Conception et implémentation in vivo chez la bactérie E. coli de commutateurs ARN synthétiques

Gabriel Joël Espel

▶ **To cite this version:**

Gabriel Joël Espel. Conception et implémentation in vivo chez la bactérie E. coli de commutateurs ARN synthétiques. Biologie structurale [q-bio.BM]. Université Grenoble Alpes [2020-..], 2020. Français. NNT : 2020GRALV051 . tel-03399389

HAL Id: tel-03399389

https://theses.hal.science/tel-03399389

Submitted on 24 Oct 2021

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.

Université Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE GRENOBLE ALPES

Spécialité : **Biologie Structurale et Nanobiologie**

Arrêté ministériel : 25 mai 2016

Présentée par

Gabriel Joël ESPEL

Thèse dirigée par **Hans GEISELMANN, PR, Université Grenoble Alpes**, et co-encadrée par **Alexandre DAWID, Maître de conférences, Université Grenoble Alpes**

préparée au sein du **Laboratoire Interdisciplinaire de Physique** dans **l'École Doctorale Chimie et Sciences du Vivant**

Conception et implémentation *in vivo* chez la bactérie *Escherichia coli* de commutateurs ARN synthétiques

Thèse soutenue publiquement le **23 octobre 2020**, devant le jury composé de :

Madame Muriel COCAIGN-BOUSQUET
Directrice de Recherche, UMR 5504 TBI, Rapportrice
Monsieur Marc BOUDVILLAIN
Directeur de Recherche, Université d'Orléans, Rapporteur
Monsieur Jérôme BONNET
Chargé de Recherche, Université de Montpellier, Examinateur
Monsieur Franz BRUCKERT
Professeur des Universités, Grenoble INP Université Grenoble Alpes, Président du jury
Monsieur Hervé GUILLOU
Maître de conférences, Université Grenoble Alpes, Examinateur
Madame Delphine ROPERS
Chargée de Recherche, INRIA Grenoble – Rhône-Alpes, Invitée
Monsieur Hans GEISELMANN
PR, Université Grenoble Alpes, Directeur de thèse
Monsieur Alexandre DAWID
Maître de conférences, Université Grenoble Alpes, Invité

Remerciements

Ainsi s'achèvent 4 années d'aventures extraordinaires auprès de personnes remarquables et pour cette occasion j'aimerais remercier toutes ces personnes qui y ont contribuées, de près ou de loin.

Tout d'abord, je tiens à remercier tous les membres du jury : Muriel Cocaign-Bousquet, Marc Boudvillain, Jérôme Bonnet, Franz Bruckert, Hervé Guillou, mais également Delphine Ropers en sa qualité d'invitée, d'avoir consacré du temps à la lecture de ce travail de thèse. Je tiens également à remercier l'Université Grenoble Alpes pour le financement de cette thèse.

Je remercie tout particulièrement mes deux directeurs de thèse, qui m'ont encadré pendant ces 4 années. Un grand merci notamment à Alexandre Dawid pour m'avoir accordé sa confiance et m'avoir donné l'opportunité de travailler sur ce sujet de thèse. Je remercie également Hans Geiselmann pour son expertise et sa disponibilité. Merci pour vos encouragements, soutiens et précieux conseils qui m'ont beaucoup aidé tout au long de cette thèse et qui me serviront même plus tard. J'ai énormément appris à vos côtés.

J'aimerais aussi remercier tous les stagiaires qui ont apporté leur contribution à la réalisation de ce travail de thèse : Margot Sonzogni et Valerio Piomponi.

Je remercie très chaleureusement tous les membres de l'équipe BIOP, notamment : Corinne, Mathilde, Antrea, Maaike, Nasab et Noël avec qui j'ai partagé des expertises scientifiques mais également des moments de convivialité que je n'oublierai jamais. Aussi je tiens à remercier Stéphan Lacour qui m'a fait profiter de son expérience non seulement dans le monde de la recherche mais aussi de l'enseignement.

Finalement, les derniers mais non les moindres, un grand merci à Irina, ma collègue de bureau, pour les discussions scientifiques…ou pas, toujours agrémentées de sa bonne humeur communicative. Un grand merci également à Célia et Charlotte avec qui j'ai passé des moments inoubliables.

Et bien sûr, un grand merci à toute ma famille, plus particulièrement ma maman pour son soutien inconditionnel.

Résumé

Ce travail de thèse s'inscrit dans le thème général de la biologie de synthèse de l'ARN et vise à construire et tester le fonctionnement de commutateurs ARN contrôlés par des petits ARN antisens. La question fondamentale correspondante concerne les relations entre séquence, repliement et fonction d'une molécule d'ARN.

Nous avons conçu des ARN commutateurs qui adoptent deux conformations alternatives : une conformation forme un terminateur de la transcription, empêchant l'expression d'un gène en aval, tandis que la conformation alternative ne conduit pas à la terminaison de la transcription. La sélection de la conformation dépend de la présence ou l'absence d'un petit ARN antisens fourni ou exprimé en *trans*. Nous avons validé le fonctionnement de commutateurs répressibles et inductibles *in vitro*. Afin d'évaluer leur fonctionnement *in vivo*, nous avons construit et optimisé une plateforme génétique chez la bactérie *Escherichia coli*. Le fonctionnement du commutateur est mesuré par l'expression d'une protéine fluorescente en aval du commutateur. Nous avons ainsi fourni la preuve de concept du fonctionnement de commutateurs activables et répressibles *in vivo*.

Nos expériences *in vitro* et *in vivo* nous ont permis de découvrir de nouveaux principes de design des ARN synthétiques, d'optimiser les conditions d'expression des commutateurs et les méthodes de mesure de l'expression *in vivo*. La gamme dynamique de la régulation par les petits ARN antisens *in vivo* doit être augmentée et nous avons identifié les pistes à explorer. Plus généralement, nos résultats ouvrent la voie à l'exploration plus systématique du lien séquence-repliement-fonction des ARN.

Table des matières

Remerciements i

Résumé iii

Partie I : Introduction : l'ARN est une molécule versatile 1

Chapitre 1 : Généralités sur les ARN 2

1. Composition de l'ARN 2

2. La synthèse des ARN 4

3. Structure des ARN 5

3.1. Structure primaire de l'ARN 5

3.2. Structure secondaire de l'ARN 8

3.3. Hélice 8

3.4. Tige-boucle 8

3.5. Boucle interne 8

3.6. Bulge (ou boucle latérale) 8

3.7. Jonction d'hélices 8

3.8. Structure tertiaire de l'ARN 9

3.9. Appariement Hoogsteen 9

3.10. Interaction triple 9

3.11. Interaction base-ribose 10

3.12. Les pseudonœuds 10

Chapitre 2 : Prédiction des structures secondaires des ARN 12

1. Méthode comparative 12

2. Algorithme combinatoire 12

3. Algorithme génétique 13

3.1. La mutation 13

3.2. La recombinaison (croisement) 13

3.3. La sélection 13

4. Algorithmes de programmation dynamique 14
5. Algorithme cinétique 14
6. Kinefold 15
6.1. Paramètres donnés en entrée 19
6.2. Résultats obtenus en sortie de simulation 19
Chapitre 3 : Ingénierie de l'ARN pour la biologie synthétique 20
1. Riboswitches naturels : localisation, structure et fonction 20
1.1. Régulation de la traduction 21
1.2. Régulation de la transcription 22
1.2.1. La terminaison de la transcription 22
1.2.2. Contrôler la formation du terminateur 23
2. Commutateurs ARN synthétiques 26
2.1. Mécanismes d'action des petits ARN 26
2.2. Rôle des ARN chaperonnes 27
2.3. Design et ingénierie des riboswitches 28
Chapitre 4 : Objectifs de la thèse 29
Partie II : Validation *in vitro* du fonctionnement des commutateurs ARN, générés par méthode de design semi-automatique 31
Chapitre 1 : Introduction 31
Chapitre 2 : Principe de mesure *in vitro* du fonctionnement des commutateurs ARN 32
1. La transcription *in vitro* 32
2. Séparation des transcrits d'ARN sur PAGE dénaturant 33
3. Visualisations des transcrits d'ARN 35
4. Analyse de l'image du gel d'électrophorèse 35
5. Quantification du taux de terminaison 36
Chapitre 3 : Résultats et discussions 37
1. Fonctionnement des commutateurs ARN inductibles 37

1.1. État nominal (absence d'antisens ARN) 37
1.2. Commutation en présence d'antisens ARN 41
2. Test d'un commutateur ARN répressible 43
2.1. État nominal (absence d'antisens ARN) 43
2.2. Commutation en présence d'antisens ARN 45
Chapitre 4 : Conclusion 49
Partie III : Développement d'une plateforme et d'une méthode pour le test et la caractérisation *in vivo* des commutateurs ARN 51
Chapitre 1 : Plateforme génétique et instrumentale pour la caractérisation *in vivo* du fonctionnement des commutateurs ARN : généralités 51
1. Principe général 51
2. Méthode de mesure et de caractérisation des commutateurs ARN 53
2.1. Conditions permettant la reproductibilité des mesures 53
2.2. Rappels sur les modes de cultures bactériennes 54
3. Spécificités visées pour la plateforme génétique 55
4. Principe de la mesure de l'expression moyenne de la protéine rapportrice par bactérie 56
4.1. Choix d'un instrument de mesure : le lecteur de microplaques 58
4.2. Fonctionnement du lecteur de microplaques : l'Infinite M200 Pro de TECAN 58
4.3. Système de mesure de la fluorescence 59
4.4. Système de mesure de l'absorbance 61
Chapitre 2 : Mise en œuvre expérimentale et analyse des données : principes 62
1. Stratégie pour atteindre une croissance équilibrée 62
2. Analyse des données de mesure 63
3. Nécessité d'homogénéiser le milieu avec des microbilles 64
4. L'outil WELLINVERTER pour le traitement et l'analyse des données de culture en microplaques, en présence de billes 66

Chapitre 3 : Plateforme génétique et instrumentale pour la caractérisation *in vivo* du fonctionnement des commutateurs ARN : mise en œuvre initiale et optimisation 68

1. Choix des composants/éléments initiaux de la plateforme 68

1.1. Origines de réplication et promoteurs de la transcription 68

1.2. Bactérie hôte : BW25113 69

1.3. Protéine fluorescente : mKate2 69

1.4. Terminateur de la transcription de l'antisens ARN : $T1_{rrnB}$ 70

2. Test des plateformes d'expression : fonctionnement des promoteurs inductibles $p_{LlacO-1}$ et $p_{LtetO-1}$ 71

2.1. Introduction 71

2.2. Induction du promoteur $p_{LlacO-1}$: contrôle du niveau de transcription du système commutateur ARN-gène rapporteur 72

2.3. Induction du promoteur $p_{LtetO-1}$: contrôle du niveau de transcription de l'antisens ARN (antisens I) 74

2.4. Restauration de la croissance des bactéries, en milieu riche défini (M9 + suppléments) 76

2.5. Induction du promoteur $p_{LtetO-1}$ par l'anhydrotetracycline (aTc) en milieu riche défini 78

3. Modification du vecteur d'expression des antisens I : expression constitutive et fort nombre de copies 79

3.1. Test du fonctionnement de la nouvelle plateforme d'expression de l'antisens (antisens II) : test du niveau d'expression de l'antisens via la protéine rapportrice mKate2 80

4. Modification du vecteur d'expression de l'antisens II : optimisation du promoteur et de sa séquence avale (antisens III) 83

4.1. Transcription *in vitro* des antisens as1 et as13 à partir de leur vecteur d'expression respectif 83

4.2. Expression de as1 et as13 via le contexte d'expression de l'antisens RNAI 85

Chapitre 4 : Optimisation du protocole de mesure pour la caractérisation *in vivo* des commutateurs ARN : l'importance de se placer en condition de croissance équilibrée 90

1. Premier test *in vivo* du fonctionnement du commutateur ARN en absence ou présence de l'antisens ARN 90

2. Amélioration des courbes de DO_{600} et de fluorescence/DO_{600} avec l'utilisation des billes 92

3. Mise au point d'un protocole permettant de se placer en condition de croissance équilibrée 94

Chapitre 5 : Conclusion 97

Partie IV : Validation *in vivo*, du fonctionnement des commutateurs ARN 100

Chapitre 1 : Introduction 100

Chapitre 2 : Optimisation des séquences des commutateurs ARN et des antisens ARN 101

1. Identification et suppression des séquences promotrices cryptiques internes dans rSw1 : obtention de rSw1.v2 101

1.1. Analyse de séquence : prédiction de séquences promotrices dans la séquence de rSw1 102

1.2. Mutations des régions consensus des promoteurs potentiels identifiés dans la séquence du commutateur rSw1 : suppression de l'activité promotrice interne 104

1.3. Suppression de l'activité des promoteurs cryptiques internes dans rSw1 : validation *in vivo* 107

2. Identification et suppression d'effets séquence-spécifiques des antisens ARN sur la croissance des bactéries 109

2.1. L'antisens ARN as13 a un effet sur la croissance des bactéries 109

2.2. Choix d'un bon contrôle de spécificité pour l'effet des antisens ARN effecteurs sur les commutateurs ARN 112

2.3. Optimisation de la séquence de l'antisens ARN as13 : réduction de l'effet sur la croissance des bactéries 113

Chapitre 3 : Validation *in vivo*, du fonctionnement des commutateurs ARN inductible iSw13 et répressible rSw1 115

1. Introduction 115
2. Fonctionnement du commutateur ARN inductible iSw13 dans la bactérie *E. coli* 116
3. Fonctionnement du commutateur ARN répressible rSw1 dans la bactérie *E. coli* 118
4. Conclusions 119
Partie V : Conclusions, discussions et perspectives 120
Conception de commutateurs ARN 120
Validation *in vitro* des commutateurs 121
Développement d'une plateforme de test *in vivo* 122
Optimisations 122
1. Mesures en présence de billes 122
2. Améliorations du design des commutateurs 123
Évaluation de deux commutateurs optimisés sur la plateforme 123
Perspectives d'améliorations 124
1. Extension du fonctionnement des commutateurs à la phase exponentielle de croissance 124
2. Amélioration des critères pour le design 125
3. Augmenter la stabilité de l'ARN antisens 125
4. Augmenter la vitesse d'hybridation entre ARN antisens et cible 125
Perspectives concernant les principes de design 126
Partie VI : Annexes 127
Annexe A : Listes des plasmides et souches utilisés 127
1. Listes des plasmides 127
2. Liste des souches de bactéries 128
3. Liste des séquences des commutateurs ARN 128
4. Liste des coeurs de séquence des antisens ARN 129
Annexe B : Méthodes expérimentales 129

1. Composition des milieux de culture utilisés 129
1.1. Milieu LB (Luria-Bertani), pour 1 L 129
1.2. Milieu minimum M9, pour 1 L 129
2. Protocoles de biologie moléculaire et microbiologie 130
2.1. Transcription *in vitro* 130
2.2. Électrophorèse sur gel de polyacrylamide 131
3. Réglages utilisés pour des mesures typiques de fluorescence et d'absorbance avec le TECAN Infinite M200 PRO 131
Bibliographie 133

Partie I : Introduction : l'ARN est une molécule versatile

La molécule centrale de mes travaux de thèse est l'ARN, acide ribonucléique. Tandis que l'ADN (acide désoxyribonucléique) stocke notre patrimoine génétique, l'ARN forme l'intermédiaire entre l'ADN et la protéine. Cette molécule possède également une importance toute particulière dans l'évolution. Selon l'hypothèse communément acceptée, les premiers stades de la vie sur terre sont caractérisés par un monde tout ARN [1]. Cette hypothèse de l'origine de la vie découle de la capacité de la molécule d'ARN de pouvoir à la fois stocker l'information génétique (ex : ARN des virus) mais aussi de réaliser des activités catalytiques ou enzymatiques (ex : les ARN ribosomaux ou les ribozymes).

Dans la cellule, les molécules d'ADN sont transcrites en ARNm (ARN messagers) par un complexe protéique appelé ARN polymérase. Certaines régions de ces ARNm, qu'on appelle régions codantes, sont ensuite traduites en protéines par la machinerie du ribosome qui est à la fois composée de protéines et d'ARN, les ARNr (ARN ribosomaux). Cette étape de traduction fait aussi appel à d'autres éléments tels que les ARNt (ARN de transfert). Les ARNt permettent de traduire un triplet de nucléotides (codon) localisé dans la séquence d'ARNm, en acides aminés. Il existe dans l'ARNm, quatre types de nucléotides, A, U, C et G. La combinatoire de 3 de ces 4 nucléotides permet le codage de 64 codons ($4^3 = 64$) différents qui peuvent coder de manière redondante 20 acides aminés différents dont l'enchaînement donne naissance à la protéine. Ce sont les ARNr de la machinerie du ribosome qui catalysent la formation des liaisons peptiques entre les acides aminés, lors de la traduction.

Rien qu'en regardant les processus de transcription et de traduction, nous pouvons déjà remarquer l'existence d'une grande diversité fonctionnelle de l'ARN. Cette grande diversité fonctionnelle est en partie liée à leur grande diversité structurale. Dans un certain nombre de processus, certaines structures d'ARN peuvent même se lier spécifiquement à des molécules cibles et changer de forme en réponse à celles-ci pour réaliser une fonction donnée.

Nous verrons dans cette introduction la composition et la synthèse des molécules d'ARN, ainsi que les différents niveaux de structure qu'elles peuvent adopter (structure primaire, structure secondaire et structure tertiaire). Nous verrons ensuite différentes méthodes bio-informatiques utilisées pour la prédiction de la structure secondaire des ARN. Enfin, nous détaillerons 2 méthodes de conception de séquences ARN synthétiques, développées par Alexandre Dawid *et al.* [2].

Chapitre 1 : Généralités sur les ARN

1. Composition de l'ARN

Comme l'ADN, l'ARN est un polymère composé d'une répétition de 4 monomères différents qu'on appelle nucléotides. Les nucléotides se décomposent en 3 modules distincts : au maximum 3 groupes phosphates liés par une liaison ester à un sucre pentose, lui-même lié à une base par une liaison β-glycosidique (Figure 1A).

Dans l'ARN le sucre est un ribose, tandis que dans l'ADN il s'agit d'un désoxyribose. Le ribose et le désoxyribose sont tous les 2 des sucres pentoses contenant une fonction aldéhyde et 5 atomes de carbones numérotés de 1' à 5' qui portent chacun une fonction hydroxyle (- OH), sauf le désoxyribose qui n'en porte pas sur le carbone 2' (Figure 1B). Cette fonction hydroxyle supplémentaire sur le carbone 2' du ribose confère à la molécule d'ARN des fonctions que l'on ne retrouve pas dans la molécule d'ADN.

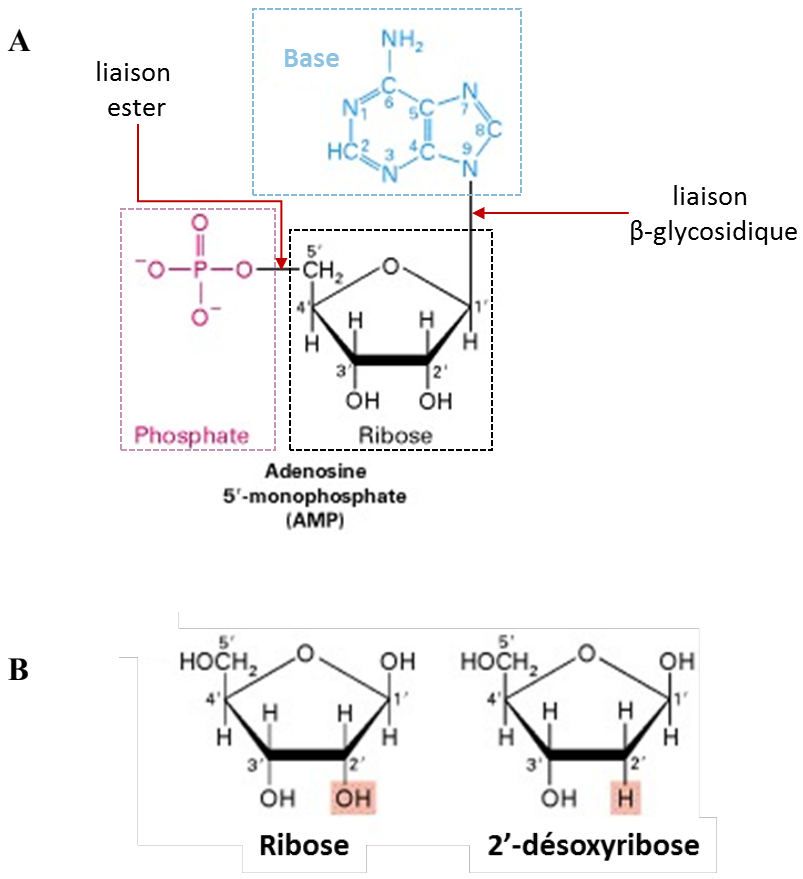


Figure 1 **(A) Structure générale d'un nucléotide.** Il s'agit ici d'une adénosine 5'-monophosphate (rAMP). **(B)** A gauche, sucre utilisé dans l'ARN et à droite, sucre utilisé dans l'ADN [3].

La distribution asymétrique des charges dans la fonction hydroxyle confère à la molécule d'ARN une propriété chimique plus réactive comparé à l'ADN qui est beaucoup plus stable, cohérent avec sa fonction de conservation de l'information génétique. Par ailleurs, la présence de la fonction hydroxyle sur le carbone 2' rajoute dans la molécule de ribose un encombrement stérique beaucoup plus important. Pour minimiser ces encombrements stériques, le ribose adopte une conformation torsadée qui est impliqué dans le type de structure secondaire adopté par l'ARN comparé à l'ADN, et que nous verrons plus loin dans ce chapitre.

Concernant les bases, ce sont des hétérocycles contenant des atomes de carbones, d'oxygènes et d'azotes, et sont au nombre de 4 : l'adénine (A) et la guanine (G) qui contiennent un double noyau aromatique et qui appartiennent à la famille des purines, l'uracile (U) et la cytosine (C) qui contiennent un seul noyau aromatique et qui appartiennent à la famille des pyrimidines. Toutes ces bases sont communes à la molécule d'ARN et d'ADN, sauf l'uracile qui est remplacée par la thymine (T) dans les molécules d'ADN (Figure 2).

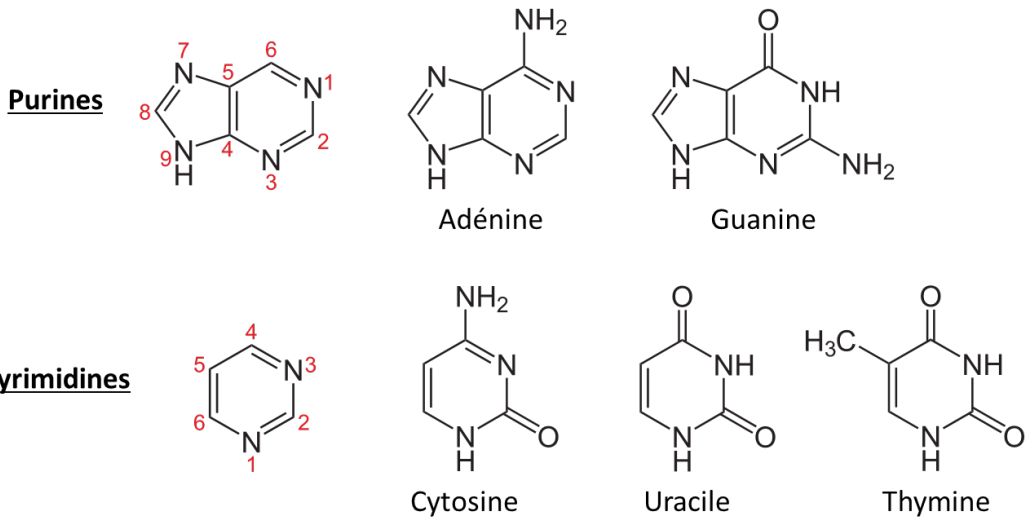


Figure 2 Structure chimique des 4 bases différentes qui composent les nucléotides.

Ces bases sont liées aux sucres par une liaison β-glycosidique, entre le carbone 1' du sucre (ribose ou désoxyribose) et l'azote 9 d'une purine ou l'azote 1 d'une pyrimidine. Cet ensemble sucre-base forme le nucléoside. Par l'action d'un acide phosphorique, la fonction hydroxyle portée par le carbone 5' du nucléoside est estérifiée pour former le nucléotide. Plus précisément, on parlera de ribonucléotide pour l'ARN et de désoxyribonucléotide pour l'ADN. Ces nucléotides peuvent porter jusqu'à 3 groupes phosphates.

Pour former un acide nucléique (ARN ou ADN), les nucléotides se polymérisent en formant une liaison ester entre le groupe hydroxyle porté le carbone 3' d'un nucléotide en amont et le

phosphate porté par le carbone 5' d'un autre nucléotide en aval. De longues chaînes linéaires de nucléotides sont ainsi reliées par des liaisons phosphodiesters séquentielles, alternant sucres et liaisons phosphodiesters, constituant le squelette de l'acide nucléique (Figure 3). Dans la notation conventionnelle, la séquence linéaire formée par cette succession de nucléotides est notée de l'extrémité 5' qui porte un phosphate libre, vers l'extrémité 3' qui porte un groupe hydroxyle libre (Figure 3), et constitue la « structure primaire » des acides nucléiques que nous verrons plus en détails dans ce même chapitre.

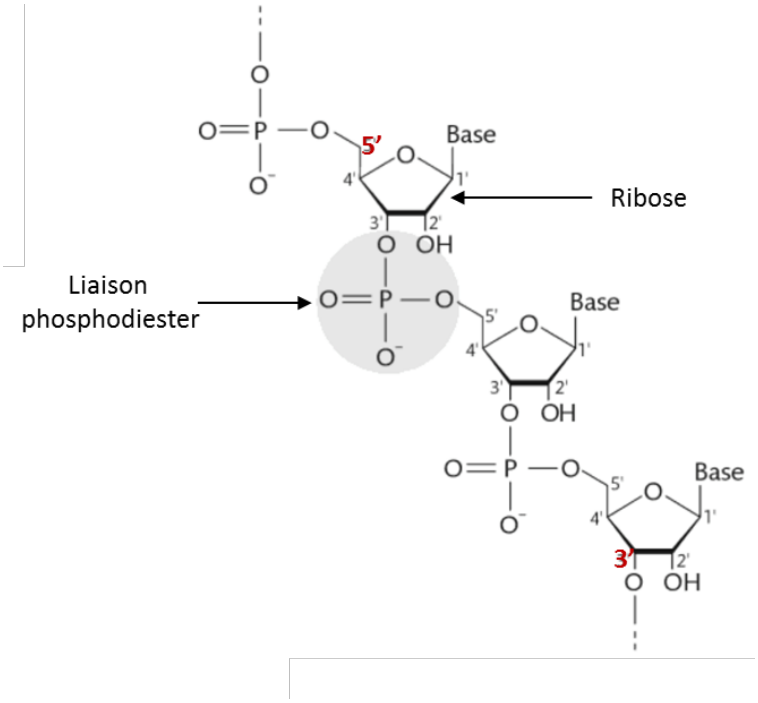


Figure 3 Composition chimique d'un squelette phosphodiester. Formation d'un acide nucléique (ici l'ARN) par polymérisation des nucléotides.

2. La synthèse des ARN

Malgré une structure primaire similaire, les structures natives des molécules d'ADN et d'ARN sont très différentes. L'ADN dans la cellule se trouve souvent sous forme de double hélice, tandis que l'ARN est le plus souvent sous forme simple brin. Cet ARN simple brin peut néanmoins contenir de courtes hélices espacées entre elles par des séquences ARN simple brin. Toute base de l'ARN peu potentiellement interagir avec une base complémentaire dans la même molécule, conduisant ainsi à la formation de structures tridimensionnelles complexes, rappelant la structure globulaire des protéines (Figure 4).

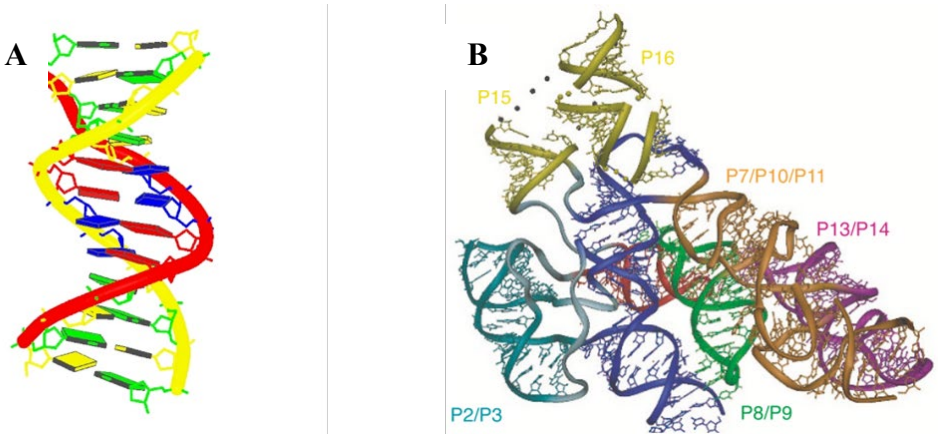


Figure 4 ADN et ARN ont une structure native différente. (A) Schéma de la structure de l'ADN de type B [4]. **(B)** Schéma de la structure de la partie ARN d'une ribonucléase P (RNase P) [5], qui correspond à une ribonucléoprotéine utilisée pour la maturation des ARN de transfert.

Cette différence structurale s'explique notamment par la façon dont l'ADN et l'ARN sont synthétisés dans la cellule. La synthèse de nouvelles molécules d'ADN se fait par réplication, tandis que la synthèse de nouvelles molécules d'ARN se fait par transcription. Lors du processus de réplication, les 2 brins antiparallèles qui composent l'ADN parent sont séparés en 2 et répliqués de sorte que 2 molécules d'ADN double brin identiques sont synthétisées. Par contre, lors de la transcription, les brins qui constituent l'ADN parent ne sont séparés que de manière transitoire et un seul de ces 2 brins est utilisé comme matrice pour synthétiser la molécule d'ARN. La transcription de l'ADN en ARN est initiée par une séquence spécifique de l'ADN, qu'on appelle promoteur, reconnue par l'ARN polymérase, un complexe enzymatique qui catalyse la polymérisation des ribonucléotides. A l'arrêt de la transcription, l'ARN nouvellement synthétisé est séparé de l'ADN matrice qui s'hybride à nouveau avec le brin d'ADN qui lui est complémentaire. L'ARN quant à lui reste simple brin. Il peut cependant former des appariements de bases intramoléculaire qui conduisent à la formation d'hélices intramoléculaires et créent un niveau supérieur d'organisation des molécules d'ARN que nous détaillerons plus tard dans ce même chapitre.

3. Structure des ARN

3.1. Structure primaire de l'ARN

Comme mentionné précédemment, la séquence d'ARN unidimensionnelle composée d'une succession de bases, représentées par les lettres A, U, G et C constitue ce que l'on appelle **la structure primaire**.

$$\Sigma = \{A, U, G, C\}$$

$$R = r_1, r_2 \ldots, rn, \qquad r_i \in R$$

Même si la molécule d'ARN est synthétisée sous la forme d'un simple brin, il est généralement observé que les structures d'ARN s'organisent autour de successions de paires de bases qui engendrent une structure hélicoïdale régulière. Contrairement à l'ADN, il existe dans l'ARN différentes manières d'arranger spatialement 2 nucléotides isolés pour former des liaisons hydrogènes. Selon Leontis et Westhof [6], il existe dans l'ARN, 168 modèles d'appariement de bases, dont 110 ont déjà été identifiés expérimentalement. Cependant, dans une hélice, on ne rencontre que les appariements qu'on appelle Watson-Crick, qui implique 2 liaisons hydrogène entre les bases A et U et 3 liaisons hydrogènes entre les bases G et C (Figure 5A et Figure 5B). On peut y retrouver également un troisième type d'appariement non-canonique qui forme 2 liaisons hydrogènes entre les bases G et U (Figure 5C). Même s'il existe différents modèles d'appariement de bases dans la molécule d'ARN, ces 3 types d'appariement sont largement majoritaires dans une hélice d'ARN pour des raisons d'encombrement stérique imposé par la structure hélicoïdale de la chaîne.

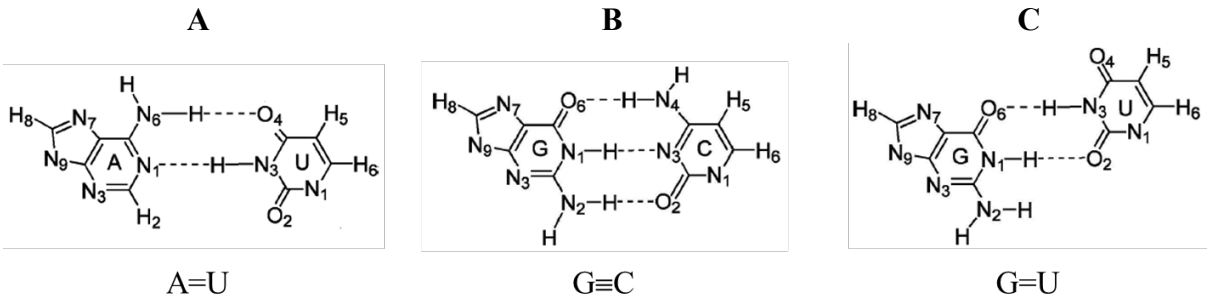


Figure 5 **Les appariements de bases qu'on retrouve dans une hélice d'ARN. (A)** et **(B)** appariement Watson-Crick. **(C)** appariement non-canonique.

Les hélices sont également stabilisées par l'empilement des bases successives (stacking) le long d'un même brin, conduisant à des interactions de Van der Waals et des interactions π–π entre les cycles aromatiques qui les composent. Ces interactions maintiennent les cycles aromatiques hydrophobes empilés à l'intérieur de la double hélice et minimisent leur surface de contact avec l'environnement aqueux (Figure 6).

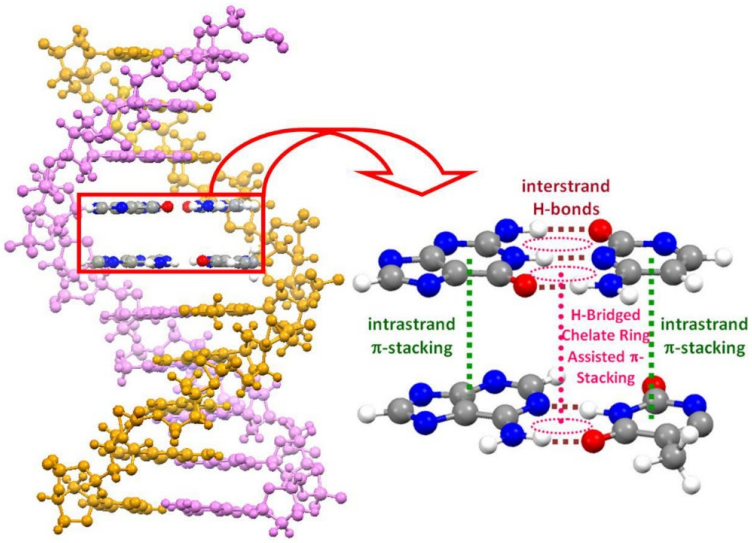


Figure 6 Formation de liaisons hydrogènes entre les paires de bases Watson-Crick et empilement des bases suivant l'axe de la double hélice.

Les doubles hélices d'ADN et d'ARN ont des propriétés géométriques légèrement différentes l'une de l'autre. Elles se présentent sous la forme B pour la double hélice d'ADN et sous la forme A pour la double hélice d'ARN. La forme B correspond à une hélice avec un pas de 36 Å et 10,4–10,5 paires de bases par tour d'hélice [7]. Les bases y sont perpendiculaires à l'axe de l'hélice et parallèles entre elles. La forme A quant à elle est légèrement plus compacte que la forme B et correspond à une hélice avec un pas de 28 Å et 11 paires de bases par tour d'hélice. Les bases y sont inclinées d'un angle de 18° par rapport à l'axe de la double hélice [8][9]. Ainsi, contrairement à l'hélice d'ADN, l'hélice d'ARN est constituée d'un petit sillon peu profond et large, et un grand sillon profond et étroit.

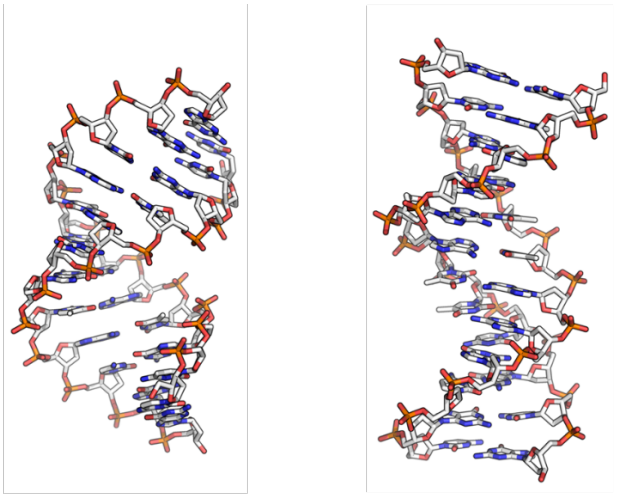


Figure 7 Forme A de l'hélice d'ARN et forme B de l'hélice d'ADN

3.2. Structure secondaire de l'ARN

En formant un certain nombre de courtes régions hélicoïdales extrêmement rigides reliées entre elles par des régions flexibles de simple brin, la structure de l'ARN obtenue donne lieu à un enchevêtrement de structure secondaire. Ainsi idéalement, une structure secondaire est représentée comme un assemblage de motifs élémentaires 2D liés entre eux par des régions non-appariées. Et selon la manière dont ces régions sont assemblées, on obtient des éléments topologiques variées, composés d'hélices et de différents types de boucles :

3.3. Hélice

C'est le motif élémentaire de base de la structure secondaire des ARN. Il correspond à l'ensemble des appariements Watson-Crick internes entre les bases de l'ARN, reliés par des liaisons hydrogènes (A=U et G≡C). Elle forme des régions double brin, antiparallèle, de conformation hélicoïdale qui tourne à droite (Figure 8).

3.4. Tige-boucle

Le motif en tige-boucle se forme lorsque la séquence de la molécule d'ARN se replie sur elle-même pour permettre à 2 régions complémentaires et antiparallèles de s'apparier entre-elles, laissant quelques bases non-appariées, formant la boucle. La boucle peut contenir au maximum 10 bases non-appariées et la plus petite boucle identifiée jusqu'à présent contient 2 bases non-appariées (Figure 8).

3.5. Boucle interne

La boucle interne se forme à l'intérieur d'une hélice et correspond à deux brins adjacents contenant le même nombre de nucléotides non-appariés (Figure 8).

3.6. Bulge (ou boucle latérale)

Le bulge se forme sur l'un des deux brins d'une hélice lorsqu'au moins un nucléotide de ce brin est non-apparié (Figure 8).

3.7. Jonction d'hélices

Les jonctions d'hélices qu'on appelle également boucle multiple connectent au moins trois hélices et constituent des points de branchement (Figure 8).

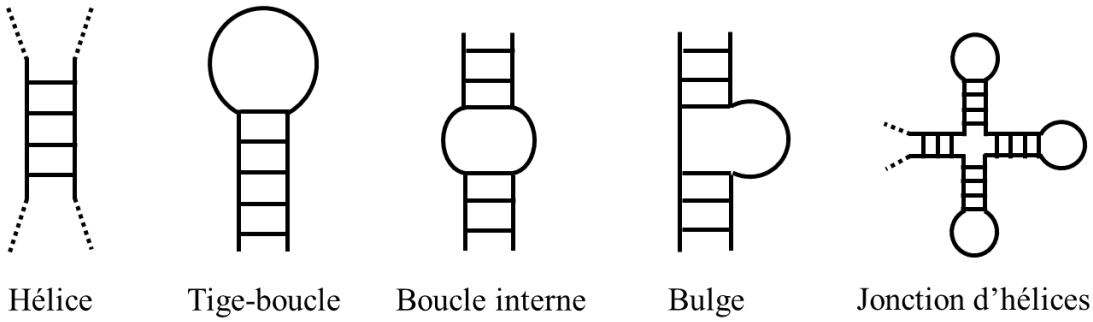


Figure 8 Motifs élémentaires de la structure secondaire

3.8. Structure tertiaire de l'ARN

Au-delà de la topologie en boucles et hélices qui composent leurs structures secondaires, les molécules d'ARN peuvent aussi se replier pour former une structure tridimensionnelle plus compacte qu'on appellera structure tertiaire de l'ARN. Leurs structures atomiques ont été identifiées grâce à des techniques de résonance magnétique nucléaire (RMN) ou de cristallographie aux rayons X. Elles ont notamment permis de montrer que les structures tertiaires de l'ARN font intervenir des interactions non-canoniques entre des motifs élémentaires de structure secondaire. On désigne par interactions non-canoniques tous les appariements qu'on différencie des appariements de bases classiques faisant intervenir des liaisons hydrogènes, entre des paires de bases Watson-Crick (A=U et G≡C) ou wobble (G=U). Leontis et Westhof [6] ont dénombré 168 modèles d'appariement de bases, impliquant toujours des liaisons hydrogènes. Voici quelques exemples de ces appariements non-canoniques :

3.9. Appariement Hoogsteen

Il s'agit d'un type d'appariement de bases pour lequel des liaisons hydrogènes se forment entre la face Watson-Crick d'une pyrimidine et la face non Watson-Crick d'une purine, qui implique en particulier les atomes d'azote et d'oxygène, portés respectivement par les carbones en positions 7 et 6 du cycle aromatique (Figure 9). En fonction de l'orientation de la pyrimidine par rapport à la purine, 2 types d'appariement Hoogsteen peuvent être observés : les paires Hoogsteen qui correspondent aux paires de bases purine-pyrimidine dont les sucres sont orientés du même côté et les paires Hoogsteen inversées qui correspondent aux paires de bases purine-pyrimidine dont les sucres sont orientés de 180° l'un par rapport à l'autre.

3.10. Interaction triple

L'interaction triple correspond à un appariement simultané de type Watson-Crick entre une purine et une pyrimidine et de type Hoogsteen entre cette même purine et une autre pyrimidine. L'appariement de type Hoogsteen utilise en effet une face de la purine indépendante de la face

Watson-Crick, ce qui permet à celle-ci d'effectuer simultanément les 2 types d'appariements. Ce type d'interaction est observé lorsqu'une région simple brin interagit avec une hélice pour former une triple hélice que l'on retrouve généralement dans une structure particulière de l'ARN appelée pseudonœud.

3.11. Interaction base-ribose

L'interaction base-ribose fait intervenir une liaison hydrogène entre le groupe hydroxyle 2'-OH d'un ribose faisant partie d'un brin d'hélice et un atome d'azote ou d'oxygène d'une base n'étant pas impliqué dans un appariement Watson-Crick (Figure 9). Par exemple l'azote en position 3 d'une purine et l'oxygène en position 2 d'une pyrimidine.

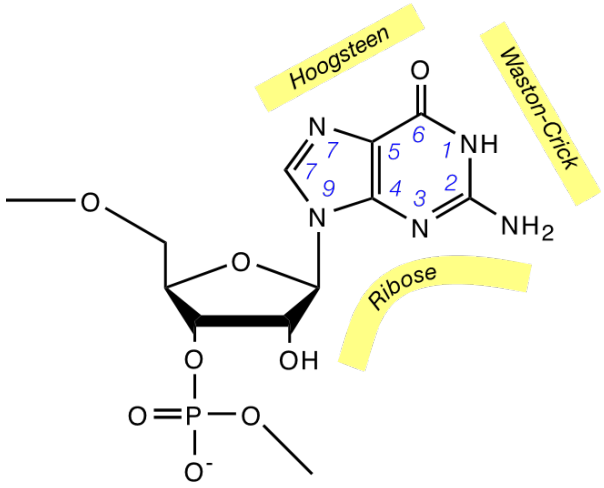


Figure 9 Les différentes faces d'appariement d'une base

3.12. Les pseudonœuds

Parmi les structures adoptées par la molécule d'ARN, on retrouve souvent les motifs en pseudonœuds. Reconnu pour la première fois dans le virus de la mosaïque du navet jaune (TYMV, Turnip Yellow Mosaic Virus) [10]. Un pseudonœud est un motif structural qui se compose au minimum de 2 segments hélicoïdaux connectés par des régions ou des boucles simples brins. Ils résultent de l'interaction d'une séquence d'ARN simple brin appartenant à une structure secondaire (ex : une boucle délimitée par une tige) avec une séquence en amont ou en aval de cette structure, dans la même molécule d'ARN, et qui est situé en dehors de cette boucle [9][11]. Les pseudonœuds permettent ainsi de lier 2 régions non-appariées l'une de l'autre. Cet appariement se fait généralement sur moins d'un tour, soit une dizaine de paires de bases et impose une rigidité tridimensionnelle à l'ensemble de la structure de l'ARN. Cette contrainte de taille est imposée par la géométrie du motif. En effet au-delà de cette limite,

l'interaction entre les 2 régions libres conduirait à un enlacement complet des brins, ce qui produirait un nœud complet au lieu d'un pseudonœud. 2 types de pseudonœud sont bien connus : les pseudonœuds en I et les pseudonœuds de Pleij ou pseudonœuds en H.

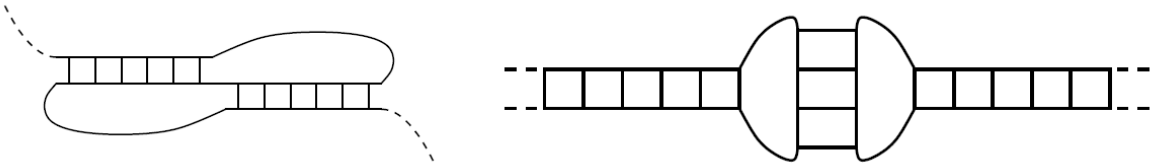


Figure 10 A gauche pseudonœud en H et à droite pseudonœud en I.

En résumé, les molécules d'ARN synthétisées sous forme de simple brin sont capables de se replier sur elles-mêmes et former des structures intramoléculaires stabilisées par des liaisons hydrogènes entre paires de bases complémentaires, notamment des paires de bases Watson-Crick A=U et G≡C et quelque fois G=U. Ces structures intramoléculaires sont généralement courtes et espacées entre-elles par des régions simples brins. Elles constituent la structure secondaire de l'ARN. Ces structures secondaires peuvent ensuite s'associer entre elles pour adopter un arrangement topologique et une géométrie spécifique. Le rapprochement de ces séquences complémentaires et de ces portions de structures secondaires dépend de la composition physicochimique du milieu, en particulier la présence de cations monovalents comme le potassium (K^+) ou divalents comme le magnésium (Mg^{2+}) et le sodium (Na^{2+}). Les cations sont essentiels pour la stabilisation thermodynamique des structures secondaires et tertiaires de l'ARN. Ces ions chargés positivement peuvent interagir avec les groupes phosphoryles chargés négativement, portés par le squelette phosphodiester. Plus précisément les cations neutralisent les répulsions électrostatiques entre les charges négatives des groupes phosphates et favorisent la stabilisation des structures secondaires et tertiaires (Figure 11).

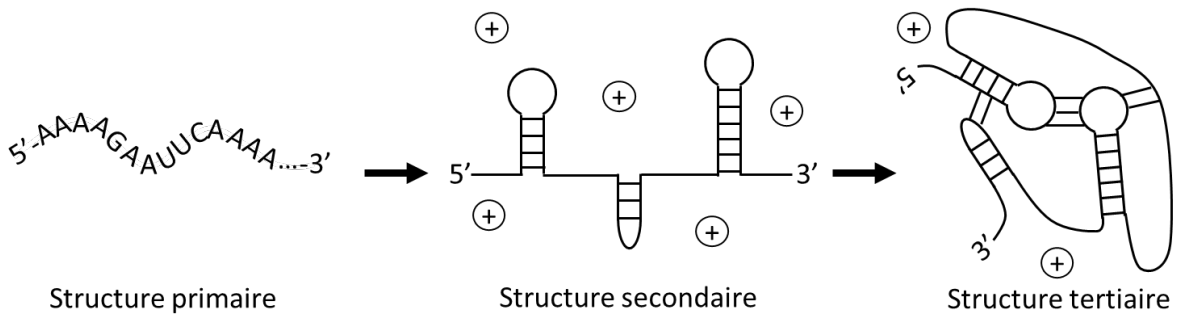


Figure 11 Les différents niveaux d'organisation de l'ARN.⊕ représentent les cations.

Comme nous avons pu le voir, la molécule d'ARN est issue de l'assemblage de 4 résidus A, U, G et C qui conduit à la formation de séquence nucléotidique de longueur variable et qui contient les informations qui définissent l'ensemble des appariements intramoléculaires possibles dans la molécule. Cette ossature s'organise ensuite en une structure tridimensionnelle qui exprime une fonction spécifique. Il existe une très grande variabilité des ces éléments structuraux expliquant la grande variabilité de fonction observée chez les ARN.

Chapitre 2 : Prédiction des structures secondaires des ARN

Les fonctions biologiques des molécules d'ARN sont très diverses et ces différentes fonctions sont influencées par les structures secondaires et tertiaires qu'elles adoptent. Les structures secondaires sont par ailleurs cruciales, soit parce qu'elles accomplissent elles-mêmes la fonction biologique de la molécule d'ARN, soit parce qu'elles conditionnent la formation de leur structure tertiaire qui donne sa fonction à la molécule d'ARN [12].

Connaitre les structures secondaires qu'adopte une molécule d'ARN est donc essentiel pour comprendre les fonctions de cette molécule. Cependant, il est parfois difficile de les déterminer expérimentalement ; c'est pourquoi plusieurs méthodes ont été développées pour prédire la structure secondaire des ARN.

1. Méthode comparative

La première méthode repose sur la comparaison des séquences d'ARN ayant la même fonction dans différentes espèces. Cette comparaison a pour but d'identifier l'importance d'une structure secondaire malgré une légère différence dans la séquence tout en respectant les règles d'appariements de base [13][14].

2. Algorithme combinatoire

Cette méthode [15] consiste à faire la liste complète de toutes les hélices qui peuvent se former dans une séquence, et ensuite de former toutes les structures secondaires potentielles en combinant toutes les hélices compatibles possibles. Chaque structure générée est ensuite évaluée en calculant son énergie libre totale. Cet algorithme suppose que les structures statistiquement sélectionnées sont celles constituées des hélices les plus stables compatibles. Cependant, le nombre de structures possibles augmente exponentiellement avec la longueur de la séquence. Par conséquent, cette méthode n'est compatible que pour des séquences

relativement courtes comme les ARN transferts et devient totalement inefficace pour les séquences beaucoup plus longues.

3. Algorithme génétique

Il s'agit d'un algorithme non-déterministe qui n'explore pas toute la combinatoire des repliements possibles mais teste par une marche aléatoire dans l'espace des séquences différentes solutions. Celles-ci sont choisies et générées par une méthode qui mime les différentes étapes de manipulations génétiques au sein d'une population lors de l'évolution, c'est pourquoi on parle d'algorithme « génétique ».

L'algorithme renvoie une collection de solutions possibles dont les meilleures seront celles qui forment des structures ayant une faible valeur d'énergie libre. Ainsi il est attendu qu'à chaque étape l'algorithme trouve progressivement des structures de plus en plus stables jusqu'à atteindre la solution optimale : la structure la plus stable.

L'exécution de l'algorithme se fait de façon cyclique et fait intervenir 3 sous-étapes qui sont répétées à chaque cycle :

3.1. La mutation

Elle consiste, de façon aléatoire, à l'échange d'une hélice par une autre, dans une solution donnée. Chaque nouvelle structure va être conservée (ou rejetée) avec une probabilité d'autant plus forte que sa stabilité est meilleure ou moins bonne comparée à la structure précédente. Si cet échange est bénéfique, alors la nouvelle structure a une forte probabilité d'être conservée pour le cycle suivant. La possibilité d'accepter des mutations défavorables permet d'éviter que l'algorithme soit piégé dans des minimums locaux. Autrement dit, si l'algorithme est coincé dans une solution localement favorable, les mutations permettront de rechercher continuellement d'autres structures potentiellement meilleures.

3.2. La recombinaison (croisement)

Deux solutions choisies aléatoirement s'échangent des parties de leurs structures pour créer une nouvelle solution. L'idée est que chaque solution possède des parties favorables et d'autres défavorables. Si un croisement se fait entre ces deux solutions, alors il est possible d'incorporer toutes les parties favorables dans une seule et même solution, et les parties défavorables dans une autre.

3.3. La sélection

Cette étape permet de sélectionner les meilleures solutions parmi les solutions obtenues précédemment. Le choix se fait en classant chaque solution selon un score qui dépend d'un

critère de sélection et les solutions sont conservées ou non pour le cycle suivant selon leur rang ou selon leur score, de façon déterministe ou probabiliste.

Cette méthode est utilisée dans différents algorithmes de prédiction des structures secondaires de l'ARN [16][17][18].

4. Algorithmes de programmation dynamique

L'algorithme de programmation dynamique[1] permet de faire l'énumération exhaustive de toutes les structures possibles sans pseudonœuds pour une séquence d'ARN de longueur N, avec un temps de calcul d'ordre N^3. Cette méthode consiste en un algorithme récursif qui décompose le calcul pour une longue séquence en un calcul pour un ensemble de séquences plus courtes. Autrement dit, une séquence peut être considérée comme une collection de sous-séquences ; ainsi, si on peut obtenir les structures optimales de toutes les sous-séquences, alors il est également possible de déterminer les structures optimales globales [19].

5. Algorithme cinétique

Plusieurs études expérimentales sur le repliement de l'intron de l'ARN ribosomique de *Tetrahymena* [20][21][22] ont permis de démontrer que la molécule d'ARN se replie à travers une succession d'évènements stochastiques d'ouverture et de fermeture d'hélices individuelles qui conduisent à la formation de structures secondaires intermédiaires en quasi-équilibres.

Se basant sur le fait que les hélices formées peuvent être très stables et prennent un temps assez long pour se défaire, l'algorithme cinétique consiste à simuler la dynamique du processus de repliement en permettant l'ajout ou la suppression d'hélice au fur et à mesure que l'ARN est transcrit, tout en favorisant la minimisation de l'énergie libre totale de la molécule. Ainsi le programme calcule toutes les transitions possibles vers les structures proches-voisines et tire aléatoirement la structure vers laquelle transiter.

Contrairement aux méthodes décrites précédemment, l'algorithme cinétique permet d'intégrer les pseudonœuds à la structure prédite. En contrepartie, il n'explore pas de façon exhaustive toutes les structures secondaires possibles : il explore de façon stochastique et de proche en proche des structures intermédiaires et leurs structures voisines.

[1] Contrairement à ce que le terme dynamique peut insinuer, ces algorithmes sont essentiellement utilisés pour le calcul des propriétés d'équilibre des molécules et ils n'ont donc pas de rapport avec la dynamique moléculaire ou la cinétique de repliement.

Différents programmes ont été développés pour simuler la dynamique de repliement des molécules d'ARN de façon recuit simulé ou co-transcriptionnelle [23][24][25][26]. Dans le cadre de cette thèse j'utiliserai le programme Kinefold développé par Isambert H. et son équipe [27] qui permet de simuler le repliement co-transcriptionnel.

6. Kinefold

Le serveur Kinefold (http://kinefold.curie.fr/) fournit une interface pour des simulations stochastiques du repliement de l'ADN et de l'ARN basées sur un algorithme de Monte Carlo cinétique introduit par Isambert et Siggia [28], puis amélioré par Xayaphoummine *et al.* [29][27]. Cette partie reprend une partie du rapport de Master de Valerio Piomponi [30] et décrit succinctement les principes de base de l'algorithme utilisé par Kinefold et comment l'outil peut être utilisé pour prédire le repliement co-transcriptionnel de l'ARN.

Le repliement de l'ARN au cours de la transcription est connu pour procéder de façon stochastique, via une succession de structures secondaires partiellement formées et en quasi-équilibre. L'algorithme Kinefold suit un chemin de repliement particulier parmi le large ensemble des chemins de repliement contenant toutes les combinaisons d'hélices encodées dans la séquence et mutuellement compatibles à chaque instant. Au cours de la transcription et de l'élongation de la séquence, des transitions élémentaires peuvent se produire, telles que l'ouverture ou la fermeture d'une hélice (**Figure 12**) éventuellement suivies d'un allongement de l'hélice. Les hélices prises en compte par Kinefold doivent être constituées de 3 paires de bases ou plus et avoir une énergie libre d'au moins 10 k_BT, où T est la température (37 °C) et $k_B = 1{,}380\ 649 \times 10^{-23}$ J K^{-1} est la constante de Boltzmann. Les transitions sont sélectionnées selon une procédure de Monte Carlo cinétique décrite par Mironov *et al.* [31], qui suggère également une méthode pour énumérer toutes les structures secondaires qu'une séquence peut former, basée sur la construction d'une matrice des hélices compatibles.

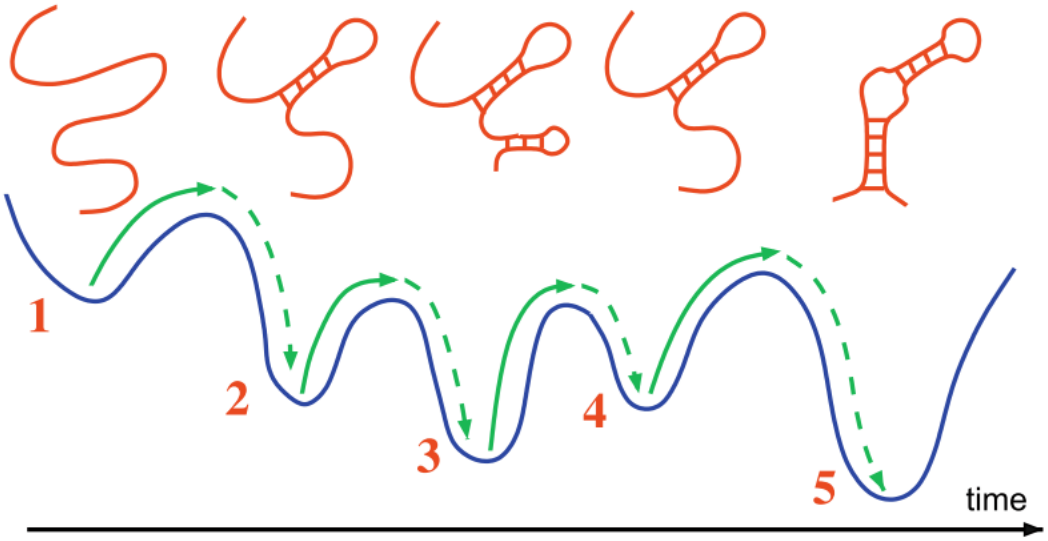


Figure 12 Chemin de repliement de l'ARN. Le repliement de la molécule évolue de façon stochastique via la formation de structures secondaires successives. Les transitions élémentaires résultent de l'ouverture ou de la fermeture d'hélices individuelles. La ligne bleue décrit comment l'énergie libre ***G*** de la structure évolue au cours du temps au fur et à mesure des transitions. Image tirée de Isambert (2009) Methods [32].

Les vitesses de transition entre deux structures sont calculées selon une formule de type Arrhenius :

$$k_+ = k° e^{-\frac{\Delta G_+}{k_B T}}$$

où k_+ est la vitesse de nucléation, $k° \sim 10^8\ \mathrm{s}^{-1}$ est une constante déterminée expérimentalement (voir ci-après), et ΔG_+ est l'énergie libre nécessaire pour franchir la barrière de l'état de transition (**Figure 13**).

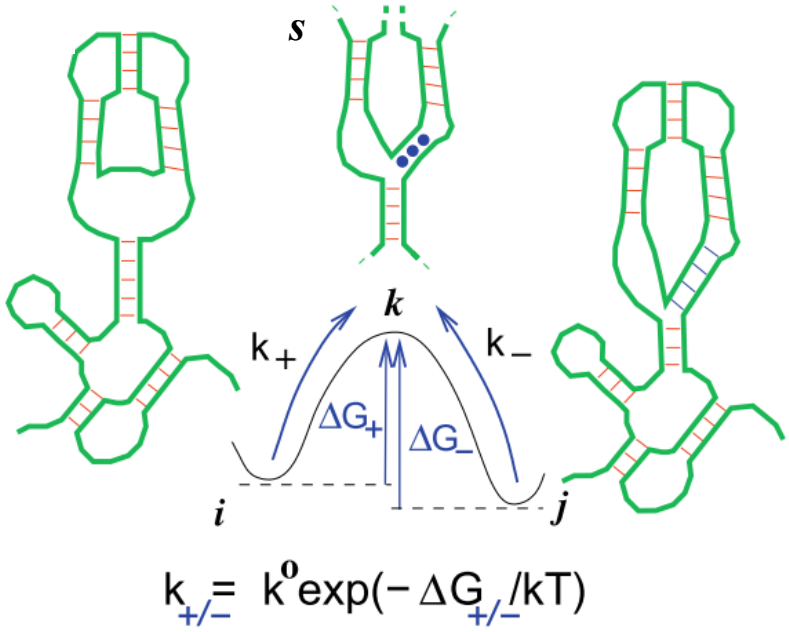


Figure 13 Énergie libre de nucléation. La vitesse de nucléation $\boldsymbol{k}_+$ est calculée selon une loi de type Arrhenius en considérant la variation d'énergie libre entre l'état initial et un état intermédiaire *s* associé au sommet de la barrière. De la même manière, la vitesse de dénaturation $\boldsymbol{k}_-$ peut être calculée à partir de $\boldsymbol{\Delta G}_-$. En général, $\boldsymbol{k}_+$ est lent comparé aux échelles de temps moléculaires typiques, mais rapide comparé à la vitesse d'élongation, de sorte que le repliement et l'élongation peuvent être considérés comme découplés. D'un autre côté, $\boldsymbol{k}_-$ peut être arbitrairement lent, conduisant à des structures métastables à temps de vie longs. Image tirée de Isambert (2009) Methods [32].

L'énergie libre G de chaque structure secondaire est calculée d'après les énergies libres des hélices individuelles et l'entropie de conformation de la structure complète, à partir des données proposées par le laboratoire de Turner [33][34]. Ce travail propose une méthode pour calculer les paramètres thermodynamiques de n'importe quelle structure secondaire ARN basée sur un modèle de plus proches-voisins (nearest-neighbors) qui suppose que la stabilité d'une paire de base donnée dépend de l'identité de paires de bases adjacentes. L'énergie libre des hélices est également calculée en tenant compte des nucléotides non-appariés aux extrémités ainsi que des nucléotides contenus dans les boucles internes et dans les boucles latérales.

Les valeurs utilisées ont été déterminées expérimentalement par des mesures de calorimétrie dans lesquelles l'échange de chaleur associée aux changements de conformation est directement mesuré, et par des expériences de spectroscopie ultraviolette, où la chaleur est utilisée pour dénaturer les molécules et où les changements de conformation sont détectés optiquement par des variations d'absorbance [35]. La justesse de l'algorithme Kinefold, comparé aux algorithmes proposés précédemment, découle de la prise en compte des effets entropiques et topologiques des pseudonœuds et des hélices enchevêtrées. Un pseudonœud

correspond à une hélice formée au sein d'une boucle. Les pseudonœuds augmentent la rigidité de l'ARN et peuvent entrainer le verrouillage topologique de certaines hélices. La **Figure 14** illustre deux topologies possibles associées à la formation d'un pseudonœud, dont une engendre une hélice enchevêtrée, topologiquement verrouillée. Les configurations en pseudonœuds sont des propriétés intrinsèquement non-locales, par conséquent, elles ne peuvent pas être représentées par de simples structures secondaires en 2D comme celles représentées sur la **Figure 12**. Kinefold est capable de prendre en compte la formation des pseudonœuds et d'évaluer leur énergie libre de façon analytique via la théorie des polymères [27].

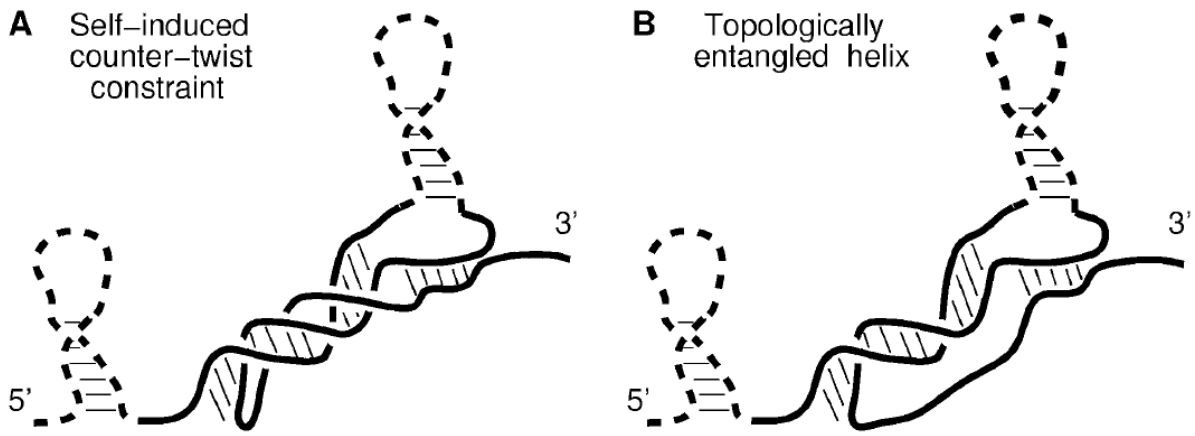


Figure 14 Contraintes topologiques et pseudonœuds. (A) Pseudonœud sans verrouillage topologique : la contre-hélicité qui évite le verrouillage topologique induit une tension du simple brin qui déstabilise entropiquement les pseudonœuds longs. **(B)** Pseudonœud avec verrouillage topologique : dans cette configuration, la dénaturation d'une hélice longue laisse un enchevêtrement topologique qui contribue à la stabilisation de l'hélice. Cette stabilisation est levée lorsque toutes les hélices qui contribuent au blocage de l'enchevêtrement sont simultanément défaites. Image tirée de Xayaphoummine *et al.* (2005) NAR [27].

Le serveur Kinefold peut simuler soit des repliements de type « renaturation » où le repliement démarre à partir d'un transcrit complet sans structures secondaires initiales, soit des repliements co-transcriptionnels. Dans ce second cas, qui est le cas qui nous intéresse, la longueur du transcrit change au cours du temps par l'addition de nucléotides en 3', un à un, à un taux d'élongation défini. En raison de cette progression, le chemin de repliement peut explorer des conformations métastables de durée de vie longue qui ne se forment pas nécessairement lors d'une renaturation.

Les prédictions de Kinefold ont été testées sur une variété de séquences courtes ou longues, et comparées à des mesures expérimentales sur des ARN [28][29][36]. De plus, les prédictions

de l'algorithme ont été directement testées via des expériences de dénaturation/renaturation mécanique en molécules uniques sur des motifs ARN [37].

6.1. Paramètres donnés en entrée

Le serveur prend tout d'abord en entrée la séquence ARN à simuler, écrite dans le sens 5' vers 3'. L'algorithme peut simuler un repliement co-transcriptionnel sur des séquences jusqu'à environs 400 nt, taille au-delà de laquelle l'espace des repliements devient trop large pour garantir l'exactitude de la simulation. Bien que Kinefold n'accepte qu'un seul brin en entrée, la simulation du repliement co-transcriptionnel d'un commutateur ARN en présence d'un antisens ARN peut être obtenue en connectant la séquence de l'antisens ARN à la séquence du commutateur par une séquence de quelques dizaines de nucléotides neutres (notés « X ») et ne participant pas au repliement.

Dans le cas d'un repliement co-transcriptionnel, un taux de transcription doit être indiqué. Pour un organisme procaryote, le temps nécessaire pour l'addition de chaque nucléotide est typiquement de 20 ms. Ensuite, le temps moléculaire de la simulation indique le temps laissé à la transcription et au processus de repliement (à ne pas confondre avec le temps nécessaire à la simulation qui peut être bien plus long). Enfin, une graine doit être donnée pour la génération de nombres pseudo-aléatoires nécessaires à la simulation de la stochasticité du processus. En conséquence, le résultat de la simulation peut varier significativement d'une simulation à l'autre, et il est essentiel de répéter la simulation plusieurs fois pour estimer la robustesse du repliement co-transcriptionnel obtenu.

6.2. Résultats obtenus en sortie de simulation

Une fois la simulation réalisée, plusieurs types de données sont disponibles. Tout d'abord, la série temporelle des structures de plus basse énergie rencontrées au cours du repliement est générée. Les structures sont représentées sous forme de diagrammes 2D, avec le temps auquel elles se sont formées et leur énergie libre. Par souci de lisibilité, les pseudonœuds, qui impliquent des régions simple-brins appariées en dehors du plan 2D, sont représentés par des lignes connectant le début et la fin de ces régions. De plus, une animation vidéo permet de visualiser l'enchainement des principales transitions au cours du repliement co-transcriptionnel. Également, un diagramme est généré indiquant l'état de formation au cours du temps des 5 hélices de plus basse énergie ainsi que l'énergie libre globale atteinte par la structure au fur et

à mesure du repliement co-transcriptionnel. L'information résumée dans ce diagramme est utile pour interpréter la stabilité des structures et hélices formées au cours de la transcription.

Chapitre 3 : Ingénierie de l'ARN pour la biologie synthétique

Un des grands défis de la biologie synthétique est le développement d'outils qui permettent de concevoir, de construire et de caractériser de manière fiable, des réseaux génétiques impliqués dans le contrôle de l'expression génique.

Les systèmes de régulation à base d'ARN offrent une simplicité et une polyvalence pour le contrôle de l'expression des gènes dans de nombreux organismes [38]. Les dispositifs synthétiques de contrôle génétiques à base d'ARN sont faciles et rapides à mettre en place car ils sont composés de seulement quatre nucléotides et ils ont une flexibilité structurelle étonnante même pour de courtes séquences. De plus, l'émergence de nouvelles méthodes de calcul de plus en plus sophistiquées facilitent la modélisation *in silico* et la prédiction du repliement des ARN.

Se basant sur ces différents avantages, de récents progrès ont été faits pour tenter de concevoir, de façon rationnelle, des régulateurs ARN entièrement synthétiques. On peut citer, la conception de ribozyme allostérique [39], la construction de fonctions de régulation génétique tels que les riboswitches qui agissent au niveau transcriptionnel [40][41] ou au niveau post-transcriptionnel [42] [43] [44].

Dans le cadre de cette thèse, je m'intéresserai plus particulièrement aux riboswitches.

1. Riboswitches naturels : localisation, structure et fonction

Les riboswitches ont été découverts pour la première fois chez les bactéries et correspondent à des séquences ARN localisées dans la région non codante de l'extrémité 5' de certains ARNm qui contrôlent l'expression du gène en aval [45] [46] [47]. L'ARN répond généralement à un stimulus externe qui correspond dans la plupart des cas à une petite molécule reconnue spécifiquement [48] [49] [50].

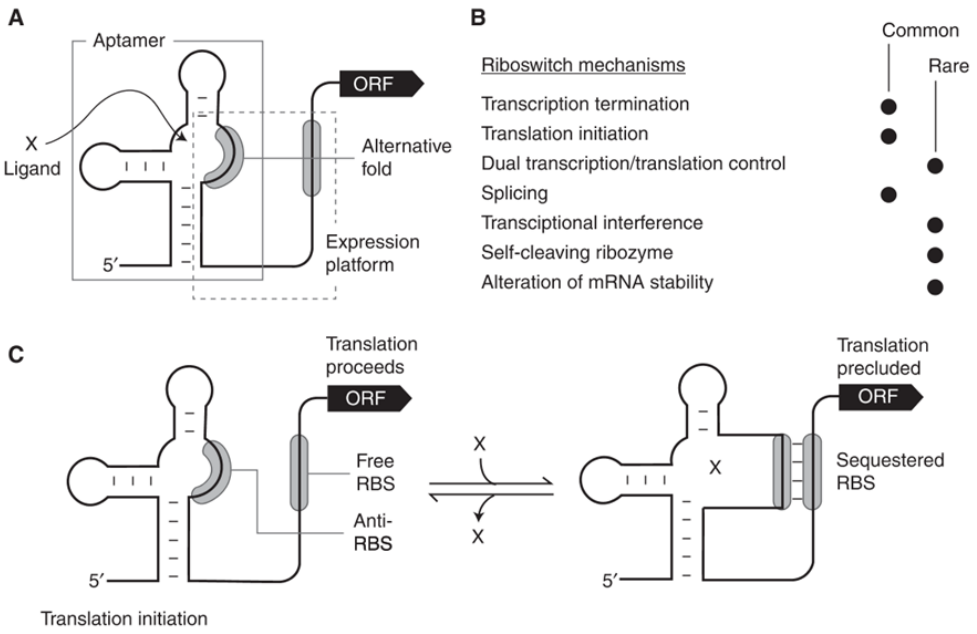


Figure 15 Fonctionnement d'un riboswitch. La séquence d'un riboswitch permet deux conformations alternatives de l'ARN et l'interaction avec une molécule signal, X (A). Dans l'exemple ici, le site de liaison du ribosome (RBS) est accessible dans une des deux conformations (C, à gauche), mais pas dans l'autre (C, à droite). La commutation entre les conformations est provoquée par une molécule X, qui peut être un métabolite, une protéine ou un autre ARN. Les processus biologiques régulés sont le plus souvent la terminaison de la transcription ou le démarrage de la traduction (B). Cependant, d'autres mécanismes restent possibles [51].

Schématiquement, les riboswitches sont constitués de deux domaines : un domaine qui lie spécifiquement le ligand, qu'on appelle aptamère et un autre domaine appelé plateforme d'expression qui agit directement sur l'expression du gène en aval, grâce à sa capacité à basculer entre deux structures secondaires différentes, en réponse à la liaison du ligand [52]. Le contrôle de l'expression du gène s'observe soit au niveau transcriptionnel, soit au niveau traductionnel. Le principe de fonctionnement d'un riboswitch est illustré à la Figure 15.

1.1. Régulation de la traduction

Pour contrôler l'expression d'un gène, les riboswitches peuvent agir au niveau de la traduction en contrôlant l'accessibilité de la séquence RBS (Ribosome Binding Site) ou séquence Shine-Dalgarno, importante pour le recrutement du ribosome et l'initiation de la traduction [53]. Un des premiers exemples décrits concerne la traduction du gène mom du bactériophage Mu. L'ARN en amont de ce gène peut adopter deux conformations alternatives dépendant de la présence ou de l'absence d'une protéine liant l'ARN, la protéine Com (Figure 16). Le thème général de ce type de régulation est de permettre au ribosome l'accès à

son site de liaison ou non (Figure 15). Ce type de régulation assure également la stœchiométrie entre ANR ribosomaux et protéines ribosomales [54].

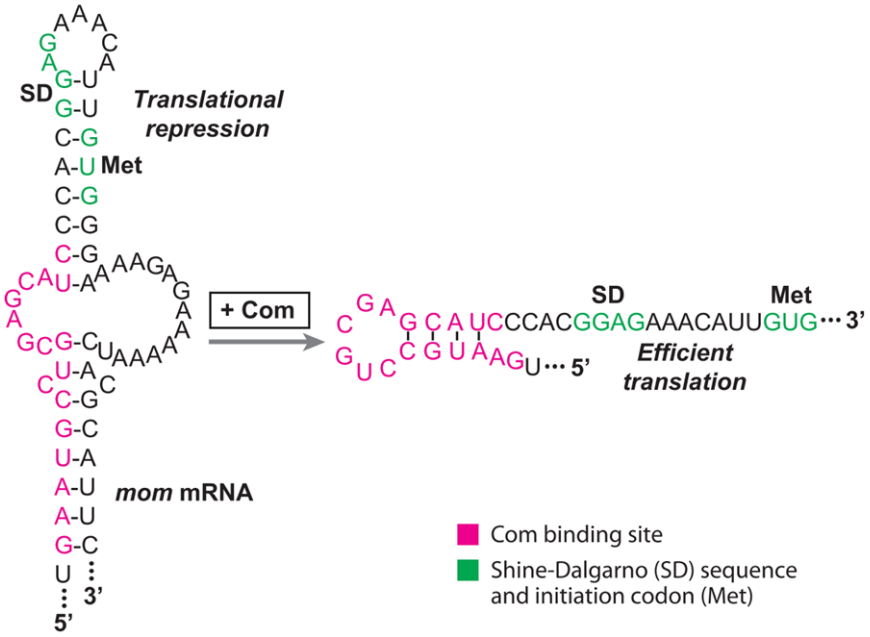


Figure 16 Contrôle de la traduction par une protéine liant l'ARN. En l'absence de la protéine Com, l'ARN en amont du gène mom adopte une conformation qui rend le site de liaison du ribosome (Shine-Dalgarno, en vert) inaccessible. En présence de la protéine, l'ARN se replie de manière à rendre la séquence SD simple brin, donc accessible au ribosome [55].

1.2. Régulation de la transcription

Dans le cadre de cette thèse, je m'intéresse plus particulièrement aux riboswitches régulant la transcription. Le principe de régulation transcriptionnelle le plus répandu (naturellement et dans les systèmes synthétiques) fait intervenir deux structures alternatives mutuellement exclusives : un « terminateur intrinsèque » qui consiste en une structure en tige-boucle, riche en G≡C, suivi d'une queue poly-U, qui arrête la transcription et d'un « anti-terminateur » qui empêche la formation du terminateur, ce qui permet la transcription de l'ARN complet [53].

1.2.1. La terminaison de la transcription

Chez *E. coli*, comme chez la plupart de bactéries, il y a deux mécanismes de terminaison de la transcription : la terminaison simple et la terminaison Rho-dépendante. Le mécanisme fondamental des deux types de terminaison est identique (Figure 17): déstabiliser le complexe d'élongation de l'ARN polymérase en déstabilisant l'hybride ADN-ARN dans la bulle de transcription [56].

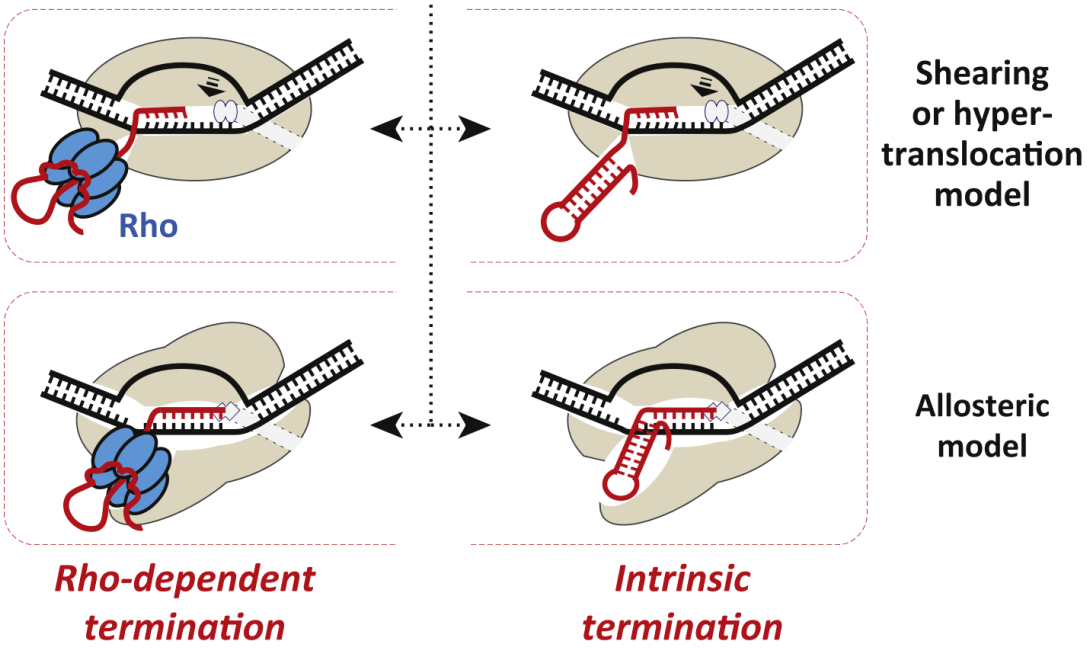


Figure 17 Mécanismes de la terminaison de la transcription. Les deux types de terminaison de la transcription (Rho-dépendante à gauche, terminaison simple ou intrinsèque à droite) partent d'un complexe d'élongation arrêté (« paused complex »), pas montré ici. Ce complexe arrêté est ensuite déstabilisé par l'action de l'hélicase Rho ou par la formation d'une tige-boucle dans l'ARN dans la terminaison simple. Différents mécanismes moléculaires peuvent expliquer la déstabilisation. Les deux classes principales invoquent un raccourcissement de l'hybride ADN-ARN (en haut) ou une action allostérique de Rho ou de la tige-boucle sur la conformation de l'ARN polymérase (en bas) [56].

L'arrêt (pause) de la transcription peut être liée à la formation d'une structure en tige-boucle dans l'ARN. Dans le cas de la terminaison simple, le complexe ternaire (ARN-polymérase, ADN, ARN) est d'avantage déstabilisé par une suite de résidus uridines en aval de la tige boucle. Les paires de bases uridine-adénine sont thermodynamiquement très instables. La règle de base pour la construction d'un terminateur simple est donc : une tige-boucle relativement stable suivie de plusieurs uridines.

1.2.2. Contrôler la formation du terminateur

Du point de vue d'ingénierie, la terminaison simple est donc probablement plus facile à mettre en œuvre que la terminaison Rho-dépendante. Nous nous focaliserons ainsi sur les systèmes synthétiques utilisant les terminateurs simples.

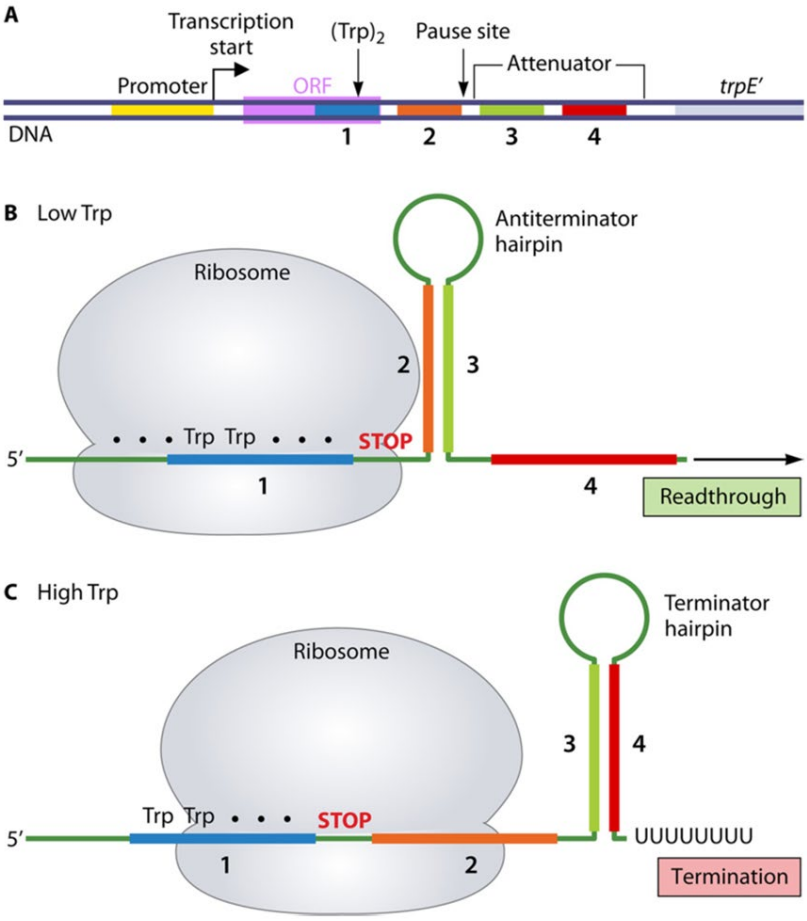


Figure 18 Atténuation de l'opéron tryptophane. L'opéron tryptophane est précédé par un court peptide « leader » qui contient deux codons trp **(A)**. Le ribosome traduit ses codons lentement quand la concentration en tryptophane dans la cellule est basse **(B)**. Le ribosome est bloqué sur la région 1 de l'ARN, ce qui favorise la formation d'une tige-boucle entre les régions 2 et 3. La région 4, également complémentaires à une partie de la région 3, ne peut pas former un terminateur simple. Lorsqu'il y a abondance de trp dans la cellule, le ribosome avance rapidement jusqu'à la séquence 2. Le seul appariement possible dans l'ARN est alors l'appariement des régions 3 et 4. Ce tige-boucle est suivi d'un répétition d'U, formant ainsi un terminateur [57].

Le principe d'ingénierie du riboswitch consiste donc à dessiner une séquence d'ARN qui peut se replier pour former, ou non, un terminateur simple. Un exemple naturel classique de formation d'un terminateur simple en fonction du contexte est l'atténuation. Ce phénomène repose sur la formation dépendant du contexte d'un terminateur entre le site de démarrage de la transcription et le cadre de lecture de la protéine régulée. La formation du terminateur simple est contrôlée par une protéine, un métabolite ou un ARN [57]. L'expression de l'opéron tryptophane, par exemple, est régulé de cette manière. La vitesse d'avancement du ribosome détermine la formation ou non d'un terminateur simple (Figure 18).

Il y a d'autres exemples dans lesquels une structure particulière d'ARN favorise ou non la formation du terminateur simple. Un des premiers cas décrits est l'opéron *tyrS* de *Bacillus*

subtilis, codant pour l'ARNt synthétase de tyrosine (Figure 19). La formation d'une structure particulière d'ARN empêche ou favorise le repliement en terminateur simple.

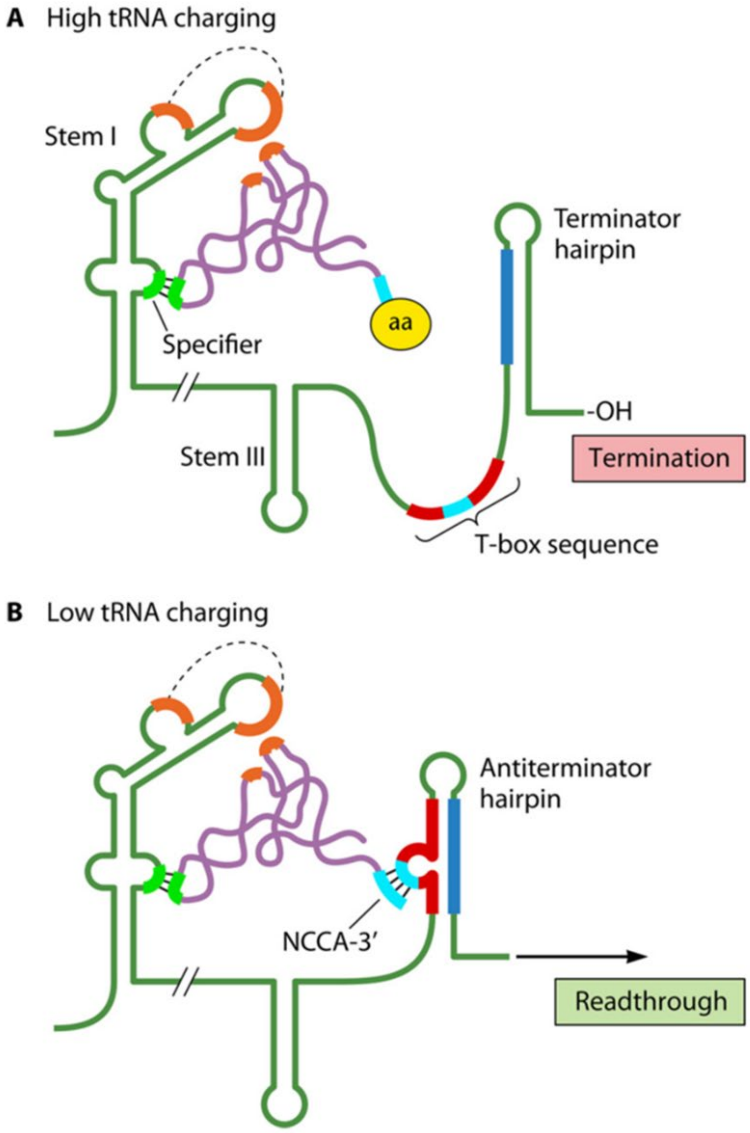


Figure 19 **(A)** Lorsque l'ARNt tyr est chargé avec un acide aminé, la séquence en 3' de cet ARNt ne peut pas interagir avec la séquence complémentaire de l'opéron tyrS (cyan). La région en bleu de l'ARN de tyrS forme un terminateur. **(B)** Lorsque l'ARNt n'est pas chargé, la séquence en 3' interagit avec une séquence appelée « boite T ». Le bulge ainsi formé crée une structure complémentaire à la séquence en bleu et empêche ainsi la formation du terminateur [57].

Dans ce travail de thèse, nous utiliserons un principe similaire en dessinant une séquence qui peut adopter deux conformations différentes en fonction de la présence ou non d'un autre ARN. C'est donc la présence (ou l'absence) de cet autre ARN qui déterminera la formation ou non du terminateur.

2. Commutateurs ARN synthétiques

Un grand nombre d'études ont été menées dans le but de construire *de novo* des riboswitches artificiels (pour une revue, voir par exemple [58]). Un focus particulier concerne la sélection et l'optimisation de la reconnaissance du ligand à détecter. Je ne décrirai pas cet aspect puisque mon travail de thèse n'aborde pas du tout ce sujet.

2.1. Mécanismes d'action des petits ARN

Chez les bactéries, en analogies avec les riboswitches naturels (voir section précédente), deux modes d'action des commutateurs ARN ont été explorés principalement : la régulation de la traduction et la régulation de la transcription via la formation de terminateurs simples. Dans mon travail de thèse, j'ai étudié ce dernier mécanisme. Plus spécifiquement, je me suis focalisé sur le changement de la conformation de l'ARN régulé par un autre petit ARN exprimé *en trans*. Le principe de ce type de commutateur ARN est illustré à la **Figure 20**.

La construction de ces commutateurs ARN est inspirée par les petits ARN naturels. Une majorité des petits ARN régulent la traduction ou la stabilité de l'ARN messager, ou servent de ligand capable de séquestrer des protéines régulatrices [59]. Il existe également des petits ARN qui modifient la terminaison de la transcription. Par exemple, le petit ARN RnaG de *Shigella flexneri* s'apparie à l'ARN messager du gène *icsA* (codant pour une protéine nécessaire à la pathogénicité de la bactérie), conduisant à la formation d'un terminateur de la transcription très similaire au schéma de la **Figure 20A** [60]. Le mécanisme de formation d'un terminateur en présence de RnaG a été démontré *in vitro* et *in vivo*. Les auteurs émettent l'hypothèse que l'action de RnaG *in vivo* ne requiert pas de protéines auxiliaires comme Hfq ou d'autres chaperonnes ARN. Ce mécanisme, par sa simplicité, est donc un candidat attractif pour le design de commutateurs ARN synthétiques. Mon travail de thèse focalise sur le design et l'analyse de ce type de régulateurs.

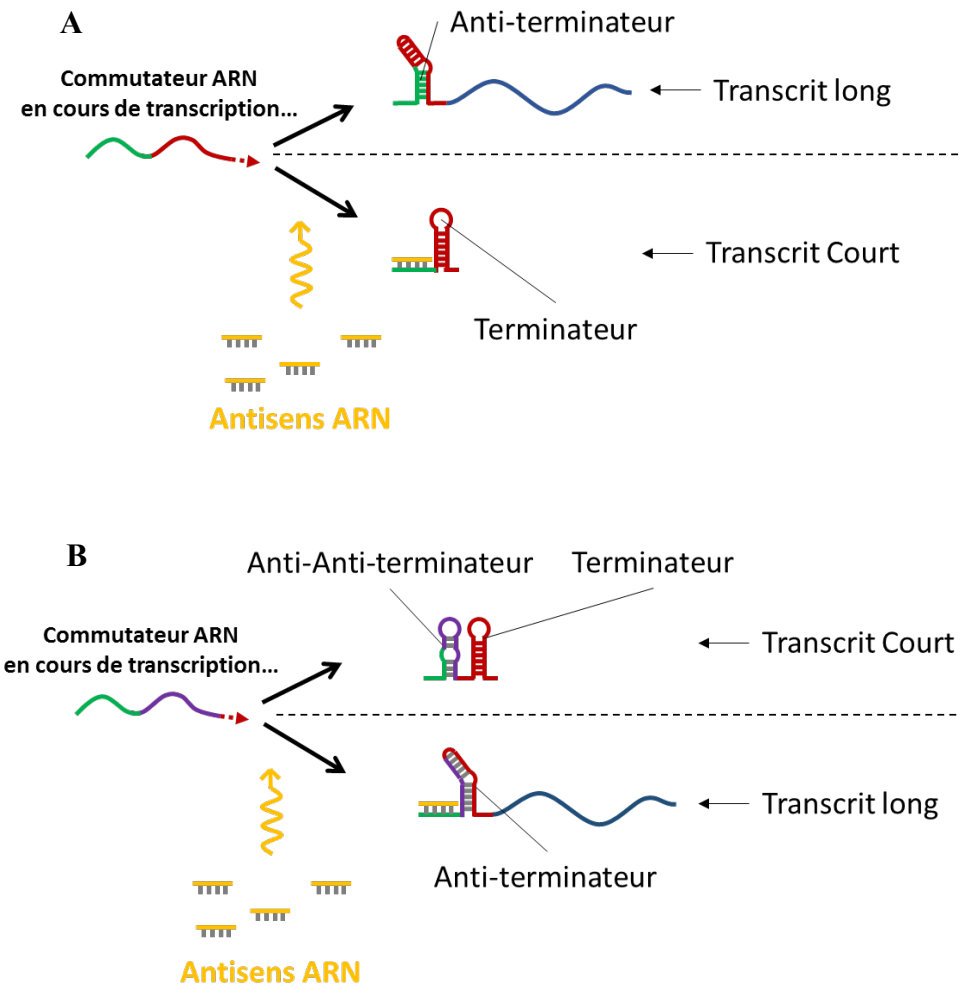


Figure 20 Schéma des repliements alternatifs adoptés par les commutateurs ARN en cours de transcription. (A) En l'absence d'un petit ARN, appelé ARN antisens, l'ARN naissant adopte une conformation qui empêche la formation d'un terminateur simple de la transcription (en haut). L'ARN antisens (en jaune), exprimé en *trans*, s'hybride à cet ARN switch « répressible ». L'hélice ainsi formée favorise une conformation alternative de l'ARN commutateur, comprenant un terminateur de la transcription. **(B)** La séquence du commutateur ARN « inductible » est telle que le terminateur simple est formé en absence de l'ARN antisens. L'expression de l'ARN antisens modifie la conformation du commutateur de manière à empêcher la formation du terminateur.

2.2. Rôle des ARN chaperonnes

Bien que l'appariement entre ARNs complémentaires soit très spécifique, lorsqu'un fragment d'ARN doit explorer un grand nombre de cibles potentielles la vitesse de l'appariement peut être lente ou de faible efficacité. Les ARN chaperonnes stabilisent le petit ARN et accélèrent l'appariement avec l'ARN cible. La chaperonne ARN majeure d'*E. coli* est la protéine Hfq [61]. Un grand nombre d'interactions entre ARN chez *E. coli* sont catalysées par Hfq et le mécanisme de l'enzyme commence à être élucidé. Afin de rester avec un système minimal, nous n'avons pas incorporé de séquences de reconnaissance de Hfq dans la

construction de nos commutateurs ARN. Un rajout de telles séquences pourrait cependant être envisagé dans la suite de ce travail de thèse.

2.3. Design et ingénierie des riboswitches

De nombreuses méthodes pour construire des commutateurs ARN ont été proposées [62]. La majorité de ces méthodes consiste à modifier des riboswitches existants, en changeant la reconnaissance du ligand, la séquence cible, etc... Peu d'études se sont concentrées sur un design *de novo* d'un système complet [63]. De plus, la plupart des riboswitches concernent la régulation de la traduction. Cependant, l'utilisation d'aptamères bien connus comme celui qui reconnaît la théophylline, a permis d'établir une méthode *in silico* suivi d'une vérification expérimentale du design, pour la conception de riboswitches régulant la terminaison de la transcription [64]. Avec ce système, les auteurs ont réussi à construire un riboswitch synthétique contrôlé par la théophylline sur une gamme dynamique de presque un facteur 3. Beaucoup de riboswitches synthétiques sont ainsi conçus à partir d'aptamères synthétiques bien connus, identifiés par une méthode appelée SELEX, comme celui qui reconnaît la théophylline [65][66].

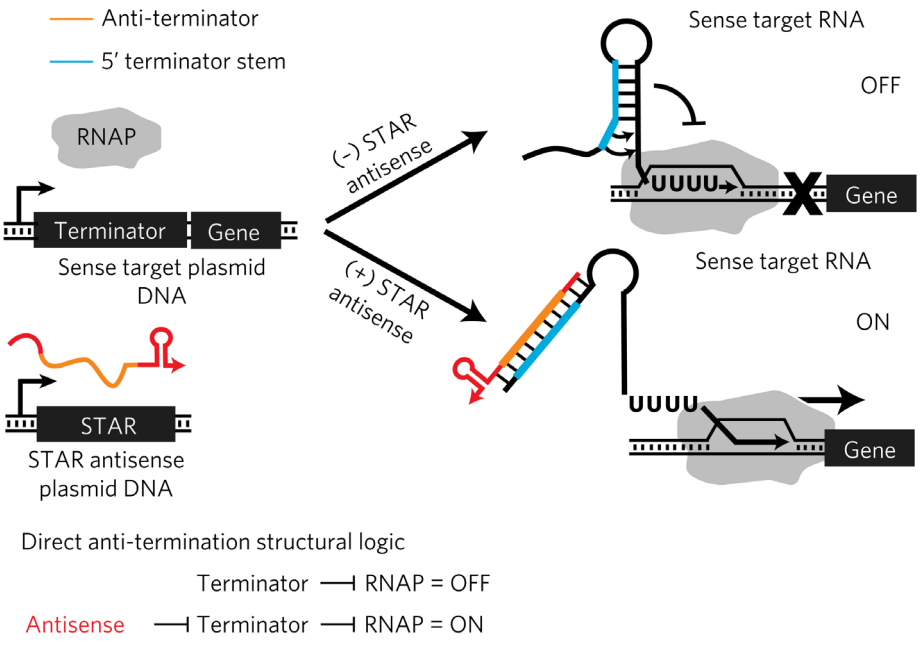


Figure 21 Design d'un commutateur ARN simple. L'ARN polymérase (ARNP) transcrit la séquence d'un terminateur simple, suivi du gène à contrôler. En l'absence du petit ARN antisens, la transcription se termine au terminateur et le gène n'est pas exprimé (en haut). En présence de l'ARN antisens (appelé STAR pour « small transcription activating RNA »), la formation de la tige-boucle du terminateur est empêchée et le gène en aval est transcrit (en bas). Figure de Chappell, J., Takahashi, M.K., and Lucks, J.B. (2015) [67].

Le système de ce type avec la meilleure performance a probablement été construite par Chappell *et al.* [67]. Cette étude s'est focalisée sur la création de commutateurs ARN activables : l'expression du petit ARN empêche la formation d'un terminateur simple et favorise ainsi la transcription du gène en aval (**Figure 21**).

Les auteurs interprètent leurs constructions et résultats avec un modèle qui incorpore trois états de l'ARN : (i) l'état initial des deux ARN repliés dans leurs structures les plus stables thermodynamiquement, (ii) l'état d'appariement parfait entre l'ARN antisens et l'ARN cible et (iii) un état intermédiaire, appelé complexe de nucléation (« seed complex »), dans lequel les interactions entre l'ARN antisens et l'ARN cible sont amorcées (avant d'atteindre l'appariement étendu de l'état (ii)). L'efficacité du régulateur est déterminée par la différence en stabilité thermodynamique (ΔG) du complexe initial (i) et du complexe intermédiaire (iii). Bien que non-testé, le système ne semble pas avoir besoin de la chaperonne Hfq pour fonctionner *in vivo*.

Dans mon travail de thèse, j'essaie de généraliser cette approche :

- en considérant des structures du commutateur plus complexes: activables et répressibles,
- en modélisant le repliement co-transcriptionnel du commutateur ARN en l'absence ou en présence de l'ARN antisens.

Au-delà de pouvoir créer des commutateurs ARN à façon, travailler sur des commutateurs ARN contrôlables par ARN antisens et commutables en cours de transcription permet plus fondamentalement d'étudier notre compréhension du lien séquence-repliement-fonction pour les ARN, c'est-à-dire d'essayer de comprendre comment l'information de repliement (et donc l'information sur la fonction) est encodée dans la séquence d'un ARN, et ce, dans un contexte biologiquement pertinent de repliement co-transcriptionnel.

Chapitre 4 : Objectifs de la thèse

L'objectif de cette thèse est de développer une plateforme pour tester *in vivo* le fonctionnement de commutateurs ARN, et de valider expérimentalement le fonctionnement des commutateurs ARN synthétiques, générés grâce aux méthodes développées par Alexandre Dawid *et al.* : méthode de design manuel [2] et méthode de design semi-automatique (en attente de publication). Nous réalisons ces validations expérimentales sur deux niveaux : *in vitro* et *in vivo*.

Pour aborder ces différents travaux, cette thèse s'articule autour de trois grandes parties :

- **Validation *in vitro* du fonctionnement de commutateurs ARN, générés par méthode de design semi-automatique.** Faisant suite aux travaux d'Alexandre Dawid *et al.* [2] sur la validation *in vitro* du fonctionnement de commutateurs ARN, générés par méthode de design manuel, nous aborderons dans cette partie la validation *in vitro* du fonctionnement de commutateurs ARN, générés par cette méthode.
- **Développement d'une plateforme de test et d'une méthode de mesure *in vivo* du fonctionnement des commutateurs ARN.** Nous avons choisi de réaliser la validation *in vivo* dans la bactérie *Escherichia coli* car c'est un organisme model qu'on peut facilement manipuler et modifier génétiquement. Pour rappel, le fonctionnement des commutateurs ARN que nous synthétisons repose sur la propriété des ARN à pouvoir se replier de manière co-transcriptionnelle. Or, expérimentalement il est encore difficile de suivre directement ce processus de repliement de l'ARN. C'est pourquoi nous utilisons un moyen indirect, qui consiste à placer le commutateur ARN en amont d'un gène rapporteur fluorescent dont l'expression sera conditionnée par le repliement adopté par le commutateur ARN utilisé. En mesurant le niveau d'expression du gène rapporteur fluorescent, nous pouvons en déduire l'état de repliement du commutateur ARN et valider son fonctionnement. Le développement de cette plateforme implique le clonage du système rapporteur ainsi que le système d'expression des antisens ARN sur des plasmides. La première version de la plateforme est ensuite optimisée par l'amélioration des plasmides concomitamment à la mise au point de protocoles expérimentaux et de procédures d'analyse pour la mesure du niveau d'expression du gène rapporteur.
- **Validation *in vivo* du fonctionnement des commutateurs ARN.** Bien que les expériences *in vitro* donnent un premier aperçu du fonctionnement des commutateurs ARN synthétisés, le but est de confirmer leur fonctionnement dans une cellule vivante. Nous décrirons dans cette partie la validation *in vivo* du fonctionnement d'un commutateur ARN répressible conçu par la méthode de design manuel, et la validation *in vivo* du fonctionnement d'un commutateur ARN inductible conçu par la méthode de design semi-automatique.

Les travaux de cette thèse vont permettre de disposer d'un ensemble d'outils pour explorer, tester et valider diverses méthodes de design de commutateurs ARN destinés à fonctionner dans les cellules vivantes.

Partie II : Validation *in vitro* du fonctionnement des commutateurs ARN, générés par méthode de design semi-automatique

Chapitre 1 : Introduction

Nous aborderons dans cette partie la validation *in vitro* du fonctionnement des commutateurs ARN générés par la méthode de design semi-automatique. Cette méthode de design a permis de générer des commutateurs ARN à partir d'un ensemble de règles heuristiques sur des motifs de séquence [68] (manuscrit en préparation). Elle a permis de générer deux types de séquences ARN commutatrices, toutes deux contrôlées par de petits antisens ARN effecteurs : des commutateurs ARN répressibles et des commutateurs ARN inductibles. Le fonctionnement de ces deux types de commutateur repose sur leur capacité à former deux repliements alternatifs, mutuellement exclusifs, qui se forment l'un en absence et l'autre en présence de l'antisens ARN effecteur spécifique du commutateur ARN. Les commutateurs ARN répressibles se replient pour former un anti-terminateur ou un terminateur, respectivement en absence et en présence d'antisens ARN (Figure 20A), tandis que les commutateurs ARN inductibles se replient pour former un terminateur ou un anti-terminateur, respectivement en absence et en présence d'antisens ARN (Figure 20B). Le terminateur est une structure de l'ARN qui arrête la transcription et conduit dans notre cas à la production de transcrits courts. L'anti-terminateur empêche la formation de ce terminateur qui conduit à la poursuite de la transcription et, par conséquent, à la production de transcrits longs. Le signal que nous détectons dans ces expériences *in vitro* est donc la longueur du transcrit généré.

Pour nos expériences de validation *in vitro* nous utilisons pour les antisens ARN des oligonucléotides d'ARN d'environ 20 nucléotides de long, comprenant une région complémentaires à leur région cible sur le commutateur ARN, encadrée de chaque côté par deux nucléotides non impliqués dans la complémentarité mais contribuant potentiellement à l'interaction avec la région cible de par leur contribution à l'énergie d'empilement des bases [69][33].

Expérimentalement, la validation *in vitro* des commutateurs consiste dans un premier temps à transcrire *in vitro*, en présence ou non de l'antisens ARN effecteur, un substrat ADN qui code pour le commutateur ARN. Dans un second temps, les tailles et quantités relatives des transcrits obtenus pour chaque condition sont analysés par PAGE (polyacrylamide gel electrophoresis).

L'effet de la présence de l'antisens ARN effecteur sur la longueur des transcrits obtenus permet de valider le fonctionnement de chaque commutateur testé : pour un commutateur ARN répressible nous nous attendons à une augmentation de transcrits courts en présence d'antisens ARN effecteur spécifique, pour un commutateur ARN inductible nous devrions observer davantage de transcrits longs en présence d'antisens ARN effecteur spécifique [2]. Le paramètre mesuré pour évaluer le fonctionnement du commutateur sous l'effet de l'antisens ARN effecteur est l'efficacité de terminaison (TE), correspondant au taux de terminaison au niveau du commutateur et décrit plus en détails plus loin.

Le fonctionnement d'un commutateur ARN repose donc sur la capacité de l'antisens ARN effecteur à altérer l'efficacité de terminaison (TE), au niveau du commutateur entre les conditions « induit » et « réprimé ». Dans la suite du manuscrit, je décrirai le principe des méthodes utilisées dans les chapitres correspondants, et le détail des matériels et méthodes pourra être trouvé en annexes.

Chapitre 2 : Principe de mesure *in vitro* du fonctionnement des commutateurs ARN

1. La transcription *in vitro*

La transcription *in vitro* est une technique couramment utilisée pour l'étude des ARN. Elle permet la synthèse *in vitro* de molécules d'ARN dans des conditions contrôlées. Les concentrations de la matrice ADN, de l'ARN polymérase, des ribonucléosides triphosphates (rNTP) et de l'antisens ARN effecteur sont connues. Pour plus de détails sur le protocole voir **Annexe B : Méthodes expérimentales, 2.1**. Les matrices d'ADN que nous utilisons sont des produits de PCR[2], obtenus après amplification d'une séquence d'ADN initialement clonée dans un plasmide circulaire. Les produits de PCR ont une taille comprise entre 400 et 600 pb selon le commutateur. Ces séquences d'ADN sont constituées d'un promoteur fort P_{T7A1} [70] (promoteur P_{A1} du phage T7) qui initie la transcription, suivi de la séquence codant pour le commutateur (répressible ou inductible), d'une séquence supplémentaire d'environ 100 pb qui joue le rôle de gène dont on régule l'expression, et d'un terminateur final constitutif qui termine la transcription de la matrice PCR (**Figure 22**).

[2] PCR : Polymerase Chain Reaction ou amplification en chaîne par polymérase, est une technique qui permet de répliquer rapidement *in vitro* une séquence d'ADN spécifique, de longueur bien précise, en de très nombreux exemplaires.

La transcription débute au +1 du promoteur, en amont de la séquence du commutateur. En fonction des conditions et du type de repliement adopté par le commutateur ARN. La transcription peut aboutir à la synthèse de trois types de transcrits (Figure 22) : (i) un transcrit interrompu au niveau du terminateur du commutateur (qualifié de « transcrit court »), (ii) un transcrit interrompu au niveau du terminateur final constitutif (qualifié de « transcrit long »), et (iii) en cas de fuite du terminateur final constitutif, des transcrits dits « run-off », résultants de la transcription complète de la matrice d'ADN et interrompu par la fin de cette dernière.

Les produits des transcriptions *in vitro* sont ensuite analysés par PAGE (polyacrylamide gel electrophoresis).

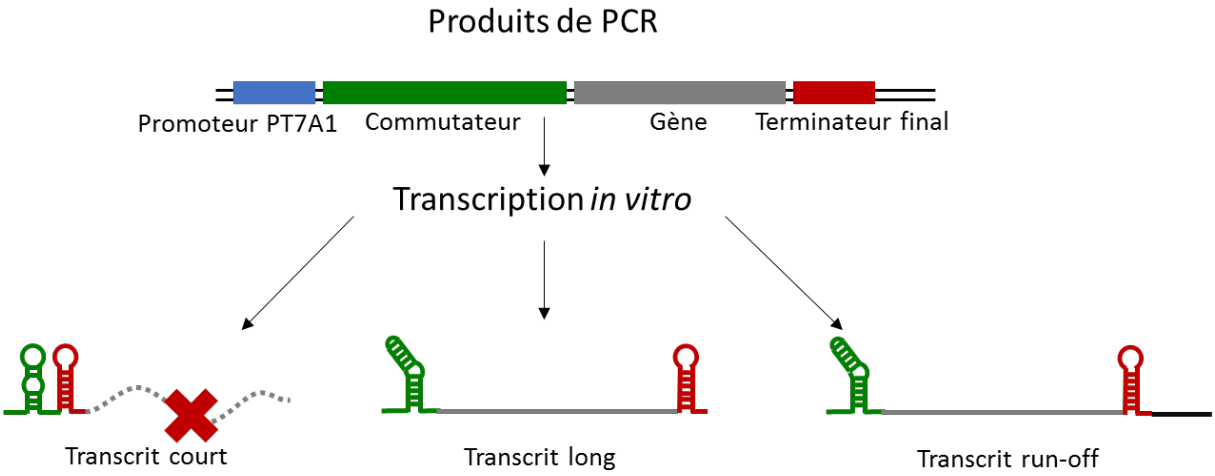


Figure 22 Schéma de la matrice d'ADN utilisée pour la transcription *in vitro* et les différents types de transcrit que l'on peut obtenir. Si le commutateur forme un terminateur (à gauche), on obtient un transcrit court. Si le commutateur se replie sous forme d'anti-terminateur, on obtient le transcrit long (au milieu). Dans le cas de fuites du terminateur final constitutif, on obtient des transcrits run-off, encore plus long (à droite). Les séquences du commutateur sont en vert. Les terminateurs (formé dans le commutateur ou constitutif) sont en rouge.

2. Séparation des transcrits d'ARN sur PAGE dénaturant

L'électrophorèse est une technique utilisée pour séparer des fragments d'acides nucléiques en fonction de leurs tailles, à travers un gel (ici un gel de polyacrylamide), sous l'effet d'un champ électrique.

Les acides nucléiques sont des molécules polyanioniques, uniformément chargées négativement. C'est grâce à cette charge négative que lors d'une électrophorèse ils vont se déplacer sous l'effet du champ électrique, et être ainsi entraînés vers l'anode du générateur. Leur taille détermine généralement leur vitesse de migration dans le gel. Plus les molécules sont longues, plus elles sont retenues par les mailles du gel, et plus lente sera leur vitesse de migration. Cependant, les molécules d'ARN adoptent le plus souvent des structures secondaires

via des appariements de bases intramoléculaires. La présence de ces structures secondaires affecte leur état de compacité et donc leur migration dans le gel. Par conséquent, afin d'obtenir un profil de migration en fonction uniquement de la taille des molécules et non en fonction de leur état de repliement, nous associons au gel des agents dénaturants tels que l'urée et le formamide et la migration est effectuée à haute température (50 °C), de façon à défaire ces structures secondaires et à maintenir la molécule d'ARN à l'état linéaire dénaturé.

La séparation de molécules par gel d'électrophorèse fonctionne alors comme un tamis moléculaire où la mobilité électrophorétique d'une molécule d'acide nucléique dans le gel dépend de sa taille. Dans le cas de l'ARN, la masse moléculaire dépend du nombre de bases qui le constituent et donc de la taille du transcrit. La mobilité électrophorétique dépend également de la concentration et du maillage (tamisage) du gel. Il existe une relation linéaire entre la mobilité électrophorétique et le logarithme de la masse moléculaire, telle que décrite par l'équation initialement proposée par Ferguson ***(I)*** [71][72] :

$$\boldsymbol{log(M) = log(M_0) - K_R T} \qquad \textbf{(I)}$$

avec :

M = mobilité électrophorétique de la molécule dans le gel ;

M_0 = mobilité électrophorétique de la molécule en absence de milieu de tamisage, et donc fonction de la charge et de la taille de la molécule ;

K_R = coefficient de retardement dû au gel, qui dépend de la masse moléculaire, en lien avec la propriété de tamisage du gel ;

T = concentration totale du gel

Le gel de polyacrylamide est un gel réticulé, formé à partir de la polymérisation de monomères d'acrylamide qui forment des chaînes d'acrylamide et un agent de réticulation le N, N– méthylène bisacrylamide qui forme des ponts entre ces chaînes d'acrylamide. La réaction de polymérisation est initiée par la formation de radicaux libres grâce à l'ajout de persulfate d'ammonium (APS) et catalysée par la N, N, N', N'-tétraméthyléthylènediamine (TEMED). La proportion relative entre l'acrylamide et l'agent de réticulation fixe la taille des mailles dans le gel, qui fixe elle-même la vitesse de migration des molécules d'ARN dans le gel. Plus il y a d'acrylamides et de bisacrylamides, plus les mailles seront petites. Dans notre cas, nous voulons séparer des fragments d'ARN allant de 20 à 800 bases.

Pour plus de détails sur le protocole voir **Annexe B : Méthodes expérimentales, 2.2**.

3. Visualisations des transcrits d'ARN

Pour visualiser les transcrits d'ARN qui ont migrés sur le gel d'électrophorèse, nous utilisons un marqueur fluorescent appelé SYBR® Gold [73]. Il s'agit d'un agent intercalant qui s'intercale entre les bases des brins d'acides nucléiques, formant ainsi un complexe fluorescent lorsqu'exposé à la lumière UV ou à la lumière bleue. Au laboratoire, nous utilisons un transilluminateur excitant dans le bleu. Le SYBR® Gold possède en effet deux pics d'excitation λ_{ex}, un à 300 nm dans le domaine de l'UV et un autre à 495 nm dans le domaine du visible. Tandis que son pic d'émission λ_{em} se trouve à 537 nm. Nous préférons le SYBR® Gold à d'autres agents intercalants à cause de sa forte brillance (50 fois plus brillant que le bromure d'éthidium) et sa sensibilité élevée (≈ 10 fois plus sensible que le bromure d'éthidium) pour la détection des ARN [73].

En fin d'électrophorèse, le gel de polyacrylamide est baigné dans une solution de TBE contenant du SYBR® Gold pour colorer les transcrits qui y ont migré. Le gel est ensuite exposé à un transilluminateur à lumière bleue et photographié à travers un filtre qui permet d'absorber la lumière d'excitation bleue, permettant ainsi de voir clairement la fluorescence émise par le SYBR® Gold.

4. Analyse de l'image du gel d'électrophorèse

L'image du gel d'électrophorèse est analysée grâce à un logiciel de traitement d'images numériques : *Gel Analyzer* (GelAnalyzer 19.1 www.gelanalyzer.com, par Istvan Lazar Jr., PhD et Istvan Lazar Sr., PhD, CSc), qui quantifie l'intensité lumineuse de chaque pixel de l'image et permet d'analyser les gels d'électrophorèse de manière quantitative. En comparant l'intensité lumineuse totale et la mobilité électrophorétique de chaque bande par rapport à une échelle ARN constituée de bandes de tailles et de masses connues, le logiciel permet d'estimer la taille (en bases) et la quantité (en masse) des transcrits présents dans chacune des différentes bandes de l'échantillon.

Théoriquement, un commutateur ARN répressible devrait majoritairement produire des transcrits longs (et éventuellement des transcrits run-off) en absence d'antisens ARN et des transcrits courts en leur présence (**Figure 20A**). A l'inverse, un commutateur ARN inductible devrait majoritairement produire des transcrits courts en absence d'antisens ARN et des transcrits longs (et éventuellement des transcrits run-off) en leur présence (**Figure 20B**).

5. Quantification du taux de terminaison

Le taux de terminaison est une mesure de la fréquence moyenne à laquelle un arrêt de la transcription est observé au niveau d'un terminateur. Pour le terminateur d'un commutateur ARN, cette mesure reflète à la fois l'état moyen de repliement du commutateur (probabilité de se replier en formant le terminateur plutôt que l'anti-terminateur) et l'efficacité de terminaison intrinsèque du terminateur. Ainsi, si un commutateur porte un terminateur ayant une efficacité de terminaison intrinsèque de 90 % et que le commutateur se replie en formant le terminateur dans 80 % des cas, alors le taux de terminaison sera de 90 % × 80 % = 72 %. On a donc la relation :

$$\text{taux de terminaison: } \boldsymbol{r}_{\text{T-com}} = \boldsymbol{f}_{\text{T-com}} \times \boldsymbol{\eta}_{\text{T-com}} \qquad (2)$$

avec :

$f_{\text{T-com}}$ = la fréquence (probabilité) avec laquelle le terminateur se forme ;

$\eta_{\text{T-com}}$ = l'efficacité intrinsèque de terminaison du terminateur

Expérimentalement, le taux de terminaison se mesure en rapportant le nombre d'événements de transcription qui se sont arrêtés au niveau du terminateur au nombre total d'événements de transcription qui ont au moins atteint le terminateur (événements de transcription interrompue au niveau du terminateur + événements de transcription qui sont allés au-delà du terminateur). En toute rigueur, un transcrit interrompu au niveau d'un terminateur pourrait aussi être le résultat non pas d'une terminaison de la transcription (détachement de l'ARN polymérase) mais simplement d'une pause de la transcription (ARN polymérase arrêtée au niveau du terminateur mais pouvant en principe reprendre la transcription). Toutefois, en pratique, comme les réactions de transcription *in vitro* se déroulent sur plusieurs heures avec recyclage des ARN polymérases et ré-initiation de la transcription sur les brins d'ADN matrices, on considère la contribution des pauses éventuelles comme négligeables comparées aux événements de terminaison qui constituent l'essentiel des transcrits interrompus au niveau d'un terminateur.

La mesure du taux de terminaison demande donc de déterminer le *nombre (relatif)* de molécules de transcrits interrompues au niveau du terminateur du commutateur, par rapport au *nombre* de molécules de transcrits ayant cette taille ou plus. Pour chaque bande dans le gel, il faut donc convertir la masse de transcrit présent en *nombre (relatif)* de molécules de transcrits

présentes dans la bande. L'intensité de fluorescence d'une bande est proportionnelle à la masse de transcrit détecté. En quantifiant l'image du gel, nous connaissons donc toutes les variables de l'équation.

On a ainsi :

$$\text{taux de terminaison: } r_{\text{T-com}} = \frac{m_{\text{T-com}}/l_{\text{T-com}}}{m_{\text{T-com}}/l_{\text{T-com}} + m_{\text{T-final}}/l_{\text{T-final}} + m_{\text{run-off}}/l_{\text{run-off}}} \quad (3)$$

avec :

m = la masse totale des transcrits présents dans une bande ;

l = la taille de ces transcrits ;

T-com = bande correspondant au terminateur du commutateur ;

T-final = bande correspondant au terminateur constitutif final sur la matrice ;

run-off = bande correspondant aux transcrits interrompus par la fin de la matrice

Chapitre 3 : Résultats et discussions

1. Fonctionnement des commutateurs ARN inductibles

1.1. État nominal (absence d'antisens ARN)

Les commutateurs ARN inductibles, qu'on appelle ici iSw3, iSw11 et iSw13, sont d'abord testés en absence d'antisens ARN pour évaluer leur état nominal (c'est-à-dire leur état par défaut, sans antisens ARN). Théoriquement, ils devraient produire majoritairement des transcrits courts, car dans ces conditions le repliement adopté par le commutateur ARN conduit à la formation d'un terminateur qui arrête la transcription au niveau du commutateur. Cependant nous constatons que les commutateurs iSw3 et iSw11 produisent majoritairement des transcrits longs et/ou run-off, 100 % pour iSw3 et 85 % pour iSw11 (Figure 23). À l'inverse iSw13 produit majoritairement des transcrits courts, comme attendu : 82 % de transcrits courts (Figure 23). Cela indique que parmi les 3 commutateurs testés, seul iSw13 est dans l'état nominal attendu d'une terminaison efficace (état OFF).

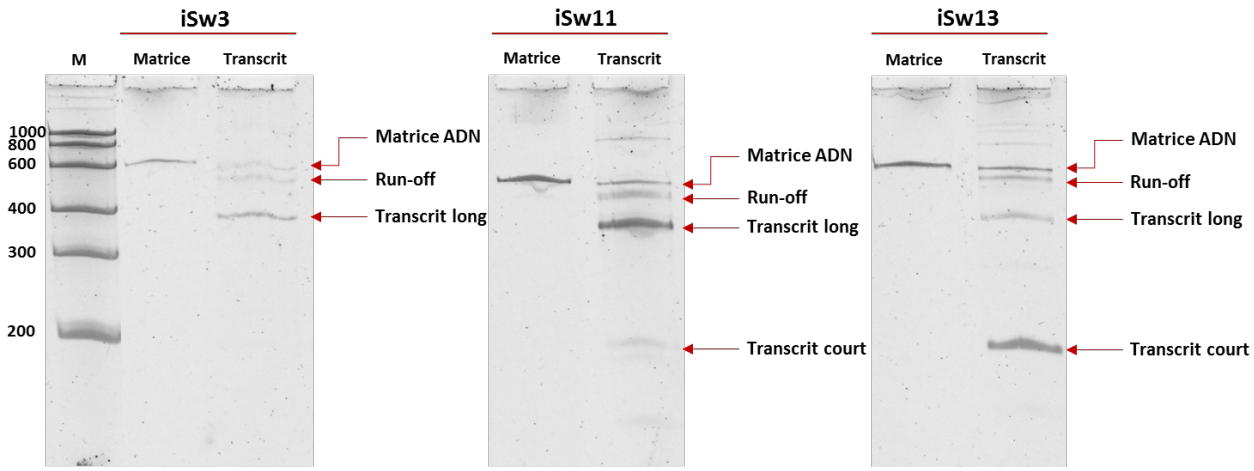


Figure 23 États nominaux des commutateurs inductibles iSw3, iSw11 et iSw13 en absence d'antisens ARN. Les pistes « matrice » correspondent à la migration des produits de PCR, utilisés comme matrices ADN pour la transcription *in vitro* et qui vont servir ici de repère pour différencier la matrice ADN des transcrits. Les pistes « transcrit » correspondent à la migration des produits de transcription *in vitro*. M ou marqueur de taille sert de référence pour identifier la taille et la masse de chaque transcrit.

Plusieurs scénarios (ainsi que les problèmes sous-jacents associés) peuvent être avancés pour expliquer pourquoi iSw3 et iSw11 ne sont pas nominalement dans l'état OFF attendu :

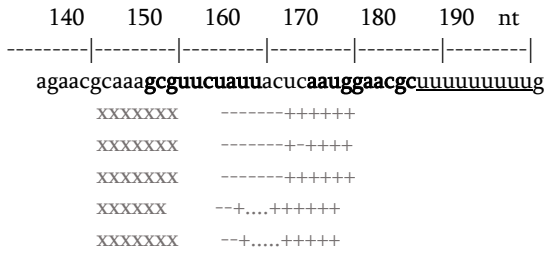
- la structure en tige-boucle du terminateur ne se forme pas au cours du repliement co-transcriptionnel (inexactitude de la simulation),
- la tige-boucle se forme mais de façon peu stable et trop furtivement pour induire une terminaison de la transcription (inexactitude de la simulation),
- la tige boucle se forme, mais partiellement ou tardivement, et ceci impacte l'efficacité de la terminaison (critère pour un bon terminateur non satisfait),
- la tige-boucle du terminateur se forme mais sa séquence ne constitue pas un bon terminateur (critère pour un bon terminateur non satisfait).

Les critères généralement admis pour constituer un terminateur de transcription sont une structure secondaire de l'ARN qui comprend une tige riche en G≡C d'environ 5 à 9 paires de bases, fermée par une courte boucle d'environ 3 à 5 nucléotides, suivie d'une queue riche en U (U-track) d'environ 7 à 9 nucléotides [74]. On pourrait penser qu'il existe une corrélation entre ces caractéristiques structurelles et l'efficacité de terminaison du terminateur. Cependant on peut trouver des terminateurs de stabilité thermodynamique égale et/ou un U-track de taille égale qui ont des efficacités de terminaison très différentes [74]. Malgré de nombreux travaux effectués sur le mécanisme de terminaison de la transcription chez *E. coli*, le mécanisme détaillé et les critères d'un bon terminateur sont encore assez mal compris. C'est pourquoi, du point de

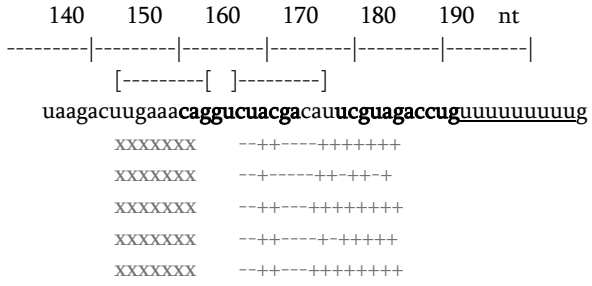
vue de la structure et de la séquence, il nous est encore difficile de trouver une explication satisfaisante pour expliquer nos résultats.

En dehors des paramètres de structure et de séquence, un autre paramètre potentiellement important pour l'efficacité de terminaison est celui de la qualité du repliement (qualité de formation) du terminateur. Pour essayer d'identifier, dans le repliement co-transcriptionnel des terminateurs, des caractéristiques pouvant expliquer les différences d'efficacité de terminaison à l'état nominal observées *in vitro* entre les différents commutateurs inductibles, nous avons analysé et comparé plus en détails les résultats de simulations de repliements co-transcriptionnels (**Tableau 1**).

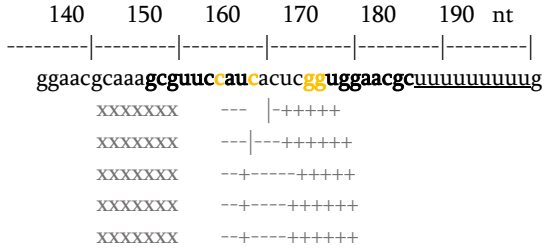
iSw1 : 0 % de terminaison *in vitro*



iSw2 : 72 % de terminaison *in vitro*



iSw3 : 0 % de terminaison *in vitro*



iSw11 : 13 % de terminaison *in vitro*

```
     140      150      160      170      180      190    nt
---------|---------|---------|---------|---------|---------|
    gcaauacaaucc**uuguauucgau**gac**guugaaugcaa**uuuuuuuuuuaag
             xxxxxx     ---+-+-++++
             xxxxxx xxxxxxxxx  xxx--        |
             xxxxxx xxxxxxxxx  xxx--        |
             xxxxxx     ---+-++-+++
             xxxxxx     ---+-++-+++
```

iSw13 : 82 % de terminaison *in vitro*

```
     140      150      160      170      180      190    nt
---------|---------|---------|---------|---------|---------|
    ccauuaccaacc**ugguaaucgag**acauu**cgauugcc**auuuuuuuuuuaag
             xxxxxx     --+-++--++
             xxxxxx     --+--+xx++
             xxxxxx     --+--+--++
             xxxxxx     --+--+--++
             xxxxxx     --+--+--++
```

Tableau 1 Analyse de la qualité du repliement co-transcriptionnel pour les terminateurs des commutateurs inductibles d'après les simulations du repliement co-transcriptionnel par Kinefold. Les commutateurs iSw1 et iSw2 ont été conçus manuellement lors d'un travail précédent [2]. iSw3 est un variant de iSw1, modifié manuellement (nucléotides en orange) dans le but d'améliorer la stabilité de la tige-boucle du terminateur. iSw11 et iSw13 ont été obtenus par conception semi-automatique. iSw13 est un variant de iSw11 (nucléotides en bleu), obtenu par l'application d'un nombre additionnel de cycles d'optimisation comparé à iSw11. Pour chaque commutateur, les derniers nucléotides de l'ARN sont représentés jusqu'au terminateur du commutateur (en gras : hélice de la tige-boucle, souligné : la plage de U). En dessous de chaque séquence sont résumés les résultats de 5 simulations de repliement du terminateur, mais effectuées sur toute la séquence du commutateur à l'aide de Kinefold [27] ; les paramètres pour les simulations sont : vitesse d'élongation : 20 ms/nt, durée simulée : 4000 ms, pseudonœuds : autorisées. Pour chaque simulation, une ligne synthétise le résultat de la simulation du repliement co-transcriptionnel.

Légende des symboles :

x : base initialement engagée dans un appariement compétitif (pouvant temporairement empêcher la formation de la tige boucle)

\+ : base s'appariant immédiatement lors de son addition au cours de l'élongation (cas idéal)

\- : base s'appariant avec retard (moment d'appariement indiqué par le symbole + ou | qui suit)

. : perte des appariements formés en aval (les appariements de la tige-boucle du terminateur précédemment initiés sont défaits)

|: symbole indiquant le moment où les appariements indiqués en amont (-) se sont formés ; | est utilisé à la place de + lorsque tous les appariements possibles ne se font pas et que l'achèvement du repliement est interrompu une ou plusieurs bases en amont de la base venant d'être ajoutée et marquée donc par le symbole |.

Les résultats des simulations de repliement co-transcriptionnel présentés dans le **Tableau 1** ne permettent pas de trancher définitivement sur le lien entre résultats des simulations et taux de terminaison mesurés *in vitro*. Toutefois, ce qui semble ressortir est que les terminateurs peu efficaces *in vitro*, semblent être associés à une tendance soit à ne pas se former du tout dans une

fraction des réplicas de simulation (iSw11 : pas formé dans 2 simulations sur 5 et 13 % de terminaison *in vitro*), soit à une initiation de formation tardive ou incomplète (iSw1 : formation initiée seulement au moment où les 2 ou 3 dernières bases de l'hélice de la tige-boucle sont ajoutées et 0 % de terminaison *in vitro*, iSw3 : initiations incomplètes (|) et complétion tardive et 0 % de terminaison *in vitro*). Les deux terminateurs présentant des efficacités de terminaison relativement bonnes *in vitro* semblent associés à une initiation de formation du terminateur précoce et à une complétion progressive du repliement au fur et à mesure de la transcription (iSw2 : 72 % de terminaison *in vitro*, iSw13 : 82 % de terminaison *in vitro*). L'ensemble de ces résultats suggère qu'un critère plus strict et plus élaboré quant à la qualité de formation des terminateurs devrait être utilisé pour la conception semi-automatique des commutateurs.

Au vu des résultats *in vitro,* nous décidons, pour la suite de nos expériences, de continuer à travailler avec le commutateur ARN inductible iSw13.

1.2. Commutation en présence d'antisens ARN

Nous testons ici le fonctionnement du commutateur iSw13 en étudiant l'effet de la présence d'une concentration croissante d'antisens ARN as13 spécifique de ce commutateur dans les mêmes conditions de transcription *in vitro* et d'électrophorèse que précédemment. Un antisens ARN non-spécifique as7 est également utilisé pour tester la spécificité de la régulation de iSw13 par l'antisens ARN as13. Les résultats sont présentés **Figure 24** pour des concentrations en as13 de 0, 1, 2, 5 et 10 µM ou pour une concentration en as7 de 10 µM.

Les résultats confirment qu'en absence d'antisens ARN as-i13, le commutateur Sw-i13 produit bien des transcrits courts (89 % de terminaison à l'état nominal). A l'inverse, lorsque le commutateur Sw-i13 est mis en présence d'une concentration croissante d'antisens ARN as-i13, nous constatons une diminution de la production de transcrits courts au profit d'une augmentation de la production de transcrits longs ou run-off : 20 % de transcrits longs + 10 % de run-off en présence de 1 µM d'antisens as-i13 et jusqu'à 87 % (60 % de transcrits longs + 27 % de run-off) pour 10 µM d'antisens as-i13. Ces résultats montrent la commutation de l'état du commutateur Sw-i13 en présence de l'antisens as-i13, c'est-à-dire le passage d'un état OFF en absence d'antisens ARN à un état ON en leur présence, avec un EC50 = 2,43 µM (EC50 = correspond à la concentration de l'antisens ARN qui induit 50 % de la réponse maximale) (**Figure 24B**). De plus, l'absence d'effet de l'antisens as7 montre que cette régulation est bien un effet spécifique dû à la séquence de l'antisens ARN as13.

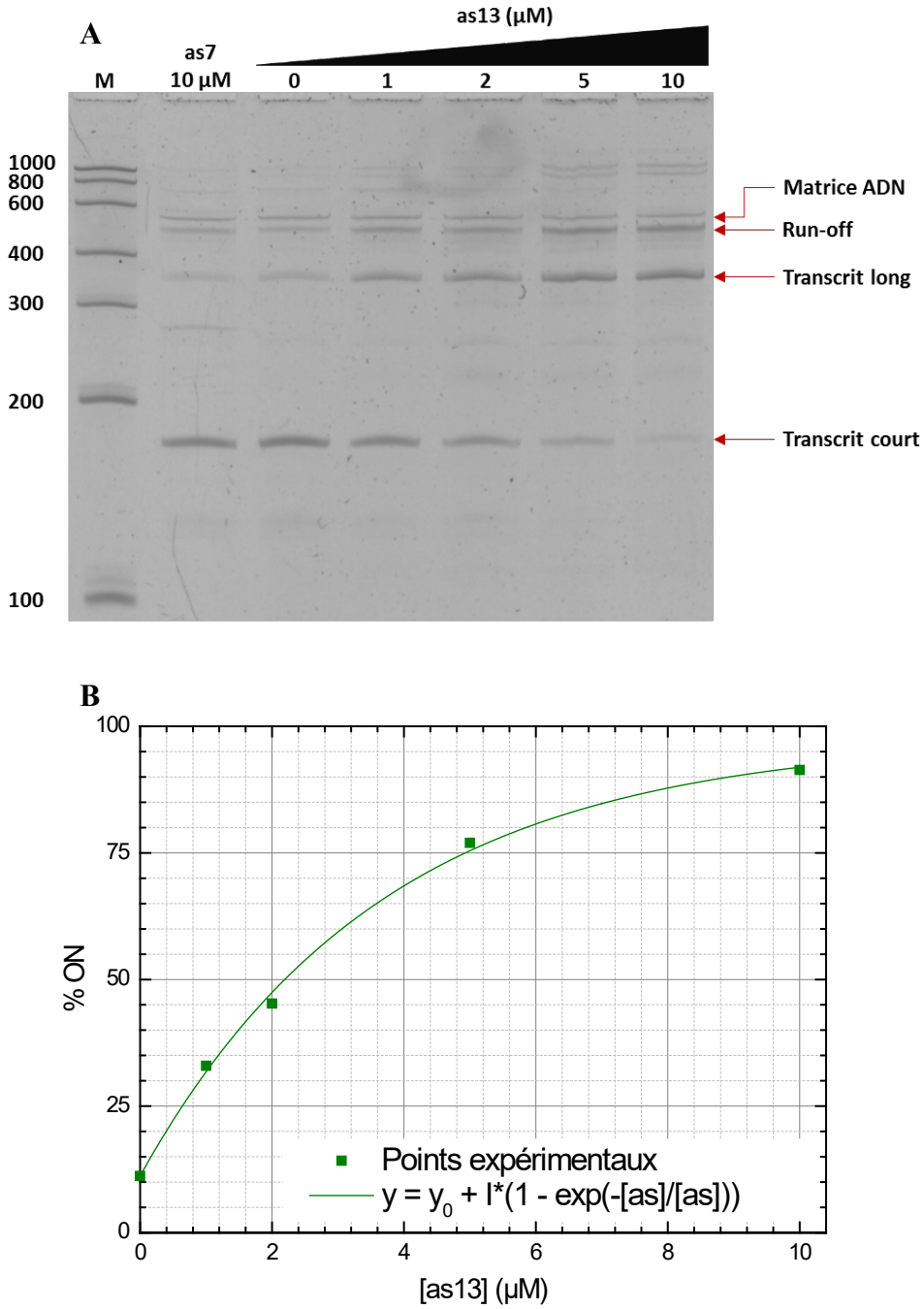


Figure 24 Fonctionnement du commutateur inductible iSw13 et induction spécifique par l'antisens ARN as13. (A) Gel d'électrophorèse montrant les tailles de transcrits obtenus par transcription *in vitro* du commutateur iSw13, en présence d'une concentration croissante de l'antisens ARN spécifique as13 ou d'un antisens ARN non-spécifique as7 (contrôle). **(B)** Niveau d'induction (% ON) obtenu avec le commutateur iSw13 en présence de différentes concentrations d'antisens as13. La courbe en trait plein est un ajustement des valeurs d'induction, construite suivant un simple modèle exponentiel, qui introduit très peu d'hypothèse sur le système que nous étudions. Les meilleurs paramètres d'ajustement sont donnés dans le **Tableau 2**. A partir de l'équation *(4)* on peut déduire la valeur de l'EC50 (EC50 = 2,43 µM) qui correspond à la concentration de l'antisens ARN qui induit 50 % de la réponse maximale, à mi-chemin entre la ligne de base et la réponse maximale.

$$y = y_0 + I \times \left(1 - \exp\left(\frac{-[as]}{[as]_0}\right)\right) \qquad (4)$$

avec :

y_0 = le pourcentage d'induction à 0 µM d'antisens ARN (ligne de base) ;

I = l'amplitude de régulation qui est égale à $Imax - y_0$, Imax étant la réponse maximale observée ;

$[as]_0$ = la concentration caractéristique de l'antisens ARN

Chi^2	4,46	
R^2	0,99	
	Valeurs	Écarts-types
y_0	11,2	1,93
I	86,42	3,42
$[as]_0$	3,68	0,40

Tableau 2 Paramètres utilisés pour l'ajustement des données d'induction avec la courbe construite à partir de l'équation (4).

L'ensemble de ces résultats montre que le commutateur iSw13 conçu de façon semi-automatique fonctionne efficacement *in vitro* (commutation de 11 % à 87 % de niveau d'induction en présence de 10 µM d'antisens ARN effecteur).

2. Test d'un commutateur ARN répressible

2.1. État nominal (absence d'antisens ARN)

Nous testons ici une seule séquence de commutateur ARN répressible conçue de façon semi-automatique : rSw7. La conception de commutateurs répressibles est soumise à moins de contraintes que la conception de commutateurs inductibles. En effet, les commutateurs inductibles sont plus complexes du fait de la présence dans leur architecture d'une structure additionnelle : un anti-anti-terminateur (Figure 20B vs. Figure 20A). De ce fait, il est beaucoup plus facile de générer de façon semi-automatique des séquences satisfaisant aux critères de sélection, et de très bons scores numériques peuvent être facilement atteints. Par ailleurs, pour la caractérisation *in vivo*, nous disposons déjà d'un commutateur répressible efficace (Dawid *et al.* 2009). C'est pourquoi nous avons décidé de ne construire et de ne tester qu'une seule séquence de commutateur répressible conçu de façon semi-automatique (Figure 25).

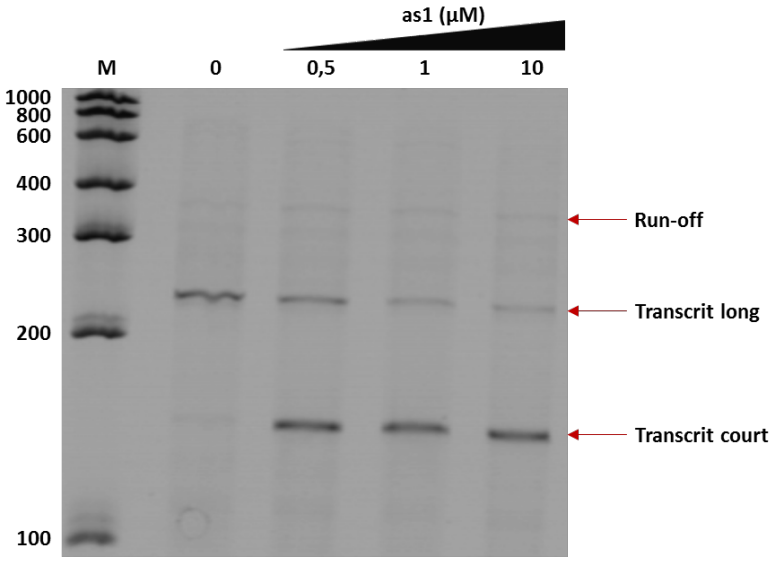


Figure 25 Fonctionnement du commutateur répressible rSw1 en présence de 0,5 ; 1 et 10 µM d'antisens ARN as1. Image du gel d'électrophorèse montrant les différents transcrits obtenus, après transcription *in vitro* du commutateur rSw1, en absence et en présence de son antisens ARN spécifique as1. Commutation de 0,6 % à 94 % du niveau de répression (production de transcrits courts) en présence de 10 µM d'antisens ARN effecteur. Ce commutateur a déjà été caractérisé précédemment [2].

Pour les commutateurs ARN répressibles, deux chemins de repliement entrent en compétition pour former un anti-terminateur ou un terminateur. En absence d'antisens ARN, ils forment un anti-terminateur qui empêche la formation du terminateur, ce qui conduit à la poursuite de la transcription et donc la production de transcrits longs. La Figure 26 illustre l'état nominal du commutateur rSw7 en conditions de transcription *in vitro*. Nous constatons que rSw7 est très majoritairement dans l'état nominal attendu (état ON à 94 %, soit 61 % de transcrits longs et 33 % de transcrits run-off) en absence d'antisens ARN as7.

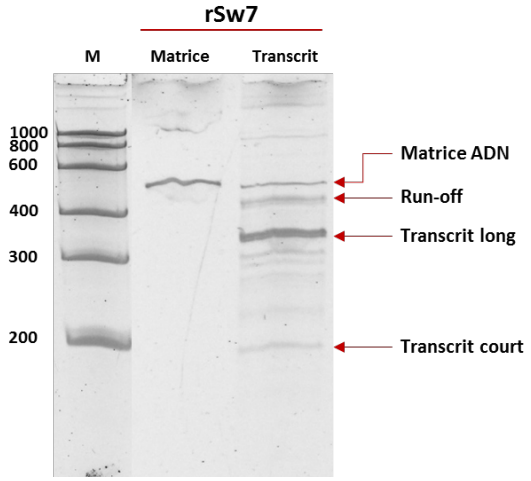


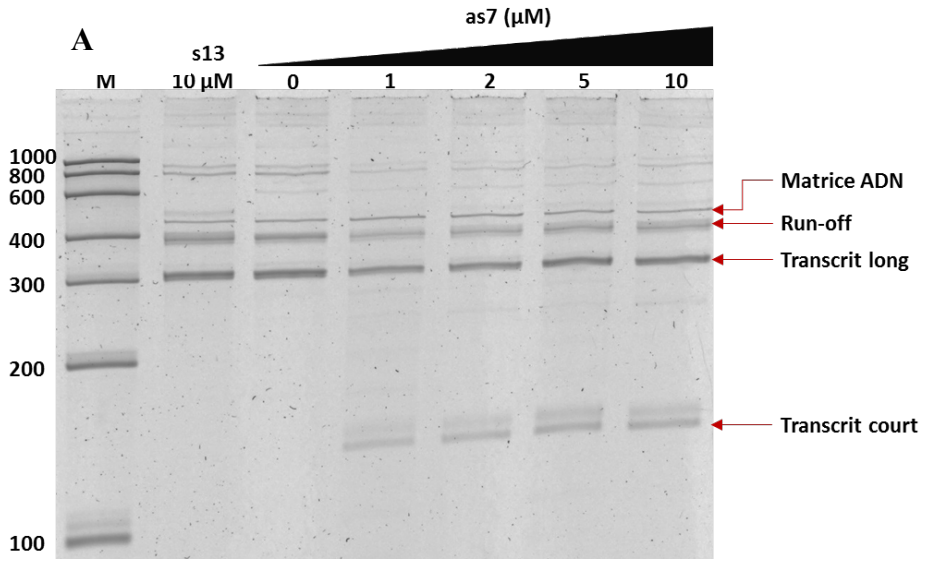
Figure 26 État nominal du commutateur rSw7 en absence d'antisens ARN as7. La piste « matrice » correspond à la migration du produit de PCR, utilisé comme matrices ADN pour la transcription *in vitro* et qui va servir ici de repère pour différencier la matrice ADN des transcrits. La piste « transcrit » correspond à la migration des produits de transcription *in vitro*. M ou marqueur de taille sert de référence pour identifier la taille et la masse de chaque transcrit.

2.2. Commutation en présence d'antisens ARN

De la même manière que pour le commutateur inductible iSw13, nous testons ici le fonctionnement du commutateur répressible rSw7 en présence d'une concentration croissante d'antisens ARN as7 : 1, 2, 5, 10 µM, spécifique de ce commutateur ; ainsi que 10 µM d'un autre antisens ARN non-spécifique : as13, pour tester la spécificité de la régulation de rSw7 par l'antisens ARN as7.

Nous confirmons qu'en absence d'antisens ARN as7, ou en présence d'antisens ARN non-spécifique as13, le commutateur rSw7 produit majoritairement des transcrits longs et run-off : état ON a 100 % (69 % de transcrits longs + 31 % de run-off) (Figure 27A).

Lorsque le commutateur rSw7 est mis en présence d'une concentration croissante d'antisens ARN as7, nous nous attendons à ce que la production de transcrits longs et run-off diminue au profit d'une augmentation de la production de transcrits courts (augmentation du taux de terminaison au niveau du commutateur), avec un effet maximal en présence de 10 µM d'antisens as7. Cependant, alors que nous observons un taux de terminaison de 30 % au niveau du commutateur en présence de 2 µM d'antisens ARN as7, cette valeur n'augmente plus lorsqu'on augmente la concentration d'antisens ARN as7 jusqu'à 10 µM (Figure 27A et Figure 27B). L'absence d'effet pour 10 µM d'antisens ARN contrôle as13 montre toutefois que l'effet de as7 est bien spécifique.



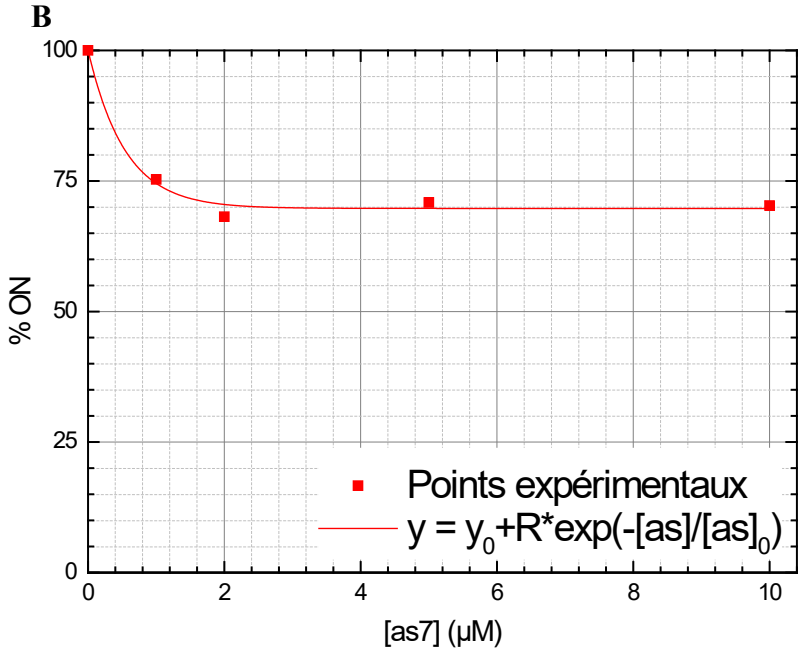


Figure 27 Fonctionnement du commutateur répressible rSw7 en présence de 1, 2, 5, 10 µM d'antisens ARN as7 et 10 µM d'antisens ARN non-spécifique as13. (A) Image du gel d'électrophorèse montrant les différents transcrits obtenus, après transcription *in vitro* du commutateur rSw7, en présence de son antisens ARN spécifique as7 et d'un autre antisens ARN non-spécifique as13. **(B)** Pourcentage de répression obtenu avec le commutateur rSw7 en présence de différentes concentrations d'antisens as7. La courbe en trait plein est un ajustement des valeurs de répression construite suivant l'équation *(5)*. Les meilleurs paramètres d'ajustement sont donnés dans le **Tableau 2.**

$$\boldsymbol{y = y_0 + R \times \exp\left(\frac{-[as]}{[as]_0}\right)} \qquad \textit{(5)}$$

avec :

y_0 : le pourcentage d'induction en présence d'antisens ARN (ligne de base)

R : l'amplitude de régulation qui est égale à $y_0 - Rmax$, $Rmax$ étant la réponse maximale observée ;

$[as]_0$: la concentration caractéristique de l'antisens ARN

Chi^2	3,81	
R^2	0,98	
	Valeurs	Écarts-types
$\boldsymbol{y_0}$	69,74	1,25
$\boldsymbol{R}$	30,32	2,30
$[as]_0$	0,54	0,14

Tableau 3 Paramètres utilisés pour l'ajustement des données de répression avec le courbe construite à partir de l'équation *(5)*.

Les résultats des transcriptions *in vitro* montrent que la commutation de rSw7 de l'état ON vers l'état OFF n'est que partielle en présence de l'antisens ARN as7. Plusieurs hypothèses peuvent expliquer ce résultat :

- faible efficacité ou faible stabilité de l'interaction antisens ARN–commutateur ARN ;
- faible efficacité de formation du terminateur (formation non-systématique, ou tardive, ou partielle, ou instable) ;
- faible efficacité de terminaison de la transcription par le terminateur (effet dû à la séquence spécifique du terminateur)

Afin de discriminer entre ces différentes hypothèses, nous décidons de modifier rSw7 en procédant à la délétion de la région ciblée par l'antisens ARN as7, ceci afin de mimer l'effet d'une interaction 100 % efficace entre l'antisens as7 et le commutateur rSw7 (**Figure 28**). Nous appelons ce nouveau commutateur rSw7_ΔAT car la délétion revient à une délétion de l'anti-terminateur présent dans le commutateur. En effet, en enlevant la séquence cible de l'antisens as7, on empêche la formation de la première hélice de l'anti-terminateur, conduisant ainsi à ce que seul le terminateur puisse se former. Les simulations par Kinefold du repliement co-transcriptionnel de rSw7_ΔAT montrent que la formation du terminateur est bien systématique. Par conséquent, on s'attend à observer un taux de transcrits courts reflétant l'efficacité de terminaison intrinsèque du terminateur du commutateur.

La **Figure 29** représente les résultats de transcriptions *in vitro* de rSw7_ΔAT comparés à la transcription de rSw7 en absence ou en présence d'antisens ARN effecteur. On observe que la transcription de rSw7_ΔAT conduit à seulement 26 % de terminaison de transcription au niveau du terminateur du commutateur. Ce taux de terminaison est donc comparable à ce que l'on obtient pour le commutateur rSw7 en présence d'une concentration saturante de 10 µM d'antisens as7 (≈ 30 % de terminaison). Ces résultats indiquent donc que la faible efficacité de terminaison du terminateur au sein du commutateur est effectivement à l'origine de la répression partielle que nous observons avec rSw7.

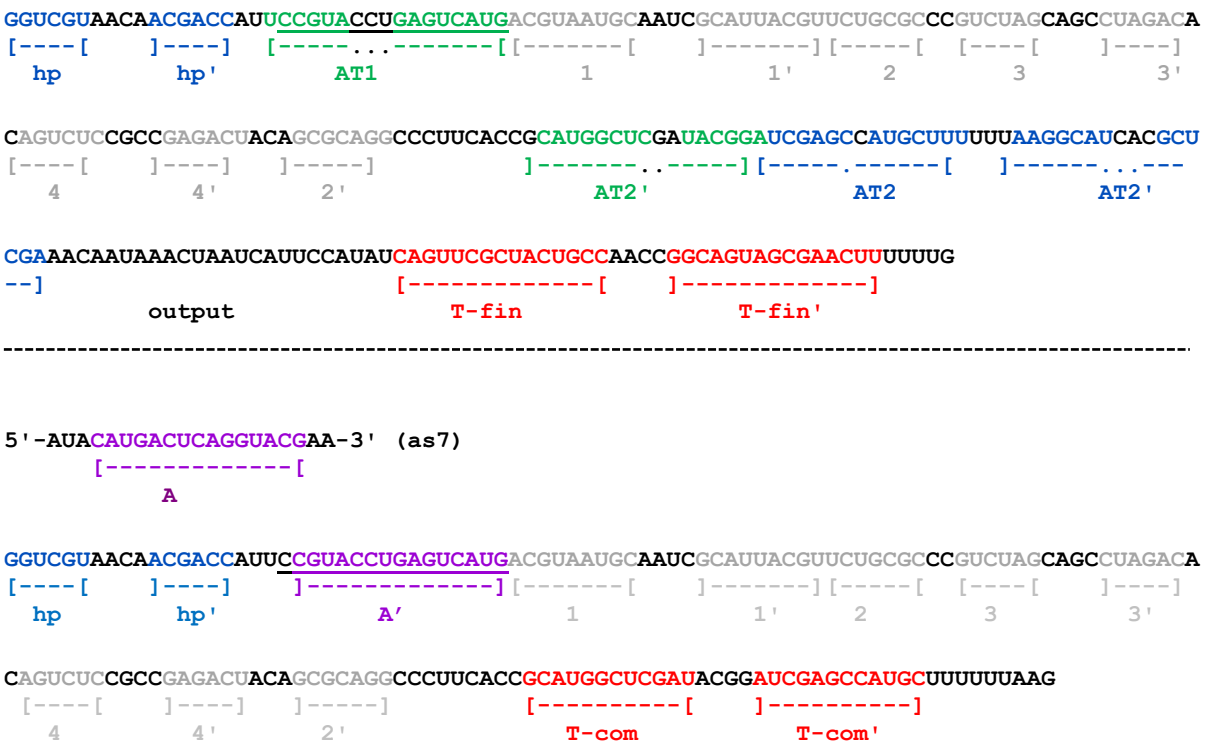


Figure 28 Séquences et structures secondaires dans le commutateur rSw7 en absence et en présence d'antisens as7. La séquence et les structures secondaires représentées sont celles du commutateur répressible rSw7. **(A)** Transcrit et structures secondaires attendu dans l'état nominal (en absence d'antisens ARN effecteur). **(B)** Transcrit et structures secondaires attendus en présence de l'antisens ARN effecteur as7. Les régions formant des appariements sont matérialisées par leur code couleur et par une paire d'éléments entre crochets (`[----[ ]----]` ; des points (`.`) indiquent des mésappariements au sein d'une hélice). Ces régions sont annotées par un nom d'hélice (`hp/hp′` : hairpin ; `AT1/AT1′` : anti-terminateur 1 ; `AT2/AT2′` : anti-terminateur 2 ; `1/1′`, `2/2′`, `3/3′` et `4/4′` : structures secondaires communes aux deux repliements ; `T-fin/T-fin′` : terminateur final constitutif ; `A/A′` : interaction entre l'antisens ARN as7 et sa région cible sur le commutateur. En **A** et **B**, la région soulignée correspond à la région retirée pour obtenir rSw7_ΔAT.

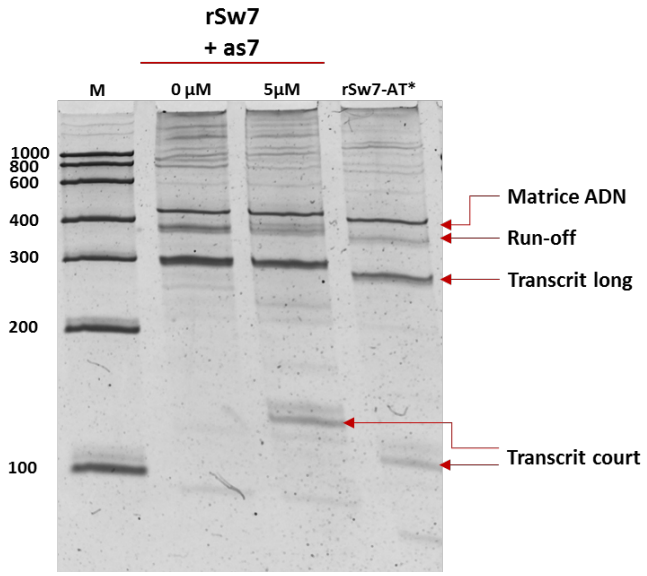


Figure 29 Détermination de l'efficacité de terminaison intrinsèque du terminateur du commutateur rSw7. Analyse par gel d'électrophorèse des transcrits obtenus par transcription *in vitro* de rSw7, en absence et en présence de son antisens ARN spécifique as7, et de rSw7_ΔAT, dans lequel la région cible de l'antisens as7 a été délétée. Bien que rSw7_ΔAT forme constitutivement le terminateur du fait de la délétion d'un élément de l'anti-terminateur, le taux de terminaison est similaire à celui observé pour rSw7 en présence de l'antisens ARN effecteur as7, indiquant que la faible efficacité de terminaison observée pour rSw7 est due à une mauvaise efficacité de terminaison intrinsèque du terminateur de rSw7.

Chapitre 4 : Conclusion

Dans cette partie, nous avons testé *in vitro* deux types de commutateurs : trois commutateurs inductibles (dont deux conçus de façon semi-automatique) et un commutateur répressible conçu de façon semi-automatique.

Parmi les trois commutateurs inductibles que nous avons testés, iSw13 fonctionne de manière satisfaisante mais les deux autres semblent être constitutivement ON dans leur état nominal. Une analyse détaillée des simulations de repliement co-transcriptionnel, comparée aux résultats expérimentaux obtenus par transcription *in vitro*, semble indiquer que la faible efficacité de terminaison obtenue pour ces commutateurs pourrait être due à une efficacité de repliement des terminateurs suboptimale. Ce résultat pourrait indiquer que des critères plus stricts sur la qualité de formation des terminateurs doivent peut-être être appliqués lors de la sélection des séquences de commutateurs conçus de façon semi-automatique.

En ce qui concerne le commutateur répressible rSw7 que nous avons testé, nous avons également observé une faible efficacité de terminaison à l'état réprimé. Dans ce cas, nous avons pu mettre en évidence un défaut dans l'efficacité de terminaison intrinsèque du terminateur du commutateur, écartant donc à priori un problème de formation du terminateur. Ce résultat pointe une limite, dans l'état actuel des connaissances, à la conception de commutateurs : il est

en effet difficile de prévoir l'efficacité de terminaison d'un terminateur à partir de sa séquence. Par ailleurs, le défaut d'efficacité de terminaison intrinsèque mis en évidence avec rSw7 pourrait aussi expliquer les résultats obtenus sur les commutateurs inductibles, à savoir un effet de séquence sur l'efficacité de terminaison intrinsèque des terminateurs des commutateurs inductible déficients.

Dans la suite de ce travail, je détaillerai la mise au point d'une plateforme de test *in vivo* pour les commutateurs ARN et les tests *in vivo* des commutateurs conçus et validés *in vitro*. Nous utiliserons donc pour la suite le commutateur inductible iSw13, conçu de façon semi-automatique, et le commutateur répressible rSw1, conçu de façon manuelle et précédemment validé par Alexandre Dawid *et al.* [2].

Partie III : Développement d'une plateforme et d'une méthode pour le test et la caractérisation *in vivo* des commutateurs ARN

Chapitre 1 : Plateforme génétique et instrumentale pour la caractérisation *in vivo* du fonctionnement des commutateurs ARN : généralités

L'objectif de cette partie de mon travail est de mettre en place une plateforme permettant d'évaluer *in vivo* le fonctionnement de commutateurs ARN. Je décrirai les améliorations successives de la plateforme construite chez *E. coli*, et je montrerai les résultats de caractérisation d'un commutateur activable (iSw13) et d'un commutateur répressible (rSw1).

1. Principe général

Nous nous sommes appuyés sur les travaux de Farren J Isaacs *et al.* [42] ou encore de Guillermo Rodrigo *et al.* [75] concernant les méthodes de caractérisation d'un riborégulateur, ainsi que les éléments génétiques utilisés pour concevoir une plateforme de test *in vivo* pour ces riborégulateurs (ici des commutateurs ARN).

Les commutateurs ARN que nous concevons, contrôlent l'expression d'un gène se trouvant à leur aval, de manière co-transcriptionnelle. Cette expression est commutable par le biais d'une interaction ARN-ARN avec un antisens ARN.

Nous choisissons d'effectuer cette caractérisation dans les bactéries, plus spécifiquement *Escherichia coli*, car elles sont facilement manipulables du point de vue génétique. De plus, le fonctionnement des commutateurs ARN que nous concevons s'inspire de mécanisme de terminaison/anti-terminaison bien connu dans certains mécanisme de régulation chez les bactéries (ex : régulation du nombre de copies de plasmide [76][77] et d'autres exemples décrits dans l'introduction).

Le type de repliement adopté par le commutateur déterminera le niveau d'expression du gène en aval, qui correspond dans cette étude à un gène qui code pour une protéine rapportrice. Dans le cas d'un commutateur ARN répressible, le gène est normalement exprimé en absence de l'antisens ARN et réprimé en sa présence (**Figure 30**). Au contraire, dans le cas d'un commutateur ARN inductible, l'expression du gène ne devrait être induite qu'en présence de l'antisens ARN.

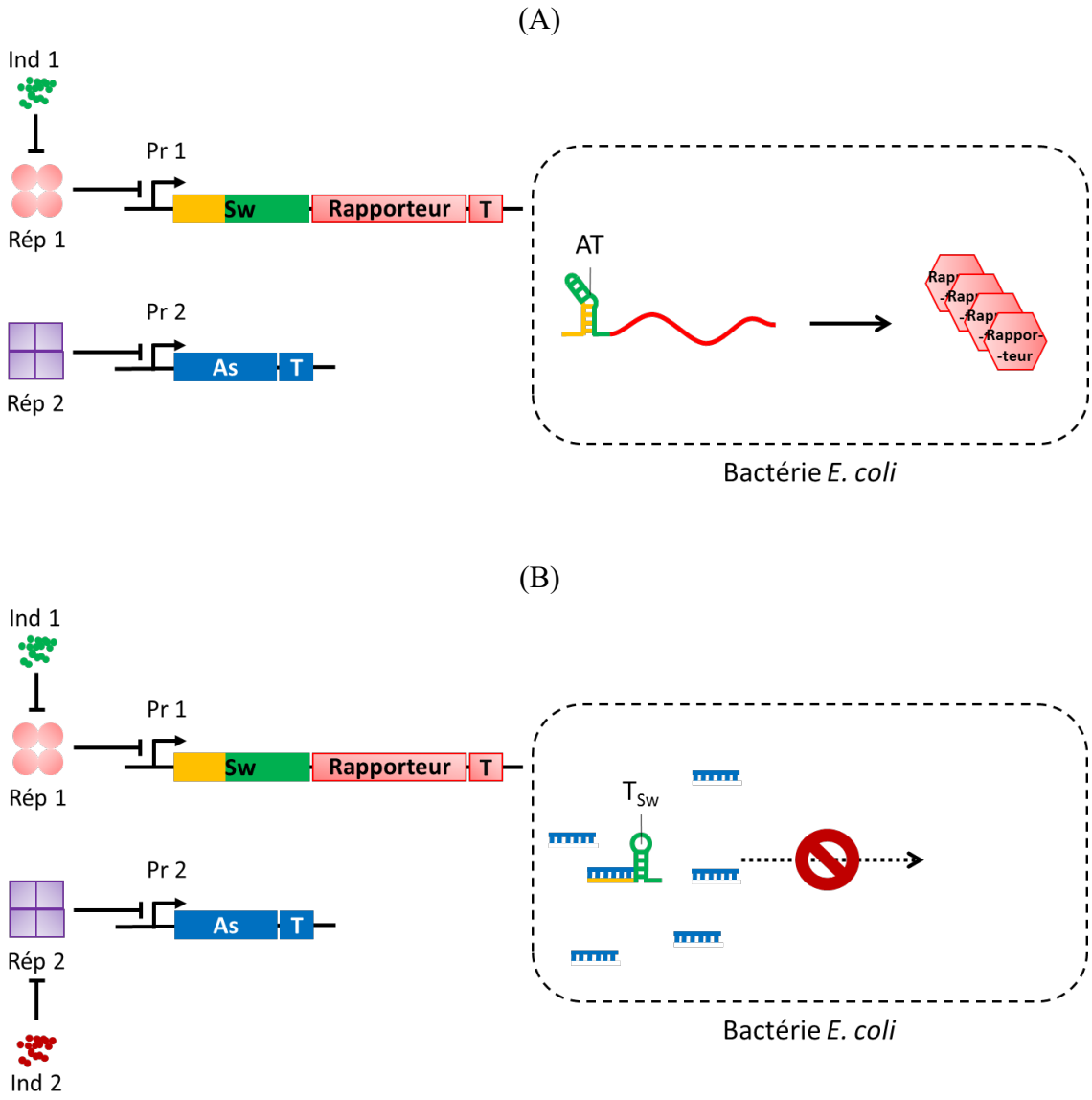


Figure 30 Schéma de principe de la plateforme de test *in vivo* initiale, illustré dans le cas d'un commutateur ARN répressible. Les transcriptions du commutateur ARN et de l'Antisens ARN sont sous le contrôle de promoteurs inductibles. A gauche : schéma des régions d'ADN contenant les gènes du système : promoteurs (Pr) et terminateurs (T), ainsi que les gènes du commutateur (Sw), de la protéine rapportrice (Rapporteur) et de l'antisens ARN (As) ; mais aussi les régulateurs de leur expression (Rép pour répresseur et Ind pour inducteur). A droite : état d'expression du système dans une bactérie. Le type de repliement adopté par le commutateur ARN détermine l'expression du gène en aval. La mesure du niveau d'expression du gène rapporteur produit par bactérie permet ainsi de déduire l'état et le fonctionnement des commutateurs ARN selon qu'ils sont en absence ou en présence d'antisens ARN. **(A)** État de régulation d'un commutateur ARN répressible en absence et **(B)** en présence d'antisens ARN. Les promoteurs inductibles Pr 1 et Pr 2 sont réprimés par leurs répresseurs respectifs (Rép 1 et Rép 2) et activés lorsque leurs inducteurs (Ind 1 et Ind 2) sont présents.

Le principe d'évaluation du fonctionnement du commutateur repose donc sur une mesure indirecte : le taux d'expression du gène en aval du commutateur. L'hypothèse implicite de cette mesure est que l'ARN antisens n'influe que sur la formation de la conformation terminateur du

commutateur et sur aucun autre paramètre (des paramètres comme la stabilité de l'ARN régulé, l'efficacité de l'ARN polymérase ou des ribosomes, la croissance des bactéries, etc.). Les effets non-spécifiques de l'ARN antisens seront évalués par l'expression d'un ARN contrôle, qui ne possède pas de séquence pouvant interagir avec le commutateur.

2. Méthode de mesure et de caractérisation des commutateurs ARN

Le but de l'approche est de déterminer l'état du commutateur en mesurant le niveau d'expression du gène rapporteur par bactérie en fonction du temps, de façon à pouvoir remonter à un taux d'expression relatif de ce dernier, et ainsi pouvoir remonter à l'état de repliement du commutateur dans une condition donnée (absence ou présence de l'antisens ARN effecteur). En effet, la mesure du niveau d'expression du gène rapporteur par bactérie correspond à une mesure indirecte de l'état de repliement du commutateur ARN. Le niveau d'expression moyen du gène rapporteur par bactérie est mesuré de manière indirecte par l'intermédiaire de la mesure de deux grandeurs qui sont la fluorescence émise par la protéine et la turbidité de la culture. Nous verrons dans la suite de ce chapitre comment nous mesurons ces deux grandeurs.

2.1. Conditions permettant la reproductibilité des mesures

Le but de l'expérience est de tester et de valider le fonctionnement des commutateurs ARN en absence ou en présence d'antisens ARN. Le niveau d'expression moyen des protéines rapportrices dans les bactéries est a priori dépendante de l'état moyen des commutateurs dans les bactéries. Toutefois elle n'est qu'une représentation indirecte de cet état moyen car, en outre l'état de repliement moyen des commutateurs, le niveau d'expression de la protéine rapportrice mesuré par bactérie dépend en principe de nombreux autres facteurs : nombre de copies de plasmides et donc nombre de copies des gènes du système, force intrinsèque des promoteurs, nombre d'ARN polymérases dans la cellule, taux de dégradation des transcrits (Antisens ARN et transcrit « commutateur + séquence codante »), efficacité du RBS, nombre de ribosomes dans la cellule, vitesse de maturation et taux de dégradation de la protéine rapportrice, mais aussi le taux de dilution et donc le taux de croissance des bactéries, etc. Pour que le niveau d'expression de la protéine rapportrice par bactérie, soit une représentation fidèle de l'état du commutateur en fonction uniquement de la présence spécifique de l'antisens ARN, il est important que les autres facteurs restent constants d'une expérience à l'autre.

Une façon d'obtenir des cellules dans un état stationnaire est de s'assurer que la culture soit en croissance équilibrée [78][79]. La croissance équilibrée correspond à la situation où toutes

les propriétés extensives d'une population de bactérie (biomasse, nombre de cellules, masse de protéine, masse d'ADN,…) augmentent avec un taux constant et identique [80]. Cette condition est atteinte lorsque la culture est en phase exponentielle de croissance pendant suffisamment longtemps de sorte que la composition moyenne des bactéries et la concentration interne moyenne de tous les constituants de la cellule deviennent constants [81]. Cette condition de croissance équilibrée signifie également que l'état moyen des bactéries ne dépend plus de l'historique de la culture mais uniquement des conditions de croissance, qui dépendent principalement du milieu dans lequel elles sont cultivées.

2.2. Rappels sur les modes de cultures bactériennes

En microbiologie, il existe trois grands modes de culture des bactéries :

- Un mode de culture en batch ou discontinue qui correspond à un système fermé dans lequel les microorganismes sont inoculés dans un volume fixe d'un milieu de culture, correspondant à une solution de nutriment stérile. La culture est ensuite maintenue dans des conditions de température, d'oxygénation et d'agitation constantes.

  Pendant le processus de croissance, les bactéries consomment les nutriments et fabriquent des produits dérivés. Par conséquent, la composition du milieu pour ce mode de culture varie constamment. On peut y observer quatre phases de croissance [82][83][84] (Figure 31) :

  - Une phase de latence pendant laquelle les bactéries s'adaptent à leur nouvel environnement. Le taux de division des bactéries y est nul, elles augmentent en taille et elles synthétisent les enzymes ainsi que des facteurs nécessaires à leur division.
  - Une phase exponentielle pendant laquelle la division cellulaire se déroule à un taux constant, de même que tous les constituants internes de la cellule. Idéalement, c'est pendant cette phase que toutes les mesures seront effectuées.
  - Une phase stationnaire pendant laquelle autant de bactéries naissent et meurent dans le milieu, en raison d'un début d'épuisement de nutriments nécessaires à leur survie ou d'un autre facteur limitant (pH trop acide, diffusion de l'oxygène limitant). Pendant cette phase, le nombre de bactéries dans le milieu est constant.
  - Une phase de déclin pendant laquelle les bactéries ne se divisent plus et meurent.
- Un mode de culture en milieu alimenté ou en fed-batch, qui est similaire à la culture en batch, mais dans laquelle le nutriment indispensable pour la croissance des

microorganismes est rajouté au fur et à mesure de sa consommation, de façon qu'il ne soit pas en excès dans le milieu de culture.

- Un mode de culture en continue ou à système ouvert, pour lequel du milieu frais est rajouté en continu et un volume équivalent de la culture est prélevé de façon que celui-ci reste constant pendant toute la durée du processus. Parce que le milieu de culture y est renouvelé régulièrement, ce mode de culture permet de garder les bactéries en croissance équilibrée pendant des jours, des semaines, et même des mois sans que les bactéries entrent en phase stationnaire de croissance.

Nous verrons plus tard quel mode de culture est plus adapté pour le type de mesure que nous envisageons de réaliser.

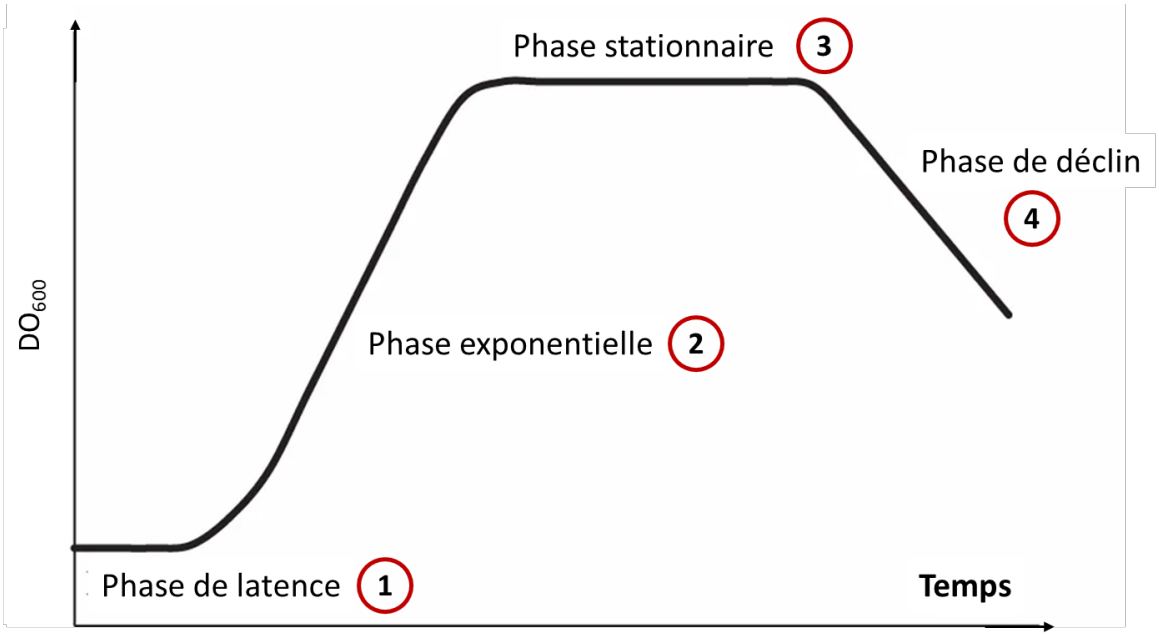


Figure 31 Courbe de croissance bactérienne. 1 – Phase de latence. Les bactéries s'adaptent au nouveau milieu de croissance. 2 – Phase exponentielle. Les bactéries se divisent à un taux constant. C'est au cours de cette phase que le régime de croissance équilibré, pour lequel la composition interne des bactéries reste constante, peut être atteint. 3 – Phase stationnaire. Épuisement des nutriments. 4 – Phase de déclin. Mort des bactéries. [85]

3. Spécificités visées pour la plateforme génétique

Dans notre plateforme de test (Figure 30), l'expression des deux espèces d'ARN (commutateur ARN + protéine rapportrice, et antisens ARN) se fait à partir de vecteurs plasmidiques différents et leurs transcriptions sont contrôlées de manière indépendante par des promoteurs distincts. Le choix du chemin de repliement des molécules d'ARN se faisant au cours de leur transcription, l'état de repliement du commutateur ARN est donc sous contrôle cinétique et non thermodynamique. Pour que la régulation par l'antisens ARN ait lieu, il faut

donc s'assurer que la fréquence de rencontre entre l'antisens ARN et sa cible sur le commutateur ARN soit suffisamment rapide pour se produire avant qu'un choix irréversible de chemin de repliement n'ait lieu.

Pour ce faire, on s'arrange pour que :

(i) la concentration en antisens ARN soit en excès devant la concentration en commutateur ARN, pour s'assurer que les antisens ARN ne soient pas épuisés même si tous les commutateurs ARN se lient à un antisens ARN ;

(ii) la cinétique d'hybridation entre l'antisens ARN et le commutateur ARN soit suffisamment rapide, pour garantir un effet de l'antisens ARN sur le repliement du commutateur ARN bien avant que la décision génétique ne soit prise – Wickiser *et al.* [86][87] montrent par exemple dans leurs études que la fonction de la plupart des riboswitches repose sur la cinétique de liaison de son ligand ainsi que sur la vitesse de transcription de l'ARN et non sur la simple affinité du ligand pour sa cible. Pour décrire le comportement d'un tel système, les paramètres tels que la constante de vitesse d'association ou de dissociation sont plus importants que la constante d'équilibre K_D.

On peut ainsi décrire le contrôle cinétique de la façon suivante :

$$\frac{d[as.Sw]}{dt} = k_{on}[as] \times [Sw] - k_{off}[as.Sw] \qquad (6)$$

avec :

k_{on} : constante de vitesse d'association

k_{off} : constante de vitesse de dissociation

$[as]$: concentration en antisens ARN libre

$[Sw]$: concentration en commutateur ARN libre

$[as.Sw]$: concentration du complexe antisens ARN_commutateur ARN

Pour cette étude, nous nous sommes cependant contentés de tout simplement vérifier que l'expression de l'antisens ARN a effectivement un effet sur l'expression du gène rapporteur et, par conséquent, sur l'état de repliement du commutateur ARN. En toute rigueur, il aurait fallu vérifier toutes ces conditions comme cela a été proposé par Wickiser *et al.* [86][87].

4. Principe de la mesure de l'expression moyenne de la protéine rapportrice par bactérie

Une des méthodes pour mesurer et caractériser l'expression d'un gène dans les cellules vivantes est l'utilisation d'une protéine rapportrice. Les protéines rapportrices peuvent

témoigner du niveau d'expression d'un gène qui serait sous le contrôle d'un promoteur d'intérêt ou d'un régulateur d'intérêt, par exemple un commutateur ARN. Plusieurs protéines rapportrices sont connues à ce jour, mais la méthode la plus simple est l'utilisation de protéine fluorescente qui permet de visualiser des évènements cellulaires dans leur environnement naturel. La fluorescence se définit comme l'émission de photons par une molécule (protéine fluorescente) en réponse à l'absorption de photons de longueur d'onde plus courte. La mesure de la fluorescence permet de faire une quantification relative du nombre de protéines synthétisées par la cellule à différents temps et dans des conditions expérimentalement maîtrisées. L'utilisation de ce type de protéine rapportrice peut cependant avoir quelques limites : lorsque l'intensité de fluorescence est trop faible il est parfois difficile de la détecter, mais aussi de la distinguer de la fluorescence émise par la bactérie, ou encore du milieu de culture (autofluorescence). Il faut alors que cette protéine fluorescente émette dans une longueur d'onde différente de celle émise par la bactérie ou du milieu de culture, il faut également qu'elle soit très brillante de manière qu'on puisse facilement la détecter. Pour ces raisons, nous avons choisi d'utiliser la protéine fluorescence mKate2 comme rapporteur. Cette protéine émet une fluorescence forte dans le rouge, une longueur d'onde à laquelle la bactérie ne montre aucune autofluorescence.

On veut mesurer une fluorescence moyenne par bactérie, que l'on obtient en rapportant l'intensité de fluorescence mesurée par le nombre de bactéries présentes dans la culture à un temps donné.

En microbiologie, la méthode standard pour estimer le nombre de bactéries est de mesurer la densité optique (ou absorbance) du milieu dans lequel elles sont en suspension. Cette densité optique, mesurée à une longueur d'onde de 600 nm grâce à un spectrophotomètre, correspond à une mesure de la lumière diffusée par la suspension bactérienne, qui, pour des densités pas trop élevées de bactéries, croît linéairement avec le nombre de bactéries. Mathématiquement elle se définit telle que :

$$\boldsymbol{DO_{600} = -log_{10}\frac{I}{I_0}} \qquad (7)$$

avec :

I_0 = intensité de la lumière incidente ;

I = intensité de la lumière qui traverse l'échantillon vue par le détecteur.

Cette méthode a cependant une certaine limite, car la relation entre la densité optique et la concentration cellulaire est linéaire jusqu'à un certain seuil, au-delà duquel la probabilité que le faisceau soit rediffusé dans le détecteur augmente à cause d'une trop forte densité cellulaire, entrainant une sous-estimation de la concentration cellulaire effective. La suspension doit alors être suffisamment diluée pour rester dans la zone de linéarité.

4.1. Choix d'un instrument de mesure : le lecteur de microplaques

Pour que les mesures de fluorescence représentent l'état du commutateur, il est souhaitable de se mettre en condition de croissance équilibrée. Ainsi les mesures d'absorbance et de fluorescence pourront représenter fidèlement le changement d'état du commutateur qui serait uniquement dû à la présence de l'antisens ARN spécifique. Une des solutions pour atteindre cette condition serait d'utiliser une culture continue qui permet de se mettre en condition de croissance équilibrée pendant un temps plus ou moins long (des jours, des semaines, voire des mois). Cependant ce mode de culture est difficile à mettre en place, car il nécessite l'utilisation de pompes, utilisées pour renouveler le milieu de culture très régulièrement ce qui pose en plus le problème de stérilité. Par ailleurs, les cultures étant faites sur une période assez longue, des mutations risques d'apparaitre dans le génome de la bactérie.

Plus simplement nous choisissons alors de réaliser les mesures avec un lecteur de microplaque qui permet non seulement de cultiver les bactéries dans un environnement maîtrisé (température, agitation), mais aussi de faire des mesures directes de fluorescence et d'absorbance au cours du temps et de suivre en temps réel l'évolution du système. On retrouve les conditions de croissance équilibrée pendant la phase exponentielle de croissance.

4.2. Fonctionnement du lecteur de microplaques : l'Infinite M200 Pro de TECAN

Pour mesurer la densité optique de la culture bactérienne et la fluorescence, nous utilisons un lecteur de microplaque « Infinite M200 PRO ». Ce lecteur de microplaque est un instrument multifonction qui permet de réaliser des mesures simultanées d'absorbance et de fluorescence. L'utilisation de microplaque 96 puits permet de réaliser des mesures de plusieurs conditions expérimentales en une seule fois. Outre cette grande flexibilité, le lecteur de microplaque « Infinite M200 PRO » réunit toutes les options nécessaires pour réaliser une culture bactérienne :

- L'agitation de la microplaque pour l'homogénéisation et l'oxygénation de la culture bactérienne

- La possibilité de contrôler la température

4.3. Système de mesure de la fluorescence

Le système optique du lecteur de microplaque utilisé pour mesurer la fluorescence est schématisé sur la Figure 32. Pour émettre une fluorescence, un fluorophore doit être excité par une lumière de longueur d'onde spécifique. Chez le TECAN Infinite M200 PRO la lumière excitatrice est fournie par une lampe à décharge au xénon de haute énergie **(1)**, elle s'éteint en quelques microsecondes et sa fréquence de clignotement est de 40 Hz. Cette même lampe est utilisée pour les mesures de fluorescence et les mesures de densité optique avec une intensité forte allant de l'ultraviolet au proche infrarouge. La lumière émise par la lampe est concentrée par un condenseur optique **(1)**, puis dirigée vers un filtre optique **(1)** pour bloquer les dispersions parasites pouvant être produites par le réseau optique. A la sortie du filtre, la lumière rentre par la fente d'un monochromateur d'excitation **(2)** utilisé pour sélectionner la longueur d'onde d'excitation souhaitée.

Le mode de fonctionnement du monochromateur est comparable à celui d'un filtre optique ajustable qui permet de régler à la fois la longueur d'onde et la bande passante. Schématisé sur la Figure 33, un monochromateur est un instrument d'optique, composé d'une fente d'entrée, d'un élément dispersif se présentant sous la forme d'un réseau optique et d'une fente de sortie. L'élément dispersif diffracte la lumière et la projette vers la fente de sortie du monochromateur. La longueur d'onde de cette lumière est définie par l'angle de rotation autour de l'axe vertical du réseau optique qui compose l'élément dispersif et la bande passante est définie par la largeur de la fente de sortie du monochromateur. Pour assurer une meilleure sélection de la longueur d'onde d'excitation, deux monochromateurs d'excitation sont connectés en série, telle que la fente de sortie du premier monochromateur joue le rôle de la fente d'entrée du deuxième monochromateur. Le double monochromateur d'excitation permet ainsi de sélectionner des longueurs d'onde comprises entre 230 et 850 nm pour la fluorescence et entre 230 et 1000 nm pour l'absorbance avec une bande passante inférieure à 9 nm si la longueur d'onde est supérieure à 315 nm et une bande passante inférieure à 5 nm si la longueur d'onde est inférieure ou égale à 315 nm.

A la sortie du monochromateur d'excitation, la lumière est guidée, à travers un faisceau de fibres optiques, vers un système de lentilles **(3)** situé à l'extrémité de la fibre et conçu pour concentrer la lumière d'excitation dans l'échantillon.

La fluorescence émise par l'échantillon est collectée en retour par le même système optique qui concentre la lumière vers le faisceau de fibres optiques et guide cette lumière vers le double monochromateur d'émission **(4)** qui permet de sélectionner des longueurs d'ondes comprises entre 280 et 850 nm avec une bande passante de 20 nm. En quittant le monochromateur d'émission, la lumière émise arrive au niveau du tube photomultiplicateur PMT **(5)**, en passant au préalable par un filtre optique **(5)**. Le tube photomultiplicateur est utilisé pour détecter la lumière associée à la fluorescence qui est ensuite convertie en courant électrique pouvant être amplifié via le réglage du gain (entre 60 et 255).

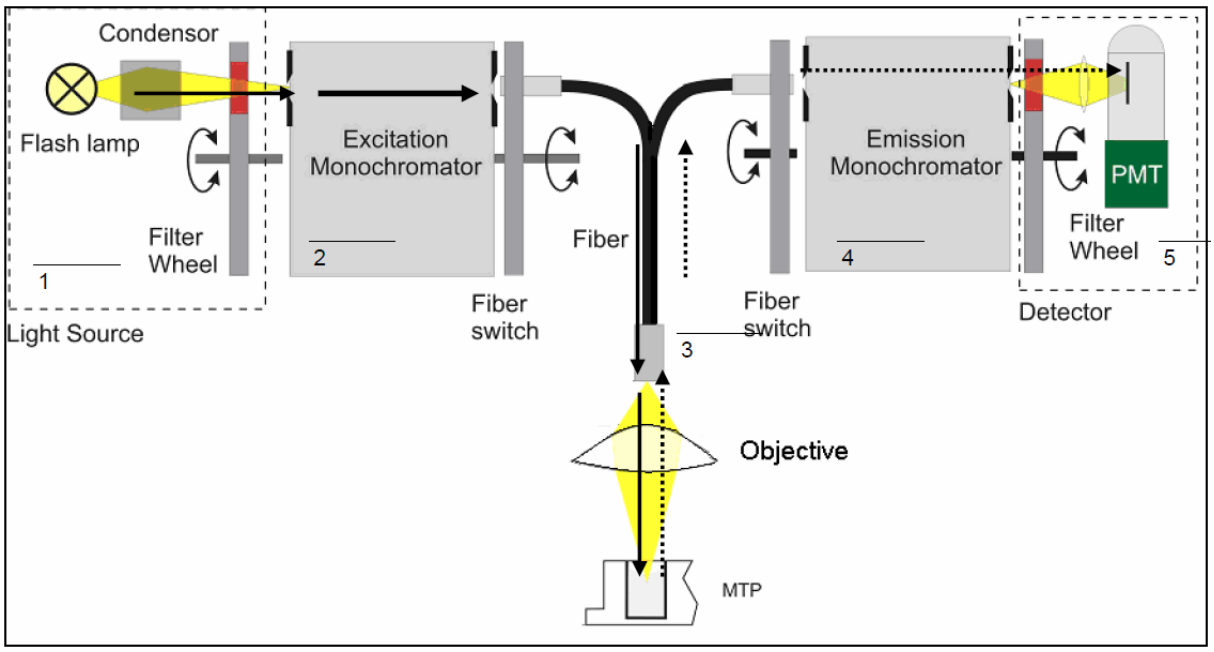


Figure 32 Système optique du lecteur de microplaque TECAN « Infinite M200 PRO » pour la mesure de fluorescence. Le système est constitué d'une source de lumière **(1)**, d'un monochromateur d'excitation **(2)**, d'un système otique qui concentre la lumière d'excitation dans l'échantillon **(3)**, d'un monochromateur d'émission **(4)** et d'un détecteur **(5)**. Les flèches en trait plein indiquent la direction de la lumière d'excitation et les flèches en pointillées indiquent la direction de la lumière d'émission. *Manuel d'utilisation du TECAN Infinite 200® PRO.*

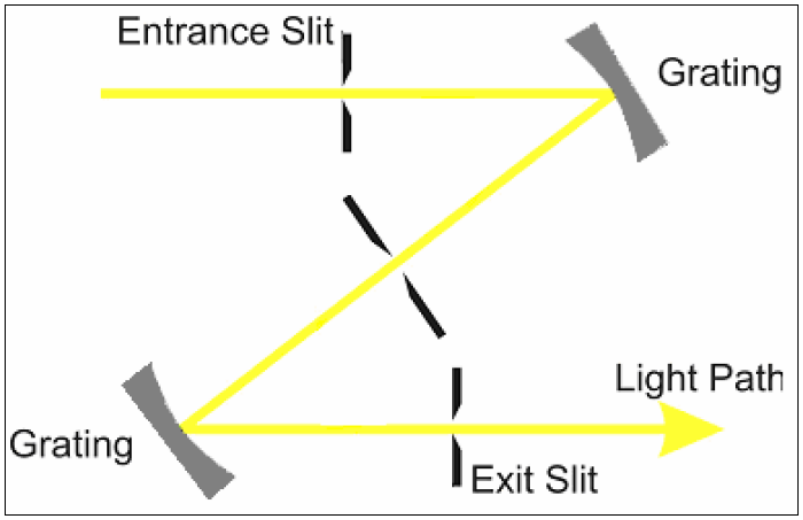


Figure 33 Schéma du réseau optique utilisé pour le monochromateur d'excitation. *Manuel d'utilisation du TECAN Infinite 200® PRO.*

4.4. Système de mesure de l'absorbance

Pour les mesures d'absorbance, un chemin optique similaire à celui de la lumière d'excitation pour les mesures de fluorescence est utilisé. La lumière provenant du monochromateur d'excitation est guidée à travers un faisceau de fibres optiques, vers une paire de lentilles qui concentre le faisceau lumineux dans un puits de la microplaque. La lumière transmise à travers l'échantillon est mesurée grâce à un dispositif optique situé sous le support de la plaque. Avant d'effectuer la mesure, une mesure de référence est effectuée avec le porte-plaque vide (Figure 34).

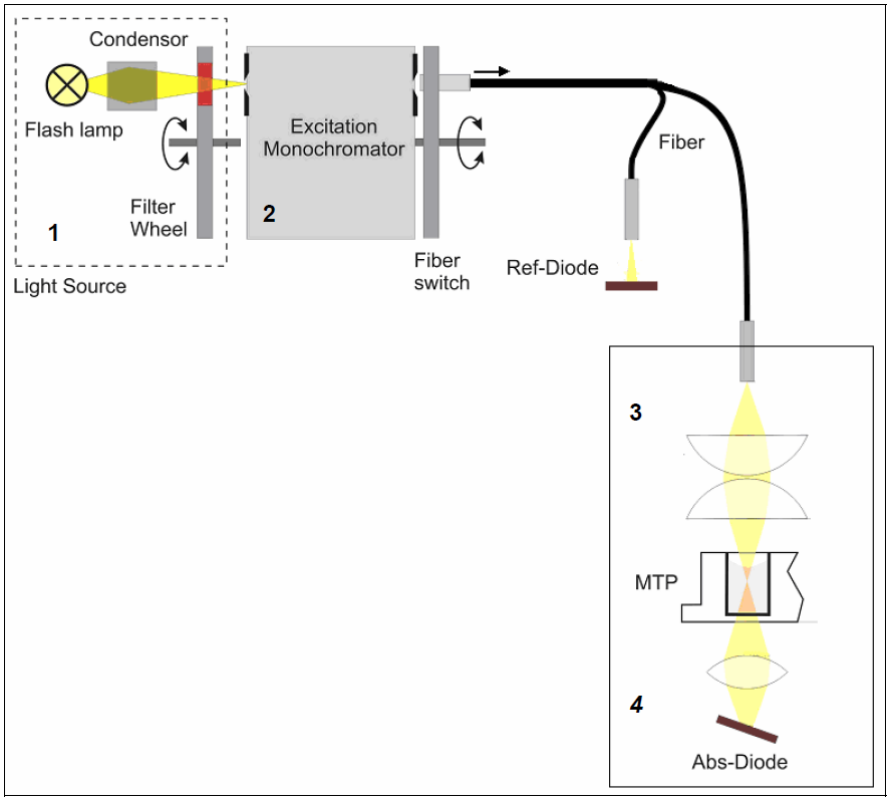


Figure 34 Système optique du lecteur de microplaque TECAN « Infinite M200 PRO » pour la mesure d'absorbance. *Manuel d'utilisation du TECAN Infinite 200® PRO.*

Chapitre 2 : Mise en œuvre expérimentale et analyse des données : principes

1. Stratégie pour atteindre une croissance équilibrée

La Figure 35 présente un premier protocole qui pourrait satisfaire aux conditions requises pour atteindre la croissance équilibrée, mentionnée au début du chapitre « Plateforme génétique et instrumentale pour la caractérisation *in vivo* du fonctionnement des commutateurs ARN (généralités) ». Dans un premier temps, les bactéries qui expriment les commutateurs à tester, stockées à – 80 °C, sont ensemencées dans un milieu minimum M9 contenant 0,2 % de glucose ainsi qu'une quantité précise d'un antibiotique donné (lorsque cet antibiotique est nécessaire). Le milieu minimum M9 est un milieu de culture avec une composition bien définie contenant une quantité suffisante en nutriments nécessaires à la survie des bactéries [88] (pour plus de détails voir Annexe B : Méthodes). L'avantage d'utiliser un tel milieu est sa faible autofluorescence nous permettant d'améliorer le rapport signal/bruit des mesures de fluorescence. On assure également une meilleure reproductibilité des expériences, grâce à sa composition bien définie [89].

Après environ 16 heures en culture, les bactéries arrivent en phase stationnaire de croissance. Or nous voulons mesurer l'intensité de fluorescence moyenne par bactérie en condition de croissance équilibrée, qui se produit en phase exponentielle de croissance [78]. C'est pourquoi cette culture de démarrage est diluée jusqu'à une DO_{600} égale à 10^{-3} ou 10^{-4} pour s'assurer que les bactéries restent suffisamment longtemps en phase exponentielle avant que la fluorescence ne soit détectable et qu'on puisse commencer les mesures.

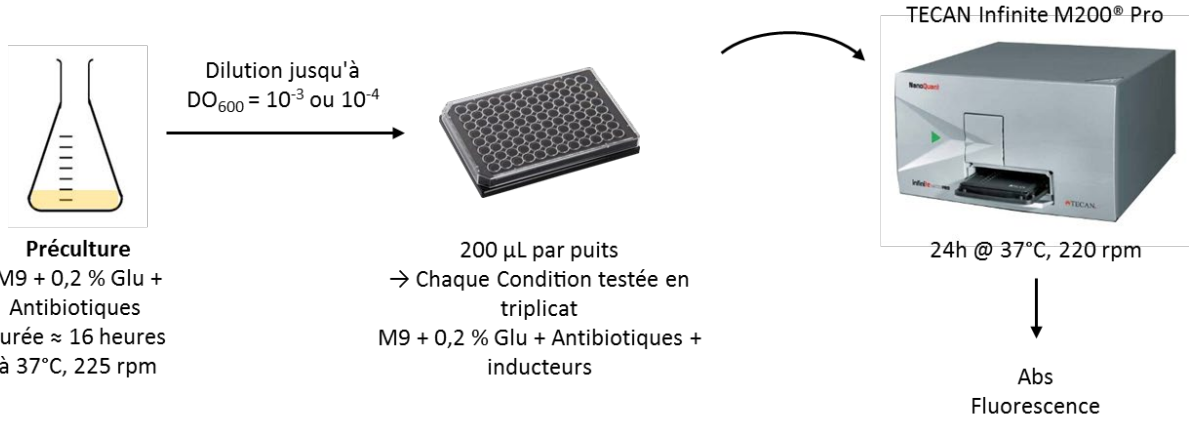


Figure 35 Préparation des échantillons bactériennes pour les mesures d'absorbance et de fluorescence. La veille, les bactéries exprimant les commutateurs à tester sont ensemencées dans un milieu minimum M9 contenant 0,2 % de glucose, ainsi que des antibiotiques lorsqu'ils sont nécessaires (40 µg/ml de kanamycine et/ou 30 µg/ml de chloramphénicol). Après environ 16 heures de culture sous agitation (225 rpm) à 37 °C, cette culture de démarrage est diluée à une DO_{600} égale à 10^{-3} ou 10^{-4}. La dilution se fait dans un milieu similaire au premier, mais dans lequel on rajoute les inducteurs chimiques utilisés pour l'induction de l'activité des promoteurs contrôlant les expressions du commutateur et de l'antisens. 200 µL de cette dilution est ensuite transféré en microplaque. Les mesures d'absorbance et de fluorescence sont réalisées à partir d'un lecteur de microplaque TECAN Infinite M200 PRO.

2. Analyse des données de mesure

Nous utilisons le logiciel OriginLab pour traiter, analyser et tracer les données sur un graphique. Pour chaque microplaque, les données issues de puits réplicats correspondant à une condition expérimentale donnée (typiquement 3 à 9 puits réplicats par condition) sont regroupées, ce qui permet de calculer la moyenne et l'écart type pour chaque condition. Les moyennes calculées à partir de groupe de puits remplis avec seulement du milieu M9 ou remplis d'une culture de bactérie n'exprimant pas la protéine fluorescente mKate2 sont respectivement utilisées comme blanc d'absorbance et blanc de fluorescence pour soustraire l'absorbance du milieu pour le premier et l'autofluorescence de la bactérie pour le second [90]. L'absorbance mesurée à partir du TECAN est ensuite multipliée par un facteur 3,4 pour la convertir en densité optique standard [91]. Cette conversion permet de convertir les données en une grandeur plus fréquemment utilisée en laboratoire et qui correspond à une mesure faite sur spectrophotomètre

avec un trajet optique standard de 1 cm. Le facteur multiplicatif 3,4 correspond à une moyenne du rapport de DO_{600} mesurées à chaque instant sur deux appareils différents (TECAN Infinite M200 PRO vs. Eppendorf BioPhotometer®) [92].

En rapportant les valeurs moyennes de fluorescence (soustrait de l'autofluorescence) au valeurs moyennes de DO_{600} (soustrait du blanc d'absorbance), nous obtenons une valeur moyenne de fluorescence/DO_{600} qui correspond à une intensité de fluorescence moyenne par bactérie. En condition de croissance équilibrée, la fluorescence moyenne par bactérie doit se présenter graphiquement comme une constante, cohérent avec la définition d'un état stationnaire.

3. Nécessité d'homogénéiser le milieu avec des microbilles

Nous verrons plus tard dans le chapitre 3 « Plateforme génétique et instrumentale pour la caractérisation *in vivo* du fonctionnement des commutateurs ARN : mise en œuvre initiale et optimisation » que le protocole utilisé ici et présenté sur la **Figure 35** ne permet pas toujours d'obtenir des courbes de fluorescence/DO_{600} constantes pendant la croissance équilibrée. Un problème en particulier est l'irrégularité observée de la courbe de croissance exponentielle (voir courbe de DO_{600} brute **Figure 36C**). Une hypothèse est que l'utilisation de puits de petit volume (200 µL) rend difficile l'homogénéisation et l'oxygénation de la culture par agitation. Les bactéries peuvent ainsi avoir tendance à s'agréger de façon inhomogène au fond des puits de la microplaque, introduisant alors un biais lors des mesures d'absorbance, et affectant finalement les valeurs de fluorescence/DO_{600}. Pour vérifier cette hypothèse, nous avons introduit des billes de 2 mm de diamètre dans chaque puits de la microplaque contenant une culture, pour favoriser l'homogénéisation. On constate en effet une suppression de la « bosse » (**Figure 36D**) pouvant être observée pendant la phase de croissance exponentielle lorsque les mesures sont faites en absence de billes (**Figure 36C**).

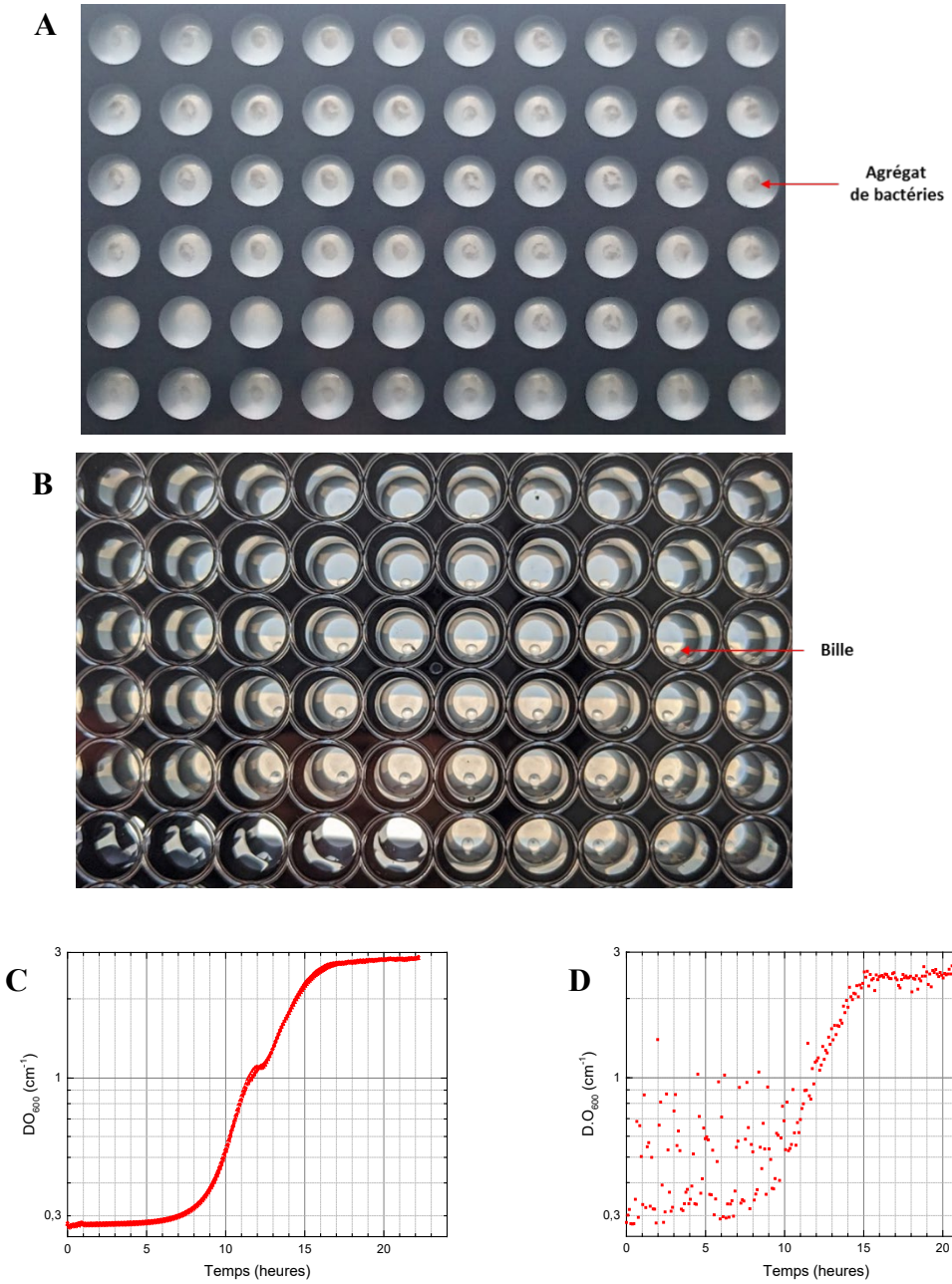


Figure 36 **(A) Image prise après une culture en microplaque en absence de billes.** Des agrégats de bactéries se déposent sur le fond de chaque puit. **(B) Image prise après une culture en microplaque en présence de billes.** Les bactéries ne se déposent plus sur le fond des puits. **(C) Courbe de DO_{600} mesurée en absence de billes et (D) en présence de billes.** En présence de billes, les courbes de DO_{600} font apparaitre un nuage de points aberrants en dehors de la trajectoire de la courbe de croissance. Ces points aberrants apparaissent à cause de la présence de billes qui peuvent se retrouver dans la trajectoire du faisceau lumineux servant à la mesure de l'absorbance. L'absorbance due aux bactéries est donc obtenue en filtrant les points aberrants [93]. Ceux-ci sont identifiés par leur écart à la courbe-tendance, et la comparaison des courbes obtenues entre répliques de cultures de bactéries faites dans des puits voisins permet un contrôle interne de la reproductibilité et la qualité du traitement appliqué.

Cependant, on observe de temps en temps, sur les courbes d'absorbance mesurées en présence de billes, des valeurs aberrantes (Figure 36D). En effet, il peut arriver qu'au moment de la mesure d'absorbance, la bille se trouve sur la trajectoire du faisceau lumineux, ce qui produit alors une valeur aberrante. A cause de ces aberrations, il devient impossible de faire l'analyse des courbes d'absorbance, à moins de pouvoir identifier et enlever ces valeurs aberrantes. L'équipe de Hidde de Jong [94][93] a développé un programme appelé WELLINVERTER qui permet de traiter et d'analyser les données d'absorbance et de fluorescence, mais surtout d'enlever les valeurs aberrantes qui peuvent apparaître sur les courbes d'absorbance mesurée en présence de billes (Figure 36D).

Nous verrons néanmoins, par la suite, que certains résultats présentés ici ont été obtenus avant la mise au point du protocole incluant les billes. Malgré cela, l'interprétation des résultats n'en sont pas affectée car les données que nous analysons sont celles obtenues en phase exponentielle qui ne sont à priori pas concernées par cette bosse qu'on observe plutôt en entrée de phase stationnaire de croissance (Figure 36C).

4. L'outil WELLINVERTER pour le traitement et l'analyse des données de culture en microplaques, en présence de billes

WELLINVERTER est un programme informatique développé par l'équipe de Hidde de Jong, qui permet de reconstruire le profil d'expression d'un gène donné, à partir de données de gène rapporteur, par une méthode numérique d'inversion linéaire [93]. Plus précisément, il permet d'estimer le taux de croissance des bactéries, le taux de synthèse protéique, ainsi que la concentration de protéine dans la bactérie avec seulement des données temporelles d'absorbance et de fluorescence. Plus important encore, cet outil permet de lire les données acquises à partir du lecteur de microplaque TECAN Infinite M200 PRO, de les traiter et de les analyser. Le programme inclut notamment la soustraction des blancs d'absorbance et de fluorescence [90], ainsi qu'un outil de suppression des valeurs aberrantes dues aux billes sur les courbes d'absorbance.

Pour la suppression de ces valeurs aberrantes, l'utilisateur a le choix entre 2 méthodes [93] :

- Une méthode manuelle qui consiste à tracer une courbe de tendance pour laquelle on fixe une plage de tolérance en dehors de laquelle on considère que les valeurs sont aberrantes (Figure 37), ou bien sélectionner manuellement des valeurs individuelles considérées comme aberrantes (non illustré ici).

- Une méthode automatique qui repose sur l'utilisation de l'algorithme de lissage de Savitsky-Golay. Différents paramètres peuvent y être spécifier, tels que :
    - La taille de la fenêtre de lissage
    - Le cut-off qui est défini comme étant l'écart type de la distance entre les différents points de données et la courbe lissée.
    - Le nombre d'itérations de la procédure de lissage

La Figure 37 montre un exemple de courbe d'absorbance sur laquelle les valeurs aberrantes ont été supprimées, grâce à la méthode manuelle utilisant la courbe de tendance.

Figure 37 Courbe d'absorbance lissée manuellement grâce à Wellinverter. La courbe de tendance (points en orange) détermine une zone (fond orange) permettant de sélectionner pour l'analyse les valeurs d'absorbance utiles (points en bleu foncé) et d'exclure les valeurs d'absorbance aberrantes (en bleu clair). La largeur de la zone de sélection (fond orange) est constante. L'échelle logarithmique de l'absorbance donne l'impression d'une tolérance plus grande à de basses valeurs d'absorbance.

Chapitre 3 : Plateforme génétique et instrumentale pour la caractérisation *in vivo* du fonctionnement des commutateurs ARN : mise en œuvre initiale et optimisation

1. Choix des composants/éléments initiaux de la plateforme

1.1. Origines de réplication et promoteurs de la transcription

Pour que la plateforme de test soit fonctionnelle et s'assurer que l'antisens ARN régule effectivement le commutateur ARN, il faut se mettre dans des conditions où l'on puisse garantir que la fréquence de rencontre entre l'antisens ARN et le commutateur ARN soit suffisamment rapide pour éviter une bifurcation possible du chemin de repliement de l'ARN. Nous avons vu dans le chapitre 1 que pour atteindre cette condition, il faudrait que la concentration en antisens ARN soit en excès devant la concentration en commutateur ARN.

Un premier paramètre qui permettrait de remplir cette condition serait l'utilisation d'origines de réplication différentes pour réguler la réplication des plasmides d'expression du commutateur ARN et de l'antisens ARN, de façon que le nombre de copies du plasmide portant l'antisens soit largement supérieur au nombre de copies du plasmide portant le commutateur.

Pour un premier essai, nous choisissons d'utiliser une origine de réplication ColE1 pour réguler le nombre de copies du plasmide d'expression de l'antisens ARN et une origine de réplication pSC101 pour le plasmide d'expression du commutateur ARN. ColE1 produit environ 50 copies du plasmide par bactérie alors que pSC101 n'en produit qu'environ 5 copies.

Un deuxième paramètre à prendre en considération pour remplir cette condition serait d'utiliser des promoteurs de forces équivalentes ou tout du moins telles que la force du promoteur qui contrôle l'expression du commutateur ARN ne soit pas tellement plus importante comparée à la force du promoteur qui contrôle l'expression de l'antisens. Et ainsi d'éviter qu'elle ne contrebalance la forte expression de l'antisens ARN favorisée par l'utilisation d'une origine de réplication conférant un fort nombre de copies du plasmide.

Un promoteur [95] est une séquence d'ADN à partir de laquelle s'initie la transcription d'un gène en ARN par l'intermédiaire d'une ARN polymérase. Le promoteur dicte le taux de transcription et donc le niveau d'expression du gène, mais aussi l'endroit à partir duquel la transcription doit commencer et qui est définit comme étant le site +1 de la transcription. Certains promoteurs peuvent déborder au-delà de ce site d'initiation (+1) à cause de la présence d'une séquence appelée DSR (Downstream Region) qui joue aussi un rôle dans la force du promoteur. Pour la plateforme que nous envisageons d'utiliser, il est cependant important que

le promoteur ne déborde pas au-delà du +1, car la séquence qui en serait transcrit pourrait interférer avec le repliement du commutateur ARN ou avec l'hybridation de l'antisens ARN. Enfin, on distingue deux types de promoteur : les promoteurs constitutifs et les promoteurs régulés. Un promoteur est constitutif lorsque son activité ne dépend que de la présence de l'ARN polymérase ; alors qu'il est régulé lorsque son activité dépend aussi d'un autre facteur appelé facteur de transcription qui permet de moduler son activité. En ce sens, nous choisissons d'utiliser deux promoteurs régulables pour contrôler de manière indépendante (i) la transcription du commutateur ARN : promoteur $p_{LlacO-1}$[3] [96] régulé par la protéine répresseur LacI et inductible par l'IPTG (isopropyl-β-D-thiogalactopyranoside) et (ii) la transcription de l'antisens ARN : promoteur $p_{LtetO-1}$ [96], régulé par la protéine répresseur TetR et inductible par l'aTc (anhydrotetracycline). L'intérêt de cette stratégie est de pouvoir ajuster de façon indépendante les taux de transcription du commutateur ARN et de l'antisens ARN, toujours dans l'optique de pouvoir obtenir une concentration en antisens ARN en excès devant la concentration en commutateur.

1.2. Bactérie hôte : BW25113

Pour notre plateforme, nous choisissons d'utiliser la bactérie *Escherichia coli* pour tester et valider le fonctionnement des commutateurs ARN. Il existe cependant plusieurs souches de la bactérie *Escherichia coli*. Dans notre cas, nous utilisons une souche BW25113 [Δ(*araD-araB*)*567* Δ(*rhaD-rhaB*)*568 ΔlacZ4787* (::rrnB-3) *hsdR514 rph-1*] [97][98] très fréquemment utilisée en laboratoire, maintenue avec une manipulation génétique minimale et par conséquent proche du génotype de la souche sauvage.

1.3. Protéine fluorescente : mKate2

Nous utiliserons comme protéine rapportrice, une protéine fluorescente appelée mKate2. Le choix de cette protéine fluorescente s'explique par sa propriété à émettre dans le rouge (pic d'excitation à λ_{ex} = 588 nm et pic d'émission, λ_{em} = 633 nm) [99] (Figure 38) ; ainsi, son spectre de fluorescence ne se recouvre pas avec la fluorescence intrinsèque de la bactérie, qu'on appelle autofluorescence verte (λ_{ex} = 470 nm et λ_{em} = 525 nm), comme cela pourrait l'être dans le cas de la GFP (Green Fluorescent Protein, λ_{ex} = 395 et 477 nm et λ_{em} = 509 nm), facilitant

[3] La séquence de $p_{LlacO-1}$ que nous utilisons est une version corrigée de celle publiée par Lutz et Bujard [96], dans laquelle l'opérateur lacO-2 a été par erreur inséré à la place de l'opérateur lacO-1, ce qui peut avoir pour conséquence une diminution de la répression du promoteur par le répresseur LacI, entraînant une fuite significative de la transcription d'ARN en absence d'IPTG [75].

grandement l'analyse des données que nous avons vu précédemment dans le chapitre 2 « Mise en œuvre expérimentale et analyse des données : principes ».

Le choix de cette protéine fluorescente se justifie également par sa forte brillance qui permet d'obtenir une meilleure sensibilité de détection et aussi pour son absence de toxicité pour la cellule [99].

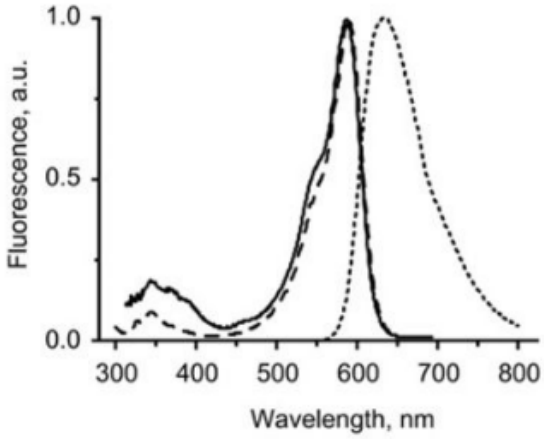


Figure 38 Spectre d'absorption (en trait plein), spectre d'excitation (en tirets) et spectre d'émission (en pointillés) de la protéine fluorescente mKate2. Longueur d'onde théorique du pic d'excitation = 588 nm et du pic d'émission = 633 nm [99].

1.4. Terminateur de la transcription de l'antisens ARN : $T1_{rrnB}$

Les antisens ARN utilisés pour notre étude, ont une taille d'environ 20 nucléotides. Pour réaliser nos tests, il est primordial que ces antisens puissent s'hybrider avec leurs cibles sur les commutateurs ARN. Dans notre système un terminateur fort est utilisé pour arrêter la transcription des antisens ARN et éviter de produire des transcrits trop longs qui risqueraient de piéger la région antisens dans des structures secondaires et qui interféreraient avec l'hybridation de l'antisens à sa région cible sur le commutateur ARN. Nous choisissons d'utiliser le terminateur $T1_{rrnb}$, un terminateur fort et dont l'efficacité de terminaison a été récemment démontrée *in vivo* par Cambray *et al.* [100].

La Figure 39 récapitule tous les composants qui ont été nécessaires pour construire les plasmides d'expression du commutateur ARN et de l'antisens ARN.

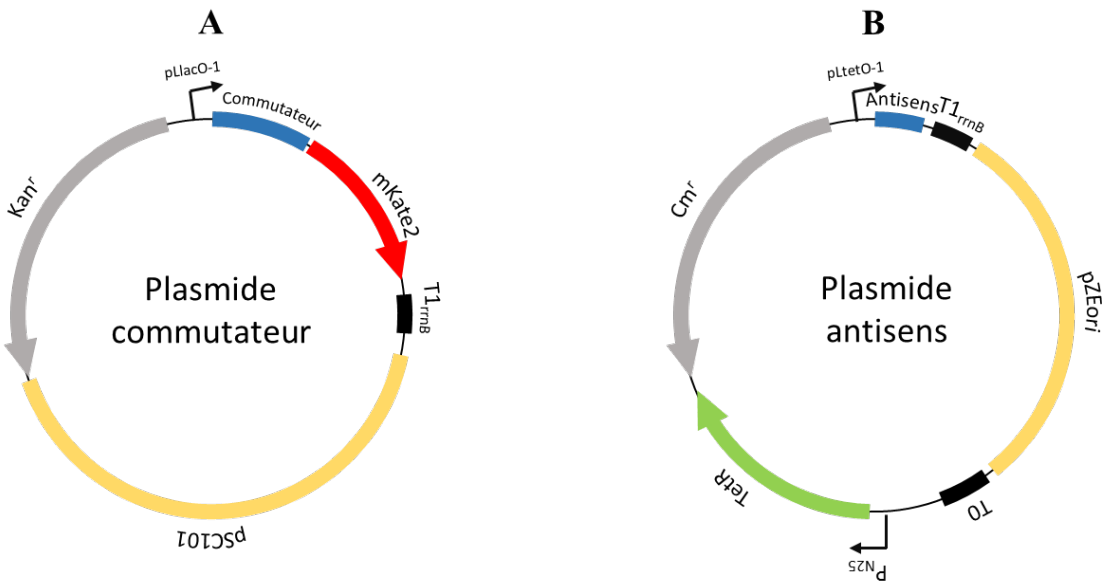


Figure 39 Schémas de l'architecture des plasmides utilisés pour la caractérisation de l'état de fonctionnement des commutateurs ARN. (A) Plasmide contenant le commutateur ARN, lui-même en aval d'un promoteur $p_{LlacO-1}$ inductible à l'IPTG. Ce commutateur est suivi d'un gène rapporteur qui code pour la protéine fluorescente mKate2. Une origine de réplication pSC101 (≈ 5–10 copies) est utilisée pour sa réplication. La présence du gène de résistance à la kanamycine, Kan^r sur la figure, confère à la bactérie portant ce plasmide une résistance à cet antibiotique. **(B)** Plasmide contenant l'antisens, lui-même en aval d'un promoteur $p_{LtetO-1}$ inductible à l'anhydrotétracycline aTc et en amont d'un terminateur fort $T1_{rrnB}$, assurant l'arrêt de la transcription. Une origine de réplication de type ColE1ori (pZEori = pBR322, Δrop/rom ≈ 50 copies) [96][101] est utilisée pour sa réplication. Le promoteur $p_{LtetO-1}$ est régulé par la protéine TetR, produite de manière constitutive à partir du même plasmide, via un promoteur constitutif p_{N25} [96][102]. La présence du gène de résistance au chloramphénicol, Cm^r sur la figure, confère à la bactérie portant ce plasmide une résistance à cet antibiotique.

2. Test des plateformes d'expression : fonctionnement des promoteurs inductibles $p_{LlacO-1}$ et $p_{LtetO-1}$

2.1. Introduction

Comme cela a été vu précédemment, nous utilisons les promoteurs inductibles $p_{LlacO-1}$ et $p_{LtetO-1}$ pour contrôler respectivement l'expression du commutateur ARN et de l'antisens ARN. Dans cette partie, nous aborderons séparément le test des deux plateformes utilisées, pour l'expression du commutateur d'une part et l'expression de l'antisens d'autre part, afin de vérifier le fonctionnement de chacune des composantes de la plateforme. Plus particulièrement, nous voulons vérifier à travers ces tests les niveaux de répression et d'induction effectifs que nous sommes capables d'obtenir pour l'expression du commutateur et de l'antisens. Nous voulons également vérifier les variations de fluorescence pouvant être attendues en cas de fonctionnement du commutateur ARN. La variation de fluorescence obtenue par la régulation de l'activité du promoteur, pour l'expression de rSw1, peut alors servir de référence à la

variation de fluorescence qu'on peut s'attendre à obtenir entre les états « ON » et « OFF » de l'ensemble des commutateurs ARN.

Ce test permet donc d'une part de vérifier si la variation de fluorescence est détectable et d'autre part d'en obtenir une amplitude typique. En effet, il est important de préciser que le niveau de fluorescence n'est qu'une lecture indirecte de l'état « ON » ou « OFF » du commutateur, du fait que cet état est déterminé au niveau de l'ARN alors que le signal de fluorescence est obtenu seulement après une étape de traduction ; or il se pourrait que des effets non-linéaires (comme par exemple une saturation de la machinerie traductionnelle) affectent la réponse dynamique entre la régulation au niveau « ARN » et la régulation au niveau « protéine ». Par la suite, nous verrons des résultats suggérant que si une non-linéarité est observée, elle n'est pas due à la gamme d'expression explorée ici.

Toujours concernant le commutateur ARN, nous voulons aussi vérifier ici qu'il est effectivement possible d'ajuster son niveau d'expression, afin, le cas échéant, de pouvoir le régler de façon à optimiser l'effet de l'antisens ARN.

2.2. Induction du promoteur $p_{LlacO-1}$: contrôle du niveau de transcription du système commutateur ARN-gène rapporteur

Nous utiliserons pour ce test une souche BW25113 transformée avec un plasmide contenant le commutateur répressible rSw1 suivi du gène rapporteur mKate2, tous les deux contrôlés par le promoteur $p_{LlacO-1}$ (pGEJ1703) (**Figure 39A** et **Annexe A : Liste des plasmides et souches utilisés**). Ici, sans l'expression de l'antisens as1, le commutateur rSw1 serait à priori dans un état « ON ». Pour vérifier le fonctionnement du promoteur $p_{LlacO-1}$, l'idéal aurait été cependant d'utiliser une plateforme à vide, c'est-à-dire sans le commutateur entre le promoteur et le gène rapporteur, ce qui n'a pas été fait dans cette étude.

Ainsi, pour vérifier les niveaux de répression et d'induction que nous sommes capables d'obtenir pour l'expression de rSw1, nous testons plusieurs concentrations d'IPTG : 0, 1, 10, 25, 50, 100 et 1000 µM.

Pour pouvoir comparer les résultats que nous obtenons, nous devons aussi nous assurer que les différentes conditions utilisées n'affectent pas la croissance des bactéries. Pour chaque condition étudiée, nous mesurons effectivement un temps de doublement[4] identique des

[4] Le temps de doublement est le temps qu'il faut pour doubler la densité bactérienne pendant la phase exponentielle. Autrement dit, il s'agit du temps qui sépare deux divisions cellulaires successives.

bactéries, d'environ 1 heure. Leur croissance n'est donc affectée ni par l'introduction du plasmide, ni par l'ajout d'IPTG.

Les intensités de fluorescence moyenne par bactérie (Fluorescence/DO_{600}) sont utilisées pour déterminer la courbe d'induction expérimentale par l'IPTG, à l'état stationnaire. Pour chaque concentration d'IPTG utilisée, nous mesurons une fluorescence moyenne par bactérie qui correspond au niveau d'expression du gène rapporteur lorsque le commutateur rSw1 est dans son état « ON » (absence ici d'antisens ARN effecteur). En absence d'IPTG, la fluorescence moyenne par bactérie que nous mesurons est étonnement élevée (≈ 2000 UA) tout en sachant que le promoteur $p_{LlacO-1}$ n'est pas induit. Ensuite, jusqu'à 10 µM d'IPTG, la fluorescence moyenne par bactérie varie très peu, et augmente rapidement entre 10 et 100 µM d'IPTG (entre 2500 et 3500 UA), avant de saturer à partir de 100 µM d'IPTG (≈ 4000 UA) (Figure 40).

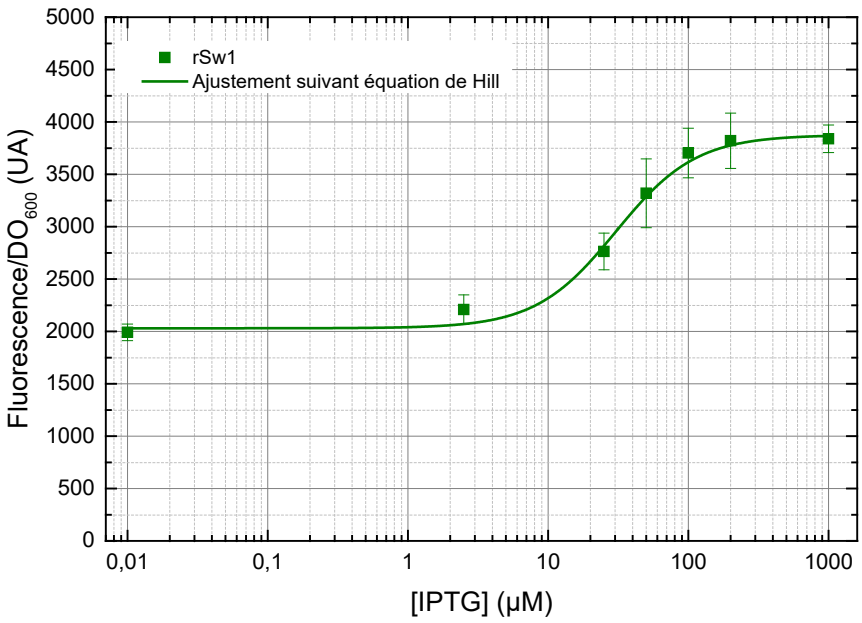


Figure 40 Courbes d'induction par l'IPTG de la plateforme rSw1-mKate2 sous contrôle du promoteur $p_{LlacO-1}$. Les fluorescences moyennes par bactérie (Fluorescence/DO_{600}), prises pendant la croissance équilibrée, sont tracées en fonction de la concentration d'IPTG. Les barres d'erreur correspondent à des écarts-types, obtenus à partir de 3 réplicats. La courbe en ligne continue est construite suivant l'équation de Hill *(8)*, les meilleurs paramètres d'ajustement sont donnés dans le Tableau 4. Le K_M est d'environ 20 µM. Croissance en milieu minimum M9 + 0,2 % glucose + 40 µg/ml de kanamycine.

Ces données sont ensuite ajustées suivant l'équation de Hill :

$$F = F0 + \frac{F\max - F0}{1 + \left(\frac{[IPTG]}{K}\right)^n} \qquad (8)$$

F et F_0 représentent respectivement les fluorescences moyennes par bactérie mesurées en présence et en absence d'IPTG. F_{max} est la fluorescence maximale. [IPTG] représente la concentration d'IPTG utilisée pour induire la production de fluorescence via le contrôle du promoteur $P_{LlacO-1}$. K correspond à la concentration d'IPTG nécessaire pour atteindre 50 % d'induction de l'expression de la fluorescence. n est le coefficient de Hill.

Chi^2	0,50	
R^2	0,99	
	Valeurs	Écarts-types
$\boldsymbol{F0}$	2030	52
$\boldsymbol{F}max$	3875	98
$\boldsymbol{K}$	30,4	5,5
$\boldsymbol{n}$	1,52	0,51

Tableau 4 Paramètres utilisés pour l'ajustement des données d'induction avec la courbe construite à partir de l'équation de Hill *(8)*.

Ces résultats indiquent que nous sommes en effet capables de réguler l'expression de la protéine rapportrice via le promoteur inductible $p_{LlacO-1}$. Cependant nous observons que la gamme dynamique de régulation que nous obtenons est assez faible, soit d'un facteur 2, due notamment à une forte expression du gène rapporteur dès 0 µM d'IPTG. Nous reviendrons sur ce problème dans la partie IV, chapitre 2 « Optimisation des commutateurs ARN et des antisens ARN ».

2.3. Induction du promoteur $p_{LtetO-1}$: contrôle du niveau de transcription de l'antisens ARN (antisens I[5])

L'expression des antisens ARN se fait au moyen d'un plasmide à relativement fort nombre de copies (ColE1ori ≈ 50 copies/bactérie) et où le gène de l'antisens ARN est sous contrôle du promoteur $p_{LtetO-1}$ régulé par le répresseur TetR (dont le gène est également présent sur le même plasmide sous contrôle d'un promoteur constitutif fort p_{N25}). Par conséquent, la transcription des antisens ARN à partir du promoteur $p_{LtetO-1}$ est inductible à l'anhydrotétracycline (aTc).

Pour tester le fonctionnement de la plateforme d'expression des antisens et notre capacité à réguler leurs niveaux d'expression via le promoteur $p_{LtetO-1}$, nous utilisons un plasmide (pGEJ1711) similaire aux plasmides utilisés pour exprimer les antisens mais sur lequel nous avons remplacé la séquence antisens par un gène rapporteur qui exprime dans notre cas la protéine fluorescente mKate2. Nous utilisons une souche BW25113 transformée avec ce

[5] Antisens I fait référence à la première version du vecteur d'expression des antisens

plasmide donnant naissance à une souche qu'on appelle GEJ078 (**Figure 39B** et **Annexe A : Liste des plasmides et souches utilisés**).

Pour vérifier la dynamique de régulation qu'il est possible d'atteindre pour ce promoteur, nous testons plusieurs concentrations d'aTc : 0 ; 0,5 ; 5 ; 12 ; 25 ; 50 et 100 ng/ml. Et bien sûr comme précédemment, il est important de vérifier que chaque condition utilisée n'affecte pas la croissance des bactéries afin de pouvoir comparer les résultats obtenus pour les différentes concentrations d'aTc. La **Figure 41** présente les courbes de croissance de bactéries portant soit un plasmide exprimant l'antisens ARN as1, soit un plasmide où l'antisens ARN as1 est remplacé par le gène codant mKate2. Comparée à la croissance de la souche sauvage BW25113, on constate que l'introduction de plasmides affecte la croissance bactérienne. Nous mesurons, pour les souches avec plasmides un temps de doublement presque 4 fois supérieur (3 h 45 min) au temps de doublement de la souche sauvage BW25113 (1 heure).

B. Glick *et al.* [**103**] expliquent que l'introduction d'ADN étranger dans un organisme hôte peut parfois nuire au fonctionnement métabolique normal de cet organisme. Il s'agirait d'une conséquence de la charge métabolique imposée à l'hôte par l'ADN qui y a été introduit. Une cellule hôte contenant un plasmide recombinant est obligée d'affecter une partie de ses ressources limitées au maintien et à la réplication du plasmide, ainsi qu'à l'expression des gènes codés par le plasmide [**104**]. Cette charge métabolique est d'autant plus importante que le nombre de copies du plasmide dans l'organisme est fort [**105**], comme c'est le cas pour notre plateforme d'expression (ColE1ori ≈ 50 copies/bactérie). A cela s'ajoute une forte expression du répresseur TetR, également exprimé à partir du même plasmide et sous le contrôle d'un promoteur constitutif fort p_{N25}, qui augmenterait d'autant plus la charge métabolique.

Tous ces facteurs combinés peuvent expliquer pourquoi les croissances de ces nouvelles souches sont très ralenties en milieu minimum M9.

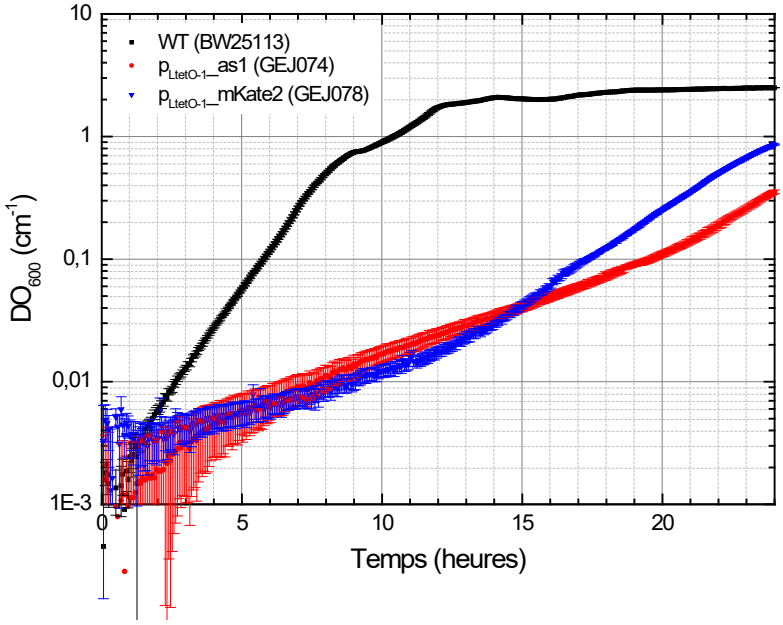


Figure 41 L'introduction du plasmide d'expression de l'antisens dans la bactérie réduit fortement sa croissance en milieu minimum M9 (DO_{600} en fonction du temps). Comparaison de la vitesse de croissance de la souche sauvage *E. coli* BW25113, de la souche GEJ074 portant un plasmide exprimant l'antisens ARN as1, et la souche GEJ078 portant un plasmide exprimant la protéine rapportrice mKate2. Les barres d'erreur correspondent aux écarts-types, obtenus à partir de 3 réplicats. Culture dans un milieu minimum M9 + 0,2 % glucose + 30 µg/ml chloramphénicol. Les temps de doublement sont de 3 h 45 min pour les souches GEJ074 et GEJ078, et de 1 h pour la souche sauvage BW25113.

Dans l'hypothèse de la charge métabolique, un apport d'une certaine quantité d'énergie permettrait donc de maintenir l'expression d'un ARN étranger, tout en limitant les effets sur la croissance. Des travaux visant à améliorer la production de protéines recombinantes à partir de plasmide recombinant montrent que l'utilisation d'un milieu de culture riche peut aider à améliorer le rendement de production des protéines tout en préservant la croissance des bactéries [104]. L'apport externe de ressources telles que des acides aminés ou des nucléotides [106] semble compenser l'épuisement des ressources fournies par la bactérie lorsqu'une charge métabolique est en jeu.

Dans la partie qui suit, nous allons montrer qu'il s'agit bien ici d'un problème de charge métabolique, probablement dû à une trop forte expression de TetR.

2.4. Restauration de la croissance des bactéries, en milieu riche défini (M9 + suppléments)

En microbiologie, les milieux de culture riches tels que le LB (Lysogeny Broth) sont très largement utilisés, car ils favorisent une croissance rapide des bactéries et un bon rendement de

croissance [80]. Cependant la composition moléculaire d'un tel milieu n'est pas toujours définie avec précision, ils sont constitués de composants complexes comme les extraits de levure qui peuvent varier d'un lot à l'autre [107].

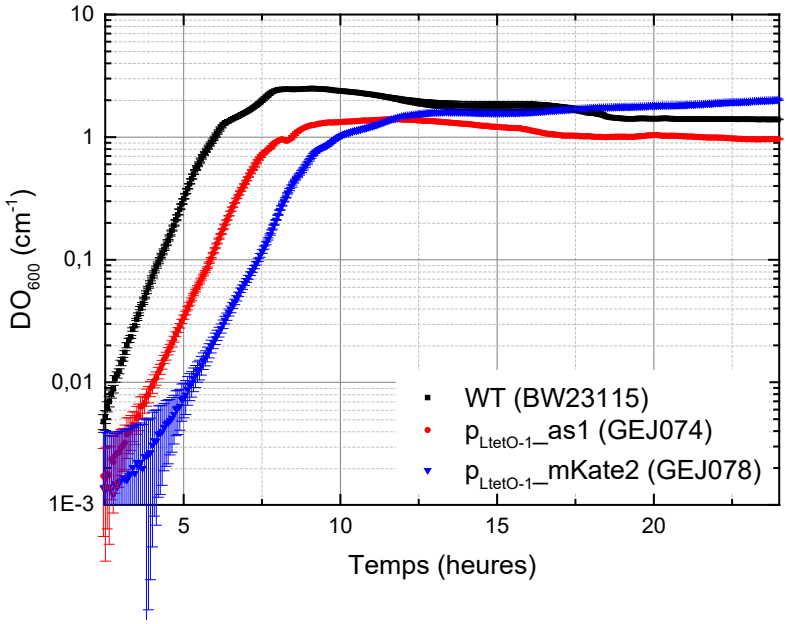


Figure 42 L'utilisation d'un milieu riche défini (M9 + suppléments) restaure la croissance des bactéries qui contiennent le plasmide d'expression de l'antisens (DO_{600} en fonction du temps). Comparaison de la vitesse de croissance de la souche sauvage *E. coli* BW25113, de la souche GEJ074 portant un plasmide exprimant l'antisens ARN as1, et la souche GEJ078 portant un plasmide exprimant la protéine rapportrice mKate2. Les barres d'erreur correspondent aux écarts-types, obtenus à partir de 3 réplicats. Culture dans un milieu minimum M9 + 0,2 % glucose + 30 µg/mL chloramphénicol + 1X EZ (acides aminés). Le temps de doublement est de ≈ 35 min pour les 3 souches.

Ainsi nous utiliserons, comme Neidhardt *et al.* [108] un milieu de culture minimum, dans notre cas un milieu minimum M9, de composition bien définie que nous complétons avec un mélange d'acides aminés et/ou de nucléotides, également de composition bien définie. On parle alors de milieu riche défini (**Annexe B : Méthodes expérimentales, 1**). Le temps de doublement des bactéries *E. coli* y est comparable au temps de doublement qu'on mesure dans une culture en milieu riche LB [106], tandis que l'autofluorescence reste très faible [108], ce qui est idéal pour les mesures de fluorescence.

Ainsi, nous mesurons un temps de doublement d'environ 35 min pour les souches GEJ074 et GEJ078, similaire au temps de doublement de la souche sauvage *E. coli* BW25113 (Figure 42).

2.5. Induction du promoteur $p_{LtetO-1}$ par l'anhydrotetracycline (aTc) en milieu riche défini

L'utilisation d'un milieu riche défini (M9 + suppléments d'acides aminés et/ou d'acides nucléiques) a permis de restaurer la croissance des bactéries contenant les plasmides d'expression de l'antisens. C'est-à-dire qu'elles ont un temps de doublement similaire à celui de la souche sauvage BW25113. Dans ces conditions, on devrait donc pouvoir vérifier une nouvelle fois les niveaux de répression et d'induction qu'il est possible d'atteindre pour le promoteur $p_{LtetO-1}$. Nous testons pour cela différentes concentrations d'aTc : 0, 12, 25, 50, 100, 200 et 400 ng/mL pour l'induction. Pour pouvoir comparer les résultats obtenus, il faut bien évidemment s'assurer que les différentes conditions utilisées n'affectent pas la croissance des bactéries. En effet, jusqu'à 200 ng/mL d'aTc, la croissance des bactéries n'est pas affectée et elles ont un temps de doublement similaire à la souche sauvage BW25113 (≈ 35 min). Au contraire, au-delà de cette concentration, leur temps de doublement devient plus lent (≈ 50 min pour 400 ng/ml d'aTc) (**Figure 43A**). Ces résultats indiquent que pour de fortes concentrations, soit l'aTc devient toxique pour les bactéries, soit l'expression du gène rapporteur devient trop forte, ralentissant ainsi leur vitesse de croissance.

Par ailleurs, le protocole utilisé pour réaliser ces expériences n'est pas encore le protocole final optimisé, avec notamment l'usage des billes ou encore les bonnes conditions pour atteindre la croissance équilibrée. Nous pouvons observer que les courbes de fluorescence moyenne par bactérie n'ont pas l'allure en plateau horizontal typique qu'on devrait observer lorsque les bactéries sont en croissance équilibrée se trouvant normalement ici entre ≈ 6h et 9h de croissance **Figure 43B**. Au contraire, nous constatons que les courbes sont parasitées par des bruits de fond, les rendant difficilement exploitables. Malgré cela, nous constatons que cette fluorescence moyenne par bactérie est particulièrement basse, même avec une forte concentration en inducteur (**Figure 43B**). D'ailleurs c'est ce qui expliquerait la mauvaise qualité du rapport signal/bruit que nous observons. Nous supposons dans ce cas précis que le promoteur $p_{LtetO-1}$ est très peu induit même en présence d'une forte concentration d'aTc. En effet, pour la plateforme que nous utilisons, le répresseur TetR est exprimé à partir d'un plasmide à fort nombre de copies (ColE1 ≈ 50 copies/bactéries), alors que Lutz et Bujard [96] le font exprimer à partir d'un chromosome bactérien (1 seule copie). On suppose que le répresseur TetR est exprimé en trop forte quantité pour que l'aTc, même à forte dose, puisse lever la répression.

Dans la partie suivante, nous allons présenter une seconde version de la plateforme d'expression des antisens ARN qui vise à corriger ces problèmes.

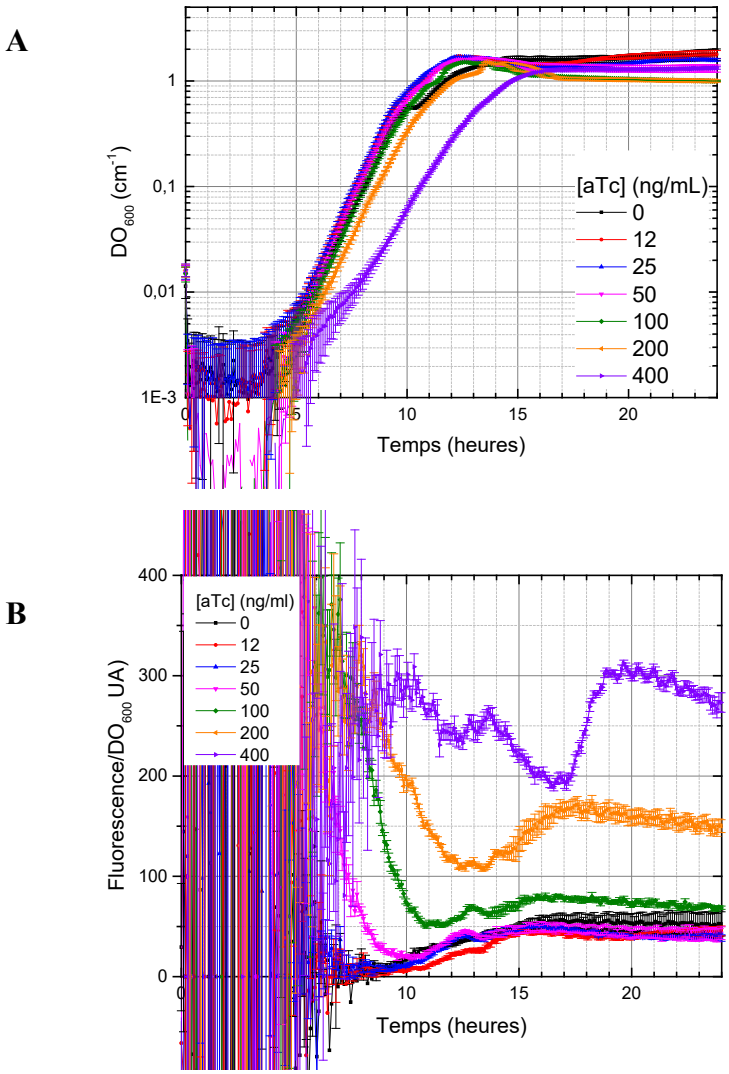


Figure 43 (A) L'effet de l'aTc sur la croissance bactérienne. Comparaison de la vitesse de croissance de la souche GEJ078 ($p_{LtetO-1}$_mKate2), pour différentes concentrations d'aTc. **(B) Niveau d'induction du promoteur $p_{LTetO-1}$ pour différentes concentrations d'aTc**. Le répresseur TetR est exprimé à partir du même plasmide d'expression que l'antisens, c'est un plasmide à fort nombre de copies (ColE1 ≈ 50 copies/bactérie). Les barres d'erreur correspondent à des écarts-types, obtenus à partir de 3 réplicats. Culture dans un milieu minimum M9 + 0,2 % glucose + 30 µg/ml chloramphénicol + 1X EZ.

3. Modification du vecteur d'expression des antisens I : expression constitutive et fort nombre de copies

Les résultats précédents montrent que la plateforme initialement pensée pour l'expression de l'antisens n'est pas encore parfaitement fonctionnelle. Il semblerait que le promoteur $p_{LtetO-1}$ soit très fortement réprimé par son répresseur TetR lui-même très fortement exprimé. A un tel point que même une forte concentration en aTc ne permette d'induire que très faiblement l'activité du promoteur.

Nous décidons alors de changer de vecteur d'expression, de telle façon à enlever la répression du promoteur $p_{LtetO-1}$ par le répresseur TetR et augmenter la quantité d'antisens exprimé dans la bactérie, grâce à l'utilisation d'une origine de réplication qui permette d'obtenir un très grand nombre de copies du plasmide (pUCori ≈ 300 – 500 copies/bactérie) (**Figure 44**). De cette manière, nous obtenons une expression constitutive de l'antisens par le promoteur $p_{LtetO-1}$ sans qu'aucune molécule inductrice (aTc) ne soit nécessaire De plus, une forte expression de l'antisens est favorisée par l'utilisation de l'origine de réplication pUCori. Nous appellerons antisens II cette version modifiée du vecteur d'expression de l'antisens.

3.1. Test du fonctionnement de la nouvelle plateforme d'expression de l'antisens (antisens II) : test du niveau d'expression de l'antisens via la protéine rapportrice mKate2

Comme précédemment, nous voulons mesurer le niveau d'expression de l'antisens que nous sommes capables d'atteindre pour cette nouvelle plateforme. Pour cela nous remplaçons l'antisens par une protéine fluorescente mKate2, dont la mesure de l'intensité permettra d'avoir une idée du niveau d'expression de l'antisens que nous sommes capables d'atteindre.

Les modifications apportées à la nouvelle plateforme d'expression de l'antisens favoriseraient à priori une expression forte et constitutive de l'antisens. Cependant, les mesures en microplaque montrent qu'aucune fluorescence n'est produite, bien que l'on observe une très forte fluorescence de la culture de démarrage (**Figure 45A**).

Par ailleurs, sur une culture en milieu solide (LB + Agar), nous constatons que ces mêmes souches peuvent produire de la fluorescence. Il semblerait cependant que cette production de fluorescence n'est pas homogène parmi tous les clones que nous obtenons (**Figure 45B**). Certains clones sont complètement rouges, alors que d'autres sont rouges au centre et blancs en périphérie et d'autres complètement blancs. De plus, en réensemençant certains clones rouges de cette même boîte, nous observant une perte progressive de la fluorescence.

Nous ne sommes pas capables de déterminer avec exactitude l'origine de cette hétérogénéité, mais nous supposons qu'il s'agit d'une toxicité due à une forte expression de la protéine fluorescente mKate2 qui, rappelons-le, est exprimée à partir d'un plasmide à fort nombre de copies (pUCori ≈ 300 – 500 copies/bactérie) et sous le contrôle d'un promoteur fort ($p_{Ltet-O1}$ dérivé du promoteur p_L du phage lambda). Par ailleurs, la perte progressive de la fluorescence après réensemencement de clone rouge renforce encore plus l'idée d'une toxicité.

Cette toxicité peut avoir plusieurs conséquences :

- Mutations qui diminueraient la transcription et/ou la traduction de la protéine
- Perte du plasmide d'expression (peu probable car les bactéries restent résistantes à l'antibiotique de sélection exprimé seulement à partir du plasmide d'expression)
- Accumulation de mutations dans la séquence de la protéine la rendant moins toxique, mais qui peuvent avoir un effet sur le repliement et par conséquent la structure native de la protéine qui est déterminant pour sa fonction et donc l'émission de fluorescence.

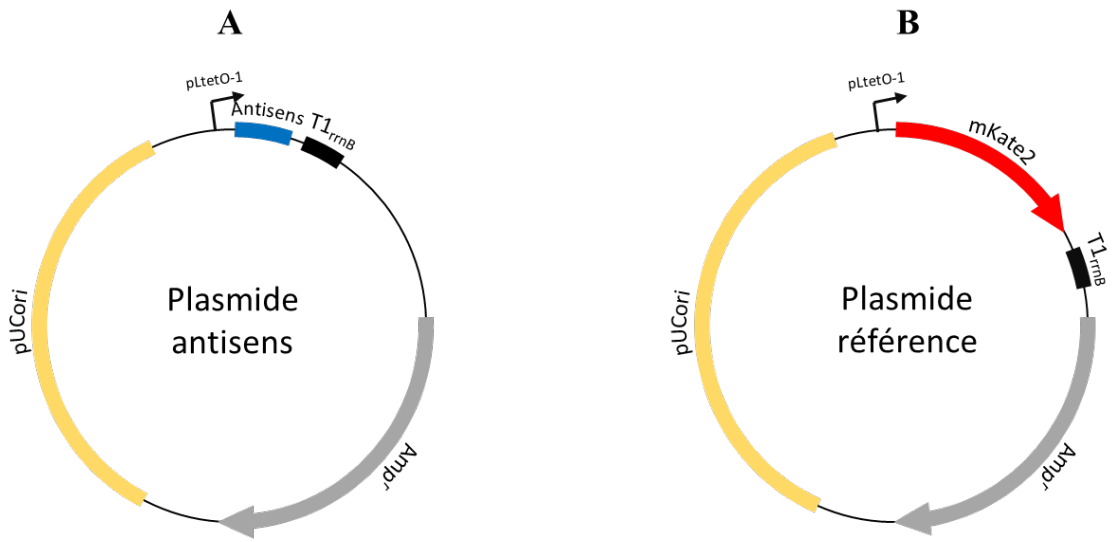


Figure 44 Schémas de la nouvelle architecture du plasmide d'expression de l'antisens. (A) Plasmide d'expression de l'antisens, lui-même en aval d'un promoteur $p_{LtetO-1}$ et en amont d'un terminateur fort $T1_{rrnB}$, assurant l'arrêt de la transcription. Une origine de réplication pUCori (≈ 300 – 500 copies/bactérie) est utilisée pour sa réplication. Ici, le promoteur $p_{LtetO-1}$ n'est plus régulé par la protéine TetR, car elle n'est plus produite dans la bactérie. La présence du gène de résistance à l'ampicilline, Amp^r sur la figure, confère à la bactérie portant ce plasmide une résistance à cet antibiotique. **(B)** Plasmide de référence (pGEJ1716) utilisé pour vérifier le niveau d'expression qu'il est possible d'atteindre avec le promoteur $p_{LtetO-1}$. Il s'agit d'un plasmide similaire au plasmide d'expression de l'antisens, sur lequel la séquence de l'antisens a été remplacé par la séquence d'un gène rapporteur fluorescent mKate2.

A **B**

Figure 45 Vérification du niveau d'expression qu'il est possible d'atteindre avec un promoteur $p_{LtetO-1}$, ici constitutivement actif (sans la répression par TetR). (A) Culture de démarrage de la souche contenant le plasmide de référence pGEJ1716 (pLtetO-1_mKate2_pUCori) avant l'ensemencement en microplaque. **(B)** On représente 2 clones différents issus de la transformation de la souche BW25113 avec le plasmide de référence pGEJ1716. On observe différents phénotypes, pourtant issus du même clone.

Avec ces résultats, il nous est impossible d'estimer le niveau d'expression *in vivo* des antisens ARN, car la fluorescence de mKate2 utilisée pour indirectement estimer cette expression se perd à cause d'une possible toxicité de la forte expression de cette protéine dans la bactérie, due probablement à une charge métabolique trop importante. Nous pourrions cependant utiliser d'autres techniques pour estimer ce niveau d'expression. On peut citer, par exemple, l'utilisation d'aptamères fluorogéniques, ou encore des méthodes telles que le Northern blot, la qRT-PCR qui permettent de doser directement l'ARN et non plus la protéine comme c'était le cas avec mKate2. L'avantage de ces différentes méthodes est qu'elles permettent d'estimer directement le niveau de transcription des antisens ARN, contrairement à la méthode précédente où l'on doit prendre en compte la traduction de l'ARN en protéine. Cependant, la mise au point d'un dosage d'une petite molécule peut s'avérer particulièrement difficile et peut prendre beaucoup de temps à mettre en place. Nous faisons alors le choix de remettre à plus tard cette estimation du niveau d'expression. En revanche, nous décidons dans la partie suivante de réaliser des transcriptions *in vitro*, en utilisant comme matrice les plasmides d'expression des antisens pour vérifier leur intégrité. Les résultats obtenus doivent cependant être pris avec précaution car les expériences faites *in vitro* et *in vivo* ne sont pas directement comparables. Par exemple, l'état de surenroulement du plasmide peut être différent *in vitro* et *in vivo* et conduire à un comportement de transcription différent pour l'ARN polymérase.

4. Modification du vecteur d'expression de l'antisens II : optimisation du promoteur et de sa séquence avale (antisens III)

4.1. Transcription *in vitro* des antisens as1 et as13 à partir de leur vecteur d'expression respectif

La caractérisation du vecteur d'expression de l'antisens faite précédemment n'a pas été concluante car la forte expression de la protéine rapportrice mKate2 utilisée pour indirectement estimer l'expression de l'antisens via le même vecteur semble être toxique pour la bactérie. Une question reste alors en suspens : est-ce qu'une forte expression de l'antisens via ce même vecteur d'expression serait également toxique pour la bactérie, à un tel point qu'on perde l'expression de l'antisens comme c'est le cas avec mKate2 ?

D'autres méthodes permettraient éventuellement de doser l'ARN *in vivo* et de répondre à cette question, toutefois nous faisons l'hypothèse que cette toxicité ne s'applique pas au plasmide exprimant l'antisens ARN. C'est pourquoi nous choisissons, pour la suite, de continuer avec le même vecteur. Il faut cependant s'assurer que l'antisens y est bien exprimé et que la transcription de l'antisens s'arrête bien au terminateur prévu à cet effet, car la production de transcrits trop longs risquerait de piéger la région antisens dans des structures secondaires qui interfèreraient avec l'hybridation de l'antisens sur le commutateur. Pour vérifier toutes ces caractéristiques, nous réalisons des transcriptions *in vitro*, qui vont non seulement permettre de vérifier l'expression de l'antisens ARN, mais aussi de vérifier qu'on produise effectivement la bonne taille du transcrit de l'antisens (≈ 83 nucléotides).

Comme cela a été décrit précédemment dans la partie II de ce manuscrit, la transcription *in vitro* est réalisée à 37 °C, avec l'ARN polymérase d'*E. coli*. Cependant, au lieu d'utiliser des produits de PCR, nous utilisons ici comme matrice le plasmide entier à partir duquel est exprimé l'antisens (**Figure 44A**), car nous voulons nous placer dans une condition proche de celle que l'on retrouverait *in vivo*. Les transcrits obtenus sont ensuite séparés sur gel de polyacrylamide en condition dénaturante. Les plasmides que nous utilisons sont : le plasmide d'expression de l'antisens as1 (pGEJ1714), le plasmide d'expression de l'antisens as13 (pGEJ1715) et le plasmide pUC19 à partir duquel nous avons construit le vecteur d'expression des deux antisens. Ce plasmide pUC19 servira de référence pour identifier les transcrits pouvant être produits, en dehors des antisens ARN as1 ou as13.

Théoriquement les tailles des transcrits des antisens as1 et as13 font respectivement 83 et 82 nucléotides qui comprennent une région DSR[6] (Downstream Region), la séquence de l'antisens (as1 ou as13) et le terminateur $T1_{rrnB}$.

Nous observons que les plasmides d'expression de as1 (pGEJ1714) et de as13 (pGEJ1715) produisent chacun un transcrit de taille identique, d'environ 95 nucléotides. On retrouve ce même transcrit parmi les transcrits de pUC19 (**Figure 46**). Or, à priori, pUC19 ne devrait pas produire les transcrits des antisens as1 et as13. En effet, il s'agirait ici d'un transcrit d'antisens ARN (RNAI) synthétisé à partir de l'origine de réplication du plasmide pUC19 (taille théorique 108 nucléotides) et qui joue un rôle important dans la régulation du nombre de copies du plasmide [**109**]. Il semblerait donc que les antisens as1 et as13 ne sont pas transcrits par leurs plasmides d'expression respectifs ou bien ils sont transcrits mais leur transcription ne s'arrête pas au niveau du terminateur $T1_{rrnB}$. En effet, nous pouvons observer une sorte de traînée sur les pistes de pGEJ1714 et pGEJ1715 comparées à la piste de pUC19 (**Figure 46**). Ces traînées pourraient correspondre à plusieurs tailles de transcrits.

Nous observons aussi que le transcrit de l'antisens RNAI est exactement ce que l'on recherche pour les antisens que nous envisageons de produire ; c'est-à-dire un petit antisens ARN intègre dont la transcription s'arrête exactement au terminateur, et un niveau d'expression élevé qui permette de se placer dans des conditions où la concentration en antisens ARN est en excès devant la concentration en commutateur ARN.

Nous décidons alors pour la suite de tester le contexte d'expression (promoteur et terminateur) de l'antisens RNAI pour exprimer les antisens as1 et as13.

[6] DSR ou Downstream Region correspond à une séquence en aval du site d'initiation de la transcription (+1), qui peut jouer un rôle important dans la force du promoteur [**131**]. De plus il s'agit ici d'une séquence qui devrait, à priori, se replier pour former une hélice lors de la transcription. Une telle structure est nécessaire pour protéger la séquence simple brin de l'antisens ARN fragile contre les RNases [**132**].

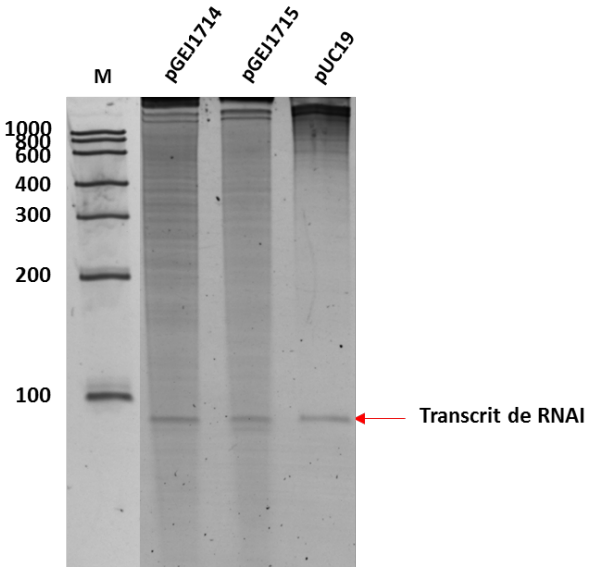


Figure 46 Transcrits d'ARN obtenus à partir de la transcription *in vitro* des plasmides d'expression de l'antisens as1 (pGEJ1714) et as13 (pGEJ1715). Image du gel de polyacrylamide pour la séparation des différents transcrits obtenus à partir de la transcription *in vitro* des plasmides d'expression des deux antisens as1 et as13. Les transcrits issus de pUC19 sont utilisés comme référence pour identifier les transcrits d'ARN produits par pGEJ1714 et pGEJ1715 en dehors des antisens as1 et as13.

4.2. Expression de as1 et as13 via le contexte d'expression de l'antisens RNAI

Afin d'obtenir un profil d'expression similaire à l'antisens RNAI (Figure 46), nous décidons d'utiliser le contexte d'expression (promoteur et/ou terminateur) de cet antisens pour l'expression de as1 et as13. Les travaux d'Hisao Masukata et Jun-ichi Tomizawa [110], ainsi que de Sue Lin-Chao *et al.* [109] ont permis d'identifier la séquence de cet antisens ARN et de son contexte d'expression. Ces études ont notamment permis d'identifier le promoteur de RNAI (p_{RNAI}), le terminateur de la transcription (T_{RNAI}), la séquence de RNAI, mais aussi une séquence DSR (Downstream Region) transcrite tout de suite après le site d'initiation +1 de la transcription (Figure 47A). Comme nous l'avons vu précédemment, les séquences DSR joue un rôle important pour la force du promoteur. Une prédiction du repliement co-transcriptionnel par Kinefold montre également que cette séquence DSR forme une hélice, qui participe probablement à la protection de l'antisens contre les RNases. Pour se placer dans les mêmes conditions d'expression que l'antisens RNAI il est donc important de garder non seulement le promoteur p_{RNAI} mais aussi la séquence DSR pour l'expression des antisens as1 et as13.

A

```
GATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGG
                                                       <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<

TTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTT
                                          [--[[--[    ]--]          ]--]      <+1
                                           1′  2′      2             1
<<<<<<<<<<<<<<<<<<<<<<<<<<<< RNAI <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<

CTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGT
  <-10<<                   <-35<<
<<<<<<<<<< pRNAI <<<<<<<<<<<<<<<<

GGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAG
```

B

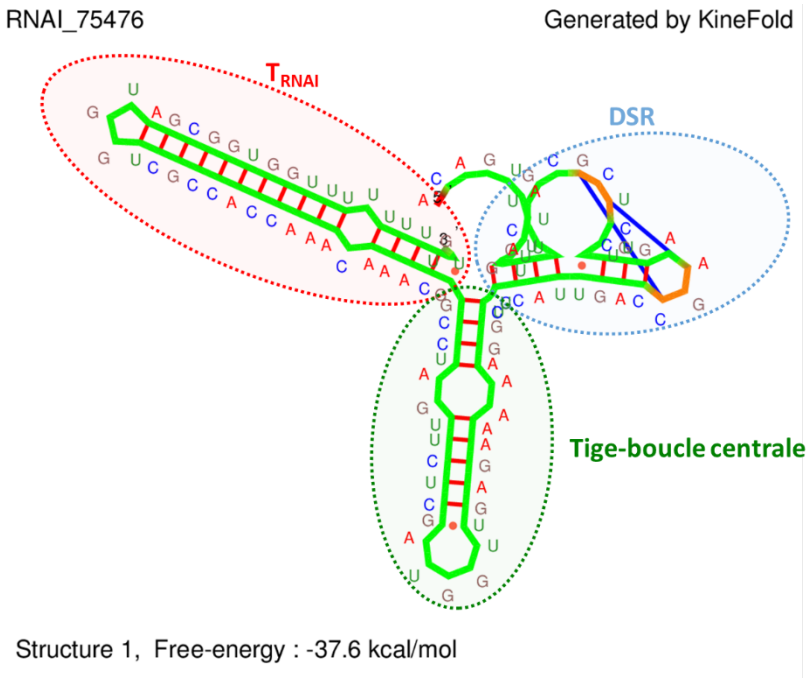
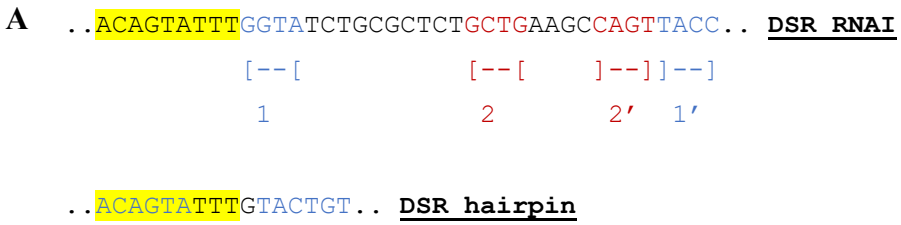


Figure 47 Séquence de l'antisens RNAI et de son contexte d'expression (promoteur et terminateur). (A) La séquence du promoteur pRNAI est représentée en mauve. Elle contient notamment les boîtes -35 et -10 dont on verra plus tard le rôle ; ainsi que le site d'initiation de la transcription +1. La séquence de la région DSR est soulignée en pointillé. Le signe < représente le sens de la transcription dans le plasmide. La séquence de l'antisens RNAI est représentée en orange. Les régions formant des appariements sont matérialisées par une paire d'éléments entre crochets (`[---[ ]---]`) et annotées par un nom d'hélice `1/1′` et `2/2′`. Le terminateur de la transcription de RNAI est souligné en gras. **(B)** Résultat de prédiction du repliement co-transcriptionnel de l'antisens RNAI. Entouré en bleu est représenté le DSR de p_{RNAI}, en orange l'antisens RNAI et en rouge le terminateur de la transcription (T_{RNAI}).

Par ailleurs, nous observons que l'hélice formée par le DSR de p_{RNAI} ne correspond pas tout à fait à une hélice en tige-boucle (Figure 47B). Nous décidons alors de tester un nouveau DSR inspiré du DSR de p_{RNAI} en essayant de garder le contexte d'expression de RNAI, c'est-à-dire garder la séquence du DSR la plus proche possible de l'originale, tout en la modifiant de façon à former une hélice en tige-boucle qu'on appellera par la suite $DSR_{hairpin}$. Les hélices en tige-boucle situées de part et d'autre d'une séquence d'ARN simple brin permettent en effet de protéger celle-ci contre les RNases. La première hélice en 5' étant le DSR et la seconde en 3' le terminateur.



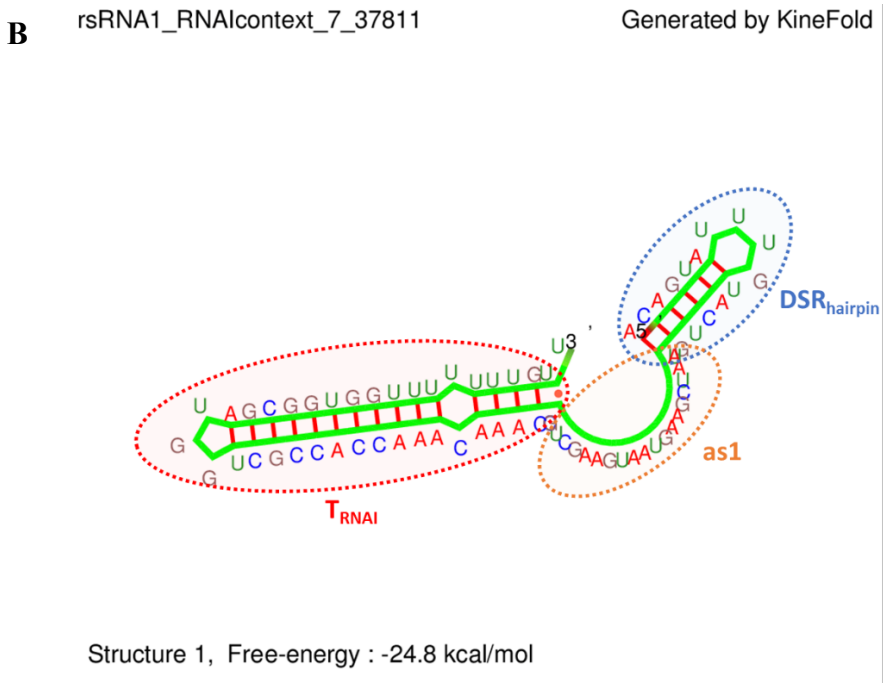


Figure 48 Modification du DSR de p_{RNAI} pour la construction d'un $DSR_{hairpin}$ en tige-boucle. **(A)** Séquence originale du DSR de pRNAI et séquence du $DSR_{hairpin}$. Les régions formant des appariements sont matérialisées par une paire d'éléments entre crochets (`[---[ ]---]`) et annotées par un nom d'hélice `1/1′` et `2/2′`. La séquence surlignée en jaune correspond à la séquence originale du DSR de p_{RNAI} que l'on garde intacte dans la séquence du $DSR_{hairpin}$. **(B)** Résultat de prédiction de repliement co-transcriptionnel du $DSR_{hairpin}$ suivi de l'antisens as1 et du terminateur T_{RNAI}.

Pour la suite, nous voulons identifier le contexte d'expression qui permet d'obtenir un fort niveau d'expression de l'antisens, avec une transcription qui s'arrête effectivement au terminateur. Pour ce faire, nous testons plusieurs combinaisons de contexte d'expression telles que décrites dans le **Tableau 5** :

Nom du plasmide testé	Promoteur	DSR	Antisens	Terminateur
pGEJ1718	p_{RNAI}	DSR de $p_{LtetO-1}$	as1	$T1_{rrnB}$
pGEJ1719	p_{RNAI}	DSR de p_{RNAI}	as1	$T1_{rrnB}$
pGEJ1720	p_{RNAI}	$DSR_{hairpin}$	as1	$T1_{rrnB}$
pGEJ1721	p_{RNAI}	$DSR_{hairpin}$	as1	T_{RNAI}

Tableau 5 Les différentes combinaisons de contexte d'expression (promoteur, DSR, terminateur) testées pour l'expression de l'antisens. Pour plus de simplicité, un nom de plasmide est donné pour chaque combinaison.

Les transcriptions *in vitro* faites à partir de ces 4 plasmides, dont le vecteur est extrait du plasmide pUC19, confirment la production d'un transcrit d'une taille d'environ 95 nucléotides. Ce transcrit, commun aux 4 plasmides, correspond au transcrit de RNAI, dont l'identité est déduite grâce à l'utilisation des transcrits de pUC19 comme référence. En plus de ce transcrit, on constate également que les plasmides pGEJ1719, pGEJ1720 et pGEJ1721 produisent d'autres transcrits qui ont une taille d'environ 80 nucléotides pour le premier et 60 nucléotides pour les deux autres. Or on estime la taille théorique de l'antisens as1, transcrit à partir de ces 3 plasmides, d'être respectivement de 102, 80 et 73 nucléotides. Cela signifie qu'à l'exception de pGEJ1718, les 3 autres plasmides produisent bien l'antisens as1. La résolution du marqueur de taille (M) ne permet pas ici de faire une prédiction fiable de la taille des transcrits.

En ce qui concerne le niveau d'expression de chacun de ces transcrits, nous constatons que le niveau d'expression de l'antisens as1 à partir de pGEJ1720 est légèrement supérieur aux niveaux d'expression obtenues à partir des deux autres (pGEJ1719 et pGEJ1721) (**Figure 49C**). D'autre part, sur la **Figure 49A** nous remarquons que toutes les bandes ne sont pas à la même hauteur, ce qui peut induire en erreur puisqu'il s'agit ici du même transcrit RNAI. Cette observation s'explique notamment par un défaut du gel qui induit un défaut de migration des ARN. En effet, sur un gel précédent, nous avons pu observer que toutes les bandes étaient à la même hauteur, mais elles paraissaient très peu intenses, car les transcriptions *in vitro* y ont été faites que pendant 1 h. C'est pourquoi nous présentons ici des résultats de transcription *in vitro* de 3 h qui permet de produire une plus grande quantité d'antisens, qui apparaissent alors plus intenses sur les gels.

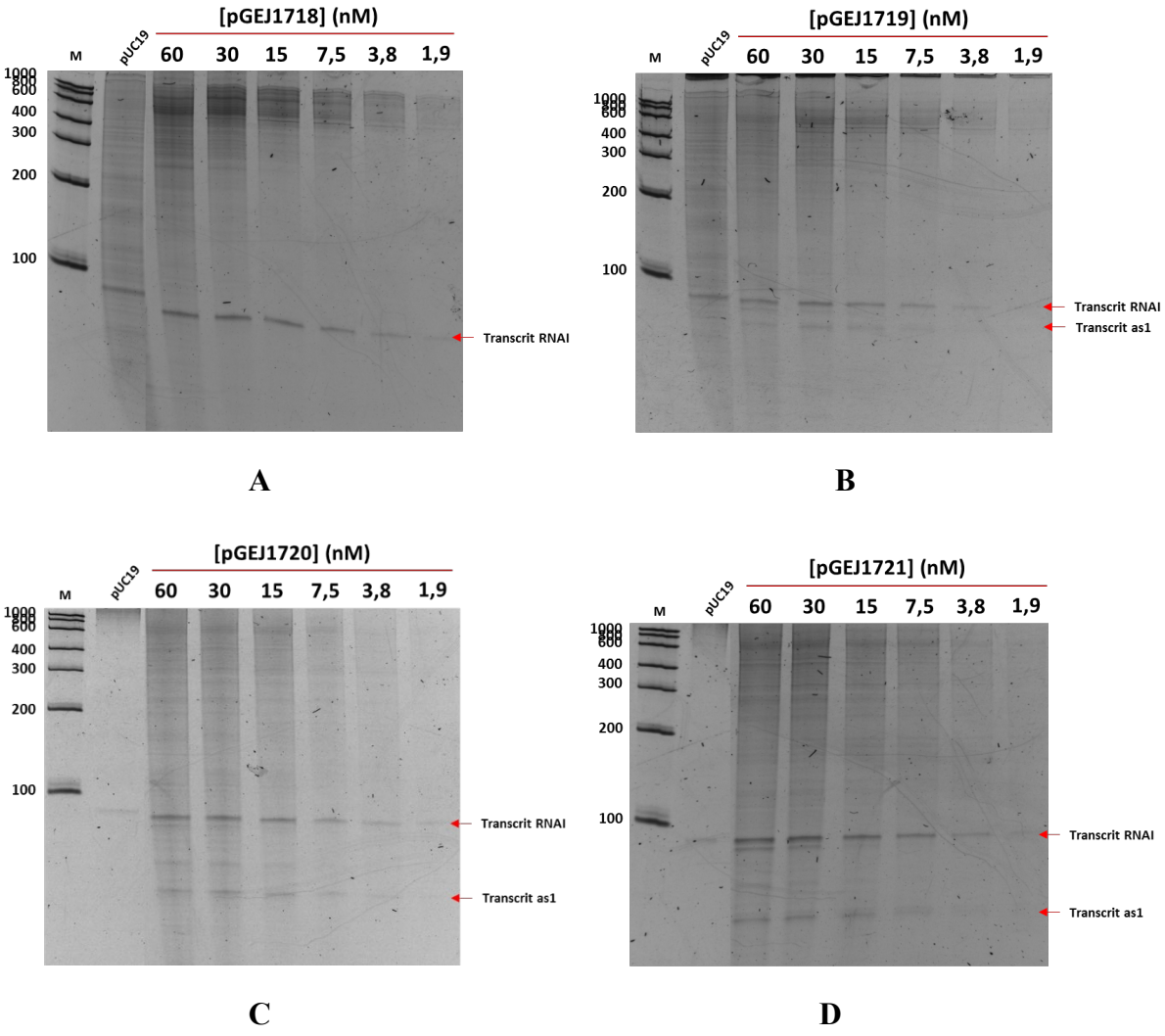


Figure 49 Expression de l'antisens as1 à partir de différentes combinaisons de contexte d'expression (promoteur, DSR et terminateur). Images de gel de polyacrylamide en condition dénaturante pour la séparation des différents transcrits. Toutes les transcriptions *in vitro* sont réalisées à 37 °C, pendant 3 heures, à partir de l'ARN polymérase d'*E. coli* (60 nM). Les matrices utilisées pour réaliser ces transcriptions *in vitro* sont des plasmides (Tableau 5). Plusieurs concentrations (60 ; 30 ; 15 ; 7,5 ; 3,8 ; 1,9 nM) de matrice sont testées pour réaliser la transcription *in vitro* afin d'identifier la concentration optimale de plasmide qu'il faudrait utiliser pour commencer à observer une transcription de l'antisens.

D'après ces résultats, il apparait que seul pGEJ1718 ne produise pas d'antisens as1. On peut attribuer cela à l'utilisation du DSR de $p_{LtetO-1}$ qui n'est probablement pas compatible avec le promoteur p_{RNAI}, ce qui confirme ici l'importance de la séquence DSR dans l'activité du promoteur.

Parmi les 3 combinaisons restantes, pGEJ1720 et pGEJ1721 semblent être de bons candidats pour exprimer l'antisens as1 car son niveau d'expression y est supérieur comparé au niveau d'expression obtenu à partir de pGEJ1719.

Par ailleurs, nous observons également des transcrits parasites, dont les niveaux d'expression sont toutefois plus faibles que ceux du transcrit d'intérêt. Ces transcrits parasites peuvent être le fruit d'une transcription ne s'étant pas arrêtée au bon terminateur.
Pour la suite, nous nous basons sur la combinaison de contexte d'expression utilisée dans pGEJ1720 pour exprimer les antisens ARN.

Chapitre 4 : Optimisation du protocole de mesure pour la caractérisation *in vivo* des commutateurs ARN : l'importance de se placer en condition de croissance équilibrée

1. Premier test *in vivo* du fonctionnement du commutateur ARN en absence ou présence de l'antisens ARN

Nous voulons tester ici le fonctionnement du commutateur ARN dans la bactérie, en absence ou en présence d'expression de l'antisens ARN. Comme cela a été vu précédemment dans la partie III de ce manuscrit, la mesure de la fluorescence moyenne par bactérie dans les deux conditions : absence ou présence de l'expression de l'antisens ARN, est un témoin indirect du fonctionnement du commutateur ARN. Or pour pouvoir comparer ces mesures, il faut que les plateformes d'expression de l'antisens et du commutateur soient présents en même temps dans la bactérie, quelles que soient les conditions (absence ou présence d'expression de l'antisens ARN). Ainsi, pour ce premier test, l'absence d'expression de l'antisens ARN est obtenu par l'utilisation d'un plasmide exprimant un antisens non spécifique du commutateur ARN ; tandis que l'autre condition est obtenue par l'utilisation d'un plasmide exprimant l'antisens spécifique.

Ainsi, nous pouvons observer sur les courbes de DO_{600} (**Figure 50A**) certaines irrégularités (absence de linéarité), en particulier à partir de la phase exponentielle de croissance (à 6 h). On retrouve ces irrégularités sur les courbes de fluorescences/DO_{600} (**Figure 50B**), notamment entre 6 h et 9 h de croissance. Or c'est pendant cette période que nous nous attendions à une croissance équilibrée qui est une condition nécessaire à l'interprétation des données. En effet, la présence de ces vagues ne permet pas d'affirmer si la présence ou l'absence d'expression de l'antisens spécifique au commutateur a effectivement un effet sur la fluorescence/DO600 et donc le repliement du commutateur ARN.

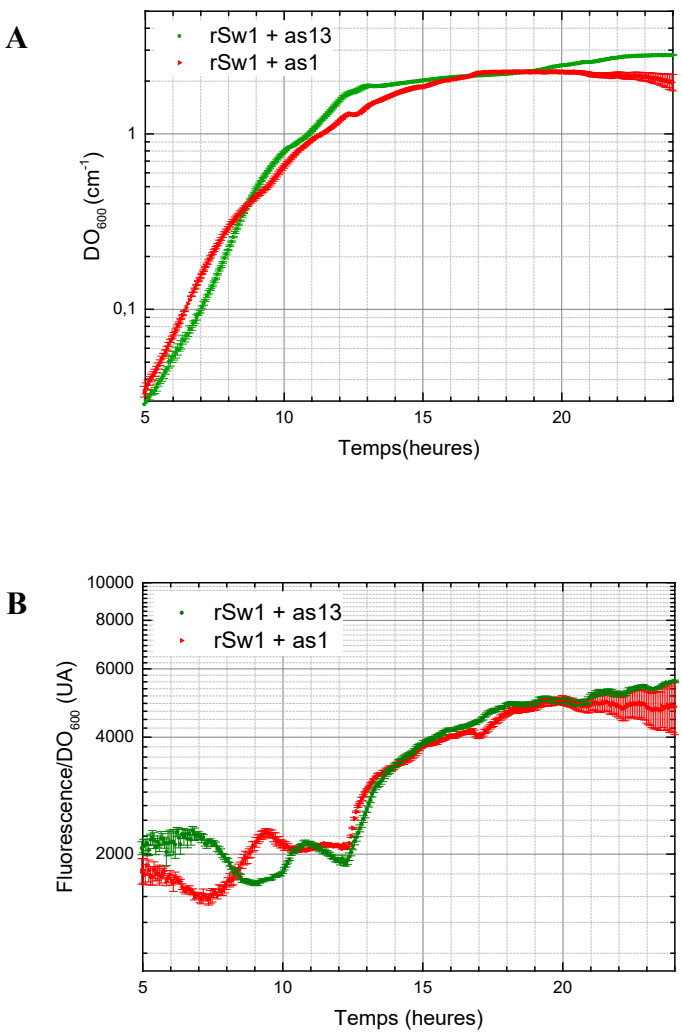


Figure 50 Fonctionnement du commutateur ARN répressible rSw1, en absence et en présence de son antisens ARN spécifique as1, dans la bactérie *E. coli*. (A) DO_{600} = f(temps). **(B)** Fluorescence/DO_{600} = f(temps). En rouge, bactérie exprimant le commutateur ARN répressible rSw1 et son antisens spécifique as1. En vert, bactérie exprimant le commutateur ARN répressible rSw1 et un antisens non spécifique as13. Les données correspondent à des moyennes et les barres d'erreur représentent des écarts-types obtenus à partir de 3 réplicats. Cultures dans un milieu minimum M9 + 0,2 % glucose + 40 µg/mL de kanamycine + 100 µg/mL d'ampicilline.

Ces résultats montrent l'importance de l'utilisation des billes pour l'homogénéisation des cultures bactériennes dans les microplaques, afin d'éviter les biais de mesure d'absorbance, principalement dû à l'agrégation des bactéries sur le fond des puits de la microplaque.

Bien que les résultats ne soient pas interprétables, nous anticipons toutefois qu'une très forte expression de l'antisens puisse être toxique pour la bactérie à cause d'une charge métabolique trop importante. L'apparition de certaines mutations peuvent cependant donner un avantage de croissance à la bactérie. Ces mutations peuvent toutefois viser le niveau de transcription de

l'antisens ARN. On peut alors se retrouver avec une souche de bactérie qui produit peu ou pas d'antisens ARN.

2. Amélioration des courbes de DO_{600} et de fluorescence/DO_{600} avec l'utilisation des billes

Nous reprenons ici l'expérience qui a été faite précédemment, mais :

- (i) une bille de verre de 2 mm de diamètre est rajoutée dans chaque puit de la microplaque contenant une culture pour assurer une bonne homogénéisation de celle-ci ;
- (ii) pour chaque condition étudiée plusieurs clones bactériens sont testés (ici 9 clones), pour espérer en sélectionner au moins un dans lequel le niveau d'expression de l'antisens est resté intact et n'est pas diminué par une quelconque mutation.

Techniquement, les souches fraîchement transformées avec le plasmide d'expression du commutateur et le plasmide d'expression de l'antisens (spécifique ou non du commutateur) sont ensemencées sur un milieu solide (LB + agar) contenant les antibiotiques (40 µg/mL de kanamycine et 100 µg/mL d'ampicilline) permettant de sélectionner les clones qui contiennent effectivement les deux plasmides.

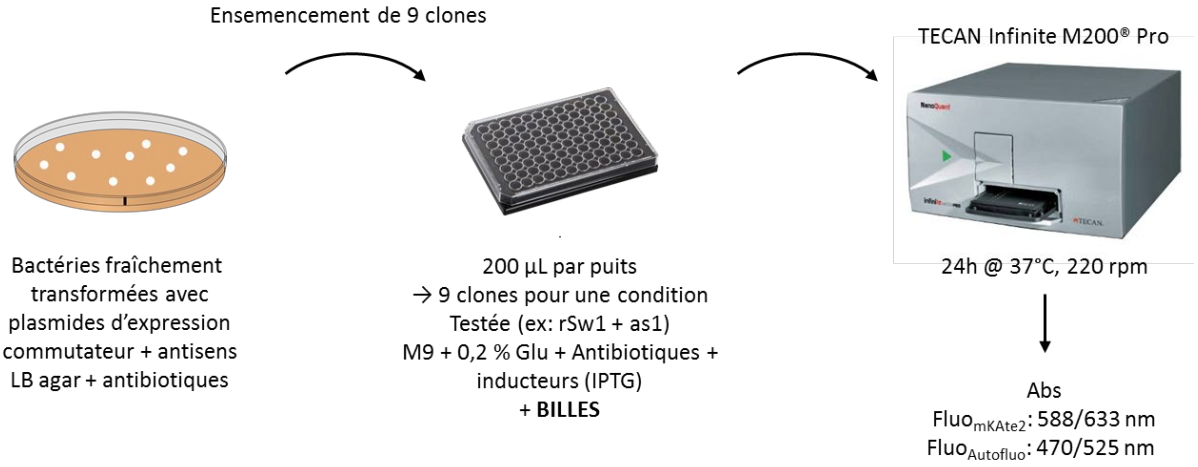


Figure 51 Première mise au point du protocole de préparation des échantillons bactériennes pour la mesure d'absorbance et de fluorescence. 9 clones issus du même transformant sont sélectionnés et directement ensemencés dans les microplaques contenant un milieu minimum M9 + 0,2 % glucose + antibiotiques (40 µg/mL de kanamycine + 100 µg/mL d'ampicilline) + inducteurs (1 mM IPTG). Une bille de verre de 2 mm de diamètre est rajoutée dans chaque puits contenant une culture. Les mesures d'absorbance et de fluorescence sont faites sur TECAN Infinite M200 Pro.

Contrairement au protocole initial, pour une condition donnée, ici 9 clones issus du même transformant sont directement ensemencés dans une microplaque, sans passer par les étapes préalables de préculture. La culture se fait dans un milieu minimum M9 contenant 0,2 % de glucose, ainsi que 40 µg/mL de kanamycine et 100 µg/mL d'ampicilline (Figure 51).

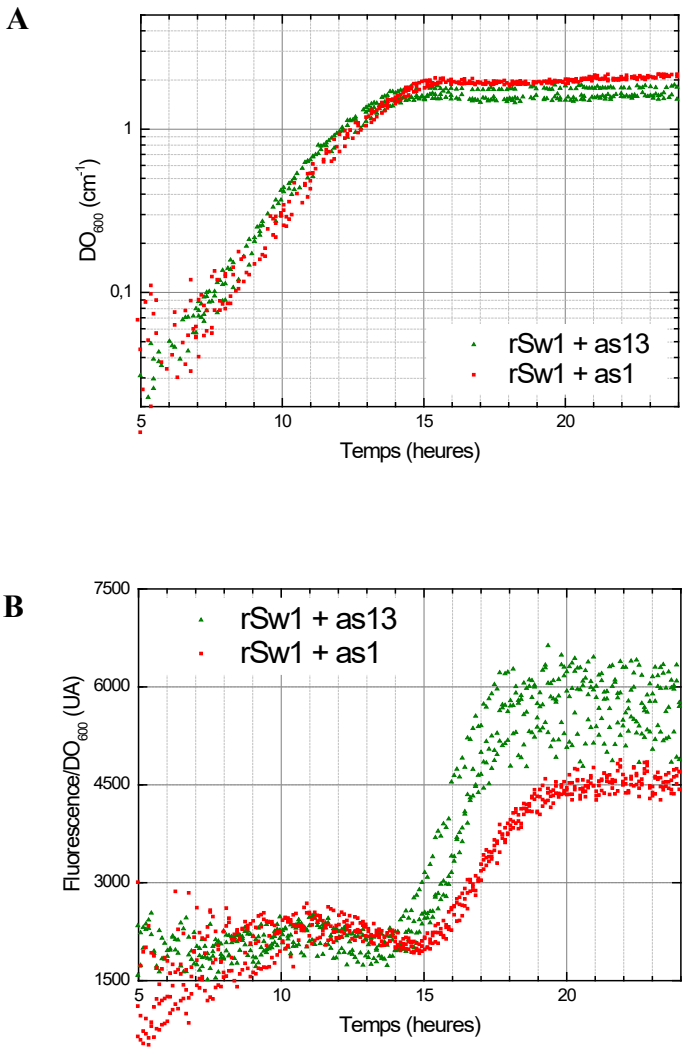


Figure 52 L'homogénéisation par des billes, du milieu de culture dans les microplaques, permet d'avoir des mesures d'absorbance plus fiables. (A) DO_{600} = f (t). (B) Fluorescence/DO_{600} = f (t). En rouge, bactéries exprimant le commutateur ARN répressible rSw1 et son antisens spécifique as1. En vert, bactéries exprimant le commutateur ARN répressible rSw1 et un antisens non spécifique as13. Les données individuelles de DO_{600} et de fluorescence/DO_{600} de chaque clone sont représentées. Cultures dans un milieu minimum M9 + 0,2 % glucose + 40 µg/mL de kanamycine + 100 µg/mL d'ampicilline. Les données de DO_{600} et de fluorescence/DO_{600} présentées ici ont été préalablement traitées et analysées sur WELLINVERTER.

Les résultats que nous obtenons confirment l'importance de l'utilisation des billes, qui limitent l'agrégation des bactéries sur le fond des puits de la microplaque. Les faisceaux utilisés

pour les mesures d'absorbance rencontrent ainsi peu d'obstacle (mis à part les billes) et permettent de réaliser des mesures plus fiables. Comparées aux courbes de DO_{600} (**Figure 50A**) obtenues à partir de mesure faite en absence de billes, celles faite en présence de billes (**Figure 52A**) sont effectivement plus régulières (pas de vagues). Il en est de même pour les courbes de fluorescence/DO_{600} (**Figure 52B**) sur lesquelles on peut même distinguer la croissance équilibrée (plateau entre 6h et 13h de croissance), contrairement aux mesures précédentes (**Figure 50B**).

Bien que la croissance équilibrée soit plus nette sur les courbes de fluorescence/DO_{600}, il semblerait qu'on n'observe pas de différence entre le fonctionnement du commutateur ARN répressible rSw1 en absence et en présence d'expression de l'antisens ARN as1.

Cependant, on commence à observer une différence à partir de l'entrée en phase stationnaire (À partir de 13 h de croissance). La fluorescence/DO_{600} en absence de l'expression de l'antisens as1 y est environ 1,3 fois supérieure à la fluorescence/DO_{600} en présence de l'expression de l'antisens. Or théoriquement, le commutateur répressible rSw1 est dans un état « ON » en absence d'expression de l'antisens as1, tandis qu'il est dans un état « OFF » en présence de son expression. Ainsi, nous obtenons un premier résultat encourageant, certes que l'on ne voit pas en condition de croissance équilibrée mais qui est tout de même en accord avec le fonctionnement théorique du commutateur. Un clone parmi ces 9 testés est stocké et utilisé pour les expériences suivantes.

3. Mise au point d'un protocole permettant de se placer en condition de croissance équilibrée

Les résultats précédents ont montré que la régulation des commutateurs par l'antisens ARN apparaissait à partir de l'entrée en phase stationnaire (**Figure 52B**), or nous évoquions l'importance de se placer en condition de croissance équilibrée pour s'assurer que la fluorescence par bactérie soit une représentation fidèle de l'état du commutateur en fonction uniquement de l'absence ou de la présence spécifique de l'antisens ARN. De plus, nous avons vu que pour atteindre cette condition, il faut que les bactéries restent suffisamment longtemps en phase exponentielle de croissance. Cependant dans le protocole précédent les bactéries sont initialement en phase stationnaire lorsqu'elles sont ensemencées en microplaque. Par conséquent, nous ne savons pas si les bactéries restent suffisamment longtemps en phase exponentielle de croissance avant qu'un signal ne puisse être détecté.

Cette nouvelle version du protocole adopte une stratégie qui permette aux bactéries d'être en phase exponentielle au moment de leur ensemencement en microplaque. Tout d'abord, les bactéries ne sont plus ensemencées directement à partir d'une colonie mais à partir d'une culture de démarrage qui correspond à une préculture faite la veille à partir d'un clone dans lequel le fonctionnement du commutateur a été partiellement validé grâce au protocole précédent. Le lendemain, cette culture de démarrage est diluée jusqu'à une DO_{600} calculée de façon qu'après 3 heures de culture, les bactéries atteignent une phase de croissance exponentielle qui commence généralement à une $DO_{600} \approx 0{,}1$. Connaissant le temps de doublement de la bactérie, ainsi que la DO_{600} finale à laquelle on souhaite se placer, on peut effectivement remonter à la DO_{600} initiale à laquelle on souhaite démarrer la culture, telle que :

$$\boldsymbol{DO_{600(finale)} = 2^{n} DO_{600(initiale)}} \qquad \textit{(9)}$$

avec :

$DO_{600(finale)}$ = densité optique mesurée à l'arrêt de la culture

$DO_{600(initiale)}$ = densité optique mesurée au démarrage de la culture

n = nombre de division

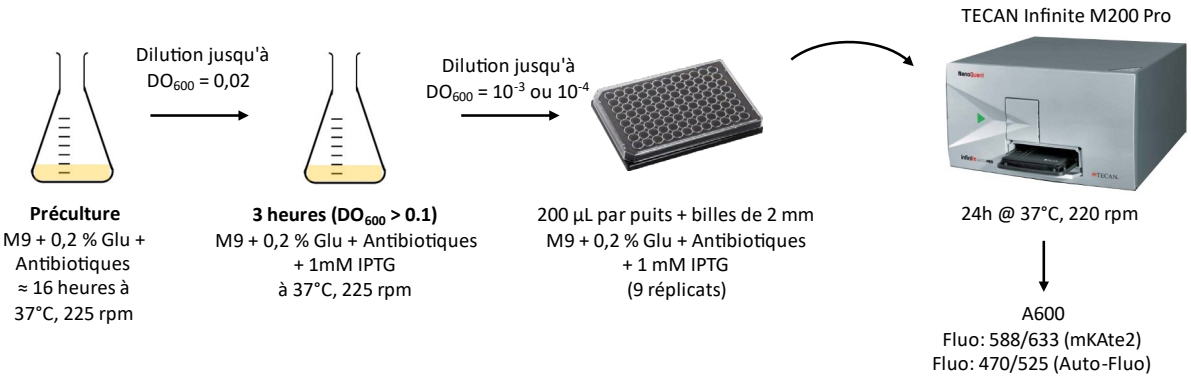


Figure 53 Protocole de préparation des échantillons bactériennes pour qu'elles atteignent une croissance équilibrée. La veille, les souches à tester sont ensemencées dans un milieu minimum M9 contenant 0,2 % de glucose, ainsi que des antibiotiques lorsqu'ils sont nécessaires (40 µg/ml de kanamycine et 100 µg/ml d'ampicilline). Après environ 16 heures de culture sous agitation (225 rpm) à 37 °C, cette culture de démarrage est diluée à une DO_{600} égale à 0,02 dans le même type de milieu que la veille, mais dans lequel on rajoute l'IPTG utilisé pour l'induction de l'activité du promoteur contrôlant l'expression du commutateur. Cette dilution est de nouveau incubée à 37 °C, sous agitation (225 rpm) pendant environ 3 h, jusqu'à ce que la DO_{600} atteigne 0,1 ou plus. Une fois atteinte, la culture est de nouveau diluée à une DO_{600} comprise entre 10^{-3} et 10^{-4} dans un milieu similaire au deuxième. 200 µL de cette dilution est ensuite transféré en microplaque. Les mesures d'absorbance et de fluorescence sont réalisées par un lecteur de microplaque TECAN Infinite M200 PRO.

Au bout de 3 heures d'incubation, la culture est diluée une nouvelle fois jusqu'à une DO_{600} comprise entre 10^{-3} et 10^{-4}, puis ensemencée en microplaque. Puisqu'elles sont très diluées dans le milieu de culture, les bactéries mettent du temps avant d'entrer en phase stationnaire et restent donc suffisamment longtemps en phase exponentielle (Figure 53).

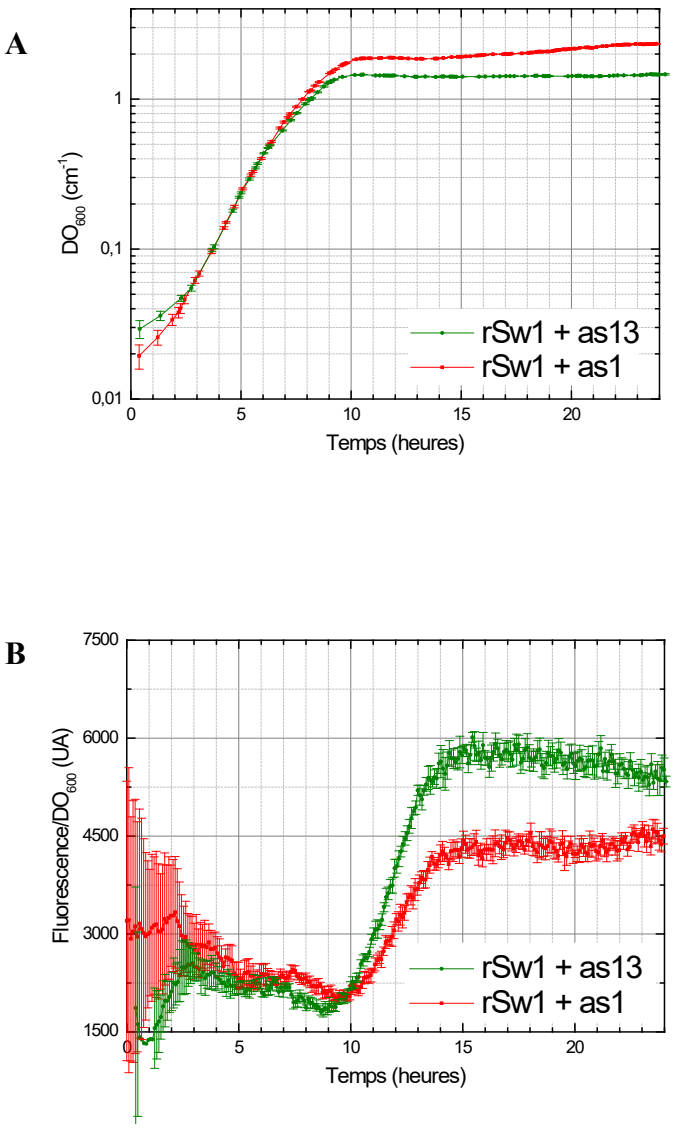


Figure 54 Obtention d'une croissance équilibrée et confirmation de l'effet de l'antisens ARN sur le fonctionnement du commutateur ARN en phase stationnaire. **(A) DO_{600} = f (t). (B) Fluorescence/DO_{600} = f (t).** En rouge, bactéries exprimant le commutateur ARN répressible rSw1 et son antisens spécifique as1. En vert, bactéries exprimant le commutateur ARN répressible rSw1 et un antisens non spécifique as13. Les données correspondent à des moyennes et les barres d'erreur représentent des écarts-types obtenus à partir de 9 réplicats. Cultures dans un milieu minimum M9 + 0,2 % glucose + 40 µg/mL de kanamycine + 100 µg/mL d'ampicilline + 1 mM d'IPTG.

Les résultats que nous obtenons confirment l'importance de l'utilisation des billes pour les cultures en microplaque. Les courbes de croissance sont en effet beaucoup plus réguliers, ce

qui permet d'identifier un problème de croissance parmi les souches de bactérie testées. Sur la **Figure 54A** nous constatons que le temps de doublement des deux souches sont certes similaires (≈ 1 heure), cependant les rendements de croissance en phase stationnaire (à partir de 9 heures de croissance) sont différents : la souche qui exprime rSw1 et as13 a un rendement de croissance inférieur à la souche exprimant rSw1 et as1. Or, c'est pendant cette même période que l'on constate l'effet de l'antisens ARN sur le commutateur (**Figure 54B**) : la fluorescence/DO_{600} en présence de l'expression de l'antisens as1 est 1,3 plus faible que la fluorescence/DO_{600} en absence de son expression. Bien que ces résultats soient en accord avec le fonctionnement théorique du commutateur, la différence de croissance à partir du début de la phase stationnaire rend la conclusion discutable, car l'analyse se porte ici sur une fluorescence rapportée à la DO_{600}.

Chapitre 5 : Conclusion

Nous avons développé une plateforme et mis en place une méthode qui permettent de tester *in vivo* le fonctionnement des commutateurs ARN. La bactérie *Escherichia coli* est choisie comme modèle pour réaliser ces tests. En effet, le fonctionnement des commutateurs que nous caractérisons est basé principalement sur des mécanismes de terminaison/anti-terminaison que l'on retrouve le plus fréquemment dans les bactéries. Le commutateur ARN ainsi que l'antisens ARN associé y sont ainsi exprimés à partir de plasmides différents, offrant la possibilité de produire une quantité d'antisens ARN en excès devant la quantité de commutateur ARN. C'est en effet l'une des conditions nécessaires pour que la régulation du commutateur ARN par l'antisens ARN se fasse efficacement.

Les premiers tests *in vivo* montrent des résultats préliminaires encourageants, en accord avec la description du fonctionnement théorique des commutateurs ARN en absence et en présence d'expression des antisens ARN qui leur sont spécifiques. Cependant ces résultats ne s'observent qu'à partir de l'entrée en phase stationnaire. Or notre objectif est de pouvoir observer cet effet de régulation au moment de la croissance équilibrée, pendant la phase exponentielle de croissance. De plus la différence de croissance, en phase stationnaire, entre la souche qui exprime le commutateur et son antisens spécifique ainsi que la souche qui exprime le commutateur et son antisens non-spécifique, rend l'interprétation des résultats discutable. En effet, l'effet de régulation que nous observons en phase stationnaire peut tout aussi bien être une conséquence de cette différence de croissance car nous comparons une fluorescence/DO_{600}.

Nous verrons dans la partie IV que cette différence de croissance en phase stationnaire est une conséquence de l'expression de l'antisens as13 dans la bactérie. Pour vérifier le fonctionnement des commutateurs en absence d'expression de leurs antisens spécifiques, nous testerons différentes alternatives qui n'auront pas d'effet sur la croissance.

Partie IV : Validation *in vivo*, du fonctionnement des commutateurs ARN

Chapitre 1 : Introduction

Le fonctionnement de la plateforme de test *in vivo*, développée et décrite dans la partie précédente, a été testée pour la première fois sur le commutateur répressible rSw1. Pour ce faire, deux conditions ont été testées : en présence de l'expression de son antisens associé, as1, et en présence de l'expression d'un antisens non-spécifique, as13. Les résultats présentés dans le chapitre 4 de la partie III sont des résultats préliminaires de ce test. Ces résultats sont cependant obtenus à partir d'une version optimisée de la séquence de rSw1. Nous verrons plus en détails dans le chapitre suivant, pourquoi et comment cette optimisation de séquence est faite.

Par ailleurs, il semblerait que l'un des deux antisens as1 ou as13 ait un effet sur la croissance des bactéries, plus particulièrement sur la turbidité finale mesurée en phase stationnaire et qui correspond a priori au rendement de la croissance (la quantité de biomasse généré pour un apport donné en nutriments). Il faut rappeler que nous avons observé l'effet de l'antisens sur l'expression du gène rapporteur en phase stationnaire, c'est à dire au moment où la turbidité apparaît stable. Par conséquent il est difficile de conclure de façon définitive sur l'effet de l'antisens sur l'expression du gène rapporteur puisqu'il peut tout aussi bien être un effet apparent dû à une modification du niveau de turbidité final. Pour la suite, nous essaierons donc d'identifier parmi les deux antisens utilisés (as1 ou as13), lequel modifie la croissance et la turbidité finale des bactéries.

Ces différentes optimisations ont pour objectifs d'écarter tout effet non-spécifique des antisens et d'assurer l'obtention de résultats fiables pour la validation *in vivo* du fonctionnement des commutateurs ARN que nous verrons par la suite.

Chapitre 2 : Optimisation des séquences des commutateurs ARN et des antisens ARN

1. Identification et suppression des séquences promotrices cryptiques internes dans rSw1 : obtention de rSw1.v2

Ce travail a été réalisé en partie par un étudiant en Master : Valerio Piomponi qui a réalisé son stage de Master 2 dans le laboratoire et que j'ai encadré en partie.

Comme cela a été vu précédemment dans le chapitre 1 de la partie III, l'expression des commutateurs ARN dans la bactérie est sous contrôle du promoteur pLlacO-1 [96]. Cependant, comparé à la caractérisation faite par Lutz *et al*., nous mesurons une gamme dynamique assez faible pour l'expression du commutateur rSw1 lorsqu'on varie la concentration de l'IPTG (**Figure 50**). Il faut toutefois prendre cette comparaison avec précaution car, entre leur système et le nôtre, le système d'expression et les conditions de mesure ne sont pas exactement les mêmes (par ex. : pas le même rapporteur, pas le même élément régulateur en 5'- UTR, pas la même souche *E. coli*). Pour une comparaison pertinente, il faudrait comparer des résultats entre systèmes d'expression similaires, exprimant le même gène rapporteur (mKate2) et sans commutateur en aval ; ce qui n'a pas été fait. Néanmoins, la faible gamme dynamique que nous obtenons reste surprenante : nous obtenons une fluorescence/DO_{600} égale à 2000 UA en absence d'IPTG et 4000 UA en présence de 1 mM d'IPTG, soit une gamme dynamique d'un facteur 2 (**Figure 40**) contre une gamme dynamique d'un facteur ~800 pour Lutz *et al*. [96]. Si la gamme dynamique du promoteur qui contrôle l'expression du commutateur est déjà faible, cela peut être le signe d'un problème au niveau de l'activité promotrice qui pourrait impacter la gamme dynamique mesurable lors du fonctionnement des commutateurs ; autrement dit, il pourrait d'agir d'un problème qui pourrait nous empêcher de mesurer la véritable amplitude de la régulation atteignable par les commutateurs. Notre premier objectif est donc de comprendre et potentiellement d'améliorer la gamme dynamique au niveau de l'activité promotrice afin de pouvoir évaluer ensuite au mieux la gamme dynamique au niveau de la régulation terminateur/anti-terminateur des commutateurs ARN.

La faible gamme dynamique que nous observons pour la régulation de l'activité du promoteur $p_{LlacO-1}$ pourrait s'expliquer par un niveau d'expression basal de la fluorescence inhabituellement haute en absence d'IPTG. Une forte fluorescence de base observée à 0 µM d'IPTG pourrait s'expliquer par une fuite du promoteur ou par une activité promotrice résiduelle. Une hypothèse possible pourrait être la création accidentelle de séquences promotrices dans la séquence-même du commutateur lors de sa conception, ce qui provoquerait

une fuite d'expression indépendante de l'activité du promoteur $p_{LlacO-1}$. Si c'est le cas, cela pourrait avoir un impact critique sur le fonctionnement du commutateur, car les transcrits issus de promoteurs internes ne porteraient très probablement qu'un commutateur tronqué d'une partie plus ou moins grande de sa région 5', ce qui conduirait à une fraction non négligeable de transcrits non-fonctionnels pour la commutation et pour la régulation par l'antisens ARN.

Dans la partie qui suit, nous présentons une méthode pour identifier ces séquences promotrices internes et les altérer par mutagenèse dirigée[7] afin de diminuer le niveau basal d'expression du gène rapporteur mKate2. Cette optimisation vise à améliorer la qualité d'expression de rSw1 et, par la même occasion, améliorer la dynamique de régulation de l'expression du rapporteur par le commutateur.

1.1. Analyse de séquence : prédiction de séquences promotrices dans la séquence de rSw1

De nombreuses études de comparaison des séquences promotrices de différents gènes chez *Escherichia coli* [111][112] ont permis d'identifier deux régions conservées dans celles-ci, situées à environ 10 et 35 pb (paires de bases) en amont du premier nucléotide transcrit (régions −10 et −35). Ce sont des régions hexanucléotidiques qui arborent respectivement les séquences consensus TATAAT et TTGACA. Ces deux régions sont espacées d'une séquence non conservée de 17 ± 1 pb. Le premier nucléotide transcrit se trouve à 7 ± 1 pb de la région −10 (Figure 55). A l'aide de cette description des promoteurs et à l'aide d'un outil d'analyse de séquence (Clone Manager – Sci-Ed Software) nous avons pu identifier dans la séquence du commutateur rSw1 d'autres promoteurs potentiels qui se superposent. En effet, une recherche de ces régions consensus est effectuée dans la séquence partant du promoteur $p_{LlacO-1}$ et allant jusqu'au commutateur rSw1. La recherche est effectuée pour chacun des espacements possibles : 16, 17 et 18 pb, et pour des séquences portant jusqu'à 4 bases de différences dans l'ensemble des deux régions −10 et −35 par rapport aux séquences hexanucléotidiques consensus. Cette recherche a permis d'identifier deux séquences promotrices internes potentielles ayant seulement 3 bases de différence par rapport à la séquence consensus avec un espacement de 17 pb, et 7 séquences promotrices internes potentielles ayant 4 bases de

[7] La mutagenèse dirigée est une méthode utilisée en biologie moléculaire pour introduire des mutations (insertions, délétions et substitutions de bases) spécifiques et ciblées au niveau d'un ADN plasmidique double brin.

différence par rapport à la séquence consensus avec un espacement de 16 ou 18 pb (Tableau 6). La séquence du promoteur $p_{LlacO-1}$ a, elle, 2 bases de différence par rapport à la séquence consensus.

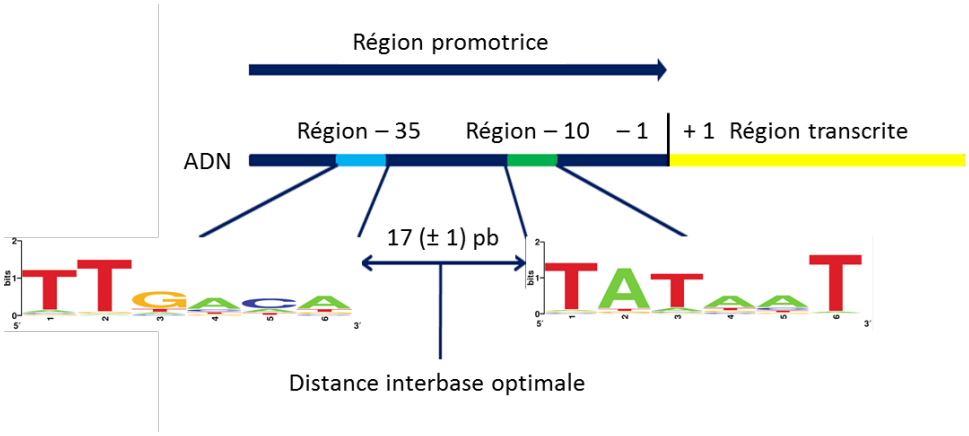


Figure 55 Structure caractéristique d'un promoteur procaryote. Les séquences consensus sont représentées sous forme de logos consensus qui permettent de définir le degré de conservation d'un nucléotide à une position donnée en fonction de la taille des lettres : plus elles sont grandes et plus le nucléotide associé est conservé.

Espacement	Pos. *	Diff. **	Séquence
			- 35 - 10
	Consensus (16 pb)		TTGACA-NNNNNNNNNNNNNNNN-TATAAT-NNNNNNNNNNNN
16 pb	–42	4	aTaACA-attgacattgtgagcg-gATAAc-aagatactgagc
	41	4	aTGAaA-tatcatagattataga-TtTtAT-caagaatacttg
	88	4	TaGAat-actatgatttaatcta-TATAtT-ataacccgaagt
			- 35 - 10
	Consensus (17 pb)		TTGACA-NNNNNNNNNNNNNNNNN-TATAAT-NNNNNNNNNNN
17 pb	–35 ($p_{LlacO-1}$)	2	TTGACA-ttgtgagcggataacaa-gATAcT-gagcacggtcg
	65	3	TTatCA-agaatacttgattttca-TAgAAT-actatgattta
	78	3	TTGAtt-ttcatagaatactatga-TtTAAT-ctatatattat
			- 35 - 10
	Consensus (18 pb)		TTGACA-NNNNNNNNNNNNNNNNNN-TATAAT-NNNNNNNNNN
18 pb	23	4	TTcAtt-acttcggaatctatgaaa-TATcAT-agattataga
	64	4	TTtAtc-aagaatacttgattttca-TAgAAT-actatgattt
	84	4	TTcAtA-gaatactatgatttaatc-TATAta-ttataacccg
	88	4	TaGAat-actatgatttaatctata-TATtAT-aacccgaagt

Tableau 6 Séquences promotrices internes potentielles détectées dans la séquence de rSw1 ainsi que dans sa région promotrice

* Pos. : positions relatives par rapport à la position +1 du promoteur $p_{LlacO-1}$, la première base de la séquence du promoteur $p_{LlacO-1}$ étant située à –55 ; **Diff. : nombre total de différences dans l'ensemble des deux régions –10 et –35 par rapport à la séquence consensus. Les bases qui diffèrent de la séquence consensus sont écrites en minuscules. La séquence identifiée en position –35 correspond au promoteur $p_{LlacO-1}$ lui-même, et la position +1 du promoteur $p_{LlacO-1}$ est souligné deux fois (g).

1.2. Mutations des régions consensus des promoteurs potentiels identifiés dans la séquence du commutateur rSw1 : suppression de l'activité promotrice interne

La force du promoteur est déterminée par le degré de conservation des régions –35 et –10 par rapport aux séquences consensus [113]. Pour tenter d'altérer les deux promoteurs les plus proches de la séquence consensus identifiés dans la séquence de rSw1, des mutations sont réalisées dans les régions –35 et –10 de chacun de ces deux promoteurs (Figure 56). Ces deux promoteurs, aux positions 65 et 78 dans le Tableau 6, possédant un espacement idéal (17 pb) entre les régions –10 et –35, produiraient des commutateurs de 99 et 112 nucléotides plus courts que le commutateur souhaité.

Pour réaliser ces mutations, nous devons nous assurer qu'elles ne déstabilisent pas le repliement prévu pour le commutateur rSw1. Pour cela, il faut d'une part procéder à des mutations qui conservent les relations de complémentarité dans les deux repliements encodés dans la séquence du commutateur (repliements en absence et en présence de l'antisens ARN), et d'autre part vérifier que les repliements co-transcriptionnels prédits par les simulations mènent toujours bien aux repliements finaux attendus dans chaque cas (absence et présence de l'antisens ARN) (Figure 57). Pour une structure de repliement donnée, la mutation d'une base doit donc entraîner la mutation couplée de la base complémentaire avec laquelle la première base est appariée. De la même façon, les appariements doivent être conservés dans la structure de repliement alternative encodée dans le commutateur. En principe, du fait de l'encodage de deux repliements différents dans la même séquence, il peut être nécessaire de modifier de proche en proche de nombreuses bases suite à la mutation d'une base initiale. Par chance, les deux séquences promotrices internes potentielles identifiées et à altérer se situent dans l'espaceur du commutateur, une région de repliement « constant » dans laquelle les appariements sont les mêmes quel que soit le repliement « ON » ou « OFF » adopté par le commutateur.

Trois paires de mutations ont ainsi été réalisées, une dans chacune des hélices 2, 3 et 4 de la région espaceur, et qui permettent d'affecter une base dans chacune des régions –10 et –35 de chacun des promoteurs internes potentiels (Figure 56).

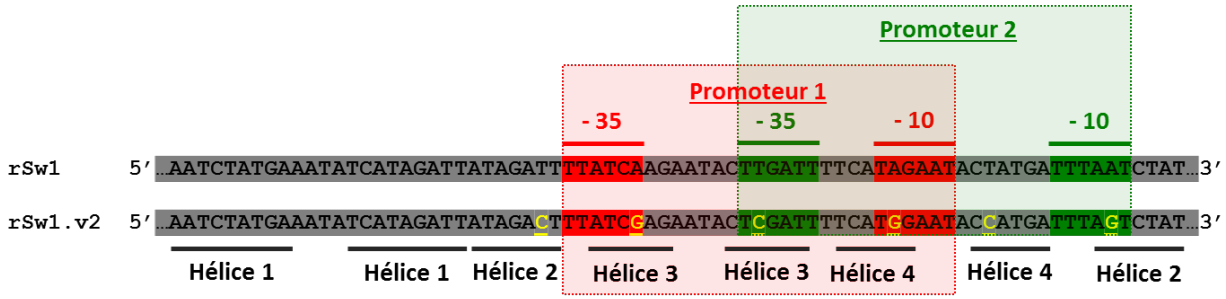


Figure 56 Un nouveau commutateur répressible rSw1.v2. Les 2 promoteurs potentiels (en rouge et en vert) dans la séquence de l'espaceur du commutateur rSw1. En jaune sont représentées les bases mutées pour l'obtention de rSw1.v2. Les traits noirs représentent les séquences des hélices qui forment la structure de repliement de l'espaceur. L'introduction de mutations ne doit pas modifier cette structure.

Nous montrons ainsi, grâce à Kinefold, que les repliements prédits pour rSw1 et rSw1.v2 (rSw1 muté) sont identiques, quel que soit le repliement « ON » ou « OFF » adopté respectivement en absence ou en présence de l'antisens ARN (Figure 57).

Aussi, une nouvelle recherche de séquences promotrices, via les régions consensus –10 et –35, dans la séquence de rSw1.v2, permet de montrer une suppression des deux séquences promotrices identifiées précédemment. Cependant, une nouvelle liste de séquences promotrices potentielles fait son apparition. Elles ont toutefois 4 bases de différence avec la séquence consensus, ce qui devrait affaiblir leur activité promotrice (Tableau 7).

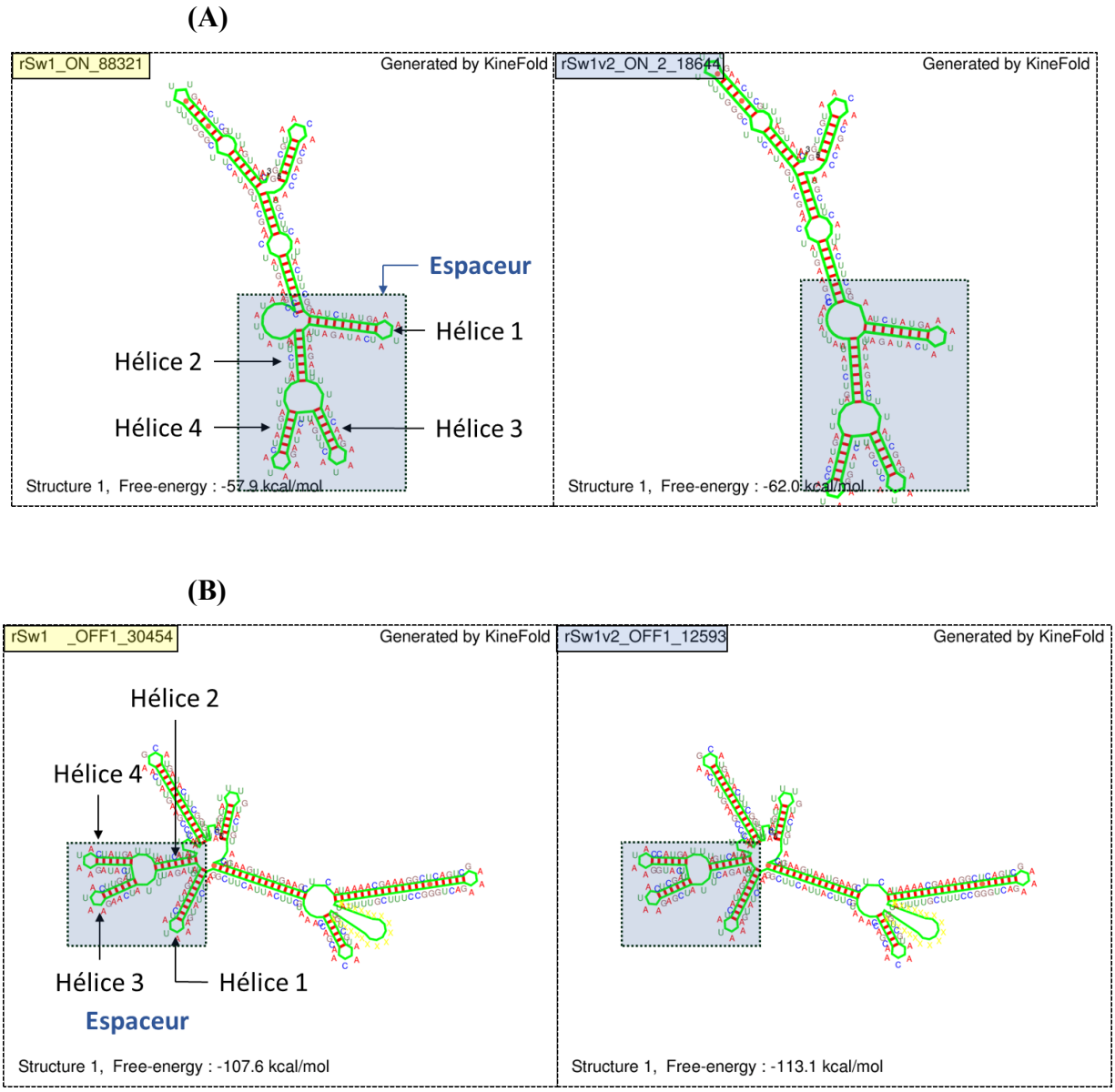


Figure 57 Comparaison de résultats de simulations du repliement de rSw1 (à gauche) et rSw1.v2 (à droite), en absence **(A)** et en présence **(B)** d'antisens ARN, obtenus grâce au programme Kinefold [27]. Dans chaque cas, 5 simulations sont réalisées et toutes donnent des résultats de repliement similaires.

Espacement	Pos. *	Diff. **	Séquence
	Consensus (16 pb)		- 35 - 10 TTGACA-NNNNNNNNNNNNNNNN-TATAAT-NNNNNNNNNNNN
16 pb	42	4	aTaACA-attgacattgtgagcg-gATAAc-aagatactgagc
	88	4	TgGAat-accatgatttagtcta-TATAtT-ataacccgaagt
	Consensus (17 pb)		- 35 - 10 TTGACA-NNNNNNNNNNNNNNNNN-TATAAT-NNNNNNNNNNN
17 pb	35	2	TTGACA-ttgtgagcggataacaa-gATAcT-gagcacggtcg
	87	4	aTGgaA-taccatgatttagtcta-TATAtT-ataacccgaag
	92	4	aTacCA-tgatttagtctatatat-TATAAc-ccgaagtatca
	Consensus (18 pb)		- 35 - 10 TTGACA-NNNNNNNNNNNNNNNNN-TATAAT-NNNNNNNNNN
18 pb	23	4	TTcAtt-acttcggaatctatgaaa-TATcAT-agattataga
	59	4	TaGACt-ttatcgagaatactcgat-TtTcAT-ggaataccat
	88	4	TgGAat-accatgatttagtctata-TATtAT-aacccgaagt

Tableau 7 Séquences promotrices internes résiduelles après mutation de rSw1 pour obtenir rSw1.v2. Les bases modifiées sont indiquées en gras et en rouge sur fond jaune. De nouvelles séquences promotrices potentielles apparaissent dans la liste. Cependant, les séquences promotrices n'ayant que 3 bases de différence avec la séquence promotrice consensus ont toutes été éliminées.

1.3. Suppression de l'activité des promoteurs cryptiques internes dans rSw1 : validation *in vivo*

Pour les mêmes conditions de mesure, lorsque nous comparons les niveaux de fluorescence mesurés dans le contexte du commutateur rSw1 initial et du commutateur altéré rSw1.v2, nous observons une réduction globale de la fluorescence moyenne par bactérie. Nous observons une baisse d'au moins un facteur 5 pour les plus faibles concentrations d'IPTG : 0 et 2 µM et une baisse d'au moins un facteur 2 pour les plus fortes concentrations d'IPTG : 20, 50, 100, 200 et 1000 µM. L'allure générale de la courbe et la différence de fluorescence entre l'état non induit (0 µM d'IPTG) et l'état complètement induit (1000 µM d'IPTG) restent inchangés dans les deux cas (**Figure 58**). Ainsi, la différence d'activité promotrice de $p_{LlacO-1}$ entre l'état non-induit et induit reste inchangée.

Ces résultats indiquent que l'altération des promoteurs, réalisée par mutagenèse dirigée dans la séquence de rSw1 a bien réduit le niveau d'expression de la protéine fluorescente mKate2 d'une quantité constante sur toute la gamme de concentration d'IPTG, confirmant ainsi notre hypothèse initiale d'un effet de promoteurs internes sur la dynamique de régulation observée. Ainsi, le fait de diminuer le niveau basal d'expression a permis d'améliorer la gamme dynamique de l'induction par l'IPTG : on passe d'un facteur 2 à un facteur 6 environ.

Enfin, comme mentionné précédemment, la suppression de séquences promotrices internes devrait également permettre une meilleure dynamique de régulation basée sur la commutation de rSw1 par son antisens ARN effecteur.[8]

Ces courbes d'induction à l'IPTG permettent également tout d'abord de tester que notre dispositif expérimental permet bien de détecter et de mesurer une régulation de l'expression de la protéine rapportrice mKate2. Et d'autre part, la régulation mesurée constitue une référence quant aux variations d'expression pouvant être escomptées pour les états « ON » et « OFF » des commutateurs répressibles.

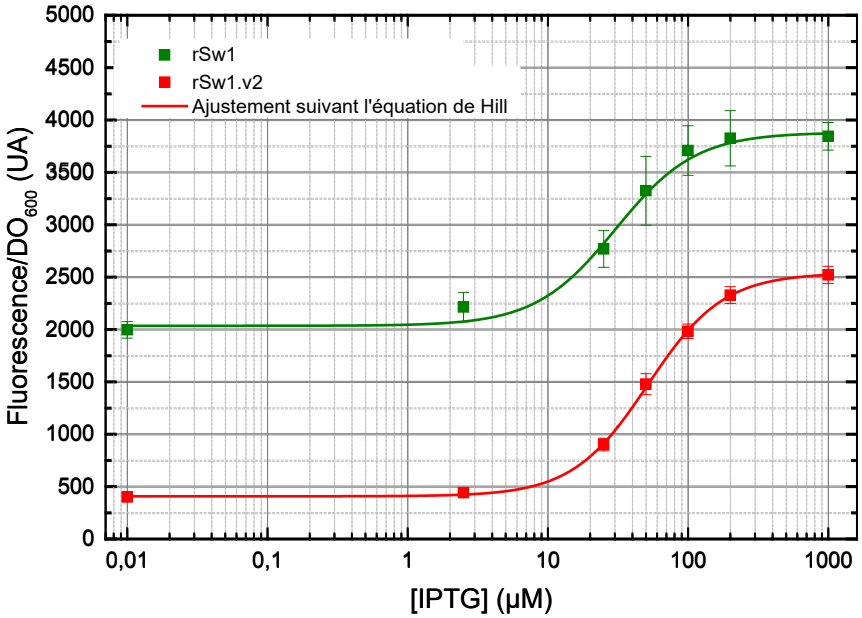


Figure 58 Courbes d'induction par l'IPTG de la plateforme rSw1-mKate2 et rSw1.v2- mKate2 sous contrôle du promoteur $p_{LlacO-1}$. Les fluorescences moyennes par bactérie (Fluorescence/DO_{600}), prises pendant la croissance équilibrée, sont tracées en fonction de la concentration d'IPTG. Les barres d'erreur correspondent à des écarts-types, obtenus à partir de 3 réplicats. La courbe en ligne continue est construite suivant l'équation de Hill *(8)*. La courbe verte est la même que la Figure 40 (page 73). Les meilleurs paramètres d'ajustement pour la courbe en rouge sont donnés dans le Tableau 8. Croissance en milieu minimum M9 + 0,2 % glucose + 40 µg/ml de kanamycine. Noter que ces mesures sont faites en absence de plasmides antisens.

[8] L'activité de ces séquences promotrices internes n'a pas été détectée dans les tests *in vitro* de rSw1. Ceci pourrait s'expliquer par l'utilisation, pour ces tests, d'un promoteur de la transcription beaucoup plus fort (p_{T7A1}) dont la forte activité promotrice a pu masquer l'activité des séquences promotrices internes.

Chi^2	0,13	
R^2	0,9994	
	Valeurs	Écarts-types
$\boldsymbol{F0}$	402,0	9,6
$\boldsymbol{F}max$	2533	30
$\boldsymbol{K}$	51,6	1,7
$\boldsymbol{n}$	1,613	0,067

Tableau 8 Paramètres utilisés pour l'ajustement des données d'induction avec la courbe construite à partir de l'équation de Hill *(8)*.

2. Identification et suppression d'effets séquence-spécifiques des antisens ARN sur la croissance des bactéries

2.1. L'antisens ARN as13 a un effet sur la croissance des bactéries

D'un point de vue méthodologique, pour s'assurer de la spécificité de l'effet de l'antisens ARN sur la régulation de l'état *in vivo* des commutateurs ARN, nous comparons l'effet d'un antisens ARN spécifique du commutateur ARN à tester à l'effet d'un antisens ARN non-spécifique, qui à priori ne s'hybride pas au commutateur ARN. En effet, puisque l'antisens est transcrit par un promoteur fort, nous ne pouvons pas utiliser la souche n'exprimant pas d'antisens du tout comme référence ; la répartition des éléments de la machinerie cellulaire (ARN polymérases, ribosomes...) ne serait pas égale dans les deux conditions. Par exemple, pour le test du fonctionnement de rSw1 en présence ou en absence de l'expression de l'antisens ARN, nous utilisons respectivement les antisens as1 et as13 ; as1 étant l'antisens spécifique de rSw1 et as13 un antisens spécifique de iSw13 qui ne devrait pas, à priori, s'hybrider sur le commutateur rSw1.

Cependant, le résultat d'un premier test du fonctionnement de rSw1 montre que l'un des deux antisens as1 ou as13 a un effet sur la croissance, visible par une turbidité différente en phase stationnaire (**Figure 54A**) : le rendement apparent[9] de croissance des bactéries exprimant as1 est supérieur par rapport à celui des bactéries exprimant as13. Ayant une conséquence sur le niveau de fluorescence/DO_{600} mesuré, cet effet sur la croissance remet en question l'effet de régulation que nous observons, également visible à partir de l'entrée en phase stationnaire (**Figure 54B**). Nous observons le même effet pour le commutateur inductible iSw13 (**Figure 59A**). Là aussi, le rendement apparent de croissance est réduit en présence de as13.

[9] L'effet sur la DO_{600} peut être dû soit à une différence dans le nombre de bactéries, et donc à une différence de biomasse, soit à une différence de forme des cellules ou de leur composition, induite par les effets sur la croissance.

Lorsqu'on calcule le taux d'expression spécifique (Fluorescence/OD_{600}), on observe l'effet attendu : une plus forte expression en présence d'as13 (Figure 59B). Cependant, les différences de rendement apparent introduisent une ambigüité dans l'interprétation des résultats. De plus, la dynamique d'expression d'iSw13 en phase stationnaire n'est pas identique dans les deux conditions : tandis que l'activité du gène rapporteur reste constante en présence d'as13, en présence d'as1 on observe une deuxième petite induction à 20h, ce qui complique encore la comparaison.

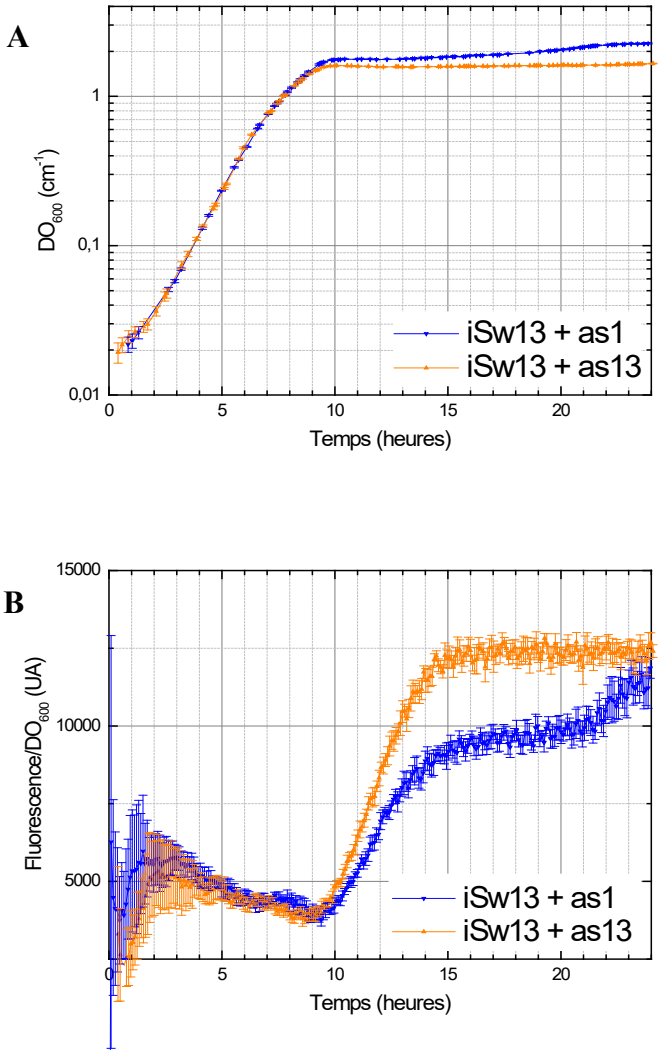


Figure 59 **L'antisens as13 réduit le rendement apparent de la croissance.** (A) DO_{600} en fonction du temps. La turbidité finale est réduite en présence d'as13. (B) Fluorescence relative, Fluorescence/DO_{600}, en fonction du temps. L'expression du gène rapporteur apparait plus forte en présence d'as13 dès l'entrée en phase stationnaire. En orange, bactéries exprimant le commutateur ARN inductible iSw13 et son antisens spécifique as13. En bleu, bactéries exprimant le commutateur ARN inductible iSw13 et un antisens non-spécifique as1. Les données correspondent à des moyennes et les barres d'erreur représentent des écarts-types obtenus à partir de 9 réplicats. Cultures dans un milieu minimum M9 + 0,2 % glucose + 40 µg/mL de kanamycine + 100 µg/mL d'ampicilline + 1 mM d'IPTG.

Nous avons cherché des causes potentielles de l'effet des ARN antisens sur le rendement de la croissance d'*E. coli*. Un alignement de séquences entre le chromosome bactérien et les séquences des deux antisens as1 et as13 montre que le transcrit de l'antisens ARN as13 peut potentiellement s'hybrider avec un transcrit du gène *hycE* (**Figure 60**) qui joue un rôle important dans le métabolisme de la bactérie, plus particulièrement en entrée de phase stationnaire où la disponibilité en oxygène devient limitée à cause d'une trop forte densité cellulaire. Il s'agit d'un gène qui code pour une protéine correspondant à la grande sous-unité d'une hydrogénase 3, faisant elle-même partie d'un complexe FHL (formate hydrogen lyase) qui dégrade le formate en CO_2 et H_2. Le formate est un sous-produit du métabolisme du glucose en condition anaérobique. Cette dégradation du formate en CO_2 et H_2 participe notamment à la désacidification du milieu de culture, permettant de réduire la mort des bactéries [114][115]. Une perturbation de l'expression de ce gène en fin de phase exponentielle pourrait alors perturber la croissance des bactéries en phase stationnaire, ce qui semble effectivement être le cas pour les différents tests que nous réalisons en présence de l'expression de l'antisens ARN as13, car le rendement de croissance en présence de l'expression de cet antisens est plus faible (**Figure 54A** et **Figure 59A**) par rapport à celui obtenu par exemple en présence de l'expression de l'antisens as1.

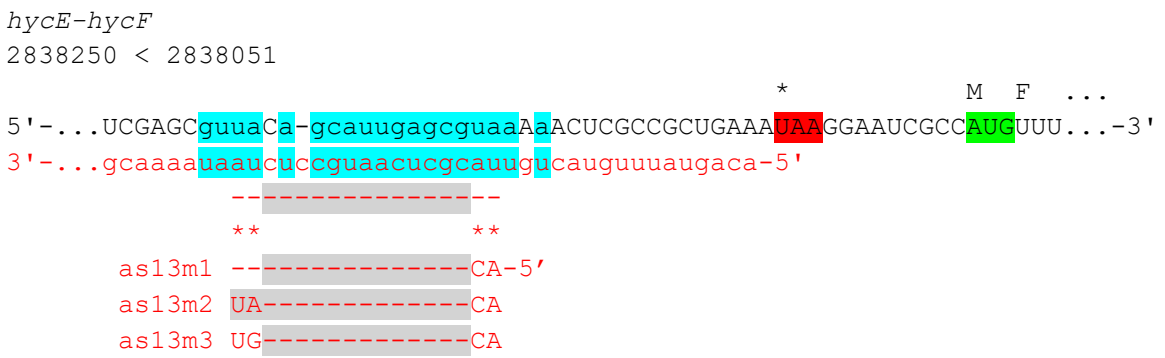


Figure 60 Alignement de séquences entre l'antisens as13 et la fin de la région ORF de *hycE* sur l'ARN messager de l'opéron *hycE-hycF*. L'ARN as13 peut s'hybrider sur l'ARN messager du gène *hycE* qui joue un rôle important dans le métabolisme de la bactérie en condition anaérobie. En bleu sont représentés les nucléotides complémentaires entre les transcrits de *hycE* et as13 ; en noir, le transcrit du gène *hycE* ; en rouge, le transcrit de l'antisens as13 ; soulignée en rouge la région contenant la séquence antisens et surlignée en gris la partie qui s'hybride effectivement au commutateur iSw13. Le codon start du gène *hycF* est surligné en vert, il code pour la méthionine (M). Le codon stop du gène *hycE* est surligné en rouge.

À la vue de ces résultats, il est donc important de trouver un antisens non-spécifique qui n'aurait pas d'effet sur la croissance. Un tel antisens pourrait servir de contrôle non-spécifique fiable pour le test du fonctionnement des commutateurs ARN en absence d'antisens ARN

spécifique. Aussi, il faudrait modifier la séquence de l'antisens ARN as13, de façon que son transcrit n'interfère plus avec le transcrit du gène *hycE* et n'affecte plus la croissance des bactéries et donc la mesure du niveau d'expression du gène rapporteur.

2.2. Choix d'un bon contrôle de spécificité pour l'effet des antisens ARN effecteurs sur les commutateurs ARN

Pour tester le fonctionnement des commutateurs ARN en absence d'expression de leurs antisens ARN spécifiques, nous envisageons d'utiliser un plasmide d'expression de l'antisens qui exprime soit un transcrit n'ayant pas de région de complémentarité avec le commutateur, soit un transcrit ne contenant que la séquence du $DSR_{hairpin}$ et la séquence de terminateur $T1_{rrnB}$ sans aucune autre séquence entre les deux, et qu'on appellera « as0 ».

Pour l'antisens ARN n'ayant pas de région de complémentarité avec le commutateur ARN, nous choisissons d'utiliser la séquence inverse de l'antisens as1, qui est donc ni complémentaire à rSw1, ni complémentaire à iSw13. Nous appellerons cet antisens « as1_inv ».

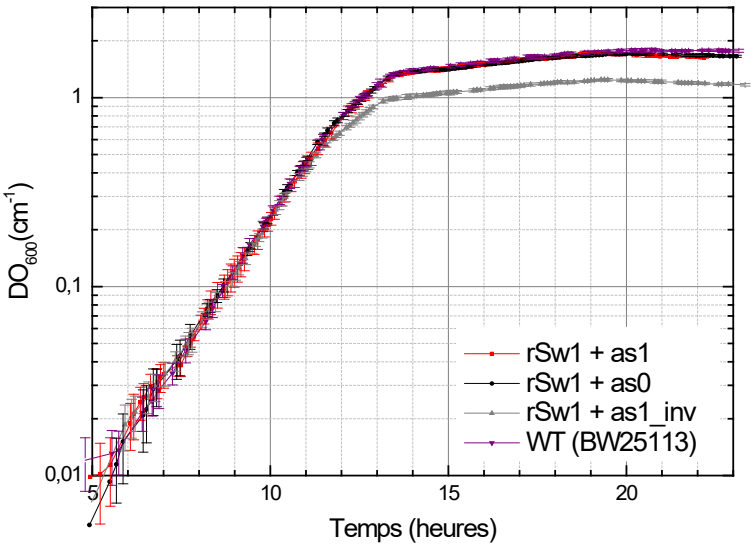


Figure 61 L'antisens as0 n'affecte pas la croissance des bactéries. Les courbes montrent la DO_{600} en fonction du temps pour la souche sauvage BW25113 sans plasmide (en mauve), la souche exprimant le commutateur rSw1 et son antisens spécifique as1 (en rouge), la souche exprimant le commutateur rSw1 et l'antisens as0 (en noir), et la souche exprimant le commutateur rSw1 et l'antisens non-spécifique as1_inv (en gris). Les données correspondent à des moyennes et les barres d'erreur représentent des écarts-types obtenus à partir de 9 réplicats. Culture dans un milieu minimum M9 + 0,2 % glucose + 40 µg/mL de kanamycine + 100 µg/mL d'ampicilline + 20 µM d'IPTG. La souche sauvage BW25113 pousse en milieu minimum M9 sans antibiotique.

La **Figure 61** compare les courbes de croissance obtenues sur une souche sauvage BW5113, une souche exprimant le commutateur et son antisens spécifique, une souche exprimant le commutateur et l'antisens non-spécifique as1_inv, et une souche exprimant le commutateur et l'antisens as0.

La similarité des courbes de croissance de la souche sauvage et de la souche exprimant le commutateur rSw1 et son antisens spécifique as1, confirme que c'est bien as13 qui affecte la croissance des bactéries. Nous rappelons que le rendement de croissance en phase stationnaire de la souche exprimant le commutateur rSw1 et l'antisens as13 est plus faible que celui de la souche exprimant le commutateur rSw1 et l'antisens as1 (**Figure 54A**). L'antisens as0 semble également ne pas affecter la croissance des bactéries puisque la courbe de croissance de la souche exprimant le commutateur rSw1 et l'antisens as0 est similaire à celle de la souche sauvage, contrairement à celle de la souche exprimant le commutateur rSw1 et l'antisens as1_inv qui a un rendement de croissance plus faible en phase stationnaire par rapport aux deux autres.

Ces résultats montrent que l'antisens as0 est le candidat idéal pour l'étude du fonctionnement des commutateurs ARN en absence de l'expression de leurs antisens ARN spécifiques.

2.3. Optimisation de la séquence de l'antisens ARN as13 : réduction de l'effet sur la croissance des bactéries

Le résultat d'alignement de séquences du chromosome bactérien avec la séquence du transcrit de l'antisens as13 ($DSR_{hairpin}$ et terminateur $T1_{rrnB}$ inclus) montre que la séquence de l'antisens ainsi que certains nucléotides qui l'encadrent peuvent s'hybrider avec le transcrit du gène *hycE* **Figure 60**. La séquence de l'antisens complémentaire à la séquence du commutateur est encadrée de part et d'autre par 4 nucléotides (respectivement 2 nucléotides en 5' et 2 nucléotides en 3') non-appariés à la séquence du commutateur. Borer *et al*. [116][117] suggère qu'une double hélice d'ARN, ici représentée par l'hybridation entre l'antisens et le commutateur, serait stabilisée par la présence de ces bases non-appariées à cause d'un phénomène d'empilement (ou stacking). Or, si elles sont non-appariées à la séquence du commutateur, nous constatons d'après les résultats d'alignement de séquences indiqués sur la **Figure 60**, qu'elles le sont avec la séquence du transcrit du gène *hycE*. La stratégie utilisée pour empêcher l'hybridation entre les transcrits de l'antisens as13 et du gène *hycE* est donc de muter

ces bases de façon à réduire la complémentarité entre les deux ARN messagers et peut-être même augmenter la complémentarité entre l'antisens et le commutateur en contrepartie.

Trois variantes de la séquence as13, notés de 1 à 3, sont ainsi réalisés (Figure 60) :

- as13**m1** : mutations de deux nucléotides (UA → AC) du côté 5' de l'antisens, qui éliminent l'appariement des deux premières bases du côté 5' du transcrit de l'antisens avec le transcrit du gène *hycE*.
- as13**m2** : mutations des deux nucléotides du côté 5' (UA → AC) et du côté 3' (UA → AU) de l'antisens, qui éliminent l'hybridation des bases encadrant l'antisens avec le transcrit du gène *hycE*. Par ailleurs, les deux bases mutées en 3' rallongent la complémentarité de l'antisens avec le commutateur.
- as13**m3** : mutations des deux nucléotides du côté 5' (UA → AC) et du côté 3' (UA → GU) de l'antisens, qui éliminent également l'hybridation des bases encadrant l'antisens avec le transcrit du gène *hycE*.

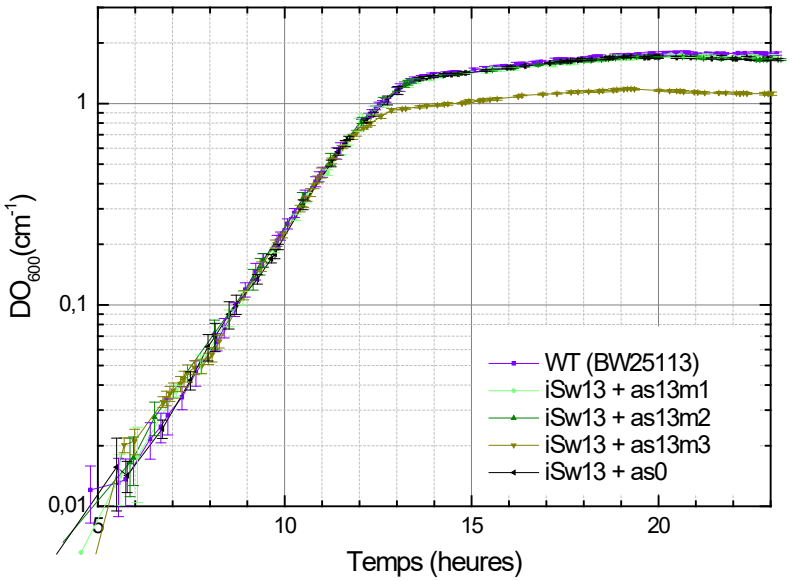


Figure 62 Comparaison des courbes de croissance des bactéries exprimant les différents mutants de l'antisens as13. Coexpression du commutateur iSw13 avec les différents mutants de l'antisens as13 : iSw13 + as13m1 (vert clair) ; iSw13 + as13m2 (vert foncé) ; iSw13 + as13m3 (vert kaki). Les courbes de croissance de la souche sauvage BW25113 (en mauve), ainsi que de la souche coexprimant le commutateur iSw13 et l'antisens non-spécifique as0 (en noir) sont également représentées. Les données correspondent à des moyennes et les barres d'erreur représentent des écarts-types obtenus à partir de 9 réplicats. Culture dans un milieu minimum M9 + 0,2 % glucose + 40 µg/mL de kanamycine + 100 µg/mL d'ampicilline + 20 µM d'IPTG. La souche sauvage BW25113 pousse en milieu minimum M9 sans antibiotiques.

A l'exception de la courbe de croissance de la souche coexprimant le commutateur iSw13 et l'antisens as13m3, les courbes de croissance des autres souches se ressemblent, avec des temps de doublement identiques. Plus précisément, le rendement de croissance de la souche coexprimant le commutateur iSw13 et l'antisens as13m3, en phase stationnaire, est plus faible comparé aux autres souches (**Figure 62**).

Les antisens as13m1 et as13m2 semblent donc être de bons candidats pour le test du fonctionnement du commutateur iSw13 en présence d'expression de l'antisens spécifique, qui peut alors être soit l'antisens as13m1, soit l'antisens as13m2.

Les séquences de ces deux antisens étant différentes l'une de l'autre, leurs cinétiques d'hybridation avec le commutateur iSw13 peuvent aussi être différentes. Par conséquent, il est possible que les efficacités de commutation du commutateur, initiée par ces deux antisens soient également différentes. Nous verrons dans le chapitre suivant lequel de ces deux antisens est meilleur pour induire une commutation du repliement du commutateur iSw13. Aussi, nous verrons plus en détails tous les résultats de test *in vivo* du fonctionnement des deux types de commutateurs : iSw13 (inductible) et rSw1 (répressible), en absence et en présence de l'expression de l'antisens qui leur est spécifique.

Chapitre 3 : Validation *in vivo*, du fonctionnement des commutateurs ARN inductible iSw13 et répressible rSw1

1. Introduction

Dans ce chapitre seront présentés les résultats de validation *in vivo* du fonctionnement des commutateurs ARN inductible iSw13 et répressible rSw1, en absence et en présence de l'expression de l'antisens qui leur est spécifique, respectivement as13 (as13m1 ou as13m2) et as1. Nous rappelons que les deux types de commutateurs testés ici sont issus de deux méthodes de design différentes : une méthode manuelle (rSw1) et une méthode semi-automatique (iSw13). Pour cette étude, le but final est bien évidemment de pouvoir valider la méthodologie générale de design semi-automatique utilisée pour la génération de séquences ARN fonctionnelles, ici des commutateurs ARN. Cependant, une partie importante de ce travail consiste également à vérifier le bon fonctionnement de la plateforme construite et utilisée pour réaliser la validation *in vivo* du fonctionnement des commutateurs ARN.

Les différentes optimisations faites précédemment ont ainsi abouties à la mise en place d'une plateforme *in vivo* de test du fonctionnement des différents commutateurs générés. Cette plateforme consiste en l'utilisation d'une bactérie *E. coli* comme organisme hôte, dans lequel

sont coexprimés le commutateur et l'antisens via deux plasmides d'expression différents. Pour le test du fonctionnement des commutateurs en absence d'expression des antisens qui leur sont spécifiques, il s'agit de coexprimer le commutateur ARN avec une séquence ARN ne contenant pas de région de complémentarité avec la séquence du commutateur. La transcription de cette séquence se fait via le même plasmide utilisé pour exprimer l'antisens. Le transcrit contient la séquence du $DSR_{hairpin}$ ainsi que la séquence du terminateur $T1_{rrnB}$.

Nous testerons d'abord le commutateur inductible, puis le commutateur répressible.

2. Fonctionnement du commutateur ARN inductible iSw13 dans la bactérie *E. coli*

La caractérisation du fonctionnement des commutateurs ARN se fait par la comparaison de l'intensité relative de fluorescence produite par la bactérie entre deux situations différentes : en coexpression avec un antisens effecteur spécifique du commutateur ou en coexpression avec un « antisens » non-spécifique (en fait, un transcrit sans région antisens). Nous rappelons qu'un rapporteur fluorescent (mKate2) est utilisé comme proxy pour déterminer le chemin de repliement emprunté par le commutateur se trouvant à son amont. Deux chemins de repliement peuvent être adoptés par celui-ci, en fonction de la présence ou non de l'antisens ARN qui lui est spécifique. Ces deux chemins de repliement du commutateur doivent conduire à la production de niveaux de fluorescence différents. Ainsi, en absence de l'antisens ARN qui lui est spécifique le commutateur ARN inductible iSw13 doit adopter un « état OFF ». Alors que l'hybridation avec l'antisens ARN qui lui est spécifique (as13m1 ou as13m2) doit induire une commutation de son chemin de repliement vers un « état ON », qui conduit à une augmentation du niveau de fluorescence.

Les résultats présentés sur la **Figure 63** montrent une augmentation de la fluorescence moyenne par bactérie en présence de l'antisens spécifique, principalement observée pendant la phase stationnaire, à partir de 13 heures de croissance environ.

Pour cette étude, deux antisens (as13m1 et as13m2) sont testés. Nous observons que l'augmentation de la fluorescence moyenne par bactérie induite en présence de l'antisens as13m2 est un petit peu plus élevée (30 %), par rapport à celle induite en présence de l'antisens as13m1 (26 %). L'efficacité de commutation initiée par l'antisens as13m2 est donc meilleure, probablement due aux mutations introduites dans la séquence de cet antisens qui ont permis de rallonger la région complémentaire au commutateur iSw13 et augmenter la stabilité ou la cinétique d'hybridation entre les deux ARN.

Par ailleurs, la similarité des courbes de croissance (Figure 63A) nous permet cette fois-ci d'affirmer que l'effet que nous observons sur la fluorescence relative en début de phase stationnaire est bien une conséquence de l'hybridation de l'antisens ARN sur le commutateur, induisant une commutation du repliement de celui-ci vers un « état ON », une expression du gène fluorescent à son aval et donc une augmentation de la fluorescence moyenne par bactérie. Il s'agit bien d'une induction de la transcription du commutateur iSw13 en présence de son antisens spécifique as13m2.

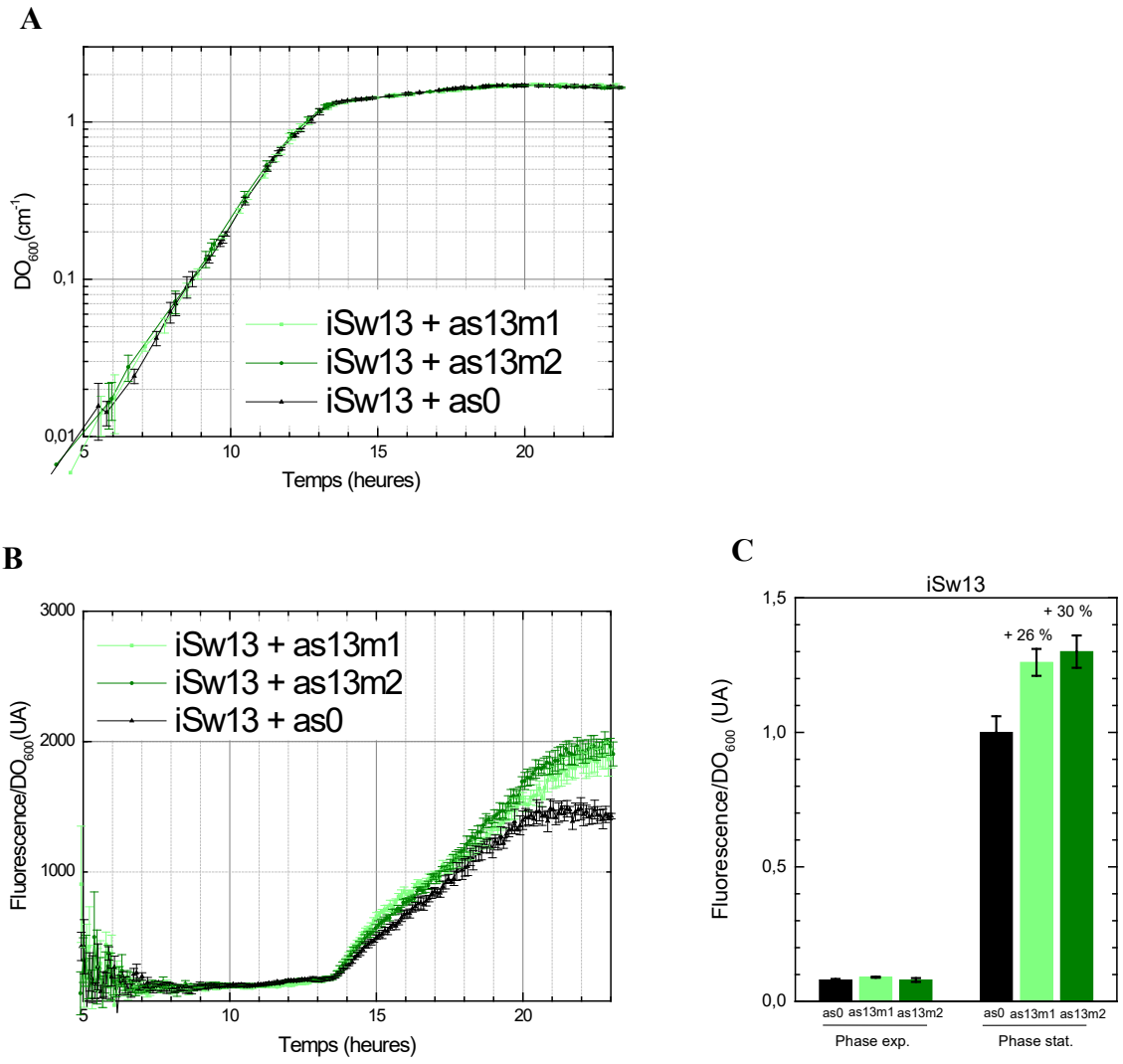


Figure 63 Le fonctionnement *in vivo* du commutateur inductible iSw13 : niveau d'expression du gène rapporteur en absence d'un antisens ARN spécifique (as13m1 ou as13m2) ou d'un transcrit sans région antisens (as0). (A) DO_{600} en fonction du temps. Les courbes de croissances de chaque culture sont identiques. (B) Fluorescences relatives, Fluorescence/DO_{600}, en fonction du temps. (C) Comparaison de la fluorescence relative en phase exponentielle et à 24h de croissance. L'intensité de la fluorescence relative en présence d'as0 en phase stationnaire est normalisée à 1. Les données correspondent à des moyennes et les barres d'erreur représentent des écarts-types obtenus à partir de 9 réplicats. Culture dans un milieu minimum M9 + 0,2 % glucose + 40 µg/mL de kanamycine + 100 µg/mL d'ampicilline + 20 µM d'IPTG.

3. Fonctionnement du commutateur ARN répressible rSw1 dans la bactérie *E. coli*

En accord avec le fonctionnement théorique des commutateurs ARN répressibles, les résultats de validation *in vivo* du fonctionnement du commutateur répressible rSw1 montrent que celui-ci se retrouve dans un « état ON » en absence de l'antisens as1. Par contre, en présence de cet antisens spécifique, la fluorescence moyenne par bactérie diminue d'environ 32 %, indiquant une commutation de repliement du commutateur rSw1 vers « l'état OFF » (Figure 64). Il s'agit bien d'une répression de la transcription du commutateur rSw1 en présence de son antisens spécifique as1.

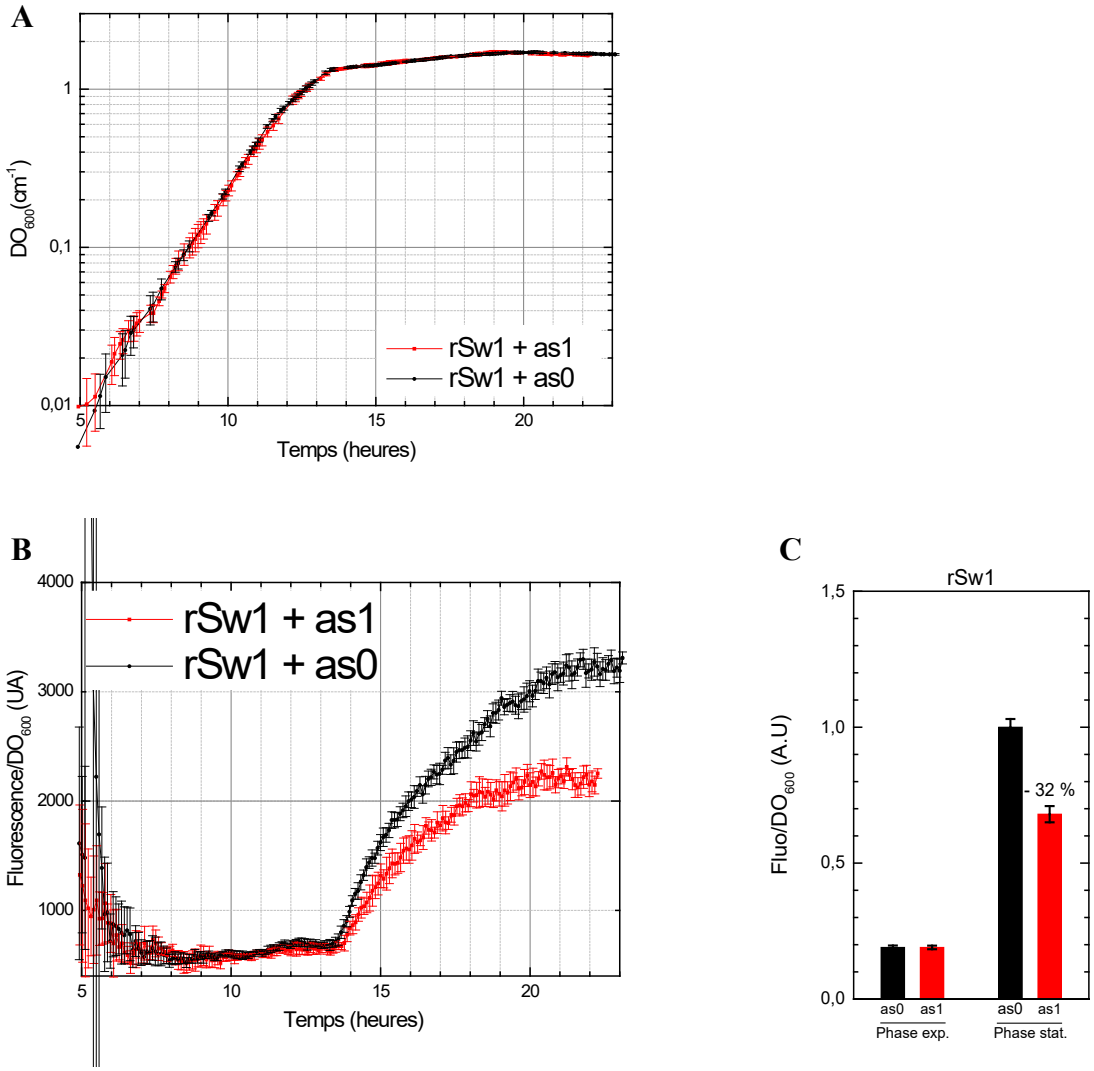


Figure 64 Le fonctionnement *in vivo* du commutateur répressible rSw1 en absence et en présence de l'antisens ARN (as1) qui lui est spécifique. (A) DO_{600} = f(t). (B) Fluorescence/DO_{600} = f(t). (C) Comparaison de l'intensité de fluorescence/DO_{600} mesurée en absence (rSw1 + as0) et présence (rSw1 + as1) de l'expression de l'antisens ARN spécifique au commutateur. Les données correspondent à des moyennes et les barres d'erreur représentent des écarts-types obtenus à partir de 9 réplicats. Culture dans un milieu minimum M9 + 0,2 % glucose + 40 µg/mL de kanamycine + 100 µg/mL d'ampicilline + 20 µM d'IPTG.

Il faut noter que ces deux dernières expériences sont faites dans un milieu de culture contenant 20 µM d'IPTG et non 1 mM comme c'était le cas dans les expériences des chapitres précédents. En diminuant le niveau d'expression du commutateur ARN, et donc en diminuant le nombre de transcrits de ce dernier, nous favorisons l'excès du transcrit de l'antisens par rapport à celui du commutateur, qui est une des conditions nécessaires pour que tout commutateur ARN puisse rencontrer avec la même probabilité un antisens ARN (pas de déplétion du stock de transcrits antisens).

4. Conclusions

Nous avons réussi à mettre en place une plateforme qui permet de valider le fonctionnement des commutateurs ARN chez la bactérie *E. coli*. Cette plateforme convient tout aussi bien à la validation du fonctionnement des commutateurs générés à partir d'une méthode manuelle de design, que des commutateurs générés à partir d'une méthode de design semi-automatique. Ainsi, nous avons réussi à valider le fonctionnement de deux commutateurs ARN, dont l'un a été généré par la méthode manuelle (rSw1) et l'autre par la méthode semi-automatique (iSw13). En accord avec les résultats de validation *in vitro*, nous montrons d'une part que le commutateur rSw1 est bien dans un « état ON » en absence de l'antisens as1 et dans un « état OFF » en sa présence ; et d'autre part, que le commutateur iSw13 est bien dans un « état OFF » en absence de l'antisens as13 et dans un « état ON » en sa présence.

Cependant, au lieu d'observer un effet de commutation en phase exponentielle de croissance, comme nous nous y attendions, c'est en phase stationnaire de croissance que ces effets sont observés. Or, nous avons vu que se placer en phase exponentielle de croissance signifie se placer en condition de croissance équilibrée, durant laquelle tous les constituants cellulaires sont synthétisés à des vitesses constantes par rapport aux autres. Ainsi on s'assure que le changement d'état du commutateur est uniquement dû à la présence de l'antisens ARN et non à d'autres paramètres. L'interprétation des effets en phase stationnaire est plus compliquée. Dans l'état actuel de nos connaissances, nous ne sommes pas capables de donner une réponse précise pour expliquer ces effets de commutation observés uniquement en phase stationnaire de croissance. Il est possible qu'il s'agisse d'un effet de régulation que le promoteur pRNAI subit pendant cette période ou la transition vers cette phase, ce qui affecterait l'expression de l'antisens et donc l'effet sur le commutateur. D'autres possibilités comprennent des changements globaux de la stabilité des petits ARN à l'arrivée en phase stationnaire ou l'implication de facteurs spécifiques de la phase stationnaire qui favoriseraient l'hybridation entre deux ARN.

En plus de la mise au point de la plateforme de mesure *in vivo* et la validation de deux commutateurs, les différentes optimisations que nous avons réalisées nous ont permis d'identifier des caractéristiques qu'il est important d'inclure dans la méthodologie de design des commutateurs ARN. Par exemple, il est important de vérifier que le commutateur ne contient pas de séquences se rapprochant des séquences consensus de promoteur. Nous avons également constaté que des séquences synthétiques, donc presque aléatoires, peuvent significativement influencer la physiologie globale de la cellule. Dans le futur, il sera probablement indiqué de minimiser la probabilité que l'antisens s'hybride sur d'autres transcrits de la bactérie, et vérifier que l'expression de l'antisens n'interfère pas avec le métabolisme de l'organisme hôte dans lequel on réalise la validation *in vivo*.

Partie V : Conclusions, discussions et perspectives

Conception de commutateurs ARN

Dans ce travail de thèse, nous présentons une méthodologie d'ingénierie [*Mladek et Dawid* (pas encore publié)][2] qui permet de concevoir de façon semi-automatique des riborégulateurs entièrement synthétiques, s'inspirant de mécanisme de terminaison/anti-terminaison que l'on retrouve chez la bactérie (par exemple la régulation du nombre de copies du plasmide pT181 chez *Staphylococcus aureus* [118][76]).

Cette approche s'appuie sur :

(i) des principes théoriques architecturaux permettant de guider le repliement co-transcriptionnel de l'ARN vers une structure fonctionnelle finale souhaitée ;

(ii) une optimisation des séquences basée sur une sélection de certains motifs de séquence ;

(iii) une validation par des simulations numériques de repliement co-transcriptionnels.

Grâce à cette méthode, deux séries de riborégulateurs ont pu être conçus : un commutateur ARN répressible qui, par défaut, est dans un état ON et réprimé en présence d'un petit ARN antisens effecteur correspondant à un antisens ARN ; trois commutateurs ARN inductibles qui, par défaut, sont dans un état OFF et induits en présence d'un petit ARN antisens effecteur.

Nous évaluons ici cette méthodologie d'ingénierie en testant les commutateurs ARN générées, d'abord par transcription *in vitro*, puis par des expériences fonctionnelles *in vivo*, plus particulièrement dans la bactérie *Escherichia coli*.

Validation *in vitro* des commutateurs

La transcription *in vitro* nous a permis de valider le fonctionnement du commutateur répressible (rSw7) et du commutateur inductible (iSw13). Cependant, nous montrons aussi que contrairement à ce que la simulation du repliement co-transcriptionnel sur Kinefold nous suggère, les commutateurs que nous testons ne fonctionnent pas toujours de manière satisfaisante.

Ainsi nous montrons, par exemple, que les deux autres commutateurs inductibles que nous testons ne fonctionnent pas correctement (**Figure 23**) : ils sont par défaut déjà dans un état ON alors que l'on s'attendrait à ce qu'ils soient dans un état OFF. Cela suggère que le terminateur présent dans chaque commutateur est déficient, laissant la transcription se poursuivre au-delà du commutateur. Une analyse plus approfondie des simulations du repliement co-transcriptionnel de ces mêmes séquences suggère que les terminateurs peu efficaces *in vitro* seraient associés à une tendance de ces derniers, soit à ne pas se former (ex : iSw11 dont le terminateur n'est pas formé dans 2 simulations sur 5 et dont la transcription *in vitro* conduit à seulement 13 % de terminaison), soit à se former, mais de façon tardive ou incomplète (ex : iSw3 dont le terminateur se forme tardivement lors des simulations et dont la transcription *in vitro* conduit à seulement 0% de terminaison). Nous parvenons toutefois à valider le fonctionnement du commutateur inductible iSw13 en présence d'antisens ARN (as13), avec 87 % d'induction (**Figure 24**).

Par ailleurs, malgré le fait qu'on observe bien une commutation ON/OFF du commutateur répressible rSw7 en présence d'antisens ARN (as7), nous n'obtenons malheureusement qu'une répression partielle (30%) de l'expression, même en présence d'une très forte concentration d'antisens ARN (**Figure 25**). Plusieurs hypothèses sont avancées pour expliquer ce défaut de régulation : il peut s'agir soit d'une faible stabilité de l'interaction entre l'antisens ARN et sa cible sur le commutateur ARN, soit d'une formation tardive ou partielle du terminateur censé se former dans le commutateur, ou bien d'une faible efficacité de terminaison de la transcription par ce terminateur. Pour discriminer entre ces différentes hypothèses, nous avons modifié la séquence de rSw7 de façon à ce que la région cible de l'antisens sur le commutateur soit délétée (rSw7_ΔAT), favorisant ainsi une formation constitutive du terminateur au cours de la transcription. Malgré la formation systématique du terminateur observée dans les simulations du repliement co-transcriptionnel par Kinefold, nous observons *in vitro* un taux de terminaison similaire au résultat précédent (rSw7 en présence d'une concentration saturante d'antisens

ARN), ce qui indique une faible efficacité intrinsèque de terminaison de la transcription par le terminateur du commutateur.

Développement d'une plateforme de test *in vivo*

Dans une seconde partie de ce travail, nous développons une plateforme de test dans la bactérie *Escherichia coli*, qui permet de valider le fonctionnement *in vivo* des commutateurs ARN précédemment validés *in vitro*. Nous utilisons pour cela, le commutateur inductible iSw13 et un commutateur répressible rSw1, de structure similaire à celle de rSw7, mais conçu à partir d'une méthodologie manuelle. Nous focalisons sur ce commutateur car il a montré une meilleure gamme dynamique de régulation *in vitro* dans un travail antérieur par *Dawid et al.* [2]. Le développement de cette plateforme aborde des défis majeurs, notamment la conception de circuits d'interaction d'ARN à l'interface avec la machinerie cellulaire. Nous avons réussi à développer une plateforme dans laquelle les expressions des deux espèces d'ARN (commutateur ARN et antisens ARN) se font à partir de vecteurs plasmidiques différents (différents nombres de copies par bactérie) et dans lesquels leurs transcriptions sont contrôlées de manière indépendante par des promoteurs distincts. Ces commutateurs contrôlent notamment l'expression d'une protéine fluorescente, de sorte que la fluorescence moyenne des bactéries reflète indirectement le fonctionnement du commutateur et donc son état de repliement.

Pour cette caractérisation *in vivo*, toutes les mesures sont réalisées dans un lecteur de microplaque. Nous mesurons alors une valeur d'intensité de fluorescence, que nous rapportons ensuite à une valeur de densité optique qui reflète la quantité de bactéries présentes dans la culture pendant la mesure. Nous mesurons ainsi une fluorescence relative, correspondant à une fluorescence moyenne par bactérie

Optimisations

1. Mesures en présence de billes

Lors des expériences en microplaques, nous constatons qu'à très forte densité bactérienne (en phase stationnaire), les bactéries ont tendance à s'agréger de façon inhomogène au fond des puits de la microplaque, introduisant un biais lors des mesures d'absorbance et affectant ainsi les valeurs finales de fluorescence par bactéries. Pour résoudre ce problème nous décidons alors d'introduire dans chaque puits de la microplaque, une bille de verre de 2 mm de diamètre, qui sert à mieux homogénéiser la culture. L'inclusion de la bille de verre améliore également la

croissance en phase exponentielle, probablement par une amélioration de l'oxygénation. La contrepartie de cette amélioration, qui nous permet d'obtenir des courbes de croissance très reproductibles, est une analyse plus compliquée afin d'éliminer les mesures d'absorbance « aberrantes » dans les cas où le faisceau de mesure frappe la bille. Ces différents aspects techniques sont très rarement abordés, voire même ignorés dans certaines études similaires, ce qui peut rendre discutable les analyses des résultats.

2. <u>Améliorations du design des commutateurs</u>

Lors de la caractérisation *in vivo*, nous avons également constaté que la séquence du commutateur répressible rSw1 contenait des séquences promotrices qui provoquaient une fuite d'expression, indépendante de l'activité du promoteur utilisé pour contrôler la transcription du commutateur. Cette fuite avait un impact critique sur le fonctionnement apparent du commutateur rSw1, car sa gamme dynamique de régulation apparente s'en trouvait artificiellement réduite.

Par ailleurs, nous avons constaté que le rendement de croissance des bactéries en phase stationnaire était plus faible pour celles qui exprimaient l'antisens ARN as13, que celles qui ne l'exprimaient pas. Nous nous sommes rendu compte qu'en raison du contexte de séquence spécifique initialement choisi, cela générait un transcrit d'as13 qui pouvait s'hybrider sur l'ARN messager du gène *hycE* du génome bactérien. Ce gène a un rôle critique dans la croissance bactérienne, notamment lorsque l'oxygène devient limitant, ce qui peut être le cas à haute densité de cellules et donc en fin de phase exponentielle ou en entrée en phase stationnaire. En modifiant le contexte de séquence de l'antisens as13, pour diminuer la complémentarité avec l'ARN messager de *hycE*, nous avons restauré la croissance au même niveau que celui observé avec la bactérie sauvage.

De nouveaux, ces résultats apportent de nouvelles suggestions, pour renforcer les performances de la méthodologie d'ingénierie de séquences de commutateur ARN, en rajoutant dans les critères de design des séquences de commutateurs l'exclusion des séquences se rapprochant des séquences consensus de promoteur naturel ; mais aussi la possibilité de vérifier que les antisens obtenus n'interfèrent pas avec le métabolisme cellulaire car cet antisens se serait hybridé avec un gène ayant un rôle critique dans le métabolisme.

Évaluation de deux commutateurs optimisés sur la plateforme

Après toutes les optimisations, nous avons testé sur la plateforme un commutateur inductible et un commutateur répressible. Il s'avère que, contrairement à ce que nous attendions,

l'effet de régulation due au fonctionnement du commutateur n'est pas observé en phase exponentielle de croissance des bactéries, et donc en régime permanent, mais seulement à partir de l'entrée en phase stationnaire. Toutefois, nous observons une régulation claire, que l'ensemble de nos contrôles permettent d'attribuer au fonctionnement des commutateurs, commutés spécifiquement par les ARN antisens effecteurs associés.

Perspectives d'améliorations

Nos expériences montrent que nous sommes capables de dessiner et construire des commutateurs ARN fonctionnant de manière co-transcriptionnelle et régulant la transcription *in vivo*. Nous avons caractérisé de tels régulateurs *in vitro*, construit et optimisé une plateforme de mesure *in vivo* et évalué deux commutateurs sur cette plateforme. Les aspects les plus importants à améliorer dans notre système sont la faible gamme dynamique de nos régulateurs et le fait que nous observons leur fonctionnement qu'à partir de l'entrée en phase stationnaire. Nous pouvons envisager différentes pistes pour tenter de résoudre ces limitations.

1. Extension du fonctionnement des commutateurs à la phase exponentielle de croissance

Plusieurs hypothèses sont avancées pour expliquer le fait que nous observons une régulation qu'à partir de l'entrée en phase stationnaire. Il est possible que la cause soit l'utilisation du promoteur p_{RNAI} pour transcrire les antisens. Il s'agit d'un promoteur utilisé pour l'expression d'un antisens ARN impliqué dans la régulation du nombre de copies des plasmides pUC19, et l'activité de ce promoteur en fonction de la phase de croissance n'est pas connue.

Une autre possibilité pour expliquer l'observation repose sur la stabilité des ARN. Sachant que les antisens ARN sont de petite taille (≈ 20 nucléotides) et sont exprimés sous forme simple brin, ce qui les rend plus fragiles face aux RNases, nous pouvons aussi nous interroger sur leur stabilité. En particulier, il n'est pas clair si leur stabilité varie en fonction de la phase de croissance. Le changement de stabilité de l'ARN est un des mécanismes de la régulation de l'expression génique [119]. Par exemple, le petit ARN MicA est régulé par la polynucléotide phosphorylase (PNPase) en fonction de la phase de croissance, dans ce cas avec une moins grande stabilité en phase stationnaire [120].

Pour vérifier ces différentes hypothèses, nous pourrions imaginer réaliser tout d'abord un Northern Blot qui permettrait de quantifier le niveau d'expression de l'antisens dans les deux phases de croissance de la bactérie (exponentielle et stationnaire), lorsque cette expression est sous le contrôle du promoteur p_{RNAI}. Avec la même méthode, nous pourrions ensuite mesurer la stabilité des antisens ARN pendant les 2 phases de croissance.

2. Amélioration des critères pour le design

Nos résultats indiquent que la méthodologie de design pourrait être renforcée en utilisant des critères plus stricts pour la sélection des séquences des commutateurs et plus particulièrement sur la qualité de formation des terminateurs telle qu'observée dans les simulations de repliement co-transcriptionnel. En ce qui concerne la faible efficacité intrinsèque de terminaison de certains des terminateurs, il semblerait qu'elle soit une conséquence d'un effet de séquence puisque les critères que nous utilisons pour constituer un terminateur sont les mêmes pour tous et sont généralement ceux admis pour constituer un bon terminateur [121]. Cependant, dans l'état actuel de nos connaissances [121][122][123], il nous est encore difficile de prévoir l'efficacité de terminaison d'un terminateur à partir de sa séquence. Une solution serait de tester l'efficacité de terminaison de différentes séquences de terminateurs dans des conditions bien contrôlées, comme le propose les travaux récents de Chen *et al.* [124] et Cambray *et al.* [100], et d'en déduire les critères importants pour constituer un terminateur efficace. Des études récentes montrent également que l'efficacité d'un terminateur *in vivo* peut dépendre des conditions de croissance [125]. La prise en compte de telles études génomiques pourrait aider à améliorer le design des terminateurs. De plus, il est connu que l'efficacité de la terminaison dépend également indirectement du promoteur [126][127] : des effets de coopération entre ARN-polymérases font qu'un promoteur fort peut diminuer l'efficacité d'un terminateur simple en aval. Dans notre design, on pourrait donc envisager non seulement d'optimiser le terminateur, mais potentiellement également le promoteur.

3. Augmenter la stabilité de l'ARN antisens

Nous n'avons pas mesuré la stabilité de nos ARN antisens *in vivo*. Il est tout à fait possible qu'ils soient dégradés rapidement. La structure de nos ARN antisens est très simple, comprenant un DSR faisant office de tige-boucle, la séquence antisens elle-même, suivie d'un terminateur simple fort. Les tiges-boucles du DSR et du terminateur ont tendance à protéger la molécule envers une dégradation en 5' et en 3'. Nous pourrions envisager de renforcer la structure secondaire côté 5' de l'ARN antisens. Également, fournir des sites de liaison de chaperonnes ARN, tels que Hfq ou ProQ, pourrait stabiliser l'ARN [128].

4. Augmenter la vitesse d'hybridation entre ARN antisens et cible

Il est tout à fait possible que la vitesse de formation de l'hybride entre ARN antisens et ARN commutateur soit limitante pour le fonctionnement du système. Un grand nombre de régulateurs ARN cible la traduction. Dans ces commutateurs sous contrôle thermodynamique, le temps entre la synthèse du commutateur et la reconnaissance par l'ARN antisens peut être

plus long. Même si la régulation ne fonctionnerait pas au premier passage des ribosomes, un temps significatif s'écoule entre deux initiations successives de la traduction. Par contre, un régulateur de la transcription doit former la structure appropriée très rapidement puisque l'ARN polymérase avance à environ un nucléotide par 20 millisecondes.

Des chaperonnes ARN spécifiques, comme Hfq, peuvent accélérer la localisation et l'hybridation de l'ARN antisens avec sa cible [61]. Pour que Hfq reconnaisse efficacement l'ARN antisens, il serait possible d'intégrer au design une séquence de nucléation, « seed sequence » [129]. Une dizaine de nucléotides en 5' sont suffisants pour constituer une séquence de nucléation efficace [130]. Bien sûr, ce rajout implique de redessiner le commutateur ARN.

D'autres tests simples pour une exploration rapide d'autres conditions thermocinétiques pour tenter d'améliorer l'hybridation consisteraient à conduire des expériences à des températures plus basses pour réduire la vitesse de transcription ou dans des souches dans lesquelles la vitesse de transcription par l'ARN polymérase est réduite. Il faut toutefois noter que la température est un paramètre très général, qui peut affecter dans un sens ou dans l'autre de nombreux processus : change la stabilité thermodynamique des structures secondaires de l'ARN et donc affecter le repliement co-transcriptionnel, affecter la cinétique des enzymes et donc les taux de transcription et de dégradation des ARN antisens, affecter la processivité des ARN polymérase ou l'efficacité de terminaison des terminateurs…

Perspectives concernant les principes de design

Un des résultats clés de ce travail est de disposer à présent d'une plateforme expérimentale complète de test de commutateurs ARN synthétiques, *in vitro* et *in vivo*. Cette situation permet à présent d'envisager d'explorer les principes de design de tels commutateurs, à savoir explorer de façon plus systématiques différentes architectures de commutateurs, différents critères de design de séquences, etc. Le but étant alors de pouvoir étudier le lien séquence–repliement–fonction pour l'ARN, dans un contexte de repliement co-transcriptionnel et par une approche d'ingénierie.

Partie VI : Annexes

Annexe A : Listes des plasmides et souches utilisés

1. Listes des plasmides

Plasmides	Description	Taille
pGEJ1701	p_{LlacO1}-**iSw13**-DSRp_L-RBS-mKate2-$T1_{rrnB}$-pSC101ori-Kan^R	4534
pGEJ1703	p_{LlacO1}-**rSw1**-DSRp_L-RBS-mKate2-$T1_{rrnB}$-pSC101ori-Kan^R	4482
pGEJ1707	p_{LtetO1}-DSRp_L-**as1**-$T1_{rrnB}$-ColE1ori-T0-tetR-P_{N25}-Cm^R	2914
pGEJ1708	p_{LtetO1}-DSRp_L-**as13**-$T1_{rrnB}$-ColE1ori-T0-tetR-P_{N25}-Cm^R	2912
pGEJ1711	p_{LtetO1}-RBS-mKate2-$T1_{rrnB}$-ColE1ori-T0-tetR-P_{N25}-Cm^R	3650
pGEJ1714	p_{LtetO1}-DSRp_L-**as1**-$T1_{rrnB}$-Amp^R-pUCori	2199
pGEJ1715	p_{LtetO1}-DSRp_L-**as13**-$T1_{rrnB}$-Amp^R-pUCori	2197
pGEJ1716	p_{LtetO1}-DSRp_{PL}-RBS-mKate2-$T1_{rrnB}$-Amp^R-pUCori	2935
pGEJ1717	p_{LlacO1}-**rSw1.v2**-DSRp_L-RBS-mKate2-$T1_{rrnB}$-pSC101ori-Kan^R	4482
pGEJ1718	p_{RNAI}-DSRp_L-**as1**-$T1_{rrnB}$-Amp^R-pUCori	2263
pGEJ1719	p_{RNAI}-DSRp_{RNAI}-**as1**-$T1_{rrnB}$-Amp^R-pUCori	2282
pGEJ1720	p_{RNAI}-$DSR_{hairpin}$-**as1**-$T1_{rrnB}$-Amp^R-pUCori	2259
pGEJ1721	p_{RNAI}-$DSR_{hairpin}$-**as1**-T_{RNAI}-Amp^R-pUCori	2194
pGEJ1722	p_{RNAI}-DSRp_{RNAI}-**as13**-$T1_{rrnB}$-Amp^R-pUCori	2280
pGEJ1723	p_{RNAI}-$DSR_{hairpin}$-**as13**-$T1_{rrnB}$-Amp^R-pUCori	2257
pGEJ1724	p_{RNAI}-$DSR_{hairpin}$-**as13**-T_{RNAI}-Amp^R-pUCori	2192
pGEJ1733	p_{RNAI}-$DSR_{hairpin}$-$T1_{rrnB}$-Amp^R-pUCori	2240
pGEJ1746	p_{RNAI}-$DSR_{hairpin}$- **as13m1**-$T1_{rrnB}$-Amp^R-pUCori	2257
pGEJ1747	p_{RNAI}-$DSR_{hairpin}$- **as13m2**-$T1_{rrnB}$-Amp^R-pUCori	2257
pGEJ1748	p_{RNAI}-$DSR_{hairpin}$- **as13m3**-$T1_{rrnB}$-Amp^R-pUCori	2257
pGEJ1749	p_{RNAI}-$DSR_{hairpin}$- **as1_inv**-$T1_{rrnB}$-Amp^R-pUCori	2255

DSRp_L : DSR de p_{LlacO1} ou p_{LtetO1}

DSRp_{RNAI} : DSR de p_{RNAI}

2. Liste des souches de bactéries

Souche	Caractéristiques
BW25113	Δ(*araD-araB*)*567* Δ(*rhaD-rhaB*)*568 ΔlacZ4787* (::rrnB-3) *hsdR514 rph-1*
GEJ074	BW25113, pGEJ1707
GEJ078	BW25113, pGEJ1716

3. Liste des séquences des commutateurs ARN

rSw1	GGUCGUAACAACGACCAAAUGCUUCAUUACUUCGGAAUCUAUGAAAUAUCAUA GAUUAUAGAUUUUAUCAAGAAUACUUGAUUUUCAUAGAAUACUAUGAUUUAAU CUAUAUAUUAUAACCCGAAGUAUCAAGCAUGAUACUUCGGGUUUUUUGAACUC GUUUAGUAUCAAUAU
rSw1.v2	GGUCGUAACAACGACCAAAUGCUUCAUUACUUCGGAAUCUAUGAAAUAUCAUA GAUUAUAGA**C**UUUAUC**G**AGAAUACU**C**GAUUUUCAU**G**GAAUACCAUGAUUUA**G**U CUAUAUAUUAUAACCCGAAGUAUCAAGCAUGAUACUUCGGGUUUUUUGAACUC GUUUAGUAUCAAUAU
rSw7	GGUCGUAACAACGACCAUUCCGUACCUGAGUCAUGACGUAAUGCAAUCGCAUU ACGUUCUGCGCCCGUCUAGCAGCCUAGACACAGUCUCCGCCGAGACUACAGCG CAGGCCCUUCACCGCAUGGCUCGAUACGGAUCGAGCCAUGCUUUUUUAAGGCA UCACGCUCGAAACAA
iSw13	GGUCGUAACAACGACCACCGAUGAGGCAUUGAGCGCAGCCAGUGAAGCACCCU UCACUGUCAUAGCUCCAGAUCCUGCCCCAGGAUCACCGAGUCCAAACGGACUC GCCGGAGCUAUGAACUACUCAAUGUCUCGUCCAUUACCAACCUGGUAAUCGAG ACAUUCGAUUGCCAUUUUUUUUUUAAGAAAUGAACGCAAUCGC
iSw11	GGUCGUAACAACGACCAAACGCGGUGAUGCUUAGAGCUAACAGUCCGAAACCG GACUGUACCUGCUAUCCCUUCGUGCCUCACGAAGCCGUCUACGCCCUCGUAGA CCCGAUAGCAGGUCAGAUAAGCGUCAUCGCGCAAUACAAUCCUUGUAUUCGAU GACGUUGAAUGCAAUUUUUUUUUUAAGAAAUUCUCGCAUUCAC
iSw3	GGUCGUAACAACGACCAAUUAUGCAUACUCAUACGUCAGAUCUAGGUAAAAAC CUAGAUUAAUUGUUAUUAUCAAGUAAUACUUGAUUUUCAUUAGAAUACUAAUG AUUUUAACAAUUAUAAUAUGAGUGUGCGUGGAACGCAAAGCGUUCCAUCACUC GGUGGAACGCUUUUUUUUUGAGAGAGCGUCAUUAAC

Les parties soulignées correspondent aux régions cibles des antisens ARN

En gras sont représentés les nucléotides mutés pour passer de rSw1 à rSw1.v2

4. Liste des coeurs de séquence des antisens ARN

as1	5′-AA**UCGAAGUAAUGAAGC**UC-3′
as7	5′-UA**CAUGACUCAGGUACG**AA-3′
as13	5′-UA**CGCUCAAUGCCUC**UA-3′
as11	5′-AA**UCUAAGCAUCACU**CC-3′
as3	5′-AU**CGUGUGAGUGUGC**UA-3′

Chaque commutateur ARN est associé à un antisens ARN portant le même numéro (exemple l'antisens as11 s'hybride avec le commutateur ARN rSw11.

Annexe B : Méthodes expérimentales

1. Composition des milieux de culture utilisés

1.1. Milieu LB (Luria-Bertani), pour 1 L

Composition	Quantité	Concentration finale
Bacto-tryptone	10 g	1 %
Extrait de levure	5 g	0,5 %
NaCl	5 g	85 mM
1 M NaOH	Ajout petit à petit jusqu'à pH = 7	

1.2. Milieu minimum M9, pour 1 L

Composition	Quantité	Concentration finale
Salts (10X)	100 mL	1 X
Glucose 20 %	10 mL	0,2 %
1 M $MgSO_4$	1 mL	1 mM
1 M $CaCl_2$	0,3 mL	0,3 mM
Vitamine B1 (1 mg/mL)	1 mL	1 µg/L
Trace elements 100X	10 mL	1X

Composition Salt (10X)

Na_2HPO_4-$2H_2O$	75,2 g/L
KH_2PO_4	30 g/L
NaCl	5 g/L
NH_4Cl	5 g/L

Trace elements 100X

EDTA	5 g/L
$FeCl_3$-$6H_2O$	0,83 g/L
$ZnCl_2$	84 mg/L
$CuCl_2$-$2H_2O$	13 mg/L
$CoCl_2$-$2H_2O$	10 mg/L
H_3BO_3	10 mg/L
$MnCl_2$-$4H_2O$	1,6 mg/L

2. Protocoles de biologie moléculaire et microbiologie

2.1. Transcription *in vitro*

Nous utilisons l'ARN polymérase d'*E. coli* (NEB M0551S) pour réaliser les réactions de transcription *in vitro*.

Pour 10 µL de volume réactionnel, mélanger les différents réactifs de la liste suivante :

	Quantité	**Concentration finale**
H_2O RNase free	qsp	
Tampon* (5X)	2 µL	1X
Matrice ADN (produits PCR)	Dépend de la concentration initiale de matrice	60 nM
rNTP 2,5 mM chacun	2 µL	0,5 mM
ARN polymérase 1000 U/ml	1 µL	1 U/µL

*La composition du tampon est donnée par le fabricant.

Laisser incuber le mélange à 37 °C pendant 1 heure, au bout de laquelle le mélange est transféré à – 20 °C pour arrêter la transcription.

2.2. Électrophorèse sur gel de polyacrylamide

Nous faisons migrer les produits de transcription *in vitro* dans un gel de polyacrylamide en condition dénaturante, c'est-à-dire en présence d'urée.

Composition du gel :

- 12 ml de polyacrylamide 6 % – urée 8M

  Cette solution est préalablement préparée en mélangent pour 100 ml de solution finale :

  – 15 ml d'acrylamide : bisacrylamide (au proportion 19 : 1) 40% (6 % final)
  – 10 ml de tampon TBE 10X (1X final)
  – qsp 100 ml H_2O RNase free
- 120 µL de persulfate d'ammonium (APS) 10 %
- 6 µL de N, N, N', N'-Tétraméthyléthylènediamine (TEMED)

Couler le mélange entre 2 plaques de verre espacées de 1,5 mm par des espaceurs pour faire un gel de dimension 16 cm x 14 cm x 1,5 mm

Une fois durci, monter l'assemblage précédent à la verticale sur une cellule d'électrophorèse remplie d'un tampon TBE 1X.

3. Réglages utilisés pour des mesures typiques de fluorescence et d'absorbance avec le TECAN Infinite M200 PRO

- Température : 37°C +/- 0,5 °C

- Mesure d'absorbance : longueur d'onde de 600 nm
  - Bande passante : 9 nm
  - Agitation avant mesure de l'absorbance pour l'homogénéisation de l'échantillon :
    - Agitation linéaire, amplitude de 3 mm, pendant 20 sec
    - Temps d'attente avant la mesure : 15 sec
  - Mesure :
    - Nombre de flash : 25

- Mesure de fluorescence de mKate2 : longueurs d'onde théoriques d'émission et d'excitation => 588/633 nm
    - Bande passante d'excitation : 9 nm
    - Bande passante d'émission : 20 nm
    - Agitation avant la mesure de fluorescence pour l'homogénéisation de l'échantillon :
        - Agitation orbitale, amplitude : 3 mm, pendant 20 sec
        - Temps d'attente avant la mesure : 15 sec
    - Mesure :
        - Nombre de flash : 25
        - Temps d'intégration : 200 µs – temps de mesure de l'intensité de fluorescence après excitation de l'échantillon. Soit 25 x 200 µs
        - Gain : 100 – ce paramètre permet d'ajuster l'amplification du signal, qui se fait lors de la conversion du signal de fluorescence en signal électronique, l'ajustement du gain permet de se placer dans la bonne gamme de détection par rapport aux niveaux d'intensité lumineuse hauts et bas à détecter.
        - Z-position : 20000 µm – ce paramètre permet d'ajuster la hauteur de l'objectif au-dessus de l'échantillon. La lumière d'excitation pouvant être réfléchie par l'échantillon, le réglage de Z-position permet de maximiser le signal rapport/bruit.

- Temps de l'expérience : 24 h

- La mesure au TECAN est gérée par un logiciel nommé I-control qui génère un fichier Excel dans lequel sont stockées les données brutes. Ces données sont présentées sous forme de tableau dans lequel chaque colonne correspond à un puits dans la microplaque et chaque ligne aux valeurs d'absorbance ou de fluorescence mesurées tous les 5 minutes environ. Pour chaque cycle de mesure, le premier temps de mesure est stocké dans la deuxième colonne.

Bibliographie

[1] T.R. Cech, The RNA worlds in context, Cold Spring Harb. Perspect. Biol. 4 (2012) 1–5. https://doi.org/10.1101/cshperspect.a006742.

[2] A. Dawid, B. Cayrol, H. Isambert, RNA synthetic biology inspired from bacteria: construction of transcription attenuators under antisense regulation., Phys. Biol. 6 (2009) 025007. https://doi.org/10.1088/1478-3975/6/2/025007.

[3] H. Lodish, A. Berk, S.L. Zipursky, P. Matsudaira, D. Baltimore, J. Darnell, Structure of Nucleic Acids, (2000). https://www.ncbi.nlm.nih.gov/books/NBK21514/#_ncbi_dlg_citbx_NBK21514 (accessed March 4, 2020).

[4] R. Wing, H. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura, R.E. Dickerson, Crystal structure analysis of a complete turn of B-DNA, Nature. 287 (1980) 755–758. https://doi.org/10.1038/287755a0.

[5] A. Torres-Larios, K.K. Swinger, A.S. Krasilnikov, T. Pan, A. Mondragón, Crystal structure of the RNA component of bacterial ribonuclease P, Nature. 437 (2005) 584–587. https://doi.org/10.1038/nature04074.

[6] N.B. Leontis, J. Stombaugh, E. Westhof, The non-Watson-Crick base pairs and their associated isostericity matrices., Nucleic Acids Res. 30 (2002) 3497–531. https://doi.org/10.1093/nar/gkf481.

[7] A. Dawid, F. Guillemot, C. Brème, V. Croquette, F. Heslot, Mechanically controlled DNA extrusion from a palindromic sequence by single molecule micromanipulation, Phys. Rev. Lett. 96 (2006) 10–13. https://doi.org/10.1103/PhysRevLett.96.188102.

[8] W. Saenger, Principles of nucleic acid structure, Springer-Verlag, 1984. https://books.google.fr/books?hl=fr&lr=&id=gaoJCAAAQBAJ&oi=fnd&pg=PA1&dq=Saenger,++W.++1984.+Principles+++of++nucleic+++acid++structure.&ots=VMkEzJtajx&sig=VPHWXtYLijxCnMOTnTgYCwv3Dm4#v=onepage&q&f=false (accessed March 11, 2020).

[9] J.R. Wyatt, I. Tinoco, RNA Structural Elements and RNA Function, 1993. www.cshlpress.com/copyright. (accessed March 11, 2020).

[10] K. Rietveld, R. Van Poelgeest, C.W.A. Pleij, J.H. Van Boom, L. Bosch, The tRNA-like structure at the 3' terminus of turnip yellow mosaic virus RNA. Differences and similarities with canonical tRNA, 1982. https://academic.oup.com/nar/article-abstract/10/6/1929/1080312 (accessed March 19, 2020).

[11] D.W. Staple, S.E. Butcher, Pseudoknots: RNA structures with diverse functions, PLoS Biol. 3 (2005) 0956–0959. https://doi.org/10.1371/journal.pbio.0030213.

[12] J.R. Wyatt, J.D. Puglisi, I. Tinoco, RNA folding: Pseudoknots, loops and bulges, BioEssays. 11 (1989) 100–106. https://doi.org/10.1002/bies.950110406.

[13] R.R. Gutell, N. Larsen, C.R. Woese, C.R. Woese2, Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective., Microbiol. Rev. 58 (1994) 10–26. http://mmbr.asm.org/ (accessed June 7, 2019).

[14] J. Parsch, J.M. Braverman, W. Stephan, Comparative sequence analysis and patterns of covariation in RNA secondary structures, Genetics. 154 (2000) 909–921.

[15] J.M.P.A.J.E.M. Institute, Method for predicting RNA secondary structure, Proc. Nat. Acad. Sci. USA Vol. 72, No. 6, Pp. 2017-2021, June 1975 Method. 72 (1975) 2017–2021.

[16] A.P.G. and C.W.A.P. F. H. D. van Batenburg, An APL-programmed Genetic Algorithm for the Prediction of RNA Secondary Structure, 1995.

[17] B.A. Shapiro, J. Navetta, A massively parallel genetic algorithm for RNA secondary structure prediction, J. Supercomput. 8 (1994) 195–207. https://doi.org/10.1007/BF01204728.

[18] G. Benedetti, S. Morosetti, A genetic algorithm to search for optimal and suboptimal RNA secondary structures, 1995.

[19] P.G. Higgs, RNA secondary structure: Physical and computational aspects, Q. Rev. Biophys. 33 (2000) 199–253. https://doi.org/10.1017/S0033583500003620.

[20] J. Pan, D. Thirumalai, S.A. Woodson, Folding of RNA Involves Parallel Pathways, 1997.

[21] M.S. Rook, D.K. Treiber, J.R. Williamson, Fast Folding Mutants of the Tetrahymena Group I Ribozyme Reveal a Rugged Folding Energy Landscape, 1998.

[22] J. Pan, M.L. Deras, S.A. Woodson, Fast Folding of a Ribozyme by Stabilizing Core Interactions: Evidence for Multiple Folding Pathways in RNA, 1999. http://www.idealibrary.com (accessed June 14, 2019).

[23] E. van B. and C.P. Jan Pieter Abrahams, Mirjam van den Berg, Prediction of RNA secondary structure, including pseudoknotting, by computer simulation, n.d. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC330835/pdf/nar00194-0180.pdf (accessed June 24, 2019).

[24] L. Wuju, W. Jiajin, Prediction of RNA secondary structure based on helical regions distribution, n.d.

[25] A.A. Mironov, V.F. Lebedev, A kinetic model of RNA folding, 1993.

[26] M. Schmitz, G. Steger, Description of RNA folding by "simulated annealing," J. Mol. Biol. (1996). https://doi.org/10.1006/jmbi.1996.0021.

[27] A. Xayaphoummine, T. Bucher, H. Isambert, Kinefold web server for RNA/DNA folding path and structure prediction including pseudoknots and knots, Nucleic Acids Res. 33 (2005) 605–610. https://doi.org/10.1093/nar/gki447.

[28] H. Isambert, E.D. Siggia, Modeling RNA folding paths with pseudoknots: application to hepatitis delta virus ribozyme., Proc. Natl. Acad. Sci. USA. 97 (2000) 6515–6520. https://doi.org/10.1073/pnas.110533697.

[29] a Xayaphoummine, T. Bucher, F. Thalmann, H. Isambert, Prediction and statistics of pseudoknots in RNA structures using exactly clustered stochastic simulations., Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 15310–15315. https://doi.org/10.1073/pnas.2536430100.

[30] V. Piomponi, Monitoring In Vivo expression of synthetic RNA Switches, (2018).

[31] A.A. Mironov, L.P. Dyakonova, A.E. Kister, A Kinetic Approach to the Prediction of RNA Secondary Structures, J. Biomol. Struct. Dyn. 2 (1985) 953–962. https://doi.org/10.1080/07391102.1985.10507611.

[32] H. Isambert, The jerky and knotty dynamics of RNA, Methods. 49 (2009) 1–7. https://doi.org/10.1016/j.ymeth.2009.06.005.

[33] D.H. Mathews, J. Sabina, M. Zuker, D.H. Turner, Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure., J. Mol. Biol. 288 (1999) 911–40. https://doi.org/10.1006/jmbi.1999.2700.

[34] M.J. Serra, D.H.B.T.-M. in E. Turner, Predicting thermodynamic properties of RNA, in: Energ. Biol. Macromol., Academic Press, 1995: pp. 242–261. https://doi.org/https://doi.org/10.1016/0076-6879(95)59047-1.

[35] J.D. Puglisi, I.B.T.-M. in E. Tinoco, Absorbance melting curves of RNA, in: RNA

Process. Part A Gen. Methods, Academic Press, 1989: pp. 304–325. https://doi.org/https://doi.org/10.1016/0076-6879(89)80108-9.

[36] V. Viasnoff, A. Meller, H. Isambert, DNA Nanomechanical Switches under Folding Kinetics Control, Nano Lett. 6 (2006) 101–104. https://doi.org/10.1021/nl052161c.

[37] S. Harlepp, T. Marchal, J. Robert, J.-F. Léger, a Xayaphoummine, H. Isambert, D. Chatenay, Probing complex RNA structures by mechanical force., Eur. Phys. J. E. Soft Matter. 12 (2003) 605–15. https://doi.org/10.1140/epje/e2004-00033-4.

[38] J. Chappell, K.E. Watters, M.K. Takahashi, J.B. Lucks, A renaissance in RNA synthetic biology: New mechanisms, applications and tools for the future, Curr. Opin. Chem. Biol. 28 (2015) 47–56. https://doi.org/10.1016/j.cbpa.2015.05.018.

[39] R. Penchovsky, R.R. Breaker, Computational design and experimental validation of oligonucleotide-sensing allosteric ribozymes, Nat. Biotechnol. 23 (2005). https://doi.org/10.1038/nbt1155.

[40] J.B. Lucks, L. Qi, V.K. Mutalik, D. Wang, A.P. Arkin, Versatile RNA-sensing transcriptional regulators for engineering genetic networks., Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 8617–22. https://doi.org/10.1073/pnas.1015741108.

[41] K. Hollands, S. Proshkin, S. Sklyarova, V. Epshtein, A. Mironov, E. Nudler, E.A. Groisman, Riboswitch control of Rho-dependent transcription termination, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 5376–5381. https://doi.org/10.1073/pnas.1112211109.

[42] F.J. Isaacs, D.J. Dwyer, C. Ding, D.D. Pervouchine, C.R. Cantor, J.J. Collins, Engineered riboregulators enable post-transcriptional control of gene expression., Nat. Biotechnol. 22 (2004) 841–7. https://doi.org/10.1038/nbt986.

[43] T.S. Bayer, C.D. Smolke, Programmable ligand-controlled riboregulators of eukaryotic gene expression, Nat. Biotechnol. 23 (2005) 337–343. https://doi.org/10.1038/nbt1069.

[44] J.M. Callura, D.J. Dwyer, F.J. Isaacs, C.R. Cantor, J.J. Collins, J.J.C. Designed, D.J.D. Performed, Tracking, tuning, and terminating microbial physiology using synthetic riboregulators, 107 (2010). https://doi.org/10.1073/pnas.1009747107.

[45] M. Mandal, R.R. Breaker, Gene regulation by riboswitches, Nat. Rev. Mol. Cell Biol. 5 (2004) 451–463. https://doi.org/10.1038/nrm1403.

[46] W.C. Winkler, R.R. Breaker, Regulation of Bacterial Gene Expression By Riboswitches, Annu. Rev. Microbiol. 59 (2005) 487–517. https://doi.org/doi:10.1146/annurev.micro.59.030804.121336.

[47] B.J. Tucker, R.R. Breaker, Riboswitches as versatile gene control elements, Curr. Opin. Struct. Biol. 15 (2005) 342–348. https://doi.org/10.1016/j.sbi.2005.05.003.

[48] A.S. Mironov, I. Gusarov, R. Rafikov, L.E. Lopez, K. Shatalin, R.A. Kreneva, D.A. Perumov, E. Nudler, Sensing small molecules by nascent RNA: A mechanism to control transcription in bacteria, Cell. 111 (2002) 747–756. https://doi.org/10.1016/S0092-8674(02)01134-0.

[49] A. Nahvi, N. Sudarsan, M.S. Ebert, X. Zou, K.L. Brown, R.R. Breaker, Genetic control by a metabolite binding mRNA, Chem. Biol. 9 (2002) 1043–1049. https://doi.org/10.1016/S1074-5521(02)00224-7.

[50] W.C. Winkler, S. Cohen-Chalamish, R.R. Breaker, An mRNA structure that controls gene expression by binding FMN, 2002. www.pnas.orgcgidoi10.1073pnas.212628899 (accessed July 1, 2019).

[51] R.R. Breaker, Riboswitches and translation control, Cold Spring Harb. Perspect. Biol. 10 (2018). https://doi.org/10.1101/cshperspect.a032797.

[52] A.D. Garst, A.L. Edwards, R.T. Batey, Riboswitches: Structures and Mechanisms, Cold Spring Harb Perspect Biol. (2011). https://doi.org/10.1101/cshperspect.a003533.

[53] J.N. Kim, R.R. Breaker, Purine sensing by riboswitches, Biol. Cell. 100 (2008) 1–11. https://doi.org/10.1042/BC20070088.

[54] F.C. (Frederick C. 1931-2016. Neidhardt, R. Curtiss, American Society for Microbiology., Escherichia coli and Salmonella : cellular and molecular biology, (n.d.) 2 v. (xx, 2822 p.) : ill. (some col.); 29 cm.

[55] P. Babitzke, C.S. Baker, T. Romeo, Regulation of translation initiation by RNA binding proteins, Annu. Rev. Microbiol. 63 (2009) 27–44. https://doi.org/10.1146/annurev.micro.091208.073514.

[56] O. Porrua, M. Boudvillain, D. Libri, Transcription Termination: Variations on Common Themes, Trends Genet. 32 (2016) 508–522. https://doi.org/10.1016/j.tig.2016.05.007.

[57] W.C. Winkler, R.R. Breaker, Regulation of Bacterial Gene Expression by Transcription Attenuation, Annu. Rev. Microbiol. 83 (2019).

[58] Z.F. Hallberg, Y. Su, R.Z. Kitto, M.C. Hammond, Engineering and in vivo applications of riboswitches, Annu. Rev. Biochem. 86 (2017) 515–539. https://doi.org/10.1146/annurev-biochem-060815-014628.

[59] T. Dutta, S. Srivastava, Small RNA-mediated regulation in bacteria: A growing palette of diverse mechanisms, Gene. 656 (2018) 60–72. https://doi.org/10.1016/j.gene.2018.02.068.

[60] M. Giangrossi, G. Prosseda, C.N. Tran, A. Brandi, B. Colonna, M. Falconi, A novel antisense RNA regulates at transcriptional level the virulence gene icsA of Shigella flexneri, Nucleic Acids Res. 38 (2010) 3362–3375. https://doi.org/10.1093/nar/gkq025.

[61] T.B. Updegrove, A. Zhang, G. Storz, Hfq: The flexible RNA matchmaker, Curr. Opin. Microbiol. 30 (2016) 133–138. https://doi.org/10.1016/j.mib.2016.02.003.

[62] A.N. Leistra, N.C. Curtis, L.M. Contreras, Regulatory non-coding sRNAs in bacterial metabolic pathway engineering, Metab. Eng. 52 (2019) 190–214. https://doi.org/10.1016/j.ymben.2018.11.013.

[63] C.M. Schmidt, C.D. Smolke, RNA Switches for Synthetic Biology, (n.d.). https://doi.org/10.1101/cshperspect.a032532.

[64] M. Wachsmuth, S. Findeiß, N. Weissheimer, P.F. Stadler, M. Mö Rl, De novo design of a synthetic riboswitch that regulates transcription termination, (n.d.). https://doi.org/10.1093/nar/gks1330.

[65] R.D. Jenison, S.C. Gill, A. Pardi, B. Polisky, High-resolution molecular discrimination by RNA, Science (80-.). 263 (1994) 1425–1429. https://doi.org/10.1126/science.7510417.

[66] A. Wrist, W. Sun, R.M. Summers, The Theophylline Aptamer: 25 Years as an Important Tool in Cellular Engineering Research, ACS Synth. Biol. 9 (2020) 682–697. https://doi.org/10.1021/acssynbio.9b00475.

[67] J. Chappell, M.K. Takahashi, J.B. Lucks, Creating small transcription activating RNAs, Nat. Chem. Biol. 11 (2015) 1–9. https://doi.org/10.1038/nchembio.1737.

[68] E. Bindewald, K. Afonin, L. Jaeger, B.A. Shapiro, Multistrand RNA secondary structure prediction and nanostructure design including pseudoknots, ACS Nano. 5 (2011) 9542–9551. https://doi.org/10.1021/nn202666w.

[69] A.E. Walter, D.H. Turner, J. Kim, M.H. Lyttle, P. Moller, D.H. Mathews, M. Zukerii, Coaxial stacking of helixes enhances binding of oligoribonucleotides and improves predictions of RNA folding, 1994. https://www.pnas.org/content/pnas/91/20/9218.full.pdf (accessed April 30, 2020).

[70] U. Deuschlel, W. Kammerer, R. Gentz, H. Bujard, Promoters of Escherichia coli: a hierarchy of in vivo strength indicates alternate structures, 5 (1986) 2987–2994.

[71] K.A. Ferguson, Starch-Gel Electrophoresis-Application to the Classification of Pituitary Proteins and Polypeptides, n.d. https://doi.org/10.1016/S0026-0495(64)80018-4.

[72] D. Rodbard, A. Chrambach, Estimation of molecular radius, free mobility, and valence using polyacrylamide gel electrophoresis, Anal. Biochem. 40 (1971) 95–134. https://doi.org/10.1016/0003-2697(71)90086-8.

[73] R.S. Tuma, M.P. Beaudet, X. Jin, L.J. Jones, C.-Y. Cheung, S. Yue, V.L. Singer, Characterization of SYBR Gold Nucleic Acid Gel Stain: A Dye Optimized for Use with 300-nm Ultraviolet Transilluminators, 1999. http://www.idealibrary.com (accessed February 6, 2020).

[74] Y. d. A. Carafa, E. Brody, C. Thermes, Prediction of rho-independent Escherichia coli transcription terminators. A statistical analysis of their RNA stem-loop structures, J. Mol. Biol. (1990). https://doi.org/10.1016/S0022-2836(99)80005-9.

[75] G. Rodrigo, T.E. Landrain, A. Jaramillo, De novo automated design of small RNA circuits for engineering synthetic riboregulation in living cells, Biophys. Comput. Biol. 109 (2012) 15271–15276. https://doi.org/10.1073/pnas.1203831109/-/DCSupplemental.

[76] C.C. Kumar, R.P. Novick, Plasmid pT181 replication is regulated by two countertranscripts (small RNAs/copy control), Biochemistry. 82 (1985) 638–642. https://doi.org/10.1073/pnas.82.3.638.

[77] R.P. Novick, Staphylococcal plasmids and their replication, Annu. Rev. Microbiol. 43:537-563 (1989). https://doi.org/https://doi.org/10.1146/annurev.mi.43.100189.002541.

[78] A. Campbell, SYNCHRONIZATION OF CELL DIVISION, n.d. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC180914/pdf/bactrev00031-0057.pdf (accessed November 25, 2019).

[79] P. Gerhardt, Methods for general and molecular bacteriology, American Society for Microbiology, 1994. https://books.google.fr/books/about/Methods_for_General_and_Molecular_Bacter.html?id=fGYXAQAAIAAJ&redir_esc=y (accessed February 18, 2020).

[80] G. Sezonov, D. Le Joseleau-Petit, R.D.' Ari, Escherichia coli Physiology in Luria-Bertani Broth, J. Bacteriol. 189 (2007) 8746–8749. https://doi.org/10.1128/JB.01368-07.

[81] I. Fishov, A.Z. I, N.B. Grover, On microbial states of growth, Mol. Microbiol. 15 (1995) 789–794.

[82] J.E. Lane-Claypon, MULTIPLICATION OF BACTERIA AND THE INFLUENCE OF TEMPERATURE AND SOME OTHER CONDITIONS THEREON, n.d. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2167167/pdf/jhyg00373-0100.pdf (accessed December 1, 2019).

[83] R.E. Buchanan, Life Phases in a Bacterial Culture, J. Infect. Dis. 23 (1918) 109–125.

[84] J. Monod, The growth of bacterial cultures, Annu. Rev. M. 3 (1949) 371–394. https://doi.org/https://doi.org/10.1146/annurev.mi.03.100149.002103.

[85] L. Wang, D. Fan, W. Chen, E.M. Terentjev, Bacterial growth, detachment and cell size control on polyethylene terephthalate surfaces, Nat. Publ. Gr. (2015). https://doi.org/10.1038/srep15159.

[86] J.K. Wickiser, W.C. Winkler, R.R. Breaker, D.M. Crothers, The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch, Mol. Cell. 18 (2005) 49–60. https://doi.org/10.1016/j.molcel.2005.02.032.

[87] J.K. Wickiser, M.T. Cheah, R.R. Breaker, D.M. Crothers, The kinetics of ligand binding by an adenine-sensing riboswitch, Biochemistry. 44 (2005) 13404–13414. https://doi.org/10.1021/bi051008u.

[88] F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, K. Struhl, Current Protocols in Molecular Biology: Preface, 2010. https://doi.org/10.1002/0471142727.mbprefs66.

[89] F.C. Neidhardt, P.L. Bloch, D.F. Smith1, Culture Medium for Enterobacteria Downloaded from, 1974. http://jb.asm.org/ (accessed November 28, 2019).

[90] H. De Jong, C. Ranquet, D. Ropers, C. Pinel, J. Geiselmann, Experimental and computational validation of models of fluorescent and luminescent reporter genes in bacteria. Experimental and computational validation of models of fluorescent and luminescent reporter genes in bacteria, Syst. Biol. (Stevenage). 4 (2010) 55. https://doi.org/10.1186/1752-0509-4-55ï.

[91] I. Mihalcescu, M. Van-Melle Gateau, B. Chelli, C. Pinel, J.L. Ravanat, Green autofluorescence, a double edged monitoring tool for bacterial growth and activity in micro-plates, Phys. Biol. 12 (2015). https://doi.org/10.1088/1478-3975/12/6/066016.

[92] B.C. Matlock, R.W. Beringer, D.L. Ash, M.W. Allen, A.F. Page, Analyzing Differences in Bacterial Optical Density Measurements between Spectrophotometers, n.d. www.thermoscientific.com/nanodrop (accessed November 29, 2019).

[93] Y. Martin, M. Page, C. Blanchet, H. De Jong, SOFTWARE Open Access WellInverter: a web application for the analysis of fluorescent reporter gene data, (n.d.). https://doi.org/10.1186/s12859-019-2920-4.

[94] V. Zulkower, M. Page, D. Ropers, J. Geiselmann, H. De Jong, Robust reconstruction of gene expression profiles from reporter gene data using linear inversion, (2015). https://doi.org/10.1093/bioinformatics/btv246.

[95] R. Of, G. Expression, B.W. Epstein, J.R. Beckwith, REGULATION OF GENE EXPRESSION WOLFGANG EpSTEIN AND JONATHAN R. BECKWITH, (1968) 411–436.

[96] R. Lutz, H. Bujard, Independent and tight regulation of transcriptional units in escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements, Nucleic Acids Res. 25 (1997) 1203–1210. https://doi.org/10.1093/nar/25.6.1203.

[97] T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K.A. Datsenko, M. Tomita, B.L. Wanner, H. Mori, Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: The Keio collection, Mol. Syst. Biol. 2 (2006). https://doi.org/10.1038/msb4100050.

[98] C. Grenier, F. Matteau, D. Baby, V. Rodrigue, Complete genome sequence of Escherichia coli BW25113, Genome Announc. 2 (2014) 1038–1052. https://doi.org/10.1128/genomeA.01038-14.

[99] T. V Chepurnykh, A.S. Shcheglov, V. Vladislav, V.Z. Pletnev, K.L. Hazelwood, P.M. Roche, A.G. Zaraisky, M.W. Davidson, M. Dmitriy, Far-red fluorescent tags protein imaging in living tissues, 418 (2010) 567–574. https://doi.org/10.1042/BJ20081949.Far-red.

[100] G. Cambray, J.C. Guimaraes, V.K. Mutalik, C. Lam, Q.A. Mai, T. Thimmaiah, J.M. Carothers, A.P. Arkin, D. Endy, Measurement and modeling of intrinsic transcription terminators, Nucleic Acids Res. 41 (2013) 5139–5148. https://doi.org/10.1093/nar/gkt163.

[101] D. Stueber, H. Bujard, H. Bujard, Transcription from efficient promoters can interfere with plasmid replication and diminish expression of plasmid specified genes, 1982. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC553223/pdf/emboj00303-0106.pdf (accessed January 15, 2020).

[102] W. Kammerer, U. Deuschle, R. Gentz1, H. Bujard, Functional dissection of Escherichia

coli promoters: information in the transcribed region is involved in late steps of the overall process, 1986. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1167252/pdf/emboj00174-0250.pdf (accessed January 15, 2020).

[103] B.R. Glick, Metabolic load and heterologous gene expression, Biotechnol. Adv. 13 (1995) 247–261. https://doi.org/10.1016/0734-9750(95)00004-A.

[104] W. Ryan, S.J. Parulekar, B.C. Stark, Expression of β-lactamase by recombinant Escherichia coli strains containing plasmids of different sizes—effects of pH, phosphate, and dissolved oxygen, Biotechnol. Bioeng. 34 (1989) 309–319. https://doi.org/10.1002/bit.260340306.

[105] S. Birnbaum, J.E. Bailey, Plasmid presence changes the relative levels of many host cell proteins and ribosome components in recombinant Escherichia coli, Biotechnol. Bioeng. 37 (1991) 736–745. https://doi.org/10.1002/bit.260370808.

[106] A. Maser, K. Peebo, R. Nahku, Avoiding amino acid depletion in a complex medium results in improved Escherichia coli BW25113 growth, (n.d.). https://doi.org/10.1099/mic.0.000742.

[107] J. Zhang, J. Reddy, B. Buckland, R. Greasham, Toward consistent and productive complex media for industrial fermentations: Studies on yeast extract for a recombinant yeast fermentation process, Biotechnol. Bioeng. 82 (2003) 640–652. https://doi.org/10.1002/bit.10608.

[108] F.C. Neidhardt, P.L. Bloch, D.F. Smith, Culture medium for enterobacteria, J. Bacteriol. 119 (1974) 736–747.

[109] S. Lin-chao, W. Chen, High copy number of the pUC piasmid results from a Rom / Rop-suppressible point mutation In RNA II, Mol. Microbiol. 6 (1992) 3385–3393.

[110] H. Masukata, J. ichi Tomizawa, Control of primer formation for ColE1 plasmid replication: Conformational change of the primer transcript, Cell. 44 (1986) 125–136. https://doi.org/10.1016/0092-8674(86)90491-5.

[111] D.K. Hawley, W.R. Mcclure, Compilation and analysis of Escherichia coli promoter DNA sequences, 1983. https://academic.oup.com/nar/article-abstract/11/8/2237/2385155 (accessed December 11, 2019).

[112] C.B. Harley, R.P. Reynolds, Analysis of E. coli promoter sequences, 1987. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC340638/pdf/nar00249-0457.pdf (accessed September 25, 2019).

[113] S. Busby', R.H. Ebrightt, Promoter Structure, Promoter Recognition, and Transcription Activation in Prokatyotes, 1994.

[114] J.S. McDowall, B.J. Murphy, M. Haumann, T. Palmer, F.A. Armstrong, F. Sargent, Bacterial formate hydrogenlyase complex, Proc. Natl. Acad. Sci. U. S. A. 111 (2014) E3948–E3956. https://doi.org/10.1073/pnas.1407927111.

[115] B. Vivijs, L.U. Haberbeck, V.B. Mfortaw Mbong, K. Bernaerts, A.H. Geeraerd, A. Aertsen, C.W. Michiels, Formate hydrogen lyase mediates stationary-phase deacidification and increases survival during sugar fermentation in acetoin-producing enterobacteria, Front. Microbiol. 6 (2015) 1–11. https://doi.org/10.3389/fmicb.2015.00150.

[116] P.N. Borer, B. Dengler, I. Tinoco, O.C. Uhlenbeck, Stability of ribonucleic acid double-stranded helices, J. Mol. Biol. 86 (1974) 843–853. https://doi.org/10.1016/0022-2836(74)90357-X.

[117] Yutaka EguchF, ANTISENSE RNA, (1991).

[118] S. Brantl, E.G.H. Wagner, Antisense RNA-mediated transcriptional attenuation: An in

vitro study of plasmid pT181, Mol. Microbiol. 35 (2000) 1469–1482. https://doi.org/10.1046/j.1365-2958.2000.01813.x.

[119] J.G. Belascop, F. Higginsb, Mechanisms of mRNA decay in bacteria : a perspective, 72 (1988) 15–23.

[120] J.M. Andrade, C.M. Arraiano, PNPase is a key player in the regulation of small RNAs that control the expression of outer membrane proteins, Rna. 14 (2008) 543–551. https://doi.org/10.1261/rna.683308.

[121] Y. d. A. Carafa, E. Brody, C. Thermes, Prediction of rho-independent Escherichia coli transcription terminators. A statistical analysis of their RNA stem-loop structures, J. Mol. Biol. 216 (1990) 835–858. https://doi.org/10.1016/S0022-2836(99)80005-9.

[122] R. Reynolds, M.J. Chamberlin, Parameters affecting transcription termination by Escherichia coli RNA. II. Construction and analysis of hybrid terminators., J. Mol. Biol. 224 (1992) 53–63. https://doi.org/10.1016/0022-2836(92)90575-5.

[123] I. Gusarov, E. Nudler, The mechanism of intrinsic transcription termination, Mol. Cell. 3 (1999) 495–504. https://doi.org/10.1016/S1097-2765(00)80477-3.

[124] Y.-J. Chen, P. Liu, A.A.K. Nielsen, J.A.N. Brophy, K. Clancy, T. Peterson, C.A. Voigt, Characterization of 582 natural and synthetic terminators and quantification of their design constraints., Nat. Methods. 10 (2013) 659–64. https://doi.org/10.1038/nmeth.2515.

[125] B. Yan, M. Boitano, T.A. Clark, L. Ettwiller, SMRT-Cappable-seq reveals complex operon variants in bacteria, Nat. Commun. 9 (2018). https://doi.org/10.1038/s41467-018-05997-6.

[126] P.R. Costa, M.L. Acencio, N. Lemke, Cooperative RNA Polymerase Molecules Behavior on a Stochastic Sequence-Dependent Model for Transcription Elongation, PLoS One. 8 (2013). https://doi.org/10.1371/journal.pone.0057328.

[127] M.A. Jacquet, C. Reiss, In vivo control of promoter and terminator efficiencies at a distance, Mol. Microbiol. 6 (1992) 1681–1691. https://doi.org/10.1111/j.1365-2958.1992.tb00893.x.

[128] A.P. Quendera, A.F. Seixas, R.F. dos Santos, I. Santos, J.P.N. Silva, C.M. Arraiano, J.M. Andrade, RNA-Binding Proteins Driving the Regulatory Activity of Small Non-coding RNAs in Bacteria, Front. Mol. Biosci. 7 (2020) 1–9. https://doi.org/10.3389/fmolb.2020.00078.

[129] J. Chen, T. Morita, S. Gottesman, Regulation of transcription termination of small RNAs and by small RNAs: Molecular mechanisms and biological functions, Front. Cell. Infect. Microbiol. 9 (2019) 1–9. https://doi.org/10.3389/fcimb.2019.00201.

[130] S.A. Gorski, J. Vogel, J.A. Doudna, RNA-based recognition and targeting: sowing the seeds of specificity, Nat. Rev. Mol. Cell Biol. 18 (2017) 215–228. https://doi.org/10.1038/nrm.2016.174.

[131] W. Kammerer, U. Deuschle, R. Gentz, H. Bujard, Functional dissection of Escherichia coli promoters: information in the transcribed region is involved in late steps of the overall process., EMBO J. 5 (1986) 2995–3000. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1167252/%5Cnhttp://www.ncbi.nlm.nih.gov/pmc/articles/PMC1167252/pdf/emboj00174-0250.pdf.

[132] T.A. Carrier, J.D. Keasling, Library of synthetic 5' secondary structures to manipulate mRNA stability in Escherichia coli, Biotechnol. Prog. 15 (1999) 58–64. https://doi.org/10.1021/bp9801143.

Abstract

This thesis work is a contribution to the general field of synthetic biology of RNA and aims at constructing and testing the functioning of RNA switches controlled by small anti-sense RNAs. The corresponding topic of fundamental research addresses the relationship between sequence, folding and function of an RNA molecule.

We have devised RNA switches that adopt two alternative conformations. In one conformation, the RNA forms a transcriptional terminator, stopping the expression of a downstream gene. The alternative conformation undoes the terminator structure and allows read-through to the downstream DNA. The choice of which conformation the RNA molecule adopts depends on the presence or absence of a small regulatory anti-sense RNA. We have validated the functioning of our repressible and inducible switches *in vitro*. In order to validate the functioning of the switches *in vivo*, we have constructed and optimized a genetic testing system in the bacterium *Escherichia coli*. The performance of the switch is measured by the expression of a fluorescent reporter gene cloned downstream of the switch. We have thus been able to provide a proof of concept of the design, construction and *in vivo* functioning of activatable and repressible RNA switches.

Our *in vitro* and *in vivo* experiments have allowed us to discover new principles of the design of synthetic RNA switches. We have also been able to optimize the expression conditions of the switches and the measurements methods of the *in vivo* experiments. The dynamic range of our RNA switches needs to be improved and we have discovered several promising avenues to accomplish this goal. More generally, our results form the basis of a more systematic exploration of the relationships between sequence, folding and functioning of RNA molecules.