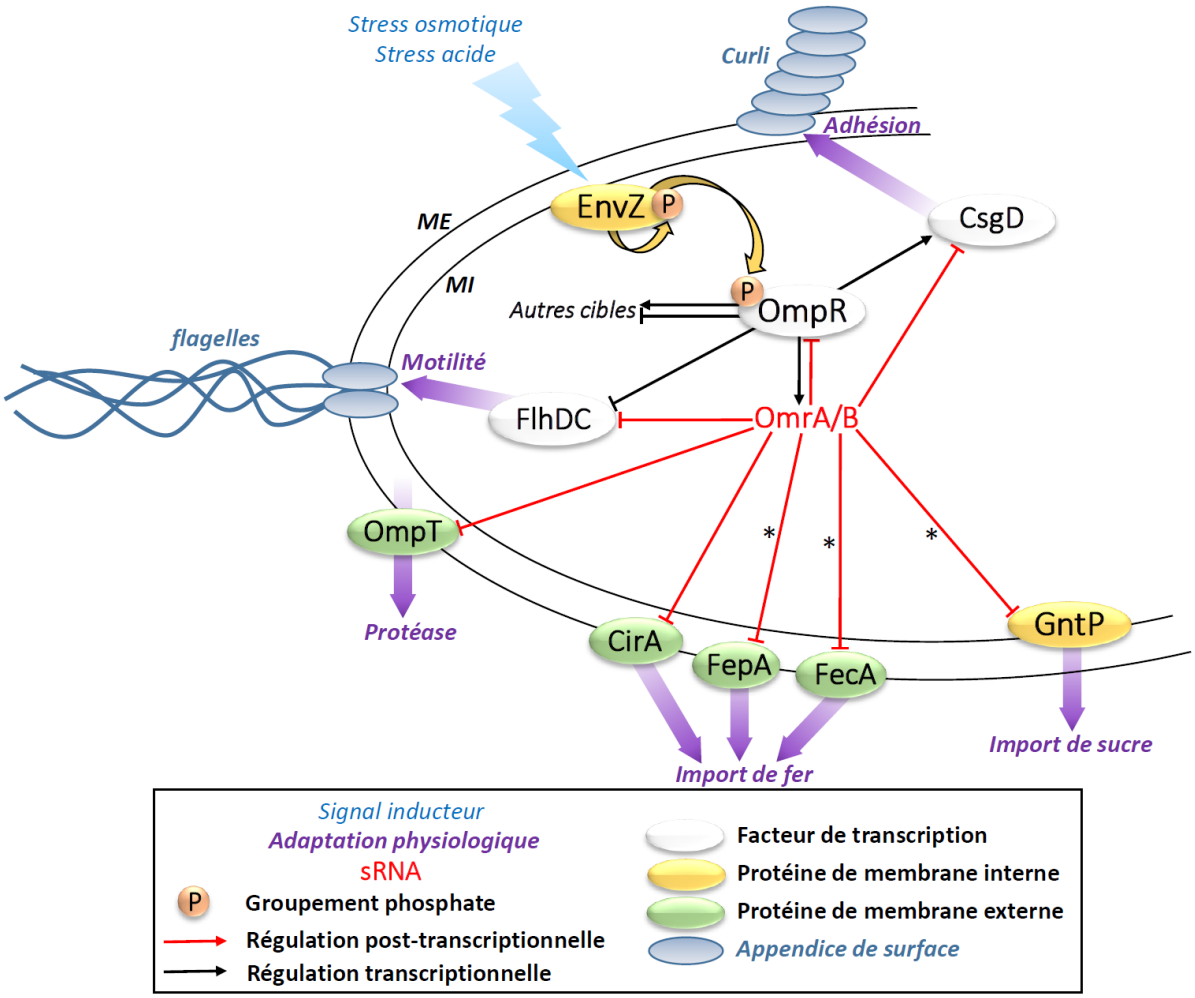


Figure 33 – Régulon des sRNA OmrA et OmrB
Une fois leur transcription activée par le SDC EnvZ-OmpR, les OmrA/B contrôlent l'expression de nombreux gènes au niveau post-transcriptionnel. Toutes les cibles connues sont communes aux deux sRNA et sont régulées négativement. On compte parmi ces cibles des OMP, des protéines de membrane interne ainsi que les régulateurs de la synthèse des flagelles et des complexes Curli. Un astérisque indique les cibles dont la régulation directe par les OmrA/B n'était pas démontrée au début de ma thèse.

Pour l'ensemble des cibles connues d'OmrA/B, l'extrémité 5' conservée des deux sRNA est importante pour la régulation et est impliquée dans l'appariement avec les cibles ARNm directes (Guillier and Gottesman, 2008). Dans la plupart des cas, la région de fixation des sRNA sur l'ARNm recoupe la région du RBS (Fig. 34). Ce genre d'appariement évoque un mécanisme classique de régulation par les sRNA Hfq dépendants, où ces derniers entrent en compétition avec la sous-unité 30S du ribosome pour la fixation au RBS, comme vu précédemment.

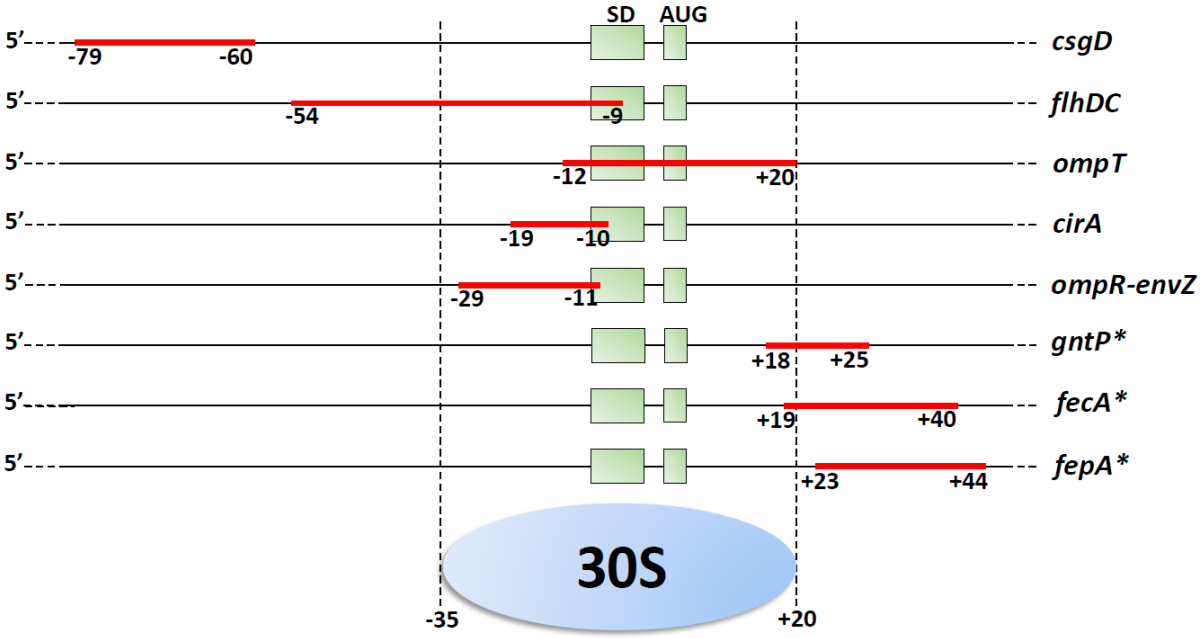


Figure 34 – Sites d'appariement d'OmrA/B sur leurs ARNm cibles
Pour chaque cible d'OmrA et OmrB, la région qui s'apparie avec l'extrémité 5' des sRNA est surlignée en rouge. Sur les cibles (indiquées par un astérisque) pour lesquelles l'appariement n'a pas été démontré, la région surlignée en rouge correspond à une prédiction d'appariement. Pour la plupart des cibles directes connues, la région d'appariement recoupe le RBS où se fixe la sous-unité ribosomique 30S pour initier la traduction.

Pour autant, il existe au moins un cas où OmrA/B s'apparient à leurs ARNm cibles en dehors du RBS. En effet, l'extrémité 5' d'OmrA/B s'apparie avec la région 5' UTR de l'ARNm *csgD*, à plus de 60 nt en amont du codon de démarrage (Holmqvist et al., 2010). Une tige-boucle en amont du RBS semble être impliquée dans cette régulation et les auteurs proposent en effet qu'en prévenant la formation de cette tige-boucle, l'appariement avec OmrA ou OmrB entraîne l'apparition d'une structure alternative inhibitrice de la traduction ; le détail de ce mécanisme reste cependant à comprendre. Les auteurs n'ont par ailleurs pas imaginé un mécanisme alternatif où la tige-boucle elle-même aurait un effet activateur de la traduction, à l'image de la structure similaire trouvée dans l'ARNm *dusB-fis* en amont de la séquence codante de *fis* et évoquée auparavant. Dès lors, un appariement avec les OmrA/B pourrait prévenir la formation de cette structure activatrice, entraînant une diminution de la synthèse de la protéine CsgD.

Pour les cibles *gntP*, *fecA* et *fepA*, les prédictions d'appariement avec OmrA/B fournies par les logiciels dédiés TargetRNA (Tjaden et al., 2006) et/ou intaRNA (Busch et al., 2008) (Fig. 34 et 35) sont tout à fait intéressantes puisqu'elles suggèrent un appariement dans la séquence codante, juste en aval du RBS. Une grande partie de mon travail de thèse a consisté

à comprendre le mécanisme de régulation de *fepA* par les OmrA/B, qui sera discuté en détail (voir **Article** et **Résultats I.D** notamment). D'autres résultats concernant la régulation de *fecA* figurent dans l'**Annexe 2.A**, tandis que l'étude du mécanisme de régulation de *gntP* se poursuit à l'heure actuelle dans le laboratoire.

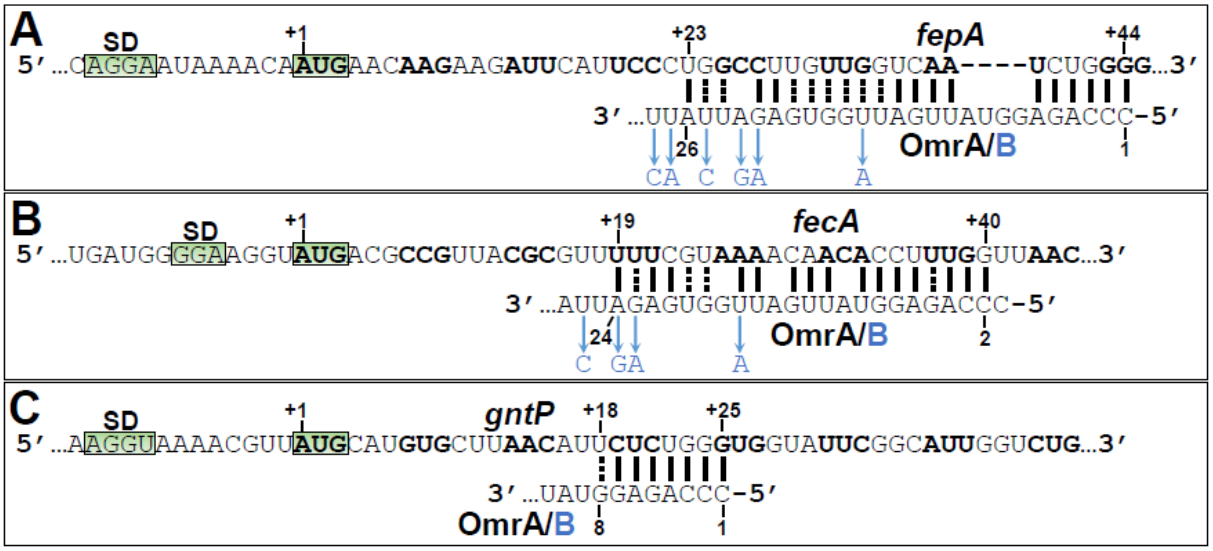


Figure 35 – Prédictions d'appariement in silico entre les OmrA/B et les ARNm fepA, fecA et gntP
Les prédictions d'appariement ont été fournies par les logiciels intaRNA et targetRNA. Les régions du RBS importantes pour le démarrage de la traduction sont encadrées en vert ; les positions délimitant les appariements sont indiquées ; la position +1 des ARNm correspond au premier nt du codon de démarrage de la traduction. Le cadre de lecture des ARNm est marqué par l'alternance de codons en caractères gras et normaux. Les flèches et nt en bleu indiquent les positions où les séquences d'OmrA et OmrB diffèrent. Un appariement Watson-Crick est représenté par un trait plein, et une paire de bases Wobble par un trait en pointillés.

Considérons à présent la physiologie cellulaire : les OmrA/B ont un impact important sur la production d'une grande diversité de protéines membranaires. Le premier rôle qui leur est logiquement attribué est la réponse au stress membranaire. En effet, la répression de la synthèse de certaines protéines membranaires peut permettre de libérer de l'espace dans les membranes pour d'autres protéines dont la présence peut être plus urgente au moment du stress. Leurs effets sur la motilité et l'adhésion des cellules sont cependant particulièrement intéressants à comprendre dans la mesure où ces deux sRNAs font partie des sRNA les plus exprimés chez *S. enterica* dans des conditions d'interaction avec des macrophages (Westermann et al., 2016). OmrA et OmrB pourraient donc être importants pour la virulence des bactéries. Etant donné que le fer est une ressource limitante dans des conditions d'interactions entre hôtes et bactéries (voir **Introduction III.C.1**), la régulation par les OmrA/B d'au moins trois OMP impliquées dans l'import du fer interpelle également et évoque à

nouveau un rôle de ces sRNA dans la survie des bactéries au sein d'un hôte. Enfin, la régulation de nombreuses OMP n'est peut-être pas anodine dans la mesure où ce sont des protéines immunogènes, souvent reconnues comme des antigènes par l'hôte. Au-delà de leur participation à la réponse au stress membranaire, les OmrA/B pourraient donc également intervenir dans l'interaction entre les bactéries et leurs hôtes. Diverses hypothèses sur les rôles de ces deux sRNA dans la physiologie de la cellule seront abordées dans la suite de ce mémoire.

2. Facteurs sigma et ARN régulateurs

Les facteurs σ sont des acteurs majeurs de la réponse au stress non seulement du fait de leurs régulons souvent très étendus, mais aussi parce que les mécanismes qui permettent leur libération et leur action dans la cellule sont très rapides. En effet, ils sont souvent déjà présents dans la cellule au moment du stress et sont simplement libérés d'une interaction inhibitrice avec un anti-σ répresseur, à l'image de σ70 avec l'ARN 6S évoqué précédemment. Cela évite à la cellule d'avoir à synthétiser le facteur σ lui-même lors du stress, au cours duquel une réponse rapide peut s'avérer primordiale. Le fait que les facteurs σ mobilisent l'ARNpol pour la synthèse des gènes de leurs régulons est également très important, puisque cela permet de focaliser l'expression génétique de la cellule sur le besoin le plus urgent de réponse au stress. Je m'attarderai ici sur le facteur σE, qui est un régulateur majeur de la réponse au stress membranaire, puis sur le lien existant entre σE et le sRNA RseX que j'ai étudié au cours de cette thèse. J'aborderai également σS et le sRNA SdsR, qui sont d'importants éléments de la réponse au stress notamment en phase stationnaire de croissance.

a. Réponse au stress membranaire : le facteur σE et le sRNA RseX

Le facteur σE est une protéine essentielle régulant l'expression d'un grand nombre de gènes directement liés à la physiologie de la membrane (Fig. 36). L'accumulation de LPS dans le périplasme (Lima et al., 2013) et d'OMP mal adressées à la membrane externe (Walsh et al., 2003) sont autant de stimuli intégrés par σE. Ces conditions correspondent à des situations de stress membranaire qui provoquent la dégradation de l'anti-σ RseA ancré à la membrane interne, entraînant le relargage de σE dans le cytoplasme (Chaba et al., 2007). σE va ainsi pouvoir activer la transcription de nombreux gènes impliqués dans la réponse au stress

membranaire (Grabowicz and Silhavy, 2017). On compte parmi les gènes régulés des protéases périplasmiques qui vont dégrader le surplus d'OMP, mais aussi des chaperonnes permettant un meilleur adressage des OMP. Les gènes du complexe Bam font d'ailleurs partie des gènes régulés par σE (Dartigalongue et al., 2001), et l'augmentation de leur expression permettra également l'intégration plus rapide d'OMP à la membrane externe. Le gène *lptD* est une autre cible de σE (Dartigalongue et al., 2001) dont l'activation permet un meilleur export du LPS à la membrane externe.

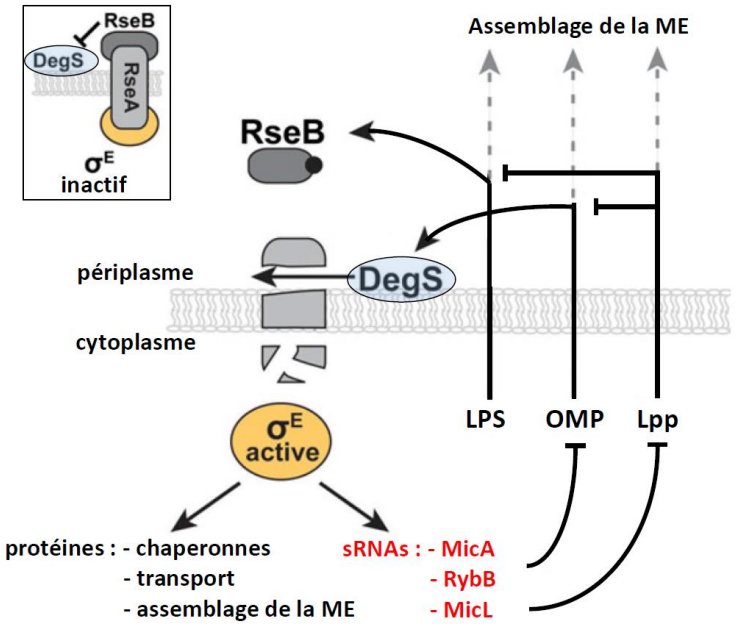


Figure 36 – Le facteur σE et la réponse au stress membranaire
En l'absence de stress, σE est ancré à la membrane via une interaction avec l'anti-σ RseA (encadré de gauche). Cet anti-σ est stabilisé par la protéine périplasmique RseB, qui réprime l'activité protéase de DegS. En cas de stress, des OMP et des LPS non intégrés à la ME (membrane externe) s'accumulent dans le périplasme et activent la protéase DegS, tandis que la protection de RseA par RseB est levée. RseA va donc être dégradée par DegS, ce qui entraîne le relargage de σE. Ce dernier peut alors activer la transcription de nombreux gènes permettant de répondre au stress, y compris trois sRNA. Ces sRNA répriment la synthèse de diverses protéines dont l'accumulation serait délétère pour la cellule. D'après (Guo et al., 2014).

Ce facteur σ induit également la transcription des sRNA Hfq-dépendants MicA, RybB et MicL (Guo et al., 2014; Johansen et al., 2006) (Fig. 36), qui répriment l'expression de gènes dont les produits aggraveraient la situation de stress. Ainsi MicA et RybB contrôlent négativement l'expression de nombreuses OMP, évitant leur accumulation inutile alors que le périplasme regorge déjà d'OMP mal adressées (Gogol et al., 2011). On compte notamment le gène d'OMP majeure *ompA* parmi leurs cibles communes, tandis qu'*ompX* est une cible spécifique de MicA et que les gènes *ompC* et *ompF* figurent parmi les cibles de RybB. La synthèse de Lpp, une autre protéine d'une grande importance dans la physiologie

membranaire, est quant à elle réprimée par MicL (Guo et al., 2014). Lpp est la plus abondante des protéines chez *E. coli*, et permet de lier la paroi peptidoglycane à la membrane externe. Un mauvais adressage de cette protéine et son accumulation au périplasme entraînent le mauvais adressage à la fois d'OMP et de LPS, d'où la nécessité de réprimer son expression précisément dans ces conditions.

Il existe cependant d'autres sRNA liés à la réponse au stress membranaire σE-dépendante, à l'image du sRNA Hfq-dépendant RseX (Fig. 37). La surexpression de ce dernier permet ainsi de supprimer la létalité d'une délétion de *degS* (Douchin et al., 2006). Cette délétion entraîne une forte stabilisation de RseA et rend impossible la libération de σE en cas de stress membranaire (Fig. 36). Les auteurs de cette étude démontrent qu'en réprimant la synthèse d'OmpA et OmpC, la surexpression de RseX rend cette mutation viable, tout comme une double délétion des gènes *ompA* et *ompC*. Cela suit une logique intéressante puisque σE réprime également indirectement la synthèse de ces deux OMP *via* l'action de *micA* et de *rybB*. On peut donc en déduire que ces effets répresseurs sont d'une grande importance en conditions de stress, et qu'en l'absence d'un facteur σE actif une trop forte expression d'*ompA* et *ompC* devient létale pour la cellule, ce qui a été démontré par la suite (Gogol et al., 2011; Papenfort et al., 2006).

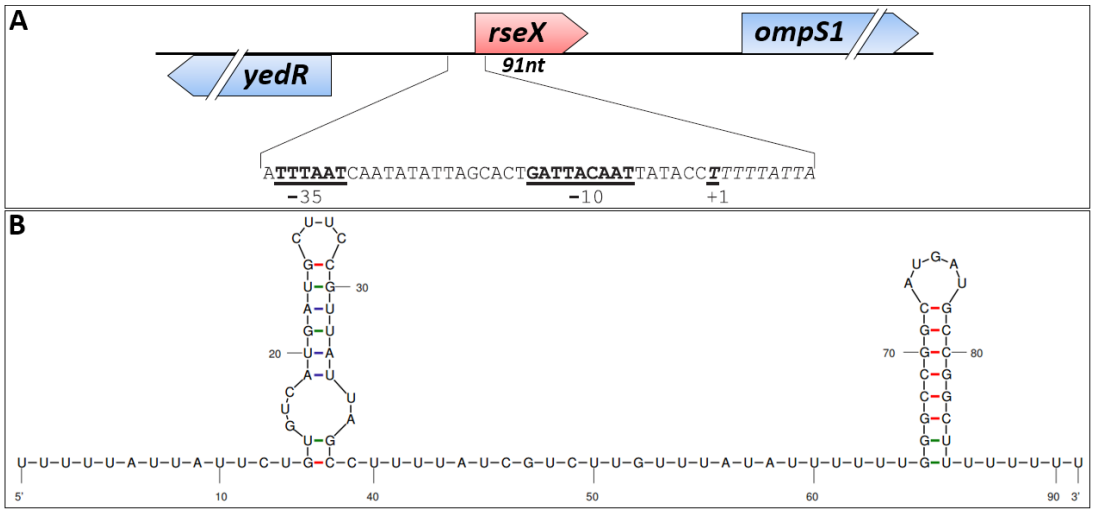


Figure 37 – Le sRNA Hfq-dépendant RseX

Locus du gène rseX (en rouge) sur le chromosome d'E. coli et des entérobactéries apparentées (A) : ce gène est situé dans la région qui sépare les gènes codants hypothétiques yedR et ompS1 (en bleu), dont la fonction n'est pas connue à ce jour. Le promoteur de rseX proposé par (Douchin et al., 2006) est indiqué avec les séquences consensus -35 et -10 ainsi que le TSS en caractères gras et soulignés. La séquence ainsi que la structure de RseX prédite par le programme mfold (Zuker, 2003) sont représentées

en (B), avec les paires de bases G-C en rouge, les paires A-U en bleu et les paires de base Wobble G-U en vert. D'après (Douchin et al., 2006).

Si l'effet de RseX est indéniablement lié à la physiologie de la membrane externe, sa propre régulation n'est toujours pas élucidée à ce jour. L'expression de cet ARN à partir de son locus chromosomique n'a d'ailleurs jamais pu être observée par la technique classique de Northern Blot. On peut cependant noter avec intérêt que ce sRNA est l'un des plus fortement exprimés chez des salmonelles infectant des fibroblastes de rat (Ortega et al., 2012). J'ai étudié au cours de ma thèse la question de la régulation de RseX, qui sera discutée dans la partie **Résultats II.B.2**.

b. Phase stationnaire de croissance : le facteur σS et le sRNA SdsR

Le facteur σS est avec σE l'un des facteurs σ les plus étudiés ; il est conservé chez la plupart des β-, ϒ- et δ-protéobactéries. Si sa présence n'est pas essentielle, il est considéré comme le régulateur majeur de la réponse au stress, notamment en stress thermique, en cas de dommages à l'ADN chromosomique ou encore en phase stationnaire de croissance (Battesti et al., 2011; Hengge, 2011) (Fig. 38). Si cette phase stationnaire correspond à des conditions de culture liquide au laboratoire, on peut considérer qu'elle se rapproche de conditions de stress fréquemment rencontrées par les bactéries. En effet, elle peut se comparer à l'épuisement de l'essentiel des ressources nécessaires à la croissance, à commencer par la ou les sources de carbone. Dans de telles conditions, il est notamment nécessaire de préparer la cellule à des conditions de stress prolongé. On trouve ainsi dans le régulon de σS des gènes impliqués dans la réponse au stress oxydatif, protégeant notamment la cellule contre des molécules causant des dommages à l'ADN, ou encore des gènes de la voie de biosynthèse du glycogène, une molécule permettant le stockage de l'énergie sur le long terme (Hengge, 2011). Cependant, vue l'étendue du régulon de σS qui comprend à peu près 10% des gènes chez *E. coli*, il semble évident que bien d'autres fonctions sont affectées par ce régulateur, à l'image de l'adhésion des cellules (Mika and Hengge, 2014).

Contrairement à σE, σS n'est pas constitutivement présent dans les cellules et sa synthèse est activée au niveau transcriptionnel et traductionnel en réponse à divers stimuli. De nombreuses cascades de régulation aboutissent ainsi à la synthèse de σS et à son association à l'ARNpol (Fig. 38), impliquant divers SDC mais aussi d'autres régulateurs de

l'expression génétique, tels que l'alarmone ppGpp, un nt modifié également important dans la réponse globale au stress. On remarque avec intérêt que de nombreux sRNA Hfq-dépendants interviennent dans la régulation de la synthèse de σS, à savoir DsrA, ArcZ, RprA et OxyS (Fig. 38, voir **Revue n°1**). Plus étonnamment, de nombreux processus régulent la stabilité de la protéine σS, ce qui constitue une voie inhabituelle de régulation (Fig. 38).

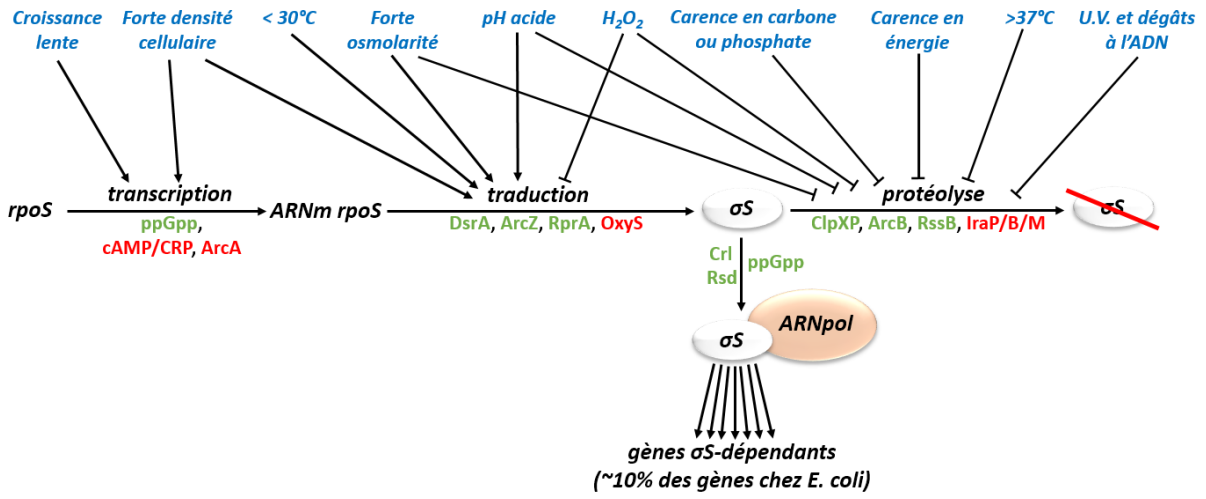


Figure 38 – Rôle central du facteur σS dans la réponse au stress
Schéma simplifié des voies de régulation de σS. Ce facteur σ répond à de nombreux stimuli (en bleu) à la fois lors de sa transcription, de sa traduction, de son association avec l'ARNpol ainsi que lors de sa dégradation. Certains nucléotides modifiés (ppGpp, cAMP), sRNA (DsrA, ArcZ, RprA, OxyS) et protéines (CRP, Crl, Rsd, ClpXP, RssB, IraP/B/M) sont indiquées respectivement en vert ou en rouge selon qu'elles agissent positivement ou négativement sur l'un de ces processus. Adapté de (Hengge, 2011).

On note également que le régulon de σS compte de nombreux sRNA Hfq-dépendants. Parmi les sRNA de ce régulon, on trouve notamment GadY, impliqué dans la réponse au stress acide chez *E. coli* (Opdyke et al., 2004), mais aussi SdsR (Frohlich et al., 2012). Ce sRNA Hfq-dépendant est tout à fait intrigant dans la mesure où il possède un ARN antisens RyeA dont la fonction reste inconnue à ce jour (Vogel, 2003) (Fig. 39). Par ailleurs, une étude globale indique qu'OmrA serait également moins exprimé dans un mutant *rpoS*- chez *S. enterica* (Lévi-Meyrueis et al., 2014), ce qui expliquerait que l'expression d'OmrB s'observe en début de phase stationnaire de croissance tandis que celle d'OmrA s'observe mieux en phase stationnaire avancée (Wassarman, 2001) (voir **Résultats II.A.1.b**). Cette même étude suggère que l'expression de RyhB, conservé chez les entérobactéries et les vibrionaceae et important en conditions de carence en fer, pourrait également être affectée par σS (Lévi-Meyrueis et al., 2014; Padalon-Brauch et al., 2008).

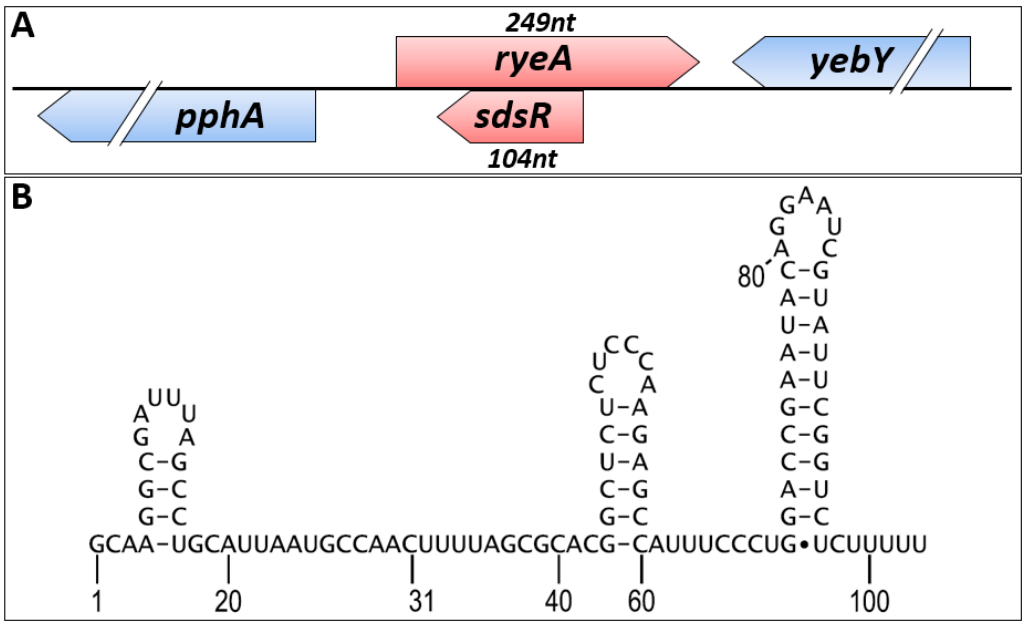


Figure 39 – Le sRNA Hfq-dépendant SdsR

A. Locus chromosomique du gène sdsR et de son antisens ryeA (en rouge) chez les entérobactéries. Ces gènes sont situés dans la région séparant le gène codant hypothétique yebY (en bleu) et le gène pphA (en bleu également) codant une phosphatase impliquée dans la réponse au stress thermique. B. Séquence de SdsR ainsi que sa structure secondaire déterminée par (Fröhlich et al., 2016) avec les paires de bases Watson-Crick représentés par des traits pleins et la paire de base Wobble représentée en pointillés. Les nt en gras correspondent à une région de SdsR s'appariant aux cibles ARNm.

Il est connu depuis peu que le régulon du sRNA SdsR est relativement large (Fröhlich et al., 2016) (Fig. 40) et inclut plusieurs OMP, à l'instar de nombreux sRNA cités précédemment, mais aussi des régulateurs globaux tels que CRP et StpA. Ont été étudiées plus en détail la répression de la synthèse de la protéine OmpD, impliquée dans la virulence chez *S. enterica* (Frohlich et al., 2012), ainsi que celle de MutS chez *E. coli* (Gutierrez et al., 2013). Cette dernière protéine est d'une grande importance dans la réparation de mésappariements lors de la réplication de l'ADN. Sa répression par SdsR peut donc sembler étonnante alors que SdsR peut être induit par σS précisément dans des conditions où l'ADN est endommagé. Les auteurs suggèrent qu'il s'agirait d'un moyen pour les bactéries de promouvoir la diversité génétique, afin d'augmenter la probabilité d'apparition de mutations bénéfiques dans des conditions précises de stress. Il s'agit d'une hypothèse intéressante du point de vue de l'évolution, dans la mesure où la production de diversité génétique serait dès lors inscrite dans le programme génétique des bactéries. Cela permettrait à ces dernières de survivre dans certains environnements hostiles non seulement *via* la modulation des gènes exprimés, mais aussi *via* l'accélération du processus de sélection naturelle.

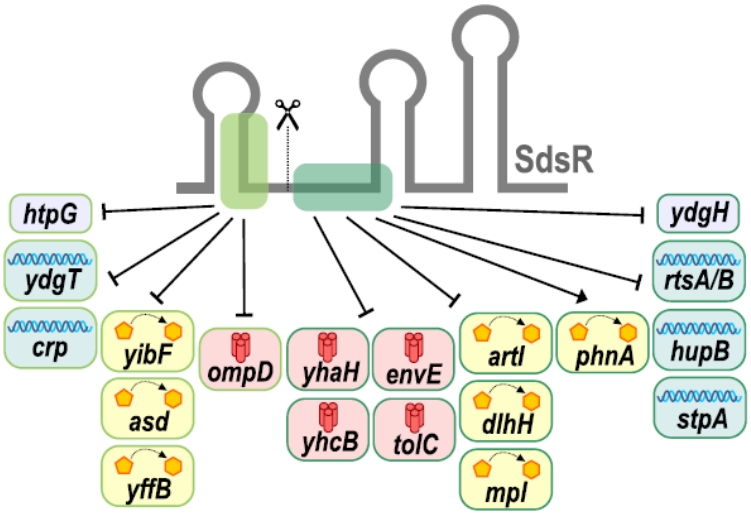


Figure 40 – Le régulon du sRNA SdsR
A. Représentation schématique de SdsR et de son régulon d'après une étude transcriptomique chez S. enterica (Fröhlich et al., 2016). Deux régions successives de SdsR (en vert clair et vert foncé) peuvent s'apparier à diverses cibles ARNm, parmi lesquelles on compte des gènes d'OMP (en rouge), des gènes impliqués dans le métabolisme (en jaune) et des gènes de protéines liant l'ADN (en bleu). La paire de ciseaux indique un clivage générant une forme courte de SdsR (en +31 par rapport au TSS). D'après (Fröhlich et al., 2016).

Les sRNA sont donc impliqués dans l'ensemble des voies de réponse au stress et des réseaux de régulation discutés jusqu'ici – il s'agit d'ailleurs plus vraisemblablement d'une règle que d'une exception, comme le suggèrent des exemples de plus en plus nombreux décrits dans la littérature (voir **Revue n°2**). Par ailleurs, il est intéressant de constater que des protéines membranaires font très souvent partie des régulons de ces sRNA. Cela semble logique dans la mesure où l'enveloppe cellulaire est, comme nous l'avons vu, au cœur de la réponse aux stress. Ainsi, s'il n'est pas certain que l'expression de tout gène soit régulée par au moins un sRNA, il serait tentant de proposer que cela puisse être le cas de l'ensemble des protéines membranaires, au moins chez les entérobactéries et peut-être même au-delà.

C. Cas de figure : la réponse à la carence en fer

De tous les stress auxquels les bactéries sont soumises, la carence en fer est certainement l'un des mieux étudiés à ce jour. J'ai choisi de porter une attention particulière à cette condition environnementale du fait de l'implication de plusieurs protéines et ARN que j'ai étudié dans l'homéostasie du fer. Le fer est une ressource capitale du fait de son utilisation dans de nombreux processus biologiques. Si c'est l'un des éléments les plus abondants sur Terre, il est cependant principalement trouvé sous sa forme non soluble et oxydée Fe^{3+}, c'est

pourquoi il existe une compétition pour y accéder, d'autant plus intense dans des conditions d'interaction entre hôte et bactéries. En effet, le fer est essentiel aussi bien pour l'hôte que pour les bactéries, qui font usage de stratégies différentes pour l'acquérir. Après avoir abordé brièvement la physiologie du fer chez les entérobactéries, je me focaliserai sur les différents moyens à disposition de ces bactéries pour importer le fer dans la cellule. Enfin, je ferai le point sur les réseaux de régulation permettant la réponse à la carence en fer.

1. Physiologie du fer chez les entérobactéries

a. Protéines à centre fer-soufre

L'un des principaux usages du fer chez les entérobactéries – et cela reste vrai dans tous les domaines du vivant – est son intégration au sein de complexes fer-soufre (Fe-S) (Imlay, 2006). On trouve ces complexes dans de nombreuses enzymes, dont l'activité dépend directement de la formation de ces complexes Fe-S et donc de la disponibilité du fer. La biosynthèse de ces complexes relève principalement de l'action des multiples protéines Isc et Suf (Fontecave et al., 2005; Lill, 2009; Mettert and Kiley, 2015; Vinella et al., 2013). L'un des nombreux intérêts des complexes Fe-S est leur fort potentiel d'oxydo-réduction, d'où leur implication dans les processus de respiration (Friedrich et al., 2016). Ils servent également de cofacteurs à des enzymes importantes dans de nombreux autres processus (Imlay, 2006). C'est par exemple le cas de la protéine MutY assurant la réparation d'erreurs dans la réplication de l'ADN (Porello et al., 1998), portant un complexe Fe-S qui joue un rôle crucial dans la reconnaissance spécifique des séquences d'ADN à corriger.

b. Fer libre et cytotoxicité

Si le fer est un composant essentiel des cellules, sa concentration cellulaire doit impérativement être très finement contrôlée. En effet, le fer libre sous forme Fe^{2+} peut rapidement devenir toxique s'il n'est pas incorporé dans un complexe Fe-S ou en interaction avec une molécule destinée à le chélater. Ainsi sous sa forme réduite et soluble Fe^{2+}, le fer peut provoquer une réaction de Fenton avec le peroxyde d'hydrogène ($Fe^{2+} + H_2O_2 \rightarrow Fe^{3+} + OH^- + {}^{\bullet}OH$), entraînant la formation de radicaux hydroxyle ${}^{\bullet}OH$ pouvant endommager la plupart des macromolécules (Imlay, 2006; Imlay and Linn, 1988). On estime toutefois que les bactéries contiennent environ 10 µM de fer libre ou du moins labile (Keyer and Imlay, 1996),

qui pourrait notamment échapper aux voies de biosynthèse des complexes Fe-S ou résulter de leur dégradation. Cette concentration monte à 80 µM dans une souche où l'import de fer n'est plus régulé par Fur (voir **Introduction III.C.3.a**), qui devient particulièrement sensible à toute source de H_2O_2. Par ailleurs et d'une manière générale, les voies de transit du fer entre son import dans la cellule et son incorporation dans les complexes Fe-S restent mal comprises à ce jour.

2. Import de fer : sidérophores et transporteurs

a. Sidérophores

Les entérobactéries sont souvent confrontées à des conditions de rude compétition pour l'accès au fer, notamment lorsqu'elles interagissent avec un hôte et avec d'autres bactéries. À cela s'ajoute le problème de rendre le fer soluble quand il se trouve sous sa forme oxydée Fe^{3+}. Pour répondre à ce problème, de nombreux organismes synthétisent des molécules ayant une très grande affinité pour le fer que l'on nomme sidérophores, pouvant former des complexes fer-sidérophores solubles (Braun and Killmann, 1999; Wandersman and Delepelaire, 2004). Les sidérophores de très loin les plus efficaces sont ceux synthétisés par les bactéries et plus particulièrement les entérobactéries, peut-être parce que ces organismes ont subi une grande pression de sélection pour l'accès au fer. Je me concentrerai principalement sur l'entérobactine, plus directement liée au travail effectué au cours de ma thèse.

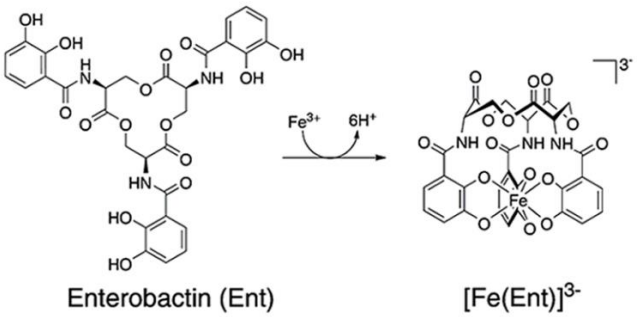


Figure 41 – La molécule d'entérobactine
Représentation schématique d'une molécule d'entérobactine. Une molécule d'entérobactine entre en interaction avec une molécule de fer Fe3+. Ce sont six atomes d'oxygène de l'entérobactine qui partagent leurs électrons avec l'atome de fer au sein d'un complexe extrêmement stable et chargé négativement. D'après (Zheng et al., 2012).

L'entérobactine (Fig. 41), synthétisée notamment par *E. coli* et *S. enterica*, a une constante d'affinité pour le fer de $10^{51}M^{-1}$: il s'agit de la molécule ayant la plus forte affinité pour le fer connue à ce jour (Carrano and Raymond, 1979; Raymond et al., 2003). L'entérobactine se décompose en trois molécules de dihydroxybenzoyl-serine (DHB-Ser), qui ont également une activité chélatrice du fer utilisée par les bactéries, quoique leur affinité pour le fer soit moindre que celle de l'entérobactine. Ce sont les gènes *entCBAEFD*, codant notamment des peptide-synthétases non ribosomiques, qui sont responsables de la biosynthèse de l'entérobactine à partir du shikimate (Kwon et al., 1996; Raymond et al., 2003). Celle-ci est ensuite exportée au périplasme par EntS puis sécrétée par l'OMP TolC (Bleuel et al., 2005). Du fait de la forte stabilité des complexes fer-entérobactine, une molécule de fer liée à l'entérobactine le restera à moins que cette dernière ne soit dégradée. C'est le rôle de l'estérase Fes qui scinde l'entérobactine en trois molécules de DHB-Ser, libérant l'atome de fer une fois le complexe fer-sidérophore intégré dans la cellule (Greenwood and Luke, 1978)

On pourra noter l'existence d'autres sidérophores semblables, comme la salmochéline synthétisée par les salmonelles, ou encore la yersiniabactine synthétisée par les bactéries du genre *Yersinia*. D'une manière générale, la voie de biosynthèse d'un sidérophore n'est souvent conservée qu'au sein d'une ou de quelques espèces bactériennes. Cela étant dit, les bactéries peuvent également tirer profit des molécules chélatant naturellement le fer à l'image du citrate, mais aussi des sidérophores et autres molécules liant le fer synthétisés non seulement par les autres bactéries, mais aussi par l'hôte, à l'image des hèmes (Wandersman and Delepelaire, 2004).

b. Transporteurs de fer

Pour les bactéries à Gram négatif, le moyen le plus simple d'importer du fer dans la cellule est certainement le transport du fer Fe^{2+} lorsqu'il est disponible, notamment en milieu anaérobie où il est moins oxydé en Fe^{3+}. En effet, cet ion diffuse librement à travers diverses porines vers le périplasme, après quoi la protéine de membrane interne FeoB est chargée de son import vers le cytoplasme (Lau et al., 2016). Néanmoins lorsque cette forme du fer vient à manquer, les bactéries disposent d'une grande diversité d'OMP dont la fonction est de reconnaître et d'importer divers chélates de fer (Wandersman and Delepelaire, 2004). C'est par exemple le cas de FecA, le transporteur spécifique du citrate de fer.

D'autres OMP sont dédiées à l'import des sidérophores synthétisés par la cellule ou par d'autres bactéries, et ce uniquement si le sidérophore est complexé avec une molécule de fer. C'est le cas de FepA, une OMP qui reconnaît très spécifiquement les complexes fer-entérobactine, avec une bien meilleure affinité que pour l'entérobactine seule (Hollifield and Neilands, 1978; Ichihara and Mizushima, 1978; Raymond et al., 2003). De façon surprenante, on la retrouve chez de nombreuses espèces incapables de synthétiser l'entérobactine (Rutz et al., 1991), ce qui démontre son importance dans la compétition pour l'accès au fer dans des environnements où ces bactéries sont amenées à se rencontrer.

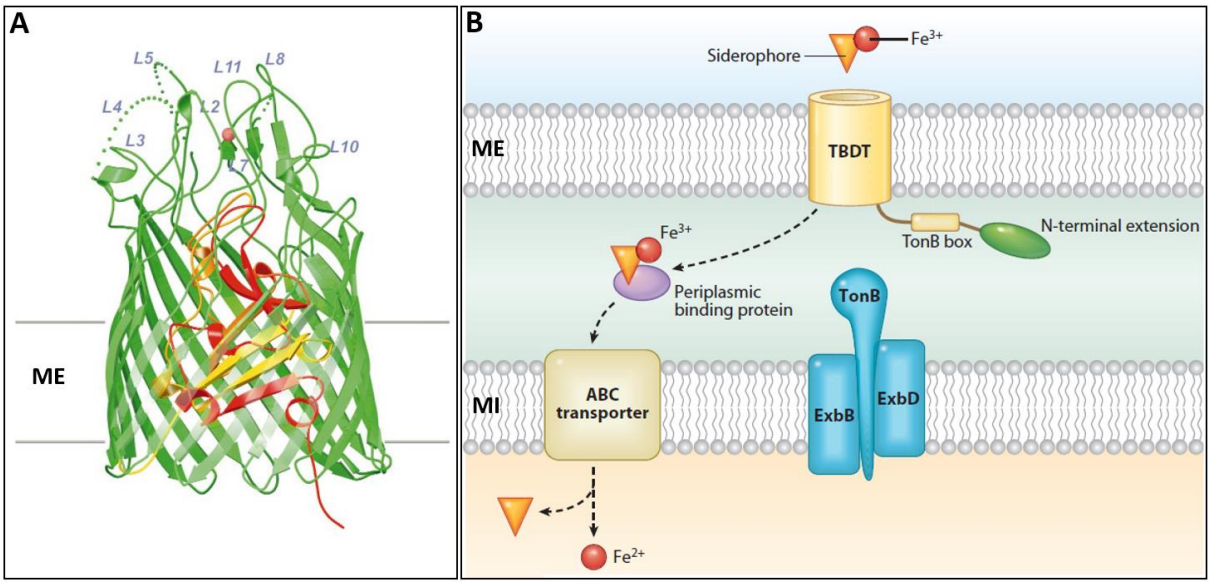


Figure 42 – FepA et l'import de fer TonB-dépendant
A. Structure tridimensionnelle de FepA. Le tonneau β de l'OMP est représenté en vert, de même que les boucles extracellulaires (L pour « loop »). Certains feuillets β ont été rendus transparents afin de laisser apparaître le domaine globulaire en rouge, jaune et orange. La sphère rouge représente l'emplacement de l'atome de fer au moment où le complexe fer-entérobactine entre en interaction avec le récepteur. D'après (Buchanan et al., 1999). B. Fonctionnement des transporteurs de fer TonB-dépendants (« TonB dependent transporter », TBDT). Ces OMP sont chacune spécifiques d'un complexe fer-sidérophore différent, et portent une « TonB box » à proximité de leur région N-terminale interagissant avec TonB. Le complexe TonB-ExbB-ExbD fournit au transporteur l'énergie nécessaire au transport grâce à la force protomotrice. Une fois dans le périplasme, le complexe fer-sidérophore interagit avec une protéine périplasmique spécifique – FepB dans le cas de l'entérobactine. Puis ce complexe est importé vers le cytoplasme par un transporteur à domaine ATPase (« ATP Binding Cassette » ABC-transporter), constitué des protéines FepC, FepD et FepE pour l'entérobactine. Enfin, le sidérophore est clivé pour être éventuellement recyclé (par Fes pour l'entérobactine), tandis que le fer est relâché sous forme Fe^{2+} dans le cytoplasme ou bien pris en charge par diverses enzymes pour la synthèse de complexes Fe-S. Adapté de (Noinaj et al., 2010). ME : membrane externe ; MI : membrane interne.

La structure de FepA a été résolue il y a une vingtaine d'années (Buchanan et al., 1999) (Fig. 42.A) : il s'agit d'une OMP à tonneau β dont le tunnel est obturé par un domaine N-terminal globulaire. Cette structure a permis de proposer puis de démontrer l'intrigant

mécanisme de transport du complexe fer-entérobactine (Ma et al., 2007). Ainsi, le domaine N-terminal globulaire de FepA se déplace vers le milieu extérieur afin d'interagir avec un complexe fer-entérobactine, puis traverse le tonneau β de FepA afin de conduire ce complexe vers le périplasme. C'est le système à transduction d'énergie Ton, impliquant les protéines TonB, ExbB et ExbD, qui fournit l'énergie nécessaire à ce mécanisme (Ma et al., 2007; Rutz et al., 1992).

On remarquera avec intérêt que le complexe Ton fonctionne avec beaucoup d'autres systèmes d'import de fer dits « TonB-dépendants » (Fig. 42.B.), et plus généralement avec un grand nombre de transporteurs spécifiques (ou TBDT pour « TonB dependent transporter ») (Noinaj et al., 2010; Wandersman and Delepelaire, 2004). Cela inclut l'OMP FecA citée auparavant, mais aussi le récepteur FhuA aux complexes fer-ferrichrome, un autre sidérophore de fer. Il en est de même pour CirA, un récepteur spécifique aux complexes fer-DHB-Ser, le précurseur de l'entérobactine. Enfin, Ton fournit également l'énergie nécessaire à l'import d'autres ions et nutriments, tels que la vitamine B12, une molécule liant un atome de cobalt et importée dans la cellule par le récepteur BtuB.

3. Réponse à la carence en fer

Le fer est donc une ressource à la fois vitale pour de nombreux processus cellulaires, et potentiellement toxique en trop grande quantité. La régulation de son import est de ce fait strictement régulée génétiquement, aussi bien en présence de fer qu'en situation de carence. Le régulateur central de l'homéostasie du fer est le répresseur transcriptionnel Fur, dont la fonction est notamment complétée par l'action du sRNA Hfq-dépendant RyhB. J'aborderai le fonctionnement de ces deux régulateurs, après quoi je consacrerai la fin de cette introduction aux régulations existant autour du gène *fepA* et des gènes impliqués dans l'utilisation de l'entérobactine.

a. Fur et RyhB : la régulation de l'homéostasie du fer

Fur est un répresseur transcriptionnel réprimant la synthèse de sidérophores de fer, de même que celle de l'ensemble des transporteurs de fer cités dans la partie précédente. (Cornelis et al., 2011; Lee and Helmann, 2007). Plus précisément, c'est la forme de Fur associée

à un ion Fe^{2+} (Fur-Fe^{2+}) qui réprime l'expression de nombreux gènes. En effet, Fur-Fe^{2+} reconnaît et s'associe à une boîte Fur de l'ADN à proximité des promoteurs de ses cibles, prévenant la fixation de l'ARNpol et donc la transcription de ces gènes (Baichoo and Helmann, 2002). La régulation de l'import de fer par Fur suit une logique simple : en présence de fer, il faut empêcher un apport trop important afin d'éviter d'atteindre des concentrations en fer toxiques pour la cellule. La levée de la répression par Fur en situation de carence en fer permet ainsi la synthèse de l'ensemble de la machinerie d'import du fer depuis les sidérophores jusqu'aux transporteurs, préparant la cellule à la compétition pour l'accès au fer. En outre, Fur réprime l'expression de gènes impliqués dans de nombreuses autres fonctions, y compris la réponse au stress oxydatif. Ainsi Fur contrôle négativement l'expression du gène *sodA* (Tardat and Touati, 1993), codant une superoxyde-dismutase produisant notamment des molécules de H_2O_2. Celles-ci sont dangereuses en présence de fer comme nous l'avons vu précédemment, d'où l'importance de la répression par Fur lorsque le fer n'est pas limitant.

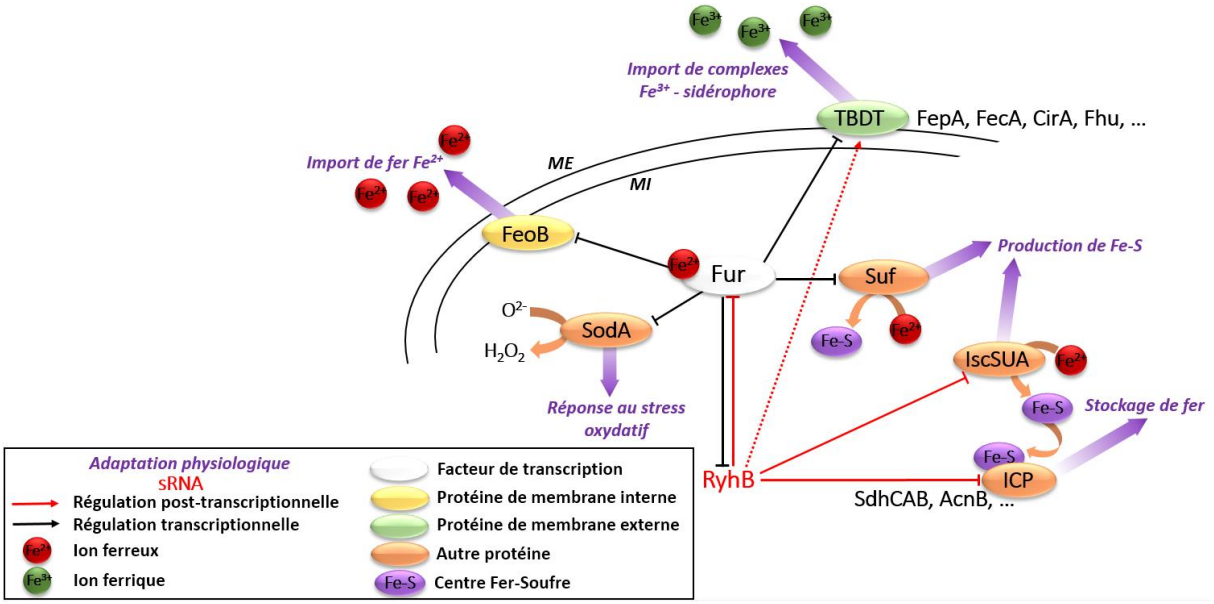


Figure 43 – Régulons simplifiés de Fur et du sRNA RyhB
Sont notamment réprimés par Fur l'ensemble des protéines permettant l'import de fer (TBDT : « TonB Dependent Transporters »), de même que le système Suf de production de centres Fe-S, ainsi que l'enzyme SodA permettant de répondre au stress oxydatif. En situation de carence en fer, ces répressions sont levées, permettant entre autres, via la synthèse de RyhB, l'activation post-transcriptionnelle de la synthèse de certains TBDT tels que CirA chez E. coli ou encore IroN, le récepteur aux complexes fer-salmochéline chez S. enterica (Balbontín et al., 2016). RyhB va par ailleurs réprimer la synthèse d'une vingtaine de protéines contenant des centres Fe-S (« Iron Containning Proteins », ICP). Enfin, on note qu'entre la présence du fer et la carence en fer, les bactéries passent du système Isc au système Suf pour la synthèse des centres Fe-S. ME : membrane externe ; MI : membrane interne.

Le régulon de Fur (Fig. 43) comprend par ailleurs un sRNA dont l'action est particulièrement importante, à savoir RyhB, qui est donc induit dans des conditions de carence en fer (Massé and Gottesman, 2002; Salvail and Massé, 2012). Ce sRNA réprime la synthèse de nombreuses protéines non-essentielles contenant du fer, à l'image de protéines à centres Fe-S telles que la succinate-déshydrogénase SdhCAB ou encore l'aconitase AcnB (Massé and Gottesman, 2002). Cela permettrait de récupérer ce fer pour assurer les fonctions les plus vitales des bactéries, à l'image du processus de respiration. RyhB active également la synthèse du transporteur de DHB-Ser CirA (Salvail et al., 2013). Enfin, RyhB exerce une répression sur la synthèse de Fur *via* un effet sur une ORF située dans le 5'UTR de l'ARNm de *fur* (Vecerek et al., 2007). Cela engendre une transition rapide entre l'expression de Fur et celle de RyhB et des cibles de Fur (Arbel-Goren et al., 2016), permettant une réponse efficace à la carence en fer.

On remarque avec intérêt la forte conservation du sRNA RyhB, et même au-delà du sRNA, de sa fonction de régulation de protéines contenant du fer en situation de carence en fer (Salvail and Massé, 2012). En effet, si cette fonction est assurée par RyhB chez les entérobactéries et les vibrionaceae, voire par deux versions de RyhB issues d'une duplication génétique chez *S. enterica* notamment (Padalon-Brauch et al., 2008), elle est également assurée par d'autres sRNA chez des bactéries plus éloignées tels que PrrH chez *Pseudomonas aeruginosa*, ou encore NrrF chez *Neisseria meningitidis* (Salvail and Massé, 2012). Cela souligne l'importance de cette fonction régulatrice, qui apporte très vraisemblablement un avantage sélectif déterminant lors d'une situation de carence en fer (Jacques et al., 2006). Cela pose également la question de l'importance de régulations exercées spécifiquement par des sRNA plutôt que par des protéines (voir **Revue n°2**).

b. Régulation de *fepA* et de l'utilisation de l'entérobactine

J'ai consacré l'essentiel de ma thèse à la compréhension de la régulation de l'expression du gène *fepA*, c'est pourquoi j'ai décidé d'employer la dernière partie de cette introduction à dresser l'état des régulations connues avant le début de cette thèse. Cette partie sera élargie à la régulation de l'utilisation de l'entérobactine, afin d'intégrer la régulation de l'expression de *fepA* dans un processus physiologique plus complet.

L'utilisation de l'entérobactine requiert la synthèse de ce sidérophore ainsi que d'un ensemble de protéines membranaires impliquées dans sa sécrétion puis sa récupération dans la cellule une fois associé au fer. La synthèse d'entérobactine ainsi que de cet ensemble de protéines est coûteuse en énergie et en ressources, c'est pourquoi elle n'est activée que dans des conditions de carence en fer, où la bactérie va devoir utiliser tous les moyens à sa disposition pour acquérir le fer dont elle a besoin. Fur-Fe^{2+} réprime ainsi la transcription des opérons *entCEBAH*, *fes-ybdZ-entF-fepE, fepDGC, fepA-entD* ainsi que des gènes *entS*, *fepB et tolC*. Ce faisant, Fur-Fe^{2+} prévient à la fois la synthèse de l'entérobactine par les synthétases Ent, sa sécrétion par EntS et TolC, l'import des complexes Fer-entérobactine par les protéines Fep ainsi que le relargage de fer dans la cellule par l'estérase Fes (Fig. 44). Le corollaire de ces régulations par Fur est qu'en absence de fer, tous ces gènes s'expriment de manière coordonnée et permettent la synthèse et l'utilisation de l'entérobactine pour accéder au fer Fe^{3+} de l'environnement. On remarque ici qu'hormis *tolC*, l'ensemble de ces gènes sont situés au même locus chromosomique, ce qui pourrait permettre une utilisation parcimonieuse et efficace de Fur dans la mesure où tous les promoteurs à réguler sont réunis au même endroit. On note en outre que la régulation par Fur-Fe^{2+} affecte également les gènes *tonB*, *exbB* et *exbD*, réprimant la synthèse du complexe Ton lorsque l'import de fer n'est pas nécessaire.

Comme beaucoup de voies de biosynthèses, les premières étapes requièrent des régulations efficaces afin d'effectuer le plus d'économies possible sur la synthèse de protéines et les modifications des précurseurs. C'est l'une des fonctions de RyhB, qui active la synthèse des OMP ShiA (Prévost et al., 2007) et CirA (Salvail et al., 2013), respectivement impliquées dans l'import du shikimate, un précurseur de l'entérobactine, et du DHB-Ser, à la fois précurseur de l'entérobactine et servant lui-même de sidérophore de fer. En conditions de carence en fer l'import de shikimate permet d'initier le processus de synthèse de l'entérobactine. Une autre régulation importante imputée à RyhB est celle du gène *cysE*, dont la répression favorise l'utilisation de serine pour la synthèse de DHB-Ser puis d'entérobactine plutôt que de cystéine (Salvail et al., 2010).

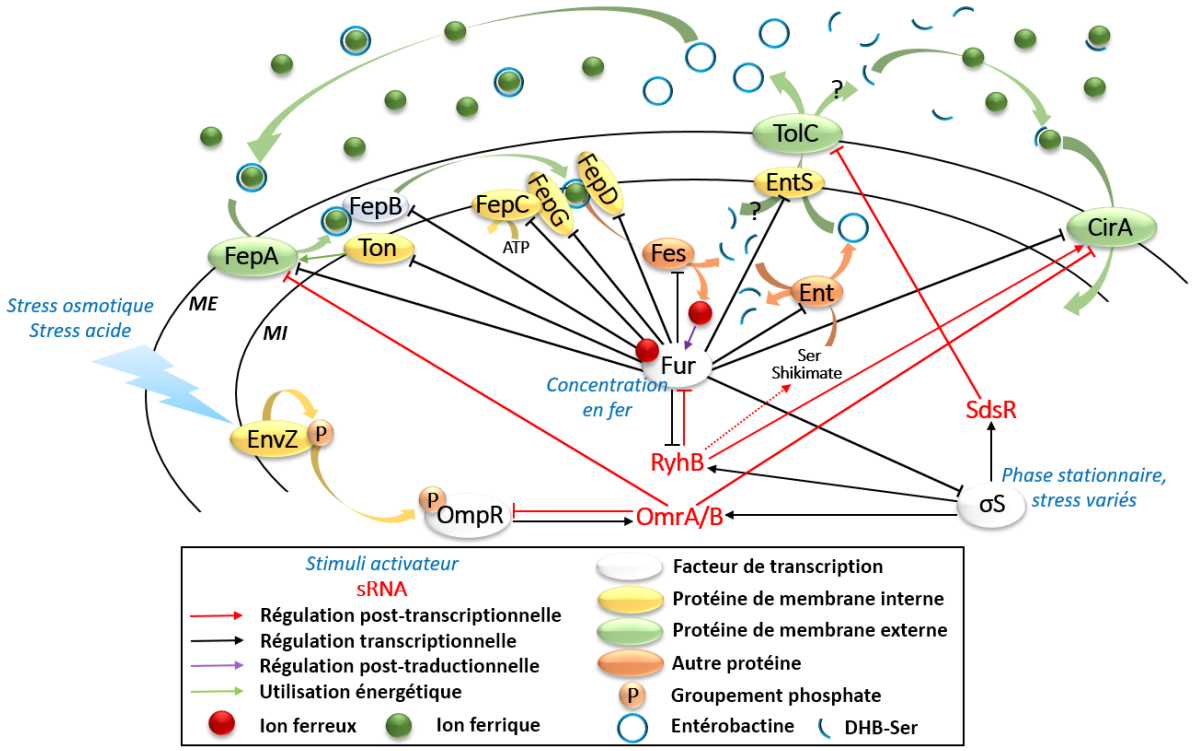


Figure 44 – Régulation de l'expression de fepA et de l'utilisation de l'entérobactine
En présence de fer, Fur-Fe^{2+} est actif et réprime l'expression de la plupart des gènes impliqués dans l'utilisation de l'entérobactine. En absence de fer, ces répressions sont levées et l'entérobactine peut être synthétisée par les protéines Ent et sécrétée par EntS-TolC. L'entérobactine peut alors chélater des ions Fe^{3+} dans le milieu environnant puis être réintégrée dans la cellule via les protéines Fep et grâce à l'énergie fournie par le complexe Ton. Enfin, l'entérobactine est clivée par Fes, qui permet à la cellule de récupérer des ions Fe^{2+} ainsi que des précurseurs de l'entérobactine pouvant être recyclés. Par ailleurs, le précurseur DHB-Ser peut lui-même être sécrété et servir de sidérophore de fer – on note que la voie de sécrétion du DHB-Ser est inconnue et pourrait correspondre à celle empruntée par l'entérobactine. En situation de carence en fer, RyhB permet l'expression de CirA et donc l'import de fer associé au DHB-Ser. RyhB active également l'import de shikimate et prévient l'utilisation de sérine (Ser) pour la synthèse de cystéine, rendant ces deux précurseurs disponibles pour la synthèse d'entérobactine. Par ailleurs, les régulateurs OmpR et σS répondent à divers stress et activent la transcription de plusieurs sRNA ayant des effets post-transcriptionnels sur l'expression des gènes fepA, cirA et tolC notamment.

Cependant le rôle des sRNA Hfq-dépendants dans l'homéostasie du fer ne s'arrête pas là, puisque l'expression de *fepA* et de *cirA* est négativement régulée par les deux sRNA OmrA/B (Guillier and Gottesman, 2006, 2008), tandis que celle de *tolC* est réprimée par SdsR (Parker and Gottesman, 2016). Ces dernières régulations peuvent être l'objet d'un questionnement vis-à-vis de leur logique physiologique : pourquoi réguler l'import de fer en conditions de pH acide pour les OmrA/B, ou en situation de stress à la fois pour OmrA et pour SdsR *via* σS? Ces questions ont été abordées au cours de cette thèse (voir **Résultats II.B**) et sont également discutées dans la conclusion (voir **Discussion B.1**).

Résultats

Chapitre I : Activation de la traduction par des structures secondaires de l'ARN messager

A. Avant-propos

Mon travail sur le mécanisme de régulation post-transcriptionnelle de l'expression de *fepA* par les sRNA OmrA/B chez *E. coli* a débuté lors de mon stage de deuxième année de Master (M2), avant d'être poursuivi en thèse. L'étude du mécanisme de régulation de l'expression du gène *fepA* par les sRNA OmrA/B a abouti à la principale découverte présentée dans ce mémoire, à savoir l'effet positif sur la fixation du ribosome, et donc sur la traduction, produit par structures secondaires d'ARNm situées en aval du RBS et au sein de l'ORF. Cette découverte a fait l'objet d'un article de recherche qui est en cours de revue dans le journal Molecular Cell (voir **Article**)

L'article sera présenté tel qu'il a été soumis après une première révision, avec une mise en page manuelle au format article. S'ensuivra une série de remarques approfondies sur certaines expérimentations qui n'ont pas été intégrées dans l'article soumis mais qui présentent un intérêt certain pour la compréhension du mécanisme de régulation impliqué et la poursuite de son étude.

B. Article de recherche: « Stem-loop structures within mRNA coding sequences activate translation initiation and mediate control by small regulatory RNAs »

Résumé:

Le démarrage de la traduction est l'étape limitante du processus de synthèse des protéines, et représente une étape cruciale de la régulation de l'expression génétique. Chez les bactéries, les structures secondaires d'ARNm ont souvent été décrites comme des éléments réduisant l'accessibilité du RBS, inhibant ainsi la fixation du ribosome et de ce fait affectant négativement la traduction.

Contre toute attente, nous avons pu mettre en évidence l'existence de structures secondaires de l'ARNm activant au contraire la fixation du ribosome et la traduction. En effet, des structures en tige-boucle ayant cet effet sont trouvées chez *E. coli* dans l'ORF des ARNm *fepA* et *bamA*, codant respectivement le récepteur aux complexes fer-entérobactine et la sous-unité essentielle du complexe d'adressage des OMP. Ces structures secondaires favorisent toutes deux la fixation du ribosome *in vitro* et la synthèse protéique *in vivo*, et ce indépendamment de leur séquence nucléotidique.

De plus, les deux sRNA Hfq-dépendants OmrA et OmrB s'apparient à la tige-boucle de l'ARNm *fepA via* un « bulge » au sein de cette structure, et répriment de ce fait la synthèse de FepA. Ces résultats nous permettent donc de proposer non seulement de nouvelles fonctions pour les structures d'ARNm, qui pourraient favoriser la traduction en permettant une fixation optimale du ribosome au RBS, mais aussi de nouveaux mécanismes d'action pour les sRNA Hfq-dépendants.

Stem-loop structures within mRNA coding sequences activate translation initiation and mediate control by small regulatory RNAs

Jonathan Jagodnik, Claude Chiaruttini and Maude Guillier*

CNRS UMR8261, Associated with University Paris Diderot, Sorbonne Paris Cité, Institut de Biologie Physico-Chimique, 75005 Paris, France

*Lead contact and corresponding author: maude.guillier@ibpc.fr

Running title: Translation activation by mRNA secondary structures

Abbreviations used: nt: nucleotide, IPTG: Isopropyl β-D-thiogalactopyranoside, dip: 2-2'-dipyridyl; ORF: open reading frame; 5' UTR: 5' untranslated region; RBS: ribosome binding site; TIR: translation initiation region; PTC: post-transcriptional control; RT: reverse transcription; CMCT: 1-cyclohexyl-3-2-(morpholinoethyl)carbodiimide metho-*p*-toluene sulfonate; DMS : dimethyl sulfate; SL: stem-loop

SUMMARY

Initiation is the rate-limiting step of translation and is thus a crucial step for translational control of gene expression. In bacteria, mRNA secondary structure is a key factor in setting the efficiency of translation initiation by occluding the ribosome binding site (RBS) and thus inhibiting ribosome binding. In striking contrast with this inhibitory effect, we report here that secondary structures located within coding sequences can instead activate translation initiation. We have found that stem-loop structures activate translation of the *Escherichia coli fepA* and *bamA* mRNAs encoding the receptor for iron-enterobactin complexes and the essential subunit of the outer membrane proteins assembly complex, respectively. Both structures promote ribosome binding *in vitro*, independently of their nucleotide sequence. Moreover, two small regulatory RNAs, OmrA and OmrB, base-pair to and most likely disrupt the *fepA* stem-loop structure, thereby repressing FepA synthesis. By expanding our understanding of how mRNA *cis*-acting elements regulate translation, these data challenge the widespread view of mRNA secondary structures as translation inhibitors, and show that translation-activating elements embedded in coding sequences can be targeted by small regulatory RNAs to inhibit gene expression.

Key words: Hfq, siderophore, translation-activating structure, regulatory RNA, non-coding RNA, outer membrane proteins, regulatory mechanism, starting block, toeprint

INTRODUCTION

Successful adaptation of bacteria to multiple niches relies on their ability to rapidly and precisely tune gene expression in response to extremely diverse environmental cues. While control can occur at virtually all stages of the gene expression pathway, regulating translation and/or the stability of target mRNAs offers several advantages for adaptive responses, in particular the speed of response by acting directly at the mRNA level, and the possibility to prevent transcriptional noise or to differentially regulate individual cistrons within an operon.

An important feature of post-transcriptional control is the wide diversity of underlying molecular mechanisms and regulators, in line with the remarkable plasticity of mRNAs, which are the controlled targets. This is particularly well illustrated by the plethora of small regulatory RNAs (sRNAs) identified in bacteria, many of which base-pair *via* short and imperfect duplexes to target mRNAs and thereby modulate their translation and/or stability (Wagner and Romby, 2015). In enterobacteria, these imperfectly pairing sRNAs require an RNA chaperone called Hfq that stabilizes many sRNAs and facilitates sRNA-mRNA duplex formation (Vogel and Luisi, 2011). Interestingly, sRNAs have been shown to negatively or positively control gene expression using a remarkable variety of molecular mechanisms (Jagodnik et al., 2017).

In the most frequent scenario, sRNAs repress the expression of targets by binding at or in the vicinity of the RBS, thereby directly competing with the 30S ribosomal subunit for binding to the mRNA. However, they can also affect the accessibility of the RBS when they base pair outside of this region by inducing structural changes that ultimately leads to a more or less accessible Shine-Dalgarno (SD) region, and thus to a positive (Mandin and Gottesman, 2010 and references therein) or negative (Heidrich et al., 2007) control. Furthermore, the primary effect of sRNA binding to mRNAs is not necessarily translational: sRNAs have been reported to mediate target mRNA destabilization by guiding the endoribonuclease RNase E to the mRNA, for instance (Bandyra et al., 2012; Pfeiffer et al., 2009), or, in contrast, to stabilize target mRNAs (Obana et al., 2010; Papenfort et al., 2013; Ramirez-Pena et al., 2010). Yet other mechanisms have been described that bring new insight into translation mechanisms. For instance, the study of IstR sRNA repression of *tisB* mRNA expression identified a ribosome stand-by site that allows translation from a downstream translation initiation region (TIR) that is sequestered in a stem-loop structure (Darfeuille et al., 2007). A variation on this theme was later proposed for the translational activation of *iroN* by the *Salmonella* RyhB sRNA homologues that would allow entry of

the ribosome at an upstream stand-by site (Balbontin et al., 2016).

OmrA and OmrB are two Hfq-dependent sRNAs conserved in enterobacteria. Their transcription is activated by the OmpR regulator, either in its phosphorylated form or under high concentrations of its non-phosphorylated form (Brosse et al., 2016). The transcription of *omrA* has also been shown to specifically respond to the RpoS sigma factor (Levi-Meyrueis et al., 2014; Peano et al., 2015). OmrA and OmrB target several mRNAs by base-pairing interaction *via* their almost identical 5'-ends, which are highly conserved in other bacteria. As a consequence, all the target mRNAs described so far, all negatively regulated, are common to both sRNAs. Several of these targets encode transcriptional regulators, such as CsgD and FlhDC, the master regulators of curli and flagella synthesis respectively, or the EnvZ-OmpR two-component system, i.e. the Omr transcriptional activator (Brosse et al., 2016; De Lay and Gottesman, 2012; Holmqvist et al., 2010). In addition, OmrA/B down-regulate the synthesis of several outer membrane proteins: OmpT, CirA, FecA and FepA (Guillier and Gottesman, 2006, 2008). While OmpT is a protease, CirA, FecA and FepA are receptors for iron-siderophore complexes involved in iron uptake. In line with this function, *cirA*, *fecA* and *fepA* genes are members of the Fur regulon and their transcription is induced under conditions of iron deprivation. In addition, synthesis of CirA also responds to iron starvation at the post-transcriptional level, this time positively, through translational activation mediated by the Fur-repressed RyhB sRNA (Salvail et al., 2013). These three receptors bind to different siderophores and FepA is likely to play a key role in iron homeostasis, as it is the receptor for iron chelated by enterobactin, i.e. the most potent siderophore.

From a mechanistic standpoint, the Omr sRNAs were shown to pair with the TIR of *ompT, cirA, ompR-envZ and flhDC* mRNAs, thus presumably repressing their translation by the classical mechanism of competition with ribosome binding. However, the question of how the Omr achieved regulation of *fecA* and *fepA* genes has not yet been addressed. This question appears all the more pertinent as (i) OmrA/B were previously shown to repress *csgD* expression by targeting the *csgD* mRNA outside of the TIR (Holmqvist et al., 2010) and (ii) there is no obvious complementarity between the Omr 5'-ends and the TIR of *fecA* or *fepA* mRNA.

In the present work, we have investigated the mechanism of *fepA* control by OmrA/B in detail and found that both sRNAs target a stem-loop (SL) structure within the *fepA* coding sequence. Unexpectedly, this SL activates translation initiation by promoting formation of the pre-initiation complex between the *fepA* mRNA, the 30S ribosomal subunit and the initiator tRNA. Introduction of compensatory changes within the SL further indicated that this effect is independent of its nt sequence. Importantly, similar SL can be predicted in several bacterial mRNAs and we have shown that the one in the essential *bamA* gene also activates translation initiation. By uncovering the positive role played by secondary structures of the mRNA, these results shed a new light on the translation initiation step in bacteria and the role of mRNA *cis*-elements in this process.

RESULTS

OmrA/B directly repress *fepA* expression by base pairing within its coding sequence

We previously showed that pulse-induction of OmrA/B led to a decrease in *fepA* mRNA levels, suggesting that *fepA* might be a direct target of these sRNAs (Guillier and Gottesman, 2006, 2008). This prompted us to search for possible base-pairing between *fepA* and OmrA/B using IntaRNA (Wright et al., 2014) and TargetRNA (Kery et al., 2014) programs. Because a mutant in the 5'-end of OmrA/B no longer affected *fepA* mRNA levels (Guillier and Gottesman, 2008), we only used the 20 conserved 5'-end nts of OmrA/B for this search; the *fepA* mRNA fragment used encompassed nts -173 (which corresponds to the transcription start site) to +80 (+1 being the first nt of *fepA* ORF). The resulting prediction, with some manual adjustment, shows that nts 1 to 26 of OmrA/B can potentially pair imperfectly with nts 23 to 44 of *fepA* open reading frame (ORF) (Fig. 1A). The existence of this interaction was further validated using a translational *fepA-lacZ* fusion, where *fepA* 5'-UTR and first 45 nts of the ORF were fused

upstream of the 10th codon of *lacZ*. This protein fusion is constitutively expressed from a P_{tet} promoter (Fig. 1B). Overproduction of OmrA or of OmrB reduced expression of this fusion by 8.3 and 4.7-fold, respectively (Fig. 1C). This reduction was no longer observed when mutations were introduced in the regions predicted to pair, either in nts 3-6 of OmrA/B or in nts 39-42 of the *fepA* ORF. Importantly, these mutant forms of the Omr sRNAs accumulate to even higher levels than their wild-type (wt) counterparts and the loss in repression is thus not due to differential levels of the different sRNAs (Fig. S1A). A combination of these mutations, which re-establishes base-pairing, partially restored control: OmrA/Bmut repressed *fepAmut-lacZ* by 2.7- and 1.8-fold, respectively (Fig. 1C). These data strongly support the *in vivo* interaction between OmrA/B 5'-ends and the proximal region of the *fepA* ORF. This regulation was confirmed by following directly the levels of the FepA protein made from its endogenous locus during a short overexpression of OmrA/B sRNA. More precisely, cells transformed with OmrA/B overexpressing plasmid were grown to mid-exponential phase, split in two, and half of the culture was treated with 250 µM 2-2'-dipyridyl (dip) and 100 µM IPTG to induce *fepA* expression and OmrA/B synthesis, respectively. As a control, the other half of the culture was left untreated. After 15 minutes, proteins and RNA samples were extracted and the levels of FepA protein and of OmrA and OmrB sRNAs were followed by western- or northern-blot analysis, respectively. In presence of the vector control, addition of dip induced FepA synthesis, as expected since *fepA* transcription is Fur-repressed. This induction was strongly reduced in presence of plasmids overexpressing either wt OmrA or OmrB, but was completely restored when the mutant versions of OmrA/B, that accumulate at similar levels than their wt counterparts, were overexpressed instead (Fig. 1D). OmrA/B sRNAs thus specifically and rapidly repress synthesis of the FepA protein *via* base-pairing interaction.

Only short regions of *fepA* mRNA and of OmrA/B are involved in the interaction

Identification of one site of interaction between an sRNA and its target mRNA does not rule out the existence of other secondary pairing sites. We therefore asked whether other regions of *fepA* or of the Omr sRNAs were required for control. We first analyzed the repression of a shorter *fepA-lacZ* fusion lacking most of the *fepA* 5'-UTR (*fepA*$_{-26}$ in Fig. 1B). Expression of this fusion was repressed by OmrA/B as efficiently as the longer fusion (Fig. 1E), showing that most of the *fepA* 5'-UTR is dispensable for control. To then determine whether the 5'-end of OmrA/B is the only region involved in pairing with *fepA* mRNA, we used a chimeric sRNA construct called OmrY which consists of the first 15 nts of OmrA/B followed by the last 61 nts of the RybB sRNA (Fig. 1F). As RybB is an Hfq-dependent sRNA that regulates most of its targets *via* its 5'-end, the resulting OmrY is expected to display the target specificity of OmrA/B while retaining all other properties of an Hfq-dependent sRNA. Overexpression of wt OmrY efficiently repressed expression of the *fepA-lacZ* fusion (Fig. 1F) as well as synthesis of the FepA protein (Fig. S1B), similar to what is observed with OmrA/B. Furthermore, repression was abolished when nts 3-6 of OmrY were mutated (this did not reduce sRNA levels (Fig. S1B)), showing that repression by OmrY also primarily relies on its 5'-end and not on regions of the RybB sRNA (Fig. 1F). Together, these data suggest that the pairing site between nts 1-15 of OmrA/B and nts 34-44 of *fepA* is the primary site of interaction, and is sufficient to provide full control. Interestingly, the region of *fepA* mRNA targeted by OmrA/B is located within the *fepA* ORF but outside of the region bound by the 30S ribosomal subunit during translation initiation (Huttenhofer and Noller, 1994; Yusupova et al., 2001), raising the question of the underlying molecular mechanism of control.

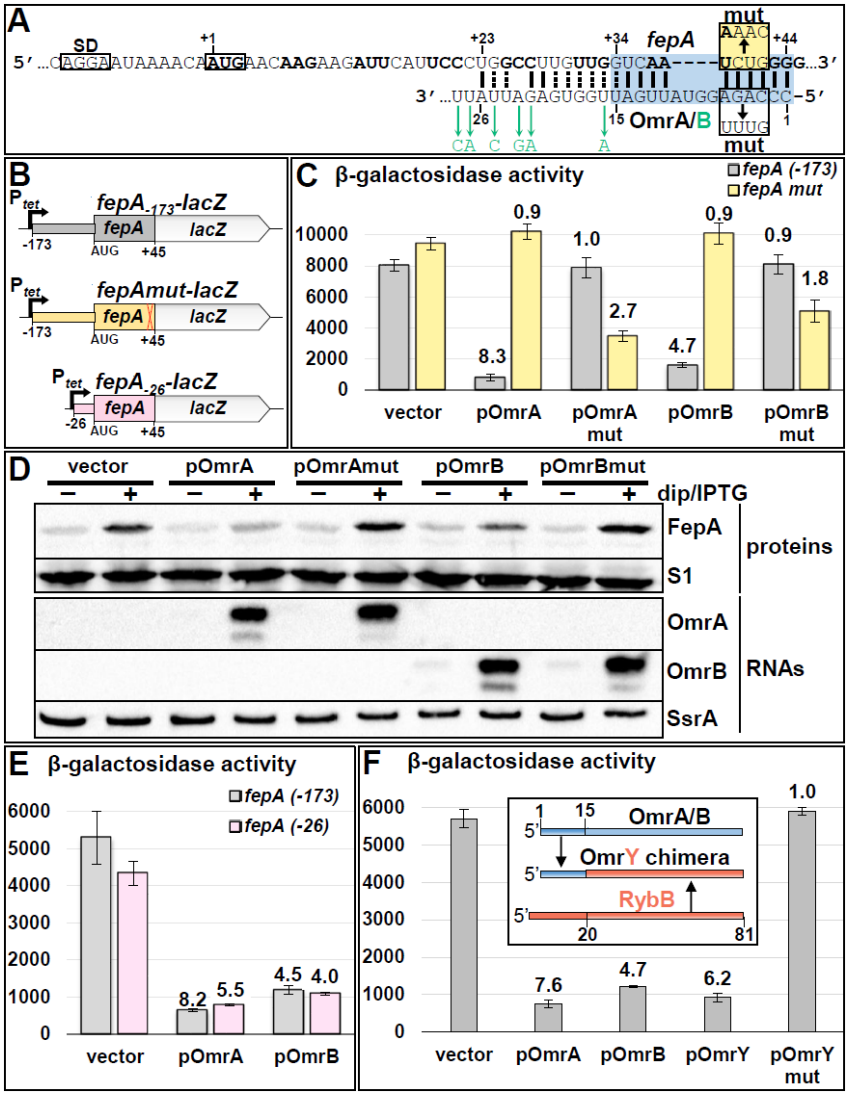


Figure 1. Direct repression of *fepA* expression by the 5'-end of OmrA/B relies on a short interaction within *fepA* ORF.

(A) Predicted base-pairing between *fepA* mRNA and OmrA/B sRNAs. Successive black and grey codons indicate *fepA* reading frame. OmrB nts that differ from OmrA are shown below the OmrA sequence. The mutations introduced in *fepA-lacZ* or in OmrA/B are boxed (mut). The pairing region shown to be necessary and sufficient for regulation is highlighted in grey. (B) Various *fepA-lacZ* translational fusions used in this study. Positions on *fepA* mRNA are given relative to the start codon. (C) β-galactosidase activities of *omrAB* deleted strains carrying the *fepA*$_{-173}$*-lacZ* (strain JJ0015) or *fepAmut-lacZ* (strain JJ0019) fusion and transformed with OmrA, OmrAmut, OmrB or OmrBmut overexpressing plasmids, or a vector control. (D) Western-Blot analysis of the FepA protein levels in response to wt or mutant OmrA or OmrB induction. Levels of the different sRNAs were followed by northern-blot in the same experiment. *E. coli* ribosomal protein S1 or SsrA RNA were used as loading controls for proteins and RNAs, respectively. (E) β-galactosidase activities of the *fepA*$_{-173}$*-lacZ* (strain MG1772) and the *fepA*$_{-26}$*-lacZ* (strain JJ0135) fusion strains transformed with OmrA or OmrB overexpressing plasmids, or a vector control. (F) β-galactosidase activity of the *fepA*$_{-173}$*-lacZ* (strain JJ0015) fusion strain transformed with OmrA, OmrB, OmrY or OmrYmut overexpressing plasmids, or a vector control. The framed schematic displays the design of the OmrY chimera: conserved nts 1 to 15 of OmrA/B are fused to nts 20 to 81 of RybB. β-galactosidase activities are expressed in Miller units. Numbers above the bars give the repression factors.

RNase E endonuclease is not required for *fepA* repression by OmrA/B

Other examples of bacterial sRNAs pairing to coding regions outside of the ribosome binding site (RBS) have been reported previously (Heidrich et al., 2007; Papenfort et al., 2013; Pfeiffer et al., 2009). In the case of the repressive MicC-*ompD* mRNA interaction in *Salmonella*, it was shown that this pairing far from the RBS did not affect translation initiation but instead induced an RNase E-dependent destabilization of the target mRNA (Bandyra et al., 2012; Pfeiffer et al., 2009). This prompted us to investigate whether RNase E was also required for OmrA/B control of *fepA* expression by following the synthesis of FepA protein in wt or *rne*^{ts} strains transformed with OmrA/B overexpressing plasmids. In this experiment, cells were grown at 37°C up to mid-exponential phase and cultures were split in two. One aliquot continued to grow at 37°C as a control; the second was shifted to 44°C for 15 minutes to inactivate the thermosensitive RNase E, in presence of 250 µM 2-2'-dipyridyl (dip) to induce FepA synthesis. Proteins and RNAs were extracted from both aliquots and

specific proteins or mRNAs were analyzed by Western- or Northern-blotting, respectively (Fig. 2). In wt cells, addition of dip induced FepA synthesis in presence of the empty vector, and induction was reduced upon OmrA/B overproduction, as expected. Strikingly, a highly similar pattern for FepA protein was observed with the empty vector or the OmrA overproducing plasmid in *rne*^{ts} cells. Inactivation of RNase E activity was confirmed by the fact that *ompT* mRNA was not degraded in response to OmrA/B induction, as previously observed (Guillier and Gottesman, 2008). Thus, RNase E is not required for the control of FepA protein levels by OmrA. For OmrB, control of FepA levels in the *rne*^{ts} mutant was not as efficient as in the *rne*$^{+}$ cells, but was nonetheless still visible, suggesting that repression still occurs.

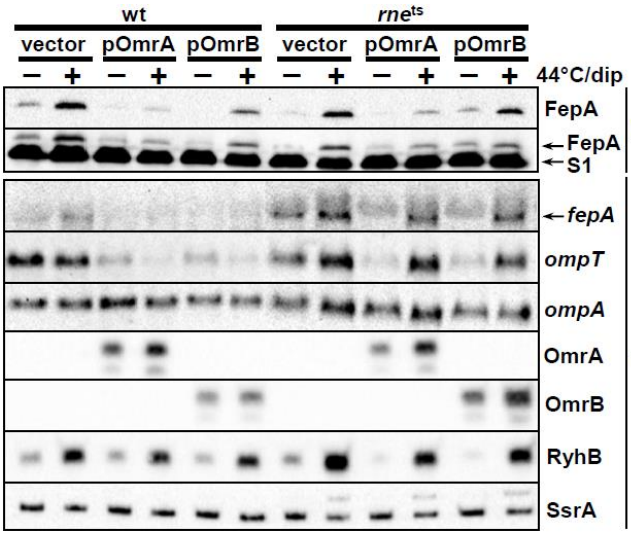


Figure 2. The regulation of FepA synthesis by OmrA/B is RNase E-independent.
The levels of FepA protein and of *fepA* mRNA were analyzed in wt or *rne*^{ts} strains overproducing OmrA or OmrB. Strains MG1325 (wt) or MG1326 (*rne*^{ts}) transformed with indicated plasmids were grown at 37°C to mid-exponential phase and either kept at 37°C (-) or shifted to 44°C in the presence of dip (+). Specific proteins and RNAs were then analyzed by western- or northern-blotting, respectively. Detection of S1 was performed as a loading control after detection of FepA, which explains the presence of the FepA signal. Probing of *ompA* and SsrA RNAs were also loading controls, while monitoring the levels of RyhB sRNA was used as a control for dip induction.

While these data are not sufficient to rule out a role for the RNase E in achieving the full range of regulation, especially for OmrB, they nonetheless show that OmrA/B still control the levels of FepA protein in the absence of RNase E. In contrast, the decrease in *fepA* mRNA levels observed upon OmrA/B overproduction clearly required a functional RNase E, suggesting that *fepA* mRNA degradation by RNase E is a consequence of the inhibition of *fepA* translation by OmrA/B.

OmrA/B 5'-end controls *fepA* translation initiation by inhibition of ribosome binding

Even though the previous result did not rule out a primary effect of OmrA/B on *fepA* mRNA stability *via* the involvement of ribonucleases other than RNase E, it nonetheless led us to investigate whether OmrA/B affected ribosome binding to the *fepA* mRNA. For this purpose, toeprinting experiments were performed on the *fepA*$_{-173}$ transcript extending from nts -173 to +129 to follow the formation of ternary translation initiation complexes *in vitro*, with and without sRNA. Addition of the 30S ribosomal subunit and initiator tRNA induced reverse-transcription (RT) stops at positions +15/+16 of the *fepA* ORF ("toeprint" signal in Fig. 3A). Despite the fact that only the wt OmrA or OmrB, and not the mutant derivatives, seemed to bind to *fepA* mRNA in this experiment, the toeprint signal was non-specifically decreased in presence of any of these sRNAs (Fig. 3A and Fig. S2). To circumvent this lack of specificity, the experiment was repeated with OmrA and OmrAmut in the presence of increasing concentrations of the Hfq RNA chaperone. Addition of Hfq alone led to a strong inhibition of the toeprint at high concentrations (Fig. 3B). This is again most likely non-specific, as it is not observed in the presence of sRNAs known to strongly bind Hfq, such as OmrA or OmrAmut. Furthermore, when added in conjunction with sRNAs, Hfq both increased formation of the OmrA-*fepA* duplex, as

previously observed for other sRNA-mRNA pairs *in vitro* (e.g. Kawamoto et al., 2006), and allowed a specific inhibition of toeprint formation by OmrA (Fig. 3B). At this stage, it was not clear whether this was simply due to the stronger binding of OmrA to *fepA* mRNA in the presence of Hfq or whether Hfq could play a more direct role in the translation inhibition. To discriminate between these possibilities, the experiment was then performed without Hfq and in presence of increasing concentrations of the chimeric OmrY sRNA or the 5'-end mutant variant. OmrY, but not OmrYmut, produced a specific inhibition of the toeprint, possibly due to a stronger binding of OmrY than OmrA/B to the *fepA* mRNA, as it induced a stronger RT-stop at nt G+44 of *fepA* ORF (Fig. 3A and Fig. S2). Finally, we also analyzed the effect of a 16-mer DNA oligonucleotide mimicking the 5'-end of OmrA/B, again in the absence of Hfq. This antisense DNA (AsDNA), referred to as As28.43, is perfectly complementary to nts 28 to 43 of the *fepA* ORF, i.e. the nts that base pair with nts 2-6 and 11-21 of OmrA/B (Fig. 3C). As28.43 DNA efficiently bound to the *fepA* mRNA *in vitro*, as shown by the RT-stop it induced at position G+43, and it strongly inhibited the toeprint, even at the lowest AsDNA/mRNA ratio used in this experiment. And as expected, neither binding nor toeprint inhibition was observed with the mutant derivative of As28.43, showing that the effect of As28.43 is specific (Fig. 3D).

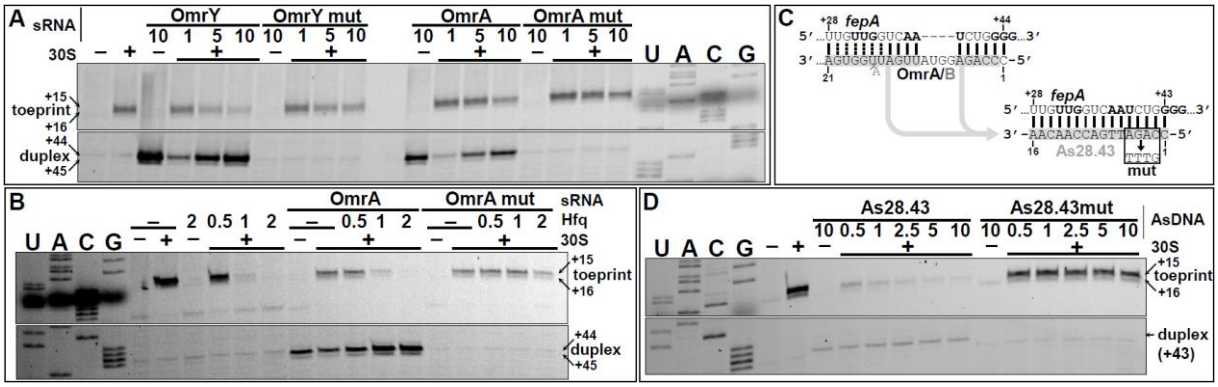


Figure 3. OmrA/B 5'-end inhibits binding of the 30S ribosomal subunit to *fepA* mRNA.
Toeprinting assays on *fepA*$_{-173}$ mRNA in the presence of increasing concentrations of wt or mutant OmrA or OmrY sRNAs (A) or of a fixed concentration of OmrA(mut) and increasing amounts of purified Hfq (B). "Toeprint" and "duplex" indicate RT stops due to 30S/initiator tRNA binding to mRNA or to pairing of the sRNA, respectively. Full-size gel of panel A is shown in Fig. S2A and Fig. S2B shows a repeat of this experiment including OmrB and OmrBmut sRNAs. (C) Design of the As28.43 DNA mimicking the Omr 5'-end interacting with *fepA* mRNA. The sequence of this AsDNA is based on the Omr nts highlighted in grey, modified in order to have only Watson-Crick bp in the mRNA-AsDNA duplex. The same mutation as that introduced in the Omr 5'-end was introduced in As28.43. (D) Toeprinting assay on *fepA*$_{-173}$ transcript in the presence of increasing concentrations of wt or mutant As28.43 AsDNA. In panels A, B and D, numbers above the gels indicate the ratio of the sRNA, hexameric Hfq or AsDNA over the *fepA* mRNA. Full-size gels of panels B and D are shown in Fig. S3 and S4, respectively.

Together, these data show that, at least in presence of Hfq, OmrA/B repress *fepA* expression at the level of translation initiation by preventing binding of the 30S ribosomal subunit. Furthermore, the results obtained with OmrY or the short AsDNA indicate that the role of Hfq is limited to facilitating and/or stabilizing duplex formation, and that the OmrA/B sequence downstream of nt +15 is dispensable for this effect. In other words, the 5'-end of OmrA/B is sufficient to inhibit translation initiation *in vitro* by pairing within *fepA* ORF downstream of nt +28. Intriguingly, this lies outside of the region bound by the 30S ribosomal subunit during initiation complex formation.

The RBS is not the only *fepA* mRNA region whose accessibility affects ribosome binding

Using an antisense scanning approach, it was shown previously that, for several *Salmonella* mRNAs, the ability of AsDNAs to compete with ribosome binding was lost when pairing occurred downstream of nt +14 of the ORF (Bouvier et al., 2008). We therefore performed a similar antisense scanning experiment on the *fepA*$_{-173}$ transcript to determine whether the *fepA* ORF displays a specific sensitivity to AsDNAs binding, as suggested by the effect of As28.43. In this experiment, the toeprint signal on *fepA* mRNA was monitored in the presence of 20 different AsDNAs, perfectly complementary to the early *fepA* ORF with similar melting temperatures and whose 3'-ends were consecutively shifted by 2 nts (Fig. 4A).

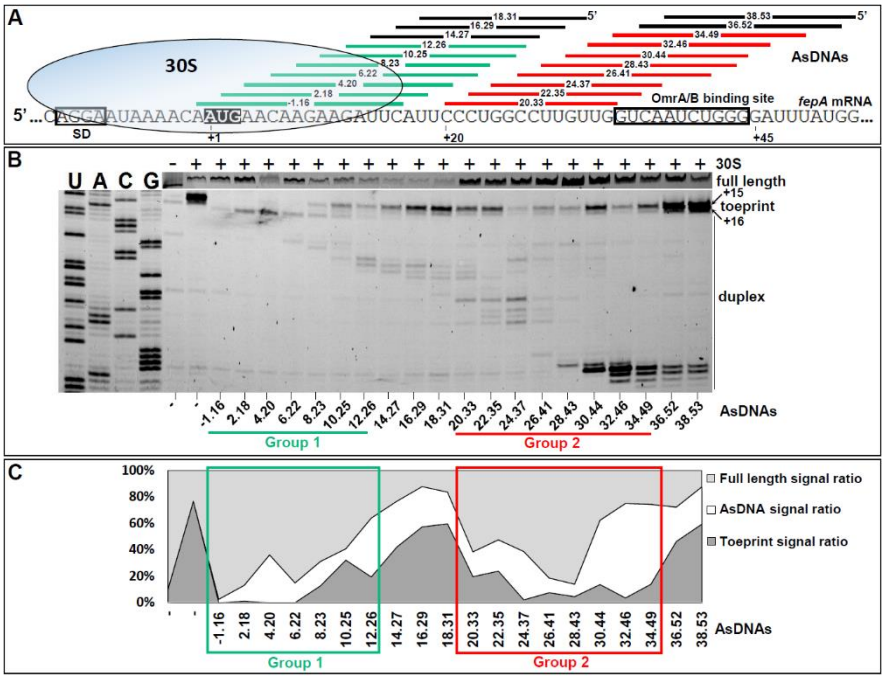


Figure 4. Formation of the *fepA* translation initiation complex depends on the accessibility of two distinct regions of the mRNA

(A) Schematic of the different oligonucleotides (AsDNA) used in the antisense-scanning experiment. AsDNAs that inhibit or not the toeprint signal are indicated in color and in black, respectively. The expected position of the 30S subunit on *fepA* mRNA in the translation initiation complex is shown schematically, together with the *fepA* SD sequence, AUG and the minimal OmrA/B binding site identified in Fig. 1. (B) Toeprinting assay on *fepA*$_{-173}$ transcript in presence of the various AsDNAs in a 10-fold molar excess over the mRNA. The corresponding full size gel is shown in Fig. S5. (C) Quantification of the toeprinting assay shown in (B). The sum of the full length, toeprint and AsDNA signals was set up at 100% for each lane. Hence, the toeprint is represented here as a percentage of total lane signals.

A first group of AsDNAs (As-1.16 to As12.26) strongly inhibited toeprint, in agreement with the fact that they bind to *fepA* RBS (Fig. 4B and 4C). In contrast, the three next AsDNAs, As14.27 to As18.31, base pair mostly downstream of the region bound by the 30S ribosomal subunit during translation initiation and, similar to what was observed previously (Bouvier et al., 2008), no longer inhibited toeprint formation. Surprisingly, however, the eight following scanning AsDNAs, As20.33 to As34.49 (group 2), recovered the ability to inhibit the toeprint, even though they base pair outside of the RBS (Fig. 4). It is worth noting that binding of the various AsDNAs to *fepA* mRNA induced RT arrests of variable intensities downstream of the toeprint position, which in theory could induce a decrease in the toeprint signal by blocking RT progression. However, our experimental data seem to exclude this possibility: the last two scanning oligos, As36.52 and As38.53, induce strong RT stops at positions +45 and +46 but do not inhibit toeprint formation. This strongly suggests that the effect of As20.33 to As28.43 on the toeprint is not due to the blocking of RT as those AsDNAs induce weaker RT stops. Even though the use of AsDNAs is artificial, these data led us to consider that the *fepA* mRNA displays two distinct regions that positively impact the binding of the 30S subunit. The first corresponds to the RBS, and the second, which is more surprising, to the region extending from nts 20 to 35 of *fepA* ORF. Interestingly, this

second region partially overlaps but is not restricted to the OmrA/B binding-site. Furthermore, a secondary structure prediction proposes the existence of an imperfect SL structure within the *fepA* ORF from nts 19 to 46 (Fig. 5A), opening the possibility that the group 2 AsDNAs, as OmrA/B, might inhibit toeprint formation *via* disruption of this structure.

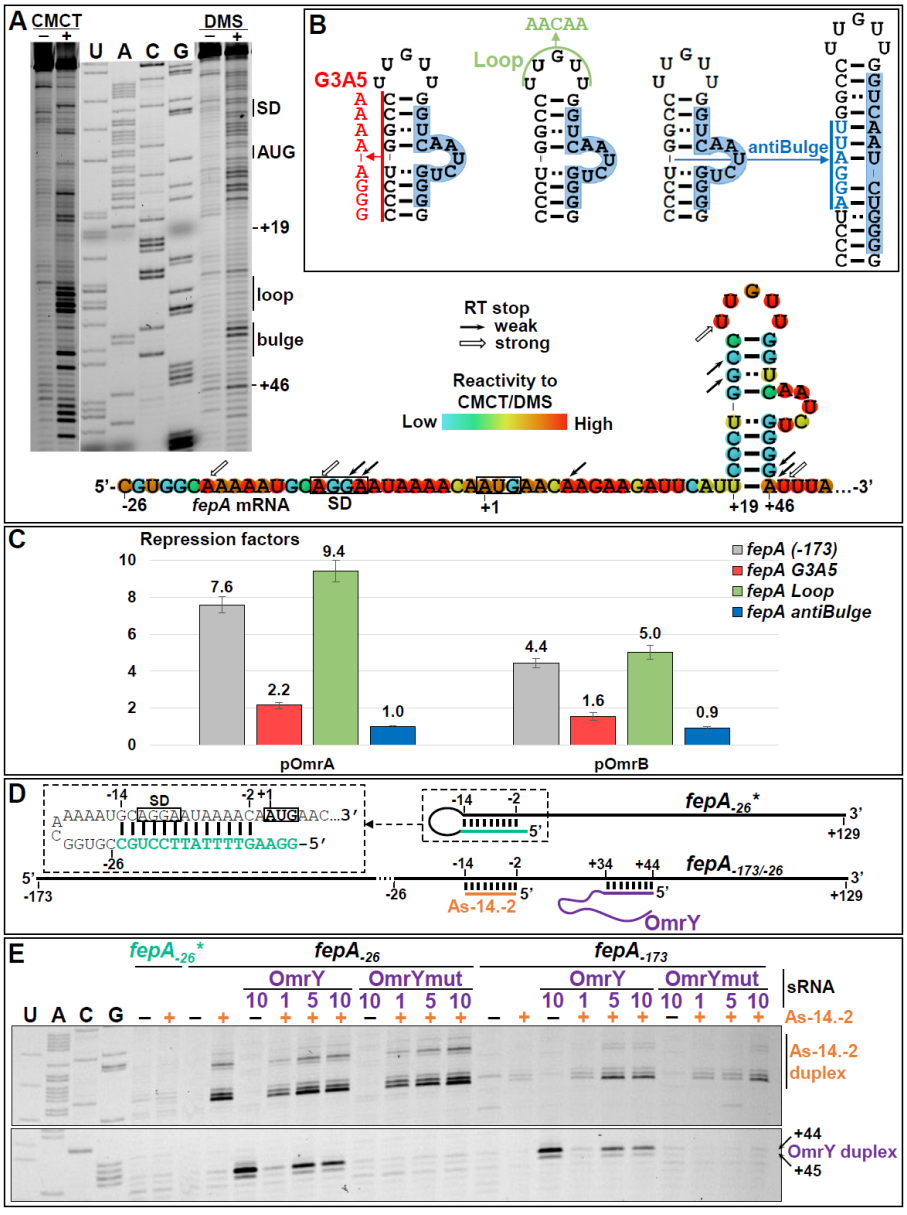


Figure 5. Regulation by OmrA/B relies on the disruption of a secondary structure in *fepA* mRNA, which does not modify RBS accessibility

(A) *In vitro* chemical probing assay of the secondary structure of the *fepA*$_{-26}$ transcript using CMCT and DMS. Stronger RT stops in the presence of the probe (+) than in its absence (-) indicate accessible nts, which are therefore considered as unfolded. Positions relative to *fepA* start codon are indicated to the right of the gel. The full size gel is reproduced in Fig. S6. A model of the secondary structure of the 5' region of *fepA*$_{-26}$ derived from probing is shown. Nts are colored according to their reactivity to DMS or CMCT while arrows indicate RT stops that are not probe-dependent. (B) Mutations designed to disrupt the predicted secondary structure (G3A5), to alter the sequence of the loop (Loop) or to pair all nts of the OmrA/B binding-site to nts of the 3' strand of the SL (antiBulge). The bp at the base of the SL (U-A in *fepA* mRNA and U-G in *fepA-lacZ*) is omitted in this representation. Nts highlighted in blue indicate the minimal Omr binding-site as defined in Fig. 1. (C) Repression by OmrA or OmrB of the *fepA*$_{-173}$*-lacZ* fusion, wt or carrying the G3A5, Loop or antiBulge mutation. Strains used in this experiment are JJ0015, JJ0286, JJ0281and MG2247; raw data used to calculate the repression factors are given in Fig. S7. (D) Principle of the RT experiment performed to assess the *fepA* mRNA RBS accessibility: following duplex formation between OmrY or OmrYmut and either *fepA*$_{-26}$ *or fepA*$_{-173}$ transcript, binding of the As-14.-2 AsDNA to the transcripts was analyzed by RT. The construct *fepA*$_{-26}$* is a modification of *fepA*$_{-26}$ carrying 16 extra nts at the 5'-end that can form a secondary structure occluding the SD region. (E) Selected portions of the gel analyzing the RT-products. Upper and lower panels show the RT-stops due to As-14.-2 binding and to OmrY-*fepA* mRNA duplex formation, respectively. The full gel is shown in Fig. S8.

The *fepA* region targeted by OmrA/B forms a stem-loop structure that is required for control by these sRNAs

To provide experimental support for the predicted SL, a structural probing analysis of the *fepA*$_{-26}$ transcript extending from nts -26 to +129 was performed *in vitro* using the chemical probes DMS (that modifies adenosine and to a lesser extent cytosine residues) and CMCT (that modifies guanosine and uridine residues). As shown in Fig. 5A, most of the nts located in the TIR of *fepA* displayed a moderate to strong reactivity towards the probes, suggesting that this region of *fepA* mRNA is mostly unstructured. In contrast, most nts between positions +19 to +46 were poorly reactive, except for nts 28-32 and 37-41 corresponding to the loop and bulge of the predicted SL. The structural probing data thus strongly support the existence of this SL. Because it encompasses the OmrA/B binding site, these results also suggest that control of *fepA* expression by these sRNAs would rely on disruption of this secondary structure.

This was more directly assessed by following repression by OmrA/B of a *fepA-lacZ* fusion where formation of the SL was impaired by changing nts +20 to +27 from CCCUGGCC to GGGAAAAA (G3A5 mutation in Fig. 5B). The ability of the Omr to repress this mutant version of *fepA-lacZ* was strongly reduced: OmrA and OmrB repressed its expression by 2.2- and 1.6-fold respectively, while they repressed expression of the wt fusion by 7.6- and 4.4-fold (Fig. 5C). Because pairing of OmrA/B to nts 20-27 of the *fepA* ORF is not required for control (Fig. 1), the effect of the G3A5 mutant is most likely due to its preventing formation of the SL. In contrast, changing nts +28 to +32 in the loop from UUGUU to AACAA (Loop mutation in Fig. 5B) did not affect the ability of OmrA/B to control *fepA* (repression factors of 9.4- and 5-fold, respectively, Fig. 5C). This result indicates that pairing of the OmrA/B downstream of nt +32 of *fepA* ORF is sufficient for the control, confirming the primary role of the Omr 5'-end in this interaction.

It is striking that several of the *fepA* nts that pair to OmrA/B are located in the bulge and we wondered whether this was a requisite for the control. We thus engineered a mutant version of the *fepA-lacZ* fusion where 6 extra nts (AGGAUU), 5 of which being perfectly complementary to the bulged nts, have been inserted in the 5' strand of the SL (Fig. 5B). In this "antiBulge" mutant, all nts of the Omr pairing region are thus expected to pair with nts from the 5' strand and, interestingly, control by OmrA/B was completely abolished (Fig. 5C). These data are not only consistent with the formation of the SL, but they also strongly suggest that the bulge is required for allowing the Omr 5'-end to interact with *fepA* mRNA, possibly serving as an anchor point for the sRNAs.

The stem-loop activates FepA synthesis in a sequence-independent manner

Because OmrA/B likely inhibit *fepA* translation by disrupting the SL, we next investigated the role of this structure in *fepA* expression. For this purpose, the effect of diverse mutations on *fepA* expression was analyzed in the context of both the short and long fusions described earlier. A first group of mutants consisted of changes of 5 or 8 nts on either side of the stem (mutants A5, U5, G3A5 and U5C3 in Fig. 6A), which should impair formation of the SL. All four mutants decreased expression of both the short and the long *fepA-lacZ* fusions, with a stronger effect observed for the 8 nts changes (Fig. 6B). These changes were then combined to re-establish the SL structure (mutants Comp1 and Comp2, Fig. 6A), which restored, at least partially, a higher *fepA* expression. In the case of the Comp1 change in the short fusion, the restoration of *fepA* expression was essentially complete (Fig. 6B). These effects were independent of the Omr sRNAs as the same results were obtained in strains deleted for *omrAB* (Fig. S9). These data strongly suggest that formation of the SL promotes FepA synthesis. Furthermore, as *fepA* expression was also increased with the Comp1 and Comp2 constructs, where most of the pairing nts of the SL have been changed, it appears that this effect is largely sequence-independent. The fact that the Comp2 change is not as efficient as Comp1 in restoring *fepA* expression could be due to a slight difference in the topology and/or the stability of the Comp2 SL compared to the wt structure.

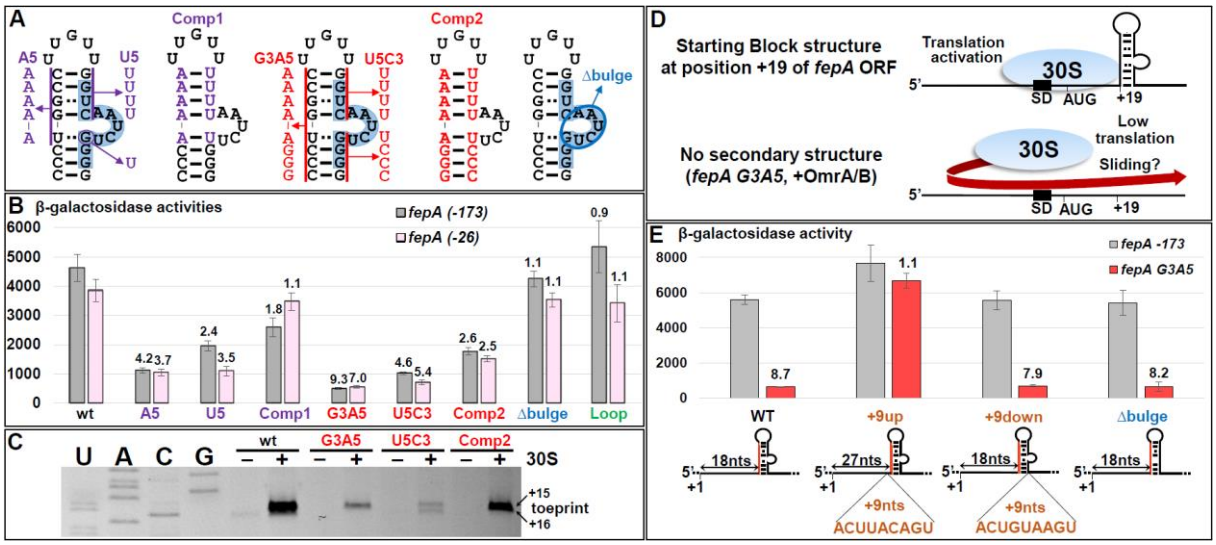


Figure 6. The *fepA* mRNA stem-loop structure activates translation in a sequence-independent manner, possibly by acting as a "starting block" structure

(A) Mutations designed to alter (mutants A5, U5, G3A5 and U5C3) or restore (mutants Comp1 and Comp2) the SL structure. The Δbulge mutant consists in the deletion of nts 37-42. The Loop mutant is depicted in figure 5B. The minimal OmrA/B binding site as defined in Fig. 1 is highlighted in blue. (B) β-galactosidase activities of strains carrying these different versions of the *fepA*$_{-173}$-*lacZ* or the *fepA*$_{-26}$-*lacZ* fusion. Strains used in this experiment are, from left to right: MG1772, JJ0150, JJ0225, JJ0190, JJ0233, JJ0229, JJ0237, JJ0241 and JJ0192 for the *fepA*$_{-173}$-*lacZ* variants; and JJ0135, JJ0120, JJ0227, JJ0180, JJ0235, JJ0231, JJ0238, JJ0243 and JJ0193 for the *fepA*$_{-26}$-*lacZ* variants. (C) Toeprinting assay on *fepA*$_{-173}$ mutant transcripts. Only the portion of the gel with the toeprint signal is shown; full size gel is reproduced in Fig. S10. (D) Model of the starting block mechanism: the SL at position +19 of *fepA* mRNA would limit scanning of the 30S ribosomal subunit at the proper position to initiate translation. Disruption of the SL by the Omr sRNAs or by the G3A5 mutation decreases translation initiation, possibly because of sliding of the 30S subunit on *fepA* mRNA. (E) Activation of *fepA* expression by the SL structure was determined after insertion of 9 nts upstream (mutant +9up) or downstream (mutant +9down) of the SL, as well as in the Δbulge variant. This activation factor, given above the red bars, is the ratio of the β-galactosidase activity of the wt SL fusion (grey bars) over the activity of the G3A5 derivative (red bars). Strains used in this experiment are from left to right: MG1772, JJ0268, JJ0271 and JJ0241 for the wt derivatives, and JJ0233, JJ0250, JJ0260 and JJ0290 for the G3A5 mutants.

The functional importance of the unpaired nts of the SL was then assayed by either deleting the bulged nts (mutant Δbulge) or using the Loop mutant (Fig. 5B). Neither of these mutations had a strong impact on *fepA* expression (Fig. 6B). Furthermore, formation of the SL still activated *fepA* expression in those two constructs as shown by the fact that the Loop mutant fusion was controlled by the Omr sRNAs similarly to the wt fusion (Fig. 5C) and that expression of the Δbulge construct was weakened by the G3A5 change to the same extent as the wt (Fig. 6E). Together, these data strongly suggest that the SL activates *fepA* expression *via* its secondary rather than primary structure, and that changes in the topology of this structure (e.g. by deleting the bulge) are permissible. Importantly, the G3A5, U5C3, Loop and Δbulge mutants show that all nts can be modified without interfering with the activating role of this structure, arguing strongly against a structural switch that would occur upon OmrA/B binding and that would ultimately affect RBS accessibility.

We nonetheless wanted to more directly assess whether binding of the Omr 5'-end would affect *fepA* RBS accessibility. For this purpose, we first repeated the probing experiment of *fepA* mRNA (Fig. 5A) in presence of an excess of OmrA or OmrY sRNA, but found that both sRNAs fail to bind to *fepA* mRNA in this experiment (Fig. S6). This is different from what we observed in other *in vitro* experiments, such as the toeprinting assays (see e.g. Fig. 3) and this is most likely due to the use of different buffers required for DMS- or CMCT-induced modifications. To assess RBS accessibility, we thus chose instead to follow the binding of the As-14.-2 AsDNA to *fepA* mRNA in an RT

experiment (Fig. 5D). Because As-14.-2 is perfectly complementary to nts -14 to -2 of *fepA*, which encompasses the SD sequence, its binding is expected to reflect the accessibility of the SD region. As a proof of principle, we first showed that this was indeed the case: while strong RT-stops indicated As-14.-2 binding to $fepA_{-26}$ or $fepA_{-173}$ transcripts, no stop was detected when we used instead a $fepA_{-26}$* transcript carrying an extra sequence at its 5'-end that can occlude the SD region. Binding of As-14.-2 was then analyzed on $fepA_{-26}$ or $fepA_{-173}$ transcripts that were beforehand denatured and renatured in presence of increasing concentrations of OmrY or OmrYmut sRNA. As visible on Fig. 5E, the specific binding of OmrY had no noticeable effect on As-14.-2 binding to *fepA* mRNA, providing yet another indication that control of *fepA* by OmrA/B does not rely on a change in RBS accessibility.

The stem-loop structure acts at the translation initiation level

Our *in vivo* results did not allow us to discriminate between an effect of the SL on *fepA* mRNA translation, stability, processing or yet another process. However, because disruption of this SL by OmrA/B sRNAs (or group 2 AsDNAs) inhibited 30S ribosomal subunit binding to *fepA*, it seemed likely that it acted at the translation initiation step. This was tested *in vitro* by performing a toeprinting assay using the wt $fepA_{-173}$ transcript or its variants G3A5, U5C3 or Comp2. The results were in remarkable agreement with what was observed *in vivo* with the fusions: disruption of the SL using the G3A5 or U5C3 mutations strongly inhibited the toeprint, while restoring the structure using the compensatory change partially recovered the toeprint signal (Fig. 6C). This demonstrates that the SL enhances the formation of the ternary translation initiation complex without requiring other actors than the *fepA* transcript, the 30S ribosomal subunit and the initiator tRNA.

How could this structure activate translation of *fepA* mRNA?

An intriguing question raised by our results is the mechanism by which the SL activates translation initiation at such an early step. One possibility is that it would do so by helping to recruit the 30S ribosomal subunit. This could be due for instance to an interaction between the SL and one of the 30S components (protein or RNA). Importantly however, the effect of this SL is independent of its nt sequence and is permissive to topological changes. This is hard to reconcile with a recruitment model at first sight; nonetheless ribosomal proteins tend to interact with the sugar-phosphate backbone on ribosomal RNA (Brodersen et al., 2002; Mangeol et al., 2011) and similar interactions could occur between proteins and the activating SL. Other elements of the *fepA* mRNA could furthermore be involved in the activating role of the SL and provide specificity.

Alternatively, one might envision that this structure activates translation initiation by providing a "starting block" to the 30S subunit (Fig. 6D). In this regard, it is striking that the 5' edge of the SL is located at position +19 of *fepA* ORF, exactly contiguous to the 3' edge of the RBS (Huttenhofer and Noller, 1994). It would thus be ideally located to block the 30S subunit at proper position for an efficient translation initiation. Results of Fig. 5E suggested that the *fepA* SD sequence is as accessible in the absence as in the presence of the SL. This is consistent with the affinity between the 30S ribosomal subunit and the *fepA* mRNA being similar regardless of the formation of the SL. If so, the increase in the toeprint signal observed in presence of the SL (Fig. 6C) could be due to an increased translation initiation rate rather than to a different binding affinity, possibly because the SL would block the 30S ribosomal subunit from sliding on the mRNA without establishing a translation initiation complex. A prediction of this "starting block" model is that increasing the distance between *fepA* start codon and the SL should prevent the ability of the latter to activate translation. This was tested by measuring the activation by the SL when 9 extra nts were introduced after position +18 (mutant +9up in Fig. 6E). While the SL activated *fepA* expression by almost 10-fold when located at nt +19, i.e. in the wt situation, activation was completely abolished when it was moved to position +28 (Fig. 6E). In contrast, the SL still activated *fepA* expression by 7.5-fold when 9 nts were introduced after nt +45, i.e. downstream of this structure (mutant +9down, Fig. 6E). Together, these results indicate that the distance between the start codon and the SL

is important for the ability of this structure to activate translation. Even though this does not necessarily rule out the recruitment model, this is in perfect agreement with the starting block hypothesis. Thus, at least in the *fepA* context, translation initiation is activated by the presence of a SL that may help the correct positioning of the 30S ribosomal subunit for efficient initiation.

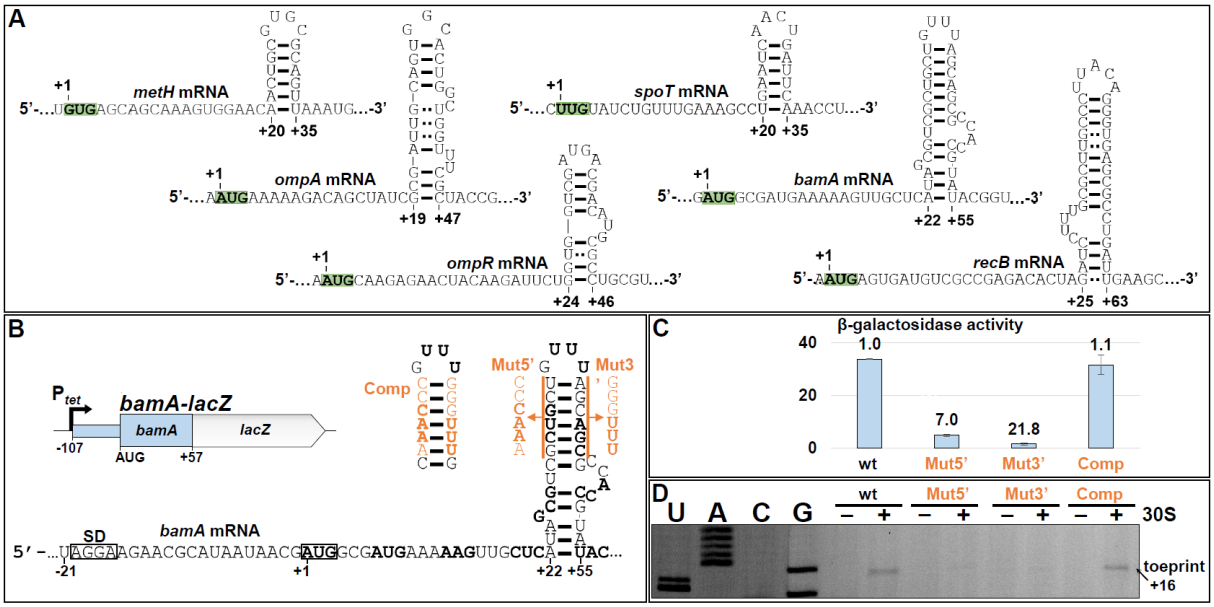


Figure 7. A similar activating stem-loop structure in the *bamA* mRNA

(A) Examples of SL predicted in a subset of *E. coli* mRNAs around the position +20 of the ORF. SD sequence and AUG start codon are indicated by green squares in each mRNA. Sequences between nt -25 and the SD are given in Fig. S11. (B) Scheme of the P_{tet}-*bamA-lacZ* fusion and of the mutations introduced in the SL to assay its role in *bamA* expression. (C) β-galactosidase activities of the different *bamA-lacZ* fusions. Strains used in this experiment are JJ0383 (wt), JJ0384 (Mut5'), JJ0385 (Mut3') and JJ0386 (Comp). (D) Toeprinting assay on wt and mutant *bamA* mRNA fragments. Only the portion of the gel with the toeprint signal is shown; full size gel is reproduced in Fig. S12.

A similar activating role for a stem-loop in *bamA* mRNA

Another obvious question raised by our results is whether translation activation by a SL is restricted to *fepA* or is in contrast a more general phenomenon. To address this, we searched for putative SL located around the nt +20 in a subset of diverse mRNAs by combining mfold prediction (Zuker, 2003) and visual inspection and found that such SL can be predicted in many mRNAs (a few examples are given in Fig. 7A). One of the most convincing SL is predicted within *bamA*, encoding the essential component of the Bam complex that allows the insertion of β-barrel integral outer membrane proteins in the outer membrane of Gram-negative bacteria (Ricci and Silhavy, 2012). Another striking feature of the *bamA* mRNA is that it possesses a putative AGGA SD sequence located 16 nts upstream of the AUG (a distance conserved in several bacteria, Fig. S13), i.e. a spacing that is not optimal for efficient initiation, and whether it is really involved in the initial binding of the ribosome can be questioned. We experimentally tested whether this predicted SL could have a role in *bamA* expression, both *in vivo* using compensatory changes introduced in a *bamA-lacZ* translational fusion and *in vitro* by a toeprinting assay (Fig. 7B-D). The results are in all points similar to those obtained with *fepA*. First, mutations Mut5' and Mut3' that are expected to disrupt the SL strongly decreased expression of the fusion, which was restored to its wt level in presence of the compensatory change (Comp) (Fig. 7C). These data support both the existence of the predicted SL and its positive role in *bamA* expression. Furthermore, formation of the translation initiation complex was detected by toeprinting for both the wt and the Comp version of *bamA*, but not for the Mut5' and Mut3' variants, showing that, as for *fepA*, the SL promotes translation initiation at a very early stage (Fig. 7D).

Importantly, while *fepA* is found only in enterobacteria, the *bamA* gene is present in much more phylogenetically distant bacteria and, despite important variations at the nt sequence level, a similar stem-loop can be predicted in *bamA* mRNA of other γ-proteobacteria such as *Vibrio* species (Fig. S13). This suggests that this novel mechanism of translation activation by a SL is conserved in a wide set of Gram-negative bacteria.

DISCUSSION

A role for mRNA secondary structures in activating translation

In bacteria, several *cis*-acting elements embedded in the mRNA are known to play important roles in translation, either at the sequence or the structure level. The most obvious example is the evolutionary conserved SD sequence but, in addition, several enhancers have been described such as (A)CA rich motifs (Sharma et al., 2007; Yang et al., 2014) or A rich regions in the mRNA (Brock et al., 2007) that improve binding of the 30S subunit.

Secondary structures in mRNA are known to impact translation as well because they represent thermodynamic and kinetic barriers that must be disrupted to allow the ribosome to bind to or move on mRNA. Secondary structures have been reported to negatively affect translation initiation by decreasing access of the ribosome to the RBS for instance (de Smit and van Duin, 1990) or by slowing down the elongation rate of the translating ribosome (e.g. Qu et al., 2011; Wen et al., 2008).

The present work shows that mRNA secondary structures located around position +20 can surprisingly promote formation of the ternary initiation complex and thus activate translation, possibly by helping to correctly position the 30S ribosomal subunit for optimal translation initiation. Why would some mRNAs require such a SL for maximal expression? In the case of *bamA*, it is tempting to speculate that this is related to the unusually long SD-AUG spacing. Regarding *fepA*, the involvement of such an activating SL in translation is more puzzling at first sight given the canonical AGGA SD sequence optimally spaced from the start codon (8 nts). This could suggest that *cis*-elements of this mRNA somehow destabilize its interaction with the 30S ribosomal subunit. In line with this, we found that, when introduced at nt +20 of a different mRNA, this SL no longer activated translation, indicating that its effect depends of specific features of the *fepA* mRNA (our unpublished data). Because the SL activates expression of the short *fepA-lacZ* fusion (Fig. 6), these specific features are most

likely located between nts -26 and +18 of *fepA* mRNA and it will be interesting to determine how they participate in *fepA* expression and its control by the SL.

Folding of the activating stem-loop structure and translation

The activating SL identified in *fepA* and *bamA* are part of ORF and are thus disrupted by translating ribosomes. Considering that disruption of the SL significantly decreases ribosome binding to those mRNAs (Fig. 6), we can assume that each translating ribosome will transiently decrease the initiation rate until the hairpin refolds. Given that the time needed to form a simple hairpin helix is less than one millisecond *in vitro* (Crothers et al., 1974), while initiation frequency for a well-translated gene like *lacZ* is on the second range (Kennell and Riezman, 1977), it is likely that the SL refolds after clearance of the translating ribosome and before binding of the next ribosome. Therefore, activation of ribosome binding by this SL can affect several rounds of translation, and might ensure that translation is kept within a given range.

Are activating stem-loops a general feature of mRNAs?

In addition to the *fepA* and *bamA* genes, it is likely that such activating SL are present in other genes given that they can be predicted around position +20 in yet more *E. coli* mRNAs (a few are shown in Fig. 7A). Also in line with the idea that our findings could apply to other mRNAs than *fepA* or *bamA*, it was previously reported that introduction of a seven bp SL at position +19 dramatically enhanced the translation of poorly expressed heterologous genes (Paulus et al., 2004). Furthermore, the analysis of *fepA* sequences from diverse enterobacteria and of *bamA* in several γ-proteobacteria indicates that these structures are likely to be conserved in several species, despite changes in the nt and possibly in the amino acid sequence (Fig. S13 and S14), suggesting that translation activation by SL structures occurs as well in these organisms.

Nonetheless, data from a recent genome wide study of RNA structures in *E. coli* did not specifically point out the existence of secondary structures near the position +20 of mRNAs (Del Campo et al., 2015). And, as mentioned above, the ability of the *fepA* SL to activate translation does not necessarily withstand a change in the surrounding mRNA sequence. Thus, translation activation by SL might be limited to a subset of mRNAs, whose identity remains to be deciphered.

The activating stem-loop as a target for regulators of fepA

We have identified at least two regulators that modulate *fepA* expression by targeting this activating SL. These regulators are the OmrA and OmrB sRNAs, whose binding-site on *fepA* mRNA corresponds to the 3' side of the SL and that accordingly repress FepA synthesis at the translation level. It is tempting to speculate that yet other post-transcriptional regulators, being sRNAs or proteins, could target this structure in *fepA* or *bamA*. In addition, these SL could also form thermosensors and integrate the temperature as an additional cue of FepA or BamA synthesis.

From a physiological standpoint, regulating *fepA* expression at the post-transcriptional level, by OmrA/B or other regulators, is expected to change FepA levels in response to different environmental cues and with different dynamic properties than the well-established transcriptional control by the Fur repressor. However, one can wonder what would be the advantage of limiting FepA synthesis by OmrA/B sRNAs that themselves respond to the EnvZ-OmpR TCS or to the RpoS sigma factor (at least for OmrA). A possible connection lies in the fact that EnvZ-OmpR is activated in response to acid stress (Quinn et al., 2014; Stincone et al., 2011), i.e. conditions that increase the solubility of iron. OmrA/B can thus participate in limiting not only the levels of FepA but also of other receptors of iron-siderophore complexes such as FecA or CirA under those conditions where iron scavenging by siderophores might be less needed, or even possibly harmful. Another important note is that both OmrA and OmrB are induced upon host infection (Westermann et al., 2016). Because FepA and other iron-dependent OMPs are recognized by the host immune system (Fernandez-Beros et al., 1989), decreasing their levels in response to the infectious process may

help to limit the host response to bacterial infection.

Together with previous reports, the study presented here highlights the wide diversity in the mRNA regions targeted by sRNAs and in the regulatory mechanisms by which sRNAs control gene expression. Importantly, it also led to the identification of translation activating SL that promote formation of the initiation ternary complex on several bacterial mRNAs, suggesting that the definition of RBS may need to be reevaluated for some transcripts.

Author Contributions

MG and JJ designed research; JJ performed experiments; MG, JJ and CC analyzed and interpreted the data, and wrote the paper.

Acknowledgements

We are grateful to Kathleen Postle and Knud Nierhaus for providing us the FepA or the S1 antibody, respectively, and to Eliane Hajnsdorf for the gift of purified Hfq. We thank Mathias Springer and Ciaran Condon for insightful comments on the manuscript. This work was supported by the CNRS, the ANR (grant ANR-14-CE10-0004-01 to MG) and by the "Initiative d'Excellence" program from the French State (Grant "DYNAMO", ANR-11-LABX-0011). JJ is a recipient of a DYNAMO Ph.D. fellowship.

STAR METHODS

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Maude Guillier (maude.guillier@ibpc.fr).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Bacterial Strains

Bacterial strains used in this study are listed in Table S1, and are derivatives of *E.coli* strain MG1655. Cells were grown in LB. When relevant, antibiotics were added at the following concentrations: kanamycin 25µg mL^{-1}, chloramphenicol 10 µg mL^{-1}, tetracycline 10 µg mL^{-1}, ampicillin 150 µg mL^{-1} in liquid cultures or 100 µg mL^{-1} on plates. The *rne*^{ts} mutant allele was obtained from E. Massé (Massé et al., 2003).

Construction of *lacZ* fusions was by recombineering following previously described λ red-based procedures (Coornaert et al., 2013). Briefly, DNA fragments corresponding to nts -173 or -26 to +45 of *fepA* mRNA (relative to start codon), or to nts -107 to +57 of *bamA*, were amplified by PCR using Phusion high fidelity DNA polymerase. The primers used introduced sequences homologous to the $P_{Ltet0-1}$ (P_{tet}) promoter upstream of *fepA* (or *bamA*) 5'-end, and to *lacZ* from nts +28 to at least +48, in frame with the coding sequence. When needed, primers introducing mutations were used instead, or in prior PCR reactions (see table S2). The final PCR products were recombined in the MG1508 strain, by replacing the *cat-sacB* cassette of a chromosomal P_{tet}*-cat-sacB-lacZ* construct. Recombinants were selected on LB plates without NaCl, supplemented with 6% sucrose to counterselect cells carrying *sacB*. Chloramphenicol-sensitive colonies were subsequently purified, and the final strains were checked by PCR and sequencing for the presence of the desired fusions.

METHOD DETAILS

Oligonucleotides and Plasmids

The oligonucleotides used in this study are listed in table S2. The plasmids are all derivatives of the pBRpLac plasmid (Guillier and Gottesman, 2006). For construction of pOmrY and pOmrYmut, DNA fragments were PCR-amplified from pBRplacRybB plamid (Mandin and Gottesman, 2010) using pBRrev2 and either AatIIOmrY or AatIIOmrYmut primers. After DpnI digestion of the template plasmids, the DNA fragments were subsequently cleaved with EcoRI and AatII and cloned into the corresponding sites on pBRplac. The recombinant plasmids were then transformed into DH5α cells, and sequenced.

pOmrAmut and pOmrBmut plasmids were constructed using the QuikChange II site-directed mutagenesis kit (Stratagene). pOmrA and pOmrB were used as templates in PCR with oligonucleotides introducing the desired mutation (see table S2).

β-galactosidase assays

Overnight cultures were diluted 500-fold in fresh LB (or LB-Amp-IPTG 100 □M for strains with plasmids) and grown at 37°C to mid-exponential phase. The β-galactosidase activity was then measured as previously described (Miller, 1992). Briefly, aliquots of 200µL (for *fepA* fusions) or 500µL (for *bamA* fusions) were mixed with 800µL or 500µL, respectively, of Z buffer and lysed with toluene. After toluene evaporation at 37°C for at least 1h, specific β-galactosidase activity was assayed with 200µL ONPG at 4mg/mL and reaction was stopped with 500µL Na_2CO_3 1M. The culture final OD at 600nm (cOD600), the culture volume in each sample (v), the reaction time (rt), and the reaction final OD at 420nm (rOD420) and 550nm (rOD550) were taken into account to calculate the specific β-galactosidase activity, following this formula: $1000 * \frac{rOD420-(1.75*rOD550)}{cOD600*rt*v}$. The results shown are the mean of at least three independent replicates, and bars represent confidence intervals.

RNA extraction and Northern Blot assays

Total RNA was extracted by the hot-phenol method (Guillier and Gottesman, 2006) either from the same cultures than those used for the β-galactosidase experiments or from cultures subjected to pulse-induction of FepA synthesis. In this latter case, overnight cultures were diluted 500-fold in fresh LB-Amp (Fig. 1D) or LB-Amp-IPTG 100 □M (Fig. 2) and grown at 37°C to mid-exponential phase, then shifted or not to 44°C for 15 minutes in presence of 250 □M dip (and IPTG for Fig. 1D); RNA was extracted after this treatment. RNA was then ethanol precipitated and equal amounts of total RNA were separated on denaturing acrylamide (for sRNAs; 3.5 µg RNA loaded) or agarose (for mRNAs; 5.5 µg RNA loaded) gels, transferred to Hybond-N+ membranes and specific RNAs were detected using specific biotinylated probes and the Ambion Brightstar detection kit.

Protein extraction and Western Blot assays

Protein samples were prepared from the same cultures as those used for RNA extraction. Samples preparation and western blotting were as described previously (Coornaert et al., 2013), except that detection was performed with the West-Femto kit (Thermo-Scientific). The anti-FepA and the anti-S1, kind gifts from Kathleen Postle (Higgs et al., 2002) and Knud Nierhaus, were used at 1:5000 and 1:10000 dilutions, respectively.

***In vitro* RNA transcription**

Templates for T7 RNA transcription were generated by PCR, using primers adding a T7 promoter upstream of the gene to be transcribed. Two G nts were added to *fepA* 5'-end to improve transcription efficiency; the 5'-end of the sRNAs was left unchanged. After purification, the PCR products were used to generate RNAs with the T7 Megascript kit (Ambion) following manufacturer's instructions. Transcripts were then phenol extracted, ethanol precipitated and finally purified with Nucleoseq columns (GE Healthcare).

***In vitro* reverse-transcription**

For those experiments, 1pmol of wt or mutant *fepA* mRNA transcripts (-173 to +129 or -26 to +129) were denatured at 80°C for 3 min in the presence of 2pmol of the Cyanine-5 labelled fepToeCy5 probe and either water or 1, 5 or 10pmol of OmrY or OmrYmut sRNA. Samples were then immediately immersed in solid CO2/ethanol for 1 min and subsequently thawed on ice. 1x final concentration of AMV-buffer (Finnzymes) was added to the reactions along with 3mM final concentration of dNTPs, and samples were subsequently incubated at 37°C for 10 min to allow for transcript-sRNA annealing and folding. 1pmol of the As-14.-2 AsDNA or water was then added to the reactions, which were incubated for another 10 min at 37°C to allow for transcript-AsDNA annealing and folding. Finally, reverse-transcription was performed with 1 unit of AMV reverse transcriptase (Finnzymes) at 37°C for 20 min, and stopped with a formamide/EDTA mixture. cDNAs were fractionated by PAGE on a 6% denaturing gel along with sequencing reactions, and visualized with a Typhoon fluorescent scanner set up for Cy5 detection.

***In vitro* toeprinting and RNA structure probing assays**

These assays were performed essentially as in (Coornaert et al., 2013), with minor modifications that are detailed thereafter. Briefly, 0.5pmol of wt or mutant *fepA* (-173 to

+129) or *bamA* (-107 to +100) transcripts were denatured at 80°C for 3 min in 10mM Tris-acetate pH7.4, 60mM NH_4Cl and 6mM β-mercaptoethanol, together with 2pmol of fepAToeCy5 or bamAToeCy5 DNA primer labeled with a Cyanine-5 at their 5'-end. When relevant, this step was performed in the presence of increasing concentrations of sRNAs or AsDNAs. Samples were then immediately immersed in solid CO2/ethanol for 1 min and subsequently thawed on ice. Magnesium was added at 10mM final concentration, and samples were incubated at 37°C for 10 min to allow for transcript-sRNA annealing and folding. Samples were then incubated at 37°C for 10 min in 0.015% triton, 3.6mM trisHCl pH7.5, 0.07mM EDTA pH8.0, 7% glycerol and 7mM NH_4Cl, or the same buffer supplemented with purified Hfq protein (a kind gift from Eliane Hajnsdorf). 5pmol of 30S ribosomal subunits were then added together with 1,5mM dNTPs and 15pmol $tRNA^{fmet}$, and the samples were incubated at 37°C for 10 min. Finally, reverse-transcription was performed by incubating the samples at 37°C for 20 min with 1 unit of AMV reverse transcriptase (Finnzymes) and stopped with a formamide/EDTA mixture. cDNAs were fractionated by PAGE on 6% denaturing gels along with sequencing reactions, and visualized with a Typhoon fluorescent scanner set up for Cy5 detection.

In vitro RNA structure probing

RNA structure probing assays were adapted from (Coornaert et al., 2013) as follows. Samples containing 3pmol of *fepA* (-26 to +129) transcript in water were heated at 80°C for 3 min, immediately immersed in solid CO2/ethanol for 1 min and thawed on ice. They were then incubated at 25°C for 10 min in a buffer containing 10mM magnesium acetate, 50mM ammonium chloride and either 50mM sodium cacodylate pH 7.5 for DMS treatment, or 50mM sodium borate pH 8.0 for CMCT treatment. 1µg of *L.lactis* 23S rRNA was added to the samples, along with 0.1 volume of either DMS (1:20 dilution in ethanol) or CMCT (100mg/mL in CMCT buffer). Probing reactions were carried out at 25°C for 5 min for DMS modification or 10 min for CMCT modification. RNAs were recovered by ethanol precipitation, resuspended in water and denatured along with 5pmol of the fepAToeCy5 primer mentioned above at 80°C for 3 min. The samples were then immediately immersed in solid CO2/ethanol for 1 min and thawed on ice. They were subsequently supplemented with 2.5mM dNTPs, 3mM DTT, 1X superscript III buffer, and 5.6 units of Superscript III reverse-transcriptase (Invitrogen). Reverse-transcription was performed at 50°C for 30min followed by incubation of the samples at 75°C for 15 min to inactivate the enzyme. After addition of a formamide /EDTA mixture, the cDNAs were analyzed as described above for the toeprinting assays.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis for β-galactosidase assays was performed as follows: for each fusion and condition, the mean of at least three independent replicates was considered as the final β-galactosidase specific activity. The 95% confidence intervals were then calculated from the standard deviation between replicates (σ), and the number of replicates (n) as follows: $(x - 1.96 * \frac{\sigma}{\sqrt{n}}, x + 1.96 * \frac{\sigma}{\sqrt{n}})$ were x is the mean value of the considered replicates.

Quantification of the cDNA bands produced in the toeprinting assay (Fig. 4.C) was carried out using the ImageJ software. For each sample, the total lane signal is the sum of the bands which correspond to the full length extension signal, the toeprint signal and the antisense-mRNA duplex signal. Each signal was then expressed as the ratio over the total lane signal.

REFERENCES

Balbontin, R., Villagra, N., Pardos de la Gandara, M., Mora, G., Figueroa-Bossi, N., and Bossi, L. (2016). Expression of IroN, the salmochelin siderophore receptor, requires mRNA activation by RyhB small RNA homologues. Mol. Microbiol. *100*, 139-155.

Bandyra, K.J., Said, N., Pfeiffer, V., Gorna, M.W., Vogel, J., and Luisi, B.F. (2012). The seed region of a small RNA drives the controlled destruction of the target mRNA by the endoribonuclease RNase E. Mol. Cell *47*, 943-953.

Bouvier, M., Sharma, C.M., Mika, F., Nierhaus, K.H., and Vogel, J. (2008). Small RNA binding to 5' mRNA coding region inhibits translational initiation. Mol. Cell *32*, 827-837.

Brock, J.E., Paz, R.L., Cottle, P., and Janssen, G.R. (2007). Naturally occurring adenines within mRNA coding sequences affect ribosome binding and expression in *Escherichia coli*. J. Bact. *189*, 501-510.

Brodersen, D.E., Clemons, W.M., Jr., Carter, A.P., Wimberly, B.T., and Ramakrishnan, V. (2002). Crystal structure of the 30 S ribosomal subunit from *Thermus thermophilus:* structure of the proteins and their interactions with 16 S RNA. J. Mol. Biol. *316*, 725-768.

Brosse, A., Korobeinikova, A., Gottesman, S., and Guillier, M. (2016). Unexpected properties of sRNA promoters allow feedback control via regulation of a two-component system. Nucleic acids research *44*, 9650-9666.

Coornaert, A., Chiaruttini, C., Springer, M., and Guillier, M. (2013). Post-transcriptional control of the *Escherichia coli* PhoQ-PhoP two-component system by multiple sRNAs involves a novel pairing region of GcvB. PLoS Genet. *9*, e1003156.

Crothers, D.M., Cole, P.E., Hilbers, C.W., and Shulman, R.G. (1974). The molecular mechanism of thermal unfolding of *Escherichia coli* formylmethionine transfer RNA. J. Mol. Biol. *87*, 63-88.

Darfeuille, F., Unoson, C., Vogel, J., and Wagner, E.G. (2007). An antisense RNA inhibits translation by competing with standby ribosomes. Mol. Cell *26*, 381-392.

De Lay, N., and Gottesman, S. (2012). A complex network of small non-coding RNAs regulate motility in *Escherichia coli*. Mol. Microbiol. *86*, 524-538.

de Smit, M.H., and van Duin, J. (1990). Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. Proc. Natl. Acad. Sci. USA *87*, 7668-7672.

Del Campo, C., Bartholomaus, A., Fedyunin, I., and Ignatova, Z. (2015). Secondary Structure across the Bacterial Transcriptome Reveals Versatile Roles in mRNA Regulation and Function. PLoS Genet. *11*, e1005613.

Fernandez-Beros, M.E., Gonzalez, C., McIntosh, M.A., and Cabello, F.C. (1989). Immune response to the iron-deprivation-induced proteins of *Salmonella typhi* in typhoid fever. Infect. Immun. *57*, 1271-1275.

Guillier, M., and Gottesman, S. (2006). Remodelling of the *Escherichia coli* outer membrane by two small regulatory RNAs. Mol. Microbiol. *59*, 231-247.

Guillier, M., and Gottesman, S. (2008). The 5' end of two redundant sRNAs is involved in the regulation of multiple targets, including their own regulator. Nucleic acids research *36*, 6781-6794.

Heidrich, N., Moll, I., and Brantl, S. (2007). *In vitro* analysis of the interaction between the small RNA SR1 and its primary target *ahrC* mRNA. Nucleic acids research *35*, 4331-4346.

Higgs, P.I., Larsen, R.A., and Postle, K. (2002). Quantification of known components of the *Escherichia coli* TonB energy transduction system: TonB, ExbB, ExbD and FepA. Mol. Microbiol. *44*, 271-281.

Holmqvist, E., Reimegard, J., Sterk, M., Grantcharova, N., Romling, U., and Wagner, E.G. (2010). Two antisense RNAs target the transcriptional regulator CsgD to inhibit curli synthesis. EMBO J. *29*, 1840-1850.

Huttenhofer, A., and Noller, H.F. (1994). Footprinting mRNA-ribosome complexes with chemical probes. EMBO J. *13*, 3892-3901.

Jagodnik, J., Brosse, A., Le Lam, T.N., Chiaruttini, C., and Guillier, M. (2017). Mechanistic study of base-pairing small regulatory RNAs in bacteria. Methods *117*, 67-76.

Kawamoto, H., Koide, Y., Morita, T., and Aiba, H. (2006). Base-pairing requirement for RNA silencing by a bacterial small RNA and acceleration of duplex formation by Hfq. Mol. Microbiol. *61*, 1013-1022.

Kennell, D., and Riezman, H. (1977). Transcription and translation initiation frequencies of the *Escherichia coli lac* operon. J. Mol. Biol. *114*, 1-21.

Kery, M.B., Feldman, M., Livny, J., and Tjaden, B. (2014). TargetRNA2: identifying targets of small regulatory RNAs in bacteria. Nucleic acids research *42*, W124-129.

Levi-Meyrueis, C., Monteil, V., Sismeiro, O., Dillies, M.A., Monot, M., Jagla, B., Coppee, J.Y., Dupuy, B., and Norel, F. (2014). Expanding the RpoS/sigmaS-network by RNA sequencing and identification of sigmaS-controlled small RNAs in *Salmonella*. PloS one *9*, e96918.

Mandin, P., and Gottesman, S. (2010). Integrating anaerobic/aerobic sensing and the general stress response through the ArcZ small RNA. EMBO J. *29*, 3094-3107.

Mangeol, P., Bizebard, T., Chiaruttini, C., Dreyfus, M., Springer, M., and Bockelmann, U. (2011). Probing ribosomal protein-RNA interactions with an external force. Proc. Natl. Acad. Sci. USA *108*, 18272-18276.

Massé, E., Escorcia, F.E., and Gottesman, S. (2003). Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*. Genes & development *17*, 2374-2383.

Miller, J.H. (1992). A Short Course in Bacterial Genetics (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Obana, N., Shirahama, Y., Abe, K., and Nakamura, K. (2010). Stabilization of *Clostridium perfringens* collagenase mRNA by VR-RNA-dependent cleavage in 5' leader sequence. Mol. Microbiol. *77*, 1416-1428.

Papenfort, K., Sun, Y., Miyakoshi, M., Vanderpool, C.K., and Vogel, J. (2013). Small RNA-mediated activation of sugar phosphatase mRNA regulates glucose homeostasis. Cell *153*, 426-437.

Paulus, M., Haslbeck, M., and Watzele, M. (2004). RNA stem-loop enhanced expression of previously non-expressible genes. Nucleic acids research *32*, e78.

Peano, C., Wolf, J., Demol, J., Rossi, E., Petiti, L., De Bellis, G., Geiselmann, J., Egli, T., Lacour, S., and Landini, P. (2015). Characterization of the Escherichia coli sigma(S) core regulon by Chromatin Immunoprecipitation-sequencing (ChIP-seq) analysis. Scientific reports *5*, 10469.

Pfeiffer, V., Papenfort, K., Lucchini, S., Hinton, J.C., and Vogel, J. (2009). Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. Nat Struct Mol Biol *16*, 840-846.

Qu, X., Wen, J.D., Lancaster, L., Noller, H.F., Bustamante, C., and Tinoco, I., Jr. (2011). The ribosome uses two active mechanisms to unwind messenger RNA during translation. Nature *475*, 118-121.

Quinn, H.J., Cameron, A.D., and Dorman, C.J. (2014). Bacterial regulon evolution: distinct responses and roles for the identical OmpR proteins of *Salmonella Typhimurium* and *Escherichia coli* in the acid stress response. PLoS Genet. *10*, e1004215.

Ramirez-Pena, E., Trevino, J., Liu, Z., Perez, N., and Sumby, P. (2010). The group A Streptococcus small regulatory RNA FasX enhances streptokinase activity by increasing the stability of the *ska* mRNA transcript. Mol. Microbiol. *78*, 1332-1347.

Ricci, D.P., and Silhavy, T.J. (2012). The Bam machine: a molecular cooper. Biochimica et biophysica acta *1818*, 1067-1084.

Salvail, H., Caron, M.P., Belanger, J., and Masse, E. (2013). Antagonistic functions between the RNA chaperone Hfq and an sRNA regulate sensitivity to the antibiotic colicin. EMBO J. *32*, 2764-2778.

Sharma, C.M., Darfeuille, F., Plantinga, T.H., and Vogel, J. (2007). A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. Genes Dev. *21*, 2804-2817.

Stincone, A., Daudi, N., Rahman, A.S., Antczak, P., Henderson, I., Cole, J., Johnson, M.D., Lund, P., and Falciani, F. (2011). A systems biology approach sheds new light on *Escherichia coli* acid resistance. Nucleic acids research *39*, 7512-7528.

Vogel, J., and Luisi, B.F. (2011). Hfq and its constellation of RNA. Nat. Rev. Microbiol. *9*, 578-589.

Wagner, E.G., and Romby, P. (2015). Small RNAs in bacteria and archaea: who they are, what they do, and how they do it. Adv. Genet. *90*, 133-208.

Wen, J.D., Lancaster, L., Hodges, C., Zeri, A.C., Yoshimura, S.H., Noller, H.F., Bustamante, C., and Tinoco, I. (2008). Following translation by single ribosomes one codon at a time. Nature *452*, 598-603.

Westermann, A.J., Forstner, K.U., Amman, F., Barquist, L., Chao, Y., Schulte, L.N., Muller, L., Reinhardt, R., Stadler, P.F., and Vogel, J. (2016). Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions. Nature *529*, 496-501.

Wright, P.R., Georg, J., Mann, M., Sorescu, D.A., Richter, A.S., Lott, S., Kleinkauf, R., Hess, W.R., and Backofen, R. (2014). CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains. Nucleic acids research *42*, W119-123.

Yang, Q., Figueroa-Bossi, N., and Bossi, L. (2014). Translation enhancing ACA motifs and their silencing by a bacterial small regulatory RNA. PLoS Genet. *10*, e1004026.

Yusupova, G.Z., Yusupov, M.M., Cate, J.H., and Noller, H.F. (2001). The path of messenger RNA through the ribosome. Cell *106*, 233-241.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic acids research *31*, 3406-3415.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-FepA	K. Postle (Higgs et al. 2002)	N/A
Rabbit polyclonal anti-S1	K. Nierhaus	N/A
Bacterial and Virus Strains		
Escherichia coli K-12 MG1655 strain	F. Blattner	RefSeq NC000913.3
Strains used in this study all derive from MG1655 and are listed in Table S1	This study or references in Table S1	N/A
Biological Samples		
N/A		
Chemicals, Peptides, and Recombinant Proteins		
DMS	Sigma-Aldrich	Cat#D186309
CMCT	Sigma-Aldrich	Cat#C106402
Critical Commercial Assays		
N/A		
Deposited Data		
N/A		
Experimental Models: Cell Lines		
N/A		
Experimental Models: Organisms/Strains		
N/A		
Oligonucleotides		
All oligonucleotides used in this study are listed in Table S2	This study	N/A
Recombinant DNA		
Plasmids pBRplac, pBRplacOmrA, pBRplacOmrB	(Guillier and Gottesman, 2006)	N/A
Plasmids pBRplacOmrAmut, pBRplacOmrBmut, pBRplacOmrY and pBRplacOmrYmut	This study	N/A
Software and Algorithms		
MFold	(Zuker, 2003)	http://unafold.rna.albany.edu/?q=mfold/RNA-Folding-Form
ImageJ	ImageJ freeware	https://imagej.nih.gov/ij/
Other		
N/A		

SUPPLEMENTARY MATERIAL

This section includes:
-Table S1: strains and plasmids used in this study
-Table S2: oligonucleotides used in this study
-Legends to supplementary figures 1 to 14

Supplementary Material & Methods

Table S1. Strains and plasmids used in this study

Strain	Description	Construction/Source
MG1655	Wild-type strain	F. Blattner
DJ480	MG1655 Δ*lacX174*	D. Jin (NIH)
DJ624	MG1655 Δ*lacX174 mal::lacI^{q}*	D. Jin (NIH)
EM1277	DJ480 *rne-3071 zce-726::Tn10*	(Massé et al., 2003)
EM1279	DJ480 *zce-726::Tn10*	(Massé et al., 2003)
MG1003	DJ480 Δ*omrAB::kan*	(Guillier and Gottesman, 2006)
MG1099	DJ624 Δ*omrAB::kan*	(Guillier and Gottesman, 2006)
MG1325	MG1099 *zce-726::Tn10*	MG1099+P1(EM1279)
MG1326	MG1099 *rne-3071 zce-726::Tn10*	MG1099+P1(EM1277)
MG1508	MG1655 *mal::lacI^{q}* P$_{tet}$-*cat-sacB-lacZ$_{+28}$, mini-λ tet^{R}*	(Coornaert et al., 2013)
MG1772	MG1508 P$_{tet}$-*fepA$_{-173+45}$-lacZ$_{+28}$*	This study
MG2246	MG1508 P$_{tet}$-*fepA$_{-173+45}$antiBulge-lacZ$_{+28}$*	This study
MG2247	MG2246 Δ*omrAB::kan*	MG2246+P1(MG1003)
JJ0015	MG1772 Δ*omrAB::kan*	MG1772+P1(MG1003)
JJ0018	MG1508 P$_{tet}$-*fepA$_{-173+45}$-mut-lacZ$_{+28}$*	This study
JJ0019	JJ0018 Δ*omrAB::kan*	JJ0018+P1(MG1003)
JJ0120	MG1508 P$_{tet}$-*fepA$_{-26+45}$-A5-lacZ$_{+28}$*	This study
JJ0135	MG1508 P$_{tet}$-*fepA$_{-26+45}$-lacZ$_{+28}$*	This study
JJ0150	MG1508 P$_{tet}$-*fepA$_{-173+45}$-A5-lacZ$_{+28}$*	This study
JJ0180	MG1508 P$_{tet}$-*fepA$_{-28+45}$-comp1-lacZ$_{+28}$*	This study
JJ0190	MG1508 P$_{tet}$-*fepA$_{-173+45}$-comp1-lacZ$_{+28}$*	This study
JJ0192	MG1508 P$_{tet}$-*fepA$_{-173+45}$-loop-lacZ$_{+28}$*	This study
JJ0193	MG1508 P$_{tet}$-*fepA$_{-26+45}$-loop-lacZ$_{+28}$*	This study
JJ0225	MG1508 P$_{tet}$-*fepA$_{-173+45}$-U5-lacZ$_{+28}$*	This study
JJ0227	MG1508 P$_{tet}$-*fepA$_{-26+45}$-U5-lacZ$_{+28}$*	This study
JJ0229	MG1508 P$_{tet1}$-*fepA$_{-173+45}$-U5C3-lacZ$_{+28}$*	This study
JJ0231	MG1508 P$_{tet}$-*fepA$_{-26+45}$-U5C3-lacZ$_{+28}$*	This study
JJ0233	MG1508 P$_{tet}$-*fepA$_{-173+45}$-G3A5-lacZ$_{+28}$*	This study
JJ0235	MG1508 P$_{tet}$-*fepA$_{-26+45}$-G3A5-lacZ$_{+28}$*	This study
JJ0237	MG1508 P$_{tet}$-*fepA$_{-26+45}$-comp2-lacZ$_{+28}$*	This study
JJ0238	MG1508 P$_{tet}$-*fepA$_{-173+45}$-comp2-lacZ$_{+28}$*	This study
JJ0241	MG1508 P$_{tet}$-*fepA$_{-173+45}$-Δbulge-lacZ$_{+28}$*	This study
JJ0243	MG1508 P$_{tet}$-*fepA$_{-26+45}$-Δbulge-lacZ$_{+28}$*	This study
JJ0250	MG1508 P$_{tet}$-*fepA$_{-173+45}$- G3A5+9up-lacZ$_{+28}$*	This study
JJ0260	MG1508 P$_{tet}$-*fepA$_{-173+45}$- G3A5+9down-lacZ$_{+28}$*	This study
JJ0268	MG1508 P$_{tet}$-*fepA$_{-173+45}$- +9up-lacZ$_{+28}$*	This study
JJ0271	MG1508 P$_{tet}$-*fepA$_{-173+45}$- +9down-lacZ$_{+28}$*	This study
JJ0281	JJ0192 Δ*omrAB::kan*	JJ0192+P1(MG1003)

JJ0282	JJ0241 Δ*omrAB::kan*	JJ0241+P1(MG1003)
JJ0286	JJ0233 Δ*omrAB::kan*	JJ0233+P1(MG1003)
JJ0287	JJ0229 Δ*omrAB::kan*	JJ0229+P1(MG1003)
JJ0288	JJ0237 Δ*omrAB::kan*	JJ0237+P1(MG1003)
JJ0290	MG1508 P_{tet}-*fepA*$_{-173+45}$-*G3A5*-Δ*bulge*-*lacZ*$_{+28}$	This study
JJ0383	MG1508 P_{tet}-*bamA*$_{-107+57}$-*lacZ*$_{+28}$	This study
JJ0384	MG1508 P_{tet}-*bamA*$_{-107+57}$-*Mut5'*-*lacZ*$_{+28}$	This study
JJ0385	MG1508 P_{tet}-*bamA*$_{-107+57}$-*Mut3'*-*lacZ*$_{+28}$	This study
JJ0386	MG1508 P_{tet}-*bamA*$_{-107+57}$-*Comp*-*lacZ*$_{+28}$	This study
plasmid (short name)		**Construction**
pBRplac (vector)	pBR322 derivative carrying a modified $P_{LlacO-1}$ promoter, *amp*^{R}, *tet*^{R}	(Guillier and Gottesman, 2006)
pBRplacOmrA (pOmrA)	*omrA* under modified $P_{LlacO-1}$ in pBRplac, *amp*^{R}, *tet*^{R}	(Guillier and Gottesman, 2006)
pBRplacOmrB (pOmrB)	*omrB* under modified $P_{LlacO-1}$ in pBRplac, *amp*^{R}, *tet*^{R}	(Guillier and Gottesman, 2006)
pBRplacOmrAmut (pOmrAmut)	*omrAmut* under modified $P_{LlacO-1}$ in pBRplac, *amp*^{R}, *tet*^{R}	This Study
pBRplacOmrBmut (pOmrBmut)	*omrBmut* under modified $P_{LlacO-1}$ in pBRplac, *amp*^{R}, *tet*^{R}	This Study
pBRplacOmrY (pOmrY)	*omrY* under modified $P_{LlacO-1}$ in pBRplac, *amp*^{R}, *tet*^{R}	This Study
pBRplacOmrYmut (pOmrYmut)	*omrYmut* under modified $P_{LlacO-1}$ in pBRplac, *amp*^{R}, *tet*^{R}	This Study

Table S2. Oligonucleotides used in this study

Underlined nucleotides indicate mutations, sequences from promoters (PT7 or Ptet) or *lacZ* are in italics or bold, respectively. The asterisks represent the deleted nts in the Δbulge construct.

Primer	5'-3' sequence	Use
Plasmids construction		
pBRomrAmutfor	TAACAAGATACTGACGTCCCGTTTGGTATTGATTGGTGAGATTA	Construction pOmrAmut
pBRomrAmutrev	TAATCTCACCAATCAATACCAAACGGGACGTCAGTATCTTGTTA	
pBRomrBmutfor	TAACAAGATACTGACGTCCCGTTTGGTATTGATAGGTGAAGTCA	Construction pOmrBmut
pBRomrBmutrev	TGACTTCACCTATCAATACCAAACGGGACGTCAGTATCTTGTTA	
AatIIOmrY	GCTAGACGTCCCCAGAGGTATTGATGTCCCCATTTTGTGGAGCCC	Construction pOmrY/pOmrYmut
AatOmrYmut	GCTAGACGTCCCGTTTGGTATTGATGTCCCCATTTTGTGGAGCCC	
pBR-rev2	GCATTGTTAGATTTCATACACGG	
Strains construction		
5'Ptet-fepA	*GATAGAGATTGACATCCCTATCAGTGATAGAGATACTGAGCAC*CTATTTGCATTTGCAATAGCG	Forward primer for construction of P_{tet}-fepA$_{-173}$ fusions
5'PtetfepA-26	*GATAGAGATTGACATCCCTATCAGTGATAGAGATACTGAGCAC*CGTGGCAAAAATGCAGGAATAAAAC	Forward primer for construction of P_{tet}-fepA$_{-26}$ fusions
3'fepAl lac	**TAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGAC**CCCCAGATTGACCAACAAGG	Reverse primers for construction of fepA$_{+45}$-*lacZ* fusions
3'fepA-lacZ	**GGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGAC**CCCCAGATTGACCAACAA	
3'fepAl lac mut4	**TAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGAC**CCCGTTTTTGACCAACAAGGCCAGGGAATG	Reverse primer for construction of fepA$_{+45}$-mut-*lacZ* fusions
3'fepAΔbulgelacZ	**TAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGAC**CCC******GACCAACAAGGCCAGGG	Reverse primer for construction of fepA$_{+45}$-Δbulge-*lacZ* fusions
3'lacfepABulG3A5	**GGGTTTTCCCAGTCACGACGTTGTAAAACGAC**CCC******GACCAACAATTTTTCCCAATGAATCTTCTTG	Reverse primer for construction of fepA$_{+45}$-Δbulge-G3A5-*lacZ* fusions
3'fepA9postHP	**TAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGAC**ACTTACAGTCCCCAGATTGACCAACAAGGCCAGGG	Reverse primer for construction of fepA$_{+45}$-+9post-*lacZ* fusions
3'fepA9postG3A5	**TAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGAC**ACTTACAGTCCCCAGATTGACCAACAATTTTCCC	Reverse primer for construction of fepA$_{+45}$-+9post-G3A5-*lacZ* fusions

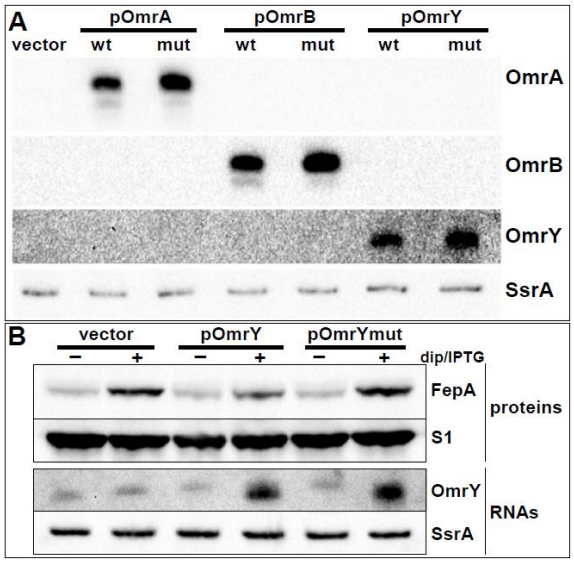
5'fepALoopMutLacZ	CCCTGGCCAACAAGGTCAATCTGGGG**GTCGTTTTACAACG**	Primers for construction of various mutant forms of the P_{tet}-*fepA-lacZ* fusions (used in multiple steps PCR reactions)
3'fepALoopMutrev	CCAGATTGACCTTGTTGGCCAGGGAATGAATCTTCTTGTTC	
5'fepAHPMutA5	GATTCATTCCCAAAAAATTGTTGGTCAATCTGGGG**GTCGTTTTACAACG**	
3'fepAHPMutA5	CCCCAGATTGACCAACAATTTTTGGGAATGAATCTTCTTGTTCATTG	
5'fepAHPComp	TCCCAAAAATTGTTTTTTAATCTTGGG**GTCGTTTTACAACGTCG**	
3'fepAHPComp	CCCAAGATTAAAAAACAATTTTTGGGAATGAATCTTCTTGTTC	
5'fepAHPU5	TCCCTGGCCTTGTTTTTTAATCTTGGG**GTCGTTTTACAACGTCG**	
3'fepAHPU5	CCCAAGATTAAAAAACAAGGCCAGGGAATGAATCTTCTTGTTC	
5'fepAHPComp2	TGGGAAAAATTGTTTTTTAATCTTCCC**GTCGTTTTACAACGTCG**	
3'fepAHPComp2	GGGAAGATTAAAAAACAATTTTTCCCAATGAATCTTCTTGTTC	
5'fepAHPG3A5	GATTCATTGGGAAAAATTGTTGGTCAATCTGGGG**GTCGTTTTACAACG**	
3'fepAHPG3A5	CCCCAGATTGACCAACAATTTTTCCCAATGAATCTTCTTGTTCATTG	
5'fepAHPU5C3	TCCCTGGCCTTGTTTTTTAATCTTCCC**GTCGTTTTACAACGTCG**	
3'fepAHPU5C3	GGGAAGATTAAAAAACAAGGCCAGGGAATGAATCTTCTTGTTC	
5'fepA9PreHP	GATTCATACTTACAGTTCCCTGGCCTTGTTGGTCAATCTG	
3'fepA9PreHP	GGCCAGGGAACTGTAAGTATGAATCTTCTTGTTCATTG	
5'fepA9preG3A5	ATACTTACAGTTGGGAAAAATTGTTGGTCAATCTG	
3'fepA9preG3A5	TTTTTCCCAACTGTAAGTATGAATCTTCTTGTTCATTG	
3'fepAantiBulge-rev	CCCAGATTGACCAACAAGGCCAATCCTAGGGAATGAATCTTCTTGTTCAT	
5'Ptet.bamA-107	*GATAGAGATTGACATCCCTATCAGTGATAGAGATACTGAGCAC*TTTTGTTATCGCATTGGCTCG	Forward primer for construction of P_{tet}-*bamA*$_{-107+57}$-*lacZ* fusion
3'lacZbamA+57	**TAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGAC**GTATACGGTGGCGCTGCTAAACAGC	Reverse primer for construction of P_{tet}-*bamA*$_{-107+57}$-*lacZ* wt fusion
3'lacZbamA-A3C3	**CCCAGTCACGACGTTGTAAAACGAC**GTATACGGTGGCGCTGCTAAACGGGTTTGACGCTATGAGCAACTTTTTCA	Reverse primer for construction of P_{tet}-*bamA*$_{-107+57}$-*lacZ* Mut5' fusion
3'lacZbamA-G3U3	**CCCAGTCACGACGTTGTAAAACGAC**GTATACGGTGGCAAACCCAAACAGCAGCGACGCTATGAGC	Reverse primer for construction of P_{tet}-*bamA*$_{-107+57}$-*lacZ* Mut3' fusion
3'lacZbamA-Comp	**CCCAGTCACGACGTTGTAAAACGAC**GTATACGGTGGCAAACCCAAACGGGTTTGACGCTATGAGCAACTTTTTC	Reverse primer for construction of P_{tet}-*bamA*$_{-107+57}$-*lacZ* Comp fusion

Jagodnik Jonathan – Thèse de doctorat - 2017

Preparation of T7 templates		
pT7OmrA	*TAATACGACTCACTATA*CCCAGAGGTATTGATTGGTG	Construction of template for *omrA/omrAmut* transcription
pT7OmrAm5	*TAATACGACTCACTATA*CCGTTTGGTATTGATTGGTG	
seqOmrA68rev+3	GCGAAAAAAAACCTGCGCATCCGC	
pT7OmrB	*TAATACGACTCACTATA*CCCAGAGGTATTGATAGGTG	Construction of template for *omrB/omrBmut* transcription
pT7OmrBm5	*TAATACGACTCACTATA*CCGTTTGGTATTGATAGGTG	
seqOmrBrev+3	GCGAAAAAAAACCTGCGCATCTGC	
5'T7OmrY12	*TAATACGACTCACTATA*CCCAGAGGTATTGATGTCCCC	Construction of template for *omrY/omrYmut* transcription
5'T7OmrY12mut5	*TAATACGACTCACTATA*CCGTTTGGTATTGATGTCCCC	
3'OmrY12	GATAACAAAAAACCCATCAACC	
pT7fepA	*TAATACGACTCACTATA*GGCTATTTGCATTTGCAATAGCG	Construction of template for fepA$_{-173+129}$/fepA$_{-26+129}$ transcription
5'pT7fepA-26G	*TAATACGACTCACTATA*GCGTGGCAAAAATGCAGGAAT	
5'pT7fepA-26*	*TAATACGACTCACTATA*GGAAGTTTTATTCCTGCCGTGGCAAAAATGCAGGAA	
seqfepA99rev	CTGCTCGGCGGCGGTAACGACAATAGTATCG	
5'fepAHPcomp2iv	GGGAAAAATTGTTTTTTAATCTTCCCATTTATGGGGTAGCGCAGG	Construction of templates for mutants of fepA$_{-173+129}$ (in multiple steps PCR, in conjunction with 3'fepAHPcomp2, 3'fepAHPU5C3 or 3'fepAHPG3A5 primers)
5'fepAHpU5C3iv	GCCTTGTTTTTTAATCTTCCCATTTATGGGGTAGCGCAGGC	
5'fepAHpG3A5iv	CAAGAAGATTCATTGGGAAAAATTGTTGGTCAATCTGGGG	
pT7bamA-107	*TAATACGACTCACTATA*GGTTTTGTTATCGCATTGGCTCG	Construction of template for bamA$_{-107+100}$ transcription
bamA+100rev	GGCCTTCGAAATGAATATCTTTCAC	
5'bamA-A3C3	TAGCGTCAAACCCGTTTAGCAGCGCCACCGTAT	Construction of templates for mutants of bamA$_{-107+100}$ (in multiple steps PCR, in conjunction with pT7bamA-107 and bamA+100rev primers)
3'bamA-A3C3	GCTAAACGGGTTTGACGCTATGAGCAACTTTTTC	
5'bamA-G3U3	GCTGTTTGGGTTTGCCACCGTATACGGTGCTG	
3'bamA-G3U3	CGGTGGCAAACCCAAACAGCAGCGACGCTATGAG	
5'bamA-Comp	TAGCGTCAAACCCGTTTGGGTTTGCCACCGTATACGGTGCTG	
3'bamA-Comp	GTGGCAAACCCAAACGGGTTTGACGCTATGAGCAACTTTTTC	
AsDNAs used in RT or toeprinting assay		
asfepA-14.-2	GTTTTATTCCTGC	
asfepA-1.16	GAATCTTCTTGTTCATC	
asfepA2.18	ATGAATCTTCTTGTTCA	
asfepA4.20	GAATGAATCTTCTTGTT	
asfepA6.22	GGGAATGAATCTTCTTG	
asfepA8.23	AGGGAATGAATCTTCT	
asfepA10.25	CCAGGGAATGAATCTT	

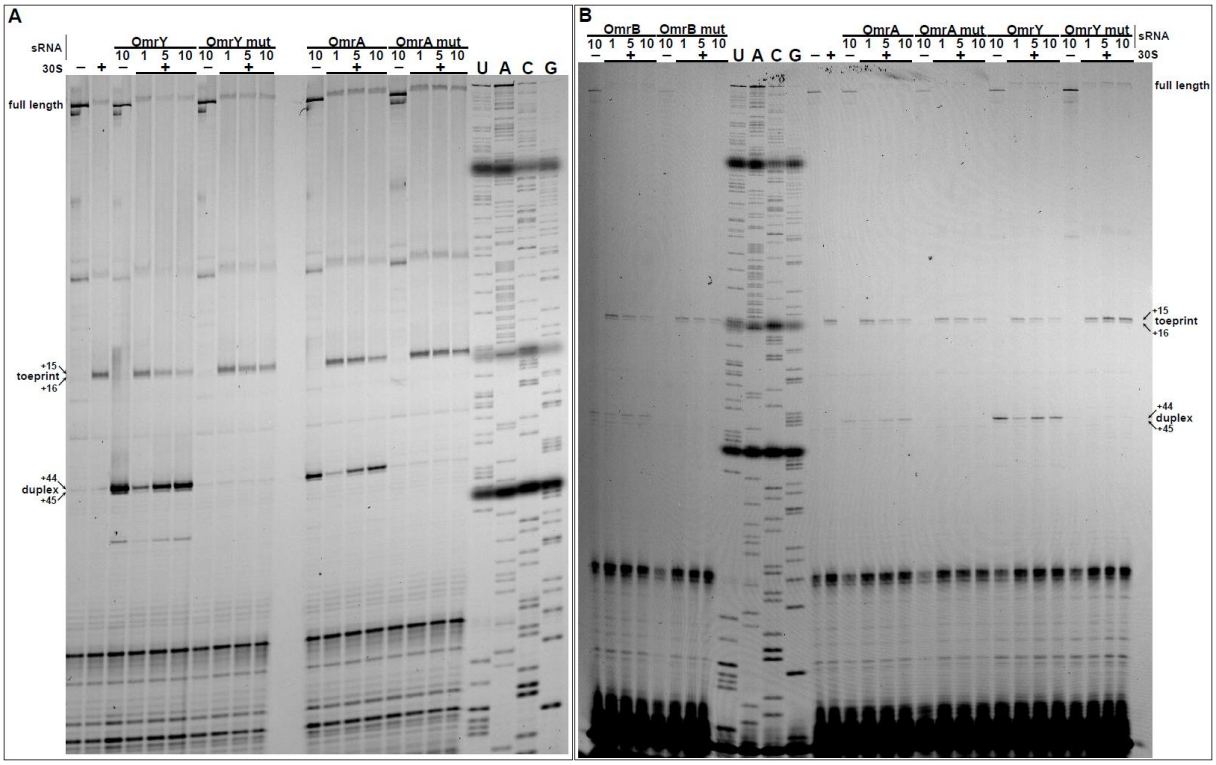
Jagodnik Jonathan – Thèse de doctorat - 2017

Supplementary figures



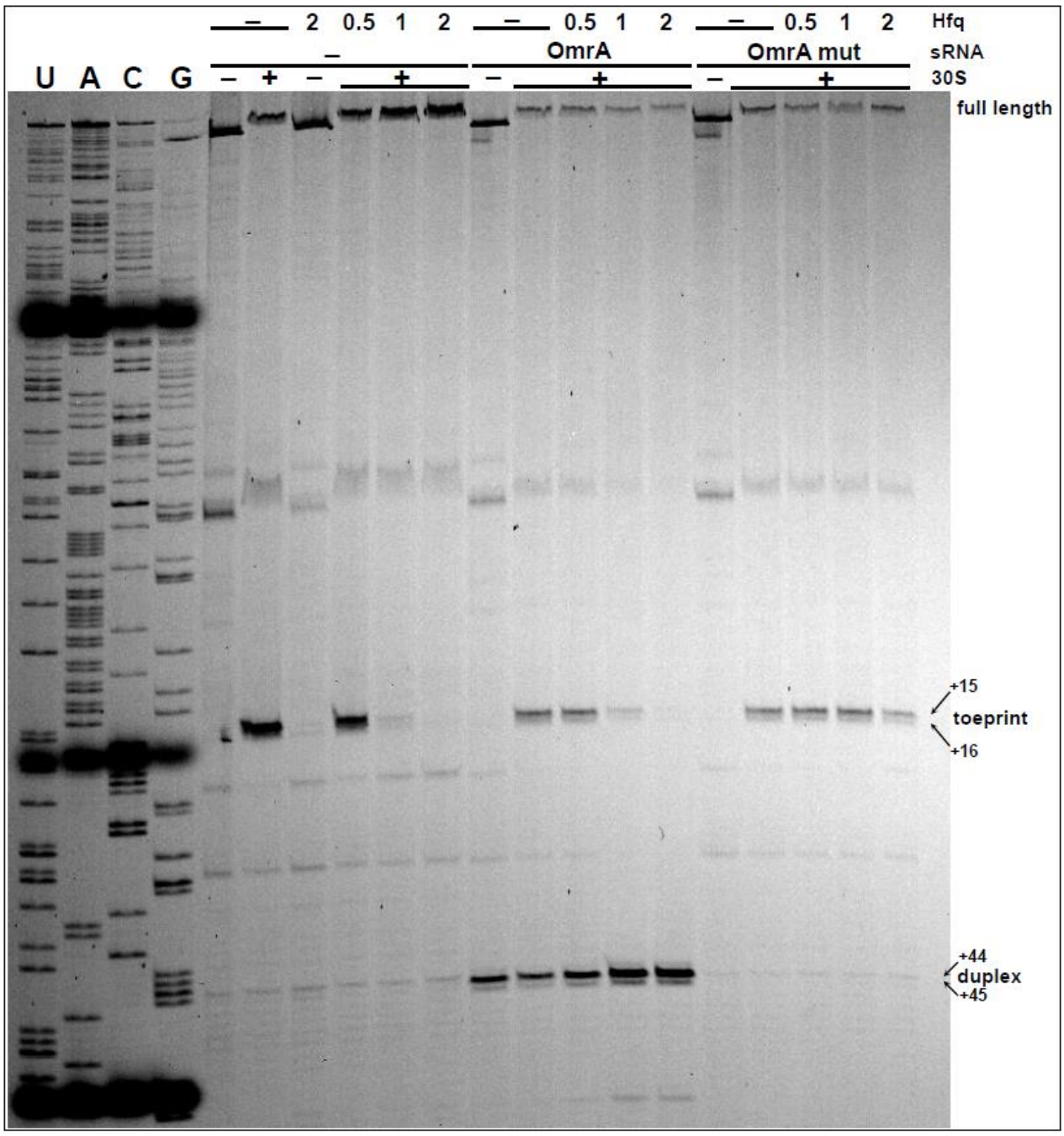
Supplementary figure 1. The mutant forms of OmrA, OmrB or OmrY accumulate at similar levels than their wt counterparts and are functional.

(A) Northern Blot analysis of OmrA, OmrB and OmrY sRNAs, wt or mutant, when overexpressed from pBRplac derivatives in strain JJ0015 (Δ*omrAB::kan*). Total RNA was extracted from cells grown in LB-Amp-IPTG 100 □M in mid-exponential phase (OD_{600} about 0.4). OmrY was detected using a biotinylated probe complementary to RybB nts 1-37, which is thus only partially complementary to OmrY; note that RybB is not detected under these conditions. Detection of SsrA is used as a loading control. (B) OmrY specifically controls FepA synthesis. Western- and northern-blot analysis of FepA protein and OmrY(mut) sRNA levels was performed in the same experiment as Fig. 1D.

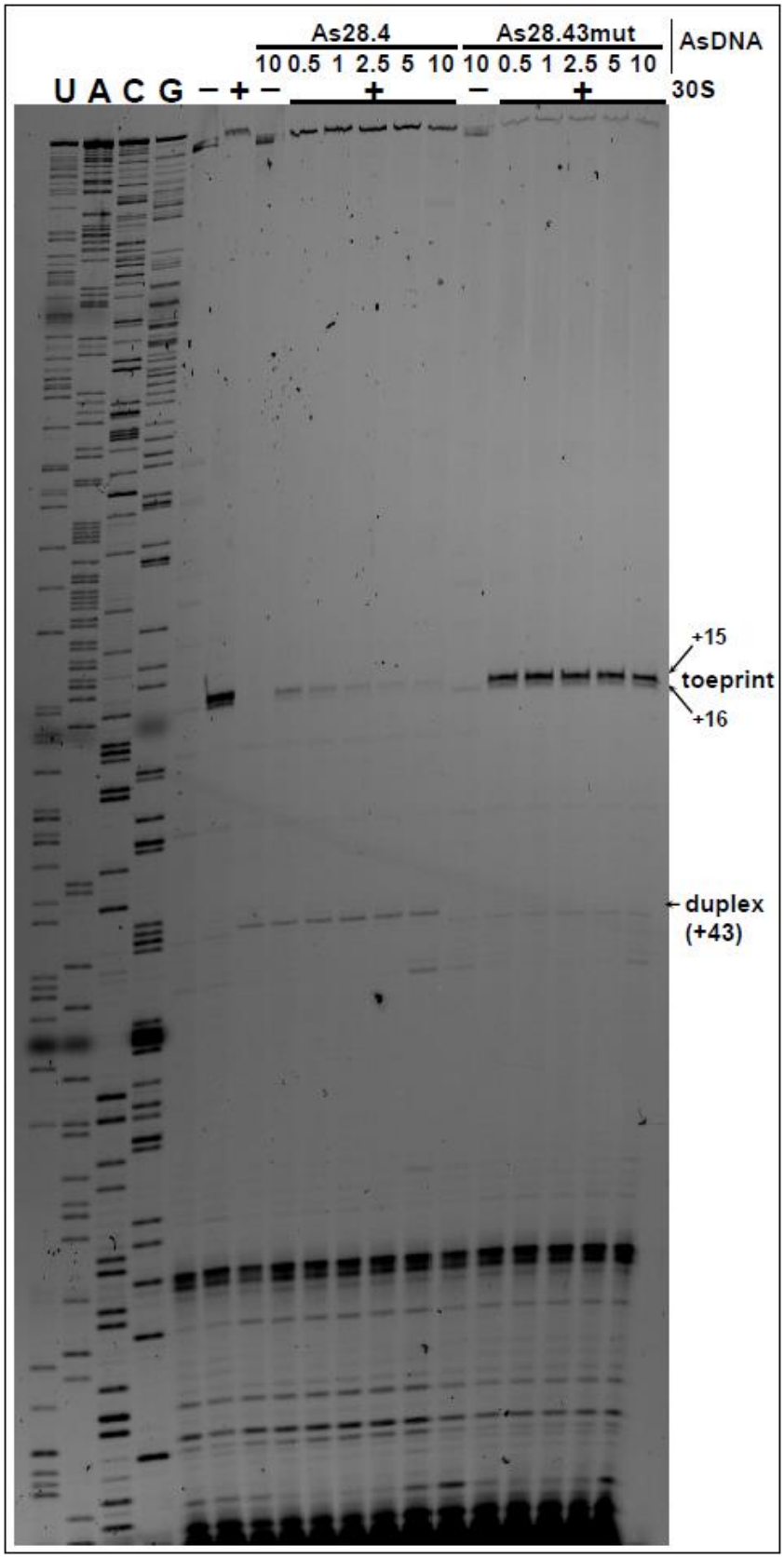


Supplementary figure 2. The Omr 5'-end represses binding of the 30S ribosomal subunit to *fepA* mRNA.

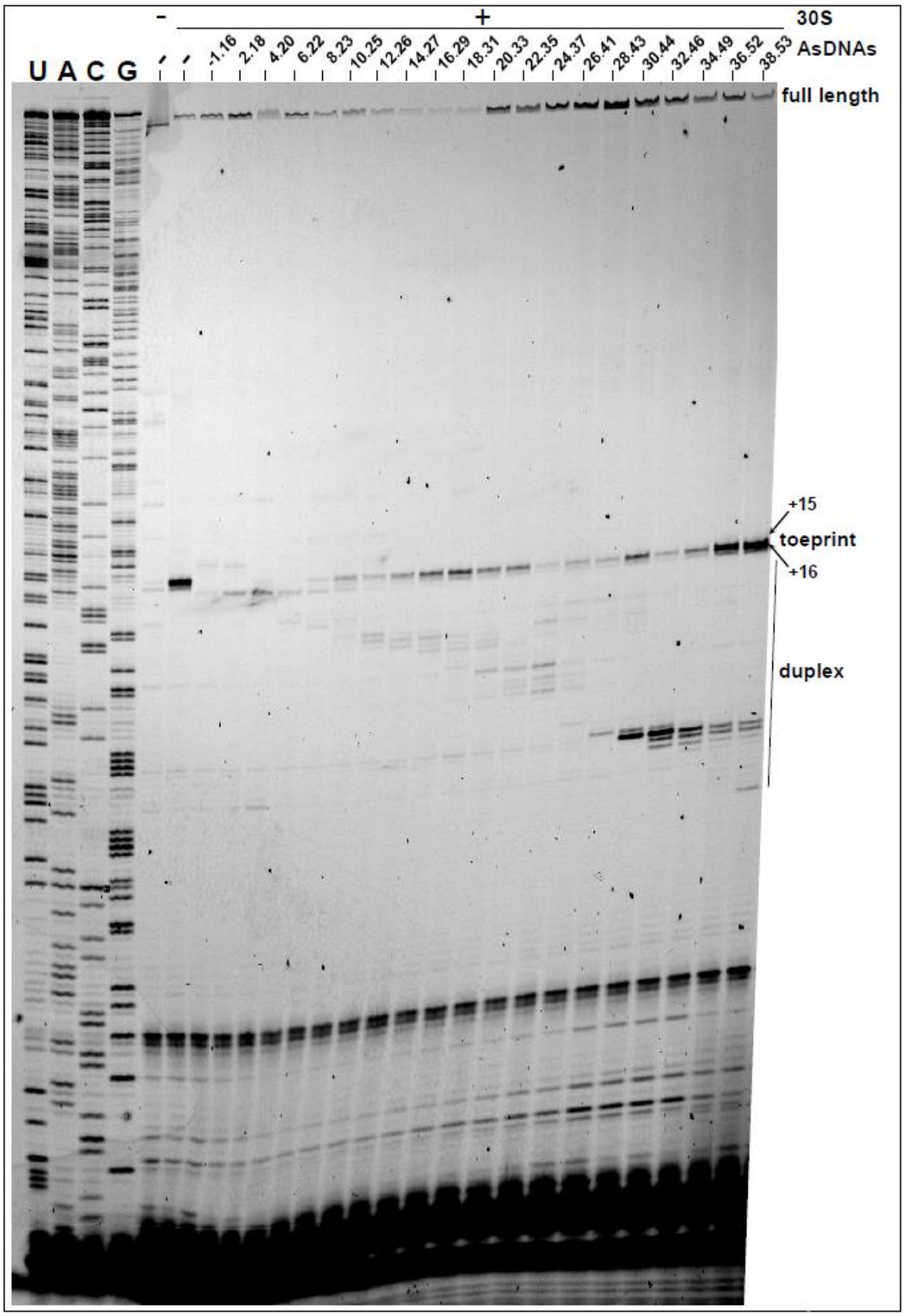
(A) Full gel for figure 3A. (B) Full gel for a repeat of the experiment of Fig. S2A, including the toeprinting analysis in presence of increasing concentrations of OmrB, wt or mutant.



Supplementary figure 3. OmrA specifically inhibits formation of the initiation complex in the presence of Hfq.
Full gel for figure 3B.

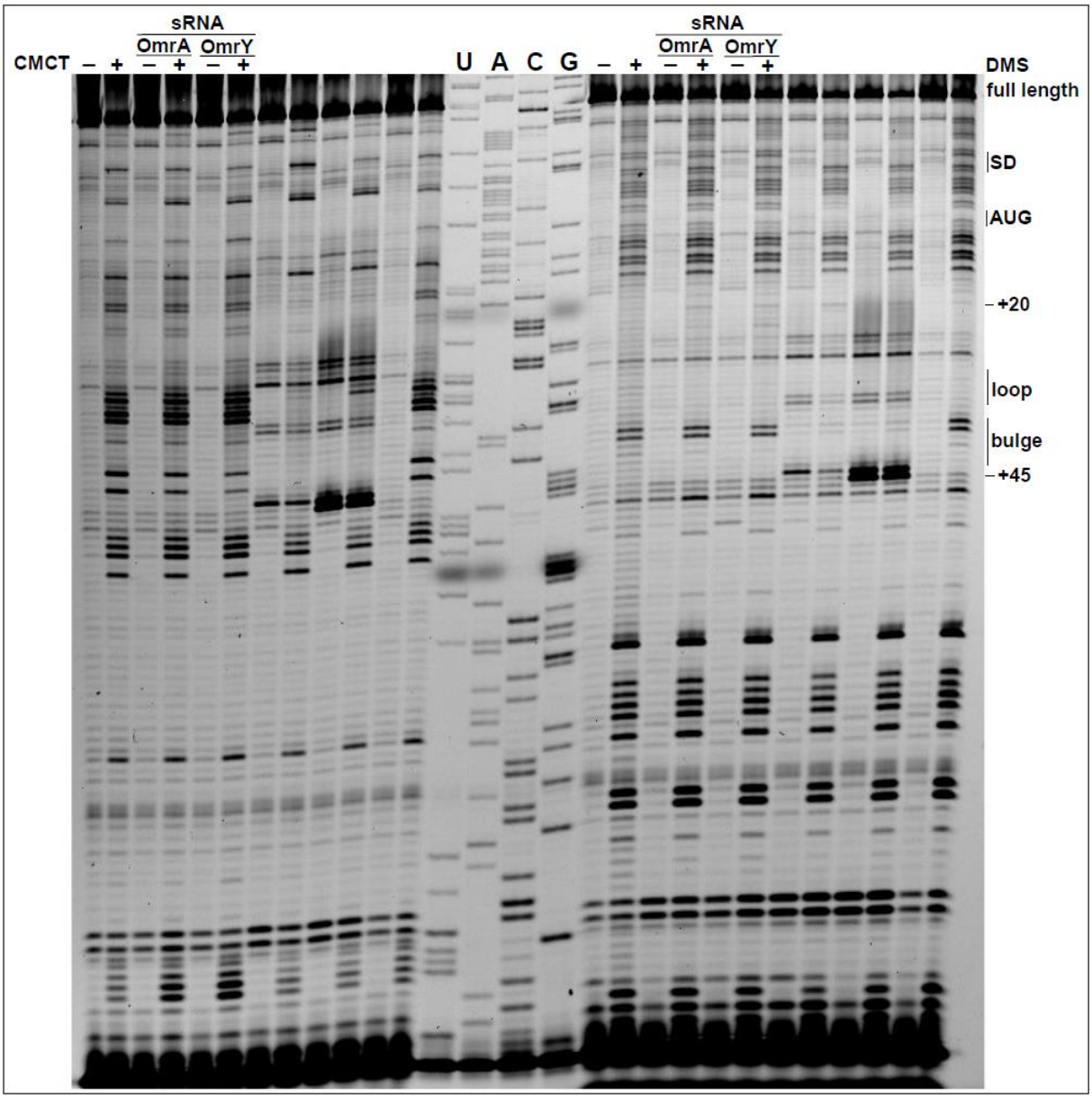


Supplementary figure 4. As28.43 DNA oligonucleotide specifically represses binding of the 30S ribosomal subunit to *fepA* mRNA.
Full gel for figure 3D.

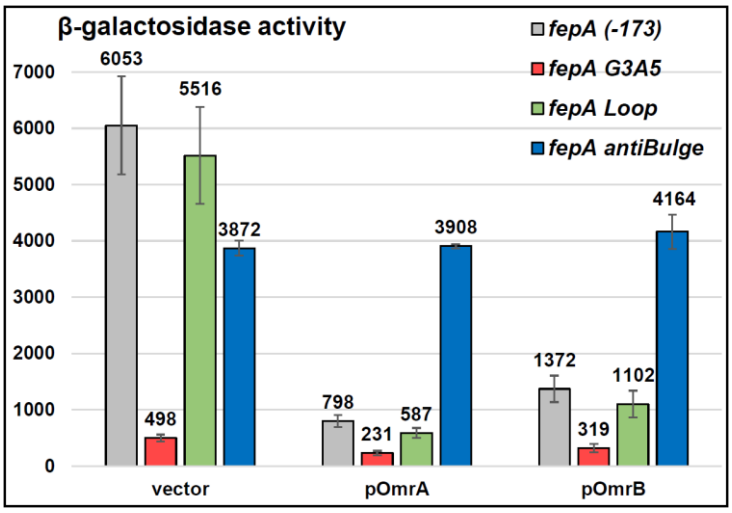


Supplementary figure 5. "Antisense-scanning" of *fepA* mRNA.

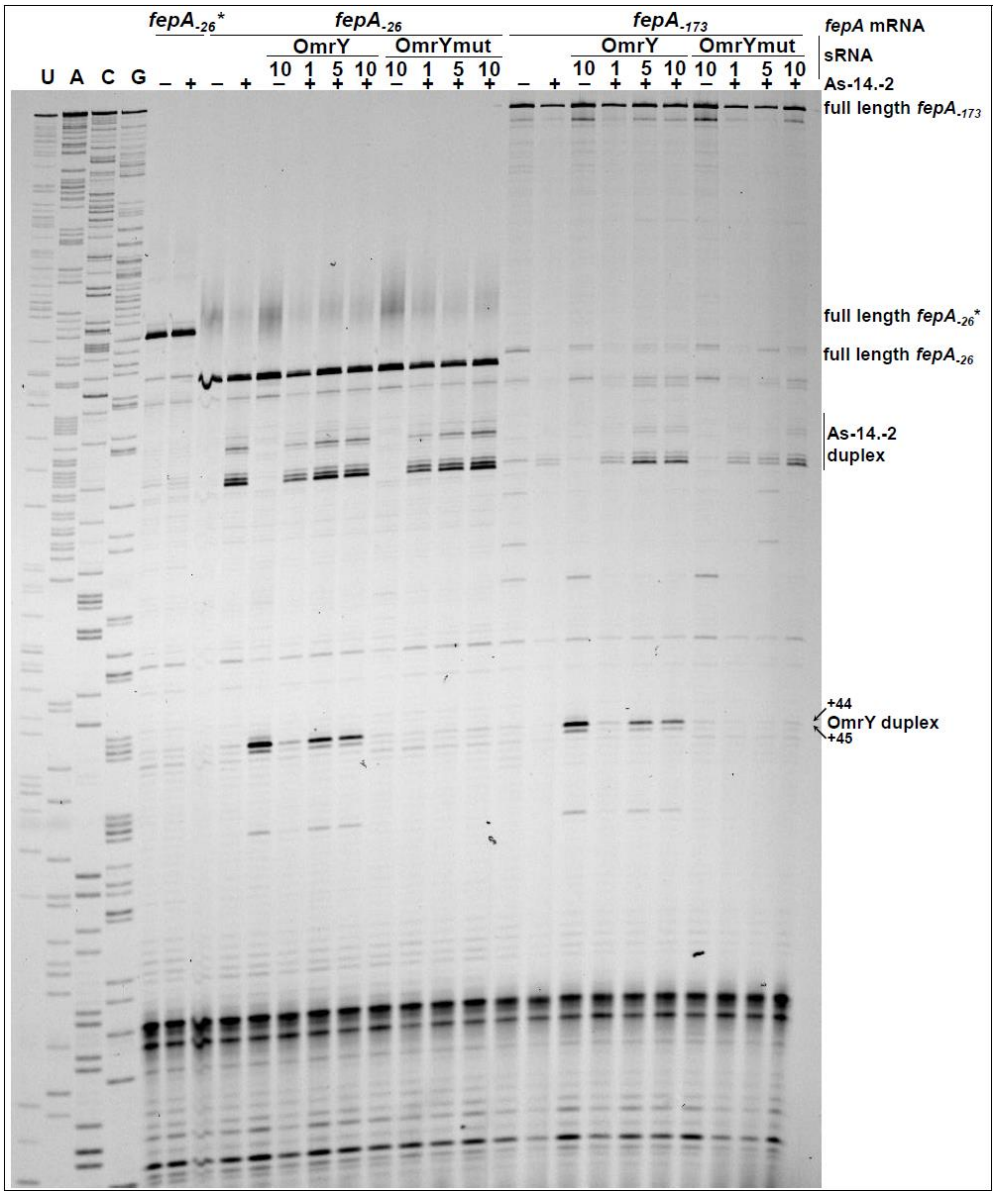
Full gel for figure 4B.



Supplementary figure 6. *In vitro* structural probing of *fepA* mRNA.
Full gel for figure 5A. The probing was also performed in presence of OmrA or OmrY sRNA, at a ratio of 10:1 relative to *fepA* transcript. Only relevant lanes are annotated. No change in RT-signals are visible in presence of either sRNA, suggesting that they did not bind to *fepA* in this experiment.

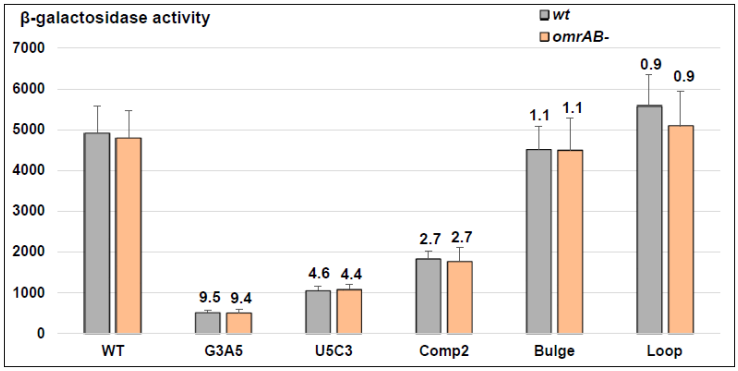


Supplementary figure 7. β-galactosidase activities of strains carrying wt, loop or G3A5 versions of the P_{tet}-*fepA*$_{-173}$-*lacZ* fusion upon overexpression of OmrA or OmrB.
These data were used to calculate the repression factors given in Fig. 5C.



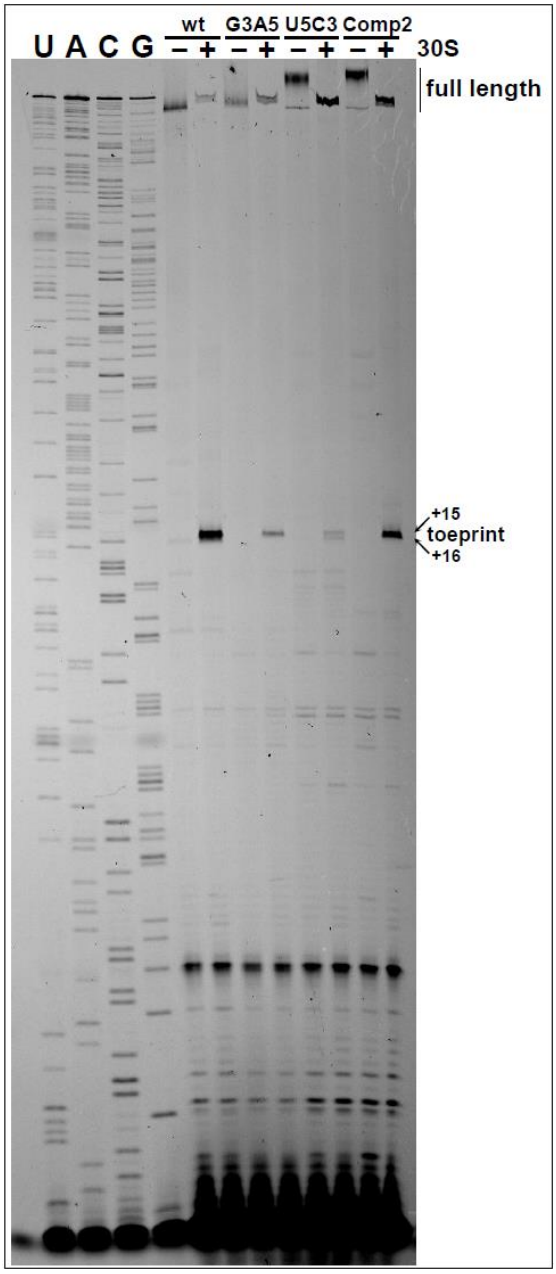
Supplementary figure 8. OmrY does not modify the accessibility of *fepA* SD sequence *in vitro*.

Full gel for figure 5E.



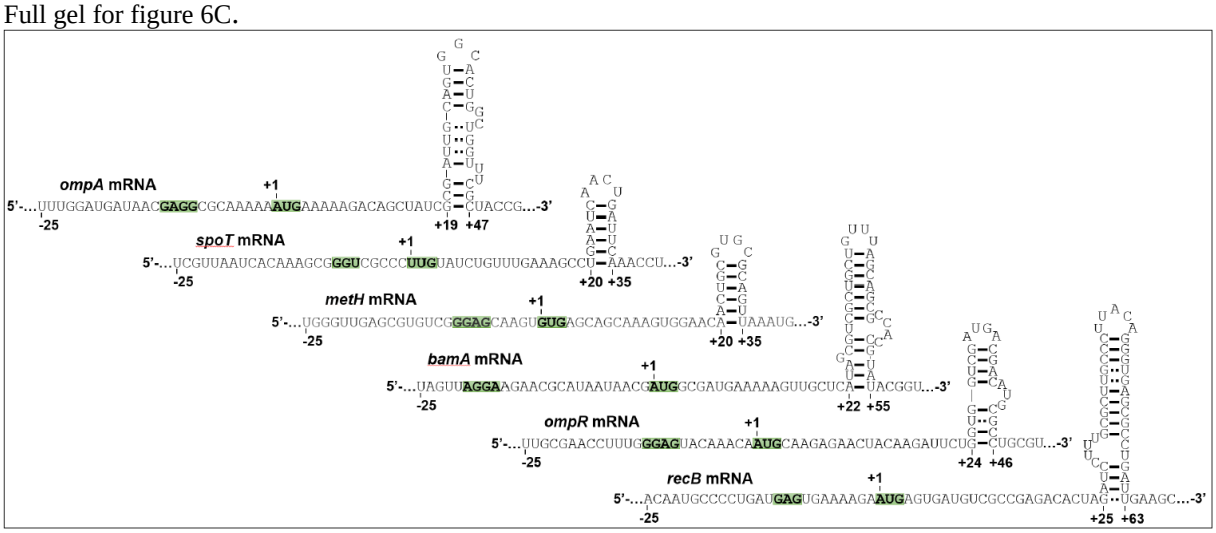
Supplementary figure 9. β-galactosidase activities (in Miller units) of strains carrying the P_{tet}-*fepA*$_{-173}$-*lacZ* fusion, wt or mutated in the SL, in *omrAB*$^{+}$ or *omrAB*$^{-}$ background.

See text and Fig. 6 for details of the mutants. Numbers above the bars indicate the fold-decrease in activity for each mutant in comparison to the wt. Strains used in this experiment are from left to right: MG1772, JJ0233, JJ0229, JJ0237, JJ0241 and JJ0192 for the wt derrivatives, and JJ0015, JJ0286, JJ0287, JJ0288, JJ0282 and JJ0281 for the *omrAB*$^{-}$ mutants.



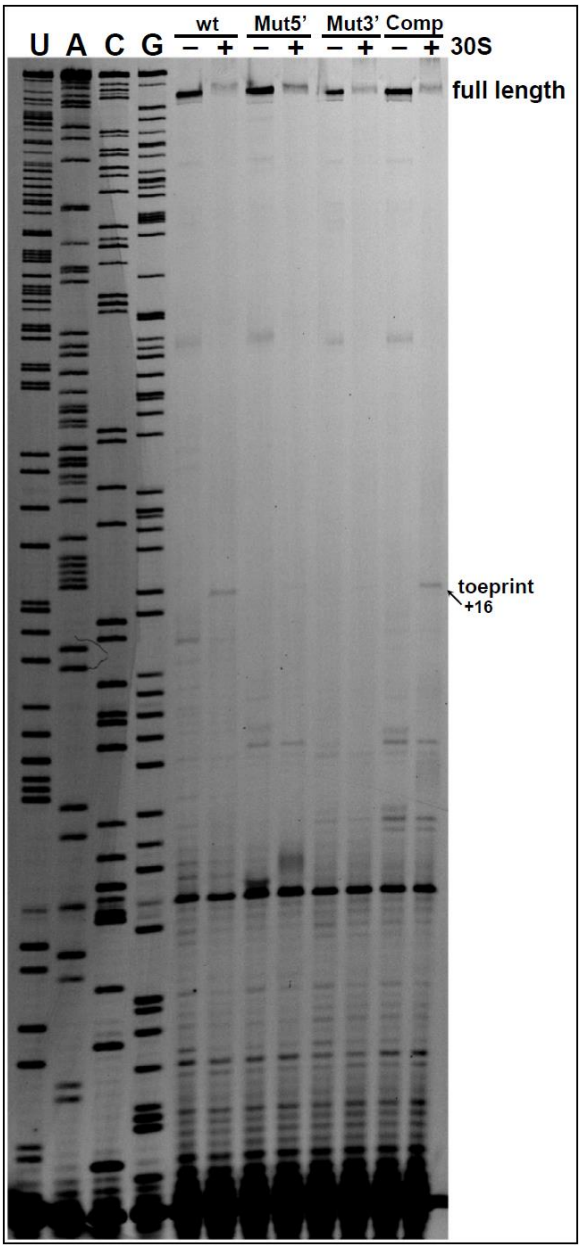
Supplementary figure 10. *In vitro* effect of *fepA* mRNA secondary structure on 30S ribosome binding.

Full gel for figure 6C.

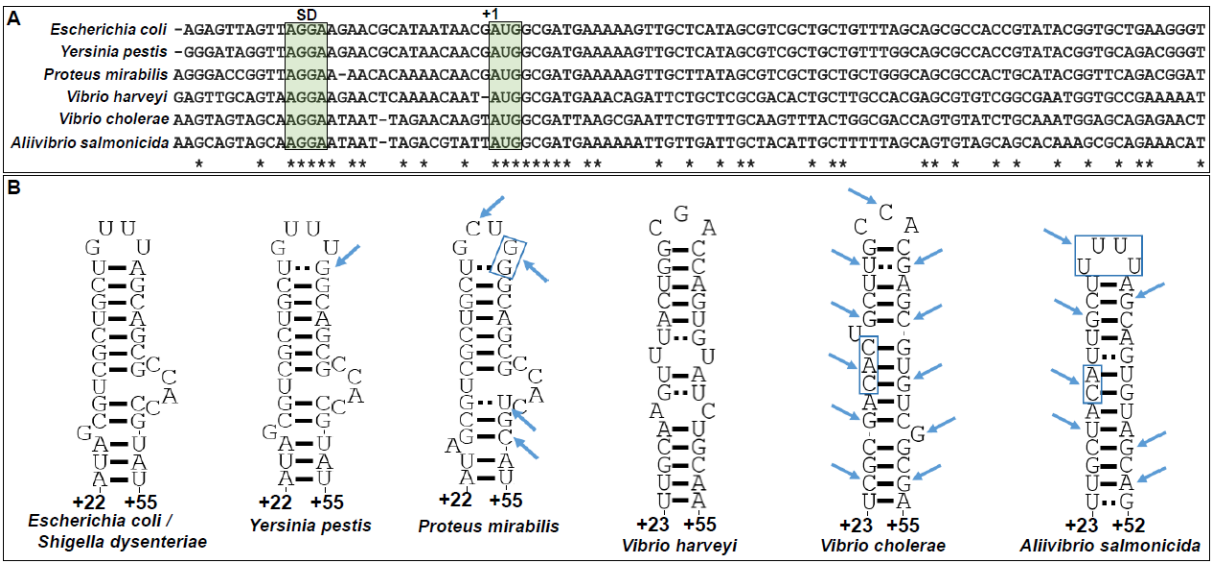


Supplementary figure 11. Stem-loop structures predicted in a subset of *E. coli* mRNAs around the position +20 of the ORF.

SD sequence and AUG start codon are indicated by green squares in each mRNA.

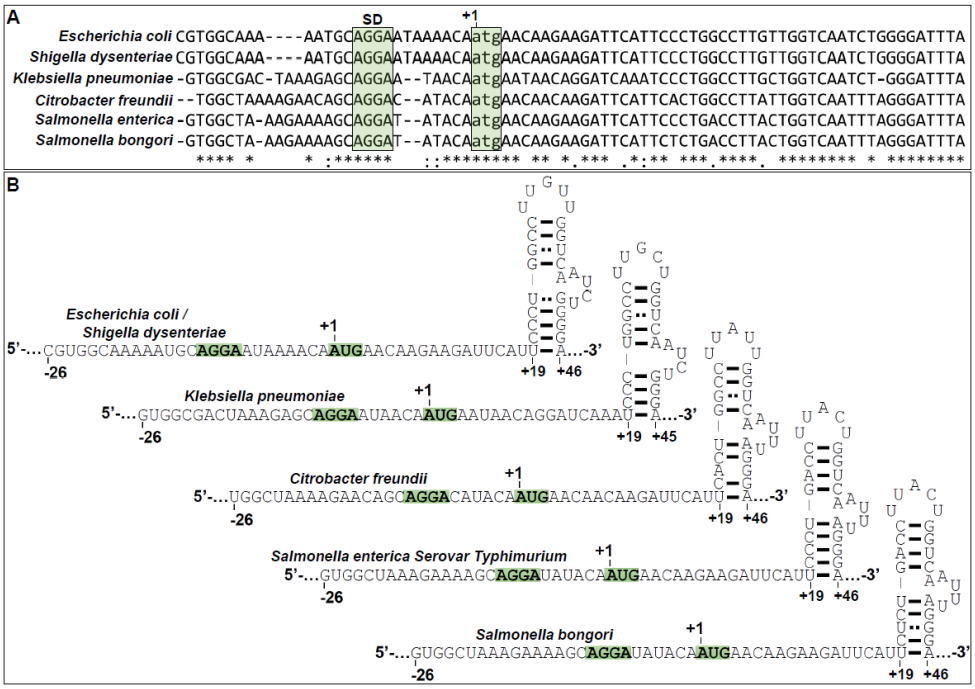


Supplementary figure 12. *In vitro* effect of *bamA* mRNA secondary structure on 30S ribosome binding.
Full gel for figure 7D.



Supplementary figure 13. Conservation of *bamA* mRNA sequence and predicted structures in γ-proteobacteria

bamA sequences from indicated bacteria extending from nts -25 to +50 relative to the start codon were aligned using Clustal Omega (A). SD sequences and AUG start codons are highlighted in green, "*" indicate positions fully conserved, ":" indicate conservation between groups of strongly similar properties and "." indicate conservation between groups of weakly similar properties, as defined in Clustal. The SL derived from these sequences are depicted in (B), where SD sequences and AUG start codons are highlighted in green. Blue arrows indicate nts whose identity differs from the *E. coli* sequence, for *Y. pestis* and *P. mirabilis*, or from the *V. harveyi* sequence, for *V. cholerae* and *A. salmonicida*.



Supplementary figure 14. Conservation of *fepA* mRNA sequence and predicted structures in enterobacteria

fepA sequences from indicated bacteria extending from nts -26 to +50 relative to the start codon were aligned using Clustal Omega (A), with the gap at position +42 of *Klebsiella pneumoniae* sequence being adjusted by hand. SD sequences and AUG start codons are highlighted in green, "*" indicate positions fully conserved, ":" indicate conservation between groups of strongly similar properties and "." indicate conservation between groups of weakly similar properties, as defined in Clustal. The SL derived from these sequences are depicted in (B), where SD sequences and AUG start codons are highlighted in green.

C. Structures secondaires et activation de la traduction

1. L'hypothèse du « switch »

Avant la mise en évidence de l'effet activateur de la structure secondaire de l'ARNm *fepA*, nous avions imaginé une alternative au mécanisme de régulation de l'expression de *fepA* par les OmrA/B. Celle-ci consistait en un changement de structure secondaire de l'ARNm, ou « switch », qui provoquerait la formation d'une structure inhibitrice de la traduction en réponse à la fixation des OmrA/B. Cette hypothèse se serait alors inscrite dans le cadre d'une compréhension plus traditionnelle de l'effet des structures secondaires d'ARNm sur la fixation du ribosome. En effet, comme cela a été discuté dans l'introduction, les structures secondaires impliquant le RBS ont généralement un effet négatif sur la fixation du ribosome, surtout lorsqu'elles occultent les séquences SD et/ou le codon de démarrage de la traduction.

Dans le cas de *fepA*, le principal argument en faveur de l'existence d'un tel switch vient des données d'expériences de sondages de structure *in vitro* (Fig. 45 et voir **Article** Fig. S6). On peut ainsi observer que certaines positions à proximité du SD et du codon de démarrage de *fepA* deviennent légèrement moins accessibles en présence des oligonucléotides antisens AsFepA28-43 et AsFepA30-44, à savoir notamment les nt C-13, A-12, A-1 et A+1 par rapport à la position de démarrage de la traduction.

Nous avons donc décidé d'analyser *in vivo* et dans le détail les deux seules structures alternatives crédibles que nous avons pu imaginer (*id est* les switches A et B, voir Fig. 45). Pour ce faire, nous avons construit des fusions traductionnelles *fepA-lacZ* portant des mutations interdisant la formation d'une ou des deux structures alternatives tout en conservant une tige-boucle en position +20 de l'ORF (mutations Switch1, Switch2 et Switch2-Loop dont les régions concernées et l'effet sur la structure secondaire sont montrées dans les Fig. 46.A, 47.A et 48.A). L'effet régulateur des sRNA OmrA et OmrB sur l'expression de ces fusions (Fig. 46.B, 47.B et 48.B) a ensuite été comparé à leur effet sur l'expression de la fusion wt (Fig. 45.B) pour savoir si cet effet était conservé alors qu'aucun « switch » n'est possible.

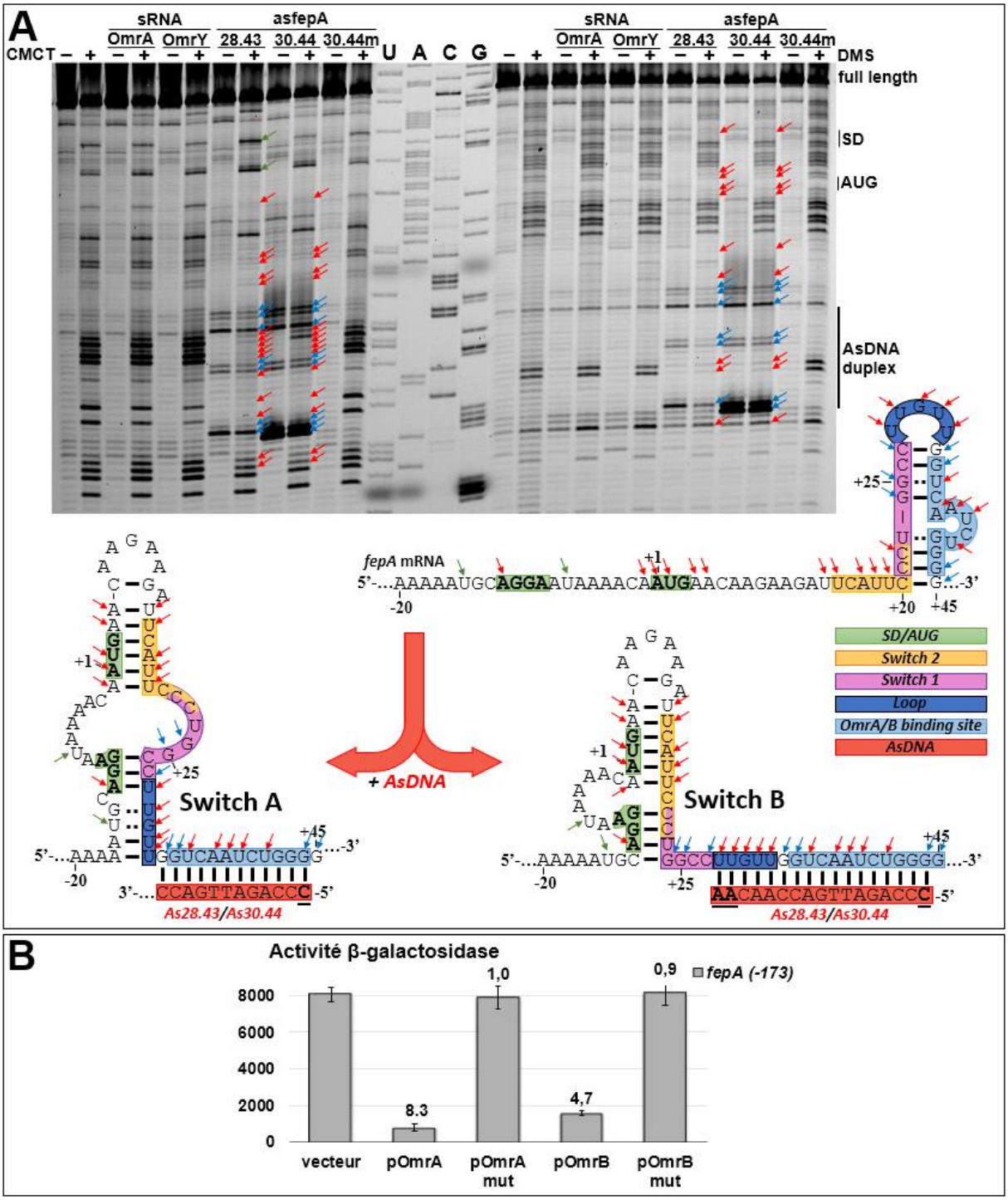


Figure 45 – Un switch de structure pour expliquer l'effet régulateur d'OmrA/B sur fepA ?
A. Sondages de la structure du transcrit fepA des nt -26 à +129 in vitro, seul ou en présence de divers sRNA ou AsDNA à un ratio 10 pour 1 ARNm. Les arrêts de RT et les nt protégés en présence des AsDNA sont indiqués respectivement par des flèches bleues et rouges, tandis que les flèches vertes correspondent aux nt dont la réactivité au CMCT augmente avec la fixation d'As28.43. Les deux switch (A et B) compatibles avec ces données de sondage sont représentés en dessous. Les nt soulignés au sein des AsDNA indiquent les nt présents uniquement dans As28.43 ou As30.44. Remarque : ces deux AsDNA s'apparient à l'ARNm fepA sur une région légèrement plus longue que la région de fixation d'OmrA/B minimale, indiquée en bleu clair sur la figure. B. Activité β-galactosidase spécifique d'une fusion fepA-$_{173}$-lacZ (souche MG1772) transformée par un plasmide contrôle (vecteur) ou surproducteur d'OmrA,

OmrAmut, OmrB ou OmrBmut. Sont indiqués au-dessus des histogrammes les facteurs de répression par rapport au vecteur contrôle ; les activités β-galactosidase sont exprimées en unités Miller.

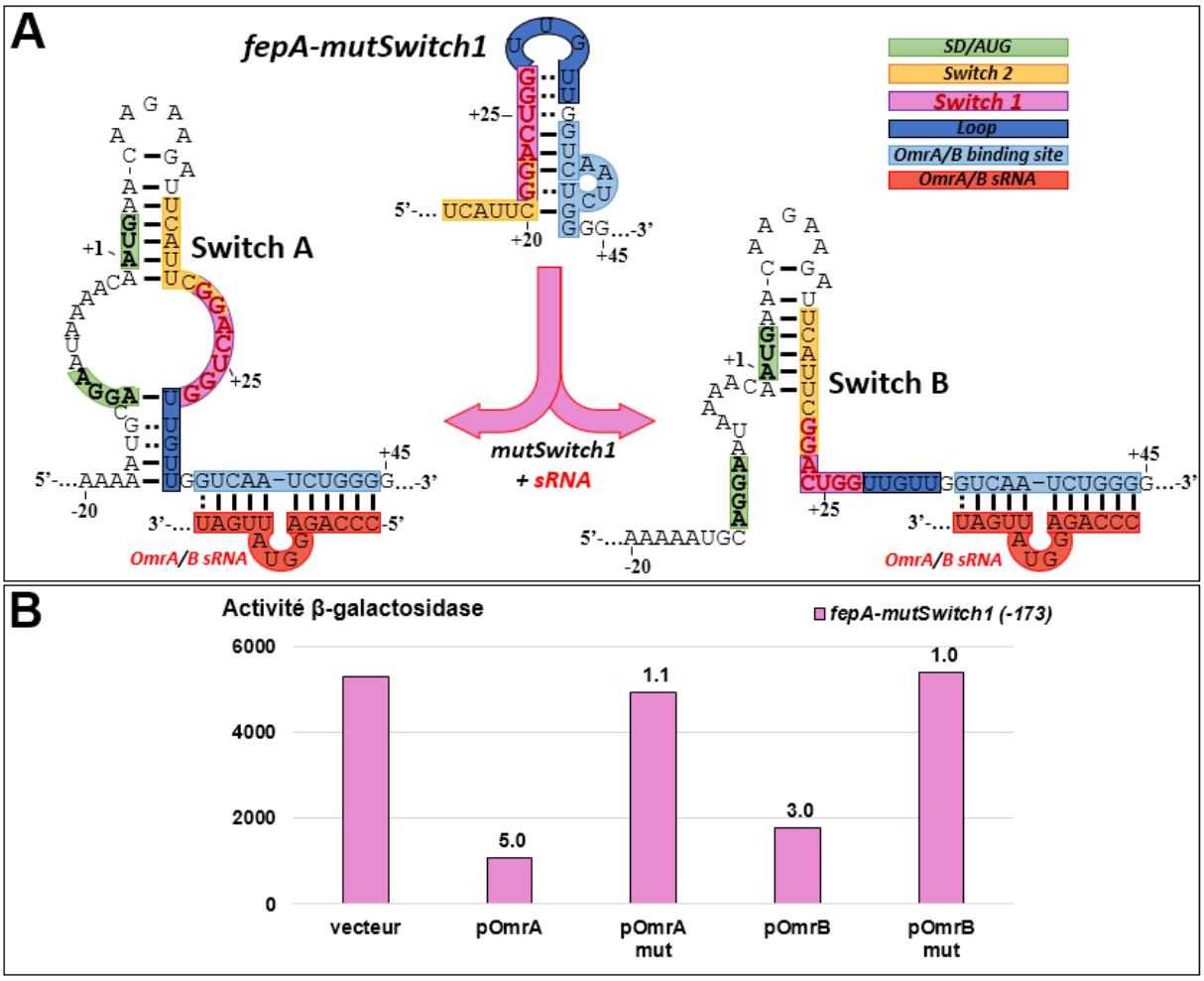


Figure 46 – Le mutant Switch1 de la fusion fepA-lacZ est toujours régulé par OmrA/B
A. Représentation schématique de la structure en tige-boucle de l'ARNm fepA-lacZ portant la mutation Switch1, en présence ou en absence d'OmrA/B. Les nt mutés sont indiqués en rouge. B. Activité β-galactosidase spécifique d'une fusion $fepA_{-173}$-mutSwitch1-lacZ (souche MG2190) transformée par un plasmide contrôle (vecteur) ou surproducteur d'OmrA, OmrAmut, OmrB ou OmrBmut. Sont indiqués au-dessus des histogrammes les facteurs de répression par rapport au vecteur contrôle.

Nous avons ainsi pu observer que quelles que soient les mutations ou la combinaison de mutations utilisées, l'expression des fusions est toujours régulée par OmrA et OmrB. Cela démontre que la régulation par OmrA/B ne dépend pas de la formation de l'une de ces structures alternatives. Par ailleurs, notons que la séquence du RBS de *fepA* est très riche en nt A (Fig. 49.A, voir **Article** Fig. 5 et Fig. S14), ce qui limite grandement la possibilité de former des structures stables pouvant empêcher la fixation du ribosome. Enfin, nous avons démontré que l'accessibilité du RBS ne dépend en aucune manière de la fixation du sRNA chimérique OmrY, ce qui constitue un autre argument fort excluant l'existence d'un tel « switch » (voir **Article**). Bien que l'hypothèse du « starting block » proposée dans l'article reste à démontrer,

par souci de concision et de clarté je nommerai « starting block » (SB) les structures en position +19/+20 de l'ARNm *fepA* et +22 de l'ARNm *bamA*.

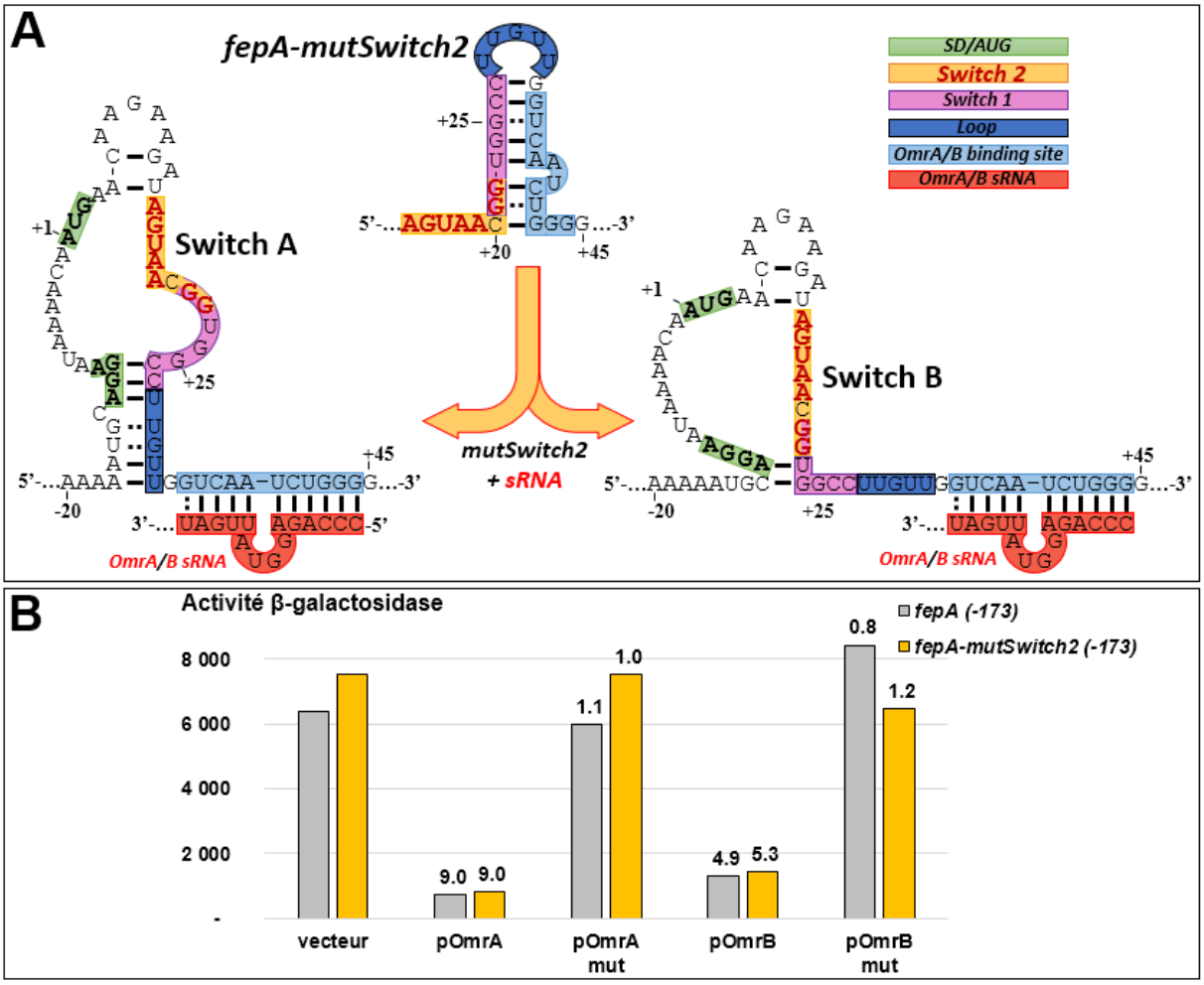


Figure 47 – Le mutant Switch2 de la fusion fepA-lacZ est toujours régulé par OmrA/B
A. Représentation schématique de la structure en tige-boucle de l'ARNm fepA-lacZ portant la mutation Switch2, en présence ou en absence d'OmrA/B. Les nt mutés sont indiqués en rouge. B. Activité β-galactosidase spécifique des fusions fepA$_{-173}$-lacZ (souche MG1772) et fepA$_{-173}$-mutSwitch2-lacZ (souche JJ0169) transformées par un plasmide contrôle (vecteur) ou surproducteur d'OmrA, OmrAmut, OmrB ou OmrBmut. Sont indiqués au-dessus des histogrammes les facteurs de répression par rapport au vecteur contrôle.

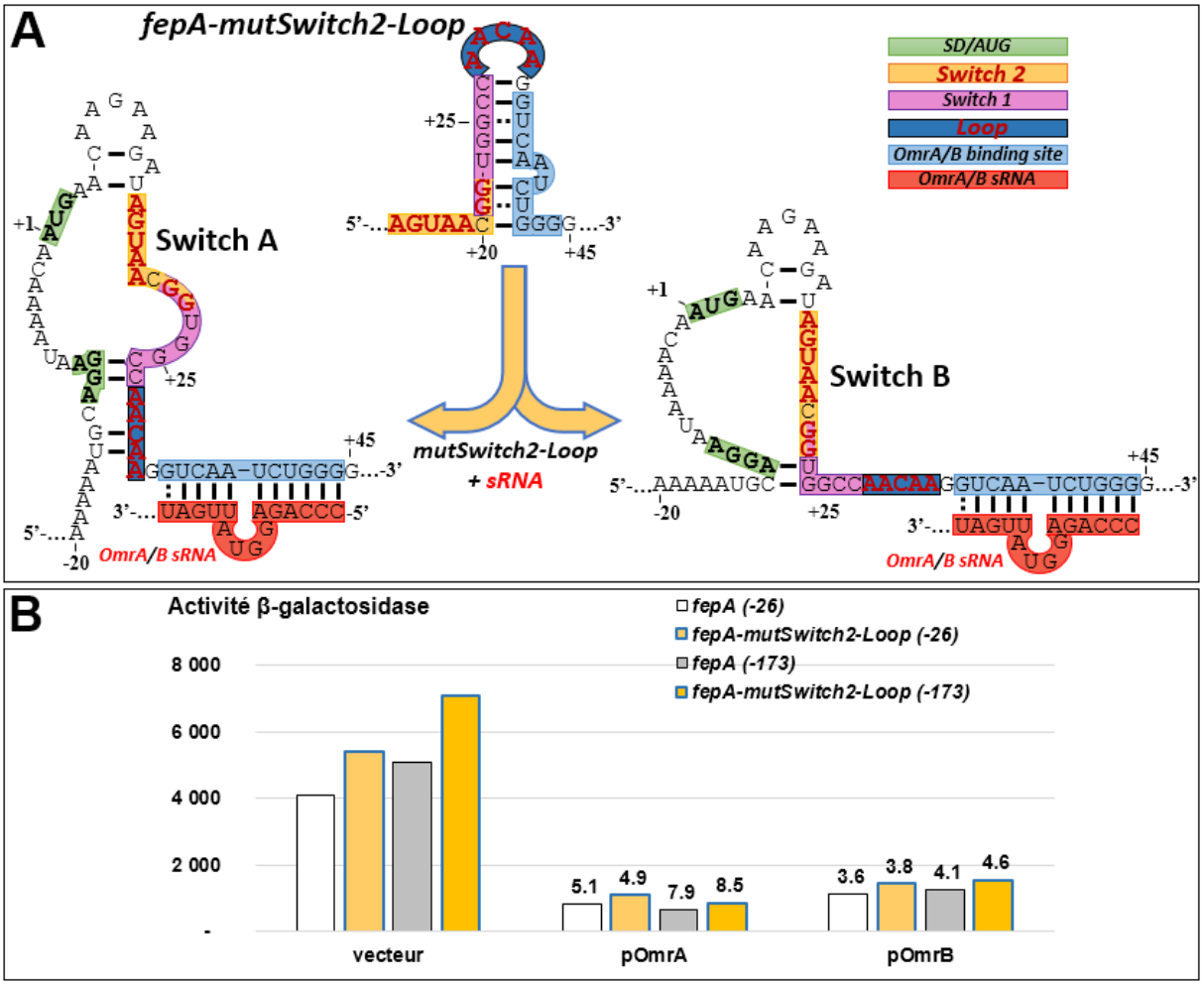


Figure 48 – Le mutant Switch2-Loop de la fusion fepA-lacZ est toujours régulé par OmrA/B
A. Représentation schématique de la structure en tige-boucle de l'ARNm fepA-lacZ portant la mutation Switch2-Loop, en présence ou en absence d'OmrA/B. Les nt mutés sont indiqués en rouge. B. Activité β-galactosidase spécifique des fusions $fepA_{-26}$-lacZ (souche JJ0135), $fepA_{-26}$-mutSwitch2-Loop-lacZ (souche JJ0217), $fepA_{-173}$-lacZ (souche MG1772) et $fepA_{-173}$-mutSwitch2-Loop-lacZ (souche JJ0169) transformées par un plasmide contrôle (vecteur) ou surproducteur d'OmrA, OmrAmut, OmrB ou OmrBmut. Sont indiqués au-dessus des histogrammes les facteurs de répression par rapport au vecteur contrôle.

2. Implication des éléments en *cis* de l'ARNm

Les structures SB de *fepA* et de *bamA* suffisent-elles à activer la traduction ou bien d'autres éléments en *cis* de ces ARNm sont-ils également nécessaires ? Pour ce qui est de *fepA* en particulier, nous savons que cet effet persiste même avec une version courte de l'ARNm, c'est-à-dire des nt -26 à +45 (voir **Article** Fig. 6). Par conséquent, s'il existe des éléments en *cis* déterminants pour l'effet de ces structures SB, il faudrait certainement les chercher dans cette région de l'ARNm.

En particulier, deux expériences que nous avons effectuées suggèrent que l'effet des structures SB dépend du contexte de l'ARNm dans lequel elles se trouvent. Ainsi, nous avons construit deux fusions traductionnelles Ptet-*phoP-fepA-lacZ* en plaçant soit la tige-boucle SB (nt +20 à +45 par rapport au démarrage de la traduction) de *fepA* après le nt +19 de l'ARNm *phoP* (fusion P_{tet}-*phoP-fepA*$_{SB}$-*lacZ*) soit un fragment de l'ORF de *fepA* (nt +1 à +45) après le nt -1 de l'ARNm *phoP* (fusion P_{tet}-*phoP-fepA*$_{ORF}$-*lacZ*) (Fig. 49.A). Les OmrA/B ont été surexprimés dans les souches portant ces fusions afin de mesurer la capacité de ces structures ainsi déplacées dans le contexte de l'ARNm *phoP* à affecter la traduction (Fig. 49.B). Sachant que les OmrA/B agissent *via* la structure SB, si leur effet persiste dans ces nouveaux contextes cela signifierait que la structure est encore activatrice indépendamment du reste du RBS de *fepA*.

La fusion *phoP-lacZ* a été utilisée comme contrôle dans cette expérience et comme attendu, aucune régulation de cette fusion par OmrA/B n'est observée. De plus, les OmrA/B n'ont que très peu voire aucun effet sur la fusion *phoP-fepA*$_{SB}$-*lacZ* (facteurs de répression de 1,7 pour OmrA et de 1,4 pour OmrB), ce qui pourrait indiquer que la déstabilisation de la structure SB par les OmrA/B n'influe pas sur l'expression génétique dans ce contexte. Cela étant dit il n'est pas possible d'exclure que dans ce contexte précis la structure ne se forme pas, ou encore que l'interaction avec OmrA/B soit impossible, ce qui expliquerait l'absence d'effet observé. Toutefois il s'agit d'une première indication que certains éléments en *cis* de l'ARNm *fepA* sont nécessaires au mécanisme d'activation de la traduction par la structure SB.

En revanche, les OmrA/B régulent l'expression de la fusion *phoP-fepA*$_{ORF}$-*lacZ*. Cette répression est approximativement deux fois moins efficace que celle d'une fusion *fepA-lacZ* (Fig. 45.B ou voir **Article** Fig. 1), avec des facteurs de répression de 3,3 pour OmrA et de 2,6 pour OmrB contre 8,3 et 4,7 avec la fusion *fepA-lacZ* wt. A nouveau, il se pourrait que dans ce contexte la structure SB se forme moins bien ou que l'interaction avec les OmrA/B soit moins efficace. Cela étant dit, le fait que la régulation soit partiellement rétablie indique qu'en présence du début de l'ORF de *fepA*, la régulation par les OmrA/B et donc l'effet de la structure SB existe. Cela suggère que certains éléments du RBS de *fepA* sont nécessaires à l'effet de la structure SB, parmi lesquels pourraient figurer les premiers codons très riches en nt A.

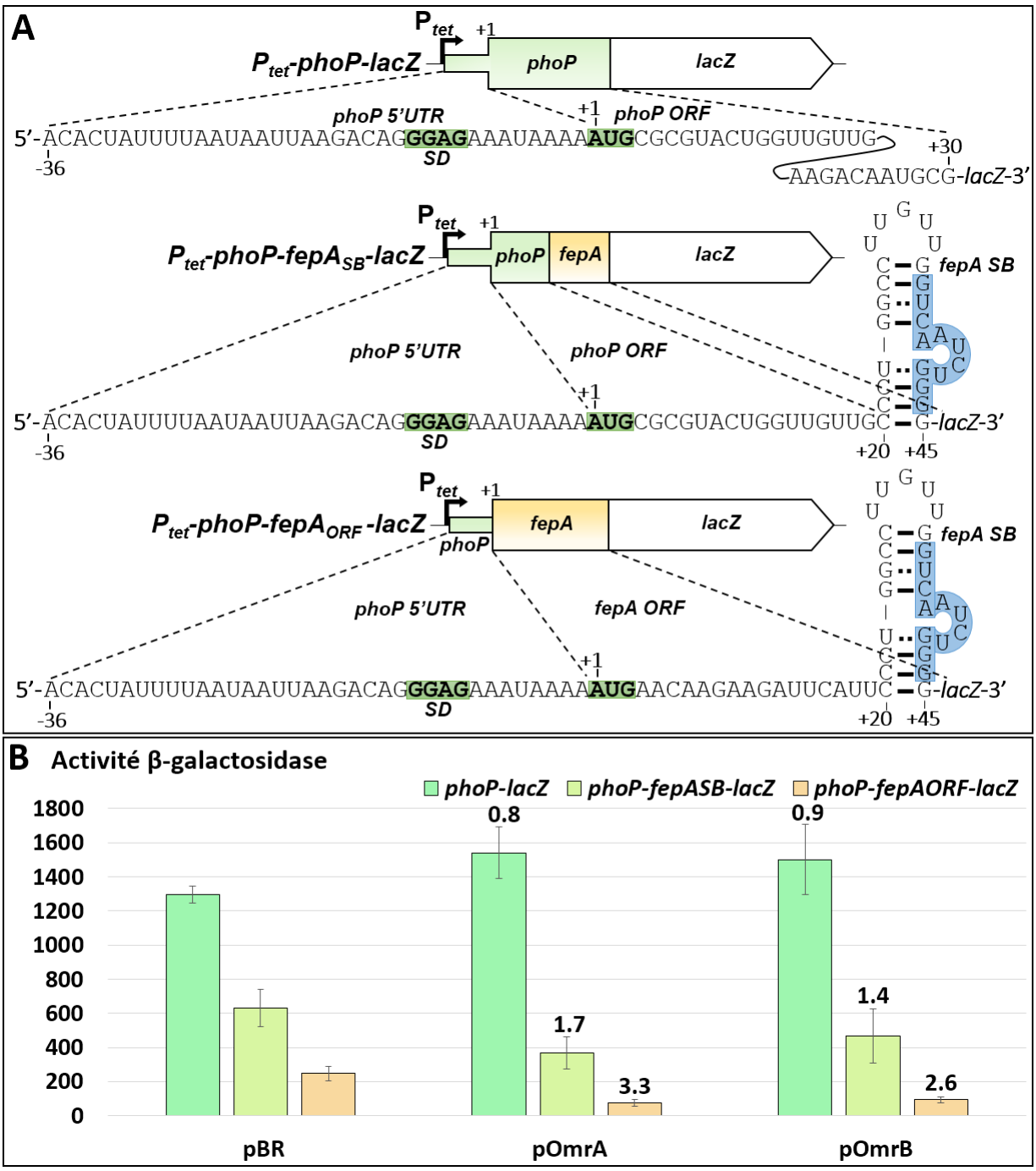


Figure 49 – Les éléments en cis de l'ARNm sont importants pour l'effet de la structure SB de fepA
A. Fusions phoP-lacZ : la fusion contrôle P_{tet}-phoP-lacZ est représentée avec la séquence de l'ARNm phoP depuis son TSS (-36) jusqu'à la position +30, fusionnée au gène lacZ à partir de sa position +28. La transcription de cette fusion est sous le contrôle du promoteur constitutif P_{tet}. Dans cette fusion, nous avons remplacé (i) les nt +20 à +30 de phoP par les nt +20 à +45 de fepA (fusion P_{tet}-phoP-$fepA_{SB}$-lacZ) et (ii) les nt +1 à +30 de phoP par les nt +1 à +45 de fepA (fusion P_{tet}-phoP-$fepA_{ORF}$-lacZ). Sont encadrés en vert la séquence SD et le codon de démarrage de la traduction, et en bleu le site de fixation minimal d'OmrA/B. B. Activités β-galactosidase spécifiques des fusions phoP-lacZ (souche MG1511), phoP-$fepA_{SB}$-lacZ (souche JJ0110) et phoP-$fepA_{ORF}$-lacZ (souche MG2213) en présence d'un plasmide contrôle (pBR) ou surproducteur d'OmrA ou OmrB. Les facteurs de répression par rapport au vecteur contrôle sont indiqués. Cette expérience a été réalisée en duplicat.

Afin de poursuivre l'étude sur l'importance du contexte de l'ARNm, nous avons voulu savoir si les structures SB de *bamA* et de *fepA* étaient interchangeables. Pour ce faire, nous avons construit une fusion *bamA-fepA*$_{SB}$*-lacZ* ou sa version mutante pour le SB *bamA-fepA*$_{SB}$*-G3A5-lacZ* (Fig. 50.A).

Contre toute attente, la fusion mutante *bamA-fepA*$_{SB}$*-G3A5-lacZ* a une activité β-galactosidase quatre fois plus élevée que la fusion *bamA-fepA*$_{SB}$*-lacZ wt*, et relativement proche de la fusion contrôle *bamA-lacZ* (Fig. 50.B). Un certain nombre d'hypothèses peuvent apporter une explication à ce résultat surprenant. Une première possibilité serait que dans le contexte de l'ARNm *bamA*, la structure de *fepA* ne se forme pas et que, dans ce cas précis, la séquence G3A5 soit plus propice à une traduction efficace de la protéine de fusion. Il est également possible d'imaginer que la structure SB de *fepA* ait un effet inhibiteur de la traduction dans ce contexte, du fait d'un mécanisme de régulation différent entre *bamA* et *fepA*, et ce malgré un effet similaire sur la fixation du ribosome *in vitro* (voir **Article** Fig. 6). On peut par exemple imaginer que la stabilité de la structure ait une importance dans le contexte de chacun des ARNm, auquel cas la structure de *fepA* n'est peut-être pas suffisamment stable ou au contraire trop stable dans le contexte de l'ARNm *bamA*.

Quoi qu'il en soit cela montre que la structure SB de *fepA* ne peut remplir la fonction occupée par la structure SB de *bamA* dans l'ARNm *bamA*. Il serait cependant intéressant de réaliser l'expérience réciproque avec la structure SB de *bamA* dans l'ARNm *fepA* avant d'en tirer plus de conclusions. Enfin si d'autres structures SB venaient à être découvertes dans d'autres ARNm, il serait certainement plus aisé d'en déduire les caractéristiques communes les plus importantes, mais aussi les éléments en *cis* les plus déterminants pour l'effet de ces structures sur la traduction de leurs ARNm (voir **Discussion A.2**).

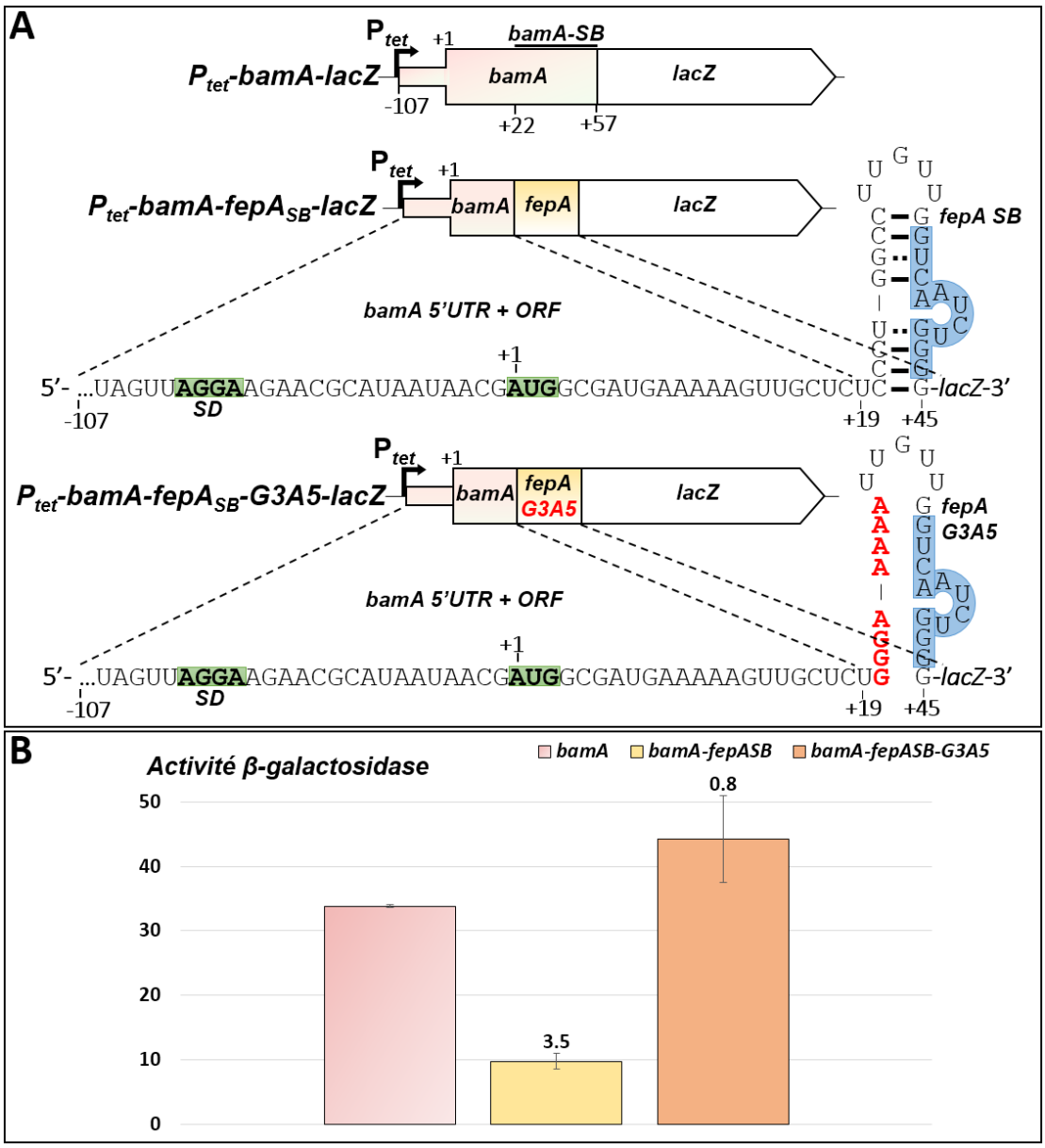


Figure 50 – La structure SB de fepA ne peut remplacer celle de bamA
A. Fusions bamA-lacZ : la fusion contrôle P_{tet}-bamA-lacZ comprend le gène bamA depuis son TSS (-107) jusqu'au nt +57, suivi du gène lacZ à partir de sa position +28. La transcription de cette fusion est sous le contrôle du promoteur constitutif P_{tet}. Dans cette fusion, les nt +18 à +55 de bamA ont été remplacés (i) par les nt +19 à +45 de fepA (fusion P_{tet}-bamA-fepA$_{SB}$-lacZ) ou bien (ii) par leur version mutée G3A5 (fusion P_{tet}-bamA-fepA$_{SB}$-G3A5-lacZ). Les nt mutés sont indiqués en rouge. Sont encadrés en vert la séquence SD et le codon de démarrage de la traduction, et en bleu le site de fixation minimal d'OmrA/B. B. Activités β-galactosidase spécifiques des fusions bamA-lacZ (souche JJ0383), bamA-fepA$_{SB}$-lacZ (souche JJ0387) et P_{tet}-bamA-fepA$_{SB}$-G3A5-lacZ (souche JJ0388). Les facteurs de répression par rapport à la fusion bamA-lacZ sont indiqués. Cette expérience a été réalisée en triplicat.

3. L'hypothèse du thermosenseur

Si la structure SB de *fepA* permet de répondre à la synthèse d'OmrA/B, une telle structure peut-elle également répondre à la température, à la manière d'un thermosenseur ?

Pour tester cette hypothèse, j'ai réalisé une expérience en utilisant des souches portant soit la fusion *fepA*$_{-173}$-*lacZ* wt, soit la version mutante du SB *fepA*$_{-173}$-*G3A5-lacZ*. La culture de ces souches à 30°C, 37°C, 40°C ou bien à 40°C avec une induction à 44°C a ainsi permis de mesurer une éventuelle réponse de *fepA* à la température dépendante du SB (Fig. 51). Cette mesure correspond au rapport entre l'activité β-galactosidase de la fusion wt et celle de la fusion mutante pour le SB.

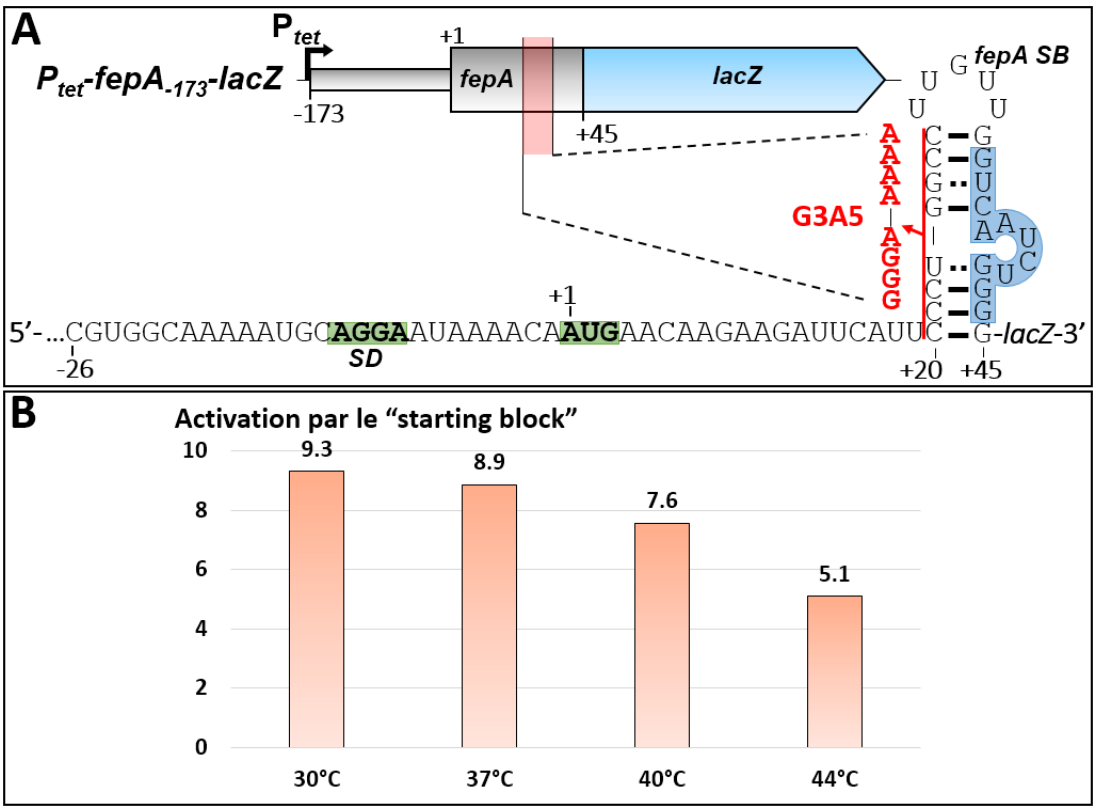


Figure 51 – La structure SB de fepA répond à la température
A. Fusions fepA-lacZ : le gène fepA des nt -173 à +45 est fusionnée au gène lacZ à partir du nt +28. La transcription de cette fusion est sous le contrôle du promoteur constitutif Ptet. La mutation G3A5, représentée en rouge, a été introduite dans la structure SB de manière à en empêcher la formation. B. Rapports des activités β-galactosidase des fusions fepA$_{-173}$*-lacZ (souche MG1772) et fepA*$_{-173}$*-G3A5-lacZ (souche JJ0233) à différentes températures. Les bactéries ont été cultivées en milieu LB aux températures indiquées (30°C, 37°C, 40°C) jusqu'à une densité optique à 600nm (OD*$_{600}$*) de 0,4 ou bien à 40°C jusqu'à OD*$_{600}$ *de 0,15 puis à 44°C jusqu'à OD*$_{600}$ *de 0,4 (44°C). Les rapports d'activation par la structure SB sont indiqués.*

Le résultat de cette expérience indique qu'il pourrait bien y avoir un effet de la température sur l'activation traductionnelle par le SB. Plus précisément, l'activation de la traduction par le SB s'amenuise progressivement à des températures supérieures à 37°C. Ainsi, au-delà de 37°C le SB pourrait donc permettre une modulation très précise de la synthèse de FepA. Cette expérience nécessite toutefois d'être reproduite et accompagnée

d'expériences d'immunodétection (« Western Blot », WB) de la protéine FepA produite à partir d'ARNm wt ou mutant. De plus, des expériences *in vitro* permettant d'analyser la structure de l'ARNm *fepA* à diverses températures pourraient aider à déterminer si cet effet est bien dû à la dénaturation progressive de la structure à haute température, ou si d'autres facteurs en *trans* sont impliqués. En effet, il n'est pas exclu que les OmrA/B soient à l'origine de cet effet sur l'expression de *fepA*, avec une induction plus importante à haute température qui pourrait être causée par σS par exemple. Cette dernière possibilité pourrait être exclue par la même expérience réalisée avec une souche délétée pour les gènes *omrA/B* (*ΔomrA/B*).

D'après cette expérience préliminaire il est toutefois possible de supposer que l'expression de *fepA* réponde à la température d'une façon SB-dépendante. Cela représente un intérêt pour la compréhension de la physiologie de FepA. En effet, la bactérie pourrait connaître des températures supérieures à 37°C à l'occasion d'une interaction avec un hôte non humain dont la température corporelle est naturellement plus élevée, ou encore avec un hôte humain dont la température corporelle est plus élevée qu'à l'accoutumée du fait d'une réponse immunitaire. Dans ces situations la production de la protéine FepA pourrait être délétère pour la bactérie, peut-être parce qu'elle est très immunogène et/ou parce qu'elle est mieux détecté par certains hôtes. Par ailleurs, des températures supérieures à 37°C induisent également un stress membranaire qui pourrait nécessiter la production d'autres OMP en priorité par rapport à FepA, d'où cette potentielle réponse à la température. A ce titre, savoir si la structure SB de l'ARNm *bamA* répond de la même façon à la température serait très intéressant, puisqu'il s'agit d'un gène dont l'expression est cruciale lors d'un stress membranaire pour intégrer correctement les OMP à la membrane externe.

4. L'existence d'autres structures secondaires dans l'ARNm *fepA*

Alors que notre article est principalement focalisé sur la structure SB, les expériences de sondage chimique de la structure de l'ARNm *fepA* (Fig. 45, voir **Article** Fig. 5) mettent en évidence l'existence d'au moins une autre structure secondaire, également prédite par le programme « mfold », en aval du SB (Fig. 52). Je nommerai cette structure SL2 (« Stem-Loop 2 ») pour plus de clarté.

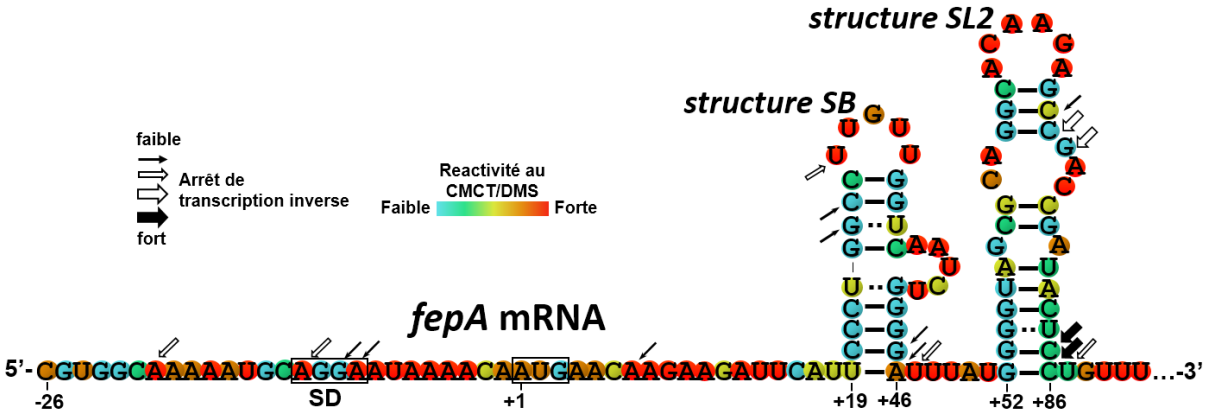


Figure 52 – Structure secondaire de l'ARNm fepA
Représentation schématique déduite des prédictions de structure fournies par mfold et des expériences de sondage in vitro (voir Fig. 45) réalisées sur un fragment de l'ARNm fepA des nt -26 à +129. La réactivité des nt au CMCT ou au DMS est représentative de leur accessibilité, tandis que les arrêts de transcription inverse indiquent une structure de l'ARNm bloquant la progression de l'enzyme. Les appariements Watson-Crick sont représentés par des traits pleins et les appariements Wobble par des pointillés. La séquence SD et le codon de démarrage de la traduction sont encadrés. Les structures SB (« starting block ») et SL2 (« stem-loop 2 ») sont indiquées.

Etant donné qu'un certain nombre d'expériences ont été réalisées avec des fusions ne portant que les nt -173 à +45 ou -26 à +45 de *fepA*, il est clair que cette structure SL2 n'est pas nécessaire à l'action d'OmrA/B et plus généralement à l'activation de la traduction de *fepA* par la structure SB. Néanmoins, la structure SL2 pourrait présenter un intérêt pour la compréhension de la régulation du gène *fepA.* En effet, nous avons poursuivi l'analyse de l'effet d'oligonucléotides antisens sur la fixation du ribosome à l'ARNm *fepA* légèrement au-delà ce qui est présenté dans l'article (Fig. 53 à comparer avec l'article Fig. 4). En l'occurrence, cette expérience a été réalisée avec deux antisens supplémentaires, à savoir AsFepA40.54 et AsFepA42.56, qui s'apparient à la base de la structure SL2. Or, ces oligonucléotides ont un effet sur la fixation du ribosome à l'ARNm *fepA* (Fig. 53.B pistes 40.54 et 42.56). En plus du groupe 1 d'oligonucléotides ayant un effet sur le toeprint parce qu'ils s'hybrident au RBS et du groupe 2 dont l'effet s'explique par la dénaturation de la structure SB, les antisens AsFepA40.54 et AsFepA42.56, et éventuellement d'autres qui s'hybrideraient plus aval, pourraient former un troisième groupe. Celui-ci pourrait agir *via* la dénaturation de la structure SL2, dont le mécanisme d'action sur la fixation du ribosome serait alors à déterminer.

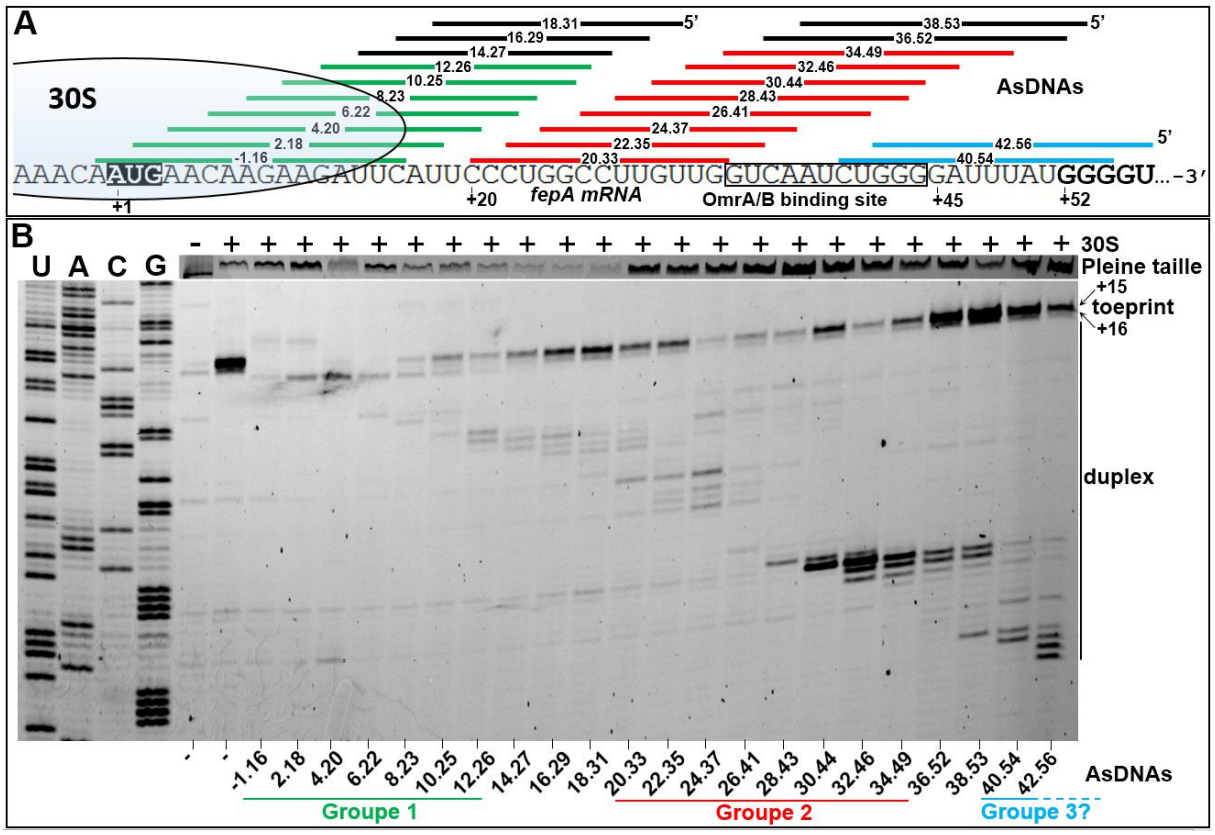


Figure 53 – Des oligonucléotides antisens ciblant la structure SL2 de fepA ont un effet sur le toeprint
A. Représentation schématique de l'ensemble des oligonucléotides ADN utilisés dans cette expérience (dénotés « -1.16 » pour AsFepA-1.16 ; « 2.18 » pour AsFepA2.18 et ainsi de suite). Les oligonucléotides en vert (groupe 1) s'appairent avec l'ARNm fepA au niveau du RBS ; ceux en rouge s'appairent à des nt de la structure SB tandis que ceux en bleu pourraient s'apparier à des nt de la structure SL2. Les nt de fepA appartenant à cette structure sont indiqués en gras ; le site de fixation d'OmrA/B est encadré et le codon de démarrage de la traduction surligné en noir. B. Toeprint réalisé sur le fragment d'ARNm fepA des nt -173 à +129. Les oligonucléotides représentés en (A) ont été utilisés dans les réactions de toeprint au ratio 10 pour 1 ARNm.

Ces effets devront être confirmés, par exemple par une expérience avec des antisens s'hybridant plus en aval, et surtout par la démonstration de l'existence de cette structure *in vivo* et de son éventuel effet sur la traduction et la fixation du ribosome. Il est toutefois tentant d'imaginer un modèle où à l'image d'un riboswitch, cette structure SL2 servirait à intégrer d'autres signaux environnementaux en changeant de conformation. Ce changement de conformation pourrait par exemple stabiliser ou déstabiliser la structure SB en fonctions de diverses conditions, agissant de ce fait sur la fixation du ribosome.

D. Le rôle des RNases et de Hfq dans la régulation par OmrA/B

1. L'implication de la RNase E

L'intervention de la RNase E dans la régulation de l'expression de *fepA* par les OmrA/B a été abordée dans l'article (voir **Article** Fig. 2). Il semblerait ainsi qu'en réponse à l'expression d'OmrA/B, l'ARNm *fepA* reste stable en l'absence de la RNase E alors qu'il est complètement dégradé en sa présence. Néanmoins, la régulation de la synthèse de FepA s'observe en l'absence de la RNase E, bien qu'elle soit moins efficace surtout pour OmrB. Cette différence entre OmrA et OmrB pose en soi question : est-ce que l'effet d'OmrB dépend plus de la RNase E que l'effet d'OmrA ?

Les conclusions de l'article permettent toutefois d'avancer que l'effet régulateur des OmrA/B est avant tout traductionnel et dépend principalement de la dénaturation de la tige-boucle SB (voir **Article** Fig. 5), puisque l'essentiel de la régulation par OmrA/B est perdu dans un mutant du SB *fepA*G3A5. On peut en déduire que la dégradation de l'ARNm *fepA* en présence de la RNase E est une conséquence de la régulation traductionnel et de l'absence de ribosomes stabilisant l'ARNm plutôt que sa cause. On note toutefois que lors d'une surexpression d'OmrA ou OmrB, il subsiste respectivement une régulation d'un facteur 2,2 et 1,6 de l'expression d'une fusion *fepA-G3A5-lacZ* (voir **Article** Fig. 5). Cela signifie que l'on observe toujours une légère régulation post-transcriptionnelle en l'absence de la tige-boucle, qui pourrait correspondre à la part de régulation imputable à l'action de la RNase E.

La RNase E pourrait donc avoir une importance pour l'efficacité de la régulation par les OmrA/B. En effet une faible fixation du ribosome persiste en absence du SB (voir **Article** Fig. 6), or si un cycle de traduction a lieu, l'interaction entre OmrA/B et *fepA* sera levée du fait du passage du ribosome. Cela entrainerait la reformation de la structure SB et une activation de la traduction jusqu'à ce qu'un nouvel appariement entre OmrA/B et *fepA* s'établisse. Aussi pourrait-on proposer que la RNase E permette d'achever l'action des OmrA/B en clivant l'ARNm, empêchant ainsi toute traduction résiduelle de l'ARNm *fepA*.

2. L'implication de la RNase III

Il existe une autre RNase dont le rôle éventuel a été éludé dans notre article, à savoir la RNase III, dont l'importance dans la régulation par certains sRNA est bien établie (voir **Revue n°1**). Nous avons choisi de ne pas en discuter dans l'article en toute connaissance de cause : la régulation de la synthèse de FepA par les OmrA/B dans une souche *Δrnc* est strictement identique à celle observée dans une souche wt (Fig. 54). Le résultat est plus étonnant pour les niveaux de l'ARNm *fepA* dans la mesure où cet ARNm est plus abondant dans la souche wt, mais semble plus sensible à l'action d'OmrA et d'OmrB dans la souche *Δrnc*. Cela étant dit, l'extrême abondance de ces deux sRNA dans le contexte *Δrnc* rend difficile l'interprétation de ces résultats sur l'ARNm *fepA*. Quoi qu'il en soit il est au moins possible de conclure que la RNase III n'est pas impliquée dans la régulation de la synthèse de FepA par les OmrA/B.

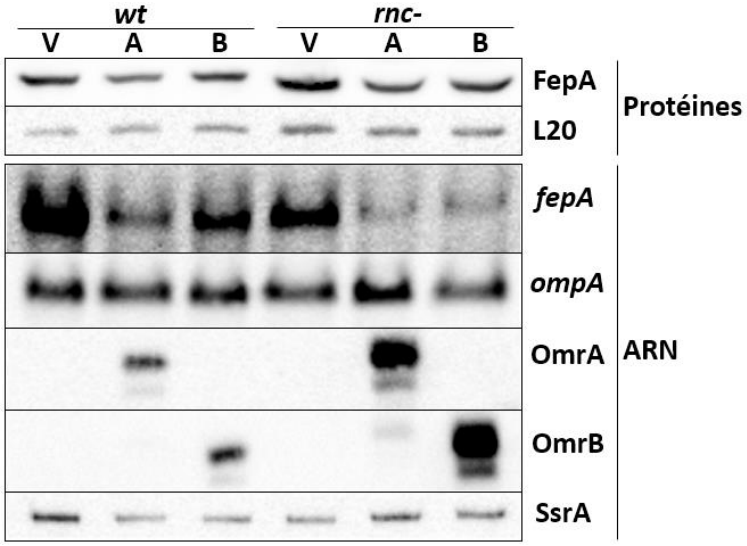


Figure 54 – La RNase III n'est pas nécessaire à la régulation de la synthèse de FepA par OmrA/B
Analyse des niveaux de la protéine FepA et de l'ARNm fepA. Les souches wt (MG1100) ou Δrnc (rnc-, MG1272) n'exprimant pas la RNase III ont été transformées par des plasmides contrôles (V pour vecteur) ou surproducteurs d'OmrA (A) ou d'OmrB (B). Les bactéries ont été cultivées en milieu LB à 37°C en présence de 100µM d'IPTG jusqu'à une OD_{600} de 0,4 puis les protéines et les ARN extraits. Les protéines ont été séparées par électrophorèse sur gel de polyacrylamide (PAGE pour « polyacrylamide gel electrophoresis ») en présence de SDS puis analysées par WB avec les anticorps monoclonaux α-FepA et α-L20. Les ARN ont été séparés soit par PAGE (OmrA, OmrB, SsrA) soit par électrophorèse sur gel d'agarose (AGE pour « agarose gel electrophoresis ») (fepA, ompA) puis analysés par « Northern Blot » (NB) avec des sondes spécifiques des ARN fepA, ompA, OmrA, OmrB et SsrA. L20, ompA et SsrA servent ici de contrôles de charge.

Cette dernière expérience revêt néanmoins un intérêt supplémentaire, justement du fait qu'elle révèle le grand impact de la RNase III sur les niveaux des sRNA OmrA et OmrB (Fig. 54). En effet, bien qu'ils soient surexprimés à partir d'un plasmide, ces sRNAs sont encore bien

plus abondants dans la souche *Δrnc*. Cela s'explique probablement par une plus grande stabilité des deux sRNA dans cette souche. On peut imaginer que la RNase III soit responsable de leur dégradation quand ils sont en interaction avec certaines cibles ARNm, ou bien lorsqu'ils sont très abondants dans la cellule et que leur niveau dépasse la capacité de prise en charge par Hfq. Ces effets sont surprenants dans la mesure où l'on s'attendrait plutôt à ce que ce soit la RNase E qui soit responsable de la dégradation d'OmrA/B, comme c'est le cas pour plusieurs autres sRNA Hfq-dépendants (Massé et al., 2003; Pfeiffer et al., 2009). Or les niveaux d'OmrA/B semblent insensibles à la déplétion de la RNase E, au moins en conditions de surproduction (voir **Article** Fig. 2). Ces résultats mériteraient donc une étude plus détaillée des mécanismes de dégradation de ces deux sRNA, à commencer par l'utilisation des mêmes mutations *Δrnc* et *rne*^{ts} pour l'observation des OmrA/B exprimés depuis leur locus chromosomique.

3. L'implication de Hfq *in vivo*

Certains résultats d'expériences de toeprinting avec les OmrA/B *in vitro* suggéraient la nécessité de la présence de la protéine chaperonne Hfq pour obtenir un appariement efficace et un effet spécifique sur la fixation du ribosome (voir **Article** Fig. 3.B). Toutefois, la formation d'un duplex suffisamment stable entre la structure SB de l'ARNm *fepA* et un ARN ou un oligonucléotide ADN semble être la seule condition requise pour obtenir un tel effet (voir **Article** Fig. 3 et 4). Nous désirions néanmoins savoir si la protéine Hfq était nécessaire à l'appariement sRNA/ARNm *in vivo*, et donc à l'action de ces sRNA. Cette question n'est pas facile à résoudre, dans la mesure où les niveaux des OmrA/B sont fortement diminués dans un mutant Δ*hfq* (Guillier and Gottesman, 2008; Holmqvist et al., 2010) (Fig. 55). Nous avons donc imaginé une expérience utilisant des souches Δ*hfq* portant également un allèle muté *rne*^{ts}, codant une RNase E thermosensible non fonctionnelle à 44°C. L'objectif était d'obtenir une souche où OmrA et OmrB seraient stables en l'absence de la chaperonne Hfq : en effet nous pensions alors que la RNase E serait la principale cause de dégradation d'OmrA/B en l'absence de Hfq, comme c'est le cas pour d'autres sRNA Hfq-dépendants. Nous avons par ailleurs utilisé comme contrôle des souches wt ou portant l'une ou l'autre des mutations, afin d'analyser l'effet d'une surexpression d'OmrA ou OmrB sur la synthèse de FepA (Fig. 55).

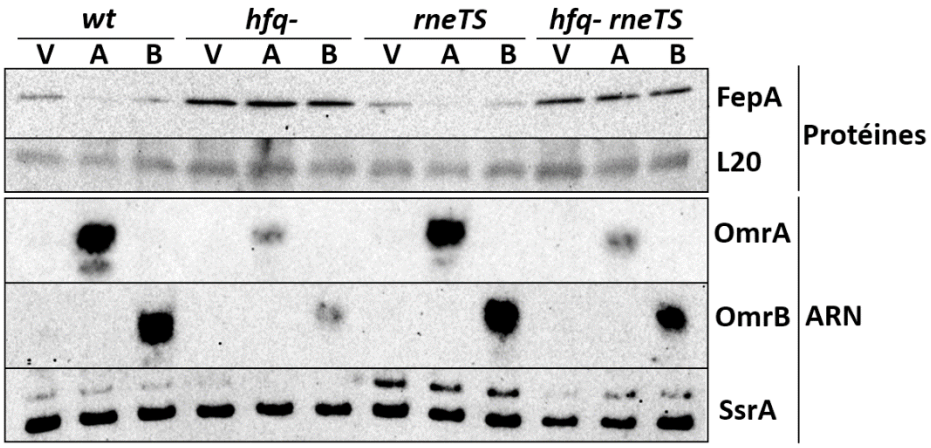


Figure 55 – Effets de Hfq et de la RNase E sur la régulation de la synthèse de FepA et sur OmrA/B
Analyse par WB des niveaux de la protéine FepA. Des souches wt (N3433), Δhfq (hfq-, IBPC953), rne^{ts} (N3431) ou Δhfq rne^{ts} (N3431x) ont été transformées par des plasmides contrôles (V) ou surproducteurs d'OmrA (A) ou d'OmrB (B). Les bactéries ont été cultivées en milieu LB à 30°C jusqu'à une OD_{600} de 0,3, puis 15 minutes supplémentaires à 44°C afin d'inactiver la RNase E thermosensible, et ce en présence de 250µM de 2,2'dipyridyl (dip), un chélateur de fer induisant l'expression de fepA, et de 100µM d'IPTG afin d'induire l'expression des sRNA. Les protéines et les ARN ont été ensuite extraits. Les protéines ont été séparées par SDS-PAGE puis analysées par WB avec les anticorps monoclonaux α-FepA et α-L20. Les ARN ont été séparés par PAGE (OmrA, OmrB, SsrA) puis analysés par NB avec des sondes spécifiques des ARN OmrA, OmrB et SsrA. L20 et SsrA servent ici de contrôles de charge.

Cette expérience permet d'observer à nouveau l'effet d'OmrA/B sur la synthèse de FepA en l'absence de la RNase E. Comme attendu les niveaux d'OmrA et OmrB baissent considérablement en l'absence de Hfq, alors même qu'ils sont surexprimés, et leur effet sur la synthèse de FepA disparaît dans ce contexte. De surcroît, la combinaison des mutations *Δhfq* et *rne^{ts}* ne permet pas de rétablir le contrôle de la synthèse de FepA par les OmrA/B, et cela s'explique notamment par le fait que ces deux sRNA ne retrouvent pas leur niveaux d'origine. Plus précisément, le niveau d'OmrA reste le même que dans un mutant simple *Δhfq*, tandis qu'OmrB est plus abondant sans l'être autant que dans une souche wt.

En résumé avec les expériences réalisées sur la RNase III, les OmrA/B semblent bien plus sensibles à la présence de la RNase III qu'à celle de la RNase E, bien qu'OmrB soit également sensible à la RNase E au moins en l'absence de Hfq. Aussi, si une expérience sur l'implication de Hfq *in vivo* devait être renouvelée, il serait certainement plus judicieux d'utiliser une mutation *Δhfq* en combinaison avec la mutation *Δrnc* plutôt que la mutation *rne^{ts}* utilisée ici. Une dernière conclusion intrigante de cette expérience est l'effet important de la délétion du gène *hfq* sur la synthèse de FepA. Cette observation suggère que d'autres sRNA Hfq-dépendants pourraient réguler l'expression de *fepA*, et nous a conduits aux expériences et résultats qui sont présentés dans le chapitre suivant.

Chapitre II : À la croisée des mécanismes et des signaux amenant à la régulation de l'expression de *fepA*

A. Quatre sRNA Hfq-dépendants régulent la synthèse de FepA

Une partie importante de ma thèse a porté sur la découverte puis la compréhension de la régulation de l'expression de *fepA* par deux autres sRNA, à savoir SdsR et RseX. Je débuterai ce chapitre par la mise en évidence de cette régulation, avant de poursuivre sur le rôle du 5'UTR de *fepA* puis sur les mécanismes de régulation par SdsR puis RseX. La plupart des expériences présentées dans cette partie sont préliminaires et demandent à être reproduites, c'est pourquoi leur interprétation demandera davantage de précautions.

1. La régulation de l'expression de *fepA* par SdsR et RseX

a. Démonstration de la régulation

Du fait de la précédente observation suggérant que l'expression de *fepA* pouvait être régulée non seulement par OmrA/B mais aussi par d'autres sRNA Hfq-dépendants (Fig. 55), j'ai entrepris d'en identifier certains. Pour ce faire, j'ai utilisé une banque de plasmides surexprimant la plupart des sRNA Hfq-dépendants connus à ce jour (Mandin and Gottesman, 2010), complétée par un plasmide surproduisant le sRNA McaS, qui a été identifié plus récemment (Jorgensen et al., 2013). Une souche portant une fusion traductionnelle P_{tet}-*fepA*$_{-173}$-*lacZ* (voir Fig. 51.A) a été transformée par chacun de ces plasmides pour observer l'effet des différents sRNA sur la synthèse de la protéine de fusion (Fig. 56).

À l'observation de ces résultats, au moins quatre sRNA semblent impliqués dans la régulation de l'expression de *fepA*, à savoir OmrA et OmrB comme attendu, mais aussi RseX et SdsR. Ces deux sRNA exercent une répression plus importante qu'OmrA/B sur l'expression de *fepA.* La répression par SdsR atteint même un facteur 37,5, ce qui est rare pour une régulation exercée par un sRNA Hfq-dépendant, y compris dans des conditions de surproduction. D'autres sRNA semblent avoir un effet plus ténu, à l'image des répresseurs SgrS et RprA, ou de l'activateur OxyS. Notons que pour OxyS, l'effet positif observé pourrait

être médié par la répression de *rpoS* par ce sRNA (Zhang, 1998), qui se répercuterait à la fois sur les niveaux d'OmrA et de SdsR (voir **Introduction III.B.1.b et III.B.2.b**).

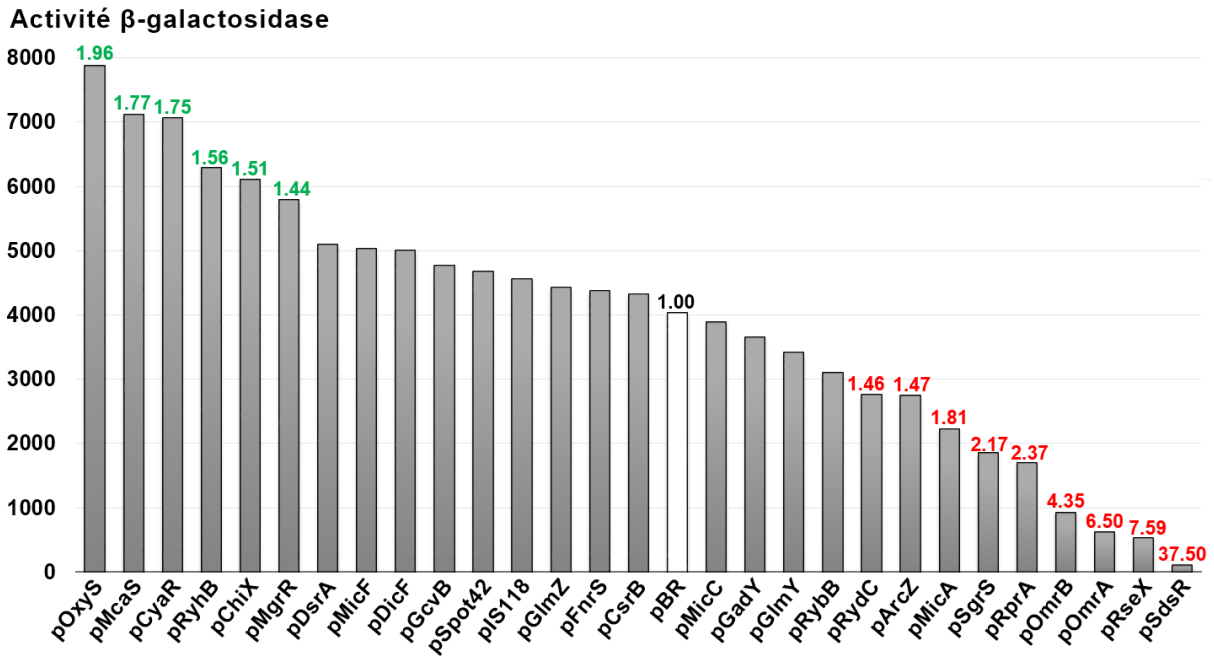


Figure 56 – Plusieurs sRNA Hfq-dépendants ont un effet sur la synthèse de FepA
Activité β-galactosidase spécifique de la fusion fepA$_{-173}$-lacZ. Une souche portant la fusion Ptet-fepA$_{-173}$-lacZ (MG1772) a été transformée par une banque de plasmides surproducteurs de divers sRNA Hfq-dépendants ou d'un plasmide contrôle (pBR). Les facteurs d'activation ou de répression par rapport au contrôle pBR sont respectivement indiqués en vert ou en rouge, au-dessus des histogrammes, quand ils sont supérieurs à 1,3.

Nous avons choisi de nous focaliser sur les régulateurs RseX et SdsR nouvellement découverts, dont les effets étaient les plus évidents, tout en poursuivant notre étude de la régulation par OmrA/B. Afin de reproduire ces résultats et de savoir si la régulation par ces sRNA dépendait de la présence de certaines régions du 5'UTR de *fepA*, nous avons choisi d'utiliser des fusions *fepA-lacZ* en raccourcissant progressivement le 5' UTR (Fig. 57). On notera que les 5'UTR débutant aux positions -173 et -146 correspondent aux deux TSS publiés de l'ARNm *fepA* (Hunt et al., 1994; Pettis et al., 1988).

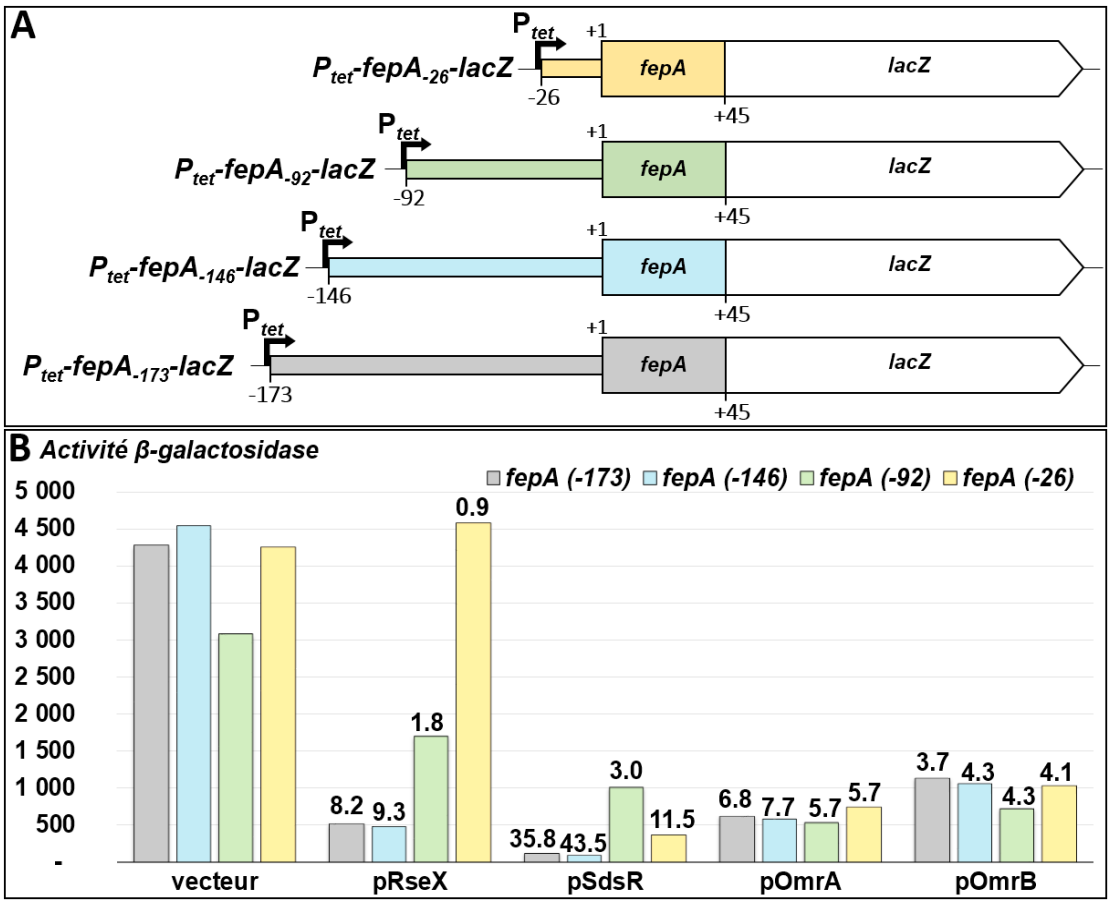


Figure 57 – La régulation par RseX et SdsR dépend du 5'UTR de l'ARNm fepA
A. Fusions fepA-lacZ. Ces fusions contiennent soit la totalité du 5'UTR de fepA (fepA-173-lacZ), soit une fraction du 5'UTR débutant à la position indiquée (-146, -92 et -26). B. Activité β-galactosidase spécifique des fusions $fepA_{-173}$-lacZ, $fepA_{-146}$-lacZ, $fepA_{-92}$-lacZ et $fepA_{-26}$-lacZ (respectivement souches MG1772, JJ0389, JJ0136 et JJ0135 transformées par un plasmide contrôle (vecteur) ou surproducteur de RseX, SdsR, OmrA ou OmrB). Les facteurs de répression par rapport au vecteur contrôle sont indiqués.

Comme attendu d'après les résultats de notre article, la longueur du 5'UTR n'a aucune influence sur la régulation par OmrA/B (voir **Article** Fig. 1). En ce qui concerne SdsR, la très forte répression sur la fusion *$fepA_{-173}$-lacZ* est reproduite ici avec un facteur 35,8, et s'observe également avec un 5'UTR légèrement plus court (fusion *$fepA_{-146}$-lacZ*) avec un facteur 43,5. Alors que la majeure partie de cette régulation est perdue avec un 5'UTR de 92nt (fusion *$fepA_{-92}$-lacZ*) avec un facteur 3 de répression seulement, on retrouve une répression d'un facteur 11,5 dans le cas du 5'UTR le plus court (fusion *$fepA_{-26}$-lacZ*). Cela pourrait signifier qu'une structure ou une autre caractéristique particulière à la fusion *$fepA_{-92}$-lacZ* empêche la régulation par SdsR. Quoi qu'il en soit, la longueur du 5'UTR de l'ARNm de *fepA* semble avoir un effet sur la régulation par SdsR, et l'on note que seules les deux fusions commençant à un TSS connu de *fepA* sont parfaitement régulées par ce sRNA. S'il s'agit d'une régulation directe

via une interaction sRNA/ARNm, il faudrait sans doute chercher une région de fixation située entre les nt -26 et +45 de l'ARNm *fepA*, puisqu'une forte régulation est toujours observée pour la fusion *fepA*$_{-26}$-*lacZ*. Cela étant dit, ces résultats indiquent également un rôle important de la région située entre les positions -146 et -92 de cet ARNm, qu'il faudra analyser attentivement. Le mécanisme de régulation par SdsR a été étudié plus en détail par la suite (voir **Résultats II.A.2.b**).

Enfin, le cas de RseX est légèrement différent puisque cette fois-ci seules les deux fusions commençant à un TSS connu de *fepA* (*fepA*$_{-173}$-*lacZ* et *fepA*$_{-146}$-*lacZ*) sont efficacement régulées par ce sRNA (Fig. 57). Il est donc probable que si ce sRNA interagit avec l'ARNm *fepA*, le site de fixation se situe entre les nt -146 et -92 du 5'UTR de *fepA*, voire éventuellement légèrement en aval du nt -92 si l'on considère que la fusion *fepA*$_{-92}$-*lacZ* est encore régulée d'un facteur 1,8. On pourrait en effet envisager que la réduction de l'effet de RseX dans cette fusion soit due à l'absence d'une partie du site de fixation de RseX. Le mécanisme de régulation de *fepA* par RseX a également été abordé plus en détail (voir **Résultats II.A.2.c**).

En conclusion, au moins quatre sRNA Hfq-dépendants régulent l'expression de *fepA*, et certaines régions du 5'UTR situées en amont du nt -92 de l'ARNm *fepA* sont importantes voire nécessaires à la régulation par SdsR et RseX. En revanche, la régulation par OmrA/B ne requiert pas la présence de la majeure partie du 5'UTR de *fepA*.

b. Le rôle du 5'UTR de l'ARNm *fepA* et l'effet d'une délétion des sRNA

Avant de poursuivre avec les mécanismes de régulation de l'expression de *fepA* par RseX et SdsR, je m'attarderai sur l'analyse du 5'UTR de *fepA*. En effet, lorsque nous nous sommes aperçus que la longueur du 5'UTR de *fepA* était importante pour la régulation par RseX et SdsR, nous avons voulu déterminer la taille du 5'UTR des transcrits *fepA in vivo.*

J'ai donc réalisé des expériences d'allongement d'amorce par transcription inverse sur des ARNm extraits à diverses phases de croissance, à partir de souches délétées du gène *fur* afin de permettre une expression constitutive de *fepA* (Fig. 58.A). L'ensemble des gènes *omrA/B*, *rseX* et *sdsR* ont également été délétés dans certaines souches (dénommées *Δ4*), dans le but de permettre une expression maximale de *fepA*, mais aussi pour révéler

d'éventuelles formes de *fepA* qui pourraient résulter d'un clivage dépendant de l'un de ces sRNA. Ces expériences sont accompagnées de la détection de la protéine FepA par WB et de l'ARNm *fepA* par NB, afin d'obtenir plus de détails sur l'expression de *fepA* dans ces conditions (Fig. 58.B).

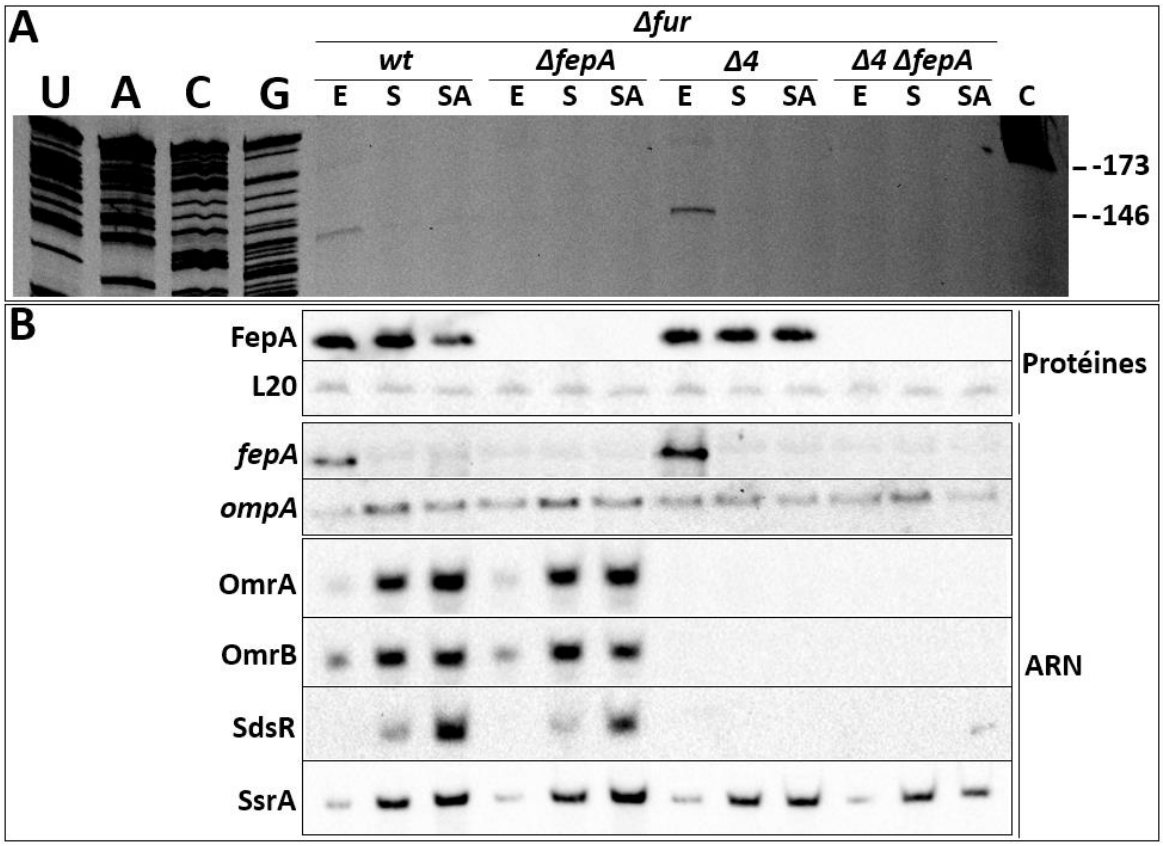


Figure 58 – Les deux formes du 5'UTR de fepA et l'effet des sRNA sur la synthèse de FepA
A. Allongement d'amorce sur des ARNm extraits de souches Δfur (MG1169), Δfur ΔfepA (JJ0171), Δfur Δ4 (JJ0188) ou Δfur Δ4 ΔfepA (JJ0199) cultivées en milieu LB jusqu'à une OD_{600} de 0,4 (E, phase exponentielle), une OD_{600} de 2,0 (S, phase stationnaire), ou cultivées pendant 3h supplémentaires après avoir atteint une OD_{600} de 2,0 (SA, phase stationnaire avancée). La réaction contrôle (C) a été réalisée sur un fragment d'ARNm fepA des nt -173 à +129, transcrit in vitro. Les principaux signaux fepA-dépendants sont indiqués (-173 et -146). Note : 5µg d'ARN ont été utilisés pour les RT « E » et « S » et 3,5µg pour les RT « SA ». B. Analyse des protéines et des ARN extraits à partir des mêmes cultures qu'en (A). Les protéines ont été séparées par SDS-PAGE puis analysées par WB. Les ARN ont été séparés soit par PAGE (OmrA, OmrB, SdsR et SsrA), soit par AGE (fepA, ompA) puis analysés par NB. RseX est indétectable sur ce NB. La protéine L20 (bande aspécifique reconnue par l'anticorps L20 ici) et les ARN ompA et SsrA servent de contrôles de charge.

Les allongements d'amorce sur l'ARNm *fepA* confirment l'existence de deux formes du 5'UTR, correspondant très vraisemblablement aux TSS publiés de *fepA*, longs respectivement de 173 et 146 nt (Fig. 58.A), bien que la forme à 146 nt semble de loin la plus abondante. Le fait que le 5'UTR de 146nt semble majoritaire *in vivo* suggère que le TSS décrit à la position -146 est le plus important pour l'expression de *fepA*.

Par ailleurs, et malgré la délétion de *fur,* on n'observe les diverses formes de l'ARNm *fepA* qu'en phase exponentielle de croissance que ce soit par allongement d'amorce ou par NB (Fig. 58). Cela pourrait indiquer que l'expression de *fepA* est régulée en phase stationnaire de croissance. Or cette phase stationnaire de croissance correspond justement aux maxima d'expression des sRNA OmrA, OmrB et SdsR ; aussi serait-il tentant d'attribuer cette régulation à ces sRNA. Cependant, le fait que l'ARNm *fepA* ne soit plus détectable en phase stationnaire y compris en contexte *Δ4* exclut une telle hypothèse. Cela suggère donc qu'en outre des régulations exercées par Fur, OmrA/B, SdsR et RseX, d'autres régulations affectent fortement les niveaux de cet ARN en fonction de la phase de croissance. On notera toutefois que les contrôles de charge *ompA* et SsrA n'étaient peut-être pas les plus appropriés pour cette expérience de NB, puisque leurs niveaux changent distinctement en fonction de la phase de croissance. Néanmoins la diminution du niveau de l'ARNm *fepA* va à contresens de celui de SsrA, tandis que l'accumulation d'OmrA/B et de SdsR en phase stationnaire a déjà été relevée dans la littérature (voir **Introduction III.B.2**), ce qui conforte les conclusions tirées de cette expérience de NB.

Par ailleurs, si le niveau de la protéine FepA reste stable jusqu'en phase stationnaire, il diminue après une phase stationnaire prolongée (Fig. 58.B). Cette fois-ci l'effet est clairement associé aux sRNA, dans la mesure où il disparaît dans le contexte *Δ4*. De même, l'ARNm *fepA* est plus abondant dans la souche *Δ*4 en phase exponentielle de croissance, ce qui suggère un effet des sRNA y compris dans cette phase de croissance où ils sont faiblement exprimés. Ces effets sur les niveaux de l'ARNm *fepA* et de la protéine FepA montrent que l'expression de ces sRNA à partir de leur locus est suffisante pour avoir un effet sur l'expression du gène *fepA*. Étant donné que le sRNA RseX est indétectable dans les conditions de cette expérience, ces effets sont très certainement principalement imputables à l'action d'OmrA, d'OmrB, de SdsR ou bien encore d'une combinaison de plusieurs de ces sRNA, ce qu'il conviendrait de confirmer avec des mutants de ces sRNA.

L'ensemble de ces expériences confirme donc l'existence de deux formes de l'ARNm *fepA in vivo* et révèle l'importance du 5'UTR de *fepA* pour la régulation par SdsR et RseX. De plus, la régulation par OmrA/B, SdsR et RseX semble suffisamment forte pour avoir un impact sur les niveaux de l'ARNm *fepA* et de la protéine FepA. Enfin, l'ARNm *fepA* n'est plus

détectable en phase stationnaire, ce qui suggère l'existence d'une régulation en fonction de la phase de croissance indépendante de Fur et des quatre sRNA discutés ici.

2. Les mécanismes de régulation de *fepA* par SdsR et RseX

a. Eléments mécanistiques communs

Après la découverte de la régulation de *fepA* par deux sRNA Hfq-dépendants supplémentaires, nous avons voulu déterminer s'il s'agissait d'une régulation directe, impliquant des appariements entre ces sRNA et l'ARNm *fepA*. Etant donné que cette régulation a été observée sur des fusions traductionnelles exprimées à partir d'un promoteur constitutif, l'intervention d'un régulateur transcriptionnel intermédiaire (tel que Fur) qui serait ciblé par l'un ou l'autre de ces sRNA est exclue. En revanche, ces sRNA pourraient tout à fait agir sur une protéine ou un ARN intermédiaire régulant *fepA* au niveau post-transcriptionnel tels que les OmrA/B, aboutissant *in fine* à la régulation de l'expression de *fepA*. Dans un premier temps nous avons donc voulu savoir si la régulation par chacun de ces quatre sRNA était indépendante. Pour OmrA/B les régulations sont directes et liées à la structure SB (voir **Article**), aussi nous avons considéré qu'il était peu probable que la régulation par OmrA/B nécessite l'intervention de RseX et/ou de SdsR. Afin de démontrer que la régulation par RseX et SdsR était également indépendante, j'ai mesuré l'activité β-galactosidase d'une fusion *fepA*$_{-173}$-*lacZ* dans une souche Δ4 lors de la surexpression de chacun de ces deux sRNA (Fig. 59).

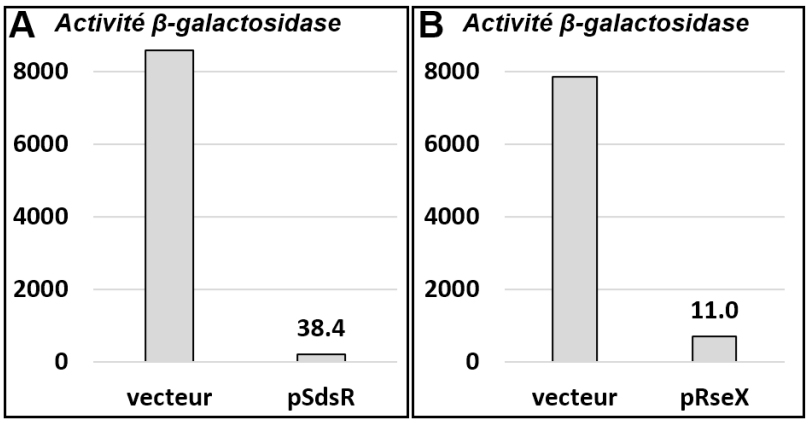


Figure 59 – La régulation de fepA par RseX ou SdsR est indépendante de l'effet d'autres sRNA
Activités β-galactosidase spécifiques de la fusions fepA$_{-173}$*-lacZ en contexte Δ4 (souche JJ0167 transformée par un plasmide contrôle (vecteur) ou surproducteur de RseX (A) ou SdsR (B)). Les facteurs de répression par rapport au vecteur contrôle sont indiqués.*

Il s'agit d'une expérience préliminaire qui demande à être reproduite, en incluant comme contrôles les effets de RseX et SdsR dans une souche wt, et idéalement la même expérience avec des plasmides surproducteurs d'OmrA et d'OmrB. Ce résultat semble toutefois exclure que la régulation par RseX ou par SdsR nécessite la présence des trois autres sRNA régulateurs de l'expression de *fepA*. La régulation exercée par chacun de ces quatre sRNA serait donc tout à fait indépendante.

Nous avons voulu aller plus loin dans notre étude et mesurer l'effet de SdsR et de RseX sur l'expression de *fepA* d'un point de vue cinétique. En effet, les expériences de β-galactosidase sont réalisées à l'équilibre, et l'accumulation des protéines de fusion, du fait de leur grande stabilité, ne permet pas la mesure du délai s'écoulant entre l'induction du sRNA et l'observation d'une régulation du gène cible. Or, cette donnée pourrait nous indiquer s'il s'agit d'une régulation directe ou bien d'une régulation indirecte transmise par un régulateur intermédiaire entre le sRNA et ce gène cible. J'ai donc suivi les niveaux de l'ARNm *fepA* dans une souche Δ*fur*, pendant les 8 minutes suivant l'induction de RseX, de SdsR ou d'OmrA comme contrôle (Fig. 60).

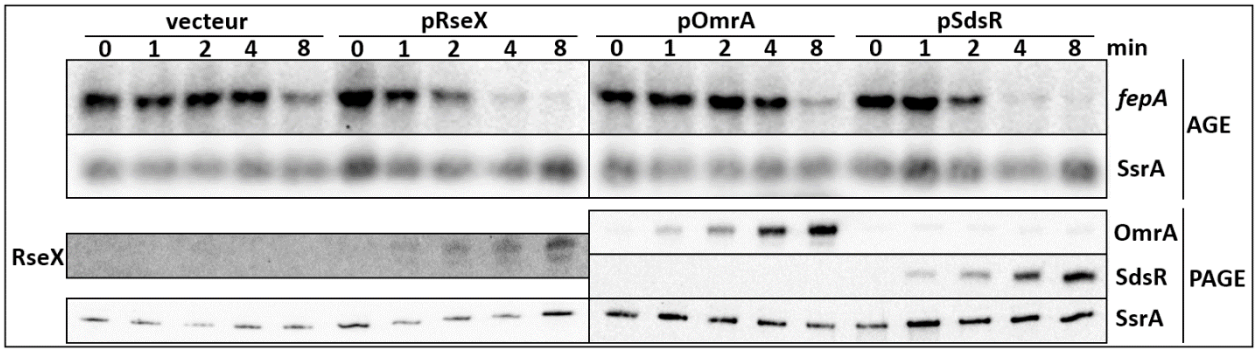


Figure 60 – RseX et SdsR ont un effet très rapide sur l'ARNm fepA
Analyse des niveaux de l'ARNm fepA au cours d'une cinétique d'induction des sRNA RseX, OmrA et SdsR. Une souche Δfur (MG1169) a été transformée par un plasmide contrôle (vecteur) ou surexprimant RseX, OmrA ou SdsR à partir d'un promoteur inductible à l'IPTG. Les bactéries ont été cultivées jusqu'à 37°C en milieu LB jusqu'à une OD_{600} de 0,4, puis l'expression des sRNA a été induite par l'addition d'IPTG à 100µM. Les ARN ont été extraits juste avant l'induction (0min) et aux temps indiqués après induction (1, 2, 4 et 8 min), puis séparés soit par PAGE (OmrA, RseX, SdsR, SsrA), soit par AGE (fepA, SsrA) avant d'être analysés par NB. SsrA sert de contrôle de charge.

Avant d'analyser le résultat de cette expérience, deux constats importants peuvent être dressés. Premièrement, dans le contrôle sans sRNAs (vecteur) on remarque une diminution du niveau de l'ARNm *fepA* huit minutes après l'addition de l'inducteur, ce qui pourrait signifier que l'on atteint un stade de croissance où le niveau de l'ARNm *fepA*

commence à décroître (comme observé auparavant, voir Fig. 58). En second lieu, le niveau de l'ARNm *fepA* ne semblent pas affecté par l'induction d'OmrA, ce qui ne correspond pas aux résultats présentés dans l'article, ni aux résultats des précédentes expériences. Cela étant dit, l'effet observé dans (Guillier and Gottesman, 2006) apparaît au bout de cinq minutes d'induction d'OmrA, c'est-à-dire entre les points à quatre minutes et huit minutes réalisés dans notre expérience, ce qui peut en partie expliquer ce résultat. Il conviendrait néanmoins de réitérer cette expérience qui n'a peut-être pas été réalisée correctement ici, en choisissant des conditions où les niveaux de l'ARNm *fepA* ne changent pas en présence du vecteur contrôle, et en se plaçant éventuellement dans une souche délétée pour les gènes des quatre sRNA régulateurs de l'expression de *fepA*.

En ce qui concerne les sRNA RseX et SdsR, il est surprenant de constater que leurs effets sur les niveaux de l'ARNm *fepA* semblent intervenir très rapidement – de l'ordre d'une minute pour les premiers effets de RseX et de deux minutes pour SdsR. S'il ne s'agit pas d'une démonstration de l'interaction directe de RseX et SdsR avec l'ARNm *fepA*, cet effet semble trop rapide pour pouvoir être expliqué par une régulation indirecte *via* d'autres régulateurs post-transcriptionnels. Dans l'ensemble, ces résultats constituent donc une première indication de l'effet direct de RseX et de SdsR sur l'expression de *fepA*.

b. Le mécanisme de régulation de l'expression de *fepA* par SdsR

La régulation de l'expression de *fepA* par SdsR est particulièrement étonnante du fait du facteur de répression important exercé par ce sRNA, dont une partie semble dépendre de la longueur du 5'UTR de *fepA* (Fig. 57). Une fois cette régulation caractérisée, j'ai cherché si des prédictions d'appariements entre SdsR et l'ARNm *fepA* étaient possibles grâce au programme informatique IntaRNA (Busch et al., 2008). Au moins deux prédictions d'appariement sont effectivement proposées par ce programme, impliquant à chaque fois la région couvrant les nt 15 à 40 de SdsR qui s'apparie aux cibles directes connues de ce sRNA (Fig. 61.A). La prédiction qui nous a semblé la plus convaincante fait intervenir une région située à proximité du SD de *fepA*, suggérant un mécanisme classique de compétition entre le sRNA et la sous-unité 30S du ribosome pour la fixation au RBS qui pourrait expliquer le fort facteur de répression de *fepA* par SdsR. C'est pourquoi nous avons voulu étudier cet appariement hypothétique de prime abord, en introduisant des mutations compensatoires

dans les deux partenaires. Les mutations *fepAmutSR* et SdsRmut ont donc été conçues afin d'abolir l'éventuel appariement au niveau du RBS de *fepA* (Fig. 61.A). La combinaison de ces mutations devrait permettre le rétablissement de l'appariement et donc de la régulation par SdsR si un appariement a bien lieu dans cette région de l'ARNm *fepA*.

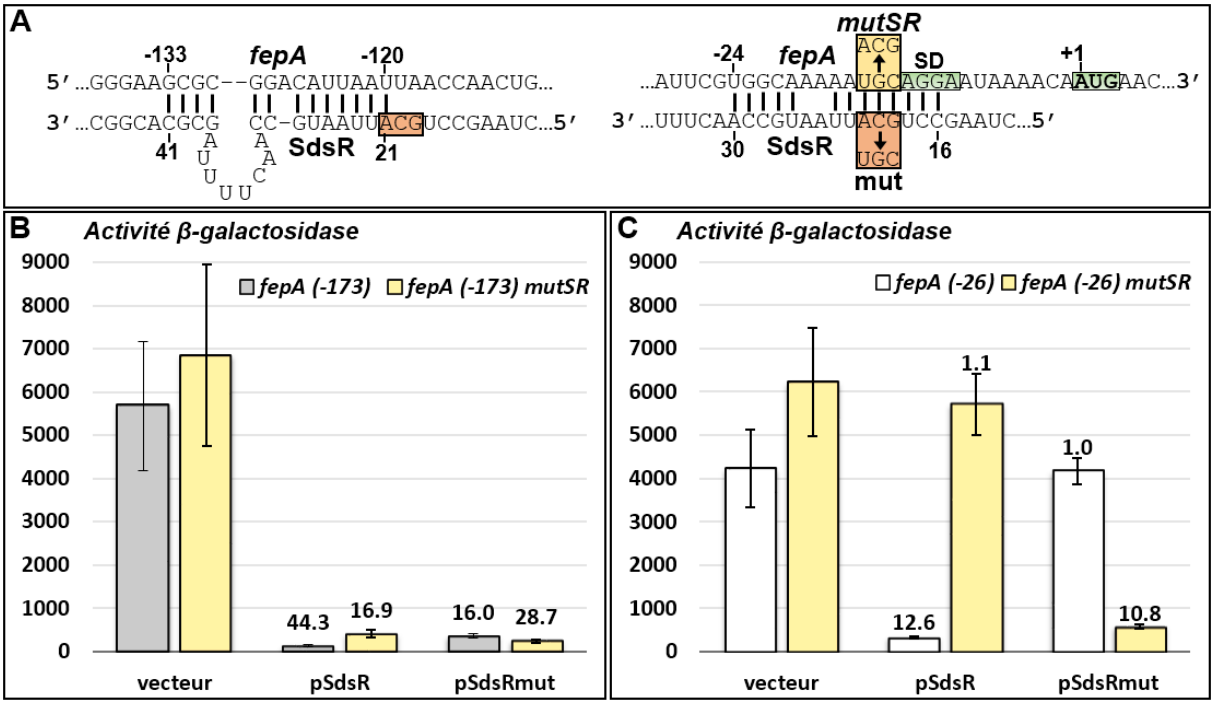


Figure 61 – SdsR s'apparie au RBS de fepA, et probablement également plus en amont au sein de son 5'UTR

A. Prédictions d'appariement entre SdsR et l'ARNm fepA obtenues in silico grâce au programme intaRNA. Les régions du RBS importantes pour le démarrage de la traduction sont encadrées en vert et les positions délimitant l'appariement sont indiquées. Les mutations introduites dans l'ARNm fepA et dans le sRNA SdsR sont encadrées respectivement en jaune et en orange : ces mutations ont été conçues pour n'affecter que l'appariement prédit à proximité de la séquence SD de fepA. B. Activités β-galactosidase spécifiques des fusions fepA$_{-173}$-lacZ (souche MG1772) et fepA$_{-173}$-mutSR-lacZ (souche JJ0141). Les bactéries ont été transformées par un plasmide contrôle (vecteur) ou surproducteur de SdsR ou SdsRmut). C. Activités β-galactosidase spécifiques des fusions fepA$_{-26}$-lacZ (souche JJ0135) et fepA$_{-26}$-mutSR-lacZ (souche JJ0146). Les bactéries ont été transformées par un plasmide contrôle (vecteur) ou surproducteur de SdsR ou SdsRmut. Les facteurs de répression par rapport au vecteur contrôle sont indiqués pour les panneaux B. et C.

Dans un premier temps, l'effet de ces mutations compensatoires a été analysé sur la fusion *fepA$_{-173}$-lacZ* comprenant l'intégralité du 5'UTR de *fepA* et permettant les deux appariements prédits entre SdsR et *fepA* (Fig. 61.B). Si l'on constate une différence entre les facteurs de répression par SdsR sur la fusion wt et mutSR (respectivement des facteurs de répression de 44,3 et 16,9), qui est compensée lors de l'usage du sRNA mutant (respectivement des facteurs de répression de 16,0 et de 28,7), un facteur de répression très important persiste quand un seul des partenaires est muté.

Nous avons voulu savoir si cette répression résiduelle importante était due à une région située en amont du 5'UTR de *fepA*, par exemple la région s'étendant des nt -133 à -120 qui forment le second appariement prédit (Fig. 61.A), lequel n'est *a priori* pas affecté par les mutations introduites ici. J'ai donc effectué la même expérience sur la fusion *fepA*$_{-26}$-*lacZ* (Fig. 61.C), pour laquelle seul l'appariement prédit au niveau du RBS de *fepA* est possible. Cette fois-ci, alors que SdsR réprime d'un facteur 12,6 l'expression de la fusion wt, toute régulation de l'expression de la fusion mutSR est perdue. De plus, une répression de l'expression de cette dernière fusion d'un facteur 10,8 est retrouvée avec SdsRmut, qui ne régule pas l'expression de la fusion wt. Cette expérience démontre donc l'interaction entre SdsR et l'ARNm *fepA* au niveau du SD, qui est responsable d'une part importante de la régulation de l'expression de *fepA* par SdsR. Cette régulation implique toutefois une seconde composante, qui pourrait correspondre au second appariement prédit. L'existence de ce second appariement reste à démontrer, mais elle est déjà soutenue par des expériences préliminaires que je ne développerai pas ici. Ce double appariement potentiel soulève des questions intéressantes qui seront abordées à la fin de cette partie dédiée à SdsR.

L'existence d'un tel appariement au niveau du RBS de *fepA* suggère que le mécanisme de régulation le plus probable soit une inhibition de la fixation du ribosome. J'ai néanmoins réalisé des expériences de toeprint afin d'en acquérir la certitude. Ces expériences ont été menées dans un premier temps avec le sRNA SdsR wt ou mutant, mais à l'image des toeprints avec les OmrA/B, des problèmes de spécificité m'ont conduit à utiliser l'oligonucléotide ADN DoSdsR (pour « DNA oligonucleotide SdsR ») imitant SdsR (Fig. 62). Il se pourrait par ailleurs que ce problème de spécificité puisse être résolu par l'usage de la protéine Hfq, comme cela a été fait avec OmrA/B (voir **Article** Fig. 3). L'expérience de toeprint présentée ici doit être réitérée dans la mesure où le contrôle avec la sous-unité 30S du ribosome seule n'a pas fonctionné correctement (Fig. 62.B puits n°2). Toutefois, l'oligonucléotide DoSdsR s'apparie de manière spécifique à la région attendue de l'ARNm *fepA*, et semble prévenir la fixation de la sous-unité 30S du ribosome si l'on s'en tient à la comparaison avec l'absence d'effet de DoSdsRmut.

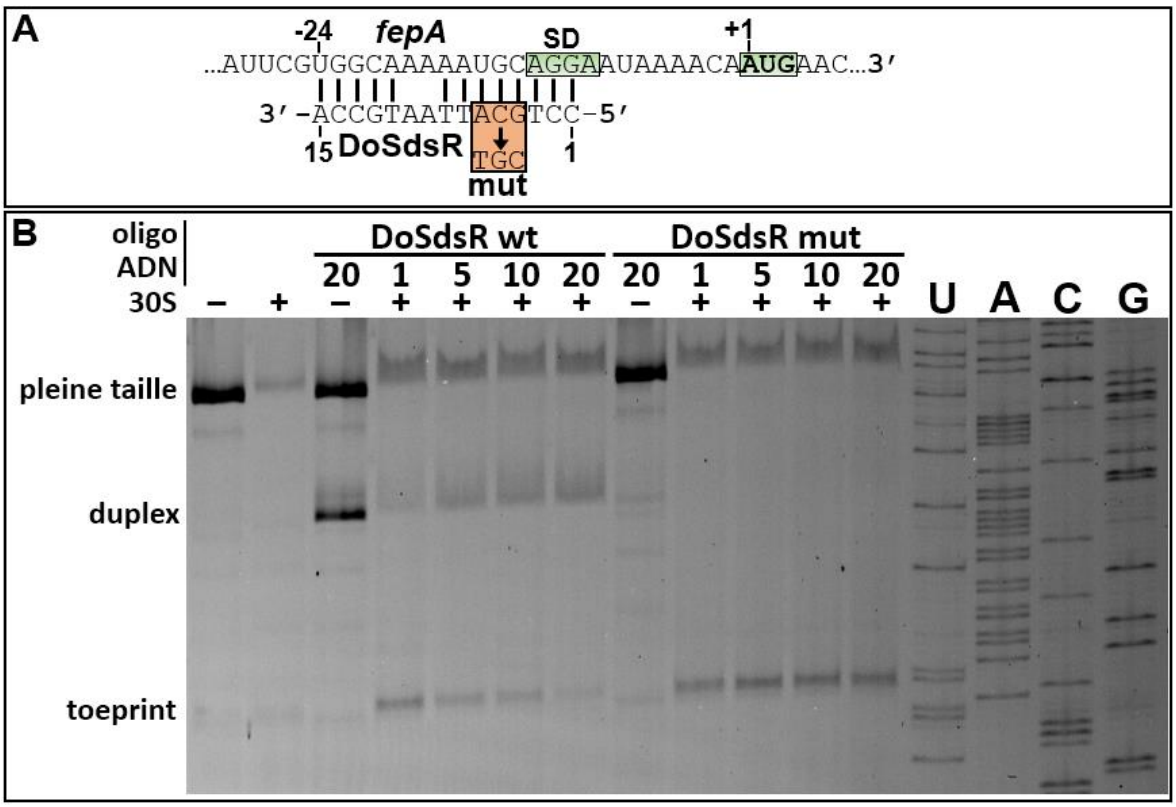


Figure 62 – L'interaction de SdsR avec le RBS de l'ARNm fepA inhibe la fixation du ribosome
A. Appariement entre l'oligonucléotide ADN DoSdsR et l'ARNm fepA. B. Expérience de toeprint sur le fragment d'ARNm fepA -26 à +129, en présence de DoSdsR wt ou mutant dans les ratios oligonucléotide : mRNA indiqués.

Cette expérience résout donc en partie la question du mécanisme de régulation de *fepA* par SdsR. Néanmoins, elle n'exclut pas qu'une partie de l'effet de *fepA* puisse être due au recrutement de la RNase E comme c'est le cas pour la régulation d'*ompD* par SdsR chez *S. enterica* (Frohlich et al., 2012), ou encore de la RNase III comme cela pourrait être le cas pour la régulation de *mutS* par SdsR chez *E. coli* (Gutierrez et al., 2013). Par ailleurs, l'existence probable du second site d'interaction plus en amont du 5'UTR de *fepA* pourrait tout à fait impliquer un second mécanisme de régulation par SdsR, combinant ou pas l'effet d'une double interaction. En effet, il est possible d'envisager qu'une molécule de sRNA interagisse partiellement avec les deux sites, ou encore que le sRNA interagisse consécutivement avec les deux sites. Alternativement, deux molécules de sRNA pourraient chacune interagir avec l'un des sites, entraînant un effet synergique sur l'expression de *fepA* qui pourrait expliquer la force inattendue de la répression par SdsR. Quoi qu'il en soit, à l'heure actuelle le mécanisme de régulation impliquant l'éventuel second site d'appariement demeure inconnu et demande de nouvelles expérimentations.

c. Le mécanisme de régulation de l'expression de *fepA* par RseX

Comme pour SdsR, nous avons cherché une prédiction d'appariement entre l'ARNm *fepA* et RseX. La première prédiction obtenue (Fig. 63.A en haut) grâce au logiciel IntaRNA (Busch et al., 2008) montre un appariement imparfait sur plus d'une trentaine de nt au sein du 5'UTR de *fepA* et éloigné du RBS d'au moins 30nt. Cette prédiction a été la première envisagée, aussi des mutations compensatoires ont été conçues à la manière de ce qui a été fait avec SdsR et avec les OmrA/B et utilisées dans des expériences préliminaires (Fig. 63). Il est à noter ici que la région de RseX impliquant la mutation RseXmut2 est également celle proposée pour l'interaction avec *ompA* et *ompC* (Douchin et al., 2006).

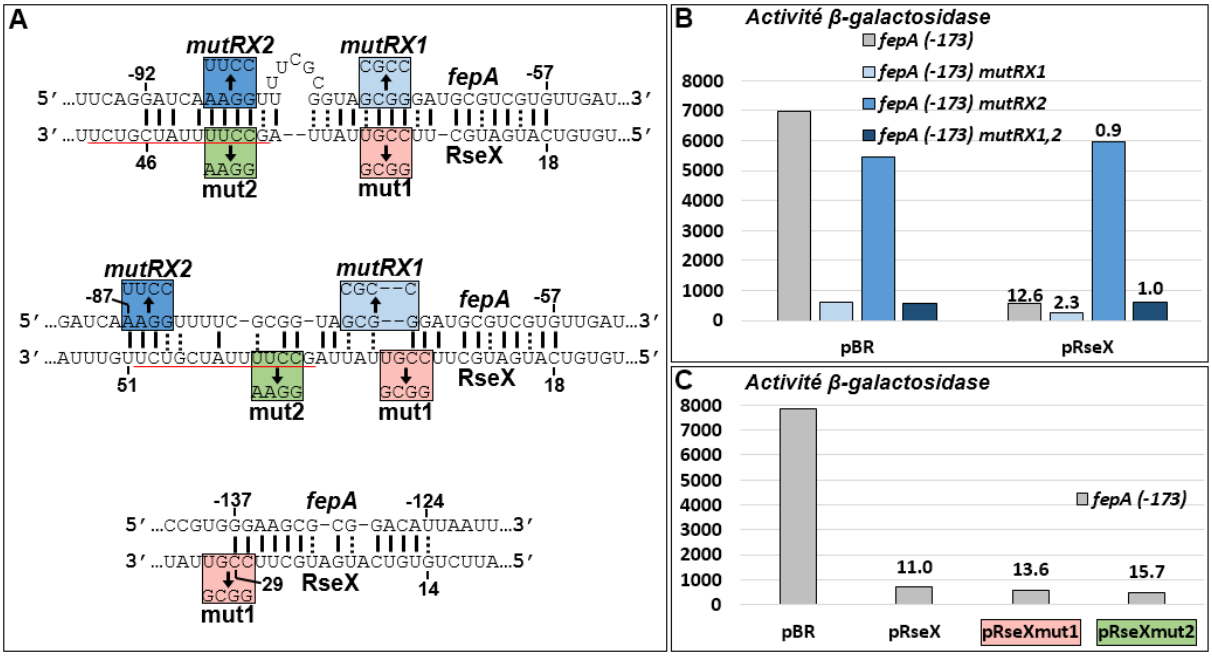


Figure 63 – L'appariement entre RseX et l'ARNm fepA reste à démontrer
A. Prédictions d'appariement entre RseX et la cible ARNm fepA fournies in silico par le programme intaRNA. Les nt soulignés en rouge correspondent à la région de RseX qui pourrait s'apparier aux ARNm ompA et ompC d'après (Douchin et al., 2006). B. Activités β-galactosidase spécifiques des fusions fepA$_{-173}$-lacZ et ses dérivées mutRX1, mutRX2 et mutRX1,2 (respectivement souches MG1772, JJ0138, JJ0147 et JJ0148). La mutation mutRX1,2 combine les deux mutations mutRX1 et mutRX2 indiquées en A. Les bactéries ont été transformées par un plasmide contrôle (vecteur) ou surproducteur de RseX. C. Activités β-galactosidase spécifiques de la fusion fepA$_{-173}$-lacZ en contexte Δ4 (souche JJ0167 transformée par un plasmide contrôle (vecteur) ou surproducteur de RseX, RseXmut1 ou RseXmut2). Les facteurs de répression par rapport au vecteur contrôle sont indiqués pour les panneaux B et C.

L'effet de RseX sur des mutants de la fusion *fepA$_{-173}$-lacZ* a été analysé de prime abord (Fig. 63.B). On constate que la mutation mutRX2 dans *fepA* empêche entièrement la régulation par RseX, tandis que la mutation mutRX1 réduit fortement l'expression de la fusion, qui reste

encore légèrement régulée par RseX d'un facteur 2,3. Enfin, le mutant mutRX1,2 combine la diminution de l'expression de *fepA* observée avec le mutant mutRX1 et la perte de la régulation par RseX observée avec le mutant mutRX2. Ces résultats ne sont donc pas en désaccord avec la première prédiction d'appariement, puisque dans l'ensemble les mutants de la fusion préviennent la régulation par RseX. Néanmoins, l'analyse de l'effet des mutants du sRNA est bien moins favorable à cette hypothèse (Fig. 63.C). En effet, les sRNA mutants RseXmut1 et RseXmut2 répriment l'expression de la fusion *fepA*$_{-173}$-*lacZ* au moins aussi efficacement que RseX.

Une adaptation manuelle de la première prédiction d'appariement (Fig. 63.A milieu) permet de proposer un appariement alternatif qui expliquerait pourquoi la fusion *fepA*$_{-173}$-*mutRX2*-*lacZ* n'est pas régulée par RseX. Dans cette hypothèse, la régulation de la fusion wt par RseXmut1 et RseXmut2 serait également plus compréhensible. Toutefois, cet appariement ainsi que le premier proposé, impliquant tous deux des régions en aval du nt -92 de l'ARNm *fepA*, prêtent à interrogation dans la mesure où la fusion *fepA*$_{-92}$-*lacZ* n'est que très faiblement régulée par RseX (Fig. 57). On pourrait imaginer ici que le 5'UTR de *fepA* en amont du nt -92 est nécessaire au mécanisme de régulation par RseX, alors même que l'appariement s'effectue en aval de cette position.

Malgré cette dernière possibilité, étant donné que l'effet répresseur de RseX sur l'expression des fusions *fepA*$_{-146}$-*lacZ* est essentiellement perdu avec la fusion *fepA*$_{-92}$-*lacZ*, j'ai effectué une nouvelle recherche d'appariement entre RseX et l'ARNm *fepA* en concentrant la prédiction sur la région -150 à -85 de l'ARNm *fepA*. Cette recherche a abouti à une troisième prédiction d'appariement (Fig. 63.A bas) qui n'implique pas ou peu les régions mut1 et mut2 de RseX. Cela pourrait expliquer pourquoi RseXmut1 et RseXmut2 répriment toujours l'expression de *fepA*. Par ailleurs, cette prédiction d'appariement est frappante dans la mesure où elle se superpose partiellement au potentiel second site d'appariement de SdsR à *fepA* (Fig. 61.A gauche). Ce site pourrait donc constituer un opérateur efficace pour la régulation par un sRNA, qu'il s'agisse de RseX ou SdsR. Cette fois-ci c'est l'absence de régulation de l'expression des fusions mutRX1 et mutRX2 par RseX qui constitue une contradiction difficile à justifier, dans la mesure où ces mutations ne devraient pas affecter l'appariement. Cette dernière observation pourrait toutefois s'expliquer par un changement de structure en aval de la région

d'appariement, prévenant l'interaction avec RseX ou bien le mécanisme de régulation qui en découle.

Pour conclure, aucune des prédictions d'appariement de RseX à l'ARNm *fepA* ne permet d'expliquer clairement les résultats de ces premières expériences. Il se pourrait qu'une combinaison des mutations mut1 et mut2 de RseX apporte une solution à ce problème si l'une des deux premières prédictions d'appariement devait s'avérer juste, et si une partie de l'appariement suffisait à obtenir une répression forte de l'expression de *fepA*. En ce qui concerne le dernier site d'appariement proposé, le plus simple serait d'effectuer de nouvelles mutations compensatoires qui permettraient de conclure à son existence ou de l'exclure. Par ailleurs, à l'instar de SdsR il se pourrait que la régulation par RseX repose sur une combinaison de deux sites d'appariement. Enfin, une explication peut-être plus anecdotique de la régulation par les mutants de RseX pourrait être qu'ils sont plus fortement exprimés ou plus stables que la forme wt dans ces expériences.

Quelle que soit la réponse à cette énigme, ces expériences confirment l'importance du 5'UTR de *fepA* pour la régulation par RseX puisque divers mutants du 5'UTR et diverses longueurs de 5'UTR modifient fortement la capacité de RseX à réprimer l'expression de *fepA*. De plus, la formation de duplex entre sRNA et ARNm dans ces régions très en amont du RBS n'a été observée que rarement (voir **Revue n°1**) et pourrait impliquer un mécanisme de régulation peu commun voire inédit, qu'il reste à comprendre. Pour finir, une alternative qui reste compatible avec l'ensemble de ces résultats est que RseX réprime la synthèse de FepA de manière indirecte, via un régulateur intermédiaire qu'il serait intéressant d'identifier.

B. Stimuli à l'origine de la régulation de l'expression de *fepA*

L'une des questions que je me suis posées au cours de ma thèse est la suivante : pourquoi les bactéries régulent-elles l'expression de *fepA*? Plus précisément, quelles conditions nécessitent une répression de l'expression de *fepA* par un sRNA? La répression transcriptionnelle de *fepA* par Fur, connue de longue date, apparaît logique d'un point de vue physiologique : en présence de fer, la bactérie n'a pas besoin de ce récepteur complexe et se dispense ainsi des coûts de la synthèse de *fepA*. Mais il existe à l'évidence d'autres régulations, comme celles observées en fonction de la température (Fig. 51), de la phase de croissance

(Fig. 58), ou encore les contrôles post-transcriptionnels par OmrA/B, SdsR et RseX. Si OmrA/B et SdsR sont fortement exprimés en phase stationnaire de croissance, l'expression de ces sRNA répond par ailleurs à des stimuli variés, tandis que les conditions d'expression de RseX restent à comprendre. Ces différents stimuli devraient donc correspondre à des situations où l'expression de *fepA* serait réprimée. J'ai eu l'occasion d'aborder cette question pour la première fois au cours de mon M2, lors duquel je me suis demandé si l'expression des OmrA/B pouvait répondre d'une manière ou d'une autre à la présence de fer, hypothèse que j'ai ensuite mise à l'épreuve de l'expérience. Les stimuli à l'origine de l'expression de RseX étant par ailleurs parfaitement inconnus à ce jour, j'ai voulu en savoir plus en étudiant le promoteur de ce gène et en effectuant un crible génétique pour en comprendre la régulation.

1. La réponse d'OmrA/B au fer

Le stimulus à l'origine de la phosphorylation d'OmpR et de la transcription d'OmrA/B, lié à l'osmolarité et au stress acide, reste mal compris à ce jour. Aussi me suis-je demandé si la présence de fer pouvait entraîner l'expression d'OmrA/B, que ce soit *via* EnvZ-OmpR ou *via* d'autres régulateurs. Cette induction permettrait par exemple de compléter la régulation transcriptionnelle exercée par Fur avec une régulation post-transcriptionnelle prévenant la traduction d'ARNm déjà présents au moment du stimulus, ou encore d'empêcher la traduction d'ARNm transcrits de manière résiduelle malgré l'action de Fur. Certaines expériences présentées ici, indiquées le cas échéant, ont été réalisées par Maxence Lejars, un étudiant qui a effectué son stage de première année de Master au laboratoire sous ma supervision ainsi que celle de Maude Guillier.

a. L'effet d'une carence en fer

L'hypothèse initiale étant que l'expression d'*omrA* et d'*omrB* pourrait être induite par la présence de fer, j'ai réalisé une expérience d'induction par le $FeSO_4$ (Fig. 64.A). Parce que les bactéries étaient cultivées en milieu LB, j'ai considéré que cela pouvait correspondre à un milieu déjà riche en fer, où l'expression des OmrA/B serait peut-être déjà activée. C'est pourquoi j'ai également ajouté une condition d'induction au dip. Comme évoqué précédemment, ce composé chimique chélate le fer, diminuant artificiellement la concentration de fer disponible aux bactéries. En suivant mon hypothèse, le dip aurait pu

diminuer l'expression des OmrA/B en réduisant les niveaux de fer, or c'est exactement l'inverse que l'on peut observer pour OmrA (Fig. 64.A).

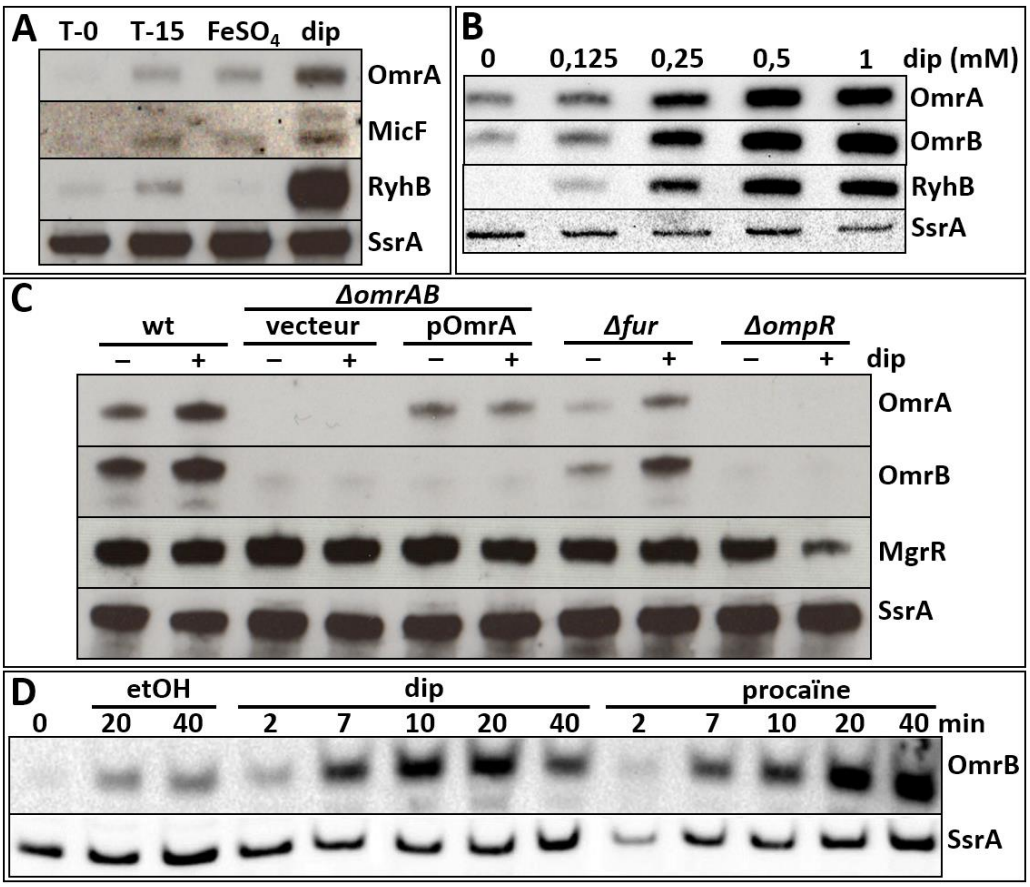


Figure 64 – Une carence en fer active l'expression d'OmrA et OmrB au niveau transcriptionnel
Analyse des niveaux des sRNA OmrA, OmrB, RyhB, MicF et MgrR par NB après séparation par PAGE. SsrA sert de marqueur de charge, tandis que RyhB sert de contrôle de l'induction au dip (A, B) et au $FeSO_4$ (A). Les bactéries ont été cultivées comme suit :
A. La souche DJ480 a été cultivée en milieu LB à 37°C. Les ARN ont été extraits à l'OD_{600} de 0,3 (T-0), après 15 min supplémentaires sans traitement (T-15) ou avec un traitement au FeSO4 à 40µM ou au dip à 250µM. B. La souche DJ624 a été cultivée comme en (A), puis traitée pendant 15 min à la concentration en dip indiquée avant extraction des ARN. C. Des bactéries wt (souche DJ480), ΔomrAB (souche MG1099) transformées par un plasmide contrôle ou surproducteur d'OmrA, Δfur (souche MG1054) ou ΔompR (souche MG1035) ont été cultivées comme en (A), puis cultivées 15 min supplémentaires avec ou sans traitement au dip à 250µM avant l'extraction des ARN. Les bactéries transformées par un plasmide ont de surcroît été traitées avec 10µM d'IPTG dès le départ de la culture pour induire une expression modérée d'OmrA. D. La souche MG1178 a été cultivée en milieu LB à 37°C jusqu'à une OD_{600} de 0,35, puis traitée pendant 40 min au dip à 500µM ou avec un volume équivalent d'éthanol à 90% (etOH) (le dip étant dissout dans une solution d'éthanol à 90%, cette condition sert de contrôle sans dip ici), ou encore à la procaïne à 10mM. Les ARN ont été extraits aux temps indiqués.

Bien que moins spectaculaire que l'induction du sRNA RyhB par le dip dans cette expérience, l'induction d'OmrA en présence de dip semble claire (Fig. 64.A). On notera que malgré l'importante induction de RyhB pouvant entraîner d'éventuels effets de saturation de la protéine Hfq (Hussein and Lim, 2011; Moon and Gottesman, 2011), cela n'empêche pas

l'induction simultanée d'OmrA, mais aussi du sRNA MicF. Etant donné que MicF et OmrA sont tous deux régulé par les SDC EnvZ/OmpR, cela pourrait indiquer que le signal de carence en fer provoqué par le dip est véhiculé par ce SDC. Quant au $FeSO_4$, il semble n'avoir aucun effet sur OmrA à la concentration utilisée ici, alors que l'on retrouve bien la répression attendue de l'expression de RyhB (Massé et al., 2003).

Afin de reproduire ce résultat sur OmrA et afin de savoir s'il en est de même pour OmrB, j'ai analysé les niveaux de ces sRNA après une induction par une gamme de concentrations croissantes de dip (Fig. 64.B). On peut s'apercevoir que les niveaux d'OmrA et d'OmrB augmentent de façon proportionnelle à la concentration au dip, à l'image des niveaux de RyhB. OmrA/B semblent donc répondre donc tous deux à la carence en fer. Cet effet est tout à fait surprenant : en effet, l'on ne s'attend pas à ce que les OmrA/B, notamment répresseurs des gènes *fepA*, *cirA* et *fecA*, soient exprimées précisément dans des conditions de carence en fer où les récepteurs FepA, CirA et FecA sont nécessaires à l'import du fer dans les cellules.

Pour comprendre si cet effet est transcriptionnel ou post-transcriptionnel, j'ai utilisé divers mutants et plasmides (Fig. 64.C). L'expression d'OmrA à partir d'un promoteur hétérologue pLac depuis un plasmide et en contexte *ΔomrAB* permet de constater que le dip n'a plus d'effet sur le niveau d'OmrA. L'effet du dip sur OmrA et probablement OmrB est donc dépendant du promoteur et est très probablement transcriptionnel. Comme Fur est un régulateur majeur de la réponse au fer, j'ai également voulu savoir si cette régulation pouvait passer par Fur. L'induction d'OmrA/B par le dip dans une souche *Δfur* est toujours aussi claire, excluant donc l'intervention de Fur. On peut en revanche noter la diminution globale du niveau d'OmrA et d'OmrB dans cette souche. L'explication la plus simple de cet effet est qu'en l'absence de Fur, plusieurs cibles d'OmrA/B sont exprimées (*i.e.* les ARNm *fepA*, *fecA*, *cirA, flhDC* et peut-être d'autres, voir **Annexe 2.B**). De ce fait, les OmrA/B vont s'apparier à ces ARNm et pourraient dans certains cas être dégradés en même temps que leurs cibles ARNm, entraînant les niveaux plus bas observés.

L'usage d'un mutant *ΔompR* permet par ailleurs de constater que l'expression d'OmrA/B requiert la présence d'OmpR dans la cellule, que ce soit avec ou sans dip. Cela n'exclut pas l'intervention d'un autre régulateur pour l'induction par la carence en fer,

néanmoins la présence d'OmpR reste indispensable pour l'expression d'OmrA/B dans ces conditions. Un contrôle notable dans cette expérience est MgrR, un sRNA Hfq-dépendant régulé par le SDC PhoP/Q en réponse à la carence en ions magnésium Mg^{2+} (Moon and Gottesman, 2009). L'absence d'effet du dip sur les niveaux de MgrR, hormis avec la souche *ΔompR* (peut-être du fait d'un problème technique dans l'expérience ou bien qui reste inexpliquée pour l'instant), permet de conclure qu'il n'y a vraisemblablement pas un effet global du dip et de la carence en fer sur les niveaux de l'ensemble des sRNA Hfq-dépendants.

Enfin, afin de comparer l'induction au dip à une induction OmpR-dépendante classique, Maxence Lejars a réalisé une expérience de cinétique en suivant l'induction d'OmrB pendant 40 minutes après une induction au dip, ou bien à la procaïne (Fig. 64.D). Ce dernier composé chimique est utilisé pour générer une induction des OmrA/B dépendante d'EnvZ/OmpR (Guillier and Gottesman, 2006; Rampersaud and Inouye, 1991). Si la cinétique et la force d'induction d'OmrB au dip est semblable à celle observée avec la procaïne, on note cependant qu'elle est légèrement plus rapide avec un niveau légèrement plus élevé au bout de 10 minutes, mais qu'elle dure moins longtemps puisqu'au bout de 40 minutes les niveaux d'OmrB commencent à décroître tandis qu'avec la procaïne l'induction continue. Ces nuances pourraient indiquer que la réponse des OmrA/B au dip est véhiculée par un mécanisme différent de leur réponse à la procaïne. Si tant est que ces deux mécanismes d'induction soient dépendants d'EnvZ/OmpR, la véritable implication de ce SDC pour l'induction par le dip est difficilement vérifiable puisque les niveaux des OmrA/B sont trop bas en l'absence de l'une ou l'autre protéine pour permettre de voir une différence. Toutefois, l'usage de fusions aux promoteurs d'*omrA/B*, ou encore d'une forme constitutivement active d'OmpR, pourrait nous permettre de répondre à cette question. Quoi qu'il en soit, l'intervention d'un autre régulateur en plus d'EnvZ/OmpR n'est pas exclue.

b. La recherche d'un régulateur et l'induction par les ions Fe^{2+}

À la suite de ces premiers résultats, nous avons effectué une recherche bibliographique afin d'essayer d'en déduire quel régulateur hormis OmpR pourrait intégrer le signal de carence en fer pour induire l'expression d'OmrA/B. Sans entrer dans le détail des expériences, nous avons vérifié l'induction d'OmrA/B dans des souches délétées des gènes *cpxR*, *phoP*, *fecI* ou *basRS* codant divers régulateurs grâce à des fusions transcriptionnelles du promoteur d'OmrB

à *lacZ*. De même, l'effet de l'inactivation des gènes *fnr*, *iscR*, *nsrR* ou *soxR* sur les niveaux d'OmrA a été analysé par NB. Dans tous les cas, l'induction par le dip reste la même, ce qui signifie qu'aucun de ces régulateurs n'est impliqué dans la réponse d'OmrA/B au dip. Brièvement, nous avons testés ces divers mutants pour les raisons suivantes :

- CpxR, le régulateur du SDC CpxA/R, est important pour la réponse à la carence en fer chez *Vibrio cholerae* (Acosta et al., 2015) ;
- les SDC PhoP/Q et surtout BasR/S sont également impliqués dans la réponse aux variations de niveaux de fer notamment chez *E. coli* (Hagiwara et al., 2004) ;
- FecI est un facteur sigma répondant à la présence de citrate de fer et activant l'expression de *fecA* (Visca et al., 2002) ;
- Fnr, IscR, NsrR et SoxR sont divers FT répondant au stress oxydatif et régulant la synthèse de centres Fer-Soufre (voir **introduction III.C.1**) (Lill, 2009; Mettert and Kiley, 2015)).

Cette recherche d'un régulateur expliquant l'effet d'une carence en fer sur les OmrA/B en testant différents gènes-candidats ayant échoué, nous avons choisi de procéder autrement et de comprendre d'abord plus en détail le signal entraînant cette induction. Maxence Lejars a donc suivi l'expression de fusions transcriptionnelles *PomrB-lacZ* afin de confirmer l'effet transcriptionnel du dip sur l'expression d'OmrB. De plus, pour déterminer si dans ce cas précis l'effet du dip pouvait être compensé par le fer, des concentrations croissantes de $FeSO_4$ ont été ajoutées simultanément au dip afin de saturer ce chélateur (Fig. 65).

Cette expérience confirme l'effet activateur du dip sur le promoteur P*omrB*, il s'agit donc bien d'un effet transcriptionnel. De plus, comme attendu l'ajout de faibles concentrations de $FeS0_4$ compense bien, au moins en partie, l'effet activateur du dip sur l'expression de la fusion *PomrB-lacZ*. Cependant, à notre grand étonnement cette expérience révèle qu'il existe aussi un effet activateur du fer $FeS0_4$ à fortes concentrations sur l'expression d'OmrB, comme le proposait l'hypothèse de départ. Il faut toutefois signaler que les concentrations de $FeS0_4$ utilisées dans cette expérience sont très élevées car nous voulions nous assurer de saturer le dip. De telles quantités de fer peuvent être toxiques pour les cellules, et entraînent certainement des effets pléiotropiques sur l'expression des gènes.

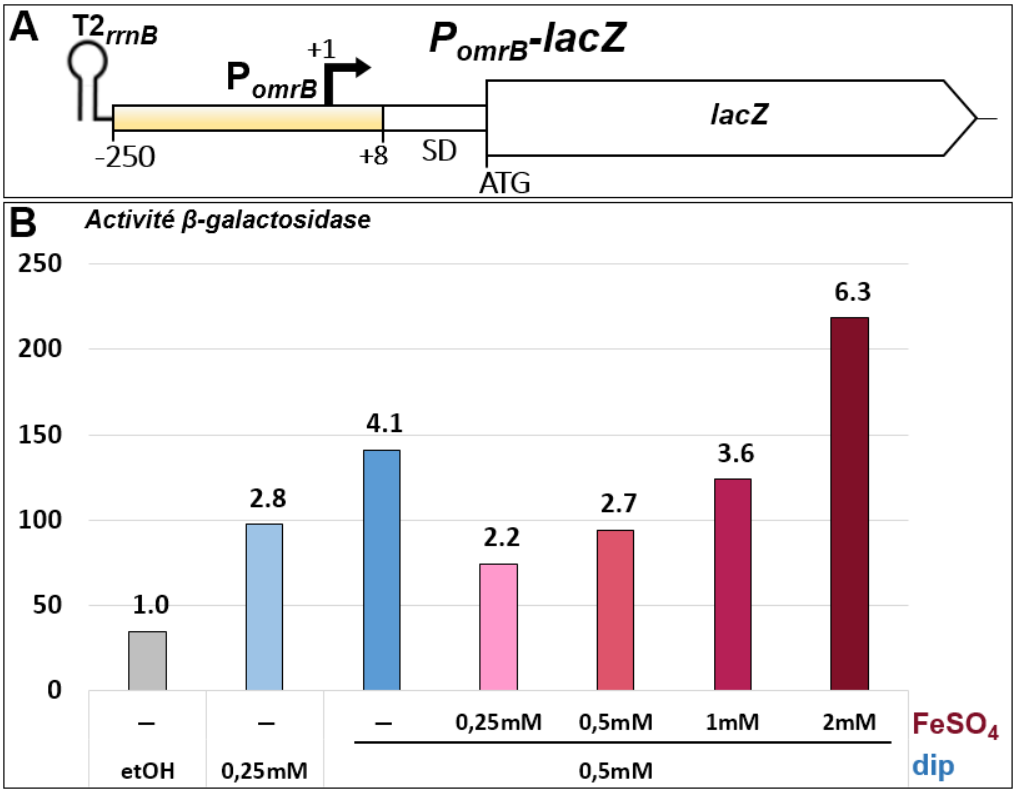


Figure 65 – Le FeSO4 compense l'effet du dip à basse concentration et semble activer la transcription d'OmrB à haute concentration

A. Fusion transcriptionnelle PomrB-lacZ. Cette fusion est composée du promoteur du gène omrB des positions -250 à +8 par rapport au TSS, suivi du gène lacZ à partir de sa position -17 par rapport au codon ATG, incluant les signaux de démarrage de la traduction de lacZ. Le terminateur T2 du gène rrnB précède le promoteur PomrB afin de prévenir toute transcription provenant de régions situées en amont des éléments régulateurs du gène omrB. B. Activités β-galactosidase spécifiques de la fusion PomrB-lacZ (souche ML04). Les bactéries ont été cultivées jusqu'à une OD_{600} de 0,3 puis ont subi 15min de traitement aux concentrations indiquées de dip et de $FeSO_4$, ou au volume équivalent d'éthanol à 90% (etOH). Les facteurs d'activation par rapport au contrôle etOH sont indiqués.

Pour confirmer cet effet, j'ai donc réalisé une nouvelle expérience en utilisant diverses sources de fer contenant des ions Fe^{2+} ou Fe^{3+} à 150µM (Fig. 66), c'est à dire des concentrations plus modérées que lors de l'expérience précédente, mais plus élevées que lors de l'expérience initiale où l'effet du fer n'était pas visible (Fig. 64.A). On observe ici une induction également plus modérée de l'expression d'OmrB. Plus étonnamment, seuls les ions Fe^{2+} semblent induire l'expression d'OmrB, tandis que les ions Fe^{3+} et des ions bivalents Mg^{2+} et Cu^{2+} ne semblent pas affecter l'expression à partir du promoteur P*omrB*. Par ailleurs, il semblerait que l'induction maximale d'OmrA/B par le FeSO4 soit atteinte à la concentration de 250µM (données non montrées ici, voir induction Fig. 67.A). La logique de cette induction par le fer, qui corrobore l'hypothèse formulée au début de ce sous-chapitre, pourrait s'expliquer plus aisément que celle observée avec le dip. En effet, en présence d'ions Fe^{2+}

facilement accessibles à la cellule par la voie du transporteur FeoAB, la synthèse des transporteurs spécifiques FepA, FecA et CirA n'est pas nécessaire et peut être réprimée par les OmrA/B, dont l'expression est induite en réponse au Fe^{2+}. En revanche, les ions Fe^{3+} nécessitent des transporteurs spécialisés comme FecA pour le citrate de fer, ou encore FepA pour les complexes Fer-entérobactine. C'est pourquoi l'expression des OmrA/B ne serait pas induite en présence de ces seuls ions Fe^{3+}, afin de permettre l'expression des gènes codant ces récepteurs. Ajoutons à titre de remarque générale que des expériences réalisées en milieu plus défini (plutôt qu'en milieu LB), et donc mieux contrôlé, permettraient d'avoir une idée plus précise des conditions de croissance des bactéries et des stimuli auxquels répondent les sRNA OmrA/B.

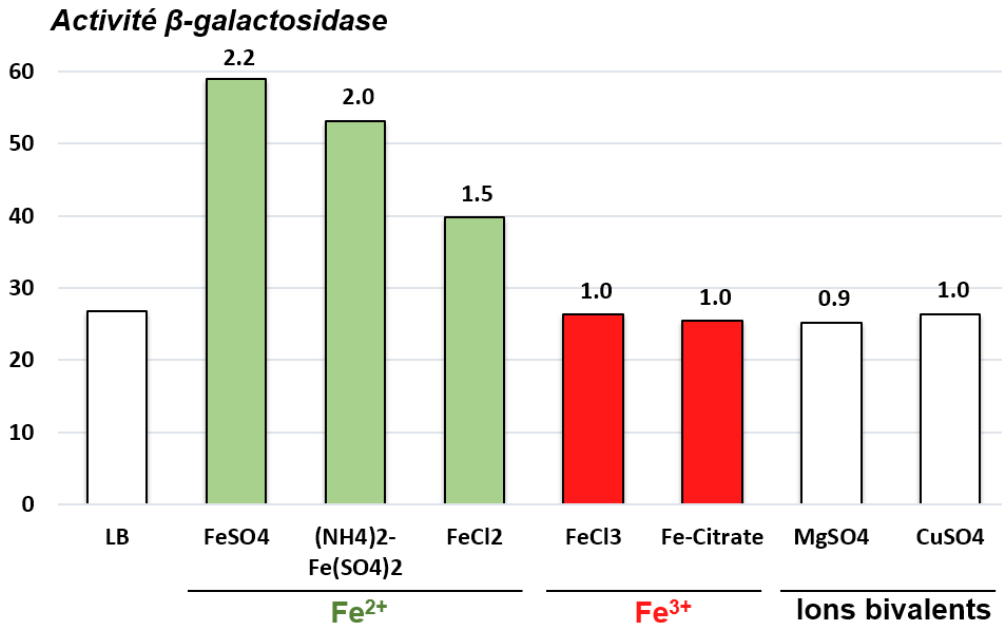


Figure 66 – Seuls les ions Fe^{2+} induisent l'expression d'OmrB
Activités β-galactosidase spécifiques de la fusion PomrB-lacZ (souche ML04). Les bactéries ont été cultivées en milieu LB à 37°C jusqu'à une OD_{600} de 0,3 puis ont subi 15 min de traitement aux ions Fe^{2+}, Fe^{3+}, Mg^{2+} ou Cu^{2+} à 150µM.

Afin de déterminer plus précisément si le SDC d'EnvZ-OmpR joue un rôle dans la régulation des OmrA/B par le dip et/ou le fer (puisque ces deux régulations pourraient tout à fait suivre deux mécanismes différents), j'ai analysé l'effet du dip et du fer sur les ARNm *ompC* et *ompF* régulés par OmpR/EnvZ (voir **Introduction III.B.1.a**) (Fig. 67). D'une part, un NB sur les ARNm *ompC* et *ompF* en conditions d'induction au $FeSO_4$ à 250µM ou au dip à 500µM (Fig. 67.A) révèle qu'*ompF* est fortement réprimé dans ces deux conditions, tandis que les niveaux d'*ompC* diminuent uniquement en présence de $FeSO_4$. On observe bien l'induction de

l'expression des OmrA/B dans les deux conditions, ce qui confirme les résultats réunis jusqu'ici. De plus, l'effet du fer et du dip sur *ompC* et *ompF* semble être post-transcriptionnel puisque des fusions transcriptionnelles aux promoteurs des gènes *ompC* et *ompF* ne sont pas affectées par la présence de dip ou de $FeSO_4$ (Fig. 67.B et C). Notons que si cette dernière expérience a été réalisée à l'équilibre, l'expression des mêmes fusions a été mesurée lors d'une induction de 15 minutes au dip ou au $FeSO_4$ aux mêmes concentrations, pour un résultat strictement identique (résultats non montrés ici). Cela semble donc exclure une intervention d'EnvZ-OmpR, et ce d'autant plus clairement que la délétion d'*envZ* n'empêche pas la régulation d'*ompF* en présence de dip (résultats non montrés ici).

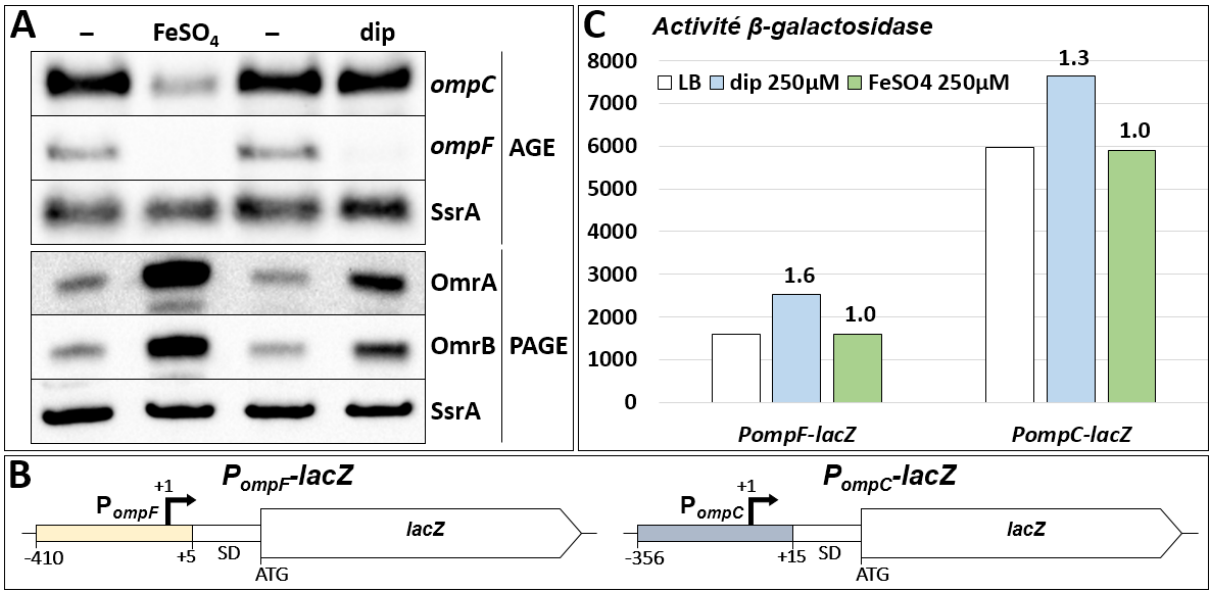


Figure 67 – Le fer et le dip affectent l'expression d'ompC et ompF au niveau post-transcriptionnel
A. Analyse des niveaux des ARNm ompC et ompF et des sRNA OmrA et OmrB par NB. La souche ML04 a été cultivée en milieu LB jusqu'à une DO_{600} de 0,3, puis ont poursuivi leur croissance pendant 15 min sans traitement (-) ou avec un traitement au FeSO4 à 250µM ou au dip à 500µM. Les ARN ont été extraits et analysés par la suite par NB. SsrA sert de marqueur de charge. B. Fusions transcriptionnelles PompF-lacZ et PompC-lacZ. Ces fusion sont composées des promoteurs des gènes ompF des positions -410 à +5 ou ompC des positions -356 à +15 par rapport au TSS (pour ompC qui a plusieurs TSS, par rapport au TSS proximal), suivis du gène lacZ à partir de sa position -17 par rapport au codon ATG, incluant donc ses signaux de démarrage de la traduction. C. Activités β-galactosidase spécifiques des fusions PompF-lacZ (souche MG1690) et PompC-lacZ (souche MG1892). Les bactéries ont été cultivées jusqu'à une OD_{600} de 0,4 à 37°C en milieu LB sans traitement ou bien en présence de dip ou de $FeSO_4$ à 250µM. Les facteurs d'activation par rapport au contrôle sans traitement (LB) sont indiqués.

La régulation de l'expression d'*ompF* étant très vraisemblablement post-transcriptionnelle, elle pourrait être due au sRNA MicF (voir **Introduction III.B.1.a**). Puisque l'expression de *micF* semble également induite en présence de dip (Fig. 64.A), cela pourrait expliquer la régulation d'*ompF* observée au moins dans cette condition. De plus, si l'expression

d'*ompF* et d'*ompC* n'est pas affectée au niveau transcriptionnel (Fig. 67.C), il devient difficile d'imaginer que la carence en fer soit perçue par le SDC EnvZ/OmpR qui régule ces deux gènes au niveau transcriptionnel (voir **Introduction III.B.1.a**). Dès lors, il est possible que ce signal soit intégré par l'un des autres régulateurs de MicF : Lrp, Rob, MarA ou SoxS (Delihas and Forst, 2001). Or Rob est un FT qui répond justement à la présence de dip (Griffith et al., 2009). Plus précisément, d'après cette étude il semblerait que Rob soit présent dans les cellules sous forme d'oligomères inactifs, chaque monomère interagissant avec les autres *via* le domaine CTD de cette protéine. En présence de dip, mais aussi de 4,4'-dipyridyl, un composé chimique similaire au dip mais ne chélatant pas le fer, les monomères Rob se sépareraient, rendant l'activité régulatrice de ce FT possible. Il est donc envisageable que MicF ne réponde pas à la carence en fer mais plus simplement à l'activation par Rob. Dès lors, on peut imaginer que la régulation des OmrA/B suive le même mécanisme de réponse à Rob plutôt qu'à la carence en fer.

Cette réponse hypothétique à Rob pourrait s'inscrire dans une logique intéressante dans la mesure où ce FT permet une réponse au stress oxydatif, lors duquel l'import de fer dans la cellule est extrêmement néfaste du fait de la production de radicaux hydroxyle $^{\bullet}OH$ par la réaction de Fenton (voir **Introduction III.C.1.b**). Ainsi l'induction des OmrA/B pourrait être nécessaire pour réprimer l'import de fer dans les conditions physiologiques où Rob serait actif dans la cellule. En renfort de cette hypothèse, on peut noter que les OmrA/B figurent parmi les sRNA les plus fortement exprimés chez des souches de *S. enterica* infectant des cellules de macrophages (Westermann et al., 2016). Or, les bactéries infectant des macrophages subissent justement un stress oxydatif particulièrement intense (Miller and Britigan, 1997) notamment *via* l'action de NADPH oxydases (Aussel et al., 2011; Hebrard et al., 2009). Si l'hypothèse de l'activation transcriptionnelle des OmrA/B par Rob s'avère exacte, cela expliquerait potentiellement la forte expression d'OmrA/B lors d'une infection bactérienne.

2. La recherche de régulateurs de RseX

A ce jour, aucun régulateur de l'expression du sRNA RseX n'a été mis en évidence, tandis que l'expression de cet ARN à partir de son locus chromosomique n'a jamais été observée par NB. La question de la régulation de RseX reste donc entière et m'a intéressé

d'autant plus que trouver les signaux qui conduisent à l'expression de RseX pourrait également permettre de mieux comprendre les conditions lors desquelles l'expression de *fepA* est régulée. J'ai donc analysé le promoteur du gène rseX dans le détail, avant d'effectuer un crible génétique afin de chercher d'éventuels régulateurs de ce gène.

a. La région promotrice de *rseX*

L'analyse des résultats d'un RNAseq visant à identifier le TSS des ARN d'*E. coli* (Fig. 68) nous a conduits à penser que RseX pourrait avoir au moins un autre TSS situé en aval du seul TSS connu actuellement (Douchin et al., 2006), générant un ARN d'environ 60 nt.

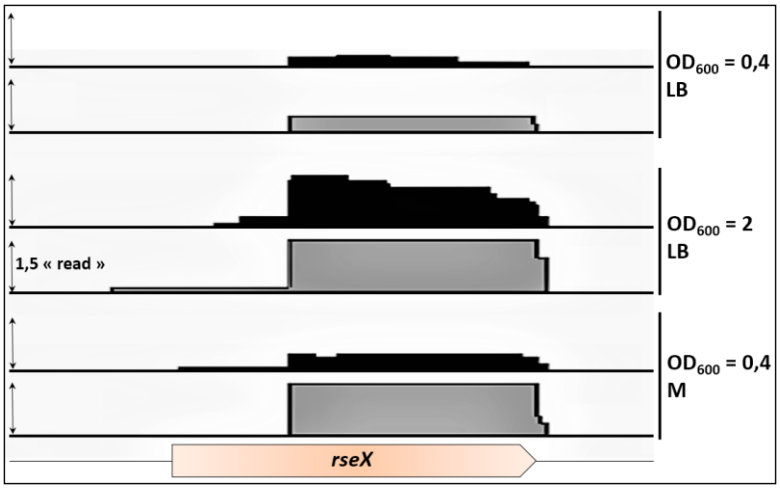


Figure 68 – Une forme plus courte de RseX pourrait exister in vivo
Analyse de RNAseq, focalisée ici sur le gène rseX. Les échantillons considérés ici sont issus de cultures de la souche MG1655 en milieu LB ou minimum (M) à 37°C jusqu'à une OD_{600} de 0,4 ou de 2. Les « reads » correspondant à RseX (i.e. nombre d'ADN complémentaires s'alignant avec le gène rseX normalisés par rapport à l'ensemble des ADN complémentaires alignés sur le génome) sont indiqués sur une échelle similaire pour chaque échantillon, et sont représentés alignés avec le gène rseX tel que proposé par (Douchin et al., 2006). Les « reads » en noir ou en gris correspondent aux résultats obtenus à partir des ARN totaux respectivement non traités, ou traités à la RNase « TEX » dégradant spécifiquement les ARN aux extrémités 5'-monophosphate, généralement issus d'un clivage in vivo (voir ***Introduction I.A.1 et II.B.2****). D'après les données de (Thomason et al., 2015).*

Un promoteur peut être prédit en amont de ce second TSS (ou TSS 2, Fig. 69.A), qui expliquerait l'expression d'une version plus courte de RseX. Pour vérifier cette hypothèse, j'ai construit deux fusions transcriptionnelles *PrseX-lacZ*, l'une ne comprenant que le promoteur connu que j'appellerai *P1-lacZ* et l'autre comprenant également le nouveau promoteur hypothétique que j'appellerai *P1-P2-lacZ*. Suivant la même logique, le promoteur connu sera

donc nommé P1 et le nouveau promoteur proposé P2. J'ai par ailleurs construit des fusions mutantes « mutP1 » où le promoteur P1 est muté et ne devrait plus permettre l'expression génétique (Fig. 69).

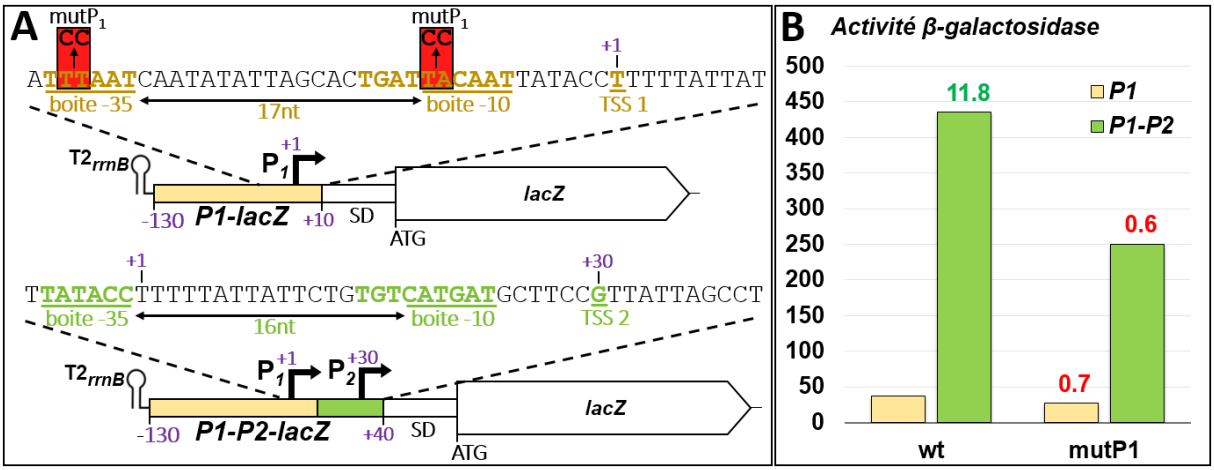


Figure 69 – Expression génétique à partir des promoteurs P1 et P2 de RseX
A. Fusions transcriptionnelles aux promoteurs de RseX P1-lacZ et P1-P2-lacZ. Ces fusions sont composées du gène RseX des positions -130 à +10 (P1-lacZ) ou -130 à +40 (P1-P2-lacZ) par rapport au TSS 1 connu (Douchin et al., 2006), suivis du gène lacZ à partir de sa position -17 par rapport au codon ATG, incluant donc ses signaux de démarrage de la traduction. Les fusions sont précédées du terminateur T2 du gène rrnB afin de prévenir toute transcription provenant de régions situées en amont des éléments régulateurs du gène rseX. Les positions par rapport au TSS 1 sont indiquées en violet pour les deux fusions ; le promoteur P1 est indiqué en jaune tandis que le promoteur P2 proposé ici est indiqué en vert, les deux ont une boîte -10 étendue ; la distance entre les boîtes -35 et -10 de chaque promoteur est fléchée ; la mutation « mutP1 » est indiquée en rouge. B. Activités β-galactosidase spécifiques des fusions P1-lacZ (souche JJ0159), P1-P2-lacZ (souche JJ0161), et de leurs dérivées P1-mutP1-lacZ et P1-P2-mutP1-lacZ (respectivement souches JJ0172 et JJ0173). Les bactéries ont été cultivées en milieu LB à 37°C jusqu'à une OD_{600} de 0,55. Le rapport d'activité observé entre la fusion P1-P2-lacZ et la fusion P1-lacZ est indiquée en vert, tandis que celui des fusions mutées par rapport aux fusions wt correspondantes est indiqué en rouge.

Le résultat de cette expérience est éloquent, puisque la mutation mutP1 permet aussi bien la transcription de la fusion *P1-lacZ* que de la fusion *P1-P2-lacZ* malgré une légère diminution de l'expression (Fig. 69.B). Cela signifie que la fusion P1-lacZ contient certainement un autre promoteur permettant l'expression de RseX, sous réserve que le terminateur $rrnB_{T2}$ en amont de cette fusion empêche bien toute transcription à partir de régions situées en amont des régions de *rseX* étudiées ici. De manière encore plus étonnante, l'expression de la fusion *P1-P2-lacZ* est plus de dix fois plus importante que celle de la fusion *P1-lacZ*, suggérant que le promoteur P2 pourrait bien exister. Il est toutefois possible d'imaginer que le ou les transcrits générés par cette fusion soient beaucoup plus stables, peut-être suite à une étape de maturation. Alternativement, cette région du gène *rseX* pourrait être importante pour la

transcription à partir d'un promoteur situé en amont de P2, mais autre que le promoteur P1 dont la mutation n'a pas d'effet ici. Il est difficile de trancher entre ces hypothèses pour l'instant, aussi construire un mutant similaire du promoteur P2 pourrait nous en apprendre plus sur les signaux transcriptionnels de *rseX*.

J'ai par ailleurs effectué une prédiction *in silico* de promoteurs sur l'ensemble de la région ADN utilisée dans la fusion *P1-P2-lacZ* à l'aide du programme « Neural Network Promoteur Prediction » (http://www.fruitfly.org/seq_tools/promoter.html). Ce programme a initialement été mis au point pour la recherche de promoteurs chez *Drosophila melanogaster* et les organismes eucaryotes (Reese, 2001) mais a été adapté au même type de recherche chez les procaryotes. Notons dans un premier temps que ce programme prédit l'existence du promoteur P1 avec un score de 0,81 sur un maximum de 1, représentatif de la ressemblance du promoteur avec l'ensemble de promoteurs qui ont servi à entraîner l'algorithme de recherche. Le programme propose toutefois un TSS à la position -1 par rapport à celle proposée par (Douchin et al., 2006), qui repose sur une expérience de « 5'-RACE ». Le programme ne prédit pas l'existence du promoteur P2, mais propose deux autres promoteurs, que j'ai nommés P3 et P4 (Fig. 70), avec des scores respectifs de 0,99 et de 0,82.

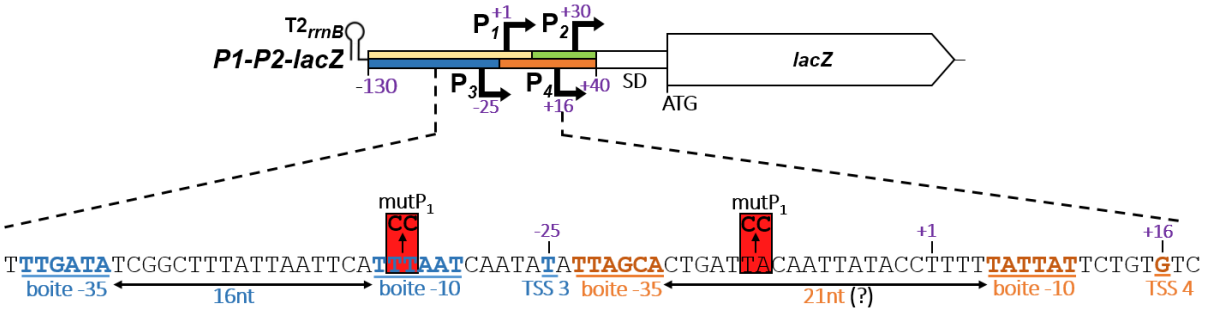


Figure 70 – Autres promoteurs hypothétiques de rseX d'après le programme « Neural Network Promoteur Prediction »
La fusion P1-P2-lacZ est représentée avec les promoteurs hypothétiques P3 et P4, respectivement en bleu et en orange. Les distances entre les boîtes -35 et -10 sont fléchées. La mutation « mutP1 » est indiquée en rouge et les positions par rapport au TSS du promoteur P1 sont indiquées en violet.

Le promoteur P3 est donc de loin celui qui a obtenu le score le plus élevé d'après le programme de prédiction, et présente effectivement des boîtes -35 et -10 très proches du consensus (correspondant respectivement à « TTGACA » et « TATAAT ») ainsi qu'une distance de 16nt entre ces deux régions, proche de la distance optimale de 17nt. L'existence de ce promoteur serait compatible avec l'hypothèse d'un transcrit plus stable en présence des

régions +10 à +40 de RseX (par rapport au TSS 1), ou bien avec la nécessité des régions +10 à +40 pour une transcription optimale. Néanmoins, l'existence de ce promoteur P3 est difficilement compatible avec l'effet de la mutation mutP1, puisque celle-ci modifie la boîte -10 de ce promoteur (Fig. 70) et devrait fortement affecter toute transcription à partir de ce promoteur, ce qui n'est pas observé dans l'expérience de β-galactosidase (Fig. 69.B). Enfin, le promoteur P4 constitue une alternative plausible au promoteur P2 pour expliquer la forte expression de la fusion *P1-P2-lacZ*. Si les boîtes -35 et -10 de ce promoteur hypothétique sont également très proches des consensus, elles semblent néanmoins trop éloignées l'une de l'autre : 21nt ici. Notons toutefois que la boîte -10 (« TATTAT », voir Fig. 70) pourrait être rapprochée de 3nt vers la boîte -35 (pour devenir « TTTTAT »), de même que le TSS. La distance serait alors de 18nt et la boîte -10 resterait relativement proche du consensus.

Un élément important à considérer en conclusion est la taille des transcrits RseX générés à partir de ces divers promoteurs. En effet, dans l'ordre de l'extrémité 5' à 3', les promoteurs P3, P1, P4 et P2 produiraient respectivement des transcrits de 116 nt, 91 nt, 76 (ou 79) nt et 62 nt en considérant le terminateur décrit dans (Douchin et al., 2006). Si toutes ces formes hypothétiques de RseX contiennent les régions nécessaires aux interactions proposées entre RseX et les ARNm *ompC* et *ompA* (Douchin et al., 2006), la forme à 62 nt ne serait en revanche plus capable de former la majeure partie des interactions avec l'ARNm *fepA* proposées dans ce manuscrit (voir Fig. 63.A). Par ailleurs, l'observation par NB d'un transcrit d'environ 100 nt lorsqu'il est exprimé à partir d'un plasmide multicopie porteur du locus chromosomique de *rseX* (Douchin et al., 2006) pourrait s'expliquer par une transcription à partir des promoteurs P3 ou P1. Nos propres résultats de NB (voir Fig. 60) sont quant à eux difficiles à interpréter ici puisque la transcription de RseX est placée sous le contrôle d'un promoteur inductible P_{lac} et non son de propre promoteur. Enfin, les résultats du RNAseq cité précédemment (Thomason et al., 2015) indiquent une forme majoritaire à environ 60 nt qui pourrait être transcrite à partir du promoteur P2 (voir Fig. 68). Cette forme semble être plus abondante dans les échantillons traités à la TEX, suggérant qu'il pourrait s'agir d'un transcrit primaire plutôt que d'un résultat du clivage d'un transcrit précurseur. On pourrait néanmoins imaginer qu'une modification de l'extrémité 5' du fragment ainsi généré ait été modifiée à la suite du clivage, le protégeant du traitement par la RNase TEX réalisé dans cette étude. La

question reste donc entière, et il serait intéressant de caractériser plus en détail l'ensemble des transcrits RseX, qu'il s'agisse de transcrits primaires ou de produits de maturation.

b. La recherche d'un régulateur transcriptionnel de RseX

Après avoir établi qu'une activité transcriptionnelle était détectable à partir de la région promotrice de *rseX*, nous avons voulu élucider la régulation génétique de son expression. La première piste que nous avons suivie est celle du SDC EnvZ-OmpR, parce que des expériences d'immunoprécipitation de chromatine suivies d'analyse sur puce à ADN (« ChIP-on-chip ») indiquaient une fixation d'OmpR au locus du gène *rseX* chez *E. coli* et *S. enterica* (Quinn et al., 2014). Cependant, les fusions aux promoteurs de *rseX P1-lacZ* et *P1-P2-lacZ* ont des activités strictement similaires en contexte wt, *ΔenvZ*, *ΔompR* ou encore lors d'un traitement à la procaïne (données non montrées ici). S'il conviendrait de s'assurer de l'absence de toute induction de RseX à partir de son locus chromosomique par NB, nous en avons toutefois déduit qu'au moins chez *E. coli*, la transcription de RseX était probablement indépendante du SDC EnvZ-OmpR.

Nous avons dès lors décidé de mettre à profit nos fusions *P1-lacZ* et *P1-P2-lacZ*, aux activités β-galactosidases respectives d'environ 40 et 440 unités Miller, dans un crible génétique suite à une mutagenèse par transposon, afin de chercher des régulateurs de *rseX* sur divers milieux. Ce crible a été réalisé sur les deux fusions au promoteur de *rseX* (voir Fig. 69.A) et est décrit en Figure 71. Sur plus de 8500 clones porteurs d'un transposon observables après étalement sur boîte, 236 ont été retenus comme ayant un phénotype, basé sur l'activité β-galactosidase, différent de la majorité des clones, et isolés par purifications successives. Ont été mesurées les activités β-galactosidase des 50 clones pour lesquels les phénotypes les plus marquants ont été observés (25 pour la fusion *P1-lacZ* et 25 pour la fusion *P1-P2-lacZ*). Parmi ces clones, ont été cartographiés les transposons des 22 clones les plus prometteurs, c'est-à-dire ceux dont les activités β-galactosidase mesurées confirmaient le plus clairement l'effet observé sur boîte et/ou dont les différences d'activités mesurées par rapport aux souches wt était les plus fortes (14 de ces clones portent la fusion *P1-lacZ* et 8 la fusion *P1-P2-lacZ*). Le tableau 1 indique pour chacun de ces clones le gène interrompu par le transposon Tn10, les rapports d'activité β-galactosidase par rapport à la souche wt dans diverses conditions de culture et les différences de coloration observées sur les milieux sélectifs du crible génétique.

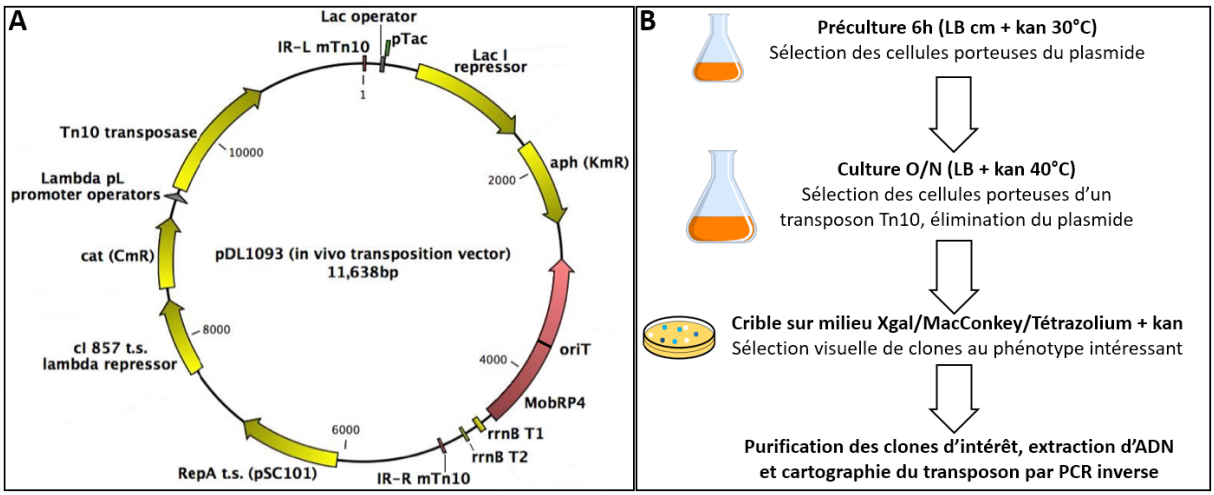


Figure 71 – L'expérience de crible génétique

A. Carte physique du plasmide pDL1093 (Lewis et al., 2011) utilisé dans cette expérience. Le transposon Tn10 a été intégré entre les séquences IR-L et IR-R. Il porte un promoteur sortant pTac et le gène lacI, ce qui place la transcription à partir de ce promoteur sous le contrôle du répresseur LacI et la rend donc inductible par l'IPTG. Ce transposon porte également une cassette aph de résistance à la kanamycine (KmR), ainsi que l'origine oriT de réplication du plasmide. Outre le transposon, le plasmide porte un allèle thermosensible du gène de la réplicase RepA empêchant la réplication du plasmide à la température non-permissive de 40°C. pDL1093 porte également l'allèle thermosensible 857 ts du gène codant le répresseur cI du phage lambda, ce qui permet de lever à température non-permissive la répression de l'expression du gène de la transposase Tn10 situé en aval. Enfin le plasmide porte une cassette cat de résistance au chloramphénicol (CmR). Ce plasmide nous a été généreusement offert par N. Figueroa-Bossi, L. Bossi et A. Camilli. Cette carte a été adaptée du manuel « Experiments in Bacterial Genetics 2016 ».

*B. Schéma du crible génétique. Des bactéries portant des fusions P1-lacZ (souche JJ0159) ou P1-P2-lacZ (souche JJ0161) transformées par le plasmide pDL1093 ont été cultivées 6h à 30°C en milieu LB + chloramphénicol + kanamycine, puis passé la nuit en culture à 40°C en LB + kanamycine. Dans cette dernière conditions le plasmide ne peut plus se répliquer, la transposase est exprimée et les bactéries sont sélectionnées pour la résistance à la kanamycine et donc pour la présence d'un transposon Tn10 intégré au chromosome. Après 12h, la plupart des cellules contiennent donc un transposon intégré au chromosome et ont perdu le plasmide pDL1093. Ce mélange de cellules a ensuite été titré puis étalé sur divers milieux contenant de la kanamycine : MacConkey, LB + X-gal ou tétrazolium (voir **Matériels et Méthodes**) afin de visualiser l'activité β-galactosidase des clones ainsi sélectionnés.*

Ces résultats demandent à être confirmés, notamment par la transduction de ces transposons dans les souches d'origine afin de vérifier si les phénotypes observés sur l'activité β-galactosidase sont toujours présents, mais aussi par la détection de RseX par NB si cela est possible avec certains de ces mutants. Toutefois, l'analyse de ces premiers résultats permet de dresser un premier bilan et d'émettre quelques hypothèses. On notera par ailleurs qu'un certain nombre de clones sont mutés dans les gènes *lacZ* et *lacY*. Ils servent ici de contrôles du bon fonctionnement du crible, dans la mesure où l'on retrouve bien des gènes influant sur l'activité β-galactosidase.

	Souche	N° clone (crible)	Gène interrompu	Fonction	Facteur de répression			Pétri	Isolé sur
					E	O/N	C		
P1-lacZ	JJ0301	35	*lacZ*	β-galactosidase	-	-	-	--	T
	JJ0303	46	*waaA*	biogenèse du LPS	1.8	0.9	1.1	-	X
	JJ0304	49	*hokC*	toxine	1.8	1.0	1.0	+	T
	JJ0305	50	*mhpA*	monooxygénase (?)	1.3	0.6	0.6	++	M
	JJ0306	52	*pfkA*	6-phosphofructokinase I	2.7	1.1	1.0	--	X
	JJ0313	100	*yegV*	"sugar kinase" (?)	2.0	0.9	1.0	-	T
	JJ0315	106	*yajO*	1-deoxyxylulose-5-phosphate synthase	1.2	1.1	0.5	++	T
	JJ0317	114	*insEF*	transposase	1.8	1.0	1.0	-	T
	JJ0319	119	*cspB*	"cold shock protein"	2.1	0.9	1.0	-	T
	JJ0321	184	*gcvP*	glycine décarboxylase	1.7	1.3	1.2	--	T
	JJ0322	188	*(rssB)-galU*	(régulateur de *rpoS*) - UTP-glucose-1-phosphate uridylyltransférase	2.2	1.2	1.1	-	T
	JJ0325	197	*dmsA*	dimethyl sulfoxide réductase (sous-unité A)	2.6	0.9	1.2	--	T
	JJ0326	211	*(rssB)-galU*	(régulateur de *rpoS*) - UTP-glucose-1-phosphate uridylyltransférase	2.5	1.4	1.1	+	T
	JJ0327	214	*waaG*	biogenèse du LPS	4.0	2.7	1.6	-	X
P1-P2-lacZ	JJ0328	87	*(arcA)-yjjY*	(régulateur de *rpoE*) - petit peptide hypothétique	0.7	0.8	1.1	+	M
	JJ0334	135	*hicB*	antitoxine (régulateur de *rpoE*)	3.5	0.9	1.6	+	M
	JJ0335	138	*(arcA)-yjjY*	(régulateur de *rpoE*) - petit peptide hypothétique	2.0	0.6	1.2	+	M
	JJ0341	167	*lacY*	transporteur de lactose	1.8	0.8	1.2	--	T
	JJ0342	168	*lacY*	transporteur de lactose	0.8	0.7	1.1	--	T
	JJ0349	217	*lacY*	transporteur de lactose	1.6	0.7	1.5	--	T
	JJ0351	220	*waaQ*	biogenèse du LPS	0.8	0.9	0.8	-	X
	JJ0353	227	*waaQ*	biogenèse du LPS	0.7	0.8	0.9	-	X

Tableau 1 – Gènes candidats à un rôle dans la régulation de RseX

Liste des clones isolés lors du crible génétique. Sont représentés les clones pour lesquels le gène interrompu par Tn10 a été cartographié. Les gènes entre parenthèse sont à proximité de l'insertion et leur régulation pourrait être affectée bien que l'ORF soit intacte. Les fonctions de ces gènes sont listées, les points d'interrogation signalent des fonctions hypothétiques. Les couleurs indiquent les voies physiologiques dans lesquelles interviennent les produits de ces gènes : la digestion du lactose en gris ; la voie σE-dépendante en jaune ; la voie σS-dépendante en rouge ; la biosynthèse de sucres en vert ; à la fois la voie σS-dépendante (rssB) et la biosynthèse de sucres (galU) en orange. Enfin gènes pour lesquels aucune voie précise n'a pu être établie sont indiqués en bleu.

Les activités β-galactosidase spécifiques des fusions P1-lacZ (souche JJ0159), P1-P2-lacZ (souche JJ0161) ainsi que les activités dans le contexte des différents mutants isolés ont été mesurées à phase exponentielle de croissance (E), après une culture sur la nuit (O/N), ou bien sur des clones de tailles similaires prélevés sur boîte de Pétri (C), et les rapports d'activité β-galactosidase (facteurs de répression) entre souche wt et souche mutante ont été calculés et indiqués ici. Ces rapports sont mis en relief par un gradient de couleur rouge ou verte respectivement lorsque l'effet de la mutation sur l'activité de la fusion est négatif ou positif, les couleurs les plus vives correspondant à l'effet le plus intense.

Les milieux sur lesquels les clones ont été isolés sont indiqués (T : tétrazolium, M : MacConkey, X : X-gal). Le phénotype associé à ces mutations est indiqué par une couleur rouge ou verte selon que l'activité β-galactosidase semble affectée négativement (-/--) ou positivement (+/++). Les clones 220 et 227 (waaQ), d'une part, et les clones 167 et 217 (lacY), d'autre part, portent la même mutation et sont vraisemblablement issus d'un seul clone d'origine.

De manière surprenante, un certain nombre de clones isolés comme activateurs sur boîte apparaissent répresseurs en culture liquide, et réciproquement. Cela pourrait être dû à une différence de physiologie bactérienne entre croissance en milieu liquide ou solide, ou bien en fonction des divers milieux utilisés. Nous avons néanmoins poursuivi l'analyse de ces

clones. On remarque ainsi que trois gènes de l'opéron *waa* (précédemment nommé *rfa*) ressortent ici, à savoir *waaG*, *waaA* et *waaQ*. Ces gènes sont impliqués dans la biogenèse des LPS, et leur mutation entraîne un défaut des LPS qui déclenche la réponse σE-dépendante (voir **Introduction III.B.2.a**) (Lima et al., 2013). Pour cette raison, mais aussi parce que RseX a été isolé comme un suppresseur multicopie de la délétion du gène *degS* qui empêche la réponse σE-dépendante, nous avons rapidement émis l'hypothèse que σE puisse affecter d'une manière ou d'une autre l'expression de RseX.

Etant donné que la surexpression d'*ompX* entraîne également la réponse σE-dépendante (Mecsas et al., 1993), j'ai analysé l'influence d'une telle surproduction ainsi que de la surproduction de *rpoE* sur l'expression des fusions *P1-lacZ* et *P1-P2-lacZ* sans observer le moindre effet (résultats non montrés ici). Bien que σE et les gènes régulés par ce facteur σ semblent de ce fait exclus pour expliquer la régulation de *rseX*, il se pourrait que cette réponse soit plus spécifiquement véhiculée par le signal d'anomalie de biogenèse des LPS, qui pourrait être indépendant par certains aspects du signal d'anomalie des OMP (Lima et al., 2013) tels que celui généré par la surexpression d'*ompX*. De plus, quatre autres gènes candidats de ce crible (*i.e. galU*, *pfkA*, *yegV* et *yajO*) sont (ou pourraient être) impliqués dans la biogenèse de sucres. Or, ces sucres peuvent servir de précurseurs à la biogénèse des LPS, ce qui renforce l'idée que la biologie des LPS et le stress membranaire lié aux LPS pourraient avoir un lien avec la régulation de RseX. Également, ce signal pourrait transiter par Rcs, un autre système qui semble important pour l'intégration du stress lié aux LPS (Klein et al., 2016). Enfin, d'autres pistes intéressantes résultant de ce crible pointent vers d'autres régulateurs, tels que *rssB* qui pourrait réguler *rseX* directement ou *via* un effet sur σS, ou encore *cspB* qui pourrait avoir un effet sur *rseX* en réponse au froid – qui correspond également à un stress membranaire et à un signal véhiculé par σS.

La régulation de l'expression *rseX* reste donc un mystère, mais ce crible génétique ouvre peut-être la voie à une étude et une compréhension plus détaillées de cette régulation, qui pourrait bien répondre d'une manière ou d'une autre – et comme de nombreux autres sRNA Hfq-dépendants – à un signal de stress membranaire. Ce type de stress pourrait ainsi entraîner la répression de l'expression de *fepA*, entre autres gènes cibles, *via* l'effet de cet ARN régulateur.

Discussion

A. Eléments de réflexion sur la traduction et les ARNm

1. Mécanismes de régulation de la traduction reposant sur une structure secondaire activatrice de la traduction

a. Le modèle du starting block

L'une des principales découvertes faites au cours de cette thèse a été l'identification et la caractérisation de structures en tige-boucle situées dans la séquence codante de deux ARNm, en aval du RBS, favorisant la fixation du ribosome et le démarrage de la traduction. Si le mécanisme de régulation impliqué n'est pas encore compris, nous proposons deux modèles dans notre article, que je me propose de commenter et d'argumenter ici.

Le premier modèle proposé, qui me semble le plus convaincant, est celui du « starting block » (Fig. 72). Ce modèle propose que la sous-unité ribosomique 30S « analyse » la séquence de l'ARNm en glissant par diffusion linéaire le long de ce dernier, à la recherche du RBS – par souci de clarté, j'associerai ce phénomène d'analyse au terme de « scan ». Dans le cas de *fepA* et de *bamA,* et peut-être d'autres ARNm, ce RBS est riche en suites de nt A qui pourraient provoquer un glissement excessif de la sous-unité 30S du ribosome lors de ce scan. Le RBS de *bamA* est par ailleurs très inhabituel dans la mesure où 16 nt séparent la séquence SD la plus probable (et conservée) du codon de démarrage de la traduction, ce qui pourrait défavoriser la formation du complexe de pré-démarrage de la traduction. Dès lors, une structure en aval du RBS pourrait empêcher la sous-unité 30S de poursuivre son scan de l'ARNm au-delà du RBS. Cela favoriserait l'interaction SD/anti-SD, la reconnaissance du codon de démarrage par l'anticodon de l'ARNt^{fMet} et de ce fait la formation du complexe de pré-démarrage.

La découverte récente de l'existence de sites de réserve (ou « standby site ») dans plusieurs ARNm d'entérobactéries (Balbontín et al., 2016; Darfeuille et al., 2007; Espah Borujeni et al., 2014) plaide fortement en faveur d'un tel scan de la sous-unité 30S le long de

l'ARNm. En effet, il existerait notamment au sein du 5'UTR des ARNm *iroN* de *S. enterica* et *tisAB* d'*E. coli* un site de fixation du ribosome en amont de la région de démarrage de la traduction (ou TIR pour « translation initiation region »), qui semble important pour l'expression de ces gènes. Dès lors, il me semble cohérent de proposer qu'une sous-unité 30S fixée au site de réserve d'un ARNm va simplement glisser le long de l'ARNm afin d'atteindre le TIR, en le scannant pour détecter un SD, et ce certainement en présence de l'ARNt^{fMet} qui permettra au complexe de détecter également la présence d'un codon de démarrage.

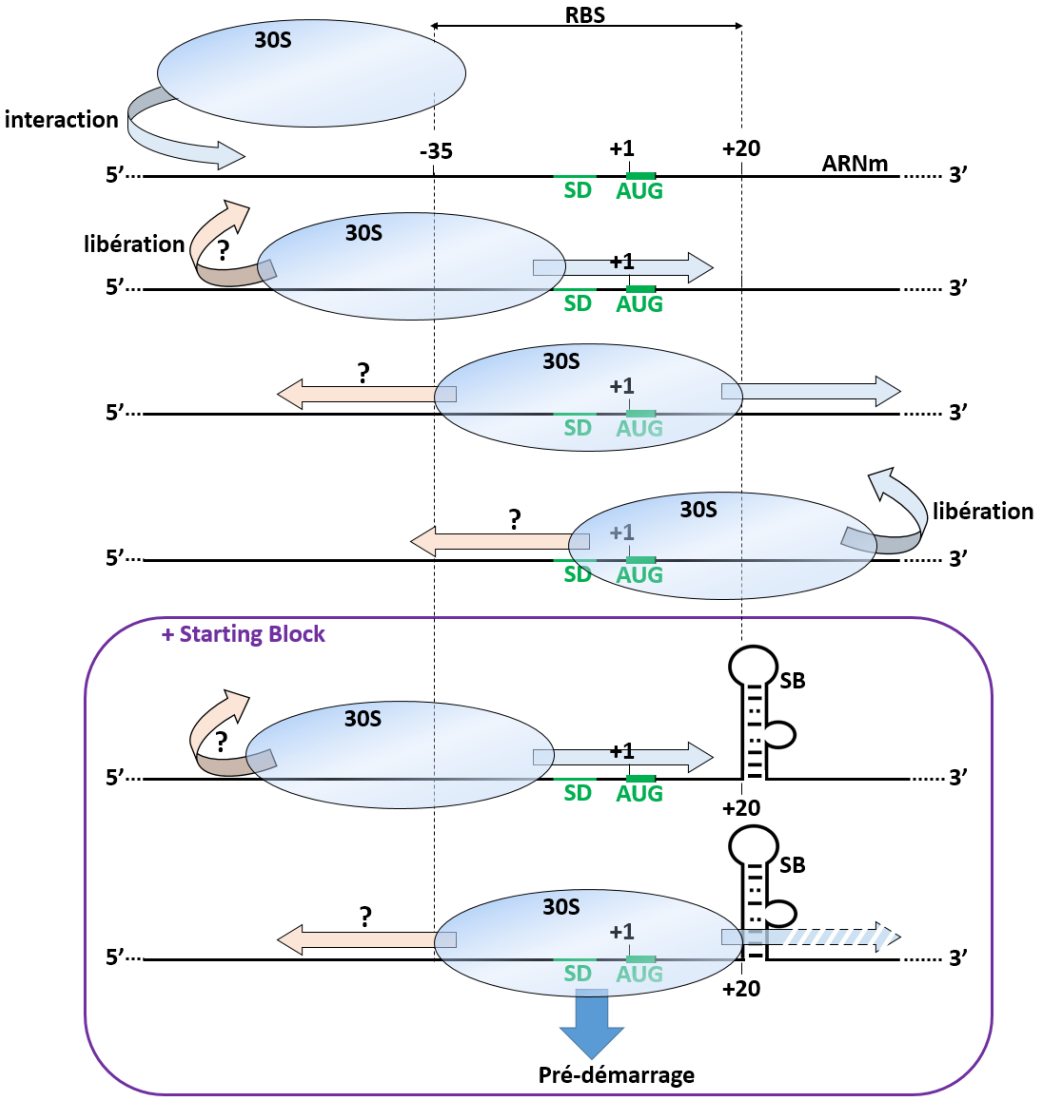


Figure 72 – Le modèle du « Starting Block »
Après une fixation aspécifique de la sous-unité 30S au 5'UTR de l'ARNm, celle-ci va scanner l'ARNm en direction du 3' (flèches bleues) et éventuellement en direction du 5' (flèches roses), possiblement par diffusion linéaire. Certaines séquences, telles que des suites de A, pourraient favoriser le glissement de la sous-unité 30S à travers le RBS. Si la sous-unité 30S n'a pu se stabiliser au niveau d'un RBS pour former un complexe de pré-démarrage de la traduction, elle peut se dissocier de l'ARNm. En présence d'un « Starting Block » (encadré violet), la structure SB va empêcher tout glissement en direction du 3',

favorisant la stabilisation de la sous-unité 30S du ribosome au niveau du RBS et la formation du complexe de pré-démarrage de la traduction.

S'il n'est pas démontré que la sous-unité 30S bactérienne effectue ce genre de glissement et de scan de l'ARNm, il s'agit d'une hypothèse crédible pour expliquer le fonctionnement des sites de réserve (Espah Borujeni et al., 2014). La démonstration de l'existence d'un tel scan est certainement nécessaire pour valider le modèle de « Starting Block ». Notons toutefois qu'il s'agit du principal mode de fonctionnement chez les organismes eucaryotes, où le scan de l'ARNm par la sous-unité ribosomique 40S associée à l'ARNt de démarrage et à certains facteurs s'effectue systématiquement à partir de l'extrémité 5' de l'ARNm (Hinnebusch, 2011). Soulignons par ailleurs que ce type de glissement et de scan des régions de démarrage de la traduction semble exister également chez les bactéries, mais avec la forme 70S du ribosome lors de la traduction d'ARNm polycistroniques (Yamamoto et al., 2016). Il est donc possible d'imaginer que le même type de mécanisme soit utilisé par la sous-unité 30S bactérienne lors d'un évènement de démarrage de la traduction à l'extrémité 5' d'un ARNm. Remarquons ici que la découverte récente de modifications de l'extrémité 5' d'ARN avec des groupements NAD (Cahová et al., 2014), NADH et desphospho-CoA (Bird et al., 2016) suggère qu'une grande variété de ces modifications existe *in vivo*. Ces modifications pourraient par exemple permettre une discrimination entre des ARNm dont la séquence est scannée par la sous-unité 30S à partir de l'extrémité 5', et des ARNm sur lesquels la sous-unité 30S se fixe directement au RBS.

Enfin, dans l'hypothèse plus spéculative où la sous-unité 30S du ribosome scannerait l'ARNm dans les deux sens à partir d'une région interne de ce dernier, une prédiction possible voudrait qu'une structure en amont du RBS puisse également activer la traduction dans certaines circonstances, en empêchant un glissement en direction du 5'. Or, les tiges-boucles en amont du RBS trouvées au sein des ARNm *dusB-fis* et *csgD*, discutées plus tôt (voir **Introduction II.B.2.c**), ont des effets activateurs non expliqués à l'heure actuelle. Celles-ci pourraient favoriser le démarrage de la traduction en bloquant une sous-unité 30S scannant l'ARNm en direction de l'extrémité 5'.

b. Le modèle du recrutement

Le second modèle que nous proposons est celui du recrutement de la sous-unité 30S par la structure SB (Fig. 73). Étant donné les changements qui peuvent être introduits dans les structures SB de *fepA* et de *bamA*, tout en conservant leur effet activateur, il est difficile d'imaginer qu'une reconnaissance spécifique d'une séquence ou d'une topologie particulière de la structure SB par la sous-unité 30S puisse avoir lieu. Néanmoins, cette reconnaissance pourrait reposer sur une interaction avec le squelette sucre-phosphate de la structure SB, sans spécificité de séquence. D'autres éléments de reconnaissance pourraient toutefois se situer dans le RBS, puisque l'effet de ces structures semble dépendant du contexte (voir **Résultats I.C.2**) : on peut ainsi imaginer que l'association des suites de nt A, évoquées précédemment, à une structure SB située immédiatement en aval, constitue l'ensemble reconnu par la sous-unité 30S. Rappelons toutefois que ce contexte pourrait aussi bien avoir un rôle important dans le cadre du modèle du starting block.

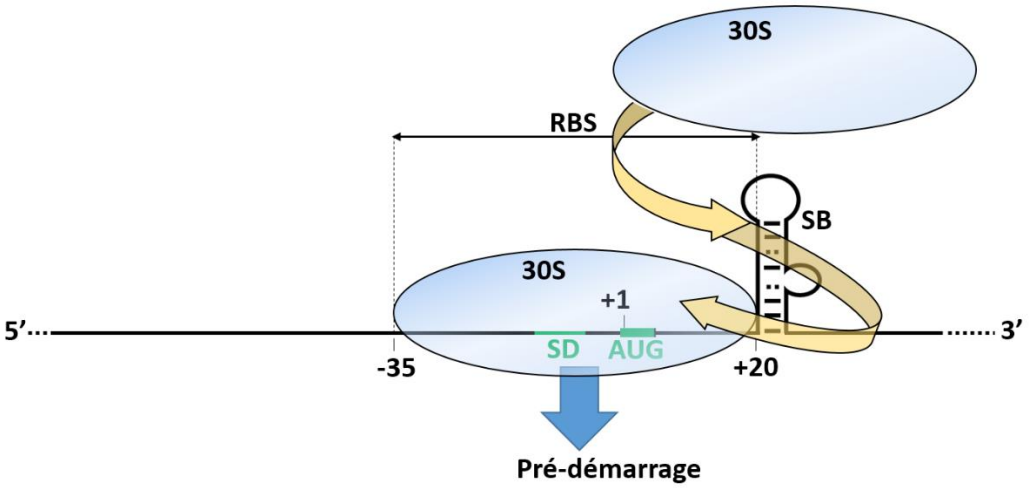


Figure 73 – Modèle de recrutement de la sous-unité ribosomique 30S
La sous-unité 30S reconnaît la structure SB directement, ou via l'intervention d'une protéine ou d'un ARN qui reste à identifier. Cette interaction permettrait de recruter la sous-unité 30S à proximité du RBS, aboutissant à la formation du complexe de pré-démarrage de la traduction.

Pour démontrer le modèle de recrutement, il faudrait donc commencer par comprendre comment la sous-unité 30S entre en interaction avec la structure SB. Puisque l'effet de la structure SB est observable *in vitro* dans des réactions de toeprint (voir **Article**), tout ce qui est nécessaire à la fixation du ribosome est vraisemblablement inclus dans ces réactions de toeprint. On peut par exemple imaginer un rôle de la protéine ribosomique S2, impliquée dans la reconnaissance d'éléments en aval du codon de démarrage de la traduction (voir Fig. 12 et **Introduction II.C.1.a**). Si l'intégralité des protéines ribosomiques interagissent

avec l'ARN, et pourraient donc être *a priori* considérées comme des candidates pour l'établissement de cette interaction, notons toutefois que l'hypothèse de l'existence d'une plateforme de reconnaissance du RBS d'ARNm structurés, composée des protéines S2, S7, S11, S18 et S21 (Marzi et al., 2007), pourrait constituer une piste sérieuse pour la validation du modèle de recrutement.

Par ailleurs, certaines régions de l'ARNr 16S semblent se situer à proximité de cette région en aval du codon de démarrage de la traduction de l'ARNm au sein du complexe de pré-démarrage (à l'image des positions 532 et 1395 de l'ARNr 16S, voir **Introduction** Fig. 12), et pourraient également participer à la reconnaissance de la structure SB. Néanmoins, le fait que l'effet des structures SB soit complètement indépendant de la séquence rend cette hypothèse assez improbable, dans la mesure où l'interaction ne pourrait pas reposer sur un appariement de séquences. Notons enfin que d'autres facteurs non ribosomiques pourraient intervenir dans le recrutement du ribosome. Dans ce cas, ces facteurs ARN et/ou protéiques ont pu être purifiés fortuitement avec les sous-unités 30S destinées aux réactions de toeprinting. On pense notamment, et inévitablement, à la protéine Hfq, qui pourrait interagir avec ces structures et les nt A situés en amont. Alternativement, ces facteurs pourraient être importants au fonctionnement des structures SB *in vivo* mais pas *in vitro*.

Enfin, une interaction directe ou indirecte *via* un facteur inconnu entre la sous-unité 30S et la structure SB n'exclurait pas pour autant le modèle du « starting block », puisque cette interaction pourrait intervenir après une première étape de scan. De même, un facteur en *trans* pourrait simplement stabiliser cette structure secondaire de l'ARNm afin d'en faire un starting block plus efficace. Si une telle interaction a lieu, la question suivante pourrait se poser : l'interaction entre la sous-unité 30S et la structure SB intervient-elle avant (modèle de recrutement) ou après (modèle de starting block) que la sous-unité 30S ait atteint le site du TIR? Nous espérons à l'avenir définir avec plus de précision le ou les mécanismes impliqués, notamment grâce à la compréhension de cette éventuelle interaction entre sous-unité 30S et structure SB. De plus, la découverte de structures SB au sein d'autres ARNm devrait nous permettre d'obtenir plus d'informations sur les éléments nécessaires au fonctionnement de ces structures.

2. Structures SB : où les trouver ?

De nombreuses questions se posent au sujet des structures SB, dont voici un échantillon : à quel point sont-elles répandues ? Sont-elles spécifiques d'une classe d'ARNm en particulier ? Quels sont les éléments en *cis* des ARNm importants pour leur effet ? Sont-elles des cibles privilégiées pour la régulation par des sRNA ?

S'il est difficile de fournir une réponse à ces questions à l'heure actuelle, l'identification et la validation d'un plus grand nombre de structures SB devrait permettre d'apporter plus d'éclaircissements. Nous avons ainsi débuté l'analyse bioinformatique des structures d'ARNm trouvées en aval du codon de démarrage de la traduction afin d'essayer de relever un biais de structure aux positions +17 à +25 des ARNm *d'E. coli*, ou au sein d'une sous-population d'ARNm. Le fait que les gènes *bamA* et *fepA* codent tous deux des OMP nous a mis sur la piste des ARNm d'OMP en particulier. Cela étant dit, nous avons pu vérifier que des structures d'ARNm prédites à environ 20 nt en aval du codon de démarrage, au sein des ARNm *lptD* et *lpp* codant respectivement une OMP essentielle et une lipoprotéine très abondante (voir **Introduction III.B.2.a**), n'ont pas d'influence particulière sur l'expression de fusions traductionnelles *lptD-lacZ* et *lpp-lacZ* (résultats non montrés ici).

C'est pourquoi l'étude de la conservation des diverses structures secondaires que nous pourrons prédire nous sera utile pour sélectionner les structures les plus conservées et le plus à même d'avoir une fonction. Une fois plus de structures SB validées par l'expérience, nous pourrons affiner notre recherche des éléments en *cis* impliqués dans la régulation, tels que les suites de nt A discutées auparavant. Notons également que cette étude pourrait révéler différentes « classes » de structures SB, affectant la traduction *via* des mécanismes différents, et dont les éléments en *cis* importants pourraient également varier. Cela permettrait peut-être d'expliquer pourquoi la structure SB de *fepA* ne peut remplacer celle de *bamA* pour la traduction de *bamA* (voir **Résultats I.C.2**)

Enfin, il serait intéressant de savoir si ces structures SB sont des opérateurs classiques pour la régulation par des sRNA. À ce titre, une première expérience à réaliser serait de

rechercher des sRNA régulateurs de l'expression de *bamA*. En effet, certains sRNA pourraient empêcher la formation de la structure SB de *bamA* en s'y appariant, à la manière des OmrA/B avec la structure SB de *fepA*. Ces sRNA pourraient notamment s'apparier au niveau du bulge situé en 3' de la structure SB de l'ARNm *bamA* (voir **Article** Fig. 7), qui évoque subtilement le bulge indispensable à l'appariement d'OmrA/B à la structure SB de l'ARNm *fepA* (voir **Article** Fig. 5).

3. La définition du RBS

La définition classique du RBS fait référence à la région couvrant les nt -35 à +20 de l'ARNm à laquelle se fixe la sous-unité 30S afin de démarrer la traduction (voir Fig. 18). Cette définition n'inclut pas d'éventuelles régions qui pourraient fortement influencer cette fixation voire même participer au RBS, à l'image des structures SB décrites dans ce mémoire. Notre vision du RBS mérite donc d'être revisitée. On pourrait ainsi faire valoir qu'en l'absence des structures SB, le site de fixation du ribosome des ARNm *fepA* et *bamA* est incomplet, puisque cela entraîne une forte diminution de la fixation du ribosome. La question se pose donc de l'extension de la définition du RBS aux structures SB dans le cas de ces deux ARNm. Il est de plus tout à fait imaginable que d'autres régions de certains ARNm, peut-être même très éloignées au niveau de la structure primaire, puissent entrer en contact avec le ribosome au gré des repliements de structure de l'ARNm et avoir une grande importance pour la fixation du ribosome. On pourrait alors définir le RBS comme l'ensemble des régions de l'ARNm nécessaires et suffisantes à une fixation optimale du ribosome.

4. L'ARNm et le code de régulation génétique

Une idée latente depuis plusieurs décennies (Beato, 1985), voire même depuis l'émergence des notions de promoteur et de programme génétique (Jacob et al., 1964) est qu'il pourrait exister un code nucléotidique de la régulation de l'expression génétique, qui se superposerait au code génétique déchiffré dans les années soixante. Ce code de la régulation génétique comprendrait l'ensemble des combinaisons de séquence et de structure qui ont un impact sur le niveau d'expression des gènes, sans affecter le message du code génétique. Cette idée d'un code de régulation génétique a fait l'objet d'intéressantes réflexions notamment en ce qui concerne la régulation transcriptionnelle (Dresch and Arnosti, 2015;

Kelley et al., 2016), sur laquelle je ne m'attarderai pas. La régulation post-transcriptionnelle sur laquelle je souhaite plus particulièrement mettre l'accent n'est pas en reste, avec notamment l'étude de l'impact (i) de l'usage en codons des ARNm chez les entérobactéries (Boël et al., 2016; Bossi and Roth, 1980), (ii) de la structure de l'ARNm également chez les entérobactéries (Del Campo et al., 2015; Righetti et al., 2016; Burkhardt et al., 2017), mais aussi (iii) de chaque nt d'un ARNm du poisson-zèbre par exemple (Yartseva et al., 2016) sur l'expression génétique. Ces études établissent à chaque fois un cadre plus précis pour ce code de régulation génétique chez une espèce donnée. En effet, contrairement au code génétique qui est universel, ce code de régulation serait beaucoup moins conservé et dépendrait de l'ensemble des mécanismes de régulation particulier à chaque organisme.

Cette question est d'autant plus d'actualité aujourd'hui avec la découverte d'un nombre croissant de régions de l'ARNm impliquées dans la régulation de l'expression génétique. On peut citer notamment les thermosenseurs et les riboswitches (voir **Introduction II.C.2**) mais aussi les sites d'appariement avec des sRNA. Ces sites d'appariement peuvent parfois être très nombreux sur un même ARNm comme l'illustrent les cas très étudiés de *csgD* et de *flhDC* (voir **Revue n°2**). De plus, la même région d'un ARNm peut réunir des sites d'interaction avec plusieurs sRNA comme c'est le cas pour l'ARNm *rpoS* avec les sRNA DsrA, RprA et ArcZ (Mandin and Gottesman, 2010; Soper et al., 2010). Par ailleurs, si le 5'UTR est confirmé année après année comme une région importante réunissant de nombreux éléments de régulation de l'expression génétique chez les bactéries, l'ORF des ARNm mais aussi leur 3'UTR se révèlent être également des régions importantes pour cette régulation (voir **Revue n°1**)

L'ARNm *fepA* est un nouvel exemple d'ARNm réunissant à la fois de multiples sites d'interaction avec divers sRNA, et au moins une structure activatrice d'une grande importance pour l'expression génétique comprenant également un site d'interaction avec les deux sRNA OmrA et OmrB. Il s'agit également d'un exemple d'ARNm régulé à la fois *via* son 5'UTR et de sa région codante. On retrouve donc tout le long de cet ARNm des éléments de structure primaire et secondaire qui pourraient correspondre au code de régulation génétique de *fepA* – et dont la mutation altère l'expression de ce dernier.

Gageons que la découverte d'un plus grand nombre de structures SB, de riboswitches et de sites d'appariements de sRNA nous permettra de nous diriger vers une ère de plus en plus prédictive, où une simple analyse des structures primaire et secondaire d'un ARNm pourrait apporter de nombreuses informations sur la régulation du gène correspondant. En ce sens, de nombreux algorithmes prédisent déjà les sites d'appariement de sRNA (Busch et al., 2008; Tjaden et al., 2006; Wright et al., 2013) ou la présence de certains riboswitches (Millman et al., 2017; Mukherjee and Sengupta, 2016). Néanmoins, ces algorithmes doivent encore être améliorés, affinés et complétés par les informations provenant des récentes et futures découvertes de nouvelles séquences et structures importantes pour la régulation de l'expression génétique. Au-delà de la compréhension des régulations qui régissent l'expression de chaque gène, cette capacité de prédiction aurait une portée considérable en permettant des modélisations de plus en plus sophistiquées des interactions génétiques qui s'établissent au sein des organismes.

B. De multiples régulations par les ARN

1. L'accumulation de signaux pour la régulation de l'expression de *fepA*

L'expression de *fepA* est régulée au niveau post-transcriptionnel par au moins quatre sRNA Hfq-dépendants, (voir **Résultats II.A**), chacun de ces sRNA répondant à divers stimuli. Ces régulations s'ajoutent à la régulation transcriptionnelle par Fur en réponse aux niveaux de fer, et peut-être aussi par d'autres régulateurs répondant à la phase de croissance (voir **Résultats II.A.1.b**). Enfin, rappelons que la structure SB répond non seulement aux OmrA/B, mais également potentiellement à la température (voir **Résultats I.C.3**). L'ensemble de ces signaux et régulateurs ont été réunis dans un schéma synthétique (Fig. 74). Ainsi, il est étonnant de constater que l'expression génétique de *fepA* répond vraisemblablement à une multitude de stimuli différents. Cela pourrait indiquer que les niveaux de la protéine FepA nécessitent une régulation d'une grande finesse pour permettre l'adaptation à une grande variété de conditions environnementales, la quantité de protéine adéquate à chaque condition étant synthétisée afin d'optimiser l'import de fer tout en limitant ses effets toxiques.

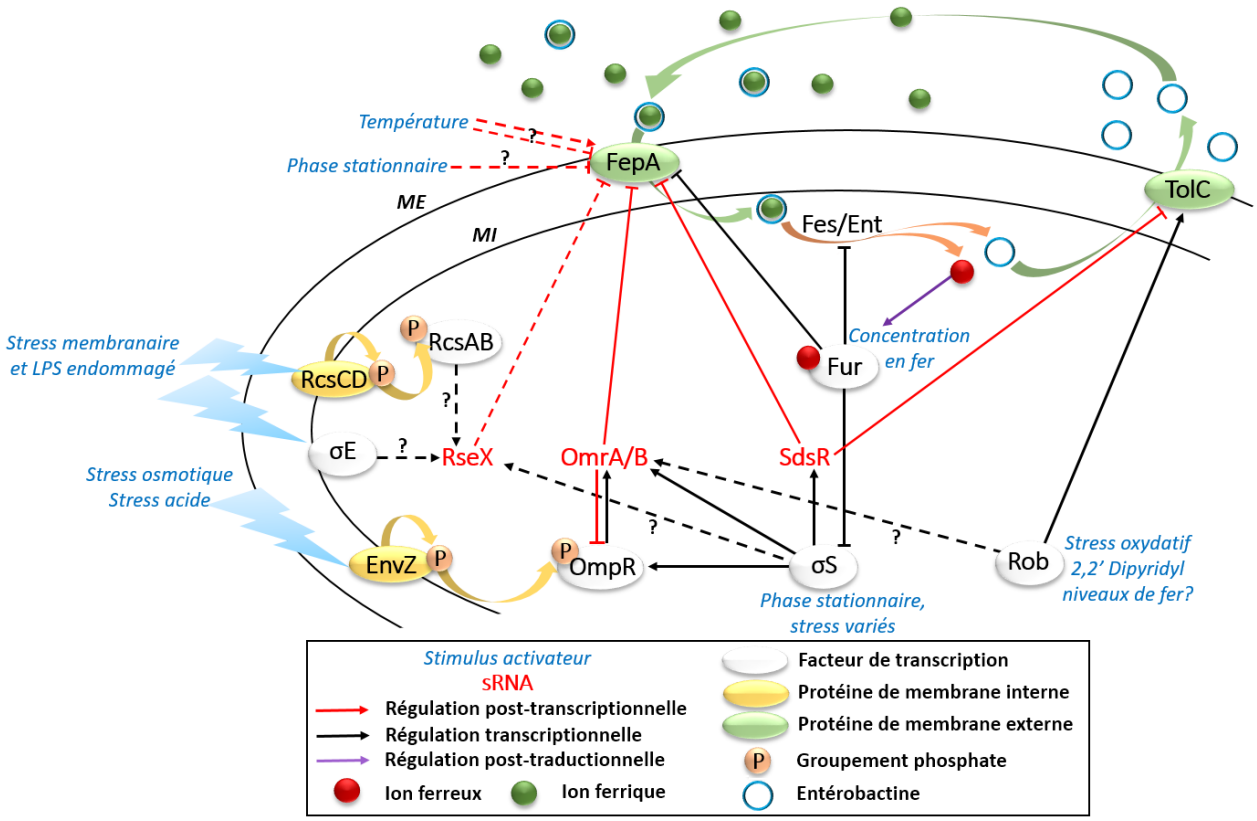


Figure 74 – La régulation de l'expression de fepA

Certains éléments de cette figure ont été présentés dans l'introduction (Fig. 44). Les flèches en pointillés indiquent une régulation qui peut être directe ou indirecte, tandis qu'un point d'interrogation signifie que la régulation concernée n'est pour l'instant qu'une hypothèse formulée à partir des résultats obtenus au cours de cette thèse, qui reste à démontrer. En présence de fer, Fur-Fe^{2+} est actif et réprime l'expression de fepA. En condition de carence en fer, cette répression est levée et les complexes fer-entérobactine peuvent être importés dans la cellule via le récepteur FepA. Les régulateurs OmpR et σS répondent à divers stress et activent la transcription des sRNA OmrA, OmrB et/ou SdsR, qui répriment quant à eux l'expression de fepA au niveau post-transcriptionnel. SdsR réprime également la synthèse de TolC, inhibant ainsi l'export de l'entérobactine, tandis que le FT Rob active sa synthèse (Zhang et al., 2008). Par ailleurs, le dip ou encore les ions Fe^{2+} activent la synthèse d'OmrA/B au niveau transcriptionnel, signal qui pourrait être intégré par Rob. Enfin, le sRNA RseX régule également l'expression de fepA au niveau post-transcriptionnel, et pourrait lui-même répondre à un défaut du LPS, peut-être par le biais d'une régulation par σE, par le système Rcs et/ou par σS.

De façon intéressante, le risque de stress oxydatif pourrait être exacerbé lorsque la bactérie subit un stress membranaire, rendant la cellule plus perméable à des molécules oxydantes par exemple. Or les sRNA OmrA/B et RseX pourraient répondre à divers types de stress membranaire et réprimer la synthèse de FepA y compris dans des situations de carence en fer, lorsque le risque qu'une réaction de Fenton s'enclenche surpasse le bénéfice de l'import de fer par FepA. Le fait que σS, qui active l'expression de SdsR, d'OmrA et peut-être aussi de RseX, réponde également au stress oxydatif (voir **Introduction III.B.2.b**) renforce cette hypothèse. Les sRNA pourraient ainsi être un moyen rapide et efficace de réguler l'expression de *fepA* lorsque la répression par la protéine Fur est levée et qu'une répression

par une autre protéine régulatrice serait trop longue à établir par dans des conditions environnementales défavorables. Les différences entre régulations par des sRNA et par des protéines constituent en soi une question très intéressante, que je ne développerai pas ici car elle est longuement discutée dans la **Revue n°2**.

Enfin, la régulation de l'expression de *fepA* en fonction de la phase de croissance reste à comprendre, et si les sRNA RseX, OmrA/B et SdsR jouent un rôle dans cette régulation, d'autres régulateurs pourraient être impliqués (voir **Résultats II.A.1.b**). On pourrait notamment penser à un répresseur transcriptionnel dont l'expression serait activée par σS en phase stationnaire de croissance, dont l'effet sur *fepA* resterait à découvrir.

2. L'évolution et la régulation par les ARN

Je voudrais à présent inviter le lecteur à une digression concernant l'évolution. En biologie, l'une des idées les plus intrigantes évoquées au cours du 20ème siècle est peut-être celle du « gène égoïste », conçue par Richard Dawkins. Ce dernier propose ainsi une alternative à la définition communément admise du gène comme unité de base de l'hérédité et constitué d'une séquence d'ADN (ou d'ARN pour certains virus à ARN) codant une molécule fonctionnelle (ARN, protéine ou autre). Selon R. Dawkins, cette définition devrait être poussée à l'extrême dans le cadre du gène égoïste : le gène resterait certes l'unité de base de l'hérédité, mais serait réduit à toute séquence d'ADN ou d'ARN en mesure de promouvoir ses propres propagation et conservation évolutive, que ce soit au sein d'un génome donné ou à travers différents génomes (ou pour citer directement R. Dawkins : «un gène peut être défini comme une portion de matériel chromosomique qui dure potentiellement pendant un nombre suffisant de générations pour servir d'unité de sélection naturelle»).

Pour donner un exemple concret, les séquences consensus des boîtes -35 et -10 des promoteurs seraient des « gènes » extrêmement conservés et répandus non seulement au sein de chaque génome où on les trouve, mais aussi à travers les génomes de nombreuses espèces. Leur caractère « égoïste » tient de leur réussite évolutive, qui a bien entendu été contrainte par la conservation de l'ARNpol et de la machinerie de transcription. Parce que ces séquences consensus sont indispensables au fonctionnement des cellules, leur conservation est assurée, et chaque fois qu'une nouvelle unité transcriptionnelle apparaît dans un génome

(que ce soit par duplication génétique, ou par l'arrivée de matériel génétique parasite), son fonctionnement et sa propre conservation dépendront de la présence de ces séquences consensus promotrices.

La définition couramment admise du gène est sans nul doute la plus adaptée à l'étude et la compréhension des phénomènes de régulation et d'expression génétique, aussi l'appellation « gène égoïste », quoiqu'attrayante, peut prêter à confusion. Néanmoins, ce concept peut être utile pour appréhender l'évolution de la régulation de l'expression des gènes, et plus particulièrement l'évolution des régions fonctionnelles des sRNA d'une part, et des régions régulatrices des ARNm d'autre part. Je parlerai donc à partir de maintenant de régions fonctionnelles et de régions régulatrices plutôt que de « gène égoïste ». Pour prendre l'exemple d'un gène de sRNA Hfq-dépendants, sa région promotrice et deux régions du sRNA transcrits à partir de ce gène sont très conservées, à savoir la région qui s'apparie aux cibles ARNm et le domaine de fixation à Hfq (voir **Introduction II.C.3.b**). Ces trois régions, bien qu'interdépendantes, sont des régions fonctionnelles qui ont évolué sous des pressions de sélection différentes et sans doute également par dérive génétique au moins au début de leur processus évolutif.

On peut tenter de retracer ici le scénario de leur évolution. Si l'on prend en considération qu'une grande partie du génome, y compris dans des régions intergéniques, est transcrite à partir de promoteurs cryptiques, une faible région promotrice est certainement apparue en premier lieu par simple dérive génétique. Une région du sRNA transcrit ainsi, qui pouvait peut-être initialement interagir avec une seule cible ARNm, a alors probablement évolué de concert avec la région ciblée de ces ARNm, au gré des avantages tirés des régulations génétiques qui en ont découlé. La conservation de cette région d'appariement aux ARNm a également dépendu des conditions environnementales dans lesquelles le sRNA était exprimé, ce qui nous conduit à l'évolution plus profonde de la région promotrice. Celle-ci a pu répondre à une forte pression de sélection pour fonctionner avec la machinerie transcriptionnelle, tandis que des régulations par divers facteurs de transcription ont pu être sélectionnées quand un avantage a été associé à ces régulations. Enfin la région de fixation à Hfq a certainement évolué au fur et à mesure, en réponse à une pression de sélection différente, à savoir celle de la stabilité du sRNA et de son efficacité d'action. S'il est difficile de

savoir à quel rythme ont eu lieu ces diverses évolutions, réfléchir à l'évolution séparée de chaque région fonctionnelle du gène de sRNA permet ainsi de retracer une histoire hypothétique de l'évolution de cet ARN régulateur et de son régulon. Chacune de ces régions fonctionnelles aurait donc évolué en réponse à une pression de sélection différente, et une fois assemblées, leur dépendance réciproque aurait abouti à la formation d'une machinerie de régulation cohérente. Une fois cette machinerie formée, la conservation de l'ensemble des régions a pu être assurée sur une échelle de temps plus grande. Par la suite, l'appariement de ces sRNA à de nouvelles cibles ARNm a pu résulter de l'évolution de ces dernières pour correspondre au site de fixation du sRNA. Une prédiction intéressante de ce scénario serait qu'en moyenne, plus un sRNA est ancien, plus il a de cibles différentes.

En se plaçant maintenant du point de vue de l'ARNm, il est intéressant de mettre en perspective l'évolution des sites de fixations des sRNAs. Pour prendre l'exemple d'OmrA/B, il est possible de générer une séquence consensus convaincante correspondant au site de fixation de l'extrémité 5' d'OmrA/B (Fig. 75). Il existe donc une pression de sélection pour l'apparition et la conservation d'une séquence (ACNUCUGGG) au sein des divers ARNm-cibles, dont la régulation post-transcriptionnelle confère un avantage à la cellule dans les conditions où OmrA/B sont exprimés. Pour soutenir la prédiction évoquée ci-dessus, il convient de remarquer que ces régions de fixation d'OmrA/B ont très certainement évolué de manière convergente, puisque la plupart des cibles ARNm d'OmrA/B sont bien plus conservées évolutivement (et donc anciennes) que les OmrA/B eux-mêmes. Voilà donc une façon de faire émerger une région régulatrice, correspondant ici au site de fixation des sRNAs, consécutive à l'apparition d'un ARN régulateur fonctionnant par appariement de séquence. L'évolution de cette région régulatrice dépendra ainsi de celle de l'extrémité 5' des OmrA/B, mais aussi du contexte dans lequel elle se situe dans l'ARNm, comme par exemple les RBS des ARNm *ompT*, *ompR*, *cirA* et *flhDC*, ou encore la structure SB de *fepA*, qui doivent notamment permettre la fixation du ribosome.

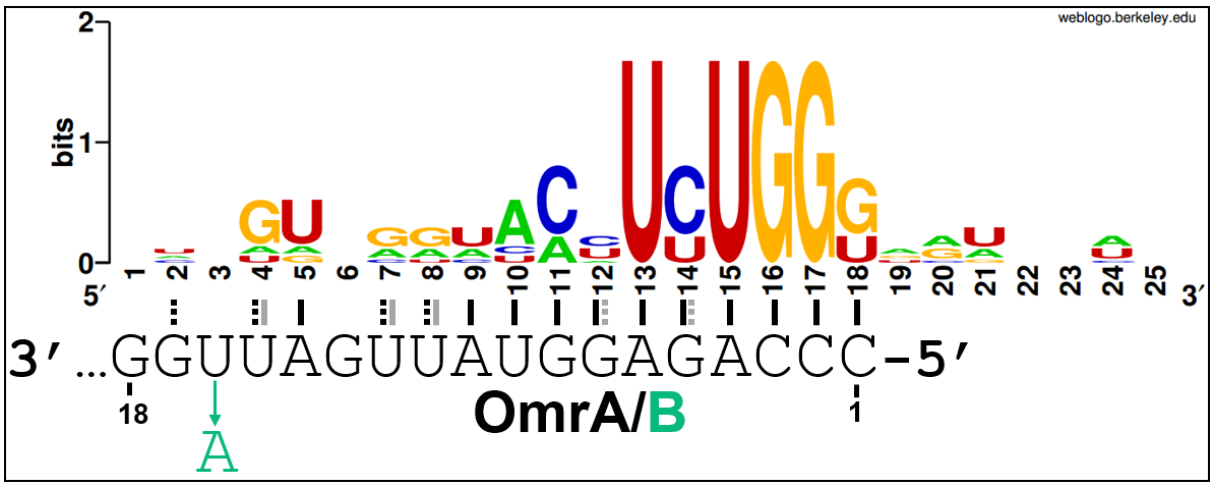


Figure 75 – La séquence consensus du site de fixation d'OmrA/B
La séquence consensus du site de fixation d'OmrA/B a été générée à l'aide du programme WebLogo (http://weblogo.berkeley.edu/logo.cgi) (Crooks et al., 2004). Le degré de conservation d'un nt à chaque position est indiqué par une échelle de 0 à 2 bits correspondant respectivement au minimum et au maximum de conservation. Un alignement visuel de 25nt dans la région du site de fixation d'OmrA/B avec des cibles ARNm démontrées (ompT, ompR, cirA, csgD, flhDC, fepA) ou prédites (fecA) a été utilisé pour créer ce motif. Ce dernier peut s'apparier virtuellement avec l'extrémité 5' des OmrA/B comme indiqué, avec en noir l'appariement le plus probable entre la position du motif et les OmrA/B, et en gris un autre appariement possible dans certains cas. Un trait plein représente un appariement Watson-Crick et un trait en pointillés un appariement Wobble. Le nt différent entre OmrA et OmrB est indiqué en vert. Note : un « gap » d'un nt au sein du motif entre les positions 14 et 15, dû à un « gap » d'alignement d'un nt entre l'ARNm cirA et les autres ARNm cibles, a été manuellement retiré ici.

Pour finir, considérons l'évolution des structures agissant sur la traduction, telles que les structures SB décrites ici. Cette fois-ci, la notion de séquence est presque secondaire, puisque c'est la structure secondaire de l'ARNm qui est mise à contribution pour générer un effet régulateur. Il existe donc une pression de conservation sur la structure, tandis que la séquence peut selon les cas évoluer avec beaucoup plus de souplesse tant que des mutations compensatoires permettent à la structure de se former. Nous décrivons un exemple de ce phénomène avec la tige-boucle SB de *bamA*, dont la séquence peut varier drastiquement d'une espèce à l'autre mais dont la structure semble toujours préservée (voir **Article** Fig. S13). Il est difficile à l'heure actuelle de comprendre les rouages évolutifs qui ont fait apparaître ces structures SB, puisqu'il faudrait d'abord comprendre le mécanisme de régulation impliqué. Néanmoins, nous proposons que ces structures SB régulent la fréquence des évènements de démarrage de la traduction (voir **Article**). Cette hypothèse permet d'imaginer que les structures SB sont apparues au sein de certains gènes pour lesquels il est important de réguler le taux de synthèse protéique, et ce malgré la pression de sélection qui existe au sein de la région codante pour la conservation des codons. Ce serait ici le contexte et la protéine synthétisée qui permettrait ou non l'apparition de cette région régulatrice, correspondant non

pas à une séquence de nt précise mais à une contrainte sur la séquence pour permettre la formation d'une structure secondaire à une distance précise en aval du codon de démarrage de la traduction.

Les régions +19 à +45 de l'ORF de l'ARNm *fepA* et +22 à +55 de celle de *bamA* ont ainsi subi une pression de sélection pour la présence d'un peptide signal permettant l'adressage de l'OMP à la membrane externe et d'une structure SB favorisant la fixation du ribosome. La structure SB de *fepA* a en outre subit une pression de sélection supplémentaire pour l'apparition d'un site de fixation des sRNA OmrA/B, permettant la régulation de l'expression de *fepA*. Il s'agit là d'une nouvelle illustration de la formidable plasticité des ARN conférée par l'évolution.

Matériels et Méthodes

A. Souches et milieux

Les souches et plasmides utilisés figurent dans le tableau 2. Ne figurent pas ici les souches répertoriées dans l'article, sauf quand elles ont servi pour la construction de souches dérivées ou comme souche wt dans une expérience. Les souches issues du crible génétique sur le promoteur de *rseX* sont répertoriées dans le tableau 1 (voir **Résultats II.B.2.b**).

Souche	Description	Construction/Source
MG1655	Souche sauvage (wt)	F. Blattner
AC0023	PM1205 P_{bad}-*fepA*$_{-173+45}$-*lacZ*$_{+28}$	A. Coornaert
AC0027	MG1432 *fecA::cat-sacB*	A. Coornaert
DJ480	MG1655 Δ*lacX174*	D. Jin (NIH)
DJ624	MG1655 Δ*lacX174 mal::lacI*^{q}	D. Jin (NIH)
EM1277	DJ480 *rne-3071 zce-726::Tn10*	(Massé et al., 2003)
EM1279	DJ480 *zce-726::Tn10*	(Massé et al., 2003)
IBPC953	N3433 Δ*hfq::tet*	E. Hajnsdorf
MG1003	DJ480 Δ*omrAB::kan*	(Guillier and Gottesman, 2006)
MG1035	DJ480 Δ*ompR::cm*	(Guillier and Gottesman, 2006)
MG1054	DJ480 Δ*fur::kan*	Cette étude (recombinaison)
MG1099	DJ624 Δ*omrAB::kan*	(Guillier and Gottesman, 2006)
MG1100	MG1099 Δ*fur::cm*	(Guillier and Gottesman, 2006)
MG1169	DJ624 Δ*fur::cm*	Cette étude (recombinaison)
MG1178	DJ624 PomrB$_{-250+50}$-*lacZ*	Cette étude (recombinaison)
MG1188	DJ624 PompT10-*lacZ* Δ*omrAB::kan*	(Guillier and Gottesman, 2008)
MG1272	MG1100 *rnc-14::Tn10*	(Guillier and Gottesman, 2008)
MG1432	DJ624 *miniλ-tet*	DJ624 + P1(NM300)
MG1462	NM1200 Δ*omrAB::tet*	Cette étude
MG1508	MG1655 *mal::lacI*^{q} P_{tet}-*cat-sacB-lacZ*$_{+28}$, *mini-λ tet*^{R}	(Coornaert et al., 2013)
MG1511	MG1508 P_{tet}-*phoP*$_{-36+30}$-*lacZ*$_{+28}$	(Coornaert et al., 2013)
MG1690	DJ624 PompF-*lacZ*	(Brosse et al., 2016)
MG1892	DJ624 PompT-*lacZ*	(Brosse et al., 2016)
MG1772	MG1508 P_{tet}-*fepA*$_{-173+45}$-*lacZ*$_{+28}$	Cette étude (recombinaison)
MG2097	NM300 Δ*fepA::amp*	Cette étude (recombinaison)
MG2114	MG1505 rrnBt2-*cat-sacB-lacZ*$_{-17}$	Cette étude (recombinaison)
MG2190	MG1508 P_{tet}-*fepA*$_{-173+45}$-*mutSwitch1-lacZ*$_{+28}$	Cette étude (recombinaison)
MG2213	MG1508 P_{tet}-*phoP*$_{-36-1}$-*fepA*$_{+1+45}$-*lacZ*$_{+28}$	Cette étude (recombinaison)

ML04	MG2114 rrnBt2-PomrB$_{-250+8}$-*lacZ*$_{-17}$	M. Lejars
N3431	N3433 *rne-3071*(ts)	(Goldblum and Apririon, 1981)
N3431x	N3431 Δ*hfq::tet*	E. Hajnsdorf
N3433	Hfr(PO1), *lacZ43*(Fs), *λ*-, *relA1*, *spoT1*, *thiE1*	(Goldblum and Apririon, 1981)
NM300	DJ480 *miniλ-tet*	N. Majdalani
Souche	**Description**	**Construction/Source**
NM1200	MG1655 *miniλ-cm*	N. Majdalani
PM1205	MG1655 P$_{bad}$-*cat-sacB-lacZ*$_{+28}$	(Mandin and Gottesman, 2009)
UGB2497	MG1655 Δ*flu::zeo*	C. Beloin
JJ0006	AC0023 Δ*omrAB::kan*	AC0023 + P1(MG1188)
JJ0015	MG1772 Δ*omrAB::kan*	MG1772 + P1(MG1003)
JJ0025	AC0027 *fecAmut2*	Cette étude (recombinaison)
JJ0049	DJ624 Δ*fecA::cat-sacB*	DJ624 + P1(AC0027)
JJ0057	MG1432 Δ*fepA::gnt*	Cette étude (recombinaison)
JJ0058	MG1432 Δ*cirA::spc*	Cette étude (recombinaison)
JJ0061	DJ624 Δ*cirA::spc*	DJ624 + P1(JJ0058)
JJ0062	JJ0061 Δ*fepA::gnt*	JJ0061 + P1(JJ0057)
JJ0063	JJ0049 Δ*fepA::gnt*	JJ0049 + P1(JJ0057)
JJ0064	JJ0058 Δ*cirA::spc*	JJ0058 + P1(JJ0058)
JJ0065	JJ0064 Δ*fepA::gnt*	JJ0064 + P1(JJ0057)
JJ0066	JJ0065 Δ*omrAB::tet*	JJ0065 + P1(MG1462)
JJ0088	JJ0025 Δ*fepA::gnt*	JJ0025 + P1(JJ0057)
JJ0089	JJ0088 Δ*cirA::spc*	JJ0088 + P1(JJ0058)
JJ0091	JJ0089 Δ*omrAB::kan*	JJ0089 + P1(MG1188)
JJ0110	MG1508 P$_{tet}$-*phoP*$_{-36+19}$-*fepA*$_{+20+45}$-*lacZ*$_{+28}$	Cette étude (recombinaison)
JJ0113	JJ0062 Δ*omrAB::kan*	JJ0062 + P1(MG1188)
JJ0135	MG1508 P$_{tet}$-*fepA*$_{-26+45}$-*lacZ*$_{+28}$	Cette étude (recombinaison)
JJ0136	MG1508 P$_{tet}$-*fepA*$_{-92+45}$-*lacZ*$_{+28}$	Cette étude (recombinaison)
JJ0138	MG1508 P$_{tet}$-*fepA*$_{-173+45}$-*mutRX-lacZ*$_{+28}$	Cette étude (recombinaison)
JJ0139	NM300 Δ*sdsR::kan*	Cette étude (recombinaison)
JJ0141	MG1508 P$_{tet}$-*fepA*$_{-173+45}$-*mutSR-lacZ*$_{+28}$	Cette étude (recombinaison)
JJ0142	MG1772 Δ*omrAB::tet*	MG1772 + P1(MG1462)
JJ0143	MG1772 Δ*sdsR::kan*	MG1772 + P1(JJ0139)
JJ0144	JJ0142 Δ*sdsR::kan*	JJ0142 + P1(JJ0139)
JJ0146	MG1508 P$_{tet}$-*fepA*$_{-26+45}$-*mutSR-lacZ*$_{+28}$	Cette étude (recombinaison)
JJ0147	MG1508 P$_{tet}$-*fepA*$_{-173+45}$-*mutRX2-lacZ*$_{+28}$	Cette étude (recombinaison)
JJ0148	MG1508 P$_{tet}$-*fepA*$_{-173+45}$-*mutRX1,2-lacZ*$_{+28}$	Cette étude (recombinaison)
JJ0156	NM300 Δ*rseX::spc*	Cette étude (recombinaison)
JJ0159	MG2114 rrnBt2-PrseX$_{140+10}$(P1)-*lacZ*$_{-17}$	Cette étude (recombinaison)
JJ0161	MG2114 rrnBt2-PrseX$_{140+40}$(P1-P2)-*lacZ*$_{-17}$	Cette étude (recombinaison)
JJ0164	MG1772 Δ*rseX::spc*	MG1772 + P1(JJ0156)
JJ0167	JJ0144 Δ*rseX::spc*	JJ0144 + P1(JJ0156)
JJ0168	JJ0147 Δ*rseX::spc*	JJ0147 + P1(JJ0156)
JJ0169	MG1508 P$_{tet}$-*fepA*$_{-173+45}$-*mutSwitch2-lacZ*$_{+28}$	Cette étude (recombinaison)
JJ0171	MG1169 Δ*fepA::gnt*	MG1169 + P1(JJ0057)
JJ0172	MG2114 rrnBt2-PrseX$_{140+10}$(P1)-mutP1-*lacZ*$_{-17}$	Cette étude (recombinaison)
JJ0173	MG2114 rrnBt2-PrseX$_{140+40}$(P1-P2)-mutP1-*lacZ*$_{-17}$	Cette étude (recombinaison)
JJ0178	MG1772 Δ*flu::zeo*	MG1772 + P1(UGB2497)
JJ0179	JJ0142 Δ*flu::zeo*	JJ0142 + P1(UGB2497)
JJ0181	JJ0143 Δ*flu::zeo*	JJ0143 + P1(UGB2497)

JJ0182	JJ0164 Δ*flu::zeo*	JJ0164 + P1(UGB2497)
JJ0183	JJ0167 Δ*flu::zeo*	JJ0167 + P1(UGB2497)
JJ0184	MG1169 Δ*OmrAB::tet*	MG1169 + P1(MG1462)
JJ0185	MG1169 Δ*sdsR::kan*	MG1169 + P1(JJ0139)
Souche	**Description**	**Construction/Source**
JJ0186	NM300 Δ*sdsR::cat*	Cette étude (recombinaison)
JJ0187	JJ0184 Δ*sdsR::kan*	JJ0184 + P1(JJ0139)
JJ0188	JJ0187 Δ*rseX::spc*	JJ0187 + P1(JJ0156)
JJ0189	JJ0181 Δ*sdsR::cat (kanS)*	JJ0181 + P1(JJ0186)
JJ0191	JJ0183 Δ*sdsR::cat (kanS)*	JJ0183 + P1(JJ0186)
JJ0199	JJ0188 Δ*fepA::gnt*	JJ0188 + P1(JJ0057)
JJ0201	JJ0178 Δ*fepA::amp*	JJ0178 + P1(MG2097)
JJ0202	JJ0179 Δ*fepA::amp*	JJ0179 + P1(MG2097)
JJ0203	JJ0182 Δ*fepA::amp*	JJ0182 + P1(MG2097)
JJ0204	JJ0189 Δ*fepA::amp*	JJ0189 + P1(MG2097)
JJ0205	JJ0191 Δ*fepA::amp*	JJ0191 + P1(MG2097)
JJ0217	MG1508 P_{tet}-*fepA*$_{-26+45}$-*mutSwitch2-Loop-lacZ*$_{+28}$	Cette étude (recombinaison)
JJ0218	MG1508 P_{tet}-*fepA*$_{-173+45}$-*mutSwitch2-Loop-lacZ*$_{+28}$	Cette étude (recombinaison)
JJ0383	MG1508 P_{tet}-*bamA*$_{-107+57}$-*lacZ*$_{+28}$	Cette étude (recombinaison)
JJ0387	MG1508 P_{tet}-*bamA*$_{-107+21}$-*fepA*$_{+19+45}$-*lacZ*$_{+28}$	Cette étude (recombinaison)
JJ0388	MG1508 P_{tet}-*bamA*$_{-107+21}$-*fepA*$_{+19+45}$-*G3A5-lacZ*$_{+28}$	Cette étude (recombinaison)
JJ0389	MG1508 P_{tet}-*fepA*$_{-146+45}$-*lacZ*$_{+28}$	Cette étude (recombinaison)
plasmide	**Description**	**Construction**
pBRplac (vecteur)	pBR322 modifié portant un promoteur $P_{LlacO-1}$, *amp*^{R}, *tet*^{R}	(Guillier and Gottesman, 2006)
pBRplacMcaS (pMcaS)	*mcaS* sous $P_{LlacO-1}$ dans pBRplac, *amp*^{R}, *tet*^{R}	(Thomason et al., 2012)
pBRplacRseX (pRseX)	*rseX* sous $P_{LlacO-1}$ dans pBRplac, *amp*^{R}, *tet*^{R}	(Mandin and Gottesman, 2010)
pBRplacRseXmut1 (pRseXmut1)	*rseXmut1* sous $P_{LlacO-1}$ dans pBRplac, *amp*^{R}, *tet*^{R}	Cette étude
pBRplacRseXmut2 (pRseXmut2)	*rseXmut2* sous $P_{LlacO-1}$ dans pBRplac, *amp*^{R}, *tet*^{R}	Cette étude
pBRplacSdsR (pSdsR)	*sdsR* sous $P_{LlacO-1}$ dans pBRplac, *amp*^{R}, *tet*^{R}	(Mandin and Gottesman, 2010)
pBRplacSdsRmut (pRseXmut1)	*sdsRmut* sous $P_{LlacO-1}$ dans pBRplac, *amp*^{R}, *tet*^{R}	Cette étude

Tableau 2 – Souches et plasmides

*Les plasmides surproducteurs d'OmrA/B wt proviennent de (Guillier and Gottesman, 2006) et les plasmides surproducteurs d'OmrAmut et d'OmrBmut de cette étude (voir **Article**). Le restant des plasmides surproducteurs de sRNA Hfq-dépendants non indiqués ici proviennent de (Mandin and Gottesman, 2010). La construction des plasmides a suivi le même protocole que celui indiqué précédemment (voir **Article**)*

a. Construction des souches

Les souches ont été construites comme décrit dans l'article. Brièvement, la souche NM300 porte un mini-lambda permettant l'induction à 42°C des gènes de recombinaison (Yu et al., 2000). Les divers mutants par délétion ont été construits dans cette souche ou une souche portant le même mini-lambda, en remplaçant le gène désiré par une cassette de

résistance à un antibiotique. Les souches mutantes ont été sélectionnées sur milieu LB additionné d'antibiotiques en fonction de la résistance choisie. Pour la construction de fusions traductionnelles, la souche MG1508 a été utilisée : elle porte une construction Ptet-*cat-sacB-lacZ* au locus *lacZ* ainsi le mini-lambda évoqué ci-dessus. Les recombinants ont été sélectionnés sur du milieu LB sans NaCl additionné de saccharose à 6% final et de X-gal. Le gène *sacB* entraînant une létalité en présence de saccharose, seuls les clones où la cassette *cat-sacB* a été remplacée poussent sur ce milieu, permettant la sélection de recombinants. De même une coloration bleue des colonies permet dans la plupart des cas une sélection plus efficace des recombinants. Les candidats ont été testés pour leur sensibilité au chloramphénicol et vérifiés par séquençage. Suivant le même principe, les fusions transcriptionnelles ont été construites à partir de la souche MG2114 portant une construction rrnBt2-*cat-sacB-lacZ*, où le terminateur rrnBt2 empêche toute transcription résiduelle à partir de signaux en amont du promoteur étudié.

b. Milieux utilisés

La plupart des milieux utilisés sont indiqués dans la partie « matériels et méthodes » de l'article. Brièvement et sauf indication contraire, les souches ont été cultivées en milieu LB en présence des concentrations suivantes d'antibiotiques et produits selon les besoins de l'expérience : ampicilline 150µg/mL, chloramphénicol 25µg/mL, gentamycine 10µg/mL, kanamycine 25µg/mL, spectinomycine 50µg/mL, tétracycline 10µg/mL, zéocine 25µg/mL, 2,2-dipyridyl (dip) 250µM, IPTG 100µM, procaïne 10mM, X-gal 0,004%. Pour le crible génétique, les milieux suivants contenant de la kanamycine à 50µg/mL ont été utilisés :

- Difco Antibiotic Medium 2 – lactose 1% - tétrazolium 77µg/mL;
- LB – agar – X-gal 0,008%;
- MacConkey – lactose 1%.

Le milieu « NBCD » pour les tests sur FecA est composé de NB – agar, Na-Citrate 50mM, dip 500µM ; le milieu agar souple pour l'observation de phénotype de motilité est composé de Bacto-tryptone 10g/L, agar 3g/L, NaCl 5g/L.

B. Oligonucléotides, sondes et anticorps

Sont répertoriés dans le tableau 3 les oligonucléotides ADN et sondes ayant servi pour les expériences d'allongements d'amorce, toeprints et NB.

Nom	Séquence	Spécificité
Allongements d'amorce		
fepAToeCy5	CGGTAACGACAATAGTATCG	5'-cyanine5 (Fig. 62)
fepAToeCy5-2L	CGTCATGTGAAACAGGAGTATCGGTCGGCTCT	5'-cyanine5 (Fig. 58)
Oligonucléotides Antisens		
AsFepA40.54	CCCATAAATCCCCAG	
AsFepA42.56	ACCCCATAAATCCCC	
DoSdsR	CCTGCATTAATGCCA	
DoSdsRmut	CCTCGTTTAATGCCA	
Sondes pour les Northern Blot		
fepA probe	CGGCTTGCCGTCAATCAAAATCAGCGTGTTTTCCGGACC	5'-biotine
ompA probe	CCATTGTTGTTGATGAAACCAGTGTCATGGTACTGGGACCAGC	5'-biotine
OmrA probe	CAGGTTGGTGCAAGAGACAGGGTACGAAGAGCGTACCG	5'-biotine
OmrB probe	CGCAGGCTGGTGTAATTCATGTGCTCAACCCGAAGTTGA	5'-biotine
RseX probe	GCCGGCATCATGCCGGCCCAAAAAATATAAACAAGACGAT	5'-biotine
RyhB probe	AAGTAATACTGGAAGCAATGTGAGCAATGTCGTGCTTTCAGGTTCTC	5'-biotine
SdsR probe	GAACACGATTCCTGTATTCGGTCCAGGGAAATGGCTCTTGGG	5'-biotine
SsrA probe	CGCCACTAACAAACTAGCCTGATTAAGTTTTAACGCTTCA	5'-biotine
Anticorps		
Anti-FepA	Origine : K. Postle	
Anti-L20	(produit au laboratoire)	

Tableau 3 – Sondes, oligonucléotides et anticorps

C. Protocoles expérimentaux

La plupart des expériences réalisées suivent très exactement le protocole détaillé dans l'article. Dans le cas où des conditions différentes ont été observées, elles sont indiquées dans les légendes des figures correspondant aux expériences concernées.

a. Allongement d'amorce

L'allongement d'amorce (Fig. 58) a été réalisé comme suit : après une culture de bactéries en milieu LB à 37°C jusqu'à une OD_{600} de 0,4, de 2 ou 3h au-delà de DO_{600}=2, les ARN ont été extraits (voir **Article**). 5µg d'ARN (OD_{600}=0,4 et OD_{600}=2), 3,5µg d'ARN (OD_{600}=2 + 3h) ou bien 0,05pmol de transcrit *fepA*$_{-173+129}$ ont été dénaturés à 80°C en présence de 2pmol d'amorce ADN fepAtoeCy5-2L pendant 3min puis plongés dans un bain d'éthanol-carboglace

pendant une minute afin d'empêcher une renaturation progressive des ARN favorisant les interactions intramoléculaires. Chaque réaction a ensuite été incubée à 37°C pendant 5min pour permettre l'appariement entre l'amorce ADN et l'ARNm *fepA*. Enfin, les réactions d'allongement d'amorce ont été réalisées à 50°C pendant 1h en présence de dNTP à 4mM, de DTT à 5mM, de tampon « First Strand » (Invitrogen) à 1x, de 20U de RNaseOut (Invitrogen) et de 100U de transcriptase inverse Superscript III (Invitrogen). Les échantillons ont ensuite été chauffés à 75°C pour inactiver la transcriptase inverse, puis dénaturé avec un tampon contenant du formaldéhyde et analysés sur gel de séquençage PAGE.

b. RNAseq (Annexe 2)

Les bactéries ont été cultivées à 37°C en LB + ampicilline jusqu'à une OD_{600} de 0,3 puis traitées avec 100µM d'IPTG pendant 10min pour induire l'expression des sRNA. Au terme de cette induction, les ARN ont été extraits, traités à la désoxyribonucléase (DNase), quantifiés, analysés par NB pour vérifier leur qualité puis envoyés à la plateforme de RNAseq « Transcriptome and Epigenome » de l'Institut Pasteur où ce RNAseq a été réalisé en collaboration. À l'Institut Pasteur, les ARN ont été traités avec un kit permettant d'éliminer plus de 99% des ARNr avant d'être fragmentés. Puis ces ARN ont subi une réaction de RT afin de produire des ADN complémentaires. Ces derniers ont été séquencés, puis l'alignement des « reads » au génome d'*E. coli* MG1655 a été réalisé par la plateforme de même qu'une analyse statistique.

Annexe 1 : **Revue n°2** « Small noncoding RNAs and stress responses »

CHAPTER 5.1

Bacterial small RNAs in mixed regulatory circuits

Jonathan Jagodnik[1], Denis Thieffry[2], and Maude Guillier[1]

[1]CNRS UMR8261 (affiliated with University Paris Diderot, Sorbonne Paris Cité), Institut de Biologie Physico-Chimique, Paris, France
[2]Institut de Biologie de l'ENS (IBENS), Ecole Normale Supérieure, UMR CNRS 8197 and INSERM 1024, Paris, France

5.1.1 Introduction

Bacteria are extremely versatile organisms that can adapt to and colonize various and changing environments. A key to this remarkable faculty resides in their ability to regulate gene expression in response to input signals, through various and very sophisticated mechanisms.

The concept of gene expression regulation was introduced around 1960 by Jacob, Monod, and coworkers who, in the operon model, proposed the existence of regulatory genes (Jacob and Monod, 1961). More than 50 years later, most of this model still holds true, even though decades of research have without surprise added further complexity with the identification of many regulators acting at different stages of gene expression.

It is, for instance, now firmly established that transcription can be regulated not only by repressors but also by activators. Furthermore, as initially discussed by Jacob and Monod, regulators can be either proteins or RNA molecules, and these regulators can target either the DNA or the messenger RNA (mRNA).

A few genes have been known for a while to be regulated by proteins at the posttranscriptional level, but this field has undergone a revolution with the identification and characterization of regulatory RNAs, many of which act at the posttranscriptional level. Interestingly, it now appears that many functions are controlled by complex regulatory circuits that involve both transcriptional control, which is mediated by proteins, and posttranscriptional regulatory steps, which rely on regulatory RNAs for the most part. In this chapter, we will briefly summarize the different regulatory RNAs identified in bacteria before describing a few examples of mixed regulatory circuits and discussing their possible advantages.

5.1.2 Bacterial regulatory RNAs

In addition to its role in gene expression (e.g., via mRNA, transfer RNA (tRNA), and ribosomal RNA (rRNA)) or in catalysis (Cech and Bass, 1986), RNA is now recognized as playing an important regulatory role in cellular physiology in all kingdoms of life. In bacteria, many regulatory RNAs are rather short, and they are therefore commonly referred to as small RNAs (sRNAs). Bacterial regulatory RNAs can be classified in several categories, based on their genomic location and mode of action (see Chapter 2.8):

Riboswitches or "sensor" mRNAs: These regulatory RNA elements act *in cis* by regulating expression of genes encoded by the same RNA molecule. Riboswitches bind a ligand (metabolite, ion, etc.) with high affinity through their aptamer domain, which induces a structural change of an expression platform and leads to transcriptional or posttranscriptional regulation. More generally, sensors denote mRNAs that respond to environmental cues such as pH or temperature via a similar structural rearrangement.

Regulatory RNAs titrating a protein: These RNAs regulate gene expression by interacting with a protein, which leads to changes in its activity. Among the best studied examples of this class are 6S or CsrB–CsrC sRNAs that, respectively, interact with the RNA polymerase (RNAP) holoenzyme or the CsrA global posttranscriptional regulator.

Regulatory RNAs acting via imperfect base pairing: sRNAs of this class base pair with target mRNA(s) and regulate, either positively or negatively, their translation and/or stability. Because they can pair to mRNAs through imperfect duplexes, these base-pairing sRNAs often have multiple targets. Furthermore, these duplexes are in general rather short and, in enterobacteria such as *Escherichia coli* or *Salmonella*, for instance, action of these sRNAs requires an RNA chaperone called Hfq that both stabilizes them and facilitates their pairing to targets (Vogel and Luisi, 2011).

Antisense RNAs (asRNAs): These RNAs are transcribed from the strand opposite of the regulated mRNA. They form a rather heterologous class, in terms of both size and mode of action (Thomason and Storz, 2010). As the previous class, asRNAs can pair to target mRNAs and modulate their translation and/or

Stress and Environmental Regulation of Gene Expression and Adaptation in Bacteria, First Edition. Edited by Frans J. de Bruijn.
© 2016 John Wiley & Sons, Inc. Published 2016 by John Wiley & Sons, Inc.

stability. One major difference, however, is that asRNAs are perfectly complementary to their target and, as a result, are thought to have a unique target. Sense–antisense pairing has also been reported to lead to premature transcription termination (Stork *et al.*, 2007). At last, an asRNA can regulate expression of the sense gene without necessarily base pairing to the mRNA, but via transcription interference, where transcription from one promoter is influenced by transcription of a second promoter (Shearwin *et al.*, 2005).

CRISPR (clustered regularly interspaced short palindromic repeat) RNAs: These are generated from CRISPR–Cas loci and are yet a distinct class of bacterial sRNAs (see Chapter 2.11). crRNAs (CRISPR RNAs processed from a long precursor CRISPR transcript containing multiple spacer repeat modules) provide adaptive immunity against phages and plasmids by guiding Cas proteins to the invading DNA through sequence complementarity (Sorek *et al.*, 2013). However, in addition to this well-documented function, CRISPR–Cas loci have recently been found to regulate endogenous gene expression as well. In the pathogen *Francisella novicida*, two partially complementary small RNAs made from a CRISPR–Cas locus (different from the crRNA mentioned here) associate with the Cas9 DNA endonuclease protein and allow degradation of the mRNA encoding a bacterial lipoprotein, through a not completely elucidated mechanism. Because lipoproteins are involved in bacterial recognition by the host immune system, Cas9 or either of these sRNAs facilitates evasion of the innate immune pathway, and these three components are required for virulence (Sampson *et al.*, 2013a, 2013b). Even though it remains to be addressed whether endogenous control of gene expression by CRISPR–Cas loci is a common process, RNA-guided Cas9 enzymes have been widely used as a tool for genome editing (Charpentier and Doudna, 2013) and to specifically control gene expression using engineered versions of Cas9 (Bikard and Marraffini, 2013).

This classification is prone to evolve with future findings, and, importantly, a single regulatory RNA can belong to multiple classes. For instance, the McaS sRNA acts via imperfect pairing on some targets while also controlling expression of other genes by sequestering the CsrA protein (Jorgensen *et al.*, 2013). Riboswitches that induce transcription termination can produce a short transcript that was reported, in at least one case, to act *in trans* by imperfect base pairing on a virulence factor gene encoded from a distinct locus (Loh *et al.*, 2009).

As many circuits involving regulatory RNAs are starting to be elucidated, it appears that transcription and translation controls are intimately connected. Indeed, many transcriptional regulators include one or several regulatory RNAs in their regulon, while many regulatory RNAs have been found to control transcriptional regulators (Gopel and Gorke, 2012; Mandin and Guillier, 2013). Several examples are discussed in the remainder of this chapter; for space constraints and because it is one of the best characterized sRNA classes, we will focus mostly on the Hfq-dependent sRNAs in enterobacteria.

5.1.3 Control of sRNAs by transcriptional regulators

In order to perform their regulatory role, regulatory RNAs must somehow respond to an input signal. In most cases, this is achieved through a change in their expression at the transcriptional level. Many of the well-established transcriptional regulators (e.g., CRP, Lrp, OxyR (see Chapter 10.5), FNR, and Fur (see Chapter 19.1), and two-component systems (TCSs; see Section 3) such as ArcB–ArcA, EnvZ–OmpR, and PhoQ–PhoP) actually include one or several sRNAs in their regulon (Table 5.1.1).

For instance, under iron-replete conditions, the Fur transcription factor (TF) is a repressor of genes encoding proteins involved in iron acquisition, but it also represses transcription of the RyhB sRNA (Massé and Gottesman, 2002; see Chapter 19.1). Interestingly, RyhB, whose synthesis is therefore induced under iron starvation, posttranscriptionally represses several genes for iron storage and iron-containing proteins and activates genes for iron scavenging (Massé *et al.*, 2005; Prevost *et al.*, 2007). This results in an increase of iron available for more important genes when iron is limiting, such as ribonucleotide reductase and dihydroxyacid dehydratase, which are involved in production of nucleic acids and amino acids, respectively (Massé *et al.*, 2007). Accordingly, RyhB is required for optimal growth and long-term viability under iron starvation (Jacques *et al.*, 2006). This is a nice example of a regulatory circuit where a sRNA contributes to the function regulated by the TF.

Not surprisingly, transcription of regulatory RNAs is also controlled by various sigma factors. Sigma factors bind to the RNAP core enzyme and direct it to a set of specific promoters. Bacteria possess a gene for a housekeeping sigma factor (e.g., RpoD in *E. coli*), often along with genes for several alternative sigma factors that are activated in response to specific conditions (Österberg *et al.*, 2011; see Chapter 2.4). The alternative sigma factor RpoS directs transcription of hundreds of genes in the stationary phase and in response to various environmental stresses, including genes for at least three regulatory RNAs: GadY, SdsR, and SraL. GadY is an asRNA that induces processing of the bicistronic *gadXW* mRNA and thereby stabilizes the *gadX* and *gadW* mRNAs, which both encode transcriptional regulators of acid resistance genes (Opdyke *et al.*, 2004). Therefore, GadY likely participates in the RpoS-mediated acid resistance. SdsR represses synthesis of OmpD, *Salmonella*'s most abundant porin (Frohlich *et al.*, 2012); and SraL has been shown to decrease the level of the *tig* mRNA for the trigger factor, a ribosome-associated chaperone important for protein folding (Silva *et al.*, 2013). In this case, the physiological connection with the RpoS dependency of SdsR and SraL promoters is less obvious. Perhaps it is beneficial to decrease OmpD levels under stress conditions, as even a slight OmpD upregulation triggers cell lysis (Frohlich *et al.*, 2012), and to slow down protein folding under nonoptimal conditions? It is, however, likely that additional targets of SdsR and SraL will be identified in the future, and these might help in establishing such a connection.

Table 5.1.1 Known connections between Hfq-binding sRNAs and TF or sigma factors in enterobacteria (adapted from Mandin and Guillier, 2013). Unless otherwise indicated, only direct or presumably direct connections are shown.

Hfq-binding sRNA	Regulatory TF or sigma factor	Regulated TF or sigma factor	References
ArcZ	ArcB–ArcA	RpoS, ArcB	(Mandin and Gottesman, 2010; Papenfort *et al.*, 2009)
CyaR	CRP		(De Lay and Gottesman, 2009; Johansen *et al.*, 2008; Papenfort *et al.*, 2008)
ChiX		DpiB–DpiA (in *E. coli*)	(Mandin and Gottesman, 2009)
DsrA		HNS, RpoS	(Lease *et al.*, 1998; Majdalani *et al.*, 1998)
DapZ	HilD (in *Salmonella*)		(Chao *et al.*, 2012)
FnrS	Fnr		(Durand and Storz, 2010)
GadY (absent in *Salmonella*)	RpoS	GadX, GadW	(Opdyke *et al.*, 2004, 2011)
GlmY	GlrK–GlrR with RpoN		(Gopel *et al.*, 2011; Reichenbach *et al.*, 2009)
GlmZ	GlrK–GlrR with RpoN (in *Salmonella* but not *E. coli*)		(Gopel *et al.*, 2011)
GcvB	GcvA, Lrp (direct?)	PhoP, CsgD (direct?), Lrp (direct?)	(Coornaert *et al.*, 2013; Jorgensen *et al.*, 2012; Modi *et al.*, 2011; Urbanowski *et al.*, 2000)
InvR (absent in *E. coli*)	HilD		(Pfeiffer *et al.*, 2007)
McaS (absent in *Salmonella*)	CRP	CsgD, FlhDC	(De Lay and Gottesman, 2012; Jorgensen *et al.*, 2012; Thomason *et al.*, 2012)
MgrR	PhoQ–PhoP		(Moon and Gottesman, 2009)
MicA	RpoE	PhoP	(Coornaert *et al.*, 2010; Figueroa-Bossi *et al.*, 2006; Gogol *et al.*, 2011; Johansen *et al.*, 2006; Papenfort *et al.*, 2006; Udekwu and Wagner, 2007)
MicF	Lrp, EnvZ–OmpR, Rob, MarA, SoxS	Lrp	(Corcoran *et al.*, 2012; Delihas and Forst, 2001; Holmqvist *et al.*, 2012)
MicL	RpoE		(Guo *et al.*, 2014)
OmrA/B	EnvZ–OmpR	CsgD, FlhDC, OmpR	(De Lay and Gottesman, 2012; Guillier and Gottesman, 2006, 2008; Holmqvist *et al.*, 2010)
OxyS	OxyR	FlhDC, FhlA	(Altuvia *et al.*, 1997, 1998; De Lay and Gottesman, 2012)
RprA	Rcs	RpoS, CsgD	(Majdalani *et al.*, 2001, 2002; Mika *et al.*, 2012)
RybB	RpoE		(Johansen *et al.*, 2006; Papenfort *et al.*, 2006; Thompson *et al.*, 2007)
RydC		CsgD	(Bordeau and Felden, 2014)
RyhB	Fur	Fur (direct and indirect)	(Massé and Gottesman, 2002; Massé *et al.*, 2005; Vecerek *et al.*, 2007)
SdsR	RpoS		(Frohlich *et al.*, 2012)
SgrS	SgrR		(Vanderpool and Gottesman, 2007)
Spot42	CRP		(Polayes *et al.*, 1988)

RpoE is another alternative sigma factor, activated upon membrane stress in response to accumulation of unassembled outer membrane proteins (OMPs) or lipopolysaccharides (LPS) in the periplasm (Lima *et al.*, 2013). In addition to many protein-coding genes, RpoE directly controls transcription of at least three sRNAs, namely MicA, RybB, and MicL. Interestingly, MicA and RybB repress expression of multiple genes, including genes for OMPs (Gogol *et al.*, 2011, and references therein). In contrast, MicL targets only the mRNA of the Lpp lipoprotein, which is one of the most abundant proteins in *E. coli* and is involved in the stabilization and integrity of the outer membrane (OM) (Guo *et al.*, 2014). By repressing synthesis of abundant proteins that have to be assembled in the OM, these three sRNAs contribute to decreasing membrane stress and lowering RpoE activation. They play a crucial role in the RpoE response, and they can rescue growth defects arising from an improper function of the RpoE pathway (Gogol *et al.*, 2011; Guo *et al.*, 2014).

5.1.4 Control of transcriptional regulators by sRNAs

Conversely, multiple examples of transcriptional regulators whose synthesis and/or activity is controlled by sRNAs have been reported. We will give a few examples of such connections for sigma factors first, then for TFs, and finally for TFs that are part of TCSs.

5.1.4.1 Sigma factors

Control of sigma factor synthesis or activity by regulatory RNAs was recognized early on and allowed pioneering discoveries in the mode of action of different regulatory RNAs.

The 6S RNA was one of the first sRNAs to be identified, because of its high abundance. It accumulates in the late stationary phase, where its concentration can reach up to 10,000 copies/cell. Although bacterial 6S are not conserved in sequence,

their structure is remarkably well conserved and resembles an open promoter. In line with this, 6S RNA interacts with the housekeeping RNAP holoenzyme (i.e., complexed with the RpoD subunit), which blocks DNA from binding. 6S therefore inhibits transcription from RpoD-dependent promoters, but surprisingly with promoter-specific effects that do not necessarily correlate with the affinities or kinetics of the various promoters for RNAP. This suggests that 6S not only acts by competition with DNA for binding RNAP, but also might have direct effects on some promoters. Furthermore, 6S activates transcription of several RpoS-dependent genes, but again with promoter-specific effects and through a not completely clear mechanism (Cavanagh and Wassarman, 2014). The other interesting point regarding 6S biology is that it not only resembles an open promoter but also can be used as a template for transcription. Indeed, upon outgrowth (shift of cells in the stationary phase into fresh medium), short RNAs, called pRNAs, are transcribed by RNAP using 6S as a template (Wassarman and Saecker, 2006). This allows release of both RNAP and 6S RNA; the latter is in a complex with pRNA, which prevents its reassociation with RNAP and makes it more prone to degradation (Beckmann *et al.*, 2012; Cavanagh *et al.*, 2012). 6S has a major impact on bacterial physiology as it leads to changes in expression of hundreds of genes. Moreover, in competition experiments, cells lacking 6S were at a disadvantage for growth and long-term survival (Cavanagh and Wassarman, 2014).

Another pioneering example of the role of sRNAs in cellular physiology comes from the recognition that multiple sRNAs influence the synthesis of the RpoS master regulator of the stationary phase. The amount of RpoS is controlled at many levels, including translation: three sRNAs, namely DsrA, RprA, and ArcZ, directly pair to *rpoS* mRNA and activate translation through opening of an inhibitory secondary structure. In addition, a fourth sRNA, OxyS, represses RpoS synthesis in an Hfq-dependent manner, presumably by inhibiting the action of the activator sRNAs by limiting Hfq availability (reviewed in Battesti *et al.*, 2011). Interestingly, all these sRNAs respond to different environmental cues: DsrA is made at low temperatures, RprA is activated by the Rcs phosphorelay involved in capsule synthesis, ArcZ is repressed by the ArcB–ArcA TCS activated under anaerobic conditions or inhibited by oxidized quinones, and OxyS responds to oxidative stress as part of the OxyR regulon (see Section 10). Therefore, these sRNAs might allow the integration of multiple signals to control a major regulator of stress response. Note that all these sRNAs target other genes as well, among them several genes for other transcriptional regulators, such as *hns* (Lease *et al.*, 1998; Majdalani *et al.*, 1998), *flhlA* (Altuvia *et al.*, 1998), and *csgD* and *flhDC* (Table 5.5.1, and discussed further in this chapter).

5.1.4.2 Transcription factors

Several regulatory RNAs have also been found to control synthesis of TFs. One of the most striking examples in this regard is the control of the switch between a sessile (community, biofilm-embedded; see Section 22) and a planktonik (individual, motile) lifestyle. This is a highly regulated and extraordinarily complex process, in which the FlhDC and CsgD regulators play a key role (Mika and Hengge, 2013). To summarize, FlhDC lies at the top of the transcriptional cascade controlling assembly of flagella, which are thin and long extracellular structures that are composed of more than 1000 protein subunits and allow bacterial motility. In contrast, CsgD promotes surface adhesion and biofilm formation by activating transcription of genes involved in synthesis of components of the biofilm extracellular matrix, such as curli and cellulose. CsgD also represses several flagellar genes (Dudin *et al.*, 2014; Ogasawara *et al.*, 2011). Not surprisingly, expression of both *csgD* and *flhDC* genes is extensively regulated. For instance, *csgD* transcription is directly controlled by RpoS and several TFs, among them MlrA, OmpR, CpxR, IHF, and RstA (Ogasawara *et al.*, 2010). Transcription of *flhDC* is also subject to extensive control (e.g., via OmpR, Rcs, H-NS, or CRP) (reviewed in Soutourina and Bertin, 2003). The second messenger c-di-GMP also plays an important role in motility or biofilm decision by activating *csgD* transcription, among other functions (Povolotsky and Hengge, 2012).

In addition to these controls, multiple sRNAs were recently identified as direct posttranscriptional regulators of *csgD* and/or *flhDC* expression by targeting their 5′-UTR and/or translation initiation region (Figure 5.1.1 and the Book Cover photo).

The CRP-activated McaS sRNA has, for instance, an opposite effect on *csgD* (repressed) and *flhDC* (activated) (De Lay and Gottesman, 2012; Jorgensen *et al.*, 2012; Thomason *et al.*, 2012). OmrA and OmrB also pair to both *csgD* and *flhDC* mRNAs, but in this case, this pairing leads to a downregulation of both TFs (De Lay and Gottesman, 2012; Holmqvist *et al.*, 2010). OmrA and OmrB are both regulated by the EnvZ–OmpR TCS, which itself directly controls *csgD* and *flhDC* transcription as well (Figure 5.1.1). Similarly, CRP also activates *flhDC* independently of McaS. The resulting regulatory motifs are called feedforward motifs (FFMs), where a regulator A controls a regulator B, and both A and B control a gene C (Figure 5.1.2). Interestingly, the two FFMs converging on *flhDC* are coherent, but it is an incoherent FFM for *csgD*. Other direct sRNA repressors of *flhDC* are OxyS and ArcZ, while RydC and GcvB repress *csgD* expression (Figure 5.1.1). This is likely due to a direct interaction as RydC and *csgD* interact *in vitro* (Bordeau and Felden, 2014), and a potential pairing interaction can be predicted between *csgD* and GcvB (Jorgensen *et al.*, 2012). Last but not least, RprA sRNA is another direct repressor of *csgD* expression, which interestingly also controls *rpoS* and *ydaM*, the gene for a diguanylate cyclase (i.e., an enzyme responsible for c-di-GMP synthesis) that indirectly activates *csgD* transcription (Mika *et al.*, 2012). RprA thus regulates *csgD* through a complex regulatory cascade that comprises both an incoherent FFM (via RpoS upregulation) and a coherent FFM (via YdaM regulation) (Figure 5.1.1). Furthermore, RprA synthesis depends on the Rcs phosphorelay, which itself plays an important role in regulating the switch from a motile to a sessile lifestyle, by

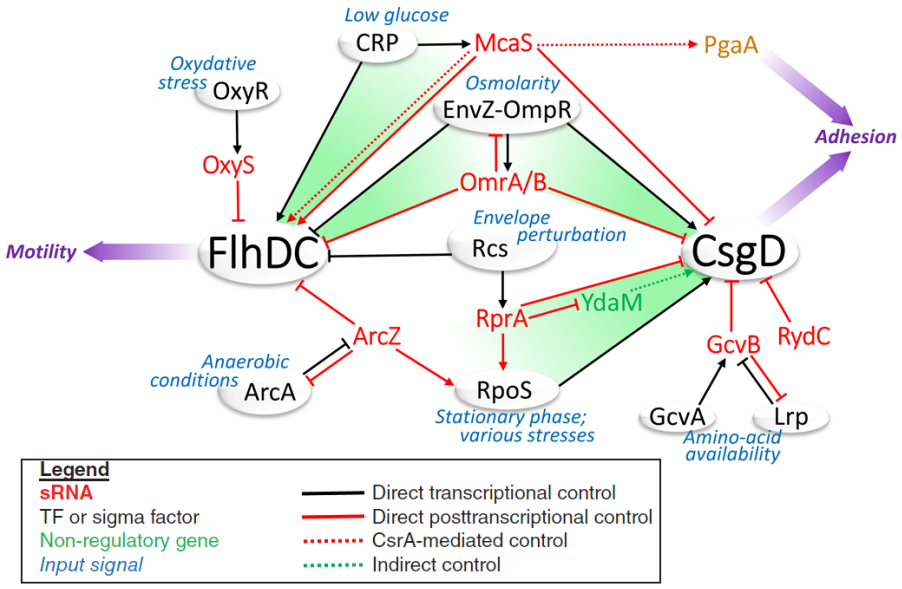


Figure 5.1.1 Involvement of Hfq-binding sRNAs in control of motility and adhesion in enterobacteria. Positive and negative controls are shown with normal and blunt-end arrows, respectively. Asterisks indicate regulatory interactions presumed to be direct. Green triangles highlight feedforward regulatory motifs.

	Regulatory motif	Examples involving sRNA and TF	Characteristics
Feedback circuits (FCs)	**Negative feedback circuits** A ⊣ B	OmpR ⊣ OmrA/B LuxO → Qrr ⊣ LuxR–HapR; Qrr ⊣ AphA	**Homeostasis and oscillations** In conjunction with negatively autoregulated TFs, Qrrs RNAs minimize fluctuations and coordinate cellular response among the population.
	Positive feedback circuits A ⊣ B *Double negative* or A → B *Double positive*	Lrp ⊣ MicF Fur ⊣ RyhB	**Bistability and differentiation**
Feedforward motifs (FFMs)	**Coherent FFM** A → B → C; A → C *(e.g., type 4)*	CRP → McaS → FlhDC; CRP → FlhDC *Type 1* OmpR ⊣ MicF → OmpF; OmpR ⊣ OmpF *Type 3* CRP → Spot42 ⊣ GltA, SrlA, NanT, FucI, FucK, SdhC, GalK, XylF….; CRP → GltA, SrlA, NanT, FucI, FucK, SdhC, GalK, XylF…. *Type 4*	**Delay in the (on or off) response, and filter transient (on or off) signals** Potentially faster response following deactivation when an FFM includes an sRNA (e.g., Spot42)
	Incoherent FFM A → B → C; A ⊣ C *(e.g., type 1)*	PhoP → AmgR ⊣ MgtC; PhoP → MgtC *Type 1* RprA → RpoS → CsgD; RprA ⊣ CsgD *Type 3*	**Speed up responses; pulses in gene expression**

Figure 5.1.2 sRNAs and TFs involved in feedback circuits and feedforward network motifs. TFs, sRNAs, and non-regulatory genes are in black, red, and green, respectively. Solid lines indicate transcriptional (black) or posttranscriptional (red) regulatory interactions, regardless of whether these interactions are direct or not. Characteristics in red are potentially specific properties of sRNAs. See text for details.

repressing *flhDC* transcription and activating synthesis of colanic acid, a major component of biofilm matrix. As for RprA, the role of McaS in biofilm formation is not restricted to regulating *csgD* and *flhDC*, since McaS was found to increase biofilm formation in a *pgaA*-dependent manner. *pgaA* is required for the export of a polysaccharide adhesin called PGA (poly-β-1,6 N-acetyl-D-glucosamine), another constituent of the biofilm matrix. Interestingly, while McaS regulates *csgD* and *flhDC* through base pairing, it activates *pgaA* indirectly through CsrA. Indeed, McaS is not only an Hfq-binding sRNA but also a CsrA-binding sRNA, and it therefore also affects biofilm formation by indirectly deregulating CsrA targets, among them *pgaA*, as well as *flhDC* and two genes for diguanylate cyclases (Figure 5.1.1) (Jorgensen *et al.*, 2013).

Despite this already extremely intricate network, sRNAs involved in biofilm formation or motility are not restricted to those mentioned so far. It is, for instance, likely that other known regulators of CsrA, such as CsrB or CsrC, could affect biofilm formation, even though it might not be true in all conditions (Jorgensen *et al.*, 2013). Furthermore, GadY and MicA sRNAs were found to affect motility while ArcZ and SdsR upregulated biofilm formation (De Lay and Gottesman, 2012; Monteiro *et al.*, 2012), via still-elusive mechanisms. It is also possible that, under specific conditions, regulators of RpoS, H-NS, or other TFs involved in the biofilm or motility pathway could have effects as well.

These sRNAs can certainly allow the integration of environmental cues, in addition to the signals that are transduced by the multiple TF regulators of *csgD* and *flhDC*, to precisely control expression of these two key regulators. They also likely contribute to restrict induction of *csgD* or *flhDC* to very specific conditions, which might be beneficial by preventing the synthesis of curli or flagella when they are not required. These multiple sRNAs could also provide unique dynamic properties, in particular fast switch-like responses under both induction and de-induction conditions, and participate in the synchrony among the population (see Section 5.1.5).

Importantly, direct control by sRNAs also applies to master regulators such as H-NS or Lrp, whose expression impacts hundreds of genes (Table 5.1.1; see also http://regulondb.ccg.unam.mx).

5.1.4.3 Two-component systems

Several bacterial TFs are part of the so-called TCSs, in which their activity is regulated by phosphorylation. TCSs are widely used in bacteria as a way to sense and respond to the environment. They usually consist of a sensor histidine kinase (HK), located at the membrane, which autophosphorylates in response to a specific stimulus and can then transfer the phosphoryl group on a conserved aspartate residue of a cognate response regulator (RR). The RR is often a DNA-binding protein, whose phosphorylation leads to changes in transcription of target genes (see Chapter 3.1). There can be variations from this simplest scheme, for instance in phosphorelay systems (such as the Rcs pathway mentioned here) where additional phosphotransfer steps and additional components are involved (Laub and Goulian, 2007). In many cases, the genes for the cognate HK and RR are organized in an operon, allowing for a common regulation. As for other TFs, regulation of several TCSs has been shown to be mediated by regulatory RNAs through various mechanisms.

Several sRNAs directly base-pair to mRNAs encoding TCSs and repress their expression. This is the case for two of the most thoroughly studied TCSs in enterobacteria, EnvZ–OmpR and PhoQ–PhoP, encoded respectively by the *ompR–envZ* and *phoPQ* operons. EnvZ–OmpR is involved in adaptation to changes in osmolarity (Pratt *et al.*, 1996), and PhoQ–PhoP in magnesium homeostasis, LPS modifications, and bacterial virulence (Groisman *et al.*, 2013). OmrA and OmrB, two seemingly redundant OmpR-activated sRNAs, repress OmpR synthesis (and presumably that of EnvZ as well), thereby creating a regulatory feedback circuit (Guillier and Gottesman, 2008). As for *phoPQ* mRNA, it is targeted by two unrelated sRNAs, MicA and GcvB, made in response to envelope stress and availability of amino acids, respectively (Coornaert *et al.*, 2010, 2013). Surprisingly, however, regulation of these TCSs by sRNAs can have drastically different effects on the regulons of the TCSs. More precisely, OmrA/B does not change the level of OmpR-P, the phosphorylated form of OmpR, in agreement with the previously reported robustness of OmpR-P to changes in *ompR* or *envZ* expression (Batchelor and Goulian, 2003). Consequently, OmrA/B does not affect the transcription of robust OmpR targets such as *ompC* or *ompF*; nevertheless, OmrA/B can limit their own synthesis by repressing *ompR* (A. Brosse *et al.*, unpublished data). In the case of PhoQ–PhoP, while both MicA and GcvB repress *phoP* translation and decrease *phoPQ* mRNA levels, only MicA has been found to indirectly repress PhoP targets (Coornaert *et al.*, 2013). While the precise molecular mechanism underlying these observations is still unclear, these findings highlight the need to systematically study the effect of a regulatory RNA controlling a given TF on this TF's regulon.

Control of TCSs by RNAs is, however, not restricted to base pairing sRNAs. Recently, the study of the regulation of the ethanolamine (EA) utilization (*eut*) genes in two Gram-positive bacteria, *Listeria monocytogenes* and *Enterococcus faecalis*, unraveled an exquisite role for a riboswitch-controlled sRNA in binding an active response regulator. EA derives from mammalian and bacterial cell membranes and is prevalent in the gastrointestinal environment. Bacteria that possess the genes for EA catabolism can use it as a source of nitrogen and carbon, and EA utilization has been linked to bacterial virulence in several studies. In diverse bacteria, expression of the *eut* genes is induced in the presence of both EA and adenosylcobalamine (AdoCbl, one of the active forms of vitamin B_{12}), a cofactor required for EA utilization. In firmicutes such as *L. monocytogenes* or *E. faecalis*, a key component of this control is the EutWV TCS, which is activated in the presence of EA. EutV is an unusual RR that, in the phosphorylated form, binds to RNA (and not DNA) elements consisting of a dual hairpin structure (also called ANTAR

elements) and promotes antitermination. Several ANTAR elements are found within the *eut* locus, including one downstream of an AdoCbl-responsive riboswitch, present toward the beginning of the locus (that comprises more than 15 genes in the two species). This riboswitch allows the production of a short transcript upon binding of AdoCbl, but in its absence, a longer transcript, that encompasses an ANTAR element, is produced. This long transcript can then sequester EutV-P, which prevents antitermination at the distal ANTAR sites. This ensures that, unless both EA and AdoCbl are present, expression of the *eut* genes is not induced (DebRoy *et al.*, 2014; Mellin *et al.*, 2014).

5.1.5 Roles and properties of mixed regulatory circuits

The recognition that many regulatory circuits in bacteria are mixed – that is, composed of both protein regulators (that act at the transcriptional level) and regulatory RNAs (that, in many cases, act at the posttranscriptional level) – raises the question of the possible advantages and specific properties of such circuits.

In general, having a regulator B in the regulon of a regulator A is expected to (i) expand the regulon of A, (ii) allow integration of multiple input signals to precisely control gene expression, (iii) allow the possibility to rewire an activator into a repressor and conversely (e.g., an activator A could indirectly repress gene expression through a repressor B), and (iv) connect different regulatory pathways. Furthermore, if A or B is a regulatory RNA, this might have other implications.

5.1.5.1 Regulating gene expression with sRNAs

An advantage of using regulatory RNAs that is often proposed is a cheap and fast synthesis in comparison with protein regulators. Maybe more importantly, they act mostly at the posttranscriptional level (i.e., at a step that is closer to protein production than transcription). This allows for a faster response than transcriptional control (Shimoni *et al.*, 2007), and, in the case of negative regulation, it can prevent expression of abundant and/or long-lived target mRNAs that, if regulation was exerted at the level of transcription, could be used in several rounds of translation (see Chapter 2.8). In this regard, it is striking that regulatory RNAs directly repress many mRNAs encoding OMPs (e.g., *ompA*, *ompC*, *ompF*, *lpp*, etc.) that have been reported to be both abundant and stable mRNAs (Bernstein *et al.*, 2002).

Possibly also contributing to the widespread occurrence of sRNAs in regulatory circuits is their ability to evolve quickly (Gottesman and Storz, 2011). For instance, many Hfq-binding sRNAs present in *E. coli* accumulate inside the *Enterobacteriales* order only (Peer and Margalit, 2014). Furthermore, these sRNAs can also rapidly acquire novel targets given the flexibility of sRNA–mRNA pairing, as exemplified by the fact that RybB controls the *ompN* gene in *Salmonella* but not in the closely related bacterium *E. coli* (Gogol *et al.*, 2011).

Another important aspect of regulatory RNAs is that they act through a noncatalytic (or stoichiometric) mechanism in many cases. In other words, they are used only once; this is true, for instance, for most base-pairing sRNAs because they are degraded along with their target mRNA, or because they form extremely stable sRNA–mRNA duplexes. This allows a threshold-linear mode of repression where, below a threshold of transcription, a target mRNA is tightly repressed, whereas above this threshold, repression decreases linearly with the difference between the transcription rate of the target mRNA and that of the sRNA (Levine *et al.*, 2007). The threshold corresponds to the situation where transcription rates of the target mRNA and of the regulatory sRNA are similar; it is therefore a dynamic threshold dependent on conditions that will affect transcription of the sRNA and its target mRNA. This mode of response has important implications: at low transcription, a target mRNA can be tightly repressed by an sRNA, and this tight repression is expected to suppress fluctuations and decrease noise in gene expression (Levine *et al.*, 2007). This can provide an important advantage for targets whose stochastic bursts of protein synthesis issued from noise in transcription can be deleterious. This can also sustain multistable genetic systems. For instance, positively autoregulated genes, such as genes for many TFs, can typically sustain two states of expression, high or low. And repression by a sRNA is expected to both maintain the bistability and increase the probability to remain in the low state over time, when compared to an autoregulated TF not controlled by a sRNA (Levine and Hwa, 2008). Interestingly, *csgD* was shown to display a bistable expression in *Salmonella* (Grantcharova *et al.*, 2010), but whether the multiple sRNAs that control *csgD* expression have a role in this pattern remains to be determined. In this case, bistability could minimize the cost of making a biofilm as only the CsgD ON cells would produce the extracellular matrix, while maintaining the possibility for a fast development of a different phenotype with the CsgD OFF cells.

An additional consequence of the threshold effect is the characteristic de-repression dynamics following the disappearance of the input signal for a repressor sRNA. Indeed, the switch from a repressed to nonrepressed state is predicted to be much more abrupt than with protein-mediated regulation (Levine and Hwa, 2008). It is also expected that stoichiometrically acting sRNAs filter transient signals while selectively responding to sustained stimuli (Legewie *et al.*, 2008; Levine and Hwa, 2008; Mehta *et al.*, 2008).

At last, the very vast majority of sRNAs acting by imperfect base pairing have multiple targets, which will compete for sRNA binding. The regulation of any target will therefore depend on expression of the whole "targetome" of the sRNA, which itself is expected to vary with changing stimuli (e.g., Plumbridge *et al.*, 2014). Furthermore, Hfq-binding sRNAs have been reported to compete for Hfq, and their abundance and activity will thus also depend on expression of all Hfq-binding sRNAs at a given time (Fender *et al.*, 2010; Hussein and Lim, 2011; Moon and Gottesman, 2011). These two types of cross-talk, between targets and

between sRNAs, are expected to set up unique patterns of regulation when conditions vary.

5.1.5.2 sRNAs in mixed regulatory circuits

Network motifs that incorporate sRNAs have been extensively described previously (Beisel and Storz, 2010); we will focus more specifically on two of those: the feedback and the feedforward motifs (Figure 5.1.2).

sRNAs commonly target genes for TFs or sigma factors in a feedback circuit (FC) (also called a *feedback loop*), in which the RNA controls expression of its own TF, and conversely. This can be mediated by direct regulatory interactions (as for the OmpR–OmrA/B FC) but also indirectly. For instance, the RpoE-dependent sRNAs indirectly reduce RpoE activity by repressing expression of multiple or abundant OMPs. Many TFs are subject to negative autoregulation (i.e., they repress their own transcription), which has been shown to increase response time and decrease cell-to-cell variations (Alon, 2007b; Thieffry *et al.*, 1998). However, examples of negative FCs based on two factors regulating each other are very rare in purely transcriptional networks, but are found in cases where the two arms of the loop act on different timescales (e.g., one arm of the loop engages in transcriptional regulation, and the other in protein–protein interaction). In such cases, a negative FC was reported to create oscillations or noise-driven excitable pulses, depending on the timescale of the two regulatory interactions (Figure 5.1.2) (Alon, 2007a; Thieffry, 2007).

Whether this is true as well for those sRNA-containing FCs is still an open question. The possible roles of negative FCs mediated by sRNAs have been addressed experimentally in the *Vibrio* quorum-sensing pathway. Quorum sensing is the process by which bacteria communicate and switch between a single-cell behavior at low cell density (LCD) and a community behavior at high cell density (HCD). In *Vibrio cholerae* and *Vibrio harveyi*, this switch is governed by a mixed regulatory circuit, whose global architecture is similar between the two species, even though some important differences have been described (reviewed in Ng and Bassler, 2009; see Section 21). To summarize, multiple sRNAs called Qrr lie at the core of this pathway and control, in response to cell density and the phosphorylated LuxO RR, the expression of two TFs: HapR (in *V. cholerae*)–LuxR (in *V. harveyi*) and AphA, which are, respectively, the HCD and the LCD master regulators of quorum sensing. Multiple FCs have been described in this pathway, in particular mixed FCs between the Qrr and LuxO, the Qrr and LuxR–HapR (indirect in the case of HapR, as it indirectly activates Qrr synthesis), and the Qrr and AphA (Figure 5.1.2). These FCs provide specific dynamic properties (they delay entry in the HCD mode while accelerating the transition from HCD to LCD) and contribute, with the negative autoregulation of LuxO, LuxR–HapR, and AphA, to minimizing the fluctuations in the levels of the different components of the pathway. This allows a precise timing of the switch between LCD and HCD, with cells coordinately responding to a specific range of input signals (autoinducers), which is likely to be the key for a successful collective behavior (Ng and Bassler, 2009, and references therein; Rutherford *et al.*, 2011; Tu *et al.*, 2010).

sRNAs also participate in positive FCs, where two repressors or two activators control themselves reciprocally: MicF inhibits synthesis of the Lrp master regulator in a double-negative FC (Corcoran *et al.*, 2012; Holmqvist *et al.*, 2012); and RyhB also forms a double-negative FC with the Fur TF, in this case through both a direct pairing and indirect control of Fur activity (Massé *et al.*, 2005; Vecerek *et al.*, 2007). A positive FC composed of two TFs has been reported to establish bistability, with both regulators being on or off (for double-positive FC) or, in contrast, with one regulator being on and the other off (for double-negative FC). In line with this, the MicF-Lrp FC was proposed to ensure that the levels of each regulator followed a reciprocal pattern with Lrp or MicF predominating in, respectively, nutrient-poor or nutrient-rich medium (Holmqvist *et al.*, 2012).

In several cases of mixed regulatory circuits, a TF controlled expression of genes both directly and indirectly through the control of a regulatory RNA, thereby creating a regulatory motif known as a feedforward motif (FFM; also called a *feedforward loop*). This is what is found, for instance, in the case of the multi-output FFM between CRP and Spot42 to control genes involved in sugar utilization, or in the control of *ompC*, *ompF*, or *flhDC* by OmpR, and by the OmpR-regulated MicC, MicF, and OmrA/B sRNAs, respectively (Figures 5.1.1 and 5.1.2). These are examples of coherent FFMs, where direct and indirect control both activate or repress gene expression. But regulatory RNAs are also found in incoherent FFMs, where two arms of the regulatory circuit have opposite effects: OmpR activates *csgD* transcription directly, while also promoting synthesis of OmrA/B, two sRNAs that repress *csgD* expression (Holmqvist *et al.*, 2010). Similarly, PhoP activates transcription of the MgtC virulence factor in *Salmonella*, but also of an asRNA named AmgR, which in turn limits MgtC production (Figure 5.1.2) (Lee and Groisman, 2010).

FFM properties have been studied in detail in purely transcriptional networks. Depending on the different interactions that can take place between the two regulators and the target, eight types of FFMs have been defined (four types of coherent FFMs and four types of incoherent ones). In general, coherent FFMs are expected to induce a delay in the (on or off) response and to filter transient (on or off) signals, whereas incoherent FFMs can increase response time and sometimes generate a pulse in gene expression (Figure 5.1.2) (Mangan and Alon, 2003). Interestingly, the mixed coherent FFMs that include transcriptional regulatory proteins and posttranscriptional repressor RNAs have been proposed to provide a tighter control of gene expression and a faster recovery following deactivation when compared to purely transcriptional FFMs (Shimoni *et al.*, 2007). In the case of the CRP/Spot42-mediated FFM, the Spot42 sRNA was shown to further repress expression of the target genes, and to affect the dynamics of the regulation. More precisely, Spot42 delayed the response to loop activation (similar to what is expected for a similar TF-based FFM) (Figure 5.1.2), but also

accelerated the response following loop deactivation (Beisel and Storz, 2011a). This last property has not been observed for purely transcriptional FFMs of the same type, and suggests that network properties could differ between mixed regulatory networks and purely transcriptional ones (Beisel and Storz, 2011b).

5.1.6 Conclusions

Regulatory RNAs allow a rapid and dynamic response of bacteria to changes in their environment, and they are consistently widespread in stress responses. In many cases, they are part of mixed regulatory circuits that combine (i) transcriptional and posttranscriptional control, and (ii) protein and RNA regulators. Although the few examples studied in detail, such as those described in this chapter, suggested that regulatory RNAs can provide specific properties over proteins, the precise role of RNA in the very vast majority of mixed circuits remains to be precisely characterized.

Furthermore, not only are many of those circuits certainly still incomplete, but also more connections between transcriptional regulators and sRNAs will undoubtedly be characterized in the future, in diverse bacteria. These findings should help answer important questions such as the role of those circuits in gene expression, the possible over- or underrepresentation of some motifs, and the conservation of circuits to control the same functions in distant bacteria where sRNAs themselves might not be conserved.

Acknowledgements

We thank Anaïs Brosse (our lab) for unpublished results and Eliane Hajnsdorf and Mathias Springer for comments on the manuscript. Our work is supported by the CNRS, ANR grants (ANR-2010-JCJC-130801 and ANR-14-CE10-0004-01) to MG and by the Initiative d'Excellence program from the French State (Grant DYNAMO, no. ANR-11-LABX-0011). JJ is the recipient of a DYNAMO fellowship.

References

Alon U. 2007a. *An introduction to systems biology: design principles of biological circuits*. Boca Raton, FL: Chapman&Hall/CRC.

Alon U. 2007b. Network motifs: theory and experimental approaches. *Nature Reviews Genetics* 8: 450–461.

Altuvia S, Weinstein-Fischer D, Zhang A, Postow L, Storz G. 1997. A small stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator. *Cell* 90: 43–53.

Altuvia S, Zhang A, Argaman L, Tiwari A, Storz G. 1998. The *Escherichia coli* OxyS regulatory RNA represses *fhlA* translation by blocking ribosome binding. *EMBO Journal* 17: 6069–6075.

Batchelor E, Goulian M. 2003. Robustness and the cycle of phosphorylation and dephosphorylation in a two-component regulatory system. *Proceedings of the National Academy of Sciences USA* 100: 691–696.

Battesti A, Majdalani N, Gottesman S. 2011. The RpoS-mediated general stress response in *Escherichia coli*. *Annual Review of Microbiology* 65: 189–213.

Beckmann BM, Hoch PG, Marz M, Willkomm DK, Salas M, *et al.* 2012. A pRNA-induced structural rearrangement triggers 6S-1 RNA release from RNA polymerase in *Bacillus subtilis*. *EMBO Journal* 31: 1727–1738.

Beisel CL, Storz G. 2010. Base pairing small RNAs and their roles in global regulatory networks. *FEMS Microbiology Reviews* 34: 866–882.

Beisel CL, Storz G. 2011a. The base-pairing RNA Spot 42 participates in a multioutput feedforward loop to help enact catabolite repression in *Escherichia coli*. *Molecular Cell* 41: 286–297.

Beisel CL, Storz G. 2011b. Discriminating tastes: physiological contributions of the Hfq-binding small RNA Spot 42 to catabolite repression. *RNA Biology* 8: 766–770.

Bernstein JA, Khodursky AB, Lin PH, Lin-Chao S, Cohen SN. 2002. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proceedings of the National Academy of Sciences USA* 99: 9697–9702.

Bikard D, Marraffini LA. 2013. Control of gene expression by CRISPR-Cas systems. *F1000Prime Reports* 5: 47.

Bordeau V, Felden B. 2014. Curli synthesis and biofilm formation in enteric bacteria are controlled by a dynamic small RNA module made up of a pseudoknot assisted by an RNA chaperone. *Nucleic Acids Research* 42: 4682–4696.

Cavanagh AT, Sperger JM, Wassarman KM. 2012. Regulation of 6S RNA by pRNA synthesis is required for efficient recovery from stationary phase in *E. coli* and *B. subtilis*. *Nucleic Acids Research* 40: 2234–2246.

Cavanagh AT, Wassarman KM. 2014. 6S RNA, a global regulator of transcription in *Escherichia coli, Bacillus subtilis*, and beyond. *Annual Review of Microbiology* 68: 45–60.

Cech TR, Bass BL. 1986. Biological catalysis by RNA. *Annual Review of Biochemistry* 55: 599–629.

Chao Y, Papenfort K, Reinhardt R, Sharma CM, Vogel J. 2012. An atlas of Hfq-bound transcripts reveals 3′ UTRs as a genomic reservoir of regulatory small RNAs. *EMBO Journal* 31: 4005–4019.

Charpentier E, Doudna JA. 2013. Biotechnology: rewriting a genome. *Nature* 495: 50–51.

Coornaert A, Chiaruttini C, Springer M, Guillier M. 2013. Post-transcriptional control of the W*Escherichia coli* PhoQ-PhoP two-component system by multiple sRNAs involves a novel pairing region of GcvB. *PLoS Genetics* 9: e1003156.

Coornaert A, Lu A, Mandin P, Springer M, Gottesman S, *et al.* 2010. MicA sRNA links the PhoP regulon to cell envelope stress. *Molecular Microbiology* 76: 467–479.

Corcoran CP, Podkaminski D, Papenfort K, Urban JH, Hinton JC, *et al.* 2012. Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA. *Molecular Microbiology.* 84: 428–445.

DebRoy S, Gebbie M, Ramesh A, Goodson JR, Cruz MR, *et al.* 2014. Riboswitches. A riboswitch-containing sRNA controls gene expression by sequestration of a response regulator. *Science* 345: 937–940.

De Lay N, Gottesman S. 2009. The Crp-activated small noncoding regulatory RNA CyaR (RyeE) links nutritional status to group behavior. *Journal of Bacteriology* 191: 461–476.

De Lay N, Gottesman S. 2012. A complex network of small non-coding RNAs regulate motility in *Escherichia coli*. *Molecular Microbiology* 86: 524–538.

Delihas N, Forst S. 2001. *MicF:* An antisense RNA gene involved in response of *Escherichia coli* to global stress factors. *Journal of Molecular Biology* 313: 1–12.

Dudin O, Geiselmann J, Ogasawara H, Ishihama A, Lacour S. 2014. Repression of flagellar genes in exponential phase by CsgD and CpxR, two crucial modulators of *Escherichia coli* biofilm formation. *Journal of Bacteriology* 196: 707–715.

Durand S, Storz G. 2010. Reprogramming of anaerobic metabolism by the FnrS small RNA. *Molecular Microbiology* 75: 1215–1231.

Fender A, Elf J, Hampel K, Zimmermann B, Wagner EG. 2010. RNAs actively cycle on the Sm-like protein Hfq. *Genes & Development* 24: 2621–2626.

Figueroa-Bossi N, Lemire S, Maloriol D, Balbontin R, Casadesus J, *et al.* 2006. Loss of Hfq activates the s^E-dependent envelope stress response in *Salmonella enterica*. *Molecular Microbiology* 62: 838–852.

Frohlich KS, Papenfort K, Berger AA, Vogel J. 2012. A conserved RpoS-dependent small RNA controls the synthesis of major porin OmpD. *Nucleic Acids Research* 40: 3623–3640.

Gogol EB, Rhodius VA, Papenfort K, Vogel J, Gross CA. 2011. Small RNAs endow a transcriptional activator with essential repressor functions for single-tier control of a global stress regulon. *Proceedings of the National Academy of Sciences USA* 108: 12875–12880.

Gopel Y, Gorke B. 2012. Rewiring two-component signal transduction with small RNAs. *Current Opinion in Microbiology* 15: 132–139.

Gopel Y, Luttmann D, Heroven AK, Reichenbach B, Dersch P, *et al.* 2011. Common and divergent features in transcriptional control of the homologous small RNAs GlmY and GlmZ in Enterobacteriaceae. *Nucleic Acids Research* 39: 1294–1309.

Gottesman S, Storz G. 2011. Bacterial small RNA regulators: versatile roles and rapidly evolving variations. *Cold Spring Harbor Perspectives in Biology* doi:10.1101/cshperspect.a003798

Grantcharova N, Peters V, Monteiro C, Zakikhany K, Romling U. 2010. Bistable expression of CsgD in biofilm development of *Salmonella enterica* serovar Typhimurium. *Journal of Bacteriology* 192: 456–466.

Groisman EA, Hollands K, Kriner MA, Lee EJ, Park SY, *et al.* 2013. Bacterial Mg homeostasis, transport, and virulence. *Annual Review of Genetics* 47: 625–646.

Guillier M, Gottesman S. 2006. Remodelling of the *Escherichia coli* outer membrane by two small regulatory RNAs. *Molecular Microbiology* 59: 231–247.

Guillier M, Gottesman S. 2008. The 5′ end of two redundant sRNAs is involved in the regulation of multiple targets, including their own regulator. *Nucleic Acids Research* 36: 6781–6794.

Guo MS, Updegrove TB, Gogol EB, Shabalina SA, Gross CA, *et al.* 2014. MicL, a new sigmaE-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. *Genes & Development* 28: 1620–1634.

Holmqvist E, Reimegard J, Sterk M, Grantcharova N, Romling U, *et al.* 2010. Two antisense RNAs target the transcriptional regulator CsgD to inhibit curli synthesis. *EMBO Journal* 29: 1840–1850.

Holmqvist E, Unoson C, Reimegard J, Wagner EG. 2012. A mixed double negative feedback loop between the sRNA MicF and the global regulator Lrp. *Molecular Microbiology* 84: 414–427.

Hussein R, Lim HN. 2011. Disruption of small RNA signaling caused by competition for Hfq. *Proceedings of the National Academy of Sciences USA* 108: 1110–1115.

Jacob F, Monod J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* 3: 318–356.

Jacques JF, Jang S, Prevost K, Desnoyers G, Desmarais M, *et al.* 2006. RyhB small RNA modulates the free intracellular iron pool and is essential for normal growth during iron limitation in *Escherichia coli*. *Molecular Microbiology* 62: 1181–1190.

Johansen J, Eriksen M, Kallipolitis B, Valentin-Hansen P. 2008. Down-regulation of outer membrane proteins by noncoding RNAs: unraveling the cAMP-CRP- and sigmaE-dependent CyaR-ompX regulatory case. *Journal of Molecular Biology* 383: 1–9.

Johansen J, Rasmussen AA, Overgaard M, Valentin-Hansen P. 2006. Conserved small non-coding RNAs that belong to the sigma(E) regulon: role in down-regulation of outer membrane proteins. *Journal of Molecular Biology* 364: 1–8.

Jorgensen MG, Nielsen JS, Boysen A, Franch T, Moller-Jensen J, et al. 2012. Small regulatory RNAs control the multi-cellular adhesive lifestyle of *Escherichia coli*. *Molecular Microbiology* 84: 36–50.

Jorgensen MG, Thomason MK, Havelund J, Valentin-Hansen P, Storz G. 2013. Dual function of the McaS small RNA in controlling biofilm formation. *Genes & Development* 27: 1132–1145.

Laub MT, Goulian M. 2007. Specificity in two-component signal transduction pathways. *Annual Review of Genetics* 41: 121–145.

Lease RA, Cusick M, Belfort M. 1998. Riboregulation in *Escherichia coli:* DsrA RNA acts by RNA:RNA interactions at multiple loci. *Proceedings of the National Academy of Sciences USA* 95: 12456–12461.

Lee EJ, Groisman EA. 2010. An antisense RNA that governs the expression kinetics of a multifunctional virulence gene. *Molecular Microbiology* 76: 1020–1033.

Legewie S, Dienst D, Wilde A, Herzel H, Axmann IM. 2008. Small RNAs establish delays and temporal thresholds in gene expression. *Biophysical Journal* 95: 3232–3238.

Levine E, Hwa T. 2008. Small RNAs establish gene expression thresholds. *Current Opinion in Microbiology* 11: 574–579.

Levine E, Zhang Z, Kuhlman T, Hwa T. 2007. Quantitative characteristics of gene regulation by small RNA. *PLoS Biology* 5: e229.

Lima S, Guo MS, Chaba R, Gross CA, Sauer RT. 2013. Dual molecular signals mediate the bacterial response to outer-membrane stress. *Science* 340: 837–841.

Loh E, Dussurget O, Gripenland J, Vaitkevicius K, Tiensuu T, *et al.* 2009. A trans-acting riboswitch controls expression of the virulence regulator PrfA in *Listeria monocytogenes*. *Cell* 139: 770–779.

Majdalani N, Chen S, Murrow J, St. John K, Gottesman S. 2001. Regulation of RpoS by a novel small RNA: the characterization of RprA. *Molecular Microbiology* 39: 1382–1394.

Majdalani N, Cunning C, Sledjeski D, Elliott T, Gottesman S. 1998. DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription. *Proceedings of the National Academy of Sciences USA* 95: 12462–12467.

Majdalani N, Hernandez D, Gottesman S. 2002. Regulation and mode of action of the second small RNA activator of RpoS translation, RprA. *Molecular Microbiology* 46: 813–826.

Mandin P, Gottesman S. 2009. A genetic approach for finding small RNAs regulators of genes of interest identifies RybC as regulating the DpiA/DpiB two-component system. *Molecular Microbiology* 72: 551–565.

Mandin P, Gottesman S. 2010. Integrating anaerobic/aerobic sensing and the general stress response through the ArcZ small RNA. *EMBO Journal* 29: 3094–3107.

Mandin P, Guillier M. 2013. Expanding control in bacteria: interplay between small RNAs and transcriptional regulators to control gene expression. *Current Opinion in Microbiology* 16: 125–132.

Mangan S, Alon U. 2003. Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Sciences USA* 100: 11980–11985.

Massé E, Gottesman S. 2002. A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proceedings of the National Academy of Sciences USA* 99: 4620–4625.

Massé E, Salvail H, Desnoyers G, Arguin M. 2007. Small RNAs controlling iron metabolism. *Current Opinion in Microbiology* 10: 140–145.

Massé E, Vanderpool CK, Gottesman S. 2005. Effect of RyhB small RNA on global iron use in *Escherichia coli*. *Journal of Bacteriology* 187: 6962–6971.

Mehta P, Goyal S, Wingreen NS. 2008. A quantitative comparison of sRNA-based and protein-based gene regulation. *Molecular Systems Biology* 4: 221.

Mellin JR, Koutero M, Dar D, Nahori MA, Sorek R, *et al.* 2014. Riboswitches. Sequestration of a two-component response regulator by a riboswitch-regulated noncoding RNA. *Science* 345: 940–943.

Mika F, Busse S, Possling A, Berkholz J, Tschowri N, *et al.* 2012. Targeting of *csgD* by the small regulatory RNA RprA links stationary phase, biofilm formation and cell envelope stress in *Escherichia coli*. *Molecular Microbiology* 84: 51–65.

Mika F, Hengge R. 2013. Small regulatory RNAs in the control of motility and biofilm formation in *E. coli* and *Salmonella*. *International Journal of Molecular Sciences* 14: 4560–4579.

Modi SR, Camacho DM, Kohanski MA, Walker GC, Collins JJ. 2011. Functional characterization of bacterial sRNAs using a network biology approach. *Proceedings of the National Academy of Sciences USA* 108: 15522–15527.

Monteiro C, Papenfort K, Hentrich K, Ahmad I, Le Guyon S, *et al.* 2012. Hfq and Hfq-dependent small RNAs are major contributors to multicellular development in *Salmonella enterica* serovar Typhimurium. *RNA Biology* 9: 489–502.

Moon K, Gottesman S. 2009. A PhoQ/P-regulated small RNA regulates sensitivity of *Escherichia coli* to antimicrobial peptides. *Molecular Microbiology* 74: 1314–1330.

Moon K, Gottesman S. 2011. Competition among Hfq-binding small RNAs in *Escherichia coli*. *Molecular Microbiology* 82: 1545–1562.

Ng WL, Bassler BL. 2009. Bacterial quorum-sensing network architectures. *Annual Review of Genetics* 43: 197–222.

Ogasawara H, Yamada K, Kori A, Yamamoto K, Ishihama A. 2010. Regulation of the *Escherichia coli csgD* promoter: interplay between five transcription factors. *Microbiology* 156: 2470–2483.

Ogasawara H, Yamamoto K, Ishihama A. 2011. Role of the biofilm master regulator CsgD in cross-regulation between biofilm formation and flagellar synthesis. *Journal of Bacteriology* 193: 2587–2597.

Opdyke JA, Fozo EM, Hemm MR, Storz G. 2011. RNase III participates in GadY-dependent cleavage of the *gadX-gadW* mRNA. *Journal of Molecular Biology* 406: 29–43.

Opdyke JA, Kang JG, Storz G. 2004. GadY, a small-RNA regulator of acid response genes in *Escherichia coli*. *Journal of Bacteriology* 186: 6698–6705.

Österberg S, del Peso-Santos T, Shingler V. 2011. Regulation of alternative sigma factor use. *Annual Review of Microbiology* 65: 37–55.

Papenfort K, Pfeiffer V, Lucchini S, Sonawane A, Hinton JC, *et al.* 2008. Systematic deletion of *Salmonella* small RNA genes identifies CyaR, a conserved CRP-dependent riboregulator of OmpX synthesis. *Molecular Microbiology* 68: 890–906.

Papenfort K, Pfeiffer V, Mika F, Luchhini S, Hinton JCD, *et al.* 2006. σ^E-dependent small RNAs of *Salmonella* respond to membrane stress by accelerating global omp mRNA decay. *Molecular Microbiology* 62: 1674–1688.

Papenfort K, Said N, Welsink T, Lucchini S, Hinton JC, *et al.* 2009. Specific and pleiotropic patterns of mRNA regulation by ArcZ, a conserved, Hfq-dependent small RNA. *Molecular Microbiology* 74: 139–158.

Peer A, Margalit H. 2014. Evolutionary patterns of *Escherichia coli* small RNAs and their regulatory interactions. *RNA* 20: 994–1003.

Pfeiffer V, Sittka A, Tomer R, Tedin K, Brinkmann V, *et al.* 2007. A small non-coding RNA of the invasion gene island (SPI-1) represses outer membrane protein synthesis from the *Salmonella core* genome. *Molecular Microbiology* 66: 1174–1191.

Plumbridge J, Bossi L, Oberto J, Wade JT, Figueroa-Bossi N. 2014. Interplay of transcriptional and small RNA-dependent control mechanisms regulates chitosugar uptake in *Escherichia coli* and *Salmonella*. *Molecular Microbiology* 92: 648–658.

Polayes DA, Rice PW, Garner MM, Dahlberg JE. 1988. Cyclic AMP-cyclic AMP receptor protein as a repressor of transcription of the *spf* gene of *Escherichia coli*. *Journal of Bacteriology* 170: 3110–3114.

Povolotsky TL, Hengge R. 2012. "Life-style" control networks in *Escherichia coli*: signaling by the second messenger c-di-GMP. *Journal of Biotechnology* 160: 10–16.

Pratt LA, Hsing W, Gibson KE, Silhavy TJ. 1996. From acids to *osmZ*: multiple factors influence synthesis of the OmpF and OmpC porins in *Escherichia coli*. *Molecular Microbiology* 20: 911–917.

Prevost K, Salvail H, Desnoyers G, Jacques JF, Phaneuf E, *et al.* 2007. The small RNA RyhB activates the translation of *shiA* mRNA encoding a permease of shikimate, a compound involved in siderophore synthesis. *Molecular Microbiology* 64: 1260–1273.

Reichenbach B, Gopel Y, Gorke B. 2009. Dual control by perfectly overlapping sigma 54- and sigma 70- promoters adjusts small RNA GlmY expression to different environmental signals. *Molecular Microbiology* 74: 1054–1070.

Rutherford ST, van Kessel JC, Shao Y, Bassler BL. 2011. AphA and LuxR/HapR reciprocally control quorum sensing in vibrios. *Genes & Development*. 25: 397–408.

Sampson TR, Saroj SD, Llewellyn AC, Tzeng YL, Weiss DS. 2013a. Corrigendum: A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature* 501: 262.

Sampson TR, Saroj SD, Llewellyn AC, Tzeng YL, Weiss DS. 2013b. A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature* 497: 254–257.

Shearwin KE, Callen BP, Egan JB. 2005. Transcriptional interference – a crash course. *Trends in Genetics* 21: 339–345.

Shimoni Y, Friedlander G, Hetzroni G, Niv G, Altuvia S, et al. 2007. Regulation of gene expression by small non-coding RNAs: a quantitative view. *Molecular Systems Biology* 3: 138.

Silva IJ, Ortega AD, Viegas SC, Garcia-Del Portillo F, Arraiano CM. 2013. An RpoS-dependent sRNA regulates the expression of a chaperone involved in protein folding. *RNA* 19: 1253–1265.

Sorek R, Lawrence CM, Wiedenheft B. 2013. CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annual Review of Biochemistry* 82: 237–266.

Soutourina OA, Bertin PN. 2003. Regulation cascade of flagellar expression in Gram-negative bacteria. *FEMS Microbiology Reviews* 27: 505–523.

Stork M, Di Lorenzo M, Welch TJ, Crosa JH. 2007. Transcription termination within the iron transport-biosynthesis operon of *Vibrio anguillarum* requires an antisense RNA. *Journal of Bacteriology* 189: 3479–3488.

Thieffry D. 2007. Dynamical roles of biological regulatory circuits. *Briefings in Bioinformatics* 8: 220–225.

Thieffry D, Huerta AM, Perez-Rueda E, Collado-Vides J. 1998. From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in *Escherichia coli*. *Bioessays* 20: 433–440.

Thomason MK, Fontaine F, De Lay N, Storz G. 2012. A small RNA that regulates motility and biofilm formation in response to changes in nutrient availability in *Escherichia coli*. *Molecular Microbiology* 84: 17–35.

Thomason MK, Storz G. 2010. Bacterial antisense RNAs: how many are there, and what are they doing? *Annual Review of Genetics* 44: 167–188.

Thompson KM, Rhodius VA, Gottesman S. 2007. s^E regulates and is regulated by a small RNA in *Escherichia coli*. *Journal of Bacteriology* 189: 4243–4256.

Tu KC, Long T, Svenningsen SL, Wingreen NS, Bassler BL. 2010. Negative feedback loops involving small regulatory RNAs precisely control the *Vibrio harveyi* quorum-sensing response. *Molecular Cell* 37: 567–579.

Udekwu KI, Wagner EG. 2007. Sigma E controls biogenesis of the antisense RNA MicA. *Nucleic Acids Research* 35: 1279–1288.

Urbanowski ML, Stauffer LT, Stauffer GV. 2000. The *gcvB* gene encodes a small untranslated RNA involved in expression of the dipeptide and oligopeptide transport systems in *Escherichia coli*. *Molecular Microbiology* 37: 856–868.

Vanderpool CK, Gottesman S. 2007. The novel transcription factor SgrR coordinates the response to glucose-phosphate stress. *Journal of Bacteriology* 189: 2238–2248.

Vecerek B, Moll I, Blasi U. 2007. Control of Fur synthesis by the noncoding RNA RyhB and iron-responsive decoding. *EMBO Journal* 26: 965–975.

Vogel J, Luisi BF. 2011. Hfq and its constellation of RNA. *Nature Reviews Microbiology* 9: 578–589.

Wassarman KM, Saecker RM. 2006. Synthesis-mediated release of a small RNA inhibitor of RNA polymerase. *Science* 314: 1601–1603.

Annexe 2 : Les autres cibles d'OmrA/B

A. L'interaction entre OmrA/B et l'ARNm *fecA*

Au cours de mon M2, j'ai pu étudier l'interaction entre les sRNA OmrA/B et l'ARNm *fecA* codant le récepteur au citrate de fer. Si de premières expériences de mutations compensatoires permettaient d'observer une légère compensation de l'effet des OmrA/B mutants sur les niveaux d'un ARNm *fecA* mutant et la synthèse de FecA à partir de cet ARNm, indiquant une régulation directe par appariement, l'effet observé au cours de ces expériences était difficilement reproductible.

J'ai donc essayé de reproduire ce résultat avec les mêmes mutations (Fig. 76.A) mais d'une manière différente, en observant la croissance de bactéries en milieu NB- citrate-dip (NBCD), où la principale source de fer disponible est le citrate de fer. Afin de vérifier que la croissance des bactéries sur ce milieu dépendait bien de la protéine FecA, j'ai utilisé des mutants de *fecA*, ainsi que des deux autres gènes de récepteurs *fepA* et *cirA* qui pourraient améliorer la survie *via* l'utilisation d'entérobactine ou de DHB-Ser pour l'import de fer (Fig. 76.B). Dans ces conditions, on observe bien une forte inhibition de croissance dépendante de la délétion de *fecA* sur milieu NBCD, d'autant plus importante en l'absence de *fepA*. On note donc que FepA semble être la principale voie d'import de fer dans ces conditions, lorsque FecA ne peut assurer l'import de citrate de fer.

Afin de poursuivre avec la démonstration de l'appariement entre les OmrA/B et l'ARNm *fecA*, j'ai analysé l'effet des surexpressions d'OmrA et d'OmrB sur la croissance en milieu NBCD+ (à la concentration en dip plus élevée, permettant une observation plus nette du phénotype provoqué par OmrA/B) (Fig. 76.C). Ces surexpressions induisent respectivement une inhibition de croissance d'un facteur 10^3 et 10 tandis que les versions mutantes d'OmrA/B n'ont pas d'effet sur la croissance. Inversement, lorsque l'ARNm *fecA* est muté au niveau du site d'appariement avec les OmrA/B, la surexpression des sRNA wt n'a qu'un très faible effet sur la croissance – d'un facteur inférieur à 10, tandis que la surexpression des sRNA mutants restaure la totalité de l'inhibition de croissance pour OmrAmut2 comme pour OmrBmut2.

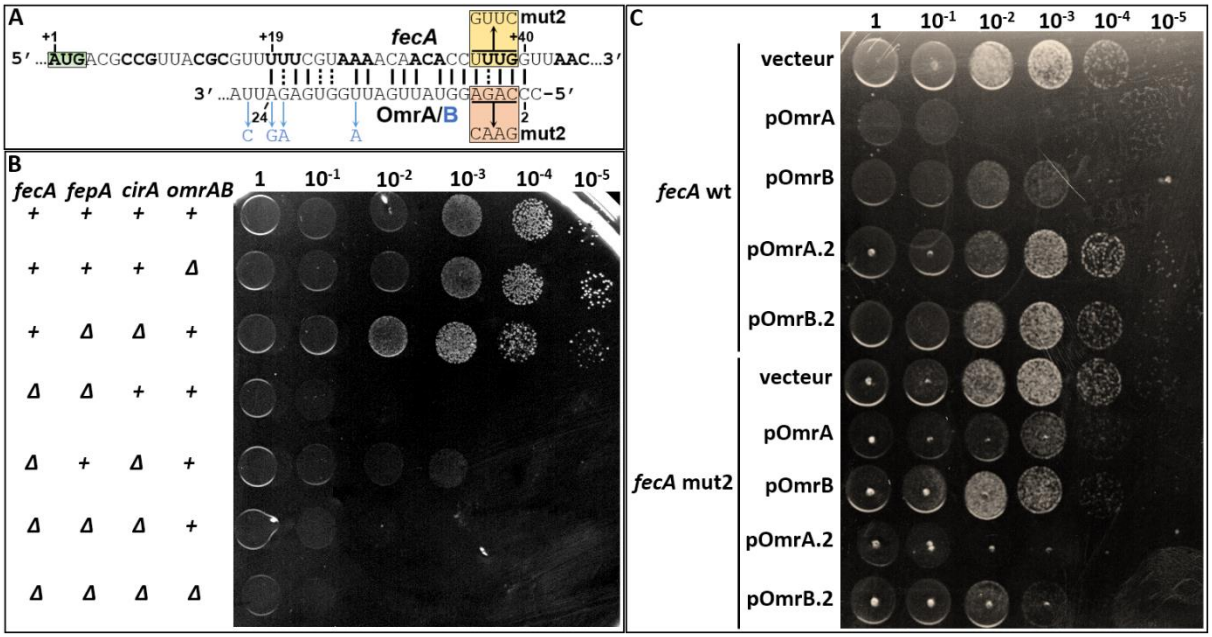


Figure 76 – OmrA/B répriment l'expression de fecA via un appariement
A. Prédiction d'appariement entre l'ARNm fecA et les sRNA OmrA et OmrB. Le codon de démarrage de la traduction est encadré en vert. Les positions délimitant l'appariement sont indiquées. B. Croissance de bactéries wt (souche DJ624), ΔOmrA/B (souche MG1099), ΔfepA ΔcirA (souche JJ0062), ΔfecA ΔfepA (souche JJ0063), ΔfecA ΔcirA (souche JJ0064), ΔfecA ΔfepA ΔcirA (souche JJ0065) et ΔfecA ΔfepA ΔcirA ΔOmrA/B (souche JJ0066) sur milieu NB – 50µM Na-Citrate – 250µM dip (NBCD). Les bactéries ont été cultivées jusqu'à saturation en LB, les cultures ont été normalisées à 2,5DO/mL puis diluées en série d'un facteur 10 par dilution. Des gouttes de 2µL ont ensuite été déposées sur boîtes de Pétri NBCD, et les bactéries ont poussé pendant 48h avant analyse C. Croissance de bactéries ΔfepA, ΔcirA, ΔOmrA/B (souche JJ0113) et de sa dérivée fecAmut2 (souche JJ0091) transformées par un plasmide contrôle (vecteur) ou surproducteur d'OmrA, OmrB, OmrAmut2 (OmrA.2) ou OmrBmut2 (OmrB.2) sur milieu NBCD+ (500µM dip au lieu de 250µM), selon la même procédure qu'en (B).

La répression de *fecA* par les OmrA/B est donc directe et implique la région 5' d'OmrA/B et une région à 40nt en aval du RBS dans l'ARNm *fecA*. Nous avons pensé à un mécanisme de régulation similaire à celui observé pour *fepA*, néanmoins aucune structure vraiment convaincante ne peut être prédite aux alentours de 20nt en aval du RBS, aussi le mécanisme de régulation de *fecA* par OmrA/B reste à déterminer. On retrouve par ailleurs un effet régulateur beaucoup plus fort d'OmrA par rapport à OmrB, comme c'est le cas pour de nombreuses cibles d'OmrA/B. Par ailleurs, cette expérience met en évidence un phénotype lié à la surexpression d'OmrA/B, or à ce jour les phénotypes associés aux sRNA Hfq-dépendants restent relativement rare. Cela permet d'imaginer que d'autres de ces sRNA pourraient provoquer des phénotypes, pour peu qu'ils soient étudiés dans les conditions de culture et les contextes génétiques appropriés (voir **Annexe 3**).

B. Recherche de nouvelles cibles d'OmrA/B par RNAseq

Le résultat d'une analyse transcriptomique réalisée après une surexpression d'OmrA/B par une méthode de puce à ADN (Guillier and Gottesman, 2006) suggère fortement qu'en plus des cibles déjà connues d'OmrA/B, de nombreuses autres restent à découvrir. Afin de tirer profit des techniques d'analyse de transcriptome plus récentes et plus puissantes, j'ai pu réaliser une expérience de RNAseq en collaboration avec la plateforme « Transcriptome and Epigenome » de l'Institut Pasteur, dirigée par Jean-Yves Coppee. Nous avons choisi de surexprimer les sRNA durant un temps court afin de nous affranchir au maximum des cibles d'une régulation indirecte. De plus, l'usage des OmrA/B wt et mut2 (voir Fig. 76.A) avait pour but de nous permettre de distinguer les cibles spécifiques de ces sRNA, puisque les autres régions d'OmrA/B n'ont jamais été impliquées dans l'appariement avec des cibles ARNm jusqu'ici. J'ai débuté l'analyse de ce RNAseq, qui a été poursuivie au laboratoire par la post-doctorante Thao Nguyen Le Lam.

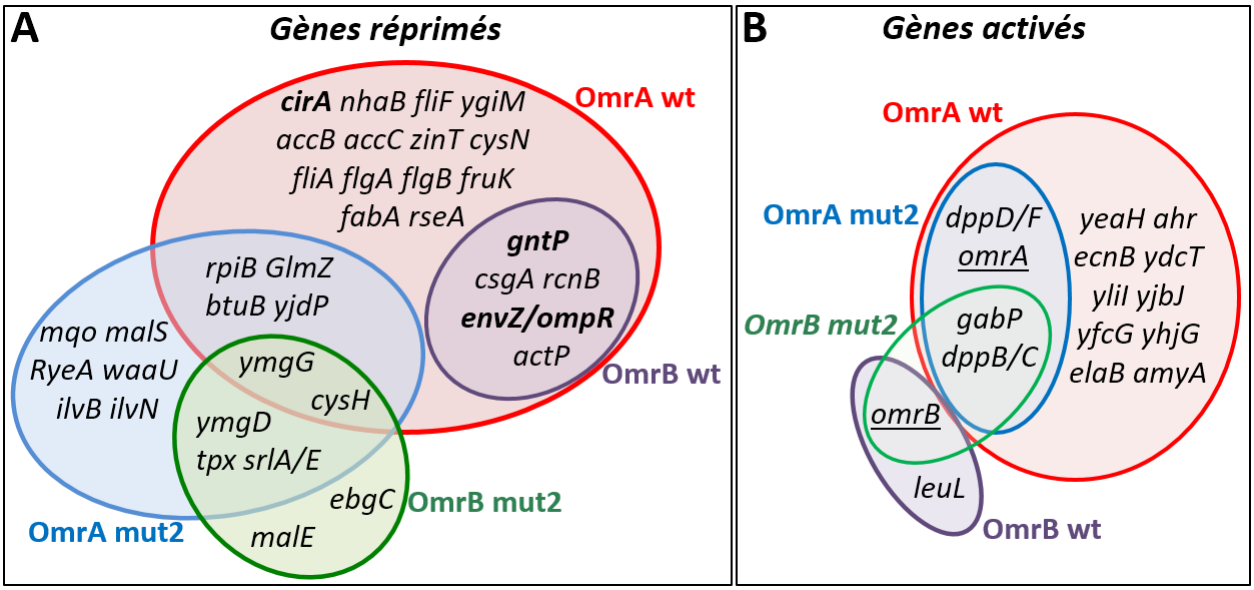


Figure 77 – Gènes dont les niveaux d'ARNm sont affectés par OmrA/B d'après les résultats de RNAseq

Représentations des gènes réprimés (A) ou activés (B) par OmrA/B, wt ou mut2. Une souche délétée pour omrA et omrB (MG1099) *a été transformée par un plasmide contrôle ou surproducteur d'OmrA, OmrB, OmrAmut2 ou OmrBmut2 sous le contrôle d'un promoteur inductible à l'IPTG. Les bactéries ont été cultivées jusqu'à une* OD_{600} *de 0,3 puis traitées avec 100µM d'IPTG pendant 10min pour induire l'expression des sRNA. Sont indiqués les gènes pour lesquels une régulation d'un facteur supérieur ou égal à 1,5 est observée avec une P-value inférieure à 0,05. Les gènes figurant dans plusieurs sous-ensemble sont régulés par plusieurs sRNA comme indiqué (i.e. gntP est réprimé par OmrA et OmrB wt par exemple). Les gènes en gras sont des cibles connues d'OmrA/B, tandis que les gènes omrA et omrB sont soulignés et correspondent à un contrôle de la surexpression de ces gènes.*

Nous retrouvons un certain nombre de gènes cibles connus d'OmrA/B parmi les gènes dont les niveaux d'ARNm sont affectés (Fig. 77), tels que *cirA*, *gntP* et *envZ-ompR*. Notons toutefois que les gènes *ompT*, *fecA* et *csgD* sont bien trouvés comme des gènes régulés par OmrA/B avec des facteurs de répression inférieurs à 1,5 et/ou des P-value supérieures 0,05, de même que *fepA* avec le sRNA OmrA. Cela témoigne de la réalisation correcte de cette expérience. L'analyse de ces résultats révèle ainsi que de nombreux gènes pourraient être régulés par au moins l'un de ces deux sRNA si ce n'est les deux sRNA OmrA et OmrB. Il convient de remarquer que les régulations d'un certain nombre de ces gènes-cibles semblent être indirectes. Ainsi, les régulations de *fliA*, *flgA* et *flgB* passent certainement par celle de *flhDC*, tandis que la régulation de l'expression de *csgA* passe vraisemblablement par celle de *csgD*.

Sans entrer dans le détail de l'ensemble de ces gènes candidats à la régulation par OmrA/B, certains gènes apparaissent comme des candidats à la fois probables et particulièrement intéressants à mettre en perspective avec le restant des résultats présentés dans ce manuscrit. Ainsi, *rcnB* apparait comme l'une des nouvelles cibles les plus convaincantes d'après ce RNAseq, avec une répression par les deux sRNA wt qui est perdue avec la version mutante des sRNA. Ce gène est régulé par Fur, et code une protéine périplasmique impliquée dans l'export de nickel et de cobalt et la résistance à ces métaux toxiques, mais aussi à la résistance au cuivre (Blériot et al., 2011, 2014). Il constitue un exemple de plus d'un gène régulé au niveau transcriptionnel par Fur et très vraisemblablement au niveau post-transcriptionnel par OmrA/B.

Dans la même catégorie de fonction on peut également compter *zinT*, une autre cible potentielle d'OmrA, mais sans doute également d'OmrB. La régulation positive de l'expression de ce gène en présence de cadmium dépend de Fur, SoxS, RelA et SpoT (Puškárová et al., 2002), tandis que son expression est négativement régulée par Zur en présence de zinc (Graham et al., 2009). Ce gène code en outre une protéine impliquée dans la résistance à ces deux métaux (Graham et al., 2009; Stojnev et al., 2007). Il est donc frappant de constater qu'à nouveau, un gène régulé par Fur l'est également par les OmrA/B – même si dans ce cas précis la régulation positive par Fur n'est pas tout à fait comprise d'un point de vue mécanistique et pourrait être indirecte. C'est également à nouveau un gène impliqué dans la résistance à des

concentrations toxiques en divers ions métalliques. Il est intéressant de noter ici qu'en plus du fer, le nickel, le cobalt et le cuivre représentent des voix possibles pour la réaction de Fenton, tandis que le zinc semble catalyser cette réaction (Friedrich et al., 2012; Lloyd and Phillips, 1999). Tout cela indique que la régulation par les OmrA/B pourrait répondre à des situations où l'homéostasie des ions métalliques, aussi bien de fer que de nickel, de cobalt, de cuivre, de zinc et de cadmium, représente une urgence importante pour la cellule, comme lors d'un stress oxydatif.

Annexe 3 : Phénotypes liés aux sRNA

A. Au moins une autre cible commune à OmrA/B et SdsR?

Au cours d'expériences faisant l'usage de souches *ΔomrA/B*, *ΔsdsR*, *ΔrseX* ou *Δ4*, j'ai pu observer que le temps de sédimentation de cultures à saturation semblait différent d'une souche à l'autre (Fig. 78). Plus particulièrement, les souches délétées des gènes *omrA/B* et *sdsR* ont un phénotype marqué et reproductible de sédimentation plus rapide. Ce phénotype est d'autant plus marqué lorsque ces mutations sont combinées dans un contexte *Δ4*, bien qu'il faille s'assurer qu'il n'y a pas un effet supplémentaire du à la délétion de *rseX*. Ce phénotype d'une simple délétion des sRNA est tout à fait étonnant, dans la mesure où la majorité des phénotypes associés à des sRNA Hfq-dépendants observés à ce jour chez des entérobactéries sont associés à la surexpression de ces ARN régulateurs. L'expression stochastique du gène *flu* codant l'antigène 43 peut provoquer un phénotype de sédimentation très rapide du fait de l'agrégation des bactéries (Diderichsen, 1980), aussi j'ai vérifié que ce phénotype était reproductible dans une souche *Δflu* (résultats non montrés ici).

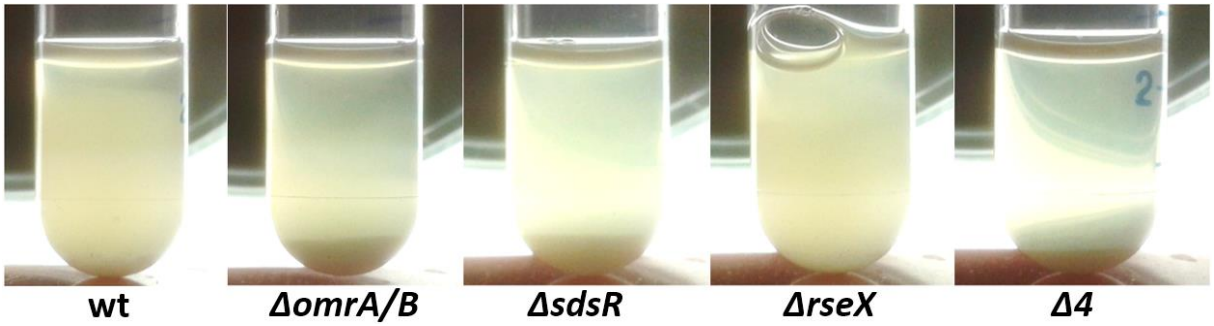


Figure 78 – Phénotype de sédimentation lié à omrA/B et sdsR
Des cultures à saturation de bactéries wt (souche MG1772), ΔomrA/B (souche JJ0142), ΔsdsR (souche JJ0143), ΔrseX (souche JJ0164) ou Δ4 (souche JJ0168) en LB à 37°C ont été agitées simultanément au vortex, puis laissées à reposer plusieurs heures à température ambiante afin de permettre la sédimentation des bactéries.

Nous avons néanmoins supposé que ce phénotype pouvait être lié à la production d'appendices de surface, ayant souvent un impact important sur la motilité des cellules (voir **Introduction III.A.3**). C'est pourquoi j'ai par la suite cherché à savoir s'il existait un phénotype de motilité lié à certaines de ces mutations, en analysant ces différents mutants sur des boîtes

d'agar souple permettant la nage des bactéries (Fig. 79). Nous avons par ailleurs choisi de réaliser cette expérience dans un contexte déléter *Δflu* afin d'éviter tout phénotype lié à l'antigène 43, mais aussi *ΔfepA* car nous suspections que ce gène, dont l'expression est régulée par OmrA/B, SdsR et RseX, pouvait intervenir dans le phénotype observé.

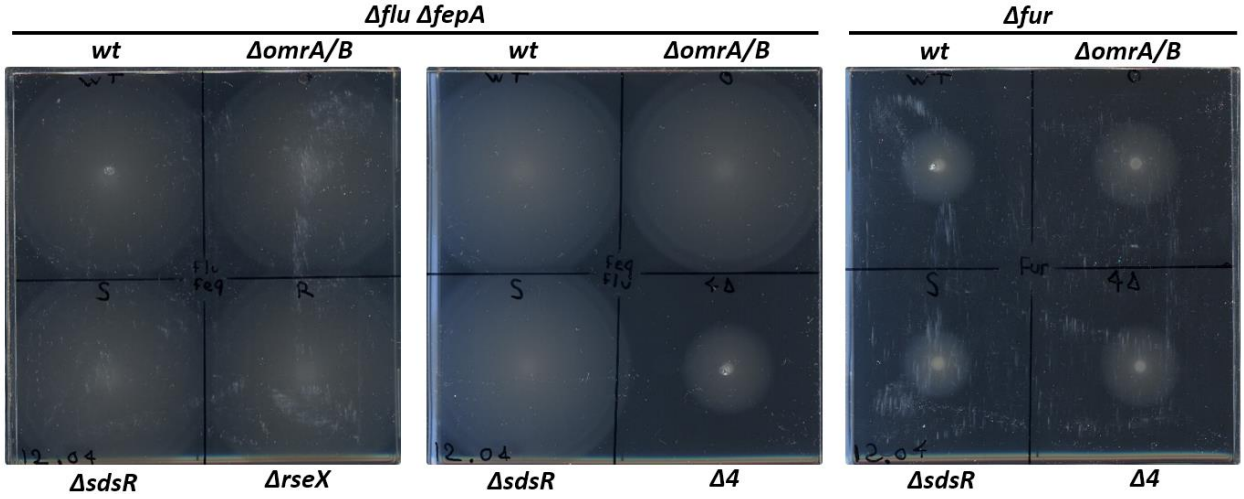


Figure 79 – Phénotype de motilité lié à OmrA/B, SdsR et RseX
Des bactéries Δflu ΔfepA (souche JJ0201) et leurs dérivées ΔomrA/B (souche JJ0202), ΔrseX (souche JJ0203), ΔsdsR (souche JJ0204) et Δ4 (souche JJ0205), ainsi que des bactéries Δfur (souche MG1169) et leurs dérivées ΔomrA/B (souche JJ0184), ΔsdsR (souche JJ0185) et Δ4 (souche JJ0188) ont été cultivées à saturation en LB à 37°C. Les cultures ont été normalisées à 2 unités de DO_{600}/mL, et des gouttes de 2µL ont été déposées sur LB – agar souple. Ces cellules ont poussé pendant 12h30 (Δflu ΔfepA) ou pendant 25h (Δfur) à 30°C avant analyse.

À nouveau un phénotype associé à la délétion des sRNA est observé, bien que cette fois-ci seule la combinaison de plusieurs délétions provoque ce phénotype. Etant donné l'expression à priori faible du sRNA RseX, ce phénotype est vraisemblablement à nouveau associé à la mutation combinée des seuls *omrA/B* et de *sdsR*, bien qu'il faille le confirmer ici aussi. Ce phénotype est indépendant de *flu*, mais aussi de *fepA*. Par ailleurs, ce phénotype n'est plus observable dans un contexte *Δfur*, ce qui peut s'expliquer par le fait que Fur semble être un répresseur de la synthèse de *flhDC* (Stojiljkovic et al., 1994). En effet, une expression trop importante de *flhDC* (ici du fait de l'absence de Fur) peut entraîner la synthèse d'une grande quantité de flagelles, aboutissant à des cellules hyperflagellées incapables de nager (Lemke et al., 2009). *flhDC* serait donc un bon candidat pour expliquer ce phénotype causé par la délétion d'*omrA/B* et de *sdsR* : en effet, sachant qu'OmrA/B sont des répresseurs de l'expression de *flhDC* et que SdsR réprime également l'expression de ce gène par un mécanisme qui reste à démontrer (De Lay and Gottesman, 2012), l'absence combinée de ces

trois sRNA pourrait provoquer une traduction relativement élevée de l'ARNm *flhDC* et aboutir à ce même phénotype de cellules hyperflagellées, à la motilité extrêmement réduite.

Il est ici intéressant de noter que la perte de motilité due à une surexpression d'OmrA/B a déjà été observée (De Lay and Gottesman, 2012) : la surexpression aussi bien que la perte de ces sRNA entraîne donc le même phénotype. Cela dénote l'importance des OmrA/B, dont il faut une quantité ni trop faible ni trop élevée pour permettre la motilité des cellules. Par ailleurs, le fait que la combinaison des délétions de plusieurs sRNA entraine un tel phénotype souligne le fait qu'il reste certainement beaucoup à apprendre sur l'impact physiologique des sRNA. En effet, la redondance des régulations par plusieurs sRNA a certainement masqué cet impact physiologique dans de nombreux autres cas. Il serait certainement possible d'en apprendre plus sur les rôles physiologiques des sRNA en généralisant ces combinaisons de mutations de sRNA, afin de rechercher d'autres phénotypes liés aux sRNA Hfq-dépendants.

B. OmrA/B et la taille des bactéries

Au cours de mon stage de M2, je me suis demandé si les OmrA/B pouvaient affecter la physiologie des cellules d'une façon qui pourrait s'observer par microscopie. J'ai donc observé des bactéries surexprimant OmrA, OmrB ou une version mutante OmrAmut ou OmrBmut (voir **Article**, mutations en Fig. 1) de ces sRNA au microscope optique (Fig. 80.A). J'ai ainsi pu m'apercevoir que lorsque les sRNA OmrA ou OmrB wt était surexprimés, les bactéries semblaient en moyenne plus petites que dans les contrôles sans sRNA ou avec des sRNA mutants. Cette impression est confirmée par la mesure de la taille des cellules et validée statistiquement (Fig. 80.B).

S'il est difficile pour l'instant d'attribuer ce phénotype à une cible des OmrA/B en particulier, il semble bien s'agir d'un effet lié à l'extrémité 5' des OmrA/B, et donc vraisemblablement à une régulation par les OmrA/B. Il serait donc intéressant de replonger dans les données de transcriptomique (Guillier and Gottesman, 2006) (voir **Annexe 2.B**) à la recherche d'une ou plusieurs cibles qui pourraient expliquer cet effet. Sans même aller jusque-là, une piste intéressante est celle de *bolA* : ce gène régulé négativement par OmpR-P a un impact négatif sur la taille des cellules (Freire et al., 2009). Ainsi, la répression d'*ompR* par les

OmrA/B pourrait provoquer une forte expression de *bolA*, entraînant ce phénotype sur la taille des bactéries. Néanmoins, le fait que *bolA* semble sensible uniquement à la forme phosphorylée d'OmpR et non aux niveaux de cette protéine n'est pas tout à fait compatible avec cette hypothèse, dans la mesure où la régulation par OmrA/B n'a de répercussion que sur les niveaux d'OmpR non phosphorylé (Brosse et al., 2016). Il serait toutefois intéressant de réitérer cette expérience dans des souches *ΔompR* et *ΔbolA* afin de vérifier si l'effet observé pourrait dépendre de l'un de ces deux gènes.

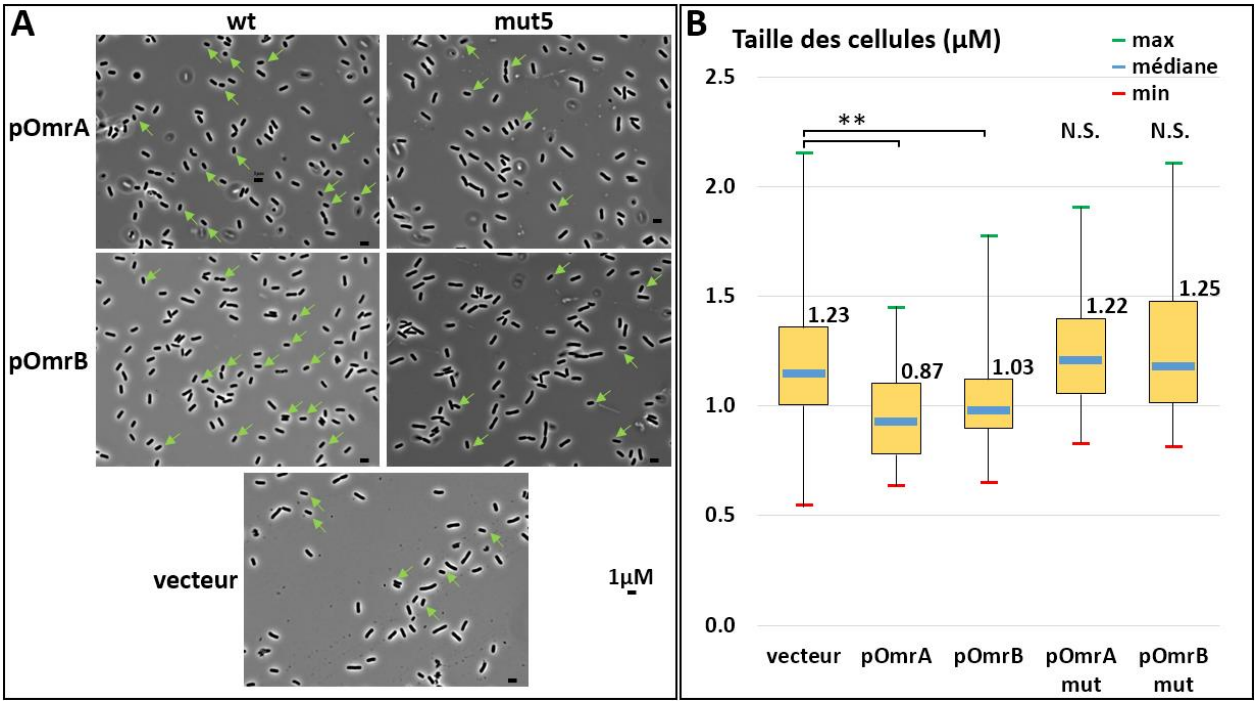


Figure 80 – OmrA/B ont un effet sur la taille d'Escherichia coli

*Des bactéries ΔomrA/B (JJ0006) transformées par un plasmide contrôle (vecteur) ou surexprimant OmrA, OmrB, OmrAmut ou OmrBmut, ont été cultivées à saturation en LB – amp – IPTG jusqu'à saturation, puis observées au microscope optique au grossissement x1000 (A). Des flèches vertes indiquent des exemples de bactéries particulièrement petites par rapport au reste de la population observée. Les tailles d'au moins 45 cellules individuelles ont été mesurées pour chaque sRNA surexprimé (B). Pour chaque population, la moyenne de taille est indiquée ; les diagrammes en boîte jaunes représentent les tailles des cellules incluses entre le premier et le troisième quartile; les tailles maximale, médiane et minimale mesurées sont respectivement désignées par un trait horizontal vert, bleu et rouge. ** : P-value < 0,001 selon un test de Student ; N.S : différence par rapport au vecteur non significative.*

Bibliographie

Acosta, N., Pukatzki, S., and Raivio, T.L. (2015). The *Vibrio cholerae* Cpx Envelope Stress Response Senses and Mediates Adaptation to Low Iron. J. Bacteriol. *197*, 262–276.

Allen, W.J., Corey, R.A., Oatley, P., Sessions, R.B., Baldwin, S.A., Radford, S.E., Tuma, R., and Collinson, I. (2016). Two-way communication between SecY and SecA suggests a Brownian ratchet mechanism for protein translocation. Elife *5*, e15598.

Antoun, A., Pavlov, M.Y., Andersson, K., Tenson, T., and Ehrenberg, M. ans (2003). The roles of initiation factor 2 and guanosine triphosphate in initiation of protein synthesis. EMBO J. *22*, 5593–5601.

Arbel-Goren, R., Tal, A., Parasar, B., Dym, A., Costantino, N., Muñoz-García, J., Court, D.L., and Stavans, J. (2016). Transcript degradation and noise of small RNA-controlled genes in a switch activated network in *Escherichia coli*. Nucleic Acids Res. *44*, 6707–6720.

Argaman, L., Hershberg, R., Vogel, J., Bejerano, G., Wagner, E.G.H., Margalit, H., and Altuvia, S. (2001). Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. Curr. Biol. *11*, 941–950.

Aussel, L., Zhao, W., Hébrard, M., Guilhon, A.-A., Viala, J.P.M., Henri, S., Chasson, L., Gorvel, J.-P., Barras, F., and Méresse, S. (2011). *Salmonella* detoxifying enzymes are sufficient to cope with the host oxidative burst: Oxidative stress adaptation by *Salmonella*. Mol. Microbiol. *80*, 628–640.

Baichoo, N., and Helmann, J.D. (2002). Recognition of DNA by Fur: a Reinterpretation of the Fur Box Consensus Sequence. J. Bacteriol. *184*, 5826–5832.

Bakelar, J., Buchanan, S.K., and Noinaj, N. (2016). The structure of the -barrel assembly machinery complex. Science *351*, 180–186.

Baker, C.S., Morozov, I., Suzuki, K., Romeo, T., and Babitzke, P. (2002). CsrA regulates glycogen biosynthesis by preventing translation of *glgC* in *Escherichia coli*. Mol. Microbiol. *44*, 1599–1610.

Balbontín, R., Villagra, N., Pardos de la Gándara, M., Mora, G., Figueroa-Bossi, N., and Bossi, L. (2016). Expression of IroN, the salmochelin siderophore receptor, requires mRNA activation by RyhB small RNA homologues: Activation of *iroN* mRNA by RyhB sRNA homologues. Mol. Microbiol. *100*, 139–155.

Banerjee, A.K., Rensing, U., and August, J.T. (1969). Replication of RNA viruses. J. Mol. Biol. *45*, 181–193.

Battesti, A., Majdalani, N., and Gottesman, S. (2011). The RpoS-Mediated General Stress Response in *Escherichia coli*. Annu. Rev. Microbiol. *65*, 189–213.

Beato, M. (1985). Modulation of gene expression through DNA-binding proteins: is there a regulatory code? Haematol. Blood Transfus. *29*, 217–223.

Berg, B.L., Baron, C., and Stewart, V. (1991). Nitrate-inducible formate dehydrogenase in *Escherichia coli* K-12. II. Evidence that an mRNA stem-loop structure is essential for decoding opal (UGA) as selenocysteine. J. Biol. Chem. *266*, 22386–22391.

Bezerra, A., Guimarães, A., and Santos, M. (2015). Non-Standard Genetic Codes Define New Concepts for Protein Engineering. Life *5*, 1610–1628.

Bird, J.G., Zhang, Y., Tian, Y., Panova, N., Barvík, I., Greene, L., Liu, M., Buckley, B., Krásný, L., Lee, J.K., Kaplan, C.D., Ebright, R.H., and Nickels, B.E. (2016). The mechanism of RNA 5'capping with NAD+, NADH and desphospho-CoA. Nature *535*, 444–447.

Blanco, L.P., Evans, M.L., Smith, D.R., Badtke, M.P., and Chapman, M.R. (2012). Diversity, biogenesis and function of microbial amyloids. Trends Microbiol. *20*, 66–73.

Blériot, C., Effantin, G., Lagarde, F., Mandrand-Berthelot, M.-A., and Rodrigue, A. (2011). RcnB Is a Periplasmic Protein Essential for Maintaining Intracellular Ni and Co Concentrations in *Escherichia coli*. J. Bacteriol. *193*, 3785–3793.

Blériot, C., Gault, M., Gueguen, E., Arnoux, P., Pignol, D., Mandrand-Berthelot, M.-A., and Rodrigue, A. (2014). Cu binding by the *Escherichia coli* metal-efflux accessory protein RcnB. Metallomics *6*, 1400.

Bleuel, C., Grosse, C., Taudte, N., Scherer, J., Wesenberg, D., Krauss, G.J., Nies, D.H., and Grass, G. (2005). TolC Is Involved in Enterobactin Efflux across the Outer Membrane of *Escherichia coli*. J. Bacteriol. *187*, 6701–6707.

Boël, G., Letso, R., Neely, H., Price, W.N., Wong, K.-H., Su, M., Luff, J.D., Valecha, M., Everett, J.K., Acton, T.B., Xiao, R., Montelione, G.T., Aalberts, D.P., and Hunt, J.F. (2016). Codon influence on protein expression in *E. coli* correlates with mRNA levels. Nature *529*, 358–363.

Boni, I.V., Musychenko, M.L., Tzareva, N.V., and others (1991). Ribosome-messenger recognition: mRNA target sites for ribosomal protein S1. Nucleic Acids Res. *19*, 155–162.

Boni, I.V., Artamonova, V.S., Tzareva, N.V., and Dreyfus, M. (2001). Non-canonical mechanism for translational control in bacteria: synthesis of ribosomal protein S1. EMBO J. *20*, 4222–4232.

Bossi, L., and Figueroa-Bossi, N. (2016). Competing endogenous RNAs: a target-centric view of small RNA regulation in bacteria. Nat. Rev. Microbiol. *14*, 775–784.

Bossi, L., and Roth, J.R. (1980). The influence of codon context on genetic code translation. Nature *286*, 123–127.

Boudvillain, M., Figueroa-Bossi, N., and Bossi, L. (2013). Terminator still moving forward: expanding roles for Rho factor. Curr. Opin. Microbiol. *16*, 118–124.

Brantl, S., Birch-Hirschfeld, E., and Behnke, D. (1993). RepR protein expression on plasmid pIP501 is controlled by an antisense RNA-mediated transcription attenuation mechanism. J. Bacteriol. *175*, 4052–4061.

Braun, V., and Killmann, H. (1999). Bacterial solutions to the iron-supply problem. Trends Biochem. Sci. *24*, 104–109.

Brosse, A., Korobeinikova, A., Gottesman, S., and Guillier, M. (2016). Unexpected properties of sRNA promoters allow feedback control via regulation of a two-component system. Nucleic Acids Res. gkw642.

Browning, D.F., and Busby, S.J.W. (2016). Local and global regulation of transcription initiation in bacteria. Nat. Rev. Microbiol. *14*, 638–650.

Bruenn, J.A. (2003). A structural and primary sequence comparison of the viral RNA-dependent RNA polymerases. Nucleic Acids Res. *31*, 1821–1829.

Buchanan, S.K., Smith, B.S., Venkatramani, L., Xia, D., Esser, L., Palnitkar, M., Chakraborty, R., van der Helm, D., and Deisenhofer, J. (1999). Crystal structure of the outer membrane active transporter FepA from *Escherichia coli*. Nat. Struct. Biol. *6*, 56–63.

Burkhardt, D.H., Rouskin, S., Zhang, Y., Li, G.-W., Weissman, J.S., and Gross, C.A. (2017). Operon mRNAs are organized into ORF-centric structures that predict translation efficiency. Elife *6*.

Busch, A., Richter, A.S., and Backofen, R. (2008). IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions. Bioinformatics *24*, 2849–2856.

Butler, J.S., Springer, M., Dondon, J., Graffe, M., and Grunberg-Manago, M. (1986). *Escherichia coli* protein synthesis initiation factor IF3 controls its own gene expression at the translational level in vivo. J. Mol. Biol. *192*, 767–780.

Cahová, H., Winz, M.-L., Höfer, K., Nübel, G., and Jäschke, A. (2014). NAD captureSeq indicates NAD as a bacterial cap for a subset of regulatory RNAs. Nature *519*, 374–377.

Callen, B.P., Shearwin, K.E., and Egan, J.B. (2004). Transcriptional interference between convergent promoters caused by elongation over the promoter. Mol. Cell *14*, 647–656.

Carrano, C.J., and Raymond, K.N. (1979). Ferric ion sequestering agents. 2. Kinetics and mechanism of iron removal from transferrin by enterobactin and synthetic tricatechols. J. Am. Chem. Soc. *101*, 5401–5404.

Caskey, C.T., Tompkins, R., Scolnick, E., Caryk, T., and Nirenberg, M. (1968). Sequential Translation of Trinucleotide Codons for the Initiation and Termination of Protein Synthesis. Science *162*, 135–138.

Caskey, C.T., Bosch, L., and Konecki, D.S. (1977). Release factor binding to ribosome requires an intact 16 S rRNA 3'terminus. J. Biol. Chem. *252*, 4435–4437.

Cassidy, L.M., Dong, Y., Joshi, P.C., Aldersley, M.F., Ferris, J.P., and McGown, L.B. (2011). Signal Enhancement of Abiotically-Synthesized RNA Oligonucleotides and other Biopolymers using Unmodified Fused Silica in MALDI-MS. J. Am. Soc. Mass Spectrom. *22*, 1100–1104.

Cavanagh, A.T., Klocko, A.D., Liu, X., and Wassarman, K.M. (2008). Promoter specificity for 6S RNA regulation of transcription is determined by core promoter sequences and competition for region 4.2 of σ^{70}. Mol. Microbiol. *67*, 1242–1256.

Chaba, R., Grigorova, I.L., Flynn, J.M., Baker, T.A., and Gross, C.A. (2007). Design principles of the proteolytic cascade governing the σE-mediated envelope stress response in *Escherichia coli*: keys to graded, buffered, and rapid signal transduction. Genes Dev. *21*, 124–136.

Chalut, C., and Egly, J.M. (1995). AUC is used as a start codon in *Escherichia coli*. Gene *156*, 43–45.

Chao, Y., Papenfort, K., Reinhardt, R., Sharma, C.M., and Vogel, J. (2012). An atlas of Hfq-bound transcripts reveals 3′ UTRs as a genomic reservoir of regulatory small RNAs. EMBO J. *31*, 4005–4019.

Chen, H., Bjerknes, M., Kumar, R., and Jay, E. (1994). Determination of the optimal aligned spacing between the Shine–Dalgarno sequence and the translation initiation codon of *Escherichia coli* m RNAs. Nucleic Acids Res. *22*, 4953–4957.

Chevance, F.F.V., and Hughes, K.T. (2008). Coordinating assembly of a bacterial macromolecular machine. Nat. Rev. Microbiol. *6*, 455–465.

Chng, S.-S., Ruiz, N., Chimalakonda, G., Silhavy, T.J., and Kahne, D. (2010). Characterization of the two-protein complex in *Escherichia coli* responsible for lipopolysaccharide assembly at the outer membrane. Proc. Natl. Acad. Sci. *107*, 5363–5368.

Clemons, W.M., May, J.L., Wimberly, B.T., McCutcheon, J.P., Capel, M.S., and Ramakrishnan, V. (1999). Structure of a bacterial 30S ribosomal subunit at 5.5 Ä resolution. Nature *400*, 833–840.

Coornaert, A., Chiaruttini, C., Springer, M., and Guillier, M. (2013). Post-Transcriptional Control of the *Escherichia coli* PhoQ-PhoP Two-Component System by Multiple sRNAs Involves a Novel Pairing Region of GcvB. PLoS Genet. *9*, e1003156.

Cornelis, P., Wei, Q., Andrews, S.C., and Vinckx, T. (2011). Iron homeostasis and management of oxidative stress response in bacteria. Metallomics *3*, 540.

Court, D.L., Gan, J., Liang, Y.-H., Shaw, G.X., Tropea, J.E., Costantino, N., Waugh, D.S., and Ji, X. (2013). RNase III: Genetics and Function; Structure and Mechanism. Annu. Rev. Genet. *47*, 405–431.

Crick, F.H.C., Barnett, L., Brenner, S., and Watts-Tobin, R.J. (1961). General Nature of the Genetic Code for Proteins. Nature *192*, 1227–1232.

Crooks, G.E., Hon, G., Chandonia, J.-M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res. *14*, 1188–1190.

Darfeuille, F., Unoson, C., Vogel, J., and Wagner, E.G.H. (2007). An Antisense RNA Inhibits Translation by Competing with Standby Ribosomes. Mol. Cell *26*, 381–392.

Dartigalongue, C., Missiakas, D., and Raina, S. (2001). Characterization of the *Escherichia coli* E Regulon. J. Biol. Chem. *276*, 20866–20875.

De Lay, N., and Gottesman, S. (2012). A complex network of small non-coding RNAs regulate motility in *Escherichia coli* Mol. Microbiol. *86*, 524–538.

Deck, C., Jauker, M., and Richert, C. (2011). Efficient enzyme-free copying of all four nucleobases templated by immobilized RNA. Nat. Chem. *3*, 603–608.

Del Campo, C., Bartholomäus, A., Fedyunin, I., and Ignatova, Z. (2015). Secondary Structure across the Bacterial Transcriptome Reveals Versatile Roles in mRNA Regulation and Function. PLOS Genet. *11*, e1005613.

Delihas, N., and Forst, S. (2001). MicF : an antisense RNA gene involved in response of *Escherichia coli* to global stress factors 1 1Edited by D. Draper. J. Mol. Biol. *313*, 1–12.

Deng, T., Vreede, F.T., and Brownlee, G.G. (2006). Different De Novo Initiation Strategies Are Used by Influenza Virus RNA Polymerase on Its cRNA and Viral RNA Promoters during Viral RNA Replication. J. Virol. *80*, 2337–2348.

Diderichsen, B. (1980). *flu*, a metastable gene controlling surface properties of *Escherichia coli*. J. Bacteriol. *141*, 858–867.

Douchin, V., Bohn, C., and Bouloc, P. (2006). Down-regulation of Porins by a Small RNA Bypasses the Essentiality of the Regulated Intramembrane Proteolysis Protease RseP in *Escherichia coli*. J. Biol. Chem. *281*, 12253–12259.

Dresch, J.M., and Arnosti, D.N. (2015). The Wisdom of Crowds: Can Mathematical Models Crack the *cis* Regulatory Code? Cell Syst. *1*, 379–380.

Dubey, A.K., Baker, C.S., Suzuki, K., Jones, A.D., Pandit, P., Romeo, T., and Babitzke, P. (2003). CsrA Regulates Translation of the *Escherichia coli* Carbon Starvation Gene, *cstA*, by Blocking Ribosome Access to the *cstA* Transcript. J. Bacteriol. *185*, 4450–4460.

Duche, D., Issouf, M., and Lloubes, R. (2008). Immunity Protein Protects Colicin E2 from OmpT Protease. J. Biochem. (Tokyo) *145*, 95–101.

Duval, M., Korepanov, A., Fuchsbauer, O., Fechter, P., Haller, A., Fabbretti, A., Choulier, L., Micura, R., Klaholz, B.P., Romby, P., Springer, M., and Marzi, S. (2013). *Escherichia coli* Ribosomal Protein S1 Unfolds Structured mRNAs Onto the Ribosome for Active Translation Initiation. PLoS Biol. *11*, e1001731.

Ero, R., Kumar, V., Chen, Y., and Gao, Y.-G. (2016). Similarity and diversity of translational GTPase factors EF-G, EF4, and BipA: From structure to function. RNA Biol. *13*, 1258–1273.

Espah Borujeni, A., Channarasappa, A.S., and Salis, H.M. (2014). Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. Nucleic Acids Res. *42*, 2646–2659.

Fabbretti, A., Pon, C.L., Hennelly, S.P., Hill, W.E., Lodmell, J.S., and Gualerzi, C.O. (2007). The Real-Time Path of Translation Factor IF3 onto and off the Ribosome. Mol. Cell *25*, 285–296.

Fedor, M.J., and Williamson, J.R. (2005). The catalytic diversity of RNAs. Nat. Rev. Mol. Cell Biol. *6*, 399–412.

Figueroa-Bossi, N., Valentini, M., Malleret, L., and Bossi, L. (2009). Caught at its own game: regulatory small RNA inactivated by an inducible transcript mimicking its target. Genes Dev. *23*, 2004–2015.

Fluman, N., Navon, S., Bibi, E., and Pilpel, Y. (2014). mRNA-programmed translation pauses in the targeting of *E. coli* membrane proteins. Elife *3*, e03440.

Fontecave, M., Choudens, S.O. de, Py, B., and Barras, F. (2005). Mechanisms of iron-sulfur cluster assembly: the SUF machinery. JBIC J. Biol. Inorg. Chem. *10*, 713–721.

Freire, P., Neves Moreira, R., and Arraiano, C.M. (2009). BolA Inhibits Cell Elongation and Regulates MreB Expression Levels. J. Mol. Biol. *385*, 1345–1351.

Freistroffer, D.V., Pavlov, M.Y., MacDougall, J., Buckingham, R.H., and Ehrenberg, M. (1997). Release factor RF3 in *E. coli* accelerates the dissociation of release factors RF1 and RF2 from the ribosome in a GTP-dependent manner. EMBO J. *16*, 4126–4133.

Friedrich, L.C., Mendes, M.A., Silva, V.O., Zanta, C.L.P.S., Machulek Jr., A., and Quina, F.H. (2012). Mechanistic implications of zinc(II) ions on the degradation of phenol by the fenton reaction. J. Braz. Chem. Soc. *23*, 1372–1377.

Friedrich, T., Dekovic, D.K., and Burschel, S. (2016). Assembly of the *Escherichia coli* NADH:ubiquinone oxidoreductase (respiratory complex I). Biochim. Biophys. Acta BBA - Bioenerg. *1857*, 214–223.

Fröhlich, K.S., Papenfort, K., Berger, A.A., and Vogel, J. (2012). A conserved RpoS-dependent small RNA controls the synthesis of major porin OmpD. Nucleic Acids Res. *40*, 3623–3640.

Fröhlich, K.S., Haneke, K., Papenfort, K., and Vogel, J. (2016). The target spectrum of SdsR small RNA in *Salmonella*. Nucleic Acids Res. gkw632.

Gerdes, K., Nielsen, A., Thorsted, P., and Wagner, E.G. (1992). Mechanism of killer gene activation. Antisense RNA-dependent RNase III cleavage ensures rapid turn-over of the stable *hok*, *srnB* and *pndA* effector messenger RNAs. J. Mol. Biol. *226*, 637–649.

Giegé, R. (2008). Toward a more complete view of tRNA biology. Nat. Struct. Mol. Biol. *15*, 1007–1014.

Giegé, R., Sissler, M., and Florentz, C. (1998). Universal rules and idiosyncratic features in tRNA identity. Nucleic Acids Res. *26*, 5017–5035.

Giessen, T.W., and Marahiel, M.A. (2012). Ribosome-independent biosynthesis of biologically active peptides: Application of synthetic biology to generate structural diversity. FEBS Lett. *586*, 2065–2075.

Gilbert, W. (1986). Origin of life: The RNA world. Nature *319*, 618–618.

Gogol, E.B., Rhodius, V.A., Papenfort, K., Vogel, J., and Gross, C.A. (2011). Small RNAs endow a transcriptional activator with essential repressor functions for single-tier control of a global stress regulon. Proc. Natl. Acad. Sci. *108*, 12875–12880.

Goh, E.-B., Siino, D.F., and Igo, M.M. (2004). The *Escherichia coli* tppB (ydgR) Gene Represents a New Class of OmpR-Regulated Genes. J. Bacteriol. *186*, 4019–4024.

Gold, L. (1988). Posttranscriptional regulatory mechanisms in *Escherichia coli*. Annu. Rev. Biochem. *57*, 199–233.

Gold, L., Polisky, B., Uhlenbeck, O., and Yarus, M. (1995). Diversity of oligonucleotide functions. Annu. Rev. Biochem. *64*, 763–797.

Goldblum, K., and Apririon, D. (1981). Inactivation of the ribonucleic acid-processing enzyme ribonuclease E blocks cell division. J. Bacteriol. *146*, 128–132.

Grabowicz, M., and Silhavy, T.J. (2017). Envelope Stress Responses: An Interconnected Safety Net. Trends Biochem. Sci. *42*, 232–242.

Graham, A.I., Hunt, S., Stokes, S.L., Bramall, N., Bunch, J., Cox, A.G., McLeod, C.W., and Poole, R.K. (2009). Severe Zinc Depletion of *Escherichia coli*: ROLES FOR HIGH AFFINITY ZINC BINDING BY ZinT, ZINC TRANSPORT AND ZINC-INDEPENDENT PROTEINS. J. Biol. Chem. *284*, 18377–18389.

Greenwood, K.T., and Luke, R.K. (1978). Enzymatic hydrolysis of enterochelin and its iron complex in *Escherichia coli* K-12. Properties of enterochelin esterase. Biochim. Biophys. Acta BBA-Enzymol. *525*, 209–218.

Greive, S.J., and von Hippel, P.H. (2005). Thinking quantitatively about transcriptional regulation. Nat. Rev. Mol. Cell Biol. *6*, 221–232.

Griffith, K.L., Fitzpatrick, M.M., Keen, E.F., and Wolf, R.E. (2009). Two Functions of the C-Terminal Domain of *Escherichia coli* Rob: Mediating "Sequestration–Dispersal" as a Novel Off–On Switch for Regulating Rob's Activity as a Transcription Activator and Preventing Degradation of Rob by Lon Protease. J. Mol. Biol. *388*, 415–430.

Grigoriadou, C., Marzi, S., Kirillov, S., Gualerzi, C.O., and Cooperman, B.S. (2007). A quantitative kinetic scheme for 70 S translation initiation complex formation. J. Mol. Biol. *373*, 562–572.

Gualerzi, C.O., and Pon, C.L. (2015). Initiation of mRNA translation in bacteria: structural and dynamic aspects. Cell. Mol. Life Sci. *72*, 4341–4367.

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

Guillier, M., and Gottesman, S. (2006). Remodelling of the *Escherichia coli* outer membrane by two small regulatory RNAs. Mol. Microbiol. *59*, 231–247.

Guillier, M., and Gottesman, S. (2008). The 5' end of two redundant sRNAs is involved in the regulation of multiple targets, including their own regulator. Nucleic Acids Res. *36*, 6781–6794.

Guo, M., and Schimmel, P. (2012). Structural analyses clarify the complex control of mistranslation by tRNA synthetases. Curr. Opin. Struct. Biol. *22*, 119–126.

Guo, M.S., Updegrove, T.B., Gogol, E.B., Shabalina, S.A., Gross, C.A., and Storz, G. (2014). MicL, a new E-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. Genes Dev. *28*, 1620–1634.

Gutierrez, A., Laureti, L., Crussard, S., Abida, H., Rodríguez-Rojas, A., Blázquez, J., Baharoglu, Z., Mazel, D., Darfeuille, F., Vogel, J., and Matic, I. (2013). β-lactam antibiotics promote bacterial mutagenesis via an RpoS-mediated reduction in replication fidelity. Nat. Commun. *4*, 1610.

Hagiwara, D., Yamashino, T., and Mizuno, T. (2004). A Genome-wide view of the *Escherichia coli* BasS-BasR two-component system implicated in iron-responses. Biosci. Biotechnol. Biochem. *68*, 1758–1767.

Haiko, J., and Westerlund-Wikström, B. (2013). The Role of the Bacterial Flagellum in Adhesion and Virulence. Biology *2*, 1242–1267.

Hall, M.N., and Silhavy, T.J. (1981). The *ompB* locus and the regulation of the major outer membrane porin proteins of *Escherichia coli* K12. J. Mol. Biol. *146*, 23–43.

Harms, J., Schluenzen, F., Zarivach, R., Bashan, A., Gat, S., Agmon, I., Bartels, H., Franceschi, F., and Yonath, A. (2001). High Resolution Structure of the Large Ribosomal Subunit from a Mesophilic Eubacterium. Cell *107*, 679–688.

Hartz, D., McPheeters, D.S., Traut, R., and Gold, L. (1988). Extension inhibition analysis of translation initiation complexes. Methods Enzymol. *164*, 419–425.

Hebrard, M., Viala, J.P.M., Meresse, S., Barras, F., and Aussel, L. (2009). Redundant Hydrogen Peroxide Scavengers Contribute to *Salmonella* Virulence and Oxidative Stress Resistance. J. Bacteriol. *191*, 4605–4614.

Heider, J., Baron, C., and Böck, A. (1992). Coding from a distance: dissection of the mRNA determinants required for the incorporation of selenocysteine into protein. EMBO J. *11*, 3759.

Hejair, H.M.A., Zhu, Y., Ma, J., Zhang, Y., Pan, Z., Zhang, W., and Yao, H. (2017). Functional role of OmpF and OmpC porins in pathogenesis of avian pathogenic *Escherichia coli*. Microb. Pathog. *107*, 29–37.

Hengge, R. (2011). Stationary-Phase Gene Regulation in *Escherichia coli*. EcoSal Plus *4*.

Hinnebusch, A.G. (2011). Molecular Mechanism of Scanning and Start Codon Selection in Eukaryotes. Microbiol. Mol. Biol. Rev. *75*, 434–467.

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

Hollifield, W.C.J., and Neilands, J.B. (1978). Ferric enterobactin transport system in *Escherichia coli* K-12. Extraction, assay, and specificity of the outer membrane receptor. Biochemistry (Mosc.) *17*, 1922–1928.

Holmqvist, E., Reimegård, J., Sterk, M., Grantcharova, N., Römling, U., and Wagner, E.G.H. (2010). Two antisense RNAs target the transcriptional regulator CsgD to inhibit curli synthesis. EMBO J. *29*, 1840–1850.

Hook-Barnard, I.G., Brickman, T.J., and McIntosh, M.A. (2007). Identification of an AU-rich Translational Enhancer within the *Escherichia coli fepB* Leader RNA. J. Bacteriol. *189*, 4028–4037.

Horne, J.E., and Radford, S.E. (2016). A growing toolbox of techniques for studying -barrel outer membrane protein folding and biogenesis. Biochem. Soc. Trans. *44*, 802–809.

Huang, K.C., Mukhopadhyay, R., Wen, B., Gitai, Z., and Wingreen, N.S. (2008). Cell shape and cell-wall organization in Gram-negative bacteria. Proc. Natl. Acad. Sci. *105*, 19282–19287.

Hunt, M.D., Pettis, G.S., and McIntosh, M.A. (1994). Promoter and operator determinants for fur-mediated iron regulation in the bidirectional *fepA-fes* control region of the *Escherichia coli* enterobactin gene system. J. Bacteriol. *176*, 3944–3955.

Hussain, T., Llácer, J.L., Wimberly, B.T., Kieft, J.S., and Ramakrishnan, V. (2016). Large-Scale Movements of IF3 and tRNA during Bacterial Translation Initiation. Cell *167*, 133–144.e13.

Hussein, R., and Lim, H.N. (2011). Disruption of small RNA signaling caused by competition for Hfq. Proc. Natl. Acad. Sci. *108*, 1110–1115.

Ichihara, S., and Mizushima, S. (1978). Identification of an outer membrane protein responsible for the binding of the Fe-enterochelin complex to *Escherichia coli* cells. J. Biochem. (Tokyo) *83*, 137–140.

Ignatova, Z., and Narberhaus, F. (2017). Systematic probing of the bacterial RNA structurome to reveal new functions. Curr. Opin. Microbiol. *36*, 14–19.

Igo, M.M., Ninfa, A.J., and Silhavy, T.J. (1989). A bacterial environmental sensor that functions as a protein kinase and stimulates transcriptional activation. Genes Dev. *3*, 598–605.

Ikemura, T. (1981). Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: A proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. J. Mol. Biol. *151*, 389–409.

Imlay, J.A. (2006). Iron-sulphur clusters and the problem with oxygen. Mol. Microbiol. *59*, 1073–1082.

Imlay, J.A., and Linn, S. (1988). DNA damage and oxygen radical toxicity. Science *240*, 1302.

Ishitani, R., Nureki, O., Nameki, N., Okada, N., Nishimura, S., and Yokoyama, S. (2003). Alternative tertiary structure of tRNA for recognition by a posttranscriptional modification enzyme. Cell *113*, 383–394.

Jacob, F., Ullman, A., and Monod, J. (1964). [THE PROMOTOR, A GENETIC ELEMENT NECESSARY TO THE EXPRESSION OF AN OPERON]. Comptes Rendus Hebd. Seances Acad. Sci. *258*, 3125–3128.

Jacob, W.F., Santer, M., and Dahlberg, A.E. (1987). A single base change in the Shine-Dalgarno region of 16S rRNA of *Escherichia coli* affects translation of many proteins. Proc. Natl. Acad. Sci. *84*, 4757–4761.

Jacques, J.-F., Jang, S., Prévost, K., Desnoyers, G., Desmarais, M., Imlay, J., and Massé, E. (2006). RyhB small RNA modulates the free intracellular iron pool and is essential for normal growth during iron limitation in *Escherichia coli*. Mol. Microbiol. *62*, 1181–1190.

Johansen, J., Rasmussen, A.A., Overgaard, M., and Valentin-Hansen, P. (2006). Conserved Small Non-coding RNAs that belong to the σE Regulon: Role in Down-regulation of Outer Membrane Proteins. J. Mol. Biol. *364*, 1–8.

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

Johnson, A.P., Cleaves, H.J., Dworkin, J.P., Glavin, D.P., Lazcano, A., and Bada, J.L. (2008). The Miller Volcanic Spark Discharge Experiment. Science *322*, 404–404.

Jorgensen, M.G., Thomason, M.K., Havelund, J., Valentin-Hansen, P., and Storz, G. (2013). Dual function of the McaS small RNA in controlling biofilm formation. Genes Dev. *27*, 1132–1145.

Karimi, R., Pavlov, M.Y., Buckingham, R.H., and Ehrenberg, M. (1999). Novel roles for classical factors at the interface between translation termination and initiation. Mol. Cell *3*, 601–609.

Kavaliauskas, D., Nissen, P., and Knudsen, C.R. (2012). The Busiest of All Ribosomal Assistants: Elongation Factor Tu. Biochemistry (Mosc.) *51*, 2642–2651.

Kawano, M., Aravind, L., and Storz, G. (2007). An antisense RNA controls synthesis of an SOS-induced toxin evolved from an antitoxin. Mol. Microbiol. *64*, 738–754.

Kelley, D.R., Snoek, J., and Rinn, J.L. (2016). Basset: learning the regulatory code of the accessible genome with deep convolutional neural networks. Genome Res. *26*, 990–999.

Keyer, K., and Imlay, J.A. (1996). Superoxide accelerates DNA damage by elevating free-iron levels. Proc. Natl. Acad. Sci. *93*, 13635–13640.

Kim, H.-K., and Tinoco, I. (2016). EF-G catalyzed translocation dynamics in the presence of ribosomal frameshifting stimulatory signals. Nucleic Acids Res. gkw1020.

Kim, S., Malinverni, J.C., Sliz, P., Silhavy, T.J., Harrison, S.C., and Kahne, D. (2007). Structure and function of an essential component of the outer membrane protein assembly machine. Science *317*, 961–964.

Kisselev, L.L., and Buckingham, R.H. (2000). Translational termination comes of age. Trends Biochem. Sci. *25*, 561–566.

Klein, G., Stupak, A., Biernacka, D., Wojtkiewicz, P., Lindner, B., and Raina, S. (2016). Multiple Transcriptional Factors Regulate Transcription of the *rpoE* Gene in *Escherichia coli* under Different Growth Conditions and When the Lipopolysaccharide Biosynthesis Is Defective. J. Biol. Chem. *291*, 22999–23019.

Klemm, B., Wu, N., Chen, Y., Liu, X., Kaitany, K., Howard, M., and Fierke, C. (2016). The Diversity of Ribonuclease P: Protein and RNA Catalysts with Analogous Biological Functions. Biomolecules *6*, 27.

Knight, R.D., and Landweber, L.F. (1998). Rhyme or reason: RNA-arginine interactions and the genetic code. Chem. Biol. *5*, R215–R220.

Kohanski, M.A., and Collins, J.J. (2008). Rewiring bacteria, two components at a time. Cell *133*, 947–948.

Komarova, A.V., Tchufistova, L.S., Supina, E.V., and Boni, I.V. (2002). Protein S1 counteracts the inhibitory effect of the extended Shine-Dalgarno sequence on translation. RNA N. Y. N *8*, 1137–1147.

Konigsberg, W., and Godson, G.N. (1983). Evidence for use of rare codons in the *dnaG* gene and other regulatory genes of *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. *80*, 687–691.

Kozak, M. (1983). Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. Microbiol. Rev. *47*, 1.

Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E., and Cech, T.R. (1982). Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. Cell *31*, 147–157.

Kukkonen, M., and Korhonen, T.K. (2004). The omptin family of enterobacterial surface proteases/adhesins: from housekeeping in *Escherichia coli* to systemic spread of *Yersinia pestis*. Int. J. Med. Microbiol. IJMM *294*, 7–14.

Kurland, C.G. (2010). The RNA dreamtime: Modern cells feature proteins that might have supported a prebiotic polypeptide world but nothing indicates that RNA world ever was. BioEssays *32*, 866–871.

Kwon, O., Hudspeth, M.E., and Meganathan, R. (1996). Anaerobic biosynthesis of enterobactin *Escherichia coli*: regulation of *entC* gene expression and evidence against its involvement in menaquinone (vitamin K2) biosynthesis. J. Bacteriol. *178*, 3252–3259.

La Teana, A., Gualerzi, C.O., and Brimacombe, R. (1995). From stand-by to decoding site. Adjustment of the mRNA on the 30S ribosomal subunit under the influence of the initiation factors. RNA N. Y. N *1*, 772–782.

Lalaouna, D., Carrier, M.-C., Semsey, S., Brouard, J.-S., Wang, J., Wade, J.T., and Massé, E. (2015). A 3' External Transcribed Spacer in a tRNA Transcript Acts as a Sponge for Small RNAs to Prevent Transcriptional Noise. Mol. Cell *58*, 393–405.

Lalaouna, D., Prévost, K., Eyraud, A., and Massé, E. (2016). Identification of unknown RNA partners using MAPS. Methods San Diego Calif.

Lau, C.K.Y., Krewulak, K.D., and Vogel, H.J. (2016). Bacterial ferrous iron transport: the Feo system. FEMS Microbiol. Rev. *40*, 273–298.

Lee, E.-J., and Groisman, E.A. (2010). An antisense RNA that governs the expression kinetics of a multifunctional virulence gene: A regulator controlling sense and antisense RNAs. Mol. Microbiol. *76*, 1020–1033.

Lee, J.-W., and Helmann, J.D. (2007). Functional specialization within the Fur family of metalloregulators. BioMetals *20*, 485–499.

Lee, E.R., Baker, J.L., Weinberg, Z., Sudarsan, N., and Breaker, R.R. (2010). An allosteric self-splicing ribozyme triggered by a bacterial second messenger. Science *329*, 845–848.

Lehman, N., and Hayden, E.J. (2011). Template-Directed RNA Polymerization: The Taming of the Milieu. ChemBioChem *12*, 2727–2728.

Lemke, J.J., Durfee, T., and Gourse, R.L. (2009). DksA and ppGpp directly regulate transcription of the *Escherichia coli* flagellar cascade: Regulation of the flagellar cascade by DksA/ppGpp. Mol. Microbiol. *74*, 1368–1379.

Lévi-Meyrueis, C., Monteil, V., Sismeiro, O., Dillies, M.-A., Monot, M., Jagla, B., Coppée, J.-Y., Dupuy, B., and Norel, F. (2014). Expanding the RpoS/σS-Network by RNA Sequencing and Identification of σS-Controlled Small RNAs in *Salmonella*. PLoS ONE *9*, e96918.

Lewis, D., Le, P., Zurla, C., Finzi, L., and Adhya, S. (2011). Multilevel autoregulation of λ repressor protein CI by DNA looping in vitro. Proc. Natl. Acad. Sci. *108*, 14807–14812.

Li, G.-W., Oh, E., and Weissman, J.S. (2012). The anti-Shine–Dalgarno sequence drives translational pausing and codon choice in bacteria. Nature *484*, 538–541.

Lill, R. (2009). Function and biogenesis of iron–sulphur proteins. Nature *460*, 831–838.

Lima, S., Guo, M.S., Chaba, R., Gross, C.A., and Sauer, R.T. (2013). Dual Molecular Signals Mediate the Bacterial Response to Outer-Membrane Stress. Science *340*, 837–841.

Liu, M.Y., Gui, G., Wei, B., Preston, J.F., Oakford, L., Yüksel, Ü., Giedroc, D.P., and Romeo, T. (1997). The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in *Escherichia coli*. J. Biol. Chem. *272*, 17502–17510.

Lloyd, D.R., and Phillips, D.H. (1999). Oxidative DNA damage mediated by copper(II), iron(II) and nickel(II) fenton reactions: evidence for site-specific mechanisms in the formation of double-strand breaks, 8-hydroxydeoxyguanosine and putative intrastrand cross-links. Mutat. Res. *424*, 23–36.

Ma, L., Kaserer, W., Annamalai, R., Scott, D.C., Jin, B., Jiang, X., Xiao, Q., Maymani, H., Massis, L.M., Ferreira, L.C.S., Newton, S.M.C., and Klebba, P.E. (2007). Evidence of Ball-and-chain Transport of Ferric Enterobactin through FepA. J. Biol. Chem. *282*, 397–406.

Mackie, G.A. (2012). RNase E: at the interface of bacterial RNA processing and decay. Nat. Rev. Microbiol. *11*, 45–57.

Makeyev, E.V., and Grimes, J.M. (2004). RNA-dependent RNA polymerases of dsRNA bacteriophages. Virus Res. *101*, 45–55.

Maldonado, R.F., Sá-Correia, I., and Valvano, M.A. (2016). Lipopolysaccharide modification in Gram-negative bacteria during chronic infection. FEMS Microbiol. Rev. *40*, 480–493.

Mandin, P., and Gottesman, S. (2010). Integrating anaerobic/aerobic sensing and the general stress response through the ArcZ small RNA. EMBO J. *29*, 3094–3107.

Marzi, S., Myasnikov, A.G., Serganov, A., Ehresmann, C., Romby, P., Yusupov, M., and Klaholz, B.P. (2007). Structured mRNAs Regulate Translation Initiation by Binding to the Platform of the Ribosome. Cell *130*, 1019–1031.

Massé, E., and Gottesman, S. (2002). A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. Proc. Natl. Acad. Sci. *99*, 4620–4625.

Massé, E., Escorcia, F.E., and Gottesman, S. (2003). Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*. Genes Dev. *17*, 2374–2383.

Massé, E., Vanderpool, C.K., and Gottesman, S. (2005). Effect of RyhB Small RNA on Global Iron Use in *Escherichia coli*. J. Bacteriol. *187*, 6962–6971.

Mecsas, J., Rouviere, P.E., Erickson, J.W., Donohue, T.J., and Gross, C.A. (1993). The activity of sigma E, an *Escherichia coli* heat-inducible sigma-factor, is modulated by expression of outer membrane proteins. Genes Dev. *7*, 2618–2628.

Meinnel, T., Sacerdot, C., Graffe, M., Blanquet, S., and Springer, M. (1999). Discrimination by *Escherichia coli* initiation factor IF3 against initiation on non-canonical codons relies on complementarity rules. J. Mol. Biol. *290*, 825–837.

Melamed, S., Peer, A., Faigenbaum-Romm, R., Gatt, Y.E., Reiss, N., Bar, A., Altuvia, Y., Argaman, L., and Margalit, H. (2016). Global Mapping of Small RNA-Target Interactions in Bacteria. Mol. Cell *63*, 884–897.

Mettert, E.L., and Kiley, P.J. (2015). How Is Fe-S Cluster Formation Regulated? Annu. Rev. Microbiol. *69*, 505–526.

Meyer, A.J., Ellefson, J.W., and Ellington, A.D. (2012). Abiotic Self-Replication. Acc. Chem. Res. *45*, 2097–2105.

Micali, G., and Endres, R.G. (2016). Bacterial chemotaxis: information processing, thermodynamics, and behavior. Curr. Opin. Microbiol. *30*, 8–15.

Mika, F., and Hengge, R. (2014). Small RNAs in the control of RpoS, CsgD, and biofilm architecture of *Escherichia coli*. RNA Biol. *11*, 494–507.

Miller, S.L. (1953). A production of amino acids under possible primitive earth conditions. Science *117*, 528–529.

Miller, R.A., and Britigan, B.E. (1997). Role of oxidants in microbial pathophysiology. Clin. Microbiol. Rev. *10*, 1–18.

Millman, A., Dar, D., Shamir, M., and Sorek, R. (2017). Computational prediction of regulatory, premature transcription termination in bacteria. Nucleic Acids Res. *45*, 886–893.

Milon, P., Carotti, M., Konevega, A.L., Wintermeyer, W., Rodnina, M.V., and Gualerzi, C.O. (2010). The ribosome-bound initiation factor 2 recruits initiator tRNA to the 30S initiation complex. EMBO Rep. *11*, 312–316.

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

Moon, K., and Gottesman, S. (2009). A PhoQ/P-regulated small RNA regulates sensitivity of *Escherichia coli* to antimicrobial peptides: sRNA regulates sensitivity of *E. coli* to antimicrobial peptides. Mol. Microbiol. *74*, 1314–1330.

Moon, K., and Gottesman, S. (2011). Competition among Hfq-binding small RNAs in *Escherichia coli*: Competition for Hfq in *E. coli*. Mol. Microbiol. *82*, 1545–1562.

Mukherjee, S., and Sengupta, S. (2016). Riboswitch Scanner: an efficient pHMM-based web-server to detect riboswitches in genomic sequences. Bioinformatics *32*, 776–778.

Myasnikov, A.G., Marzi, S., Simonetti, A., Giuliodori, A.M., Gualerzi, C.O., Yusupova, G., Yusupov, M., and Klaholz, B.P. (2005). Conformational transition of initiation factor 2 from the GTP- to GDP-bound state visualized on the ribosome. Nat. Struct. Mol. Biol. *12*, 1145–1149.

Nafissi, M., Chau, J., Xu, J., and Johnson, R.C. (2012). Robust Translation of the Nucleoid Protein Fis Requires a Remote Upstream AU Element and Is Enhanced by RNA Secondary Structure. J. Bacteriol. *194*, 2458–2469.

Naville, M., and Gautheret, D. (2010). Transcription attenuation in bacteria: theme and variations. Brief. Funct. Genomics *9*, 178–189.

Nikaido, H. (2003). Molecular Basis of Bacterial Outer Membrane Permeability Revisited. Microbiol. Mol. Biol. Rev. *67*, 593–656.

Nirenberg, M., Leder, P., Bernfield, M., Brimacombe, R., Trupin, J., Rottman, F., and O'neal, C. (1965). RNA codewords and protein synthesis, VII. On the general nature of the RNA code. Proc. Natl. Acad. Sci. *53*, 1161–1168.

Nocker, A., Hausherr, T., Balsiger, S., Krstulovic, N.P., Hennecke, H., and Narberhaus, F. (2001). A mRNA-based thermosensor controls expression of rhizobial heat shock genes. Nucleic Acids Res. *29*, 4800–4807.

Noinaj, N., Guillier, M., Barnard, T.J., and Buchanan, S.K. (2010). TonB-Dependent Transporters: Regulation, Structure, and Function. Annu. Rev. Microbiol. *64*, 43–60.

Noinaj, N., Gumbart, J.C., and Buchanan, S.K. (2017). The β-barrel assembly machinery in motion. Nat. Rev. Microbiol.

O'Connor, M., and Dahlberg, A.E. (2001). Enhancement of translation by the epsilon element is independent of the sequence of the 460 region of 16S rRNA. Nucleic Acids Res. *29*, 1420–1425.

O'Connor, M., Asai, T., Squires, C.L., and Dahlberg, A.E. (1999). Enhancement of translation by the downstream box does not involve base pairing of mRNA with the penultimate stem sequence of 16S rRNA. Proc. Natl. Acad. Sci. *96*, 8973–8978.

Ogasawara, H., Yamada, K., Kori, A., Yamamoto, K., and Ishihama, A. (2010). Regulation of the *Escherichia coli csgD* promoter: interplay between five transcription factors. Microbiology *156*, 2470–2483.

Olins, P.O., and Rangwala, S.H. (1989). A novel sequence element derived from bacteriophage T7 mRNA acts as an enhancer of translation of the *lacZ* gene in *Escherichia coli*. J. Biol. Chem. *264*, 16973–16976.

Omotajo, D., Tate, T., Cho, H., and Choudhary, M. (2015). Distribution and diversity of ribosome binding sites in prokaryotic genomes. BMC Genomics *16*.

Opdyke, J.A., Kang, J.-G., and Storz, G. (2004). GadY, a Small-RNA Regulator of Acid Response Genes in *Escherichia coli*. J. Bacteriol. *186*, 6698–6705.

Orme, R., Douglas, C.W.I., Rimmer, S., and Webb, M. (2006). Proteomic analysis of *Escherichia coli* biofilms reveals the overexpression of the outer membrane protein OmpA. PROTEOMICS *6*, 4269–4277.

Ortega, Á.D., Gonzalo-Asensio, J., and García-del Portillo, F. (2012). Dynamics of *Salmonella* small RNA expression in non-growing bacteria located inside eukaryotic cells. RNA Biol. *9*, 469–488.

Österberg, S., Peso-Santos, T. del, and Shingler, V. (2011). Regulation of Alternative Sigma Factor Use. Annu. Rev. Microbiol. *65*, 37–55.

Padalon-Brauch, G., Hershberg, R., Elgrably-Weiss, M., Baruch, K., Rosenshine, I., Margalit, H., and Altuvia, S. (2008). Small RNAs encoded within genetic islands of *Salmonella* typhimurium show host-induced expression and role in virulence. Nucleic Acids Res. *36*, 1913–1927.

Paetzel, M. (2014). Structure and mechanism of *Escherichia coli* type I signal peptidase. Biochim. Biophys. Acta *1843*, 1497–1508.

Pallen, M.J., and Matzke, N.J. (2006). From The Origin of Species to the origin of bacterial flagella. Nat. Rev. Microbiol. *4*, 784–790.

Pallesen, J., Hashem, Y., Korkmaz, G., Koripella, R.K., Huang, C., Ehrenberg, M. ans, Sanyal, S., and Frank, J. (2013). Cryo-EM visualization of the ribosome in termination complex with apo-RF3 and RF1. Elife *2*, e00411.

Papenfort, K., Pfeiffer, V., Mika, F., Lucchini, S., Hinton, J.C.D., and Vogel, J. (2006). Sigma E-dependent small RNAs of *Salmonella* respond to membrane stress by accelerating global omp mRNA decay. Mol. Microbiol. *62*, 1674–1688.

Parker, A., and Gottesman, S. (2016). Small RNA Regulation of TolC, the Outer Membrane Component of Bacterial Multidrug Transporters. J. Bacteriol. *198*, 1101–1113.

Paulus, M., Haslbeck, M., and Watzele, M. (2004). RNA stem-loop enhanced expression of previously non-expressible genes. Nucleic Acids Res. *32*, e78–e78.

Peng, T., Berghoff, B.A., Oh, J.-I., Weber, L., Schirmer, J., Schwarz, J., Glaeser, J., and Klug, G. (2016). Regulation of a polyamine transporter by the conserved 3′ UTR-derived sRNA SorX confers resistance to singlet oxygen and organic hydroperoxides in *Rhodobacter sphaeroides*. RNA Biol. *13*, 988–999.

Peske, F., Rodnina, M.V., and Wintermeyer, W. (2005). Sequence of Steps in Ribosome Recycling as Defined by Kinetic Analysis. Mol. Cell *18*, 403–412.

Pettis, G.S., Brickman, T.J., and McIntosh, M.A. (1988). Transcriptional mapping and nucleotide sequence of the *Escherichia coli fepA-fes* enterobactin region. Identification of a unique iron-regulated bidirectional promoter. J. Biol. Chem. *263*, 18857–18863.

Pfeiffer, V., Papenfort, K., Lucchini, S., Hinton, J.C.D., and Vogel, J. (2009). Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. Nat. Struct. 38 Mol. Biol. *16*, 840–846.

Porello, S.L., Cannon, M.J., and David, S.S. (1998). A substrate recognition role for the [4Fe-4S] 2+ cluster of the DNA repair glycosylase MutY. Biochemistry (Mosc.) *37*, 6465–6475.

Powner, M.W., Gerland, B., and Sutherland, J.D. (2009). Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. Nature *459*, 239–242.

Powner, M.W., Sutherland, J.D., and Szostak, J.W. (2010). Chemoselective Multicomponent One-Pot Assembly of Purine Precursors in Water. J. Am. Chem. Soc. *132*, 16677–16688.

Prévost, K., Salvail, H., Desnoyers, G., Jacques, J.-F., Phaneuf, É., and Massé, E. (2007). The small RNA RyhB activates the translation of *shiA* mRNA encoding a permease of shikimate, a compound involved in siderophore synthesis. Mol. Microbiol. *64*, 1260–1273.

Prevost, K., Desnoyers, G., Jacques, J.-F., Lavoie, F., and Massé, E. (2011). Small RNA-induced mRNA degradation achieved through both translation block and activated cleavage. Genes Dev. *25*, 385–396.

Proshkin, S., Rahmouni, A.R., Mironov, A., and Nudler, E. (2010). Cooperation Between Translating Ribosomes and RNA Polymerase in Transcription Elongation. Science *328*, 504–508.

Puškárová, A., Ferianc, P., Kormanec, J., Homerova, D., Farewell, A., and Nyström, T. (2002). Regulation of *yodA* encoding a novel cadmium-induced protein in *Escherichia coli*. Microbiology *148*, 3801–3811.

Qin, Y., Polacek, N., Vesper, O., Staub, E., Einfeldt, E., Wilson, D.N., and Nierhaus, K.H. (2006). The Highly Conserved LepA Is a Ribosomal Elongation Factor that Back-Translocates the Ribosome. Cell *127*, 721–733.

Quax, T.E.F., Claassens, N.J., Söll, D., and van der Oost, J. (2015). Codon Bias as a Means to Fine-Tune Gene Expression. Mol. Cell *59*, 149–161.

Quinn, H.J., Cameron, A.D.S., and Dorman, C.J. (2014). Bacterial Regulon Evolution: Distinct Responses and Roles for the Identical OmpR Proteins of *Salmonella Typhimurium* and *Escherichia coli* in the Acid Stress Response. PLoS Genet. *10*, e1004215.

Radhakrishnan, A., and Green, R. (2016). Connections Underlying Translation and mRNA Stability. J. Mol. Biol. *428*, 3558–3564.

RajBhandary, U.L. (1994). Initiator transfer RNAs. J. Bacteriol. *176*, 547.

Ramesh, V., Gite, S., Li, Y., and RajBhandary, U.L. (1997). Suppressor mutations in *Escherichia coli* methionyl-tRNA formyltransferase: role of a 16-amino acid insertion module in initiator tRNA recognition. Proc. Natl. Acad. Sci. *94*, 13524–13529.

Ramos, A., Lane, A.N., Hollingworth, D., and Fan, T.W.-M. (2004). Secondary structure and stability of the selenocysteine insertion sequences (SECIS) for human thioredoxin reductase and glutathione peroxidase. Nucleic Acids Res. *32*, 1746–1755.

Rampersaud, A., and Inouye, M. (1991). Procaine, a local anesthetic, signals through the EnvZ receptor to change the DNA binding affinity of the transcriptional activator protein OmpR. J. Bacteriol. *173*, 6882–6888.

Ratje, A.H., Loerke, J., Mikolajka, A., Brünner, M., Hildebrand, P.W., Starosta, A.L., Dönhöfer, A., Connell, S.R., Fucini, P., Mielke, T., Whitford, P.C., Onuchic, J.N., Yu, Y., Sanbonmatsu K.Y., Hartmann, R.K., Penczek, P.A., Wilson, D.N., and Spahn, C.M.T. (2010). Head swivel on the ribosome facilitates translocation by means of intra-subunit tRNA hybrid sites. Nature *468*, 713–716.

Raymond, K.N., Dertz, E.A., and Kim, S.S. (2003). Enterobactin: an archetype for microbial iron transport. Proc. Natl. Acad. Sci. *100*, 3584–3588.

Reese, M.G. (2001). Application of a time-delay neural network to promoter annotation in the Drosophila melanogaster genome. Comput. Chem. *26*, 51–56.

Righetti, F., Nuss, A.M., Twittenhoff, C., Beele, S., Urban, K., Will, S., Bernhart, S.H., Stadler, P.F., Dersch, P., and Narberhaus, F. (2016). Temperature-responsive in vitro RNA structurome of *Yersinia pseudotuberculosis*. Proc. Natl. Acad. Sci. *113*, 7237–7242.

Ringquist, S., Shinedling, S., Barrick, D., Green, L., Binkley, J., Stormo, G.D., and Gold, L. (1992). Translation initiation in *Escherichia coli*: sequences within the ribosome-binding site. Mol. Microbiol. *6*, 1219–1229.

Rollauer, S.E., Sooreshjani, M.A., Noinaj, N., and Buchanan, S.K. (2015). Outer membrane protein biogenesis in Gram-negative bacteria. Philos. Trans. R. Soc. B Biol. Sci. *370*, 20150023.

Romeo, T., Gong, M.I.N., Liu, M.Y., and Brun-Zinkernagel, A.-M. (1993). Identification and molecular characterization of csrA, a pleiotropic gene from *Escherichia coli* that affects glycogen biosynthesis, gluconeogenesis, cell size, and surface properties. J. Bacteriol. *175*, 4744–4755.

Romero, A., and García, P. (1991). Initiation of translation at AUC, AUA and AUU codons in *Escherichia coli*. FEMS Microbiol. Lett. *68*, 325–330.

Ross, W., Sanchez-Vazquez, P., Chen, A.Y., Lee, J.-H., Burgos, H.L., and Gourse, R.L. (2016). ppGpp Binding to a Site at the RNAP-DksA Interface Accounts for Its Dramatic Effects on Transcription Initiation during the Stringent Response. Mol. Cell *62*, 811–823.

Ruiz, N., Kahne, D., and Silhavy, T.J. (2009). Transport of lipopolysaccharide across the cell envelope: the long road of discovery. Nat. Rev. Microbiol. *7*, 677–683.

Rutz, J.M., Abdullah, T., Singh, S.P., Kalve, V.I., and Klebba, P.E. (1991). Evolution of the ferric enterobactin receptor in gram-negative bacteria. J. Bacteriol. *173*, 5964–5974.

Rutz, J.M., Liu, J., Lyons, J.A., Goranson, J., Armstrong, S.K., McIntosh, M.A., Feix, J.B., and Klebba, P.E. (1992). Formation of a gated channel by a ligand-specific transport protein in the bacterial outer membrane. Science *258*, 471–475.

Salvail, H., and Massé, E. (2012). Regulating iron storage and metabolism with RNA: an overview of posttranscriptional controls of intracellular iron homeostasis: Regulating iron storage and metabolism with RNA. Wiley Interdiscip. Rev. RNA *3*, 26–36.

Salvail, H., Lanthier-Bourbonnais, P., Sobota, J.M., Caza, M., Benjamin, J.-A.M., Mendieta, M.E.S., Lepine, F., Dozois, C.M., Imlay, J., and Massé, E. (2010). A small RNA promotes siderophore production through transcriptional and metabolic remodeling. Proc. Natl. Acad. Sci. *107*, 15223–15228.

Salvail, H., Caron, M.-P., Bélanger, J., and Massé, E. (2013). Antagonistic functions between the RNA chaperone Hfq and an sRNA regulate sensitivity to the antibiotic colicin. EMBO J. *32*, 2764–2778.

Sankaran, N. (2016). The RNA World at Thirty: A Look Back with its Author. J. Mol. Evol. *83*, 169–175.

Sauert, M., Wolfinger, M.T., Vesper, O., Müller, C., Byrgazov, K., and Moll, I. (2016). The MazF-regulon: a toolbox for the post-transcriptional stress response in *Escherichia coli*. Nucleic Acids Res. *44*, 6660–6675.

Sauter, C. (2003). Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein from *Escherichia coli*. Nucleic Acids Res. *31*, 4091–4098.

Scharff, L.B., Childs, L., Walther, D., and Bock, R. (2011). Local Absence of Secondary Structure Permits Translation of mRNAs that Lack Ribosome-Binding Sites. PLoS Genet. *7*, e1002155.

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

Schneider, T., and Sahl, H.-G. (2010). An oldie but a goodie - cell wall biosynthesis as antibiotic target pathway. Int. J. Med. Microbiol. IJMM *300*, 161–169.

Schu, D.J., Zhang, A., Gottesman, S., and Storz, G. (2015). Alternative Hfq-sRNA interaction modes dictate alternative mRNA recognition. EMBO J. *34*, 2557–2573.

Segré, D., Ben-Eli, D., Deamer, D.W., and Lancet, D. (2001). The lipid world. Orig. Life Evol. Biosph. *31*, 119–145.

Sengupta, J., Agrawal, R.K., and Frank, J. (2001). Visualization of protein S1 within the 30S ribosomal subunit and its interaction with messenger RNA. Proc. Natl. Acad. Sci. *98*, 11991–11996.

Serganov, A., and Nudler, E. (2013). A Decade of Riboswitches. Cell *152*, 17–24.

Shean, C.S., and Gottesman, M.E. (1992). Translation of the prophage lambda cl transcript. Cell *70*, 513–522.

Shen, P.S., Park, J., Qin, Y., Li, X., Parsawar, K., Larson, M.H., Cox, J., Cheng, Y., Lambowitz, A.M., Weissman, J.S., Brandman, O., and Frost, A. (2015). Protein synthesis. Rqc2p and 60S ribosomal subunits mediate mRNA-independent elongation of nascent chains. Science *347*, 75–78.

Sherman, F., Stewart, J.W., and Tsunasawa, S. (1985). Methionine or not methionine at the beginning of a protein. Bioessays *3*, 27–31.

Shin, S., and Park, C. (1995). Modulation of flagellar expression in *Escherichia coli* by acetyl phosphate and the osmoregulator OmpR. J. Bacteriol. *177*, 4696–4702.

Shine, J., and Dalgarno, L. (1974). The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. *71*, 1342–1346.

Silhavy, T.J., Kahne, D., and Walker, S. (2010). The Bacterial Cell Envelope. Cold Spring Harb. Perspect. Biol. *2*, a000414–a000414.

Simonetti, A., Marzi, S., Myasnikov, A.G., Fabbretti, A., Yusupov, M., Gualerzi, C.O., and Klaholz, B.P. (2008). Structure of the 30S translation initiation complex. Nature *455*, 416–420.

Simonetti, A., Marzi, S., Jenner, L., Myasnikov, A., Romby, P., Yusupova, G., Klaholz, B.P., and Yusupov, M. (2009). A structural view of translation initiation in bacteria. Cell. Mol. Life Sci. *66*, 423–436.

Skorski, P., Leroy, P., Fayet, O., Dreyfus, M., and Hermann-Le Denmat, S. (2006). The Highly Efficient Translation Initiation Region from the *Escherichia coli* rpsA Gene Lacks a Shine-Dalgarno Element. J. Bacteriol. *188*, 6277–6285.

Smirnov, A., Wang, C., Drewry, L.L., and Vogel, J. (2017). Molecular mechanism of mRNA repression in *trans* by a ProQ-dependent small RNA. EMBO J. *36*, 1029–1045.

de Smit, M.H., and van Duin, J. (1994). Translational initiation on structured messengers: Another role for the shine-dalgarno interaction. J. Mol. Biol. *235*, 173–184.

Soper, T., Mandin, P., Majdalani, N., Gottesman, S., and Woodson, S.A. (2010). Positive regulation by small RNAs and the role of Hfq. Proc. Natl. Acad. Sci. *107*, 9602–9607.

Sprengart, M.L., Fatscher, H.P., and Fuchs, E. (1990). The initiation of translation in *E. coli*: apparent base pairing between the 16S rRNA and downstream sequences of the mRNA. Nucleic Acids Res. *18*, 1719–1723.

Sprengart, M.L., Fuchs, E., and Porter, A.G. (1996). The downstream box: an efficient and independent translation initiation signal in *Escherichia coli*. EMBO J. *15*, 665.

Steitz, T.A. (1998). A mechanism for all polymerases. Nature *391*, 231–232.

Sternberg, S.H., Fei, J., Prywes, N., McGrath, K.A., and Gonzalez, R.L. (2009). Translation factors direct intrinsic ribosome dynamics during translation termination and ribosome recycling. Nat. Struct. Mol. Biol. *16*, 861–868.

Stojiljkovic, I., Bäumler, A.J., and Hantke, K. (1994). Fur regulon in gram-negative bacteria. Identification and characterization of new iron-regulated *Escherichia coli* genes by a fur titration assay. J. Mol. Biol. *236*, 531–545.

Stojnev, T., Harichová, J., Ferianc, P., and Nyström, T. (2007). Function of a Novel Cadmium-Induced YodA Protein in *Escherichia coli*. Curr. Microbiol. *55*, 99–104.

Storz, G., Vogel, J., and Wassarman, K.M. (2011). Regulation by Small RNAs in Bacteria: Expanding Frontiers. Mol. Cell *43*, 880–891.

Sussman, J.K., Simons, E.L., and Simons, R.W. (1996). *Escherichia coli* translation initiation factor 3 discriminates the initiation codon in vivo. Mol. Microbiol. *21*, 347–360.

Takahashi, S., Furusawa, H., Ueda, T., and Okahata, Y. (2013). Translation Enhancer Improves the Ribosome Liberation from Translation Initiation. J. Am. Chem. Soc. *135*, 13096–13106.

Takyar, S., Hickerson, R.P., and Noller, H.F. (2005). mRNA Helicase Activity of the Ribosome. Cell *120*, 49–58.

Tardat, B., and Touati, D. (1993). Iron and oxygen regulation of *Escherichia coli* MnSOD expression: competition between the global regulators Fur and ArcA for binding to DNA. Mol. Microbiol. *9*, 53–63.

Tedin, K., Moll, I., Grill, S., Resch, A., Graschopf, A., Gualerzi, C.O., and Bläsi, U. (1999). Translation initiation factor 3 antagonizes authentic start codon selection on leaderless mRNAs. Mol. Microbiol. *31*, 67–77.

Thomason, M.K., and Storz, G. (2010). Bacterial Antisense RNAs: How Many Are There, and What Are They Doing? Annu. Rev. Genet. *44*, 167–188.

Thomason, M.K., Fontaine, F., De Lay, N., and Storz, G. (2012). A small RNA that regulates motility and biofilm formation in response to changes in nutrient availability in *Escherichia coli*. Mol. Microbiol. *84*, 17–35.

Thomason, M.K., Bischler, T., Eisenbart, S.K., Förstner, K.U., Zhang, A., Herbig, A., Nieselt, K., Sharma, C.M., and Storz, G. (2015). Global Transcriptional Start Site Mapping Using Differential RNA Sequencing Reveals Novel Antisense RNAs in *Escherichia coli*. J. Bacteriol. *197*, 18–28.

Tian, B., Bevilacqua, P.C., Diegelman-Parente, A., and Mathews, M.B. (2004). The double-stranded-RNA-binding motif: interference and much more. Nat. Rev. Mol. Cell Biol. *5*, 1013–1023.

Timmermans, J., and Van Melderen, L. (2010). Post-transcriptional global regulation by CsrA in bacteria. Cell. Mol. Life Sci. *67*, 2897–2908.

Tjaden, B., Goodwin, S.S., Opdyke, J.A., Guillier, M., Fu, D.X., Gottesman, S., and Storz, G. (2006). Target prediction for small, noncoding RNAs in bacteria. Nucleic Acids Res. *34*, 2791–2802.

Trotochaud, A.E., and Wassarman, K.M. (2005). A highly conserved 6S RNA structure is required for regulation of transcription. Nat. Struct. Mol. Biol. *12*, 313–319.

Tsai, A., Petrov, A., Marshall, R.A., Korlach, J., Uemura, S., and Puglisi, J.D. (2012). Heterogeneous pathways and timing of factor departure during translation initiation. Nature *487*, 390–393.

Tuorto, F., and Lyko, F. (2016). Genome recoding by tRNA modifications. Open Biol. *6*, 160287.

Updegrove, T.B., Zhang, A., and Storz, G. (2016). Hfq: the flexible RNA matchmaker. Curr. Opin. Microbiol. *30*, 133–138.

Vecerek, B., Moll, I., and Bläsi, U. (2007). Control of Fur synthesis by the non-coding RNA RyhB and iron-responsive decoding. EMBO J. *26*, 965–975.

Vesper, O., Amitai, S., Belitsky, M., Byrgazov, K., Kaberdina, A.C., Engelberg-Kulka, H., and Moll, I. (2011). Selective Translation of Leaderless mRNAs by Specialized Ribosomes Generated by MazF in *Escherichia coli*. Cell *147*, 147–157.

Vinella, D., Loiseau, L., de Choudens, S.O., Fontecave, M., and Barras, F. (2013). *In vivo* [Fe-S] cluster acquisition by IscR and NsrR, two stress regulators in *Escherichia coli*: [Fe-S] cluster assembly and stress response in *E. coli*. Mol. Microbiol. *87*, 493–508.

Visca, P., Leoni, L., Wilson, M.J., and Lamont, I.L. (2002). Iron transport and regulation, cell signalling and genomics: lessons from *Escherichia coli* and *Pseudomonas*. Mol. Microbiol. *45*, 1177–1190.

Vogel, J. (2003). RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. Nucleic Acids Res. *31*, 6435–6443.

Voorhees, R.M., and Ramakrishnan, V. (2013). Structural Basis of the Translational Elongation Cycle. Annu. Rev. Biochem. *82*, 203–236.

Wagner, E.G.H., and Romby, P. (2015). Small RNAs in Bacteria and Archaea. In Advances in Genetics, (Elsevier), pp. 133–208.

Wagner, E.G.H., Altuvia, S., and Romby, P. (2002). 12-Antisense RNAs in Bacteria and Their Genetic Elements. Adv. Genet. *46*, 361–398.

Walsh, N.P., Alba, B.M., Bose, B., Gross, C.A., and Sauer, R.T. (2003). OMP peptide signals initiate the envelope-stress response by activating DegS protease via relief of inhibition mediated by its PDZ domain. Cell *113*, 61–71.

Wandersman, C., and Delepelaire, P. (2004). Bacterial Iron Sources: From Siderophores to Hemophores. Annu. Rev. Microbiol. *58*, 611–647.

Wang, J., Rennie, W., Liu, C., Carmack, C.S., Prévost, K., Caron, M.-P., Massé, E., Ding, Y., and Wade, J.T. (2015). Identification of bacterial sRNA regulatory targets using ribosome profiling. Nucleic Acids Res. gkv1158.

Wassarman, K.M. (2001). Identification of novel small RNAs using comparative genomics and microarrays. Genes Dev. *15*, 1637–1651.

Wassarman, K.M., and Storz, G. (2000). 6S RNA regulates *E. coli* RNA polymerase activity. Cell *101*, 613–623.

Wen, J.-D., Lancaster, L., Hodges, C., Zeri, A.-C., Yoshimura, S.H., Noller, H.F., Bustamante, C., and Tinoco, I. (2008). Following translation by single ribosomes one codon at a time. Nature *452*, 598–603.

Westermann, A.J., Förstner, K.U., Amman, F., Barquist, L., Chao, Y., Schulte, L.N., Müller, L., Reinhardt, R., Stadler, P.F., and Vogel, J. (2016). Dual RNA-seq unveils noncoding RNA functions in host–pathogen interactions. Nature *529*, 496–501.

Wimberly, B.T., Brodersen, D.E., Clemons, W.M., Morgan-Warren, R.J., Carter, A.P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000). Structure of the 30S ribosomal subunit. Nature *407*, 327–339.

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

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

Woese, C.R., Fox, G.E., Zablen, L., Uchida, T., Bonen, L., Pechman, K., Lewis, B.J., and Stahl, D. (1975). Conservation of primary structure in 16S ribosomal RNA. Nature *254*, 83–86.

Wright, P.R., Richter, A.S., Papenfort, K., Mann, M., Vogel, J., Hess, W.R., Backofen, R., and Georg, J. (2013). Comparative genomics boosts target prediction for bacterial small RNAs. Proc. Natl. Acad. Sci. *110*, E3487–E3496.

Wu, T., Malinverni, J., Ruiz, N., Kim, S., Silhavy, T.J., and Kahne, D. (2005). Identification of a Multicomponent Complex Required for Outer Membrane Biogenesis in *Escherichia coli*. Cell *121*, 235–245.

XiaoLong, Z., and EnDuo, W. (2013). Transfer RNA: A dancer between charging and mis-charging for protein biosynthesis. Sci. China Life Sci. *56*, 921–932.

Yaku, H., and Mizuno, T. (1997). The membrane-located osmosensory kinase, EnvZ, that contains a leucine zipper-like motif functions as a dimer in *Escherichia coli*. FEBS Lett. *417*, 409–413.

Yamamoto, H., Qin, Y., Achenbach, J., Li, C., Kijek, J., Spahn, C.M.T., and Nierhaus, K.H. (2013). EF-G and EF4: translocation and back-translocation on the bacterial ribosome. Nat. Rev. Microbiol. *12*, 89–100.

Yamamoto, H., Wittek, D., Gupta, R., Qin, B., Ueda, T., Krause, R., Yamamoto, K., Albrecht, R., Pech, M., and Nierhaus, K.H. (2016). 70S-scanning initiation is a novel and frequent initiation mode of ribosomal translation in bacteria. Proc. Natl. Acad. Sci. *113*, E1180–E1189.

Yamamoto, K., Nagura, R., Tanabe, H., Fujita, N., Ishihama, A., and Utsumi, R. (2000). Negative regulation of the *bolA1p* of *Escherichia coli* K-12 by the transcription factor OmpR for osmolarity response genes. FEMS Microbiol. Lett. *186*, 257–262.

Yartseva, V., Takacs, C.M., Vejnar, C.E., Lee, M.T., and Giraldez, A.J. (2016). RESA identifies mRNA-regulatory sequences at high resolution. Nat. Methods *14*, 201–207.

Yu, D., Ellis, H.M., Lee, E.-C., Jenkins, N.A., Copeland, N.G., and others (2000). An efficient recombination system for chromosome engineering in *Escherichia coli*. Proc. Natl. Acad. Sci. *97*, 5978–5983.

Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H.D., and Noller, H.F. (2001). Crystal structure of the ribosome at 5.5 Å resolution. Science *292*, 883–896.

Yusupova, G.Z., Yusupov, M.M., Cate, J.H.D., and Noller, H.F. (2001). The path of messenger RNA through the ribosome. Cell *106*, 233–241.

Zavialov, A.V., Hauryliuk, V.V., and Ehrenberg, M. (2005). Splitting of the Posttermination Ribosome into Subunits by the Concerted Action of RRF and EF-G. Mol. Cell *18*, 675–686.

Zhang, A. (1998). The OxyS regulatory RNA represses *rpoS* translation and binds the Hfq (HF-I) protein. EMBO J. *17*, 6061–6068.

Zhang, A., Rosner, J.L., and Martin, R.G. (2008). Transcriptional activation by MarA, SoxS and Rob of two *tolC* promoters using one binding site: a complex promoter configuration for *tolC* in *Escherichia coli*. Mol. Microbiol. *69*, 1450–1455.

Zhang, A., Schu, D.J., Tjaden, B.C., Storz, G., and Gottesman, S. (2013). Mutations in Interaction Surfaces Differentially Impact *E. coli* Hfq Association with Small RNAs and Their mRNA Targets. J. Mol. Biol. *425*, 3678–3697.

Zheng, T., Bullock, J.L., and Nolan, E.M. (2012). Siderophore-Mediated Cargo Delivery to the Cytoplasm of *Escherichia coli* and *Pseudomonas aeruginosa* : Syntheses of Monofunctionalized Enterobactin Scaffolds and Evaluation of Enterobactin–Cargo Conjugate Uptake. J. Am. Chem. Soc. *134*, 18388–18400.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. *31*, 3406–3415.