



HAL
open science

Modélisation mathématique de la dégradation des ARNm bactériens et intégration de données omiques

Thibault Étienne

► **To cite this version:**

Thibault Étienne. Modélisation mathématique de la dégradation des ARNm bactériens et intégration de données omiques. Statistiques [math.ST]. Université de Lyon, 2020. Français. NNT : 2020LYSE1271 . tel-03306903

HAL Id: tel-03306903

<https://theses.hal.science/tel-03306903v1>

Submitted on 29 Jul 2021

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

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



N°d'ordre NNT :

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée au sein de
l'Université Claude Bernard Lyon 1
Ecole Doctorale N° ED 341
(Évolution, Écosystèmes, Microbiologie, Modélisation (E2M2))
Spécialité de doctorat : Bio-mathématiques
Discipline : Biostatistique

Soutenue publiquement le 01/12/2020, par :

Thibault Etienne

**Modélisation mathématique de la dégradation
des ARNm bactériens et intégration de
données omiques**

Devant le jury composé de :

Pujo-Menjouet Laurent
Pères Sabine
Guedj Jérémie
Carpousis Agamenon
Charles Hubert
Martin-Magniette Marie-Laure
Ropers Delphine
Cocaïgn-Bousquet Muriel

MCF, Univ. Lyon 1
MCF, Univ. Paris-Saclay
CR, INSERM
DR, CNRS
PR, INSA
DR, INRAE
CR, Inria
DR, INRAE

Président
Rapportrice
Rapporteur
Examinateur
Examinateur
Examinatrice
Directrice de thèse
Co-directrice de thèse

Université Claude Bernard – LYON 1

Administrateur provisoire de l'Université	M. Frédéric FLEURY
Président du Conseil Académique	M. Hamda BEN HADID
Vice-Président du Conseil d'Administration	M. Didier REVEL
Vice-Président du Conseil des Etudes et de la Vie Universitaire	M. Philippe CHEVALLIER
Vice-Président de la Commission de Recherche	M. Jean-François MORNEX
Directeur Général des Services	M. Pierre ROLLAND

COMPOSANTES SANTE

Département de Formation et Centre de Recherche en Biologie Humaine	Directrice : Mme Anne-Marie SCHOTT
Faculté d'Odontologie	Doyenne : Mme Dominique SEUX
Faculté de Médecine et Maïeutique Lyon Sud - Charles Mérieux	Doyenne : Mme Carole BURILLON
Faculté de Médecine Lyon-Est	Doyen : M. Gilles RODE
Institut des Sciences et Techniques de la Réadaptation (ISTR)	Directeur : M. Xavier PERROT
Institut des Sciences Pharmaceutiques et Biologiques (ISBP)	Directrice : Mme Christine VINCIGUERRA

COMPOSANTES & DEPARTEMENTS DE SCIENCES & TECHNOLOGIE

Département Génie Electrique et des Procédés (GEP)	Directrice : Mme Rosaria FERRIGNO
Département Informatique	Directeur : M. Behzad SHARIAT
Département Mécanique	Directeur M. Marc BUFFAT
Ecole Supérieure de Chimie, Physique, Electronique (CPE Lyon)	Directeur : Gérard PIGNAULT
Institut de Science Financière et d'Assurances (ISFA)	Directeur : M. Nicolas LEBOISNE
Institut National du Professorat et de l'Education	Administrateur Provisoire : M. Pierre CHAREYRON
Institut Universitaire de Technologie de Lyon 1	Directeur : M. Christophe VITON
Observatoire de Lyon	Directrice : Mme Isabelle DANIEL
Polytechnique Lyon	Directeur : Emmanuel PERRIN
UFR Biosciences	Administratrice provisoire : Mme Kathrin GIESELER
UFR des Sciences et Techniques des Activités Physiques et Sportives (STAPS)	Directeur : M. Yannick VANPOULLE
UFR Faculté des Sciences	Directeur : M. Bruno ANDRIOLETTI

Remerciements

Avant toute chose, il convient de remercier l'INRAE et l'Inria pour avoir rendu cette thèse possible.

Une thèse, c'est long et ça se fait rarement tout seul. Tout d'abord, je remercie très chaleureusement Delphine et Muriel, qui ont accepté d'encadrer ma thèse et qui n'ont pas cessé, depuis quatre ans, de me supporter, dans tous les sens du terme, avec une patience dont moi-même j'aurais été incapable. Vous m'avez énormément appris.

Ensuite, il faut mentionner l'équipe BLADE de l'institut de biotechnologie de Toulouse, ainsi que l'équipe IBIS de l'Inria, pour m'avoir accueilli pendant tout ce temps. Je tiens tout particulièrement à remercier les membres de l'équipe IBIS (Inria) : Eugenio et Aline pour leur aide apportée en ce qui concerne mes nombreuses interrogations mathématiques, Alain pour ses conseils en bioinformatique et son humour aux pauses repas, et je n'oublie pas non plus ma collègue de bureau, Antrea pour ses conseils en biologie et son soutien apporté durant cette longue épreuve qu'est une thèse.

Dans le même couloir, je remercie l'équipe MISTIS, en particulier Julyan pour m'avoir fait découvrir sous un autre angle les statistiques non paramétriques. Je remercie enfin les doctorants Daria, Karina et Maria qui m'ont poussé à travailler mon anglais et qui ont été d'un grand soutien tout au long de cette thèse. Je tiens aussi à remercier le personnel administratif : les assistantes de l'équipe IBIS qui ont su m'aider à surmonter ma phobie administrative.

Je ne dois pas non plus oublier Fabienne Macro de l'école doctorale qui m'a aidé dans les démarches de ma fin thèse (sélection du jury et dossier administratif pour le dépôt de ce manuscrit).

Je remercie les membres de mon comité de thèse, Jean-Pierre Mazat, Hubert Charles, Adeline Samson et Laurence Girbal, pour avoir suivi mon travail au cours de toutes ces années. Je remercie également l'ensemble de mon jury de thèse : le président de jury, Laurent Pujot-Menjouet ; les deux rapporteurs de ma thèse, Sandrine Pérès et Jérémie Guedj, pour avoir accepté de relire mon manuscrit ; les examinateurs Marie-Laure Martin-Magnette, Agamemnon Carpousis et Hubert Charles.

Je remercie mes amis de longue date Aurélien, Christophe, Clémence, Enora et Floriane qui ont toujours su m'apporter aide et soutien.

Enfin, je tiens à adresser mes plus sincères remerciements à ma famille, qui a toujours su me soutenir et m'encourager à distance, peu importe les problèmes que j'ai pu rencontrer.

Résumé

L'adaptation de la physiologie bactérienne à l'environnement met en jeu des régulations de l'expression génique. Le contrôle de la concentration des ARNm est l'une d'entre elles. Cette concentration dépend à la fois d'une balance entre les vitesses de transcription et de dégradation des messagers. Si ces deux mécanismes sont bien étudiés à l'échelle moléculaire chez *Escherichia coli*, leurs rôles respectifs dans la régulation du niveau des transcrits à l'échelle du génome restent encore mal compris. L'apparition de données omiques en dynamique permet de mieux caractériser la régulation de la dégradation des ARNm en réponse à des changements environnementaux. Cette masse de données omiques est souvent sous-exploitée. Des approches sans a priori sont généralement utilisées pour analyser ces données. Cependant la réduction des coûts computationnels et l'avancée des connaissances biologiques du processus de dégradation des ARNm bactériens et de la modélisation mécanistique et statistique permettent d'aller plus loin.

Au cours de cette thèse, différentes approches pour modéliser et analyser des données de transcriptomique en dynamique précédemment obtenues par T. Esquerré chez *E. coli* sont abordées sous le prisme de la biologie des systèmes. L'idée est d'étendre les travaux déjà effectués sur le sujet à l'aide de modèles mathématiques et statistiques afin de pousser notre compréhension de ce problème biologique. Plus précisément, un modèle mécanistique a été créé pour modéliser la dégradation des ARNm bactériens. Ce modèle permet de montrer le rôle de la compétition entre les ARNm dans la variabilité de la stabilité des ARNm et l'apparition de délais avant la dégradation. Une méthodologie statistique utilisant les modèles non linéaires à effets mixtes a été adaptée au modèle mécanistique et éprouvée pour l'estimation des paramètres cinétiques de ce modèle à partir de données omiques dynamiques. L'interprétation biologique des paramètres estimés montre que la compétition joue un rôle majeur dans l'adaptation de la stabilité des ARNm en couplant étroitement la dégradation à la transcription. Il s'agit d'un nouveau mécanisme global de la dégradation. Seuls 10% des ARNm sont en plus régulés par des facteurs comme les petits ARN.

Mots clés : Dégradation des ARNm, modélisation déterministe, modèles non linéaires à effets mixtes, biologie des systèmes, données omiques dynamiques.

Abstract

Adaptation of bacterial physiology to environment involves regulations of gene expression. The control of mRNA concentrations is one of them. It results from a balance between mRNA transcription and degradation rates. While these two mechanisms are well-studied at the molecular level in *Escherichia coli*, their contributions to transcript levels at the genome-wide level remain poorly understood. The emergence of dynamic omics data opens up new opportunities to better understand the regulation of mRNA degradation in response to environmental changes. This mass of omics data is often under-utilised. Approaches without a priori are generally used to analyse these data. However, computational cost reduction and knowledge advancements about bacterial mRNA degradation and both mechanistic and statistical modelling should permit deeper investigations.

In this work, different approaches to model and analyse dynamic transcriptomic data previously obtained by T. Esquerré in *E. coli* are approached with the angle of systems biology. The idea is to extend past work on the subject with mathematical and statistical modelling to push further our understanding of this biological problem. More precisely, a mechanistic model of bacterial mRNA degradation was created. This model allows to establish the key role of competition between mRNAs in the variability of mRNA stability and the occurrence of delays before degradation. A statistical methodology based on non-linear mixed-effect models was adapted to the mechanistic model and tested for the estimation of model parameters from dynamic omics data. Biological interpretation of the estimated parameters shows the pivotal role of competition in tuning mRNA stability through the tight coupling of mRNA degradation to transcription. This is a new global regulatory mechanism of mRNA decay. Sole 10% of mRNAs are regulated by additional factors such as small RNAs.

Keywords : mRNA degradation, deterministic modelling, non-linear mixed effect models, system biology, time-series omics data.

Table des matières

1	Introduction	13
1.1	Préambule	13
1.2	Les approches omiques	14
1.3	La biologie des systèmes	16
1.3.1	Définition et défis	16
1.3.2	Approches top-down	18
1.3.3	Approches bottom-up	19
1.4	La dégradation des ARNm chez <i>E. coli</i>	20
1.4.1	La RNase E : acteur majeur de la dégradation des ARNm	21
1.4.2	Mécanismes généraux de la dégradation des ARNm chez <i>E. coli</i>	22
1.4.3	Régulation de la dégradation des ARNm chez <i>E. coli</i>	23
1.4.4	Couplage de la dégradation avec d'autres processus cellulaires	25
1.5	Modélisation de la dégradation des ARNm	25
1.5.1	Modélisation du mécanisme général	25
1.5.2	Modélisation du couplage de la dégradation avec d'autres processus cellulaires	27
1.6	Problématique et plan	31
2	Approches statistiques d'analyse de la dégradation des ARNm	35
2.1	Mesure expérimentale de la dégradation des ARNm	35
2.1.1	Quantification des ARNm par transcriptomique dynamique après arrêt de transcription	36
2.1.2	Méthodes expérimentales alternatives	38
2.2	Pré-traitement des données de transcriptome dynamique	38
2.2.1	Correction du bruit de fond	39
2.2.2	Prise en compte des bruits liés aux réplicats et aux sondes	42
2.2.3	Normalisation entre les taux de croissance	43
2.3	Clustering de séries temporelles omiques	44
2.3.1	Méthodes de classification de séries temporelles	44
2.3.2	Clustering des patterns de dégradation des ARNm	45
3	Modélisation cinétique déterministe de la dégradation des ARNm	53
3.1	Étude de la cinétique des réactions biochimiques	53
3.1.1	La loi d'action de masse	53
3.1.2	Réactions enzymatiques : Henri-Michaelis-Menten	55
3.2	Modélisation mécanistique de la dégradation des ARNm	57
3.2.1	Réaction de Michaelis-Menten	58

3.2.2	Réaction de Michaelis-Menten avec inhibition compétitive des ribosomes	58
3.2.3	Réaction de Michaelis-Menten avec inhibition non-compétitive des ribosomes	61
3.2.4	Quel modèle pour quelles données?	64
3.2.5	Choix des hypothèses de quasi-équilibre pour modéliser la dégradation des ARNm	65
3.3	Modélisation de la dégradation des ARNm par système enzymatique compétitif à plusieurs substrats	66
3.4	Discussion sur l'origine du délai en début de cinétique de dégradation	88
4	Modèle statistique pour l'estimation des paramètres cinétiques	89
4.1	Les modèles non linéaires à effets mixtes : définition générale	90
4.1.1	Modèle d'observation	90
4.1.2	Modélisation des paramètres individuels	91
4.1.3	Modélisation des paramètres individuels avec ajout d'une covariable .	93
4.1.4	Modèle d'erreur résiduelle	94
4.1.5	Implémentation des modèles non linéaires à effets mixtes	96
4.2	Application à l'estimation des paramètres cinétiques du modèle de dégradation	97
5	Annexes	117
5.1	Densité des intensité brutes	117
5.2	Article de revus	121

Chapitre 1

Introduction

1.1 Préambule

Mes travaux de thèse se situent dans le domaine de la biologie des systèmes computationnelle et ont fait appel à des approches biostatistiques, biomathématiques, bioinformatiques et de biologie moléculaire. Mon objectif a été de pousser plus loin l'analyse de données omiques dynamiques telles qu'elles sont couramment employées pour répondre à des questions biologiques. J'ai ainsi complété les approches statistiques à grande échelle par des approches de modélisation mathématique intégrant des connaissances de processus biologiques pour en apprendre plus sur les mécanismes de régulation sous-jacents. J'ai appliqué ces approches à l'étude de la régulation de la dégradation des ARNm de la bactérie *Escherichia coli* dans différentes conditions environnementales. Financé par une bourse de thèse INRAE-Inria, j'ai réalisé mes travaux au sein des équipes BLADE (Bacterial Adaptation, Diversity and Engineering) du Toulouse Biotechnology Institute et Ibis du centre de recherche Inria Grenoble - Rhône-Alpes. L'équipe BLADE a pour objectif de mieux comprendre, chez *E. coli* et *Lactococcus lactis*, la dynamique d'expression des gènes et ses conséquences dans l'adaptation bactérienne. L'équipe s'intéresse notamment aux mécanismes de régulation de la stabilité des ARNm qu'elle étudie par des expériences de transcriptomique dynamique. L'équipe de biologie des systèmes Ibis étudie les mécanismes moléculaires impliqués dans l'adaptation des micro-organismes à l'environnement par des approches de modélisation mathématique et expérimentales. Elle développe en particulier des modèles mathématiques dont l'estimation des paramètres à partir de données expérimentales renseignent sur les mécanismes de régulation impliqués dans l'adaptation bactérienne à l'environnement.

Les travaux de thèse que je vais décrire dans ce manuscrit ont couvert des aspects variés d'analyse de profils de dégradation des ARNm cellulaires, de la modélisation cinétique du mécanisme de dégradation, ainsi que de l'estimation et l'interprétation des paramètres cinétiques de ces modèles. Les concepts méthodologiques nécessaires à la compréhension de ces différents aspects seront décrits dans les chapitres dédiés à ces différents travaux. Dans ce premier chapitre, j'introduis de façon générale ce qui forme le contexte de ma thèse, à savoir l'utilisation d'approches de biologie des systèmes pour l'analyse de données omiques sur la dégradation des ARNm.

1.2 Les approches omiques

Les approches omiques à haut débit connaissent un succès grandissant au cours des trois dernières décennies et sont en constante évolution en raison de l'avancée des connaissances et des technologies. En effet on retrouve plus de 800 termes différents utilisant le suffixe -omique sur Google. Un autre marqueur de l'intérêt grandissant porté aux approches omiques est le nombre croissant de publications sur ce sujet (*cf* figure 1.1).

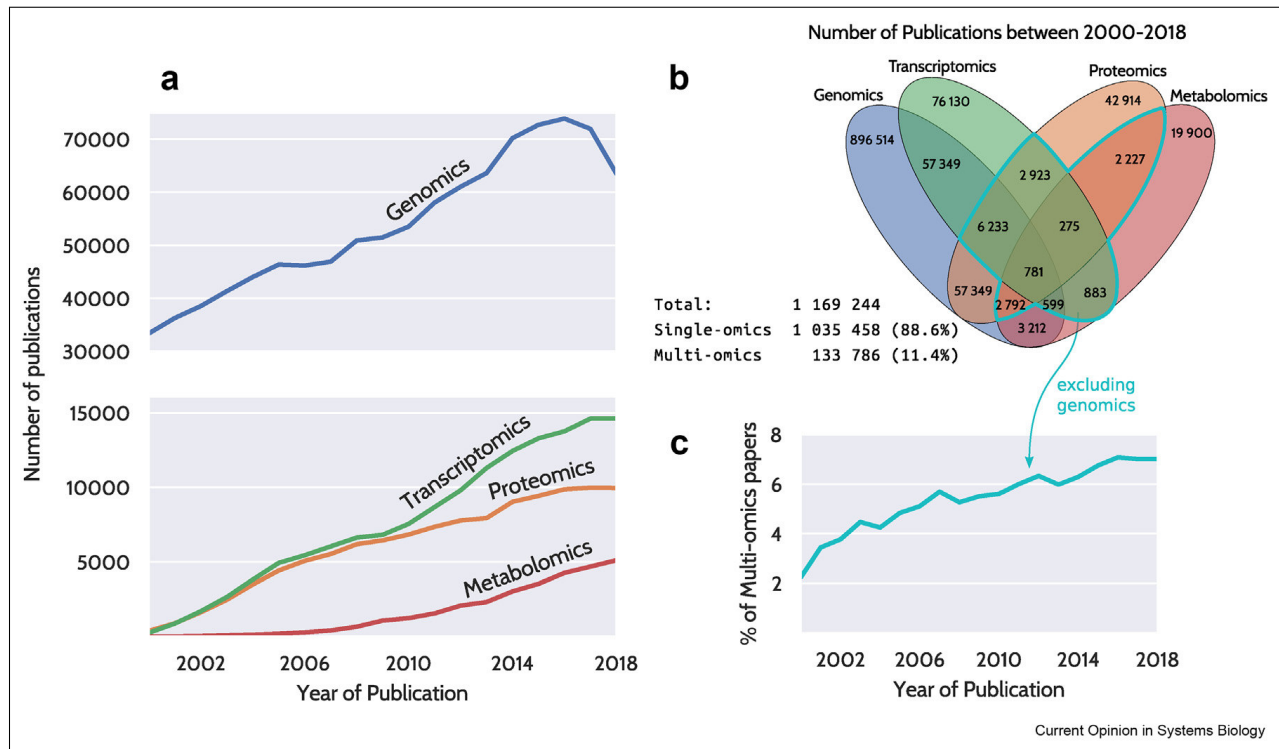


FIGURE 1.1 – **A.** Nombre total d'articles publiés par année depuis 2000. **B.** Diagramme de Venn représentant l'intersection des articles publiés entre 2000 et 2017 incluant un ou plusieurs termes omiques. **C.** Pourcentage des publications traitant des données multi-omiques rapporté au nombre total de publications. Les auteurs n'ont pas pris en compte le terme "génomique" car celui-ci apparaît très souvent dans d'autres contextes. (D'après Noor et al. 2019)

Lorsqu'on parle d'approches omiques, cela peut devenir assez rapidement confus. En effet, aujourd'hui on a tendance à utiliser ce fameux suffixe -omique à tort et à travers. Pour clarifier le propos, revenons à l'origine étymologique du mot. Il dérive du terme "Ome" provenant de la langue sanscrit qui signifie "totalité". Lorsque l'on combine des objets du vivant tels que les gènes, les protéines... avec le suffixe -ome, on obtient "génomique", "transcriptomique", "protéomique", "métabolomique" qui signifient respectivement l'ensemble des gènes, des ARN, des protéines ou des métabolites d'un système biologique donné (*cf.* figure 1.2). Les termes "génomique", "protéomique", "métabolomique" etc. définissent quand à eux les méthodes analytiques qui permettent d'explorer, sans à priori, l'ensemble de ces différents objets biologiques.

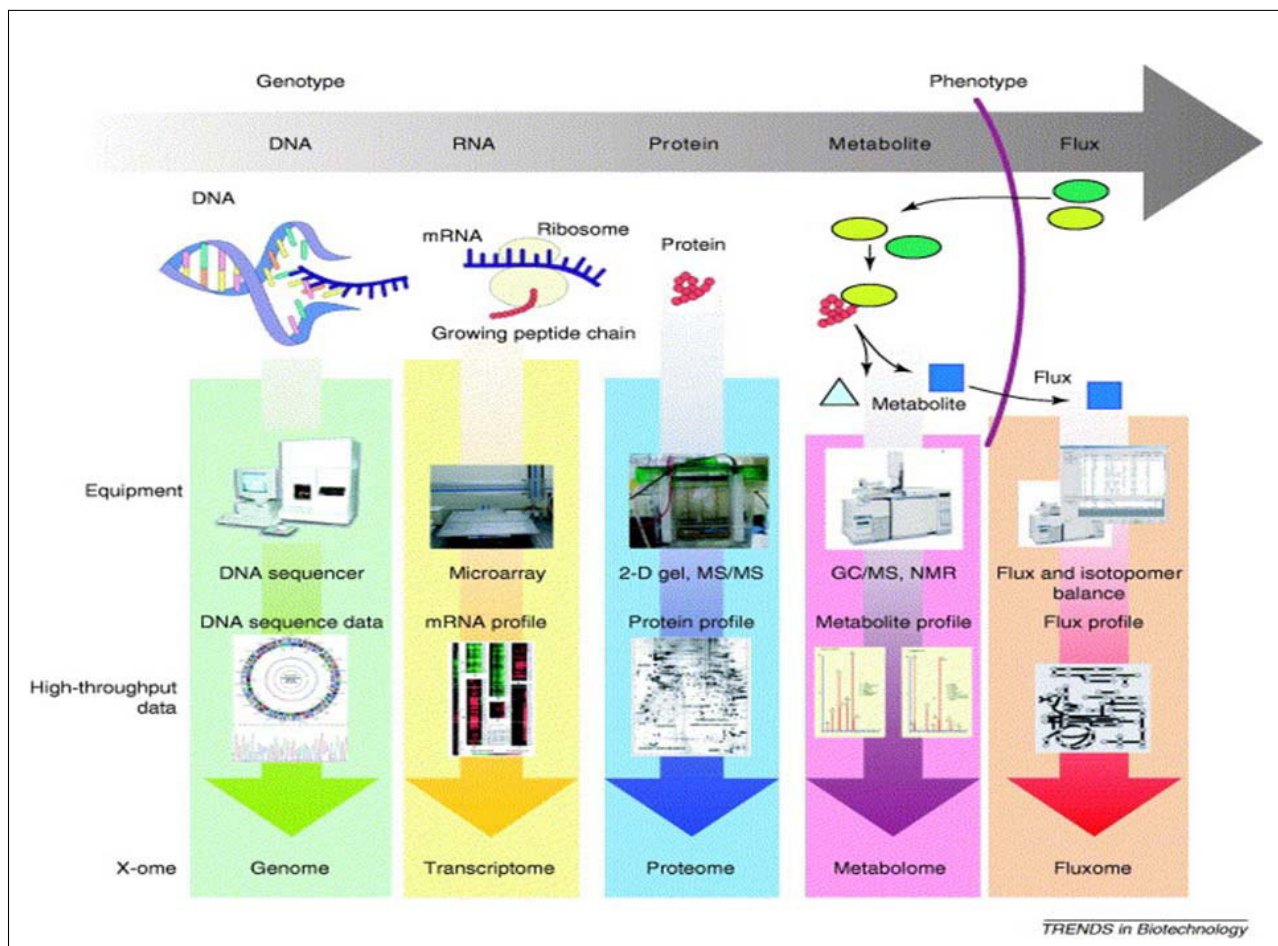


FIGURE 1.2 – Représentation schématique des approches omiques les plus répandues associées à leurs techniques expérimentales. La génomique est associée au séquençage d’ADN à haut débit. Le transcriptome s’obtient à l’aide de puces à ADN ou d’expériences de RNAseq qui sont maintenant la référence. Les données de protéome s’acquièrent par l’électrophorèse sur gel 2D associée à la spectrographie de masse (MALDI-TOF) et plutôt par LC MS-MS (chromatographie liquide) désormais. L’analyse du métabolome progresse également rapidement avec le développement de meilleures technologies GC/MS, LC/MS et RMN. Enfin le fluxome s’obtient avec un substrat marqué isotopiquement qui permet la détermination des profils de flux dans la cellule (fluxome). (D’après Lee et al. 2005)

Plus récemment, de nouveaux types de données omiques ont fait leur apparition. En effet jusqu’à maintenant les données omiques ont été obtenues à partir d’une population de cellules. Une mesure omique comme par exemple la concentration d’un ARNm représente donc une mesure moyenne prise sur une population de cellules. Par conséquent, la variabilité qui existe entre les différentes cellules n’est pas prise en compte. Les données de cellules uniques (single cell) offrent à présent de nouvelles opportunités, notamment pour étudier les variations de concentrations entre les cellules d’une même espèce [Wang and Bodovitz, 2010]. Enfin dû à l’affluence des données omiques de différentes sortes, de nouvelles approches ont été développées permettant de combiner différents types de données omiques au sein d’une même analyse. On parle alors d’approches multi-omiques.

La démocratisation de ces approches et la baisse des coûts permettent également l'obtention de données omiques dynamiques. C'est typiquement le cas des données de stabilité des ARNm étudiées dans cette thèse. Elles ont été obtenues par transcriptomique dynamique encore appelée stabilomique.

1.3 La biologie des systèmes

1.3.1 Définition et défis

Autant par son rythme effréné que par son impact fondamental sur la recherche biologique, la révolution des "omiques" dans son ensemble est tout sauf une révolution tranquille. Grisantes par leur potentiel en apparence illimité, les méthodes "omiques" ne sont pas sans poser des défis majeurs, aussi bien sur le plan scientifique que sur celui de l'organisation structurelle (stockage des données, moyens financiers, humains, *etc*). Parmi les difficultés le plus souvent évoquées, on peut lister deux défis majeurs. Premièrement, un défi intellectuel qui a pour objectif d'appréhender la complexité du vivant et formuler de meilleures/nouvelles hypothèses à partir de la masse de données générées. Deuxièmement, un défi méthodologique visant à contrôler la qualité des données produites, à comparer de larges ensembles de données obtenues dans des conditions expérimentales différentes, à disséminer les résultats (visualisation) ou encore à gérer l'émergence des données omiques en dynamique.

L'émergence de ces nouvelles données et de leurs défis associés a eu un impact très important sur la façon d'approcher une problématique biologique. Pour répondre à ces différents défis, les chercheurs se sont alors tournés vers l'utilisation des mathématiques et le développement d'outils informatiques associés. Ces thématiques étaient auparavant réservées aux problèmes de biologie théorique et souffraient cruellement du manque de données expérimentales pour la validation des hypothèses biologiques étudiées. L'arrivée en masse des données omiques a permis l'avènement d'une nouvelle approche hautement interdisciplinaire qu'on appelle aujourd'hui biologie systémique plus communément appelée biologie des systèmes [Kitano, 2002a].

Donner une définition précise et unique de la biologie est compliqué en raison de ses racines dans divers domaines de recherche. Plusieurs chercheurs s'y sont essayé. Par exemple, Kitano [2002b] la définit comme la compréhension des systèmes biologiques complexes nécessitant l'intégration de la recherche expérimentale et informatique. Alors que selon Cassman [2005], "l'objectif de la biologie des systèmes se définit comme la compréhension du comportement des réseaux, et en particulier de leurs aspects dynamiques, qui nécessite l'utilisation de la modélisation mathématique étroitement liée à l'expérience". Le point important qui ressort de ces deux définitions est que la biologie des systèmes n'est pas un domaine de recherche en tant que tel mais plutôt une approche méthodologique pour aborder des questions biologiques. En ce qui me concerne, je pense que chaque chercheur utilisant cette approche possède sa propre définition qui est biaisée par son domaine de prédilection. Ce qui fait consensus est que l'utilisation de la biologie des systèmes nécessite des connaissances variées et des collaborations entre différentes équipes pour mener à bien un projet de recherche dans ce domaine. La biologie des systèmes fait émerger des problématiques méthodologiques variées

comme l'observation de la réponse cellulaire (production et analyse des données), l'intégration de données pour l'identification d'interactions (méthodes statistiques, apprentissage), la modélisation formelle et numérique du comportement du système (modèles formels, modèles symboliques, modèles numériques), l'étude de modèles de processus cellulaires (mathématiques appliquées, informatique théorique), ou encore le contrôle d'un système biologique dans l'optique d'une application biotechnologique. Chacun de ces thèmes peuvent être vu comme faisant partie intégrante d'un cycle itératif caractéristique de la biologie des systèmes (figure 1.3).

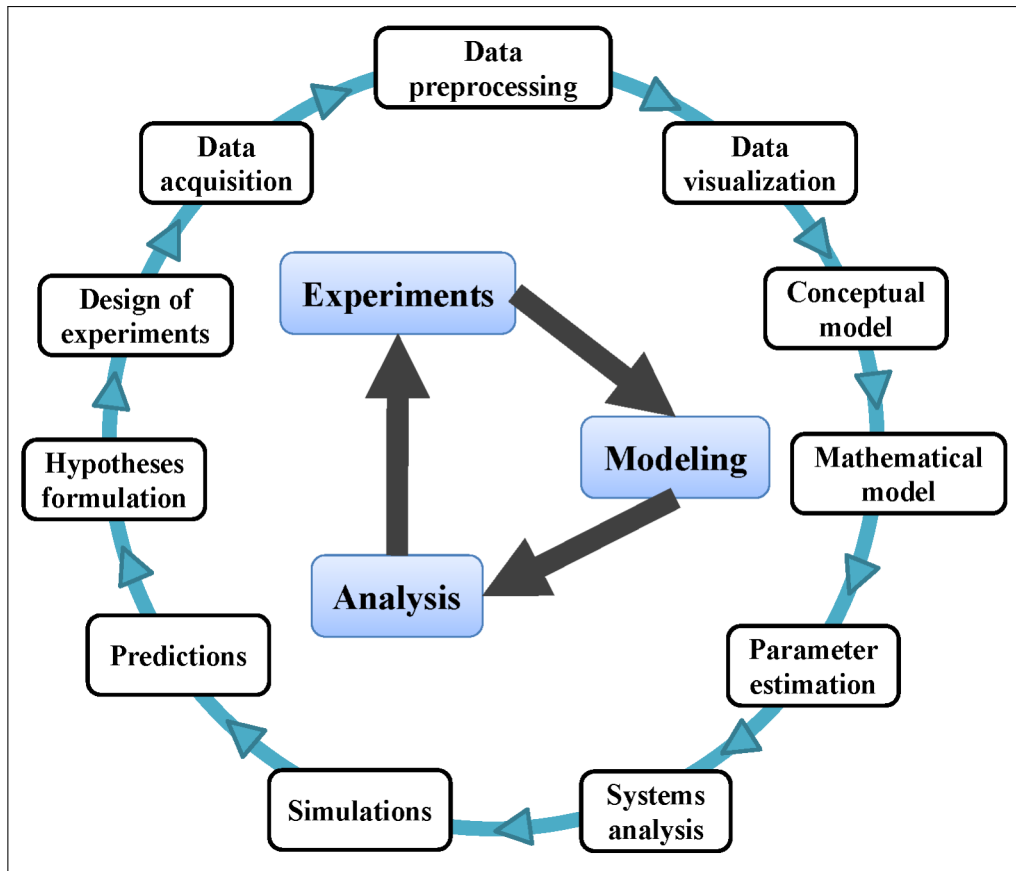


FIGURE 1.3 – Schéma méthodologique représentant les différentes étapes d'un projet de recherche en biologie des systèmes. Ce cycle itératif est construit à travers trois étapes majeures : les expériences (design expérimental et production de données), la modélisation du système biologique étudié (construction d'un modèle mathématique permettant la simulation d'un système biologique sous différentes hypothèses plus ou moins réalistes) et l'analyse des résultats (interprétation des simulations, confrontation avec les données expérimentales et validation) (D'après Jayavelu 2015).

Deux visions complémentaires de la biologie des systèmes se dégagent de la figure 1.3 : l'approche top-down qui est basée sur les données (production de données/analyses statistiques/interprétation biologique) et celle qu'on appellera "bottom-up" qui se base en priorité sur les connaissances disponibles dans la littérature (modélisation/confrontation aux données expérimentales/ validation des hypothèses biologiques). Ces deux visions sont décrites plus en détails dans les deux sections suivantes.

1.3.2 Approches top-down

Les approches top-down de la biologie des systèmes sont à l'interface entre la bioinformatique et les biostatistiques. Celles-ci permettent d'explorer, sans a priori, l'ensemble des différents objets biologiques à l'aide de méthodes statistiques et informatiques. Ces méthodes requièrent peu de connaissances sur le système biologique étudié et, le plus souvent, les données utilisées sont statiques par opposition aux données dynamiques (suivi de l'évolution d'une concentration au cours du temps). Nous pouvons classer ces approches en trois catégories : les approches statistiques qui regroupent l'analyse différentielle de l'expression génique, l'analyse des corrélations ou l'analyse d'enrichissement fonctionnel. Une deuxième classe inclut les méthodes non supervisées qui ont pour but principal la classification et la réduction de dimension. Enfin on regroupera dans la dernière classe les méthodes d'apprentissage automatique supervisé. Ces dernières sont très utiles pour gérer les grandes quantités de données biologiques [Tarca et al., 2007, Camacho et al., 2018] (pour revue [Zeng and Lumley, 2018]). Les applications sont variées comme par exemple l'étude des interactions microbiennes [DiMucci et al., 2018], la prédiction de combinaisons de médicaments [Weinstein et al., 2017] ou la découverte de nouveaux diagnostics [Perakakis et al., 2018]. Notons toutefois que ces méthodes sont encore limitées par le manque de données déjà validées pour entraîner l'algorithme et mesurer son succès.

Les données de transcriptome (quantification des concentrations relatives des ARNm à l'échelle du génome) produites par puce à ADN sont un exemple de données couramment analysées par des approches de la première catégorie [Hatfield et al., 2003]. Ce type de puce est un ensemble de molécules d'ADN fixées en rangées ordonnées sur une petite surface qui peut être du verre, du silicium ou du plastique. Cette technologie permet d'analyser le niveau d'expression des gènes (transcrits) dans une cellule, un tissu, un organe ou encore un organisme à un moment donné et dans un état donné par rapport à un échantillon de référence [Schena et al., 1995]. L'analyse différentielle de l'expression génique permet d'identifier les gènes différemment exprimés entre deux conditions en utilisant la théorie des statistiques inférentielles (test statistique). Ce type d'analyse a aussi été étendu à des données dynamiques (courte séries temporelles ou des données de survie) [Ernst and Bar-Joseph, 2006]. Nous aurons l'occasion de revenir plus en détail sur les aspects techniques de ces approches dans le chapitre 2. Comme expliqué précédemment, cette méthodologie sans a priori s'appuie sur des bases statistiques solides permettant l'obtention de résultats biologiques importants. La recherche de biomarqueurs en est un exemple notable [McDermott et al., 2013, Ghosh and Poisson, 2009]. Une grande partie de ces méthodes sont de plus implémentées en R ou en Python, ce qui les rend accessibles à un grand nombre d'utilisateurs (environnement Bioconductor [Gentleman et al., 2004] de R, Biopython [Cock et al., 2009] en python et les bibliothèques spécifiques associées).

Cependant, les approches top-down sont limitées, d'une part par la complexité des systèmes biologiques. En effet les multiples interactions (directes et indirectes) entre différents objets biologiques (régulations pléiotropiques/spécifiques, effets spatiaux temporels dans les processus biologiques, etc) peuvent être identifiés, cependant l'approche top-down révèle rarement une causalité ou des preuves d'interactions directes sans intégration de connaissances supplémentaires [Kuppe et al., 2020]. D'autre part, l'impossibilité d'observer certains processus biologiques dans leur intégralité, comme par exemple l'étude d'un sous-ensemble de gènes

qui ne représente pas la population ou encore l'utilisation de covariables incomplètes (données manquantes), peuvent limiter l'analyse et la biaiser si les résultats portent sur une population [Mills et al., 2019, Fang et al., 2018]. Pour palier à cela, il existe une autre approche dite "bottom up" qui est le plus souvent associée aux aspects prédictifs de la biologie des systèmes.

1.3.3 Approches bottom-up

Pour parvenir à une compréhension mécanistique du problème biologique étudié, il faut aller au-delà des approches purement basées sur les données et utiliser les connaissances préalables sur les interactions des composants. Il existe de nombreuses approches de modélisation basées sur des connaissances, dont la mise en oeuvre varie selon le problème biologique, la dimension du problème et l'objectif souhaité. Par exemple la modélisation des réseaux d'interactions protéines-protéines utilise le formalisme booléen [Wang et al., 2012] qui permet de gérer de manière efficace le nombre important d'interactions dans un temps computationnel raisonnable. La modélisation d'un réseau métabolique à l'échelle de la cellule utilise quand à elle les modèles basés sur les relations stœchiométriques contraintes entre composants chimiques. On utilisera dans ce cas des techniques d'analyse de flux pour étudier la structure du réseau métabolique [Terzer et al., 2009]. Enfin pour maximiser la puissance prédictive et les connaissances mécanistiques au niveau moléculaire, les simulations d'équations différentielles (ou stochastiques) basées sur des modèles physiques de liaison et de catalyse restent la référence [Segel, 1975, Cornish-Bowden, 2013a, Gillespie, 1977, Wilkinson, 2018]. La combinaison de ces modèles mécanistiques pour décrire l'ensemble des systèmes qui existent dans une cellule est l'un des plus grands défis de la biologie des systèmes. Un modèle très précis d'une cellule humaine entière, par exemple, pourrait potentiellement rendre obsolètes toutes les expériences sur des cultures de cellules vivantes [Yurkovich and Palsson, 2018, Karr et al., 2015]. Cependant, la simulation de grands systèmes d'équations est une tâche intensive en calcul, c'est pourquoi cette approche est le plus souvent utilisée sur des "petits" systèmes biologiques. Notons toutefois que différentes méthodes sont capables de gérer ce problème et permettent de passer un peu plus à l'échelle. On citera par exemple les méthodes booléennes ou la séparation des échelles de temps (hypothèse du quasi équilibre) qui simplifient la forme mathématique des fonctions de vitesse et/ou réduisent le nombre de paramètres [Segel and Slemrod, 1989].

Un exemple classique d'application de l'approche bottom-up est l'étude du comportement dynamique des réactions biochimiques [Segel, 1975, Cornish-Bowden, 2013a]. Le plus souvent, le biologiste a connaissance des "inputs" (concentrations initiales de substrats) et des outputs (concentrations des produits). Cependant il est nettement plus compliqué (au niveau expérimental en particulier) d'étudier comment le substrat se transforme en produit. Grâce à la biologie des systèmes (thermodynamique, théorie cinétique des réactions biochimiques), le chercheur est à même de mieux comprendre comment une molécule A est transformée en une molécule B et par conséquent d'intégrer des connaissances supplémentaires sur le système sans pour autant avoir besoin des observations complètes du processus biologique. En d'autres termes, l'approche bottom-up permet une compréhension plus précise du système biologique étudié en incorporant des connaissances de la littérature. Nous reviendrons sur la modélisations des réactions enzymatiques dans le chapitre 3 et 4, lorsque nous nous intéresserons à modéliser la cinétique de dégradation des ARNm.

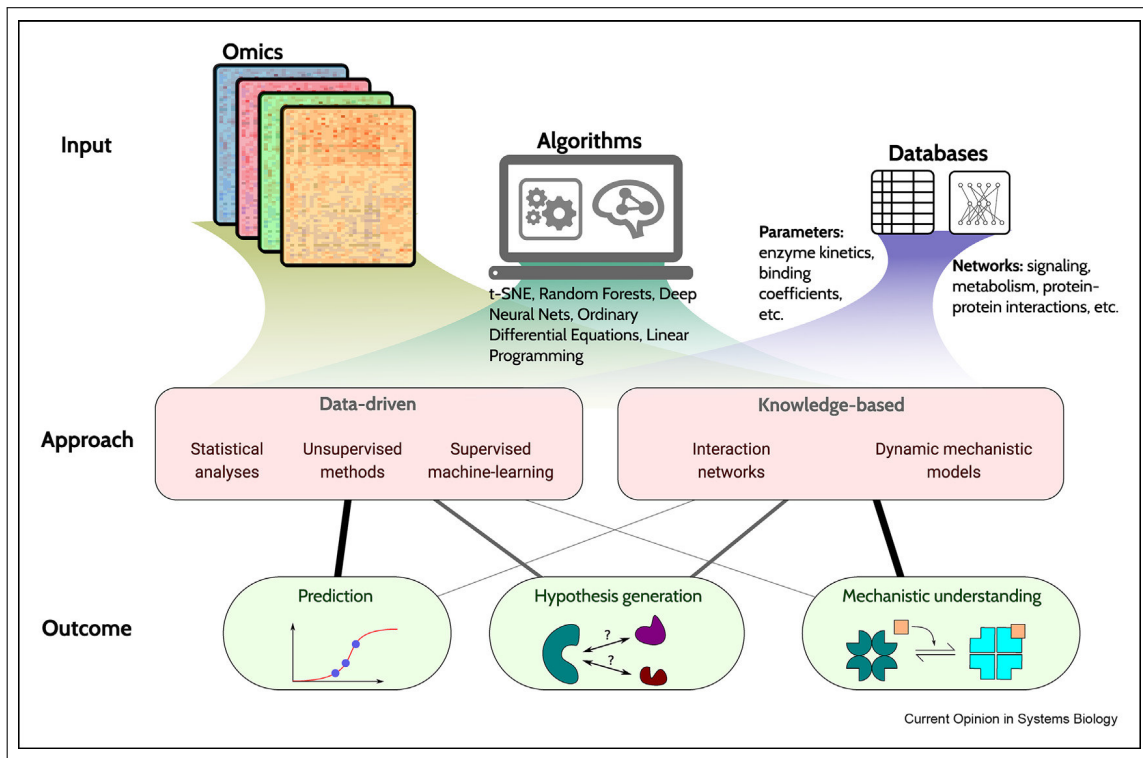


FIGURE 1.4 – Représentation schématique des deux approches majeures d’analyse de données omiques. Les approches basées sur les données (approches "top-down" ou "data driven") choisissent des modèles qui nécessitent moins de paramètres ou de connaissances préalables, offrant ainsi une représentation abstraite des interactions entre les entités. Les approches basées sur les connaissances (approches "bottom-up" ou "knowledge driven") nécessitent plus d’hypothèses et dépendent fortement des informations de la littérature pour alimenter un modèle mécanistique concret, mais elles peuvent fournir une compréhension plus précise. Ces deux approches peuvent être complémentaires en terme de méthodologie et/ou dans l’interprétation des résultats. (D’après Noor et al. 2019)

1.4 La dégradation des ARNm chez *E. coli*

Dans le cadre de ma thèse, je me suis intéressé à la dégradation des ARNm chez *E. coli*. Ce processus fait partie intégrante du cycle de vie des ARNm et son rôle dans la régulation de l’expression génique est souvent sous-estimé [Kim and Jacobs-Wagner, 2018, Régnier and Arraiano, 2000, Swain et al., 2002]. Il participe à l’élimination des transcrits superflus ou défectueux et contribue au recyclage du pool de ribonucléotides dans les cellules. En général, la transcription est vue comme l’unique responsable des variations de messagers dans les cellules. Cependant, de nombreux travaux ont montré ces dernières années que la dégradation des ARNm participe également à la régulation de la synthèse protéique [Esquerré et al., 2015]. La concentration intracellulaire des ARNm résulte en effet d’une balance entre transcription et dégradation (la dilution par la croissance cellulaire étant négligeable). Dans cette section, nous décrivons dans un premier temps les acteurs principaux de la dégradation des ARNm chez *E. coli*, le processus biologique en lui-même, ses régulations et son couplage avec d’autres processus cellulaires comme la transcription et la traduction.

1.4.1 La RNase E : acteur majeur de la dégradation des ARNm

Le mécanisme de dégradation des ARNm et la machinerie de dégradation sont relativement bien décrits chez *E. coli* [Mackie, 2013, Arraiano et al., 2010, Richards and Belasco, 2019]. Chez cet organisme, la RNase E est l'acteur central de la dégradation des ARNm. Elle est également impliquée dans la maturation des ARN. Elle initie ainsi le processing des petits ARN et de la majorité des pre-tRNAs. Elle participe aussi à la maturation des ARN ribosomiaux. Ce rôle important dans la maturation des ARN explique le fait que le gène *rne* codant cette enzyme est essentiel chez *E. coli*. La structure primaire d'un monomère Rne (1061 résidus d'acides aminés) est représentée dans la figure 1.5(A). La partie catalytique se trouve dans le domaine N-terminal (NTH) couvrant les résidus 1–529. La partie C-terminale (CTH), des résidus 530 à 1061, est le domaine d'interaction macromoléculaire. Il contient des sites d'ancrage dans la membrane (segment A) et des sites de liaison pour l'hélicase RhlB, l'énolase et la polynucléotide phosphorylase (PNPase). L'association de la RNase E avec ces différentes protéines forme un complexe appelé dégradosome [Callaghan et al., 2005, Carpousis, 2007, Bandyra et al., 2013]. *In vivo*, la RNase E existe sous la forme d'un tétramère composé d'un dimère de dimères se fixant sur la membrane interne (Figure 1.5 (B)) [Strahl et al., 2015, Khemici et al., 2008].

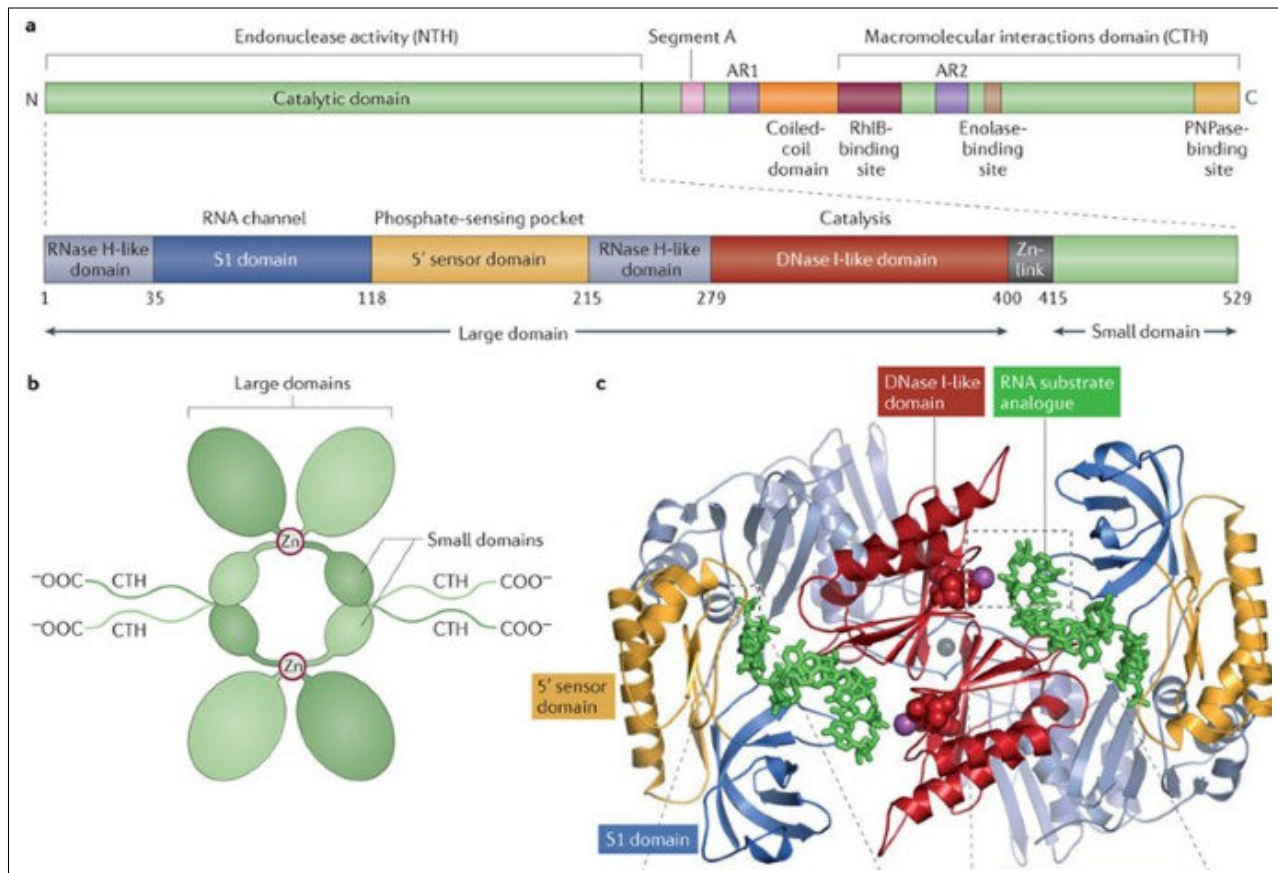


FIGURE 1.5 – (A) Structure primaire d'un monomère Rne. (B) Représentation de la forme tétramérique de la RNase E. (C) Structure cristalline (tertiaire) d'un dimère Rne en complexe avec un analogue de substrat. (Adaptée de Mackie 2013).

1.4.2 Mécanismes généraux de la dégradation des ARNm chez *E. coli*

La dégradation est initiée par un clivage endonucléolytique des ARNm, impliquant principalement la RNase E. Dans des cas moins fréquents, d'autres RNases peuvent intervenir, comme la RNase III ou la RNase G. Le clivage endonucléolytique est suivi de la conversion des fragments coupés en mononucléotides par des 3'-exoribonucléases (cf. Figure 1.6) [Mackie, 2013, Arraiano et al., 2010, Richards and Belasco, 2019]. Deux mécanismes de reconnaissance des substrats par la RNase E ont été identifiés. Dans la voie d'entrée 5' dépendante (Figure 1.6(A)), la RNase E se fixe à l'extrémité 5', ce qui permet d'aligner correctement le site de clivage sur l'ARNm dans le site actif de l'enzyme. Cela nécessite au préalable la conversion par la RppH de l'extrémité 5' triphosphate en extrémité 5' monophosphate (Celesnik et al., 2007; Deana et al., 2008; Garrey and Mackie, 2011). Dans le mode d'entrée direct (Figure 1.6(B)), la RNase E se fixe directement à une région structurée de l'ARNm, ce qui permet l'alignement du site de clivage dans le site catalytique indépendamment de la nature de l'extrémité 5' [Bandyra et al., 2018]. Ce mode de clivage semble le plus fréquent [Clarke et al., 2014] Richards et al., 2012), mais les deux mécanismes ne sont pas exclusifs [Bandyra et al., 2018].

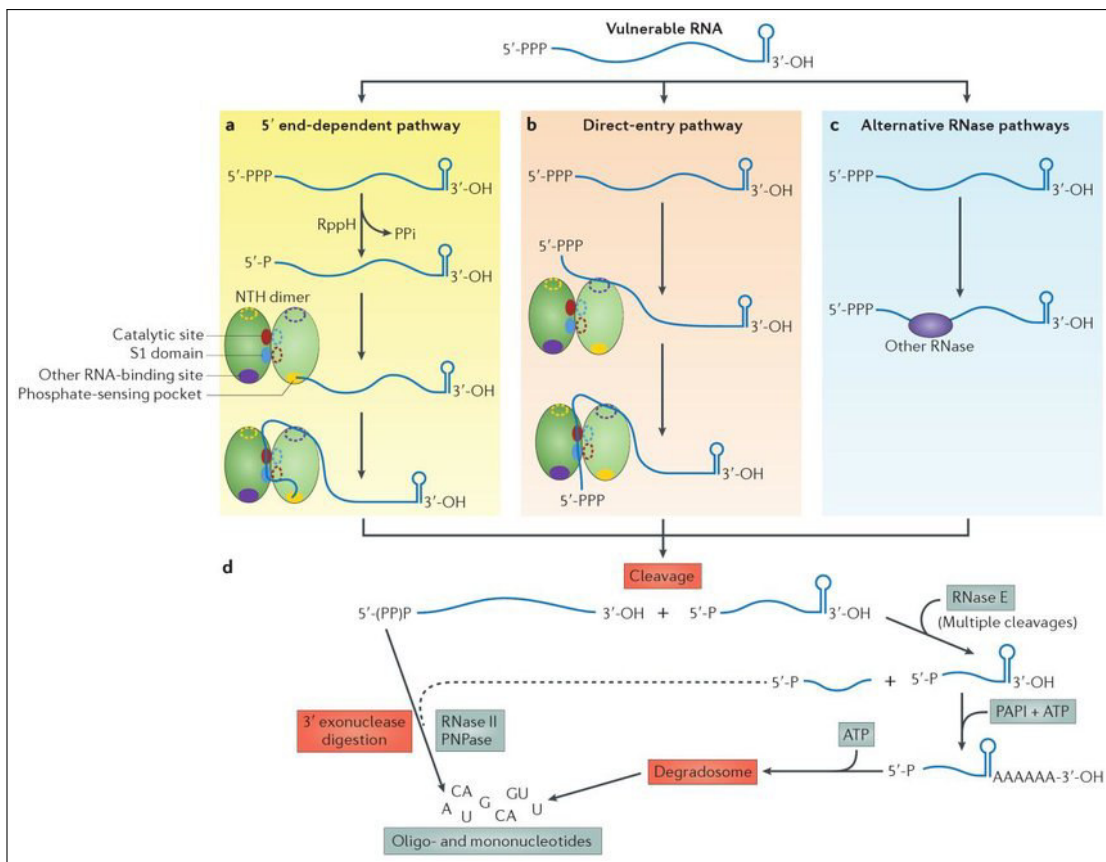


FIGURE 1.6 – Mécanismes généraux de dégradation des ARNm. **a)** La "voie d'entrée 5'-dépendante" de la RNase E. **b)** "Entrée directe" de la RNase E. **c)** Initiation de la dégradation par d'autres endoribonucléases. **d)** Dégradation des fragments engendrés par les étapes précédentes faisant intervenir les exonucléases. (D'après Mackie 2013)

1.4.3 Régulation de la dégradation des ARNm chez *E. coli*

Plusieurs mécanismes de régulation modulent la vitesse de dégradation des ARNm. Par exemple la concentration et l'activité de la RNase E sont régulées. Un mécanisme d'auto-régulation contrôle le niveau intracellulaire de la protéine, en contrôlant la dégradation de son propre ARNm [Jain and Belasco, 1995]. Les protéines régulatrices RraA et RraB réprimeraient aussi l'activité de la RNase E *in vivo* lorsqu'elles sont sur-exprimées notamment durant la phase stationnaire de croissance, mais leur rôle physiologique exact reste à établir [Lee et al., 2003]. La réaction de dégradation est elle-même régulée. Par exemple, un mécanisme d'inhibition commun à un grand nombre d'ARNm est leur protection par les ribosomes en cours de traduction. Deux cas sont envisageables selon le mode d'action de la RNase E (cf. figure 1.7). Dans le cas d'un accès direct de l'enzyme (figure 1.7(A)), la traduction ininterrompue de la région codante par des ribosomes étroitement espacés peut masquer les sites de clivage voisins et empêcher la fixation de la RNase E. Dans le cas d'une reconnaissance de l'extrémité 5', la présence d'un ribosome au niveau du RBS ou d'une tige-boucle à l'extrémité 5' peut empêcher l'accès de l'enzyme et protéger ainsi l'ARNm d'une attaque.

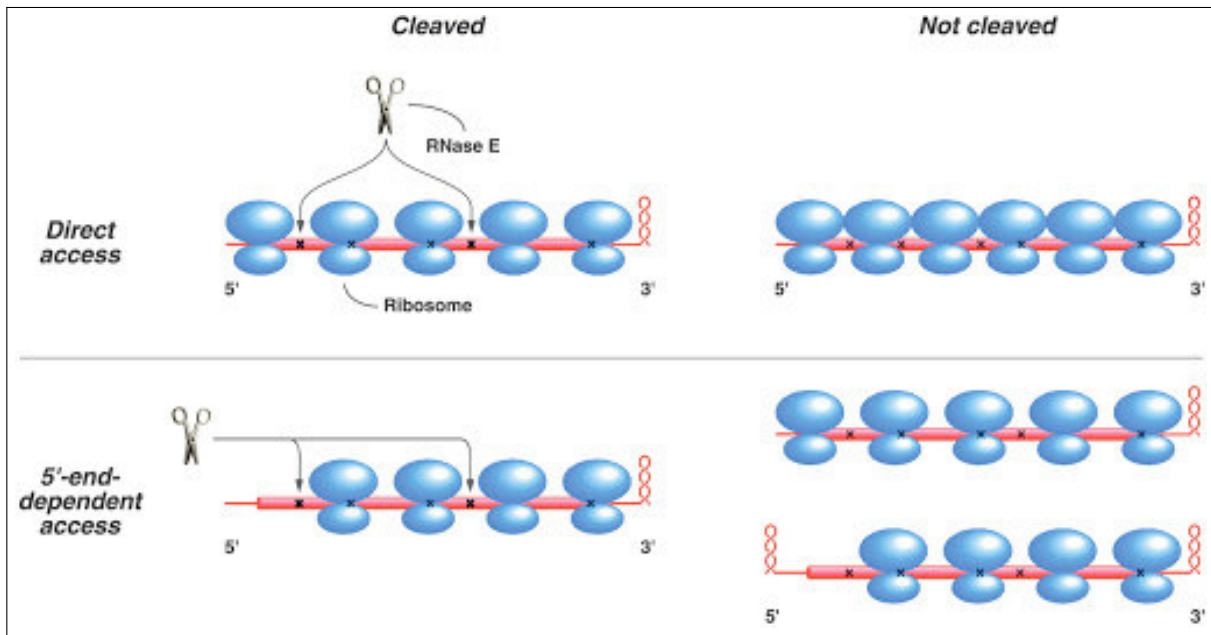


FIGURE 1.7 – Représentation schématique de la régulation du processus de dégradation des ARNm par la protection des ribosomes dans le cas d'un accès direct (panneau du haut) ou dans le cas d'un accès par l'extrémité 5' (panneau du bas). Légende : (ciseaux) RNase E; (cylindre) région codant la protéine; (X) site de clivage potentiel de la RNase E. Les flèches identifient les sites de clivage accessibles. (D'après Deana and Belasco 2005)

D'autres facteurs de régulation comme les petits ARN contrôlent la fixation de la RNase E aux ARNm bactériens [Richards and Belasco, 2019]. Ainsi, en réponse à un stress, la fixation d'un petit ARN (sRNA) à un ARNm au niveau du RBS peut inhiber l'initiation de sa traduction, ce qui est souvent associé à une déstabilisation accrue de l'ARNm (Figure 1.8(a-c)). C'est le mécanisme de régulation le plus fréquent. Dans le cas d'une activation de la traduction et de l'inhibition de la dégradation, le petit ARN agit en se fixant dans la région 5' non traduite, en amont du RBS (Figure 1.8(d)).

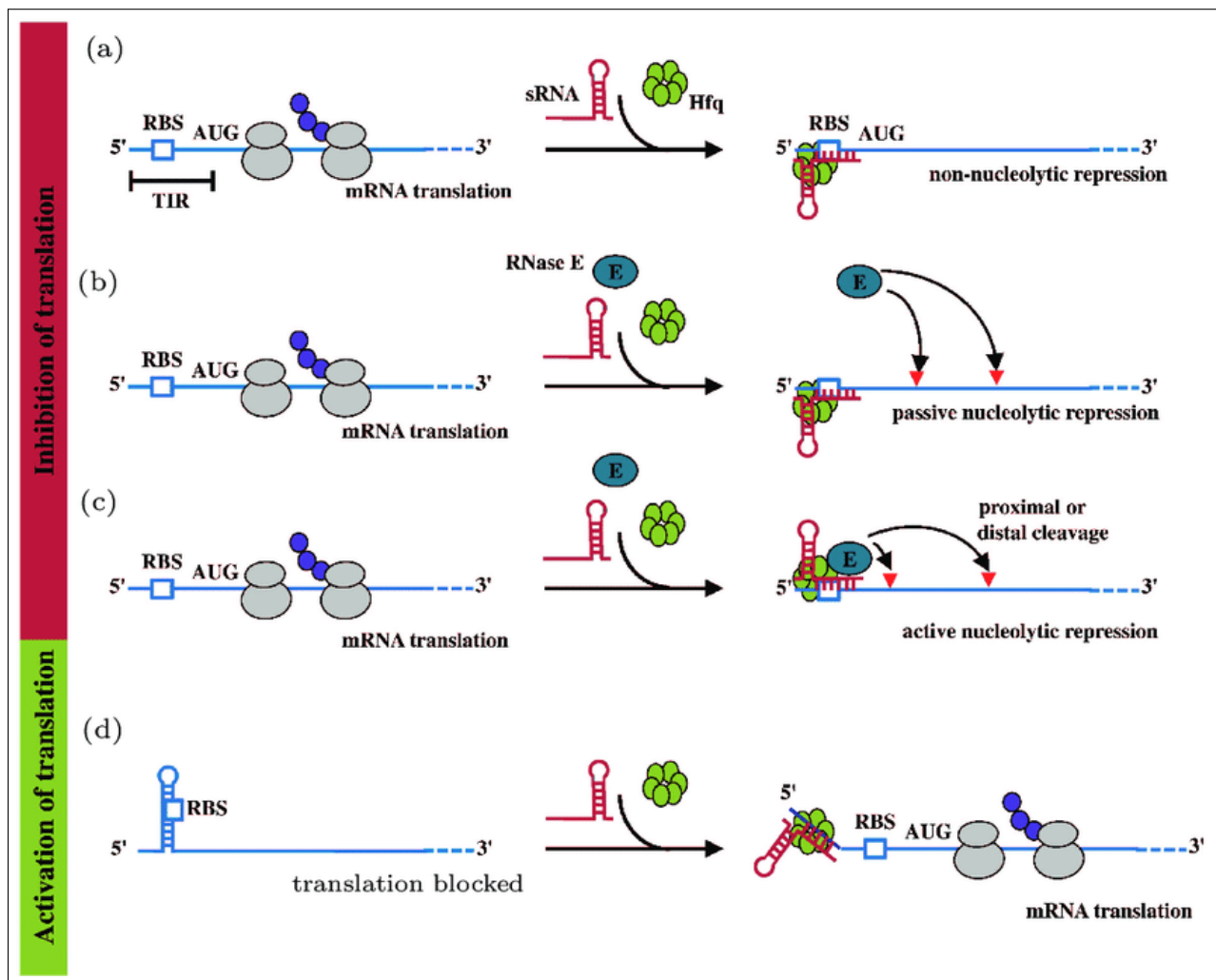


FIGURE 1.8 – Régulation de la traduction et dégradation par les petits ARN. (a) Le sRNA bloque la traduction de l'ARNm en se liant directement au site de liaison du ribosome (RBS) sur l'ARNm. (b) et (c) La liaison du sRNA sur l'ARNm séquestre le ribosome, ce qui laisse les sites d'endonuclease exposés à la dégradation par la RNase E de manière passive ou active. (d) La région RBS de l'ARNm est enfermée dans une structure secondaire en l'absence du sRNA qui inhibe la traduction de l'ARNm. La liaison du sRNA libère le RBS pour la traduction par le ribosome. Les ovales gris indiquent les ribosomes impliqués dans la traduction de l'ARNm. Les cercles violets indiquent les protéines générées. AUG est le codon d'initiation de la traduction. (D'après Xiao-Xue et al. 2015).

Plus d'une dizaine de petits ARN ont été expérimentalement identifiés chez *E. coli* [Kessler et al., 2016] et chacun d'entre eux a de multiples cibles. La régulation par les petits ARN implique souvent la protéine chaperone Hfq, qui facilite la liaison du sRNA à l'ARNm (pour revue, [Santiago-Frangos and Woodson, 2018, Jørgensen et al., 2020]) ou la protéine ProQ ([Melamed et al., 2020]). La protéine CsrA (Carbon Storage Regulator A) est un autre régulateur post-transcriptionnel [Romeo and Babitzke, 2018]. Sa fixation à l'ARNm va également empêcher la fixation des ribosomes et stimuler la dégradation lorsque son site de reconnaissance se trouve au niveau du RBS. Des cas de stabilisation de l'ARNm par CsrA ont aussi été décrits, lorsque la fixation de la protéine bloque l'accès de la RNase à son site de clivage. L'activité de CsrA est régulée par séquestration de la protéine par les petits ARN CsrB/C

et CsrD.

1.4.4 Couplage de la dégradation avec d'autres processus cellulaires

Les premières études de la régulation de la dégradation ont vite pu établir que ce processus est étroitement lié aux autres mécanismes cellulaires. Ainsi, comme on a pu le voir, traduction et dégradation sont étroitement connectées par la présence sur les ARNm de ribosomes en cours de traduction. Des couplages étroits avec la transcription, le métabolisme et la localisation cellulaire ont également été observés. Ils s'influencent mutuellement, ce qui ajoute une couche de complexité importante pour la compréhension de la physiologie bactérienne (cf. figure 1.9).

Une étude bibliographique plus approfondie des couplages a fait l'objet d'un article de revue en cours de soumission dont je suis deuxième auteur 5.2. Ma contribution la plus importante à cet article concerne la section 3 sur les approches de biologie des systèmes pour analyser la dégradation des ARNm. Ces approches sont détaillées dans la section suivante. J'y décris notamment les modèles mathématiques employés pour étudier la cinétique de dégradation, au côté des techniques expérimentales de suivi de la disparition des ARNm.

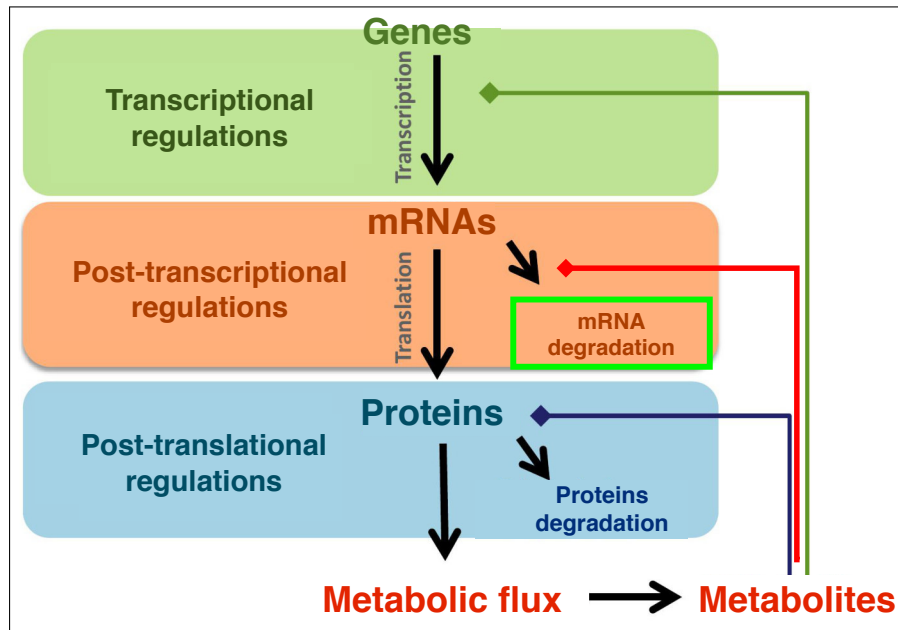


FIGURE 1.9 – Représentation schématique des différents régulations inter-connectées impactant le processus de dégradation des ARNm

1.5 Modélisation de la dégradation des ARNm

1.5.1 Modélisation du mécanisme général

La biologie des systèmes fournit des approches puissantes pour aborder la complexité des interactions entre la dégradation de l'ARNm et d'autres processus cellulaires. À cet égard, la

modélisation mathématique a longtemps été un outil compagnon des études expérimentales sur la cinétique de disparition de l'ARNm, que ce soit pour extraire les taux de dégradation des données ou pour aider à l'intégration et à l'analyse de nouvelles découvertes expérimentales. Nous décrirons le type de données obtenues dans le chapitre 2. Dans cette section, nous illustrerons avec quelques exemples ce qui a été réalisé jusqu'à présent sur la modélisation mathématique de la dégradation de l'ARNm et son interaction avec d'autres processus cellulaires.

Le modèle le plus simple et le plus largement utilisé décrit le changement dans le temps de la concentration $m(t)$ d'un ARNm donné comme la différence entre son taux de synthèse, v_s , et son taux de dégradation v_d . Cette évolution temporelle est décrite à l'aide d'une équation différentielle ordinaire (EDO) :

$$\frac{m(t)}{dt} = v_s - v_d. \quad (1.1)$$

Le choix d'un modèle déterministe peut s'expliquer par le fait que tous les ARNm transcrits à partir du même gène ont le même sort. Dans le cas d'un arrêt de transcription ($v_s = 0$), l'équation 1.1 se simplifie en :

$$\frac{m(t)}{dt} = -v_d. \quad (1.2)$$

La vitesse de dégradation (v_d) est le plus souvent décrite par une réaction du premier ordre avec une constante de dégradation k_{deg} , transformant l'équation 1.2 en équation 1.3 :

$$\frac{m(t)}{dt} = -k_{deg} \times m(t). \quad (1.3)$$

Cette description de la vitesse de dégradation dans l'équation 1.3 est très souvent employée pour des raisons pratiques car elle possède une solution analytique :

$$m(t) = m(0) \times e^{(-k_{deg} \times t)}. \quad (1.4)$$

Elle reflète le fait que le profil de dégradation ressemble à une courbe exponentielle décroissante, comme le montre l'exemple de l'ARNm *thrL* dans la figure 1.10, extrait des données de Esquerré et al. [2014]. L'ajustement de cette fonction exponentielle aux courbes de décroissance mesurées expérimentalement permet la détermination du paramètre k_{deg} et de la demi-vie de l'ARNm : $t_{1/2} = \log(2)/k_{deg}$ (figure 1.10). Ce modèle exponentiel et ses différentes variantes est utilisé depuis des décennies pour estimer les demi-vies individuelles d'ARNm, par exemple à partir de mesures de qRT-PCR ou de données à haut débit (Bernstein et al. 2002, Nouaille et al. 2017 ; pour revue, Laguerre et al. 2018). Plus récemment, des variants de ce modèle ont été développés de façon à prendre en compte le délai avant dégradation attribué à l'effet de la rifampicine et un plateau dans les données, correspondant à une concentration d'ARNm résiduel en fin d'expérience [Chen et al., 2015, Moffitt et al., 2016, Laguerre et al., 2018].

Le modèle exponentiel suppose que la dégradation de l'ARNm a lieu indépendamment des autres processus cellulaires et en une seule étape (l'étape de fixation, de coupure initiale puis la dégradation complète de l'ARNm sont indifférenciés). Cependant l'observation des données disponibles ont montré que la cinétique de dégradation ne suit pas forcément un profil exponentiel. D'autres modèles ont alors fait leur apparition, représentant la dégradation d'un

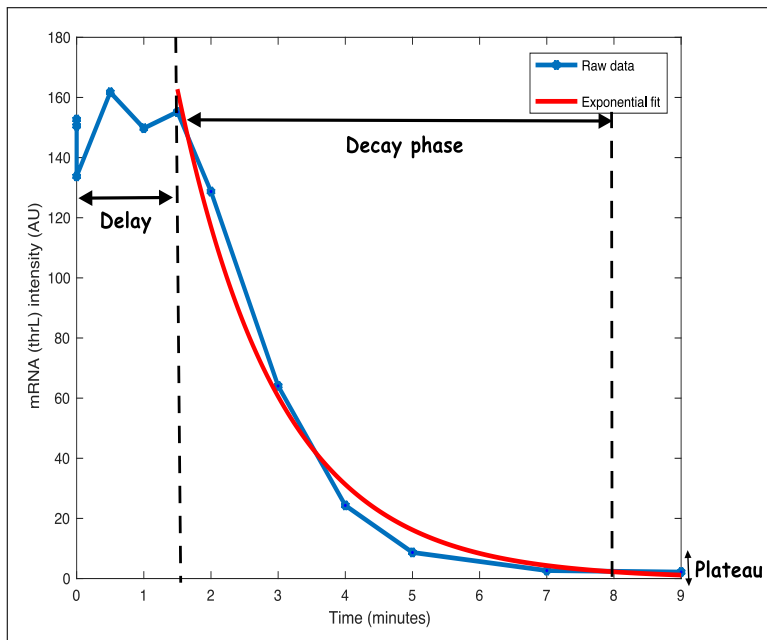


FIGURE 1.10 – Exemple de profil de dégradation de l'ARNm *thrL* (ligne bleue) et son fit par une fonction exponentielle décroissante (ligne rouge) [Esquerré et al., 2014]

ARNm comme un processus à plusieurs étapes [Deneke et al., 2013a] en utilisant la théorie des chaîne de Markov (figure 1.11 B). Pour des raisons pratiques, ce type de modèle se limite le plus souvent à deux étapes (états) [Sin et al., 2016] permettant d'obtenir une expression analytique de la loi stationnaire comme une combinaison de deux fonctions exponentielles. Cela facilite notamment le fit aux données expérimentales.

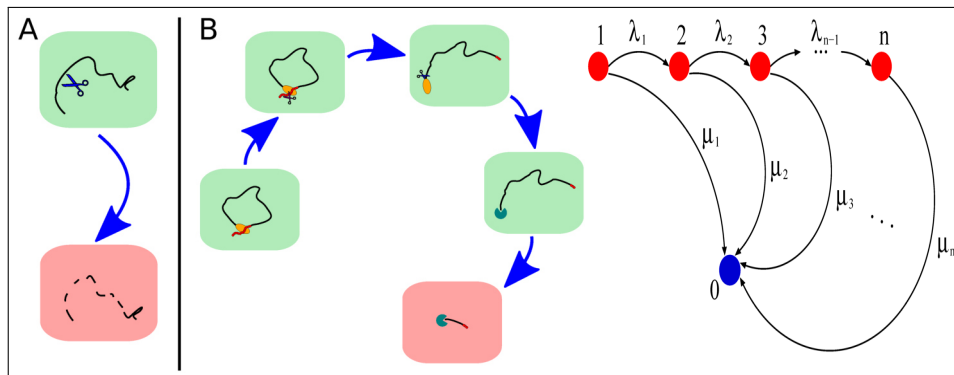


FIGURE 1.11 – **A.** Processus de dégradation d'un ARNm en une seule étape. **B.** Processus de dégradation d'un ARNm en plusieurs étapes. La figure de droite représente la modélisation d'une chaîne de Markov du processus à plusieurs étapes. (D'après Deneke et al. 2013a)

1.5.2 Modélisation du couplage de la dégradation avec d'autres processus cellulaires

La prise en compte des relations complexes de la dégradation de l'ARNm avec le métabolisme, la transcription et la traduction a poussé les efforts de modélisation au développement

de modèles de complexité variable. Un exemple bien étudié est le couplage entre dégradation et traduction des ARNm. Les modèles de traduction-dégradation décrivent la dynamique des ribosomes le long de l'ARNm avec une variété de formalismes (déterministe ou stochastique) [Deneke et al., 2013b, Mier-y Terán-Romero et al., 2013]. L'effet stabilisant des ribosomes conduit à des profils de dégradation prédits divergeant du profil exponentiel. Ces prédictions décrivent mieux la variété de profils observés en transcriptomique dynamique chez *E. coli* [Deneke et al., 2013b]. Mier-y Terán-Romero et al. [2013] modélisent la traduction comme un flux de ribosomes se déplaçant sur l'ARNm à l'aide de la loi d'action de masse pour obtenir un modèle d'équation différentielle avec délais. Les mécanismes inclus dans ce modèle sont représentés schématiquement dans la figure 1.12.

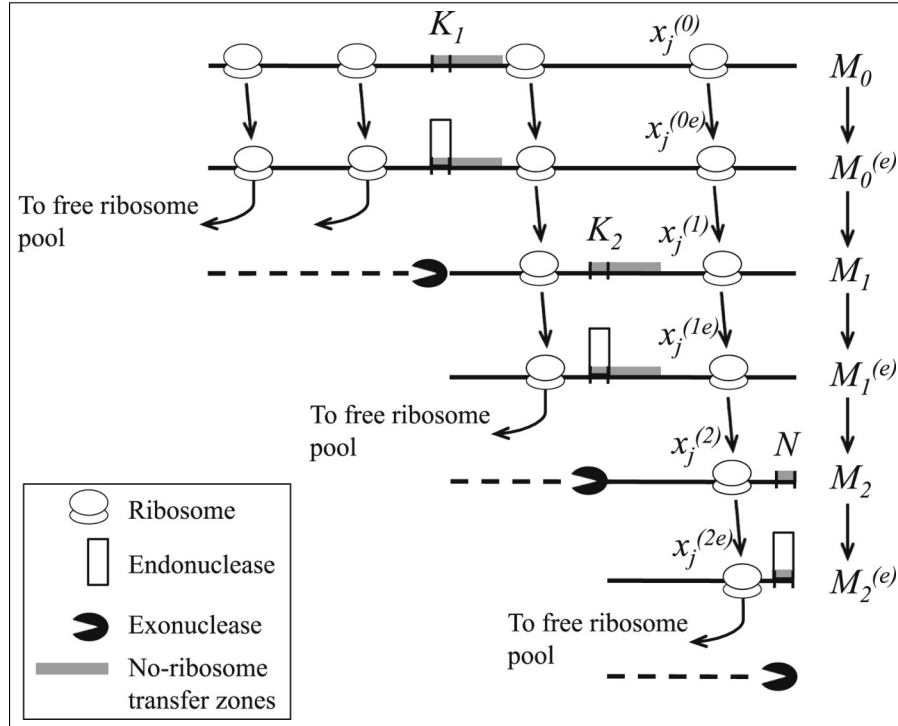


FIGURE 1.12 – Diagramme schématisant le mécanisme de dégradation couplé à la traduction par l'effet stabilisateur des ribosomes. Les variables $M_i = M_i(t)$ et $M_i(e)$ et $M_i(e)(t)$, avec $i = 0, 1, 2$ représentent la concentration des brins d'ARNm à différents stades de dégradation ; les quantités $x_j^{(l)} = x_j^{(l)}(t)$, pour $\downarrow = 0, 0e, 1, 1e, 2, 2e$, représentent les probabilités d'occupation des codons. (D'après Mier-y Terán-Romero et al. 2013)

Les modèles ci-dessus ont permis l'analyse du couplage de dégradation de l'ARNm à la transcription et/ou à la traduction. Une intégration plus poussée de ces processus a aidé à étudier différents mécanismes comme la propagation du bruit lors de l'expression de protéines bactériennes et l'apparition de "bursts" dans leur production [Friedman et al., 2006, McAdams and Arkin, 1997, Paulsson, 2005, Kim and Jacobs-Wagner, 2018]. Par exemple dans [Kim and Jacobs-Wagner, 2018], les auteurs ont étudié les effets de la dégradation de l'ARNm et des pauses transcriptionnelles sur le bruit d'expression des protéines. A l'aide de ce modèle, les auteurs ont montré que la courte durée de vie des ARNm bactériens facilite la production efficace de protéines. Pour modéliser ce couplage les auteurs généralisent et complexifient le modèle TASEP (Totally Asymmetric Simple Exclusion Process) communément utilisé pour modéliser la traduction (pour revue [Zur and Tuller, 2016]). Une description schématique du

modèle est disponible dans la figure 1.13.

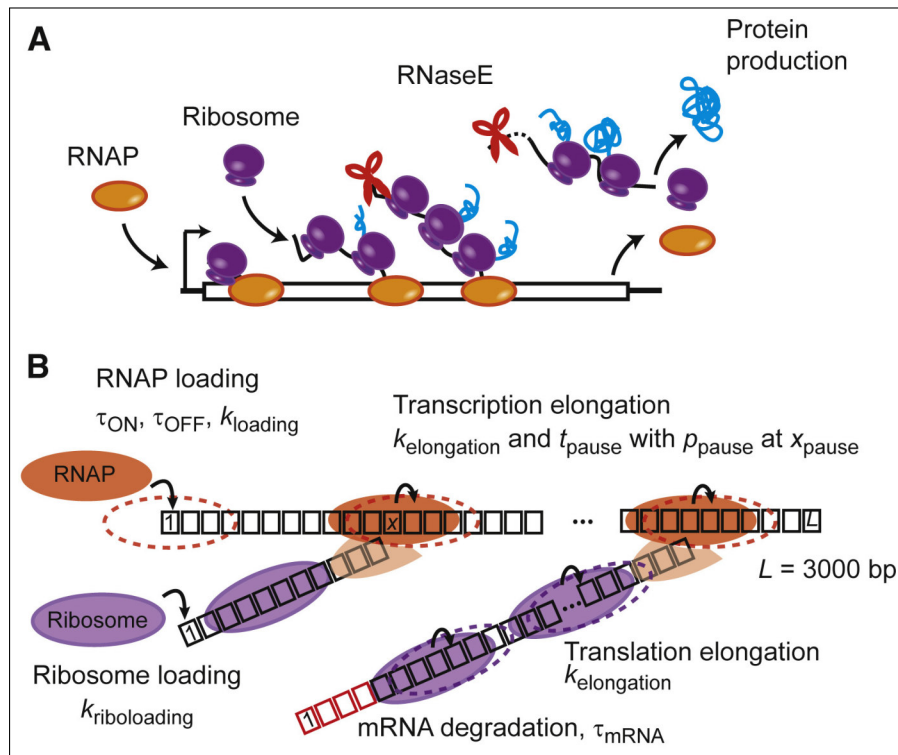


FIGURE 1.13 – Modèle intégré d'expression génique. (A) Couplage entre transcription, traduction et dégradation de l'ARNm chez *E. coli*. L'ARN polymérase (RNAP) se lie au promoteur et démarre la transcription. Une fois la séquence RBS transcrite, les ribosomes s'y fixent et commencent à traduire. La dégradation de l'ARNm par la RNase E peut commencer sur l'ARNm naissant. Le premier ribosome maintient le contact avec le RNAP tout au long de l'allongement. Dans ce modèle, une protéine complète est produite immédiatement après qu'un ribosome atteint la fin de l'ARNm. (B) Schéma montrant les différentes étapes incluses dans le modèle basé sur un modèle TASEP de l'expression génique bactérienne. La transcription commence à la première paire de bases de l'ADN, puis vient l'élongation par la RNAP. La traduction commence à la première base de l'ARNm (considérée ici comme le RBS), et l'élongation de la traduction se produit par le ribosome évoluant sur chaque nucléotide le long de l'ARNm. La dégradation de l'ARNm commence à partir de l'extrémité 5' de l'ARNm et se poursuit en même temps que le mouvement du dernier ribosome sur le transcrit. Les ribonucléotides dégradés sont représentés en rouge. (D'après Kim and Jacobs-Wagner 2018)

D'autres modèles (déterministes ou stochastiques) ont examiné quantitativement la régulation de la dégradation de l'ARNm par les petits ARN lors de réponses aux stress, à travers la modulation de leur concentration, l'implication de la protéine HFQ ou leur interaction avec les ribosomes [Levine et al., 2007, Arbel-Goren et al., 2016, Adamson and Lim, 2011, Lavi-Itzkovitz et al., 2014]. Dans [Lavi-Itzkovitz et al., 2014] par exemple, des expériences quantitatives et des modélisations stochastiques ont révélé que les ribosomes liés facilitent la fixation d'un petit ARN chez *E. coli*. L'augmentation du taux de traduction renforce ainsi la répression exercée par les sRNA RyhB, OmrA et DsrA, ce qui conduit à une déstabilisation de l'ARNm cible (cf. figure 1.14).

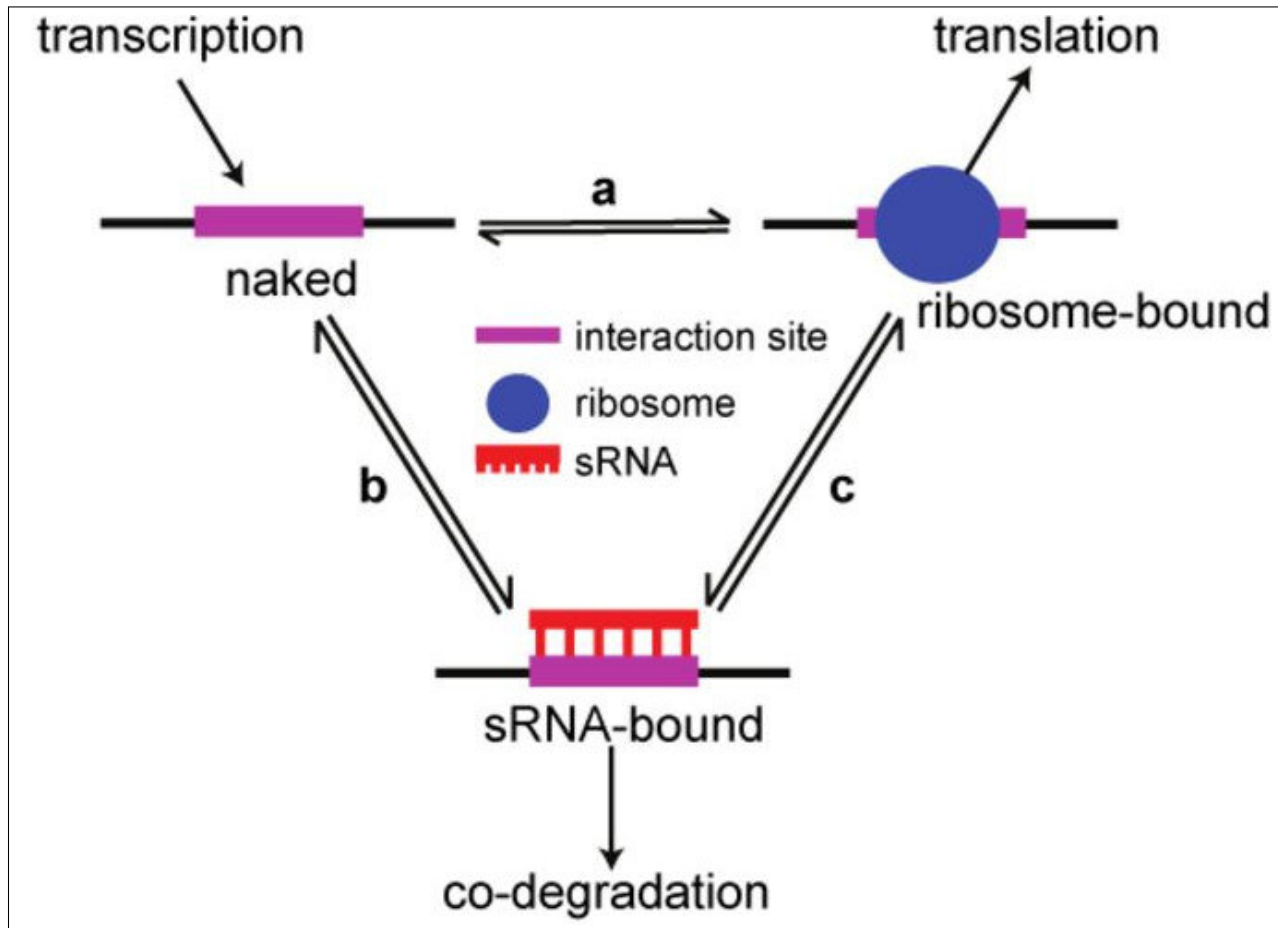


FIGURE 1.14 – Modèle à trois états pour l'interaction des ribosomes et petits ARN avec l'ARNm. Le ribosome (a) ou le petit ARN (sRNA) (b) peuvent se lier aux ARNm nus transcrits, conduisant respectivement à la traduction de l'ARNm ou à la co-dégradation du complexe sRNA-ARNm. Les ribosomes liés et le complexe sRNA peuvent interagir, directement ou indirectement (c). D'après [Lavi-Itzkovitz et al., 2014]

1.6 Problématique et plan

L'apparente simplicité du processus de dégradation des ARNm cache une réalité complexe et amène à sous-estimer le rôle important de ce mécanisme dans la régulation génique. Comme nous l'avons montré dans la section précédente, l'étude de ce processus nécessite à la fois des données expérimentales à l'échelle du génome et un modèle mathématique permettant d'étudier son fonctionnement. Le plus souvent, un modèle phénoménologique de type exponentiel est utilisé pour analyser les profils de dégradation. L'interprétation mécanistique et l'intégration de connaissances dans ce type de modèle se sont avérées jusqu'à présent peu fructueuses, car un seul paramètre, la constante de dégradation, résume l'ensemble de la cinétique de dégradation. Ceci limite l'interprétation biologique et n'exploite pas complètement l'information contenue dans les données. En outre, les modifications récentes de ces modèles (*cf* Section 1.5.1), avec l'ajout d'un délai ou d'un plateau, complique l'interprétation des résultats et leur comparaison avec d'autres études de la littérature. En effet, ces modifications ont un impact non négligeable sur la constante de dégradation et donc le temps de demi-vie, qui est la variable utilisée pour analyser et interpréter la vitesse de dégradation [Laguerre et al., 2018]. Idéalement, dans cette thèse, nous voudrions confronter aux données un modèle mécanistique, permettant de nous donner de nouvelles informations sur les mécanismes de régulation régissant la cinétique de dégradation des ARNm.

Les problèmes d'identification des paramètres à partir des données expérimentales ont justifié l'utilisation des modèles exponentiels plus simples. Cependant, de nombreux profils sont laissés de côté lors de l'estimation de la constante de dégradation, généralement lorsqu'ils divergent de l'exponentielle (par exemple, Bernstein et al. 2002, Esquerré et al. 2014). Ceci réduit de façon drastique le nombre de profils de dégradation étudiés pour lesquels on peut obtenir une estimation robuste et poursuivre l'analyse, comme de l'enrichissement fonctionnel pour identifier des fonctions cellulaires affectées par la stabilité de groupes d'ARNm. Par exemple, sur les 4254 profils de dégradation obtenus par Thomas Esquerré pendant sa thèse, seuls 2954 ont pu être fittés, soit 70% des profils. Il s'agit pourtant d'un jeu de données plutôt exhaustif, ayant repoussé l'état de l'art du domaine. Cela s'explique en partie par une limite de la méthode expérimentale choisie pour mesurer la concentration résiduelle des ARNm, lorsque les concentrations sont faibles, mais pas uniquement. On observe dans ces jeux de données des profils plus plats avec des concentrations se détachant de la limite de détection qui pourraient refléter un mécanisme biologique. Cela se complique lorsqu'on étudie l'effet de l'environnement ou de mutations sur le temps de demi-vie, ce qui nécessite d'avoir pu estimer une constante de dégradation pour chaque ARNm dans chaque condition environnementale. Cela réduit encore plus le pool d'ARNm pour lequel le changement de demi-vie peut être étudié. Sachant que de plus en plus de données de stabilité des ARNm à l'échelle de la cellule sont disponibles dans la littérature, il serait intéressant de pouvoir disposer d'une méthode permettant une analyse exhaustive des données.

L'approche actuelle d'estimation repose sur la détermination du temps de demi-vie, séparément pour chaque ARNm dans chaque condition. Les auteurs du domaine calculent a posteriori les paramètres de la population d'ARNm comme la moyenne ou médiane et la variance du temps de demi-vie. Cette méthodologie engendre pourtant des biais dans l'interprétation et le plus souvent surestime la variabilité au sein de la population [Dharmarajan et al., 2019, González-Vargas et al., 2016]. Une méthode d'estimation prenant en compte ces aspects de population d'ARNm, afin d'obtenir une bonne estimation des paramètres de la population, serait donc un atout. Un dernier aspect concerne la dimension des données qui engendre des problèmes computationnels. Cela nécessite une attention particulière lors de l'estimation des paramètres et de disposer d'outils adaptés.

Tous ces éléments sont autant de verrous, complexifiant l'analyse des données omiques, qui engendrent de nouveaux défis méthodologiques qui sont au coeur de la problématique de ma thèse. Comment exploiter au maximum les données omiques disponibles pour répondre à des questions biologiques? Peut-on en apprendre plus sur les régulations de la dégradation des ARNm chez *E. coli*, en combinant l'approche basée sur les connaissances (modélisation mécanistique) avec une méthode permettant l'inférence de paramètres cinétiques? Ce questionnement est au coeur de ma thèse. Il s'agit d'un véritable défi méthodologique car peu d'équipes ont abordé ce thème de recherche jusqu'à présent.

La réponse à cette problématique est décomposée en deux parties fortement inter-connectées. La première partie sera le développement d'un modèle mécanistique de la dégradation de l'ARNm, capable d'intégrer de grands ensembles de données et d'étudier les mécanismes de régulation cellulaire. Pour cela il est nécessaire de choisir un niveau d'abstraction adapté (booléen, stochastique ou déterministe; Klipp et al. 2016) et de construire un modèle mathématique basé sur les connaissances du processus de dégradation des ARNm chez *E. coli*. Ce modèle doit être le plus parcimonieux possible pour réduire les complications liées à la dimension des données durant l'étape d'inférence des paramètres de ce dit modèle. Par la suite, nous simulerons numériquement le(s) modèle(s) avec des paramètres cinétiques les plus réalistes, pour comparer les cinétiques temporelles avec les données expérimentales et interpréter les profils de dégradation. La deuxième partie aura pour objectif d'utiliser le modèle mécanistique, préalablement construit dans la première partie, pour décrire des données expérimentales obtenues par T. Esquerré pendant sa thèse dans l'équipe BLADE. L'idée est de découvrir les principaux mécanismes de régulation ajustant les demi-vies d'ARNm dans quatre régimes différents de taux de croissance d'*E. coli* via l'inférence des paramètres cinétiques de dégradation des ARNm. Pour cela il nous faudra choisir une méthode d'inférence adaptée à nos différentes contraintes : dimension des données, normalisation, extraction d'informations au niveau de la population (ensemble des ARNm) et au niveau individuel (sous-ensemble de la population), et gestion du bruit dans les données expérimentales. Les méthodes utilisées relèvent d'approches "bottom-up" et "top-down" de la biologie des systèmes (section 1.3).

Dans le chapitre 2 de ce manuscrit, je présente les données de T. Esquerré et des résultats d'analyse statistique sans a priori des profils de dégradation afin de dégager leurs principales caractéristiques. Le chapitre 3 est consacré à la modélisation mécanistique de la dégradation des ARNm et l'utilisation du modèle dans des études de simulation numérique pour comprendre l'importance du phénomène de compétition entre ARNm pour leur fixation à la RNase E. Ces travaux ont fait l'objet d'une publication dans *The Journal Of Theore-*

tical Biology dans laquelle je suis 1^{er} auteur. Dans le chapitre 4, nous nous intéresserons à la problématique de l'estimation des paramètres du modèle mécanistique, ainsi qu'à la recherche d'approches adaptées à notre modèle et à nos données. L'application de l'approche aux données de transcriptomique dynamique chez *E. coli* est décrite dans le chapitre 5. Elle fait l'objet d'un article en préparation pour un journal de biologie dans lequel je suis 1^{er} auteur. Elle nous a notamment permis d'identifier la compétition comme un mécanisme majeur d'adaptation de la stabilité des ARNm aux changements de conditions de croissance. Dans les conditions expérimentales étudiées, les régulations spécifiques par des petits ARNm ne concernent qu'une petite partie des ARNm cellulaires. Dans le chapitre 6, je discute les résultats obtenus et présente des perspectives. L'annexe 1 inclut l'article, sur le couplage entre la dégradation et d'autres processus cellulaires, qui sera prochainement soumis et dans lequel je suis 2^e auteur.

Chapitre 2

Approches statistiques d'analyse de la dégradation des ARNm

Ce qu'on peut faire en terme d'analyse de données omiques dynamiques à l'aide de modèles mécanistiques dépend étroitement de la façon dont ces données ont été obtenues et traitées. En effet, la technique utilisée pour la mesure de concentration des ARNm et le design expérimental choisis sont des éléments importants à prendre en considération dans la modélisation. Dans ce chapitre, nous allons discuter de différentes approches top-down, basées sur des méthodes statistiques sans a priori, pour analyser des données similaires aux nôtres. Dans une première section, nous aborderons la production des données expérimentales pour l'étude de la dégradation des ARNm chez les prokaryotes. Nous nous intéresserons en particulier à la méthode utilisée par l'équipe BLADE pour produire les données étudiées dans cette thèse [Esquerré et al., 2014]. Dans une deuxième section, nous étudierons les pré-traitements des données brutes. Nous discuterons notamment des méthodes de normalisation et de correction du bruit de fond. Enfin dans une troisième et dernière partie, nous nous intéresserons aux méthodes de classification des données omiques dynamiques. A titre d'illustration, nous appliquerons à notre jeu de données une des méthodes décrites chez *E. coli*.

2.1 Mesure expérimentale de la dégradation des ARNm

Diverses techniques expérimentales peuvent être utilisées pour étudier la dégradation des ARNm au cours du temps. Nous nous intéresserons en particulier aux principales approches : les expériences basées sur un arrêt de transcription et de la mesure de la concentration résiduelle en ARNm par transcriptomique ou qRT-PCR et les expériences de pulse-chase labelling suivant le devenir d'un pool d'ARNm marqués. Ces données à l'échelle de la cellule sont encore peu abondantes dans la littérature, même pour un organisme bien caractérisé comme *E. coli* [Bernstein et al., 2002, Chen et al., 2015, Sin et al., 2016, Esquerré et al., 2014, Moffitt et al., 2016, Morin et al., 2020].

2.1.1 Quantification des ARNm par transcriptomique dynamique après arrêt de transcription

L'une des méthodes les plus répandues est l'arrêt de la synthèse des ARNm et la mesure de leur quantité résiduelle au cours du temps, par exemple par des expériences de microarray [Bernstein et al., 2002, Esquerré et al., 2014] ou de RNAseq [Chen et al., 2015]. Elles fournissent des profils de dégradation sous la forme de séries temporelles représentant l'évolution de la concentration d'ARNm au cours du temps. Dans ce type d'expérience un antibiotique, la rifampicine, est utilisé pour arrêter la synthèse de nouveaux ARNm dans les cellules bactériennes en inhibant l'élongation des transcrits par l'ARN polymérase. Cette méthode permet de découpler la dégradation de l'ARNm du processus de transcription.

Un exemple de cette procédure expérimentale est représenté en figure 2.1. Elle a été réalisée par T. Esquerré au cours de sa thèse dans l'équipe BLADE au Toulouse Biotechnology Institute [Esquerré et al., 2014]. Les données obtenues sont celles sur lesquelles j'ai travaillé au cours de ma thèse. Elles ont permis l'observation par puce à ADN de la dégradation de 4254 ARNm d'*E. coli* dans 4 conditions de croissance. Une première étape a consisté à cultiver une population de cellules à différents taux de croissance à l'aide d'un procédé de culture en continue. Ainsi 3 expériences en chemostat ont été réalisées pour les croissances à 0.1 h^{-1} , 0.2 h^{-1} et 0.4 h^{-1} et une expérience en batch a été menée pour le taux de croissance le plus élevé à 0.63 h^{-1} (il n'est pas possible de cultiver les cellules en chémostat à ce taux de croissance). La rifampicine a ensuite été ajoutée pour stopper l'élongation par l'ARN polymérase, puis plusieurs expériences de microarrays ont été menées pour l'obtention des profils de dégradation au cours du temps.

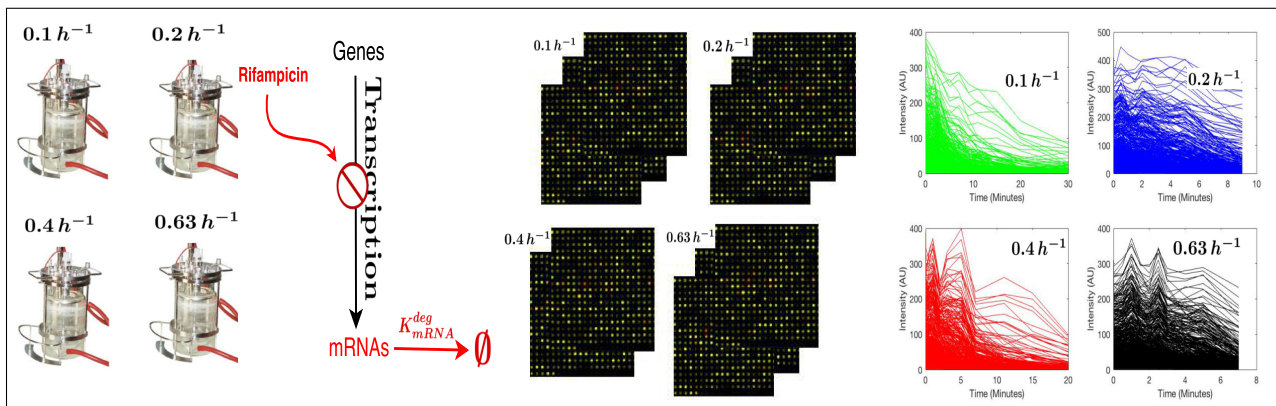


FIGURE 2.1 – Principe de la mesure expérimentale de la stabilité d'ARNm dans des cultures continues d'*E. coli*. Cette procédure a été employée par T. Esquerré pour l'obtention du jeu de données utilisé au cours de cette thèse [Esquerré et al., 2014].

Afin d'obtenir un nombre suffisant de points de mesures et pour des raisons pratiques, 3 cultures indépendantes ont été nécessaires pour chaque taux de croissance. 4 prélèvements ont ainsi été effectués pour chacune de ces 3 cultures, donnant au total 12 points de mesure par condition de croissance : 3 points de mesure avant ajout de la rifampicine appelées "temps 0" et 9 points pour décrire la phase de dégradation (cf. tableau 2.1). Un exemple représentatif de ces 3 réplicats est représenté dans la figure 2.2 pour l'ARNm *thrL* [Esquerré et al., 2014]. Le seul temps de mesure commun à ces 3 cultures est celui avant l'ajout de la rifampicine.

L'échantillonnage est le même entre les ARNm d'un même taux de croissance, mais diffère entre taux de croissance. Enfin le pas de temps n'est pas uniforme (cf. tableau 2.1).

Taux de croissances	Temps de mesure
$0.1h^{-1}$	[0, 0, 0, 2, 3, 5, 7, 9, 11, 15, 20, 30]
$0.2h^{-1}$	[0, 0, 0, 0.5, 1, 1.5, 2, 3, 4, 5, 7, 9]
$0.4h^{-1}$	[0, 0, 0, 0.5, 1, 2, 3, 5, 7, 11, 15, 20]
$0.63h^{-1}$	[0, 0, 0, 0.5, 1, 1.5, 2, 2.5, 3, 4, 5, 7]

TABLE 2.1 – Temps de prélèvement pour chaque taux de croissance

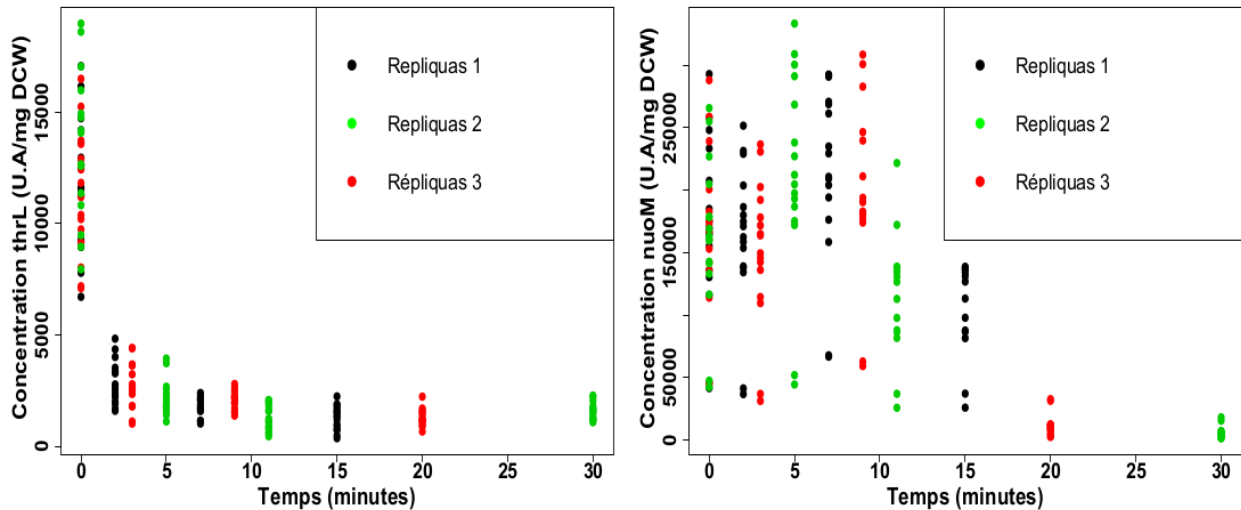


FIGURE 2.2 – Représentation graphique des données de trois réplicats composant le profil de dégradation de l'ARNm *thrL* et *nuoM* au taux de croissance $0.1h^{-1}$ [Esquerré et al., 2014]. Les différents points correspondent aux mesures d'intensité de chaque échantillon d'ARN par 16 sondes différentes

A l'issue des prélèvements, les ARN totaux sont extraits et les ARNm, quantifiés par puce à ADN grâce à l'utilisation de sondes spécifiques. Un total de 16 sondes est utilisé pour chaque échantillon. Ces sondes correspondent à un des segments différents couvrant de manière prédominante l'extrémité 5' tout en essayant d'avoir une couverture raisonnable de la région codante. Ces choix expérimentaux ont été faits dans le but d'obtenir une cinétique de dégradation la plus longue possible à chaque taux de croissance et de permettre la mesure d'ARNm complètement transcrits (en principe, on ne détecte pas les ARNm à moitié dégradés). Cette méthode expérimentale a pour avantage l'obtention d'un grand nombre de profils de dégradation, avec une couverture à l'échelle du génome. Elle est très bien maîtrisée par l'équipe BLADE qui possède une grande expertise dans le contrôle des cultures bactériennes et l'analyse des ARNm par puces à ADN. Cette méthode expérimentale est plus propice à une modélisation déterministe pour l'analyse de données car les mesures sont effectuées sur une population de cellules et les résultats générés sont des concentrations moléculaires. Cette méthode peut être cependant stressante pour la cellule, entraînant potentiellement un changement de morphologie du nucléoïde et de la distribution spatiale des ribosomes causé par l'ajout de rifampicine [Bakshi et al., 2014, Moffitt et al., 2016].

2.1.2 Méthodes expérimentales alternatives

Des méthodes alternatives de suivi de dégradation sont disponibles. On citera les expériences de qRT-PCR qui s'effectuent sur un nombre plus restreint d'ARNm après l'arrêt de transcription par ajout d'antibiotique. Elles sont souvent utilisées comme méthodes de validation expérimentale des résultats obtenus par puces à ADN ou RNAseq. Cela a été le cas du jeu de données étudié dans cette thèse [Esquerré et al., 2014]. Les expériences dites de "pulse-chase" ont un principe très différent [Sin et al., 2016]. Elles se déroulent en deux grandes phases.

Dans une première phase de "pulse" ou impulsion, les cellules sont exposées à un nucléotide marqué introduit dans le milieu de culture. Celui-ci entre dans les cellules où il est utilisé comme précurseur pour la synthèse de nouveaux ARNm. Dans la deuxième phase, la phase de "chase", le même composé sous forme non marquée est ajouté en excès, remplaçant la forme marquée. Les ARNm nouvellement synthétisés ne sont plus marqués. La quantité d'ARNm marqués pendant la phase de pulse diminue en raison des processus de dégradation. Le principe de cette approche est illustré dans la figure 2.3A à quatre stades différents d'avancement de l'expérience [Sin et al., 2016]. Au stade I (30 a.u.), le pulse vient de commencer. Les points blancs représentent la population de molécules déjà présentes dans la cellule avant l'impulsion. Toutes les molécules nouvellement synthétisées (points rouges) sont marquées et mesurables par l'expérimentateur. Alors que le pulse continue dans le stade II (60 a.u.), nous voyons plus de molécules marquées apparaître. Pendant ce temps, les molécules marquées et non marquées se dégradent. Au stade III (90 a.u.), l'impulsion est terminée depuis un certain temps. Les molécules nouvellement synthétisées à partir de ce moment ne sont plus marquées. Encore une fois, les molécules marquées et non marquées se dégradent. Dans le stade IV (200 a.u.), toutes les molécules marquées se sont dégradées. Les molécules non marquées continuent d'être synthétisées et dégradées. Dans la figure 2.3 textbfB., Sin et al. [2016] représentent les distributions d'âge des molécules marquées. Chaque densité correspond à une phase décrite dans le panneau A. Au stade I (courbe bleue), l'impulsion vient juste de commencer et toutes les molécules marquées lors de cette impulsion ont comme âge le temps écoulé depuis le début de l'impulsion. Dans le stade II (courbe verte), le pulse est appliqué depuis un certain temps; certaines molécules marquées peuvent être assez anciennes. Dans le stade III (courbe rouge), les molécules ne peuvent pas être plus jeunes que le temps écoulé depuis l'arrêt du pulse. S'il restait des molécules au stade IV, elles auraient la distribution d'âge représentée sur la courbe cyan. Cette méthode de mesure est plus propice à une modélisation stochastique du processus de dégradation car elle permet la mesure (en nombre) d'ARNm et génère des distributions d'âges. Comme pour la technique de mesure utilisant la rifampicine, l'ajout de matériel marqué et le contrôle de la durée du pulse constituent aussi une limite à l'utilisation de ce type d'expériences [Wada and Becskei, 2017].

2.2 Pré-traitement des données de transcriptome dynamique

Notre objectif est de pouvoir analyser la variabilité biologique à l'origine des différences de profil de dégradation des ARNm. Il faut donc pouvoir éliminer de nos données brutes la variabilité technique qui pourrait gêner l'interprétation des données.

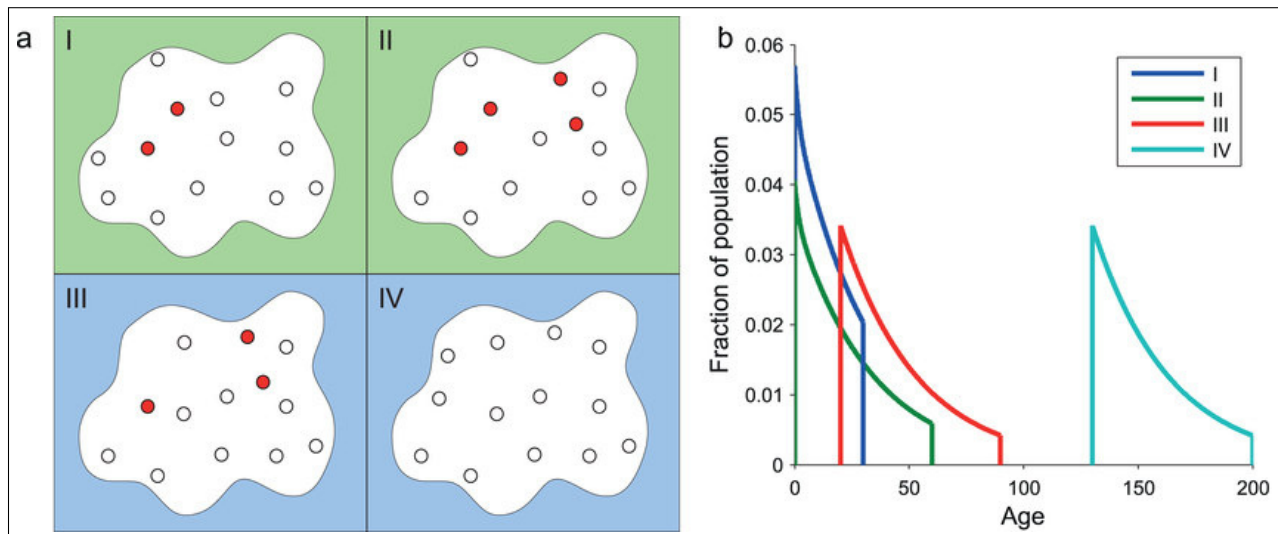


FIGURE 2.3 – **A.** Représentation des molécules dans une cellule pendant une expérience de "pulse-chase" avec une durée d'impulsion de 70 en unités arbitraires (a.u.). Quatre stades différents d'avancement de l'expérience sont représentés (I à IV). **B.** Distribution d'âge des molécules marquées. (D'après Sin et al. 2016)

Les méthodes de normalisation sont le plus souvent appliquées à des données de transcriptome en statique produites par puce à ADN indépendantes, par exemple pour la comparaison d'un mutant à un sauvage. Pour ces données statiques et pour des données dynamiques comme les nôtres, les étapes de pré-traitement des données brutes vont chercher à gérer au mieux les différentes sources de variabilité. On en distingue généralement deux types : la variabilité technique qui va dépendre du design expérimental et de la technique de mesure utilisée, et la variabilité biologique liée à l'étude d'une population de cellules ou au processus biologique étudié. Pour des données de transcriptomique (statique et dynamique), la variabilité technique va par exemple correspondre au bruit de fond et aux bruits entre les réplicats et les sondes. La normalisation de données dynamiques de transcriptomique n'est pas triviale. Ainsi, dans la revue [Bar-Joseph et al., 2012], les auteurs nous mettent en garde sur le choix des méthodes de normalisation, en particulier pour l'analyse du processus de dégradation des ARNm après arrêt de la transcription. Une étude chez la levure du couplage transcription-dégradation [Shalem et al., 2008] (manipulant donc des profils de dégradation) a en effet montré que les procédures classiques englobant une étape de correction du bruit de fond et normalisation entre réplicats et entre les sondes telles que la Robust Multi-array Average (RMA) [Irizarry et al., 2003] ne sont pas applicables car la quantité d'ARNm varie au cours du temps. Nous décrivons ci-dessous les stratégies alternatives que nous avons déployées pour prendre en compte la variabilité technique. Ce travail de pré-traitement avait été réalisé par Thomas Esquerré [Esquerré et al., 2014], mais nous avons dû le revoir pour nos besoins d'estimation des modèles mécanistiques afin d'explorer les données et d'améliorer qualité de fit

2.2.1 Correction du bruit de fond

Deux solutions principales s'offrent à nous dans notre contexte Bar-Joseph et al. 2012. La première serait d'utiliser la méthode des invariants bas qui consiste à déterminer un sous-set

de sondes (spécifiques et/ou aléatoires) que l'on considère non exprimées. Cette sélection de sous-set invariants est effectuée pour les 12 points de mesure et permet d'obtenir un ensemble de sondes classifiées comme invariants bas. La médiane de leur intensité est alors calculée. C'est l'approche utilisée dans [Esquerré et al., 2014]. Une solution équivalente et beaucoup plus facile à mettre en oeuvre consiste à calculer la médiane des intensités des sondes non spécifiques à chaque temps de mesure. Une fois que la valeur du bruit de fond commun à tous les ARNm d'un taux de croissance est déterminée, celle-ci peut être soustraite ou divisée pour supprimer le bruit des intensités spécifiques. Le choix de la méthode de suppression du bruit de fond est relativement sans impact, sauf lorsque les données sont multipliées par un paramètre de normalisation. C'est notre cas de figure comme nous le verrons en section 2.2.3.

Dans cette thèse j'ai choisi d'estimer le bruit de fond de chaque taux de croissance à l'aide des sondes non spécifiques qui sont en nombre suffisant (4013 sondes sur un total de 72037) pour obtenir une bonne estimation. Pour les mêmes raisons de qualité d'estimation, la suppression s'effectuera par soustraction. Je reviendrai sur ce point dans la sous-section 2.2.3. Pour illustration nous représentons les densités des sondes spécifiques (block 1 (rouge) et block 2 (vert)) et non-spécifiques (bleu) dans la figure 2.4. A travers cette figure nous remarquons qu'il n'y pas d'effets liés à la localisation des sondes (la densité rouge (block 1) est complètement recouverte par celle en vert (block 2)). Les densités des intensité brutes pour les autres taux de croissances sont disponibles en annexe section 5.1.

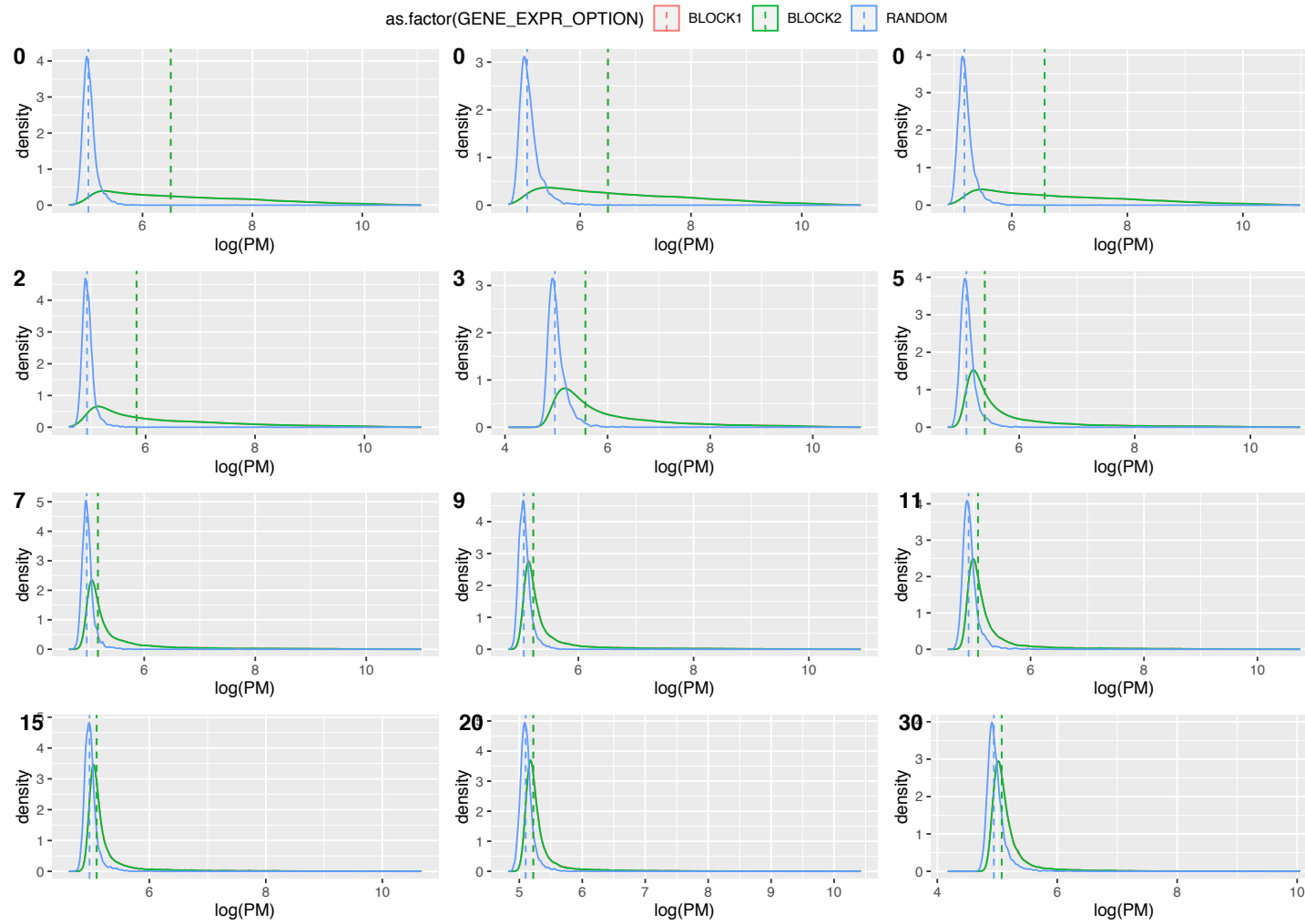


FIGURE 2.4 – Densité des intensités des sondes spécifiques en rouge pour le block 1 et en vert pour le block 2) et non spécifiques (bleu). Les valeurs de densité sont données pour chaque temps de mesure au taux de croissance $0.1h^{-1}$. Les intensités sont données en abscisse sur une échelle logarithmique

2.2.2 Prise en compte des bruits liés aux réplicats et aux sondes

La prise en compte et la réduction du bruit liés aux réplicats et aux sondes est importante. Spécialement lorsque l'on veut estimer des paramètres cinétiques d'un modèle mécanistique. En effet, si on veut que les variations de paramètres cinétiques nous renseignent sur le mécanisme de régulation, il faut préalablement s'assurer que ces variations ne proviennent pas du bruit expérimental.

La normalisation par quantiles est l'approche couramment employée pour l'analyse des puces à ADN mais, comme expliqué plus haut, elle n'est pas applicable dans notre contexte. Dans nos données, nous avons de la variabilité du fait des différents réplicats, à laquelle s'ajoute celle liée à l'emploi de 16 sondes pour chaque ARNm. Une méthode classique est de simplement prendre la médiane des intensités de ces 16 sondes. Une façon de réduire à la fois la variabilité entre les réplicats et les sondes est d'utiliser plutôt une méthode de lissage, par exemple par une fonction semi-paramétrique de type spline basée sur des polynômes. C'est l'approche que nous avons privilégiée dans notre travaux car elle améliore grandement la qualité des estimations de paramètres (cf. chapitre 4). La fonction est définie par morceaux par des polynômes sur des intervalles entre noeuds. Le degré de la spline est défini par le polynôme de plus haut degré. Par exemple, dans le cas d'une spline de degré 1, les noeuds sont tous reliés par une droite. Dans notre cas, nous utilisons 6 noeuds et des splines cubiques de lissage pour décrire au mieux les données. L'intervalle de temps de l'expérience est divisé en 5 segments sur lesquels les données sont fittées par des splines cubiques composées de polynômes de degré 0 à 3. En sortie d'estimation nous obtenons les valeurs des coefficients estimés pour chaque spline (intersection, coefficients de chaque spline utilisée pour représenter chaque segment). Toute la subtilité de cette méthode de lissage est de choisir la régularisation adéquate et le nombre de noeuds communs à chaque profil de dégradation sans pour autant compromettre les spécificités des dynamiques individuelles. Nous avons choisi ici un paramètre de régularisation commun à toutes les courbes d'ARNm d'une part parce qu'elles sont assez monotones et d'autres part, pour éviter un sur-estimation des courbes.

Dans la figure ci-dessous nous présentons l'application de notre approche de normalisation par spline aux données de quantification de deux ARNm, *thrL* dans le panneau de gauche et *nuoM* à droite. Dans le cas d'un ARNm comme *thrL* dont le profil de dégradation n'est pas trop bruité, la normalisation prenant en compte les sondes random et les invariants bas ou l'utilisation d'une spline donne des résultats sensiblement identiques. Dans le cas d'un ARNm dont les données sont bruitées, comme pour *nuoM*, l'utilisation de la spline permet de lisser légèrement les données en plus de prendre en compte les bruits liés aux réplicats et aux sondes.

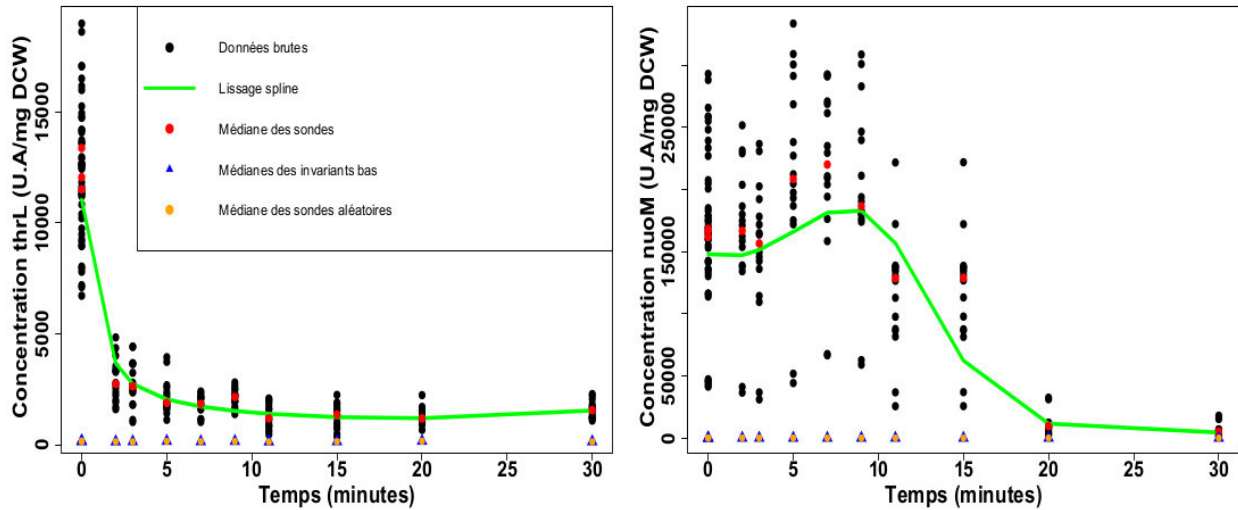


FIGURE 2.5 – Données de dégradation des ARNm *thrL* (panneau de gauche) et *nuoM* (panneau de droite). Les points orange représentent la valeur du bruit de fond à chaque temps de mesure calculé via la médiane des sondes non spécifiques et en bleu les valeurs du bruits de fond identifiées par la méthode des invariant bas. Les points noir les intensités brutes. La ligne verte représente le lissage par une spline cubique et les points rouges la médiane des 16 sondes.

2.2.3 Normalisation entre les taux de croissance

Le fait que nous cherchons à modéliser la dégradation des ARNm à quatre taux croissance différents requiert de prendre en compte des différences de physiologie des populations de cellules entre ces taux de croissance. En effet, si la même quantité d'ARN total est déposée sur les puces à ADN pour les différentes conditions, la concentration en ARN cellulaires, y compris les ARNm, varie d'un taux de croissance à un autre [Esquerré et al., 2014]. Il faut donc rendre compte de ces différences afin de pouvoir comparer les données de concentration d'ARNm entre taux de croissance.

Un bon indicateur de ces différences de physiologie est la mesure du rendement d'extraction qui représente la quantité d'ARN total extrait (en microgramme) par masse de cellules sèches (en milligramme, mgDCW) (μg ARN total / mgDCW). Comme la proportion d'ARN stables est relativement constante (elle représente 98% des ARN totaux), la variation positive des rendements d'extraction entre les taux de croissance représente alors l'augmentation moyenne de la concentration des ARNm au sein d'une population de cellules. Les intensités brutes en unités arbitraires pour une certaine quantité d'ARN total déposée (UA) sont donc converties en UA / mgDCW par multiplication avec les rendements d'extraction.

Cependant, les seules mesures robustes pour les rendements d'extraction sont obtenues pour les répliquats à $t = 0$. Par conséquent l'ensemble des points de mesure des cinétiques de dégradation des ARNm est multiplié par une seule et même valeur. Cela engendre donc une erreur proportionnelle supplémentaire qui est commune à tous les ARNm pour un taux de croissance donné. Puisque 3 cultures sont réalisées pour chaque taux de croissance, nous avons accès à la variation des rendements d'extraction qui est représentée avec les valeurs de rendement à $t = 0$ dans le tableau suivant :

Taux de croissance	$0.1h^{-1}$	$0.2h^{-1}$	$0.4h^{-1}$	$0.63h^{-1}$
Temps de doublement	$6.9h$	$3.5h$	$2.3h$	$1.1h$
Rendements d'extraction (μg d'ARN / mg DCW)	$13.1 \pm 2, 2$	$22.5 \pm 2, 5$	$25, 2 \pm 2, 6$	$45.2 \pm 8.1, 2$

TABLE 2.2 – Tableau des rendements d'extraction mesurés et leurs variations pour chaque taux de croissance [Esquerré et al., 2014]

Cette étape de normalisation entre taux de croissance peut s'effectuer en amont ou en aval des étapes de pré-traitement décrites précédemment. Cependant il ne faut pas négliger son impact, notamment au niveau de la suppression du bruit de fond. La suppression du bruit de fond par division suivie d'une multiplication par les rendements d'extraction génère un plateau artéfactuel dans les données. A noter que la transformation logarithmique de ces données comme dans [Esquerré et al., 2014] ne corrige pas ce problème. En effet cette étape ajoute une variabilité proportionnelle dans le bruit des données gênante pour l'estimation et dont il faut pouvoir tenir compte. Nous reparlerons plus précisément des effets de ce traitement et de leur implication sur le modèle d'erreur d'un modèle de régression dans le chapitre 4. Dans notre cas nous avons multiplié en premier les données par les rendements d'extraction avant la suppression du bruit de fond par soustraction (sa division éliminerait la normalisation entre taux de croissance).

2.3 Clustering de séries temporelles omiques

Le clustering de gènes avec une dynamique similaire révèle un plus petit ensemble de types de réponse qui peuvent ensuite être explorées et analysées pour des fonctions distinctes. C'est particulièrement utile quand on dispose de grands jeux de données. Avant de confronter les données traitées à un modèle mécanistique, il est donc intéressant de les explorer par des approches de classification sans a priori de ce type pour identifier les différents types de pattern de dégradation. Le point clé est que les variations dynamiques doivent être assez flagrantes pour être captées et pour pouvoir discriminer les profils temporels en différents clusters. Dans la section suivante nous passons brièvement en revue les différentes méthodes disponibles dans la littérature. Nous choisirons celle qui nous semble la plus adaptée à notre contexte. Cette analyse nous sera utile plus tard, notamment pour réduire la dimension de notre jeu de données (cf chapitre 4).

2.3.1 Méthodes de classification de séries temporelles

Deux grandes classes d'algorithmes sont utilisées pour classifier des séries temporelles omiques [Grigorov, 2011] : les algorithmes de discrimination et les algorithmes génératifs (pour revue [Kuenzel, 2010]). La première classe d'algorithme est basée sur des fonctions mesurant la similitude entre deux profils d'expression. Ces fonctions de similitude vont de la distance point par point à des fonctions plus complexes basées sur des caractéristiques générales extraites du profil [Di Camillo et al., 2005, Lampert et al., 2018, Qian et al., 2001]. La deuxième classe d'algorithmes de clustering correspond aux algorithmes génératifs. L'idée fondamentale qui motive ces algorithmes est qu'un ensemble fini de processus ou de modèles

génère les données d'expression génique. Plutôt que de mesurer directement la similitude entre les paires de profils d'expression comme le font les méthodes discriminantes, les algorithmes génératifs utilisent les données pour déterminer les paramètres optimaux des modèles sous-jacents, puis identifient comme similaires tous les profils générés par le même modèle [Peddada et al., 2003, Kim and Kim, 2007, McDowell et al., 2018]. Plusieurs algorithmes ont l'avantage d'être disponibles pour la communauté, parmi lesquels cinq paquets R : `limma` [Smyth, 2005], `timecourse` [Tai, 2007], `gprege` [Kalaitzis and Lawrence, 2011], `TTCA` [Albrecht et al., 2017], `Edge` [Leek et al., 2006] et `BETR` [Aryee et al., 2009]). Pour déterminer les avantages et les inconvénients de chacun des algorithmes précédents listés, je vous renvoie à une étude comparative effectuée en fonction du nombre d'échantillons (un ou deux) avec des données simulées [Moradzadeh et al., 2019].

2.3.2 Clustering des patterns de dégradation des ARNm

L'objet de cette section sera d'explorer les dynamiques des profils de dégradation via une approche top-down basée sur des approches de classification. Il est en effet intéressant d'utiliser ce type d'approche pour avoir une connaissance plus approfondie des caractéristiques des 4254 courbes de dégradation pour 4 conditions de croissance, soit un total de 17016 profils.

Il existe une grande variété de méthodes de clustering des séries temporelles. Chaque méthode a ses particularités dépendantes des données expérimentales. Dans notre cas les méthodes de type "model based" semblent les plus adaptées pour clusteriser les profils de dégradation d'un ARNm. Cela s'explique majoritairement par la nécessité de prendre en compte la dépendance temporelle entre les points, chose qui est possible à travers une représentation fonctionnelle des données. Ce type d'approche pour clusteriser des séries temporelles omiques se décompose en deux parties : une représentation fonctionnelle des données et l'application d'un algorithme de discrimination (k-means, mixture gaussienne, clustering hiérarchique, ...) sur les caractéristiques estimées de la représentation fonctionnelle (constante de dégradation pour un modèle exponentiel). Notons que ces caractéristiques doivent être suffisamment représentatives de la dynamique de la série temporelle pour permettre un clustering des profils efficace.

Dans notre cas une représentation semi-paramétrique des profils de dégradation semble la plus adaptée. Nous allons tirer partie des travaux effectués lors de l'étape de pré-traitement en particulier le lissage des données brutes par ces fonctions définies par morceaux que sont les splines (cf section 2.2.2). Cette étape, en plus de réduire le bruit lié au réplicats et aux sondes, nous permet d'obtenir une représentation fonctionnelle semi-paramétrique. Comme toute méthode de lissage un compromis est à effectuer entre un lissage efficace et représentatif de la dynamique temporelle observée et les questions de sur-estimation. Dans notre cas on souhaite conserver la dynamique temporelle tout en réduisant le bruit. Ce compromis s'exprime à travers le choix du paramètre de régularisation et le nombre de noeuds choisi pour segmenter les données. Cette procédure de lissage nous donne accès à 6 coefficients dont l'interprétation des valeurs n'est pas triviale. Chacun de ces 6 coefficients est associé à un noeud et à une base fonctionnelle (dans le sens mathématique). En effet chaque coefficient ne caractérise pas à lui seul le lissage obtenu car celui-ci dépend aussi du nombre de noeuds et du paramètre de régularisation. Cependant, même s'il est difficile de donner un sens pratique à ces coefficients, ceux ci contiennent des informations relatives à la dynamique des données.

Dans une première étape nous effectuerons une analyse par composante principale pour réduire les dimensions du set de coefficients estimés lors du lissage des données brutes (cf. section 2.2.2) et identifier les composantes principales permettant de discriminer les profils de dégradation. Une seconde étape utilisera les coordonnées en deux ou trois dimensions obtenues durant l'étape précédente (par projection orthogonale) pour appliquer une méthode de clustering (mixture gaussienne finie) et discriminer les profils de dégradation à travers les deux composantes principales retenues. Enfin nous discuterons des résultats et de leurs implications dans les chapitres suivant.

Les résultats d'analyse par composante principale nous permettent de réduire les dimensions du set de paramètres (6 coefficients \times 4254 ARNm \times 4 conditions). Dans la figure 2.6, nous représentons le pourcentage de la variance expliquée en fonction du nombre de composantes pour le taux de croissance $0.1h^{-1}$.

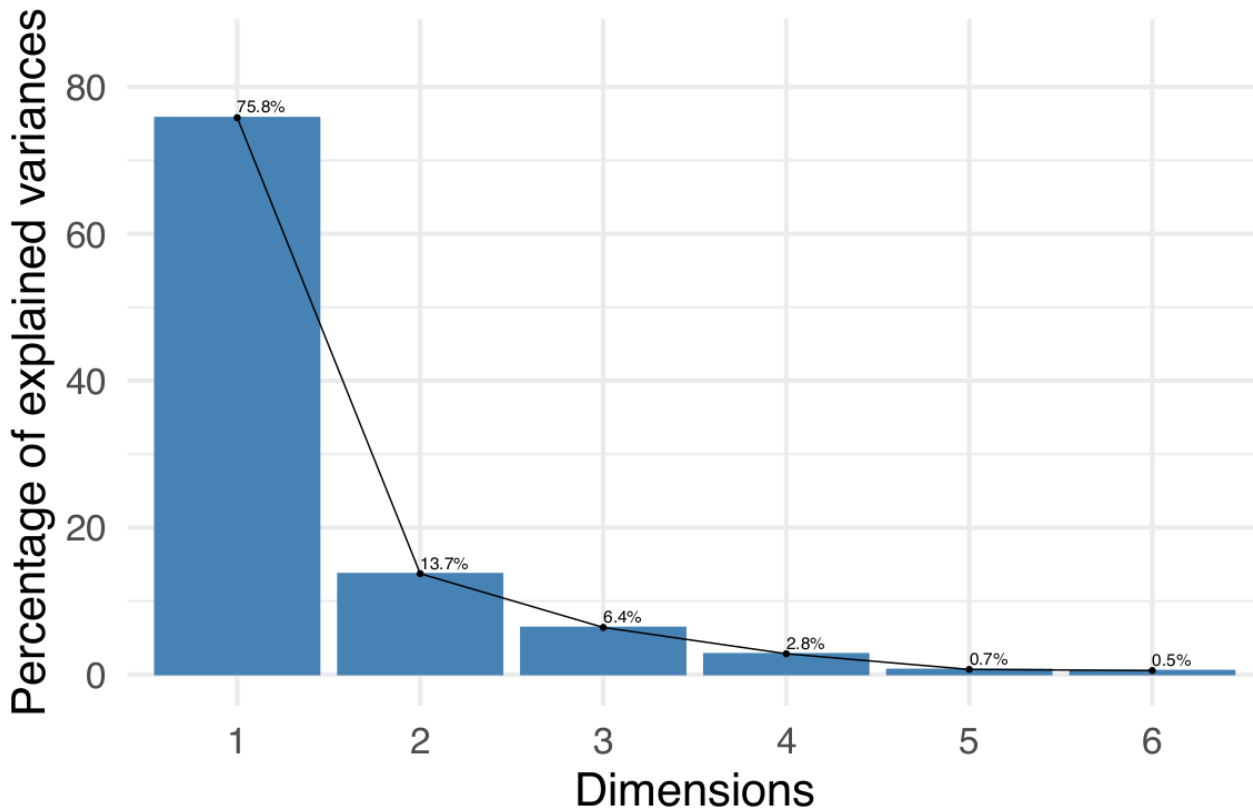


FIGURE 2.6 – Pourcentage de variances expliquées selon la dimension

Il apparaît clairement que les 2 premières composantes décrivent la majeure partie de la variabilité des coefficients ($\sim 90\%$). Comme nous le voyons dans la figure 2.6, la composante 1 explique environ 76% de la variance des coefficients. Le reste de la variabilité est majoritairement expliquée par la composante 2 ($\sim 14\%$) et la composante 3 ($\sim 7\%$). De plus au sein de la composante 1 on observe dans la figure 2.7 une forte corrélation entre les variables (coefficients) ce qui peut laisser penser à une dépendance entre les concentrations initiales et la dynamique de la phase de dégradation.

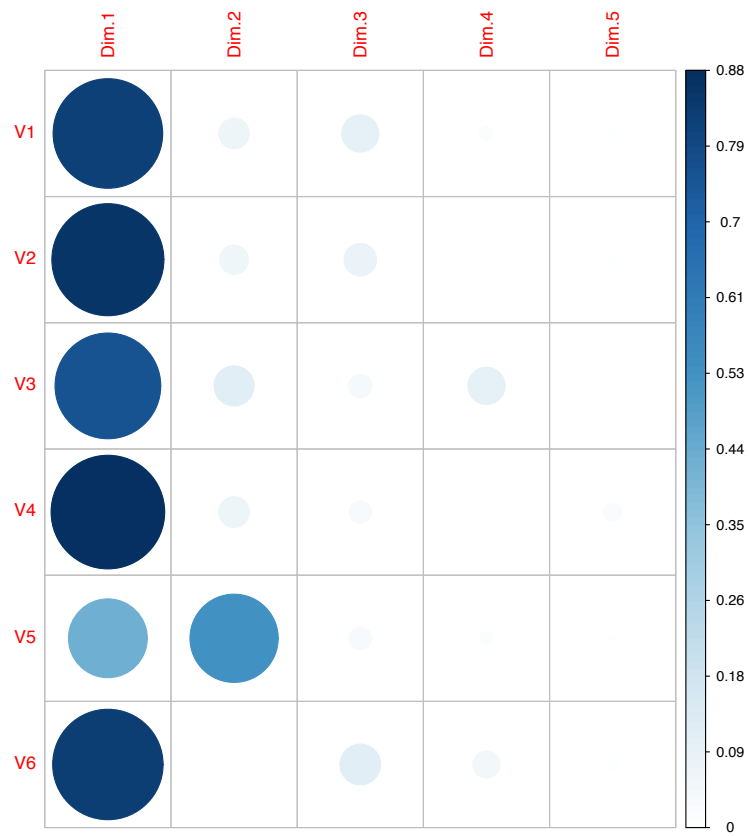


FIGURE 2.7 – Corrélations entre les coefficients selon les composantes principales

Quelles sont les contributions de chaque coefficient de la spline aux deux premières dimensions ? Grâce à la représentation graphique de la figure 2.8, nous remarquons premièrement que la composante 1 (axe horizontale) est décrite par deux groupes. Le coefficient $V5$ a en effet une contribution opposée aux autres coefficients : $V1$, $V2$, $V3$, $V4$ et $V6$. En ce qui concerne la composante 2, on remarquera que les variables $V1$, $V2$, $V3$ et $V5$ vont dans le sens même alors la variable $V4$ y contribue dans le sens opposé. Ces résultats nous confortent dans l'idée qu'il existe différentes dynamiques de profils de dégradation.

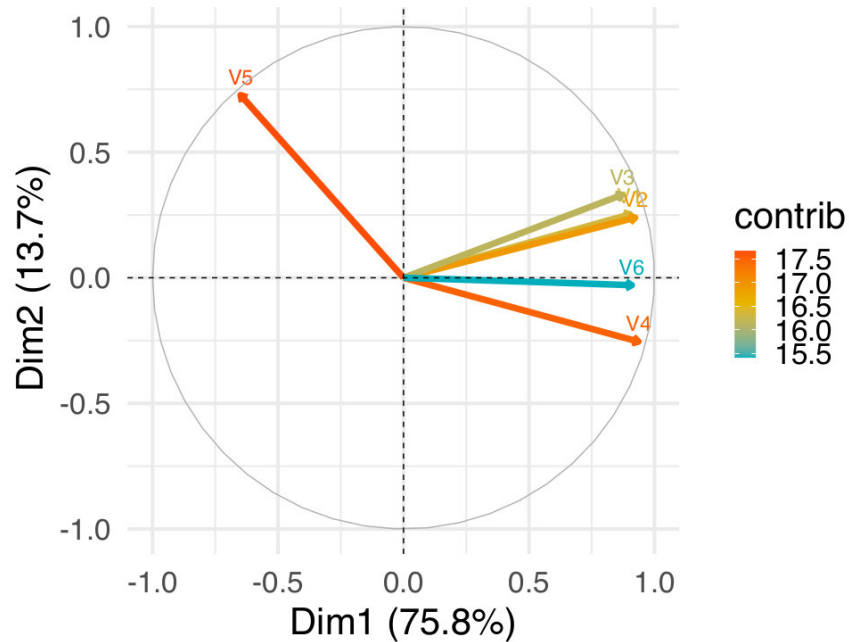


FIGURE 2.8 – Contribution de chaque variable aux deux premières dimensions

A l'aide des résultats précédents, nous utilisons les nouvelles coordonnées dans le plan orthogonal en 3 dimensions obtenus par l'analyse des composantes principales pour effectuer un clustering des profils de dégradation. L'intégration de la troisième composante renforce en effet la description des dynamiques temporelles à l'aide des coefficients. Nous utilisons pour ce faire la modélisation par mixture gaussienne finie implémentée dans le package R Mclust [Scrucca et al., 2016]. Dans un premier temps nous déterminerons le nombre de clusters optimal et le type de paramétrisation des mixtures qui convient le mieux (moyennes/variances différentes ou non entre clusters, corrélations,...) via une procédure automatique basée sur l'indicateur BIC (Bayesian Indicator Criteria).

D'après la figure 2.9 le nombre optimal de clusters semble être 5 ou 6 selon la paramétrisation des mixtures gaussiennes pour les configurations de la matrice de covariance entre les clusters en fonction de 4 critères : distribution, volume, forme et orientation (cf. [Scrucca et al., 2016] pour plus d'informations). Dans notre cas nous obtenons une valeur de l'indicateur la plus grande pour les 3 configurations suivantes : VEV (distribution = ellipsoïdale, volume = variable, forme = variable, orientation = variable) avec 6 clusters et un BIC égal à -34335), EVV (distribution = ellipsoïdale, volume = égale, forme = variable, orientation = variable) avec 5 clusters et le BIC égal à -34367) et enfin un troisième choix EVV avec 6 clusters et un BIC égal -34371. Pour le reste de l'analyse de clustering nous utiliserons la première configuration (VEV).

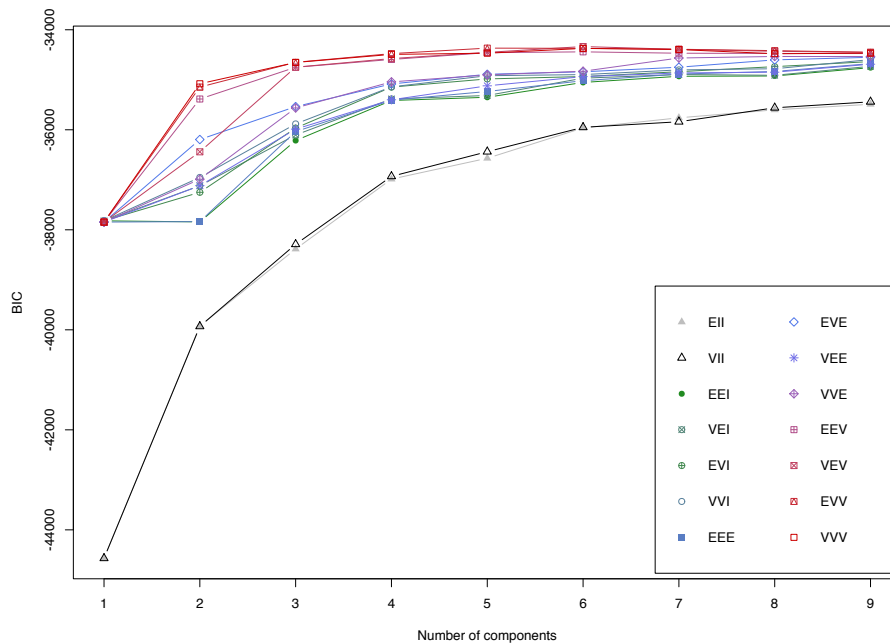


FIGURE 2.9 – Indicateur BIC en fonction du nombre de clusters initial et la paramétrisation des mixtures gaussiennes finies

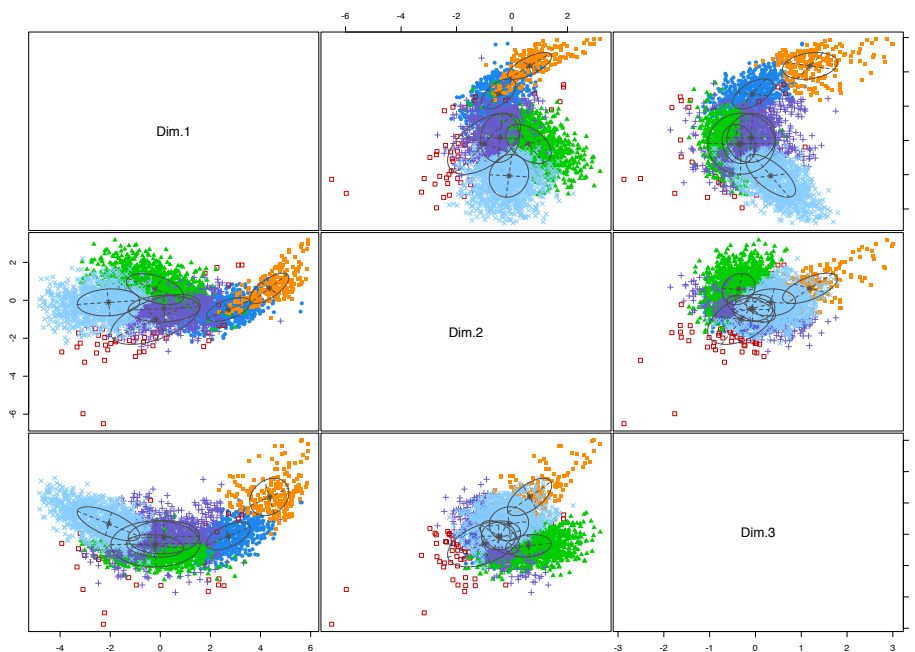


FIGURE 2.10 – Résultats de clustering des coefficients dans plan orthogonal en 3 dimensions au taux de croissance $0.1h^{-1}$

La figure 2.10 représente les résultats de clustering des coordonnées dans l'espace en 3 dimensions du vecteur des coefficients estimés lors du lissage des données brutes (cf. section 2.2.2). Nous remarquons que seulement quelques clusters se distinguent proprement comme, par exemple, le cluster orange qui est en relation avec les coefficients $V4$ et $V5$ (les deux coefficients qui évoluaient dans le sens opposé aux autres paramètres dans les 2 premières composantes principales représentées dans la figure 2.8). La qualité relative de ce clustering peut s'expliquer par la forte corrélation des coefficients au sein de la composante 1 qui, rappelons-le, est composante principale expliquant la plus grande partie de la variabilité du set des coefficients au taux de croissance $0.1h^{-1}$.

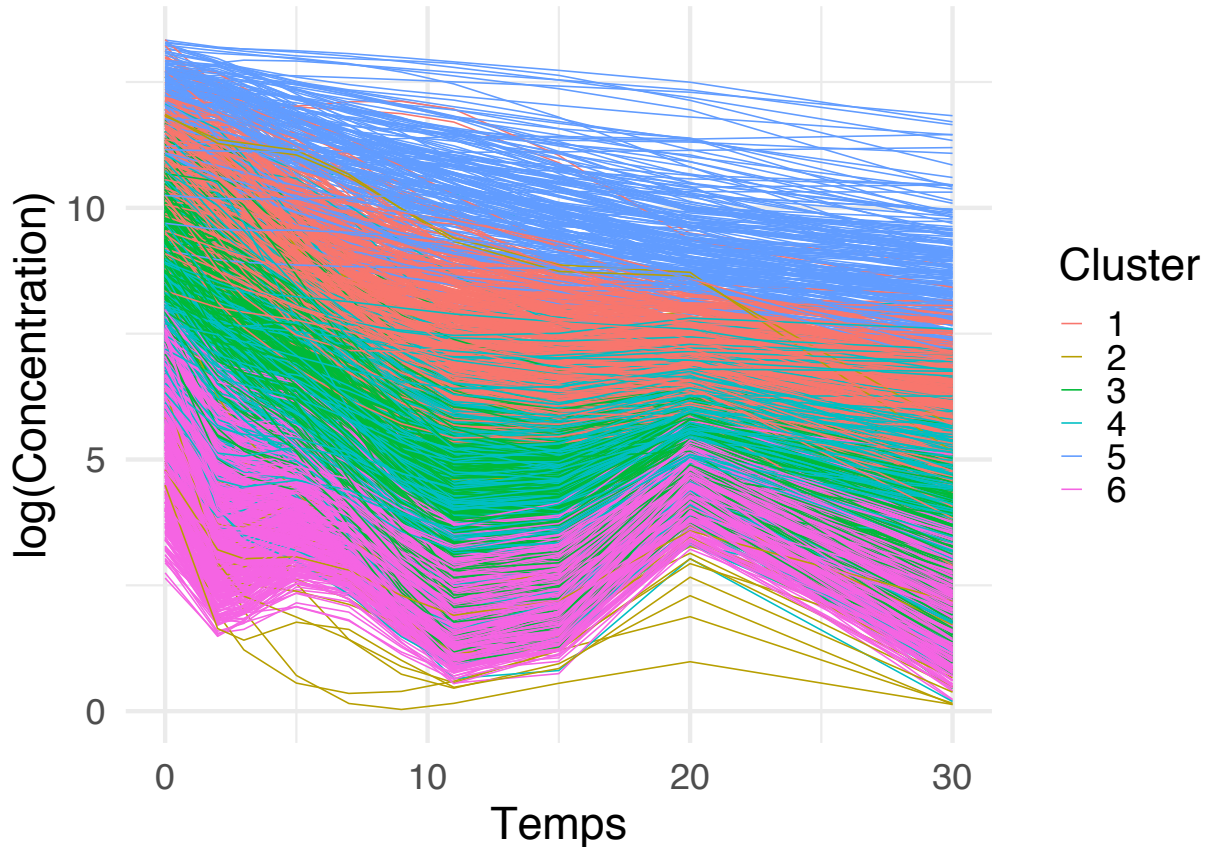


FIGURE 2.11 – Résultats de clustering des profils de dégradation au taux de croissance $0.1h^{-1}$

On observe que les profils de dégradation se classifient en premier lieu par la concentration à $t = 0$. La dynamique des profils de dégradation est fortement liée à ces concentrations initiales comme l'a déjà souligné Esquerré et al. [2014]. Cela peut s'expliquer car à $t = 0$ la rifampicine ne fait pas encore effet par conséquent la valeur est la balance entre la vitesse de transcription et de dégradation (ie processus que l'on cherche à étudier). Le deuxième critère est la forte corrélation entre cette concentration initiale et le reste de la dynamique (cf. figure 2.7). Enfin nous remarquons aussi pour certains profils, notamment les profils ayant des concentrations initiales faibles, un manque de régularité autour du point de mesure à 20 minutes. Cela peut être mis en relation avec la contribution opposée du coefficient $V5$ dans la composante 1 par rapport aux autres coefficients. La même chose (dans une moindre mesure) peut être supposée pour la contribution du coefficient $V4$ dans la composante 2 (cf 2.8).

Les résultats de la procédure top-down précédemment exposés sont pour un seul taux de croissance. Nous sommes bien conscients que cette méthodologie pour classifier des séries temporelles n'est pas la plus perfectionnée (comparée à celle de McDowell et al. [2018] par exemple) et mériterait certains ajustements. Cependant le but premier de cette analyse est d'explorer le nombre important de profils de dégradation à travers une approche top-down et d'en extraire les caractéristiques flagrantes et discriminantes composant celui-ci. Notons que ces résultats préliminaires sur l'analyse des dynamiques des profils de dégradation pourront être utilisés comme prior notamment dans l'approche bottom-up de modélisation mécanistique de la dégradation. C'est elle que nous allons aborder dans le prochain chapitre.

Chapitre 3

Modélisation cinétique déterministe de la dégradation des ARNm

Le jeu de données dont nous disposons décrit l'évolution au cours du temps de la concentration d'ARNm d'une population de cellules dans différentes conditions de croissance. La nature de ces données impose l'utilisation d'une modélisation cinétique déterministe. Dans ce contexte, il est coutumier de décrire des vitesses de réactions biochimiques à l'aide de la loi d'action de masse ou de l'équation de Michaelis-Menten qui en dérive. Dans ce chapitre, nous passons donc en revue ces lois de vitesse et appliquons ces concepts à un premier essai de modélisation de la dégradation des ARNm bactériens. Nous construisons des modèles qui varient dans leurs niveaux de description et dont l'étude nous permet de définir le niveau d'abstraction adapté à la nature des données disponibles. L'idée est de développer un modèle mathématique qui reproduit correctement les profils de dégradation des ARNm sans être trop gourmand en données expérimentales. Nous utilisons ensuite ce modèle dans des études de simulation pour analyser le rôle que pourrait jouer la compétition entre ARNm pour leur fixation à la RNase E. Nous montrons qu'il s'agit d'un phénomène qui a un impact non négligeable dans la cinétique de dégradation et que cela pourrait expliquer les particularités de nombreux profils de dégradation.

3.1 Étude de la cinétique des réactions biochimiques

3.1.1 La loi d'action de masse

En chimie, la loi d'action de masse est la proposition que la vitesse de la réaction chimique est directement proportionnelle au produit des activités ou des concentrations des réactifs (pour revue, Érdi and Tóth 1989). Elle explique et prédit les comportements des solutions en équilibre dynamique. En effet, cela implique que pour un mélange réactionnel chimique en équilibre, le rapport entre la concentration des réactifs et des produits est constant à une température donnée.

Deux aspects importants sont impliqués dans la formulation initiale de la loi : premièrement l'aspect équilibre, concernant la composition d'un mélange réactionnel à l'équilibre via l'hypothèse d'homogénéité (les molécules se diffusent de manière homogènes dans le milieu) et deuxièmement, l'aspect cinétique concernant les équations de vitesse pour les réactions élémentaires. Ces deux aspects découlent des recherches effectuées par Cato M. Guldberg et

Peter Waage entre 1864 et 1879 dans lesquelles des constantes d'équilibre ont été dérivées en utilisant des données cinétiques et l'équation de vitesse qu'ils avaient proposée (voir Guldberg and Waage 1864 et sa traduction en anglais : Waage and Gulberg 1986). La loi est donc un énoncé sur l'équilibre d'une réaction chimique et donne une expression pour la constante d'équilibre.

La première application de la loi d'action de masse a été pour un système bimoléculaire : une molécule A s'associe à une molécule B pour donner un complexe AB. Celle-ci se généralise à des réactions de différentes moléculaires. Dans le tableau qui suit (tableau 3.1), nous donnons l'écriture mathématique de la vitesse de réaction via la loi d'action de masse pour quelques exemples classiques de réactions :

Molécularité	Réaction	loi de la vitesse
uni-moléculaire	$A \rightarrow B$	$k \times A$
bimoléculaire	$A + B \rightarrow C$	$k \times A \times B$
bimoléculaire	$A + A \rightarrow B$	$k \times A^2$
bimoléculaire réversible	$A + B \rightleftharpoons C$	$k \times A \times B - k_{-}C$
trimoléculaire	$A + B + C \rightarrow D$	$k \times A \times B \times C$

TABLE 3.1 – Modélisation cinétique de réactions moléculaires simples par la loi d'action de masse

Notons qu'une réaction multi-moléculaire peut se décomposer en réactions unimoléculaires, comme dans le cas de la réaction bimoléculaire réversible $A + B \rightleftharpoons C$. Cela permet de généraliser la loi d'action de masse pour n réactants et m produits.

Il existe de nombreuses applications de cette loi en plus des réaction biochimiques et on sous-estime souvent son impact [Voit et al., 2015]. En effet la loi d'action de masse se retrouve aussi dans la formulation stochastique de la vitesse d'une réaction chimique [Gillespie, 1977] et en écologie également, ou plus généralement dans l'étude des dynamiques de population d'objets biologiques [Brauer et al., 2012]. L'écriture du système d'équations différentielles Lotka-Volterra en est un exemple [Lotka, 2002, Volterra, 1927]. Ces équations décrivent la dynamique des systèmes proies-prédateurs. Le taux de prédation sur la proie est supposé être proportionnel à la vitesse à laquelle les prédateurs et la proie se rencontrent. Ce taux est évalué comme $x \times y$, où x est le nombre de proies et y le nombre de prédateurs. Ceci est un exemple typique de la loi de l'action de masse. Une autre application de cette loi est la construction de modèles mathématiques en épidémiologie. La loi d'action de masse forme la base du modèle compartimental de propagation de la maladie en épidémiologie mathématique, dans lequel une population d'humains, d'animaux ou d'autres individus est divisée en catégories de sensibles (S), infectés (I) et récupérés (R) (dans le sens immunisés). Le modèle SIR est une abstraction utile de la dynamique de la maladie qui s'applique bien à de nombreux systèmes pathologiques et fournit des résultats utiles dans de nombreuses circonstances lorsque le principe d'action de masse s'applique. Cependant, à l'inverse des molécules qui sont considérées se mélanger de manière homogène, les individus des populations humaines ou animales obéissent moins souvent à cette hypothèse, rendant les modèles SIR invalides du fait de la non applicabilité de la loi d'action de masse [Kermack and McKendrick, 1932, 1991b,a]. Les modèles pharmacocinétique/pharmacodynamique utilisent eux aussi les

modèles par compartiments et la loi d'action de masse pour décrire mathématiquement la dynamique des médicaments [Rajman, 2008].

On retrouve également cette loi dans d'autres domaines comme la physique des semiconducteurs par exemple, la modélisation de la diffusion dans la matière condensée ou encore la sociophysique qui utilise des outils et des concepts issus de la physique et de la chimie physique pour décrire certains aspects du comportement social et politique.

3.1.2 Réactions enzymatiques : Henri-Michaelis-Menten

L'équation d'Henri-Michaelis-Menten est fondamentale en cinétique enzymatique. Elle permet de décrire des phénomènes de saturation dans des processus enzymatiques en solution suivant le mécanisme réactionnel représenté dans la figure 3.1 ci-dessous ([Michaelis and Menten, 1913]; pour revue, [Cornish-Bowden, 2013b]) :



FIGURE 3.1 – Schéma réactionnel du processus enzymatique caractérisé par L. Michaelis

où E et ES désignent respectivement l'enzyme libre et le complexe enzyme : substrat, S désigne le substrat libre et P désigne le produit, k_{+1} est la constante de vitesse de formation du complexe, k_{-1} la constante de vitesse de sa dissociation et k_2 est la constante catalytique. Pour les réactions enzymatiques dans des systèmes fermés (pas de production de substrat libre et d'enzyme libre), le schéma cinétique dans la figure 3.1 implique les relations de conservation suivantes :

$$E_T = E + ES \quad (3.1)$$

$$S_T = S + ES + P \quad (3.2)$$

où E_T et S_T sont, respectivement, les concentrations totales de l'enzyme et du substrat dans le système. Appliquer la loi de l'action de masse au schéma cinétique de la figure 3.1, et en utilisant l'équation 3.1 pour éliminer la concentration de substrat libre, donne l'ensemble suivant d'équations différentielles :

$$\frac{dS}{dt} = -k_1 \times (E_T - ES) \times S + k_{-1} \times ES \quad (3.3)$$

$$\frac{dES}{dt} = k_1 \times (E_T - ES) \times S + K_m \times ES \quad (3.4)$$

$$\frac{dP}{dt} = k_2 \times ES. \quad (3.5)$$

En règle générale, les réactions enzymatiques sont étudiées sous réserve des conditions initiales suivantes :

$$(E, S, ES, P) = (E_T, S_T, 0, 0), \quad t = 0 \quad (3.6)$$

Les équations 3.3-3.5 ne sont intégrables que sous forme fermées pour le cas trivial d'une vitesse de catalyse nulle, $k_2 = 0$ et $P \equiv 0$, auquel cas le problème se réduit à simple liaison bimoléculaire. Néanmoins, l'utilisation de l'approximation de l'état quasi-stable (quasi-steady-state approximation, QSSA) dont Briggs and Haldane [1925] étaient les auteurs des premiers travaux, s'est avérée très fructueuse dans l'analyse des équations différentielles 3.3-3.5, donnant des solutions analytiques approximatives permettant l'estimation de paramètres. Depuis, ces résultats ont été étendus à d'autres types d'approximation QSSA [Segel and Slemrod, 1989, Borghans et al., 1996, Schnell and Mendoza, 1997, 2000]. Nous en distinguons trois. L'approximation Briggs-Haldane, appelée sQSSA, est obtenue en supposant qu'une courte période transitoire précède l'état de quasi-état stable tel que $\frac{dES}{dt} \approx 0$, ce qui permet d'obtenir l'équation différentielle la plus connue décrivant la vitesse du substrat (S) catalysé en produit (P) (pour le développement complet de l'expression, voir Klipp et al. 2016).

$$\frac{dS}{dt} = \frac{k_2 \times E_T \times S}{Km + S} \quad (3.7)$$

Avec $Km = \frac{k_{-1} + k_2}{k_1}$ et comme condition de validité :

$$E_T \leq S_T + Km \quad (3.8)$$

Une autre version des hypothèses du quasi-équilibre (QSSA) considère plutôt $\frac{dS}{dt} \approx 0$; Appelée "reverse QSSA" (rQSSA), elle a été introduite pour la première fois par Segel and Slemrod [1989]. Des travaux plus récents ont étendu cette hypothèse, par exemple en considérant un scénario où la concentration d'enzyme totale est très largement en excès [Schnell, 2000], le cas de multiples substrats alternatifs [Schnell and Mendoza, 2000], ou encore le développement d'approximations (solution fermée de l'équation 3.7 à l'aide la fonction W de Lambert [Schnell and Mendoza, 1997]).

$$\frac{dS}{dt} = \frac{k_2 \times E_T \times S}{Km + E_T} \quad (3.9)$$

Schnell and Maini [2002] ont dérivé les conditions nécessaires suivantes pour la validité de la rQSSA :

$$E_T \geq S_T \quad \text{et} \quad E_T \geq K \quad (3.10)$$

avec K la constante de Van Slyke.

Enfin, Borghans et al. [1996] a étendu la validité de la sQSSA en utilisant le changement de variable suivant : $S + ES = S_T - P$. Cette dernière version des hypothèses de quasi-équilibre est appelée "total QSSA" (tQSSA). Dans [Tzafiriri and Edelman, 2004], les auteurs étudient cette approximation plus précisément en étendant le domaine de validité initialement décrit par Borghans et al. [1996] et en donnant une forme analytique appelée approximation du premier ordre donnée dans l'équation 3.11 ci-dessous. :

$$\frac{dS_T}{dt} = -\frac{k_2 \times E_T \times S_T}{Km + E_T + S_T} \quad (3.11)$$

Des extensions de la tQSSA ont également été développées, comme par exemple l'adaptation avec plusieurs substrats en compétition Pedersen et al. [2007]. Une étude comparative entre les différentes approximations montre que la tQSSA recouvre quasiment les domaines

de validité de la sQSSA et de la rQSSA [Pedersen and Bersani, 2009]. Ses conditions de validité sont les suivantes :

$$S_T + Km \gg E_T \quad \text{et} \quad K \ll Km \quad (3.12)$$

$$E_T + Km \gg S_T \quad \text{et} \quad E_T \gg Km \approx K \quad (3.13)$$

L'utilisation de cette hypothèse a démontré son intérêt notamment pour la modélisation de systèmes complexes faisant intervenir des couplages [Pedersen et al., 2008, Ciliberto et al., 2007, Bersani et al., 2019]. Les avantages sont à la fois théoriques comme par exemple l'obtention d'une approximation uniforme pour les systèmes de type Goldbeter-Koshland Switch [Bersani et al., 2020] ou plus pratique comme la simplification des modèles de la décomposition microbienne [Tang and Riley, 2013]. Il s'agira de la forme d'équation que nous adopterons par la suite pour décrire la dégradation des ARNm.

Dans la section suivante nous discuterons plus en détails la modélisation de la dégradation des ARNm via des réactions enzymatiques. Pour cela nous explorons dans un premier temps le choix de l'approximation du quasi-état stable en fonction du mécanisme biologique et des données expérimentales disponibles.

3.2 Modélisation mécanistique de la dégradation des ARNm

Notre objectif est de développer un modèle mécanistique de la dégradation des ARNm dont la confrontation avec des données dynamiques de la concentration résiduelle des ARNm nous renseignent sur les mécanismes de régulation sous-jacents et nous permettent de comprendre les différences de profils de dégradation entre ARNm et à différents taux de croissance. Ce modèle doit être assez détaillé pour être informatif mais pas trop pour limiter le nombre de paramètres, ce qui peut être problématique lors de l'étape d'estimation.

Nous nous focaliserons ici sur différents modèles que nous avons développés pour décrire la cinétique de la réaction. Il s'agissait d'une première étude de modélisation pour cerner le problème et définir la méthodologie adaptée aux données à disposition. Ces différents modèles considèrent tous que la RNase E est l'enzyme principale de la dégradation des ARNm, comme nous l'avons vu en section 1.4. C'est elle qui assure la première coupure dans les ARNm et qui contribue donc à la disparition des ARNm complets, ceux dont la concentration est mesurée dans les expériences de transcriptomique. Cette observation nous a conduit à assimiler le mécanisme de dégradation des ARNm à une macro-réaction catalysée par la RNase E. Les modèles développés diffèrent en revanche par la nature des mécanismes mis en oeuvre, de la réaction de base de la dégradation à sa régulation par les ribosomes en cours de traduction ou des petits ARNm via différents mécanismes réactionnels. Nous décrivons ci-dessous les différents modèles envisagés et les choix qui ont conduit à la modélisation de la dégradation par le modèle le plus parcimonieux possible et sa simplification par l'hypothèse sQSSA. L'article décrivant ce modèle final, inséré en section 3.3, reprend plus en détails toutes les hypothèses de modélisation.

3.2.1 Réaction de Michaelis-Menten

En l'absence de tout autre mécanisme de régulation et en assimilant la dégradation à une macro-réaction catalysée par la RNase E, nous pouvons décrire la macro-réaction comme la réaction de Michaelis-Menten de la figure 3.1. La RNase E forme un complexe enzyme-substrat avec un ARNm cible et le coupe (figure 3.2). La cinétique de la réaction est décrite par la réaction de Michaelis-Menten dans l'équation 3.7. Le paramètre Km va dépendre de l'affinité de l'enzyme pour son ARNm cible et peut-être modulé par les caractéristiques structurales de l'ARNm ou sa séquence (selon que le site de fixation de l'enzyme est proche ou non de la séquence consensus).

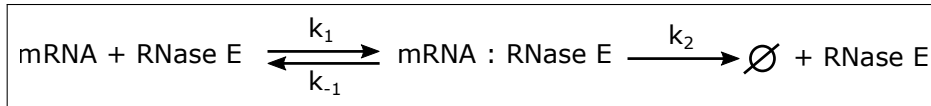


FIGURE 3.2 – Schéma réactionnel de la macro-réaction catalysée par la RNase E

Pour ce système nous ne détaillerons pas ici le système d'équations différentielles car celui-ci est déjà décrit dans la section 3.1.2 et dans l'article 1 3.3.

Avec ce modèle, nous pouvons simuler la disparition des ARNm en utilisant des valeurs de paramètres raisonnables d'un point de vue biologique. Un exemple de simulation est donné en figure 3.3. Dans ce scénario, nous sommes partis d'une concentration initiale unique et avons fait varier largement la valeur de la constante catalytique et à degré moindre celle du Km . Cela donne des courbes qui parviennent à reproduire une diversité de courbes semblable à celles observées dans les données expérimentales, avec des ARNm se dégradant vite (ils sont à moitié dégradés au bout de 2 minutes) et d'autres beaucoup moins (il faut quasiment 10 minutes pour être à moitié dégradés).

Par la suite, on reprendra et étendra ce type de modèle dans notre modélisation finale. Notamment nous considérerons que le Km est en réalité un Km apparent, englobant l'effet inhibiteur des ribosomes en cours de traduction, les régulations par des petits ARN ou des protéines régulatrices comme HFQ. Dans les modèles qui suivent, nous avons essayé de prendre en compte explicitement cet effet inhibiteur de la fixation des ribosomes ou des petits ARN. Nous envisageons deux mécanismes d'inhibition distincts, l'inhibition compétitive et l'inhibition non-compétitive.

3.2.2 Réaction de Michaelis-Menten avec inhibition compétitive des ribosomes

Ainsi que nous l'avons décrit dans la section 1.4, la fixation des ribosomes protège un ARNm en empêchant sa dégradation par la RNase E car le site de fixation de cette dernière n'est plus accessible. C'est souvent le cas lorsque la RNase E se fixe à proximité du RBS. En enzymologie, ce mécanisme s'apparente à une inhibition compétitive selon le mécanisme réactionnel décrit dans la figure 3.4.

La cinétique de chaque réaction élémentaire (réaction de premier ordre) peut se décrire par la loi d'action de masse.

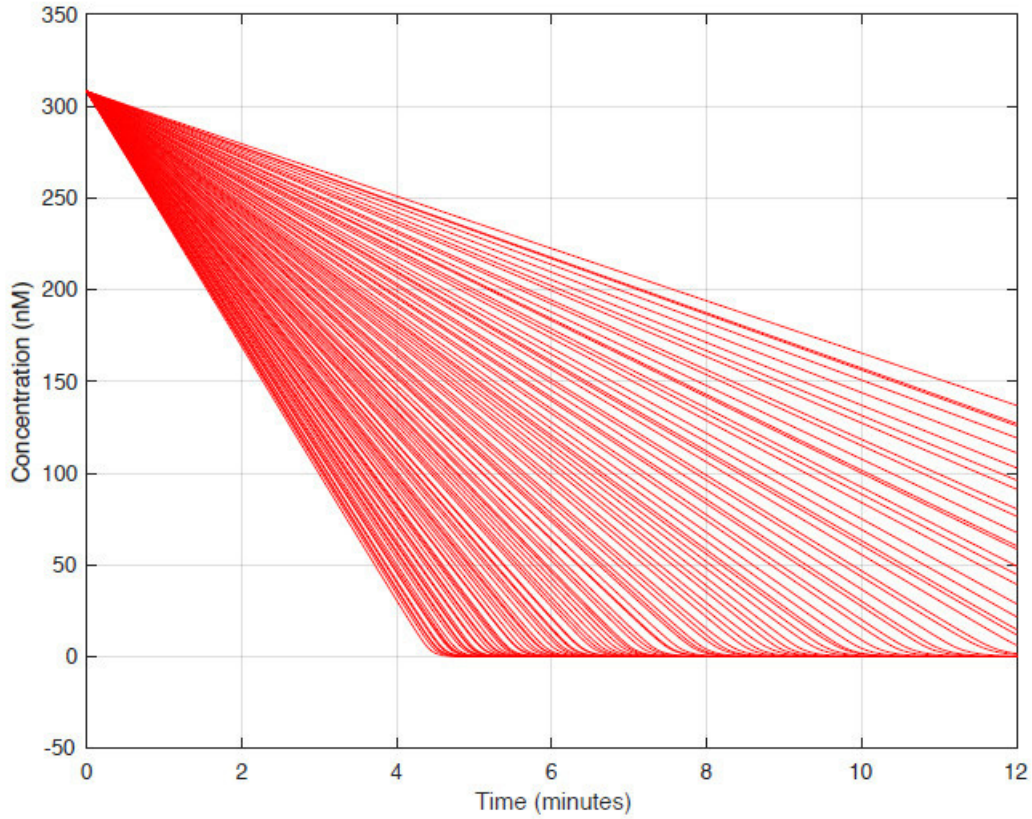


FIGURE 3.3 – Exemple de 100 simulations du modèle Michaelis-Menten de la dégradation avec des paramètres tirés aléatoirement selon une loi uniforme dans des intervalles biologiquement raisonnables

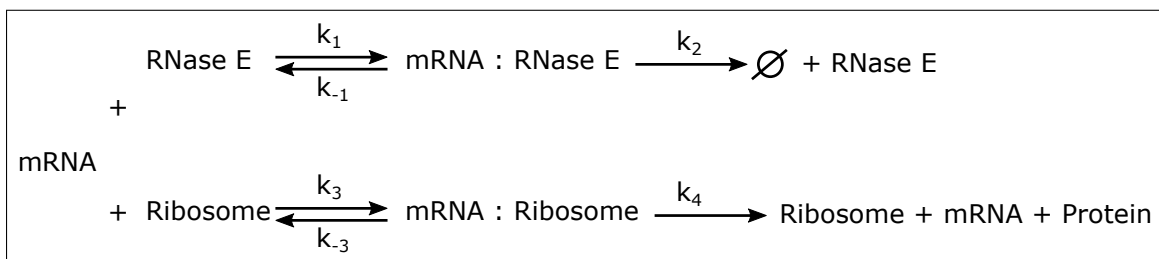


FIGURE 3.4 – Inhibition compétitive de la dégradation des ARNm par les ribosomes. Le schéma réactionnel est le même dans le cas d'un petit ARN régulateur dont la fixation stabilise l'ARNm en empêchant la fixation de la RNase E

$$\begin{cases} r_1 = k_1 \times [mRNA_{free}] \times [RNaseE_{free}] \\ r_{-1} = k_{-1} \times [mRNA : RNaseE] \\ r_2 = k_2 \times [mRNA : RNaseE] \\ r_3 = k_3 \times [mRNA] \times [Ribosome] \\ r_{-3} = k_{-3} \times [mRNA : Ribosome] \\ r_4 = k_4 \times [mRNA : Ribosome] \end{cases} \quad (3.14)$$

Ceci nous donne le système d'équations différentielles suivant pour décrire l'évolution des concentrations des différentes variables du modèle :

$$\begin{cases} \frac{d[mRNA_{free}]}{dt} = -r_1 - r_3 + r_{-1} + r_{-3} + r_4 \\ \frac{d[RNaseE_{free}]}{dt} = -r_1 + r_2 + r_{-1} \\ \frac{d[mRNA : RNaseE]}{dt} = r_1 - r_{-1} - r_2 \\ \frac{d\emptyset}{dt} = r_2 \\ \frac{d[Ribosome_{free}]}{dt} = -r_3 + r_4 + r_3 \\ \frac{d[mRNA : Ribosome]}{dt} = r_3 - r_{-3} - r_4 \\ \frac{d[Protein]}{dt} = r_4 \end{cases} \quad (3.15)$$

Afin de simplifier la forme mathématique du modèle, et de façon semblable à ce qui est fait pour l'équation de Michaelis-Menten, nous utilisons la loi de conservation de la matière et appliquons des hypothèses de quasi-équilibre sur la formation des complexes. Nous écrivons ainsi les équations de conservation suivantes pour les différentes formes d'ARNm, de RNase E et de ribosomes :

$$\begin{cases} [mRNA_{tot}] = [mRNA_{free}] + [mRNA : RNaseE] + [mRNA : Ribosome] - \emptyset \\ [RNaseE_{tot}] = [RNaseE_{free}] + [mRNA : RNaseE] \approx [RNaseE_{free}] \\ [Ribosome_{tot}] = [Ribosome_{free}] + [mRNA : Ribosome]. \end{cases} \quad (3.16)$$

L'hypothèse de quasi-équilibre est appliquée à la concentration des complexes enzyme-ARNm et ARNm-ribosome :

$$\begin{cases} \frac{d[mRNA : RNaseE]}{dt} = r_1 - r_{-1} - r_2 = 0 \\ \frac{d[mRNA : Ribosome]}{dt} = r_3 + r_{-5} - r_{-3} - r_4 - r_5 = 0. \end{cases} \quad (3.17)$$

Nous introduisons deux nouveaux paramètres K_E et K_B (ie constantes de dissociation). Nous obtenons la forme finale suivante pour la concentration des deux complexes enzyme-mRNA and mRNA-ribosome :

$$\left\{ \begin{array}{l} [mRNA : RNaseE] = [mRNA_{tot}] \times \frac{[RNaseE_{tot}]}{[RNaseE_{tot}] + K_E \times \left(1 + \frac{[Ribosome_{free}]}{K_B}\right)} \\ [mRNA : Ribosome] = [mRNA_{tot}] \times \frac{[Ribosome_{free}]}{[Ribosome_{free}] + K_B \times \left(1 + \frac{[Ribosome_{free}]}{K_E}\right)} \end{array} \right. \quad (3.18)$$

La concentration de complexe mRNA :RNase E nous permet de calculer la vitesse r_2 de dégradation de l'ARNm.

La figure 3.5 montre des résultats de simulation de ce modèle. Les paramètres communs au modèle de Michaelis-Menten utilisés en figure 3.3 ont été conservés. Dans ce scénario, l'effet inhibiteur des ribosomes est visible par le ralentissement de la dégradation que cela engendre. On observe également une plus grande diversité de profils avec les paramètres utilisés, avec des courbes aussi bien de type exponentiel que linéaire.

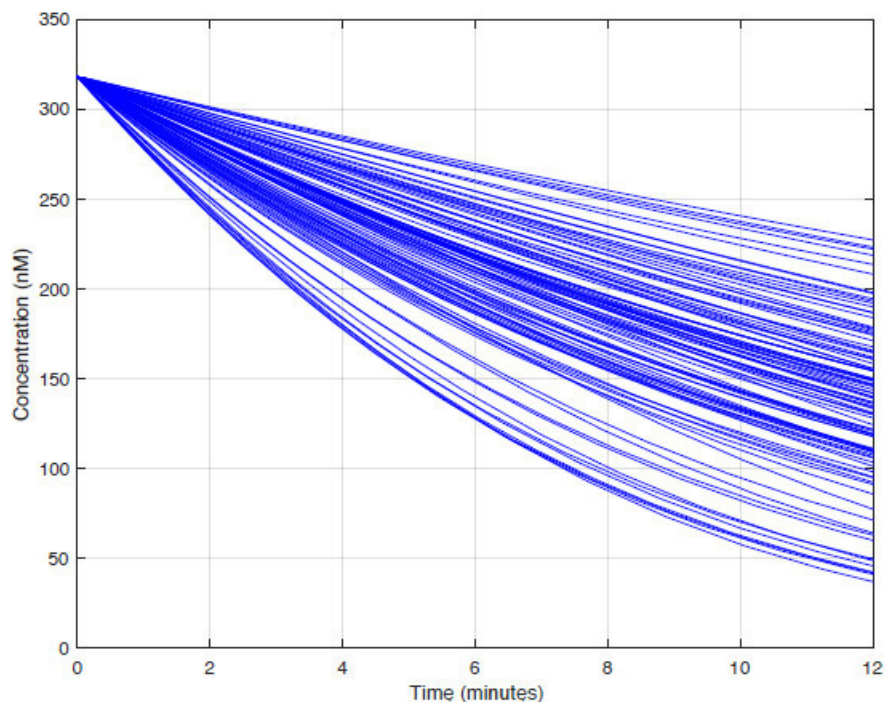


FIGURE 3.5 – Exemple de 100 simulations du modèle d'inhibition compétitive avec des paramètres tirés aléatoirement selon une loi uniforme dans des intervalles biologiquement raisonnables. Les paramètres communs avec le modèle Michaelis-Menten sont conservés

3.2.3 Réaction de Michaelis-Menten avec inhibition non-compétitive des ribosomes

Dans ce scénario, la présence de ribosomes sur l'ARNm ne gêne pas la RNase E. Ce cas peut se produire lorsque les ribosomes ne masquent pas le site de fixation de l'enzyme et

qu'elle peut s'y lier sans qu'il y ait d'encombrement stérique. Ainsi, trois cas peuvent se produire : 1) la RNase E se fixe en premier et l'ARNm est dégradé, 2) la RNase E parvient à se fixer à l'ARNm en cours de traduction et 3) le ribosome se fixe et poursuit la traduction sans qu'il y ait de coupure de l'ARNm. Ce mécanisme est décrit dans la figure 3.6.

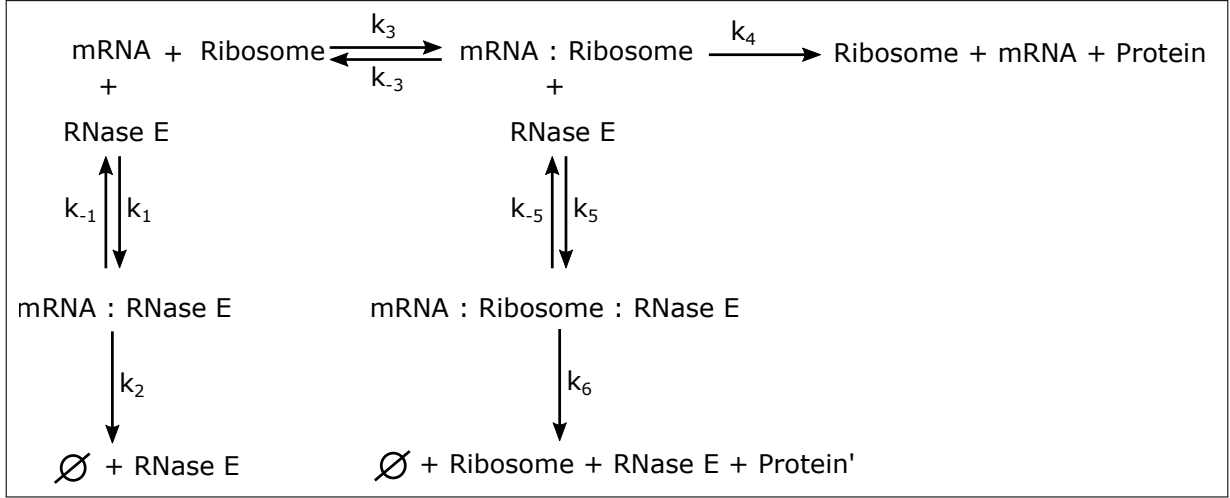


FIGURE 3.6 – Inhibition non-compétitive de la dégradation des ARNm par les ribosomes. Le schéma réactionnel est le même dans le cas d'un petit ARN régulateur dont la fixation stabilise l'ARNm

Comme précédemment, nous pouvons décrire à l'aide de la loi d'action de masse les cinétiques de chaque réaction de notre modèle :

$$\left\{ \begin{array}{l}
 r_1 = k_1 \times [\text{mRNA}] \times [\text{RNaseE}] \\
 r_{-1} = k_{-1} \times [\text{mRNA : RNaseE}] \\
 r_2 = k_2 \times [\text{mRNA : RNaseE}] \\
 r_3 = k_3 \times [\text{mRNA}] \times [\text{Ribosome}] \\
 r_{-3} = k_{-3} \times [\text{mRNA : Ribosome}] \\
 r_4 = k_4 \times [\text{mRNA : Ribosome}] \\
 r_5 = k_5 \times [\text{mRNA : Ribosome}] \times [\text{RNaseE}] \\
 r_{-5} = k_{-5} \times [\text{mRNA : Ribosome : RNaseE}] \\
 r_6 = k_6 \times [\text{mRNA : Ribosome : RNaseE}]
 \end{array} \right. \quad (3.19)$$

Le modèle EDO suivant décrit l'évolution des concentrations des différentes variables du système :

$$\left\{ \begin{array}{l}
\frac{d[mRNA]}{dt} = -r_1 - r_3 + r_{-1} + r_{-3} + r_4 \\
\frac{d[RNaseE]}{dt} = -r_1 - r_5 + r_2 + r_{-1} + r_{-5} + r_6 \\
\frac{d[mRNA : RNaseE]}{dt} = r_1 - r_{-1} - r_2 \\
\frac{d\emptyset}{dt} = r_2 + r_6 \\
\frac{d[Ribosome]}{dt} = -r_3 + r_4 + r_{-3} + r_6 \\
\frac{d[mRNA : Ribosome]}{dt} = r_3 + r_{-5} - r_{-3} - r_4 - r_5 \\
\frac{d[Protein]}{dt} = r_4 \\
\frac{d[mRNA : Ribosome : RNaseE]}{dt} = r_5 - r_{-5} - r_6 \\
\frac{d[Protein']}{dt} = r_6
\end{array} \right. \quad (3.20)$$

Nous appliquons également la loi de conservation de la matière et des approximations de quasi-équilibre pour simplifier le modèle :

$$\left\{ \begin{array}{l}
[mRNA]_{tot} = [mRNA]_{free} + [mRNA : RNaseE] + [mRNA : Ribosome] \\
\quad + [mRNA : Ribosome : RNaseE] - \emptyset \\
[RNaseE]_{tot} = [RNaseE]_{free} + [mRNA : RNaseE] + [mRNA : Ribosome : RNaseE] \\
[Ribosome]_{tot} = [Ribosome]_{free} + [mRNA : Ribosome] + [mRNA : Ribosome : RNaseE]
\end{array} \right. \quad (3.21)$$

et :

$$\left\{ \begin{array}{l}
\frac{d[mRNA : RNaseE]}{dt} = r_1 - r_{-1} - r_2 = 0 \\
\frac{d[mRNA : Ribosome]}{dt} = r_3 + r_{-5} - r_{-3} - r_4 - r_5 = 0 \\
\frac{d[mRNA : Ribosome : RNaseE]}{dt} = r_5 - r_{-5} - r_6 = 0
\end{array} \right. \quad (3.22)$$

Nous introduisons trois nouveaux paramètres K_E et K_B K_{EB} .

$$\left\{ \begin{array}{l}
[mRNA : RNaseE] = [mRNA]_{tot} \times \frac{[RNaseE]_{tot}}{[RNaseE]_{tot} + K_E \times \left(1 + \frac{[Ribosome]_{free}}{K_B}\right)} \\
[mRNA : Ribosome] = \frac{k_3 \times [mRNA] \times [Ribosome] + k_{-5} \times [mRNA : Ribosome : RNaseE]}{k_3 + k_4 + k_5 \times [RNaseE]} \\
[mRNA : Ribosome : RNaseE] = \frac{k_5 \times [mRNA : Ribosome] \times [RNaseE]}{k_{-5} + k_6}
\end{array} \right. \quad (3.23)$$

La concentration de complexe mRNA : RNase E nous permet de calculer la vitesse r_2 de dégradation de l'ARNm.

Nous montrons en figure 3.7 des résultats de simulation obtenus avec ce modèle. La pente dans la phase de décroissance est nettement plus forte dans le sens où le coude de la courbe est plus prononcé par rapport aux modèles précédents. Cela s'explique par le fait qu'il y a deux voies de dégradation possibles contrairement aux deux modèles précédents.

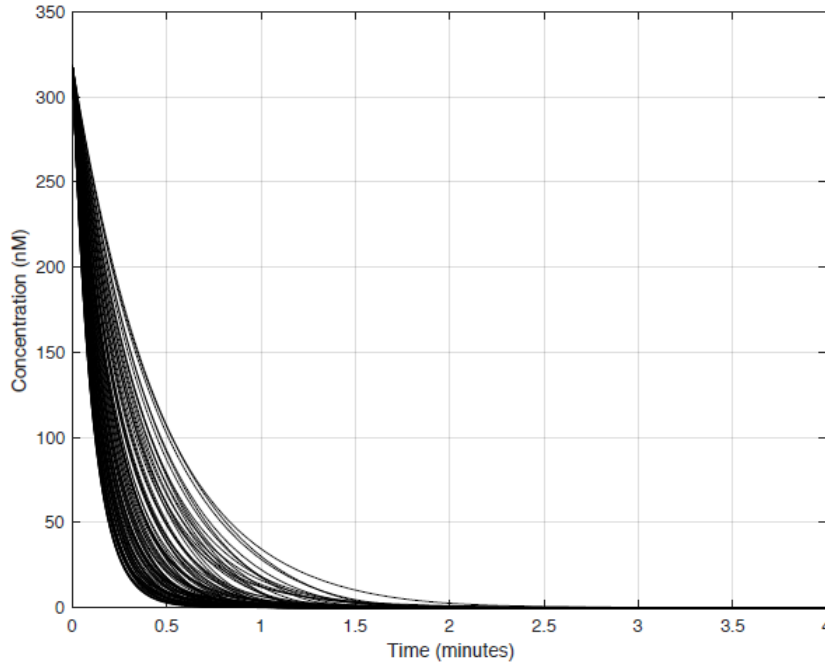


FIGURE 3.7 – Exemple de 100 simulations du modèle d'inhibition non-compétitive avec des paramètres tirés aléatoirement selon une loi uniforme dans des intervalles biologiquement raisonnables. Les paramètres communs avec le modèle d'inhibition compétitive sont conservés

3.2.4 Quel modèle pour quelles données ?

Les courbes de dégradation prédites avec les trois modèles reproduisent globalement les profils de dégradation observés expérimentalement. Avec les valeurs de paramètres employées, la complexification des modèles permet de produire des courbes plus variées, ressemblant à une exponentielle ou à une droite, et affichant des courbures plus ou moins prononcées dans les premières minutes de la cinétique. Dans ces modèles, le paramètre K_m de l'enzyme reste le même quelle que soit l'hypothèse de quasi-équilibre choisie et peut donc être interprété de façon identique.

Ces modèles ont l'avantage qu'il est relativement facile de les complexifier en introduisant plus de détails, comme des inhibitions/ activations ou des couplages avec d'autres objets/processus biologiques (processus de traduction ou encore régulation par des petits ARN non codants). Cependant, le manque cruel d'informations dans la littérature sur les paramètres cinétiques et les valeurs de concentration pose problème dès l'étape de simulation numérique des modèles. Par exemple, nous ne disposons pas de toutes les valeurs de paramètres et concentrations nécessaires pour simuler ces trois modèles et nous avons dû utiliser

des valeurs qui semblaient raisonnables d'un point de vue biologique. Ce n'est pas un bon signe pour la suite, dans l'optique de l'estimation de paramètres, car ces modèles demandent trop d'informations pour lesquelles nous n'avons pas de données expérimentales, comme la mesure de la concentration des ribosomes dans les conditions expérimentales d'étude de la cinétique de dégradation et les paramètres cinétiques des différentes réactions. Pour ces raisons, on préférera le modèle le plus parcimonieux basé sur la réaction "de base" de Michaelis-Menten pour faciliter l'étape d'estimation des paramètres cinétiques tout en réduisant le risque des problèmes d'identification. Il nous servira de base pour l'élaboration d'un nouveau modèle qui sera décrit dans la section 3.3. L'effet de régulation sera pris en compte de la même façon qu'en enzymologie : le K_m mesuré est généralement un paramètre apparent, qui englobe l'effet inhibiteur ou activateur des effecteurs de la réaction, et dont la valeur va fortement impacter le profil de dégradation. Par rapport au modèle exponentiel classiquement utilisé pour décrire la cinétique de la réaction, la réaction de Michaelis-Menten va permettre de décrire une plus grande diversité de profils du fait du phénomène de saturation pris en compte dans le mécanisme.

3.2.5 Choix des hypothèses de quasi-équilibre pour modéliser la dégradation des ARNm

Maintenant que le type de mécanisme réactionnel a été défini, il est important de pouvoir choisir l'hypothèse de quasi-équilibre adaptée à notre contexte parmi les trois décrites dans la section 3.1. Ce choix est essentiellement dicté par la nature des données à notre disposition.

Premièrement la technique de mesure utilisée (microarray) ne permet pas avec certitude de déterminer si on mesure la concentration d'un substrat libre et intact (un ARNm libre, sans RNase E fixé dessus) ou alors un la concentration de l'ensemble des formes libre et complexée. C'est vraisemblablement le dernier cas qui se produit, du fait des traitements chimiques réalisés lors de l'expérience. Ceci ne nous permet donc pas d'utiliser l'hypothèse de QSSA classiquement utilisée dans l'équation de Michaelis-Menten. Deuxièmement, la concentration de RNase E est estimée à moins de 10% de la concentration totale d'ARN à l'échelle de la population ($\sum m_i > E_t$), d'après des données de protéomique de la littérature obtenues dans des conditions de croissance similaires [Valgepea et al., 2013]. A l'échelle individuelle, la concentration d'un ARNm est très largement inférieure à la concentration totale de la RNase E ($m_i \ll E_t$). Là encore, cette observation disqualifie l'hypothèse classique de la QSSA. Cette observation aura par ailleurs une conséquence importante par la suite, car cela suggère qu'en l'absence d'excès de RNase E, des mécanismes de compétition entre ARNm peuvent subvenir lors de la dégradation des ARNm cellulaires. Troisièmement il s'avère qu'une des conditions de l'hypothèse tQSSA $K \ll K_m$ est vérifiée pour un nombre non négligeable d'enzymes métaboliques (voir chapitre 5 du livre [Atkinson, 1977]) et cette hypothèse a le plus grand domaine de validité que la QSSA classique et la rQSSA. Par conséquent, ces trois éléments militent pour l'utilisation de la tQSSA comme hypothèse de quasi-équilibre pour la modélisation de la dégradation des ARNm dans la suite de notre étude.

3.3 Modélisation de la dégradation des ARNm par système enzymatique compétitif à plusieurs substrats

L'étude de modélisation mécanistique que nous avons décrite dans la section précédente nous a permis de déterminer le type de cinétique à considérer pour décrire de façon parcimonieuse la dégradation des ARNm et leur régulation, ainsi que l'hypothèse de quasi-équilibre à utiliser, la tQSSA. Cependant, ces efforts de modélisation sont issus de la même démarche que la modélisation basée sur une exponentielle, dans la mesure où la dégradation de chaque ARNm est considérée indépendamment de celle des autres ARNm. Les données de protéomique citées en section 3.2.5 suggèrent qu'il n'y a pas assez de RNase E dans les cellules pour que la dégradation des ARNm se fasse de façon indépendante. Nous avons donc modifié notre modèle réactionnel simple de façon à prendre en compte que la dégradation des ARNm est couplée parce qu'ils partagent la même machinerie de dégradation. Ces travaux sont décrits dans un article publié dans *Journal of Theoretical Biology* inséré ci-après, dont je suis premier auteur. Ils nous ont permis de montrer par des études de simulation des modèles développés l'importance de la compétition entre ARNm pour générer une grande diversité de profils de dégradation, tels qu'on les observe expérimentalement.

Competitive effects in bacterial mRNA decay

Thibault A. Etienne, Muriel Coccagn-Bousquet and Delphine Ropers
Journal of Theoretical Biology, 504 : 110333, 2020



Competitive effects in bacterial mRNA decay

Thibault A. Etienne^{a,b}, Muriel Coccagn-Bousquet^a, Delphine Ropers^{b,*}

^a TBI, Université de Toulouse, CNRS, INRA, INSA, Toulouse, France

^b Univ. Grenoble Alpes, Inria, 38000 Grenoble, France



ARTICLE INFO

Article history:

Received 28 October 2019

Revised 8 May 2020

Accepted 12 May 2020

Available online 29 June 2020

Keywords:

Michaelis–Menten kinetics

Quasi-steady-state assumption

Substrate competition

mRNA turnover

Regulation of gene expression

ABSTRACT

In living organisms, the same enzyme catalyses the degradation of thousands of different mRNAs, but the possible influence of competing substrates has been largely ignored so far. We develop a simple mechanistic model of the coupled degradation of all cell mRNAs using the total quasi-steady-state approximation of the Michaelis–Menten framework. Numerical simulations of the model using carefully chosen parameters and analyses of rate sensitivity coefficients show how substrate competition alters mRNA decay. The model predictions reproduce and explain a number of experimental observations on mRNA decay following transcription arrest, such as delays before the onset of degradation, the occurrence of variable degradation profiles with increased non linearities and the negative correlation between mRNA half-life and concentration. The competition acts at different levels, through the initial concentration of cell mRNAs and by modifying the enzyme affinity for its targets. The consequence is a global slow down of mRNA decay due to enzyme titration and the amplification of its apparent affinity. Competition happens to stabilize weakly affine mRNAs and to destabilize the most affine ones. We believe that this mechanistic model is an interesting alternative to the exponential models commonly used for the determination of mRNA half-lives. It allows analysing regulatory mechanisms of mRNA degradation and its predictions are directly comparable to experimental data.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

Intracellular levels of bacterial mRNAs result from a balance between transcription and degradation. It has long been believed that mRNA decay plays a minor role in the regulation of bacterial gene expression, most likely because bacterial transcription has been early shown the target of so many regulations. mRNA degradation has gained renewed attention when studies have evidenced tight modulations of mRNA half-life by environmental conditions, through the action of small RNAs and RNA-binding proteins (see e.g. Bobrovskyy et al., 2015; Mohanty and Kushner, 2016). Short mRNA half-lives allow both a rapid remodelling of gene expression when cells undergo environmental perturbations and a recycling of the ribonucleotide pool (Carpousis, 2007). In addition, a complicated multi-protein machinery called degradosome is responsible for both mRNA instability and RNA processing (see Marcaida et al., 2006; Lykke-Andersen et al., 2009 for reviews).

Experimental approaches for analysing mRNA decay are based on transcriptional arrest by antibiotics addition (e.g. rifampicin). The kinetics of mRNA degradation is monitored by measuring

mRNA levels along time. With the advent of high-throughput technologies, this is nowadays achieved at the whole cell level using microarrays or RNA sequencing. Fitting the degradation profiles to an exponential model provides half-life values for cell mRNAs. They typically vary between 3 and 8 min in the model bacterium *Escherichia coli*, much shorter than the duration of the cell cycle (Bernstein et al., 2004; Chen et al., 2015; Esquerré et al., 2014). By comparison, ribosomal and transfer RNAs are stable and function over several generations. In some extreme cases in the stationary phase, half-lives as high as 50–100 min are measured in *E. coli* (Dressaire et al., 2018) and up to 90 min in the budding yeast (Wang et al., 2002).

These exponential models of mRNA decay fall short of providing mechanistic interpretations to the degradation kinetics. They are based on the assumption that mRNA decay follows a first-order reaction, that is, the reaction rate depends linearly on the mRNA concentration (Bernstein et al., 2002; Chen et al., 2015). They do not explicitly include regulations of mRNA decay, as this would require complexifying both the model and the procedure of parameter estimation. Notably, one largely ignored regulatory mechanism is the possible implication of substrate competition, while experimental studies point at its possible involvement (Esquerré et al., 2015; Nouaille et al., 2017).

* Corresponding author.

E-mail address: delphine.ropers@inria.fr (D. Ropers).

Substrate competition is encountered in many biochemical processes, from signal transduction to gene expression or transport in living organisms (e.g., Cookson et al., 2011; De Vos et al., 2011; Kiel and Serrano, 2012). It occurs when several substrates compete for binding to the same active site of an enzyme, which possibly leads to complex dynamical behaviours. For instance, the substrates of the MAP kinases (MAPK), which regulate various cell fate decisions (proliferation, differentiation, and apoptosis), compete for binding to MAPK in early *Drosophila* embryos (Kim et al., 2010; Kim et al., 2011; Legewie et al., 2007). Their local concentration and their relative binding affinities for the MAPK is predicted to control both MAPK activity and the choice of substrate. This enables the same MAPK pathway to operate in different modes and elicit distinct cell type responses.

In the case of mRNA degradation, the same machinery is responsible for the fate of messenger RNAs. In *E. coli* for instance, the main degradation pathway is the internal cleavage of mRNAs by RNase E, an enzyme also involved in RNA processing (Clarke et al., 2014). In the alternative, so-called 5'-end-dependent, pathway, mRNA cleavage by RNase E occurs after the prior recognition and removal of two of the three phosphate groups at the 5' end of the transcript by the pyrophosphohydrolase RppH. The fragments of mRNAs are then rapidly degraded by exoribonucleases. RNase E is abundant in *E. coli* cells (Mackie, 2013; Valgepea et al., 2013), at least in excess with respect to individual substrates, but the total substrate concentration largely exceeds that of the enzyme. As a consequence, pools of RNase E are limited in cells. It is thus expected that differences of substrate affinities and concentrations play a determinant role in the degradation kinetics of mRNAs. Experimental observations seem to agree with this hypothesis, at least for mRNA initial concentrations, for which a strong negative correlation was found with the mRNA lifetime (Esquerré et al., 2015; Nouaille et al., 2017).

In this manuscript, we study the possible role of substrate competition on mRNA degradation kinetics. We need for this a mechanistic modelling of mRNA degradation describing the binding of RNase E to competing substrates and the subsequent titration of the enzyme. The model should make a trade-off between complexity and biological realism, and have physiologically interpretable parameters. In addition model predictions should be directly comparable to experimental data. Common models of mRNA degradation kinetics do not meet this requirement. We thus developed a mechanistic model of mRNA decay in this study and analysed the effect of mRNA competition on the degradation kinetics.

In what follows, we briefly survey mRNA degradation (Section 2). We show in Section 3 how to assimilate mRNA degradation by the degradosome to a macro-reaction catalysed by the limiting enzyme RNase E. The kinetics of the reaction is described by means of a Michaelis–Menten equation. To take into account the fact that RNase E is in limited concentration with respect to all mRNA concentrations, we used the tQSSA form of the Michaelis–Menten equation for describing the degradation kinetics. By extending previous works by Pedersen et al. (2007) and Tang and Riley (2013), we developed an approximated version of the tQSSA form which includes competition between all cell mRNAs. We used this approximation in Section 4 to study the contribution of competition to mRNA decay by means of numerical simulations, based on carefully chosen parameter values and distributions. Our results show that the introduction of competition leads to variability, delays and increased non linearities in the degradation curves, similarly to experimental observations. Such diversity and non linearities are difficult if not impossible to reproduce in the absence of competition. By means of rate response coefficients in Section 5 and 6, we show how competition globally affects degradation by slowing it down and how it differentially affects the fate of mRNAs. We discuss these results in Section 7.

2. A short survey of mRNA degradation: mechanism, monitoring and kinetic modelling

2.1. The mechanism of mRNA degradation and its regulation

In many bacteria, RNase E (EC 3.1.26.12) is a central player of mRNA degradation and RNA processing. It initiates the processing of small RNAs and of a majority of pre-tRNAs, participates to the maturation of ribosomal RNAs in addition to degrading mRNAs. The tetrameric enzyme forms a multi-protein complex called the RNA degradosome, together with the RNA helicase B (RhB), the glycolytic enzyme enolase (Eno), and the phosphorolytic exonuclease polynucleotide phosphorylase (PNPase) (Carpousis, 2007). It is mostly anchored to the cytoplasmic inner membrane, where it forms highly mobile foci diffusing all over the membrane (Strahl et al., 2015). These foci disappear upon inhibition of transcription and the subsequent disappearance of mRNAs.

mRNA degradation results from the combined action of endo- and exo-ribonucleases (Fig. 1). As an endoribonuclease, RNase E

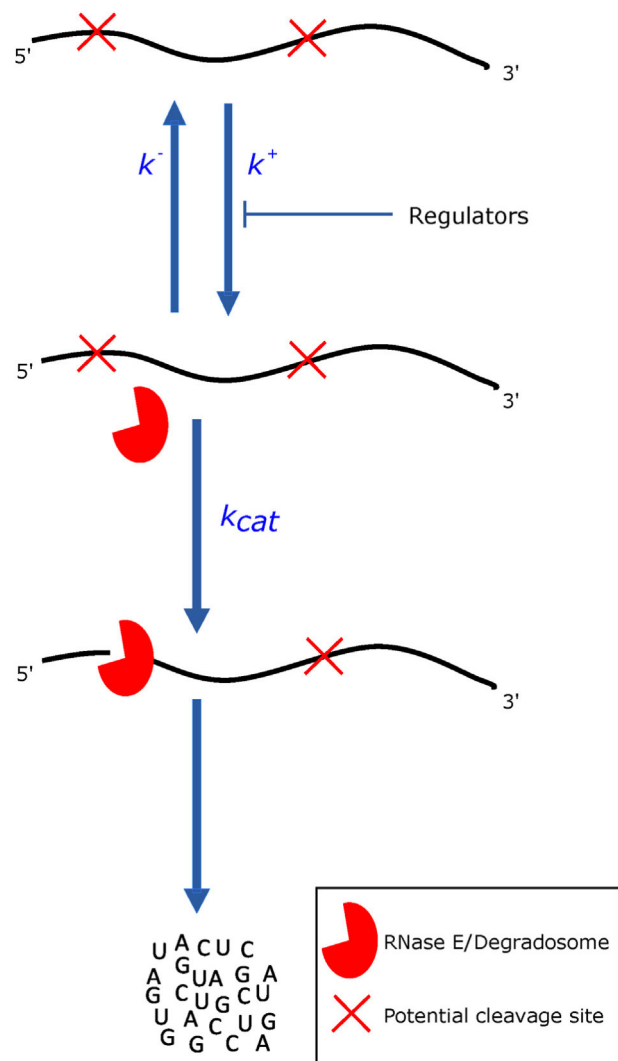


Fig. 1. RNA degradation in Prokaryotes. Endoribonucleolytic cleavage by RNase E generates fragments that are further converted to nucleotides by oligoribonucleases. RNase E binds to a A-U rich sequence of the messenger RNA, often downstream of a stem loop. The binding can be strengthened or weakened by regulatory proteins or RNAs. k^+ and k^- are the rate constants for the association and dissociation of RNase E to the mRNA, respectively. k_{cat} is the catalytic constant of the degradation reaction.

catalyses the first internal cleavage, followed by a rapid degradation of the RNA fragments by exoribonucleases. This initial cleavage is considered to be the limiting step in mRNA degradation (Kushner, 2002). RNase E has also an indirect effect on mRNA decay, through the processing of small regulatory RNAs that regulate the stability of their target mRNAs (Chao et al., 2017). This is for instance the case of CsrB and C, which control the activity of the post-transcriptional regulator CsrA. Their stabilization or destabilization alters the half-life of many CsrA-regulated genes (Suzuki et al., 2006).

The gene coding RNase E is essential, presumably due to its role in RNA processing. Knocking down RNase E activity *in vivo* or decreasing its cellular level result in a slowdown in the turnover of bulk mRNAs (Bernstein et al., 2004; Sousa et al., 2001; Stead et al., 2010). Microarray analyses have shown RNase E to initiate the decay of 60% of *E. coli* mRNAs (Bernstein et al., 2004; Stead et al., 2010). Recently RNA-seq experiments have mapped thousands of RNase E cleavage sites in both *E. coli* and *Salmonella* transcriptomes (Chao et al., 2017; Clarke et al., 2014). Other endoribonucleases have been identified, such as RNase III and the paralog of RNase E in *E. coli*, RNase G, but they have a much more limited set of substrates (Carpousis, 2007; Kushner, 2002).

RNase E prefers binding to single-stranded RNA regions that are typically A-U rich, with a strong preference for U at position +2 with respect to the scissile phosphate (Chao et al., 2017; Clarke et al., 2014; Huang et al., 1998; McDowall et al., 1994; McDowall et al., 1995). Two determinants of substrate recognition are the 5'-monophosphate termini – known to activate degradation through a higher enzyme affinity (Kime et al., 2010; Mackie, 1998) – and the presence of a stem-loop upstream of the cleavage site in most mRNAs cleaved by RNase E (Del Campo et al., 2015). The current model proposes two modes of substrate recognition, which do not need to be mutually exclusive (Bandyra et al., 2018). In the 5'-sensing mode, RNase E binds to the 5' monophosphate and the first bases of the 5' termini, causing an appropriate alignment at the active site of the single-stranded region containing the cleavage site. In this case, degradation by RNase E requires a preliminary processing by RppH of the 5'-triphosphate termini to 5'-monophosphate (Celesnik et al., 2007; Deana et al., 2008; Garrey and Mackie, 2011). In the 5' bypass mode, which seems to be the most frequent one (Clarke et al., 2014; Richards et al., 2012), the direct binding of RNase E to a structured region of the mRNA allows to align the single-strand region at the active site independently of the nature of the 5' end (Bandyra et al., 2018).

The binding of RNase E to individual bacterial mRNAs is also regulated by trans-acting factors (Fig. 1) (Richards and Belasco, 2019). The base pairing of a small RNA with a mRNA can lead for instance to translational activation or repression in response to stresses. The later case is often associated to an increased destabilization of the messenger RNA. In many cases the process involves the protein chaperone Hfq, which facilitates the base pairing between the sRNA and the mRNA (see for review Santiago-Frangos and Woodson, 2018). Ribosomes bound to the ribosome-binding site protect mRNAs from degradation by RNase E (for review, Deana and Belasco, 2005). Changes in the level and activity of RNase E also modulates the degradation rate. An auto-regulation mechanism has been shown sensing the intracellular level of RNase E (Jain and Belasco, 1995). However, under normal circumstances there appears to be excess RNase E in the cell, because enzyme levels have to be significantly reduced to see any phenotypic effect (Jain and Belasco, 1995). The regulatory proteins RraA and RraB were also shown to repress RNase E activity *in vivo* when over-expressed (Lee et al., 2003), but their physiological role remains to be determined.

2.2. Monitoring mRNA degradation kinetics

Current experimental methods monitor mRNA decay in Prokaryotes and Eukaryotes through three different approaches: general inhibition of transcription by drugs (e.g. rifampycin) or mutation, transcriptional control by regulatable promoters and pulse-chase labelling (Lugowski et al., 2018). Transcription is inhibited in the two first cases, either that of all cellular genes or of a single gene, following which the mRNA level starts to decay. Pulse-chase labelling relies on the cell incorporation of labelled nucleotides into newly synthesized RNAs. The addition of natural nucleotides wash out the labelled nucleotides. mRNA decay is then monitored by following the decline of labelled mRNAs.

Transcription inhibition is often used in Prokaryotes (see Bernstein et al., 2002; Dressaire et al., 2018; Esquerré et al., 2014; Hundt et al., 2007 for a few examples), coupled nowadays to high throughput technologies for the quantification over time of residual mRNA concentrations. Large scale data obtained with micro-arrays are by far the most abundant in the literature (for review Laguerre et al., 2018). At the time of transcription inhibition and in the following minutes, the culture is sampled, mRNAs are extracted, purified, and used as matrix for the synthesis of complementary cDNAs. The latter are then quantified by hybridization to DNA chips. As an illustration, mRNA degradation profiles obtained in *E. coli* with this approach are shown in Appendix A.1. Alternatively, qRT-PCR approaches allow obtaining medium scale data through the amplification of selected cDNAs. RNA sequencing approaches starts also being used (Chen et al., 2015; Laguerre et al., 2018; Moffitt et al., 2016; Potts et al., 2017). They consist in the depletion of ribosomal RNAs and the retro-transcription of mRNAs into cDNAs, before the latter are sequenced.

Depending on the method chosen for the quantification of residual mRNAs, different types of data normalization can be applied (Laguerre et al., 2018). The objective is to compensate for technical variations between samples at different time points. The mRNA levels can be further multiplied by RNA extraction yields to obtain mRNA concentrations in arbitrary units per gram dry weight. Mathematical models are subsequently used to estimate mRNA half-lives from these normalized data, as we shall see in the following section. Other modelling efforts include mRNA degradation beside other processes, some models being used to analyse regulatory mechanisms of mRNA degradation kinetics.

2.3. Kinetic modelling of mRNA degradation in literature

Various mathematical models of mRNA degradation kinetics have been proposed in the literature, the most common and simplest one being the exponential model. This mathematical description simplifies the mechanism of mRNA degradation to a first-order reaction modelled by the following linear ordinary differential equation (ODE):

$$\frac{dm_i(t)}{dt} = -k_{deg_i} \times m_i(t), \quad m_i(0) = m_{i0} \quad (1)$$

where $m_i(t)$ is the concentration at time t of the messenger RNA of gene i . k_{deg_i} is the degradation constant, from which mRNA half-life can be determined: $t_{1/2_i} = \log(2)/k_{deg_i}$. This model is valid once transcription has been stopped at time 0, which means that the transcription rate does no longer counterbalance the degradation rate of mRNAs.

This mathematical modelling implies that the degradation kinetics depends linearly on the initial concentration of mRNA i , according to the following analytical solution of the ODE system:

$$m_i(t) = m_{i0} \times e^{(-k_{deg_i} \times t)}. \quad (2)$$

The exponential model in Eq. (2) is commonly used to estimate mRNA half-life by linear regression on a semilog plot of the concentration levels (for review, Laguerre et al. 2018). The approach has been used for decades to estimate individual mRNA half-lives. It has been extended more recently to genome-wide data (e.g. Bernstein et al., 2002, 2004; Esquerré et al., 2014; Potts et al., 2017). Other related models are now emerging. They extend the exponential model with a delay before the onset of the decay (Chen et al., 2015; Dar and Sorek, 2018; Moffitt et al., 2016) and a residual mRNA concentration at the end of the degradation process (Moffitt et al., 2016). The delay is generally attributed to the time required by RNA polymerase to complete transcription initiated prior to the drug addition.

First-order models provide valuable information on reaction rates through the determination of mRNA half-lives. However, they fell short of describing the biological mechanisms at work. In the literature, mechanistic models including mRNA degradation are generally broader in scope. With a relative simple description of the degradation mechanism, they relate several cellular processes such as the synthesis and degradation of mRNAs and proteins, the DNA replication and the cell growth... (e.g. Tadmor and Tlusty, 2008; Thomas et al., 2018; Weiße et al., 2015). In (Thomas et al., 2018) for instance, mRNA degradation is identified as one source of growth heterogeneity at slow grow rates, when ribosomes are in too low numbers to protect mRNA from degradation by endonucleases. This leads to variations in biomass production and division times. Some modelling efforts are more focused on the mechanism of mRNA degradation and its regulation, such as the 5' binding and sliding of endonucleases on mRNAs and the mRNA protection by ribosomes in bacteria (Carrier and Keasling, 1997; Deneke et al., 2013), or else the modes of action of miRNAs on protein translation and mRNA degradation in Eukaryotes (Morozova et al., 2012). For instance, Carrier and Keasling (1997) and Deneke et al. (2013) have developed models of this type using a stochastic framework. However, the simulation of these models comes with a computational cost, which makes it inappropriate to analyse the coupled degradation of all cell mRNAs.

Knowing that large scale data are mostly available for mRNA degradation nowadays, these mechanistic modelling studies tell us that a compromise has to be reached in terms of model complexity and biological realism. The ideal model should i) be realistic from a biological point of view and useful to analyse regulatory mechanisms of mRNA degradation including competition; ii) alleviate the computational cost of numerical simulations; iii) include parameters that can be interpreted from a biological point of view; and iv) predict degradation kinetics that can be directly confronted to experimental data. We develop in the following section a kinetic model of the coupled degradation of all cell mRNAs, which meets these constraints as much as possible. We justify how we simplified the complexity of mRNA degradation to a macro-reaction catalysed by the limiting enzyme RNase E. We modelled this reaction and its regulations with the total quasi-steady-state approximation of the Michaelis–Menten equation. We incorporate substrate competition among the regulatory mechanisms. This required extending earlier modelling works on substrate competition to the case when multiple substrates compete for the same enzyme and when individual substrates are in excess of enzyme, but not the total substrate concentration.

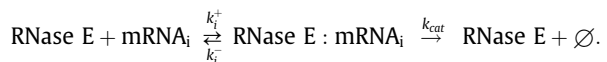
3. Model development

3.1. Simplified description of mRNA degradation

As explained in Section 2.3, developing a model that can be confronted to experimental data constrains the level of description and the nature of the mechanisms to consider. We described in

Section 2.2 how high throughput approaches are used to quantify residual mRNA concentrations after transcriptional arrest. These approaches quantify full length mRNAs through a good coverage of mRNA sequences. For instance, qRT-PCR and microarray experiments make use of oligonucleotides distributed along mRNAs (see Esquerré et al., 2014; Nouaille et al., 2017 for example), and read counts for each gene are determined in RNAseq experiments. In other words, messenger RNAs that have been cut once or several times by endo- and exo-nucleases are not tractable in these experiments. Especially since the rapidity of degradation by exonucleases makes it unlikely that partially degraded mRNAs could be confused with full length mRNAs. According to Fig. 1, this implies that our model does not need to describe reaction steps posterior to the initial cleavage by the endoribonuclease. In most cases, the endoribonuclease is RNase E and we will refer to that enzyme hereafter, although in some cases other endoribonucleases like RNase G are known to specifically degrade some target mRNAs. Our model remains valid for these alternative enzymes as well.

Experiments monitor the disappearance of full length mRNAs following the binding of RNase E to these mRNAs and their internal cleavage. As summarized in Section 2.1, this is the limiting step in mRNA degradation. The initial cleavage by RNase E is followed by the complete degradation of mRNAs into ribonucleotides, which implies that the reaction is irreversible and we do not expect end-product inhibition. So essentially, the simplest mechanistic description of mRNA degradation would simplify the reaction steps in Fig. 1 to the binding of free RNase E to a given mRNA i and its subsequent cleavage according to the general biochemical equation:



The rate constants k_i^+ and k_i^- denote respectively the rate of association or dissociation of mRNA_i with the enzyme. Sequence and structure specificities may affect the rate at which RNase E binds to or unbinds from mRNAs. k_{cat} is the catalytic constant of the reaction. The nature of the 5' extremity of some mRNAs is known to affect this constant when the 5' sensing mode is the mechanism of substrate recognition. In this case, RNase E can bind to mRNAs with 5' triphosphate termini but do not cleave them due to steric hindrance (Mackie, 1998; Callaghan et al., 2005). These mRNAs need a pre-processing by RppH to convert the 5' triphosphate into a 5' monophosphate termini. The conversion of the 5' end by RppH seems to concern only a small fraction of each transcript (Luciano et al., 2017; Richards et al., 2012), but the fact has not been established for all cell mRNAs. Given that the direct entry mechanism is the most frequently used and that it functions independently of the 5' end, we will thus consider for simplicity that the catalytic constant is the same for all mRNAs. According to this view, the recognition of the 5' end is expected to act on the enzyme affinity only, in agreement with Kime et al. (2010).

The binding of RNase E to its target mRNA can be modulated, for instance by small RNAs, regulatory factors such as the proteins Hfq and CsrA, translating ribosomes or else, as we shall see later, by competition between mRNAs. In the following section, we will show how to model the biochemical equation and its regulation. This equation is reminiscent of enzymatic reactions driven by Michaelis–Menten kinetics. However, we will see that the classical Michaelis–Menten expression based on the standard quasi-steady-state approximation does not always hold for this particular system along the degradation kinetics and for all mRNAs.

3.2. Kinetic modelling using quasi-steady-state approximations

Measurements of residual mRNA concentrations are performed experimentally on populations of sampled cells. We will thus use a

deterministic modelling framework to describe the degradation kinetics. The model will predict averaged intracellular mRNA concentrations over the population of cells that can be related directly to experimental data.

In biochemistry, Michaelis–Menten–Henri kinetics is classically used to model enzymatic reactions (for review, Cornish-Bowden, 2013). The expression is derived from mass-action law and additional simplifying hypotheses introduced by Briggs and Haldane (1925), based on a standard quasi-steady-state approximation (sQSSA). It relates the reaction rate to the substrate concentration. This gives the following formula with our system:

$$v_i = -\frac{dm_i(t)}{dt} = \frac{Vm \times m_i(t)}{Km_i + m_i(t)}, \quad (3)$$

where $m_i(t)$ is the total concentration of mRNA i , assumed to be equal to the free substrate concentration: $m_i(t) \approx m_{i,free}(t)$. $Vm = k_{cat} \times E_0$ is the maximum velocity achieved by the system at saturating concentration of mRNA i and E_0 is the total enzyme concentration. The value of the Michaelis constant $Km_i = (k_{cat} + k_i^-)/k_i^+$ is defined as the substrate concentration at half maximum velocity. Kinetic parameters and concentrations are in standard units: molar units for Km values and intracellular concentrations, and min^{-1} for the k_{cat} .

The sQSSA form of Michaelis–Menten kinetics in Eq. (3) reduces not only the model complexity but also allows relating model predictions with experimental observations. This canonical equation is commonly used in biochemical models, without regard to the validity of the underlying assumptions. Under the sQSSA, however, the substrate concentration must be greater than the enzyme concentration and the following relation should hold in our conditions: $Km_i + m_i^T \gg E_0$, where m_i^T is the total concentration of mRNA i corresponding to the concentrations of the free form of mRNA i , its form complexed to RNase E and the degradation product. This condition is verified in most *in vitro* enzymatic assays, where the enzyme concentration is chosen much lower than the substrate concentration.

However, the sQSSA condition can break down in *in vivo* conditions (e.g. Albe et al., 1990; Sols and Marco, 1970). Intra-cellular enzyme concentrations are usually higher or at least of the same magnitude as their substrates. Consequently, a significant fraction of substrates can be bound as complexes (Bennett et al., 2009). What happens to mRNA degradation? The difficulty to verify sQSSA validity lies in the fact that Km_i values are unknown. Available values in the literature were obtained in *in vitro* assays using synthetic RNAs and resulted in variable values whose validity in *in vivo* conditions is questionable (Jeske et al., 2018; Jiang and Belasco, 2004; Kim et al., 2004; Kime et al., 2008; Kime et al., 2010). We expect that Km values *in vivo* vary between mRNAs due to sequence and structure specificities. Regulations by small RNAs or RNA-binding proteins may affect as well the affinity of the enzyme for its substrates. Quantitative data are available for the concentrations of RNase E and mRNAs though. For example, Valgepea et al. (2013) have quantified these concentrations in *E. coli* cultures growing at various rates in steady state. It appears that the median mRNA concentration at $0.48h^{-1}$ (0.5 nM) is way smaller than the concentration of the tetrameric form of RNase E (411 nM). This enzyme excess with respect to individual mRNAs is no longer observed at the mRNA population level, since the total concentration of mRNAs is much larger (5927 nM).

An alternative to sQSSA is the total quasi-steady-state approximation (tQSSA) (Borghans et al., 1996; Tzafirri, 2003). It replaces the concentration of free substrate with its total concentration. Unmistakably, the total substrate cannot be depleted by the formation of a complex with the enzyme. As a consequence of this simple modification, tQSSA is valid for almost any substrate and enzyme

concentrations. Its domain of validity is also larger than a third approximation, the reverse quasi-steady state approximation (rQSSA), which assumes that the substrate concentration is in quasi-steady-state with respect to that of the enzyme-substrate complex (Schnell and Maini, 2000; Segel and Slemrod, 1989; Tzafirri and Edelman, 2004). The rQSSA is valid for high enzyme concentrations and would be appropriate to describe the degradation of individual mRNAs but not for a population of mRNAs. For these reasons of larger domain of validity, we will thus use the total QSSA of the Michaelis–Menten equation. Moreover, this in line with the experimental procedure, which most likely quantifies the total pool of mRNAs due to their deproteination following cell lysis. In the next section, we describe the modelling of the degradation of individual mRNAs using total QSSA. We will extend further the model in Section 3.4 with mRNA competition.

3.3. Kinetic modelling of the enzymatic degradation of individual mRNAs

We start here with the simplest case of mRNA degradation, where mRNAs are isolated and do not compete for binding to RNase E. This allows us to develop a simple model for the degradation of isolated mRNAs, which we will refer to as 'isolated system' in the rest of the article. We denote $m_{i,free}(t)$ the free concentration of mRNA i . $c_i(t)$ is the concentration of complex RNase E : mRNA $_i$. We have the following mass conservation relation for the total concentration of mRNA i :

$$m_i(t) = m_{i,free}(t) + c_i(t). \quad (4)$$

Employing the total substrate concentration instead of the free concentration results in the so-called single reaction tQSSA (Borghans et al., 1996; Tzafirri, 2003). This equation is virtually valid for any substrate and enzyme concentrations:

$$\frac{dm_i(t)}{dt} = -k_{cat} \times \frac{E_0 + Km_i + m_i(t)}{2} - \frac{\sqrt{(E_0 + Km_i + m_i(t))^2 - 4 \times E_0 \times m_i(t)}}{2}, \quad m_i(0) = m_{i0}. \quad (5)$$

From the somewhat complex expression in (5), Tzafirri (2003) developed an approximation called first order tQSSA. It resembles the sQSSA form of Michaelis–Menten kinetics:

$$\frac{dm_i(t)}{dt} = \frac{k_{cat} \times E_0 \times m_i(t)}{Km_i + E_0 + m_i(t)} \quad (6)$$

The Eq. (6) is valid if either of the following criteria is satisfied. Let K_i be the individual constant of Van Slyke defined by $K_i = k_{cat}/k_i^+$:

$$E_0 + Km_i \gg m_{i0} \text{ and } K_i \ll Km_i \text{ or :} \quad (7)$$

$$E_0 \gg m_{i0} \text{ and } E_0 \gg Km_i \approx K_i \quad (8)$$

As for the sQSSA form of the Michaelis–Menten, model parameters are expressed in the same international standard units. Note that the comparison of model predictions generated by these models with experimental data may require preliminary conversions or normalizations. Indeed, some experimental protocols used for mRNA quantification generate data in arbitrary units. For instance, microarrays quantifying residual mRNAs generate concentrations in arbitrary units per gram dry weight (see Esquerré et al., 2014 for instance), and Ct values for qRT-PCR data (e.g. Nouaille et al., 2017).

3.4. Model extension with competition between mRNAs

The tQSSA derived in the previous section is well adapted for the description of the degradation of isolated mRNAs. However, the reality is that cells include a large number of mRNAs that share

One model describes the reaction kinetics in situations where mRNAs are isolated and degraded independently of other mRNAs (Eq. (6)). The second model includes the effect of competition between substrates, where the formation of complexes between RNase E and mRNAs possibly affects the formation of other complexes (Eq. (10)). Whether this is a true phenomenon in the control of mRNA turnover or not has not been established. Two recent studies suggest that this might be the case, since mRNAs with higher concentrations are destabilized in the bacteria *E. coli* and *Lactococcus lactis* (Esquerré et al., 2015; Nouaille et al., 2017). It has been proposed that higher mRNA concentrations increase the probability that RNase E encounters its substrates in these two bacterial strains.

Although our two models are relatively simple at first glance, the introduction of competition between mRNAs couples the ordinary differential equations. This makes the system behaviour difficult to understand intuitively, in particular when studying degradation of all cell mRNAs. We will rather rely on numerical simulations for our purpose. This requires setting parameters to reasonable physiological values in order to predict degradation kinetics. Since literature data on the subject are generally scarce, we will use data obtained in the bacterium *E. coli* for which more knowledge and quantitative information are available. However, the model structure is sufficiently general and the mechanisms considered, well conserved among bacteria, for the model to be valid in other biological backgrounds.

4.1. Model parameters and initial conditions

Our objective with the models is to generate progress curves comparable to those obtained experimentally. The experimental curves vary a lot, from exponential to more linear shapes (see Appendix A.1 as an example). Their fitting by an exponential function allows determining mRNA half-lives ranging from one up to a dozen of minutes. This variability in mRNA stability reflects the diversity of mRNA sequence and structure determinants that modulate the enzyme affinity for its targets. It also results from regulatory mechanisms, such as regulations by small RNAs or mRNA protection by ribosomes to name but a few, which also impact the binding of RNase E to mRNAs. According to our modelling hypotheses described in Section 3.1, this results in an expected large variability of Km values because the k_{cat} value is held constant between mRNAs.

To account for this variability, we constructed a parameter distribution for Km values based on diverse biological information. First, we chose a log-normal distribution to guarantee positive

Km values (Fig. 3), where the mean μ_{Km} is set to the empirical median value of 13 different Km values estimated from experimental data (Jeske et al., 2018; Jiang and Belasco, 2004; Kim et al., 2004; Kime et al., 2008; Kime et al., 2010) (see also Table B.1 for some values). The dispersion of these data did not reflect the expected variability of Km values *in vivo*. We thus chose the standard deviation σ_{Km} such that it mirrors the variability of mRNA concentrations measured experimentally: $\sigma_{Km} = \sigma_{m0}$, similarly to what is found generally for many enzymes and their substrates (e.g. Bennett et al., 2009). The chosen k_{cat} and Km_i values result in a mean catalytic efficiency constant of $6 \cdot 10^5 \text{ s}^{-1} \cdot \text{M}^{-1}$, in the average of values determined experimentally for RNase E (Table B.1). The mRNA concentrations at time 0, together with the RNase E concentration, correspond to transcriptomics and proteomics data in (Valgepea et al., 2013), which we discussed in Section 3.2. The mRNA concentrations form a log-normal distribution illustrated on a logarithmic scale in Panel (a) of Fig. 3. The corresponding distribution for Km values is shown in Panel (b). On the logarithmic scale, the distribution is normal, while it is characterized by a longer tail towards higher Km values on a linear scale. It reflects the proportion of mRNAs potentially affected by specific regulations antagonizing RNase E (less than 10% with the chosen variance). As well, an even smaller fraction of mRNAs have low Km values, reflecting possible regulatory mechanisms activating mRNA decay. This is altogether a reasonable assumption, given the similar fraction of mRNAs shown or predicted to be regulated at the degradation level (Keseler et al., 2017; Wright et al., 2014). The distribution of initial mRNA concentrations is also characterized by a longer tail, where few mRNAs have high concentrations. Most of them are present at low intracellular levels. For the catalytic constant, we used the mean value estimated from experimental data obtained with various mRNAs (Jeske et al., 2018). Overall, the distributions and parameter values set verify the criteria of validity for the tQSSA described in Sections 3.3 and 3.4. Remember that in these data, the individual mRNA concentrations are well below the RNase E concentration, while the total mRNA concentration is in large excess (Fig. 3).

4.2. Competition between mRNAs slows down the degradation kinetics and increases the variability of progress curves

With the parameter values and concentrations described above, we simulated the kinetic models of the isolated and competitive systems. The transcriptomics data used for the initial conditions includes information for 4312 mRNAs. We thus ran as many simulations. We took every time a Km value from the log-normal distri-

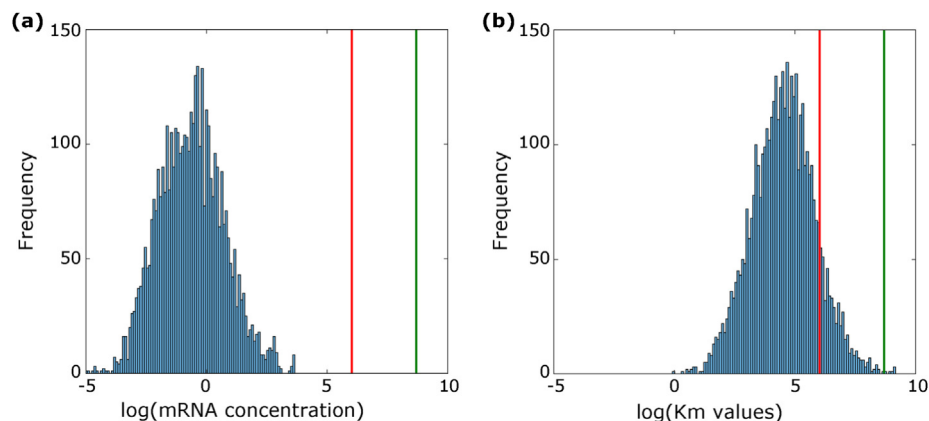


Fig. 3. Log-normal distributions of initial concentrations (a) and Km values (b). The x-axes are on a logarithmic scale. The red and green lines denote the total concentrations of RNase E and mRNA, respectively.

bution as well as the fixed catalytic constant and enzyme concentration. The same values of parameters and initial conditions were used to simulate the models with and without competition. All simulations were performed with the function `ode15s` of Matlab (Mathworks).

Remember that the simulated models are approximated forms of the tQSSA, valid under the criteria described in Sections 3.3 and 3.4. We could not exclude that some combinations of parameters and initial concentrations may violate these criteria. Beforehand we thus compared the predictions generated by the various tQSSA models to make sure that approximations of poorest quality did not impact significantly the degradation profiles (see Appendix D.1). However, the strong resemblance of the model predictions prompted us to use the approximated versions in the rest of the study.

Fig. 4 describes the simulation results obtained with the approximated tQSSA models. Predicted mRNA concentrations are shown for the competitive system in Panel (a) and the isolated system in Panel (b) (see Fig. D.2 for a log scale representation of these results). The curves differ by the initial mRNA concentration and the K_m value. However, the variability of the mRNA concentrations

is such that it hides the effect of differences of RNase E affinity, represented by curves from red (high K_m values) to blue (low K_m). These differences are more visible in Panels (c) and (d), where the same curves normalized to the initial mRNA concentrations are shown. They differ by their K_m value only. A striking difference between the isolated and competitive systems is the rapidity of the decay in the absence of competition. After two minutes, all mRNAs are at least half degraded in the isolated system (Panel (d)), whereas it takes almost five minutes to the competitive system to reach the same state (Panel (c)). This slower rate of degradation is consistent with measurements of mRNA half-lives in *E. coli* cells (Bernstein et al., 2002; Esquerré et al., 2014). For instance in Esquerré et al., 2014, mRNAs have median half-lives ranging from 2.9 to 4.2 min in *E. coli* cells at various doubling rates. This general slowdown of mRNA degradation results from the reduced RNase E affinity in a competitive system compared to an isolated one (Section 3.4). This can be easily seen in Fig. E.1, where the distribution at time 0 of the apparent K_m values (including the competition term) is shifted to a higher median value compared to the distribution of K_m values.

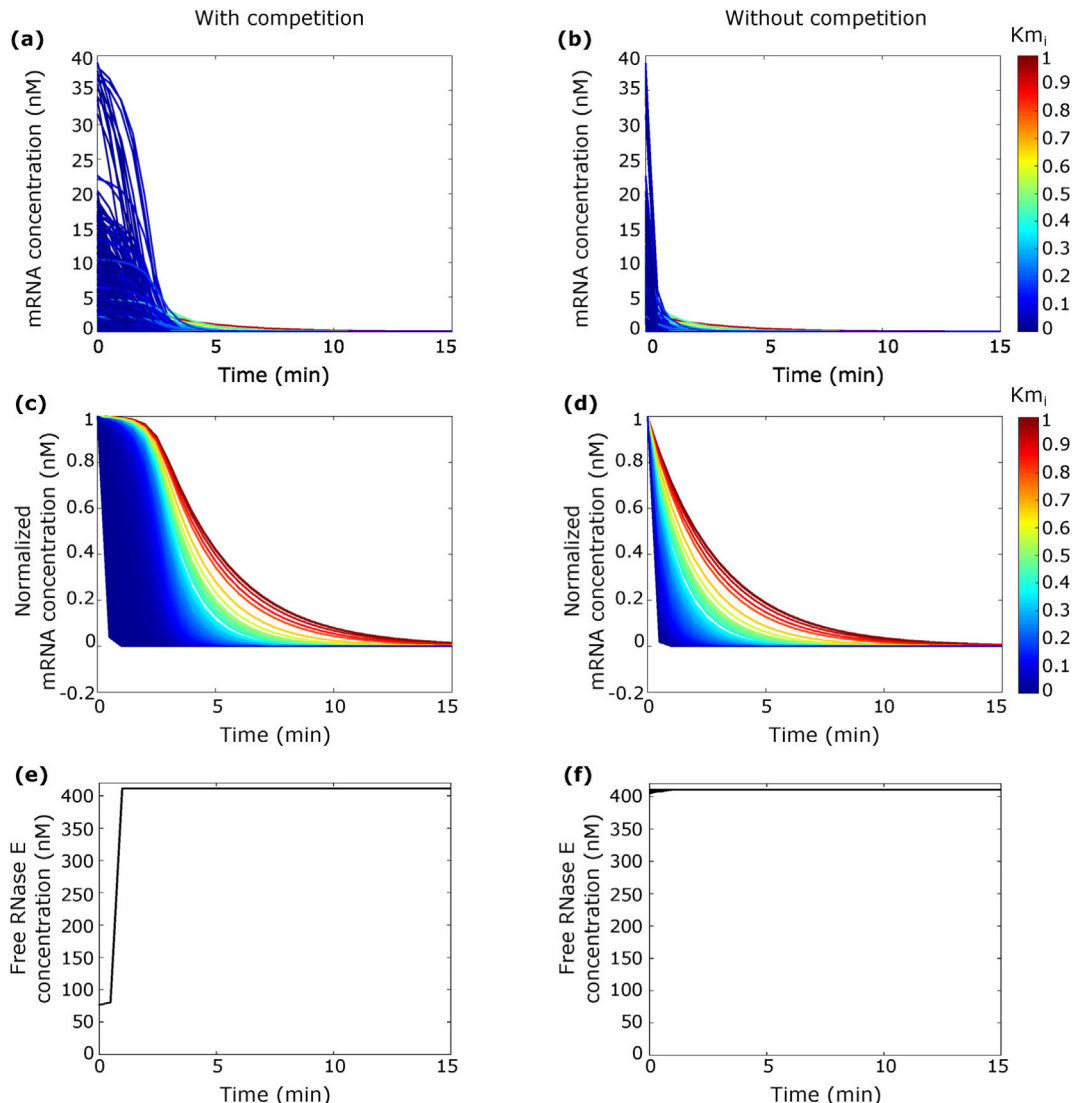


Fig. 4. Numerical simulation of mRNA degradation kinetics in isolated and competitive systems. Predicted profiles for the competitive (a) and isolated (b) systems. The profiles are normalized to their respective initial concentrations for the competitive (c) and isolated (d) systems. Predicted free RNase E concentrations for (e) the competitive system ($E_{free} = E_0 - \sum_i 1nc_i(t)$) and (f) the isolated one ($E_{i,free} = E_0 - c_i(t)$). 4312 curves are displayed in this case, due to the lack of coupling between mRNAs. The colour bars on the right side represent the normalized gradient of K_m values, on a scale from zero (the minimal K_m value) to one (maximal value).

The competitive effects result in more variable and more non-linear degradation profiles. Hence, the competition affects how degradation is launched. It starts immediately in isolated systems while it is delayed in competitive ones (Fig. 4(a,b)). This delay is slightly more pronounced with the approximated tQSSA, but still present in the absence of approximation (see Fig. D.1 for comparison). It is therefore a feature of competing mRNAs. The occurrence of delays in mRNA degradation kinetics is a known phenomenon. The primary reason is the common use of rifampicin as a drug to block the initiation of transcription (e.g. Chen et al., 2015). Elongating RNA polymerase is insensitive to the drug and pursues transcription. During a transient time, the transcription rate counterbalances the degradation rate. It is thus expected that longer transcripts have longer delays before mRNA levels start decreasing (Chen et al., 2015; Esquerré et al., 2015; Moffitt et al., 2016). Are there other determinants to the delay before degradation?

In our study we have not modelled the delay explicitly, at least not the time required for the completion of transcription. The predicted delay results from the titration of RNase E by competing mRNAs. As shown in Panel (e), the intracellular pool of free RNase E is strongly depleted in the first minutes of the degradation kinetics in the competitive system. While the enzyme level is indeed much higher than those of individual mRNAs, the total mRNA concentration is bigger and the formation of complexes between the enzyme and the mRNAs limits the enzyme availability. This implies that sole mRNAs at high initial concentration or with low Km values are more likely to be degraded by this small enzymatic pool. The intracellular pool of RNase E raises to higher levels when mRNAs are degraded to a large extent. At this time of the degradation kinetics, RNase E is in excess and all mRNAs can be degraded. This is what happens during the whole kinetics for mRNAs in the isolated system, since the enzyme concentration is never limiting (Panel (f)). The mRNAs can all bind to RNase E and be degraded soon afterwards.

Another striking difference is the variability of the degradation profiles driven by competing mRNAs (compare Panels (a) and (b) or (c) and (d)). Even though we set the variance of the Km distribution to quite a large value, it is clearly not enough to produce a large diversity of curves in the isolated system. With the same parameters and initial concentrations, the introduction of the competition term gives rise to curves with variable shapes. This is especially true for the curves in dark blue in Panel (d), characterized by low Km values. Addition of the competition term spread them out, as can be seen in Panel (c). Cyan to red curves, characterized by a high Km, are not scattered but simply shifted. The corresponding mRNAs seem merely to be stabilized.

Where do these variability and increased non linearity come from? As can be seen in Panels (a) and (b) for both isolated and competitive systems, many mRNAs have a low initial concentration. Many of them have also higher Km values, as shown in Panels (c) and (d) (lines coloured from cyan to red). In these conditions, the mRNA concentration becomes negligible with respect to the Km and/or the RNase concentration: $m_{i0} \ll Km_i + E_0$. Therefore, the kinetics in Eq. (6) can be approximated by the following system for isolated mRNAs:

$$\frac{dm_i(t)}{dt} \approx \frac{k_{cat} \times E_0}{Km_i + E_0} \times m_i(t). \quad (11)$$

For mRNAs in competition, it becomes possible to approximate the degradation kinetics in Eq. (10) by:

$$\frac{dm_i(t)}{dt} \approx \frac{k_{cat} \times E_0}{Km_i \left(1 + \sum_{j \neq i} \frac{m_j(t)}{Km_j} \right) + E_0} \times m_i(t). \quad (12)$$

In both expressions, the degradation rate varies linearly with the mRNA concentration. Note as a side comment that the solution of these ODEs is an exponential function comparable to the classical exponential model used to determine mRNA half-life (see Eq. (2)). It is thus appropriate to fit the degradation profiles of these particular mRNAs (higher Km and/or lower initial concentration) with an exponential model. The decrease of enzyme affinity resulting from the increased apparent Km value in the competitive system (Fig. E.1(a)) globally shifts the curves to the right compared to the isolated system (look for instance at the cyan to red curves in Panel (c) versus (d)). All mRNAs with higher Km and/or lower initial concentrations will react in the same manner. While they tend to be stable in the isolated system, competition stabilizes them even more.

The situation is different for mRNAs characterized by higher initial concentrations and/or lower Km values. In these conditions, the approximation in Eqs. (11) and (12) does not hold any more. RNase E is closer to saturation at the beginning of the degradation kinetics (see Fig. 4), the substrates start to compete and their degradation rates depend on each other. The degradation rate is no longer linear in the mRNA concentration and the non linearity is amplified by the multiplication of the Km value with the competitive term. We thus expect that these mRNAs react differently to the competition. This is visible in Fig. E.1(b), where normalized values of the Km and apparent Km for each mRNA are compared. We observe the global rescaling of the apparent Km values at time 0 in the competitive system. However, for a non negligible part of the mRNAs, the trend of the rescaling is different. mRNAs with higher Km values in the isolated system can have lower apparent Km values in the competitive system. It is only later in the kinetics, when mRNAs are degraded in large parts, that the curves for the isolated and competitive systems are more alike. In these conditions, the mRNA concentration becomes negligible. The degradation kinetics become again linear in the mRNA concentration and can be approximated by the simplified expressions in Eqs. (11) and (12).

In summary, competition between mRNAs acts globally on mRNA decay by decreasing RNase E affinity for its substrates and by retarding the onset of degradation. It also acts on each mRNA individually. To better understand how the own characteristics of mRNAs in terms of enzyme affinity and initial concentrations lead to variable responses to competition, we determined rate sensitivity coefficients for the two models, as we shall see in the following section.

5. Determination of rate sensitivity coefficients disentangles the effect of substrate concentration and enzyme affinity in competitive systems

In our models, mRNAs start to decay following the transcription arrest at time zero. The initial velocity of the degradation reaction depends on the saturation of RNase E. Any change in the initial concentrations of mRNAs and the enzyme affinity is expected to affect the degree of saturation. This is exactly what bacteria experience in nature. They constantly adjust their growth rate to changing environmental conditions. This involves a wide reorganization of gene expression, which includes, among others, changes in mRNA abundance and half-life resulting notably from post-transcriptional regulations (see Bernstein et al., 2004; Esquerré et al., 2014 for instance). The latter affect the enzyme affinity for its substrates in our modelling framework. The consequence is that mRNA stabilities should react differently to these changes according to our simulation results in Section 4.2.

The determination of scaled rate sensitivity coefficients allow assessing how changes of initial concentrations and RNase E affin-

ity impact the degradation rates. For each model, we derived such a coefficient for the initial velocity, by determining the scaled change of the initial degradation rate of each mRNA i upon a variation of mRNA concentrations at time 0 or of Km value. These coefficients are the equivalent of the elasticities and control coefficients determined at steady state in Metabolic Control Analysis (Heinrich and Schuster, 1996; Ingalls and Sauro, 2003). Their analytical expressions are shown in Table 1.

We first examined how changing the initial concentration (m_{i0}) or Km value (Km_i) of a messenger RNA i affects the initial velocity (v_i) of its degradation reaction. The sign of the analytical expressions of the response coefficients in Table 1 shows that the direction of the change is the same in both isolated and competitive systems: increasing the initial concentration destabilizes the mRNA ($R_{m_i}^{v_i} \geq 0$), while increasing the Km value (and thus decreasing the enzyme affinity) stabilizes it ($R_{Km_i}^{v_i} \leq 0$). However, the amplitude of the change may differ between both systems. We quantified it, by evaluating the analytical expressions of the response coefficients for the various pairs of mRNA concentration and Km used for the numerical simulations in Section 4.2. The linear interpolations of the coefficients are reported as contour plots in Fig. 5. We also displayed as points the pairs of concentration and Km value used to generate the degradation profiles in Section 4.2.

The response coefficients with respect to the change in mRNA concentrations are close to unity for both models (see Panels (a) and (b)). This means that increasing the initial concentration by 10 % results in a similar ~10% increase in the rate of degradation and conversely, a decrease in the time at which mRNAs are half degraded. The coefficients tend to be even higher for the competitive system. The initial velocity of mRNA degradation is thus essentially determined by the initial messenger concentration and this effect is reinforced in competitive systems. This agrees well with experimental observations, where mRNA concentration has been found negatively correlated with mRNA half time in *E. coli* and *L. lactis* bacteria (Esquerré et al., 2015; Nouaille et al., 2017). To reach this conclusion, the authors of these studies analysed large data sets previously obtained in these bacteria growing in environmental conditions leading to changes in mRNA concentrations and half-lives. Increasing specifically the concentrations of four mRNAs by means of inducible promoters elicited again the same negative relationship between mRNA concentration and stability (Nouaille et al., 2017). Similarly to our observations, the correlation is more

pronounced at low to middle mRNA concentrations, but less at high mRNA concentration ($R_{m_i}^{v_i}$ coefficients are a bit smaller).

We then studied how the isolated and competitive systems react to changes in enzyme affinity. The corresponding response coefficients are shown in Panels (c) and (d) of Fig. 5. For the pairs of initial concentration and Km value used for the numerical simulations in Fig. 4, changing the Km value by 10% leads, on average, to a 3% decrease of the initial velocity in the isolated system ($R_{Km_i}^{v_i} \approx -0.3$) and to a 9% decrease in the competitive system ($R_{Km_i}^{v_i} \approx -0.9$). This effect does not depend on the initial mRNA concentration. It is stronger at higher Km, but the range of Km values for which the system response is highly negative is much larger in the competitive system than in the isolated one. It encompasses the majority of the pairs of initial concentration and Km value used in the numerical simulations.

It is interesting to realize that initial concentrations impact the degradation kinetics at two different levels in competitive systems. First of all, as described above, they strongly influence mRNA stability due to the strong, positive response of the degradation rate to changes of concentration. Secondly, they have an antagonistic effect on the degradation rate in the competitive system, due to the strong negative response to Km changes. This response to Km changes is low in the isolated system, which implies that the enzyme affinity hardly compensates for the effect of the initial concentration ($R_{Km_i}^{v_i} \approx -0.3$ and $R_{m_i}^{v_i} \approx 1$ on average for the pairs of Km and initial concentration used in the numerical simulations). The degradation kinetics of an isolated system is thus almost exclusively driven by the initial mRNA concentration. Now, in the competitive system, the response coefficients are of opposite sign and similar amplitude ($R_{Km_i}^{v_i} \approx -0.9$ and $R_{m_i}^{v_i} \approx 1$ on average). The combination of these two antagonistic effects explain the non linearities observed in the predicted degradation curves in Fig. 4(a,c), in particular the occurrence of delayed degradation when mRNAs compete between them. Indeed, as long as the total mRNA concentration is high at the beginning of the kinetics, it augments the apparent Km value to sufficiently high values to compensate for the effect of the initial mRNA concentrations. RNase E is totally saturated and the initial velocity is almost null. It is only when the total mRNA concentration decreases significantly that the weight of the competitive term in Eq. (10) fades away. RNase E is no longer saturated and the competitive system resembles the isolated one, with less sensitivity to changes of enzyme affinity.

With these analyses, we were able to disentangle the contribution of Km and mRNA concentration to the degradation kinetics in competitive systems. One question remains open though. Since the degradation kinetics of all mRNAs are coupled due to the competition, are they impacted by the cell composition in mRNAs? The competition term in Eq. (10) includes the affinity of cell mRNAs j different from the mRNA i of interest, as well as their concentration. Given that there are 4312 different mRNAs in our biological system, the intuition would say that the change of concentration or Km of a single mRNA should not affect the degradation kinetics of another one. But how true is it? We further study this question in the following section.

6. Effect of coupling on mRNA degradation kinetics

We analysed a new set of rate sensitivity coefficients to analyse how the changes of concentration or Km value of another mRNA j affect the degradation kinetics of a given mRNA i . The analytical expressions for these new coefficients are shown in Table 1. The isolated system is characterized by the absence of coupling between mRNAs. Consequently, the response coefficient to changes of Km value or initial concentration of mRNA j is zero

Table 1

Scaled rate sensitivity coefficients of the initial velocity v_i with respect to changes in initial mRNA concentrations and Km values.

Model	Coefficients
Isolated system (Eq. (6))	Effect of changes of m_{i0} and Km_i on v_i :
	$R_{m_i}^{v_i} = \frac{\partial v_i}{\partial m_{i0}} \frac{m_{i0}}{v_i} = \frac{Km_i + E_0}{Km_i + E_0 + m_{i0}}$
	$R_{Km_i}^{v_i} = \frac{\partial v_i}{\partial Km_i} \frac{Km_i}{v_i} = \frac{-Km_i}{Km_i + E_0 + m_{i0}}$
	Effect of changes of m_{j0} and Km_j on v_i :
Competitive system (Eq. (10))	Effect of changes of m_{i0} and Km_i on v_i :
	$R_{m_i}^{v_i} = \frac{\partial v_i}{\partial m_{i0}} \frac{m_{i0}}{v_i} = \frac{Km_i \times \left(1 + \sum_{j \neq i} \frac{m_{j0}}{Km_j}\right) + E_0}{Km_i \times \left(1 + \sum_{j \neq i} \frac{m_{j0}}{Km_j}\right) + E_0 + m_{i0}}$
	$R_{Km_i}^{v_i} = \frac{\partial v_i}{\partial Km_i} \frac{Km_i}{v_i} = \frac{-Km_i \times \left(1 + \sum_{j \neq i} \frac{m_{j0}}{Km_j}\right)}{Km_i \times \left(1 + \sum_{j \neq i} \frac{m_{j0}}{Km_j}\right) + E_0 + m_{i0}}$
	Effect of changes of m_{j0} and Km_j on v_i :
	$R_{m_j}^{v_i} = \frac{\partial v_i}{\partial m_{j0}} \frac{m_{j0}}{v_i} = \frac{-m_{j0} \times \frac{Km_i}{Km_j}}{Km_i \times \left(1 + \sum_{j \neq i} \frac{m_{j0}}{Km_j}\right) + E_0 + m_{i0}}$
	$R_{Km_j}^{v_i} = \frac{\partial v_i}{\partial Km_j} \frac{Km_j}{v_i} = -R_{m_j}^{v_i}$

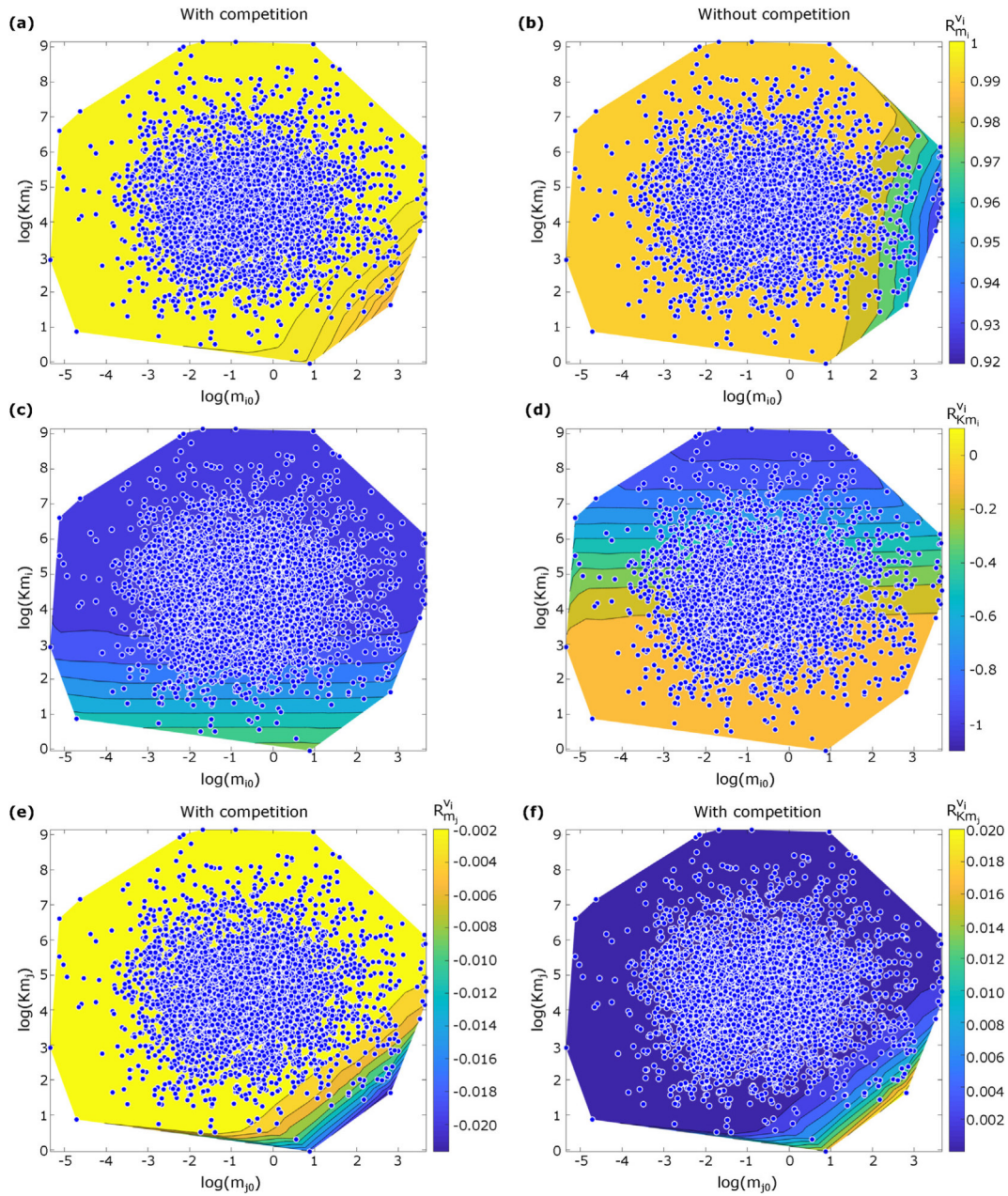


Fig. 5. Scaled rate sensitivity coefficients for the initial velocity of mRNA i degradation. The points correspond to the pairs of initial concentration and Km used in the numerical simulations in Fig. 4. For each pair, the contour plots display by means of linear interpolation the corresponding response coefficients with respect to (a) the initial concentrations ($R_m^{v_i}$) or (c) the Km ($R_{Km}^{v_i}$), in a competitive system or in the isolated system ((b) and (d), respectively). Colour bars indicate the value of the elasticities. Panels (e) and (f) display the median response coefficients to changes of mRNA j concentration ($R_m^{v_i}$) or Km ($R_{Km}^{v_i}$), respectively. The x- and y-axes are on a logarithmic scale.

($R_m^{v_i} = R_{Km}^{v_i} = 0$). In the competitive system, the analytical expression depends on the concentration of the other mRNA j and the relative affinities of mRNAs j and i . The sign of the response coefficients $R_m^{v_i}$ and $R_{Km}^{v_i}$ is opposed to the previous coefficients $R_m^{v_i}$ and $R_{Km}^{v_i}$. It shows that increasing the initial concentration of mRNA j stabilizes mRNA i ($R_m^{v_i} \leq 0$), while increasing the Km value of mRNA j (and thus decreasing the enzyme affinity) destabilizes it ($R_{Km}^{v_i} \geq 0$). As previously, this is a matter of enzyme titration. If a given mRNA j titrates more RNase E due to a high initial concentration or low Km value, less RNase E will be available for mRNA i , which is stabilized.

We evaluated the response coefficients by determining how a change of 10% of the initial concentration or Km of each of the

4312 mRNAs j affects the initial velocity of each of the 4311 other mRNAs i and j itself. The results are shown as heatmaps in Fig. F.1. In most cases, the mRNA j marginally affects the other mRNAs (response coefficients close to zero, see the yellow area in Panel (b) for $R_m^{v_i}$ and the dark blue area in Panel (c) for $R_{Km}^{v_i}$). However, when it does impact the initial velocity, it does so for all mRNAs i (see horizontal lines coloured from orange to blue in Panel (b) of Fig. F.1, and from yellow to cyan in Panel (c)). To facilitate the analysis of the response coefficients, we thus took the median response coefficients of all mRNAs i to a change of either the Km value or the initial concentration of a given mRNA j (Fig. F.1). They express how changing the characteristics of mRNA j impacts the average mRNA i . The resulting coefficients are shown in Panels (e) and (f) of Fig. 5. Each point (corresponding to a pair of initial concentration and Km

value from the numerical simulations in Fig. 4) represents a given mRNA j . The response coefficients are way smaller than their counterparts $R_{m_i}^{v_i}$ and $R_{K_{m_i}}^{v_i}$ in Panels (a) and (c), because they only concern the effect of an external mRNA j on the degradation of the average mRNA i . But interestingly, they are not negligible. The competitive effect is the strongest when the mRNA j has a high concentration or a low Km value. Increasing its concentration by 10% in this case reduces by 0.2% the initial velocity of mRNA i degradation ($R_{m_j}^{v_i} \approx -0.02$; Panel (e)). The effect is exactly the opposite when increasing the Km by 10%. This stimulates the initial degradation rate of the average mRNA i by 0.2% ($R_{K_{m_j}}^{v_i} \approx 0.02$; Panel (f)) and reduces the time at which it is half degraded. Hence, a highly competitive mRNA, which titrates more RNase E due to a high initial concentration or a low Km value, can affect the degradation rate of all cell mRNAs, with an average reduction of 0.2% of the velocity.

The competing effects that mRNAs exert on each other appear limited at first glance. However, these effects do exist whenever a single competing mRNA is considered and could be even stronger if several competing mRNAs are subject to changes. Indeed mRNA abundances and half-lives are known to massively adjust to environmental conditions (see Bernstein et al., 2004; Schneider et al., 1999 for instance). More precisely, modifications of the total mRNA concentration and the relative abundance of individual mRNAs have been observed. Changes in enzyme affinity are also expected for mRNAs specifically regulated by small RNAs and other post-transcriptional regulations. All these combined changes should affect the decay of all cell mRNAs according to our computed response coefficients. The global increase of the total mRNA concentration globally decrease the enzyme affinity, leading to a global but non uniform decrease of the individual degradation rates, since they are sensitive to various degrees to changes of Km values as we have seen in Panel (c). Among these mRNAs, those with the lowest Km values ($K_{m_i} < K_{m_j}$) and/or the highest initial concentrations ($m_{i0} > m_{j0}$) are less sensitive to changes in the competing mRNA pool, because they are saturated by RNase E (Panels (a) and (c)). They behave similarly to mRNAs in an isolated system. However, the vast majority of cell mRNAs has a low concentration (see Fig. 3) and most likely also a relatively weak affinity (Fig. 3 (b)), such that they are not saturated by RNase E. Consequently, it is well possible that a majority of cell mRNAs can be significantly affected by changes in competing mRNA pools, in particular if the latter are highly concentrated and/or affine for RNase E (panels (e) and (f)).

7. Discussion

We have developed in this study a mechanistic model of the coupled degradation of all cell mRNAs. The model represents a compromise in terms of complexity and biological realism. Its predictions can be compared directly to degradation profiles obtained experimentally. We hence assimilated the degradation mechanism by the degradosome machinery to a macro-reaction catalysed by RNase E. Literature data motivated the use of the tQSSA form of the Michaelis–Menten equation for describing the degradation kinetics. We further approximated the tQSSA model by extending previous work by Pedersen et al. (2007) and Tang and Riley (2013). The resulting simplified model allows analysing the effect of regulatory mechanisms on mRNA decay, which are difficult to test experimentally. The possible implication of mRNA competition is one of them. While there is no direct experimental proof for significant titration of RNase E, a body of evidences – on its limited concentration relative that of mRNAs and its mean catalytic efficiency – suggest that competition between mRNAs could well

impact the mRNA degradation kinetics (Section 3.4). We analysed the contribution of competition to mRNA decay by means of numerical simulations and rate sensitivity coefficients, based on carefully chosen parameter values and distributions. To the best of our knowledge, this is the first study focusing on the role of competition in mRNA turnover. A similar study was conducted on the maturation of transfer RNAs by RNase P in conditions of limiting substrates (Yandek et al., 2013), but the number of alternative substrates was low compared to our system. In addition the authors of the study opted for the classical sQSSA form of the Michaelis–Menten framework, which would not be suitable in our case, given its limits of validity.

Our results provide a plausible explanation to the observed delays and variable degradation profiles in experimental studies. Under the effect of competition, the profiles vary from seemingly exponential to more linear profiles. Such diversity is difficult to reproduce in the absence of competition, while it is observed experimentally (e.g., see Fig. A.1). The explanation lies in the titration of RNase E at the beginning of the kinetics and as long as the total mRNA concentration remains sufficiently high, which delays the onset of degradation. During this period of time, mRNAs compete between them. The consequence is that the degradation rate varies along time, from zero during the delay to maximal when free RNase E is available, and back to zero towards the end of the kinetics. Such variation is visible in experimental data (e.g., see Fig. A.1).

The lag effect caused by mRNA competition is a striking result. Residual transcription is generally associated with delayed degradation, in particular for long genes or genes towards the 3' end of operons (Chen et al., 2015). Could mRNA competition be an additional mechanism? In the absence of direct evidence of this phenomenon, through measurements of the free or bound concentration of RNase E for instance, we cannot answer formally the question. However, an analysis of available dynamical transcriptomics data in *E. coli* indicates that residual transcription alone does not suffice to explain the duration of the delay before degradation kicks in (see Appendix G for details). We estimated the delay before degradation for 3140 mRNAs of this data set (Esquerré et al., 2014), as well as the time needed to transcribe each of them. In the absence of information on the position of RNA polymerase at the time of rifampicin addition, determining the time needed for transcribing an entire mRNA gives an upper bound to the residual transcription time. Among the 2454 mRNAs that are not immediately degraded after rifampicin addition, 51% of them have a delay before degradation larger than the time needed for transcription. The delay is even twice larger for 21% of them (Fig. G.2). While this quick analysis clearly underestimates the number of mRNAs for which the transcription elongation takes a shorter time than the delay before degradation, it indicates that residual transcription is not the sole determinant for the delay in the data set studied. Competition between mRNAs could well be another one.

The competition explains also the negative correlation observed between mRNA half-life and concentration (Esquerré et al., 2015; Nouaille et al., 2017). Our model has the advantage that it provides a mechanistic interpretation to the correlation, through the coupling of the degradation kinetics of each mRNA. This results from the fact that the competition acts at two different levels. First of all, the degradation rate and thus the time at which the mRNA is half degraded is extremely sensitive to mRNA concentration (Fig. 5(a)). This explains why the correlation is the highest in the experimental data, in particular for mRNAs with low to intermediate concentrations (Nouaille et al., 2017). At higher mRNA levels, the degradation rate becomes less sensitive to the concentrations while it can be still strongly sensitive to the Km values (Fig. 5(a, b)). This is where the correlation between half-life and concentra-

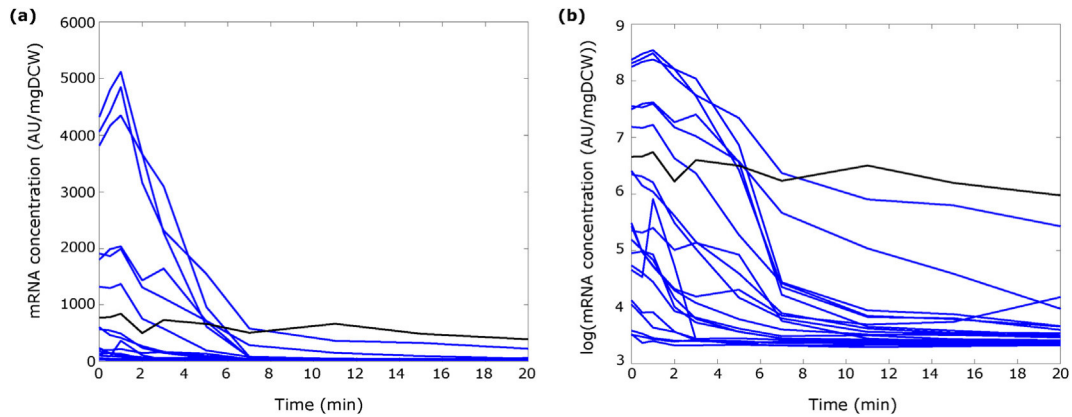


Fig. A.1. Experimental monitoring of mRNA decay in *E. coli* cells growing in continuous culture at 0.4 h^{-1} . For readability, only 21 out of the 4254 generated profiles are displayed, both on linear (a) and semi-logarithmic (b) scales. At time 0 of the experiment, rifampicin was added to the culture and residual mRNA concentrations were quantified by DNA micro-arrays. Data were taken from (Esquerré et al., 2014). After delays of various duration, most mRNAs have a degradation kinetics resembling an exponential decay (blue curves), while other mRNAs can be slowly decaying, like the murein lipoprotein mRNA *lpp* (black curve).

tion is smaller: the mRNA half-lives no longer vary with high mRNA concentrations.

Our simulation results and the response coefficients show that mRNA competition amplifies the effect of the K_m value in the degradation kinetics. mRNAs that have intermediate to low affinities compete for binding to RNase E. The less affine mRNAs will have less chance to be degraded: the competition term in their apparent K_m value in Eq. (10) increases due to the presence of competing mRNAs with higher affinities (the ratio $\sum_{j \neq i} \frac{m_j(t)}{K_{mi}(t)}$ takes higher values). This results in a higher apparent K_m reflecting a stabilisation of these mRNAs (Fig. E.1(b)). Hence, the more stable mRNAs are stabilized by competition. On the contrary, the most affine mRNAs are saturated by RNase E, in both isolated and competitive systems. They are among the less sensitive to the competition and can bind to the low enzyme levels. The ratio $\sum_{j \neq i} \frac{m_j(t)}{K_{mi}(t)}$ in the competition term is among the smallest throughout the population of mRNAs, because the competing mRNAs have higher affinities. These mRNAs will have the smallest apparent K_m values among the population of mRNAs (Fig. E.1(b)). The consequence is that these unstable mRNAs in the isolated system are destabilized by competing mRNAs in the competitive system.

By affecting differently the mRNAs, the competition brings about a diversity of effects that is responsible for the larger variability of degradation profiles in a competitive system with respect to an isolated one. Surprisingly, a single mRNA can affect the degradation profiles of the other cell mRNAs, provided it is sufficiently concentrated or affine for RNase E. All these observations are interesting because it implies that cells can efficiently control the stability of their entire population of mRNAs through the sole use of a competitive mechanism and a limiting enzyme concentration. This suggests that mRNA competition allows cells to adjust degradation to transcription. Similar coupling effects are observed in Eukaryotes and Prokaryotes like *E. coli*. Buffering mechanisms against changes in mRNA stability or transcription ensure that intracellular levels of mRNAs are maintained for instance, although the molecular mechanisms have not been fully elucidated (Dori-Bachash et al., 2011, 2012; Esquerré et al., 2014; Shalem et al., 2011; Sun et al., 2012; see Hartenian and Glaunsinger, 2019 for review). In our simulation results, the positive coupling between transcription and degradation results from the higher probability that mRNAs with increased concentration will be degraded. However, we also observe a counter-acting effect that negatively correlates transcription and degradation rates. The increase of mRNA concentration through transcription increases RNase E apparent

affinity and slows down degradation, although the amplitude of the effect varies between mRNAs.

The occurrence of non linearities resulting from enzyme titration is a well studied phenomenon in biochemistry. It has been observed in many biological systems, ranging from the MAP kinases to transcription and translation (e.g. Buchler and Cross, 2009; De Vos et al., 2011; Huang and Ferrell, 1996; Mehra and Hatzimanikatis, 2006). Quite often, this results in ultrasensitive responses. We have not really observed such phenomenon in our simulation study. The response coefficients that we determined from the pairs of K_m value and initial concentrations used for the numerical simulations are generally between -1 and 1 . Sole responses above one in absolute value reflect an ultrasensitive behaviour. Values slightly below -1 were sometimes obtained in specific conditions, when the K_m value of more affine mRNAs i was varied. This does not mean that no ultra-sensitivity occurs in mRNA degradation since the amplitude of the response coefficients strongly depends on the chosen parametrization.

While we were able to retrieve quite precise values for the enzyme concentration, the mRNA concentrations and the catalytic constant in the literature, this prove to be more difficult for the K_m value. Available data for the enzyme affinity were obtained with a limited number of substrates in (often) *in vitro* assays. While this gave a possible average affinity of RNase E for its substrates, it did not reflect the variety of cell mRNAs. The latter were expected having different affinities, depending on their structural and sequence determinants, as well as their possible regulations by small RNAs and other regulatory proteins. This is the reason why we chose a distribution of K_m as large as the distribution of measured initial concentrations to take into account as much as possible these different situations. This said, using different K_m distributions would only tune the amplitude of the competition effect. Indeed there will be always an antagonism between the contributions of initial concentrations and enzyme affinities to the competition effect. Depending on their relative distributions, one may become prevailing. In our simulations, the choice of distributions with similar variances lead to contributions of initial concentrations and K_m of similar strength in many cases.

Estimating the K_m value from experimental data, and possibly also the catalytic constant, could help refine our analysis on the role of mRNA competition in mRNA decay. This is beyond the scope of this study and will be done elsewhere (T.A. Etienne, L. Girbal, M. Coccagn-Bousquet, D. Ropers, in preparation). Using a total quasi-steady-state approximation rather than the classical quasi-steady-state approximation is known to improve the estimation

Table B.1Catalytic efficiency of RNase E. Kinetic parameters k_{cat} and K_m were estimated from the same assays. References: (1) – Jiang and Belasco (2004); (2) – Kim et al. (2004).

Ref.	Experimental conditions	Substrates	k_{cat} (s^{-1})	K_m (M)	k_{cat}/K_m ($M^{-1} \cdot s^{-1}$)
(1)	pH 7.5, 25 °C, recombinant N-terminal domain	5' hydroxylated fluorogenic oligonucleotide	0.015	$3.3 \cdot 10^{-7}$	45455
(1)	pH 7.5, 25 °C, recombinant N-terminal domain	5' monophosphorylated fluorogenic oligonucleotide	0.014	$2.3 \cdot 10^{-7}$	60870
(2)	pH 7.5, 30 °C	AAUUU-containing RNA oligonucleotide	0.192	$7.3 \cdot 10^{-7}$	263,014
(2)	pH 7.5, 30 °C	CAUUU-containing RNA oligonucleotide	0.150	$4.6 \cdot 10^{-7}$	326,087
(2)	pH 7.5, 30 °C	GAUUU-containing RNA oligonucleotide	0.765	$1.2 \cdot 10^{-6}$	616,935
(2)	pH 7.5, 30 °C	GUUUU-containing RNA oligonucleotide	0.121	$3.7 \cdot 10^{-7}$	327,027
(2)	pH 7.5, 30 °C	UAUUU-containing RNA oligonucleotide	0.209	$8.1 \cdot 10^{-7}$	258,025
(2)	pH 7.5, 30 °C	UUUUU-containing RNA oligonucleotide	0.023	$3.2 \cdot 10^{-6}$	71,875
		Mean	0.186 (11.2 min ⁻¹)	$5.6 \cdot 10^{-7}$	246,161
		Median	0.136 (8.1 min ⁻¹)	$4.1 \cdot 10^{-7}$	260,519
		Standard deviation	0.247	$3.4 \cdot 10^{-7}$	191,168

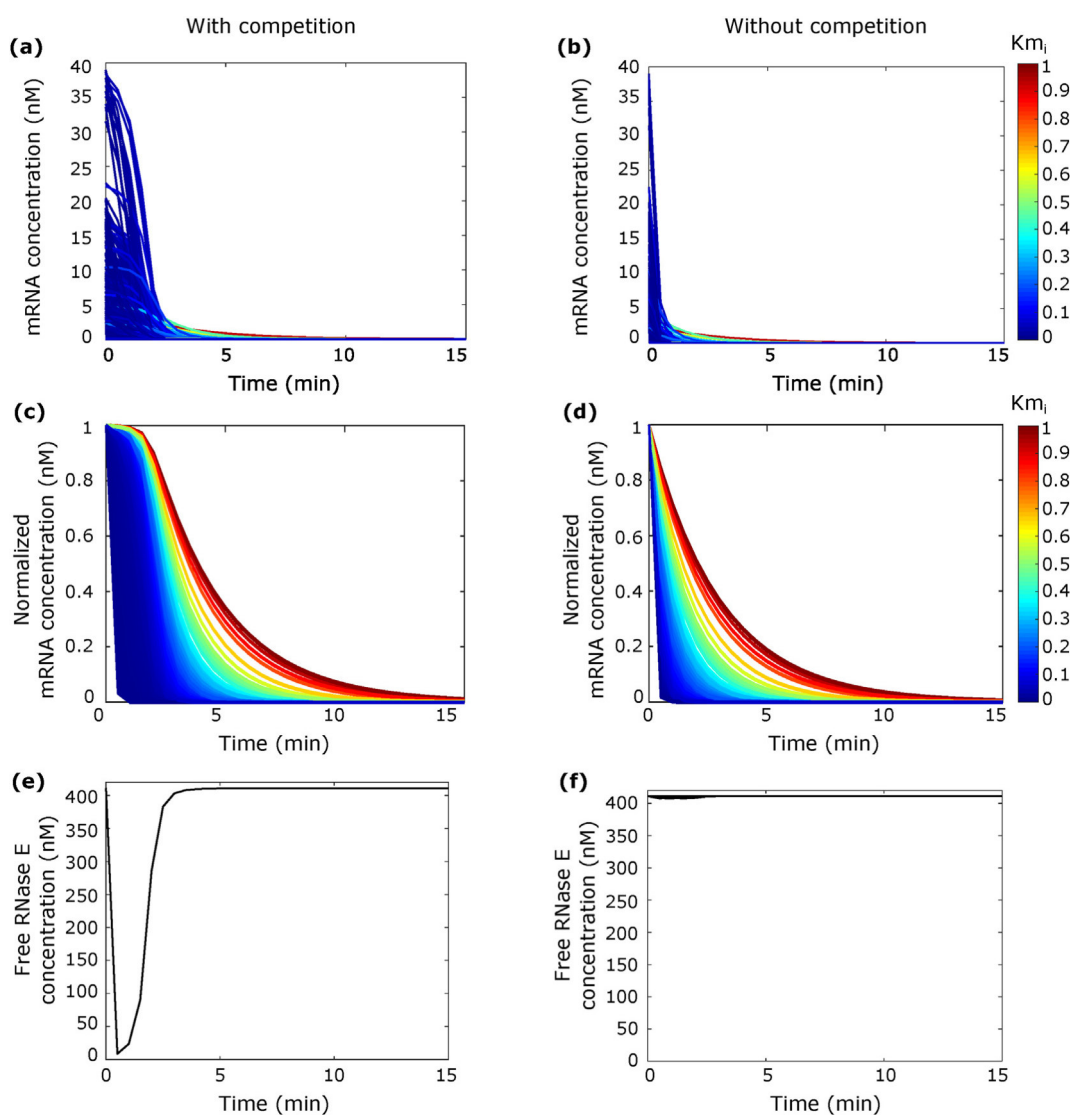


Fig. D.1. Numerical simulation of mRNA degradation kinetics in isolated and competitive systems using Eqs. (5) and (9). Predicted profiles for the competitive (a) and isolated (b) systems. The profiles are normalized to their respective initial concentrations for the competitive (c) and isolated (d) systems. Predicted free RNase E concentrations for (e) the competitive system and (f) the isolated one. 4312 curves are displayed in this case, due to the lack of coupling between mRNAs. The colour bars on the right side represent the normalized gradient of K_m values, on a scale from zero (the minimal K_m value) to one (maximal value). Note that the total concentration of RNase E, but not the free concentration, is a variable in Eqs. (5) and (9). We thus calculated a posteriori the concentration of free RNase E from the total RNase E and complex concentrations. Since the model at time zero is not at steady state, we could not determine a steady-state value for the initial complex concentration, which would have given a more realistic value for the initial concentration of free RNase E. Instead, we used the same initial condition for the complex concentration as in the work of (Pedersen et al., 2007), with $c_i(0) = 0$. This effect is transitory only, since the free RNase E concentration predicted with Eq. (9) becomes quickly similar to that obtained with the approximated tQSSA model in Fig. 4(a).

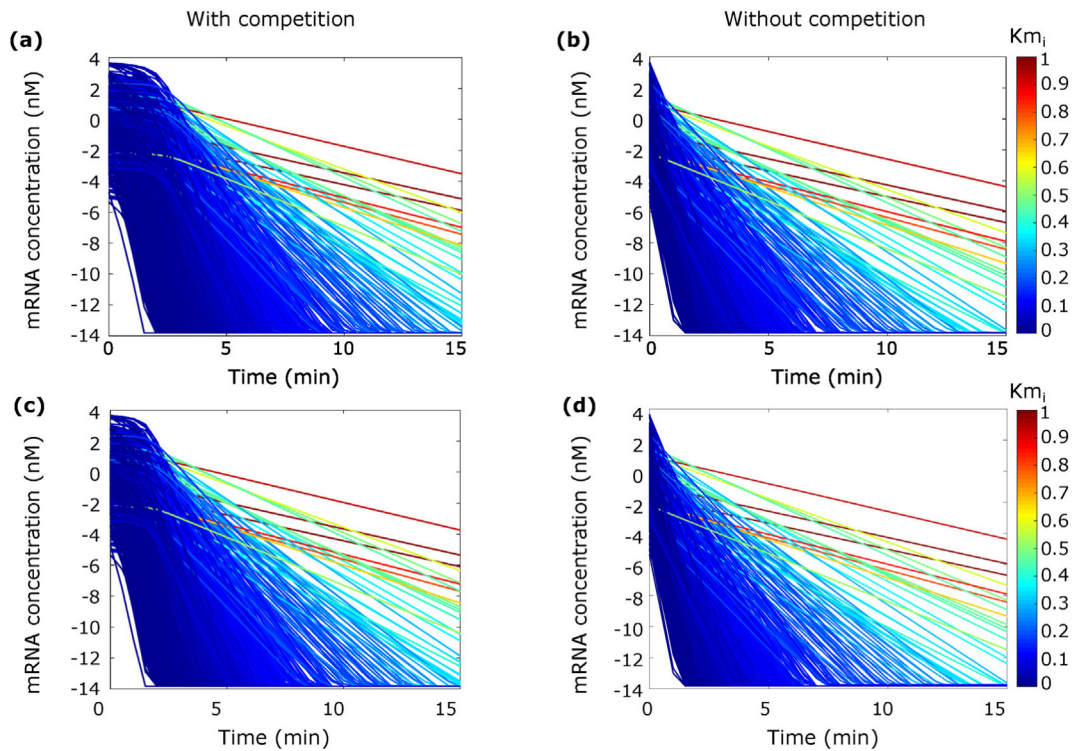


Fig. D.2. Semi-logarithmic representation of simulation results shown in Figs. 4(a,b) (Panels (a) and (b)) and D.1(a,b) (Panels (c) and (d)). The predicted profiles for the competitive (a) and isolated (b) systems were obtained with the approximated tQSSA models Eqs. (9) and (5), respectively. The predicted profiles obtained with the original models Eqs. (10) and (6) are shown in Panels (c) and (d) respectively.

of parameter values by introducing the enzyme concentration in the equation (Choi et al., 2017). In addition to helping with the identifiability problems related to the use of Michaelis–Menten kinetics, estimating the parameter values of such competition model allows to determine the true value of the K_m and not an apparent constant that includes a competition term.

The model developed is sufficiently general to be applied to other organisms and regulatory mechanisms of mRNA degradation. For instance, the apparent K_m value in our equations may be extended to include regulations by small RNAs or the competition by stable RNAs, which are also known to be matured by RNase E. The fact that RNase E is active both in the cytoplasm and bound to the membrane as part of the degradosome allows to interpret the K_m in a different manner. It also reflects the diffusion of substrates to the membrane and allows to study the effect of substrate localization in the degradation kinetics (Tang and Riley, 2013; Tzafiriri et al., 2002). Overall, we believe that our mechanistic model of mRNA degradation and its regulation by competition and other mechanisms is an interesting alternative to the classically used exponential models.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Agence Nationale de la Recherche under project RibEco (ANR-18-CE43-0010), as well as by the Institut National de la Recherche Agronomique and Inria, from which T.A.E received a doctoral fellowship. The authors thank Hidde de Jong and Marco Mauri (Ibis, Inria) for discussions, and Laurence Girbal (TBI, CNRS) for discussions as well as comments

on a previous version of the manuscript. The authors also thank an anonymous reviewer for helpful comments.

Appendix A. Example of observed degradation kinetics

We illustrate the type of degradation profiles obtained experimentally with data obtained for mRNAs in *E. coli* cells growing at 0.4 h^{-1} in continuous cultures. We display in Fig. A.1 21 degradation profiles out of the 4254 profiles obtained in (Esquerré et al., 2014).

Appendix B. Kinetic parameters of RNase E

Table B.1.

Appendix C. Development of the tQSSA model with competition

We first describe below the development of the mass-action law model in Eq. (9) and its subsequent approximations into tQSSA and first-order tQSSA models.

With $i = 1 \dots, n$, we can write the following mass-action law system:

$$\begin{aligned} \frac{dm_i(t)}{dt} &= -k_i^+ \times E_{free}(t) \times m_{i,free}(t) + k_i^- \times c_i(t) \\ \frac{dE_{i,free}(t)}{dt} &= \sum_i^n (-k_i^+ \times E_{free}(t) \times m_{i,free}(t) + (k_i^- + k_{cat}) \times c_i(t)) \end{aligned} \quad (C.1)$$

$$\frac{dc_i(t)}{dt} = k_i^+ \times [E_{free}(t) \times m_{i,free}(t)] - (k_i^- + k_{cat}) \times c_i(t)$$

$$\frac{dP_i(t)}{dt} = k_{cat} \times c_i(t),$$

with $i = 1 \dots, n$ and initial conditions at $t = 0 : m_{i,free}(0), E_{free}(0), c_i(0), P_i(0)$.

Using the conservation laws

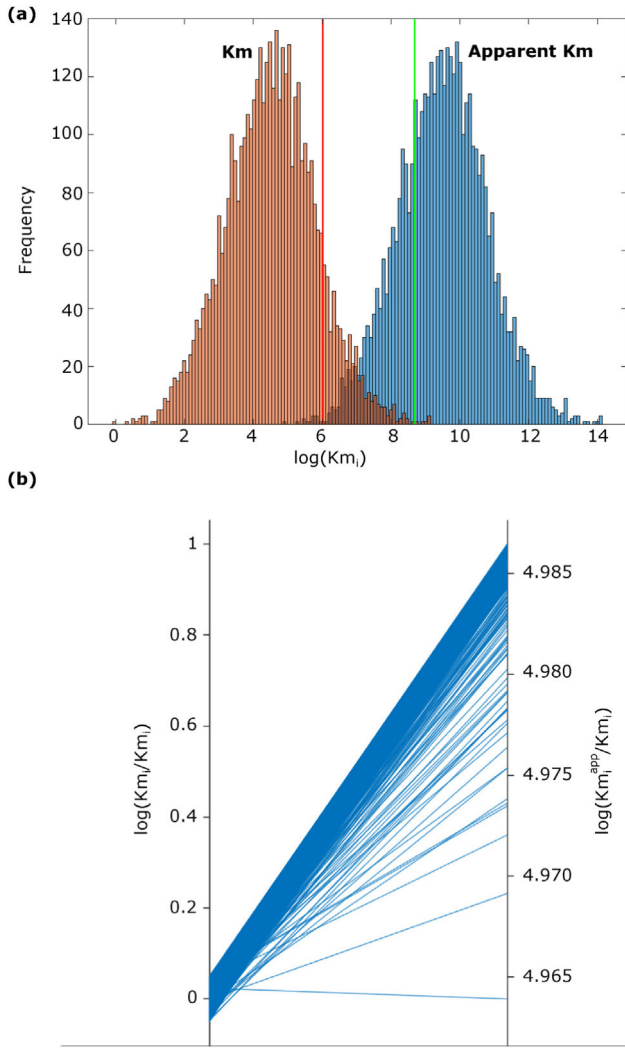


Fig. E.1. Changes in enzyme affinity upon competition. (a) Log-normal distributions of Km (red) and apparent Km (blue) values. The x-axes are on a logarithmic scale. The red and green lines denote the total concentrations of RNase E and mRNA, respectively. (b) Parallel coordinate plot of Km_i (red) and Km_i^{app} values (blue) normalized to the Km value. Axes are on a logarithmic scale.

$$\frac{dE_{i,free}(t)}{dt} + \sum_i^n \frac{dc_i(t)}{dt} = 0 \Rightarrow E_{free}(t) + \sum_i^n c_i(t) = E_0$$

$$\frac{dm_{i,free}(t)}{dt} + \frac{dc_i(t)}{dt} + \frac{dP_i(t)}{dt} = 0 \Rightarrow m_{i,free}(t) + c_i(t) + P_i(t) = m_{i,free}(t)$$

we can reduce the previous system:

$$\frac{dm_{i,free}(t)}{dt} = -k_i^+ \times (E_0 - \sum_{j=1}^n c_j(t)) \times m_{i,free} + k_i^- \times c_i(t) \quad (C.2)$$

$$\frac{dc_i(t)}{dt} = k_i^+ \times [(E_0 - \sum_{j=1}^n c_j(t)) \times m_{i,free}(t) - (k_i^- + k_{cat}) \times c_i(t)]$$

We apply the total quasi-steady-state approximation $m_i(t) = m_{i,free}(t) + c_i(t)$ and obtain:

$$\frac{dm_i(t)}{dt} = \frac{dm_{i,free}(t)}{dt} + \frac{dc_i(t)}{dt} = k_{cat} \times c_i(t) \quad (C.3)$$

$$\frac{dc_i(t)}{dt} = k_i^+ \times [(E_0 - \sum_{j=1}^n c_j(t)) \times (m_i(t) - c_i(t)) - Km_i \times c_i(t)],$$

with Km_i the Michaelis constant:

$$Km_i = \frac{k_{cat} + k_i^-}{k_i^+} \quad (C.4)$$

We apply the quasi-steady-state assumption: $\frac{dc_i(t)}{dt} \approx 0$ and obtain:

$$(E_0 - \sum_{j=1}^n c_j(t)) \times (m_i(t) - c_i(t)) = Km_i \times c_i(t) \quad (C.5)$$

and

$$c_i(t) = \frac{(E_0 - \sum_{j=1}^n c_j(t)) \times (m_i(t) - c_i(t))}{Km_i} \quad (C.6)$$

Two alternative methods have been proposed in the literature to solve the non-linear system (C.6) for c_i . Tang and Riley (2013, 2017) use perturbation theory applied to sQSSA to derive an approximated solution for the tQSSA system. Pedersen et al. (2007) develop their approximation for a competing system with two substrates by means of the total two Padé approximation. Applied to our system, these approaches allow obtaining the following formulation for the approximated concentration of complex c_i :

$$\frac{dc_i(t)}{dt} = \frac{E_0 \times m_i(t)}{Km_i \times \left(1 + \sum_{j \neq i} \frac{m_j(t)}{Km_j}\right) + E_0 + m_i(t)} \quad (C.7)$$

with $j = 1 \dots, n$.

Appendix D. Numerical simulations

D.1. Numerical simulations will full system

To exclude the fact that poor combinations of mRNA concentrations and Km values may violate the validity of the tQSSA approximation described in Section 3.4, we simulated the full system in Eqs. (5) and (9) and compared the simulation results with the profiles obtained with the approximated tQSSA in Fig. 4. Simulation results are shown in Fig. D.1.

D.2. Log-scale representation of simulation results

Appendix E. Changes of enzyme affinity upon competition

Competition rescale the enzyme affinity through the multiplication of the Km value with a term of competition. The amplitude of the changes is illustrated in Fig. E.1.

Appendix F. Heatmaps of rate sensitivity coefficients

In this section, we assess the effect of changing the initial concentration or Km value of a competing mRNA j on the initial velocity of the degradation of a given mRNA i . To that aim, we determine rate sensitivity coefficients for the effect of each of the 4312 mRNAs j on the degradation kinetics of each of the 4312 mRNAs i (they include the mRNA j itself). We obtain a square matrix of 4312×4312 coefficients, represented as heatmap in Fig. F.1.

Appendix G. Contribution of transcription elongation to the delay before degradation

Elongation of transcription is known to delay degradation, but to which extent? We used a set of dynamic transcriptomics data from Esquerré et al., 2014 to answer the question. In this study, mRNA half-lives were determined in continuous cultures of

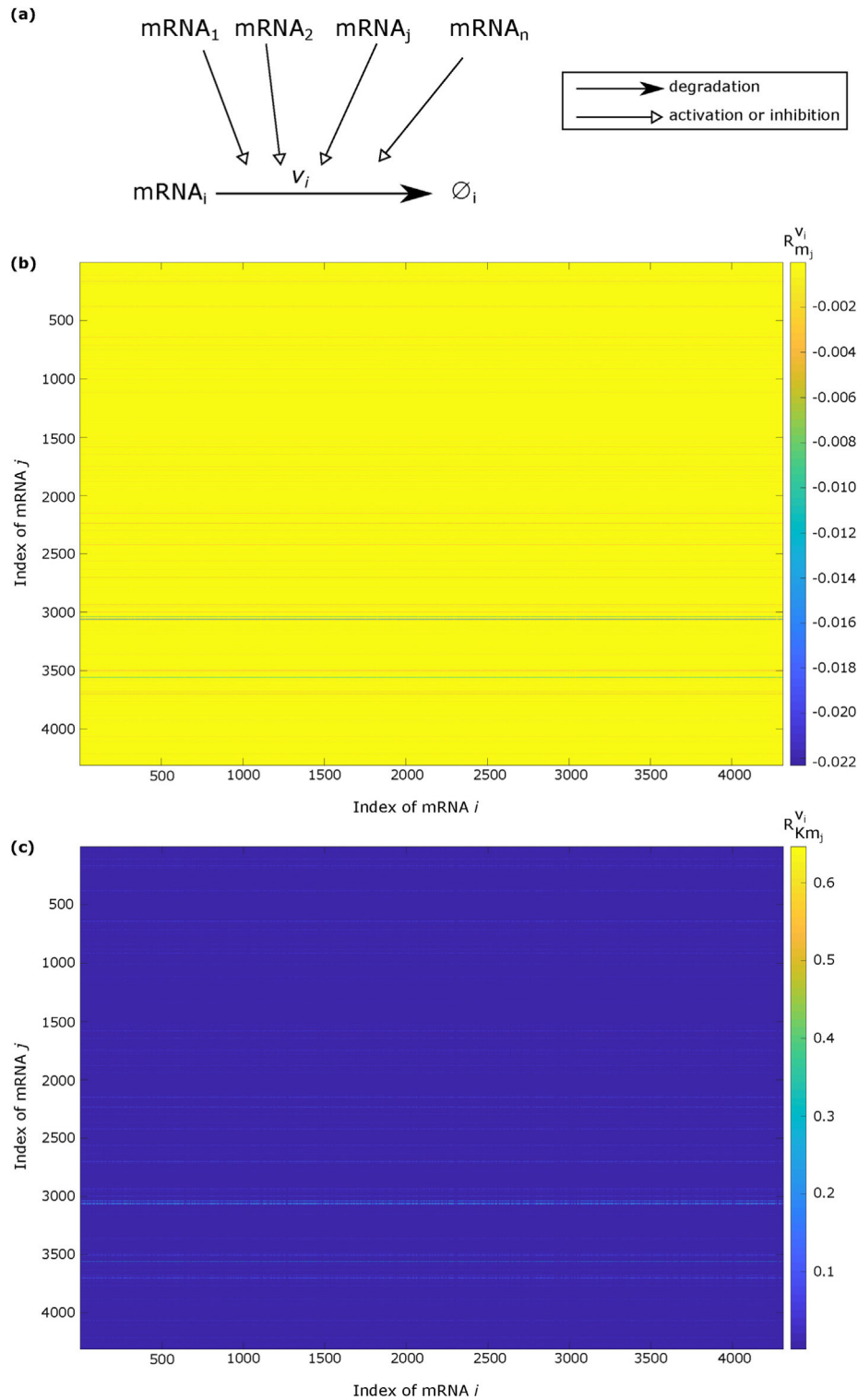


Fig. F.1. Scaled rate sensitivity coefficients for the initial velocity of mRNA i degradation. (a) Schematic drawing of the coupling between the degradation of mRNA i and the other cell mRNAs j . We determined, for each of the 4312 *E. coli* mRNAs i , the effect of a 10% change in (b) the initial concentration and (c) the K_m value of mRNA j on the initial degradation rate of each of the 4312 mRNAs i . Colour bars indicate the value of the response coefficients.

E. coli cells growing at four different rates. We chose one condition (0.4 h^{-1}) and spline fitted the logarithm of the 4254 degradation profiles using the `smooth.spline` function of the R package `stat`. Further analysis was possible only for 3868 mRNAs, for which the spline fit was satisfactory and information on the gene length and promoter localization was available. For each of these 3868

profiles, we determined the delay before degradation by simply finding the time point at which the sign of the first derivative of the spline function changes, from positive or null to negative due to the decrease of mRNA concentration. We observed the same phenomena as [Chen et al. \(2015\)](#) in their study: the size of mRNAs and their position towards the end of an operon increases the dura-

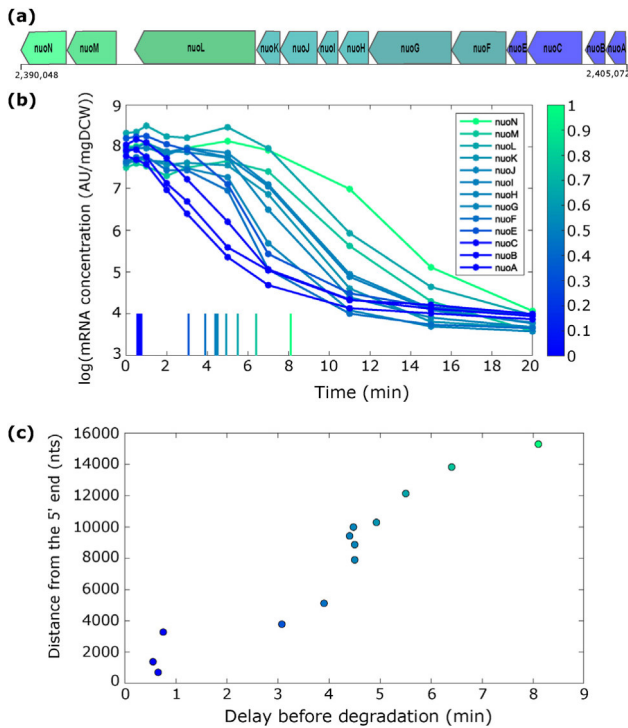


Fig. G.1. The distance from the 5' end of the operon *nuo* scales proportionally to the duration of the delay before start of RNA degradation. (a) Organization of the *nuo* operon. Numbers indicate genome position in base. (b) RNA concentration over time of exponential phase MG1655 cells growing at 0.4 h^{-1} on a semilogarithmic scale. Vertical lines indicate the duration of the delay. (c) Delay before degradation increases with the distance of the gene transcript to the 5' end of the operon.

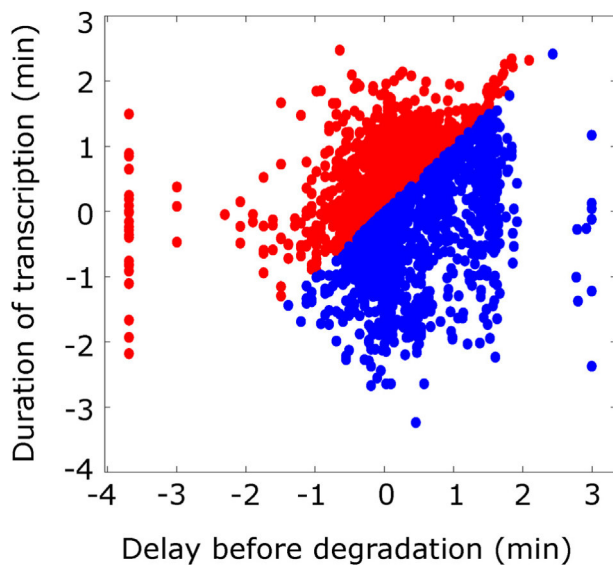


Fig. G.2. Role of residual transcription in the retardation of degradation. On log scales and for each mRNA with a non negligible delay, the duration of transcription is plotted versus the delay before degradation. mRNAs with a delay higher than the time needed for transcription are displayed in blue and red, otherwise.

tion of the delay. This is illustrated in Fig. G.1 in the case of the operon *nuo*. Genes at the beginning of the operon *nuo* have shorter delays before start of degradation than genes near the 3' end (Panel (b)). The same trend is observed when plotting the distance from the 5' end of the operon to the 3' end of each gene (Panel (c)). However, this is not a general rule. For instance, we noticed that many monocistronic mRNAs of low to medium size have also a delay.

We then assessed the contribution of residual transcription to the total duration of the delay before degradation. In principle, this would require to know the position of RNA polymerase when rifampicin inhibits the holoenzyme. Such data is not available, which makes it impossible to determine the residual transcription time precisely. However, we can set an upper limit to this time, by determining the time needed to transcribe entire mRNAs. If this transcription time is smaller than the delay before degradation, this means that residual transcription cannot be the sole determinant of the delay. Results are more difficult to interpret when the maximal transcription time is larger than the delay before degradation. They can reflect co-transcriptional degradation as in (Chen et al., 2015) for instance, but also include false negative: mRNAs with low residual transcription time, but total gene transcription time larger than the delay, should be counted as mRNAs for which residual transcription is not the sole determinant of the delay. To determine the time needed for transcription, we multiplied the elongation rate of transcription determined at the genome scale by Chen and co-authors (25 nts/s on average), with the length of each mRNA. When a gene is located within an operon, we used the promoter operon to determine a maximal size for that mRNA. We focused on the 2454 out of 3868 mRNAs with a delay before the onset of degradation (the remaining 1414 are immediately degraded). As shown in Fig. G.2, 1255 out of 2454 mRNAs have a delay before degradation larger than the maximal theoretical time needed for transcription (points in blue). 512 of them have a delay before degradation twice as large as the transcription time. Overall these results show that, for a vast majority of genes, transcription alone does not explain the whole duration of the delay with the transcription rate considered. This does not prove the role of competition experimentally, but does not exclude that this phenomenon could play a role in the retardation of degradation. Note we cannot rule out that heterogeneity in the polymerisation rate in transcription or in degradation may affect the estimated transcription times and delays before degradation. However, our analysis is based on population data for which information on heterogeneity is not available.

References

- Albe, K.R., Butler, M.H., Wright, B.E., 1990. Cellular concentrations of enzymes and their substrates. *J. Theor. Biol.* 143 (2), 163–195.
- Bandyra, K.J., Wandzik, J.M., Luisi, B.F., 2018. Substrate recognition and autoinhibition in the central ribonuclease RNase E. *Mol. Cell* 72 (2), 275–285.
- Bennett, B.D., Kimball, E.H., Gao, M., Osterhout, R., Van Dien, S.J., Rabinowitz, J.D., 2009. Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat. Chem. Biol.* 5 (8), 593–599.
- Bernstein, J.A., Khodursky, A.B., Lin, P.-H., Lin-Chao, S., Cohen, S.N., 2002. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl. Acad. Sci. U.S.A.* 99 (15), 9697–9702.
- Bobrovskyy, M., Vanderpool, C.K., Richards, G.R., 2015. Small RNAs regulate primary and secondary metabolism in Gram-negative bacteria. In: *Metabolism and Bacterial Pathogenesis*. American Society of Microbiology, pp. 59–94.
- Bernstein, J.A., Lin, P.H., Cohen, S.N., Lin-Chao, S., 2004. Global analysis of *Escherichia coli* RNA degradosome function using DNA microarrays. *Proc. Natl. Acad. Sci. U.S.A.* 101 (9), 2758–2763.
- Borghans, J.A.M., de Boer, R.J., Segel, L.A., 1996. Extending the quasi-steady state approximation by changing variables. *Bull. Math. Biol.* 58 (1), 43–63.
- Briggs, G.E., Haldane, J.B.S., 1925. A note on the kinetics of enzyme action. *Biochem. J.* 19 (2), 338.
- Buchler, N.E., Cross, F.R., 2009. Protein sequestration generates a flexible ultrasensitive response in a genetic network. *Mol. Syst. Biol.* 5 (1).
- Callaghan, A.J., Marcaida, M.J., Stead, J.A., McDowall, K.J., Scott, W.G., Luisi, B.F., 2005. Structure of *Escherichia coli* RNase E catalytic domain and implications for RNA turnover. *Nature* 437 (7062), 1187.
- Carpousis, A.J., 2007. The RNA degradosome of *Escherichia coli*: an mRNA-degrading machine assembled on RNase E. *Annu. Rev. Microbiol.* 61, 71–87.
- Carrier, T.A., Keasling, J., 1997. Mechanistic modeling of prokaryotic mRNA decay. *J. Theor. Biol.* 189 (2), 195–209.
- Celesnik, H., Deana, A., Belasco, J.G., 2007. Initiation of RNA decay in *Escherichia coli* by 5' pyrophosphate removal. *Mol. Cell.* 27 (1), 79–90.
- Chao, Y., Li, L., Girodat, D., Förstner, K.U., Said, N., Corcoran, C., Šmiga, M., Papenfort, K., Reinhardt, R., Wieden, H.-J., et al., 2017. In vivo cleavage map illuminates the

- central role of RNase E in coding and non-coding RNA pathways. *Mol. Cell.* 65 (1), 39–51.
- Chen, H., Shiroguchi, K., Ge, H., Xie, X.S., 2015. Genome-wide study of mRNA degradation and transcript elongation in *Escherichia coli*. *Mol. Syst. Biol.* 11 (1), 781.
- Choi, B., Rempala, G.A., Kim, J.K., 2017. Beyond the Michaelis-Menten equation: Accurate and efficient estimation of enzyme kinetic parameters. *Sci. Rep.* 7 (1), 17018.
- Clarke, J.E., Kime, L., Romero, A., McDowall, D., 2014. Direct entry by RNase E is a major pathway for the degradation and processing of RNA in *Escherichia coli*. *Nucleic Acids Res.* 42 (18), 11733–11751.
- Cookson, N.A., Mather, W.H., Danino, T., Mondragón-Palmino, O., Williams, R.J., Tsimring, L.S., Hasty, J., 2011. Queueing up for enzymatic processing: correlated signaling through coupled degradation. *Mol. Syst. Biol.* 7 (1).
- Cornish-Bowden, A., 2013. The origins of enzyme kinetics. *FEBS Lett.* 587 (17), 2725–2730.
- Dar, D., Sorek, R., 2018. Extensive reshaping of bacterial operons by programmed mRNA decay. *PLoS Genet.* 14, (4) e1007354.
- De Vos, D., Bruggeman, F.J., Westerhoff, H.V., Bakker, B.M., 2011. How molecular competition influences fluxes in gene expression networks. *PLoS ONE* 6 (12), e28494.
- Deana, A., Belasco, J., 2005. Lost in translation: the influence of ribosomes on bacterial mRNA decay. *Genes Dev.* 19, 2526–2533.
- Deana, A., Celesnik, H., Belasco, J.G., 2008. The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal. *Nature* 451 (7176), 355.
- 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 Genet.* 11, (10) e1005613.
- Deneke, C., Lipowsky, R., Valleriani, A., 2013. Effect of ribosome shielding on mRNA stability. *Phys. Biol.* 10, (4) 046008.
- Dori-Bachash, M., Shalem, O., Manor, Y., Pilpel, Y., Tirosh, I., 2012. Widespread promoter-mediated coordination of transcription and mRNA degradation. *Genome Biol.* 13 (12), R114.
- Dori-Bachash, M., Shema, E., Tirosh, I., 2011. Coupled evolution of transcription and mRNA degradation. *PLoS Biol.* 9, (7) e1001106.
- Dressaire, C., Pobre, V., Laguerre, S., Girbal, L., Arraiano, C.M., Coccagn-Bousquet, M., 2018. PNPase is involved in the coordination of mRNA degradation and expression in stationary phase cells of *Escherichia coli*. *BMC Genomics* 19 (1), 848.
- Esquerré, T., Laguerre, S., Turlan, C., Carpousis, A.J., Girbal, L., Coccagn-Bousquet, M., 2014. Dual role of transcription and transcript stability in the regulation of gene expression in *Escherichia coli* cells cultured on glucose at different growth rates. *Nucleic Acids Res.* 42 (4), 2460–2472.
- Esquerré, T., Moisan, A., Chiapello, H., Arike, L., Vilu, R., Gaspin, C., Coccagn-Bousquet, M., Girbal, L., 2015. Genome-wide investigation of mRNA lifetime determinants in *Escherichia coli* cells cultured at different growth rates. *BMC Genom.* 16 (1), 275.
- Garrey, S.M., Mackie, G.A., 2011. Roles of the 5'-phosphate sensor domain in RNase E. *Mol. Microbiol.* 80 (6), 1613–1624.
- Hartenian, E., Glaunsinger, B.A., 2019. Feedback to the central dogma: cytoplasmic mRNA decay and transcription are interdependent processes. *Crit. Rev. Biochem. Mol. Biol.* 54 (4), 385–398.
- Heinrich, R., Schuster, S., 1996. *The Regulation of Cellular Systems*. Chapman and Hall, New-York.
- Huang, C.-Y., Ferrell, J.E., 1996. Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proc. Natl. Acad. Sci. U.S.A.* 93 (19), 10078–10083.
- Huang, H., Liao, J., Stanley, N.C., 1998. Poly (A)- and poly (U)-specific RNA 3' tail shortening by *E. coli* ribonuclease E. *Nature* 391 (6662), 99.
- Hundt, S., Zaigler, A., Lange, C., Soppa, J., Klug, G., 2007. Global analysis of mRNA decay in *Halobacterium salinarum* NRC-1 at single-gene resolution using DNA microarrays. *J. Bacteriol.* 189 (19), 6936–6944.
- Ingalls, B.P., Sauro, H.M., 2003. Sensitivity analysis of stoichiometric networks: an extension of metabolic control analysis to non-steady state trajectories. *J. Theor. Biol.* 222 (1), 23–36.
- Jain, C., Belasco, J., 1995. Autoregulation of RNase E synthesis in *Escherichia coli*. *Nucleic Acids Symp. Series*. No. 33, 85–88.
- Jain, C., Belasco, J.G., 1995. RNase E autoregulates its synthesis by controlling the degradation rate of its own mRNA in *Escherichia coli*: unusual sensitivity of the *me* transcript to RNase E activity. *Genes Dev.* 9 (1), 84–96.
- Jeske, L., Placzek, S., Schomburg, I., Chang, A., Schomburg, D., 2018. BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Res.* 47 (D1), D542–D549.
- Jiang, X., Belasco, J.G., 2004. Catalytic activation of multimeric RNase E and RNase G by 5'-monophosphorylated RNA. *Proc. Natl. Acad. Sci. U.S.A.* 101 (25), 9211–9216.
- Keseler, I.M., Mackie, A., Santos-Zavaleta, A., Billington, R., Bonavides-Martínez, C., Caspi, R., Fulcher, C., Gama-Castro, S., Kothari, A., Krummenacker, M., et al., 2017. The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Res.* 45 (D1), D543–D550.
- Kiel, C., Serrano, L., 2012. Challenges ahead in signal transduction: MAPK as an example. *Curr. Opin. Biotechnol.* 23 (3), 305–314.
- Kim, K.-S., Sim, S., Ko, J.-H., Cho, B., Lee, Y., 2004. Kinetic analysis of precursor M1 RNA molecules for exploring substrate specificity of the N-terminal catalytic half of RNase E. *J. Biochem.* 136 (5), 693–699.
- Kim, Y., Andreu, M.J., Lim, B., Chung, K., Terayama, M., Jiménez, G., Berg, C.A., Lu, H., Shvartsman, S.Y., 2011. Gene regulation by MAPK substrate competition. *Dev. Cell* 20 (6), 880–887.
- Kim, Y., Coppey, M., Grossman, R., Ajuria, L., Jiménez, G., Paroush, Z., Shvartsman, S. Y., 2010. MAPK substrate competition integrates patterning signals in the *Drosophila* embryo. *Curr. Biol.* 20 (5), 446–451.
- Kime, L., Jourdan, S.S., McDowall, K.J., 2008. Identifying and characterizing substrates of the RNase E/G family of enzymes. *Methods Enzymol.* 447, 215–241.
- Kime, L., Jourdan, S.S., Stead, J.A., Hidalgo-Sastre, A., McDowall, K.J., 2010. Rapid cleavage of RNA by RNase E in the absence of 5' monophosphate stimulation. *Mol. Microbiol.* 76 (3), 590–604.
- Kushner, S.R., 2002. mRNA decay in *Escherichia coli* comes of age. *J. Bacteriol.* 184 (17), 4658–4665.
- Laguerre, S., González, I., Nouaille, S., Moisan, A., Villa-Vialaneix, N., Gaspin, C., Bouvier, M., Carpousis, A.J., Coccagn-Bousquet, M., Girbal, L., 2018. Large-scale measurement of mRNA degradation in *Escherichia coli*: To delay or not to delay. *Methods Enzymol.* 612, 47–66.
- Lee, K., Zhan, X., Gao, J., Qiu, J., Feng, Y., Meganathan, R., Cohen, S.N., Georgiou, G., 2003. RraA: a protein inhibitor of RNase E activity that globally modulates RNA abundance in *E. coli*. *Cell* 114 (5), 623–634.
- Legewie, S., Schoeberl, B., Blüthgen, N., Herzog, H., 2007. Competing docking interactions can bring about bistability in the mapk cascade. *Biophys. J.* 93 (7), 2279–2288.
- Luciano, D.J., Vasilyev, N., Richards, J., Serganov, A., Belasco, J.G., 2017. A novel RNA phosphorylation state enables 5' end-dependent degradation in *Escherichia coli*. *Mol. Cell* 67 (1), 44–54.
- Lugowski, A., Nicholson, B., Rissland, O.S., 2018. Determining mRNA half-lives on a transcriptome-wide scale. *Methods* 137, 90–98.
- Lykke-Andersen, S., Brodersen, D.E., Jensen, T.H., 2009. Origins and activities of the eukaryotic exosome. *J. Cell Sci.* 122 (10), 1487–1494.
- Mackie, G.A., 1998. Ribonuclease E is a 5'-end-dependent endonuclease. *Nature* 395 (6703), 720.
- Mackie, G.A., 2013. RNase E: at the interface of bacterial RNA processing and decay. *Nat. Rev. Microbiol.* 11 (1), 45.
- Marcaida, M.J., DePristo, M.A., Chandran, V., Carpousis, A.J., Luisi, B.F., 2006. The RNA degradosome: life in the fast lane of adaptive molecular evolution. *Trends Biochem. Sci.* 31 (7), 359–365.
- McDowall, K.J., Kabardin, V.R., Wu, S.-W., Cohen, S.N., Lin-Chao, S., 1995. Site-specific RNase E cleavage of oligonucleotides and inhibition by stem-loops. *Nature* 374 (6519), 287.
- McDowall, K.J., Lin-Chao, S., Cohen, S.N., 1994. A+U content rather than a particular nucleotide order determines the specificity of RNase E cleavage. *J. Biol. Chem.* 269 (14), 10790–10796.
- Mehra, A., Hatzimanikatis, V., 2006. An algorithmic framework for genome-wide modeling and analysis of translation networks. *Biophys. J.* 90 (4), 1136–1146.
- Moffitt, J.R., Pandey, S., Boettiger, A.N., Wang, S., Zhuang, X., 2016. Spatial organization shapes the turnover of a bacterial transcriptome. *Elife* 5, e13065.
- Mohanty, B.K., Kushner, S.R., 2016. Regulation of mRNA decay in bacteria. *Ann. Rev. Microbiol.* 70, 25–44.
- Morozova, N., Zinovyev, A., Nonne, N., Pritchard, L.-L., Gorban, A.N., Harel-Bellan, A., 2012. Kinetic signatures of microRNA modes of action. *RNA* 18 (9), 1635–1655.
- Nouaille, S., Mondeil, S., Finoux, A.-L., Moulis, C., Girbal, L., Coccagn-Bousquet, M., 2017. The stability of an mRNA is influenced by its concentration: a potential physical mechanism to regulate gene expression. *Nucleic Acids Res.* 45 (20), 11711–11724.
- Pedersen, M.G., Bersanib, A.M., Bersanic, E., 2007. The total quasi-steady-state approximation for fully competitive enzyme reactions. *Bull. Math. Biol.* 69 (1), 433.
- Potts, A.H., Vakulskas, C.A., Pannuri, A., Yakhnin, H., Babitzke, P., Romeo, T., 2017. Global role of the bacterial post-transcriptional regulator csra revealed by integrated transcriptomics. *Nature Comm.* 8 (1), 1596.
- Richards, J., Belasco, J.G., 2019. Obstacles to scanning by RNase E govern bacterial mRNA lifetimes by hindering access to distal cleavage sites. *Mol. Cell* 74 (2), 284–295.
- Richards, J., Luciano, D.J., Belasco, J.G., 2012. Influence of translation on RppH-dependent mRNA degradation in *Escherichia coli*. *Mol. Microbiol.* 86 (5), 1063–1072.
- Santiago-Frangos, A., Woodson, S.A., 2018. Hfq chaperone brings speed dating to bacterial sRNA. *RNA* 9, (4) e1475.
- Schneider, R., Travers, A., Kutateladze, T., Muskhelishvili, G., 1999. A DNA architectural protein couples cellular physiology and DNA topology in *Escherichia coli*. *Mol. Microbiol.* 34 (5), 953–964.
- Schnell, S., Maini, P., 2000. Enzyme kinetics at high enzyme concentration. *Bull. Math. Biol.* 62 (3), 483–499.
- Segel, L.A., Slemrod, M., 1989. The quasi-steady-state assumption: a case study in perturbation. *SIAM Rev.* 31 (3), 446–477.
- Shalem, O., Groisman, B., Choder, M., Dahan, O., Pilpel, Y., 2011. Transcriptome kinetics is governed by a genome-wide coupling of mRNA production and degradation: a role for RNA Pol II. *PLoS Genet.* 7 (9).
- Sols, A., Marco, R., 1970. Concentrations of metabolites and binding sites. implications in metabolic regulation. In: *Curr Topics Cellular Reg.* vol. 2. Elsevier, pp. 227–273.
- Sousa, S., Marchand, I., Dreyfus, M., 2001. Autoregulation allows *Escherichia coli* RNase E to adjust continuously its synthesis to that of its substrates. *Mol. Microbiol.* 42 (3), 867–878.
- Stead, M.B., Marshall, S., Mohanty, B.K., Mitra, J., Castillo, L.P., Ray, D., Van Bakel, H., Hughes, T.R., Kushner, S.R., 2010. Analysis of *Escherichia coli* RNase E and RNase III activity in vivo using tiling microarrays. *Nucleic Acids Res.* 39 (8), 3188–3203.

- Strahl, H., Turlan, C., Khalid, S., Bond, P.J., Kebalo, J.-M., Peyron, P., Poljak, L., Bouvier, M., Hamoen, L., Luisi, B.F., et al., 2015. Membrane recognition and dynamics of the RNA degradosome. *PLoS Genet.* 11, (2) e1004961.
- Sun, M., Schwab, B., Schulz, D., Pirkl, N., Etzold, S., Larivière, L., Maier, K.C., Seizl, M., Tresch, A., Cramer, P., 2012. Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. *Genome Res.* 22 (7), 1350–1359.
- Suzuki, K., Babitzke, P., Kushner, S.R., Romeo, T., 2006. Identification of a novel regulatory protein (CsrD) that targets the global regulatory RNAs CsrB and CsrC for degradation by RNase E. *Genes Dev.* 20 (18), 2605–2617.
- Tadmor, A., Tlustý, T., 2008. A coarse-grained biophysical model of *E. coli* and its application to perturbation of the rRNA operon copy number. *PLoS Comput. Biol.* 4. e1000038–e1000038.
- Tang, J., Riley, W., 2013. A total quasi-steady-state formulation of substrate uptake kinetics in complex networks and an example application to microbial litter decomposition. *Biogeosciences* 10 (12), 8329–8351.
- Tang, J., Riley, W., 2017. SUPECA kinetics for scaling redox reactions in networks of mixed substrates and consumers and an example application to aerobic soil respiration. *Geosci. Model Dev.* 10, 3277–3295.
- Thomas, P., Terradot, G., Danos, V., Weiße, A.Y., 2018. Sources, propagation and consequences of stochasticity in cellular growth. *Nat. Commun.* 9 (1), 4528.
- Tzafiriri, A., 2003. Michaelis-Menten kinetics at high enzyme concentrations. *Bull. Math Biol.* 65 (6), 1111–1129.
- Tzafiriri, A., Bercovier, M., Parnas, H., 2002. Reaction diffusion model of the enzymatic erosion of insoluble fibrillar matrices. *Biophys. J.* 83 (2), 776–793.
- Tzafiriri, A., Edelman, E., 2004. The total quasi-steady-state approximation is valid for reversible enzyme kinetics. *J. Theor. Biol.* 226 (3), 303–313.
- Valgepea, K., Adamberg, K., Seiman, A., Vilu, R., 2013. *Escherichia coli* achieves faster growth by increasing catalytic and translation rates of proteins. *Mol. Biosyst.* 9 (9), 2344–2358.
- Wang, Y., Liu, C.L., Storey, J.D., Tibshirani, R.J., Herschlag, D., Brown, P.O., 2002. Precision and functional specificity in mRNA decay. *Proc. Natl. Acad. Sci. U.S.A.* 99 (9), 5860–5865.
- Weiße, A.Y., Oyarzún, D.A., Danos, V., Swain, P.S., 2015. Mechanistic links between cellular trade-offs, gene expression, and growth. *Proc. Natl. Acad. Sci. U.S.A.* 112 (9), E1038–E1047.
- Wright, P.R., Georg, J., Mann, M., Sorescu, D.A., Richter, A.S., Lott, S., Kleinkauf, R., Hess, W.R., Backofen, R., 2014. CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains. *Nucleic Acids Res.* 42 (W1), W119–W123.
- Yandek, L.E., Lin, H.-C., Harris, M.E., 2013. Alternative substrate kinetics of *Escherichia coli* ribonuclease P determination of relative rate constants by internal competition. *J. Biol. Chem.* 288 (12), 8342–8354.

3.4 Discussion sur l'origine du délai en début de cinétique de dégradation

Notre travaux sur la modélisation de la dégradation exposés dans la section 3.3 apportent un regard différent sur l'origine du délai communément observé avant le démarrage de la dégradation. Dans la vision "classique" de la littérature, ce délai résulte du couplage de la transcription à la dégradation dans les analyses de stabilité de l'ARNm. Sa prise en compte reproduit la transcription résiduelle observée par l'ARN polymérase après l'arrêt de la transcription [Chen et al., 2015, Moffitt et al., 2016, Dar and Sorek, 2018] et les niveaux d'ARNm résiduels à la fin du processus de dégradation. Pour reprendre les notations de la section 1.5, le taux de transcription v_s est non nul durant une courte période et compense le taux de dégradation v_d jusqu'à ce que l'ARN polymérase termine l'élongation de l'ARNm. La vitesse de transcription devient alors nulle et les concentrations d'ARNm décroissent de manière exponentielle. De cette façon le délai est la conséquence de la transcription résiduelle induit par l'utilisation de rifampicine, un inhibiteur de l'initialisation de la transcription par l'ARN polymérase mais pas de son élongation. Par conséquent cette caractéristique est alors considérée par Chen et al. [2015], Moffitt et al. [2016], Dar and Sorek [2018] comme une conséquence du design expérimental, c'est-à-dire une variation extrinsèque au processus de dégradation en lui même.

Dans nos travaux, le délai est une conséquence du phénomène de compétition entre les ARNm pour se fixer à la RNase E. Par conséquent la phase transitoire observée avant la phase de dégradation est considérée comme une variation intrinsèque au processus de dégradation des ARNm. Notons que ces deux interprétations sont fortement connectées. En effet dans l'interprétation extrinsèque, le délai est directement corrélé à la distance entre l'extrémité 5' et la position de l'ARN polymérase au moment de l'ajout de rifampicine [Chen et al., 2015, Moffitt et al., 2016, Dar and Sorek, 2018]. Cette même position a directement à voir avec la vitesse de transcription et la force du promoteur. Plus le promoteur est fort, plus il y aura de RNAP en cours de transcription et par conséquent plus la concentration de l'ARNm en question sera élevée. Dans la deuxième interprétation nous montrons qu'un ARNm ayant une concentration initiale élevée (avant ajout de rifampicine) est susceptible de se dégrader plus rapidement tout en étant plus sensible à la titration de l'enzyme par les autres ARNm d'autant plus que la vitesse moyenne de dégradation des ARNms est relativement rapide. Pour ces ARNm en forte concentration, le délai a donc une forte chance de résulter de la combinaison d'un effet transcriptionnel et de l'effet de compétition dans la dégradation (cf. annexe de l'article 3.3).

Chapitre 4

Modèle statistique pour l'estimation des paramètres cinétiques

L'estimation des paramètres d'un modèle mécanistique déterministe est un défi, car cela suppose de contrôler le bruit des données et celui lié aux simulations numériques. Ce défi se rencontre dans différents domaines où les modèles mathématiques jouent un rôle prépondérant dans la compréhension du problème étudié (section 3.1.1).

Plusieurs questions cruciales doivent se poser. Il faut tout d'abord se demander si les données nécessaires existent. Ensuite, lorsque les données sont effectivement disponibles, se pose la question de leur qualité : est-elle suffisante pour mener à bien l'estimation des paramètres ? En effet, l'un des problèmes classiques quand on cherche à estimer des paramètres cinétiques est celui de l'identification théorique et pratique. La question se pose ensuite de savoir si on souhaite estimer des paramètres cinétiques moyens sur un jeu de données, ou, au contraire, obtenir des estimations pour chaque individu. Ce modèle statistique ne devra pas simplement avoir pour but l'estimation des paramètres, il devra également chercher à ajouter des connaissances supplémentaires. Pour répondre à ces questions, il nous faut définir alors un modèle statistique flexible qui pourra être utilisé avec différents types de données existantes, mais dans lequel on pourra également intégrer des informations supplémentaires, permettant le raffinement ou l'extension du modèle primaire. Notons que dans ce chapitre nous utiliserons volontairement un vocabulaire statistique (population, individus, variations globales/variations spécifiques).

Dans une première section, nous nous intéresserons aux modèles non linéaires à effets mixtes qui semblent être prometteurs pour l'estimation des paramètres cinétiques d'un modèle mécanistique et qui remplissent une majorité des critères définis ci-dessus. Nous aborderons en premier lieu la formulation mathématique générale de ce type de modèles et les outils computationnels disponibles. Enfin dans une deuxième section, nous appliquerons le modèle à effets mixtes non linéaire à l'estimation des paramètres cinétiques du modèle mécanistique développé dans la section 3.3.

4.1 Les modèles non linéaires à effets mixtes : définition générale

Qu'est ce qu'un modèle à effets mixtes non linéaire ? Un modèle à effets mixtes est un modèle statistique qui incorpore à la fois des effets fixes et des effets aléatoires à travers la modélisation des paramètres individuels. Les effets fixes sont décrits par les paramètres de population supposés être communs à tous les individus. Les effets aléatoires sont des variables aléatoires associées à chaque individu d'une population. Les modèles à effets mixtes sont souvent utilisés pour faire l'inférence sur les caractéristiques sous-jacentes aux profils de mesures individuelles d'une population d'intérêt. Comme pour tous les modèles de régression, leur objectif est de décrire une variable de réponse en fonction des variables prédictives (indépendantes). Cependant, les modèles à effets mixtes reconnaissent les corrélations au sein d'une population, offrant un compromis raisonnable entre ignorer complètement les groupes de données, perdant ainsi des informations précieuses, et ajuster chaque individu séparément, ce qui nécessite beaucoup plus de points de données et un biais dans la quantification a posteriori des valeurs moyennes d'une population. Les modèles non linéaires à effets mixtes sont appliqués dans de nombreux domaines, notamment en médecine, santé publique, pharmacologie et écologie. Il y a peu d'exemples en ce qui concerne l'expression génique [Karlsson et al., 2015, Fröhlich et al., 2018].

4.1.1 Modèle d'observation

Soit y_{ij} les observations pour l'individu i au temps de mesure j , le modèle d'observation d'un modèle non linéaire à effets mixtes s'écrit de la façon suivante :

$$y_{ij} = f(t_{ij}, \psi_i) + g(t_{ij}, \psi_i) \times \epsilon_{ij}, \quad 1 \leq i \leq N, \quad 1 \leq j \leq n_i. \quad (4.1)$$

N représente le nombre d'individus dans la population et n_i est le nombre de points de mesure. f est appelé modèle structurel. Cela correspond au type de courbe de base que nous soupçonnons être suivie par les données, par exemple, linéaire, logarithmique, exponentielle, etc. Parfois, un modèle des processus biologiques associés conduit à des équations différentielles et par conséquent $f(t_{ij}, \psi_i)$ représente la solution analytique ou numérique. $\psi = (\psi_1, \psi_2, \dots, \psi_d)$ est le vecteur de l'ensemble des paramètres $1 \cdot d$ influençant la fonction f . $(\epsilon_i1, \epsilon_i2, \dots, \epsilon_in)$ sont les erreurs résiduelles normalisées au point de mesure j pour l'individu i . Nous supposons que ceux-ci proviennent d'une distribution de probabilité centrée et de variance unitaire ($\mathbb{E}(\epsilon_j) = 0$ et $\text{Var}(\epsilon_j) = 1$). Dans ce cas $f(t_{ij}, \psi_i)$ et $g(t_{ij}, \psi_i)$ sont l'espérance et la déviation standard des y_{ij} (cf équations 4.3).

$$\mathbb{E}(y_{ij}|\psi_i) = f(t_{ij}, \psi_i) \quad (4.2)$$

$$Sd(y_{ij}|\psi_i) = g(t_{ij}, \psi_i) \quad (4.3)$$

L'hypothèse que la distribution de toute observation y_{ij} est symétrique autour de sa valeur prédite est très forte. Si cette hypothèse ne tient pas, nous pouvons décider de transformer les données pour les rendre plus symétriques autour de leur valeur prédite (transformée). Dans d'autres cas, les contraintes sur les valeurs que peuvent prendre les observations peuvent également nous conduire à vouloir transformer les données. Soit u la transformation appliquée

aux données la formulation générale d'un modèle à effets mixtes non linéaire (cf. équation 4.1) devient alors :

$$u(y_{ij}) = u(f(t_{ij}, \psi_i)) + g(t_{ij}, \psi_i) \times \epsilon_{ij} \quad (4.4)$$

Par exemple, si y_{ij} prend des valeurs non négatives, une transformation logarithmique peut être utilisée :

$$\log(y_{ij}) = \log(f(t_{ij}, \psi_i)) + g(t_{ij}, \psi_i)\epsilon_{ij}, \quad (4.5)$$

$$y_{ij} = f(t_{ij}, \psi_i) e^{-g(t_{ij}, \psi_i)\epsilon_{ij}}. \quad (4.6)$$

4.1.2 Modélisation des paramètres individuels

Comme évoqué précédemment, le modèle d'observation fait intervenir les paramètres individuels ψ_i . Comme nous voulons étudier l'aspect population, nous supposons que ψ_i provient d'une certaine distribution de probabilité p_{ψ_i} .

Généralement nous assumons que les individus sont indépendants. Si p_{ψ_i} est définie comme une distribution paramétrique qui dépend du vecteur des paramètres de la population (θ) et d'un ensemble de covariables $c_i = c_{i1}, c_{i2}, c_{i3}, \dots, c_{iL}$ avec L le nombre de covariables, les paramètres individuels se définissent de la manière suivante :

$$\psi_i \sim p_{\psi_i}(\cdot; c_i, \theta). \quad (4.7)$$

La distribution p_{ψ_i} joue un rôle fondamentale car elle décrit la variabilité inter-individuelle (entre les individus). En substance, cela signifie que modéliser les paramètres individuels revient à caractériser le vecteur de paramètres de la population (θ).

Au lieu d'écrire ψ_i en utilisant une distribution de probabilité, nous pouvons aussi utiliser l'équation suivante :

$$\psi_i = M(\beta, c_i, \eta_i) \quad (4.8)$$

où β est le vecteur des effets fixes (moyenne, médiane, mode) et η_i le vecteur des effets aléatoires. Cette description dans l'équation 4.8 est très utilisée pour les applications pratiques car elle requiert moins de connaissances en probabilité et est plus facilement interprétable. Elle est cependant moins générale que la précédente (cf. équation 4.7) car cette dernière est majoritairement utilisée lors de développements théoriques (mathématiques) du fait de sa généralité.

Pour illustrer notre propos nous allons aborder dans la suite de cette section l'exemple d'une définition des paramètres individuels par une distribution gaussienne et son extension par transformation logarithmique (Log-normal). Le modèle gaussien pour les paramètres individuels a plusieurs avantages. En particulier pour cet exemple les deux définitions (équation 4.7 et 4.8) sont parfaitement équivalentes en terme de représentation mathématique. Ceci s'explique par le fait qu'une distribution normale est complètement caractérisée par ses moments d'ordre 1 et 2. A l'inverse pour les distributions de Cauchy par exemple, les deux représentations mathématiques ne sont pas strictement équivalentes.

$$\psi_i \sim \mathcal{N}(\mu, \sigma^2) \quad (4.9)$$

La forme dans l'équation 4.9 nous donne une description explicite de la distribution du vecteur des paramètres individuels ψ qui nous permet de déduire la fonction de la distribution de probabilité (PDF). Comme précisé précédemment nous pouvons aussi caractériser les paramètres individuels en utilisant l'équation (4.8) :

$$\psi_i = \mu + \eta_i \quad \eta \sim \mathcal{N}(0, \sigma^2) \quad (4.10)$$

L'équation 4.10 sépare les paramètres ψ en deux parties. La première partie (μ) représente les effets fixes et la deuxième partie (η_i), les effets aléatoires. Notons que les effets fixes peuvent tout aussi bien être caractérisés par une autre statistique que la moyenne. Par exemple, on peut utiliser la médiane ou le mode (propriété de la loi normale).

Cependant, toutes les distributions ne sont pas gaussiennes. La loi normale a comme support \mathbb{R} , alors que certains paramètres prennent leurs valeurs dans un intervalle précis. Par exemple, certains paramètres ont seulement des valeurs positives (volume, paramètres cinétiques,...) et d'autres sont sur un intervalle borné (proportion, biodisponibilité). De plus la loi normale est symétrique (propriété qui n'est pas partagée par toutes les distributions). Une possibilité d'étendre l'utilisation de la loi normale est de considérer une transformation de celle-ci. Supposons qu'il existe une fonction monotone h telle que $h(\psi)$ est distribuée selon une loi normale, alors il existe un μ et un σ tels que :

$$h(\psi_i) = h(\mu) + \eta_i \quad \eta \sim \mathcal{N}(0, \sigma^2) \quad (4.11)$$

$$h(\psi_i) \sim \mathcal{N}(h(\mu), \sigma^2) \quad (4.12)$$

Notons qu'il est alors possible de redéfinir la fonction de distribution de probabilité (PDF) telle que :

$$p_\psi(\psi) = h'(\psi) \times p_{h(\psi)}(h(\psi)) \quad (4.13)$$

Une extension classique pour la définition de la distribution de probabilité des paramètres individuels positifs est d'utiliser une transformation logarithmique. Dans ce cas les paramètres individuels peuvent se redéfinir comme :

$$\log(\psi_i) = \log(\mu) + \eta_i \quad \eta \sim \mathcal{N}(0, \sigma^2) \quad (4.14)$$

$$\log(\psi_i) \sim \mathcal{N}(\log(\mu), \sigma^2) \quad (4.15)$$

Notons qu'il existe d'autres transformations telles que la transformation de Box-cox ou encore l'utilisation des distributions probit-normal et logit-normal.

4.1.3 Modélisation des paramètres individuels avec ajout d'une covariable

Dans certain cas des données supplémentaires sont disponibles pour permettre de mieux caractériser les variations inter-individuelles. Ces données qu'on appellera covariables peuvent être de plusieurs types. Dans cette section nous aborderons l'intégration de covariables continues (par exemple, la taille d'un individu) et catégorielles (le sexe) dans la modélisation des paramètres individuels car c'est le type de covariables utilisé durant cette thèse.

La forme la plus simple pour intégrer une covariable est de choisir un modèle linéaire où le paramètre individuel est modélisé comme une combinaison linéaire. Nous considérons une distribution normale du paramètre individuel pour simplifier le propos. Nous pouvons alors réécrire les équations (4.9) et (4.10) pour une covariable continue de la manière suivante :

$$\psi_i \sim \mathcal{N}(\mu + \beta \times c, \sigma^2). \quad (4.16)$$

Dans le cas d'une covariable catégorielle, la covariable peut se présenter par une fonction indicatrice ($\mathbf{1}$). Notons qu'il nous faut alors définir une catégorie de référence (cat_{ref}) pour laquelle la valeur de l'indicatrice est égale à 0 :

$$\psi_i \sim \mathcal{N}(\mu + \beta_{cat} \times \mathbf{1}_{cat_{ref}=0}, \sigma^2). \quad (4.17)$$

Une extension à l'ajout d'une covariable est de prendre en compte un niveau de variabilité supplémentaire [Lavielle, 2014, Panhard and Samson, 2009] :

$$\psi_i = \mu + \beta_{cat} \times \mathbf{1}_{cat_{ref}=0} + \eta_{i,cat_{ref}} + \eta_{i,cat} \quad (4.18)$$

avec $\eta_{i,cat_{ref}} \sim \mathcal{N}(0, \sigma^2)$ et $\eta_{i,cat} \sim \mathcal{N}(0, \omega^2)$. Cette extension sera développée plus en détail en section 4.2 où nous l'utiliserons pour décrire la variabilité importante de la concentration des ARNm aux différents taux de croissance et entre les différents transcrits.

4.1.4 Modèle d'erreur résiduelle

Il existe plusieurs modèles usuellement utilisés pour décrire l'erreur résiduelle des données (voir tableau 4.1). On appelle $g(t_{ij}, \psi_i) \geq 0$ la fonction caractérisant le modèle d'erreur résiduelle des données dépendant des temps de mesures et des paramètres individuels pour l'individu i .

Modèle d'erreur résiduelle	$g(t_{ij}, \psi_i) \times \epsilon_{ij}$
Additif	$a_i \times \epsilon_{ij}$
proportionnel	$b_i \times f(t_{ij}, \psi_i) \times \epsilon_{ij}$
Combiné 1	$(a_i + b_i \times f(t_{ij}, \psi_i)) \times \epsilon_{ij}$
Combiné 2	$\sqrt{a_i^2 + b_i^2 f^2(t_{ij}, \psi_i)} \times \epsilon_{ij}$

TABLE 4.1 – Tableau récapitulatif des différents modèles d'erreur résiduelle usuels

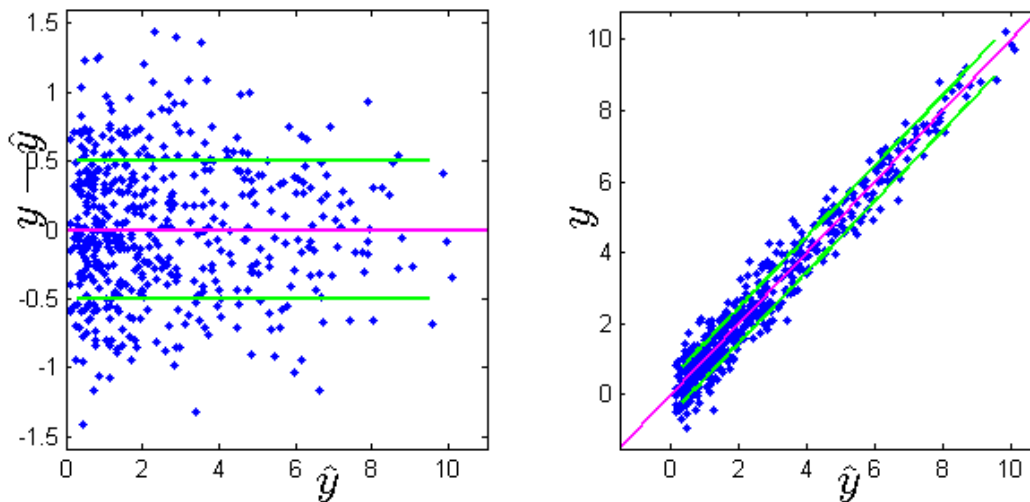


FIGURE 4.1 – Représentation graphique du modèle d'erreur constant. Sur la figure de droite on représente sur l'axe des ordonnées l'erreur résiduelle (donnée - prédiction). Sur la figure de gauche le graphique représente le modèle d'erreur en affichant (prédiction versus observation)

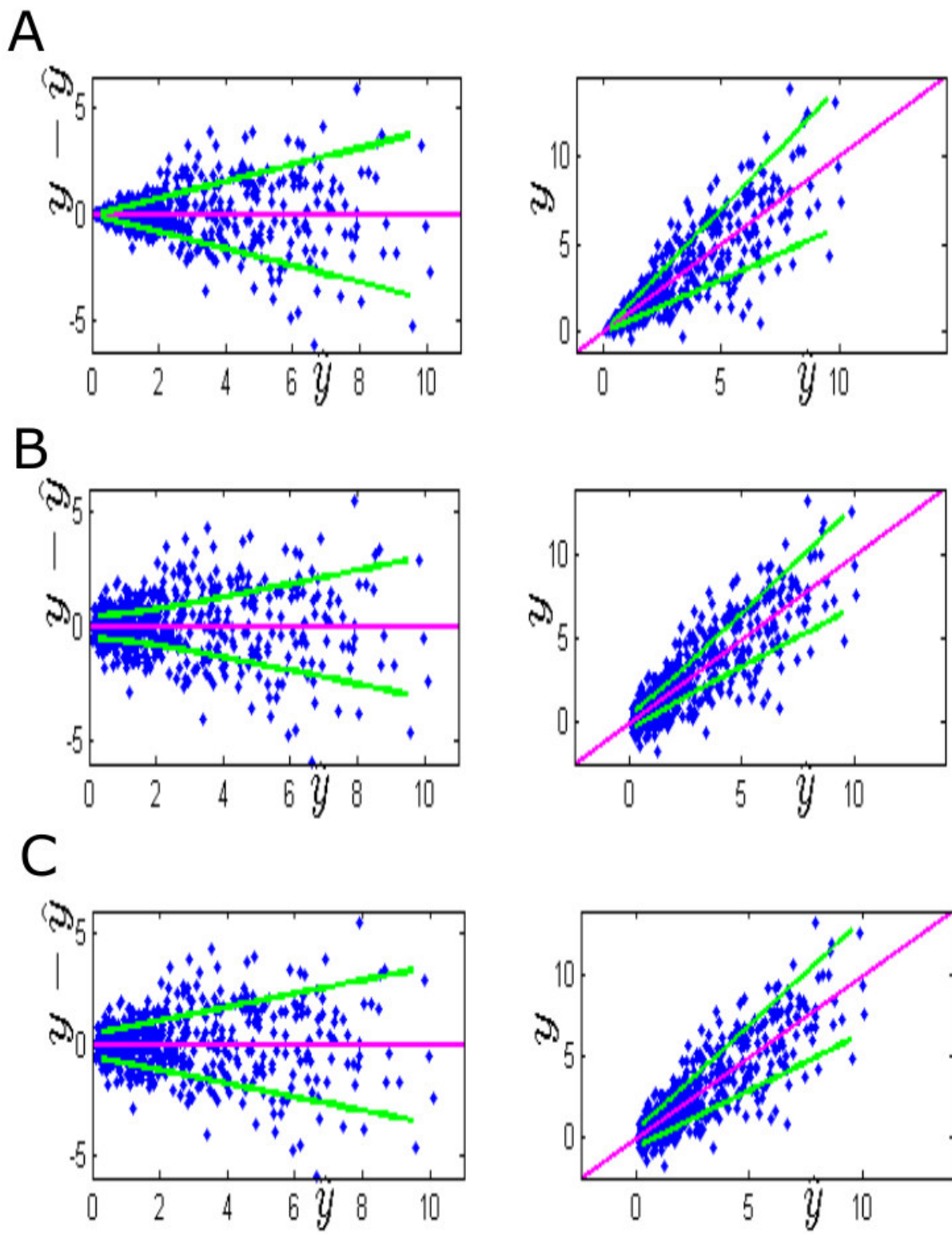


FIGURE 4.2 – Illustration graphique des différents modèles d’erreur résiduelle : **A.** proportionnel, **B.** combiné 1 et **C.** combiné 2 (C). 4.1

4.1.5 Implémentation des modèles non linéaires à effets mixtes

Dans la littérature il existe deux grand types de méthodes pour estimer les paramètres d'un modèle à effet mixte non linéaire. La première méthode consiste à linéariser la vraisemblance puis la maximiser par des méthodes de gradients telles que l'algorithme itératif de Newton Raphson. Ce principe a été implémenté dans le premier logiciel disponible, NONMEM, développé dans les années 70. Cette méthodologie et ses extensions sont également disponibles dans divers langages de programmation tels que SAS (proc NLMIXED, SAS Institute) et R (R Development Core Team (2006)), ou le package nlme [Pinheiro and Bates, 1995, Davidian and Giltinan, 1995] qui fait partie de l'installation de base.

Les méthodes basées sur la linéarisation présentent cependant des lacunes à la fois statistiques et pratiques. Premièrement elles ne convergent pas vers les estimations du maximum de vraisemblance. Alors que le biais est généralement mineur pour les effets fixes, les composantes de la variance peuvent être significativement biaisées, surtout avec une grande variabilité inter-individuelle [Comets and Mentré, 2001, Bertrand et al., 2008]. Deuxièmement, ces méthodes souffrent de biais sévères lorsqu'elles sont appliquées à des données non continues, comme le montrent Molenberghs and Verbeke [2006]. En pratique, les algorithmes basés sur la linéarisation présentent également des problèmes de convergence et peuvent être compliqués à utiliser avec des modèles complexes. Il a été démontré en revanche que les algorithmes stochastiques fournissent des estimations non biaisées [Savic et al., 2011].

Au cours de la dernière décennie, de nouveaux algorithmes d'estimation plus performants ont donc été proposés [Lavielle, 2014]. Ces algorithmes calculent la vraisemblance par approximations statistiques, pour préserver les propriétés statistiques des estimateurs du maximum de vraisemblance. Un exemple est l'approximation de Laplace, qui équivaut à une quadrature gaussienne et a été implémentée dans NONMEM. Dans R, la bibliothèque lme4 utilise cette approximation [Barr, 2013] avec une approche des moindres carrés pénalisés pour l'algorithme d'estimation.

Une alternative à la minimisation basée sur le gradient est l'algorithme EM, qui est également un algorithme itératif. Il a été développé dans le contexte des données manquantes [Dempster et al., 1977]. L'algorithme SAEM (Stochastic Approximation Expectation Maximization) combinant une approximation stochastique de la vraisemblance avec un algorithme EM s'est avéré très efficace pour les modèles à effets mixtes non linéaires. Il converge rapidement vers les estimateurs du maximum de vraisemblance [Delyon et al., 1999] et est plus performant que les algorithmes basés sur la linéarisation Girard and Mentré [2005]. Cette méthode été implémentée dans le logiciel Monolix [Lavielle and Mboogning, 2014], qui bénéficie d'une utilisation de plus en plus répandue. ce type de solution est aussi dans la toolbox "statistique" de MATLAB (nlmefitsa.m), et est également disponible en NONMEM version 7. L'équivalent en R est également disponible via le package saemix [Comets et al., 2011].

Suite à cette revue succincte des différents algorithmes d'estimation des paramètres individuels d'un modèle à effets mixtes non linéaire, il apparaît très clairement que l'estimation des paramètres d'un tel modèle doit s'effectuer via l'algorithme SAEM. En effet cette méthode a démontré son efficacité aux niveaux théorique et pratique.

4.2 Application à l'estimation des paramètres cinétiques du modèle de dégradation

Ayant décrit la modélisation non-linéaire à effets mixtes en détails, nous allons à présent l'appliquer à l'étude de données omiques dynamiques de la dégradation des ARNm chez *E. coli*.

Comme décrit dans la partie précédente, nous avons besoin de plusieurs éléments pour mener à bien l'estimation des paramètres. Premièrement il nous faut définir une fonction structurelle (f) qui sera dans notre cas une simplification du modèle mécanistique développé dans la section 3.3. Cette fonction est ici définie par une équation différentielle décrivant la vitesse de la dégradation. Deuxièmement il nous faut définir un modèle d'erreur. Enfin nous avons besoin d'une implémentation. Nous décrivons dans l'article suivant le développement du modèle statistique, son implémentation et les résultats obtenus avec cette approche. Il s'agit d'une première version d'un article que nous rédigeons actuellement pour valoriser ces résultats. Cette version est encore en train d'évoluer mais décrit de façon générale les informations mécanistiques que nous avons pu retirer de l'analyse du jeu de données omiques à différents taux de croissance. Nous devons encore consolider nos résultats mais les points forts de notre étude qui se dégagent déjà sont d'avoir pu analyser de façon exhaustive les jeux de données utilisés, là où classiquement les analyses de profils de dégradation se concentrent sur ceux se rapprochant de l'exponentielle. Nous avons pu gérer le bruit des données avec notre modèle d'erreur de façon à obtenir une estimation satisfaisante des paramètres de la population et des paramètres individuels. Enfin, nous avons identifié un nouveau mécanisme couplant la dégradation à la transcription des ARNm, basé sur la compétition entre ARNm. Ce mécanisme affecte l'ensemble des ARNm en réponse à des changements de taux de croissance alors que seulement 10% des ARNm apparaissent être régulés spécifiquement.

A mechanistic model informed by dynamic omics data reveals the physiological control of bacterial mRNA decay

Thibault A. Etienne, Eugenio Cinquemani, Laurence Girbal, Muriel Cocaign-Bousquet and Delphine Ropers

Manuscript in preparation

A mechanistic model informed by dynamic omics data reveals the physiological control of bacterial mRNA decay

Thibault A. Etienne^{1,2}, Eugenio Cinquemani², Laurence Girbal¹, Muriel Cocaign-Bousquet¹,
and Delphine Ropers^{*,2}

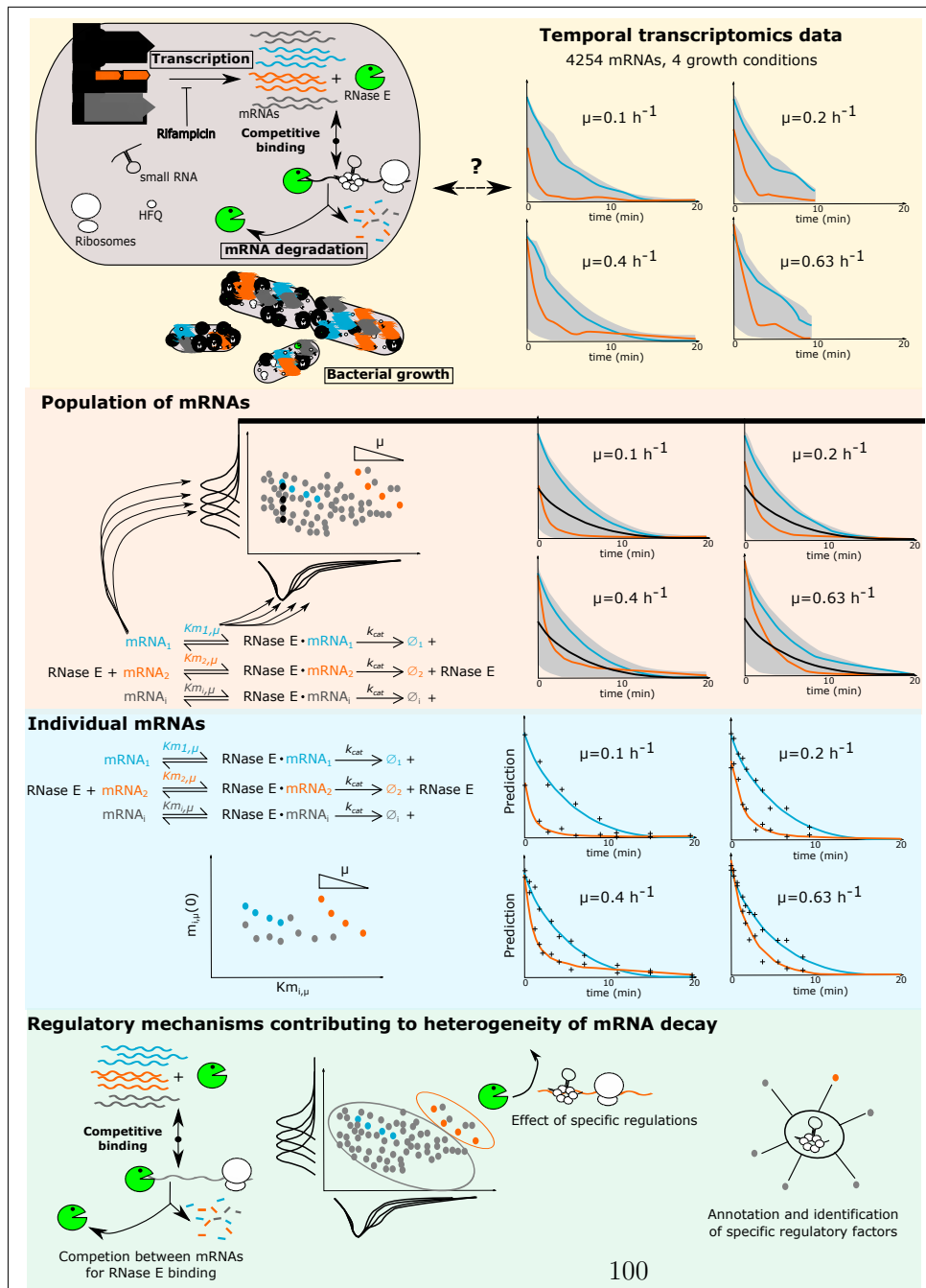
¹TBI, Université de Toulouse, CNRS, INRA, INSA, Toulouse, France

²Univ. Grenoble Alpes, Inria, 38000 Grenoble, France

October 9, 2020

*Correspondence: delphine.ropers@inria.fr

Graphical abstract



Highlights

- Model-based interpretation of dynamic transcriptomic data pose unique opportunities and challenges
- Non-linear mixed-effect models integrating dynamic transcriptomics data disentangle sources of variability of mRNA decay
- Competition between mRNAs is shown to be a new regulatory mechanism coupling mRNA degradation to transcription at the whole-cell level
- Sole 10% of mRNAs are additionally regulated by other specific factors such as small RNAs
- Our modelling and estimation approach is modular and scalable to more complex models and multi-omics data

Summary

Intracellular levels of bacterial mRNAs are regulated from all sides. Myriads of regulatory factors adjust transcription and decay rates to environment. Contrary to transcription, little is known about how cells coordinate mRNA stability changes. Dynamic transcriptomics data provide unique opportunities to address this question but their mechanistic interpretation remains a daunting task. We develop an approach based on the mechanistic modelling of bacterial mRNA degradation at the whole-cell level. We estimate distributions of physiologically-interpretable parameters using non-linear mixed effect modelling and dynamic omics data. Results reveal a new global regulatory mechanism of mRNA decay, competition between mRNAs for binding to the degradation machinery. It accounts for the massive adjustment of mRNA stabilities when environment changes. Sole 10% of mRNAs are additionally regulated by specific factors such as small RNAs. Our approach is able to cope with noisy omics data. It is modular and scalable to more complex models and data.

1 Introduction

Temporal omics data provide unprecedented views of cellular inner workings. They account for the molecular responses of living organisms to perturbations, resulting from regulatory mechanisms that ensure adaptation and survival. The challenge nowadays is to make the most out of these data and uncover key biological features of cellular systems. Statistical modelling tools are commonly used to analyse and interpret these data, from the clustering of gene expression profiles according to their similarity or characteristics, to the identification of differentially-expressed genes or network inference (for reviews, Grigorov 2011; Noor et al. 2019; ?). Achieving mechanistic understanding requires moving from data-driven to mechanistic modelling approaches, by including prior knowledge on the relations between molecular responses and cell physiology (Androulakis et al., 2007; Noor et al., 2019). Mechanistic models have physiologically-interpretable parameters that inform us about cell mechanisms involved in adaptation and survival.

Integrating temporal omics data into mechanistic models is, however, a daunting task and examples of these models are poorly abundant in the literature (Noor et al., 2019). Combinations of technological and practical limitations result in small numbers of biological and technical replicates that can be used and of time points that can be measured. Appropriate normalization of the data is required to distinguish the different sources of noise, resulting from measurement errors and biological process heterogeneity, within and across studies (Liu et al., 2019). This makes the mechanistic models hardly identifiable due their non linearity and their numerous parameters. Enriching the omics data with heterogeneous data improves parameter estimation in principle, but introduces extra noise to be dealt with. The consequence of these shortcomings has been a coarse-graining of the models to help with estimation and a limitation of the information to be extracted from the data. The challenge now is how to overcome these limitations and improve the information content of temporal omics data with more detailed models.

We tackle these issues in our study, by extending a modelling and estimation framework based on non-linear mixed effect (NLME) models for the inference of mechanistic model parameters from time-series omics data. NLME modelling is a well-established and versatile approach for studying variability between individuals in a population. Frequently used in pharmacokinetics-pharmacodynamics studies of the variability of patient responses to drug absorption and elimination, it has been successfully applied in other scientific fields as well (epidemiology, agriculture...) and, more recently, in cell biology to study cell-to-cell variability (Almquist et al., 2015; Davidian and Giltinan, 2003; Davidian, 2009; Hasenauer et al., 2014; Lindstrom and Bates, 1990; Llamasi et al., 2016; Loos et al., 2018). In NLME modelling, the same model is used to describe a physical or biological phenomena observed for each individual and a statistical model describes why the model parameters vary between individuals. The added value of this framework is the consideration of the information for all

individuals simultaneously, which improves parameter estimation (González-Vargas et al., 2016; Oddi et al., 2019). The outcome is a distribution for each model parameter that best fits the data and its estimated statistical parameters (an average value and some variability). In addition, the associated error model describing residual errors allows to handle noisy temporal data and include additional heterogeneous data sets.

Unlike other NLME approaches, we model the diversity of intracellular cell responses to perturbation. Biological processes involved in adaptation and survival share the same cellular machineries, which strongly impacts the molecular composition of cells. For instance, mRNA concentrations raise on average when bacterial growth rate is increased, reflecting a more active transcription by RNA polymerase (Dennis and Bremer, 2008; Dressaire et al., 2013; Esquerré et al., 2014). Deviations from this average behaviour may reflect additional specific regulations, which further tune individual mRNA concentrations (Berthoumieux et al., 2013; Gerosa et al., 2013). With this view, NLME modelling becomes a suitable approach to analyse the diversity of molecular cell responses using temporal omics data, both at the single cell and population level. In this study, we apply the NLME approach to a prototypical example of the issues described so far, bacterial mRNA decay, and analyse its regulation by environmental growth conditions using time-series micro-array data.

In eukaryotic as in prokaryotic cells, mRNA stability is now recognized as a key parameter to adjust gene expression to environment (Dressaire et al. 2013; Esquerré et al. 2014; for review, Carpousis 2007). It is generally determined following transcriptional arrest, by means of microarrays or RNAseq experiments to monitor at the genome-wide level the residual mRNA concentrations along time. The resulting degradation profiles are fitted each by a simple exponential model, assuming a first-order reaction rate and independent degradation reactions (for review, Laguerre et al., 2018). This provides half-life values for cell mRNAs, typically between 3 and 8 minutes in the model bacterium *Escherichia coli* (Bernstein et al., 2002; Chen et al., 2015; Esquerré et al., 2014). A previous micro-array analysis of mRNA decay in continuous cultures of *E. coli* allowed to show that mRNA decay contributes significantly to setting mRNA concentrations at high growth rate (Esquerré et al., 2014). mRNAs become less stable as the growth rate is increased from 0.1 to 0.63 h⁻¹, with observed negative correlations between mRNA concentrations and half lives indicating that more abundant mRNAs are degraded more rapidly (Esquerré et al., 2014; Esquerré et al., 2015; Nouaille et al., 2017). The underlying molecular mechanism remains unknown, although it could involve competition between mRNAs for binding to the degradation machinery (Etienne et al., 2020; Esquerré et al., 2014; Esquerré et al., 2015; Nouaille et al., 2017).

What can we learn more of these data that we already know? Ideally, one would also like to uncover the principles of the physiological control of bacterial mRNA stability and being able to mechanistically explain the degradation profiles observed at various growth rates. Apart from a few exceptions, mRNAs are degraded by the same machinery and additional regulatory mechanisms involving small RNAs, ribonucleoproteins, translating ribosomes, competition between mRNAs... contribute to tune mRNA stability (Etienne et al., 2020; Richards and Belasco, 2019; ?). How do these combined regulatory mechanisms shape mRNA degradation at the genome-wide level? How does this explain the observed negative correlations? Such molecular understanding requires to move from the simple exponential model to a mechanistic model including these regulatory mechanisms. We have recently shown that modelling mRNA processing by the same degradation machinery, the degradosome, and their competition for binding to cell ribonucleases accounts for the observed variability of degradation profiles (Etienne et al., 2020). This simple ordinary differential equation model can be easily extended to include regulatory mechanisms.

In what follows, we adapt our mechanistic model of mRNA degradation and its regulation to the NLME framework. The latter is well adapted for the study of mRNA decay, since mRNA half-lives lie in a narrow range of half-lives and most curves have similar profiles, reflecting the use of a common machinery for their processing. Temporal omics data obtained in (Esquerré et al., 2014) were adequately preprocessed to allow for comparison between mRNAs and conditions, and facilitate their further model integration. The data were enriched by additional data sets, on the concentration of the main degradosome component, RNase E, obtained in a proteomics study (ref) and kinetic parameter values of degradation reactions from biochemical assays (refs), used as priors for estimation. Estimated parameter distributions were obtained for the 4254 degradation curves obtained at four different growth curves, which goes far beyond the traditional analyses that focus on a subset of well-fitted degradation profiles. By investigating the sources of variability of the degradation profiles, we discovered a novel regulatory mechanism of mRNA stability based on the competition between mRNAs for their degradation by RNase E. The regulatory mechanisms allow to couple degradation to transcription, by adjusting mRNA stability to transcript levels. The mechanism affects all mRNA indistinctively but amplifies enzyme affinities for some transcripts. mRNAs with weak affinity will be stabilized by competition and destabilized when affinity is higher. Additional specific regulations tune the stability of 10% of the mRNAs with a fifth of these mRNAs being known targets of small RNAs and validated experimentally. From a methodological side, we show step by step how the non-linear mixed effect framework can be applied to interpret time-series omic data sets and analyse the diversity of molecular cell responses to environmental and genetic perturbations.

2 Results

2.1 Modelling of mRNA decay diversity

A mechanistic model of mRNA degradation describing time-series omics data

We use a deterministic framework based on ordinary differential equations (ODEs) to allow the integration of transcriptomics data obtained from populations of cells at different time points. Each ODE thus describes the change over time of the concentration of a given mRNA i , averaged over the population of cells. Four growth conditions are studied, corresponding to continuous cultures of *E. coli* cells growing at a constant growth rate μ of 0.1, 0.2, 0.4, and 0.63 h⁻¹ (Esquerré et al., 2014). For each μ , we have measurements of the residual concentration over time of 4254 mRNAs ($i = 1, 2, \dots, 4254$). This brings us to a system of 17,016 ODEs describing mRNA decay at the genome-wide level in four growth conditions.

The kinetic rate law in each ODE should describe mRNA decay in a parsimonious manner, with more details than simpler versions assuming first-order reaction rate, while keeping minimal complexity to limit identifiability issues. We use the total quasi-steady-state approximation of the Michaelis-Menten equation for this purpose, as described previously in (Etienne et al., 2020). We showed in this study how to assimilate mRNA degradation by the degradosome to a macro-reaction catalysed by the limiting enzyme RNase E. The reaction describes the binding of free RNase E to a target mRNA and its subsequent cleavage. Sole this initial cleavage is tractable in current transcriptomics approaches that quantify the residual concentration of full length mRNAs. For this reason our model does not describe the following steps during which mRNA fragments are converted to nucleotides by oligoribonucleases (Etienne et al., 2020).

The intracellular concentration of mRNAs is much larger than that of RNase E. As a consequence, mRNAs compete for binding to RNase E and the chance that a given mRNA is degraded by RNase E depends on its own concentration and RNase E affinity, relatively to that of other mRNAs. The enzyme affinity is determined by the mRNA sequence and structure characteristics. It is also modulated by additional regulatory factors, such as small RNAs, proteins (Hfq, CsrA,...), and translating ribosomes (refs). This apparent affinity will be denoted $Km_{i,\mu}$ hereafter. We consider that the value may vary between growth rates, reflecting different regulatory effects in various environments. In the same manner, the initial concentration of each mRNA at the time of transcription inhibition, $m_{i,\mu}(0)$, varies with the environmental growth conditions. This gives the following mechanistic model for the degradation of a given mRNA i with a concentration $m_{i,\mu}(t)$ at time t and growth rate μ (Etienne et al., 2020):

$$\frac{dm_{i,\mu}}{dt} = - \frac{kcat \times E_T \times m_{i,\mu}(t)}{Km_{i,\mu} \times (1 + \sum_{j \neq i} \frac{m_{j,\mu}(t)}{Km_{j,\mu}}) + E_T + m_{i,\mu}(t)}. \quad (1)$$

The catalytic constant $kcat$ is considered to be the same for all mRNAs due to characteristics of the degradation process (see Etienne et al. 2020 for details). E_T is the total concentration of RNase E, including the free form and the enzyme-mRNA complexes. Model units are adapted to the concentration data to be integrated, expressed in arbitrary units per gram of dry cell weight (UA/gDW). This rescales the Km values and the enzyme and mRNA concentrations to the same units (see SI xx for details: assumption of same conversion factor between growth rates, explain how ET is determined). The catalytic constant is given in min⁻¹. The values of $kcat$, $Km_{i,\mu}$, and $m_{i,\mu}(0)$ are unknown and will be later estimated. We have observations for E_T and $m_{i,\mu}(t)$, including $m_{i,\mu}(0)$. The competitive term $(1 + \sum_{j \neq i} \frac{m_{j,\mu}(t)}{Km_{j,\mu}})$ changes RNase E affinity from the apparent value $Km_{i,\mu}$ to a much larger value, which we will denote $Km_{comp,i,\mu}$ hereafter:

$$Km_{comp,i,\mu} = Km_{i,\mu} \times (1 + \sum_{j \neq i} \frac{m_{j,\mu}(t)}{Km_{j,\mu}}) \quad (2)$$

This term slows down mRNA decay. On a methodological side, it creates a dependence between the affinity $Km_{comp,i,\mu}$ of RNase E for its target mRNA i and the concentration and affinity of the 4253 other cell mRNAs j at each growth rate. This couples the 4254 ODEs and makes the problem of inferring model parameters difficult, if not impossible to solve. We rather simplify our model to decouple the equations and allow its further calibration, based on results obtained in Etienne et al. (2020). By means of rate sensitivity coefficients and in agreement with previous experimental studies (Esquerré et al., 2014; Esquerré et al., 2015; Nouaille et al., 2017), we showed that the strongest effect of competition involving mRNA concentration impacts all mRNAs indistinctively. For instance, an increase of the total mRNA concentration slows down mRNA decay at the cell level by increasing the $Km_{comp,i,\mu}$ value of all mRNAs. Specific competition effects on the degradation kinetics of a given mRNA i are much smaller, for instance when the mRNA i has a very high concentration compared to all other mRNAs j or a $Km_{i,\mu}$ very different from the other $Km_{j,\mu}$ values. In other words, the strength of the competition is not strongly affected by differences of Km in the competitive term ($Km_{i,\mu} \approx Km_{j,\mu}$) and

the concentration $m_{i,\mu}(t)$ is negligible with respect to the total mRNA concentration. This allows the following simplification:

$$Km_{comp,i,\mu} \approx Km_{i,\mu} + \sum_{j \neq i} m_{j,\mu}(t) \approx Km_{i,\mu} + \sum_{i=1}^{4254} m_{i,\mu}(t), \quad (3)$$

which gives the following decoupled kinetics model for mRNA decay:

$$\frac{dm_{i,\mu}}{dt} = \frac{kcat \times E_T \times m_{i,\mu}(t)}{Km_{i,\mu} + \sum_{j \neq i} m_{j,\mu}(t) + E_T + m_{i,\mu}(t)} \quad (4)$$

The mechanistic model has now a more suitable form for inferring kinetic parameters and initial conditions. This is a difficult problem though, due to the non linearity of the model and its dimension. With 17,016 ODEs, this requires to estimate 17,016 $Km_{i,\mu}$ values, as much initial conditions, and the $kcat$ parameter. Our estimation strategy is twofold: i) to use literature data for setting RNase E concentration and as prior information for the unknown catalytic constant (see SIxx for details). This information and the use of a total quasi-steady-state approximation of the Michaelis-Menten equation are known to give better estimation results (Choi et al., 2017); ii) to use a so-called population approach exploiting the information for all mRNAs to improve numerical identification of model parameters for each individual mRNA. Among the candidate population approaches, we have chosen the NLME statistical framework for its demonstrated ability to improve parameter estimation relatively to approaches treating each individual independently, and distinguish the sources of heterogeneity of the measurements, between measurement noise and biological variability (Davidian and Giltinan, 2003; Dharmarajan et al., 2019; González-Vargas et al., 2016; Lindstrom and Bates, 1990). The parameters for an individual mRNA result from average parameter values of the population and additional individual-specific deviations from the average. We will interpret parameter variability biologically, between mRNAs and between growth rates, to understand the observed heterogeneity of the degradation curves in various growth conditions.

Modelling and estimation of mRNA decay heterogeneity

The following measurement model relates the model predictions to the experimental observations:

$$Y_{i,t_j,\mu} = f(m_{i,\mu}(t_j,\mu), \Phi_{i,\mu}) + g(f(m_{i,\mu}(t_{i,j,\mu}), \Phi_{i,\mu}), \theta) \times \epsilon_{i,j,\mu} \text{ with } \epsilon_{i,j,\mu} \sim \mathcal{N}(0,1). \quad (5)$$

$Y_{i,t_j,\mu}$ corresponds to the measurement by transcriptomics of the concentration of a given mRNA i at time point $t_{i,j,\mu}$ in the growth condition μ . The original data are published in (Esquerré et al., 2014). They concern the *E. coli* MG1655 strain, grown in glucose-limited cultures in minimal medium and maintained at steady state at four different growth rates (see SIx for more details). At time zero, the transcriptional inhibitor rifampicin was added and the residual mRNA concentrations were monitored by micro-arrays. The sampling time $t_{j,\mu}$ corresponds to 10 time points that differ between growth conditions. A new pre-processing of the raw data was required for the purpose of this study, described in SI (see also Chapter 2 of PhD manuscript). The function f is the numerical solution of our mechanistic model in Eq. 4. It depends on the set of estimated parameters $\Phi_{i,\mu}$, which includes the kinetic parameters and initial conditions: $\Phi = \{k_{cat}, m_{i,\mu}(0), Km_{i,\mu}\}$. Measurement errors $\epsilon_{i,j,\mu}$ are classically assumed to be independent and identically distributed. The function g describes the data measurement noise, which we assume linear. It depends both on the predicted concentrations and on a vector of noise parameters θ , which are identical for all mRNAs but varies between growth rates (see SIx):

$$g(f(m_{i,\mu}(t_j,\mu), \Phi_{i,\mu})) = \theta_\mu^{(0)} + \theta_\mu^{(1)} \times f(m_{i,\mu}(t_j,\mu), \Phi_{i,\mu}) \quad (6)$$

We now turn to the population aspect of our model to account for the variability of mRNA decay in cell cultures growing at different rates. More precisely, we would like to estimate the variability of affinity and initial condition of a given mRNA at a growth rate of reference, which we chose to be 0.1 h^{-1} , and how these variabilities are affected in response to a change of growth rate from 0.1 h^{-1} to $0.2, 0.4$ or 0.63 h^{-1} . This is translated in the following expressions:

$$\begin{cases} kcat_{i,\mu} & = kcat_{pop} \\ Km_{app,i,\mu} & = \log(Km_{pop}) + \eta_i^{Km} + \eta_{i,\mu}^{Km} \\ m_{i,\mu}(0) & = \log(m_{pop}(0)) + \log(\beta_{\mu \neq 0.1}) + \eta_i^{m(0)} + \eta_{i,\mu}^{m(0)} \end{cases} \quad (7)$$

The catalytic constant is set to a fixed value for all mRNAs, while the Km and initial concentrations are allowed to vary. More precisely, the $Km_{i,\mu}$ parameter is described as a combination of a so-called fixed effect, Km_{pop} , corresponding to the average population parameter identical to all mRNAs and all growth rates, and random effects $\eta_i^{Km} + \eta_{i,\mu}^{Km}$ describing the individual variability of mRNA i . The first random effect, η_i^{Km} , models the variability between mRNAs at the reference growth rate (0.1 h^{-1}) ($\eta_i^{Km} \sim N(0, \sigma_{Km}^2)$). It may



Figure 1: Estimation of NLME model for mRNA degradation kinetics at the whole-cell level. The distributions of initial mRNA concentrations and K_m values explain the heterogeneity of the degradation profiles between mRNAs and between growth rates.

result from different mRNA sequence and structure characteristics and regulations by small RNAs and/or RNA-binding proteins. The second one, $\eta_{i,\mu}^{K_m}$ ($\eta_{i,\mu}^{K_m} \sim N(0, \omega_{K_m}^2)$), models the additional variability brought about by the change of growth rate. Concentration changes of a regulatory factor may typically affect this value. Competitive effects that specifically modify enzyme affinity may also be included in the random effect due to the simplification introduced by the decoupling of equations in our final mechanistic model in (4). Both random effects are normally distributed with zero mean and variances denoted σ_{K_m} and ω_{K_m} , respectively. We also estimate the initial concentration of mRNAs for which signal intensities from the micro-array experiments are most noisy at time 0. The initial concentrations can change a lot between growth rates, with an average concentration that takes higher value as the growth rate increases. We model this with a fixed effect that can change between growth rates through a latent categorical variable ($\beta_{\mu \neq 0.1}$). We also add random effects to express differences of mRNA concentrations between the reference growth rate and the other growth rates. The particular choice of a log-normal distribution enforces the positivity of the estimated parameters and reflects the distribution of the measured initial concentrations. We eventually estimate the correlation between random effects (i.e. $cor(\eta_{i,K_m}, \eta_{i,m(0)})$ and $cor(\eta_{i,\mu,K_m}, \eta_{i,\mu,m(0)})$) (see STAR methods section for more details). Parameter values are estimated with a high certainty, with a small RSE for most parameters in all of the four experiments (below 10%; see Section 5.1). Given the good quality of the estimation, we will pursue our analysis with the 4254 degradation profiles at the 4 growth rates in the rest of our study. To best of our knowledge, this is the first time that such an exhaustive analysis of time-series omics data sets is conducted, without discarding a single degradation profile.

2.2 A novel global regulatory mechanism based on competitive effects adjusts mRNA decay to environment

Half-lives of mRNAs are known to be adjusted to environmental changes (Bernstein et al., 2002; Esquerré et al., 2014). A key question is how this happens, that is, by which regulatory mechanism cells control stability of the entire mRNA pool. Small RNAs and regulatory proteins have been shown regulating mRNA decay but whether the sole combination of these regulations explains the adaptation of mRNA decay to environmental perturbations is still unknown. Our NLME model allows to address this question because it engulfs a description of the degradation mechanism at the level of the mRNA population. With our settings, two parameters can vary when growth rate is changed from the reference value $0.1h^{-1}$ to $0.2h^{-1}$, $0.4h^{-1}$ and $0.63h^{-1}$: the initial concentration of mRNAs and their affinity for RNase E. The estimated distributions of these parameters are shown in Figure 1 for the four growth rates.

Parameter distributions show that mRNA concentrations increase with growth rate. The augmentation is

the result of two combined phenomena. The first one is a global adjustment of mRNA concentrations at time 0, reflected by the value of the latent categorical variable $\beta_{\mu \neq 0.1}$. The increase of this variable is the manifestation that the vast majority of mRNAs accumulate inside cells when bacterial growth speeds up. This has been analysed in a previous study with the same data (Esquerré et al., 2014), where increased transcription rates have been shown to mainly contribute to the elevation of mRNA concentrations at higher growth rates. This most likely reflects the higher activity of the transcription machinery to adjust cell macromolecular composition (Dennis and Bremer, 2008; Berthoumioux et al., 2013; Gerosa et al., 2013). The second phenomena is a small rearrangement of certain mRNA concentrations. They increase the variability of intracellular mRNA levels, reflected by larger distributions of mRNA concentrations when cells grow faster. This most likely results from transcriptional regulations that adapt the gene expression program to the higher growth rates.

RNase E affinity is modulated by the sequence and structure characteristics of its target mRNA and possible regulations by small RNAs or proteins like Hfq when they have binding sites on the mRNA. The enzyme affinity is represented by the parameter K_m in our model. No categorical latent variable was needed to adapt its value at the population level when the growth rate is changed. This indicates that on average at the whole-scale level the enzyme affinity does not change in response to environmental perturbations. This can be seen on the estimated distributions of K_m values that overlap at the different growth rates. These distributions show a large width, reflecting a diversity of RNase E affinities for its target mRNAs. As can be seen from the scatter plot in Figure 1, a good proportion of mRNAs have high concentrations and smaller K_m . They will have more chance to be degraded by the enzyme. Some modifications in the shape of the distributions suggest that enzyme affinity can change for some mRNAs when the growth rate varies. This may result from regulatory mechanisms that we will analyse in more details in Section 2.3.

Within cells, mRNAs are numerous and at any rate they outnumber RNase E enzymes (Etienne et al., 2020). This potentially titrates the enzyme at the beginning of the kinetics, which one becomes mainly available for mRNAs with stronger affinities or larger concentrations. In a numerical simulation study using the same mechanistic model and physiologically-relevant parameters, we have shown that competition between mRNAs for their binding to RNase E elicits a large variability of mRNA degradation profiles through the amplification of enzyme affinity: competition stabilizes weakly affine mRNAs and destabilizes the most affine ones (Etienne et al., 2020). Our NLME model provides a unique opportunity to study the reality of this phenomenon *in vivo*. The change of growth rates perturbs the concentration levels of the 4254 mRNAs, which allows to probe the contribution of this mechanism to stability changes in the four different growth conditions studied. In our model the competitive effect is represented by the apparent parameter $Km_{comp,i,\mu}$ resulting from the multiplication of a given $Km_{i,\mu}$ parameter with a term depending on the concentration of all other cell mRNAs.

Based on the estimation of initial mRNA concentrations and individual K_m parameters, we were able to calculate a posteriori the apparent enzyme affinity for each mRNA and each growth rate. Results are shown in Figure 2 as distributions of $Km_{comp,i,\mu}$ values evaluated at time 0. The distributions are shifted to larger values when the growth rate is increased. The phenomenon cannot be attributed to changes in the real K_m values that on average do not vary with growth rates as seen above. The shift reflects the amplification of competitive effects at higher growth rates due to the global increase of mRNA concentrations, which tends to stabilize mRNAs.

We further analysed whether the competitive effect affects mRNAs in the same way. The distributions in Figure 2 show that $Km_{comp,i,\mu}$ values become more variable as growth changes from low to high rates. They reflect that individual $Km_{i,\mu}$ values are affected by the competition. This is confirmed with the important correlation between random effects of individual mRNA concentrations and K_m values in Figure 4. Remember that the change of these K_m values between growth rates reflects at least partially the effect of competition at the individual level due to the decoupling of model equations (Section 2.1). Hence, the effect of competition on individual mRNAs is that at high concentration they will also have lower $Km_{i,\mu}$ values and thus have more chances to be degraded. At low concentration with higher K_m values, they will be stabilized. This is in agreement with the observation by (Esquerré et al., 2014) of a negative correlation between mRNA concentrations and half-lives, further confirmed experimentally by (Nouaille et al., 2017). Our model provides a molecular explanation to this phenomenon, through the prism of mRNA competition.

In summary, competition between mRNAs strongly impacts mRNA decay at the whole-cell level through the titration of RNase E in the minutes following the onset of degradation. The effect is direct through the initial concentration of cell mRNAs and indirect by modifying the enzyme affinity for its targets. The existence of such global regulatory mechanism is new in mRNA degradation studies. It implies that it is enough for cells to orchestrate the adaptation of their mRNA levels to environment through transcriptional regulations. The coupling of degradation to transcription through competition elicits at once a global regulation of the degradation kinetics and more specific regulations of mRNA half-lives.

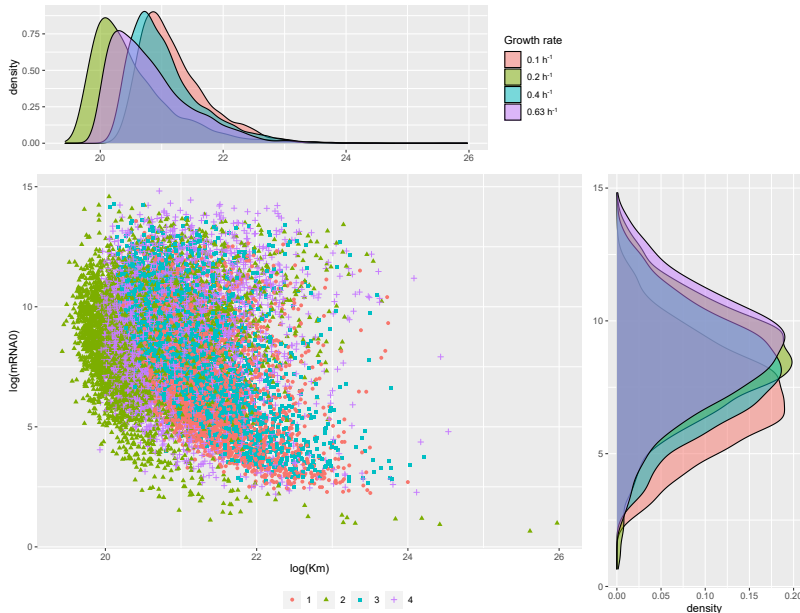


Figure 2: Estimation of NLME model for mRNA degradation kinetics at the whole-cell level. The distributions of initial mRNA concentrations and $Km_{comp,i,\mu}$ values explain the heterogeneity of the degradation profiles between mRNAs and between growth rates.

2.3 Specific regulations complement global competitive effects for 10% of mRNAs only

At this stage we have established the importance of competition for adjusting mRNA decay to transcription. Does that mean that additional regulations by small RNAs or RNA-binding proteins are negligible in this picture? We investigated this question by analysing most extreme Km values. Although we know that their slight decrease between growth rates reflects competitive effects at the individual scale, Km values between individuals at the reference growth rate result from different mRNA specificities in terms of sequence, secondary structure and regulations. A given mRNA without any regulation is expected to have a $Km_{i,\mu}$ at 0.1 h^{-1} close to the population parameter Km_{pop} . The inhibition of mRNA decay with a small RNA, RNA-binding protein or translating ribosomes will increase significantly the $Km_{i,\mu}$ value. In case of activation, the decrease will be noticeable as well.

We use a clustering approach with a Gaussian mixture model to identify mRNA sub-populations with regulated mRNA decay (see STAR methods). Among the 4 clusters identified in Figure 3 for the reference growth rate at 0.1 h^{-1} , 2 of them are organized along the diagonal. This reflects the strong correlation between initial concentrations and Km values that we have attributed to competitive effects at the individual level. The 2 most distant clusters, coloured in blue and purple in the figure, include points characterized by larger Km values and large or small mRNA concentrations at time 0. The cluster with smaller mRNA concentrations include 468 mRNAs. They have flat degradation curves suggesting that the process could be inhibited but their concentration is so close to the background that it is not certain at this stage whether they are transcribed at all. In this case the flat degradation curve and corresponding high Km value reflect that there is almost no mRNA to be degraded. This does not result from a regulatory mechanism. We will thus omit this cluster from the rest of our study. Note that further analyses at other growth rates are planned to verify if some of these mRNAs can be transcribed in other conditions. The cluster in blue represents the mRNAs targeted by specific regulatory mechanisms fine-tuning the global competition effects. Surprisingly, this cluster includes 409 mRNAs, which means that sole 10% of cell mRNAs should be the target of specific mechanisms. Functional annotation of this cluster of mRNAs shows that they belong to the central carbon metabolism providing carbon and energy precursors needed for cell growth (SIX section, to be completed). This result is in line with previous observations with the same data set that mRNAs in this pathway are mainly regulated at the stability level (Esquerré et al., 2014).

The central carbon metabolism is essential for cell growth and survival in a variety of growth conditions. Recent studies have pointed at its regulation by post-transcriptional factors, such as small RNAs, protein Hfq or the carbon storage regulator (Esquerré et al., 2014; Morin et al., 2016, 2017; ?, e.g.). Their regulatory effect

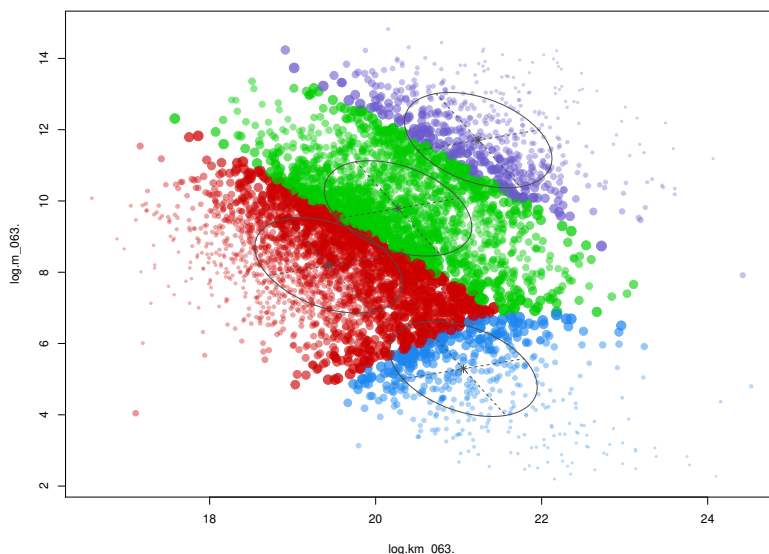


Figure 3: Clustering with a Gaussian mixture model of individual model parameters. Axes are on a logarithmic scale. Results obtained with best model based on BIC values are shown. Four clusters are identified, represented by different colours. Two of them (red and green) are organized along the diagonal reflecting the competitive effects. The cluster in purple is associated to mRNAs that are seemingly not described in cells at the growth rate 0.63 h^{-1} . Points from the blue cluster are distant and seem to result from specific regulations inhibiting mRNA decay. The diameter of the points reflects the uncertainty of the clustering.

should complement the global effect induced by competition and even counteract it. Indeed further analysis of Km parameters at other growth rates show that active regulations identified at the reference growth rate are still functional at higher growth rates (SIx). As a consequence, the increased concentration of this sub-population of mRNAs is antagonized by specific regulations alleviating the acceleration of mRNA decay imposed by the concentration changes. Why an additional layer of regulation is needed is not clear, although this seems to be a general phenomenon given that a combination of specific and global effects regulate transcription rates as well in *E. coli* (Berthoumieux et al., 2013; Dennis and Bremer, 2008; Gerosa et al., 2013).

2.4 Specific regulations by small RNAs contribute to the heterogeneity of degradation profiles between mRNAs and between growth rates

The knowledge of the key role played by mRNA degradation in the adjustment of intracellular mRNA concentrations is recent. As a consequence, we still do not have an exhaustive repertoire of the target mRNAs and their regulatory factors. It is thus out of reach to validate the list of the 400 mRNAs regulated specifically in *E. coli*. However this list should contain at least known targets from the literature. We thus gathered a list of experimentally confirmed regulatory factors and target mRNAs such as RhyB, GcvB and OmrA from (Keseler et al., 2017; ?). Out of the 400 mRNAs, 150 are known to be regulated by small RNAs based on our list and 90 of these 150 mRNAs exhibit large individual Km values according to our estimation results (SIx). The prediction of a regulation for these 90 mRNAs is thus validated by experimental data. The 60 other mRNAs may correspond to situations where the antagonistic competition and regulatory effects annihilate each other, resulting in individual Km values close to the average population parameter. Hence, our NLME approach allows to identify mRNAs specifically regulated at the degradation level. More generally, it allows to disentangle the variability of the degradation profiles due to competition and that due to specific regulations. The approach also predicts new target mRNAs, 250 in our case, awaiting further experimental studies for confirmation and to investigate the regulatory mechanism involved.

Small RNAs are themselves regulated and their concentration can be tuned when the environment changes, which suggests that their regulon should be tuned accordingly (Kaberina et al., 2019, e.g.). Can we observe such phenomenon in our degradation profiles? To answer the question, we focused on one specific small RNA, GcvB. This Hfq-dependent RNA regulates the stability of transporters of amino acids and short peptides, amino-acid biosynthesis proteins and major transcription factors such as Lrp, PhoP, as well as CsgD (refs).

GcvB is highly abundant in rich media supporting high growth rates. In numerous cases, GcvB stabilizes its target mRNAs. A closer inspection of the estimated individual Km values for the target in our list of 400 mRNAs revealed that the Km values increase with the growth rate. Most likely, this phenomenon reflects the elevation of GcvB pools when cells grow faster which concomitantly stabilize the target mRNAs. Our NLME approach thus allows us to identify the consequences on mRNA decay of environmental perturbations of the regulator levels.

3 Discussion

With technological advances, more and more omics data accumulate and more recently, the data have turned dynamic. How to gain a mechanistic understanding from them is challenging. This type of question is best addressed by mechanistic models but they often fail to describe data due to noise and identification problems to name but a few. As a result a deeper interpretation of time-series omics data is still lagging behind. We have overcome these issues in our study by using a non-linear mixed-effect framework which we applied to the analysis of mRNA degradation. Time-series data are available to describe the dynamics of this process but analyses have not gone beyond the determination of mRNA half-lives so far. Can we extract more information from the profiles and learn about the underlying regulatory mechanisms? Based on a mechanistic model of mRNA degradation including the omics data and additional heterogeneous information, the mixed effect framework allowed us to deal with the noise of the data and to estimate reliably model parameters with physiological interpretation at the population and individual levels. This provides us with a global picture to which extent variabilities of parameter values account for the observed heterogeneity of the degradation profiles.

By investigating the sources of variability of the profiles, we identified a novel regulatory mechanism of mRNA decay. It relies on the competition between mRNAs for their binding to RNase E. In the first minutes after the onset of degradation, the enzyme is titrated by the myriad of cell mRNAs. Sole those with higher concentrations and/or higher affinities have more chances to be degraded. Competition on the one hand tends to stabilize mRNAs by increasing the competitive effect at the whole-cell level. On the other hand, it affects mRNAs individually through the modification of their affinity: competition stabilizes mRNAs with weaker affinities and destabilizes those with stronger affinities. This global mechanism allows to adjust degradation rates to the intracellular mRNA concentrations. The latter are in large part regulated at the transcriptional level (Esquerré et al., 2014). This means that degradation is directly coupled to transcription and that mRNA competition is the cornerstone of this regulatory mechanism. Surprisingly, sole 10% of cell mRNAs are the target of additional specific regulations. This shared control of a biological process by a combination of global and specific effects is reminiscent of a similar mechanism observed for transcription (Berthoumieux et al., 2013; Gerosa et al., 2013).

While sub-sets of time-series data are generally analysed in studies of mRNA degradation, our approach is exhaustive since all cell mRNAs are included and their degradation profiles fitted. This allows to draw conclusions valid at the entire cell level. On a methodological side, we have shown step by step how the non-linear mixed effect framework can be applied to interpret time-series omic data sets and how to analyse the diversity of molecular cell responses to environmental and genetic perturbations. The approach is modular and scalable. It can be extended to other types of models and to other single-cell or population data, as well as additional heterogeneous data. In the current state of our study, our model estimation do not appropriately address the problem of delay before the onset of degradation. We are currently working on these aspects and refining our estimation results to be able to study the contribution of competition to the retardation of degradation. In our former study, we predicted that this could be an important mechanism in addition to the delay attributed to the completion of transcription elongation.

4 Star method

To be completed

5 Supplementary information

5.1 Estimation of population parameters

Figure 4 shows the results of the estimation of the NLME model with the time-series omics data :

	VALUE	STOCH. APPROX.	
		S.E.	R.S.E.(%)
Fixed Effects			
kcat_pop	11		
Km_pop	8.57e+8	1.32e+7	1.54
mRNA0_pop	1.35e+3	93.7	6.96
beta_mRNA0_cat3_2	1.25	0.0993	7.91
beta_mRNA0_cat3_3	0.935	0.0229	2.44
beta_mRNA0_cat3_4	1.57	0.104	6.61
Standard Deviation of the Random Effects			
omega_Km	0.369	nan	nan
omega_mRNA0	2.04	0.027	1.33
gamma_Km	0.968	0.03	3.1
gamma_mRNA0	0.405	0.0549	13.5
Correlations			
corr1_mRNA0_Km	0.559	0.0507	9.06
corr2_mRNA0_Km	-0.382	0.214	55.9
Error Model Parameters			
b1	0.236	0.00093	0.394
b2	0.171	0.000699	0.409
b3	0.189	0.000738	0.391
b4	0.165	0.000679	0.411

Elapsed time (seconds)¹²: 5.9e+3
CPU time (seconds)¹⁰: 1.1e+5

5.2 To be completed

References

- Joachim Almquist, Loubna Bendrioua, Caroline Beck Adiels, Mattias Goksör, Stefan Hohmann, and Mats Jirstrand. A nonlinear mixed effects approach for modeling the cell-to-cell variability of *mig1* dynamics in yeast. *PLoS one*, 10(4):e0124050, 2015.
- I.P. Androulakis, E. Yang, and R.R. Almon. Analysis of time-series gene expression data: Methods, challenges, and opportunities. *Annual Review of Biomedical Engineering*, 9(1):205–228, Aug 2007. ISSN 1545-4274. doi: 10.1146/annurev.bioeng.9.060906.151904.
- Jonathan A Bernstein, Arkady B Khodursky, Pei-Hsun Lin, Sue Lin-Chao, and Stanley N Cohen. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc Natl Acad Sci USA*, 99(15):9697–9702, 2002.
- S. Berthoumieux, H. de Jong, G. Baptist, C. Pinel, C. Ranquet, D. Ropers, and J. Geiselmann. Shared control of gene expression in bacteria by transcription factors and global physiology of the cell. *Mol. Syst. Biol.*, 9: 634, 2013.
- Agamemnon J Carpousis. The RNA degradosome of *Escherichia coli*: an mRNA-degrading machine assembled on RNase E. *Annu Rev Microbiol*, 61:71–87, 2007.
- Huiyi Chen, Katsuyuki Shiroguchi, Hao Ge, and Xiaoliang Sunney Xie. Genome-wide study of mRNA degradation and transcript elongation in *Escherichia coli*. *Mol Syst Biol*, 11(1):781, 2015.
- Boseung Choi, Grzegorz A Rempala, and Jae Kyoung Kim. Beyond the michaelis-menten equation: Accurate and efficient estimation of enzyme kinetic parameters. *Scientific reports*, 7(1):1–11, 2017.
- Marie Davidian. Non-linear mixed effect models. *Longitudinal data analysis*, pages 107–142, 2009.
- Marie Davidian and David M Giltinan. Nonlinear models for repeated measurement data: an overview and update. *Journal of agricultural, biological, and environmental statistics*, 8(4):387, 2003.
- Patrick P Dennis and Hans Bremer. Modulation of chemical composition and other parameters of the cell at different exponential growth rates. *EcoSal Plus*, 3(1):1–49, 2008.
- Lekshmi Dharmarajan, Hans-Michael Kaltenbach, Fabian Rudolf, and Joerg Stelling. A simple and flexible computational framework for inferring sources of heterogeneity from single-cell dynamics. *Cell systems*, 8(1): 15–26, 2019.
- Clémentine Dressaire, Flora Picard, Emma Redon, Pascal Loubière, Isabelle Queinnec, Laurence Girbal, and Muriel Coccagn-Bousquet. Role of mRNA stability during bacterial adaptation. *PLoS one*, 8(3):e59059, 2013.
- Thomas Esquerré, Sandrine Laguerre, Catherine Turlan, Agamemnon J Carpousis, Laurence Girbal, and Muriel Coccagn-Bousquet. Dual role of transcription and transcript stability in the regulation of gene expression in *Escherichia coli* cells cultured on glucose at different growth rates. *Nucleic Acids Res*, 42(4):2460–2472, 2014.
- Thomas Esquerré, Annick Moisan, Hélène Chiapello, Liisa Arike, Raivo Vilu, Christine Gaspin, Muriel Coccagn-Bousquet, and Laurence Girbal. Genome-wide investigation of mrna lifetime determinants in *escherichia coli* cells cultured at different growth rates. *BMC Genomics*, 16(1), Apr 2015. ISSN 1471-2164. doi: 10.1186/s12864-015-1482-8. URL <http://dx.doi.org/10.1186/s12864-015-1482-8>.
- Thibault A Etienne, Muriel Coccagn-Bousquet, and Delphine Ropers. Competitive effects in bacterial mrna decay. *Journal of Theoretical Biology*, page 110333, 2020.
- L. Gerosa, K. Kochanowski, M. Heinemann, and U. Sauer. Dissecting specific and global transcriptional regulation of bacterial gene expression. *Mol. Syst. Biol.*, 9:658, 2013.
- Andrés M González-Vargas, Eugenio Cinquemani, and Giancarlo Ferrari-Trecate. Validation methods for population models of gene expression dynamics. *IFAC-PapersOnLine*, 49(26):114–119, 2016.
- Martin G Grigorov. Analysis of time course omics datasets. In *Bioinformatics for Omics Data*, pages 153–172. Springer, 2011.
- Jan Hasenauer, Christine Hasenauer, Tim Hucho, and Fabian J Theis. Ode constrained mixture modelling: a method for unraveling subpopulation structures and dynamics. *PLoS Comput Biol*, 10(7):e1003686, 2014.

- Anna Chao Kaberdina, Olatz Ruiz-Larrabeiti, Sue Lin-Chao, and Vladimir R Kaberdin. Reprogramming of gene expression in *Escherichia coli* cultured on pyruvate versus glucose. *Mol Genet Genomics*, 294(5):1359–1371, 2019.
- Ingrid M Keseler, Amanda Mackie, Alberto Santos-Zavaleta, Richard Billington, César Bonavides-Martínez, Ron Caspi, Carol Fulcher, Socorro Gama-Castro, Anamika Kothari, Markus Krummenacker, et al. The EcoCyc database: reflecting new knowledge about escherichia coli K-12. *Nucleic Acids Res*, 45(D1):D543–D550, 2017.
- Sandrine Laguerre, Ignacio González, Sebastien Nouaille, Annick Moisan, Nathalie Villa-Vialaneix, Christine Gaspin, Marie Bouvier, Agamemnon J Carpousis, Muriel Coccagn-Bousquet, and Laurence Girbal. Large-scale measurement of mRNA degradation in *Escherichia coli*: To delay or not to delay. *Methods Enzymol*, 612:47–66, 2018.
- Mary J Lindstrom and Douglas M Bates. Nonlinear mixed effects models for repeated measures data. *Biometrics*, pages 673–687, 1990.
- Xueyan Liu, Nan Li, Sheng Liu, Jun Wang, Ning Zhang, Xubin Zheng, Kwong-Sak Leung, and Lixin Cheng. Normalization methods for the analysis of unbalanced transcriptome data: a review. *Frontiers in Bioengineering and Biotechnology*, 7, 2019.
- Artémis Llamosi, Andres M Gonzalez-Vargas, Cristian Versari, Eugenio Cinquemani, Giancarlo Ferrari-Trecate, Pascal Hersen, and Gregory Batt. What population reveals about individual cell identity: single-cell parameter estimation of models of gene expression in yeast. *PLoS computational biology*, 12(2):e1004706, 2016.
- Carolin Loos, Katharina Moeller, Fabian Fröhlich, Tim Hucho, and Jan Hasenauer. A hierarchical, data-driven approach to modeling single-cell populations predicts latent causes of cell-to-cell variability. *Cell Systems*, 6(5):593–603, 2018.
- M. Morin, D. Ropers, F. Letisse, S. Laguerre, J.-C. Portais, M. Coccagn-Bousquet, and B. Enjalbert. The post-transcriptional regulatory system csr controls the balance of metabolic pools in upper glycolysis of escherichia coli. *Mol Microbiol*, 100(4):686–700, 2016.
- M. Morin, D. Ropers, E. Cinquemani, J.-C. Portais, B. Enjalbert, and M. Coccagn-Bousquet. The CSR system regulates *Escherichia coli* fitness by controlling glycogen accumulation and energy levels. *mBio*, 8(5), 2017.
- Elad Noor, Sarah Cherkaoui, and Uwe Sauer. Biological insights through omics data integration. *Current Opinion in Systems Biology*, 15:39 – 47, 2019. ISSN 2452-3100. doi: <https://doi.org/10.1016/j.coisb.2019.03.007>. URL <http://www.sciencedirect.com/science/article/pii/S2452310019300125>. Gene regulation.
- Sébastien Nouaille, Sophie Mondeil, Anne-Laure Finoux, Claire Moulis, Laurence Girbal, and Muriel Coccagn-Bousquet. The stability of an mRNA is influenced by its concentration: a potential physical mechanism to regulate gene expression. *Nucleic Acids Res*, 45(20):11711–11724, 2017.
- Facundo J Oddi, Fernando E Miguez, Luciana Ghermandi, Lucas O Bianchi, and Lucas A Garibaldi. A nonlinear mixed-effects modeling approach for ecological data: Using temporal dynamics of vegetation moisture as an example. *Ecology and Evolution*, 9(18):10225–10240, 2019.
- Jamie Richards and Joel G Belasco. Obstacles to scanning by rnase e govern bacterial mrna lifetimes by hindering access to distal cleavage sites. *Molecular cell*, 74(2):284–295, 2019.

Conclusions et discussion

Au cours de cette thèse, différentes approches pour modéliser et analyser des données de transcriptomique en dynamique précédemment obtenues par T. Esquerré chez *E. coli* sont abordées sous le prisme de la biologie des systèmes. Dans la figure 4.3, nous représentons schématiquement les différentes étapes (correspondant aux différents chapitres de ce manuscrit) de la méthodologie mise en place. L'épaisseur des flèches correspond à l'intensité des dépendances entre ces étapes.

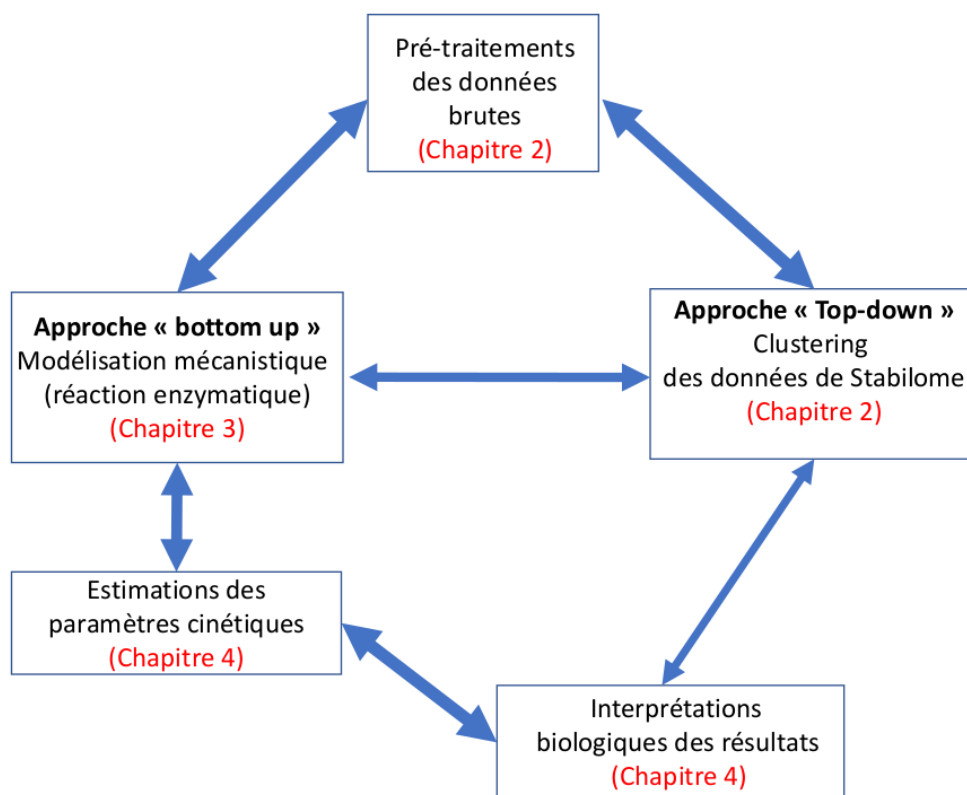


FIGURE 4.3 – Méthodologie mise en place pour l'étude des données de transcriptome en dynamique

L'idée a été d'étendre les travaux déjà effectués sur le sujet à l'aide de modèles mathématiques et statistiques afin de pousser notre compréhension de ce problème biologique. A l'aide d'une approche sans a priori (top-down), nous avons exploré les dynamiques des profils de dégradation (cf. chapitre 2). Cela a permis de confirmer qu'une majorité d'ARNm semble suivre des profils de dégradation monotones qu'on peut associer à une exponentielle décroissante. Cependant, une part significative d'entre eux ont des dynamiques plus complexes qu'il ne

faut pas négliger. Pour mieux comprendre l'origine et les mécanismes sous-jacents induisant cette variété de profils de dégradation, une approche bottom-up basée sur le développement d'un modèle mécanistique a été mise en place avec, comme objectif, la modélisation de la dégradation des ARNm bactériens (chapitre 3). Ce modèle permet de montrer, en outre, le rôle de la compétition entre les ARNm dans la variabilité de la stabilité des ARNm et l'apparition de délais avant la dégradation. Ces travaux ont fait l'objet d'un article publié dans *Journal of Theoretical Biology* dont je suis 1^{er} auteur.

Une méthodologie statistique (cf. chapitre 4) utilisant les modèles non linéaires à effets mixtes a ensuite été adaptée au modèle mécanistique et éprouvée pour l'estimation des paramètres cinétiques de ce modèle à partir de données omiques dynamiques. L'interprétation biologique des paramètres estimés montre que la compétition joue un rôle majeur dans l'adaptation de la stabilité des ARNm en couplant étroitement la dégradation à la transcription. Ce couplage intervient à travers les composantes aléatoires des paramètres estimés (concentrations initiales et constantes de Michaelis-Menten). Ces travaux ont donné lieu à une première version d'un article (chapitre 4) dont je suis 1^{er} auteur. Notons que celui-ci nécessite encore quelques ajustements, en particulier sur l'interprétation et la robustesse des corrélations estimées. Il est également prévu de renforcer l'interprétation biologique à l'aide d'une analyse d'enrichissement fonctionnel utilisant les effets aléatoires estimés, entre les ARNm aux taux de croissance de référence et entre les taux de croissance pour un ARNm donné. Les premiers résultats conduisent à l'émergence d'un nouveau mécanisme global de la dégradation permettant une réponse plus efficace des cellules dans leur adaptation aux changements d'environnement. Il s'agit d'un travail novateur à deux points de vue. Sur le plan biologique, il nous a permis de mettre en évidence un nouveau mécanisme de régulation de la dégradation des ARNm et d'observer qu'une fraction des ARNm cellulaires seulement (10%) est régulée par des mécanismes spécifiques additionnels dans les conditions expérimentales étudiées. Il sera d'ailleurs intéressant de voir si ces conclusions sont spécifiques à l'adaptation du taux de croissance ou s'ils restent valides dans d'autres conditions expérimentales comme des changements d'environnement nutritif ou des contextes génétiques différents voire dans d'autres micro-organismes. Sur le plan méthodologique, ce travail nous a permis de pousser plus loin l'analyse de données omiques dynamiques à l'aide de modèles mécanistiques plus informatifs, sans laisser de côté des jeux de données au profil non exponentiel comme c'est le cas généralement.

Les travaux sur la dégradation des ARNm décrits dans ce manuscrit ont demandé un travail bibliographique conséquent, tant sur les études expérimentales que celles de modélisation mathématique. Ils ont été valorisés par un article de revue en cours de finalisation auquel j'ai contribué en tant que 2^e auteur (section 5.2). Nous discutons dans cet article qu'en plus du mécanisme de couplage entre la transcription et la dégradation mis en évidence dans cette thèse, de nombreux autres couplages lient la dégradation des ARNm à d'autres processus cellulaires comme la traduction, le métabolisme ou la localisation cellulaire. Ces différents processus s'influencent mutuellement, ce qui rend la compréhension de la physiologie bactérienne très complexe. Ceci justifie l'utilisation d'approches de modélisation intégrant des données expérimentales pour mieux caractériser la contribution de ces différents processus cellulaires à la régulation de la stabilité des ARNm.

Les différents travaux exposés dans ce manuscrit ont été spécifiquement discutés dans les

articles consacrés. Ils nous ouvrent plusieurs perspectives, notamment en terme d’extension. Premièrement, la méthodologie basée sur les modèles non linéaires à effets mixtes offre un cadre flexible permettant de faire évoluer la nature des données et les modèles mécanistiques pris en compte. Elle est ainsi applicable à d’autres types de données, comme par exemple les données de RNAseq qui sont devenues le standard dans les études de transcriptomique. En outre, le développement des données omiques permet d’enrichir notre modèle mécanistique par leur intégration sous forme de covariable, comme par exemple des données de ribosome footprint ou encore des mesures dynamiques des concentrations de petits ARN et de la RNase E. La prise en compte de ces différentes données permettrait d’améliorer la qualité des estimations et de raffiner notre connaissance des mécanismes de régulation de la dégradation. La contrainte est de pouvoir gérer le bruit résultant de l’utilisation de données hétérogènes, ce que le cadre de la modélisation non linéaire à effets mixtes permet de faire.

Dans nos travaux, nous n’avons pas pris en compte d’aspects spatiaux. Cependant des études récentes soulignent l’importance de la localisation à la fois de la RNase E mais aussi des ARNm [Moffitt et al., 2016, Hadjeras et al., 2019, Al-Husini et al., 2020]. Comment les inclure dans nos approches de modélisation ? La compréhension des incidences spatiales ont été abordées au niveau théorique par des modèles de réaction enzymatique intégrant le phénomène de diffusion de l’enzyme ou du substrat (modèle de réaction-diffusion) [Frank et al., 2018]. Cependant ces modèles ne sont pas applicables en tant que tels dans notre contexte. Certaines pistes sont donc à explorer pour leur intégration dans nos approches comme, par exemple, l’utilisation des hypothèses de quasi équilibre pour approximer ce type de modèle. Une autre piste serait d’utiliser des modèles de dégradation formulés à l’aide d’une équation stochastique différentielle puis de les confronter aux données via les modèles non linéaires à effet mixtes [Donnet and Samson, 2013].

Enfin, l’observation d’un couplage étroit entre le processus de dégradation des ARNm et les régulations métaboliques semble aussi très prometteur pour pousser plus loin notre étude (section 5.2). Pour cela, le couplage des méthodes de fluxomique et l’intégration de données de métabolomique à notre analyse peut affiner notre compréhension de l’adaptation de la stabilité des ARNm à des changements d’environnement.

Les différentes extensions envisagées ne sont pas mutuellement exclusives. Elles concourent à pousser l’intégration dans des modèles de données expérimentales multiples sur les différents processus et mécanismes modulant la stabilité des ARNm. Cela nous permettrait au final d’obtenir une vision intégrée de la régulation de la dégradation des ARNm.

Chapitre 5

Annexes

5.1 Densité des intensité brutes

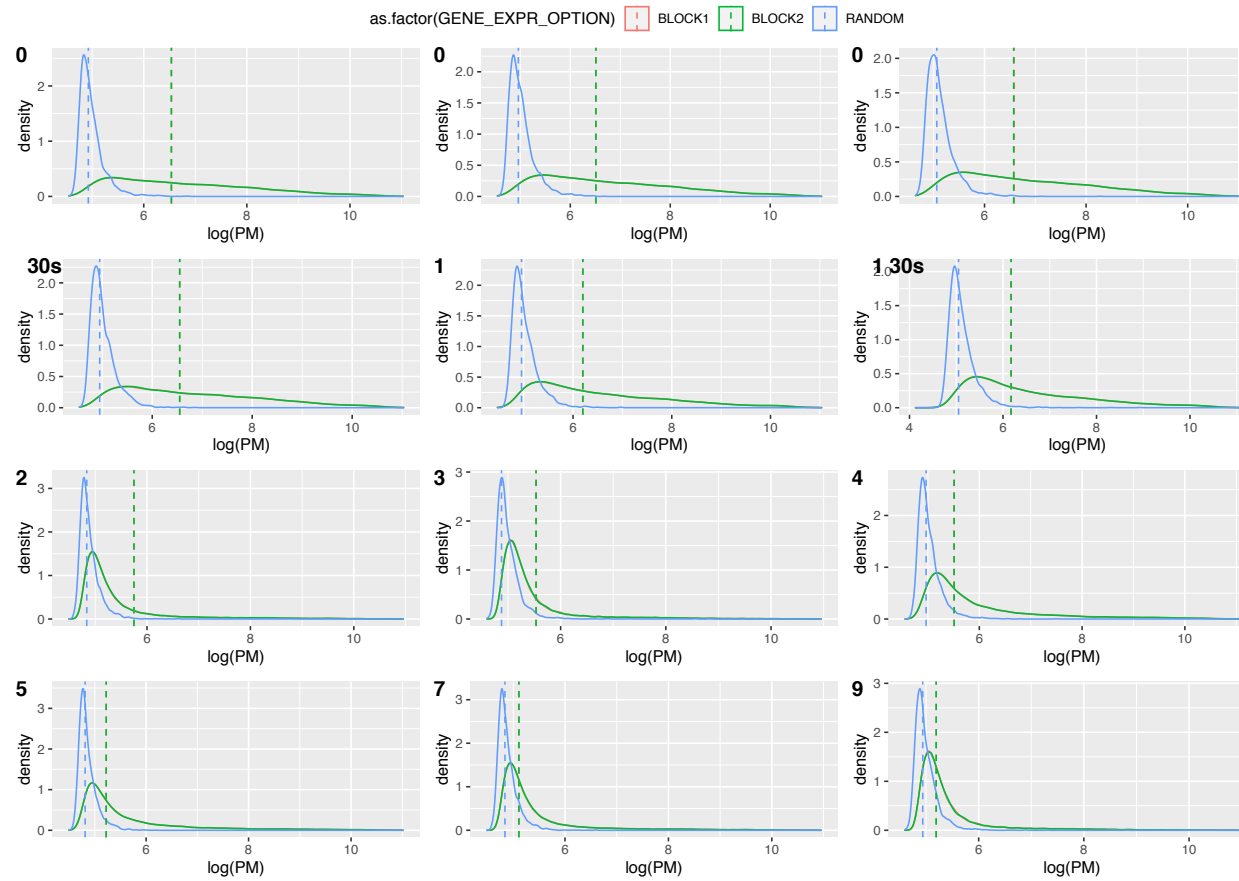


FIGURE 5.1 – Densité des intensités des sondes spécifiques (vert et rouge) et non spécifiques (bleu). Les valeurs de densité sont données pour chaque temps de mesure au taux de croissance $0.2h^{-1}$. Les intensités sont données en abscisse sur une échelle logarithmique

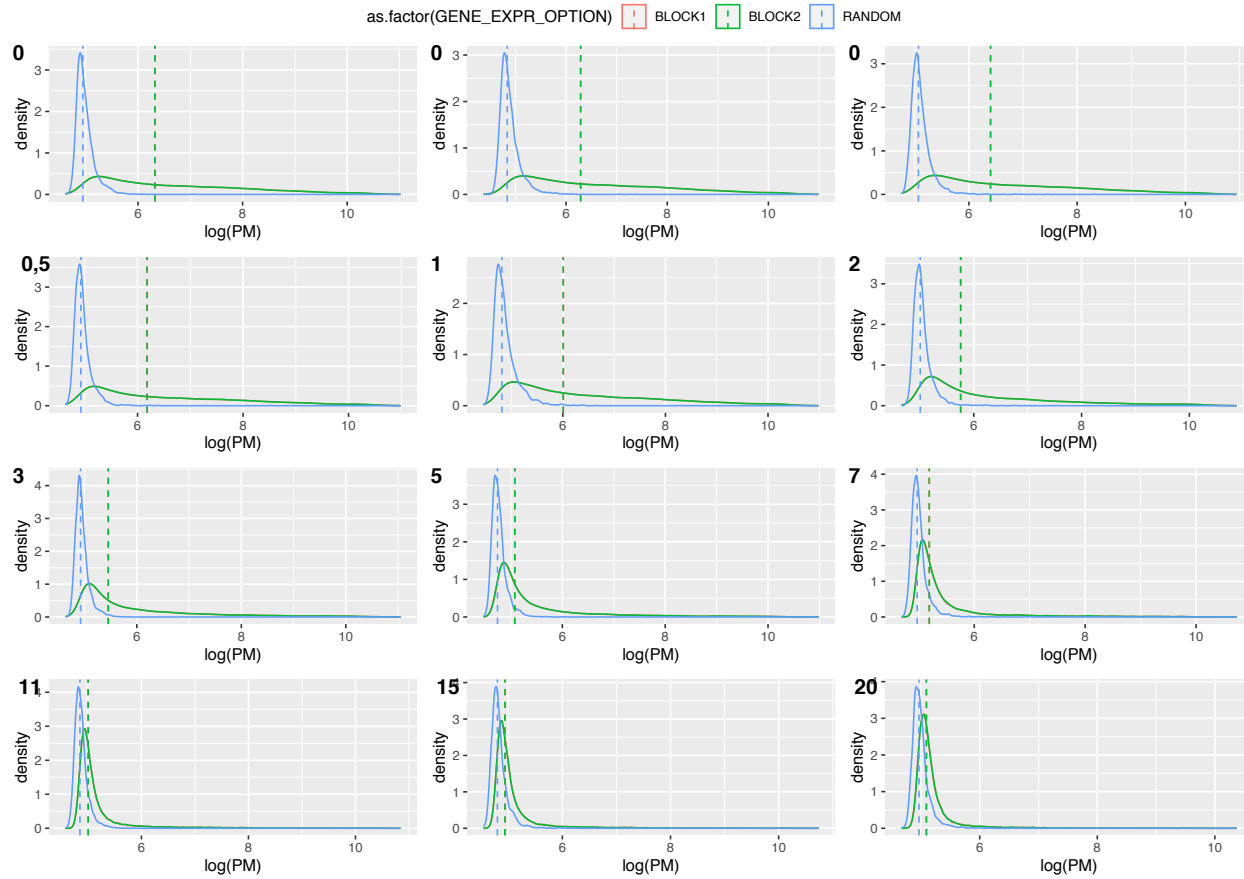


FIGURE 5.2 – Densité des intensités des sondes spécifiques (vert et rouge) et non spécifiques (bleu). Les valeurs de densité sont données pour chaque temps de mesure au taux de croissance $0.4h^{-1}$. Les intensités sont données en abscisse sur une échelle logarithmique

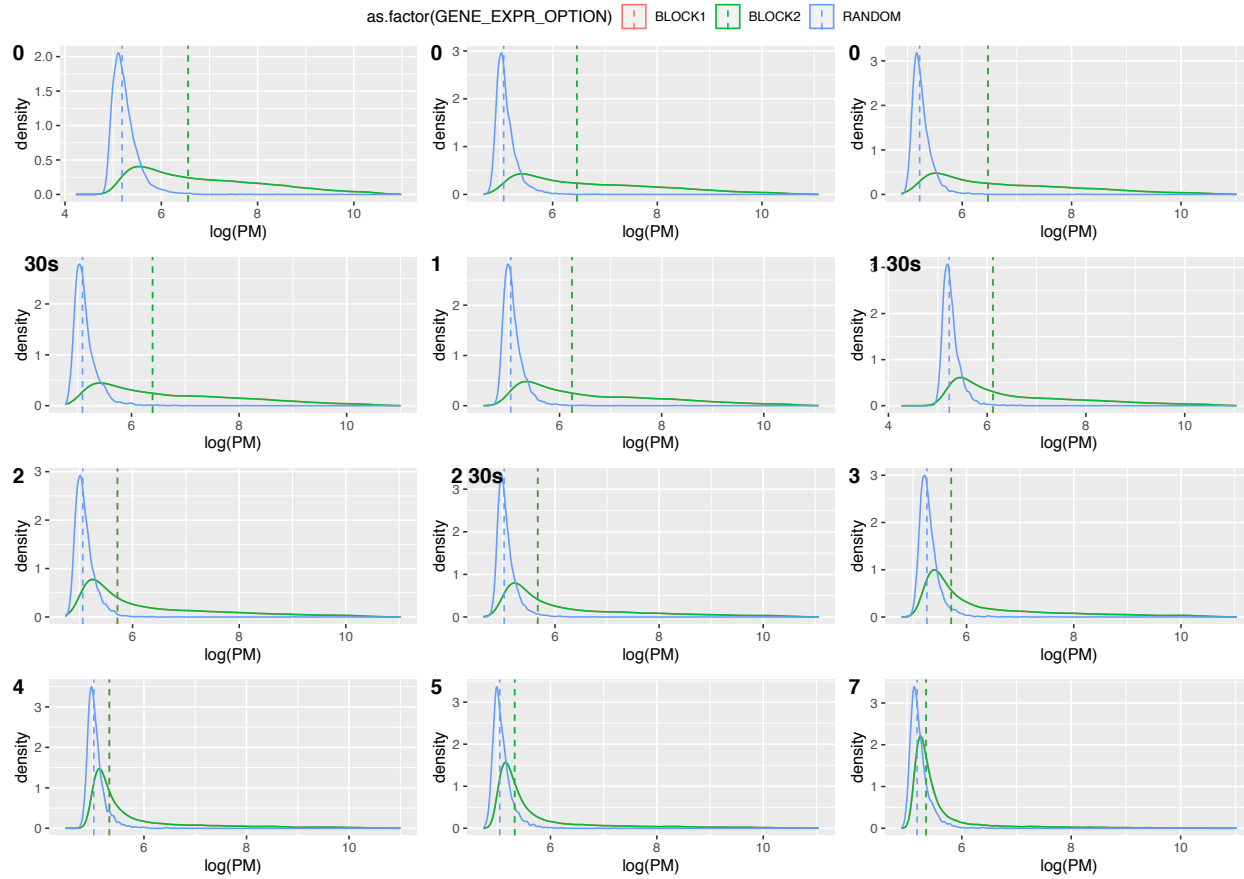


FIGURE 5.3 – Densité des intensités des sondes spécifiques (vert et rouge) et non spécifiques (bleu). Les valeurs de densité sont données pour chaque temps de mesure au taux de croissance $0.63h^{-1}$. Les intensités sont données en abscisse sur une échelle logarithmique

5.2 Article de revus

Connecting mRNA degradation, metabolism and cellular processes: a cornerstone in understanding *Escherichia coli* physiology

Charlotte Roux^{1,2}, Thibault Etienne³, Eliane Hajnsdorf², Delphine Ropers³, AJ Carpousis⁴, Muriel Cocaign-Bousquet^{1 #}, Laurence Girbal^{1 #*}

¹ TBI, Université de Toulouse, CNRS, INRAE, INSA, Toulouse, France

² UMR8261, CNRS, Université de Paris, Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, 75005 Paris, France

³ Univ. Grenoble Alpes, Inria, 38000 Grenoble, France

⁴

Key words:

Co-last authors

*Corresponding author

Introduction

The process of mRNA degradation is part of the mRNA life cycle. Degradation of mRNA participates in the elimination of defective or superfluous transcripts and contributes to the recycling of the nucleotide pool in the cell. It also takes part in the regulation of protein synthesis since intracellular mRNA concentration is a balance between transcription and degradation (dilution by growth is generally neglected during cell proliferation). From an evolution point of view, mRNA instability is a more suitable way to regulate protein levels than protein instability because of the lower cost of mRNA synthesis compared to protein synthesis (Pérez-Ortín et al., 2019). In the model bacterium *Escherichia coli*, large-scale studies have shown that mRNA degradation varies significantly with the growth rate (Esquerré et al., 2014). At low growth rates and also in stationary phase, mRNA are globally stabilized (Chen et al., 2015 ; Dressaire et al., 2018 ; Morin et al., 2020 in press). Such omics data confirm the growth rate dependency of the stability of particular mRNAs such as *ompA*, *cat* and *fumA* described in earlier works (Nilsson et al., 1984 ; Lin et al., 2012). These findings highlight a cross talk between the regulation of mRNA degradation and the growth state and thus link mRNA degradation to metabolism and cell physiology.

The molecular process and components of the mRNA degradation machinery are well described in *E. coli* (Carpousis et al., 2002 ; Andrade et al., 2009; Gordon et al., 2017). mRNA degradation is initiated by an RNase E- or RNase III-mediated endonucleolytic cleavage followed by the conversion of cleaved fragments to mononucleotides by 3'-exoribonucleases. Regulations of specific metabolic pathways but also large-scale regulatory circuits of metabolism involved in *E. coli* adaptation to environmental conditions are also well known. They are based on more or less complex interactions between metabolites, enzymes and transcriptional regulators (Chubukov et al., 2014). Generally, mRNA degradation regulations are not taken into account in study of *E. coli* physiology although growing evidence shows implication of mRNA degradation regulation in control of specific metabolic pathways. We believe that to fully understand *E. coli* physiology it is now necessary to consider mRNA degradation as an indispensable regulatory level of metabolism and cellular processes and to answer the following question. Are mRNA degradation regulations coordinated or not with regulations of other cellular processes and metabolism and how? This automatically leads to the next challenge on how to include mRNA degradation in system biology approaches to provide a complete understanding of cell systems. In this review, our ambition is to place the

process of mRNA degradation at the core of metabolic regulations and cell adaptation to provide an integrated vision of *E. coli* physiology. Although virulence, cellular morphology, biofilm formation and motility are part of *E. coli* physiology, they will not be discussed here. We will focus on the role of mRNA degradation regulations in normal and stressful growth environments to show (i) that mRNA degradation is in close interaction with other intracellular processes also involved in gene expression regulation, *ie.* transcription, translation and subcellular organisation; (ii) that metabolism acts on the process of mRNA degradation and *vice versa*; (iii) how important it is to integrate mRNA degradation in mathematical models to decipher metabolic network regulation in systems biology and (iv) finally that knowledge on the control of mRNA degradation can be applied to enhance syntheses of proteins of interest in new strategies of synthetic biology.

1. Coupling between mRNA degradation and other cellular processes

mRNA degradation can be co-transcriptional and can counterbalance transcription

The mRNA concentration depends on both processes of transcription and mRNA degradation. Transcription and mRNA degradation are generally assumed to act independently. However, several works show that these two processes are linked to each other.

A coordinated decrease in mRNA stability and transcription is shown when a higher susceptibility of *lacZ* mRNA to be degraded leads to premature transcription termination (Pedersen et al., 2019 ; Eriksen et al., 2017). A co-transcriptional degradation of mRNA is also described in *E. coli* for the *trp* and *lac* operons, with a 5'-end degradation of the transcript before the synthesis of its 3' extremity (Morikawa and Imamoto, 1969 ; Morse et al., 1969 ; Cannistraro and Kennell, 1985). Cleavage within the *ahpCF* operon was also recently proposed to be co-transcriptional (Lacoux et al., 2020). Recently the co-transcriptional mRNA degradation of 263 transcripts was modelled as a stochastic event, suggesting that this could be widespread in the cell (Chen et al., 2015) (**Figure 1**). This idea conflicts with earlier observation of mRNA mostly full-sized in the cell (Montero Llopis et al., 2010) and the compartmentalisation of DNA, ribosomes and RNases in the cell (Mackie 2013). In fact, co-transcriptional mRNA degradation is more likely to be operated by free RNases in the nucleoid that could be different from the anchored degradosome in the membrane (see below section "Spatial distribution influences mRNA degradation")

Modelling and experimental works also propose that mRNA concentration is a key determinant of its stability, by showing an inverse relationship between mRNA half-lives and concentrations (Bernstein et al., 2002 ; Mohanty and Kushner, 2003 ; Esquerré et al., 2015 ; Nouaille et al., 2017 ; Dressaire et al., 2018). Note that neither the RNA sequence nor the expression of degradosome machinery components are involved in this correlation (Nouaille et al., 2017). This inverse correlation was proposed to reflect the counter balanced controls of mRNA stability and transcription (Mohanty and Kushner, 2003, Nouaille et al., 2017). Inverse relationship between mRNA half-lives and concentrations has been attributed to a higher probability of RNase to bind to concentrated mRNA (Nouaille et al., 2017). This phenomenon was recently modelled through competition effects between mRNAs for binding to the same limited pool of RNase E (Etienne et al JTB 2020). However, the precise mechanism and its physiological implication are still unknown and require further experimentations to be fully elucidated. A recent study shows that for 30% of *E. coli* mRNAs the processing frequency peaks when RNAs accumulate (Lacoux et al., 2020). We can speculate that the higher stability of mRNAs at low concentration could be an economic strategy of mRNA turnover in limiting growth conditions, by permitting translation of highly stable mRNA despite their low abundance. In this case, it could avoid total degradation of mRNA that would lead to inability of restoring growth in better environmental conditions.

Ribosome protection drives the interaction between degradation and translation for mRNAs at the individual scale but not at the genome scale

First, a substantial amount of work suggests that ribosomes could interact with the RNA degradosome. Indeed, RNase E co-purifies with 16S and 23S rRNAs (Bassarab et al., 1998), directly interacts with the 50S subunit ribosomal L4 protein (Singh et al., 2009) and can form *in vivo* a stable complex with an isolated ribosome (Tsai et al., 2012). The other members of the degradosome, PNPase, RhlB and enolase can also bind ribosomal proteins of both the 30S and 50S subunits (Butland et al., 2005). Singh et al. showed that interaction of the 50S subunit ribosomal L4 protein with RNase E inhibits *in vitro* RNase E endonucleolytic activity and is associated *in vivo* with increased levels of many mRNAs (Singh et al., 2009). In contrast, polysomes could facilitate the degradation of mRNAs, by acting as antennae to enhance the rates of capture of degradosomes (Tsai et al., 2012).

Translation depends on both translation initiation and elongation efficiencies. The initiation step is tightly coupled to mRNA stability (see the review of Kaberdin and Bläsi, 2006). A global study of mRNA stability parameters using synthetic library in *E. coli* reveals a

significant correlation between mRNA stability and the ribosome binding site (RBS) strength (Kosuri et al., 2013). Indeed, the stability of *lacZ* and *ompA* mRNAs are decreased with a mutated less-efficient RBS and restored in the *rne-50* inactivation mutant, suggesting that binding the 30S ribosome subunit to the Shine Dalgarno sequence of an mRNA has a protective effect against RNase E cleavage (Yarchuk et al., 1992 ; Wagner et al., 1994, Arnolds et al., 1998 ; Eriksen et al., 2017) (**Figure 1**). In addition, ribosomes can mask an RNA feature, 10-20 nucleotides upstream of the processing site, which contributes to the RNase E cleavage efficiency (Braun et al 1998). Ribosome protection is also suggested to explain the relationships between translation initiation and mRNA stability using a library of synthetic translation initiation regions (Cambray et al., 2018). Concerning the importance of the 5'-UTR segment of an mRNA for its stability, the particular case of leaderless mRNA is interesting. One study shows that the low-translated *tetR* leaderless mRNA is unstable due to the lack of 5'-UTR (Baumeister et al., 1991).

Although translation initiation is important to regulate the stability of an mRNA, elongation also plays a significant role (**Figure 1**). Density of translating ribosome is positively correlated with mRNA stability as exemplified for particular mRNAs (*IS10*, *lacZ*, *bla*, *ompA* and *thrS* mRNAs (Jain and Kleckner, 1993 ; Deana et al., 2005 ; Iost and Dreyfus, 1995 ; Nogueira et al., 2001 ; Nilsson et al., 1987 ; Arnold et al., 1998). High density of ribosomes can thus protect mRNA from degradation. Focusing on the first 45 codons of *lacZ*, Pedersen and co-workers (Pedersen et al., 2019) recently model that slow ribosomes at the beginning of the coding region stabilize the mRNA molecule. Under the assumption that the presence of slow ribosomes will increase the local ribosome density, this work supports the idea of the protective effect of ribosome density. The corollary is also true: ribosome-free coding regions are more instable, nevertheless a ribosome-free segment in the coding sequence does not systematically induce mRNA destabilization (Nilsson et al., 1987). These studies highlighting that (i) the positive correlation between ribosome density and mRNA stability can concern only part of the coding sequence (for instance from codons 20 to 45 in the case of *lacZ* (Pedersen et al., 2011) and (ii) the position of the ribosome-free segment along the coding sequence determines the effect on mRNA stability, show that the rules of mRNA degradation regulation are difficult to predict. Recent studies at the omics scale challenged the idea of mRNA protection by ribosome density. Two genome-wide studies report a positive correlation between codon optimality and mRNA stability (Boël et al., 2016 ; Esquerré et al., 2015) meaning that mRNAs encoded by less optimized codons tend to be degraded more rapidly. If we consider that less optimized codons are associated with slow translating ribosomes and thus higher ribosome density, these

studies do not confirm at the large-scale the positive correlation between ribosome density and mRNA stability. These results do not exclude that particular mRNAs could behave differently than the average of the mRNA population. The phenomenon of lower stability of mRNA related to slow ribosomes was recently named “slowness-mediated decay” (Rak et al., 2018).

Regarding the underlying mechanisms, the mechanism responsible for the slowness-mediated decay is not known. One might assume that it is related to the ability of ribosomes to recruit the RNA degradosome (Tsai et al., 2012). Most of the literature describes how ribosomes protect mRNA from degradation by interfering with the endonucleolytic cleavage by RNase E. In the current model for mRNA decay in *E. coli*, RNase E is the first and limiting step, cleaving mRNA by either 5'-end independent (direct) entry or by 5'-end dependent pathways (Kushner et al., 2002 ; Clarke et al., 2014). Recent works suggest that translation could hinder both 5'-end dependent and independent recognitions and so mRNA cleavage by RNase E, because ribosomes and RNase E compete to access the transcript in the translation initiation region and/or in the ORF (Richards and Belasco, 2012 ; Richards and Belasco, 2019). Altogether, these findings suggest that (1) the presence of ribosome in the translation initiation region can impede 5'-end mRNA recognition by RNase E and so directly inhibits the 5'-end dependent degradation pathway and (2) high ribosome traffic in the ORF can hinder linear scanning of RNase E from the 5'-end (5'-end-dependent pathway) and also recognition of internal cleavage sites (direct pathway) (**Figure 1**).

During translation elongation, a particular sequence of the nascent polypeptide can promote ribosome pausing, leading to the mRNA cleavage at the A-site of the paused ribosome that creates a non-stop mRNA, lacking stop codon (Sunohara et al., 2004 J. Biol. Chem ; Sunohara et al., 2004, RNA). This non-stop codon is recognized by the tmRNA system for trans-translation mechanism initiation, leading to ribosome rescue and mRNA and protein quality controls (Pedersen et al., 2003 ; Hayes and Sauer, 2003) (**Figure 1**). The current hypothesis is that the ribosome cleaves the mRNA its-self, via its proteins or rRNA. Supporting this idea is the endonuclease activity of the S16 ribosomal protein (Oberto et al., 1996), and the fact that neither the three major endoribonucleases of *E. coli* RNase E, G and III (Sunohara et al., 2004 J. Biol. Chem) nor tmRNA itself (Hayes and Sauer, 2003) seem to be required for this cleavage.

Spatial distribution influences mRNA degradation

The reason why RNase E and the RNA degradosome are membrane-localized is one of the most intriguing questions when studying the dynamics of mRNA degradation. In Gram-positive bacteria (*Bacillus subtilis* and *Staphylococcus aureus*), RNase Y, the functional RNase E homolog, is also membrane-bound (Hunt et al., 2006 ; Shahbadian et al., 2009), although the two RNases share no phylogenetic common ancestor (Lehnik-Habrink et al., 2011). This highlights a potential strong biological relevance of membrane-anchorage of the RNA degradation machinery in bacteria although in the Gram-negative bacterium *Caulobacter crescentus*, the RNase E is cytoplasmic and co-localizes with the nucleoid (Montero Llopis et al., 2010 ; Bayas et al., 2018).

RNase E rapidly diffuses into the *E. coli* inner membrane and forms small short-lived foci containing several molecules of RNase E. Depleting mRNA by rifampicin treatment decreases the number of foci indicating that RNase E foci formation requires direct interaction with its mRNA substrate (Khemici et al., 2008 ; Strahl et al., 2015). mRNA molecules thus seem to be synthesized and degraded in two different subcellular localizations, in the nucleoid and at the cellular membrane, suggesting a displacement of mRNA in the intracellular cytoplasm. Detachment of the RNase E from the membrane by deleting its MTS anchor leads to a global slowdown of mRNA degradation despite the formation of an active cytoplasmic RNA degradosome (Moffitt et al., 2016 ; Hadjeras et al., 2019). Moreover, artificial membrane-localization of mRNA induces their destabilization (Moffitt et al., 2016). Interestingly, under anaerobic stress conditions, the glycolytic enzyme enolase of the degradosome alters the subcellular localization of RNase E from the membrane to the cytosol (Murashko and Lin-Chao, 2017), suggesting a metabolic control of RNase E localization under stress conditions. RhlB, PNPase and the poly(A) polymerase PAP I are enriched at the membrane, and this localization depends on the membrane anchoring of RNase E (Moffitt et al., 2016 ; Strahl et al., 2015). PNPase also forms a cytosolic complex with RNase PH (Wong et al., 2013) and with RhlB (Lin et al., 2005) but the physiological roles of these degradosome-independent interactions are not fully understood yet. RNase P is also presumed to be localized at *E. coli* membrane but the significance of this localization remains unknown (Mickaz et al., 1991).

First studies concerning intracellular mRNA localization in *E. coli* showed that mRNA molecules are not fixed in space but move constantly with a subdiffusive behaviour, a random movement constrained by variations of cytoplasm density (Golding and Cox, 2004 ; Golding and Cox, 2006). Localization of most of mRNA molecules in translation appears to be outside the nucleoid due to a density exclusion phenomenon of ribosome-rich mRNAs (**Figure 1**). The more an mRNA molecule is bound by ribosomes the more it is modelled to be excluded from

the nucleoid, creating a flux of mRNA from the nucleoid to the membrane. To compensate there is an inverse flux of free ribosomes, which are small enough to penetrate the highly dense nucleoid and reach the translation initiation sites (Castellana et al., 2016). This circular flux appears to depend on the RNA life cycle of constant cytoplasmic synthesis and membrane-localized degradation (Figure 1). The density exclusion model from the nucleoid is consistent with the experimental finding of mRNA enrichment at the periphery of the cells along the long cell axis (Valencia-Burton et al., 2007 ; Valencia-Burton et al., 2009). The spatial localization of mRNA is independent of the gene loci position (Moffitt et al., 2016) but rather differs on their function, and more precisely on the cellular compartment of the encoded proteins (Valencia-Burton and al., 2007 ; Nevo-Dinur et al., 2011 ; Kannaiah et al., 2019). For inner-membrane proteins, the corresponding transcripts are enriched near the membrane in a translation dependent manner, and this also applies to operons as shown for the *bgl* operon with all cistrons co-localized at the membrane (Nevo-Dinur et al., 2011 ; Moffitt et al., 2016 ; Kannaiah et al., 2019).

These results indicate that there is a control of the spatial distribution of the mRNA molecules and the components of the degradation machinery in *E. coli* despite the lack of membrane-bound compartments like organelles in bacteria. The membrane-attachment of RNase E and RNA degradosome protects free transcripts from premature degradation giving them time to be translated during their trip from the nucleoid to the pole (Figure 1).

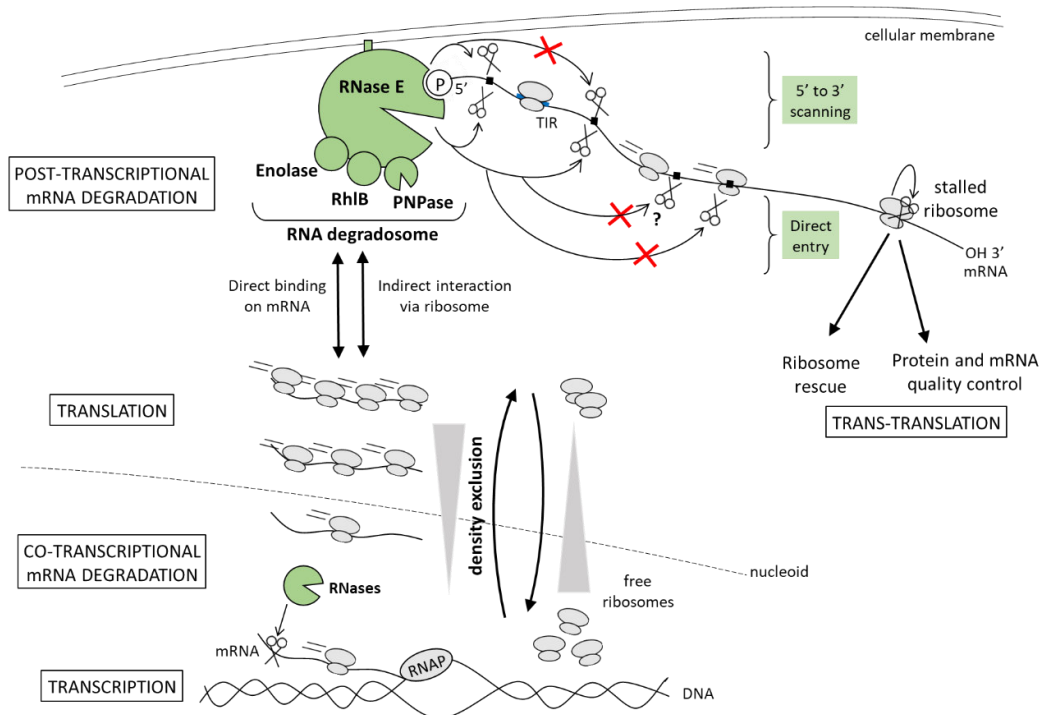


Figure 1: Schematic of the coordination between mRNA degradation and transcription, translation, and cellular localization in *E. coli*. Small black boxes on mRNA represent RNase E cleavage sites, the TIR sequence is in blue. The two lines at the left of ribosomes indicate translating ribosomes along the mRNA. Thin arrows ending with a scissor indicate cleavage activity by RNases or ribosomes (in the case of trans-translation). Large arrows mean actions specified by the text written right next to it.

2. mRNA degradation and metabolism

The environments in which cells grow often change rapidly. Cells evolved their metabolism for adjusting their growth to environmental modifications. Growth is fast when nutritional conditions are optimised while it is reduced until no growth at all in stationary phase under nutritional deprivation or physicochemical stress. Here we will focus on the interconnection between mRNA degradation and changes in metabolism associated to modifications of growth conditions (*e.g.* in fast growing cells *versus* cells under nutritional/physicochemical stress). These interactions are widespread over the metabolism, more particularly in the central carbon and energetic metabolisms, but also in other metabolisms such amino acid, ion and other stress-related metabolisms. Interactions between mRNA degradation and metabolism are bidirectional: a control of the activity of the mRNA degradation machinery by metabolism and more frequently a control of metabolism by mRNA degradation. All these interactions are shortly described in Table 1 and presented in more detail in the following subsections.

Metabolism regulates the activity of the mRNA degradation machinery

Despite the major role of RNase E in the mRNA degradation process (Carpousis et al., 2007) and its postulated role in regulating the ribosome biosynthesis network (Thiele et al., 2010), almost no clear experimental evidence of direct regulation of RNase E by actors of the metabolism is available in the literature. The only report is an *in vitro* inhibition of RNase E by glucosamine-6-phosphate, a precursor of bacterial cell envelope peptidoglycans and lipopolysaccharides, (MARDLE Bioch Biophy Rep2020). The apparent constant RNase E activity in *E. coli* cells could be due to the strong autoregulation of RNase E at the post-transcriptional level that buffers any change in RNase E stability and activity without a tight control by metabolism (Leroy et al., 2002 ; Hadjeras et al., 2019). For the other RNA degradosome components, PNPase, enolase, and RhlB and for RppH and PAP I, higher mRNA concentration is measured with an increasing growth rate (Esquerré et al., 2014). For all of them

plus the endoribonuclease toxins and the decapping enzyme ApaH, metabolic elements regulating each enzyme have been identified in central carbon metabolism, nucleotide and energy metabolism, amino acid metabolism and in stress condition, illustrating how the activity of the RNA degradation machinery can be controlled by the metabolism (**Table 1**, arrows in green).

The exoribonuclease activity of PNPase is negatively regulated by citrate, a tricarboxylic acid (TCA) cycle metabolite, at physiological concentrations (Nurmohamed et al., 2011). It is interesting to note that intracellular citrate concentrations vary with the carbon sources used for *E. coli* growth, suggesting that this control is important *in vivo* (Bennett et al., 2009). Moreover, attachment of PNPase to the RNA degradosome seems to be necessary for this action, since its detachment leads to decoupling the two halves of the TCA cycle (Nurmohamed et al., 2011). This citrate-mediated inhibition of PNPase, linking mRNA degradation and central carbon metabolism, may be widespread across all living organisms, since citrate was predicted to bind PNPase in human, bacterial *Streptomyces antibioticus* and archaeal *Sulfolobus solfataricus* (Stone et al., 2017).

In addition, PNPase activity is directly regulated by nucleotides. Indeed, PNPase activity is modulated *in vitro* by cyclic di-GMP (Tuckerman et al., 2011) and is allosterically inhibited by ATP *in vitro* for both its phosphorolytic and polymerization activities (Del Favero et al., 2008). This finding indicates that PNPase acts mainly at low energy charge *in vivo* and connects mRNA turnover to the cellular energy state. The current model for the cytosolic RNase E-independent RhlB/PNPase complex proposes that RhlB hydrolyses ATP to unwind mRNA secondary structures and to favour access and activity of PNPase on targeted mRNA due to the decrease of local ATP concentration (Carpousis et al., 2007). The finding that PNPase is involved in the *ompA* transcript degradation in stationary phase (Andrade et al., 2006) could also reflect the ATP inhibition of PNPase. In the same idea, the hydrolytic degradation of mRNA by RNase II, which competes PNPase, is energetically more expensive than the phosphorolytic degradation by PNPase (Deutscher, 1993 ; Deutscher and Reuven, 1993), and may be the reason why PNPase is favoured at low temperature, a low energy state (Phadtare et al., 2012).

Enolase is an abundant dual-function enzyme in *E. coli*, responsible for the conversion of 2-phospho-D-glycerate into phosphoenolpyruvate during glycolysis and also one of the four major components of the RNA degradosome in *E. coli* (Py et al., 1996 ; Liou et al., 2002). Other

bacterial organisms display such interactions between RNA degradosome and metabolic enzymes, such as aconitase in *Caulobacter crescentus* (Hardwicks et al., 2011) and enolase and phosphofructokinase in *B. subtilis* (Commichau et al., 2009 ; Newman et al., 2012). The association of metabolic enzymes into complexes dedicated to mRNA degradation in both Gram positive and negative bacteria likely highlights a certain biological relevance. Enolase activity and recruitment in the *E. coli* RNA degradosome appears to be required for normal mRNA turnover *in vivo* (Bernstein et al., 2004). Mutation in the *eno* gene encoding enolase stabilizes many mRNAs of the glycolysis and TCA cycle and thus involved in energy generating pathways (Bernstein et al., 2004), suggesting a potential role of sensor for enolase, connecting mRNA degradation to the energetic state of the cell (Carpousis, 2007). Moreover, the recruitment of enolase by RNase E in the degradosome is required to respond to metabolic stress (accumulation of glucose-6-phosphate; Morita et al., 2004) and to anaerobic condition (Murashko and Lin-Chao, 2017). In anaerobic condition, stability of *eno* mRNA is enhanced by the reduction of RNase G abundance and its translation favoured following RNase III-mediated cleavage (Lee et al., 2019). In turn enolase controls RNase E subcellular localization in anaerobic condition and therefore RNase E activity (Murashko and Lin-Chao, 2017). This last result is one rare example of regulation of RNase E activity in relationship with metabolism. Altogether, although the precise role of enolase presence in the RNA degradosome of *E. coli* has not been fully elucidated, it remains clear that this metabolic enzyme has a precise and significant role in mRNA turnover, and raises the question of a possible control of mRNA degradation by the carbon and energetic metabolisms.

RppH is a central enzyme of the mRNA degradation process. It removes the orthophosphate from dephosphorylated intermediates to generate monophosphorylated transcripts, ready to be degraded by RNase E through the 5'-end dependent pathway (Luciano et al., 2017). It was recently shown that the DapF enzyme, involved in lysine biosynthesis, directly interacts with RppH and accelerates its activity of phosphate decapping (Lee et al., 2004 ; Gao et al., 2018). For instance, disrupting RppH-DapF interaction stabilizes *yeiP* mRNA. The regulation of the RNA 5'-pyrophosphohydrolase activity by DapF is another evidence of the interaction between a metabolic enzyme and the RNA degradation pathway. Moreover, in stationary phase, the presence of methylated dinucleoside polyphosphate caps protecting RNAs from RppH cleavage was proposed to contribute to mRNA stabilization in this condition (Hudeček et al., 2020).

The poly(A) polymerase PAP I is responsible for the addition of a poly(A) tail to the 3'-end of RNA using ATP as donor in order to facilitate RNA degradation by exoribonucleases (Regnier et al., 2009). The *pcnB* gene encoding PAP I is under the control of five distinct promoters (Jasiecki et al., 2006 Plasmid). The activity of at least two of them is inversely correlated to *E. coli* growth rate, with higher PAP I abundance and RNA poly-adenylated in slow growing cells (Jasiecki et al., FEMS Microbiol Lett ; Jasiecki et al., 2003). Moreover, the activity of the major *pcnB* promoter is drastically reduced in stationary phase (Nadratowska-Wesolowska et al., 2010). These apparently opposite regulations of the different promoters could be the solution to avoid a waste of ATP in form of poly(A) tail in non-growing cells but still provide sufficient pools of nucleotides for *de novo* mRNA synthesis in slow growing cells via degradation of poly-adenylated mRNA. The PAP I synthesis seems thus in close connection with the energetic and nucleotide metabolisms.

A little aside concerns the activation of the endoribonuclease toxins under stress condition. For instance, when cells encounter amino acid starvation, (p)ppGpp enhances the proteolysis of the labile antitoxin freeing the endoribonuclease toxin that will cleave a large part of bulk mRNA and provoke inhibition of essential cellular processes and growth arrest (Sauert et al., Tian et al., 2016). In contrast, example of enzyme inactivation was reported under oxidative stress in cadmium-treated cells when the decapping enzyme ApaH is inactivated by disulfide bond formation (Luciano et al., 2019).

Changes in mRNA degradation regulate central carbon and energetic metabolisms

In the following section, we will show how metabolic pathways of the central carbon and energetic metabolisms are controlled by mRNA degradation regulation (Table 1 arrows in blue). This direction of interactions (mRNA degradation regulates metabolism) seems more frequent than the other direction (metabolism regulates mRNA degradation).

E. coli is able to uptake and utilize different carbon sources to produce precursor metabolites and energy necessary for cell formation. Carbon source utilization can depend on modulation of mRNA stability in *E. coli*. RNA degradosome formation is not necessary for glucose utilization but is crucial for utilization of acids like fumarate and succinate (Tamura et al., 2013). The galactose utilization is controlled by stability regulation of *galK* transcript encoding galactokinase. *galK* mRNA is destabilized by binding of the small RNA (sRNA) *spot42*. *spot42* is encoded by *spf* gene and its expression varies with the carbon source.

Destabilization of *galK* mRNA discoordinates the whole *gal* operon and inhibits galactose utilization (Møller et al., 2002). The deletion of *spf* only has a slight effect on *E. coli* growth but its overexpression strongly impairs cell growth. A *spf* deleted strain is unviable on a minimum media with succinate as sole carbon source (Rice and Dahlberg, 1982).

One very well studied connection between mRNA degradation and carbon utilisation is the retro-control of the sugar uptake expression by mRNA degradation. This regulation is mediated by the toxic accumulation of sugar phosphate when the carbon source entry is too fast and more particularly in *E. coli*, the glucose-6-phosphate (G6P) accumulation. The *ptsG* gene encodes the main glucose transporter of *E. coli*, EIICB^{Glc}, and its expression is regulated in response to glycolytic flux through G6P (Kimata et al., 2001 ; Morita et al., 2003). G6P accumulation leads to *ptsG* mRNA destabilization. Accumulation of G6P activates SgrS sRNA synthesis that base pairs with *ptsG* mRNA in the RBS sequence to repress translation and targets the transcript to RNase E-mediated degradation (Vanderpool et al., 2004 ; Kawamoto et al., 2006). This *ptsG* downregulation impedes glucose import into the cell and thus prevents G6P accumulation. The SgrS sRNA can also bind to *manX* mRNA encoding the mannose transporter that can be used as a secondary transporter for glucose uptake and stimulates degradation of the *manXYZ* transcript by the RNA degradosome (Rice and Vanderpool, 2011). Destabilization of *manXYZ* mRNA seems to be also regulated by the activity of CsrA (Esquerré et al., 2016).

Glycolysis and TCA cycle are strongly regulated by changes in mRNA degradation. A large-scale study of mRNA stability shows that in exponential growth the levels of mRNAs involved in the TCA cycle are controlled by changes in transcript stability (Esquerré et al., 2014). Several components of the mRNA degradation machinery (RNase E, RNase III, RNase G, enolase and PNPase) are involved in regulating glycolysis and TCA cycle. The defective function of RNase E in a double *rne deaD* mutant strain is responsible for the inability of the strain to grow on minimum media with glucose as sole carbon source while growth is allowed with pyruvate or pyruvate/arabinose mix (Tamura et al., 2012). The connection between glycolysis and TCA cycle is broken preventing glycolysis products to enter the TCA cycle. There is a drastic reduction of the phosphoenolpyruvate carboxylase expression, the glycolytic enzyme responsible for the conversion of phosphoenolpyruvate into oxaloacetic acid (Tamura et al., 2013). In the same idea, loss of the RNase E scaffold region reduces transcript abundance of glycolytic enzymes (Lee et al., 2002). In addition the deletion of the *rng* gene encoding the RNase E homolog RNase G, leads to increased steady-state levels of several glycolytic enzyme transcripts, including the *eno* gene encoding enolase (Lee et al., 2002 ; Gordon et al., 2017).

RNase III, the second major endoribonuclease of *E. coli* after RNase E, with a double-strand specificity (Altuvi et al., 2018), also affects abundance of glycolytic mRNA, like *ptsA* transcript encoding the sugar phosphotransferase system (Sim et al., 2010). A $\Delta rhlB$ *E. coli* strain as well as a Δpnp strain show strong mRNA expression changes for TCA cycle enzymes (Zhou et al., 2013). More precisely, in stationary phase a Δpnp strain shows a global destabilization of mRNAs except for some mRNAs related to glycolysis and TCA cycle (Bernstein et al., 2004 ; Dressaire et al., 2018). The well-studied carbon storage regulator (CSR) system is also a major regulator of the central carbon metabolism. The RNA-binding CsrA protein is the central actor and can be sequestered by two sRNA, CsrB and CsrC (Liu et al., 1997 ; Weilbacher et al., 2003). The expression of these two sRNAs is mediated by the BarA/UvrY two component system (Gudapaty et al., 2001 ; Suzuki et al., 2002) and repressed by the CRP/cAMP system linking the catabolite repression to the CSR system (Pannuri et al., 2016). CsrA is reported to favour glycolysis by stabilizing mRNAs like *pfkA* and *tpiA* and repress gluconeogenesis by destabilizing mRNAs such as *maeA*, *maeB*, and *pck* (Esquerré et al., 2016 ; Potts et al., 2017).

The acetate metabolism is regulated by the sRNA SdhX (first called RybD) that is cleaved from the 3'-UTR of the TCA operon *sdhCDAB-sucABCD* transcript by RNase E (or by RNase III in a slighter way) (De Mets et al., 2019 ; Miyakoshi et al., 2018). SdhX thus links the TCA cycle and acetate pathways. SdhX negatively regulates the *ackA* gene encoding the acetate kinase that converts acetyl phosphate into acetate and inversely. A mutation in SdhX increases acetate excretion in the medium during the overflow metabolism but also impedes acetate utilization when acetate is used as sole carbon source (De Mets et al., 2019).

Glycogen accumulation is the main storage form of glucose in *E. coli*. Genes essential for glycogen biosynthesis are *glgB*, *glgC* and *glgA* encoding glycogen branching enzyme, ADP-glucose pyrophosphorylase and glycogen synthase, respectively. It is well-known that the CSR system represses glycogen metabolism (Romeo et al., 1993). CsrA binds the *glgCAP* leader transcript and promotes its degradation (Backer et al., 2002 ; Liu et al., 1995). The activity of CsrA also decreases *glgB* mRNA stability and expression (Esquerré et al., 2016).

mRNA degradation regulation also contributes to other metabolic responses

It is interesting to note that RhlB/PNPase complex is linked to *E. coli* amino-acid cysteine biosynthesis. An RhlB^{P238L} mutation, allowing interaction of RhlB with RNase E but

not with PNPase, leads to *cysB* down-regulation, and alters by extension the transcript abundance of cysteine regulon genes (Tseng et al., 2015). This study provides the first link between RhlB-PNPase complex and cysteine biosynthesis.

Iron homeostasis genes in *E. coli* are under the control of Fur (Ferric uptake regulator), a Fe²⁺ repressor targeting several genes including uptake iron genes that become inactive during iron depletion. When iron is depleted, RyhB sRNA expression increases and down-regulates mRNA levels of genes of the Fur regulon including an iron containing superoxide dismutase gene (*sodB*) and some genes of the TCA cycle (Massé et al., 2002 ; Massé et al., 2005). Base pairing of RyhB with its target mRNA facilitates their degradation. Recent reports show that inactivation of *pcnB* gene encoding PAP I destabilizes RyhB and impairs *sodB* and *sdhCDAB* repression (Sinha et al., 2018). This finding suggests that RyhB stability and thus function are controlled by polyadenylation. Expression of genes involved in iron metabolism is also repressed by CsrA but more likely indirectly after changes in the intracellular iron content (Potts et al., 2017). A tight control of intracellular iron content is required to avoid production in the presence of oxygen of highly reactive and damaging reactive oxygen species. In addition, protection against oxidative stress involves PNPase. Deletion of *pnp* gene leads to hypersensitivity to H₂O₂, a common reactive oxygen species (Wu et al., 2009). H₂O₂ leads to an increase of damaged 8-hydroxyguanine RNA and PNPase is able to bind such an oxidized form of RNA to favour growth in oxidative condition (Hayakawa et al., 2001 ; Wu et al., 2009). It is interesting to note that this PNPase protective effect does not require RhlB or RNase E association, and is thus an intrinsic property of PNPase (Wu et al., 2009). Consistent with this finding, a truncated RNase E lacking its scaffold region is hyper resistant to oxidative stress, probably by releasing PNPase from the RNA degradosome (Nanbu-Wakao et al., 2000). Under oxidative stress in cadmium-treated cells, the inactivation of the decapping enzyme ApaH stabilizes capped *trxB* mRNA that encodes thioredoxin reductase. Increase in thioredoxin reductase activity could contribute to respond to disulfide stress (Luciano et al., 2019). Resistance to osmotic stress is related to stabilization of *betT* mRNA encoding an influx transporter of choline. Cleavage of *betT* message is performed by RNase III whose activity is decreased under conditions of osmotic stress (Sim et al., 2014).

The metabolic adaptation to temperature involves many components of mRNA degradation machinery, such as RNase III, PNPase, RNase II, RNase R, helicases, RNA chaperone. PNPase is essential in *E. coli* for growth at low temperature and is induced during

the acclimation phase (Jones et al., 1987 ; Zangrossi et al., 2000). To better allow its expression at low temperature, RNase III activity which negatively controls PNPase stability, is reduced by interacting with YmdB (Kim et al. 2008). A cold shock causes growth stop and induces production of cold shock proteins. mRNAs of cold shock proteins are selectively destabilized by PNPase at the end of the acclimation phase to promote cell growth reinitiation (Yamanaka et al., 2001). PNPase ribonuclease activity but not its polymerization activity is required for PNPase protective effect against cold (Awano et al., 2008). It is interesting to note that RNase II can complement the loss of PNPase in such cold conditions (Awano et al., 2008), but not RNase R, which is nevertheless also induced at low temperature (Cairrão et al., 2003). However, during the acclimation phase after shift to cold temperature, RNase R is part of a two-component surveillance system involving cold shock proteins that allows correct global mRNA degradation (Zhang et al., 2018). RhlB may also play a role in response to cold-shock via its RNA helicase function. It was suggested that at high temperature mRNAs were easily unwound by thermal energy only, and that RhlB helicase function was more relevant at lower temperature (Iost et al., 2013). However, at very low temperature RhlB is replaced by another DEAD-box protein, DeaD, which is thought to be more efficient for helicase activity (Prud'homme-Généreux et al., 2004). Moreover, deletion of *rppH* gene encoding RppH enhanced cellular growth under low temperature conditions (Choi et al., 2017). At the opposite high temperature, a lower concentration of the RNA chaperone CspC reduces transcript stability of heat shock genes (*groEL*, *dnaK*, *hslV* and *htpG*) and allows *E. coli* adaptation to elevated temperatures (Shenhar et al., 2009).

Table 1: Cross talk between mRNA degradation and metabolism in *E. coli*. This table presents examples of interconnections between mRNA degradation and metabolic pathways or stress responses. A green arrow shows a regulation of an mRNA degradation actor by metabolic elements, and a blue arrow a regulation of a metabolic pathway by mRNA degradation actor(s). CSP: cold shock proteins. *non specifically targeted mRNAs.

Metabolism		Direction	Actor(s) of the degradation	Type of interaction	References
Pathway	Element (mRNA or metabolite)				
Central carbon metabolism					
Sugar uptake and utilization	<i>galk</i>	←	mediated by Spot42	mRNA destabilization	Møller et al., 2002
	<i>ptsG</i>	←	RNase E mediated by SgrS	mRNA destabilization	Vanderpool et al., 2004 ;

					Kamamoto et al., 2006
	<i>manXYZ</i>	←	RNase E mediated by SgrS and CsrA	mRNA destabilization	Rice et al., 2012 ; Esquerré et al., 2016
	<i>ptsA</i>	←	RNase III	mRNA destabilization	Sim et al., 2010
Glycolysis	<i>eno</i>	→	RNA degradosome	Component	Py et al., 1996 ; Liou et al., 2002
	<i>eno</i>	←	RNase G	mRNA destabilization	Lee et al., 2002 ; Gordon et al., 2017
	<i>eno</i>	←	RNase III	mRNA processing	Lee et al., 2019
	<i>eno</i>	→	RNase E	Enzyme localization	Murashko and Lin-Chao, 2017
	<i>pgi, pgk, ldhA</i>	←	PNPase, enolase	mRNA destabilization	Bernstein et al., 2004 ; Dressaire BMC et al., 2018,
	<i>tpiA</i>	←	Mediated by CsrA	mRNA stabilization	Esquerré et al., 2016
	<i>pfkA</i>	←	PNPase, enolase, mediated by CsrA	mRNA stabilization	Bernstein et al., 2004, Dressaire et al., 2018, Esquerré et al., 2016
Neoglucogenesis	<i>pck</i>	←	PNPase, enolase, mediated by CsrA	mRNA destabilization	Bernstein et al., 2004, Dressaire et al., 2018 ; Potts et al., 2017
	<i>maeA, maeB</i>	←	Mediated by CsrA	mRNA destabilization	Potts et al., 2017
Glycogen biosynthesis	<i>glgCAP, glgB</i>	←	Mediated by CsrA	mRNA destabilization	Romeo et al., 1993, Liu et al., 1995, Baker et al., 2002, Esquerré et al., 2016
TCA	Citrate	→	PNPase	Enzyme inhibition	Nurmohamed et al., 2011 (<i>E. coli</i>), Stone et al., 2017 (<i>S. antibioticus</i> and <i>S. solfataricus</i>)
	<i>sucA</i>	←	PNPase, enolase	mRNA destabilization	Bernstein et al., 2004, Dressaire et al., 2018
	<i>sucABCD</i>	←	RNase E or RNase III	mRNA processing	De Mets et al., 2018, Miyakoshi et al., 2018
	<i>gltA</i>	←	PNPase, enolase	mRNA destabilization	Bernstein et al., 2004, Dressaire et al., 2018
	<i>sdhCDAB</i>	←	Mediated by RyhB and PAP I	mRNA destabilization	Massé et al., 2002, Sinha et al., 2018

	<i>acnA, fumA</i>	←	Mediated by RyhB	mRNA destabilization	Massé et al., 2002
Acetate metabolism	<i>ackA</i>	←	RNase E mediated by SdhX	mRNA destabilization	De Mets et al., 2018
Nucleotide and energy metabolism					
Nucleotide and energy metabolism	ATP	→	PNPase	Enzyme inhibition	Del Favero et al., 2008
	Cyclic di-GMP	→	PNPase	Enzyme inhibition	Tuckerman et al., 2011
	ATP	→	PAP I, RhIB	Enzyme substrate	Regnier et al., 2009, Carpousis 2007
Amino acid metabolism					
Lysine biosynthesis	DapF	→	RppH	Enzyme activation	Lee et al., 2014, Gao et al., 2018
Amino acid starvation	(p)ppGpp	→	Endoribonucleases toxins	Indirect activation	Tian et al., 2016
Cystein biosynthesis	<i>cysB</i>	←	PNPase/RhIB complex	mRNA stabilization	Tseng et al., 2015
Iron homeostasis, osmotic and oxidative stress					
Low oxygen level	<i>eno</i>	←	RNase III	mRNA processing	Lee et al., 2019
	<i>eno</i>	←	RNase G	mRNA stabilization	Lee et al., 2019
ROS detoxification	<i>sodB</i>	←	Mediated by RyhB and PAP I	mRNA destabilization	Massé et al., 2002, Massé et al., 2004, Sinha et al., 2018
	hydroxy-G mRNAs	←	Free PNPase	mRNA destabilization	Hayakawa et al., 2001, Wu et al., 2009
Iron metabolism	<i>Fur</i> regulon	←	RNase E mediated by RyhB and CsrA	mRNA destabilization	Massé et al., 2002, Massé et al., 2004, Potts et al., 2017
Osmotic stress	<i>betT</i>	←	RNase III	mRNA destabilization	Sim et al., 2014
Oxidative stress	Cd ²⁺	→	ApaH	Enzyme inhibition	Luciano et al., 2019
	<i>trxB</i>	←	ApaH	mRNA destabilization	Luciano et al., 2019
Adaptation to temperature					
Cold temperature	CSPs	←	Free PNPase, RNase II, RNase III	mRNA destabilization	Jones et al., 1987, Luttinger et al., 1996, Yamanaka et al., 2001, Awano et al., 2008, Kim et al., 2008
	bulk mRNAs *	←	RNase R, RhIB and Dead	mRNA destabilization	Zhang et al., 2018, Prud'homme-Généreux et al., 2004, Iost et al., 2013

High temperature	<i>groEL</i> , <i>dnaK</i> , <i>hslV</i> , <i>htpG</i>	←	CspC	mRNA stabilization	Shenhar et al., 2009
------------------	--	---	------	-----------------------	-------------------------

3. Integration of mRNA degradation in systems biology approaches

Systems biology provides powerful approaches to address the complexity of the cross talks between mRNA decay and other cell processes. In that respect, mathematical modelling has long been a companion tool to experimental studies of mRNA degradation, whether for extracting decay rates from data or for helping with the integration and analysis of new experimental findings. In this section, we will illustrate with few examples what has been achieved so far about the mathematical modelling of mRNA degradation and its interplay with other cell processes.

The simplest and most widely used model describes the change in time of the concentration $x(t)$ of a given mRNA as the difference between its synthesis rate, v_s , and its degradation rate, v_d : $\frac{dx(t)}{dt} = v_s - v_d$. The ordinary differential equation (ODE) model is deterministic since all mRNAs transcribed from the same gene have the same fate. In the case of transcription arrest ($v_s = 0$), the model simplifies to: $\frac{dx(t)}{dt} = -v_d$ (Equation 1). Assimilating the degradation mechanism to a first-order reaction with degradation constant k_{deg} : $v_d = k_{deg} \times x(t)$, leads to the ODE $\frac{dx(t)}{dt} = -k_{deg} \times x(t)$ and its solution $x(t) = x(0) \times e^{(-k_{deg} \times t)}$. Fitting this exponential function to experimentally measured decay curves allows to determine k_{deg} and mRNA half-life: $t_{1/2} = \frac{\log(2)}{k_{deg}}$. This exponential model has been used for decades to estimate individual mRNA half-lives, for instance from qRT-PCR measurements or high-throughput data (e.g., Bernstein et al 2002; Nouaille et al 2017; for review, Laguerre et al., 2018). Using alternative expressions for v_d gives rise to a larger diversity of decay curves. For instance, non-exponential decay can be obtained when considering that the degradation of individual mRNAs is not independent from that of other cell mRNAs, because they compete for the same limited pool of endonucleases (Etienne et al 2020 JTB). In this model, v_d is based on the Michaelis-Menten expression including the concentration of endonuclease and cell mRNAs.

The exponential model assumes that mRNA decay takes place independently of other cell processes. Considering the complex relationships of mRNA degradation with metabolism, transcription and translation, has pushed modelling efforts to the development of models of varying complexity. They allowed to demonstrate the importance of the cross talks. For instance, coupling of transcription to decay in mRNA stability analyses reproduced well the observed residual transcription by RNA polymerase following transcription arrest (Chen et al., 2015; Dar & Sorek, 2018; Moffitt et al., 2016) and residual mRNA levels at the end of the degradation process (Moffitt et al., 2016). The transcription rate v_s is non-null in this case and compensate for the degradation rate v_d during the time needed by RNA polymerase to complete mRNA elongation. It is switched to zero afterwards and mRNA concentrations decay exponentially.

Another well-studied example of cross talk is the sensitivity of mRNA stability to ribosome loading and translation rate. Translation-degradation models describe the dynamics of ribosomes along the mRNA with a variety of formalisms (*e.g.* Carrier and Keasling 1997; Deneke et al 2013; Mier-y-Terán-Romero 2013; for review, Zur & Tuller, 2016; von der Haar, 2012). In Deneke et al (2013), a stochastic model combining these two processes describes that at each time during elongation of translation, either the ribosome moves to another position on the mRNA or the latter is degraded if it is accessible to endonucleolytic enzymes. The stabilizing effect of translating ribosomes leads to predicted degradation profiles diverging from exponential decay, which better agrees with dynamical transcriptomics data in *E. coli*.

Other stochastic models examined quantitatively the regulation of mRNA degradation by small RNAs during stress responses, through the modulation of their concentration, the involvement of protein Hfq or their interplay with ribosomes (*e.g.* Levine et al. 2007; Arbel-Goren et al. 2016; Lim and Adamson, 2011; Lavi-Itzkovitz et al 2014). In Lavi-Itzkovitz et al for instance, quantitative experiments and stochastic modelling revealed that bound ribosomes facilitate the binding of a small RNA. The increased translation rates thus result in strengthened repression by the small RNA and destabilization of the target mRNA.

The above models enabled the analysis of mRNA degradation coupling to either transcription or translation. Further integration of these processes helped study (among others) the propagation of noise during bacterial protein expression and the occurrence of bursts in protein production (*e.g.*, Friedman et al 2006 ; McAdams and Arkin, 1997; Paulsson, 2005; Thattai and van Oudenaarden, 2001; Kim and Jacobs-Wagner, 2018). If the transcriptional origins of bursts have been well studied, modelling the coupling of transcription, translation, degradation and their dependencies (co-transcriptional degradation and mRNA stabilization by

ribosomes) helped demonstrate that the short lifetime of bacterial mRNAs facilitates the production of protein bursts as well (Kim and Jacobs-Wagner, 2018).

The realization that metabolism and mRNA degradation are tightly coupled is fairly recent. This is reflected by the relative scarcity of modelling efforts in this direction. Production and degradation of mRNAs is often omitted in metabolic models that include gene expression. When they do so, the models describe mRNA decay with simple first-order kinetics without analyzing the metabolism-degradation cross talk (*e.g.* Usuda et al 2010). Notable exceptions concern studies that analyze the impact of mRNA degradation on cell processes with simplified models. A first study combined quantitative experiments and modelling of the cross talk between translation, degradation and energetic metabolism (Mitarai et al 2008). The model allowed to study conditions leading to ribosomes colliding and queuing, which slows down translation. Cells appear to minimize the risk of collision by selecting codon usage and making mRNAs unstable. The energetic cost of wasting mRNAs (producing and recycling them, and having more RNA polymerases to make up for the mRNA loss) is less than synthesizing ribosomes that compensate for the ones colliding, for reasons of saved translation time. In Thomas et al, 2018, a stochastic coarse-grained model of bacterial growth was used to show that fluctuations in the degradation of mRNAs coding for proteins involved in metabolism is a major source of growth heterogeneity at small growth rate. Another deterministic model described at the genome-scale level the production of the transcription-translation machinery from amino acids and nucleotides, under environmental and biological constraints related to nutrient availability in the growth medium and reaction directionality (Thiele et al 2010). In this constraint-based model, genes involved in transcription, mRNA degradation, protein folding and active tRNA availability appear to be determinant for the optimal production of ribosomes, notably the gene *rne* coding for RNase E.

The mechanistic models cited above describe mRNA degradation and its coupling with cell processes at various depths of molecular details. Models focusing on a given mRNA are generally more precise and informative, but difficult to scale up when several if not all cell mRNAs need to be considered. Their cleavage by the same degradation machinery makes the kinetics of their degradation dependent on each other. Without speaking of the other cellular cross talks, this leads to the development of large mechanistic models with coupled equations, relying on the use of numerous parameters to reproduce the experimental observations (Etienne et al in preparation). In addition, these models are cumbersome to handle, to parameterize and

become quickly intractable. This is problematic as most systems biology studies rely nowadays on the acquisition of high throughput data. Coarse-grained models lie on the other side of the spectrum. They simplify the molecular details by lumping together molecular constituents (*e.g.* Thomas et al 2018) or by simplifying the kinetics of the reaction (*e.g.* Thiele et al 2010). They permit the analysis of grand principles involved in the control of mRNA degradation and inter-dependent cell processes, but lack the molecular details that one would need to finely understand how cell processes influence the fate of individual mRNAs. For instance, predictions with these models cannot be directly confronted to transcriptomics, proteomics, and metabolomics data. In addition, these models rely on very different formalisms, stochastic, deterministic, or constraint based, which ones vary in their capacity to describe cell processes, from the noise of biochemical reactions to the flux of matter in the constraint-based models. Efforts to bring all these formalisms in one single model exist. For instance, whole-cell models of *Mycoplasma genitalium* and *E. coli* combine deterministic constraint-based models for metabolism, deterministic ODE models for cell division, and stochastic models for transcription, translation, and degradation of mRNAs and proteins (Karr et al 2012; Macklin et al 2020).

The tight coupling of mRNA degradation to cell processes thus poses a challenge to modellers in systems biology approaches. There is no ideal model but clearly, it should be tailored to maximally exploit the available experimental data and be informative on the cross talks studied. This poses a challenge to experimentalists as well. Experiments need to be carefully designed to produce quantitative and informative multi-omics data reporting on the state of metabolism and gene expression. At the end of the journey, combining these quantitative experiments and mathematical modelling should help identify missing connections between mRNA degradation, cell metabolism and other cell processes, and grasp how these intertwined connections shape the dynamics of mRNA decay. The resulting models can be used for understanding purposes or as test bed to predict strategies aiming at modulating mRNA stability for biotechnological purposes.

4. Playing with mRNA degradation to improve protein synthesis

Because mRNA degradation plays an essential role in *E. coli* in regulating gene expression, and therefore protein synthesis, recent researches in microbial engineering aim at developing control strategies of mRNA degradation to regulate synthesis of proteins of interest.

Regulation of the degradation of specific mRNAs to control their expression

A first strategy is to use *trans*-acting control of mRNA degradation to down-regulate specifically gene expressions detrimental to recombinant protein synthesis (**Figure 2A**). In biotechnological processes, acetate production inhibits *E. coli* growth due to its toxicity and limits recombinant protein synthesis. Approaches based on both native and synthetic sRNAs base-pairings are developed to reduce acetate production. Over-expression of the native SgrS sRNA destabilizes *ptsG* mRNA decreasing glucose uptake and therefore reduces 5-fold acetate production (Negrete et al., 2013). Synthetic antisense RNAs targeting the *pta* and *ackA* genes encoding respectively the phosphotransacetylase and acetate kinase, two key enzymes for acetate synthesis, are also used to decrease mRNA levels (by 10-20%) and thus acetate production by 2.5 mM (Kim et al., 2003). In this case, acetate production reduction results in a 2-fold increased recombinant protein synthesis. A high-throughput screening strategy is now available to design any desired synthetic antisense RNA to down-regulate in *trans* any endogenous gene expression (Sharma et al., 2012).

A second approach of mRNA degradation regulation consists in a *cis*-acting control of degradation to regulate the synthesis of the corresponding protein (**Figure 2B**). Such approaches are often based on RNA elements localized at the mRNA extremities. Differential half-lives and thus levels of a reporter mRNA are achieved only by changing the 5'UTR sequence of the molecule (Nouaille et al., 2017 ; Viegas et al., 2018). Small size sets of native and artificial 5'UTR allow scalable expression of a reporter protein (around 4.5 and 8-fold ranges for native and artificial UTR, respectively). At a larger scale, design of thousands of synthetic sequences of the 5'UTR and the first codons of the reporter gene results in a range of mRNA half-lives mostly between 2 to 11 minutes (Cambray et al., 2018). This strategy can be used to increase protein synthesis when the stabilization of the mRNA does not provoke secondary structures impeding the translation. From a molecular point of view, it is known that riboswitch (for example the *lysC* riboswitch upon lysine binding) at the 5'-end extremity can directly influence the corresponding protein level by modulating mRNA degradation when RNase E access is impaired (Caron et al., 2012). At the 3' mRNA extremity, insertion of synthetic repetitive palindromic sequences to prevent 3'-5' exonuclease degradation improves the level of a fluorescent protein between 23 and 54% depending on the interval length between the repetitive palindromic sequence and stop codon (Deng et al., 2019). Features of the protein-coding sequence also contribute to a *cis*-control of mRNA stability and protein level. At the omics scale of *E. coli* endogenous genes, there is a correlation between codon sequence (defined

by a codon-influence metric or the codon adaptation index), mRNA half-life and protein level (Boël et al., 2016 ; Esquerré et al., 2015). The sign and strength of the correlation differ with the growth phase and rate. In the case of polycistronic operons, protective RNA structures within the coding sequence provide differential decay of the mRNA molecule and thus differential protein expressions (Dar et al., 2018). For instance in the *tatABC* operon, a 25-fold higher expression of the TatA protein compared to TatBC proteins is associated with a 2.5-fold stabilization of the *tatA* mRNA segment compared to the *tatBC* mRNA region. Insertion of such stabilizing RNA elements within synthetic operon can be a solution to promote non-uniform protein stoichiometry when expressing artificial metabolic pathway in synthetic biology.

Large-scale regulation of mRNA degradation to enhance protein synthesis

An alternative strategy to regulate protein synthesis is to act on mRNA degradation at the cellular level (**Figure 2C**). The straight and simple idea to increase global protein synthesis would be to use ribonuclease-deficient host strains to prevent mRNA degradation. In this line, a BL21 strain (BL21 Star™(DE3)) containing the *rne131* allele expressing a truncated RNase E (lacking the C-terminal domain) is widely used to overproduce heterologous proteins in *E. coli*. However, this solution is not so simple since the mRNA stabilization effect differs between mRNAs, the less stable and less translated mRNAs being more stabilized (Lopez et al. 1999) and a phenomenon of autoregulation of the RNase E partially compensates the loss of activity (Leroy et al. 2002). In the case of 3'-5' exoribonucleases (except the oligoribonuclease), overlapping activities (Andrade et al., 2009) complicate the construction of strains with reduced exoribonuclease activity. Nevertheless, Venturelli and coworkers successfully play with genome-scale mRNA degradation to improve expression of synthetic pathways (Venturelli et al., 2017). They develop an elegant approach in which the endoribonuclease MazF toxin degrades almost all mRNAs except the mRNA of interest specifically protected by the absence of MazF recognition sites in the sequence. The cellular resource is then redirected from growth towards the expression of the recombinant protein. Depending on the MazF induction level, the recombinant protein level increases between to 3 to 5 times.

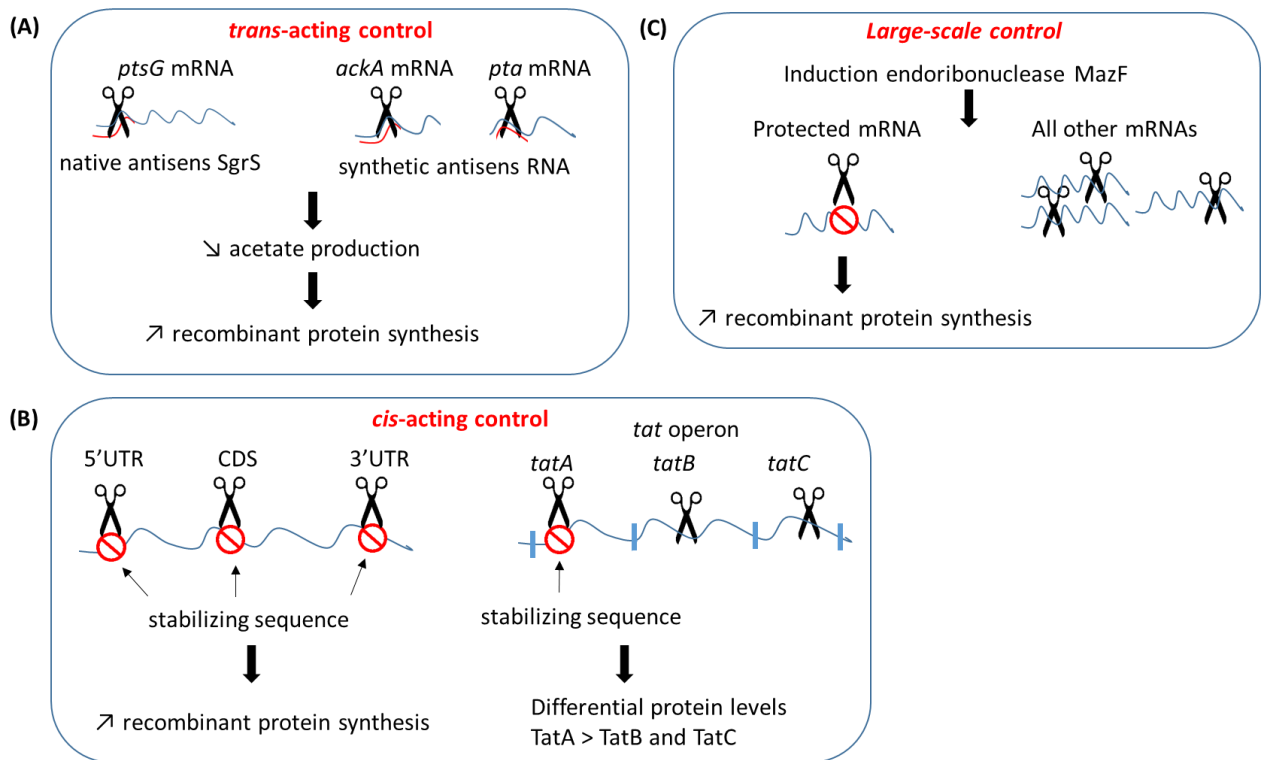


Figure 2 : Control strategies of mRNA degradation to regulate synthesis of proteins of interest

Conclusions

This review demonstrates that the process of mRNA degradation can be placed at the heart of metabolic regulations and cell adaptation. Regulation of mRNA degradation constitutes an indispensable level of regulation in systems biology approaches, at the centre of a physiological regulatory loop: mRNA degradation impacts the main processes of gene expression and many metabolic pathways and stress responses but is its-self regulated by the carbon and energetic metabolisms (**Figure 3**). Intuitively, the most obvious contributor of the loop is enolase, both a glycolytic enzyme and a component of the RNA degradosome but its role is not clear yet. Nevertheless these interconnections involve very different actors of the degradation, such as RNases and other components of the RNA degradation machinery, as well as RNA-binding proteins (global regulator and chaperone) and small RNAs. The elements of the metabolism involved in the interactions are also multiple including different metabolic pathways, targeted mRNAs, metabolites or enzymes. This regulatory cycle opens new perspectives in microbial engineering to act on mRNA degradation regulators to optimize the production of protein of interest. Finally, we focused in this review on the tight interaction

between the mRNA degradation and cell physiology but the growth condition dependency of *E. coli* “stable” RNA, ribosomal RNA (Fessler et al., 2020) and transfer RNA (Sørensen et al., 2018 ; Svenningsen et al., 2017) extends the interaction to the general process of RNA degradation highlighting further the need to integrate the RNA degradation process in *E. coli* physiological studies.

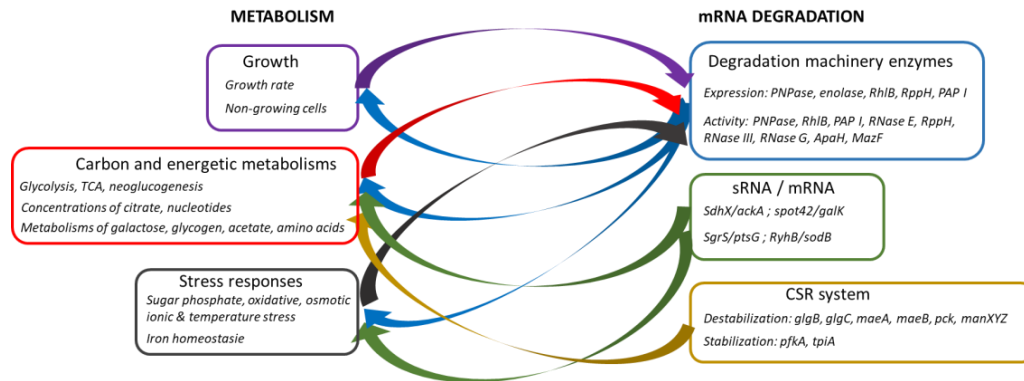


Figure 3 : Regulatory loop of interactions between mRNA degradation and metabolism in *E. coli*.

References

- Altuvia Y**, Bar A, Reiss N, Karavani E, Argaman L, Margalit H. In vivo cleavage rules and target repertoire of RNase III in *Escherichia coli*. *Nucleic Acids Res.* 2018 Nov 2;46(19):10380-10394.
- Andrade JM**, Cairrão F, Arraiano CM. RNase R affects gene expression in stationary phase: regulation of *ompA*. *Mol Microbiol.* 2006 Apr;60(1):219-28.
- Andrade JM**, Pobre V, Silva IJ, Domingues S, Arraiano CM. The role of 3'-5' exoribonucleases in RNA degradation. *Prog Mol Biol Transl Sci.* 2009;85:187-229.
- Arnold TE**, Yu J, Belasco JG. mRNA stabilization by the *ompA* 5' untranslated region: two protective elements hinder distinct pathways for mRNA degradation. *RNA.* 1998 Mar;4(3):319-30.
- Awano N**, Inouye M, Phadtare S. RNase activity of polynucleotide phosphorylase is critical at low temperature in *Escherichia coli* and is complemented by RNase II. *J Bacteriol.* 2008 Sep;190(17):5924-33.
- Baker CS**, Morozov I, Suzuki K, Romeo T, Babitzke P. CsrA regulates glycogen biosynthesis by preventing translation of *glgC* in *Escherichia coli*. *Mol Microbiol.* 2002 Jun;44(6):1599-610.
- Baumeister R**, Flache P, Melefors O, von Gabain A, Hillen W. Lack of a 5' non-coding region in Tn1721 encoded *tetR* mRNA is associated with a low efficiency of translation and a short half-life in *Escherichia coli*. *Nucleic Acids Res.* 1991 Sep 11;19(17):4595-600.
- Bayas CA**, Wang J, Lee MK, Schrader JM, Shapiro L, Moerner WE. Spatial organization and dynamics of RNase E and ribosomes in *Caulobacter crescentus*. *Proc Natl Acad Sci U S A.* 2018 Apr 17;115(16):E3712-E3721.
- Bennett BD**, Kimball EH, Gao M, Osterhout R, Van Dien SJ, Rabinowitz JD. Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat Chem Biol.* 2009 Aug;5(8):593-9.
- Bernstein JA**, Khodursky AB, Lin PH, Lin-Chao S, Cohen SN. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc Natl Acad Sci U S A.* 2002 Jul 23;99(15):9697-702.
- Bernstein JA**, Lin PH, Cohen SN, Lin-Chao S. Global analysis of *Escherichia coli* RNA degradosome function using DNA microarrays. *Proc Natl Acad Sci U S A.* 2004 Mar 2;101(9):2758-63. Epub 2004 Feb 23.
- Bessarab DA**, Kaberdin VR, Wei CL, Liou GG, Lin-Chao S. RNA components of *Escherichia coli* degradosome: evidence for rRNA decay. *Proc Natl Acad Sci U S A.* 1998 Mar 17;95(6):3157-61.
- Boël G**, Letso R, Neely H, Price WN, Wong KH, Su M, Luff J, Valecha M, Everett JK, Acton TB, Xiao R2,4, Montelione GT, Aalberts DP, Hunt JF. Codon influence on protein expression in *E. coli* correlates with mRNA levels. *Nature.* 2016 Jan 21;529(7586):358-363.
- Butland G**, Peregrín-Alvarez JM, Li J, Yang W, Yang X, Canadien V, Starostine A, Richards D, Beattie B, Krogan N, Davey M, Parkinson J, Greenblatt J, Emili A. Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature.* 2005 Feb 3;433(7025):531-7.

Cairrão F, Cruz A, Mori H, Arraiano CM. Cold shock induction of RNase R and its role in the maturation of the quality control mediator SsrA/tmRNA. *Mol Microbiol.* 2003 Nov;50(4):1349-60.

Cambray G, Guimaraes JC, Arkin AP. Evaluation of 244,000 synthetic sequences reveals design principles to optimize translation in *Escherichia coli*. *Nat Biotechnol.* 2018 Nov;36(10):1005-1015.

Cannistraro VJ, Kennell D. Evidence that the 5' end of lac mRNA starts to decay as soon as it is synthesized. *J Bacteriol.* 1985 Feb;161(2):820-2.

Carpousis AJ. The *Escherichia coli* RNA degradosome: structure, function and relationship in other ribonucleolytic multienzyme complexes. *Biochem Soc Trans.* 2002 Apr;30(2):150-5.

Carpousis AJ. The RNA degradosome of *Escherichia coli*: an mRNA-degrading machine assembled on RNase E. *Annu Rev Microbiol.* 2007;61:71-87.

Castellana M, Hsin-Jung Li S, Wingreen NS. Spatial organization of bacterial transcription and translation. *Proc Natl Acad Sci U S A.* 2016 Aug 16;113(33):9286-91.

Caron MP, Bastet L, Lussier A, Simoneau-Roy M, Massé E, Lafontaine DA. Dual-acting riboswitch control of translation initiation and mRNA decay. *Proc Natl Acad Sci U S A.* 2012 Dec 11;109(50):E3444-53.

Chen H, Shiroguchi K, Ge H, Xie XS. Genome-wide study of mRNA degradation and transcript elongation in *Escherichia coli*. *Mol Syst Biol.* 2015 Jan 12;11(1):781.

Choi U, Park YH, Kim YR, Seok YJ, Lee CR. Effect of the RNA pyrophosphohydrolase RppH on envelope integrity in *Escherichia coli*. *FEMS Microbiol Lett.* 2017 Aug 15;364(15).

Chubukov V, Gerosa L, Kochanowski K, Sauer U. Coordination of microbial metabolism. *Nat Rev Microbiol.* 2014 May;12(5):327-40.

Clarke JE, Kime L, Romero A D, McDowall KJ. Direct entry by RNase E is a major pathway for the degradation and processing of RNA in *Escherichia coli*. *Nucleic Acids Res.* 2014 Oct;42(18):11733-51.

Commichau FM, Rothe FM, Herzberg C, Wagner E, Hellwig D, Lehnik-Habrink M, Hammer E, Völker U, Stülke J. Novel activities of glycolytic enzymes in *Bacillus subtilis*: interactions with essential proteins involved in mRNA processing. *Mol Cell Proteomics.* 2009 Jun;8(6):1350-60.

Dar D, Sorek R. Extensive reshaping of bacterial operons by programmed mRNA decay. *PLoS Genet.* 2018 Apr 18;14(4):e1007354.

Deana A, Belasco JG. Lost in translation: the influence of ribosomes on bacterial mRNA decay. *Genes Dev.* 2005 Nov 1;19(21):2526-33.

De Mets F, Van Melderen L, Gottesman S. Regulation of acetate metabolism and coordination with the TCA cycle via a processed small RNA. *Proc Natl Acad Sci U S A.* 2019 Jan 15;116(3):1043-1052.

Del Favero M, Mazzantini E, Briani F, Zangrossi S, Tortora P, Dehò G. Regulation of *Escherichia coli* polynucleotide phosphorylase by ATP. *J Biol Chem.* 2008 Oct 10;283(41):27355-9.

Deng C, Lv X, Li J, Liu Y, Du G, Amaro RL, Liu L. Synthetic repetitive extragenic palindromic (REP) sequence as an efficient mRNA stabilizer for protein production and metabolic engineering in prokaryotic cells. *Biotechnol Bioeng.* 2019 Jan;116(1):5-18.

Deutscher MP, Reuven NB. Enzymatic basis for hydrolytic versus phosphorolytic mRNA degradation in *Escherichia coli* and *Bacillus subtilis*. *Proc Natl Acad Sci U S A.* 1991 Apr 15;88(8):3277-80.

- Deutscher MP.** Ribonuclease multiplicity, diversity, and complexity. *J Biol Chem.* 1993 Jun 25;268(18):13011-4.
- Dressaire C,** Pobre V, Laguerre S, Girbal L, Arraiano CM, Coccagn-Bousquet M. PNPase is involved in the coordination of mRNA degradation and expression in stationary phase cells of *Escherichia coli*. *BMC Genomics.* 2018 Nov 29;19(1):848.
- Eriksen M,** Sneppen K, Pedersen S, Mitarai N. Occlusion of the Ribosome Binding Site Connects the Translational Initiation Frequency, mRNA Stability and Premature Transcription Termination. *Front Microbiol.* 2017 Mar 14;8:362.
- Esquerré T,** Laguerre S, Turlan C, Carpousis AJ, Girbal L, Coccagn-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 Feb;42(4):2460-72.
- Esquerré T,** Moisan A, Chiapello H, Arike L, Vilu R, Gaspin C, Coccagn-Bousquet M, Girbal L. Genome-wide investigation of mRNA lifetime determinants in *Escherichia coli* cells cultured at different growth rates. *BMC Genomics.* 2015 Apr 9;16:275.
- Esquerré T,** Bouvier M, Turlan C, Carpousis AJ, Girbal L, Coccagn-Bousquet M. The Csr system regulates genome-wide mRNA stability and transcription and thus gene expression in *Escherichia coli*. *Sci Rep.* 2016 Apr 26;6:25057.
- Etienne T. A.,** M. Coccagn-Bousquet, D. Ropers. Competitive effects in bacterial mRNA decay. *Journal of Theoretical Biology* (in revision).
- Fessler M,** Gummesson B, Charbon G, Svenningsen SL, Sørensen MA. Short-term kinetics of rRNA degradation in *Escherichia coli* upon starvation for carbon, amino acid or phosphate. *Mol Microbiol.* 2020 Jan 20.
- Gao A,** Vasilyev N, Luciano DJ, Levenson-Palmer R, Richards J, Marsiglia WM, Traaseth NJ, Belasco JG, Serganov A. Structural and kinetic insights into stimulation of RppH-dependent RNA degradation by the metabolic enzyme DapF. *Nucleic Acids Res.* 2018 Jul 27;46(13):6841-6856.
- Golding I, Cox EC.** RNA dynamics in live *Escherichia coli* cells. *Proc Natl Acad Sci U S A.* 2004 Aug 3;101(31):11310-5. Epub 2004 Jul 26.
- Golding I, Cox EC.** Physical nature of bacterial cytoplasm. *Phys Rev Lett.* 2006 Mar 10;96(9):098102. Epub 2006 Mar 10.
- Gordon GC,** Cameron JC, Pflieger BF. RNA Sequencing Identifies New RNase III Cleavage Sites in *Escherichia coli* and Reveals Increased Regulation of mRNA. *mBio.* 2017 Mar 28;8(2). pii: e00128-17.
- Gudapaty S,** Suzuki K, Wang X, Babitzke P, Romeo T. Regulatory interactions of Csr components: the RNA binding protein CsrA activates *csrB* transcription in *Escherichia coli*. *J Bacteriol.* 2001 Oct;183(20):6017-27.
- Hadjeras L,** Poljak L, Bouvier M, Morin-Ogier Q, Canal I, Coccagn-Bousquet M, Girbal L, Carpousis AJ. Detachment of the RNA degradosome from the inner membrane of *Escherichia coli* results in a global slowdown of mRNA degradation, proteolysis of RNase E and increased turnover of ribosome-free transcripts. *Mol Microbiol.* 2019 Jun;111(6):1715-1731.
- Hardwick SW,** Chan VS, Broadhurst RW, Luisi BF. An RNA degradosome assembly in *Caulobacter crescentus*. *Nucleic Acids Res.* 2011 Mar;39(4):1449-59.

- Hayakawa H**, Kuwano M, Sekiguchi M. Specific binding of 8-oxoguanine-containing RNA to polynucleotide phosphorylase protein. *Biochemistry*. 2001 Aug 21;40(33):9977-82.
- Hayes CS**, Sauer RT. Cleavage of the A site mRNA codon during ribosome pausing provides a mechanism for translational quality control. *Mol Cell*. 2003 Oct;12(4):903-11.
- Hudeček O**, Benoni R, Reyes-Gutierrez PE, Culka M, Šanderová H, Hubálek M, Rulíšek L, Cvačka J, Krásný L, Cahová H. Dinucleoside polyphosphates act as 5'-RNA caps in bacteria. *Nat Commun*. 2020 Feb 26;11(1):1052.
- Hunt A**, Rawlins JP, Thomaides HB, Errington J. Functional analysis of 11 putative essential genes in *Bacillus subtilis*. *Microbiology*. 2006 Oct;152(Pt 10):2895-2907.
- Iost I, Dreyfus M**. The stability of *Escherichia coli* lacZ mRNA depends upon the simultaneity of its synthesis and translation. *EMBO J*. 1995 Jul 3;14(13):3252-61.
- Iost I**, Bizebard T, Dreyfus M. Functions of DEAD-box proteins in bacteria: current knowledge and pending questions. *Biochim Biophys Acta*. 2013 Aug;1829(8):866-77.
- Jain C**, Kleckner N. IS10 mRNA stability and steady state levels in *Escherichia coli*: indirect effects of translation and role of rne function. *Mol Microbiol*. 1993 Jul;9(2):233-47.
- Jasiecki J**, Wegrzyn G. Growth-rate dependent RNA polyadenylation in *Escherichia coli*. *EMBO Rep*. 2003 Feb;4(2):172-7.
- Jasiecki J**, Wegrzyn G. Transcription start sites in the promoter region of the *Escherichia coli* pcnB (plasmid copy number) gene coding for poly(A) polymerase I. *Plasmid*. 2006 Mar;55(2):169-72. Epub 2005 Dec 2.
- Jasiecki J**, Wegrzyn G. Phosphorylation of *Escherichia coli* poly(A) polymerase I and effects of this modification on the enzyme activity. *FEMS Microbiol Lett*. 2006 Aug;261(1):118-22.
- Jones PG**, VanBogelen RA, Neidhardt FC. Induction of proteins in response to low temperature in *Escherichia coli*. *J Bacteriol*. 1987 May;169(5):2092-5.
- Kaberđin VR, Bläsi U**. Translation initiation and the fate of bacterial mRNAs. *FEMS Microbiol Rev*. 2006 Nov;30(6):967-79. Epub 2006 Sep 21.
- Kannaiah S**, Livny J, Amster-Choder O. Spatiotemporal Organization of the *E. coli* Transcriptome: Translation Independence and Engagement in Regulation. *Mol Cell*. 2019 Nov 21;76(4):574-589.e7.
- Kawamoto H**, Koide Y, Morita T, Aiba H. Base-pairing requirement for RNA silencing by a bacterial small RNA and acceleration of duplex formation by Hfq. *Mol Microbiol*. 2006 Aug;61(4):1013-22. Epub 2006 Jul 12.
- Khemici V**, Poljak L, Luisi BF, Carpousis AJ. The RNase E of *Escherichia coli* is a membrane-binding protein. *Mol Microbiol*. 2008 Nov;70(4):799-813.
- Kim JY**, Cha HJ. Down-regulation of acetate pathway through antisense strategy in *Escherichia coli*: improved foreign protein production. *Biotechnol Bioeng*. 2003 Sep 30;83(7):841-53.
- Kimata K**, Tanaka Y, Inada T, Aiba H. Expression of the glucose transporter gene, ptsG, is regulated at the mRNA degradation step in response to glycolytic flux in *Escherichia coli*. *EMBO J*. 2001 Jul 2;20(13):3587-95.

Kosuri S, Goodman DB, Cambray G, Mutalik VK, Gao Y, Arkin AP, Endy D, Church GM. Composability of regulatory sequences controlling transcription and translation in *Escherichia coli*. *Proc Natl Acad Sci U S A*. 2013 Aug 20;110(34):14024-9.

Kushner SR. mRNA decay in *Escherichia coli* comes of age. *J Bacteriol*. 2002 Sep;184(17):4658-65; discussion 4657.

Lacoux C, Fouquier d'Hérouël A, Wessner-Le Bohec F, Innocenti N, Bohn C, Kennedy SP, Rochat T, Bonnin RA, Serror P, Aurell E, Bouloc P, Repoila F. Dynamic insights on transcription initiation and RNA processing during bacterial adaptation. *RNA*. 2020 Apr;26(4):382-395

Lee K, Bernstein JA, Cohen SN. RNase G complementation of *rne* null mutation identifies functional interrelationships with RNase E in *Escherichia coli*. *Mol Microbiol*. 2002 Mar;43(6):1445-56.

Lee CR, Kim M, Park YH, Kim YR, Seok YJ. RppH-dependent pyrophosphohydrolysis of mRNAs is regulated by direct interaction with DapF in *Escherichia coli*. *Nucleic Acids Res*. 2014 Nov 10;42(20):12746-57.

Lee M, Joo M, Sim M, Sim SH, Kim HL, Lee J, Ryu M, Yeom JH, Hahn Y, Ha NC, Cho JC, Lee K. The coordinated action of RNase III and RNase G controls enolase expression in response to oxygen availability in *Escherichia coli*. *Sci Rep*. 2019 Nov 21;9(1):17257.

Lehnik-Habrink M, Newman J, Rothe FM, Solovyova AS, Rodrigues C, Herzberg C, Commichau FM, Lewis RJ, Stülke J. RNase Y in *Bacillus subtilis*: a Natively disordered protein that is the functional equivalent of RNase E from *Escherichia coli*. *J Bacteriol*. 2011 Oct;193(19):5431-41.

Leroy A, Vanzo NF, Sousa S, Dreyfus M, Carpousis AJ. Function in *Escherichia coli* of the non-catalytic part of RNase E: role in the degradation of ribosome-free mRNA. *Mol Microbiol*. 2002 Sep;45(5):1231-43.

Lin PH, Lin-Chao S. RhIB helicase rather than enolase is the beta-subunit of the *Escherichia coli* polynucleotide phosphorylase (PNPase)-exoribonucleolytic complex. *Proc Natl Acad Sci U S A*. 2005 Nov 15;102(46):16590-5. Epub 2005 Nov 7.

Lin HH, Lin CH, Hwang SM, Tseng CP. High growth rate downregulates *fumA* mRNA transcription but is dramatically compensated by its mRNA stability in *Escherichia coli*. *Curr Microbiol*. 2012 May;64(5):412-7

Liou GG, Chang HY, Lin CS, Lin-Chao S. DEAD box RhIB RNA helicase physically associates with exoribonuclease PNPase to degrade double-stranded RNA independent of the degradosome-assembling region of RNase E. *J Biol Chem*. 2002 Oct 25;277(43):41157-62. Epub 2002 Aug 13.

Luciano DJ, Vasilyev N, Richards J, Serganov A, Belasco JG. A Novel RNA Phosphorylation State Enables 5' End-Dependent Degradation in *Escherichia coli*. *Mol Cell*. 2017 Jul 6;67(1):44-54.e6.

Luciano DJ, Levenson-Palmer R, Belasco JG. Stresses that Raise Np4A Levels Induce Protective Nucleoside Tetraphosphate Capping of Bacterial RNA. *Mol Cell*. 2019 Sep 5;75(5):957-966.e8.

Liu MY, Yang H, Romeo T. The product of the pleiotropic *Escherichia coli* gene *csrA* modulates glycogen biosynthesis via effects on mRNA stability. *J Bacteriol*. 1995 May;177(10):2663-72.

Liu MY, Gui G, Wei B, Preston JF 3rd, Oakford L, Yüksel U, Giedroc DP, Romeo T. The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in *Escherichia coli*. *J Biol Chem*. 1997 Jul 11;272(28):17502-10.

- Mackie GA.** RNase E: at the interface of bacterial RNA processing and decay. *Nat Rev Microbiol.* 2013 Jan;11(1):45-57.
- Massé E,** Gottesman S. A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc Natl Acad Sci U S A.* 2002 Apr 2;99(7):4620-5. Epub 2002 Mar 26.
- Massé E,** Vanderpool CK, Gottesman S. Effect of RyhB small RNA on global iron use in *Escherichia coli*. *J Bacteriol.* 2005 Oct;187(20):6962-71.
- Miczak A,** Srivastava RA, Apirion D. Location of the RNA-processing enzymes RNase III, RNase E and RNase P in the *Escherichia coli* cell. *Mol Microbiol.* 1991 Jul;5(7):1801-10.
- Miyakoshi M,** Matera G, Maki K, Sone Y, Vogel J. Functional expansion of a TCA cycle operon mRNA by a 3' end-derived small RNA. *Nucleic Acids Res.* 2019 Feb 28;47(4):2075-2088.
- Mohanty BK,** Kushner SR. Genomic analysis in *Escherichia coli* demonstrates differential roles for polynucleotide phosphorylase and RNase II in mRNA abundance and decay. *Mol Microbiol.* 2003 Oct;50(2):645-58.
- Moffitt JR,** Pandey S, Boettiger AN, Wang S, Zhuang X. Spatial organization shapes the turnover of a bacterial transcriptome. *Elife.* 2016 May 20;5. pii: e13065.
- Montero Llopis P,** Jackson AF, Sliusarenko O, Surovtsev I, Heinritz J, Emonet T, Jacobs-Wagner C. Spatial organization of the flow of genetic information in bacteria. *Nature.* 2010 Jul 1;466(7302):77-81.
- Morikawa N,** Imamoto F. Degradation of tryptophan messenger. On the degradation of messenger RNA for the tryptophan operon in *Escherichia coli*. *Nature.* 1969 Jul 5;223(5201):37-40.
- Morin M.,** Enjalbert B., Ropers D., Girbal L., and Coccagn-Bousquet M. Genome-wide stabilization of mRNA during a 'feast-to-famine' growth transition in *Escherichia coli*. (2020) *mSphere* (in press).
- Morita T,** El-Kazzaz W, Tanaka Y, Inada T, Aiba H. Accumulation of glucose 6-phosphate or fructose 6-phosphate is responsible for destabilization of glucose transporter mRNA in *Escherichia coli*. *J Biol Chem.* 2003 May 2;278(18):15608-14. Epub 2003 Feb 10.
- Morita T,** Kawamoto H, Mizota T, Inada T, Aiba H. Enolase in the RNA degradosome plays a crucial role in the rapid decay of glucose transporter mRNA in the response to phosphosugar stress in *Escherichia coli*. *Mol Microbiol.* 2004 Nov;54(4):1063-75.
- Morse DE,** Mosteller RD, Yanofsky C. Dynamics of synthesis, translation, and degradation of trp operon messenger RNA in *E. coli*. *Cold Spring Harb Symp Quant Biol.* 1969;34:725-40.
- Møller T,** Franch T, Udesen C, Gerdes K, Valentin-Hansen P. Spot 42 RNA mediates discoordinate expression of the *E. coli* galactose operon. *Genes Dev.* 2002 Jul 1;16(13):1696-706.
- Murashko ON, Lin-Chao S.** *Escherichia coli* responds to environmental changes using enolase degradosomes and stabilized DicF sRNA to alter cellular morphology. *Proc Natl Acad Sci U S A.* 2017 Sep 19;114(38):E8025-E8034.
- Nadratowska-Wesołowska B,** Słomińska-Wojewódzka M, Łyżeń R, Wegrzyn A, Szalewska-Pałasz A, Wegrzyn G. Transcription regulation of the *Escherichia coli* *pcnB* gene coding for poly(A) polymerase I: roles of ppGpp, DksA and sigma factors. *Mol Genet Genomics.* 2010 Oct;284(4):289-305.

- Nanbu-Wakao R**, Asoh S, Nishimaki K, Tanaka R, Ohta S. Bacterial cell death induced by human pro-apoptotic Bax is blocked by an RNase E mutant that functions in an anti-oxidant pathway. *Genes Cells*. 2000 Mar;5(3):155-67.
- Negrete A**, Majdalani N, Phue JN, Shiloach J. Reducing acetate excretion from *E. coli* K-12 by over-expressing the small RNA SgrS. *N Biotechnol*. 2013 Jan 25;30(2):269-73.
- Nevo-Dinur K**, Nussbaum-Shochat A, Ben-Yehuda S, Amster-Choder O. Translation-independent localization of mRNA in *E. coli*. *Science*. 2011 Feb 25;331(6020):1081-4.
- Newman JA**, Hewitt L, Rodrigues C, Solovyova AS, Harwood CR, Lewis RJ. Dissection of the network of interactions that links RNA processing with glycolysis in the *Bacillus subtilis* degradosome. *J Mol Biol*. 2012 Feb 10;416(1):121-36.
- Nilsson G**, Belasco JG, Cohen SN, von Gabain A. Growth-rate dependent regulation of mRNA stability in *Escherichia coli*. *Nature*. 1984 Nov 1-7;312(5989):75-7.
- Nilsson G**, Belasco JG, Cohen SN, von Gabain A. Effect of premature termination of translation on mRNA stability depends on the site of ribosome release. *Proc Natl Acad Sci U S A*. 1987 Jul;84(14):4890-4.
- Nogueira T**, de Smit M, Graffe M, Springer M. The relationship between translational control and mRNA degradation for the *Escherichia coli* threonyl-tRNA synthetase gene. *J Mol Biol*. 2001 Jul 20;310(4):709-22.
- Nouaille S**, Mondeil S, Finoux AL, Moulis C, Girbal L, Coccagn-Bousquet M. The stability of an mRNA is influenced by its concentration: a potential physical mechanism to regulate gene expression. *Nucleic Acids Res*. 2017 Nov 16;45(20):11711-11724.
- Nurmohamed S**, Vincent HA, Titman CM, Chandran V, Pears MR, Du D, Griffin JL, Callaghan AJ, Luisi BF. Polynucleotide phosphorylase activity may be modulated by metabolites in *Escherichia coli*. *J Biol Chem*. 2011 Apr 22;286(16):14315-23.
- Oberto J**, Bonnefoy E, Mouray E, Pellegrini O, Wikström PM, Rouvière-Yaniv J. The *Escherichia coli* ribosomal protein S16 is an endonuclease. *Mol Microbiol*. 1996 Mar;19(6):1319-30.
- Pannuri A**, Vakulskas CA, Zere T, McGibbon LC, Edwards AN, Georgellis D, Babitzke P, Romeo T. Circuitry Linking the Catabolite Repression and Csr Global Regulatory Systems of *Escherichia coli*. *J Bacteriol*. 2016 Oct 7;198(21):3000-3015. Print 2016 Nov 1.
- Pedersen K**, Zavialov AV, Pavlov MY, Elf J, Gerdes K, Ehrenberg M. The bacterial toxin RelE displays codon-specific cleavage of mRNAs in the ribosomal A site. *Cell*. 2003 Jan 10;112(1):131-40.
- Pedersen M**, Nissen S, Mitarai N, Lo Svenningsen S, Sneppen K, Pedersen S. The functional half-life of an mRNA depends on the ribosome spacing in an early coding region. *J Mol Biol*. 2011 Mar 18;407(1):35-44.
- Pedersen S**, Terkelsen TB, Eriksen M, Hauge MK, Lund CC, Sneppen K, Mitarai N. Fast Translation within the First 45 Codons Decreases mRNA Stability and Increases Premature Transcription Termination in *E. coli*. *J Mol Biol*. 2019 Mar 15;431(6):1088-1097.
- Pérez-Ortín JE**, Tordera V, Chávez S. Homeostasis in the Central Dogma of molecular biology: the importance of mRNA instability. *RNA Biol*. 2019 Dec;16(12):1659-1666.

Phadtare S. Escherichia coli cold-shock gene profiles in response to over-expression/deletion of CsdA, RNase R and PNPase and relevance to low-temperature RNA metabolism. *Genes Cells.* 2012 Oct;17(10):850-74.

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 Nov 17;8(1):1596.

Prud'homme-Généreux A, Beran RK, Iost I, Ramey CS, Mackie GA, Simons RW. Physical and functional interactions among RNase E, polynucleotide phosphorylase and the cold-shock protein, CsdA: evidence for a 'cold shock degradosome'. *Mol Microbiol.* 2004 Dec;54(5):1409-21.

Py B, Higgins CF, Krisch HM, Carpousis AJ. A DEAD-box RNA helicase in the Escherichia coli RNA degradosome. *Nature.* 1996 May 9;381(6578):169-72.

Rak R, Dahan O, Pilpel Y. Repertoires of tRNAs: The Couplers of Genomics and Proteomics. *Annu Rev Cell Dev Biol.* 2018 Oct 6;34:239-264.

Régnier P, Hajnsdorf E. Poly(A)-assisted RNA decay and modulators of RNA stability. *Prog Mol Biol Transl Sci.* 2009;85:137-85.

Rice PW, Dahlberg JE. A gene between polA and glnA retards growth of Escherichia coli when present in multiple copies: physiological effects of the gene for spot 42 RNA. *J Bacteriol.* 1982 Dec;152(3):1196-210.

Rice JB, Vanderpool CK. The small RNA SgrS controls sugar-phosphate accumulation by regulating multiple PTS genes. *Nucleic Acids Res.* 2011 May;39(9):3806-19.

Richards J, Luciano DJ, Belasco JG. Influence of translation on RppH-dependent mRNA degradation in Escherichia coli. *Mol Microbiol.* 2012 Dec;86(5):1063-72.

Richards J, Belasco JG. Obstacles to Scanning by RNase E Govern Bacterial mRNA Lifetimes by Hindering Access to Distal Cleavage Sites. *Mol Cell.* 2019 Apr 18;74(2):284-295.e5.

Romeo T, Gong M, Liu MY, Brun-Zinkernagel AM. Identification and molecular characterization of csrA, a pleiotropic gene from Escherichia coli that affects glycogen biosynthesis, gluconeogenesis, cell size, and surface properties. *J Bacteriol.* 1993 Aug;175(15):4744-55.

Sauert M, Wolfinger MT, Vesper O, Müller C, Byrgazov K, Moll I. The MazF-regulon: a toolbox for the post-transcriptional stress response in Escherichia coli. *Nucleic Acids Res.* 2016 Aug 19;44(14):6660-75.

Shahbadian K, Jamalli A, Zig L, Putzer H. RNase Y, a novel endoribonuclease, initiates riboswitch turnover in Bacillus subtilis. *EMBO J.* 2009 Nov 18;28(22):3523-33.

Sharma V, Yamamura A, Yokobayashi Y. Engineering artificial small RNAs for conditional gene silencing in Escherichia coli. *ACS Synth Biol.* 2012 Jan 20;1(1):6-13.

Shenhar Y, Rasouly A, Biran D, Ron EZ. Adaptation of Escherichia coli to elevated temperatures involves a change in stability of heat shock gene transcripts. *Environ Microbiol.* 2009 Dec;11(12):2989-97.

Sim SH, Yeom JH, Shin C, Song WS, Shin E, Kim HM, Cha CJ, Han SH, Ha NC, Kim SW, Hahn Y, Bae J, Lee K. Escherichia coli ribonuclease III activity is downregulated by osmotic stress: consequences for the degradation of bdm mRNA in biofilm formation. *Mol Microbiol.* 2010 Jan;75(2):413-25.

Sim M, Lim B, Sim SH, Kim D, Jung E, Lee Y, Lee K. Two tandem RNase III cleavage sites determine betT mRNA stability in response to osmotic stress in Escherichia coli. PLoS One. 2014 Jun 23;9(6):e100520.

Singh D, Chang SJ, Lin PH, Averina OV, Kaberdin VR, Lin-Chao S. Regulation of ribonuclease E activity by the L4 ribosomal protein of Escherichia coli. Proc Natl Acad Sci U S A. 2009 Jan 20;106(3):864-9.

Sinha D, Matz LM, Cameron TA, De Lay NR. Poly(A) polymerase is required for RyhB sRNA stability and function in Escherichia coli. RNA. 2018 Nov;24(11):1496-1511.

Sunohara T, Jojima K, Tagami H, Inada T, Aiba H. Ribosome stalling during translation elongation induces cleavage of mRNA being translated in Escherichia coli. J Biol Chem. 2004 Apr 9;279(15):15368-75. Epub 2004 Jan 26.

Sunohara T, Jojima K, Yamamoto Y, Inada T, Aiba H. Nascent-peptide-mediated ribosome stalling at a stop codon induces mRNA cleavage resulting in nonstop mRNA that is recognized by tmRNA. RNA. 2004 Mar;10(3):378-86.

Stone CM, Butt LE, Bufton JC, Lourenco DC, Gowers DM, Pickford AR, Cox PA, Vincent HA, Callaghan AJ. Inhibition of homologous phosphorolytic ribonucleases by citrate may represent an evolutionarily conserved communicative link between RNA degradation and central metabolism. Nucleic Acids Res. 2017 May 5;45(8):4655-4666.

Strahl H, Turlan C, Khalid S, Bond PJ, Kebalo JM, Peyron P, Poljak L, Bouvier M, Hamoen L, Luisi BF, Carpousis AJ. Membrane recognition and dynamics of the RNA degradosome. PLoS Genet. 2015 Feb 3;11(2):e1004961.

Suzuki K, Wang X, Weilbacher T, Pernestig AK, Melefors O, Georgellis D, Babitzke P, Romeo T. Regulatory circuitry of the CsrA/CsrB and BarA/UvrY systems of Escherichia coli. J Bacteriol. 2002 Sep;184(18):5130-40.

Svenningsen SL, Kongstad M, Stenum TS, Muñoz-Gómez AJ, Sørensen MA. Transfer RNA is highly unstable during early amino acid starvation in Escherichia coli. Nucleic Acids Res. 2017 Jan 25;45(2):793-804.

Sørensen MA, Fehler AO, Lo Svenningsen S. Transfer RNA instability as a stress response in Escherichia coli: Rapid dynamics of the tRNA pool as a function of demand. RNA Biol. 2018;15(4-5):586-593.

Tamura M, Kers JA, Cohen SN. Second-site suppression of RNase E essentiality by mutation of the dead RNA helicase in Escherichia coli. J Bacteriol. 2012 Apr;194(8):1919-26.

Tamura M, Moore CJ, Cohen SN. Nutrient dependence of RNase E essentiality in Escherichia coli. J Bacteriol. 2013 Mar;195(6):1133-41.

Thiele I, Fleming RM, Bordbar A, Schellenberger J, Palsson BØ. Functional characterization of alternate optimal solutions of Escherichia coli's transcriptional and translational machinery. Biophys J. 2010 May 19;98(10):2072-81.

Tian C, Roghanian M, Jørgensen MG, Sneppen K, Sørensen MA, Gerdes K, Mitarai N. Rapid Curtailing of the Stringent Response by Toxin-Antitoxin Module-Encoded mRNases. J Bacteriol. 2016 Jun 27;198(14):1918-1926.

Tsai YC, Du D, Domínguez-Malfavón L, Dimastrogiovanni D, Cross J, Callaghan AJ, García-Mena J, Luisi BF. Recognition of the 70S ribosome and polysome by the RNA degradosome in *Escherichia coli*. *Nucleic Acids Res.* 2012 Nov 1;40(20):10417-31.

Tseng YT, Chiou NT, Gogiraju R, Lin-Chao S. The Protein Interaction of RNA Helicase B (RhB) and Polynucleotide Phosphorylase (PNPase) Contributes to the Homeostatic Control of Cysteine in *Escherichia coli*. *J Biol Chem.* 2015 Dec 11;290(50):29953-63.

Tuckerman JR, Gonzalez G, Gilles-Gonzalez MA. Cyclic di-GMP activation of polynucleotide phosphorylase signal-dependent RNA processing. *J Mol Biol.* 2011 Apr 15;407(5):633-9.

Valencia-Burton M, McCullough RM, Cantor CR, Broude NE. RNA visualization in live bacterial cells using fluorescent protein complementation. *Nat Methods.* 2007 May;4(5):421-7. Epub 2007 Apr 1.

Valencia-Burton M, Shah A, Sutin J, Borogovac A, McCullough RM, Cantor CR, Meller A, Broude NE. Spatiotemporal patterns and transcription kinetics of induced RNA in single bacterial cells. *Proc Natl Acad Sci U S A.* 2009 Sep 22;106(38):16399-404.

Vanderpool CK, Gottesman S. Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system. *Mol Microbiol.* 2004 Nov;54(4):1076-89.

Venturelli OS, Tei M, Bauer S, Chan LJG, Petzold CJ, Arkin AP. Programming mRNA decay to modulate synthetic circuit resource allocation. *Nat Commun.* 2017 Apr 26;8:15128.

Viegas SC, Apura P, Martínez-García E, de Lorenzo V, Arraiano CM. Modulating Heterologous Gene Expression with Portable mRNA-Stabilizing 5'-UTR Sequences. *ACS Synth Biol.* 2018 Sep 21;7(9):2177-2188.

Wagner LA, Gesteland RF, Dayhuff TJ, Weiss RB. An efficient Shine-Dalgarno sequence but not translation is necessary for lacZ mRNA stability in *Escherichia coli*. *J Bacteriol.* 1994 Mar;176(6):1683-8.

Weilbacher T, Suzuki K, Dubey AK, Wang X, Gudapaty S, Morozov I, Baker CS, Georgellis D, Babitzke P, Romeo T. A novel sRNA component of the carbon storage regulatory system of *Escherichia coli*. *Mol Microbiol.* 2003 May;48(3):657-70.

Wong AG, McBurney KL, Thompson KJ, Stickney LM, Mackie GA. S1 and KH domains of polynucleotide phosphorylase determine the efficiency of RNA binding and autoregulation. *J Bacteriol.* 2013 May;195(9):2021-31.

Wu J, Jiang Z, Liu M, Gong X, Wu S, Burns CM, Li Z. Polynucleotide phosphorylase protects *Escherichia coli* against oxidative stress. *Biochemistry.* 2009 Mar 10;48(9):2012-20.

Yamanaka K and Inouye M. Selective mRNA Degradation by Polynucleotide Phosphorylase in Cold Shock Adaptation in *Escherichia coli*. *J Bacteriol.* 2001 May; 183(9): 2808–2816.

Yarchuk O, Jacques N, Guillerez J, Dreyfus M. Interdependence of translation, transcription and mRNA degradation in the lacZ gene. *J Mol Biol.* 1992 Aug 5;226(3):581-96.

Zangrossi S, Briani F, Ghisotti D, Regonesi ME, Tortora P, Dehò G. Transcriptional and post-transcriptional control of polynucleotide phosphorylase during cold acclimation in *Escherichia coli*. *Mol Microbiol.* 2000 Jun;36(6):1470-80.

Zhang Y, Burkhardt DH, Rouskin S, Li GW, Weissman JS, Gross CA. A Stress Response that Monitors and Regulates mRNA Structure Is Central to Cold Shock Adaptation. *Mol Cell*. 2018 Apr 19;70(2):274-286.e7.

Zhou L, Zhang AB, Wang R, Marcotte EM, Vogel C. The proteomic response to mutants of the *Escherichia coli* RNA degradosome. *Mol Biosyst*. 2013 Apr 5;9(4):750-7.

References à rajouter

Gordon GC, Cameron JC, Pflieger BF. 2017. RNA sequencing identifies new RNase III cleavage sites in *Escherichia coli* and reveals increased regulation of mRNA. *mBio* 8:e00128-17. <https://doi.org/10.1128/mBio.00128-17>

Etienne et al JTB 2020

Braun et al 1998, *EMBO J* 17(16):4790-7

MARDLE Identification and analysis of novel small molecule inhibitors of RNase E Implications for antibacterial targeting and regulation of RNase E *Bioch Biophys Rep*2020

Kim et al. 2008 *Genes and Dev* 22, pp 3497-508

Laguerre S., González I., Nouaille S., Moisan A., Villa-Vialaneix N., Gaspin C., Bouvier M., Carpousis A.J., Cocaign-Bousquet M., GIRBAL L. 2018. Large-scale measurement of mRNA degradation in *Escherichia coli*: to delay or not to delay. *Methods in Enzymology*, 612:47-66.

Carrier, T.A., Keasling, J., 1997. Mechanistic modeling of prokaryotic mRNA decay. *J Theor Biol* 189, 195-209.

Deneke, C., Lipowsky, R., Valleriani, A., 2013. Effect of ribosome shielding on mRNA stability. *Phys Biol* 10, 046008.

Mier-y-Terán-Romero, L., Silber, M., & Hatzimanikatis, V. (2013). Mechanistically Consistent Reduced Models of Synthetic Gene Networks. *Biophysical journal*, 104(9), 2098-2109.

Zur & Tuller, 2016 *NAR* <https://doi.org/10.1093/nar/gkw764>;

von der Haar, 2012 <https://doi.org/10.5936/csbj.201204002>

Levine et al. 2007; <https://doi.org/10.1371/journal.pbio.0050229>;

Arbel-Goren et al. 2016 *Nucleic Acids Research*, Volume 44, Issue 14, Pages 6707–6720, <https://doi.org/10.1093/nar/gkw273>;

Lim and Adamson, 2011, *PLoS Comp Biol* <https://doi.org/10.1371/journal.pcbi.1002138>;

Lavi-Itzkovitz et al 2014 *Nucleic Acids Research*, Volume 42, Issue 19, Pages 12200–12211

Friedman et al 2006 *Physical review letters* <https://doi.org/10.1103/PhysRevLett.97.168302>;

McAdams and Arkin, 1997, *PNAS* 94 (3) 814-819 <https://doi.org/10.1073/pnas.94.3.814>

Paulsson, 2005, *Physics of Life Reviews*, Volume 2, Issue 2, Pages 157-175 <https://doi.org/10.1016/j.plrev.2005.03.003>;

Thattai and van Oudenaarden, 2001, PNAS, 98 (15) 8614-8619

<https://doi.org/10.1073/pnas.151588598>;

Kim and Jacobs-Wagner, 2018, Biophysical Journal 1718-1729

<https://doi.org/10.1016/j.bpj.2018.02.010>

Kim, S., and Jacobs-Wagner, C. (2018). Effects of mRNA degradation and site-specific 491transcriptional pausing on protein expression noise. *Biophys. J.*114, 1718-1729

Usuda et al 2010 *Journal of Biotechnology* 147 17–30; Wong et al 2008 *Biotechnol Prog*

<https://doi.org/10.1021/bp970003o>

Mitarai et al 2008 *J Mol Biol* Volume 382, Issue 1, Pages 236-245

<https://doi.org/10.1016/j.jmb.2008.06.068>

Thomas et al, 2018 *Nature Communications* volume 9, Article number: 4528

Karr et al 2012 *Cell* 150, 389–401 doi:10.1016/j.cell.2012.05.044;

Macklin et al., *Science* 369, 390 (2020) Simultaneous cross-evaluation of heterogeneous *E. coli* datasets via mechanistic simulation

LOPEZ et al The C-terminal half of RNase E, which organizes the *Escherichia coli* degradosome, participates in mRNA degradation but not rRNA processing in vivo *Mol Mic* 1999

LEROY et al Function in *Escherichia coli* of the non-catalytic part of RNase E_role in the degradation of ribosome-free mRNA *Mol Mic* 2002

Bibliographie

- D. N. Adamson and H. N. Lim. Essential requirements for robust signaling in Hfq dependent small rna networks. *PLoS Computational Biology*, 7(8) :e1002138, 2011.
- N. Al-Husini, D. T. Tomares, Z. J. Pfaffenberger, N. S. Muthunayake, M. A. Samad, T. Zuo, O. Bitar, J. R. Aretakis, M.-H. M. Bharmal, A. Gega, et al. Br-bodies provide selectively permeable condensates that stimulate mrna decay and prevent release of decay intermediates. *Molecular Cell*, 2020.
- M. Albrecht, D. Stichel, B. Müller, R. Merkle, C. Sticht, N. Gretz, U. Klingmüller, K. Breuhahn, and F. Matthäus. TTCA : an R package for the identification of differentially expressed genes in time course microarray data. *BMC Bioinformatics*, 18(1) :33, 2017.
- R. Arbel-Goren, A. Tal, B. Parasar, A. Dym, N. Costantino, J. Muñoz-García, D. L. Court, and J. Stavans. Transcript degradation and noise of small RNA-controlled genes in a switch activated network in *Escherichia coli*. *Nucleic Acids Research*, 44(14) :6707–6720, 2016.
- C. M. Arraiano, J. M. Andrade, S. Domingues, I. B. Guinote, M. Malecki, R. G. Matos, R. N. Moreira, V. Pobre, F. P. Reis, M. Saramago, et al. The critical role of RNA processing and degradation in the control of gene expression. *FEMS Microbiology Reviews*, 34(5) : 883–923, 2010.
- M. J. Aryee, J. A. Gutiérrez-Pabello, I. Kramnik, T. Maiti, and J. Quackenbush. An improved empirical bayes approach to estimating differential gene expression in microarray time-course data : Betr (bayesian estimation of temporal regulation). *BMC Bioinformatics*, 10 (1) :1–10, 2009.
- D. E. Atkinson. *Cellular energy metabolism and its regulation*. Academic Press, 1977.
- S. Bakshi, H. Choi, J. Mondal, and J. C. Weisshaar. Time-dependent effects of transcription- and translation-halting drugs on the spatial distributions of the *Escherichia coli* chromosome and ribosomes. *Molecular Microbiology*, 94(4) :871–887, 2014.
- K. J. Bandyra, M. Bouvier, A. J. Carpousis, and B. F. Luisi. The social fabric of the rna degradosome. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1829 (6-7) :514–522, 2013.
- K. J. Bandyra, J. M. Wandzik, and B. F. Luisi. Substrate recognition and autoinhibition in the central ribonuclease RNase E. *Molecular Cell*, 72(2) :275–285, 2018.
- Z. Bar-Joseph, A. Gitter, and I. Simon. Studying and modelling dynamic biological processes using time-series gene expression data. *Nature Reviews Genetics*, 13(8) :552–564, 2012.

- D. J. Barr. Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in psychology*, 4 :328, 2013.
- J. A. Bernstein, A. B. Khodursky, P.-H. Lin, S. Lin-Chao, and S. N. Cohen. 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 of the United States of America*, 99(15) :9697–9702, 2002.
- A. Bersani, A. Borri, A. Milanese, G. Tomassetti, and P. Vellucci. A study case for the analysis of asymptotic expansions beyond the tqssa for inhibitory mechanisms in enzyme kinetics. *Communications in Applied and Industrial Mathematics*, 10(1) :162–181, 2019.
- A. Bersani, A. Borri, A. Milanese, G. Tomassetti, and P. Vellucci. Uniform asymptotic expansions beyond the tqssa for the goldbeter–koshland switch. *SIAM Journal on Applied Mathematics*, 80(3) :1123–1152, 2020.
- J. Bertrand, E. Comets, and F. Mentre. Comparison of model-based tests and selection strategies to detect genetic polymorphisms influencing pharmacokinetic parameters. *Journal of biopharmaceutical statistics*, 18(6) :1084–1102, 2008.
- J. A. M. Borghans, R. J. de Boer, and L. A. Segel. Extending the quasi-steady state approximation by changing variables. *Bulletin of Mathematical Biology*, 58(1) :43–63, Jan 1996. ISSN 1522-9602. doi: 10.1007/bf02458281. URL <http://dx.doi.org/10.1007/BF02458281>.
- F. Brauer, C. Castillo-Chavez, and C. Castillo-Chavez. *Mathematical models in population biology and epidemiology*, volume 2. Springer, 2012.
- G. E. Briggs and J. B. S. Haldane. A note on the kinetics of enzyme action. *Biochemical journal*, 19(2) :338–339, 1925.
- A. J. Callaghan, M. J. Marcaida, J. A. Stead, K. J. McDowall, W. G. Scott, and B. F. Luisi. Structure of *Escherichia coli* RNase E catalytic domain and implications for RNA turnover. *Nature*, 437(7062) :1187, 2005.
- D. M. Camacho, K. M. Collins, R. K. Powers, J. C. Costello, and J. J. Collins. Next-generation machine learning for biological networks. *Cell*, 173(7) :1581–1592, 2018.
- A. J. Carpousis. The RNA degradosome of *Escherichia coli* : an mRNA-degrading machine assembled on RNase E. *Annual Review of Microbiology*, 61 :71–87, 2007.
- M. Cassman. Barriers to progress in systems biology. *Nature*, 438(7071) :1079–1079, 2005.
- H. Chen, K. Shiroguchi, H. Ge, and X. S. Xie. Genome-wide study of mRNA degradation and transcript elongation in *Escherichia coli*. *Molecular Systems Biology*, 11(1) :781, 2015.
- A. Ciliberto, F. Capuani, and J. J. Tyson. Modeling networks of coupled enzymatic reactions using the total quasi-steady state approximation. *PLoS Comput Biol*, 3(3) :e45, 2007.
- J. E. Clarke, L. Kime, D. Romero A, and K. J. McDowall. Direct entry by RNase E is a major pathway for the degradation and processing of RNA in *Escherichia coli*. *Nucleic Acids Research*, 42(18) :11733–11751, 2014.

- P. J. Cock, T. Antao, J. T. Chang, B. A. Chapman, C. J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, B. Wilczynski, et al. Biopython : freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11) :1422–1423, 2009.
- E. Comets and F. Mentré. Evaluation of tests based on individual versus population modeling to compare dissolution curves. *Journal of Biopharmaceutical Statistics*, 11(3) :107–123, 2001.
- E. Comets, A. Lavenu, and M. Lavielle. Saemix, an r version of the saem algorithm. In *20th meeting of the Population Approach Group in Europe, Athens, Greece. Abstr*, volume 2173, 2011.
- A. Cornish-Bowden. *Fundamentals of enzyme kinetics*. John Wiley & Sons, 2013a.
- A. Cornish-Bowden. The origins of enzyme kinetics. *FEBS Letters*, 587(17) :2725–2730, 2013b.
- D. Dar and R. Sorek. Extensive reshaping of bacterial operons by programmed mRNA decay. *PLoS genetics*, 14(4) :e1007354, 2018.
- M. Davidian and D. M. Giltinan. *Nonlinear models for repeated measurement data*, volume 62. CRC press, 1995.
- A. Deana and J. G. Belasco. Lost in translation : the influence of ribosomes on bacterial mRNA decay. *Genes & development*, 19(21) :2526–2533, 2005.
- B. Delyon, M. Lavielle, and E. Moulines. Convergence of a stochastic approximation version of the em algorithm. *Annals of statistics*, pages 94–128, 1999.
- A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society : Series B (Methodological)*, 39(1) :1–22, 1977.
- C. Deneke, R. Lipowsky, and A. Valleriani. Complex degradation processes lead to non-exponential decay patterns and age-dependent decay rates of messenger RNA. *PLoS One*, 8(2) :e55442, 2013a.
- C. Deneke, R. Lipowsky, and A. Valleriani. Effect of ribosome shielding on mRNA stability. *Physical Biology*, 10(4) :046008, 2013b.
- L. Dharmarajan, H.-M. Kaltenbach, F. Rudolf, and J. Stelling. A simple and flexible computational framework for inferring sources of heterogeneity from single-cell dynamics. *Cell systems*, 8(1) :15–26, 2019.
- B. Di Camillo, F. Sanchez-Cabo, G. Toffolo, S. K. Nair, Z. Trajanoski, and C. Cobelli. A quantization method based on threshold optimization for microarray short time series. *BMC Bioinformatics*, 6(4) :1–7, 2005.
- D. DiMucci, M. Kon, and D. Segrè. Machine learning reveals missing edges and putative interaction mechanisms in microbial ecosystem networks. *mSystems*, 3(5), 2018.

- S. Donnet and A. Samson. A review on estimation of stochastic differential equations for pharmacokinetic/pharmacodynamic models. *Advanced drug delivery reviews*, 65(7) :929–939, 2013.
- P. Érdi and J. Tóth. *Mathematical models of chemical reactions : theory and applications of deterministic and stochastic models*. Manchester University Press, 1989.
- J. Ernst and Z. Bar-Joseph. Stem : a tool for the analysis of short time series gene expression data. *BMC Bioinformatics*, 7(1) :191, 2006.
- T. Esquerré, S. Laguerre, C. Turlan, A. J. Carpousis, L. Girbal, and M. Coccagn-Bousquet. 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, 2014.
- T. Esquerré, A. Moisan, H. Chiapello, L. Arike, R. Vilu, C. Gaspin, M. Coccagn-Bousquet, and L. Girbal. Genome-wide investigation of mRNA lifetime determinants in *Escherichia coli* cells cultured at different growth rates. *BMC Genomics*, 16(1), Apr 2015. ISSN 1471-2164. doi: 10.1186/s12864-015-1482-8. URL <http://dx.doi.org/10.1186/s12864-015-1482-8>.
- Z. Fang, T. Ma, G. Tang, L. Zhu, Q. Yan, T. Wang, J. C. Celedón, W. Chen, and G. C. Tseng. Bayesian integrative model for multi-omics data with missingness. *Bioinformatics*, 34(22) :3801–3808, 2018.
- M. Frank, C. Lax, S. Walcher, and O. Wittich. Quasi-steady state reduction for the michaelis-menten reaction-diffusion system. *Journal of Mathematical Chemistry*, 56(6) :1759–1781, 2018.
- N. Friedman, L. Cai, and X. S. Xie. Linking stochastic dynamics to population distribution : an analytical framework of gene expression. *Physical Review Letters*, 97(16) :168302, 2006.
- F. Fröhlich, A. Reiser, L. Fink, D. Woschée, T. Ligon, F. J. Theis, J. O. Rädler, and J. Hase-nauer. Multi-experiment nonlinear mixed effect modeling of single-cell translation kinetics after transfection. *NPJ systems biology and applications*, 4(1) :1–12, 2018.
- R. C. Gentleman, V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, et al. Bioconductor : open software development for computational biology and bioinformatics. *Genome Biology*, 5(10) :R80, 2004.
- D. Ghosh and L. M. Poisson. “Omics” data and levels of evidence for biomarker discovery. *Genomics*, 93(1) :13–16, 2009.
- D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25) :2340–2361, 1977.
- P. Girard and F. Mentré. A comparison of estimation methods in nonlinear mixed effects models using a blind analysis. *PAGE 14Abstr*, 834, 2005.
- A. M. González-Vargas, E. Cinquemani, and G. Ferrari-Trecate. Validation methods for population models of gene expression dynamics. *IFAC-PapersOnLine*, 49(26) :114–119, 2016.

- M. G. Grigorov. Analysis of time course omics datasets. In *Bioinformatics for Omics Data*, pages 153–172. Springer, 2011.
- C. Guldberg and P. Waage. Studies concerning affinity. forhandlinger : Videnskabs-selskabet i christinia. *Norwegian Academy of Science and Letters*, 35 :1864, 1864.
- L. Hadjeras, L. Poljak, M. Bouvier, Q. Morin-Ogier, I. Canal, M. Cocaign-Bousquet, L. Girbal, and A. J. Carpousis. Detachment of the rna degradosome from the inner membrane of escherichia coli results in a global slowdown of mrna degradation, proteolysis of rnase e and increased turnover of ribosome-free transcripts. *Molecular microbiology*, 111(6) : 1715–1731, 2019.
- G. W. Hatfield, S.-p. Hung, and P. Baldi. Differential analysis of DNA microarray gene expression data. *Molecular Microbiology*, 47(4) :871–877, 2003.
- R. A. Irizarry, B. Hobbs, F. Collin, Y. D. Beazer-Barclay, K. J. Antonellis, U. Scherf, and T. P. Speed. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, 4(2) :249–264, 2003.
- C. Jain and J. G. Belasco. RNase E autoregulates its synthesis by controlling the degradation rate of its own mRNA in *Escherichia coli* : unusual sensitivity of the *rne* transcript to RNase E activity. *Genes & Development*, 9(1) :84–96, 1995.
- N. D. Jayavelu. *Integrative Systems Biology Approaches for Analyzing High-Throughput Data Applications to modeling of gene regulatory networks*. PhD thesis, Norwegian University of Science and Technology, 2015.
- M. G. Jørgensen, J. S. Pettersen, and B. H. Kallipolitis. sRNA-mediated control in bacteria : An increasing diversity of regulatory mechanisms. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1863(5) :194504, 2020.
- A. A. Kalaitzis and N. D. Lawrence. A simple approach to ranking differentially expressed gene expression time courses through gaussian process regression. *BMC Bioinformatics*, 12(1) :180, 2011.
- M. Karlsson, D. L. Janzén, L. Durrieu, A. Colman-Lerner, M. C. Kjellsson, and G. Cedersund. Nonlinear mixed-effects modelling for single cell estimation : when, why, and how to use it. *BMC systems biology*, 9(1) :1–15, 2015.
- J. R. Karr, K. Takahashi, and A. Funahashi. The principles of whole-cell modeling. *Current opinion in microbiology*, 27 :18–24, 2015.
- W. Kermack and A. McKendrick. Contributions to the mathematical theory of epidemics—iii. further studies of the problem of endemicity. *Bulletin of mathematical biology*, 53(1-2) : 89–118, 1991a.
- W. O. Kermack and A. G. McKendrick. Contributions to the mathematical theory of epidemics. ii.—the problem of endemicity. *Proceedings of the Royal Society of London. Series A, containing papers of a mathematical and physical character*, 138(834) :55–83, 1932.
- W. O. Kermack and A. G. McKendrick. Contributions to the mathematical theory of epidemics—i. *Bulletin of mathematical biology*, 53(1-2) :33–55, 1991b.

- I. M. Keseler, A. Mackie, A. Santos-Zavaleta, R. Billington, C. Bonavides-Martínez, R. Caspi, C. Fulcher, S. Gama-Castro, A. Kothari, M. Krummenacker, M. Latendresse, L. Muñiz-Rascado, Q. Ong, S. Paley, M. Peralta-Gil, P. Subhraveti, D. A. Velázquez-Ramírez, D. Weaver, J. Collado-Vides, I. Paulsen, and P. D. Karp. The EcoCyc database : reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Research*, 45(D1) : D543–D550, 11 2016.
- V. Khemici, L. Poljak, B. F. Luisi, and A. J. Carpousis. The RNase E of *Escherichia coli* is a membrane-binding protein. *Molecular Microbiology*, 70(4) :799–813, 2008.
- J. Kim and J. H. Kim. Difference-based clustering of short time-course microarray data with replicates. *BMC Bioinformatics*, 8(1) :253, 2007.
- S. Kim and C. Jacobs-Wagner. Effects of mRNA degradation and site-specific transcriptional pausing on protein expression noise. *Biophysical Journal*, 114(7) :1718–1729, 2018.
- H. Kitano. Looking beyond the details : a rise in system-oriented approaches in genetics and molecular biology. *Current Genetics*, 41(1) :1–10, 2002a.
- H. Kitano. Systems biology : a brief overview. *Science*, 295(5560) :1662–1664, 2002b.
- E. Klipp, W. Liebermeister, C. Wierling, and A. Kowald. *Systems biology : a textbook*. John Wiley & Sons, 2016.
- L. Kuenzel. Gene clustering methods for time series microarray data. *Biochemistry*, 218, 2010.
- C. Kuppe, M. Sciacovelli, E. Gjerga, K. B. Emdal, D. B. Bekker-Jensen, J. Kranz, E. J. Bindels, S. Costa, J. V. Olsen, C. Frezza, et al. Causal integration of multi-omics data with prior knowledge to generate mechanistic hypotheses. *bioRxiv*, 2020.
- S. Laguerre, I. González, S. Nouaille, A. Moisan, N. Villa-Vialaneix, C. Gaspin, M. Bouvier, A. J. Carpousis, M. Cocaign-Bousquet, and L. Girbal. Large-scale measurement of mRNA degradation in *Escherichia coli* : to delay or not to delay. In *Methods in Enzymology*, volume 612, pages 47–66. Elsevier, 2018.
- T. Lampert, B. Lafabregue, N. Serrette, G. Forestier, B. Crémilleux, C. Vrain, P. Gancarski, et al. Constrained distance based clustering for time-series : a comparative and experimental study. *Data Mining and Knowledge Discovery*, 32(6) :1663–1707, 2018.
- A. Lavi-Itzkovitz, N. Peterman, D. Jost, and E. Levine. Quantitative effect of target translation on small RNA efficacy reveals a novel mode of interaction. *Nucleic Acids Research*, 42(19) :12200–12211, 2014.
- M. Lavielle. *Mixed effects models for the population approach : models, tasks, methods and tools*. CRC press, 2014.
- M. Lavielle and C. Mbogning. An improved saem algorithm for maximum likelihood estimation in mixtures of non linear mixed effects models. *Statistics and Computing*, 24(5) : 693–707, 2014.

- K. Lee, X. Zhan, J. Gao, J. Qiu, Y. Feng, R. Meganathan, S. N. Cohen, and G. Georgiou. RraA : a protein inhibitor of RNase E activity that globally modulates RNA abundance in *E. coli*. *Cell*, 114(5) :623 – 634, 2003.
- S. Y. Lee, D.-Y. Lee, and T. Y. Kim. Systems biotechnology for strain improvement. *Trends in Biotechnology*, 23(7) :349–358, 2005.
- J. T. Leek, E. Mosen, A. R. Dabney, and J. D. Storey. Edge : extraction and analysis of differential gene expression. *Bioinformatics*, 22(4) :507–508, 2006.
- E. Levine, Z. Zhang, T. Kuhlman, and T. Hwa. Quantitative characteristics of gene regulation by small RNA. *PLoS Biol*, 5(9) :e229, 2007.
- A. J. Lotka. Contribution to the theory of periodic reactions. *The Journal of Physical Chemistry*, 14(3) :271–274, 2002.
- G. A. Mackie. RNase E : at the interface of bacterial RNA processing and decay. *Nature Reviews Microbiology*, 11(1) :45, 2013.
- H. H. McAdams and A. Arkin. Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 94(3) :814–819, 1997.
- J. E. McDermott, J. Wang, H. Mitchell, B.-J. Webb-Robertson, R. Hafen, J. Ramey, and K. D. Rodland. Challenges in biomarker discovery : combining expert insights with statistical analysis of complex omics data. *Expert Opinion on Medical Diagnostics*, 7(1) :37–51, 2013.
- I. C. McDowell, D. Manandhar, C. M. Vockley, A. K. Schmid, T. E. Reddy, and B. E. Engelhardt. Clustering gene expression time series data using an infinite gaussian process mixture model. *PLoS Computational Biology*, 14(1) :e1005896, 2018.
- S. Melamed, P. P. Adams, A. Zhang, H. Zhang, and G. Storz. RNA-RNA interactomes of ProQ and Hfq reveal overlapping and competing roles. *Molecular Cell*, 77(2) :411–425, 2020.
- L. Michaelis and M. L. Menten. The kinetics of the inversion effect. *Biochemische Zeitung*, 49 :333–369, 1913.
- L. Mier-y Terán-Romero, M. Silber, and V. Hatzimanikatis. Mechanistically consistent reduced models of synthetic gene networks. *Biophysical Journal*, 104(9) :2098–2109, 2013.
- H. L. Mills, J. Heron, C. Relton, M. Suderman, and K. Tilling. Methods for dealing with missing covariate data in epigenome-wide association studies. *American Journal of Epidemiology*, 188(11) :2021–2030, 2019.
- J. R. Moffitt, S. Pandey, A. N. Boettiger, S. Wang, and X. Zhuang. Spatial organization shapes the turnover of a bacterial transcriptome. *Elife*, 5 :e13065, 2016.
- G. Molenberghs and G. Verbeke. *Models for discrete longitudinal data*. Springer Science & Business Media, 2006.
- K. Moradzadeh, S. Moein, N. Nickaeen, and Y. Gheisari. Analysis of time-course microarray data : Comparison of common tools. *Genomics*, 111(4) :636–641, 2019.

- M. Morin, B. Enjalbert, D. Ropers, L. Girbal, and M. Coccagn-Bousquet. Genomewide stabilization of mRNA during a “feast-to-famine” growth transition in *Escherichia coli*. *mSphere*, 5(3), 2020.
- E. Noor, S. Cherkaoui, and U. Sauer. Biological insights through omics data integration. *Current Opinion in Systems Biology*, 15 :39–47, 2019.
- S. Nouaille, S. Mondeil, A.-L. Finoux, C. Moulis, L. Girbal, and M. Coccagn-Bousquet. The stability of an mRNA is influenced by its concentration : a potential physical mechanism to regulate gene expression. *Nucleic Acids Research*, 45(20) :11711–11724, 2017.
- X. Panhard and A. Samson. Extension of the saem algorithm for nonlinear mixed models with 2 levels of random effects. *Biostatistics*, 10(1) :121–135, 2009.
- J. Paulsson. Models of stochastic gene expression. *Physics of Life Reviews*, 2(2) :157–175, 2005.
- S. D. Peddada, E. K. Lobenhofer, L. Li, C. A. Afshari, C. R. Weinberg, and D. M. Umbach. Gene selection and clustering for time-course and dose–response microarray experiments using order-restricted inference. *Bioinformatics*, 19(7) :834–841, 2003.
- M. G. Pedersen and A. M. Bersani. Introducing total substrates simplifies theoretical analysis at non-negligible enzyme concentrations : pseudo first-order kinetics and the loss of zero-order ultrasensitivity. *Journal of Mathematical Biology*, 60 (2) :267–283, Mar 2009. ISSN 1432-1416. doi: 10.1007/s00285-009-0267-6. URL <http://dx.doi.org/10.1007/s00285-009-0267-6>.
- M. G. Pedersen, A. M. Bersani, E. Bersani, and G. Cortese. The total quasi-steady-state approximation for complex enzyme reactions. *Mathematics and Computers in Simulation*, 79(4) :1010–1019, 2008.
- M. G. Pedersena, A. M. Bersanib, and E. Bersanic. The total quasi-steady-state approximation for fully competitive enzyme reactions. *Bulletin of mathematical biology*, 69(1) :433, 2007.
- N. Perakakis, A. Yazdani, G. E. Karniadakis, and C. Mantzoros. Omics, big data and machine learning as tools to propel understanding of biological mechanisms and to discover novel diagnostics and therapeutics. *Metabolism-Clinical and Experimental*, 87 :A1–A9, 2018.
- J. C. Pinheiro and D. M. Bates. Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of computational and Graphical Statistics*, 4(1) :12–35, 1995.
- J. Qian, M. Dolled-Filhart, J. Lin, H. Yu, and M. Gerstein. Beyond synexpression relationships : local clustering of time-shifted and inverted gene expression profiles identifies new, biologically relevant interactions. *Journal of Molecular Biology*, 314(5) :1053–1066, 2001.
- I. Rajman. Pk/pd modelling and simulations : utility in drug development. *Drug discovery today*, 13(7-8) :341–346, 2008.
- P. Régnier and C. M. Arraiano. Degradation of mRNA in bacteria : emergence of ubiquitous features. *Bioessays*, 22(3) :235–244, 2000.

- J. Richards and J. G. Belasco. Obstacles to scanning by RNase E govern bacterial mRNA lifetimes by hindering access to distal cleavage sites. *Molecular Cell*, 74(2) :284–295, 2019.
- T. Romeo and P. Babitzke. Global regulation by csra and its rna antagonists. *Regulating with RNA in Bacteria and Archaea*, pages 339–354, 2018.
- A. Santiago-Frangos and S. A. Woodson. Hfq chaperone brings speed dating to bacterial sRNA. *Wiley Interdisciplinary Reviews : RNA*, 9(4) :e1475, 2018.
- R. M. Savic, F. Mentré, and M. Lavielle. Implementation and evaluation of the saem algorithm for longitudinal ordered categorical data with an illustration in pharmacokinetics–pharmacodynamics. *The AAPS journal*, 13(1) :44–53, 2011.
- M. Schena, D. Shalon, R. W. Davis, and P. O. Brown. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, 270(5235) :467–470, 1995.
- S. Schnell. Enzyme kinetics at high enzyme concentration. *Bulletin of Mathematical Biology*, 62(3) :483–499, May 2000. ISSN 0092-8240. doi: 10.1006/bulm.1999.0163. URL <http://dx.doi.org/10.1006/bulm.1999.0163>.
- S. Schnell and P. K. Maini. Enzyme kinetics far from the standard quasi-steady-state and equilibrium approximations. *Mathematical and Computer Modelling*, 35(1-2) :137–144, 2002.
- S. Schnell and C. Mendoza. Closed form solution for time-dependent enzyme kinetics. *Journal of theoretical Biology*, 187(2) :207–212, 1997.
- S. Schnell and C. Mendoza. Enzyme kinetics of multiple alternative substrates. *Journal of Mathematical Chemistry*, 27(1/2) :155–170, 2000. ISSN 0259-9791. doi: 10.1023/a :1019139423811. URL <http://dx.doi.org/10.1023/A:1019139423811>.
- L. Scrucca, M. Fop, T. B. Murphy, and A. E. Raftery. mclust 5 : clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal*, 8(1) :289–317, 2016. URL <https://doi.org/10.32614/RJ-2016-021>.
- I. H. Segel. *Enzyme kinetics : behavior and analysis of rapid equilibrium and steady state enzyme systems*, volume 115. Wiley New York :, 1975.
- L. A. Segel and M. Slemrod. The quasi-steady-state assumption : a case study in perturbation. *SIAM review*, 31(3) :446–477, 1989.
- O. Shalem, O. Dahan, M. Levo, M. R. Martinez, I. Furman, E. Segal, and Y. Pilpel. Transient transcriptional responses to stress are generated by opposing effects of mrna production and degradation. *Molecular systems biology*, 4(1) :4, 2008.
- C. Sin, D. Chiarugi, and A. Valleriani. Degradation parameters from pulse-chase experiments. *PLoS One*, 11(5) :e0155028, 2016.
- G. K. Smyth. Limma : linear models for microarray data. In *Bioinformatics and computational biology solutions using R and Bioconductor*, pages 397–420. Springer, 2005.

- H. Strahl, C. Turlan, S. Khalid, P. J. Bond, J.-M. Kebalo, P. Peyron, L. Poljak, M. Bouverier, L. Hamoen, B. F. Luisi, et al. Membrane recognition and dynamics of the RNA degradosome. *PLoS Genetics*, 11(2) :e1004961, 2015.
- P. S. Swain, M. B. Elowitz, and E. D. Siggia. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 99(20) :12795–12800, 2002.
- Y. Tai. Timecourse : statistical analysis for developmental microarray time course data. *R package version*, 1(0), 2007.
- J. Tang and W. Riley. A total quasi-steady-state formulation of substrate uptake kinetics in complex networks and an example application to microbial litter decomposition. *Network*, 10 :10615–10683, 2013.
- A. L. Tarca, V. J. Carey, X.-w. Chen, R. Romero, and S. Drăghici. Machine learning and its applications to biology. *PLoS Computational Biology*, 3(6) :e116, 2007.
- M. Terzer, N. D. Maynard, M. W. Covert, and J. Stelling. Genome-scale metabolic networks. *Wiley Interdisciplinary Reviews : Systems Biology and Medicine*, 1(3) :285–297, 2009.
- A. Tzafiri and E. Edelman. The total quasi-steady-state approximation is valid for reversible enzyme kinetics. *Journal of Theoretical Biology*, 226(3) :303 – 313, 2004. ISSN 0022-5193. doi: <https://doi.org/10.1016/j.jtbi.2003.09.006>. URL <http://www.sciencedirect.com/science/article/pii/S0022519303003382>.
- K. Valgepea, K. Adamberg, A. Seiman, and R. Vilu. *Escherichia coli* achieves faster growth by increasing catalytic and translation rates of proteins. *Molecular BioSystems*, 9(9) : 2344–2358, 2013.
- E. O. Voit, H. A. Martens, and S. W. Omholt. 150 years of the mass action law. *PLoS Comput Biol*, 11(1) :e1004012, 2015.
- V. Volterra. *Variazioni e fluttuazioni del numero d'individui in specie animali conviventi*. C. Ferrari, 1927.
- P. Waage and C. M. Gulberg. Studies concerning affinity. *Journal of chemical education*, 63 (12) :1044, 1986.
- T. Wada and A. Becskei. Impact of methods on the measurement of mRNA turnover. *International Journal of Molecular Sciences*, 18(12) :2723, 2017.
- D. Wang and S. Bodovitz. Single cell analysis : the new frontier in ‘omics’. *Trends in Biotechnology*, 28(6) :281–290, 2010.
- R.-S. Wang, A. Saadatpour, and R. Albert. Boolean modeling in systems biology : an overview of methodology and applications. *Physical Biology*, 9(5) :055001, 2012.
- Z. B. Weinstein, A. Bender, and M. Cokol. Prediction of synergistic drug combinations. *Current Opinion in Systems Biology*, 4 :24–28, 2017.
- D. J. Wilkinson. *Stochastic modelling for systems biology*. CRC press, 2018.

- C. Xiao-Xue, X. Liu-Fang, and S. Hua-Lin. Theoretical studies on sRNA-mediated regulation in bacteria. *Chinese Physics B*, 24(12) :128703, 2015.
- J. T. Yurkovich and B. O. Palsson. Quantitative-omic data empowers bottom-up systems biology. *Current Opinion in Biotechnology*, 51 :130–136, 2018.
- I. S. L. Zeng and T. Lumley. Review of statistical learning methods in integrated omics studies (an integrated information science). *Bioinformatics and Biology Insights*, 12 :1–16, 2018.
- H. Zur and T. Tuller. Predictive biophysical modeling and understanding of the dynamics of mRNA translation and its evolution. *Nucleic Acids Research*, 44(19) :9031–9049, 2016.