



**HAL**  
open science

# Modèles physiques pour l'étude de la cinétique de transport des ribosomes sur l'ARN messenger lors de la traduction génétique et l'interprétation des données de ribosome profiling

Carole Chevalier

► **To cite this version:**

Carole Chevalier. Modèles physiques pour l'étude de la cinétique de transport des ribosomes sur l'ARN messenger lors de la traduction génétique et l'interprétation des données de ribosome profiling. Génétique. Université de Montpellier, 2022. Français. NNT : 2022UMONS075 . tel-04095361v2

**HAL Id: tel-04095361**

**<https://theses.hal.science/tel-04095361v2>**

Submitted on 11 May 2023

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

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

# THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En physique

École doctorale 166 Information, Structures et Systèmes (I2S)

Unité de recherche : Laboratoire Charles Coulomb

Modèles physiques pour l'étude de la cinétique de  
transport des ribosomes sur l'ARN messager lors de la  
traduction génétique et l'interprétation des données de  
*ribosome profiling*

Présentée par Carole CHEVALIER

Le 15 décembre 2022

Sous la direction de Andrea PARMEGGIANI  
et Jean-Charles WALTER

Devant le jury composé de

Didier FELBACQ, professeur, Laboratoire Charles Coulomb (L2C)

Aleksandra WALCZAK, directrice de recherche, ENS Paris

Marc LEFRANC, professeur, Laboratoire de Physique des Lasers, Atomes et Molécules (PhLAM)

Adélaïde RAGUIN, *Junior Group Leader*, Institute for Computational Cell Biology

Jean-Charles WALTER, chargé de recherche, Laboratoire Charles Coulomb (L2C)

Andrea PARMEGGIANI, professeur, Laboratoire Charles Coulomb (L2C)

Président du jury

Rapporteuse

Rapporteur

Examinatrice

Co-directeur de thèse

Directeur de thèse



UNIVERSITÉ  
DE MONTPELLIER





## Résumé

L'environnement intracellulaire est densément peuplé et extrêmement organisé. Plusieurs processus de transport y ont lieu : la marche des moteurs moléculaires sur les microtubules et les microfilaments d'actine, ainsi que la translocation d'autres moteurs moléculaires comme l'ARN polymérase sur l'ADN et les ribosomes sur l'ARN messager (ARNm). Dans cette thèse, nous étudions la translocation du ribosome sur l'ARNm permettant la traduction du code génétique en protéines et qui, d'un point de vue physique, est un processus de transport dirigé à une dimension. Il s'agit d'un phénomène non-linéaire, stochastique et hors-équilibre thermodynamique qui relève de la physique statistique hors-équilibre.

La traduction génétique est une étape de l'expression des gènes régulée par plusieurs mécanismes qui sont encore mal connus. En apportant une étude physique de la cinétique de traduction, nous espérons contribuer à une meilleure compréhension de la régulation de la production des protéines pendant cette étape. Une telle compréhension devrait impacter également la recherche médicale en ce qui concerne les maladies liées à une dérégulation de l'expression génétique comme le cancer et les maladies neurodégénératives.

Un premier chapitre est consacré au contexte biologique de la traduction ainsi qu'à la brève description de l'expérience de ribosome profiling, une méthode de séquençage haut débit effectuée par nos collègues biologistes. Elle permet, depuis une dizaine d'années, une avancée considérable de la recherche sur la traduction, mais la complexité du système mesuré nécessite un travail de modélisation physico-mathématique plus poussé.

Le deuxième chapitre décrit le modèle paradigmatique des transports dirigés à une dimension, le modèle de gaz sur réseau Totally Asymmetric Simple Exclusion Process (TASEP) et ses variantes, plus proches de la translocation ribosomale, sont étudiés. Nous montrons qu'une particule étendue engendre des effets d'exclusions non négligeables en phase basse densité. En particulier, l'exclusion au départ de la traduction a l'effet le plus important sur la densité moyenne le long du réseau lorsque le rapport du taux d'initiation sur le taux de saut de la particule est inférieur ou égal à 0,1, tandis que lorsque ce rapport est plus élevé, un effet de bord en sortie du réseau devient aussi non négligeable.

Dans le troisième chapitre, nous modélisons l'ensemble d'ARNm, séquencé dans l'expérience de ribosome profiling (ou Ribo-Seq), par un modèle balistique où les ribosomes sont des particules ponctuelles se déplaçant de manière déterministe sur des filaments ayant, ce qui est rarement considéré, une durée de vie finie. Nous montrons que, pour des paramètres ayant des valeurs biologiques typiques, l'effet de temps de vie fini sur la densité de ribosomes est négligeable lorsque l'on considère l'ensemble des ARNm, les polysomes, mais qu'il devient important lorsque l'on considère des populations d'ARNm séparées par leur nombre de ribosomes, les k-somes. Nous caractérisons précisément cet effet par une analyse paramétrique, ce qui nous permet de définir trois régimes de dégradation de l'ARNm.

Dans le dernier chapitre, nous mettons en évidence pour la première fois à notre connaissance cet effet de temps de vie fini sur des profils Ribo-Seq de monosomes, disomes, trisomes et tétrasomes (k-somes pour  $k=1,2,3$  et 4). Nous comparons ensuite notre modèle avec plusieurs milliers de profils Ribo-Seq effectués par nos collaborateurs biologistes de l'IGF et bioinformaticiens du LIRMM pour observer que notre modèle donne une bonne description des profils k-somes. Enfin, nous proposons dans ce chapitre une méthode de déduction des paramètres cinétiques de la traduction, le taux d'initiation et le taux de saut du ribosome à l'aide de la densité monosome et de la densité polysome.

## Abstract

The intracellular environment is densely populated and extremely organized. Several transport processes take place within this environment : the walking and translocation of molecular motors on microfilaments, as well as on DNA and messenger RNA (mRNA). In this thesis we study the translocation of ribosomes on mRNA ,which permits the translation of the genetic code into proteins and which from a physical point of view is an example of a one dimensional directed transport process. It is a non-linear, stochastic and non-equilibrium thermodynamic phenomenon that belongs to non-equilibrium statistical physics. Gene translation is the major steps involved in gene expression, which is regulated by several mechanisms that are still poorly understood. By providing a physical study of translation kinetics, we hope to contribute to a better understanding of the regulation of protein production during this step. Such an understanding should also impact medical research on diseases related to gene expression deregulation such as cancer and neurodegenerative diseases.

A first chapter is devoted to the biological context of translation and to a brief description of ribosome profiling experiments, a deep sequencing method performed by our biologist colleagues. It has allowed, over the last ten years, considerable progress in translation research, but the complexity of the studied system requires a more advanced physico-mathematical modeling approach.

The second chapter describes the paradigmatic model of one-dimensional directed transport, the Totally Asymmetric Simple Exclusion Process (TASEP) lattice gas model, and its variants, closer to the actual ribosomal translocation process, are studied. We show that an extended particle generates non-negligible exclusion effects in the low density phase. In particular, exclusion at the translation start (initiation) has the largest effect on the average density along the lattice when the ratio of the initiation rate to the hopping rate of the particle is less than or equal to 0.1, while when this ratio is higher, such an edge effect at the lattice exit also becomes non-negligible.

In the third chapter, we model the set of mRNAs, sequenced in the ribosome profiling experiment (or Ribo-Seq), by a ballistic model where ribosomes are point particles moving deterministically on filaments having, what is rarely considered, a finite lifetime. We show that, for parameters with typical biological values, the finite lifetime effect is negligible when considering the whole set of mRNAs (polysomes), but becomes important when considering populations of mRNAs sorted by their number of ribosomes, the k-somes. We precisely characterize this effect by a parametric analysis of the model and define three regimes of degradation depending on to the translation initiation rate, the degradation rate of mRNAs , and the time for the ribosomes to cross the mRNA.

In the last chapter, we demonstrate, for the first time to our knowledge, this finite lifetime effect on the Ribo-Seq profiles of monosomes, disomes, trisomes and tetrasomes (k-somes for  $k=1,2,3$  and 4). We then compare our model with several thousands of Ribo-Seq profiles obtained by our collaborators (biologists from the IGF and bioinformaticians from the LIRMM). We then compare our model with several thousand Ribo-Seq profiles performed by our biologist collaborators at IGF and bioinformaticians at LIRMM to observe that our model gives a good description of the k-some profiles. Finally, in this chapter we propose a method for inferring the kinetic parameters of translation, the initiation rate and the ribosome hopping rate using the monosome and the polysome densities.



# Remerciements

*Je tiens tout d'abord à remercier le jury qui m'a donné le sentiment d'avoir réellement mérité le titre de docteur.*

*Je remercie ma famille et tout particulièrement Geoffrey qui a été d'un soutien précieux et déterminant par sa présence chaleureuse et son écoute. Merci à Yasashi qui a adouci la période d'écriture avec son ronronnement apaisant et sa fourrure douillette.*

*Cette thèse n'aurait pas pu se faire sans le soutien du Professeur Nicolas Destainville qui m'a permis de surmonter mes difficultés, aussi grâce à la mission handicap de l'université Paul Sabatier, et qui m'a recommandé lors de ma recherche de projet de thèse. La reconnaissance de ma situation de handicap n'aurait pu se faire sans l'insistance du docteur Boyer-Vidal malgré ma propre réticence. Merci au CNRS de m'avoir choisi pour leur contrat doctoral handicap et à toutes les personnes et institutions qui se battent pour l'inclusion des personnes en situation de handicap.*

*Je remercie toute l'équipe SCPN du laboratoire Charles Coulomb dans laquelle j'ai effectué ma thèse. Je suis particulièrement reconnaissante à Jérôme Dorignac qui a beaucoup contribué à cette thèse. Merci à Jean-Charles Walter et Andrea Parmeggiani de m'avoir proposé cette thèse. L'écoute et la considération d'Andrea en particulier m'ont permis de la mener terme. Merci à John Palmeri de s'être particulièrement impliqué bien que n'étant pas un de mes superviseurs. Je remercie également Frédéric Geniet pour sa participation et son soutien.*

*Merci aux collaborateurs biologistes et bioinformaticiens, Alexandre David, Armelle Choquet, Éric Rivals et Julie Ripoll, de m'avoir confié leurs données pour que je puisse en extraire la biophysique moelle. Merci de leur patience lors des discussions riches d'échanges de connaissances et d'expertises.*

*Enfin, je me dois de remercier la musique et surtout les nombreux musiciens qui la créent, du métal au classique en passant par la pop et le jazz, qui masse mon cerveau et mes émotions dès que le besoin s'en fait ressentir.*

Dr. Carole Chevalier



*À ma sœur, Sophie, qui avait de grandes ambitions.*

# Table des matières

Résumé . . . . .	iv
Abstract . . . . .	v
<b>Table des matières</b>	<b>x</b>
<b>Liste des figures</b>	<b>xii</b>
<b>Liste des tableaux</b>	<b>xiv</b>
<b>Liste des symboles</b>	<b>xv</b>
<b>Introduction</b>	<b>1</b>
<b>Chapitre 1 : Contexte Biologique</b>	<b>3</b>
1 Introduction . . . . .	4
2 Expression de l'information génétique : de l'ADN aux protéines . . . . .	4
3 L'expérience de <i>ribosome profiling</i> . . . . .	16
4 Conclusion . . . . .	17
5 Références . . . . .	17
<b>Chapitre 2 : Modélisation du transport des ribosomes par gaz sur réseaux : état de l'art et résultats préliminaires</b>	<b>21</b>
1 Introduction . . . . .	22
2 Historique et état de l'art des modèles physiques de la traduction génétique . . . . .	22
3 TASEP . . . . .	23
4 Évolution du modèle : $\ell$ -TASEP inhomogène . . . . .	31
5 Résultats préliminaires à propos de l'effet de la particule étendue sur la densité de particules . . . . .	35
6 Simulations Monte Carlo . . . . .	41
7 Conclusion et perspectives . . . . .	42
8 Références . . . . .	43
<b>Chapitre 3 : Modélisation d'un ensemble d'ARN messagers à temps de vie finis</b>	<b>47</b>
1 Introduction . . . . .	48
2 Séparation par nombre de ribosomes . . . . .	48
3 Distribution des âges des ARNm . . . . .	49
4 Traduction modélisée par TASEP inhomogène . . . . .	51
5 Modèle de transport balistique . . . . .	54
6 Analyse de l'effet de durée de vie finie . . . . .	59
7 Diagramme de phase . . . . .	69
8 Conclusion et perspectives . . . . .	70
9 Références . . . . .	71
<b>Chapitre 4 : Étude des données de ribosome profiling et méthode de fit</b>	<b>75</b>
1 Introduction . . . . .	76
2 Les profils de densité polysome n'apportent pas assez d'information . . . . .	76

3	Information apportée par les densités <i>k</i> -somes . . . . .	77
4	Comparaison avec les profils <i>k</i> -somes obtenus par l'expérience Ribo-Seq sur un gène	78
5	Analyse des profils Ribo-Seq sur les <i>k</i> -somes lissés . . . . .	80
6	Déduction des paramètres cinétiques de la traduction des densités polysomes et monosomes . . . . .	85
7	Conclusion et perspectives . . . . .	88
8	Références . . . . .	89
	<b>Conclusion</b>	<b>91</b>
	<b>Annexes</b>	<b>I</b>
1	Calculs annexes . . . . .	I
2	Tableaux et figures annexes . . . . .	III
3	Références . . . . .	X

# Liste des figures

1	Illustration par David S. Goodsell de l'environnement cellulaire avec ses processus de transport. . . . .	2
1.1	Cellules eucaryote et procaryote . . . . .	4
1.2	Structure en double hélice de l'ADN . . . . .	5
1.3	Modèle de l'expression génétique . . . . .	6
1.4	Schéma de la transcription . . . . .	6
1.5	Image de microscopie électronique de la transcription . . . . .	7
1.6	L'ARN messenger procaryote et eucaryote . . . . .	8
1.7	L'ARN messenger procaryote polycistronique . . . . .	8
1.8	L'épissage de l'ARN messenger . . . . .	8
1.9	Le code génétique . . . . .	9
1.10	Deux vues de la structure cristallographique du ribosome procaryote . . . . .	10
1.11	Distribution des longueurs des empreintes de ribosomes . . . . .	10
1.12	Structure de l'ARN de transfert . . . . .	11
1.13	Schéma de la traduction . . . . .	12
1.14	Mécanisme de la traduction . . . . .	12
1.15	Image de microscopie électronique de la traduction . . . . .	12
1.16	Schéma de l'initiation de la traduction . . . . .	13
1.17	Schéma de la terminaison de la traduction . . . . .	13
1.18	Différents facteurs biophysiques de régulation de la traduction . . . . .	14
1.19	Schéma de principe de la technique Ribo-seq . . . . .	16
1.20	Profil Ribo-Seq . . . . .	17
2.1	Schéma du modèle TASEP avec bords ouverts . . . . .	24
2.2	Schéma du processus markovien pour un TASEP de $N = 2$ sites. . . . .	25
2.3	Temps de calcul des densités du TASEP exact en fonction du nombre de sites $N$ du réseau . . . . .	27
2.4	Les 3 phases de TASEP . . . . .	28
2.5	Comparaison des profils de densité TASEP en phase basse et haute densité obtenus en simulation Monte Carlo et en champ moyen. . . . .	31
2.6	Schéma $\ell$ -TASEP pour $\ell = 5$ sites . . . . .	32
2.7	Diagramme de phases du $\ell$ -TASEP homogène . . . . .	35
2.8	Comparaison entre le $\ell$ -TASEP homogène en simulation Monte Carlo et les équations de champ moyen . . . . .	37
2.9	Schéma $\ell$ -TAESEP pour $\ell = 5$ sites . . . . .	38
2.10	Comparaison des profils de densités TAESEP obtenus en simulation Monte Carlo et en formulation analytique exacte. . . . .	39
2.11	Comparaison des densités moyennées sur le réseau pour différentes valeurs de $\alpha$ et $\ell$ . . . . .	40
3.1	Schéma de la séparation des polysomes et monosomes par gradient de sucrose. . . . .	49
3.2	Exemple de mesure d'absorbance du gradient de sucrose contenant les polysomes et monosomes . . . . .	49
3.3	Distribution des temps de demi-vies des ARNm . . . . .	50

3.4	Schéma du modèle de transport balistique incluant la dégradation . . . . .	54
3.5	Diagramme spatio-temporel illustrant les trajectoires des ribosomes entrant à des moments aléatoires sur un brin d'ARNm né à $t_b = 0$ . . . . .	55
3.6	Diagramme spatio-temporel illustrant les trajectoires des ribosomes entrant à des moments aléatoires dans un brin d'ARNm né à $t_b = 0$ . . . . .	56
3.7	Tracé de la fonction d'échelle $f(\tilde{\omega})$ . . . . .	61
3.8	Comparaison de la densité polysome à durée de vie finie et infinie . . . . .	62
3.9	La distribution de population $P_k$ . . . . .	63
3.10	Profils de densité des $k$ -somes et du polysome . . . . .	66
3.11	Courbes de croisement $\mathcal{R}_k(0) = 1$ pour $k$ certaines densités (lignes solides : $k = 1$ (rouge), $k = 2$ (bleu), $k = 3$ (vert), $k = 4$ (noir)). Les lignes pointillées de même couleur sont les expressions asymptotiques données par (3.84) à partir de $\tilde{\alpha} = k$ . . . . .	68
3.12	Contributions transitoire et stationnaire de la densité de monosomes $\rho_1(x)$ pour les trois régimes illustrés dans la Fig. 3.10 . . . . .	69
3.13	Diagramme de phase du modèle balistique dans le plan $(\tilde{\alpha}, \tilde{\omega})$ . . . . .	70
4.1	Test de fit des densités $k$ -some du modèle balistique homogène avec les équations du modèle . . . . .	78
4.2	Fit homogène des données de l'expérience Ribo-Seq sur les $k$ -somes $\rho_k$ ( $k=1$ à 4) lissées, avec $\alpha$ et $\omega$ fixés aux valeurs trouvées dans la littérature. . . . .	79
4.3	Fit homogène des données de l'expérience Ribo-Seq sur les $k$ -somes $\rho_k$ ( $k=1$ à 4) lissées, sans fixer $\alpha$ et $\omega$ . . . . .	80
4.4	Histogrammes des erreurs faites sur le fit des données $k$ -somes lissées sur 19 codons par les densités du modèle balistique en considérant une vitesse $p$ constante le long de l'ARNm. . . . .	82
4.5	Histogrammes des erreurs faites sur le fit des données $k$ -somes lissées sur 19 codons par les densités du modèle balistique en considérant une vitesse $p$ constante le long de l'ARNm . . . . .	83
4.6	La fonction score du fit avec le modèle balistique lorsque $\alpha = 0.08$ et $p_i \in [0.2, 4[$ . . .	87
4.7	Diagramme de phase du modèle balistique dans le plan $(\tilde{\alpha}, \tilde{\omega})$ 3.13 ajusté à la faisabilité du fit . . . . .	88
A.1	Diagramme de phase du $\ell$ -TASEP inhomogène . . . . .	III
A.2	Densités monosomes pour $\mathcal{R}_1(0) = \mathcal{F}_1(0)/\mathcal{S}_1 = 1.3$ correspondant à l'effet de durée de vie finie faible sur les $k$ -somes et pour $\mathcal{R}_1(0) = 1.3 \times 10^4$ correspondant à l'effet de durée de vie finie élevé sur les $k$ -somes . . . . .	IV
A.3	Densités Ribo-Seq de monosomes fittées par le modèle balistique homogène. . . . .	V
A.4	Densités Ribo-Seq de disomes fittées par le modèle balistique homogène. . . . .	VI
A.5	Densités Ribo-Seq de trisomes fittées par le modèle balistique homogène. . . . .	VII
A.6	Densités Ribo-Seq de tétrasomes fittées par le modèle balistique homogène. . . . .	VIII
A.7	Densités Ribo-Seq de $k$ -somes, pour $k = 1$ à 4, fittées par le modèle balistique homogène. . . . .	IX

# Liste des tableaux

3.1	Taux cinétiques de traduction et propriétés des ARNm chez les espèces pour lesquelles ces valeurs ont été mesurées, à notre connaissance . . . . .	60
3.2	Paramètres sans dimension chez les espèces pour lesquels les valeurs des taux d'initiation et d'élongation de la traduction, et les temps de demi-vies des ARNm ont été mesurés à notre connaissance . . . . .	60
A.1	Nombre des gènes de l'expérience Ribo-Seq sur les $k$ -somes analysés dans la construction des histogrammes 4.4 et 4.5. . . . .	IV

# Liste des symboles

- $a$  Age de l'ARNm au temps d'observation  $t_{ob}$ .
- $\alpha$  Taux d'entrée sur le réseau / taux d'initiation de la traduction.
- $\tilde{\alpha}$  Taux d'initiation adimensionnel,  $\tilde{\alpha} = \alpha \mathcal{T}(L)$ .
- $\beta$  Taux de sortie du réseau / taux de terminaison de la traduction.
- $\ell$  Longueur de la particule en nombre de sites.
- $J$  Courant de particules.
- $k$  Nombre de ribosomes sur l'ARNm.
- $\langle k \rangle$  Nombre de ribosomes moyen sur l'ARNm.
- $\langle k \rangle_\infty$  Nombre de ribosomes moyen sur l'ARNm à temps de vie infini.
- $L$  Longueur de l'ARNm.
- $\omega$  Taux de dégradation de l'ARNm.
- $\tilde{\omega}$  Taux de dégradation adimensionnel,  $\tilde{\omega} = \omega \mathcal{T}(L)$ .
- $p$  Taux de saut / taux d'élongation homogène de la traduction.
- $\Phi_A(a)$  Distribution des âges des ARNm au temps d'observation  $t_{ob}$ .
- $p_i$  Taux de saut / taux d'élongation inhomogène de la traduction au site  $i$ .
- $P_k$  Probabilité qu'un ARNm soit un  $k$ -some.
- $P(k|a)$  Probabilité qu'un ARNm d'âge  $a$  soit un  $k$ -some.
- $\tilde{p}(x)$  Taux de saut / d'alongation inhomogène au point  $x$  adimensionnel,  $\tilde{p}(x) = p(x) \mathcal{T}(L)$ .
- $\rho_i$  Densité de particules au site  $i$  / densité de ribosomes au codon  $i$ .
- $\rho_k(x)$  Densité de ribosomes à la position  $x$  du  $k$ -some.
- $\rho_k(x|a)$  Densité de ribosomes à la position  $x$  du  $k$ -some d'âge  $a$ .
- $\mathcal{R}_k(0)$  Mesure de l'effet de temps de vie fini sur la densité  $k$ -some.
- $t_b$  Instant de naissance de l'ARNm.
- $t_{ob}$  Instant d'observation.
- $\theta$  Durée de vie de l'ARNm.
- $\mathcal{T}(x)$  Temps nécessaire au ribosome pour aller du point 0 au point  $x$  de l'ARNm.



# Introduction

« En effet l'homme sage - pour connaître manifestement tout ce qui se peut dire contre lui, pour défendre sa position contre tous les contradicteurs, pour savoir que loin d'éviter les objections et les difficultés, il les a recherchées et n'a négligé aucune lumière susceptible d'éclairer tous les aspects du sujet - l'homme sage a le droit de penser que son jugement vaut mieux que celui d'un autre ou d'une multitude qui n'ont pas suivi le même processus. »

---

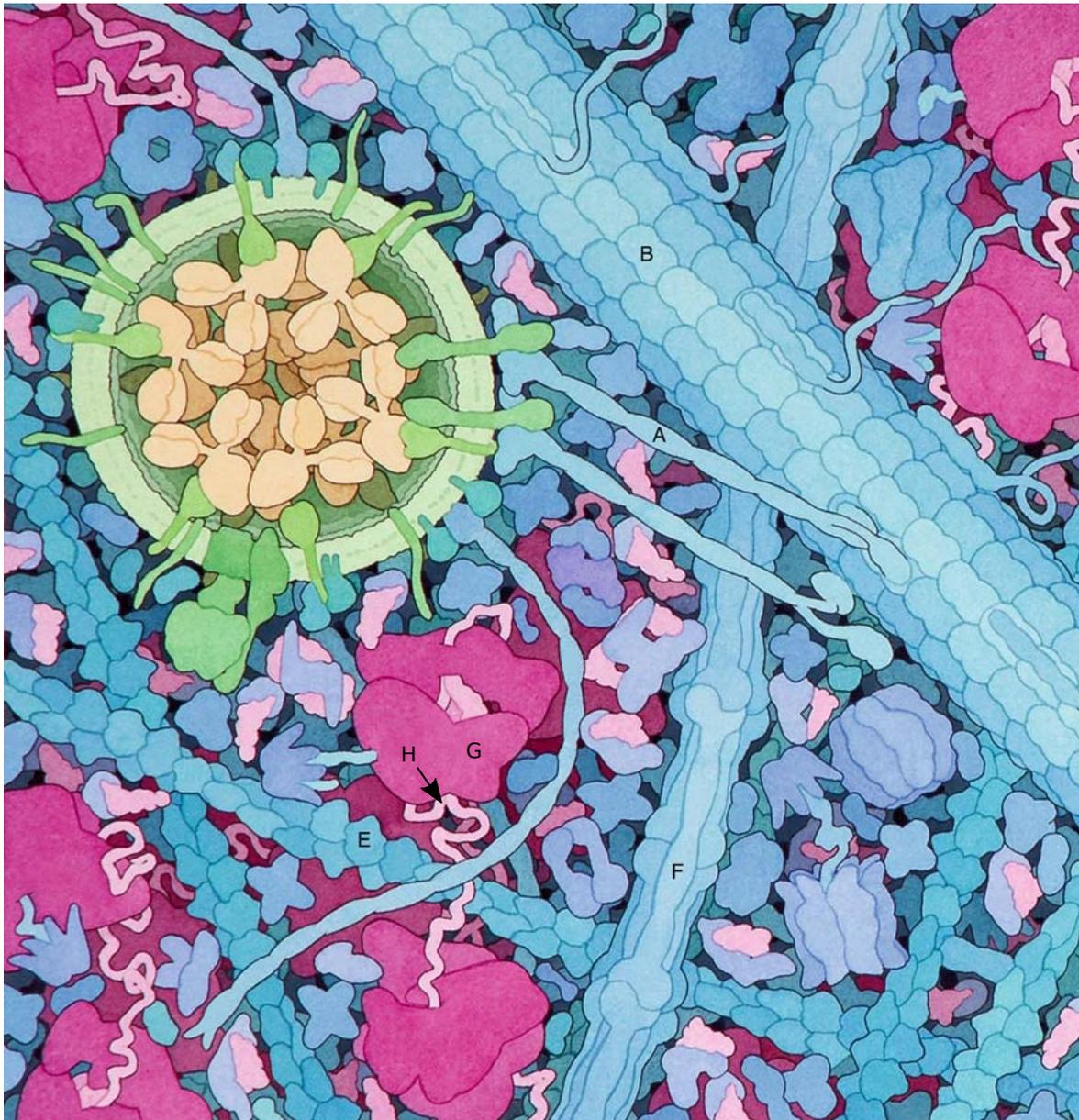
John Stuart Mill (1859),  
*De la liberté* (trad. Laurence  
Lenglet, 1990)

Cette thèse est née dans le contexte interdisciplinaire d'une collaboration entre des biologistes de l'IGF, des informaticiens du LIRMM et des physiciens du Laboratoire Charles Coulomb autour du thème de la génomique. On s'intéresse particulièrement à la régulation de l'expression des protéines à partir du génome, un processus fondamental sans lequel aucune cellule ni organisme biologique ne peuvent fonctionner. En effet, la régulation de l'expression génétique permet la différenciation des cellules d'un organisme ainsi que l'adéquation des cellules à leur environnement.

Dans cette thèse nous allons étudier la traduction génétique qui d'un point de vue physique est un processus stochastique de transport dirigé à une dimension. Comment un tel processus peut-il être efficace dans l'environnement intracellulaire très encombré (Fig. 1)? Cette question fascine la communauté des biophysiciens.

De plus, la traduction génétique est une étape de l'expression des gènes régulée par plusieurs mécanismes qui sont encore mal connus. En apportant une étude physique de la cinétique de traduction, nous espérons contribuer à une compréhension complète de la régulation de la production des protéines pendant cette étape. La régulation de l'expression génétique est aussi un sujet central du domaine médical. Sa compréhension précise devrait impacter également la recherche médicale en ce qui concerne les maladies liées à une dysrégulation de l'expression génétique comme le cancer et les maladies neurodégénératives.

Dans le premier chapitre, nous allons préciser le contexte biologique de cette thèse. Un chapitre sera ensuite consacré au modèle paradigmatique des transports dirigés à une dimension, TASEP (*Totally asymmetric simple exclusion process*) et ses variantes. Dans le troisième chapitre, nous étudierons un ensemble d'ARNm à temps de vie fini afin de pouvoir caractériser les données de l'expérience de séquençage à haut débit des ARNm nommée *ribosome profiling*. Ces données seront ensuite comparées à la modélisation dans le dernier chapitre et une méthode de déduction des paramètres cinétiques de la traduction sera proposée.



**FIGURE 1** – Illustration par David S. Goodsell de l'environnement cellulaire avec des processus de transport.<sup>1</sup> "The cytoplasm is criss-crossed with a network of filaments that form a cytoskeleton that supports the cell and provides a railway for delivery of materials. These include thin actin filaments (E), thicker intermediate filaments (F), and huge microtubules (B)."<sup>2</sup> La vésicule en vert contient des anticorps qui doivent être livrés hors de la cellule. Elle est transportée par les kinésines (A), des moteurs moléculaires qui se déplacent le long des microtubules (B). La traduction est représentée en rose avec le ribosome (G) qui traduit l'ARN messager (H). De nombreuses autres protéines assurent le fonctionnement de la cellule et en font un environnement très dense.

1. Goodsell, David S. 2009. *The Machinery of Life*. Second Edition. Springer Science & Business Media.

2. *Ibid.* (Il s'agit notamment de minces filaments d'actine (E), de filaments intermédiaires plus épais (F) et de volumineux microtubules (B) [traduction libre]).

# Chapitre 1

## Contexte Biologique

« *Il y a plus de choses sur la terre et dans le ciel, Horatio, qu'il n'en est rêvé dans votre philosophie.* »

---

William Shakespeare (1601),  
*Hamlet* (trad. François-Victor  
Hugo, 1866)

### Sommaire

---

<b>1</b>	<b>Introduction</b>	<b>4</b>
<b>2</b>	<b>Expression de l'information génétique : de l'ADN aux protéines</b>	<b>4</b>
2.1	Cellules eucaryotes et procaryotes	4
2.2	Acide désoxyribonucléique	5
2.3	Schéma général de l'expression génétique	5
2.4	Aperçu de la transcription	6
2.5	Focus sur la traduction	9
2.5.1	Le code génétique	9
2.5.2	Le ribosome et l'ARN de transfert	9
2.5.3	Initiation, élongation et terminaison de la traduction	11
2.6	Régulation de l'expression génétique	14
<b>3</b>	<b>L'expérience de <i>ribosome profiling</i></b>	<b>16</b>
<b>4</b>	<b>Conclusion</b>	<b>17</b>
<b>5</b>	<b>Références</b>	<b>17</b>

---

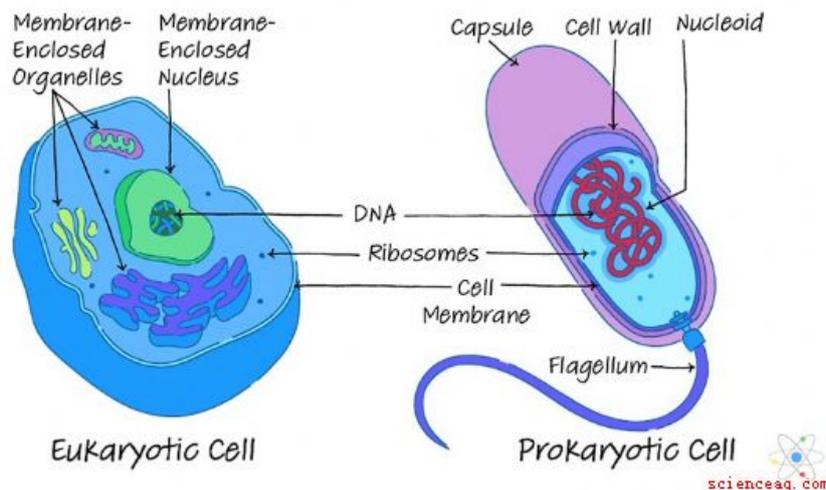
## 1 Introduction

En première partie de ce chapitre, nous allons présenter les phénomènes fondamentaux de biologie moléculaire qui sont à la base du vivant, en se focalisant particulièrement sur la traduction des gènes en protéines qui sera modélisée et étudiée d'un point de vue physique tout au long de cette thèse. Dans une deuxième partie, nous allons décrire les expériences appelées *ribosome profiling* (ou Ribo-Seq) faites par nos collaborateurs biologistes, car elles constituent l'objet de modélisation principal de cette thèse.

## 2 Expression de l'information génétique : de l'ADN aux protéines

### 2.1 Cellules eucaryotes et procaryotes

La cellule est un compartiment délimité par une membrane lipidique dans lequel sont regroupés tous les acteurs moléculaires du vivant. Tout organisme vivant est soit une cellule isolée, soit une association de plusieurs cellules. Il existe deux grands types de cellules. Les premières apparues au cours de l'évolution sont les cellules procaryotes. Ce sont des organismes unicellulaires qui regroupent les bactéries et les archées tandis que les cellules eucaryotes peuvent être unicellulaires, comme les levures, ou pluricellulaires comme les plantes et les animaux. Les cellules eucaryotes constituent les organismes les plus évolués. Comme on peut le voir sur la figure 1.1, les cellules procaryotes se différencient des eucaryotes principalement par l'absence de noyau et d'organites (structures subcellulaires). Les cellules procaryotes sont aussi habituellement plus petites puisque leur taille typique est de 1 – 10 $\mu$ m de diamètre alors que les eucaryotes ont plutôt un diamètre de l'ordre de 10 – 100 $\mu$ m. Dans les cellules eucaryotes, le noyau protège les molécules d'ADN stockées sous forme de chromosomes. Tandis que, dans les cellules procaryotes, l'ADN stocké est une molécule circulaire généralement unique, le nucléoïde. Des copies de cet ADN, appelées plasmides, sont cependant présentes et destinées à être transmises à d'autres bactéries.

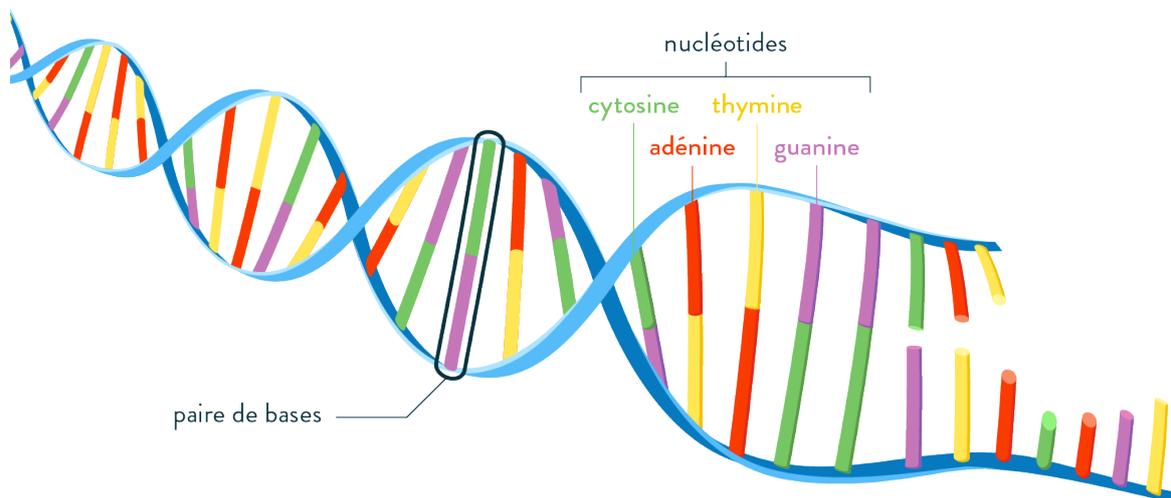


**FIGURE 1.1** – Cellules eucaryote (*eukaryotic cell* en anglais) et procaryote (*prokaryotic cell* en anglais). Les cellules procaryotes se différencient des eucaryotes principalement par l'absence de noyau et d'organites (structures subcellulaires, appelée *organelles* en anglais). Le nucléoïde (*nucleoid* en anglais) est la région où se trouve le matériel génétique. Ici, la cellule procaryote a été représentée avec un flagelle (*flagellum* en anglais) bien qu'il ne soit pas présent chez toutes les cellules procaryotes et que certaines cellules eucaryotes en possèdent également.

## 2.2 Acide désoxyribonucléique

ADN est l'abréviation d'acide désoxyribonucléique. Cette molécule (ou ensemble de molécules) est le support de l'information génétique, c'est-à-dire qu'elle contient la liste de tous les gènes de l'organisme, appelé génome. Les gènes sont encodés par une séquence de molécules appelées nucléotides ou bases azotées. Il existe 4 bases différentes dans la composition de l'ADN : adénine, thymine, cytosine et guanine abrégées par A,T,C et G. La structure en double hélice caractéristique de l'ADN est la conséquence de l'appariement en paires des bases. L'adénine s'apparie avec la thymine et la cytosine s'apparie avec la guanine (Fig. 1.2). L'étude de [MANDELKERN et al. \[1981\]](#) mesurant les paramètres géométriques de l'ADN en solution donne un diamètre de 2,2 à 2,6 nm avec une longueur par nucléotide de 0,33 nm. L'affinité entre bases est fondamentale, car elle va permettre d'une part la construction de séquences complémentaires afin de transmettre l'information génétique aux cellules produites, c'est le processus de réplication de l'ADN, et d'autre part de produire les protéines qui assurent une multitude de fonctions au sein de la cellule, par les processus de transcription et de traduction que nous décrivons dans les parties suivantes.

Ces deux processus constituent les deux étapes fondamentales de l'expression de l'information génétique. C'est le mécanisme auquel s'intéresse particulièrement l'équipe de physiciens théoriciens Systèmes Complexes et Phénomènes non Linéaires (SCPN) au sein de laquelle j'ai effectué ma thèse.



**FIGURE 1.2** – Structure en double hélice de l'ADN. Les brins sont constitués d'un squelette pentose-phosphate et de 4 bases azotées : adénine (A), thymine (T), guanine (G) et cytosine (C). Ils vont se lier par l'appariement des bases A avec les bases T et des bases C avec les bases G grâce à des liaisons hydrogènes. © Schoolmouv.

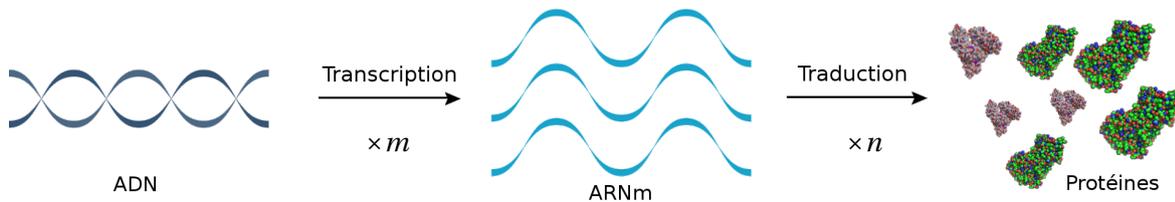
## 2.3 Schéma général de l'expression génétique

L'expression génétique s'effectue en deux étapes importantes résumées en figure 1.3. D'abord, les gènes stockés par l'ADN devant être exprimés sont transcrits en chaînes nucléotidiques appelées acides ribonucléiques messagers (ARNm) par appariement des bases. Il s'agit de copies simples brins de portions de l'ADN ayant une durée de vie de quelques minutes à quelques heures. Dans le cas des eucaryotes, les ARNm sont synthétisés dans le noyau de la cellule, puis sont transloqués hors du noyau avant d'être traduit en protéines. Plus précisément, la traduction désigne celle de la séquence nucléotidique en séquence d'acides aminés qui formera la protéine en se repliant.

Lors de ces deux phénomènes, de nombreux mécanismes de régulations peuvent intervenir et faire varier l'expression des gènes d'une cellule à l'autre. Beaucoup de ces phénomènes sont encore mal compris et font l'objet de nombreux travaux de recherches. Nous ne pouvons pas pré-

dire à l'heure actuelle le nombre de copies d'un même ARNm qui va être produit à partir d'un gène présent sur l'ADN, ni le nombre de protéines qui vont être produites à partir d'un seul brin d'ARNm. Or, le bon fonctionnement de la cellule dépend très fortement des concentrations en les différentes protéines ce qui fait de la régulation de leur production un processus profondément fondamental de la vie cellulaire.

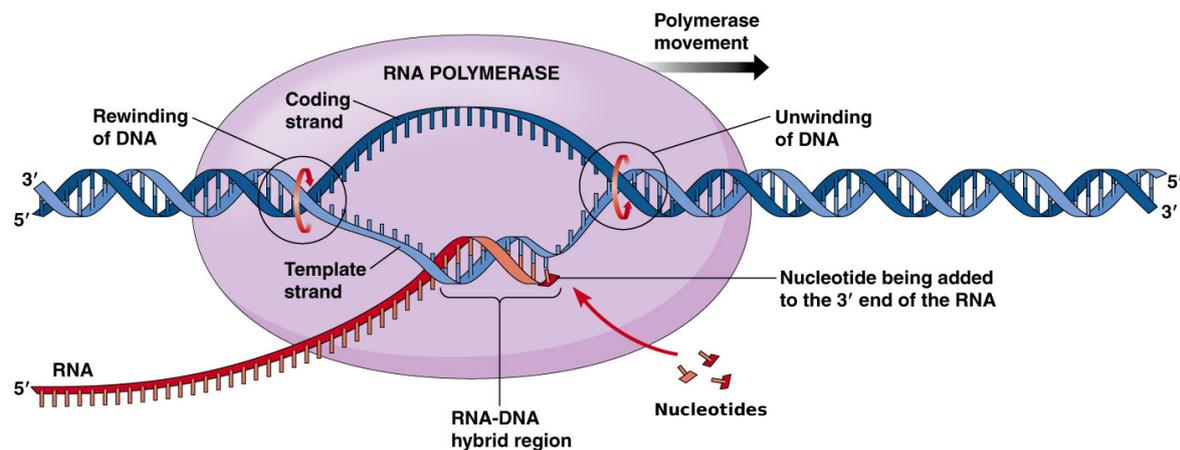
La transcription de l'ADN et la traduction de l'ARNm vont être détaillées dans les parties suivantes. Le sujet de ma thèse portant sur la traduction, un niveau de détails minimal sera apporté à la transcription tandis que le processus de traduction sera plus explicité.



**FIGURE 1.3** – Modèle de l'expression génétique. La double hélice d'ADN contient des gènes qui vont être transcrits en ARNm. Chaque ARNm va être traduit en protéine. Le nombre d'ARNm  $m$  produits à partir d'un gène de l'ADN et le nombre de protéines  $n$  produites à partir d'un gène de l'ADN ne peuvent pas être prédits en l'état actuel des connaissances scientifiques.

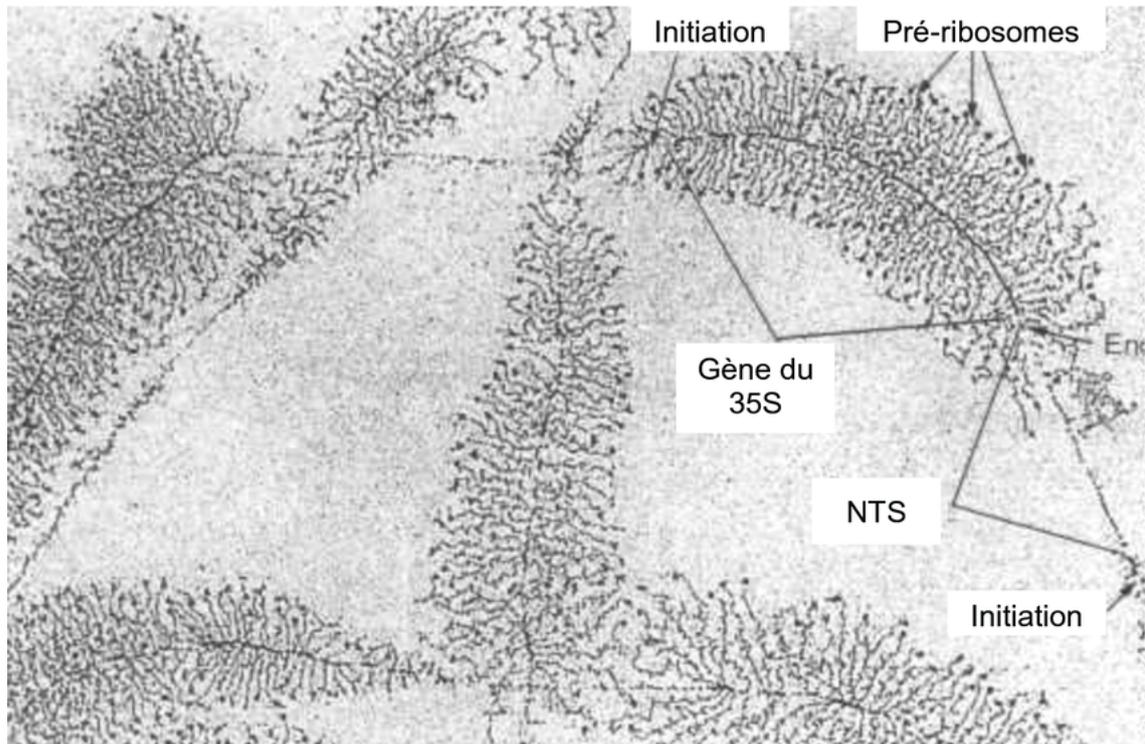
## 2.4 Aperçu de la transcription

**Le processus.** Pour effectuer la transcription de l'ADN en ARN, un complexe macromoléculaire, l'ARN polymérase, va ouvrir la double hélice, et associer à chaque nucléotide d'un brin de l'ADN, le nucléotide complémentaire (Fig. 1.4). Une séquence de l'ADN sert de signal d'initiation de la transcription et une autre de signal de terminaison de la transcription. Afin de transcrire la totalité de la portion d'ADN compris entre ces deux séquences signalisatrices, l'ARN polymérase va se déplacer le long de la double hélice, engendrant des torsions qui vont être corrigées par les enzymes topoisomérases. Il est intéressant de pouvoir observer la transcription en microscopie électronique (Fig. 1.5).



**FIGURE 1.4** – Schéma de la transcription. L'ARN polymérase ouvre la double hélice d'ADN afin de synthétiser un brin d'ARN messenger complémentaire à partir d'un brin de l'ADN originel. Dans la zone hybride, les nucléotides sont recrutés pour allonger l'ARNm. L'ouverture de l'ADN nécessite le déroulement de l'hélice au front et lors de la progression de l'ARN polymérase, l'hélice va s'enrouler à nouveau à l'arrière. © 2012 Pearson Education, Inc.

**L'ARN.** L'ARN n'a en réalité pas exactement les mêmes bases que l'ADN. Le nucléotide uracile, noté U, remplace le nucléotide T dans l'appariement avec le nucléotide A. L'ADN a bénéficié de

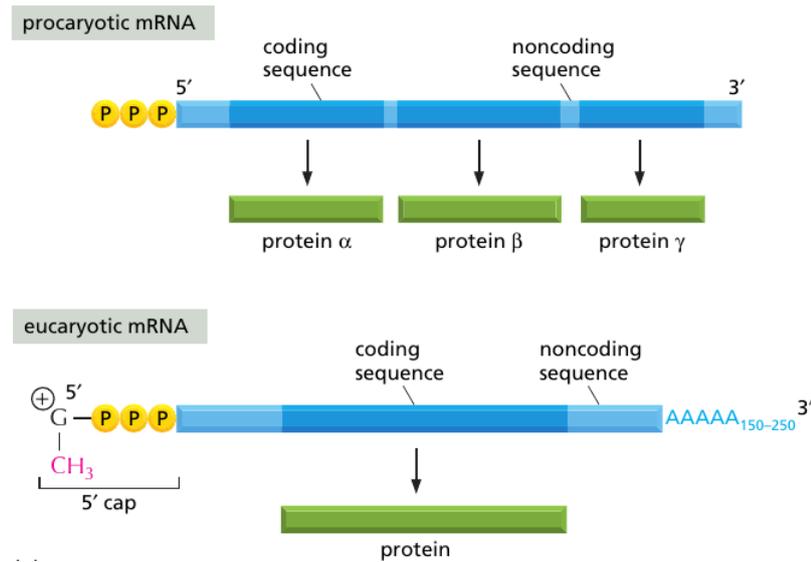


**FIGURE 1.5** – Image de microscopie électronique à transmission de la transcription d'ADN ribosomique [CROSS, 1995]. Le complexe à une forme caractéristique dite en "arbre de Noël", ou "plume". "Le sommet de l'arbre correspond au site d'initiation de la transcription, les branches sont les transcrits primaires d'ARNr, dont la maturation commence" [LEBRETON, 2006]. Les régions NTS (Non-Transcribed Spacer) sont des régions non codantes entre gènes.

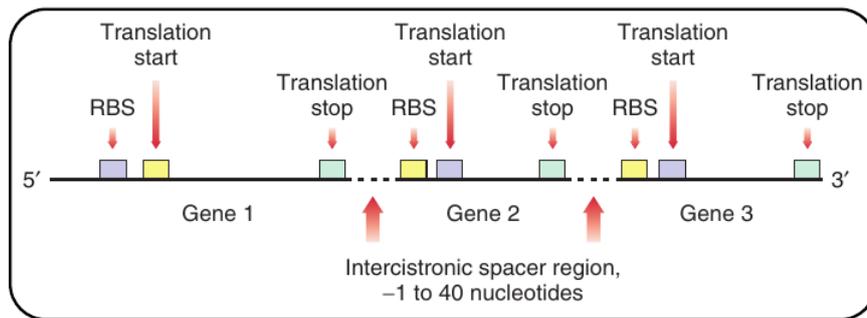
cette évolution le rendant plus stable, ce qui est particulièrement important pour son rôle de conservation de l'information génétique [POOLE et al., 2001].

Tous les ARN ne sont pas des ARN messagers. Il existe deux autres types principaux d'ARN, qui ne sont pas destinés à être traduits, mais sont des acteurs fondamentaux de la traduction : l'ARN ribosomique et l'ARN de transfert. Leur rôle est décrit dans la partie suivante.

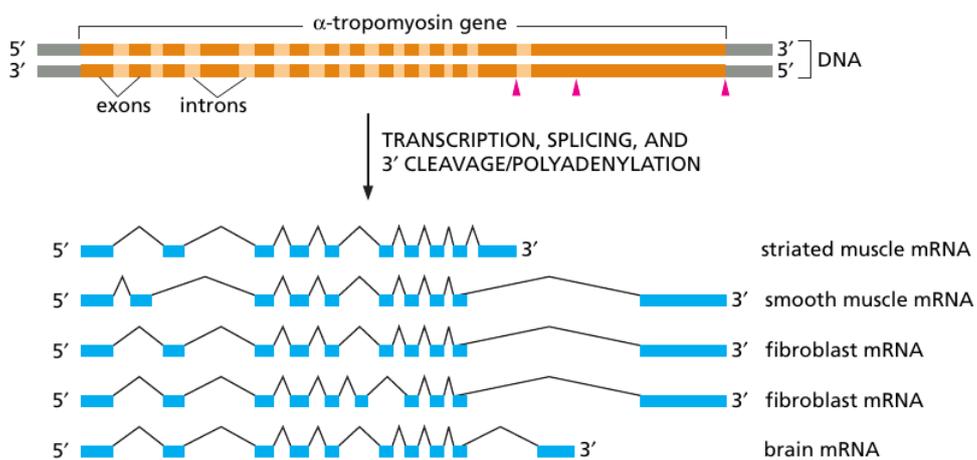
Concernant les ARN messagers, il existe une différence importante entre ceux des cellules eucaryotes et ceux des cellules procaryotes. D'abord, un seul ARNm code une seule protéine à la fois chez les eucaryotes, alors qu'un ARNm procaryote peut coder plusieurs protéines à la fois. On dit que les ARNm eucaryotes sont "monocistroniques" tandis que les ARNm procaryotes peuvent être "polycistroniques". Les séquences de plusieurs protéines se succèdent sur un même ARNm eucaryote en étant séparés par des parties non codantes comme schématisé sur la figure 1.6 [KUSHNER, 2018]. Pour chaque gène, on retrouve un site d'attachement du ribosome (RBS pour *Ribosome Binding Site*) et codon *start* et un codon *stop* 1.7. Les séquences sur un même ARNm peuvent donc être traduites simultanément. Nous précisons cela car un autre processus permet la traduction de protéines différentes à partir d'un seul ARNm. Ce processus est présent chez les procaryotes et les eucaryotes. Il s'agit de séquences "alternatives" qui correspondent à un décalage d'un ou deux nucléotides de sorte que la séquence de codons se trouve changée [ORR et al., 2020]. De plus, contrairement aux ARNm procaryotes, les ARNm eucaryotes issus de la transcription ne sont pas les ARNm finaux qui vont être traduits en protéines. Il existe un processus appelé "épissage" durant lequel des parties non codantes, appelées introns, de l'ARNm vont être soustraites. Les parties codantes, appelées exons, vont ensuite pouvoir être reliées afin de former l'ARN messager mature prêt à être traduit. Ce processus d'épissage permet d'obtenir des protéines différentes à partir du même gène en utilisant certains exons et pas d'autres (Fig. 1.8). Il serait aussi un avantage évolutif en favorisant l'apparition de nouvelles protéines grâce à des exons de différents gènes [ALBERTS et al., 2015].



**FIGURE 1.6** – L'ARN messenger procaryote est généralement polycistronique, il peut contenir les séquences codantes de plusieurs gènes à la suite, tandis que l'ARNm eucaryote est monocistronique [ALBERTS et al., 2015].



**FIGURE 1.7** – Sur l'ARN messenger procaryote polycistronique un site de liaison du ribosome (RBS), un codon *start* et un codon *stop* est présent sur chaque gène, ce qui permet leur traduction simultanée [KUSHNER, 2018].



**FIGURE 1.8** – L'épissage de l'ARN messenger [ALBERTS et al., 2015]. Exemple du gène  $\alpha$ -tropomyosine du rat. Chez les eucaryotes, l'ARNm est composé de parties non codantes, les introns, qui séparent les parties codantes, les exons. L'épissage consiste à retirer les introns et à recombinaison les exons. Il existe plusieurs recombinaisons possibles permettant de produire des ARNm mature différents et donc des protéines différentes.

## 2.5 Focus sur la traduction

### 2.5.1 Le code génétique

Pour former les protéines, la séquence nucléotidique de l'ARNm va être traduite en séquence d'acides aminés. Il existe donc une traduction du gène en protéine. Cette traduction est donnée par le code génétique représenté en figure 1.9. Chaque triplet de nucléotides, appelé codon, code pour un acide aminé ou le codon stop qui termine le gène. Une propriété remarquable du code génétique est qu'il est dégénéré. C'est-à-dire que plusieurs codons peuvent coder le même acide aminé. Parmi les 20 acides aminés, seuls la méthionine et le tryptophane ne sont codés que par un seul codon. La recherche des raisons de cette dégénérescence est un champ très riche en biologie moléculaire et bénéficie de l'apport considérable de la bioinformatique pour l'analyse des combinaisons de codons formant les gènes.

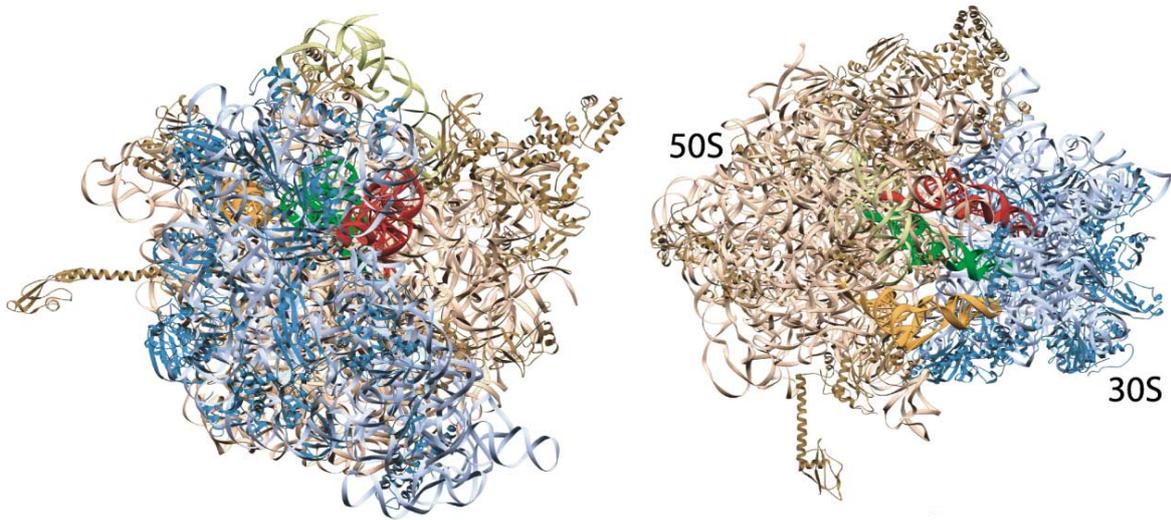
		ARN messenger Codon : deuxième base azotée				
		U	C	A	G	
ARN messenger Codon : première base azotée	U	Phe	Ser	Tyr	Cys	U
		Phe	Ser	Tyr	Cys	C
		Leu	Ser	STOP	STOP	A
		Leu	Ser	STOP	Trp	G
	C	Leu	Pro	His	Arg	U
		Leu	Pro	His	Arg	C
		Leu	Pro	Gln	Arg	A
		Leu	Pro	Gln	Arg	G
	A	Ile	Thr	Asn	Ser	U
		Ile	Thr	Asn	Ser	C
		Ile	Thr	Lys	Arg	A
		Met	Thr	Lys	Arg	G
	G	Val	Ala	Asp	Gly	U
		Val	Ala	Asp	Gly	C
		Val	Ala	Glu	Gly	A
		Val	Ala	Glu	Gly	G
		ARN messenger Codon : troisième base azotée				

**FIGURE 1.9** – Le code génétique. Les codons de l'ARNm formés par trois bases azotées A,U,C et G codent pour les 20 acides aminés Glycine (Gly), Proline (Pro), Alanine (Ala), Valine (Val), Leucine (Leu), Isoleucine (Ile), Methionine (Met), Cysteine (Cys), Phenylalanine (Phe), Tyrosine (Tyr), Tryptophan (Trp), Histidine (His), Lysine (Lys), Arginine (Arg), Glutamine (Gln), Asparagine (Asn), Glutamic Acid (Glu), Aspartic Acid (Asp), Serine (Ser), Threonine (Thr) et le signal d'arrêt de la traduction (codon STOP). Image de Quo-Fata FERUNT - CC BY-SA.

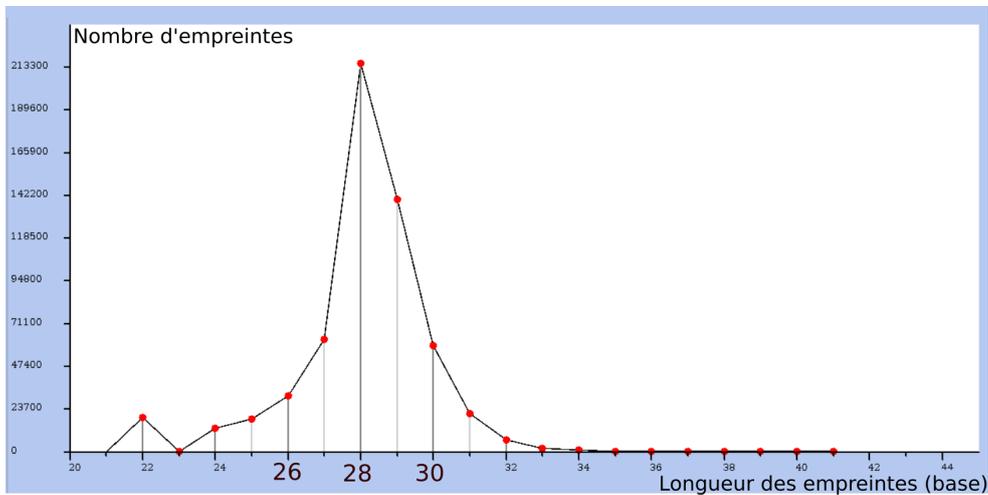
### 2.5.2 Le ribosome et l'ARN de transfert

La traduction est guidée par un complexe macromoléculaire appelé ribosome. La structure des ribosomes procaryotes a été déterminée à l'échelle atomique en l'an 2000 (Fig. 1.10) tandis que celle des ribosomes eucaryotes a été déterminée en 2010, montrant une grande similitude avec celle des procaryotes. Le ribosome est composé d'ARN ribosomique et de protéines. Il comporte deux sous-unités se fixant de part et d'autre de l'ARNm, mais détachées l'une de l'autre quand le ribosome n'est pas en train de traduire. Bien que le ribosome possède 3 sites d'interaction avec les codons, il en recouvre en réalité 9 environ chez les eucaryotes. Nos collaborateurs ont développé un logiciel permettant de visualiser les données dites "Ribo-Seq", que nous allons décrire ensuite,

appelé RNA-Ribo Explorer (RRE) [PAULET et al., 2021]. Avec ce logiciel, nous pouvons observer la taille des ribosomes. Nous donnons un exemple avec la distribution des tailles des ribosomes sur un gène humain en Fig. 1.11. Chez les deux types cellulaires, les ribosomes traduisant l'ARNm sont libres dans le cytoplasme, mais dans les eucaryotes certains sont liés à la membrane du réticulum endoplasmique, qu'on appelle alors réticulum endoplasmique granuleux.

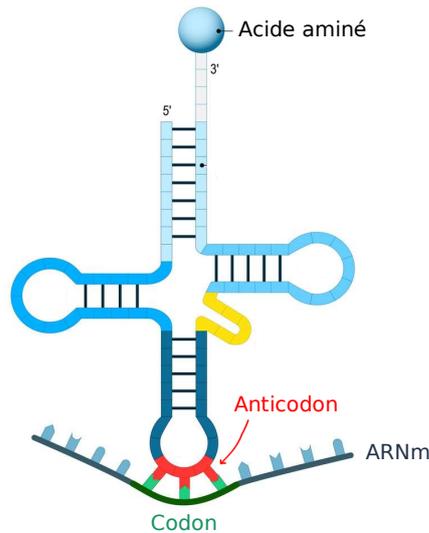


**FIGURE 1.10** – Deux vues de la structure cristallographique du ribosome procaryote donné par YUSUPOV et al. [2001]. On distingue la grosse sous-unité 50S, la petite sous unité 30S, ainsi que les trois sites accueillant les ARNt (site A en rouge, site P en vert et site E en jaune).



**FIGURE 1.11** – Distribution des longueurs des empreintes de ribosomes pour le gène humain SRR18N2152 obtenue par le traitement des données Ribo-Seq par le logiciel RNA-Ribo Explorer (RRE) [PAULET et al., 2021]. On en déduit que le ribosome recouvre environ 28 nucléotides, soit 9 - 10 codons.

Les acides aminés formant la protéine sont appariés aux codons par une autre molécule, l'ARN de transfert (ou ARNt). Il s'agit d'un ARN dont la séquence s'organise en forme de trèfle et comporte le codon complémentaire à celui correspondant à l'acide aminé qui doit être apporté. Celui-ci est porté par une extrémité de l'ARNt comme indiqué sur la figure 1.12.



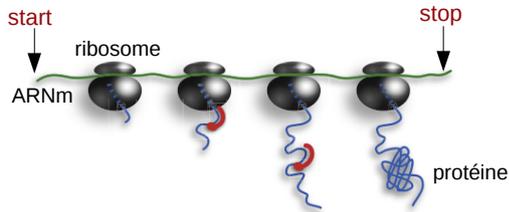
**FIGURE 1.12** – Structure de l'ARN de transfert. L'ARN de transfert (ARNt) apporte l'acide aminé correspondant au codon en train d'être traduit par le ribosome. Grâce à l'anticodon constitué des bases complémentaires au codon, il le reconnaît et se fixe sur celui-ci. © designua, Fotolia.

### 2.5.3 Initiation, élongation et terminaison de la traduction

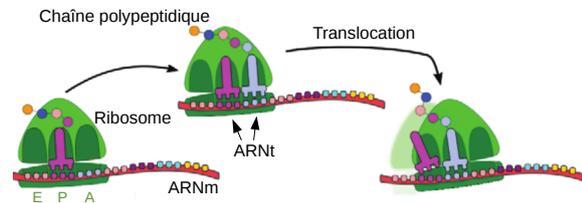
Un gène codant pour une protéine va être lu à partir du codon "start" jusqu'au codon "stop" (Fig. 1.13). Pendant la traduction, à chaque triplet de nucléotide, nommé "codon", sera associé un acide aminé, la brique élémentaire des protéines. Le ribosome effectue la traduction séquentielle de l'ARNm grâce à 3 sites d'interactions, désigné E, P et A.

Pour que la traduction commence, les deux extrémités de l'ARNm doivent être jointes grâce aux facteurs eIF4. La petite sous-unité du ribosome portant l'ARNt initiateur au niveau du site A et le facteur d'initiation eIF2 va s'attacher à l'extrémité 5' de l'ARNm et se déplacer le long de l'ARNm jusqu'à ce que l'ARNt initiateur se retrouve sur le codon *start* AUG. Les facteurs d'initiations vont alors se dissocier de la petite sous-unité pour laisser place à la grande sous-unité (Fig. 1.16). Les deux sous-unités du ribosome ainsi rassemblées, la traduction de la séquence commence avec l'arrivée du second ARNt avec le second acide aminé au site A du ribosome. Le ribosome va catalyser la liaison peptique entre le nouvel acide aminé et celui porté par l'ARNt initiateur (méthionine) resté au site P du ribosome. Ensuite le ribosome va se déplacer d'un codon sur l'ARNm. L'ARNt initiateur se retrouve alors au site E du ribosome où il pourra être libéré et le site A du ribosome, étant maintenant vide, peut pour accueillir un nouvel ARNt. L'opération va se répéter (Fig. 1.14) jusqu'à ce que le ribosome rencontre le codon stop. Ici, un facteur de terminaison va occuper le site A du ribosome, provoquant le détachement de la chaîne d'acides aminés (appelé aussi chaîne polypeptidique), puis des deux sous-unités du ribosome et du dernier ARNt (Fig. 1.17). Nous trouvons remarquable de pouvoir observer la traduction au sein des cellules et produire des images comme celle de la Fig. 1.15.

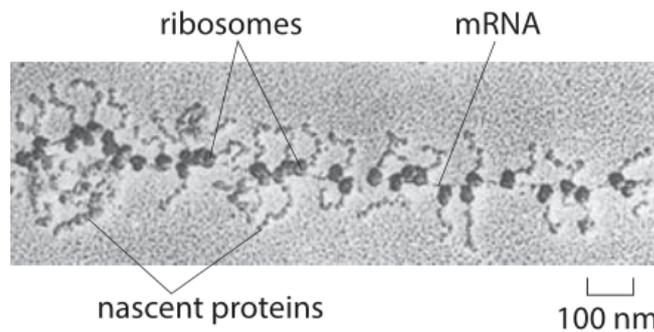
La chaîne polypeptidique peut commencer à se replier avant la fin de la traduction ou peut être aidée par des molécules appelées "chaperon" afin d'obtenir la conformation tridimensionnelle finale qui va minimiser l'énergie de la protéine.



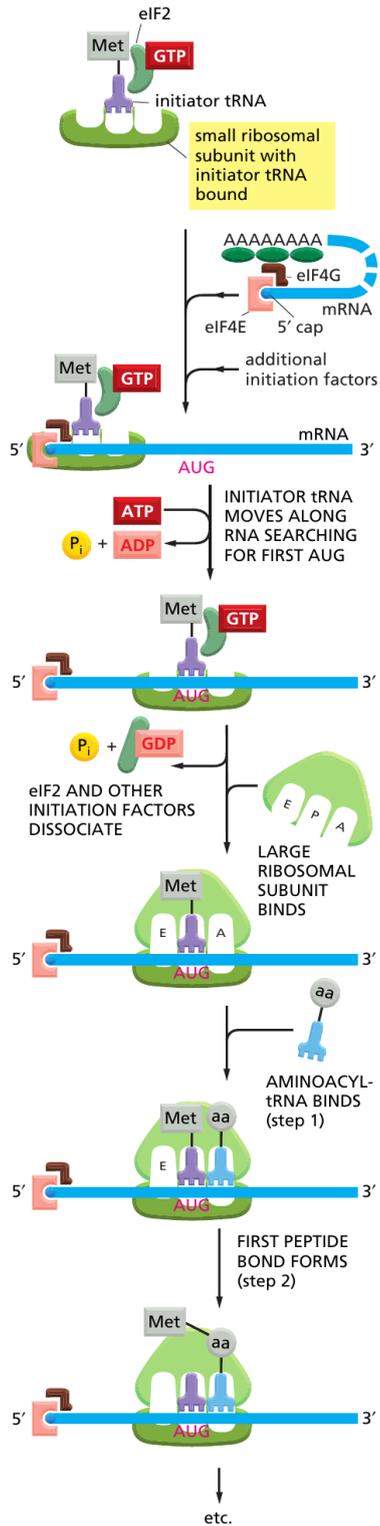
**FIGURE 1.13** – Schéma de la traduction [OH et al., 2011]. Les ribosomes vont se déplacer du codon “start” jusqu’au codon “stop” pour traduire l’information d’un gène contenu dans un ARNm en protéines. Une telle structure, formée de l’ARNm et des ribosomes, se nomme polysome.



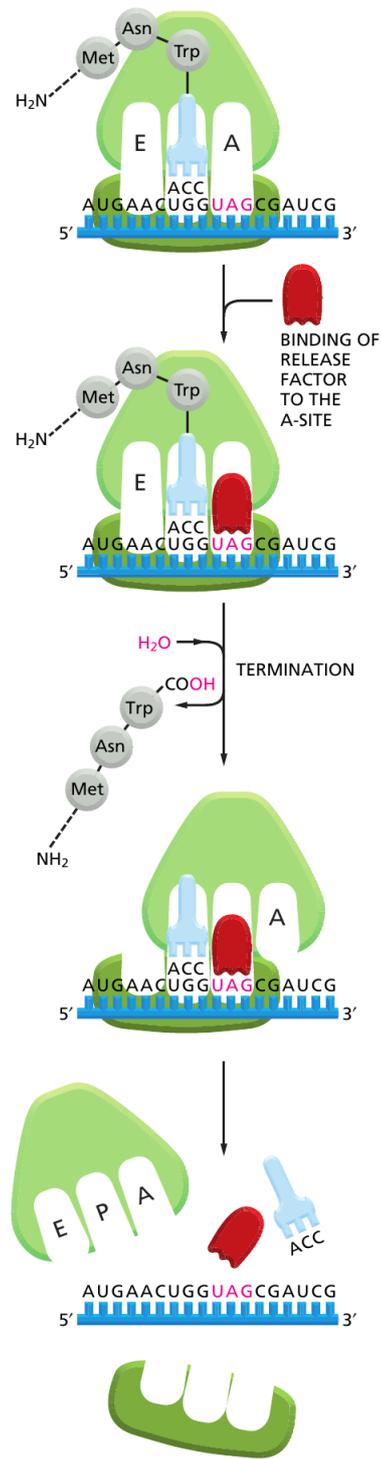
**FIGURE 1.14** – Mécanisme de la traduction CIAN-DRINI et al. [2013]. Le ribosome va permettre aux ARNt d’apporter successivement les acides aminés correspondant aux codons (triplets de nucléotides) et de former des liaisons entre les acides aminés. Les chaînes d’acides aminés ou chaînes polypeptidiques formeront les protéines en se repliant.



**FIGURE 1.15** – Image de microscopie électronique de la traduction [MILO et PHILLIPS, 2016]. On distingue clairement les ribosomes et les chaînes polypeptidiques (désignées ici par “nascent proteins”).



**FIGURE 1.16** – Schéma de l'initiation de la traduction [ALBERTS et al., 2015]. Plusieurs facteurs d'initiations permettent de déclencher la traduction. Le facteur eIF2 est présent sur la petite sous-unité du ribosome tandis que d'autres facteurs (eIF4) sont présents sur l'ARNm et lient ses extrémités. La petite sous-unité du ribosome se fixe au niveau de l'extrémité 5' de l'ARNm puis se déplace jusqu'au codon *start*. Puis, la grande sous-unité du ribosome s'associe à la petite sous-unité et la traduction peut commencer.

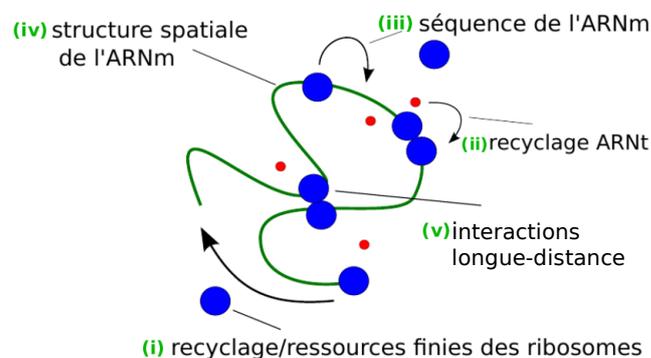


**FIGURE 1.17** – Schéma de la terminaison de la traduction [ALBERTS et al., 2015]. Lorsque le site A du ribosome arrive au niveau du codon *stop*, un facteur de terminaison vient permettre le détachement de la chaîne polypeptidique produite, puis le détachement du ribosome et du dernier ARNt.

## 2.6 Régulation de l'expression génétique

Avant, pendant, entre et après la transcription et la traduction, de nombreux processus de vérifications et de régulations ont lieu. Les vérifications assurent que le gène ait été correctement exprimé en protéines et peut conduire à l'avortement de la traduction avec destruction de la chaîne polypeptidique naissante. La régulation de l'expression génétique désigne tous les processus qui sont à l'œuvre dans le type et la quantité des protéines traduites à chaque instant. Elle est fondamentale pour la différenciation cellulaire ainsi que pour l'adéquation des cellules à leur environnement. Elle permet en particulier, aux cellules d'un même individu d'exprimer différemment le même génome. Une mauvaise régulation de production de protéines est à l'origine de nombreuses maladies. On compte les maladies neurodégénératives comme Alzheimer et la sclérose latérale amyotrophique (SLA appelée maladie de Charcot), l'obésité et les cancers [Har, 2019; POWERS et al., 2009].

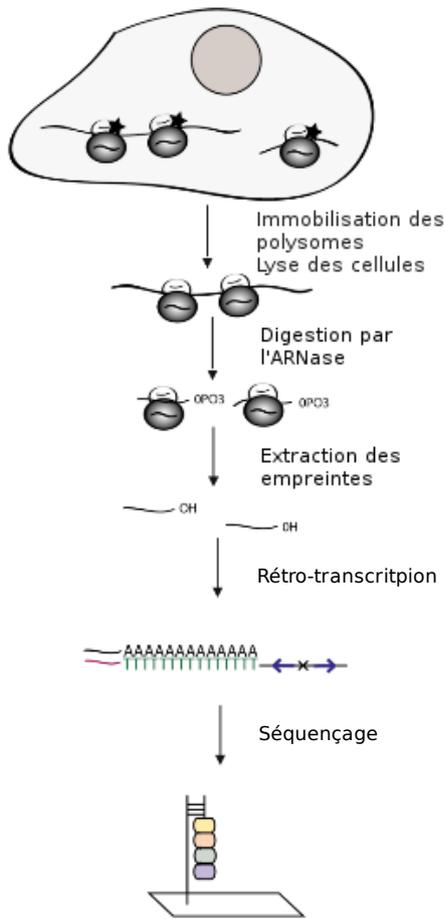
Certains mécanismes de régulation font actuellement l'objet d'études [ZUR et TULLER, 2016]. Nous citons dans ce paragraphe les mécanismes biophysiques de régulation de la traduction et les résumons dans la Fig. 1.18. (i) Le taux de traduction des ARNm peut être diminué à cause d'un nombre restreint de ribosomes, mais cette limitation peut être palliée par un mécanisme de recyclage. L'influence de ressources finies en ribosome sur la densité et le flux de ribosomes sur l'ARNm a été décrite par ADAMS et al. [2008]; COOK et ZIA [2009]; GREULICH et al. [2012], et avec l'effet du recyclage des ribosomes par CHOU [2003]; FERNANDES et al. [2017], optimisé par une courte distance bout à bout de l'ARNm (distance entre les deux extrémités de l'ARNm). Cependant, d'après DAULOUDET et al. [2021], la diffusion des ribosomes dans le cytoplasme étant beaucoup plus rapide que la traduction d'un ARNm, un mécanisme d'optimisation du recyclage ne serait pas pertinent. (ii et iii) Un autre phénomène de ressource fini et de recyclage a lieu au niveau des ARNt. Le ribosome précédant directement un autre ribosome pourra profiter des ARNt qui viennent d'être utilisés. Il a aussi été montré que les vitesses des ribosomes diffèrent selon les codons et que le principal acteur de cette variation est la concentration en ARNt portant les codons complémentaires [GARDIN et al., 2014]. (iii) Un autre responsable de cette variation serait les interactions électrostatiques entre le canal du site P du ribosome et les acides aminés de la chaîne polypeptidique s'extrayant par ce canal [DUC et SONG, 2018]. De plus, il semblerait que les codons plus riches en nucléotides A et U soient traduits plus vite que les codons plus riches en nucléotides G et C [GARDIN et al., 2014]. (iv) La structure spatiale des chaînes nucléotidiques peut aussi influencer la vitesse des protéines motrices en formant des obstacles, des boucles par exemples [ZUR et TULLER, 2016]. (v) Sur un polymère comme l'ARNm, une interaction longue distance désigne une interaction entre deux zones éloignées du polymère. De telles interactions ont lieu entre les deux extrémités de l'ARNm pour effectuer sa circularisation et réguler l'initiation de la traduction NICHOLSON et WHITE [2014]. Une autre source de régulation non étudiée à notre connaissance, pourrait être les interactions à longues distances entre les ribosomes.



**FIGURE 1.18** – Différents facteurs biophysiques de régulation de la traduction. Ces mécanismes peuvent impacter la vitesse de traduction par (i) les ressources finies en ribosomes et (ii) des ARNt ainsi que leur recyclage, (iii) les différents codons constituant l'ARNm, (iv) la structure de l'ARNm dans l'espace 3D et (v) les interactions à longue distance entre deux zones éloignées de la séquence nucléotidique.

L'étude de la traduction génétique et de sa régulation bénéficie d'un nouveau type d'expérimentation, appelé *ribosome profiling*, depuis les années 2010. La partie suivante est consacrée à la description de son protocole.

### 3 L'expérience de *ribosome profiling*



**FIGURE 1.19** – Schéma de principe de la technique Ribo-seq. Après isolation des polysomes, l'enzyme ARNase dégrade les parties des ARNm qui ne sont pas protégées par les ribosomes. Les fragments protégés par les ribosomes sont récupérés et séquencés afin de construire le profil de densité de ribosomes le long de l'ARNm (figure 1.20).

cléotides, le reste étant traité comme du “non-spécifique”. Il est à noter ici que la présence de clusters de ribosomes (ribosomes côte à côte) sera omise lors de la construction du profil Ribo-seq. Ensuite, on se sert de la reverse transcriptase pour synthétiser des brins d'ADN complémentaires à partir des fragments d'ARNm, afin de les séquencer. Par traitement informatique des données, les fragments d'ARNm recueillis vont être associés à la position qu'ils occupaient dans le gène.

Pour un gène donné, le but du traitement bioinformatique est de comptabiliser combien de fragments couvrent une position donnée dans la séquence codante de l'ARNm. Après nettoyage, chaque séquence des fragments est alignée sur la séquence de l'ARNm afin de déterminer la position d'origine du fragment - il s'agit d'une tâche classique de comparaison de séquences appelée *mapping* - à l'aide d'un logiciel efficace [STADLER et FIRE, 2011]. Seuls les fragments ayant un alignement de haute qualité et une position d'origine unique sont traités. On enregistre la position d'alignement de la première base du fragment. Sur la première position de chaque fragment, nous

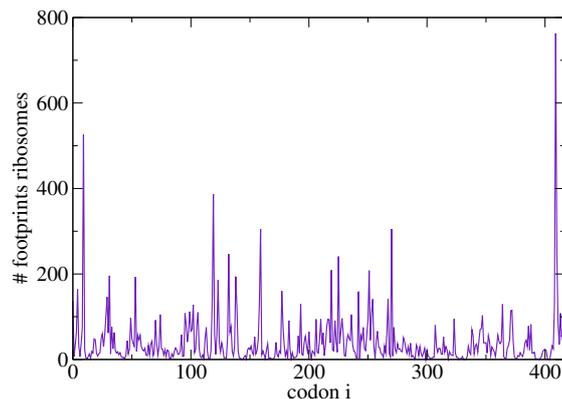
L'expérience de *ribosome profiling* ou Ribo-Seq, est une expérience de séquençage haut débit qui permet d'avoir une vision sur l'ensemble des ARN messagers traduits d'un type de cellules. Elle est donc complémentaire à des expériences de molécules uniques qui permettent d'observer une seule molécule. Plus précisément, il s'agit du comptage des ribosomes sur chaque codon de l'ARNm.

Elle a été introduite par Nicholas Ingolia il y a une dizaine d'années et peut maintenant être appliquée à la majorité des espèces, dont l'homme. Une variante de la méthode d'origine a été inventée par GAO et al. [2016] (Fig. 1.19). C'est cette méthode que pratique l'équipe de biologistes de l'IGF qui collabore avec nous sur ce projet au sein du consortium GEM. Cette partie du rapport présente une description de cette méthode.

Le séquençage est dit haut débit dans le sens où les polysomes (le complexe formé par l'ARNm et l'ensemble des ribosomes fixés dessus) de millions de cellules sont analysés. Les ribosomes sont d'abord immobilisés sur l'ARNm par l'injection d'un inhibiteur d'élongation, l'émetine, et d'un inhibiteur d'initiation, l'haringtonine. Cela va permettre d'avoir un instantané de la traduction pour toutes les cellules. Les cellules sont ensuite lysées afin d'isoler leurs polysomes. Des molécules qui clivent les liaisons phosphodiester des chaînes nucléotidiques entre deux nucléotides, les ARNases, vont être incorporées afin qu'elles dégradent les zones qui ne sont pas recouvertes par un ribosome. Ensuite, les ribosomes sont soustraits de la solution par purification au Ribo-Zero, un produit commercialisé par l'entreprise *illumina* qui permet l'élimination enzymatique de l'ARN ribosomique [ILLUMINA]. Les zones qui étaient en train d'être traduites puisque recouvertes par un ribosome, sont alors isolées par gel d'extraction. Puis, les fragments sont sélectionnés par taille grâce à un gel d'électrophorèse. Le plus souvent, il n'est gardé que les fragments faisant la taille d'un ribosome, c'est-à-dire environ 28 nu-

avons appliqué un décalage de 12 nucléotides pour assigner le fragment à la position du site P des ribosomes [PAULET et al., 2017]. Sur les analyses précédentes, il a été observé que 12 nucléotides est de loin la distance la plus fréquente entre le début du fragment et le codon recouvert par le P-site [PHILIPPE et al., 2013], qui détermine la position du codon en train d’être traduit. Enfin, chaque position de P-site recouverte est enregistrée dans un tableau.

Cette expérience récente, qui permet la détermination des positions des ribosomes sur l’ARNm, est un outil considérable pour étudier les propriétés cinétiques de la traduction, particulièrement en comparant les données de ces expériences avec les résultats des modélisations physico-mathématiques.



**FIGURE 1.20** – Profil Ribo-Seq obtenu après traitements bioinformatique des données expérimentales obtenues chez l’homme. Nombre de ribosomes détectés sur chaque codon d’un gène par séquençage haut débit sur des millions de cellules.

## 4 Conclusion

Dans cette section, nous avons présenté les principes généraux de l’expression des gènes. L’ensemble de l’information génétique est stockée par l’ADN qui constitue donc une véritable bibliothèque des gènes, protégée dans le noyau de la cellule chez la cellule eucaryote. Seulement certains gènes sont exprimés selon le type de cellule et ses besoins. Pour être exprimé, un gène, qui est une séquence nucléotidique de l’ADN, est transcrit en ARN messager, une séquence nucléotidique similaire. L’information génétique de ce gène pourra alors être déplacée au lieu où devront être produites les protéines.

Ces principes généraux fondamentaux sont bien connus, ainsi que de nombreux détails sur leur fonctionnement, mais la question de la régulation de l’expression génétique est encore largement non élucidée alors qu’il s’agit d’un mécanisme vital pour les cellules biologiques. En particulier, la régulation traductionnelle reste actuellement mal comprise. Mais l’émergence de la technique de séquençage haut débit Ribo-Seq permet une nouvelle investigation de la cinétique de traduction.

Nous avons vu que la traduction est effectuée par un processus de transport du ribosome le long du brin d’ARNm. Il s’agit d’un processus de transport dirigé à une dimension très intéressante d’un point de vue physique. Durant cette thèse, on s’est intéressé aux modèles suggérés par la traduction pour en étudier sa physique et nous avons cherché à faire bénéficier la recherche en biologie des résultats obtenus. Le chapitre suivant est consacré aux modèles stochastiques de transports dirigés à une dimension inspirés par la traduction génétique.

## 5 Références

2019, «Platform Communications : Abstract Book - 30th International Symposium on ALS/MND (Complete printable file)», *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*,

- vol. 20, n° sup1, doi :10.1080/21678421.2019.1646546, p. 1–99, ISSN 2167-8421. URL <https://doi.org/10.1080/21678421.2019.1646546>, publisher : Taylor & Francis. 14
- ADAMS, D. A., B. SCHMITTMANN et R. K. P. ZIA. 2008, «Far-from-equilibrium transport with constrained resources», *Journal of Statistical Mechanics : Theory and Experiment*, vol. 2008, n° 06, doi :10.1088/1742-5468/2008/06/P06009, p. P06009, ISSN 1742-5468. URL <http://arxiv.org/abs/0804.2717>, arXiv :0804.2717 [cond-mat]. 14
- ALBERTS, B. M., A. JOHNSON, J. LEWIS, D. O. MORGAN, M. C. RAFF, K. ROBERTS et P. WALTER. 2015, *Molecular biology of the cell*, 6<sup>e</sup> éd., Garland Science, Taylor & Francis Group, New York [etc., ISBN 978-0-8153-4432-2. 7, 8, 13
- CHOU, T. 2003, «Ribosome Recycling, Diffusion, and mRNA Loop Formation in Translational Regulation», *Biophysical Journal*, vol. 85, n° 2, p. 755–773, ISSN 0006-3495. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1303200/>. 14
- CIANDRINI, L., I. STANSFIELD et M. C. ROMANO. 2013, «Ribosome Traffic on mRNAs Maps to Gene Ontology : Genome-wide Quantification of Translation Initiation Rates and Polysome Size Regulation», *PLOS Computational Biology*, vol. 9, n° 1, doi :10.1371/journal.pcbi.1002866, p. e1002866, ISSN 1553-7358. URL <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002866>. 12
- COOK, L. J. et R. K. P. ZIA. 2009, «Feedback and Fluctuations in a Totally Asymmetric Simple Exclusion Process with Finite Resources», *Journal of Statistical Mechanics : Theory and Experiment*, vol. 2009, n° 02, doi :10.1088/1742-5468/2009/02/P02012, p. P02012, ISSN 1742-5468. URL <http://arxiv.org/abs/0811.1543>, arXiv :0811.1543 [cond-mat]. 14
- CROSS, C. E. 1995, «Bloom and Fawcett : A Textbook of Histology», *JAMA*, vol. 274, n° 4, doi :10.1001/jama.1995.03530040080050, p. 352, ISSN 0098-7484. URL <https://doi.org/10.1001/jama.1995.03530040080050>. 7
- DAULOUDET, O., I. NERI, J.-C. WALTER, J. DORIGNAC, F. GENIET et A. PARMEGGIANI. 2021, «Modelling the effect of ribosome mobility on the rate of protein synthesis», *The European Physical Journal E*, vol. 44, n° 2, doi :10.1140/epje/s10189-021-00019-8, p. 19, ISSN 1292-895X. URL <https://doi.org/10.1140/epje/s10189-021-00019-8>. 14
- DUC, K. D. et Y. S. SONG. 2018, «The impact of ribosomal interference, codon usage, and exit tunnel interactions on translation elongation rate variation», *PLOS Genetics*, vol. 14, n° 1, doi :10.1371/journal.pgen.1007166, p. e1007166, ISSN 1553-7404. URL <https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1007166>. 14
- FERNANDES, L. D., A. P. S. D. MOURA et L. CIANDRINI. 2017, «Gene length as a regulator for ribosome recruitment and protein synthesis : theoretical insights», *Scientific Reports*, vol. 7, n° 1, doi :10.1038/s41598-017-17618-1, p. 17409, ISSN 2045-2322. URL <http://www.nature.com/articles/s41598-017-17618-1>, bandiera\_abtest : a Cc\_license\_type : cc\_by Cg\_type : Nature Research Journals Number : 1 Primary\_atype : Research Publisher : Nature Publishing Group Subject\_term : Biological physics;Systems biology;Translation Subject\_term\_id : biological-physics;systems-biology;translation. 14
- GAO, X., J. WAN et S.-B. QIAN. 2016, «Genome-Wide Profiling of Alternative Translation Initiation Sites», *Methods in Molecular Biology (Clifton, N.J.)*, vol. 1358, doi :10.1007/978-1-4939-3067-8\_19, p. 303–316, ISSN 1940-6029. 16
- GARDIN, J., R. YEASMIN, A. YUROVSKY, Y. CAI, S. SKIENA et B. FUTCHER. 2014, «Measurement of average decoding rates of the 61 sense codons in vivo», *eLife*, vol. 3, doi :10.7554/eLife.03735, p. e03735, ISSN 2050-084X. URL <https://doi.org/10.7554/eLife.03735>, publisher : eLife Sciences Publications, Ltd. 14

- GREULICH, P., L. CIANDRINI, R. J. ALLEN et M. C. ROMANO. 2012, «Mixed population of competing TASEPs with a shared reservoir of particles», *Physical Review E*, vol. 85, n° 1, doi : 10.1103/PhysRevE.85.011142, p. 011 142, ISSN 1539-3755, 1550-2376. URL <http://arxiv.org/abs/1111.1775>, arXiv :1111.1775 [cond-mat]. 14
- ILLUMINA. «Illumina Ribo-Zero Plus rRNA Depletion Kit | Stand-alone rRNA depletion», URL <https://emea.illumina.com/products/by-type/accessory-products/ribo-zero-plus-rrna-depletion.html>. 16
- KUSHNER, S. R. 2018, «Messenger RNA in Prokaryotes», dans *eLS*, John Wiley & Sons, Ltd, ISBN 978-0-470-01590-2, p. 1–11, doi :10.1002/9780470015902.a0000874.pub4. URL <http://onlinelibrary.wiley.com/doi/abs/10.1002/9780470015902.a0000874.pub4>, \_eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470015902.a0000874.pub4>. 7, 8
- LEBRETON, A. 2006, *Dynamique des facteurs pré-ribosomiques au cours de la biogenèse de la grande sous-unité ribosomique chez S. cerevisiae*, thèse de doctorat. 7
- MANDELKERN, M., J. G. ELIAS, D. EDEN et D. M. CROTHERS. 1981, «The dimensions of DNA in solution», *Journal of Molecular Biology*, vol. 152, n° 1, doi :10.1016/0022-2836(81)90099-1, p. 153–161, ISSN 0022-2836. URL <https://www.sciencedirect.com/science/article/pii/0022283681900991>. 5
- MILO, R. et R. B. PHILLIPS. 2016, *Cell biology by the numbers*, Garland Science, Taylor & Francis Group, C 2016, New York, NY Abingdon, ISBN 978-0-8153-4537-4. 12
- NICHOLSON, B. L. et K. A. WHITE. 2014, «Functional long-range RNA–RNA interactions in positive-strand RNA viruses», *Nature Reviews Microbiology*, vol. 12, n° 7, doi :10.1038/nrmicro3288, p. 493–504, ISSN 1740-1534. URL <http://www.nature.com/articles/nrmicro3288>, number : 7 Publisher : Nature Publishing Group. 14
- OH, E., A. H. BECKER, A. SANDIKCI, D. HUBER, R. CHABA, F. GLOGE, R. J. NICHOLS, A. TYPAS, C. A. GROSS, G. KRAMER, J. S. WEISSMAN et B. BUKAU. 2011, «Selective Ribosome Profiling Reveals the Cotranslational Chaperone Action of Trigger Factor In Vivo», *Cell*, vol. 147, n° 6, doi : 10.1016/j.cell.2011.10.044, p. 1295–1308, ISSN 0092-8674. URL <https://www.sciencedirect.com/science/article/pii/S0092867411013481>. 12
- ORR, M. W., Y. MAO, G. STORZ et S.-B. QIAN. 2020, «Alternative ORFs and small ORFs : shedding light on the dark proteome», *Nucleic Acids Research*, vol. 48, n° 3, doi :10.1093/nar/gkz734, p. 1029–1042, ISSN 0305-1048. URL <https://doi.org/10.1093/nar/gkz734>. 7
- PAULET, D., A. DAVID et E. RIVALS. 2017, «Ribo-seq enlightens codon usage bias», *DNA research : an international journal for rapid publication of reports on genes and genomes*, vol. 24, n° 3, doi :10.1093/dnares/dsw062, p. 303–210, ISSN 1756-1663. 17
- PAULET, D., A. DAVID et E. RIVALS. 2021, «RNA-Ribo Explorer : interactive mining and visualisation of Ribosome profiling data», preprint, Bioinformatics, doi :10.1101/2021.03.23.436679. URL <http://biorxiv.org/lookup/doi/10.1101/2021.03.23.436679>. 10
- PHILIPPE, N., M. SALSON, T. COMMES et E. RIVALS. 2013, «CRAC : an integrated approach to the analysis of RNA-seq reads», *Genome Biology*, vol. 14, n° 3, doi :10.1186/gb-2013-14-3-r30, p. R30, ISSN 1474-760X. URL <https://doi.org/10.1186/gb-2013-14-3-r30>. 17
- POOLE, A., D. PENNY et B.-M. SJÖBERG. 2001, «Confounded cytosine! Tinkering and the evolution of DNA», *Nature Reviews Molecular Cell Biology*, vol. 2, n° 2, doi :10.1038/35052091, p. 147–151, ISSN 1471-0080. URL <http://www.nature.com/articles/35052091>, number : 2 Publisher : Nature Publishing Group. 7

- POWERS, E. T., R. I. MORIMOTO, A. DILLIN, J. W. KELLY et W. E. BALCH. 2009, «Biological and Chemical Approaches to Diseases of Proteostasis Deficiency», *Annual Review of Biochemistry*, vol. 78, n° 1, doi :10.1146/annurev.biochem.052308.114844, p. 959–991. URL <https://doi.org/10.1146/annurev.biochem.052308.114844>. 14
- STADLER, M. et A. FIRE. 2011, «Wobble base-pairing slows in vivo translation elongation in metazoans», *RNA*, vol. 17, n° 12, doi :10.1261/rna.02890211, p. 2063–2073, ISSN 1355-8382. URL <http://rnajournal.cshlp.org/cgi/doi/10.1261/rna.02890211>. 16
- YUSUPOV, M. M., G. Z. YUSUPOVA, A. BAUCOM, K. LIEBERMAN, T. N. EARNEST, J. H. CATE et H. F. NOLLER. 2001, «Crystal structure of the ribosome at 5.5 Å resolution», *Science (New York, N.Y.)*, vol. 292, n° 5518, doi :10.1126/science.1060089, p. 883–896, ISSN 0036-8075. 10
- ZUR, H. et T. TULLER. 2016, «Predictive biophysical modeling and understanding of the dynamics of mRNA translation and its evolution», *Nucleic Acids Research*, vol. 44, n° 19, doi :10.1093/nar/gkw764, p. 9031–9049, ISSN 0305-1048. URL <https://doi.org/10.1093/nar/gkw764>. 14

## Chapitre 2

# Modélisation du transport des ribosomes par gaz sur réseaux : état de l'art et résultats préliminaires

### Sommaire

---

<b>1</b>	<b>Introduction</b> . . . . .	<b>22</b>
<b>2</b>	<b>Historique et état de l'art des modèles physiques de la traduction génétique</b> . .	<b>22</b>
<b>3</b>	<b>TASEP</b> . . . . .	<b>23</b>
3.1	Le modèle . . . . .	24
3.2	Résolution exacte . . . . .	24
3.2.1	Équation maîtresse & Matrice de Markov . . . . .	24
3.2.2	Caractérisation du TASEP (Ansatz matriciel) . . . . .	27
3.3	Approximation de champ moyen . . . . .	29
<b>4</b>	<b>Évolution du modèle : <math>\ell</math>-TASEP inhomogène</b> . . . . .	<b>31</b>
4.1	Le modèle TASEP avec des taux de sauts site-spécifiques et des particules étendues . . . . .	31
4.2	Densité et courant à l'état stationnaire . . . . .	32
4.2.1	Équations exactes sur la variation de l'occupation moyenne . . . . .	32
4.2.2	Équations de champ moyen . . . . .	33
4.3	Diagramme de phases . . . . .	34
<b>5</b>	<b>Résultats préliminaires à propos de l'effet de la particule étendue sur la densité de particules</b> . . . . .	<b>35</b>
5.1	$\ell$ -TASEP homogène : effet de bord sur la densité de particules en phase basse densité . . . . .	35
5.2	$\ell$ -TAESEP . . . . .	38
5.3	Comparaison des modèles $\ell$ -TASEP et $\ell$ -TAESEP homogènes pour étudier l'influence de la particule étendue . . . . .	39
<b>6</b>	<b>Simulations Monte Carlo</b> . . . . .	<b>41</b>
6.1	Algorithme de Gillespie . . . . .	41
6.2	Simulation du TASEP à l'état stationnaire . . . . .	42
<b>7</b>	<b>Conclusion et perspectives</b> . . . . .	<b>42</b>
<b>8</b>	<b>Références</b> . . . . .	<b>43</b>

---

## 1 Introduction

Ce chapitre est consacré à la modélisation des processus de transports stochastiques et unidirectionnels à une dimension par gaz sur réseau. Ce processus physique convient particulièrement à la modélisation de la traduction génétique pendant laquelle le ribosome se déplace codon par codon d'une extrémité à l'autre de la séquence de l'ARNm à l'aide de plusieurs processus chimiques. Ce système d'apparence simple est particulièrement intéressant parce qu'il est hors équilibre et présente de nombreuses propriétés complexes que nous allons décrire le long de ce chapitre. Après un bref historique des modèles physiques de la traduction, nous explorerons différentes variantes de ce processus en partant du modèle paradigmatique du transport dirigé à une dimension pour aller vers des modèles plus spécifiques à la traduction.

Le modèle paradigmatique du transport dirigé à une dimension est le TASEP pour *Totally Asymmetric Simple Exclusion Process* dans lequel une particule occupant un site se déplace avec un taux de saut constant. Nous décrivons l'approche Markovienne pour une résolution exacte des équations qui permet de complètement caractériser le système, c'est-à-dire prédire le comportement précis de la densité et du courant en fonction des paramètres cinétiques et les classer dans un diagramme de phase. Puis nous démontrerons que la densité et le courant de particules en basse et haute densité peuvent être retrouvés facilement par une approche en champ moyen. Ensuite, nous étudierons les deux évolutions fondamentales rapprochant le modèle TASEP de la complexité de la traduction. Il s'agit de la particule étendue, une particule occupant plusieurs sites du réseau à la fois tout en se déplaçant toujours d'un seul site à la fois, et de l'inhomogénéité des taux sauts qui n'ont plus même la valeur en chaque site. On parlera alors de  $\ell$ -TASEP inhomogène,  $\ell$  désignant la longueur de la particule. Nous démontrerons les équations de champ moyen pour ce modèle, puis nous étudierons spécifiquement l'effet de la particule étendue sans l'effet de l'inhomogénéité des taux de saut. En particulier, nous remarquerons que contrairement au cas du TASEP classique, la densité du  $\ell$ -TASEP homogène ne peut plus être considérée constante le long du réseau en phase basse densité pour tous les paramètres en raison d'un effet de bord trop important. De l'étude de cet effet de bord, nous en déduisons un critère très simple indiquant si la variation de la densité induite peut être négligée en fonction des paramètres cinétiques du modèle.

Enfin, nous introduisons une nouvelle variante pour étudier l'exclusion que nous avons appelée  $\ell$ -TAESEP pour *Totally Asymmetric Entry Boundary Exclusion Process* avec particule étendue. Nous en donnerons une résolution exacte, mais non Markovienne, afin d'en déduire les expressions du courant et de la densité de particules. Nous montrerons que l'exclusion à l'entrée est de nature différente de l'exclusion dans le reste du réseau et qu'elle a une influence plus importante sur la densité pour notre règle d'entrée sur le réseau qui correspond à la traduction génétique. Pour cela, nous utiliserons des simulations Monte Carlo avec l'algorithme de Gillespie qui sera décrit dans la dernière section.

## 2 Historique et état de l'art des modèles physiques de la traduction génétique

Le premier article concernant la modélisation de la traduction date de 1963 [SIMHA et al., 1963]. Pour cette première tentative, il manquait un élément essentiel du processus de traduction connu aujourd'hui : l'ARN messager. En effet, l'ADN en tant que matériel génétique a été confirmé par Alfred Hershey et Martha Chase en 1952 [DAHM, 2008]. Le ribosome a été découvert dans les années 50 par George Emil Palade [PALADE, 2007]. Mais on ne savait pas de quelle façon l'information génétique était traduite en protéine. Une première supposition était que l'information génétique est contenue dans les ribosomes. Il n'y avait donc pas de processus de transport. L'idée d'un "messenger" de l'information génétique a été émise et étudiée par JACOB et MONOD [1961], mais pas encore confirmée. Les chimistes de la première modélisation ne l'avaient pas pris en compte. C'est en 1965 qu'est publiée la première modélisation de la traduction génétique comme proces-

sus de transport du ribosome sur l'ARN messager [GERST et LEVINE, 1965]. Le ribosome diffusait déjà dans un seul sens sur l'ARNm. Le système était défini par le nombre de ribosomes sur l'ARNm et le nombre constant de codons séparant chaque ribosome. On s'intéressait à la dynamique de formation de toutes les chaînes polypeptidiques en cours à la fois. C'est en 1968 que naît le modèle qui va devenir le modèle paradigmatique de la traduction [MACDONALD et al., 1968]. Ce modèle n'a pas encore de nom. Il s'agit d'un réseau sur lequel se déplace de manière stochastique un ribosome qui couvre plusieurs sites tout en se déplaçant d'un site à la fois. Le ribosome peut avancer ou reculer sur le réseau, mais il suffit de mettre le taux de recul à 0 pour retrouver un modèle où le ribosome ne peut qu'avancer. Ce modèle de transport sera massivement étudié pour divers domaines et nommé ASEP pour Asymmetric Simple Exclusion Process. "Asymmetric" indique que le taux d'avancée est différent du taux de recul, "Simple" indique que le ribosome, ou plus généralement la particule, ne peut se déplacer que d'un site à la fois, "Exclusion" indique que 2 particules ne peuvent se trouver sur le même site en même temps. Dans le cas de la traduction, la variante "TASEP" sera retenue, le -T pour "Totally" indiquant que le ribosome ne peut se déplacer que dans une seule direction. Ce sont quatre physiciens statisticiens qui donneront la résolution exacte du modèle à l'état stationnaire en 1992 [DERRIDA et al., 1992; SCHÜTZ et DOMANY, 1993]. Avec ce modèle, on s'intéresse à la densité et au flux de particules sur chaque site du réseau lors de l'état stationnaire. La densité au site  $i$  est en réalité la densité de probabilité qu'une particule se trouve au site  $i$  à n'importe quel instant puisqu'on est à l'état stationnaire. C'est cette grandeur que nous allons majoritairement étudier dans cette thèse. En 1969, donc peu de temps après la naissance du TASEP, est imaginé un modèle qui donne un temps de vie à l'ARNm [SINGH, 1969]. On s'intéresse alors à la manière dont la traduction se termine et au lien avec le processus précédent qui donne naissance à l'ARNm, la transcription. Pour cette approche, le processus de transport n'est pas un TASEP, mais un processus déterministe et sans exclusion, ce qui permet de cibler spécifiquement la question de l'influence de la durée de vie de l'ARNm sur la production des protéines et la distribution des populations d'ARNm. Cette approche a nettement moins été utilisée que le TASEP. Pourtant, avec la naissance de l'expérience Ribo-Seq qui donne accès à la statistique des positions des ribosomes sur l'ARNm, il devient important de s'intéresser à l'effet du temps de vie des ARNm et à la manière dont le complexe se dégrade. C'est ce qu'a fait une équipe de biophysiciens à partir de 2010 avec un article qui s'intéresse à l'effet de la durée de vie finie des ARNm sur la densité de ribosomes sur l'ARNm [NAGAR et al., 2011]. Quand on considère un ensemble d'ARNm à temps de vie finie et naissant à différents instants, ce qui correspond à ce qui est mesuré dans l'expérience Ribo-Seq, les systèmes ARNm + ribosomes, que nous appellerons "polysomes" ne sont pas tous à l'état stationnaire, ce qui a une incidence non-négligeable sur le profil de densité de ribosomes sur chaque codon de l'ARNm.

Dans ce chapitre et le suivant, nous allons nous intéresser à ces deux modèles importants (transport stochastique sur un ARNm à temps de vie infini et transport déterministe sur un ARNm à temps de vie fini) et tenter de les combiner afin de comprendre ce système physique particulier et de donner une interprétation riche et physiquement pertinente des expériences Ribo-Seq.

### 3 TASEP

Le modèle TASEP est le modèle paradigmatique des phénomènes de transports dirigés à une dimension. La revue de LIU et al. [2012] indique que le modèle TASEP et ses extensions ont été appliqués en " *biology, physics, and chemistry* [SCHÜTZ, 2001], *such as gel electrophoresis* [WIDOM et al., 1991], *protein synthesis* [SCHÜTZ, 1997; SHAW et al., 2003], *mRNA translation* [CHOU, 2003], *motion of molecular motors along cytoskeletal filaments* [KLUMPP et LIPOWSKY, 2003; PARMEGGIANI et al., 2004], *the depolymerization of microtubules by special enzymes* [KLEIN et al., 2005], *and the vehicular traffic* [CHOWDHURY et al., 2000; HELBING, 2001]." Le modèle TASEP a aussi été utilisé par SUGDEN et al. [2007] pour modéliser le transport des vésicules dans les hyphes des champignons et par JOHN et al. [2009] pour modéliser les mouvements collectifs des fourmis. La

liste des publications citées n'est pas exhaustive. <sup>1</sup>.

### 3.1 Le modèle

TASEP est l'acronyme de Totally Asymmetric Simple Exclusion Process. "Totally Asymmetric" indique que le ribosome ne peut se déplacer que dans une seule direction, "Simple" indique que la particule ne peut se déplacer que d'un site à la fois, "Exclusion" indique que plusieurs particules ne peuvent se trouver sur le même site au même instant. Nous allons nous intéresser au cas des bords ouverts, c'est-à-dire que les particules sont injectées d'un côté du réseau et éjectées de l'autre côté du réseau. C'est le cas qui correspond au processus biologique de la traduction génétique. Sur ce réseau, une particule entre avec un taux  $\alpha$  sur le site 1 s'il n'est pas déjà occupé, se déplace par saut avec un taux  $p$  sur le site suivant s'il n'est pas occupé et sort du réseau avec un taux  $\beta$ .

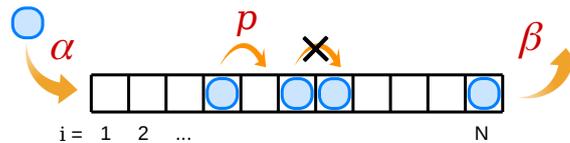


FIGURE 2.1 – Schéma du modèle TASEP avec bords ouverts. Les particules ne peuvent entrer que sur le premier site, avec un taux  $\alpha$ . La particule sur le site  $i$  peut se déplacer sur le site suivant seulement s'il n'est pas déjà occupé, avec au taux  $p$ . Les particules ne peuvent se déplacer que d'un site à la fois. Elles ne peuvent quitter le réseau que si elles se trouvent sur le dernier site, le site  $N$ , avec un taux  $p_N = \beta$ .

Ce modèle a été particulièrement étudié afin de connaître la densité et le flux de particules à l'état stationnaire en fonction des paramètres cinétiques  $\alpha$ ,  $\beta$  et  $p$ . Nous allons d'abord expliquer comment établir la résolution exacte et présenter la caractérisation du TASEP qui en découle. Puis, nous établirons la résolution en champ moyen avec laquelle on retrouve les caractéristiques essentielles du TASEP à l'état stationnaire, de manière considérablement plus simple et rapide.

### 3.2 Résolution exacte

Pour caractériser précisément le système, on effectue une résolution exacte qui permet l'établissement du diagramme de phase avec la densité et le flux associé à chaque phase à l'état stationnaire. Cette caractérisation est possible grâce à l'utilisation d'un Ansatz Matriciel.

#### 3.2.1 Équation maîtresse & Matrice de Markov

Dans le TASEP, les probabilités de changement entre deux états sont déterminées par les différents taux d'événements par unité de temps (entrée, sortie, saut d'un site au site suivant). Ces taux sont des constantes par rapport au temps et ne dépendent pas des événements précédents. La distribution de probabilité des états futurs ne dépend ainsi que de l'état présent, c'est le caractère markovien. Le TASEP est donc ce qu'on appelle un processus de Markov. Il est gouverné par une équation maîtresse sur les probabilités des états, qui est l'équation de l'évolution des probabilités des états en fonction du temps. Elle s'écrit

$$\frac{d|P(t)\rangle}{dt} = M|P(t)\rangle \quad (2.1)$$

où  $|P(t)\rangle$  est le vecteur des probabilités au temps  $t$  de tous les états  $j$  de configuration  $\mathcal{E}_j$  possibles et  $M$  la matrice de transition du système en un pas de temps infinitésimal  $dt$ . Puisqu'il y a 2 états

1. biologie, physique et chimie [SCHÜTZ, 2001], comme les gels d'électrophorèses [WIDOM et al., 1991], la synthèse des protéines [SCHÜTZ, 1997; SHAW et al., 2003], la traduction de l'ARNm [CHOU, 2003], le mouvement des moteurs moléculaires sur les filaments du cytosquelette [KLUMPP et LIPOWSKY, 2003; PARMEGGIANI et al., 2004], la dépolymérisation des microtubules par des enzymes spéciales [KLEIN et al., 2005] et le trafic routier [CHOWDHURY et al., 2000; HELBING, 2001][traduction libre]

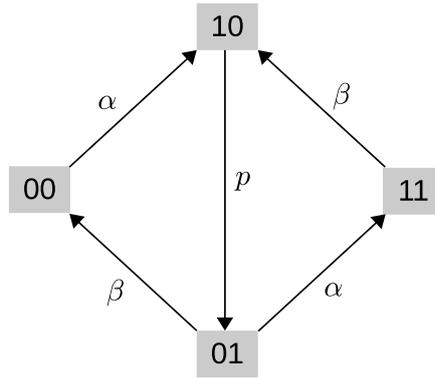


FIGURE 2.2 – Schéma du processus markovien pour un TASEP de  $N = 2$  sites.

possibles pour chaque site, pour un réseau de  $N$  sites il y a  $2^N$  états possible au total.  $|P(t)\rangle$  à donc  $2^N$  composantes et s'écrit :

$$|P(t)\rangle = \begin{pmatrix} P_{\mathcal{C}_1}(t) \\ P_{\mathcal{C}_2}(t) \\ \vdots \\ P_{\mathcal{C}_{2^N}}(t) \end{pmatrix}. \quad (2.2)$$

La matrice  $M$  contient les probabilités de changement de chaque état  $k$  vers chaque état  $j$ ,  $P_{\mathcal{C}_k \rightarrow \mathcal{C}_j}$ , tel que

$$P_{\mathcal{C}_j}(t + dt) = \sum_{k=1}^{2^N} P_{\mathcal{C}_k \rightarrow \mathcal{C}_j} P_{\mathcal{C}_k}(t). \quad (2.3)$$

$M$  est donc de taille  $2^N \times 2^N$ . Pour la calculer, on transforme l'Eq. (2.3) pour chaque  $j$  de manière à retrouver l'Eq. (2.1), soit

$$P_{\mathcal{C}_j}(t + dt) = P_{\mathcal{C}_j \rightarrow \mathcal{C}_j} P_{\mathcal{C}_j}(t) + \sum_{k=1, k \neq j}^{2^N} P_{\mathcal{C}_k \rightarrow \mathcal{C}_j} P_{\mathcal{C}_k}(t) \quad (2.4)$$

$P_{\mathcal{C}_j \rightarrow \mathcal{C}_j}$  étant la probabilité que l'état  $j$  persiste et est donc aussi égale à  $1 - \sum_{l=1, l \neq j}^{2^N} P_{\mathcal{C}_j \rightarrow \mathcal{C}_l} P_{\mathcal{C}_j}(t)$ , ce qui permet d'écrire

$$\frac{dP_{\mathcal{C}_j}(t)}{dt} = \lim_{\Delta t \rightarrow 0} \left\{ - \left( \sum_{l=1, l \neq j}^{2^N} \frac{P_{\mathcal{C}_j \rightarrow \mathcal{C}_l}}{\Delta t} \right) P_{\mathcal{C}_j}(t) + \sum_{k=1, k \neq j}^{2^N} \frac{P_{\mathcal{C}_k \rightarrow \mathcal{C}_j}}{\Delta t} P_{\mathcal{C}_k}(t) \right\}. \quad (2.5)$$

On en déduit que  $-\lim_{\Delta t \rightarrow 0} \sum_{l=1, l \neq j}^{2^N} \frac{P_{\mathcal{C}_j \rightarrow \mathcal{C}_l}}{\Delta t}$  est la composante  $(j, j)$  de  $M$  et  $\lim_{\Delta t \rightarrow 0} \frac{P_{\mathcal{C}_k \rightarrow \mathcal{C}_j}}{\Delta t}$  les composantes  $(j, k)$  de  $M$  pour  $k$  allant de 1 à  $2^N$  et différent de  $j$ . En écrivant les équations pour chaque état  $j$  possible du réseau, on obtient de cette manière toutes les composantes de  $M$ . Une fois  $M$  déterminée, on peut en déduire les probabilités à l'état stationnaire. Cela permet de calculer notamment la densité de particule en chaque site  $i$  à l'état stationnaire qui s'écrit

$$\rho_i = \sum_{\{\mathcal{C}_j | n_i = 1\}} P_{\mathcal{C}_j} \quad (2.6)$$

où  $\{\mathcal{C}_j | n_i = 1\}$  est l'ensemble des configurations des états  $j$  où le site  $i$  est occupé.

Par exemple, pour un réseau de  $N = 2$  sites, on a 4 états dont les configurations sont 00, 01, 10, 11. En s'aidant du graphe du processus 2.2, on obtient l'évolution des probabilités des états correspondant à l'Eq. (2.3) :

$$\begin{cases} P_{00}(t + \Delta t) = (1 - \alpha\Delta t)P_{00}(t) + \beta\Delta tP_{01}(t), \\ P_{01}(t + \Delta t) = (1 - (\beta + \alpha)\Delta t)P_{01}(t) + p\Delta tP_{10}(t), \\ P_{10}(t + \Delta t) = (1 - p\Delta t)P_{10}(t) + \alpha\Delta tP_{00}(t) + \beta\Delta tP_{11}(t), \\ P_{11}(t + \Delta t) = (1 - \beta\Delta t)P_{11}(t) + \alpha\Delta tP_{01}(t), \end{cases} \quad (2.7)$$

et en faisant apparaître la différentielle de chaque probabilité, on en déduit la matrice de Markov

$$M = \begin{pmatrix} -\alpha & \beta & 0 & 0 \\ 0 & -(\alpha + \beta) & p & 0 \\ \alpha & 0 & -p & \beta \\ 0 & \alpha & 0 & -\beta \end{pmatrix}. \quad (2.8)$$

On remarque que chaque somme des composantes sur une colonne est nulle, c'est en effet un critère nécessaire facile à vérifier. En vertu de la conservation de la probabilité totale, la somme des probabilités des états doit valoir 1, la dérivée de cette somme est nulle et d'après l'Eq. (2.1) on obtient cette propriété (seuls les coefficients d'une même probabilité  $P_{n_1 n_2}$  peuvent s'annuler). On s'intéresse en général à l'état stationnaire, ce qui permet de résoudre aisément l'Eq.(2.1). Dans notre exemple on obtient

$$|P\rangle = \frac{1}{\alpha^2 p + \beta(\alpha + \beta)(p + \alpha)} \begin{pmatrix} p\beta^2 \\ p\alpha\beta \\ \alpha\beta(\alpha + \beta) \\ p\alpha^2 \end{pmatrix} \quad (2.9)$$

et les densités

$$\begin{aligned} \rho_1 &= \frac{\alpha(\alpha p + (\alpha + \beta)\beta)}{\alpha^2 p + \beta(\alpha + \beta)(p + \alpha)}, \\ \rho_2 &= \frac{\alpha p(\alpha + \beta)}{\alpha^2 p + \beta(\alpha + \beta)(p + \alpha)}. \end{aligned} \quad (2.10)$$

Bien que nous ne l'ayons pas utilisé et que nous n'allons pas la démontrer, il existe une formule pour M particulièrement utile si l'on veut calculer numériquement les probabilités et les grandeurs physiques qui en dépendent [RAGOUCY, 2017]. Pour un nombre de sites N quelconque, on peut décomposer la matrice M selon la forme

$$M = A \otimes \mathbb{1}^{N-1} + \sum_{i=0}^{N-2} p_i \mathbb{1}^i \otimes m \otimes \mathbb{1}^{N-2-i} + \mathbb{1}^{N-1} \otimes B \quad (2.11)$$

où  $\mathbb{1}$  est la matrice identité de taille  $2 \times 2$  et  $\mathbb{1}^n$  est le produit tensoriel de  $n$  matrice identité. Les matrices A, B, et m sont

$$A = \begin{pmatrix} -\alpha & 0 \\ \alpha & 0 \end{pmatrix}; \quad B = \begin{pmatrix} -\beta & \beta \\ 0 & -\beta \end{pmatrix}; \quad m = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \quad (2.12)$$

Malgré la simplification du calcul grâce à cette formule, la matrice de Markov étant une matrice carrée de taille  $2^N$ , elle augmente donc exponentiellement avec le nombre de sites N et va devenir déraisonnablement longue à calculer pour des réseaux ayant un nombre de sites de l'ordre de cent, comme le nombre de codons d'un ARNm. Nous avons estimé la loi exponentielle du temps de calcul des densités exactes avec notre programme écrit sous Octave à 3,  $1.10^{-7} e^{1.2N}$  (Fig. 2.3). Pour un réseau de 100 sites, il faudrait donc  $4.10^{45}$  secondes soit  $10^{38}$  années ! Un nombre que même les superordinateurs ne peuvent pas diminuer suffisamment, y compris en utilisant un logiciel optimisé pour la rapidité de calcul. À savoir qu'ils sont environ  $10^6$  fois plus rapides que les ordinateurs de notre laboratoire.

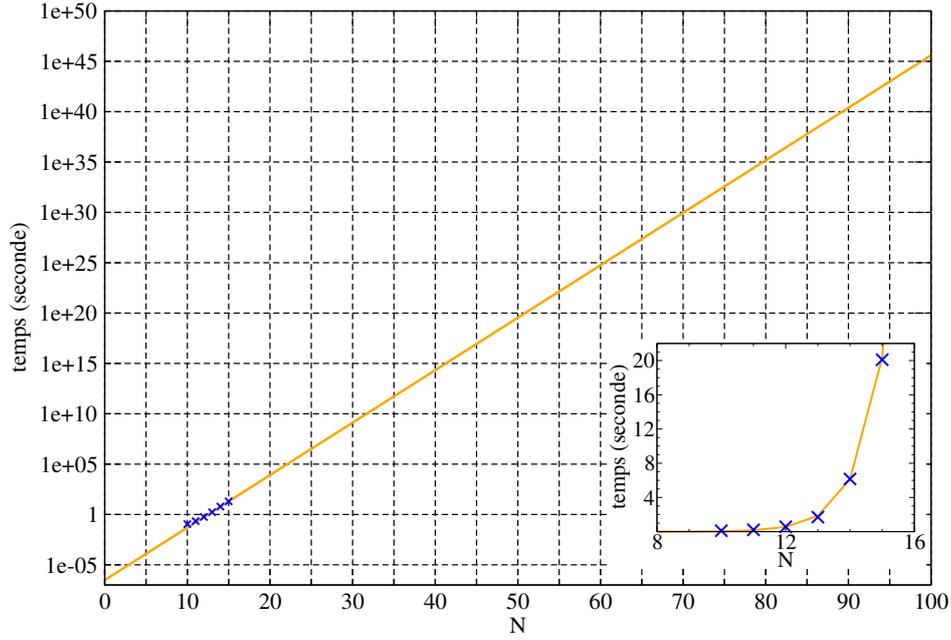


FIGURE 2.3 – Temps de calcul des densités du TASEP exact en fonction du nombre de sites  $N$  du réseau. Les croix bleues sont les temps de calculs obtenus par notre programme Octave, en orange, la loi exponentielle passant par ces points. Échelle semi-logarithmique. Un zoom sur les points fittés est visible dans un insert en échelle linéaire.

### 3.2.2 Caractérisation du TASEP (Ansatz matriciel)

Une méthode efficace pour trouver les caractéristiques du système est celle de l'ansatz matriciel. Il s'agit d'une hypothèse sur la forme des probabilités des états en fonction des paramètres du système qui va se révéler juste après résolution du système. DERRIDA et al. [1993] retrouvent par cette méthode le diagramme de phase précédemment trouvé par DERRIDA et al. [1992] de manière plus efficace, ainsi que les valeurs des densités aux bords et les formes des couches limites (densité vers les bords). Nous allons nous intéresser ici seulement aux résultats élémentaires qui caractérisent le TASEP.

Le diagramme de phases contient trois phases suivant les valeurs de  $\alpha/p$  et  $\beta/p$  (Fig. 2.4b) :

$$\left\{ \begin{array}{ll} \text{Basse densité (LD pour } Low\ Density) \text{ si} & \beta/p > \alpha/p, \alpha/p < 0,5 \\ \text{Haute densité (HD pour } High\ Density) \text{ si} & \alpha/p > \beta/p, \beta/p < 0,5 \\ \text{Courant maximum (MC pour } Maximal\ Current) \text{ si} & \alpha/p > 0,5, \beta/p > 0,5 \end{array} \right.$$

La transition entre la phase LD et la phase HD est caractérisée par ce qu'on appelle un "mur de domaine" qui sépare la basse et la haute densité. Ce mur de domaine fluctue au cours du temps, il diffuse entre les bords du réseau. Sur la Fig. 2.4c, nous avons tracé le profil de densité moyen à la transition entre les deux phases, c'est-à-dire à  $\alpha = \beta$  inférieur à 0.5. La moyenne est faite sur un très grand nombre de configurations à l'état stationnaire ce qui donne le mur de domaine moyen qui s'étale sur tout le réseau.

Avec les trois phases et la transition entre les phases LD et HD, nous avons quatre régimes bien définis par les valeurs aux bords et loin des bords des densités et le courant  $J$  à l'état stationnaire. Dans le cas où  $p=1$ , la résolution exacte nous donne [DERRIDA et al., 1993] :

$$\left\{ \begin{array}{llll} \text{LD:} & \rho_1 = \alpha, & \rho_N = \frac{\alpha(1-\alpha)}{\beta}, & \rho_{bulk} \approx \alpha, & J = \alpha(1-\alpha) \\ \text{HD:} & \rho_1 = 1 - \frac{\beta(1-\beta)}{\alpha}, & \rho_N = 1 - \beta, & \rho_{bulk} \approx 1 - \beta, & J = \beta(1-\beta) \\ \text{MC:} & \rho_1 = 1 - \frac{1}{4\alpha}, & \rho_N = \frac{1}{4\beta}, & \rho_{bulk} \approx \frac{1}{2}, & J = \frac{1}{4} \\ \text{SP:} & \rho_1 = \alpha, & \rho_N = 1 - \alpha, & \rho_{i,bulk} \approx \alpha + i(1-2\alpha), & J = \alpha(1-\alpha) \end{array} \right. \quad (2.13)$$

Ainsi que les lois aux couches limites. Nous n'allons pas les détailler ici, mais simplement noter que pour la phase MC la variation de la densité suit une loi de puissance et pour les phases LD et HD, des lois exponentielles.

Une autre caractéristique importante du TASEP est sa relation fondamentale densité-courant. Elle est donnée par la résolution exacte et peut être retrouvée en utilisant l'approximation de champ moyen. On la comprend aisément pour un réseau TASEP en conditions périodiques. Si on considère un anneau par exemple, il y a invariance par translation des sites sur le réseau. Donc, à la stationnarité, chaque site est équivalent tel que  $\rho_i = \rho_{i+1} = \bar{\rho}$ . La densité sur le réseau est homogène et de valeur  $\bar{\rho} = \frac{k}{N}$ , où  $k$  est le nombre de particules dans le réseau. Les courants entrant et sortant au site  $i$  et au temps  $t$  sont définis par

$$\begin{cases} J_i^+(t) = p \langle n_{i-1}(t)(1 - n_i(t)) \rangle, \\ J_i^-(t) = p \langle n_i(t)(1 - n_{i+1}(t)) \rangle. \end{cases} \quad (2.14)$$

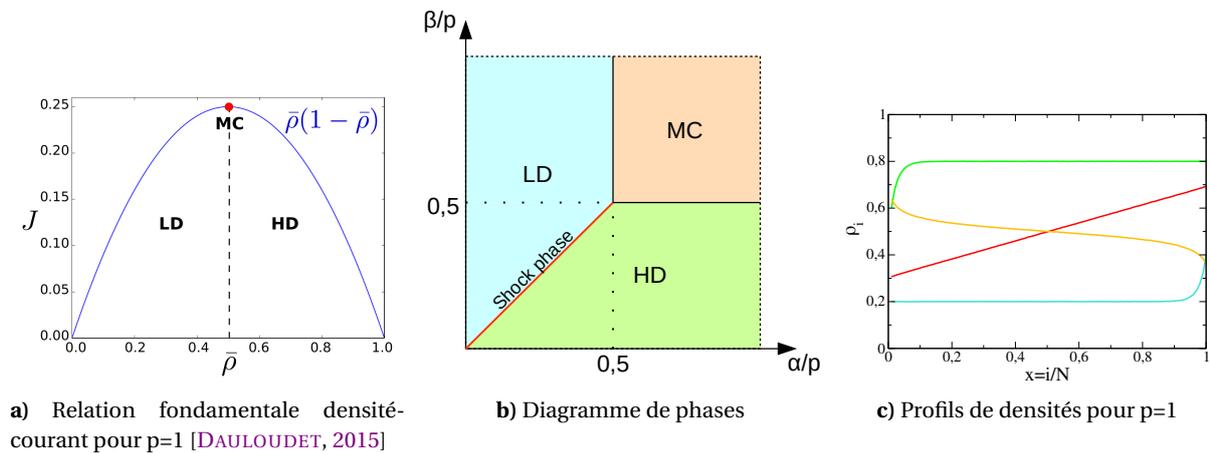
où  $n_i$  est l'occupation au site  $i$  tel que  $\langle n_i \rangle = \rho_i$  et peut donc prendre seulement les valeurs 0 ou 1. L'approximation de champ moyen consiste à négliger les corrélations entre particules en découplant les termes avant de les moyenner. Par exemple,  $\langle n_{i-1}(1 - n_i) \rangle$  s'approxime par  $\langle n_{i-1} \rangle (1 - \langle n_i \rangle)$ . Le courant devient donc à l'état stationnaire :

$$J = p\bar{\rho}(1 - \bar{\rho}). \quad (2.15)$$

Tracer la relation fondamentale densité-courant permet d'avoir une visualisation physique des trois phases (Fig. 2.4a). La parabole obtenue montre aussi qu'il y a une symétrie dans le système. On l'appelle la symétrie trou-particule. Elle laisse les équations de TASEP invariantes par les transformations [PARMEGGIANI et al., 2004] :

$$\begin{cases} n_i \leftrightarrow 1 - n_{N-i} \\ \alpha \leftrightarrow \beta \end{cases} \quad (2.16)$$

Ainsi, on peut considérer que  $\beta$  est le taux d'injection des trous et que  $1 - \alpha$  est le taux de sortie des trous.



**FIGURE 2.4** – Les 3 phases de TASEP. (a) On distingue trois phases : basse densité (LD), haute densité (HD), courant maximum (MC). (b) Suivant les paramètres  $\alpha$ ,  $\beta$ ,  $p$ , le système se trouve dans l'une des trois phases ou dans la phase de choc des phases LD et HD. (c) (bleu) LD :  $\alpha = 0.2$ ,  $\beta = 0.4$ , (vert) HD :  $\alpha = 0.4$ ,  $\beta = 0.2$ , (orange) MC :  $\alpha = 0.7$ ,  $\beta = 0.7$ , (rouge) SP :  $\alpha = 0.3$ ,  $\beta = 0.3$ . Obtenu par simulation Monte Carlo.

La symétrie trou-particule et ces valeurs de la densité permettent d'expliquer les quatre régimes. Pour plus de clarté, considérons le cas  $p=1$  et l'on pourra se référer aux Fig. 2.4a, 2.4b et 2.4c pour l'explication suivante. (LD) Quand  $\alpha < \beta$  et  $\alpha < 0,5$ , il est injecté plus de trous que de particules. La densité de particules dans le *bulk* (la zone du réseau qui n'est ni l'entrée ni la sortie)

est donc faible (inférieure à 0,5) et le paramètre limitant est  $\alpha$ . Ce qui est cohérent avec la valeur exacte de la densité à l'entrée du réseau en régime LD (Eq. (2.13)) qui ne dépend que de  $\alpha$  et qui est aussi la valeur de la densité dans le *bulk*. (HD) Quand  $\alpha > \beta$  et  $\beta < 0,5$ , il est injecté plus de particules que de trous. La densité de particule dans le *bulk* sera donc élevée (supérieure à 0,5). Le facteur limitant est alors  $\beta$  comme le montre la valeur exacte de la densité à la sortie du réseau en régime HD (Eq. (2.13)) qui ne dépend que de  $\beta$  et qui est aussi la valeur de la densité dans le *bulk*. (MC) Dans le cas où  $\alpha > 0,5$  et  $\beta > 0,5$ , il est introduit un grand nombre de particules et un grand nombre de trous. A la stationnarité, les particules s'organisent telles que la densité moyenne vaut 0,5. La densité ainsi que le courant plafonnent. (SP) Quand  $\alpha = \beta$  et  $\alpha, \beta < 0,5$ , il se forme une compétition trou-particule et le système se trouve dans la phase de choc entre les phases haute densité et basse densité. Cette phase met en évidence l'existence d'un mur de domaines (DW pour *Domain Wall*) qui sépare les régimes LD, à sa droite et le régime HD, à sa gauche.

On remarque que lorsque le système se trouve en phase LD, le DW est repoussé à la sortie du réseau puisque la basse densité se trouve à sa gauche. On peut dire que la compétition trou-particules est remportée par les trous. De même, lorsque le système se trouve en phase HD, le DW est repoussé à l'entrée du réseau puisque la basse densité se trouve à sa droite. La compétition trou-particules est remportée par les particules.

### 3.3 Approximation de champ moyen

On s'intéresse maintenant aux équations maîtresses du système non pas sur les probabilités, mais sur la densité de particules. Ceci permettra d'obtenir des expressions des densités à l'état stationnaire en fonction de la position sur le réseau.

Soit  $n_i$  la variable qui définit l'occupation du site  $i$  tel que  $n_i = 0$  si le site  $i$  est vide et  $n_i = 1$  si une particule occupe le site  $i$ . Nous allons déduire la variation d'occupation moyenne au cours du temps de chaque site grâce aux détails de leur mouvement stochastique :

- Une particule entre sur le réseau avec un taux  $\alpha$  et s'il n'y a pas déjà une particule sur le premier site. La probabilité que la particule entre sur le réseau est donc  $\alpha dt$  multipliée par la probabilité que le premier site ne soit pas occupé.
- Une particule sur le site  $i$  saute sur le site suivant avec un taux  $p$  s'il n'est pas occupé. La probabilité de saut d'une particule du site  $i$  au site  $i + 1$  est donc  $p dt$  multiplié par la probabilité que le site  $i$  soit occupé, mais pas le site  $i + 1$ .
- Une particule sort du réseau avec un taux  $\beta$ . La probabilité qu'une particule sorte du réseau est donc  $\beta dt$  multiplié par probabilité que le dernier site soit occupé.

La probabilité d'occupation d'un site est obtenue grâce à la moyenne sur un grand nombre de configurations, ce qui nous permet d'écrire les équations des variations temporelles d'occupation moyenne en chaque site comme

$$\frac{d}{dt} \langle n_1 \rangle = \alpha(1 - \langle n_1 \rangle) - p \langle n_1(1 - n_2) \rangle, \quad (2.17a)$$

$$\frac{d}{dt} \langle n_i \rangle = p \langle n_{i-1}(1 - n_i) \rangle - p \langle n_i(1 - n_{i+1}) \rangle, \quad i = 2, 3, \dots, N-1 \quad (2.17b)$$

$$\frac{d}{dt} \langle n_N \rangle = p \langle n_{N-1}(1 - n_N) \rangle - \beta \langle n_N \rangle. \quad (2.17c)$$

Effectuons maintenant l'approximation de champ moyen qui consiste à négliger les corrélations entre proches voisins afin de pouvoir moyenner sur l'occupation d'un seul site à la fois. Par exemple, le terme  $\langle n_{i-1}(1 - n_i) \rangle$  s'approxime par  $\langle n_{i-1} \rangle(1 - \langle n_i \rangle)$ . En notant  $\rho_i \equiv \langle n_i \rangle$ , les équations (2.17c)

deviennent donc

$$\frac{d}{dt}\rho_1 = \alpha(1 - \rho_1) - p\rho_1(1 - \rho_2), \quad (2.18a)$$

$$\frac{d}{dt}\rho_i = p\rho_{i-1}(1 - \rho_i) - p\rho_i(1 - \rho_{i+1}), \quad i = 2, 3, \dots, N-1 \quad (2.18b)$$

$$\frac{d}{dt}\rho_N = p\rho_{N-1}(1 - \rho_N) - \beta\rho_N. \quad (2.18c)$$

Nous allons voir que la loi de conservation de la matière est bien respectée. En effet, puisque le courant entrant s'écrit

$$\begin{cases} J_1^+ = \alpha(1 - \rho_1), \\ J_i^+ = p\rho_{i-1}(1 - \rho_i), \quad i = 2, \dots, N \end{cases} \quad (2.19)$$

et le courant sortant s'écrit

$$\begin{cases} J_i^- = p\rho_i(1 - \rho_{i+1}), \quad i = 1, \dots, N-1 \\ J_N^- = \beta\rho_N, \end{cases} \quad (2.20)$$

on a bien

$$\frac{d}{dt}\rho_i = J_i^+ - J_i^-, \quad \forall i. \quad (2.21)$$

On retrouve aisément les densités et le courant pour la phase basse densité donnés par la résolution exacte (Eq. (2.13)). En effet, lorsque  $\frac{\alpha}{p} \ll 1$ , on peut considérer que la densité est négligeable devant 1 dans l'Eq. (2.18a), ce qui donne l'expression de  $\rho_1$ . On peut ensuite injecter  $\rho_1$  dans l'expression de  $J_1^+$  (Eq.(2.19)) pour obtenir le courant, puis finalement utiliser l'égalité entre  $J_1^+$  et  $J_N^-$  (Eq. (2.20)) pour retrouver  $\rho_N$ . On a donc

$$J = \alpha \left(1 - \frac{\alpha}{p}\right) \approx \alpha, \quad (2.22a)$$

$$\rho_1 = \rho_{bulk} = \frac{\alpha}{p}, \quad (2.22b)$$

$$\rho_N = \frac{\alpha}{\beta} \left(1 - \frac{\alpha}{p}\right) \approx \frac{\alpha}{\beta}, \quad (2.22c)$$

ce qui correspond aux Eq. (2.13) quand  $p = 1$  en phase LD. On retrouve maintenant également la densité de la phase HD en effectuant la symétrie trou-particule (2.16) pour obtenir  $\rho_{N-1}$  à partir de l'Eq. (2.22b). Comme le bord limitant est la sortie du réseau, on considère que la densité dans le *bulk* est égale à  $\rho_N$  et donc  $\rho_N = \rho_{N-1} = \rho_{bulk}$ . Par l'égalité entre le courant entrant  $J_1^+$  et le courant sortant  $J_N^-$  on obtient ensuite  $\rho_1$ . On a donc

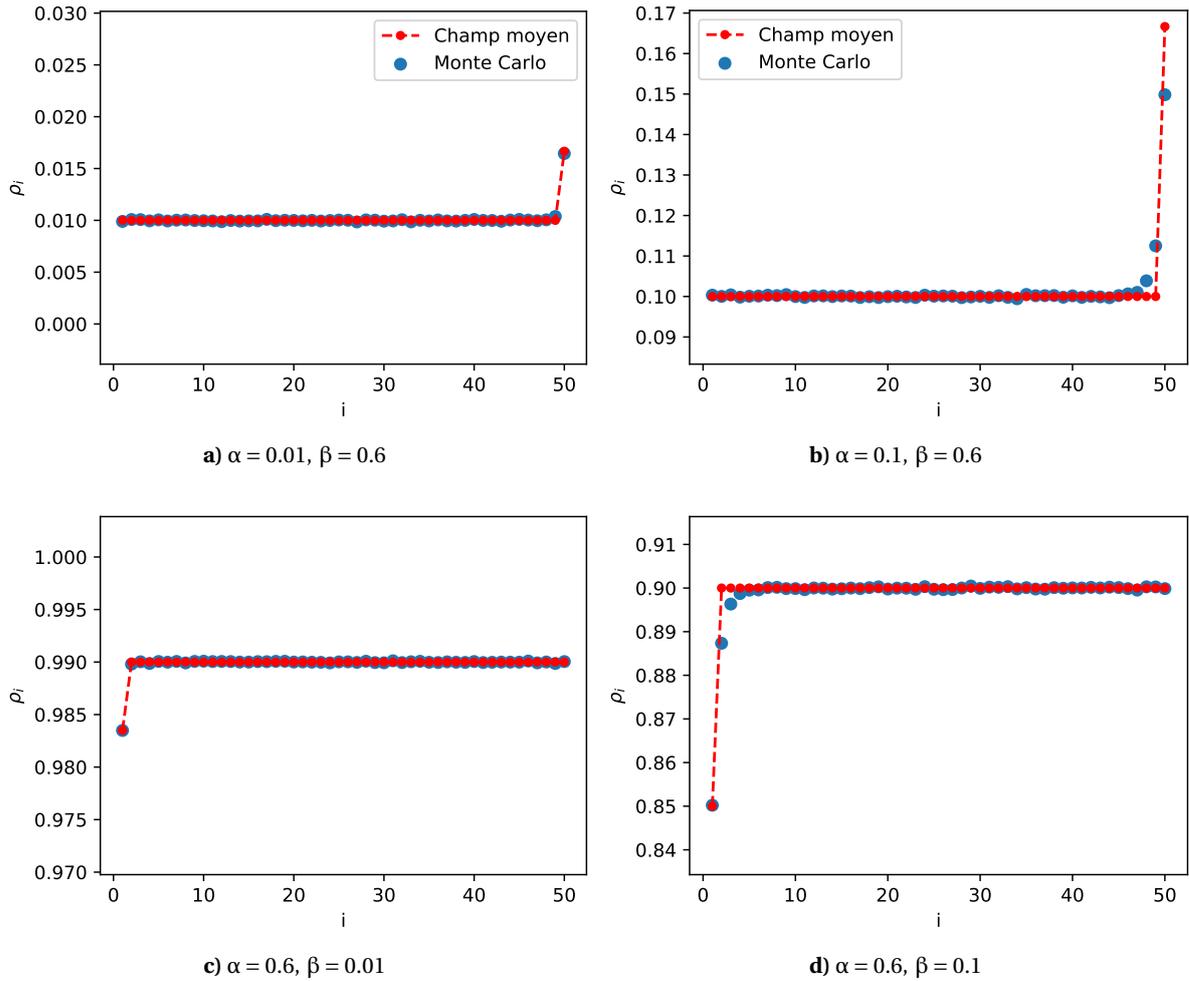
$$J = \beta \left(1 - \frac{\beta}{p}\right), \quad (2.23a)$$

$$\rho_1 = 1 - \frac{\beta}{\alpha} \left(1 - \frac{\beta}{p}\right), \quad (2.23b)$$

$$\rho_N = \rho_{bulk} = 1 - \frac{\beta}{p}, \quad (2.23c)$$

ce qui correspond aux équations exactes (2.13) quand  $p = 1$  en phase HD. Nous remarquons sur la Fig. 2.5 que l'effet de bord n'est pas étendu et qu'il peut être négligé pour un réseau assez grand. Les équations de champ moyen suffisent alors à décrire le courant et la densité du modèle TASEP.

Maintenant que nous avons décrit et caractérisé l'essentiel du modèle TASEP, on va le modifier de manière à ce qu'il corresponde mieux à ce qu'on sait du processus de traduction génétique.



**FIGURE 2.5** – Comparaison des profils de densité TASEP en phase basse et haute densité obtenus en simulation Monte Carlo avec l’algorithme de Gillespie et en champ moyen ( $\rho_{bulk} = \frac{\alpha}{\beta}$ ,  $\rho_N = \frac{\alpha}{\beta}$  pour les graphes a et b) avec  $p = 1$ . On remarque que les profils de densité TASEP en phase basse et haute densité sont très bien décrits par les équations de champ moyen, sauf au niveau de la couche limite. La couche limite est cependant très courte et pourra être négligée pour des réseaux assez grands.

## 4 Évolution du modèle : $\ell$ -TASEP inhomogène

### 4.1 Le modèle TASEP avec des taux de sauts site-spécifiques et des particules étendues

Dès 1969, le modèle TASEP est étendu pour intégrer des caractéristiques essentielles du phénomène de transport qu’est la traduction [MACDONALD et GIBBS, 1969]. La particule recouvre plusieurs sites tout en se déplaçant d’un seul site à la fois. Sa longueur est  $\ell$  et le modèle nommé  $\ell$ -TASEP. Il faut alors choisir la manière dont la particule entre et sort du réseau bien que ce choix ait un impact minime sur la cinétique et les grandeurs physiques associées. Nous avons fait ce choix en fonction de ce qu’on sait du mécanisme biologique de la traduction. Pour modéliser la traduction et ensuite les expériences Ribo-Seq, le choix naturel est de considérer la partie codant la protéine, *open reading frame (ORF)*, avec le codon *start* et le codon *stop*, car c’est sur le codon *start* que les sous-unités du ribosome s’assemblent et sur le codon *stop* qu’elles se désassemblent en leur milieu. Sachant que le ribosome recouvre environ 9 codons, et pour éviter des complications techniques peu intéressantes, nous allons considérer le cas de  $\ell$  impair uniquement. Notre particule entrera donc sur le réseau en recouvrant les  $(\ell + 1)/2$  premiers sites et pourra en sortir lorsque son milieu sera sur le dernier site. Les taux de sauts ne sont plus constants le long du réseau, mais sont distribués aléatoirement à chaque site. On dit que les taux de saut sont "site spécifique" ou plus succinctement "inhomogènes". Nous modélisons donc la traduction génétique

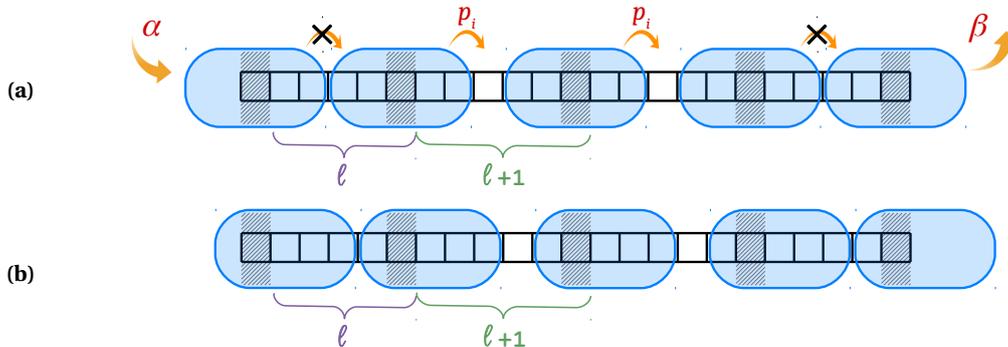
avec le modèle  $\ell$ -TASEP inhomogène.

## 4.2 Densité et courant à l'état stationnaire

Nous allons calculer la densité et le flux en considérant d'abord les équations exactes sur la variation de l'occupation moyenne, puis en effectuant une approximation de champ moyen sur ces équations. Considérer des taux de saut site-spécifiques n'apporte pas de difficulté à la formulation analytique. Il suffit de remplacer le taux de saut  $p$  par  $p_i$  avec  $i$  le site correspondant. Mais les résultats des densités et du courant à l'état stationnaire vont être considérablement changés. Considérer une particule étendue, au contraire, est plus difficile à formuler alors que son impact sur la densité et le courant est beaucoup plus trivial. En effet, certains états ne sont plus possibles et il faut donc définir une nouvelle matrice de Markov. De plus, le processus cinétique n'est plus le même, ce qui va modifier les équations maîtresses, donc la matrice de Markov et les équations de champ moyen. Mais le diagramme des phases n'est modifié que très légèrement (Fig. 2.7). Le diagramme de phase du  $\ell$ -TASEP inhomogène, prenant en compte les deux évolutions au modèle TASEP, a été déterminé très récemment par ERDMANN-PHAM et al. [2020]. Nous ne l'avons pas étudié dans cette thèse, mais nous en donnons la représentation en annexe (Fig. A.1).

### 4.2.1 Équations exactes sur la variation de l'occupation moyenne

La dynamique du  $\ell$ -TASEP est illustrée sur la Fig. 2.6 et décrite comme suit. Une particule est comptée par son centre ce qui signifie que l'occupation  $n_i(t)$  est égale à 1 si le centre d'une particule est sur le site  $i$  et 0 sinon. Pour  $\ell$  égal à un nombre impair de sites, le centre de la particule est évidemment  $\frac{\ell+1}{2}$ . On dit que le site de comptage est le site  $\frac{\ell+1}{2}$  si  $\ell$  est un nombre impair. La particule entre directement dans le réseau en fixant son centre sur le premier site et quitte directement le réseau lorsque son centre se trouve sur le dernier site. Une propriété importante de ce modèle est que les équations restent les mêmes pour tout site de comptage tant que la particule entre et sort du réseau par son site de comptage.



**FIGURE 2.6** – Schéma  $\ell$ -TASEP pour  $l = 5$  sites. Les taux de saut  $p_i$  peuvent être différents pour chaque site  $i = 1 \dots N$ .  $\alpha \equiv p_0$  est le taux d'entrée et  $\beta \equiv p_N$  le taux de sortie. Les particules ne peuvent pas se couvrir les unes les autres. (a) Une particule est comptée par son centre (partie hachurée). Elle entre directement dans le réseau en fixant son centre sur le premier site et sort directement du réseau lorsque son centre est sur le dernier site. (b) Une particule est comptée par une autre partie d'elle-même (partie hachurée), entre directement dans le réseau en attachant cette partie sur le premier site et quitte directement le réseau lorsque cette partie est sur le dernier site. On peut voir que le comptage ne change pas par rapport à (a).

La variation de l'occupation moyenne du site  $i$ ,  $\langle n_i(t) \rangle$ , doit être décrite par 4 équations en raison des différentes corrélations locales :

- Site 1 : les  $\ell$  premiers sites doivent valoir 0 pour que la particule puisse entrer sur le réseau.
- Sites  $i = 2 \dots N - \ell$  : le site  $i - 1 + \ell$  doit valoir 0 pour qu'une particule se déplace sur le site  $i$  et le site  $i + \ell$  doit valoir 0 pour que la particule quitte le site  $i$ .

- Site  $N - \ell + 1$  : le site  $N$  doit valoir 0 pour qu'une particule se déplace sur le site  $N - \ell + 1$ , mais une particule sur ce site ne peut pas être gênée pour avancer.
- Sites  $N - \ell + 2 \dots N$  : la particule n'est gênée ni pour avancer ni pour quitter le site.

Les équations de variation de l'occupation moyenne sur le site  $i$  sont donc :

$$\begin{aligned} \frac{d\langle n_1(t) \rangle}{dt} &= p_0 \langle \prod_{j=1}^{\ell} (1 - n_j(t)) \rangle - p_1 \langle n_1(t)(1 - n_{1+\ell}(t)) \rangle, \\ \frac{d\langle n_i(t) \rangle}{dt} &= p_{i-1} \langle n_{i-1}(t)(1 - n_{i-1+\ell}(t)) \rangle - p_i \langle n_i(t)(1 - n_{i+\ell}(t)) \rangle, \quad i = 2 \dots N - \ell \\ \frac{d\langle n_{N-\ell+1}(t) \rangle}{dt} &= p_{N-\ell} \langle n_{N-\ell}(t)(1 - n_N(t)) \rangle - p_{N-\ell+1} \langle n_{N-\ell+1}(t) \rangle, \\ \frac{d\langle n_i(t) \rangle}{dt} &= p_{i-1} \langle n_{i-1}(t) \rangle - p_i \langle n_i(t) \rangle, \quad i = N - \ell + 2 \dots N, \end{aligned} \quad (2.24)$$

avec  $p_0$  le taux d'entrée (ou taux d'initiation qu'on note aussi  $\alpha$ ),  $p_N$  le taux de sortie (noté aussi  $\beta$ ),  $p_i$  les taux de saut et  $n_0(t) = 1$  (réservoir infini). Mais, à cause de l'exclusion des particules,

$$n_i(t)n_j(t) = 0 \quad \text{pour } |i - j| < \ell, \forall i > 0, \forall j > 0. \quad (2.25)$$

Par conséquent, le terme d'exclusion à l'entrée est linéarisé :

$$\langle \prod_{j=1}^{\ell} (1 - n_j(t)) \rangle = 1 - \sum_{j=1}^{\ell} \langle n_j(t) \rangle \quad (2.26)$$

et nous pouvons réécrire l'occupation moyenne du site 1 comme

$$\frac{d\langle n_1(t) \rangle}{dt} = p_0 \left( 1 - \sum_{j=1}^{\ell} \langle n_j(t) \rangle \right) - p_1 \langle n_1(t)(1 - n_{1+\ell}(t)) \rangle. \quad (2.27)$$

De ces équations, nous déduirons les équations en champ moyen du  $\ell$ -TASEP inhomogène ainsi que les équations d'un modèle similaire qui sera introduit plus loin, le TAEBEP.

#### 4.2.2 Équations de champ moyen

Dans ce paragraphe, on va retrouver l'expression du courant de [MACDONALD et al. \[1968\]](#) avec une explication adaptée au formalisme choisi ici. On en déduira ensuite les expressions des densités en chaque point pour le cas stationnaire. On peut lire sur la Fig. 2.6 qu'il y a déplacement au site  $i$  si celui-ci est occupé et si le site  $i + \ell$  est libre. Le courant s'écrit donc

$$J_i^+(t) = p_i \rho_i(t) f_{i+\ell}(t) \quad (2.28)$$

où  $f_{i+\ell}(t)$  est la probabilité que le site  $i + \ell$  soit vide sachant que le site  $i$  est occupé. Cette probabilité étant difficile à calculer, nous faisons une approximation de champ moyen qui consiste à la remplacer par une probabilité facile à calculer. On peut suggérer la présence de la particule en  $i$  en considérant que les sites  $i + 1$  à  $i + \ell - 1$  doivent être vides. On définit la probabilité conditionnelle  $F_{i+\ell}$  que les sites  $i + 1$  à  $i + \ell$  soient vides sachant que les sites  $i + 1$  à  $i + \ell$  sont vides ou que seul le site  $i + \ell$  est occupé. Pour décrire cette probabilité, on écrit les états des sites  $i + 1$  à  $i + \ell$  sous la forme  $00\dots n$  avec  $n$  égal à 0 ou 1. D'après le théorème de Bayes, on a

$$F_{i+\ell}(t) = P(00\dots 0 \mid [00\dots 0 \cup 00\dots 1])(t) = \frac{P(00\dots 0 \cap [00\dots 0 \cup 00\dots 1])(t)}{P(00\dots 0 \cup 00\dots 1)(t)} = \frac{P(00\dots 0)(t)}{P(00\dots 0 \cup 00\dots 1)(t)} \quad (2.29)$$

qu'on sait écrire en terme d'occupation moyenne. En effet,

$$\begin{aligned} P(00\dots 0)(t) &= \langle \prod_{j=1}^{\ell} (1 - n_{i+j}(t)) \rangle, \\ P(00\dots 0 \cup 00\dots 1)(t) &= \langle \prod_{j=1}^{\ell} (1 - n_{i+j}(t)) \rangle + \langle n_{i+\ell} \prod_{j=1}^{\ell-1} (1 - n_{i+j}(t)) \rangle, \end{aligned} \quad (2.30)$$

qui devient d'après la propriété (2.25)

$$\begin{aligned} P(00\dots 0)(t) &= 1 - \sum_{j=1}^{\ell} \langle n_{i+j}(t) \rangle, \\ P(00\dots 0 \cup 00\dots 1)(t) &= 1 - \sum_{j=1}^{\ell} \langle n_{i+j}(t) \rangle + \langle n_{i+\ell}(t) \rangle. \end{aligned} \quad (2.31)$$

En considérant  $f_{i+\ell}(t) = F_{i+\ell}(t)$  en champ moyen, on obtient donc l'expression du courant

$$J_i^+(t) = p_i \rho_i(t) \frac{1 - \sum_{j=1}^{\ell} \rho_{i+j}(t)}{1 - \sum_{j=1}^{\ell} \rho_{i+j}(t) + \rho_{i+\ell}(t)} \quad (2.32)$$

Nous en déduisons avec les Eq. (2.24) et (2.27) les expressions de la densité à l'état stationnaire,

$$J = \alpha \left( 1 - \sum_{s=1}^{\ell} \rho_s \right), \quad (2.33a)$$

$$\rho_i = \frac{J}{p_i} \frac{1 - \sum_{j=1}^{\ell} \rho_{i+j} + \rho_{i+\ell}}{1 - \sum_{j=1}^{\ell} \rho_{i+j}}, \quad i = 1, 2, \dots, N - \ell \quad (2.33b)$$

$$\rho_i = \frac{J}{p_i}, \quad i = N - \ell + 1, N - \ell + 2, \dots, N \quad (2.33c)$$

où  $J$  est le courant stationnaire défini à l'entrée du réseau et  $p_N \equiv \beta$ . Ce sont les équations données aussi par SHAW et al. [2004].

### 4.3 Diagramme de phases

Les équations de champ moyen ci-dessus ne sont pas explicites. Elles ne suffisent donc pas à la détermination d'un diagramme de phase. LAKATOS et CHOU [2003] ont proposé une résolution dans le cas du  $\ell$ -TASEP homogène à l'aide du modèle du gaz de Tonks et d'une approche "champ moyen amélioré" ("*Refined mean-field theory*") (Fig. 2.7). Cette approche de champ moyen améliorée a été pensée pour trouver les valeurs explicites de la densité et du courant du modèle  $\ell$ -TASEP homogène des phases où le courant est limité par les bords, c'est-à-dire les phases de haute et de basse densité. Dans la phase LD, et pour le cas homogène, la densité peut être considérée comme constante vers l'entrée du réseau. Il est alors possible d'évaluer cette densité en exprimant la probabilité que les  $\ell$  premiers sites soient vident [LAKATOS et CHOU, 2003]. Un raisonnement similaire permet d'obtenir aussi la densité et le courant en fonction des paramètres  $\alpha$ ,  $\beta$  et  $\ell$ . On donne les valeurs du courant et de la densité en chaque phase pour un taux de saut  $p = 1$  :

$$\left\{ \begin{array}{ll} \text{LD :} & \rho_{bulk} = \frac{\alpha}{1+(\ell-1)\alpha}, \quad J = \frac{\alpha(1-\alpha)}{1+(\ell-1)\alpha} \\ \text{HD :} & \rho_{bulk} = \frac{1-\beta}{1+(\ell-1)\beta}, \quad J = \frac{\beta(1-\beta)}{1+(\ell-1)\beta} \\ \text{MC :} & \rho_{bulk} = \frac{1}{\sqrt{\ell}(\sqrt{\ell}+1)}, \quad J = \frac{1}{(\sqrt{\ell}+1)^2} \end{array} \right. \quad (2.34)$$

Sans étonnement, on constate que le courant et la densité ne sont pas fondamentalement changés, mais modulés à la baisse par la longueur de la particule. Les valeurs de  $\alpha$  et  $\beta$  délimitant les phases sont obtenues en prenant respectivement l'égalité entre les courant en basse densité - courant maximum, haute densité - courant maximum. Il est intéressant de noter que bien que le diagramme de phase garde une symétrie par rapport à l'axe  $\alpha = \beta$ , la symétrie trou-particule est brisée pour  $\ell > 1$ . En effet, on doit toujours considérer qu'un trou fait la taille d'un site pour avoir un système où la particule étendue se déplace d'un seul site à la fois. Ainsi, quand une particule entre (sort),  $\ell$  trous sortent (entrent), ce qui brise la symétrie établie pour le TASEP (2.16).

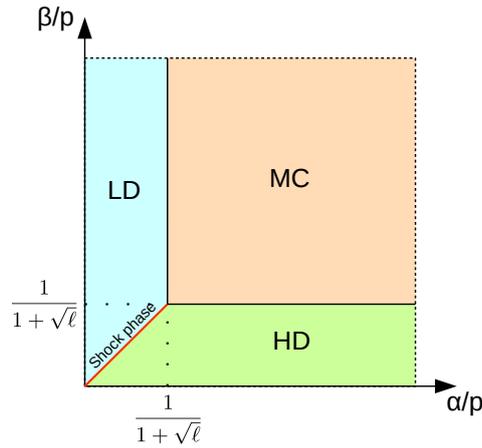


FIGURE 2.7 – Diagramme de phases du  $\ell$ -TASEP homogène où  $\ell$  est la longueur de la particule en nombre de sites. La seule différence avec le diagramme de phases du TASEP homogène est la valeur seuil pour  $\alpha$  et  $\beta$  qui délimite les 3 phases (HD : haute densité, LD : basse densité, MC : courant maximum).

## 5 Résultats préliminaires à propos de l'effet de la particule étendue sur la densité de particules

Dans cette partie, nous décrivons deux résultats que nous avons obtenus concernant l'effet de la particule étendue sur la densité de particules du  $\ell$ -TASEP homogène. Pour cela, nous utilisons les équations en champ moyen du  $\ell$ -TASEP inhomogène, des simulations Monte Carlo et nous proposons un modèle simplifié du  $\ell$ -TASEP inhomogène où l'exclusion entre particules n'est considérée qu'à l'entrée du réseau.

### 5.1 $\ell$ -TASEP homogène : effet de bord sur la densité de particules en phase basse densité

Nous définissons la couche limite comme étant la longueur sur laquelle un effet de bord (ou de taille finie) est ressenti sur la densité de particule. Dans le modèle  $\ell$ -TASEP, la couche limite peut être très importante dans la phase LD jusqu'à être aussi grande que le réseau comme on peut le voir sur la Fig. 2.8. Nous voulons savoir quand est ce que nous pouvons négliger l'effet de bord afin de pouvoir considérer la densité majoritairement constante le long du réseau. Ceci afin de comparer cette densité caractéristique à celle d'autres modèles. Pour cela, nous allons utiliser les équations du TASEP champ moyen dans le cas homogène. En considérant la densité constante sur les premiers sites, le courant du modèle TASEP inhomogène en champ moyen (Eq. (2.33a)) devient  $J = \alpha(1 - \ell\rho)$ , et la densité avant la couche limite (Eq. (2.33b)) devient :

$$\rho = \frac{1}{\frac{\beta}{\alpha} + \ell - 1}. \quad (2.35)$$

On peut calculer une estimation de la densité sur les  $\ell$  derniers sites de la couche limite en utilisant la densité en champ moyen inhomogène (Eq. (2.33b)) puisque nous connaissons la densité pour  $i \geq N - \ell$ . En effet, on calcule d'abord la densité en  $N - \ell$  qui ne dépend que de densités connues, puis nous pouvons calculer la densité en  $N - \ell - 1$ , puis en  $N - \ell - 2$ , et ainsi de suite jusqu'à  $i = N - 2\ell - 1$ . En résumé, les densités du  $\ell$ -TASEP homogène,  $\ell \geq 2$ , en champ moyen et en considérant

l'effet de bord jusqu'à  $i = N - 2\ell + 1$ , peuvent s'écrire

$$\left\{ \begin{array}{l} \rho_i = \frac{1}{\frac{p}{\alpha} + \ell - 1}, \quad i = 1, 2, \dots, N - 2\ell \\ \rho_i = \frac{J}{p} \frac{1 - \sum_{j=i+1}^{N-\ell} \rho_j - [i - (N - 2\ell + 1)] \frac{J}{p}}{1 - \sum_{j=i+1}^{N-\ell} \rho_j - [i - (N - 2\ell)] \frac{J}{p}}, \quad i = N - 2\ell + 1, \dots, N - \ell - 1 \\ \rho_{N-\ell} = \frac{J}{p} \frac{1 - (\ell - 1) \frac{J}{p}}{1 - (\ell - 1) \frac{J}{p} - \frac{J}{\beta}}, \\ \rho_i = \frac{J}{p}, \quad i = N - \ell + 1, \dots, N - 1 \\ \rho_N = \frac{J}{\beta}, \end{array} \right. \quad (2.36)$$

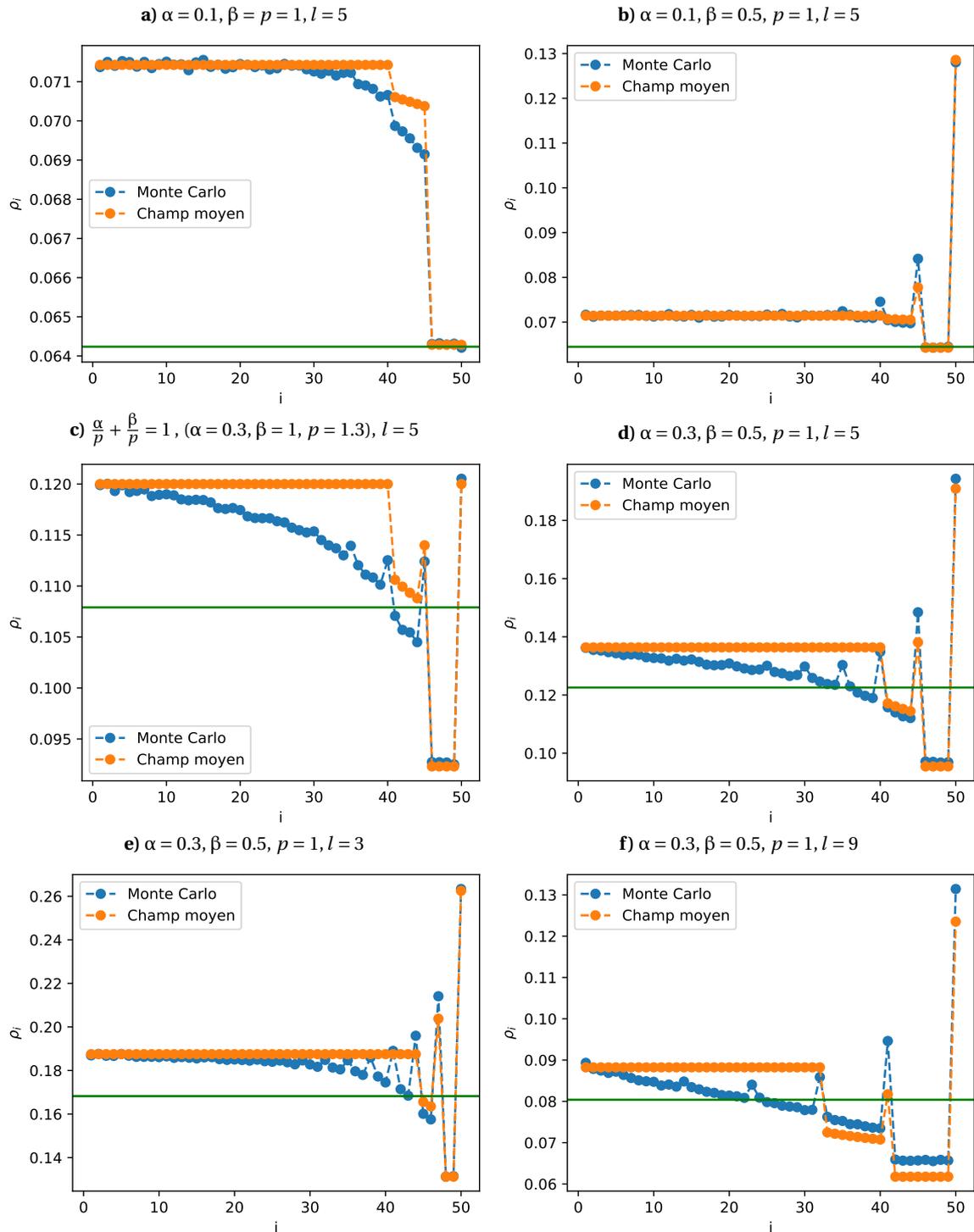
où  $J = \alpha \left(1 - \frac{\ell}{\frac{p}{\alpha} + \ell - 1}\right)$ . Pour  $p = 1$ , on retrouve ainsi le courant et la densité en basse densité de **LAKATOS** et **CHOU** [2003] (Eq. (2.34)). Sur la Fig. 2.8 on observe un bon accord entre ces équations et la simulation du TASEP concernant la densité dans le *bulk* et la fin de la couche limite. Nous pourrions continuer à calculer la densité sur les sites précédents de la même manière, mais leur complexité augmenterait et il faudrait alors procéder à des approximations pour pouvoir les exploiter. Or, nous avons trouvé une manière simple de caractériser la couche limite.

On observe sur la Fig. 2.8 que l'effet de bord évolue par paliers de taille  $\ell - 1$  traduisant un embouteillage de particules de taille  $\ell$  en fin de réseau. Nous appelons dernier palier, le palier le plus proche de la fin du réseau. On peut considérer que l'effet de bord (et sa couche limite) est trop important pour être négligé quand son dernier palier est inférieur ou supérieur de 10% à la densité dans le *bulk*. La quantité à observer est donc la différence relative absolue entre  $\rho_{bulk} = \rho_1$  et  $\rho_{N-1}$ , soit

$$\Delta\rho_{bulk-CL} = \frac{\alpha}{p}. \quad (2.37)$$

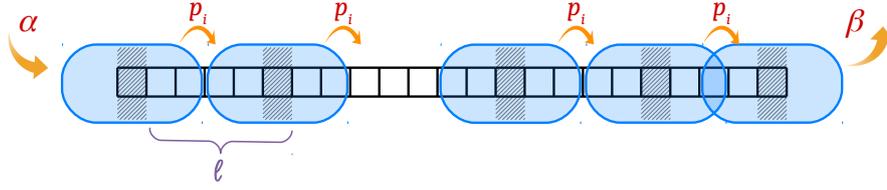
Il est intéressant de constater que cette quantité ne dépend pas de  $\beta$ . Ainsi, la densité au dernier site peut être très différente de celle dans le *bulk* sans que la couche limite ait une longueur importante. Elle ne dépend pas non plus de  $\ell$  ( $\ell > 1$ , le terme de la couche limite que nous regardons disparaît pour  $\ell = 1$ ). Nous considérons donc que la couche limite est négligeable lorsque  $\frac{\alpha}{p} < 0.1$ . Nous illustrons cette dépendance exclusive de la longueur de la couche limite en  $\frac{\alpha}{p}$  dans la Fig. 2.8 représentant la densité TASEP homogène obtenue par simulation avec l'algorithme de Gillespie et celle donnée par les équations de champ moyen (2.36). Sur chaque graphe, une ligne a été ajoutée pour indiquer la valeur à moins 10% de la densité dans le *bulk*.

Sur les graphes 2.8a et 2.8b, on vérifie que l'écart entre le dernier palier de l'effet de bord est de 10% quand  $\frac{\alpha}{p} = 0.1$  quelque soit  $\beta$  et donc la densité au dernier site. Sur le graphe 2.8c, on montre que les amplitudes de l'effet de bord et de sa couche limite peuvent être très importants bien que la densité au dernier site soit égale à la densité dans le *bulk*. Avec les graphes 2.8d, 2.8d et 2.8e entre lesquels seule la taille de la particule change, on voit que la taille des paliers est toujours égale à  $\ell - 1$ . Cependant, on observe que la densité semble toujours évoluer sur toute la longueur du réseau, ce qui suggère que la taille de la couche limite ne dépend pas non plus de  $\ell$ . On vérifie de plus que le dernier palier est 30% en dessous de la valeur de la densité dans le *bulk* comme prédit par l'Eq. (2.37).



**FIGURE 2.8** – Comparaison entre le  $\ell$ -TASEP homogène en simulation Monte Carlo par l’algorithme de Gillespie ( $10^8$  itérations) et les équations de champ moyen (Eq. (2.36)). La ligne verte indique la valeur de la densité du *bulk* prédite par les équations de champ moyen abaissée de 10%. On en déduit que l’effet de bord est bien prédit par l’expression de la densité en champ moyen et on vérifie que la différence absolue entre  $\rho_{bulk} = \rho_1$  et le dernier palier de l’effet de bord est bien  $\Delta\rho_{bulk-CL} = \frac{\alpha}{p}$ . L’effet de bord est calculé et représenté ici qu’à partir de  $i = N - 2\ell + 1$ , ce qui explique l’écart entre la densité obtenue par la simulation et celle obtenue en champ moyen lorsque la couche limite est importante par rapport à la taille du réseau. Nous voyons aussi que la couche limite peut être très importante dans la phase basse densité, jusqu’à être aussi grande que le réseau. On remarque que la hauteur et la longueur de l’effet de bord varient indépendamment de  $\beta$ .

## 5.2 $\ell$ -TAEBEP



**FIGURE 2.9** – Schéma  $\ell$ -TAEBEP pour  $\ell = 5$  sites. Comme dans le modèle  $\ell$ -TASEP, les taux de saut  $p_i$  peuvent être différents pour chaque site  $i = 1 \dots N$ ,  $\alpha$  est le taux d'entrée et  $\beta \equiv p_N$  le taux de sortie, les particules entrent directement dans le réseau en attachant leur centre sur le premier site et quittent directement le réseau lorsque leur centre est sur le dernier site. Mais elles peuvent se recouvrir l'une l'autre sauf sur les 3 premiers sites (ce qui est généralisé par  $\frac{\ell+1}{2}$  si  $\ell$  est un nombre impair).

Nous supposons que l'exclusion entre particules étendues à l'entrée a un impact plus important sur la cinétique du système que l'exclusion sur le reste du réseau. Pour étudier cette hypothèse, nous considérons maintenant qu'il n'y a pas d'exclusion à part à l'entrée. Le système n'est pas un simple processus d'exclusion, nous ne l'appelons donc plus un TASEP. Un nom approprié peut être TAEBEP pour Totally Asymmetric Entrance Boundary Exclusion Process et plus précisément  $\ell$ -TAEBEP pour mentionner la particule étendue. L'inhomogénéité des taux de saut le long du réseau est également considérée.

Nous déduisons les équations de variation de l'occupation moyenne du  $\ell$ -TAEBEP de celles du  $\ell$ -TASEP inhomogène (Eq. (2.24) et (2.27)) en supprimant les termes d'exclusion sauf pour l'entrée, ce qui conduit à

$$\begin{aligned} \frac{d\langle n_1(t) \rangle}{dt} &= \alpha \left( 1 - \sum_{j=1}^{\ell} \langle n_j(t) \rangle \right) - p_1 \langle n_1(t) \rangle, \\ \frac{d\langle n_i(t) \rangle}{dt} &= p_{i-1} \langle n_{i-1}(t) \rangle - p_i \langle n_i(t) \rangle, \quad i = 2 \dots N. \end{aligned} \quad (2.38)$$

C'est un système linéaire en  $\langle n_i(t) \rangle = \rho_i(t)$ ,  $i = 1 \dots N$ , la densité de particule. Par conséquent, nous n'avons pas besoin de faire une approximation de champ moyen pour avoir des équations linéaires sur la densité des particules lorsque l'exclusion est négligée sauf à l'entrée.

Nous déduisons des Eq.(2.38) le courant et la densité à l'état stationnaire :

$$\begin{aligned} J &= \alpha \left( 1 - \sum_{j=1}^{\ell} \rho_j \right) = \frac{\alpha}{1 + \alpha \sum_{j=1}^{\ell} \frac{1}{p_j}}, \\ \rho_i &= \frac{J}{p_i}, \quad i = 1 \dots N. \end{aligned} \quad (2.39)$$

Nous montrons en Fig. 2.10 que l'expression sur la densité correspond effectivement exactement à la densité trouvée avec la simulation Monte Carlo en Gillespie du modèle  $\ell$ -TAEBEP.

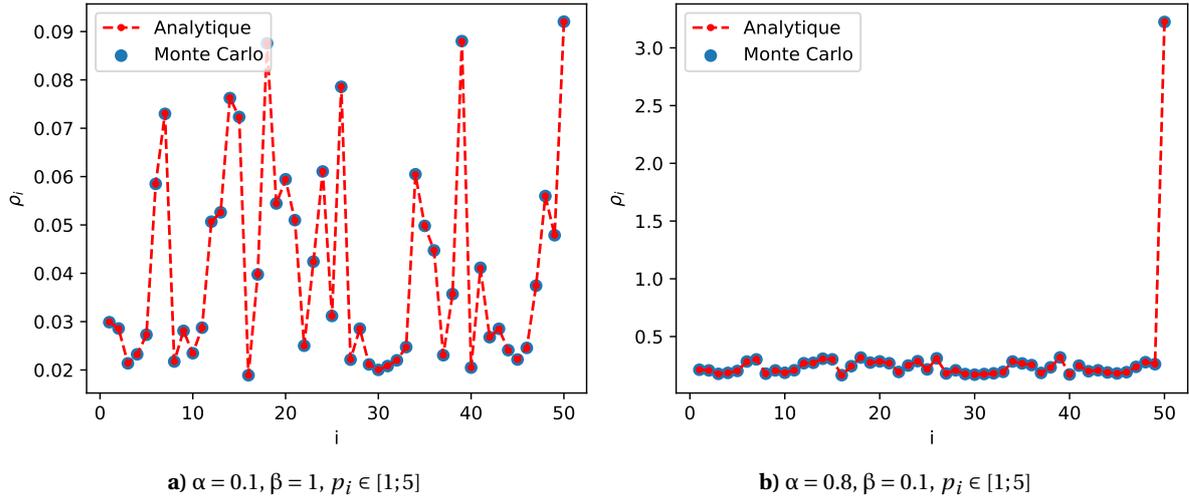


FIGURE 2.10 – Comparaison des profils de densités TAEBP obtenus en simulation Monte Carlo et en formulation analytique exacte. Les séquences de  $p_i$  sont obtenus par tirage aléatoire selon une distribution uniforme.

### 5.3 Comparaison des modèles $\ell$ -TASEP et $\ell$ -TAEBP homogènes pour étudier l'influence de la particule étendue

Nous supposons que, en phase basse densité, l'exclusion à l'entrée doit avoir un impact plus important sur la densité et le courant de particules que l'exclusion sur le reste du réseau. Ceci à cause du fait que les  $\ell$  premiers sites doivent être libres pour permettre à une nouvelle particule d'entrer, alors qu'un seul site devant la particule doit être libre pour que celle-ci puisse avancer. Cet effet doit être d'autant plus important que  $\ell$  est grand. Pour vérifier cela, nous avons tracé la densité moyennée sur le réseau pour trois modèles différents : le modèle  $\ell$ -TASEP, le modèle  $\ell$ -TAEBP et le modèle sans aucune exclusion, pour le cas homogène (Fig.2.11). Utiliser la densité moyennée sur le réseau permet de négliger les fluctuations locales. Nous limitons par conséquent cette étude aux cas où l'effet de bord peut être négligé, c'est-à-dire quand  $\frac{\alpha}{p} \leq 0.1$  d'après ce que nous avons démontré à la section précédente.

Nous voyons que dans le cas où  $\ell = 1$ , la densité  $\ell$ -TASEP est plus proche de la densité du modèle sans exclusion que de la densité  $\ell$ -TAEBP. L'exclusion dans le *bulk* doit donc avoir un effet opposé à l'exclusion à l'entrée, de manière à ce que les effets des deux exclusions s'annulent. L'exclusion à l'entrée entraîne une diminution de la densité moyenne, ce qui n'est pas étonnant. Mais, l'exclusion dans le *bulk* engendre une gêne dans le mouvement des particules, ce qui les fait rester plus longtemps sur le réseau. La densité moyenne augmente. Ceci explique aussi pourquoi la densité  $\ell$ -TASEP est toujours supérieure ou égale à la densité  $\ell$ -TAEBP, quel que soit  $\ell$ . Nous voyons ensuite, que dès que la particule est plus grande que son pas d'avancement, la densité  $\ell$ -TASEP est plus proche de la densité  $\ell$ -TAEBP que de celle du modèle sans exclusion. L'impact de l'exclusion à l'entrée sur la densité moyenne est donc plus important que celui de l'exclusion dans le *bulk*. De plus, nous voyons que la différence entre la densité  $\ell$ -TASEP et la densité  $\ell$ -TAEBP est toujours inférieure à 10%, elle peut donc être considérée comme négligeable. Nous concluons que le modèle 9-TAEBP est une bonne approximation du modèle 9-TASEP homogène.

Nous pouvons comparer les équations  $\ell$ -TASEP et  $\ell$ -TAEBP dans le *bulk* afin de voir plus précisément comment varie la différence entre les 2 densités. Dans le cas homogène, la densité du modèle  $\ell$ -TAEBP devient constante sur tout le réseau puisqu'elle ne varie qu'avec les  $p_i$  (Eq. (2.39)). On obtient

$$\rho = \frac{1}{\frac{p}{\alpha} + \ell}. \quad (2.40)$$

Mais, si on considère le taux de sortie  $\beta$  différent de  $p$ , la densité aura une valeur différente au

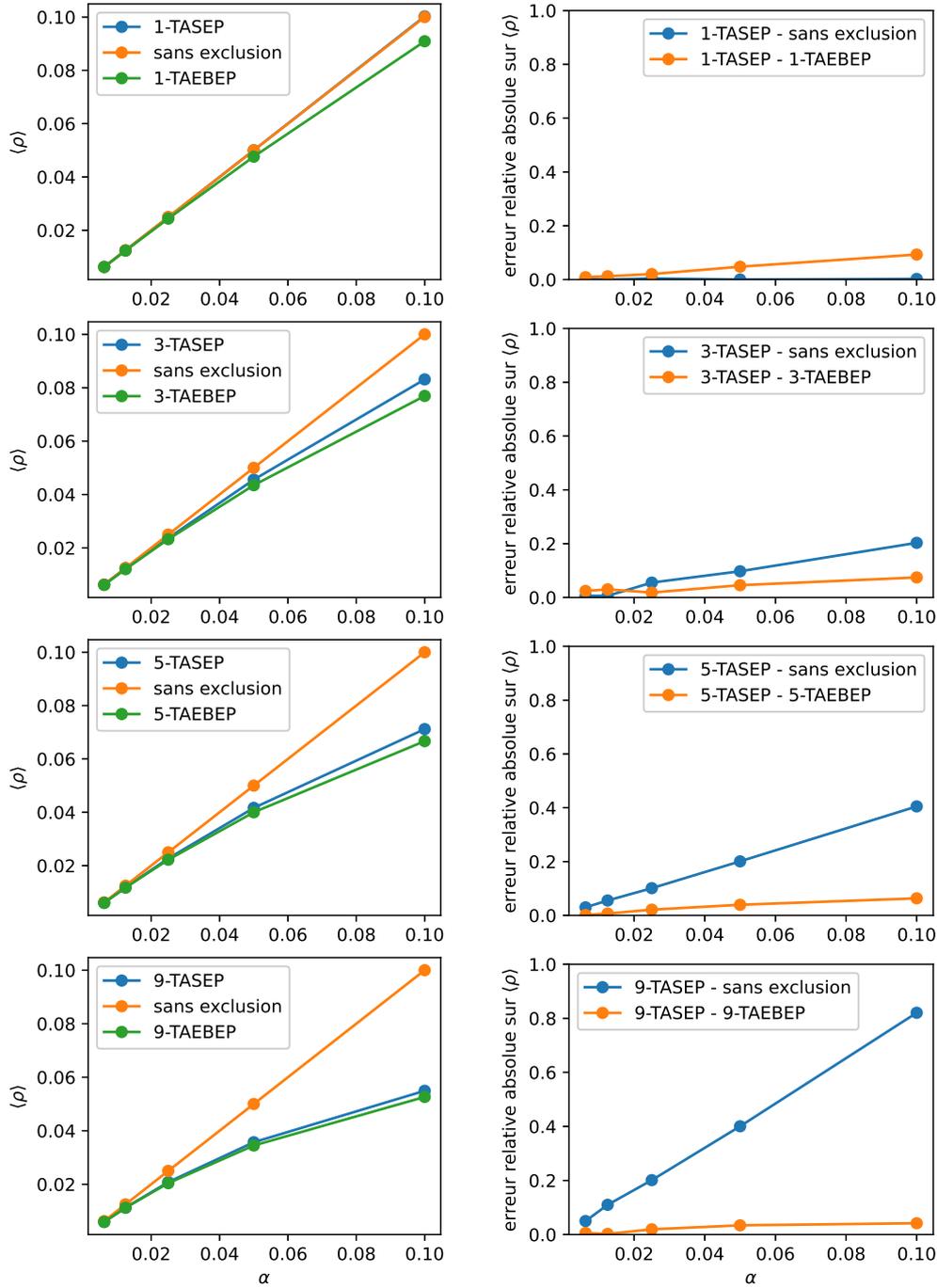


FIGURE 2.11 – Comparaison des densités moyennées sur le réseau pour différentes valeurs de  $\alpha$  et  $\ell$  avec  $p = 1$ ,  $L = 100$ . Temps de simulation =  $10^6$ ,  $dt = 50$ ,  $t_0 = 100$ . On voit que l'erreur du TAESEP sur le TASEP reste à moins de 10%.

dernier site. On obtient

$$\rho = \frac{p}{\beta} \frac{1}{\frac{p}{\alpha} + \ell}. \quad (2.41)$$

Pour les sites précédents la couche limite, l'erreur relative du  $\ell$ -TAESEP sur le  $\ell$ -TASEP en champ moyen (TASEP-CM) est donc facile à calculer :

$$\Delta\rho = \frac{|\rho_{\text{TASEP-CM}} - \rho_{\text{TAESEP}}|}{\rho_{\text{TASEP-CM}}} = \frac{1}{\frac{p}{\alpha} + \ell}. \quad (2.42)$$

On fera donc une erreur de plus de 10% sur ces sites quand  $\frac{p}{\alpha} + \ell < 10$ . Ce qui est cohérent avec la Fig. 2.11 puisqu'on a au minimum  $\frac{p}{\alpha} + \ell = 11$  et que l'erreur est toujours inférieure à 10%.

Pour savoir si c'est une bonne approximation du  $\ell$ -TASEP pour modéliser le transport des ribosomes sur l'ARN messenger, il faudra aussi étudier l'effet de l'inhomogénéité des taux de saut. Ce résultat est cependant cohérent avec les mesures faites en biologie qui montrent qu'une stratégie de "rampe d'accélération" où la vitesse du ribosome augmente sur les premiers codons des ARNm semble exister [MCKINNON et al., 2021; TULLER et al., 2010; VERMA et al., 2019] afin d'éviter les collisions entre ribosomes à l'initiation et de favoriser le recrutement de nouveaux ribosomes [SZAVITS-NOSSAN et CIANDRINI, 2020]. Cette stratégie mise en place seulement à l'entrée de la séquence suggère que l'exclusion entre les ribosomes à l'entrée a un effet important sur la cinétique de traduction.

## 6 Simulations Monte Carlo

### 6.1 Algorithme de Gillespie

Les simulations des différents modèles stochastiques ont été programmées à l'aide de l'algorithme de Gillespie. Cet algorithme a été inventé par GILLESPIE [1976] pour simuler de manière exacte l'évolution du nombre de molécules identiques d'un ensemble de molécules pouvant réagir. On note  $\vec{X}(t)$  le vecteur qui contient le nombre  $X_i(t)$  de molécules  $i$  au temps  $t$  sachant que le système était à l'état  $\vec{X}(t_0)=\vec{x}_0$  à l'instant initial  $t_0$ . L'évolution du système stochastique à chaque instant dépend de deux variables aléatoires. Le temps d'attente jusqu'à la prochaine réaction  $\tau$ , et la nature de la réaction  $j$ . On s'intéresse donc à la fonction [GILLESPIE, 2007]

$$p(\tau, j|\vec{x}, t) \equiv \text{probabilité que la prochaine réaction ait lieu dans l'intervalle de temps infinitésimal } [t + \tau, t + \tau + d\tau[, \text{ et soit une réaction } R_j, \quad (2.43)$$

$$\text{sachant que } \vec{X}(t) = \vec{x}.$$

Étant donné les taux de réaction  $a_j(\vec{x})$  définis par

$$a_j(\vec{x})dt \equiv \text{probabilité qu'une réaction } R_j \quad (2.44)$$

$$\text{ait lieu dans l'intervalle de temps } [t, t + dt[, \text{ sachant que } \vec{X}(t) = \vec{x},$$

on peut montrer que [GILLESPIE, 1977]

$$p(\tau, j|\vec{x}, t) = a_j(\vec{x})e^{-a_0(\vec{x})\tau} \quad (2.45)$$

où

$$a_0(\vec{x}) \equiv \sum_{k=1}^M a_k(\vec{x}), \quad (2.46)$$

et  $M$  le nombre de réactions possibles. Par inversion de la probabilité (2.45), on montre qu'en tirant deux nombres  $(r_1, r_2)$  d'une distribution uniforme sur l'intervalle unité, on obtient des valeurs pour les deux variables aléatoires par GILLESPIE [1976]

$$\tau(r_1) = \frac{1}{a_0(\vec{x})} \ln\left(\frac{1}{r_1}\right), \quad (2.47)$$

$$j = \text{le plus petit entier satisfaisant } \sum_{k=1}^j a_k(\vec{x}) > r_2 a_0(\vec{x}).$$

L'algorithme permettant de calculer la trajectoire du système en termes de réaction et de population est donc

1. Initialisation du système  $(\vec{x}_0, t_0)$ .
2. Calculer la somme  $a_0(\vec{x})$ .
3. Tirer  $(r_1, r_2)$  sur  $]0; 1]$  uniforme pour générer les valeurs de  $\tau(r_1)$  et  $j$  avec les équations (2.47).
4. Effectuer la réaction sélectionnée.
5. Enregistrer  $(\vec{x}, t)$  et recommencer à l'étape 2 ou arrêter la simulation.

## 6.2 Simulation du TASEP à l'état stationnaire

Dans le cas du TASEP et des autres modèles de transport, on veut connaître la densité de particule en chaque site. Ce qu'on appelle "densité" ici est la moyenne d'occupation sur les configurations enregistrées, qu'on peut écrire

$$\rho_i = \frac{1}{N_E} \sum_{e=1}^{N_E} n_i^{(e)} \quad (2.48)$$

où  $N_E$  est le nombre d'enregistrements et  $n_i^{(e)}$ , égal à 0 ou 1, est l'occupation du site  $i$  à l'enregistrement  $e$ . En vertu du principe d'ergodicité, et s'intéressant au système à l'état stationnaire, on fait évoluer un seul système jusqu'à l'état stationnaire, puis on effectue un enregistrement à chaque pas de temps  $\Delta t$ . On gagne ainsi un temps de simulation considérable puisqu'on effectue une seule simulation pour obtenir une moyenne statistique sur un grand nombre de réseaux. On peut donc écrire l'Eq. (2.48) comme

$$\rho_i = \frac{1}{N_E} \sum_{e=0}^{N_E-1} n_i(t_m + e\Delta t) \quad (2.49)$$

où  $n_i$  est fonction de  $e$ ,  $\Delta t$ , et  $t_m$ .  $t_m$  est le temps, supérieur au temps nécessaire pour atteindre l'état stationnaire, auquel on commence les mesures. Le pas de temps doit être assez élevé pour que deux configurations successives soient considérées comme décorréelées. Pour le respecter, il faut modifier légèrement l'algorithme de Gillespie. En effet, puisque le temps du système saute de  $\tau(r_1)$  à chaque itération, il faut savoir si le pas de temps  $\Delta t$  va être dépassé au prochain changement d'état pour savoir s'il faut enregistrer la configuration actuelle. L'algorithme devient donc

1. Initialisation du système  $(\vec{x}_0, t_0)$ .
2. Calculer la somme  $a_0(\vec{x})$ .
3. Tirer  $(r_1, r_2)$  sur  $]0; 1]$  uniforme pour générer les valeurs de  $\tau(r_1)$  et  $j$  avec les équations (2.47).
4. Calculer le temps  $t$ .
5. Si  $t > t_m$ , enregistrer la configuration  $k$  fois avec  $k$  le plus grand entier tel que  $\tau(r_1) \geq k\Delta t$ .
6. Effectuer l'événement sélectionné et recommencer à l'étape 2 ou arrêter la simulation.

Effectuer les enregistrements en respectant le pas de temps fixé  $\Delta t$  est particulièrement important quand les taux cinétiques sont faibles, puisque plus les taux sont faibles, plus le temps réel s'éloigne du "temps Gillespie" qui correspond à un changement d'état à chaque itération ou pas de "temps Gillespie".

## 7 Conclusion et perspectives

Dans ce chapitre, nous avons décrit le modèle paradigmatique des transports unidirectionnels à une dimension avec bords ouverts afin de discuter son extension au modèle  $\ell$ -TASEP inhomogène qu'on peut considérer comme le modèle paradigmatique de la traduction génétique. Nous avons décrit la résolution exacte du TASEP qui a été très utile pour caractériser de manière complète le système, mais dont le temps de calcul de la matrice de Markov augmente exponentiellement avec la taille du système, ce qui empêche le calcul pour des réseaux correspondant à la plupart des ARNm qui ont une taille typique de l'ordre de quelques centaines de codons. Nous avons ensuite montré que la résolution en champ moyen, considérablement plus simple et rapide que la résolution exacte, donne de très bonnes prédictions de la densité et du courant de particules. De plus, cette approche est aussi effective pour le  $\ell$ -TASEP inhomogène contrairement à l'approche Markovienne exacte. En conclusion de l'étude de ces deux modèles, il est intéressant de remarquer que la cinétique du TASEP est contrôlée par les bords du réseau avec une symétrie trou particule, mais que la particule étendue va briser la symétrie et abaisser la limite entre la basse et la haute densité et le courant maximum.

Nous avons montré que l'effet de bords en  $\ell$ -TASEP basse densité est beaucoup plus important que pour le modèle TASEP simple, et que son amplitude dépend seulement du ratio entre le taux d'initiation et la taux de saut  $\frac{\alpha}{p}$ . Il serait intéressant de continuer l'étude de cet effet par la caractérisation de sa couche limite, qui semble aussi peu dépendante du taux de terminaison  $\beta$  et de la taille de la particule  $\ell$  d'après nos simulations, contrairement à la variation de son amplitude le long du réseau.

Grâce à l'introduction d'une nouvelle variante du modèle dans laquelle l'exclusion entre particules existe seulement à l'entrée du réseau, le  $\ell$ -TAESEP, et à l'étude de l'effet de bords en  $\ell$ -TASEP basse densité, nous avons montré l'importance de l'exclusion à l'entrée dans le cas homogène. Ici, le modèle  $\ell$ -TASEP homogène est bien approximé par le modèle  $\ell$ -TAESEP quand  $\frac{p}{\alpha} + \ell > 10$  et  $\alpha \leq 0.1$ . La limite  $\alpha \leq 0.1$  peut cependant être négligée pour les très grands réseaux où les effets deviennent négligeables pour des réseaux de plusieurs milliers de codons. On comprend alors qu'une stratégie de "rampe d'accélération" où la vitesse du ribosome augmente sur les premiers codons des ARNm ait été développée par les cellules afin d'éviter les collisions entre ribosomes à l'initiation et de favoriser le recrutement de nouveaux ribosomes. Un grand nombre de ribosomes sur l'ARNm pourrait néanmoins générer de l'encombrement ce qui engendrerait des collisions entre ribosomes sur les codons de la séquence où la vitesse diminue. L'étape suivante sera d'étudier la variation de l'exclusion à l'entrée lorsque les taux de saut sont inhomogènes avec la considération d'un modèle  $\ell$ -TAESEP inhomogène. En conclusion, nous pouvons dire que nous avons introduit une variante simplifiée du  $\ell$ -TASEP qui a permis de contribuer à la caractérisation de l'effet de l'exclusion d'une particule étendue sur la densité de particules.

## 8 Références

- CHOU, T. 2003, «Ribosome Recycling, Diffusion, and mRNA Loop Formation in Translational Regulation», *Biophysical Journal*, vol. 85, n° 2, doi :10.1016/S0006-3495(03)74518-4, p. 755–773, ISSN 0006-3495. URL <https://www.sciencedirect.com/science/article/pii/S0006349503745184>. 23, 24
- CHOWDHURY, D., L. SANTEN et A. SCHADSCHNEIDER. 2000, «Statistical physics of vehicular traffic and some related systems», *Physics Reports*, vol. 329, n° 4, doi :10.1016/S0370-1573(99)00117-9, p. 199–329, ISSN 0370-1573. URL <https://www.sciencedirect.com/science/article/pii/S0370157399001179>. 23, 24
- DAHM, R. 2008, «Discovering DNA : Friedrich Miescher and the early years of nucleic acid research», *Human Genetics*, vol. 122, n° 6, doi :10.1007/s00439-007-0433-0, p. 565–581, ISSN 1432-1203. URL <https://doi.org/10.1007/s00439-007-0433-0>. 22
- DAULOUDET, O. 2015, «Étude théorique des phénomènes de transport intracellulaire hors-équilibre thermodynamique : rôle du couplage entre transport actif et diffusif en volume confiné», . 28
- DERRIDA, B., E. DOMANY et D. MUKAMEL. 1992, «An exact solution of a one-dimensional asymmetric exclusion model with open boundaries», *Journal of Statistical Physics*, vol. 69, n° 3, doi :10.1007/BF01050430, p. 667–687, ISSN 1572-9613. URL <https://doi.org/10.1007/BF01050430>. 23, 27
- DERRIDA, B., M. R. EVANS, V. HAKIM et V. PASQUIER. 1993, «Exact solution of a 1D asymmetric exclusion model using a matrix formulation», *Journal of Physics A : Mathematical and General*, vol. 26, n° 7, doi :10.1088/0305-4470/26/7/011, p. 1493–1517, ISSN 0305-4470, 1361-6447. URL <http://stacks.iop.org/0305-4470/26/i=7/a=011?key=crossref.3f450f9745d574cb00579a3790aaf843>. 27

- ERDMANN-PHAM, D. D., K. DAO DUC et Y. S. SONG. 2020, «The Key Parameters that Govern Translation Efficiency», *Cell Systems*, vol. 10, n° 2, doi :10.1016/j.cels.2019.12.003, p. 183–192.e6, ISSN 2405-4712. URL <http://www.sciencedirect.com/science/article/pii/S2405471219304648>. 32
- GERST, I. et S. N. LEVINE. 1965, «Kinetics of protein synthesis by polyribosomes», *Journal of Theoretical Biology*, vol. 9, n° 1, doi :10.1016/0022-5193(65)90054-8, p. 16–36, ISSN 0022-5193. URL <https://www.sciencedirect.com/science/article/pii/0022519365900548>. 23
- GILLESPIE, D. T. 1976, «A general method for numerically simulating the stochastic time evolution of coupled chemical reactions», *Journal of Computational Physics*, vol. 22, n° 4, doi :10.1016/0021-9991(76)90041-3, p. 403–434, ISSN 0021-9991. URL <http://www.sciencedirect.com/science/article/pii/0021999176900413>. 41
- GILLESPIE, D. T. 1977, «Exact stochastic simulation of coupled chemical reactions», *The Journal of Physical Chemistry*, vol. 81, n° 25, doi :10.1021/j100540a008, p. 2340–2361, ISSN 0022-3654. URL <https://doi.org/10.1021/j100540a008>. 41
- GILLESPIE, D. T. 2007, «Stochastic Simulation of Chemical Kinetics», *Annual Review of Physical Chemistry*, vol. 58, n° 1, doi :10.1146/annurev.physchem.58.032806.104637, p. 35–55. URL <https://doi.org/10.1146/annurev.physchem.58.032806.104637>, \_eprint : <https://doi.org/10.1146/annurev.physchem.58.032806.104637>. 41
- HELBING, D. 2001, «Traffic and related self-driven many-particle systems», *Reviews of Modern Physics*, vol. 73, n° 4, doi :10.1103/RevModPhys.73.1067, p. 1067–1141. URL <https://link.aps.org/doi/10.1103/RevModPhys.73.1067>, publisher : American Physical Society. 23, 24
- JACOB, F. et J. MONOD. 1961, «Genetic regulatory mechanisms in the synthesis of proteins», *Journal of Molecular Biology*, vol. 3, n° 3, doi :10.1016/S0022-2836(61)80072-7, p. 318–356, ISSN 0022-2836. URL <https://www.sciencedirect.com/science/article/pii/S0022283661800727>. 22
- JOHN, A., A. SCHADSCHNEIDER, D. CHOWDHURY et K. NISHINARI. 2009, «Trafficlike Collective Movement of Ants on Trails : Absence of a Jammed Phase», *Physical Review Letters*, vol. 102, n° 10, doi :10.1103/PhysRevLett.102.108001, p. 108001. URL <https://link.aps.org/doi/10.1103/PhysRevLett.102.108001>, publisher : American Physical Society. 23
- KLEIN, G. A., K. KRUSE, G. CUNIBERTI et F. JÜLICHER. 2005, «Filament Depolymerization by Motor Molecules», *Physical Review Letters*, vol. 94, n° 10, doi :10.1103/PhysRevLett.94.108102, p. 108102. URL <https://link.aps.org/doi/10.1103/PhysRevLett.94.108102>, publisher : American Physical Society. 23, 24
- KLUMPP, S. et R. LIPOWSKY. 2003, «Traffic of Molecular Motors Through Tube-Like Compartments», *Journal of Statistical Physics*, vol. 113, n° 1, doi :10.1023/A:1025778922620, p. 233–268, ISSN 1572-9613. URL <https://doi.org/10.1023/A:1025778922620>. 23, 24
- LAKATOS, G. et T. CHOU. 2003, «Totally asymmetric exclusion processes with particles of arbitrary size», *Journal of Physics A : Mathematical and General*, vol. 36, n° 8, doi :10.1088/0305-4470/36/8/302, p. 2027–2041, ISSN 0305-4470. URL <https://doi.org/10.1088/0305-4470/36/8/302>, publisher : IOP Publishing. 34, 36
- LIU, M.-Z., S.-D. LI et W. RUI-LI. 2012, «Asymmetric simple exclusion processes with complex lattice geometries : A review of models and phenomena», vol. 21, n° 9, doi :10.1088/1674-1056/21/9/090510, p. 090510, ISSN 1674-1056. URL <https://doi.org/10.1088/1674-1056/21/9/090510>, publisher : IOP Publishing. 23

- MACDONALD, C. T. et J. H. GIBBS. 1969, «Concerning the kinetics of polypeptide synthesis on polyribosomes», *Biopolymers*, vol. 7, n° 5, doi :10.1002/bip.1969.360070508, p. 707–725, ISSN 1097-0282. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/bip.1969.360070508>, \_eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1002/bip.1969.360070508>. 31
- MACDONALD, C. T., J. H. GIBBS et A. C. PIPKIN. 1968, «Kinetics of biopolymerization on nucleic acid templates», *Biopolymers*, vol. 6, n° 1, doi :10.1002/bip.1968.360060102, p. 1–25, ISSN 0006-3525, 1097-0282. URL <http://doi.wiley.com/10.1002/bip.1968.360060102>. 23, 33
- MCKINNON, L. M., J. B. MILLER, M. F. WHITING, J. S. K. KAUWE et P. G. RIDGE. 2021, «A comprehensive analysis of the phylogenetic signal in ramp sequences in 211 vertebrates», *Scientific Reports*, vol. 11, n° 1, doi :10.1038/s41598-020-78803-3, p. 622, ISSN 2045-2322. URL <http://www.nature.com/articles/s41598-020-78803-3>, number : 1 Publisher : Nature Publishing Group. 41
- NAGAR, A., A. VALLERIANI et R. LIPOWSKY. 2011, «Translation by Ribosomes with mRNA Degradation : Exclusion Processes on Aging Tracks», *Journal of Statistical Physics*, vol. 145, n° 5, doi :10.1007/s10955-011-0347-z, p. 1385–1404, ISSN 0022-4715. WOS :000297593900013. 23
- PALADE, G. E. 2007, «TRIBUTE TO», *Journal of Cellular and Molecular Medicine*, vol. 11, n° 1, doi :10.1111/j.1582-4934.2007.00018.x, p. 2–3, ISSN 1582-1838, 1582-4934. URL <https://onlinelibrary.wiley.com/doi/10.1111/j.1582-4934.2007.00018.x>. 22
- PARMEGGIANI, A., T. FRANOSCH et E. FREY. 2004, «Totally asymmetric simple exclusion process with Langmuir kinetics», *Physical Review E*, vol. 70, n° 4, doi :10.1103/PhysRevE.70.046101, p. 046101. URL <https://link.aps.org/doi/10.1103/PhysRevE.70.046101>, publisher : American Physical Society. 23, 24, 28
- RAGOUCY, E. 2017, «Integrability in out-of-equilibrium systems», *Journal of Physics : Conference Series*, vol. 804, doi :10.1088/1742-6596/804/1/012037, p. 012037, ISSN 1742-6588, 1742-6596. URL <http://arxiv.org/abs/1609.08560>, arXiv : 1609.08560. 26
- SCHÜTZ, G. et E. DOMANY. 1993, «Phase transitions in an exactly soluble one-dimensional exclusion process», *Journal of Statistical Physics*, vol. 72, n° 1, doi :10.1007/BF01048050, p. 277–296, ISSN 1572-9613. URL <https://doi.org/10.1007/BF01048050>. 23
- SCHÜTZ, G. M. 1997, «The Heisenberg Chain as a Dynamical Model for Protein Synthesis - Some Theoretical and Experimental Results», *International Journal of Modern Physics B*, vol. 11, n° 01n02, doi :10.1142/S0217979297000265, p. 197–202, ISSN 0217-9792. URL <https://www.worldscientific.com/doi/abs/10.1142/S0217979297000265>, publisher : World Scientific Publishing Co. 23, 24
- SCHÜTZ, G. M. 2001, «1 - Exactly Solvable Models for Many-Body Systems Far from Equilibrium», dans *Phase Transitions and Critical Phenomena*, vol. 19, édité par C. Domb et J. L. Lebowitz, Academic Press, p. 1–251, doi :10.1016/S1062-7901(01)80015-X. URL <https://www.sciencedirect.com/science/article/pii/S106279010180015X>. 23, 24
- SHAW, L. B., J. P. SETHNA et K. H. LEE. 2004, «Mean field approaches to the totally asymmetric exclusion process with quenched disorder and large particles», *Physical Review E*, vol. 70, n° 2, doi :10.1103/PhysRevE.70.021901, p. 021901, ISSN 1539-3755, 1550-2376. URL <http://arxiv.org/abs/cond-mat/0403523>, arXiv : cond-mat/0403523. 34
- SHAW, L. B., R. K. P. ZIA et K. H. LEE. 2003, «Totally asymmetric exclusion process with extended objects : A model for protein synthesis», *Physical Review E*, vol. 68, n° 2, doi :10.1103/PhysRevE.68.021910, p. 021910. URL <https://link.aps.org/doi/10.1103/PhysRevE.68.021910>, publisher : American Physical Society. 23, 24

- SIMHA, R., J. M. ZIMMERMAN et J. MOACANIN. 1963, «Polymerization Kinetics of Biological Macromolecules on Templates», *The Journal of Chemical Physics*, vol. 39, n° 5, doi :10.1063/1.1734422, p. 1239–1246, ISSN 0021-9606, 1089-7690. URL <http://aip.scitation.org/doi/10.1063/1.1734422>. 22
- SINGH, U. N. 1969, «Polyribosomes and unstable messenger RNA : A stochastic model of protein synthesis», *Journal of Theoretical Biology*, vol. 25, n° 3, doi :10.1016/S0022-5193(69)80032-9, p. 444–460, ISSN 0022-5193. URL <https://www.sciencedirect.com/science/article/pii/S0022519369800329>. 23
- SUGDEN, K. E. P., M. R. EVANS, W. C. K. POON et N. D. READ. 2007, «Model of hyphal tip growth involving microtubule-based transport», *Physical Review E*, vol. 75, n° 3, doi :10.1103/PhysRevE.75.031909, p. 031 909. URL <https://link.aps.org/doi/10.1103/PhysRevE.75.031909>, publisher : American Physical Society. 23
- SZAVITS-NOSSAN, J. et L. CIANDRINI. 2020, «Inferring efficiency of translation initiation and elongation from ribosome profiling», *Nucleic Acids Research*, doi :10.1093/nar/gkaa678, ISSN 0305-1048, 1362-4962. URL <https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkaa678/5895331>. 41
- TULLER, T., A. CARMI, K. VESTSIGIAN, S. NAVON, Y. DORFAN, J. ZABORSKE, T. PAN, O. DAHAN, I. FURMAN et Y. PILPEL. 2010, «An Evolutionarily Conserved Mechanism for Controlling the Efficiency of Protein Translation», *Cell*, vol. 141, n° 2, doi :10.1016/j.cell.2010.03.031, p. 344–354, ISSN 0092-8674. URL <http://www.sciencedirect.com/science/article/pii/S0092867410003193>. 41
- VERMA, M., J. CHOI, K. A. COTTRELL, Z. LAVAGNINO, E. N. THOMAS, S. PAVLOVIC-DJURANOVIC, P. SZCZESNY, D. W. PISTON, H. S. ZAHER, J. D. PUGLISI et S. DJURANOVIC. 2019, «A short translational ramp determines the efficiency of protein synthesis», *Nature Communications*, vol. 10, n° 1, doi :10.1038/s41467-019-13810-1, p. 1–15, ISSN 2041-1723. URL <https://www.nature.com/articles/s41467-019-13810-1>. 41
- WIDOM, B., J. L. VIOVY et A. D. DEFONTAINES. 1991, «Repton model of gel electrophoresis and diffusion», *Journal de Physique I*, vol. 1, n° 12, doi :10.1051/jp1:1991239, p. 1759–1784, ISSN 1155-4304, 1286-4862. URL <http://dx.doi.org/10.1051/jp1:1991239>, publisher : EDP Sciences. 23, 24

## Chapitre 3

# Modélisation d'un ensemble d'ARN messagers à temps de vie finis

### Sommaire

---

<b>1</b>	<b>Introduction</b> . . . . .	<b>48</b>
<b>2</b>	<b>Séparation par nombre de ribosomes</b> . . . . .	<b>48</b>
<b>3</b>	<b>Distribution des âges des ARNm</b> . . . . .	<b>49</b>
<b>4</b>	<b>Traduction modélisée par TASEP inhomogène</b> . . . . .	<b>51</b>
4.1	Profil de densité polysome . . . . .	51
4.2	Profil de densité $k$ -some . . . . .	51
4.3	Exemple d'un ARNm à 2 codons . . . . .	52
4.3.1	Solution du système (3.7) . . . . .	52
4.3.2	Profil de densité polysome . . . . .	53
4.3.3	Profils de densités $k$ -somes . . . . .	53
<b>5</b>	<b>Modèle de transport balistique</b> . . . . .	<b>54</b>
5.1	Le modèle . . . . .	54
5.2	Probabilité qu'un ARNm né à $t_b$ soit un $k$ -some à $t_{ob}$ . . . . .	55
5.3	Probabilité qu'un ARNm de la soupe soit un $k$ -some . . . . .	56
5.4	Statistique des $k$ -somes . . . . .	56
5.5	Densité $k$ -some en fonction de son âge . . . . .	57
5.6	Profil de densité $k$ -some . . . . .	58
5.7	Profil de densité polysome . . . . .	58
<b>6</b>	<b>Analyse de l'effet de durée de vie finie</b> . . . . .	<b>59</b>
6.1	Quelques taux de la traduction génétique . . . . .	59
6.2	Densité polysome et nombre moyen de ribosomes sur l'ARN messenger, $\langle k \rangle$ . . . . .	59
6.3	Erreur sur les estimations des taux de saut avec les modèles à temps de vie infini . . . . .	61
6.4	Distribution des $k$ -somes . . . . .	63
6.5	Profils de densités $k$ -somes et polysome à vitesse constante . . . . .	65
6.5.1	Analyse générale . . . . .	65
6.5.2	Point de croisement entre $\rho_k(x)$ et $\rho_k^\infty$ . . . . .	66
6.6	Rapport entre les densités des ARNm à l'état transitoire et des ARNm à l'état stationnaire . . . . .	67
<b>7</b>	<b>Diagramme de phase</b> . . . . .	<b>69</b>
<b>8</b>	<b>Conclusion et perspectives</b> . . . . .	<b>70</b>
<b>9</b>	<b>Références</b> . . . . .	<b>71</b>

---

## 1 Introduction

Depuis environ une décennie, la technique de séquençage haut débit, Ribo-Seq (ou *ribosome profiling*), a révolutionné l'étude de la traduction génétique. Cette technique, décrite au premier chapitre de cette thèse, permet de compter les ribosomes présents sur chaque codon des ARN messagers à un instant donné. Nous appelons "système de traduction" l'ARNm et les ribosomes le traduisant, modélisés le plus souvent par le système  $\ell$ -TASEP inhomogène que nous avons décrit dans le chapitre précédent. Pour ce modèle est étudié la densité seulement à l'état stationnaire et pourrait donc ne pas correspondre aux profils Ribo-Seq. Dans ce chapitre, nous allons adapter la modélisation de la traduction aux systèmes mesurés par cette expérience. Nous pensons que considérer les temps de vie finis des ARNm conduit à un système ayant des propriétés intéressantes encore très peu étudiées. En particulier, la mesure s'effectuant à un instant donné sur de nombreux ARNm qui ne sont pas traduits de manière synchronisée, il est important de considérer que chaque système de traduction n'est pas à l'état stationnaire à un instant donné. Les ARNm fraîchement nés seront en phase de remplissage de ribosome que nous appellerons régime transitoire. Pour cela, nous allons considérer la durée de vie des ARNm. La non-synchronisation des traductions sera assurée par des naissances réparties dans le temps selon une loi exponentielle.

Dans ce chapitre, nous allons modéliser les résultats de l'expérience Ribo-Seq effectuée sur deux populations différentes. La première est la population habituelle, c'est-à-dire tous les ARN messagers de millions de cellules. Nous l'appellerons la population des "polysomes". Dans la seconde, les ARNm sont séparés suivant le nombre  $k$  de ribosomes qui sont en train de les traduire. Nous l'appellerons la population des  $k$ -somes. Nous allons montrer dans une première approche qu'on peut simplement intégrer la description Markovienne sur un ensemble d'ARNm à durée de vie finie. Puis, nous développerons un modèle dit balistique en raison de la non stochasticité du mouvement du ribosome et de sa modélisation par une particule ponctuelle. Cette simplification nous permettra de caractériser l'effet de temps de vie fini sur les profils de densité  $k$ -some et polysome ainsi que sur la distribution des différentes populations  $k$ -some. Nous allons définir différents régimes de dégradation et situer celui du vivant en discutant des valeurs biologiques présentes dans la littérature pour les paramètres d'intérêts. Enfin, nous synthétiserons nos résultats dans un *diagramme de phase*.

Une comparaison de ce modèle avec les données Ribo-Seq obtenues par nos collaborateurs biologistes et bioinformaticiens du consortium GEM à Montpellier pour les polysomes et les  $k$ -somes sera faite au chapitre suivant et motivera la proposition d'une nouvelle stratégie de mesure des paramètres cinétiques de la traduction grâce à ces expériences.

## 2 Séparation par nombre de ribosomes

Des discussions avec les biologistes de l'IGF et les bioinformaticiens du LIRMM, une nouvelle expérience sur les ARNm a émergée. Lors de ces discussions, nous avons souligné le manque de données par rapport au nombre d'inconnus. Typiquement, on veut déterminer la vitesse sur chaque codon et pour cela, nous écrivons la densité de ribosome en chaque codon. Mais nous n'avons pas réellement accès à la densité. Le nombre d'ARNm mesurés ne peut pas être extrait et nous ne pouvons donc pas normaliser les profils de *footprint* de ribosomes. Une normalisation arbitraire peut-être faite permettant ainsi de mesurer des rapports de vitesses du ribosome sur chaque codon d'un même gène. Mais, comme nous le démontrerons, les vitesses propres à chaque codon ne pourront pas être déterminées.

De leur côté, les biologistes avaient le désir d'observer les vitesses de ribosomes selon le nombre de ribosomes qui traduisent simultanément un ARNm. Ils ont donc eu l'idée de séparer les polysomes avant la mesure Ribo-Seq. Pour cela, ils plongent les polysomes dans un gradient de sucrose dans lequel les polysomes vont se ranger par ordre croissant (Fig. 3.1), puis ils extraient les fractions de monosome, disome, trisome et tétrasome à l'aide d'une mesure d'absorbance (Fig.3.2). On peut ensuite faire un séquençage Ribo-Seq sur chaque fraction.

Nous avons pensé qu'il serait très intéressant d'étudier ces systèmes, qui n'ont pas été modélisés ni étudiés d'un point de vue biophysique. C'est en effet l'objet de ce chapitre et du suivant, et plusieurs perspectives de travaux de recherches pourraient en découler. Nous allons montrer notamment, et pour la première fois, que les densités de ribosomes pour ces populations sont particulièrement sensibles à la durée de vie des ARNm.

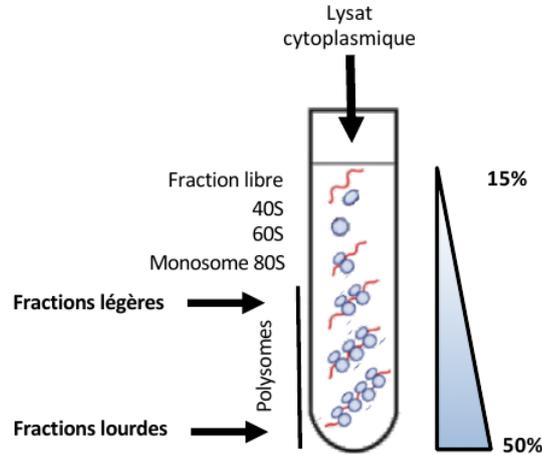


FIGURE 3.1 – Schéma de la séparation des polysomes et monosomes par gradient de sucrose.

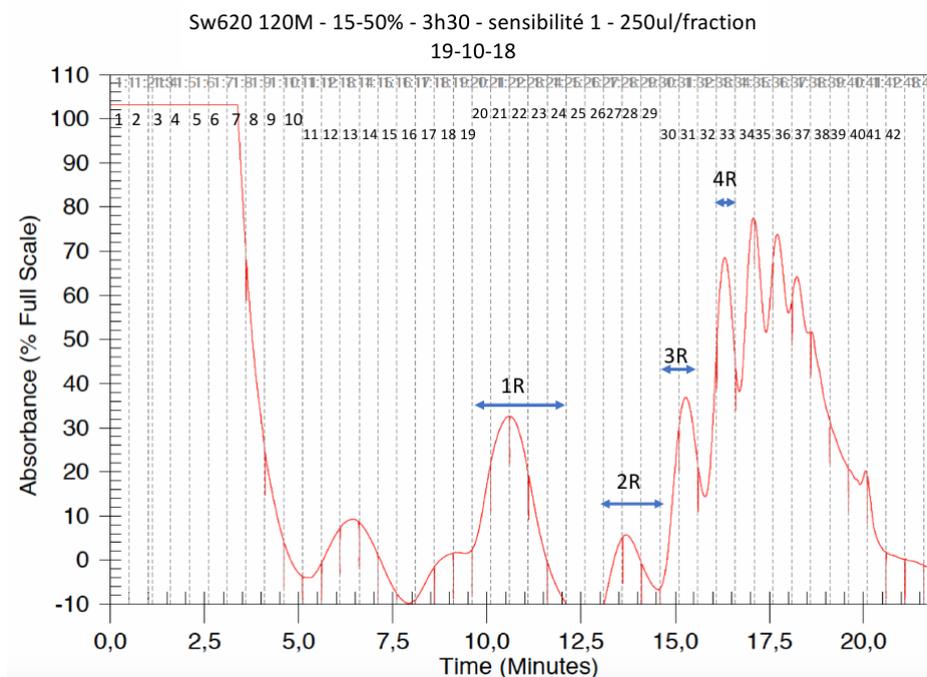


FIGURE 3.2 – Exemple de mesure d'absorbance du gradient de sucrose contenant les polysomes et monosomes. Les intervalles bleus indiquent les fractions de monosomes, disomes, trisomes et tétrasomes qui vont être extraites afin d'être séquencées par Ribo-Seq.

### 3 Distribution des âges des ARNm

Nous appelons "soupe" l'ensemble des brins d'ARNm présents dans l'échantillon au moment  $t_{ob}$  où ils sont immobilisés pour être observés par la méthode Ribo-Seq. Nous considérons tous les brins d'ARNm d'un même gène de longueur  $L$ . En raison des processus de dégradation, la durée de vie de l'ARNm  $\theta$  est une variable aléatoire régie par une fonction de densité de probabilité

(PDF),  $\Phi_\theta(\theta)$ . En supposant en outre que le volume des ARNm est plus ou moins constant dans le temps (stationnarité du processus de "renouvellement" des ARNm dans le cytosol), on peut montrer [COX, 1970] que la distribution statistique des âges des ARNm au temps  $t_{ob}$  a une PDF,  $\Phi_A(a)$ , donnée par

$$\Phi_A(a) = \frac{1}{\langle \theta \rangle} \int_a^\infty \Phi_\theta(\theta) d\theta \quad (3.1)$$

où le temps de vie moyen d'un ARNm  $\langle \theta \rangle$  est défini par

$$\langle \theta \rangle = \int_0^\infty \theta \Phi_\theta(\theta) d\theta. \quad (3.2)$$

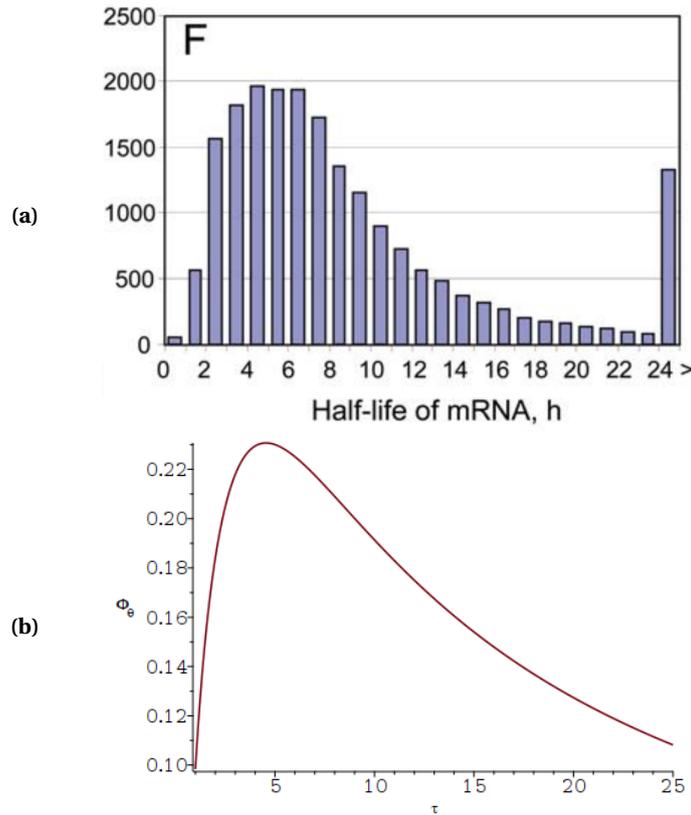
La distribution des temps de vie des ARNm à la forme

$$\Phi_\theta(\theta) = \omega e^{-\omega\theta}, \quad \theta \geq 0, \quad (3.3)$$

avec un temps de vie moyen  $\langle \theta \rangle = \omega^{-1}$ . La distribution de l'âge de l'ARNm est donc également exponentielle et peut s'écrire comme

$$\Phi_A(a) = \omega e^{-\omega a}, \quad a \geq 0. \quad (3.4)$$

Nous montrons sur la Fig. 3.3 qu'une telle distribution est cohérente avec les temps de demi-vies des ARNm mesurés chez la souris par SHAROVA et al. [2009]. Nous considérerons dans notre modèle que les ARNm sont traduits durant tout leur temps de vie.



**FIGURE 3.3** – (a) Distribution des temps de demi-vies des ARNm de souris obtenue par SHAROVA et al. [2009] sur tous les gènes. La médiane est de 7h. La dernière barre de l'histogramme doit contenir tous les temps de demi-vies supérieurs à 24h. (b) Distribution des demi-vies  $\tau$  obtenues par la distribution exponentielle des temps de vie 3.3 en injectant  $\omega = \ln(2)/\tau$ . Somme  $\Phi_\theta(3) + \Phi_\theta(6) + \Phi_\theta(10) + \Phi_\theta(12) + \Phi_\theta(15)$  de manière à considérer plusieurs gènes avec une médiane des temps de vie de 10h correspondant à la médiane des temps de demi-vie de 7h ( $\theta = \tau/\ln(2)$ ).

## 4 Traduction modélisée par TASEP inhomogène

Nous allons montrer dans cette section qu'on peut facilement intégrer la description Markovienne du modèle TASEP inhomogène sur la distribution des âges. Comme nous l'avons montré dans le chapitre 2 (Fig. 2.3), le temps de calcul augmente exponentiellement, ce qui permettra d'utiliser cette approche pour étudier seulement des petits systèmes, comme des ARNm de quelques codons.

### 4.1 Profil de densité polysome

Nous appelons "polysome" la population de tous les ARNm quel que soit le nombre de ribosomes en train de les traduire, à discerner des  $k$ -somes qui désignent les ARNm sur lesquels se trouve exactement  $k$  ribosomes. Nous allons simplement intégrer l'équation de la résolution exacte sur l'ensemble des âges. On a donc

$$\rho_i = \int_0^\infty \Phi_A(a) \rho_i(a) da = \omega \int_0^\infty e^{-\omega a} \rho_i(a) da = \omega \hat{\rho}_i(\omega) \quad (3.5)$$

où le symbole  $\hat{\phantom{x}}$  représente la transformée de Laplace. Dans la description Markovienne, la densité s'écrit donc d'après l'Eq. (2.6)

$$\rho_i = \omega \sum_{\{\mathcal{C} | n_i=1\}} \hat{P}_{\mathcal{C}}(\omega). \quad (3.6)$$

où  $P_{\mathcal{C}}$  est la probabilité que le réseau ait la configuration  $\mathcal{C}$  et  $\{\mathcal{C}_j | n_i = 1\}$  l'ensemble des configurations des états  $j$  où le site  $i$  est occupé. En prenant la transformée Laplace de l'équation maîtresse (2.1) avec un réseau initialement vide, on obtient

$$(\omega \mathbb{1} - M) |\hat{P}(\omega)\rangle = |P(0)\rangle. \quad (3.7)$$

où  $\mathbb{1}$  est la matrice identité. Une fois résolu, ce système algébrique fournit les quantités nécessaires pour calculer la densité polysome (voir Eq. (3.6)) mesurée dans les expériences Ribo-Seq. Il n'est donc pas nécessaire de résoudre l'équation de Markov elle-même, à savoir de calculer le spectre de la matrice de Markov.

### 4.2 Profil de densité $k$ -some

Il est également possible de calculer la densité moyenne le long de  $k$ -somes. Rappelons que les  $k$ -somes sont des ARNm sur lesquels se trouve exactement  $k$  ribosomes. La densité de ribosome au codon  $i$  d'un  $k$ -some à l'âge  $a$ ,  $\rho_k(i|a)$ , est obtenue en sélectionnant les configurations à  $k$  ribosomes, c'est-à-dire les configurations obéissant à la contrainte  $\sum_{j=1}^L n_j = k$ . On a donc

$$\rho_k(i|a) = \frac{1}{P(k|a)} \sum_{\{\mathcal{C} | n_i=1\}} P_{\mathcal{C}}(a) \delta_{\sum_{j=1}^L n_j, k}, \quad (3.8)$$

où  $P(k|a)$  est la probabilité qu'un ARNm soit un  $k$ -some à l'âge  $a$ , et s'écrit

$$P(k|a) = \sum_{\{\mathcal{C}\}} P_{\mathcal{C}}(a) \delta_{\sum_{j=1}^L n_j, k}. \quad (3.9)$$

On vérifie que la densité (3.8) est correctement normalisée quelque soit  $a$ , c'est à dire que

$$\sum_{i=1}^L \rho_k(i|a) = k \quad \forall a. \quad (3.10)$$

Pour calculer la densité moyennée sur les âges, nous devons d'abord calculer la densité de probabilité  $p(a|k)$  qu'un ARNm ait un âge  $a$  sachant qu'il est un  $k$ -some. En utilisant le théorème de Bayes, nous obtenons la distribution des âges des  $k$ -somes dans la soupe comme suit

$$p(a|k) = \frac{\Phi_A(a) P(k|a)}{P_k}. \quad (3.11)$$

où la probabilité  $P(k|a)$  qu'un ANRm soit un  $k$ -some à l'âge  $a$  est donnée par (3.9) et sa moyenne sur les âges par

$$P_k = \int_0^\infty \Phi_A(a) P(k|a) da = \omega \widehat{P}(k|\omega). \quad (3.12)$$

Le profil de densité moyenne d'un  $k$ -some indépendamment de son âge, défini par

$$\rho_k(i) = \int_0^\infty p(a|k) \rho_k(i|a) da, \quad (3.13)$$

est donc donnée par (voir Eq. (3.8), (3.11) and (3.12))

$$\rho_k(i) = \frac{1}{\widehat{P}(k|\omega)} \sum_{\{\mathcal{C}|n_i=1\}} n_i \widehat{P}_{\mathcal{C}}(\omega) \delta_{\sum_{j=1}^L n_j, k}. \quad (3.14)$$

Cette formule, dont on peut facilement vérifier la normalisation à  $k$ , est exacte. Elle ne contient que les transformées de Laplace de certaines composantes du vecteur de probabilité  $P(t)$  qui sont des solutions du système algébrique (3.7). Par conséquent, en résolvant ce système, on connaît le profil exact de la densité des  $k$ -somes dans les expériences RiboSeq. Cette formule est intéressante pour vérifier des résultats sur de petits réseaux seulement, car le temps de calcul est beaucoup trop long pour des réseaux correspondant à la longueur des ARN messagers puisqu'il augmente exponentiellement avec la longueur du réseau  $L$  (cf Fig. 2.3).

### 4.3 Exemple d'un ARNm à 2 codons

#### 4.3.1 Solution du système (3.7)

Afin de faciliter la compréhension de la démarche calculatoire, illustrons ces calculs sur l'exemple spécifique d'un ARNm à 2 codons. En appelant  $p$  le taux de saut entre les codons 1 et 2, la matrice de Markov  $M$  (2.11) se lit comme suit

$$M = \begin{pmatrix} -\alpha & \beta & 0 & 0 \\ 0 & -(\alpha + \beta) & p & 0 \\ \alpha & 0 & -p & \beta \\ 0 & \alpha & 0 & -\beta \end{pmatrix}. \quad (3.15)$$

Le vecteur probabilité et sa transformée de Laplace sont

$$|P(a)\rangle = \begin{pmatrix} P_{00}(a) \\ P_{01}(a) \\ P_{10}(a) \\ P_{11}(a) \end{pmatrix} \quad \text{and} \quad |\widehat{P}(\omega)\rangle = \begin{pmatrix} \widehat{P}_{00}(\omega) \\ \widehat{P}_{01}(\omega) \\ \widehat{P}_{10}(\omega) \\ \widehat{P}_{11}(\omega) \end{pmatrix}. \quad (3.16)$$

Comme l'ARNm est initialement vide, toutes les probabilités sont nulles sauf  $P_{00}(0)$  :

$$|P(0)\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \quad (3.17)$$

Résolvons maintenant le système (3.7) pour  $|\widehat{P}(\omega)\rangle$ , ce qui mène à

$$\widehat{P}_{00}(\omega) = \frac{(\alpha + \beta + \omega)(p + \omega)(\beta + \omega) - \alpha p \beta}{|\omega \mathbb{1} - M|}, \quad (3.18)$$

$$\widehat{P}_{01}(\omega) = \frac{\alpha p (\beta + \omega)}{|\omega \mathbb{1} - M|}, \quad (3.19)$$

$$\widehat{P}_{10}(\omega) = \frac{\alpha (\beta + \omega)(\alpha + \beta + \omega)}{|\omega \mathbb{1} - M|}, \quad (3.20)$$

$$\widehat{P}_{11}(\omega) = \frac{\alpha^2 p}{|\omega \mathbb{1} - M|}, \quad (3.21)$$

où  $|\omega\mathbb{1} - M|$  est le déterminant de  $(\omega\mathbb{1} - M)$ , c'est-à-dire la somme des numérateurs de (3.18-3.21) multipliés par <sup>1</sup>  $\omega$  qui peut être écrit sous la forme

$$|\omega\mathbb{1} - M| = \omega [\alpha p(\alpha + \omega) + (\alpha + \beta + \omega)(p + \omega + \alpha)(\beta + \omega)]. \quad (3.22)$$

### 4.3.2 Profil de densité polysome

Selon (3.6), nous pouvons calculer le profil de densité des polysomes en sélectionnant les configurations ayant un ribosome sur un codon spécifique. On obtient ,

$$\rho_1 = \omega(\widehat{P}_{10} + \widehat{P}_{11}) \quad \text{et} \quad \rho_2 = \omega(\widehat{P}_{01} + \widehat{P}_{11}) \quad (3.23)$$

soit

$$\rho_1 = \frac{\alpha(\alpha p + (\alpha + \beta + \omega)(\beta + \omega))}{\alpha p(\alpha + \omega) + (\alpha + \beta + \omega)(p + \omega + \alpha)(\beta + \omega)} \quad (3.24)$$

et

$$\rho_2 = \frac{\alpha p(\alpha + \beta + \omega)}{\alpha p(\alpha + \omega) + (\alpha + \beta + \omega)(p + \omega + \alpha)(\beta + \omega)} \quad (3.25)$$

Notons que lorsque la durée de vie moyenne d'un ARNm devient très grande (c'est-à-dire quand  $\omega \rightarrow 0$ ) les expressions ci-dessus deviennent

$$\rho_1 \rightarrow \frac{\alpha(\alpha p + (\alpha + \beta)\beta)}{\alpha^2 p + \beta(\alpha + \beta)(p + \alpha)} \quad (3.26)$$

et

$$\rho_2 \rightarrow \frac{\alpha p(\alpha + \beta)}{\alpha^2 p + \beta(\alpha + \beta)(p + \alpha)} \quad (3.27)$$

qui sont précisément les densités de ribosomes obtenues pour un ARNm dans son régime stationnaire (cf Eq. (2.10) du chapitre précédent). Ceci est bien sûr vrai pour tout nombre de codons dans la mesure où le comportement de la transformée de Laplace d'une fonction pour  $\omega \rightarrow 0$  fournit son asymptote à long terme ( $t \rightarrow \infty$ ), c'est-à-dire ici l'état stationnaire du processus de Markov.

### 4.3.3 Profils de densités $k$ -somes

Pour un ARNm à deux codons, ce dernier est soit vide (configuration 00), soit rempli d'un ribosome (configurations 01 et 10), soit complètement rempli de 2 ribosomes (configuration 11). Il est donc facile de calculer le profil de densité de  $k$ -somes pour les monosomes et les disomes. D'après l'Eq. 3.14, nous avons pour les monosomes

$$\rho_1(1) = \frac{\widehat{P}_{10}}{\widehat{P}_{10} + \widehat{P}_{01}} = \frac{\alpha + \beta + \omega}{\alpha + \beta + \omega + p} \quad (3.28)$$

et

$$\rho_1(2) = \frac{\widehat{P}_{01}}{\widehat{P}_{10} + \widehat{P}_{01}} = \frac{p}{\alpha + \beta + \omega + p}. \quad (3.29)$$

Notons que les densités de monosomes s'additionnent correctement à 1 ribosome et que, pour  $\omega \rightarrow 0$ , elles convergent vers leur valeur stationnaire. Pour le disome, évidemment

$$\rho_2(1) = \frac{\widehat{P}_{11}}{\widehat{P}_{11}} = 1 = \rho_2(2). \quad (3.30)$$

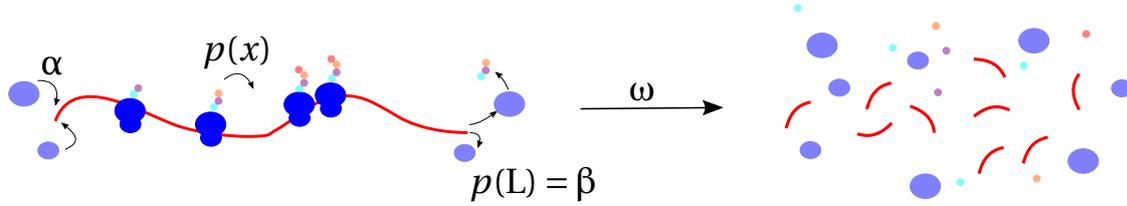
Le profil de densité est donc uniforme et s'additionne à 2 ribosomes, comme prévu. Le fait que le profil de densité soit uniforme est évidemment vrai pour toute configuration complètement remplie, quels que soient le nombre de codons et l'espérance de vie de l'ARNm.

1. Sachant que le vecteur de probabilité est normalisé à l'unité (la somme de toutes ses composantes est 1), la normalisation de sa transformée de Laplace est  $\omega^{-1}$ .

## 5 Modèle de transport balistique

Comme SINGH [1969] et NAGAR et al. [2011] nous avons choisi un mouvement balistique des ribosomes ainsi qu'un taux de dégradation exponentiel pour pouvoir donner une description simple des effets du temps de vie des ARNm sur les mesures Ribo-Seq.

### 5.1 Le modèle



**FIGURE 3.4** – Schéma du modèle de transport balistique incluant la dégradation. L'ARNm est représenté par la courbe rouge et est de longueur  $L$ . Les ribosomes sont des particules ponctuelles se déplaçant sur l'ARNm avec un taux  $p(x)$  dans une seule direction. Ils sont néanmoins représentés par leur deux sous-unités qui se lient à l'entrée de l'ARNm pour initier la traduction avec un taux  $\alpha$  et se détachent à la sortie avec le taux  $p(L) = \beta$ . L'ARNm peut se dégrader selon une loi exponentielle de paramètre  $\omega$ .

Une fois recruté par un brin d'ARNm, un ribosome est supposé avoir une trajectoire déterministe et avancer avec une vitesse spécifique au codon. Dans un cadre continu, nous appelons  $x$  la coordonnée le long de l'ARNm et  $p(x)$  la vitesse du ribosome au point  $x$ . La position  $x(a)$  d'un ribosome entrant dans un brin d'ARNm à  $a = 0$  est donc donnée par  $\dot{x}(a) = p(x(a))$ , et

$$\int_0^{x(a)} \frac{dx}{p(x)} = a. \quad (3.31)$$

Nous appelons  $\mathcal{T}(L)$  le temps que met le ribosome pour traverser tout l'ARNm. On a,

$$\mathcal{T}(L) = \int_0^L \frac{dx}{p(x)}. \quad (3.32)$$

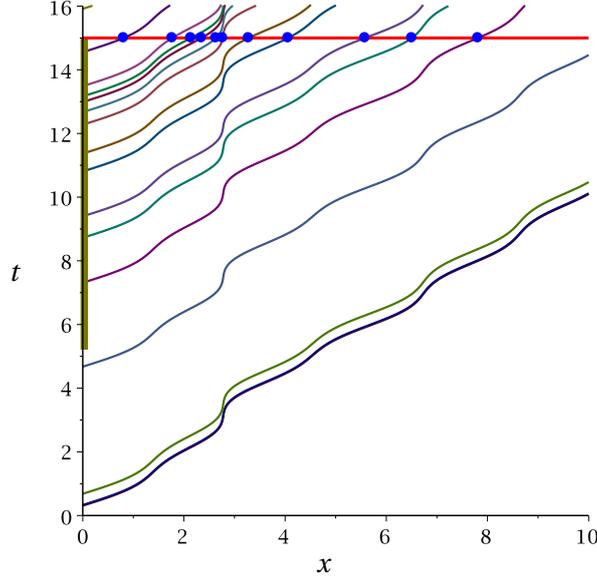
Ensuite, nous supposons que les ribosomes sont recrutés par l'ARNm à des moments aléatoires avec des temps entre chaque entrée distribués de façon exponentielle. Par conséquent, le  $i$ -ème ribosome entre sur un ARNm né à  $t_b$  au temps

$$t_i = t_b + \sum_{j=1}^i T_j, \quad (3.33)$$

où la variable aléatoire indépendante et identiquement distribuée  $T_i$  est gouvernée par la densité de probabilité  $q(T) = \alpha \exp(-\alpha T)$ ,  $T \geq 0$ .

## 5.2 Probabilité qu'un ARNm né à $t_b$ soit un $k$ -some à $t_{ob}$

Les ribosomes entrent dans un ARNm né à  $t_b$  aux dates  $t_i = t_b + \sum_{j=1}^i T_j$ . Il y a deux possibilités. Si le temps d'observation  $t_{ob}$  est inférieur à  $\mathcal{T}(L)$ , les ribosomes se trouvant sur l'ARNm à  $t_{ob}$  sont ceux qui sont entrés aux temps  $t_i \in [t_b, t_{ob}]$ . Mais, si le temps d'observation  $t_{ob}$  est plus grand que le temps nécessaire pour qu'un ribosome traverse la totalité de l'ARNm,  $\mathcal{T}(L)$ , les ribosomes qui se trouvent sur l'ARNm à  $t_{ob}$  sont ceux qui sont entrés aux temps  $t_i \in [t_{ob} - \mathcal{T}(L), t_{ob}]$ . En effet, ceux qui sont entrés avant  $t_{ob} - \mathcal{T}(L)$  sont déjà sortis de l'ARNm. La Fig. 3.5 illustre ce processus.



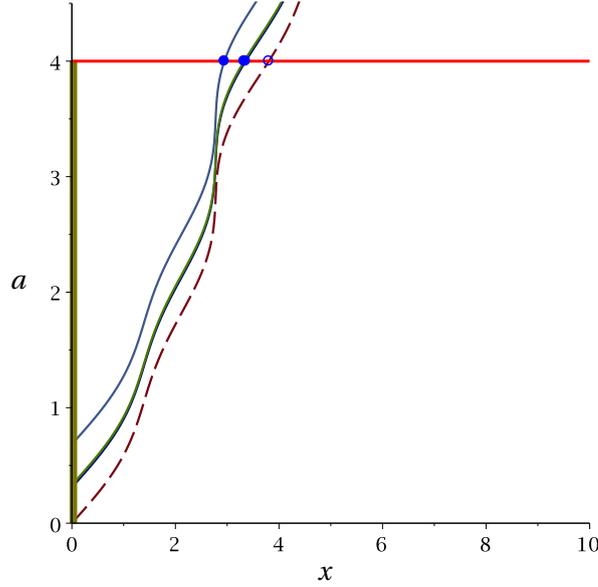
**FIGURE 3.5** – Diagramme spatio-temporel illustrant les trajectoires des ribosomes entrant à des moments aléatoires sur un brin d'ARNm né à  $t_b = 0$  et de longueur  $L = 10$  (unités arbitraires). Au temps d'observation  $t_{ob} = 15$ , la ligne rouge horizontale épaisse représente l'ARNm. Les disques bleus précisent l'emplacement des ribosomes dans l'ARNm à  $t_{ob} = 15$ . Les ribosomes qui sont entrés sur l'ARNm dans l'intervalle de temps désigné par la ligne épaisse vert olive sur l'axe des ordonnées, c'est-à-dire entre  $t_{ob} - \mathcal{T}(L)$  et  $t_{ob}$ , seront visible sur l'ARNm à  $t_{ob}$ .

Ce diagramme spatio-temporel montre un brin d'ARNm né à  $t_b = 0$  et de longueur  $L = 10$  (unités arbitraires) sur l'axe  $x$  et des ribosomes entrant dans cet ARNm à des moments aléatoires  $t_i$  (axe  $t$ ). La bande vert olive sur l'axe  $a$  représente la fenêtre temporelle  $[t_{ob} - \mathcal{T}(L), t_{ob}]$  pour  $t_{ob} = 15$ . La dynamique de ribosome choisie est telle que le temps nécessaire pour traverser l'ARNm entier est  $\mathcal{T}(L = 10) \approx 9.73$ . Les trajectoires des ribosomes sont représentées par des courbes pleines colorées. Les ribosomes situés sur l'ARNm à  $t = t_{ob}$  sont ceux dont les trajectoires intersectent la ligne rouge horizontale épaisse qui matérialise l'ARNm. Ils sont représentés par des disques bleus. Leurs positions sont données par  $x_i = x(t_{ob} - t_i)$ . Lorsque  $t_{ob} - t_b$  est inférieur à  $\mathcal{T}(L)$  tous les ribosomes qui ont pénétré dans l'ARNm entre  $t_b$  et  $t_{ob}$  y sont encore. Cette situation est représentée sur la Fig. 3.6.

Étant donné que les temps d'interarrivée  $T_i$  sont distribués de manière exponentielle, le nombre d'arrivées  $t_i$  compris dans la fenêtre temporelle suit une loi de Poisson avec une espérance proportionnelle à sa largeur, indépendamment de sa position spécifique dans le temps (voir par exemple [ENGEL, 2011]). La probabilité qu'un brin d'ARNm né à  $t_b$  héberge  $k$  ribosomes au temps  $t_{ob}$  est donc donnée par

$$P(k|a) = \begin{cases} \frac{(\alpha a)^k}{k!} e^{-\alpha a}, & a \leq \mathcal{T}(L), \\ \frac{(\alpha \mathcal{T}(L))^k}{k!} e^{-\alpha \mathcal{T}(L)}, & a \geq \mathcal{T}(L), \end{cases} \quad (3.34)$$

où  $a = t_{ob} - t_b$  est l'âge de l'ARNm au temps  $t_{ob}$ .



**FIGURE 3.6** – Diagramme spatio-temporel illustrant les trajectoires des ribosomes entrant à des moments aléatoires dans un brin d'ARNm né à  $t_b = 0$ . Au moment de l'observation  $t_{ob} = 4$ , l'épais trait horizontal rouge représente l'ARNm. Ce temps d'observation est inférieur à  $\mathcal{T}(L = 10) \simeq 9.73$ , le temps nécessaire pour traverser tout l'ARNm. Les disques bleus précisent l'emplacement des ribosomes au sein de l'ARNm à  $t_{ob} = 4$ . La trajectoire en pointillés représente un ribosome qui serait entré immédiatement dans l'ARNm lorsque l'ARNm est né à  $a = 0$  et le cercle bleu, l'emplacement de ce ribosome à  $t_{ob}$ . Au-delà de cette position, aucun ribosome ne peut être présent : la densité de ribosomes devient nulle.

### 5.3 Probabilité qu'un ARNm de la soupe soit un $k$ -some

Pour déterminer la probabilité de trouver un  $k$ -some dans la soupe, nous faisons la moyenne de la probabilité qu'un ARNm soit un  $k$ -some à l'âge  $a$ ,  $P(k|a)$  (Eq. (3.34)) sur la distribution d'âge  $\Phi_A(a)$  (Eq. (3.1)) :

$$P_k = \int_0^\infty P(k|a)\Phi_A(a) da. \quad (3.35)$$

On obtient

$$P_k = \frac{\omega}{\alpha + \omega} \left( \frac{\alpha}{\alpha + \omega} \right)^k \frac{\gamma(k+1, (\alpha + \omega)\mathcal{T}(L))}{k!} + \frac{(\alpha\mathcal{T}(L))^k}{k!} e^{-(\alpha + \omega)\mathcal{T}(L)}, \quad (3.36)$$

où  $\gamma(k, z)$  est la fonction gamma incomplète définie par

$$\gamma(k, z) = \int_0^z e^{-t} t^{k-1} dt. \quad (3.37)$$

Le terme de  $P_k$  à gauche du signe "+" correspond aux ARNm en état transitoire de traduction

### 5.4 Statistique des $k$ -somes

Bien que cela ne nous ait pas servi pour la suite de la thèse, nous profitons de l'expression de la probabilité d'avoir un  $k$ -some pour calculer des grandeurs statistiques sur les  $k$ -somes qui pourraient être utiles à une analyse future des données Ribo-Seq.

**Probabilité que la soupe contienne  $N_k$   $k$ -somes** Disons que la soupe contient  $N$  ARNm. La probabilité qu'elle contienne exactement  $N_k$   $k$ -somes peut être évaluée comme suit. Un ARNm donné a une probabilité  $P_k$  d'être un  $k$ -some et  $Q_k = 1 - P_k$  de ne pas être un  $k$ -some. La probabilité que la soupe contienne exactement  $N_k$   $k$ -somes est donc binomiale.

$$\mathcal{P}^{(k)}(N_k) = \binom{N}{N_k} P_k^{N_k} Q_k^{N-N_k}. \quad (3.38)$$

La fonction génératrice correspondante s'écrit

$$\mathcal{N}^{(k)}(z) := \sum_{n=0}^{\infty} \mathcal{P}^{(k)}(n) z^n = (zP_k + Q_k)^N. \quad (3.39)$$

**Moyenne et variance du nombre de  $k$ -somes dans la soupe** En utilisant la fonction génératrice (3.39), Le nombre moyen de  $k$ -somes dans la soupe,  $\langle N_k \rangle$  est donné par

$$\langle N_k \rangle = \left. \frac{\partial \mathcal{N}^{(k)}}{\partial z} \right|_{z=1} = NP_k, \quad (3.40)$$

comme attendu, et comme

$$\langle N_k^2 \rangle = \left. \frac{\partial^2 \mathcal{N}^{(k)}}{\partial z^2} \right|_{z=1} - \langle N_k \rangle, \quad (3.41)$$

sa variance est

$$\sigma^2(N_k) = \langle N_k^2 \rangle - \langle N_k \rangle^2 = NP_k Q_k. \quad (3.42)$$

### 5.5 Densité $k$ -some en fonction de son âge

Appelons les temps  $t_i$  auxquels les ribosomes s'accrochent sur un ARNm donné, les temps d'"arrivées". Comme nous l'avons déjà vu, les ribosomes qui se trouvent encore sur l'ARNm à l'instant d'observation  $t_{ob}$  sont soit ceux qui ont pénétré dans l'ARNm aux arrivées  $t_i \in [t_{ob} - \mathcal{F}(L), t_{ob}]$  si  $t_{ob} - \mathcal{F}(L) > t_b$  ou bien ceux qui sont entrés à  $t_i \in [t_b, t_{ob}]$  si  $t_{ob} - \mathcal{F}(L) < t_b$ . Ici, comme précédemment,  $\mathcal{F}(L)$  est le temps mis par un ribosome pour traverser tout l'ARNm et  $t_b$ , la date de naissance de l'ARNm.

Nous utilisons maintenant un théorème important pour les processus de Poisson : si nous savons qu'il existe  $k$  temps d'arrivées  $t_i$  pendant l'intervalle de temps  $[t - \tau, t]$ , leur densité de probabilité est uniforme sur cet intervalle de temps (voir par exemple le livre de DALEY et al. [2008] pour un énoncé mathématique rigoureux de ce théorème). En d'autres termes, la distribution des  $k$   $t_i$  s'effectuant sur l'intervalle de temps  $[t - \tau, t]$  peut être obtenue en tirant  $k$  variables uniformément sur  $[t - \tau, t]$  puis en les ordonnant chronologiquement. Ceci fournit un moyen très simple d'évaluer le profil de densité des ribosomes d'un  $k$ -some. En effet, à un  $t_i$  donné correspond une position du ribosome à  $t_{ob}$  donnée par  $x_i := x(t_{ob} - t_i)$ . Par définition, le profil de densité des ribosomes sachant que l'ARNm né à  $t_b$  est un  $k$ -some à  $t_{ob}$  est donné par

$$\rho_k(x, t_{ob} - t_b) := \left\langle \sum_{i=1}^k \delta(x - x_i) \right\rangle = k \langle \delta(x - x_1) \rangle, \quad (3.43)$$

où la dernière espérance est censée s'écouler sur un "temps d'arrivée"  $t_1$  qui est uniformément distribué sur l'intervalle de temps  $[\max(t_b, t_{ob} - \mathcal{F}(L)), t_{ob}]$ .

Nous montrons dans l'annexe 1.1 que ceci donne le profil de densité de ribosomes suivant :

$$\rho_k(x|a) = \begin{cases} \frac{k}{ap(x)} H(x(a) - x), & a \leq \mathcal{F}(L), \\ \frac{k}{\mathcal{F}(L)p(x)}, & a \geq \mathcal{F}(L), \end{cases} \quad (3.44)$$

où  $a = t_{ob} - t_b$  est l'âge de l'ARNm à  $t_{ob}$  et où  $p(x)$  est la vitesse d'un ribosome situé à  $x$  sur l'ARNm.  $H(x)$  représente la fonction de Heaviside. Dans la première ligne de (3.44),  $x(a)$  est la distance maximale qu'un ribosome peut parcourir le long d'un ARNm d'âge  $a < \mathcal{F}(L)$ .

La première expression de  $\rho^{(k)}(x)$  tient compte du fait que les ARNm "fraîchement" nés, c'est-à-dire nés trop près de la date d'observation pour qu'un ribosome puisse le traverser entièrement, ont un profil de densité nulle pour les positions supérieures à la distance maximale qu'un ribosome est capable de parcourir dans le temps  $t_{ob} - t_b$ , à savoir  $x = x(t_{ob} - t_b)$  (voir Fig. 3.6). La deuxième expression montre que, dès lors que les ribosomes ont la possibilité de traverser toute

la longueur  $L$  de l'ARNm, le profil de densité des ribosomes est simplement proportionnel à l'inverse de sa vitesse le long de l'ARNm.

Nous soulignons qu'il est possible de distinguer deux états pour les ARNm que les ribosomes traduisent. Le premier est l'état transitoire, pour lequel les ARNm jeunes avec  $a < \mathcal{T}(L)$  ont des densités  $\rho_k(x|a)$  qui dépendent explicitement de l'âge  $a$ , contrairement à l'état stationnaire pour lequel  $a > \mathcal{T}(L)$  et  $\rho_k(x|a)$  ne dépend plus explicitement de  $a$ . Nous verrons plus loin que la proportion d'ARNm dans l'état transitoire plutôt que dans l'état stationnaire est déterminante dans la caractérisation de la densité moyennée sur les âges.

On vérifie que dans les deux cas, le profil de densité est correctement normalisé. En effet, en considérant (3.31),  $\int_0^L \rho_k(x|a) dx = k$ .

## 5.6 Profil de densité $k$ -some

Le profil de densité pour un  $k$ -some né à  $t_b$  et observé à  $t_{ob}$ , c'est-à-dire avec un âge  $a = t_{ob} - t_b$ , a été calculé en Eq. (3.44). Pour obtenir le profil de densité moyen d'un  $k$ -some dans la soupe, on calcule d'abord la densité de probabilité  $p(a|k)$  qu'un ARNm ait un âge  $a$  déjà définie dans l'approche Markovienne (Eq. (3.11)) où la distribution des âges des ARNm  $\Phi_A(a)$  est donnée par (3.1), la probabilité  $P(k|a)$  qu'un ARNm soit un  $k$ -some sachant que son âge est  $a$  est donnée par (3.34) et la probabilité  $P_k$  qu'un ARNm soit un  $k$ -some quel que soit son âge est donnée par (3.36).

Le profil de densité des  $k$ -somes dans la soupe est maintenant obtenu en faisant la moyenne du profil de densité de  $k$ -somes d'âge  $a$  sur la distribution d'âge de ces  $k$ -somes  $p(a|k)$ . Cela donne

$$\rho_k(x) := \int_0^\infty \rho_k(x|a) p(a|k) da \quad (3.45)$$

$$= \frac{\omega}{P_k^p(x)} \left( \frac{\alpha}{\alpha + \omega} \right)^k \frac{\gamma(k, (\alpha + \omega)\mathcal{T}(L)) - \gamma(k, (\alpha + \omega)\mathcal{T}(x))}{(k-1)!} + \frac{\alpha}{P_k p(x)} \frac{(\alpha\mathcal{T}(L))^{k-1}}{(k-1)!} e^{-(\alpha + \omega)\mathcal{T}(L)} \quad (3.46)$$

où

$$\mathcal{T}(x) = \int_0^x \frac{dy}{p(y)} \quad (3.47)$$

et où  $\gamma(k, z)$  a été défini en (3.37). Comme attendu, nous avons  $\int_0^L \rho_k(x) dx = k$ . Le terme à gauche du signe "+" correspond aux ARNm en état transitoire tandis que le terme à droite du signe "+" correspond aux ARNm en état stationnaire.

Notons que, pour une durée de vie de l'ARNm qui dépasse largement  $\mathcal{T}(L)$ , on retrouve la valeur "stationnaire" de la densité, à savoir

$$\rho_k(x) \sim \frac{k}{\mathcal{T}(L)p(x)}, \quad \omega \rightarrow 0. \quad (3.48)$$

## 5.7 Profil de densité polysome

Nous procédons maintenant au calcul du profil global de densité de ribosomes sur l'ARNm, le profil polysome  $\rho(x)$ . Il peut être obtenu en sommant les contributions des différents  $k$ -somes, ou en intégrant le profil de densité des polysome d'âge  $a$  sur les âges. On a donc

$$\rho(x) = \sum_{k=0}^\infty \rho_k(x) P_k = \int_0^\infty \rho(x|a) \Phi_A(a) da, \quad (3.49)$$

où  $\rho(x|a)$  est le profil de densité des polysome d'âge  $a$  (au temps d'observation  $t_{ob}$ ) qui est obtenu en sommant les contributions de tous les  $k$ -somes d'âge  $a$ ,

$$\begin{aligned} \rho(x|a) &= \sum_{k=0}^\infty P(k|a) \rho_k(x|a) \\ &= \begin{cases} \frac{\alpha}{p(x)} H(x(a) - x), & a \leq \mathcal{T}(L) \\ \frac{\alpha}{p(x)}, & a \geq \mathcal{T}(L) \end{cases} \end{aligned} \quad (3.50)$$

(voir Eqs. (3.34) et (3.44)). En calculant la moyenne sur tous les âges possibles, on obtient

$$\rho(x) = \int_0^\infty \rho(x|a)\Phi_A(a) da = \frac{\alpha}{p(x)} e^{-\omega\mathcal{F}(x)}, \quad (3.51)$$

où  $\mathcal{F}(x)$  est défini en Eq. (3.47). On a retrouvé l'expression de VALLERIANI et al. [2010], pour un taux de détachement nul avant la fin de l'ARNm. On constate qu'elle correspond à celle du modèle TASEP en basse densité (Eq. (2.22c)), ce qui confirme que le modèle balistique correspond au modèle TASEP lorsque la densité de particule est suffisamment basse pour que les corrélations entre particules soient négligeables. En intégrant la densité globale sur la longueur  $L$  de l'ARNm, on obtient le nombre moyen de ribosomes sur les ARNm de la soupe

$$\langle k \rangle = \int_0^L \rho(x) dx = \frac{\alpha}{\omega} (1 - e^{-\omega\mathcal{F}(L)}). \quad (3.52)$$

## 6 Analyse de l'effet de durée de vie finie

Dans cette partie, nous allons étudier quelle est l'influence de la durée de vie finie sur les grandeurs physiques en fonction des paramètres du modèle. Pour cela, nous utiliserons  $\mathcal{F}(L)$  comme échelle de temps. Nous définissons les taux sans dimension comme les taux originaux multipliés par  $\mathcal{F}(L)$  et utilisons un symbole tilde pour les désigner :

$$\tilde{\alpha} = \alpha\mathcal{F}(L); \quad \tilde{\omega} = \omega\mathcal{F}(L); \quad \tilde{p}(x) = p(x)\mathcal{F}(L). \quad (3.53)$$

Nous montrerons que  $\tilde{\alpha}$  et  $\tilde{\omega}$  sont en effet les grandeurs qui régissent l'effet de temps de fini sur les densités  $k$ -somes et polysome. Ce travail nous permettra notamment d'interpréter qualitativement les données des expériences Ribo-Seq, puis de définir une méthode de déduction de chaque paramètre cinétique de la traduction pour chaque gène dans de futures analyses quantitatives. Cette méthode sera décrite dans le chapitre suivant.

### 6.1 Quelques taux de la traduction génétique

Dans le tableau 3.1 nous avons reporté toutes les mesures moyennes que nous avons trouvées dans la littérature des paramètres en jeu dans notre modèle. Elles nous permettent d'appréhender les ordres de grandeurs des taux afin que nous puissions valider les modèles biophysiques et déterminer un régime théorique correspondant à la situation paramétrique des cellules vivantes.

Il est intéressant de remarquer, déjà avec ces quatre espèces, qu'une seule grandeur varie de manière significative : la durée de vie de l'ARN messenger, tandis que les trois autres grandeurs, les taux d'élongation et d'initiation de la traduction, et la longueur médiane des gènes varient peu, y compris entre espèces simples et espèces évoluées. Cela pose la question de la raison d'une évolution en faveur de temps de vie plus long pour les ARNm.

Dans le tableau 3.2, nous calculons les taux adimensionnels qui sont pertinents pour ce système comme nous allons le voir tout au long de ce chapitre.

### 6.2 Densité polysome et nombre moyen de ribosomes sur l'ARN messenger, $\langle k \rangle$

Nous discutons dans cette sous-section, de l'effet de la dégradation sur l'ensemble de la population d'ARNm, les polysomes. Pour cela, nous considérerons la fonction d'échelle  $f(\tilde{\omega})$  du ratio du nombre moyen de ribosomes  $\langle k \rangle$  défini dans l'équation (3.52) sur la même quantité pour un temps de vie infini,  $\langle k \rangle_\infty = \tilde{\alpha}$  :

$$f(\tilde{\omega}) = \frac{\langle k \rangle}{\langle k \rangle_\infty} = \frac{\langle k \rangle}{\tilde{\alpha}} = \frac{1}{\tilde{\omega}} (1 - e^{-\tilde{\omega}}). \quad (3.54)$$

Espèce	$t_{1/2}$ ARNm (médiane)	$p$ (aa/s)	L (médiane, en codons)	$\alpha$ ( $s^{-1}$ )	$\mathcal{T}(L)$ (s)
<i>E. coli</i>	5 min	10-20	256	-	13-26
<i>S. cerevisiae</i>	20 min	3-10	399	0.1	40-133
<i>M. musculus</i>	7 h	6	493	-	82
<i>H. sapiens</i>	10 h	3	481	0.06	160

**TABEAU 3.1** – Taux cinétiques de traduction et propriétés des ARNm chez les espèces pour lesquelles ces valeurs ont été mesurées, à notre connaissance. Les médianes de la longueur des gènes L sont calculées à partir des génomes entiers disponibles dans la base de données NCBI. Demi-vies des ARNm  $t_{1/2}$  [BERNSTEIN et al., 2002; SHAROVA et al., 2009; WANG et al., 2002; YANG et al., 2003], taux d'élongation de la chaîne polypeptidique  $p$  en acides aminés par seconde [INGOLIA et al., 2011; KARPINETS et al., 2006; YAN et al., 2016; YOUNG et BREMER, 1976] et taux d'initiation  $\alpha$  [CIANDRINI et al., 2013; TRÖSEMEIER et al., 2019] trouvés dans la littérature. Le temps de traduction d'un gène est  $\mathcal{T}(L) \approx L/p$ .

Espèce	$\tilde{\omega} = \omega \mathcal{T}(L)$	$\tilde{\alpha} = \alpha \mathcal{T}(L)$	$\alpha/\omega$	$\mathcal{R}_1(0)$
<i>E. coli</i>	0.04-0.07	-	-	-
<i>S. cerevisiae</i>	0.02-0.08	4-13	170	0.3-3700
<i>M. musculus</i>	0.002	-	-	-
<i>H. sapiens</i>	0.003	9.6	$3.1 \times 10^3$	4.8

**TABEAU 3.2** – Paramètres sans dimension chez les espèces pour lesquels les valeurs des taux d'initiation et d'élongation de la traduction, et les temps de demi-vies des ARNm ont été mesurés à notre connaissance. Calculs faits à partir des valeurs du tableau 3.1 du chapitre 1. Le paramètre  $\mathcal{R}_1(0)$  sera défini en section 6.6.

Nous traçons  $f(\tilde{\omega})$  en fonction de  $\tilde{\omega}$  dans la Fig. 3.7. Un développement en série pour les petits  $\tilde{\omega}$  conduit à

$$f(\tilde{\omega}) = 1 - \frac{\tilde{\omega}}{2} + \mathcal{O}(\tilde{\omega}^2), \quad (3.55)$$

tandis que lorsque  $\tilde{\omega} \rightarrow \infty$ , on obtient

$$f(\tilde{\omega}) = \frac{1}{\tilde{\omega}} + \mathcal{O}\left(\frac{1}{\tilde{\omega}^2}\right). \quad (3.56)$$

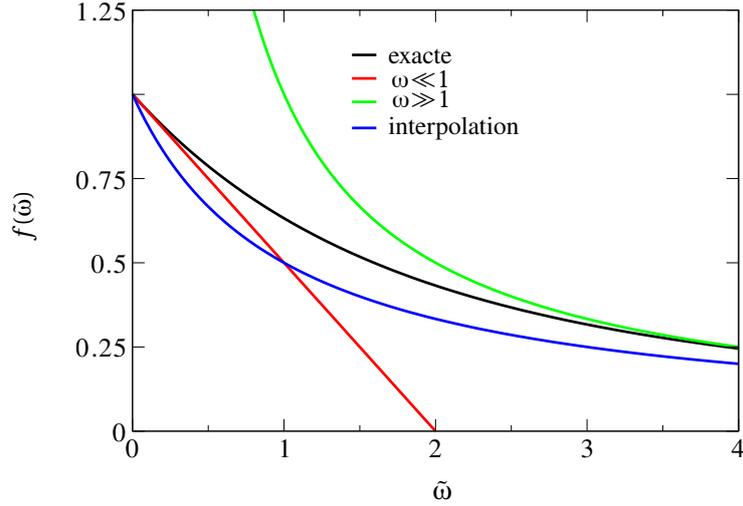
Nous donnons une explication heuristique des rôles joués par les trois paramètres du modèle  $\alpha$ ,  $\mathcal{T}(L)$  et  $\omega$  dans la détermination de la densité moyenne des polysomes. Dans le cadre du modèle balistique, l'inverse du temps nécessaire pour traverser l'ARNm,  $1/\mathcal{T}(L)$  et le taux de dégradation  $\omega$  de l'ARNm jouent le rôle de taux de mortalité effectifs qui compensent le taux d'initiation  $\alpha$ , conduisant ainsi à un nombre moyen de ribosomes  $\langle k \rangle$  sur l'ARNm. Les deux taux de mortalité,  $1/\mathcal{T}(L)$  et  $\omega$ , sont en compétition et c'est le plus élevé qui détermine principalement la valeur de  $\langle k \rangle$ . Une interpolation simple prend la forme

$$\frac{1}{\langle k \rangle_I} \equiv \frac{1}{\alpha \mathcal{T}(L)} + \frac{\omega}{\alpha}, \quad (3.57)$$

où  $\langle k \rangle_I$  est la valeur interpolée du nombre moyen de ribosomes sur le polysome. Cela conduit aux formes limites correctes,  $\langle k \rangle = \tilde{\alpha} = \alpha \mathcal{T}(L)$ , lorsque  $\tilde{\omega} = \omega \mathcal{T}(L) \ll 1$ , et  $\alpha/\omega$  lorsque  $\tilde{\omega} \gg 1$ . On remarque que le passage d'un régime à l'autre est déterminé uniquement par  $\tilde{\omega}$ , qui ne fait pas intervenir  $\alpha$ . Cette formule d'interpolation peut être mise sous une forme de mise à l'échelle

$$f_I(\tilde{\omega}) = \frac{\langle k \rangle_I}{\tilde{\alpha}} = \frac{1}{1 + \tilde{\omega}}. \quad (3.58)$$

Dans la Fig. 3.7, nous comparons les formes de mise à l'échelle de différentes expressions pour  $f(\tilde{\omega})$  : valeur exacte, développement en série à petite et grande échelle de  $f(\tilde{\omega})$  et l'interpolation  $f_I(\tilde{\omega})$ . Nous en déduisons que l'interpolation fournit une approximation raisonnable sur toute la gamme de  $\tilde{\omega}$ .



**FIGURE 3.7** – Tracé de la fonction d'échelle  $f(\tilde{\omega})$  définie dans l'équation (3.54) en fonction de  $\tilde{\omega} = \omega \mathcal{T}(L)$  (ligne noire). Les comportements asymptotiques lorsque  $\tilde{\omega} \rightarrow 0$ , c'est-à-dire,  $f(\tilde{\omega}) \sim 1 - \tilde{\omega}/2$  (ligne rouge) et lorsque  $\tilde{\omega} \rightarrow \infty$ , c'est-à-dire,  $f(\tilde{\omega}) \sim 1/\tilde{\omega}$  (ligne verte) sont également représentés. La courbe bleue est l'interpolation phénoménologique  $f_1(\tilde{\omega})$  donnée par l'Eq. (3.57). On remarque un bon accord entre la formule exacte et la formule interpolée, et on vérifie les comportements asymptotiques. En particulier,  $f(\tilde{\omega})$  est dans sa forme asymptotique à  $\tilde{\omega} \rightarrow 0$  dans les cellules vivantes puisque  $\tilde{\omega} \sim 10^{-3} - 10^{-2}$  (Tab. 3.2).

Nous estimons le changement de régime à  $\omega \approx \mathcal{T}(L)^{-1}$ . Puisque le temps de vie moyen est  $1/\omega$ , nous pouvons dire que le croisement correspond à un temps de vie moyen égale au temps  $\mathcal{T}(L)$  qu'il faut au ribosome pour traduire tout l'ARNm. Il s'agit donc aussi d'une limite à la viabilité du système, puisqu'au-delà, l'ARNm ne peut pas être traduit avant d'être dégradé et donc la protéine n'est pas produite.

Les valeurs biologiques issues de la littérature présentées dans le Tab. 3.1 et 3.2 nous permettent de quantifier  $\tilde{\omega}$  dans différentes espèces : chez la levure *S. cerevisiae* et la bactérie *E. Coli*,  $\tilde{\omega} \sim 10^{-2}$ , ce qui est un ordre de grandeur plus élevé que chez les mammifères (*H. sapiens* et *M. musculus*) pour lesquels  $\tilde{\omega} \sim 10^{-3}$ . Cependant, dans tous les cas, la limite  $\tilde{\omega} \ll 1$  est vérifiée pour la majorité des gènes. Nous pouvons donc dire que dans les cellules vivantes, le nombre de ribosomes moyen  $\langle k \rangle = \langle k \rangle_\infty = \tilde{\alpha}$  et ne dépend pas de la durée de vie de l'ARNm.

Nous concluons par une discussion sur le profil de densité du polysome,  $\rho(x)$ , qui est la quantité expérimentalement accessible mesurée dans les expériences Ribo-seq. Comme la densité moyenne globale est égale à

$$\rho_{ave} = \langle k \rangle / L, \quad (3.59)$$

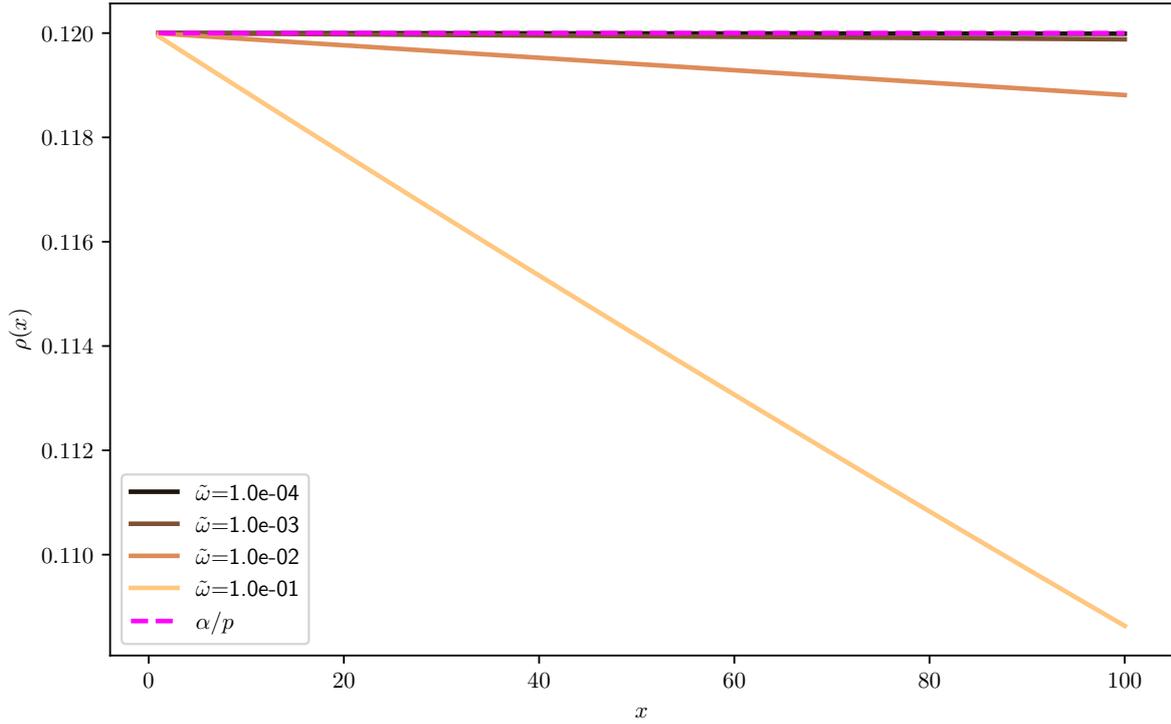
$f(\tilde{\omega})$  peut être interprété comme le rapport entre cette quantité et la même quantité à durée de vie infinie,  $f(\tilde{\omega}) = \rho_{ave} / \rho_{ave}^\infty$ . Nous voyons d'après l'équation (3.51) que l'influence de la dégradation de l'ARNm sur le profil Ribo-Seq du polysome est maximisée au dernier site où  $\rho(L) / \rho^\infty(L) = e^{-\tilde{\omega}}$ , ce qui conduit à

$$\frac{|\rho(x) - \rho^\infty(x)|}{\rho^\infty(x)} \leq 1 - e^{-\tilde{\omega}} \approx \tilde{\omega}, \quad (3.60)$$

où l'approximation est valable pour les petits  $\tilde{\omega}$ . Par conséquent, les différents intervalles de  $\omega$  discutés ci-dessus pour la fonction d'échelle  $f(\tilde{\omega})$  sont également valables pour  $\rho(x)$ . En particulier, nous observons dans le tableau 3.2 que  $\tilde{\omega} < 10^{-1}$ , une valeur qui conduit à une erreur dans l'estimation du taux de saut inférieure à 10%. Nous allons préciser cette erreur dans la partie suivante.

### 6.3 Erreur sur les estimations des taux de saut avec les modèles à temps de vie infini

La traduction et les données Ribo-Seq correspondent à la version discrète du modèle que nous donnons en annexe, section 1.4. A chaque codon correspond une densité de ribosome  $\rho_i$  et un



**FIGURE 3.8** – Comparaison de la densité polysome à durée de vie finie et infinie.  $\alpha = 0.06$  et  $p = 0.5$ . Dès  $\tilde{\omega} = 10^{-4}$ , le profil se confond avec celui à durée de vie infinie. Une différence d'environ 10% n'est visible qu'à partir de  $\tilde{\omega} = 0.1$ .

taux de saut  $p_i$ . Nous obtenons la densité discrète au codon  $i$  en intégrant la densité continue de  $i - 1$  à  $i$ . Pour la densité polysome, nous obtenons

$$\rho_i = \frac{\alpha}{\omega} \left( e^{-\omega \mathcal{F}(i-1)} - e^{-\omega \mathcal{F}(i)} \right). \quad (3.61)$$

Les taux de saut valent donc

$$p_i = - \frac{\omega}{\ln \left( 1 - \frac{\omega}{\alpha} e^{\omega \mathcal{F}(i-1)} \rho_i \right)}. \quad (3.62)$$

L'erreur relative absolue faite sur l'estimation des taux de saut quand on ne tient pas compte du temps de vie fini est donc

$$Err_{LT}(p_i) = -1 - \frac{\alpha}{\omega \rho_i} \ln \left( 1 - \frac{\omega}{\alpha} e^{\omega \mathcal{F}(i-1)} \rho_i \right) = -1 - \frac{\tilde{\alpha}}{\tilde{\omega} \rho_i} \ln \left( 1 - \frac{\tilde{\omega}}{\tilde{\alpha}} e^{\tilde{\omega} \frac{\mathcal{F}(i-1)}{\mathcal{F}(L)}} \rho_i \right). \quad (3.63)$$

Nous observons une dépendance en  $\tilde{\alpha}$  qui disparaît cependant si on considère  $\tilde{\omega} \ll 1$  et  $\frac{\tilde{\omega}}{\tilde{\alpha}} \ll 1$ , ce qui doit nécessairement être le cas pour que la séquence soit entièrement traduite. Nous obtenons alors

$$Err_{LT}(p_i) = e^{\tilde{\omega} \frac{\mathcal{F}(i-1)}{\mathcal{F}(L)}} - 1. \quad (3.64)$$

Nous remarquons que nous obtenons le même résultat avec la version continue, en regardant la différence relative de  $\frac{\alpha}{\rho(x)} e^{\omega \mathcal{F}(x)}$  avec  $\frac{\alpha}{\rho(x)}$  et, soit

$$Err_{LT}(p(x)) = e^{\tilde{\omega} \frac{\mathcal{F}(x)}{\mathcal{F}(L)}} - 1. \quad (3.65)$$

Donc pour une erreur inférieure à 10%, on doit avoir :

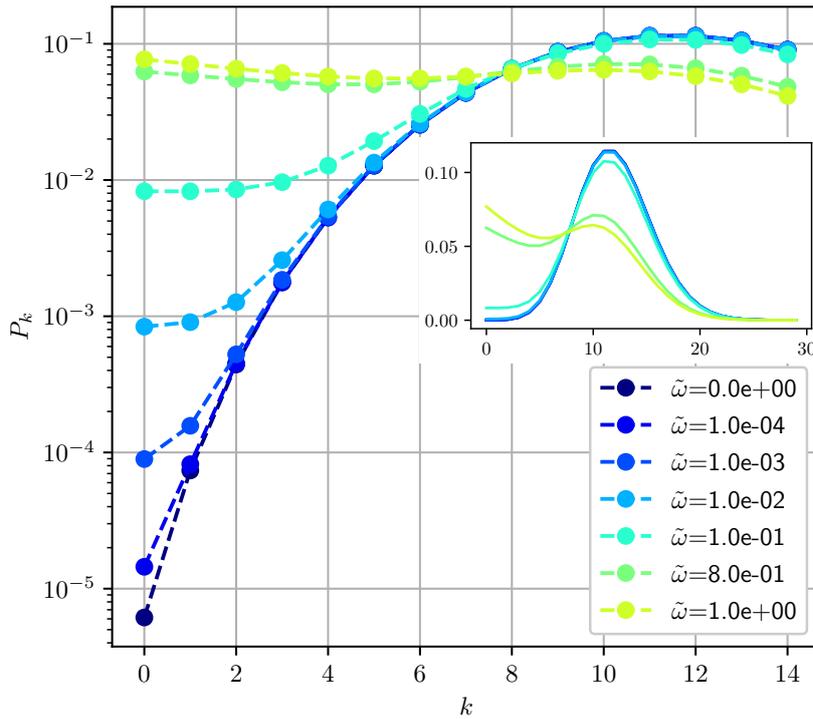
$$\tilde{\omega} \frac{\mathcal{F}(x)}{\mathcal{F}(L)} < \ln(1.1) \approx 0.095. \quad (3.66)$$

Comme  $\tilde{\omega} \frac{\mathcal{F}(x)}{\mathcal{F}(L)}$  est au maximum en  $x = L$ , nous pouvons dire que  $\tilde{\omega} < 0.1$  assure une erreur de moins de 10% sur l'estimation des taux de sauts sans tenir compte de la durée de vie finie (Fig. 3.8).

Cette analyse numérique montre que les travaux précédents estimant les taux de saut dans *S. cerevisiae* à partir de polysomes et avec un modèle qui ne prend pas en compte les durées de vie des ARNm constituent une approximation valide [CIANDRINI et al., 2013; DUC et SONG, 2018]. Précisément, l'erreur d'estimation des taux de saut est de 10% si  $\tilde{\omega}$  est de l'ordre de  $10^{-1}$ , ce qui est un ordre de grandeur supérieur à la valeur typique que nous avons trouvée dans la littérature (Tab. 3.2).

Cependant, un effet est visible sur la densité pour  $\tilde{\omega} \sim 10^{-2}$  et il est donc possible qu'il soit aussi visible sur des profils Ribo-Seq. En biologie, une hypothèse de "rampe d'accélération" est étudiée [VERMA et al., 2019]. Il s'agit d'une augmentation progressive de la vitesse du ribosome sur les premiers codons de l'ARNm. La décroissance de la densité a pu être interprétée comme une augmentation de la vitesse et donc constituer une preuve de cette rampe d'accélération. Nous montrons donc ici que cela peut être confondu avec l'effet de temps de vie fini. Pour conclure sur cette rampe d'accélération, il faudrait estimer plus précisément la durée de vie et la décroissance de la densité polysome des ARNm étudiés.

#### 6.4 Distribution des $k$ -somes



**FIGURE 3.9** – La distribution de population  $P_k$  pour des valeurs finies de  $\tilde{\omega}$  (de 0 à 1) et les autres paramètres fixés :  $\mathcal{F}(L) = 200s$ ,  $p = 0,5s^{-1}$ ,  $L = 100$  codons et  $\alpha = 0,06s^{-1}$ . La valeur  $\langle k \rangle = \tilde{\alpha} = 12$  correspond au maximum de la distribution à  $\tilde{\omega} = 0$ . Dans le graphique principal (échelle lin-log), on peut apprécier l'augmentation de la population d'ARNm vides ( $k = 0$ ) et la faible valeur de  $k$  lorsque  $\tilde{\omega}$  augmente. Dans l'encart (lin-lin), nous pouvons voir que les deux maxima (à  $k = 0$  et  $k = \tilde{\alpha}$ ) sont égaux pour  $\tilde{\omega} \approx 0,8$ .

En fonction des paramètres adimensionnels  $\tilde{\alpha}$  et  $\tilde{\omega}$ , la probabilité  $P_k$  donnée par l'Eq. (3.36) qu'un ARNm soit un  $k$ -some s'écrit

$$P_k = \frac{\tilde{\omega}}{\tilde{\alpha} + \tilde{\omega}} \left( \frac{\tilde{\alpha}}{\tilde{\alpha} + \tilde{\omega}} \right)^k \frac{\gamma(k+1, \tilde{\alpha} + \tilde{\omega})}{k!} + \frac{\tilde{\alpha}^k}{k!} e^{-(\tilde{\alpha} + \tilde{\omega})}. \quad (3.67)$$

A  $\tilde{\omega} = 0$  ou temps de vie infini,  $P_k$  se réduit à une distribution Poissonienne :

$$P_k^\infty = \frac{\tilde{\alpha}^k e^{-\tilde{\alpha}}}{k!}. \quad (3.68)$$

Pour la plupart des régimes biologiquement pertinents,  $\tilde{\omega}$  est petit comparé à  $\tilde{\alpha}$  et il est intéressant d'effectuer une expansion linéaire de  $P_k$  en  $\tilde{\omega}/\tilde{\alpha}$ . Nous utilisons pour cela l'approximation de la fonction  $\gamma$  indiquée en annexe (A.3). Pour  $\tilde{\alpha} \gg 1$  et  $k \ll \tilde{\alpha}$ , on obtient

$$P_k \approx P_k^\infty + \frac{\tilde{\omega}}{\tilde{\alpha}} + \mathcal{O}\left(\frac{\tilde{\omega}}{\tilde{\alpha}}\right)^2, \quad (3.69)$$

Le terme  $P_k^\infty$  augmentant avec  $k$ ,  $\frac{\tilde{\omega}}{\tilde{\alpha}}$  est d'autant plus négligeable que  $k$  est grand. Cela explique que la correction à la Poissonienne  $P_k^\infty$  est importante pour  $k$  assez petit comme on pourra l'observer sur la Fig. 3.9 décrite au paragraphe suivant. Nous précisons qu'en utilisant l'approximation de Stirling

$$k! \approx \sqrt{2\pi k} k^k e^{-k}, \quad (3.70)$$

valide pour  $k > 1$ , on obtient la gaussienne vers laquelle tend  $P_k^\infty$ , soit

$$P_k^\infty \approx \frac{\exp\left[-(k - \tilde{\alpha} + 1/2)^2 / (2\tilde{\alpha})\right]}{\sqrt{2\pi\tilde{\alpha}}} \quad (3.71)$$

où la dernière approximation, valable pour  $\tilde{\alpha} \gg 1$ , provient d'une simple approximation du point col.

Sur la Fig. 3.9, nous traçons la distribution de la population de  $k$ -somes  $P_k$  de l'équation (3.67). Nous comparons les distributions obtenues avec des paramètres de valeurs qui correspondent au contexte des cellules vivantes (cf. Tab. 3.1) :  $\mathcal{T}(L) = 200s$ ,  $p = 0,5s^{-1}$ ,  $L = 100$  codons,  $\alpha = 0,06s^{-1}$  ( $\tilde{\alpha} = 12$ ) et différentes valeurs de  $\tilde{\omega}$  entre 0 (aucune dégradation), et 1 (passage au régime de forte dégradation). Dans le graphique principal, nous utilisons une échelle log-lin afin de voir plus clairement l'évolution de la population  $k$ -some avec de petits  $k$ . Dans l'encadré, les mêmes données sont représentées sur une échelle lin-lin et avec une gamme plus large de  $k$ . Pour  $\tilde{\omega} \leq 10^{-1}$ , nous observons une distribution gaussienne presque symétrique, comme indiqué dans (3.71) pour  $\tilde{\omega} = 0$ , de la population centrée autour d'un maximum proche de la valeur stationnaire  $\langle k \rangle = \tilde{\alpha} = 12$ .

Pour  $\tilde{\omega} = 0.8$ , on observe que le maximum à  $k = 0$  atteint la même hauteur que le pic local secondaire à  $k \approx 12$ . Nous donnons maintenant une explication heuristique de l'inversion des maxima de  $P_k$  observés sur la Fig. 3.9 lorsque  $\tilde{\omega}$  augmente et atteint une valeur de croisement  $\tilde{\omega}_{co}$  pour laquelle un nouveau maximum local à  $k = 0$  devient égal au maximum (dépendant de la dégradation) près de  $k = \tilde{\alpha}$ . La position du maximum de  $P_k$  pour une dégradation finie, mais pas trop élevée, ne se déplace pas beaucoup par rapport à la valeur de dégradation nulle  $k \approx \tilde{\alpha}$ , et nous pouvons donc estimer  $\tilde{\omega}_{co}$  en égalisant  $P_{\tilde{\alpha}}$  avec  $P_0 \approx \tilde{\omega}/\tilde{\alpha} \ll 1$ . Cette approximation conduit à une relation entre  $\tilde{\omega}_{co}$  et  $\tilde{\alpha}$  (voir annexe 1.3), valable pour  $\tilde{\alpha} > 2$ , montrant que  $\tilde{\omega}_{co} \simeq 1$  n'augmente que lentement avec  $\tilde{\alpha}$  :

$$\tilde{\omega}_{co} \approx \ln\left(\sqrt{\frac{\tilde{\alpha}}{2\pi}} + \frac{1}{2}\right). \quad (3.72)$$

Pour  $\tilde{\alpha} \gg 1$ , le résultat  $\tilde{\omega}_{co} \simeq 1$  est ce que l'on aurait naïvement attendu puisque les ARNm ne vivent pas assez longtemps pour que les ribosomes puissent atteindre la sortie. C'est le même seuil pour lequel la durée de vie finie des ARNm a un impact important sur le profil de densité des polysomes et par conséquent sur la valeur moyenne de  $k$  dans l'Eq.(3.54) : globalement, l'échange des maxima de  $P_k$  est cohérent avec le croisement des polysomes à  $\tilde{\omega} \approx 1$ , une valeur pour laquelle la densité globale des ribosomes est diminuée par la dégradation, car les  $k$ -somes avec  $k \ll \tilde{\alpha}$  commencent à représenter une fraction significative de la population globale.

Il serait extrêmement intéressant de comparer ces profils théoriques des  $P_k$  avec une mesure expérimentale de la distribution des  $k$ -somes afin d'en déduire les variations de la cinétique de traduction par rapport à celle du modèle balistique. Nous observons en particulier que pour des

valeurs de  $\tilde{\omega}$  de l'ordre de  $10^{-3}$  et  $\tilde{\alpha} = 12$ ,  $P_1$  est presque de l'ordre de  $10^{-4}$ . Les expériences Ribo-Seq pouvant se faire sur des milliards d'ARNm, la proportion de monosomes doit être de  $10^5$  tout gène confondu.

## 6.5 Profils de densités $k$ -somes et polysome à vitesse constante

### 6.5.1 Analyse générale

Dans cette sous-section, nous effectuons une analyse paramétrique des  $k$ -somes et discutons de leur sensibilité à la dégradation de l'ARNm. Par comparaison avec l'effet de la dégradation de la densité polysome, nous allons définir trois régimes de dégradation.

Dans la Fig. 3.10, colonne de gauche, nous traçons  $\rho_k(x)$  obtenu à partir du modèle balistique pour  $k = 1$  à 4 et pour une valeur uniforme du taux d'élongation  $p(x) = p$ . Les profils en traits pleins sont donnés par l'équation (3.45) avec  $L = 100$  codons,  $\alpha = 0,06s^{-1}$ ,  $p = 0,5s^{-1}$  et donc  $\mathcal{T}(L) = 200s$ . Dans la colonne de gauche, nous avons tracé les mêmes densités, mais normalisées à 1 afin de mieux pouvoir les comparer entre elles.

En développant l'Eq. (3.45) au premier ordre en  $\omega$  pour  $k$  fixe, on obtient

$$\rho_k(x) = \frac{k}{\tilde{p}(x)} \left[ 1 + \frac{\tilde{\omega}}{kP_k^\infty} \left( \frac{\gamma(k, \tilde{\alpha})}{(k-1)!} - \frac{\gamma(k+1, \tilde{\alpha})}{\tilde{\alpha}(k-1)!} - \frac{\gamma(k, \tilde{\alpha} \frac{\mathcal{T}(x)}{\mathcal{T}(L)})}{(k-1)!} + \mathcal{O}\left(\frac{\tilde{\omega}}{\tilde{\alpha}}\right) \right) \right], \quad (3.73)$$

où nous rappelons que  $P_k^\infty = \frac{\tilde{\alpha}^k e^{-\tilde{\alpha}}}{k!}$ .

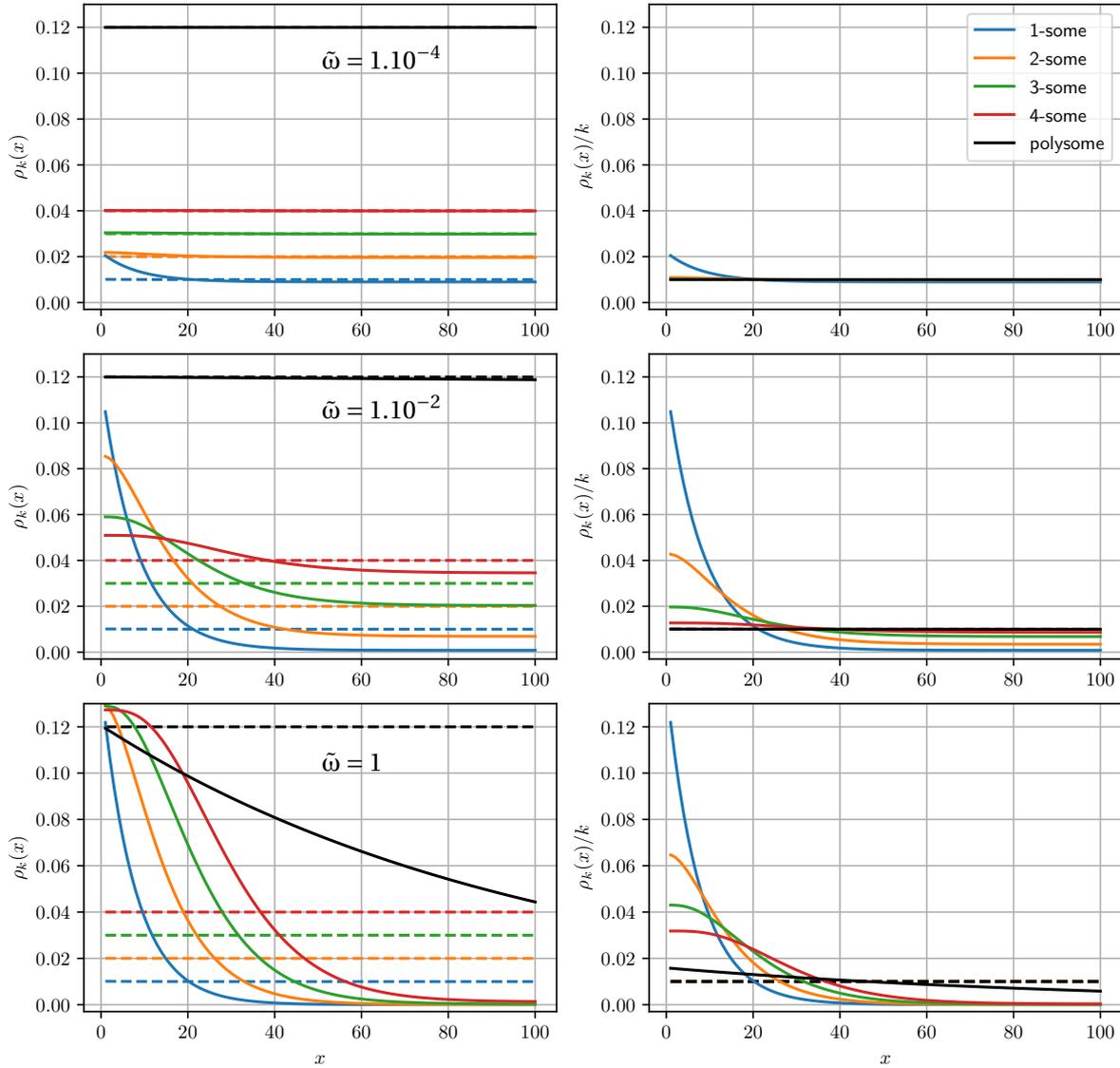
Dans chaque graphique, les lignes pointillées représentent les profils pour  $\tilde{\omega} = 0$  qui sont plats avec une magnitude  $\rho_k^\infty = k/L$  d'après l'Eq. (3.73). La conséquence est que lorsque  $\tilde{\omega} = 0$ , tous les profils de  $k$ -some sont donc linéairement dépendants. Le profil des polysomes dans le cas où  $\tilde{\omega} = 0$  est également plat et de magnitude  $\rho_{ave}^\infty = \alpha/p$ .

Les valeurs de  $\tilde{\omega}$  utilisées dans les panneaux successifs augmentent d'un facteur  $10^2$  (de haut en bas), elles sont  $\tilde{\omega} = 10^{-4}$  ( $\omega = 5.10^{-7}s^{-1}$ , durée de vie de 555h),  $\tilde{\omega} = 10^{-2}$  ( $\omega = 5.10^{-5}s^{-1}$ , durée de vie de 5h) et  $\tilde{\omega} = 1$  ( $\omega = 5.10^{-3}s^{-1}$ , durée de vie de 3min), respectivement. Dans le premier cas, tous les profils de densité sont proches de la limite  $\tilde{\omega} = 0$  et on remarque à peine une petite déviation pour les monosomes, au début du filament. Ainsi, dans ce régime de faible dégradation, nous ne nous attendons pas à ce que les  $k$ -somes apportent plus d'information que les polysomes d'une expérience Ribo-Seq habituelle. Dans le second cas, nous ne constatons aucun effet sur le profil des polysomes, mais un effet clair sur les  $k$ -somes. Ce découplage fournira des informations supplémentaires sur les paramètres du modèle. Il est intéressant de noter que ce régime intermédiaire ( $\tilde{\omega} \sim 10^{-2}$ ) tombe dans la gamme des valeurs biologiques. Dans le dernier cas, la dégradation est élevée ( $\tilde{\omega} = 1$ ). Tous les profils de densité sont modifiés, même celui du polysome. Ce régime n'est probablement pas pertinent pour le processus de traduction, mais nous le décrivons par souci d'intérêt théorique et d'exhaustivité.

En conclusion, il existe trois régimes de dégradation : (i) une faible dégradation, où  $k$ -somes et polysomes sont linéairement liés. Dans ce régime, les  $k$ -somes ne devraient donc pas fournir plus d'informations que les polysomes ; (ii) un régime intermédiaire, dans lequel certains gènes sont susceptibles de tomber, où le profil des polysomes est plat, mais où les profils des  $k$ -somes ne sont plus linéairement couplés. (iii) un régime de forte dégradation qui n'est pas pertinent pour la traduction et dans lequel tous les profils sont modifiés par la dégradation.

Les transitions entre ces régimes sont, en première approximation, pilotées par  $\tilde{\omega}$  pour les polysomes et par  $\tilde{\omega}/(kP_k^\infty)$  pour les  $k$ -somes d'après l'Eq (3.73) où ce terme apparaît en facteur de la correction à  $\rho_k^\infty$ . Pour les valeurs biologiques typiques  $\tilde{\alpha} \in [5, 15]$ , cela signifie que le changement de régime pour le monosome se produit pour des valeurs de  $\tilde{\omega}$  qui sont de 2 à 6 ordres de grandeurs plus petites que pour les polysomes puisqu'alors  $P_1^\infty = \tilde{\alpha}e^{-\tilde{\alpha}} \in [5.10^{-6}, 3.10^{-2}]$ .

Dans la prochaine sous-section 6.6, nous utiliserons une approche légèrement différente pour dériver une condition de changement de régime similaire pour les  $k$ -somes. Mais avant cela, on va s'intéresser au point de croisement spécifique de la densité à durée de vie infinie  $\rho_k^\infty$  et  $\rho_k(x)$ .



**FIGURE 3.10** – Les densités  $\rho_k$  de  $k$ -somes ( $k = 1$  à  $4$ ) sont représentées en fonction de la coordonnée génomique  $x$  (en unités de codon) pour  $\alpha = 0,06s^{-1}$  et  $p = 1/2s^{-1}$  pour un filament de taille  $L = 100$  sites (en unités de codon). Des lignes complètes sont obtenues pour des valeurs finies croissantes de  $\omega = 5.10^{-7}s^{-1}$  ( $\tilde{\omega} = 10^{-4}$ ),  $\omega = 5.10^{-5}s^{-1}$  ( $\tilde{\omega} = 10^{-2}$ ) et  $\omega = 5.10^{-3}s^{-1}$  ( $\tilde{\omega} = 1$ ) de haut en bas tandis que les lignes pointillées sont obtenues pour  $\omega = 0s^{-1}$ . Sur le panel de droite ce sont les mêmes profils, mais normalisés à 1 afin de faire ressortir la différence entre les densités  $k$ -somes entre elles ainsi qu’avec la densité polysome. On observe que considérer une durée de vie finie des ARNm entraîne le découplage de ces différentes densités d’autant plus que  $\tilde{\omega}$  est grand. Ce découplage est dû à une diminution de la sensibilité à l’effet de durée de vie finie avec l’augmentation de  $k$  et entre les  $k$ -somes et les polysomes.

### 6.5.2 Point de croisement entre $\rho_k(x)$ et $\rho_k^\infty$

Contrairement à ce qui se passe pour les polysomes (voir Fig. 3.8), la densité sur les sites les plus proches de l’entrée des ribosomes est plus sensible que sur les sites plus éloignés le long de la séquence. Comme on le voit sur la Fig. 3.10,  $\rho_k(x)$  doit être supérieure à  $\rho_k^\infty$  pour  $x < x_k^{co}$  et inférieure à  $\rho_k^\infty$  pour  $x > x_k^{co}$ , où  $x_k^{co}$  est la position de croisement entre  $\rho_k(x)$  et  $\rho_k^\infty$  dépendante de  $k$ .

Nous allons considérer plus en détail la densité des monosomes ( $k = 1$ ), car elle est la plus sensible à l’effet de temps de vie fini et aussi la plus simple des densités  $k$ -somes ( $k > 0$ ). A partir

de l'équation (3.45), nous obtenons

$$\rho_1(x) = \frac{(\tilde{\omega} + \tilde{\alpha}) (\tilde{\alpha} + \tilde{\omega} e^{(\tilde{\omega} + \tilde{\alpha})[1 - \mathcal{F}(x)/\mathcal{F}(L)]})}{\tilde{p}(x) [\tilde{\alpha}(\tilde{\omega} + \tilde{\alpha}) + \tilde{\omega}(e^{\tilde{\omega} + \tilde{\alpha}} - 1)]}, \quad (3.74)$$

et d'après l'équation (3.73), son expansion au premier ordre en  $\omega$  s'écrit

$$\rho_1(x) = \frac{1}{\tilde{p}(x)} \left[ 1 - \frac{\tilde{\omega}}{\tilde{\alpha} P_1^\infty} \left( 1 - \tilde{\alpha} e^{-\tilde{\alpha} \frac{\mathcal{F}(x)}{\mathcal{F}(L)}} - e^{-\tilde{\alpha}} + \mathcal{O}\left(\frac{\tilde{\omega}}{\tilde{\alpha}}\right) \right) \right]. \quad (3.75)$$

En égalisant  $\rho_1(x)$  à cet ordre et  $\rho_1^\infty(x)$  nous trouvons  $x_1^{\text{co}}$  comme la solution de

$$\frac{\mathcal{F}(x_1^{\text{co}})}{\mathcal{F}(L)} = \frac{x_1^{\text{co}}}{L} \approx \frac{1}{\tilde{\alpha}} \ln\left(\frac{\tilde{\alpha}}{1 - e^{-\tilde{\alpha}}}\right), \quad (3.76)$$

où l'égalité est valable pour un taux d'élongation  $p$  uniforme. Ce résultat qui montre que  $x_1^{\text{co}}$  diminue avec l'augmentation de  $\tilde{\alpha}$ , et est indépendante de la dégradation comme on peut le vérifier sur la Fig. 3.10. Dans ce cas simple mais important du monosome, nous pouvons illustrer la différence de sensibilité à la dégradation due à la position en écrivant  $\rho_1(x)$  à  $x = 0$  et  $L$  de manière à faire ressortir le signe de la correction au premier ordre :

$$\rho_1(0) = \frac{1}{\tilde{p}(0)} \left[ 1 + \frac{\tilde{\omega}}{P_1^\infty} \left( 1 - \frac{1}{\tilde{\alpha}} (1 - e^{-\tilde{\alpha}}) + \mathcal{O}\left(\frac{\tilde{\omega}}{\tilde{\alpha}}\right) \right) \right] \quad (3.77)$$

et

$$\rho_1(L) = \frac{1}{\tilde{p}(L)} \left[ 1 - \frac{\tilde{\omega}}{\tilde{\alpha} P_1^\infty} \left( 1 - e^{-\tilde{\alpha}} (\tilde{\alpha} - 1) + \mathcal{O}\left(\frac{\tilde{\omega}}{\tilde{\alpha}}\right) \right) \right]. \quad (3.78)$$

D'après cette expression, nous voyons que pour les monosomes, le passage entre les régimes de faible et de forte dégradation à proximité du début de l'ARNm dépend du paramètre  $\tilde{\omega}/P_1^\infty$  et est positif, tandis que pour qu'à la fin de l'ARNm, il est négatif et plus lent puisqu'il dépend de  $\tilde{\omega}/(\tilde{\alpha} P_1^\infty)$ .

## 6.6 Rapport entre les densités des ARNm à l'état transitoire et des ARNm à l'état stationnaire

Pour caractériser l'ensemble de la population d'ARNm mesurée lors d'une expérience Riboseq, nous les divisons en deux groupes distincts en fonction de leur âge  $a$  : les ARNm en régime transitoire de remplissage ( $a \leq \mathcal{F}(L)$ ) et les ARNm en régime stationnaire ( $a \geq \mathcal{F}(L)$ ). Cette distinction a déjà été faite dans la description analytique du modèle. Nous écrivons donc la densité  $k$ -some comme la somme de deux contributions :

$$\rho_k(x) = \mathcal{F}_k(x) + \mathcal{S}_k(x), \quad (3.79)$$

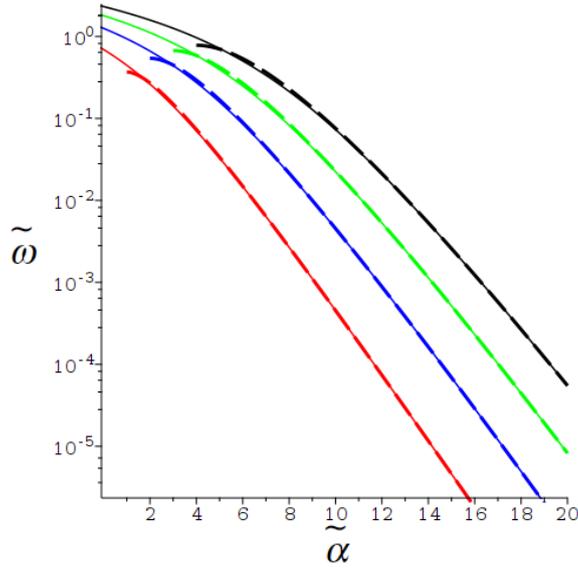
où  $\mathcal{F}_k(x)$  est la densité des ARNm dans le régime transitoire de remplissage

$$\mathcal{F}_k(x) = \frac{\tilde{\omega}}{P_k \tilde{p}(x)} \left( \frac{\tilde{\alpha}}{\tilde{\alpha} + \tilde{\omega}} \right)^k \frac{\gamma(k, (\tilde{\alpha} + \tilde{\omega})) - \gamma(k, (\tilde{\alpha} + \tilde{\omega}) \frac{\mathcal{F}(x)}{\mathcal{F}(L)})}{(k-1)!}, \quad (3.80)$$

et où  $\mathcal{S}_k(x)$  est la densité des ARNm dans leur état stationnaire

$$\mathcal{S}_k(x) = \frac{\tilde{\alpha}^k}{P_k \tilde{p}(x)} \frac{e^{-(\tilde{\alpha} + \tilde{\omega})}}{(k-1)!}. \quad (3.81)$$

$\mathcal{S}_k(x)$  n'ayant pas d'autre dépendance en  $x$  que le taux d'élongation, la majeure partie des effets de la durée de vie finie de l'ARNm est supportée par le terme  $\mathcal{F}_k(x)$ . Nous choisissons donc de



**FIGURE 3.11** – Courbes de croisement  $\mathcal{R}_k(0) = 1$  pour  $k$  certaines densités (lignes solides :  $k = 1$  (rouge),  $k = 2$  (bleu),  $k = 3$  (vert),  $k = 4$  (noir)). Les lignes pointillées de même couleur sont les expressions asymptotiques données par (3.84) à partir de  $\tilde{\alpha} = k$ .

quantifier les effets de la dégradation des ARNm sur les  $k$ -somes en définissant un rapport ARNm transitoire/stationnaire comme suit,

$$\mathcal{R}_k(x) = \frac{\mathcal{F}_k(x)}{\mathcal{S}_k(x)} = \frac{\tilde{\omega}}{(\tilde{\alpha} + \tilde{\omega})^k} [\gamma(k, \tilde{\alpha} + \tilde{\omega}) - \gamma(k, (\tilde{\alpha} + \tilde{\omega})\mathcal{F}(x)/\mathcal{F}(L))] e^{\tilde{\alpha} + \tilde{\omega}}. \quad (3.82)$$

Pour se débarrasser d'une dépendance complexe de  $x$ , on utilise sa valeur maximale obtenue à  $x = 0$ ,

$$\mathcal{R}_k(0) = \tilde{\omega} \frac{\gamma(k, \tilde{\omega} + \tilde{\alpha})}{(\tilde{\omega} + \tilde{\alpha})^k} e^{\tilde{\omega} + \tilde{\alpha}}. \quad (3.83)$$

Or, le passage entre les régimes de faible et de forte influence de la dégradation sur  $\rho_k(x)$  est déterminé par des valeurs de  $\mathcal{R}_k(0)$  qui sont de l'ordre de l'unité. Lorsque  $\tilde{\omega} + \tilde{\alpha} \gg 1$  et  $\tilde{\alpha} \gg \tilde{\omega}$ , une hypothèse raisonnable pour les systèmes biologiques, la condition de changement de régime pour les  $k$ -somes se simplifie en

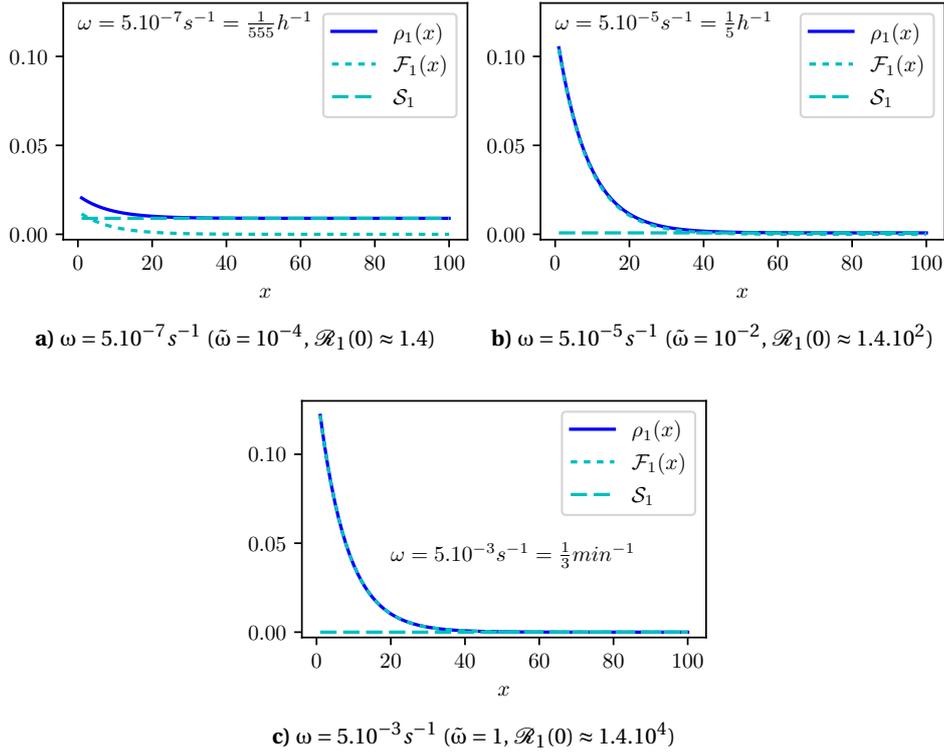
$$\tilde{\omega}_c \sim \mathcal{R}_k(0) \frac{\tilde{\alpha}^k e^{-\tilde{\alpha}}}{(k-1)!}, \quad \tilde{\alpha} > k, \quad (3.84)$$

où nous avons utilisé l'approximation de la fonction gamma donnée par l'Eq. (A.5). Comme indiqué, cette expression est typiquement valable pour  $\tilde{\alpha} > k$ . Les courbes de croisement pour des densités de  $k$  caractérisées par  $\mathcal{R}_k(0) = 1$  sont représentées sur la Fig. 3.11 par des lignes pleines. Les expressions asymptotiques, commençant précisément aux valeurs  $\tilde{\alpha} = k$ , et affichées sous forme de lignes pointillées de même couleur, confirment leur domaine de validité. Notons que ces expressions sont valables jusqu'à des valeurs relativement élevées puisque pour  $\tilde{\alpha} = k$ ,  $\tilde{\omega}_c \approx \mathcal{R}_k(0) \sqrt{k/2\pi}$ .

On observe sur la Fig. 3.11 que, plus  $k$  est faible, plus les densités de  $k$ -somes sont sensibles à la dégradation de l'ARNm. La densité de monosomes (courbe rouge) est de loin la plus sensible à la dégradation. Son rapport exact se calcule facilement comme

$$\mathcal{R}_1(0) = \frac{\tilde{\omega}}{\tilde{\omega} + \tilde{\alpha}} (e^{\tilde{\omega} + \tilde{\alpha}} - 1) \approx \frac{\tilde{\omega}}{\tilde{\alpha} e^{-\tilde{\alpha}}}, \quad \text{pour } \tilde{\alpha} + \tilde{\omega} \gg 1, \quad (3.85)$$

et permettra de définir la limite entre le régime où les densités sont linéairement couplées et le régime où elles commencent à sentir l'effet de durée de vie finie qui les découple. Dans la Fig. 3.12, nous illustrons la décomposition de  $\rho_1(x)$  pour le modèle balistique avec taux d'élongation  $p$  constant. Les trois régimes affichés dans cette figure correspondent aux trois cas de la



**FIGURE 3.12** – Contributions transitoire et stationnaire de la densité de monosomes  $\rho_1(x)$  pour les régimes illustrés dans la Fig. 3.10 ( $\alpha = 0.06s^{-1}$  et  $p = 1/2s^{-1}$  pour un filament de taille  $L = 100$  codons). On voit que la contribution du terme  $\mathcal{F}_1(x)$  à la densité est majoritaire quand  $\mathcal{R}_1(0) \gg 1$ .

Fig. 3.10. Pour ce modèle balistique à  $p$  constant, la contribution non linéaire à la densité de monosome provient uniquement de  $\mathcal{F}_1(x)$ . Lorsque l'effet de durée de vie finie est relativement faible,  $\mathcal{R}_1(0) \sim 1$ , la densité des monosomes dans leur régime transitoire de remplissage,  $\mathcal{F}_1(x)$ , correspond à peine à celle de ceux dans leur état stationnaire de remplissage,  $\mathcal{S}_1$ , au voisinage de l'entrée de l'ARNm ( $x = 0$ ) et reste beaucoup plus faible sur le reste du filament. En revanche, lorsque l'effet de durée de vie finie est élevé,  $\mathcal{R}_1(0) \gg 1$ ,  $\mathcal{F}_1(x)$  est grand par rapport à  $\mathcal{S}_1$  sur vers l'entrée de l'ARNm, et la densité  $\rho_1(x)$  est bien approximée par  $\mathcal{F}_1(x)$  presque partout.

Afin d'illustrer la pertinence de ce rapport pour déterminer l'état du système, nous présentons une comparaison de profils avec différents paramètres mais le même  $\mathcal{R}_1(0)$  dans la Fig. A.2.

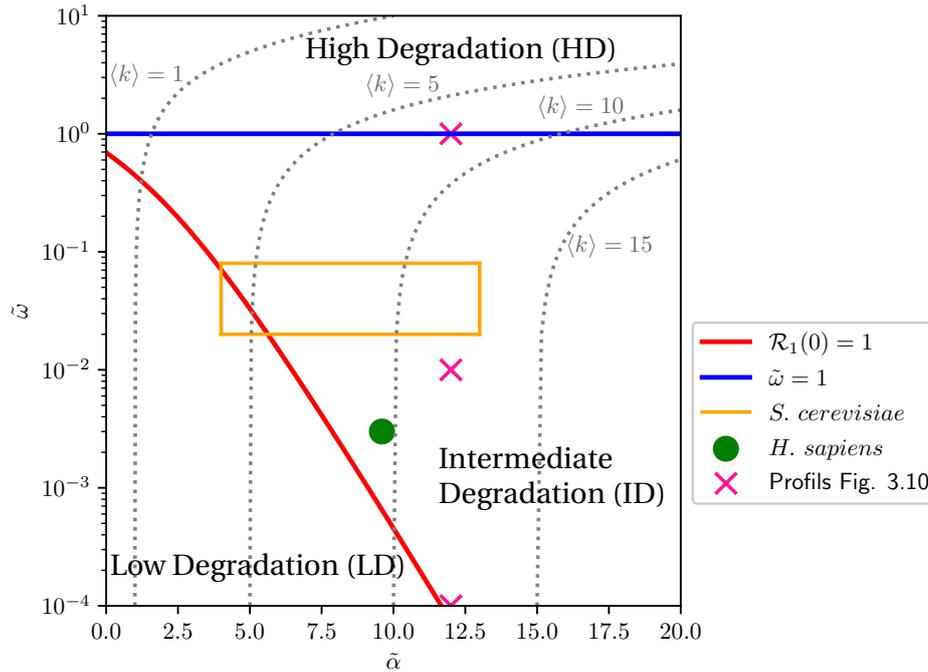
## 7 Diagramme de phase

Les résultats de l'analyse de l'effet de durée de vie finie sur la distribution des  $k$ -somes et les densités de ribosomes sur les  $k$ -somes de la section précédente nous amènent à définir un diagramme de phase du modèle balistique de la traduction génomique dans le plan  $(\tilde{\alpha}, \tilde{\omega})$  (Fig. 3.13). Trois phases sont définies : la phase de faible dégradation (LD) où ni les profils de densité polysome, ni  $k$ -somes ne sont affectés par l'effet de durée de vie finie, la phase de dégradation intermédiaire (ID) où seuls les profils de densité  $k$ -somes sont affectés par l'effet de durée de vie finie, et la phase de haute dégradation (HD) où les profils de densité polysome et  $k$ -somes sont affectés. Les phases LD et ID sont séparées par notre indicateur de l'effet de durée de vie finie sur les monosomes,  $\mathcal{R}_1(0)$  car le monosome est le plus sensible à cet effet.  $\mathcal{R}_1(0) = 1$  signifie que la population des monosomes en état stationnaire et des monosomes en état transitoire ont une contribution égale à la densité au premier codon, ce qui signifie que l'effet de temps de vie fini sur la densité monosome devient non négligeable au premier site et donc sur la densité, puisque c'est toujours le premier site qui est le plus impacté. Les phases ID et HD sont séparées par la ligne  $\tilde{\omega} = 1$  puisque

nous avons montré dans la section 6.2 que c'est la limite entre une densité polysome subissant un effet de durée de vie finie négligeable et une densité polysome subissant un effet de durée de vie finie non négligeable.

Les lignes pointillées sont les lignes à nombre de ribosomes moyen  $\langle k \rangle = \tilde{\alpha}/\tilde{\omega}(1 - e^{-\tilde{\omega}})$  constant. Lorsque la dégradation augmente, nous observons un écart par rapport à  $\tilde{\alpha}$  qui se produit autour de la limite entre le régime intermédiaire et le régime haute dégradation.

Nous avons ajouté dans ce diagramme, les résultats de notre recherche bibliographique sur les mesures déjà existantes de  $\tilde{\alpha}$  et  $\tilde{\omega}$  pour montrer que le régime de dégradation intermédiaire est celui qui semble prédominer en biologie, même si les valeurs sont proches de la limite avec le régime LD.



**FIGURE 3.13** – Diagramme de phase du modèle balistique dans le plan  $(\tilde{\alpha}, \tilde{\omega})$ . Le diagramme de phase est divisé en trois régimes de dégradation avec les lignes d'équation  $\tilde{\omega} = 1$  et  $\mathcal{R}_1(0) = 1$ . Les valeurs sont extraites de la littérature (cf Tab. 3.2). D'après ces mesures, qui sont les seules que nous ayons trouvées pour l'instant, on peut dire que le régime de dégradation intermédiaire est le régime du vivant.

## 8 Conclusion et perspectives

Dans ce chapitre, nous avons étudié un système inspiré par un dispositif expérimental où les ARNm sont séparés en fonction du nombre de ribosomes qui leur sont attachés avant d'être séquencés par l'expérience Ribo-Seq que nous avons présenté au premier chapitre. Pour ce système encore jamais étudié à notre connaissance, nous avons développé des modèles de transport des ribosomes sur un ensemble d'ARNm à temps de vie fini.

Dans un premier temps, nous avons montré que la résolution exacte du modèle TASEP décrit au chapitre précédent est facilement intégrable sur un ensemble d'ARNm à temps de vie fini. Le calcul exact des densités étant déraisonnablement long pour des réseaux de l'ordre de 100 codons, taille typique des ARNm (Tab. 3.1), ce modèle pourra être utilisé pour étudier des cas particuliers de petits ARNm.

Dans un deuxième temps, nous avons simplifié le processus de transport afin de mener une étude exhaustive de l'effet de temps de vie fini sur les grandeurs physiques du système. Nous avons développé un modèle dit "ballistique" car les ribosomes sont modélisés par des particules ponctuelles qui se déplacent de manière déterministe. En étudiant le nombre moyen de ribosomes  $\langle k \rangle$ ,

ainsi que les densités de polysomes et de  $k$ -somes, nous caractérisons le système en identifiant trois régimes qui émergent d'une analyse paramétrique du modèle balistique. Nous arrivons à l'image globale suivante : dans le régime de faible dégradation (LD), l'effet de la durée de vie finie de l'ARNm sur  $\langle k \rangle$  et les densités est négligeable ; dans le régime de dégradation intermédiaire (ID), la durée de vie finie de l'ARNm a un fort impact sur les densités  $k$ -somes, mais pas sur les densités polysomes et sur  $\langle k \rangle$  ; dans le régime de dégradation élevée (HD), toutes les densités, ainsi que  $\langle k \rangle$ , sont affectées par la durée de vie finie de l'ARNm. Le régime HD est délimité par  $\bar{\omega} \equiv \omega \mathcal{T}(L) = 1$ . Cette limite ne correspond probablement à aucun régime biologique habituel, ce qui nous amène à la conclusion que les densités de polysomes ne devraient jamais être affectées par la durée de vie finie de l'ARNm (sauf peut-être dans des conditions rares ou pathologiques).

L'étude de l'effet de durée de vie finie aboutie au diagramme de phase du modèle balistique dans le plan  $(\bar{\alpha}, \bar{\omega})$ , où les phases sont les trois régimes : LD, ID et HD (Fig. 3.13). A partir des paramètres trouvés dans la littérature, nous montrons que différentes espèces vivantes (la levure avec *S. cerevisiae*, les bactéries avec *E. coli* et les mammifères avec *M. musculus* et *H. sapiens*) tombent dans le régime ID. Par conséquent, ce régime s'avère être le régime biologique.

Dans ce cadre, nous pouvons étendre le modèle balistique à une variété d'effets physiques potentiellement importants tels que les ressources finies et la diffusion des ribosomes [DAULOUDET et al., 2021] et des effets biologiques tels que le biais d'usage des codons [PAULET et al., 2017] et l'ARNm polycistronique [KUSHNER, 2018]. D'un point de vu physique il pourrait aussi être intéressant de comparer le modèle TASEP et le modèle balistique afin d'étudier l'effet de l'exclusion entre particules sur les densités déjà affectées par l'effet du temps de vie fini. Il faut cependant noter que la particule étendue n'est pas considérée alors qu'elle engendrerait une exclusion plus proche de la réalité biologique. Une autre possibilité serait alors d'utiliser des simulations de Monte Carlo d'un ensemble d'ARNm modélisé par le modèle  $\ell$ -TASEP inhomogène. Nous précisons aussi que nous avons réfléchi à intégrer sur un ensemble de réseaux à temps de vie fini le modèle  $\ell$ -TASEP avec sa formulation markovienne et de champ moyen et le modèle  $\ell$ -TAESEP, mais que les résultats n'ont pas été concluants.

## 9 Références

- «National Center for Biotechnology Information», URL <https://www.ncbi.nlm.nih.gov/>. 60
- BERNSTEIN, J. A., A. B. KHODURSKY, P.-H. LIN, S. LIN-CHAO et S. N. COHEN. 2002, «Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays», *Proceedings of the National Academy of Sciences*, vol. 99, n° 15, doi :10.1073/pnas.112318199, p. 9697–9702, ISSN 0027-8424, 1091-6490. URL <https://www.pnas.org/content/99/15/9697>, publisher : National Academy of Sciences Section : Biological Sciences. 60
- CIANDRINI, L., I. STANSFIELD et M. C. ROMANO. 2013, «Ribosome Traffic on mRNAs Maps to Gene Ontology : Genome-wide Quantification of Translation Initiation Rates and Polysome Size Regulation», *PLOS Computational Biology*, vol. 9, n° 1, doi :10.1371/journal.pcbi.1002866, p. e1002866, ISSN 1553-7358. URL <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002866>. 60, 63
- COX, D. R. 1970, *Renewal theory*, M. S. Bartlett, F.R.S (Methuen & Co. LTD, Science Paperbacks). 50
- DALEY, D. J., D. VERE-JONES et D. VERE-JONES. 2008, *An Introduction to the Theory of Point Processes*, Springer New York : Springer e-books, New York, NY, Etats-Unis d'Amérique, ISBN 978-0-387-49835-5. ISSN : 2297-0398. 57
- DAULOUDET, O., I. NERI, J.-C. WALTER, J. DORIGNAC, F. GENIET et A. PARMEGGIANI. 2021, «Modelling the effect of ribosome mobility on the rate of protein synthesis», *The European Physical Journal E*, vol. 44, n° 2, doi :10.1140/epje/s10189-021-00019-8, p. 19, ISSN 1292-895X. URL <https://doi.org/10.1140/epje/s10189-021-00019-8>. 71

- DUC, K. D. et Y. S. SONG. 2018, «The impact of ribosomal interference, codon usage, and exit tunnel interactions on translation elongation rate variation», *PLOS Genetics*, vol. 14, n° 1, doi : 10.1371/journal.pgen.1007166, p. e1007 166, ISSN 1553-7404. URL <https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1007166>. 63
- ENGEL, A. 2011, *Processus aléatoires pour les débutants*, Cassini, Paris, France, ISBN 978-2-84225-090-4. ISSN : 2259-8030. 55
- INGOLIA, N., L. LAREAU et J. WEISSMAN. 2011, «Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes», *Cell*, vol. 147, n° 4, doi : 10.1016/j.cell.2011.10.002, p. 789–802, ISSN 0092-8674. URL <http://www.sciencedirect.com/science/article/pii/S0092867411011925>. 60
- KARPINETS, T. V., D. J. GREENWOOD, C. E. SAMS et J. T. AMMONS. 2006, «RNA :protein ratio of the unicellular organism as a characteristic of phosphorous and nitrogen stoichiometry and of the cellular requirement of ribosomes for protein synthesis», *BMC Biology*, vol. 4, doi : 10.1186/1741-7007-4-30, p. 30, ISSN 1741-7007. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1574349/>. 60
- KUSHNER, S. R. 2018, «Messenger RNA in Prokaryotes», dans *eLS*, John Wiley & Sons, Ltd, ISBN 978-0-470-01590-2, p. 1–11, doi :10.1002/9780470015902.a0000874.pub4. URL <http://onlinelibrary.wiley.com/doi/abs/10.1002/9780470015902.a0000874.pub4>, \_eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470015902.a0000874.pub4>. 71
- NAGAR, A., A. VALLERIANI et R. LIPOWSKY. 2011, «Translation by Ribosomes with mRNA Degradation : Exclusion Processes on Aging Tracks», *Journal of Statistical Physics*, vol. 145, n° 5, doi : 10.1007/s10955-011-0347-z, p. 1385–1404, ISSN 0022-4715. WOS :000297593900013. 54
- PAULET, D., A. DAVID et E. RIVALS. 2017, «Ribo-seq enlightens codon usage bias», *DNA research : an international journal for rapid publication of reports on genes and genomes*, vol. 24, n° 3, doi :10.1093/dnares/dsw062, p. 303–210, ISSN 1756-1663. 71
- SHAROVA, L. V., A. A. SHAROV, T. NEDOREZOV, Y. PIAO, N. SHAIK et M. S. KO. 2009, «Database for mRNA Half-Life of 19 977 Genes Obtained by DNA Microarray Analysis of Pluripotent and Differentiating Mouse Embryonic Stem Cells», *DNA Research*, vol. 16, n° 1, doi :10.1093/dnares/dsn030, p. 45–58, ISSN 1340-2838. URL <https://doi.org/10.1093/dnares/dsn030>. 50, 60
- SINGH, U. N. 1969, «Polyribosomes and unstable messenger RNA : A stochastic model of protein synthesis», *Journal of Theoretical Biology*, vol. 25, n° 3, doi :10.1016/S0022-5193(69)80032-9, p. 444–460, ISSN 0022-5193. URL <https://www.sciencedirect.com/science/article/pii/S0022519369800329>. 54
- TRÖSEMEIER, J.-H., S. RUDORF, H. LOESSNER, B. HOFNER, A. REUTER, T. SCHULENBORG, I. KOCH, I. BEKEREDJIAN-DING, R. LIPOWSKY et C. KAMP. 2019, «Optimizing the dynamics of protein expression», *Scientific Reports*, vol. 9, n° 1, doi :10.1038/s41598-019-43857-5, ISSN 2045-2322. URL <http://www.nature.com/articles/s41598-019-43857-5>. 60
- VALLERIANI, A., Z. IGNATOVA, A. NAGAR et R. LIPOWSKY. 2010, «Turnover of messenger RNA : Polysome statistics beyond the steady state», *EPL (Europhysics Letters)*, vol. 89, n° 5, doi :10.1209/0295-5075/89/58003, p. 58 003, ISSN 0295-5075, 1286-4854. URL <http://stacks.iop.org/0295-5075/89/i=5/a=58003?key=crossref.2c6752a49b7f8210e613ba411218a5b2>. 59
- VERMA, M., J. CHOI, K. A. COTTRELL, Z. LAVAGNINO, E. N. THOMAS, S. PAVLOVIC-DJURANOVIC, P. SZCZESNY, D. W. PISTON, H. S. ZAHER, J. D. PUGLISI et S. DJURANOVIC. 2019, «A short translational ramp determines the efficiency of protein synthesis», *Nature Communications*, vol. 10, n° 1, doi :10.1038/s41467-019-13810-1, p. 1–15, ISSN 2041-1723. URL <https://www.nature.com/articles/s41467-019-13810-1>. 63

- WANG, Y., C. L. LIU, J. D. STOREY, R. J. TIBSHIRANI, D. HERSCHLAG et P. O. BROWN. 2002, «Precision and functional specificity in mRNA decay», *Proceedings of the National Academy of Sciences*, vol. 99, n° 9, doi :10.1073/pnas.092538799, p. 5860–5865. URL <https://www.pnas.org/doi/abs/10.1073/pnas.092538799>, publisher : Proceedings of the National Academy of Sciences. 60
- YAN, X., T. HOEK, R. VALE et M. TANENBAUM. 2016, «Dynamics of Translation of Single mRNA Molecules In Vivo», *Cell*, vol. 165, n° 4, doi :10.1016/j.cell.2016.04.034, p. 976–989, ISSN 0092-8674. URL <http://www.sciencedirect.com/science/article/pii/S0092867416304779>. 60
- YANG, E., E. V. NIMWEGEN, M. ZAVOLAN, N. RAJEWSKY, M. SCHROEDER, M. MAGNASCO et J. E. DARNELL. 2003, «Decay Rates of Human mRNAs : Correlation With Functional Characteristics and Sequence Attributes», *Genome Research*, vol. 13, n° 8, doi :10.1101/gr.1272403, p. 1863–1872, ISSN 1088-9051, 1549-5469. URL <https://genome.cshlp.org/content/13/8/1863>, company : Cold Spring Harbor Laboratory Press Distributor : Cold Spring Harbor Laboratory Press Institution : Cold Spring Harbor Laboratory Press Label : Cold Spring Harbor Laboratory Press Publisher : Cold Spring Harbor Lab. 60
- YOUNG, R. et H. BREMER. 1976, «Polypeptide-chain-elongation rate in Escherichia coli B/r as a function of growth rate», *The Biochemical Journal*, vol. 160, n° 2, doi :10.1042/bj1600185, p. 185–194, ISSN 0264-6021. 60



## Chapitre 4

# Étude des données de *ribosome profiling* et méthode de déduction des taux cinétiques de la traduction

« Suivant l'enseignement de l'École  
Ekuménique, lorsque l'action cesse  
d'être payante, recueil de  
l'information; lorsque  
l'information cesse d'être payante,  
dors. »

---

Ursula K. Le Guin (1969), *La Main  
gauche de la nuit* (trad. Jean  
Bailhache, 1971)

### Sommaire

---

<b>1</b>	<b>Introduction</b> . . . . .	<b>76</b>
<b>2</b>	<b>Les profils de densité polysome n'apportent pas assez d'information</b> . . . . .	<b>76</b>
<b>3</b>	<b>Information apportée par les densités <i>k</i>-somes</b> . . . . .	<b>77</b>
<b>4</b>	<b>Comparaison avec les profils <i>k</i>-somes obtenus par l'expérience Ribo-Seq sur un gène</b> . . . . .	<b>78</b>
<b>5</b>	<b>Analyse des profils Ribo-Seq sur les <i>k</i>-somes lissés</b> . . . . .	<b>80</b>
<b>6</b>	<b>Déduction des paramètres cinétiques de la traduction des densités polysomes et monosomes</b> . . . . .	<b>85</b>
6.1	Un profil de densité <i>k</i> -some apporte l'information manquante . . . . .	85
6.2	Fonction score . . . . .	85
6.3	Diagramme de phases . . . . .	87
<b>7</b>	<b>Conclusion et perspectives</b> . . . . .	<b>88</b>
<b>8</b>	<b>Références</b> . . . . .	<b>89</b>

---

## 1 Introduction

Différentes approches utilisent des analyses Ribo-Seq pour étudier la cinétique des ribosomes sur l'ARN messenger. Des "sites de poses" sont déterminés. Ils correspondent à des sites de l'ARNm sur lesquels un grand nombre de ribosomes a été détecté par rapport au reste de la séquence [HAN et al., 2014; INGOLIA et al., 2019]. D'autres études s'intéressent plus précisément à la vitesse du ribosome. Une première approche de INGOLIA et al. [2011] a été de mesurer la vitesse moyenne sur un brin d'ARNm entier. Ensuite, une méthode a été mise au point pour déterminer une vitesse moyenne sur quelques codons de l'ARNm seulement [DANA et TULLER, 2012]. Puis, la précision du codon a voulu être atteinte.

Dans toutes ces études, l'interprétation la plus basique des expériences Ribo-Seq, qui est de considérer que la vitesse des ribosomes est inversement proportionnelle à leur densité et donc à leur comptage en chaque codon a été utilisé. Les équations de champ moyen (Eq. 2.33) montrent en effet que dans le cas très basse densité où on néglige la densité en tout point devant 1, ce qui revient à négliger complètement toutes interactions entre ribosomes, on a :

$$\rho_i = \frac{\alpha}{p_i}, \quad (4.1)$$

Cette interprétation est la plus utilisée en biologie. Le problème est que le facteur de normalisation permettant de connaître la densité à partir du comptage n'est pas mesurable actuellement, ce qui ne permet de mesurer que les valeurs relatives entre elles des taux de saut  $p_i$  (ou vitesse) sur chaque codon  $i$  à partir des expériences Ribo-Seq.

Dans ce chapitre, nous allons explorer la mise au point d'une nouvelle méthode utilisant les densités monosomes et polysome pour déterminer les paramètres  $\alpha/\omega$  et  $p_i\omega$  pour chaque codon  $i$ . Une expérience supplémentaire pourra être faite pour déterminer les taux de dégradation  $\omega$  afin de déduire les taux absolus d'initiation  $\alpha$  et de saut  $p_i$  sur chaque codon et pour chaque ARNm. Une base de données des temps de demi-vie des ARNm produits chez la souris est déjà disponible grâce à SHAROVA et al. [2009], et pourrait être mesurée pour d'autres espèces. Nous allons d'abord préciser quelle information est apportée par chaque profil Ribo-Seq. Nous montrerons ensuite la correspondance entre les densités  $k$ -somes du modèle balistique et celles données par l'expérience Ribo-Seq d'un gène histone pour lequel une bonne statistique a été obtenue. Puis nous décrirons une analyse des profils Ribo-Seq sur les  $k$ -somes pour observer leur concordance avec le modèle balistique, ce qui nous permettra d'en déduire que la densité monosome est la meilleure densité  $k$ -somes à considérer pour notre méthode. Enfin, nous montrerons un test de la méthode de fit sur le modèle balistique lui-même, donc sans aucun bruit statistique ou issu des approximations de la modélisation.

## 2 Les profils de densité polysome n'apportent pas assez d'information

Les expériences Ribo-Seq habituelles ne concernent que les polysomes et leur modélisation ne tient pas compte du taux de dégradation de l'ARNm,  $\omega$ . Nous avons montré dans la partie 6.2 du chapitre précédent, que cela est justifié dans la mesure où, pour la plupart des cellules vivantes, la densité des polysomes est à peine sensible à la dégradation et où cette approximation ne conduit pas à des erreurs significatives dans l'estimation des taux de saut,  $p_i$ . Notre évaluation a cependant été faite avec des mesures médianes et moyennes de l'ensemble du génome, ce qui n'exclut pas qu'il existe certains ARNm à durée de vie extrêmement courte impliquant un effet de durée de vie finie non négligeable sur les polysomes. Mais le principal problème de l'estimation des taux cinétiques à partir des polysomes seuls est qu'il y a plus d'inconnues que d'équations. Même si nous supposons que le taux de dégradation  $\omega$  est connu grâce à des expériences indépendantes, nous avons toujours besoin d'estimer  $(L + 1)$  paramètres, l'ensemble  $\{p_i\}$  où  $i \in [1, L]$  et le taux d'initiation  $\alpha$ . Comme la densité des polysomes, par contre, ne fournit que  $L$  données,  $\rho(i)$  pour chaque  $i \in [1, L]$ , c'est-à-dire  $L$  équations, on se retrouve avec une inconnue supplémentaire.

L'estimation des paramètres cinétiques pose un autre problème : le nombre d'ARNm à partir duquel le profil Ribo-Seq est obtenu n'est pas connu, ce qui empêche une normalisation correcte de la densité Ribo-Seq. Une estimation paramétrique repose donc maintenant sur une densité normalisée arbitrairement. Nous choisissons de normaliser simplement à l'unité. La densité  $\bar{\rho}(x)$  telle que  $\int_0^L \bar{\rho}(x) dx = 1$  est

$$\bar{\rho}(x) = \frac{\omega}{p(x)} \frac{e^{-\omega \mathcal{T}(x)}}{1 - e^{-\omega \mathcal{T}(L)}}. \quad (4.2)$$

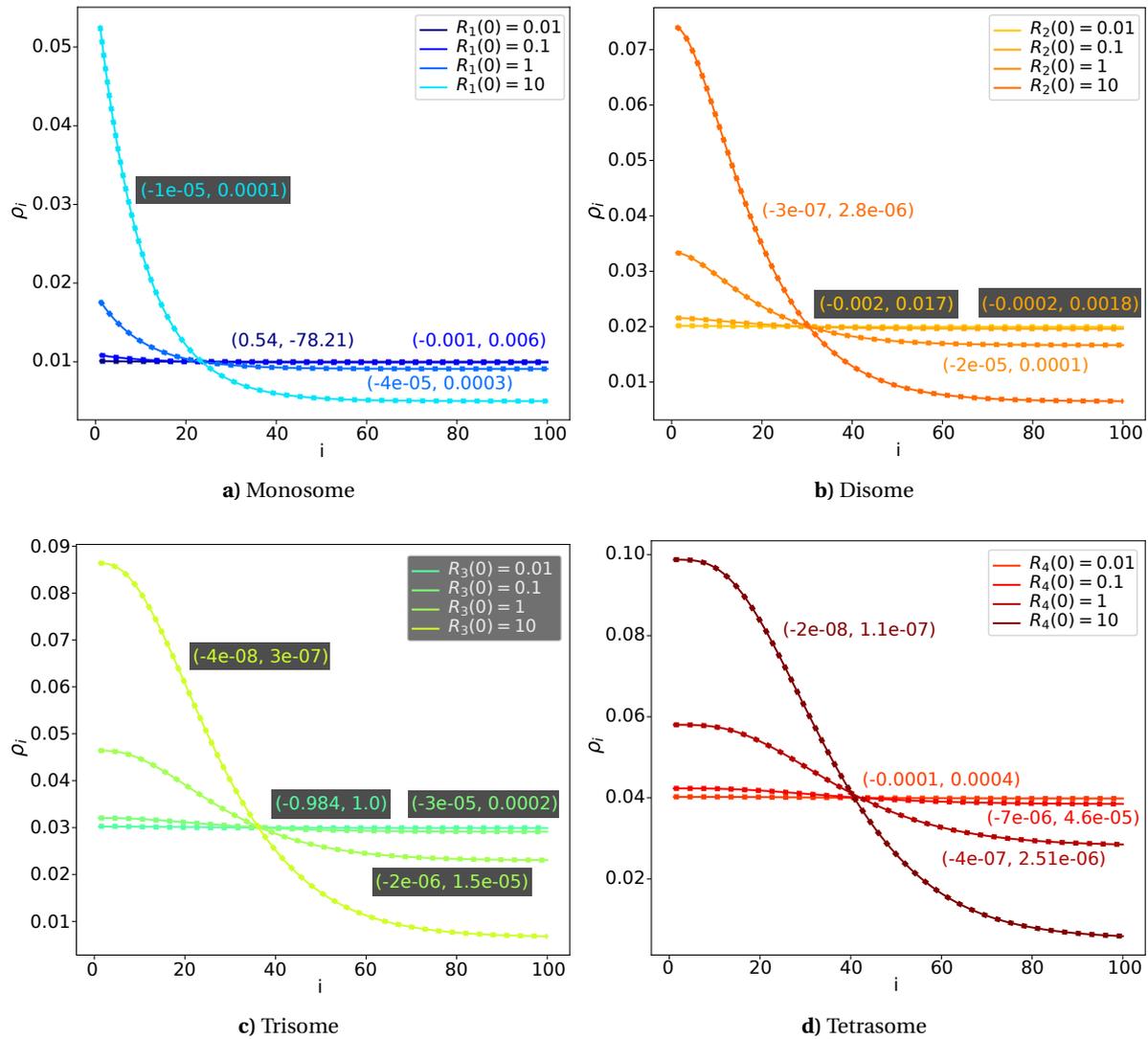
Cette dernière ne contient plus le paramètre  $\alpha$ , mais cela ne résout pas le problème puisque la nouvelle contrainte de normalisation supprime l'indépendance entre les densités aux différentes positions  $x$ .

Une stratégie utilisée par [DUC et SONG \[2018\]](#) pour contourner ce problème consiste à normaliser les profils Ribo-Seq par une vitesse moyenne des ribosomes mesurée à partir d'expériences indépendantes. Cette méthode présente toutefois plusieurs inconvénients : (i) la vitesse de ribosome utilisée est une moyenne sur l'ensemble d'un génome au lieu d'être spécifique à un gène, (ii) les résultats sont obtenus pour des espèces différentes de celles impliquées dans l'expérience Ribo-Seq, et (iii) les expériences sont réalisées indépendamment dans des conditions différentes. Dans l'ensemble, la normalisation obtenue par cette procédure est probablement sujette à de grandes fluctuations.

### 3 Information apportée par les densités $k$ -somes

Dans cette partie, nous allons montrer de manière empirique que les densités  $k$ -somes du modèle balistique nous donnent l'information sur les paramètres  $\tilde{\alpha}$  et  $\tilde{\omega}$  attendus en raison de leur sensibilité à l'effet de temps de vie fini, et que cette information est d'autant plus accessible que l'effet de temps de vie fini est important. Par cette occasion, nous verrons aussi la sensibilité remarquable de l'algorithme L-BFGS utilisé pour chaque fit. L'algorithme L-BFGS est une adaptation de la méthode quasi-Newton Broyden-Fletcher-Goldfarb-Shanno dans le but d'utiliser une quantité limitée de mémoire informatique. Nous l'utilisons encodé dans une fonction de la bibliothèque *SciPy* du langage de programmation interprété Python 3 [[VIRTANEN et al., 2020](#)].

Les densités étant non linéaires, il n'est pas simple de prouver mathématiquement que les expressions des densités  $k$ -somes sont injectives et nous allons simplement montrer que le fit sans bruit fonctionne ce qui signifie que les expressions des densités  $k$ -somes sont bien injectives. Sur la figure 4.1 nous présentons plusieurs fit du modèle balistique par le modèle lui-même. Nous avons fait des tests de fits des densités  $k$ -somes homogènes ( $p_i$  constant le long de l'ARNm) pour savoir s'il est possible de déduire les paramètres en jeu dans l'effet de durée de vie finie,  $\tilde{\alpha} = \alpha \mathcal{T}(L)$  et  $\tilde{\omega} = \omega \mathcal{T}(L)$ , d'un seul profil  $k$ -some quand il n'y a pas un excès d'inconnues par rapports aux équations dû à la vitesse non constante du ribosome le long de l'ARNm. Pour cela, nous avons fitté pour chaque  $k$ -some la densité  $\rho_k(i)$  pour  $\mathcal{R}_k(0) = 0.01, 0.1, 1, 10$  et regardé quelle erreur relative a été faite. On observe que le fit s'améliore effectivement avec l'augmentation de  $\mathcal{R}_k(0)$ . On remarque que l'erreur relative faite sur les paramètres est inférieure à 10% dès  $\mathcal{R}_k(0) = 0.1$ , ce qui témoigne de la grande performance de l'algorithme L-BFGS. Nous verrons par la suite que pour les tests de fit sur les densités inhomogènes, cette erreur augmentera d'un ordre de grandeur.

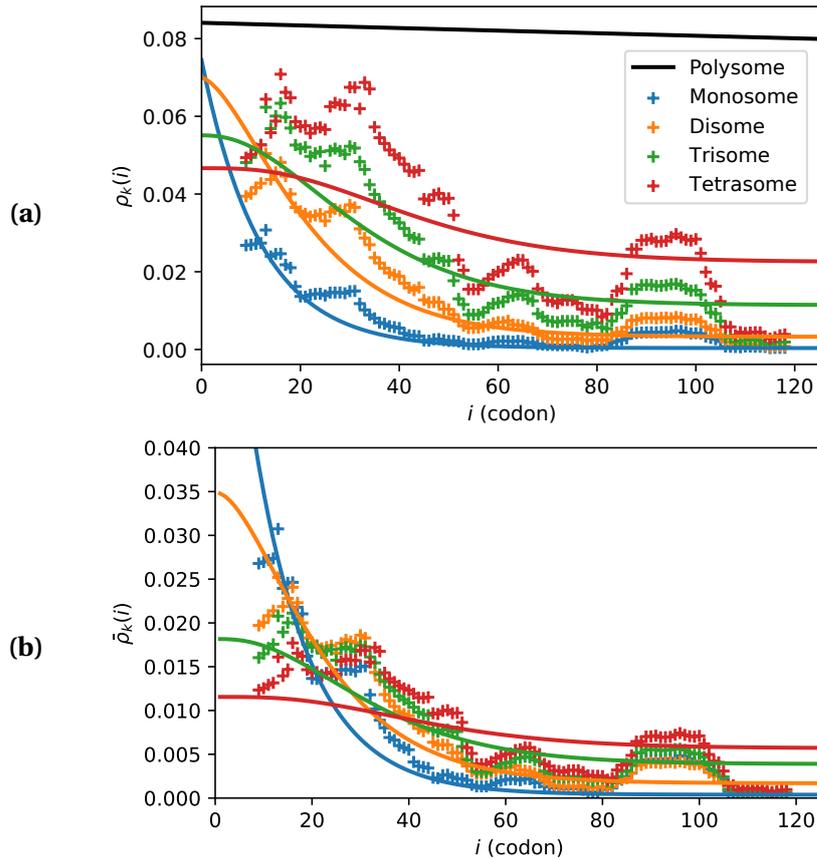


**FIGURE 4.1** – Test de fit des densités  $k$ -some du modèle balistique homogène avec les équations du modèle. Les paramètres fittés sont  $\tilde{\alpha}$  et  $\tilde{\omega}$ . Il est indiqué pour chaque profil les erreurs relatives sous la forme  $(\text{Err}(\tilde{\alpha}), \text{Err}(\tilde{\omega}))$ . On remarque (i) que le fit fonctionne dans la plupart des cas et (ii) que le fit s’améliore avec l’augmentation de  $R_1(0)$ . (Les cadres gris ont été mis pour faire ressortir les couleurs claires.)

## 4 Comparaison avec les profils $k$ -somes obtenus par l’expérience Ribo-Seq sur un gène

Dans cette partie, nous comparons la phénoménologie du modèle balistique pour les  $k$ -somes aux données expérimentales avec un gène codant pour une histone. Cette comparaison montrera que l’effet de la dégradation est visible dans certains gènes et que la sensibilité à la dégradation est accrue pour les  $k$ -somes.

Les figures 4.2 et 4.3 affichent les densités de  $k$ -somes,  $\rho_k$  ( $k=1$  à 4), en fonction des coordonnées génomiques obtenues à partir des expériences Ribo-Seq pour le gène d’histone HIST1H2BM (symboles) et à partir d’un ajustement de ces données à l’aide du modèle balistique homogène (lignes pleines). Les deux graphes 4.2a et 4.3a présentent les données Ribo-Seq lissées sur 19 codons en faisant la moyenne du signal dans la fenêtre de glissement et ont été normalisées de telle sorte que l’intégrale de chaque densité  $\rho_k$  corresponde à  $k$ . Lorsque la coordonnée génomique augmente, nous observons une décroissance exponentielle globale des densités, un phénomène prédit par le modèle balistique lorsque la dégradation de l’ARNm est prise en compte. Un autre effet intéressant peut être observé sur les graphes 4.2b et 4.3b de la figure. Ici, nous affichons les

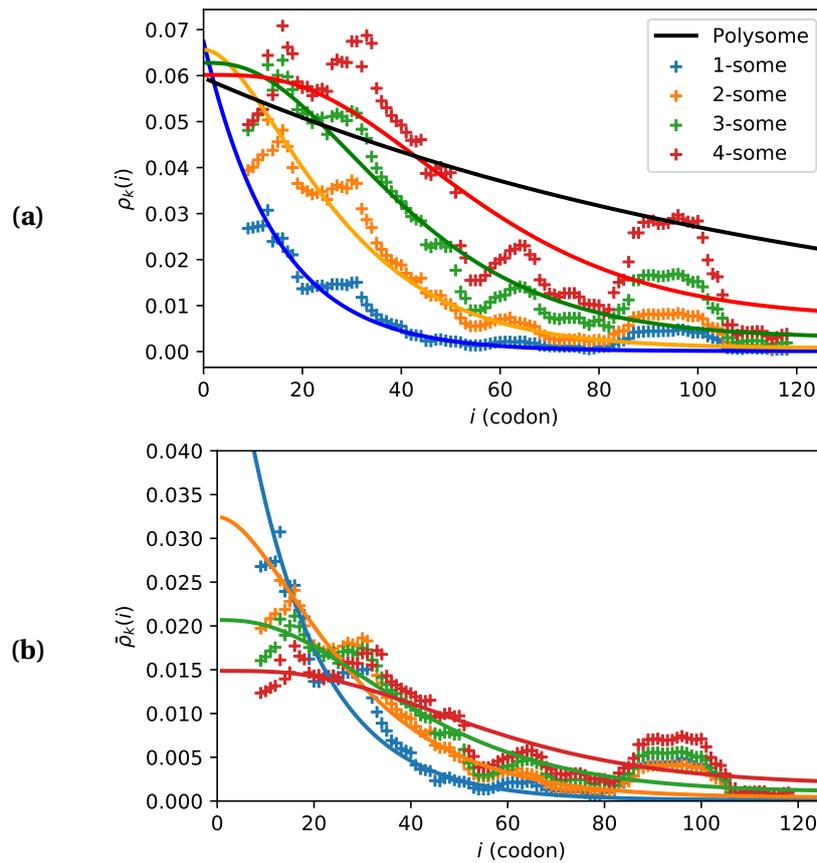


**FIGURE 4.2** – Fit homogène des données de l’expérience Ribo-Seq sur les  $k$ -somes  $\rho_k$  ( $k=1$  à 4) lissées, avec  $\alpha$  et  $\omega$  fixés aux valeurs trouvées dans la littérature. (a) Les densités de  $k$ -somes  $\rho_k$  ( $k=1$  à 4) obtenues à partir du modèle balistique homogène (lignes pleines) et des expériences Ribo-Seq (symboles) sont tracées en fonction de la coordonnée génomique. Les données Ribo-Seq sont lissées par une fenêtre glissante qui calcule la densité moyenne sur 19 codons. La densité des polysomes est représentée en noir pour le modèle balistique uniquement (absence de données expérimentales). Les données expérimentales correspondent au gène histone HIST1H2BM et le fit est basé sur le modèle balistique avec des paramètres fixes,  $\alpha = 0,06s^{-1}$  et  $\omega = 1/60 \text{ min}^{-1} \approx 3.10^{-4}s^{-1}$ . Le fit donne alors  $\mathcal{T}(L) = 178s$  et donc  $p = 0,7 \text{ codons/s}$ . (b) Les mêmes densités de  $k$ -somes sont normalisées à 1, c’est-à-dire  $\bar{\rho}_k = \rho_k/k$ , afin de comparer les écarts entre eux et de montrer que nous retrouvons le croisement caractéristique entre les courbes de  $k$ -somes balistiques.

densités  $k$ -somes Ribo-Seq normalisées à l’unité. Ces profils montrent très clairement que, près de l’entrée de l’ARNm, sur une distance génomique de 30 à 40 codons, les profils de densités  $k$ -somes diminuent avec l’augmentation de  $k$  alors qu’ils augmentent avec  $k$  au-delà de cette distance. Ce croisement des courbes de densité est à nouveau un effet prédit par le modèle balistique.

Nous utilisons d’abord (Fig. 4.2) la demi-vie spécifique des ARNm d’histones de 40 min, soit  $\omega = 1/60 \text{ min}^{-1} \approx 3.10^{-4}s^{-1}$  [HEINTZ et al., 1983] et le taux d’initiation pour *H. sapiens* (Tab. 3.1),  $\alpha = 0.06s^{-1}$ . Enfin,  $p$  sera estimé depuis  $\mathcal{T}(L) = L/p$  par le fit des quatre profils Ribo-Seq  $k$ -somes avec le modèle balistique. Nous trouvons finalement  $\mathcal{T}(L) = 178 \text{ s}$ , la valeur utilisée pour tracer les densités balistiques  $k$ -some correspondantes sur la Fig. 3.10. Notez qu’avec ces paramètres, le nombre moyen de ribosomes sur l’ARNm en régime stationnaire est de  $\langle k \rangle = \alpha \mathcal{T}(L) = 0,06 \times 178 \approx 10$ .

Globalement, nous observons que les tendances qualitatives présentées par les données Ribo-seq sont décrites par le modèle balistique et que l’inversion des profils de densité  $k$ -some près du début de l’ARNm donne une signature claire de l’influence de la dégradation de l’ARNm sur les profils Ribo-Seq. Enfin, la densité théorique pour les polysomes (ligne noire continue dans le panneau intermédiaire de la figure) n’a pas de contrepartie expérimentale mais, d’après les paramètres estimés, elle doit être quasiment plate et non influencée par la dégradation, comme on l’a



**FIGURE 4.3** – Fit homogène des données de l’expérience Ribo-Seq sur les  $k$ -somes  $\rho_k$  ( $k=1$  à 4) lissées, sans fixer  $\alpha$  et  $\omega$ . Les données et le lissage sont les mêmes que sur la Fig. 4.2. On trouve une meilleure correspondance entre les données et le modèle que lorsqu’on fixe  $\alpha$  à  $0.06s^{-1}$  et  $\omega$  à  $3.10^{-4}s^{-1}$ . Elle est trouvée pour  $\tilde{\alpha} = 7,6$  et  $\tilde{\omega} = 1$ . (a) Densités proprement normalisées à  $k$ . (b) Densités normalisées à 1. Le croisement avec les courbes trisomie et tétrasomie sont mieux retrouvés que sur la Fig. 4.2.

déjà vu dans le régime de dégradation intermédiaire présenté dans la Fig. 3.10.

Néanmoins, nous voyons un écart qui s’accroît avec  $k$  entre les données et le modèle. Pour cette raison, nous avons effectué un autre ajustement sans fixer de paramètre au préalable. On obtient alors une meilleure correspondance en particulier pour le trisomie et le tétrasomie (Fig. 4.3). Elle est obtenue pour  $\tilde{\alpha} = 7,6$  et  $\tilde{\omega} = 1$ . La valeur de  $\tilde{\omega}$  est inattendue puisqu’elle correspond à  $\mathcal{T} = 1/\omega$ , ce qui signifie qu’à peine une seule traduction de l’ARNm aurait le temps de se faire.

Ce résultat fournit une motivation supplémentaire forte pour la nécessité de diviser la population d’ARNm en  $k$ -somes afin d’augmenter la sensibilité à la dégradation de l’ARNm. On note que d’après notre analyse d’échelle de  $f(\tilde{\omega})$  décrite à la section 6.2 du chapitre 3, on s’attendait à ce que dans le cas présent, la dégradation de l’ARNm ait peu d’influence sur la densité des polysomes puisque  $\tilde{\omega} = \omega \mathcal{T}(L) = 3.10^{-4} \times 178 \approx 0.05 \ll 1$ . Enfin, nous concluons cette section en remarquant la nécessité d’introduire un  $p(x)$  inhomogène, car nous pouvons observer des fluctuations fortes et systématiques le long des profils qui ne peuvent être expliquées par les fluctuations expérimentales, mais plutôt par la distribution de  $p(x)$ .

## 5 Analyse des profils Ribo-Seq sur les $k$ -somes lissés

Armelle Choquet et Alexandre David, nos collègues biologistes de l’IGF (Institut de Génomique Fonctionnelle) ont effectué l’expérience Ribo-Seq sur les  $k$ -somes de l’ensemble du génome humain dont les données ont été traitées par Julie Ripoll et Eric Rivals, nos collègues bioinformaticiens du LIRMM (Laboratoire d’Informatique, de Robotique et de Microélectronique de Montpellier). Nous avons donc pu commencer par comparer le modèle avec les profils  $k$ -somes humains

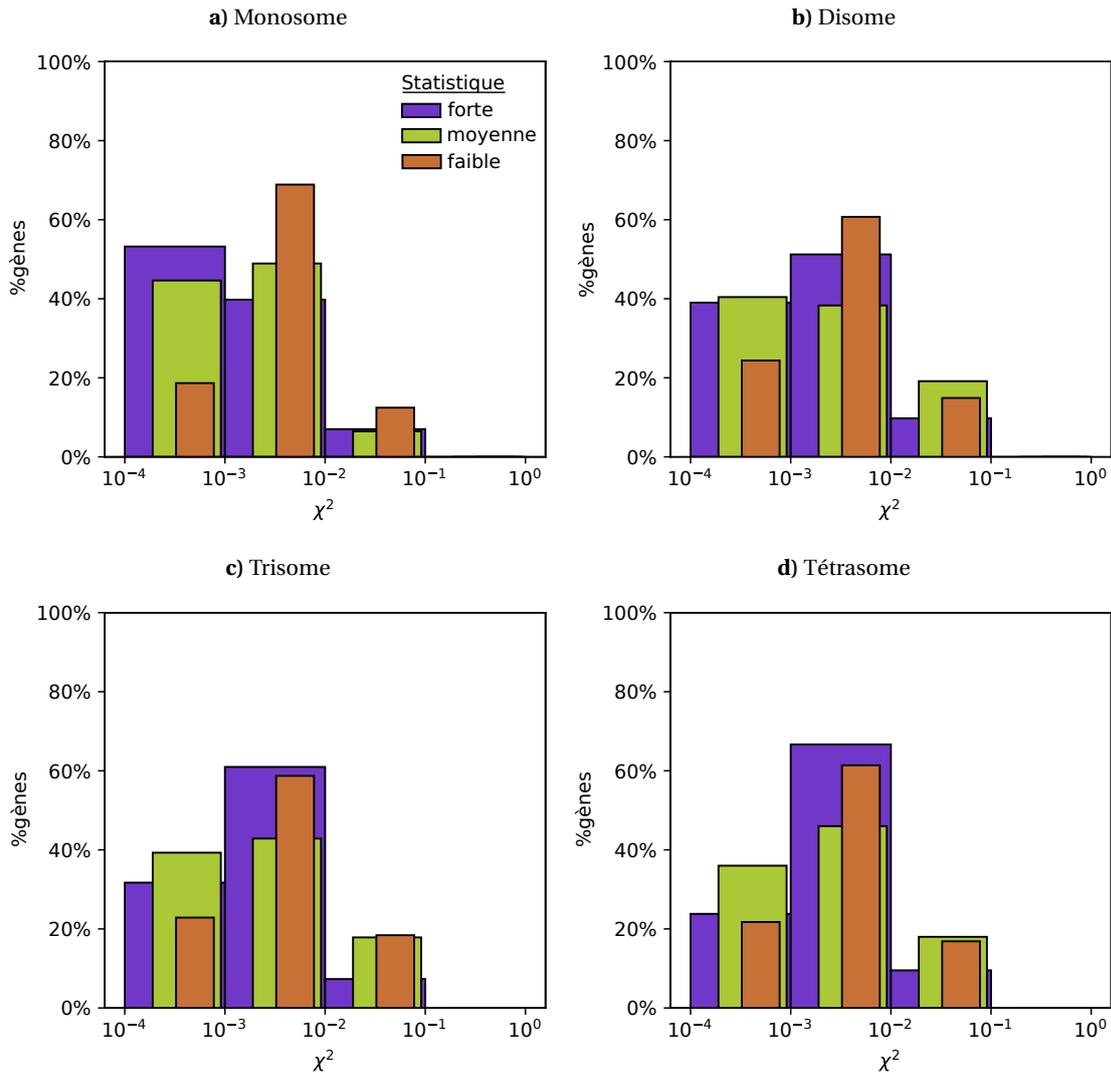
dans le but d'observer la correspondance du modèle au système des  $k$ -somes mesuré. Pour cela, nous avons effectué un fit des données lissées. Nous avons d'abord fitté les  $k$ -somes pour chaque  $k$  allant de 1 à 4 indépendamment quel que soit le gène, puis nous avons ensuite fitté l'ensemble des  $k$ -somes pour  $k$  allant de 1 à 4, la fonction de minimisation devant trouver un seul jeu de paramètres  $(\tilde{\alpha}, \tilde{\omega})$  par gène correspondant aux quatre  $k$ -somes, comme pour le test de la partie 4 précédente. Nous avons analysé seulement les gènes codant pour des protéines du noyau pour nous assurer que la traduction mesurée n'est pas faite par des ribosomes attachés au réticulum endoplasmique, ce qui pourrait modifier la dynamique du processus.

Nous avons étudié séparément les 3 groupes de gènes définis par leur couverture en ribosomes par nos collègues biologistes selon la procédure suivante : Pour chaque gène, ils calculent la couverture moyenne Ribo-Seq entre les codons 20 et 200 (en comptant à partir du codon de départ, et au moins 20 codons avant le codon d'arrêt). Ils éliminent les gènes ayant moins de 10 empreintes Ribo-seq. Ensuite, ils ordonnent les gènes en fonction de la couverture moyenne décroissante. À partir de cette liste, ils divisent l'ensemble des gènes en deux groupes, de sorte que leur couverture totale soit égale. Cela détermine un point médian dans la liste, qui sépare la liste en gènes fortement et faiblement exprimés. Pour accentuer le contraste entre ces deux groupes, ils excluent les 5% supérieurs des gènes faiblement exprimés et les 5% inférieurs des gènes fortement exprimés. Ces 10% de gènes centraux en termes de couverture médiane sont dits avoir une "couverture intermédiaire". On note que le nombre de gènes dans les groupes de gènes faiblement et fortement traduits ne doit pas être égal.

Ces catégories ont été définies à l'origine pour différencier les gènes fortement traduits des faiblement traduits et étudier quelles caractéristiques ces deux groupes partagent ou s'ils diffèrent par d'autres caractéristiques. Dans cette partie, nous allons reprendre ces groupes pour leur différence statistique puisque pour les  $k$ -somes, le nombre de copies d'ARNm mesurés pour un gène est proportionnel au nombre de ribosomes détectés. Précisément, la mesure d'une copie d'ARNm à l'état de  $k$ -some correspond à la détection de  $k$  ribosomes. Les gènes peu couverts ayant un nombre de positions de ribosomes détectés plus faibles, leur qualité statistique est moindre que celles des gènes très couverts. Le gène présenté dans la partie précédente appartient à la catégorie *haute couverture*, il a donc une des meilleures qualités statistiques. Afin d'analyser le plus grand nombre de gènes possible, et étant donné que cela ne gêne pas notre étude, nous avons souhaité ne pas exclure les gènes de couverture intermédiaire dans l'étude statistique de l'adéquation du modèle aux densités Ribo-Seq. Finalement, nous considérerons les trois groupes de gènes *haute couverture*, *basse couverture* et *couverture intermédiaire* définis par les bioinformaticiens comme trois groupes de gènes dont la mesure Ribo-Seq a une qualité statistique différente : *forte statistique*, *basse statistique* et *statistique moyenne*.

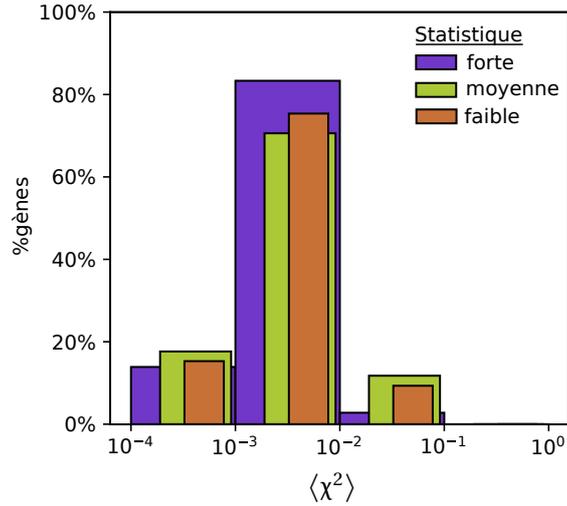
Pour chaque catégorie de gènes, on s'intéresse cette fois-ci à la qualité du fit en regardant quel est l'écart entre la densité du modèle balistique homogène et le profil de données Ribo-Seq lissé. A cause des densités nulles obtenues par l'expérience, nous calculons la différence non relative à la densité Ribo-Seq bien qu'une différence relative serait plus précise. Nous définissons cette différence simplement par  $\chi^2 = \sum_{i=1}^N |\bar{\rho}_{RS}(i) - \bar{\rho}_{balistique}(i)|^2$ , où l'indice "RS" signifie "Ribo-Seq" et  $N$  est le nombre de codons considérés pour le profil lissé. Nous utilisons les densités normalisées à 1 pour que les densités fluctuent autour des mêmes valeurs quelque soit  $k$  et qu'on puisse ainsi comparer les valeurs de  $\chi_{DL}$  pour les différents  $k$ -somes entre elles. Les gènes étant en majorité de l'ordre de  $10^2$  codons (avec une médiane à environ 500 codons, voir Tab. 3.1), leurs densités normalisée à 1 est de l'ordre  $10^{-3}$ . Compte tenu des fluctuations qui sont négligées dans ce fit préliminaire, mais qui sont prévues par le modèle balistique, nous considérons que les données et le modèle s'accordent bien lorsque  $\chi_{DL}^2$  est de l'ordre de  $10^{-3}$  ou moins.

Sur la Fig. 4.4, nous avons représenté sous forme d'histogramme les résultats pour chaque groupe de gènes et pour chaque  $k$ -some. L'histogramme de la Fig. 4.5 représente les résultats en fittant de la même manière, mais en cherchant, pour chaque paramètre, une seule valeur commune aux quatre densités  $k$ -somes pour  $k = 1, 2, 3, 4$  d'un gène. Les densités pour quelques gènes sont visibles en annexes A.3, A.4, A.5, A.6 et A.7.



**FIGURE 4.4** – Histogrammes des erreurs faites sur le fit des données  $k$ -somes lissées sur 19 codons par les densités du modèle balistique en considérant une vitesse  $p$  constante le long de l'ARNm.  $\tilde{\alpha}$  et  $\tilde{\omega}$  sont les paramètres recherchés sans qu'aucun des paramètres ne soit fixé. Les 3 groupes de gènes séparés par la qualité statistique de leur mesure ont été analysés séparément. Les monosomes, disomes, trisomes et tétrasomes ont été fittés indépendamment. Chaque profil Ribo-Seq a été fitté indépendamment. Les histogrammes représentent le pourcentage de chaque ordre de grandeurs de  $\chi^2_{DL}$  lors des fits. La largeur des barres a été modifiée de façon à ce que trois histogrammes soient visibles en étant superposés. Les nombres de gènes analysés pour chaque histogramme sont indiqués en Tab. A.1.

**Monosomes** Nous voyons dans un premier temps sur la Fig. 4.4 que seulement 12% des gènes ayant une forte statistique, 6% des gènes ayant une statistique moyenne et 7% des gènes ayant une statistique faible ne correspondant pas au modèle balistique homogène dans leur forme lissée. Ce sont généralement des gènes qui possèdent des variations trop importantes de la densité pour que le lissage ait pu faire ressortir une enveloppe décroissante. De plus, nous voyons de manière remarquable que la correspondance du modèle avec les données augmente avec la qualité statistique. En effet, une majorité de gènes à faible statistique ont été fittés avec un  $\chi^2$  de l'ordre de  $10^{-3}$ . Pour les gènes à statistique moyenne, il a été obtenu quasiment autant de  $\chi^2$  d'ordre de  $10^{-3}$  que de  $\chi^2$  d'ordre de  $10^{-4}$ . Enfin, les gènes à forte statistique ont obtenu le meilleur résultat avec une majorité de fit avec un  $\chi^2$  d'ordre de  $10^{-4}$ . Cela constitue un argument en faveur de la pertinence du modèle pour décrire les expériences Ribo-Seq et la traduction de l'ARNm.



**FIGURE 4.5** – Histogrammes des erreurs faites sur le fit des données  $k$ -somes lissées sur 19 codons par les densités du modèle balistique en considérant une vitesse  $p$  constante le long de l'ARNm. Les 3 groupes de gènes séparés par la qualité statistique de leur mesure ont été analysés séparément. Pour un gène donné, les quatre profils monosome disome trisome et tétrasome ont été fittés ensemble pour un seul jeu de paramètres  $(\bar{\alpha}, \bar{\omega})$ . Les histogrammes représentent le pourcentage de chaque ordre de grandeurs de  $\chi_{DL}^2$  lors des fits. La largeur des barres a été modifiée de façon à ce que trois histogrammes soient visibles en étant superposés. Les nombres de gènes analysés sont indiqués en Tab. A.1.

**Disomes** Lorsqu'on regarde l'histogramme du  $\chi^2$  du fit homogène des disomes, on remarque que le fit de la majorité des gènes ayant une forte statistique a un  $\chi^2$  qui bascule vers des valeurs de l'ordre de  $\sim 10^{-3}$ , mais il y a toujours une proportion importante de  $\chi^2 \sim 10^{-4}$ . Pour les gènes à statistique moyenne, il y a surtout un déplacement d'une partie des  $\chi^2 \sim 10^{-3}$  vers les  $\chi^2 \sim 10^{-2}$ . Paradoxalement, il y a une augmentation des  $\chi^2 \sim 10^{-4}$  pour les gènes à faible statistique, mais seulement de quelques pourcents, ce qu'on peut considérer comme non significatif pour ces gènes de moins bonne statistique. La proportion de  $\chi^2 \geq 10^{-2}$  étant de moins de 20% pour les gènes de statistique moyenne et basse, et de moins de 10% pour les gènes de forte statistique, nous pouvons considérer que la forme globale des densités disomes est bien prédite par le modèle balistique.

**Trisomes** Concernant les trisomes, on observe à nouveau un décalage des  $\chi^2 \sim 10^{-4}$  vers les  $\chi^2 \sim 10^{-3}$  pour les gènes très couverts, mais très peu de variations pour les gènes de statistique moyenne et faible. La proportion de  $\chi^2 \geq 10^{-2}$  reste de moins de 20% pour les gènes de statistique moyenne et basse, et de moins de 10% pour les gènes de forte statistique, ce qui permet de valider le modèle également pour les trisomes.

**Tétrasomes** La tendance de la qualité du fit à se dégrader légèrement avec l'augmentation de  $k$  mais avec des  $\chi^2$  qui restent inférieur à  $10^{-2}$  se confirme avec l'histogramme sur les tétrasomes. Le modèle semble donc correspondre aux densités  $k$ -somes quelque soit  $k$ . Il serait intéressant de regarder par la suite si cette différence entre les fit des densités  $k$ -somes ne sont pas seulement dues aux fluctuations qui doivent augmenter avec  $k$  en raison de la diminution de l'effet de temps de vie fini qui écrase les densités vers la fin du réseau.

**Fit simultané des  $k$ -somes pour  $k$  allant de 1 à 4** Cette fois, les  $k$ -somes pour  $k = 1, 2, 3, 4$  d'un gène sont fittés en même temps comme dans l'exemple de l'histone présenté en Fig. 4.3. On s'intéresse à la différence quadratique  $\chi^2$  moyennée sur les quatre  $k$ -somes ( $k = 1, 2, 3, 4$ ), soit

$$\langle \chi^2 \rangle = \frac{1}{4} \sum_{k=1}^4 \sum_{i=1}^N |\bar{\rho}_{k,RS}(i) - \bar{\rho}_{k,balistique}(i)|^2 \quad (4.3)$$

de manière à pouvoir la comparer aux fits précédents. Pour les gènes à forte statistique, nous observons que la grande majorité, plus de 80%, ont un  $\langle \chi^2 \rangle \sim 10^{-3}$ , et donc peu de gènes ont un  $\langle \chi^2 \rangle$  de l'ordre de  $10^{-4}$  (environ 15%) et  $10^{-2}$  (moins de 5%). L'histogramme ne correspond donc pas à la moyenne des histogrammes précédents des fits des  $k$ -somes pour  $k = 1, 2, 3, 4$  pris séparément où la proportion de  $\chi^2 \sim 10^{-3}$  varie progressivement entre environ 40% et 60%, la proportion de  $\chi^2 \sim 10^{-4}$  varie progressivement entre environ 20% et 50% et la proportion de  $\chi^2 \sim 10^{-2}$  reste autour de 10%. Une plus grande proportion de  $\chi^2 \sim 10^{-3}$  et une plus faible proportion de  $\chi^2 \sim 10^{-4}$  montrent une concordance moins bonne que précédemment. Cependant, la très faible proportion de  $\chi^2 \sim 10^{-2}$ , c'est-à-dire au-dessus du seuil de validité, nous permet de conclure que le modèle décrit correctement les tendances données par les mesures expérimentales des densités de l'ensemble des  $k$ -somes.

Cette comparaison entre le modèle et les données Ribo-Seq sur les  $k$ -somes nous indique une bonne prédiction de l'effet du temps de vie fini sur les densités de ribosomes. Pour préciser cette comparaison, il serait intéressant d'effectuer des comparaisons plus précises en utilisant le modèle balistique inhomogène, ce qui pourrait permettre de comprendre pourquoi la correspondance est moins bonne quand on considère les  $k$ -somes pour  $k=1,2,3,4$  plutôt que séparément.

Premièrement, on pourra savoir si les différences de correspondance entre le modèle homogène et les données lissées ne sont tout simplement pas dues aux différences d'amplitude des fluctuations. En effet, comme on peut le voir sur la Fig. 4.3 et sur les exemples de fit donnés en annexe (A.3, A.4, A.5, A.6), l'amplitude des fluctuations a tendance à augmenter avec le nombre de ribosomes  $k$ . Nous expliquons cela par un effet de temps de vie fini qui augmente quand  $k$  diminue et qui engendre un écrasement des densités vers la fin de l'ARNm.

Si les variations d'amplitude des fluctuations ne suffisent pas à expliquer l'écart entre le modèle et les données Ribo-Seq, nous pensons qu'il pourrait s'agir d'un effet d'exclusion non négligeable entre les ribosomes. Il a été découvert chez les eucaryotes un processus de contrôle de la traduction par la détection et l'élimination de ribosomes qui calent, *ribosome stalling* en anglais, appelé *no-go decay* [DOMA et PARKER, 2006]. Si ce processus était très efficace, il pourrait empêcher tout embouteillage et donc la diminution de la vitesse d'un ribosome à cause d'un autre. Mais, SIMMS et al. [2017] ont trouvé que le processus s'enclenche quand au moins trois ribosomes sont embouteillés. L'exclusion stérique peut donc être une interaction ayant un effet non négligeable sur la vitesse, la densité et le flux de ribosomes. En effet, une étude biophysique de SZAVITS-NOSSAN et CIANDRINI [2020] a aussi mis en évidence la présence de site d'exclusion local sur certains gènes de la levure *S. cerevisiae*, même s'ils considèrent qu'il y a peu d'exclusion en moyenne. Les collisions entre ribosomes pourraient aussi engendrer une variation de la durée de vie en fonction du nombre de ribosomes présent sur l'ARNm. Lorsque des collisions entre ribosomes ont lieu, des mécanismes interviennent pour interrompre la traduction et détruire l'ARNm [MORRIS et al., 2021]. Le risque de collisions augmentant avec le nombre de ribosomes, des ARNm sur lequel il y a plus de ribosome vivrait moins longtemps.

Une étude future serait donc d'inclure de l'exclusion dans ce modèle. Cependant, on perdrait beaucoup en simplicité et il semble compliqué d'établir un modèle d'ensemble d'ARNm à temps de vie fini avec de l'exclusion entre particules qui reste efficace. Nous concluons que ce modèle se révèle pertinent pour l'étude des monosomes. Cependant, interpréter la densité polysome pourrait être juste étant donné que leur durée de vie finie a un effet négligeable sur la densité. Pour vérifier que la cinétique des ribosomes n'est pas modifiée significativement par les interactions entre ribosomes, on pourra étudier la corrélation entre les pics des densités monosome et polysome. Une étude plus précise pourra ensuite être faite en étudiant les corrélations entre les pics des densités  $k$ -somes à différent  $k$ .

## 6 Déduction des paramètres cinétiques de la traduction des densités polysomes et monosomes

### 6.1 Un profil de densité $k$ -some apporte l'information manquante

Puisque la demi-vie de l'ARNm  $\tau_{1/2}$  peut être mesurée par des expériences standards, nous considérerons dorénavant que  $\omega$  est connu, puisque  $\omega = \ln(2)/\tau_{1/2}$  par définition.

Nous avons également montré que le profil de densité du polysome est beaucoup moins sensible à  $\omega$  que les profils de  $k$ -some pour  $k = 1, 2, 3, 4$  (Fig. 3.10). Cela signifie que le profil de densité normalisé des polysomes dépend principalement du profil de la vitesse de saut,  $p(x)$ , et seulement légèrement de  $\omega$  (puisque nous négligeons les effets d'obstruction des ribosomes dans le modèle balistique). Le profil de densité normalisé des polysomes est donc une bonne mesure des variations de  $p_i$  indépendamment des autres paramètres.

Dans la section 6.5 du chapitre 3 et dans la section 3 du présent chapitre, nous avons montré que des informations supplémentaires sont disponibles en augmentant l'effet de durée de vie finie lorsque les différents  $\rho_k$  deviennent de moins en moins linéairement liés. De plus, nous avons montré que pour un certain niveau de dégradation de l'ARNm, le signal des profils de densité de  $k$ -some disparaît près de la fin de l'ARNm et nous perdons des informations, mais cela se produit pour des paramètres éloignés des valeurs biologiquement viables. Dans la section précédente, nous avons montré que la densité monosome est la mieux prédite par le modèle balistique. Nous avons donc décidé d'ajouter au profil de densité normalisé du polysome l'information fournie par le profil de densité du monosome. De plus, il s'agit du profil  $k$ -some le plus affecté par la durée de vie finie de l'ARNm à l'inverse du profil polysome qui est le moins affecté. Le profil monosome présente le plus grand ratio d'ARNm à l'état transitoire par rapport à l'ARNm à l'état stationnaire, une quantité dont nous avons montré qu'elle était directement liée à l'enveloppe des profils de densité des  $k$ -somes dans la section 6.6 du chapitre 3.

### 6.2 Fonction score

Cette section est ensuite consacrée à la détermination des valeurs du taux d'initiation  $\alpha$  et des taux de saut  $p_i$  de tous les codons constituant l'ARNm. Nous trouvons ces paramètres cinétiques en minimisant la différence entre le modèle balistique et, simultanément, les données Ribo-Seq du profil de densité normalisée à 1 du polysome et du profil de densité du monosome. Pour être précis, nous minimisons la quantité

$$\chi^2 = \chi_{\text{polysome}}^2 + \chi_{1\text{-some}}^2, \quad (4.4)$$

où  $\chi_{\text{polysome}}^2$  est la somme de l'erreur relative quadratique de la densité normalisée des polysomes sur chaque codon :

$$\chi_{\text{polysome}}^2 = \sum_{i=1}^L \left| \frac{\rho_i^{\text{data}} - \rho_i^{\text{ballistic}}}{\rho_i^{\text{data}}} \right|^2, \quad (4.5)$$

et  $\chi_{1\text{-some}}^2$  est la somme de l'erreur relative quadratique de la densité des monosomes sur chaque codon définie de la même manière que  $\chi_{\text{polysome}}^2$ . La minimisation est effectuée par l'algorithme L-BFGS encodé dans la bibliothèque *SciPy* du langage de programmation interprété Python 3 [VIR-TANEN et al., 2020].

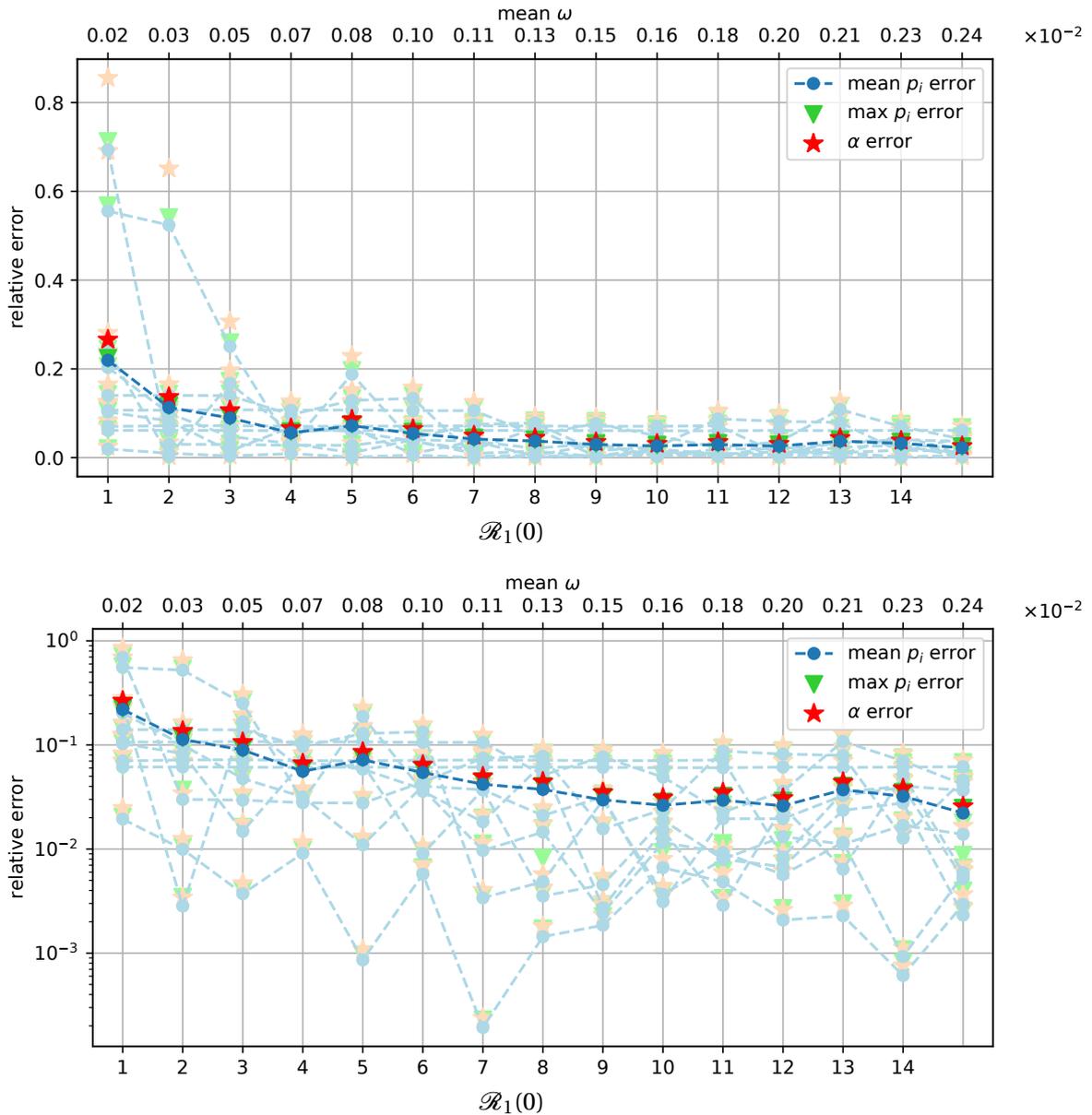
Afin d'évaluer notre méthode de fit et le critère de qualité défini ci-dessus, nous traçons ce que nous appelons la "fonction score". Il s'agit de l'erreur relative sur les valeurs des paramètres cinétiques trouvées par la procédure de minimisation en fonction du ratio  $\mathcal{R}_1(0)$  (Eq. (3.85)). La procédure de minimisation est effectuée sur des données artificielles créées avec le modèle balistique. L'objectif est d'observer si la procédure de minimisation récupère de manière précise les paramètres d'entrée connus en fonction des valeurs de  $\mathcal{R}_1(0)$ . Les erreurs relatives sont simplement définies par

$$Err_Y = \frac{|Y - Y_{\text{minimization}}|}{Y}, \quad (4.6)$$

où  $Y$  est le paramètre évalué. Nous évaluons  $\alpha$  et chaque  $p_i$ . Pour que la fonction de score soit lisible, nous choisissons concernant les  $p_i$ , de nous concentrer sur l'erreur moyenne sur les  $L$  valeurs de  $p_i$  et l'erreur relative maximale parmi les  $L$  valeurs de  $p_i$ .

Nous calculons la fonction de score pour dix profils partageant le même  $\alpha = 0.08$  et le même  $\omega$ . Les valeurs de  $p_i$  sont déterminées par un tirage aléatoire d'une distribution uniforme entre  $p_{\min} = 0.2$  et  $p_{\max} = 4$ . Le tirage aléatoire est la variable entre les dix profils. L'objectif est de lisser la variabilité de l'erreur due à la disposition et à la spécificité de  $p_i$  en faisant la moyenne de la fonction score sur les dix profils. Pour chaque profil, nous effectuons la minimisation pour 20 valeurs  $\mathcal{R}_1(0)$ . Cela signifie que nous exécutons la procédure de minimisation cent fois. Le paramètre ajusté correspondant à la valeur déterminée de  $\mathcal{R}_1(0)$  est  $\omega$ .

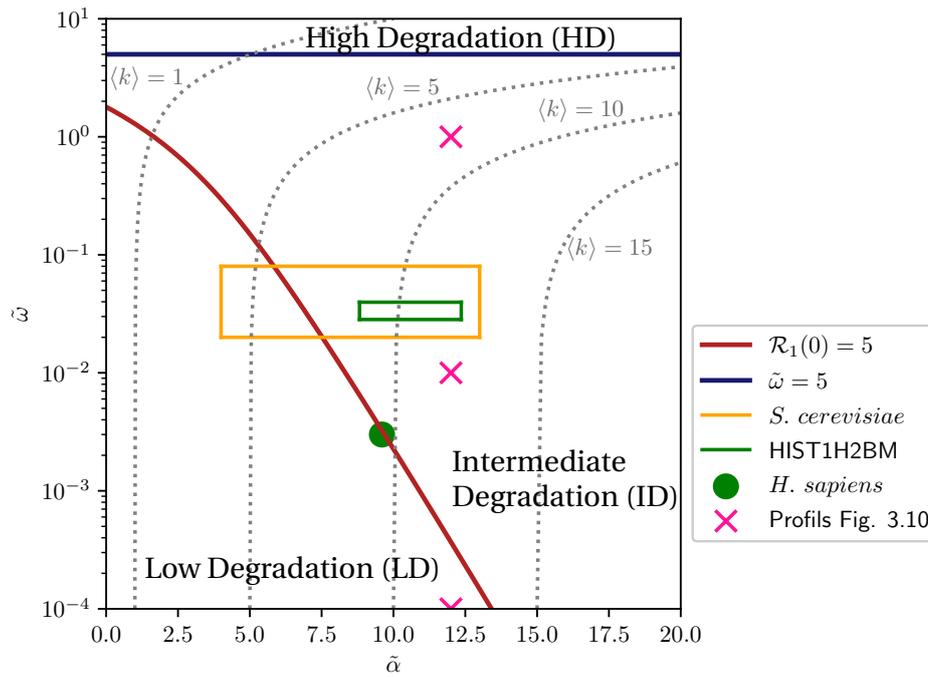
En observant la figure 4.6, nous pouvons constater que les erreurs relatives diminuent sur  $\mathcal{R}_1(0)$  de 1 à 6 et restent plutôt constantes pour les plus grandes valeurs de  $\mathcal{R}_1(0)$ . Ceci confirme la pertinence du critère que nous avons choisi pour évaluer la qualité de l'ajustement. On constate par ailleurs que les erreurs relatives sont inférieures à 10% à partir de  $\mathcal{R}_1(0) = 4$ , ce qui constitue une validation de notre méthode d'ajustement. Les  $\mathcal{R}_1(0) = 4$  correspondent aux paramètres  $\alpha = 0.08s^{-1}$ ,  $\mathcal{T}(L) \approx 81s$  et  $\omega \approx 0.0005s^{-1} \approx \frac{1}{33}min^{-1}$  qui sont des valeurs biologiquement raisonnables pour la traduction des ARNm. Plus précisément, nous mettons en évidence et déterminons la faisabilité d'une telle méthode de fit pour déterminer les paramètres cinétiques de la traduction de l'ARNm.



**FIGURE 4.6** – La fonction score du fit avec le modèle balistique lorsque  $\alpha = 0.08$  et  $p_i \in [0.2, 4]$ . Le résultat est la moyenne de la minimisation de l'équation 4.4 sur dix profils de densité balistique. Représentations et échelle lin-lin et en échelle lin-log. Ils sont obtenus par tirage aléatoire à partir d'une distribution uniforme de taux de saut compris entre 0,2 et 4.  $\mathcal{T}(L) = 81,35s$  en moyenne sur les dix distributions de  $p_i$ . La valeur de  $\mathcal{R}_1(0)$  est augmentée en augmentant  $\omega$ . Les erreurs sur les taux  $\alpha$  et  $p_i$  sont calculées avec l'équation 4.6.

### 6.3 Diagramme de phases

Nous terminons ce chapitre en reprenant le diagramme de phase du chapitre 3 pour l'adapter aux résultats de la méthode de fit. D'après nos tests, nous voyons sur la fonction score de la Fig. 4.6 que pour notre méthode avec l'algorithme de minimisation que nous utilisons, l'effet de temps de vie fini sur la densité monosome est suffisamment prononcé pour que les valeurs des paramètres puissent être correctement estimés lorsque  $\mathcal{R}_1(0) \geq 4$ . Nous arrondissons ce chiffre à 5 pour considérer de manière plus générale un ordre de grandeur pour la synthèse que constitue le diagramme de phase. La limite entre le régime LD et le régime ID sera donc légèrement déplacé vers les  $\bar{\alpha}$  et  $\bar{\omega}$  grands. Nous avons aussi testé le fit pour des  $\bar{\omega}$  de l'ordre de 1 pour les paramètres  $\alpha = 0.08$ ,  $p \in [0.2; 4]$  et constaté que l'algorithme de minimisation est assez efficace pour évaluer les vitesses  $p_i$  même quand la densité polysome est écrasée en bout de séquence à cause de l'effet de durée de vie finie, tant que la densité polysome reste d'un ordre supérieur à  $10^{-5}$ . Nous pou-



**FIGURE 4.7** – Diagramme de phase du modèle balistique dans le plan  $(\tilde{\alpha}, \tilde{\omega})$  3.13 ajusté à la faisabilité du fit. Le diagramme de phase est divisé en trois régimes de dégradation avec les lignes d'équation  $\tilde{\omega} = 5$  et  $\mathcal{R}_1(0) = 5$ . Le régime ID est celui où notre méthode de fit utilisant les profils monosome et polysome est effective. Les valeurs biologiques obtenues pour l'histone HIST1H2BM (humaine) sont données par le fit homogène du monosome au tétrasome (Fig. 4.2). Les autres valeurs sont extraites de la littérature (cf Tab. 3.2).

vons donc déplacer la ligne de séparation entre les phases ID et HD à  $\tilde{\omega} = 5$  où la densité peut être de l'ordre  $10^{-5}$ . On peut finalement considérer que l'effectivité du fit correspond au régime ID à 5 points près.

Nous avons la preuve, d'après les résultats présentés dans la figure, qu'au moins certains gènes historiques (rectangle vert), ainsi que la valeur moyenne du génome de la levure (*S. cerevisiae*, rectangle jaune), tombent en la phase ID. La moyenne du génome de *H. sapiens* semble toutefois se situer près de la limite entre la phase LD et la phase ID (ligne rouge) et pourrait être à la limite de la faisabilité de notre méthode de fit.

Il est intéressant de noter que dans le régime ID, la méthode que nous proposons pour extraire les paramètres cinétiques de la traduction exploite un profil de densité de monosomes qui est expérimentalement accessible malgré la probabilité peut-être beaucoup plus faible de trouver un monosome qu'un  $k$ -some avec  $k \approx \langle k \rangle_\infty \approx \tilde{\alpha}$  ( $P_1 \ll P_\alpha^\infty$ ). En ce sens, un séquençage "profond" (des milliards d'ARNm correspondant à des millions de cellules) est nécessaire pour étudier les monosomes car il en faut des milliers pour assurer une analyse statistiquement pertinente (voir la discussion de la section 6.5 du chapitre 3).

## 7 Conclusion et perspectives

Ce chapitre a été consacré à la comparaison des profils Ribo-Seq de  $k$ -somes avec ceux du modèle balistique décrits au chapitre précédent et à l'élaboration d'une méthode de déduction des valeurs précises des paramètres cinétiques de la traduction.

Nous avons d'abord montré quelles sont les informations apportées par les profils polysomes et les profils  $k$ -somes. Ces deux types de population apportent des informations complémentaires dans un régime de dégradation intermédiaire où celle-ci est assez élevée pour que les densités perdent leur couplage linéaire comme montré au chapitre 3.

En comparant qualitativement les profils Ribo-Seq sur les  $k$ -somes, nous déduisons que la

phénoménologie du système est bien décrite par le modèle balistique et nous démontrons pour la première fois la présence d'un effet de temps de vie fini sur des densités de ribosomes mesurées expérimentalement. Nous faisons ensuite une comparaison plus exhaustive en analysant l'ensemble des profils  $k$ -somes à notre disposition. Nous observons une très bonne concordance avec les profils de densités monosomes, qui se dégrade légèrement lorsque le nombre de ribosomes  $k$  augmente et quand on cherche les mêmes paramètres pour les quatre  $k$ -somes  $k = 1, 2, 3, 4$  d'un même gène. Une hypothèse est que l'augmentation de l'amplitude des fluctuations avec  $k$  diminue la qualité du fit homogène sur les données lissées qui a été effectué ici. Pour vérifier cette hypothèse, il sera nécessaire d'effectuer des fits des données non lissées avec le modèle inhomogène bien que cela soit beaucoup plus gourmand en temps. Si cette hypothèse n'est pas vérifiée, nous en déduisons qu'il existe un mécanisme important non considéré dans le modèle balistique qui engendre la variation d'un des paramètres en fonction du nombre de ribosomes présent sur l'ARNm. Il s'agit peut-être de corrélations non négligeables entre ribosomes. Pour étudier cette hypothèse, on pourra utiliser, dans un travail futur, des simulations de Monte Carlo d'un ensemble d'ARNm modélisé par le modèle  $\ell$ -TASEP inhomogène afin de comparer les densités de particules de ce modèle avec les données Ribo-Seq.

Nous élaborons finalement une méthode de déduction des paramètres cinétiques de la traduction qui utilise le profil monosome puisqu'ils sont très bien décrits par le modèle et le profil polysome ainsi que les mesures de temps de demi-vies des ARNm. Nous présentons les résultats d'un test de cette méthode sous forme de "fonction score" qui montre que comme prévu la qualité du fit décroît avec l'effet de durée de vie fini. En particulier, nous montrons que le ratio  $\mathcal{R}_1(0) \gtrsim 5$  est suffisant pour que notre méthode, utilisant l'algorithme L-BFGS, fonctionne. D'après les valeurs biologiques trouvées dans la littérature, notre méthode doit être effective pour de nombreux gènes. La méthode étant efficace pour des paramètres qui correspondent approximativement au régime ID défini dans le chapitre précédent, nous proposons finalement une adaptation du diagramme de phase du modèle balistique à la méthode de fit afin de voir pour quels paramètres cette méthode est effective. Pour poursuivre ce travail, il faudra d'abord évaluer la résistance de la méthode au bruit des données expérimentales, puis trouver une stratégie de traitement des profils Ribo-Seq qui présentent des densités nulles sur certains codons. En effet, sur certains codons, aucun ribosome n'a été détecté et un fit d'un tel profil engendrerait une vitesse infinie du ribosome ce qui n'est pas biologiquement réaliste.

## 8 Références

- DANA, A. et T. TULLER. 2012, «Determinants of Translation Elongation Speed and Ribosomal Profiling Biases in Mouse Embryonic Stem Cells», *PLOS Computational Biology*, vol. 8, n° 11, doi : 10.1371/journal.pcbi.1002755, p. e1002755, ISSN 1553-7358. URL <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002755>. 76
- DOMA, M. K. et R. PARKER. 2006, «Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation», *Nature*, vol. 440, n° 7083, doi :10.1038/nature04530, p. 561–564, ISSN 1476-4687. 84
- DUC, K. D. et Y. S. SONG. 2018, «The impact of ribosomal interference, codon usage, and exit tunnel interactions on translation elongation rate variation», *PLOS Genetics*, vol. 14, n° 1, doi : 10.1371/journal.pgen.1007166, p. e1007166, ISSN 1553-7404. URL <https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1007166>. 77
- HAN, Y., X. GAO, B. LIU, J. WAN, X. ZHANG et S.-B. QIAN. 2014, «Ribosome profiling reveals sequence-independent post-initiation pausing as a signature of translation», *Cell Research*, vol. 24, n° 7, doi :10.1038/cr.2014.74, p. 842–851, ISSN 1748-7838. URL <http://www.nature.com/articles/cr201474>, number : 7 Publisher : Nature Publishing Group. 76

- HEINTZ, N., H. L. SIVE et R. G. ROEDER. 1983, «Regulation of human histone gene expression : kinetics of accumulation and changes in the rate of synthesis and in the half-lives of individual histone mRNAs during the HeLa cell cycle», *Molecular and Cellular Biology*, vol. 3, n° 4, doi : 10.1128/mcb.3.4.539-550.1983, p. 539–550. URL <https://journals.asm.org/doi/10.1128/mcb.3.4.539-550.1983>, publisher : American Society for Microbiology. 79
- INGOLIA, N., L. LAREAU et J. WEISSMAN. 2011, «Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes», *Cell*, vol. 147, n° 4, doi : 10.1016/j.cell.2011.10.002, p. 789–802, ISSN 0092-8674. URL <http://www.sciencedirect.com/science/article/pii/S0092867411011925>. 76
- INGOLIA, N. T., J. A. HUSSMANN et J. S. WEISSMAN. 2019, «Ribosome Profiling: Global Views of Translation», *Cold Spring Harbor Perspectives in Biology*, vol. 11, n° 5, doi :10.1101/cshperspect.a032698, p. a032 698, ISSN 1943-0264. WOS :000466468300003. 76
- MORRIS, C., D. CLUET et E. P. RICCI. 2021, «Ribosome dynamics and mRNA turnover, a complex relationship under constant cellular scrutiny», *WIREs RNA*, vol. 12, n° 6, doi :10.1002/wrna.1658, p. e1658, ISSN 1757-7012. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/wrna.1658>, \_eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1002/wrna.1658>. 84
- SHAROVA, L. V., A. A. SHAROV, T. NEDOREZOV, Y. PIAO, N. SHAIK et M. S. H. KO. 2009, «Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells», *DNA research : an international journal for rapid publication of reports on genes and genomes*, vol. 16, n° 1, doi :10.1093/dnares/dsn030, p. 45–58, ISSN 1756-1663. 76
- SIMMS, C. L., L. L. YAN et H. S. ZAHER. 2017, «Ribosome Collision Is Critical for Quality Control during No-Go Decay», *Molecular Cell*, vol. 68, n° 2, doi :10.1016/j.molcel.2017.08.019, p. 361–373.e5, ISSN 1097-2765. URL <http://www.sciencedirect.com/science/article/pii/S1097276517306160>. 84
- SZAVITS-NOSSAN, J. et L. CIANDRINI. 2020, «Inferring efficiency of translation initiation and elongation from ribosome profiling», *Nucleic Acids Research*, doi :10.1093/nar/gkaa678, ISSN 0305-1048, 1362-4962. URL <https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkaa678/5895331>. 84
- VIRTANEN, P., R. GOMMERS, T. E. OLIPHANT, M. HABERLAND, T. REDDY, D. COURNAPEAU, E. BUROVSKI, P. PETERSON, W. WECKESSER, J. BRIGHT, S. J. VAN DER WALT, M. BRETT, J. WILSON, K. J. MILLMAN, N. MAYOROV, A. R. J. NELSON, E. JONES, R. KERN, E. LARSON, C. J. CAREY, Í. POLAT, Y. FENG, E. W. MOORE, J. VANDERPLAS, D. LAXALDE, J. PERKTOLD, R. CIMRMAN, I. HENRIKSEN, E. A. QUINTERO, C. R. HARRIS, A. M. ARCHIBALD, A. H. RIBEIRO, F. PEDREGOSA, P. VAN MULBREGT et SCIPY 1.0 CONTRIBUTORS. 2020, «SciPy 1.0 : Fundamental Algorithms for Scientific Computing in Python», *Nature Methods*, vol. 17, doi :10.1038/s41592-019-0686-2, p. 261–272. 77, 85

# Conclusion

En biologie, la traduction reste encore peu étudiée par rapport à la transcription. C'est pourtant une étape extrêmement importante de l'expression génétique puisque c'est le moment où l'information stockée dans les gènes est traduite en matière fonctionnelle, les protéines, qui permettent à la cellule de vivre puisqu'elles assurent toutes ses fonctions. D'un point de vue physico-chimique, la traduction est modélisée depuis 1963, peu de temps après sa découverte dans les années 1950. Elle a donc toujours été considérée par les physiciens et nourri la physique statistique hors équilibre en tant que phénomène de transport stochastique dirigé à une dimension et hors équilibre thermodynamique.

Dans le chapitre 2, nous avons présenté le modèle paradigmatique de ce phénomène, le modèle TASEP pour en montrer les propriétés principales et jeter les bases pour l'étude de ses variantes plus proches de la traduction génétique. Celles-ci incluent une particule étendue sur plusieurs sites qui se déplace toujours d'un seul site à la fois et des taux de saut inhomogènes le long du réseau. Ces caractéristiques induisent des effets d'exclusions intéressants et encore mal compris. En particulier, nous faisons remarquer un impact non négligeable de la particule étendue en entrée et en sortie du réseau dans la phase basse densité pour laquelle l'exclusion est pourtant la plus faible. En effet, nous avons montré, dans le cas des taux de sauts homogènes, que considérer l'exclusion à l'entrée suffit à approximer la densité du modèle  $\ell$ -TASEP lorsque le ratio du taux d'initiation sur le taux de saut  $\alpha/p \leq 0,1$  alors qu'elle s'éloigne considérablement de la densité TASEP lorsque la longueur de la particule  $\ell$  et  $\alpha/p$  augmentent ( $0,01 < \alpha/p \leq 0,1$ ). Il n'est alors pas étonnant que la cellule mette en place une stratégie d'évitement des collisions à l'entrée de l'ARN messenger comme le montrent plusieurs études, d'autant plus que le courant est également abaissé par l'exclusion à l'entrée. Pour  $\alpha/p > 0,1$ , l'exclusion en sortie du réseau devient elle aussi non négligeable en engendrant l'augmentation considérable de la couche limite de l'effet de bord. Par les équations de champ moyen du modèle  $\ell$ -TASEP, nous montrons que l'amplitude de cet effet de bord dépend majoritairement de  $\alpha/p$  seulement.

La recherche expérimentale est nécessaire à l'étude de la traduction et ce n'est que récemment qu'elle s'est particulièrement développée. Les expériences de séquençage haut débit de *ribosome profiling* (ou Ribo-Seq), ont été mises au point au début des années 2000 et ont permis de faire considérablement avancer la connaissance du processus de traduction génétique. Cependant, l'interprétation de ce type d'expérience n'est pas évidente. La durée de vie finie des ARNm est rarement prise en compte dans les modèles. Pourtant, cela implique qu'une partie importante des ARNm ne sont pas dans leur état stationnaire puisqu'ils commencent juste à être traduits par le ribosome tandis que d'autres ont été dégradés. En considérant un processus de renouvellement qui maintient une population constante d'ARNm et en simplifiant le processus de transport des ribosomes sur l'ARNm par un mouvement balistique (i.e. déterministe), nous étudions l'effet de durée de vie finie sur un tel ensemble dans le chapitre 3. Nous introduisons pour cela l'étude d'ARNm séparés par leur nombre de ribosomes allant de 1 à 4, les monosomes, disomes, trisomes et tétrasomes (ou  $k$ -somes,  $k = 1, 2, 3, 4$ ). Nous montrons en particulier que, lorsque le temps de vie est infini, les densités des  $k$ -somes sont linéairement couplées entre elles, et avec la densité polysome. Considérer un temps de vie fini va engendrer le découplage de ces densités par une décroissance de type exponentielle le long du filament qui sera d'autant plus importante que  $k$  est petit, et sera négligeable pour le polysome. Nous définissons un diagramme de phase présentant trois régimes de dégradation en fonction des deux paramètres adimensionnels régissant l'effet de

temps de vie fini sur les profils de densités :  $\alpha \mathcal{T}(L)$  et  $\omega \mathcal{T}(L)$  où  $\alpha$  est le taux d'initiation de la traduction,  $\omega$  le taux de dégradation de l'ARNm et  $\mathcal{T}(L)$  le temps que met le ribosome pour traverser l'ARNm. Ces trois régimes sont : le régime de faible dégradation (LD) où ni les densités polysomes, ni les densités  $k$ -some ne ressentent l'effet de temps de vie fini, le régime de dégradation intermédiaire (ID) où seules les densités  $k$ -somes ressentent l'effet de temps de vie fini et enfin le régime de haute dégradation (HD) où l'effet de temps de vie fini devient non négligeable aussi sur les densités polysomes. Grâce aux mesures déjà existantes dans la littérature chez la levure *S. cerevisiae* et l'humain, nous situons le régime biologique au niveau du régime de dégradation intermédiaire.

Dans le chapitre 4, nous analysons les données Ribo-Seq sur les  $k$ -somes produites par nos collaborateurs biologistes de l'IGF et bioinformaticiens du LIRMM. Nous comparons ces données avec le modèle balistique décrit au chapitre 3 afin d'en déduire le niveau de correspondance entre le modèle et le système biologique mesuré. Grâce à une comparaison qualitative, nous démontrons pour la première fois la présence d'un effet de temps de vie fini des ARNm sur des densités de ribosomes mesurées expérimentalement. Une analyse plus quantitative montre en revanche qu'un des paramètres de la traduction semble varier avec le nombre de ribosomes présent sur l'ARNm. Une analyse plus précise sera nécessaire pour confirmer ce résultat dans un travail futur. Si ce résultat est confirmé, il ouvrira la voie à la recherche d'une caractéristique manquante pour la modélisation précise des expériences Ribo-Seq. Nous pensons qu'il pourrait s'agir de l'exclusion stérique ou autre type d'interaction entre les ribosomes. Bien que la cellule semble mettre en place une stratégie pour éviter les collisions entre ribosomes, nous ne savons pas précisément à quel degré (Est-il prévenu seulement les encombrements entre ribosomes, ou le moindre contact entre ribosome est-il évité ?) et il peut être important de considérer des particules qui occupent plusieurs sites sans se recouvrir.

Nous proposons également une méthode de déduction des paramètres cinétiques de la traduction : le taux d'initiation  $\alpha$ , et les taux d'élongation  $p_i$  pour chaque codon. Pour cela, nous montrons d'abord que le profil de densité polysome n'apporte pas assez d'information par rapport au nombre d'inconnues, et nous choisissons d'associer l'information de la densité monosome, qui est la mieux décrite par le modèle et la plus découplée linéairement de la densité polysome. Nous tirons parti du fait que la densité monosome est la densité qui ressent le plus l'effet de temps de vie fini, tandis qu'il est négligeable sur la densité polysome. Cette méthode est donc effective principalement dans le régime de dégradation intermédiaire déterminé au chapitre 3. Elle pourra donc être utilisée pour analyser une grande proportion des ARNm des cellules biologiques. Nous pensons qu'une méthode quantitative servira à déchiffrer les mécanismes microscopiques de la traduction ce qui aiderait à comprendre le rôle dans l'expression des gènes dans l'apparition de pathologies comme le cancer, les maladies neurodégénératives et l'obésité.

En conclusion, nous pouvons dire que nous avons travaillé à la modélisation physico-mathématique de la traduction génétique en considérant deux caractéristiques qui se révèlent être importantes en particulier pour les profils de densités de ribosomes sur les ARN messagers et donc pour l'interprétation des profils de données Ribo-Seq. Il s'agit d'une part de l'effet de la particule étendue en basse densité, et d'autre part de l'effet de temps de vie fini sur un ensemble d'ARN messagers traduit de manière non synchronisée. Nous avons effectué une analyse statistique des données de séquençage haut débit, Ribo-Seq, sur les monosomes, disomes, trisomes et tétratsomes, qui nous a permis de mettre en évidence pour la première fois l'effet de temps de vie fini prédit par notre modèle sur les densités de ribosomes, avec un très bon accord pour les monosomes. Dans un travail futur, il serait intéressant de construire un modèle qui tient compte à la fois du temps de vie fini et des interactions entre ribosomes afin de voir s'il décrit plus quantitativement les profils de densités Ribo-Seq mesurés sur les  $k$ -somes à  $k > 1$ .

# Annexes

## 1 Calculs annexes

### 1.1 Moyenne sur les positions

Par définition, si  $t_o - t_b \geq \mathcal{T}(L)$  alors  $t_1$  est uniformément distribué sur l'intervalle  $[t_o - \mathcal{T}(L), t_o]$  et a la densité de probabilité  $q(t_1) = 1/\mathcal{T}(L)$  sur cet intervalle. Donc,

$$\begin{aligned}
 \langle \delta(x - x_1) \rangle &= \langle \delta(x - x(t_o - t_1)) \rangle \\
 &= \int_{t_o - \mathcal{T}(L)}^{t_o} \delta(x - x(t_o - t_1)) \frac{dt_1}{\mathcal{T}(L)} \\
 &= \int_0^{\mathcal{T}(L)} \delta(x - x(t)) \frac{dt}{\mathcal{T}(L)} \\
 &= \int_0^{\mathcal{T}(L)} \frac{\delta(\mathcal{T}(x) - t)}{\dot{x}(\mathcal{T}(x))} \frac{dt}{\mathcal{T}(L)} \\
 &= \frac{1}{p(x)\mathcal{T}(L)}
 \end{aligned} \tag{A.1}$$

où nous avons utilisé  $\dot{x}(t) = p(x(t))$  et où  $\mathcal{T}(x)$  est le temps pour qu'un ribosome entré sur l'ARNm à  $t = 0$  met pour atteindre la position  $x$ . Nous l'avons défini en (3.47).

Si  $t_o - t_b \leq \mathcal{T}(L)$ , alors  $t_1$  est distribué uniformément sur l'intervalle  $[t_b, t_o]$  et a la densité de probabilité  $q(t_1) = 1/(t_o - t_b)$  sur cet intervalle. Donc,

$$\begin{aligned}
 \langle \delta(x - x_1) \rangle &= \langle \delta(x - x(t_o - t_1)) \rangle \\
 &= \int_{t_b}^{t_o} \delta(x - x(t_o - t_1)) \frac{dt_1}{t_o - t_b} \\
 &= \int_0^{t_o - t_b} \delta(x - x(t)) \frac{dt}{t_o - t_b} \\
 &= \int_0^{t_o - t_b} \frac{\delta(\mathcal{T}(x) - t)}{\dot{x}(\mathcal{T}(x))} \frac{dt}{t_o - t_b} \\
 &= \frac{1}{p(x)(t_o - t_b)} H(x(t_o - t_b) - x)
 \end{aligned} \tag{A.2}$$

### 1.2 Approximation de la fonction gamma

Pour les entiers non négatifs  $k$ , la fonction gamma peut être exprimée en terme de fonctions élémentaires comme

$$\frac{\Upsilon(k+1, z)}{k!} = 1 - e^{-z} \sum_{n=0}^k \frac{z^n}{n!} \tag{A.3}$$

$$= 1 - \frac{z^k e^{-z}}{k!} \sum_{m=0}^k \frac{k!}{(k-m)! z^m}. \tag{A.4}$$

([ABRAMOWITZ et STEGUN, 1972], Eq. 6.5.32). Pour  $k \ll z$ , c'est une bonne approximation de ne garder que les deux premiers termes dans la somme de (A.3) :

$$\frac{\Upsilon(k+1, z)}{k!} \approx 1 - \frac{z^k e^{-z}}{k!} \left(1 + \frac{k}{z}\right), \quad (\text{A.5})$$

un résultat exact pour  $k = 0$  et 1. Pour  $k = z \gg 1$ , l'approximation ci-dessus doit être remplacée par

$$\frac{\Upsilon(z+1, z)}{z!} \approx \frac{1}{2} \left(1 - \frac{4}{3\sqrt{2\pi z}}\right). \quad (\text{A.6})$$

([ABRAMOWITZ et STEGUN, 1972], Eq. 6.5.35).

### 1.3 Inversion des populations

Lorsque  $\tilde{\omega}$  augmente et atteint une valeur de croisement  $\tilde{\omega}_{\text{co}}$ , la valeur de  $P_0$  devient plus élevée que le maximum près de  $k = \tilde{\alpha}$ , ce qui signifie que les  $k$ -somes à petit  $k$  prennent le dessus sur ceux dont le nombre moyen de ribosomes est proche de  $\langle k \rangle_\infty = \tilde{\alpha}$ . Pour déterminer approximativement  $\tilde{\omega}_{\text{co}}$ , on note que, pour une dégradation finie mais pas trop élevée, la position du maximum de  $P_k$  reste proche de  $k \approx \tilde{\alpha}$ . Nous pouvons donc estimer  $\tilde{\omega}_{\text{co}}$  en égalisant  $P_{\tilde{\alpha}}$  avec  $P_0 \approx \tilde{\omega}/\tilde{\alpha} \ll 1$  :

$$P_{\tilde{\alpha}} \approx \left(\frac{\tilde{\omega}}{2\tilde{\alpha}} + \frac{1}{\sqrt{2\pi\tilde{\alpha}}}\right) e^{-\tilde{\omega}} = P_0 \approx \frac{\tilde{\omega}}{\tilde{\alpha}}, \quad (\text{A.7})$$

où nous avons utilisé (A.6) pour obtenir l'approximation de  $P_{\tilde{\alpha}}$ . Résoudre (A.7) pour  $\tilde{\omega}$  conduit à une relation entre  $\tilde{\omega}_{\text{co}}$  et  $\tilde{\alpha}$  valide pour  $\tilde{\alpha} > 2$  :

$$\tilde{\alpha} \approx 2\pi\tilde{\omega}_{\text{co}}^2 \left(e^{\tilde{\omega}_{\text{co}}} - \frac{1}{2}\right)^2. \quad (\text{A.8})$$

Cette relation révèle que  $\tilde{\omega}_{\text{co}}$  augmente lentement avec  $\tilde{\alpha}$  et que  $\tilde{\omega}_{\text{co}} \sim 1$  pour  $10 < \tilde{\alpha} < 100$ , ce qui permet d'obtenir une estimation plus simple mais toujours valide,

$$\tilde{\omega}_{\text{co}} = \ln \left( \sqrt{\frac{\tilde{\alpha}}{2\pi\tilde{\omega}_{\text{co}}} + \frac{1}{2}} \right) \approx \ln \left( \sqrt{\frac{\tilde{\alpha}}{2\pi} + \frac{1}{2}} \right). \quad (\text{A.9})$$

### 1.4 Approche discrète

#### 1.4.1 Densités discrètes : définition et normalisation

Le modèle balistique a été formulé dans le texte principal en espace continu  $x \in [0, L]$ , où  $L$  est la longueur de l'ARNm en unités de codons. Dans le domaine biologique de l'ARNm, cependant, les codons sont des unités séparées qui doivent être considérées comme des sites formant un réseau avec  $L$  sites. Les densités de ribosomes (pour les polysomes comme pour les  $k$ -somes) sont alors naturellement définies sur ce réseau comme des quantités discrètes. Afin de retrouver ce caractère discret, nous définirons simplement la densité au site  $i$  comme étant l'intégrale de la densité continue sur l'intervalle précédent  $x \in [i-1, i]$ , soit

$$\rho_i = \int_{i-1}^i \rho(x) dx. \quad (\text{A.10})$$

La normalisation de cette densité discrète est donc la même que celle de sa contrepartie continue

$$\sum_{i=1}^L \rho_i = \int_0^L \rho(x) dx. \quad (\text{A.11})$$

### 1.4.2 Densité polysome discrète

La densité du polysome est fournie par l'équation (3.51). En utilisant (A.10), nous obtenons sa contrepartie discrète,

$$\rho_i = \frac{\alpha}{\omega} \left( e^{-\omega \mathcal{T}(i-1)} - e^{-\omega \mathcal{T}(i)} \right), \quad (\text{A.12})$$

où le temps intégré pris par un ribosome pour aller du site 1 à  $i$  ( $i \in [1, L]$ ) est maintenant :

$$\mathcal{T}_i = \int_0^i \frac{dx}{p(x)} = \sum_{j=1}^i \frac{1}{p_j}, \quad (\text{A.13})$$

où  $p(x)$  est considérée comme une fonction constante par morceaux :  $p(x) = p_i$  pour  $x \in ]i-1, i]$ .

### 1.4.3 Densité $k$ -some discrète

L'expression de la densité  $k$ -some au site  $i$  est donnée par

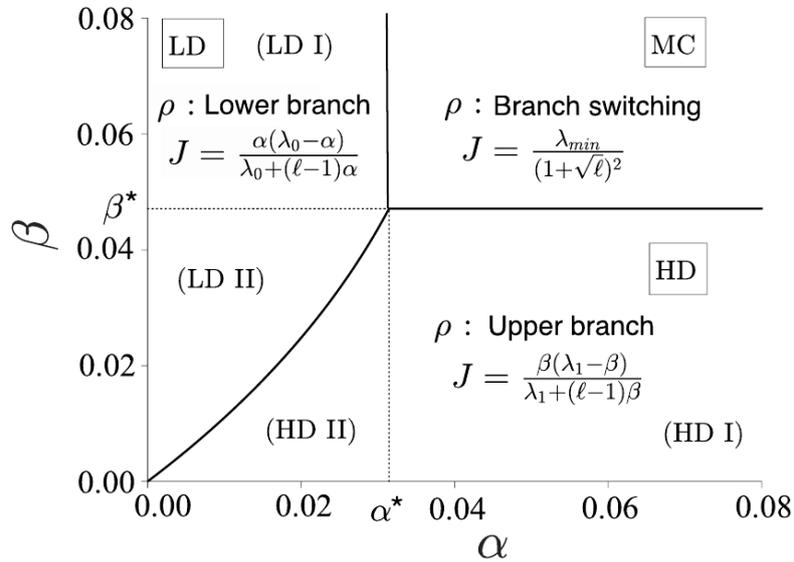
$$\rho_{k;i} = \left( \frac{\tilde{\alpha}}{\tilde{\alpha} + \tilde{\omega}} \right)^k \frac{\tilde{\omega}}{P_k(k-1)!} \left( \frac{\Upsilon(k, (\tilde{\alpha} + \tilde{\omega}))}{\tilde{p}_i} - (\mathcal{I}_i - \mathcal{I}_{i-1}) \right) + \frac{1}{\tilde{p}_i P_k(k-1)!} \tilde{\alpha}^k e^{-(\tilde{\alpha} + \tilde{\omega})}, \quad (\text{A.14})$$

où

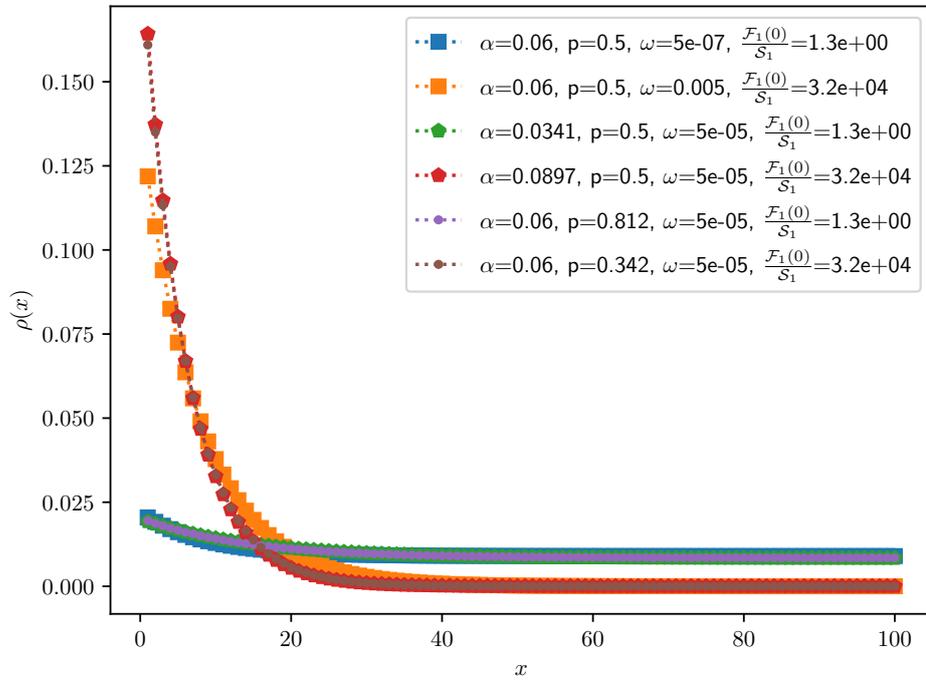
$$\mathcal{I}_i = \tau_i \Upsilon(k, (\tilde{\alpha} + \tilde{\omega}) \tau_i) - \frac{\Upsilon(k+1, (\tilde{\alpha} + \tilde{\omega}) \tau_i)}{\tilde{\alpha} + \tilde{\omega}}. \quad (\text{A.15})$$

avec  $\tilde{\alpha} = \alpha \mathcal{T}_L$ ,  $\tilde{\omega} = \omega \mathcal{T}_L$ ,  $\tilde{p} = p \mathcal{T}_L$  et  $\tau_i = \mathcal{T}_i / \mathcal{T}_L$ . L'expression de  $P_k$  reste la même que dans la cas continu puisqu'elle ne dépend pas de la position. Dans la limite du temps de vie infini ( $\omega \rightarrow 0$ ) la densité  $k$ -some devient  $\rho_{k;i}^\infty = k / \tilde{p}_i$  et la densité polysome  $\rho_i^\infty = \alpha / p_i$ .

## 2 Tableaux et figures annexes



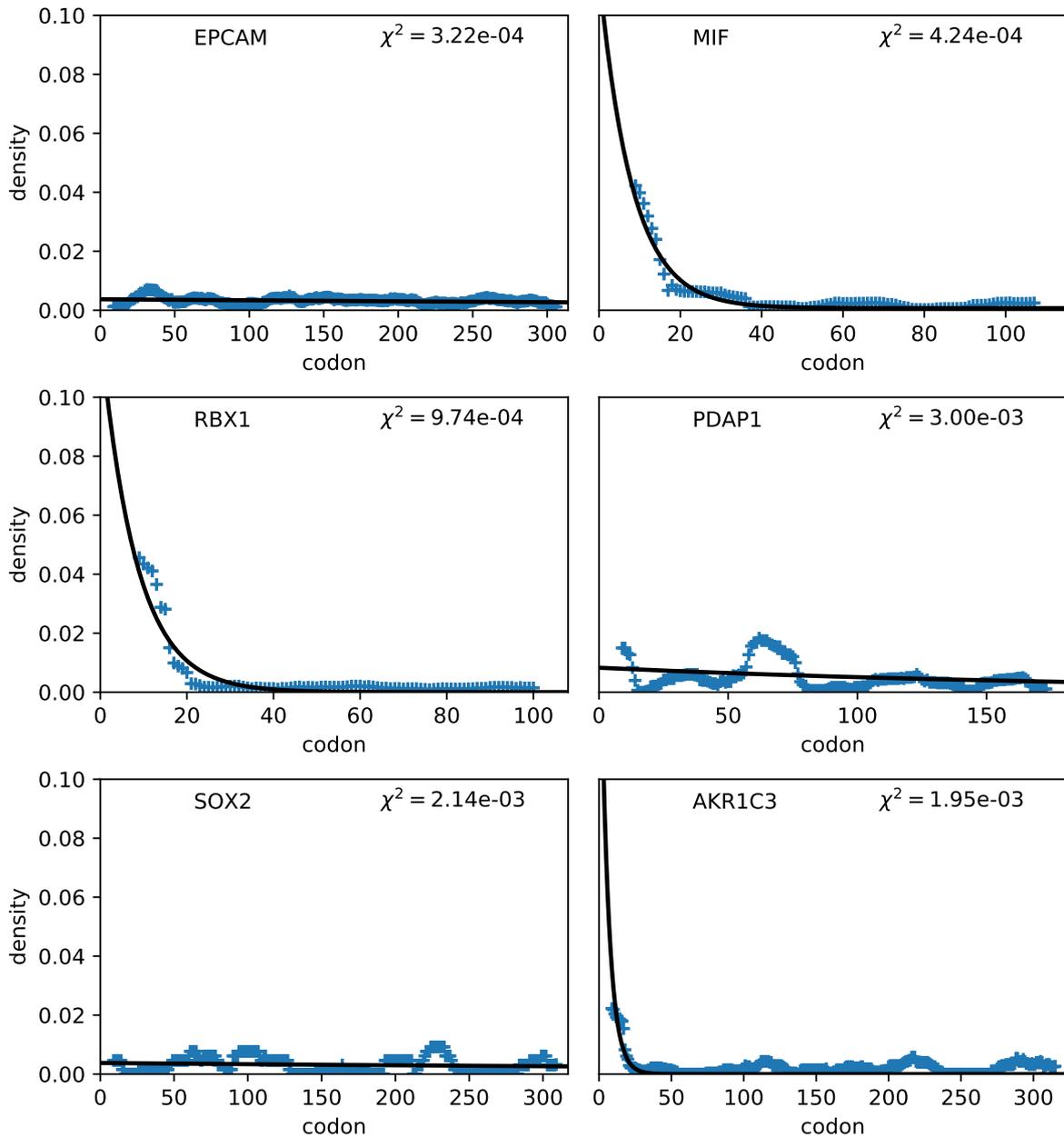
**FIGURE A.1** – Diagramme de phase du  $\ell$ -TASEP inhomogène donné par ERDMANN-PHAM et al. [2020].  $\lambda$  est une version différentiable des taux de saut.  $\lambda_0$  correspond au taux de saut à l'entrée du filament et  $\lambda_1$  à la sortie du filament,  $\lambda_{min}$  est le taux de saut minimal du filament. "The phase diagram is completely determined by  $\lambda_0, \lambda_1, \lambda_{min}$ , and  $\ell$ . In this example,  $(\lambda_0, \lambda_1, \lambda_{min}, \ell) = (0.9; 0.3; 0.1; 10)$ . All phase transitions are continuous in  $J$ , and, unless  $\lambda_{min}$  coincides with  $\lambda_0$  or  $\lambda_1$ , discontinuous in  $\rho$ ." [ERDMANN-PHAM et al., 2020]



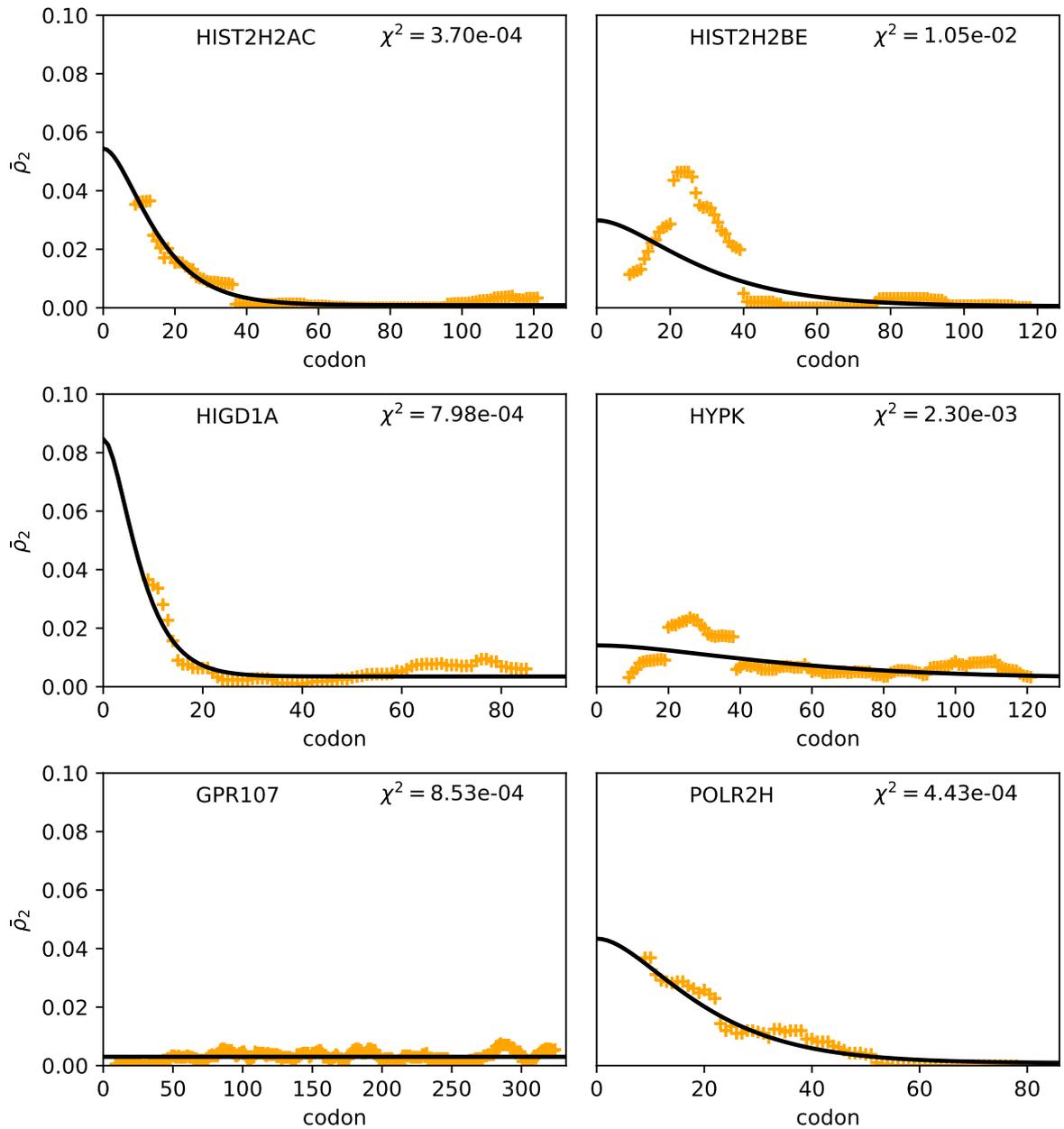
**FIGURE A.2** – Densités monomes pour  $\mathcal{R}_1(0) = 1.3$  correspondant à l'effet de durée de vie finie faible sur les  $k$ -somes et pour  $\mathcal{R}_1(0) = 1.3 \times 10^4$  correspondant à l'effet de durée de vie finie élevé sur les  $k$ -somes. Les régimes peuvent être obtenus par différentes valeurs de  $\alpha$ ,  $p$  et  $\omega$  alors que  $\mathcal{R}_1(0)$  reste constant.

$k$	<i>low coverage</i>	<i>intermediate coverage</i>	<i>high coverage</i>
1	2642	186	171
2	443	47	41
3	516	56	41
4	492	50	42
1-4	268	17	36

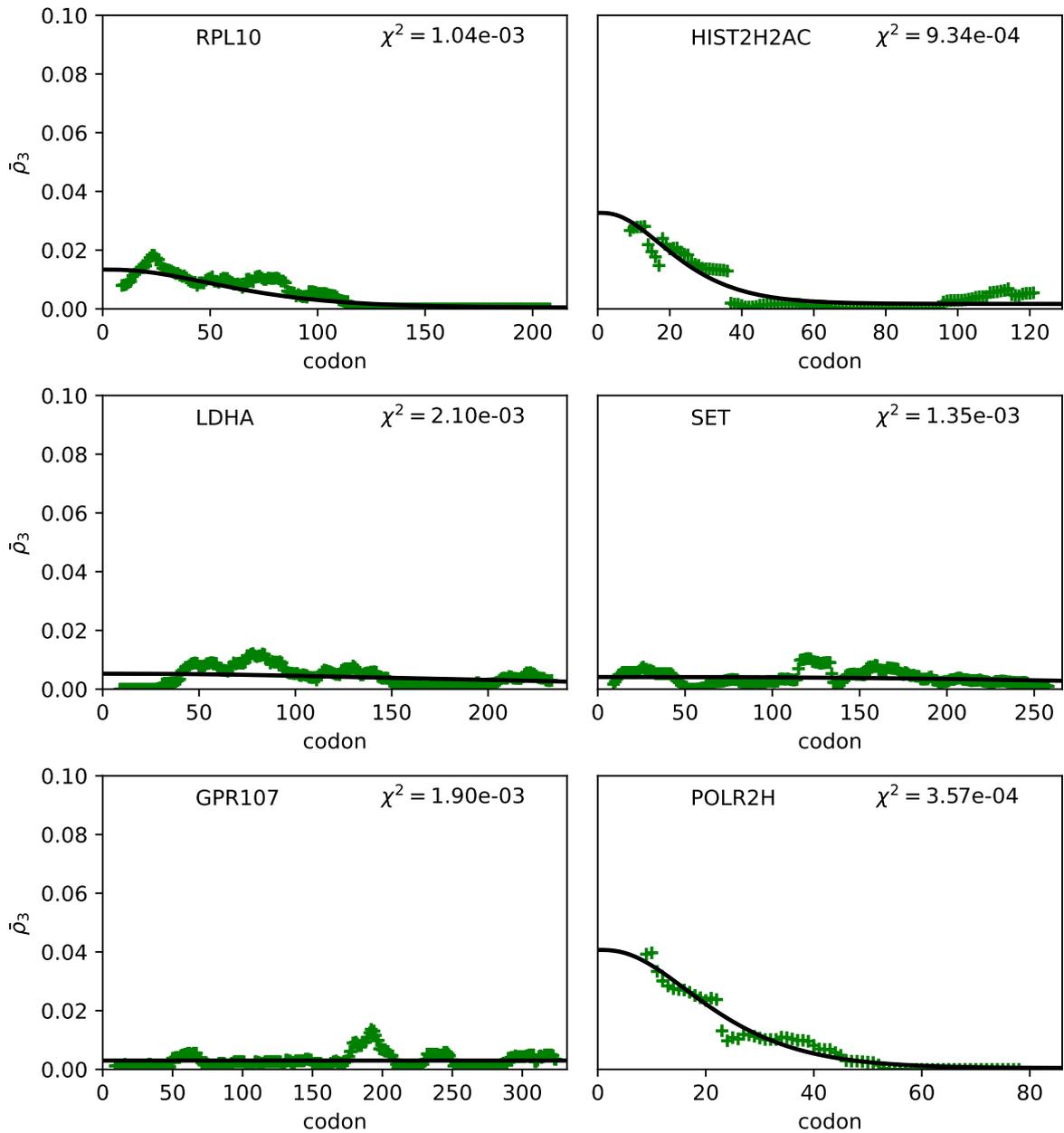
**TABLEAU A.1** – Nombre des gènes de l'expérience Ribo-Seq sur les  $k$ -somes analysés dans la construction des histogrammes 4.4 et 4.5.



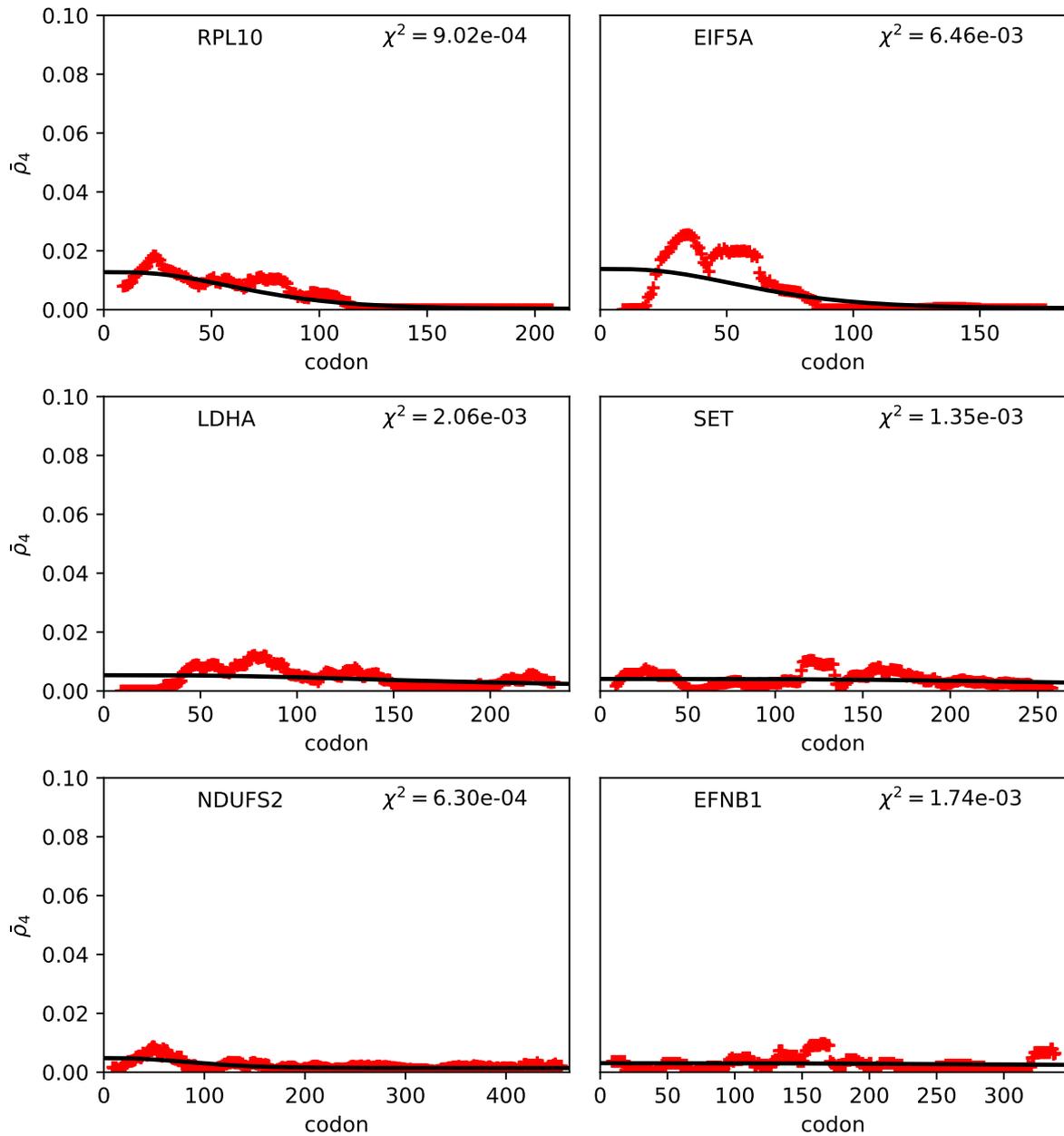
**FIGURE A.3** – Densités Ribo-Seq de monosomes fittées par le modèle balistique homogène. Les données sont lissées par une fenêtre glissante qui fait la moyenne sur 19 codons. Sur la ligne du haut, il s'agit de deux gènes à haute couverture, sur la deuxième ligne, de gènes à couverture intermédiaire, et sur la ligne du bas de gènes à faible couverture. Sur chaque graphe, est indiqué en haut à gauche le nom du gène, et en haut à droite l'écart entre les données lissées et le modèle.



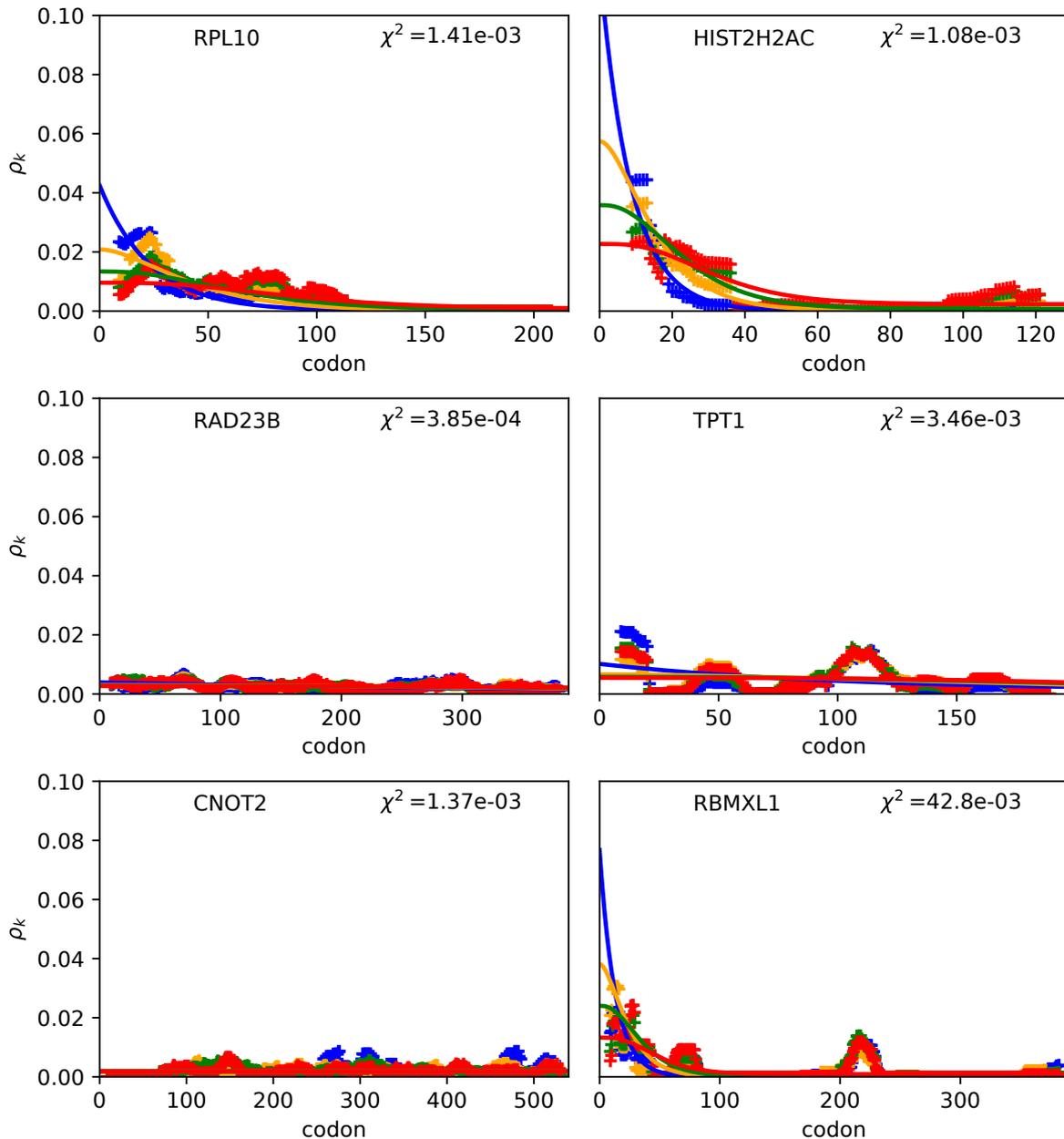
**FIGURE A.4** – Densités Ribo-Seq de disomes fittées par le modèle balistique homogène. Les données sont lissées par une fenêtre glissante qui fait la moyenne sur 19 codons. Sur la ligne du haut, il s’agit de deux gènes à haute couverture, sur la deuxième ligne, de gènes à couverture intermédiaire, et sur la ligne du bas de gènes à faible couverture. Sur chaque graphe, est indiqué en haut à gauche le nom du gène, et en haut à droite l’écart entre les données lissées et le modèle.



**FIGURE A.5** – Densités Ribo-Seq de trisomes fittées par le modèle balistique homogène. Les données sont lissées par une fenêtre glissante qui fait la moyenne sur 19 codons. Sur la ligne du haut, il s'agit de deux gènes à haute couverture, sur la deuxième ligne, de gènes à couverture intermédiaire, et sur la ligne du bas de gènes à faible couverture. Sur chaque graphe, est indiqué en haut à gauche le nom du gène, et en haut à droite l'écart entre les données lissées et le modèle.



**FIGURE A.6** – Densités Ribo-Seq de tétrasomes fittées par le modèle balistique homogène. Les données sont lissées par une fenêtre glissante qui fait la moyenne sur 19 codons. Sur la ligne du haut, il s’agit de deux gènes à haute couverture, sur la deuxième ligne, de gènes à couverture intermédiaire, et sur la ligne du bas de gènes à faible couverture. Sur chaque graphe, est indiqué en haut à gauche le nom du gène, et en haut à droite l’écart entre les données lissées et le modèle.



**FIGURE A.7** – Densités Ribo-Seq de  $k$ -somes, pour  $k = 1$  à 4 fittées par le modèle balistique homogène. Pour chaque paramètre, une seule valeur commune pour les quatre  $k$ -somes est recherchée. Les données sont lissées par une fenêtre glissante qui fait la moyenne sur 19 codons. Sur la ligne du haut, il s'agit de deux gènes à haute couverture, sur la deuxième ligne, de gènes à couverture intermédiaire, et sur la ligne du bas de gènes à faible couverture. Sur chaque graphe, est indiqué en haut à gauche le nom du gène, et en haut à droite l'écart entre les données lissées et le modèle.

### 3 Références

ABRAMOWITZ, M. et I. A. STEGUN. 1972, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, tenth printing éd., U.S. Government Printing Office, Washington, DC, USA. II

ERDMANN-PHAM, D. D., K. DAO DUC et Y. S. SONG. 2020, «The Key Parameters that Govern Translation Efficiency», *Cell Systems*, vol. 10, n° 2, doi :10.1016/j.cels.2019.12.003, p. 183–192.e6, ISSN 2405-4712. URL <http://www.sciencedirect.com/science/article/pii/S2405471219304648>. III