



**HAL**  
open science

# Rôle des régulations de la stabilité des ARN messagers dans l'adaptation d'*Escherichia coli* à son environnement

Thomas Esquerre

## ► To cite this version:

Thomas Esquerre. Rôle des régulations de la stabilité des ARN messagers dans l'adaptation d'*Escherichia coli* à son environnement. Biologie moléculaire. INSA de Toulouse, 2014. Français. NNT : 2014ISAT0016 . tel-01186712

**HAL Id: tel-01186712**

**<https://theses.hal.science/tel-01186712>**

Submitted on 25 Aug 2015

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

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



# THÈSE

## DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut National des Sciences Appliquées de Toulouse (INSA de Toulouse)

---

**Présentée et soutenue par :**

Thomas Esquerré

Le mardi 01 juillet 2014

Titre :

Rôle des régulations de la stabilité des ARN messagers  
dans l'adaptation d'*Escherichia coli* à son environnement

---

**École doctorale et spécialité :**

ED SEVAB : Ingénieries microbienne et enzymatique

**Unités de recherche :**

Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés – UMR 5504  
Laboratoire de Microbiologie et de Génétique Moléculaires – UMR 5100

**Directeurs de Thèse :**

Dr. Laurence Girbal  
Dr. Agamemnon J. Carpousis

**Jury de Thèse :**

Dr. Philippe Bouloc, Directeur de recherche, CNRS, Orsay (Rapporteur)  
Dr. Hans Geiselman, Professeur d'Université, Université Joseph Fournier, Grenoble (Rapporteur)  
Dr. Harald Putzer, Directeur de recherche, CNRS, Paris (Examinateur)  
Dr. Agamemnon J. Carpousis, Directeur de recherche, CNRS, Toulouse  
Dr. Laurence Girbal, Chargée de recherche, CNRS, Toulouse  
Dr. Muriel Cocaign-Bousquet, Directrice de recherche, INRA, Toulouse



Cette thèse a été financée avec le soutien de  
la **région Midi-Pyrénées**



et de l'**Université de Toulouse**







## Rôle des régulations de la stabilité des ARNm dans l'adaptation d'*Escherichia coli* à son environnement

L'adaptation des bactéries à leur environnement résulte de régulations de l'expression génique pour optimiser leur physiologie aux conditions de culture. Le contrôle de la concentration des ARNm constitue l'une de ces régulations. Il dépend à la fois des variations de transcription et de dégradation des messagers. Si ces deux mécanismes sont bien étudiés à l'échelle moléculaire chez *E. coli*, leurs poids respectifs sur la régulation du niveau des transcrits à l'échelle du génome restent inconnus en raison de l'absence de données omiques relatives à la dégradation des ARNm lors de changements environnementaux. D'autre part, les paramètres déterminant la stabilité des messagers sont mal identifiés et n'ont jamais été hiérarchisés.

Au cours de cette thèse, la stabilité de chacun des ARNm d'*E. coli* a été mesurée par la détermination du stabilome. Plus précisément, le temps de demi-vie de près de 70 % de tous les messagers a pu être déterminé de façon fiable pour quatre taux de croissance différents obtenus dans les mêmes conditions de culture à l'aide de chimostats. Pour la première fois, cette étude démontre qu'une croissance bactérienne plus rapide entraîne une augmentation globale de la dégradation des transcrits. L'analyse conjointe de ces données avec les données transcriptomiques montre que même si la transcription est le mécanisme principal de régulation du niveau des messagers, la dégradation exerce un effet inverse dans la plupart des cas. De plus, le rôle de la dégradation dans le contrôle de la concentration des ARNm s'accroît de façon significative avec l'augmentation du taux de croissance et affecte particulièrement les gènes impliqués dans le métabolisme carboné central.

À partir des données de stabilité générées à différents taux de croissance, des approches de biologie intégrative ont permis d'identifier et de hiérarchiser les déterminants de la dégradation des ARNm. Ainsi, la concentration des messagers qui est le principal paramètre, mais aussi le biais de codon, la longueur de la séquence codante et la présence de certains motifs de séquence déterminent la stabilité d'un ARNm. Toutefois, si la hiérarchie des déterminants identifiés reste identique avec la variation du taux de croissance, la stabilité des ARNm de certaines catégories fonctionnelles en est dépendante. D'autres déterminants du temps de demi-vie des messagers, en particulier à fort taux de croissance, restent néanmoins à être identifiés.

La protéine CsrA, appartenant au système Csr, est un exemple de régulateur post-transcriptionnel qui contrôle positivement ou négativement l'expression d'ARNm par divers mécanismes qui peuvent modifier leur stabilité. Toutefois, l'étendue de l'action de CsrA sur la stabilité des ARNm à l'échelle omique n'a jamais été étudiée. En comparant les stabilomes et transcriptomes d'une souche sauvage et d'une souche où l'activité de CsrA est diminuée, les effets indirects transcriptionnels de CsrA ont été mesurés et de nouveaux ARNm cibles de CsrA dont la stabilité est régulée, en majorité positivement, ont été identifiés. De plus, il a été démontré au cours de cette thèse que la protéine CsrD, régulateur de la stabilité des ARN non codants CsrB/C, n'est pas impliquée dans la régulation de la stabilité des ARNm, mais agit sur la transcription de nombreux gènes indépendamment de son rôle au sein du système Csr.

En conclusion, ces travaux ont permis de mieux appréhender les régulations de la stabilité des ARNm, en identifiant leurs déterminants et en caractérisant leur rôle et portée dans le contrôle de la concentration des messagers. Ils soulignent en particulier l'importance de ces régulations dans le processus d'adaptation bactérien.



## Role of mRNA stability regulation in *Escherichia coli* adaptation to environment

Bacterial adaptation to environment results from regulations of gene expression to optimize cell physiology to growth conditions. Control of mRNA concentration is one of those regulations. It depends on both variations of transcription and transcript degradation. Although these two mechanisms are well defined at the molecular level in *E. coli*, their respective impact on mRNA level regulation is still unknown at the genome scale because of a lack of omic data on mRNA stability during changing environment. Moreover, parameters determining messenger stability are not yet clearly identified and have never been ranked.

During this PhD, the stability of each of the *E. coli* mRNAs was measured through stabilome determination. More precisely, the half-life of around 70 % of all messengers was reliably determined at four different growth rates obtained in the same growth conditions in chemostats. For the first time, this study demonstrated that increase of growth rate led to global increase of transcript degradation. Joint analysis of these data with transcriptomic data showed that although transcription was the main mechanism which regulated mRNA level, messenger degradation exerted an opposite effect in most of the cases. The role of messenger degradation in the control of mRNA concentration was significantly accentuated with increasing growth rate and affected particularly genes involved in central carbon metabolism.

Using mRNA stability data produced at different growth rates, integrative biology approaches allowed identification and ranking of the determinants of messenger stability. mRNA concentration was the main parameter, but also codon bias, length of the coding sequence, sequence motifs contributed to transcript stability. However, although the hierarchy of determinants remained identical with variations of growth rate, the stability of mRNAs belonging to specific functional categories differed with the growth rate. Nevertheless, other determinants of messenger half-life, in particular at high growth rates still remain to be discovered.

The CsrA protein, which belongs to the Csr system, is one example of a post-transcriptional regulator. CsrA positively or negatively controls expression of several mRNAs by mechanisms able to modify transcript stability. Nevertheless, the extent of CsrA effect on mRNA stability at the omic level has never been studied. By comparing stabilomes and transcriptomes of the wild type strain with a strain with reduced CsrA activity, the indirect transcriptional effects of CsrA were measured and new mRNAs whose stability was targeted by CsrA (mostly stabilized) were identified. Moreover, the CsrD protein, a regulator of CsrB/C small RNA stability, was not involved in mRNA stability regulation, but played a role in transcriptional regulation of many genes independently of its role in the Csr system.

To conclude, this work provides a better understanding of the regulation of the mRNA stability. It identifies mRNA stability determinants and characterizes the role and extent of mRNA stability regulation in the control of messenger concentration. The study underlines the importance of this regulation in the process of bacterial adaptation.



## Remerciements

Je souhaite tout d'abord remercier l'ensemble des membres de mon jury de thèse pour avoir accepté d'évaluer ces travaux malgré des emplois du temps chargés : **Hans Geiselmann, Philippe Bouloc et Harald Putzer**.

J'ai toujours dit que j'avais eu de la chance lors de mes précédents stages et expériences professionnelles. Une fois de plus, ma bonne étoile a fait des merveilles ! Une fois de plus, mes encadrants ainsi que mes collègues ont été **exceptionnels** et ont rendu cette aventure si agréable à vivre ! Pour faire simple, séparons les remerciements en fonction des lieux géographiques si vous le voulez bien !

### Tout d'abord au LMGM :

Je souhaite en premier lieu dire chaleureusement merci à **A. J. Carpousis**, mon co-directeur de thèse au LMGM pour ses propositions et ses conseils éclairés vis-à-vis de mon travail ainsi que pour sa disponibilité malgré son statut de directeur de laboratoire. Malheureusement, mes séjours au LMGM ont été trop rares mais j'y ai toujours trouvé : écoute, coups de pouce et bienveillance. Les expériences menées là-bas ont été un vrai régal tant du point de vue professionnel (la chance du débutant : toutes les manip ont fonctionné du premier coup !) que pour l'ambiance qu'il y régnait.

Je souhaite tout particulièrement remercier **Catherine Turlan** pour l'exceptionnelle qualité de son encadrement même si celui-ci fut trop court... J'espère que ta nouvelle vie t'apportera toutes les satisfactions que tu peux souhaiter.

Évidemment, je ne peux oublier de remercier **Marie Bouvier, Léonora Poljak, Béatrice Clouet-d'Orval** ainsi que **Dana Rinaldi** (qui m'a ensuite rejoint à l'INSA) pour leur soutien, leur aide et leur bonne humeur. Enfin, je souhaite bon courage à **Duy Khanh Phung** pour la suite de sa thèse même si je ne me fais aucun souci pour lui !

Je ne peux également pas oublier toutes les personnes de l'équipe aux côtés desquelles j'ai travaillé : **Frédéric Angles** qui m'a initié au ski nautique, **Jean-Marie Kebalo** et **Pascale Salek** qui ont choisi tous les deux par la suite, d'autres voies que la recherche : bon vent à tous !

### Ensuite à l'INSA de Toulouse :

Je veux maintenant exprimer toute ma gratitude, toute mon amitié, et un très très grand merci à **Laurence Girbal**<sup>#</sup> et **Muriel Cocaïgn-Bousquet**<sup>#</sup> dans cet ordre et dans l'autre (<sup>#</sup> Equally thanked by Thomas) car je ne veux/peux pas choisir d'en distinguer une plus que l'autre comme on nous l'impose bien souvent dans les papiers. Comme Flora le disait déjà avant moi : votre complémentarité est exemplaire ! Bien qu'ayant des visions différentes des choses, vous trouvez toujours un terrain d'entente qui fait ressortir la meilleure solution. Cette complémentarité prouve encore une fois que le travail collectif est indispensable pour faire de la « bonne » science. C'était un véritable privilège d'avoir été votre « apprenti » mais aussi votre collègue car vous m'avez toujours proposé une écoute attentive. Laurence, merci pour cette rigueur dans laquelle je me suis retrouvé mais aussi pour ta patience (oh oui, il lui en a fallu !) et ton calme lors de mes assauts pour te prouver que j'avais raison (même si j'avais tort !) ! Muriel, merci pour ton optimisme sans faille et ce que Flora appelait ton « enthousiasme rassurant » alors même que les résultats paraissent décevants (le thésard, avide de bons résultats, a besoin d'entendre dire que ses résultats sont super !). Votre côté humain à

toutes les deux a fait que ces trois ans se sont déroulés dans les meilleures conditions possibles.

Ensuite, je veux remercier chaudement mes collègues de bureau (les « jeunes » ou « précaires » dans notre jargon local) qui ont évidemment animé mon quotidien. Cette entraide entre nous fait partie des meilleurs moments de ces trois ans. Ceux qui m'ont aidé à finir en premier lieu : « **Tutur** » (Arthur Raymond) et son oustiti, « **Maman** » (Anne-Laure Finoux) et son catalogue de mes frasques les plus célèbres, « **Popo** » (Pauline Goodwin), mon dictionnaire Anglais/Français en direct, **Manon Morin**, la cinéphile et pâtissière, **Marie Izac**, ma binôme de running (car c'est tout de même plus sympa de courir accompagné !), **Cécile Philippe**, la transfuge ainsi que **Stefano Nebbia et Nina Hemery** les tous derniers stagiaires de l'équipe. Merci également à **Rozenn Le Guellec** qui vient d'arriver. Mais avant eux, d'autres se sont relayés. En premier lieu : **Amandine Dhaisne** – camarade de terminale – retrouvée ici après 6 ans, toujours aussi travailleuse (une véritable locomotive pour moi !) et appliquée tant dans sa connaissance que dans l'aide qu'elle m'a fournie, **Stéphanie Da Silva** – camarade de promotion à l'IUP – hyperactive et le cœur sur la main, **Jérôme Le Hir**, à l'humour plus que douteux (que j'apprécie tant !) mais aussi **Ly Tran Thi**, **Flora Picard** (mon aïeule), **Clémentine Dressaire** (ma « grande » aïeule), **Delphine Passerini**, **Marie-Anne Gauriat**, **Lam Doan Thanh**, **Laurie Watterlot**, **Mélanie Rayme**, **Jade Riba**, **Claire Lanusse-Crousse**, **Claire Derlot**, **Lucie Peloux**, **Alessandra Fontana**, **Shyam Suwal** et **Chiara Porta**.

De la même façon, je souhaite remercier sincèrement et chaleureusement l'ensemble des « permanents » de l'équipe (ou « vieux » toujours dans notre jargon local) qui ont toujours répondu présents pour donner un coup de main et animé nos repas de midi ! Il y a notamment, les fans de rugby : **Brice Enjalbert** et **Mickaël Castelain** qui se sont affrontés une nouvelle fois cette année (Castres vs Toulon), **Sébastien Nouaille**, notre consultant "Stabilité et Bio Mol" à qui on peut tout dire et demander (et qui paraît tellement jeune qu'on le prend pour un post-doc !), **Sandrine Laguerre**, notre statisticienne ADOREE (merci pour ton aide oh combien précieuse à certains moments !!!), **Marie-Pierre Duviau**, la multicasquette de l'équipe dont la patience avec les étudiants n'est plus à démontrer, **Sophie Mondeil**, notre aventurière transatlantique, **Valérie Laroute**, la reine de l'HPLC, **Nic Lindley**, notre directeur de laboratoire et **Pascal Loubière** et **Myriam Mercade** qui sont partis momentanément vers d'autres horizons ! Je souhaite également la bienvenue à **Pascal Lebourgeois** et **Marie-Line Mingot** qui se sont déjà intégrés dans l'équipe !

Il est maintenant l'heure de parler des autres personnes que j'ai cotoyées au cours de ces trois ans : un très grand merci en premier lieu à **Lidwine Trouilh** pour son expertise en microarray mais aussi pour son écoute et ses précieux conseils. Merci également au reste de la plateforme Biopuces ainsi qu'à tous les membres de MetaSys avec qui j'ai fait un bout de chemin autour et en dehors des prélèvements chémostats : **Pierre Millard**, **Edern Cahoreau**, **Lara Gales**, **Juliette Pointot**, **Lindsay Peyriga**, **Fabien Létisse**, **Jean-Charles Portais** et tous les autres... Je veux remercier aussi nos collaborateurs pour la qualité de leur travail : **Hélène Chiapello**, **Annick Moisan**, **Christine Gaspin**, **Liisa Arike**, **Kaspar Valgepea** et **Raivo Vilu** !

Je souhaite également dire merci à **Aaron Peter Green** pour avoir été mon premier « vrai » stagiaire avec qui, il y a eu des hauts et des bas ! Malgré tout, je garde un très bon

souvenir de cette première expérience d'enseignement tout comme avec les étudiants de première année de l'INSA à qui j'ai eu l'honneur de dispenser des enseignements de.....chimie ! Merci également à toute l'équipe enseignante des TP de chimie 1<sup>ère</sup> année et à **Claude Maranges** pour son écoute et ses conseils avisés ! Enfin, je remercie également tout le personnel du laboratoire, scientifique et non scientifique, pour ces trois superbes années.

Je voudrais maintenant ajouter une mention spéciale pour..... mon **PC de bureau** ! Oui, Oui ! Finalement, c'est avec lui que j'ai passé le plus de temps au cours de ces trois dernières années (analyses de données omiques !!) et je le remercie de ne jamais m'avoir lâché !

Enfin, merci du fond du coeur à mes proches : **parents, frère et soeur, grands parents, famille** et **amis** pour m'avoir poussé et soutenu dans cette aventure. Pour finir, mon dernier merci est pour **Sarah** qui partage ma vie. Pendant ces trois ans, tu as supporté mes hauts, mes bas, l'euphorie, le désespoir, l'enthousiasme, le doute... mais tu as toujours été là pour me soutenir, me rassurer, m'épauler et me pousser même dans les moments les plus difficiles. Je te remercie pour tout ça... De nouvelles aventures nous attendent mais je sais depuis longtemps que je peux compter sur toi !





# **SOMMAIRE**



---

<b>INTRODUCTION.....</b>	<b>1</b>
<b>ETUDE BIBLIOGRAPHIQUE.....</b>	<b>7</b>
<b>I. La bactérie <i>Escherichia coli</i>.....</b>	<b>9</b>
I. 1. Description générale.....	9
I. 2. Pathogénicité et classification.....	9
I. 3. La souche <i>E. coli</i> K-12 MG1655.....	11
<b>II. Les différents niveaux de régulation de l'expression génique chez <i>Escherichia coli</i> en réponse à l'environnement.....</b>	<b>13</b>
II. 1. S'adapter à l'environnement.....	13
II. 2. Les régulations transcriptionnelles.....	15
II. 2. 1. La régulation globale de la transcription.....	15
II. 2. 2. Les régulations spécifiques de la transcription.....	17
II. 3. Les régulations post-transcriptionnelles.....	20
II. 3. 1. Les régulations de la stabilité des ARN.....	20
II. 3. 2. Les régulations de la traduction.....	22
II. 4. Les régulations post-traductionnelles.....	25
II. 4. 1. Les régulations de la stabilité des protéines.....	25
II. 4. 2. Les régulations de la disponibilité ou de l'activité protéique.....	27
<b>III. Les mécanismes de dégradation des ARNm chez <i>Escherichia coli</i>.....</b>	<b>29</b>
III. 1. Les principales ribonucléases.....	29
III. 1. 1. Les endoribonucléases.....	29
III. 1. 2. Les exonucléases.....	31
III. 2. Les mécanismes de dégradation des ARNm.....	32
III. 3. Le dégradosome.....	34
III. 4. Les mécanismes régulant la stabilité des ARNm.....	36
III. 4. 1. Les régulations de la RNase E.....	36
III. 4. 2. Les effets de la traduction.....	36
III. 4. 3. L'interaction Hfq / ARNnc / RNase E.....	37
III. 4. 4. Les autres molécules se liant à l'ARNm.....	39
III. 4. 5. Les systèmes toxine-antitoxine.....	40
III. 4. 6. L'influence de la séquence des ARNm.....	40
III. 5. Les déterminants de stabilité des ARNm à l'échelle omique.....	41
<b>IV. Le système Csr chez <i>Escherichia coli</i>.....</b>	<b>43</b>
IV. 1. Description du système Csr.....	43

IV. 1. 1. Le régulateur post-transcriptionnel global CsrA .....	43
IV. 1. 2. Les ARNnc séquestrant la protéine CsrA .....	46
IV. 1. 3. La protéine CsrD .....	47
IV. 2. Les mécanismes de régulation par CsrA .....	48
IV. 2. 1. Les mécanismes de répression .....	48
IV. 2. 2. Les mécanismes de régulation positive .....	50
IV. 3. Les ARNm cibles du système Csr .....	51
IV. 4. Le réseau de régulation du système Csr .....	53
IV. 4. 1. Régulation de CsrA .....	53
IV. 4. 2. Régulation des petits ARN non codants .....	53
IV. 4. 3. Régulation de CsrD .....	54
IV. 4. 4. Liens avec d'autres systèmes de régulation .....	55
V. Problématique de la thèse .....	57

## CHAPITRE I

<b>Rôle de la transcription et de la stabilité des transcrits dans la régulation de l'expression génique à différents taux de croissance d'<i>Escherichia coli</i> sur glucose .....</b>	<b>61</b>
--	-----------

## CHAPITRE II

<b>Identification des déterminants de la stabilité des ARNm à différents taux de croissance .....</b>	<b>83</b>
---	-----------

## CHAPITRE III

<b>Détermination à l'échelle omique de l'étendue du contrôle de l'expression et de la stabilité des ARNm par le système Csr chez <i>Escherichia coli</i> .....</b>	<b>119</b>
--	------------

<b>CONCLUSION ET PERSPECTIVES .....</b>	<b>151</b>
---	------------

<b>REFERENCES BIBLIOGRAPHIQUES .....</b>	<b>165</b>
--	------------

<b>PUBLICATIONS ET COMMUNICATIONS .....</b>	<b>187</b>
---	------------

<b>ANNEXES .....</b>	<b>193</b>
----------------------	------------

## INDEX DES ABREVIATIONS

### Micro-organismes

<i>B. cereus</i>	<i>Bacillus cereus</i>
<i>E. coli</i>	<i>Escherichia coli</i>
<i>H. salinarum</i>	<i>Halobacterium salinarum</i>
<i>L. lactis</i>	<i>Lactococcus lactis</i>
<i>M. tuberculosis</i>	<i>Mycobacterium tuberculosis</i>
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>

### Français

<b>3'UTR</b>	Région 3' non traduite
<b>5'UTR</b>	Région 5' non traduite
<b>AAA</b>	ATPases associées à diverses activités cellulaires
<b>ADN</b>	Acide désoxyribonucléique
<b>AHLs</b>	N-acyl-L homoserine lactones
<b>AMP</b>	Adénosine monophosphate
<b>ARN</b>	Acide ribonucléique
<b>ARNm</b>	ARN messenger
<b>ARNnc</b>	ARN non codant
<b>ARNr</b>	ARN ribosomique
<b>ARNt</b>	ARN de transfert
<b>ARNtm</b>	ARN de transfert-messenger
<b>ATP</b>	Adénosine tri-phosphate
<b>c-di-GMP</b>	Diguanosine monophosphate cyclique
<b>CAI</b>	Biais de codon
<b>HSP</b>	Protéine de choc thermique
<b>k</b>	Constante de vitesse de dégradation
<b>Kb</b>	Kilo base (1 Kb = 10 <sup>3</sup> bases)
<b>Mb</b>	Méga base (1 Mb = 10 <sup>6</sup> bases)
<b>NAP</b>	Protéine associée au nucléoïde
<b>ORF</b>	Cadre ouvert de lecture
<b>PGA</b>	Poly-β-1,6-N-acetyl-D-glucosamine
<b>ppGpp</b>	Guanosine tétraphosphate
<b>RBS</b>	Site de fixation des ribosomes
<b>RNase</b>	Ribonucléase
<b>SELEX</b>	Évolution systématique de ligands par enrichissement exponentiel
<b>t<sub>1/2</sub></b>	Temps de demi-vie
<b>μ</b>	Taux de croissance

### English

<b>3'UTR</b>	3' Untranslated region
<b>5'UTR</b>	5' Untranslated region
<b>AIC</b>	Akaike information criterion
<b>bp</b>	Base pair
<b>c-di-GMP</b>	Cyclic diguanosine monophosphate
<b>CAI</b>	Codon adaptation index

## *Abréviations*

---

<b>cAMP</b>	Cyclic adenosine monophosphate
<b>CDS</b>	Coding DNA sequence
<b>COG</b>	Cluster of orthologous group
<b>DNA</b>	Deoxyribonucleic acid
<b>GRAVY</b>	Grand average of hydropathicity score
<b>iBAQ</b>	Intensity-based absolute quantification
<b><i>k</i></b>	Degradation rate constant
<b><i>k<sub>trans</sub></i></b>	Transcription rate constant
<b>mRNA</b>	Messenger ribonucleic acid
<b>NAD</b>	Nicotinamide adenine dinucleotide
<b>ORF</b>	Open reading frame
<b>PGA</b>	Poly-β-1,6-N-acetyl-D-glucosamine
<b>ppGpp</b>	Guanosine tetraphosphate
<b><math>\rho_D</math></b>	Degradation regulation coefficient
<b><math>\rho_T</math></b>	Transcription regulation coefficient
<b><i>q<sub>Acetate</sub></i></b>	Specific acetate production rate
<b><i>q<sub>s</sub></i></b>	Specific glucose consumption rate
<b><math>R^2</math></b>	Determination coefficient
<b>RBS</b>	Ribosome binding site
<b>RNA</b>	Ribonucleic acid
<b>RNase</b>	Ribonuclease
<b>rRNA</b>	Ribosomal ribonucleic acid
<b>SELEX</b>	Systematic evolution of ligands by the exponential enrichment
<b>SILAC</b>	Stable isotope labeling by amino acids in cell culture
<b>sRNA</b>	Small non coding ribonucleic acid
<b><math>t_{1/2}</math></b>	Half-life
<b>tAI</b>	Transfer ribonucleic acid adaptation index
<b>TCA</b>	Tricarboxylic acid cycle
<b>tRNA</b>	Transfer ribonucleic acid
<b>TSS</b>	Transcription starting site
<b><math>\mu</math></b>	Growth rate
<b><math>\mu_{max}</math></b>	Maximal growth rate







# **INTRODUCTION**



La bactérie *Escherichia coli* est une des bactéries les plus couramment retrouvées au sein de la flore intestinale. Elle constitue un groupe très divers de bactéries entériques avec des souches commensales et pathogènes. Ce bacille Gram-négatif est capable de se développer dans des conditions très variables et notamment sur un grand nombre de substrats carbonés différents. Ces données témoignent que la bactérie dispose d'une importante faculté d'adaptation à son environnement associée à un réseau complexe de régulation de l'expression des gènes. Le contrôle de l'expression génique intègre plusieurs niveaux de régulation depuis la transcription d'un gène jusqu'à la disponibilité ou l'activité de la protéine correspondante. Une meilleure compréhension des mécanismes sous-jacents est donc indispensable pour pouvoir anticiper l'adaptation de la bactérie en fonction de son environnement. Ces connaissances devraient permettre à plus long terme le développement de nouveaux traitements antimicrobiens non-conventionnels comme par exemple la construction de souches probiotiques pour prévenir l'implantation de souches pathogènes opportunistes.

La régulation de l'expression génique implique des processus stratifiés et coordonnés qui étaient auparavant difficilement appréhendables dans leur globalité. L'avènement de l'ère omique a changé la donne et permet aujourd'hui d'étudier les différents niveaux de régulation des gènes à l'échelle de génomes entiers. Ces données incluent notamment le génome, le transcriptome, le protéome ou encore le métabolome (reflétant respectivement les caractéristiques de l'ensemble des gènes sur le chromosome et les niveaux des ARNm, protéines et métabolites présents au sein de la cellule dans une condition donnée). Les travaux de cette thèse se distinguent par le fait qu'ils se focalisent sur un autre type de données en « ome » : le « stabilome ». La mesure de ce dernier, basée sur le même principe que le transcriptome, donne accès à la stabilité de l'ensemble des ARNm d'un organisme. Le contrôle de la stabilité des ARNm correspond à un niveau de régulation de l'expression génique très peu étudié à l'échelle omique malgré le fait que plusieurs études sur des ARNm particuliers montrent que ce processus varie significativement en fonction des conditions environnementales.

Toutefois, bien qu'informatrice, l'analyse indépendante de chacun des niveaux de régulation de l'expression des gènes ne procure qu'une vision parcellaire de la physiologie d'adaptation de la bactérie et ne permet pas de comprendre intégralement les réponses à des changements environnementaux. Pour cela, des outils mathématiques et statistiques permettant l'intégration de différents jeux de données ont été développés

au cours de cette thèse afin d'améliorer notre compréhension du « système » bactérien dans son ensemble. Ce type d'approches appartient à un domaine particulier de la biologie appelé Biologie des Systèmes ou Biologie Intégrative.

Ainsi cette thèse s'appuie sur une quantification du stabilome, donnée encore rare aujourd'hui, et sur son intégration multi-échelle dans le but d'améliorer notre compréhension des mécanismes d'adaptation de la bactérie *E. coli*. Pour cela, l'impact post-transcriptionnel des régulations de la stabilité des ARNm sur l'expression des gènes a été étudié par deux approches distinctes : en modifiant les conditions environnementales (faisant varier le taux de croissance) et en affectant génétiquement un régulateur post-transcriptionnel global impliqué dans l'adaptation bactérienne (le système Csr).

Ce manuscrit de thèse débute par une introduction bibliographique qui reprend brièvement les caractéristiques de la bactérie *E. coli*. La deuxième partie de cette synthèse s'intéresse aux différents niveaux de régulation de l'expression génique lors de l'adaptation de la bactérie à son environnement. A l'issue de la description de ces différents niveaux, le mécanisme de dégradation des ARNm est détaillé dans une troisième partie. L'étude bibliographique se poursuit par la description du système de régulation post-transcriptionnelle global affectant la stabilité des ARNm : le système Csr. Enfin, la dernière partie de cette synthèse définit la problématique de la thèse.

Les résultats obtenus lors de ces travaux sont présentés en anglais sous forme de publications avec en préambule de chaque article, un résumé étendu en français suivi d'un schéma récapitulatif. Le premier chapitre s'intéresse au poids de la transcription et de la dégradation des ARNm dans la régulation de l'expression génique lorsque le taux de croissance de la bactérie change. Puis, les données de stabilité des ARNm sont analysées dans le deuxième chapitre pour identifier les principaux déterminants du temps de demi-vie des messagers en fonction du taux de croissance. Enfin, le troisième chapitre évalue l'effet du régulateur post-transcriptionnel CsrA sur la stabilité des ARNm à l'échelle omique.

Enfin, l'ensemble des résultats obtenus est repris dans une conclusion générale pour dresser une image complète des régulations de la stabilité des messagers à l'échelle omique. Cette conclusion termine sur les perspectives scientifiques de la thèse.





# **ETUDE**

# **BIBLIOGRAPHIQUE**





## **I. La bactérie *Escherichia coli***

### ***I. 1. Description générale***

La bactérie *Escherichia coli* est un bacille Gram négatif, anaérobie facultatif non sporulant de quelques micromètres de longueur. Elle a été isolée pour la première fois en 1885 par Theodor Escherich dans les fèces d'un nouveau-né. La classification de cette bactérie est Bacteria - Proteobacteria - Gammaproteobacteria - Enterobacteriales - Enterobacteriaceae - Escherichia - Escherichia coli (pour Règne - Embranchement - Classe - Ordre - Famille - Genre - Espèce respectivement). La bactérie *E. coli* est classiquement retrouvée dans le gros intestin d'humains et d'animaux qu'elle colonise lors de la naissance de l'individu et dès ses premiers jours de vie. On la retrouve notamment dans le cæcum et le colon où elle niche dans la couche de mucus qui tapisse les cellules épithéliales. Se nourrissant notamment de plusieurs sucres dérivés du mucus (e.g. gluconate, N-acetylglucosamine, fucose, deoxyribose, ribose, galactose, maltose,...), ses capacités métaboliques sont très importantes pour lui permettre de s'adapter et de coloniser cette niche (Chang et al., 2004). Sa prévalence chez l'homme est supérieure à 90 % (Tenailon et al., 2010), c'est pourquoi sa détection est utilisée comme un indicateur de pollution fécale. Toutefois, en proportion, cette bactérie ne fait pas partie des bactéries les plus représentées au sein du microbiote intestinal (Eckburg et al., 2005). En raison de sa facilité de culture, la bactérie *E. coli* est devenu un organisme d'étude de choix pour de nombreux domaines de la biologie.

### ***I. 2. Pathogénicité et classification***

Le groupe des *E. coli* contient de nombreuses souches dont la plupart sont commensales. Cela signifie qu'elles se nourrissent de l'interaction avec leur hôte sans lui porter préjudice. Au contraire, ces souches commensales permettent de protéger la barrière intestinale contre d'autres pathogènes bactériens (Hudault et al., 2001). Néanmoins, certaines souches d'*E. coli* sont des pathogènes pouvant entraîner diverses maladies responsables d'environ deux millions de décès par an (Russo & Johnson, 2003). Ces souches pathogènes sont classées en deux catégories : les InPEC et ExPEC qui provoquent respectivement des pathologies intra-intestinales (diarrhées hémorragiques, gastro-entérites...) et extra-intestinales (infections urinaires,

méningites, septicémies...) (**Table 1**). Bien que la bactérie *E. coli* soit une bactérie entérique, les ExPEC sont des souches opportunistes qui, au contact d'autres compartiments anatomiques, sont capables d'induire des infections qui peuvent être fatales. On distingue plusieurs variétés pathogènes (ou pathovars) au sein des souches InPEC et ExPEC en fonction des syndromes infectieux qu'elles provoquent et des facteurs de pathogénicité exprimés (Croxen & Finlay, 2010). Pour le groupe des *E. coli* provoquant des pathologies intestinales (InPEC), on retrouve les *E. coli* entérohémorragiques EHEC, les *E. coli* entéropathogènes EPEC, les *E. coli* entérotoxigènes ETEC, les *E. coli* entéroinvasives EIEC, les *E. coli* entéroaggrégatives EAEC et les *E. coli* à adhésion diffuse DAEC. En ce qui concerne les ExPEC, seuls deux pathovars ont été identifiés : les *E. coli* uropathogènes UPEC responsables des infections du tractus urinaire et les *E. coli* associées à la méningite et aux septicémies NMEC (Kaper et al., 2004).

**Table 1** : Récapitulatif des différents types de souches pathogènes d'*E. coli*

Souches pathogènes intra-intestinales (InPEC)		Souches pathogènes extra-intestinales (ExPEC)	
Souches entérohémorragiques	EHEC	Souches uropathogènes	UPEC
Souches entéropathogènes	EPEC	Souches associées à la méningite et aux septicémies	NMEC
Souches entérotoxigènes	ETEC		
Souches entéroinvasives	EIEC		
Souches entéroaggrégatives	EAEC		
Souches à adhésion diffuse	DAEC		

Toutefois, malgré cette classification, les souches pathogènes sont généralement identifiées en fonction des versions de trois antigènes qu'elles présentent à leur surface : l'antigène somatique O contenu au niveau des lipopolysaccharides de la paroi bactérienne, l'antigène flagellaire H qui correspond à une protéine de structure du flagelle et l'antigène capsulaire K de nature polysaccharidique. Le sérotype des souches d'*E. coli* est déterminé par la version de l'antigène O uniquement (ex : O157) alors que le sérotype est déterminé par la combinaison des versions des antigènes O et H (ex : O157:H7) voire parfois K (lorsqu'une capsule est présente).

La plasticité du génome d'*E. coli*, via la perte ou le gain d'éléments génétiques, joue un rôle important dans l'acquisition des caractères de virulence des souches pathogènes. En effet, des regroupements de plusieurs gènes de virulence appelés îlots de

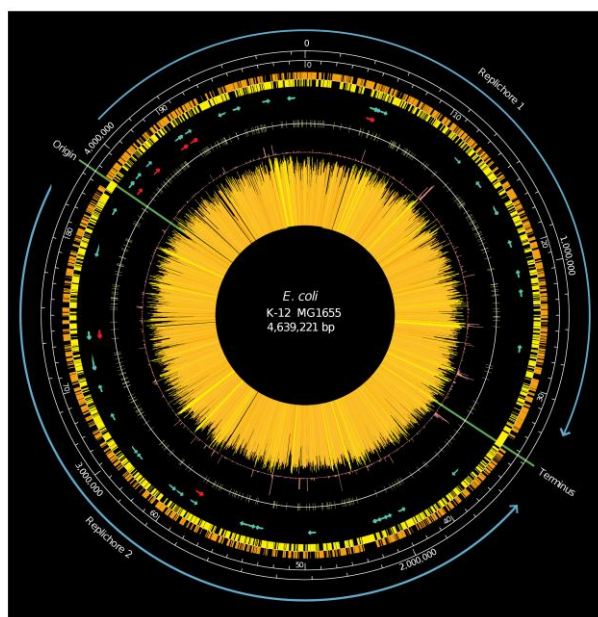
pathogénicité (de 10 à 200 Kb) sont retrouvés sur des plasmides ou intégrés dans le chromosome des souches pathogènes alors qu'ils sont généralement absents des souches non virulentes. Ainsi, les génomes des différentes souches d'*E. coli* peuvent être de tailles très différentes (entre 4.5 et 5.5 Mb) bien qu'elles disposent toutes d'un génome « cœur » composé d'environ 2200 gènes (Croxen & Finlay, 2010).

Enfin, en se basant sur les séquences génomiques entières, ou les séquences de certains gènes de plusieurs souches d'*E. coli* isolées chez l'humain (commensales et pathogènes), des analyses phylogénétiques ont mis en évidence quatre groupes phylogénétiques majeurs (A, B1, B2, D1) et d'autres groupes mineurs (C, D2, E) (Chaudhuri & Henderson, 2012). Ces groupes phylogénétiques diffèrent selon plusieurs paramètres parmi lesquels on retrouve l'histoire de leur évolution, les différentes niches écologiques occupées, la capacité à consommer différents substrats ou à résister différemment aux antibiotiques ou encore le taux de croissance. De plus, les souches ExPEC appartiennent généralement aux groupes B2 et D, les souches commensales aux groupes A et B1 alors que des InPEC se retrouvent à la fois dans les groupes A, B1 et D (Carlos et al., 2010).

### ***I. 3. La souche E. coli K-12 MG1655***

La souche *E. coli* K-12 MG1655 est une souche non pathogène très étudiée appartenant au groupe phylogénétique A qui provient de la souche K-12 ancestrale. Cette dernière a été initialement isolée en 1922 à partir des selles d'un patient convalescent d'une diphtérie à Palo Alto en Californie. En 1925, elle a été déposée dans la collection de souches du Département de Bactériologie à l'université de Stanford où elle a été nommée « K-12 ». Cette souche a ensuite été utilisée lors de multiples études et a progressivement perdu son antigène O (à cause de la mutation *rfb*<sup>-50</sup>), le bactériophage  $\lambda$  (irradiation aux ultraviolets) et le plasmide F (traitement à l'acridine orange) pour donner la souche MG1655 (Bachmann, 1996). La souche MG1655 présente d'autres différences génétiques par rapport à la souche ancestrale K-12 ( $\lambda^+$  F<sup>+</sup>). En effet, son génotype complet est F<sup>-</sup>  $\lambda^-$  *ilvG*<sup>-</sup> *rfb*<sup>-50</sup> *rph*<sup>-1</sup> et l'ensemble de ses mutations est bien caractérisé aujourd'hui. Bien que cette souche ait fait l'objet d'un très grand nombre de publications, Blattner et ses collaborateurs ont considéré qu'elle avait subi très peu de manipulations génétiques en comparaison d'autres souches. C'est la raison pour laquelle

c'est la première souche d'*E. coli* à avoir été séquencée (Blattner et al., 1997) (**Figure 1**). La taille de son chromosome est de 4.64 Mb et son pourcentage en GC est de 50.8 %. L'annotation du génome d'*E. coli* K-12 MG1655 a fait l'objet de nombreuses publications (Blattner et al., 1997; Riley et al., 2006; Karp et al., 2007) et est mise à jour régulièrement au travers de la base de données EcoCyc (Keseler et al., 2013). Le génome de cette souche contiendrait environ 4500 gènes dont 4300 coderaient pour des protéines. Parmi ces gènes, on estime qu'à peu près 300 d'entre eux sont essentiels (Baba et al., 2006).



**Figure 1 :** Structure globale du génome d'*E. coli* K-12 MG1655. D'après (Blattner et al., 1997)

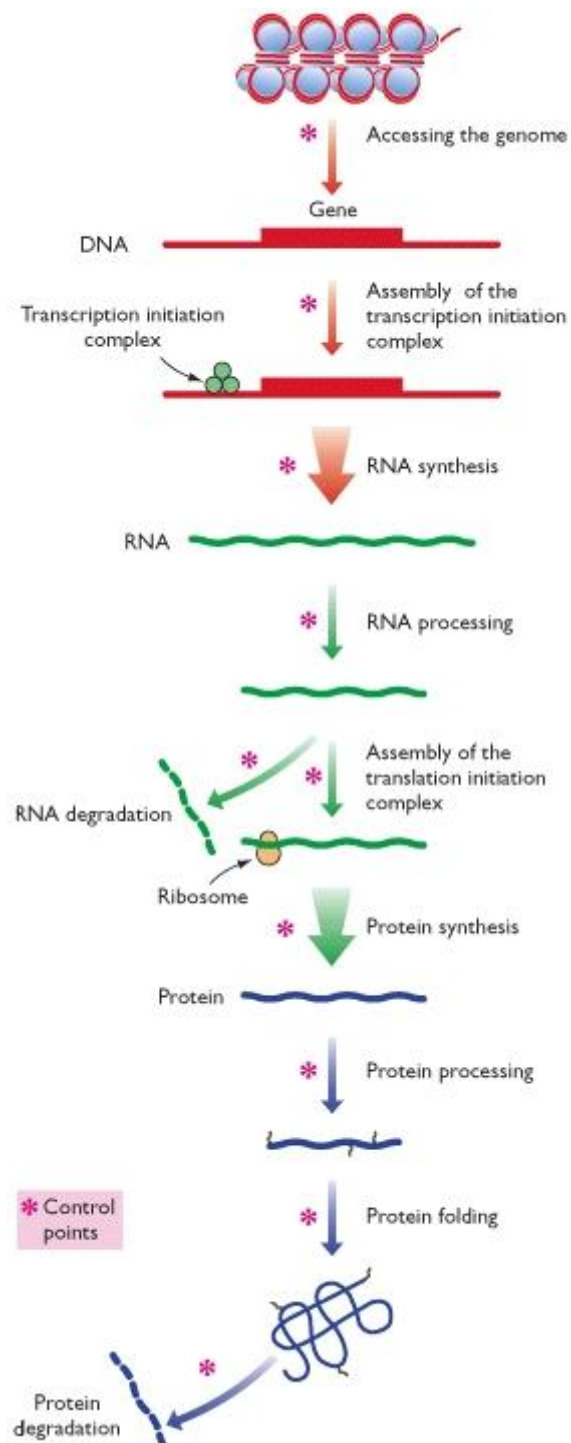
La caractérisation de son phénotype montre que cette souche se développe aussi bien en milieu riche qu'en milieu minimum à des taux de croissance assez rapides : environ  $1.6 \text{ h}^{-1}$  en milieu LB ( $g = 25 \text{ min}$ ) (Baev et al., 2006) et  $0.6 \text{ h}^{-1}$  en milieu minimum M9 supplémenté en glucose ( $g = 1 \text{ h}$ ) (Volkmer & Heinemann, 2011). La croissance de cette souche est également possible sur un grand nombre de substrats différents : sucres, acides aminés, acides faibles... (Yoon et al., 2012).

Finalement, la grande quantité d'informations disponibles sur la biologie de cette souche, son innocuité et sa capacité à se multiplier sur un grand nombre de substrats témoignant d'importantes capacités d'adaptation (Baev et al., 2006) nous ont mené à sélectionner la souche *E. coli* K-12 MG1655 pour notre étude.

## **II. Les différents niveaux de régulation de l'expression génique chez *Escherichia coli* en réponse à l'environnement**

### ***II. 1. S'adapter à l'environnement***

L'adaptation est définie comme un processus au cours duquel des changements anatomiques et/ou physiologiques permettent à un individu, à une lignée ou à une espèce d'augmenter ses chances de survie et de reproduction dans le milieu où il vit. En microbiologie, l'adaptation d'une bactérie met en jeu de multiples mécanismes cellulaires permettant à la cellule d'optimiser sa croissance ou sa survie dans certaines conditions nutritives, de température, de pH, d'oxygénation et/ou d'osmolarité données. Ainsi, les bactéries les mieux adaptées dans un contexte donné ont non seulement plus de chances de survivre, mais sont également plus susceptibles de remporter la compétition qui les oppose aux autres micro-organismes environnants. Pour s'adapter, les bactéries utilisent les mécanismes de régulation de l'expression génique qui consistent à optimiser l'expression ou l'activité de protéines en adéquation avec les conditions environnementales. Pour synthétiser une protéine, le gène codant pour cette dernière est transcrit en ARNm qui sera lui-même traduit en protéine. Cependant, toutes les bactéries ne possèdent pas les mêmes capacités d'adaptation du fait de leur patrimoine génétique mais les niveaux de régulation de l'expression génique restent relativement bien conservés d'une espèce à l'autre et intègrent des régulations transcriptionnelles, post-transcriptionnelles et post-traductionnelles (**Figure 2**) qui seront abordées dans cette partie pour la bactérie modèle *E. coli*.



**Figure 2 :** Les différents niveaux de régulation de l'expression des gènes. D'après (Brown, 2002).

## **II. 2. Les régulations transcriptionnelles**

La transcription correspond au processus qui permet à un gène d'être transcrit en ARN. Lors de l'adaptation, les régulations transcriptionnelles impliquent deux types de mécanismes (Berthoumieux et al., 2013). Le premier mécanisme correspond à la régulation globale de la transcription liée à l'état physiologique de la cellule (ex : le taux de croissance) (Klumpp et al., 2009). Le second mécanisme implique lui des régulateurs plus spécifiques et dépendants des conditions environnementales.

### **II. 2. 1. La régulation globale de la transcription**

Chez *E. coli*, la transcription de l'ADN en ARN est assurée par une seule ARN polymérase. Cette dernière est composée de l'enzyme cœur, constituée de deux sous-unités  $\alpha$ , une sous-unité  $\beta$  et une sous-unité  $\beta'$ , qui se lie à un des sept facteurs  $\sigma$  pour former l'holoenzyme (Burgess et al., 1969). En plus de permettre l'initiation de la transcription (Hinkle & Chamberlin, 1972), les facteurs  $\sigma$  confèrent à l'holoenzyme la capacité de reconnaître et de cibler des promoteurs spécifiques (Ishihama, 2000). La transcription se compose de trois étapes : l'initiation, l'élongation et la terminaison (Ghosh et al., 2010). La transcription est initiée par la fixation de l'holoenzyme sur le promoteur du gène au niveau de deux séquences relativement conservées situées environ à -35 et -10 nucléotides par rapport au point +1 de transcription correspondant au premier nucléotide de l'ARN. L'enzyme écarte ensuite la double hélice d'ADN au niveau de la région -10 pour former un complexe ouvert et commence de synthétiser l'ARN en associant les ribonucléosides sur le modèle de la matrice d'ADN. En quittant le promoteur, l'holoenzyme libère le facteur  $\sigma$  (Shimamoto et al., 1986) pour ne conserver que l'enzyme cœur et l'élongation de la transcription se poursuit jusqu'à rencontrer un signal de terminaison de transcription dépendant ou non de la protéine Rho (Nudler & Gottesman, 2002).

Le taux de croissance  $\mu$  (ou le temps de doublement  $g$  tel que  $g = \ln(2)/\mu$ ) reflète l'activité métabolique des cellules en réponse au contenu génétique de la souche bactérienne, à l'historique de l'inoculum et à l'environnement actuel. Ces trois critères déterminent également la composition macromoléculaire de la cellule (Bremer & Dennis, 1996; Ehrenberg et al., 2013). En conséquence, les variations de taux de



croissance impactent la transcription modifiant ainsi globalement l'expression des gènes.

Par exemple, plus le taux de croissance augmente, plus les fourches de réplication de l'ADN se multiplient au niveau du chromosome bactérien (Bremer & Dennis, 1996). De ce fait, la transcription des gènes localisés à proximité de l'origine de réplication est accentuée. C'est le cas par exemple pour des opérons ribosomiaux (Jin et al., 2012). D'autre part, il a été démontré que le nombre d'ARN polymérases par cellule ainsi que leur activité augmentent avec le taux de croissance (Ehrenberg et al., 2013). La vitesse de synthèse d'ARNm croît également, notamment pour les gènes dits « constitutifs » (c'est-à-dire non régulés par un facteur spécifique) (Berthoumieux et al., 2013; Gerosa et al., 2013; Klumpp et al., 2009). Cette augmentation est liée à l'accroissement du pourcentage d'ARN polymérase libre et fonctionnelle avec le taux de croissance (Bremer et al., 2003; Klumpp & Hwa, 2008; Klumpp et al., 2009).

Il existe cependant une compétition entre la transcription des ARN stables et celle des autres ARN. En effet, il a été observé que plus le taux de croissance augmente, plus la transcription des ARN stables est privilégiée par rapport à celle des autres ARN (Cabrera & Jin, 2006; Jin & Cabrera, 2006; Klumpp & Hwa, 2008). Ce phénomène de répartition de l'ARN polymérase entre ces différents types de gènes serait majoritairement contrôlé par le système de la réponse stringente impliquant l'alarmone ppGpp (Cashel et al., 1996). Ce système a d'ailleurs été suggéré comme étant le principal mécanisme de contrôle du taux de croissance (Potrykus et al., 2011). Le ppGpp est synthétisé à partir de GTP par les produits des gènes *relA* et *spoT* (ce dernier disposant également d'une fonction permettant d'hydrolyser le ppGpp). Il est produit en réponse à l'accumulation d'ARNt dé-acétylés, c'est-à-dire non chargés d'un acide aminé ainsi que lors de carences diverses en acides aminés, carbone, fer, phosphate... (Potrykus & Cashel, 2008). Plus intéressant, la concentration en ppGpp dans la cellule diminue alors que le taux de croissance augmente (Bremer & Dennis, 1996). En fait, l'accumulation de ppGpp diminue le ratio entre les ARN stables et les ARN totaux en inhibant la synthèse des ARN stables. Bien que tous les mécanismes de régulation par le ppGpp ne soient pas identifiés, il semble que ce dernier interagisse directement avec l'ARN polymérase de pair avec le co-régulateur DksA et inhibe globalement la transcription des gènes stables tout en favorisant l'expression de gènes spécifiques (Paul et al., 2004; Durfee et al., 2008; Potrykus & Cashel, 2008). Ainsi, en contrôlant la distribution de l'ARN polymérase entre

les gènes codant pour les ARN stables et le reste du génome en fonction du taux de croissance, le système de la réponse stringente régule globalement la transcription chez *E. coli*.

## II. 2. 2. Les régulations spécifiques de la transcription

Au-delà de la régulation globale de la transcription, il existe des mécanismes de régulations plus spécifiques qui répondent à des conditions environnementales données.

Comme évoqué dans la partie précédente, l'holoenzyme de l'ARN polymérase est notamment composée d'un des sept facteurs  $\sigma$  qui permettent à l'enzyme de reconnaître et d'interagir avec des promoteurs qui leur sont spécifiques. Ces sept facteurs  $\sigma$ , contrôlant l'expression de gènes spécifiques et exprimés selon certaines conditions sont :  $\sigma^{70}$  (ou  $\sigma^D$ , gène : *rpoD*),  $\sigma^{54}$  (ou  $\sigma^N$ , gène : *rpoN*),  $\sigma^{38}$  (ou  $\sigma^S$ , gène : *rpoS*),  $\sigma^{32}$  (ou  $\sigma^H$ , gène : *rpoH*),  $\sigma^{28}$  (ou  $\sigma^F$ , gène : *rpoF*),  $\sigma^{24}$  (ou  $\sigma^E$ , gène : *rpoE*) et  $\sigma^{18}$  (ou  $\sigma^{FecI}$ , gène : *fecI*) (Ishihama, 2000; Gruber & Gross, 2003). La liste des différents facteurs  $\sigma$  ainsi que les nombres et types de gènes qu'ils régulent sont récapitulés dans la **Figure 3**.

Sigma subunit	Genes under the control of each sigma
RpoD (613 aa)	Growth-related genes (~1,000)
RpoN (477 aa)	Nitrogen-regulated/stress response genes (~15)
RpoS (330 aa)	Stationary phase/stress response genes (~100)
RpoH (284 aa)	Heat shock/stress response genes (~40)
RpoF (239 aa)	Flagella-chemotaxis genes (~40)
RpoE (202 aa)	Extreme heat shock/extracytoplasmic genes (~5)
FecI (173 aa)	Ferric citrate transport/extracytoplasmic genes (~5)

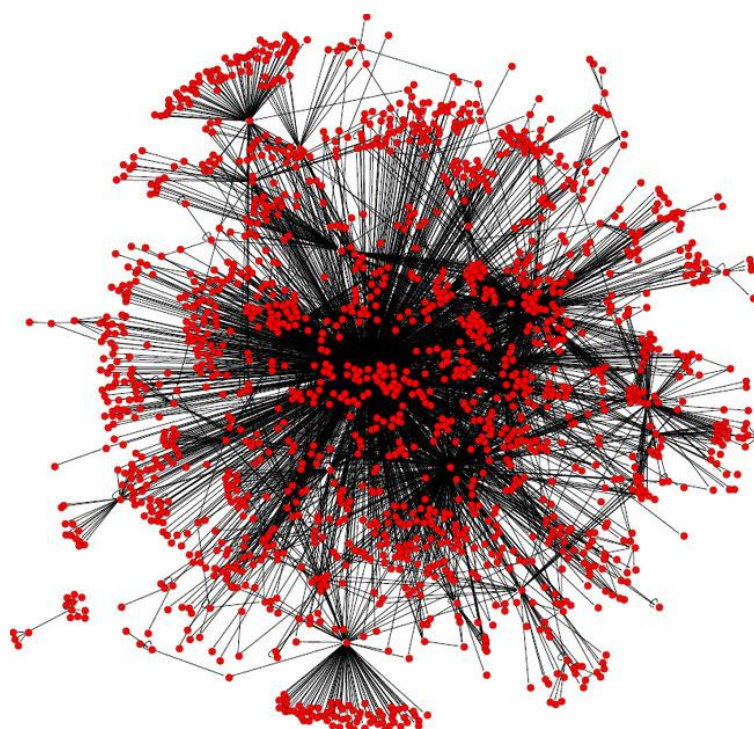
Figure 3 : Liste des facteurs  $\sigma$  chez *E. coli*. D'après (Ishihama, 2000).

A part le facteur  $\sigma^{70}$  dont la concentration cellulaire est la plus forte quelles que soient les conditions de culture, l'expression des autres facteurs  $\sigma$  alternatifs répond généralement à des stimuli particuliers (Jishage et al., 1996). D'autre part, la régulation de l'expression génique par les facteurs  $\sigma$  dépend également de leur affinité pour

l'enzyme cœur (Maeda et al., 2000) dans la mesure où sa quantité de forme libre est limitante et inférieure à la quantité de tous les facteurs  $\sigma$  réunis. Même si le facteur  $\sigma^{70}$  est le plus concentré et afin pour l'enzyme cœur, la bactérie a développé plusieurs stratégies pour permettre aux autres facteurs  $\sigma$  d'accéder à l'enzyme cœur dans certaines conditions. Par exemple, le ppGpp et DksA, la protéine Rsd ou encore l'ARN 6S favorisent l'expression de promoteurs dépendants de facteurs  $\sigma$  alternatifs en diminuant l'interaction de  $\sigma^{70}$  avec l'enzyme cœur ou en piégeant l'holoenzyme  $\sigma^{70}$  (Sharma & Chatterji, 2010; Österberg et al., 2011). L'activité des facteurs  $\sigma$  est également régulée par des facteurs anti- $\sigma$  spécifiques qui sont capables de séquestrer et/ou d'entraîner la dégradation des sous-unités  $\sigma$  dans des conditions particulières (Österberg et al., 2011; Trevino-Quintanilla et al., 2013). Enfin, l'expression même des gènes codant pour les facteurs  $\sigma$  est finement régulée par des mécanismes transcriptionnels et post-transcriptionnels (Ishihama, 2000).

L'intervention de facteurs de transcription correspond à un autre des principaux mécanismes de régulation spécifique de la transcription. Les facteurs de transcription sont des protéines qui activent et/ou répriment généralement l'initiation de la transcription d'un gène sans appartenir à l'ARN polymérase par des mécanismes très divers (Zhou & Yang, 2006). Leur fonctionnement nécessite une interaction de leur domaine de liaison à l'ADN avec une séquence spécifique en *cis* du promoteur du gène régulé. Tout comme pour les facteurs  $\sigma$ , l'activité des facteurs de transcription dépend de signaux environnementaux ou de signaux internes reflétant un changement intra/extracellulaire (Martínez-Antonio, 2011). Il existe plusieurs mécanismes de régulation de l'action des facteurs de transcription par ces changements environnementaux : (i) de petits ligands peuvent modifier l'affinité de leur domaine de liaison à l'ADN, (ii) leur activité peut être modulée par des modifications covalentes comme dans le cas des systèmes à double composante où la protéine détectrice peut, en réponse à un stimulus, activer ou inactiver son facteur de réponse associé (Hoch, 2000), (iii) et enfin l'expression de ces facteurs de transcription peut être elle-même régulée (Browning & Busby, 2004). À ce jour, chez *E. coli* K-12, le site RegulonDB (<http://regulondb.ccg.unam.mx/>) recense pas moins de 200 facteurs de transcription mis en évidence expérimentalement (Salgado et al., 2013) et en prédit environ 160 supplémentaires (Pérez-Rueda & Collado-Vides, 2000, 2001). Au total, 35 % des facteurs de transcription seraient des activateurs, 43 % des répresseurs et 22 % auraient les

deux activités. Parmi ces facteurs, il existe des régulateurs globaux (par exemple : Crp, Fnr, IHF, Fis, ArcA ou Lrp) qui contrôlent environ 50 % de tous les gènes régulés chez *E. coli* alors qu'environ 60 facteurs de transcription ne contrôlent qu'un seul et unique gène (Martínez-Antonio, 2011). Ces régulateurs globaux contrôlent généralement aussi d'autres facteurs de transcription. Par ailleurs, si un facteur de transcription peut réguler plusieurs gènes, un unique gène peut également être contrôlé par plusieurs facteurs de transcription. Ainsi le réseau de régulation auquel appartiennent les facteurs transcriptionnels est extrêmement complexe (**Figure 4**).



**Figure 4** : Schéma du réseau de régulations transcriptionnelles chez *E. coli*. Les nœuds correspondant aux points rouges sont les gènes et les liaisons noires correspondent aux interactions. D'après (Freyre-Gonzalez & Trevino-Quintanilla, 2010).

Finalement, parmi ces facteurs de transcription, certains d'entre eux sont des protéines associées au nucléoïde (ou NAP). Le nucléoïde correspond à la région cytoplasmique où se situe le chromosome bactérien. Les NAPs sont très probablement responsables de l'organisation structurale et compactée du chromosome même si les mécanismes exacts ne sont pas encore bien connus (Dame, 2005). Chez *E. coli*, on trouve par exemple Fis, IHF, H-NS, HU, StpA et Dps (Azam & Ishihama, 1999). L'expression de ces protéines varie en fonction des conditions de croissance et les changements structuraux qu'elles induisent au niveau du chromosome peuvent entraîner l'activation

ou la répression de certains promoteurs (McLeod & Johnson, 2001; Dillon & Dorman, 2010).

Il est important de conserver à l'esprit que les promoteurs des gènes sont sensibles à la fois aux régulations globales et spécifiques de la transcription. La détermination des poids respectifs de ces deux types de régulations lors de changements environnementaux a fait l'objet d'études récentes utilisant des approches de biologie des systèmes (Berthoumieux et al., 2013; Gerosa et al., 2013; Keren et al., 2013). Ces études ont mesuré l'activité de promoteurs « constitutifs » à l'aide de gènes rapporteurs pour évaluer l'évolution des régulations globales de la transcription dans différentes conditions de croissance et ainsi déterminer la part des régulations spécifiques. Ces travaux concluent généralement que la régulation globale de la transcription, dépendante de la physiologie cellulaire, serait prépondérante par rapport aux régulations spécifiques. Ces dernières resteraient néanmoins nécessaires pour compléter ou moduler les capacités globales des promoteurs.

### ***II. 3. Les régulations post-transcriptionnelles***

Après avoir décrit les régulations de la transcription, les régulations post-transcriptionnelles en réponse à un changement d'environnement seront abordées dans cette partie.

#### **II. 3. 1. Les régulations de la stabilité des ARN**

Au sein d'une cellule, la concentration d'un ARN est contrôlée par trois mécanismes. Le premier correspond à la transcription dont les régulations lors de l'adaptation ont été abordées précédemment. Les second et troisième mécanismes correspondent respectivement à la dégradation de ces ARN ainsi qu'à leur dilution lors de la division cellulaire. Le phénomène de dilution, de l'ordre de l'heure, est généralement négligeable devant la forte vitesse de dégradation des ARNm qui présente des cinétiques de l'ordre de la minute. Au contraire, les ARN dits « stables » (correspondant aux ARNr et ARNt) présentent une très faible dégradation. La stabilité/dégradation des ARN est mesurée au travers de leur temps de demi-vie qui correspond à la durée au bout de laquelle la concentration initiale en messager a été divisée par deux.

Initialement imaginé comme un simple mécanisme nécessaire pour recycler le pool de ribonucléotides, la dégradation des ARNm est maintenant considérée comme un processus régulant l'expression des gènes lors de l'adaptation. En effet plusieurs travaux ont mis en évidence que la stabilité de certains ARNm variait en fonction des conditions environnementales. Par exemple, la stabilité de l'ARNm de *ompA* diminue avec le taux de croissance (Nilsson et al., 1984; Vytvytska et al., 1998) alors que celle de *fumA* augmente (Lin et al., 2012). Par contre, le temps de demi-vie des transcrits *lpp* et *bla* ne varient pas avec le taux de croissance. Ces résultats montrent que la dégradation des messagers en fonction du taux de croissance n'est pas un mécanisme universel. Il a été démontré que le stress induit par la phase stationnaire peut également influencer la régulation de la stabilité de certains messagers (Albertson & Nyström, 1994; Kuzj et al., 1998) tout comme la croissance en anaérobie (Georgellis et al., 1993). D'autre part, alors que l'ARNm du gène *cspA* est extrêmement instable à 37°C ( $t_{1/2} < 12$  sec), sa stabilité augmente fortement lors d'une diminution de la température à 15°C ( $t_{1/2} > 20$  min) (Brandi et al., 1996; Goldenberg et al., 1996; Gualerzi et al., 2003). Ces données suggèrent donc que la stabilité des ARNm peut également varier en fonction de la température. Par ailleurs, la dégradation des messagers joue aussi un rôle important pour l'expression différentielle des produits d'un opéron comme par exemple pour *malEFG* (Newbury et al., 1987). La plupart de ces observations chez *E. coli* sont également retrouvées chez d'autres micro-organismes (Takayama & Kjelleberg, 2000).

Malgré l'avènement des technologies omiques, assez peu d'études se sont intéressées à la stabilité de l'ensemble des ARNm d'une cellule (aussi appelé « Stabilome ») et de surcroît à son évolution en fonction des conditions de croissance. En conséquence, très peu de conclusions générales sur le rôle des régulations de la stabilité des messagers ont pu être dégagées. En 1970, Pato et ses collaborateurs reportent la tendance que la stabilité des ARNm pourrait augmenter alors que le taux de croissance diminue (Pato & von Meyenburg, 1970). Au contraire, la première étude à l'échelle omique chez *E. coli* indique que la stabilité moyenne des ARNm n'est pas différente lorsque les cellules se développent en milieu riche ou pauvre à 30°C suggérant que le taux de croissance n'impacte pas la dégradation des transcrits (Bernstein et al., 2002). Cependant, d'autres études à l'échelle omique chez *Lactococcus lactis* et *Mycobacterium tuberculosis* indiquent le contraire (Dressaire et al., 2013; Rustad et al., 2013). Concernant des conditions de stress, à notre connaissance, aucune étude omique comparative sur la

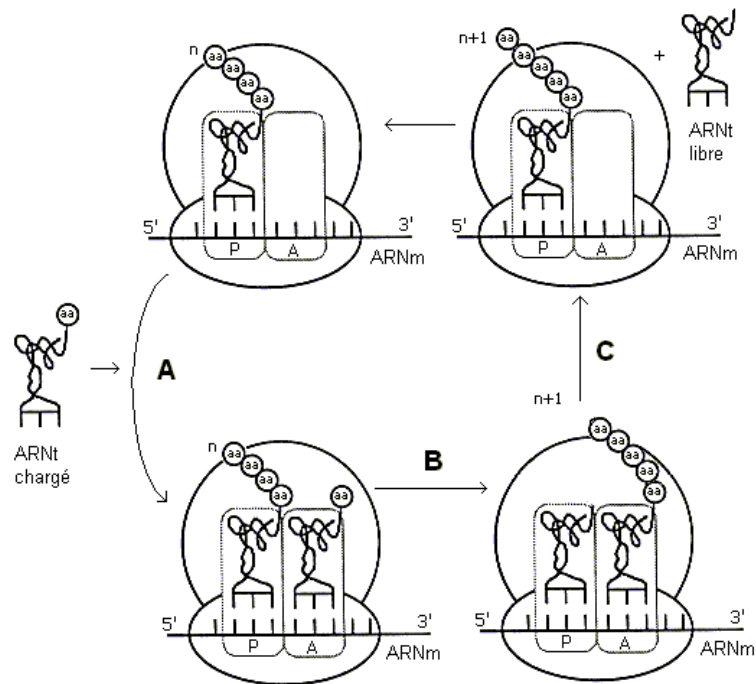
stabilité des ARNm chez *E. coli* n'a été publiée. Par contre, les ARNm chez *Lactococcus lactis* sont globalement stabilisés lors d'une carence en carbone et lorsque le taux de croissance diminue (Redon et al., 2005; Dressaire et al., 2013). D'autre part, la faible oxygénation ou la diminution de température du milieu de culture entraînent une stabilisation globale des ARNm chez *Mycobacterium tuberculosis* (Rustad et al., 2013). Globalement, ces études suggèrent une stabilisation des ARNm en cas de stress (température, carences nutritives...) qui permettrait de maintenir un « pool » minimal de transcrits, utilisables rapidement par les ribosomes en cas de conditions de croissance plus favorables.

Les mécanismes de dégradation des ARNm chez *E. coli* ainsi que leurs régulations seront développés plus en détail dans la partie III de cette étude bibliographique.

### **II. 3. 2. Les régulations de la traduction**

Parmi les mécanismes de régulations post-transcriptionnelles, les régulations de la traduction permettent d'ajuster la production de protéine à partir d'une concentration donnée en ARNm dans la cellule.

La traduction des ARNm en protéines est réalisée par les ribosomes. Les ribosomes sont constitués de deux sous-unités de tailles différentes qui correspondent à des complexes ribonucléoprotéiques mêlant ARNr et protéines. La plus grande des sous-unités se compose de deux ARNr (23S et 5S) associés à 33 protéines pour un coefficient de sédimentation de 50S. De son côté, la petite sous-unité, dont le coefficient de sédimentation est égal à 30S, est constituée de l'ARNr 16S associé à 21 protéines. Entier, le ribosome possède un coefficient de sédimentation de 70S (Kaczanowska & Rydén-Aulin, 2007). Brièvement, l'initiation de la traduction implique la reconnaissance d'une séquence spécifique de l'ARNm, appelée séquence Shine-Dalgarno correspondant au RBS. Cette séquence située à environ 10 nucléotides en amont du codon d'initiation de la traduction (ou codon start généralement AUG) est complémentaire de l'extrémité 3' de l'ARNr 16S de la petite sous-unité 30S du ribosome (Shine & Dalgarno, 1974; Chen et al., 1994). Cette complémentarité permet à la sous-unité 30S de reconnaître la séquence Shine-Dalgarno et de bien positionner le ribosome sur l'ARNm avec l'aide de facteurs d'initiation. Une fois la petite sous-unité positionnée, la grande sous-unité s'y associe pour former le ribosome. S'en suit l'étape d'élongation schématisée dans la **Figure 5**.



**Figure 5 :** Schéma de l'élongation de la traduction. Seuls les sites A et P sont représentés.

Le ribosome est subdivisé en trois sites dont les fonctions sont : d'interagir avec l'ARNt chargé de l'acide aminé spécifique du codon en cours de lecture (site A pour « Acceptor » - Réaction A), de catalyser la réaction peptidyl-transférase qui lie cet acide aminé au peptide en cours de synthèse (site P pour « Peptidyl » - Réaction B) et de permettre la libération de l'ARNt dé-acétylé (site E pour « Exit »). Ainsi, en se déplaçant dans le sens 5' vers 3' de l'ARNm, le ribosome permet au peptide naissant de basculer du site A au site P et à l'ARNt dé-acétylé de passer du site P au site E entraînant sa libération (Réaction C). En répétant autant de fois cette étape que le nombre de codons de l'ARNm en cours de traduction, le ribosome synthétise la protéine sur le modèle de la séquence codante du transcrit jusqu'à la terminaison de la traduction dont le signal est un codon stop.

Une fois de plus, en impactant le contenu macromoléculaire de la cellule, le taux de croissance va réguler le niveau de traduction. En effet, il a été démontré que la quantité de ribosomes augmente avec l'augmentation du taux de croissance (Bremer & Dennis, 1996; Ehrenberg et al., 2013). Ceci est notamment lié à la diminution de la concentration en ppGpp qui inhibe la synthèse des ARNr et des messagers de protéines ribosomales comme expliqué précédemment (II.2.1). A l'opposé, en présence d'un stress lié à une carence, le phénomène inverse est observable (Srivatsan & Wang, 2008). De plus, une



étude montre que le ppGpp peut directement inhiber la traduction en bloquant l'action d'un des facteurs d'initiation suggérant un double niveau de régulation de la machinerie de traduction par ce métabolite (Milon et al., 2006).

D'autre part, les conditions de stress et la phase stationnaire ont également un effet sur la traduction via le mécanisme d'hibernation des ribosomes. Dans ces conditions, les complexes ribosomiaux de 70S se dimérisent en présence des facteurs RMF et HPF pour former des particules de 100S dépourvues d'activité traductionnelle en raison du blocage de l'activité peptidyl-transférase (Wada, 1998; Yoshida et al., 2004). Toutefois, la formation de ces structures est réversible. Lorsque les cellules sont transférées dans un milieu non carencé, les facteurs RMF et HPF sont rapidement libérés ainsi que les complexes 70S qui redeviennent opérationnels. Ce phénomène permet à la cellule de remobiliser rapidement sa machine de traduction dans des conditions favorables évitant de devoir re-synthétiser *de novo* les ribosomes.

Un autre mécanisme contrôlant le processus de traduction en réponse à l'environnement correspond aux protéines se liant à l'ARN. Par exemple, le régulateur post-transcriptionnel CsrA est notamment connu pour inhiber, lors de la phase exponentielle de croissance, la traduction de gènes exprimés en phase stationnaire (Baker et al., 2002; Romeo et al., 2012). Le fonctionnement du système Csr dans son ensemble sera détaillé dans la partie IV de cette étude bibliographique.

L'évolution de la structure secondaire d'un ARNm en fonction de la température peut également moduler sa traduction. En effet, dans le cas du gène *rpoH*, il a été démontré qu'à 30°C, les structures secondaires que forme son ARNm bloquent l'initiation de sa traduction. En revanche, à 42°C, la déstabilisation de cette structure entraîne une augmentation de l'expression de la protéine (Morita et al., 1999a, 1999b). De plus, à faible température, il a été prouvé que la protéine CspA se fixe sur les régions 5' non traduites de certains ARNm et inhibe la formation de structures secondaires. Dans ce cas, les auteurs suggèrent que la protéine CspA favoriserait la traduction de ses cibles à faible température (Jiang et al., 1997).

La traduction peut également être modulée par les petits ARN non codants (Gottesman, 2004). Ces derniers, synthétisés en réponse à un stress, s'hybrident généralement sur un ARNm cible au niveau de la région d'initiation de la traduction. Cette liaison peut soit inhiber soit favoriser l'initiation de la traduction régulant ainsi l'expression du gène cible. Par exemple, l'ARNnc OxyS, exprimé en réponse à un stress

oxydatif, peut inhiber la traduction du gène *fhIA* (codant pour un facteur de transcription) en bloquant l'accès des ribosomes au niveau de la séquence Shine-Dalgarno de son ARNm (Altuvia et al., 1998). A l'inverse, l'ARNnc DsrA est synthétisé lors d'une phase exponentielle de croissance à faible température (Repoila & Gottesman, 2001). Il peut alors s'hybrider avec le 5'UTR de l'ARNm du gène *rpoS* et favoriser sa traduction en levant la répression traductionnelle exercée par une structure secondaire préexistante (Majdalani et al., 1998).

Enfin, le mécanisme de riboswitch est également un moyen pour la cellule de contrôler la traduction lors de changements environnementaux. Il s'agit de réarrangements allostériques de l'ARNm qui s'articulent autour d'un ligand de faible poids moléculaire dont la concentration est déterminante pour la formation de structures secondaires (Vitreschak et al., 2004). Par exemple, la liaison de l'adénosylcobalamine (dérivé de la vitamine B<sub>12</sub>) sur le 5'UTR de l'ARNm de *btuB* (impliqué dans l'import de la vitamine B<sub>12</sub>) inhibe la traduction du messager en bloquant l'interaction des ribosomes avec le transcrit via la formation d'une tige-boucle (Nou & Kadner, 2000).

Si les mécanismes de régulation de la stabilité des ARNm et d'initiation de la traduction ont été séparés au cours de cet exposé, ces deux processus restent néanmoins extrêmement liés. En effet, l'inhibition de la traduction d'un ARNm mène souvent à sa déstabilisation par un mécanisme qui sera abordé dans la partie III. D'autre part, la stabilité influence évidemment la traduction puisqu'elle détermine le niveau de transcrits dans la cellule (Kaberdin & Bläsi, 2006).

## ***II. 4. Les régulations post-traductionnelles***

Une fois l'ARNm traduit en protéine, la cellule dispose encore de processus lui permettant de réguler non seulement la concentration de la protéine mais également sa disponibilité ou son activité en fonction de l'environnement. Ces types de régulations sont appelés régulations post-traductionnelles et seront détaillés dans la partie suivante.

### **II. 4. 1. Les régulations de la stabilité des protéines**

La régulation de la stabilité des protéines est un mécanisme qui permet de moduler la concentration des protéines dans la cellule. Elle fait intervenir des complexes protéiques généralement composés d'une protéine à activité ATP dépendante

appartenant à la famille des ATPases AAA (ex : ClpA, ClpX et ClpY) avec une protéase à activité ATP indépendante (ex : ClpP et ClpQ). L'ATPase AAA a pour rôle de déplier la protéine substrat via des changements structuraux faisant suite à l'hydrolyse d'ATP (Sauer et al., 2004). Puis, elle adresse la protéine déstructurée à la protéase qui la dégradera progressivement en acides aminés (Chandu & Nandi, 2004). Seules les associations ClpAP, ClpXP et ClpYQ semblent exister chez *E. coli* (Picard et al., 2009). La dégradation des protéines peut également faire intervenir des protéines « tout en un » comme Lon et FtsH qui possèdent à la fois les domaines ATPasique et protéolytique (Gottesman, 2003). Le choix des protéines à dégrader est gouverné par plusieurs mécanismes : (i) présence de séquences spécifiques au niveau des extrémités N-terminale ou C-terminale de la protéine, (ii) présence d'une étiquette SsrA ajoutée par un ARNtm sur un peptide naissant lorsque la traduction rencontre une difficulté (ex : ARNm endommagé) ou (iii) action d'une protéine modulatrice/adaptatrice permettant de « diriger » les protéines à recycler vers les protéases (ex : RssB pour l'adressage du facteur  $\sigma^S$  à la protéase ClpXP).

En condition de stress ou de changements environnementaux, la dégradation protéique permet de réguler rapidement la concentration d'une protéine en la stabilisant ou en la déstabilisant selon les conditions. Par exemple, l'état d'activation de la protéine adaptatrice RssB en réponse à un stress détermine son affinité pour le facteur  $\sigma^S$  régulant ainsi la dégradation de ce dernier par la protéase ClpXP (Hengge, 2009a). De la même façon, l'activité du facteur  $\sigma^E$  est augmentée lors de l'altération de l'intégrité de la membrane cellulaire grâce à une cascade protéolytique qui aboutit à la dégradation du facteur anti- $\sigma$  RseA qui séquestre  $\sigma^E$  dans les conditions normales de croissance (Ades, 2008).

D'autre part, la protéolyse impacte également les systèmes toxine-antitoxine de type II en fonction des conditions de croissance. Ce système consiste en un opéron codant généralement pour deux protéines liées : la toxine correspondant à un « poison » et l'antitoxine qui correspond à son « antidote ». En conditions normales de croissance, l'antitoxine est généralement en excès par rapport à la toxine menant à l'inactivité de cette dernière. Si dans la plupart des cas la toxine est stable, l'antitoxine sous forme libre est une protéine instable ciblée particulièrement par des protéases. En revanche, lors de son interaction avec sa toxine associée, l'antitoxine adopte une conformation

structurelle qui la rend stable. Lors de conditions particulières de stress, la suractivité des protéases va entraîner une diminution du pool d'antitoxine libre qui va modifier l'équilibre toxine-antitoxine et libérer la toxine (Van Melderen & Aertsen, 2009; Brzozowska & Zielenkiewicz, 2013).

#### **II. 4. 2. Les régulations de la disponibilité ou de l'activité protéique**

Bien que la concentration d'une protéine dans la cellule soit évidemment importante pour refléter son rôle *in vivo*, il est important de garder à l'esprit que toutes les protéines présentes dans le cytoplasme de la cellule à un instant *t* ne sont pas disponibles ou actives en fonction des conditions environnementales.

Par exemple, les interactions protéine-protéine jouent un rôle important dans la régulation de la disponibilité protéique. Les systèmes toxine-antitoxine de type II décrits dans la partie II.4.1 en sont le parfait exemple puisque le niveau de toxine libre est contrôlé via l'interaction toxine-antitoxine qui est dépendante de conditions environnementales ou de stress donnés (Brzozowska & Zielenkiewicz, 2013). Parmi les interactions protéine-protéine, on retrouve également les protéines chaperonnes dont les plus étudiées sont TF, DnaK et GroEL/ES. Ces protéines, dont beaucoup correspondent à des protéines HSP exprimées lors d'un choc thermique, préviennent les mauvais repliements des protéines et leur agrégation. Elles ont un rôle essentiel dans les mécanismes de repliement protéiques *de novo*, dans le maintien du repliement protéique en cas de stress, dans l'assemblage et le transport de polypeptides naissants ainsi que dans le processus de dégradation protéique (Hartl & Hayer-Hartl, 2009). Ainsi, elles permettent à certaines protéines de rester disponibles ou de conserver leur activité catalytique malgré des conditions environnementales qui entraîneraient leur déstructuration.

Si les interactions protéine-protéine contrôlent la disponibilité de certaines protéines, il a également été démontré que d'autres molécules peuvent en faire autant. En effet certains ARN sont capables de piéger des protéines et de les empêcher d'exercer leur fonction en « mimant » leurs cibles (Marzi & Romby, 2012). C'est par exemple le cas de l'ARN 6S, qui s'accumule en phase stationnaire et dont la structure secondaire « mime » la conformation ouverte d'une région promotrice lors de l'initiation de la transcription. Cette structure lui permet alors de séquestrer l'holoenzyme  $\sigma^{70}$  et l'empêche d'activer la transcription de gènes de ménage durant la phase stationnaire

(Sharma & Chatterji, 2010). Le système Csr, décrit dans la partie IV de cet exposé, illustre lui aussi un autre exemple de ce type de régulation avec les ARNnc CsrB et CsrC.

L'activité catalytique de certaines protéines peut également être régulée à travers des modifications post-traductionnelles comme des acétylations ou des phosphorylations dépendantes des conditions environnementales. L'acétylation des protéines au niveau de lysines toucherait une centaine de protéines dont un grand nombre d'enzymes métaboliques (Zhang et al., 2009). Par exemple, l'activité catalytique de l'acétyl-coA synthétase (Acs codée par *acs*), impliquée dans la consommation de l'acétate, est favorisée par sa désacétylation par CobB alors qu'au contraire, son acétylation par PatZ la réduit (Castaño-Cerezo et al., 2011). D'autre part, la mise en place du shunt glyoxylate lors de la croissance d'*E. coli* en présence d'acétate, dépend par exemple de l'inactivation de l'isocitrate déshydrogénase (Icd codée par *icd*) impliquant une phosphorylation par l'isocitrate déshydrogénase kinase (AceK codée par *aceK*) (Cozzone & El-Mansi, 2005). En effet, la phosphorylation de Icd inhibe l'activité de l'enzyme en empêchant la liaison de l'isocitrate au niveau de son site catalytique.

Enfin, les régulations allostériques permettent également de moduler l'activité catalytique des protéines. Au contraire de la phosphorylation et de l'acétylation qui correspondent à des modifications covalentes, les régulations allostériques sont des mécanismes au cours desquels une molécule effectrice se lie de façon non covalente à la protéine au niveau d'un site spécifique. Cette liaison modifie la conformation de la protéine de telle sorte que cela affecte l'activité catalytique protéique. Pour conserver l'exemple du shunt glyoxylate, la protéine AceK détient en fait l'activité kinase mais également l'activité phosphatase permettant de déphosphoryler Icd. Il a été démontré qu'un changement de structure de AceK lié à la liaison allostérique d'une molécule d'AMP favorise l'activité phosphatase de l'enzyme et inhibe son activité kinase (Zheng & Jia, 2010).

Ainsi, il est généralement admis que les différents niveaux de régulation de l'expression génique (transcriptionnels, post-transcriptionnels et post-traductionnels) sont impliqués selon les conditions environnementales pour optimiser l'adaptation de la bactérie afin de favoriser sa croissance et/ou sa survie. Néanmoins, des études multi-échelles sont nécessaires pour comprendre comment s'articulent ces différents niveaux de régulation les uns par rapport aux autres (synergie ? antagonisme ?).

### **III. Les mécanismes de dégradation des ARNm chez *Escherichia coli***

La dégradation des ARNm est un processus important pour réguler la concentration des transcrits dans la cellule. Comme décrit précédemment, il appartient aux multiples niveaux de régulation de l'expression génique qui permettent à la bactérie de s'adapter aux conditions environnementales. En effet, selon ces conditions, des variations de stabilité des ARNm ont été identifiées (voir partie II.3.1). Dans cette partie, les acteurs, les mécanismes et les régulations de la dégradation des ARNm chez *E. coli* seront détaillés.

#### ***III. 1. Les principales ribonucléases***

La dégradation des ARNm chez *E. coli* impliquent des enzymes appelées ribonucléases. Selon leur activité, elles sont regroupées comme des endoribonucléases ou des exoribonucléases. Les endoribonucléases sont capables d'induire un clivage intramoléculaire au sein même des séquences d'ARN au niveau des liaisons phosphodiester. Les exonucléases sont des enzymes qui dégradent les ARN par leur extrémité en libérant des mononucléotides. Chez *E. coli*, seules des exonucléases 3'-5' ont pu être identifiées (Zuo & Deutscher, 2001). Au total, on distingue 21 RNases chez *E. coli* K-12 cependant toutes ne sont pas impliquées dans la dégradation des ARNm et l'implication de certaines est très limitée (Mackie, 2013). Seules les enzymes majeures de la dégradation des ARNm seront détaillées dans cette partie.

##### **III. 1. 1. Les endoribonucléases**

L'endoribonucléase la plus importante pour la dégradation des ARNm est la RNase E, codée par le gène essentiel *rne* (Carpousis et al., 2009). La RNase E clive les séquences d'ARN simple brin et son action globale sur les ARNm a été démontrée grâce au mutant thermolabile *rne-1* qui présente une stabilisation générale des messagers à une température non permissive (Babitzke & Kushner, 1991). Elle participe également à la maturation des ARNr 5S, 16S, ARNtm et ARNt (Roy et al., 1983; Li et al., 1999; Lin-Chao et al., 1999; Li & Deutscher, 2002) ainsi qu'à la maturation de précurseurs de certains ARNm polycistroniques (Nilsson & Uhlin, 1991). La RNase E est une protéine de 1061 acides aminés possédant deux domaines fonctionnels distincts. En effet, l'extrémité

N-terminale de la protéine (1 – 529) forme une structure globulaire qui correspond au siège de l'activité catalytique de la RNase alors que la partie C-terminale (530 – 1061) plutôt non structurée sert de support pour la formation d'un complexe protéique appelé dégradosome qui sera décrit dans la partie III.3 (Callaghan et al., 2004, 2005). La RNase E est une protéine localisée à la membrane (Khemici et al., 2008) qui s'organise en tétramère par l'association de deux dimères (Mackie, 2013). Il est clair aujourd'hui que l'état de phosphorylation de l'extrémité 5' des ARNm affecte l'activité de la RNase E. En effet, l'enzyme clive préférentiellement les ARNm dont l'extrémité 5' est monophosphorylée par rapport aux messagers tri-phosphorylés (Mackie, 1998). Ainsi la protéine RppH, codée par le gène *rppH*, a une grande influence sur la stabilité des ARNm (Deana et al., 2008). En effet, il a été démontré que cette protéine, en déphosphorylant les ARNm triphosphates en ARNm monophosphates, favorisait le clivage des messagers par la RNase E.

Parmi les endoribonucléases ayant également un rôle dans la dégradation des ARNm, on trouve la RNase III, codée par le gène *rnc*, qui clive spécifiquement les doubles brins d'ARN (Nicholson, 1999). Cette enzyme, impliquée lors de la maturation des ARNr, intervient également dans le clivage de certains messagers ainsi que des régions intergéniques de plusieurs ARNm polycistroniques (Shen et al., 1982; Portier et al., 1987; Lim et al., 2014). Elle favorise la dégradation des transcrits disposant de structures secondaires qui empêchent l'action immédiate de la RNase E. Cependant, une étude utilisant un mutant de la RNase III montre qu'elle n'est absolument pas indispensable pour la viabilité cellulaire, la maturation des ARNr et la dégradation générale des ARNm (Babitzke et al., 1993).

Enfin, certains ARNm peuvent être spécifiquement clivés par l'endoribonucléase RNase G, codée par le gène *rng*. C'est le cas par exemple des ARNm des gènes *adeH* et *eno* qui s'accumulent dans une souche délétée pour *rng* (Umitsuki et al., 2001; Kaga et al., 2002). Cette enzyme non essentielle est considérée comme un paralogue de la RNase E en raison d'une très forte homologie avec son domaine catalytique globulaire (Okada et al., 1994). Comme la RNase E, la RNase G est, elle aussi, plus sensible à la présence d'une extrémité 5' monophosphate et est impliquée dans la maturation de l'ARNr 16S (Li et al., 1999; Jiang & Belasco, 2004). Toutefois, elle ne dispose pas de domaine C-terminal comme son paralogue. Il a été démontré qu'une surexpression très importante de *rng* permet de restaurer partiellement la viabilité d'une souche délétée pour *rne* (Lee et al.,

2002). Toutes ces observations indiquent que la RNase G exerce des fonctions redondantes à la RNase E bien que son rôle soit beaucoup plus limité.

### **III. 1. 2. Les exonucléases**

Parmi les exoribonucléases identifiées chez *E. coli*, la RNase II, la PNPase et la RNase R, codées respectivement par les gènes *rnb*, *pnp* et *rnr* sont les principales exonucléases impliquées dans la dégradation des ARNm (Régnier & Arraiano, 2000; Cheng & Deutscher, 2005). Ces enzymes sont des exoribonucléases 3'-5' processives. Cela signifie qu'elles dégradent l'ARNm par son extrémité 3' vers son extrémité 5' en restant liées au messenger entre chaque étape catalytique. Alors que la RNase II et la RNase R libèrent des ribonucléotides monophosphates, la PNPase fournit des ribonucléotides diphosphates. Cette différence est liée aux types de dégradation des enzymes : hydrolytique pour la RNase II et la RNase R mais phosphorolytique pour la PNPase. Une étude a montré qu'environ 90 % de l'activité exoribonucléotidique chez *E. coli* est de nature hydrolytique (réalisée par la RNase II et RNase R) contre 10 % phosphorolytique (réalisée par la PNPase) sur l'ensemble des ARN (Deutscher & Reuven, 1991). Cependant, l'inactivation de la PNPase, qui est un des composants majeur du dégradosome (III.3), entraîne la stabilisation d'un plus grand nombre d'ARNm par rapport à l'inactivation de la RNase II suggérant que cette dernière enzyme n'est pas majoritaire lors de la dégradation des ARNm (Mohanty & Kushner, 2003). Malgré tout, la délétion simultanée de ces deux enzymes n'est pas viable pour la cellule et mène à l'accumulation d'intermédiaires de dégradation (Donovan & Kushner, 1986) tout comme la délétion simultanée de la PNPase et de la RNase R (Cheng & Deutscher, 2005). D'autre part, la PNPase et la RNase II sont inhibées par des structures secondaires au niveau du messenger (au contraire de la RNase R) et ne sont pas capables de se lier à des ARN de très petites tailles (Guarneros & Portier, 1991; Spickler & Mackie, 2000). Les structures secondaires au niveau des 3'UTR peuvent, par exemple, être des tiges-boucles correspondant aux séquences terminatrices de la transcription (Rho-indépendante). Il existe néanmoins des enzymes permettant de surpasser ces difficultés : la RNase R qui peut dégrader les ARN avec ce type de structures (Cheng & Deutscher, 2005), la Poly(A) polymérase I (PAP I codée par le gène *pcnB*) et les hélicases. L'enzyme PAP est capable de synthétiser une queue poly(A) à l'extrémité 3' de ce type de messagers pour permettre aux 3' exonucléases de se lier et de dégrader les ARN de



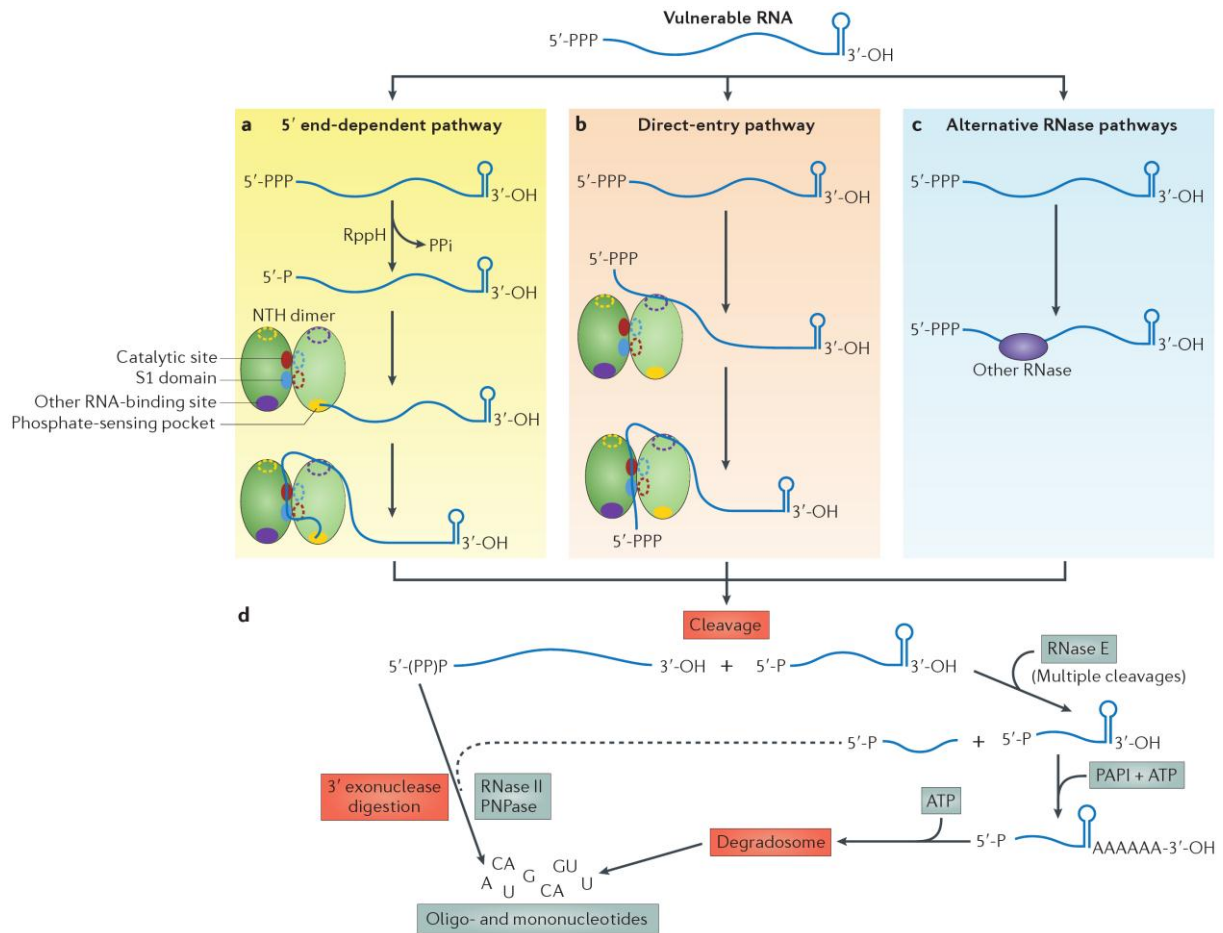
petites tailles. Plusieurs cycles de polymérisation et de dégradation successifs par PAP et la PNPase permettent également d'éliminer des transcrits disposant de structures secondaires stables (Blum et al., 1999; Régnier & Arraiano, 2000). Les hélicases, comme RhlB (codée par *rhlB*) qui fait partie du dégradosome, peuvent aussi, en hydrolysant l'ATP, déstabiliser ces structures en les « déroulant » pour permettre aux 3' exonucléases de dégrader les ARNm résistants (Coburn et al., 1999; Carpousis et al., 2008). Enfin, lorsque de petits oligoribonucléotides ne peuvent plus être ciblés par la RNase II et la PNPase, l'oligoribonucléase, codée par le gène essentiel *orn*, dégrade ces ARN en mononucléotides (Ghosh & Deutscher, 1999).

### **III. 2. Les mécanismes de dégradation des ARNm**

A partir des nombreuses études réalisées chez *E. coli* sur la dégradation des ARNm, il a été possible de définir un modèle global décrivant les mécanismes mis en jeu durant ce processus. L'ensemble de ces mécanismes sont récapitulés sur la **Figure 6**.

Il est aujourd'hui communément admis que dans des conditions normales de croissance, la dégradation des ARNm est initiée par un clivage endoribonucléolytique impliquant principalement la RNase E (Carpousis et al., 2009). Cette coupure peut également être réalisée par la RNase G pour certains transcrits ou par la RNase III au niveau de structures en double brin d'ARN. Ce clivage, qui correspond à l'étape limitante de la dégradation des messagers, entraîne la libération de produits de dégradation correspondant à des fragments d'ARNm. Ces fragments sont ensuite pris en charge par les exonucléases 3'-5' pour être réduits à l'état de monoribonucléotides (Carpousis, 2002). Si des structures secondaires gênent la progression des exonucléases, l'activité d'hélicases et/ou de la PAP permettent de surpasser ces difficultés comme décrit précédemment (III.1.2) (Régnier & Arraiano, 2000; Mackie, 2013). De plus, à l'issue du premier clivage, le fragment initialement situé en 3' de la zone de coupure présente une extrémité 5' monophosphate qui peut favoriser une nouvelle action de la RNase E qui va libérer à nouveau un fragment 5' monophosphate. Ce processus peut alors se répéter entraînant des produits de dégradation de plus en plus courts. Ainsi, bien qu'il n'existe pas d'exonucléase 5'-3' chez *E. coli*, il est suggéré que la dégradation des ARNm suive un sens 5'-3' comme proposé pour les ARNm polycistroniques (Selinger et al., 2003). Deux « voies d'entrée » initiales de la RNase E sur l'ARNm sont proposées (Bouvier & Carpousis, 2011) : soit le transcrit présente une extrémité 5' monophosphate favorisant

un premier clivage de la RNase E de façon 5'-dépendante (par exemple suite à l'action RppH) ou soit l'extrémité 5' n'est pas reconnue (à cause d'une triphosphorylation ou d'une structure secondaire piégeant l'extrémité 5') et la RNase E utilise la voie d'entrée « directe » en clivant le message de façon indépendante de l'extrémité 5'.

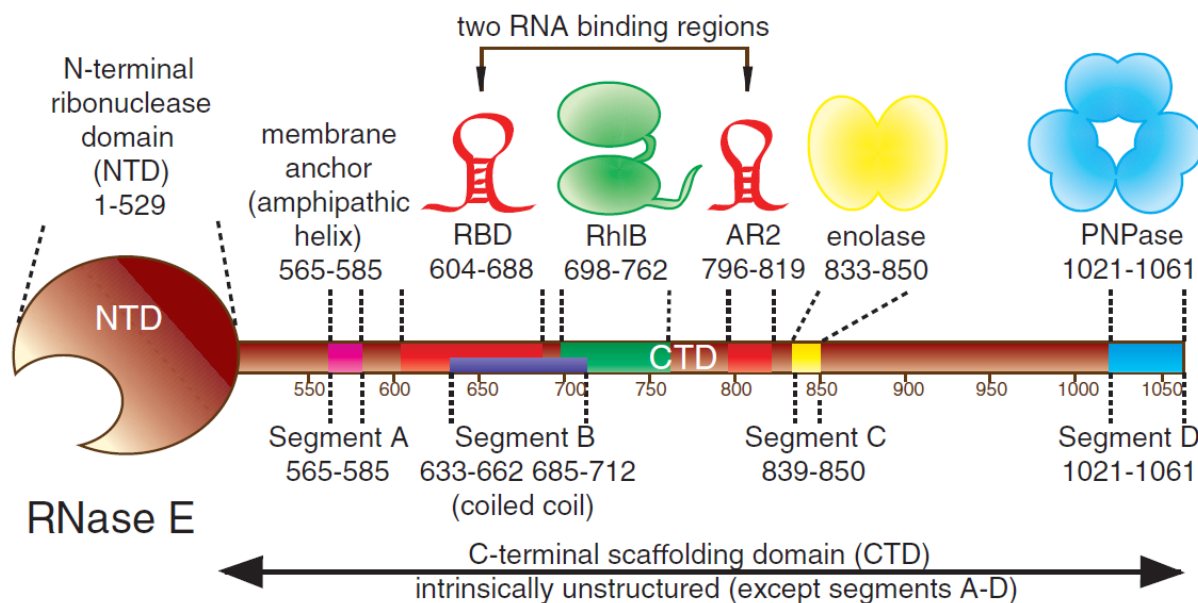


**Figure 6 :** Mécanismes généraux de dégradation des ARNm. a) La « voie d'entrée 5'-dépendante » de la RNase E est liée à la détection de l'extrémité 5' monophosphate par la RNase E. b) « Entrée directe » de la RNase E indépendante de l'extrémité 5' de l'ARNm. c) Initiation de la dégradation par d'autres endoribonucléases (RNase G et RNase III). d) Les trois « voies » convergent vers les mêmes étapes tardives de dégradation des messages faisant intervenir les exonucléases et un complexe protéique : le dégradosome qui sera détaillé en III.3.1. Seuls deux monomères de la RNase E sont représentés en vert. Le site catalytique est représenté en rouge, le domaine S1 qui forme un canal pour l'ARNm lors d'une reconnaissance 5'-dépendante du transcrite est représenté en bleu, un deuxième site d'interaction avec l'ARNm est représenté en violet et le site spécifique de reconnaissance de l'extrémité 5' monophosphate est représenté en jaune. Les sites sont colorés pour un monomère et représentés en pointillés pour l'autre. D'après (Mackie, 2013).

### III. 3. Le dégradosome

Le modèle global du processus de dégradation des ARNm chez *E. coli* montre que plusieurs partenaires protéiques interviennent. Bien qu'ils puissent agir de façon séparée et indépendante, il a été démontré que certaines de ces protéines interagissent et coopèrent pour former un complexe protéique appelé dégradosome.

Le dégradosome a initialement été découvert lors de la co-purification de deux nucléases majeures de la dégradation des ARNm : la RNase E et la PNPase (Carpousis et al., 1994; Py et al., 1994). Il a ensuite été démontré que ce complexe contenait également majoritairement deux autres protéines : la DEAD-box hélicase RhlB et l'enzyme glycolytique Enolase (codée par le gène *eno*) (Py et al., 1996). Comme indiqué précédemment, c'est la RNase E qui sert de support à ce complexe protéique. En effet, son domaine non structuré C-terminal possède des sites d'interaction protéine-protéine pour chacun des membres du dégradosome permettant la formation du complexe protéique (Callaghan et al., 2004) (**Figure 7**).



**Figure 7** : Schéma du domaine C-terminal de la RNase E et des différents domaines d'interaction protéine-protéine permettant la formation du dégradosome. D'après (Górna et al., 2012).

Un seul monomère de RhlB se fixe au complexe alors que l'Enolase qui est un dimère se lie à la RNase E avec un ratio de 2:1. Le cas de la PNPase est plus compliqué car la protéine fonctionne en trimère et chaque monomère contient un site de liaison à la RNase E. En conséquence, un trimère de PNPase pourrait interagir avec trois RNase E

simultanément. Etant donné que la PNPase et l'Enolase sont plus abondantes que la RNase E, il est probable que cette dernière n'existe qu'en complexe (Mackie, 2013).

L'interaction de l'endoribonucléase RNase E, de l'hélicase RhlB et de l'exonucléase PNPase au sein du dégradosome est supposée faciliter la dégradation des ARNm suivant le modèle de dégradation des ARNm (III.2) (Górna et al., 2012). En revanche, si le rôle exact de l'Enolase au sein du dégradosome n'est pas encore bien compris, il semble que la protéine jouerait un rôle dans la stabilité d'ARNm spécifiques de voies métaboliques produisant de l'énergie (Bernstein et al., 2004).

Le domaine C-terminal de la RNase E n'est pas essentiel à la cellule au contraire de son domaine catalytique N-terminal (Leroy et al., 2002). De ce fait, la non formation du dégradosome n'est pas létale. Cependant, la croissance chez ce type de mutant est largement affectée par rapport à la souche sauvage. De plus, une étude à l'échelle omique a démontré que chacun des quatre constituants majeurs du dégradosome (RNase E, PNPase, RhlB et Enolase) était nécessaire pour une dégradation normale des ARNm chez *E. coli* (Bernstein et al., 2004). D'autre part, cette étude a montré que le dégradosome intervenait bien dans la dégradation des ARNm mais pas pour la totalité d'entre eux.

D'autres protéines ont pu également être co-purifiées avec le dégradosome mais dans des quantités sous stœchiométriques. C'est le cas par exemple de la protéine Hfq (discuté en III.4.3), de l'exoribonucléase RNase R, de la PAP, de la Polyphosphate kinase PPK, des protéines chaperonnes GroEL et DnaK ainsi que des protéines ribosomales (Górna et al., 2012). Si les rôles de l'exonucléase et de la PAP peuvent être facilement envisagés dans l'action du dégradosome (III.1.2), celui des autres protéines n'est pas encore compris. D'autre part, la composition du dégradosome peut également être différente en phase stationnaire ou durant des stress de température (Carabetta et al., 2010). C'est par exemple le cas des hélicases alternatives comme CsdA, RhlE ou SrmB qui peuvent aussi s'associer avec le dégradosome bien que les raisons de ces interactions ne soient pas comprises (Carpousis, 2007).

Enfin, trois protéines interagissant avec la partie C-terminale de la RNase E ont montré leur capacité à réguler la composition du dégradosome et/ou son activité (Gao et al., 2006; Singh et al., 2009). Il s'agit des protéines RraA, RraB et de la protéine ribosomale L4 dont l'expression serait dépendante des conditions de croissance

permettant un ajustement de la dégradation d'ARNm spécifiques (Kaberdin & Lin-Chao, 2009).

### ***III. 4. Les mécanismes régulant la stabilité des ARNm***

Les mécanismes de dégradation des ARNm décrits précédemment peuvent être modulés pour ajuster de façon optimale la stabilité des messagers (et donc leur concentration) en cas d'adaptation à un nouvel environnement. Les processus permettant cette régulation seront détaillés dans cette partie.

#### **III. 4. 1. Les régulations de la RNase E**

Le rôle majeur de l'endoribonucléase RNase E dans le processus de dégradation des ARNm explique que sa régulation soit un moyen efficace pour ajuster le niveau de dégradation dans la cellule. Tout d'abord, l'expression même de la RNase E est soumise à une boucle d'auto-régulation puisque la RNase E régule la dégradation de son propre ARNm (Jain & Belasco, 1995). D'autre part, nous avons précédemment vu que la protéine RppH peut substantiellement modifier l'état de phosphorylation de l'extrémité 5' des ARNm et ainsi réguler l'activité de la RNase E (Deana et al., 2008). Également abordé précédemment, les protéines RraA, RraB et L14 régulent l'activité du dégradosome en modifiant sa composition via leur interaction avec le domaine C-terminal de la RNase E (Kaberdin & Lin-Chao, 2009). Enfin, une étude a démontré que la surexpression de la Poly(A) polymérase I entraîne une augmentation de la polyadénylation des ARNm ainsi qu'une diminution globale de leur stabilité bien que certains transcrits comme celui de la RNase E et de la PNPase soient stabilisés (Mohanty & Kushner, 1999). Les auteurs suggèrent qu'il existerait un mécanisme de régulation permettant de maintenir un équilibre entre polyadénylation et dégradation des ARNm.

#### **III. 4. 2. Les effets de la traduction**

Le processus de traduction joue un rôle sur la stabilité des messagers. En effet, la fixation de ribosomes au niveau du RBS localisé dans le 5'UTR des ARNm lors de l'initiation de la traduction ainsi que leur présence le long de la séquence codante lors de l'élongation peuvent perturber l'action des RNases et ainsi influencer la dégradation des messagers (Deana & Belasco, 2005; Dreyfus, 2009). Par exemple, des mutations affectant la séquence Shine-Dalgarno, le codon d'initiation de la traduction et l'ajout ou

retrait de codon stop sont associés à des variations de la stabilité de l'ARNm via leur effet sur la traduction (Arnold et al., 1998; Iost & Dreyfus, 1995; Braun et al., 1998). Globalement, lorsque l'initiation ou l'efficacité de traduction sont réduites, les ARNm dépourvus de ribosome sont déstabilisés et réciproquement. Ces résultats indiquent que les ribosomes ont généralement un effet protecteur contre les attaques de la RNase E, le plus souvent en bloquant l'accès de l'enzyme au niveau de ses sites de coupures. Cet effet est d'ailleurs d'autant plus fort que l'espace entre deux ribosomes en cours de traduction est faible (Pedersen et al., 2011). Cependant, il a été démontré que la présence de ribosomes uniquement à proximité de l'extrémité 5' d'un ARNm pouvait également le protéger de la dégradation (Nilsson et al., 1987; Rapaport & Mackie, 1994). Sachant qu'un des mécanismes de dégradation des ARNm nécessite la détection de l'extrémité 5' par la RNase E, ces résultats seraient expliqués par le fait que la présence des ribosomes dans cette région pourrait inhiber cette détection et donc la dégradation complète du messenger (Dreyfus, 2009).

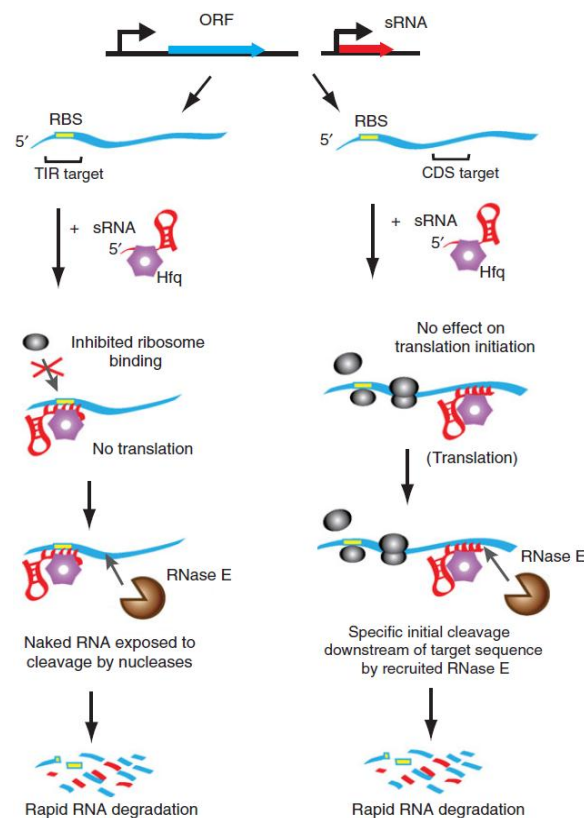
Toutefois, des preuves mettent en évidence que les ribosomes peuvent également avoir un effet délétère sur la stabilité des ARNm. En effet, si le déplacement des ribosomes le long de l'ARNm le protège contre les attaques nucléotidiques, les ribosomes en pause sur le messenger peuvent entraîner son clivage (Hayes & Sauer, 2003; Sunohara et al., 2004). Tout comme les mécanismes impliquant les ARNtm, ces phénomènes sont supposés assurer le contrôle qualité de la synthèse protéique. Enfin, une étude récente a mis en lumière une interaction du dégradosome avec le ribosome seul ou sous forme de polysomes. Les auteurs suggèrent que cette interaction permettrait un arrêt précoce de la traduction par la dégradation du messenger en fonction des conditions via l'intervention ou non d'un ARNnc (Tsai et al., 2012).

L'ensemble de ces résultats explique pourquoi un certain nombre de mécanismes impliqués dans la régulation de la traduction affectent également la stabilité des ARNm.

### **III. 4. 3. L'interaction Hfq / ARNnc / RNase E**

La protéine Hfq est une protéine se liant à l'ARN. Elle est considérée comme un des acteurs majeurs de la régulation post-transcriptionnelle car elle modifie l'expression d'ARNm en favorisant leur appariement avec des ARNnc (Vogel & Luisi, 2011). Parmi les multiples mécanismes identifiés pour l'action de Hfq, des travaux ont prouvé que ce régulateur interagit directement avec le domaine C-terminal de la RNase E et permet de

cibler des transcrits spécifiques à l'aide d'ARNnc pour les dégrader (Massé et al., 2003; Morita et al., 2005). C'est par exemple le cas de l'ARNm *ptsG* qui code pour un des transporteurs majeurs du glucose chez *E. coli*. Lors d'un stress entraînant l'accumulation de sucres phosphorylés, l'expression de l'ARNnc SgrS est induite (Vanderpool & Gottesman, 2004). Ce petit ARN va pouvoir alors s'apparier avec *ptsG* grâce à la protéine Hfq qui est elle-même associée à la RNase E. Ces interactions forment un complexe ribonucléoprotéique qui permet le clivage rapide de *ptsG* et SgrS par la RNase E. Ainsi le rôle des ARNnc dans la dégradation des ARNm serait de guider la RNase E vers leur cible via la protéine adaptatrice Hfq (Morita et al., 2005). Toutefois, il a été démontré que l'appariement d'un ARNnc sur son messenger cible peut s'effectuer dans la région d'initiation de la traduction ainsi que dans la région codante (Bouvier et al., 2008; Pfeiffer et al., 2009). Si un clivage endonucléolytique direct de l'ARNm par la RNase E semble indispensable lorsque l'ARNnc se fixe dans la séquence codante du messenger, il semble que ce soit l'inhibition de l'initiation de la traduction qui entraîne la dégradation de l'ARNm lorsque l'ARNnc se fixe à proximité des sites de fixation des ribosomes (Figure 8).



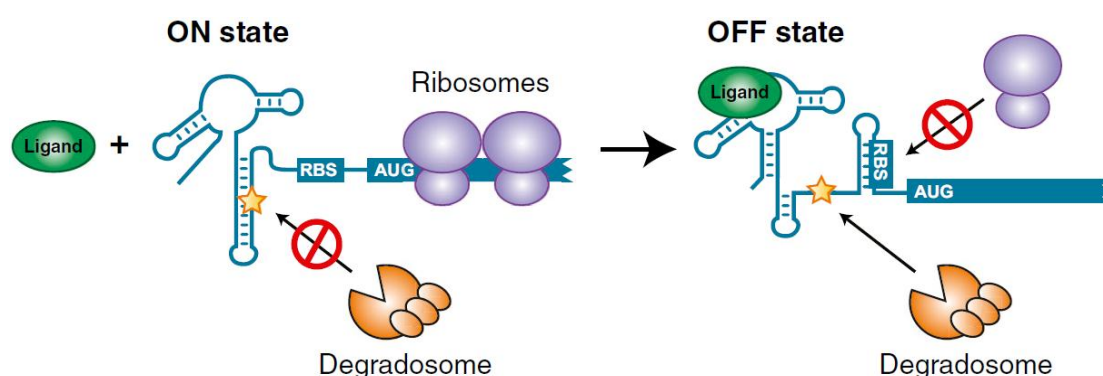
**Figure 8** : Schéma des différents mécanismes de la dégradation des ARNm par les petits ARN non codants, la protéine Hfq et la RNase E. D'après (Górna et al., 2012).

Enfin, une étude récente met en évidence que l'interaction de la protéine Hfq seule avec l'ARNm *cirA* induit sa déstabilisation en inhibant l'initiation de sa traduction. Au contraire, la présence de l'ARNnc RhyB favorise la traduction du messenger en raison de changements structuraux et le stabilise (Salvail et al., 2013). Ces résultats montrent que la protéine Hfq et les ARNnc peuvent avoir seuls des effets antagonistes sur la stabilité de leur cible.

### III. 4. 4. Les autres molécules se liant à l'ARNm

D'autres molécules que les ARNnc et Hfq peuvent influencer directement ou indirectement la stabilité de certains ARNm chez *E. coli*. C'est par exemple le cas du régulateur post-transcriptionnel global CsrA qui peut entraîner la stabilisation ou la déstabilisation de plusieurs ARNm et dont les mécanismes seront détaillés dans la partie IV de cet exposé (Romeo et al., 2012).

La liaison de ligands de faible poids moléculaire au niveau des ARNm peut mener à des changements de structures secondaires du messenger et réguler son expression généralement au niveau de l'initiation de la traduction (Serganov & Nudler, 2013). En conséquence, ce processus appelé riboswitch régule généralement indirectement la stabilité d'un ARNm. Cependant, Caron et ses collaborateurs ont récemment mis en évidence qu'un riboswitch au niveau de l'ARNm *lysC* contrôlait à la fois l'initiation de la traduction ainsi que la stabilité de l'ARNm de façon indépendante (Caron et al., 2012). En effet, l'interaction de la lysine au niveau du 5'UTR du messenger entraîne simultanément le piégeage du RBS au sein d'une tige-boucle (inhibant la traduction du transcrit) et la libération d'un site de clivage par la RNase E diminuant ainsi la stabilité de l'ARNm (**Figure 9**).



**Figure 9** : Schéma des régulations de la traduction et de la stabilité de l'ARNm *lysC* par le riboswitch sensible à la lysine. D'après (Caron et al., 2012).



### **III. 4. 5. Les systèmes toxine-antitoxine**

Si les systèmes toxine-antitoxine de type II correspondent à un type de régulation post-traductionnelle (II.4.1), d'autres types sont liés à des régulations post-transcriptionnelles au cours desquelles la stabilité d'ARNm est modifiée (Goeders & Van Melderen, 2014). Par exemple, dans le système toxine-antitoxine de type I GadY / *gadX-gadW*, l'antitoxine GadY correspondant à un ARNnc stabilise l'ARNm *gadX-gadW* de la toxine car leur appariement mène à des modifications de la séquence de l'ARNm via l'action de la RNase III (Opdyke et al., 2011). Toutefois, dans la plupart des autres systèmes de type I, l'antitoxine inhibe la traduction de l'ARNm de la toxine entraînant sa dégradation (Brantl, 2012). D'autre part, la plupart des toxines des systèmes de type II libérées en condition de stress correspondent à des endonucléases qui clivent de façon spécifiques certains ARNm (Goeders & Van Melderen, 2014). Enfin, il a été récemment découvert un nouveau type de système toxine-antitoxine, appelé type V, pour lequel, l'antitoxine GhoS correspond à une endonucléase qui dégrade l'ARNm codant pour la toxine GhoT (Wang et al., 2012).

### **III. 4. 6. L'influence de la séquence des ARNm**

La séquence ribonucléotidique des ARNm joue également un rôle dans la régulation de leur dégradation. En effet, la séquence d'un messenger décide de l'appariement ou non d'un ARNnc avec toutes les conséquences que cela implique au niveau de la stabilité de l'ARNm (III.4.3). Ensuite, bien qu'il n'existe aucune séquence consensus permettant de définir un site de clivage par la RNase E, McDowall et ses collaborateurs ont montré que l'enzyme clive préférentiellement les transcrits au niveau de régions non structurées riches en AU (McDowall et al., 1994) et que les structures secondaires localisées à proximité de ces sites peuvent diminuer l'efficacité de l'enzyme (McDowall et al., 1995). D'autre part, la présence d'un G, deux nucléotides en amont du site de coupure, favoriserait l'intervention de l'enzyme (Kaberdin, 2003). La séquence d'un ARNm peut être à l'origine de la formation de structures secondaires, tout particulièrement au niveau des 5'UTR et 3'UTR des ARNm. Plusieurs études mettent en évidence que ces conformations particulières peuvent moduler la stabilité des ARNm. Par exemple, il a été démontré qu'une tige-boucle au niveau du 5'UTR de l'ARNm de *ompA* favorise la stabilisation du transcrit (Arnold et al., 1998; Emory et al., 1992). En effet, la fusion du

5'UTR du gène *ompA* sur le messenger *bla* entraîne une très forte stabilisation de ce dernier.

### ***III. 5. Les déterminants de stabilité des ARNm à l'échelle omique***

Les études sur la dégradation des ARNm ont permis d'identifier les acteurs participant à cette tâche mais également ses principaux mécanismes. Ces travaux ont également conduit à l'identification de multiples voies de régulation de ce processus mais rarement à l'échelle de tous les messagers. En effet, uniquement trois études omiques se sont intéressées à la stabilité des ARNm chez *E. coli* (Bernstein et al., 2002; Selinger et al., 2003; Bernstein et al., 2004). Les deux premières études ont mis en évidence une stabilité très variable selon les transcrits, avec des temps de demi-vie de l'ordre de la minute jusqu'à plus d'une heure. La troisième étude a démontré que l'ensemble des éléments majeurs du dégradosome (RNase E, PNPase, RhlB et Enolase) sont nécessaires pour une dégradation « normale » des ARNm.

À partir de ces données, les auteurs ont tenté d'identifier quels pourraient être les déterminants majeurs des  $t_{1/2}$  des messagers. Ainsi, Bernstein et ses collaborateurs ont pu identifier une très faible corrélation négative entre stabilité et abondance suggérant que les ARNm les plus abondants sont les plus dégradés. De plus, la stabilité des transcrits appartenant à une même catégorie fonctionnelle semble similaire et dépendante de la catégorie en question (Bernstein et al., 2002; Selinger et al., 2003). Toutefois, les auteurs n'ont pu relier leurs données de stabilité avec aucun des autres paramètres testés : longueurs de la séquence codante et des UTR, pourcentages en GC, prédictions de sites de clivage de la RNase E, prédictions de structures secondaires et biais de codon (Bernstein et al., 2002). Dans leur étude, Selinger et ses collaborateurs ont pu déterminer qu'il existe un « sens » de dégradation 5' vers 3' au sein des opérons bien que ce « sens » ne soit pas exclusif (Selinger et al., 2003). Cependant, des études statistiques reprenant ces données de stabilité ont émis des conclusions différentes que celles de ces premiers auteurs soulevant de nouvelles questions et de nouveaux doutes. Par exemple, il existerait un lien entre la stabilité des ARNm et la longueur de leur séquence codante, le pourcentage de GC, le biais de codon ainsi que la formation de structures secondaires (Feng & Niu, 2007; Lenz et al., 2011). Enfin, des mesures de  $t_{1/2}$  des ARNm à l'échelle omique ont également été publiées pour plusieurs autres micro-organismes mais démontrent que les déterminants de la stabilité ne seraient pas tous

identiques d'une espèce à l'autre (Wang et al., 2002; Hambraeus et al., 2003; Redon et al., 2005; Hundt et al., 2007; Kristoffersen et al., 2012; Rustad et al., 2013; Dressaire et al., 2013).

## IV. Le système Csr chez *Escherichia coli*

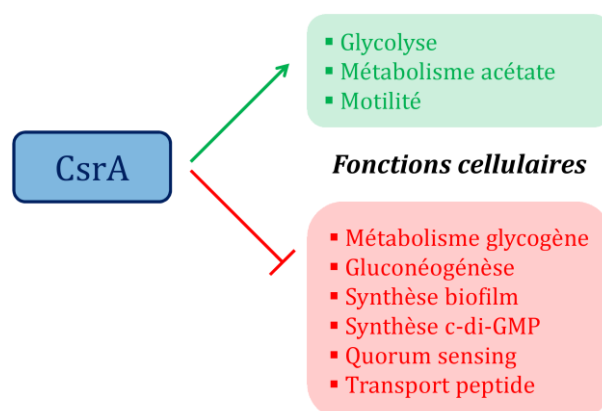
Comme décrit dans la partie II, la bactérie *E. coli* a développé de multiples niveaux de régulation pour moduler l'expression des gènes en fonction des conditions environnementales. Plus particulièrement, l'inhibition de l'initiation de la traduction (partie II) ainsi que la régulation de la dégradation des ARNm (partie III) font partie des mécanismes de régulation post-transcriptionnelle. Dans cette nouvelle partie, un exemple de régulation post-transcriptionnelle alliant ces deux mécanismes sera détaillé : il s'agit du système Csr.

### IV. 1. Description du système Csr

Le système Csr (pour Carbon Storage Regulator) est un système de régulation post-transcriptionnelle global qui contrôle la traduction et/ou la stabilité d'ARNm de gènes impliqués dans de nombreux processus d'adaptation physiologiques et métaboliques. Il est constitué de plusieurs éléments dont nous ferons la description ci-dessous.

#### IV. 1. 1. Le régulateur post-transcriptionnel global CsrA

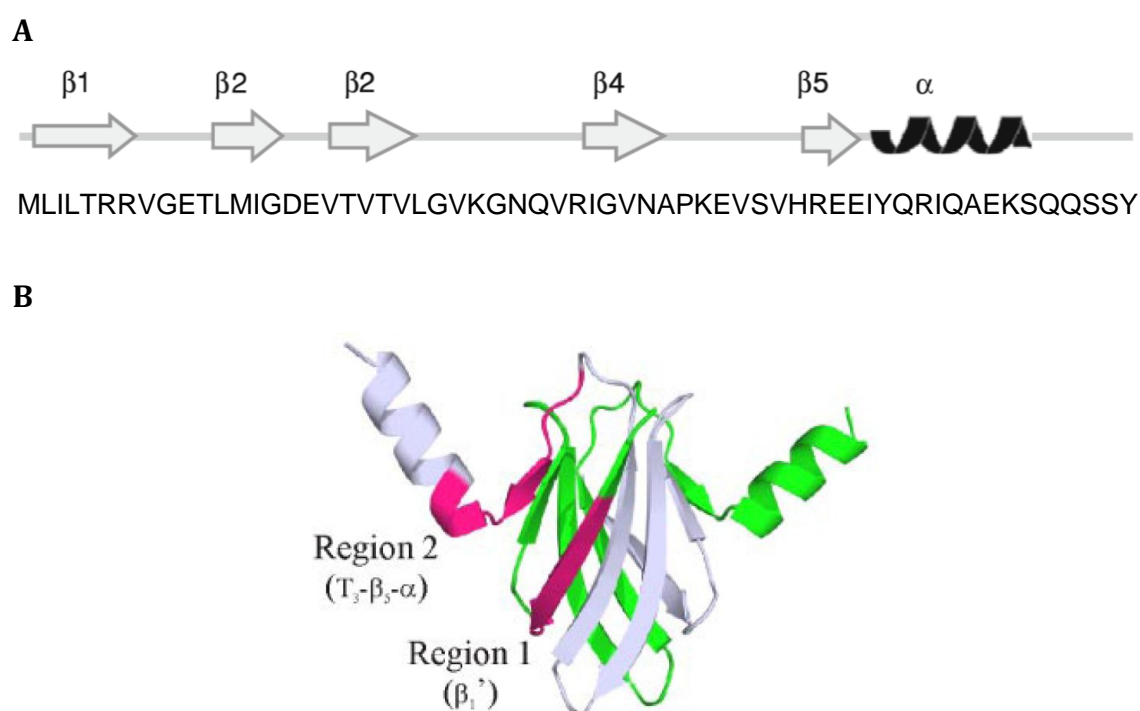
L'élément principal du système de régulation Csr est la protéine CsrA. Découverte en 1993 par Tony Roméo (Romeo et al., 1993) lors de la recherche de nouveaux gènes impliqués dans la synthèse du glycogène, la protéine CsrA a depuis été largement étudiée et démontrée comme régulant positivement ou négativement un grand nombre de fonctions cellulaires (Timmermans & Van Melderen, 2010) (**Figure 10**).



**Figure 10** : Fonctions cellulaires régulées par la protéine CsrA. La couleur verte désigne une activation alors que la couleur rouge indique une répression.

En effet, CsrA régule notamment négativement le métabolisme du glycogène (Yang et al., 1996), la gluconéogenèse (Romeo, 1996), la synthèse de biofilm par différentes voies (Jackson et al., 2002; Wang et al., 2005; Pannuri et al., 2012), la synthèse de c-di-GMP par les protéines GGDEF-EAL (Jonas et al., 2008), le transport de peptide (Dubey et al., 2003) ou encore le quorum sensing (Yakhnin et al., 2011a). A l'opposé, la protéine contrôle positivement plusieurs fonctions cellulaires dont la glycolyse (Sabnis et al., 1995), le métabolisme de l'acétate (Wei et al., 2000) ou la motilité (Wei et al., 2001). Il est clair maintenant que la protéine CsrA contrôle fortement le métabolisme carboné central et qu'elle réprime en fait, un grand nombre de gènes normalement exprimés en phase stationnaire (Romeo et al., 2012).

La protéine CsrA est une petite protéine de 61 acides aminés qui se lie spécifiquement à l'ARN (Liu & Romeo, 1997). Constituée de cinq feuillets  $\beta$  et une hélice  $\alpha$ , deux protéines CsrA s'assemblent en homodimère grâce au positionnement inversé des feuillets  $\beta$  qui forme un cœur hydrophobe stabilisateur (Gutierrez et al., 2005) (Figure 11).



**Figure 11** : (A) Représentation et séquence protéique de la protéine CsrA. D'après (Timmermans & Van Melderen, 2010). (B) Dimérisation de la protéine CsrA et région importante pour l'interaction avec l'ARN. Les deux monomères sont colorés en blanc et vert. Les acides aminés importants pour l'interaction sont localisés dans la région colorée en rouge (un seul site est coloré). D'après (Mercante et al., 2006)

Une étude comparative impliquant un criblage de mutations a par ailleurs démontré qu'il existe plusieurs acides aminés critiques pour l'interaction de CsrA avec l'ARN. Ces derniers sont localisés au niveau des feuillets  $\beta 1$  et  $\beta 5$  et à proximité de l'hélice  $\alpha$  de chaque monomère (Gutierrez et al., 2005; Mercante et al., 2006). D'autre part, les positions des deux monomères étant inversées, un homodimère de CsrA contient deux sites fonctionnels d'interaction avec l'ARN (Schubert et al., 2007; Mercante et al., 2009). En effet, chacun de ces sites fonctionnels est formé des acides aminés critiques du feuillet  $\beta 1$  du monomère A (Région 1 de la **Figure 11**) associées à ceux du feuillet  $\beta 5$  et de l'hélice  $\alpha$  du monomère B (Région 2 de la **Figure 11**) (et vice-versa pour le second site).

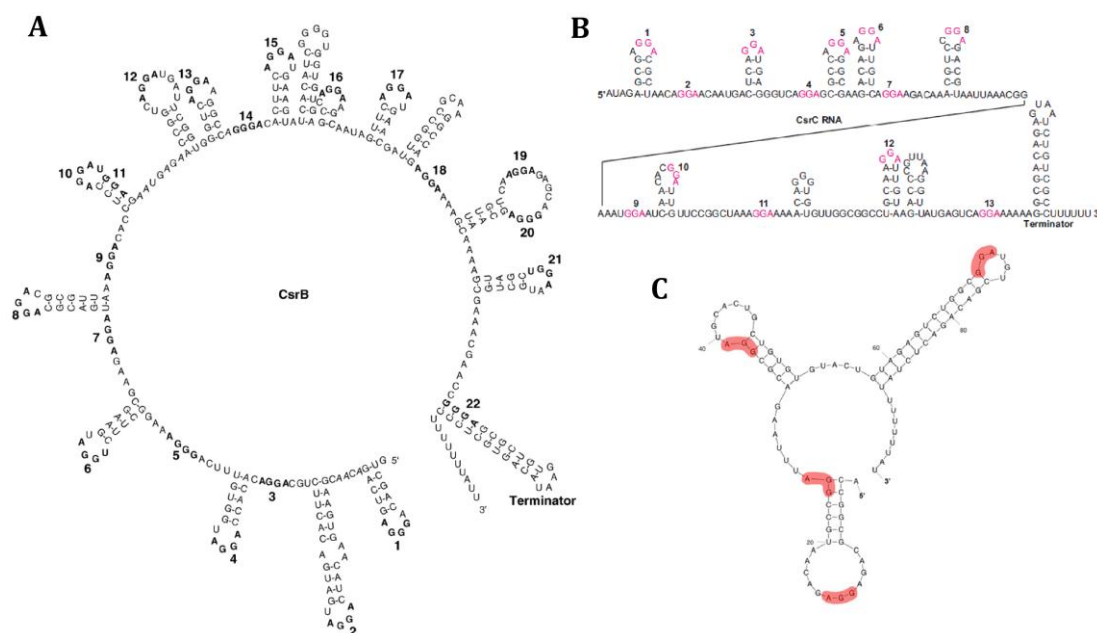
Le gène *csrA* est un gène essentiel chez les souches *E. coli* K-12 (Baba et al., 2006). En effet, la délétion complète du gène n'est pas viable sur source glycolytique en raison d'une accumulation trop importante de glycogène alors que la viabilité peut être restaurée par un double mutant  $\Delta csrA$  et  $\Delta glgCAP$  (Timmermans & Van Melderen, 2009). C'est la raison pour laquelle l'ensemble des études entreprises pour étudier le rôle de la protéine CsrA chez *E. coli* ont utilisé le mutant T1-5 *csrA::kan* de Tony Romeo obtenu par mutagenèse aléatoire (Romeo et al., 1993). Ce mutant contient l'insertion d'un transposon après le 50<sup>ème</sup> acide aminé de la protéine. Cette mutation mène à la disparition de la structure en hélice  $\alpha$  supposée par Mercante et ses collaborateurs comme participant à la reconnaissance de l'ARN (Mercante et al., 2006). Malgré cela, les multiples travaux réalisés avec ce mutant confirment que ce dernier conserve une activité réduite qui le rend viable.

Enfin, la protéine CsrA régule la traduction et/ou la stabilité d'ARNm. Elle cible ces derniers au niveau d'une séquence consensus. La séquence consensus de fixation de CsrA a été identifiée chez *E. coli* par la méthode de SELEX comme étant 5'-RUACARGGAUGU-3' (R correspondant à une purine) (Dubey et al., 2005). Au sein de cette séquence, il a été déterminé que le motif GGA était extrêmement conservé. D'autre part, la présence d'une tige-boucle contenant ce motif ainsi que la conservation de certains nucléotides favorisent d'autant plus l'affinité de la protéine pour sa cible (Dubey et al., 2005; Mercante et al., 2009; Duss et al., 2014). Il a été démontré que CsrA est capable de se fixer simultanément sur deux sites au sein d'un même transcrite de manière coopérative (Dubey et al., 2003; Wang et al., 2005). Cela signifie que lorsqu'un dimère de CsrA se fixe sur un premier site de liaison, cela stimule l'interaction du dimère au niveau

d'un site additionnel de plus faible affinité (Mercante et al., 2009). D'autre part, la distance minimale entre deux sites de fixation de CsrA sur un même ARN pour permettre la formation d'un pont serait de 10 nucléotides alors que la distance optimale serait de 18 nucléotides (Mercante et al., 2009).

#### IV. 1. 2. Les ARNnc séquestrant la protéine CsrA

Outre la protéine CsrA, le système Csr contient d'autres éléments qui permettent de réguler son activité. En effet, trois petits ARN non codants ont été identifiés à ce jour comme pouvant former un complexe ribonucléoprotéique avec la protéine CsrA. Ces petits ARN non codants sont CsrB, CsrC et McaS (Liu et al., 1997; Weilbacher et al., 2003; Babitzke & Romeo, 2007; Jørgensen et al., 2013). L'interaction de la protéine CsrA avec ces ARN entraîne une séquestration du régulateur post-transcriptionnel global qui ne peut alors se fixer sur ses messagers cibles. Cette régulation est rendue possible par le fait que la protéine CsrA présente une plus grande affinité pour ces petits ARN que pour les ARNm cibles. Ceci peut être expliqué par la structure semblable de ces différents ARN : ils possèdent tous de multiples sites de fixation de CsrA majoritairement en forme de tige-boucle bien que leur nombre varie selon l'ARN (Babitzke & Romeo, 2007; Duss et al., 2014) (**Figure 12**).



**Figure 12 :** (A) Structure prédite de CsrB. Les sites potentiels de fixation de CsrA sont représentés en gris. D'après (Babitzke & Romeo, 2007). (B) Structure prédite de CsrC. Les sites potentiels de fixation de CsrA sont représentés en rose. D'après (Weilbacher et al., 2003). (C) Structure prédite de McaS. Les sites potentiels de fixation de CsrA sont représentés en rouge. D'après (Jørgensen et al., 2013).

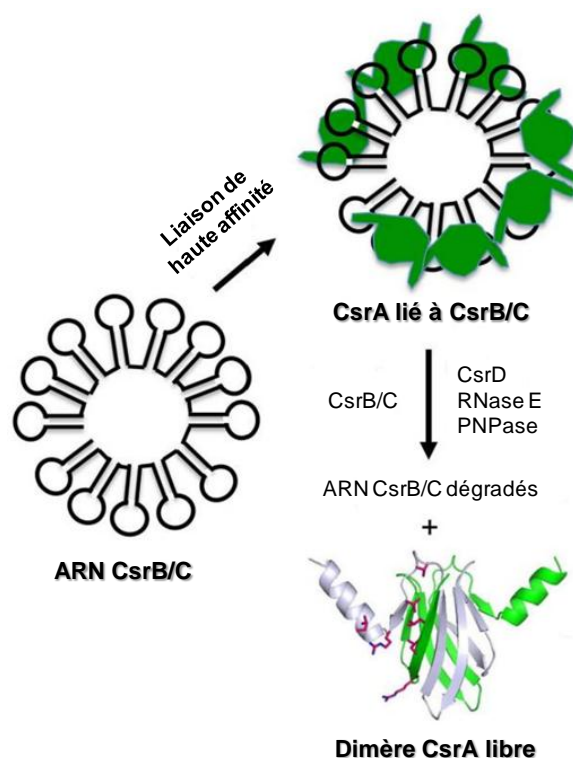
En effet, CsrB dispose d'environ 22 sites potentiels de fixation de CsrA pouvant séquestrer environ 9 dimères de CsrA alors que CsrC n'en présente que 13. Récemment, le petit ARN McaS, connu pour réguler l'expression de *csgD* via son interaction avec Hfq (Thomason et al., 2012), a été identifié comme un nouveau régulateur de CsrA avec seulement 4 sites potentiels de fixation de la protéine. Enfin, l'affinité de CsrA pour les ARN CsrB et McaS est similaire ( $K_d \approx 1$  nM) alors qu'elle est dix fois moins importante pour CsrC ( $K_d \approx 9$  nM) (Weilbacher et al., 2003; Jørgensen et al., 2013). Pour des cellules en croissance dans un milieu enrichi en glucose, il est estimé qu'environ 30 % des dimères de CsrA sont liés à CsrB alors qu'un bien moindre pourcentage serait lié à CsrC (Gudapaty et al., 2001; Weilbacher et al., 2003). Toutefois, la concentration cellulaire de CsrA (6 – 17  $\mu$ M) étant largement supérieure aux constantes de dissociation pour ses ARNm cibles, Roméo et ses collaborateurs proposent que la plupart des molécules de CsrA sont probablement liées dans la cellule (soit aux ARNnc, soit à ses ARNm cibles) (Romeo et al., 2012). La redondance des petits ARN CsrB/C et McaS suggère qu'ils contrôlèrent la quantité « active » de CsrA à des moments distincts de la croissance cellulaire, en fonction de la dynamique de leur expression (Jørgensen et al., 2013).

#### **IV. 1. 3. La protéine CsrD**

Un autre acteur clé du système Csr est la protéine CsrD (**Figure 13**). CsrD est prédite comme une protéine associée à la membrane et possède des domaines GGDEF et EAL (Jonas et al., 2006; Suzuki et al., 2006) qui sont généralement impliqués dans la synthèse et la dégradation de c-di-GMP (Ryjenkov et al., 2005; Schmidt et al., 2005). Il a été démontré que la protéine CsrD est essentielle à la dégradation des petits ARN CsrB et CsrC (Suzuki et al., 2006). Pourtant, les auteurs ont démontré que CsrD n'est pas une ribonucléase et qu'elle déstabilise ces petits ARN via l'action de la RNase E (et de la PNPase pour CsrB) par un mécanisme encore inconnu. D'autre part, bien que les domaines GGDEF et EAL soient indispensables à l'activité de la protéine, CsrD ne produit ni ne dégrade de c-di-GMP, probablement en raison d'une forte dégénérescence (Suzuki et al., 2006). La protéine CsrD est donc capable de contrôler le niveau des petits ARN CsrB et CsrC et donc indirectement l'activité de CsrA. À ce jour, seuls CsrB et CsrC ont été identifiés comme des cibles de CsrD, mais une revue récente suggère que CsrD pourrait cibler d'autres ARN (Romeo et al., 2012). D'un autre côté, CsrD stimule également



l'expression du facteur de transcription CsgD de façon indépendante du système Csr par un mécanisme qu'il reste encore à élucider (Sommerfeldt et al., 2009).



**Figure 13 :** Modèle actuel du fonctionnement du système Csr. Les ARN non codants possèdent de multiples sites de fixation de CsrA au sein de tiges-boucles pour séquestrer la protéine CsrA. Des dimères de CsrA « libres » seraient relâchés lors de la dégradation des petits ARN non codants impliquant CsrD et la RNase E (et PNPase pour CsrB). D'après (Romeo et al., 2012).

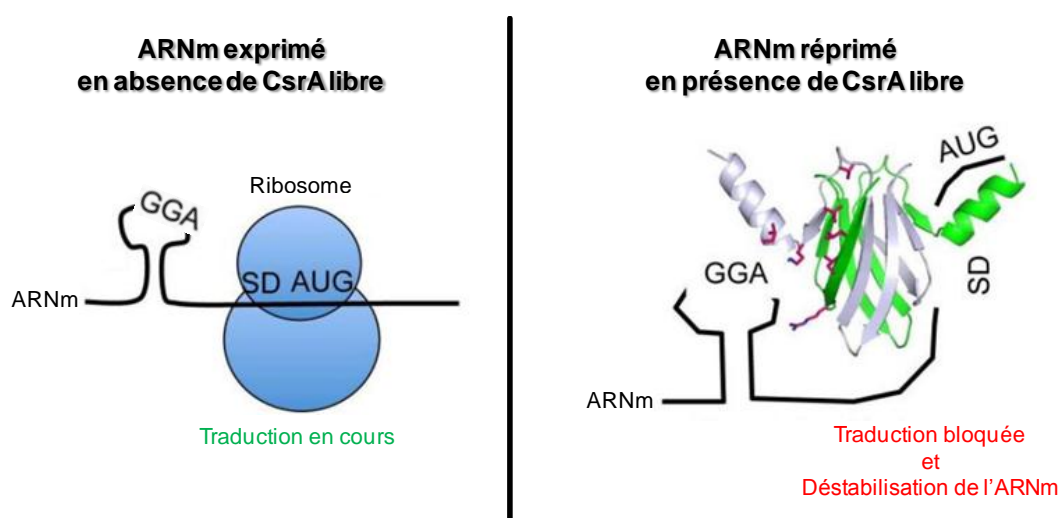
## IV. 2. Les mécanismes de régulation par CsrA

Le régulateur post-transcriptionnel CsrA régule soit positivement soit négativement l'expression d'une trentaine de transcrits bien que les mécanismes moléculaires n'aient été entièrement étudiés que pour quelques cibles. Dans cette partie, ces différents mécanismes de régulation par CsrA seront détaillés.

### IV. 2. 1. Les mécanismes de répression

À ce jour, la plupart des cibles identifiées de CsrA pour lesquelles le mécanisme moléculaire a été décrit sont des cibles négativement régulées par le régulateur global. Malgré certaines variations, ce mécanisme de régulation négative reste similaire pour toutes les cibles (**Figure 14**). Cette répression s'exerce via la fixation d'un dimère de

CsrA au niveau d'une séquence consensus détaillée au paragraphe IV.1.1. Cette séquence consensus est généralement présente 1 à 6 fois au niveau du 5'UTR et/ou au début de la séquence codante des ARNm cibles (Baker et al., 2002; Wang et al., 2005; Yakhnin et al., 2011a). Dans la plupart des cas, au moins un des sites de fixation de CsrA se superpose avec la séquence Shine-Dalgarno de fixation du ribosome (Seyll & Van Melderen, 2013). En conséquence, la fixation de CsrA sur cette séquence réprime la traduction de l'ARNm en empêchant la liaison du ribosome sur le messager (Baker et al., 2002) et entraîne dans la plupart des cas étudiés une déstabilisation de l'ARNm (Liu et al., 1995; Wang et al., 2005; Jonas et al., 2008).



**Figure 14 :** Schéma récapitulatif des mécanismes de régulation négatifs par CsrA. En absence de CsrA libre, l'ARNm cible peut être normalement traduit. En présence de CsrA libre, le dimère de CsrA se fixe sur l'ARNm au niveau de plusieurs sites de liaison à CsrA dont un se superpose avec la séquence Shine-Dalgarno (SD) inhibant la traduction et menant à la déstabilisation de la cible.

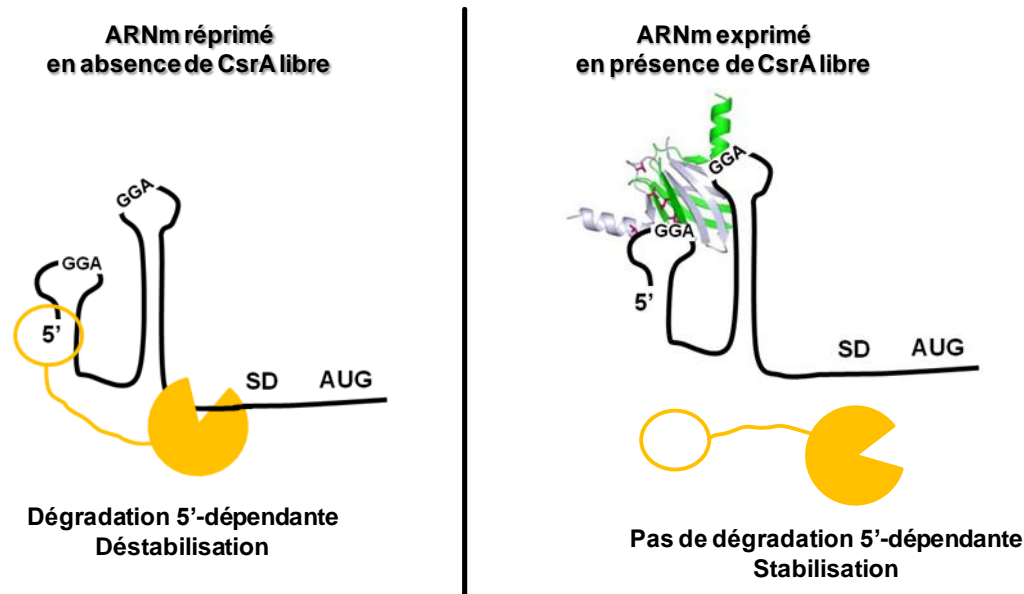
D'après (Romeo et al., 2012).

Des petites variations de ce mécanisme ont été identifiées pour certains transcrits. Par exemple, le gène *hfq* est le seul exemple de cible identifiée pour lequel l'ARNm ne présente qu'un site de fixation à CsrA. D'autre part, bien que la traduction du messager soit inhibée par CsrA, cela n'entraîne pas de déstabilisation de l'ARNm (Baker et al., 2007). Enfin, les deux sites de liaison de CsrA du transcrit du gène *sdiA* sont tous deux situés au niveau de la séquence codante du messager. Pour autant, les auteurs ont bien confirmé une compétition de CsrA avec la sous-unité 30S du ribosome entraînant une inhibition de la traduction (Yakhnin et al., 2011a).

#### **IV. 2. 2. Les mécanismes de régulation positive**

La capacité de la protéine CsrA à réguler positivement l'expression d'un gène a été décelée dès 1995 avec l'étude de l'impact de CsrA sur le métabolisme carboné central (Sabnis et al., 1995). Les auteurs ont démontré que l'activité de plusieurs enzymes de la glycolyse est favorisée par la protéine CsrA. En 2001, Wei et ses collaborateurs découvraient que CsrA régule positivement l'ARNm du gène *flhDC* en se fixant sur son 5'UTR et le stabilise (Wei et al., 2001) mais le mécanisme sous-jacent n'a été décrypté que très récemment (Yakhnin et al., 2013) (**Figure 15**). Deux sites de fixation de la protéine CsrA ont été identifiés dans le 5'UTR du messenger du gène *flhDC*. Ces sites de fixation, relativement éloignés de la séquence Shine-Dalgarno, résident néanmoins à proximité de l'extrémité 5' du messenger et de plusieurs sites de clivage de l'ARNm par la RNase E (Yakhnin et al., 2013). La fixation de CsrA sur ces sites entraînerait la séquestration de l'extrémité 5' de l'ARNm, bloquant ainsi le clivage du transcrit par la RNase E via la voie 5'-dépendante (partie III). Ce mécanisme serait donc suffisant pour stabiliser le messenger et favoriser ainsi son expression (Wei et al., 2001; Yakhnin et al., 2013).

Le deuxième exemple de régulation positive par CsrA concerne le gène *moaA* qui code pour une enzyme de biosynthèse du Moco (pour Molybdate cofactor), cofacteur nécessaire au fonctionnement des enzymes du métabolisme anaérobie. L'expression de ce gène est soumise à un rétrocontrôle négatif en présence de Moco. En effet, le 5'UTR de son ARNm pourrait former un riboswitch, qui par interaction avec le cofacteur, bloquerait sa traduction (Regulski et al., 2008). Il a été démontré que la protéine CsrA peut se fixer sur ce même riboswitch induisant une activation de la traduction (Patterson-Fortin et al., 2013). Bien que le mécanisme exact ne soit pas identifié, les auteurs suggèrent que cette activation serait liée à une modification structurelle de la région 5' lors de la fixation de CsrA (Patterson-Fortin et al., 2013).



**Figure 15** : Schéma récapitulatif du mécanisme de régulation de *flhDC* par CsrA. La RNase E reconnaît l'extrémité 5' et dégrade le transcrite. En présence de CsrA libre, le dimère de CsrA se fixe sur l'ARNm et séquestre l'extrémité 5' du messager inhibant la dégradation 5'-dépendante de la RNase E.

### IV. 3. Les ARNm cibles du système Csr

Ces mécanismes moléculaires de répression et d'activation de l'expression génique par CsrA ont été mis en évidence chez un certain nombre d'ARNm mais pas pour toutes les cibles « suggérées » de CsrA (**Table 2**).

À ce jour, il a été démontré que la protéine CsrA réprimait par le mécanisme détaillé précédemment, les ARNm suivants : *glgCAP* impliqué dans la synthèse de glycogène (Baker et al., 2002), *cstA* qui code pour un transporteur de peptidique (Dubey et al., 2003), *pgaABCD* qui permet la synthèse du PGA favorisant la formation de biofilm (Wang et al., 2005), *hfq* qui est un régulateur post-transcriptionnel (Baker et al., 2007), *csrA* (Yakhnin et al., 2011b), *sdiA* qui code pour un récepteur aux AHLs impliqué dans le quorum sensing (Yakhnin et al., 2011a) et *nhaR* qui est un facteur de transcription favorisant notamment l'expression de *pgaABCD* (Pannuri et al., 2012). D'autres ARNm ont également été identifiés comme interagissant avec la protéine CsrA et réprimés post-transcriptionnellement mais sans que le mécanisme exact soit déterminé. C'est le cas par exemple de *relA* qui est impliqué dans le système de réponse stringente (Edwards et al., 2011) ou *ydeH* et *ycdT* qui sont des protéines GGDEF-EAL (Jonas et al., 2008). En ce qui concerne les mécanismes de régulations positives exercés par CsrA, seuls les transcrits *flhDC* et *moaA* ont été identifiés comme cibles (Yakhnin et al., 2013; Patterson-Fortin et al., 2013).

**Table 2 :** Liste des cibles de la protéine CsrA chez la bactérie *E. coli* K-12.

Fonction	Blattner	Gène	Réponse (+ / -) <sup>a</sup>	Critères <sup>b</sup>	Références <sup>c</sup>
Biosynthèse glycogène	<i>b3430</i>	<i>glgC</i>	Répression (-)	A - F	1 - 3
	<i>b3429</i>	<i>glgA</i>	Répression (-)	C	4
	<i>b3432</i>	<i>glgB</i>	Répression (-)	C	1
	<i>b3049</i>	<i>glgS</i>	Répression (-)	C	4
Catabolisme glycogène	<i>b3428</i>	<i>glgP</i>	Répression (-)	C, F	4
	<i>b3417</i>	<i>malP</i>	Répression (-)	F	4
Gluconéogenèse	<i>b0688</i>	<i>pgm</i>	Répression (-)	F	5
	<i>b3403</i>	<i>pck</i>	Répression (-)	C, F	1
	<i>b1702</i>	<i>ppsA</i>	Répression (-)	F	5
	<i>b4232</i>	<i>fbp</i>	Répression (-)	F	5
Glycolyse	<i>b4025</i>	<i>pgi</i>	Activation (+)	F	5
	<i>b3916</i>	<i>pfkA</i>	Activation (+)	F	5
	<i>b1723</i>	<i>pfkB</i>	Répression (-)	F	5
	<i>b3919</i>	<i>tpiA</i>	Activation (+)	F	5
	<i>b2779</i>	<i>eno</i>	Activation (+)	F	5
	<i>b1676</i>	<i>pykF</i>	Activation (+)	F	5
Shunt glyoxylate	<i>b4015</i>	<i>aceA</i>	Activation (+)	F	6
	<i>b4014</i>	<i>aceB</i>	Activation (+)	F	6
Métabolisme acétate	<i>b4069</i>	<i>acs</i>	Activation (+)	C, F	7
Synthèse PGA (Biofilm)	<i>b1024</i>	<i>pgaA</i>	Répression (-)	A - E	8
	<i>b1023</i>	<i>pgaB</i>	Répression (-)	B	8
	<i>b1022</i>	<i>pgaC</i>	Répression (-)	B	8
	<i>b1021</i>	<i>pgaD</i>	Répression (-)	B	8
Métabolisme c-di-GMP (Biofilm)	<i>b1535</i>	<i>ydeH</i>	Répression (-)	(A), B, D	9
	<i>b1025</i>	<i>ycdT</i>	Répression (-)	(A), B, D	9
Facteur transcription (Biofilm)	<i>b0020</i>	<i>nhaR</i>	Répression (-)	B - F	10
Facteur transcription (Motilité)	<i>b1892</i>	<i>flhD</i>	Activation (+)	A - D, F	11, 12
	<i>b1891</i>	<i>flhC</i>	Activation (+)	A - D, F	11, 12
Régulateur post-transcriptionnel	<i>b4172</i>	<i>hfq</i>	Répression (-)	A, B, D - F	13
	<i>b2696</i>	<i>csrA</i>	Répression (-)	C - F	14
Quorum sensing	<i>b1916</i>	<i>sdiA</i>	Répression (-)	C - F	15
Transport peptidique	<i>b0598</i>	<i>cstA</i>	Répression (-)	C - E	16
Réponse stringente - ppGpp	<i>b2784</i>	<i>relA</i>	Répression (-)	B - D, F	17
Synthèse du cofacteur au molybdène	<i>b0781</i>	<i>moaA</i>	Activation (+)	B - D	18

**a.** CsrA réprime (-) ou active (+) l'expression du gène ou produit de gène considéré

**b.** Manipulations ayant permis d'identifier l'effet : mesure de la stabilité (A) ; dosage de l'ARNm par NorthernBlot, RT-qPCR et/ou puce à ADN (B) ; fusion transcriptionnelle et/ou traductionnelle avec *lacZ* (C) ; Interaction de la protéine avec le transcrit (D) ; Compétition de CsrA avec le ribosome (E) ; mesure de l'activité et/ou dosage protéique (F).

**c.** Les références listées sont : 1, (Romeo et al., 1993) ; 2, (Liu et al., 1995) ; 3, (Baker et al., 2002), 4, (Yang et al., 1996) ; 5, (Sabnis et al., 1995), 6, (Romeo, 1998) ; 7, (Wei et al., 2000) ; 8, (Wang et al., 2005) ; 9, (Jonas et al., 2008) ; 10, (Pannuri et al., 2012) ; 11, (Wei et al., 2001) ; 12, (Yakhnin et al., 2013) ; 13, (Baker et al., 2007) ; 14, (Yakhnin et al., 2011b) ; 15, (Yakhnin et al., 2011a) ; 16, (Dubey et al., 2003) ; 17, (Edwards et al., 2011) ; 18, (Patterson-Fortin et al., 2013).

Toutefois, d'autres ARNm sont considérés comme des cibles du système Csr alors même que leur mécanisme de régulation n'est pas encore connu. Les auteurs se sont alors basés sur des expériences de fusions transcriptionnelles/traductionnelles et/ou de mesure d'activité enzymatique dans un mutant *csrA* pour les identifier. C'est le cas par exemple, de plusieurs gènes de la gluconéogenèse réprimés par CsrA et de certains gènes impliqués dans la glycolyse et le métabolisme de l'acétate favorisés par la protéine (Romeo et al., 1993; Sabnis et al., 1995; Romeo, 1998; Wei et al., 2000).

#### **IV. 4. Le réseau de régulation du système Csr**

Au sein du système Csr, il a été précédemment démontré (IV.1) que l'activité de la protéine CsrA est régulée en particulier par les petits ARN non codants CsrB/C dont la dégradation est contrôlée par la protéine CsrD. Cependant, pour que la réponse adaptative apportée par le système Csr soit cohérente vis-à-vis de changements environnementaux, ces mécanismes appartiennent à un réseau de régulation plus complexe qui sera détaillé dans cette partie. Un modèle du réseau de régulation du système Csr est représenté par la **Figure 16**.

##### **IV. 4. 1. Régulation de CsrA**

La régulation de l'activité de CsrA en fonction des conditions environnementales constitue la clé de voûte du système Csr (**Figure 16**). L'activité du régulateur global dépend à la fois des petits ARN CsrB/C et McaS mais également de l'expression de la protéine. Yakhnin et ses collaborateurs ont démontré que la transcription du gène *csrA* était sous le contrôle de cinq promoteurs différents essentiellement sensibles au facteur  $\sigma^{38}$  ( $\sigma^S$ ) et/ou au facteur  $\sigma^{70}$  ( $\sigma^D$ ) (Yakhnin et al., 2011b). De façon surprenante, l'expression de *csrA* augmente fortement lors de la transition entre phase exponentielle de croissance et phase stationnaire. Cette augmentation est liée à l'accumulation bien connue du facteur  $\sigma^{38}$  ( $\sigma^S$ ) (Battesti et al., 2011) lors de cette transition qui active un des promoteurs du gène *csrA* (Yakhnin et al., 2011b). De plus, alors que la protéine CsrA est capable de favoriser indirectement sa transcription par un mécanisme encore inconnu, son ARNm est lui-même régulé négativement de façon post-transcriptionnelle par la protéine (Yakhnin et al., 2011b). Bien que cette double autorégulation opposée de CsrA ne soit pas bien comprise, il est probable que la mise en place de la régulation post-transcriptionnelle du messenger soit plus rapide que celle de la régulation transcriptionnelle indirecte. En conséquence, cela éviterait une synthèse trop importante de la protéine dans le cas où la concentration de sa forme « libre » serait critique dans la cellule.

##### **IV. 4. 2. Régulation des petits ARN non codants**

Le niveau des petits ARN non codants CsrB/C et McaS détermine la quantité de dimères de CsrA séquestrés et donc l'activité de la protéine. La dégradation de CsrB et

CsrC est contrôlée par la protéine CsrD qui les adresse à la RNase E (et PNPase pour CsrB). D'autre part, il a également été démontré que CsrA induisait indirectement leur transcription (Gudapaty et al., 2001; Weilbacher et al., 2003) créant ainsi une nouvelle boucle d'autorégulation (**Figure 16**). Au contraire, l'expression de McaS est indépendante de l'activité de CsrA et de la régulation transcriptionnelle de CsrB et CsrC (Jørgensen et al., 2013). Le contrôle qu'exerce CsrA sur la synthèse des ARN CsrB/C nécessite que la protéine active indirectement le système à double composante BarA-UvrY (Suzuki et al., 2002). Ce système à deux composantes est constitué d'une kinase liée à la membrane qui détecte les changements environnementaux (ex : BarA) et qui selon les conditions, active par phosphorylation un régulateur de réponse associé (ex : UvrY) capable de se lier à l'ADN et de réguler la transcription de certains gènes (ex : *csrB* et *csrC*) (Hoch, 2000; Pernestig et al., 2001). La transcription du gène *csrC* semble toutefois légèrement moins dépendante de l'activation de UvrY que celle de *csrB* (Weilbacher et al., 2003; Jonas et al., 2006) suggérant que les mécanismes d'activation transcriptionnelle des deux ARN soient légèrement différents. De plus, bien que BarA et UvrY fonctionnent généralement de pair, certains résultats montrent que CsrA serait capable d'induire indirectement l'activation de UvrY indépendamment de BarA (Suzuki et al., 2002). Enfin, ces auteurs ont également démontré que la protéine SdiA, dont l'ARNm est une cible réprimée par CsrA, active la transcription de *uvrY* et participe également à la synthèse de CsrB/C.

Tout comme CsrA, l'expression des petits ARN est fortement augmentée lors de la transition entre la phase exponentielle de croissance et la phase stationnaire (Jonas & Melefors, 2009). Par contre, cette expression est plus forte lorsque les cellules se développent en milieu pauvre par rapport au milieu riche. Le déclenchement de cette expression peut être expliqué par le fait que le système BarA-UvrY s'active en absence de certains éléments nutritifs dans le milieu et en présence d'acides carboxyliques issus du métabolisme carboné central comme le formate ou l'acétate (Chavez et al., 2010). Ce dernier métabolite est notamment connu pour s'accumuler dans le milieu de culture en fin de phase exponentielle (Wolfe, 2005).

#### **IV. 4. 3. Régulation de CsrD**

Tout comme les autres éléments du système Csr, le contrôle de l'expression de CsrD est déterminant car la protéine est responsable de la dégradation des petits ARN CsrB/C

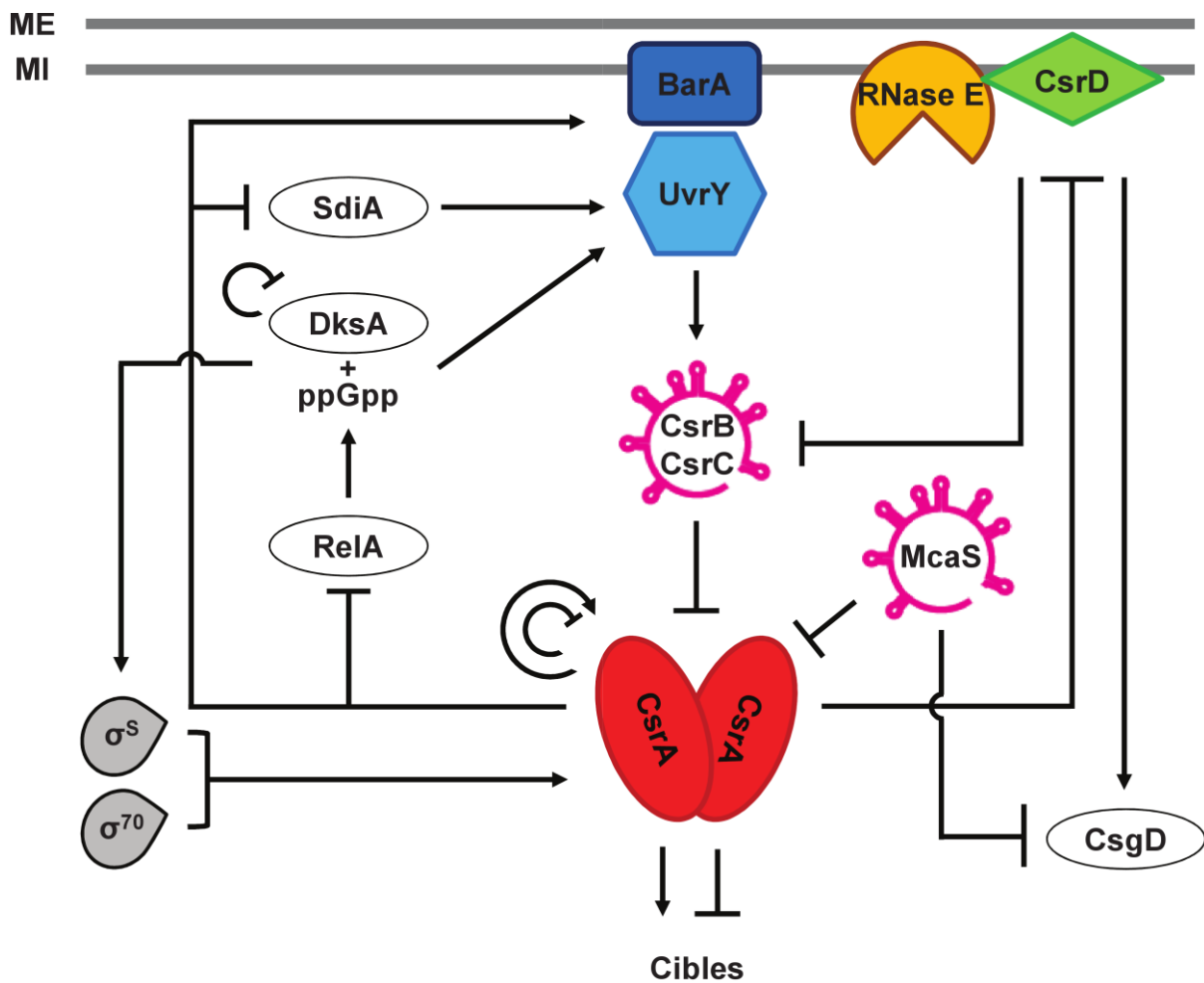
et donc participe à la régulation de leur concentration dans la cellule. La surexpression de CsrA montre que le niveau de l'ARNm de CsrD est diminué dans la cellule (Suzuki et al., 2006; Jonas et al., 2008). Cependant, ce contrôle n'intervient pas au niveau post-transcriptionnel : CsrA régule indirectement la transcription de *csrD*. Il est supposé qu'en cas de surexpression de CsrA, la protéine CsrD dégraderait moins les petits ARN qui séquestreeraient davantage CsrA, ajoutant ainsi une nouvelle boucle d'autorégulation complexifiant le réseau (**Figure 16**).

#### **IV. 4. 4. Liens avec d'autres systèmes de régulation**

Au travers des cibles dont il affecte l'expression, le système Csr est connecté à d'autres systèmes de régulation élargissant ainsi considérablement son réseau. Par exemple, un lien étroit entre Csr et le système de réponse stringente a été identifié. En effet, la protéine CsrA réprime post-transcriptionnellement *relA* inhibant ainsi la synthèse de ppGpp (Edwards et al., 2011). D'autre part, les auteurs ont clairement démontré que le ppGpp et la protéine associée DksA activent la transcription des petits ARN CsrB/C via UvrY mais indépendamment de CsrA par un mécanisme inconnu (Edwards et al., 2011). En conséquence, les deux systèmes sont interconnectés et se régulent l'un l'autre. De plus, les systèmes Csr et de réponse stringente exercent des effets opposés sur certaines cibles comme *glgCAP* (Romeo et al., 1990; Yang et al., 1996) et *flhDC* (Wei et al., 2001; Lemke et al., 2009), permettant probablement une régulation fine de ces messagers. Le gène *flhDC* correspond d'ailleurs à un facteur de transcription affectant lui-même l'expression d'un très grand nombre de gènes (Smith & Hoover, 2009). Comme indiqué précédemment, CsrA inhibe également directement l'expression de deux facteurs de transcription : SdiA qui régule des gènes impliqués dans la division cellulaire (Yakhnin et al., 2011a) et NhaR qui régule entre autres *pgaABCD* (Pannuri et al., 2012). D'autre part, en impactant la synthèse de c-di-GMP via la régulation de protéines à domaines GGDEF-EAL (Jonas et al., 2008), CsrA peut notamment contrôler les mécanismes d'adhésion et de formation de biofilm (Hengge, 2009b). Edwards et ses collaborateurs ont également découvert que plus de 700 ARN interagissaient potentiellement avec la protéine CsrA dont plusieurs sont des régulateurs globaux (Edwards et al., 2011). Par exemple, l'expression du régulateur post-transcriptionnel global Hfq est réprimée par la protéine CsrA (Baker et al., 2007) amplifiant en conséquence la portée des régulations induites par le système Csr. De son côté, la



protéine CsrD impacte potentiellement l'ensemble du régulon du facteur transcriptionnel CsgD en régulant son expression (Sommerfeldt et al., 2009).



**Figure 16 :** Modèle du réseau de régulation du système Csr. ME correspond à la membrane externe et MI correspond à la membrane interne (ou membrane plasmique). D'après (Seyll & Van Melderren, 2013).

## **V. Problématique de la thèse**

Cette étude bibliographique met en avant le rôle important et encore mal caractérisé de la stabilité des ARNm dans la régulation de l'expression des gènes chez *E. coli*. Si de nombreux travaux se sont attachés à la description des mécanismes moléculaires associés à ce processus, peu d'études se sont intéressées à la quantification et au rôle des régulations de la stabilité des ARNm à l'échelle omique. À l'inverse, avec l'avènement de l'ère omique, les données d'expression des gènes se sont multipliées. Toutefois, il est important de garder à l'esprit que les données transcriptomiques sont le reflet du niveau des messagers dans la cellule. Elles correspondent donc au résultat de l'équilibre entre synthèse et dégradation des ARNm. En conséquence, l'observation de différences transcriptomiques entre deux conditions peut être le fruit de variations de transcription et/ou de stabilité des ARNm même si ce dernier aspect est souvent négligé.

Parce que la dégradation des ARNm participe à l'ajustement de leur concentration au même titre que leur synthèse, l'étude des régulations de la stabilité des messagers ne peut pas être abordée sans prendre en compte l'impact des variations de transcription. Ainsi, en associant systématiquement les données de « stabilomes » et de « transcriptomes » de la bactérie par des approches intégratives, cette thèse évaluera l'impact de la stabilité des ARNm sur leur concentration en fonction de différents états physiologiques induits soit par des variations environnementales (via l'étude de l'effet du taux de croissance, dans le Chapitre I) ou génétiques (via l'étude de mutations du système Csr de régulation post-transcriptionnelle impliqué dans l'adaptation bactérienne, dans le Chapitre III).

Dans le premier chapitre de cette thèse, l'évolution de la stabilité des ARNm sera donc étudiée en fonction de variations de taux de croissance de la bactérie via l'utilisation de cultures continues. Il est en effet important de comprendre le réseau de régulation mis en place lors du contrôle du taux de croissance car ce dernier varie très fréquemment lors de perturbations environnementales ou de modifications génétiques. La compréhension de cet effet global améliorera donc plus largement celle de l'adaptation de la bactérie en permettant de démêler les réponses observées lors d'autres conditions.

Au-delà de cet effet « global » de dégradation des ARNm, la stabilité des transcrits peut également être régulée par des processus plus spécifiques ciblant uniquement

certaines messagers. C'est la raison pour laquelle nous avons décidé d'étudier le cas de la régulation post-transcriptionnelle par le système Csr impliqué lors de l'adaptation bactérienne. Bien qu'elle soit très souvent suggérée, l'implication de la protéine CsrA dans la modification de la stabilité de ses « cibles » n'a été expérimentalement démontrée que de façon marginale. De plus, l'étendue du régulon Csr n'a pas encore été évaluée. En conséquence, le troisième chapitre de cette thèse étudiera, à l'aide de mutants du système Csr, l'impact de ce dernier sur la dégradation des ARNm à l'échelle omique.

La compréhension globale des paramètres qui influent sur la stabilité d'un ARNm est nécessaire pour comprendre pourquoi cette dernière varie en fonction des conditions environnementales. Toutefois, l'étude bibliographique a révélé que les déterminants de la diversité de stabilité des messagers restent encore peu étudiés et donc peu compris notamment parce que les paramètres intrinsèques (longueur, pourcentage en GC...) et extrinsèques (concentration, essentialité...) des transcrits correspondent à des données très hétérogènes qu'il est difficile d'analyser ensemble et de hiérarchiser. Avec l'aide des mathématiques, certaines approches intégratives permettent d'identifier les principaux paramètres participant à la détermination d'une variable malgré la diversité de leur nature. Nous utiliserons donc, au cours d'un chapitre intermédiaire (Chapitre II), ce type d'outil afin d'identifier et de hiérarchiser les principaux déterminants de la stabilité des ARNm à l'échelle omique. Les données de stabilité obtenues à différents taux de croissance de la bactérie permettront d'étudier si ces déterminants évoluent en fonction de ce paramètre physiologique majeur.





# **CHAPITRE I**

**Rôle de la transcription et de la  
stabilité des transcrits dans la  
régulation de l'expression génique à  
différents taux de croissance  
d'*Escherichia coli* sur glucose**



La régulation de la concentration des ARNm dans la cellule fait partie des leviers de régulation de l'expression génique permettant d'ajuster le taux de croissance en réponse à des changements environnementaux. Si l'effet de la vitesse de croissance sur le mécanisme de transcription est relativement bien documenté pour la bactérie *E. coli*, aucune étude à l'échelle omique n'a évalué l'effet spécifique d'une variation de taux de croissance sur la stabilité des ARNm bien que plusieurs travaux révèlent que la dégradation de certains transcrits varie en réponse à un stress (Nilsson et al., 1984; Lin et al., 2012; Albertson & Nyström, 1994; Kuzj et al., 1998).

En utilisant les techniques de culture continue (chémostat) et batch, nous avons pu faire varier le taux de croissance de la bactérie *E. coli* K-12 MG1655 en utilisant le même milieu de croissance (milieu minimum M9 supplémenté en glucose) à 0.10, 0.20, 0.40 et 0.63 h<sup>-1</sup> (correspondant à des temps de doublement de 6.9, 3.5, 2.3 et 1.1 h respectivement). Puis, à l'aide d'une approche basée sur des puces à ADN, les concentrations (transcriptome) et les t<sub>1/2</sub> (stabilome) de 70 % des ARNm d'*E. coli* ont pu être mesurés de manière fiable et ce pour les quatre taux de croissance étudiés.

Ainsi, il a été démontré pour la première fois que la dégradation des ARNm augmente globalement avec le taux de croissance. Le temps de demi-vie médian diminue en effet de 4.2 à 2.8 min entre 0.10 et 0.63 h<sup>-1</sup>. On peut cependant noter que la dégradation reste significative même à faible taux de croissance avec un temps de demi-vie médian relativement faible. Cette déstabilisation qui concerne la majorité des transcrits est associée à une surexpression globale des messagers de la machinerie de dégradation. Toutefois, environ un millier d'ARNm ne suivent pas cette tendance globale puisque leur stabilité reste constante ou augmente avec l'accroissement du taux de croissance (de façon continue ou uniquement à  $\mu = 0.63$  h<sup>-1</sup>). D'autre part, bien que plusieurs profils de dégradation soient observés au sein des ARNm polycistroniques, le taux de croissance a généralement peu d'influence sur la stabilité différentielle des cistrons qui le composent.

Même si les ARNm sont majoritairement déstabilisés à fort taux de croissance, les données transcriptomiques mettent en évidence que la concentration des ARNm dans les cellules augmente alors qu'elles se multiplient plus rapidement. Afin de déterminer pour chaque transcrit, la contribution relative des processus de transcription et de dégradation sur la variation de concentration de l'ARNm entre deux taux de croissance, les données transcriptomiques et de stabilomes ont été analysées par une approche

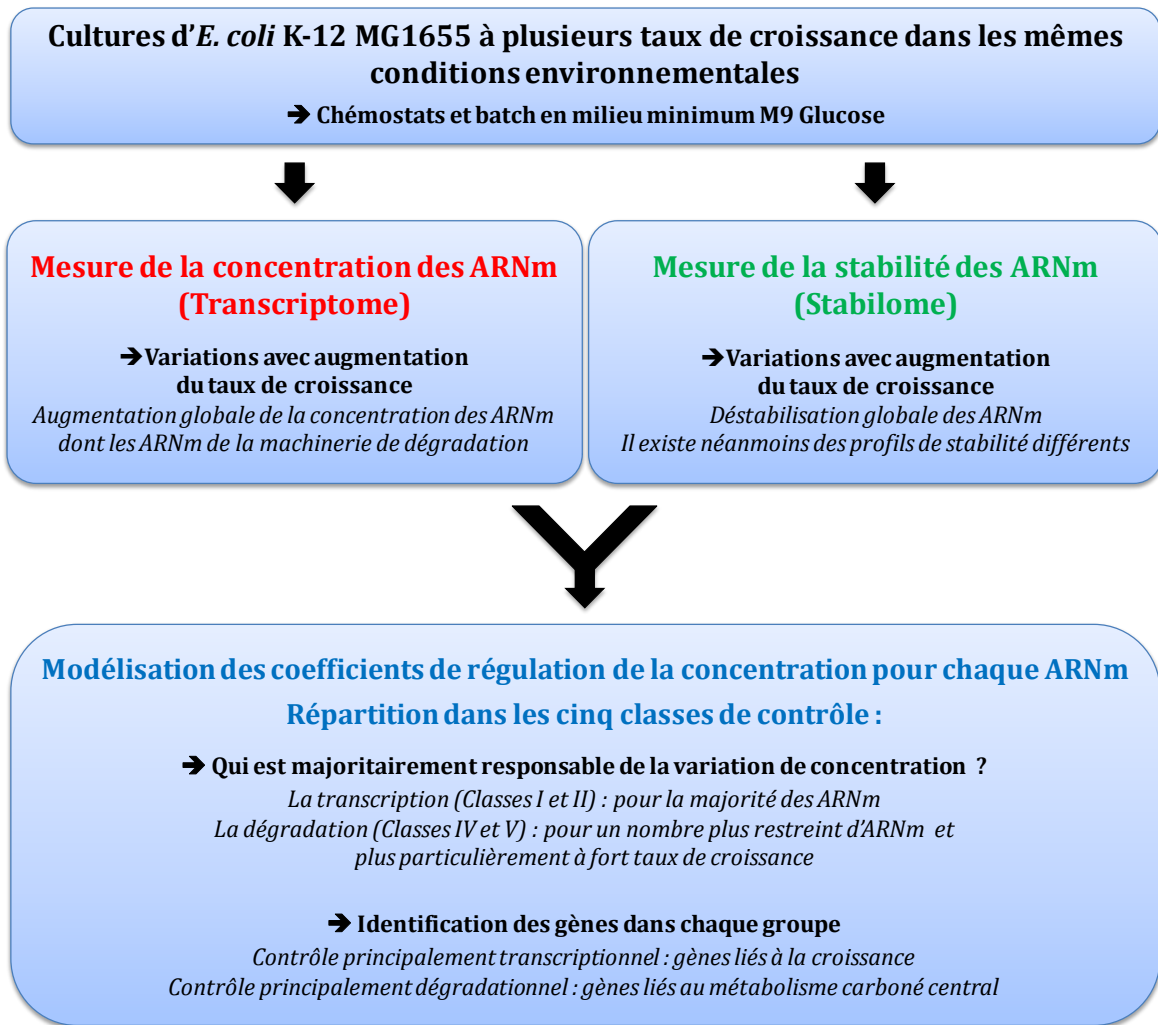


intégrative basée sur le calcul de coefficients de régulation selon une approche dérivée des travaux de ter Kuile and Westerhoff (ter Kuile & Westerhoff, 2001). Selon leurs valeurs, ces coefficients ont permis d'identifier cinq classes différentes de régulation de la concentration des messagers allant d'un contrôle majoritairement transcriptionnel à un contrôle majoritairement dégradatif.

Les résultats de ces expériences mettent en évidence que dans la plupart des cas, l'augmentation de la transcription avec le taux de croissance surpasse la déstabilisation des ARNm et contrôle donc majoritairement leur niveau, même si la dégradation participe à cet équilibre. Toutefois, le contrôle majoritairement dégradatif affecte un nombre non négligeable de transcrits et augmente de façon très significative à fort taux de croissance. Ceci indique que la stabilité des ARNm impacte davantage la concentration des transcrits lorsque les cellules se multiplient rapidement. L'identification de ces transcrits contrôlés majoritairement par la dégradation révèle qu'un grand nombre d'entre eux est impliqué dans le métabolisme carboné central soulignant que la dégradation des ARNm pourrait jouer un rôle majeur dans l'adaptation du métabolisme d'*E. coli* lorsque les conditions environnementales varient.

En conclusion, cette étude démontre clairement chez *E. coli* les effets partagés et associés des régulations de la transcription et de la dégradation des ARNm dans le contrôle de leur expression en fonction du taux de croissance.

Ces travaux ont fait l'objet d'une publication en 2014 dans le journal *Nucleic Acids Research* et sont présentés ici sous cette forme. Veuillez noter qu'une petite erreur s'est glissée dans la Figure 4 de l'article. Les symboles carrés et triangulaires correspondent aux taux de croissance  $\mu = 0.20$  et  $0.40 \text{ h}^{-1}$  respectivement (et non l'inverse). Seule une partie des données supplémentaires est présentée dans le manuscrit (Figure S1, Figure S2 et Table S3). Les autres données, de trop grande taille pour être présentées ici, sont néanmoins disponibles sur la version numérique de la publication (Table S1 et Table S2 disponibles sur <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3936743/>).



# Dual role of transcription and transcript stability in the regulation of gene expression in *Escherichia coli* cells cultured on glucose at different growth rates

Thomas Esquerré<sup>1,2,3,4</sup>, Sandrine Laguerre<sup>1,2,3</sup>, Catherine Turlan<sup>4</sup>, Agamemnon J. Carpousis<sup>4</sup>, Laurence Girbal<sup>1,2,3,\*</sup> and Muriel Cacaïgn-Bousquet<sup>1,2,3</sup>

<sup>1</sup>Université de Toulouse; INSA, UPS, INP; LISBP, 135, avenue de Rangueil, 31077 Toulouse, France, <sup>2</sup>INRA, UMR792 Ingénierie des systèmes biologiques et des procédés, 31400 Toulouse, France, <sup>3</sup>CNRS, UMR5504, 31400 Toulouse, France and <sup>4</sup>Laboratoire de Microbiologie et Génétique Moléculaires, UMR5100, Centre National de la Recherche Scientifique et Université Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse, France

Received July 11, 2013; Revised October 23, 2013; Accepted October 25, 2013

## ABSTRACT

Microorganisms extensively reorganize gene expression to adjust growth rate to changes in growth conditions. At the genomic scale, we measured the contribution of both transcription and transcript stability to regulating messenger RNA (mRNA) concentration in *Escherichia coli*. Transcriptional control was the dominant regulatory process. Between growth rates of 0.10 and 0.63 h<sup>-1</sup>, there was a generic increase in the bulk mRNA transcription. However, many transcripts became less stable and the median mRNA half-life decreased from 4.2 to 2.8 min. This is the first evidence that mRNA turnover is slower at extremely low-growth rates. The destabilization of many, but not all, transcripts at high-growth rate correlated with transcriptional upregulation of genes encoding the mRNA degradation machinery. We identified five classes of growth-rate regulation ranging from mainly transcriptional to mainly degradational. In general, differential stability within polycistronic messages encoded by operons does not appear to be affected by growth rate. We show here that the substantial reorganization of gene expression involving downregulation of tricarboxylic acid cycle genes and acetyl-CoA synthetase at high-growth rates is controlled mainly by transcript stability. Overall, our results demonstrate that the control of transcript stability has an important role in fine-tuning mRNA concentration during changes in growth rate.

## INTRODUCTION

The doubling time (or growth rate) of bacterial cells is determined by their metabolic activity, which is in turn dependent on genetic content and environmental conditions (including growth medium composition and temperature) (1). Studies with bacterial cells growing at imposed doubling times in continuous cultures report overall adaptation of the gene expression network in response to changing growth rate (2–5). Gene expression is controlled by the global effect of the activity of the transcription machinery and the specific effect of transcription factors (6). The rate of transcription of all genes is affected by the growth-rate dependent activity of (free) RNA polymerase and its partitioning into the synthesis of stable RNA species (rRNA and tRNA) and of messenger RNA (mRNA) (7). On the contrary, global regulators [transcriptional regulators such as the Crp, Fis and FNR (8) or sigma factors such as RpoS (9)] specifically activate or inhibit transcription of target genes to enable *Escherichia coli* to adapt rapidly to changing environment.

However, the gene expression as measured in transcriptomic experiments reflects RNA metabolism, the balance between mRNA synthesis by transcription and both mRNA degradation and dilution as cells divide. mRNA degradation participates in the elimination of defective or superfluous transcripts and contributes to a recycling of the nucleotide pool in the cell. In *E. coli*, most of mRNA turnover is mediated by RNase E, which is the major endoribonuclease responsible for initiating the decay of messengers (10). Then, 3'-exonucleases are believed to perform most of the conversion of RNase E-cleavage products to mononucleotides. RNase

\*To whom correspondence should be addressed. Tel: +33 561 559 419; Fax: +33 561 559 400; Email: girbal@insa-toulouse.fr

© The Author(s) 2013. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

E forms the RNA degradosome by associating with polynucleotide phosphorylase (PNPase), RNA helicase B (RhlB) and the glycolytic enzyme enolase through its C-terminal scaffold region (11). RNase E self-regulates its level of expression to meet growth requirements (12). A number of proteins (cold shock helicase CsdA, putative regulators of ribonuclease activity RraA, RraB) that occasionally co-purified with the *E. coli* degradosome in substoichiometric amounts are believed to be involved in modulation of degradosome composition, activity or specificity during bacterial adaptation to changing environments (11,13,14). Distinct from the RNA degradosome, RNase E also forms ribonucleoprotein complexes with Hfq/small RNAs to destabilize target mRNAs under specific physiological conditions (for example with Hfq/SgrS in response to the accumulation of glucose phosphates) (15). Other enzymes also contribute to mRNA decay in *E. coli*. In particular, the pyrophosphohydrolase RppH converts the 5' triphosphate end of a primary transcript to a 5' monophosphate favoring RNase E-mediated cleavage (10,16). Polyadenylation of mRNAs by the poly(A) polymerase I is believed to stimulate mRNA degradation (17–19).

For a given growth condition, the rate of dilution of each mRNA can be easily determined from growth kinetics. Large-scale determination of mRNA half-lives is more complex and therefore, studies of the global effect of growth rate on mRNA degradation are rare in *E. coli*. In the 70s, Pato *et al.* (20) were the first to report a trend of bulk mRNA stabilization as the growth rate decreases. Using microarray-based techniques, Bernstein *et al.* (21) reported large-scale analysis of mRNA half-life in *E. coli* cells growing quickly on rich medium and slowly on minimal medium: they did not identify any significant difference in average mRNA half-lives between the two conditions. In contrast, the stability of a few particular transcripts was demonstrated to vary with the growth rate in *E. coli* cells (22,23). Moreover, there was evidence that mRNA stability depends on the growth rate in mycobacteria and lactic acid bacteria (24,25).

To elucidate the effects of growth rate on overall mRNA stability in *E. coli*, we performed continuous cultures on glucose minimal medium. We measured the half-lives of the transcripts of 70% of the genes of *E. coli* at four different growth rates, corresponding to a large range of doubling times from 6.9 to 1.1 h. At the global regulatory level, we examined the growth rate effect on the expression of the mRNA degradation machinery. At the transcript-specific regulatory level, we showed how these exhaustive data sets can be used to identify new regulation of transcript stability (even within operons) in response to changing growth rate. We used a formal mathematical method to combine the data obtained with determinations of mRNA concentrations at the different growth rates. This genomic method provides, for the first time, insight into the relative contributions of mRNA degradation and transcription to the control of gene expression in *E. coli*. Both transcription and degradation contribute. Contrary to general practice in transcriptomic work studies, mRNA stability cannot be neglected when considering the regulation of gene

expression. These results were further analyzed in view of gene functionalities, and the significance of mRNA degradation on global metabolic adaptation was highlighted. At high-growth rates, changes in mRNA stability controlled the concentration of numerous transcripts of genes involved in central carbon catabolism. Our results demonstrated that the control of transcript stability played an important role in fine-tuning mRNA concentration during changes in growth rate.

## MATERIALS AND METHODS

### Strains and culture conditions

*E. coli* K-12 MG1655 strain ( $\lambda$ -F-*rph-1*) was grown in M9 minimal medium supplemented with glucose with a composition as follows: in g l<sup>-1</sup>: Na<sub>2</sub>HPO<sub>4</sub> 12H<sub>2</sub>O 3.48, KH<sub>2</sub>PO<sub>4</sub> 0.606, NaCl 0.51, NH<sub>4</sub>Cl 2.04, MgSO<sub>4</sub> 0.098, CaCl<sub>2</sub> 0.00438, thiamine hydrochloride 0.1, glucose 3.0 and trace elements; in mg l<sup>-1</sup>: Na<sub>2</sub>EDTA·2H<sub>2</sub>O 15, ZnSO<sub>4</sub>·7H<sub>2</sub>O 4.5, CoCl<sub>2</sub>·6H<sub>2</sub>O 0.3, MnCl<sub>2</sub>·4H<sub>2</sub>O 1, H<sub>3</sub>BO<sub>3</sub> 1, Na<sub>2</sub>MoO<sub>4</sub>·2H<sub>2</sub>O 0.4, FeSO<sub>4</sub>·7H<sub>2</sub>O 3, CuSO<sub>4</sub>·5H<sub>2</sub>O 0.3; 10 μl l<sup>-1</sup> of PPG2000 was added to prevent foam formation. All cultures were inoculated from overnight precultures in M9 minimal medium. For chemostat experiments, cells were grown in batch mode for at least 3 h before switching on the feed/efflux pumps. Cultures were grown in 0.51 Sartorius bioreactors controlled by a Biostat Qplus device. A working volume of 400 ml was used. The temperature was maintained at 37°C and pH was regulated automatically to 7.0 by addition of NaOH (2.5 mol l<sup>-1</sup>). Constant stirring at 600 rpm and a volumetric aeration rate at 0.451 min<sup>-1</sup> (corresponding to 1.2 vol/vol/min) were sufficient to maintain dissolved oxygen >30% for all growth rates. Biomass was estimated from absorbance at 600 nm (1 U of absorbance was equal to 0.42 g of dry cell weight l<sup>-1</sup> in chemostat samples, and 0.35 g of dry cell weight l<sup>-1</sup> in exponential phase samples). Specific growth rates of 0.10, 0.20, 0.30 and 0.40 h<sup>-1</sup> (corresponding to doubling times of 6.9, 3.5, 2.3 and 1.7 h, respectively) were obtained in chemostat mode. The maximal growth rate (with the medium and temperature used) of 0.63 h<sup>-1</sup> (corresponding to a doubling time of 1.1 h) was obtained during the exponential phase in batch culture. In chemostat mode, cells were harvested after seven working volumes. Each culture was repeated three times to provide independent biological replicates.

### Analytical methods

Glucose and acetate concentrations were measured by HPLC coupled to a refractometer and with ultraviolet detection. The device was equipped with a Bio-Rad HPX87H column maintained at a temperature of 48°C and 5 mM H<sub>2</sub>SO<sub>4</sub> was used as the eluent, at a flow rate of 0.5 ml min<sup>-1</sup>.

### Sampling and RNA extraction

When the desired steady state or exponential growth rate was reached, samples were collected for transcriptomic



analysis; this time point was the reference time point (T0) for the half-life determination procedure. Subsequently, rifampicin ( $500 \mu\text{g ml}^{-1}$ ) was added to inhibit the initiation of transcription, and cells were harvested 0.5, 1, 1.5, 2, 3, 4, 5, 7 and 11 min later and immediately frozen in liquid nitrogen. After thawing and centrifugation steps, the cells were suspended in Tris-EDTA (TE) buffer and lysed mechanically with glass beads in the presence of phenol. Total RNA was extracted with TRI Reagent (Sigma-Aldrich) from 12 samples from the three biological replicates as follows: T0, 0.5, 2 and 5 min for the first replicate; T0, 1, 3 and 7 min for the second replicate; and T0, 1.5, 4 and 11 min for the third replicate. Potential DNA contamination was eliminated with the TURBO DNA-free™ Kit (Ambion). Total RNA concentration and integrity were measured using a Nanodrop® spectrophotometer and Agilent BioAnalyzer, respectively.

### Microarray procedures

A double-stranded complementary DNA (cDNA) synthesis kit (InvitroGen) was used to produce cDNA from 2- $\mu\text{g}$  aliquots of total RNA. Aliquots of 1  $\mu\text{g}$  of cDNA were labeled using the one color DNA labeling kit and 2  $\mu\text{g}$  of labeled cDNA were hybridized onto *E. coli* K-12 gene expression arrays (Nimblegen, Roche) for 17 h at 42°C according to the manufacturers' instructions. Arrays were washed and then scanned with a MS200 Microarray Scanner (Nimblegen, Roche). The images were analyzed with DEVA 1.2.1 software. Only raw data were used for further analyses. All array procedures were performed by the GeT-Biopuces platform (<http://get.genotoul.fr>).

### mRNA concentration and mRNA half-life determination by microarray experiments

For transcriptomic analysis, raw probe intensities (three replicates at four growth rates) were processed and analyzed with the R computing environment (26) using the affy (27) and limma (28) packages of Bioconductor. Raw data were submitted to an Robust Multi-array Average (RMA)-based background correction (29); after background correction, intra-replicate quantile normalization was performed at each growth rate. A set of probes in the background for which the ranks were roughly invariant across all 12 arrays was selected. The median value of the invariant probes intensities in each condition was used as a scaling factor for normalization between growth rate conditions. After normalization, the intensity of a transcript was calculated by an RMA-summarization procedure (29) within each condition. Intensity values were multiplied by the total RNA extraction yield (in microgram total RNA per milligram of dry cell weight) to provide the mRNA concentration value in arbitrary units per milligrams of dry cell weight. RNA extraction yields were  $13.1 \pm 2.2$ ,  $22.5 \pm 2.5$ ,  $25.2 \pm 2.6$  and  $45.2 \pm 8.1 \mu\text{g RNA/mg}$  of dry cell weight for growth rate of 0.10, 0.20, 0.40 and  $0.63 \text{ h}^{-1}$ , respectively. These data agreed well with those in the literature (7). Differences in mRNA concentration were evaluated with a modified *t*-test in conjunction with an empirical Bayes

method (30). The *P*-values were adjusted for multiple testing by the 'Benjamini and Hochberg' (BH) false discovery rate method (31). A *P*-value threshold of 1% was used for significance of differences in mRNA concentration.

For mRNA half-life determinations, 12 arrays (3 reference T0 samples and 9 time points after addition of rifampicin) were used. Only the normalization between arrays according to the invariant probes intensities was performed. In each array, transcript-specific intensity was computed as the median value of the 16 targeting probe intensities. The linear regression coefficient, *k*, of  $\ln(\text{mRNA})$  versus time (12 points) and its associated coefficient of variation (standard error of slope/estimation of slope) were calculated for each mRNA species. The determination of *k* was considered as reliable only if the associated coefficient of variation was <30%. The linear regression coefficient *k* corresponding to the degradation rate constant was inversely proportional to the mRNA half-life  $t_{1/2}$ ,  $k = \frac{\ln 2}{t_{1/2}}$ . The statistical significance of differences in half-life was evaluated using the probability value of interaction between time and growth rate in a global model of linear regression. A statistical threshold of 10% was used for adjusted *P*-values by the 'BH' false discovery rate method (31). The 4254 mRNA concentrations and 2947 confident half-lives for the four growth rates are listed in the Supplementary Table S1.

### mRNA half-life determination by qPCR

Half-life was determined by quantitative real-time PCR (qPCR) for a subset of 11 selected genes using primers designed to give 100 bp amplicons (Supplementary Figure S1). Non-labeled cDNAs prepared from T0, 2, 3, 5, 7 and 11 min samples (from the  $0.10 \text{ h}^{-1}$  growth-rate experiment) for microarray experiments were used as templates. Half-life was calculated from cycle threshold (Ct) values for samples from each time point. qPCR experiments confirmed half-life ranking between transcripts obtained with microarrays, although the absolute values were smaller (Supplementary Figure S1). One possible explanation for the discrepancy is that in qPCR experiments, mRNA half-life was determined only for a small region (around 100 bp) of the mRNA molecule, whereas, in the microarray-based technique, multiple probes were used to increase the coverage of the mRNA molecule, improving the accuracy of the determination.

### Determination of regulation coefficients

Assuming that a steady state was established, the following equation applies:

$$\begin{aligned} V_t &= k_{trans} \cdot \text{genecopynumber} \\ &= k \cdot [mRNA] + \mu \cdot [mRNA] \end{aligned} \quad (1)$$

where  $V_t$  is the transcription rate,  $k_{trans}$  the transcription rate constant, *genecopynumber* the number of copies of the gene in the cell,  $[mRNA]$  the transcript concentration,  $\mu$  the growth rate and *k* the degradation rate constant. The rate of dilution of messengers due to cellular growth can be neglected in this context because the generation time

was much longer than most of the mRNA half-lives. Assuming that  $V_i$  and  $k$  were independent, a difference in mRNA concentration between two steady states can be described as a function of the differences in the transcription and degradation rates:

$$d[mRNA] = \frac{1}{k} \cdot dV_i - \frac{V_i}{k^2} \cdot dk \quad (2)$$

This equation is mathematically equivalent to

$$1 = \frac{d \ln V_i}{d \ln [mRNA]} - \frac{d \ln k}{d \ln [mRNA]} \quad (3)$$

Two regulation coefficients were defined from Equation (3):  $\rho_D$ , the degradation coefficient, which represents the influence of decay on the mRNA concentration ( $-\frac{d \ln k}{d \ln [mRNA]}$ ) and  $\rho_T$ , the transcription coefficient, which describes the influence of the transcription rate on the mRNA concentration ( $\frac{d \ln V_i}{d \ln [mRNA]}$ ). The degradation coefficient  $\rho_D$  can be calculated as the negative value of the slope of the double-logarithmic plot of the degradation rate constant  $k$  against the initial mRNA concentration (before rifampicin treatment) in the three growth rate comparisons studied. The transcription coefficient  $\rho_T$  can then be estimated from  $\rho_D$  (values in Supplementary Table S2) as

$$\rho_D + \rho_T = 1 \quad (4)$$

$\rho_D$  and  $\rho_T$  values for transcripts included in the 5% lowest  $d \ln [mRNA]$  or  $d \ln k$  of all growth rate comparisons were not calculated. Regulatory classes named I–V were defined according to the  $\rho_D$  value:  $\rho_D < 0$ , class I;  $0 < \rho_D < 0.4$ , class II;  $0.4 < \rho_D < 0.6$ , class III;  $0.6 < \rho_D < 1$ , class IV and  $1 < \rho_D$ , class V.

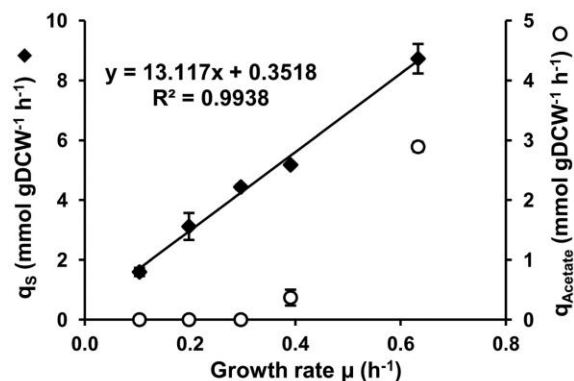
### Clustering and enrichment methods

R free statistical software ([www.r-project.org](http://www.r-project.org)) with the Ward classification method was used for clustering. The number of classes was determined graphically from the dendrogram. Functional categories enriched in transcript subgroups were determined using hypergeometric test of data using the Biological Process of Gene Ontology annotation database (GO project; <http://www.geneontology.org/>).

## RESULTS

### Steady-state cultures of *E. coli* at various growth rates

To investigate the effects of growth rate on mRNA stability, we cultured *E. coli* MG1655 in synthetic medium in glucose-limited chemostat cultures maintained at steady state at four different growth rates ( $\mu$ ): 0.10, 0.20, 0.30 and  $0.40 \text{ h}^{-1}$ . To reach a higher growth rate without any wash out of the cells, batch culture was used: in these conditions of exponential growth, the maximum growth rate was  $0.63 \text{ h}^{-1}$ . To compare cell physiology in batch and chemostat cultures, macrokinetic parameters were determined. Metabolic organization differs depending on



**Figure 1.** Effect of growth rate on the glucose consumption rate and the acetate production rate. The specific glucose consumption rate ( $q_S$  – solid diamonds) and acetate production rate ( $q_{Acetate}$  – empty circles) with their respective standard deviations are plotted against the growth rate. The  $q_S$ -associated linear regression equation is indicated on the graph. The inverse of the slope is equal to the growth yield (in g of dry cell weight per mole of glucose) which corresponds to the quantity of produced biomass per quantity of consumed glucose without maintenance associated processes. The intercept is the non-growth related maintenance (in mmol per gram of dry cell weight  $\text{h}^{-1}$ ).

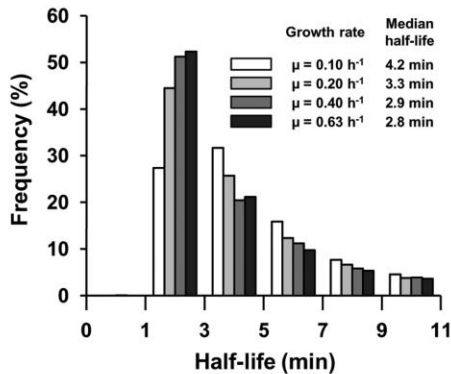
the growth rate, and acetate production (also referred to as acetate overflow) increased with growth rate (Figure 1). In our experiments, the specific production rate of acetate  $q_{Acetate}$  increased from  $0.37 \pm 0.13 \text{ mmol/g}$  of dry cell weight  $\text{h}^{-1}$  at  $\mu = 0.40 \text{ h}^{-1}$  to  $2.89 \pm 0.08 \text{ mmol/g}$  of dry cell weight  $\text{h}^{-1}$  at  $\mu = 0.63 \text{ h}^{-1}$ . These rates of acetate production are consistent with previous reports (5,9,32,33) and are necessary to maintain a cellular redox balance (2). We confirmed that acetate production started above the threshold  $q_S$  (specific glucose consumption rate) value of  $\sim 5 \text{ mmol/g}$  of dry cell weight  $\text{h}^{-1}$ , which was reached in our experiment in the  $0.30$  and  $0.40 \text{ h}^{-1}$  interval (34).

In chemostat experiments, cells have to adapt their physiology to the imposed growth rate in the presence of a limiting glucose concentration in the feeding medium. This adaptation involves adjusting the flow of glucose into the cell,  $q_S$ . Plotting  $q_S$  as a function of  $\mu$  for the five growth rates (Figure 1) showed a linear increase indicating a constant growth yield (76.2 g of dry cell weight per mole of glucose) in chemostat and batch cultures and a non-growth-related maintenance coefficient (0.35 mmol of glucose per g of dry cell weight  $\text{h}^{-1}$ ). Both values are in agreement with published data for continuous cultures of *E. coli* (2,32). These results indicate that cells in exponential growth can be considered to be at steady state in a physiological condition similar to that of cells in chemostat cultures.

### Effect of growth rate on mRNA stability

Rifampicin was added to cultures growing at 0.10, 0.20, 0.40 and  $0.63 \text{ h}^{-1}$  and the half-life of each mRNA was determined by microarray-based analysis. Half-lives of a subset of 11 selected mRNAs were measured by qPCR to confirm the findings (see 'Materials and Methods' section). Only transcript half-lives considered as reliable at all four growth rates (corresponding to 2947 genes, 70%





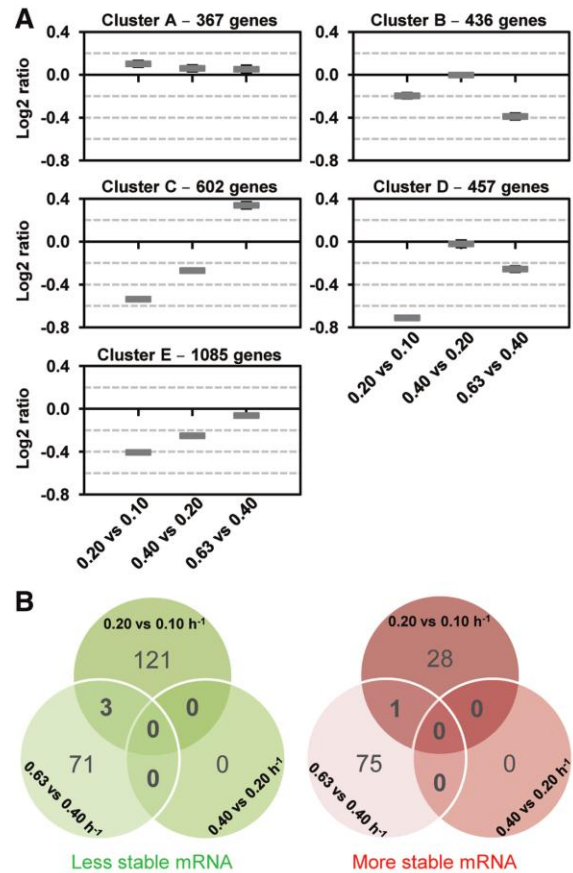
**Figure 2.** Effect of growth rate on mRNA stability. The histogram shows the relative frequency of half-lives for the four growth rates. Respective median half-lives are listed in the inset.

of the 4254 *E. coli* genes) were included in subsequent investigations (see ‘Materials and Methods’ section). The mRNAs excluded were mostly extremely stable transcripts (with average half-lives >40 min) and were not accurately quantified with the ranges of time-points used for stability measurements (see ‘Materials and Methods’ section). An analysis of the distribution of mRNA lifetimes indicated that the subset of 2947 genes selected can be used as a representative sample of the entire mRNA population in the cells (results not shown).

The range of half-lives was from 1 to 53 min. However, in all growth conditions, ~90% of mRNA half-lives were <11 min (Figure 2). The mRNA half-life distribution shifted to shorter times as the growth rate increased: the median mRNA half-life was 4.2 min at a growth rate of 0.10 h<sup>-1</sup> and 2.8 min at 0.63 h<sup>-1</sup> (Figure 2). The difference in mRNA half-life between 0.20 and 0.10 h<sup>-1</sup> was greater than that between 0.40 and 0.20 h<sup>-1</sup>; between 0.63 and 0.40 h<sup>-1</sup> the difference was the smallest (only the first two differences were significant, Mann–Whitney test for *P*-value < 0.05).

To investigate the different patterns of mRNA stability changes with increasing growth rate, we used a clustering approach to analyze 2947 half-life log<sub>2</sub> ratios (0.20 versus 0.10 h<sup>-1</sup>, 0.40 versus 0.20 h<sup>-1</sup> and 0.63 versus 0.40 h<sup>-1</sup>). Five clusters of patterns were defined, named A (367 genes), B (436 genes), C (602 genes), D (457 genes) and E (1085 genes) (Figure 3A). Most transcripts classified in clusters B, D and E, displaying decreasing half-life with increasing growth rate (with negative log<sub>2</sub> median value for each comparison between growth rates), were consistent with the general trend. These three clusters differed in the magnitude of the decrease between growth rates. Cluster A exhibited mRNA stabilization with increasing growth rate (with positive log<sub>2</sub> ratios for all growth rate comparisons). Cluster C included transcripts with a particular pattern: less stable at 0.20 than 0.10 h<sup>-1</sup> and at 0.40 than 0.20 h<sup>-1</sup> but more strongly stabilized at 0.63 than at 0.40 h<sup>-1</sup>.

We used a statistical test (*P*-value < 0.1) to identify transcripts with significantly different stabilities at different growth rates (Figure 3B): 284 transcripts (from the total

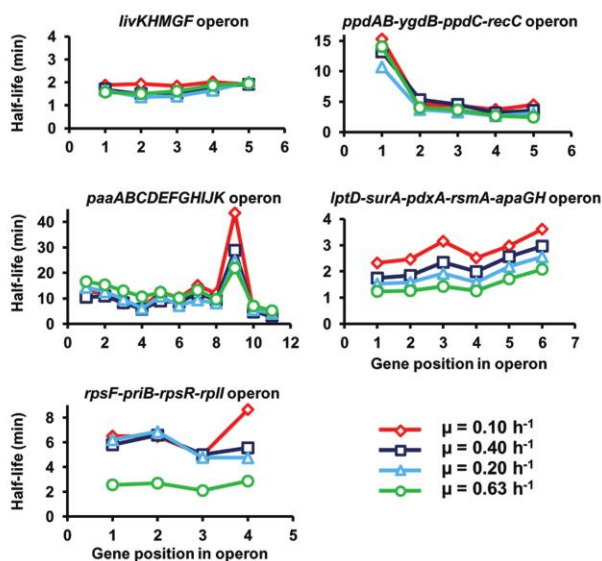


**Figure 3.** (A) Clusters of mRNA stability. Mean values and confidence intervals at 95% of the ratios of mRNA stability are given in the five clusters (cluster A, cluster B, cluster C, cluster D, cluster E) from left to right 0.20 versus 0.10 h<sup>-1</sup>, 0.40 versus 0.20 h<sup>-1</sup> and 0.63 versus 0.40 h<sup>-1</sup>. (B) Venn diagrams identify common destabilized (green diagram) and stabilized mRNAs (red diagram) between the different growth rates (*P*-value < 0.1).

of 2947) showed a significant difference in stability for at least one growth rate comparison studied. Of the 153 transcripts with significantly different half-lives between 0.20 and 0.10 h<sup>-1</sup>, 80% (124/153) were destabilized. No transcript with significantly different stability between 0.40 and 0.20 h<sup>-1</sup> was detected. Of those differing in stability between 0.63 and 0.40 h<sup>-1</sup>, equivalent numbers were more and less stable. No transcript was found to be significantly stabilized or destabilized across the whole range of growth rates. These results demonstrate that the relationship of significant changes in mRNA stability with growth rate is not simple.

The complexity of the link between mRNA stability and growth rate was confirmed by comparing stability values of cistrons within polycistronic mRNAs of operons (Figure 4). As the growth rate increased, diverse behaviors were observed: unaffected half-lives of all the cistrons (*livKHMgf* and *ppdAB-ygdB-ppdC-recC*), progressive destabilization of only one cistron (*paaI* the ninth gene in the operon *paaABCDEFGHIJK*) or of all the cistrons (*lptD-*





**Figure 4.** Half-lives of cistrons within polycistronic mRNAs. The half-lives are depicted according to the gene position in the operon for five selected operons.

*surA-pdxA-rsmA-apaGH*) and destabilization of all the cistrons but only at one growth rate (*rpsF-priB-rpsR-rplI* at the highest value of  $0.63 \text{ h}^{-1}$ ). These results identified entire polycistronic mRNAs whose stability was regulated by the growth rate. In contrast, the differential stability of cistrons within polycistronic messages does not appear generally to be affected by growth rate.

These analyses show that as the growth rate increases, the overall tendency is for mRNA to be destabilized. However, some mRNAs are stabilized and some show opposite regulations (destabilization and stabilization) in different ranges of growth rate.

#### Effect of growth rate on mRNA concentration

We investigated the contribution of mRNA stability to the control of mRNA concentration. The concentrations of the mRNAs of all 4254 genes were measured in T0 samples (before addition of rifampicin) at each growth rate in the mRNA stability experiment. First, the average mRNA concentration was higher at higher growth rates (Figure 5A). This resulted directly from a higher total RNA content per dry cell weight (see RNA extraction yields in 'Materials and Methods' section). However, mRNA as a proportion of total RNA was not significantly affected by growth rate (result not shown).

Differences in mRNA concentrations were compared between the same three growth rate pairs as for the mRNA stability study (Figure 5B). For each growth rate comparison, the number of mRNAs with upregulated concentration was at least 4-fold higher than the number of downregulated. The concentrations of 1440 transcripts were systematically upregulated and 45 systematically downregulated with increasing growth rate. The general trend of transcript concentration upregulation was confirmed by a clustering approach. Transcripts were

classified into three different groups (Figure 5C). One large cluster (cluster 3) included 3209 of the 4254 transcripts for which the concentration increased in correlation with increasing growth rate. The concentrations of the mRNAs in cluster 1 were generally similar, or decreased slightly between  $0.20$  and  $0.10 \text{ h}^{-1}$ , and between  $0.63$  and  $0.40 \text{ h}^{-1}$ . The mRNAs of cluster 2 exhibited opposite regulations: increases in concentration between  $0.20$  and  $0.10 \text{ h}^{-1}$  and  $0.63$  and  $0.40 \text{ h}^{-1}$ , but decreases between  $0.40$  and  $0.20 \text{ h}^{-1}$ .

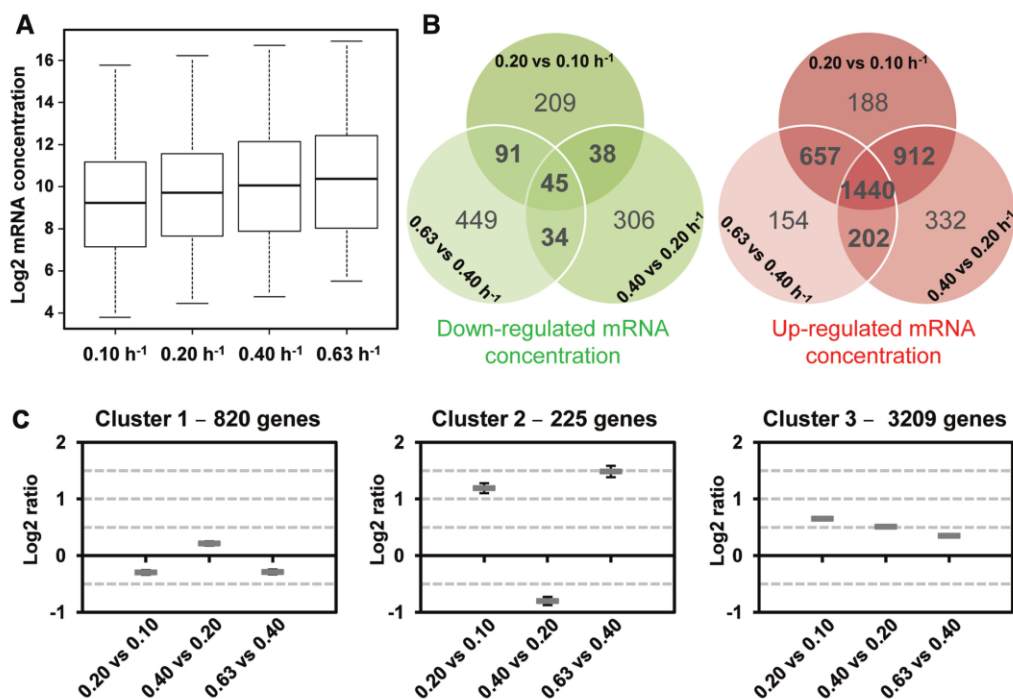
These results indicate that the concentration of a large majority of transcripts increases concomitantly with the growth rate.

#### Transcriptional and degradational controls of mRNA level

A difference in mRNA concentration between two growth rates could be the consequence of different rates of mRNA synthesis (transcription) and/or elimination. Elimination includes not only enzymatic degradation, but also, dilution due to cell growth and division. However, most mRNA half-lives were  $<11$  min in our experiments, whereas all doubling times were in the hour range. Consequently, the mRNA dilution effect was neglected in the following analyses. Thus, differences in mRNA concentration were considered to be dependent on differences in the rates of only transcription and degradation. According to Equation (1) in 'Materials and Methods' section, the transcription rate is defined as the number of mRNA molecules (in arbitrary units) per unit of time and per unit of dry mass produced from all the copies of a gene present in a cell. Changes in transcription rate of an individual gene include both variation in the activity of individual promoter (included in the transcription rate constant  $k_{trans}$ ) and variation in gene copy number. We used a formal mathematical method to evaluate the contributions of transcriptional and degradational control to determining mRNA concentration ('Materials and Methods' section and 25,35). This method allows the relative difference in transcription rate of each transcript between two steady states to be estimated from the relative differences in the degradation rate constant,  $k$ , and the transcript concentration. Two coefficients were extracted from the model:  $\rho_D$ , the degradation coefficient, and  $\rho_T$ , the transcription coefficient that are measures of the contribution of degradation and transcription, respectively, to the mRNA concentration.

$\rho_D$  and  $\rho_T$  were computed for the 2947 genes with confident mRNA half-life values for the comparisons  $0.20$  versus  $0.10 \text{ h}^{-1}$ ,  $0.40$  versus  $0.20 \text{ h}^{-1}$  and  $0.63$  versus  $0.40 \text{ h}^{-1}$ . Five regulatory classes of mRNA concentration, numbered I to V, were defined according to the  $\rho_D$  value (see 'Materials and Methods' section): two classes of mRNA for which the concentration was mainly under transcriptional control with either opposite (I) or co-directional (II) effects of degradation and transcription, two classes mainly under degradational control with either co-directional (IV) or opposite (V) effects of degradation and transcription, and one class (III) with control almost equally divided between degradation and transcription with the two acting in the same direction.





**Figure 5.** Effect of growth rate on transcript concentration. mRNA concentrations (4254 genes) are expressed in arbitrary units per g of dry cell weight. (A) Box plots of transcript concentrations at the four growth rates. Values are separated into four quartiles by horizontal bars. The central bar (in the middle of the rectangle) represents the median value. (B) Venn diagrams identify down-regulated transcript concentrations (green diagram) and up-regulated transcript concentrations (red diagram) common to the different growth rates ( $P$ -value  $< 0.01$ ). (C) Clusters of mRNA concentration. Mean values and confidence intervals at 95% of the ratios of mRNA concentration are given in the three clusters (cluster 1, cluster 2, cluster 3) from left to right 0.20 versus 0.10 h<sup>-1</sup>, 0.40 versus 0.20 h<sup>-1</sup> and 0.63 versus 0.40 h<sup>-1</sup>.

The differences in the concentrations of most mRNAs between the growth rates were due to differences in the rates of transcription (classes I and II, Figures 6). Most mRNAs were in class I and many fewer in class II; 1176 transcripts were in class I for all three comparisons between growth rates. For these mRNAs, differences in the concentrations were mainly due to different rates of transcription, and although the half-lives varied, the contribution of transcriptional control was greater than that of degradational control. A lower but significant number of mRNAs were under degradational control (classes IV and V, Figures 6). This group included two transcripts, *ybiB*, encoding a predicted transferase/phosphorylase and *yedZ* encoding a reductase that exhibited a predominant degradational control in all the three comparisons; 49 mRNAs exhibited a predominant degradational control in two comparisons and 662 in at least one. Degradational control was mostly observed at high growth rates (Figures 6): there were fewer class I and more class V transcripts for the 0.63 versus 0.40 h<sup>-1</sup> comparison than for the other comparisons.

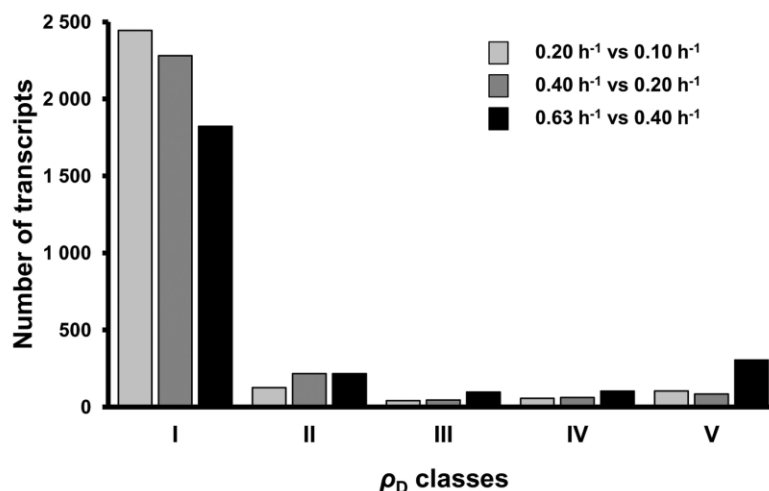
To correlate the regulatory classes of mRNA concentration to patterns of stability and expression, we superimposed the stability and concentration clusters on the  $\rho_D$  class for each mRNA (data not shown). Transcriptionally controlled mRNAs (class I) were mostly destabilized (~70% of these transcripts are in clusters B, D and E) but up-expressed (~80% are in

cluster 3) with increasing growth rate. However, class I also contained stabilized mRNAs that were mainly down-expressed with increasing growth rate. In both cases, the regulation by mRNA degradation was not sufficient to counteract regulation by transcription. By contrast, most transcripts for which degradation controlled the difference between 0.63 and 0.40 h<sup>-1</sup> (class V) were stabilized (49% of these transcripts are in cluster C) and up-expressed (84% are in cluster 3) at higher growth rates; the transcription rate for these mRNAs decreased but the mRNA concentration increased with growth rate due to a larger increase in stability.

These results clearly show that changes in rates of both transcription and degradation contribute to regulating mRNA concentration, and thus gene expression, as *E. coli* cells adapt to changing growth rate.

#### Transcriptional and degradational controls of *E. coli* metabolism

We used function enrichment tests to identify functions for which mRNA concentrations were determined more by changes in transcription than mRNA degradation as the growth rate increased (1176 transcripts of class I under the control of transcription in the three comparisons,  $P$ -value  $< 0.05$ ). Functions related to growth (DNA [replication and repair], RNA [modification and methylation, rRNA metabolic process], cell wall [phospholipid and



**Figure 6.** Distribution of transcripts in the five  $\rho_D$  classes. The degradation coefficient  $\rho_D$  was calculated for 2947 individual transcripts for the three growth rate comparisons (bars from left to right: 0.20 versus 0.10  $\text{h}^{-1}$ , 0.40 versus 0.20  $\text{h}^{-1}$  and 0.63 versus 0.40  $\text{h}^{-1}$ ). Regulatory classes named I to V were defined according to the  $\rho_D$  value:  $\rho_D < 0$ , class I;  $0 < \rho_D < 0.4$ , class II;  $0.4 < \rho_D < 0.6$ , class III;  $0.6 < \rho_D < 1$ , class IV and  $1 < \rho_D$ , class V. The signification of the  $\rho_D$  classes is: I – Mainly transcriptional control with differences in degradation and transcription having opposite effects, II – Mainly transcriptional control with differences in degradation and transcription acting in the same direction, III – Control almost equally divided between degradation and transcription, with the two acting in the same direction, IV – Mainly degradational control with differences in degradation and transcription acting in the same direction, V – Mainly degradational control with differences in degradation and transcription acting in opposite directions.

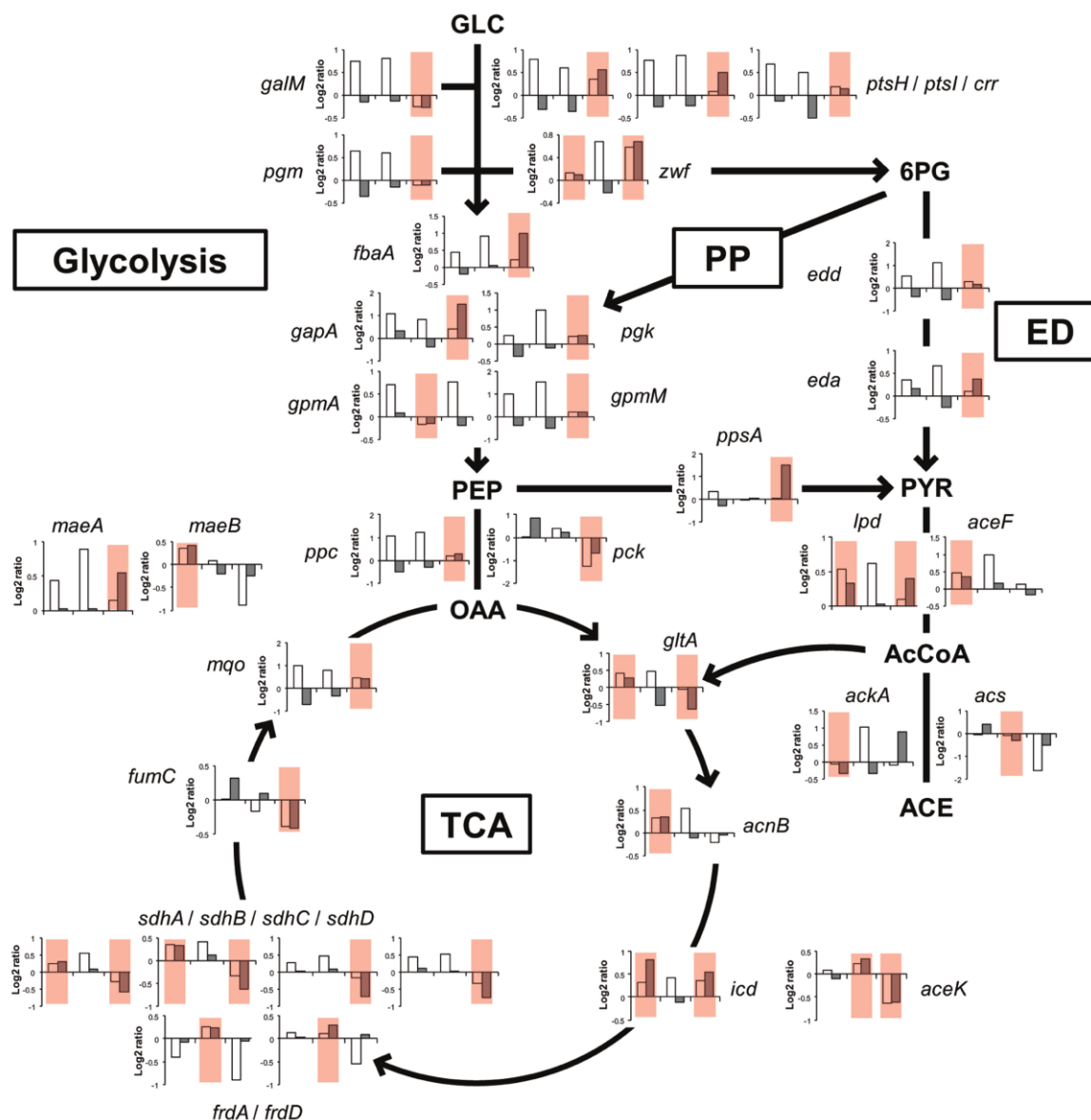
macromolecule biosynthesis, cell shape regulation], cofactor supply [ubiquinone, menaquinone, NAD]) were enriched in this group of up-expressed but destabilized mRNAs. In the group of transcriptionally-regulated transcripts which were down-expressed but stabilized (46 transcripts), we found functions related to the transport and catabolism of alternative carbon sources (sugars, *i.e.* xylose, ribose, melibiose, idose, rhamnose, gluconate; monocarboxylic acids; and peptides and amino acids). Opposite regulation of the synthesis of macromolecules and cofactors, and of the metabolism of alternative carbon substrates is consistent with their function: at high growth rates, the demand for cellular constituents and cofactors is significantly enhanced whereas the expression of an alternative carbon metabolism is unnecessary in our conditions (the presence of glucose and the absence of any alternative carbon sources). Thus, for both some essential and facultative functions, transcription is regulated rather than mRNA degradation to adapt gene expression to growth rate in *E. coli*.

Interestingly, degradation determines the difference in concentration of 34 transcripts involved in central carbon metabolism between at least one pair of growth rates (Figure 7). For seven of these transcripts, degradational control determined the difference between two pairs of growth rates. Degradational control was more frequently found in fast growing cells (0.63 versus 0.40  $\text{h}^{-1}$ ) and applied to transcripts involved in glucose transport (*ptsH*, *ptsI* and *err*), glucose inter-conversion (*galM* and *pgm*), glycolysis (*fbaA*, *gapA*, *pgk* and *gpmM*), anaplerotic pathways (*ppc*, *pck*, *ppsA* and *maeA*), and pentose phosphate and Entner–Doudoroff pathways (*zwf*, *edd* and *eda*). Destabilization was the most frequent regulatory

mode in *E. coli* cells for decreasing the concentration of transcripts of the tricarboxylic acid (TCA) cycle (*gltA*, *aceK*, *sdhA*, *sdhB*, *sdhC*, *sdhD* and *fumC*). In contrast, transcripts of the biosynthesis pathway of acetate were stabilized in fast growing cells. The mRNAs for *lpd* (belonging to pyruvate dehydrogenase complex) and *ackA* (acetate kinase), even if the latter transcript was not formally demonstrated to be under a degradational control, were indeed stabilized between 0.63 and 0.40  $\text{h}^{-1}$ . The *acs* mRNA (acetyl-CoA synthetase) responsible for acetate uptake was destabilized in the two last comparisons. These multiple regulations of the stability of mRNAs associated with the acetyl-CoA node presumably contribute to increase the net acetate production at growth rates above 0.40  $\text{h}^{-1}$ . Degradational control was also responsible for differences between one or two pairs of growth rates in the expression of enzymes involved in biosynthesis of several amino acids, purine and fatty acids, all associated with intermediates of the central carbon metabolic pathway (Supplementary Table S3).

#### Expression of the mRNA degradation machinery

The general trend was for mRNA stability to decrease as the growth rate increased. We investigated the relationship between mRNA stability and the expression of the machinery of mRNA degradation. In *E. coli*, the RNA degradosome is a multiprotein complex with four major components: RNase E (*rne*), PNPase (*pnp*), RhlB helicase (*rhlB*) and enolase (*eno*) in stoichiometric amounts. The concentrations of the transcripts encoding these proteins increased overall with increasing growth rate, although their half-lives also shortened (Figure 8A). These transcripts were under transcriptional control in most of the



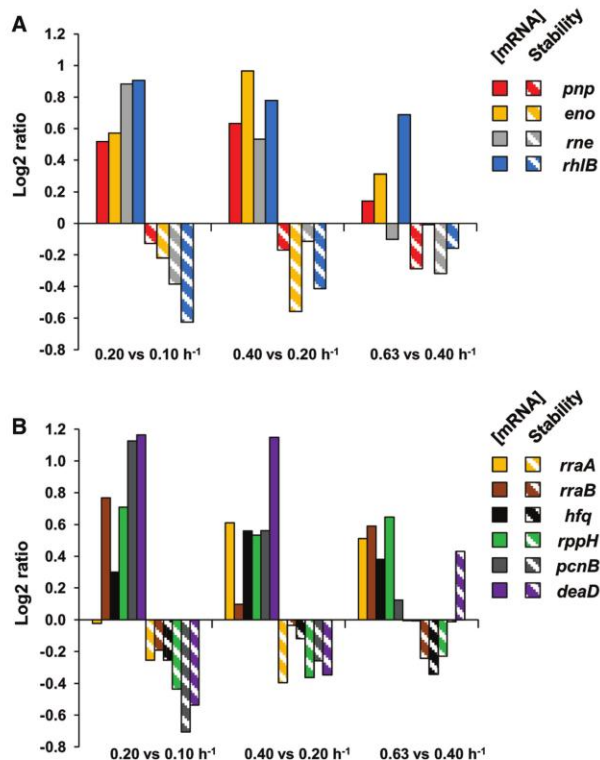
**Figure 7.** Control of the concentration of transcripts for enzymes involved in central carbon metabolism. Ratios of mRNA concentrations (white bar) and mRNA half-lives (grey bar) are represented for the three growth rate comparisons (left to right: 0.20 versus 0.10 h<sup>-1</sup>, 0.40 versus 0.20 h<sup>-1</sup> and 0.63 versus 0.40 h<sup>-1</sup>) for 34 genes encoding enzymes of the central carbon metabolism. Growth rate comparisons for which the mRNA concentration is controlled mainly by degradation (p<sub>D</sub> classes IV and V) are colored in red. The pathway names are given in boxes (PP: pentose phosphate pathway, ED: Entner–Doudoroff pathway, TCA: tricarboxylic acid cycle) and the direction of flow are indicated by arrows. Metabolites, GLC: glucose, PEP: phosphoenolpyruvate, 6PG: 6-phosphogluconate, PYR: pyruvate, OAA: oxaloacetate, AcCoA: acetyl-CoA and ACE: acetate. Enzymes, *galM*: galactose-1-epimerase, *ptsH/ptsI/crr*: PTS system, *pgm*: phosphoglucomutase, *zwf*: glucose 6-phosphate-1-dehydrogenase, *fbaA*: fructose biphosphate aldolase, *gapA*: glyceraldehyde 3-phosphate dehydrogenase, *pgk*: phosphoglycerate kinase, *gpmA*: 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase, *gpmM*: 2,3-bisphosphoglycerate-independent phosphoglycerate mutase, *edd*: phosphogluconate dehydratase, *eda*: Entner-Doudoroff aldolase, *ppc*: PEP carboxylase, *pck*: PEP carboxykinase, *ppsA*: PEP synthetase, *lpd/aceF*: pyruvate dehydrogenase, *ackA*: acetate kinase, *acs*: acetyl-CoA synthetase, *gltA*: citrate synthase, *acnB*: aconitase B, *icd*: isocitrate dehydrogenase, *aceK*: isocitrate dehydrogenase kinase/phosphatase, *sdhA/sdhB/sdhC/sdhD*: succinate dehydrogenase, *frdA/frdD*: fumarate reductase, *fumC*: fumarase C, *mqa*: malate:quinone oxidoreductase and *maeA/maeB*: malic enzymes.

growth-rate comparisons (class I). Overall, this indicates that transcription plays a major role in regulating the concentrations of the degradation machinery components. Higher transcript concentrations of proteins involved in the mRNA degradation machinery with increasing

growth rate are in agreement with the concomitant general mRNA destabilization observed.

In the comparison 0.63 versus 0.40 h<sup>-1</sup>, the increase of PNPase and enolase transcript concentrations were less marked than at slower growth rates; the concentration





**Figure 8.** Effect of growth rate on mRNA concentration and stability of the degradosome components, partners and mRNA decay associated proteins. (A) Histograms of ratios of mRNA concentrations (solid bars) and mRNA half-lives (hatched bars) are represented for the three growth rate comparisons (left to right: 0.20 versus 0.10 h<sup>-1</sup>, 0.40 versus 0.20 h<sup>-1</sup> and 0.63 versus 0.40 h<sup>-1</sup>) for degradosome components: *pnp*: PNPase, *eno*: enolase, *rhIB*: RhlB, *rne*: RNase E. (B) Histograms of ratios of mRNA concentrations (solid bars) and mRNA half-lives (hatched bars) are represented for the three growth rate comparisons (left to right: 0.20 versus 0.10 h<sup>-1</sup>, 0.40 versus 0.20 h<sup>-1</sup> and 0.63 versus 0.40 h<sup>-1</sup>) for degradosome partners and mRNA decay associated proteins: *rppH*: RppH, *rraA*: RraA, *rraB*: RraB, *hfq*: Hfq, *pcnB*: poly(A) polymerase I, *deaD*: CsdA (also known as DeaD).

of *rne* transcript was even lower at 0.63 than 0.40 h<sup>-1</sup>. Bernstein et al. reported that RNase E, PNPase and enolase are necessary for normal mRNA turnover in *E. coli* (36). The observed regulation of *rne*, *pnp* and *eno* transcript concentration is may be related to the smaller destabilization of most transcripts observed at the highest growth rates.

Concerning the RNase E regulators (Figure 8B), we observed that with changing growth rates, the concentrations of the *rraA* and *rraB* transcripts were affected, but not in a coordinated way. With respect to other enzymes associated with the mRNA degradation process, *hfq*, *rppH* and *pcnB* mRNA concentrations were always up-regulated with growth rate, which could favor mRNA degradation. For CsdA (encoded by *deaD* gene), which is involved with the degradosome at low temperature, we confirmed the presence of its transcript under normal growth conditions and showed that it was strongly stabilized at high growth rate in contrast to the general trend.

Altogether, these results show that the expression and stability of transcripts encoding proteins involved in the mRNA degradation process differ with growth rate.

## DISCUSSION

Using steady-state cultures of *E. coli* at four very different growth rates (between 0.10 and 0.63 h<sup>-1</sup>) with only glucose as the carbon source, we demonstrated that mRNA stability was specifically dependent on the growth rate. Overall, the median mRNA half-lives were 2.4 min at the highest growth rate and 4.2 min at the lowest. The increase of mRNA half-lives (1.8-fold) is small compared to the decrease of growth rate (6.3-fold). Therefore, the process of mRNA degradation is reduced but still very active even at an extremely low growth rate. Bernstein et al. (21) did not identify any significant differences in mean half-lives of mRNAs in *E. coli* grown at high and low growth rates (1.38 and 0.46 h<sup>-1</sup>). We used more time points and a wider range of growth rates. This may explain, at least in part, the different results. In addition, the experiments of Bernstein et al. were performed using different growth media (rich and minimal) whereas we used the same medium for all cultures. Nevertheless, we generally confirmed the relative ranking of half-lives within functional categories at similar growth rates (Supplementary Figure S2). We believe that our exhaustive and reliable datasets represent a significant advance compared to previous estimates of mRNA lifetimes in *E. coli*. Growth rate-dependent mRNA stability has also been observed in lactic acid bacteria and mycobacteria, with transcripts having shorter average lifetimes at higher doubling times (24,25).

Our results show that the general trend for mRNA destabilization as the growth rate increases is associated with higher concentrations of the transcripts for enzymes of the mRNA degradation process (canonical degradosome components, RppH and poly(A) polymerase I). In addition to the self-post-transcriptional regulation of RNase E and PNPase (12,37), we provide the first evidence that transcriptional regulation also contributes to the control of RNase E and PNPase expression. Transcriptional regulation is presumably a long-term adaptation for both genes/enzymes compared to post-transcriptional autoregulation. At higher growth rates, RNase E transcript concentration was the only one to decrease. Since degradosome composition and activity have generally been studied under conditions of fast growth, it would be interesting to determine if its composition and targets differ at slow growth rates. Furthermore, specific regulation of other partners of the degradosome (*RraA*, *RraB* and *CsdA*) could also contribute to modulation of the degradosome composition with growth rate. Indeed, *RraA*, *RraB* and *CsdA* were previously suggested to act on the stoichiometry of degradosome components (11,38).

Although overall mRNA stability decreases with growth rate, some mRNAs showed different patterns (for example genes of stability clusters A and C). We confirmed previous reports of stabilization of particular transcripts with increasing growth rate: the half-life of *ompA*

mRNA was 9 min at  $0.10 \text{ h}^{-1}$  and 17 min at  $0.63 \text{ h}^{-1}$  (22) and of *fumA* mRNA was 2.6 min at  $0.40 \text{ h}^{-1}$  and 5.9 min at  $0.63 \text{ h}^{-1}$  (23). Differences between the regulation of transcript half-life could be related to accessibility of RNase E sensitive sites. Ribosome binding can either protect against RNase E attack by shielding cleavage sites (39,40) or favor RNase E attack by exposing cleavage sites (41). Similarly, the association of regulatory RNA-binding proteins and/or small RNAs with the leader transcript can either stabilize mRNA (42) or trigger RNase E-dependent degradation (43). It would be interesting to determine how these various regulatory mechanisms contribute to the variability of mRNA stability. In addition, our results identified growth rate effect on the stability of entire polycistronic mRNAs. On the contrary, differential transcript stability within operon does not appear generally to be growth rate dependent. Differences in translation between cistrons within polycistronic mRNA could contribute to differential stability although the possible influence of *cis*-acting RNA elements cannot be excluded. Our results on the specific regulation of mRNA stability highlight new targets for mechanistic studies.

We demonstrated that both transcription and mRNA stability contribute to the adaptation of mRNA concentration, and thus gene expression, to growth rate. For 40% of the transcripts for which the regulation of concentration was characterized, the control was systematically transcriptional with counteracting mRNA stability (class I). Concentrations of messenger RNAs encoding growth-related proteins were up-regulated with increasing growth rate due to increased transcription outweighing increased degradation. In contrast, the mRNA concentrations of genes encoding proteins involved in transport and catabolism of alternative carbon sources were down-regulated at higher growth rates, mainly by lower rates of transcription outweighing transcript stabilization. This down-regulation is associated with catabolic repression: the cAMP concentration decreases at high growth rates (33,44). The stabilization of these transcripts limits the decline in their mRNA concentration that would result from lower rates of transcription and thus could provide a selective advantage to the bacterium in case of future alternative carbon substrate availability.

We found that the expression of numerous genes was under degradational control. Indeed, for 662 mRNAs, variation in concentrations was determined by changes in degradation in at least one pair of growth rates. Only two transcripts were under major degradational control in all three growth rate comparisons. As the growth rate increases, control of transcript concentration by degradation does not therefore constitute a simple and general response; degradational control appears to be a specific and flexible adaptive response. The number of mRNAs under mainly degradational control may, however, be underestimated by our analysis since many stable transcripts were excluded from the dataset due to the difficulty in accurately measuring the half-life of long-lived mRNAs. Interestingly, many transcripts controlled by degradation were involved in central carbon metabolism. This essential metabolism provides energy and anabolic precursors

required for *E. coli* survival and growth. More particularly, degradational control was most often evidenced at high growth rates ( $0.63$  versus  $0.40 \text{ h}^{-1}$ ) when the high rate of glucose uptake resulted in acetate overflow. This phenomenon is a consequence of the imbalance between glucose uptake and TCA cycle capacity (45), and has been shown to be associated with a substantial reorganization of gene expression involving down regulation of TCA genes and acetyl-CoA synthetase (2,5,33). Here, we show that the regulation of mRNA stability makes a major contribution to this reprogramming. The mechanism(s) governing this global response remains to be elucidated.

We investigated the conditions under which *E. coli* cells used degradational rather than transcriptional control to modify mRNA concentration. Transcripts under degradational control were mainly up-expressed and stabilized with increasing growth rate (classes IV and V). For transcripts for which transcription and degradation regulation are co-directional (classes IV), transcription is up-regulated with increasing growth rate; mRNA stabilization may allow faster increases in transcript levels when for example transcription is maximally up-regulated. For mRNAs for which transcription and degradation regulation act in opposite directions (class V), transcription is down-regulated. As the growth rate increases, it was shown that an adjustment in RNA polymerase partitioning occurred to favor synthesis of stable RNAs over that of mRNAs (1,46). Consequently, for mRNAs of cluster V, stabilization may counteract and postpone the decrease of their concentration due to RNA polymerase allocation towards stable RNA synthesis.

We report that global and individual mRNA stabilities differ between steady-state cultures of *E. coli* growing at different growth rates. This genome-scale analysis shows that regulation of mRNA degradation makes a significant contribution, in addition to transcription, to the control of mRNA concentration. These findings need to be taken into account by bacterial physiologists when interpreting gene expression differences for the same strain in different growth conditions. They are also of interest to molecular biologists when comparing gene expression between wild-type and mutant strains. Moreover, analysis of mRNA half-life variation can provide potential new targets for mechanistic studies of mRNA decay. In our chemostat experiments, *E. coli* physiology was fully adapted to the environment and imposed growth rate. Further studies are required to investigate the response of mRNA stability and its contribution to the regulation of gene expression during dynamic changes in growth rate.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online, including [47].

## FUNDING

Université de Toulouse and Région Midi-Pyrénées. Funding for open access charge: Université de Toulouse and Région Midi-Pyrénées.



Conflict of interest statement. None declared.

## REFERENCES

- Ehrenberg, M., Bremer, H. and Dennis, P.P. (2013) Medium-dependent control of the bacterial growth rate. *Biochimie*, **95**, 643–658.
- Vemuri, G.N., Altman, E., Sangurdekar, D.P., Khodursky, A.B. and Eiteman, M.A. (2006) Overflow metabolism in *Escherichia coli* during steady-state growth: transcriptional regulation and effect of the redox ratio. *Appl. Environ. Microbiol.*, **72**, 3653–3661.
- Ishii, N., Nakahigashi, K., Baba, T., Robert, M., Soga, T., Kanai, A., Hirasawa, T., Naba, M., Hirai, K., Hoque, A. et al. (2007) Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. *Science*, **316**, 593–597.
- Dressaire, C., Redon, E., Milhem, H., Besse, P., Loubiere, P. and Coccain-Bousquet, M. (2008) Growth rate regulated genes and their wide involvement in the *Lactococcus lactis* stress responses. *BMC Genomics*, **9**, 343.
- Nahku, R., Valgepea, K., Lahtvee, P.-J., Erm, S., Abner, K., Adamberg, K. and Vilu, R. (2010) Specific growth rate dependent transcriptome profiling of *Escherichia coli* K12 MG1655 in accelerostat cultures. *J. Biotechnol.*, **145**, 60–65.
- Berthoumieux, S., de Jong, H., Baptist, G., Pinel, C., Ranquet, C., Ropers, D. and Geiselmann, J. (2013) Shared control of gene expression in bacteria by transcription factors and global physiology of the cell. *Mol. Syst. Biol.*, **9**, 634.
- Bremer, H. and Dennis, P.P. (1996) Modulation of chemical composition and other parameters of the cell by growth rate. 2nd edn. In: Neidhardt FC (ed.), *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, ASM Press, Washington, pp. 1553–1569.
- Martínez-António, A. (2011) *Escherichia coli* transcriptional regulatory network. *New. Biol.*, **1**, 21–33.
- Ihssen, J. and Egli, T. (2004) Specific growth rate and not cell density controls the general stress response in *Escherichia coli*. *Microbiol. Read. Engl.*, **150**, 1637–1648.
- Carpousis, A.J., Luisi, B.F. and McDowall, K.J. (2009) Endonucleolytic initiation of mRNA decay in *Escherichia coli*. *Prog. Mol. Biol. Transl. Sci.*, **85**, 91–135.
- Carpousis, A.J. (2007) The RNA degradosome of *Escherichia coli*: an mRNA-degrading machine assembled on RNase E. *Annu. Rev. Microbiol.*, **61**, 71–87.
- Jain, C. and Belasco, J.G. (1995) RNase E autoregulates its synthesis by controlling the degradation rate of its own mRNA in *Escherichia coli*: unusual sensitivity of the rne transcript to RNase E activity. *Genes Dev.*, **9**, 84–96.
- Kaberlin, V.R. and Lin-Chao, S. (2009) Unraveling new roles for minor components of the *E. coli* RNA degradosome. *RNA Biol.*, **6**, 402–405.
- Mackie, G.A. (2013) RNase E: at the interface of bacterial RNA processing and decay. *Nat. Rev. Microbiol.*, **11**, 45–57.
- Morita, T., Maki, K. and Aiba, H. (2005) RNase E-based ribonucleoprotein complexes: mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. *Genes Dev.*, **19**, 2176–2186.
- Deana, A., Celesnik, H. and Belasco, J.G. (2008) The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal. *Nature*, **451**, 355–358.
- Blum, E., Carpousis, A.J. and Higgins, C.F. (1999) Polyadenylation promotes degradation of 3'-structured RNA by the *Escherichia coli* mRNA degradosome *in vitro*. *J. Biol. Chem.*, **274**, 4009–4016.
- Mohanty, B.K. and Kushner, S.R. (1999) Analysis of the function of *Escherichia coli* poly(A) polymerase I in RNA metabolism. *Mol. Microbiol.*, **34**, 1094–1108.
- Mohanty, B.K. and Kushner, S.R. (2006) The majority of *Escherichia coli* mRNAs undergo post-transcriptional modification in exponentially growing cells. *Nucleic Acids Res.*, **34**, 5695–5704.
- Pato, M.L. and von Meyenburg, K. (1970) Residual RNA synthesis in *Escherichia coli* after inhibition of initiation of transcription by rifampicin. *Cold Spring Harb. Symp. Quant. Biol.*, **35**, 497–504.
- Bernstein, J.A., Khodursky, A.B., Lin, P.H., Lin-Chao, S. and Cohen, S.N. (2002) Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl Acad. Sci. USA*, **99**, 9697–9702.
- Nilsson, G., Belasco, J.G., Cohen, S.N. and von Gabain, A. (1984) Growth-rate dependent regulation of mRNA stability in *Escherichia coli*. *Nature*, **312**, 75–77.
- Lin, H.H., Lin, C.H., Hwang, S.M. and Tseng, C.P. (2012) High growth rate downregulates fumA mRNA transcription but is dramatically compensated by its mRNA stability in *Escherichia coli*. *Curr. Microbiol.*, **64**, 412–417.
- Rustad, T.R., Minch, K.J., Brabant, W., Winkler, J.K., Reiss, D.J., Baliga, N.S. and Sherman, D.R. (2013) Global analysis of mRNA stability in *Mycobacterium tuberculosis*. *Nucleic Acids Res.*, **41**, 509–517.
- Dressaire, C., Picard, F., Redon, E., Loubiere, P., Queinnee, I., Girbal, L. and Coccain-Bousquet, M. (2013) Role of mRNA Stability during bacterial adaptation. *PLoS One*, **8**, e59059.
- R Core Team (2012) R: A Language and Environment for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/> (6 November 2013, date last accessed).
- Gautier, L., Cope, L., Bolstad, B.M. and Irizarry, R.A. (2004) affy—analysis of affymetrix GeneChip data at the probe level. *Bioinformatics*, **20**, 307–315.
- Smyth, G.K. (2005) Limma: linear models for microarray data. In: Gentleman, R., Carey, V.J., Huber, W., Irizarry, R.A. and Dudoit, S. (eds), *Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Statistics for Biology and Health*. Springer, New York, pp. 397–420.
- Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U. and Speed, T.P. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostat.*, **4**, 249–264.
- Smyth, G.K. (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.*, **3**, article 3.
- Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.*, **57**, 289–300.
- Nanchen, A., Schicker, A. and Sauer, U. (2006) Nonlinear dependency of intracellular fluxes on growth rate in miniaturized continuous cultures of *Escherichia coli*. *Appl. Environ. Microbiol.*, **72**, 1164–1172.
- Valgepea, K., Adamberg, K., Nahku, R., Lahtvee, P.J., Arike, L. and Vilu, R. (2010) Systems biology approach reveals that overflow metabolism of acetate in *Escherichia coli* is triggered by carbon catabolite repression of acetyl-CoA synthetase. *BMC Syst. Biol.*, **4**, 166.
- Renilla, S., Bernal, V., Fuhrer, T., Castaño-Cerezo, S., Pastor, J.M., Iborra, J.L., Sauer, U. and Cánovas, M. (2012) Acetate scavenging activity in *Escherichia coli*: interplay of acetyl-CoA synthetase and the PEP-glyoxylate cycle in chemostat cultures. *Appl. Microbiol. Biotechnol.*, **93**, 2109–2124.
- Redon, E., Loubière, P. and Coccain-Bousquet, M. (2005) Role of mRNA stability during genome-wide adaptation of *Lactococcus lactis* to carbon starvation. *J. Biol. Chem.*, **280**, 36380–36385.
- Bernstein, J.A., Lin, P.H., Cohen, S.N. and Lin-Chao, S. (2004) Global analysis of *Escherichia coli* RNA degradosome function using DNA microarrays. *Proc. Natl Acad. Sci. USA*, **101**, 2758–2763.
- Carzaniga, T., Briani, F., Zangrossi, S., Merlino, G., Marchi, P. and Deho, G. (2009) Autogenous regulation of *Escherichia coli* polynucleotide phosphorylase expression revisited. *J. Bacteriol.*, **191**, 1738–1748.
- Gao, J., Lee, K., Zhao, M., Qiu, J., Zhan, X., Saxena, A., Moore, C.J., Cohen, S.N. and Georgiou, G. (2006) Differential modulation of *E. coli* mRNA abundance by inhibitory proteins that alter the composition of the degradosome. *Mol. Microbiol.*, **61**, 394–406.
- Kaberlin, V.R. and Bläsi, U. (2006) Translation initiation and the fate of bacterial mRNAs. *FEMS Microbiol. Rev.*, **30**, 967–979.
- Richards, J., Luciano, D.J. and Belasco, J.G. (2012) Influence of translation on RppH-dependent mRNA degradation in *Escherichia coli*. *Mol. Microbiol.*, **86**, 1063–1072.

2472 *Nucleic Acids Research*, 2014, Vol. 42, No. 4

41. Tsai, Y.C., Du, D., Dominguez-Malfavon, L., Dimastrogiovanni, D., Cross, J., Callaghan, A.J., Garcia-Mena, J. and Luisi, B.F. (2012) Recognition of the 70S ribosome and polysome by the RNA degradosome in *Escherichia coli*. *Nucleic Acids Res.*, **40**, 10417–10431.
42. Yakhnin, A.V., Baker, C.S., Vakulskas, C.A., Yakhnin, H., Berezin, I., Romeo, T. and Babitzke, P. (2013) CsrA activates flhDC expression by protecting flhDC mRNA from RNase E-mediated cleavage. *Mol. Microbiol.*, **87**, 851–866.
43. Morita, T. and Aiba, H. (2011) RNase E action at a distance: degradation of target mRNAs mediated by an Hfq-binding small RNA in bacteria. *Genes Dev.*, **25**, 294–298.
44. Ferenci, T. (2008) Bacterial physiology, regulation and mutational adaptation in a chemostat environment. *Adv. Microb. Physiol.*, **53**, 169–229.
45. Wolfe, A.J. (2005) The acetate switch. *Microbiol. Mol. Biol. Rev.*, **69**, 12–50.
46. Jin, D.J., Cagliero, C. and Zhou, Y.N. (2012) Growth rate regulation in *Escherichia coli*. *FEMS Microbiol. Rev.*, **36**, 269–287.
47. Serres, M.H., Goswami, S. and Riley, M. (2004) GenProtEC: an updated and improved analysis of functions of *Escherichia coli* K-12 proteins. *Nucleic Acids Res.*, **32**, D300–D302.

Genes	Microarray half-life	qPCR half-life
<i>lpp</i>	43.4	30.5
<i>ompC</i>	20.2	11.5
<i>ompA</i>	9.1	4.6
<i>acs</i>	8.0	3.7
<i>glgS</i>	7.9	8.7
<i>infA</i>	4.4	2.5
<i>ihfB</i>	4.4	4.8
<i>edd</i>	4.3	2.0
<i>maeA</i>	2.1	2.9
<i>ppsA</i>	1.9	1.6
<i>sdhD</i>	1.6	1.3

**Figure S1.** Comparison of mRNA half-lives measured by the microarray-based technique and by qPCR. mRNA half-lives are expressed in minutes.

Gene class	This study		Bernstein et al. 2002	
	M9 medium 37°C / $\mu = 0.63 \text{ h}^{-1}$		M9 medium 30°C / $\mu = 0.47 \text{ h}^{-1}$	
	Number of genes in category	Median half-life	Number of genes in category	Mean half-life
Amino acid synthesis	155	2.0	96	3.8
Macromolecule synthesis	180	2.1	206	4.4
Nucleotide synthesis	33	2.2	30	4.6
Synthesis of cofactors/carriers	148	2.1	106	4.8
Cell division	53	2.1	37	5.0
Macromolecule degradation	74	2.3	60	5.0
Fatty acid synthesis	16	2.4	20	5.1
Regulation	334	2.8	49	5.1
Adaptation to stress	94	3.1	25	5.1
Ribosome constituents	53	2.4	50	5.4
Central intermediary metabolism	273	3.2	170	5.6
Protection responses	104	3.0	44	6.0
Transport/Binding proteins	429	3.0	393	6.2
Energy metabolism, carbon	165	3.7	173	6.3

**Figure S2.** Gene function and mRNA stability. This table compares the median half-lives of gene-family memberships that we measured with the results published by Bernstein et al, (Bernstein et al, 2002, Proc. Natl. Acad. Sci. 99: 9697–9702) (21). In both studies half-lives were measured in *E. coli* cells in the exponential phase of growth on minimal medium (but at different temperatures in the two studies). The functional classification is based on Riley GenProtEC annotation (Serres et al, 2004, Nucleic Acids Res. 32: D300-D302) (47).



Functional category	Two comparisons	One comparison
<b>Fatty acid biosynthesis</b>	-	<i>accA, acpP, fabF, fabG, fabI, fabR, fabZ, lipA</i>
<b>Purine nucleotide biosynthesis</b>	-	<i>adk, atpE, atpF, atpH, flil, guaA, guaB, kdpB, purA, purB, purC, purD, purF, purM, purN, purT</i>
<b>Amino acid biosynthesis</b>		
- Val/Leu/Ile	<i>leuB, leuC</i>	<i>asd, ilvB, ilvC, ilvD, ilvE, ilvN</i>
- His	<i>hisB, hisH, hisJ</i>	<i>hisD, hisI</i>
- Arg	-	<i>argA, argC, argE, argF, argG, carB</i>
- Cys/Met	-	<i>cysH, cysI, cysJ, cysK, iscS, metB, metC, metE, metJ, mmuM, mtn</i>
- Ser/Gly	-	<i>serA, serC, glyA, sdaA</i>
- Asp/Asn/Thr/Lys	<i>aspC</i>	<i>asd, asnA, asnB, dapA, dapD, thrA, thrB</i>

**Table 1.** Functional categories enriched in the group of mRNAs with concentration under mainly degradational control (p-value < 0.07). For each category, the associated genes are listed according to the number of comparisons (one or two) in which the degradational control is present.





# **CHAPITRE II**

## **Identification des déterminants de la stabilité des ARNm à différents taux de croissance**



Le Chapitre I a permis de démontrer que la dégradation des ARNm occupe un rôle important dans la régulation de leur concentration en réponse à un changement de taux de croissance. Une grande disparité de stabilité selon les ARNm a pu être mise en évidence avec des  $t_{1/2}$  de l'ordre de la minute et d'autres supérieurs à 50 min. À ce jour, si de multiples mécanismes moléculaires de régulation de la dégradation des ARNm ont été identifiés pour quelques transcrits (Picard et al., 2009), les déterminants de la stabilité des messagers à l'échelle omique chez la bactérie *E. coli* restent encore mal connus en raison de données anciennes et d'études contradictoires. De plus, ces études n'ont jamais évalué si ces déterminants varient en fonction de changements environnementaux ou de taux de croissance.

Au travers d'un modèle d'analyse de covariance, les données de stabilité des ARNm mesurées lors de la première partie de cette thèse associées à de multiples informations compilées pour chacun des messagers (concentration, longueur, pourcentage de GC, catégorie fonctionnelle...) ont permis d'identifier les déterminants du  $t_{1/2}$  des transcrits et de les hiérarchiser. Une recherche de motifs stabilisateurs ou déstabilisateurs au sein des 5'UTR des messagers a également été conduite en collaboration avec Christine Gaspin, Annick Moisan et Hélène Chiapello de l'équipe « Statistiques et Algorithmique pour la Biologie » à l'INRA de Castanet-Tolosan. Enfin, l'impact des régulations de la dégradation des ARNm sur la concentration de protéine a également été étudié dans ce chapitre. Les données de protéomique (obtenues à partir des mêmes cultures bactériennes que celles utilisées pour les études de stabilome et transcriptome) ont été générées sur la plateforme « Cambridge Centre for Proteomics » à Cambridge dans le cadre d'une seconde collaboration avec l'équipe de Raivo Vilu du « Competence Center of Food and Fermentation Technologies » à Tallin en Estonie. L'ensemble de ces analyses a été conduit pour les quatre taux de croissance étudiés au cours du premier chapitre (0.10, 0.20, 0.40 et 0.63 h<sup>-1</sup>) afin d'évaluer si ces déterminants évoluent en fonction de ce paramètre.

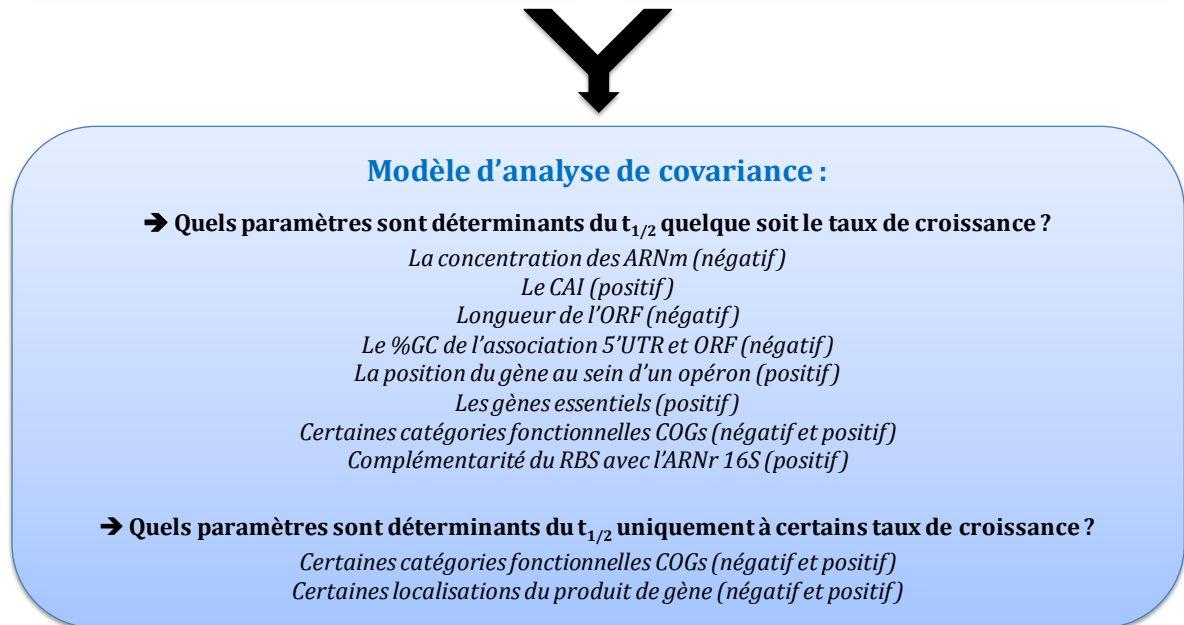
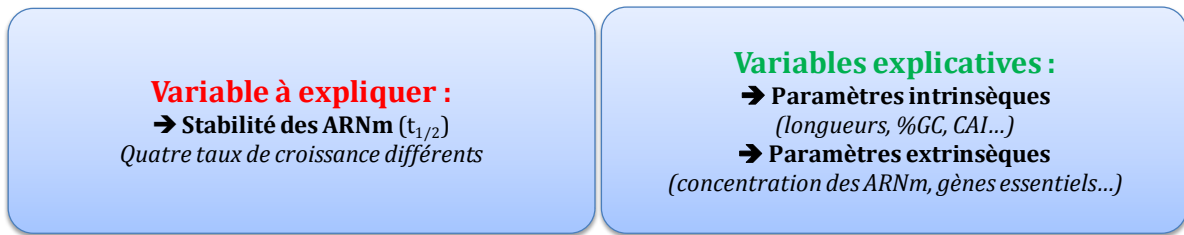
Les résultats de ces expériences ont premièrement démontré que les modèles d'analyse de covariance utilisés permettent d'expliquer entre 51 % et 74 % de la variance des  $t_{1/2}$  des ARNm pour chaque taux de croissance étudié. Ceci justifie l'approche et indique que les paramètres inclus dans nos modèles sont des déterminants majeurs de la stabilité des messagers. La concentration des ARNm apparaît notamment comme le paramètre ayant le plus de poids dans la détermination du  $t_{1/2}$  d'un transcrit

suggérant une part importante de dégradation non ciblée des messagers. Globalement, lorsqu'on compare l'ensemble des ARNm à un taux de croissance donné, plus un transcrit est concentré dans la cellule, plus il est dégradé. De plus, nos résultats démontrent que la capacité maximale de dégradation des ARNm est accentuée avec le taux de croissance. En effet, lorsqu'on compare tous les transcrits aux différents taux de croissance, pour une concentration donnée, la valeur maximale de la constante de vitesse de dégradation  $k$  augmente lorsque les cellules se multiplient plus rapidement (plus généralement, le « front » maximal de dégradation augmente quelle que soit la concentration). D'autres caractéristiques comme le CAI, la longueur de l'ORF, le pourcentage de GC de l'association 5'UTR et ORF ou la catégorie fonctionnelle à laquelle appartiennent les ARNm participent également à la détermination de leur  $t_{1/2}$  indépendamment du taux de croissance des cellules. De plus, pour la grande majorité des ARNm polycistroniques, la stabilité de leur dernier cistron est supérieure à celle du premier quelle que soit la vitesse de croissance de la bactérie. Ce résultat soutient le modèle de dégradation 5' vers 3' de l'ARNm chez *E. coli*. La recherche de motifs à l'échelle omique a démontré que les transcrits avec les RBS dont les séquences sont les plus complémentaires de l'extrémité 3' de l'ARNr 16S sont généralement les plus stables soulignant le lien entre traduction et stabilité de l'ARNm. Seules certaines fonctions ou localisations des produits de gènes semblent être déterminantes de la stabilité des transcrits de façon dépendante du taux de croissance. Enfin, il semblerait que l'impact de la stabilité des ARNm sur la concentration des protéines soit faible et dépendant du taux de croissance. En effet, de façon surprenante, des corrélations opposées entre ces deux paramètres ont été observées à 0.10 et 0.63 h<sup>-1</sup>. Nos modèles ont par ailleurs confirmé que la concentration de l'ARNm et le CAI sont les déterminants ayant le plus de poids dans la détermination du niveau des protéines.

En conclusion, même s'il reste à découvrir d'autres caractéristiques qui contribuent aussi au niveau de stabilité des ARNm, cette étude a permis d'identifier et de hiérarchiser, parmi les paramètres étudiés, les déterminants majeurs du temps de demi-vie des messagers

Cette étude a été rédigée sous la forme d'un article qui sera très prochainement soumis dans une revue à comité de lecture.

## Détermination des principaux déterminants du $t_{1/2}$ des ARNm



## Le $t_{1/2}$ des ARNm est-il à son tour déterminant de la concentration des protéines ?





## Identification of mRNA half-life determinants in *Escherichia coli* cells at different growth rates

Thomas Esquerre<sup>a,b,c,d</sup>, Annick Moisan<sup>e</sup>, H el ene Chiapello<sup>e</sup>, Liisa Arike<sup>f</sup>, Raivo Vilu<sup>f,g</sup>, Christine Gaspin<sup>e</sup>, Muriel Cocaign-Bousquet<sup>a,b,c,\*</sup> and Laurence Girbal<sup>a,b,c</sup>

<sup>a</sup> Universit e de Toulouse; INSA, UPS, INP; LISBP, 135, avenue de Rangueil, 31077 Toulouse, France

<sup>b</sup> INRA, UMR792 Ing enierie des syst emes biologiques et des proc ed es, 31400 Toulouse, France

<sup>c</sup> CNRS, UMR5504, 31400 Toulouse, France

<sup>d</sup> Laboratoire de Microbiologie et G en etique Mol eculaires, UMR5100, Centre National de la Recherche Scientifique et Universit e Paul Sabatier, 118 Route de Narbonne, 31062, Toulouse, France.

<sup>e</sup> INRA, UR 875, F-31320 Castanet-Tolosan, France.

<sup>f</sup> Competence Center of Food and Fermentation Technologies, Akadeemia tee 15A, 12618 Tallinn, Estonia.

<sup>g</sup> Tallinn University of Technology, Department of Chemistry, Akadeemia tee 15, 12618 Tallinn, Estonia

\* To whom correspondence should be addressed. Tel: (33) 561 559 465; Fax: (33) 561 559 400; Email: cocaign@insa-toulouse.fr

### ABSTRACT

mRNA degradation is one of the post-transcriptional mechanism allowing adjustment of mRNA concentration in cells in response to environmental changes and subsequent growth rate adaptation. However, to date, determinants of messenger half-life are poorly understood at the genomic scale and have never been investigated at different growth rates. Using the most recent and largest dataset of mRNA half-lives in *E. coli* at four growth rates, linear covariance models identified

and ranked intrinsic and extrinsic transcript parameters involved in messenger lifetime determination. The determination coefficients of the models were relatively high indicating that the selected parameters explained the major part of mRNA stability. mRNA concentration was found to be the major determinant of mRNA stability: more an mRNA is concentrated, shorter the half-life is. This result suggests that, *in vivo*, a significant part of messenger degradation is independent of their intrinsic features at the omic level. CAI, ORF length, GC content in 5'UTR + ORF and some functional categories were also found as explicative parameters independently of the growth rate. Only particular functional categories and gene product locations were found to be growth rate-dependent and were thus involved in growth-related adaptation process. Looking for stabilizing/destabilizing sequence motifs, we confirmed at the omic level that ribosome binding site with increased complementarity with the 16S rRNA conferred more stability to mRNA probably by favoring translation initiation. Finally, mRNA stability was demonstrated to have in turn small impact on protein level compared to mRNA concentration and CAI. This study provides the most complete genome-wide analysis of mRNA half-life determinants in *E. coli* opening the discussion of new mechanistic role of regulation of mRNA lifetime.

### INTRODUCTION

Cell adaptation to changing environment involves regulation of gene expression and in particular of the messenger (mRNA) level which depends on variations of both mRNA synthesis and degradation. In a recent study, mRNA decay regulation has been demonstrated to significantly affect the level of mRNA as the growth rate changed (1). In the literature, several molecular mechanisms regulating mRNA stability have been described involving mRNA binding molecules (small RNA, protein and/or metabolite). For example, binding of SgrS sRNA coupled with Hfq and RNase E is known to destabilize *ptsG* mRNA (2). CsrA protein favors expression of *flhDC* transcript by preventing its degradation by RNase E (3). Riboswitches can either sequester or release RNase E cleavage sites like in *lysC* messenger upon lysine binding (4). Association of ribosomes (ribonucleo-protein complexes) with messengers was also reported to have a dual role, impeding or accelerating bacterial mRNA degradation (5, 6). These mechanisms indirectly underlined the role played by mRNA sequence determinants such as specific recognition sequences (7) as well as

secondary structures (8). These multiple mechanisms were generally described for particular transcripts and consequently represent mRNA-specific determinants. However, global determinants of mRNA stability related to widespread regulation targeting a large number of messengers were not well characterized yet. This global regulation and associated determinants could be discovered in large-scale data of mRNA half-lives available now with the development of the “omic” era. However, such approaches remained quite rare and less than ten studies dedicated to mRNA half-life determinants were found in the literature for all micro-organisms (i.e. *Escherichia coli* (9), *Saccharomyces cerevisiae* (10), *Bacillus subtilis* (11), *Lactococcus lactis* (12, 13), *Halobacterium salinarum* (14), *Bacillus cereus* (15) or more recently *Mycobacterium tuberculosis* (16)). From these studies, it appears large disparities in methods for data collection or determinant statistical identification leading to important differences between the different models and even inconsistent results for the same micro-organism. For example, the effect of length on the mRNA stability was contradictory: no length effect in *S. cerevisiae*, *B. subtilis* and *cereus*, *H. salinarum* and *M. tuberculosis* or negative effect in *L. lactis* (10–16). In the particular case of *E. coli*, mRNA stability data were produced in 2002 (9), unlike other micro-organisms which have benefited of more advanced methods, and were reanalyzed by several groups leading to inconsistent results with for example negative or no effect of length on the mRNA stability depending on the study (9, 17). Moreover, whichever their nature, the influence of various parameters on mRNA stability has been almost systematically studied through individual linear correlation analyses which did not allow to rank determinants one from the others. Furthermore, global determinants of messenger lifetimes were generally investigated for only one growth condition although global variations in mRNA stability were observed at different growth rates in *E. coli* (1), *M. tuberculosis* (16) and *L. lactis* (13). Consequently, we still do not know whether messenger stability determinants vary with this major parameter.

In this study, we have identified the global determinants of transcript lifetimes in the model bacterium *E. coli* at various growth rates using the largest and the most recent dataset available for this micro-organism (1). All mRNA half-lives were modeled statistically taking into account both qualitative and quantitative biological parameters in order to make a hierarchy of the determinants according to their global effect on mRNA stability. This allowed the influence of growth rate on the contribution

of each determinant to be analyzed. At last, the role of mRNA stability regulation in the protein level determination was also investigated from new proteomic dataset produced at various growth rates. Overall, the results provide invaluable guidelines on the global mechanisms involved in regulating mRNA stability in *E. coli* and their interplays with changing growth rate.

## **MATERIALS AND METHODS**

### **Data compilation**

Stability and expression data of 2947 genes of *E. coli* at four growth rates ( $\mu = 0.10, 0.20, 0.40$  and  $0.63 \text{ h}^{-1}$ ) were obtained from (1). ORF lengths and sequences were retrieved from the Ecocyc database (<http://www.ecocyc.org/>). 5'UTR length and sequence of each mRNA was experimentally determined in (18). When several alternative Transcription Starting Sites (TSS) were described for one transcript, the TSS located the furthest from the site of translation initiation was selected to avoid any loss of sequence information. The nature of the second nucleotide (purine/pyrimidine) of each messenger was determined in these 5'UTR sequences. 5'UTR + ORF sequences and lengths were acquired by concatenating the 5'UTR and ORF described above. Transcription units allowing determination of the number of genes in operons and their positions within the polycistronic mRNA were retrieved from RegulonDB database (<http://regulondb.ccg.unam.mx/>) and confirmed by Ecocyc. COG annotations and gene-product cell locations were obtained from *E. coli* K-12 annotation (19). Essential genes were identified in (20).

### **Calculation of sequence properties**

GC content was calculated as the percentage of G and C nucleotides in each sequence. Prediction of peptide signal in ORF sequences was determined using SignalP4.1 software (<http://www.cbs.dtu.dk/services/SignalP/>). Calculation of Codon Adaptation Index (CAI) was performed as described in (21) using 55 ribosomal proteins as references. We used the Zscore as a measure for thermodynamic stability of each -30 to +24 bp region relative to CDS's start codon. The Zscore was defined for each region as the minimum free folding energy of the native sequence minus the mean of the minimum free folding energy of its randomized sequences divided by the standard deviation of minimum free folding energy of its randomized

sequences. Randomized sequences were computed by shuffling 100 times each region of interest, using Ushuffle software (22) which conserves dinucleotide composition. The minimum free folding energy of native and randomized sequences was estimated using RNAfold of the Vienna Package (23). The Grand average of hydropathicity score (GRAVY) of the linear polypeptide sequence was calculated as the sum of hydropathy values of all amino acids, divided by the number of residues in the sequence (24).

### **Protein concentration determination**

Cells from *E. coli* steady-state continuous cultures at four growth rates (1) were centrifuged and cell pellets were washed with ice-cold PBS buffer before new step of centrifugation. Cell pellets were frozen in liquid nitrogen and stored at -80°C. Cells were then resuspended in lysis buffer (4 % SDS, 100 mM Tris-HCl pH 8, 100 mM DTT) during 5 min at 95°C before sonication to break down cell membrane. Total protein concentration was then measured using tryptophan absorbance. A SILAC approach (for Stable Isotope Labeling by Amino acids in Cell culture) was used to get relative quantification of proteins for the four growth rates. Absolute protein quantification was then performed by iBAQ (intensity-Based Absolute Quantification) method adding UPS2 Proteomic Standards (Sigma-Aldrich) in samples (25).

### **Linear covariance models**

A linear analysis of covariance model was carried out to identify the major determinants of mRNA stability at four growth rates using various quantitative and qualitative parameters as described previously (26). Gene parameters such as mRNA concentration ( $[mRNA]$ ), length (*ORF length*, *5'UTR length*, *5'UTR + ORF length*) and GC content (*%GC ORF*, *%GC 5'UTR* and *%GC 5'UTR+ORF* respectively), CAI and Zscore for folding energy (*Zscore*) were considered as quantitative parameters. They were log-transformed to obtain a normal distribution (except for Zscore which is normally distributed by construction) and all were centered and reduced. This normalization was applied in order to adjust their level and allow comparison of model coefficients. The number of genes within operon in which is located the gene (*Nb of gene in operon*), the gene essentiality (*Essential*), the COG membership, the cell location of gene product, the presence of peptide signal in mRNA and the nature of the 2nd nucleotide of the messenger (*Nature of*

2nd nt) were considered as qualitative parameters. Equation [1] described the models established to explain the mRNA half-lives at each growth rate.

$$\begin{aligned}
 \ln(mRNA\ stability_{(i,j)}) &= \alpha + \beta_{[mRNA]_{(i,j)}} \ln([mRNA]_{(i,j)}) + \beta_{ORF\ length_{(i)}} \ln(ORF\ length_{(i)}) \\
 &+ \beta_{5'UTR\ length_{(i)}} \ln(5'UTR\ length_{(i)}) \\
 &+ \beta_{5'UTR+ORF\ length_{(i)}} \ln(5'UTR+ORF\ length_{(i)}) \\
 &+ \beta_{\%GC\ ORF_{(i)}} \ln(\%GC\ ORF_{(i)}) + \beta_{\%GC\ 5'UTR_{(i)}} \ln(\%GC\ 5'UTR_{(i)}) \\
 &+ \beta_{\%GC\ 5'UTR+ORF_{(i)}} \ln(\%GC\ 5'UTR+ORF_{(i)}) + \beta_{CAI_{(i)}} \ln(CAI_{(i)}) \\
 &+ \beta_{Zscore_{(i)}} \ln(Zscore_{(i)}) + \lambda_{Nb\ of\ gene\ in\ operon_{(i)}} + \lambda_{Essential_{(i)}} + \lambda_{COG_{(i)}} \\
 &+ \lambda_{Cell\ location_{(i)}} + \lambda_{signal\ peptide_{(i)}} + \lambda_{Nature\ of\ 2nd\ nt_{(i)}} + \varepsilon_{(i)}
 \end{aligned}$$

[1]

where  $mRNA\ stability_{(i,j)}$  is the measured level of the  $i$  th value for the variable of interest,  $j$  is the  $j$  th growth rates,  $\alpha$  was the intercept,  $\beta$  and  $\lambda$  were the coefficients associated to the quantitative and qualitative parameters respectively and  $\varepsilon_{(i)}$  was the error term for the  $i$  th half-life. A complete matrix (without missing value) for all the parameters and half-lives at the four growth rates was obtained for 1589 messengers. In order to select the best model without any *a priori* subjective parameter selection, the Akaike model selection criterion (AIC for Akaike Information Criterion) was used (26). In a correct analysis of the model results, the coefficients of the qualitative parameters cannot be compared to the coefficients of the quantitative parameters. Coefficients related to quantitative parameters influence the slope of the model whereas those linked to qualitative parameters modify the intercept. The influence of each quantitative parameter can thus be estimated by the ranking of the absolute values of associated coefficient. Coefficients linked to qualitative parameters had to be interpreted differently. A positive value of the coefficient means higher half-life of the corresponding mRNA compared with the overall average transcript half-lives. Inversely, a negative value was associated to lower messenger stability. Each parameter coefficient was given with a P-value.

### Simple linear correlation

Simple correlations were estimated by calculating Pearson correlation coefficient and associated P-value using R free statistical software. Significant correlation displayed P-value < 0.05.

### Motifs research

The existence of stabilizing and destabilizing sequence motifs was explored in the top (most stable) and bottom (less stable) quartiles of the mRNA half-life distribution. Four datasets were constructed containing each the 270 most (respectively less) stable mRNAs at separately  $\mu = 0.10 \text{ h}^{-1}$  and  $\mu = 0.63 \text{ h}^{-1}$ . We searched for motif contrasts between unstable and stable mRNAs in the 5'UTR region upstream mRNA sequence (from -100 to -1 bp relative to CDS' start codon) of these subgroups using RMES software (<http://migale.jouy.inra.fr/?q=rmes>). We assessed 5-nt length motifs over- and under-representation in these sequences using a Markov Model of order 1 (M1) and a compound Poisson approximation. RMES computes P-values and scores for all the 1024 possible motifs of length 5 by comparing observed count of each motif to the count expected in random sequences showing the same composition in length 2 words. P-values measure the probability, under our model, to observe as much (or as few) occurrences of each motif. RMES scores are derived from P-values using the standard one-to-one probit transformation and allow discriminating exceptionally frequent motifs (exhibiting high positive scores) from exceptionally rare motifs (exhibiting high negative scores).

## RESULTS

To identify the general determinants of mRNA stability in *E. coli*, we used linear covariance model to explain large-scale mRNA half-life datasets (1) from quantitative (mRNA concentration, length, sequence composition and structure, CAI...) and qualitative (gene essentiality, gene function, cell location of gene product, predicted peptide signal, polycistronic transcript...) parameters (complete list in Materials and Methods section). Since a previous study showed different regulatory patterns of mRNA stability with changing growth rate (1), we decided to carry out four models at different growth rates. In each model, 1589 half-life values (~ 37 % of all *E. coli* coding genes) were available: these are the largest sets of mRNA half-lives ever

used to identify mRNA stability determinants. The parameters selected by the AIC model criterion are listed for each model in Table 1. They were both quantitative and qualitative reflecting either intrinsic features (e.g. CAI, length, GC%, function) and extrinsic parameters dependent on the environment (e.g. [mRNA], essentiality of the gene function). These parameter selections led to the relatively good fitting quality of the four models with determination coefficients (adjusted  $R^2$ ) ranging between 0.74 at  $\mu = 0.10 \text{ h}^{-1}$  and 0.51 at  $\mu = 0.63 \text{ h}^{-1}$ . These results indicated that mRNA stability can be relatively well explained with the selected parameters especially at the lowest growth rates.

To analyze the relative contribution of the parameters and the dependency on the growth-rate value, we defined as a significant determinant of mRNA stability only selected parameters with an associated P-value  $< 0.05$ . The determinants of the mRNA stability present in the four models were defined as not dependent on the growth rate; a determinant present in at least one but not all the models was considered as dependent on the growth rate.

### **Determinants of mRNA stability independent of the growth rate**

Seven determinants ([mRNA], CAI, ORF length, GC% in 5'UTR+ORF, number of genes in operon, function and essentiality) were effective at all the growth rates. Among the quantitative determinants, the most influent was always the mRNA concentration with negative coefficients lower than  $< -0.79$  (P-value  $< 10^{-204}$ ). The negative sign indicated that more an mRNA is concentrated, shorter the half-life is. This corresponds to a counterintuitive finding since it could be believed that increase of mRNA concentration would be associated with an increase of mRNA stability. The second determinant was CAI with positive coefficients ranging between 0.13 and 0.25 (P-value  $< 10^{-14}$ ). These results showed that a higher content of optimal codons led to a higher half-life value. CAI is a well-known determinant of the protein level in *E. coli* (27–30) and was confirmed in this study (see below) but its effect on mRNA stability has been only rarely investigated before. We also could determine that only the ORF length (neither the length of 5'UTR nor the length of 5'UTR + ORF) was systematically negatively correlated with mRNA half-life (coefficients  $< -0.12$ , P-value  $< 10^{-2}$ ). Finally, a weak negative effect of the GC content of 5'UTR + ORF on mRNA stability was also obtained (coefficients  $< -0.05$ , P-value  $< 10^{-3}$ ).



Regarding the qualitative determinants of mRNA half-lives, the essential genes displayed significantly higher transcript stability than non-essential genes for all the growth rates (P-value < 0.05). Three COG categories had significant coefficients for the four models. Messengers of genes belonging to [D] Cell cycle control, cell division, chromosome partitioning and [J] Translation, ribosomal structure and biogenesis were shown to be more stable than the rest of the other mRNAs whereas messengers of genes belonging to [I] Lipid transport and metabolism were shown to be less stable.

We observed that mRNA of genes belonging to operons containing 3 genes (coefficients < -0.16, P-value < 0.05) were significantly less stable than messengers of genes belonging to operons containing 12 genes (coefficients > 0.45, P-value <  $10^{-2}$ ) suggesting that cistron stability is higher in longer operons at all the growth rates. To determine whether a link between cistron stability and its position in the operon exists, we compared the stability of the first and the last cistrons in each operon for which we could collect all stability data (431 operons out of 836 in *E. coli*). The results showed that for each growth rate, the median half-life of a cistron was increasing with its ranking in the polycistronic mRNA. The representative example of operons with four cistrons is given in Figure 1. The last cistron was in average 1.5 times more stable than the first cistron in at least 62 % of operons studied (> 268 operons). These data suggested that globally, more a cistron is distant from the 5'-end extremity of a transcript, more it is stable.

Furthermore we searched for the presence of sequence motifs possibly related to mRNA stability. We used the RMES software to assess over- and under-representation of all possible five letter motifs in the 100 bp 5'UTR regions of the 25 % most stable and the 25 % most unstable mRNAs at the two extreme growth rates (0.10 and 0.63 h<sup>-1</sup>). RMES scores of exceptionality of each motif were computed and compared between the unstable and stable mRNAs (Figure 2). At the two growth rates, the plots indicated a strong correlation of the motif exceptionality scores between the two groups of stable and unstable mRNAs: most of the motifs were located on the diagonal line and no motif was found over-represented in one group but under-represented in the other one. Nevertheless, six motifs (highlighted in the figure) were quite distant from the scatter plot and exhibited distinct patterns of exceptionality in stable and unstable mRNAs. These motifs could be split into two groups: a set of three A/G rich motifs including AAGGA, AGGAG and CAGGA (with a

common core-motif = AGGA) and a set of three C/G rich motifs including CGCTG, GCTGG and CTGGC (with a common core-motif = CTG). Interestingly, the AAGGA and AGGAG motifs were globally more over-represented in the stable mRNAs compared to the unstable ones, while these differences were less visible for the CAGGA motif. On the other hand, the G/C rich motifs (CTGGC, GCTGG and CGCGT) were all more frequent in the stable mRNAs than in the unstable ones (Figure 2). Analysis of the side extension of AGGAG motif showed that A and C were the most frequent upstream nucleotides indicating that CAGGA and AAGGA could partially overlap with AGGAG (Figure S1). In contrast, the analysis of the side context of the GCTGG motif (Figure S1) did not exhibit any motif that could overlap to form a longer conserved motif. To validate at the entire population level these observations we compared the half-life distributions for the whole mRNA population and for the different subpopulations of mRNAs whose 5'UTR region contained motifs (Figure 3). The subpopulations of transcripts with respectively the AGGAG, AAGGA, CTGGC, CGCTG or GCTGG motifs were confirmed to be globally more stable than the entire mRNA population whereas the CAGGA motif did not affect significantly mRNA stability. The three motifs AAGGA, AGGAG and CAGGA were mostly located in the Shine-Dalgarno region (between 7 bp and 20 bp upstream of the start codon) with no other "hot spot" region along the 5'UTR even when the motifs were duplicated (up to 4 times). Interestingly, AAGGA motif, but not the CAGGA motif, is the exact complement of the 3'-terminus of the 16S rRNA of *E. coli*. The three G/C motifs, which could also be duplicated up to 4 times, were randomly located in 100 bp 5'UTR.

Finally, it was quite surprising to not find the Zscore of the folding energy in the -30 / +24 bp region of mRNA as a general determinant of mRNA stability in none of the models. Indeed, secondary structure in the 5'UTR was reported to impact decay of specific mRNAs by impeding access to RNase (8, 32). To look closely to the relationship between mRNA stability and Zscore, we compared the stability of mRNAs displaying the lowest Zscores (< -2, highest probability of stable secondary structure formation) to those of the rest of transcripts (Figure S2). No significant difference in the mRNA half-life distribution of the two groups was obtained. These results indicate that potential mRNA secondary structure in -30 / +24 bp region of mRNA is not predictive of messenger stability at the omic level.

### Evolution of the mRNA stability determinants with the growth rate

The significance of several determinants of mRNA stability was dependent on the growth rate. Four COG categories were significant from  $\mu = 0.10$  to  $0.40 \text{ h}^{-1}$  but not at  $0.63 \text{ h}^{-1}$  (Table 1). In contrast, three COGs became only significant for the model at  $\mu = 0.63 \text{ h}^{-1}$ . At the low growth rates, the mRNAs of genes belonging to [H] Coenzyme transport and metabolism and [U] Intracellular trafficking, secretion, and vesicular transport were more stable while transcripts from [N] Cell motility and [Q] Secondary metabolites biosynthesis, transport and catabolism were more unstable compared to all transcripts. At high growth rate, mRNAs of [G] Carbohydrate and [F] Nucleotide transport and metabolism were more stable than the rest of messengers whereas [E] Amino acid transport and metabolism displayed lower half-lives. These last data are perfectly coherent with a previous study which highlighted the control of gene expression of carbon central and amino acid metabolism by mRNA stability at high growth rate (1). The results show that the regulation of mRNA stability depends on the gene function but also for these functional categories on the growth rate value. The transcripts of lipoproteins located at the outer membrane were more stable than the other transcripts only at  $\mu = 0.10$  and  $0.20 \text{ h}^{-1}$ . The interpretation of this finding is unclear.

The mRNA concentration was the most influent determinant in the four models but a slight decrease of its coefficient as the growth rate increases was observed suggesting a slighter influence of the mRNA level in the determination of mRNA half-life. In order to better understand direct correlation between mRNA concentration and stability, we plotted the degradation rate constant  $k$  ( $\ln(2)/t_{1/2}$ ) as a function of the concentration for all the transcripts at the four growth rates (Figure 4). This profile confirmed the previous trend since data were more dispersed as the growth rate increases. Moreover, one can see that the maximal front of degradation rate was higher as the growth rate increases. So, independently of the mRNA level, the cells raised their ability to degrade mRNA at high growth rate. This could be due to the involvement of other determinants such as an increased expression of the degradation machinery when the growth rate increases (1). This could explain why mRNA concentration was less contributive of mRNA stability at high growth rate.

**Relationships between mRNA stability and protein level**

To investigate whether mRNA stability could be a determinant of protein level in *E. coli*, new linear covariance models were used to explain large scale proteomic data (643 protein levels) at different growth rates. Additionally to the previous parameters used in the half-life determination models, mRNA half-life, GRAVY score and the percentages of each amino acid in a protein were considered as explicative variables. Since we showed in the part above that mRNA concentration and stability were strongly correlated, the mRNA concentrations were not included in the models.

The adjusted  $R^2$  of protein concentration determination models with mRNA half-life data were around 0.50, and quite similar to those of models for which the mRNA half-life data were excluded (data not shown). This result showed that the impact of mRNA stability on protein level was small. However, in the models with mRNA half-life data, this parameter was selected by AIC criterion at  $\mu = 0.10 \text{ h}^{-1}$  and  $\mu = 0.63 \text{ h}^{-1}$  (Table 2). The correlation coefficients were significant but small and with opposite sign at the two growth rates (-0.14, and +0.08 at  $\mu = 0.10$  and  $0.63 \text{ h}^{-1}$ , respectively). These data suggested that the weak correlation between mRNA half-life and protein level could be dependent on the growth rate. The low impact of mRNA stability on protein level and the sign inversion at high growth rate were quite surprising since (i) protein level and mRNA concentration were positively well correlated in this data set of 643 mRNAs and proteins (model correlation coefficient  $> 0.50$ , P-value  $< 10^{-55}$ ) and (ii) mRNA concentration and half-life were strongly negatively correlated (as demonstrated above in the half-life determination models for a 1589 mRNA population). This intriguing finding could be explained by the fact that the mRNAs of the 643 dataset (used in the protein determination models) were not exactly representative of the larger dataset of 1589 mRNAs (used in the half-life determination models) as shown by the non-overlapping densities of mRNA expression level obtained in the two datasets (Figure S3). Nevertheless, our models of protein level determination confirmed the significant positive influence of CAI on protein level (coefficient  $> 0.48$ , P-value  $< 10^{-31}$ ) (33). Indeed, the CAI which is a parameter related to translation efficiency was often demonstrated to be an important positive determinant of protein concentration (27–30).

## DISCUSSION

We used the largest and most recent datasets of mRNA half-lives in *E. coli* and a large number of parameters in linear covariance models to identify and rank the determinants of mRNA stability at four different growth rates. The determination coefficients  $R^2$  of the four models were relatively high ( $> 0.51$ ) indicating that a major part of mRNA stability can be explained by the combination of the selected parameters.

The determinants of mRNA stability can be grouped according to the regulatory mode of interference. Some determinants are directly linked to action of RNase E and the machinery of degradation. The weak but significant negative correlation between ORF length and mRNA half-lives is consistent with the general believe that the endonucleolytic attack by RNase E is the rate-limiting step of mRNA degradation and that cleavage sites used during the alternative direct entry pathway could be distributed evenly along the messenger (34). However, it cannot be totally excluded that mechanical damages make the long mRNAs less stable than the short ones (17).

We validated for a majority of polycistronic mRNA that more a cistron is distant from the 5'-end extremity, more it is stable (35) but we also demonstrated for the first time that this phenomenon is generally independent of the growth rate. These results are perfectly consistent with the current model of 5' to 3' mRNA degradation initiated via the 5' end dependent RNase E endonucleolytic pathway (34).

It was shown on a model RNA that RNase E has a preference for AU rich regions (7). However, it has never been proven at omic level that AU rich motif and GC content in the 5'UTR were determinants of mRNA stability (9). Here, our linear models demonstrated that a negative correlation existed between the GC content and half-life values but only for the combination 5'UTR + ORF sequence and not for the 5'UTR or ORF alone. Therefore we confirm at the global level the absence of correlation between the GC content of the 5'UTR and the transcript stability. Our results differed slightly from those of Lenz and coworkers who described a negative correlation between the GC content and the mRNA half-life but for the ORF alone (31). These authors showed that the correlation between GC content and half-life values was not simple and fluctuated along the mRNA molecule. This could explain the difficulty to find this parameter as a determinant of mRNA stability. Given that

formation of secondary structures is often linked to the GC content, it was not surprising to find that the Zscore for the folding energy of the -30 / +24 bp region of mRNA was never selected as a contributive parameter of the mRNA stability. In addition, we showed in this study that the G/C rich motif GCTGG located in the 100 bp 5'UTR region was stabilizing mRNAs. GCTGG is a sub-motif of the CGCTGG motif enriched in CDS sequences of mRNAs likely targeted by the RNA degradosome (36). The opposite effect of the GCTGG motif in the two studies on mRNA half-life regulation could reflect the involvement of different regulatory mechanisms according to the motif localization. It can also be noticed that the GCTGG motif is a sub-motif of the Chi (Crossover Hotspot Instigator, GCTGGTGG) motif involved in recombinational repair of genomic breaks in *E. coli* (37). Given its over-representation in stable mRNAs, it is tempting to hypothesize that sequence evolution in order to simultaneously favor DNA conservation and mRNA stability of a gene could occur in *E. coli*.

We demonstrated in this study that mRNA concentration was the main determinant to explain the mRNA stability and was negatively correlated with the messenger lifetime. In other words, the more concentrated mRNAs will be the less stable ones supporting the previous idea that mRNA degradation tends in general to counter balance transcription (1). This counter intuitive idea was also reported in other bacteria (9, 13, 16). It can be explained by the assumption that generally, a high mRNA concentration leads to a higher probability to meet RNases and consequently lowers the mRNA half-life. Furthermore, cell ability to degrade mRNA was shown to be enhanced as the growth rate increases. This result supports the assumption proposed above since at high growth rate, a global up-expression of the degradation machinery and a global up-regulation of the mRNA concentration occur (1).

Additionally to these determinants directly linked to action of RNases, we identified determinants related to translation interfering with the action of ribosomes. Surprisingly we found that CAI positively correlated with mRNA half-lives. This observation is in contradiction with tAI and half-life simple correlation reported in *E. coli* (31) but coherent with results in *L. lactis* (13). A high CAI is believed to lead to efficient translation elongation by reducing ribosome pausing (38, 39). Knowing the well-documented protective effect of ribosomes against RNase attacks (40), we can postulate that efficient translation elongation leading to continuous and fast ribosome

motion along the mRNA will impede the interaction of the nucleases with its substrate and thus will protect mRNA from degradation. A second hypothesis which could explain the results uses the opposite idea of a deleterious effect of ribosome on mRNA stability by its ability to recruit the RNA degradosome (41). When translation is initiation-limited, a high CAI would lead to a lower number of ribosomes per messenger. Thus, less degradosome complexes will be recruited per mRNA molecule increasing consequently the mRNA stability. Also related to ribosome protection, we observed that increasing the sequence complementary of the ribosome binding site and the ribosome 16S rRNA stabilized the mRNA. This result suggests for the first time at the omic level that a higher strength of ribosome binding on the 5'UTR favors mRNA stability conferring a protective effect of ribosome binding against mRNA degradation.

Others determinants of mRNA stability have been identified in this study without elucidating yet the regulations involved. These new determinants were linked to gene functions probably dependent on cell physiology requirement and bacterial adaptation. Messengers of functional category related to growth like [D] Cell cycle control, cell division, chromosome partitioning and [J] Translation, ribosomal structure and biogenesis were significantly more stable at all the growth rates. As mRNA stability determines the time of accessibility of a messenger for translation, decay rate of these growth-related mRNAs is probably maintained low by the cell to keep a basal expression of these messengers. This idea was strengthened by the fact that essential genes were demonstrated for the first time to display significant higher mRNA half lives than the other genes. In the same way, the stability of cistrons of long operons was higher than those of shorter cistrons. Given that operon are often considered as functional structures, it is highly probable that the stability of these cistrons is programmed to satisfy requirements related to the function of the products they encode as demonstrated in yeast (10).

Expression of a messenger needs generally to be in line with the constraints of the growth conditions. Highly concentrated mRNAs were more sensitive to degradation. Consequently, the rapid turnover of a highly expressed messenger is probably necessary to help rapid cell physiology adaptation in case of changing environment. In a similar manner, mRNA half-lives of some functional categories were specifically regulated at certain growth rates. For example, stabilization of

transcripts from functional categories such as [G] Carbohydrate and [F] Nucleotide transport and metabolism at the highest growth rate could support higher metabolism (1).

Taken together, these data suggest that functional genomics probably results of the setting of specific regulations of mRNA stability depending on the involvement of gene function in the cell physiology and adaptation.

Our linear covariance model displayed good determination coefficient but always lower than 1. It means that some non-investigated parameters also participate to mRNA half-life determination. Furthermore, since the  $R^2$  decreased as the growth rate increases, the contribution of these non-investigated parameters was more important at higher growth rate. These parameters remain to be discovered and could be related to RNA binding proteins, riboswitches or sRNA. Moreover, their contribution would be growth-rate dependent (42, 43). Finally, we showed that the contribution of the mRNA stability in the determination of the protein level was low whatever the growth rate value. A significant but also small correlation between mRNA half-life and protein abundance was previously reported in *E. coli* (31). This result underlines the complexity of protein level regulation which involved in addition to mRNA stability and transcription other complex and highly regulated processes such as translation and protein turn-over.

## FUNDING

The funding of this work was provided by Université de Toulouse, Région Midi-Pyrénées and Prime XS consortium. Funding for open access charge: Université de Toulouse, Région Midi-Pyrénées.



## REFERENCES

1. Esquerre, T., Laguerre, S., Turlan, C., Carpousis, A.J., Girbal, L. and Cocaign-Bousquet, M. (2014) Dual role of transcription and transcript stability in the regulation of gene expression in *Escherichia coli* cells cultured on glucose at different growth rates. *Nucleic Acids Res.*, 42, 2460–2472.
2. Morita, T., Maki, K. and Aiba, H. (2005) RNase E-based ribonucleoprotein complexes: mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. *Genes Dev.*, 19, 2176–2186.
3. Yakhnin, A.V., Baker, C.S., Vakulskas, C.A., Yakhnin, H., Berezin, I., Romeo, T. and Babitzke, P. (2013) CsrA activates *flhDC* expression by protecting *flhDC* mRNA from RNase E-mediated cleavage. *Mol. Microbiol.*, 87, 851–866.
4. Caron, M.-P., Bastet, L., Lussier, A., Simoneau-Roy, M., Masse, E. and Lafontaine, D.A. (2012) Dual-acting riboswitch control of translation initiation and mRNA decay. *Proc. Natl. Acad. Sci. U. S. A.*, 109, E3444–E3453.
5. Deana, A. and Belasco, J.G. (2005) Lost in translation: the influence of ribosomes on bacterial mRNA decay. *Genes Dev.*, 19, 2526–2533.
6. Dreyfus, M. (2009) Killer and Protective Ribosomes. In Ciaran Condon (ed), *Progress in Molecular Biology and Translational Science, Molecular Biology of RNA Processing and Decay in Prokaryotes*. Academic Press, Vol. Volume 85, pp. 423–466.
7. McDowall, K.J., Lin-Chao, S. and Cohen, S.N. (1994) A+U content rather than a particular nucleotide order determines the specificity of RNase E cleavage. *J. Biol. Chem.*, 269, 10790–10796.
8. Arnold, T.E., Yu, J. and Belasco, J.G. (1998) mRNA stabilization by the ompA 5' untranslated region: two protective elements hinder distinct pathways for mRNA degradation. *RNA*, 4, 319–330.
9. Bernstein, J.A., Khodursky, A.B., Lin, P.-H., Lin-Chao, S. and Cohen, S.N. (2002) Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl. Acad. Sci. U. S. A.*, 99, 9697–9702.
10. Wang, Y., Liu, C.L., Storey, J.D., Tibshirani, R.J., Herschlag, D. and Brown, P.O. (2002) Precision and functional specificity in mRNA decay. *Proc. Natl. Acad. Sci.*, 99, 5860–5865.
11. Hambræus, G., von Wachenfeldt, C. and Hederstedt, L. (2003) Genome-wide survey of mRNA half-lives in *Bacillus subtilis* identifies extremely stable mRNAs. *Mol. Genet. Genomics MGG*, 269, 706–714.
12. Redon, E., Loubière, P. and Cocaign-Bousquet, M. (2005) Role of mRNA stability during genome-wide adaptation of *Lactococcus lactis* to carbon starvation. *J. Biol. Chem.*, 280, 36380–36385.

13. Dressaire, C., Picard, F., Redon, E., Loubiere, P., Queinnec, I., Girbal, L. and Coccagn-Bousquet, M. (2013) Role of mRNA Stability during Bacterial Adaptation. PLoS ONE, 8, e59059.
14. Hundt, S., Zaigler, A., Lange, C., Soppa, J. and Klug, G. (2007) Global Analysis of mRNA Decay in *Halobacterium salinarum* NRC-1 at Single-Gene Resolution Using DNA Microarrays. J. Bacteriol., 189, 6936–6944.
15. Kristoffersen, S.M., Haase, C., Weil, M.R., Passalacqua, K.D., Niazi, F., Hutchison, S.K., Desany, B., Kolstø, A.-B., Tourasse, N.J., Read, T.D., et al. (2012) Global mRNA decay analysis at single nucleotide resolution reveals segmental and positional degradation patterns in a Gram-positive bacterium. Genome Biol., 13, R30.
16. Rustad, T.R., Minch, K.J., Brabant, W., Winkler, J.K., Reiss, D.J., Baliga, N.S. and Sherman, D.R. (2013) Global analysis of mRNA stability in *Mycobacterium tuberculosis*. Nucleic Acids Res., 41, 509–517.
17. Feng, L. and Niu, D.-K. (2007) Relationship between mRNA stability and length: an old question with a new twist. Biochem. Genet., 45, 131–137.
18. Kim, D., Hong, J.S.-J., Qiu, Y., Nagarajan, H., Seo, J.-H., Cho, B.-K., Tsai, S.-F. and Palsson, B.Ø. (2012) Comparative Analysis of Regulatory Elements between *Escherichia coli* and *Klebsiella pneumoniae* by Genome-Wide Transcription Start Site Profiling. PLoS Genet, 8, e1002867.
19. Riley, M., Abe, T., Arnaud, M.B., Berlyn, M.K.B., Blattner, F.R., Chaudhuri, R.R., Glasner, J.D., Horiuchi, T., Keseler, I.M., Kosuge, T., et al. (2006) *Escherichia coli* K-12: a cooperatively developed annotation snapshot—2005. Nucleic Acids Res., 34, 1–9.
20. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L. and Mori, H. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol., 2, 2006.0008.
21. Sharp, P.M. and Li, W.H. (1987) The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res., 15, 1281–1295.
22. Jiang, M., Anderson, J., Gillespie, J. and Mayne, M. (2008) uShuffle: A useful tool for shuffling biological sequences while preserving the k-let counts. BMC Bioinformatics, 9, 192.
23. Gruber, A.R., Lorenz, R., Bernhart, S.H., Neubock, R. and Hofacker, I.L. (2008) The Vienna RNA Websuite. Nucleic Acids Res., 36, W70–W74.
24. Kyte, J. and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. J. Mol. Biol., 157, 105–132.

25. Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W. and Selbach, M. (2011) Global quantification of mammalian gene expression control. *Nature*, 473, 337–342.
26. Dressaire, C., Laurent, B., Loubière, P., Besse, P. and Coccagn-Bousquet, M. (2010) Linear covariance models to examine the determinants of protein levels in *Lactococcus lactis*. *Mol. Biosyst.*, 6, 1255.
27. Lithwick, G. and Margalit, H. (2003) Hierarchy of Sequence-Dependent Features Associated With Prokaryotic Translation. *Genome Res.*, 13, 2665–2673.
28. Lu, P., Vogel, C., Wang, R., Yao, X. and Marcotte, E.M. (2007) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat. Biotechnol.*, 25, 117–124.
29. Ishihama, Y., Schmidt, T., Rappsilber, J., Mann, M., Hartl, F.U., Kerner, M.J. and Frishman, D. (2008) Protein abundance profiling of the *Escherichia coli* cytosol. *BMC Genomics*, 9, 102.
30. Valgepea, K., Adamberg, K., Seiman, A. and Vilu, R. (2013) *Escherichia coli* achieves faster growth by increasing catalytic and translation rates of proteins. *Mol. Biosyst.*, 9, 2344–2358.
31. Lenz, G., Doron-Faigenboim, A., Ron, E.Z., Tuller, T. and Gophna, U. (2011) Sequence features of *E. coli* mRNAs affect their degradation. *PloS One*, 6, e28544.
32. Hambræus, G., Karhumaa, K. and Rutberg, B. (2002) A 5' stem-loop and ribosome binding but not translation are important for the stability of *Bacillus subtilis* aprE leader mRNA. *Microbiology*, 148, 1795–1803.
33. Guimaraes, J.C., Rocha, M. and Arkin, A.P. (2014) Transcript level and sequence determinants of protein abundance and noise in *Escherichia coli*. *Nucleic Acids Res.*, 10.1093/nar/gku126.
34. Bouvier, M. and Carpousis, A.J. (2011) A tale of two mRNA degradation pathways mediated by RNase E. *Mol. Microbiol.*, 82, 1305–1310.
35. Selinger, D.W., Saxena, R.M., Cheung, K.J., Church, G.M. and Rosenow, C. (2003) Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. *Genome Res.*, 13, 216–223.
36. Zhou, L., Zhang, A.B., Wang, R., Marcotte, E.M. and Vogel, C. (2013) The proteomic response to mutants of the *Escherichia coli* RNA degradosome. *Mol. Biosyst.*, 9, 750–757.
37. Halpern, D., Chiapello, H., Schbath, S., Robin, S., Hennequet-Antier, C., Gruss, A. and El Karoui, M. (2007) Identification of DNA Motifs Implicated in Maintenance of Bacterial Core Genomes by Predictive Modeling. *PLoS Genet*, 3, e153.

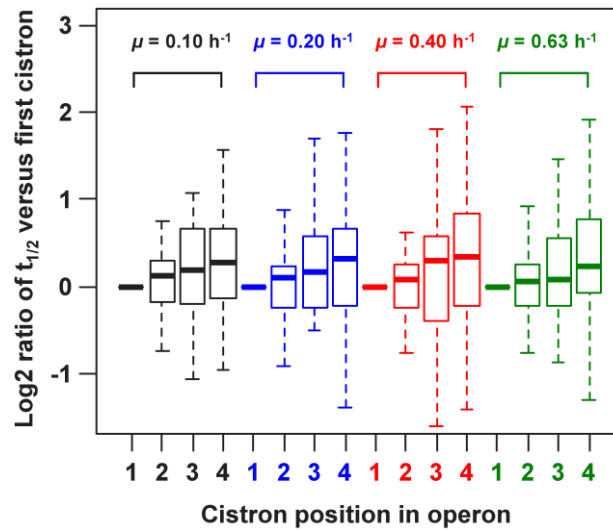
38. Tuller, T., Waldman, Y.Y., Kupiec, M. and Ruppin, E. (2010) Translation efficiency is determined by both codon bias and folding energy. *Proc. Natl. Acad. Sci.*, 10.1073/pnas.0909910107.
39. Plotkin, J.B. and Kudla, G. (2011) Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.*, 12, 32–42.
40. Kaberdin, V.R. and Bläsi, U. (2006) Translation initiation and the fate of bacterial mRNAs. *FEMS Microbiol. Rev.*, 30, 967–979.
41. Tsai, Y.-C., Du, D., Dominguez-Malfavon, L., Dimastrogiovanni, D., Cross, J., Callaghan, A.J., Garcia-Mena, J. and Luisi, B.F. (2012) Recognition of the 70S ribosome and polysome by the RNA degradosome in *Escherichia coli*. *Nucleic Acids Res.*, 40, 10417–10431.
42. Vytvytska, O., Jakobsen, J.S., Balcunaite, G., Andersen, J.S., Baccarini, M. and von Gabain, A. (1998) Host factor I, Hfq, binds to *Escherichia coli* ompA mRNA in a growth rate-dependent fashion and regulates its stability. *Proc. Natl. Acad. Sci. U. S. A.*, 95, 14118–14123.
43. Grunberg-Manago, M. (1999) Messenger Rna Stability and Its Role in Control of Gene Expression in Bacteria and Phages. *Annu. Rev. Genet.*, 33, 193–227.
44. Chambers, J.M. ed. (1983) Graphical methods for data analysis Wadsworth International Group; Duxbury Press, Belmont, Calif.; Boston.
45. Crooks, G.E., Hon, G., Chandonia, J.-M. and Brenner, S.E. (2004) WebLogo: A Sequence Logo Generator. *Genome Res.*, 14, 1188–1190.

**Table 1.** Estimated coefficients of growth-rate dependent mRNA half-life models

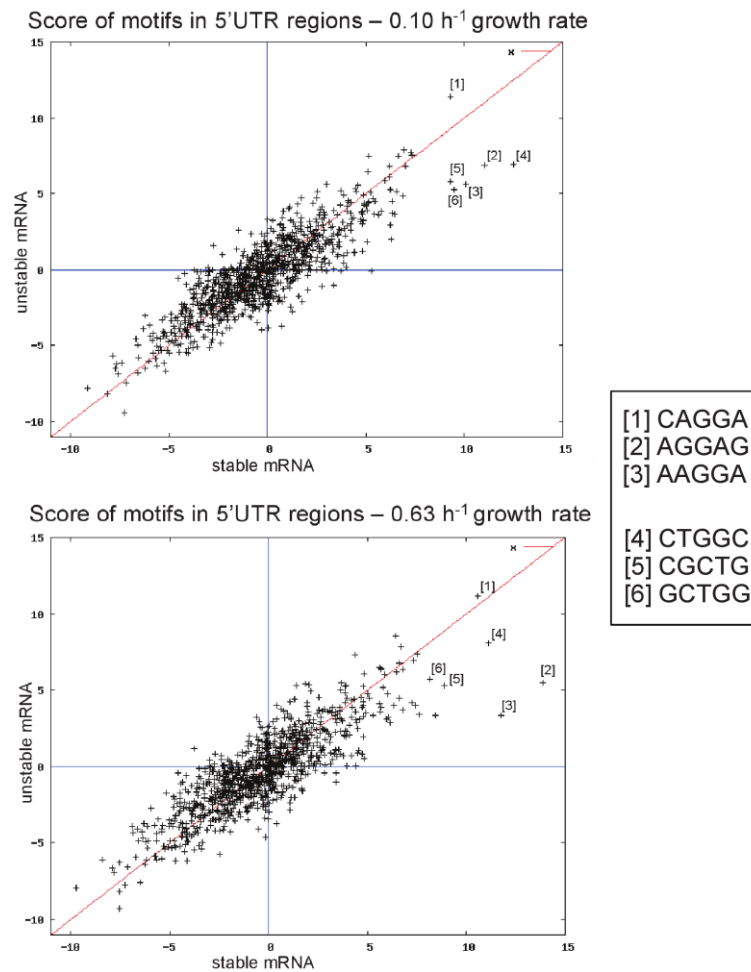
Parameter	Variable to explain : half-life time							
	$\mu = 0.10 \text{ h}^{-1}$		$\mu = 0.20 \text{ h}^{-1}$		$\mu = 0.40 \text{ h}^{-1}$		$\mu = 0.63 \text{ h}^{-1}$	
	Est. Coeff.	P-value	Est. Coeff.	P-value	Est. Coeff.	P-value	Est. Coeff.	P-value
[mRNA]	<b>-0.91</b>	$< 10^{-226}$	<b>-0.80</b>	$1.4E10^{-223}$	<b>-0.80</b>	$2.9E10^{-225}$	<b>-0.79</b>	$4.3E10^{-205}$
CAI	<b>0.13</b>	$2.6E10^{-15}$	<b>0.18</b>	$1.7E10^{-15}$	<b>0.18</b>	$5.5E10^{-17}$	<b>0.25</b>	$7.9E10^{-28}$
ORF length	<b>-0.14</b>	$2.8E10^{-5}$	<b>-0.16</b>	$2.2E10^{-4}$	<b>-0.19</b>	$1.1E10^{-5}$	<b>-0.12</b>	$7.7E10^{-3}$
5'UTR + ORF length	0.05	$9.4E10^{-2}$	0.06	$1.3E10^{-1}$	<b>0.14</b>	$9.1E10^{-4}$	0.09	$5.2E10^{-2}$
5'UTR + ORF GC%	<b>-0.05</b>	$4.4E10^{-4}$	<b>-0.07</b>	$3.1E10^{-4}$	<b>-0.10</b>	$1.3E10^{-7}$	<b>-0.09</b>	$8.0E10^{-6}$
Essential genes	<b>0.06</b>	$8.1E10^{-3}$	<b>0.12</b>	$4.4E10^{-5}$	<b>0.10</b>	$2.7E10^{-4}$	<b>0.06</b>	$4.5E10^{-2}$
<b>Nb of genes in operon</b>								
1	0.29	$2.3E10^{-1}$	0.55	$8.1E10^{-2}$	0.46	$1.4E10^{-1}$	0.43	$2.0E10^{-1}$
2	<b>-0.15</b>	$2.4E10^{-3}$	<b>-0.16</b>	$1.1E10^{-2}$	<b>-0.29</b>	$3.8E10^{-6}$	-0.09	$1.7E10^{-1}$
3	<b>-0.16</b>	$1.3E10^{-3}$	<b>-0.21</b>	$2.0E10^{-3}$	<b>-0.31</b>	$2.3E10^{-6}$	<b>-0.16</b>	$2.1E10^{-2}$
4	-0.10	$8.2E10^{-2}$	-0.10	$1.9E10^{-1}$	<b>-0.19</b>	$9.0E10^{-3}$	-0.11	$1.5E10^{-1}$
5	-0.02	$7.8E10^{-1}$	-0.06	$4.3E10^{-1}$	-0.14	$8.1E10^{-2}$	-0.06	$4.6E10^{-1}$
6	<b>-0.14</b>	$3.2E10^{-2}$	<b>-0.09</b>	$3.1E10^{-1}$	-0.12	$1.6E10^{-1}$	-0.11	$2.2E10^{-1}$
7	0.03	$7.1E10^{-1}$	0.08	$4.0E10^{-1}$	0.06	$5.4E10^{-1}$	0.08	$4.3E10^{-1}$
8	-0.07	$6.3E10^{-1}$	-0.07	$6.9E10^{-1}$	-0.04	$8.3E10^{-1}$	-0.01	$9.4E10^{-1}$
9	-0.08	$3.9E10^{-1}$	0.04	$7.2E10^{-1}$	0.04	$7.4E10^{-1}$	-0.03	$8.3E10^{-1}$
10	-0.13	$3.1E10^{-1}$	-0.11	$5.2E10^{-1}$	-0.10	$5.1E10^{-1}$	0.05	$7.5E10^{-1}$
11	-0.29	$4.1E10^{-1}$	-0.52	$2.4E10^{-1}$	-0.43	$3.3E10^{-1}$	-0.65	$1.7E10^{-1}$
12	<b>0.45</b>	$3.3E10^{-4}$	<b>0.50</b>	$2.2E10^{-3}$	<b>0.74</b>	$4.0E10^{-6}$	<b>0.84</b>	$1.3E10^{-6}$
13	<b>0.48</b>	$1.8E10^{-4}$	<b>0.49</b>	$3.3E10^{-3}$	<b>0.59</b>	$2.9E10^{-4}$	0.31	$8.1E10^{-2}$
16	-0.12	$6.2E10^{-1}$	-0.36	$2.6E10^{-1}$	-0.28	$3.7E10^{-1}$	-0.48	$1.6E10^{-1}$
<b>COG annotation</b>								
[C]	-0.00	$1.0E10^0$	0.02	$8.8E10^{-1}$	0.05	$7.5E10^{-1}$	-0.01	$9.3E10^{-1}$
[D]	<b>0.39</b>	$8.2E10^{-13}$	<b>0.53</b>	$5.8E10^{-14}$	<b>0.59</b>	$6.2E10^{-18}$	<b>0.54</b>	$1.3E10^{-13}$
[E]	-0.21	$6.5E10^{-2}$	-0.23	$1.3E10^{-1}$	-0.23	$1.2E10^{-1}$	<b>-0.41</b>	$1.0E10^{-2}$
[F]	0.01	$8.4E10^{-1}$	0.05	$3.1E10^{-1}$	0.06	$2.7E10^{-1}$	<b>0.23</b>	$1.8E10^{-5}$
[G]	-0.13	$6.6E10^{-2}$	-0.16	$7.5E10^{-2}$	-0.16	$8.4E10^{-2}$	<b>0.24</b>	$1.4E10^{-2}$
[H]	<b>0.12</b>	$2.0E10^{-2}$	<b>0.20</b>	$4.7E10^{-3}$	<b>0.20</b>	$3.6E10^{-3}$	0.14	$6.3E10^{-2}$
[I]	<b>-0.12</b>	$4.4E10^{-2}$	<b>-0.18</b>	$1.7E10^{-2}$	<b>-0.24</b>	$1.4E10^{-3}$	<b>-0.22</b>	$6.1E10^{-3}$
[J]	<b>0.20</b>	$6.2E10^{-3}$	<b>0.21</b>	$2.6E10^{-2}$	<b>0.23</b>	$1.3E10^{-2}$	<b>0.31</b>	$2.4E10^{-3}$
[K]	0.10	$7.2E10^{-2}$	0.04	$5.5E10^{-1}$	0.06	$3.8E10^{-1}$	-0.03	$6.8E10^{-1}$
[L]	-0.01	$8.4E10^{-1}$	-0.05	$4.4E10^{-1}$	-0.07	$2.7E10^{-1}$	-0.12	$6.6E10^{-2}$
[M]	<b>0.13</b>	$2.4E10^{-2}$	0.11	$1.6E10^{-1}$	0.02	$7.5E10^{-1}$	-0.02	$8.3E10^{-1}$
[N]	<b>-0.12</b>	$3.3E10^{-2}$	<b>-0.15</b>	$4.3E10^{-2}$	<b>-0.14</b>	$3.6E10^{-2}$	-0.11	$1.3E10^{-1}$
[O]	0.16	$3.0E10^{-1}$	0.25	$2.0E10^{-1}$	0.25	$1.9E10^{-1}$	0.28	$1.7E10^{-1}$
[P]	-0.10	$8.7E10^{-1}$	-0.12	$1.1E10^{-1}$	-0.11	$1.5E10^{-1}$	-0.13	$1.2E10^{-1}$
[Q]	<b>-0.19</b>	$2.4E10^{-3}$	<b>-0.25</b>	$2.2E10^{-3}$	<b>-0.19</b>	$1.5E10^{-2}$	-0.12	$1.5E10^{-1}$
[R]	-0.08	$5.4E10^{-1}$	-0.12	$4.6E10^{-1}$	-0.20	$2.2E10^{-1}$	-0.16	$3.5E10^{-1}$
[S]	0.01	$8.8E10^{-1}$	-0.01	$8.0E10^{-1}$	-0.04	$4.9E10^{-1}$	-0.03	$6.1E10^{-1}$
[T]	<b>-0.10</b>	$2.7E10^{-2}$	-0.12	$5.4E10^{-2}$	-0.10	$1.1E10^{-1}$	-0.12	$5.5E10^{-2}$
[U]	<b>0.15</b>	$4.2E10^{-2}$	<b>0.21</b>	$2.6E10^{-2}$	<b>0.30</b>	$1.3E10^{-3}$	0.05	$5.9E10^{-1}$
[V]	-0.19	$9.8E10^{-2}$	-0.22	$1.4E10^{-1}$	<b>-0.29</b>	$4.8E10^{-2}$	<b>-0.31</b>	$5.0E10^{-2}$
<b>Cell location</b>								
Cytoplasmic	-0.07	$4.0E10^{-1}$	-0.08	$4.3E10^{-1}$	N.S		N.S	
Inn. Mb. Lipo.	-0.12	$1.1E10^{-1}$	-0.11	$2.6E10^{-1}$	N.S		N.S	
Int. Mb. Pr.	-0.20	$3.9E10^{-1}$	-0.27	$3.8E10^{-1}$	N.S		N.S	
Mb. Anchored	0.04	$6.3E10^{-1}$	0.02	$8.6E10^{-1}$	N.S		N.S	
Mb. Lipo.	-0.04	$6.6E10^{-1}$	-0.11	$3.4E10^{-1}$	N.S		N.S	
Out. Mb. B-Bar Pr.	0.48	$1.4E10^{-1}$	0.55	$1.9E10^{-1}$	N.S		N.S	
Out. Mb. Lipo.	<b>0.34</b>	$3.7E10^{-2}$	<b>0.48</b>	$2.4E10^{-2}$	N.S		N.S	
Periplasmic	-0.20	$7.4E10^{-1}$	-0.21	$1.5E10^{-1}$	N.S		N.S	
Unknown	<b>-0.23</b>	$3.8E10^{-2}$	<b>-0.27</b>	$1.0E10^{-2}$	N.S		N.S	
<b>Peptide Signal</b>	-0.09	$5.4E10^{-2}$	-0.12	$5.8E10^{-2}$	N.S		N.S	
<b>Adjusted R<sup>2</sup></b>	<b>0.74</b>		<b>0.56</b>		<b>0.57</b>		<b>0.51</b>	

**[mRNA]:** mRNA concentration. **CAI:** Codon Adaptation Index. **5'UTR + ORF GC%:** percentage of G and C bases in combined 5'UTR and ORF sequences. **Nb of genes in operon:** Number of genes in the operon in which is located the gene. **COG annotation** – [C]: Energy production and conversion, [D]: Cell cycle control, cell division, chromosome partitioning, [E] Amino acid transport and metabolism, [F]:

Nucleotide transport and metabolism, [G]: Carbohydrate transport and metabolism, [H]: Coenzyme transport and metabolism, [I]: Lipid transport and metabolism, [J]: Translation, ribosomal structure and biogenesis, [K]: Transcription, [L]: Replication, recombination and repair, [M]: Cell wall/membrane/envelope biogenesis, [N]: Cell motility, [O]: Posttranslational modification, protein turnover, chaperones, [P]: Inorganic ion transport and metabolism, [Q]: Secondary metabolites biosynthesis, transport and catabolism, [R]: General function prediction only, [S]: Function unknown, [T]: Signal transduction mechanisms, [U]: Intracellular trafficking, secretion, and vesicular transport, [V]: Defense mechanisms. **Cell location:** Gene product cell location – Inn. Mb. Lipo.: Inner membrane Lipoprotein, Int. Mb. Pr.: Integral Membrane Protein, Mb. Anchored: Membrane Anchored, Mb. Lipo.: Membrane Lipoprotein, Out. Mb. B-Bar Pr.: Outer Membrane B-barrel protein, Out. Mb. Lipo.: Outer membrane Lipoprotein. N.S means that the parameter was not selected by AIC algorithm. Only parameters selected at least once by AIC algorithm are represented. Model determinants (P-value  $\leq 0.05$ ) are displayed in bold and highlighted in grey.

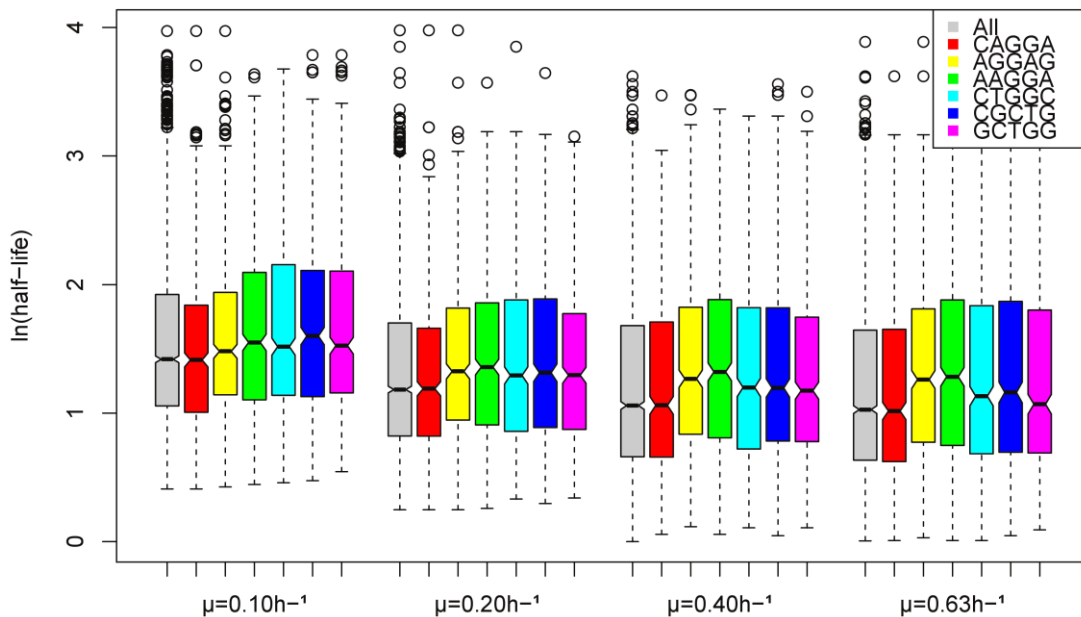


**Figure 1.** Cistron position dependent-half-life in polycistronic mRNA for four growth rates. This figure represents boxplot distribution of log<sub>2</sub> ratios of the cistron half-lives at position 2, 3 and 4 versus the half-life of the first cistron for the four growth rates ( $\mu = 0.10, 0.20, 0.40$  and  $0.63 \text{ h}^{-1}$  in black, blue, red and green, respectively). Only operons with four cistrons are represented in this figure (47 operons). This figure is representative of all the operons with other sizes.

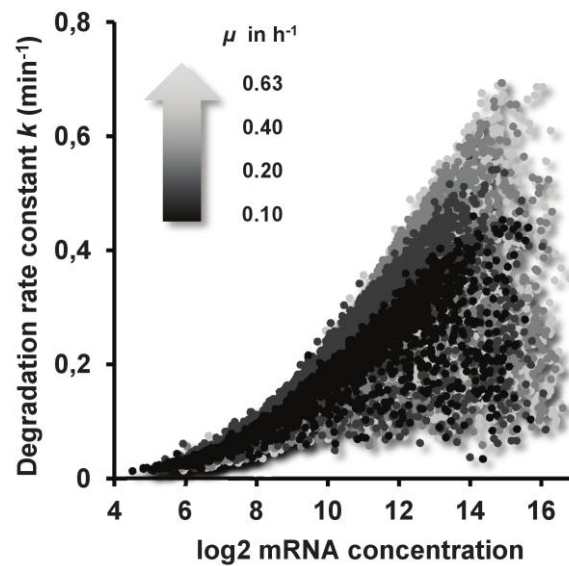


**Figure 2.** Motif frequencies in 5'UTR regions of stable and unstable mRNAs. RMES scores of all 5-nt length motifs are plotted in the most stable (x-axis) and unstable mRNAs (y-axis) at low and high growth rates (0.10 and 0.63 h<sup>-1</sup>, respectively). Exceptional motifs exhibit either low score values (under-represented motifs) or high score values (over-represented motifs). Exceptional motifs that are found distant from the scatter plot in one of the two conditions exhibit distinct patterns of exceptionality in stable and unstable mRNAs; they are tagged and described in the dedicated box.





**Figure 3.** Notched boxplot (44) of half-lives of transcripts containing motifs in 5'UTR. Distributions of mRNA half-lives of the entire mRNA population “All” (grey) (2882 sequences) can be compared to subpopulations of mRNAs whose 5'UTR region contained motif “CAGGA” (red) (#589), “AGGAG” (yellow) (#547), “AAGGA” (green) (#613), “CTGGC” (sky blue) (#592), “CGCTG” (deep blue) (#528) or “GCTGG” (purple) (#535) at growth-rate  $\mu = 0.10, 0.20, 0.40$  and  $0.63 \text{ h}^{-1}$ . The box shows the interquartile range of the data, while the notch displays the 95 % confidence interval around the median. When the intervals of two groups do not overlap, then their medians can be assumed to be different with 95 % confidence.



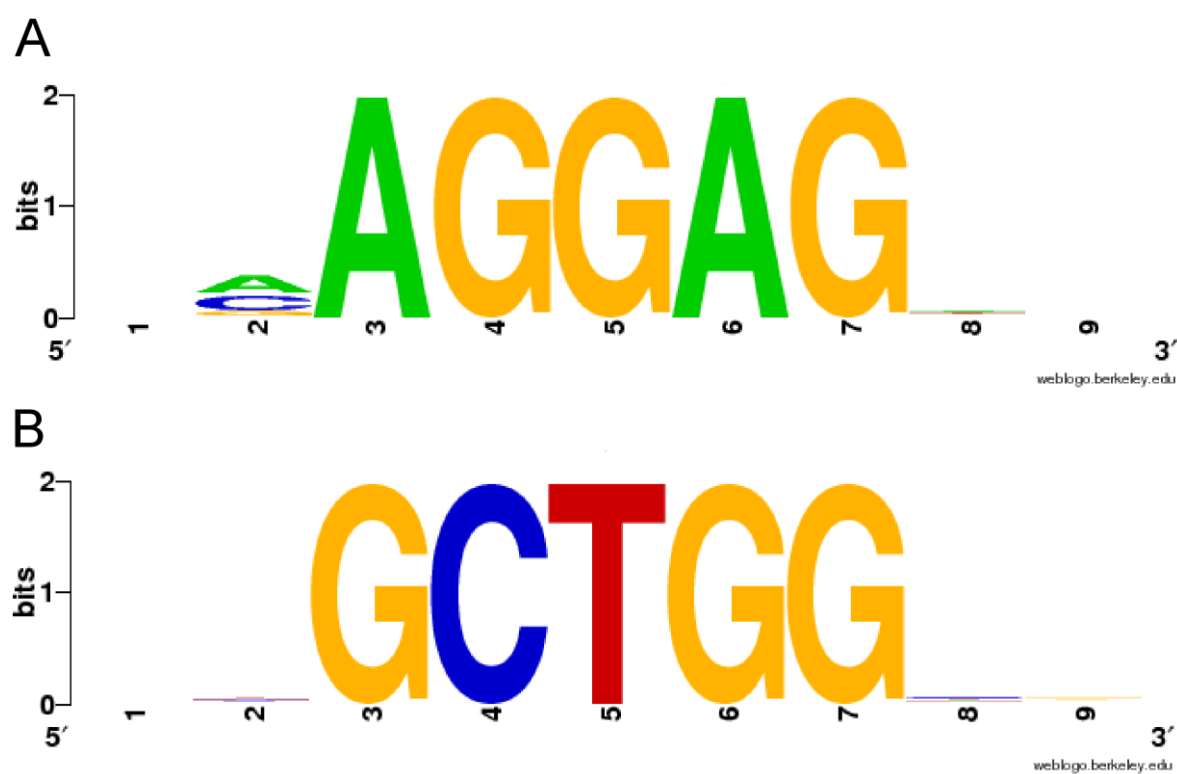
**Figure 4.** Effect of the growth rate on the relationship between the degradation rate constant  $k$  and the mRNA concentration at different growth rates  $\mu = 0.10, 0.20, 0.40$  and  $0.63 \text{ h}^{-1}$  (corresponding to gradient from black to grey).

**Table 2.** Estimated coefficients of protein concentration models for a selection of explicative variables

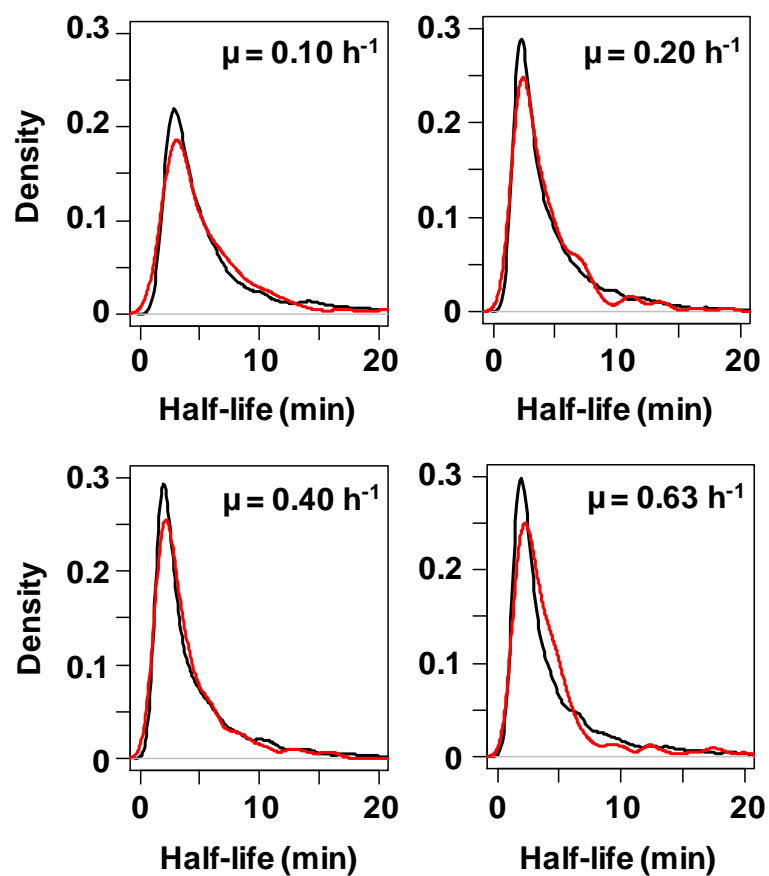
Parameter	Variable to explain : protein concentration							
	$\mu = 0.10 \text{ h}^{-1}$		$\mu = 0.20 \text{ h}^{-1}$		$\mu = 0.40 \text{ h}^{-1}$		$\mu = 0.63 \text{ h}^{-1}$	
	Est. Coeff.	<i>P-value</i>	Est. Coeff.	<i>P-value</i>	Est. Coeff.	<i>P-value</i>	Est. Coeff.	<i>P-value</i>
Half-life time	<b>-0,14</b>	<b><math>1.7E^{-05}</math></b>	N.S		N.S		<b>0.08</b>	<b><math>1.1E^{-02}</math></b>
CAI	<b>0.48</b>	<b><math>2.4E^{-30}</math></b>	<b>0.50</b>	<b><math>3.1E^{-37}</math></b>	<b>0.56</b>	<b><math>2.6E^{-46}</math></b>	<b>0.54</b>	<b><math>5.4E^{-43}</math></b>
Adjusted R <sup>2</sup>	<b>0.46</b>		<b>0.47</b>		<b>0.49</b>		<b>0.50</b>	

**CAI:** Codon Adaptation Index. N.S means that the parameter was not selected by AIC algorithm. Model determinants ( $P\text{-value} \leq 0.05$ ) are displayed in bold and highlighted in grey.

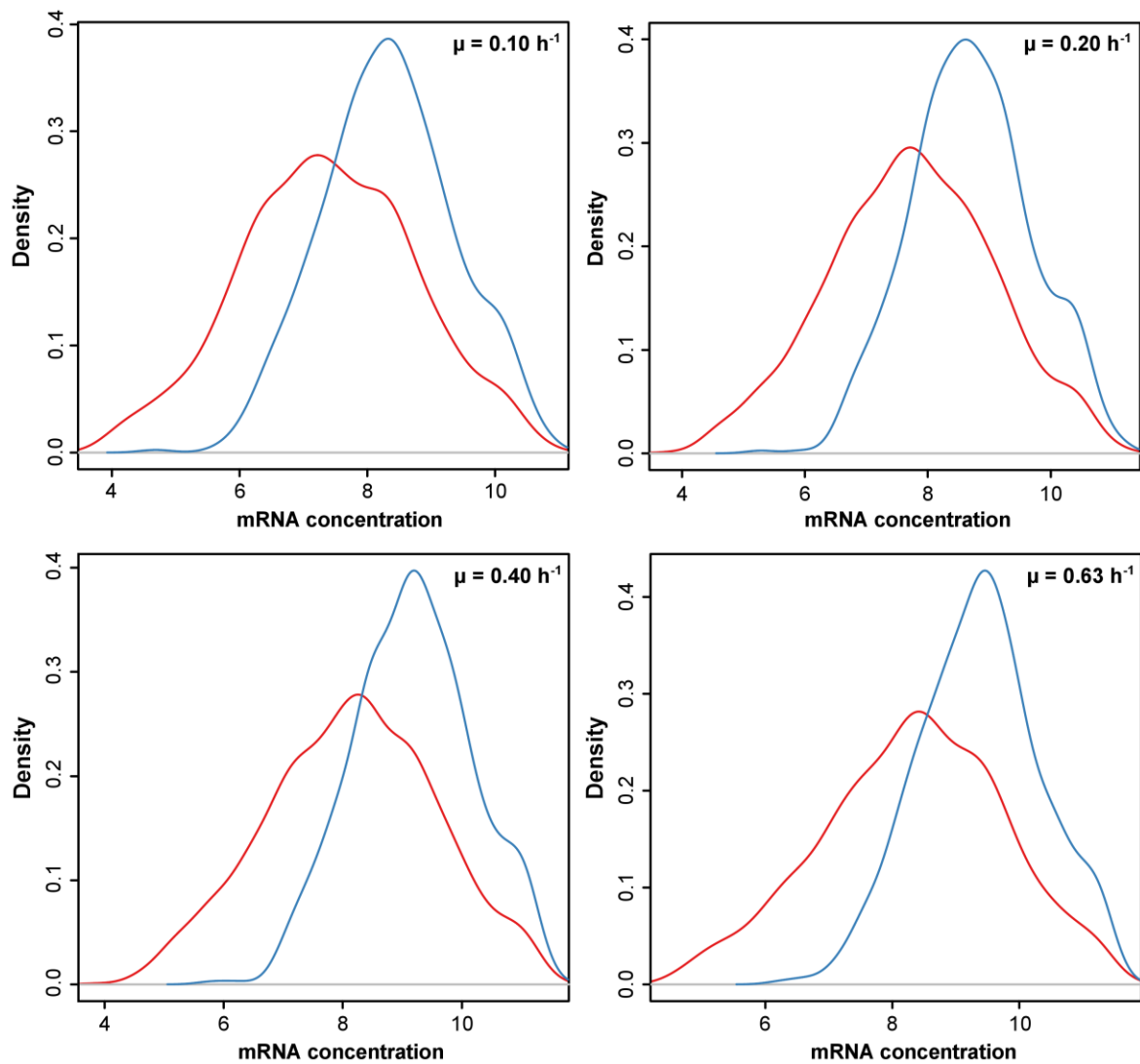
SUPPLEMENTARY FIGURES



**Figure S1.** Flanking sequence consensus of (A) AGGAG and (B) GCTGG motifs in the 5'UTR region of the entire mRNA population. Consensus was represented as sequence logos created by alignment of the 9-nucleotide subsequences overlapping each motif using WebLogo software (45).



**Figure S2.** Half-life densities of mRNAs with low Zscores ( $< -2$ ) (black) and the rest of the mRNA population (red) at different growth rates.



**Figure S3.** Density curves of mRNA levels. Density curve of the 643 mRNA levels included in the protein level model is represented in blue. Density curve of the 1589 mRNA levels included in the half-life model is represented in red. Density curves are represented for each growth rate studied.



# **CHAPITRE III**

**Détermination à l'échelle omique de  
l'étendue du contrôle de l'expression et  
de la stabilité des ARNm par le système  
*Csr* chez *Escherichia coli***





La protéine CsrA est un régulateur post-transcriptionnel qui modifie l'expression de certains ARNm en ajustant l'initiation de leur traduction et/ou en modifiant leur stabilité. Toutefois, l'implication de la protéine dans la régulation de la dégradation de ses « cibles » est souvent suggérée mais uniquement démontrée pour un nombre très restreint de messagers. Ce nouveau chapitre a pour but d'identifier l'impact d'une mutation de la protéine CsrA sur le stabilome d'*E. coli* afin de déterminer la portée d'action du régulateur à l'échelle du génome. D'autre part, l'effet de la délétion du gène *csrD* codant pour un régulateur de l'activité de la protéine CsrA sur les données de transcriptome et de stabilome a aussi été analysé dans ce contexte.

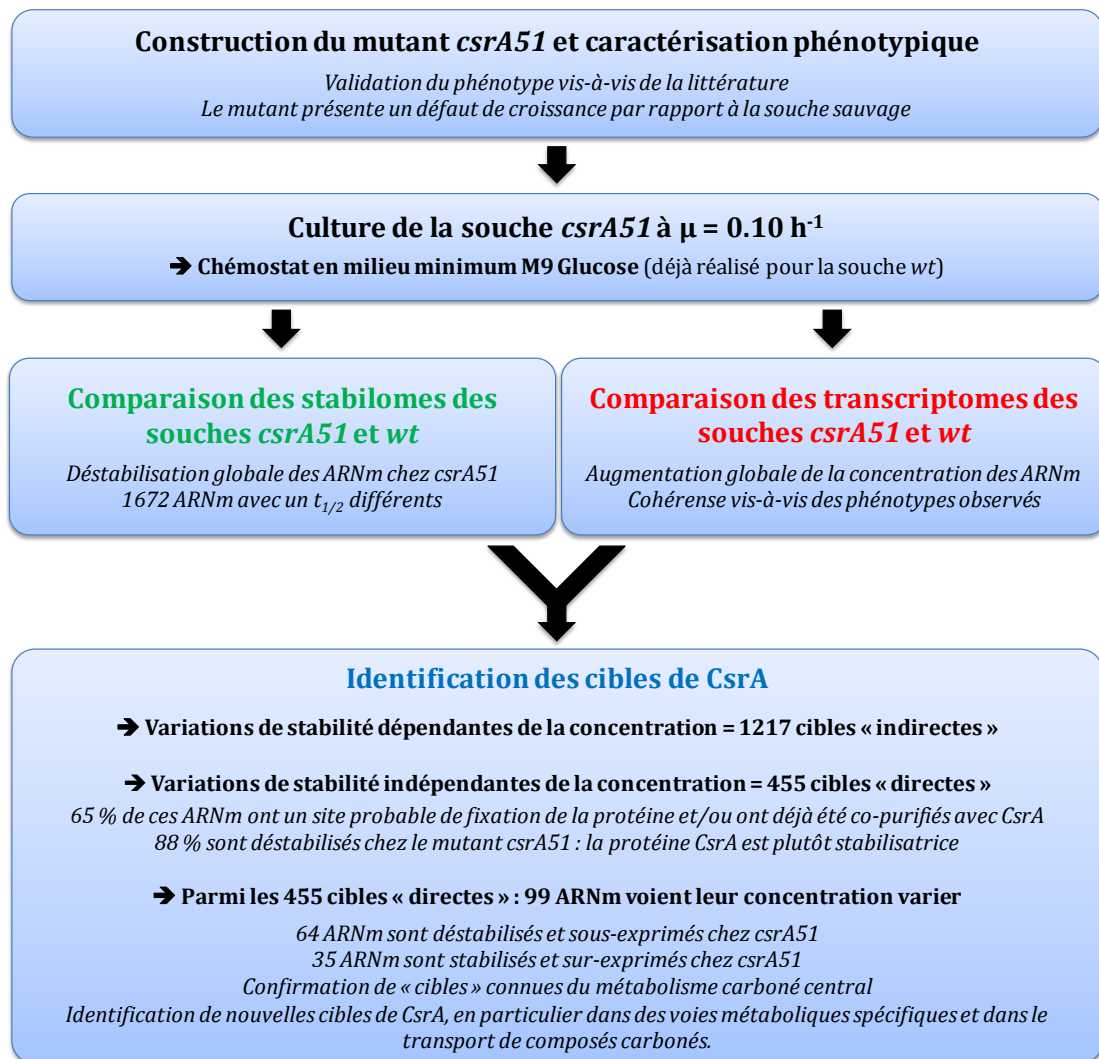
Cette étude a tout d'abord nécessité la construction d'un nouveau mutant de délétion partielle du gène *csrA* réalisée au Laboratoire de Microbiologie et de Génétique Moléculaires dans l'équipe « ARN et ribonucléases » dirigée par A.J. Carpousis. La délétion totale du gène n'était pas envisageable en raison de sa létalité. La construction de ce mutant, nommé *csrA51*, entraînant une diminution de l'activité de la protéine CsrA a été confirmée par caractérisation phénotypique. En effet, les fonctions cellulaires régulées par la protéine CsrA sont affectées dans la souche *csrA51* de façon cohérente vis-à-vis des données de la littérature (Timmermans & Van Melderen, 2010). Par ailleurs, cette caractérisation a mis en évidence un fort défaut de croissance de ces cellules en milieu minimum supplémenté en glucose. Étant donné que les travaux du premier chapitre de cette thèse ont souligné un impact du taux de croissance sur la stabilité des ARNm, la comparaison des stabilomes des souches *wt* et *csrA51* a nécessité d'utiliser des méthodes de cultures continues (chémostat) pour pouvoir comparer les deux souches à un taux de croissance identique. Ainsi, seules les différences de temps de demi-vie de messagers liées à la mutation du gène *csrA* ont été mesurées. Globalement, la mutation *csrA51* mène à une dégradation supérieure des ARNm suggérant un effet plutôt stabilisateur de la protéine CsrA à l'échelle omique. Toutefois, le deuxième chapitre de cette thèse a montré que la stabilité d'un ARNm dépend fortement de sa concentration (corrélation négative). Ainsi, l'analyse transcriptomique, qui montre une surexpression globale des transcrits, a permis d'isoler 455 ARNm dont la variation de stabilité entre les souches *wt* et *csrA51* n'était pas liée à une variation du niveau d'ARNm dans la cellule (exclusion des gènes qui présentent des variations opposées de niveaux d'ARNm et de stabilité). Ces 455 ARNm ont donc été considérés comme des cibles « directes » de CsrA. De plus, une grande majorité d'entre eux dispose d'au moins un site

probable de fixation de CsrA et/ou avait déjà été co-purifiée avec la protéine dans une étude précédente. D'autre part, 88 % de ces messagers sont déstabilisés dans la souche *csrA51* par rapport à la souche *wt* confirmant que la protéine CsrA favoriserait davantage la stabilité de ses cibles que l'inverse. Toutefois, dans nos conditions de croissance, nous avons recherché les cibles prépondérantes de CsrA définies comme celles pour lesquelles la protéine exerce aussi un contrôle sur la concentration des ARNm. Ces transcrits présentent à la fois une variation significative de stabilité et de concentration et la plupart d'entre eux ont été démontrés comme étant principalement régulés par la dégradation d'après l'analyse des coefficients de contrôle définie dans le Chapitre I. Ainsi, 99 messagers parmi les 455 correspondant à des cibles majeures de CsrA ont pu être identifiés. La majorité d'entre eux est déstabilisée et sous-exprimée dans le mutant *csrA51*. L'analyse fonctionnelle de ces gènes a permis de confirmer certaines « cibles » déjà connues du métabolisme carboné central mais également de mettre en évidence de nouvelles implications de CsrA dans la régulation de voies métaboliques spécifiques (acides aminés, acides gras...) et de transports de sucres. Enfin, l'analyse des transcriptomes et stabilomes d'une souche  $\Delta csrD$  montre pour la première fois que la protéine CsrD régulerait la concentration d'un grand nombre d'ARNm indépendamment de leur stabilité et donc de l'activité de CsrA. De plus, les réponses transcriptomiques des souches *csrA51* et  $\Delta csrD$  présentent des signatures opposées. Nos résultats suggèrent ainsi que la surexpression globale des ARNm dans le mutant *csrA51* serait partiellement liée à la surexpression de CsrD dont l'ARNm est très fortement surexprimé dans la souche *csrA51*.

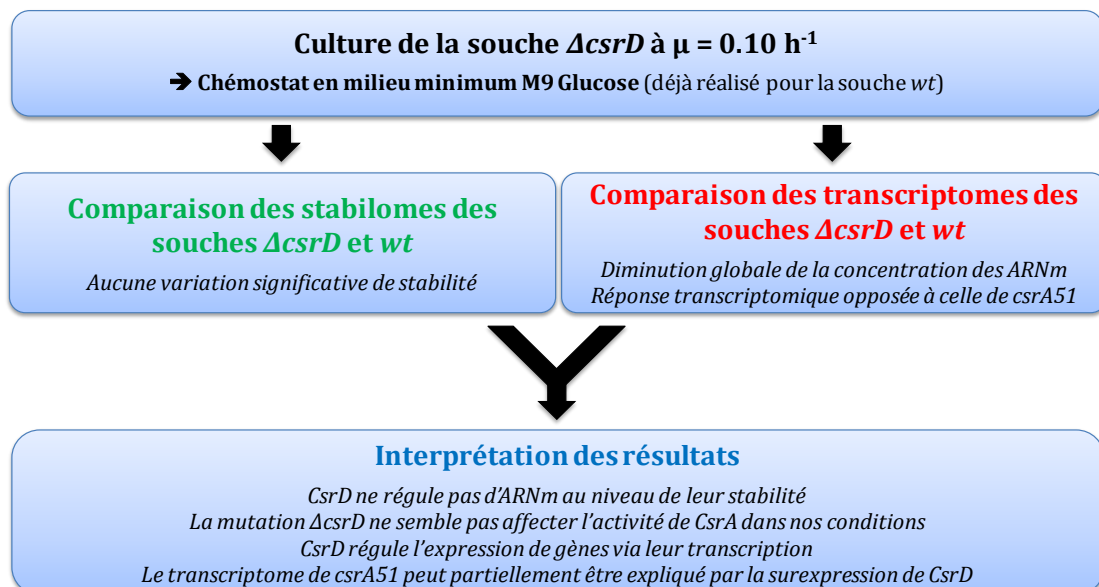
En conclusion, cette étude a permis d'évaluer pour la première fois l'impact d'une mutation du gène *csrA* sur la stabilité des ARNm à l'échelle omique. Ces travaux fournissent un aperçu de l'étendue de l'action de CsrA à la fois au niveau transcriptionnel et post-transcriptionnel. D'autre part, ce chapitre met également en lumière un rôle de la protéine CsrD dans la régulation de l'expression génique, probablement indépendamment du système Csr.

Cette étude a été rédigée sous la forme d'un article qui sera très prochainement soumis dans une revue à comité de lecture. Seule la Figure S1 est présentée à la suite du manuscrit dans un souci de clarté. Les Tables S1 et S2, de très grandes tailles sont présentées dans la partie Annexes de la thèse.

## Évaluation de l'impact d'une mutation de *csrA* sur la stabilité des ARNm



## Évaluation de l'impact de CsrD dans le mutant *csrA51*



## Genome-wide extent of the Csr system on mRNA expression and stability in *Escherichia coli*

Thomas Esquerré<sup>a,b,c,d</sup>, Catherine Turlan<sup>d</sup>, Marie Bouvier<sup>d</sup>, Agamemnon J. Carpousis<sup>d</sup>, Muriel Cocaign-Bousquet<sup>a,b,c,\*</sup> and Laurence Girbal<sup>a,b,c</sup>

<sup>a</sup> Université de Toulouse; INSA, UPS, INP; LISBP, 135, avenue de Rangueil, 31077 Toulouse, France

<sup>b</sup> INRA, UMR792 Ingénierie des systèmes biologiques et des procédés, 31400 Toulouse, France

<sup>c</sup> CNRS, UMR5504, 31400 Toulouse, France

<sup>d</sup> Laboratoire de Microbiologie et Génétique Moléculaires, UMR5100, Centre National de la Recherche Scientifique et Université Paul Sabatier, 118 Route de Narbonne, 31062, Toulouse, France.

\* To whom correspondence should be addressed. Tel: (33) 561 559 465; Fax: (33) 561 559 400; Email: cocaign@insa-toulouse.fr

### ABSTRACT

The post-transcriptional regulator CsrA was demonstrated to control gene expression in *E. coli* by regulating in certain specific cases the transcript stability. This study investigated for the first time at the genome scale the effect of the CsrA protein on messenger stability by comparing the half-lives of each mRNA in the *wt* and *csrA* mutant strains. The reduced CsrA activity in the mutant *csrA51* strain led to a lower growth rate in minimal medium supplemented with glucose. Using continuous cultures of the two strains at the same growth rate, we showed surprisingly that mRNA stability was in general lower in the *csrA51* cells than in the *wt* cells suggesting a large scale stabilizing effect of CsrA. For a large part of the messengers, CsrA controlled mRNA stability indirectly through transcriptional regulation of mRNA level. In contrast, for 455 mRNAs, change in half-life was

independent of mRNA concentration and was likely linked to a post-transcriptional control of mRNA stability by CsrA. For most of these mRNAs, this control was attributed to a physical interaction between CsrA and the mRNA molecule. In our growth conditions, 99 out of the 455 mRNAs were identified as preponderant targets of CsrA since they showed significant mRNA stability and concentration variations and were mainly associated to a degradational control of mRNA level. In addition to confirm known CsrA targets, the results demonstrated the implication of CsrA in regulating specific metabolic pathways and carbohydrate transporters. Finally, the CsrD protein was demonstrated to have a strong effect on mRNA concentration via transcriptional regulation which could partially explain the transcriptomic response observed in the *csrA51* mutant.

## INTRODUCTION

Bacterial cells have developed multiple mechanisms to adapt their physiology and metabolism to changing environment in order to compete with others species. One of these mechanisms is the regulation of gene expression which involves different levels such as transcriptional, post-transcriptional and post-translational regulations. In *Escherichia coli* cells, the Csr system (for Carbon Storage Regulator) post-transcriptionally regulates gene expression of several cell functions. For example, Csr system represses glycogen metabolism, gluconeogenesis, c-di-GMP synthesis, biofilm formation or quorum sensing while it controls positively glycolysis or motility (1). Several elements form the Csr system. The main component is the global regulatory protein CsrA which regulates translation and/or stability of mRNA targets (2, 3). The *csrA* gene is essential for cell viability in *E. coli* K-12 strains on glycolytic sources (4, 5). However, the *csrA::kan* mutant, corresponding to a transposon insertion after the position 50 of the 61 amino acid protein, was viable and used in numerous works studying the role of CsrA (6–9). CsrA is a RNA binding protein (10) targeting specific sequences whose consensus was determined by SELEX experiments as RUACARGGAUGU (11). In this sequence, the GGA motif was highly conserved among identified targets and its location in the loop of a hairpin increased CsrA affinity for the substrate (11). Non-coding small RNAs (sRNA) CsrB, CsrC and the recently discovered McaS, act as negative regulators of CsrA activity by sequestering the protein (12–15). Finally, the GGDEF-EAL CsrD protein indirectly regulates CsrA activity by targeting CsrB and CsrC to degradation by the RNase E

nuclease (16). CsrD was also suggested to focus other targets in the cells (3) and demonstrated to stimulate expression of the CsgD transcriptional factor (17).

The regulation of the Csr system is extremely complex because it involves multiple autoregulatory circuits (3). The *csrA* gene displays  $\sigma^{70}$  and  $\sigma^S$ -dependent promoters and the protein is able to post-transcriptionally negatively regulate itself and to promote its own transcription by an unknown mechanism (18). Moreover, CsrA was also demonstrated to indirectly activate the transcription of its antagonistic sRNAs CsrB and CsrC through the BarA-UvrY two component signal transduction system (13, 19–21). Although the two sRNAs are able to compensate each other, emerging evidence shows that CsrB and CsrC transcriptional activations by BarA-UvrY are slightly different (22, 23). Furthermore, CsrA negatively controls the expression of the *csrD* gene adding a new feedback loop in the regulation of Csr system (16, 24). Finally, CsrA acts on global regulators such as *hfq* (25) or the stringent response system (26) further complicating the regulation network.

Up to date, two molecular mechanisms of post-transcriptional regulation by CsrA are described in the literature. On one hand, CsrA represses gene expression by binding to the 5'UTR and/or the initially translated region and inhibits translation by competing with ribosome binding (27, 28). Inhibition of translation by CsrA can lead to rapid decay of the transcript (29, 30). On the other hand, CsrA can favor expression of some genes such as those of the glycolytic enzymes, *acs* and *flhDC* (7, 31, 32). In the case of *flhDC* mRNA, CsrA stabilizes the transcript by binding to the 5'UTR and sequestering the cleavage site of the RNase E (33).

These studies clearly demonstrated the role of CsrA on the mRNA decay but only on a limited number of targets. However, the extent and physiological importance of mRNA stability regulation by CsrA at the omic level is not yet clearly identified. Therefore, we have investigated at the genome-wide scale the effect of CsrA activity on mRNA stability. The half-lives of each mRNA in the *wt* and *csrA* mutant strains were determined and compared. Since we previously showed that the half-life value of an mRNA could be partially dependent on its concentration (Esquerré et al, in preparation, [Chapitre II]), we used transcriptomic analysis to discriminate between mRNA half-life variations related indirectly to mRNA level change to those directly regulated by CsrA activity. For more than 450 mRNAs, the stability was demonstrated to be directly controlled by CsrA. For these mRNAs, we examined the potentially of a CsrA binding and the contribution of the mRNA stability

variation in the physiology of the cell via mRNA level variation and regulation coefficient analyses. In addition, the mRNA half-life and level variations observed in the *csrA* mutant strain were further investigated in relation with a putative role of the CsrD protein.

## MATERIALS AND METHODS

### Bacterial Strains and Plasmids

*E. coli* K-12 MG1655 strain ( $\lambda^-$   $F^-$   $rph^{-1}$ ) was used as a genetic background for mutation of *csrA* and *csrD* genes. Mutations were carried out using  $\lambda$  red system recombination (34). As the *csrA* gene is essential (4, 5), the ORF has been mutated using as a model of the *csrA* mutant obtained by Romeo and coworkers (2) Our *csrA* mutant strain corresponded to a partial deletion of the 11 last amino acids of the protein and insertion of a stop codon. It was named the *csrA51* strain. The  $\Delta csrD$  strain contained a total deletion of the *csrD* gene obtained from KEIO collection (4). The mutations were transferred into *E. coli* K-12 MG1655 *wt* background by P1 phage-mediated transduction before removing antibiotic-resistance cassette by FLP recombination. The constructions were controlled by sequencing. Complementation of *csrA51* mutation corresponding to the *csrA+* strain was only possible by transforming the *csrA51* strain with a low copy number pSAB11 plasmid (35) containing the wild type version of the *csrA* gene.

### Culture conditions

Cells were grown in M9 minimal medium supplemented with glucose as described in (36). In chemostat mode, the *wt*, *csrA51* and  $\Delta csrD$  strains were grown at the same growth rate  $\mu = 0.10 \text{ h}^{-1}$  (corresponding to doubling times of 6.9 h). Biomass was estimated from absorbance at 600 nm (Libra S4, Biochrom): 1 unit of absorbance corresponding to 0.42 g of dry cell weight  $\text{l}^{-1}$  for the *wt* strain, 0.27 g of dry cell weight  $\text{l}^{-1}$  for the *csrA51* strain and 0.44 g of dry cell weight  $\text{l}^{-1}$  for the  $\Delta csrD$  strain. The maximal growth rate of each strain was obtained during the exponential phase in batch culture (with identical medium and environment such as oxygen, pH and temperature). Each culture was repeated three times to provide independent biological replicates.



### Sampling and RNA extraction

Sampling and RNA extraction were conducted as described in (36) with additional cell samplings after rifampicin addition at 9, 15, 20 and 30 min.

### Northern Blot

For each sample, 10 µg of total RNA were loaded onto 6 % polyacrylamide gel containing 7 M urea for 1 h 30 min migration at 250 V. RNA was transferred to charged nylon membrane (Amersham Hybond-XL – GE Healthcare) and cross-linked with UV light.  $\alpha$ -<sup>32</sup>P UTP radio-labeled riboprobes targeting CsrB and CsrC were hybridized to the membrane at 65 °C overnight. 5S rRNA was used as loading control and was targeted by 5'  $\gamma$ -<sup>32</sup>P radio-labeled oligoprobe. After several washings, the membrane was dried and exposed overnight with storage phosphor screen (BAS MS Imaging Plates – Fujifilm). Results were read using storage phosphor autoradiography imager (Typhoon Trio – Amersham Biosciences) and scanned images were analyzed and quantified with MultiGauge software (Fujifilm).

### Microarray procedures

A double-stranded cDNA synthesis kit (InvitroGen) was used to produce cDNA from 2 µg aliquots of total RNA. Aliquots of 1 µg of cDNA were labeled using the one color DNA labeling kit and 2 µg of labeled cDNA were hybridized onto *E. coli* K-12 gene expression arrays (Nimblegen – Roche) for 17 h at 42 °C according to the manufacturers' instructions. Arrays were then washed, and scanned with a MS200 Microarray Scanner (Nimblegen – Roche). The images were analyzed with DEVA 1.2.1 software. Only raw data were used for further analyses. All array procedures were performed by the GeT-Biopuces platform (<http://get.genotoul.fr>).

### mRNA concentration and half-life determination by microarray experiments

mRNA concentration determination by transcriptomic analysis was conducted as described in (36). Intensity values were multiplied by the total RNA extraction yield (in µg total RNA per mg of dry cell weight) to provide mRNA concentration value in arbitrary units per mg of dry cell weight. RNA extraction yields were  $13.1 \pm 2.2$ ,  $8.7 \pm 0.9$  and  $8.9 \pm 1.5$  µg RNA per mg of dry cell weight for the *wt*, *csrA51* and  $\Delta$ *csrD* strains, respectively. Differences in mRNA concentration were evaluated as

described in (36). The P-values were adjusted for multiple testing by the 'Benjamini and Hochberg' (BH) false discovery rate method. (37). Differences in mRNA concentration were considered as significant for adjusted P-values lower than 1 % and  $\log_2$  fold change ( $\log_2$  FC) higher than 0.5 or lower than -0.5. mRNA half-life determination was performed as previously described (36). The statistical significance of differences in half-life was evaluated using the probability value of interaction between time and the type of strain (*wt*, *csrA51* or  $\Delta$ *csrD*) in a global model of linear regression. A statistical threshold of 10 % was used for adjusted P-values by the "BH" false discovery rate method (37).

#### **Determination of mRNA concentration regulatory coefficients**

For the selected mRNAs, the coefficient  $\rho_D$ , corresponding to the relative contribution of the mRNA stability regulation in the control of mRNA concentration was calculated as previously described (36) using data of mRNA stability and concentration obtained in the *csrA51* and *wt* strains. Regulatory classes were defined according to the  $\rho_D$  value:  $0 < \rho_D < 0.6$ , mRNA concentration mainly controlled by transcription or controlled by both transcription and mRNA stability;  $0.6 < \rho_D$ , mRNA concentration mainly controlled by mRNA stability.

#### **Enrichment methods**

R free statistical software ([www.r-project.org](http://www.r-project.org)) was used for enrichment methods. Functional categories enriched in transcript subgroups were determined using an hypergeometric test on data using the Biological Process of Gene Ontology annotation database (GO project; <http://www.geneontology.org/>). Only GO terms with associated P-value  $\leq 0.05$  were considered as significant.

#### **In silico research of CsrA binding sites**

To determine if an mRNA contained a potential CsrA binding site, the software PatScan was used (38). The SELEX-derived consensus RUACARGGAUGU was researched in -100 / +100 bp region (+1 corresponding to the start codon) of listed mRNAs allowing a maximum of 5 mismatches. Subsequently, only sequences which displayed a conserved GGA motif were considered as a potential CsrA binding site.

## Phenotype assays

Glycogen staining by iodine vapor, biofilm formation by cristal violet staining and motility assays on 0.3 % agar plate were carried out as previously described (39). Three independent experiments were performed for all the assays.

## RESULTS

### Construction and phenotype of the *csrA51* mutant

Although inspired by the TR1-5 *csrA* mutant strains of Romeo and coworkers (2), the construction of our *csrA51* strain was quite different since it was not based on transposon insertion. The *csrA51* strain was obtained by removing the last 11 amino acids of the CsrA protein and inserting a stop codon after the 50th codon to force arrest of translation. To check the phenotype of the *csrA51* strain, well-known cell functions controlled by the Csr system were investigated. We found that *csrA51* mutant strain produced much more glycogen than the *wt* strain and this phenotype was abolished by complementation with a *csrA wt* gene (Figure 1A). The *csrA51* strain displayed significantly less motility and more biofilm formation compared to the *wt* strain (Figure 1B). Measurement by Northern Blot of CsrB and CsrC levels in the *csrA51* strain showed the absence of expression of the two sRNAs (Figure 1C) confirming the role of CsrA in transcriptional activation of CsrB and CsrC. All these results confirmed that the activity of CsrA was drastically diminished in our *csrA51* strain affecting several well-known physiological functions under the control of the protein (1). In addition, when cultured in M9 minimal medium supplemented with glucose, *csrA51* cells exhibited a two-fold reduced maximal growth rate  $\mu_{\max} \text{ } csrA51 = 0.30 \text{ h}^{-1}$  compared to the reference  $\mu_{\max} \text{ } wt = 0.63 \text{ h}^{-1}$ ; a similar growth rate as the *wt* was restored by complementation with a *csrA wt* gene (Figure 1D,1E). This is the first quantification of the deleterious effect of removing the last 11 amino acids of CsrA on the growth rate in minimum medium.

### mRNA half-life analysis in the *csrA51* mutant strain

To determine the role of CsrA at the genome-wide scale on the mRNA half-life regulation, the half-life of each mRNA was determined and compared between the *csrA51* and *wt* strains. Since the *csrA51* and *wt* strains differed strongly in their growth rate in batch cultures and since the growth rate affects the mRNA stability

(36), half-life measurements were performed during chemostat cultures of both strains maintained at a constant growth rate of  $\mu = 0.10 \text{ h}^{-1}$ . Therefore, change in mRNA half-life between the *csrA51* and *wt* strains was only linked to the CsrA mutation and not to growth rate variation. 3028 mRNAs (71 % of the 4254 *E. coli* genes) exhibited reliable half-lives in both strains. In the two strains, at least 91 % mRNA half-lives were lower than 12 min, but the transcripts were globally less stable in *csrA51* than in the *wt* strain, the median half-life decreasing from 4.1 min in the *wt* to 2.8 min in the mutant strain (Figure 2A). More precisely, a statistical test (P-value  $\leq 0.1$ ) identified 1672 messengers with differential half-life between the *csrA51* and *wt* strains: 1618 were significantly destabilized in the mutant strain whereas only 54 were significantly stabilized (Figure 2B). Interestingly, these results suggested a large scale stabilizing effect of CsrA.

In order to go deeper in the description of the relationship between CsrA and mRNA half-life regulation, we wanted to separate the “indirect” and “direct” effects of CsrA on mRNA stability. In a previous study, we have demonstrated that globally, more an mRNA is concentrated, less stable it is and vice versa (Esquerré et al, in preparation, [Chapitre II]). So we will define as an “indirect” effect of CsrA on mRNA stability, half-life variation that can be mainly attributed to mRNA level variation. In contrast, we will define as a “direct” CsrA effect on mRNA stability, half-life variation independent of the mRNA level. To do so, we used transcriptomic analysis to measure mRNA concentration changes of each messenger between the *csrA51* and *wt* strains.

### **Effect of the *csrA51* mutation on mRNA concentrations**

Transcriptomic analyses showed a global up-regulation of the mRNA concentrations in the *csrA51* mutant strain with 2195 mRNA concentrations significantly up-regulated compared to the *wt* strain and only 249 mRNA concentrations significantly down-regulated (Figure 3). Functional category enrichments in the two groups were consistent with the phenotypic characterization of the mutant strain and the literature. The functions “Flagellum organization” and “Ciliary or flagellar motility” (containing *flhDC* gene, *fli* and *flg* operons) were enriched in the down-regulated mRNA levels correlating with reduced motility (32, 40). GO term “Single species biofilm formation” (including *csg* and *ydeH* genes) was enriched in the group of up-regulated mRNA concentrations as well as *yddV*, *yjjC* and *yliE* mRNA encoding GGDEF-EAL proteins

involved in c-di-GMP synthesis (8, 24). Moreover, numerous transcripts whose expression was known to be controlled by CsrA including *pck*, *cstA*, *hfq*, *relA* and *spoT* adopted consistent behavior with the literature (2, 25, 26, 41). The concentrations of all the transcripts involved in glycogen synthesis (*glgC*, *glgB*, *glgA* and *glgP/Y*) were at least 3-fold up-regulated. Finally, the strong transcriptomic response in *csrA51* was correlated with the enrichment of the function “Transcription, DNA-dependent” in the group of up-regulated mRNA concentrations. In particular were up-regulated in the *csrA51* strain mRNAs of the RNA polymerase ( $\beta$  and  $\beta'$  subunits, *rpoB* and *rpoC*) and sigma factors (*rpoD*, *rpoE*, *rpoH*, *rpoN*, *rpoS* and *fecI*), mRNAs related to the stringent response (*relA*, *spoT* and *dksA*), and those of global transcriptional regulators (*crp* concomitantly with *cyaA* encoding the adenylate cyclase, *fruR/cra* for regulating the carbon flow, *fnr* and *arcA* involved in regulation of the aerobic/anaerobic metabolisms and *ihfA* and *ihfB* encoding the IHF factor involved in chromosomal structure) (Table 1). These results suggested that the global up-regulation of mRNA concentrations in the mutant background was likely related to higher expression of several transcriptional factors.

#### **Identification of mRNA half-lives directly and indirectly regulated by CsrA**

Because we previously demonstrated that a mRNA half-life time could be dependant on its concentration (Esquerré et al, in preparation), we used both stabilome and transcriptome data to identify mRNA whose stability is indirectly and directly controlled by CsrA. Among the 1672 mRNAs with differential half-life between the *csrA51* and *wt* strains, 1217 mRNAs exhibited opposite variations of mRNA half-life and concentration. For these mRNAs, half-life variation could be mainly attributed to mRNA level variation: this group corresponded to mRNAs whose stabilities were indirectly controlled by CsrA. For the remaining 455 mRNAs, the half-lives varied independently of the mRNA level. Consequently they constituted the group of mRNA stabilities directly controlled by CsrA. The presence of a consensus CsrA binding site was investigated in the -100 / +100 bp sequences (+1 corresponding to the start codon) of these transcripts (see Material and Methods). We found that 180 mRNAs displayed unique CsrA while 78 had at least 2 sites with hot spot localization in the 5'UTR region. A cross-comparison of our list of CsrA direct targets with the sets of mRNAs previously shown to co-purify with CsrA (26) indicated that 93 of the 455 direct targets can bind CsrA. Taken together, more than 65 % (297/455) CsrA direct

targets were found to bind CsrA and/or to have a CsrA binding site. These results indicated that most of the identified CsrA direct targets likely interact physically with the CsrA protein (Table S1). We found that more than 88 % (403/455) of the CsrA direct targets were destabilized in the *csrA51* strain compared to the *wt* (Table S1). These results surprisingly demonstrate that CsrA had a more stabilizing rather than destabilizing effect on its direct targets.

### Functional impact of mRNA stability regulated by CsrA

Among the 455 mRNAs with stability directly regulated by CsrA between the *csrA51* and *wt* strains, 356 mRNA stability regulations were not associated with significant change in mRNA level. On the contrary, for 99 transcripts, mRNA stability regulation was associated with differential mRNA concentration (Table S2). For these mRNAs, we wanted to determine the relative contribution of the mRNA half-life regulation compared to transcription in the control of the mRNA concentration. We analyzed the control of mRNA concentration, through mRNA stability versus transcription, by calculating regulatory coefficients (36). The results showed that for 85 mRNAs, half-life regulation predominantly controlled mRNA concentration (degradational control), whereas for 14 mRNAs, the control was equally divided between mRNA stability and transcription or imposed by transcription (Table S2). These results underline the significant contribution of the mRNA stability regulation by CsrA in the control of mRNA concentration changes.

More precisely, 64 mRNAs were destabilized and down-expressed whereas 35 mRNAs were stabilized and up-expressed between the *csrA51* and *wt* strains (Table S2). These genes constituted the targets whose stability is the most influenced by the CsrA protein. Among the destabilized and down-expressed mRNAs, we found *tpiA* in agreement with previous results showing that the triose-phosphate isomerase activity was positively regulated by CsrA activity (31). We also identified the *zwf* transcript, previously reported to co-purify with CsrA (26), but its expression was never shown to be regulated by the Csr system. Our results clearly showed that, by regulating the mRNA half-life and concentration, CsrA promoted the expression of carbohydrate transport (*araF*, *malE*, *alsB*, *idnT*, *melB*, *glpF*, *srlAEB* and *agaVW*) and certain metabolic pathways, e.g. idonate/gluconate metabolic process (*idnDOT*) and galacturonate catabolic process (*uxaCA*). mRNA stability and concentration of a large number of metabolic enzymes (*aldA*, *ldhA*, *udp*, *yfaU*, *hdhA*,

*glpK*, *dmlA*, *gloB*, *aes*, *fdhF*, *speB*, *tpx*, *thiH*, *ydiB*, *ilvC*, *pptA*, *ebgC* and *frlB*) were found to be positively regulated by CsrA .

The 35 mRNAs stabilized and up-expressed in the *csrA51* strain corresponded to mRNAs with lower stability and concentration in the presence of CsrA (Table S2). At the metabolic level, several genes belonging to the glycogenesis and gluconeogenesis pathways were found in this group: *maeA*, *maeB*, *pck*, *gapA*, *pgm* and *glgB*. The expressions and/or activity of some of these genes/proteins were known to be repressed by CsrA (2, 31). Our results provide a better understanding of the CsrA effect on these targets, the regulator acting at the levels of both mRNA stability and concentration. Several genes of the Krebs cycle and the amino acid metabolism were found to be repressed by CsrA protein through mRNA stability: *acnB*, *fumA*, *sdhC*, *leuB*, *leuC*, *glnA* and *asd*. We observed that the messengers of the mannose PTS permease (*manXYZ*), of the pyridine nucleotide transhydrogenase (*pntAB*) and of the fatty acid synthesis (*fabD* and *fabF*) were destabilized and repressed by CsrA. CsrA seemed to also destabilize and repress mRNAs related to transcription (*rpoD* of the primary sigma factor  $\sigma^{70}$  of RNA polymerase) and to translation (ribosomal *rplP*, *rplN* and *rpsS* transcripts). This could partially explain the slightly lower biomass yield in the *wt* strain compared to the mutant ( $Y_{x/s} = 36 \pm 1$  and  $41 \pm 2$  g of dry cell weight per g glucose in the *wt* and *csrA51* strains respectively). These results provide new insights into how CsrA-mediated regulations of mRNA stability contribute to metabolic pathway regulation.

#### **Role of CsrD in the *csrA51* genetic background**

As expected according to the feedback loop of the CsrA protein on the *csrD* gene expression (16, 24), *csrD* mRNA concentration was found strongly higher in the *csrA51* strain compared to the *wt* (fold change higher than 4.6). Previously, it was suggested that CsrD could regulate the stability of other targets than the sRNA CsrB and CsrC (3) and demonstrated that CsrD stimulated expression of the CsgD transcriptional factor (17). Consequently, analyses of mRNA half-lives and transcriptome in the  $\Delta csrD$  mutant strain were performed to investigate the possible role of the CsrD protein in the responses observed in the *csrA51* mutant background. The mRNA half-life and transcriptome measurements were carried out in the  $\Delta csrD$  mutant strain cultivated in a chemostat at  $\mu = 0.10 \text{ h}^{-1}$ . The growth rate chosen was

the same as the experiments with the *wt* and *csrA51* strains to eliminate the growth rate effects on the mRNA half-lives and concentrations.

Although the median mRNA half-life in  $\Delta csrD$  was slightly higher than in the *wt* strain, no mRNA with a significant variation of stability was found between the  $\Delta csrD$  and *wt* strains. This observation was quite surprising since according to the regulatory network of the Csr system, CsrD is believed to positively regulate CsrA activity via CsrB and CsrC. Consequently, a stabilization as already observed for *csrA51* strain was expected in the  $\Delta csrD$  strain. As previously shown (16), the half-lives of both CsrB and CsrC sRNAs were higher in  $\Delta csrD$  than in the *wt* strain (Figure S1) but surprisingly the regulations of their levels were very different: the CsrB level was increased around two fold whereas the CsrC level was divided around two fold (Figure 4A). Therefore, CsrA must exhibit rather similar activity in the  $\Delta csrD$  and *wt* strains. Consequently, in our growth conditions, the presence/absence of CsrD did not affect any mRNA stability neither via CsrD targeting nor via CsrA activity.

Nevertheless, a strong transcriptomic response was observed in the  $\Delta csrD$  strain compared to the *wt* (Figure 4B). The large majority of the mRNA concentrations (3345/4254) was down-regulated and only 61 transcript levels were significantly up-regulated. Comparison of the transcriptomic responses in the *csrA51* and  $\Delta csrD$  strains showed that 77 % (1878/2444) of differentially expressed mRNAs in *csrA51* were also differentially expressed in  $\Delta csrD$  and for 88 % of them (1653/1878) the regulations were in opposite directions: most of these mRNA were up-regulated in *csrA51* but down-regulated in  $\Delta csrD$ . More particularly for transcriptional regulators, up-regulated mRNA concentrations in the *csrA51* strain (where the *csrD* transcript was highly induced) were strongly down-regulated in the absence of CsrD (Table 2). Consequently, the transcriptomic response observed in the *csrA51* strain could be at least partially linked to the higher *csrD* gene expression in this mutant. Our results show that the CsrD protein regulates many mRNA concentrations preferentially through regulation of gene transcription since no regulation of mRNA stability was present in  $\Delta csrD$ .

## DISCUSSION

The regulatory Csr system is structured around the global post-transcriptional regulator CsrA. In this study, we showed for the first time that the *csrA51* mutation (resulting in the deletion of the last 11 amino acids in the CsrA protein) strongly



reduced compared to the *wt* strain the maximal growth rate of the strain on M9 minimal medium supplemented with glucose. Our results confirmed previously *csrA51* mutation-associated phenotypes and related regulations of mRNA stability and concentration e.g. the stabilization and increase of expression of the *glgCAP* operon for glycogen accumulation (2, 29).

To go deeper in the understanding of the role of CsrA in gene expression as a post-transcriptional regulator, we have performed the first genome-wide study of mRNA half-life regulation in the *csrA51* mutant strain. By choosing to center our approach on CsrA-associated mRNA stability regulation, this study did not investigate the targets whose mRNA degradation was not affected by CsrA, as this is the case for the *hfq* transcript (25).

At the same growth rate (imposed  $\mu = 0.10 \text{ h}^{-1}$  in continuous cultures), this study identified more than 1600 mRNAs with differential half-lives between the *csrA51* and *wt* strains. Around 96 % of the half-lives were shorter in the *csrA51* than in the *wt* cells. This result was surprising because, to date, most of the known targets of CsrA are repressed and/or destabilized, and at the molecular level, repression mechanisms by CsrA are more often described than the positive regulation (3, 27, 28, 33, 42). Thus, contrary to earlier studies on selected targets, the genome-wide analysis emphasized the significance of the CsrA protein as a positive regulator of mRNA stability.

The results demonstrated clearly a dual control of mRNA stability by CsrA. For 72 % of the mRNAs, change in half-life could mainly be attributed to an indirect effect of CsrA via mRNA level variation. An increase/decrease in concentration will result in a lower/higher half-life, respectively. This type of regulation requires necessarily a strong transcriptional control possibly via a transcriptional regulator whose expression depends on CsrA. We found many transcriptional regulators with CsrA-dependent expression; this could explain the large extent of such indirect regulation. In contrast for 28 % of the mRNAs, change in half-life was independent of mRNA concentration and was likely linked to a direct control of mRNA stability by CsrA. For the majority of these mRNAs, this control was attributed to a physical interaction between CsrA and the mRNA molecule. For a limited number of mRNAs, CsrA could act through regulation of expression of other post-transcriptional regulators (like the RNA-binding protein Hfq). Altogether these results underline the complexity of CsrA

contribution to mRNA stability control acting either indirectly via transcriptional regulation or directly via post-transcriptional regulation.

In term of gene expression regulation, for a hundred of mRNAs, the direct regulation of mRNA stability by CsrA led to a significant variation in mRNA concentration. These genes were considered as the major targets of CsrA at the level of mRNA stability. However for a large majority of mRNAs, change in mRNA stability was compensated by transcriptional regulation leading to no significant variation of the mRNA concentration and thus to no real physiological impact. These results reflect the flexibility of the *E. coli* gene expression regulatory network to adapt gene expression to growth conditions. The growth condition used in this study was a continuous culture on glucose at low growth rate. Utilization of other growth conditions (different substrate or growth rate value) would probably modify the number and nature of the mRNAs which were direct stability targets of CsrA with associated variation in concentration. Nevertheless, in our growth condition, we confirmed previously identified targets of CsrA such as *glgB*, *pgm*, *tpiA* and *pck* mRNAs (2, 31). In addition, many new direct mRNA targets of CsrA were discovered with a potential metabolic impact via mRNA concentration changes. Functional analysis of these transcripts underlined that the Csr system is strongly involved in the control of *E. coli* metabolism by regulating expression of enzymes involved in the central carbon metabolism (Krebs cycle, gluconeogenesis, glycolysis) as previously demonstrated (31) but also in other specific metabolic pathways (idonate and galacturonate metabolism, syntheses of amino acids and fatty acids) and different carbohydrate transporters. Consequently, these results enlarge the list of the metabolic CsrA regulon (43).

This study provides new insights into the Csr regulatory network. In our conditions, the transcriptomic response in the *csrA* mutant strain could be partially explained by an overexpression of CsrD. The CsrD protein was demonstrated for the first time at a genome-wide scale to favor mRNA expression. We showed that CsrD did not control the stability of any mRNA meaning that CsrD regulated mRNA expression at the transcriptional level. This could occur via transcriptional factors whose expression is CsrD-dependent. Indeed, we observed many down-regulated expressions of transcriptional regulators in the absence of CsrD as it was previously reported for *csgD* expression (17). Concerning the mechanism, we could suggest that CsrD targets the stability of small RNAs others than CsrB and CsrC (16) which

could in turn control transcriptional regulator expression. In our experiments, the stabilizations of CsrB and CsrC in the absence of CsrD led to opposite behaviors of their respective concentration leading to a similar activity of CsrA in the presence/absence of CsrD. This result supports the conclusion that although CsrB and CsrC are functionally related, they differ quantitatively in the regulation of their expression (16). The mechanisms allowing differential transcriptional regulation of CsrB and CsrC remain to be elucidated.

To conclude, this study has contributed to better define the nature and physiological significance of the CsrA regulon and to improve our understanding of the Csr system. In particular, the regulations mediated by the Csr system have highlighted the complex connection of the transcriptional and post-transcriptional regulatory networks. Sharing together the control of the mRNA level in the cell, transcription and mRNA stability regulations are interconnected processes acting either co-directionally or in opposite directions. Consequently, both regulatory levels have to be taken into account to decipher the control of mRNA concentration by a post-transcriptional regulator.

### FUNDING

The funding of this work was provided by Université de Toulouse and Région Midi-Pyrénées. Funding for open access charge: Université de Toulouse and Région Midi-Pyrénées.

### REFERENCES

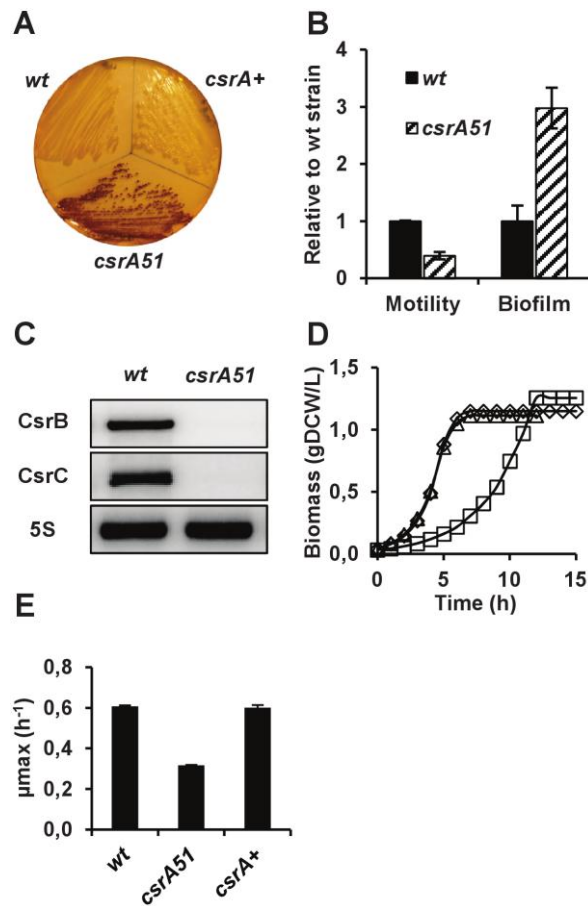
1. Timmermans, J. and Van Melderen, L. (2010) Post-transcriptional global regulation by CsrA in bacteria. *Cell. Mol. Life Sci. CMLS*, 67, 2897–2908.
2. Romeo, T., Gong, M., Liu, M.Y. and Brun-Zinkernagel, A.M. (1993) Identification and molecular characterization of *csrA*, a pleiotropic gene from *Escherichia coli* that affects glycogen biosynthesis, gluconeogenesis, cell size, and surface properties. *J. Bacteriol.*, 175, 4744–4755.
3. Romeo, T., Vakulskas, C.A. and Babitzke, P. (2012) Post-transcriptional regulation on a global scale: form and function of Csr/Rsm systems. *Environ. Microbiol.*, 10.1111/j.1462-2920.2012.02794.x.
4. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L. and Mori, H. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.*, 2, 2006.0008.

5. Timmermans, J. and Van Melder, L. (2009) Conditional essentiality of the *csrA* gene in *Escherichia coli*. *J. Bacteriol.*, 191, 1722–1724.
6. Yang, H., Liu, M.Y. and Romeo, T. (1996) Coordinate genetic regulation of glycogen catabolism and biosynthesis in *Escherichia coli* via the *csrA* gene product. *J. Bacteriol.*, 178, 1012–1017.
7. Wei, B., Shin, S., LaPorte, D., Wolfe, A.J. and Romeo, T. (2000) Global Regulatory Mutations in *csrA* and *rpoS* Cause Severe Central Carbon Stress in *Escherichia coli* in the Presence of Acetate. *J. Bacteriol.*, 182, 1632–1640.
8. Jackson, D.W., Suzuki, K., Oakford, L., Simecka, J.W., Hart, M.E. and Romeo, T. (2002) Biofilm Formation and Dispersal under the Influence of the Global Regulator CsrA of *Escherichia coli*. *J. Bacteriol.*, 184, 290–301.
9. Pannuri, A., Yakhnin, H., Vakulskas, C.A., Edwards, A.N., Babitzke, P. and Romeo, T. (2012) Translational repression of NhaR, a novel pathway for multi-tier regulation of biofilm circuitry by CsrA. *J. Bacteriol.*, 194, 79–89.
10. Liu, M.Y. and Romeo, T. (1997) The global regulator CsrA of *Escherichia coli* is a specific mRNA-binding protein. *J. Bacteriol.*, 179, 4639–4642.
11. Dubey, A.K., Baker, C.S., Romeo, T. and Babitzke, P. (2005) RNA sequence and secondary structure participate in high-affinity CsrA-RNA interaction. *RNA*, 11, 1579–1587.
12. Liu, M.Y., Gui, G., Wei, B., Preston, J.F., Oakford, L., Yüksel, Ü., Giedroc, D.P. and Romeo, T. (1997) The RNA Molecule CsrB Binds to the Global Regulatory Protein CsrA and Antagonizes Its Activity in *Escherichia coli*. *J. Biol. Chem.*, 272, 17502–17510.
13. Weilbacher, T., Suzuki, K., Dubey, A.K., Wang, X., Gudapaty, S., Morozov, I., Baker, C.S., Georgellis, D., Babitzke, P. and Romeo, T. (2003) A novel sRNA component of the carbon storage regulatory system of *Escherichia coli*. *Mol. Microbiol.*, 48, 657–670.
14. Babitzke, P. and Romeo, T. (2007) CsrB sRNA family: sequestration of RNA-binding regulatory proteins. *Curr. Opin. Microbiol.*, 10, 156–163.
15. Jørgensen, M.G., Thomason, M.K., Havelund, J., Valentin-Hansen, P. and Storz, G. (2013) Dual function of the McaS small RNA in controlling biofilm formation. *Genes Dev.*, 10.1101/gad.214734.113.
16. Suzuki, K., Babitzke, P., Kushner, S.R. and Romeo, T. (2006) Identification of a novel regulatory protein (CsrD) that targets the global regulatory RNAs CsrB and CsrC for degradation by RNase E. *Genes Dev.*, 20, 2605–2617.
17. Sommerfeldt, N., Possling, A., Becker, G., Pesavento, C., Tschowri, N. and Hengge, R. (2009) Gene expression patterns and differential input into curli fimbriae regulation of all GGDEF/EAL domain proteins in *Escherichia coli*. *Microbiology*, 155, 1318–1331.

18. Yakhnin, H., Yakhnin, A.V., Baker, C.S., Sineva, E., Berezin, I., Romeo, T. and Babitzke, P. (2011) Complex regulation of the global regulatory gene *csrA*: CsrA-mediated translational repression, transcription from five promoters by E $\sigma$ 70 and E $\sigma$ (S), and indirect transcriptional activation by CsrA. *Mol. Microbiol.*, 81, 689–704.
19. Gudapaty, S., Suzuki, K., Wang, X., Babitzke, P. and Romeo, T. (2001) Regulatory Interactions of Csr Components: the RNA Binding Protein CsrA Activates *csrB* Transcription in *Escherichia coli*. *J. Bacteriol.*, 183, 6017–6027.
20. Suzuki, K., Wang, X., Weilbacher, T., Pernestig, A.-K., Melefors, Ö., Georgellis, D., Babitzke, P. and Romeo, T. (2002) Regulatory Circuitry of the CsrA/CsrB and BarA/UvrY Systems of *Escherichia coli*. *J. Bacteriol.*, 184, 5130–5140.
21. Chavez, R.G., Alvarez, A.F., Romeo, T. and Georgellis, D. (2010) The physiological stimulus for the BarA sensor kinase. *J. Bacteriol.*, 192, 2009–2012.
22. Jonas, K., Tomenius, H., Römling, U., Georgellis, D. and Melefors, Ö. (2006) Identification of YhdA as a regulator of the *Escherichia coli* carbon storage regulation system. *FEMS Microbiol. Lett.*, 264, 232–237.
23. Jonas, K. and Melefors, Ö. (2009) The *Escherichia coli* CsrB and CsrC small RNAs are strongly induced during growth in nutrient-poor medium. *FEMS Microbiol. Lett.*, 297, 80–86.
24. Jonas, K., Edwards, A.N., Simm, R., Romeo, T., Römling, U. and Melefors, Ö. (2008) The RNA binding protein CsrA controls cyclic di-GMP metabolism by directly regulating the expression of GGDEF proteins. *Mol. Microbiol.*, 70, 236–257.
25. Baker, C.S., Eöry, L.A., Yakhnin, H., Mercante, J., Romeo, T. and Babitzke, P. (2007) CsrA Inhibits Translation Initiation of *Escherichia coli hfq* by Binding to a Single Site Overlapping the Shine-Dalgarno Sequence. *J. Bacteriol.*, 189, 5472–5481.
26. Edwards, A.N., Patterson-Fortin, L.M., Vakulskas, C.A., Mercante, J.W., Potrykus, K., Vinella, D., Camacho, M.I., Fields, J.A., Thompson, S.A., Georgellis, D., et al. (2011) Circuitry Linking the Csr and Stringent Response Global Regulatory Systems. *Mol. Microbiol.*, 10.1111/j.1365-2958.2011.07663.x.
27. Baker, C.S., Morozov, I., Suzuki, K., Romeo, T. and Babitzke, P. (2002) CsrA regulates glycogen biosynthesis by preventing translation of *glgC* in *Escherichia coli*. *Mol. Microbiol.*, 44, 1599–1610.
28. Yakhnin, H., Baker, C.S., Berezin, I., Evangelista, M.A., Rassin, A., Romeo, T. and Babitzke, P. (2011) CsrA represses translation of *sdiA*, encoding the N-acylhomoserine-L-lactone receptor of *Escherichia coli*, by binding exclusively within the coding region of *sdiA* mRNA. *J. Bacteriol.*, 10.1128/JB.05975-11.

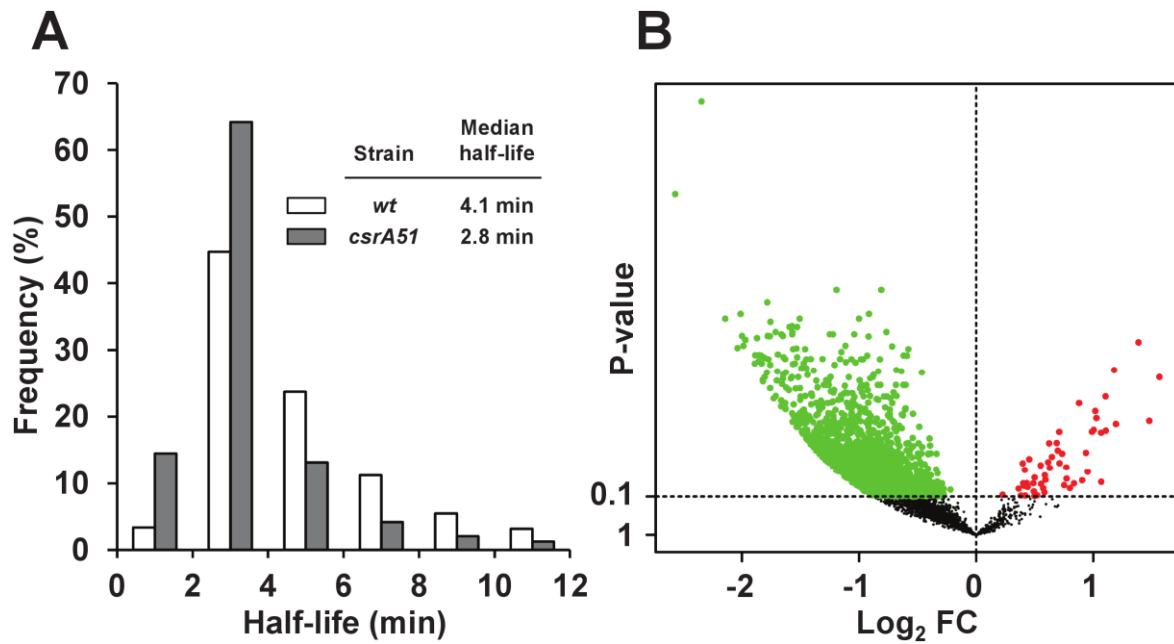
29. Liu, M.Y., Yang, H. and Romeo, T. (1995) The product of the pleiotropic *Escherichia coli* gene *csrA* modulates glycogen biosynthesis via effects on mRNA stability. *J. Bacteriol.*, 177, 2663–2672.
30. Wang, X., Dubey, A.K., Suzuki, K., Baker, C.S., Babitzke, P. and Romeo, T. (2005) CsrA post-transcriptionally represses *pgaABCD*, responsible for synthesis of a biofilm polysaccharide adhesin of *Escherichia coli*. *Mol. Microbiol.*, 56, 1648–1663.
31. Sabnis, N.A., Yang, H. and Romeo, T. (1995) Pleiotropic regulation of central carbohydrate metabolism in *Escherichia coli* via the gene *csrA*. *J. Biol. Chem.*, 270, 29096–29104.
32. Wei, B.L., Brun-Zinkernagel, A.-M., Simecka, J.W., Prüß, B.M., Babitzke, P. and Romeo, T. (2001) Positive regulation of motility and *flhDC* expression by the RNA-binding protein CsrA of *Escherichia coli*. *Mol. Microbiol.*, 40, 245–256.
33. Yakhnin, A.V., Baker, C.S., Vakulskas, C.A., Yakhnin, H., Berezin, I., Romeo, T. and Babitzke, P. (2013) CsrA activates *flhDC* expression by protecting *flhDC* mRNA from RNase E-mediated cleavage. *Mol. Microbiol.*, 87, 851–866.
34. Datsenko, K.A. and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci.*, 97, 6640–6645.
35. Aït-Bara, S. and Carpousis, A.J. (2010) Characterization of the RNA Degradosome of *Pseudoalteromonas haloplanktis*: Conservation of the RNase E-RhlB Interaction in the Gammaproteobacteria. *J. Bacteriol.*, 192, 5413–5423.
36. Esquerre, T., Laguerre, S., Turlan, C., Carpousis, A.J., Girbal, L. and Coccagn-Bousquet, M. (2014) Dual role of transcription and transcript stability in the regulation of gene expression in *Escherichia coli* cells cultured on glucose at different growth rates. *Nucleic Acids Res.*, 42, 2460–2472.
37. Benjamini, Y. and Hochberg, Y. (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.*, 57, 289–300.
38. Dsouza, M., Larsen, N. and Overbeek, R. (1997) Searching for patterns in genomic data. *Trends Genet.*, 13, 497–498.
39. Revelles, O., Millard, P., Nougayrède, J.-P., Dobrindt, U., Oswald, E., Létisse, F. and Portais, J.-C. (2013) The Carbon Storage Regulator (Csr) System Exerts a Nutrient-Specific Control over Central Metabolism in *Escherichia coli* Strain Nissle 1917. *PLoS ONE*, 8, e66386.
40. Lawhon, S.D., Frye, J.G., Suyemoto, M., Porwollik, S., McClelland, M. and Altier, C. (2003) Global regulation by CsrA in *Salmonella typhimurium*. *Mol. Microbiol.*, 48, 1633–1645.

41. Dubey, A.K., Baker, C.S., Suzuki, K., Jones, A.D., Pandit, P., Romeo, T. and Babitzke, P. (2003) CsrA Regulates Translation of the *Escherichia coli* Carbon Starvation Gene, *cstA*, by Blocking Ribosome Access to the *cstA* Transcript. *J. Bacteriol.*, 185, 4450–4460.
42. Patterson-Fortin, L.M., Vakulskas, C.A., Yakhnin, H., Babitzke, P. and Romeo, T. (2013) Dual Posttranscriptional Regulation via a Cofactor-Responsive mRNA Leader. *J. Mol. Biol.*, 425, 3662–3677.
43. Seyll, E. and Van Melderen, L. (2013) The Ribonucleoprotein Csr Network. *Int. J. Mol. Sci.*, 14, 22117–22131.
44. Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., et al. (1997) The Complete Genome Sequence of *Escherichia coli* K-12. *Science*, 277, 1453 – 1462.

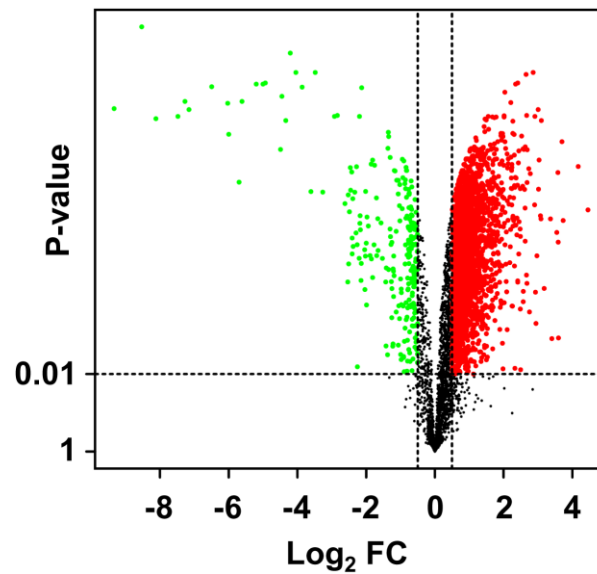


**Figure 1.** Phenotypic characterization of the *csrA51* mutant strain. (A) Glycogen staining assay in the *wt*, *csrA51* and *csrA*<sup>+</sup> strains. (B) Motility and biofilm formation assays in *csrA51*. (C) Northern Blots of CsrB and CsrC small RNAs in the *csrA51* strain. (D) Growth curves of the *wt* (◇), *csrA51* (□) and complemented *csrA*<sup>+</sup> (△) strains in M9 minimal medium supplemented with glucose and (E) maximal growth rates  $\mu_{max}$ .





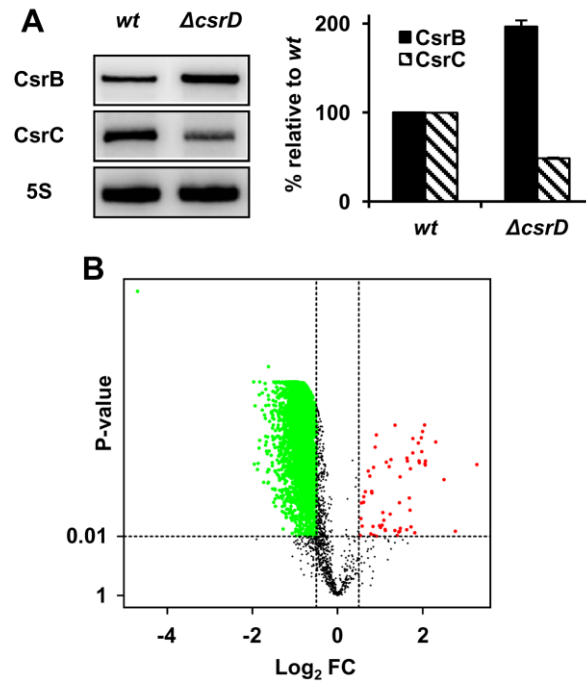
**Figure 2.** Effect of the *csrA51* mutation on mRNA half-life. (A) The histogram shows the relative frequency of the half-lives lower than 12 min in the *wt* (white bars) and *csrA51* (grey bars) strains. Respective median half-lives are listed in the inset. (B) VolcanoPlot of the log<sub>2</sub> fold change (Log<sub>2</sub> FC) of mRNA half-lives between the *csrA51* and *wt* strains. A P-value ≤ 0.1 was required for fold change significance (above the horizontal dashed line). The 54 significantly stabilized mRNAs in *csrA51* compared to the *wt* strain are colored in red whereas the 1618 significantly destabilized mRNAs are colored in green.



**Figure 3.** Effect of the *csrA51* mutation on mRNA concentration. VolcanoPlot of the  $\log_2$  fold change ( $\text{Log}_2 \text{FC}$ ) of mRNA levels between the *csrA51* and *wt* strains. A  $P\text{-value} \leq 0.01$  (above the horizontal dashed line) and a  $\log_2 \text{FC}$  higher than 0.5 or lower than -0.5 (outside the vertical dashed lines) were required for fold change significance. The 2195 significantly up-regulated concentrations in the *csrA51* strain were colored in red whereas the 249 significantly down-regulated concentrations were colored in green.

Geneld	GeneSym	FC	P-value
b0080	<i>fruR/cra</i>	1.40	1.6E <sup>-05</sup>
b0145	<i>dksA</i>	1.67	3.3E <sup>-05</sup>
b0912	<i>ihfB</i>	1.43	9.9E <sup>-05</sup>
b1334	<i>fnr</i>	1.31	4.1E <sup>-03</sup>
b1712	<i>ihfA</i>	1.30	2.8E <sup>-03</sup>
b2573	<i>rpoE</i>	1.36	4.3E <sup>-03</sup>
b2741	<i>rpoS</i>	1.28	1.9E <sup>-02</sup>
b2784	<i>relA</i>	2.05	1.0E <sup>-05</sup>
b3067	<i>rpoD</i>	1.26	3.5E <sup>-05</sup>
b3202	<i>rpoN</i>	2.06	1.3E <sup>-07</sup>
b3357	<i>crp</i>	1.39	4.1E <sup>-05</sup>
b3461	<i>rpoH</i>	2.28	1.2E <sup>-05</sup>
b3650	<i>spoT</i>	2.31	1.2E <sup>-05</sup>
b3806	<i>cyaA</i>	1.73	4.9E <sup>-07</sup>
b3987	<i>rpoB</i>	1.13	5.5E <sup>-04</sup>
b3988	<i>rpoC</i>	1.14	2.9E <sup>-02</sup>
b4293	<i>fecI</i>	2.13	2.3E <sup>-05</sup>
b4401	<i>arcA</i>	1.16	2.7E <sup>-02</sup>

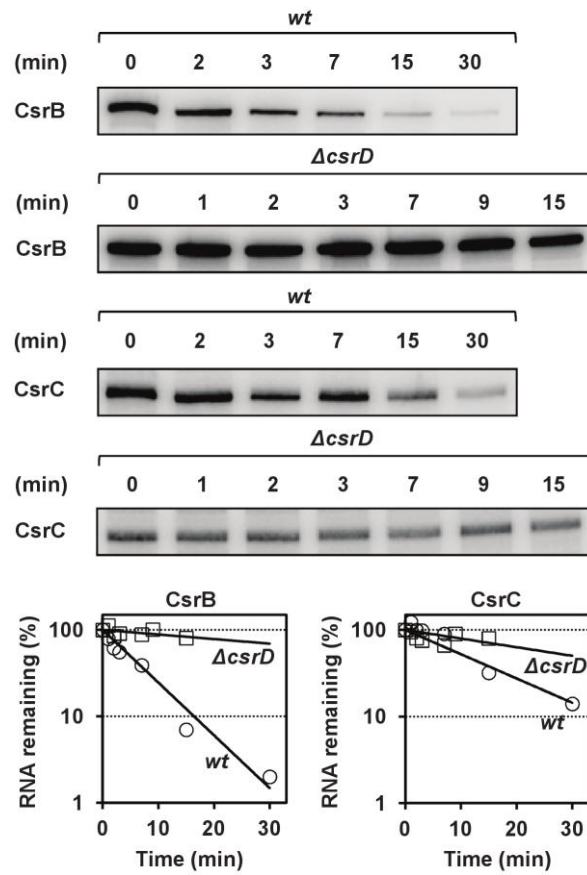
**Table 1.** Effect of the *csrA51* mutation on the mRNA concentration of some transcriptional regulators. Geneld corresponds to the gene Blattner number (44), GeneSym to the gene name, FC to the fold change of mRNA concentrations between the *csrA51* and *wt* strains and P-value is the associated adjusted P-value.



**Figure 4.** Effect of the  $\Delta csrD$  mutation on the concentrations of CsrB, and CsrC RNAs. (A) Steady state levels of CsrB and CsrC in the *wt* and  $\Delta csrD$  strains observed in Northern Blot and quantified by PhosphorImager. (B) VolcanoPlot of the log<sub>2</sub> fold change (Log<sub>2</sub> FC) of mRNA levels between the  $\Delta csrD$  and *wt* strains. A P-value  $\leq 0.01$  (above the horizontal dashed line) and a log<sub>2</sub> FC higher than 0.5 or lower than -0.5 (outside the vertical dashed lines) were required for fold change significance. The 61 significantly up-regulated concentrations in the  $\Delta csrD$  strain were colored in red whereas the 3345 significantly down-regulated concentrations were colored in green.

Geneld	GeneSym	FC	P-value
b0080	<i>fruR/cra</i>	0.50	2.4E <sup>-07</sup>
b0145	<i>dksA</i>	0.53	7.4E <sup>-06</sup>
b0912	<i>ihfB</i>	0.60	8.2E <sup>-06</sup>
b1334	<i>fnr</i>	0.55	3.0E <sup>-05</sup>
b1712	<i>ihfA</i>	0.58	3.0E <sup>-05</sup>
b2573	<i>rpoE</i>	0.51	3.7E <sup>-05</sup>
b2741	<i>rpoS</i>	0.49	4.0E <sup>-05</sup>
b2784	<i>relA</i>	0.53	1.6E <sup>-05</sup>
b3067	<i>rpoD</i>	0.55	1.8E <sup>-07</sup>
b3202	<i>rpoN</i>	0.54	2.5E <sup>-07</sup>
b3357	<i>crp</i>	0.48	4.0E <sup>-07</sup>
b3461	<i>rpoH</i>	0.51	3.0E <sup>-05</sup>
b3650	<i>spoT</i>	0.58	1.2E <sup>-04</sup>
b3806	<i>cyaA</i>	0.53	2.2E <sup>-07</sup>
b3987	<i>rpoB</i>	0.69	5.8E <sup>-07</sup>
b3988	<i>rpoC</i>	0.53	3.7E <sup>-06</sup>
b4293	<i>fecI</i>	0.54	6.6E <sup>-05</sup>
b4401	<i>arcA</i>	0.50	3.3E <sup>-06</sup>

**Table 2.** Effect of the *csrD* deletion on the mRNA concentration of some transcriptional regulators. Geneld corresponds to the gene Blattner number (44), GeneSym to the gene name, FC to the fold change of mRNA concentrations between the  $\Delta csrD$  and *wt* strains and P-value is the associated adjusted P-value.



**Figure S1.** Effect of the  $\Delta csrD$  mutation on the stability of CsrB, and CsrC RNAs. CsrB and CsrC stability in *wt* and  $\Delta csrD$  strains was measured in Northern Blot and quantified by PhosphorImager. All the experiments were repeated at least twice.



**CONCLUSION**

**ET**

**PERSPECTIVES**





### **Conclusion**

La compréhension des mécanismes d'adaptation des bactéries à leur environnement reste un des grands défis des sciences biologiques actuelles, notamment pour des raisons de santé publique. En effet, la maîtrise des régulations de l'expression des gènes permettrait par exemple, de moduler les mécanismes d'implantation et/ou de persistance de micro-organismes pathogènes chez l'Homme à des fins thérapeutiques.

La bactérie *Escherichia coli* est probablement la bactérie la plus étudiée en microbiologie en raison de sa culture aisée et des facilités de modification de son génome. D'autre part, elle constitue un groupe de bactéries très divers regroupant à la fois des souches commensales et pathogènes qui aujourd'hui encore, induisent de graves problèmes sanitaires. L'ensemble de ces raisons expliquent pourquoi *E. coli* a été sélectionnée comme modèle d'étude pour ces travaux de thèse qui portent sur les régulations de l'expression des gènes en réponse à l'environnement.

La régulation de l'expression génique comprend plusieurs niveaux de contrôle dont fait partie la régulation de la concentration des ARNm. Le niveau des ARNm dans la cellule est déterminé à la fois par leur synthèse, leur dégradation ainsi que leur dilution lorsque la cellule se divise. Dans le cadre de cette étude, nous avons vu que la dilution des ARNm pouvait être négligée car le temps de doublement de la bactérie était au minimum de l'ordre de l'heure alors qu'au moins 90 % de tous les messagers étudiés avaient des  $t_{1/2}$  inférieurs à 11 min dans toutes les conditions testées. Les régulations de cet équilibre entre synthèse et dégradation sont donc dépendantes des régulations de la transcription ainsi que de la stabilité des ARNm en fonction des conditions environnementales. Si les régulations transcriptionnelles ont été beaucoup étudiées, l'impact de la dégradation des ARNm à l'échelle omique en réponse à des changements environnementaux reste encore méconnu.

En utilisant les techniques de cultures continues et de batch, la stabilité des ARNm d'une souche sauvage d'*E. coli* K-12 MG1655 cultivée en milieu minimum supplémenté en glucose a été mesurée à quatre taux de croissance différents. Pour la première fois, cette étude démontre une déstabilisation générale des ARNm alors que le taux de croissance de la bactérie augmente. Toutefois, le facteur de variation de stabilité des ARNm n'est pas aussi important que celui du taux de croissance. En effet, au taux de croissance très faible  $\mu = 0.10 \text{ h}^{-1}$  (temps de génération de 7 h), bien que la machinerie de dégradation soit ralentie, nos résultats prouvent que cette dernière reste

particulièrement active (Chapitre I). L'observation d'une déstabilisation alors que les cellules se multiplient plus rapidement est retrouvée chez d'autres micro-organismes comme *Lactococcus lactis* ou même *Mycobacterium tuberculosis* (Dressaire et al., 2013; Rustad et al., 2013) et pourrait être expliquée par plusieurs éléments. Tout d'abord, nous avons observé que cette déstabilisation était corrélée à une augmentation de l'expression des messagers codant pour les protéines du dégradosome suggérant que le complexe protéique serait plus exprimé à taux de croissance élevé. Cette hypothèse semble cohérente avec le fait que nos données indiquent que la capacité maximale de dégradation des messagers augmente également. En effet, le front maximal de dégradation des ARNm croît avec le taux de croissance (Chapitre II). Enfin, la concentration des ARNm semble être le principal déterminant de leur stabilité. Les résultats du deuxième chapitre mettent en évidence que globalement, plus un ARNm est concentré, moins il est stable. Or, les données transcriptomiques indiquent que la plupart des transcrits voient leur concentration augmenter avec le taux de croissance. Ceci pourrait donc expliquer l'accélération de leur dégradation. Ces observations vont dans le sens d'une dégradation globalement « non ciblée » des ARNm où la probabilité de rencontre avec les RNases serait déterminante de la stabilité d'un grand nombre d'ARNm : quand la concentration en ARNm augmente, sa stabilité diminue et réciproquement. Ainsi, le rôle de la dégradation pour une grande majorité de messagers serait de contrebalancer l'effet des régulations de la transcription pour « équilibrer » la concentration des ARNm dans la cellule comme cela a été démontré pour le gène *fumA* (Lin et al., 2012). L'important nombre de transcrits contrôlés principalement par la transcription avec un effet opposé de la dégradation mis en évidence dans la première partie de cette thèse semble donc valider cette hypothèse. Les analyses d'enrichissement de catégories fonctionnelles montrent que ce sont plutôt les gènes impliqués dans la croissance cellulaire qui sont régulés de cette façon.

Cependant, la classification hiérarchique ascendante réalisée à partir des  $t_{1/2}$  dans le Chapitre I démontre qu'environ un millier de transcrits adoptent des profils de stabilité différents de la tendance générale. C'est le cas par exemple de l'ARNm du gène *ompA* dont la stabilité est connue pour augmenter avec le taux de croissance (Nilsson et al., 1984). Ces données confirment donc que la dégradation des ARNm n'est pas un mécanisme identique et universel pour tous les messagers et que des régulations plus spécifiques, ciblées et indépendantes de la concentration des ARNm interviennent aussi.

En effet, pour un grand nombre de transcrits dont le profil de stabilité s'écarte de la tendance générale, nous avons montré que ce sont les variations de stabilité qui contrôlent majoritairement les variations de leur concentration et non les variations de transcription. Bien que seuls deux messagers aient été identifiés sous le contrôle permanent de la dégradation, les concentrations de plus de 650 ARNm ont été régulées par leurs variations de stabilité pour au moins une comparaison de taux de croissance étudiée. De plus, d'autres paramètres intrinsèques et extrinsèques des ARNm (différents de leur concentration) ont pu être identifiés comme des déterminants de leur stabilité. C'est le cas par exemple du CAI, de la longueur de l'ORF, du pourcentage en GC de l'association 5'UTR et ORF, de l'appartenance à certaines catégories fonctionnelles ou encore de la complémentarité de séquence du RBS avec l'ARNr 16S. La mise en évidence de ces paramètres démontre que d'autres processus comme la traduction ou l'implication de structures secondaires peuvent intervenir dans la régulation de la stabilité des transcrits indépendamment de leur concentration. D'autre part, même si nos modèles de prédiction de la stabilité des ARNm ont pu expliquer plus de 50 % de la variabilité des  $t_{1/2}$ , des déterminants supplémentaires restent à être identifiés. Dans le futur, la liste exhaustive des ARNm régulés par des ARNnc permettra certainement de combler une grande partie de ce manque. La part de ces déterminants inconnus est d'ailleurs plus importante à taux de croissance élevé puisque nos modèles ont été moins prédictifs dans ces conditions (diminution aussi de l'effet majeur de la concentration des ARNm). Parallèlement, c'est lors de la transition entre les deux taux de croissance les plus forts que la part d'ARNm contrôlés par la dégradation est la plus importante. Pris tous ensemble, ces résultats suggèrent que lorsque les cellules se multiplient plus rapidement, l'impact des déterminants des  $t_{1/2}$  des ARNm indépendants de leur concentration augmente menant à des variations de stabilité chez un nombre non négligeable de transcrits, qui à leur tour entraînent des variations du niveau de ces messagers surpassant les effets de la transcription. Nos analyses ont par ailleurs montré que parmi les transcrits les plus contrôlés par la dégradation, les ARNm des enzymes du métabolisme carboné central sont particulièrement représentés démontrant que la stabilité des ARNm joue un rôle encore plus important dans la régulation du métabolisme de la bactérie à fort taux de croissance.

Les deux premières études de cette thèse ont donc mis en évidence deux grands types de régulation de la dégradation des ARNm : le premier, très majoritaire, où la

stabilité des messagers dépend principalement de leur concentration et donc des régulations de la transcription ; le second, où les variations de stabilité des transcrits, induites par d'autres mécanismes, entraînent des variations de concentration.

Si les deux premiers chapitres de cette thèse se sont intéressés aux régulations globales de la stabilité des ARNm en fonction de changements de conditions de croissance, le troisième étudie un système de régulation post-transcriptionnelle plus ciblé impliqué dans l'adaptation bactérienne : le système Csr. Si les mécanismes par lesquels la protéine CsrA exerce un contrôle sur la dégradation des ARNm en fonction des conditions de croissance ne sont plus à démontrer, l'étendue de l'effet de la protéine sur la stabilité des messagers à l'échelle du génome n'a jamais été étudiée. Ces travaux ont nécessité la construction d'un mutant de délétion partielle du gène de la protéine, nommé *csrA51*, sur le modèle d'une étude précédente en raison du fait que le gène *csrA* est essentiel dans nos conditions. En plus des caractéristiques phénotypiques attendues (liées à une diminution de l'activité de CsrA), le mutant obtenu présente un défaut de croissance important par rapport à la souche sauvage. Or, nous avons démontré que la dégradation des ARNm varie en fonction du taux de croissance (Chapitre I). En conséquence, le mutant *csrA51* a été cultivé à un même temps de doublement que la souche sauvage dans le but de s'affranchir de l'effet « taux de croissance » sur la stabilité des ARNm. La comparaison des  $t_{1/2}$  des messagers du mutant *csrA51* avec ceux de la souche *wt* a ainsi permis d'identifier les messagers dont la stabilité est sensible à la mutation. Les résultats ont montré que globalement, les ARNm sont davantage déstabilisés dans le mutant *csrA51*. Cependant, l'analyse combinée du transcriptome et du stabilome de la souche a mis en évidence que de nombreux messagers sont plus dégradés en raison d'une concentration plus importante dans la souche *csrA51*. C'est la première fois qu'une étude révèle un impact fortement transcriptionnel de la protéine CsrA, probablement via la régulation post-transcriptionnelle de régulateurs de la transcription (ex : *rpoD*). D'autre part, une fois ces cibles « indirectes » identifiées à l'aide des conclusions issues de la première partie de la thèse (Chapitre II), plus de 450 ARNm ont été révélés comme des cibles « directes » de CsrA. En effet, la variation de leur stabilité dans le mutant *csrA51* n'est pas liée à une variation de concentration. Nous avons démontré qu'une grande majorité d'entre eux interagit très probablement avec la protéine CsrA et est déstabilisée dans la souche *csrA51* démontrant qu'à l'échelle du génome, la protéine CsrA a un effet plutôt stabilisateur. Toutefois, dans nos conditions

de culture (milieu minimum M9 supplémenté en glucose à un taux de croissance  $\mu = 0.10 \text{ h}^{-1}$ ), une centaine d'ARNm parmi ces cibles « directes » présentent à la fois une variation de stabilité et de concentration. La modification de la concentration de messenger est un critère important pour affiner la liste des cibles de CsrA car ces variations mènent en général à des variations de concentration protéique qui participent à l'adaptation de la bactérie à son environnement. Cette liste de transcrits, dont la stabilité régulée par CsrA entraîne une variation de concentration, est considérée comme la liste des cibles majeures de CsrA. Néanmoins, nous suggérons que cette liste évolue en fonction des conditions de croissance pour permettre à la bactérie de mieux s'adapter à son nouvel environnement. En conséquence, il sera nécessaire de la compléter dans le futur. Si plusieurs ARNm cibles connus de CsrA appartenant au métabolisme carboné central ont été confirmés lors de cette étude, nos travaux révèlent que la protéine joue également un rôle dans les régulations de stabilité et de concentration de messagers de voies métaboliques particulières et de transports de sucres. Enfin, nous avons démontré que la protéine CsrD, connue pour adresser les petits ARNnc CsrB et CsrC à la dégradation, n'est pas capable de faire de même pour les ARNm. Cependant, nos résultats montrent pour la première fois que le mutant CsrD régule la transcription d'un grand nombre d'ARNm par un mécanisme encore inconnu mais indépendant de l'activité de CsrA. Nous suggérons même que sa surexpression dans le mutant *csrA51*, pourrait expliquer une partie de la réponse transcriptionnelle de ce mutant soulevant de nouvelles interrogations sur les mécanismes de ce système.

En conclusion, ces travaux de thèse ont répondu aux questions proposées en préambule. En effet, les deux premiers chapitres ont permis : (i) d'évaluer l'effet du taux de croissance sur la stabilité et la concentration des ARNm, (ii) de quantifier l'impact des régulations de la dégradation des messagers sur leur concentration, (iii) de comparer cet effet à celui des régulations transcriptionnelles et (iv) d'identifier et hiérarchiser les principaux déterminants du temps de demi-vie des ARNm en fonction du taux de croissance de la bactérie. A l'aide d'approches issues de la Biologie des Systèmes, ces travaux permettent de mieux appréhender le rôle et l'impact des régulations de la stabilité des ARNm dans la régulation de leur concentration en réponse à des variations de conditions de croissance (taux de croissance). La compréhension de ce mécanisme est particulièrement intéressante puisqu'elle permettra de démêler les réponses

cellulaires observées dans d'autres conditions environnementales induisant des variations de taux de croissance (mutants, carences nutritionnelles, etc...). Enfin, au cours du dernier chapitre de ces travaux, nous avons, à l'aide des conclusions précédentes, étudié un exemple de régulation post-transcriptionnelle en évaluant l'effet des protéines CsrA et CsrD sur la stabilité et la concentration des ARNm à l'échelle omique dans le but d'étendre nos connaissances du régulon du système Csr.

### **Perspectives**

En perspective, l'ensemble de ce travail a proposé de nouveaux concepts concernant les régulations de la stabilité des ARNm qui méritent d'être confirmés par des expériences supplémentaires.

Nos résultats suggèrent notamment que la concentration d'un ARNm est un déterminant très fort de sa stabilité. Cette notion déjà évoquée chez *E. coli* (Bernstein et al., 2002) et chez *M. tuberculosis* (Rustad et al., 2013) doit néanmoins être confirmée expérimentalement. Pour cela, des mesures de stabilité d'un ARNm modèle sous le contrôle d'un promoteur inductible permettraient de vérifier si l'augmentation du niveau d'induction influence négativement la valeur du  $t_{1/2}$  du transcrit. D'autre part, le couplage de ce type d'expériences avec des variations de séquences au niveau du messenger modèle permettrait de confirmer les déterminants de la stabilité identifiés par notre approche statistique. Par exemple, il est possible d'envisager de valider les effets des motifs putatifs identifiés de façon bioinformatique, de placer le gène sous différents 5'UTR, de changer le pourcentage de GC de certaines régions du transcrit, d'ajouter des structures secondaires à des positions différentes... Néanmoins, il sera important de rester dans la même gamme de concentrations que les conditions témoins (en jouant sur le niveau d'induction) pour comparer les  $t_{1/2}$ . Étant donné que la concentration d'une protéine est déterminée en grande partie par le niveau de son ARNm chez *E. coli*, le contrôle artificiel de la stabilité de ce dernier représente un levier innovant (car différent d'un contrôle via la transcription) et particulièrement intéressant pour des applications biotechnologiques. En effet, l'identification de déterminants spécifiques de la stabilité des messagers entraînant une augmentation de leur concentration pourrait permettre de favoriser la production de protéines d'intérêt par cette bactérie.

Malheureusement, la réponse à la question : « qu'est-ce qui détermine qu'un ARNm est plus stable qu'un autre ? » n'est toujours pas complète. En effet, nos données suggèrent que d'autres paramètres encore inconnus et ayant un rôle non négligeable dans la dégradation des ARNm à l'échelle omique restent à être identifiés. Notamment, il est encore très difficile aujourd'hui de prédire les sites de clivage des messagers par les RNases à l'échelle omique même si des travaux sur certains ARN modèles ont donné certaines pistes (McDowall et al., 1994, 1995). Pour répondre à cette question, le développement des techniques de marquage des extrémités des transcrits clivés et de la technologie de séquençage RNA-Seq mènera dans un futur proche à l'identification de



tous les sites de coupures des ARNm. À partir de ces résultats omiques, une analyse bioinformatique devrait permettre d'identifier des séquences communes de coupure ou de prédire l'environnement favorable au clivage d'un ARNm par les endonucléases. D'autre part, la détermination expérimentale des 3'UTR faciliterait également l'intégration de ces données pour évaluer leur effet potentiel sur la stabilité des ARNm. Plus globalement, la mesure de la stabilité des ARNm à l'échelle omique est très intéressante pour améliorer notre compréhension des mécanismes moléculaires impliqués dans la dégradation des messagers à l'échelle du génome. Bernstein et ses collaborateurs ont par exemple utilisé ce type d'approche pour étudier le rôle du dégradosome (Bernstein et al., 2004). Toutefois, nous avons vu que l'étude transcriptomique est indispensable pour comprendre et analyser les données de stabilité des transcrits. En conséquence, l'analyse simultanée du transcriptome et de la stabilité des ARNm chez des souches délétées ou mutées pour certains gènes impliqués dans la dégradation des messagers (tels que les gènes de la RNase R, RNase G ou la ré-analyse des données de Bernstein en ajoutant le transcriptome) permettrait par exemple, de mesurer leur effet sur l'ensemble des transcrits et/ou sur leur mode d'action.

L'application de cette stratégie d'étude nous a permis d'identifier un certain nombre de nouvelles cibles potentielles du régulateur post-transcriptionnel CsrA. Si des indications d'une interaction probable de CsrA avec ces messagers ont été apportées, des analyses supplémentaires de type gel-retard permettraient de confirmer et de valider définitivement ces nouvelles cibles. A la suite de cela, les mécanismes de régulation positive ou négative par la protéine CsrA devraient être également étudiés pour déterminer s'ils correspondent à ceux déjà décrits dans la littérature ou si l'action de la protéine CsrA implique de nouveaux mécanismes.

Jusqu'à aujourd'hui, deux mécanismes principaux de régulations post-transcriptionnelles de la protéine CsrA ont été décrits dans la littérature. Lors de la répression de l'expression d'un transcrit, la protéine CsrA inhibe généralement l'initiation de la traduction qui peut mener à la dégradation du transcrit. Lors de l'activation de l'expression d'un ARNm, la protéine stabiliserait ce dernier en inhibant sa dégradation 5'-dépendante par la RNase E. Au cours de nos travaux, en se concentrant sur l'effet de CsrA sur la dégradation des ARNm, nous avons mis en évidence que la

protéine aurait un rôle davantage activateur que répresseur. Toutefois, l'exemple de l'ARNm du gène *hfq* montre que l'expression d'un messager peut être réprimée sans déstabilisation de ce dernier (Baker et al., 2007). En conséquence, si l'analyse de la stabilité des transcrits est particulièrement indiquée et justifiée pour l'identification des cibles activées par CsrA, ceci n'est peut-être pas le cas pour l'ensemble des cibles qu'elle réprime. De ce fait, la comparaison des protéomes des souches *csrA51* et *wt* devrait permettre dans un premier temps de confirmer certaines de nos cibles et d'identifier les protéines dont l'expression est réprimée par CsrA sans affecter la stabilité de leur ARNm. Cependant, les résultats protéomiques peuvent masquer des régulations de la traduction à cause de régulations post-traductionnelles. Il serait donc plus intéressant encore de comparer dans un deuxième temps, les « traductomes » de ces deux souches. Ces données omiques donnent en effet accès à l'état de traduction de chaque ARNm de la cellule, en déterminant leur pourcentage de molécules engagées en traduction ainsi que leur densité en ribosomes. En conséquence, ce type d'expérience nous permettrait de compléter le régulon de CsrA. Dans un contexte plus général, l'association des résultats de traductome à ceux du stabilome donneraient des informations sur l'interaction entre stabilité et traduction des messagers. Par exemple, une corrélation simple entre les  $t_{1/2}$  et la densité en ribosomes caractériserait à l'échelle omique un effet protecteur ou délétère des ribosomes dans une condition donnée.

D'autre part, s'il a été démontré que la protéine CsrD pouvait réguler l'expression du facteur CsgD (Sommerfeldt et al., 2009), nos travaux suggèrent que cette protéine régule l'expression d'un nombre très important d'ARNm indépendamment de l'activité de CsrA. Cependant, contrairement à son action sur les petits ARNnc CsrB et CsrC, CsrD n'affecte pas la stabilité des messagers : son contrôle est donc transcriptionnel. Il serait donc très intéressant de déterminer si CsrD peut jouer le rôle d'un facteur de transcription ou s'il agit indirectement sur la transcription via la régulation d'ARNnc. Cette seconde hypothèse semble la plus probable puisque Romeo et ses collaborateurs ont suggéré récemment que CsrD pouvait cibler d'autres petits ARN que CsrB et CsrC (Romeo et al., 2012). Si cela se vérifiait, il serait alors nécessaire d'établir la liste des ARNnc dont la stabilité et/ou l'expression sont régulées par la protéine CsrD à l'aide d'approches de RNA-Seq par exemple.

Il est important de souligner que les travaux présentés ici ont évalué l'effet de la dégradation des ARNm au sein de cellules cultivées en continu lors de chémostats. Ce mode de culture offre plusieurs avantages : le taux de croissance du micro-organisme peut être facilement modulé à partir d'un même milieu de culture et lors de l'état d'équilibre, la physiologie des cellules est parfaitement adaptée aux conditions environnementales constantes. Si la régulation du taux de croissance reste un atout incontestable de cette étude, le rôle et l'impact global de la dégradation des ARNm sur leur concentration lors de l'adaptation dynamique (ou transition) entre deux conditions environnementales différentes restent à être déterminés. En effet, les régulations post-transcriptionnelles proposent généralement des mécanismes de contrôle plus rapides que les régulations transcriptionnelles (Jonas et al., 2008; Adamson & Lim, 2013) et faciliteraient à court terme l'adaptation du micro-organisme à son nouvel environnement. Nous supposons donc que l'importance de la dégradation des ARNm pourrait avoir été sous-estimée dans nos travaux. Toutefois, l'évaluation des régulations spécifiques de la stabilité des ARNm lors de transitions dynamiques sera extrêmement complexe pour plusieurs raisons. Tout d'abord, une telle analyse représente un challenge technique en raison de la rapidité d'évolution des cellules et de la nécessité d'avoir un grand nombre d'échantillons pour couvrir la dynamique avec un outil aussi complexe que le stabilome. De plus, ces transitions entraînent très souvent des changements de taux de croissance qui, nous l'avons vu dans ces travaux, modifient la stabilité des messagers. En conséquence, l'analyse de ces données nécessitera de se tourner vers des méthodes d'intégration et de modélisation des régulations pour pouvoir démêler les différentes réponses observées. Par exemple, ce type d'approche serait intéressant pour étudier l'impact du régulateur post-transcriptionnel CsrA lors d'une transition métabolique.

Enfin, plus largement, nous avons vu lors de l'étude bibliographique que de nombreux autres niveaux de régulation de l'expression des gènes existent. En conséquence, la production des différentes données omiques à partir de cultures bactériennes identiques représente des sources d'informations de grande qualité pour comprendre les mécanismes globaux de régulation des gènes. En effet, l'intégration du stabilome avec d'autres niveaux de données omique tels que le transcriptome, le traductome (qu'il reste encore à réaliser), le protéome et la stabilité des protéines (en cours de réalisation en collaboration avec Raivo Vilu de l'Université de Tallin en Estonie)

permettra de répondre à de nombreuses questions qui se posent encore aujourd'hui à propos de la régulation de l'expression des gènes à savoir : les différents niveaux de régulation ont-ils des rôles équivalents ? Sinon, comment se fait la répartition du contrôle ? Est-ce que certains niveaux de régulation sont davantage influents dans certaines conditions de croissance ?



# **REFERENCES**

# **BIBLIOGRAPHIQUES**

**Étude bibliographique**  
**Conclusion et perspectives**



- Adamson, D.N., Lim, H.N. (2013). Rapid and robust signaling in the CsrA cascade via RNA–protein interactions and feedback regulation. *Proc. Natl. Acad. Sci.* *110*, 13120–13125.
- Ades, S.E. (2008). Regulation by destruction: design of the  $\sigma$ E envelope stress response. *Curr. Opin. Microbiol.* *11*, 535–540.
- Albertson, N.H., Nyström, T. (1994). Effects of starvation for exogenous carbon on functional mRNA stability and rate of peptide chain elongation in *Escherichia coli*. *FEMS Microbiol. Lett.* *117*, 181–187.
- Altuvia, S., Zhang, A., Argaman, L., Tiwari, A., Storz, G. (1998). The *Escherichia coli* OxyS regulatory RNA represses *fhlA* translation by blocking ribosome binding. *EMBO J.* *17*, 6069–6075.
- Arnold, T.E., Yu, J., Belasco, J.G. (1998). mRNA stabilization by the *ompA* 5' untranslated region: two protective elements hinder distinct pathways for mRNA degradation. *RNA* *4*, 319–330.
- Azam, T.A., Ishihama, A. (1999). Twelve Species of the Nucleoid-associated Protein from *Escherichia coli*. *J. Biol. Chem.* *274*, 33105–33113.
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., Mori, H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* *2*, 2006.0008.
- Babitzke, P., Kushner, S.R. (1991). The Ams (altered mRNA stability) protein and ribonuclease E are encoded by the same structural gene of *Escherichia coli*. *Proc. Natl. Acad. Sci.* *88*, 1–5.
- Babitzke, P., Romeo, T. (2007). CsrB sRNA family: sequestration of RNA-binding regulatory proteins. *Curr. Opin. Microbiol.* *10*, 156–163.
- Babitzke, P., Granger, L., Olszewski, J., Kushner, S.R. (1993). Analysis of mRNA decay and rRNA processing in *Escherichia coli* multiple mutants carrying a deletion in RNase III. *J. Bacteriol.* *175*, 229–239.
- Bachmann, B.J. (1996). Derivation and genotypes of some mutant derivatives of *Escherichia coli* K-12. In *Escherichia Coli and Salmonella Typhimurium: Cellular and Molecular Biology*, 2nd Edition, (Washington: ASM Press), pp. 2460–2488.
- Baev, M.V., Baev, D., Radek, A.J., Campbell, J.W. (2006). Growth of *Escherichia coli* MG1655 on LB medium: monitoring utilization of sugars, alcohols, and organic acids with transcriptional microarrays. *Appl. Microbiol. Biotechnol.* *71*, 310–316.
- Baker, C.S., Morozov, I., Suzuki, K., Romeo, T., Babitzke, P. (2002). CsrA regulates glycogen biosynthesis by preventing translation of *glgC* in *Escherichia coli*. *Mol. Microbiol.* *44*, 1599–1610.



- Baker, C.S., Eöry, L.A., Yakhnin, H., Mercante, J., Romeo, T., Babitzke, P. (2007). CsrA Inhibits Translation Initiation of *Escherichia coli* hfq by Binding to a Single Site Overlapping the Shine-Dalgarno Sequence. *J. Bacteriol.* *189*, 5472–5481.
- Battesti, A., Majdalani, N., Gottesman, S. (2011). The RpoS-Mediated General Stress Response in *Escherichia coli*. *Annu. Rev. Microbiol.* *65*, 189–213.
- Bernstein, J.A., Khodursky, A.B., Lin, P.-H., Lin-Chao, S., Cohen, S.N. (2002). Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl. Acad. Sci. U. S. A.* *99*, 9697–9702.
- Bernstein, J.A., Lin, P.-H., Cohen, S.N., Lin-Chao, S. (2004). Global analysis of *Escherichia coli* RNA degradosome function using DNA microarrays. *Proc. Natl. Acad. Sci. U. S. A.* *101*, 2758–2763.
- Berthoumieux, S., de Jong, H., Baptist, G., Pinel, C., Ranquet, C., Ropers, D., Geiselman, J. (2013). Shared control of gene expression in bacteria by transcription factors and global physiology of the cell. *Mol. Syst. Biol.* *9*, 634.
- Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., et al. (1997). The Complete Genome Sequence of *Escherichia coli* K-12. *Science* *277*, 1453–1462.
- Blum, E., Carpousis, A.J., Higgins, C.F. (1999). Polyadenylation Promotes Degradation of 3'-Structured RNA by the *Escherichia coli* mRNA Degradosome in Vitro. *J. Biol. Chem.* *274*, 4009–4016.
- Bouvier, M., Carpousis, A.J. (2011). A tale of two mRNA degradation pathways mediated by RNase E. *Mol. Microbiol.* *82*, 1305–1310.
- Bouvier, M., Sharma, C.M., Mika, F., Nierhaus, K.H., Vogel, J. (2008). Small RNA binding to 5' mRNA coding region inhibits translational initiation. *Mol. Cell* *32*, 827–837.
- Brandi, A., Pietroni, P., Gualerzi, C.O., Pon, C.L. (1996). Post-transcriptional regulation of CspA expression in *Escherichia coli*. *Mol. Microbiol.* *19*, 231–240.
- Brantl, S. (2012). Bacterial type I toxin-antitoxin systems. *RNA Biol.* *9*, 1488–1490.
- Braun, F., Le Derout, J., Regnier, P. (1998). Ribosomes inhibit an RNase E cleavage which induces the decay of the rpsO mRNA of *Escherichia coli*. *EMBO J.* *17*, 4790–4797.
- Bremer, H., Dennis, P.P. (1996). Modulation of Chemical Composition and Other Parameters of the Cell by Growth Rate. In *Escherichia Coli and Salmonella Typhimurium: Cellular and Molecular Biology*, 2nd Edition, F.C. Neidhardt, ed. (Washington: ASM Press), pp. 1553–1569.
- Bremer, H., Dennis, P., Ehrenberg, M. (2003). Free RNA polymerase and modeling global transcription in *Escherichia coli*. *Biochimie* *85*, 597–609.
- Brown, T.A. (2002). *Genomes*, 2nd edition (Oxford: Wiley-Liss).

- Browning, D.F., Busby, S.J.W. (2004). The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol.* 2, 57–65.
- Brzozowska, I., Zielenkiewicz, U. (2013). Regulation of toxin–antitoxin systems by proteolysis. *Plasmid* 70, 33–41.
- Burgess, R.R., Travers, A.A., Dunn, J.J., Bautz, E.K. (1969). Factor stimulating transcription by RNA polymerase. *Nature* 221, 43–46.
- Cabrera, J.E., Jin, D.J. (2006). Active Transcription of rRNA Operons Is a Driving Force for the Distribution of RNA Polymerase in Bacteria: Effect of Extrachromosomal Copies of *rrnB* on the In Vivo Localization of RNA Polymerase. *J. Bacteriol.* 188, 4007–4014.
- Callaghan, A.J., Aurikko, J.P., Ilag, L.L., Günter Grossmann, J., Chandran, V., Kühnel, K., Poljak, L., Carpousis, A.J., Robinson, C.V., Symmons, M.F., et al. (2004). Studies of the RNA Degradosome-organizing Domain of the Escherichia coli Ribonuclease RNase E. *J. Mol. Biol.* 340, 965–979.
- Callaghan, A.J., Marcaida, M.J., Stead, J.A., McDowall, K.J., Scott, W.G., Luisi, B.F. (2005). Structure of Escherichia coli RNase E catalytic domain and implications for RNA turnover. *Nature* 437, 1187–1191.
- Carabetta, V.J., Silhavy, T.J., Cristea, I.M. (2010). The Response Regulator SprE (RssB) Is Required for Maintaining Poly(A) Polymerase I-Degradosome Association during Stationary Phase. *J. Bacteriol.* 192, 3713–3721.
- Carlos, C., Pires, M.M., Stoppe, N.C., Hachich, E.M., Sato, M.I., Gomes, T.A., Amaral, L.A., Ottoboni, L.M. (2010). Escherichia coli phylogenetic group determination and its application in the identification of the major animal source of fecal contamination. *BMC Microbiol.* 10, 161.
- Caron, M.-P., Bastet, L., Lussier, A., Simoneau-Roy, M., Masse, E., Lafontaine, D.A. (2012). Dual-acting riboswitch control of translation initiation and mRNA decay. *Proc. Natl. Acad. Sci. U. S. A.* 109, E3444–E3453.
- Carpousis, A.J. (2002). The Escherichia coli RNA degradosome: structure, function and relationship in other ribonucleolytic multienzyme complexes. *Biochem. Soc. Trans.* 30, 150–155.
- Carpousis, A.J. (2007). The RNA degradosome of Escherichia coli: an mRNA-degrading machine assembled on RNase E. *Annu. Rev. Microbiol.* 61, 71–87.
- Carpousis, A.J., Van Houwe, G., Ehretsmann, C., Krisch, H.M. (1994). Copurification of E. coli RNAase E and PNPase: Evidence for a specific association between two enzymes important in RNA processing and degradation. *Cell* 76, 889–900.
- Carpousis, A.J., Khemici, V., Poljak, L. (2008). Assaying DEAD-box RNA Helicases and Their Role in mRNA Degradation in Escherichia coli. In *Methods in Enzymology*, Lynne E. Maquat and Cecilia M. Arraiano, ed. (Academic Press), pp. 183–197.

- Carpousis, A.J., Luisi, B.F., McDowall, K.J. (2009). Endonucleolytic Initiation of mRNA Decay in *Escherichia coli*. *Prog. Mol. Biol. Transl. Sci. Volume 85*, 91–135.
- Cashel, M., Gentry, D.R., Hernandez, V.J., Vinella, D. (1996). The Stringent Response. In *Escherichia Coli and Salmonella Typhimurium: Cellular and Molecular Biology*, 2nd Edition, (Washington: ASM Press), pp. 1458–1496.
- Castaño-Cerezo, S., Bernal, V., Blanco-Catalá, J., Iborra, J.L., Cánovas, M. (2011). cAMP-CRP co-ordinates the expression of the protein acetylation pathway with central metabolism in *Escherichia coli*. *Mol. Microbiol.* *82*, 1110–1128.
- Chandu, D., Nandi, D. (2004). Comparative genomics and functional roles of the ATP-dependent proteases Lon and Clp during cytosolic protein degradation. *Res. Microbiol.* *155*, 710–719.
- Chang, D.-E., Smalley, D.J., Tucker, D.L., Leatham, M.P., Norris, W.E., Stevenson, S.J., Anderson, A.B., Grissom, J.E., Laux, D.C., Cohen, P.S., et al. (2004). Carbon nutrition of *Escherichia coli* in the mouse intestine. *Proc. Natl. Acad. Sci. U. S. A.* *101*, 7427–7432.
- Lin-Chao, S., Wei, C.L., Lin, Y.T. (1999). RNase E is required for the maturation of *ssrA* RNA and normal *ssrA* RNA peptide-tagging activity. *Proc. Natl. Acad. Sci. U. S. A.* *96*, 12406–12411.
- Chaudhuri, R.R., Henderson, I.R. (2012). The evolution of the *Escherichia coli* phylogeny. *Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis.* *12*, 214–226.
- Chavez, R.G., Alvarez, A.F., Romeo, T., Georgellis, D. (2010). The physiological stimulus for the BarA sensor kinase. *J. Bacteriol.* *192*, 2009–2012.
- Chen, H., Bjercknes, M., Kumar, R., Jay, E. (1994). Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of *Escherichia coli* mRNAs. *Nucleic Acids Res.* *22*, 4953–4957.
- Cheng, Z.-F., Deutscher, M.P. (2005). An Important Role for RNase R in mRNA Decay. *Mol. Cell* *17*, 313–318.
- Coburn, G.A., Miao, X., Briant, D.J., Mackie, G.A. (1999). Reconstitution of a minimal RNA degradosome demonstrates functional coordination between a 3' exonuclease and a DEAD-box RNA helicase. *Genes Dev.* *13*, 2594–2603.
- Cozzone, A.J., El-Mansi, M. (2005). Control of isocitrate dehydrogenase catalytic activity by protein phosphorylation in *Escherichia coli*. *J. Mol. Microbiol. Biotechnol.* *9*, 132–146.
- Croxen, M.A., Finlay, B.B. (2010). Molecular mechanisms of *Escherichia coli* pathogenicity. *Nat. Rev. Microbiol.* *8*, 26–38.
- Dame, R.T. (2005). The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. *Mol. Microbiol.* *56*, 858–870.
- Deana, A., Belasco, J.G. (2005). Lost in translation: the influence of ribosomes on bacterial mRNA decay. *Genes Dev.* *19*, 2526–2533.

- Deana, A., Celesnik, H., Belasco, J.G. (2008). The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal. *Nature* 451, 355–358.
- Deutscher, M.P., Reuven, N.B. (1991). Enzymatic basis for hydrolytic versus phosphorolytic mRNA degradation in *Escherichia coli* and *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U. S. A.* 88, 3277–3280.
- Dillon, S.C., Dorman, C.J. (2010). Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat. Rev. Microbiol.* 8, 185–195.
- Donovan, W.P., Kushner, S.R. (1986). Polynucleotide phosphorylase and ribonuclease II are required for cell viability and mRNA turnover in *Escherichia coli* K-12. *Proc. Natl. Acad. Sci. U. S. A.* 83, 120–124.
- Dressaire, C., Picard, F., Redon, E., Loubiere, P., Queinnec, I., Girbal, L., Coccagn-Bousquet, M. (2013). Role of mRNA Stability during Bacterial Adaptation. *PLoS ONE* 8, e59059.
- Dreyfus, M. (2009). Killer and Protective Ribosomes. In *Progress in Molecular Biology and Translational Science*, Ciaran Condon, ed. (Academic Press), pp. 423–466.
- Dubey, A.K., Baker, C.S., Suzuki, K., Jones, A.D., Pandit, P., Romeo, T., Babitzke, P. (2003). CsrA Regulates Translation of the *Escherichia coli* Carbon Starvation Gene, *cstA*, by Blocking Ribosome Access to the *cstA* Transcript. *J. Bacteriol.* 185, 4450–4460.
- Dubey, A.K., Baker, C.S., Romeo, T., Babitzke, P. (2005). RNA sequence and secondary structure participate in high-affinity CsrA-RNA interaction. *RNA* 11, 1579–1587.
- Durfee, T., Hansen, A.-M., Zhi, H., Blattner, F.R., Jin, D.J. (2008). Transcription Profiling of the Stringent Response in *Escherichia coli*. *J. Bacteriol.* 190, 1084–1096.
- Duss, O., Michel, E., Konté, N.D. dit, Schubert, M., Allain, F.H.-T. (2014). Molecular basis for the wide range of affinity found in Csr/Rsm protein–RNA recognition. *Nucleic Acids Res.* gku141.
- Eckburg, P.B., Bik, E.M., Bernstein, C.N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S.R., Nelson, K.E., Relman, D.A. (2005). Diversity of the Human Intestinal Microbial Flora. *Science* 308, 1635–1638.
- Edwards, A.N., Patterson-Fortin, L.M., Vakulskas, C.A., Mercante, J.W., Potrykus, K., Vinella, D., Camacho, M.I., Fields, J.A., Thompson, S.A., Georgellis, D., et al. (2011). Circuitry Linking the Csr and Stringent Response Global Regulatory Systems. *Mol. Microbiol.* no–no.
- Ehrenberg, M., Bremer, H., Dennis, P.P. (2013). Medium-dependent control of the bacterial growth rate. *Biochimie* 95, 643–658.
- Emory, S.A., Bouvet, P., Belasco, J.G. (1992). A 5'-terminal stem-loop structure can stabilize mRNA in *Escherichia coli*. *Genes Dev.* 6, 135–148.
- Feng, L., Niu, D.-K. (2007). Relationship between mRNA stability and length: an old question with a new twist. *Biochem. Genet.* 45, 131–137.

- Freyre-Gonzalez, J.A., Trevino-Quintanilla, L.G. (2010). Analyzing Regulatory Networks in Bacteria. *Nat. Educ.* 3, 24.
- Gao, J., Lee, K., Zhao, M., Qiu, J., Zhan, X., Saxena, A., Moore, C.J., Cohen, S.N., Georgiou, G. (2006). Differential modulation of *E. coli* mRNA abundance by inhibitory proteins that alter the composition of the degradosome. *Mol. Microbiol.* 61, 394–406.
- Georgellis, D., Barlow, T., Arvidson, S., von Gabain, A. (1993). Retarded RNA turnover in *Escherichia coli*: a means of maintaining gene expression during anaerobiosis. *Mol. Microbiol.* 9, 375–381.
- Gerosa, L., Kochanowski, K., Heinemann, M., Sauer, U. (2013). Dissecting specific and global transcriptional regulation of bacterial gene expression. *Mol. Syst. Biol.* 9, 658.
- Ghosh, S., Deutscher, M.P. (1999). Oligoribonuclease is an essential component of the mRNA decay pathway. *Proc. Natl. Acad. Sci. U. S. A.* 96, 4372–4377.
- Ghosh, T., Bose, D., Zhang, X. (2010). Mechanisms for activating bacterial RNA polymerase. *FEMS Microbiol. Rev.* 34, 611–627.
- Goeders, N., Van Melderen, L. (2014). Toxin-Antitoxin Systems as Multilevel Interaction Systems. *Toxins* 6, 304–324.
- Goldenberg, D., Azar, I., Oppenheim, A.B. (1996). Differential mRNA stability of the *cspA* gene in the cold-shock response of *Escherichia coli*. *Mol. Microbiol.* 19, 241–248.
- Górna, M.W., Carpousis, A.J., Luisi, B.F. (2012). From conformational chaos to robust regulation: the structure and function of the multi-enzyme RNA degradosome. *Q. Rev. Biophys.* 45, 105–145.
- Gottesman, S. (2003). Proteolysis in Bacterial Regulatory Circuits. *Annu. Rev. Cell Dev. Biol.* 19, 565–587.
- Gottesman, S. (2004). The small RNA regulators of *Escherichia coli*: Roles and Mechanisms. *Annu. Rev. Microbiol.* 58, 303–328.
- Gruber, T.M., Gross, C.A. (2003). Multiple Sigma Subunits and the Partitioning of Bacterial Transcription Space. *Annu. Rev. Microbiol.* 57, 441–466.
- Gualerzi, C.O., Maria Giuliadori, A., Pon, C.L. (2003). Transcriptional and Post-transcriptional Control of Cold-shock Genes. *J. Mol. Biol.* 331, 527–539.
- Guarneros, G., Portier, C. (1991). Different specificities of ribonuclease II and polynucleotide phosphorylase in 3'mRNA decay. *Biochimie* 73, 543–549.
- Gutierrez, P., Li, Y., Osborne, M.J., Pomerantseva, E., Liu, Q., Gehring, K. (2005). Solution Structure of the Carbon Storage Regulator Protein CsrA from *Escherichia coli*. *J Bacteriol* 187, 3496–3501.

- Hambraeus, G., von Wachenfeldt, C., Hederstedt, L. (2003). Genome-wide survey of mRNA half-lives in *Bacillus subtilis* identifies extremely stable mRNAs. *Mol. Genet. Genomics* **269**, 706–714.
- Hartl, F.U., Hayer-Hartl, M. (2009). Converging concepts of protein folding in vitro and in vivo. *Nat. Struct. Mol. Biol.* **16**, 574–581.
- Hayes, C.S., Sauer, R.T. (2003). Cleavage of the A Site mRNA Codon during Ribosome Pausing Provides a Mechanism for Translational Quality Control. *Mol. Cell* **12**, 903–911.
- Hengge, R. (2009a). Proteolysis of  $\sigma^S$  (RpoS) and the general stress response in *Escherichia coli*. *Res. Microbiol.* **160**, 667–676.
- Hengge, R. (2009b). Principles of c-di-GMP signalling in bacteria. *Nat. Rev. Microbiol.* **7**, 263–273.
- Hinkle, D.C., Chamberlin, M.J. (1972). Studies of the binding of *Escherichia coli* RNA polymerase to DNA. I. The role of sigma subunit in site selection. *J. Mol. Biol.* **70**, 157–185.
- Hoch, J.A. (2000). Two-component and phosphorelay signal transduction. *Curr. Opin. Microbiol.* **3**, 165–170.
- Hudault, S., Guignot, J., Servin, A. (2001). *Escherichia coli* strains colonising the gastrointestinal tract protect germfree mice against *Salmonella typhimurium* infection. *Gut* **49**, 47–55.
- Hundt, S., Zaigler, A., Lange, C., Soppa, J., Klug, G. (2007). Global Analysis of mRNA Decay in *Halobacterium salinarum* NRC-1 at Single-Gene Resolution Using DNA Microarrays. *J. Bacteriol.* **189**, 6936–6944.
- Iost, I., Dreyfus, M. (1995). The stability of *Escherichia coli* lacZ mRNA depends upon the simultaneity of its synthesis and translation. *EMBO J.* **14**, 3252–3261.
- Ishihama, A. (2000). Functional Modulation of *Escherichia coli* RNA Polymerase. *Annu. Rev. Microbiol.* **54**, 499–518.
- Jackson, D.W., Suzuki, K., Oakford, L., Simecka, J.W., Hart, M.E., Romeo, T. (2002). Biofilm Formation and Dispersal under the Influence of the Global Regulator CsrA of *Escherichia coli*. *J. Bacteriol.* **184**, 290–301.
- Jain, C., Belasco, J.G. (1995). RNase E autoregulates its synthesis by controlling the degradation rate of its own mRNA in *Escherichia coli*: unusual sensitivity of the rne transcript to RNase E activity. *Genes Dev.* **9**, 84–96.
- Jiang, X., Belasco, J.G. (2004). Catalytic activation of multimeric RNase E and RNase G by 5'-monophosphorylated RNA. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 9211–9216.
- Jiang, W., Hou, Y., Inouye, M. (1997). CspA, the major cold-shock protein of *Escherichia coli*, is an RNA chaperone. *J. Biol. Chem.* **272**, 196–202.

- Jin, D.J., Cabrera, J.E. (2006). Coupling the distribution of RNA polymerase to global gene regulation and the dynamic structure of the bacterial nucleoid in *Escherichia coli*. *J. Struct. Biol.* *156*, 284–291.
- Jin, D.J., Cagliero, C., Zhou, Y.N. (2012). Growth rate regulation in *Escherichia coli*. *FEMS Microbiol. Rev.* *36*, 269–287.
- Jishage, M., Iwata, A., Ueda, S., Ishihama, A. (1996). Regulation of RNA polymerase sigma subunit synthesis in *Escherichia coli*: intracellular levels of four species of sigma subunit under various growth conditions. *J. Bacteriol.* *178*, 5447–5451.
- Jonas, K., Melefors, Ö. (2009). The *Escherichia coli* CsrB and CsrC small RNAs are strongly induced during growth in nutrient-poor medium. *FEMS Microbiol. Lett.* *297*, 80–86.
- Jonas, K., Tomenius, H., Römling, U., Georgellis, D., Melefors, Ö. (2006). Identification of YhdA as a regulator of the *Escherichia coli* carbon storage regulation system. *FEMS Microbiol. Lett.* *264*, 232–237.
- Jonas, K., Edwards, A.N., Simm, R., Romeo, T., Römling, U., Melefors, Ö. (2008). The RNA binding protein CsrA controls cyclic di-GMP metabolism by directly regulating the expression of GGDEF proteins. *Mol. Microbiol.* *70*, 236–257.
- Jørgensen, M.G., Thomason, M.K., Havelund, J., Valentin-Hansen, P., Storz, G. (2013). Dual function of the McaS small RNA in controlling biofilm formation. *Genes Dev.*
- Kaberdin, V.R. (2003). Probing the substrate specificity of *Escherichia coli* RNase E using a novel oligonucleotide-based assay. *Nucleic Acids Res.* *31*, 4710–4716.
- Kaberdin, V.R., Bläsi, U. (2006). Translation initiation and the fate of bacterial mRNAs. *FEMS Microbiol. Rev.* *30*, 967–979.
- Kaberdin, V.R., Lin-Chao, S. (2009). Unraveling new roles for minor components of the *E. coli* RNA degradosome. *RNA Biol.* *6*, 402–405.
- Kaczanowska, M., Rydén-Aulin, M. (2007). Ribosome Biogenesis and the Translation Process in *Escherichia coli*. *Microbiol. Mol. Biol. Rev.* *71*, 477–494.
- Kaga, N., Umitsuki, G., Nagai, K., Wachi, M. (2002). RNase G-dependent degradation of the *eno* mRNA encoding a glycolysis enzyme enolase in *Escherichia coli*. *Biosci. Biotechnol. Biochem.* *66*, 2216–2220.
- Kaper, J.B., Nataro, J.P., Mobley, H.L. (2004). Pathogenic *Escherichia coli*. *Nat. Rev. Microbiol.* *2*, 123–140.
- Karp, P.D., Keseler, I.M., Shearer, A., Latendresse, M., Krummenacker, M., Paley, S.M., Paulsen, I., Collado-Vides, J., Gama-Castro, S., Peralta-Gil, M., et al. (2007). Multidimensional annotation of the *Escherichia coli* K-12 genome. *Nucleic Acids Res.* *35*, 7577–7590.

Keren, L., Zackay, O., Lotan-Pompan, M., Barenholz, U., Dekel, E., Sasson, V., Aidelberg, G., Bren, A., Zeevi, D., Weinberger, A., et al. (2013). Promoters maintain their relative activity levels under different growth conditions. *Mol. Syst. Biol.* *9*.

Keseler, I.M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., Gama-Castro, S., Bonavides-Martinez, C., Fulcher, C., Huerta, A.M., Kothari, A., Krummenacker, M., et al. (2013). EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res.* *41*, D605–D612.

Khemici, V., Poljak, L., Luisi, B.F., Carpousis, A.J. (2008). The RNase E of *Escherichia coli* is a membrane-binding protein. *Mol. Microbiol.* *70*, 799–813.

Klumpp, S., Hwa, T. (2008). Growth-rate-dependent partitioning of RNA polymerases in bacteria. *Proc. Natl. Acad. Sci.* *105*, 20245–20250.

Klumpp, S., Zhang, Z., Hwa, T. (2009). Growth Rate-Dependent Global Effects on Gene Expression in Bacteria. *Cell* *139*, 1366–1375.

Kristoffersen, S.M., Haase, C., Weil, M.R., Passalacqua, K.D., Niazi, F., Hutchison, S.K., Desany, B., Kolstø, A.-B., Tourasse, N.J., Read, T.D., et al. (2012). Global mRNA decay analysis at single nucleotide resolution reveals segmental and positional degradation patterns in a Gram-positive bacterium. *Genome Biol.* *13*, R30.

Ter Kuile, B.H., Westerhoff, H.V. (2001). Transcriptome meets metabolome: hierarchical and metabolic regulation of the glycolytic pathway. *FEBS Lett.* *500*, 169–171.

Kuzj, A.E.S., Medberry, P.S., Schottel, J.L. (1998). Stationary phase, amino acid limitation and recovery from stationary phase modulate the stability and translation of chloramphenicol acetyltransferase mRNA and total mRNA in *Escherichia coli*. *Microbiology* *144*, 739–750.

Lee, K., Bernstein, J.A., Cohen, S.N. (2002). RNase G complementation of *rne* null mutation identifies functional interrelationships with RNase E in *Escherichia coli*. *Mol. Microbiol.* *43*, 1445–1456.

Lemke, J.J., Durfee, T., Gourse, R.L. (2009). DksA and ppGpp Directly Regulate Transcription of the *Escherichia coli* Flagellar Cascade. *Mol. Microbiol.* *74*, 1368–1379.

Lenz, G., Doron-Faigenboim, A., Ron, E.Z., Tuller, T., Gophna, U. (2011). Sequence features of *E. coli* mRNAs affect their degradation. *PloS One* *6*, e28544.

Leroy, A., Vanzo, N.F., Sousa, S., Dreyfus, M., Carpousis, A.J. (2002). Function in *Escherichia coli* of the non-catalytic part of RNase E: role in the degradation of ribosome-free mRNA. *Mol. Microbiol.* *45*, 1231–1243.

Li, Z., Deutscher, M.P. (2002). RNase E plays an essential role in the maturation of *Escherichia coli* tRNA precursors. *RNA N. Y.* *8*, 97–109.

Li, Z., Pandit, S., Deutscher, M.P. (1999). RNase G (CafA protein) and RNase E are both required for the 5' maturation of 16S ribosomal RNA. *EMBO J.* *18*, 2878–2885.



- Lim, B., Ahn, S., Sim, M., Lee, K. (2014). RNase III Controls mltD mRNA Degradation in *Escherichia coli*. *Curr. Microbiol.* *68*, 518–523.
- Lin, H.-H., Lin, C.-H., Hwang, S.-M., Tseng, C.-P. (2012). High growth rate downregulates *fumA* mRNA transcription but is dramatically compensated by its mRNA stability in *Escherichia coli*. *Curr. Microbiol.* *64*, 412–417.
- Liu, M.Y., Romeo, T. (1997). The global regulator CsrA of *Escherichia coli* is a specific mRNA-binding protein. *J. Bacteriol.* *179*, 4639–4642.
- Liu, M.Y., Yang, H., Romeo, T. (1995). The product of the pleiotropic *Escherichia coli* gene *csrA* modulates glycogen biosynthesis via effects on mRNA stability. *J. Bacteriol.* *177*, 2663–2672.
- Liu, M.Y., Gui, G., Wei, B., Preston, J.F., Oakford, L., Yüksel, Ü., Giedroc, D.P., Romeo, T. (1997). The RNA Molecule CsrB Binds to the Global Regulatory Protein CsrA and Antagonizes Its Activity in *Escherichia coli*. *J. Biol. Chem.* *272*, 17502–17510.
- Mackie, G.A. (1998). Ribonuclease E is a 5'-end-dependent endonuclease. *Nature* *395*, 720–724.
- Mackie, G.A. (2013). RNase E: at the interface of bacterial RNA processing and decay. *Nat. Rev. Microbiol.* *11*, 45–57.
- Maeda, H., Fujita, N., Ishihama, A. (2000). Competition among seven *Escherichia coli*  $\sigma$  subunits: relative binding affinities to the core RNA polymerase. *Nucleic Acids Res.* *28*, 3497–3503.
- Majdalani, N., Cuning, C., Sledjeski, D., Elliott, T., Gottesman, S. (1998). DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription. *Proc. Natl. Acad. Sci. U. S. A.* *95*, 12462–12467.
- Martínez-Antonio, A. (2011). *Escherichia coli* transcriptional regulatory network. *Netw. Biol.* *1*, 21–33.
- Marzi, S., Romby, P. (2012). RNA mimicry, a decoy for regulatory proteins. *Mol. Microbiol.* *83*, 1–6.
- Massé, E., Escorcia, F.E., Gottesman, S. (2003). Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*. *Genes Dev.* *17*, 2374–2383.
- McDowall, K.J., Lin-Chao, S., Cohen, S.N. (1994). A+U content rather than a particular nucleotide order determines the specificity of RNase E cleavage. *J. Biol. Chem.* *269*, 10790–10796.
- McDowall, K.J., Kaberdin, V.R., Wu, S.-W., Cohen, S.N., Lin-Chao, S. (1995). Site-specific RNase E cleavage of oligonucleotides and inhibition by stem-loops. *Nature* *374*, 287–290.

- McLeod, S.M., Johnson, R.C. (2001). Control of transcription by nucleoid proteins. *Curr. Opin. Microbiol.* *4*, 152–159.
- Van Melderen, L., Aertsen, A. (2009). Regulation and quality control by Lon-dependent proteolysis. *Res. Microbiol.* *160*, 645–651.
- Mercante, J., Suzuki, K., Cheng, X., Babitzke, P., Romeo, T. (2006). Comprehensive alanine-scanning mutagenesis of *Escherichia coli* CsrA defines two subdomains of critical functional importance. *J. Biol. Chem.* *281*, 31832–31842.
- Mercante, J., Edwards, A.N., Dubey, A.K., Babitzke, P., Romeo, T. (2009). Molecular Geometry of CsrA (RsmA) Binding to RNA and Its Implications for Regulated Expression. *J. Mol. Biol.* *392*, 511–528.
- Milon, P., Tischenko, E., Tomsic, J., Caserta, E., Folkers, G., La Teana, A., Rodnina, M.V., Pon, C.L., Boelens, R., Gualerzi, C.O. (2006). The nucleotide-binding site of bacterial translation initiation factor 2 (IF2) as a metabolic sensor. *Proc. Natl. Acad. Sci. U. S. A.* *103*, 13962–13967.
- Mohanty, B.K., Kushner, S.R. (1999). Analysis of the function of *Escherichia coli* poly(A) polymerase I in RNA metabolism. *Mol. Microbiol.* *34*, 1094–1108.
- Mohanty, B.K., Kushner, S.R. (2003). Genomic analysis in *Escherichia coli* demonstrates differential roles for polynucleotide phosphorylase and RNase II in mRNA abundance and decay. *Mol. Microbiol.* *50*, 645–658.
- Morita, M., Kanemori, M., Yanagi, H., Yura, T. (1999a). Heat-Induced Synthesis of  $\sigma_{32}$  in *Escherichia coli*: Structural and Functional Dissection of rpoH mRNA Secondary Structure. *J. Bacteriol.* *181*, 401–410.
- Morita, M.T., Tanaka, Y., Kodama, T.S., Kyogoku, Y., Yanagi, H., Yura, T. (1999b). Translational induction of heat shock transcription factor  $\sigma_{32}$ : evidence for a built-in RNA thermosensor. *Genes Dev.* *13*, 655–665.
- Morita, T., Maki, K., Aiba, H. (2005). RNase E-based ribonucleoprotein complexes: mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. *Genes Dev.* *19*, 2176–2186.
- Newbury, S.F., Smith, N.H., Higgins, C.F. (1987). Differential mRNA stability controls relative gene expression within a polycistronic operon. *Cell* *51*, 1131–1143.
- Nicholson, A.W. (1999). Function, mechanism and regulation of bacterial ribonucleases. *FEMS Microbiol. Rev.* *23*, 371–390.
- Nilsson, P., Uhlin, B.E. (1991). Differential decay of a polycistronic *Escherichia coli* transcript is initiated by RNaseE-dependent endonucleolytic processing. *Mol. Microbiol.* *5*, 1791–1799.
- Nilsson, G., Belasco, J.G., Cohen, S.N., von Gabain, A. (1984). Growth-rate dependent regulation of mRNA stability in *Escherichia coli*. *Nature* *312*, 75–77.

- Nilsson, G., Belasco, J.G., Cohen, S.N., von Gabain, A. (1987). Effect of premature termination of translation on mRNA stability depends on the site of ribosome release. *Proc. Natl. Acad. Sci. U. S. A.* *84*, 4890–4894.
- Nou, X., Kadner, R.J. (2000). Adenosylcobalamin inhibits ribosome binding to *btuB* RNA. *Proc. Natl. Acad. Sci. U. S. A.* *97*, 7190–7195.
- Nudler, E., Gottesman, M.E. (2002). Transcription termination and anti-termination in *E. coli*. *Genes Cells* *7*, 755–768.
- Okada, Y., Wachi, M., Hirata, A., Suzuki, K., Nagai, K., Matsushashi, M. (1994). Cytoplasmic axial filaments in *Escherichia coli* cells: possible function in the mechanism of chromosome segregation and cell division. *J. Bacteriol.* *176*, 917–922.
- Opdyke, J.A., Fozo, E.M., Hemm, M.R., Storz, G. (2011). RNase III Participates in GadY-Dependent Cleavage of the *gadX-gadW* mRNA. *J. Mol. Biol.* *406*, 29–43.
- Österberg, S., Peso-Santos, T. del, Shingler, V. (2011). Regulation of Alternative Sigma Factor Use. *Annu. Rev. Microbiol.* *65*, 37–55.
- Pannuri, A., Yakhnin, H., Vakulskas, C.A., Edwards, A.N., Babitzke, P., Romeo, T. (2012). Translational repression of *NhaR*, a novel pathway for multi-tier regulation of biofilm circuitry by *CsrA*. *J. Bacteriol.* *194*, 79–89.
- Pato, M.L., von Meyenburg, K. (1970). Residual RNA Synthesis in *Escherichia coli* after Inhibition of Initiation of Transcription by Rifampicin. *Cold Spring Harb. Symp. Quant. Biol.* *35*, 497–504.
- Patterson-Fortin, L.M., Vakulskas, C.A., Yakhnin, H., Babitzke, P., Romeo, T. (2013). Dual Posttranscriptional Regulation via a Cofactor-Responsive mRNA Leader. *J. Mol. Biol.* *425*, 3662–3677.
- Paul, B.J., Barker, M.M., Ross, W., Schneider, D.A., Webb, C., Foster, J.W., Gourse, R.L. (2004). *DksA*: A Critical Component of the Transcription Initiation Machinery that Potentiates the Regulation of rRNA Promoters by ppGpp and the Initiating NTP. *Cell* *118*, 311–322.
- Pedersen, M., Nissen, S., Mitarai, N., Svenningsen, S.L., Sneppen, K., Pedersen, S. (2011). The Functional Half-Life of an mRNA Depends on the Ribosome Spacing in an Early Coding Region. *J. Mol. Biol.* *407*, 35–44.
- Pérez-Rueda, E., Collado-Vides, J. (2000). The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. *Nucleic Acids Res.* *28*, 1838–1847.
- Pérez-Rueda, E., Collado-Vides, J. (2001). Common history at the origin of the position-function correlation in transcriptional regulators in archaea and bacteria. *J. Mol. Evol.* *53*, 172–179.
- Pernestig, A.-K., Melefors, Ö., Georgellis, D. (2001). Identification of UvrY as the Cognate Response Regulator for the BarA Sensor Kinase in *Escherichia coli*. *J. Biol. Chem.* *276*, 225–231.

- Pfeiffer, V., Papenfort, K., Lucchini, S., Hinton, J.C.D., Vogel, J. (2009). Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. *Nat. Struct. Mol. Biol.* *16*, 840–846.
- Picard, F., Dressaire, C., Girbal, L., Coccagn-Bousquet, M. (2009). Examination of post-transcriptional regulations in prokaryotes by integrative biology. *C. R. Biol.* *332*, 958–973.
- Portier, C., Dondon, L., Grunberg-Manago, M., Régnier, P. (1987). The first step in the functional inactivation of the *Escherichia coli* polynucleotide phosphorylase messenger is a ribonuclease III processing at the 5' end. *EMBO J.* *6*, 2165–2170.
- Potrykus, K., Cashel, M. (2008). (p)ppGpp: Still Magical? *Annu. Rev. Microbiol.* *62*, 35–51.
- Potrykus, K., Murphy, H., Philippe, N., Cashel, M. (2011). ppGpp is the major source of growth rate control in *E. coli*. *Environ. Microbiol.* *13*, 563–575.
- Py, B., Causton, H., Mudd, E.A., Higgins, C.F. (1994). A protein complex mediating mRNA degradation in *Escherichia coli*. *Mol. Microbiol.* *14*, 717–729.
- Py, B., Higgins, C.F., Krisch, H.M., Carpousis, A.J. (1996). A DEAD-box RNA helicase in the *Escherichia coli* RNA degradosome. *Nature* *381*, 169–172.
- Rapaport, L.R., Mackie, G.A. (1994). Influence of translational efficiency on the stability of the mRNA for ribosomal protein S20 in *Escherichia coli*. *J. Bacteriol.* *176*, 992–998.
- Redon, E., Loubière, P., Coccagn-Bousquet, M. (2005). Role of mRNA stability during genome-wide adaptation of *Lactococcus lactis* to carbon starvation. *J. Biol. Chem.* *280*, 36380–36385.
- Régnier, P., Arraiano, C.M. (2000). Degradation of mRNA in bacteria: emergence of ubiquitous features. *BioEssays News Rev. Mol. Cell. Dev. Biol.* *22*, 235–244.
- Regulski, E.E., Moy, R.H., Weinberg, Z., Barrick, J.E., Yao, Z., Ruzzo, W.L., Breaker, R.R. (2008). A widespread riboswitch candidate that controls bacterial genes involved in molybdenum cofactor and tungsten cofactor metabolism. *Mol. Microbiol.* *68*, 918–932.
- Repoila, F., Gottesman, S. (2001). Signal transduction cascade for regulation of RpoS: temperature regulation of DsrA. *J. Bacteriol.* *183*, 4012–4023.
- Riley, M., Abe, T., Arnaud, M.B., Berlyn, M.K.B., Blattner, F.R., Chaudhuri, R.R., Glasner, J.D., Horiuchi, T., Keseler, I.M., Kosuge, T., et al. (2006). *Escherichia coli* K-12: a cooperatively developed annotation snapshot—2005. *Nucleic Acids Res.* *34*, 1–9.
- Romeo, T. (1998). Global regulation by the small RNA-binding protein CsrA and the non-coding RNA molecule CsrB. *Mol. Microbiol.* *29*, 1321–1330.
- Romeo, T., Black, J., Preiss, J. (1990). Genetic regulation of glycogen biosynthesis in *Escherichia coli*: In vivo effects of the catabolite repression and stringent response systems on *ingl* gene expression. *Curr. Microbiol.* *21*, 131–137.

- Romeo, T., Gong, M., Liu, M.Y., Brun-Zinkernagel, A.M. (1993). Identification and molecular characterization of *csrA*, a pleiotropic gene from *Escherichia coli* that affects glycogen biosynthesis, gluconeogenesis, cell size, and surface properties. *J. Bacteriol.* *175*, 4744–4755.
- Romeo, T., Vakulskas, C.A., Babitzke, P. (2012). Post-transcriptional regulation on a global scale: form and function of Csr/Rsm systems. *Environ. Microbiol.*
- Roy, M.K., Singh, B., Ray, B.K., Apirion, D. (1983). Maturation of 5-S rRNA: ribonuclease E cleavages and their dependence on precursor sequences. *Eur. J. Biochem. FEBS* *131*, 119–127.
- Russo, T.A., Johnson, J.R. (2003). Medical and economic impact of extraintestinal infections due to *Escherichia coli*: focus on an increasingly important endemic problem. *Microbes Infect.* *5*, 449–456.
- Rustad, T.R., Minch, K.J., Brabant, W., Winkler, J.K., Reiss, D.J., Baliga, N.S., Sherman, D.R. (2013). Global analysis of mRNA stability in *Mycobacterium tuberculosis*. *Nucleic Acids Res.* *41*, 509–517.
- Ryjenkov, D.A., Tarutina, M., Moskvina, O.V., Gomelsky, M. (2005). Cyclic Diguanylate Is a Ubiquitous Signaling Molecule in Bacteria: Insights into Biochemistry of the GGDEF Protein Domain. *J. Bacteriol.* *187*, 1792–1798.
- Sabnis, N.A., Yang, H., Romeo, T. (1995). Pleiotropic regulation of central carbohydrate metabolism in *Escherichia coli* via the gene *csrA*. *J. Biol. Chem.* *270*, 29096–29104.
- Salgado, H., Peralta-Gil, M., Gama-Castro, S., Santos-Zavaleta, A., Muniz-Rascado, L., Garcia-Sotelo, J.S., Weiss, V., Solano-Lira, H., Martinez-Flores, I., Medina-Rivera, A., et al. (2013). RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucleic Acids Res.* *41*, D203–D213.
- Salvail, H., Caron, M.-P., Bélanger, J., Massé, E. (2013). Antagonistic functions between the RNA chaperone Hfq and an sRNA regulate sensitivity to the antibiotic colicin. *EMBO J.* *32*, 2764–2778.
- Sauer, R.T., Bolon, D.N., Burton, B.M., Burton, R.E., Flynn, J.M., Grant, R.A., Hersch, G.L., Joshi, S.A., Kenniston, J.A., Levchenko, I., et al. (2004). Sculpting the Proteome with AAA+ Proteases and Disassembly Machines. *Cell* *119*, 9–18.
- Schmidt, A.J., Ryjenkov, D.A., Gomelsky, M. (2005). The Ubiquitous Protein Domain EAL Is a Cyclic Diguanylate-Specific Phosphodiesterase: Enzymatically Active and Inactive EAL Domains. *J. Bacteriol.* *187*, 4774–4781.
- Schubert, M., Lapouge, K., Duss, O., Oberstrass, F.C., Jelesarov, I., Haas, D., Allain, F.H.-T. (2007). Molecular basis of messenger RNA recognition by the specific bacterial repressing clamp RsmA/CsrA. *Nat. Struct. Mol. Biol.* *14*, 807–813.
- Selinger, D.W., Saxena, R.M., Cheung, K.J., Church, G.M., Rosenow, C. (2003). Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. *Genome Res.* *13*, 216–223.

- Serganov, A., Nudler, E. (2013). A Decade of Riboswitches. *Cell* 152, 17–24.
- Seyll, E., Van Melderen, L. (2013). The Ribonucleoprotein Csr Network. *Int. J. Mol. Sci.* 14, 22117–22131.
- Sharma, U.K., Chatterji, D. (2010). Transcriptional switching in *Escherichia coli* during stress and starvation by modulation of  $\sigma$ 70 activity. *FEMS Microbiol. Rev.* 34, 646–657.
- Shen, V., Imamoto, F., Schlessinger, D. (1982). RNase III cleavage of *Escherichia coli* beta-galactosidase and tryptophan operon mRNA. *J. Bacteriol.* 150, 1489–1494.
- Shimamoto, N., Kamigochi, T., Utiyama, H. (1986). Release of the sigma subunit of *Escherichia coli* DNA-dependent RNA polymerase depends mainly on time elapsed after the start of initiation, not on length of product RNA. *J. Biol. Chem.* 261, 11859–11865.
- Shine, J., Dalgarno, L. (1974). The 3'-Terminal Sequence of *Escherichia coli* 16S Ribosomal RNA: Complementarity to Nonsense Triplets and Ribosome Binding Sites. *Proc. Natl. Acad. Sci. U. S. A.* 71, 1342–1346.
- Singh, D., Chang, S.-J., Lin, P.-H., Averina, O.V., Kaberdin, V.R., Lin-Chao, S. (2009). Regulation of ribonuclease E activity by the L4 ribosomal protein of *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 106, 864–869.
- Smith, T.G., Hoover, T.R. (2009). Chapter 8 Deciphering Bacterial Flagellar Gene Regulatory Networks in the Genomic Era. In *Advances in Applied Microbiology*, S.S. and G.M.G. Allen I. Laskin, ed. (Academic Press), pp. 257–295.
- Sommerfeldt, N., Possling, A., Becker, G., Pesavento, C., Tschowri, N., Hengge, R. (2009). Gene expression patterns and differential input into curli fimbriae regulation of all GGDEF/EAL domain proteins in *Escherichia coli*. *Microbiology* 155, 1318–1331.
- Spickler, C., Mackie, G.A. (2000). Action of RNase II and Polynucleotide Phosphorylase against RNAs Containing Stem-Loops of Defined Structure. *J. Bacteriol.* 182, 2422–2427.
- Srivatsan, A., Wang, J.D. (2008). Control of bacterial transcription, translation and replication by (p)ppGpp. *Curr. Opin. Microbiol.* 11, 100–105.
- Sunohara, T., Jojima, K., Tagami, H., Inada, T., Aiba, H. (2004). Ribosome Stalling during Translation Elongation Induces Cleavage of mRNA Being Translated in *Escherichia coli*. *J. Biol. Chem.* 279, 15368–15375.
- Suzuki, K., Wang, X., Weilbacher, T., Pernestig, A.-K., Melefors, Ö., Georgellis, D., Babitzke, P., Romeo, T. (2002). Regulatory Circuitry of the CsrA/CsrB and BarA/UvrY Systems of *Escherichia coli*. *J. Bacteriol.* 184, 5130–5140.
- Suzuki, K., Babitzke, P., Kushner, S.R., Romeo, T. (2006). Identification of a novel regulatory protein (CsrD) that targets the global regulatory RNAs CsrB and CsrC for degradation by RNase E. *Genes Dev.* 20, 2605–2617.
- Takayama, K., Kjelleberg, S. (2000). The role of RNA stability during bacterial stress responses and starvation. *Environ. Microbiol.* 2, 355–365.

- Tenaillon, O., Skurnik, D., Picard, B., Denamur, E. (2010). The population genetics of commensal *Escherichia coli*. *Nat. Rev. Microbiol.* *8*, 207–217.
- Thomason, M.K., Fontaine, F., De Lay, N., Storz, G. (2012). A small RNA that regulates motility and biofilm formation in response to changes in nutrient availability in *Escherichia coli*. *Mol. Microbiol.* *84*, 17–35.
- Timmermans, J., Van Melderen, L. (2009). Conditional essentiality of the *csrA* gene in *Escherichia coli*. *J. Bacteriol.* *191*, 1722–1724.
- Timmermans, J., Van Melderen, L. (2010). Post-transcriptional global regulation by CsrA in bacteria. *Cell. Mol. Life Sci. CMLS* *67*, 2897–2908.
- Trevino-Quintanilla, L.G., Freyre-Gonzalez, J.A., Martinez-Flores, I. (2013). Anti-Sigma Factors in *E. coli*: Common Regulatory Mechanisms Controlling Sigma Factors Availability. *Curr. Genomics* *14*, 378–387.
- Tsai, Y.-C., Du, D., Dominguez-Malfavon, L., Dimastrogiovanni, D., Cross, J., Callaghan, A.J., Garcia-Mena, J., Luisi, B.F. (2012). Recognition of the 70S ribosome and polysome by the RNA degradosome in *Escherichia coli*. *Nucleic Acids Res.* *40*, 10417–10431.
- Umitsuki, G., Wachi, M., Takada, A., Hikichi, T., Nagai, K. (2001). Involvement of RNase G in in vivo mRNA metabolism in *Escherichia coli*. *Genes Cells* *6*, 403–410.
- Vanderpool, C.K., Gottesman, S. (2004). Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system. *Mol. Microbiol.* *54*, 1076–1089.
- Vitreschak, A.G., Rodionov, D.A., Mironov, A.A., Gelfand, M.S. (2004). Riboswitches: the oldest mechanism for the regulation of gene expression? *Trends Genet.* *20*, 44–50.
- Vogel, J., Luisi, B.F. (2011). Hfq and its constellation of RNA. *Nat Rev Micro* *9*, 578–589.
- Volkmer, B., Heinemann, M. (2011). Condition-dependent cell volume and concentration of *Escherichia coli* to facilitate data conversion for systems biology modeling. *PloS One* *6*, e23126.
- Vytvytska, O., Jakobsen, J.S., Balcunaite, G., Andersen, J.S., Baccarini, M., von Gabain, A. (1998). Host factor I, Hfq, binds to *Escherichia coli* ompA mRNA in a growth rate-dependent fashion and regulates its stability. *Proc. Natl. Acad. Sci. U. S. A.* *95*, 14118–14123.
- Wada, A. (1998). Growth phase coupled modulation of *Escherichia coli* ribosomes. *Genes Cells* *3*, 203–208.
- Wang, X., Dubey, A.K., Suzuki, K., Baker, C.S., Babitzke, P., Romeo, T. (2005). CsrA post-transcriptionally represses *pgaABCD*, responsible for synthesis of a biofilm polysaccharide adhesin of *Escherichia coli*. *Mol. Microbiol.* *56*, 1648–1663.

- Wang, X., Lord, D.M., Cheng, H.-Y., Osbourne, D.O., Hong, S.H., Sanchez-Torres, V., Quiroga, C., Zheng, K., Herrmann, T., Peti, W., et al. (2012). A new type V toxin-antitoxin system where mRNA for toxin GhoT is cleaved by antitoxin GhoS. *Nat. Chem. Biol.* *8*, 855–861.
- Wang, Y., Liu, C.L., Storey, J.D., Tibshirani, R.J., Herschlag, D., Brown, P.O. (2002). Precision and functional specificity in mRNA decay. *Proc. Natl. Acad. Sci.* *99*, 5860–5865.
- Wei, B., Shin, S., LaPorte, D., Wolfe, A.J., Romeo, T. (2000). Global Regulatory Mutations in *csrA* and *rpoS* Cause Severe Central Carbon Stress in *Escherichia coli* in the Presence of Acetate. *J. Bacteriol.* *182*, 1632–1640.
- Wei, B.L., Brun-Zinkernagel, A.-M., Simecka, J.W., Prüß, B.M., Babitzke, P., Romeo, T. (2001). Positive regulation of motility and *flhDC* expression by the RNA-binding protein CsrA of *Escherichia coli*. *Mol. Microbiol.* *40*, 245–256.
- Weilbacher, T., Suzuki, K., Dubey, A.K., Wang, X., Gudapaty, S., Morozov, I., Baker, C.S., Georgellis, D., Babitzke, P., Romeo, T. (2003). A novel sRNA component of the carbon storage regulatory system of *Escherichia coli*. *Mol. Microbiol.* *48*, 657–670.
- Wolfe, A.J. (2005). The Acetate Switch. *Microbiol. Mol. Biol. Rev.* *69*, 12–50.
- Yakhnin, A.V., Baker, C.S., Vakulskas, C.A., Yakhnin, H., Berezin, I., Romeo, T., Babitzke, P. (2013). CsrA activates *flhDC* expression by protecting *flhDC* mRNA from RNase E-mediated cleavage. *Mol. Microbiol.* *87*, 851–866.
- Yakhnin, H., Baker, C.S., Berezin, I., Evangelista, M.A., Rassin, A., Romeo, T., Babitzke, P. (2011a). CsrA represses translation of *sdiA*, encoding the N-acylhomoserine-L-lactone receptor of *Escherichia coli*, by binding exclusively within the coding region of *sdiA* mRNA. *J. Bacteriol.*
- Yakhnin, H., Yakhnin, A.V., Baker, C.S., Sineva, E., Berezin, I., Romeo, T., Babitzke, P. (2011b). Complex regulation of the global regulatory gene *csrA*: CsrA-mediated translational repression, transcription from five promoters by  $E\sigma^{70}$  and  $E\sigma(S)$ , and indirect transcriptional activation by CsrA. *Mol. Microbiol.* *81*, 689–704.
- Yang, H., Liu, M.Y., Romeo, T. (1996). Coordinate genetic regulation of glycogen catabolism and biosynthesis in *Escherichia coli* via the CsrA gene product. *J. Bacteriol.* *178*, 1012–1017.
- Yoon, S.H., Han, M.-J., Jeong, H., Lee, C.H., Xia, X.-X., Lee, D.-H., Shim, J.H., Lee, S.Y., Oh, T.K., Kim, J.F. (2012). Comparative multi-omics systems analysis of *Escherichia coli* strains B and K-12. *Genome Biol.* *13*, R37.
- Yoshida, H., Yamamoto, H., Uchiumi, T., Wada, A. (2004). RMF inactivates ribosomes by covering the peptidyl transferase centre and entrance of peptide exit tunnel. *Genes Cells* *9*, 271–278.
- Zhang, J., Sprung, R., Pei, J., Tan, X., Kim, S., Zhu, H., Liu, C.-F., Grishin, N.V., Zhao, Y. (2009). Lysine Acetylation Is a Highly Abundant and Evolutionarily Conserved Modification in *Escherichia coli*. *Mol. Cell. Proteomics* *8*, 215–225.



Zheng, J., Jia, Z. (2010). Structure of the bifunctional isocitrate dehydrogenase kinase/phosphatase. *Nature* 465, 961–965.

Zhou, D., Yang, R. (2006). Global analysis of gene transcription regulation in prokaryotes. *Cell. Mol. Life Sci. CMLS* 63, 2260–2290.

Zuo, Y., Deutscher, M.P. (2001). Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. *Nucleic Acids Res.* 29, 1017–1026.





# **PUBLICATIONS ET COMMUNICATIONS**



### **Publications**

**Esquerré T.**, Laguerre S., Turlan C., Carpousis A.J., Girbal L., Coccagn-Bousquet M. (2014) Dual role of transcription and transcript stability in the regulation of gene expression in *Escherichia coli* cells cultured on glucose at different growth rates  
*Nucleic Acids Research*, 42 (4): 2460-2472.

### **Manuscrits en cours de finalisation :**

**Esquerré T.**, Moisan A., Chiapello H., Arike L., Vilu R., Gaspin C., Coccagn-Bousquet M., Girbal L. Identification of mRNA half-life determinants in *Escherichia coli* cells at different growth rates

**Esquerré T.**, Turlan C., Bouvier M., Carpousis A.J., Coccagn-Bousquet M., Girbal L. Genome-wide extent and physiological significance of mRNA stability controlled by CsrA in *Escherichia coli*

### **Communications orales**

**Esquerré T.** (2011) Role of Csr system in regulations of mRNA stability in *Escherichia coli* adaptation to environment  
Séminaire d'unité, LMGM, Toulouse, Octobre 2011.

**Esquerré T.** (2013) Transcript stability: a factor controlling gene expression in *E. coli*  
5ème édition du séminaire Microbiotoul, Toulouse, Octobre 2013.

**Esquerré T.** (2013) Role of mRNA stability regulation in *Escherichia coli* gene expression at different growth rates  
Séminaire du Centre de Biologie Intégrative, LMGM, Toulouse, Novembre 2013.

**Communication par posters**

**Esquerré T.**, Turlan C., Carpousis A.J., Coccagn-Bousquet M., Girbal L. (2012) Role of post-transcriptional regulator CsrA and mRNA stability in *Escherichia coli* adaptation to environment

Journée des Microbiologistes de l'INRA, L'Isle-sur-la-Sorgue, Novembre 2012

Dressaire C., **Esquerré T.**, Nouaille S., Girbal L., Coccagn-Bousquet M. (2013) Bacterial adaptation: role of mRNA degradation

3rd International Conference on Regulating with RNA in Bacteria, Wurtzbourg, Allemagne, Juin 2013

**Esquerré T.**, Dressaire C., Nouaille S., Girbal L., Coccagn-Bousquet M. (2013) Role of mRNA degradation in bacterial adaptation

5th Congress of European Microbiologists FEMS, Leipzig, Allemagne, Juillet 2013

**Esquerré T.**, Laguerre S., Turlan C., Carpousis A.J., Girbal L., Coccagn-Bousquet M. (2014) Dual role of transcription and transcript stability in the regulation of gene expression in bacteria

EMBO Conference on Microbiology after the genomics revolution: Genomes 2014, Paris, Juin 2014







# **ANNEXES**



Cette partie Annexes contient les Table S1 et Table S2 du Chapitre III qui ne sont pas encore accessibles sur Internet.

**Table S1. List of the 455 mRNAs whose stability was directly controlled by CsrA.**

**CsrA regulation:** (-) negative and (+) positive in the *wt* cells.

**Number of CsrA binding site:** the number indicates the number of potential CsrA binding site; (-) means that no binding site was found.

**CsrA binding site location:** (5'UTR), 5' untranslated region; (CDS), coding sequence; (5'UTR-CDS), CsrA binding site overlapping both regions; if several binding sites are found, locations are separated by (/); (-) means that no binding site was found.

**Co-purified with CsrA-His<sub>6</sub>:** (Yes) if the mRNA was found to co-purify with CsrA-His<sub>6</sub> in Edwards et al, Mol. Mic., 2011; (-) means the mRNA was not found to co-purify with CsrA-His<sub>6</sub>.

Blattner	Gene Symbol	CsrA regulation	Fold change of half-lives <i>csrA51/wt</i>	Number of CsrA binding site	CsrA binding site location	Co-purified with CsrA-His <sub>6</sub>
b0019	<i>nhaA</i>	+	0,70	-	-	-
b0034	<i>caiF</i>	+	0,57	1	CDS	-
b0040	<i>caiT</i>	+	0,37	-	-	Yes
b0041	<i>fixA</i>	+	0,41	1	CDS	-
b0042	<i>fixB</i>	+	0,39	-	-	-
b0061	<i>araD</i>	+	0,60	-	-	-
b0062	<i>araA</i>	+	0,71	1	5'UTR	-
b0063	<i>araB</i>	+	0,68	-	-	-
b0065	<i>yabl</i>	+	0,61	-	-	-
b0072	<i>leuC</i>	-	1,50	1	5'UTR	-
b0073	<i>leuB</i>	-	1,53	1	5'UTR	Yes
b0118	<i>acnB</i>	-	2,15	1	CDS	Yes
b0197	<i>metQ</i>	+	0,71	1	5'UTR	-
b0212	<i>gloB</i>	+	0,69	-	-	-
b0221	<i>fadE</i>	+	0,71	-	-	-
b0228	<i>yafM</i>	+	0,55	1	5'UTR	Yes
b0245	<i>ykfl</i>	+	0,65	-	-	-
b0246	<i>yafW</i>	+	0,68	1	CDS	-

b0252	<i>yafZ</i>	+	0,60	1	5'UTR	-
b0258	<i>ykfC</i>	+	0,67	1	5'UTR	-
b0260	<i>mmuP</i>	+	0,69	1	5'UTR	-
b0268	<i>yagE</i>	+	0,58	1	5'UTR	Yes
b0269	<i>yagF</i>	+	0,60	-	-	-
b0270	<i>yagG</i>	+	0,56	-	-	Yes
b0271	<i>yagH</i>	+	0,45	-	-	-
b0287	<i>yagU</i>	+	0,70	-	-	-
b0303	<i>ykgI</i>	+	0,51	-	-	-
b0304	<i>ykgC</i>	+	0,54	-	-	-
b0306	<i>ykgE</i>	+	0,59	-	-	Yes
b0307	<i>ykgF</i>	+	0,69	1	5'UTR	-
b0308	<i>ykgG</i>	+	0,69	-	-	-
b0320	<i>yahF</i>	+	0,48	1	5'UTR	-
b0321	<i>yahG</i>	+	0,45	-	-	-
b0330	<i>prpR</i>	+	0,53	-	-	-
b0342	<i>lacA</i>	+	0,58	-	-	-
b0343	<i>lacY</i>	+	0,66	2	5'UTR / CDS	-
b0344	<i>lacZ</i>	+	0,50	-	-	-
b0353	<i>mhpT</i>	+	0,55	-	-	-
b0357	<i>frmR</i>	+	0,78	-	-	-
b0365	<i>tauA</i>	+	0,65	-	-	Yes
b0366	<i>tauB</i>	+	0,69	1	5'UTR	-
b0367	<i>tauC</i>	+	0,52	1	5'UTR	-
b0368	<i>tauD</i>	+	0,56	-	-	-
b0384	<i>psiF</i>	+	0,50	1	5'UTR	-
b0391	<i>yaiE</i>	+	0,68	-	-	-
b0406	<i>tgt</i>	+	0,54	-	-	-
b0426	<i>yajQ</i>	+	0,70	2	5'UTR	-
b0440	<i>hupB</i>	+	0,34	2	5'UTR	Yes
b0443	<i>fadM</i>	+	0,36	2	CDS	-
b0476	<i>aes</i>	+	0,72	1	5'UTR	Yes
b0507	<i>gcl</i>	+	0,48	1	CDS	-
b0511	<i>ybbW</i>	+	0,68	1	5'UTR	-
b0516	<i>allC</i>	+	0,76	-	-	-
b0518	<i>fdrA</i>	+	0,50	1	5'UTR	-
b0525	<i>ppiB</i>	+	0,75	1	5'UTR	-
b0553	<i>nmpC</i>	+	0,73	-	-	-
b0565	<i>ompT</i>	+	0,52	-	-	Yes
b0567	<i>ybcH</i>	+	0,53	-	-	-
b0573	<i>cusF</i>	+	0,58	-	-	-
b0610	<i>rnk</i>	+	0,61	1	CDS	-
b0623	<i>cspE</i>	+	0,64	-	-	Yes
b0627	<i>tatE</i>	+	0,55	-	-	-
b0688	<i>pgm</i>	-	1,41	-	-	Yes

b0721	<i>sdhC</i>	-	1,17	1	5'UTR	-
b0755	<i>gpmA</i>	+	0,75	1	CDS	-
b0767	<i>pgl</i>	+	0,82	-	-	-
b0771	<i>ybhJ</i>	+	0,44	-	-	-
b0784	<i>moaD</i>	+	0,53	1	5'UTR-CDS	-
b0823	<i>ybiW</i>	+	0,46	1	5'UTR	-
b0860	<i>artJ</i>	+	0,71	-	-	-
b0865	<i>ybjP</i>	+	0,67	2	5'UTR	-
b0875	<i>aqpZ</i>	+	0,53	-	-	-
b0882	<i>clpA</i>	+	0,48	1	5'UTR	-
b0889	<i>lrp</i>	+	0,66	2	5'UTR / CDS	-
b0894	<i>dmsA</i>	+	0,64	-	-	Yes
b0895	<i>dmsB</i>	+	0,65	1	5'UTR	-
b0896	<i>dmsC</i>	+	0,66	4	5'UTR / 5'UTR-CDS / CDS	-
b0897	<i>ycaC</i>	+	0,29	1	5'UTR-CDS	-
b0903	<i>pflB</i>	+	0,48	1	CDS	Yes
b0906	<i>ycaP</i>	+	0,63	-	-	-
b0929	<i>ompF</i>	+	0,35	-	-	Yes
b0934	<i>ssuC</i>	+	0,62	1	5'UTR	-
b0935	<i>ssuD</i>	+	0,54	-	-	-
b0936	<i>ssuA</i>	+	0,50	-	-	-
b0937	<i>ssuE</i>	+	0,49	3	5'UTR / CDS	-
b0966	<i>hspQ</i>	+	0,63	2	5'UTR / CDS	-
b0981	<i>etk</i>	+	0,54	1	CDS	Yes
b0982	<i>etp</i>	+	0,60	-	-	-
b1042	<i>csgA</i>	+	0,46	-	-	-
b1056	<i>yceI</i>	+	0,39	2	5'UTR	-
b1061	<i>dinI</i>	+	0,51	-	-	-
b1088	<i>yceD</i>	+	0,62	-	-	-
b1089	<i>rpmF</i>	+	0,63	1	5'UTR	-
b1092	<i>fabD</i>	-	1,40	2	5'UTR	-
b1095	<i>fabF</i>	-	1,61	-	-	-
b1101	<i>ptsG</i>	-	1,36	-	-	-
b1103	<i>hinT</i>	+	0,74	2	5'UTR / CDS	-
b1108	<i>ycfP</i>	+	0,80	2	5'UTR / CDS	-
b1127	<i>pepT</i>	+	0,58	2	5'UTR-CDS / CDS	-
b1146	<i>croE</i>	+	0,39	1	CDS	-
b1147	<i>ymfL</i>	+	0,51	-	-	-
b1148	<i>ymfM</i>	+	0,50	-	-	-
b1149	NA	+	0,57	-	-	-
b1150	<i>ymfR</i>	+	0,51	-	-	-
b1151	<i>beeE</i>	+	0,56	1	5'UTR	-
b1178	<i>pliG</i>	+	0,56	1	CDS	-

b1195	<i>ymgE</i>	+	0,45	-	-	-
b1197	<i>treA</i>	+	0,78	-	-	Yes
b1205	<i>yehH</i>	+	0,67	1	5'UTR	-
b1219	<i>yehN</i>	+	0,72	-	-	-
b1221	<i>narL</i>	+	0,63	1	5'UTR	Yes
b1237	<i>hns</i>	+	0,36	-	-	-
b1324	<i>tpx</i>	+	0,54	-	-	-
b1336	<i>abgT</i>	+	0,61	1	CDS	-
b1337	<i>abgB</i>	+	0,63	1	CDS	-
b1338	<i>abgA</i>	+	0,65	-	-	-
b1349	<i>recT</i>	+	0,59	1	CDS	-
b1359	<i>ydaU</i>	+	0,49	-	-	-
b1361	<i>ydaW</i>	+	0,52	1	CDS	-
b1380	<i>ldhA</i>	+	0,60	-	-	-
b1383	<i>ydbL</i>	+	0,53	-	-	-
b1385	<i>feaB</i>	+	0,67	-	-	-
b1387	<i>paaZ</i>	+	0,45	1	CDS	Yes
b1389	<i>paaB</i>	+	0,63	2	5'UTR	-
b1390	<i>paaC</i>	+	0,61	-	-	-
b1391	<i>paaD</i>	+	0,62	1	CDS	-
b1392	<i>paaE</i>	+	0,48	1	5'UTR	-
b1406	<i>ydbC</i>	+	0,76	-	-	-
b1415	<i>aldA</i>	+	0,65	1	CDS	Yes
b1419	<i>ydcA</i>	+	0,55	2	5'UTR / CDS	Yes
b1423	<i>ydcJ</i>	+	0,72	-	-	-
b1431	<i>ydcL</i>	+	0,67	-	-	-
b1433	<i>ydcO</i>	+	0,58	1	5'UTR	-
b1440	<i>ydcS</i>	+	0,78	1	5'UTR	-
b1449	<i>yncB</i>	+	0,64	2	5'UTR	Yes
b1461	<i>pptA</i>	+	0,53	-	-	-
b1479	<i>maeA</i>	-	1,41	2	5'UTR	-
b1482	<i>osmC</i>	+	0,35	1	CDS	-
b1487	<i>ddpA</i>	+	0,72	-	-	-
b1489	<i>dosP</i>	+	0,62	1	5'UTR	-
b1511	<i>lsrK</i>	+	0,72	2	5'UTR	-
b1517	<i>lsrF</i>	+	0,72	-	-	-
b1518	<i>lsrG</i>	+	0,65	1	5'UTR	-
b1521	<i>uxaB</i>	+	0,62	-	-	-
b1538	<i>dcp</i>	+	0,79	1	5'UTR	-
b1541	<i>ydfZ</i>	+	0,46	-	-	-
b1562	<i>hokD</i>	+	0,51	-	-	-
b1563	<i>relE</i>	+	0,51	-	-	-
b1588	<i>ynfF</i>	+	0,56	-	-	-
b1596	<i>ynfM</i>	+	0,55	1	5'UTR	-
b1602	<i>pntB</i>	-	1,55	-	-	-

b1603	<i>pntA</i>	-	1,54	-	-	-
b1609	<i>rstB</i>	+	0,63	-	-	-
b1612	<i>fumA</i>	-	1,32	-	-	-
b1619	<i>hdhA</i>	+	0,64	1	5'UTR	-
b1642	<i>slyA</i>	+	0,59	1	5'UTR	-
b1654	<i>grxD</i>	+	0,72	1	5'UTR	-
b1667	<i>ydhR</i>	+	0,53	-	-	-
b1692	<i>ydiB</i>	+	0,57	2	5'UTR / 5'UTR-CDS	Yes
b1694	<i>ydiF</i>	+	0,60	2	5'UTR / CDS	-
b1695	<i>ydiO</i>	+	0,65	1	5'UTR-CDS	-
b1698	<i>ydiR</i>	+	0,53	-	-	-
b1718	<i>infC</i>	-	1,37	3	5'UTR / CDS	Yes
b1719	<i>thrS</i>	-	1,64	2	5'UTR / CDS	Yes
b1724	<i>ydiZ</i>	+	0,43	1	CDS	-
b1725	<i>yniA</i>	+	0,69	-	-	Yes
b1732	<i>katE</i>	+	0,65	2	5'UTR / CDS	-
b1744	<i>astE</i>	+	0,58	1	5'UTR	-
b1746	<i>astD</i>	+	0,61	1	5'UTR	-
b1747	<i>astA</i>	+	0,68	-	-	-
b1772	<i>ydjH</i>	+	0,55	-	-	-
b1773	<i>ydjI</i>	+	0,63	1	CDS	-
b1774	<i>ydjJ</i>	+	0,57	-	-	-
b1779	<i>gapA</i>	-	1,66	-	-	Yes
b1782	<i>mipA</i>	-	1,41	1	5'UTR	-
b1783	<i>yeaG</i>	+	0,49	2	5'UTR / CDS	-
b1784	<i>yeaH</i>	+	0,58	-	-	-
b1789	<i>yeaL</i>	+	0,45	-	-	-
b1800	<i>dmlA</i>	+	0,69	4	5'UTR / CDS	Yes
b1817	<i>manX</i>	-	2,02	1	5'UTR	-
b1818	<i>manY</i>	-	2,29	1	5'UTR	-
b1819	<i>manZ</i>	-	2,10	1	CDS	-
b1820	<i>yobD</i>	+	0,64	1	CDS	-
b1823	<i>cspC</i>	+	0,64	1	5'UTR	-
b1824	<i>yobF</i>	+	0,45	-	-	-
b1852	<i>zwf</i>	+	0,79	-	-	Yes
b1854	<i>pykA</i>	-	1,35	-	-	-
b1874	<i>cutC</i>	+	0,69	-	-	-
b1884	<i>cheR</i>	+	0,54	-	-	-
b1885	<i>tap</i>	+	0,66	1	5'UTR	-
b1886	<i>tar</i>	+	0,60	-	-	-
b1889	<i>motB</i>	+	0,66	2	5'UTR / CDS	-
b1901	<i>araF</i>	+	0,54	-	-	-
b2001	<i>yeeR</i>	+	0,60	1	CDS	-
b2009	<i>sbmC</i>	+	0,66	1	5'UTR	-



b2038	<i>rfbC</i>	+	0,55	1	5'UTR	-
b2080	<i>yegP</i>	+	0,42	1	5'UTR	-
b2094	<i>gatA</i>	-	1,62	2	5'UTR	Yes
b2095	<i>gatZ</i>	-	1,84	1	5'UTR	Yes
b2098	<i>yegT</i>	+	0,68	-	-	-
b2099	<i>yegU</i>	+	0,60	3	5'UTR / CDS	-
b2100	<i>yegV</i>	+	0,67	2	5'UTR	-
b2104	<i>thiM</i>	+	0,58	1	5'UTR	-
b2158	<i>yeiH</i>	+	0,59	1	5'UTR	-
b2159	<i>nfo</i>	+	0,70	-	-	-
b2174	<i>lpxT</i>	+	0,49	1	CDS	-
b2194	<i>ccmH</i>	+	0,53	-	-	-
b2195	<i>ccmG</i>	+	0,59	1	CDS	-
b2196	<i>ccmF</i>	+	0,58	2	5'UTR / CDS	-
b2205	<i>napG</i>	+	0,68	1	5'UTR	Yes
b2220	<i>atoC</i>	+	0,68	2	5'UTR / CDS	-
b2223	<i>atoE</i>	+	0,51	-	-	-
b2226	<i>yfaQ</i>	+	0,57	1	5'UTR	-
b2245	<i>yfaU</i>	+	0,57	-	-	-
b2246	<i>rhmT</i>	+	0,48	1	5'UTR	-
b2248	<i>rhmR</i>	+	0,56	2	5'UTR	-
b2266	<i>elaB</i>	+	0,45	1	5'UTR	-
b2267	<i>elaA</i>	+	0,66	1	5'UTR	-
b2309	<i>hisJ</i>	+	0,62	1	5'UTR	Yes
b2310	<i>argT</i>	+	0,63	-	-	-
b2333	<i>yfcP</i>	+	0,46	-	-	-
b2343	<i>yfcZ</i>	+	0,51	1	CDS	Yes
b2360	<i>yfdQ</i>	+	0,68	-	-	-
b2371	<i>yfdE</i>	+	0,48	-	-	-
b2386	<i>fryC</i>	+	0,47	1	5'UTR	-
b2387	<i>fryB</i>	+	0,56	1	5'UTR	-
b2407	<i>xapA</i>	+	0,43	1	5'UTR	-
b2431	<i>yfeX</i>	+	0,75	1	CDS	Yes
b2437	<i>eutR</i>	+	0,55	1	5'UTR	-
b2447	<i>yffP</i>	+	0,59	-	-	-
b2450	<i>yffS</i>	+	0,56	2	CDS	-
b2463	<i>maeB</i>	-	2,26	3	5'UTR / 5'UTR-CDS / CDS	Yes
b2464	<i>talA</i>	+	0,56	2	5'UTR / CDS	-
b2465	<i>tktB</i>	+	0,62	-	-	-
b2496	<i>hda</i>	+	0,51	1	5'UTR	-
b2521	<i>sseA</i>	-	1,32	1	5'UTR	-
b2528	<i>iscA</i>	+	0,80	-	-	-
b2535	<i>csiE</i>	+	0,61	-	-	-

b2539	<i>hcaF</i>	+	0,71	2	5'UTR	-
b2541	<i>hcaB</i>	+	0,65	1	CDS	-
b2542	<i>hcaD</i>	+	0,73	1	5'UTR	Yes
b2546	<i>yphD</i>	+	0,46	-	-	Yes
b2547	<i>yphE</i>	+	0,52	1	CDS	-
b2548	<i>yphF</i>	+	0,52	-	-	-
b2549	<i>yphG</i>	+	0,52	1	CDS	Yes
b2579	<i>yfiD</i>	+	0,63	-	-	-
b2599	<i>pheA</i>	+	0,79	-	-	-
b2632	<i>yfjP</i>	+	0,57	-	-	Yes
b2675	<i>nrdE</i>	+	0,58	-	-	-
b2679	<i>proX</i>	+	0,62	1	5'UTR	-
b2698	<i>recX</i>	+	0,57	1	5'UTR	-
b2702	<i>srlA</i>	+	0,31	1	CDS	-
b2703	<i>srlE</i>	+	0,28	2	CDS	-
b2704	<i>srlB</i>	+	0,32	-	-	-
b2705	<i>srlD</i>	+	0,33	1	5'UTR	-
b2716	<i>ascB</i>	+	0,58	1	5'UTR	-
b2741	<i>rpoS</i>	+	0,82	1	5'UTR	-
b2774	<i>ygcW</i>	+	0,60	1	CDS	-
b2779	<i>eno</i>	+	0,66	-	-	-
b2787	<i>gudD</i>	+	0,53	1	5'UTR	-
b2789	<i>gudP</i>	+	0,44	-	-	Yes
b2799	<i>fucO</i>	-	2,96	2	5'UTR / 5'UTR-CDS	-
b2801	<i>fucP</i>	+	0,66	1	5'UTR	-
b2802	<i>fucl</i>	+	0,51	-	-	-
b2833	<i>ygdR</i>	+	0,53	2	5'UTR / 5'UTR-CDS	-
b2841	<i>araE</i>	+	0,56	-	-	-
b2843	<i>kdul</i>	+	0,75	1	5'UTR	-
b2875	<i>yqeB</i>	+	0,58	1	5'UTR	-
b2877	<i>mocA</i>	+	0,50	2	5'UTR / CDS	-
b2904	<i>gcvH</i>	+	0,73	-	-	Yes
b2914	<i>rpiA</i>	+	0,69	3	5'UTR / CDS	-
b2933	<i>cmtA</i>	+	0,69	-	-	-
b2937	<i>speB</i>	+	0,53	1	5'UTR	-
b2938	<i>speA</i>	-	1,49	1	CDS	Yes
b2940	<i>yqgC</i>	-	1,50	1	CDS	-
b2943	<i>galP</i>	+	0,54	-	-	Yes
b2962	<i>yggX</i>	+	0,53	2	5'UTR / CDS	-
b2989	<i>yghU</i>	+	0,59	-	-	-
b3008	<i>metC</i>	+	0,64	-	-	-
b3011	<i>yqhD</i>	+	0,50	1	5'UTR	-
b3029	<i>ygiN</i>	+	0,68	1	5'UTR	-

b3067	<i>rpoD</i>	-	2,79	1	5'UTR	Yes
b3068	<i>mug</i>	+	0,68	1	CDS	-
b3070	<i>yqjH</i>	+	0,70	-	-	-
b3071	<i>yqjI</i>	+	0,61	1	5'UTR	-
b3076	<i>ebgA</i>	+	0,51	1	5'UTR	-
b3077	<i>ebgC</i>	+	0,57	-	-	-
b3078	<i>ygjI</i>	+	0,52	1	5'UTR	-
b3089	<i>sstT</i>	-	1,35	2	5'UTR	-
b3091	<i>uxaA</i>	+	0,66	-	-	-
b3092	<i>uxaC</i>	+	0,52	-	-	-
b3115	<i>tdcD</i>	+	0,70	1	5'UTR	Yes
b3116	<i>tdcC</i>	+	0,68	2	5'UTR-CDS / CDS	Yes
b3117	<i>tdcB</i>	+	0,63	-	-	Yes
b3124	<i>garK</i>	+	0,32	3	5'UTR / CDS	-
b3126	<i>garL</i>	+	0,43	1	5'UTR	-
b3128	<i>garD</i>	+	0,76	-	-	-
b3132	<i>kbaZ</i>	+	0,58	1	CDS	Yes
b3133	<i>agaV</i>	+	0,67	-	-	-
b3134	<i>agaW</i>	+	0,73	1	5'UTR	Yes
b3135	<i>agaA</i>	+	0,56	2	5'UTR / CDS	-
b3139	<i>agaC</i>	+	0,35	1	CDS	Yes
b3165	<i>rpsO</i>	+	0,66	1	5'UTR	Yes
b3223	<i>nanE</i>	+	0,71	1	5'UTR	Yes
b3234	<i>degQ</i>	+	0,67	-	-	-
b3253	<i>yhdH</i>	+	0,69	1	CDS	Yes
b3255	<i>accB</i>	-	1,43	1	5'UTR-CDS	Yes
b3268	<i>yhdW</i>	+	0,66	1	5'UTR	-
b3309	<i>rplX</i>	-	1,33	-	-	Yes
b3310	<i>rplN</i>	-	1,29	1	5'UTR	Yes
b3312	<i>rpmC</i>	-	1,74	-	-	-
b3313	<i>rplP</i>	-	2,10	2	5'UTR / CDS	-
b3314	<i>rpsC</i>	-	1,87	1	CDS	Yes
b3315	<i>rplV</i>	-	1,68	1	5'UTR	-
b3316	<i>rpsS</i>	-	1,93	-	-	-
b3317	<i>rplB</i>	-	1,71	-	-	-
b3318	<i>rplW</i>	-	1,64	-	-	Yes
b3319	<i>rplD</i>	-	1,51	1	5'UTR	Yes
b3320	<i>rplC</i>	-	1,57	-	-	-
b3321	<i>rpsJ</i>	-	1,47	2	5'UTR	-
b3336	<i>bfr</i>	+	0,47	2	5'UTR	-
b3340	<i>fusA</i>	-	2,15	1	5'UTR	Yes
b3341	<i>rpsG</i>	-	2,04	-	-	Yes
b3342	<i>rpsL</i>	-	2,01	2	5'UTR	Yes
b3364	<i>tsgA</i>	+	0,53	3	CDS	-

b3367	<i>nirC</i>	+	0,63	2	5'UTR	-
b3371	<i>frlB</i>	+	0,59	1	CDS	-
b3401	<i>hslO</i>	+	0,72	-	-	-
b3403	<i>pck</i>	-	2,61	-	-	Yes
b3405	<i>ompR</i>	+	0,74	-	-	Yes
b3411	<i>yhgA</i>	+	0,52	1	5'UTR	-
b3414	<i>nfuA</i>	+	0,74	1	CDS	-
b3432	<i>glgB</i>	-	1,92	2	5'UTR / CDS	Yes
b3433	<i>asd</i>	-	1,47	-	-	Yes
b3447	<i>ggt</i>	+	0,52	1	CDS	-
b3451	<i>ugpE</i>	+	0,63	-	-	Yes
b3452	<i>ugpA</i>	+	0,66	-	-	Yes
b3453	<i>ugpB</i>	+	0,72	-	-	Yes
b3460	<i>livJ</i>	+	0,20	3	5'UTR / 5'UTR-CDS / CDS	-
b3479	<i>nikD</i>	+	0,51	-	-	Yes
b3491	<i>yhiM</i>	+	0,57	2	5'UTR	-
b3506	<i>slp</i>	+	0,38	1	5'UTR	-
b3509	<i>hdeB</i>	+	0,56	-	-	-
b3510	<i>hdeA</i>	+	0,56	1	5'UTR	-
b3529	<i>yhjK</i>	+	0,57	1	5'UTR	-
b3532	<i>bcsB</i>	+	0,64	1	CDS	-
b3547	<i>yhjX</i>	+	0,60	1	CDS	Yes
b3550	<i>yiaC</i>	+	0,71	-	-	-
b3565	<i>xylA</i>	+	0,60	-	-	-
b3566	<i>xylF</i>	+	0,54	-	-	-
b3573	<i>ysaA</i>	+	0,57	1	CDS	-
b3579	<i>yiaO</i>	+	0,61	1	5'UTR	Yes
b3580	<i>lyxK</i>	+	0,69	3	5'UTR / CDS	-
b3583	<i>yiaS</i>	+	0,43	-	-	-
b3600	<i>mtlD</i>	+	0,75	-	-	-
b3605	<i>lldD</i>	+	0,66	2	5'UTR / CDS	-
b3709	<i>tnaB</i>	+	0,75	2	5'UTR / CDS	-
b3721	<i>bglB</i>	+	0,53	3	5'UTR / CDS	-
b3748	<i>rbsD</i>	+	0,51	-	-	-
b3774	<i>ilvC</i>	+	0,82	-	-	Yes
b3829	<i>metE</i>	+	0,69	3	5'UTR / CDS	Yes
b3831	<i>udp</i>	+	0,51	-	-	-
b3858	<i>yihD</i>	+	0,47	1	5'UTR	-
b3860	<i>dsbA</i>	+	0,68	-	-	-
b3870	<i>glnA</i>	-	1,71	-	-	Yes
b3874	<i>yihN</i>	+	0,54	1	5'UTR	-
b3878	<i>yihQ</i>	+	0,53	1	5'UTR	-
b3887	<i>dtd</i>	+	0,77	1	CDS	-

b3888	<i>yiiD</i>	+	0,78	-	-	Yes
b3908	<i>sodA</i>	+	0,57	1	CDS	-
b3909	<i>kdgT</i>	+	0,66	-	-	Yes
b3919	<i>tpiA</i>	+	0,83	-	-	Yes
b3926	<i>glpK</i>	+	0,65	1	5'UTR	-
b3927	<i>glpF</i>	+	0,57	-	-	-
b3928	<i>zapB</i>	+	0,66	2	5'UTR / CDS	-
b3929	<i>rraA</i>	+	0,63	-	-	Yes
b3936	<i>rpmE</i>	+	0,61	-	-	-
b3981	<i>secE</i>	-	1,34	1	CDS	Yes
b3990	<i>thiH</i>	+	0,55	1	CDS	-
b4000	<i>hupA</i>	+	0,57	1	5'UTR	-
b4014	<i>aceB</i>	-	1,33	1	5'UTR	Yes
b4023	<i>yjbD</i>	+	0,52	-	-	Yes
b4034	<i>malE</i>	+	0,51	3	5'UTR	Yes
b4051	<i>qorA</i>	+	0,65	1	5'UTR	-
b4069	<i>acs</i>	+	0,62	1	5'UTR	-
b4076	<i>nrfG</i>	+	0,44	-	-	-
b4079	<i>fdhF</i>	+	0,75	-	-	-
b4084	<i>alsK</i>	+	0,75	1	CDS	-
b4085	<i>alsE</i>	+	0,71	2	5'UTR / CDS	-
b4087	<i>alsA</i>	+	0,59	-	-	-
b4088	<i>alsB</i>	+	0,66	-	-	-
b4090	<i>rpiB</i>	+	0,59	3	5'UTR	-
b4107	<i>yjdN</i>	+	0,55	-	-	-
b4119	<i>melA</i>	+	0,64	2	5'UTR	-
b4120	<i>melB</i>	+	0,66	3	5'UTR / CDS	-
b4122	<i>fumB</i>	+	0,67	-	-	Yes
b4129	<i>lysU</i>	+	0,73	2	5'UTR / CDS	Yes
b4131	<i>cadA</i>	+	0,53	2	5'UTR	-
b4138	<i>dcuA</i>	-	1,99	1	5'UTR	Yes
b4177	<i>purA</i>	+	0,71	-	-	-
b4190	<i>yjfP</i>	+	0,59	1	5'UTR	-
b4192	<i>ulaG</i>	+	0,58	1	5'UTR	-
b4193	<i>ulaA</i>	+	0,68	-	-	-
b4194	<i>ulaB</i>	+	0,72	1	5'UTR	-
b4196	<i>ulaD</i>	+	0,71	2	5'UTR	-
b4198	<i>ulaF</i>	+	0,77	1	CDS	-
b4203	<i>rplI</i>	+	0,64	-	-	-
b4211	<i>qorB</i>	+	0,67	-	-	-
b4214	<i>cysQ</i>	+	0,60	1	CDS	Yes
b4216	<i>ytfJ</i>	+	0,54	-	-	Yes
b4219	<i>msrA</i>	+	0,59	1	CDS	Yes
b4227	<i>ytfQ</i>	+	0,53	-	-	-
b4230	<i>ytfT</i>	+	0,53	1	CDS	-

b4231	<i>yjfF</i>	+	0,59	1	5'UTR	-
b4234	<i>yjgA</i>	+	0,55	-	-	-
b4237	<i>nrdG</i>	+	0,49	1	5'UTR	-
b4238	<i>nrdD</i>	+	0,70	-	-	-
b4243	<i>yjgF</i>	+	0,66	-	-	-
b4244	<i>pyrI</i>	+	0,68	2	5'UTR	-
b4245	<i>pyrB</i>	+	0,67	-	-	-
b4251	<i>yjgJ</i>	+	0,53	-	-	-
b4254	<i>argI</i>	+	0,74	1	5'UTR	-
b4255	<i>rraB</i>	+	0,67	-	-	-
b4265	<i>idnT</i>	+	0,64	-	-	Yes
b4266	<i>idnO</i>	+	0,67	-	-	-
b4267	<i>idnD</i>	+	0,61	-	-	-
b4269	<i>yjgB</i>	+	0,61	1	5'UTR	-
b4291	<i>fecA</i>	+	0,59	-	-	-
b4298	<i>yjhH</i>	+	0,69	1	5'UTR	-
b4301	<i>sgcE</i>	+	0,71	3	5'UTR / CDS	-
b4302	<i>sgcA</i>	+	0,69	2	5'UTR / 5'UTR-CDS	-
b4306	<i>yjhP</i>	+	0,57	2	5'UTR / 5'UTR-CDS	-
b4334	<i>yjiL</i>	+	0,67	2	5'UTR / CDS	-
b4348	<i>hsdS</i>	+	0,53	2	CDS	-
b4354	<i>yjiY</i>	+	0,67	-	-	-
b4355	<i>tsr</i>	+	0,65	-	-	-
b4356	<i>yjiL</i>	+	0,65	-	-	-
b4378	<i>yjiV</i>	+	0,66	1	5'UTR	-
b4381	<i>deoC</i>	+	0,66	1	CDS	-
b4401	<i>arcA</i>	+	0,80	1	5'UTR	-
b4412	<i>hokC</i>	+	0,50	1	5'UTR	-
b4465	<i>yggP</i>	+	0,59	1	CDS	Yes
b4478	<i>dgoD</i>	+	0,62	-	-	-
b4485	<i>ytfR</i>	+	0,56	3	5'UTR / CDS	-
b4500	<i>yfaS</i>	+	0,61	2	5'UTR	-
b4504	<i>ykfH</i>	+	0,52	-	-	-
b4512	<i>ybdD</i>	+	0,74	-	-	-
b4525	<i>ymjC</i>	+	0,59	1	5'UTR	-
b4543	<i>ypaA</i>	+	0,67	1	5'UTR	-
b4551	<i>yheV</i>	+	0,77	1	5'UTR	-
b4572	<i>ylbE</i>	+	0,57	1	CDS	-
b4598	<i>yncl</i>	+	0,40	-	-	-
b4601	<i>ydgU</i>	+	0,48	-	-	-
b4613	<i>dinQ</i>	+	0,42	1	CDS	-
b4656	<i>yjhY</i>	+	0,57	-	-	-

**Table S2. List of the 99 mRNAs whose mRNA stability was controlled by CsrA and the mRNA stability regulation was associated with differential mRNA concentration.**

**CsrA regulation:** (-) negative and (+) positive in the wt cells.

**Number of CsrA binding site:** the number indicates the number of potential CsrA binding site; (-) means that no binding site was found.

**CsrA binding site location:** (5'UTR), 5' untranslated region; (CDS), coding sequence; (5'UTR-CDS), CsrA binding site overlapping both regions; if several binding sites are found, locations are separated by (/); (-) means that no binding site was found.

**Co-purified with CsrA-His<sub>6</sub>:** (Yes) if the mRNA was found to co-purify with CsrA-His<sub>6</sub> in Edwards et al, Mol. Mic., 2011; (-) means the mRNA was not found to co-purify with CsrA-His<sub>6</sub>.

**Degradational coefficient  $\rho_D$  value:** if  $\rho_D \geq 0,6$  (bold), mRNA concentration is mainly controlled by degradation; if  $\rho_D \leq 0,6$  (italic), mRNA concentration is equally controlled by degradation and transcription or if mainly controlled by transcription.

Blattner	Gene Symbol	CsrA regulation	Fold change of half-lives <i>csrA51/wt</i>	Fold change of mRNA level <i>csrA51/wt</i>	Number of CsrA binding site	CsrA binding site location	Co-purified with CsrA-His <sub>6</sub>	Degradational coefficient $\rho_D$ value
b0072	<i>leuC</i>	-	1,50	1,28	1	5'UTR	-	<b>1,67</b>
b0073	<i>leuB</i>	-	1,53	1,34	1	5'UTR	Yes	<b>1,45</b>
b0118	<i>acnB</i>	-	2,15	1,12	1	CDS	Yes	<b>6,83</b>
b0212	<i>gloB</i>	+	0,69	0,65	-	-	-	<b>0,88</b>
b0391	<i>yaiE</i>	+	0,68	0,62	-	-	-	<b>0,82</b>
b0406	<i>tgt</i>	+	0,54	0,82	-	-	-	<b>3,01</b>
b0476	<i>aes</i>	+	0,72	0,71	1	5'UTR	Yes	<b>0,99</b>
b0688	<i>pgm</i>	-	1,41	3,19	-	-	Yes	<i>0,29</i>
b0721	<i>sdhC</i>	-	1,17	1,13	1	5'UTR	-	<b>1,28</b>
b0897	<i>ycaC</i>	+	0,29	0,61	1	5'UTR-CDS	-	<b>2,54</b>
b1042	<i>csgA</i>	+	0,46	0,77	-	-	-	<b>3,04</b>
b1089	<i>rpmF</i>	+	0,63	0,79	1	5'UTR	-	<b>1,97</b>
b1092	<i>fabD</i>	-	1,40	1,85	2	5'UTR	-	0,55
b1095	<i>fabF</i>	-	1,61	1,28	-	-	-	<b>1,94</b>
b1178	<i>pliG</i>	+	0,56	0,65	1	CDS	-	<b>1,33</b>

b1219	<i>yehN</i>	+	0,72	0,58	-	-	-	<b>0,60</b>
b1324	<i>tpx</i>	+	0,54	0,77	-	-	-	<b>2,31</b>
b1380	<i>ldhA</i>	+	0,60	0,82	-	-	-	<b>2,62</b>
b1415	<i>aldA</i>	+	0,65	0,93	1	CDS	Yes	<b>5,58</b>
b1440	<i>ydcS</i>	+	0,78	0,50	1	5'UTR	-	<i>0,35</i>
b1461	<i>pptA</i>	+	0,53	0,61	-	-	-	<b>1,27</b>
b1479	<i>maeA</i>	-	1,41	1,32	2	5'UTR	-	<b>1,25</b>
b1538	<i>dcp</i>	+	0,79	0,78	1	5'UTR	-	<b>0,92</b>
b1602	<i>pntB</i>	-	1,55	1,32	-	-	-	<b>1,56</b>
b1603	<i>pntA</i>	-	1,54	1,21	-	-	-	<b>2,25</b>
b1612	<i>fumA</i>	-	1,32	1,12	-	-	-	<b>2,48</b>
b1619	<i>hdhA</i>	+	0,64	0,77	1	5'UTR	-	<b>1,70</b>
b1667	<i>ydhR</i>	+	0,53	0,19	-	-	-	<i>0,38</i>
b1692	<i>ydiB</i>	+	0,57	0,64	2	5'UTR / 5'UTR- CDS	Yes	<b>1,23</b>
b1694	<i>ydiF</i>	+	0,60	0,76	2	5'UTR / CDS	-	<b>1,84</b>
b1695	<i>ydiO</i>	+	0,65	0,73	1	5'UTR- CDS	-	<b>1,35</b>
b1718	<i>infC</i>	-	1,37	1,29	3	5'UTR / CDS	Yes	<b>1,23</b>
b1719	<i>thrS</i>	-	1,64	1,59	2	5'UTR / CDS	Yes	<b>1,06</b>
b1779	<i>gapA</i>	-	1,66	1,39	-	-	Yes	<b>1,54</b>
b1782	<i>mipA</i>	-	1,41	1,36	1	5'UTR	-	<b>1,12</b>
b1800	<i>dmlA</i>	+	0,69	0,67	4	5'UTR / CDS	Yes	<b>0,92</b>
b1817	<i>manX</i>	-	2,02	1,71	1	5'UTR	-	<b>1,32</b>
b1818	<i>manY</i>	-	2,29	1,25	1	5'UTR	-	<b>3,74</b>
b1819	<i>manZ</i>	-	2,10	1,37	1	CDS	-	<b>2,37</b>
b1823	<i>cspC</i>	+	0,64	0,28	1	5'UTR	-	<i>0,35</i>
b1824	<i>yobF</i>	+	0,45	0,37	-	-	-	<b>0,79</b>
b1852	<i>zwf</i>	+	0,79	0,93	-	-	Yes	<b>3,19</b>
b1901	<i>araF</i>	+	0,54	0,51	-	-	-	<b>0,90</b>
b2001	<i>yeeR</i>	+	0,60	0,06	1	CDS	-	<i>0,18</i>
b2094	<i>gatA</i>	-	1,62	1,19	2	5'UTR	Yes	<b>2,79</b>
b2095	<i>gatZ</i>	-	1,84	1,22	1	5'UTR	Yes	<b>3,06</b>
b2245	<i>yfaU</i>	+	0,57	0,82	-	-	-	<b>2,96</b>
b2246	<i>rhmT</i>	+	0,48	0,69	1	5'UTR	-	<b>2,02</b>
b2360	<i>yfdQ</i>	+	0,68	0,49	-	-	-	<i>0,55</i>
b2463	<i>maeB</i>	-	2,26	1,40	3	5'UTR / 5'UTR- CDS / CDS	Yes	<b>2,43</b>
b2521	<i>sseA</i>	-	1,32	3,10	1	5'UTR	-	<i>0,24</i>
b2542	<i>hcaD</i>	+	0,73	0,84	1	5'UTR	Yes	<b>1,84</b>
b2702	<i>srlA</i>	+	0,31	0,53	1	CDS	-	<b>1,83</b>



b2703	<i>srlE</i>	+	0,28	0,42	2	CDS	-	<b>1,47</b>
b2704	<i>srlB</i>	+	0,32	0,39	-	-	-	<b>1,21</b>
b2799	<i>fucO</i>	-	2,96	4,07	2	5'UTR / 5'UTR- CDS	-	<b>0,77</b>
b2937	<i>speB</i>	+	0,53	0,73	1	5'UTR	-	<b>2,02</b>
b2938	<i>speA</i>	-	1,49	1,19	1	CDS	Yes	<b>2,27</b>
b2989	<i>yghU</i>	+	0,59	0,65	-	-	-	<b>1,23</b>
b3008	<i>metC</i>	+	0,64	0,73	-	-	-	<b>1,41</b>
b3011	<i>yqhD</i>	+	0,50	0,72	1	5'UTR	-	<b>2,12</b>
b3067	<i>rpoD</i>	-	2,79	1,26	1	5'UTR	Yes	<b>4,37</b>
b3077	<i>ebgC</i>	+	0,57	0,66	-	-	-	<b>1,34</b>
b3078	<i>ygjI</i>	+	0,52	0,68	1	5'UTR	-	<b>1,69</b>
b3089	<i>sstT</i>	-	1,35	2,32	2	5'UTR	-	<b>0,36</b>
b3091	<i>uxaA</i>	+	0,66	0,80	-	-	-	<b>1,86</b>
b3092	<i>uxaC</i>	+	0,52	0,80	-	-	-	<b>2,93</b>
b3133	<i>agaV</i>	+	0,67	0,70	-	-	-	<b>1,10</b>
b3134	<i>agaW</i>	+	0,73	0,75	1	5'UTR	Yes	<b>1,12</b>
b3253	<i>yhdH</i>	+	0,69	0,65	1	CDS	Yes	<b>0,87</b>
b3310	<i>rplN</i>	-	1,29	1,13	1	5'UTR	Yes	<b>2,06</b>
b3313	<i>rplP</i>	-	2,10	1,17	2	5'UTR / CDS	-	<b>4,63</b>
b3316	<i>rpsS</i>	-	1,93	1,11	-	-	-	<b>6,10</b>
b3371	<i>friB</i>	+	0,59	0,67	1	CDS	-	<b>1,30</b>
b3403	<i>pck</i>	-	2,61	1,25	-	-	Yes	<b>4,35</b>
b3432	<i>glgB</i>	-	1,92	3,58	2	5'UTR / CDS	Yes	<b>0,51</b>
b3433	<i>asd</i>	-	1,47	1,84	-	-	Yes	<b>0,63</b>
b3460	<i>livJ</i>	+	0,20	0,48	3	5'UTR / 5'UTR- CDS / CDS	-	<b>2,20</b>
b3491	<i>yhiM</i>	+	0,57	0,76	2	5'UTR	-	<b>2,05</b>
b3774	<i>ilvC</i>	+	0,82	0,80	-	-	Yes	<b>0,88</b>
b3831	<i>udp</i>	+	0,51	0,63	-	-	-	<b>1,47</b>
b3870	<i>glnA</i>	-	1,71	1,33	-	-	Yes	<b>1,88</b>
b3919	<i>tpiA</i>	+	0,83	0,76	-	-	Yes	<b>0,71</b>
b3926	<i>glpK</i>	+	0,65	0,41	1	5'UTR	-	<b>0,49</b>
b3927	<i>glpF</i>	+	0,57	0,28	-	-	-	<b>0,44</b>
b3981	<i>secE</i>	-	1,34	1,35	1	CDS	Yes	<b>0,98</b>
b3990	<i>thiH</i>	+	0,55	0,87	1	CDS	-	<b>4,30</b>
b4034	<i>malE</i>	+	0,51	0,59	3	5'UTR	Yes	<b>1,26</b>
b4079	<i>fdhF</i>	+	0,75	0,65	-	-	-	<b>0,65</b>
b4088	<i>alsB</i>	+	0,66	0,62	-	-	-	<b>0,86</b>
b4120	<i>melB</i>	+	0,66	0,76	3	5'UTR / CDS	-	<b>1,53</b>
b4138	<i>dcuA</i>	-	1,99	4,33	1	5'UTR	Yes	<b>0,47</b>

## Annexes

---

b4211	<i>qorB</i>	+	0,67	0,83	-	-	-	<b>2,17</b>
b4216	<i>ytfJ</i>	+	0,54	0,66	-	-	Yes	<b>1,48</b>
b4234	<i>yjgA</i>	+	0,55	0,82	-	-	-	<b>2,99</b>
b4265	<i>idnT</i>	+	0,64	0,86	-	-	Yes	<b>2,96</b>
b4266	<i>idnO</i>	+	0,67	0,46	-	-	-	<i>0,51</i>
b4267	<i>idnD</i>	+	0,61	0,56	-	-	-	<b>0,86</b>
b4525	<i>ymjC</i>	+	0,59	0,87	1	5'UTR	-	<b>3,89</b>