

## SYNTHESE EN FRANCAIS

UNIVERSITE PARIS-SUD

ÉCOLE DOCTORALE : GÈNES GÉNOMES, CELLULES (ED 426)

Laboratoire : “Régulation génique chez Salmonella et ses phages”

Centre de Génétique Moléculaire du CNRS, UPR3404

**DISCIPLINE** : GÉNÉTIQUE MOLÉCULAIRE

THÈSE DE DOCTORAT

soutenue le 19/09/2013

par

Qi YANG

Régulation du gène *yifK* par le petit ARN  
multi-cible GcvB chez *Salmonella*

**Directeur de thèse :**

Lionello BOSSI

Directeur de Recherche, CNRS

**Composition du jury :**

*Président du jury :*

Philippe BOULOC

Directeur de Recherche, CNRS, UMR8621, Orsay

*Rapporteurs :*

Philippe BOULOC

Directeur de Recherche, CNRS, UMR8621, Orsay

Francis REPOILA

Directeur de Recherche, INRA, Jouy en Josas

*Examineurs :*

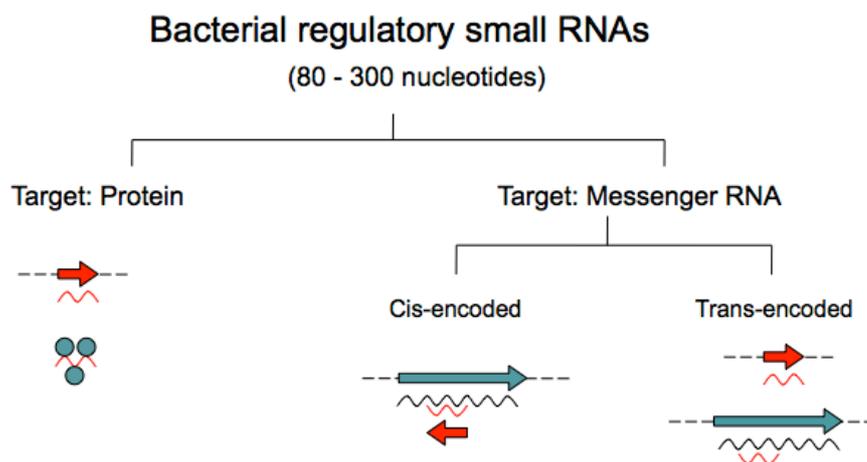
Claudio GUALERZI

Professeur, Université de Camerino, Italie

# I. Introduction

L'étude de modèles bactériens a constitué une étape clé pour la compréhension des mécanismes moléculaires à la base du fonctionnement de la cellule. L'un de ces modèles utilisé avec succès est la bactérie *Salmonella enterica* (précédemment appelée *Salmonella typhimurium*). C'est en travaillant avec cet organisme que Norton Zinder et Joshua Lederberg, dans les années 50, découvrirent le phénomène de la transduction – un des premiers exemples de transfert génétique horizontale (Zinder & Lederberg, 1952) – et que Milislav Demerec développa les premières techniques de cartographie génétique. Plus récemment, *Salmonella* s'est affirmé comme modèle de choix pour l'étude des bases génétiques de la pathogénie bactérienne (Ohl & Miller, 2001). Les salmonelles sont des  $\gamma$ -protéobactéries de la famille des *Enterobacteriaceae* au même titre qu'*Escherichia coli*. Ce sont des bactéries à Gram négatif, en forme de bâtonnets des dimensions 1  $\mu\text{m}$  sur 2  $\mu\text{m}$ , mésophiles, chimio-organotrophes et aérobies facultatives.

Comme la plupart des bactéries entériques, *Salmonella* peut endurer des variations très marquées dans ses conditions de croissance, par exemple lors du passage du milieu intestinale au milieu intracellulaire ou au milieu naturel. L'adaptabilité des bactéries à ces changements environnementaux repose sur leur capacité de réguler l'expression de leur génome en fonction de la composition du milieu. La régulation de l'expression des gènes peut se faire à deux niveaux : le niveau transcriptionnel qui va agir sur la synthèse de l'ARNm avec notamment l'utilisation de facteurs sigma alternatifs ou des protéines régulatrices (activateurs ou répresseurs) et le niveau post-transcriptionnel impliquant la capacité de l'ARNm à être traduit ou sa susceptibilité à la dégradation. Ce deuxième type de régulation fait souvent appel à l'activité de petits ARN non codants.



### **Figure 1 : Présentation des 2 classes principales d'ARNs régulateurs.**

Les petits ARN qui interagissent avec des protéines et ceux qui agissent par appariement de bases.

Parmi ces derniers, deux sous-groupes:

a) Les petits ARN codés en «Cis» (en rouge) sont synthétisés à partir du brin complémentaire à celui codant l'ARNm cible (en vert).

b) Les petits ARN codés en «Trans» (en rouge) sont synthétisés à partir d'une région chromosomique différente de leurs ARNm cibles (en vert).

## **I.1 Les ARN non codants**

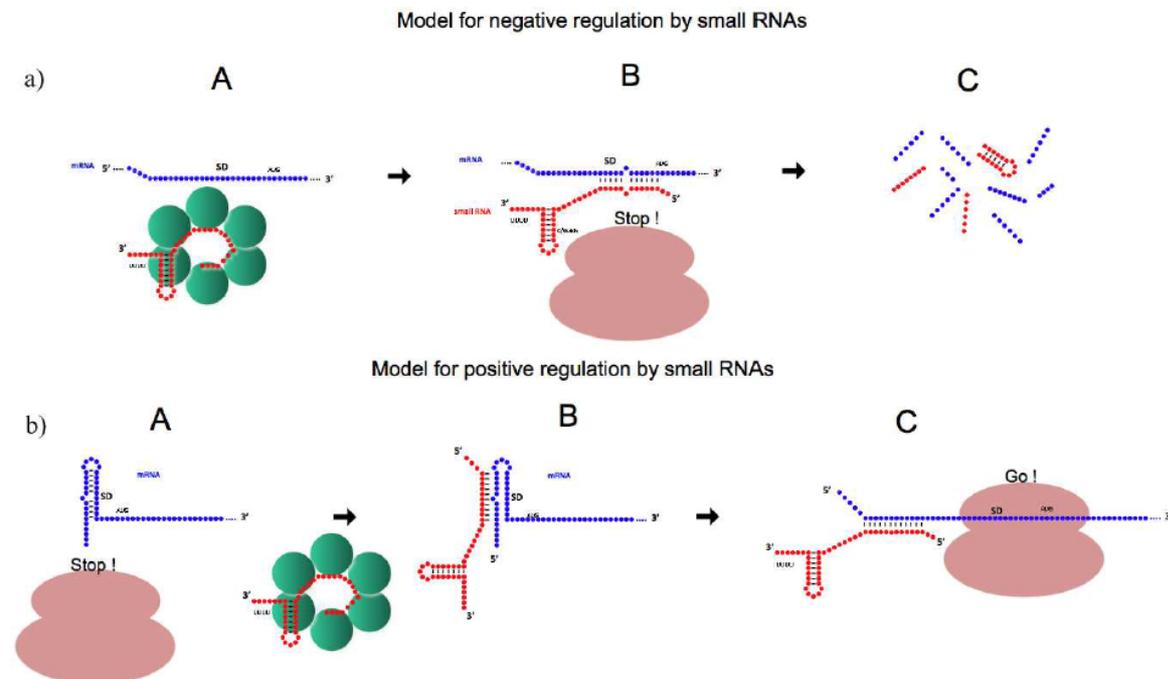
Les ARN non codants (ARNnc) sont présents aussi bien chez les eucaryotes que chez les procaryotes. Certains de ces ARNnc sont extrêmement bien conservés sur l'arbre du vivant, c'est le cas notamment des ARN ribosomiaux ou des ARN de transferts alors que d'autres ne sont présents que chez certaines espèces. Les fonctions de ces ARN ne sont pas toutes connues, mais ils jouent un rôle de plus en plus prépondérant dans la cellule. Ils sont des acteurs dans la régulation des gènes, l'épissage des ARNm ou encore dans la défense antivirale. La régulation de l'expression génique par des ARNnc est extrêmement variable d'un organisme à l'autre. Chez les eucaryotes, les microARNs (miARN), qui font entre 21 et 24 paires de bases, vont se lier avec les protéines de la famille Argonaute. Ce complexe va interagir avec un ARNm cible complémentaire au miARN et va entraîner la dégradation ou l'inhibition de la traduction de ce dernier. Chez les bactéries, la régulation post-transcriptionnelle se fait en partie grâce à de petits ARN régulateurs.

## **I.2 Les petits ARN régulateurs bactériens**

Les petits ARN régulateurs bactériens sont des molécules d'ARN simple brin, généralement non-codantes et de taille variable, entre 50 et 300 nucléotides. On peut distinguer deux familles : ceux qui fonctionnent en interagissant avec des protéines et ceux qui interagissent avec d'autres molécules d'ARN par un mécanisme d'appariement de bases (Figure 1).

Dans la première des deux familles, les petits ARN qui séquestrent des protéines capables de se lier à l'ARN (**RNA-Binding Proteins**, RBP), modulent ainsi l'expression des gènes au niveau post-transcriptionnel (Marzi & Romby, 2012): en absence des petits ARN, les RBP s'associent à des structures spécifiques sur leurs ARNm cibles, souvent proches du site de liaison du ribosome (**Ribosome Binding Site**, RBS) et bloquent la traduction. Quand les petits ARN capables de lier les RBP sont transcrits, les RBP correspondantes sont titrés à l'aide des motifs structuraux du petit ARN, libérant le site de fixation du ribosome et permettant l'initiation de la traduction.

En général, afin d'améliorer leur capacité de titration, les petits ARN possèdent des multiples motifs répétés, qui miment les motifs reconnus par les RBP sur leurs ARNm cibles.



**Figure 2 : Deux modèles de régulation par les petits ARNs.**

L'ARNm est représenté en bleu; le petit ARN est représenté en rouge ; la protéine Hfq est représentée par des sphères vertes et le ribosome par des ovales rouges ;

a) Le petit ARN par l'intermédiaire d'Hfq, s'apparie à l'ARNm au niveau de la séquence Shine-Dalgarno (SD), ce qui va empêcher le ribosome de se fixer à l'ARNm et va entraîner la dégradation de ce dernier ;

b) Le ribosome ne peut accéder à la séquence SD présent sur l'ARNm à cause de la structure secondaire présent sur ce dernier. La fixation du petit ARN sur l'ARNm déstabilise la structure secondaire présente sur l'ARNm ce qui va permettre la traduction de ce dernier.

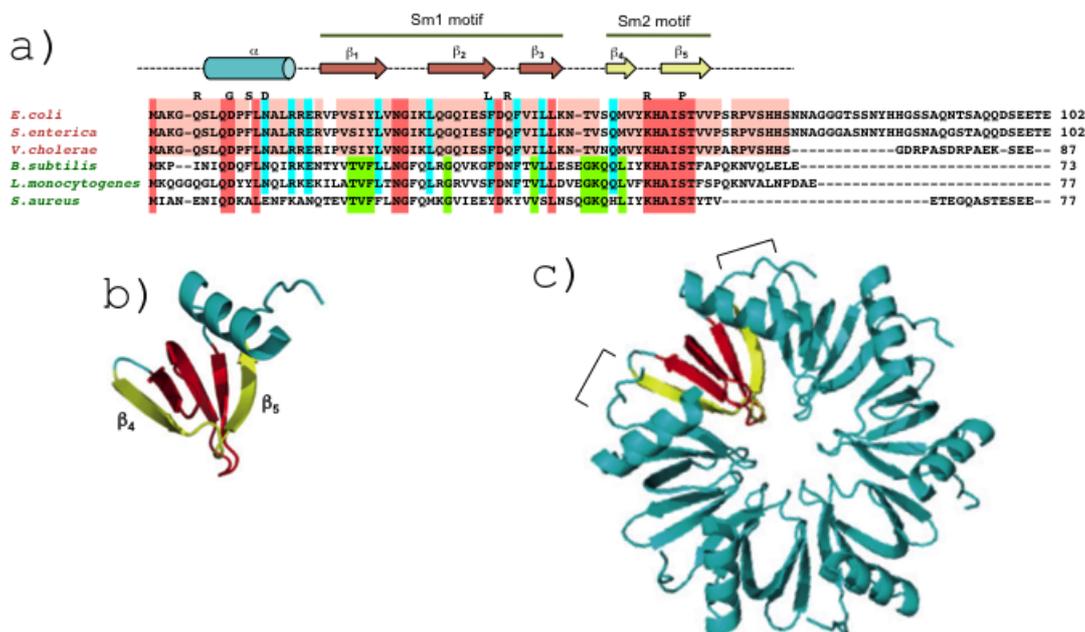
Chez *Escherichia coli*, le système du "Carbon storage regulator" Csr, contrôle plusieurs messagers cibles à travers le répresseur traductionnel CsrA et deux petits ARN CsrB et CsrC (Romeo, 1998) Ces deux ARN possèdent 22 et 13 motifs GGA respectivement, situés au niveau des boucles de structures en epingle à cheveux courtes (Babitzke & Romeo, 2007). Lorsque CsrB/C sont synthétisés, ils séquestrent la protéine CsrA à l'aide de leurs multiples motifs GGA, entraînant la libération du RBS des ARNm. En l'absence des ARN CsrB/C, CsrA semble promouvoir la dégradation rapide de son ARNm cible à travers l'inhibition de la traduction (Babitzke & Romeo, 2007). Cela indique que le modulateur traductionnel peut aussi modifier la stabilité de l'ARNm.

Les petits ARN de la deuxième famille, qui agissent par appariement de bases, forment deux sous-groupes selon la position de leur gène par rapport au locus de la cible: ceux codés en *cis*, et ceux codés en *trans*. Les gènes des petits ARNs en *cis* ont été découverts dans un premier temps sur les éléments extra-chromosomiques et plus récemment sur le chromosome bactérien (Brantl, 2007). Cette classe de régulateurs se caractérise par leur complémentarité complète à l'ARNm cible, car ils sont codés sur le brin d'ADN opposé à celui spécifiant l'ARNm cible. Les petits ARNs en *trans* sont synthétisés majoritairement à partir de régions intergéniques. Ils n'ont qu'une complémentarité partielle à l'ARNm cible, généralement comprise entre 7 et 20 paires de bases et à la différence des petits ARNs codés en *cis*, ils peuvent avoir plusieurs ARNm cibles et plusieurs effets sur ces ARNm.

Dans un bon nombre de cas, l'expression des petits ARNs est régulée en réponse à un stress environnemental ou à des variations des conditions de croissance, telle qu'une limitation en fer, un stress oxydatif, une faible concentration en magnésium, etc.

Dans la plupart des cas, la régulation est de type négatif. L'appariement entre le petit ARN et l'ARNm a pour effet de limiter l'accessibilité du ribosome pour se fixer sur la séquence Shine-Dalgarno (SD) et/ou de reconnaître le codon d'initiation (Figure 2a). Il en résulte une inhibition de la traduction qui est généralement accompagnée d'une augmentation du taux de dégradation de l'ARNm par des ribonucléases. Plus rarement, la régulation peut être positive, comme pour l'ARNm *rpoS* (un facteur Sigma alternatif), dont le site de fixation du ribosome est séquestré dans une structure secondaire en forme de tige-boucle. L'appariement avec le petit ARN DsrA déstabilise cette structure et libère ainsi la séquence SD qui devient par la suite accessible aux ribosomes (Figure 2b). Chez *Escherichia coli* et *Salmonella enterica*, la fonction de la plupart des petits ARNs codés en « *trans* » dépend étroitement de la protéine chaperonne d'ARN Hfq.

Figure 3 : Topologie de la protéine Hfq (d'après(Vogel & Luisi, 2011))



a) Topologie en suivant un alignement des séquences de la protéine Hfq de différentes espèces bactérienne.

b) Structure tridimensionnelle du monomère d'Hfq d'*E. coli* avec en rouge le motif Sm1 composé des feuillets  $\beta_1\beta_2\beta_3$  et en jaune le motif Sm2 composé des feuillets  $\beta_4\beta_5$ .

c) Structure tridimensionnelle de l'hexamère d'Hfq d'*E. coli*.

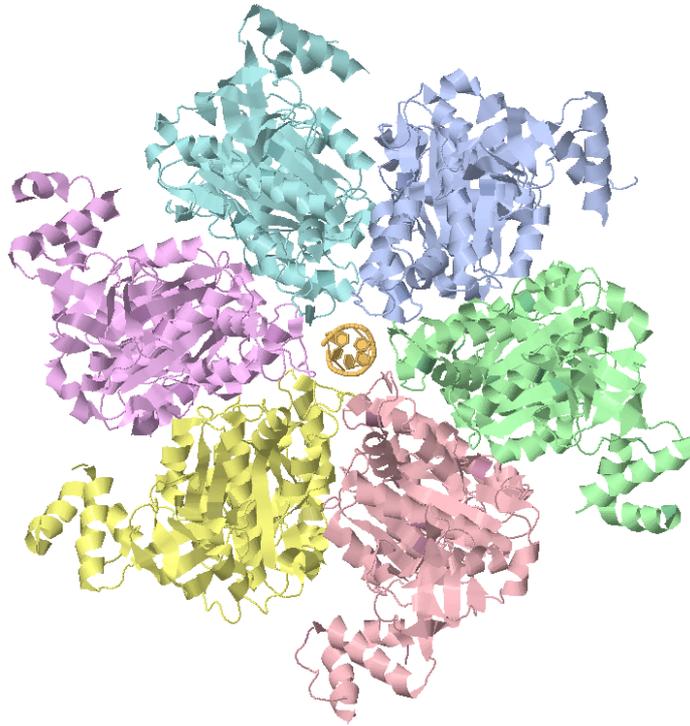
### I.3 La protéine chaperonne d'ARN Hfq

Hfq, est une protéine chaperonne des ARN, membre de la famille des Sm-like (LSm), un groupe de protéines présente dans toutes les domaines du vivant. Hfq est une protéine abondante, cytoplasmique constitué par 70 à 102 aa selon l'espèce bactérienne, composé de 6 sous unités qui s'assemblent pour former un homo-hexamère (Figure 3c). Elle capable de se fixer aux ARNm ainsi qu'aux petits ARN. Elle joue un rôle clé dans la stabilisation des petits ARN ainsi que dans l'appariement entre les petits ARN et l'ARNm.

La protéine Hfq fut identifiée pour la première fois en 1968 chez *E. coli* infectée par le bactériophage Q $\beta$  et son rôle proposé était d'initier la synthèse du brin positif de l'ARN du phage Q $\beta$  en facilitant la reconnaissance du brin négatif par la réplicase du phage (Hfq – Host Factor required for phage Q $\beta$ ). De nombreuses études ultérieures ont montré que chez *E. coli*, la protéine Hfq joue un rôle très important dans la physiologie de la bactérie; la délétion du

gène *hfq* entraîne des effets pléiotropiques et un défaut de croissance (Tsui *et al.*, 1994). Le premier rôle cellulaire reporté pour Hfq a été sa participation à la régulation du gène *rpoS* (codant pour le facteur  $\sigma^S$ ) qui intervient dans la régulation de nombreux gènes de réponses aux stress chez la bactérie, notamment en phase stationnaire. Cela est mieux compris maintenant en vue de la participation des petits ARNs à la régulation de *rpoS* (Repoila *et al.*, 2003).

Des études de cristallographie ont permis de déterminer la structure des homo-hexamères d'Hfq de différentes espèces bactériennes, notamment *E.coli*, *S.aureus*, et *P.aeruginosa*. Ces travaux ont montré qu'Hfq, comme d'autres protéines Sm/LSm forme un anneau environ 70Å de diamètre, d'une épaisseur de 25Å et d'un pore interne de 8 à 12Å (Figure 3) (Sauter *et al.*, 2003). Des expériences de liaison d'Hfq à l'ARN montrent chez *S. aureus* qu'un heptamère d'ARN de séquences A(U)<sub>5</sub>G se lie à la protéine au niveau du centre de l'anneau, sur la face proximale (électropositive) et que chez *E. coli*, une séquence poly(A) de 18 à 27 nucléotides se lie à la protéine Hfq sur sa face distale ainsi que sur son flanc (Brennan & Link, 2007). La structure de la protéine Hfq est caractérisée par une hélice  $\alpha$  en N-terminal suivie par 5 feuillets béta, dont la topologie est  $\beta 5\alpha 1\beta 1\beta 2\beta 3\beta 4$  (Figure 3a). Son appartenance à la superfamille des SM-like protéines est illustré par le fait que Hfq possède notamment deux motifs très conservés de cette famille : les motifs SM1 et SM2 : SM1 englobe les 3 premiers feuillets  $\beta$  ( $\beta 1\beta 2\beta 3$ ), alors que le motif SM2 est composé du feuillet  $\beta 4$  et  $\beta 5$  (Figure 3b).



**Figure 4 : Structure tridimensionnelle de l'homo-hexamère Rho.**

Les 6 sous-unités de la protéine Rho sont représenté chacun d'une couleur différente et au centre en orange est représenté un ARN(Rabhi *et al.*, 2011).

En plus de sa participation à la régulation post transcriptionnelle par les petits ARN, Hfq joue également un rôle sur la stabilité des ARNm, par sa capacité de recruter des RNases spécifiques au niveau du duplex petit ARN:ARNm. Par ailleurs, Hfq peut se lier aux extrémités poly(A) des ARNm, et entraîner une plus forte poly adénylation par la poly(A) polymérase I et conduire ainsi l'ARNm hyperpoly(A) à être dégradé par les RNases II et III. (Hajnsdorf & Regnier, 2000).

Enfin, des études récentes proposent un rôle nouveau pour Hfq, au niveau de la terminaison de la transcription. En effet, des travaux biochimiques *in vitro* ont montré que Hfq était capable d'agir comme anti-terminateur de la transcription par son interaction directe avec le facteur de terminaison de la transcription Rho (Rabhi et al., 2011).

## II . RÉSULTATS

### II.1 . Caractérisation d'*yifK* , un locus régulé par Hfq chez *Salmonella*.

Il y a quelques années, notre laboratoire a développé un crible génétique pour identifier les gènes régulés par la protéine chaperon d'ARNs Hfq chez *Salmonella* (Figueroa-Bossi et al., 2006). Cette approche visait à identifier de nouveaux gènes régulés par petits ARN régulateurs Hfq-dépendants. La stratégie qui a été mise au point pour réaliser ce crible employée un élément transposable dérivé du phage Mu, MudK. Ce transposon contient l'opéron lac dont le gène *lacZ* est sans promoteur et sans codon AUG initiateur. L'insertion de MudK, de façon aléatoire, dans le chromosome bactérien peut donner lieu à fusion traductionnelle (Hughes & Roth, 1988). Ces événements de transpositions sont sélectionnés positivement grâce à la présence d'un marqueur de sélection, gène de résistance à la kanamycine ( $Kan^R$ ), dans le transposon. Cette expérience de transposition a été réalisée dans une souche dont la copie chromosomique sauvage d'Hfq a été délétée et où une seconde copie de ce même gène a été placée sous le contrôle du promoteur de l'opéron arabinose. De ce fait, la souche utilisée est un mutant conditionnel d'Hfq, elle est Hfq- en absence d'arabinose et Hfq+ en présence de ce dernier. Les colonies ont été sélectionnées sur boîtes LB additionnée de kanamycine puis ont été répliquées sur milieu indicateur (MacConkey-lactose ou LB-Xgal) contenant ou non de l'arabinose. Les colonies dont la couleur diffère en fonction de la présence ou non de l'arabinose ont été caractérisées ultérieurement. Ces travaux ont conduit à l'identification de 19 fusions lac ayant une expression accrue en absence d'arabinose et 4 fusions dont l'expression diminue dans ces mêmes conditions (Figueroa-Bossi et al., 2006). Les membres du premier groupe ont été classés comme des candidats à une régulation négative par des petits ARN Hfq-dépendants alors que le second groupe a été classé comme candidat à une régulation positive par des petits ARN Hfq-dépendants. En fait, trois des quatre membres du second groupe sont des gènes transcrits à l'aide du facteur sigma alternatif  $\sigma^S$ , connus pour être activés par un petit ARN régulateur Hfq-dépendant (Repoila et al., 2003).

Quand j'ai rejoint le laboratoire pour préparer ma thèse, la participation des petits ARN dans la régulation de la plupart des loci identifiés ci-dessus n'était pas encore connue. Mon travail a donc été de déterminer l'implication de petits ARN dans la régulation de certains de ces loci. Un des loci étudiés a été celui d'*yifK*, gène codant pour une protéine très conservée

annoté comme « transporteur putatif d'acides aminés » dans les bases de données de séquences d'ADN. Deux insertions MudK en générant des fusions *lacZ* en phase avec le gène *yifK* ont été isolées au laboratoire lors de la précédente étude, l'une à la position 143 et l'autre à la position 284 de la séquence codante. L'expression de ces fusions est augmentée d'un facteur 5 à 7 dans une souche délétée pour *hfq*. La fusion lac à la position 143, exprimant le plus fortement l'activité  $\beta$ -galactosidase, a été choisie pour la plupart des expériences décrites dans cette thèse et afin de faciliter la compréhension, cette construction sera appelée *yifK-lacZY* même s'il faut se rappeler la présence de la cassette kanR en aval de la fusion. Il est également important de noter que la présence de l'ensemble de l'opéron lac dans le transposon MudK va permettre de suivre l'ensemble du parcours de l'utilisation du lactose – et pas seulement le produit du gène *lacZ*. Ceci aura son importance dans de nombreuses expériences décrites ici.

### II.1.1 . *yifK* est réprimée par GcvB

Chez *Salmonella* et *E. coli*, le petit ARN GcvB est connu pour réprimer la traduction de nombreux petits ARN codant pour différents transporteurs d'acides aminés et de peptides. Le gène *yifK* appartenant à cette classe (Voir ci-dessus), GcvB semblait donc un candidat probable pour être un régulateur d'*yifK*. J'ai pu montrer que la délétion de *gcvB* entraîne une augmentation d'un facteur 3 à 5 de l'expression de la fusion *yifK-lacZY*. La présence d'une séquence complémentaire à GcvB dans la région immédiatement en amont de la séquence de Shine-Dalgarno (SD) d'*yifK* suggère que GcvB réprime la traduction d'*yifK* par appariement de bases. GcvB utilise une séquence riche en GU appelée R1 pour s'apparier avec la plupart de ces cibles. Et cela semble être également le cas avec *yifK*, même si la partie de la région R1 impliquée dans l'appariement avec *yifK* diffère de celle utilisée par d'autres cibles.

### II.1.2 . GcvB régule l'expression d'*yifK*

Nous avons ensuite réalisé de nombreuses mutations par la technique de mutagenèse dirigée. Afin de vérifier nos suppositions, nous avons changé un segment de quatre bases dans la région d'appariement : la séquence UGUU du petit ARN GcvB a été changée en ACAC ; le changement réciproque a également été effectué sur *yifK* (ACAC en UGUG). Les effets de

ces modifications sur l'expression de la fusion *yifK-lacZY* ont été analysés en mesurant l'activité  $\beta$ -galactosidase(Figure 5).

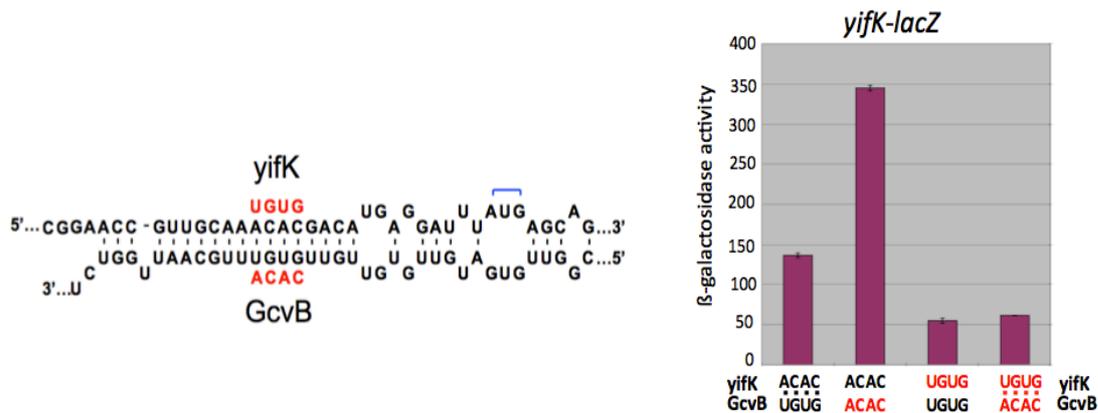


Figure 5. Mutant des régions d'appariement entre *yifK* et *gcvB*.

### II.1.3 . GcvB inhibe la traduction *yifK* lors de son appariement au niveau d'un élément activateur

De façon inattendue, lorsque la séquence ACAC est mutée dans la séquence de l'ARNm d'*yifK*, l'expression de la fusion *yifK-lacZY* n'est pas augmenté comme attendu mais elle est au contraire diminué (Figure 5). Afin d'essayer de clarifier ces observations, nous avons mutagénéisé séparément des parties de la région d'intérêt. Comme le montre la figure 6A, la mutation d'un triplé AAA en UGU au milieu de la séquence cible d'*yifK* ou en réalisant la mutation inverse sur GcvB (UGU en AAA), permet d'inhiber la répression de la fusion *yifK-lacZY*. Par contre, l'utilisation de la combinaison d'allèles compensatoire permet de rétablir la répression. Ainsi, cette partie de la séquence cible se comporte comme ce qui était prévu, et le phénotype du mutant contenant les allèles compensatoire suggère fortement que GcvB réprime *yifK* par appariement de base. Lors des expériences de Toeprint in vitro, j'ai pu montrer que l'appariement de GcvB sur *yifK* inhibé la fixation de la sous-unité 30S du ribosome de manière spécifique et dépendant de la dose (données non présentées ). Ceci fournit un soutien supplémentaire à cette conclusion.

Cependant, lorsque j'ai muté le doublet CA en UC en 3' du triplet AAA, le résultat qui a été obtenu ne correspond pas au schéma classique : comme lors de la mutation du quadruplet ACACA présenté auparavant, l'expression de la fusion *yifK-lacZ* s'est trouvé diminué au lieu d'augmenter, devenant en plus insensible à un mutant de GcvB portant le changement compensatoire (Figure 6B), ou à la délétion de GcvB ou d'Hfq (Figure 6C). L'interprétation

de ces résultats nous a suggérer que la mutation du doublet CA en UC affecté la traduction et que, dans ces conditions, l'action de GcvB sur la régulation d'*yifK* n'était plus limitant pour l'expression d'*yifK*. Les effets de la mutation sur l'efficacité de traduction ont été étudiés in vitro en utilisant un système reconstitué. Les résultats de ces expériences ont pu montrer qu'une fusion de la région 5'UTR d'*yifK* avec le gène codant la protéine Cat taggué d'un épitope s'accumuler de façon significative par rapport à une construction équivalente portant la mutation du doublet CA en UC (Figure 7). Ces différentes données confirment que le doublet CA stimule la traduction et suggère que l'efficacité de GcvB dans la régulation d'*yifK* reflète le ciblage de cet élément activateur de la traduction.

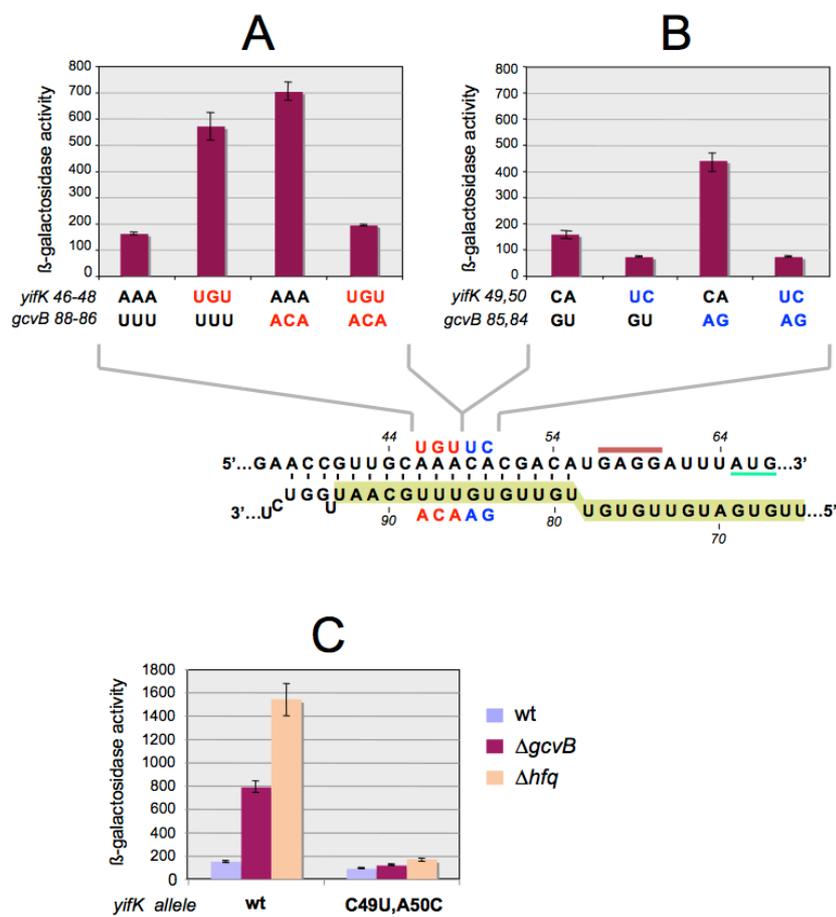


Figure 6. Effets différentiels de mutations de la séquence d'appariement d'*yifK* sur l'expression *yifK-lacZ* et sa réponse aux mutants compensatoires dans GcvB.

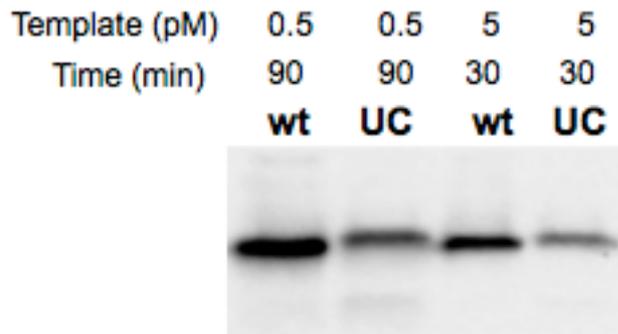
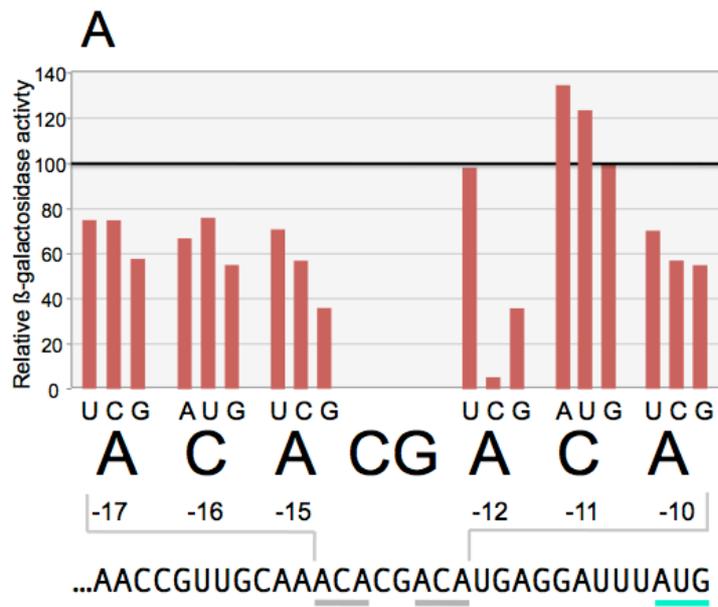


Figure 7. Effet des modifications de la région 5'UTR d'*yifK* sur la traduction de l'ARNm.

## II.1.4 . Anatomie d'un activateur traductionnel

Pour caractériser l'élément activateur (enhancer), une partie définie de la séquence de fixation du ribosome d'*yifK* a été modifiée en changeant individuellement et systématiquement chaque résidu par chacune des trois autres bases. En particulier, nous nous sommes concentrés sur deux triplets ACA, le premier correspondant à l'emplacement décrit ci-dessus (positions -17 à -15 par rapport à l'AUG initiateur) et le second quasi-adjacent à la séquence de SD (-12 à -10). Cette analyse a été entièrement réalisée dans une souche où le gène *GcvB* a été délété. Comme le montre la figure 8A, n'importe quel changement de nucléotide dans le premier triplet ACA entraîne une baisse de l'expression d'*yifK-lacZY*. Les variations que nous observons dans l'expression fluctuent entre 23% et 62%, et où le résidu G exerce les effets les plus néfastes à n'importe quelle position. Ses effets semblent être additifs puisque lors d'une expérience distincte dans laquelle les trois bases dans le triplet ACA ont été randomisés ont permis d'obtenir des allèles ayant une diminution de l'expression d'*yifK-lacZY* de 92% (Figure 8B).

Les modifications du second triplet ACA produisent des résultats quelque peu différents. Les variations du résidu C central sont soit neutres (G) ou stimulantes (A ou U) pour l'expression d'*yifK-lacZY*, alors qu'avoir un résidu C à la première position de ce triplet est très délétère entraînant une réduction de près de 95 % de l'activité  $\beta$ -galactosidase (figure 8A). Malgré le fait que nous ne sachions pas les bases physiques de ces effets, les données de la figure 6 indiquent clairement un rôle du triplet ACA dans la traduction d'*yifK*.



**B**

-17 -15 triplet	Relative activity
AGG	34
CGG	35
GGG	8
ACC	58
UCC	33
CCC	29
ACA	100

Figure 8. Analyse des mutations des triplets ACA.

### III . Discussion

Au cours de ce travail, nous avons caractérisé la régulation du locus *yifK* chez *Salmonella*, gène codant pour un transporteur putatif d'acides aminées hautement conservé chez les entérobactéries. Nos travaux ont montré qu'*yifK* est contrôlé négativement au niveau post-transcriptionnel par le petit ARN régulateur GcvB.

L'analyse génétique des interactions entre le petit ARN régulateur GcvB et l'ARNm *yifK* a révélé que la séquence cible de GcvB sur l'ARNm d'*yifK* contient un enhancer. Curieusement, les mutations qui vont perturber l'enhancer – diminuant l'expression d'*yifK* – rendent l'expression d'*yifK* totalement insensible à la répression induite par GcvB. Ces résultats suggèrent que l'efficacité de la régulation par GcvB est dépendante de la fonction de l'enhancer et que lorsque ce dernier est supprimé, la répression liée à GcvB ne constitue plus une étape limitante dans l'expression d'*yifK*. Sharma et ses collaborateurs (2007) ont montré précédemment que la séquence cible de GcvB dans le gène *gtII* de *Salmonella* agit comme un enhancer transférable. Contrairement à notre étude, les effets de GcvB comme répresseur de la traduction sont beaucoup plus importants que les effets de la suppression de l'enhancer, ce qui a conduit les auteurs à conclure que GcvB n'a pas bloqué l'effet de l'enhancer (Sharma *et al.*, 2007). Il est possible que ces résultats soient dus à la nature plasmidique du gène *gcvB* dans l'étude réalisée par Sharma et ses collaborateurs fait que la répression induite par le petit ARN soit plus forte que lorsque ce dernier est exprimé sur le chromosome. D'autre part, il est possible que l'enhancer soit moins efficace chez *gtII* que chez *yifK*. L'enhancer de *gtII*, localisé 45nt en amont du codon d'initiation, est caractérisé en tant qu'une partie d'une région 27nt et n'a pas été étudié en détail (Sharma *et al.*, 2007). Dans notre cas, nous avons constaté que le remplacement dans l'un ou l'autre des triplets ACA au sein de la séquence d'appariement de GcvB sur l'ARNm d'*yifK* peut entraîner une diminution de plus de 90 de l'expression d'*yifK*. Bien que nos données ne nous permettent pas de définir les contours de l'enhancer, ils identifient très clairement le motif ACA comme un facteur déterminant pour son activité. Nous avons également pu constater que l'activité de l'enhancer est maintenue après un changement dans la position de ce dernier de 7nt par rapport au site d'initiation, ce qui suggère l'absence d'exigences spatiales strictes pour le fonctionnement de cet enhancer. Ceci est cohérent avec les données obtenues après les travaux sur le gène *gtII* et avec une étude montrant que les répétitions du doublet CA stimulent la traduction même lorsqu'il est placé en aval du codon d'initiation (Martin-Farmer & Janssen, 1999).

L'efficacité d'initiation de traduction est connue pour varier fortement en fonction du contexte de la séquence de la région d'initiation (Dreyfus, 1988, Yarchuk *et al.*, 1992). L'analyse informatique des séquences entourant les sites d'initiation de traduction de gènes d'*E. coli* a montré que la distance entre la SD et le codon d'initiation affecte la conservation de la séquence de la SD et son motifs. Cette étude n'a pas révélé de biais importants en dehors de ces principaux éléments (Shultzaberger *et al.*, 2001); mais des motifs conservés qui se produisent à des positions variables sont difficiles à identifier en analyse statistique. En effet, des lignes de vue différentes pointent le rôle du motif ACA dans l'initiation de la traduction.

Ce motif est retrouvé dans d'autres séquences activatrices de la traduction (Komarova *et al.*, 2002, McCarthy *et al.*, 1985) et, comme la répétition ACAA, il a été montré comme favorisant l'initiation de la traduction, en l'absence d'une séquence SD (Tzareva *et al.*, 1994). Le motif ACA se retrouve également dans les boucles des pseudonœuds formés entre l'ARN ligand et la protéine ribosomal S1, obtenus par Evolution Systématique de Ligands par enrichissement Exponentiel (SELEX) (Ringquist *et al.*, 1995) et fait partie de la séquence consensus déterminé par la méthode SELEX de liaison à la protéine CsrA, un régulateur traductionnelle (Dubey *et al.*, 2005). Enfin, le motif ACA est la séquence de reconnaissance de l'endonucléase MazF qui inactive les ARNm d'*E. coli* en clivant préférentiellement à proximité du codon d'initiation de la traduction (Vesper *et al.*, 2011).

Le manque d'exigences dans sa position pour la fonction de l'enhancer suggère que son rôle est de fournir un point d'ancrage à la sous-unité 30S du ribosome de manière à faciliter la reconnaissance ultérieure de la séquence SD. Certains éléments de preuve examinés ci-dessus identifient provisoirement la protéine S1 comme candidat possible pour cette interaction. Des études de liaison de la protéine S1 *in vitro* avec certains des mutants construits dans le cadre de ces travaux devraient permettre de tester cette hypothèse. Combiné avec l'analyse mutationnelle d'autres ARNm régulé par GcvB, cette approche pourrait fournir plus de renseignements sur la façon dont le motif ACA participe à l'étape d'initiation de la traduction.

# Bibliographie

- Babitzke, P. & T. Romeo, (2007) CsrB sRNA family: sequestration of RNA-binding regulatory proteins. *Current opinion in microbiology* **10**: 156-163.
- Brantl, S., (2007) Regulatory mechanisms employed by cis-encoded antisense RNAs. *Current opinion in microbiology* **10**: 102-109.
- Brennan, R.G. & T.M. Link, (2007) Hfq structure, function and ligand binding. *Current opinion in microbiology* **10**: 125-133.
- Dreyfus, M., (1988) What constitutes the signal for the initiation of protein synthesis on Escherichia coli mRNAs? *Journal of molecular biology* **204**: 79-94.
- Dubey, A.K., C.S. Baker, T. Romeo & P. Babitzke, (2005) RNA sequence and secondary structure participate in high-affinity CsrA-RNA interaction. *RNA* **11**: 1579-1587.
- Figueroa-Bossi, N., S. Lemire, D. Maloriol, R. Balbontin, J. Casadesus & L. Bossi, (2006) Loss of Hfq activates the sigmaE-dependent envelope stress response in Salmonella enterica. *Molecular microbiology* **62**: 838-852.
- Hajnsdorf, E. & P. Regnier, (2000) Host factor Hfq of Escherichia coli stimulates elongation of poly(A) tails by poly(A) polymerase I. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 1501-1505.
- Hughes, K.T. & J.R. Roth, (1988) Transitory cis complementation: a method for providing transposition functions to defective transposons. *Genetics* **119**: 9-12.
- Komarova, A.V., L.S. Tchufistova, E.V. Supina & I.V. Boni, (2002) Protein S1 counteracts the inhibitory effect of the extended Shine-Dalgarno sequence on translation. *RNA* **8**: 1137-1147.
- Martin-Farmer, J. & G.R. Janssen, (1999) A downstream CA repeat sequence increases translation from leadered and unleadered mRNA in Escherichia coli. *Molecular microbiology* **31**: 1025-1038.
- Marzi, S. & P. Romby, (2012) RNA mimicry, a decoy for regulatory proteins. *Molecular microbiology* **83**: 1-6.
- McCarthy, J.E., H.U. Schairer & W. Sebald, (1985) Translational initiation frequency of atp genes from Escherichia coli: identification of an intercistronic sequence that enhances translation. *The EMBO journal* **4**: 519-526.
- Ohl, M.E. & S.I. Miller, (2001) Salmonella: a model for bacterial pathogenesis. *Annual review of medicine* **52**: 259-274.
- Rabhi, M., O. Espeli, A. Schwartz, B. Cayrol, A.R. Rahmouni, V. Arluison & M. Boudvillain, (2011) The Sm-like RNA chaperone Hfq mediates transcription antitermination at Rho-dependent terminators. *The EMBO journal* **30**: 2805-2816.
- Repoila, F., N. Majdalani & S. Gottesman, (2003) Small non-coding RNAs, co-ordinators of adaptation processes in Escherichia coli: the RpoS paradigm. *Molecular microbiology* **48**: 855-861.
- Ringquist, S., T. Jones, E.E. Snyder, T. Gibson, I. Boni & L. Gold, (1995) High-affinity RNA ligands to Escherichia coli ribosomes and ribosomal protein S1: comparison of natural and unnatural binding sites. *Biochemistry* **34**: 3640-3648.
- Romeo, T., (1998) Global regulation by the small RNA-binding protein CsrA and the non-coding RNA molecule CsrB. *Molecular microbiology* **29**: 1321-1330.
- Sauter, C., J. Basquin & D. Suck, (2003) Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein from Escherichia coli. *Nucleic acids research* **31**: 4091-4098.

- Sharma, C.M., F. Darfeuille, T.H. Plantinga & J. Vogel, (2007) A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. *Genes & development* **21**: 2804-2817.
- Shultzaberger, R.K., R.E. Bucheimer, K.E. Rudd & T.D. Schneider, (2001) Anatomy of Escherichia coli ribosome binding sites. *Journal of molecular biology* **313**: 215-228.
- Tsui, H.C., H.C. Leung & M.E. Winkler, (1994) Characterization of broadly pleiotropic phenotypes caused by an hfq insertion mutation in Escherichia coli K-12. *Molecular microbiology* **13**: 35-49.
- Tzareva, N.V., V.I. Makhno & I.V. Boni, (1994) Ribosome-messenger recognition in the absence of the Shine-Dalgarno interactions. *FEBS letters* **337**: 189-194.
- Vesper, O., S. Amitai, M. Belitsky, K. Byrgazov, A.C. Kaberdina, H. Engelberg-Kulka & I. Moll, (2011) Selective translation of leaderless mRNAs by specialized ribosomes generated by MazF in Escherichia coli. *Cell* **147**: 147-157.
- Vogel, J. & B.F. Luisi, (2011) Hfq and its constellation of RNA. *Nature reviews. Microbiology* **9**: 578-589.
- Yarchuk, O., N. Jacques, J. Guillerez & M. Dreyfus, (1992) Interdependence of translation, transcription and mRNA degradation in the lacZ gene. *Journal of molecular biology* **226**: 581-596.
- Zinder, N.D. & J. Lederberg, (1952) Genetic exchange in Salmonella. *Journal of bacteriology* **64**: 679-699.

**Remerciement :** Ce résumé en français a été réalisé avec l'aide de M. Benoit Guillemardet.