Study of regulatory factors of translation in Escherichia coli

Huong Le Nguyen

Université Fédérale
Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

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

Présentée et soutenue par :
Huong Nguyen
le 19 Avril 2019

Titre :

Etude des facteurs régulateurs de la traduction chez Escherichia coli

---

École doctorale et discipline ou spécialité :
ED SEVAB : Ingénieries microbienne et enzymatique

Unité de recherche :
Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés (LISBP)

Directeur/trice(s) de Thèse :

Laurence GIRBAL
Muriel COCAIGN-BOUSQUET

Jury :
Alice LEBRETON (Rapportrice) - Chargée de recherche INRA, Paris
Matthieu JULES (Rapporteur) - Professeur à AgroParisTech, Paris
Eliane HAJNSDORF (Examinatrice) - Directrice de recherche CNRS, Paris
Laurence GIRBAL (Co-directrice) - Chargée de recherche CNRS, Toulouse
Muriel COCAIGN-BOUSQUET (Co-directrice) - Directrice de recherche INRA, Toulouse

Etude des facteurs régulateurs de la traduction chez *Escherichia coli*

L'expression génique chez les bactéries est finement régulée à chaque étape entre le gène et la protéine. L'analyse des régulations de l'expression des gènes permet de mieux comprendre l'adaptation des bactéries à leur environnement et dans un contexte de biologie de synthèse de pouvoir optimiser la production microbienne de molécules d'intérêt. Nous avons choisi la bactérie *Escherichia coli* pour étudier le système de régulations de l'expression génique car elle possède une importante flexibilité métabolique qui lui permet de s'adapter à différents substrats carbonés. Par ailleurs *E. coli* est aussi un hôte largement utilisé en biologie synthétique.

Le rôle de la régulation traductionnelle dans l'expression génique est de plus en plus étudié suite à l'observation de faibles corrélations entre les niveaux de transcription et de protéines. Plusieurs études moléculaires lors de l'adaptation métabolique ont démontré la présence de régulations de la machinerie de traduction ou de mécanismes de régulations spécifiques de la traduction de certains gènes. Cependant, des analyses à l'échelle globale de la traduction de l'ensemble des gènes sont encore peu nombreuses. Ainsi, l'objectif de cette thèse a été d'étudier la traduction au niveau du génome ainsi que ses relations avec les autres processus cellulaires par une approche de biologie des systèmes.

Tout d'abord l'activité de traduction à l'échelle omique (appelée le traductome) a été mesurée par le couplage des méthodes de polysome profiling et de RNA-Seq afin d'obtenir pour chacun des ARN messagers, son pourcentage de copies en traduction et sa densité en ribosomes. Pour la première fois, une image complète de l'état traductionnel de *E. coli* en croissance rapide a été obtenue, caractérisée par une majorité de transcrits avec un très fort pourcentage de copies en traduction mais faiblement chargés en ribosomes. Ces résultats couplés avec l'observation de corrélations positives entre le niveau de protéines et les deux paramètres, pourcentage de copies en traduction et densité en ribosomes ont suggéré que l'initiation serait la principale étape limitante de la traduction chez *E. coli* en croissance rapide.

Nous avons ensuite voulu prédire les déterminants majeurs de la variabilité de traduction entre les gènes chez *E. coli* en croissance rapide. Nous avons intégré nos données du traductome avec des données génomique, transcriptomique, de fonction et localisation des protéines, les cibles de régulations spécifiques… dans un modèle statistique de régression linéaire multiple. Notre modèle a identifié des facteurs liés à la séquence comme déterminants de la traduction mais de façon plus surprenante, le modèle prédit aussi le rôle important d'un paramètre physiologique : la concentration en ARNm. Ainsi plus un ARNm est concentré, plus il aurait un fort pourcentage de copies en traduction et une densité en ribosomes élevée. Pour la première fois, cet effet de la transcription sur la traduction a été validé à l'échelle moléculaire sur plusieurs gènes. Pour cela, nous avons développé une méthode de multiplexage de polysome profiling afin d'étudier la traduction de plusieurs souches en parallèle. Nous avons montré qu'une augmentation de la concentration d'un ARNm par induction transcriptionnelle entrainait une augmentation du pourcentage de copies en traduction et de la charge en ribosomes.

L'ensemble de nos résultats -omiques et moléculaires souligne le rôle crucial de l'état physiologique dans la régulation traductionnelle, la traduction étant régulée de façon co-directionnelle avec la transcription en fonction des besoins de la cellule.

Study of regulatory factors of translation in *Escherichia coli*

Gene expression in bacteria is finely regulated at each step between the gene and the protein. The analysis of gene expression regulation is necessary to better understand bacterial adaptation to environment and to be able in a context of synthetic biology to optimize the production of molecules of interest. We chose to study the regulatory system of gene expression in *Escherichia coli* because *E. coli* has a high metabolic flexibility and is capable to grow on different carbon substrates. Moreover, *E. coli* is also widely used as a host in synthetic biology.

The role of translational regulation in gene expression is increasingly studied after the discovery of weak correlations between transcription and protein levels. Several molecular studies during metabolic adaptation have demonstrated regulations of the translational machinery or specific regulatory mechanisms of translation of certain genes. However, analyses of global translation of all the genes are still very few. Thus, the goal of this thesis was to study translation at the genome-wide level and its relationship to other cellular processes using a systems biology approach.

First, translation activity at the -omic scale (called the traductome) was measured by coupling polysome profiling and RNA-Seq methods to provide for each messenger RNAs, its percentage of copies in translation and ribosome density. For the first time, a complete picture of the translational state in fast growing *E. coli* cells was obtained, characterized by a majority of transcripts with a very high percentage of copies in translation but a low ribosome density. These results coupled with the observation of positive correlations between protein level and both parameters, percentage of copies in translation and ribosome density, suggested that initiation would be the major limiting step of translation in rapidly growing *E. coli* cells.

We then wanted to predict the major determinants of translation variability among genes in rapidly growing *E. coli.* We integrated our traductome data with genomic, transcriptomic, functional and protein localization data, specific regulatory targets ... in a statistical model of multiple linear regression. Our model identified sequence-related factors as determinants of translation but, more surprisingly, the model predicted the important role of a physiological parameter: the mRNA concentration. Thus, more concentrated mRNA would have higher percentage of copies in translation and higher ribosome density. For the first time, this effect of transcription on translation has been validated at the molecular level on several genes. For this, we developed a polysome profiling multiplexing method to study the translation of several strains in parallel. We showed that an increase in mRNA concentration by transcriptional induction results in increases in percentage of copies in translation and in ribosome load.

All of our -omic and molecular results underline the crucial role of the physiological state in translation regulation, translation being co-directionally regulated with the transcription according to the cellular needs.

REMERCIEMENTS

Cette thèse a été pour moi une magnifique aventure, plonger dans la physiologie bactérienne et chercher à comprendre un petit nœud du fonctionnement de ce micro-organisme. La thèse m'a apporté des connaissances et savoir-faire scientifiques mais avant tout des relations humaines précieuses. Un grand grand merci à tous mes facteurs régulateurs positifs, les personnes qui ont permis le bon déroulement de la thèse.

Je souhaite remercier les membres du jury Alice Lebreton, Matthieu Jules et Eliane Hajnsdorf pour l'évaluation de ce travail et les discussions enrichissantes lors de la soutenance. Je remercie Isabelle Iost d'avoir été membre de mes comités de thèse et d'avoir apporté de précieuses remarques pour le projet.

Je profite de l'opportunité unique de pouvoir remercier du fond du cœur Muriel et Laurence de m'avoir accueillie au sein de votre équipe et surtout d'être mes directrices de thèse. Vous m'avez donné votre confiance et cette bourse de thèse pour laquelle vous vous étiez battues. A mon arrivée, la lecture des remerciements de Flora et Thomas sur votre duo légendaire m'a bien rassurée. Maintenant, après trois ans et demie, je reconfirme encore une fois tout ce qui a été dit. Je referais tout le même chemin pour arriver au même point qu'aujourd'hui et vous avoir comme directrices. Merci infiniment Laurence pour ta rigueur, ton écoute, ta patience inimaginable pour moi et tout ce que tu m'as appris dans le travail. Merci Muriel pour ta confiance, ton optimisme et tes idées excitantes pour creuser toujours plus la science. Vos qualités et ma gratitude dépassent les mots, je pense que mes vrais remerciements seront de m'améliorer, de tirer profit à fond de votre formation et de l'appliquer dans la réalisation de mes futurs projets.

Dans ce projet, j'adresse ma reconnaissance à mes collaborateurs qui sont aussi des membres formidables de l'EAD4. J'ai eu la chance de commencer dès mon arrivée par analyser des résultats, mais ce travail ne pourrait jamais pu être bien effectué sans les fondations statistiques solides mises en place par Sandrine. Je te remercie d'être le « problème solveur » de stats et R, toujours disponible pour répondre clairement aux questions. Par ailleurs, tu es la française la plus pro-asiatique que j'ai rencontrée, nos échanges sur la nourriture asiatique m'ont fait beaucoup plaisir ! Je remercie beaucoup Seb, mon point d'appui très rassurant en bio mol. Tu as été là pour les idées, le design et les questions à tous les moments. Je t'admire pour ta clarté incroyable, ton esprit synthétique lors des réunions et toujours de bons conseils pour les présentations. J'ai apprécié sincèrement ton aide pour la préparation de la soutenance. Enfin, Marie-Pierre, merci de m'avoir formée et accompagnée dans les expériences délicates de traductome. Les plus jolis résultats ont été obtenus par tes mains. Je te remercie de tout cœur pour ton écoute et tes encouragements précieux à chaque moment de mes excitations ou dépressions. Ta bienveillance m'aidera à prendre confiance en moi lors de mes prochains pas.

Merci à tous les membres de l'EAD4, vous formez une équipe « extra » ordinaire, dans le sens propre car il y a ici un enrichissement de bonne humeur, d'échanges, de stimulations et d'entraides. Merci Brice pour tes blagues lors des repas d'équipe et ton accueil amical à mon arrivée, merci Valérie pour le travail de gestion. Merci Marie-Line et Pascal pour votre passion pour la science tout en restant très cools. Parfois, je souhaiterais retourner en master pour vous avoir comme profs.

Je veux maintenant exprimer toute mon amitié aux jeunes qui ont rendu le labo si agréable tous les jours. Merci beaucoup Nuria et Virginie, votre arrivée était comme une pluie après la sécheresse et votre accompagnement a été pour moi une période très sympa et agréable. Merci Pierre et Anthony, pendant vos six mois de stage, nous avons eu des moments joyeux ensemble que j'ai vraiment appréciés. Merci Manon pour ta gentillesse et plein d'autres qualités scientifiques : clarté, organisation et détermination… Merci Marie-Aurore pour ton esprit sportif et ton enthousiasme pour le travail. Et Kevin, tu vois, je ne t'oublie certainement pas, le petit dernier de l'équipe et mon voisin de paillasse plein de bonne humeur. Je suis très contente que Pascal t'ait eu comme thésard et l'inverse, vous êtes un duo extra. Peut-être tu ne t'en rends pas compte mais tu nous apportes beaucoup d'énergie, d'enthousiasme et de positivité. Merci Marjorie de venir rejoindre la colonie de vacances et d'amener la bonne ambiance. Pour Charlotte, tu es dans de bonnes mains et je suis sûre que tu vas passer des moments formidables pendant ta thèse et t'épanouir dans le travail. Je vais garder de très beaux souvenirs de toi, de tes fiches et surtout de tes câlins, qui marchent mieux que des vitamines pour mon moral. Fan, c'est sympa d'avoir un autre asiatique dans l'équipe, je te souhaite beaucoup de courage et de réussite pour la thèse. Enfin, surtout je n'oublierai pas Théo, Pedro et Martine pour votre soutien à la fin de ma rédaction, votre gentillesse m'a beaucoup touchée. Pour tous les stagiaires, thésards et anciens de l'EAD4, gardez vos qualités et restez tels quels. Je me trouve chanceuse de vous avoir rencontrés et vous souhaite beaucoup de motivation, de meilleurs résultats pendant votre temps ici et de beaux projets pour la suite. J'espère vraiment que plusieurs parmi vous deviendront de grands scientifiques !

Je remercie toutes les personnes du LISBP que j'ai croisées pendant ces trois ans, pour les échanges, conversations et encouragements.

A cette occasion, je veux exprimer ma reconnaissance à l'association ASPF et M. Odon Vallet, des mécènes qui m'ont donné l'opportunité de faire des études en France. J'ai reçu également le soutien et les encouragements de mes directeurs de stage et enseignants de AgroParisTech pour effectuer une thèse. Ils m'ont donné le goût pour la science et la confiance en moi, je remercie chaleureusement Françoise Rul, Michel Popoff, Aurélie Baliarda et Colin Tinsley.

Enfin, je tiens à remercier toutes mes proches pour leur accompagnement continu durant toutes ces années. Merci à mes amies qui me soutiennent de loin et particulièrement à Anh. Nous avons passé la thèse ensemble et pendant trois ans, tu as écouté avec patience toutes mes histoires quotidiennes, mes découvertes, mes joies ainsi que mes démotivations et mes doutes. Grâce à toi, ces trois ans se sont passés avec beaucoup de bonheur. Merci la compagnie précieuse de Tam pendant les derniers mois assez tendus. Merci sincèrement Anne et Louis, sans vous je ne serais jamais arrivée jusque-là. Vous ne m'avez jamais lâchée, vous m'avez suivie, accompagnée et poussée pendant neuf ans avec des conseils toujours justes et au bon moment.

Merci à mon frère qui m'a apporté de l'énergie et de la motivation. Je remercie infiniment mes parents, qui sont toujours des exemples pour moi dans la vie. Vous m'avez toujours encouragée et soutenue et vous m'avez donné tous les moyens pour que je puisse faire les études que j'ai souhaitées.

Merci à tous d'être venus dans ma vie !

SOMMAIRE

INTRODUCTION

La bactérie *Escherichia coli* est la bactérie modèle des bactéries à Gram négative en microbiologie. Elle est caractérisée par un temps de génération rapide en condition favorable et par une importante flexibilité métabolique lui permettant de se développer sur des différents substrats carbonés. Sa grande faculté d'adaptation aux fluctuations de l'environnement est permise par un réseau complexe de régulations de l'expression des gènes. L'expression des gènes comprend plusieurs étapes menant du gène à la protéine correspondante dont chaque niveau est rigoureusement contrôlé. Les régulations peuvent être transcriptionnelles (de la transcription), post-transcriptionnelles (de la dégradation des ARNm et de la traduction) et post-traductionnelle (de la dégradation des protéines et de leur activité). Les connaissances fondamentales sur l'ensemble de ces régulations permettraient de comprendre l'adaptation des bactéries à leur environnement. Par ailleurs, dans un contexte de biologie de synthèse, cette compréhension de la globalité du système de régulation de l'expression des gènes permettrait de mieux contrôler les processus métaboliques et d'optimiser la production de molécules d'intérêt.

Dans ce réseau complexe, la régulation de la traduction constitue un nœud crucial. Etant le processus le plus coûteux en énergie dans la cellule, la traduction doit être contrôlée précisément, pour augmenter l'efficacité de la cellule. Une attention croissante est portée au rôle important de la traduction dans la régulation de la synthèse des protéines, en particulier lorsque des variations de la concentration en ARNm seule ne sont pas suffisantes pour expliquer des variations du niveau de protéines. Les régulations de la traduction peuvent faire intervenir des acteurs extrinsèques (les petits ARNs, les protéines se fixant à l'ARN, les métabolites) et sont souvent en lien avec des caractéristiques intrinsèques de la séquence de l'ARNm (comme la force du RBS, la structure secondaire, l'usage de codon…) qui modulent la fréquence d'initiation et la vitesse d'élongation de la traduction. Malgré la multiplication d'exemples de mécanismes de régulations spécifiques de la traduction, l'image au niveau cellulaire des régulations de la traduction reste à préciser. Pour cela, l'étude de la traduction ne devrait pas être abordée de façon indépendante pour chaque ARN mais pour l'ensemble des ARNm de la cellule et intégrée dans le réseau complexe de régulations de l'expression génique.

Dans ce contexte, cette thèse se distingue par une approche de biologie des systèmes pour appréhender les régulations de la traduction chez *E. coli*. Ce travail

étudie les variables de la traduction à l'échelle du génome (le traductome) chez *E. coli* en phase exponentielle de croissance, dans son état physiologique optimal. Pour l'ensemble des ARNm de la cellule, le pourcentage de copies en traduction (copies associées avec les ribosomes) et la densité en ribosomes ont été mesurés. Pour la première fois, ces données de l'état de traduction à l'échelle globale ont été obtenues par la méthode de polysome profiling chez cette bactérie. L'intégration de ces résultats avec d'autres données -omiques (transcriptome, stabilome, protéome…) ont permis d'explorer la place de la régulation traductionnelle dans la synthèse protéique finale, et ses interactions avec les autres processus cellulaires. Des règles générales de régulation de la traduction ont ainsi été identifiées. Jouer sur ces régulations de la traduction pourrait constituer de nouvelles voies d'amélioration de la production de protéines d'intérêt chez *E. coli.*

Ce manuscrit débute par une introduction bibliographique qui fournit des informations nécessaires pour la compréhension de ces travaux. Après une présentation générale d'*E. coli* et de la régulation de l'expression génique, les étapes de la traduction sont décrites. Ensuite, les principales régulations traductionnelles sont présentées, celles de la traduction de manière globale et celles dépendantes de chaque transcrit. Un focus sur le rôle central de la traduction par son couplage avec les autres processus cellulaires sera aussi présenté. Enfin, cette étude bibliographique se termine par une présentation des méthodes pour étudier la traduction à l'échelle -omique et les objectifs de la thèse.

Les trois chapitres de résultats obtenus lors de cette thèse sont présentés en anglais, dont deux sous forme d'articles scientifiques (chapitres I et III), précédés d'un résumé en français. Ces trois chapitres contribuent à l'objectif de la thèse qui est de décrire l'état traductionnel global chez *E. coli* en croissance rapide et d'identifier les principaux facteurs qui le régulent. Pour cela, les deux premiers chapitres décrivent des méthodologies développées au cours de la thèse. Dans le chapitre I, le développement d'une méthode de multiplexage de polysome profiling est présenté. Cette méthode permet d'obtenir l'état traductionnel de plusieurs ARNm cibles en parallèle et elle est utilisée pour valider l'effet de la concentration en ARNm sur la traduction dans le chapitre III. Dans le chapitre II, des méthodes d'alignement de séquences et de normalisation sont comparées pour analyser les données de RNA-Seq dans le cadre de l'étude de la traduction. Dans le chapitre III, l'état traductionnel de

l'ensemble des ARNm chez *E. coli* est décrit. Des expériences de polysome profiling couplé à la quantification des ARN par RNA-Seq permettent l'estimation du pourcentage de copies en traduction et la densité en ribosomes de chaque ARNm. Des approches statistiques et moléculaires démontrent dans ce chapitre le rôle clé de la concentration en ARNm dans la régulation de la traduction.

Enfin, la conclusion reprend l'ensemble des résultats, discute le rôle de la traduction ainsi que de ses déterminants, en particulier l'effet du niveau de l'ARNm. Cette dernière partie s'ouvrira sur différentes perspectives à plus ou moins long terme.

ETUDE BIBLIOGRAPHIQUE

I) Escherichia Coli

I.1) Overall presentation

Our bacterial model *Escherichia Coli* is a Gram-negative Enterobacteriaceae of about few µm in length. It was first isolated by Theodor Escherich in 1885 while studying infant gut microbes. It is found primarily in mammalian gut and less commonly in water, plants and food. *E. coli* has the ability to catabolize different nutrients due to its diverse niches. In gut, *E. coli* is located at the thin mucus layer surrounding epithelium cells. It has important metabolism capacity to use different substrates from this rich food supply of mucus (e.g gluconate, N-acetylglucosamine, fucose, deoxyribose, ribose, galactose, maltose…) (Chang et al., 2004). *E. coli* is mainly aerobe but it is also a facultative anaerobic bacterium that can grow in absence of oxygen (as it is the major case in gut). It consumes oxygen and promotes anoxic environment for other anaerobic bacteria. This bacillus cannot sporulate nor grow in extreme temperature or pH. The majority of *E. coli* strains are non-pathogenic, but some can cause diarrhea and urinary infection. We note the dangerous enterohemorrhagic strain *E. coli* O157:H7 which produces shiga-toxin. *E. coli* has an important genome flexibility and its pathogenicity could be acquired by horizontal transfer (Blount, 2015).

I.2) Utilization of E. coli in biotechnology

E. coli stands out as a model organism largely used in academic research to study mechanism of many fundamental biological processes. Important findings in molecular biology and other basic aspects of life, among them Nobel prizes were elucidated with *E. coli*. We can account the discoveries of genetic code (Crick et al., 1961), mechanism of DNA replication (Lehman et al., 1958) transcription (Stevens, 1960), gene regulation and organization via operon (Jacob and Monod, 1961), restriction enzymes (Linn and Arber, 1968; Meselson and Yuan, 1968), structure and function of ATP synthase (Capaldi et al., 2000), etc. These fundamental findings allowed then diverse biotechnology applications, such as transcriptional regulation, genetic engineering technologies using restriction enzymes, molecular cloning and recombinant DNA (Cohen et al., 1973).

As *E. coli* has a well-studied and relatively simple molecular machinery, and is easily transformable and modified, it is involved in production of recombinant therapeutic proteins: insulin, interleukin-2, human interferon-B, human growth

hormone, taxol, pegloticase, cerolizumab, etc (Reviewed in (Huang et al., 2012; Kamionka, 2011) which presents huge economic interest. Moreover, *E. coli* is used as a host to synthetize biofuels and industrial chemicals such as phenol, ethanol, mannitol. Its rapid growth in cheap culture media and absence of aggregation provide additional advantages in industrial production of these molecules. All implications of *E. coli* in pharmaceutical production and biotechnology industry contributed to 500 billion to the global economy in 2011 (Blount, 2015).

I.3) Strain E. coli K-12 MG1655

The strain K-12 MG1655 is a non-pathogenic strain derived from the ancestral K-12 strain and widely used in basic research studies. It was the first *E. coli* strain which has its complete genome sequenced (Blattner et al., 1997). Its genome size is 4.639.221 bp with a GC content of 50.8%. It contains around 4500 genes with around 4300 genes coding for proteins (Blattner et al., 1997; Karp et al., 2007; Riley et al., 2006). Among them, around 300 genes are suggested to be essential (Baba et al., 2006).

Genetic, genomic information and biochemical features such as chromosomal position, gene sequence, transcriptional promoters, transcriptional start site, organization in operons, gene regulation, protein function, category, metabolic pathway in *E. coli* K-12 are available in several databases: EcoCyc, RegulonDB, Kegg, PortEco….

Moreover, in recent years, increasing number of studies in systems biology: transcriptome, stabilome, proteome and metabolome have been performed in *E. coli*, providing extensive genome-wide knowledge. Its small genome and relatively simple metabolism make this bacterium a model to study at the genome scale metabolic adaptation to new environments (Valgepea et al., 2010) or protein synthesis regulation (Guimaraes et al., 2014; Esquerré et al., 2015).

The high interest of *E. coli* in biotechnology and the massive information about its biological functioning led us to use it as our model organism. In the context of systems biology, we would like to understand how translation is regulated at the genome-wide scale, in relationships with transcription, mRNA degradation and protein synthesis, using transcriptome, stabilome and proteome data already available in our team (Esquerré et al., 2014, 2015).

II) The global process of gene expression

II.1) The main steps of the gene expression process

Gene expression consists in cellular processes from gene to protein (Fig. 1). Which proteins are present, and their precise amounts are crucial for cell performance in terms of growth, adaptation and survival.

The first step of gene expression is transcription, the process of RNA synthesis from DNA (Fig. 1). Catalyzed by RNA polymerase, each gene in double-stranded DNA is transcribed to numerous copies of single-stranded RNA, in which the deoxyribose sugar is replaced by ribose and the thymine base by uracil. The three steps of transcription are initiation, elongation and termination. Briefly, during transcription initiation, RNA polymerase binds to the promoter located at -35 and -10 nucleotides upstream of the transcriptional start site and separates the DNA strands, providing the single-stranded template needed for transcription. During elongation, the RNA polymerase "reads" the template strand one base at a time and builds from 5' to 3' an RNA molecule out of the complementary nucleotides. Once the terminator signals are transcribed, they cause the transcript to be released from the RNA polymerase.

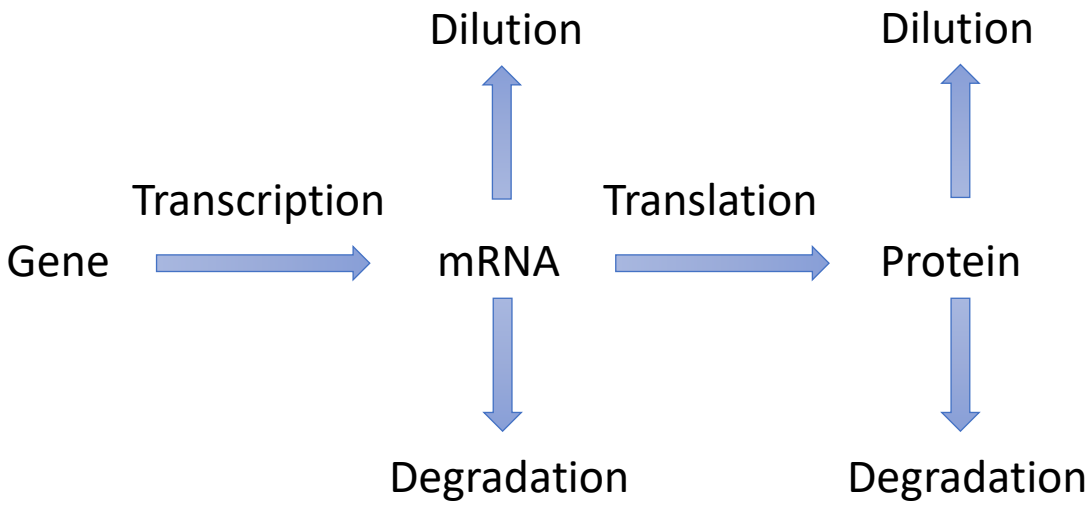


Figure 1: Schematic process of gene expression.

Second, translation occurs to synthetize a polypeptide chain (a protein) from the mRNA template. Translation is carried out by ribosomes. In prokaryotes, the ribosome is composed of two subunits, the small 30S subunit and the large 50S subunit, leading to a sedimentation coefficient of 70S for the assembled ribosome. Translation

is composed of three steps, initiation, elongation and termination. The process of translation will be described in more details in section III.

In time, the mRNA concentration varies because of mRNA degradation (Fig. 1). In prokaryotes, and particularly in *E. coli*, the average mRNA half-life is relatively short, about few minutes (Bernstein et al., 2002). The degradosome is a large protein complex which catalyzes most of mRNA degradation in *E. coli* (Carpousis, 2007). The main components of *E. coli* degradosome are the RNase E, the polynucleotide phosphorylase (PNPase), the RNA helicase B (RhlB) and the enolase. RNase E is an endoribonuclease that preferentially cleaves single-stranded A/U rich regions, PNPase is a 3'- 5' exoribonuclease and RhlB unwinds RNA. The role of enolase is not yet well defined. In *E. coli*, initiation of mRNA degradation starts with the endonucleolytic cleavage by the RNase E (Bouvier and Carpousis, 2011).

As the mRNA molecule, protein can also be degraded over time (Fig. 1). However, proteins have relatively longer stability than mRNAs with half-lives in the range of hours. Protein degradation is mediated by protein complexes usually composed of a protein with an ATPase activity (from the AAA family such as ClpA, ClpX andt ClpY) and a protease for the proteolysis (for example ClpP and ClpQ) (Chandu and Nandi, 2004). The role of the ATPase is to unfold the protein via conformational changes driven by ATP hydrolysis and to address the unstructured protein to the protease. Then the protease degrades the protein into smaller peptides and, finally, into amino acids, which are recycled into the cellular pool. Only the ClpAP, ClpXP et ClpYQ associations seem to be present in *E. coli* (Picard et al., 2009). In some cases (like for Lon and FtsH) the enzymes harbor in the same polypeptide the two domains conferring the ATPase and proteolysis activities.

At last, both the mRNAs and proteins are diluted over time as cells divide (Fig.1). The rate of dilution of mRNAs and proteins due to cellular growth depends on the generation time of the microbial cells.

II.2) Regulations of gene expression

Gene expression is constantly regulated in response to environmental changes, (e.g. nutrient availability, physico-chemical conditions, etc). In prokaryotes, transcriptional regulation has been thought for many years as the main regulation of

protein synthesis. Transcription regulation involves both global and specific regulations (Berthoumieux et al., 2014; Gerosa et al., 2014). An example of global regulation that affects the transcription of all the genes, is the increase of the RNA polymerase quantity and activity with the growth rate (Ehrenberg et al., 2013; Klumpp and Hwa, 2008; Klumpp et al., 2009). Specific regulation affects gene expression of a rather limited number of genes via transcriptional factors binding to specific promoters (Martínez-Antonio, 2011). The transcriptional factors act either as enhancer, repressor or inducer of transcription initiation in response to environmental changes.

However, it was shown that post-transcriptional regulations are also involved in mRNA level regulation. mRNA stability differs between *E. coli* transcripts depending on sequence determinants such as 5'end secondary structure (Arnold et al., 1998) or codon bias (Esquerré et al., 2015). The mRNA half-life is also dependent on growth conditions, it increases in *E. coli* when the growth rate decreases (Chen et al., 2015; Esquerré et al., 2014). Concerning the dilution of the mRNA molecules due to cell division, this post-transcriptional process is often neglected in *E. coli* because the generation time (in the range of hours) is generally much longer than most of the mRNA half-lives (in the range of minutes).

Variation in mRNA level is however not sufficient to explain protein level variability. Protein level is not totally correlated with the mRNA abundance in bacteria. Correlation coefficients between 0.2 and 0.6 were reported between the proteome and transcriptome data in in *E. coli* (Lu et al., 2007), *Lactococcus lactis* (Dressaire et al., 2009), and *Desulfovibrio vulgaris* (Nie et al., 2006). Taniguchi and colleagues confirmed in *E. coli* at the single cell level this low correlation (Taniguchi et al., 2010). Therefore, the role of translational regulations in gene expression regulation was largely suspected.

II.3) Stochasticity of gene expression

Stochastic gene expression is one reason of phenotype heterogeneity between genetically identical cells. Variability of gene expression between cells is manifested at the levels of transcription and translation. Taniguchi and colleagues performed a fluorescent reporter library to measure transcriptome and proteome in *E. coli* and showed different mRNA and protein copy numbers between isogenic cells (Taniguchi

et al., 2010). The heterogeneity of proteome, which was not correlated to transcriptome, suggests the existence of translation variation between cells. Measurement of translatome at the single-cell level could be performed to quantify this phenomenon. Transcription and translation, like other biological processes, are inherently stochastic based on the random encounter between molecules (for example between RNA polymerase and gene promoter for transcription or ribosome and mRNA for translation). This probability of encounter depends on the numbers of molecules (RNA polymerase, ribosomes, mRNAs) which could be different between cells.

Stochasticity of gene expression may occur at another level: within the same cell. Indeed, the translational levels of mRNA copies of the same transcript could be different. Experiments using fluorescence method to monitor translation of single molecule in a single human cell (Morisaki et al., 2014) showed that there was a variability of translation activity between mRNA copies of one mRNA. Within the same cell, heterogeneity may come from the stochastic diffusion of molecules. The local concentrations of ribosomes and mRNAs are crucial factors for the random encounter between mRNAs and ribosomes during initiation. Ribosomes and mRNAs are not homogeneously localized within the cell: polysomes are enriched at the polar zones, spatially excluded from the nucleoid (Bakshi et al., 2012, 2015) whereas mRNAs are synthetized at the nucleoid. In bacteria, mRNAs could be present at a very low copy number and sparsely distributed in the cell, the main source of variability according to some models (Kaern et al., 2005). Moreover, cells are crowed with macromolecules that could limit the diffusion and random encounter between mRNAs and ribosomes (Klumpp et al., 2013).

In the next sections, we will describe the process of translation (III), its regulation (IV) and the interactions of translation with other cellular processes (V).

III) How does translation work?

III.1) Ribosome structure

The principal actor of the translational machinery is the ribosome, a ribonucleoprotein complex composed of around 60% of ribosomal RNA (rRNA) and 40% of proteins. Ribosome is composed of two subunits: the large one is constituted from two rRNAs, 5S and 23S (116 nts and 2900 nts respectively) and 33 ribosomal proteins (L1 to L33). The small subunit includes the 16s rRNA and the 21 ribosomal proteins (S1 to S21) (Fig. 2).

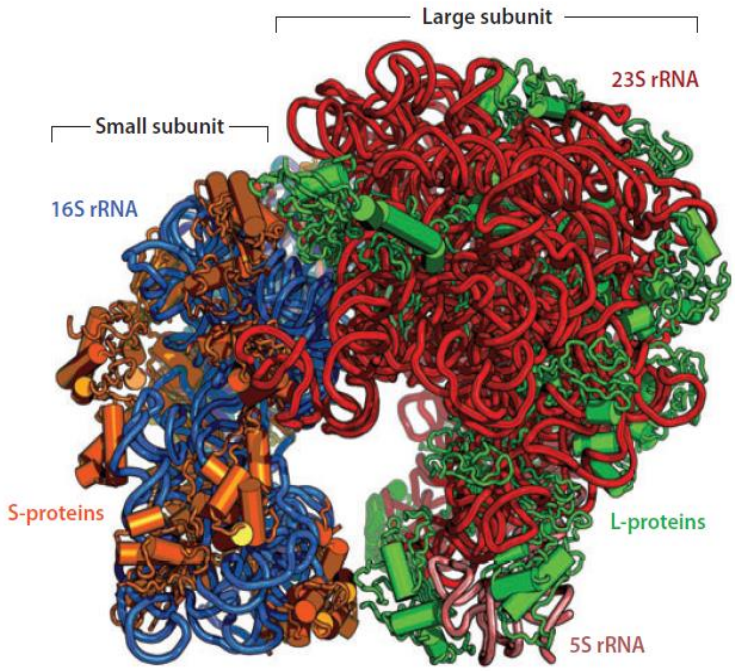


Figure 2: Structure of *E. coli* ribosome. The small subunit is shown in the left with the 16S rRNA in blue and the small (S) ribosomal proteins in orange. The large subunit is in the right with the 5S and 23S rRNAs in red and the large (L)-proteins in green (Shajani et al., 2011).

The small 30S subunit assures the ribosome binding to mRNA and controls translation fidelity via precise binding between mRNA and tRNA. The large 50S subunit catalyzes the peptide bond formation. The tRNA makes the connection between the mRNA sequence read by nucleotide triplet (corresponding to a unit named codon) and the neosynthesized peptide. One charged tRNA comprises at one end the anticodon, a sequence complementary to the codon read on the mRNA and at the other end the

corresponding amino acid. Note that there is a tRNA specific to the start codon: noted fMet-tRNA, this initiator tRNA is loaded with a formylated methionine amino acid.

In *E. coli*, genes coding for rRNAs are organized in operons and present in 7 copies. Each *rrn* operon contains 3 genes coding for the 5S, 16S and 23S rRNAs. These operons are under the control of two promoters P1 and P2. The two promoters are positively regulated by the transcription factor Fis and negatively regulated by the H-NS factor. The concentrations of Fis and H-NS depend on the growth rate with a low Fis concentration at the low growth rate or on the onset of the stationary phase whereas H-NS level increases in stationary phase (Schneider et al., 2003).

The translation process comprises three steps: initiation, elongation and termination shown in the figure 3 and detailed below.

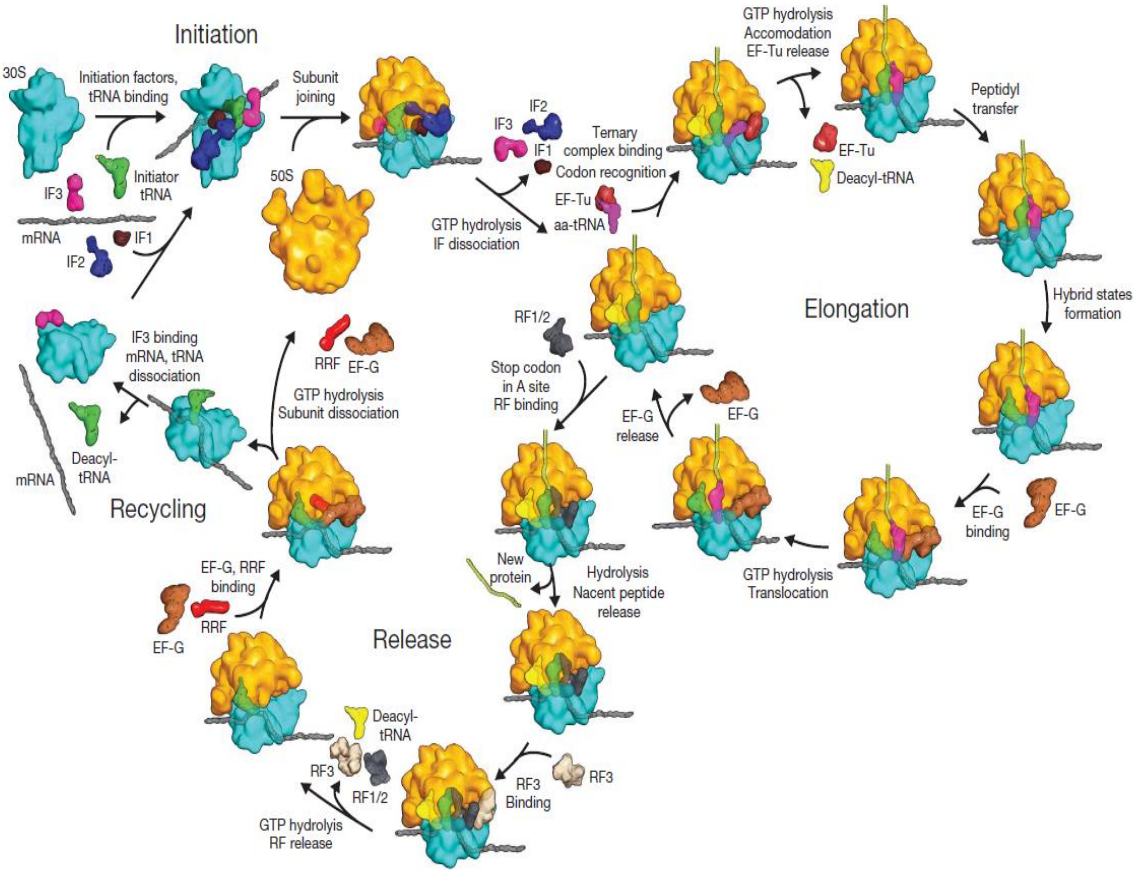


Figure 3: The different steps of translation in prokaryotes (Schmeing and Ramakrishnan, 2009).

III.2) Translation steps in prokaryotes

III.2.1) Initiation

Initiation is the process of assembling the two ribosomal subunits on the mRNA near the start codon (Fig. 4). The first step of initiation consists in the fixation of the 30S subunit at the correct location on the mRNA via the interaction of the Shine Dalgarno (SD) sequence (Shine and Dalgarno, 1974) with its complementary sequence (anti-SD) located at the 3'end of the 16S rRNA. The ribosome binding site (RBS) is purine rich, of 4-9 pb length and located at around 10 nucleotides upstream of the start codon. This alignment ensures that the start codon is located at the right position within the ribosome (at the P site). A consensus SD sequence is found for each species, in *E. coli*, it is GGAGG (Shine and Dalgarno, 1974), since the 3'end of 16S rRNA is 3'- CCUCCAC-5'. The RBS strength, which means the degree of complementary SD / anti-SD is largely studied and is thought to have a critical role on the initiation rate (Chang et al., 2006).

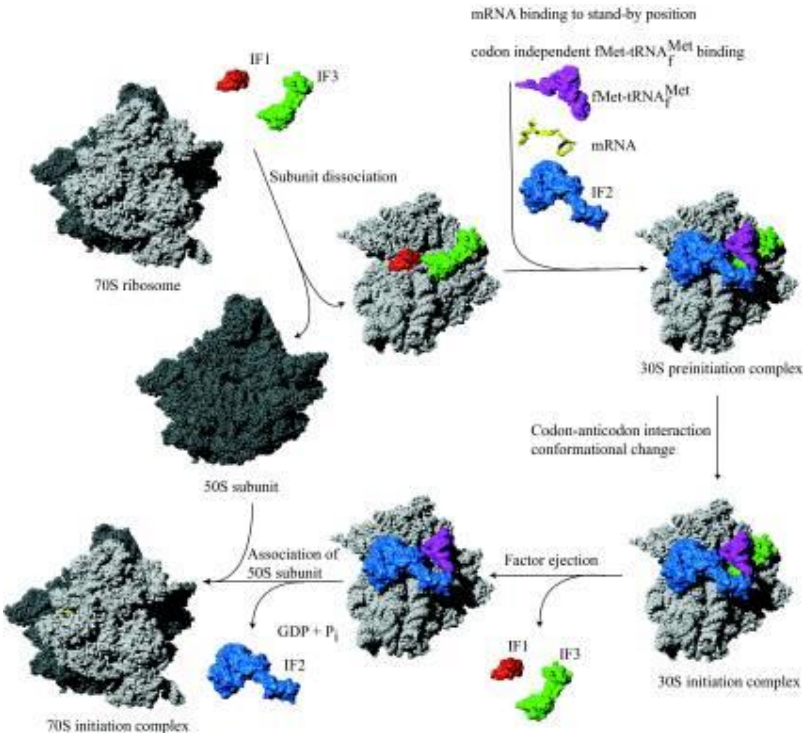


Figure 4: Steps of translation initiation in bacteria (Laursen et al., 2005).

The formation of the 30S pre-initiation complex (30 PIC) and the correct binding of the initiator fMet-tRNA on the start codon involve three initiation factors. IF1 and IF3 help to stabilize the 30S subunit while IF2 brings fMet-tRNA and mRNA together to form the 30 PIC. IF3 checks the correct position of fMet-tRNA at the P site of the 30S subunit and ensures the fidelity of initiation while IF2 hydrolyzes GTP.

Assembling of the 50S subunit occurs after release of the three initiation factors. Some "leaderless" mRNAs lack a 5'UTR region and therefore an RBS. For theses mRNAs, the ribosome already comprising fMet-tRNA binds directly to the mRNA to initiate translation (Moll et al., 2002; Udagawa et al., 2004).

The initiation efficiency depends on 5'UTR sequence-related parameters such as secondary structure, like riboswitches, SD sequences, the bindings of small RNAs and proteins as described in section IV.

III.2.2) Elongation

Once initiation is completed, elongation of the polypeptide chain consists in a cycle of reactions where the cognate aminoacyl-tRNA (aa-tRNA) corresponding to the present codon is recruited. The corresponding amino acid is added to the nascent polypeptide chain by the formation of a peptide bond between adjacent amino acids and then the ribosome translocates to the next codon (Fig. 5). Right after the initiation, the initiator fMet-tRNA is located directly at the P site of the ribosome, the A site is thus free for access of the aa-tRNA corresponding to the next codon. For each cycle, tRNA proceeds through the three functional sites of the ribosome: A (Aminoacyl tRNA binding), P (Peptidyl site) and E (Exit site).

Elongation requires three elongation factors EF-Tu, EF-G and EF-T that act as cofactors during the different stages of the elongation cycle and hydrolyzes 2 GTP for energy (Fig. 5). Recruitment of the aa-tRNAs occurs with the help of the elongation factor EF-Tu. Together with the aa-tRNA and GTP, EF-Tu forms a complex and binds to the A site. GTP is only hydrolyzed when the anticodon matches the codon. This quality control is ensured by EF-Tu. After this quality control, EF-Tu is released and recycled.

The formation of the peptide bond between adjacent amino acids at the P and A sites is catalyzed by a peptidyl transferase, a catalytic RNA ribozyme, located in the 50S ribosomal subunit. After the formation of the peptide bond, the amino acid is released from the tRNA at the P site and the nascent polypeptide chain is transferred to the tRNA in the A site. Then, the ribosome moves 3 nucleotides forward to the next codon with the help of the EF-G factor. Hydrolyzing of a GTP is required for this ribosome translocation. During translocation the tRNA without an amino acid moves from the P site to the E site. The tRNA with the nascent polypeptide chain that was in the A site is

moved to the P site. The uncharged tRNA leaves the E site and the A site is again open for the recruitment of the new aa-tRNA corresponding to next codon. The peptidyl transferase activity is controlled by the 23S rRNA and the L2 and L3 proteins.

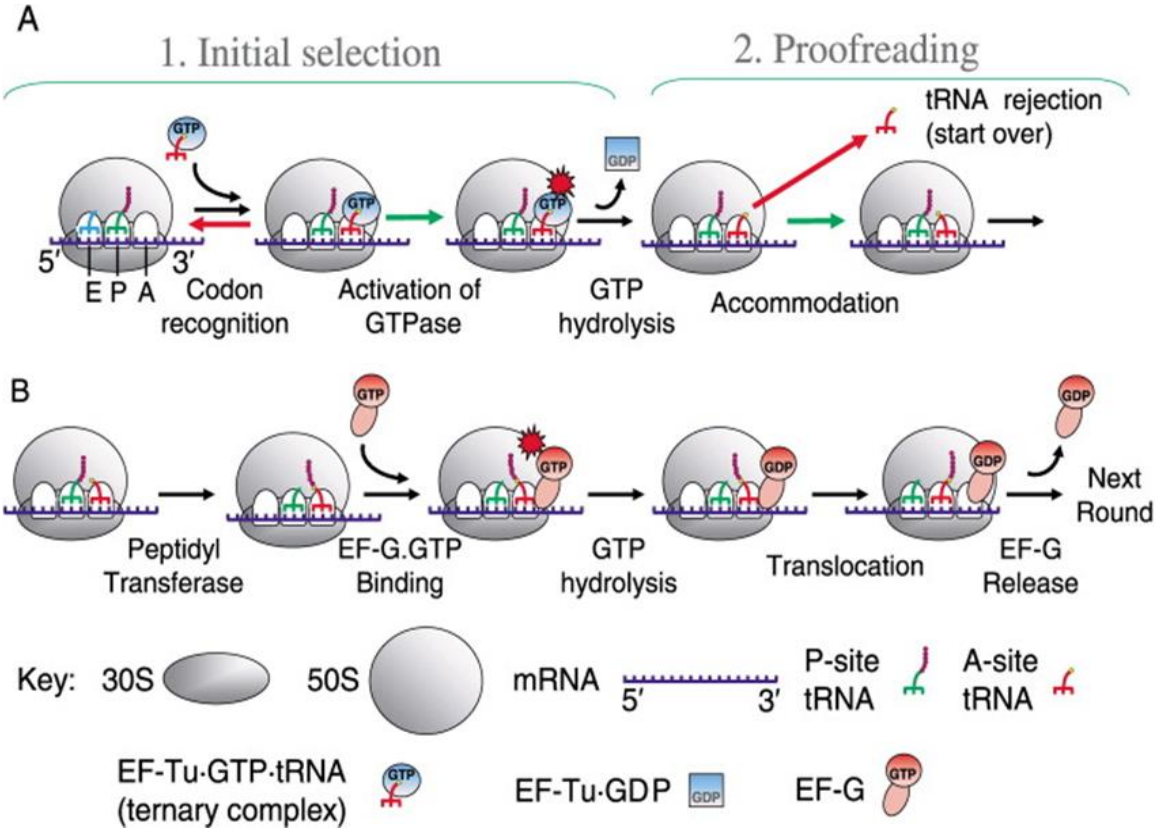


Figure 5: Steps of translation elongation in bacteria (Ramakrishnan, 2002).

During elongation, some sequence features can lead to ribosome pausing. Ribosome stalling was demonstrated at anti-SD like motifs in the coding sequence (Li et al., 2012) and after translation of consecutive prolines (partially compensated by the elongation factor P) (Qi et al., 2018). Moreover, the presence of rare codons, low tRNA pool, or unfavorable secondary structure can also slow down translation elongation as detailed in section IV.

III.2.3) Termination

The termination of translation occurs when the ribosome encounters a nonsense codon (UAA, UAG, or UGA) for which there is no complementary tRNA. Termination is mediated by three release factors. RF1 recognizes UAA and UAG, RF2 recognizes UAA and UAG codons and RF3 is a GTP-binding protein that facilitates the

binding/unbinding of RF1 and RF2. When a nonsense codon is in the A site, RF1 or RF2 hydrolyze the ester bond between the amino acid and the tRNA at the P site. The nascent polypeptide is released, and the two ribosomal subunits dissociate from the mRNA (Fig. 3). The two ribosomal subunits are recycled for another initiation. A global analysis of termination in *E. coli* reveals that the majority of terminations is accurate but there is a number of genes with read-through and frameshifting translation (Baggett et al., 2017).

III.2.4) The rate-limiting step of translation

For most endogenous genes under optimal growth condition, it is assumed that initiation is the rate-limiting step that controls protein synthesis (Bulmer, 1991; Plotkin and Kudla, 2011; Rocha et al., 1999; Salis et al., 2009). However, some studies proposed that initiation is not the only rate-limiting step for all the genes and in all the conditions (Hu et al., 2015; Plotkin and Kudla, 2011; Zouridis and Hatzimanikatis, 2007). Elongation can also be involved in translation limitation and an elongation limitation will lead to a lower protein synthesis (Fig. 6a).

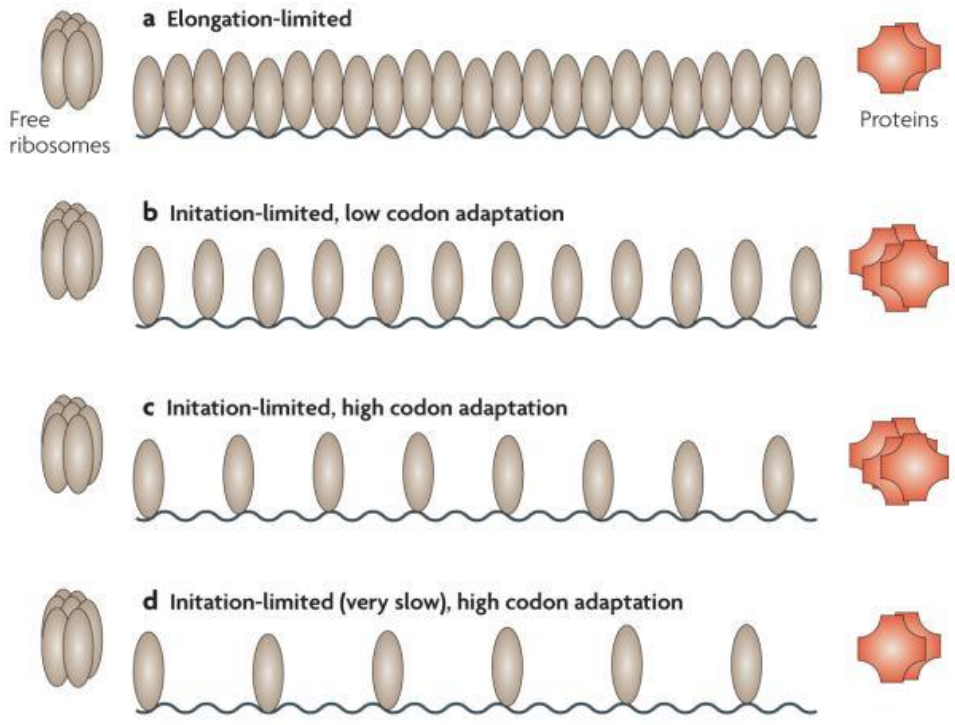


Figure 6: Different scenarios of translation rate-limiting steps and effect on the protein level (Plotkin and Kudla, 2011).

Studies that found elongation related factors (e.g. codon usage bias) as major determinants of translation comparing to initiation related factors (e.g. SD sequence) proposed that elongation is the rate-limiting step (Lithwick, 2003). In condition of low amino acid supply, elongation was also reported to limit translation (Hu et al., 2015). Depending on the ribosome density, the limiting step is modeled to be initiation and/or elongation or termination (Zouridis and Hatzimanikatis, 2007). Translation is predicted to be initiation-limited at low ribosome density, while elongation becomes also a limitation at higher ribosome density. Translation can be therefore limited by both initiation and elongation in the same time. Termination limitation is predicted to occur at extremely high ribosome density and therefore to be very rare.

IV) Translation regulations

Translation regulations can be divided in two types: global regulation that simultaneously affects all the mRNAs and transcript-specific regulation that affects only a subset of mRNAs. Figure 7 summarizes the global and specific translation regulations described in the following paragraphs.

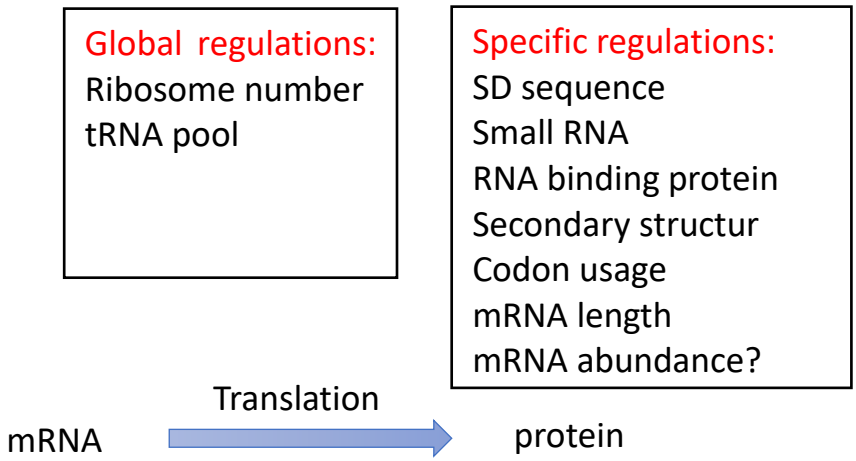


Figure 7: Examples of global and transcript-specific regulations of translation.

IV.1) Global regulations

Variation in nutrient availability, environmental parameters (pH, temperature) or exposure to a stress molecule (antibiotics) can have an effect on growth rate and reduce the protein demand. Regulation of the translation machinery, for example via changes in the number of ribosomes and tRNA pools, is hence a way to control the global process of translation in response to the cell demand.

IV.1.1) Regulation of ribosome number

The number of ribosomes varies proportionally with the growth rate when *E. coli* cells grow on different media (Bremer and Dennis, 1996). The number of ribosomal proteins (r-proteins) can be adjusted via a retrocontrol. When translation activity is down-regulated, the ribosomal proteins not rapidly incorporated into assembling ribosomes are in excess and then repress their own synthesis. One example is the auto-regulation of the ribosomal proteins L1/L11 by playing on the translation of the operon *rplA-rplK* which codes for these proteins. Upon high level of L1, L1 plays as repressor and binds to the stem loop structure near the SD site of its own mRNA, preventing the binding of the 30S subunit and hence the translation of the *rplA-rplK* transcript (Yates and Nomura, 1981). Translation inhibition by L1 seems to activate the degradation of *rplA-rplK* mRNA (Cole and Nomura, 1986).

Another mechanism of ribosome number regulation is the control of the ribosomal protein synthesis via a feedback loop in which guanosine tetraphosphate (ppGpp) acts as a feedback signal (Fig. 8). This strategy is a response to the available resources. When amino acid concentration drops (at late growth or during depletion following a high-level synthesis of a heterologous protein), the protein synthesis rate per ribosome declines and the production of the small sensor molecule ppGpp via two synthetases SpoT and RelA is activated. ppGpp accumulates, binds to the promoter P1 of the *rrn* operons turning off rRNA synthesis and thus reducing de novo ribosome synthesis. As a consequence, the number of ribosomes decreases resulting in a lower rate of amino acid consumption (Bremer and Dennis, 2008).

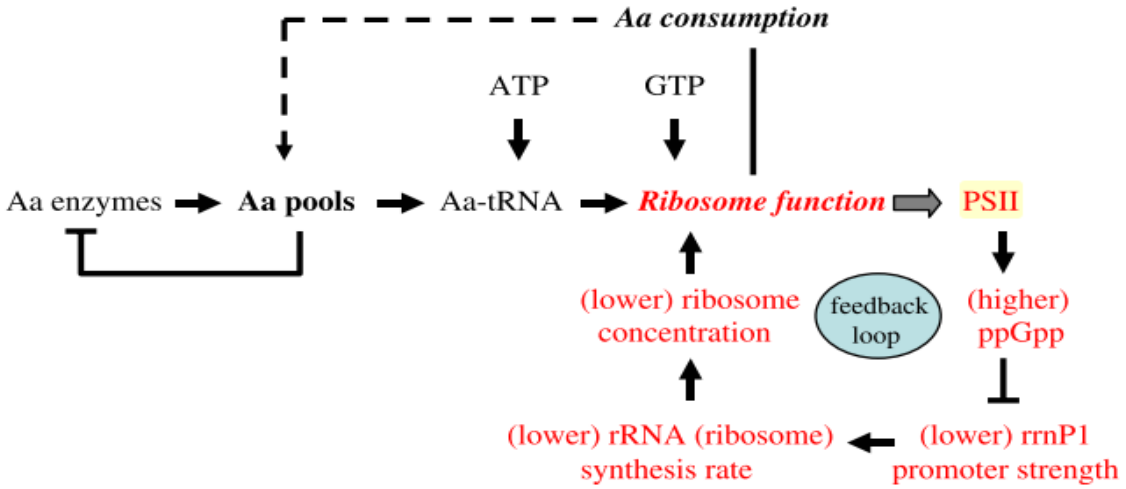


Figure 8: Feedback loop of ribosome synthesis regulation (Bremer and Dennis, 2008).

IV.1.2) Regulation of the tRNA pool

Beside rRNA, tRNAs are also dynamically regulated according to cell demand in different cellular states or growth conditions. The tRNAs cognate to abundant codons increase in concentration as the growth rate increases (Dong et al., 1996). Furthermore, tRNA pool regulation during amino acid starvation is the issue of several studies. The majority of tRNAs are rapidly degraded after exposure to amino acid starvation affecting the tRNA pools (Svenningsen et al., 2017). tRNA decay is suggested to be not only specific to the amino acid starvation but also to the general stress response (Sorensen et al., 2018), as it was also observed after rifampicin treatment (Svenningsen et al., 2017) and exposure to oxidative stress (Zhong et al., 2015). These tRNA degradations affect almost all tRNA species, aa-tRNAs, uncharged tRNAs, tRNAs cognate and non-cognate to the lacking amino acid. Rapid non-specific degradation of the surplus tRNAs is suggested to be the first response to cope with stress in order to limit the competition between cognate and near-cognate charged tRNAs (Sorensen et al., 2018; Svenningsen et al., 2017). Indeed, unfavorable conditions lead to higher rate of mistranslation which can be toxic for the cell. tRNA degradation seems to be ppGpp-independent (Svenningsen et al., 2017) and involves probably many different RNases not yet identified (Sorensen et al., 2018; Zhong et al., 2015).

IV.2) Transcript-specific regulations

As shown before, variation in translation activity between different growth conditions can result from global regulation of the translation machinery. In addition, intra-condition variability of transcript translation can be explained by the presence of transcript-dependent regulation. We describe below some mRNA-specific regulations which mainly target either translation initiation (SD-sequences, riboswitch, small RNAs, RNA-binding proteins), translation elongation (codon usage) or both initiation and elongation (secondary structure, mRNA length and abundance).

IV.2.1) Shine-Dalgarno sequence

As the initiation of translation consists in the binding of the ribosomal 30S subunit to the complementary RBS sequence, the RBS region is generally assumed to play a crucial role in translation regulation. The presence of an SD sequence, its degree of complementarity with the anti-SD sequence and the distance to the start codon affect the initiation efficiency in *E. coli, B. subtilis* (Osada et al., 1999) and *D. vulgaris* (Nie et

al., 2006). A synthetic library of 111 RBS sequences coupled with the reporter GFP protein shows that RBS strength can explain near 30% of variability of the protein level (Kosuri et al., 2013). Interestingly, a recent study highlights that an intermediate level rather than the highest level of complementary of SD-aSD sequence provides a better prediction of translation efficiency. They proposed that a too high aSD sequence complementary results in a strong binding between the ribosome and the RBS, hence increases the transition time to proceed elongation. The overall translation rate is therefore non-optimal (Hockenberry et al., 2017). Moreover, the frequency of SD-like sequences also affects the translation initiation rate: mRNAs with multiple SD sequences have higher translation efficiency than mRNAs with a single SD sequence (Evfratov et al., 2017). Bioinformatics tools such as RBS Calculator, developed by Salis lab (Salis et al., 2009) assume that the RBS strength is the major modulator of the translation initiation rate and therefore the primary contributor to gene expression.

IV.2.2) Small RNAs

Small RNAs (sRNAs) are non-coding regulatory RNAs which act by base-pairing with their mRNA targets. The base-pairing can be in *cis* (Brantl, 2007), when the sRNA is completely complementary to the target sequence. This occurs for instance when the sRNA is synthetized from the opposite strand of the mRNA target. The base-pairing can also be in *trans* when the sRNA is encoded by distinct loci from the mRNA target. There are around one hundred sRNAs experimentally identified or predicted in *E. coli* (Ruiz-Larrabeiti et al., 2016). Table 1 shows some main characteristics of sRNAs. The sRNA action can provide both translation repression and translation activation. sRNAs mainly target translation initiation by binding to the 5'UTR region of the regulated mRNA and act through a variety of mechanisms often in close interaction with the RNA degradation machinery (Storz et al., 2011). Most sRNAs are known to act in stress conditions or during changes in environment (e.g. DsrA on *rpoS* mRNA in response to cold shock or RyhB on *sodB* mRNA when iron is limiting). However some sRNAs also act in optimal growth condition. For example the IstR sRNA is as an antitoxin that represses the translation of the TisB toxin in optimal growth condition and Spot42 represses GalK expression on glucose (when GalK is not needed). Trans-acting sRNAs generally require the Hfq cofactor to stabilize the mRNA-sRNA base-pairing but not always as shown with the Hfq-independent binding of IstR on *tisB* mRNA.

Name	Condition	Effect	Mechanism/target sequence	Hfq
sRNA: RyhB target: *sodB* mRNA (Storz et al., 2011)	Low iron level	R	Binding to RBS site, changing conformation, disrupting stem-loop structure and preventing the formation of 30SIC complex	Yes
sRNA: DsrA target: *rpoS* mRNA (Resch et al., 2008; Soper and Woodson, 2008)	Cold shock	A	Base-pairing with 5'UTR and hairpin loop, unmasking RBS site and enabling initiation	Yes
sRNA: IstR target: *tisB* mRNA (Darfeuille et al., 2007; Vogel et al., 2004)	Optimal growth no stress	R	Base-pairing to 5'UTR 100 nucleotides upstream the RBS and preventing the binding of 30S subunit	No
sRNA: Spot42 target: *galK* mRNA (Møller et al., 2002)	Optimal growth	R	Binding to SD sequence	Yes

Table 1: Examples of sRNAs with their mRNA targets. The conditions of action, the effects (R: repression, A: activation), the mechanisms and the required association with Hfq are also shown.

IV.2.3) RNA binding proteins

As previously mentioned in the case of Hfq, RNA-binding proteins are involved in translation regulation in bacteria, by mainly targeting translation initiation (Holmqvist and Vogel, 2018). Hfq is a well-known post-transcriptionally acting RNA chaperone which facilitates base-pairing of sRNA with their mRNA targets. Hfq is a hexameric ring protein, which binds preferentially to single-stranded RNA, interacting with U-rich RNA on the proximal side of its central pore and with A-rich RNA on its distal face. In *E. coli* and *Salmonella*, Hfq may recognize more than 25% of all mRNAs (Holmqvist and Vogel, 2018). Recently a new role of Hfq in ribosome biogenesis was reported in *E. coli* (Andrade et al., 2018). Hfq binds to the 17S rRNA (16S precursor RNA) and facilitates its processing and folding to mature 16S rRNA.

A second well-studied RNA-binding protein acting on *E. coli* translation is CsrA (for "Carbon Storage Regulator"). CsrA is a 61 amino acid protein that regulates gene expression of many important cellular functions. CsrA represses glycogen metabolism, gluconeogenesis, biofilm formation and quorum sensing while it activates glycolysis, cell motility, virulence and pathogenesis as demonstrated in γ –Proteobacteria such as *Escherichia, Salmonella, Erwinia, Pseudomonas* or *Vibrio* (Timmermans and Van Melderen, 2010). CsrA is an essential protein in *E. coli* K-12 for growth on glucose (Baba et al., 2006). Translation repression by CsrA is mainly performed at the level of translation initiation. CsrA generally binds to multiple sites in the RBS region of the targeted mRNA and one of the binding sites overlaps the Shine-Dalgarno sequence to block ribosome binding (as exemplified for the *glgCAP* transcript (Baker et al., 2002)).

Some RNA-binding proteins are more specific to a growth condition. In *E. coli* the cold-shock protein CspA promotes translation during acclimation at low temperature (10 °C) by unfolding mRNA structure in the 5'UTR (Zhang et al., 2018).

IV.2.4) mRNA secondary structure

Another widely studied factor influencing translation initiation is the formation of secondary structure in the 5'UTR of mRNA. In silico, the strength of secondary structure is estimated by calculating the minimal Gibbs free energy (ΔG). Low folding energy corresponds to strong secondary structure and vice versa. First studies showed that unfolded structure surrounding the RBS site correlate positively with translation efficiency (de Smit and van Duin, 1990). Since then, other studies based on modified natural RBS through mutagenesis (Ringquist et al., 1992) and on synthetic randomized RBS sequences (at low scale (Barrick et al., 1994) and with larger size library (Cambray et al., 2018; Del Campo et al., 2015; Kudla et al., 2009)) also demonstrated that less structured regions upstream the start codon facilitate translation initiation of a reporter mRNA. Two genome-wide studies have measured *in vivo* the mRNA secondary structure of native genes. In *E. coli*, an enrichment of reduced structure in the 5'UTR is associated with an increased translation initiation (Del Campo et al., 2015). A similar correlation between translation efficiency and structure near the start codon is reported in yeast (Kertesz et al., 2010).

A special focus can be made on riboswitches, particular secondary structures located in the 5'UTR of an mRNA. Riboswitch generally regulates in response to changes in concentration of a small molecule ligand (for instance a nuclear base, an

amino acid or a sugar) the translation initiation of a protein involved in the metabolism of this ligand. Riboswitch is a common strategy of regulation of expression of genes involved in amino acid synthesis. In *E. coli*, examples are the regulation of the translation of enzymes involved in thiamine and lysine syntheses (Fig. 9). In the presence of the ligand (thiamin pyrophosphate and lysine, respectively), ligand binding to a specific structure of the 5'UTR (called aptamer) allows allosteric rearrangement of this structure, the RBS site is then sequestered in a stem-loop structure and translation initiation (of *thiM* and *lysC*, respectively) is inhibited. The riboswitch can have a dual role in *E. coli*. In addition to inhibit translation initiation, they modulate either Rho-dependent transcription termination (in the case of *thiM*) or RNase E accessibility (in the case of *lysC*) (Bastet et al., 2018).

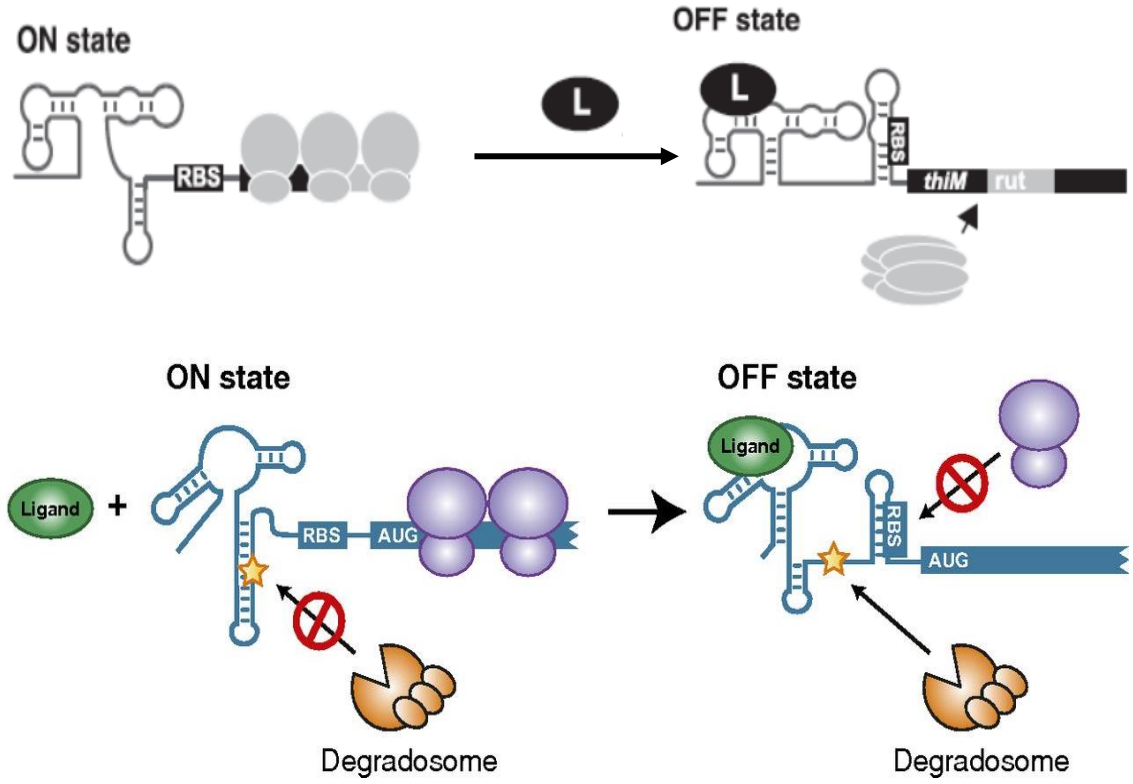


Figure 9: **Riboswitch modulation of translation initiation**. (A) Riboswitch of the thiamine synthesis: binding of thiamin pyrophosphate (L) results in RBS sequestration and inhibition of translation initiation. Concomitantly a *rut* sequence recognized by Rho is revealed leading to transcription termination (Bastet et al., 2018). B. Riboswitch of the lysine synthesis: binding of lysine (L) results in RBS sequestration and inhibition of translation initiation. In parallel a RNase E cleavage site is unmasked that facilitates transcript degradation (Caron et al., 2012).

Fewer studies have investigated the effect of secondary structure within the coding sequence (CDS) on translation elongation. Burkhardt and colleagues (Burkhardt et al., 2017) have experimentally measured *in vivo* the mRNA secondary structure in operons using DMS (dimethyl sulfate), a reagent that reacts with unpaired adenosine and cytosine nucleotides. They quantified the accessibility of RNA to DMS using DMS-seq, a high accessibility meaning a high number of unpaired adenosine and cytosine nucleotides and hence a low secondary structure. They found that adjacent CDS on the same mRNA molecule are structurally distinct and this structure is one major determinant of translation efficiency of each CDS. Strong negative correlation was shown between the level of secondary structure of the whole CDS and the translation efficiency. However, in another study of *in vivo* mRNA structure in *E. coli*, CDS secondary structure did not appear to generally affect translation elongation (Del Campo et al., 2015) except for a small group of genes coding for membrane proteins with secondary structures that stall ribosomes (Del Campo et al., 2015). This study used a different method than Burkhardt, called PARS. The enzyme RNase V1 is used to cleave double-stranded structured RNA and RNases A and T1 to cleave single stranded RNA at unpaired guanosine, cytosine and uracil. The ratio of double stranded/single stranded unpaired reads reflects the level of secondary structure. Using a synthetic approach based on a library of 224.000 sequences Cambray and colleagues (Cambray et al., 2018) support the Burkhardt results: the secondary structures around the start codon (-30 +30 bp) and in the gene (+1 to +60 bp and +31 to +90 bp from the start codon) control protein synthesis in *E. coli.*

IV.2.5) Codon usage

The genetic code is degenerated and different codons coding for the same amino acid are not used at the same frequency: this is called the bias in codon usage. Codon usage is one of the most intensively studied factors of gene expression as it generally correlates positively with the protein level (Guimaraes et al., 2014). Classic explanation is from an evolutionary point of view: optimal codons may have been selected for highly expressed genes (e.g. genes coding for ribosomal proteins) for higher fitness and lower cost for the cell (Kudla et al., 2009; Plotkin and Kudla, 2011). Codon usage has been thought to be tightly correlated to tRNA availability. Codons having highly abundant cognate tRNAs are more frequently used. Indeed, a positive correlation exists between the codon adaptation index (CAI) (Sharp and Li, 1987) and the tRNA

Adaptation Index (tAI). CAI attributes a coefficient to each codon according to its usage frequency in a set of reference genes coding for highly expressed proteins (which are supposed to have efficient translation). tAI estimates the amount of adaptation of a gene to the genomic tRNA pool and represents the co-adaptation between the coding sequences and the tRNA pool (Reis, 2004). Optimal codons are assumed to be translated faster and more efficiently, hence increasing elongation speed and translation efficiency (Berg and Kurland, 1997). Two recent studies demonstrate that rare codons with less abundant tRNAs are decoded slower, in *E. coli* and yeast (Gardin et al., 2014; Gorochowski et al., 2015). In contrast, some studies in yeast report no correlation between the codon usage frequency and ribosome speed, the more frequent codons being translated at the same rate as the rare codons (Pop et al., 2014; Qian et al., 2012). These studies suggest that the codon bias is adapted to the tRNA pools as an evolution strategy to balance the tRNA supply and demand.

In addition, the effect of the codon usage on translation depends on different parameters such as the scale of the analysis (at the codon level or the gene level), the study of homologous or heterogeneous genes and the growth conditions. These three parameters are described below.

Several studies propose a local effect of the codon usage on translation. An enrichment of rare codons at the N-terminal part facilitates in yeast translation initiation and therefore gene expression (Allert et al., 2010; Tuller et al., 2010a). Goodman and colleagues (Goodman et al., 2013) using 14000 constructions with codon variants at the N-terminal extremity confirm in *E. coli* that N-terminal rare codons instead of common ones increase expression of a reporter protein by ~14-fold. Nakahighashi et al. (Nakahigashi et al., 2014) also confirm in *E. coli* that at the codon level less abundant codons cause higher ribosome footprint density, but when average at the gene level, they report an opposite result: genes with high average CAI tend to have higher ribosome footprint density. The proposed explanation is that translation initiation rather than translation elongation (and thus codon usage) determines translation efficiency.

Another debate is provided by the study of translation of endogenous versus heterologous genes. Tuller and colleagues find a positive correlation in *E. coli* between tAI and translation efficiency (ratio protein on mRNA level) for endogenous genes (Tuller et al., 2010b). In contrast, no positive correlation between protein expression

and CAI is reported using a synthetic library of 154 variants of the heterologous GFP gene in *E. coli* (Kudla et al., 2009). One hypothesis is that GFP has higher GC content and secondary structure than the native genes and hence the codon bias is not the major factor for GFP (Tuller et al., 2010b). However, this is may be not so simple since different results are also reported respectively for the native and heterologous genes by other authors. No correlation between codon bias and protein/mRNA ratio is reported for *E. coli* native genes (Lu et al., 2007) whereas optimizing the overall codon sequence of heterologous genes results in increased expression (Burgess-Brown et al., 2008; Gustafsson et al., 2004).

The effect of codon usage differs also with the growth condition. A correlation between codon bias and local ribosome footprint density is only reported in *E. coli* cells during amino acid starvation (Subramaniam et al., 2014). The effect of non-optimal codons could be more important under stress or limited growth (Wohlgemuth et al., 2013). Indeed under optimal growth condition, initiation is assumed to be the main rate limiting step of translation and not elongation. In this case, although codon optimality of the coding sequence may have an impact on the elongation rate, it will not significantly affect the final protein expression (Cambray et al., 2018; Plotkin and Kudla, 2011). In contrast, under amino acid starvation, elongation may become the limiting step of translation allowing the effect of the codon usage on the final protein expression to be seen.

In conclusion, although the majority of studies uses synthetic approach with heterologous genes and report positive correlation between codon optimality and translation efficiency, the effect of codon usage on translation remains an active debate.

The determination of the codon usage effect on translation is even more challenging by the fact that the codon usage and the formation of secondary structure are intrinsically coupled. Several studies investigate the simultaneous effects of codon usage and secondary structure and contrasted results were obtained. In one hand, some studies suggest that only the secondary structure is the driving force of translation (Bentele et al., 2014; Goodman et al., 2013; Kudla et al., 2009). In this case, the effect of the codon bias on translation is not direct but mediated through the effect of the secondary structure (Evfratov et al., 2017). On the other hand, some studies identify the codon usage as the main determinant of translation compared to the influence of the secondary structure (Boël et al., 2016; Tuller et al., 2010). Tuller et al.

propose that secondary structures may only modulate the effect of the codon usage (Tuller et al., 2010b). At last, Gorochowski et al., propose that both codon usage and secondary structure are affecting *E. coli* translation (Gorochowski et al., 2015). Since highly abundant codons are found in regions with high secondary structure and vice versa, the two factors appear to have opposite effect on translation elongation rate. The codon usage and secondary structure interaction is suggested to smooth the overall translational speed (Gorochowski et al., 2015).

IV.2.6) mRNA length

Another potential feature regulating translation efficiency is mRNA length. Shorter genes have generally higher ribosome density, a universal mechanism conserved from unicellular to high eukaryotes: in *S. cerevisae* (Arava et al., 2003; Ingolia et al., 2009), *L. lactis* (Picard et al., 2012), *P. falciparum* (Lacsina et al., 2011) and human (Hendrickson et al., 2009). From an evolutionary point of view, highly needed proteins such as ribosomal proteins have short length and high translation efficiency, a selection during evolution to fulfil the cell demand (Shah et al., 2013).

One possible mechanism of length-depend translation is an increase of translation initiation in short genes. Fernandes and colleagues proposed a mechanistic model based on physical principles to explain how CDS length can affect translation in yeast (Fernandes et al., 2017). According to them, in short genes, ribosome recycling rate is higher via a feedback mechanism and contributes to higher local concentration of ribosomes at the 5'UTR and thus enhances initiation rate. Short genes have a smaller distance between their 3' end and 5' UTR and thus a higher probability of the terminating ribosome to be reused for initiation. A second explanation consists in a higher rate of ribosome drop off for long genes (Fernandes et al., 2017). The failure of ribosomes to complete the synthesis of a full-length protein is reported in *E. coli* under both stressful and non-stressing conditions (Sin et al., 2016).

IV.2.7) mRNA abundance

Besides sequence-related determinants, mRNA abundance is mentioned as a potential translational regulator in some studies. A positive correlation between mRNA abundance and translation efficiency (ratio of protein synthesis rate/mRNA level) is shown in *S. cerevisiae* (Pop et al., 2014). In *E. coli*, mRNA concentration is also reported to have a positive effect on the ribosome footprint abundance, a proxy of protein

synthesis (Bartholomäus et al., 2016; Morgan et al., 2017). Regarding the ribosome density, a negative effect of mRNA concentration is described in *L. lactis* (Picard et al., 2012) whereas in yeast, no correlation and a positive correlation are reported in *S. cerevisiae* and *S. pombe*, respectively (Lackner et al., 2007). The regulatory effect of the mRNA abundance on translation remains thus unclear in microorganisms. In our study, we will investigate in *E. coli* the impact of mRNA concentration on translation at the genome-wide level.

V) Coupled regulations of translation and other cellular processes

Efficient, rapid, and energetically beneficial fine-tuning of protein concentrations requires multilayer regulation involving interactions between different cellular processes. In the next paragraphs, we will describe how translation regulation can influence transcription, mRNA degradation and other protein-related processes such as the formation of protein complex, protein folding and localization, and resource allocation.

V.1) Coupling of translation and transcription

In prokaryotes, translation is tightly coupled with transcription. Since there is no nucleus and spatial separation, the two processes occur quite simultaneously. During the first round of translation, ribosome binds to the nascent mRNA and follows the RNA polymerase (RNAP). The ribosome and RNAP can be physically coupled by the small subunit ribosomal protein S10 bridged or not by the coupling agent NusG (or its paralog RfaH) (Artsimovitch, 2018). The ribosome and RNAP coupling is assumed to control transcription rate and ensure accuracy. The presence of translating ribosomes provides efficient transcription by pushing forward RNAP and preventing spontaneous backtracking and pausing of RNAP (Kohler et al., 2017; McGary and Nudler, 2013). Backtracking slows down transcription and generates antisense untranslated RNA. Proshkin and colleagues showed that the transcription rate depends on the translation rate (Proshkin et al., 2010). Moreover, the loss of the ribosome-RNAP coupling can lead to premature termination of transcription by the Rho factor (Artsimovitch, 2018). In absence of ribosome-RNAP coupling, naked mRNA fragments located between RNAP

and the ribosome facilitate Rho association and the formation of hairpin structure causing premature termination. The decoupling of ribosome-RNAP can be due to unwanted events causing ribosome stalling (e.g. rare codon, proline, anti-Shine Dalgarno like sequence…) or accelerated RNAP (Fig. 10). When translation encounters accident, termination of transcription can be an economical strategy to ensure only synthesis of high quality proteins.

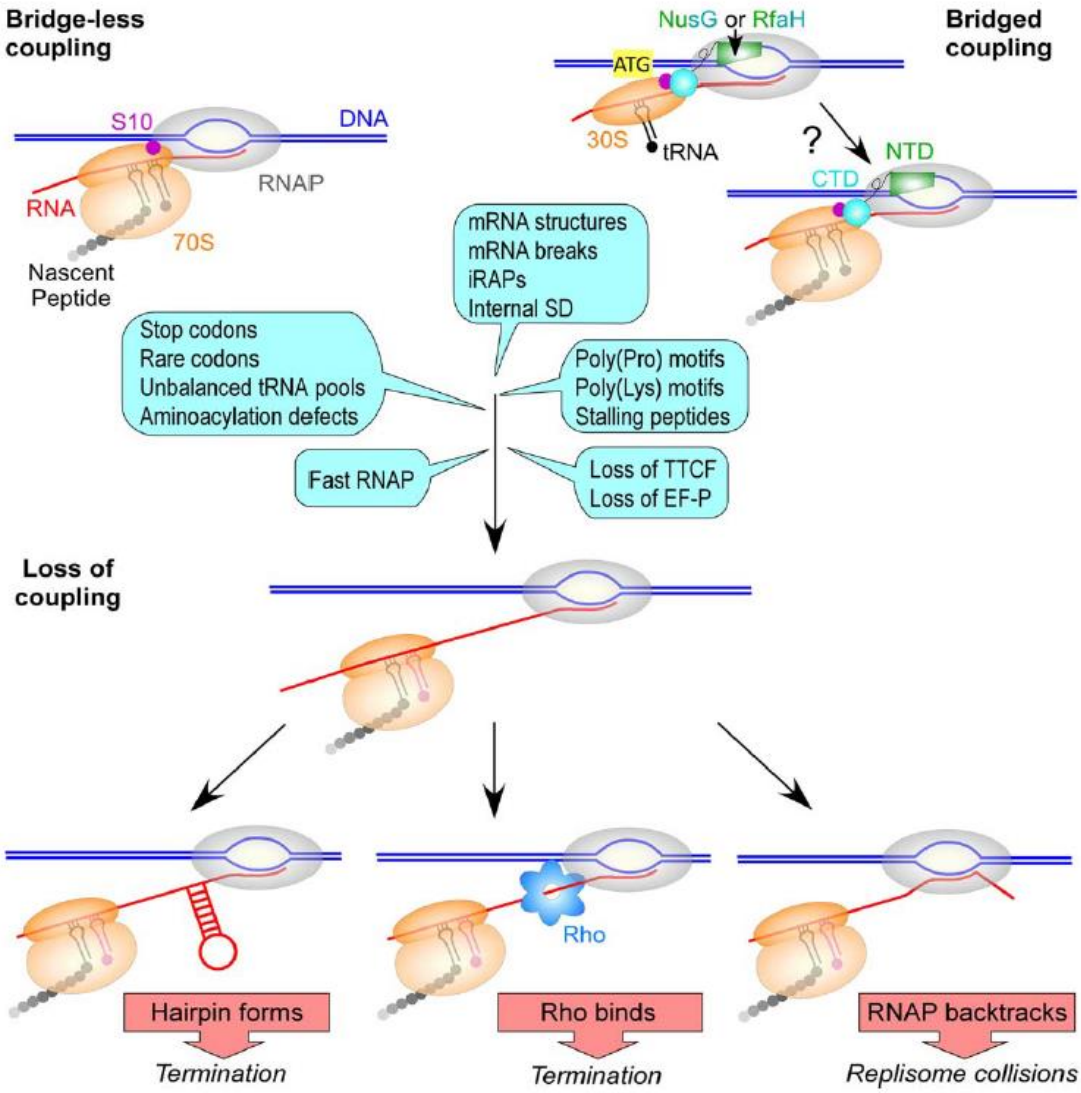


Figure 10: Transcription – translation coupling via the association of ribosome and RNAP. Ribosome and RNAP coupling can be bridge-less via the ribosomal protein S10 or bridged via NusG (or RfaH). Loss of coupling can result from different reasons and has different consequences (Artsimovitch, 2018).

V.2) Coupling of translation and mRNA degradation

Another largely studied interaction is the effect of translation on mRNA stability. Generally, in *E. coli*, ribosome binding on mRNA has a protecting effect against degradation. Binding of ribosome at the RBS in the 5'UTR prevents the fixation of RNases, as shown in the cases of *ermA*, *ermC* and *phage82* mRNAs (Joyce and Dreyfus, 1998). Mutations affecting translation initiation lead to variation of mRNA stability (Arnold et al 1998; Iost and Dreyfus 1995). Moreover, close spaced ribosomes during elongation mask cleavage sites of RNase E (Deana and Belasco, 2005; Dreyfus, 2009). Protecting effect of ribosome depends on the distance between adjacent ribosomes (Pederson 2011) and is effective when the distance is lower than 24 nucleotides (Carrier and Keasling, 1997).

However, ribosomes can also promote transcript degradation. Ribosome stalling (because of aberrant codon or the presence of rare codons…) can activate trans-translation and leads to mRNA decay (Hayes and Sauer, 2003). This phenomenon is supposed to control the quality of the protein synthesis to avoid deleterious effects on the cell (Keiler 2008). In addition, a potential direct interaction between translating polysomes and the mRNA degradation machinery (RNase E and degradosome) is shown in *E. coli* suggesting that polysome may recruit RNase E and facilitate mRNA degradation (Tsai et al., 2012). At last, a recent large-scale study suggests that inefficiently translated mRNAs (corresponding to a low quantity of synthetized protein measured on SDS-page) in *E. coli* are more rapidly degraded (Boël et al., 2016). This can indicate that a higher number of slow ribosomes facilitates transcript degradation. Therefore, the authors propose a kinetic competition between protein elongation and mRNA degradation (Boël et al., 2016).

The relationships between translation and mRNA degradation is also a matter of research in yeast. Chan and colleagues (Chan et al., 2018) support the model of translation protection of mRNA rather than the stalled ribosome-triggered decay (Figure 11). They found that elongation inhibition with antibiotics or rare codons increases mRNA stability. Reducing elongation rate causes higher ribosome density and hence higher protection against RNases.

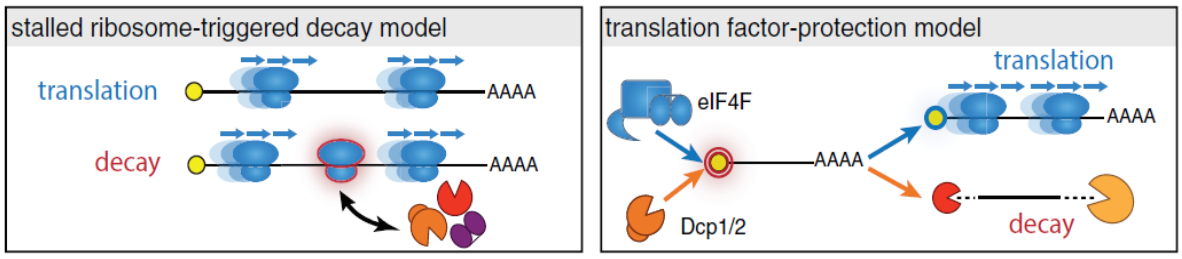


Figure 11: Two models of effect of translation on degradation: stalled ribosome-triggered decay and ribosome-protective model in yeast (Chan et al., 2018).

Presnyak et al (Presnyak et al., 2015) obtain opposite results. They showed that optimal codons enhance translation elongation rate and also stabilize the mRNA molecule. Vice versa, slowing ribosome movement by inserting rare codons promotes mRNA decay. Thus, translation speed seems to positively protect mRNA from decay in agreement with Boel et al. in *E. coli*.

V.3) Coupling of translation and other protein-related processes

V.3.1) Formation of multiple protein complex

In bacteria, proteins with related functions can form a multiple protein complex. Their genes are often clustered in an operon, a polycistronic structure. Operon allows a co-regulated expression of these genes at the transcriptional level to rapidly adapt their expression to environmental changes. For some enzymatic complexes, despite the same transcription level, enzyme subunits are produced unequally. Different studies investigated the proportional protein synthesis according to the complex stoichiometry and they suggested that not transcription but translation is responsible of this uneven protein synthesis. Indeed, analyses of some protein complexes (e.g. ATP synthase, ribosomal operon) in three bacteria show a positive correlation between ribosome density and protein stoichiometry in *E. coli* and *B. subtilis* (Quax et al., 2013). Genes coding for the highest subunit stoichiometry have also the highest initiation rate (regulated by secondary structure) and elongation rate (via high CAI). Mechanisms such as *de novo* internal initiation and translational coupling are suggested to contribute to this uneven production (Levin-Karp et al., 2013; Tian and Salis, 2015) (Fig. 12). The position in the operon can have an impact on translation via translational coupling while *de novo* internal initiation is position-independent. In agreement with these results, Li at al. also demonstrated that the protein synthesis rate (measured by ribosome footprints) is proportional to the complex stoichiometry (Li et al., 2014). In

the case of the ATP synthase complex, the ATP operon has evolved to synthesize the appropriate ratio of subunit proteins, ranging from 1- to 10-fold. This is observed for numerous protein complexes (for example the formate dehydrogenase, the methionine transporter, the bacterial condensing…) with very diverse cellular functions (Li et al., 2014).

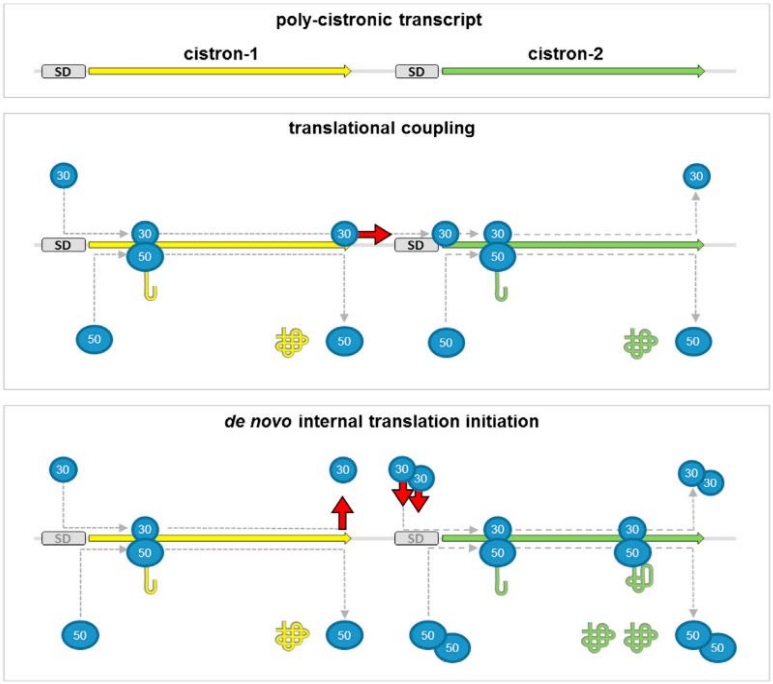


Figure 12: Model for translational coupling and *de novo* internal translation initiation which can contribute to differential translation in polycistronic transcript (Quax et al., 2013).

V.3.2) Protein folding and localization

Translation also contributes to protein folding. Slowing down elongation by introducing non-optimal codons facilitates in *Neurospora* the cotranslationnal folding of the luciferase protein increasing the proportion of functional enzyme (Yu et al., 2015). The translation-dependent folding of a protein plays also a role in the protein localization to the membrane. Membrane proteins are extremely hydrophobic and require to be inserted into the membrane cotranslationally. To do so, they must be targeted to the translocon early in translation before large polypeptide portions are synthesized. Membrane targeting of ribosome-translating proteins is mediated by the signal recognition particle (SRP) system which recognizes the N-terminal sequence and helps protein translocation to the membrane. Slowing down elongation (for example by designing ribosome pausing at the beginning of the transcript) improves

the protein folding and facilitates the protein targeting to the membrane (Fluman et al., 2014). A modelling study proposes that translation dynamics affects the spatial distribution of proteins: playing with the translation rate will result in more or less protein clusters on the membrane (Korkmazhan et al., 2017).

V.3.3) Resource allocation

Translation is also involved in resource allocation issues. By increasing translation efficiency (a higher number of proteins produced per mRNA using 5' sequence variants), *E. coli* cells can produce the same amount of a protein from a lower RNA level. In this case, the cells minimize the expression cost per protein produced by reducing the transcription cost for the mRNA synthesis (Frumkin et al., 2017). Frumkin et al (Frumkin et al., 2018) shows also that the translation level of the very highly expressed genes influences the efficiency of global protein translation in *E. coli* cells. Conversion of abundant codons into rare codons in highly expressed genes leads to a lower expression of these genes but also to a lower translation efficiency of other genes which include these rare codons. At last, non negligible amounts of proteins are translated in *E. coli* either in excess (like ribosomal proteins (Mori et al., 2017)) or when they are even not needed in the used growth condition (22% of the proteome do not have any benefit during growth in minimal medium (Price et al., 2016)). The unused protein expression is a source of fitness cost but it can provide a "reservoir" strategy for the cells to quickly adapt to changing growth conditions (O'Brien et al., 2016).

In conclusion, translation through its interaction with many cellular processes plays a central role in cell physiology. Different approaches have therefore been developed for translation characterization at the genome-wide scale (called translatome). Translatome experiments are described in the following paragraph VI.

VI) Large-scale study of translatome

Recent years are marked by the expansion of large-scale studies of translation regulation, through the advance in high-throughput methods. Genome-wide experiments based on microarrays and sequencing techniques allow to investigate the *in vivo* translational status of all the genes, they are called translatomes. Comparison of translatomes is commonly used to study the translational response between two conditions (between different growth phases) or after environmental changes, mostly

perturbation (stress, starvation) (Kuhn et al., 2001; Lange et al., 2007; Melamed et al., 2008; Warringer et al., 2010)

VI.1) Polysome profiling

First studies allowing to access translatome data were based on the polysome profiling technique (Arava et al., 2003; Melamed et al., 2008). For each mRNA in the cell, the polysomal profiling method provides answers to the two questions: is an mRNA engaged in translation? and if yes, how many ribosomes are bound. After translation arrest by antibiotic addition and cell lysis, mRNA-ribosome complexes are separated according to the charge in ribosomes in a sucrose gradient (Fig.13). The different fractions differ in mRNA translational status: free mRNAs not bound to any ribosome, mRNAs bound to the small or large ribosomal subunits, mRNAs bound to 1 ribosome and mRNAs bound to polysomes. mRNA levels are identified and quantified in each fractions by genome-wide method such as microarray (Picard et al., 2012; Melamed et al., 2008).

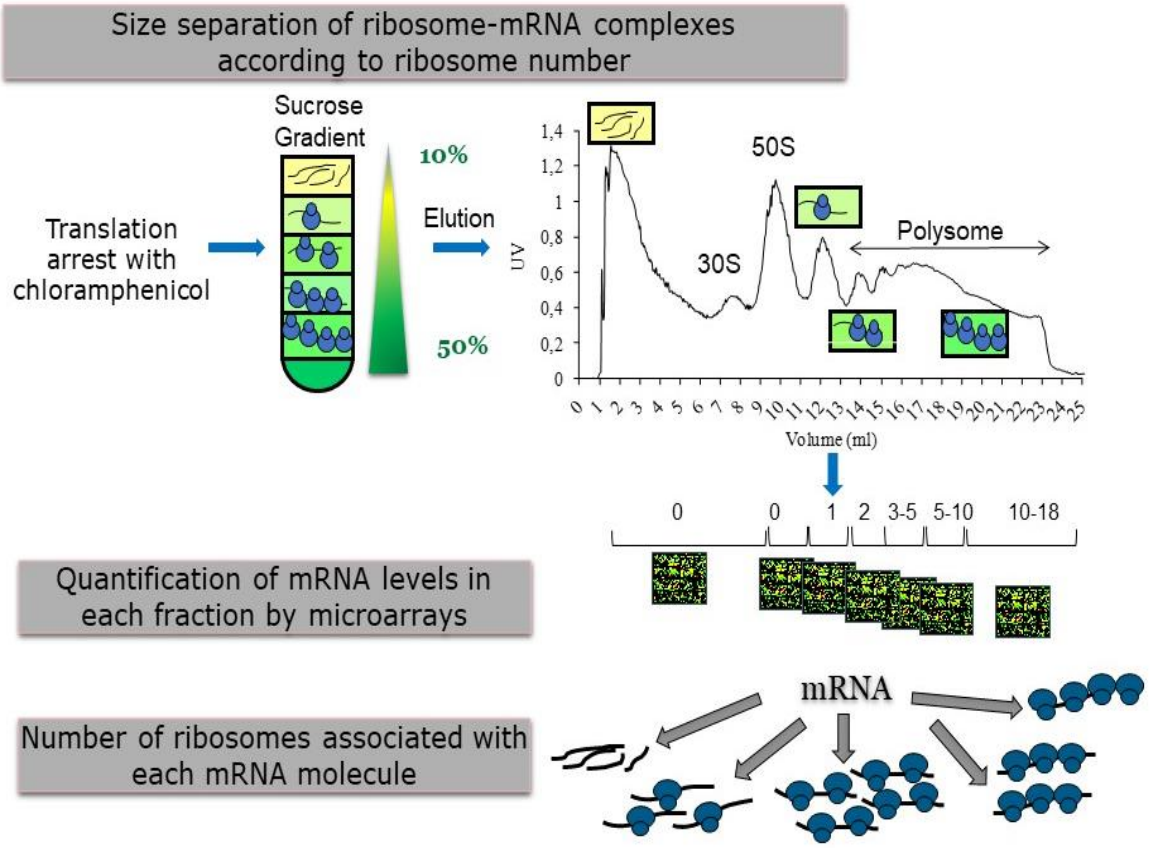


Figure 13: Scheme of the polysome profiling method developed in *L. lactis* (Picard et al., 2012)**.**

Fraction assembling depends on the needed translatome resolution. The low-resolution assembling corresponds to two mRNA groups: polysomal versus non-polysomal groups (the polysomal fraction corresponding to the strongly-translated mRNAs loaded with several ribosomes, and the nonpolysomal one including weakly or untranslated mRNAs). The low resolution is sufficient to study the translational parameter: proportion of mRNA engaged in translation also called 'ribosome occupancy'. This parameter assesses the level of translation (yes/no) of an mRNA without focusing on the exact number of bound ribosomes. Assembling for a high translatome resolution provides a higher number of mRNA groups classified with respect to the precise number of loaded ribosomes. The high resolution allows to measure another translational parameter: 'ribosome density' which is the most frequent number of bound ribosomes per transcript length unit (Arava et al., 2003; Picard et al., 2012, 2013).

Ribosome occupancy is suggested to be mainly related to translation initiation as demonstrated in yeast by the significant correlation between ribosome occupancy and effectiveness of the AUG context to promote translational initiation (Lackner et al., 2007). Ribosome density depends on both rates of initiation and elongation. For example, a low ribosome density can be associated to low rates of initiation and elongation (Fig. 14) (when initiation is the rate limiting step) or to high rates of initiation and elongation, whereas a high ribosome density can come from high initiation and low elongation rate (when elongation is limiting the translation rate).

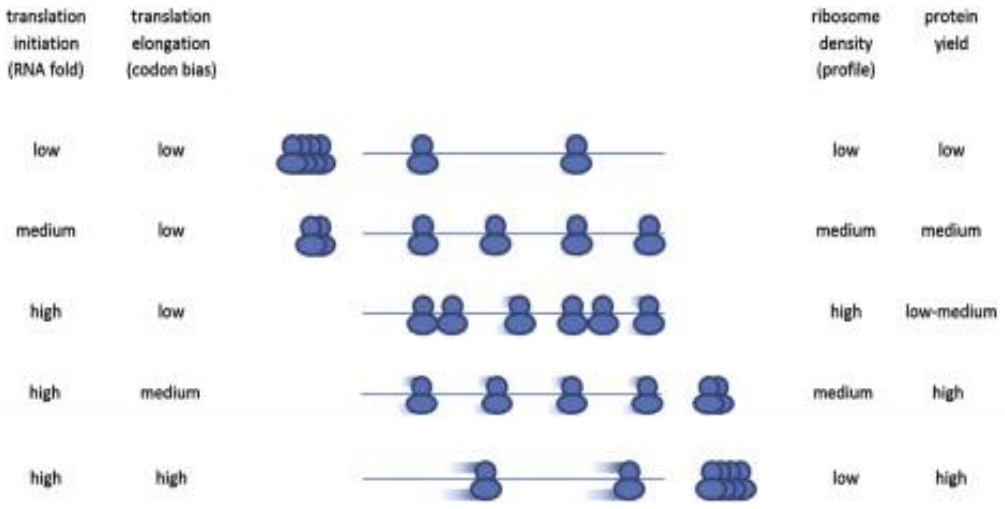


Figure 14: Different levels of ribosome density resulting from different combinations of translation initiation and elongation rates (Quax et al., 2013)**.**

VI.2) Ribosome profiling

As shown in the previous paragraph, polysome profiling provides an overall translational status with the number of ribosomes bound to an mRNA but not the precise ribosome distribution along the transcript. Ribosomes could be homogenously distributed along the transcript or preferentially located at some specific regions of the mRNA creating differential local ribosome densities. The ribosome profiling method was first developed by Ingolia and colleagues in yeast to measure local ribosome density (Ingolia et al., 2009). Ribosome profiling is based on a deep sequencing of ribosome protected mRNA fragments (Fig. 15). mRNA fragments bound with ribosomes are protected against RNase I degradation. Deep sequencing is used to quantify ribosome footprint fragments. This technique was initiated by Huttenhofer 2004 and became a powerful tool with the development of high-throughput sequencing techniques for short RNA fragments. Ribosome footprints are then mapped to the genome sequence and the exact position of ribosome can be visualized.

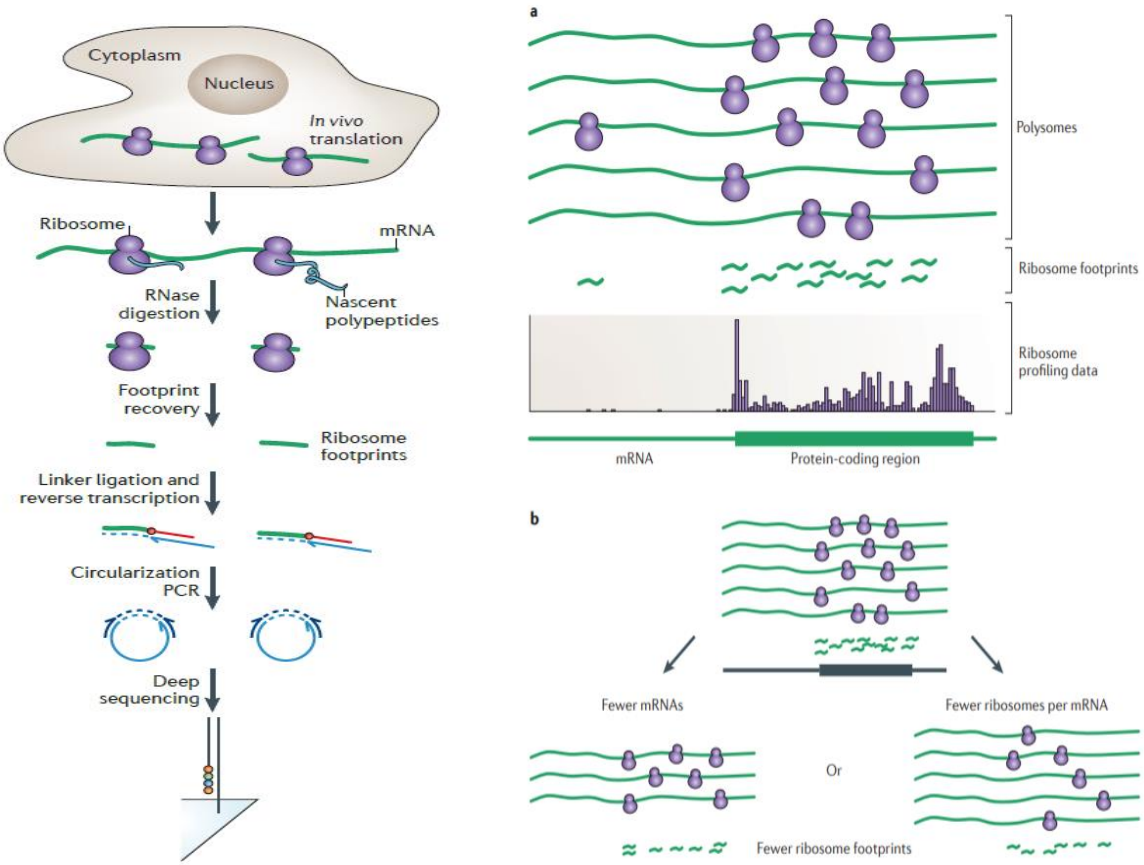


Figure 15 : Method of ribosome profiling: experimental steps (left) and visualization and analysis of ribosome profiling data (right) (Ingolia et al., 2009).

In *E. coli,* the first ribosome profiling study was carried out by Oh and colleagues to study the recruitment of the chaperone trigger factor (TF) on nascent protein (Oh et al., 2011). Later, Li et al. used ribosome profiling to show a propensity of ribosome pausing at internal SD like sequences (GGUGGUG) (Li et al., 2012). By discovering ribosome accumulation at specific locations, ribosome profiling is applied to investigate the effect of mRNA features (for example codon usage) on translation elongation.

At the gene level, the ribosome profiling method also provides the overall number of ribosomes bound to an mRNA species. The ribosome footprints can be used as a good indicator of protein synthesis (Ingolia et al., 2009; Li et al., 2014). Normalizing the ribosome footprints by the mRNA level results in the average ribosome density per mRNA species. Comparison of the results obtained by the polysome profiling and the ribosome profiling in yeast shows a correlation between the two methods, the genes peaking in the heavier fraction in the polysome profiling experiment tend to have high ribosome density by ribosome profiling (Arava et al., 2003; Ingolia et al., 2009). In *E. coli*, this method is used to study translation efficiency of each gene in optimal growth, to identify key factors regulating translation (Burkhardt et al., 2017; Li et al., 2014). Others studies investigate the variation of translation efficiency in condition of stress (Morgan et al., 2018; Zhang et al., 2017).

VI.3) Comparison of polysome profiling and ribosome profiling

The polysome profiling and ribosome profiling methods are complementary to study translation with their own advantages/disadvantages. A brief comparison is presented in Table 2.

Polysome profiling provides a real picture of the translational status by describing the heterogeneity in the number of bound ribosomes between the mRNA copies. This is not possible by ribosome profiling that does not differentiate the different mRNA copies. From this heterogeneity, the percentage of mRNA copies engaged in translation (ribosome occupancy) can be calculated. The polysome profiling allows also the calculation of the physical ribosome density of each mRNA copies. On the other hand, the ribosome profiling allows to access local ribosome density to study for example translation initiation or ribosome pausing.

	Polysome profiling	Ribosome profiling
Heterogeneity between mRNA copies	**+**	**-**
Ribosome occupancy	**+**	**-**
Physical ribosome density	**+**	**-**
Local ribosome density	**-**	**+**
Study of untranslated regions	**+**	**-**
Technical difficulty	**-**	**+**

Table 2: Major advantages and disadvantages of polysome profiling and ribosome profiling methods

Polysome profiling can be used to study translation of full-length mRNAs including untranslated regions in contrast to ribosome profiling which it restricts to mRNA regions protected by ribosomes (Chassé et al., 2016). However, the major difficulties of polysome profiling are technical: to cleanly resolve the different fractions during the fractionation process and to handle an important number of samples (Ingolia 2014). Methods need to be developed to reduce this experimental effort.

VII) OBJECTIFS DE LA THESE

Cette étude bibliographique souligne le rôle central de la traduction dans la régulation de l'expression génique. Etape clé dans le réseau complexe de régulations entre le gène et la protéine, la traduction est en étroites interactions avec les autres processus cellulaires (la transcription, la dégradation de l'ARNm et d'autres processus liés au devenir de la protéine). La traduction est grande consommatrice d'énergie, elle est donc l'objet de régulations pour répondre aux besoins cellulaires tout en optimisant l'utilisation des ressources. En effet, la traduction est régulée à l'échelle globale via la machinerie de traduction (le nombre de ribosomes et la disponibilité des tRNAs) pour s'adapter aux conditions environnementales. Elle est aussi activée ou inhibée de façon plus spécifique (via les petits ARNs, des protéines se fixant à l'ARN, les riboswitches…) pour des sous-groupes de gènes dont les expressions sont nécessaires pour l'adaptation et la croissance dans certaines conditions.

Les données de la littérature montrent la complexité de la régulation traductionnelle : il existe des variations du niveau de traduction non seulement entre différentes conditions de croissance mais aussi entre les transcrits dans une même condition de croissance. Les déterminants de la traduction les plus connus sont des paramètres liés à la séquence de l'ARNm (la force du RBS, le CAI) mais bien que très étudiés dans le cas d'ARNm particuliers, leurs effets sur la traduction dans sa globalité ne sont pas encore complètement éclaircis ni hiérarchisés. Par ailleurs, l'analyse de la traduction ne devrait pas être effectuée seule car la traduction est directement liée aux autres processus cellulaires et à l'état physiologique de la cellule. Ainsi, l'originalité notre étude est d'utiliser une approche de biologie intégrative en associant les régulations de la traduction avec d'autres données -omiques telles que les caractéristiques du génome, le transcriptome et le protéome.

Cette thèse vise à compléter l'image globale de la traduction chez *E. coli* en décrivant l'état traductionnel de l'ensemble des gènes et leurs régulations. La condition de croissance choisie pour *E. coli* est une croissance exponentielle sur glucose correspondant à une vitesse de multiplication maximale des cellules. Nous voudrions savoir comment la traduction est régulée en concordance avec les besoins cellulaires nécessaires à cette croissance optimale. Pour cela, la première étape consistera à la génération de données de traductome, la mesure expérimentale à l'échelle -omique du

niveau de traduction de chaque ARNm (chapitre III). Les données du traductome seront déterminées pour la première fois chez *E. coli* par la méthode de polysome profiling suivie de la quantification des ARN par RNA-Seq. Dans la littérature, le traductome chez *E. coli* a été réalisé soit par couplage du polysome profiling avec la technique des microarrays mais pour une étude de la traduction à faible résolution (ARN libres versus pools des fractions polysomiques) (Lange et al., 2017) soit par la méthode du ribosome profiling (Li et al., 2014). Comme montré dans notre étude bibliographique, le polysome profiling présente plusieurs avantages par rapport au ribosome profiling et fournit une photo différente de la traduction. Le polysome profiling permet de fractionner les ARNm en fonction de leurs charges en ribosomes et le RNA-Seq de quantifier la part de chaque ARNm dans chacune des fractions. L'activité de traduction sera exprimée par deux variables : la « ribosome occupancy » (RO) et la densité en ribosomes (RD). RO correspond à la proportion de copies d'un ARNm donné occupées par au moins un ribosome, c'est-à-dire le pourcentage de copies d'un ARNm en traduction. RD est le nombre de ribosomes par unité de longueur.

A partir de ces données de traductome, le deuxième objectif sera d'identifier les déterminants majeurs qui pourraient être à l'origine d'une variabilité de traduction inter-gènes (chapitre III). Les études de la littérature utilisant le polysome profiling ou le ribosome profiling chez *E. coli* avaient pour but d'étudier par exemple la traduction des ARN « leaderless » (Lange et al., 2017) ou la pause des ribosomes sur des motifs antiSD dans l'ORF (Li et al., 2012) mais pas d'identifier les déterminants généraux qui expliqueraient la variabilité entre les gènes dans une condition donnée. Un modèle statistique de régression linéaire multiple sera donc utilisé dans nos travaux pour hiérarchiser l'effet simultané de différents facteurs sur les RO et RD de l'ensemble des gènes. Les paramètres généraux pris en compte comprendront ceux souvent étudiés (la structure secondaire, l'usage des codons) mais également ceux liés aux caractéristiques des protéines correspondantes (fonction, localisation,…) et à l'état physiologique de la cellule via la concentration en ARNm. En effet, notre hypothèse est que la traduction est finement régulée en fonction des besoins cellulaires. L'analyse statistique des données de traductome requerra au préalable le choix des méthodes d'alignement des séquences et de normalisation des données de RNA-Seq (chapitre II).

En dernier lieu, nous voudrons valider au niveau moléculaire l'effet du facteur le plus intéressant prédit par l'analyse statistique (chapitre III). Pour cela, nous

utiliserons des ARNm rapporteurs pour lesquels le paramètre sélectionné sera modifié de manière artificielle. Afin de mesurer de manière simultanée l'effet du paramètre sur les variables RO et RD, une méthode de multiplexage de polysome profiling sera développée afin de mélanger les souches exprimant ces différents ARNm rapporteurs (chapitre I).

L'ensemble de ces résultats permettra sur le plan fondamental de connaitre le niveau traductionnel de l'ensemble des gènes chez *E. coli* en croissance rapide et d'éclaircir le rôle de la régulation traductionnelle au niveau cellulaire. Sur un plan plus appliqué, ces travaux devraient permettre de proposer des perspectives d'amélioration de la production de protéines d'intérêt chez *E. coli*.

REFERENCES

Allert, M., Cox, J. C., & Hellinga, H. W. (2010). Multifactorial Determinants of Protein Expression in Prokaryotic Open Reading Frames. *Journal of Molecular Biology*, *402*(5), 905–918.

Andrade, J. M., dos Santos, R. F., Chelysheva, I., Ignatova, Z., & Arraiano, C. M. (2018). The RNA-binding protein Hfq is important for ribosome biogenesis and affects translation fidelity. *The EMBO Journal*, *37*(11), e97631.

Arava, Y., Wang, Y., Storey, J. D., Liu, C. L., Brown, P. O., & Herschlag, D. (2003). Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. TL - 100. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(7), 3889–3894.

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.

Artsimovitch, I. (2018). Rebuilding the bridge between transcription and translation: NusG as a transcription-translation coupling factor. *Molecular Microbiology*, *108*(5), 467–472.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., … Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*, *2*, 2006.0008.

Baggett, N. E., Zhang, Y., & Gross, C. A. (2017). Global analysis of translation termination in E. coli. *PLOS Genetics*, *13*(3), e1006676.

Baker, C. S., Morozov, I., Suzuki, K., Romeo, T., & Babitzke, P. (2002). CsrA regulates glycogen biosynthesis by preventing translation of glgC in Escherichia coli: Regulation of glycogen biosynthesis in E. coli. *Molecular Microbiology*, *44*(6), 1599–1610.

Bakshi, Somenath, Heejun Choi, and James Weisshaar. (2015). The Spatial Biology of Transcription and Translation in Rapidly Growing Escherichia Coli. *Frontiers in Microbiology* 6 (July).

Bakshi, Somenath, Albert Siryaporn, Mark Goulian, and James C. Weisshaar. (2012). Superresolution Imaging of Ribosomes and RNA Polymerase in Live Escherichia Coli Cells. *Molecular Microbiology* 85 (1): 21–38.

Barrick, D., Villanueba, K., Childs, J., Kalil, R., Schneider, T. D., Lawrence, C. E., … Stormo, G. D. (1994). Quantitative analysis of ribosome binding sites in E.coli. *Nucleic Acids Research*, *22*(7), 1287–1295.

Bartholomäus, A., Fedyunin, I., Feist, P., Sin, C., Zhang, G., Valleriani, A., & Ignatova, Z. (2016). Bacteria differently regulate mRNA abundance to specifically respond to various stresses. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, *374*(June), 20150069-.

Bastet, L., Turcotte, P., Wade, J. T., & Lafontaine, D. A. (2018). Maestro of regulation: Riboswitches orchestrate gene expression at the levels of translation, transcription and mRNA decay. *RNA Biology*, 1–4.

Bentele, K., Saffert, P., Rauscher, R., Ignatova, Z., & Bluthgen, N. (2014). Efficient translation initiation dictates codon usage at gene start. *Molecular Systems Biology*, *9*(1), 675–675.

Berg, O. G., & Kurland, C. G. (1997). Growth Rate-optimised tRNA Abundance and Codon Usage. *Journal of Molecular Biology*, *270*, 544–550.

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. *Proceedings of the National Academy of Sciences*, *99*(15), 9697–9702.

Berthoumieux, S., de Jong, H., Baptist, G., Pinel, C., Ranquet, C., Ropers, D., & Geiselmann, J. (2014). Shared control of gene expression in bacteria by transcription factors and global physiology of the cell. *Molecular Systems Biology*, *9*(1), 634–634.

Blattner, F. R., Plunkett, G., Bloch, C., Perna, N., Burland, V., & Riley, M. (1997). The Complete Genome Sequence of Escherichia coli K-12. *Science*, *277*(5331), 1453–1462.

Blount, Z. D. (2015). The unexhausted potential of E. coli. *ELife*, *4*.

Boël, G., Letso, R., Neely, H., Price, W. N., Wong, K., Su, M., … Hunt, J. F. (2016). Codon influence on protein expression in E. coli correlates with mRNA levels. *Nature*, *529*, 358–363.

Bouvier, M., & Carpousis, A. J. (2011). A tale of two mRNA degradation pathways mediated by RNase E: RNase E pathways. *Molecular Microbiology*, *82*(6), 1305–1310.

Brantl, S. (2007). Regulatory mechanisms employed by cis-encoded antisense RNAs. *Current Opinion in Microbiology*, *10*(2), 102–109.

Bremer, H., & Dennis, P. (2008). Feedback control of ribosome function in Escherichia coli. *Biochimie*, *90*(3), 493–499.

Bremer, Hans, & Dennis, P. P. (1996). Modulation of Chemical Composition and Other Parameters of the Cell by Growth Rate. *ASM Press Washington*, 1553–1569.

Bulmer, M. (1991). The Selection-Mutation-DriftTheory of Synonymous Codon Usage. *Genetics*, *129*, 897–907.

Burgess-Brown, N. A., Sharma, S., Sobott, F., Loenarz, C., Oppermann, U., & Gileadi, O. (2008). Codon optimization can improve expression of human genes in Escherichia coli: A multi-gene study. *Protein Expression and Purification*, *59*(1), 94–102.

Burkhardt, D. H., Rouskin, S., Zhang, Y., Li, G.-W., Weissman, J. S., & Gross, C. A. (2017). Operon mRNAs are organized into ORF-centric structures that predict translation efficiency. *ELife*, *6*, e22037.

Cambray, G., Guimaraes, J. C., & Arkin, A. P. (2018). Evaluation of 244,000 synthetic sequences reveals design principles to optimize translation in Escherichia coli. *Nature Biotechnology*.

Capaldi, R. A., Schulenberg, B., Murray, J., & Aggeler, R. (2000). Structure and function of E. coli ATP synthase. *The Journal of Experimental Biology*, *203*, 29–33.

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. *Proceedings of the National Academy of Sciences*, *109*(50), E3444–E3453.

Carpousis, A. J. (2007). The RNA Degradosome of *Escherichia coli* : An mRNA-Degrading Machine Assembled on RNase E. *Annual Review of Microbiology*, *61*(1), 71–87.

Carrier, T. A., & Keasling, J. D. (1997). Controlling Messenger RNA Stability in Bacteria: Strategies for Engineering Gene Expression. *Biotechnology Progress*, *13*(6), 699–708.

Chan, L. Y., Mugler, C. F., Heinrich, S., Vallotton, P., & Weis, K. (2018). Non-invasive measurement of mRNA decay reveals translation initiation as the major determinant of mRNA stability. *ELife*, *7*(e32536), 32.

Chandu, D., & Nandi, D. (2004). Comparative genomics and functional roles of the ATP-dependent proteases Lon and Clp during cytosolic protein degradation. *Research in Microbiology*, *155*(9), 710–719.

Chang, B., Halgamuge, S., & Tang, S.-L. (2006). Analysis of SD sequences in completed microbial genomes: Non-SD-led genes are as common as SD-led genes. *Gene*, *373*, 90–99.

Chang, D.-E., Smalley, D. J., Tucker, D. L., Leatham, M. P., Norris, W. E., Stevenson, S. J., … Conway, T. (2004). Carbon nutrition of Escherichia coli in the mouse intestine. *Proceedings of the National Academy of Sciences*, *101*(19), 7427–7432.

Chassé, H., Boulben, S., Costache, V., Cormier, P., & Morales, J. (2016). Analysis of translation using polysome profiling. *Nucleic Acids Research*, 45(3), e15.

Chen, H., Shiroguchi, K., Ge, H., & Xie, X. S. (2015). Genome-wide study of mRNA degradation and transcript elongation in Escherichia coli. *Molecular Systems Biology*, *11*(1), 781–781.

Cohen, S. N., Chang, A. C. Y., Boyer, H. W., & Helling, R. B. (1973). Construction of Biologically Functional Bacterial Plasmids In Vitro. *Proceedings of the National Academy of Sciences of the United States of America*, *70*(11), 3240–3244.

Cole, J. R., & Nomura, M. (1986). Changes in the half-life of ribosomal protein messenger RNA caused by translational repression. *Journal of Molecular Biology*, *188*(3), 383–392.

Crick, F. H. C., Barnett, L., Brenner, S., & Watts-Tobin, R. J. (1961). General Nature of the Genetic Code for Proteins. *Nature*, *192*(4809), 1227–1232.

Darfeuille, F., Unoson, C., Vogel, J., & Wagner, E. G. H. (2007). An Antisense RNA Inhibits Translation by Competing with Standby Ribosomes. *Molecular Cell*, *26*(3), 381–392.

de Smit, M. H., & van Duin, J. (1990). Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proceedings of the National Academy of Sciences*, *87*(19), 7668–7672.

Deana, A., & Belasco, J. G. (2005). Lost in translation: the influence of ribosomes on bacterial mRNA decay. *Genes & Development*, *19*(21), 2526–2533.

Del Campo, C., Bartholomäus, A., Fedyunin, I., & Ignatova, Z. (2015). Secondary Structure across the Bacterial Transcriptome Reveals Versatile Roles in mRNA Regulation and Function. *PLoS Genetics*, *11*(10), 1–24.

Dong, H., Nilsson, L., & Kurland, C. G. (1996). Co-variation of tRNA Abundance and Codon Usage inEscherichia coliat Different Growth Rates. *Journal of Molecular Biology*, *260*(5), 649–663.

Dressaire, C., Gitton, C., Loubière, P., Monnet, V., Queinnec, I., & Cocaign-Bousquet, M. (2009). Transcriptome and Proteome Exploration to Model Translation Efficiency and Protein Stability in Lactococcus lactis. *PLoS Computational Biology*, *5*(12), e1000606.

Dreyfus, M. (2009). *Chapter 11 Killer and Protective Ribosomes*. *Progress in Molecular Biology and Translational Science* (1st ed., Vol. 85). Elsevier Inc.

Ehrenberg, M., Bremer, H., & Dennis, P. P. (2013). Medium-dependent control of the bacterial growth rate. *Biochimie*, *95*(4), 643–658.

Esquerré, T., Laguerre, S., Turlan, C., Carpousis, A. J., Girbal, L., & 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 Research*, *42*(4), 2460–2472.

Esquerré, T., Moisan, A., Chiapello, H., Arike, L., Vilu, R., Gaspin, C.,Girbal, L. (2015). Genome-wide investigation of mRNA lifetime determinants in Escherichia coli cells cultured at different growth rates. *BMC Genomics*, *16*(1), 275.

Evfratov, S. A., Osterman, I. A., Komarova, E. S., Pogorelskaya, A. M., Rubtsova, M. P., Zatsepin, T. S.,Dontsova, O. A. (2017). Application of sorting and next generation sequencing to study 5′-UTR influence on translation efficiency in Escherichia coli. *Nucleic Acids Research*, *45*(6), 3487–3502.

Fernandes, L. D., Moura, A. P. S. D., & Ciandrini, L. (2017). Gene length as a regulator for ribosome recruitment and protein synthesis: Theoretical insights. *Scientific Reports*, *7*(1), 1–11.

Fluman, N., Navon, S., Bibi, E., & Pilpel, Y. (2014). mRNA-programmed translation pauses in the targeting of E. coli membrane proteins. *ELife*, *3*(August2014), 1–19.

Frumkin, I., Lajoie, M. J., Gregg, C. J., Hornung, G., Church, G. M., & Pilpel, Y. (2018). Codon usage of highly expressed genes affects proteome-wide translation efficiency. *Proceedings of the National Academy of Sciences*, *115*(21), E4940–E4949.

Frumkin, I., Schirman, D., Rotman, A., Li, F., Zahavi, L., Mordret, E., Pilpel, Y. (2017). Gene Architectures that Minimize Cost of Gene Expression. *Molecular Cell*, *65*(1), 142–153.

Gardin, J., Yeasmin, R., Yurovsky, A., Cai, Y., Skiena, S., & Futcher, B. (2014). Measurement of average decoding rates of the 61 sense codons in vivo. *ELife*, *3*, 1–20.

Gerosa, L., Kochanowski, K., Heinemann, M., & Sauer, U. (2014). Dissecting specific and global transcriptional regulation of bacterial gene expression. *Molecular Systems Biology*, *9*(1), 658–658.

Goodman, D. B., Church, G. M., & Kosuri, S. (2013). Causes and Effects of N-Terminal Codon Bias in Bacterial Genes. *Science*, *342*(6157), 475–479.

Gorochowski, T. E., Ignatova, Z., Bovenberg, R. A. L., & Roubos, J. A. (2015). Trade-offs between tRNA abundance and mRNA secondary structure support smoothing of translation elongation rate. *Nucleic Acids Research*, *43*(6), 3022–3032.

Guimaraes, J. C., Rocha, M., & Arkin, A. P. (2014). Transcript level and sequence determinants of protein abundance and noise in Escherichia coli. *Nucleic Acids Research*, *42*(8), 4791–4799.

Gustafsson, C., Govindarajan, S., & Minshull, J. (2004). Codon bias and heterologous protein expression. *Trends in Biotechnology*, *22*(7), 346–353.

Hayes, C. S., & Sauer, R. T. (2003). Cleavage of the A Site mRNA Codon during Ribosome Pausing Provides a Mechanism for Translational Quality Control. *Molecular Cell*, *12*(4), 903–911.

Hendrickson, D. G., Hogan, D. J., McCullough, H. L., Myers, J. W., Herschlag, D., Ferrell, J. E., & Brown, P. O. (2009). Concordant Regulation of Translation and mRNA Abundance for Hundreds of Targets of a Human microRNA. *PLoS Biology*, *7*(11), e1000238.

Hockenberry, A. J., Pah, A. R., Jewett, M. C., & Amaral, L. A. N. (2017). Leveraging genome-wide datasets to quantify the functional role of the anti-Shine-Dalgarno sequence in regulating translation efficiency. *Open Biology*, *7*(1), 160239.

Holmqvist, E., & Vogel, J. (2018). RNA-binding proteins in bacteria. *Nature Reviews Microbiology*, *16*(10), 601–615.

Hu, X.-P., Yang, Y., & Ma, B.-G. (2015). Amino Acid Flux from Metabolic Network Benefits Protein Translation: the Role of Resource Availability. *Scientific Reports*, *5*, 11113.

Huang, C.-J., Lin, H., & Yang, X. (2012). Industrial production of recombinant therapeutics in Escherichia coli and its recent advancements. *Journal of Industrial Microbiology & Biotechnology*, *39*(3), 383–399.

Ingolia, N. T. (2014). Ribosome profiling: new views of translation, from single codons to genome scale. *Nature Reviews Genetics*, *15*(3), 205–213.

Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S., & Weissman, J. S. (2009). Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science (New York, N.Y.)*, *324*(5924), 218–223.

Iost, I., & Dreyfus, M. (1995). The stability of Escherichia coli lacZ mRNA depends upon the simultaneity of its synthesis and translation. *The EMBO Journal*, *14*(13), 3252–3261.

Jacob, F., & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, *3*(3), 318–356.

Joyce, S. A., & Dreyfus, M. (1998). In the Absence of Translation, RNase E can Bypass 5H mRNA Stabilizers in Escherichia coli. *Journal of Molecular Biology*, *282*, 241–254.

Kærn, Mads, Timothy C. Elston, William J. Blake, and James J. Collins. (2005). Stochasticity in Gene Expression: From Theories to Phenotypes. *Nature Reviews Genetics* 6(6):451–64.

Kamionka, M. (2011). Engineering of Therapeutic Proteins Production in Escherichia coli. *Current Pharmaceutical Biotechnology*, *12*(2), 268–274.

Karp, P. D., Keseler, I. M., Shearer, A., Latendresse, M., Krummenacker, M., Paley, S. M., … Ingraham, J. (2007). Multidimensional annotation of the Escherichia coli K-12 genome. *Nucleic Acids Research*, *35*(22), 7577–7590.

Kertesz, M., Wan, Y., Mazor, E., Rinn, J. L., Nutter, R. C., Chang, H. Y., & Segal, E. (2010). Genome-wide measurement of RNA secondary structure in yeast. *Nature*, *467*(7311), 103–107.

Klumpp, S., & Hwa, T. (2008). Growth-rate-dependent partitioning of RNA polymerases in bacteria. *Proceedings of the National Academy of Sciences*, *105*(51), 20245–20250.

Klumpp, Stefan, Zhang, Z., & Hwa, T. (2009). Growth-rate dependent global effects on gene expression in bacteria. *Cell*, *139*(7), 1366.

Klumpp, S., M. Scott, S. Pedersen, and T. Hwa. (2013). Molecular Crowding Limits Translation and Cell Growth. *Proceedings of the National Academy of Sciences* 110(42):16754–59.

Kohler, R., Mooney, R. A., Mills, D. J., Landick, R., & Cramer, P. (2017). Architecture of a transcribing-translating expressome. *Science*, *356*(6334), 194–197.

Korkmazhan, E., Teimouri, H., Peterman, N., & Levine, E. (2017). Dynamics of translation can determine the spatial organization of membrane-bound proteins and their mRNA. *Proceedings of the National Academy of Sciences*, *114*(51), 201700941.

Kosuri, S., Goodman, D. B., Cambray, G., Mutalik, V. K., Gao, Y., Arkin, A. P., … Church, G. M. (2013). Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. *Proceedings of the National Academy of Sciences*, *110*(34), 14024–14029.

Kudla, G., Murray, A. W., Tollervey, D., & Plotkin, J. B. (2009). Coding-Sequence Determinants of Gene Expression in Escherichia coli, *324*, 4.

Kuhn, K. M., DeRisi, J. L., Brown, P. O., & Sarnow, P. (2001). Global and Specific Translational Regulation in the Genomic Response of Saccharomyces cerevisiae to a Rapid Transfer from a

Fermentable to a Nonfermentable Carbon Source. *Molecular and Cellular Biology*, *21*(3), 916–927.

Lackner, D. H., Beilharz, T. H., Marguerat, S., Mata, J., Watt, S., Schubert, F., Bähler, J. (2007). A Network of Multiple Regulatory Layers Shapes Gene Expression in Fission Yeast. *Molecular Cell*, *26*(1), 145–155.

Lacsina, J. R., LaMonte, G., Nicchitta, C. V., & Chi, J.-T. (2011). Polysome profiling of the malaria parasite Plasmodium falciparum. *Molecular and Biochemical Parasitology*, *179*(1), 42–46.

Lange, C., Lehr, M., Zerulla, K., Ludwig, P., Schweitzer, J., Polen, T., Soppa, J. (2017). Effects of Kasugamycin on the Translatome of Escherichia coli. *PLOS ONE*, *12*(1), e0168143.

Lange, C., Zaigler, A., Hammelmann, M., Twellmeyer, J., Raddatz, G., Schuster, S. C., Soppa, J. (2007). Genome-wide analysis of growth phase-dependent translational and transcriptional regulation in halophilic archaea. *BMC Genomics*, *8*, 1–17.

Laursen, B. S., Sørensen, H. P., Mortensen, K. K., & Sperling-Petersen, H. U. (2005). Initiation of Protein Synthesis in Bacteria. *Microbiology and Molecular Biology Reviews*, *69*(1), 101–123.

Lehman, I. R., Zimmerman, S. B., Adler, J., Bessman, M. J., Simms, E. S., & Kornberg, A. (1958). Enzymatic synthesis of deoxyribonucleic acid. *Proceedings of the National Academy of Sciences of the United States of America*, *44*(12), 1191–1196.

Levin-Karp, A., Barenholz, U., Bareia, T., Dayagi, M., Zelcbuch, L., Antonovsky, N., Milo, R. (2013). Quantifying translational coupling in E. coli synthetic operons using RBS modulation and fluorescent reporters. *ACS Synthetic Biology*, *2*(6), 327–336.

Li, G. W., Burkhardt, D., Gross, C., & Weissman, J. S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell*, *157*(3), 624–635.

Li, G.-W., Oh, E., & Weissman, J. S. (2012). The anti-Shine–Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature*, *484*(7395), 538–541.

Linn, S., & Arber, W. (1968). Host specificity of DNA produced by Escherichia coli, X. In vitro restriction of phage fd replicative form. *Proceedings of the National Academy of Sciences of the United States of America*, *59*(4), 1300–1306.

Lithwick, G. (2003). Hierarchy of Sequence-Dependent Features Associated With Prokaryotic Translation. *Genome Research*, *13*(12), 2665–2673.

Lu, P., Vogel, C., Wang, R., Yao, X., & Marcotte, E. M. (2007). Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nature Biotechnology*, *25*(1), 117–124.

Martínez-Antonio, A. (2011). Escherichia coli transcriptional regulatory network. Network Biology, 2011, 1(1):21-33.

McGary, K., & Nudler, E. (2013). RNA polymerase and the ribosome: the close relationship. *Current Opinion in Microbiology*, *16*(2), 112–117.

Melamed, D., Pnueli, L., & Arava, Y. (2008). Yeast translational response to high salinity: global analysis reveals regulation at multiple levels. *RNA (New York, N.Y.)*, *14*(7), 1337–1351.

Meselson, M., & Yuan, R. (1968). DNA Restriction Enzyme from E. coli. *Nature*, *217*(5134), 1110.

Moll, I., Grill, S., Gualerzi, C. O., & Bläsi, U. (2002). Leaderless mRNAs in bacteria: surprises in ribosomal recruitment and translational control. *Molecular Microbiology*, *43*(1), 239–246.

Møller, T., Franch, T., Udesen, C., Gerdes, K., & Valentin-Hansen, P. (2002). Spot 42 RNA mediates discoordinate expression of the E. coli galactose operon. *Genes & Development*, *16*(13), 1696–1706.

Morgan, G. J., Burkhardt, D. H., Kelly, J. W., & Powers, E. T. (2018). Translation efficiency is maintained at elevated temperature in *Escherichia coli*. *Journal of Biological Chemistry*, *293*(3), 777–793.

Mori, M. (2017). Quantifying the benefit of a proteome reserve in fluctuating environments. *Nature Communications*, *8*(1).

Morisaki, T., K. Lyon, K. F. DeLuca, J. G. DeLuca, B. P. English, Z. Zhang, L. D. Lavis, J. B. Grimm, S. Viswanathan, L. L. Looger, T. Lionnet, and T. J. Stasevich. (2016). Real-Time Quantification of Single RNA Translation Dynamics in Living Cells. *Science* 352(6292):1425–29.

Nakahigashi, K., Takai, Y., Shiwa, Y., Wada, M., Honma, M., Yoshikawa, H., … Mori, H. (2014). Effect of codon adaptation on codon-level and gene-level translation efficiency in vivo. *BMC Genomics*, *15*(1), 1115.

Nie, L., Wu, G., & Zhang, W. (2006). Correlation of mRNA Expression and Protein Abundance Affected by Multiple Sequence Features Related to Translational Efficiency in *Desulfovibrio vulgaris* : A Quantitative Analysis. *Genetics*, *174*(4), 2229–2243.

O'Brien, E. J., Utrilla, J., & Palsson, B. O. (2016). Quantification and Classification of E. coli Proteome Utilization and Unused Protein Costs across Environments. *PLOS Computational Biology*, *12*(6), e1004998.

Oh, E., Becker, A. H., Sandikci, A., Huber, D., Chaba, R., Gloge, F., … Bukau, B. (2011). Selective Ribosome Profiling Reveals the Cotranslational Chaperone Action of Trigger Factor In Vivo. *Cell*, *147*(6), 1295–1308.

Osada, Y., Saito, R., & Tomita, M. (1999). Analysis of base-pairing potentials between 16S rRNA and 5' UTR for translation initiation in various prokaryotes. *Bioinformatics*, *15*(7), 578–581.

Picard, F., Dressaire, C., Girbal, L., & Cocaign-Bousquet, M. (2009). Examination of post-transcriptional regulations in prokaryotes by integrative biology. *Comptes Rendus Biologies*, *332*(11), 958–973.

Picard, F., Loubière, P., Girbal, L., & Cocaign-Bousquet, M. (2013). The significance of translation regulation in the stress response. *BMC Genomics*, *14*(1), 588.

Picard, F., Milhem, H., Loubière, P., Laurent, B., Cocaign-Bousquet, M., & Girbal, L. (2012). Bacterial translational regulations: high diversity between all mRNAs and major role in gene expression. *BMC Genomics*, *13*(1), 528.

Plotkin, J. B., & Kudla, G. (2011). Synonymous but not the same: the causes and consequences of codon bias. *Nature Reviews Genetics*, *12*(1), 32–42.

Pop, C., Rouskin, S., Ingolia, N. T., Han, L., Phizicky, E. M., Weissman, J. S., & Koller, D. (2014). Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. *Molecular Systems Biology*, *10*(12), 770–770.

Presnyak, V., Alhusaini, N., Chen, Y. H., Martin, S., Morris, N., Kline, N., … Coller, J. (2015). Codon optimality is a major determinant of mRNA stability. *Cell*, *160*(6), 1111–1124.

Price, M. N., Wetmore, K. M., Deutschbauer, A. M., & Arkin, A. P. (2016). A Comparison of the Costs and Benefits of Bacterial Gene Expression. *PLOS ONE*, 22.

Proshkin, S., Rahmouni, R., Mironov, A., & Nudler, E. (2010). Cooperation between Translating Ribosomes and RNA Polymerase in Transcription Elongation. *Science (New York, N.Y.)*, *328*(5977), 504–508. 9

Qi, F., Motz, M., Jung, K., Lassak, J., & Frishman, D. (2018). Evolutionary analysis of polyproline motifs in Escherichia coli reveals their regulatory role in translation. *PLOS Computational Biology*, *14*(2), e1005987.

Qian, W., Yang, J.-R., Pearson, N. M., Maclean, C., & Zhang, J. (2012). Balanced Codon Usage Optimizes Eukaryotic Translational Efficiency. *PLOS Genetics*, *8*(3), e1002603.

Quax, T. E. F., Wolf, Y. I., Koehorst, J. J., Wurtzel, O., VanderOost, R., Ran, W., … VanderOost, J. (2013). Differential translation tunes uneven production of operon-encoded proteins. *Cell Reports*, *4*(5), 938–944.

Ramakrishnan, V. (2002). Ribosome Structure and the Mechanism of Translation. *Cell*, *108*(4), 557–572.

Reis, M. d. (2004). Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Research*, *32*(17), 5036–5044.

Resch, A., Afonyushkin, T., Lombo, T. B., McDowall, K. J., Bläsi, U., & Kaberdin, V. R. (2008). Translational activation by the noncoding RNA DsrA involves alternative RNase III processing in the rpoS 5'-leader. *RNA*, *14*(3), 454–459.

Riley, M., Abe, T., Arnaud, M., Berlyn, M., Blattner, F., & Chaudhuri, R. (2006). Escherichia coli K-12: a cooperatively developed annotation snapshot--2005. *Nucleic Acids Research*, *34*(1), 1–9.

Ringquist, S., Shinedling, S., Barrick, D., Green, L., Binkley, J., Stormo, G. D., & Gold, L. (1992). Translation initiation in Escherichia coli: sequences within the ribosome-binding site. *Molecular Microbiology*, *6*(9), 1219–1229.

Rocha, E. P. C., Danchin, A., & Viari, A. (1999). Translation in Bacillus subtilis: roles and trends of initiation and termination, insights from a genome analysis. *Nucleic Acids Research*, *27*(17), 3567–3576.

Ruiz-Larrabeiti, O., Plágaro, A. H., Gracia, C., Sevillano, E., Gallego, L., Hajnsdorf, E., & Kaberdin, V. R. (2016). A new custom microarray for sRNA profiling in *Escherichia coli*. *FEMS Microbiology Letters*, *363*(13), fnw131.

Salis, H. M., Mirsky, E. A., & Voigt, C. A. (2009). Automated design of synthetic ribosome binding sites to control protein expression. *Nature Biotechnology*, *27*(10), 946–950.

Schmeing, T. M., & Ramakrishnan, V. (2009). What recent ribosome structures have revealed about the mechanism of translation. *Nature*, *461*(7268), 1234–1242.

Schneider, D. A., Ross, W., & Gourse, R. L. (2003). Control of rRNA expression in Escherichia coli. *Current Opinion in Microbiology*, *6*(2), 151–156.

Shah, P., Ding, Y., Niemczyk, M., Kudla, G., & Plotkin, J. B. (2013). Rate-limiting steps in yeast protein translation. *Cell*, *153*(7), 1589–1601.

Shajani, Z., Sykes, M. T., & Williamson, J. R. (2011). Assembly of Bacterial Ribosomes. *Annual Review of Biochemistry*, *80*(1), 501–526.

Sharp, P. M., & Li, W. (1987). The codon adaptation index - a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research*, *15*(3), 1281–1295.

Shine, J., & Dalgarno, L. (1974). The 3'-Terminal Sequence of Escherichia coli 16S Ribosomal RNA: Complementarity to Nonsense Triplets and Ribosome Binding Sites. *Proceedings of the National Academy of Sciences of the United States of America*, *71*(4), 1342–1346.

Sin, C., Chiarugi, D., & Valleriani, A. (2016). Quantitative assessment of ribosome drop-off in *E. coli*. *Nucleic Acids Research*, *44*(6), 2528–2537.

Soper, T. J., & Woodson, S. A. (2008). The rpoS mRNA leader recruits Hfq to facilitate annealing with DsrA sRNA. *RNA*, *14*(9), 1907–1917.

Sorensen, M. A., Fehler, A. O., & Lo Svenningsen, S. (2018). Transfer RNA instability as a stress response in *Escherichia coli* : Rapid dynamics of the tRNA pool as a function of demand. *RNA Biology*, *15*(4–5), 586–593.

Stevens, A. (1960). Incorporation of the adenine ribonucleotide into RNA by cell fractions from E. coli. *Biochemical and biophysical research communications, 3*(1), 5.

Storz, G., Vogel, J., & Wassarman, K. M. (2011). Regulation by Small RNAs in Bacteria: Expanding Frontiers. *Molecular Cell*, *43*(6), 880–891.

Subramaniam, A. R., Zid, B. M., & O'Shea, E. K. (2014). An Integrated Approach Reveals Regulatory Controls on Bacterial Translation Elongation. *Cell*, *159*(5), 1200–1211.

Svenningsen, S. L., Kongstad, M., Stenum, T. S., Muñoz-Gómez, A. J., & Sørensen, M. A. (2017). Transfer RNA is highly unstable during early amino acid starvation in *Escherichia coli*. *Nucleic Acids Research*, *45*(2), 793–804.

Taniguchi, Y., Choi, P. J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Xie, X. S. (2010). Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *Science*, *329*(5991), 533–538.

Tian, T., & Salis, H. M. (2015). A predictive biophysical model of translational coupling to coordinate and control protein expression in bacterial operons. *Nucleic Acids Research*, *43*(5), 1–15.

Timmermans, J., & Van Melderen, L. (2010). Post-transcriptional global regulation by CsrA in bacteria. *Cellular and Molecular Life Sciences*, *67*(17), 2897–2908.

Tsai, Y.-C., Du, D., Domínguez-Malfavón, L., Dimastrogiovanni, D., Cross, J., Callaghan, A. J., Luisi, B. F. (2012). Recognition of the 70S ribosome and polysome by the RNA degradosome in Escherichia coli. *Nucleic Acids Research*, *40*(20), 10417–10431.

Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pilpel, Y. (2010). An Evolutionarily Conserved Mechanism for Controlling the Efficiency of Protein Translation. *Cell*, *141*(2), 344–354.

Tuller, T., Waldman, Y. Y., Kupiec, M., & Ruppin, E. (2010). Translation efficiency is determined by both codon bias and folding energy. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(8), 3645–3650.

Udagawa, T., Shimizu, Y., & Ueda, T. (2004). Evidence for the Translation Initiation of Leaderless mRNAs by the Intact 70 S Ribosome without Its Dissociation into Subunits in Eubacteria. *Journal of Biological Chemistry*, *279*(10), 8539–8546.

Valgepea, K., Adamberg, K., Nahku, R., Lahtvee, P.-J., Arike, L., & 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 Systems Biology*, *4*(1), 166.

Vogel, J., Argaman, L., Wagner, E. G. H., & Altuvia, S. (2004). The Small RNA IstR Inhibits Synthesis of an SOS-Induced Toxic Peptide. *Current Biology*, *14*(24), 2271–2276.

Warringer, J., Hult, M., Regot, S., Posas, F., & Sunnerhagen, P. (2010). The HOG Pathway Dictates the Short-Term Translational Response after Hyperosmotic Shock. *Molecular Biology of the Cell*, *21*(17), 3080–3092.

Yates, J. L., & Nomura, M. (1981). Feedback regulation of ribosomal protein synthesis in E. coli: localization of the mRNA target sites for repressor action of ribosomal protein L1. *Cell*, *24*(1), 243–249.

Yu, C.-H., Dang, Y., Zhou, Z., Wu, C., Zhao, F., Sachs, M. S., & Liu, Y. (2015). Codon Usage Influences the Local Rate of Translation Elongation to Regulate Co-translational Protein Folding. *Molecular Cell*, *59*(5), 744–754.

Zhang, Yan, Burkhardt, D. H., Rouskin, S., Li, G.-W., Weissman, J. S., & Gross, C. A. (2018). A Stress Response that Monitors and Regulates mRNA Structure Is Central to Cold Shock Adaptation. *Molecular Cell*, *70*(2), 274-286.e7.

Zhang, Yanqing, Xiao, Z., Zou, Q., Fang, J., Wang, Q., Yang, X., & Gao, N. (2017). Ribosome Profiling Reveals Genome-wide Cellular Translational Regulation upon Heat Stress in Escherichia coli. *Genomics, Proteomics & Bioinformatics*, *15*(5), 324–330.

Zhong, J., Xiao, C., Gu, W., Du, G., Sun, X., He, Q. Y., & Zhang, G. (2015). Transfer RNAs Mediate the Rapid Adaptation of Escherichia coli to Oxidative Stress. *PLoS Genetics*, *11*(6), 1–24.

Zouridis, H., & Hatzimanikatis, V. (2007). A model for protein translation: Polysome self-organization leads to maximum protein synthesis rates. *Biophysical Journal*, *92*(3), 717–730.

- Chapitre 1 -

Multiplexing of polysome profiling experiments to study translation in *Escherichia coli*

Dans ce premier chapitre, nous présentons une méthode de multiplexage de polysome profiling. Le polysome profiling est la méthode classique largement utilisée pour étudier l'état traductionnel des ARNm. Le polysome profiling permet de séparer les ARNm en fonction de leurs charges en ribosomes et ainsi d'accéder au niveau de traduction des ARNm. Un premier niveau d'analyse de polysome profiling est de séparer les ARNm en deux groupes, les ARNm libres et ceux associés aux ribosomes, ce qui permet d'étudier des différences majeures de traduction (variation de la proportion d'ARNm en traduction ou non). Cependant, pour pouvoir observer plus finement des différences de traduction, il est nécessaire d'avoir un niveau de résolution plus élevé. Pour cela, le polysome est divisé en un nombre plus élevé de fractions. Cela génère un grand nombre d'échantillons à gérer lors des étapes d'élimination des protéines, d'extraction et quantification des ARN, ainsi augmente les biais techniques, le coût et le temps de l'expérimentation. Nous avons donc développé pour la première fois une méthode de multiplexage de polysome profiling afin de réduire les biais techniques, le temps et le coût des manipulations. Ainsi l'approche de polysome profiling pourra être plus facilement utilisée pour étudier l'état traductionnel de plusieurs ARNm cibles en parallèle.

Cette méthode de multiplexage consiste à assembler différents extraits cellulaires qui expriment des ARNm rapporteurs différents en séquence avant de les charger sur le même gradient de sucrose. Une quantification spécifique des ARNm cibles est effectuée par RTqPCR. Dans un premier temps, le multiplexage a été validé avec le gène *ihfB* car les mêmes valeurs de pourcentage de copies en traduction (RO) et de densité en ribosomes (RD) ont été obtenues avec et sans multiplexage.

Dans un deuxième temps, nous avons appliqué cette méthode de multiplexage à l'étude de l'effet du facteur « concentration en ARNm » sur la traduction. Différentes souches d'*E. coli* ont été construites afin de pouvoir exprimer différents ARNm rapporteurs sous le contrôle d'un promoteur inductible. La concentration de l'ARNm rapporteur est augmentée par induction transcriptionnelle et les profils polysomiques de l'ARNm à faible concentration (en absence d'induction) et à forte concentration (après induction) sont comparés. La réponse traductionnelle des deux gènes *cysZ* et *lacZ* a été analysée simultanément avec la méthode de multiplexage. Pour les deux gènes, nous avons obtenu le même type de réponse traductionnelle face à une augmentation de la concentration de leur ARNm : leur niveau de traduction augmente

avec un pourcentage de copies en traduction et des charges en ribosomes plus élevés. Le multiplexage a aussi permis d'observer une certaine spécificité des réponses. En effet, l'identité des fractions enrichies en ribosomes et l'amplitude des réponses diffèrent entre *lacZ* et *cysZ*.

En conclusion, ce travail a permis de valider l'utilisation d'une approche de multiplexage de polysome profiling pour étudier la réponse traductionnelle de plusieurs gènes en parallèle. Cette méthode sera utilisée dans le chapitre III pour valider au niveau moléculaire l'effet prédit par une analyse statistique de la concentration de l'ARNm sur sa traduction. Les réponses traductionnelles de six gènes rapporteurs à une augmentation de leur concentration en ARNm seront mesurées de manière simultanée.

L'utilisation de cette méthode de multiplexage de polysome profiling pourrait être étendue à l'étude d'autres déterminants de la traduction, comme des paramètres de séquence de l'ARNm (la séquence du RBS, l'usage des codons, la présence de mutation ponctuelle entre allèles de souches génétiquement proches). Elle peut aussi être étendue à l'étude de la traduction chez d'autres microorganismes.

Cette étude a fait l'objet d'une publication en 2019 dans le journal Plos One (cf Annexe) et est présentée ici sous la forme du manuscrit accepté.

Multiplexing polysome profiling experiments to study translation in *Escherichia coli*

Huong Le Nguyen, Marie-Pierre Duviau, Muriel Cocaign-Bousquet*, Sébastien Nouaille, Laurence Girbal*

LISBP, Université de Toulouse, CNRS, INRA, INSA, Toulouse, France

* Corresponding authors
Email: cocaign@insa-toulouse.fr (MCB)
Email: girbal@insa-toulouse.fr (LG)

KEYWORDS: Multiplexing, polysome profiling, translational response, ribosome density, ribosome occupancy, mRNA level, transcriptional level

Abstract

Polysome profiling is a widely used method to monitor the translation status of mRNAs. Although it is theoretically a simple technique, it is labor intensive. Repetitive polysome fractionation rapidly generates a large number of samples to be handled in the downstream processes of protein elimination, RNA extraction and quantification. Here, we propose a multiplex polysome profiling experiment in which distinct cellular extracts are pooled before loading on the sucrose gradient for fractionation. We used the multiplexing method to study translation in *E. coli*. Multiplexing polysome profiling experiments provided similar mRNA translation status to that obtained with the non-multiplex method with comparable distribution of mRNA copies between the polysome profiling fractions, similar ribosome occupancy and ribosome density. The multiplexing method was used for parallel characterization of gene translational responses to changing mRNA levels. When the mRNA level of two native genes, *cysZ* and *lacZ* was increased by transcription induction, their global translational response was similar, with a higher ribosome load leading to increased ribosome occupancy and ribosome densities. However the pattern and the magnitude of the translational response were gene specific. By reducing the number of polysome profiling experiments, the multiplexing method saved time and effort and reduced cost and technical bias. This method would be useful to study the translational effect of mRNA sequence-dependent parameters that often require testing multiple samples and conditions in parallel.

Introduction

In the last decade, interest in the role of regulating translation has been increasing. Translation regulation plays an essential role in fine-tuning gene expression and protein level (1). It allows a rapid response to extracellular stimuli, which can be crucial for adaptation to different environmental conditions (2). Polysome profiling is a widely used method to study translation status. For each individual gene, the method quantifies the number of ribosomes bound to each copy of the mRNA molecule and provides a detailed distribution of the mRNA copies per number of bound ribosomes (proportions of free mRNA copies, of monosome-bound and polysome-bound mRNA copies (3–5). Polysome profiling enables the definition of two translational variables, ribosome occupancy (RO) and ribosome density (RD) (6–8). RO is the proportion of mRNA copies bound with at least one ribosome and can be used as a proxy for

translation initiation. RD can be calculated as the number of mRNA-bound ribosomes in each polysome profiling fraction normalized to the coding sequence length; it reflects the level of translation initiation, elongation and termination. Thus, polysome profiling not only assesses heterogeneity in the number of bound ribosomes within the copies of an mRNA, but also accesses physical ribosome density by measuring joint binding of multiple ribosomes on the same transcript. This technique is complementary to the recently expanding method of ribosome profiling that quantifies the heterogeneity of ribosome position occupation averaged over the population of mRNA copies.

The polysome profiling method can be used to study translational status in different organisms at different stages of growth and development (e.g. *Arabidopsis thaliana* (9), in sea urchin (10), in halophilic archaea (3)). It is also used to understand translational response to various stresses: marine organisms under oxidative stress (11), yeast and *A. thaliana* in high salinity conditions (12,13), yeast under Zn limitation (14) and *Lactococcus lactis* and yeast under nutrient starvation (8,15). It is also frequently used in mechanistic studies of translation regulation, for instance to characterize the effects of elongation factors (16,17). The role of mRNA sequence related parameters such as 5'UTR and codon usage on translation has also been investigated using polysome profiling (18–20).

These studies are usually limited to a small number of samples and conditions because polysome profiling is labor intensive. The drawbacks of the technique remain the main difficulty in handling many samples in parallel (21). Separation of mRNA-polysome complexes according to bound ribosomal loading consists in polysome fractionation on the sucrose gradient. This step generates numerous samples to be handled in the downstream processes of protein elimination, RNA extraction and quantification. This may introduce technical bias between different polysome profiling experiments and also entails rather expensive and time consuming experiments. We developed a multiplexing method for polysome profiling experiments that makes it possible to assemble six different cell free extracts before loading on the sucrose gradient. The RNA in each fraction was quantified by RT-qPCR with an optimized amount of exogenous RNA spike-ins. A challenge in multiplexing experiment is to differentiate the cell free extract origin of the measured mRNAs. In this study, we identified the origin of an mRNA of interest by strongly overexpressing this mRNA in

only one cell free extract of the mixture. The multiplexing polysome profiling method was then used to study translation regulation in the bacterial model *Escherichia coli* K12 MG1655. The translational states of different genes were simultaneously characterized by measuring their ribosome occupancies and ribosome densities. We also accessed the translational response of these genes to changing gene expression, a situation that may be encountered when *E. coli* cells need to adapt to variable growth environments.

Materials and methods

Plasmids and strain construction

All strains were constructed in the genetic background *E. coli* MG1655 Δ*araFGH*, Ω*pcp18*::*araE533* (22). MET734 and MET739 carried the *cysZ* and *lacZ* genes, respectively, on a plasmid under the P_{BAD} inducible promoter (Table 1). We selected these two native genes of *E. coli* MG1655 for their low level of expression in exponential growth in synthetic medium (23). For each, the 5'UTR + ORF fragment was amplified by PCR and cloned into the pBAD/myc/His plasmid (Invitrogen) to obtain the constructs: pBAD– 5'UTR+$ORF_{selected\ gene}$ – myc/His tag. The pBAD-5'UTR+ORF_{lacZ}-myc/His tag was introduced into the *E. coli* variant where the chromosomal copy of *lacZ* was deleted (24). Four other "filling" strains were used to mix their cell free extract with those of MET734 and MET739 for multiplexing purposes. In the same genetic background, the four strains contained genes (*yeeZ, inaA*, *ucpA* and *yjc*) not related to this study, under the control of $P_{BAD.}$

Table 1: Strain description

Strain	Description	Source
DCT2202	*E. coli* MG1655 Δ*araFGH*, Ω*pcp18*::*araE533*	(22)
MET345	DCT2202 Δ*lacZ*	(24)
MET739	MET345 with plasmid (pBAD-lacZ-*myc*-his)	This work
MET734	*E. coli* DCT2202 with plasmid (pBAD-cysZ-*myc*-his)	This work

Culture and preparation of cell lysate

Each strain was individually grown in chemically defined minimum medium M9 supplemented with 3 g/L glucose (23) , 0.1 mg/mL ampicillin, at 37 °C, under shaking (150 rpm). Arabinose was added at a final concentration of 0.001% (w/v) when the

culture reached an OD_{600} of 1 (exponential growth). After 30 minutes of induced gene expression, chloramphenicol was added at a final concentration of 0.1 mg/mL to stop translation elongation.

Cell culture was immediately transferred on ice for one minute. A fixed volume corresponding to 320 ml per OD unit was collected and centrifuged at 6,300 g, at 4 °C for 15 minutes. The supernatant was discarded and the cell pellet washed twice with cold lysis buffer (20 mM Tris HCl pH 8, 140 mM KCl, 40 mM $MgCl_2$, 0.5 mM DTT, 100 µg/mL chloramphenicol, 1 mg/mL heparin, 20 mM EGTA, 1% (v/v) Triton X-100) and resuspended in 1.2 mL of cold lysis buffer.

The cell suspension was transferred in cold screw-capped tubes containing 0.1 g of glass beads (0.1 mm diameter, Sigma) and disrupted using a FastPrep®-24 (MP Biomedicals). We observed that more RNA was obtained when we performed four 30 s cycles at 6.5 m/s with at least one minute on ice between cycles rather than two cycles. The lysate was first centrifuged for 5 minutes at 2,100 g at 4 °C to remove the glass beads. The supernatant was collected and centrifuged again for 5 minutes at 8,600 g, at 4 °C. Clarified lysate was gently collected, immediately frozen in liquid nitrogen and stored at -80 °C. The concentration of protein in the lysate was measured using a NanoDrop® ND-1000 UV spectrophotometer (Nanodrop Technologies). Starting from around 110 mg of dry cell weight, the protein yield was around 50-60 mg/mL. All the steps were performed at 4 °C and the samples were kept on ice.

Polysome profiling experiments

According to ribosome loading, the mRNA-ribosome complexes were separated on 10-50% (w/v) linear sucrose gradient (24 mL) prepared in cold lysis buffer. In the non-multiplexing experiment, individual cell free extract (≈ 2.4 mL) was loaded on the sucrose gradient, whereas in the multiplexing experiment, six cell free extracts were pooled to reach an equivalent protein amount per strain in a final volume of 2.4 mL. Ultracentrifugation was performed in a Sorvall WX80 (ThermoScientific) using a swing rotor AH-629, for 16h30min, at 23,700 g, at 4 °C. The sucrose gradient was eluted with cold buffer (55% sucrose (w/v), 0.5 mM Tris HCl pH 8, 4 µg/mL Bromophenol blue) in 24 sub-fractions at 2.5 mL/minute. Absorbance was continuously measured at 254 nm with a UV detector (UPC900 Amersham Pharmacia Biotech).

Protein elimination and RNA extraction

Protein denaturation and nucleic acid precipitation were performed by adding one volume of 8 M guanidium-HCl, two volumes of absolute ethanol and overnight storage at -20°C. After 30 minutes centrifugation at 13,300 g at 4 °C, the supernatant containing the free proteins was gently removed and the pellet of nucleic acids (including mRNA loaded with ribosomes) were washed with cold 75% ethanol (v/v) and resuspended in TE buffer (10 mM Tris HCl pH 8, 1 mM EDTA). Total RNA extraction and purification were performed using the extraction RNeasy Midi kit (Qiagen). Genomic DNA was removed by on-column DNase digestion using 90U of RNase-free DNase I (Qiagen) for 15 minutes at room temperature. The total RNA concentration was measured using NanoDrop® ND-1000. RNA integrity was validated and 16S and 23S rRNA were quantified using Bioanalyzer 2100 (Agilent Technologies®). Total RNA samples were stored at -80 °C.

Reverse transcription and RNA quantification by real-time quantitative PCR

Total RNA (5 µg) was reverse-transcribed to yield cDNA using 200U of SuperScript II reverse transcriptase (Invitrogen) as previously described (25) . cDNA was quantified by Real Time PCR Detection system (Bio-Rad) in 96-well plates as previously described (26). cDNA dilutions of 10^{-1} and 10^{-2} were used to provide quantifiable signals, i.e. cycle threshold (Ct) of between 15 and 25. For large numbers of samples, a high-throughput qPCR technique was applied using Biomark™ HD System (Fluidigm Corporation, CA, USA) as previously described (24).

To account for variability of the reverse transcription and the qPCR steps between samples and experiments, control Ambion™ ERCC RNA Spike-In mix was used as external normalizer. For each fraction, an equal amount of ERCC was added to a constant amount of total RNA and they were then reverse transcribed together. To improve the efficiency and reproducibility of the reverse transcription of ERCC, 0.2 µM of reverse primers specific to the four most concentrated ERCCs (ERCC 130, ERCC 002, ERCC 074 and ERCC 096) were added during reverse transcription in addition to random primers.

A total of 12 different mRNAs and four ERCCs were quantified in this work (S1 Table). Quantification of *lacZ* mRNA is the average value obtained from five primer pairs. Primers for qPCR were designed for these 20 genes using Vector NTI advance

v11 (Life Technologies) using a melting temperature of 59-61 °C, length of 18-20 bp and 50-70% GC content (S1 Table). Amplicon sizes ranged between 80 and 150 bp. The reaction efficiency was tested on cDNA serial dilutions and focused around 100%.

Data normalization and analysis

To calculate the relative amount of a target mRNA in each fraction, two normalizations were applied. First, relative mRNA abundance compared to a constant quantity of ERCC was calculated using the method of fold change ΔCt values (27). As only 5 µg of the total RNA amount extracted in each fraction was used in the RT-qPCR experiment, we normalized the relative mRNA abundance by the total RNA quantity extracted in each fraction to obtain the relative initial mRNA abundance in each fraction.

For each target gene, its relative initial mRNA abundance compared to ERCC in fraction *i* was calculated as follows:

$$Relative\ initial\ mRNA\ abundace = 2^{(Ct\ ERCC_i - Ct\ target_i)} * \frac{total\ RNA\ quantity_i}{5}$$

To obtain the distribution of the abundance of mRNA copies, the proportion of mRNA copies in each fraction was calculated by dividing the relative initial abundance in one fraction by the sum of the abundances in all the fractions:

$$Proportion\ in\ fraction\ i = \frac{Relative\ initial\ abundance\ in\ fraction\ i}{\sum_{j=1}^{7} Relative\ initial\ abundance\ in\ fraction\ j} * 100$$

Calculation of ribosome occupancy and ribosome density

For each gene, the ribosome occupancy was the proportion of the mRNA copies undergoing translation. It was calculated by summing the proportions of mRNA copies in fractions containing mRNAs bound to at least one ribosome (from fractions B to G in Fig. 1). For each gene, we calculated the ribosome density in each fraction as the number of ribosomes bound to the mRNA copies normalized with respect to the coding sequence length.

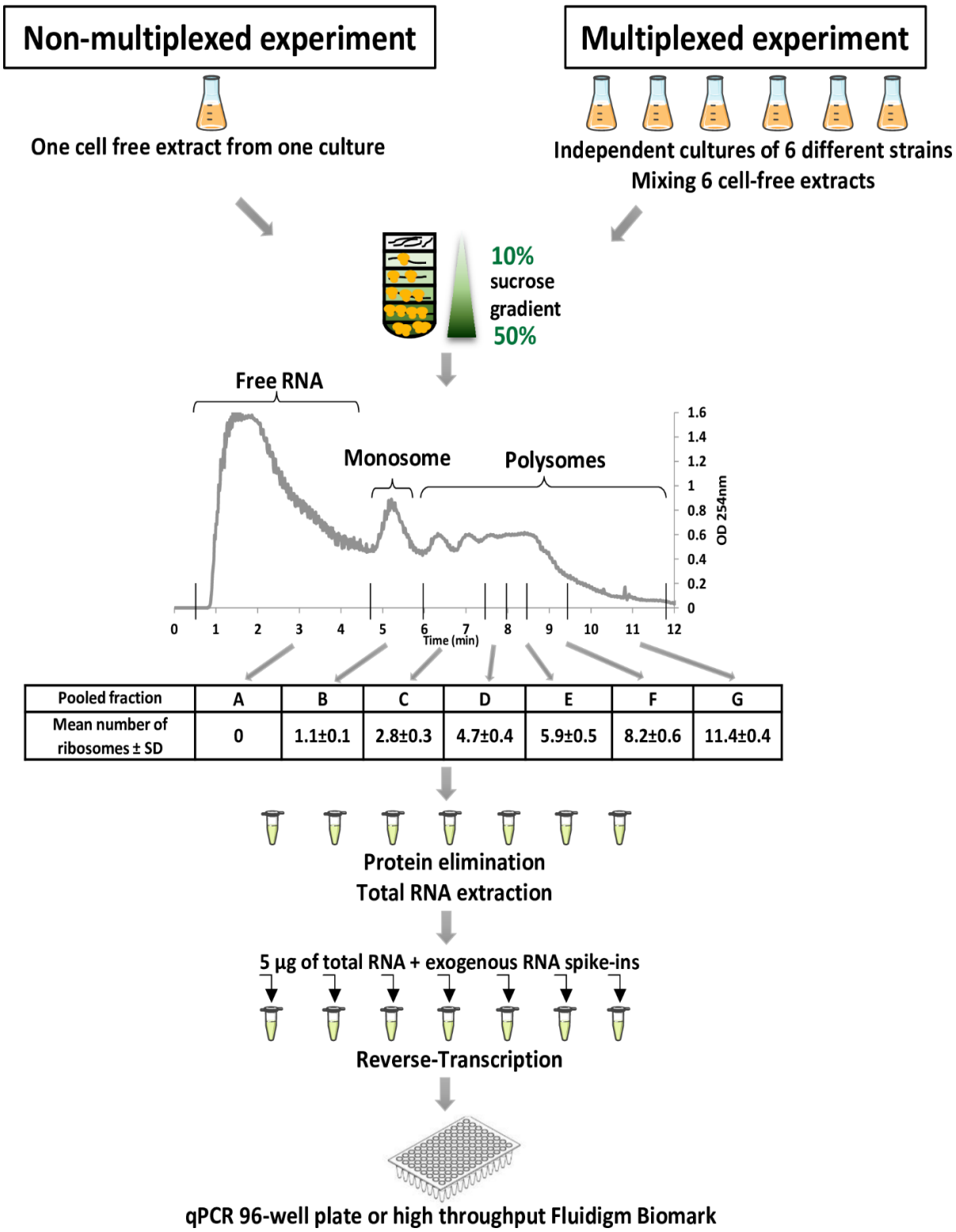


Figure 1: Polysome profiling experiment at a glance. All cultures and polysome profiling experiments were repeated twice to provide independent biological replicates.

Results

Polysome profiling experiment and translational parameters

The polysome profiling method was applied to *E. coli* MET739 cells during exponential growth on glucose in M9 minimum medium (Fig. 1). Translation elongation was stopped by adding chloramphenicol and the cells were washed and lysed. The cell free extract was loaded on a sucrose gradient to fractionate the mRNA-ribosome complexes according to the number of bound ribosomes. A representative polysome profile is shown in Fig. 1. The polysome profile was eluted in a total of 24 sub-fractions. After protein elimination and total RNA extraction, peaks were assigned by estimating the ratios between the 23S rRNA and 16S rRNA in each sub-fraction (S1 Fig). Once the components of each sub-fraction were identified, sub-fractions were grouped in seven fractions from A to G (Fig. 1). Fraction A comprised sub-fractions containing DNA, free RNAs and the small and large ribosomal subunits. The two small and large ribosomal subunits were respectively identified through the high 16S/23S rRNA and 23S/16S rRNA ratios (S1 Fig). In fractions B to G, the 23S/16S rRNA ratios were constant around 1.8 and matched entire ribosomes. The 2nd peak (fraction B) was attributed to the monosome. The 3rd and 4th peaks corresponding to two and three ribosomes were grouped in fraction C. The 5th peak corresponded to four ribosomes. The number of ribosomes in the following fractions was extrapolated as previously described (7). The mean value of the number of bound ribosomes in fractions B to G was calculated from four independent experiments (Fig. 1). We chose to exemplify the estimations of the translational parameters RO and RD using the *ihfB* gene. It is a well-expressed gene in *E. coli* coding for an integration host factor β-subunit commonly used as an internal normalization control in RT-qPCR experiments (24,28). To estimate *ihfB* RO and RD, we quantified the abundance of *ihfB* mRNA copies in fractions A to G.

A first normalization of *ihfB* mRNA abundance was performed using the ERCC RNA spike-ins. The ERCC RNA spike-ins consisted in a mix of 92 transcripts with a wide range of lengths, GC contents and concentrations. A constant quantity of the ERCC RNA spike-in mix was introduced in all the total RNA samples before the reverse transcription step. Different total RNA/ERCC ratios (in µg/µL) were tested: 2.5/0.01, 4/0.01, 5/0.01, 2/0.02, 5/0.03 and the very high ratio of 5/0.001. For the 5/0.001 ratio, oligonucleotides specific to the ERCC RNA spike-ins were added to increase their reverse transcription. The highest reverse transcription efficiencies of ERCC and RNA

were obtained with the very high ratio of 5/0.001. This ratio was thus chosen for all the ERCC normalization steps. The *ihfB* mRNA abundance was then normalized by the abundance of the four most concentrated ERCCs (ERCC 130, ERCC 002, ERCC 074 and ERCC 096) and by the initial mRNA abundance in each fraction to provide the relative initial *ihfB* mRNA abundance in each fractions.

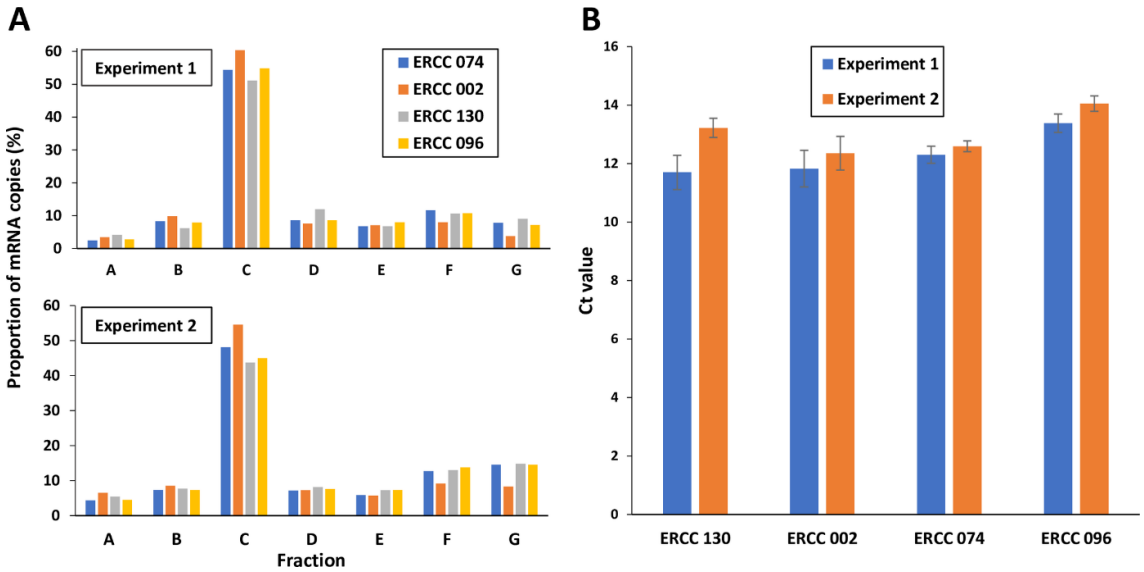


Figure 2: Distribution of *ihfB* mRNA copies in polysome fractions. (A) Proportion of *ihfB* mRNA copies in fractions A to G in two independent polysome profiling experiments from MET739 cell free extract when normalized by the four most concentrated ERCC (ERCC 130, ERCC 002, ERCC 074 and ERCC 096) using an RNA/ERCC ratio of 5/0.001 in µg/µL. (B) Variations in the levels of ERCC 130, ERCC 002, ERCC 074 and ERCC 096 estimated in two independent experiments. Mean values and standard deviations were calculated using the level values in the seven fractions (from A to G) of the same experiment.

The distributions of *ihfB* mRNA copy abundances between fractions A and G were then calculated to provide the typical plots of translational status (Fig. 2A). Normalizations using any of the four ERCC led to similar distribution of *ihfB* mRNA copies between the polysome fractions, so any of the four ERCC can be used to analyze a translational status. For further analyses, ERCC 074 (522pb, 35% GC, $15x10^{-21}$ mole/µL) was selected as normalizing ERCC, as it displayed the smallest variability between fractions and experiments (Fig. 2B). The translational status of *ihfB* was characterized by an RO of 96.6 ± 0.4% corresponding to 3.4 ± 0.4% of ribosome-free

mRNA copies not undergoing translation. Half the *ihfB* mRNA copies (in fraction C) were loaded with around 2.8 ribosomes corresponding to a ribosome density of 1 ribosome/100 nt and around 22% of the *ihfB* mRNA copies were heavily-loaded (in fractions F and G) with more than 8.2 bound ribosomes corresponding to a RD higher than 2.9 ribosomes/100 nt.

Multiplexing polysome profiling experiment does not alter the translational status of an mRNA

To validate a multiplexing method of polysome profiling experiments, we compared the translational status of *ihfB* without multiplexing and after multiplexing with cell extracts from distinct strains. We chose to multiplex up to six cell-free extracts. The translational status without multiplexing corresponds to the polysome profiling experiment described above when only the cell-free extract of *E. coli* MET739 was loaded on the sucrose gradient. In the multiplexing polysome profiling experiment (Fig. 1), the cell free extract of *E. coli* MET739 was mixed before loading on the sucrose gradient with five other cell free extracts: one from MET734 (overexpressing *cysZ*) and four from unrelated *E. coli* MG1655 constructs. Each cell free extract was produced from an individual culture in the exponential phase in M9 glucose medium. Quantification of *ihfB* mRNA copies in each fraction of the polysome was very reproducible within each experiment of polysome profiling (with and without multiplexing) (Fig. 3). In addition, the distributions of *ihfB* mRNA copies were very similar before and after multiplexing (Fig. 3). Consequently, comparable values of ribosome occupancy (96.6 ± 0.4% versus 96.6 ± 1.1%) and similar distributions of the ribosome density were obtained, the most frequent RD of the *ihfB* mRNA copies still being 1 ribosome/100 nt after multiplexing. We concluded that the multiplexing polysome profiling method did not affect the translational status observed for an mRNA and that this method can therefore be used to study translation statuses in parallel in multiple strains.

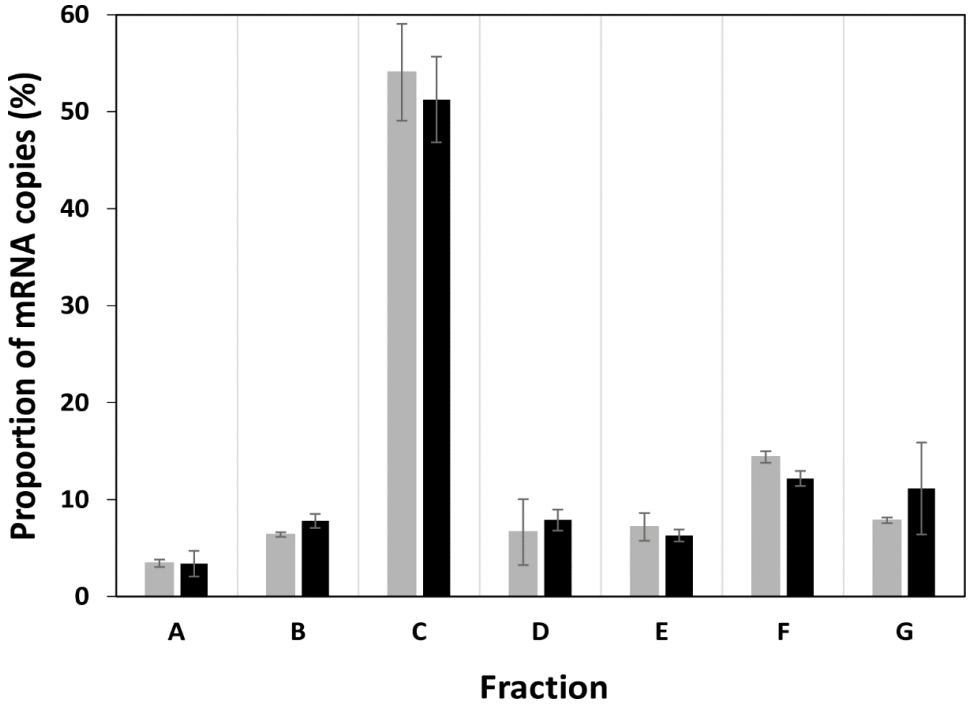


Figure 3: Distribution of *ihfB* mRNA copies between fractions A to G without multiplexing (grey) and after multiplexing (black). Fraction A, free mRNA not undergoing translation. Fractions B to G, mRNA bound with ascending number of ribosomes from 1 to 11. Mean values and standard deviations of two independent biological replicates are presented in the figure. Results were obtained with only MET739 cell-free extract in the non-multiplexing experiment or with MET739 cell free extract mixed with cell free extracts from five other strains in the multiplexing experiment.

Multiplexing polysome profiling experiment can be used to monitor the translational response between two conditions

We wanted to know if the translational response of a gene between two conditions was similar using the classical non-multiplexing and our multiplexing polysome profiling methods. Using the two methods, we thus investigated the translational response of the *cysZ* gene when its mRNA level was increased. The *cysZ* gene codes for a high-affinity, high-specificity sulfate transporter that provides the sulfur source for the synthesis of cysteine (29). Sulfate uptake by CysZ is essential for the survival of *E. coli* under low sulfate conditions. We focused on the *E. coli* MET734 strain in which the *cysZ* gene is under the transcriptional control of the arabinose inducible P_{BAD} promoter. *E. coli* MET734 was cultured in M9 glucose without arabinose (low *cysZ* mRNA level) and with arabinose transcriptional induction. Our measurements showed that arabinose

induction led to a 27 ± 6 fold induction of *cysZ* mRNA in MET734. On the other hand, the mRNA level from the chromosomal copy of *cysZ* in MG1655 wild type was four times lower than the level observed in MET734 without arabinose induction, and about 100 times lower than in MET734 with arabinose induction. In the pool of *cysZ* mRNA copies, the part originating from the chromosomal copy of *cysZ* can thus be neglected. Therefore in the multiplexing experiments, the *cysZ* mRNAs were only assigned to strain MET734. The five other strains not carrying the *cysZ* gene on the plasmid were also cultured independently with and without arabinose induction. Four polysome profiling experiments were then performed: one non-multiplexing experiment (only the cell free extract of *E. coli* MET734 was loaded on sucrose gradient) and the multiplexing experiment (loading of *E. coli* MET734 mixed with the five other cell free extracts), with and without arabinose transcriptional induction. The translational responses of *cysZ* between low and high mRNA expression level using the two methods are shown in Fig. 4.

Comparison of the distributions of *cysZ* mRNA copies using the standard (grey bars) and multiplexing method (black bars between A and B in Figure 4) showed equivalent distributions in the different fractions. This confirmed what was observed with *ihfB* that multiplexing cell free extracts did not alter the analysis of the distribution of a particular mRNA. With both the non-multiplexing and multiplexing methods, induction of *cysZ* expression led to a shift in the mRNA copies from the free mRNA fraction (fraction A) toward the more heavily ribosome bound fractions (mainly fraction G). With a high level of mRNA, ribosome occupancy increased from 62 ± 4.1% to 96 ± 3.0% without multiplexing and from 55.7 ± 0.7% to 89.6 ± 5.2% after multiplexing, reflecting the marked decrease in free mRNAs (lower proportions in fraction A). Consequently, ribosome density increased at high mRNA level to reach 1.5 ribosomes/100 nt in around 30% of the *cysZ* mRNA copies (in fraction G). These results showed that the multiplexing polysome profiling experiment allowed similar *cysZ* translational responses to higher mRNA levels: a decrease in free *cysZ* mRNA fraction and a higher ribosome load.

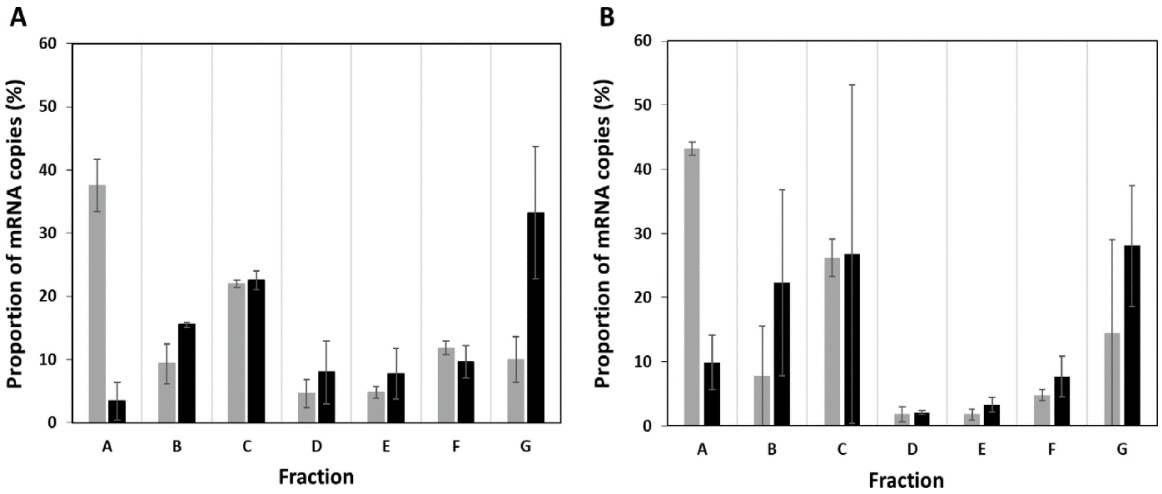


Figure 4: Distribution of *cysZ* mRNA copies between fractions in the two conditions: "without induction" (grey) and "with induction" (black) in (A) standard and (B) multiplexing polysome profiling experiments. Fraction A consists in free mRNA copies not undergoing translation. Fractions B to G contain mRNA copies bound with ascending numbers of ribosomes from 1 to 11. Mean values and standard deviations of two independent biological replicates are given. Results were obtained using only strain MET734 in the non-multiplexing experiment or using strain MET734 mixed with 5 other cell free extracts.

Parallel characterization of translational responses to changing transcription level

We used the multiplexing polysome profiling experiment to simultaneously study the translational responses of the *cysZ* and *lacZ* genes at two transcriptional levels. The *lacZ* gene codes for a β-D-galactosidase enzyme involved in lactose and other β-galactoside catabolism (30). In *E. coli* MET739, the *lacZ* gene is under the transcriptional control of the arabinose inducible P_{BAD} promoter. Without arabinose, the mRNA level from the chromosomal copy of *lacZ* in MG1655 wild type was 14 times lower than the level observed in MET739. Arabinose induction led to a 50 ± 23 times higher *lacZ* mRNA level in MET739. Therefore in the multiplexing experiments, the *lacZ* mRNAs were only assigned to strain MET739. Using the two multiplexing polysome profiling experiments described in the previous section, without arabinose (low *cysZ* and *lacZ* mRNA levels) and with arabinose (high *cysZ* and *lacZ* mRNA levels), we assessed the translational responses of *cysZ* and *lacZ* to changing mRNA levels (Fig. 5). We calculated the ratio of the proportion of *cysZ* and *lacZ* mRNA copies with high mRNA

levels to the proportion of mRNA copies with low mRNA levels. In both genes, when the mRNA expression was induced, we observed a significant reduction of the proportion of the free mRNA copies (ratios much lower than 1 in fraction A) and therefore an increase in the more heavily ribosome-loaded mRNA copies. However the pattern of translational responses of *cysZ* and *lacZ* differed. The *cysZ* mRNA copies spread all over the ribosome loaded fractions (all fractions from B to G exhibited ratios higher than 1) whereas the *lacZ* mRNA copies shifted more preferentially to fractions E and F. The magnitude of the translational response was more than 6-fold higher for *lacZ* than for *cysZ.* These results show that the multiplexing polysome profiling experiment allowed the parallel characterization of translational responses of two genes after increasing their mRNA level. For both *cysZ* and *lacZ*, the global translational responses to increased mRNA levels consisted in a shift toward the more heavily ribosome-loaded mRNA copies but the pattern and magnitude of the responses differed.

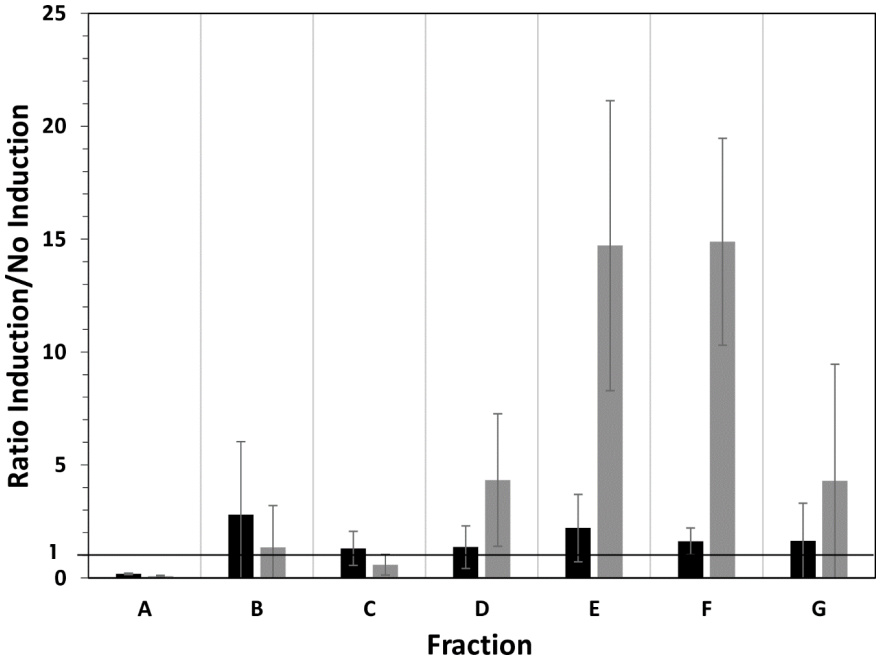


Figure 5: Translational response of *cysZ* (black) and *lacZ* (grey) in two mRNA expression conditions using the multiplexing polysome profiling method. For each fraction, the ratio of the proportion of mRNA copies with high mRNA levels to the proportion of mRNA copies with low mRNA levels was determined. Fraction A consists in free mRNA copies not undergoing translation while fractions B to G contain mRNA copies bound with ascending number of ribosomes from 1 to 11. Mean values and standard deviations of two independent biological replicates are given.

Specificity of the translational response

To check the specificity of the translational response to changing the mRNA level measured for *cysZ* and *lacZ*, we compared their responses to those of 10 chromosomally encoded genes not under the control of P_{BAD} (namely *eno, ihfB, rpsJ, rpsD, rplK, rplV, rpsL, trmJ, rnpA* and *rppH,* S1 Table). Ratios of mRNA copies proportions for the non-inducible genes were calculated in the two conditions, i.e. with and without arabinose, using the multiplexing experiments and were compared to the values of *cysZ* and *lacZ* (Fig. 6).

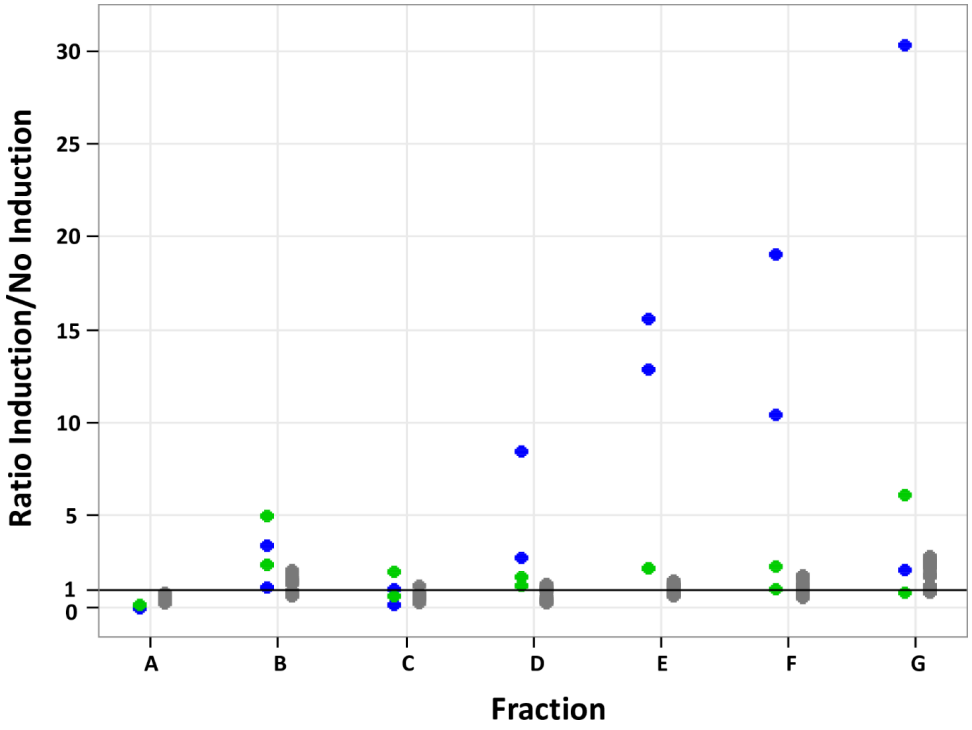


Figure 6: Translational response of the *cysZ* (green), *lacZ* (blue) and of 10 other unregulated genes (grey), with and without arabinose induction, using multiplexing polysome profiling experiments. For each fraction, the ratio was calculated between the proportion of mRNA copies with arabinose induction (high mRNA level) and the proportion of mRNA copies without arabinose induction (low mRNA level). Fraction A consists in free mRNA copies not undergoing translation while fractions B to G contain mRNA copies bound with ascending number of ribosomes from 1 to 11. Ratios were calculated from two independent biological replicates of each condition (2 dots are plotted for *cysZ* and *lacZ* and 20 dots for the unregulated genes in each fraction).

As expected, the non-inducible genes showed no significant difference in their translational response between with and without arabinose, since their mRNA level

did not significantly differ between the two conditions (variations in the expression of the 10 non-inducible genes were in the range of the technical error). With arabinose, the decreases in the proportions of both *cysZ* and *lacZ* copies in fraction A were higher than those in the non-inducible genes. The increases in the proportions of *lacZ* copies in the fractions D, E and F were considerably higher than the increases in the non-inducible genes; the increase was only slight for *cysZ* in fraction B compared to the non-inducible genes. These results confirmed that the multiplexing method can be used to specifically measure the translational response of a gene after changing its mRNA level in the cell.

Discussion

In this work, we developed and validated a multiplexing polysome profiling method to study the translation status of mRNAs and its variations in *E. coli.* The distribution of mRNA copies between the different polysome profiling fractions, and consequently ribosome occupancy and ribosome density, was similar using the standard non-multiplexing method and the new multiplexing method. The multiplexing method allows parallel quantification of the specific translational response to changing gene expression. In this study, we present the translational response of two genes, *cysZ* and *lacZ,* but our multiplexing approach allows simultaneous analysis of the translational response of up to six genes. In the case of *cysZ* and *lacZ*, we demonstrated a similar overall effect of the concentration of mRNA on the translation status with a higher ribosome load at higher mRNA levels but with a gene-specific pattern and magnitude of the responses. This result demonstrates co-transcriptional regulation of translation for these two genes. We hypothesize that co-transcriptional regulation of *cysZ* and *lacZ* translation contributes to physiological adaptation when cells regulate the mRNA level of genes to adapt to environmental changes (such as a low sulfate conditions (29) for *cysZ* or the availability of lactose as a carbon source for *lacZ* (30)). Further studies are now required to confirm these findings at the genome scale. Applied here to explore *E. coli* translation regulation, the multiplexing polysome profiling method can be expanded to any other organism. Using multiplexing saves time and effort and reduces the cost and technical bias that may result when large numbers of samples have to be handled.

In this study, we simultaneously studied the translational response of different genes to the same stimulus (i.e. their mRNA level). We used different molecular constructs in the same genetic background to trigger changes in mRNA levels. In the multiplexing experiment, we mixed cell free extracts of different strains generated in the same conditions, either without induction or with transcription induction. Another possible application of the multiplexing method is studying the translational response of one gene to different stimuli related to changes in the growth environment or to modifications in the genetic background. In this case, the multiplexing experiment will mix cell free extracts generated in different conditions. The only constraint will be that the gene of interest has to be specifically tagged (for example using barcode tagging to provide specific hybridization for qPCR) in each condition to differentiate the mRNA copies originating from each condition in the polysome profiling fractions.

Our multiplexing method will be of particular interest for the study of translation regulation at the mechanistic level. The effect of mRNA sequence-related parameters as potential regulators of translation initiation and elongation has been investigated using molecular approaches. The effect of codon usage on translation in yeast (19) or the one of 5'UTR sequence on polysome distribution in *Arabidopsis* (20) have already been investigated using polysome profiling. As these studies were limited to the characterization of only one mRNA with only three different sequence modifications, multiplexing polysome profiling experiment could easily be extended to the analysis of more genes and sequence modifications. When many samples were analyzed by the standard polysome profiling technique like in (18) to investigate the effect of eight different 5'UTR structures on the translation of a reporter mRNA in yeast, using our multiplexing method would have saved time, money and effort by analyzing a single multiplexed polysome. In the study of the effect of codon usage on translation elongation using the ribosome profiling method (31), the implementation of a complementary experiment of multiplex polysome profiling would have provided additional information on translation heterogeneity in the copies of the reporter mRNA.

Additional regulatory features of mRNA sequences on translation can be explored with the multiplexing method coupled with a high resolution PCR technique such as the TaqMan RT-PCR. Analysis of the translational response to small differences, from some nucleotides to single point mutation, in sequences suspected of being

involved in translation regulation (like sequence motif (conserved pattern (7) and SD-like motif (32)), secondary structure (33,34) for the binding sequences of regulatory ncRNA (35) and proteins (for instance CsrA (36)) could be performed more easily and quickly using the multiplexing method coupled with highly specific TaqMan probes. The technique could be also used to study the translational effect of natural single nucleotide polymorphism, for example between different alleles (37,38), to tackle the long-term evolution of translation regulation.

In conclusion, the multiplexing polysome profiling method is a low scale method mainly useful to study translation of (i) several reporter mRNAs (with different expression level, 5'and 3' UTR sequences or coding sequence), (ii) endogenous genes in a strain when they have been previously tagged with specific artificial sequences and (iii) different natural alleles of a gene found in closely related strains or species. Furthermore, at the genome-wide scale, this method coupled with RNA sequencing can also be used when mixing microorganisms with distinct genetic backgrounds. In this case, the possibility to assign the sequenced reads to the specific genes of each microorganism will allow the translation of these genes to be studied. The multiplexing method could open the way for "metatranslatomics" analyses.

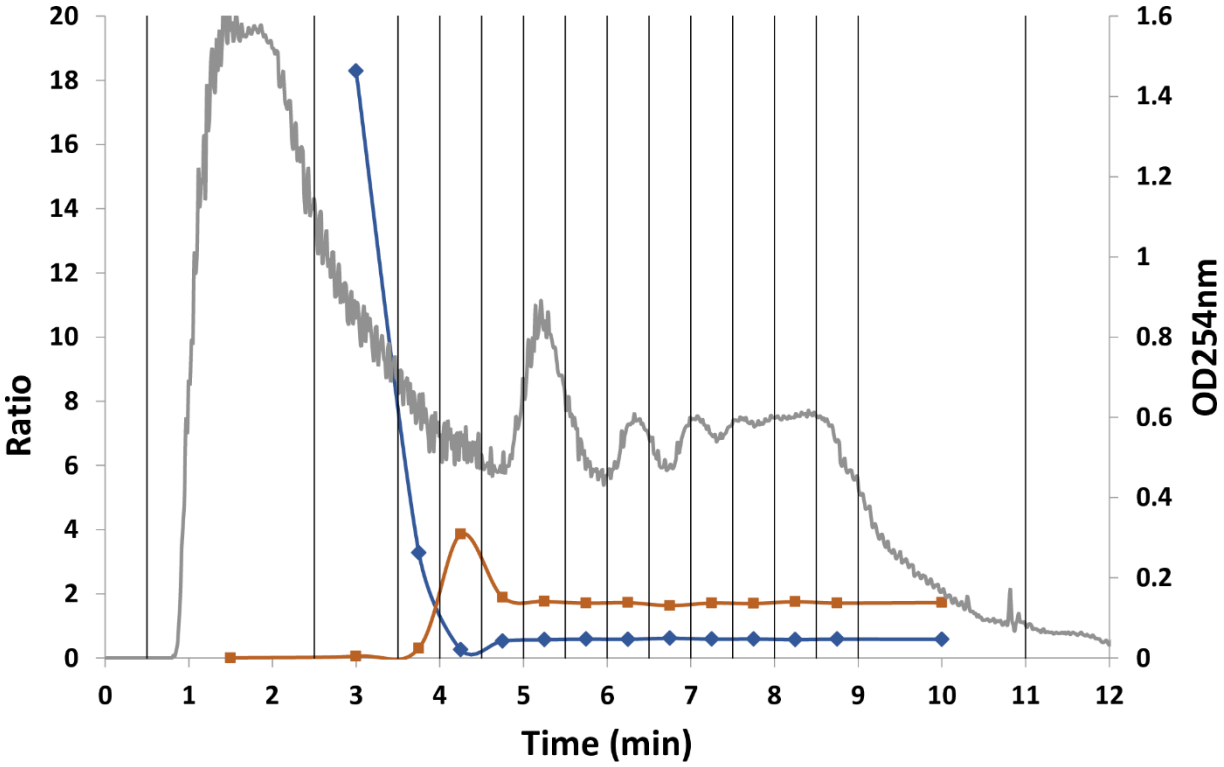
References

1. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet. 2013;13(4):227–32.
2. Duval M, Simonetti A, Caldelari I, Marzi S. Multiple ways to regulate translation initiation in bacteria: Mechanisms, regulatory circuits, dynamics. Biochimie. 2015;114:18–29.
3. Lange C, Zaigler A, Hammelmann M, Twellmeyer J, Raddatz G, Schuster SC, et al. Genome-wide analysis of growth phase-dependent translational and transcriptional regulation in halophilic archaea. BMC Genomics. 2007;8:1–17.
4. Fisunov GY, Evsyutina D V., Garanina IA, Arzamasov AA, Butenko IO, Altukhov IA, et al. Ribosome profiling reveals an adaptation strategy of reduced bacterium to acute stress. Biochimie. 2017;132:66–74.
5. Chassé H, Boulben S, Costache V, Cormier P, Morales J. Analysis of translation using polysome profiling. Nucleic Acids Res. 2017;45(3):e15.
6. Arava Y, Wang Y, Storey JD, Liu CL, Brown PO, Herschlag D. Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A. 2003;100(7):3889–94.
7. Picard F, Milhem H, Loubière P, Laurent B, Cocaign-Bousquet M, Girbal L. Bacterial translational regulations: high diversity between all mRNAs and major role in gene expression. BMC Genomics. 2012;13(1):528.
8. Picard F, Loubière P, Girbal L, Cocaign-Bousquet M. The significance of translation regulation in the stress response. BMC Genomics. 2013;14(1):588.

9. Yamasaki S, Matsuura H, Demura T, Kato K. Changes in Polysome Association of mRNA Throughout Growth and Development in *Arabidopsis thaliana*. Plant Cell Physiol. 2015 Sep 26;56(11):pcv133.
10. Chassé H, Aubert J, Boulben S, Le Corguillé G, Corre E, Cormier P, et al. Translatome analysis at the egg-to-embryo transition in sea urchin. Nucleic Acids Res. 2018 May 18;46(9):4607–21.
11. Pytharopoulou S, Grintzalis K, Sazakli E, Leotsinidis M, Georgiou CD, Kalpaxis DL. Translational responses and oxidative stress of mussels experimentally exposed to Hg, Cu and Cd: One pattern does not fit at all. Aquat Toxicol. 2011 Sep;105(1-2):157–65.
12. Melamed D, Pnueli L, Arava Y. Yeast translational response to high salinity: global analysis reveals regulation at multiple levels. RNA. 2008 Jul;14(7):1337–51.
13. Matsuura H, Ishibashi Y, Shinmyo A, Kanaya S, Kato K. Genome-Wide Analyses of Early Translational Responses to Elevated Temperature and High Salinity in Arabidopsis thaliana. Plant Cell Physiol. 2010 Mar 1;51(3):448–62.
14. Wu Y-H, Taggart J, Song PX, MacDiarmid C, Eide DJ. An MSC2 Promoter-lacZ Fusion Gene Reveals Zinc-Responsive Changes in Sites of Transcription Initiation That Occur across the Yeast Genome. PLoS One. 2016 Sep 22;11(9):e0163256.
15. Smirnova JB, Selley JN, Sanchez-Cabo F, Carroll K, Eddy AA, McCarthy JEG, et al. Global gene expression profiling reveals widespread yet distinctive translational responses to different eukaryotic translation initiation factor 2B-targeting stress pathways. Mol Cell Biol. 2005 Nov;25(21):9340–9.
16. Alves LR, Oliveira C, Goldenberg S. Eukaryotic translation elongation factor-1 alpha is associated with a specific subset of mRNAs in Trypanosoma cruzi. BMC Microbiol. 2015 May 19;15:104.
17. Liu B, Chen C. Translation Elongation Factor 4 (LepA) Contributes to Tetracycline Susceptibility by Stalling Elongating Ribosomes. Antimicrob Agents Chemother. 2018 Aug 1;62(8):e02356–17.
18. Saglioccos FA, Vega Lasoj MR, Zhull D, Tuitell MF, Mccarthyj JEG, Brownsii AJP. The Influence of 5'-Secondary Structures upon Ribosome Binding to mRNA during Translation in Yeast. J Biol Chem. 1993;268(35):26522–30.
19. Presnyak V, Alhusaini N, Chen YH, Martin S, Morris N, Kline N, et al. Codon optimality is a major determinant of mRNA stability. Cell. 2015;160(6):1111–24.
20. Matsuura H, Takenami S, Kubo Y, Ueda K, Ueda A, Yamaguchi M, et al. A Computational and Experimental Approach Reveals that the 5′-Proximal Region of the 5′-UTR has a Cis-Regulatory Signature Responsible for Heat Stress-Regulated mRNA Translation in Arabidopsis. Plant Cell Physiol. 2013 Apr 1;54(4):474–83.
21. King HA, Gerber AP. Translatome profiling: Methods for genome-scale analysis of mRNA translation. Brief Funct Genomics. 2016;15(1):22–31.
22. Ah-Seng Y, Rech J, Lane D, Bouet J-Y. Defining the Role of ATP Hydrolysis in Mitotic Segregation of Bacterial Plasmids. Viollier PH, editor. PLoS Genet. 2013 Dec 19;9(12):e1003956.
23. Esquerré T, Laguerre S, Turlan C, Carpousis AJ, Girbal L, Cocaign-Bousquet M. 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. 2014;42(4):2460–72.

24. Nouaille S, Mondeil S, Finoux AL, Moulis C, Girbal L, Cocaign-Bousquet M. The stability of an mRNA is influenced by its concentration: a potential physical mechanism to regulate gene expression. Nucleic Acids Res. 2017;45(20):11711–24.
25. Redon E, Loubière P, Cocaign-Bousquet M. Role of mRNA stability during genome-wide adaptation of Lactococcus lactis to carbon starvation. J Biol Chem. 2005 Oct 28;280(43):36380–5.
26. Maligoy M, Mercade M, Cocaign-Bousquet M, Loubiere P. Transcriptome analysis of Lactococcus lactis in coculture with Saccharomyces cerevisiae. Appl Environ Microbiol. 2008 Jan 15;74(2):485–94.
27. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 2001 May 1;29(9):e45.
28. Morin M, Ropers D, Cinquemani E, Portais JC, Enjalbert B, Cocaign-Bousquet M. The Csr system regulates Escherichia coli fitness by controlling glycogen accumulation and energy levels. MBio. 2017;8(5):e01628–17.
29. Zhang L, Jiang W, Nan J, Almqvist J, Huang Y. The Escherichia coli CysZ is a pH dependent sulfate transporter that can be inhibited by sulfite. Biochim Biophys Acta. 2014;1838:1809–16.
30. Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol. 1961;3(3):318–56.
31. Yu CH, Dang Y, Zhou Z, Wu C, Zhao F, Sachs MS, et al. Codon Usage Influences the Local Rate of Translation Elongation to Regulate Co-translational Protein Folding. Mol Cell. 2015;59(5):744–54.
32. Li GW, Oh E, Weissman JS. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. Nature. 2012;484(7395):538–41.
33. Evfratov SA, Osterman IA, Komarova ES, Pogorelskaya AM, Rubtsova MP, Zatsepin TS, et al. Application of sorting and next generation sequencing to study 5′-UTR influence on translation efficiency in Escherichia coli. Nucleic Acids Res. 2017 Apr 7 ;45(6):3487–502.
34. Giuliodori AM, Di Pietro F, Marzi S, Masquida B, Wagner R, Romby P, et al. The cspA mRNA is a thermosensor that modulates translation of the cold-shock protein CspA. Mol Cell. 2010 Jan 15;37(1):21–33.
35. Gottesman S, Storz G. Bacterial small RNA regulators: versatile roles and rapidly evolving variations. Cold Spring Harb Perspect Biol. 2011 Dec 1;3(12):a003798.
36. Potts AH, Vakulskas CA, Pannuri A, Yakhnin H, Babitzke P, Romeo T. Global role of the bacterial post-transcriptional regulator CsrA revealed by integrated transcriptomics. Nat Commun. 2017 Dec 17;8(1):1596.
37. Artieri CG, Fraser HB. Evolution at two levels of gene expression in yeast. Genome Res. 2014;24(3):411–21.
38. Wang Z, Sun X, Zhao Y, Guo X, Jiang H, Li H, et al. Evolution of gene regulation during transcription and translation. Genome Biol Evol. 2015;7(4):1155–67.

Supporting information



S1 Fig: Ratios of the 23S rRNA and 16S rRNA amounts in sub-fractions of the polysome profiling experiment (grey line). The ratio of 16S/23S rRNA amounts is in blue and the ratio of 23S/16S rRNA amounts is in brown. The sub-fractions are delimited by vertical black lines.

S1 Table: Sequence of qPCR primers used to quantify 12 endogenous genes of *E. coli* MG1655 and four ERCC RNA spike-ins.

Gene	Function	Forward primer	Reverse primer
cysZ	Sulfate transporter	CATCATTCACATCTGCCCCACG	GCGCCCCCCATCAACAAAATAT
lacZ	β-D-galactosidase		
pair_1		TCCGTGACGTCTCGTTGCTGC	TCACGCAACTCGCCGCACAT
pair_2		CCCGCATCTGACCACCAGCG	CAGCGGCGTCAGCAGTTGTT
pair_3		GTCGTGACTGGGAAAACCCTGG	AACTGTTGGGAAGGGCGATCG
pair_4		AACAACTTTAACGCCGTGCGCT	CACCATGCCGTGGGTTTCAATA
pair_5		CAGCTGGCGCAGGTAGCAGAG	GGCAGATCCCAGCGGTCAAA
ihfB	Integration host factor	GCCAAGACGGTTGAAGATGC	CAAAGAGAAACTGCCGAAACC
trmJ	Methyltransferase	TCGAATTGTGCTGGTGGAGACG	AATCGCCTGGGAGTCGGGTT
eno	Enolase	GTCGTGAAATCATCGACTCC	CAGCTTTGGTTACGCCTTTA
rppH	Pyrophosphohydrolase	AGAAGTAGGATTAAGCCGCA	GTCAAACTCTGGTGTACTGC
rnpA	RNase P protein component	AACGCAATCGGATTAAACGT	GGCGCCATAATTTTTCCAAC
rpsD	30S ribosomal subunit protein	CGTGAGGGCACCGACTTATTC	AGTCAGACAGACGCGGTTTACG
rpsJ	30S ribosomal subunit protein	CGTACTCACTTGCGTCTGGTTG	AGGCTGATCTGCACGTCTACAC
rpsL	30S ribosomal subunit protein	CATCGGTGGTGAAGGTCACAAC	CTTTAACGCCGGAGCAGTCAAG
rplK	50S ribosomal subunit protein	AGCGGCTGGTATCAAGTCTGG	GTCATGTCGGCAGCTTTGGTC
rplV	50S ribosomal subunit protein	CGCGGTAAGAAAGTGTCGCAG	CAGCGCCATCGTTGTGTTCAG
ERCC 130	RNA spike-ins	GTGAAGATGATTGACCGCACGC	TGCATATTGCAGCTGAGCCAGC
ERCC 002	RNA spike-ins	CCGTCGGCTGATCGTGGTTT	CGACCGTACAGCTCTGGAACCC
ERCC 074	RNA spike-ins	GCCTTGGTAGGGATAGATAGCCACC	CTGGGGTTATGAGTAGGGATGAGCA
ERCC 096	RNA spike-ins	CGTAACCAAACATGCACAGCGG	TCGCGTCATCGATCCGGGT

- Chapitre 2 -

Analysis of RNA-Seq data to study translatome

Dans le premier chapitre, nous avons présenté le principe d'analyse par la méthode de polysome profiling de l'état traductionnel de quelques ARNm cibles. Dans ce cas, la quantification des ARN était réalisée par RT-qPCR mais cette technique ne permet d'analyser qu'un petit nombre d'ARNm sélectionnés. Dans les deux prochains chapitres, nous allons passer à l'étude de l'état traductionnel de l'ensemble des ARNm pour une analyse de la traduction à l'échelle omique.

L'étude à l'échelle -omique des processus de l'expression génique (transcription, traduction, dégradation…) requiert des méthodes de séquençage et quantification à haut débit des ARNs comme le microarray et plus récemment le RNA-Seq. La méthode puissante de RNA-Seq présente de nombreux avantages et elle est de plus en plus utilisée chez les microorganismes y compris les bactéries. Cette méthode de RNA-Seq couplée à la technique de polysome profiling a été utilisée pour la première fois dans l'équipe juste avant le début de ma thèse, donc mon premier objectif a été de comparer différentes méthodes de bioinformatique et statistiques pour analyser correctement les données RNA-Seq.

Ce chapitre présente le déroulement des étapes de traitement des résultats du RNA-Seq, dont les deux points clés sont l'alignement des séquences et la normalisation. La première étape de traitement des données du RNA-Seq est d'attribuer les séquences lues aux gènes correspondants. Pour cela, une méthode rapide et répandue pour aligner les fragments courts, bwa, a été comparée avec l'outil fourni par l'appareil Ion Torrent (tmap). Les deux outils se basent sur le même algorithme de recherche de séquences et ont donc donné des résultats très similaires. Le choix de bwa a été motivé par sa facilité à enlever les séquences multi-mappées (qui s'alignent à plusieurs endroits).

En second lieu, pour chaque gène, nous voulions comparer la quantité de ses copies d'ARNm entre les différentes fractions générées lors du fractionnement des complexes ARNm-ribosomes. Avant de pouvoir comparer les niveaux de chaque ARNm entre les fractions, il faut normaliser les données. En effet, des biais techniques systémiques pourraient être introduits lors de l'expérimentation : par exemple différentes efficacités de la ribodéplétion, de la rétro-transcription en ARNc lors de la préparation des librairies et du séquençage. L'expression différentielle entre les fractions pourrait alors être liée à des variations techniques et non biologiques. La

normalisation a donc pour objectif d'enlever ces biais et de rendre les échantillons comparables. Dans ce travail, nous avons considéré deux stratégies de normalisation : la première est basée sur l'approche DESeq (une méthode couramment utilisée) et la deuxième dénommée RUV pour Remove Unwanted Variation moins fréquemment utilisée, est basée sur l'utilisation de contrôles externes (dans notre cas les ERCC). La méthode DESeq rend comparable les fractions en rapprochant les médianes après le séquençage. Cette normalisation doit être complétée par la normalisation par la quantité d'ARN après la ribodéplétion. Pour la deuxième méthode RUV, les contrôles externes ERCC ont été introduits en même quantité dans chaque fraction avant la ribodéplétion. En théorie, sans la présence de biais techniques, leur expression devrait être la même dans toutes les fractions. Les variations d'expressions des ERCC seront utilisées dans la méthode RUV pour déterminer les variations systématiques non désirables. Les effets de chaque méthode sur la variabilité des données ont été vérifiés et comparés. Les deux méthodes ont réduit l'hétérogénéité entre les fractions (l'ensemble des gènes pour DESeq et seulement les contrôles ERCC pour RUV). Les deux méthodes ont permis d'obtenir des résultats cohérents : les valeurs de RO et RD obtenues sont très corrélées. Les modèles linéaires de covariance sont utilisés pour expliquer la variabilité de RO et RD par des facteurs quantitatifs et qualitatifs. Ces modèles permettent d'identifier les déterminants de RO et RD et de les hiérarchiser pour pouvoir sélectionner les plus significatifs. Les deux méthodes de normalisation ont abouti à la même sélection de déterminants de RO et RD. En prenant en compte d'autres critères comme le nombre de gènes analysés, le nombre de variables significatives sélectionnées et l'hypothèse de l'invariance de l'expression des gènes, nous avons choisi la méthode RUV pour la normalisation.

Comme le RNA-Seq permet d'accéder aux régions non traduites, nous avons aussi effectué un focus sur le rôle du 5'UTR dans la régulation de la traduction. Cette région est susceptible d'être impliquée par exemple via l'accessibilité des ribosomes aux RBS et à la fixation de régulateurs (petits ARNs, protéines ou des métabolites). Nous avons effectué le comptage des séquences non seulement sur l'ORF mais sur toute la longueur de l'ARNm (5'UTR et ORF). Nous n'avons pas observé de différences significatives au niveau des valeurs de RO et RD lors du comptage sur 5'UTR + ORF. Les variables liées au 5'UTR comme le pourcentage de GC, le deltaG de la région -30 +24 nucléotides, la longueur du 5'UTR et la nature du 2^{e} nucléotide ont ensuite été prises

en compte dans les modèles linéaires pour étudier leur effet sur les variables de traductome. Les modèles de covariance ont donné les mêmes facteurs significatifs en prenant en compte ou non le 5'UTR lors du comptage, aucun des facteurs liés au 5'UTR n'a pas été trouvé comme significatif de RO et RD.

En conclusion, ce travail a permis de choisir les méthodes de bioinformatique et statistiques adaptées au protocole expérimental. La normalisation a réduit la variabilité des résultats due aux erreurs techniques. Ce travail a aussi montré la robustesse des résultats et choisi la méthode appropriée pour garder un grand nombre de gènes pour l'analyse biologique.

Introduction

The approaches of systems biology are powerful to study bacterial physiology. It provides multi-level knowledge on genome-wide gene expression regulation and hence allows to explore the complex regulation of metabolic adaptation. Our team is interested in the relationship between transcriptome, stabilome and translatome. These genome-wide data were previously obtained by microarray technology. Recently high-throughput sequencing technology like RNA-Seq was developed to study large-scale regulation of gene expression with a number of benefits compared to microarray based methods. As sequencing technology does not require pre-selection of transcript probes, it enables to detect novel transcripts, study untranslated regions and discover unknown genome events, for e.g. insertions or deletions (Zhong et al. 2009). Several studies used RNA-Seq to identity transcriptional start sites in *Escherichia coli* and *Helicobacter pylori* (Kim et al. 2012; Sharma et al. 2010; Thomason et al. 2015) and study small non coding RNAs in prokaryotes (Croucher and Thomson 2010, Sittka et al. 2008). Moreover, RNA-Seq does not suffer from the saturation of light detection as for microarrays and hence has no upper limit of quantification for highly expressed genes. It also provides higher sensitivity and lower background in detecting rare and low abundant transcripts (Zhong et al. 2009). Its broader quantification range is useful to identify differential expression to study transcriptional and translational changes between growth conditions, for example to compare transcriptomes between growth phases in *Bacillus anthracis* (Passalacqua et al. 2009) or compare translatomes without or with amino acid starvation in yeast (Ingolia 2009).

For these reasons, our team wanted to use RNA-Seq technology coupled to polysome profiling experiment to study genome-wide regulation of gene expression in *E. coli*. Applied to study translation regulation, RNA-Seq allows to access (i) non-coding RNAs and untranslated regions (e.g. 5'UTR), which could be involved in translation regulation (Tuller and Zur 2015) and (ii) RNA integrity by visualizing potential partial degradation of transcript. The analyze of transcript integrity using RNA-Seq permitted for example the discovery of a new mechanism of translation regulation based on partial RNA fragmentation of the 3' end (Feng et al. 2015). Besides, as mRNA-ribosome complexes are fractionated in several fractions during the polysome profiling

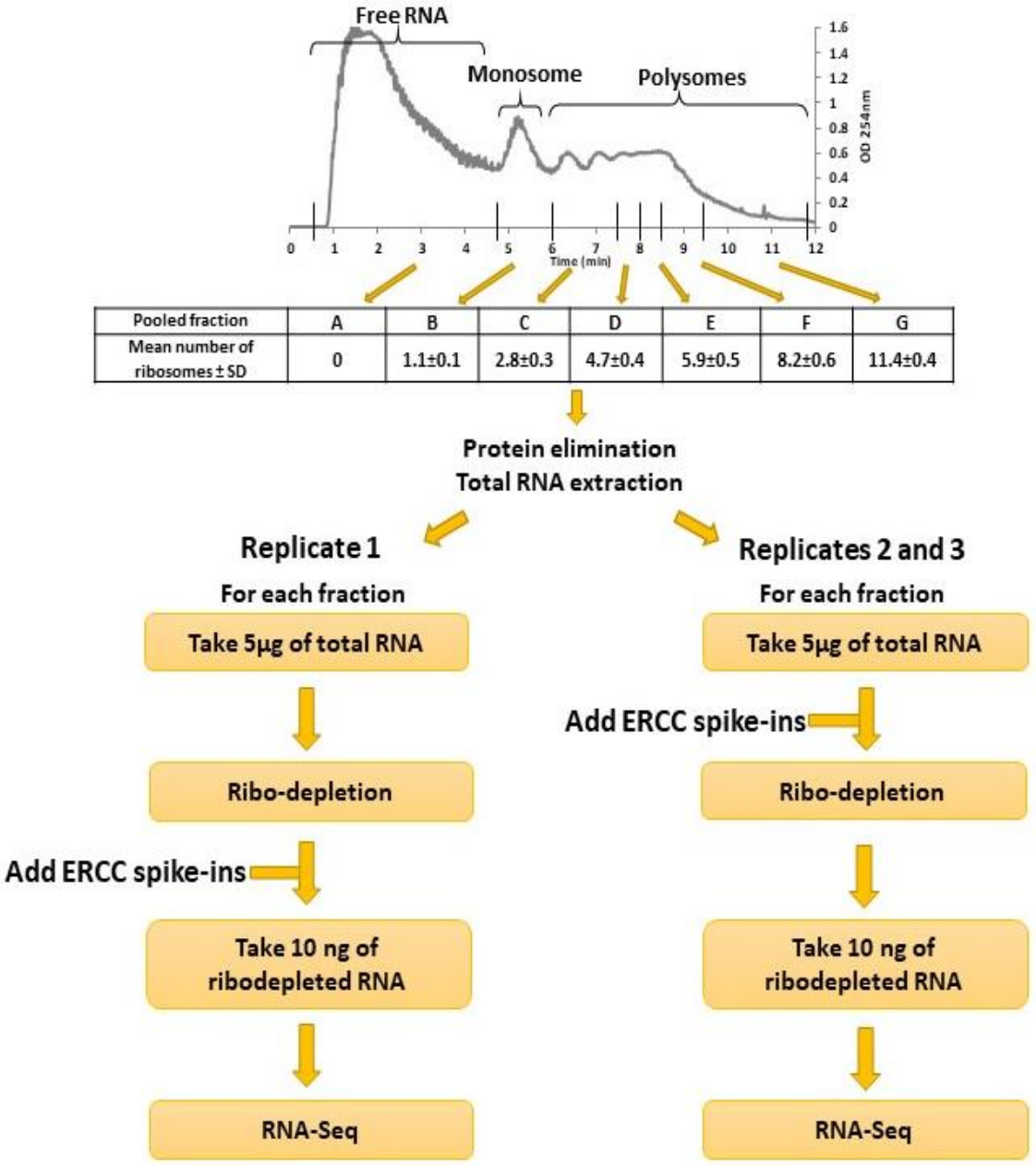
experiment, mRNA abundances in some fractions might be very low. RNA-Seq allows therefore to measure these low mRNA levels with more accuracy.

Since our team used for the first time the RNA-Seq technology in translatome study while I was starting my PhD, the first objective of my thesis was to define a methodology of analyzing RNA-Seq data in the framework of a polysome profiling experiment. The major challenges consisted in comparing methods of read mappings and data normalizations to provide the two translational parameters, ribosome occupancy and ribosome density of all the transcripts present in fast growing *E. coli* cells.

I) Data generation

Polysome profiling experiments were performed in the team before my arrival by Antoine Picavet and Marie-Pierre Duviau. *E. coli* MG1655 was cultured in batch culture in minimum medium M9 supplemented with 3g/L glucose (Esquerré et al. 2014). Translation was stopped at mid-exponential phase by adding chloramphenicol and mRNA-polysome complexes were separated according to ribosome load on a sucrose gradient. Different eluted sub-fractions were grouped into 7 fractions (from A to G) with ascending number of mRNA bound ribosomes (Fig. 1). Three polysome profiling experiments from three independent batch cultures were carried out. After extraction of total RNA, rRNAs were depleted using Ribo-Zero^{RM} Magnetic Kit (for gram-negative bacteria). To account for technical variability, control Ambion™ ERCC RNA spike-in mix was used as external normalizer. For each fraction, an equal amount of ERCC spike-ins was added to a constant amount of RNA. ERCC spike-ins were added to RNA before ribodepletion for replicates 2 and 3 and after ribodepletion for replicate 1 (Fig. 1). 10 ng of ribodepleted RNA of each fraction were used for sequencing library construction using the Ion Total RNA-Seq Kit v2 kit and the Ion Xpress™ RNA-Seq Barcode 1-16 Kit following the manufacturer's instructions. Briefly, after enzymatic (RNase III) fragmentation, RNA fragments were purified. After ligation of adaptors, complementary DNA (cDNA) was synthesized using the reverse transcriptase SuperScript III and purified. Then, the fixation of the barcodes were done during the amplification of the cDNA and the libraries were purified and quantified. 100 pmol/L of 7 libraries were pooled together, amplified on beads by emulsion PCR and enriched on the Ion OneTouch™ 2 System (Thermofisher) using the Ion PI Template OT2 kit V3.

The template was sequenced (100 bp oriented single end) on the Ion Torrent Proton using the Ion P1 chip. Library construction and sequencing were performed by the GeT-BioPuces platform (http://get-biopuces.insa-toulouse.fr/). In final, we had 21 RNA-Seq data corresponding to the 7 fractions and the 3 biological replicates.



Pooled fraction	A	B	C	D	E	F	G
Mean number of ribosomes ± SD	0	1.1±0.1	2.8±0.3	4.7±0.4	5.9±0.5	8.2±0.6	11.4±0.4

Figure 1: Overview of the RNA-Seq sample preparation in the framework of a polysome profiling experiment.

The following steps of bioinformatical analysis of RNA-Seq raw data (mapping and read counting) were performed by Sabrina Rodriguez from Toulouse White Biotechnology. Methods of statistical data analysis and normalization were developed with the help of Sandrine Laguerre, biostatistician in our team.

II) Comparison of two mapping methods: tmap and bwa

As the RNA-Seq technique relies on mapping read sequences on genome, the alignment step is crucial for accurate quantification. We compared two mapping tools : tmap (torrent mapping alignment program), a software specifically developed for the Ion Torrent™ and coupled with the machine and the algorithm bwa-mem version 0.7.12-r1069 of the Burrows-Wheeler Alignment tool (BWA) which is a tool based on backward search with Burrows-Wheeler Transform (Li and Durbin 2009, Li 2012).
We used *E. coli* MG1655 genome of length 4.641.652 bp (version 3 U00096.3 in GenBank https://www.ncbi.nlm.nih.gov/nuccore/U00096.3, corresponding to release 30, version ASM584v2, GCA_000005845.2 in Ensembl (www.bacteria.ensembl.org)) which merges different exons of the same transcript and results in 4497 transcripts. After mapping, reads counting was performed with HTSeq Count version 0.6.1p1 using the « intersection non empty » mode (Anders et al. 2015).

To compare the two pipelines, we plotted the read counts obtained with tmap as a function of those generated with bwa for the 21 datasets (7 fractions in triplicates) (Fig. 2). All the plots nearly matched the line y=x with very high correlation coefficients (all above 0.99). These results showed that both pipelines gave very similar results. We chose bwa because it allows to distinguish easily multi-mapped reads and later on to remove them easily.

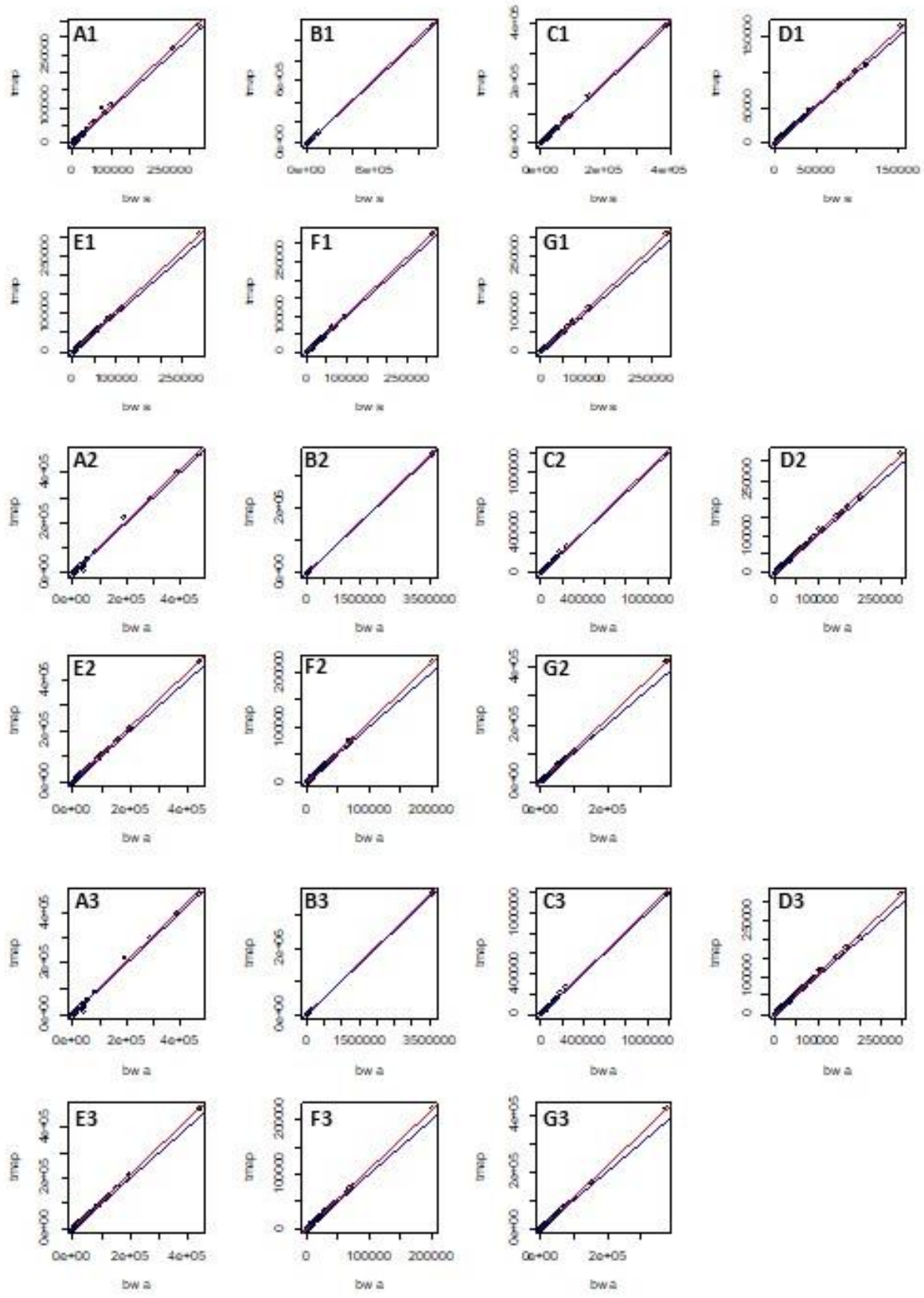


Figure 2: Scatter plots of read counts aligned with tmap and bwa for the 7 fractions (A to G) and the replicates 1, 2 and 3. The linear regression is in red and the line y=x is in blue.

III) Sequencing results

The length and quantity of total reads and the percentage of unmapped and multi-mapped reads for the 21 RNA-Seq data are shown in table 1.

Table 1: Summary of RNA-Seq data.

Fraction	Mean read length (bp)	Total reads	Unmapped reads	Multi-mapped reads
A-1	121	5 175 752	5%	25%
B-1	108	4 973 907	6%	8%
C-1	101	6 443 529	5%	7%
D-1	111	4 990 527	5%	10%
E-1	101	8 013 480	6%	9%
F-1	94	6 852 230	6%	8%
G-1	109	7 316 431	6%	10%
A-2	96	8 000 803	3%	19%
B-2	111	9 018 579	3%	9%
C-2	126	11 123 465	2%	14%
D-2	93	10 814 614	4%	10%
E-2	88	14 046 750	5%	9%
F-2	93	6 245 816	3%	12%
G-2	96	8 461 871	3%	13%
A-3	92	7 591 197	4%	16%
B-3	94	9 022 654	5%	7%
C-3	84	8 222 286	5%	5%
D-3	97	11 508 394	4%	8%
E-3	89	11 423 585	4%	6%
F-3	93	12 093 823	3%	8%
G-3	97	13 553 717	4%	10%

The reads length was around 100 bp while the total reads varied from ≈ 5 to 13.5 million depending on the samples. The percentage of unmapped reads was quite constant in the different fractions (from 3 to 6%). The percentage of multi-mapped

reads was more variable between the fractions (from 5 to 25%). rRNAs being a potential source of multi-mapped reads, one explanation of different percentages of multi-mapped reads could be differences in the ribodepletion efficiency within the samples which would result in variable rRNA amounts in the ribo-depleted RNA samples.

We also checked for the presence of DNA contamination, which could disrupt RNA counts. We used IGV software (Robinson et al. 2011) to visualize reads mapping on the genome. In case of DNA contamination, a background of read covering all over the genome would be present. Here this was not the case: for any of the 21 samples because in all the samples we found genes and regions without any mapped read as exemplified in figure. 3.

Figure 3: Visualization of the reads coverage of a selected genome region using the IGV software. The region in red had no mapped read indicating the absence of DNA contamination.

Additionally, we estimated the linearity between RNA concentration and read count using known concentrations of sequenced ERCC spike-ins. The read counts of ERCC spike-ins were correlated with their concentrations for the counts above 10 reads (Fig. 4). Therefore, we defined 10 reads as the minimum threshold below which the RNA expression was considered in the technical noise and not quantitative enough to be further analyzed. 380 transcripts exhibiting a maximum read count in all the fractions below 10 were discarded because of their too low expression.

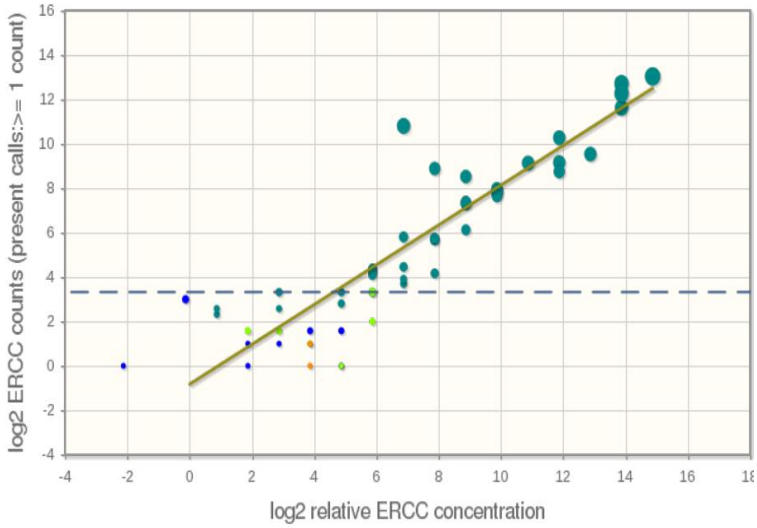


Figure 4: Plot of the ERCC spike-ins read counts and concentrations. Horizontal dotted line indicates the threshold below which there is no more a linearity.

IV) Normalization methods

Our objective was to quantify and compare the mRNA copy concentration of a given gene in the different polysome profiling fractions. However, part of the observed differences in mRNA copy concentration may not result of biological differences but were rather due to experimental variations between samples and technical bias. To lower such non-biological effects, normalization is required to correct data.

We used two methods of normalization (Fig. 5). Globally, method 1 consisted in normalizing the data by the library size (number of total reads in each sample) using the DESeq2 R package and after that by the RNA quantity after ribodepletion, whereas method 2 used the ERCC spike-ins for data normalization with the RUVseq R package. The two methods had the same last step, a normalization by the initial total RNA quantity. Each step of the two methods is described below.

The first step of method 1 was the normalization by the RNA-Seq library size implemented in the DESeq2 R package. DESeq2 provides normalized counts which correspond to the ratio of the initial counts Y_{ij} (read count associated to gene i in sample j) on a normalized factor s_j. This normalized factor is previously estimated as the median of $(\frac{Y_{1j}}{Y_1^R}, \ldots, \frac{Y_{ij}}{Y_i^R}, \ldots, \frac{Y_{gj}}{Y_g^R})$ where $(Y_1^R, \ldots, Y_i^R, \ldots, Y_m^R)$ is a pseudo-reference sample. This pseudo reference sample was previously defined as the geometric mean of Y_{ij} $(Y_i^R = \prod_{j=1}^{m} Y_{ij}^{1/m})$.

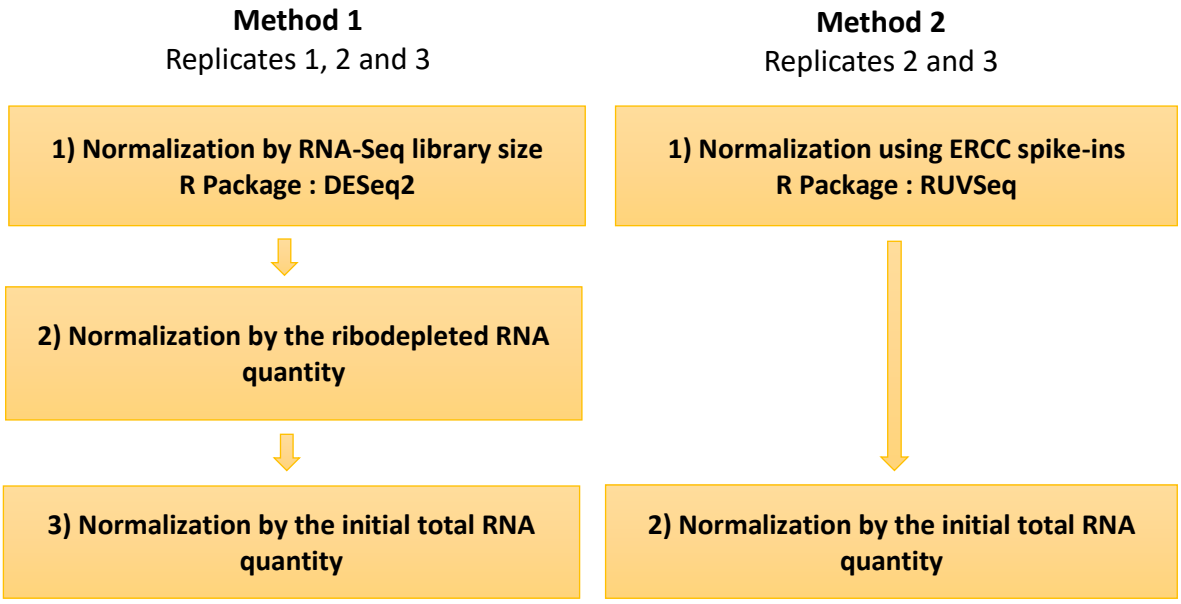


Figure 5: Normalization methods.

The second step of method 1 is a normalization using the RNA quantity after ribodepletion. The ribodepletion step provided different ribodepleted RNA amounts between the samples (Fig. 6) but only 10 ng of each fraction were used for library construction. So the read counts were multiplied by the ribodepleted RNA quantity (in ng) of each fraction and divided by 10.

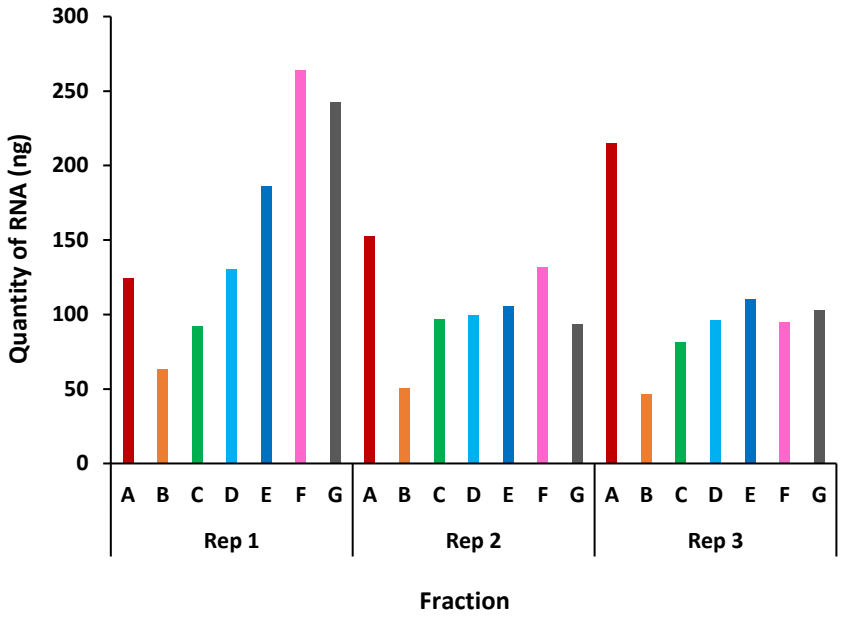


Figure 6: Ribodepleted RNA quantity for the 7 fractions and the 3 replicates.

The second method (RUVg function from RUVSeq R package) (Gagnon-Bartsch and Speed 2012; Risso et al. 2014) considers a log-linear regression model for the read counts. Variations of the ERRC counts are used to estimate a matrix of hidden factors of unwanted variation. Then the matrix of normalized counts (wanted variations) is defined as the residuals from ordinary least squares regression of the read counts on the estimation of the unwanted variations.

Last necessary step of both methods is the normalization by the initial total RNA quantity. At the end of the polysome profiling experiment, total RNA extraction provided different total RNA quantity in each fraction (Fig. 7), but only 5 µg of total RNA of each fraction was used for ribodepletion (Fig. 1). Therefore, the read counts were multiplied by the initial total RNA quantity (in µg) of each fraction and divided by 5.

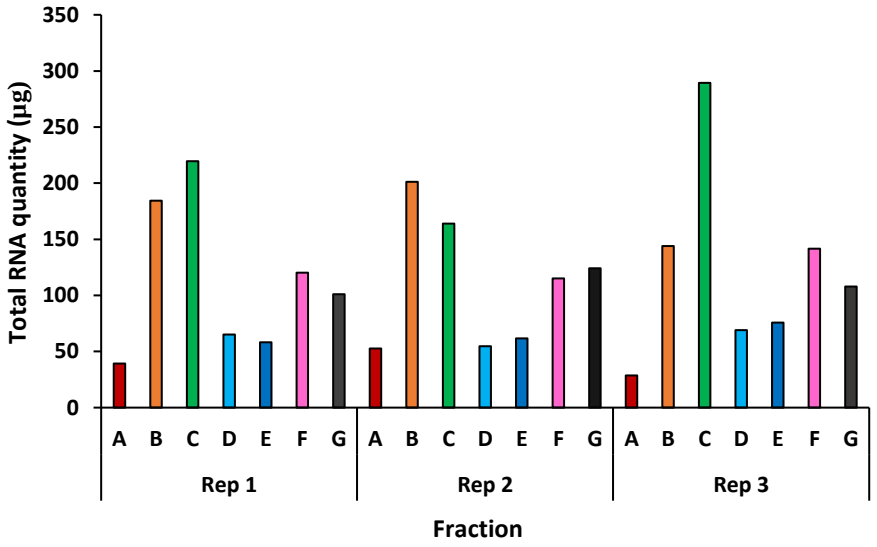


Figure 7: Initial total RNA quantity for the 7 fractions and the 3 replicates.

Method 1 was applied to all three replicates. Method 2 was only applied to replicates 2 and 3 since the ERCC spike-ins in replicate 1 were only added after the ribodepletion step (Fig. 1).

V) Normalization results

To visualize the data transformation by the two normalization methods we used as the metric the relative log-expression (RLE). RLE consists in relative comparison of the transcript level in a given fraction to the levels in all the fractions and it is thus useful to visualize variability between samples (Gandolfo and Speed 2018). For each gene in

each fraction, the ratio between its level in this fraction and the median level in all fractions was calculated. The box plots in figureure 8 show the RLE values of all the genes in each fraction before and after normalization by the RNA-Seq library size in method 1. After normalization, the RLE plots were all centered on 0, and the fractions became more comparable.

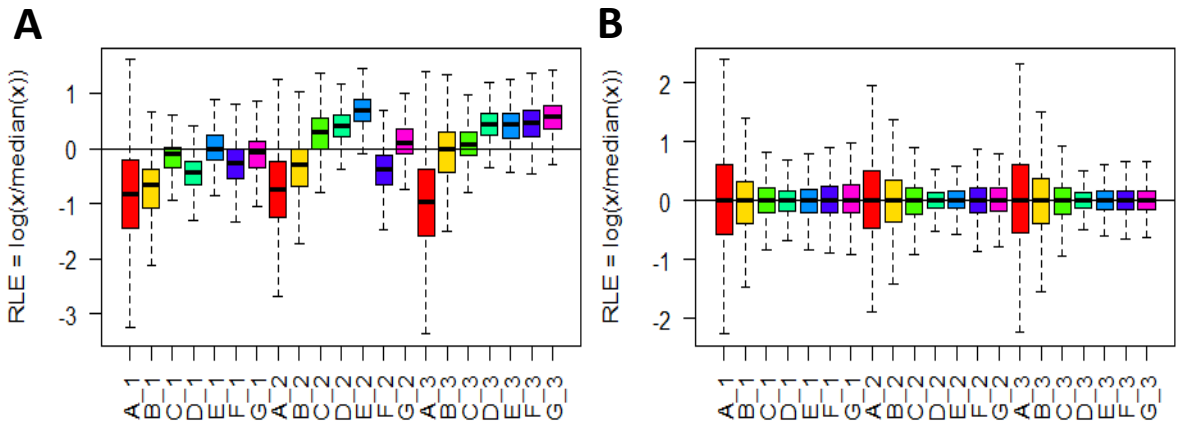


Figure 8: RLE plots of all reads counts (A) before and (B) after normalization by the RNA-Seq library size (using DESeq) in method 1, for the 7 fractions (from A to G) and the 3 replicates.

In method 2, although the same amount of ERCC spike-ins was added in each fraction, we obtained an important variability in ERCC read counts before normalization (Fig. 9A). This demonstrated the presence of technical bias. Normalization by RUVg reduced heterogeneity between fractions and centered on 0 almost all the ERCC spike-in counts as shown by the RLE plots (Fig. 9B).

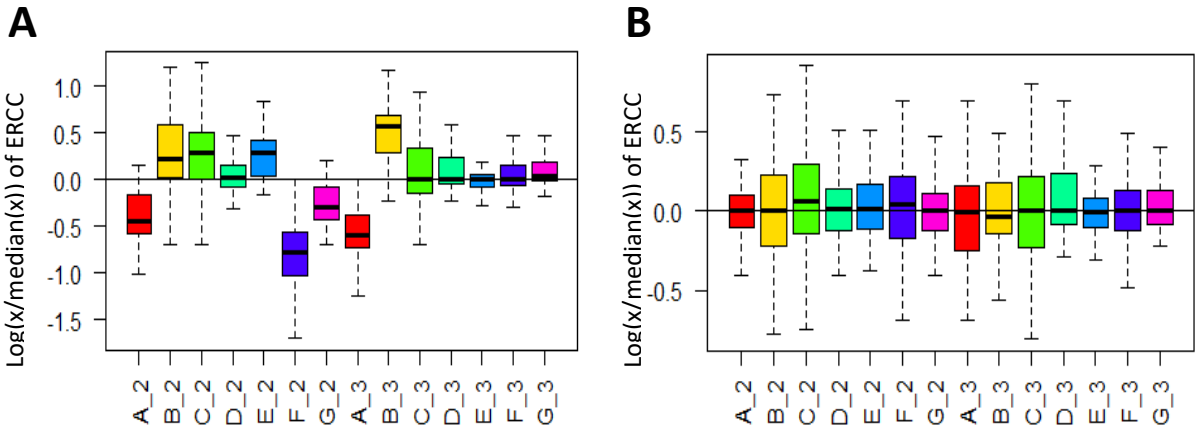


Figure 9: RLE plots of ERCC spike-ins read counts (A) before and (B) after removing unwanted variation (by RUVg) in method 2.

Another way to visualize the disparity between samples is the principal component analysis (PCA). The clustering of samples is a good indicator of the reproducibility between replicates. After normalization, we expected reduced distances between replicates. Figure 10 shows the PCA plots of the raw data and the normalized results after all normalization steps by methods 1 and 2. Normalization by both methods effectively improved the clustering of the different replicates for fractions B, D, E and G when compared to non-normalized data. The two normalization methods reduced non-biological variation.

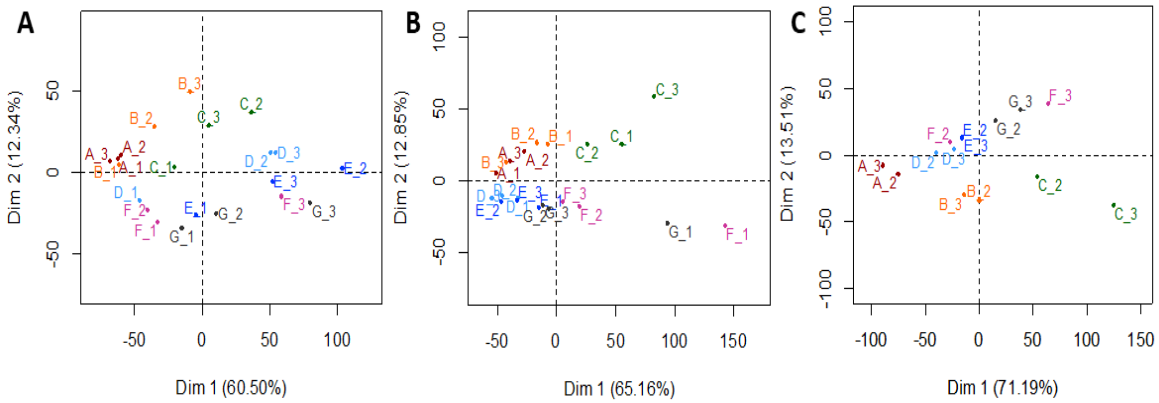


Figure 10: PCA plots showing (A) non-normalized data, (B) normalized data with method 1 (replicates 1, 2 and 3) and (C) normalized data with method 2 (replicates 2 and 3). Different replicates of same fraction are shown with the same color. The distance between replicates indicates the variability between replicates.

Method 1 based on the quantification of very low amounts of ribodepleted RNA in the range of ng was expected to be not so accurate. Method 2 was thought to be more accurate because it is a direct normalization step based on control ERCC spike-ins. However, method 2 was applied to only two replicates whereas method 1 used the three replicates. Therefore, to help our choice of a normalization method, we compared the impacts of two normalizations on the translatome variables, ribosome occupancy and ribosome density.

VI) Impacts of the normalization methods on ribosome occupancy and ribosome density

Normalized data by the two methods were used to calculate ribosome occupancy (RO) and ribosome density (RD) of each gene. We then compared RO and RD values, distribution and correlation. We restricted our analysis to the 1654 monocistronic genes of *E. coli* MG1655.

The ribosome occupancy of a gene is the proportion of its mRNA copies associated with at least one ribosome. Both normalization methods provided the same number of genes with a calculated RO (1563 genes). They both led to very high RO values (median values above 90%) but their distributions slightly differed, method 2 resulting in average in higher RO values (Fig. 11 A and B). Although absolute RO values differed between methods, the RO values obtained with both methods were highly correlated (Pearson correlation coefficient of 0.96, p-value $<2.2\ 10^{-16}$) (Fig. 11C).

For the determination of ribosome density, we first identified the peak fraction of each gene (see the Material and Methods section of chapter III for more details on RD calculation). The peak fraction corresponds to the highest mRNA copy proportion within fractions B to G of the polysome profiling experiment (Fig. 1). Using normalized data by either method 1 or method 2, the peak fraction was determined by a bootstrap method on residuals with a confidence interval fixed at 95% (Picard et al. 2012). The distributions of the peak fractions obtained by the two normalizations are shown in figure 12. In both cases, most of the genes had an identified peak fraction in fraction C (between 2 and 4 ribosomes) and only few genes exhibited a peak fraction in the more heavily ribosome loaded fractions. The major difference was the number of genes with an identified peak fraction: a higher number was obtained with method 2 (1077 genes) than with method 1 (732 genes). The two methods shared 599 common genes whereas 233 genes had only a peak fraction when normalized with method 1 (5, 225 and 3 genes in fractions B, C and F, respectively) and 473 genes had only a peak fraction when normalized with method 2 (13, 432 and 28 genes in fractions B, C and G, respectively).

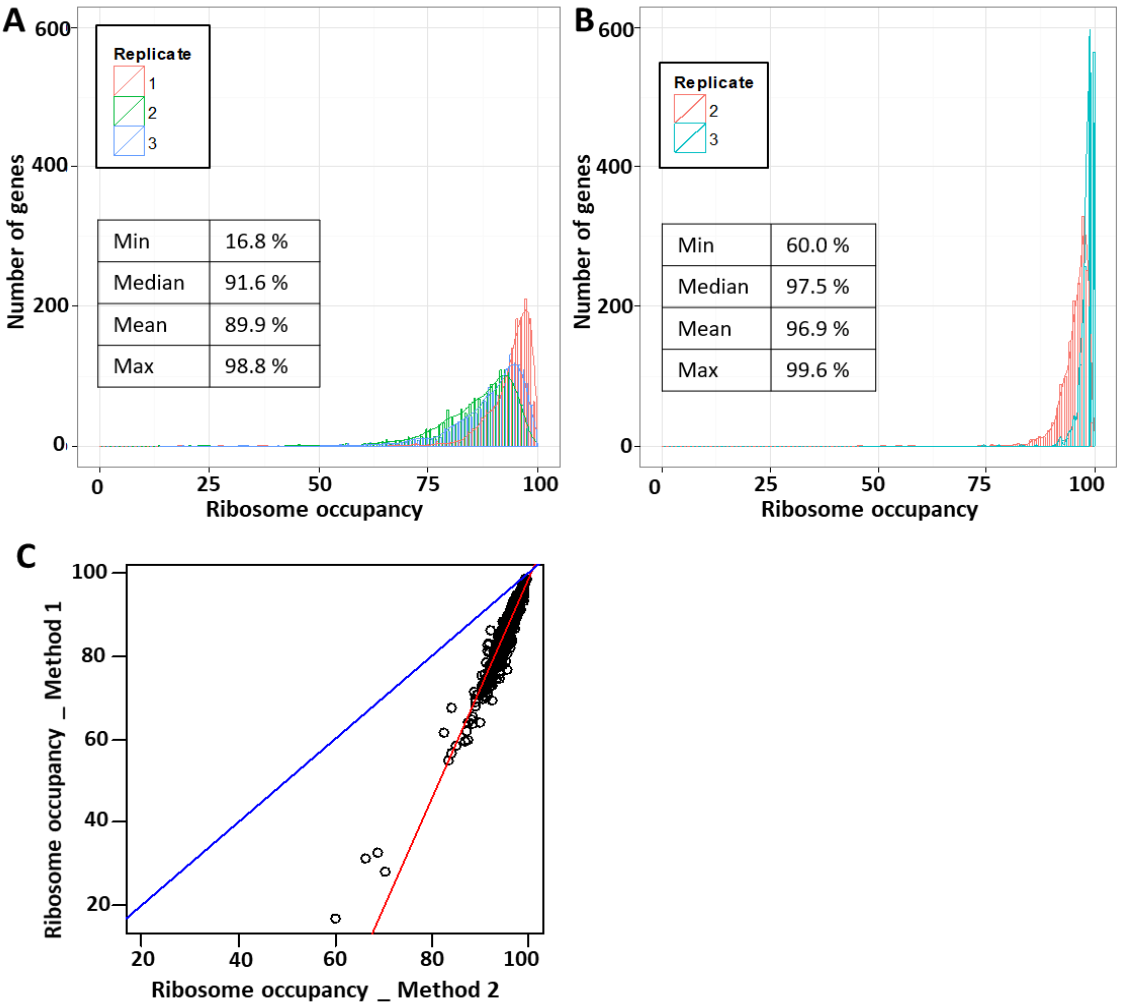


Figure 11: Comparison of the ribosome occupancy values using the two normalization methods. Distribution of RO values calculated from normalized data (A) using method 1 and (B) using method 2. For each gene, the mean RO value was calculated from three replicates in method 1 and from two replicates in method 2. (C) Scatter plot between RO values obtained by methods 1 and 2 (1563 common genes). The linear regression is in red and the line y=x line is in blue.

The ribosome density corresponds to the number of bound ribosomes in the identified peak fraction normalized to the ORF length. Consequently, more genes had a RD value (1072 genes) with method 2 than with method 1 (732 genes). Despite different gene number, the RD distributions were similar with the two methods with rather low RD values (median values around 0.4-0.5 ribosomes/100 nts) (Fig. 13). RD values were in average slightly higher for method 1, this might come from the higher number of genes having a peak fraction in fraction F, a heavily ribosome loaded fraction (Fig. 12).

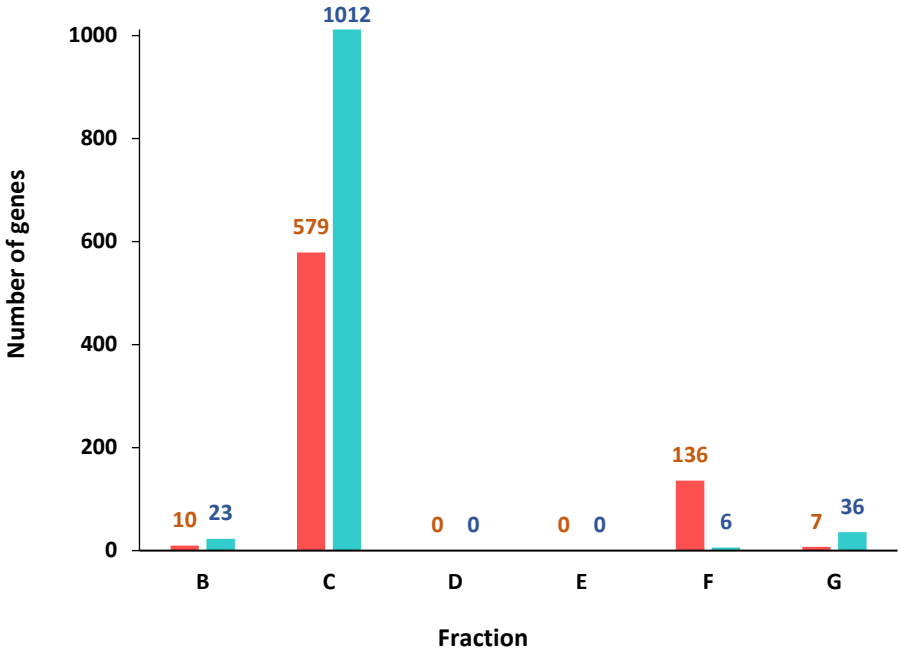


Figure 12: Distribution of peak fraction obtained by method 1 (orange) and method 2 (blue). Peak fraction was identified using a bootstrap method from three replicates (method 1) and two replicates (method 2).

We found a very high correlation (Pearson correlation coefficient of 0.98, p-value $<2.2\ 10^{-16}$) between RD values calculated by the two methods (Fig. 13C). Indeed, for the 599 common genes having an identified peak fraction by the two methods, all, except 8, had the same identified peak fraction and therefore exactly the same ribosome density.

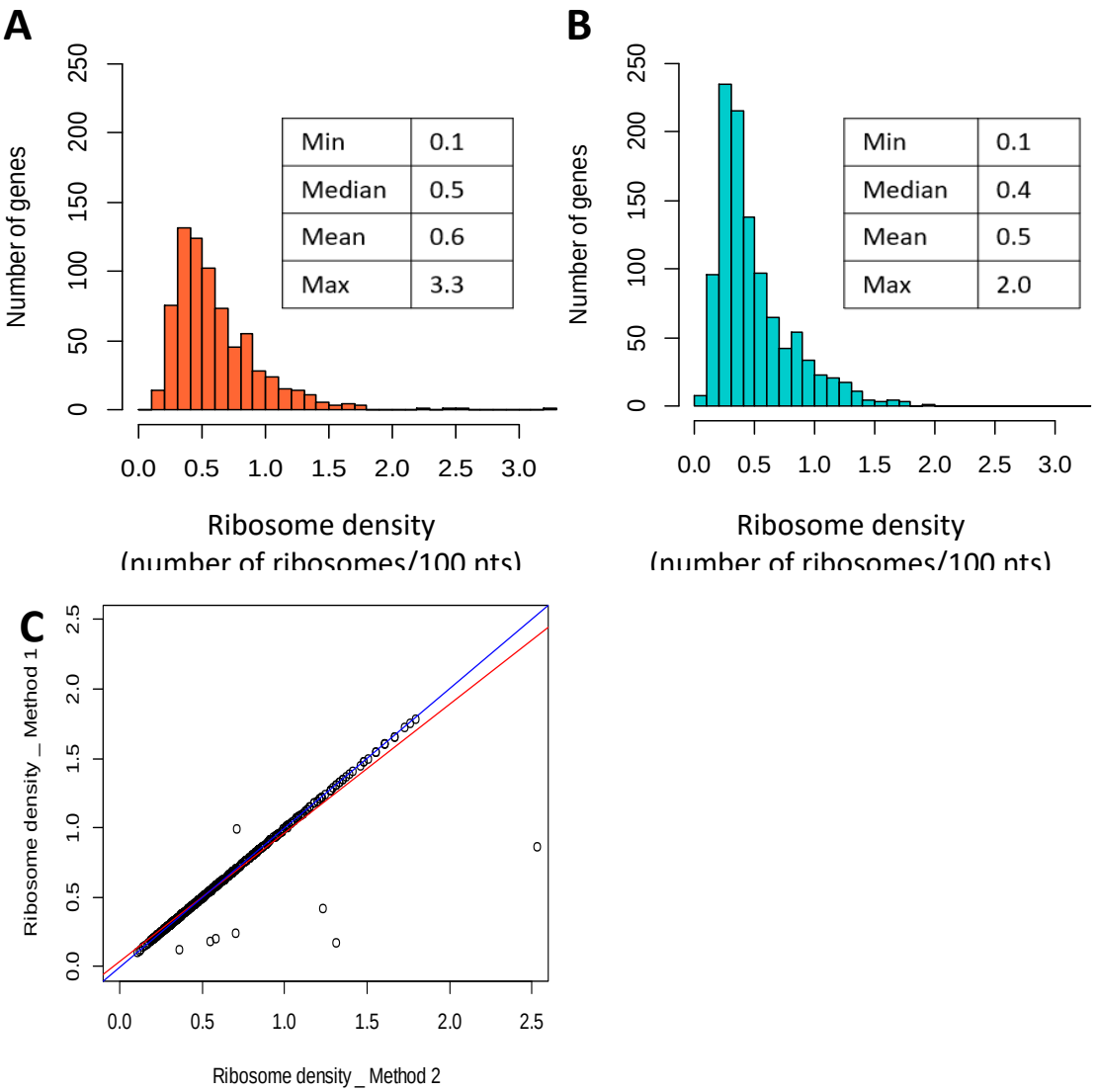


Figure 13: Comparison of the ribosome density values using the two normalization methods. Distribution of RD values calculated from normalized data (A) using method 1 and (B) using method 2. (C) corresponding mathematical values. (C) Scatter plot between RD values obtained by methods 1 and 2 (599 common genes). The linear regression is in red and the line y=x line is in blue.

Altogether, we showed very consistent results for the translatome variables RO and RD independently of the normalization methods. Although the two normalizations comprised different steps and packages, the RO and RD results were quite robust and consistent. However, while the same number of genes with a RO value was obtained with both methods, method 2 allowed to obtain substantially more genes with an identified peak fraction and therefore more RD values. To be able to select one of the two methods, we further analyzed the normalization effect on the identification of the major determinants of RO and RD.

VII) Impacts of the normalization methods on the key determinants of RO and RD

To identify the major determinants of RO and RD, we used linear covariance models to select and rank the quantitative and qualitative parameters which respectively explained RO and RD variability between genes. Relevant parameters were selected by the Akaike Information Criteria (AIC) and significant determinants were identified with a p-value < 0.05 (Dressaire et al. 2010).The method of linear covariance model is described in more details in the Material and Methods section of chapter III.

For RO, the two covariance models based on the two normalization methods had the same number of genes (1100 genes) and R^2 (0.26). Similar significant parameters (mRNA concentration, having a signal peptide, hydrophobicity, being a target of CsrA) were selected with the same positive or negative effect on RO (Fig. 14A). Related to protein localization, the two parameters "being an inner membrane protein" and "being located at the membrane" showed low RO values with one of the two methods. Some parameters were only significant with one normalization method (ORF GC% with method 2, position on chromosome with method 1). This high similarity between the results of the RO models was expected because the two normalization methods provided very correlated data for 1563 genes.

In the models of RD, the set of genes with method 1 (501 genes) was lower than the one (745 genes) with method 2. The adjusted R^2 obtained were very low (<0.2) with both methods and the one with method 1 (0.08) was half the one obtained with method 2 (0.16). In term of significant parameters for RD, some concordances were observed between the two methods (being an inner membrane protein or a CsrA target and the ORF GC%) but less significant parameters were identified for method 1 (Fig. 14).

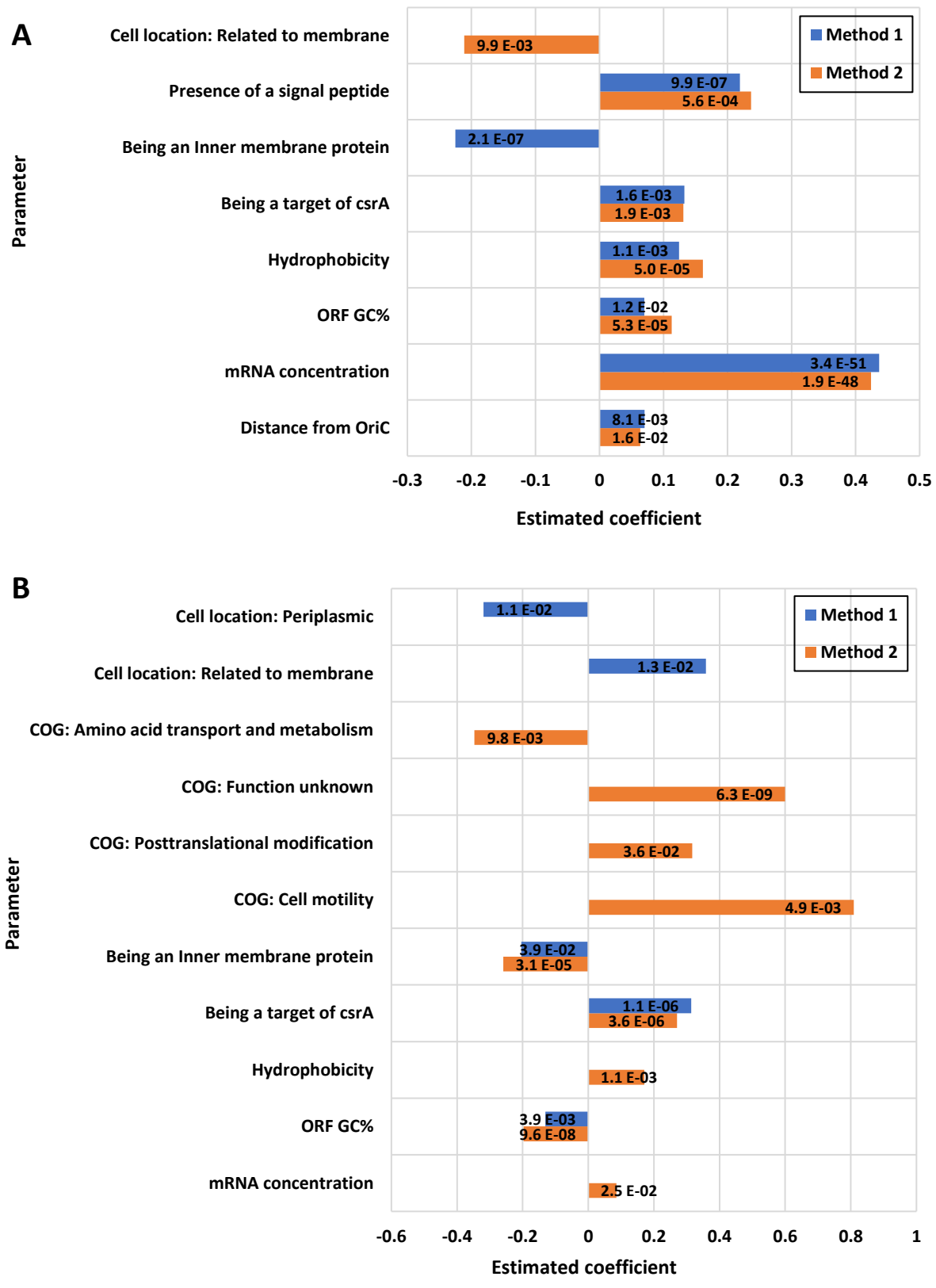


Figure 14: Significant parameters selected by the linear covariance models to explain the variability of (A) ribosome occupancy and (B) ribosome density using normalized data obtained with method 1 (orange) and with method 2 (blue). Length of bar corresponds to the value of estimated coefficient and corresponding p-value is shown on the bar.

Although until now, the results were rather consistent with the two normalization methods it is the higher number of selected factors for method 2 in the RD covariance model that led us to select method 2 for normalizing our RNA-Seq data. This is probably related to the larger set of RD values available with method 2 for searching RD determinants.

VIII) Effect of 5'UTR

5' untranslated regions are involved in translation regulation for example through riboswitch, RBS strength and binding of sRNA and RNA-binding proteins (Baker et al. 2002; Caron et al. 2012; Kosuri et al. 2013; Storz et al. 2011). To analyze the influence on RO and RD values of counting or not the reads on the 5'UTR of the mRNA molecules, we compared the RO and RD values when counting on the 5'UTR+ORF or only on the ORF (as performed until now). We restricted the analysis to the monocistronic genes that have an experimentally identified transcriptional start site (Kim et al. 2012; Thomason et al. 2015) leading to smaller sets of 1568 genes for RO and 861 genes for RD.

There was no significant difference on the RO values (mean and median values around 97%) by including the 5'UTR in the read counting. Concerning RD, the identified peak fractions were the same with or without the 5'UTR. Dividing now the ribosome number of the peak fraction by the 5'UTR+ORF length instead of the ORF length resulted in a slight decrease of RD values. In the linear covariance models to identify the major determinants of RO and RD we now included new 5'UTR-related parameters, the ΔG of the -30 +24 bp region starting from the start codon, the 5'UTR length, the 5'UTR GC% and the nature of 2nd nucleotide of the mRNA molecule. Similar results were obtained for both counting methods on the selection of RO determinants (for the 5'UTR+ORF counting, the model included 873 genes and corresponded to an adjusted R^2 of 0.29) (Fig. 15A). For RD, the results were consistent for most of the selected determinants although some parameters were only selected for one of the two counting methods (5'UTR GC%, essentiality and location at the periplasm when counting on the ORF and signal peptide, cell motility and CAI when counting on the 5'UTR+ORF) (Fig. 15B). These discrepancies may be explained by different sets of genes included in the two RD models (607 genes when counting on the 5'UTR+ORF against 745 genes when counting only on the ORF) and the slight differences in RD values.

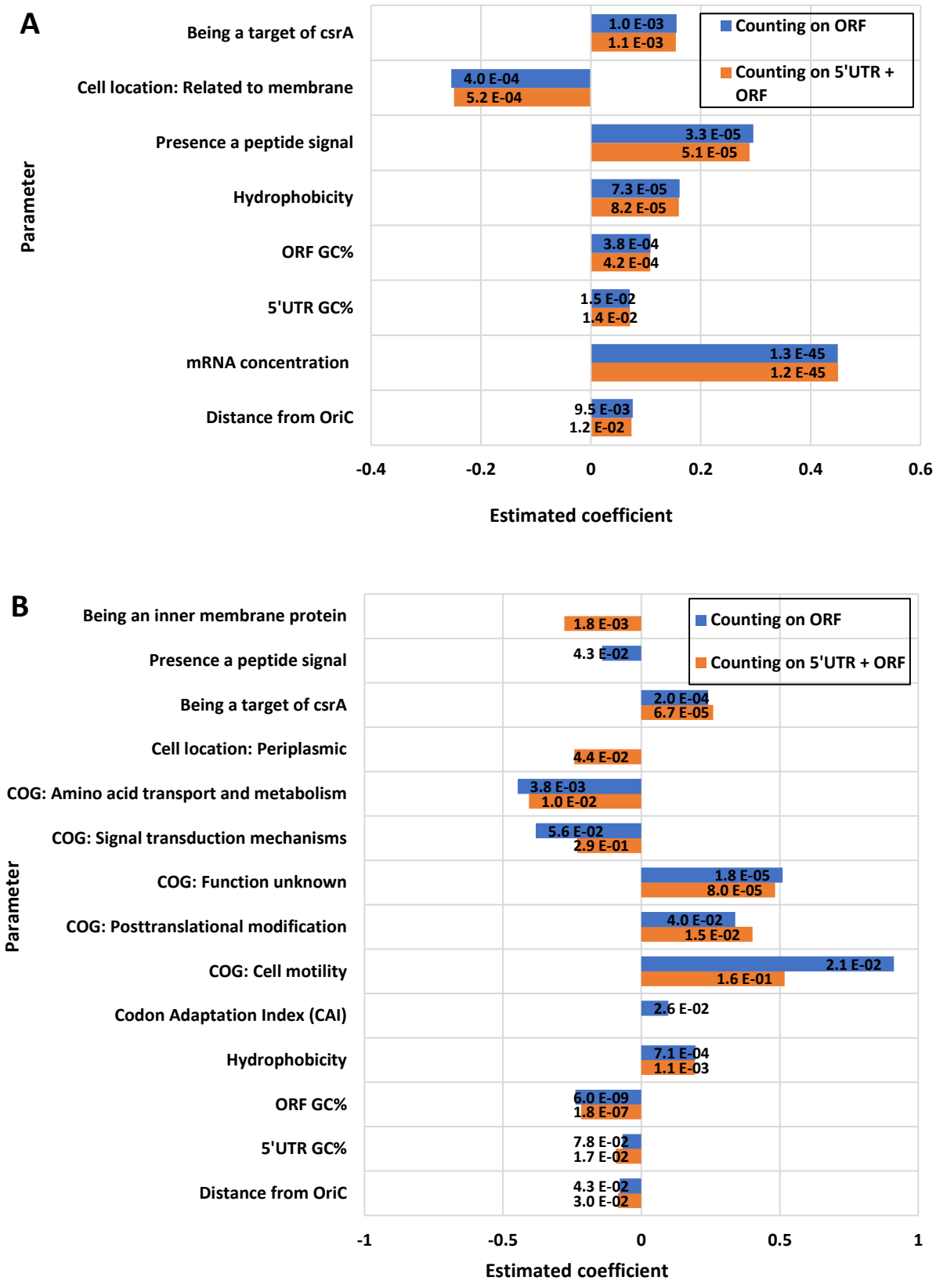


Figure 15: Significant parameters selected by the linear covariance models to explain the variability of (A) ribosome occupancy and (B) ribosome density using read counting on the ORF only (orange) and on the 5'UTR + ORF (blue). Length of bar corresponds to the value of estimated coefficient and the corresponding p-value is shown on the bar.

Altogether, we did not find significant changes on RO and RD value when the counting on the 5'UTR was included. The main counting effect was only on some determinants of RD. Counting on the 5'UTR changed the absolute read counts of an mRNA but did not significantly affect the relative proportions of the mRNA copies between the fractions. We decided to keep counting only on the ORF because the integration of 5'UTR decreased the numbers of genes included in the analyses.

CONCLUSION and DISCUSSION

RNA-Seq is a sensitive tool to detect and quantify RNA. Critical steps of RNA-Seq data analysis are read mapping and data normalizations. They must be accurate and appropriate to the experimental protocol. Here for each step, we considered two methods and selected the better one. For mapping, we compared tmap, the tool provided by the ion torrent machine and bwa an independent commonly used alignment tool. As tmap includes bwa and closed methods, therefore counts obtained by the two methods were almost the same. Different options between the two pipelines resulted in small differences. We decided to use bwa because it is easier to exclude multi-mapped reads. Based on counts with bwa, we also verified the absence of DNA contamination and removed mRNAs with too low counts.

For normalization, two strategies were compared, one included DESeq and the other one based on RUV method. Both relied on the hypothesis of unvaried genes and corrected the data by removing potential non-biological variations. The first method applied this hypothesis to all the genes whereas the second restricted the hypothesis to the control ERCC spike-ins. However, the hypothesis of non-differential expression for all the genes between the fractions as suggested by DESeq was less likely since each gene had different mRNA copy proportions in the different fractions. The hypothesis of DESeq would be more appropriate for normalizations only within replicates of the same fraction. Moreover, method 1 required normalization by the RNA quantity after ribodepletion, which could be inaccurate due to experimental errors when measuring very low RNA amounts. In contrast, method 2 corrected data in one step without the intermediate measurement of ribodepleted RNA amounts. Nevertheless, the drawback of method 2 was the reduced number of replicates used in the analysis (only 2 instead of 3 for method 1). Despite these differences, both methods showed high concordance in ribosome occupancy and ribosome density values, which underlines the robustness

of our results. Finally, the identification of the key factors involved in RO and RD determination showed very consistent results with the two methods but method 2 provided a higher number of selected determinants for RD. Therefore, we chose method 2 for RNA-Seq data normalization.

In conclusion, bwa mapping and RUV normalization were selected to provide reliable data for further biological analysis. We also showed that counting on the 5'UTR+ORF did not significantly modify the RO and RD results but reduced the set of analyzed genes. Since our objective was to understand the translational regulation of as much as possible genes we restricted the read counting to the ORF.

REFERENCES

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

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: Regulation of glycogen biosynthesis in E. coli. Molecular Microbiology *44*, 1599–1610.

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. Proceedings of the National Academy of Sciences *109*, E3444–E3453.

Croucher, N.J., and Thomson, N.R. (2010). Studying bacterial transcriptomes using RNA-seq. Current Opinion in Microbiology *13*, 619–624.

Dressaire, C., Laurent, B., Loubière, P., Besse, P., and Cocaign-Bousquet, M. (2010). Linear covariance models to examine the determinants of protein levels in Lactococcus lactis. Molecular BioSystems *6*, 1255.

Esquerré, 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 Research *42*, 2460–2472.

Feng, H., Zhang, X., and Zhang, C. (2015). MRIN for direct assessment of genome-wide and gene-specific mRNA integrity from large-scale RNA-sequencing data. Nature Communications *6*, 1–10.

Gagnon-Bartsch, J.A., and Speed, T.P. (2012). Using control genes to correct for unwanted variation in microarray data. Biostatistics (Oxford, England) *13*, 539–552.

Gandolfo, L.C., and Speed, T.P. (2018). RLE plots: Visualizing unwanted variation in high dimensional data. PLOS ONE *13*, e0191629.

Ingolia (2009). Genome wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *29*, 5095–5101.

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 <italic>Escherichia coli</italic> and <italic>Klebsiella pneumoniae</italic> by Genome-Wide Transcription Start Site Profiling. PLoS Genet *8*, e1002867.

Kosuri, S., Goodman, D.B., Cambray, G., Mutalik, V.K., Gao, Y., Arkin, A.P., Endy, D., and Church, G.M. (2013). Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. Proceedings of the National Academy of Sciences *110*, 14024–14029.

Li, H. (2012). Exploring single-sample snp and indel calling with whole-genome de novo assembly. Bioinformatics *28*, 1838–1844.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *25*, 1754–1760.

Passalacqua, K.D., Varadarajan, A., Ondov, B.D., Okou, D.T., Zwick, M.E., and Bergman, N.H. (2009). Structure and Complexity of a Bacterial Transcriptome. Journal of Bacteriology *191*, 3203–3211.

Risso, D., Ngai, J., Speed, T.P., and Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. Nature Biotechnology *32*, 896–902.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nature Biotechnology *29*, 24–26.

Sharma, C.M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiß, S., Sittka, A., Chabas, S., Reiche, K., Hackermüller, J., Reinhardt, R., et al. (2010). The primary transcriptome of the major human pathogen Helicobacter pylori. Nature *464*, 250–255.

Sittka, A., Lucchini, S., Papenfort, K., Sharma, C.M., Rolle, K., Binnewies, T.T., Hinton, J.C.D., and Vogel, J. (2008). Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. PLoS Genetics *4*.

Storz, G., Vogel, J., and Wassarman, K.M. (2011). Regulation by Small RNAs in Bacteria: Expanding Frontiers. Molecular Cell *43*, 880–891.

Thomason, M.K., Bischler, T., Eisenbart, S.K., Förstner, K.U., Zhang, A., Herbig, A., Nieselt, K., Sharma, C.M., and Storza, G. (2015a). Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in Escherichia coli. Journal of Bacteriology *197*, 18–28.

Thomason, M.K., Bischler, T., Eisenbart, S.K., Förstner, K.U., Zhang, A., Herbig, A., Nieselt, K., Sharma, C.M., and Storz, G. (2015b). Global Transcriptional Start Site Mapping Using Differential RNA Sequencing Reveals Novel Antisense RNAs in Escherichia coli. Journal of Bacteriology *197*, 18–28.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics *10*, 57–63.

- Chapitre 3 -

Co-directional regulations of transcription and translation in *Escherichia coli*: more concentrated mRNAs are more efficiently translated

Dans ce dernier chapitre, l'état traductionnel de l'ensemble des ARNm chez *E. coli* en croissance rapide est présenté. Pour chacun des 1654 gènes monocistroniques, deux variables de traduction ont été estimées : le pourcentage de copies en traduction d'un ARNm (RO) et la densité en ribosomes (RD). Ces données ont été obtenues par la méthode du polysome profiling (présentée dans le chapitre I) suivi de la quantification par RNA-Seq. La méthode d'alignement des séquence bwa et la normalisation RUV qui utilisent les contrôles externes ERCC ont été appliquées pour le traitement des données RNA-Seq (comme présenté dans le chapitre II).

Nous avons montré que le pourcentage de copies en traduction d'un ARNm était corrélé avec la vitesse d'initiation (estimée par RBS calculator) et la densité en ribosomes avec la vitesse globale de traduction (estimée par le logiciel Transim). RO reflèterait l'efficacité d'initiation de la traduction et RD serait dépendante des efficacités de l'initiation et de l'élongation de la traduction. Pour la plupart des ARNm, nous avons montré que le pourcentage de copies en traduction d'un ARNm était très élevé, alors que la densité en ribosomes était plutôt faible. Cela signifierait que pour la majorité des gènes l'initiation de la traduction est plutôt efficace et que l'élongation n'est pas une étape fortement limitante de la traduction. En effet, une faible densité pourrait résulter d'une vitesse d'élongation rapide alors qu'une très forte densité (proche de la densité maximale théorique) pourrait être due à un embouteillage de ribosomes. Par ailleurs, les deux variables RO et RD sont corrélées positivement avec la teneur en protéines. L'ensemble de ces résultats suggère que la traduction d'un ARNm serait limitée par l'initiation avec un effet possible mais dans une moindre mesure de l'élongation.

Même si l'initiation et l'élongation sont efficaces pour la majorité des gènes, nous notons néanmoins une variabilité des valeurs de RO et RD entre les gènes. Pour essayer d'expliquer cette hétérogénéité, nous avons utilisé une approche de régressions linéaires multiples pour identifier et hiérarchiser les déterminants majeurs de la régulation de RO et RD à l'échelle du génome. Nous avons identifié comme déterminants de la traduction des facteurs liés à la séquence de l'ARNm (motifs conservés dans le 5'UTR et au niveau des tout premiers codons après l'AUG et le %GC de la séquence codante) et aux caractéristiques des protéines (fonction et localisation).

De façon plus surprenante, la concentration en ARNm, un paramètre qui dépend de la physiologie de la cellule et des conditions de croissance, a été identifié comme le principal facteur positif de RO avec un effet positif sur RD. L'effet de la concentration de l'ARNm sur la traduction a d'abord été testé à l'échelle moléculaire avec le gène chromosomique *lacZ*. La concentration de l'ARNm *lacZ* a été augmentée par induction transcriptionnelle (promoteur P_{lac}) et les profils polysomiques de l'ARNm *lacZ* à faible concentration (en absence d'induction) et à forte concentration (après induction à l'IPTG) ont été comparés. Nous avons mis en évidence une augmentation de l'activité de traduction de *lacZ* lors de l'augmentation du niveau de son ARNm. Nous avons observé le même type de réponse traductionnelle lors de l'augmentation de la transcription du gène *lacZ* exprimé à partir d'un plasmide sous la dépendance d'un promoteur inductible à l'arabinose. La proportion de copies en traduction et les charges en ribosomes augmentent de façon importante. Dans les deux cas, nous avons noté un déplacement de la charge en ribosomes de la fraction des ARNm libres (sans ribosome, fraction A) vers les fractions correspondantes aux ARNm les plus chargés (fractions F et G).

Nous avons alors construit cinq autres souches exprimant de manière plasmidique sous la dépendance d'un promoteur inductible à l'arabinose cinq autres gènes cibles *(cysZ, inaA, yjcO, ucpA* et *yeeZ*) pour étendre la validation de l'effet de la concentration de l'ARNm sur la traduction à plus de gènes. L'étude simultanée de plusieurs souches est réalisée par la méthode de multiplexage de polysome profiling présentée dans le chapitre I. Nous avons montré l'effet positif du niveau de l'ARNm sur la traduction de tous les gènes sélectionnés. L'augmentation de la concentration en ARNm provoque généralement des augmentations du pourcentage de copies en traduction et de la charge en ribosomes. Cette réponse n'est toutefois pas exactement identique pour tous les gènes. Ces validations moléculaires permettent de mettre en évidence pour la première fois le lien de causalité entre la transcription et la traduction d'un gène.

En conclusion, ces travaux ont permis d'obtenir une image globale de la traduction chez *E. coli* en croissance rapide. Lors de cette croissance, la traduction est efficace caractérisée par une fréquence d'initiation élevée et une vitesse d'élongation rapide afin de pouvoir répondre aux forts besoins cellulaires en protéines. Par ailleurs, la recherche des déterminants de traduction au niveau –omique montre l'importance

du paramètre physiologique « concentration de l'ARNm ». Son effet positif sur la traduction a été prédit au niveau du génome par l'analyse statistique et validé à l'échelle moléculaire pour un groupe de six gènes. La concentration de l'ARNm est une variable peu étudiée par rapport aux facteurs liés à la séquence et l'effet de causalité de la transcription sur la traduction n'a jamais été mis en évidence. Nos résultats démontrent la codirectionalité des régulations traductionnelle et transcriptionnelle dans le contrôle de l'expression génique. Une hypothèse serait que plus un ARNm est concentré, plus sa probabilité de rencontre avec les ribosomes est élevée. Pour mieux comprendre le lien entre concentration en ARNm et traduction, nous aurions souhaité intégrer les données de stabilome, temps de demi-vie de l'ensemble des ARNm que nous avions à disponibilité mais ce paramètre n'a pas pu être inclus dans le modèle linéaire car le temps de demi-vie d'un ARNm est très corrélé à sa concentration. Le mécanisme précis liant concentration en ARN et traduction reste donc à étudier.

Ce travail est rédigé sous la forme d'un article qui sera prochainement soumis dans une revue à comité de lecture.

Co-directional regulations of transcription and translation in Escherichia coli: more concentrated mRNAs are more efficiently translated

H. L. Nguyen et al.

ABSTRACT

In this study, translation regulation was investigated at the genome scale in fast growing *E. coli* cells. Using polysome profiling method, the ribosome occupancy (RO) and ribosome density (RD) of 1654 genes were calculated. For most genes, nearly all the mRNA copies were undergoing translation but were loaded with a rather low ribosome number. Both RO and RD were positively correlated with the protein level, RO through the positive correlation with translation initiation and RD with the translation rate. A multiple linear regression approach identified key parameters involved in the genome-wide regulation of RO and RD. Unexpectedly, the physiological parameter, mRNA concentration was identified as the main positive factor of RO and with a positive effect on RD. Using a selected set of genes, we showed that increasing mRNA concentration by transcriptional induction resulted in increase of ribosome occupancy and ribosome load. These results demonstrated the co-directional regulation of transcription and translation in *E. coli.* The co-transcriptional regulation of translation may contribute to physiological adaptation when cells regulate the mRNA level of genes to adapt to environmental changes.

INTRODUCTION

Protein synthesis is one of the most important biological process. Translation of mRNA into proteins is the largest energy consumer process in bacterial cell (Russell and Cook 1995). Translational control constitutes a non-negligible part of gene expression and is usually involved in responses to environmental cues and stress (Bartholomäus et al. 2016; Gallant 1979; Picard et al. 2012, 2013). Deciphering how translation is regulated is central to understand microbial physiology and adaptation and to further design and

optimize homologous/heterologous protein synthesis in the context of synthetic biology.

Studying translational regulation at the genome scale was possible with the development of high throughput methods such as microarray and RNA-Seq. By providing a global view of translation status of all genes, these methods allowed to access translational efficiency and explore the diversity within microbial genomes (Arava et al. 2003; Ingolia 2009; Lackner et al. 2007; Li et al. 2014; Oh et al. 2011; Picard et al. 2012). Experimental translation status at the genome-wide scale is generally measured by the ribosome profiling method (Dana and Tuller 2012; Ingolia 2009; Li et al. 2014; Liu et al. 2013; Oh et al. 2011). This method provides an average ribosome density that is defined by the total number of bound ribosomes for the entire population of mRNA copies. However it has been previously shown that for one mRNA the ribosome density is heterogeneous since the number of bound ribosomes differs among the mRNAs copies (Arava et al. 2003; Picard et al. 2012). Only the polysome profiling method provides the detailed load in ribosomes of each mRNA copy (Arava et al. 2003; Picard et al. 2012). Furthermore, this polysome profiling method allows for each gene, the proportion of its mRNA copies undergoing translation defined as the ribosome occupancy (RO) to be quantified.

Noting the variability of translation status between genes, many studies search for general rules which control translation. Factors largely studied and modelled for global roles in translation regulation are mainly sequence related features (Reuveni et al. 2011; Shaham and Tuller 2017) such as the level of secondary structure of 5'UTR, sequence affecting initiation (strength of ribosome binding site in prokaryotes), local and global codon usage bias of the coding sequence. Their effects on translation are not fully understood and were led to contradictory results (Kudla 2009; Nakahigashi et al. 2014; Osterman et al. 2013; Tuller et al. 2010).

In this study, we investigated translation regulation in *E. coli* in exponential growth. The genome-wide translation status was analyzed by the polysome profiling approach. Each gene was characterized by the ribosome occupancy and ribosome density which were respectively correlated to translation initiation and translation rate. We used a multiple linear regression approach to study and hierarchy effects of multiple parameters on the ribosome occupancy and ribosome density values. We

identified factors related to gene sequence (conserved motif, GC% of the ORF), to protein features (functional category and localization) and more unexpectedly the mRNA concentration, a factor which depends on cell physiology and growth conditions. The positive effect of mRNA concentration on translation was confirmed by transcriptional induction of a selected set of genes.

MATERIALS and METHODS

Strains and growth conditions

E. coli MG1655 was grown in bioreactor under aerobic condition in batch culture at 37 °C, pH 7.0 and 350 rpm, in a M9 medium supplemented with 3 g/L glucose as previously described (Esquerré et al. 2015).

Experiments of chromosomal induction of *lacZ* transcription (coding for a β-galactosidase) were performed in *E. coli* MG1655 flask cultures in M9 glucose medium at 37°C and 150 rpm in exponential phase at an OD of 0.6 and induced by 1 mM IPTG for 2 hours. For plasmidic transcription induction, we selected five genes of *E. coli* in addition to *lacZ*: *cysZ* (sulfate transporter), *inaA* (putative lipopolysaccharide kinase), *ucpA* (putative NAD binding oxidoreductase), *yeeZ* (putative epimerase) and *yjcO* (unknown function). They are all non-essential monocistronic genes with a known promoter and an experimentally identified primary Transcription start sites (TSS) (after combining two experimental datasets in minimum medium (Kim et al. 2012; Thomason et al. 2015) and selecting a TSS length < 300 nts). These genes also exhibited a low mRNA level during exponential growth in M9 glucose (Esquerré et al. 2014) and low values of RO and RD in this study. For each of gene, the 5'UTR + ORF fragment was amplified by PCR and cloned in a P_{BAD}/*myc*/His vector (Invitrogen) to obtain the construction: P_{BAD} inducible promoter – 5'UTR+ORF – *myc*/His tag. The construction with *lacZ* was introduced in the strain *E. coli* MG1655 *ΔlacZ* Δ*araFGH*, Ω*pcp18*::*araE533* (Nouaille et al. 2017) *whereas the other five plasmids were* transformed individually into modified strain *E. coli* MG1655 Δ*araFGH*, Ω*pcp18*::*araE533* (Ah-Seng 2013). All strains were grown in flasks in M9 glucose medium with 0.1 mg/mL ampicillin, at 37°C, under shaking (150 rpm). Arabinose was added at a final concentration of 0.001% (w/v) when the culture reached an OD_{600} of 1 (exponential growth). After 30 minutes of induction, translation elongation was arrested by addition of 0.1 mg/mL chloramphenicol. Multiplexing polysome profiling experiments as previously

described in Nguyen et al. (Nguyen et al. 2019) were performed: one experiment mixing cell free extracts of the six strains (*cysZ*, *inaA*, *ucpA*, *yeeZ*, *yjcO* and *lacZ*) obtained without arabinose induction and one experiment mixing cell free extracts of the six strains with arabinose induction.

Polysome profiling experiment

Classic and multiplexed polysome profiling experiments were performed as previously described in Nguyen et al. (Nguyen et al. 2019). Briefly, translation elongation in cells in exponential phase was arrested by adding 0.1 mg/mL chloramphenicol and cells were collected on ice. Cells were harvested, resuspended in lysis buffer and washed twice. After cell disruption using glass beads, mRNA-ribosome complexes were size-fractionated on sucrose gradient in 24 sub-fractions. Total RNA of all 24 sub-fractions was extracted using the extraction RNeasy Midi kit (Qiagen), 16S and 23S rRNAs were quantified with Bioanalyzer 2100 (Agilent). The 23S/16S and 16S/23S ratios were calculated for each sub-fraction. Once identified the component of each sub-fraction, sub-fractions were regrouped into 7 fractions from A to G. Fraction A comprised sub-fractions containing the DNAs, free RNAs, free small and free large ribosome subunit. From fractions B to G, 23S/16S ratios were constant around 1.8 and match to entire ribosomes. The 1st peak (fraction B) was attributed to the monosome. The 2nd, 3rd and 4th peaks corresponded to 2, 3 and 4 ribosomes respectively. The number of ribosomes in other fractions was extrapolated as previously described (Picard et al. 2012). The percentage of ribosomes engaged in translation was estimated by area integration of the polysomal profile as described in Picard 2012. Protein denaturation, nucleic acid precipitation and total RNA extraction were performed in each fraction as previously described (Nguyen et al. 2019).

RNA quantification by RNA-Seq

An equal amount of Ambion™ ERCC RNA Spike-In mix (1µL of ERCC mix diluted at 1/10 per 5 µg of total RNA) was added to each fraction. Ribosomal RNA was removed by Ribo-Zero^{RM} Magnetic Kit (for gram-negative bacteria) following manufacturer's protocol. RNA was finally purified by RNeasy mini kit (Qiagen). 10 ng of ribodepleted RNA of each fraction was used for sequencing library construction as described in second chapter. Library construction and sequencing were performed by the GeT-BioPuces platform (http://get-biopuces.insa-toulouse.fr/). These experiments were repeated with two independent cultures.

Reads issued from RNA-Seq sequencing were mapped on *E. coli* genome version U0009.3 (GenBank) using BWA version 0.7.12-r1069 (Burrows-Wheeler Alignment) method (Li and Durbin 2009). Counting was performed with HTSeq-count version 0.6.1p1 using the « intersection non empty » mode. Transcripts with read counts less than the reliable quantification threshold of 10 were excluded from the analysis as described in chapter 2. To correct experimental variations from ribodepletion to sequencing, a first normalization based on the method of remove unwanted variation (RUV, RUVSeq package of R software) was applied using the ERCC spike-ins introduced in each fraction before the ribodepletion step. We noted $N_{i,j,k\ NormRUVg}$ the count of the gene *i*, in the fraction *j* and replicate *k* after normalizing with the RUV method, with $i \in \{1,\ldots,4497\}$, $j \in \{A,B,C,D,E,F,G\}$ and $k \in \{1,2\}$. A second normalization was applied to take into account the variability of initial amount of total RNA between fractions as a constant 5 µg of total RNA was taken in each fraction for ribodepletion. We obtained the final normalized count of the gene *i*, in the fraction *j* and replicate *k*:

$$N_{i,j,k} = N_{i,j,k\ NormRUVg}\, x\, \frac{total\ RNA\ quantity\ j,k}{5}$$

For each gene *i*, we calculated the proportion of mRNA copies in each fraction *j*:

$$mRNA\ proportion\ in\ fraction_j\ (\%) = \frac{Ni,j,k}{\sum_{j=A}^{G} Ni,j,k} \text{ x } 100$$

RNA quantification by real-time quantitative PCR

Total RNA (5 µg) was reverse-transcribed to yield cDNA using 200U of SuperScript II reverse transcriptase (Invitrogen) as previously described (Redon et al. 2005). Quantification of cDNA was carried out by Real Time PCR Detection system (Bio-Rad) in 96 well-plate as previously described (Maligoy et al. 2008). For large number of samples, a high-throughput qPCR technique was applied using Biomark™ HD System (Fluidigm Corporation, CA, USA) as previously described (Nouaille et al. 2017). To take into account the variability between samples and experiments of the reverse transcription and the qPCR steps, control Ambion™ ERCC RNA Spike-In mix was used as external normalizers (Nguyen et al. 2019). A total of 8 different genes and 4 ERCCs were quantified in this work (Table S1). Quantification of *lacZ* mRNA was the average value obtained from five primer pairs. Primers for qPCR were designed using Vector NTI advance v11 (Life Technologies) with melting temperature of 59-61°C, length of 18-20 bp and 50-70% GC content (Table S1). Amplicon sizes ranged between 80 and

150 bp. The reaction efficiency was tested on cDNA serial dilutions and focused around 100%. To calculate in each fraction the relative amount of a target mRNA, two normalizations were applied (Nguyen et al. 2019). First, relative mRNA abundance compare to constant quantity of ERCC was calculated by the method of fold-change ΔCt values (Pfaffl 2001). As only 5 µg of the total RNA amount extracted in each fraction was used in the RT-qPCR experiment, we normalized the relative mRNA abundance by the total RNA quantity extracted in each fraction to obtain the relative initial mRNA abundance in each fraction.

Ribosome occupancy and ribosome density calculations

For each gene *i*, ribosome occupancy (RO) is the proportion of its mRNA copies undergoing translation. It is calculated as the sum of proportions of fractions containing RNAs bound with at least one ribosome (fractions B to G):

$$Ribosome\ occupancy_{\ gene\ i,k} = \sum_{j=B}^{G} mRNA\ proportion\ in\ fraction\ j$$

Final ribosome occupancy for each gene was calculated as the mean of RO of two replicates.

For each gene, the peak fraction corresponds to the highest mRNA proportion within fractions B to G containing mRNA undergoing translation. The peak fraction was determined by a bootstrap method on residuals with a confidence interval fixed at 95% (Picard et al. 2012). When the 95% bootstrap confidence interval was not confined to a single fraction, the definition of the peak fraction was widened from only one to two (or more) adjacent fractions and a search for the maximum was initiated again. Ribosome density per 100 nucleotides was the number of bound ribosomes in the peak fraction normalized to the ORF length.

$$Ribosome\ density_{\ gene\ i} = \frac{number\ of\ ribosomes\ of\ peak\ fraction}{Length\ of\ Open\ Reading\ Frame\ gene\ i} \text{ x } 100$$

Ribosome density and ribosome occupancy values are available in Table S2.

Linear covariance model

A multiple linear regression model with various quantitative and qualitative variables was used to identify the major determinant of RO and RD levels (Dressaire et al. 2010; Picard et al. 2012, 2013). Sets of 1100 and 745 genes were respectively included in the models for RO and RD.

$$RO_{(i)} = \alpha + \beta_{[mRNA]} \log([mRNA]_{(i)}) + \beta_{\text{ORF length}} \log(ORF\ length_{(i)}) + \beta_{\text{ORF GC\%}} ORF\ GC\%_{(i)} + \beta_{\text{CAI}} \log(CAI_{(i)}) + \beta_{\text{Chrom.location}} Chrom.location_{(i)} + \beta_{\text{Hydrophobicity}} Hydrophobicity_{(i)} + \lambda_{\text{Strand}} + \lambda_{\text{Essentiality}} + \lambda_{\text{Presence signal peptide}} + \lambda_{\text{Inner mbr protein}} + \lambda_{\text{Cell location}} + \lambda_{\text{COG}} + \lambda_{\text{Target of CsrA}} + \xi_{(i)}$$

$$\log(RD_{(i)}) = \alpha + \beta_{[mRNA]} \log([mRNA]_{(i)}) + \beta_{\text{ORF GC\%}} ORF\ GC\%_{(i)} + \beta_{\text{CAI}} \log(CAI_{(i)}) + \beta_{\text{Chrom.location}} Chrom.location_{(i)} + \beta_{\text{Hydrophobicity}} Hydrophobicity_{(i)} + \lambda_{\text{Strand}} + \lambda_{\text{Essentiality}} + \lambda_{\text{Presence signal peptide}} + \lambda_{\text{Inner mbr protein}} + \lambda_{\text{Cell location}} + \lambda_{\text{COG}} + \lambda_{\text{Target of CsrA}} + \xi_{(i)}$$

Where $RO_{(i)}$ and $RD_{(i)}$ are the measured levels of the *i*th value for the variable of interest, α is the intercept, β et λ are the coefficients associated to each quantitative and qualitative parameter, $\xi_{(i)}$ the error term of the *i*th value.

Features related to gene (length, GC%, strand, chromosome position (distance from OriC), codon adaptation index (CAI)) were provided or calculated from GenBank annotation version 3. Other variables such as gene essentiality and target of CsrA were taken from Baba et al. (Baba et al. 2006) and Esquerré et al. (Esquerré et al. 2016) respectively. Hydrophobicity (GRAVY score) was calculated as the average of hydrophobicity (sum of gravy score of all amino acids divided by the number of amino acids (Kyte and Doolittle 1982). Parameters related to protein features clusters of orthologous groups of protein (COG) annotations and the cellular distributions of proteins obtained from the *E. coli* K-12 annotation (Riley 2006). Inner membrane proteins and proteins with a signal peptide were predicted according to Moffitt et al. (Moffitt et al. 2016) and SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP/) respectively. The genome-wide mRNA concentrations measured in exponential phase in the same standardized growth conditions (cultures in bioreactor in M9 glucose medium at 37°C and pH 7) were included (Esquerré et al. 2015).

Positive quantitative variables were log-transformed when necessary to obtain a normal distribution. Then, all quantitative variables were centered and reduced. To select the most significant parameters, we used the Akaike Information Criteria

(Dressaire et al. 2010). Since RD is by definition inversely proportional to the ORF length, the ORF length was removed from the RD model. For each variable, its influence and significance on RO and RD levels were quantified by the estimated coefficient and p-value. The influence of quantitative parameter was estimated by ranking the absolute values of the associated coefficients. For a qualitative parameter, the coefficients of one sub-level was only compared to the overall mean value of all sub-levels of the same parameter.

When 5'UTR related features were included in the models, the set of genes was restricted to the monocistronic genes which exhibited an identified TSS according to the criteria described above (873 and 607 genes were included for RO and RD, respectively). The -30 +24 nt sequence region of these genes was extracted and the folding energy of this region was calculated using Mfold software. The second nucleotide (purine/pyrimidine) of the 5'UTR was also identified for these genes.

Research of conserved motifs

Nucleotide sequences near the start codon (from -30 to +24 bp, -50 to +24 bp and -100 to +24 bp relatives to ATG) were extracted for the monocistronic genes with an identified TSS according to the TSS criteria described above. The sequences were used to find significant specific patterns of 4 to 8 nucleotide length using oligo-analysis of RSA-Tools (http://www.rsat.eu/).

Simple linear correlation

Simple linear correlations between two variables were estimated by calculating Pearson correlation coefficient and corresponding p-value. Significance threshold was fixed at p-value < 0.01.

Enrichment methods.

R free statistical software (www.r-project.org) was used for enrichment methods. Functional categories enriched in transcript subgroups were determined by a 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 ≤ 0.05 were considered as significant.

RESULTS

Study of translation efficiency using RO and RD parameters

To characterize the translational status of *E. coli* cells at the genome-wide scale, we used the polysome profiling method. mRNA-polysome complexes were fractionated in 7 fractions (from A to G) in function of ribosome load. In each fraction RNA quantification of all transcripts was performed by RNA-Seq. For each gene, we obtained a distribution of mRNA copy proportion in each fraction (Fig. S1). From this distribution, two experimental variables of translation were calculated, the ribosome occupancy and ribosome density. RO is the proportion of the mRNA copies undergoing translation and RD is the ribosome density averaged on length of coding sequence (ribosomes per 100 nucleotides).

We investigated the distribution of RO ad RD values over all the monocistronic genes (1654 genes from RegulonDB database). Polycistronic genes were excluded of the analysis since it was not possible to assign the ribosome load to one or the other cistron. Among the 1654 genes, 91 with too low read counts were excluded from our analysis. The distribution of RO of the remaining 1563 genes is shown in Figure 1.

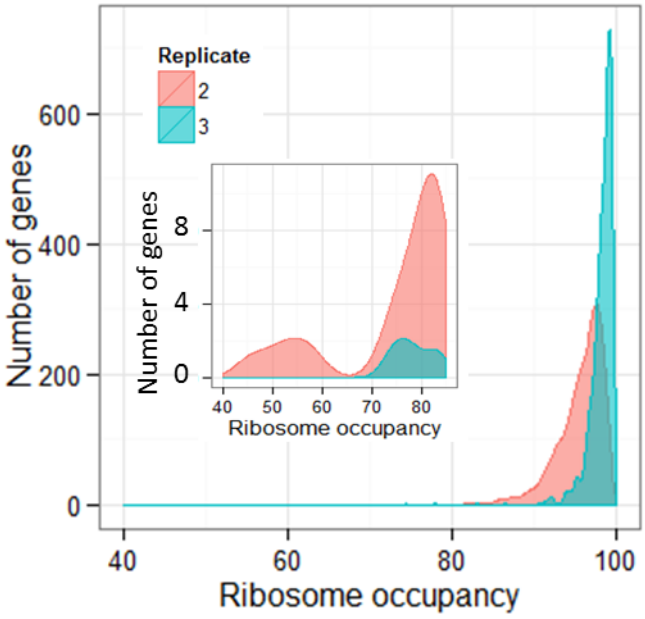


Figure 1: **Distribution of ribosome occupancy for 1563 genes for two replicates. The insert is a zoom on the low RO values between 40 and 80%.**

Our results showed non-gaussian distributions of RO with an average lowest value of 60.0 ± 20%. The average mean and median values were 97.0 ± 2% and 98.0 ±

2%, respectively. These values showed that for the majority of genes nearly all the mRNA copies underwent translation in fast growing *E. coli* cells. The RO values in *E. coli* are particularly high when compared to the ones of other microorganisms (mean RO values of 66, 71 and 77% in *Lactococcus lactis*, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, respectively (Arava et al. 2003; Lackner et al. 2007; Picard et al. 2012)).

Two thirds of the genes (1077/1563 genes) had only one identified peak fraction (Figure 2A). For the large majority (1012 genes) the peak fraction was in fraction C corresponding to a relative low number of bound ribosomes (in average 2.8 ± 0.3 ribosomes). Only a small set of genes had a peak fraction with very low or very high number of ribosomes (23 and 36 genes with 1.1 ± 0.1 and 11.4 ± 0.4 in fractions B and G, respectively). Although a homogeneity between genes was observed for the peak fraction, a very variable ribosome density was obtained when normalized by the ORF length (Figure 2B). The RD span from 0 to a maximum of 2.0 ribosomes/100 nts. Genes generally exhibited relatively low RD with a median of 0.4 and a mean at 0.5 ribosomes/100 nts, far lower than the theoretical maximum density of 3.3 ribosomes/100 nts (under the assumption that a ribosome occupies around 30 nucleotides). These values are in the same range of those in yeast (*S. cerevisae* mean 0.64 ribosomes/100 nts (Arava et al. 2003), *S. pombe* 0.45 ribosomes/100 nts (Lackner et al. 2007)) and much lower than in *L. lactis* (average 1.3 ribosomes/100 nts (Picard et al. 2012)). Only five genes presented aberrant RD (>3.3 ribosomes/100 nts). It was *icdC* a pseudogene, *ymgF* an inner membrane protein, *ybgE* with unknown function), *yoaI* (putative PhoB regulon) and *cydX* subunit of cytochrome ubiquinol oxidase and they were removed from further analysis. We checked that the low RD values were not due to a limitation in free ribosomes since the total ribosomal content not undergoing translation was estimated at 44 ± 4%. This value was estimated by area integration of the absorbance of the polysome profile (Picard et al. 2012). For one third of the genes, the peak fraction was assigned to two fractions, adjacent or not. A close look of the read mapping on these genes did not identify unusual density distribution (for example differential local densities provided by two types of transcript) that could explain such as a behavior. Genes with bimodal profiles were not further analyzed.

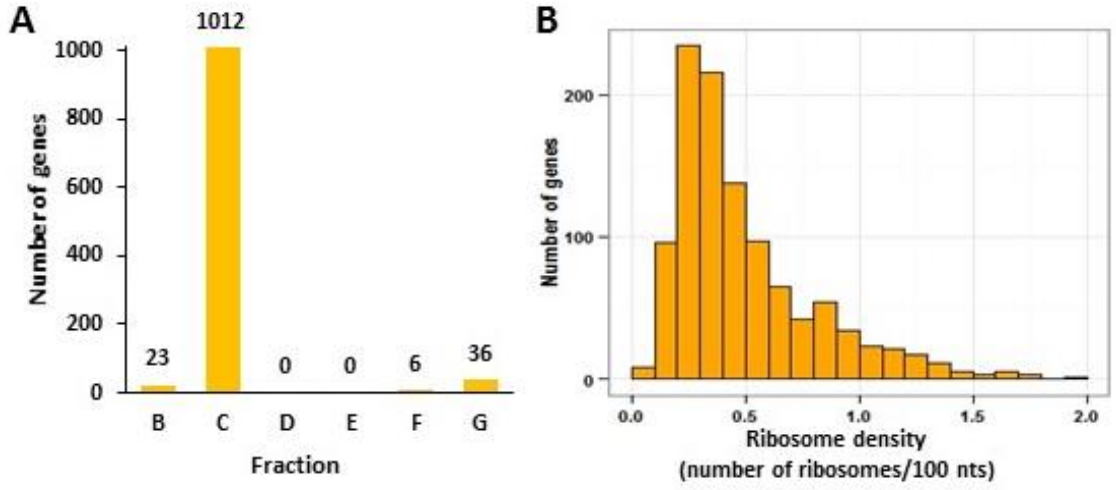


Figure 2: **Distribution of (A) the peak fraction calculated from 2 replicates and (B) the ribosome density for 1077 genes**

RO and RD as translation efficiency estimators

We would like to identify how RO and RD are related to translation efficiency and protein level. Since RO specifies the proportion of mRNA copies involved in translation, a high RO could indicate a high initiation rate and thus could lead to a high protein level. The relationship between the ribosome density RD and translation efficiency is more complex. Indeed, although increasing the ribosome number might favor protein synthesis, a too high ribosome density could reflect ribosome pausing and traffic jam and be not optimal for protein synthesis (Mitarai, Sneppen, and Pedersen 2008; Zarai, Margaliot, and Tuller 2016). To elucidate the relationship between RO, RD and translation efficiency, we investigated at the genome-wide scale the correlations between RO and RD and several translational variables: proteomic data that we previously measured in our lab in fast growing *E. coli* cells (Esquerré et al. 2015), translation efficiency measured by ribosome profiling (Burkhardt et al. 2017; Li et al. 2014) and estimated translation rates.

First, the proteomics dataset is previously generated in our group in the same physiological condition as this study (Esquerré et al. 2015). Both RO and RD were significantly positively correlated with protein levels (Table 1). This confirmed that as expected higher ribosome occupancy and ribosome density were associated with higher protein level.

Variables	Estimated coefficient	P-Value
Ribosome occupancy	0.24	5.1×10^{-5}
Ribosome density	0.17	2.7×10^{-3}

Table 1: Linear covariance model to explain protein level (set of 262 proteins)

When we compared our RO and RD values with the translation efficiency (TE) reported in the literature by the ribosome profiling method (Burkhardt et al. 2017; Li et al. 2014), positive correlations were also obtained (between RO and TE: coeff=0.35, p-value = 1.4×10^{-27} on 934 genes and between RD and TE: coeff=0.36, p-value=4.4×10^{-21} on 660 genes). When we searched for a correlation with the translational efficiency estimated using the number of protein molecules produced per mRNA molecule (ratio of our data of protein and mRNA levels), positive correlations of the ratio with RO (coeff=0.16, p-value =2×10^{-3} on 382 genes) and RD (coeff=0.34, p-value =4×10^{-8} on 262 genes) were obtained. One explanation for these moderate correlations and lower significances could be that the protein level/mRNA level ratio takes into account protein degradation in addition to the translational control.

To verify the assumption that RO reflects translation initiation, we studied the correlations between RO with the RBS score, a widely used metric of translation initiation rate (RBS calculator (Borujeni et al. 2014)). First, genes in the bottom 5% of RO had significantly lower RBS scores than those in the top 5% (Kolmogorov-Smirnov test: p-value of 2.4×10^{-8}) (Fig. 3A). This positive relationship between RO and the RBS score was still valid at a larger-scale (848 genes) with a positive correlation (coeff=0.11, p-value =2.6×10^{-3}).

RD depends on both translation initiation and elongation (Shaham and Tuller 2018). So we checked the correlation of RD with the translation rate estimated with the Transim software which takes into account both the initiation and elongation processes (Shaham and Tuller 2018). The estimated translation rate of genes in bottom 5% of RD was significantly lower than those in the top 5% (Kolmogorov-Smirnov test p-value =1.7×10^{-5}) (Fig. 3B). Large-scale analysis (848 genes) confirmed this result with a positive correlation between RD and the translation rate (coeff=0.10, p-value =8.8×10^{-3}).

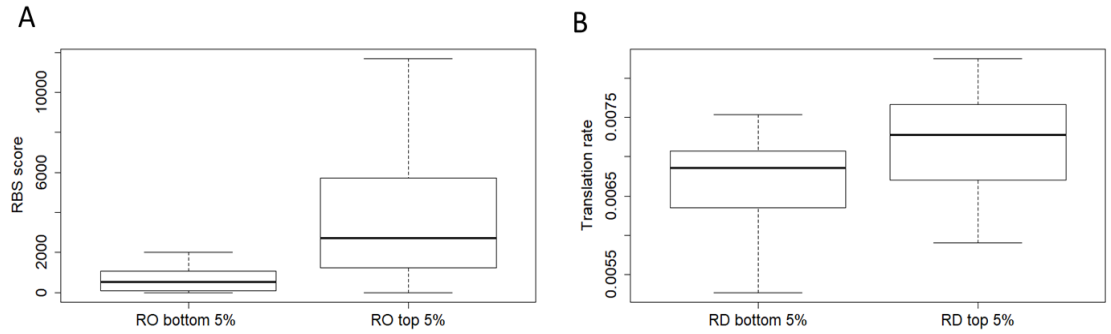


Figure 3: Box plots of RO and RD. (A) Box plots of the bottom 5% and top 5% of RO and the RBS score (62 genes) and (B) Box plots of the bottom 5% and top 5% of RD an d the estimated translation rate (44 genes).

Altogether, these results demonstrated positive correlations between RO and RD and different translation features such as efficiency, initiation rate and translation rate. High ribosome occupancy and density lead to high translation efficiency and rate, resulting in higher protein synthesis. Therefore, in *E. coli* cells, we demonstrated that RO and RD can therefore be considered as estimators of translation. This result is in agreement with previous works in other microorganisms: both RO and RD were positively correlated with protein level in *S. pombe* (Lackner et al. 2007) and RO in *L. lactis* (Picard et al. 2012).

Factors regulating RO and RD

Noting the diversity of RO and RD between all mRNAs, we searched for factors explaining this variability to understand how ribosome occupancy and ribosome density were regulated.

First we checked if RO and RD depended on gene function. We studied if functional categories were overrepresented in genes having extreme values of RO and RD (bottom 10% and top 10%). In the top of RO, there was an enrichment of genes involved in amino acid metabolism (GO terms related to biosynthetic pathways of arginine, lysine and aspartate family amino acids were enriched) and production of energy (more precisely the two GO terms glucose catabolic process and tricarboxylic acid cycle). Genes related to the functional category named signal transduction mechanism (for example the SOS response) were overrepresented in the bottom 10 % of both RO and RD. No striking functional category was significantly enriched in the

genes part of the RD top 10%. These results suggest that in fast growing cells, genes related to important functions for growth tend to have high RO while genes having less necessary roles in rapid growth have low values of RO and RD.

Then, we searched for sequence motifs that may contribute to low and high values of RO and RD. We found the conserved motif ACAGGAG in mRNAs of the top 10% of RO (Fig. 4A). This motif was localized in the ribosome binding region in the 5'UTR. This motif included part of the Shine-Dalgarno sequence but it was not the exact complement of the 3'-terminus of the *E. coli* 16S rRNA (which is 5' TAA**GGAGG**TGTAT 3' with in bold the Shine-Dalgarno sequence). For the top 10% of RD, we found the motif of 4 adenines following the ATG start site (ATGAAAA) (Fig. 4B). Having as a 2nd codon AAA and AAU was reported to enhance the translation efficiency in *E. coli* (Sato et al. 2001). Conserved motifs were only found in genes in the top of RO and RD and not for genes in the bottom suggesting that translational regulation in exponential growth consists mainly in positive control.

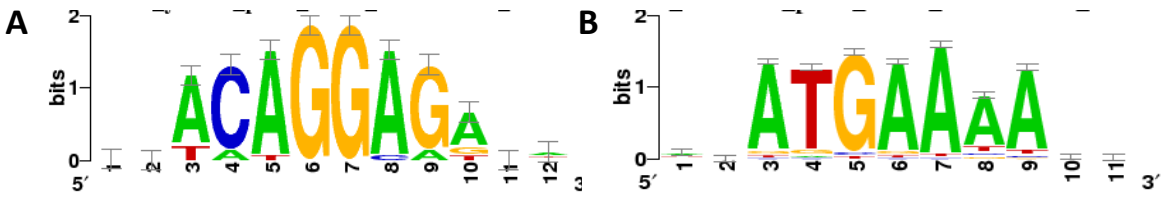


Figure 4: Conserved motifs in mRNAs of (A) the top 10% of RO (90 mRNAs) and (B) the top 10% of RD (62 mRNAs) using RSA-Tools

After the analyses of extreme values, we identified systemic key factors regulating the level of RO and RD using linear covariance models. Estimated coefficients of variables selected by Akaike information criterion and with significance p-value < 0.05 are shown in Table 2. This approach allowed to quantify and classify without any *a priori* simultaneous contributions to RO and RD of several parameters related to features of gene and protein and to cell physiology (see Material and Methods section for the parameter list).

We obtained relatively low R^2 for the two models (0.26 for RO and 0.16 for RD). So a non-negligible variation of RO and RD was explained by our model, but other unidentified parameters (i.e. RNA-binding proteins and regulatory small RNAs) could also contribute to the complex and multiple-layer regulation of translation.

Parameter	Variable to explain: Ribosome occupancy		Variable to explain: Ribosome density	
	Est. coeff.	p-value	Est. coeff.	p-value
mRNA concentration	0.42	1.9×10^{-48}	0.09	2.5×10^{-2}
ORF GC%	0.11	5.3×10^{-5}	-0.20	9.7×10^{-8}
Hydrophobicity	0.16	5.0×10^{-5}	0.17	1.1×10^{-3}
Distance from OriC	0.06	1.6×10^{-2}	/	/
Being a target of CsrA	0.13	1.9×10^{-3}	0.27	3.7×10^{-6}
Being an Inner membrane protein	/	/	-0.26	3.1×10^{-5}
Cell location: Related to membrane	-0.21	9.9×10^{-3}	/	/
Presence of a signal peptide	0.24	5.6×10^{-4}	/	/
COG: Cell motility	/	/	0.81	5.0×10^{-3}
COG: Amino acid transport and metabolism	/	/	-0.35	9.9×10^{-3}
COG: Function unknown	/	/	0.60	6.3×10^{-9}
COG: Posttranslational modification	/	/	0.32	3.6×10^{-2}

Table 2: Estimated coefficients and p-values of ribosome occupancy and ribosome density models.

Some factors influenced both RO and RD. The first one is the mRNA level with positive effects on both RO and RD. It is the most significant factor for RO, with the highest coefficient and significance (coeff= 0.45 and p-value < $2e^{-16}$). Genes with high mRNA concentration had significantly higher value of RO and slightly higher value of RD. The second factor was the GC% of ORF, which was the most important factor of RD. A negative coefficient for RD in contrast to a positive coefficient for RO indicates that genes with higher GC content tend to have lower ribosome density but a higher proportion of mRNA copies undergoing translation. We also found that genes targeted by the RNA binding protein CsrA had higher RO and RD than the others. Additional experiments would be required to better understand the suggested positive effect of CsrA binding on translation. Additionally, protein localization seemed to be a factor of

translation regulation since being a membrane protein, an inner membrane protein and having a signal peptide were significant parameters to explain RO and RD values. However, contrasted results were obtained, with for example a positive effect of the presence of a peptide signal but a negative effect of the protein location at the membrane on RO. Genes involved in amino acid transport and metabolism were found with a negative coefficient of RD (meaning a lower RD value compared to the average RD of all the COG categories). When added to the models, the effects of 5'UTR related parameters such as length, GC%, folding energy of -30 +24 bp region and nature of 2nd nucleotide were not significant (p-value>0.05).

These results showed the diversity of parameters involved in translation regulation. We identified factors related to gene sequence (the conserved motif, GC% of the ORF), to protein features (the functional category and localization) and more surprisingly related as well to cell physiology and growth condition (mRNA concentration). Indeed, sequence-related parameters (5'UTR secondary structure, CAI…) are well-studied translational factors that affect translation initiation and elongation whereas the effect of mRNA concentration to explain translation variation between genes in a same condition is not commonly studied. The transcriptional response is usually taken into account in the translational response in conditions of stress but hardly to study the mechanism of the direct effect of transcription on translation. When taken in account, the effect of mRNA abundance on translation is not clear: no effect (Arava et al. 2003), positive effect (Lackner et al. 2007) or negative effect (Picard et al. 2012) on translatome variables were reported.

Validation of the translational response to changing mRNA concentration

Since translation regulation by the physiological factor mRNA concentration was unexpected, we decided to validate that mRNA concentration is an important factor to determine RO and has a positive effect on RD. Our strategy was to artificially increase mRNA synthesis of selected genes by transcriptional induction and observe the impact on polysome profiles. The first candidate was the native *lacZ* gene because chromosomal induction of *lacZ* transcription was easily achieved by addition of IPTG. Two *lacZ* mRNA concentration levels were studied, the basal low concentration without induction and a 200-fold increased concentration level after IPTG induction. Figure 3 shows the distribution of *lacZ* mRNA copies in fractions A to G in the two conditions. At low concentration, most of the *lacZ* mRNA copies were ribosome-free

and not undergoing translation (fraction A) corresponding to a low RO (34 ± 19%). At high concentration, we observed at the translational level a dramatic decrease of *lacZ* mRNA copies in fraction A, resulting in an increase of RO to 99 ± 0.6%. This was combined with an important shift of the mRNA copies to fractions with high number of ribosomes (E, F and G). These results showed that increasing *lacZ* mRNA concentration resulted in nearly all the mRNA copies occupied by ribosomes and bound with more ribosomes. These results confirmed our observation at –omics scale of a positive correlation between the mRNA concentration and RO and the ribosome load.

Next, we investigated the translation response of more genes to changing mRNA concentration by cloning genes on a plasmid under an inducible promoter. In addition to *lacZ*, we selected five monocistronic *E. coli* genes, *cysZ*, *inaA*, *ucpA*, *yeeZ* and *yjcO* which exhibited low mRNA concentration during exponential growth in M9 glucose (Esquerré et al BMC Genomics). The six genes were cloned in a plasmid under the arabinose inducible promoter. In the presence of arabinose, between 18-83 fold higher mRNA concentrations were obtained. Since for the *cysZ*, *inaA*, *ucpA*, *yeeZ* and *yjcO* recombinant strains, the chromosomal copy of the selected genes was still present, we checked that the chromosomal expression was negligible (it was between 20-800 fold lower than the induced plasmidic expression). We have previously developed a multiplexing polysome profiling method (Nguyen et al. 2019) for parallel characterization of translational responses of multiple genes to changing mRNA concentrations. Using two multiplexing polysome profiling experiments, one without arabinose (low mRNA concentrations) and one with arabinose (high mRNA concentrations), we assessed the translational responses of six genes (*lacZ*, *cysZ*, *inaA*, *ucpA*, *yeeZ* and *yjcO*) to changing mRNA concentrations (Fig. 6).

First we compared using the *lacZ* gene that the translational response was similar using a chromosomal induction with IPTG and a plasmidic induction with arabinose (Fig. 5B). Plasmidic induction led to a sharp increase of RO (from 41% without induction to 91% after arabinose induction) and a notable shift of the mRNA copies towards the more ribosome heavily-loaded fractions (E, F and G). Therefore, very similar translational responses were obtained for the chromosomal and plasmidic inductions. This result validated the use of plasmidic induction to study the translational response of the five other genes to changing mRNA concentration.

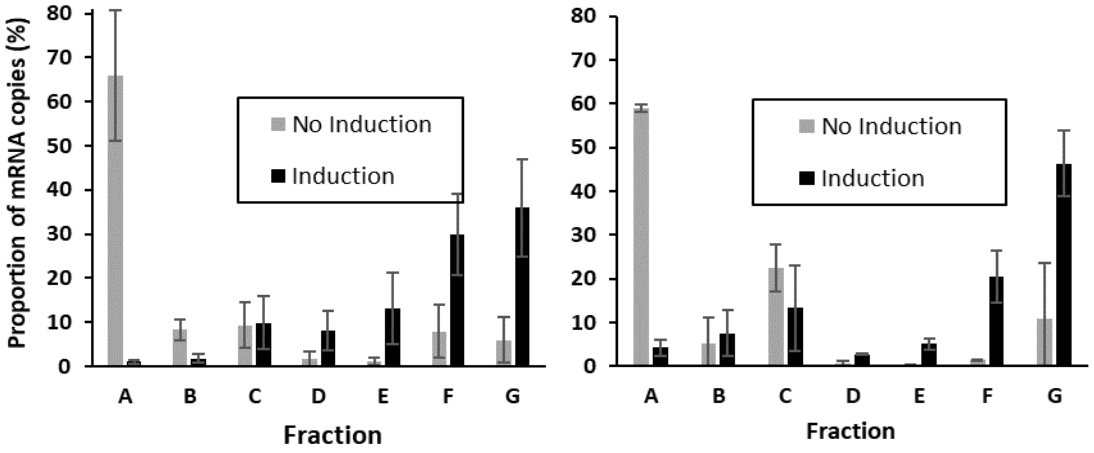


Figure 5: Distribution of *lacZ* mRNA copies between fractions in two conditions of "without induction" (grey bars) and "with induction" (black bars). Experiments were performed with strains containing *lacZ* (A) on the chromosome under the native promoter P_{Lac} and (B) on plasmid under the promoter P_{BAD}. Fraction A consists in free mRNA copies not undergoing translation while fractions B to G contain mRNA copies bound with ascending numbers of ribosomes from 1 to 11. Each culture and the polysome profiling experiment were repeated two times to provide independent biological replicates.

Like *lacZ*, for the five genes *cysZ, inaA, ucpA, yeeZ and yjcO,* when the mRNA expression was induced we observed a substantial decrease of free mRNA copies (fraction A). Therefore ribosome occupancy strongly increased at high mRNA level (61, 95, 44, 37 and 69% increased RO values for *cysZ, inaA, ucpA, yeeZ* and *yjcO,* respectively). mRNA copies shifted from fraction A to fractions corresponding to higher loads in ribosomes. Two shift patterns were observed: (i) mRNA copies spread all over the ribosome-loaded fractions, all fractions from B to G having higher proportions (for *cysZ, ucpA* and *yjcO*), and (ii) mRNA copies shifted preferentially towards the highest ribosome loaded fractions (E, F and G like for *lacZ, inaA* and *yeeZ*). We have excluded that the observed higher load in ribosomes resulted of an increased number of ribosomes in the condition of arabinose induction; in the six strains, the concentrations of the ribosomal protein mRNAs (RpsJ and RplK) were similar without and with arabinose (Table S3). Similarly, an effect of a decrease in total RNAs in the condition of arabinose induction was also excluded since the total RNA contents were similar in the six strains without and with arabinose (Table S3). Altogether, these results demonstrated a similar effect of mRNA concentration on the translational

status of the six genes with a higher load in ribosomes at higher mRNA level even if differences in the pattern of the response can exist between the genes.

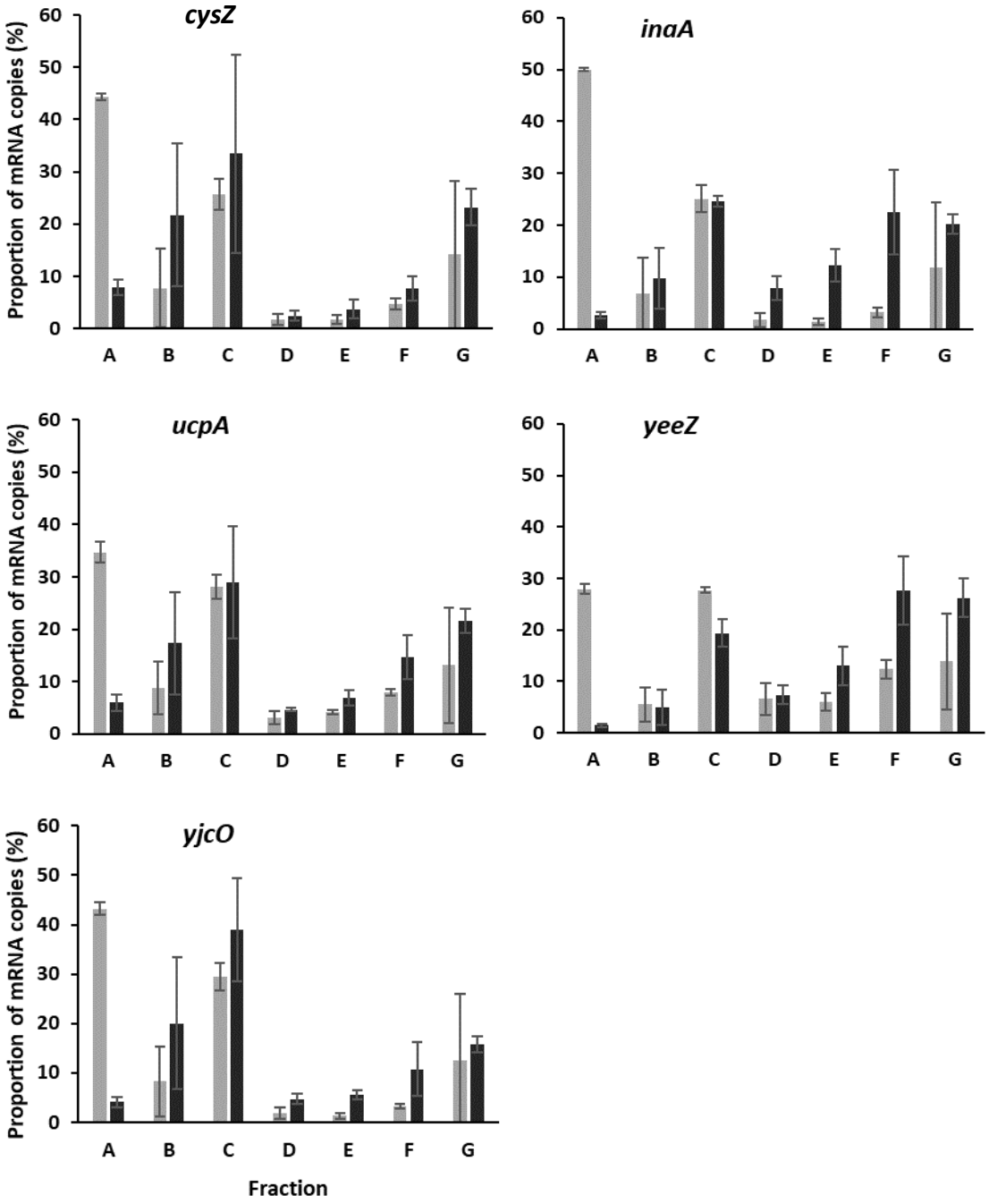


Figure 6: Distribution of mRNA copies between fractions in the two conditions of "without induction" grey bars) and "with induction" (black bars) for the five genes. Fraction A consists in free mRNA copies not undergoing translation while fractions B to G contain mRNA copies bound with ascending numbers of ribosomes from 1 to 11. Each culture and the polysome profiling experiment were repeated two times to provide independent biological replicates.

DISCUSSION

In this work, we provided a genome-wide study of the translation status in *E. coli* in exponential growth. Translation was described through two parameters, ribosome occupancy and ribosome density. Analysis of RO and RD over the genome gives an overall picture of translation activity in fast growing *E. coli* cells. First, for most genes, nearly all the mRNA copies were undergoing translation. The RO values were correlated with the estimated rate of translation initiation and the protein levels. These results indicate a high efficiency of translation initiation in response to the high demand in protein expression in our conditions when *E. coli* cells need to grow rapidly. This is particularly the case for proteins having extreme values of RO (top 10%) and namely proteins related to the metabolism and energy generation that are crucial for bacterial growth. However most of the mRNAs were loaded with a rather low number of ribosomes. Ribosome density maximizing protein translation rate was modelled and values lower than the maximal theoretical density of 80 % or 50 % were found respectively by (Zouridis and Hatzimanikatis 2007) and (Zarai et al. 2016). Higher ribosome densities were predicted to lead to ribosome collisions and traffic jam that slow down translation. At lower ribosome densities both the initiation and elongation steps were predicted to limit the translation rate. We demonstrated here at the genome-wide scale that RD was positively correlated with the estimated translation rate and protein level. RD reflecting both translation initiation and elongation, we cannot exclude that elongation is not also limiting. However our finding of a rather low ribosome density for most of the genes (in average 15% of the maximum density) excluded a strong limitation of translation by elongation (Plotkin and Kudla 2011; Zouridis and Hatzimanikatis 2007) and reinforce the assumption of initiation limitation.

By exploring the diversity of ribosome occupancy and ribosome density values within the mRNA population, systemic key factors involved in translation regulation were sought. As expected, we identified sequence-related factors influencing translation. First, high translation was found to be associated with conserved motifs respectively in the 5'UTR to lead to high ribosome occupancy and in the beginning of the coding sequence to lead to high ribosome density. Second, we observed that the GC content of the ORF had opposite effects on RO and RD: genes rich in GC tended to have

lower ribosome density but higher ribosome occupancy. Since high GC content sequences are more prone to form stable secondary structure our result on RD is in agreement with the reported negative correlation between the in vivo secondary structure of ORF and the ribosome density (Burkhardt et al. 2017). It should be noted that in our models the GC% of the 5'UTR and the ΔG of the translation initiation region were not selected as translational determinant. mRNA folding in the ribosome binding site was reported to interact with ribosome accessibility as demonstrated in *E. coli* (Kudla 2009; Seo et al. 2009; Tuller et al. 2010) but at this stage we do not have a clear explanation of our positive effect of the ORF GC content on RO. We did not select the CAI as a determinant of RO and RD. This in in agreement with the little impact of CAI recently reported on translation efficiency in *E. coli* (Cambray et al. 2018) and on RO and RD in *L. lactis* (Picard et al. 2012). Our previous finding of a small limitation of translation by the elongation step may be the explanation. We identified also parameters related to protein features such as their localization. We showed that membrane related proteins had lower RO. This is contrasted with the results of Andreeva et al. who proposed that the high initiation rate of the outer membrane lipoprotein Lpp resulted of a mechanism of early recruitment of second ribosome (Andreeva et al. 2018). We also found that genes coding for inner membrane proteins had lower RD. Lower translation efficiency of inner membrane protein was also found by Morgan and coworkers (Morgan et al. 2017) and they proposed the effect on translational efficiency of the cotranslational translocation into the inner membrane via the SRP pathway. But in contrast, some other authors proposed a higher ribosome density of mRNAs of inner membrane proteins. This conclusion was supported by the two following facts: mRNAs encoding inner-membrane proteins are enriched at the membrane (Moffitt et al. 2016) while ribosome rich mRNAs are modelled to preferentially diffuse to the cell poles (Castellana et al. 2016). A role of the translation rate in the spatial organization of proteins in the membrane was also described (Korkmazhan et al. 2017).

Our linear covariance model identified mRNA concentration as a determinant of ribosome occupancy and ribosome load in *E. coli*. Experiments of molecular validation on selected genes demonstrated the causal positive effect of transcriptional induction on translation. This result was rather unexpected according to contrasting results in the literature. Working with the same parameters (RO and RD) used in this

study, opposite small effects of mRNA concentration on RO and RD were described in *L. lactis* (Picard et al. 2012) whereas in yeast, no correlation and positive correlations were reported in *S. cerevisiae* and *S. pombe,* respectively (Lackner et al. 2007). Moreover, a positive correlation between mRNA abundance and the other parameter of translation efficiency (ratio of protein synthesis rate/mRNA level) was shown in *S. cerevisiae* (Pop et al. 2014). Finally, mRNA concentration was described in several studies as a primary determinant of translational status defined as the amount of ribosome footprint. mRNA abundance was positively correlated with the quantity of ribosome footprint in *Mycoplasma gallisepticum* (Fisunov et al. 2017) and *E. coli* (Bartholomäus et al. 2016; Morgan et al. 2017). However, the quantity of ribosome footprint is rather a proxy of protein level than estimators of translation efficiency and the relationship between protein level and mRNA level was already well-known (Esquerré et al. 2015; Guimaraes et al. 2014).

Different mechanisms are possible to explain the positive correlations between mRNA level and RO and RD. First, higher intracellular mRNA concentration will increase the probability of the mRNA copies to encounter ribosomes directly impacting RO and translation initiation. Since initiation was suggested to be the limiting step of translation, an increase in initiation will result also in an increase in RD. Second, the presence of translating ribosomes was proposed to prevent the backtracking of RNA polymerase and hence to enhance transcription and therefore mRNA level (Proshkin et al. 2010). This transcription-translation coupling could also impact the observed positive correlation between mRNA level and RO and RD. Third, translating ribosomes may protect mRNA from degradation (Chan et al. 2018). Low ribosome density will lead to destabilized mRNA and decreased mRNA level while high ribosome density will increase mRNA stability and thus mRNA level. We have previously demonstrated at the genomic and molecular levels that the most concentrated mRNAs were the least stable in *E. coli* (Esquerré et al. 2015; Nouaille et al. 2017). The strong inverse correlation between mRNA concentration and half-life does not support the last hypothesis of the protective effect of ribosomes.

When facing new environmental conditions, the metabolism of *E. coli* cells must adapt by generally adjusting gene expressions. From a physiological point of view, the coordinated variations in mRNA concentration, RO and RD make sense to efficiently control protein expression levels. For example, co-directional transcriptional and

translational regulation has been previously described in yeast and *L. lactis* in stress conditions (Melamed et al. 2008; Picard et al. 2012; Preiss et al. 2003; Serikawa et al. 2003). It would now be interesting to apply our genome-wide analysis of ribosome occupancy and ribosome density to *E. coli* cells in stress condition to support the physiological significance of our finding of a co-transcriptional regulation of translation.

Acknowledgements

We would like to thank Lidwine Trouilh from GeT-Biopuces at the Genopole in Toulouse for the sequencing experiments. We would also like to thank Sabrina Legoueix Rodriguez from Toulouse White Biotechnology for bioinformatical analysis of the RNAseq data.

REFERENCES

Ah-Seng, Y., Rech, J., Lane, D., and Bouet, J.-Y. (2013). Defining the Role of ATP Hydrolysis in Mitotic Segregation of Bacterial Plasmids. PLOS Genetics *9*, e1003956.

Andreeva, I., Belardinelli, R., and Rodnina, M. V (2018). Translation initiation in bacterial polysomes through ribosome loading on a standby site on a highly translated mRNA. Proceedings of the National Academy of Sciences of the United States of America *115*, 4411–4416.

Arava, Y., Wang, Y., Storey, J.D., Liu, C.L., Brown, P.O., and Herschlag, D. (2003). Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A *100*, 3889–3894.

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. Molecular Systems Biology *2*.

Bartholomäus, A., Fedyunin, I., Feist, P., Sin, C., Zhang, G., Valleriani, A., and Ignatova, Z. (2016). Bacteria differently regulate mRNA abundance to specifically respond to various stresses. Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences *374*, 20150069-.

Burkhardt, D.H., Rouskin, S., Zhang, Y., Li, G.-W., Weissman, J.S., and Gross, C.A. (2017). Operon mRNAs are organized into ORF-centric structures that predict translation efficiency. ELife *6*, e22037.

Cambray, G., Guimaraes, J.C., and Arkin, A.P. (2018). Evaluation of 244,000 synthetic sequences reveals design principles to optimize translation in Escherichia coli. Nature Biotechnology.

Castellana, M., Hsin-Jung Li, S., and Wingreen, N.S. (2016). Spatial organization of bacterial transcription and translation. Proceedings of the National Academy of Sciences *113*, 9286–9291.

Chan, L.Y., Mugler, C.F., Heinrich, S., Vallotton, P., and Weis, K. (2018). Non-invasive measurement of mRNA decay reveals translation initiation as the major determinant of mRNA stability. ELife *7*, 32.

Dana, A., and Tuller, T. (2012). Determinants of Translation Elongation Speed and Ribosomal Profiling Biases in Mouse Embryonic Stem Cells. PLoS Comput Biol *8*.

Dressaire, C., Laurent, B., Loubière, P., Besse, P., and Cocaign-Bousquet, M. (2010). Linear covariance models to examine the determinants of protein levels in Lactococcus lactis. Molecular BioSystems *6*, 1255.

Espah Borujeni, A., Channarasappa, A.S., and Salis, H.M. (2014). Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. Nucleic Acids Research *42*, 2646–2659.

Esquerré, 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 Research *42*, 2460–2472.

Esquerré, T., Moisan, A., Chiapello, H., Arike, L., Vilu, R., Gaspin, C., Cocaign-Bousquet, M., and Girbal, L. (2015). Genome-wide investigation of mRNA lifetime determinants in Escherichia coli cells cultured at different growth rates. BMC Genomics *16*, 275.

Esquerré, T., Bouvier, M., Turlan, C., Carpousis, A.J., Girbal, L., and Cocaign-Bousquet, M. (2016). The Csr system regulates genome-wide mRNA stability and transcription and thus gene expression in Escherichia coli. Scientific Reports *6*.

Fisunov, G.Y., Evsyutina, D. V., Garanina, I.A., Arzamasov, A.A., Butenko, I.O., Altukhov, I.A., Nikitina, A.S., and Govorun, V.M. (2017). Ribosome profiling reveals an adaptation strategy of reduced bacterium to acute stress. Biochimie *132*, 66–74.

Gallant (1979). Stringent control in Escherichia coli. Annu. Rev. Genet, 13 (1979), pp. 393-415

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 Research *42*, 4791–4799.

Ingolia (2009). Genome wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *29*, 5095–5101.

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 Genetics *8*, e1002867.

Korkmazhan, E., Teimouri, H., Peterman, N., and Levine, E. (2017). Dynamics of translation can determine the spatial organization of membrane-bound proteins and their mRNA. Proceedings of the National Academy of Sciences *114*, 13424–13429.

Kudla (2009). Coding sequence determinants of gene expression. Science 255–258.

Kyte, J., and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology *157*, 105–132.

Lackner, D.H., Beilharz, T.H., Marguerat, S., Mata, J., Watt, S., Schubert, F., Preiss, T., and Bähler, J. (2007a). A Network of Multiple Regulatory Layers Shapes Gene Expression in Fission Yeast. Molecular Cell *26*, 145–155.

Lackner, D.H., Beilharz, T.H., Marguerat, S., Mata, J., Watt, S., Schubert, F., Preiss, T., and Bähler, J. (2007b). A Network of Multiple Regulatory Layers Shapes Gene Expression in Fission Yeast. Molecular Cell *26*, 145–155.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *25*, 1754–1760.

Li, G.W., Burkhardt, D., Gross, C., and Weissman, J.S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell *157*, 624–635.

Liu, B., Han, Y., and Qian, S.-B. (2013). Co-Translational Response to Proteotoxic Stress by Elongation Pausing of Ribosomes. Mol Cell *49*, 453–463.

Maligoy, M., Mercade, M., Cocaign-Bousquet, M., and Loubiere, P. (2008). Transcriptome analysis of Lactococcus lactis in coculture with Saccharomyces cerevisiae. Applied and Environmental Microbiology *74*, 485–494.

Melamed, D., Pnueli, L., and Arava, Y. (2008). Yeast translational response to high salinity: Global analysis reveals regulation at multiple levels. RNA *14*, 1337–1351.

Mitarai, N., Sneppen, K., and Pedersen, S. (2008). Ribosome Collisions and Translation Efficiency: Optimization by Codon Usage and mRNA Destabilization. Journal of Molecular Biology *382*, 236–245.

Moffitt, J.R., Pandey, S., Boettiger, A.N., Wang, S., and Zhuang, X. (2016). Spatial organization shapes the turnover of a bacterial transcriptome. ELife *5*.

Morgan, G.J., Burkhardt, D.H., Kelly, J.W., and Powers, E.T. (2017). Translation efficiency is maintained at elevated temperature in *E. coli*. Journal of Biological Chemistry *293*, jbc.RA117.000284.

Nakahigashi, K., Takai, Y., Shiwa, Y., Wada, M., Honma, M., Yoshikawa, H., Tomita, M., Kanai, A., and Mori, H. (2014). Effect of codon adaptation on codon-level and gene-level translation efficiency in vivo. BMC Genomics *15*, 1115.

Nguyen, H.L., Duviau, M.-P., Cocaign-Bousquet, M., Nouaille, S., and Girbal, L. (2019). Multiplexing polysome profiling experiments to study translation in Escherichia coli. PLOS ONE *14*, e0212297.

Nouaille, S., Mondeil, S., Finoux, A.L., Moulis, C., Girbal, L., and Cocaign-Bousquet, M. (2017). The stability of an mRNA is influenced by its concentration: a potential physical mechanism to regulate gene expression. Nucleic Acids Research *45*, 11711–11724.

Oh, E., Becker, A.H., Sandikci, A., Huber, D., Chaba, R., Gloge, F., Nichols, R.J., Typas, A., Gross, C.A., Kramer, G., et al. (2011). Selective Ribosome Profiling Reveals the Cotranslational Chaperone Action of Trigger Factor In Vivo. Cell *147*, 1295–1308.

Osterman, I.A., Evfratov, S.A., Sergiev, P. V., and Dontsova, O.A. (2013). Comparison of mRNA features affecting translation initiation and reinitiation. Nucleic Acids Research *41*, 474–486.

Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Research *29*, e45.

Picard, F., Milhem, H., Loubière, P., Laurent, B., Cocaign-Bousquet, M., and Girbal, L. (2012). Bacterial translational regulations: high diversity between all mRNAs and major role in gene expression. BMC Genomics *13*, 528.

Picard, F., Loubière, P., Girbal, L., and Cocaign-Bousquet, M. (2013). The significance of translation regulation in the stress response. BMC Genomics *14*, 588.

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.

Pop, C., Rouskin, S., Ingolia, N.T., Han, L., Phizicky, E.M., Weissman, J.S., and Koller, D. (2014). Causal signals between codon bias, mRNA structure, and the efficiency of translation and elongation. Molecular Systems Biology *10*, 770.

Preiss, T., Baron-Benhamou, J., Ansorge, W., and Hentze, M.W. (2003). Homodirectional changes in transcriptome composition and mRNA translation induced by rapamycin and heat shock. Nature Structural & Molecular Biology *10*, 1039–1047.

Proshkin, S., Rahmouni, A.R., Mironov, A., and Nudler, E. (2010). Cooperation between translating ribosomes and RNA polymerase in transcription elongation. Science (New York, N.Y.) *328*, 504–508.

Redon, E., Loubière, P., and Cocaign-Bousquet, M. (2005). Role of mRNA stability during genome-wide adaptation of Lactococcus lactis to carbon starvation. The Journal of Biological Chemistry *280*, 36380–36385.

Reuveni, S., Meilijson, I., Kupiec, M., Ruppin, E., and Tuller, T. (2011). Genome-Scale Analysis of Translation Elongation with a Ribosome Flow Model. PLoS Computational Biology *7*, e1002127.

Riley, M. (2006). Escherichia coli K-12: a cooperatively developed annotation snapshot--2005. Nucleic Acids Research *34*, 1–9.

Russell, J.B., and Cook, G.M. (1995). Energetics of Bacterial Growth: Balance of Anabolic and Catabolic Reactions.

Sato, T., Terabe, M., Watanabe, H., Gojobori, T., Hori-Takemoto, C., and Miura, K. (2001). The Journal of biochemistry. (Japanese Biochemical Society).

Seo, S.W., Yang, J., and Jung, G.Y. (2009). Quantitative correlation between mRNA secondary structure around the region downstream of the initiation codon and translational efficiency in Escherichia coli. Biotechnology and Bioengineering *104*, 611–616.

Serikawa, K.A., Xu, X.L., MacKay, V.L., Law, G.L., Zong, Q., Zhao, L.P., Bumgarner, R., and Morris, D.R. (2003). The Transcriptome and Its Translation during Recovery from Cell Cycle Arrest in *Saccharomyces cerevisiae*. Molecular & Cellular Proteomics *2*, 191–204.

Shaham, G., and Tuller, T. (2018). Genome scale analysis of Escherichia coli with a comprehensive prokaryotic sequence-based biophysical model of translation initiation and elongation. DNA Research *25*, 195–205.

Thomason, M.K., Bischler, T., Eisenbart, S.K., Förstner, K.U., Zhang, A., Herbig, A., Nieselt, K., Sharma, C.M., and Storz, G. (2015). Global Transcriptional Start Site Mapping Using Differential RNA Sequencing Reveals Novel Antisense RNAs in Escherichia coli. Journal of Bacteriology *197*, 18–28.

Tuller, T., Waldman, Y.Y., Kupiec, M., and Ruppin, E. (2010). Translation efficiency is determined by both codon bias and folding energy. Proceedings of the National Academy of Sciences *107*, 3645–3650.

Zarai, Y., Margaliot, M., and Tuller, T. (2016). On the Ribosomal Density that Maximizes Protein Translation Rate. PLOS ONE *11*, e0166481.

Zouridis, H., and Hatzimanikatis, V. (2007). A Model for Protein Translation: Polysome Self-Organization Leads to Maximum Protein Synthesis Rates. Biophysical Journal *92*, 717–730.

Supporting information

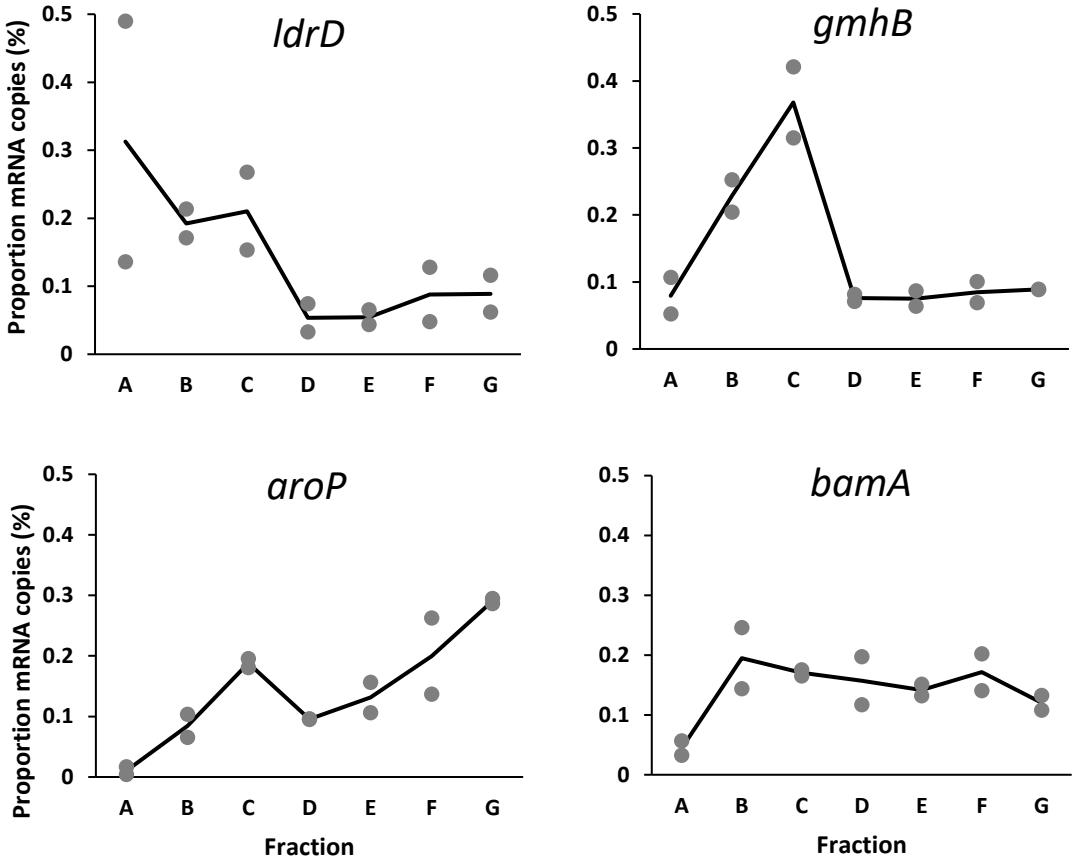


Figure S1: **Examples of mRNA proportions between fractions A to G**. Dots represent data from the two independent experiences. Black line is the average value of two experiments.

Table S1: Sequence of qPCR primers used to quantify 8 endogenous genes of *E. coli* MG1655 and four ERCC RNA spike-ins.

Gene	Function	Forward primer	Reverse primer
cysZ	Sulfate transporter	CATCATTCACATCTGCCCCACG	GCGCCCCCCATCAACAAAATAT
inaA	Lipopolysaccharide kinase	TGGGCAACAGAGGGCGACTG	CGGCCGAACGGATAACGTACG
ucpA	NAD binding oxidoreductase	CTCACGGGCAAGACAGCACTGA	TCCGCCAGCTTTTCGATCTCAG
yeeZ	Epimerase	CGGGTTAGGGTGGTTAGGCATG	GGCTCCATGCGAAGCAGATAGC
yjcO	Unknown function	TTTGACATTCTTTGCCCACGCC	TCAAATCGCCGGAGCTAAACCA
lacZ	β-D-galactosidase		
pair_1		TCCGTGACGTCTCGTTGCTGC	TCACGCAACTCGCCGCACAT
pair_2		CCCGCATCTGACCACCAGCG	CAGCGGCGTCAGCAGTTGTT
pair_3		GTCGTGACTGGGAAAACCCTGG	AACTGTTGGGAAGGGCGATCG
pair_4		AACAACTTTAACGCCGTGCGCT	CACCATGCCGTGGGTTTCAATA
pair_5		CAGCTGGCGCAGGTAGCAGAG	GGCAGATCCCAGCGGTCAAA
rpsJ	30S ribosomal subunit protein	CGTACTCACTTGCGTCTGGTTG	AGGCTGATCTGCACGTCTACAC
rplK	50S ribosomal subunit protein	AGCGGCTGGTATCAAGTCTGG	GTCATGTCGGCAGCTTTGGTC
ERCC 130	RNA spike-ins	GTGAAGATGATTGACCGCACGC	TGCATATTGCAGCTGAGCCAGC
ERCC 002	RNA spike-ins	CCGTCGGCTGATCGTGGTTT	CGACCGTACAGCTCTGGAACCC
ERCC 074	RNA spike-ins	GCCTTGGTAGGGATAGATAGCCACC	CTGGGGTTATGAGTAGGGATGAGCA
ERCC 096	RNA spike-ins	CGTAACCAAACATGCACAGCGG	TCGCGTCATCGATCCGGGT

Table S2: RO and RD values for 1563 monocistronic genes with read count above 10 reads (cf Annexe)

Table S3: Fold-change of ribosomal protein mRNA (*rpsJ* and *rplK*) concentrations in the six recombinant strains after arabinose induction

	rpsJ	***rplK***
***yeeZ* recombinant strain**	1.02	0.76
***inaA* recombinant strain**	1.79	2.00
***cysZ* recombinant strain**	1.01	1.31
***ucpA* recombinant strain**	0.61	0.66
***yjcO* recombinant strain**	1.64	1.33
***lacZ* recombinant strain**	0.59	0.61
Mean value and SD	1.11 ± 0.51	1.11 ± 0.54

CONCLUSION

ET

PERSPECTIVES

CONCLUSION – DISCUSSION

La compréhension de la régulation de l'expression génique des bactéries permet non seulement de comprendre les mécanismes d'adaptation à l'environnement mais aussi de mieux contrôler la croissance et la production de molécules d'intérêt dans un contexte de biologie synthétique.

E. coli est un microorganisme qui possède une capacité métabolique diversifiée et efficace qui lui permet d'exploiter des ressources nutritionnelles variées pour se développer dans différents environnements. Ce modèle d'étude bactérien a permis d'importantes découvertes fondamentales, par exemple sur les mécanismes de régulation de l'expression des gènes, qui ont abouti au développement de nombreux outils pour la biotechnologie (comme des promoteurs régulés types P_{BAD}, P_{lac}…)(Marschall, Sagmeister, and Herwig 2017). La bactérie elle-même est aussi utilisée comme hôte pour la production de composés d'intérêt économique dans les secteurs pharmaceutique et énergétique. Elle présente de nombreux avantages comme une croissance rapide, une facilité à être cultivée, transformée et modifiée génétiquement et une quantité importante de connaissances sur sa génétique et son métabolisme. L'optimisation des voies de synthèse (natives et hétérologues) nécessite la compréhension des différents niveaux de régulations présents à chaque étape de l'expression génique. De plus, comme la cellule doit en permanence maintenir son équilibre métabolique, la modification de l'expression d'un gène peut affecter l'expression des autres gènes. Il est donc nécessaire de connaitre non seulement les régulations de l'expression du ou des gènes d'intérêt, mais aussi celles des autres gènes de la cellule. Pour obtenir cette vue d'ensemble, des approches de biologie des systèmes sont mises en place pour étudier au niveau de la cellule les différentes régulations de l'expression des gènes et leurs interactions.

Vue globale de la traduction chez *E. coli* en croissance rapide

Dans ce travail nous avons effectué une étude à l'échelle -omique de l'état traductionnel d'*E. coli* en croissance rapide sur glucose. Nous voulions étudier comment la traduction est régulée à l'échelle de la cellule pour permettre cette croissance optimale. Pour étudier l'état traductionnel de l'ensemble des gènes (*i.e.* obtenir les données de traductome), nous avons déterminé pour chaque gène deux variables de leur traduction, la « ribosome occupancy » (RO) et leur densité en ribosomes (RD). RO

correspond au pourcentage de copies d'un ARNm donné occupées par au moins un ribosome (soit le pourcentage de copies en traduction) et RD est le nombre de ribosomes par unité de longueur.

Nous avons utilisé la méthode de polysome profiling pour fractionner les ARNm en fonction de leurs charges en ribosomes puis la technique de RNA-Seq pour quantifier les ARNm dans chaque fraction. Cette méthode de séquençage à haut débit présente de nombreux avantages pour étudier le traductome qui sont une haute sensibilité et les accès aux régions non codantes et à l'intégrité de l'ARNm. Le couplage des méthodes de polysome profiling et de RNA-Seq a déjà été utilisé pour étudier la traduction chez des eucaryotes supérieurs ((Chassé et al. 2016), eukaryote cell (Spangenberg et al. 2013)), chez la levure (Aboulhouda et al. 2017), et chez la bactérie *E. coli* (Cambray et al. 2018; Sauert et al. 2016). Pour *E. coli*, dans la première étude, le polysome profiling est effectué à faible résolution (ARN libres versus liés aux ribosomes) et le RNA-Seq est utilisé pour comparer l'expression différentielle entre deux conditions (+/- MazF). La deuxième étude utilise le couplage polysome profiling et RNA-Seq pour calculer la densité en ribosomes de séquences synthétiques. Notre étude effectue un polysome à résolution élevée (fractionnement en plus de fractions) et compare la traduction des gènes natifs au sein d'une condition. Ainsi, pour chaque gène natif, nous pouvons calculer les proportions de ses copies d'ARNm en fonction de la charge en ribosomes. Pour analyser correctement les différences entre les échantillons, nous devons être capables de distinguer et enlever les variations techniques non spécifiques. Le premier chapitre a décrit la comparaison de méthodes bioinformatiques et statistiques pour analyser les données de RNA-Seq. Les étapes cruciales sont l'alignements des séquences et la normalisation. Pour chaque étape, différentes méthodes ont été comparées. Les méthodes d'alignement bwa et de normalisation RUV (basée sur l'utilisation de contrôles externes, dans notre cas les ERCC) ont été choisies pour obtenir des données de traductome fiables. Ces méthodes ont été sélectionnées sur la base de leurs hypothèses d'utilisation et leur efficacité à réduire la variabilité technique des données. A notre connaissance, il s'agit de la première étude de polysome qui utilise les contrôles externes de type ERCC pour normaliser les données de RNA-Seq.

Dans nos travaux, l'analyse des valeurs de RO et RD de l'ensemble des gènes de la cellule donne une photo globale de l'activité traductionnelle chez *E. coli* en croissance

rapide. Les études précédentes sur le traductome se sont intéressées à d'autres problématiques telles que la réponse à un stress ou à la traduction d'un groupe de gènes particuliers (« ARNm leaderless» (Lange et al. 2017), de complexes protéiques (Li et al. 2014)) ou à la pause des ribosomes sur les motifs anti-SD (Li et al. 2012). Aujourd'hui encore la compréhension chez *E. coli* des différences de traduction entre messagers est très parcellaire. Pour la majorité des gènes, nous avons observé que presque toutes les copies d'ARNm sont engagées dans la traduction. Le pourcentage de copies en traduction RO est corrélé positivement avec la vitesse d'initiation de la traduction et le niveau de protéine. Ce résultat montre une efficacité élevée de l'initiation de la traduction en réponse à la demande élevée en protéines quand la bactérie a besoin de croître rapidement. Il suggère aussi que c'est l'étape d'initiation qui limite la traduction. Cependant, les ARNm ont généralement un petit nombre de ribosomes. La densité moyenne en ribosomes est bien plus faible que la densité théorique qui maximise la production de protéines (Zarai et al. 2016; Zouridis and Hatzimanikatis 2007). Cette densité optimale est prédite comme la moitié de la densité maximale de 3.3 ribosomes/100 nucléotides alors que nous avons mesuré une densité moyenne autour de 12% de la densité maximale. Une densité très élevée pourrait provoquer des collisions de ribosomes et des embouteillages qui ralentiraient la traduction (Mitarai et al. 2008; Subramaniam et al. 2014; Zarai et al. 2016). En dessous de la densité optimale, une double limitation par l'initiation et l'élongation est prédite (Racle et al. 2013 ; Zouridis and Hatzimnikatis 2007). Néanmoins, l'élongation de la traduction semble plutôt très efficace dans nos conditions (RD très faible) et nous n'avons pas montré de corrélation entre la densité en ribosomes et le paramètre d'élongation qu'est l'usage de codons. L'ensemble de nos résultats tend donc à minimiser le poids du contrôle par l'élongation et à privilégier l'hypothèse de la limitation de la traduction majoritairement par l'initiation.

Recherche des déterminants de la régulation traductionnelle

La recherche systématique des facteurs clés pour expliquer la variabilité du pourcentage de copies en traduction et de la densité en ribosomes entre les gènes n'a pas été effectuée dans les autres études de traductome de la littérature. Nous avons donc utilisé une approche de régressions linéaires multiples pour identifier les déterminants généraux de la traduction. Cette approche d'analyse multifactorielle est souvent utilisée pour expliquer le niveau de protéines mais plus rarement pour

expliquer l'efficacité de traduction. Fréquemment, les études de la traduction étudient l'effet d'un seul ou d'un petit nombre de paramètres et effectuent des corrélations deux à deux. Or, notre hypothèse est que la traduction est régulée non pas par un seul déterminant mais par de multiples facteurs. De plus, ces facteurs peuvent être liés entre eux, leurs effets sur les variables de la traduction peuvent être redondants. Les corrélations deux à deux ne permettent pas de distinguer ces redondances. Notre approche de régressions multiples permet d'étudier simultanément un grand nombre de paramètres et de hiérarchiser leur importance. Bien que non exhaustive, notre étude a inclus de nombreux paramètres, liés à la fois à l'initiation et à l'élongation de la traduction, aux caractéristiques des protéines correspondantes (fonction, localisation,…) et à l'état physiologique de la cellule via la concentration en ARNm.

Nous avons trouvé comme déterminants de la traduction des facteurs attendus, liés à la séquence de l'ARNm. Le début de la séquence de l'ARNm semble important pour la traduction avec la présence d'un motif particulier dans la région 5'UTR associé à un RO plus élevé et l'influence de la nature des tout premiers codons sur la densité en ribosomes. Deuxièmement, nous avons observé que le pourcentage de GC de la séquence codante avait des effets opposés sur RO et RD : les gènes riches en GC ont tendance à avoir une faible densité mais un RO élevé. Notre résultat est en accord avec le travail de Burkhardt et al (Burkhardt 2017) qui a montré une corrélation négative entre la structure secondaire de l'ORF et la densité en ribosomes obtenue par ribosome profiling.

Cependant, deux paramètres liés à la séquence de l'ARNm largement étudiés tels que la structure de la région -30 +24 nts à partir du codon start et l'usage des codons de la séquence codante ne sont pas sélectionnés par nos modèles comme paramètres significatifs de la traduction. Or, la structure secondaire du RBS a été trouvée comme jouant un rôle sur l'initiation de la traduction via l'accessibilité du ribosome (Kudla et al. 2009; Seo et al. 2009; Tuller et al. 2010). L'effet du CAI reste encore un sujet de débat, certaines études trouvent un effet positif alors que d'autres ne trouvent aucun effet. Les études qui ont montré un effet du CAI utilisent souvent des gènes synthétiques (Boël et al. 2016; Presnyak et al. 2015) alors que les études de la variabilité de traduction entre des gènes natifs (Arava et al. 2003; Picard et al. 2012) comme dans notre cas, n'ont pas trouvé l'effet de ce paramètre sur la densité en ribosomes. Une hypothèse est que pour les gènes natifs, l'usage des codons a été

sélectionné au cours de l'évolution pour être optimal par rapport à la disponibilité en ARNt dans la cellule (Kudla et al. 2009). Le CAI pourrait avoir un effet local sur certaines régions de l'ARNm mais pas un effet global sur l'ARNm en entier.

Le résultat le plus surprenant de notre approche de régressions multiples a été d'identifier la concentration en ARNm comme un facteur important d pourcentage de copies en traduction et dans une moindre mesure de la densité en ribosomes. Bien que ce facteur soit très étudié pour expliquer le niveau de protéines, peu d'études analysent son effet sur l'efficacité de traduction. Ces analyses ont abouti à des résultats différents. Utilisant aussi la méthode polysome profiling et les mêmes variables, l'effet du niveau de l'ARNm sur RD n'est pas significatif chez *S. cerevisae* (Arava et al. 2003), a un effet positif sur RO et RD chez *S. pombe* (Lackner et al. 2007) et un effet positif sur RO mais négatif sur RD chez *L. lactis* (Picard et al. 2012).

Nos modèles de régressions multiples de RO et RD ont des R^2 plutôt faibles. Cela signifie qu'il existe encore d'autres facteurs importants de la régulation traductionnelle à identifier. En particulier pour la densité en ribosomes, pour mieux la comprendre et l'expliquer, il faudrait compléter nos travaux avec (i) un modèle mécanistique/biophysique (Racle et al. 2013; Shaham and Tuller 2018) et (ii) la détermination de densités locales en ribosomes par la méthode de ribosome profiling. En effet certains paramètres peuvent avoir un effet seulement local et non pas sur tout l'ARNm comme cela a déjà été montré pour le CAI, les structures secondaires, des motifs de pause des ribosomes (de type SD-like ou polyprolines). En complément, il faudrait considérer des valeurs locales de CAI (au niveau d'un codon ou de quelques codons) et aussi travailler sur une fenêtre glissante le long de la séquence de l'ARNm pour la prédiction de structures secondaires.

Validation moléculaire de l'effet de l'expression sur la traduction

Nous avons ensuite validé expérimentalement à l'échelle moléculaire l'effet positif de l'expression sur la traduction qui était prédit par notre approche de régressions multiples. Nous avons augmenté la concentration d'ARNm cibles par induction transcriptionnelle et comparé leur pourcentage de copies en traduction et leur charge en ribosomes entre deux niveaux de concentrations.

Afin d'étudier simultanément six ARNm cibles, nous avons développé une méthode de multiplexage de polysome profiling. Cette méthode a permis de diminuer

le temps, le coût et les erreurs techniques, tout en donnant les mêmes résultats que les analyses non-multiplexées. Cette méthode a été utilisée pour étudier l'effet de la concentration sur la traduction mais elle pourrait être étendue à l'étude d'autres facteurs liés à la séquence de l'ARNm (par exemple la présence d'un motif, d'un peptide signal, la séquence du RBS, l'usage des codons). Elle pourrait aussi être utilisée pour étudier les différences de traduction des allèles d'un même gène présents dans des souches génétiquement proches ou la traduction des gènes spécifiques d'espèces génétiquement éloignées.

Nos expériences ont montré pour six gènes que l'augmentation de la concentration de leur ARNm entrainait l'augmentation de leur pourcentage de copies en traduction et de leur charge en ribosomes. Il s'agit de la première validation à l'échelle moléculaire de l'effet de la concentration de l'ARNm sur des paramètres de traduction. La corrélation positive entre traduction et le niveau de l'ARNm pourrait venir de différents mécanismes, la traduction étant en relation avec les autres processus cellulaires comme la transcription et la stabilité des ARNm (voir l'étude bibliographique). Ici, nous avons montré la causalité entre le niveau de transcription d'un ARNm et l'efficacité de sa traduction. Le mécanisme de contrôle reste à identifier mais une hypothèse pourrait être l'implication d'un simple mécanisme physique. En effet, on pourrait considérer la rencontre entre ribosomes et ARNm comme une interaction enzyme-substrat. Nous avons estimé à partir des profils polysomiques que 44% des sous-unités ribosomales n'étaient pas impliquées dans la traduction. Ce résultat indique que bien que les ribosomes soient en excès ils ne seraient pas en très large excès. Il pourrait y avoir une compétition entre les ARNm pour rencontrer les ribosomes libres, d'autant plus que des problèmes de diffusion des macromolécules dans le cytoplasme de *E. coli* ont été démontrés par des approches en modélisation (Ando and Skolnick 2010). Ainsi, plus un ARNm est concentré, plus sa probabilité de rencontre avec les ribosomes libres serait élevée, d'où une fréquence d'initiation plus forte. Pour savoir si les ribosomes libres sont vraiment limitants, il faudrait quantifier les protéines ribosomales par spectrométrie de masse en faisant la part des protéines provenant de ribosomes libres et de ribosomes en traduction.

Notre résultat sur l'effet positif de l'expression sur la traduction confirme le rôle important de la concentration des ARNm dans la régulation de l'expression génique. En effet, notre équipe a récemment démontré que la concentration en ARNm est un

facteur important de la régulation de la stabilité des ARNm (Nouaille et al. 2017). Plus un ARNm est concentré, plus sa vitesse de dégradation est élevée. Ceci soulève la question fréquemment posée de la relation entre la stabilité des ARNm et la traduction. Dans ce travail, nous avons montré que plus un ARNm est concentré, plus il est occupé par les ribosomes. Ainsi, la traduction aurait-elle tendance à activer la dégradation des ARNm? Sur le plan physiologique, cela voudrait dire un effet antagoniste de la traduction et de la dégradation des ARNm dans la régulation de l'expression génique. Comme la concentration et la stabilité des ARNm sont très (négativement) corrélées, il est mathématiquement déconseillé d'introduire simultanément ces deux paramètres dans notre méthode de régressions multiples. Cependant, nous avons pu examiner les corrélations simples entre les stabilités des ARNm et les paramètres de traduction. Nous avons trouvé une faible corrélation négative entre RO et les temps de demi-vie (coefficient -0.13, p-value $<2\ 10^{-16}$) mais une faible corrélation positive entre RD et les temps de demi-vie (coefficient 0.09 avec p-value $2\ 10^{-05}$). Ces résultats contrastés demandent donc des expériences complémentaires pour comprendre le lien stabilité et traduction. Afin d'éclaircir la relation entre la stabilité des ARNm et la traduction, notre équipe est en train de réaliser des expériences moléculaires afin de changer la charge en ribosomes d'un ARNm rapporteur et de mesurer les conséquences sur sa stabilité. Deux approches pour modifier la charge en ribosomes sont mises en œuvre, la première consiste à modifier la force du RBS afin de moduler la fréquence d'initiation de la traduction et la deuxième correspond à l'introduction d'un codon stop dans la séquence codante afin de mettre à nu (sans ribosome) une partie plus ou moins importante de l'ARNm.

L'autre originalité de nos résultats est que nous avons démontré des régulations co-directionnelles de la transcription et de la traduction même en condition de croissance optimale. La coordination entre les régulations transcriptionnelle et traductionnelle a été reportée précédemment dans les conditions de stress chez la levure et *L. lactis* (Melamed et al. 2008; Racle et al. 2013). Cet effet amplificateur de la traduction a été étudié par deux travaux de modélisation chez la levure (Csárdi et al. 2015; Li, 2017). Basées sur une approche statistique des données de ribosome profiling, ces études proposent que l'effet amplificateur de la traduction dans le sens de la transcription pourrait en partie correspondre à la part de la quantité de protéines non explicable par une relation linéaire avec la quantité d'ARNm. L'efficacité de

traduction (nombre de protéines par ARNm) d'un gène varierait à hauteur de 20% en fonction de la valeur de la concentration de son ARNm et pour 80 % de manière indépendante de la concentration en ARNm (Li et al. 2017). Les régulations traductionnelles ont donc un fort potentiel de régulation de l'expression génique et apparaissent comme un levier original d'optimisation de micro-organismes pour des applications en biotechnologie.

Par ailleurs, la concentration en ARNm, à la différence des paramètres liés à la séquence, n'est pas figée, elle dépend de l'état physiologique de la cellule et donc des conditions de croissance. Nos résultats indiquent que la régulation de la traduction n'est pas que déterministe mais elle doit aussi être adaptée à l'environnement. Afin de savoir si la concentration des ARNm joue toujours un rôle important dans la régulation des ARNm en dehors de la croissance rapide, nous étudions dans l'équipe la corrélation entre les données de transcriptome et traductome en condition de croissance très lente (μ=0.04 h^{-1}). Cette croissance très lente est obtenue chez *E. coli* lors d'une culture batch sur milieu M9 plus glucose, une heure après l'épuisement du glucose quand les cellules croissent en re-consommant l'acétate produit pendant la phase de consommation de glucose.

Nos expériences moléculaires ont démontré l'accroissement de la densité en ribosomes lors de l'augmentation de la concentration de l'ARNm. Sachant que RD est dépendante à la fois de l'initiation et l'élongation de la traduction, une modification de RD peut venir d'un changement de l'initiation et/ou de la vitesse d'élongation. Nous pourrions compléter nos travaux par des expériences de type 'ribosome run-off' (Presnyak et al. 2015) pour mesurer des variations éventuelles de la vitesse d'élongation en fonction de la concentration en ARNm. Au cours des expériences de 'ribosome run-off', l'initiation est arrêtée par un passage sur un milieu sans source de carbone, puis l'allure du profil polysomique est analysée avant et après l'arrêt de l'initiation de la traduction. Si par exemple on observe une disparition plus importante des ribosomes des fractions très chargées quand la concentration en ARNm augmente alors cela indiquera que la vitesse d'élongation est supérieure à forte concentration en ARNm.

PERSPECTIVES

Modulations de l'effet de la concentration sur la traduction

Nous avons remarqué lors de l'analyse omique (Fig 1) et lors des expériences moléculaires que tous les gènes ne répondent pas de la même façon au niveau traductionnel à l'effet de la concentration.

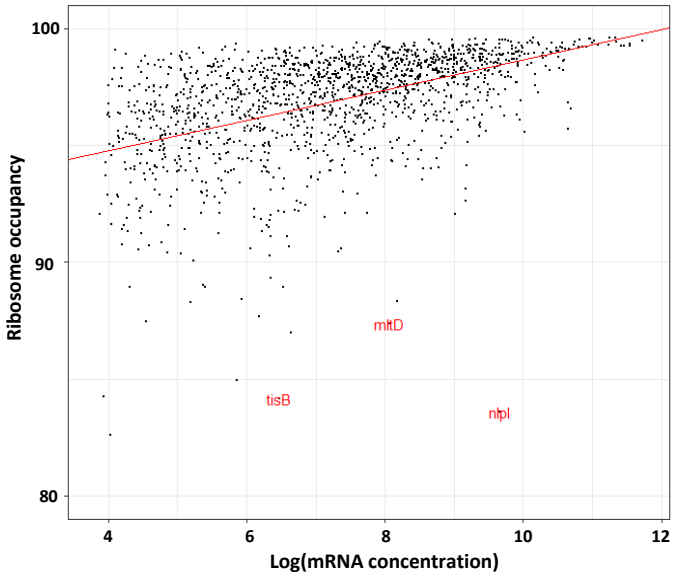


Figure 1 : Corrélation entre le pourcentage de copies en traduction (RO) et la concentration en ARNm. Les gènes en rouge sont quelques exemples de gènes qui sortent très fortement de la corrélation entre RO et la concentration en ARNm.

Pour étudier cette hétérogénéité des réponses traductionnelles, nous avons d'abord regardé à l'échelle du génome si la relation entre la concentration en ARNm et RO variait en fonction d'un autre facteur. Un autre facteur pourrait avoir un effet modulateur de la relation entre ces deux variables. En effet, l'affinité entre un ARNm et les ribosomes peut être plus ou moins forte si la structure secondaire du 5'UTR ou la force de RBS sont plus ou moins fortes. Pour tester l'effet d'un facteur sur la relation entre la concentration en ARNm et RO nous avons comparé la pente entre RO et la concentration pour les différentes modalités de ce facteur. Par exemple, la figure 2A montre l'évolution de la corrélation entre concentration en ARNm et RO pour les 4 modalités possibles du facteur « localisation de la protéine ». Avec cette approche, nous avons testé des facteurs qualitatifs (la localisation dans la cellule, être une protéine membranaire, être cible de CsrA, être un gène essentiel et les catégories fonctionnelles COG) et des facteurs quantitatifs relatifs à l'initiation de traduction (DeltaG de la région

-30 + 24 nts, %GC du 5'UTR, Longueur du 5'UTR, force du RBS estimée par RBS Calculator). Pour les facteurs quantitatifs, nous divisons les gènes en 4 groupes comme le montre pour exemple la figure 2B avec le facteur quantitatif du score RBS. Aucune de ces facteurs n'a un effet modulateur significatif sur la pente entre la concentration en ARNm et RO. A ce stade, nous n'avons donc pas identifié de facteur influençant l'effet de la concentration en ARNm sur le pourcentage de copies en traduction. Nous n'avons pas testé ce type d'approche sur la relation entre la densité en ribosomes et la concentration en ARNm car leur corrélation est très faible au niveau omique.

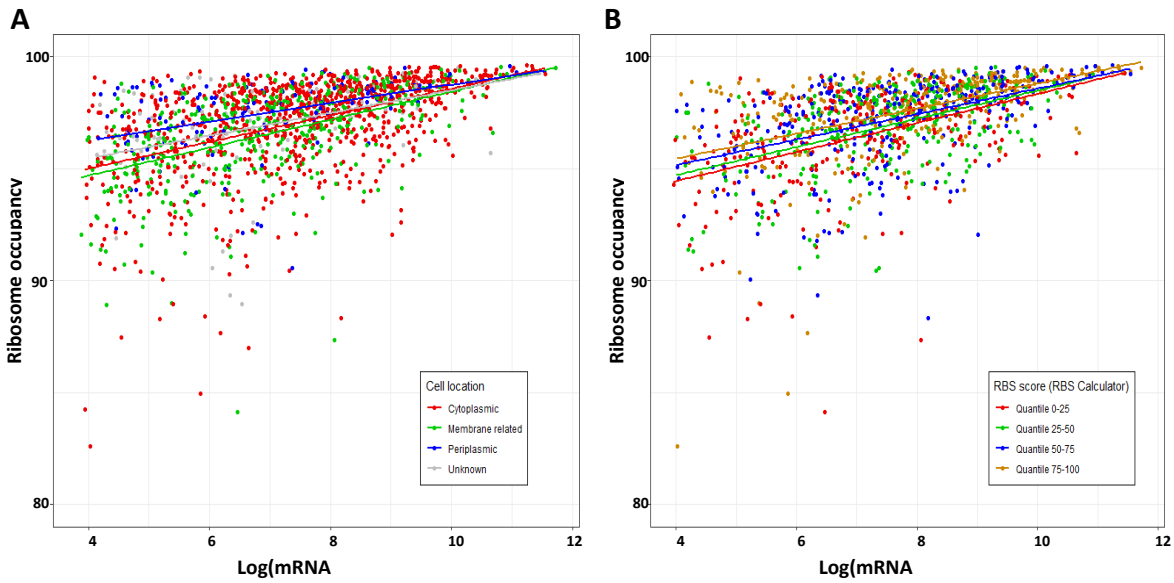


Figure 2 : Corrélation entre le pourcentage de copies en traduction (RO) et la concentration en ARNm. (A) En fonction de la localisation de la protéine. (B) En fonction des quantiles du score RBS estimé par RBS calculator.

Par ailleurs, nous pouvons aussi émettre l'hypothèse que les gènes qui s'éloignent de la tendance générale sont soumis à des régulations traductionnelles spécifiques importantes. Ces régulations spécifiques masqueraient l'effet de la concentration en ARNm sur le pourcentage de copies en traduction. Par exemple un RO élevé bien que la concentration en ARNm soit faible pourrait résulter d'un mécanisme d'activation spécifique de la traduction. Au contraire, un RO faible malgré une concentration en ARNm élevée (par exemple comme pour les gènes *tisB*, *mltD* et *nlpI*) (Fig. 1) pourrait signifier la présence d'un mécanisme d'inhibition de la traduction particulier. C'est par exemple le cas pour le gène *tisB* codant pour une toxine dont la traduction est inhibée dans les conditions de croissance optimale par le petit ARN

antitoxine *IstR*. Nos résultats permettraient donc de lister les gènes soumis à des régulations spécifiques importantes et l'analyse plus en détail de la traduction de ces gènes pourrait identifier leur mécanisme de régulation de la traduction (par exemple pour *mltD* et *nlpI* (deux lipoprotéines de la membrane externe) où il n'y a pas encore de régulation spécifique connue).

Cas particuliers des opérons

Une limite de notre étude est l'analyse des gènes polycistroniques. En effet, les gènes d'un même opéron sont à priori transcrits sur un même ARNm, or notre méthode de polysome profiling ne permet pas de distinguer le nombre de ribosomes exact liés à chaque cistron. En effet, nous obtenons une densité en ribosomes moyenne sur pour l'ensemble de l'ARNm (en divisant le nombre de ribosomes chargés par la longueur de l'ARNm) mais pas la densité en ribosomes spécifique de chaque cistron de l'opéron. Chez *E. coli*, la proportion de gènes polycistroniques est importante : 2517 gènes sont organisés en 849 opérons. Nous avons remarqué que pour 22% des opérons, tous les cistrons ont le même comportement dans le profil polysomique. Pour 183 opérons, tous les gènes de l'opéron se retrouvent majoritairement dans la même fraction du polysome. Cela est en accord avec l'hypothèse que ces gènes sont tous exprimés sur le même transcrit. Par contre pour 16 opérons, les cistrons d'un même opéron ne sont pas tous élués dans la même fraction préférentielle. Cela suppose que tous les cistrons ne sont plus exprimés sur un même transcrit. La présence de plusieurs transcrits pour un même opéron peut provenir de promoteurs internes ou de sites de clivage internes.

Pour accéder à la densité en ribosomes des gènes polycistroniques, il faut utiliser la méthode complémentaire de ribosome profiling. Cette méthode a été utilisée chez *E. coli* en milieu minimum MOPS (Li et al. 2014) et les données sont disponibles. Nous avons montré dans le chapitre III que pour les gènes monocistroniques, la densité en ribosomes obtenue par ribosome profiling est partiellement corrélée avec celle obtenue par notre méthode de polysome profiling. Nous avons donc utilisé les valeurs de RD de 377 gènes polycistroniques obtenues par la méthode de ribosome profiling pour rechercher les déterminants de la densité en ribosomes des gènes polycistroniques. Avec notre approche de régressions multiples, nous avons trouvé que le biais d'usage des codons est la variable majeure du RD des gènes polycistroniques avec un coefficient positif significatif (coefficient de corrélation de 0.70 et une p-value de $3.67\ 10^{-38}$). La concentration en ARNm a un effet très peu

significatif (coefficient de corrélation de -0.07 et une pval de 0.08). La différence de ces résultats avec ceux obtenus pour les gènes monocistroniques pourrait venir (i) de différents types de densités en ribosomes mesurées en ribosome profiling et en polysome profiling. Même si ces densités sont partiellement corrélées elles ne sont pas équivalentes. Dans la technique du ribosome profiling, la densité est moyennée sur l'ensemble des copies d'un ARNm alors qu'en polysome profiling on choisit la densité la plus fréquente parmi les copies d'un ARNm; et (ii) de mécanismes différents de régulation de la charge en ribosomes entre les gènes polycistroniques et monocistroniques.

Quant au pourcentage de copies en traduction RO des gènes polycistroniques, il peut être surestimé par la méthode de polysome profiling. En effet, si le transcrit d'un opéron est occupé par au moins un ribosome, alors tous les gènes de l'opéron sont considérés comme aussi associés à un ribosome. Or rien ne prouve que ce soit vraiment le cas. Le pourcentage de copies en traduction ne peut pas être obtenu avec la méthode de ribosome profiling seule qui ne différencie pas les différentes copies d'ARNm. La seule solution serait de combiner le polysome profiling et ribosome profiling de façon à fractionner les complexes ARNm – polysomes à l'aide du gradient de sucrose et à effectuer le footprint au sein de chaque fraction.

Implications de nos travaux en biotechnologie

Nos travaux ont montré un effet amplificateur de la régulation de la traduction sur celle de la transcription. Ainsi lors d'expériences de surexpression de gène (aussi bien homologue qu'hétérologue) en biologie synthétique, l'augmentation de la synthèse de l'ARNm devrait être associée à une augmentation du pourcentage de copies en traduction et de la charge en ribosomes. Cependant, l'augmentation du niveau de transcription d'un gène ou des gènes d'une voie de synthèse pourrait induire un titrage des ribosomes libres et influencer le niveau de traduction des autres gènes. En effet, un travail de modélisation a proposé que les traductions des différents ARNm soient interdépendantes car elles partagent le même pool de ribosomes libres (Raveh et al. 2015). Un changement de la vitesse d'initiation ou de l'élongation d'un gène peut réduire ou augmenter le nombre de ribosomes disponibles et affecter la traduction des autres gènes. L'interdépendance de traduction entre les gènes est démontrée expérimentalement chez *E. coli* : la surexpression d'un gène contenant des codons rares affecte le niveau de traduction des gènes contenant ces mêmes codons rares

(Frumkin et al. 2018). Pour alimenter le pool en ribosomes libres lors de la surexpression de gènes, une stratégie a été proposée chez *E. coli* basée sur l'introduction de ribosomes orthogonaux spécifiquement dédiés à la surexpression des gènes d'intérêt (Darlington et al. 2018). Grâce à nos travaux, nous pourrions aussi imaginer de diminuer l'expression des gènes dont les ARNm sont fortement chargés en ribosomes et dont la fonction n'est pas nécessaire dans les conditions de croissance. En effet certains gènes sont transcrits en excès ou ne sont pas forcément utilisés dans certaines conditions de croissance (Price et al. 2016). Une diminution de l'expression des ARNm non nécessaires et séquestrant beaucoup de ribosomes devrait libérer des ribosomes pour la traduction des autres gènes.

REFERENCES

Aboulhouda, S., Di Santo, R., Therizols, G., and Weinberg, D. (2017). Accurate, Streamlined Analysis of mRNA Translation by Sucrose Gradient Fractionation. Bio Protoc *7*.

Ando, T., and Skolnick, J. (2010). Crowding and hydrodynamic interactions likely dominate in vivo macromolecular motion. Proceedings of the National Academy of Sciences *107*, 18457–18462.

Arava, Y., Wang, Y., Storey, J.D., Liu, C.L., Brown, P.O., and Herschlag, D. (2003). Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A *100*, 3889–3894.

Boël, G., Letso, R., Neely, H., Price, W.N., Wong, K.-H., Su, M., Luff, J.D., Valecha, M., Everett, J.K., Acton, T.B., et al. (2016). Codon influence on protein expression in E. coli correlates with mRNA levels. Nature *529*, 358–363.

Cambray, G., Guimaraes, J.C., and Arkin, A.P. (2018). Evaluation of 244,000 synthetic sequences reveals design principles to optimize translation in Escherichia coli. Nature Biotechnology.

Chassé, H., Boulben, S., Costache, V., Cormier, P., and Morales, J. (2016). Analysis of translation using polysome profiling. Nucleic Acids Research gkw907.

Csárdi, G., Franks, A., Choi, D.S., Airoldi, E.M., and Drummond, D.A. (2015). Accounting for Experimental Noise Reveals That mRNA Levels, Amplified by Post-Transcriptional Processes, Largely Determine Steady-State Protein Levels in Yeast. PLOS Genetics *11*, e1005206.

Darlington, A.P.S., Kim, J., Jiménez, J.I., and Bates, D.G. (2018). Dynamic allocation of orthogonal ribosomes facilitates uncoupling of co-expressed genes. Nat Commun *9*.

Frumkin, I., Lajoie, M.J., Gregg, C.J., Hornung, G., Church, G.M., and Pilpel, Y. (2018). Codon usage of highly expressed genes affects proteome-wide translation efficiency. Proceedings of the National Academy of Sciences *115*, E4940–E4949.

Kudla, G., Murray, A.W., Tollervey, D., and Plotkin, J.B. (2009). Coding-Sequence Determinants of Gene Expression in Escherichia coli. *324*, 4.

Lackner, D.H., Beilharz, T.H., Marguerat, S., Mata, J., Watt, S., Schubert, F., Preiss, T., and Bähler, J. (2007). A Network of Multiple Regulatory Layers Shapes Gene Expression in Fission Yeast. Molecular Cell *26*, 145–155.

Lange, C., Lehr, M., Zerulla, K., Ludwig, P., Schweitzer, J., Polen, T., Wendisch, V.F., and Soppa, J. (2017). Effects of Kasugamycin on the Translatome of Escherichia coli. PLOS ONE *12*, e0168143.

Li, G.-W., Oh, E., and Weissman, J.S. (2012). The anti-Shine–Dalgarno sequence drives translational pausing and codon choice in bacteria. Nature *484*, 538–541.

Li, G.-W., Burkhardt, D., Gross, C., and Weissman, J.S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. Cell *157*, 624–635.

Li, J.J., Chew, G.-L., and Biggin, M.D. (2017). Quantitating translational control: mRNA abundance-dependent and independent contributions and the mRNA sequences that specify them. Nucleic Acids Research *45*, 11821–11836.

Marschall, L., Sagmeister, P., and Herwig, C. (2017). Tunable recombinant protein expression in E. coli: promoter systems and genetic constraints. Applied Microbiology and Biotechnology *101*, 501–512.

Melamed, D., Pnueli, L., and Arava, Y. (2008). Yeast translational response to high salinity: Global analysis reveals regulation at multiple levels. RNA *14*, 1337–1351.

Mitarai, N., Sneppen, K., and Pedersen, S. (2008). Ribosome Collisions and Translation Efficiency: Optimization by Codon Usage and mRNA Destabilization. Journal of Molecular Biology *382*, 236–245.

Nouaille, S., Mondeil, S., Finoux, A.-L., Moulis, C., Girbal, L., and Cocaign-Bousquet, M. (2017). The stability of an mRNA is influenced by its concentration: a potential physical mechanism to regulate gene expression. Nucleic Acids Research *45*, 11711–11724.

Picard, F., Milhem, H., Loubière, P., Laurent, B., Cocaign-Bousquet, M., and Girbal, L. (2012). Bacterial translational regulations: high diversity between all mRNAs and major role in gene expression. BMC Genomics *13*, 528.

Presnyak, V., Alhusaini, N., Chen, Y.-H., Martin, S., Morris, N., Kline, N., Olson, S., Weinberg, D., Baker, K.E., Graveley, B.R., et al. (2015). Codon Optimality Is a Major Determinant of mRNA Stability. Cell *160*, 1111–1124.

Price, M.N., Wetmore, K.M., Deutschbauer, A.M., and Arkin, A.P. (2016). A Comparison of the Costs and Benefits of Bacterial Gene Expression. PLOS ONE 22.

Racle, J., Picard, F., Girbal, L., Cocaign-Bousquet, M., and Hatzimanikatis, V. (2013). A Genome-Scale Integration and Analysis of Lactococcus lactis Translation Data. PLoS Computational Biology *9*, e1003240.

Raveh, A., Margaliot, M., Sontag, E.D., and Tuller, T. (2015). A Model for Competition for Ribosomes in the Cell. ArXiv:1508.02408 [q-Bio].

Sauert, M., Wolfinger, M.T., Vesper, O., Müller, C., Byrgazov, K., and Moll, I. (2016). The MazF-regulon: a toolbox for the post-transcriptional stress response in Escherichia coli. Nucleic Acids Res *44*, 6660–6675.

Seo, S.W., Yang, J., and Jung, G.Y. (2009). Quantitative correlation between mRNA secondary structure around the region downstream of the initiation codon and translational efficiency in Escherichia coli. Biotechnology and Bioengineering *104*, 611–616.

Shaham, G., and Tuller, T. (2018). Genome scale analysis of Escherichia coli with a comprehensive prokaryotic sequence-based biophysical model of translation initiation and elongation. DNA Research *25*, 195–205.

Spangenberg, L., Shigunov, P., Abud, A.P.R., Cofré, A.R., Stimamiglio, M.A., Kuligovski, C., Zych, J., Schittini, A.V., Costa, A.D.T., Rebelatto, C.K., et al. (2013). Polysome profiling shows extensive posttranscriptional regulation during human adipocyte stem cell differentiation into adipocytes. Stem Cell Research *11*, 902–912.

Subramaniam, A.R., Zid, B.M., and O'Shea, E.K. (2014). An Integrated Approach Reveals Regulatory Controls on Bacterial Translation Elongation. Cell *159*, 1200–1211.

Tuller, T., Waldman, Y.Y., Kupiec, M., and Ruppin, E. (2010). Translation efficiency is determined by both codon bias and folding energy. Proceedings of the National Academy of Sciences *107*, 3645–3650.

Zarai, Y., Margaliot, M., and Tuller, T. (2016). On the Ribosomal Density that Maximizes Protein Translation Rate. PLOS ONE *11*, e0166481.

Zouridis, H., and Hatzimanikatis, V. (2007). A Model for Protein Translation: Polysome Self-Organization Leads to Maximum Protein Synthesis Rates. Biophysical Journal *92*, 717–730.

ANNEXES

Table S2 (Chapitre III) : RO (percentage of mRNA copies) and RD (ribosomes/100 nucleotides) values of 1563 monocistronic genes

GeneID	GeneSymbol	RO	RD
b0005	yaaX	97.9	NA
b0006	yaaA	98.8	0.4
b0007	yaaJ	91.9	0.2
b0008	talB	99.1	NA
b0009	mog	98.6	0.5
b0010	satP	96.7	NA
b0011	yaaW	95.5	0.4
b0023	rpsT	96.7	NA
b0030	rihC	98.6	NA
b0031	dapB	99.2	0.3
b0034	caiF	98.9	0.7
b0045	yaaU	94.4	0.2
b0048	folA	96.5	0.6
b0055	djlA	96.7	0.3
b0060	polB	94.3	0.1
b0064	araC	97.7	0.3
b0065	yabI	95.0	0.4
b0076	leuO	95.6	0.3
b0080	cra	96.5	0.3
b0104	guaC	98.6	NA
b0109	nadC	95.0	0.3
b0112	aroP	99.0	0.8
b0117	yacH	95.5	NA
b0118	acnB	99.4	NA
b0119	yacL	98.7	0.8
b0123	cueO	98.7	NA
b0124	gcd	98.7	0.1
b0125	hpt	98.2	0.5
b0126	can	97.9	0.4
b0129	yadI	95.4	0.6
b0130	yadE	94.6	0.2
b0131	panD	99.1	0.7
b0132	yadD	95.2	NA
b0147	ligT	94.4	NA
b0148	hrpB	96.7	NA
b0149	mrcB	98.0	NA
b0154	hemL	98.5	NA
b0155	clcA	95.3	NA
b0156	erpA	96.4	NA
b0160	dgt	97.5	0.2
b0161	degP	98.7	NA
b0162	cdaR	96.1	NA
b0163	yaeH	97.9	0.7
b0164	yaeI	97.0	0.3
b0171	pyrH	98.9	NA
b0172	frr	97.4	0.5
b0173	dxr	96.2	0.2
b0174	ispU	96.6	0.4
b0175	cdsA	97.4	NA
b0176	rseP	97.3	NA
b0185	accA	98.4	0.3
b0328	yahN	97.5	0.4
b0329	yahO	98.7	1.0

GeneID	GeneSymbol	RO	RD
b0186	ldcC	98.8	NA
b0187	yaeR	98.7	0.7
b0188	tilS	98.5	0.2
b0193	yaeF	94.6	1.4
b0194	proS	99.2	NA
b0200	gmhB	92.1	0.5
b0207	dkgB	97.7	NA
b0208	yafC	99.2	0.3
b0211	mltD	87.4	NA
b0212	gloB	98.6	0.4
b0213	yafS	95.0	0.4
b0214	rnhA	96.4	0.6
b0215	dnaQ	97.3	0.4
b0217	yafT	95.7	0.4
b0219	yafV	96.8	0.4
b0220	ivy	99.1	0.6
b0221	fadE	95.5	NA
b0222	gmhA	98.8	0.5
b0223	yafJ	98.7	0.4
b0224	yafK	97.5	0.4
b0227	yafL	95.3	0.4
b0228	rayT	98.4	0.6
b0229	lfhA	94.8	NA
b0230	lafU	94.6	NA
b0237	pepD	98.9	NA
b0238	gpt	98.3	0.6
b0239	frsA	97.3	0.2
b0240	crl	96.8	0.2
b0241	phoE	97.6	0.3
b0254	perR	96.3	NA
b0258	ykfC	97.8	0.3
b0272	yagI	98.4	NA
b0273	argF	99.5	NA
b0276	yagJ	92.9	0.3
b0277	yagK	90.4	NA
b0280	yagN	98.0	0.6
b0281	intF	98.8	0.2
b0287	yagU	99.2	0.5
b0288	ykgJ	97.2	0.8
b0293	ecpA	98.7	0.5
b0295	ykgL	93.1	1.2
b0300	ykgA	98.0	0.4
b0304	rclA	98.7	NA
b0305	rclR	94.4	0.3
b0314	betT	97.1	0.1
b0315	yahA	97.5	0.3
b0316	yahB	92.5	0.3
b0317	yahC	95.4	NA
b0325	yahK	99.1	0.3
b0326	yahL	84.2	NA
b0327	yahM	87.5	1.1
b0465	mscK	96.6	NA
b0468	ybaN	96.1	0.7

b0330	prpR	99.0	NA
b0338	cynR	99.0	0.3
b0353	mhpT	97.3	0.9
b0354	yaiL	97.3	0.5
b0363	yaiP	94.8	NA
b0369	hemB	98.5	0.3
b0375	yaiV	82.6	0.2
b0376	ampH	97.7	0.2
b0379	yaiY	94.9	NA
b0380	yaiZ	91.5	NA
b0381	ddlA	97.9	0.3
b0382	iraP	97.0	1.1
b0385	yaiC	96.4	NA
b0386	proC	98.9	0.3
b0387	yaiI	95.2	NA
b0391	yaiE	97.5	1.0
b0392	ykiA	97.6	0.8
b0393	rdgC	98.8	0.3
b0394	mak	99.2	0.3
b0396	araJ	97.2	0.2
b0403	malZ	98.8	NA
b0404	acpH	96.9	0.5
b0405	queA	97.6	0.3
b0410	yajD	98.4	0.8
b0411	tsx	98.7	0.3
b0412	yajI	96.9	0.5
b0423	thiI	98.6	0.2
b0426	yajQ	98.5	0.6
b0427	yajR	97.4	NA
b0433	ampG	98.0	NA
b0434	yajG	98.8	0.5
b0435	bolA	97.8	0.9
b0436	tig	98.4	0.2
b0440	hupB	98.8	1.0
b0441	ppiD	98.5	NA
b0442	ybaV	98.3	0.7
b0443	fadM	97.0	NA
b0444	queC	97.0	0.4
b0445	ybaE	96.7	0.2
b0446	cof	93.7	0.3
b0447	ybaO	95.8	0.6
b0452	tesB	97.5	0.3
b0453	ybaY	99.0	0.5
b0454	ybaZ	95.3	0.7
b0456	ybaA	90.0	0.8
b0457	ylaB	93.8	0.2
b0458	ylaC	97.0	0.6
b0459	maa	96.9	0.5
b0464	acrR	87.0	NA
b0608	ybdR	97.5	0.2
b0610	rnk	97.3	0.7
b0611	rna	97.3	0.3
b0612	citT	94.2	0.2
b0621	dcuC	94.0	0.2
b0622	pagP	91.9	0.5
b0623	cspE	96.9	1.3
b0624	flc	97.1	0.7

b0469	apt	98.1	0.5
b0470	dnaX	96.5	0.1
b0473	htpG	99.5	NA
b0474	adk	98.1	0.4
b0475	hemH	97.3	0.3
b0476	aes	97.3	NA
b0477	gsk	97.4	0.2
b0478	ybaL	97.2	NA
b0479	fsr	98.0	NA
b0480	ushA	98.9	NA
b0481	ybaK	98.4	0.6
b0482	ybaP	95.0	NA
b0483	ybaQ	95.8	0.8
b0484	copA	98.6	0.1
b0487	cueR	97.9	0.7
b0492	ybbN	99.5	NA
b0502	ylbG	96.6	0.7
b0503	mnmH	93.6	NA
b0504	allS	92.1	1.2
b0505	allA	97.4	0.6
b0506	allR	98.5	0.3
b0526	cysS	98.8	0.2
b0527	ybcI	98.8	NA
b0537	intD	96.1	0.2
b0544	ybcK	93.7	NA
b0551	quuD	94.0	0.7
b0553	nmpC	99.5	NA
b0557	borD	98.7	0.9
b0559	ybcW	96.7	1.3
b0561	tfaD	95.4	0.2
b0562	ybcY	90.8	NA
b0564	appY	97.5	NA
b0567	ybcH	99.1	NA
b0576	pheP	98.2	NA
b0577	ybdG	98.0	0.2
b0578	nfsB	98.9	0.4
b0579	ybdF	97.7	0.8
b0580	ybdJ	94.4	1.1
b0581	ybdK	96.4	NA
b0591	entS	99.0	0.2
b0592	fepB	97.6	NA
b0598	cstA	98.9	0.5
b0599	ybdH	97.4	0.3
b0600	ybdL	98.6	NA
b0601	ybdM	96.7	0.4
b0602	ybdN	96.6	0.2
b0603	ybdO	96.4	0.3
b0604	dsbG	98.6	0.4
b0607	uspG	99.2	0.6
b0800	ybiB	99.1	NA
b0801	ybiC	99.5	NA
b0802	ybiJ	98.6	1.1
b0803	ybiI	91.1	NA
b0804	ybiX	98.3	0.4
b0805	fiu	99.1	0.5
b0806	mcbA	98.3	1.1
b0807	rlmF	98.6	0.3

b0627	tatE	96.6	1.4
b0628	lipA	95.6	0.3
b0629	ybeF	95.5	0.3
b0632	dacA	98.7	0.2
b0643	ybeL	98.8	0.6
b0644	ybeQ	95.4	NA
b0650	hscC	94.4	NA
b0651	rihA	98.1	0.3
b0661	miaB	98.7	0.2
b0662	ubiF	98.4	NA
b0674	asnB	99.4	NA
b0679	nagE	98.5	0.6
b0680	glnS	98.2	NA
b0685	ybfE	96.6	0.9
b0686	ybfF	98.3	0.4
b0689	ybfP	93.9	NA
b0699	ybfA	94.0	0.5
b0709	dtpD	98.5	0.2
b0715	abrB	96.2	0.3
b0719	ybgD	97.3	0.5
b0720	gltA	99.3	NA
b0730	mngR	93.1	NA
b0735	ybgE	99.5	NA
b0752	zitB	96.5	NA
b0753	ybgS	98.2	0.7
b0754	aroG	99.1	NA
b0755	gpmA	98.9	0.4
b0762	acrZ	94.7	0.7
b0766	ybhA	98.9	0.3
b0767	pgl	99.5	0.8
b0768	ybhD	92.8	0.3
b0770	ybhI	94.1	NA
b0771	ybhJ	92.1	NA
b0772	ybhC	98.9	0.2
b0773	ybhB	98.0	0.6
b0774	bioA	97.4	0.2
b0779	uvrB	97.2	NA
b0780	ybhK	96.2	0.3
b0786	ybhL	97.9	0.4
b0791	ybhQ	97.1	0.7
b0797	rhlE	98.1	0.2
b0798	ybiA	95.5	0.6
b0799	dinG	95.8	NA
b0901	ycaK	98.4	0.5
b0902	pflA	97.4	0.4
b0905	ycaO	98.6	NA
b0906	ycaP	98.1	0.4
b0909	ycaL	96.7	0.4
b0919	ycbJ	94.9	0.3
b0920	elyC	98.0	0.4
b0925	ldtD	97.9	0.2
b0928	aspC	99.2	NA
b0929	ompF	99.5	NA
b0930	asnS	98.1	NA
b0931	pncB	96.2	0.2
b0932	pepN	98.3	NA
b0938	elfA	97.2	0.5

b0808	ybiO	95.5	0.1
b0812	dps	99.2	0.6
b0813	rhtA	97.2	0.3
b0814	ompX	99.1	0.5
b0815	opgE	97.3	0.2
b0819	ldtB	98.7	0.3
b0820	ybiT	98.5	0.2
b0821	ybiU	98.0	NA
b0822	ybiV	97.3	0.3
b0825	fsaA	98.5	0.4
b0835	rimO	98.3	0.2
b0836	bssR	93.8	NA
b0837	yliI	96.4	NA
b0838	gstB	99.1	0.4
b0839	dacC	98.6	0.2
b0840	deoR	94.3	0.4
b0841	ybjG	92.1	NA
b0842	mdfA	98.0	NA
b0843	ybjH	98.3	1.0
b0846	rcdA	95.8	0.5
b0847	ybjL	96.2	NA
b0848	ybjM	94.0	0.7
b0849	grxA	98.2	1.1
b0860	artJ	99.2	0.4
b0865	ybjP	99.1	0.5
b0868	ybjS	98.3	0.3
b0874	ybjE	98.0	NA
b0875	aqpZ	97.9	0.4
b0876	ybjD	97.7	0.2
b0877	ybjX	97.5	0.3
b0880	cspD	97.9	1.2
b0881	clpS	96.3	0.9
b0882	clpA	98.5	NA
b0884	infA	96.8	NA
b0885	aat	96.7	0.4
b0888	trxB	98.8	0.3
b0889	lrp	98.7	0.6
b0890	ftsK	96.9	0.1
b0893	serS	97.1	0.2
b0897	ycaC	99.3	0.4
b0898	ycaD	97.5	NA
b0899	ycaM	92.7	0.2
b0900	ycaN	96.7	0.3
b1050	yceK	94.1	0.5
b1051	msyB	99.5	0.7
b1053	mdtG	96.6	0.2
b1054	lpxL	96.0	0.3
b1055	yceA	98.5	0.3
b1060	bssS	87.7	NA
b1061	dinI	94.8	1.1
b1062	pyrC	99.1	NA
b1063	yceB	98.2	0.5
b1064	grxB	99.0	0.4
b1065	mdtH	96.8	NA
b1066	rimJ	98.7	NA
b1067	yceH	99.3	NA
b1068	yceM	99.2	NA

b0945	pyrD	98.0	NA
b0946	zapC	93.5	NA
b0947	ycbX	98.1	0.3
b0952	pqiC	99.2	NA
b0953	rmf	97.0	1.7
b0954	fabA	98.0	0.5
b0955	ycbZ	97.6	NA
b0956	matP	97.5	0.6
b0957	ompA	99.3	0.3
b0958	sulA	95.1	0.5
b0959	sxy	94.8	NA
b0962	helD	98.2	0.1
b0963	mgsA	98.4	0.6
b0964	yccT	96.7	0.4
b0965	yccU	99.0	0.7
b0966	hspQ	96.3	0.9
b0967	rlmI	98.2	0.2
b0968	yccX	98.2	1.0
b0969	tusE	97.8	0.8
b0970	yccA	97.4	0.4
b0987	gfcA	97.5	0.9
b0989	cspH	88.3	NA
b0990	cspG	97.5	NA
b0992	yccM	95.2	NA
b0993	torS	97.4	NA
b0994	torT	96.5	0.3
b0995	torR	97.4	0.4
b1002	agp	98.8	NA
b1013	rutR	96.5	0.4
b1014	putA	98.5	0.3
b1015	putP	99.0	NA
b1020	phoH	99.0	0.3
b1025	ycdT	93.1	NA
b1033	ghrA	98.9	NA
b1036	ycdZ	97.2	0.6
b1044	ymdA	99.2	0.9
b1047	opgC	94.4	0.2
b1192	ldcA	98.5	0.3
b1193	emtA	98.4	0.5
b1194	ycgR	98.2	0.4
b1195	ymgE	98.0	1.1
b1196	ycgY	95.5	0.2
b1197	treA	98.8	NA
b1201	dhaR	97.1	0.1
b1202	ycgV	94.3	NA
b1205	ychH	96.4	1.0
b1206	dauA	96.9	NA
b1216	chaA	97.1	0.3
b1219	ychN	98.3	0.8
b1220	ychO	96.1	0.2
b1223	narK	90.4	NA
b1232	purU	98.0	0.3
b1233	ychJ	96.4	0.6
b1236	galU	98.9	0.3
b1237	hns	97.4	0.7
b1238	tdk	97.4	0.5
b1241	adhE	99.1	0.4

b1069	murJ	98.0	0.2
b1084	rne	98.7	0.1
b1086	rluC	96.5	0.3
b1087	yceF	98.1	0.5
b1101	ptsG	99.2	0.8
b1102	fhuE	98.1	NA
b1109	ndh	98.3	0.2
b1110	ycfJ	97.8	0.5
b1111	ycfQ	97.3	0.4
b1112	bhsA	98.7	1.1
b1113	ldtC	98.3	0.3
b1114	mfd	99.0	NA
b1127	pepT	96.3	NA
b1128	roxA	98.7	0.2
b1133	mnmA	95.9	0.3
b1136	icd	99.4	NA
b1139	lit	98.5	NA
b1143	ymfI	98.5	0.8
b1144	ymfJ	96.4	0.9
b1145	cohE	97.2	0.4
b1158	pinE	99.1	NA
b1159	mcrA	95.3	0.3
b1161	ycgX	96.7	0.7
b1162	bluR	94.4	0.4
b1163	bluF	96.0	0.2
b1168	ycgG	91.4	0.1
b1177	ycgJ	98.7	0.8
b1178	pliG	99.4	0.7
b1179	ycgL	98.8	0.9
b1180	ycgM	99.0	0.4
b1181	ycgN	97.7	0.6
b1182	hlyE	96.5	NA
b1185	dsbB	98.2	0.5
b1186	nhaB	98.3	NA
b1187	fadR	95.9	0.4
b1188	ycgB	97.6	0.2
b1191	cvrA	94.0	NA
b1330	ynaI	97.6	NA
b1332	ynaJ	97.9	1.1
b1333	uspE	98.4	0.3
b1334	fnr	96.6	0.4
b1339	abgR	96.3	0.3
b1340	smrA	98.0	0.5
b1341	ydaM	97.9	NA
b1342	ydaN	97.6	0.3
b1344	ttcA	97.8	0.3
b1356	racR	90.7	NA
b1363	trkG	95.1	NA
b1370	insH1	97.7	0.3
b1376	uspF	97.9	0.6
b1379	hslJ	97.7	0.7
b1380	ldhA	98.0	0.3
b1384	feaR	91.8	0.3
b1385	feaB	97.2	NA
b1386	tynA	96.2	0.1
b1387	paaZ	96.4	NA
b1406	ydbC	98.5	NA

b1242	ychE	96.3	NA
b1250	kch	96.7	0.2
b1251	yciI	98.7	0.9
b1252	tonB	96.8	0.4
b1253	yciA	94.2	NA
b1256	ompW	97.8	0.4
b1269	rluB	97.9	0.3
b1272	sohB	98.1	0.3
b1273	yciN	97.1	1.1
b1274	topA	98.1	0.1
b1275	cysB	98.3	NA
b1276	acnA	99.1	0.4
b1277	ribA	97.6	0.5
b1278	pgpB	95.0	0.4
b1283	osmB	99.2	1.3
b1285	gmr	98.4	0.1
b1286	rnb	99.0	NA
b1287	yciW	98.4	0.2
b1288	fabI	98.7	0.4
b1289	ycjD	97.8	0.8
b1295	ymjA	95.8	1.1
b1296	puuP	97.7	NA
b1297	puuA	98.7	NA
b1303	pspF	96.6	0.3
b1319	ompG	92.3	0.3
b1320	ycjW	96.3	0.3
b1324	tpx	99.4	0.5
b1325	ycjG	98.6	0.3
b1327	ycjY	99.0	0.9
b1328	pgrR	95.4	NA
b1329	mppA	97.7	0.2
b1461	pptA	98.4	1.2
b1462	yddH	97.5	0.5
b1463	nhoA	98.2	0.3
b1464	yddE	97.7	0.3
b1469	narU	96.2	0.2
b1473	yddG	96.2	0.3
b1477	yddM	96.9	NA
b1478	adhP	98.3	NA
b1479	maeA	99.0	NA
b1482	osmC	98.0	0.6
b1491	yddW	96.5	0.2
b1494	pqqL	98.5	0.4
b1498	ydeN	98.4	NA
b1501	ydeP	98.4	NA
b1502	ydeQ	96.3	0.3
b1520	yneE	95.7	0.3
b1521	uxaB	96.0	NA
b1522	yneF	94.0	NA
b1525	sad	98.5	NA
b1526	yneJ	98.0	0.3
b1527	yneK	94.1	NA
b1528	ydeA	96.6	NA
b1529	marC	96.7	0.4
b1533	eamA	98.2	0.3
b1534	ydeE	97.5	0.2
b1535	dgcZ	93.9	0.3

b1407	ydbD	94.9	NA
b1412	azoR	99.1	0.5
b1413	hrpA	97.7	0.1
b1414	ydcF	98.5	0.3
b1415	aldA	98.8	0.8
b1418	cybB	98.6	0.5
b1419	ydcA	97.9	1.6
b1420	mokB	66.2	NA
b1421	trg	96.9	NA
b1422	ydcI	97.7	0.3
b1423	ydcJ	98.1	NA
b1424	opgD	97.5	0.2
b1426	ydcH	96.8	1.2
b1427	rimL	93.2	NA
b1428	ydcK	97.7	0.3
b1431	ydcL	98.6	0.4
b1432	insQ	88.4	NA
b1433	ydcO	97.4	NA
b1434	ydcN	96.7	0.5
b1435	ydcP	96.6	0.1
b1436	yncJ	96.7	1.2
b1439	ydcR	96.9	0.2
b1444	patD	98.6	0.8
b1445	ydcX	97.6	1.6
b1446	ydcY	98.3	1.2
b1449	curA	99.0	NA
b1450	mcbR	96.2	0.4
b1451	yncD	98.5	0.1
b1452	yncE	99.1	0.3
b1453	ansP	95.8	NA
b1454	yncG	96.4	0.5
b1604	ydgH	98.6	0.3
b1607	ydgC	96.1	0.8
b1610	tus	95.3	0.3
b1613	manA	98.7	0.2
b1614	ydgA	98.7	NA
b1618	uidR	95.4	0.5
b1619	hdhA	99.2	0.4
b1620	mall	96.3	0.3
b1623	add	97.5	NA
b1624	ydgJ	97.6	0.3
b1625	cnu	97.9	1.3
b1634	dtpA	93.8	NA
b1635	gstA	99.2	0.5
b1639	mliC	97.0	0.8
b1640	anmK	97.1	0.3
b1641	slyB	98.5	0.6
b1642	slyA	94.3	0.6
b1646	sodC	98.6	0.5
b1647	ydhF	98.5	NA
b1648	ydhL	96.5	0.5
b1651	gloA	99.4	0.7
b1654	grxD	98.2	0.8
b1655	mepH	98.0	0.3
b1656	sodB	99.4	0.5
b1657	ydhP	97.8	0.2
b1658	purR	97.7	0.3

b1536	ydeI	98.7	0.7
b1537	ydeJ	97.3	0.5
b1538	dcp	99.2	NA
b1539	ydfG	99.0	0.4
b1540	rspR	94.4	0.4
b1541	ydfZ	94.0	NA
b1542	ydfI	96.6	NA
b1550	gnsB	98.3	1.6
b1551	ynfN	94.3	1.8
b1552	cspI	94.6	0.5
b1557	cspB	95.1	0.5
b1558	cspF	93.1	NA
b1561	rem	95.8	1.1
b1565	ydfV	93.9	0.9
b1566	flxA	96.8	0.8
b1570	dicA	96.0	0.7
b1582	ynfA	96.3	0.9
b1585	ynfC	98.0	0.4
b1586	ynfD	98.2	0.9
b1592	clcB	95.7	NA
b1595	ynfL	97.5	NA
b1596	ynfM	96.0	0.2
b1597	asr	97.6	0.9
b1598	ydgD	97.4	0.3
b1601	tqsA	92.1	NA
b1726	yniB	93.0	NA
b1727	yniC	97.2	0.4
b1728	ydjM	92.8	NA
b1729	ydjN	98.9	0.8
b1731	cedA	94.5	1.1
b1732	katE	99.4	NA
b1739	osmE	99.3	0.8
b1740	nadE	99.3	0.3
b1741	cho	92.2	NA
b1742	ves	96.4	0.5
b1743	spy	95.6	0.6
b1749	xthA	98.7	0.3
b1757	ynjE	96.8	0.2
b1758	ynjF	95.0	0.4
b1759	nudG	98.8	0.7
b1760	ynjH	98.2	1.0
b1761	gdhA	99.5	NA
b1762	ynjI	93.6	NA
b1766	sppA	96.9	NA
b1769	ydjE	94.5	NA
b1770	ydjF	94.6	0.4
b1777	yeaC	98.0	1.0
b1778	msrB	96.9	0.7
b1781	yeaE	98.5	0.3
b1782	mipA	98.9	0.4
b1785	yeaI	91.6	0.2
b1786	yeaJ	96.0	NA
b1787	yeaK	97.0	0.6
b1788	yoaI	97.3	NA
b1789	yeaL	97.9	0.6
b1790	yeaM	94.8	0.3
b1791	yeaN	96.8	NA

b1659	ydhB	97.5	0.3
b1660	ydhC	95.8	0.2
b1661	cfa	95.7	0.2
b1662	ribC	98.2	0.4
b1663	mdtK	98.6	0.2
b1664	ydhQ	98.6	NA
b1667	ydhR	97.5	0.9
b1668	ydhS	98.0	NA
b1675	ydhZ	94.9	1.3
b1676	pykF	99.3	NA
b1677	lpp	97.6	1.2
b1678	ldtE	99.3	NA
b1688	ydiK	98.6	NA
b1696	ydiP	97.2	0.3
b1702	ppsA	99.1	NA
b1703	ppsR	98.5	NA
b1704	aroH	98.1	0.3
b1705	ydiE	92.5	1.5
b1706	ydiU	97.3	NA
b1707	ydiV	93.5	0.4
b1708	nlpC	98.5	0.6
b1722	ydiY	98.2	0.4
b1723	pfkB	99.1	NA
b1724	ydiZ	98.5	1.0
b1725	yniA	98.0	0.3
b1825	yebO	93.9	1.0
b1826	mgrB	92.6	NA
b1827	kdgR	97.9	0.4
b1828	yebQ	95.9	NA
b1829	htpX	99.2	NA
b1832	msrC	98.4	0.6
b1835	rsmF	96.4	0.2
b1836	yebV	98.9	1.2
b1837	yebW	90.4	NA
b1838	pphA	93.0	0.4
b1842	holE	97.5	1.2
b1845	ptrB	97.6	NA
b1846	yebE	97.2	0.4
b1847	yebF	97.2	0.8
b1848	yebG	95.8	1.0
b1849	purT	99.3	0.2
b1852	zwf	98.7	NA
b1853	yebK	97.2	0.3
b1854	pykA	98.7	NA
b1855	lpxM	95.9	0.3
b1856	mepM	97.4	NA
b1857	znuA	96.1	0.3
b1866	aspS	98.8	NA
b1874	cutC	94.4	0.4
b1875	yecM	97.0	0.5
b1876	argS	98.8	0.2
b1877	yecT	96.5	0.6
b1895	uspC	97.7	0.6
b1902	ftnB	98.2	0.6
b1904	yecR	97.5	0.9
b1905	ftnA	99.0	0.6
b1906	yecH	93.4	1.2

b1792	yeaO	98.0	0.8
b1793	yoaF	98.4	1.1
b1794	yeaP	94.9	0.3
b1795	yeaQ	99.2	1.1
b1798	leuE	94.6	0.4
b1799	dmlR	97.5	0.3
b1800	dmlA	98.2	NA
b1803	yeaX	97.5	NA
b1804	rnd	96.5	0.2
b1808	yoaA	95.5	0.1
b1809	yoaB	98.9	0.8
b1810	yoaC	97.7	0.9
b1811	yoaH	95.6	1.5
b1814	sdaA	98.5	0.2
b1815	yoaD	95.8	0.2
b1816	yoaE	98.8	NA
b1820	yobD	98.5	NA
b1821	mntP	98.4	0.5
b1822	rlmA	97.8	0.3
b1959	yedA	92.2	0.3
b1967	hchA	99.3	NA
b1970	hiuH	98.6	0.7
b1973	zinT	96.0	0.4
b1974	yodB	98.1	0.5
b1976	mtfA	97.4	0.3
b1978	yeeJ	97.0	NA
b1981	shiA	97.0	0.2
b1982	amn	99.5	NA
b1983	yeeN	98.9	0.4
b1985	yeeO	96.3	NA
b1987	cbl	98.4	0.3
b1988	nac	97.9	0.3
b1990	ldtA	97.6	0.3
b1994	insH1	95.7	0.3
b1999	yeeP	95.3	NA
b2000	flu	99.6	0.4
b2007	yeeX	98.1	0.8
b2008	yeeA	95.6	0.3
b2009	sbmC	97.1	0.6
b2010	dacD	95.5	0.2
b2011	sbcB	97.8	0.2
b2014	plaP	98.9	NA
b2015	yeeY	98.1	0.3
b2016	yeeZ	98.0	0.3
b2027	wzzB	96.6	0.3
b2028	ugd	95.1	0.2
b2029	gnd	99.4	NA
b2042	wcaN	98.7	0.3
b2063	yegH	98.9	NA
b2064	asmA	98.0	NA
b2067	yegE	94.4	NA
b2068	alkA	97.4	0.3
b2069	yegD	99.0	0.2
b2070	yegI	97.1	0.1
b2071	yegJ	96.2	0.6
b2080	yegP	99.0	0.8
b2081	yegQ	99.1	0.2

b1907	tyrP	98.2	0.9
b1908	yecA	98.6	0.4
b1912	pgsA	96.8	0.5
b1915	yecF	95.0	1.2
b1916	sdiA	97.2	0.4
b1923	fliC	99.2	NA
b1927	amyA	96.7	NA
b1928	yedD	98.7	0.7
b1931	yedK	95.7	0.4
b1932	yedL	98.7	0.6
b1936	intG	94.7	NA
b1937	fliE	97.6	0.9
b1951	rcsA	98.3	0.4
b1952	dsrB	96.9	1.5
b1953	yodD	97.9	1.2
b1955	yedP	98.7	0.3
b1956	yedQ	96.1	NA
b1957	yodC	98.4	1.5
b1958	yedI	98.0	0.3
b2133	dld	99.4	0.5
b2134	pbpG	98.0	0.3
b2135	yohC	97.5	0.5
b2136	yohD	95.6	0.5
b2137	yohF	98.2	0.4
b2139	mdtQ	95.8	NA
b2140	dusC	98.1	0.3
b2143	cdd	96.7	0.3
b2151	galS	96.7	0.3
b2154	yeiG	98.9	NA
b2155	cirA	98.6	NA
b2156	lysP	98.3	NA
b2157	yeiE	96.3	0.3
b2158	yeiH	97.7	0.3
b2159	nfo	98.4	0.3
b2160	yeiI	98.3	NA
b2161	nupX	96.5	NA
b2162	rihB	96.7	0.3
b2164	psuT	93.9	NA
b2170	setB	95.2	NA
b2171	yeiP	98.5	0.5
b2172	yeiQ	95.6	NA
b2175	mepS	93.7	0.5
b2176	rtn	93.9	NA
b2181	yejG	97.5	0.8
b2182	bcr	97.4	0.2
b2183	rsuA	96.0	0.4
b2184	yejH	96.5	0.2
b2185	rplY	98.9	1.0
b2186	yejK	98.9	0.3
b2190	yejO	97.0	NA
b2193	narP	96.7	0.4
b2209	eco	98.5	0.6
b2210	mqo	98.7	NA
b2211	yojI	97.9	0.2
b2214	apbE	96.2	0.3
b2215	ompC	99.4	NA
b2218	rcsC	97.4	0.1

b2082	ogrK	97.0	1.3
b2086	yegS	95.7	0.3
b2097	fbaB	99.2	NA
b2101	yegW	96.4	0.4
b2102	yegX	94.8	NA
b2105	rcnR	92.8	NA
b2112	yehE	98.2	1.0
b2113	mrp	95.8	0.3
b2114	metG	98.7	NA
b2123	yehR	95.3	0.6
b2124	yehS	95.7	NA
b2127	mlrA	96.3	0.4
b2132	bglX	98.2	NA
b2270	yfbK	94.8	NA
b2272	yfbM	96.5	0.6
b2273	yfbN	96.8	0.4
b2274	yfbO	98.4	0.7
b2275	yfbP	95.8	NA
b2289	lrhA	95.4	0.3
b2292	yfbS	94.9	NA
b2295	yfbV	97.2	0.6
b2298	yfcC	93.8	0.2
b2299	yfcD	99.1	0.5
b2300	yfcE	98.3	0.5
b2301	yfcF	97.4	NA
b2302	yfcG	98.1	0.4
b2305	yfcI	98.0	0.3
b2316	accD	97.7	0.3
b2321	flk	96.4	0.3
b2322	yfcJ	98.1	NA
b2323	fabB	98.7	0.2
b2324	mnmC	98.3	0.1
b2331	smrB	97.2	0.5
b2340	sixA	97.9	0.6
b2343	yfcZ	92.9	0.4
b2344	fadL	99.0	0.2
b2345	yfdF	97.8	0.3
b2346	mlaA	98.1	0.4
b2347	yfdC	97.7	0.3
b2349	intS	96.7	0.2
b2353	tfaS	97.7	1.0
b2364	dsdC	97.0	0.3
b2371	yfdE	97.5	NA
b2372	yfdV	94.2	NA
b2373	oxc	97.1	NA
b2374	frc	98.3	NA
b2375	yfdX	98.7	NA
b2377	yfdY	97.4	1.1
b2378	lpxP	96.7	0.3
b2379	alaC	98.6	0.2
b2388	glk	97.8	0.3
b2389	yfeO	97.2	0.2
b2390	ypeC	97.8	0.9
b2392	mntH	97.0	0.2
b2393	nupC	98.7	NA
b2395	yfeA	95.5	0.1
b2400	gltX	98.5	0.2

b2231	gyrA	96.0	NA
b2232	ubiG	96.0	0.4
b2233	yfaL	97.6	NA
b2237	inaA	97.0	0.4
b2249	yfaY	98.6	0.2
b2250	yfaZ	98.3	0.5
b2251	nudI	94.8	0.7
b2252	ais	97.9	0.5
b2259	pmrD	91.9	NA
b2266	elaB	99.1	0.9
b2267	elaA	98.2	0.6
b2268	rbn	98.1	0.3
b2269	elaD	92.2	0.7
b2469	narQ	97.1	0.2
b2470	acrD	98.7	0.1
b2473	ypfH	98.5	0.4
b2476	purC	98.8	0.4
b2479	gcvR	98.0	0.5
b2480	bcp	98.3	0.6
b2493	yfgO	98.4	0.3
b2496	hda	92.9	NA
b2503	yfgF	93.5	NA
b2504	yfgG	95.0	0.6
b2509	xseA	94.4	0.2
b2510	yfgJ	98.7	NA
b2513	yfgM	98.2	NA
b2514	hisS	96.3	NA
b2516	rodZ	96.0	0.3
b2517	rlmN	98.4	NA
b2518	ndk	98.6	0.6
b2521	sseA	97.4	0.3
b2522	sseB	98.2	NA
b2523	pepB	99.6	NA
b2532	trmJ	97.0	0.4
b2533	suhB	96.2	0.3
b2534	yfhR	96.2	NA
b2535	csiE	96.1	NA
b2536	hcaT	98.2	0.2
b2543	yphA	97.1	0.7
b2549	yphG	97.4	NA
b2550	yphH	96.9	0.2
b2551	glyA	99.3	NA
b2552	hmp	99.2	NA
b2555	yfhG	96.8	NA
b2556	glrK	95.3	NA
b2557	purL	99.0	NA
b2558	mltF	94.1	NA
b2561	yfhH	96.8	0.3
b2562	yfhL	93.9	NA
b2574	nadB	97.9	NA
b2575	trmN	97.6	0.4
b2576	srmB	98.2	0.2
b2577	yfiE	96.3	0.3
b2578	eamB	92.4	0.5
b2579	grcA	97.4	0.7
b2580	ung	98.9	0.4
b2581	yfiF	98.3	0.3

b2405	xapR	94.8	0.3
b2408	yfeN	95.8	NA
b2409	yfeR	97.8	0.3
b2410	yfeH	95.4	NA
b2412	zipA	97.6	NA
b2413	cysZ	97.7	0.4
b2414	cysK	99.3	NA
b2602	yfiL	96.5	0.8
b2610	ffh	96.5	0.2
b2611	ypjD	96.7	0.4
b2614	grpE	98.0	0.5
b2615	nadK	96.6	0.3
b2616	recN	94.4	0.2
b2617	bamE	98.1	0.8
b2620	smpB	95.4	0.6
b2622	intA	97.7	0.2
b2623	yfjH	94.6	0.3
b2624	alpA	98.9	1.3
b2626	yfjJ	96.2	0.4
b2629	yfjM	96.4	NA
b2642	yfjW	98.8	NA
b2647	ypjA	96.5	0.1
b2664	csiR	97.8	0.4
b2665	ygaU	99.1	0.6
b2666	yqaE	97.2	1.8
b2669	stpA	98.9	0.7
b2670	alaE	96.9	0.6
b2671	ygaC	96.6	0.8
b2672	ygaM	98.4	0.8
b2687	luxS	98.4	0.5
b2688	gshA	97.7	NA
b2696	csrA	97.6	1.5
b2697	alaS	98.6	NA
b2700	pncC	98.3	0.6
b2701	mltB	98.4	0.3
b2709	norR	98.7	NA
b2714	ascG	98.0	0.3
b2732	ygbA	93.4	0.8
b2733	mutS	97.7	0.1
b2734	pphB	91.9	0.2
b2735	ygbI	94.1	0.4
b2740	ygbN	93.5	NA
b2745	truD	98.5	0.3
b2748	ftsB	94.3	NA
b2749	ygbE	94.8	0.3
b2753	iap	98.0	0.3
b2761	ygcB	95.5	0.1
b2765	queD	97.9	0.8
b2771	ygcS	95.3	NA
b2777	queE	94.6	0.4
b2785	rlmD	96.8	0.2
b2786	barA	95.4	0.1
b2793	syd	92.2	0.5
b2794	queF	98.0	0.3
b2795	ygdH	98.3	0.2
b2798	ygdG	95.7	0.4
b2806	rlmM	98.4	0.3

b2582	trxC	98.5	0.7
b2585	pssA	97.8	0.2
b2586	yfiM	96.0	0.9
b2587	kgtP	96.6	NA
b2592	clpB	99.1	NA
b2595	bamD	96.9	0.4
b2597	raiA	97.3	0.8
b2808	gcvA	95.5	NA
b2809	ygdI	97.4	1.2
b2812	tcdA	97.3	0.3
b2813	mltA	97.3	0.3
b2817	amiC	96.2	0.2
b2818	argA	99.0	0.2
b2831	mutH	93.4	NA
b2832	ygdQ	98.0	0.4
b2833	ygdR	95.1	NA
b2834	tas	98.9	0.3
b2837	galR	97.0	0.3
b2838	lysA	99.3	NA
b2839	lysR	96.4	NA
b2840	ygeA	98.7	0.4
b2841	araE	95.9	NA
b2842	kduD	98.2	NA
b2843	kduI	95.4	0.3
b2844	yqeF	98.0	0.2
b2845	yqeG	96.3	0.2
b2863	ygeQ	92.1	0.2
b2865	ygeR	93.8	0.4
b2869	ygeV	95.2	0.2
b2870	ygeW	96.1	0.2
b2871	ygeX	96.5	0.2
b2872	ygeY	96.8	NA
b2873	hyuA	95.8	NA
b2874	yqeA	95.1	0.3
b2875	yqeB	97.6	0.2
b2876	yqeC	95.0	0.4
b2877	mocA	95.6	NA
b2889	idi	94.6	0.5
b2895	fldB	97.4	0.5
b2898	ygfZ	98.9	0.3
b2899	yqfA	94.0	0.4
b2900	yqfB	97.7	0.9
b2901	bglA	98.9	NA
b2902	ygfF	98.6	NA
b2910	zapA	88.3	NA
b2913	serA	99.3	NA
b2914	rpiA	99.1	0.4
b2915	yqfE	95.6	0.9
b2916	argP	96.3	0.3
b2921	ygfI	95.6	NA
b2922	yggE	98.8	0.4
b2923	argO	98.6	0.4
b2924	mscS	98.8	NA
b2935	tktA	99.2	NA
b2936	loiP	98.8	0.4
b2939	yqgB	98.2	NA
b2940	yqgC	92.6	NA

b2807	ygdD	98.2	0.7
b2943	galP	97.9	NA
b2944	yggI	98.6	NA
b2945	endA	97.3	0.4
b2950	yggR	98.3	1.2
b2956	yggM	96.3	0.3
b2957	ansB	97.3	0.3
b2958	yggN	98.5	0.4
b2965	speC	97.7	NA
b2966	yqgA	93.5	0.4
b2971	yghG	98.6	0.7
b2972	pppA	91.2	0.3
b2980	glcC	94.4	0.4
b2981	yghO	93.9	0.3
b2983	yghQ	95.2	0.1
b2987	pitB	96.9	0.2
b2988	gss	97.9	NA
b2989	yghU	98.2	0.3
b2998	yghW	98.9	1.0
b3001	gpr	97.9	NA
b3002	yqhA	96.9	0.6
b3003	yghA	98.7	NA
b3008	metC	99.3	NA
b3009	yghB	94.6	0.4
b3010	yqhC	97.4	0.3
b3013	yqhG	96.8	0.3
b3014	yqhH	97.8	1.1
b3017	ftsP	98.4	0.2
b3018	plsC	98.1	0.4
b3019	parC	97.3	NA
b3020	ygiS	97.4	NA
b3023	ygiV	97.9	0.6
b3024	ygiW	99.0	0.7
b3028	mdaB	98.7	0.5
b3029	ygiN	98.3	0.9
b3039	ygiD	96.5	0.4
b3040	zupT	97.1	0.4
b3042	yqiC	97.3	1.0
b3043	ygiL	97.3	0.5
b3049	glgS	96.6	1.4
b3052	hldE	97.6	0.2
b3057	bacA	98.5	0.3
b3058	folB	98.4	0.8
b3059	plsY	94.7	NA
b3060	ttdR	91.6	NA
b3064	tsaD	98.0	0.3
b3068	mug	98.2	0.5
b3070	yqjH	98.7	0.4
b3071	yqjI	99.4	0.4
b3072	aer	97.2	NA
b3073	patA	98.8	NA
b3075	ebgR	96.8	0.3
b3252	csrD	97.1	NA
b3253	acuI	97.8	0.3
b3262	yhdJ	96.4	NA
b3263	yhdU	91.3	1.5
b3267	yhdV	98.4	1.3

b2942	metK	98.9	NA
b3081	fadH	95.6	0.1
b3084	rlmG	98.1	0.2
b3085	ygjP	96.2	NA
b3086	ygjQ	97.6	0.4
b3087	ygjR	98.0	0.3
b3088	alx	95.9	NA
b3089	sstT	98.6	NA
b3090	ygjV	96.2	NA
b3093	exuT	95.1	NA
b3094	exuR	94.5	0.4
b3101	yqjF	98.1	0.7
b3102	yqjG	98.1	0.3
b3103	yhaH	98.7	0.8
b3105	yhaJ	97.9	0.3
b3128	garD	97.2	NA
b3131	agaR	96.5	0.3
b3142	yraH	96.1	0.5
b3146	rsmI	96.8	0.3
b3147	lpoA	96.7	0.1
b3148	yraN	98.1	NA
b3149	diaA	97.5	0.5
b3150	yraP	98.6	0.5
b3151	yraQ	98.3	0.3
b3152	yraR	98.5	0.4
b3153	yhbO	97.3	0.5
b3154	yhbP	96.5	NA
b3155	yhbQ	95.1	0.9
b3160	yhbW	99.1	0.3
b3161	mtr	97.3	NA
b3163	nlpI	83.6	NA
b3172	argG	99.4	NA
b3173	yhbX	95.0	NA
b3180	yhbY	94.0	0.4
b3181	greA	98.3	0.6
b3182	dacB	97.4	0.2
b3187	ispB	97.9	0.3
b3188	sfsB	98.0	1.0
b3207	yrbL	97.5	0.4
b3210	arcB	98.2	NA
b3211	yhcC	97.1	0.3
b3220	yhcG	90.5	NA
b3226	nanR	94.7	0.4
b3227	dcuD	95.7	NA
b3232	zapE	95.9	0.2
b3233	yhcB	97.1	0.7
b3236	mdh	99.3	NA
b3237	argR	96.4	0.6
b3238	yhcN	98.8	1.1
b3239	yhcO	95.9	1.0
b3243	aaeR	97.2	0.3
b3244	tldD	97.6	0.2
b3470	tusA	96.2	1.1
b3471	yhhQ	98.3	0.4
b3472	dcrB	97.9	0.5
b3473	yhhS	97.7	NA
b3474	yhhT	97.5	0.3

b3279	yrdA	97.2	0.5
b3284	smg	97.1	0.6
b3289	rsmB	98.7	NA
b3290	trkA	98.7	NA
b3291	mscL	98.4	0.7
b3338	chiA	95.1	NA
b3347	fkpA	98.5	0.3
b3348	slyX	97.4	1.3
b3349	slyD	98.2	0.5
b3355	prkB	97.7	0.3
b3356	yhfA	97.2	0.7
b3357	crp	96.1	0.4
b3358	yhfK	96.2	NA
b3359	argD	98.8	NA
b3363	ppiA	98.0	0.5
b3364	tsgA	97.9	NA
b3369	yhfL	92.1	1.7
b3391	hofQ	98.1	0.2
b3396	mrcA	96.5	NA
b3397	nudE	96.7	0.5
b3398	yrfF	95.2	NA
b3402	yhgE	96.8	NA
b3403	pck	97.9	NA
b3406	greB	95.2	0.6
b3407	yhgF	99.5	NA
b3411	yhgA	97.6	0.3
b3412	bioH	98.2	0.4
b3415	gntT	96.0	NA
b3418	malT	97.8	0.1
b3422	rtcR	96.3	NA
b3426	glpD	94.4	NA
b3427	yzgL	91.3	NA
b3433	asd	99.2	NA
b3434	yhgN	98.0	0.5
b3439	yhhW	98.4	0.4
b3440	yhhX	98.8	NA
b3441	yhhY	98.3	0.6
b3446	yrhB	96.5	1.0
b3447	ggt	99.2	NA
b3448	yhhA	98.2	0.6
b3459	panM	98.4	0.7
b3460	livJ	99.6	NA
b3461	rpoH	95.7	NA
b3467	yhhM	97.6	NA
b3468	yhhN	97.7	0.4
b3469	zntA	98.7	NA
b3588	aldB	97.3	0.7
b3589	yiaY	92.5	1.0
b3592	yibF	98.1	0.5
b3595	rhsJ	92.5	0.3
b3602	yibL	96.7	0.8
b3606	trmL	97.9	0.6
b3607	cysE	98.0	NA
b3610	grxC	98.8	1.1
b3611	yibN	97.9	0.6
b3615	waaH	95.1	NA
b3618	yibB	89.3	0.3

b3475	acpT	96.2	NA
b3482	rhsB	97.5	NA
b3488	yhiJ	90.8	0.2
b3491	yhiM	96.5	0.3
b3492	yhiN	96.7	0.2
b3493	pitA	95.6	0.2
b3494	uspB	96.8	0.8
b3495	uspA	99.1	0.6
b3496	dtpB	97.2	NA
b3499	rlmJ	98.1	NA
b3500	gor	98.3	NA
b3511	hdeD	98.4	NA
b3518	yhjA	91.6	0.2
b3519	treF	97.7	0.2
b3520	yhjB	98.4	0.5
b3521	yhjC	98.6	0.3
b3522	yhjD	97.3	0.3
b3523	yhjE	99.4	0.9
b3524	yhjG	97.6	NA
b3525	yhjH	96.8	0.4
b3526	kdgK	98.4	0.3
b3527	yhjJ	97.9	0.2
b3528	dctA	98.6	0.2
b3529	yhjK	95.1	0.1
b3534	bcsQ	95.9	0.4
b3535	yhjR	94.7	1.5
b3539	yhjV	95.5	NA
b3546	eptB	95.7	NA
b3547	yhjX	95.2	0.2
b3548	yhjY	98.5	0.4
b3551	bisC	97.4	NA
b3552	yiaD	98.8	0.4
b3553	ghrB	99.2	NA
b3554	yiaF	97.2	0.4
b3555	yiaG	98.8	1.0
b3556	cspA	96.3	NA
b3557	insJ	96.3	0.5
b3558	insK	96.7	NA
b3561	wecH	97.5	0.3
b3570	bax	90.6	NA
b3571	malS	96.6	0.1
b3572	avtA	98.7	0.2
b3573	ysaA	96.8	0.6
b3574	yiaJ	96.9	0.3
b3584	yiaT	96.2	NA
b3585	yiaU	90.6	0.1
b3775	ppiC	96.4	1.0
b3778	rep	98.2	0.1
b3779	gpp	96.2	0.2
b3780	rhlB	96.0	0.2
b3781	trxA	96.2	0.8
b3795	yifK	98.5	NA
b3800	aslB	96.2	NA
b3801	aslA	94.0	NA
b3806	cyaA	94.1	NA
b3807	cyaY	97.2	NA
b3813	uvrD	97.9	0.1

b3639	dfp	98.2	0.2
b3644	yicC	98.4	0.3
b3645	dinD	95.1	0.3
b3646	yicG	98.1	0.5
b3647	ligB	97.1	NA
b3648	gmk	93.5	0.4
b3653	gltS	97.7	0.2
b3654	xanP	98.1	0.2
b3655	yicH	98.9	NA
b3659	setC	94.0	NA
b3660	yicL	95.5	NA
b3661	nlpA	99.3	0.3
b3662	nepI	95.9	NA
b3663	yicN	96.3	0.6
b3664	adeQ	95.4	0.9
b3665	adeD	99.0	NA
b3666	uhpT	95.5	NA
b3673	emrD	95.5	0.2
b3677	yidI	95.5	0.6
b3680	yidL	95.3	0.3
b3684	yidP	96.8	NA
b3685	yidE	96.8	NA
b3688	yidQ	97.9	0.8
b3689	yidR	99.2	0.2
b3690	cbrA	96.4	0.3
b3696	yidX	89.0	0.2
b3697	yidA	99.0	0.3
b3698	yidB	98.2	0.7
b3699	gyrB	96.8	NA
b3706	mnmE	97.2	NA
b3710	mdtL	97.4	0.2
b3711	yidZ	94.8	NA
b3714	adeP	97.6	0.2
b3715	yieH	95.6	0.4
b3744	asnA	98.7	NA
b3747	kup	95.2	NA
b3764	yifE	95.0	0.3
b3765	yifB	94.1	NA
b3773	ilvY	96.5	0.3
b3774	ilvC	99.6	NA
b3924	fpr	97.4	0.4
b3928	zapB	98.7	1.1
b3934	cytR	96.2	0.3
b3935	priA	96.9	0.1
b3936	rpmE	98.9	1.3
b3937	yiiX	95.2	NA
b3938	metJ	94.9	NA
b3941	metF	98.5	0.3
b3942	katG	98.9	0.5
b3943	yijE	95.0	NA
b3944	yijF	95.9	NA
b3953	frwD	98.3	0.8
b3954	yijO	95.4	NA
b3955	eptC	98.6	0.7
b3956	ppc	99.4	0.4
b3957	argE	99.1	NA
b3961	oxyR	97.6	0.3

b3816	corA	97.9	0.3
b3819	rarD	95.8	NA
b3820	yigI	97.9	0.6
b3821	pldA	97.3	0.3
b3822	recQ	96.5	0.2
b3823	rhtC	97.8	0.4
b3824	rhtB	96.9	0.4
b3827	yigM	96.2	0.3
b3828	metR	99.4	0.3
b3829	metE	99.6	0.5
b3830	ysgA	99.2	0.3
b3831	udp	98.2	0.4
b3832	rmuC	95.5	0.2
b3842	rfaH	94.0	NA
b3843	ubiD	96.5	0.2
b3844	fre	98.6	0.4
b3858	yihD	98.6	1.0
b3861	yihF	95.7	NA
b3862	yihG	94.4	NA
b3863	polA	98.9	NA
b3865	yihA	98.3	0.4
b3866	yihI	98.6	0.5
b3867	hemN	98.2	0.2
b3871	typA	98.4	0.2
b3874	yihN	96.6	NA
b3878	yihQ	98.9	NA
b3890	yiiF	97.9	1.3
b3895	fdhD	95.7	0.3
b3896	yiiG	97.6	0.3
b3907	rhaT	95.6	NA
b3908	sodA	99.3	0.4
b3909	kdgT	96.0	0.3
b3910	yiiM	97.9	0.4
b3915	fieF	96.4	0.3
b3916	pfkA	98.5	0.3
b3917	sbp	99.1	NA
b3918	cdh	97.9	0.4
b3919	tpiA	97.7	0.4
b3920	yiiQ	98.3	0.5
b3921	yiiR	97.8	0.6
b4059	ssb	97.3	0.5
b4060	yjcB	95.7	1.0
b4061	yjcC	93.9	NA
b4062	soxS	94.9	0.9
b4063	soxR	96.4	0.6
b4064	ghxP	98.1	NA
b4065	yjcE	98.5	0.7
b4077	gltP	98.9	0.9
b4078	yjcO	97.0	0.4
b4079	fdhF	90.3	NA
b4083	yjcS	95.8	NA
b4084	alsK	97.4	0.3
b4090	rpiB	95.2	0.6
b4107	yjdN	97.8	0.6
b4108	yjdM	95.2	0.8
b4111	proP	97.2	NA
b4114	eptA	97.6	0.2

b3962	sthA	96.9	0.2
b3965	trmA	96.5	0.3
b3974	coaA	96.9	0.3
b3989	yjaZ	91.8	NA
b3995	rsd	98.1	0.6
b3999	yjaG	98.5	0.5
b4000	hupA	99.2	1.0
b4001	yjaH	98.8	0.4
b4002	zraP	99.1	0.7
b4011	yjaA	93.5	NA
b4012	yjaB	97.8	0.6
b4013	metA	98.9	0.3
b4018	iclR	95.2	NA
b4019	metH	99.1	NA
b4020	yjbB	97.1	NA
b4021	pepE	96.9	0.4
b4022	rluF	98.1	0.3
b4023	yjbD	96.8	1.0
b4024	lysC	99.3	NA
b4025	pgi	99.0	0.7
b4030	psiE	94.4	0.7
b4031	xylE	88.9	0.2
b4041	plsB	98.0	0.1
b4042	dgkA	98.2	0.8
b4045	yjbJ	84.9	NA
b4046	zur	93.6	NA
b4049	dusA	97.3	0.3
b4050	pspG	96.1	1.1
b4051	qorA	98.7	0.3
b4052	dnaB	96.7	0.2
b4053	alr	97.9	0.3
b4054	tyrB	99.3	NA
b4055	aphA	93.9	0.4
b4058	uvrA	93.8	NA
b4207	fklB	98.5	0.4
b4208	cycA	98.6	0.8
b4209	ytfE	93.0	0.4
b4210	ytfF	95.3	0.3
b4211	qorB	98.8	NA
b4212	ytfH	96.5	0.7
b4213	cpdB	98.0	NA
b4214	cysQ	98.7	0.4
b4216	ytfJ	98.2	0.5
b4217	ytfK	94.8	1.3
b4218	ytfL	97.1	0.2
b4219	msrA	99.3	0.4
b4226	ppa	98.4	0.5
b4232	fbp	98.7	NA
b4233	mpl	98.6	0.2
b4234	yjgA	96.0	0.5
b4235	pmbA	97.7	0.2
b4236	cybC	96.7	NA
b4241	treR	95.4	0.3
b4243	ridA	99.1	0.7
b4248	yjgH	97.7	0.7
b4249	bdcA	95.7	0.4
b4251	bdcR	98.7	0.5

b4115	adiC	96.1	0.2
b4116	adiY	97.2	0.4
b4117	adiA	94.1	0.1
b4118	melR	93.1	0.3
b4121	yjdF	97.2	0.4
b4129	lysU	97.8	NA
b4130	dtpC	97.3	NA
b4133	cadC	92.9	0.2
b4137	cutA	94.8	0.8
b4140	fxsA	97.4	NA
b4141	yjeH	95.1	0.2
b4144	yjeI	99.1	0.8
b4145	yjeJ	98.0	0.3
b4146	epmB	95.4	0.3
b4147	efp	98.4	0.5
b4148	sugE	97.0	0.9
b4149	blc	99.2	0.5
b4150	ampC	97.5	0.2
b4155	epmA	92.1	0.3
b4156	yjeM	96.4	NA
b4161	rsgA	95.6	0.3
b4162	orn	98.5	0.5
b4166	queG	91.6	NA
b4176	yjeT	95.2	1.4
b4177	purA	97.3	0.2
b4187	aidB	98.6	NA
b4188	yjfN	97.1	1.0
b4189	bsmA	93.6	0.8
b4190	yjfP	99.0	0.4
b4191	ulaR	97.7	0.4
b4192	ulaG	96.4	NA
b4199	yjfY	98.3	1.0
b4204	yjfZ	94.2	0.4
b4206	ytfB	95.3	0.4
b4332	yjiJ	92.9	NA
b4333	yjiK	96.8	NA
b4340	yjiR	98.3	0.2
b4342	yjiT	96.4	0.2
b4347	symE	94.4	0.8
b4351	mrr	97.6	0.3
b4354	yjiY	97.3	0.5
b4355	tsr	99.2	0.7
b4356	lgoT	95.2	NA
b4357	lgoR	98.1	0.3
b4358	lgoD	97.5	NA
b4359	opgB	96.3	0.1
b4364	yjjP	93.1	NA
b4367	fhuF	96.1	NA
b4371	rsmC	97.5	0.3
b4375	prfC	99.0	NA
b4376	osmY	99.5	NA
b4379	yjjW	98.4	NA
b4380	yjjI	95.0	0.2
b4385	yjjJ	98.0	0.2
b4391	ettA	99.0	NA
b4392	slt	98.9	0.1
b4393	trpR	97.0	0.9

b4252	tabA	99.1	0.6
b4253	yjgL	97.6	NA
b4254	argI	99.3	NA
b4255	rraB	97.1	0.7
b4256	yjgM	97.9	0.6
b4257	yjgN	96.4	NA
b4260	pepA	99.0	0.2
b4263	yjgR	98.7	NA
b4268	idnK	97.5	0.5
b4269	ahr	99.3	NA
b4271	intB	97.0	0.2
b4277	yjgZ	96.9	NA
b4278	insG	98.5	NA
b4281	yjhD	89.0	NA
b4282	yjhE	90.7	NA
b4294	insA	89.0	NA
b4295	yjhU	96.3	0.3
b4296	yjhF	94.7	0.2
b4306	yjhP	99.3	NA
b4308	yjhR	94.3	NA
b4309	nanS	99.2	NA
b4312	fimB	98.8	0.5
b4313	fimE	94.9	0.5
b4321	gntP	96.8	NA
b4324	uxuR	96.4	0.4
b4326	iraD	92.5	0.7
b4327	hypT	94.9	NA
b4331	kptA	97.1	0.5
b4520	ymgF	96.3	NA
b4529	ydbJ	98.8	1.0
b4535	yniD	92.0	NA
b4536	yobH	95.1	1.2
b4537	yecJ	97.9	1.1
b4538	yoeF	96.4	NA
b4542	yohO	97.5	1.0
b4550	arfA	98.9	1.3
b4553	ysaB	96.5	NA
b4554	yibT	98.4	1.3
b4555	yicS	98.5	0.9
b4567	yjjZ	93.4	NA
b4568	ytjA	99.1	1.7
b4580	yaiT	91.4	NA
b4582	yoeA	94.5	0.2
b4595	yciY	92.4	0.6
b4598	yncL	97.8	NA
b4599	yneM	95.1	NA
b4600	ydfJ	91.2	0.0
b4606	ypfM	93.9	NA
b4613	dinQ	60.0	1.3
b4618	tisB	84.1	NA
b4660	yhiL	92.9	NA
b4518	ymdF	95.8	NA
b4519	icdC	98.2	NA

b4394	yjjX	97.6	0.5
b4395	ytjC	98.3	0.4
b4396	rob	97.1	0.3
b4401	arcA	96.9	0.4
b4402	yjjY	93.1	2.0
b4403	yjtD	98.3	0.4
b4409	blr	91.1	NA
b4415	hokE	94.1	NA
b4428	hokB	70.5	0.7
b4453	ldrD	68.7	NA
b4461	yfjD	96.7	NA
b4463	ygcU	94.2	NA
b4466	sslE	96.2	NA
b4469	ygiQ	98.5	0.1
b4472	yhdP	96.6	NA
b4473	smf	97.6	0.2
b4480	hdfR	97.2	0.3
b4482	yigE	98.0	0.4
b4484	cpxP	92.4	0.6
b4486	yjiV	95.4	NA
b4487	yjdP	95.8	0.8
b4501	torI	94.9	1.4
b4502	yeiW	92.3	1.1
b4508	peaD	96.7	0.9
b4512	ybdD	97.9	NA
b4515	cydX	99.2	NA

Citation: Nguyen HL, Duviau M-P, Cocaign-
Bousquet M, Nouaille S, Girbal L (2019) Multiplexing polysome profiling experiments to study translation in *Escherichia coli*. PLoS ONE 14(2): e0212297. https://doi.org/10.1371/journal. pone.0212297

Editor: Thomas Preiss, John Curtin School of
Medical Research, AUSTRALIA

Received: November 20, 2018

Accepted: January 30, 2019

Published: February 19, 2019

DataAvailabilityStatement: All relevant data are within the manuscript and its Supporting Information files.

Funding: HLN was supported by a pre-doctoral fellowship from INRA (french national institute for agricultural research). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competinginterests: The authors have declared that no competing interests exist.

RESEARCH ARTICLE

Multiplexing polysome profiling experiments to study translation in *Escherichia coli*

Huong Le Nguyen, Marie-Pierre Duviau, Muriel Cocaign-Bousquet*, Se´bastien Nouaille, Laurence Girbal *

LISBP, Universite´ de Toulouse, CNRS, INRA, INSA, Toulouse, France

* cocaign@insa-toulouse.fr (MCB); girbal@insa-toulouse.fr (LG)

Abstract

Polysome profiling is a widely used method to monitor the translation status of mRNAs. Although it is theoretically a simple technique, it is labor intensive. Repetitive polysome fractionation rapidly generates a large number of samples to be handled in the downstream processes of protein elimination, RNA extraction and quantification. Here, we propose a multiplex polysome profiling experiment in which distinct cellular extracts are pooled before loading on the sucrose gradient for fractionation. We used the multiplexing method to study translation in *E*. *coli*. Multiplexing polysome profiling experiments provided similar mRNA translation status to that obtained with the non-multiplex method with comparable distribution of mRNA copies between the polysome profiling fractions, similar ribosome occupancy and ribosome density. The multiplexing method was used for parallel characterization of gene translational responses to changing mRNA levels. When the mRNA level of two native genes, *cysZ* and *lacZ* was increased by transcription induction, their global translational response was similar, with a higher ribosome load leading to increased ribosome occupancy and ribosome densities. However the pattern and the magnitude of the translational response were gene specific. By reducing the number of

polysome profiling experiments, the multiplexing method saved time and effort and reduced cost and technical bias. This method would be useful to study the translational effect of mRNA sequence-dependent parameters that often require testing multiple samples and conditions in parallel.

Introduction

In the last decade, interest in the role of regulating translation has been increasing. Translation regulation plays an essential role in fine-tuning gene expression and protein level [1]. It allows a rapid response to extracellular stimuli, which can be crucial for adaptation to different environmental conditions [2]. Polysome profiling is a widely used method to study translation status. For each individual gene, the method quantifies the number of ribosomes bound to each copy of the mRNA molecule and provides a detailed distribution of the mRNA copies per number of bound ribosomes (proportions of free mRNA copies, of monosome-bound and polysome-bound mRNA copies [3–5]. Polysome profiling enables the definition of two translational variables, ribosome occupancy (RO) and ribosome density (RD) [6–8]. RO is the proportion of mRNA copies bound with at least one ribosome and can be used as a proxy for translation initiation. RD can be calculated as the number of mRNA-bound ribosomes in each polysome profiling fraction normalized to the coding sequence length; it reflects the level of translation initiation, elongation and termination. Thus, polysome profiling not only assesses heterogeneity in the number of bound ribosomes within the copies of an mRNA, but also accesses physical ribosome density by measuring joint binding of multiple ribosomes on the same transcript. This technique is complementary to the recently expanding method of ribosome profiling that quantifies the heterogeneity of ribosome position occupation averaged over the population of mRNA copies.

The polysome profiling method can be used to study translational status in different organisms at different stages of growth and development (e.g. *Arabidopsis thaliana* [9], in sea urchin [10], in halophilic archaea [3]). It is also used to understand translational response to various stresses: marine organisms under oxidative stress [11], yeast and *A. thaliana* in high salinity conditions [12,13], yeast under Zn limitation [14] and *Lactococcus lactis* and yeast under nutrient starvation [8,15]. It is also frequently used in mechanistic studies of translation regulation, for instance to characterize the effects of elongation factors [16,17]. The role of mRNA sequence related parameters such as 5'UTR and codon usage on translation has also been investigated using polysome profiling [18–20].

These studies are usually limited to a small number of samples and conditions because polysome profiling is labor intensive. The drawbacks of the technique remain the main difficulty in handling many samples in parallel [21]. Separation of

mRNA-polysome complexes according to bound ribosomal loading consists in polysome fractionation on the sucrose gradient. This step generates numerous samples to be handled in the downstream processes of protein elimination, RNA extraction and quantification. This may introduce technical bias between different polysome profiling experiments and also entails rather expensive and time consuming experiments. We developed a multiplexing method for polysome profiling experiments that makes it possible to assemble six different cell free extracts before loading on the sucrose gradient. The RNA in each fraction was quantified by RT-qPCR with an optimized amount of exogenous RNA spike-ins. A challenge in multiplexing experiment is to differentiate the cell free extract origin of the measured mRNAs. In this study, we identified the origin of an mRNA of interest by strongly overexpressing this mRNA in only one cell free extract of the mixture. The multiplexing polysome profiling method was then used to study translation regulation in the bacterial model *Escherichia coli* K12 MG1655. The translational states of different genes were simultaneously characterized by measuring their ribosome occupancies and ribosome densities. We also accessed the translational response of these genes to changing gene expression, a situation that may be encountered when *E*. *coli* cells need to adapt to variable growth environments.

Materials and methods

Plasmids and strain construction

All strains were constructed in the genetic background *E*. *coli* MG1655 Δ*araFGH*, *Opcp18*:: *araE533* [22]. MET734 and MET739 carried the *cysZ* and *lacZ* genes, respectively, on a plasmid under the P_{BAD} inducible promoter (Table 1). We selected these two native genes of *E*. *coli* MG1655 for their low level of expression in exponential growth in synthetic medium [23]. For each, the 5'UTR + ORF fragment was amplified by PCR and cloned into the pBAD/myc/

His plasmid (Invitrogen) to obtain the constructs: pBAD– 5'UTR+$ORF_{selected\ gene}$— myc/His tag. The pBAD-5'UTR+ORF_{lacZ}-myc/His tag was introduced into the *E*. *coli* variant where the chromosomal copy of *lacZ* was deleted [24]. Four other "filling" strains were used to mix their cell free extract with those of MET734 and MET739 for multiplexing purposes. In the same

Table 1. Strain description.

Strain	Description	Source
DCT2202	*E*. *coli* MG1655 Δ*araFGH*, *Opcp18*::*araE533*	[22]
MET345	DCT2202 Δ*lacZ*	[24]
MET739	MET345 with plasmid (pBAD-lacZ-*myc*-his)	This work
MET734	*E*. *coli* DCT2202 with plasmid (pBAD-cysZ-*myc*-his)	This work

https://doi.org/10.1371/journal.pone.0212297.t001

genetic background, the four strains contained genes (*yeeZ*, *inaA*, *ucpA* and *yjc*) not related to this study, under the control of P_{BAD}.

Culture and preparation of cell lysate

Each strain was individually grown in chemically defined minimum medium M9 supplemented with 3 g/L glucose [23], 0.1 mg/mL ampicillin, at 37 ˚C, under shaking (150 rpm). Arabinose was added at a final concentration of 0.001% (w/v) when the culture reached an OD_{600} of 1 (exponential growth). After 30 minutes of induced gene expression, chloramphenicol was added at a final concentration of 0.1 mg/mL to stop translation elongation.

Cell culture was immediately transferred on ice for one minute. A fixed volume corresponding to 320 ml per OD unit was collected and centrifuged at 6,300 g, at 4 ˚C for 15 minutes. The supernatant was discarded and the cell pellet washed twice with cold lysis buffer

(20 mM Tris HCl pH 8, 140 mM KCl, 40 mM $MgCl_2$, 0.5 mM DTT, 100 µg/mL chloramphenicol, 1 mg/mL heparin, 20 mM EGTA, 1% (v/v) Triton X-100) and resuspended in 1.2 mL of cold lysis buffer.

The cell suspension was transferred in cold screw-capped tubes containing 0.1 g of glass beads (0.1 mm diameter, Sigma) and disrupted using a FastPrep-24 (MP Biomedicals). We observed that more RNA was obtained when we performed four 30 s cycles at 6.5 m/s with at least one minute on ice between cycles rather than two cycles. The lysate was first centrifuged for 5 minutes at 2,100 g at 4 ˚C to remove the glass beads. The supernatant was collected and centrifuged again for 5 minutes at 8,600 g, at 4 ˚C. Clarified lysate was gently collected, immediately frozen in liquid nitrogen and stored at -80 ˚C. The concentration of protein in the lysate was measured using a NanoDrop ND-1000 UV spectrophotometer (Nanodrop Technologies). Starting from around 110 mg of dry cell weight, the protein yield was around 50– 60 mg/mL. All the steps were performed at 4 ˚C and the samples were kept on ice.

Polysome profiling experiments

According to ribosome loading, the mRNA-ribosome complexes were separated on 10– 50% (w/v) linear sucrose gradient (24 mL) prepared in cold lysis buffer. In the non- multiplexing experiment, individual cell free extract (2.4 mL) was loaded on the sucrose gradient, whereas in the multiplexing experiment, six cell free extracts were pooled to reach an equivalent protein amount per strain in a final volume of 2.4 mL. Ultracentrifugation was performed in a Sorvall WX80 (ThermoScientific) using a swing rotor AH-629, for 16h30min, at 23,700 g, at 4 ˚C. The sucrose gradient was eluted with cold buffer (55% sucrose (w/v), 0.5 mM Tris HCl pH 8, 4 µg/mL Bromophenol blue) in 24 sub-fractions at 2.5 mL/minute. Absorbance was continuously measured at 254 nm with a UV detector (UPC900 Amersham Pharmacia Biotech).

Protein elimination and RNA extraction

Protein denaturation and nucleic acid precipitation were performed by adding one volume of

8 M guanidium-HCl, two volumes of absolute ethanol and overnight storage at -20˚C. After

30 minutes centrifugation at 13,300 g at 4 ˚C, the supernatant containing the free proteins was gently removed and the pellet of nucleic acids (including mRNA loaded with ribosomes) were washed with cold 75% ethanol (v/v) and resuspended in TE buffer (10 mM Tris HCl pH 8, 1 mM EDTA). Total RNA extraction and purification were performed using the extraction RNeasy Midi kit (Qiagen). Genomic DNA was removed by on-column DNase digestion using 90U of RNase-free DNase I (Qiagen) for 15 minutes at room temperature. The total RNA concentration was measured using NanoDrop ND-1000. RNA integrity was validated and 16S and 23S rRNA were quantified using Bioanalyzer 2100 (Agilent Technologies). Total RNA samples were stored at -80 ˚C.

Reverse transcription and RNA quantification by real-time quantitative PCR

Total RNA (5 µg) was reverse-transcribed to yield cDNA using 200U of SuperScript II reverse transcriptase (Invitrogen) as previously described [25]. cDNA was quantified by Real Time PCR Detection system (Bio-Rad) in 96-well plates as previously described [26]. cDNA dilutions of 10^{-1} and 10^{-2} were used to provide quantifiable signals, i.e. cycle threshold (Ct) of between 15 and 25. For large numbers of samples, a high-throughput qPCR technique was applied using Biomark HD System (Fluidigm Corporation, CA, USA) as previously described

[24].

To account for variability of the reverse transcription and the qPCR steps between samples and experiments, control Ambion ERCC RNA Spike-In mix was used as external normalizer. For each fraction, an equal amount of ERCC was added to a constant amount of total RNA and they were then reverse transcribed together. To improve the efficiency and reproducibility of the reverse transcription of ERCC, 0.2 µM of reverse primers specific to the four most concentrated ERCCs (ERCC 130, ERCC 002, ERCC 074 and ERCC 096) were added during reverse transcription in addition to random primers.

A total of 12 different mRNAs and four ERCCs were quantified in this work (S1 Table). Quantification of *lacZ* mRNA is the average value obtained from five primer pairs. Primers for qPCR were designed for these 20 genes using Vector NTI advance v11 (Life Technologies) using a melting temperature of 59–61 ˚C, length of 18–20 bp and 50–70% GC content (S1 Table). Amplicon sizes ranged between 80 and 150 bp. The reaction efficiency was tested on cDNA serial dilutions and focused around 100%.

Data normalization and analysis

To calculate the relative amount of a target mRNA in each fraction, two normalizations were applied. First, relative mRNA abundance compared to a constant quantity of ERCC was calculated using the method of fold change ΔCt values [27]. As only 5 µg of the total RNA amount extracted in each fraction was used in the RT-qPCR experiment, we normalized the relative mRNA abundance by the total RNA quantity extracted in each fraction to obtain the relative initial mRNA abundance in each fraction.

For each target gene, its relative initial mRNA abundance compared to ERCC in fraction *i* was calculated as follows:

$$\text{Relative initial mRNA abundace} = 2^{\delta Ct\ ERCC_i \quad Ct\ target_i} \times \frac{\text{total RNA quantity}_i}{5}$$

To obtain the distribution of the abundance of mRNA copies, the proportion of mRNA copies in each fraction was calculated by dividing the relative initial abundance in one fraction by the sum of the abundances in all the fractions:

$$\text{Proportion in fraction } i = \frac{\text{Relative initial abundance in fraction } i}{\sum_{j=1}^{100} \text{Relative initial abundance in fraction } j} \quad P_7$$

Calculation of ribosome occupancy and ribosome density

For each gene, the ribosome occupancy was the proportion of the mRNA copies undergoing translation. It was calculated by summing the proportions of mRNA copies in fractions containing mRNAs bound to at least one ribosome (from fractions B to G in Fig 1). For each gene, we calculated the ribosome density in each fraction as the number of ribosomes bound to the mRNA copies normalized with respect to the coding sequence length.

Results

Polysome profiling experiment and translational parameters

The polysome profiling method was applied to *E*. *coli* MET739 cells during exponential growth on glucose in M9 minimum medium (Fig 1). Translation elongation was stopped by adding chloramphenicol and the cells were washed and lysed. The cell free extract was loaded on a sucrose gradient to fractionate the mRNA-ribosome complexes according to the number of bound ribosomes. A representative polysome profile is shown in Fig 1. The polysome profile was eluted

in a total of 24 sub-fractions. After protein elimination and total RNA extraction, peaks were assigned by estimating the ratios between the 23S rRNA and 16S rRNA in each sub-fraction (S1 Fig). Once the components of each sub-fraction were identified, sub-fractions were grouped in seven fractions from A to G (Fig 1). Fraction A comprised sub-fractions containing DNA, free RNAs and the small and large ribosomal subunits. The two small and large ribosomal subunits were respectively identified through the high 16S/23S rRNA and 23S/16S rRNA ratios (S1 Fig). In fractions B to G, the 23S/16S rRNA ratios were constant around 1.8 and matched entire ribosomes. The 2^{nd} peak (fraction B) was attributed to the monosome. The 3^{rd} and 4^{th} peaks corresponding to two and three ribosomes were grouped in fraction C. The 5^{th} peak corresponded to four ribosomes. The number of ribosomes in the following fractions was extrapolated as previously described [7]. The mean value of the number of bound ribosomes in fractions B to G was calculated from four independent experiments (Fig 1). We chose to exemplify the estimations of the translational parameters RO and RD using the *ihfB* gene. It is a well-expressed gene in *E*. *coli* coding for an integration host factor β-subunit commonly used as an internal normalization control in RT-qPCR experiments [24,28]. To estimate *ihfB* RO and RD, we quantified the abundance of *ihfB* mRNA copies in fractions A to G.

A first normalization of *ihfB* mRNA abundance was performed using the ERCC RNA spike-ins. The ERCC RNA spike-ins consisted in a mix of 92 transcripts with a wide range of lengths, GC contents and concentrations. A constant quantity of the ERCC RNA spike-in mix was introduced in all the total RNA samples before the reverse transcription step. Different total RNA/ERCC ratios (in µg/µL) were tested: 2.5/0.01, 4/0.01, 5/0.01, 2/0.02, 5/0.03 and the very high ratio of 5/0.001. For the 5/0.001 ratio, oligonucleotides specific to the ERCC RNA spike-ins were added to increase their reverse transcription. The highest reverse transcription efficiencies of ERCC and RNA were obtained with the very high ratio of 5/0.001. This ratio was thus chosen for all the ERCC normalization steps. The *ihfB* mRNA abundance was then normalized by the abundance of the four most concentrated ERCCs (ERCC 130, ERCC 002, ERCC 074 and ERCC 096) and by the initial mRNA abundance in each fraction to provide the

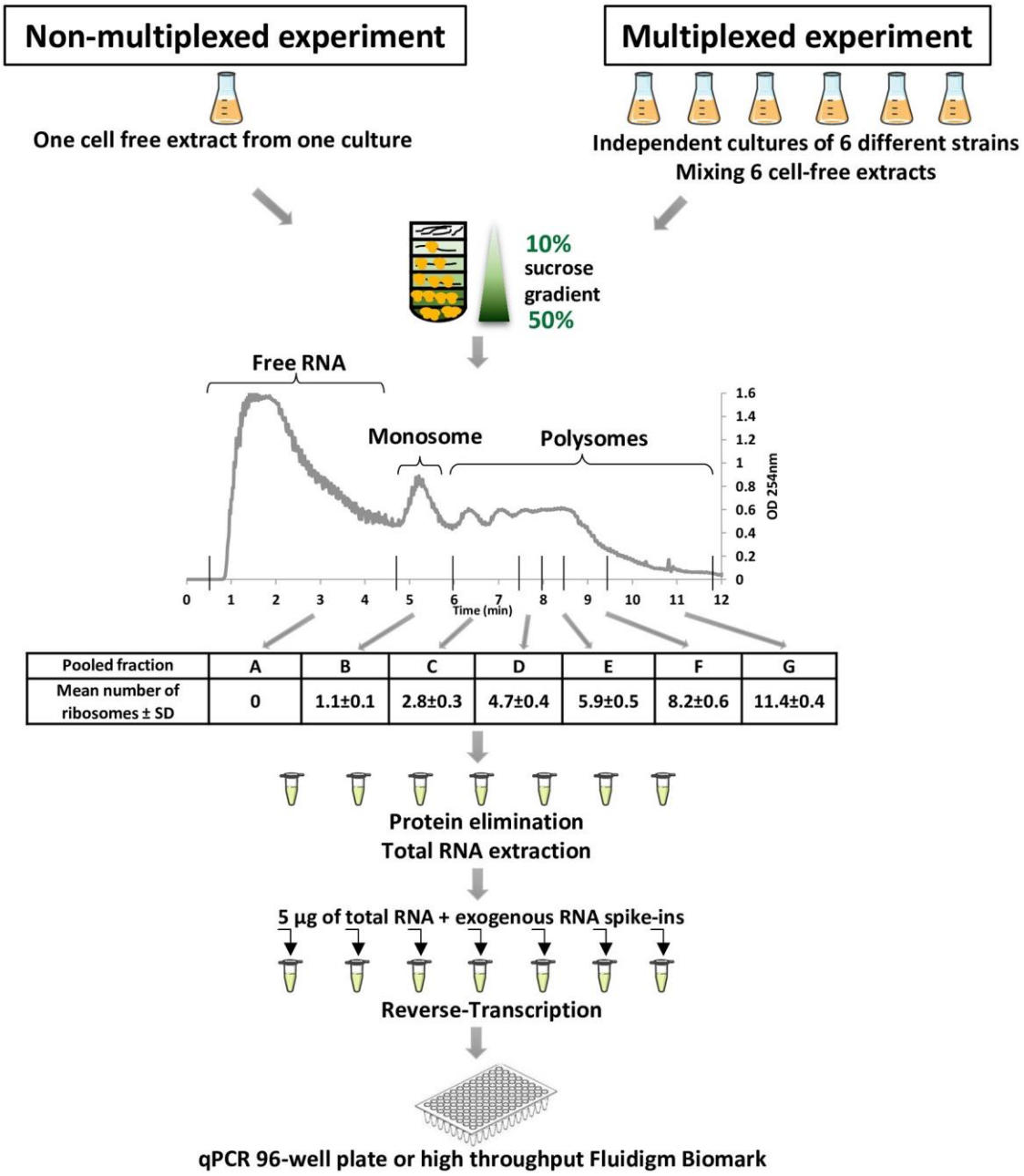


Fig 1. Polysome profiling experiment at a glance. All cultures and polysome profiling experiments were repeated twice to provide independent biological replicates.

https://doi.org/10.1371/journal.pone.0212297.g001

relative initial *ihfB* mRNA abundance in each fractions. The distributions of *ihfB* mRNA copy abundances between fractions A and G were then calculated to provide the typical plots of translational status (Fig 2A). Normalizations using any of the four ERCC led to similar distribution of *ihfB* mRNA copies between the

polysome fractions, so any of the four ERCC can be used to analyze a translational status. For further analyses, ERCC 074 (522pb, 35% GC, $15x10^{-21}$ mole/µL) was selected as normalizing ERCC, as it displayed the smallest variability between fractions and experiments (Fig 2B). The translational status of *ihfB* was characterized

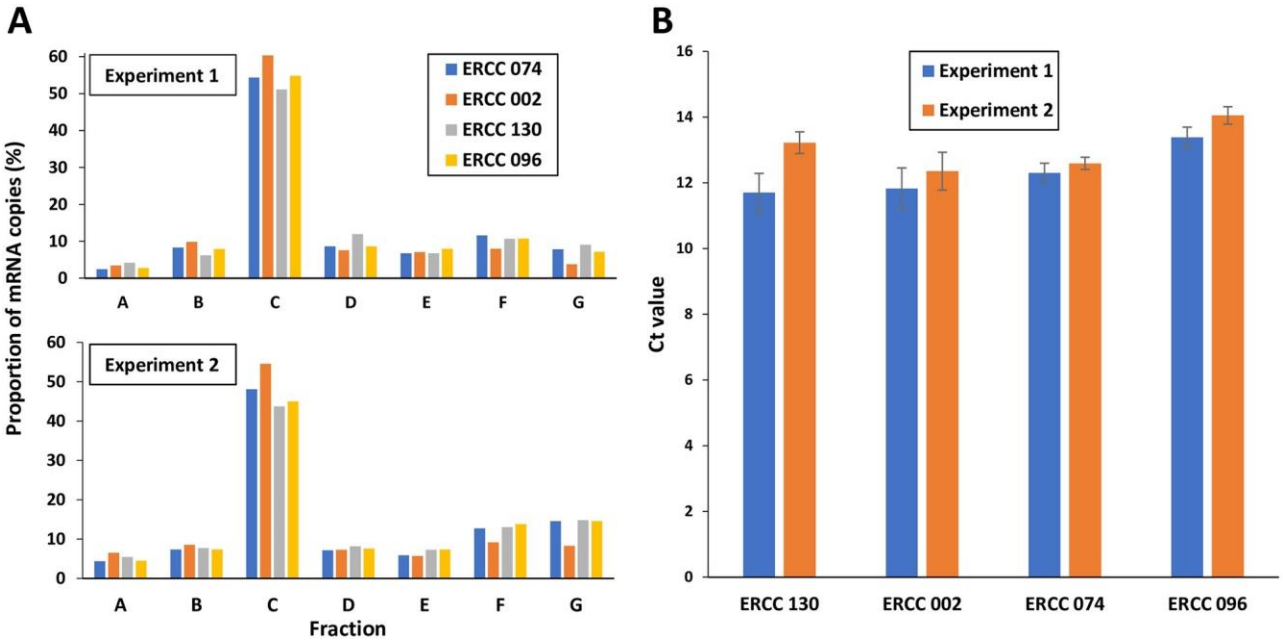


Fig 2. Distribution of *ihfB* mRNA copies in polysome fractions. (A) Proportion of *ihfB* mRNA copies in fractions A to G in two independent polysome profiling experiments from MET739 cell free extract when normalized by the four most concentrated ERCC (ERCC 130, ERCC 002, ERCC 074 and ERCC 096) using an RNA/ERCC ratio of 5/0.001 in µg/µL. (B) Variations in the levels of ERCC 130, ERCC 002, ERCC 074 and ERCC 096 estimated in two independent experiments. Mean values and standard deviations were calculated using the level values in the seven fractions (from A to G) of the same experiment.

https://doi.org/10.1371/journal.pone.0212297.g002

by an RO of 96.6 ± 0.4% corresponding to 3.4 ± 0.4% of ribosome-free mRNA copies not undergoing translation. Half the *ihfB* mRNA copies (in fraction C) were loaded with around 2.8 ribosomes corresponding to a ribosome density of 1 ribosome/100 nt and around 22% of the *ihfB* mRNA copies were heavily-loaded (in fractions F and G) with more than 8.2 bound ribosomes corresponding to a RD higher than 2.9 ribosomes/100 nt.

Multiplexing polysome profiling experiment does not alter the translational status of an mRNA

To validate a multiplexing method of polysome profiling experiments, we compared the translational status of *ihfB* without multiplexing and after multiplexing with cell extracts from distinct strains. We chose to multiplex up to six cell-free extracts. The translational status without multiplexing corresponds to the polysome profiling

experiment described above when only the cell-free extract of *E*. *coli* MET739 was loaded on the sucrose gradient. In the multiplexing polysome profiling experiment (Fig 1), the cell free extract of *E*. *coli* MET739 was mixed before loading on the sucrose gradient with five other cell free extracts: one from MET734 (overexpressing *cysZ*) and four from unrelated *E*. *coli* MG1655 constructs. Each cell free extract was produced from an individual culture in the exponential phase in M9 glucose medium. Quantification of *ihfB* mRNA copies in each fraction of the polysome was very reproducible within each experiment of polysome profiling (with and without multiplexing) (Fig 3). In addition, the distributions of *ihfB* mRNA copies were very similar before and after multiplexing (Fig 3). Consequently, comparable values of ribosome occupancy (96.6 ± 0.4% versus 96.6 ± 1.1%) and similar distributions of the ribosome density were obtained, the most frequent RD of the *ihfB* mRNA copies still being 1 ribosome/100 nt after multiplexing. We concluded that the

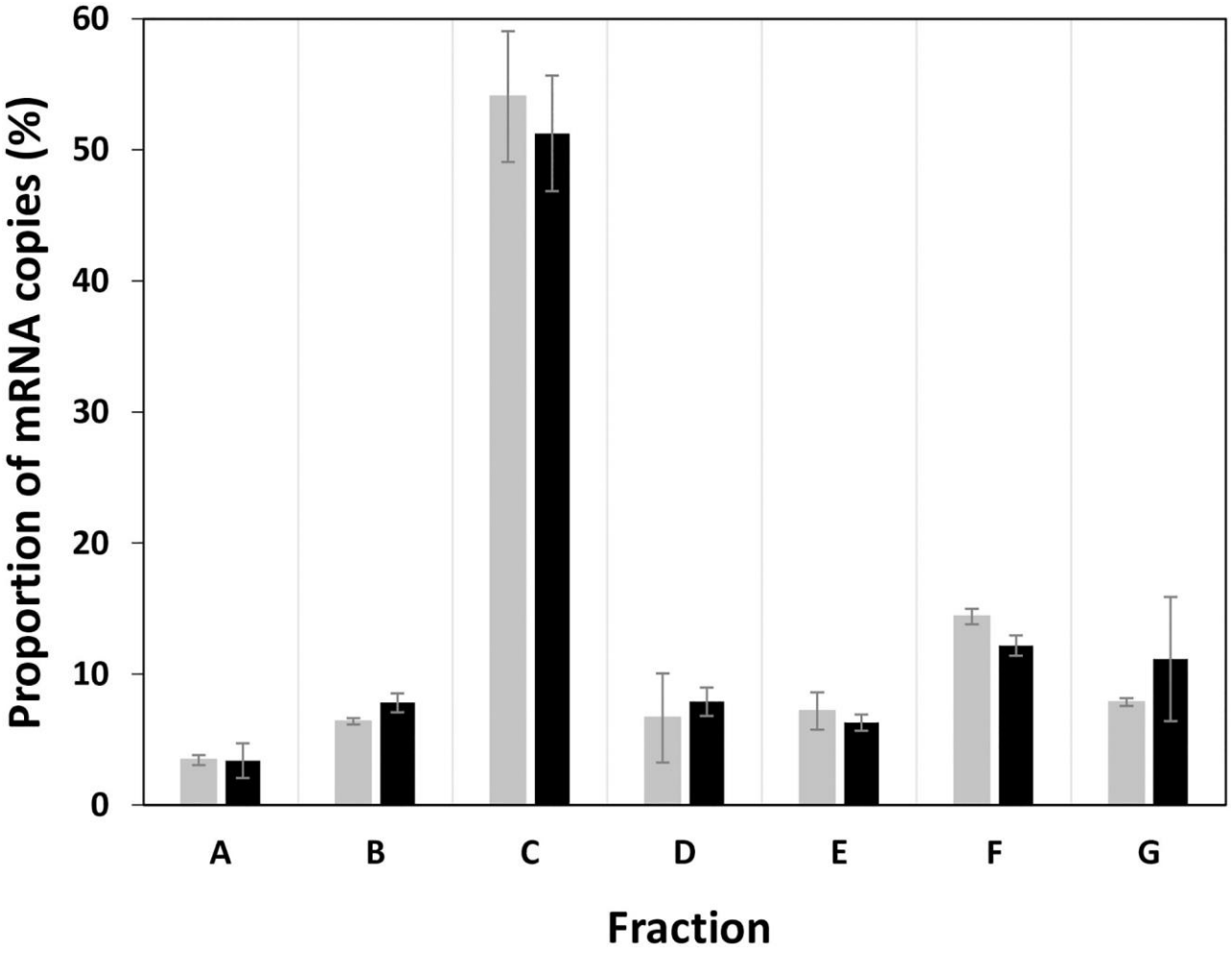


Fig 3. Distribution of *ihfB* mRNA copies between fractions A to G without multiplexing (grey) and after multiplexing (black). Fraction A, free mRNA not undergoing translation. Fractions B to G, mRNA bound with ascending number of ribosomes from 1 to 11. Mean values and standard deviations of two independent biological replicates are presented in the figure. Results were obtained with only MET739 cell-free extract in the nonmultiplexing experiment or with MET739 cell free extract mixed with cell free extracts from five other strains in the multiplexing experiment.

https://doi.org/10.1371/journal.pone.0212297.g003

multiplexing polysome profiling method did not affect the translational status observed for an mRNA and that this method can therefore be used to study translation statuses in parallel in multiple strains.

Multiplexing polysome profiling experiment can be used to monitor the translational response between two conditions

We wanted to know if the translational response of a gene between two conditions was similar using the classical non-multiplexing and our multiplexing polysome profiling methods. Using the two methods, we thus investigated the translational response of the *cysZ* gene when its mRNA level was increased. The *cysZ* gene codes for a high-affinity, high-specificity sulfate transporter that provides the sulfur source for the synthesis of cysteine [29]. Sulfate uptake by CysZ is essential for the survival of *E*. *coli* under low sulfate conditions. We focused on the *E*. *coli* MET734 strain in which the *cysZ* gene is under the transcriptional control of the arabinose inducible P_{BAD} promoter. *E*. *coli* MET734 was cultured in M9 glucose without arabinose (low *cysZ* mRNA level) and with arabinose transcriptional induction. Our measurements showed that arabinose induction led to a 27 ± 6 fold induction of *cysZ* mRNA in MET734. On the other hand, the mRNA level from the chromosomal copy of *cysZ* in MG1655 wild type was four times lower than the level observed in MET734 without arabinose induction, and about 100 times lower than in MET734 with arabinose induction. In the pool of *cysZ* mRNA copies, the part originating from the chromosomal copy of *cysZ* can thus be neglected. Therefore in the multiplexing experiments, the *cysZ* mRNAs were only assigned to strain MET734. The five other strains not carrying the *cysZ* gene on the plasmid were also cultured independently with and without arabinose induction. Four polysome profiling experiments were then performed: one non-multiplexing experiment (only the cell free extract of *E*. *coli* MET734 was loaded on sucrose gradient) and the multiplexing experiment (loading of *E*. *coli* MET734 mixed with the five other cell free extracts), with and without arabinose transcriptional induction. The translational responses of *cysZ* between low and high mRNA expression level using the two methods are shown in Fig 4.

Comparison of the distributions of *cysZ* mRNA copies using the standard (grey bars) and multiplexing method (black bars between A and B in Fig 4) showed equivalent distributions in the different fractions. This confirmed what was observed with *ihfB* that multiplexing cell free extracts did not alter the analysis of the distribution of a particular mRNA. With both the nonmultiplexing and multiplexing methods, induction of *cysZ* expression led to a shift in the mRNA copies from the free mRNA fraction (fraction A) toward the more heavily ribosome bound fractions (mainly fraction G). With a high level of mRNA, ribosome

occupancy increased from 62 ± 4.1% to 96 ± 3.0% without multiplexing and from 55.7 ± 0.7% to

89.6 ± 5.2% after multiplexing, reflecting the marked decrease in free mRNAs (lower proportions in fraction A). Consequently, ribosome density increased at high mRNA level to reach 1.5 ribosomes/100 nt in around 30% of the *cysZ* mRNA copies (in fraction G). These results showed that the multiplexing polysome profiling experiment allowed similar *cysZ* translational

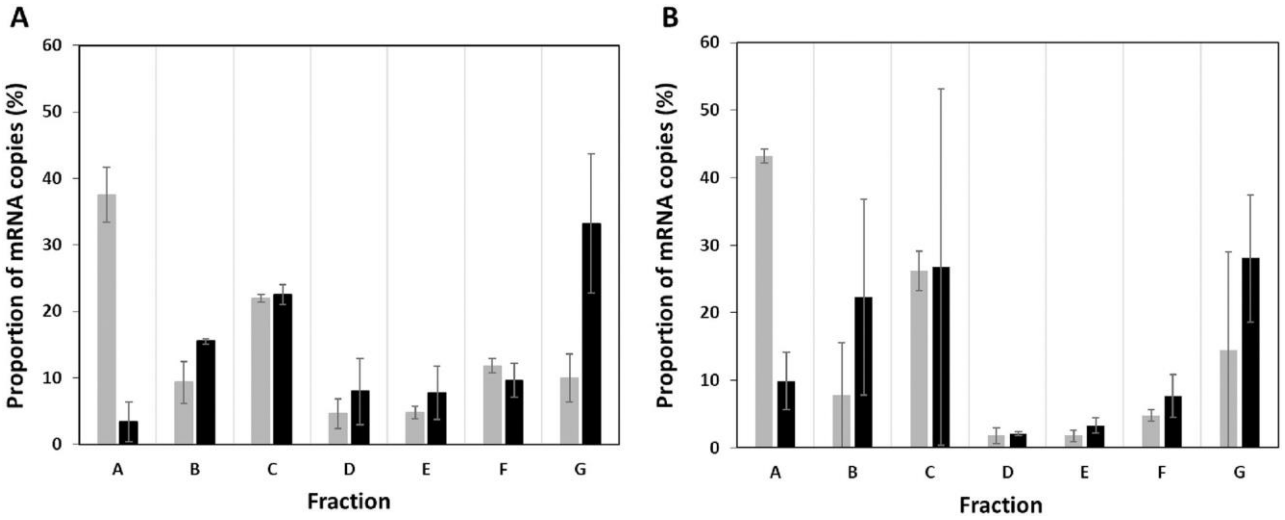


Fig 4. Distribution of *cysZ* mRNA copies between fractions in the two conditions: "without induction" (grey) and "with induction" (black) in (A) standard and (B) multiplexing polysome profiling experiments. Fraction A consists in free mRNA copies not undergoing translation. Fractions B to G contain mRNA copies bound with ascending numbers of ribosomes from 1 to 11. Mean values and standard deviations of two independent biological replicates are given. Results were obtained using only strain MET734 in the non-multiplexing experiment or using strain MET734 mixed with 5 other cell free extracts.

https://doi.org/10.1371/journal.pone.0212297.g004

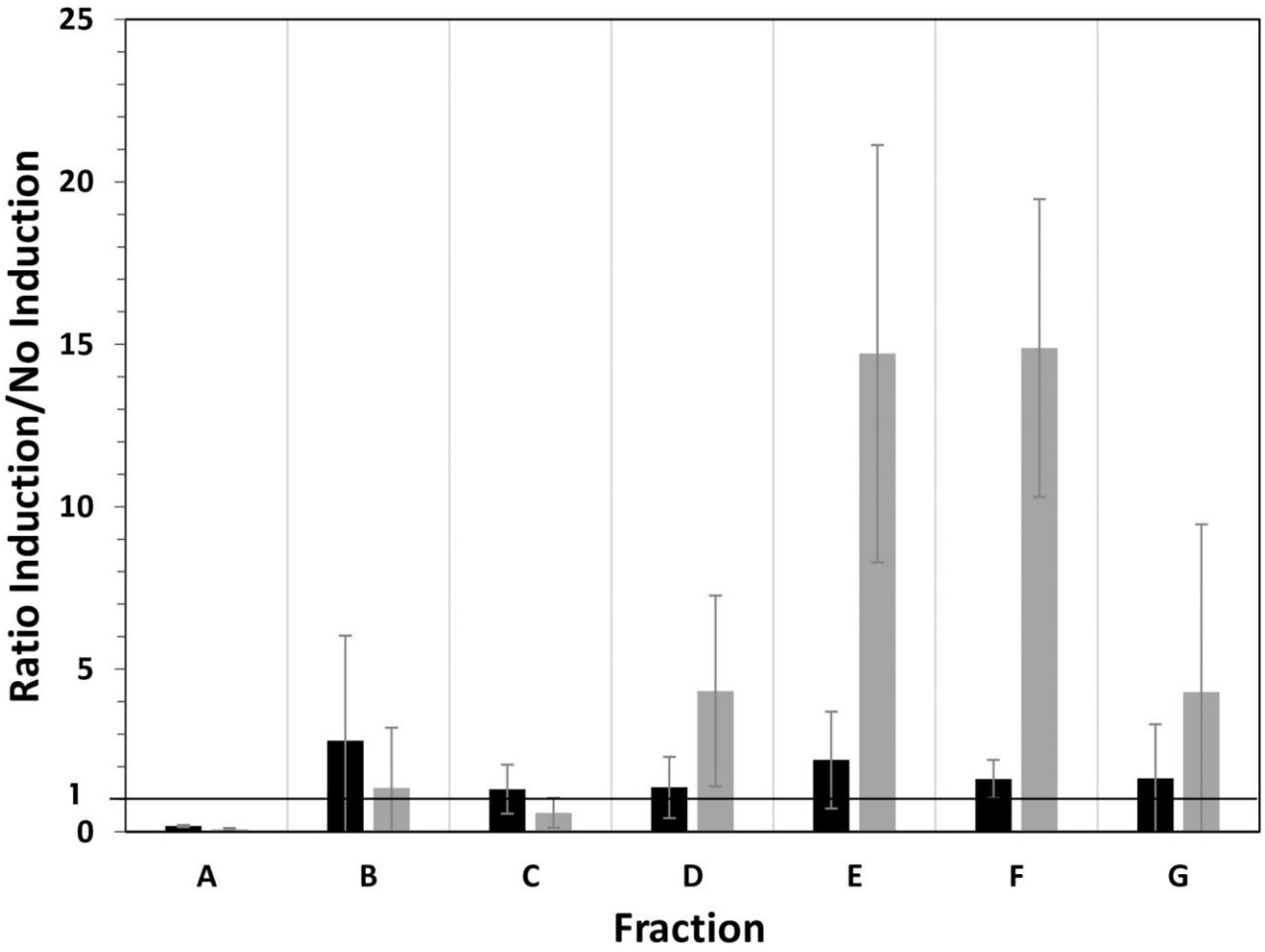


Fig 5. Translational response of *cysZ* (black) and *lacZ* (grey) in two mRNA expression conditions using the multiplexing polysome profiling method. For each fraction, the ratio of the proportion of mRNA copies with high mRNA levels to the proportion of mRNA copies with low mRNA levels was determined. Fraction A consists in free mRNA copies not undergoing translation while fractions B to G contain mRNA copies bound with ascending number of ribosomes from 1 to 11. Mean values and standard deviations of two independent biological replicates are given.

https://doi.org/10.1371/journal.pone.0212297.g005

responses to higher mRNA levels: a decrease in free *cysZ* mRNA fraction and a higher ribosome load.

Parallel characterization of translational responses to changing transcription level

We used the multiplexing polysome profiling experiment to simultaneously study the translational responses of the *cysZ* and *lacZ* genes at two transcriptional levels. The *lacZ* gene codes for a β-D-galactosidase enzyme involved in lactose and other β-galactoside catabolism [30]. In

E. *coli* MET739, the *lacZ* gene is under the transcriptional control of the arabinose inducible

P_{BAD} promoter. Without arabinose, the mRNA level from the chromosomal copy of *lacZ* in MG1655 wild type was 14 times lower than the level observed in MET739. Arabinose induction led to a 50 ± 23 times higher *lacZ* mRNA level in MET739. Therefore in the multiplexing experiments, the *lacZ* mRNAs were only assigned to strain MET739. Using the two multiplexing polysome profiling experiments described in the previous section, without arabinose (low *cysZ* and *lacZ* mRNA levels) and with arabinose (high *cysZ* and *lacZ* mRNA levels), we assessed the translational responses of *cysZ* and *lacZ* to changing mRNA levels (Fig 5). We calculated the ratio of the proportion of *cysZ* and *lacZ* mRNA copies with high mRNA levels to the proportion of mRNA copies with low mRNA levels. In both genes, when the mRNA expression was induced, we observed a significant reduction of the proportion of the free mRNA copies (ratios much lower than 1 in fraction A) and therefore an increase in the more heavily ribosome-loaded mRNA copies. However the pattern of translational responses of *cysZ* and *lacZ* differed. The *cysZ* mRNA copies spread all over the ribosome loaded fractions (all fractions from B to G exhibited ratios higher than 1) whereas the *lacZ* mRNA copies shifted more preferentially to fractions E and F. The magnitude of the translational response was more than 6-fold higher for *lacZ* than for *cysZ*. These results show that the multiplexing polysome profiling experiment allowed the parallel characterization of translational responses of two genes after increasing their mRNA level. For both *cysZ* and *lacZ*, the global translational responses to increased mRNA levels consisted in a shift toward the more heavily ribosome-loaded mRNA copies but the pattern and magnitude of the responses differed.

Specificity of the translational response

To check the specificity of the translational response to changing the mRNA level measured for *cysZ* and *lacZ*, we compared their responses to those of 10 chromosomally encoded genes not under the control of P_{BAD} (namely *eno*, *ihfB*, *rpsJ*, *rpsD*, *rplK*, *rplV*, *rpsL*, *trmJ*, *rnpA* and *rppH*, S1 Table). Ratios of mRNA copies proportions for the non-inducible genes were calculated in the two conditions, i.e. with and without arabinose, using the multiplexing experiments and were compared to the values of *cysZ* and *lacZ* (Fig 6). As expected, the noninducible genes showed no significant difference in their translational response between with and without arabinose, since their mRNA level did not significantly differ between the two conditions (variations in the expression of the 10 non-inducible genes were in the range of the technical error). With arabinose, the decreases in the proportions of both *cysZ* and *lacZ* copies in fraction A were higher than those in the non-inducible genes. The increases in the proportions of *lacZ* copies in the fractions D, E and F were considerably higher than the increases in the non-inducible genes; the increase was only slight for *cysZ* in fraction B compared to the non-inducible genes. These results confirmed that the multiplexing method can be used to specifically measure the translational response of a gene after changing its mRNA level in the cell.

Discussion

In this work, we developed and validated a multiplexing polysome profiling method to study the translation status of mRNAs and its variations in *E*. *coli*. The distribution of mRNA copies between the different polysome profiling fractions, and consequently ribosome occupancy and ribosome density, was similar using the standard non-multiplexing method and the new multiplexing method. The multiplexing method allows parallel quantification of the specific translational response to changing gene expression. In this study, we present the translational response of two genes, *cysZ* and *lacZ*, but our multiplexing approach allows simultaneous analysis of the translational response of up to six genes. In the case of *cysZ* and *lacZ*, we demonstrated a similar overall effect of the concentration of mRNA on the translation status with a higher ribosome load at higher mRNA levels but with a gene-specific pattern and magnitude of the responses. This result demonstrates co-transcriptional regulation of translation for these two genes. We hypothesize that co-transcriptional regulation of *cysZ* and *lacZ* translation contributes to physiological adaptation when cells regulate the mRNA level of genes to adapt to environmental changes (such as a low sulfate conditions [29] for *cysZ* or the availability of

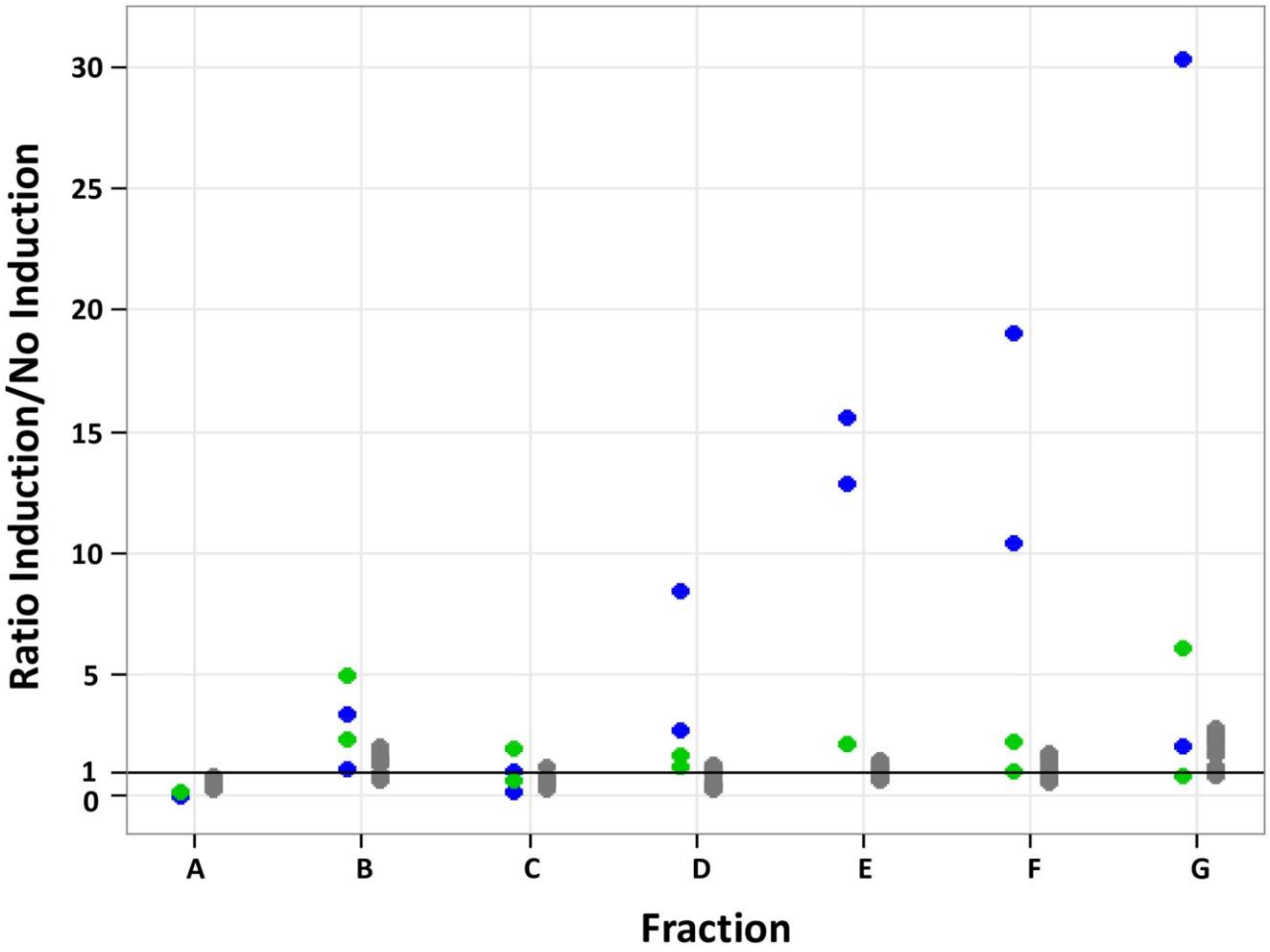


Fig 6. Translational response of the *cysZ* (green), *lacZ* (blue) and of 10 other unregulated genes (grey), with and without arabinose induction, using multiplexing polysome profiling experiments. For each fraction, the ratio was calculated between the proportion of mRNA copies with arabinose induction (high mRNA level) and the proportion of mRNA copies without arabinose induction (low mRNA level). Fraction A consists in free mRNA copies not undergoing translation while fractions B to G contain mRNA copies bound with ascending number of ribosomes from 1 to 11. Ratios were calculated from two independent biological replicates of each condition (2 dots are plotted for *cysZ* and *lacZ* and 20 dots for the unregulated genes in each fraction).

https://doi.org/10.1371/journal.pone.0212297.g006

lactose as a carbon source for *lacZ* [30]). Further studies are now required to confirm these findings at the genome scale. Applied here to explore *E*. *coli* translation regulation, the multiplexing polysome profiling method can be expanded to any other organism. Using multiplexing saves time and effort and reduces the cost and technical bias that may result when large numbers of samples have to be handled.

In this study, we simultaneously studied the translational response of different genes to the same stimulus (i.e. their mRNA level). We used different molecular constructs in the same genetic background to trigger changes in mRNA levels. In the

multiplexing experiment, we mixed cell free extracts of different strains generated in the same conditions, either without induction or with transcription induction. Another possible application of the multiplexing method is studying the translational response of one gene to different stimuli related to changes in the growth environment or to modifications in the genetic background. In this case, the multiplexing experiment will mix cell free extracts generated in different conditions.

The only constraint will be that the gene of interest has to be specifically tagged (for example using barcode tagging to provide specific hybridization for qPCR) in each condition to differentiate the mRNA copies originating from each condition in the polysome profiling fractions.

Our multiplexing method will be of particular interest for the study of translation regulation at the mechanistic level. The effect of mRNA sequence-related parameters as potential regulators of translation initiation and elongation has been investigated using molecular approaches. The effect of codon usage on translation in yeast [19] or the one of 5'UTR sequence on polysome distribution in *Arabidopsis* [20] have already been investigated using polysome profiling. As these studies were limited to the characterization of only one mRNA with only three different sequence modifications, multiplexing polysome profiling experiment could easily be extended to the analysis of more genes and sequence modifications. When many samples were analyzed by the standard polysome profiling technique like in [18] to investigate the effect of eight different 5'UTR structures on the translation of a reporter mRNA in yeast, using our multiplexing method would have saved time, money and effort by analyzing a single multiplexed polysome. In the study of the effect of codon usage on translation elongation using the ribosome profiling method [31], the implementation of a complementary experiment of multiplex polysome profiling would have provided additional information on translation heterogeneity in the copies of the reporter mRNA.

Additional regulatory features of mRNA sequences on translation can be explored with the multiplexing method coupled with a high resolution PCR technique such as the TaqMan RT-PCR. Analysis of the translational response to small differences, from some nucleotides to single point mutation, in sequences suspected of being involved in translation regulation (like sequence motif (conserved pattern [7] and SD-like motif [32]), secondary structure [33,34] for the binding sequences of regulatory ncRNA [35] and proteins (for instance CsrA [36]) could be performed more easily and quickly using the multiplexing method coupled with highly specific TaqMan probes. The technique could be also used to study the translational effect of natural single nucleotide polymorphism, for example between different alleles [37,38], to tackle the long-term evolution of translation regulation.

In conclusion, the multiplexing polysome profiling method is a low scale method mainly useful to study translation of (i) several reporter mRNAs (with different

expression level, 5'and 3' UTR sequences or coding sequence), (ii) endogenous genes in a strain when they have been previously tagged with specific artificial sequences and (iii) different natural alleles of a gene found in closely related strains or species. Furthermore, at the genome-wide scale, this method coupled with RNA sequencing can also be used when mixing microorganisms with distinct genetic backgrounds. In this case, the possibility to assign the sequenced reads to the specific genes of each microorganism will allow the translation of these genes to be studied. The multiplexing method could open the way for "metatranslatomics" analyses.

Supporting information

S1 Fig. Ratios of the 23S rRNA and 16S rRNA amounts in sub-fractions of the polysome profiling experiment (grey line). The ratio of 16S/23S rRNA amounts is in blue and the ratio of 23S/16S rRNA amounts is in brown. The sub-fractions are delimited by vertical black lines. (TIF)

S1 Table. Sequence of qPCR primers used to quantify 12 endogenous genes of *E. coli* MG1655 and four ERCC RNA spike-ins.
(DOCX)

Author Contributions

Conceptualization: Se´bastien Nouaille, Laurence Girbal.

Data curation: Huong Le Nguyen.

Formal analysis: Huong Le Nguyen, Marie-Pierre Duviau.

Funding acquisition: Muriel Cocaign-Bousquet.

Methodology: Marie-Pierre Duviau, Se´bastien Nouaille, Laurence Girbal.

Project administration: Muriel Cocaign-Bousquet.

Supervision: Laurence Girbal.

Writing – original draft: Huong Le Nguyen.

Writing – review & editing: Muriel Cocaign-Bousquet, Se´bastien Nouaille, Laurence Girbal.

References

1. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet. 2013; 13(4):227–32.
2. Duval M, Simonetti A, Caldelari I, Marzi S. Multiple ways to regulate translation initiation in bacteria:

Mechanisms, regulatory circuits, dynamics. Biochimie. 2015; 114:18–29. https://doi.org/10.1016/j.biochi.2015.03.007 PMID: 25792421

3. Lange C, Zaigler A, Hammelmann M, Twellmeyer J, Raddatz G, Schuster SC, et al. Genome-wide analysis of growth phase-dependent translational and transcriptional regulation in halophilic archaea. BMC Genomics. 2007; 8:1–17.

4. Fisunov GY, Evsyutina D V., Garanina IA, Arzamasov AA, Butenko IO, Altukhov IA, et al. Ribosome profiling reveals an adaptation strategy of reduced bacterium to acute stress. Biochimie. 2017; 132: 66–74. https://doi.org/10.1016/j.biochi.2016.10.015 PMID: 27984202

5. Chasse´ H, Boulben S, Costache V, Cormier P, Morales J. Analysis of translation using polysome profiling. Nucleic Acids Res. 2017; 45(3):e15. https://doi.org/10.1093/nar/gkw907 PMID: 28180329

6. Arava Y, Wang Y, Storey JD, Liu CL, Brown PO, Herschlag D. Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A. 2003; 100(7):3889–94. https://doi. org/10.1073/pnas.0635171100 PMID: 12660367

7. Picard F, Milhem H, Loubière P, Laurent B, Cocaign-Bousquet M, Girbal L. Bacterial translational regulations: high diversity between all mRNAs and major role in gene expression. BMC Genomics. 2012; 13(1):528.

8. Picard F, Loubière P, Girbal L, Cocaign-Bousquet M. The significance of translation regulation in the stress response. BMC Genomics. 2013; 14(1):588.

9. Yamasaki S, Matsuura H, Demura T, Kato K. Changes in Polysome Association of mRNA Throughout Growth and Development in *Arabidopsis thaliana*. Plant Cell Physiol. 2015 Sep 26; 56(11):pcv133.

10. Chasse´ H, Aubert J, Boulben S, Le Corguille´ G, Corre E, Cormier P, et al. Translatome analysis at the egg-to-embryo transition in sea urchin. Nucleic Acids Res. 2018 May 18; 46(9):4607–21. https://doi.org/ 10.1093/nar/gky258 PMID: 29660001

11. Pytharopoulou S, Grintzalis K, Sazakli E, Leotsinidis M, Georgiou CD, Kalpaxis DL. Translational responses and oxidative stress of mussels experimentally exposed to Hg, Cu and Cd: One pattern does not fit at all. Aquat Toxicol. 2011 Sep; 105(1–2):157–65. https://doi.org/10.1016/j.aquatox.2011.06.007 PMID: 21718659

12. Melamed D, Pnueli L, Arava Y. Yeast translational response to high salinity: global analysis reveals regulation at multiple levels. RNA. 2008 Jul; 14(7):1337–51. https://doi.org/10.1261/rna.864908 PMID: 18495938

13. Matsuura H, Ishibashi Y, Shinmyo A, Kanaya S, Kato K. Genome-Wide Analyses of Early Translational Responses to Elevated Temperature and High Salinity in Arabidopsis thaliana. Plant Cell Physiol. 2010 Mar 1; 51(3):448–62. https://doi.org/10.1093/pcp/pcq010 PMID: 20089509

14. Wu Y-H, Taggart J, Song PX, MacDiarmid C, Eide DJ. An MSC2 Promoter-lacZ Fusion Gene Reveals Zinc-Responsive Changes in Sites of Transcription Initiation That Occur across the Yeast Genome. PLoS One. 2016 Sep 22; 11(9):e0163256. https://doi.org/10.1371/journal.pone.0163256 PMID: 27657924

15. Smirnova JB, Selley JN, Sanchez-Cabo F, Carroll K, Eddy AA, McCarthy JEG, et al. Global gene expression profiling reveals widespread yet distinctive translational responses to different eukaryotic translation initiation factor 2B-targeting stress pathways. Mol Cell Biol. 2005 Nov; 25(21):9340–9. https://doi.org/10.1128/MCB.25.21.9340-9349.2005 PMID: 16227585

16. Alves LR, Oliveira C, Goldenberg S. Eukaryotic translation elongation factor-1 alpha is associated with a specific subset of mRNAs in Trypanosoma cruzi. BMC Microbiol. 2015 May 19; 15:104. https://doi. org/10.1186/s12866-015-0436-2 PMID: 25986694

17. Liu B, Chen C. Translation Elongation Factor 4 (LepA) Contributes to Tetracycline Susceptibility by Stalling Elongating Ribosomes. Antimicrob Agents Chemother. 2018 Aug 1; 62(8):e02356–17. https:// doi.org/10.1128/AAC.02356-17 PMID: 29784847

18. Saglioccos FA, Vega Lasoj MR, Zhull D, Tuitell MF, Mccarthyj JEG, Brownsii AJP. The Influence of 5'-
Secondary Structures upon Ribosome Binding to mRNA during Translation in Yeast. J Biol Chem. 1993; 268(35):26522–30. PMID: 8253781

19. Presnyak V, Alhusaini N, Chen YH, Martin S, Morris N, Kline N, et al. Codon optimality is a major determinant of mRNA stability. Cell. 2015; 160(6):1111–24. https://doi.org/10.1016/j.cell.2015.02.029 PMID: 25768907

20. Matsuura H, Takenami S, Kubo Y, Ueda K, Ueda A, Yamaguchi M, et al. A Computational and Experimental Approach Reveals that the 5⁰-Proximal Region of the 5⁰-UTR has a Cis-Regulatory Signature Responsible for Heat Stress-Regulated mRNA Translation in Arabidopsis. Plant Cell Physiol. 2013 Apr 1; 54(4):474–83. https://doi.org/10.1093/pcp/pcs189 PMID: 23314753

21. King HA, Gerber AP. Translatome profiling: Methods for genome-scale analysis of mRNA translation. Brief Funct Genomics. 2016; 15(1):22–31. https://doi.org/10.1093/bfgp/elu045 PMID: 25380596

22. Ah-Seng Y, Rech J, Lane D, Bouet J-Y. Defining the Role of ATP Hydrolysis in Mitotic Segregation of Bacterial Plasmids. Viollier PH, editor. PLoS Genet. 2013 Dec 19; 9(12):e1003956. https://doi.org/10. 1371/journal.pgen.1003956 PMID: 24367270

23. Esquerre´ T, Laguerre S, Turlan C, Carpousis AJ, Girbal L, Cocaign-Bousquet M. 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. 2014; 42(4):2460–72. https://doi.org/10.1093/nar/ gkt1150 PMID: 24243845

24. Nouaille S, Mondeil S, Finoux AL, Moulis C, Girbal L, Cocaign-Bousquet M. The stability of an mRNA is influenced by its concentration: a potential physical mechanism to regulate gene expression. Nucleic Acids Res. 2017; 45(20):11711–24. https://doi.org/10.1093/nar/gkx781 PMID: 28977619

25. Redon E, Loubière P, Cocaign-Bousquet M. Role of mRNA stability during genome-wide adaptation of Lactococcus lactis to carbon starvation. J Biol Chem. 2005 Oct 28; 280(43):36380–5. https://doi.org/10. 1074/jbc.M506006200 PMID: 16131490

26. Maligoy M, Mercade M, Cocaign-Bousquet M, Loubiere P. Transcriptome analysis of Lactococcus lactis in coculture with Saccharomyces cerevisiae. Appl Environ Microbiol. 2008 Jan 15; 74(2):485–94. https://doi.org/10.1128/AEM.01531-07 PMID: 17993564

27. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 2001 May 1; 29(9):e45. PMID: 11328886

28. Morin M, Ropers D, Cinquemani E, Portais JC, Enjalbert B, Cocaign-Bousquet M. The Csr system regulates Escherichia coli fitness by controlling glycogen accumulation and energy levels. MBio. 2017; 8(5): e01628–17. https://doi.org/10.1128/mBio.01628-17 PMID: 29089432

29. Zhang L, Jiang W, Nan J, Almqvist J, Huang Y. The Escherichia coli CysZ is a pH dependent sulfate transporter that can be inhibited by sulfite. Biochim Biophys Acta. 2014; 1838:1809–16. https://doi.org/ 10.1016/j.bbamem.2014.03.003 PMID: 24657232

30. Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol. 1961; 3(3): 318–56.

31. Yu CH, Dang Y, Zhou Z, Wu C, Zhao F, Sachs MS, et al. Codon Usage Influences the Local Rate of Translation Elongation to Regulate Co-translational Protein Folding. Mol Cell. 2015; 59(5):744–54. https://doi.org/10.1016/j.molcel.2015.07.018 PMID: 26321254

32. Li GW, Oh E, Weissman JS. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. Nature. 2012; 484(7395):538–41. https://doi.org/10.1038/nature10965 PMID: 22456704

33. Evfratov SA, Osterman IA, Komarova ES, Pogorelskaya AM, Rubtsova MP, Zatsepin TS, et al. Application of sorting and next generation sequencing to study 5⁰-UTR influence on translation efficiency in Escherichia coli. Nucleic Acids Res. 2017 Apr 7; 45(6):3487–502. https://doi.org/10.1093/nar/gkw1141 PMID: 27899632

34. Giuliodori AM, Di Pietro F, Marzi S, Masquida B, Wagner R, Romby P, et al. The cspA mRNA is a thermosensor that modulates translation of the cold-shock protein CspA. Mol Cell. 2010 Jan 15; 37(1): 21–33. https://doi.org/10.1016/j.molcel.2009.11.033 PMID: 20129052

35. Gottesman S, Storz G. Bacterial small RNA regulators: versatile roles and rapidly evolving variations.
Cold Spring Harb Perspect Biol. 2011 Dec 1; 3(12):a003798. https://doi.org/10.1101/cshperspect.a003798 PMID: 20980440

36. Potts AH, Vakulskas CA, Pannuri A, Yakhnin H, Babitzke P, Romeo T. Global role of the bacterial posttranscriptional regulator CsrA revealed by integrated transcriptomics. Nat Commun. 2017 Dec 17; 8(1):
1596. https://doi.org/10.1038/s41467-017-01613-1 PMID: 29150605

37. Artieri CG, Fraser HB. Evolution at two levels of gene expression in yeast. Genome Res. 2014; 24(3):
411–21. https://doi.org/10.1101/gr.165522.113 PMID: 24318729

38. Wang Z, Sun X, Zhao Y, Guo X, Jiang H, Li H, et al. Evolution of gene regulation during transcription and translation. Genome Biol Evol. 2015; 7(4):1155–67.
https://doi.org/10.1093/gbe/evv059 PMID: 25877616