Stochasticité dans la réponse d'individus bactériens à une perturbation : étude dynamique

Edith Grac

▶ **To cite this version:**

Edith Grac. Stochasticité dans la réponse d'individus bactériens à une perturbation : étude dynamique. Autre [cond-mat.other]. Université de Grenoble, 2012. Français. NNT : 2012GRENY021 . tel-00728293

HAL Id: tel-00728293

https://theses.hal.science/tel-00728293

Submitted on 5 Sep 2012

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.

UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Physique pour les Sciences du Vivant**

Arrêté ministériel : 7 Août 2006

Présentée par

Grac Edith

Thèse dirigée par **Mihalcescu Irina**

préparée au sein **Laboratoire Interdisciplinaire de Physique**
et de **l'Ecole Doctorale de Physique de Grenoble**

Stochasticité dans la Réponse d'Individus bactériens à une Perturbation : Etude Dynamique.

Thèse soutenue le **16 Février 2012**,
devant le jury composé de :

Mr, Hans Geiselmann
Professeur UJF, Président
Mr, Didier Chatenay
DR1 CNRS, Rapporteur
Mr, Hervé Isambert
DR2 CNRS, Rapporteur
Mr, Olivier Gandrillon
DR2 CNRS, Examinateur
Mme, Irina Mihalcescu
Professeur UJF, Directeur de thèse

Contenu

Deux cellules génétiquement identiques placées dans le même environnement peuvent afficher des comportements différents en raison de l'aléa inhérent à l'expression génique. Cette stochasticité d'expression génique est de plus en plus étudiée, mais les études se limitent souvent à des environnements extracellulaires stables. On se demande alors comment évolue cet aléa dans un environnement dynamique. Dans ce but, nous nous proposons d'étudier l'évolution du bruit d'expression génique lorsque des bactéries *E. Coli* transitent d'un état de stress nutritionnel vers un état de croissance exponentielle. Cette transition d'état est initiée par l'ajout de glucose au milieu de culture défaillant en sources carbonnées. Elle est gérée par un réseau de régulation complexe, et nous nous focalisons ici sur un noeud clé du réseau : la protéine GyrA.

L'état transcriptionnel du module d'expression génique GyrA est sondé par un système de rapporteurs transcriptionnels plasmidiques. La vidéo-microscopie de fluorescence permet alors de mesurer ce niveau d'activation sur chaque cellule et pour chaque instant de la réponse.

Nous décrivons ici un protocole expérimental qui permet de réaliser ces mesures : nous présentons la procédure suivie pour préparer les échantillons de cellules individuelles, et nous donnons une description détaillée du système d'imagerie utilisé. Les images collectées sont ensuite nettoyées des différents bruits instrumentaux qu'elles contiennent, et les données de fluorescence des cellules individuelles sont extraites. Elles seront corrigées du bruit d'autofluorescence cellulaire avant d'être analysées.

Comme notre protocole expérimentale contient un nombre important de paramètres, nous avons réalisé une première étude qui vise à identifier les paramètres critiques. Une fois identifiés, ceux-ci ont été réglés avec minutie pour obtenir la reproductibilité expérimentale. A partir de là, on a pu quantifier la variabilité intercellulaire (*i.e.* le bruit d'expression génique) en traçant, pour chaque instant de la réponse, la distribution des densités de fluorescence cellulaire au sein de la population.

Les premiers résultats font alors apparaître des distributions bimodales. Dans ces distributions, la première bosse est générée par des cellules qui ne se divisent pas pendant la durée de l'expérience. Ce sont, soit des cellules mortes, soit des cellules qui répondent plus tardivement à l'ajout de glucose. Par la suite, elles ne sont pas considérées. Dans ces conditions, les résultats quantitatifs montrent que tout le long de la reprise de croissance, le bruit d'expression génique peut être décrit par une distribution gamma dont seuls les paramètres changent au cours du temps.

Nous avons alors chercher à exploiter cette signature du bruit pour extraire des informations quantitatives sur la dynamique du module d'expression génique GyrA. Dans ce but, nous avons étudié par simulation un modèle simplifié de l'expression génique (transcription-traduction). Les paramètres de ce modèle sont les taux de réaction, et ils sont suceptibles d'évoluer dans le temps. Pour réaliser les simulations, nous avons donc implémenté une version modifiée de l'algorithme de Gillepsie qui intégre ces variations temporelles.
Les paramètres du modèle ont été inférrés à partir de données expérimentales, mais deux inconnues subsistent : les durées des rampes qui vairent les taux de production k_m et k_p. Ces deux paramètres inconnues du modèle sont ajustés sur les données expérimentales.
Les résultats des simulations montrent alors que le modèle simplifié proposé permet de décrire globalement les résultats expérimentaux obtenus. On peut donc, à partir de nos mesures de bruit, déterminer la dynamique des taux de production du module GyrA lors de la transition stress nutritionnel/croissance exponentielle. Le modèle utilisé pour tirer ces conclusions est par ailleurs discuté, avec une dernière étude sur les sources extrinsèques de bruit négligées dans notre modèle.

Remerciements

En premier et avant tout, je tiens à remercier mes parents, et plus largement ma famille : ce sont eux qui m'ont donné les armes nécessaires pour en arriver là.

Je tiens aussi à remercier Irina pour m'avoir accueillie, guidée, s'être montrée toujours présente pour discuter les résultats et la littérature, et pour m'avoir toujours encouragée. J'ai beaucoup appris avec elle, et elle m'a montré un soutien infaillible tout le long de cette thèse : je lui en suis profondément reconnaissante.

J'ai aussi beaucoup apprécier les réunions avec nos collaborateurs biologistes du LAPM et bioinformaticiens de l'INRIA. Merci pour votre implication et pour ces "tables rondes" où tous les drapeaux ont une place. Je tiens ici à remercier particuliérement Corinne Pinel et Hans Geiselmann pour leur implication totale dans la construction des doubles rapporteurs chromosomaux.

Mes remerciements vont également à Bahram Ouchmandzadeh dont l'aide a été précieuse pour simuler mon modèle. Je tiens aussi à exprimer ma gratitude à Marc Joyeux qui m'a soutenue et beaucoup aidée dans l'organisation du temps de rédaction. Je suis reconnaissante envers Noêl Scaramozzino pour toutes les fautes d'orthographe qu'il m'a corrigées.

Je souhaite aussi remercier Didier Chatenay et Hervé Isambert pour avoir accepter d'être les rapporteurs de cette thèse, et plus largement, pour le vif intérêt qu'ils ont porté à mes travaux de recherche. Un grand merci aussi à Olivier Gandrillon pour sa participation à mon jury, ses remarques, et pour toute son implication. Je ne saurai également que remercier Hans Geiselmann, non seulement pour avoir présider mon jury, mais aussi pour ses qualités pédagogiques.

J'aimerai aussi remercier Lavinia et Purvi for their appreciated support, mon ancien cobureau Eric pour sa chaleur, Catherine dont les blagues légères n'enlèvent rien à la profondeur, Alexandre qui a su se montrer compréhensif sur mes humeurs enfumées de rédaction, Christophe pour les apéros d'après boulot à EVE, ma compagne de bord de fin de rédaction Aparna, Bernard pour ses repas, Clément, Salima, Titoun, Mario, et toutes les personnes qui ont égayé mes journées au travers d'un sourire, un café, une pause clop, une discussion, un repas. Merci aussi à toutes les personnes qui sont venues me soutenir le jour de ma soutenance !

Je passe ensuite une dédicace spéciale au "maininkitie gang" who always believed in me, and whose memories helped me keeping confidance when I was lacking. Pour son écoute et sa grande humanité je remercie chaudement Jessie Sitbon : les discussions de vie partagées

ont été un vrai plaisir, et souvent même une bouffée d'air frais. J'ai également une pensée pour David et les personnes de mon pays qui m'ont permis de relativiser quand je voyais noir. Enfin et pas des moindres, merci à la team du "sirop de la rue" qui m'a accompagnée tout le long de cette aventure ! En particulier Cécile et Lucas qui ont souvent été là pour me remonter et me changer les idées. Merci !

Table des matières

I. Introduction 1

I.1. Le hasard dans la cellule 2

I.1.1. L'expression génique 3
I.1.2. Expression stochastique des gènes 7
I.1.3. Formulation de la variabilité d'expression génique 13
I.1.4. Mise en évidence expérimentale et caractérisation 19
I.1.5. Rôle du bruit stochastique d'expression génique 36

I.2. Système d'étude 39

II. Dispositif expérimental 47

II.1. Biologie 48

II.1.1. Système rapporteur 48
II.1.2. Préparation des cellules 59
II.1.3. Préparation des échantillons pour la microscopie 62
II.1.4. Injection du glucose 63

II.2. Microscopie 64

II.2.1. Le microscope et son environnement 64
II.2.2. Acquisition des images 65
II.2.3. Caractéristiques de la caméra 69
II.2.4. Intensité d'éclairage de la lampe fluorescente 75
II.2.5. Autofluorescence cellulaire 83

III. Analyse des données 86

III.1. Traitement des images 87
III.1.1. Correction des artefacts d'imagerie 88
III.1.2. Programme d'analyse d'images (Schnitzcells) 97

III.2. Traitement des données 101
III.2.1. Statistiques de population 101
III.2.2. Ajustement numérique des statistiques 104
III.2.3. Equivalence de deux statistiques 106

III.3. Déconvolution de l'autofluorescence 108
III.3.1. Déconvolution matricielle 109
III.3.2. Régularisation du problème 116
III.3.3. Conclusion 120

III.4. Simulations 122
III.4.1. L'algorithme de Gillespie 122
III.4.2. Extension : taux de réaction qui varient dans le temps 126

IV. Résultats 136

IV.1. Question posée et démarche suivie 137

IV.2. Conditions expérimentales standard 142
IV.2.1. Paramètres expérimentaux 142
IV.2.2. Cellules endommagées 148

IV.3. Résultats expérimentaux 155
IV.3.1. Reproductibilité des expériences 155
IV.3.2. Bruit en sortie du module : caractérisation 159

IV.4. Signaux d'entrée de la fonction de transfert : valeurs expérimentales 176
IV.4.1. Taux de dégradation: composante induite par la dilution cellulaire . . . 176
IV.4.2. Taux de dégradation 188
IV.4.3. Taux de production 189

IV.5. Résultats de la Modélisation 194
IV.5.1. Simulations 194
IV.5.2. Ajustement du modèle aux données expérimentales 197

IV.5.3. Résultats 199

IV.6. Discussions **208**

IV.7. Conclusions et perspectives **224**

V. ANNEXES 229

A. Composition du milieu M9 **230**

B. Correspondance entre absorbance et concentration cellulaire **231**

C. Programme de segmentation des cellules sur les images de contraste de phase **233**

Première partie .

Introduction

I.1. Le hasard dans la cellule

Qu'est ce qui différencie deux organismes vivants ? On peut commencer par citer la carte d'identité génétique. Celle-ci est imprimée sur l'ADN : c'est un code barre inné, qui une fois lu, affiche les traits de caractère. Mais ce n'est pas le seul facteur : l'environnement dans lequel un être vivant évolue influencera son comportement. On parle d'expérience acquise. Ainsi, la réponse du vivant est entièrement déterminée par le code génétique et l'histoire personnelle : c'est l'idée qui a longtemps prévalu.

Mais c'est laisser peu de place au hasard ! Après tout, la machine vivante est composée de molécules nanométriques qui interagissent pour dessiner un trait macroscopique. Ces interactions sont souvent de nature biochimique. Or, dès lors que l'on parle de réaction à l'échelle microscopique, on ne peut plus décrire les phénomènes de manière déterministe : les réactants sont soumis à l'agitation thermique, ils se déplacent de manière aléaoire, et donc, se rencontrent au bout d'un temps probabiliste. Et ceci devient critique pour les processus dynamiques où la loi des grands nombres ne s'applique pas : on ne peut pas moyenner l'aléa par un nombre élevé d'évenements "rencontre" si il y a peu de molécules en jeu. Dans ces conditions, l'échelle de temps associée à une réaction reste probabiliste, et la dynamique du processus est sujette au hasard.

C'est le cas pour l'unité de la vie, une cellule individuelle : les processus cellulaires sont implémentés par des réactions biochimiques qui font souvent intervenir des espèces moléculaires en faible quantité. Par conséquent, le produit d'une réaction est sensible au bruit thermique. Le processus dynamique renferme donc de l'aléa. Et ceci est d'autant plus vrai, que les transformations chimiques ont lieu dans un micro environnement dynamique qui conditionne les taux de réaction.

Ainsi, contrairement à l'idée conventionnellement véhiculée, les réactions intra-cellulaires ne sont pas déterministes et il convient d'y introduire de la stochasticité, c'est à dire de l'aléa. Ceci est particulièrement vrai pour les réactions qui impliquent l'ADN : celui-ci se trouve en copie unique dans la cellule, ou du moins en très faible nombre de copies pour les cellules sur le point de se diviser. Ces réactions sur l'ADN composent ce qu'on appelle l'*expression génique*, un mécanisme de lecture de l'ADN qui aboutit à la formation de molécules fonctionnelles. Et ces dernières molécules, avec l'ADN, sont les principaux composants des réseaux d'interactions responsables du fonctionnement cellulaire.

Dans ce contexte, il devient crucial de caractériser la stochasticité dans l'expression génique et de comprendre comment cette stochasticité se propage à travers les différents composants du réseau d'interaction. C'est le but de notre recherche.

Dans cette partie, nous posons le contexte de notre étude : après avoir défini les notions d'expression génique et de réseau de régulation génique, nous caractériserons le bruit d'expression génique à travers les travaux clés de confrères. Puis nous en viendrons à nos travaux de recherche avec une decription détaillée du système d'étude choisi.

I.1.1. L'expression génique

Chez la plupart des organismes vivants, l'information génétique est portée par l'ADN. Nous nous plaçons dans ce contexte général, et nous ne nous intéressons ici qu'aux organismes unicellulaires qui vérifient cette hypothèse.

L'ADN est un polymère dont les monomères (nucléotides) peuvent prendre quatre valeurs, complémentaires deux à deux. La séquence de nucléotides définit le programme génétique : un programme qui implémente la réplication de l'ADN, mais qui code aussi pour toutes les fonctions cellulaires. Ce programme est héréditaire : une fois dupliqué (réplication), la cellule mère le transmet aux deux cellules filles.
Les gènes, eux, sont des unités d'information génétique : ce sont des portions d'ADN qui codent pour les principaux constituants et bâtisseurs cellulaires. Ils sont délimités sur l'ADN par des marqueurs placés en amont (promoteur) et en aval (terminateur) de la séquence codante du gène. Ces marqueurs permettent à la cellule de ne sélectionner que l'information stockée dans le gène ciblé.

Dans ce contexte, l'expression génique désigne l'ensemble des transformations chimiques qui permettent de décoder l'information stockée dans un gène, pour aboutir à la formation de molécules indispensables au fonctionnement cellulaire.
Ces transformations se produisent en série : le gène est d'abord activé, puis l'information qu'il porte est transcrite en acides ribonucléiques (ARN). Les molécules d'ARN ainsi crées peuvent être fonctionnelles [1] ou servir de matrice (ARNm) pour la synthèse de peptides, notamment les protéines qui sont des polypeptides. Dans ce cas, l'expression génique requiert une étape supplémentaire où l'information stockée dans l'ARNm est traduite en protéine.

Ici, nous employons le terme "expression génique" pour décrire l'ensemble des processus associés au transfert d'information de l'ADN, vers les protéines. A savoir : l'activation

1. Par exemple, certains ARN (les ribosymes) ont une fonction enzymatique. D'autres, les ARN de transfert, jouent le rôle de transporteur et de contremaître : ils permettent de transporter l'unité de base des polypeptides vers le lieu d'assemblage, et de lire le plan d'assemblage.

génique, la transcription et la traduction. Ces processus sont représentés sur la figure I.1.1 ci-dessous.

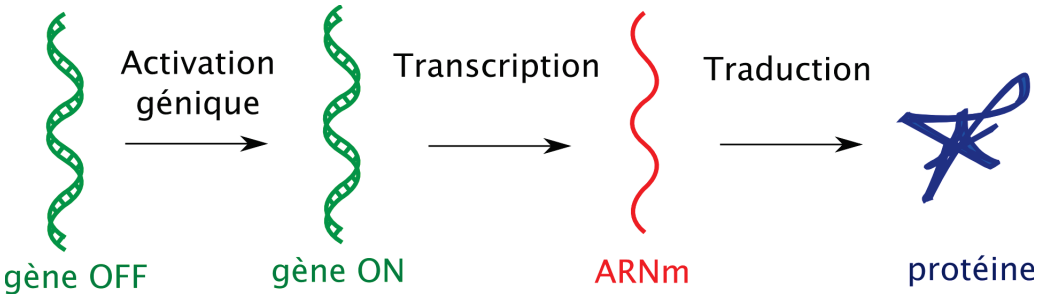


FIGURE I.1.1.: Expréssion génique : Une fois que le gène est prêt à être exprimé (gène ON), l'information génétique est transférée de l'ADN vers l'ARNm (transcription), puis de l'ARNm vers les protéines (traduction).

L'activation génique

L'activation génique désigne le passage du gène, d'un état de veille, à un état actif où l'information qu'il porte (séquence codante) peut être lue.

Ce passage est dicté par la liaison de la machine de transcription (ARN polymérase) au gène. La liaison se fait au niveau du promoteur, le marqueur qui indique le début de la séquence codante : celui-ci contient une séquence de nucléotides qui est reconnue par l'ARN polymérase. Elle vient s'y fixer.

De plus, outre cette séquence commune à tous les gènes, le promoteur contient aussi des séquences spécifiques : les opérateurs. Ces derniers permettent de moduler l'efficacité de la transcription, via des protéines régulatrices. En effet, ces régulateurs transcriptionnels se lient à l'opérateur correspondant, et leur présence rend le site promoteur plus ou moins attractif à l'ARN polymérase.

Nous reviendrons sur ce mécanisme de régulation transcriptionnelle dans la section I.1.1. Pour l'heure, il faut retenir que lorsque l'ARN polymérase (ARNp) se fixe au promoteur P_x du gène x, le gène est prêt à être transcrit. On dira qu'il est activé.

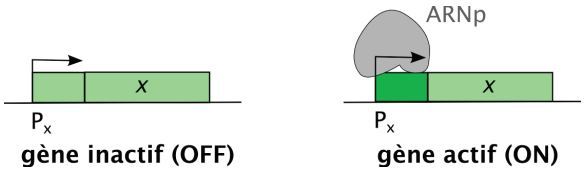


FIGURE I.1.2.: Activation génique.

La transcription

La molécule d'ARN polymérase (ARNp) commence à lire l'information stockée dans le gène et elle en crée un négatif, l'ARNm.
Cette opération dure jusqu'à ce que l'ARNp arrive au site de terminaison, le marqueur qui indique la fin de la séquence codante. A partir de là, la machine de transcription se décroche de l'ADN et elle libère l'ARNm synthétisé. Ces étapes sont résumées sur la figure ci-dessous :

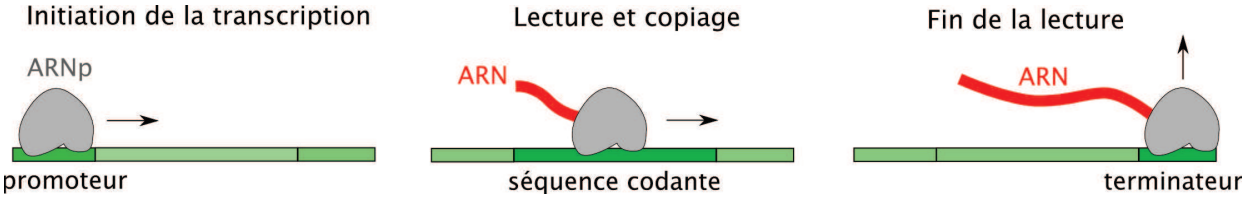


FIGURE I.1.3.: Transcription du gène (représenté en vert) par l'ARN polymérase.

Les ARNm ainsi produits peuvent être dégradés ou remplir leur rôle de messager. Dans ce cas, ils transportent l'information génétique depuis son lieu de stockage (ADN), vers le lieu de synthèse des protéines (ribosomes).

La traduction

Les protéines sont des longues chaînes de peptides : les acides aminés [2] sont assemblés par les ribosomes selon le plan fourni par l'ARNm.
Tout d'abord, le ribosome s'attache à l'ARNm au niveau d'une séquence spécifique appelée le "Site de Liaison Ribosomal" (RBS). A partir de là, il commence à lire le plan fourni par l'ARNm : sur ce messager chaque triplet de nucléotides (codon) code pour un acide aminé. La lecture se fait donc codon par codon. Et pour traduire un codon en terme d'acides aminés, le ribosome utilise les ARN de transfert (ARNt) : ces molécules portent un acide aminé et une séquence de trois nucléotides, l'anticodon. Cette séquence est spécifique de l'acide aminé, et elle peut s'appareiller à un codon de l'ARNm lorsqu'elle est complémentaire.
Le ribosome teste alors, pour chaque ARNt, la complémentarité de l'anticodon avec le codon lu sur l'ARNm. Lorsque l'anticodon est complémentaire, c'est que l'ARNt porte l'acide aminé inscrit sur le codon : le ribosome a trouvé la pièce qui correspond au codon lu. Il se déplace alors pour lire le codon suivant, et il recherche la nouvelle pièce codée. Lorsqu'il l'a trouvée, il l'accole à la précédente. Il recommence ces opérations jusqu'à ce qu'une séquence "stop" soit reconnue. Il s'agit d'un codon aquel aucun ARNt n'est associé.

2. unité de base des peptides

A ce moment, le ribosome se détache de l'ARNm et la chaîne d'acides aminés est libérée dans le cytoplasme. Cette chaîne peptidique peut alors être dégradée, ou se replier sur elle même pour adopter la structure spatiale qui la rend opérationnelle (maturation). Dans ce cas, la protéine devient fonctionnelle : elle sera dégradée ou remplira sa fonction.

Réseau de régulation génique

La fonction d'une protéine peut être de nature diverse. Par exemple, certaines protéines structurent le cytosquelette et les "muscles cellulaires", d'autres assurent la signalisation et les livraisons, d'autres encore ont une fonction régulatrice.
En fait, les protéines sont les briques et le cerveau de la cellule : ce sont elles, qui en grande partie, composent les processus cellulaires. Dans ces conditions, la production des différentes protéines doit être régulée pour n'intervenir que lorsque le processus cellulaire associé est en marche. C'est le rôle des réseaux de régulation génique.

Les réseaux de régulation génique sont des réseaux d'interactions moléculaires qui contrôlent l'expression des gènes. Autrement dit, ils orchestrent la production des différentes protéines associées à un processus.
Dans ces réseaux, l'expression des gènes est souvent contrôlée par des interactions entre les acides nucléiques et les protéines : les protéines régulatrices vues précédemment se lient au promoteur du gène cible, favorisant ainsi ou non sa transcription. Lorsque l'interaction rend le promoteur plus attractif à l'ARN polymérase, la régulation est dite positive. Lorsqu'au contraire elle inhibe la transcription du gène par des liaisons compétitives avec le promoteur, la régulation est dite négative. Ces deux configurations sont représentées sur la figure I.1.4 suivante.

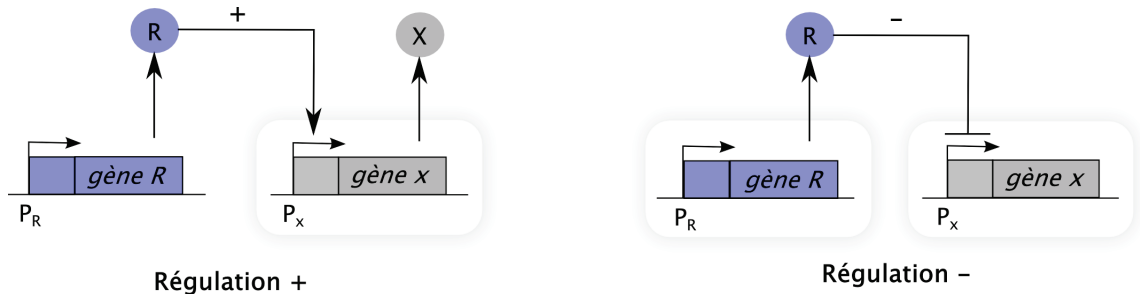


FIGURE I.1.4.: Principe de la régulation positive (+) et négative (-) : le facteur de transcription (R) se lie au promoteur P_x du gène cible (gène x), contrôlant ainsi l'expression de la protéine X.

Ces deux mécanismes forment la base des réseaux de régulation génique. La base et non le tout, car les réseaux ne se limitent pas à deux gènes en interaction comme dans la figure I.1.4 : ils font intervenir plusieurs gènes, chacun d'entre eux étant en interaction

avec le reste du génôme. Ce sont donc des circuits complexes d'interactions qui dictent les processus cellulaires. Et ce, d'autant plus que les réseaux de régulation génique sont souvent imbriqués dans d'autres réseaux, tels que les réseaux métaboliques ou les réseaux de transduction du signal. Les premiers assurent la connection entre les petites molécules de la cellule (métabolites) à travers des réactions enzymatiques, tandis que la transduction des signaux passent, par exemple par la modification post-traductionnelle (maturation) des protéines, ou le transfert de phosphorylation.

Relier l'architecture régulatrice de ces réseaux au comportement dynamique des cellules s'avérent donc d'une grande complexité. Et pour mener à bien ce travail, les scientifiques ont développé des outils "génômiques", mais aussi des outils informatiques. En effet, les biologistes réalisent des expériences dictées par leur intuition : ils sont ainsi capables d'identifier et de connecter différents composants du réseau. Mais parfois, l'intuition est freinée par la complexité du réseau. C'est ici qu'interviennent les bioinformaticiens : ils lancent des simulations qui permettent d'orienter les nouvelles expériences. Les modèles proposés sont alors testés par les biologistes, et les données recueillies permettent d'ajuster le modèle. Ainsi, en faisant appel à deux disciplines complémentaires, les scientifiques sont capables de comprendre et de prédire les processus cellulaires, enfin, du moins souvent. Souvent mais pas toujours, car outre la complexité du problème, les deux approches supposent généralement que les composants du réseau sont déterministes. Or de plus en plus d'études montrent qu'au contraire, l'expression génique est stochastique...

I.1.2. Expression stochastique des gènes

L'expression génique peut être vue comme un module électronique qui renferme l'ensemble des réactions vues au I.1.1. Le signal de sortie correspond alors à la quantité de protéines présentes dans la cellule, et les réactions internes peuvent être modélisées par des taux de réaction (voir figure I.1.5).

Dans ce module, chaque réaction nécessite que les molécules soient suffisamment proches pour interagir. Or ces molécules sont soumises à l'agitation thermique. Les réactions sont donc le résultat de rencontres aléaoires : les taux de réaction sont probabilistes.

Cependant lorsque les événements "rencontre" sont fréquents, on peut moyenner cet aléa par la loi des grands nombres. C'est d'ailleurs sur cette hypothèse d'événements fréquents que l'on s'est longtemps appuyé. Cela permet en effet de simplifier les réactions d'expression génique par des mécanismes déterministes. On peut alors étudier le module d'expression génique sur une population uniforme de cellules, le comportement moyen étant révélateur d'un individu (c'est l'approche suivie par beaucoup de biologistes).

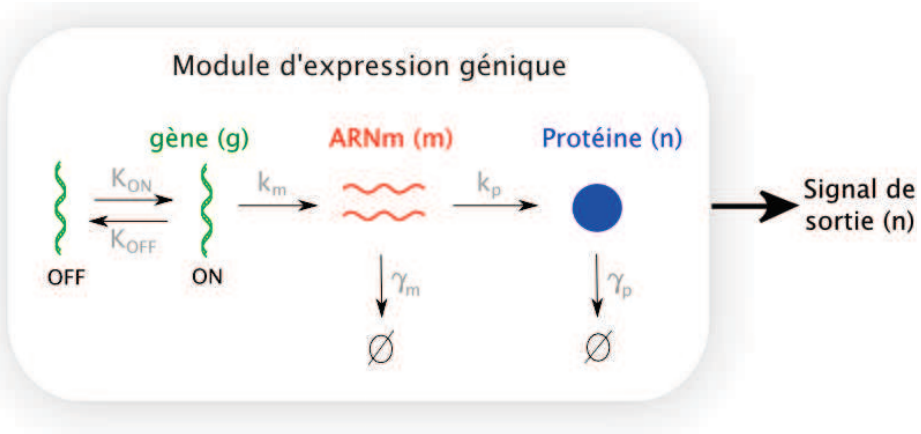


FIGURE I.1.5.: Un module d'expression génique : le gène est activé et désactivé avec des taux respectifs K_{ON} et K_{OFF}. Puis il est transcrit avec un taux k_m. L'ARNm ainsi synthétisé peut être dégradé (taux γ_m) ou traduit en protéine avec un taux k_p. Ces dernières sont alors soit dégradées (taux γ_p), soit utilisées par la cellule (signal de sortie).

Mais le vivant n'est pas si prévisible, avec des événements de rencontre qui sont souvent rares. En effet, les espéces moléculaires en jeux se trouvent bien souvent en faible quantité. Qui plus est, la dynamique de rencontre peut être ralentie par d'autres espéces [3]. Dans ces conditions, les réactions sont probabilistes. Et ce, d'autant plus qu'elles font intervenir des molécules dont la quantité, discrète, est souvent aléaoire.

Ainsi, un même module d'expression génique peut générer des signaux variés, malgré l'uniformité des conditions. Cette variabilité stochastique est souvent qualifiée de bruit et elle se traduit de deux manières :

- Au sein d'une même cellule, la quantité de protéines varie aléaoirement d'un instant à un autre. On parle de *fluctuations dynamiques*.
- Au sein d'une population de cellules identiques, la quantité de protéines, à un instant donné, varie de cellule à cellule. Ainsi, deux cellules génétiquement identiques et exposées aux mêmes conditions peuvent présenter des différences dans la quantité d'une protéine donnée. On parlera de *fluctuations statiques*.

Cette stochasticité depuis longtemps préssentie [2, 1], a pendant longtemps souffert du manque de techniques expérimentales. Mais avec le développement récent de nouveaux outils de génie génétique et de microscopie, elle a été remise au goût du jour en 2000 [3], et elle est maintenant clairement démontrée.

3. Par exemple, la dynamique d'association/dissociation d'une protéine régulatrice au site promoteur peut être ralentie par la présence d'un autre facteur de transcription. Celui-ci exerce alors des forces attractives sur le régulateur lié au promoteur, et des forces répulsives lorsque le régulateur est détaché du promoteur.

De plus en plus de scientifiques se penchent alors sur la question, tentant de comprendre son origine et de caractériser sa signature. Et même si les études expérimentales sont encore peu nombreuses, avec l'aide de travaux théoriques menés en parallèle, elles permettent d'ores et déjà de mieux comprendre cette stochasticité. On assiste ainsi à l'émergence d'une première image qui fait intervenir différentes sources de variabilité.

Les sources de variabilité dans l'expression génique

Ces sources de variabilité sont multiples, et les expériences suggèrent de les classer selon leur appartenance au module d'expression génique [4, 5, 6]. On distingue ainsi :

- Les sources intrinsèques. Ce sont toutes les réactions internes au module qui génèrent de l'aléa dans le signal de sortie.
- Les sources extrinsèques. Ce sont tous les facteurs externes qui introduisent de l'aléa dans le module. Parmi ces facteurs, on distingue ceux qui sont spécifiques du module (protéines régulatrices), et ceux qui agissent simultanément sur tous les modules - *i.e* sur tous les gènes- (facteurs globaux).

Ces différentes sources de variabilité sont hierarchisées sur la figure ci-dessous, et elles sont détaillées dans les paragraphes suivants.

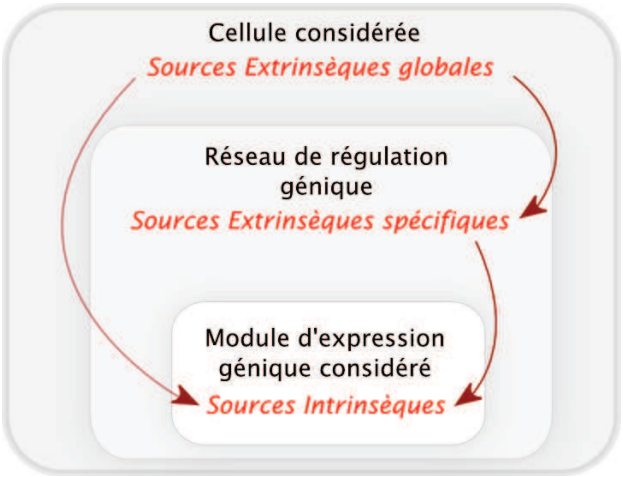


FIGURE I.1.6.: Différentes sources de variabilité dans l'expression génique et échelle associée. Les sources intrinsèques sont internes au module d'expression génique considéré, alors que les sources extrinsèques sont externes au module. Elles peuvent affecter le module de manière spécifique (proteines régulatrices du gène considéré), ou "aspécifique" (facteurs globaux).

On note que cette hierarchisation dépend de la définition du module, et que celle-ci varie parfois d'un scientifique à l'autre. Dans notre étude, les sources intrinsèques se refèrent à la définition de la figure I.1.5.

Sources Intrinsèques

Chaque réaction interne au module -l'activation génique, la transcription et la traduction- est soumise à l'aléa de rencontre. Les variables intrinsèques g, m, et n (cf. figure I.1.5)sont donc aléaoires, et l'aléa sur une variable se propage sur la prochaine. Ainsi, chaque réaction contribue à rendre le signal de sortie stochastique, et on sépare souvent cette stochasticité selon son origine [4]. On parle alors de bruit d'activation génique, de bruit de transcription et de bruit de traduction.

Le bruit d'activation génique concerne surtout l'état du promoteur d'un gène avec les facteurs de transcription, souvent peu nombreux, qui se lient et se délient aléaoirement à l'ADN. Chez les eucaryotes, une autre contribution s'ajoute : les modifications aléaoires de la structure d'un gène. Du réarrangement des portions du gène aux modifications dans la structure de la chromatine, les changements de conformation de l'ADN obéissent aussi à un processus aléaoire dicté par l'agitation thermique. Et comme la structure de l'ADN détermine l'état d'activation du promoteur du gène, celui-ci se trouve aléaoirement activé ou désactivé. Ce bruit d'activation génique est non négligeable puisque les molécules d'ADN se trouvent en faible quantité dans la cellule : le nombre de gènes g actifs est souvent 0 ou 1.

Le bruit de transcription traduit l'aléa dans la production et la dégradation de l'ARNm. On rappelle que lors de la transcription, une molécule d'ARN polymérase vient se lier à l'ADN pour former un complexe ARNpolymérase-ADN, qui une fois ouvert entraîne la production d'ARNm. On trouve donc deux causes à la stochasticité de production de l'ARNm : d'une part, le faible nombre de molécules d'ARNpolymérase qui, soumises au mouvement brownien, se lient aléaoirement à l'ADN ; d'autre part, l'isomérisation aléaoire des complexes ARNpolymérase-ADN en état transcriptionnel compétent. Une fois transcrit, le brin d'ARNm est traduit ou dégradé. Le phénomène de dégradation passe par la liaison du brin avec la RNase [4]. Il introduit donc aussi des fluctuations dans la quantité m d'ARNm. Ces fluctuations vont se transmettre, via la traduction, aux protéines.

Le bruit de traduction concerne la production et la dégradation des protéines. De manière analogue à la dégradation de l'ARNm, les protéines se lient aléaoirement aux protéases avant d'être détruites par protéolyse. Le nombre de protéines étant souvent faible ($\lesssim$1000 molécules par cellule), ce processus reste stochastique. Quant à la production, elle est engendrée par la liaison de l'ARNm au ribosome. Et encore une fois, la rencontre entre les deux molécules se fait de manière aléaoire sous l'effet de l'agitation thermique.

4. Les ribonucléases (ou RNase) sont une classe de nucléase qui catalysent la dégradation de l'ARN en plus petits éléments. Pour ce faire, ces enzymes coupent les liaisons phosphodiester des brins d'acides nucléiques entre deux nucléotides.

Comme l'ARNm est généralement présent en faible quantité dans la cellule (10 à 20 copies), ce processus stochastique a une influence non négligeable sur la concentration n des protéines.

Ces trois contributions sont représentées sur la figure ci-dessous :

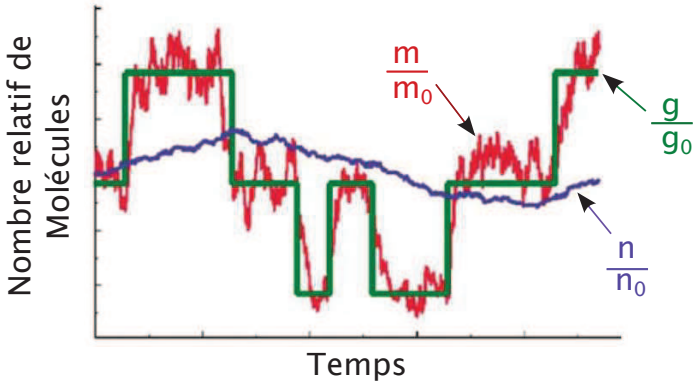


FIGURE I.1.7.: Fluctuations dynamique d'expression génique : représentation des trois sources intrinsèques de variabilité. En vert, l'évolution temporelle de l'activité génique qui bascule aléaoirement de l'état ON vers l'état OFF. Ce bruit d'activation génique se transmet aux protéines via la transcription et la traduction. En rouge, les fluctuations consécutives dans le nombre d'ARNm : ce bruit de transcription se transmet aux protéines via la traduction. Enfin, en bleu, les fluctuations dans le nombre de protéines : ces fluctuations sont le résultat des deux bruits précédemment évoqués et du bruit de traduction qui concerne directement les protéines. D'après J.Paulsson [4].

Sources Extrinsèques

Les réactions précédemment décrites reposent sur des interactions entre les variables intrinsèques (g, m, et n sur la figure I.1.5) et des variables externes au module (nombre de protéines régulatrices, nombre de ribosomes, ...). Ces variables externes, qualifiées d'extrinsèques, déterminent les taux de réaction, et elles peuvent aussi être aléaoires, transmettant dans ce cas leur aléa à la réaction. Elles ajoutent une autre contribution (extrinsèque) à la stochasticité en sortie du module.
Parmi ces variables extrinsèques, on distingue celles qui sont spécifiques d'un module (*i.e* d'un gène), et celles qui affectent tous les modules d'expression génique. On parle alors de bruit spécifique et de bruit global.

Le bruit spécifique concerne surtout les protéines régulatrices : celles-ci sont spécifiques d'un gène ou d'un ensemble de gènes. En tant que protéines, elles sont produites de manière stochastique, et en tant que régulateurs, elles régulent l'activation du gène cible. Le mécanisme d'activation génique ciblé est donc soumis à la stochasticité des régulateurs. Dans ces conditions, les protéines régulatrices représentent une source extrinsèque

de variabilité pour le module associé au gène cible.

Le bruit global concerne les facteurs globaux d'expression génique. Par exemple, le nombre d'ARNp, de RNase, de ribosomes, d'ARNt, ou de protéases affectent simultanément tous les modules d'expression génique. Or ces facteurs peuvent fluctuer dans le temps : ils varient au cours du cycle cellulaire et leur production est sujette à l'aléa de la transcription et/ou de la traduction d'un autre module. Ils représentent donc des sources extrinsèques de variabilité pour le module considéré. Egalement, la division cellulaire introduit de l'aléa dans le module : elle répartit aléaoirement les variables intrinsèques et extrinsèques dans les deux cellules filles.

Toutes ces sources extrinsèques sont représentées sur la figure ci-après, sans distinction sur la spécificité vis à vis du module.

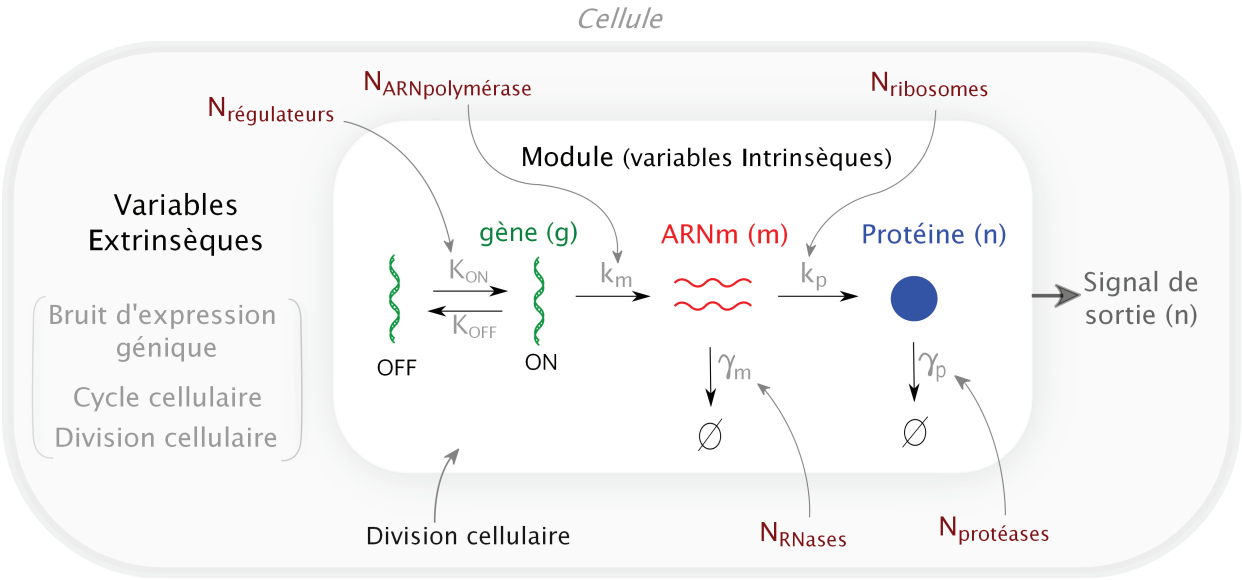


FIGURE I.1.8.: Sources de bruit externes au module considéré (variables extrinsèques). Les variables bordeaux interagissent avec les variables intrinsèques g, m et n pour définir les taux de réaction. Ces taux sont intrinsèquement probabilistes (événements rares de rencontre), et comme les variables bordeaux se trouvent en quantité aléaoire dans la cellule (cycle cellulaire, bruit d'expression génique, partitionnement lors de la division cellulaire), cette stochasticité intrinsèque est soumise à l'aléa des variables externes. Ces variables représentent donc des sources extrinsèques de stochasticité. La division cellulaire, quant à elle, répartit le matériel cellulaire aléaoirement dans les deux cellules filles. Pour les variables intrinsèques, elle représente donc, aussi, une source extrinsèque de stochasticité.

I.1.3. Formulation de la variabilité d'expression génique

En se basant sur le modèle de la figure précédente, des confrères ont formulé le caractère aléaoire de l'expression génique. Ces travaux théoriques ont fortement été influencés par des résultats expérimentaux, et ils ont permis de comprendre comment les différentes sources de bruit opèrent sur la variable n (nombre de protéines/cellule) en sortie du module d'expression génique.

Sources Intrinsèques

Les sources intrinsèques sont modélisées par le module de la figure I.1.5 : ce module caractérise l'état (n, m, g) de la cellule.

Equation maîtresse pour le modèle de transcription-traduction

Dans ce module, les réactions d'expression génique peuvent être formulées en terme de probabilités de transition : les taux de réaction représentent la probabilité de transiter d'un état vers un autre pendant la durée infinitésimale dt. Par exemple, k_m représente la probabilité, par unité de temps, de transiter d'un état (n, m, g) vers un état $(n, m+1, g)$ à partir d'un gène actif.
Cette formulation s'appuie sur quatres hypothèses : 1-l'état de la cellule est entiérement décrit par ses variables intrinsèques (n, m, g) , 2-les réactions impliquées sont instantanées, 3- ces réactions suivent des processus de Poisson (*i.e.* probabilité d'occurence/unité de temps constante), 4- les variables intrinsèques sont réparties uniformément dans l'espace intra-cellulaire.
Sous ces hypothèses, le module d'expression génique se réduit aux transitions ci-dessous :

$$\begin{aligned}
&\text{activation génique :} && \text{g} \xrightarrow{K_{ON}(g^{max}-g)} \text{g-1}\\
&\text{inactivation génique :} && \text{g} \xrightarrow{K_{OFF}\ g} \text{g+1}\\
&\text{transcription :} && \text{m} \xrightarrow{gk_m} \text{m+1}\\
&\text{dégradation ARNm :} && \text{m} \xrightarrow{m\gamma_m} \text{m-1}\\
&\text{traduction :} && \text{n} \xrightarrow{m.k_p} \text{n+1}\\
&\text{dégradation protéine :} && \text{n} \xrightarrow{n\gamma_p} \text{n-1}
\end{aligned}$$

A partir de ce système de réactions, on peut expliciter la probabilité que la cellule se trouve à l'instant t dans un état donné : $P(g, m, n, t)$ s'exprime en fonction des probabilités de transition associées aux réactions.

Sans aller plus loin dans les détails, nous allons nous focaliser sur le cas simplifié où l'on n'a que deux variables m et n, le gène étant toujours actif. Dans ce cas, les mécanismes considérés se réduisent à la transcription et à la traduction, comme schématisé sur la figure ci-dessous :

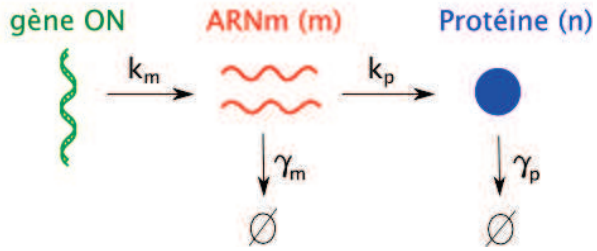


FIGURE I.1.9.: Mécanismes d'expression génique considérés dans un modèle simplifié : le gène est toujours activé, et les taux de réactions (k_m, k_p, γ_m, γ_p) ne sont pas soumis à l'aléa des sources extrinsèques. Ces taux désignent la probabilité d'occurence d'une réaction par unité de temps.

On exprimera dans ce cas la variation de la probabilité d'avoir (dans une cellule) m molécules d'ARNm et n molécules de protéines au temps t selon l'équation maîtresse :

$$\begin{aligned}\frac{dP_{m,n}(t)}{dt} &= \gamma_m\Big((m+1)P_{m,n}(t) - mP_{m,n}(t)\Big) + k_m\Big(P_{m-1,n}(t) - P_{m,n}(t)\Big) \\ &+ \gamma_p\Big((n+1)P_{m,n+1}(t) - nP_{m,n}(t)\Big) + k_p\Big(mP_{m,n-1}(t) - mP_{m,n}(t)\Big)\end{aligned}$$

Cette équation a été résolue par V.Shahrezaei et P.S.Swain[7] dans le cas particulier où $\gamma_m \gg \gamma_p$ (comme c'est souvent le cas dans la nature). Dans ce cas, sur une échelle de temps τ associée aux protéines, les fluctuations dynamiques sur l'ARNm sont moyennées. Dans ces conditions, la distribution de probabilité jointe $P_{m,n}$ est piquée autour de $m = \langle m \rangle$. La probabilité sur l'échelle de temps des protéines $P_{m,n}(\tau)$ s'exprime alors comme une probabilité fixe de m :$P_{m,n}(\tau) \approx P_{n,\langle m \rangle}(\tau)$. En utilisant un changement d'espace par les fonctions génératrices $F(z,z') = \sum_{n,m} z'^m z^n P_{m,n}$, on peut ainsi résoudre l'équation maîtresse pour $P_n(\tau)$. Ces fonctions permettent de passer de l'espace (n,m) vers l'espace (z,z') où l'équation se simplifie en une équation aux dérivées partielles du premier ordre.

Ainsi, à l'équilibre statistique, la quantité de protéines par cellule est distribuée selon une loi binomiale négative :

$$P_n = \frac{\Gamma(a+n)}{\Gamma(n+1)\Gamma(a)} \times \left(\frac{b}{1+b}\right)^n \times \left(1 - \frac{b}{1+b}\right)^a \tag{I.1.1}$$

où $b = k_p/\gamma_m$ et $a = k_m/\gamma_p$.

Pour n grand, cette loi tend vers la distribution gamma :

$$P_n \underset{n\gg 1}{\longrightarrow} \frac{1}{b^a \Gamma(a)} n^{a-1} e^{-\frac{n}{b}}$$

Et si on s'intéresse maintenant à la dynamique de la loi de probabilité, on ne peut plus considérer qu'on est à l'équilibre statistique. Dans ce cas, la distribution change, et pour formuler ces changements, V.Shahrezaei et P.S.Swain ont résolu l'équation en dynamique pour des temps grands devant la durée de vie de l'ARNm ($t \times \gamma_m \gg 1$). Les distributions générées sont présentées sur la figure ci-dessous :

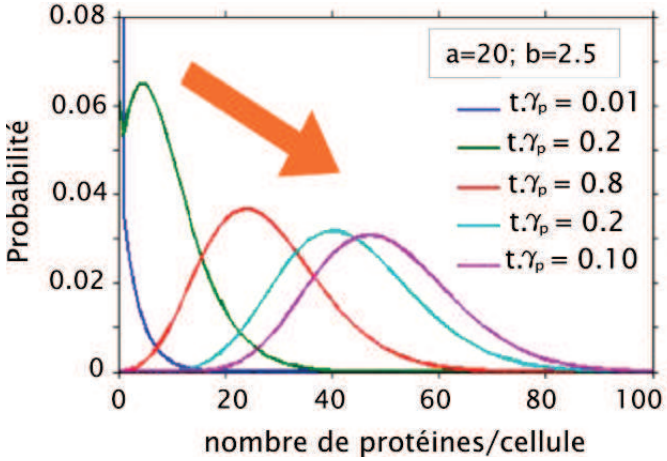


FIGURE I.1.10.: Résolution analytique de l'équation maîtresse qui formule un module à deux étages (transcription/traduction). Les courbes sont tracées pour différents temps, avec des valeurs fixes $a = 20$ et $b = 2,5$. Le nombre intial de protéines dans les cellules est fixé à 0. D'après V.Shahrezaei et P.S.Swain [7].

Etude dynamique : Bouffées de production protéique

Pour comprendre la loi binomiale et sa limite de distribution gamma, nous nous intéressons au paramètre $\sigma_n^2/\langle n\rangle$ tiré de P_n. Il quantifie la déviation du processus observé par rapport à un processus poissonien. En effet, pour un processus poissonnien, la variance et la moyenne du nombre n de protéines/cellule sont reliées par $\sigma_n^2/\langle n\rangle = 1$. Dans le cas du modèle précédent à deux étages (transcription/traduction), on a à l'équilibre statistique [5] :

$$\frac{\sigma_n^2}{\langle n\rangle} = 1 + b = 1 + \frac{k_p}{\gamma_m}$$

Si ce b est $\gg 1$, le bruit généré par les sources intrinsèques (transcription et traduction) sur le nombre de protéines/cellule n'est pas poissonien. Dans ce cas, les protéines ne sont pas issues d'événements de production indépendants. Elles sont produites par paquets, comme schématisé sur la figure I.1.11 : plusieurs protéines sont synthétisées en rafale

5. Pour une distribution binomiale négative de paramètres a et b, les moments du 1^{er} et du $2^{ième}$ ordre valent : $\langle n\rangle = ab$ et $\sigma_n^2 = ab(1+b)$.

à partir d'un brin d'ARNm, et ces courtes périodes sont encadrées par de longs temps d'attente où la traduction est inhibée par l'absence d'ARNm.

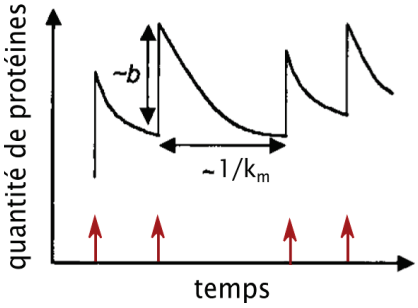


FIGURE I.1.11.: Représentation schématique de la production de protéines par bouffée : chaque pic pointé par une fléche correspond à une bouffée de traduction d'un ARNm. L'amplitude des bouffées est donnée par b, et la fréquence est inversement proportionelle à k_m. D'après M.Thattai et al [8].

Et la propension d'un module à générer des bouffées de production protéique sera caractérisée par l'écart b par rapport à un processus poissonnien. Pour chaque module, les bouffées de production sont caractérisées par leur amplitude et leur fréquence d'apparition.

Fréquence des bouffées : La fréquence est donnée en unité de temps de demi-vie de la protéine : c'est le nombre de bouffées produites pendant la durée de demi-vie de la protéine. Et dans le modèle à deux étages où les probabilités de réaction sont constantes par unité de temps, la transcription de l'ARNm suit un processus poissonnien. Dans ces conditions, l'intervalle de temps qui sépare la création de deux brins ARNm suit une distribution exponentielle, de paramèter k_m. Et le nombre de bouffées présentes pendant la durée de demi-vie de la protéine suit une loi poissonienne de moyenne $a = k_m/\gamma_p$.

Amplitude des bouffées : L'amplitude des bouffées correspond au nombre de protéines crées par impulsion. C'est une variable aléaoire, et on peut montrer qu'elle suit une distribution géométrique [2, 7]. En effet, dans le cadre du modèle d'expression génique à deux étages, les ARNm peuvent se lier, soit aux ribosomes pour être traduits, soit aux RNases pour être dégradés. Ces deux réactions sont donc en compétition. Et si on les considère indépendantes, la probabilité que l'ARNm se lie au ribosome est de :

$$p = \frac{k_p}{\gamma_m + k_p} = \frac{b}{1+b}$$

où $b = k_p/\gamma_m$. Si l'hypothèse $\gamma_m \gg \gamma_p$ est maintenue, alors sur l'échelle de temps associée aux protéines, toutes les protéines sont produites simultanément à partir d'un ARNm (bouffée). Dans ces conditions, la probabilité que n_1 protéines soient produites à partir

d'un ARNm avant qu'il ne soit dégradé est donnée par une distribution géométrique, à savoir :

$$P(n_1) = p^{n_1} \times (1-p)$$

Et pour a bouffées/demi vie de la protéine, on retrouve le nombre total des protéines dans la cellule comme une variable aléaoire sommant la contribution de chaque bouffée :

$$n = \sum_{i=1}^{a} n_i$$

On retrouve donc la distribution binomiale négative de l'équation I.1.1 précédente.

Relation avec les fluctuations statiques : N.Friedman et al[10] ont développé un modèle qui traduit ces fluctuations dynamiques en terme de fluctuations statiques. Les hypothèses sont les mêmes que celles utilisés par V.Shahrezaei et P.S. Swain pour décrire les fluctuations statiques, mais ils considèrent en plus, que les protéines sont produites par bouffées. Cela permet de résumer les réactions de transcription, dégradation de l'ARNm, et traduction en une seule étape. Dans ces conditions, la probabilité $P_{n,t}$ de trouver n protéines dans la cellule à l'instant t s'écrit :

$$P_{n,t+\triangle t} - P_{n,t} = w_{n\to n+1} P_{n+1,t} - w_{n\to n-1} P_{n,t} + \sum_{bouffées} \Big(P_{n,B,t} - P_{n,t} \Big)$$

où $P_{n,B}$ représente la probabilité d'avoir n protéines après une bouffée de traduction B. Et $w_{n\to n+1}$ (resp. $w_{n\to n-1}$) représente la probabilité de passer, pendant $\triangle t$, d'un état $n+1$ (resp. n) vers un état n (resp. $n-1$). Ces changements d'état traduisent l'effet de la dégradation protéique et ils peuvent se réécrire comme suit :

$$\frac{\triangle P_{n,t}}{\triangle t} = \gamma_p \Big((n+1) P_{n+1,t} - n P_{n,t} \Big) + \sum_{bouffées} \Big(P_{n,B} - P_{n,t} \Big)$$

dégradation production

Cette équation a été résolue en passant à la limite continu : la variable n'est plus la quantité de protéines n, mais la concentration intracellulaire de protéines x. On peut ainsi montrer qu'à l'équilibre statistique (*i.e.* $dP(x)/dt = 0$), cette concentration est décrite par une distribution gamma :

$$P(x) = \frac{1}{b^a \Gamma(a)} x^{a-1} e^{-(x/b)}$$

avec $b = k_p/\gamma_m$ qui définit l'amplitude des bouffées, et $a = k_m/\gamma_p$ qui en définit la fréquence.

Empreinte de la fréquence et de l'amplitude des bouffées sur les fluctuations dynamiques : M.Thattai et al[8] ont montré à travers des simulations, que lorsque les bouffées sont rares et de grande amplitude (faible transcription et forte traduction), les fluctuations dynamiques intrinsèques sont importantes. Au contraire, lorsque ces bouffées sont fréquentes et de faible amplitude, les fluctuations sont faibles. Ainsi, en jouant conjointement sur l'efficacité de la traduction et de la transcription, les cellules sont capables de moduler le bruit intrinsèque d'expression génique.

Modèle plus complexe : activation génique-transcription-traduction

A. Coulon et al[9] ont montré que l'étage d'activation génique permet de moduler plus efficacement le bruit d'expression génique et qu'il offre à la cellule plus de flexibilité pour contrôler ce bruit. Ces résultats s'appuient sur un modèle qui intégre l'activation génique et l'architecture régulatrice du gène. Le promoteur est décrit par une multitude d'états, chacun correspondant à la liaison d'un ou plusieurs facteurs de transcription. Ces facteurs peuvent se lier au promoteur de manière coopérative ou compétitive : la constante d'association dépend non seulement du facteur de transcription, mais aussi de l'état initial du promoteur. Ce comportement coopératif/compétitif est étendu aux molécules d'ARNpolymérase, ce qui se traduit par une efficacité de la transcription qui dépend elle aussi de l'état initial du promoteur. On a donc un taux de transcription qui peut prendre plusieurs valeurs, chaque valeur étant dictée par l'état du promoteur. Ce modèle plus large permet notamment de reproduire la complexité observée sur le promoteur dans les organismes eucaryotes.

Formulation de la variabilité extrinsèque d'expression génique

On souligne que les travaux théoriques précédents ne considèrent que les sources intrinsèques de bruit. Dans les organismes vivants, d'autres sources (extrinsèques) peuvent contribuer à la variabilité observée sur n. Dans ces conditions, il est difficile d'inférer le sens physique des paramètres a et b de la distribution gamma : ceux-ci ne reflètent pas forcément (respectivement) k_m/γ_p qui traduit le nombre de protéines créées par ARNm, et k_p/γ_m qui traduit le nombre d'ARNm créés par temps de vie de la protéine.
Et pour intégrer ces sources dans les modèles précédents, on rajoute des fluctuations aléaoires sur les taux de réaction. Pour les études statiques, cela se traduit par une distribution des taux de réaction au sein de la population de cellules. Et pour les études dynamiques, cela se traduit par des fluctuations temporelles stochastiques sur les taux de réaction, jusqu'alors déterministes. En suivant cette deuxième approche, V.Shahrezaei et P.S.Swain ont montré, par simulation, que les sources extrinsèques peuvent changer l'al-

lure de la distribution [11, 12]. Les simulations sont faites pour un modèle à trois étages qui prend en compte l'activation génique. Et elles sont réalisées avec l'algorithme de Gillespie qui intègre les fluctuations dynamiques des taux de réaction [6] [13]. Deux simulations sont lancées : l'une ne considère que les sources intrinsèques avec des taux de réaction constants dans le temps l'autre considère en plus les fluctuations extrinsèques sur k_m. Les résultats donnés sur la figure I.1.12 ci-après montrent clairement que les fluctuations extrinsèques dans le nombre d'ARNpolymérase (k_m) affectent l'allure de la distribution.

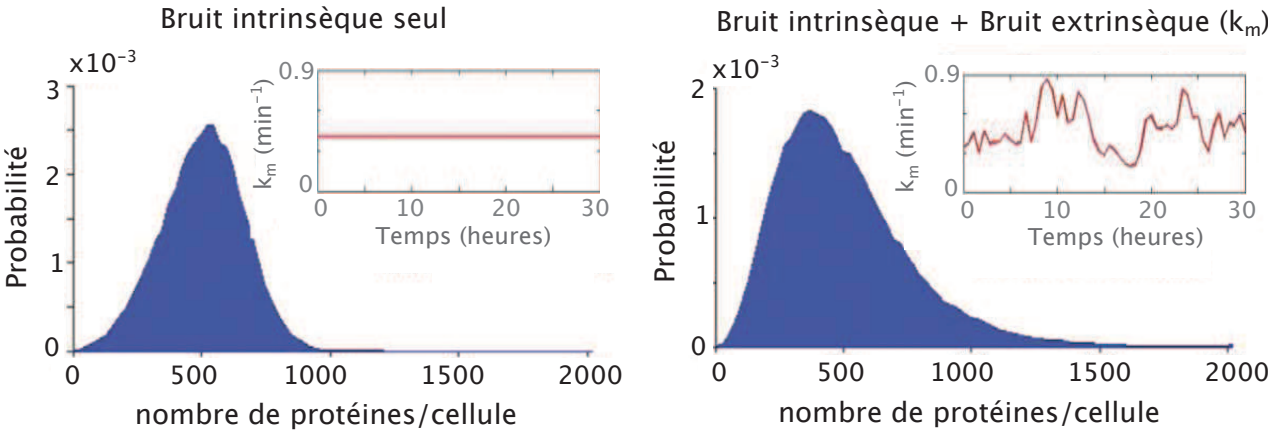


FIGURE I.1.12.: Distribution de la quantité de protéines par cellule à l'équilibre statistique. Les statistiques sont obtenues par simulation avec l'algorithme de Gillespie. Le modèle simulé est celui de la figure I.1.9 avec l'étage d'activation génique qui est rajouté : le promoteur peut fluctuer entre état actif et inactif avec des probabilités de transition par unité de temps : $K_{ON} = 0,012$ min^{-1} et $K_{OFF} = 0,036$ min^{-1}. Quant aux taux de transcription et traduction, ils valent $k_m = 0,402$ min^{-1}, $k_p = 0,036$ protéine.min^{-1}, $\gamma_m = 0,06$ min^{-1}, et $\gamma_p = 0,012$ min^{-1}. A gauche, résultat de la simulation lorsque les taux de réactions sont constants dans le temps. A droite, résultat de la simulation lorsque k_m fluctue aléaoirement dans le temps. Ces fluctuations sont considérées distribuées selon une loi log-normale d'écart type $\sigma = 0.5\langle k_m \rangle$, et corrélées pendant 167 min. Elles sont représentées en rouge dans l'encadré du coin droit. D'après V.Shahrezaei et al [12].

I.1.4. Mise en évidence expérimentale et caractérisation

Pour mettre en évidence et caractériser ces différentes sources de bruit, des expériences ont été menées sur différents modules, dans différents organismes unicellulaires, et dans différentes conditions.

Dans ces expériences, la stochasticité est souvent étudiée à travers les variations intercellulaires de la quantité de protéines (fluctuations statiques). Ces variations sont alors

6. Voir la section III.4.1 pour plus d'informations sur cet algorithme.

décrites par des distributions, qui donnent la loi de probabilité associée au nombre de protéines dans une cellule. Et plus communément, mais moins informatif, elles sont décrites par le paramètre :

$$\eta^2_{total} = \frac{\sigma_n^2}{\langle n \rangle^2}$$

où $\langle n \rangle$ désigne le nombre moyen de protéines et σ_n la déviation standard associée au nombre de protéines.
On trouve néammoins quelques études dynamiques qui, comme nous le verrons, apportent un autre regard sur le bruit d'expression génique. Dans les deux cas, les expériences requièrent des outils technologiques de pointe.

Techniques expérimentales

Pour sonder un module d'expression génique en particulier, les scientifiques utilisent généralement des rapporteurs fluorescents [7]. Développés par des biologistes, ce sont des gènes qui codent pour une protéine fluorescente : ils sont intégrés dans la cellule. Celle-ci émet alors un signal de fluorescence qui témoigne du niveau d'expression de la variable considérée. Cette variable d'étude est souvent le nombre de protéines en sortie du module, mais elle peut aussi être la quantité d'ARNm dans le module [14].
Pour pouvoir sonder la stochasticité dans la variable choisie, les expériences sont réalisées sur des cellules individuelles : on peut ainsi observer la variabilité inter-cellulaire (fluctuations statiques) ou intra-cellulaire (fluctuations dynamiques). Deux techniques peuvent être utilisées pour cette opération : la cytométrie de flux ou la microscopie.

- La cytométrie de flux permet de mesurer à un instant donné la fluorescence de chaque cellule. Cette technique présente l'avantage de pouvoir sonder facilement un grand nombre de cellules, mais elle distingue moins finement que la microscopie les cellules en bonne santé, des cellules mortes, agrégées, ou encore des débris de taille cellulaire [15]. De plus, elle perd l'information sur l'état d'une cellule aux différents instants.
- La microscopie, quant à elle, permet de sonder moins de cellules (généralement de 1 à 500 [8]), mais elle offre une visualisation directe des cellules, et elle permet de mesurer l'évolution temporelle de leur fluorescence. Elle fournit donc des informations plus précises [15], et aussi plus complètes, avec des données à la fois statiques (population de cellules imagées au même instant) et dynamiques (suivie d'une cellule individuelle dans le temps).

  Cependant, elle est plus longue et difficile à mettre en oeuvre. En effet, elle nécessite une étape supplémentaire de traitement d'image qui devient vite fastidieuse

7. Pour une description plus détaillée de ce système rapporteur, voir la section II.1.1 .
8. Cette quantité dépend de la concentration de l'échantillon imagé, et elle est limitée par les risques de photoblanchiment encourus lorsqu'on "image" trop de zones de l'échantillon.

lorsqu'elle est faite avec rigueur [9]. De plus, elle requiert une étape préalable où les cellules sont, d'une part séparées dans l'espace pour être distinguables les unes des autres, d'autre part immobilisées pour être reconnaissables à différents instants. Cette opération est souvent réalisée avec le gel d'agarose, un gel de culture cellulaire qui permet d'immobiliser les cellules. On peut aussi utiliser une colle plus efficace, la poly-L-Lysine [10]. Dans les deux cas, le suivi temporel est limité par la croissance cellulaire. En effet, les cellules se multiplient, et elles finissent par occuper tout l'espace, devenant alors indistinguables les unes des autres. Pour remédier à ce problème et pouvoir observer les cellules sur des temps plus longs [11], les scientifiques commencent à développer des circuits microfluidiques : ils confinent les cellules dans un espace restreint (séparation des cellules), et ils régulent la densité de cellules grâce à des canaux d'évacuation (pas de surpopulation de l'espace) [16, 17].

Ces deux outils essentiels permettent ainsi d'observer les fluctuations cellulaires. Et pour pouvoir imputer la variabilité observée à l'aléa d'expression génique, les mesures sont réalisées dans des conditions uniformes. En particulier :

- Les expériences sont réalisées sur des cellules monoclonales : issues d'un même ancêtre, elles sont génétiquement identiques, exception faite des mutations qui peuvent apparaître. Ces mutations sont extrêmement rares sur les échelles de temps associées aux expériences (durée de l'ordre d'une dizaine de divisions cellulaires), et la probabilité qu'une des cellules sondées soit un mutant est infime. Dans ces conditions, le hasard des mutations peut être négligé devant la stochasticité issue de l'expression génique.
- Les cellules sont exposées aux mêmes conditions environnementales : le milieu de culture est uniforme dans l'espace. Et pour les études de fluctuations dynamiques, le milieu de culture reste invariant pendant les mesures (milieu de culture infini ou rafraîchissement constant du milieu).

Grâce à ces techniques, des expériences récentes ont permis de quantifier les niveaux de protéines et d'ARNm à l'échelle d'une cellule individuelle. Les différentes sources de bruit dans l'expression génique ont alors pu être investiguées, comme nous allons voir à travers quelques expériences clés.

9. Elimination manuelle des fausses détections cellulaires sur les images acquises (débris, cellules mortes, ...). Pour plus d'informations, voir la section III.1.2

10. Dans ce cas, les cellules sont cultivées, non pas dans un milieu solide comme sur l'agarose, mais dans un milieu liquide. Pour plus d'informations, voir la section II.1.3.

11. Long signifie ici supérieur à 5 cycles de division cellulaire.

Fluctuations générées sur la variable n

Production de protéines par bouffées

Les bouffées de production protéique évoquées dans la section précédente ont été mises en évidence en 2006 par L.Cai, N.Friedman et X.Sunney Xie [18] : ces chercheurs ont développé un système expérimental qui permet de compter, en temps réel, le nombre de protéines présentes dans les cellules *E. Coli*.
Les expériences sont réalisées sur la $\beta-$galactosidase, une protéine naturelle qui est codée par le gène *lacZ* placé sous le contrôle du promoteur P_{lac}. Cette protéine présente des propriétés enzymatiques très intéressantes : elle est capable d'hydrolyser un substrat fluorogène, catalysant ainsi la production d'une molécule fluorescente (la fluoresceine). En présence de l'enzyme, le signal de fluorescence est donc considérablement amplifié. Et comme la fluoresceine n'est pas retenue à l'intérieur de la cellule, L.Cai et al ont développé un système microfluidique qui piège les cellules dans une chambre close : le signal de fluorescence émis par cette chambre traduit alors le nombre de molécules de fluoresceine.
Ils ont ainsi pu mesurer l'activité enzymatique de la $\beta-$galactocidase pour plusieurs cellules. Et lorsqu'ils tracent l'histogramme des taux d'hydrolyse, ils observent des pics discrets : ces pics sont le résultat du nombre discret de molécules $\beta-$galactocidase. Une calibration "taux d'hydrolyse↔nombre d'enzyme" a ainsi pu être effectuée. A partir de cette calibration, la quantité d'enzyme a été mesurée en temps réél lorsque l'expression de la protéine est réprimée. Les résultats obtenus montrent une production de protéines par paquet, avec le nombre de molécules synthétisées par bouffée qui suit une loi exponentielle (voir figure I.1.13 ci-après).

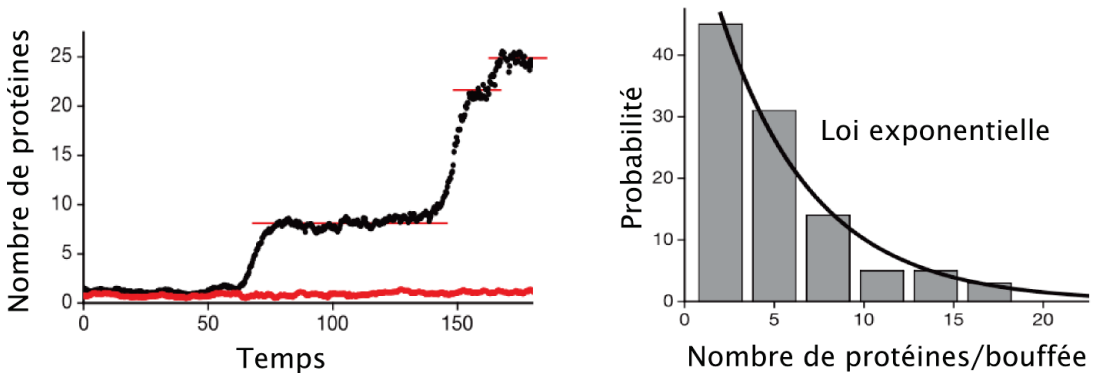


FIGURE I.1.13.: Mesures en temps réél du nombre de $\beta-$galactocidase dans des cellules *E. Coli* où l'expression de la protéine est réprimée. a) Evolution temporelle de la quantité de protéines, avec des bouffées se produisant à une fréquence moyenne $a = 0,11 \pm 0,03$ bouffée/cycle cellulaire. b) Histogramme du nombre de protéines créées par bouffées : la statistique suit une loi exponentielle de moyenne $b = 5 \pm 2$ enzymes/ARNm. D'après L.Cai, N.Friedman et al [18] .

En accord avec le modèle présenté dans la section I.1.3 précédente, lorsqu'ils s'intéressent à la variabilité inter-cellulaire d'équilibre statistique, ils mesurent des distributions gamma dont les paramètres a et b sont proches des valeurs trouvées par l'approche dynamique : $a_{statique} = 0,16 \pm 0,05$ et $b_{statique} = 7,8 \pm 2,6$.

Ce comportement en bouffées de traduction a été retrouvé par J.Yu, J.Xiao et al[19] sur la *Yellow fluorescent Protein* (YFP) [12], encore une fois activée par le promoteur P_{lac}. Pour pouvoir visualiser les protéines une par une, la protéine rapporteuse a été fusionnée [13] avec une protéine membranaire : les protéines fluorescentes se localisent ainsi sur la membrane de la cellule. Et comme l'expression de la YFP est réprimée, il y a peu de molécules rapporteuses : le signal de fluorescence est distinguable entre deux fluorophores appartenant à la même cellule, comme montré sur la figure I.1.14. Les mesures en temps réél de la quantité de YFP par cellule font alors apparaître des bouffées de production protéique. L'amplitude de ces bouffées y suit une distribution géométrique de moyenne b $= 4,2$ YFP/ARNm, et la fréquence décrit une loi de probabilité poissonienne, de moyenne $a = 1,2$ bouffées/cycle cellulaire [14].

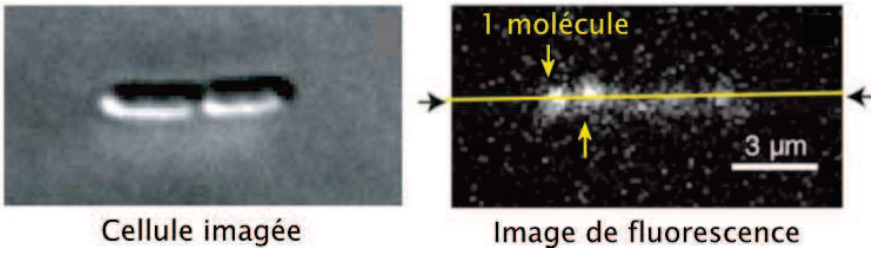


FIGURE I.1.14.: Identification d'une molécule YFP positionnée sur la membrane cellulaire. A gauche, la cellule est imagée en DIC. A droite, la même cellule est imagée en fluorescence : les deux spots blancs pointés par une fléche représentent une molécule de YFP. D'après J.Yu, J.Xiao et al [19].

La propension du module à générer des bouffées est fortement corrélée à k_p

Les expériences précédentes se placent dans des conditions particulières : elles sont menées sur une protéine dont l'ARNm se trouve en faible quantité dans la cellule (activation génique inhibée). Et on souligne que la prodution par bouffée nécessite des conditions spéciales sur les taux de réactions intrinsèques.

Pour déterminer ces conditions, Ozbudak et al[20] ont mené des expéricences sur *B.subtilis* : ils observent comment la propension du module à générer des bouffées de protéines dépend

12. Pour plus d'informations sur cette protéine fluorescente, voir section II.1.1. Dans les expériences de J.Yu, J.Xiao et al, la protéine fluorescente utilisé est un dérivé de la YFP plus brillant, "Venus".

13. Voir la section II.1.1 pour plus d'informations sur la fusion protéique.

14. On note ici que lorsque la dégradation protéique est dominée par la dilution cellulaire, le temps de demi-vie de la protéine est donné par la durée du cycle cellulaire. Dans ce cas, a s'exprime en bouffée/cycle cellulaire.

du niveau moyen d'activation génique et du taux de traduction moyen. Pour ce faire, ils ont introduit un gène rapporteur codant pour la *Green Fluorescent Protein* [15] (GFP) dans le chromosome des cellules. Ce gène est activé par un promoteur dont l'état (actif/inactif) peut être régulé par une molécule ajoutée au milieu de culture (IPTG). La concentration de cette molécule régulatrice détermine alors l'efficacité d'activation génique. Quant à la traduction, elle est ajustée en introduisant des mutations ponctuelles dans la séquence du gène qui code pour le site de liaison ribosomal [16].

Les fluctuations statiques ont alors été mesurées par cytométrie de flux pour un gène quasi-inactivé, puis fortement activé, et quatre taux de traduction différents. Les résultats obtenus mettent en évidence une forte corrélation entre la propension du module à générer des bouffées de protéines, et l'efficacité de traduction. En effet, sur la figure I.1.15-a, on constate que plus l'efficacité de traduction est importante, plus $\sigma_n^2/\langle n\rangle$ s'éloigne d'un processus de Poisson. Et sur la figure I.1.15-b, le paramètre $\sigma_n^2/\langle n\rangle$ semble indépendant de l'état du promoteur. La propension du module à génerer des bouffées de production protéique apparaît donc proportionnelle au taux moyen de traduction : $b =\sim k_p$.

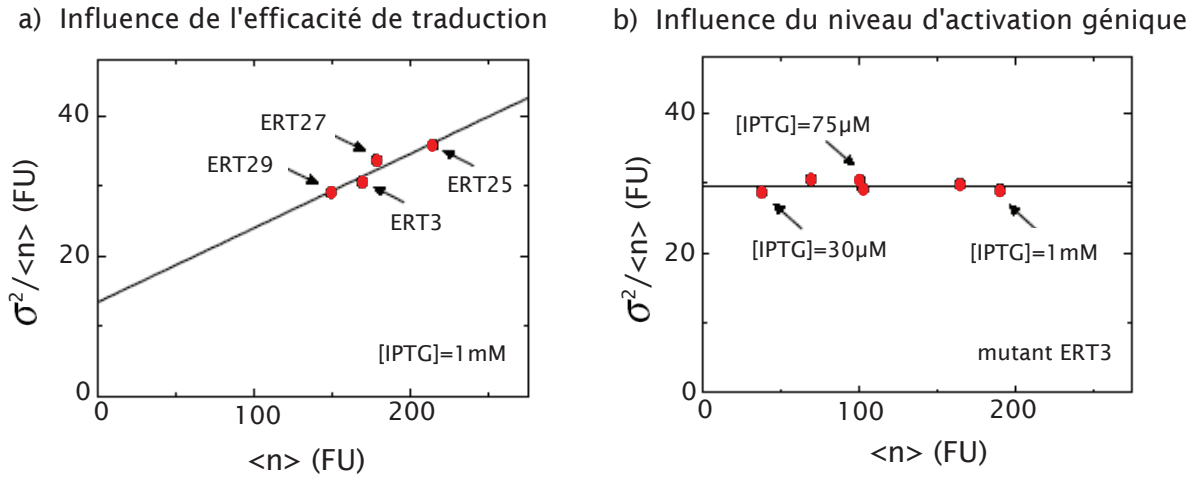


FIGURE I.1.15.: Propension du module à produire les protéines par bouffée : plus le paramètre $\sigma_n^2/\langle n\rangle$ est grand devant 1, plus la production de protéines se fait par paquet. a) Propension mesurée pour un gène activé par la même concentration d'inducteur (IPTG), mais muté à différentes efficacités de traduction (mutants ERT3, ERT25, ERT27 et ERT29). b) Propension mesurée pour un gène muté à la même efficacité de traduction, mais activé avec différentes forces : la concentration d'inducteur (IPTG) est variée. D'après Ozbudak et al[20].

Ces résultats sont en accord avec le modèle à deux étages formulé dans la section I.1.3 précédente. On note toutefois que dans les données reccueillies par Ozbudak et al, le gène n'est pas toujours actif, alors qu'il est fixé à son niveau d'induction maximal dans le modèle à deux étages. De plus, on ne sait pas si les sources extrinsèques sont réelement négligeables

15. Pour plus d'informations sur cette protéine fluorescente, voir section II.1.1.
16. voir paragraphe "La traduction" de la section I.1.1 précédente.

devant les sources intrinsèques. Qui plus est, un modèle qui negligerait l'aléa de rencontre dans la traduction et la dégradation protéique fournit la même dépendance en k_p [4]. En effet, dans ce cas $\sigma_n^2/\langle n\rangle = k_p/\gamma_m$. Grande prudence donc sur des conclusions hâtives quant à l'adéquation du modèle avec les données expérimentales. On retient toutefois que cette étude montre que les processus d'expression génique dévient fortement d'un processus poissonnien, et que cette déviation est fortement corrélée à l'étage de traduction.

Signature des fluctuations statiques sur l'ensemble du protéasome d'*E. Coli*

Les résultats expérimentaux précédemment établis sur *E. Coli* sont retreints à un gène particulier, avec le promoteur P_{lac}. Pour connaître la signature du bruit d'expression génique sur d'autres gènes, Y.Taniguchi et al [21] ont mesuré la variabilité d'expression génique sur un large nombre de protéines. Ces mesures ont été effectuées sur des cellules *E. Coli* en phase stationnaire de croissance exponentielle. Pour sonder cette variabilité inter-cellulaire dans le nombre de protéines, ils ont inséré des rapporteurs YFP dans le chromosome des bactéries. L'insertion est telle que la protéine fluorescente fusionne avec la protéine d'intérêt : le niveau de fluorescence détecté par microscopie renseigne directement sur la quantité de protéines d'intérêt. De plus, ce niveau de fluorescence a été approximé en nombre de protéines. Et pour pouvoir sonder plusieurs protéines en parallèle, Y.Taniguchi et al ont développé un circuit microfluidique où chaque canal contient des cellules qui rapportent l'activité d'une protéine différente. Les cellules sont collées dans le canal avec de la poly-L-lysine et elles sont sondées lorsqu'elles sont en pleine croissance exponentielle.
Les résultats obtenus montrent que les variations inter-cellulaire dans la quantité de protéines peuvent être décrites par une distribution gamma, comme illustré sur la figure I.1.16 ci-dessous pour deux protéines différentes.

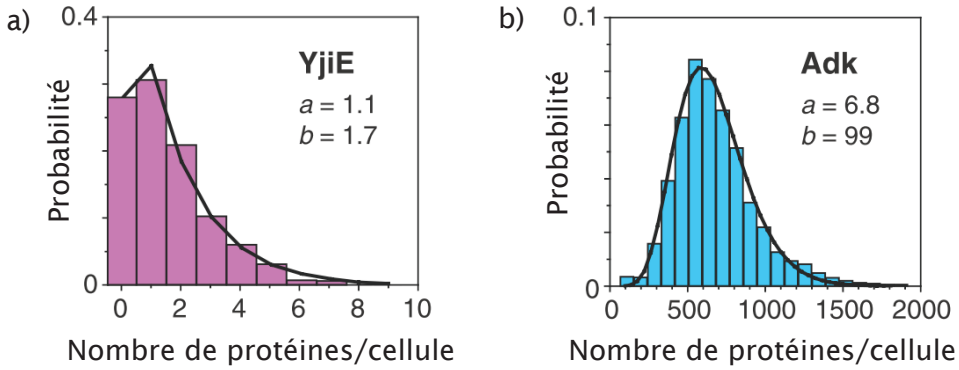


FIGURE I.1.16.: Etude statique du bruit d'expression génique : distribution du nombre de pro--téines par cellule. a) distribution pour la protéine Adk fusionnée avec la YFP. b) distribution pour la protéine YjiE fusionnée avec la YFP. D'après Y.Taniguchi et al [21].

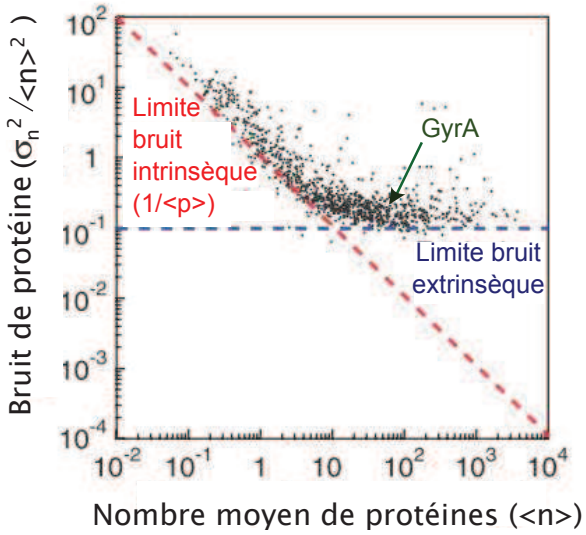


FIGURE I.1.17.: D'après Y.Taniguchi et al [21].

Cette étude a été menée sur 1018 protéines différentes, et pour beaucoup (99%) on retouve une description par la distribution gamma. Ces travaux font aussi apparaître une forte dépendance du niveau de bruit η^2 sur la concentration moyenne de protéines par cellule. En effet, sur la figure I.1.17 ci-contre, chaque point donne le niveau de bruit mesuré sur une protéine différente (par exemple, le point fléché caractérise la protéine GyrA). Ce niveau de bruit est tracé en fonction de la quantité de protéines présentes dans la cellule, et on constate qu'il est inversement proportionnel à la quantité de protéines sondées/cellule (la relation inverse limite est donnée par la courbe rouge). Cette dépendance inverse s'explique par la composante intrinsèque du bruit : dans un module d'expression génique qui ne considère que les sources intrinsèques de variabilité, le bruit d'équilibre statistique est inversement proportionnel au nombre de protéines/cellule. Et lorsque ce nombre de protéines devient important, le bruit généré par les sources intrinsèques devient faible, et le bruit est dominé par les sources extrinsèques. Dans ce cas, la précision du module est limitée par les sources extrinsèques. Et dans ces sources extrinsèques, celles qui sont globales affectent tous les gènes de la même maniére. Cela se traduit sur la figure I.1.17 par un plateau (en bleu) au dessous duquel le niveau de bruit ne peut pas descendre.

Fluctuations générées sur la variable m

La production d'ARNm peut, elle aussi, se faire par bouffées

I. Golding et al[14] se sont intéressés directement au bruit stochastique sur la variable m : ils ont mesuré les fluctuations statiques, et dynamiques, sur le nombre d'ARNm/cellule. Ces mesures ont été réalisées sur des cellules *E. Coli* pour un gène contrôlé par le promoteur P_{lac}. Et pour pouvoir mesurer *in-vivo* le nombre d'ARNm synthétisés dans une cellule, ils ont utilisé des rapporteurs fluorescents plasmidiques. Leur système rapporteur est constitué de deux éléments :

- La *Green Fluorescent Protein* (GFP) qui rapporte le nombre d'ARNm présents dans une cellule : la séquence codante de la GFP est fusionnée à une protéine capable de reconnaître l'ARNm ciblé. Cette protéine de reconnaissance est dénommée MS2, et la fusion protéique est insérée sur un plasmide.

- L'ARNm cible, qui contient 96 sites de liaison pour la protéine MS2. La transcription de cet ARNm est contrôlée par le promoteur P_{lac}, et il code pour la *Red Fluorescent Protein* (RFP). Ce système rapporteur est porté par un plasmide indépendant, régulé à 1 copie/genôme.

L'expression de chacun de ces éléments peut être contrôlée de manière indépendante par l'ajout de molécules inductrices au milieu de culture. Et lorsque les fusions protéiques "MS2-GFP" sont exprimées en parallèle de l'ARNm cible, elle viennent s'y accoler en nombre, générant ainsi un spot lumineux qui se dénote du reste de la cellule. L'intensité de ce spot reflète le nombre de GFP excitées, et comme le nombre moyen de fluorophores liés à un brin d'ARNm est connu, elle indique le nombre de transcrits compris dans le spot lumineux. Ainsi, en mesurant l'intensité des différents spots lumineux, I.Golding et al ont pu quantifier le nombre d'ARNm présents dans une cellule.

A partir de là, ils ont mesuré la variabilité inter-cellulaire dans le nombre d'ARNm/cellule. Ils montrent que cette variabilité ne suit pas un processus de Poisson : le paramètre $\sigma_m^2/\langle m \rangle$ est plus grand que 1, il vaut $\approx 4.1 \pm 0.5$. Dans ces conditions, les ARNm ne sont pas produits par des événements indépendants : ils sont produits par bouffées, avec de courtes périodes pendant lesquelles le gène est actif. Une étude dynamique qui mesure la quantité d'ARNm/cellule au cours du temps, confirme d'ailleurs ce modèle de production par bouffées. Et tout comme pour les protéines de J.Yu et J.Xiao et al [19], la fréquence des bouffées suit une loi de probabilité poissonienne, et l'amplitude une loi de probabilité géométrique. On souligne ici que dans l'étude de I. Golding et al, le module d'expression génique étudié se résume à l'activation génique et à la transcription :

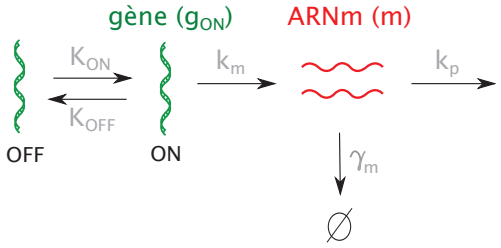


FIGURE I.1.18.: Modélisation des mécanismes impliqués dans la production et la dégradation des ARNm.

La propension du module à synthétiser les ARNm par bouffées est dictée par le nombre moyen d'ARNm

Les mesures précédentes concernent un gène particulier, avec le promoteur P_{lac}. Pour comprendre le bruit de transcription de manière plus large, Lok-hang So et al[22] ont récémment mesuré les fluctuations du nombre d'ARNm/cellule pour 20 promoteurs diffé-

rents chez *E. Coli*, chacun induit à différents niveaux (le nombre moyen d'ARNm résultant varie sur 3 ordres de grandeurs).

Le système rapporteur utilisé est semblable à celui du paragraphe précédent, mais il n'autorise pas le suivi dynamique des cellules. En effet, à présent le système de reconnaissance n'est plus produit par la bactérie, mais il est inséré par le scientifique. Il est constitué d'oligonucléotides [17] marqués par un colorant fluorescent. Ces petits morceaux d'ADN sont insérés dans le milieu de culture, et une fois intégrés par la cellule, ils reconnaissent un morceau spécifique du gène cible.

Avec ce système, Lok-hang So et al ont mesuré la variabilité intercellulaire pour les différents promoteurs. Cette variabilité a été étudiée à travers le niveau de bruit η^2_{total} et le paramètre $\sigma^2_m/\langle m\rangle$ qui quantifie la déviation par rapport à un processus poissonnien. Et en s'appuyant sur la valeur de $\sigma^2_m/\langle m\rangle$ par rapport à 1, Lok-hang So et al ont pu retrouver la production d'ARNm par bouffées précédément observées sur le promoteur P_{lac}.

Mais ils montrent aussi que cette production par bouffées n'est pas une régle générale. Et les études menées sur des modules de transcription différents montrent que la propension à synthétiser les ARNm par bouffées est indépendante de l'architecture régulatrice du gène étudié, et de la force de régulation exercée sur le gène. Elle est par contre, fortement liée au nombre moyen d'ARNm/cellule. Leurs données expérimentales sont présentées sur la figure ci-dessous, et elles suggèrent que lorsque le nombre moyen d'ARNm/cellule est inférieur à ≈ 0.2, le bruit de transcription suit un simple processus poissonnien.

Le bruit de transcription est dominé par les sources intrinsèques de bruit

Les données de Lok-hang So et al suggèrent aussi que le bruit de transcription est dominé par les sources intrinsèques. En effet, sur les figures précédentes, on observe que :

- Pour chaque promoteur, le niveau de bruit η^2 est inversement proportionnel au nombre moyen d'ARNm. Cette signature est spécifique d'un modèle à deux étages (ici, le modèle de la figure I.1.18) qui ne considère que les sources intrinséques de variabilité.
- Lorsque le nombre moyen d'ARNm est faible, les processus stochastiques en jeux s'approchent d'un processus Poissonien : $\sigma^2_m/\langle m\rangle \approx 1$. Ce résultat est typique des sources intrinsèques.
- Dans la limite des fortes quantités d'ARNm/cellule, le niveau de bruit décroît abruptement, au lieu d'atteindre un plateau, comme c'est le cas dans la figure I.1.17 précédente qui concerne le bruit des protéines.

Ces données montrent aussi que plusieurs copies du même gène (générées par la réplication de l'ADN avant la division cellulaire) génèrent des fluctuations statiques sur m qui sont

17. petite séquence d'ADN.

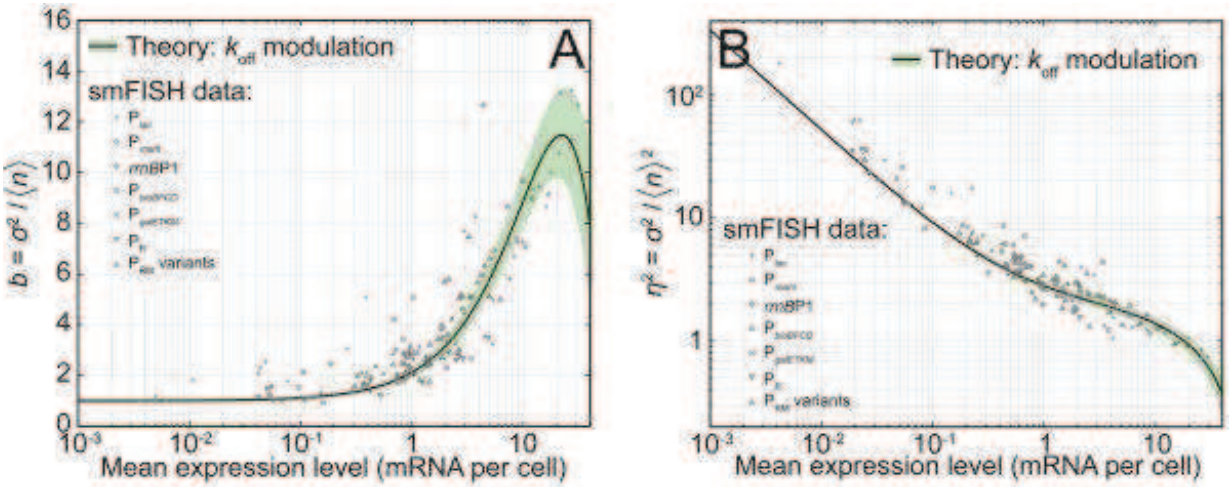


FIGURE I.1.19.: Mesures des fluctuations générées sur le nombre d'ARNm/cellule par différents modules de transcription. Chaque point représente un module de transcription. Les différents modules représentent 20 promoteurs différents, étudiés chacuns sous différents niveaux d'induction. a) Propension du module à synthétiser les ARNm par bouffées. Cette propension est caractérisée par $\sigma_m^2/\langle m \rangle$, et elle est tracée en fonction du nombre moyen d'ARNm/cellule,$\langle m \rangle$. b) Niveau de bruit η^2 en fonction du nombre moyen d'ARNm/cellule. D'après Lok-hang So et al[22].

corrélées. En effet, sur les cellules en fin de cycle cellulaire (longues cellules), la quantité d'ARNm est doublée [18], mais pas la variance : elle est triplée. Or, si le gène et son répliquat sont indépendants dans la cellule mère, on doit trouver : $\sigma^2_{m1+m2} = \sigma^2_{m1} + \sigma^2_{m2} = 2\sigma^2_{m1}$ où m_1 désigne la quantité d'ARNm issue du gène initial de la cellule mère, et m_2 la quantité d'ARNm issue du réplicat du même gène. Ici, la variance n'est pas doublée lorsqu'elle est mesurée sur les cellules en fin de cycle cellulaire. Donc les deux gènes ne sont pas indépendants. D'ailleurs, Lok-hang So et al trouve une covariance positive et non nulle, qui après normalisation au nombre d'ARNm/cellule, vaut en moyenne :

$$\frac{\sigma^2_{2\text{ copies}} - 2\sigma^2_{1\text{ copie}}}{2\langle m \rangle_{1\text{ copie}}} = 0.4 \pm 0.1$$

où $\sigma^2_{2\text{ copies}}$ désigne la variance (dans le nombre d'ARNm/cellule) mesurée sur les cellules en fin de cycle cellulaire. C'est donc la variance mesurée sur les cellules qui contiennent deux copies du même gène.

Séparation des différentes composantes

L'étude précédente montre que les sources extrinsèques ont un effet négligeable sur l'étage de transcription. Mais lorsqu'on s'intéresse au bruit sur la quantité de protéines/cellule,

18. En fin de cycle cellulaire, l'ADN est répliqué pour transmettre le matériel génétique aux deux cellules filles. La cellule mère contient alors deux copies du même gène. Cela se traduit sur les mesures d'ARNm, par une quantité d'ARNm/cellule qui est $\approx$ doublée.

ces sources extrinsèques peuvent être non négligeables, comme montré dans la figure I.1.17 précédente.

Bruit intrinsèque et bruit extrinsèque

Ainsi, en 2002 M.B.Elowitz et al.[5, 23] ont mis en évidence la présence de deux contributions (intrinsèque et extrinsèque) dans la variabilité inter-cellulaire observée (fluctuations statiques). Pour ce faire, ils ont introduit deux gènes rapporteurs dans le même chromosome de cellules *E. Coli.* Ces deux gènes rapporteurs codent pour deux protéines fluorescentes (presques identiques) aux propriétés spectrales différentes, la *Cyan Fluorescent Protein* (CFP) et la *Yellow Fluorescent protein* (YFP) [19] : cela permet de comparer le niveau d'expression des deux gènes dans la même cellule. De plus, les deux gènes rapporteurs sont activés par le même promoteur. Ils sont donc soumis au même environnement intra-cellulaire, c'est à dire aux mêmes variables extrinsèques. Dans ces conditions, si on regarde une cellule à un instant donné (microscopie), la différence d'expression entre les deux gènes traduit principalement les sources intrinsèques de bruit [5, 6]. Lorsqu'on regarde maintenant deux cellules au même instant, les variations d'expression entre les deux cellules traduisent toutes les sources de bruit (intrinsèques et extrinsèques). Elles sont caractérisées par η^2_{total}. M.B.Elowitz et al suggèrent que ce bruit total peut se décomposer en deux, avec une composante induite par les sources intrinsèques (η^2_{int}), et une autre par les sources extrinsèque (η^2_{ext}) :

$$\eta^2_{total} = \eta^2_{int} + \eta^2_{ext}$$

A partir de là, deux définitions expérimentales ont été établies :

- Le bruit intrinsèque η^2_{int} est la différence d'expression de deux gènes identiques, placés dans la même cellule (même environnement intra-cellulaire). Il traduit principalement les sources intrinsèques.

  $$\eta^2_{int} = \frac{\langle (c-y)^2 \rangle}{2\langle c \rangle \langle y \rangle}$$

  où c et y désignent les niveaux de fluorescence mesurés respectivement sur les rapporteurs CFP et YFP. Pour être comparables, ces niveaux de fluorescence ont été normalisés à leur valeur moyenne.
- Le bruit extrinsèque η^2_{ext} est la covariance des deux rapporteurs :

  $$\eta^2_{ext} = \frac{\langle cy \rangle - \langle c \rangle \langle y \rangle}{\langle c \rangle \langle y \rangle}$$

19. Pour plus d'information sur ces protéines fluorescentes, voir sectionII.1.1.

Il définit à quel point les variations intra-cellulaires des deux rapporteurs sont corrélées. Autrement dit, il indique à quel point les variations intra-cellulaires sont induites par les changements d'environnement intra-cellulaire.

Tout ceci est illustré sur la figure I.1.20.

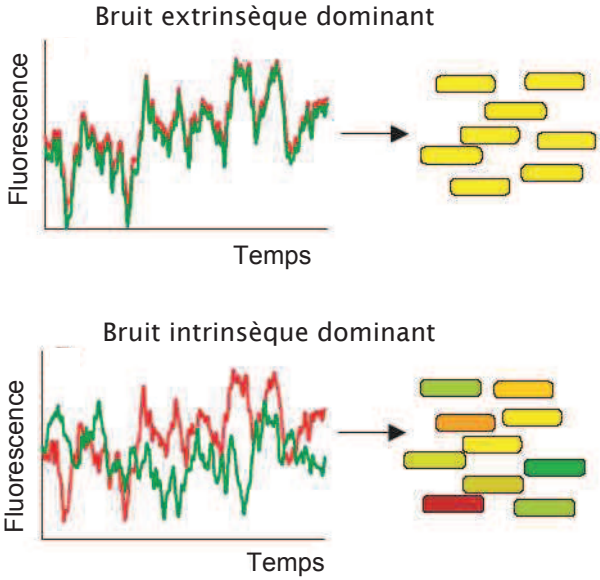


FIGURE I.1.20.: Bruit intrinsèque et bruit extrinsèque. Le bruit intrinsèque caractérise les différences d'expression génique dans la même cellule : lorsqu'il est dominant, les fluctuations dynamiques sont décorrélées dans la même cellule. Le bruit extrinsèque, quant à lui, caractérise la variabilité inter-cellulaire générée par le changement d'environnement intra-cellulaire : lorsqu'il est dominant, les fluctuations dynamiques sont corrélées dans la même cellule. D'après M.B.Elowitz et al [5].

Ces deux composantes du bruit ont été mesurées sur le promoteur P_{lac} qui est inhibé par la protéine LacI. Plusieurs mesures ont été faites, chacune correspondant à une situation différente où entre autre, le gène rapporteur est soit fortement activé, soit inhibé. Les résultats obtenus permettent de quantifier le bruit d'expression génique, avec η_{int} allant de 5 à 55% et η_{ext} de $2,4$ à 80% selon la situation. Ainsi, le bruit d'expression génique peut être non négligeable, avec des contributions à la fois intrinsèques et extrinsèques.

Des bruits intrinsèques et extrinsèques mal évalués ?

Le système de M.B.Elowitz et al suppose que les sources intrinsèques et extrinsèques apportent des contributions orthogonales au niveau total de bruit (η^2_{total}). Or des études théoriques récentes [11, 6] suggèrent que ce n'est pas toujours le cas. En effet, si les sources extrinsèques fluctuent rapidement devant l'échelle de temps associée à réactions d'expression génique (intrinsèques), alors l'environnement cellulaire perçu par les deux

gènes n'est pas le même. Dans ces conditions, les deux gènes rapporteurs ne sont pas soumis au même environnement intracellulaire. Et le bruit intrinsèque traduit les sources intrinsèques, mais aussi l'effet des fluctuations extrinsèques rapides sur ces sources. Tout ceci est illustré sur la figure I.1.21.

De plus, l'étude de Lok-hang So et al [22] (présentée deux paragraphes plus hauts) montre que les sources intrinsèques ne décorrèlent pas complétement la quantité d'ARNm produite par les deux copies du même gène (dans cette étude, les deux copies sont issues de la réplication de l'ADN avant la division cellulaire). Ce dernier résultat suggère que le niveau de bruit η^2_{ext} mesuré par la covariance de deux rapporteurs peut sur-évaluer en effet les sources extrinsèques. Qui plus est, les deux rapporteurs fluorescents de M.B.Elowitz ne sont pas vraiment équivalents : une étude récente[44] a permis de mesurer le temps de maturation des deux protéines rapporteuses. Et nous verrons dans la section II.1.1 que ces temps différent de $\approx 38\%$ pour les deux rapporteurs. Dans ces conditions, la variabilité dans le niveau de fluorescence détecté sur les deux rapporteurs, traduit non seulement les sources extrinsèques, mais aussi les différences dans le temps de maturation. Dans ce cas, ce système de double rapporteur sur-évalue le bruit intrinsèque. D'où des conclusions difficiles à tirer sur l'ampleur réele du bruit extrinsèque par rapport au bruit intrinsèque.

Bruit extrinsèque global

Dans la limite du système présenté dans le paragraphe précédent, N.Rosenfeld et al.[24] ont montré que le bruit extrinsèque global pouvait être non négligeable chez *E. Coli*. Pour ce faire, ils ont construit une cascade de gènes, avec un gène qui régule négativement l'expression d'un autre gène. Le répresseur y est fusionné à la protéine fluorescente YFP, et le gène régulé exprime une autre protéine fluorescente, la CFP. Ce système permet donc d'observer simultanément la quantité de protéines régulatrices et la quantité de protéines régulées. Par ailleurs, l'expression de la protéine régulatrice peut être activée par l'ajout d'une molécule au milieu de culture. Lorsqu'elle n'est pas activée, la concentration du régulateur diminue avec la division cellulaire, et la production de protéines CFP est de moins en moins inhibée. Des mesures de microscopie en temps réel permettent alors d'étudier la réponse de la CFP à une diminution de la concentration de répresseurs, autrement dit la Fonction de Régulation Génique (GRF) qui est représentée sur la figure I.1.22 ci-après.

Les mesures effectuées sur des cellules individuelles montrent que cette GRF varie d'une cellule à l'autre. Ces variations inter-cellulaires sont caractérisées par la déviation standard de la GRF autour de sa valeur moyenne : elles sont de 55%. Cette variabilité peut provenir des sources intrinsèques, mais aussi de sources extrinsèques telles que le cycle cellulaire. En effet, ce dernier introduit des variations dans le nombre d'ADN, avec les cellules sur

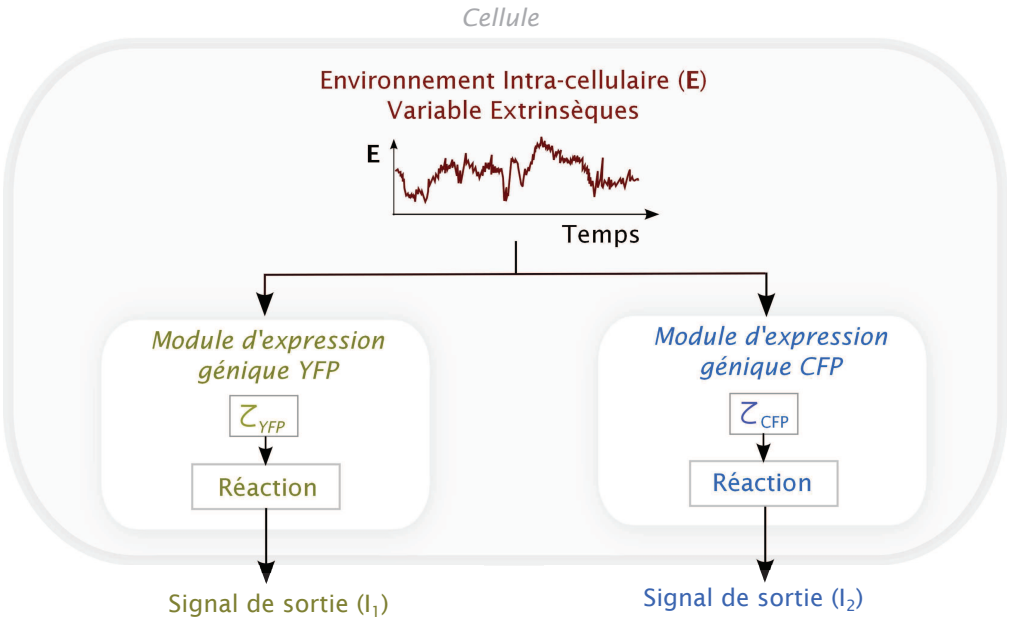


FIGURE I.1.21.: Effet des fluctuations extrinsèques sur le bruit η^2_{int} mesuré dans le système de M.B.Elowitz. Deux gènes rapporteurs codant pour la CFP et la YFP sont insérés dans le chromosome des cellules : ces deux gènes sont caractérisés par des modules d'expression génique équivalents. Dans ces modules, les réactions probabilistes se produisent à des temps alétoires. Les temps d'occurence des réactions d'expression génique sont donc différents pour le gène YFP et le gène CFP. Ces différences sont ici modélisées par un retard τ qui diffère pour les deux modules d'expression génique. Les signaux d'entrée des deux modules, eux, sont les mêmes : ce sont les variables Extrinsèques qui reflétent l'environnement intracelluliare. Ces variables fluctuent aléaoirement au cours du temps ; et comme le retard τ n'est pas le même pour les deux modules, les réactions d'expression des gènes YFP et CFP utilisent des valeurs différentes pour les variables extrinsèques. Dans ces conditions, les deux modules n'intègrent pas le même environnement cellulaire. Cela se traduira sur le signal en sortie du module par une variabilité qui diffère, non seulement à cause de l'aléa inhérent aux réactions intrinsèques d'expression génique, mais aussi à cause de l'environnement intra-cellulaire qui n'est pas perçu de la même manière par les deux gènes rapporteurs.

le point de se diviser qui ont répliqué leur matériel génétique. Dans ces cellules, il est montré que la production de CFP est à peu près le double de lorsqu'elle est mesurée sur des cellules fraîchement divisées. Pour s'affranchir de ces variations, la GRF a donc été normalisée par la phase du cycle cellulaire : la variabilité est réduite, mais elle reste conséquente comme montré sur la figure I.1.22. Elle vaut 40%.

Pour investiguer l'origine de ces variations (intrinsèques ou extrinsèques), la construction de M.B.Elowitz et al. [5] a été rajoutée : le répresseur régule maintenant l'expression de la CFP, mais aussi celle d'un rapporteur supplémentaire. Les deux rapporteurs sont soumis au même environnement cellulaire, ils ne différent donc que par la stochasticité intrinsèque d'expression génique. Les mesures montrent alors qu'au sein d'une même cellule, la production de protéines varie de 20% entre les deux rapporteurs. Cela suggère que

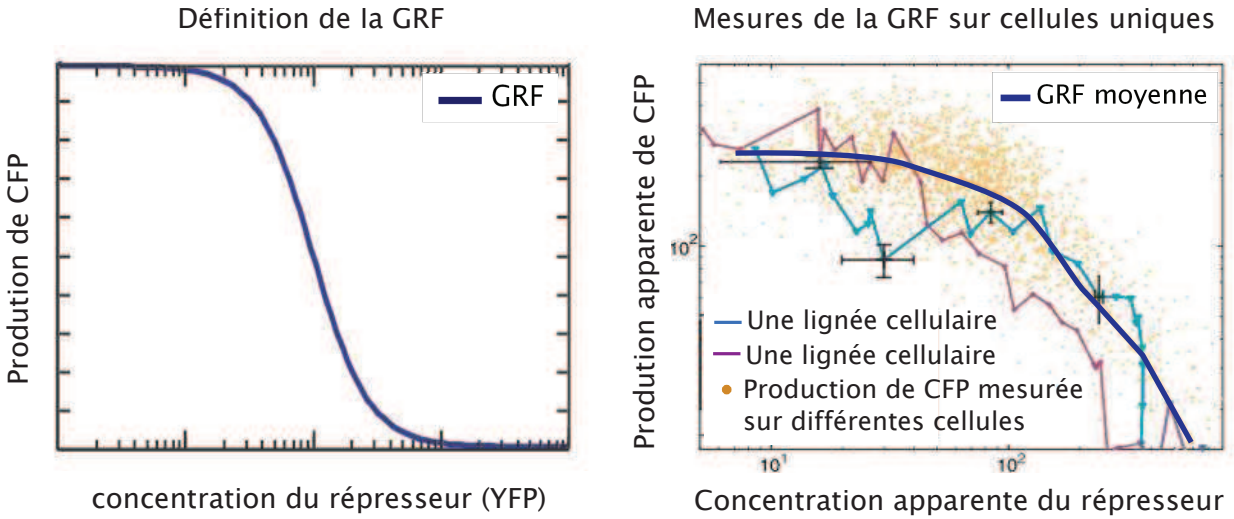


FIGURE I.1.22.: A gauche, définition de la fonction GRF : elle exprime la relation entre la concentration de protéine régulatrice, et le taux d'expression de la protéine régulée. A droite, les mesures de GRF effectuées sur cellules uniques : les données montrées sont normalisées par la phase du cycle cellulaire pour s'affranchir du bruit généré par le cycle cellulaire, et chaque point orangé représente la mesure d'une cellule. Ces points sont dispersés autour de la GRF moyenne (bleu marine). Cette dispersion peut être visualisée plus en détail sur deux lignées cellulaires différentes (courbes bleu-cyan et mauve) imagées toutes les 9 minutes. D'après N.Rosenfeld et al [24].

les 55% de variabilité inter-cellulaire observée sur la GRF sont principalement induits par des sources extrinsèques. Et comme la GRF englobe la concentration du répresseur, cette variabilité extrinsèque n'est pas le fruit de fluctuations dans la quantité de répresseur : elle n'est pas spécifique, mais globale, générée plutôt par les fluctuations dans les facteurs globaux d'expression génique, tels que les ribosomes.

Dynamique du bruit extrinsèque global

Cette étude a aussi mis en évidence la dynamique lente du bruit extrinsèque. En effet, sur la figure I.1.22 précédente, on peut voir que les lignées individuelles dévient de la GRF moyenne pendant des périodes relativement longues. D'ailleurs, lorsque Rosenfeld et al. regarde le temps d'autocorrélation de la production de CFP, ils trouvent une valeur de $\tau_{corr} = 40 \pm 10$ min, ce qui est proche de la durée d'une cycle cellulaire ($\tau_{cycle} = 45 \pm 10$ min). Par conséquent, les fluctuations persistent sur un cycle cellulaire, donnant ainsi une *mémoire* à la cellule. De plus, une étude supplémentaire menée sur les doubles rapporteurs chromosomaux montre que dans la même cellule, les fluctuations des deux rapporteurs sont corrélées sur une durée $\tau_{corr,int} < 10$ min. Par conséquent, ces fluctuations lentes sont principalement le résultat des sources extrinsèques globales.

Bruit introduit par les différences de cycle cellulaire

Dans la population de cellules sondées, les individus se trouvent dans des états cellulaires différents : certains sont avancés dans le cycle cellulaire, et d'autres viennent tout juste de naître. Nous avons vu à travers les travaux de N.Rosenfeld et al [24], que ces différences d'état cellulaire contribuaient à générer de la variabilité sur le nombre de protéines/cellule. Et pour s'affranchir de cette contribution, les données de fluorescence sont souvent normalisées à la taille cellulaire. Y.Taniguchi et al[21] ont ainsi montré qu'après normalisation à la taille cellulaire, le niveau de bruit des protéines est inchangé lorsqu'il est mesuré sur les cellules fraîchement divisées, et lorsqu'il est mesuré sur les cellules qui sont sur le point de se diviser. Ces résultats suggèrent que le bruit introduit par la phase du cycle cellulaire peut être corrigé par une simple normalisation à la taille de la cellule.

Toutefois, les mesures de bouffées de production protéique réalisées par J.Yu et al[19] montrent que la fréquence a des bouffées dépend du cycle cellulaire : pour les cellules en fin de cycle cellulaire, les bouffées de production protéiques apparaissent plus fréquemment. Ils interprétent ce résultat comme l'effet de la réplication de l'ADN, avec un gène qui se trouve en copies multiples dans les cellules sur le point de se diviser, et en copie unique dans les cellules fraîchement divisées. Ils notent que cette dépendance n'affecte pas l'allure poissonienne de la distribution de la fréquence des bouffées. Mais on retient que l'état cellulaire peut modifier les caractéristiques du bruit mesuré en sortie du module d'expression génique.

Bruit de partitionnement lors de la division cellulaire

Le système rapporteur utilisé dans les premières expériences de I. Golding et al[14] a permis de mesurer, en plus du nombre d'ARNm/cellule, le nombre de protéines/cellule. I. Golding et al ont ainsi pu étudier la corrélation entre le nombre d'ARNm/cellule, et le nombre de protéines/cellule. Ce niveau de corrélation a été mesuré sur les cellules longues qui ne se sont pas encore divisées [20], et sur les petites cellules qui viennent de se diviser. Et leurs résultats montrent que les deux grandeurs sont moins corrélées après la division cellulaire : le coefficient de corrélation vaut 0.23 ± 0.05 lorsqu'il est mesuré sur les grandes cellules, et il est inférieur à 0.04 lorsqu'il est mesuré sur les cellules fraîchement divisées. Cela suggère que la division cellulaire, avec le partitionnement binomial du matériel génétique[14, 24], décorrèle les deux grandeurs. Pour les protéines qui se trouvent en grand nombre dans la cellule, la division cellulaire répartit de manière $\approx$ équitable les protéines dans les cellules filles. Mais pour l'ARNm qui se trouve en faible quantité dans la cellule, le partitionnement binomial introduit de grandes différences entre les cellules

20. Une grande cellule signifie ici une cellule dont la taille est supérieure à la valeur médiane de taille cellulaire.

filles : la quantité d'ARNm n'est pas répartie équitablement entre les deux cellules filles. Donc, le partitionnement induit par la division cellulaire introduit de la variabilité dans les variables intrinsèques au module d'expression génique : il représente une source extrinsèque de bruit. Cet aléa de partitionnement est souvent négligé dans les modèles qui décrivent les fluctuations mesurées dans les expériences. Pourtant cette conclusion n'est pas triviale. En effet, une étude théorique[25] montre que ce partitionnement peut, à lui seul, générer la même signature que les sources intrinsèques.

Bruit du plasmide

Pour les systèmes rapporteurs portés par des plasmides, le module d'expression génique contient une source supplémentaire de bruit : le bruit introduit par les plasmides. En effet, ceux-ci se trouvent en quantité inégale dans la population de cellules : la division cellulaire répartie aléaoirement ces plasmides entre les deux cellules filles, générant des différences dans le nombre de système rapporteurs présents. Et ces différences dans le nombre de copies du gène rapporteur peuvent contribuer à la variabilité générée en sortie du module d'expression génique.
J.Wong Ng et al [26] ont mesuré ces différences dans le nombre de systèmes rapporteurs plasmidiques chez *E. Coli*. Pour ce faire, ils ont utilisé un système de doubles rapporteurs aux propriétés statistiques équivalentes, mais aux propriétés spectrales différentes : un gène codant pour la mOrange est inséré sur un plasmide, et un gène codant pour la EFGP est inséré directement dans le chromosome des cellules. La variabilité inter-cellulaire est alors mesurée sur le système plasmidique, et le système chromosomal permet de soustraire la composante du bruit induite par les mécanismes d'expression génique. Une fois purifiée de l'aléa d'expression génique, la variabilité mesurée sur le système plasmidique traduit les différences dans le nombre de plasmides/cellule. Ils montrent ainsi que pour des plasmides présents en grand nombre dans les cellules, la variabilité dans le nombre de copies du plasmides est réduite : elle est de 25% pour des mini-plasmides qui possèdent une origine de réplication ColE1.

I.1.5. Rôle du bruit stochastique d'expression génique

Le bruit d'expression génique génère de l'imprécision dans les réseaux de régulation protéique. Il peut représenter une nuisance pour la cellule, et elle a du développer des mécanismes de contrôle pour fonctionner de manière fiable. De plus, nous allons voir en ce qui suit que cet aléa peut s'avérer avantageux [27] : il permet aussi d'implémenter certaines fonctionnalités.

De nombreuses études suggèrent ainsi que ce bruit peut servir de "catalyseur" pour

initier une transition d'état. Par exemple, lorsque les bactéries *B. subtilis* rentrent en phase stationnaire, une partie ($\approx 10 - 20\%$) développe la capacité d'intégrer l'ADN de l'environnement extra-cellulaire (cellules compétentes). La transition vers l'état de compétence est gérée par une boucle d'autorégulation positive, et deux études expérimentales distinctes [28] montrent qu'en réduisant le bruit sur le module autorégulé, on diminue la fraction de cellules compétentes. Ces résultats suggèrent donc que les fluctuations d'expression génique larges sont à l'origine de la transition d'état. D'ailleurs, plus récemment, T.L.To et al [29] ont utilisé un système synthétique pour montrer que, là où les études déterministes prédisent un comportement uniforme des cellules, le bruit stochastique permet de générer la population de cellules hétérogène observée. Leurs expériences ont été menées chez *S. Cerevisae* pour un module d'expression génique qui s'autorégule positivement. Les résultats obtenus font alors apparaître des distributions bimodales du niveau d'expression génique dans la population de cellules. Cette bimodalité est observée à l'équilibre statistique, et elle semble caractérisée par des bouffées de production protéique de grande amplitude et peu fréquentes. Dans ce cas, c'est l'amplitude large des bouffée qui permet de faire passer le système dans un état de forte activation génique (via la boucle d'autorégulation). Quant à la fréquence faible des bouffées, combinée à l'instabilité du régulateur, elle permet de faire transiter le module dans son état de faible expression avant la prochaine bouffée. On a donc coexistance de deux états métastables au sein de la population de cellules isogéniques.

Cette bistabilité a été observée dans de nombreux autres systèmes, avec notamment l'exemple historique du réseau d'utilisation du lactose chez *E. Coli*. Dans ce cas, le réseau génère deux états cellulaires stables qui correspondent à la métabolisation ou non du lactose [30, 31]. Dans les réseaux métaboliques tels que celui présenté, cette bistabilité peut permettre d'anticiper l'arrivée/carence de nouvelles sources de nutriments [32]. En effet, une partie des cellules se trouvent dans un état qui n'est pas favorisé par l'environnement actuel, mais lorsque cet environnement varie brutalement, ces cellules sont déjà adaptées aux nouvelles conditions. Dans des réseaux lents de transduction des signaux soumis aux fluctuations environementales, cette stratégie rapide d'adaptation peut être plus avantageuse [33].

Pour les bactéries pathogènes, on peut imaginer que la même stratégie est utilisée pour résister aux antibiotiques. En effet, après traitement antibiotique d'une infection bactérienne, la majorité des bactéries est éliminée rapidement, mais une petite fraction de cellules subsiste. Ces individus "persistants" ont une croissance cellulaire ralentie, ce qui leur permet de résister plus longtemps au traitement [34]. Lorsque l'antibiotique est supprimé de l'environnement, ces cellules résistantes transitent aléatoirement vers un régime de croissance, ce qui permet de relancer l'infection. Les individus persistants sont

parfois présents avant le traitement antibiotique [35]. Ils anticipent alors les risques de stress antibiotique au détriment de la prolifération en milieu favorable. On peut donc supposer que c'est la même stratégie de survie qui est à l'oeuvre, avec une population de cellules bistable (croissance cellulaire normale/ralentie).

Et plus récemment, une étude a montré que des pulses d'activation génique permettent de retarder la réponse cellulaire de plusieurs cycles de division [36]. Dans cette étude menée chez *B. subtilis*, le phénomène de sporulation [21] apparaît plusieurs cycles cellulaires après la déplétion de nutriments. Il a été montré que ce délai long de réponse est réalisé par une activation pulsée du régulateur clé. Ces pulses apparaissent environ une fois par cycle cellulaire, et une boucle d'autorégulation positive à retard permet d'amplifier l'amplitude d'un pulse à l'autre. Il faut donc un certain nombre de cycles cellulaires pour que le régulateur clé devienne suffisament abondant pour initier la transition vers l'état de sporulation.

Il existe de nombreuses autres études qui mettent en évidence le rôle fonctionnel du bruit d'expression génique. Pour de plus amples informations, nous vous invitons à lire la revue de A.Raj et A.Van Oudernaarden [37]. On retiendra simplement que les cellules ont appris à exploiter l'aléa d'expression génique pour développer des stratégies de survie et de développement plus efficaces. L'étude et la compréhension de ce bruit est donc nécessaire pour mieux comprendre le fonctionnement du vivant.

21. Suite la déplétion de nutriments, les bactéries se différentient en spores plus résistants à l'environnement

I.2. Système d'étude

Les travaux présentés précédemment se limitent à des environnements statiques : le bruit est mesuré lorsque l'environnement extra-cellulaire est stable et le système cellulaire en régime stationnaire (au sens dynamique). Pourtant dans la nature, cet environnement est variable et la cellule fonctionne rarement en régime stationnaire. Ici, nous nous proposons de déterminer la signature du bruit d'expression génique lorsque les individus bactériens répondent à un signal externe. Etant donné que cette signature reflète les mécanismes cellulaires en jeu, elle peut fournir des informations sur les processus cellulaires qui génèrent la réponse. Pour mener à bien cette étude, nous avons choisi un système bien caractérisé : nous étudions le bruit dans le réseau de régulation génique qui gére la réponse au stress nutritionnel des bactéries *E. Coli*. Ce réseau est étudié en parallèle par nos collaborateurs biologistes du Laboratoire d'Adaptation et Pathogénie des Microorganismes de Grenoble (L.A.P.M) et bioinformaticiens de l'INRIA-Grenoble, ce qui fournit une bonne indication sur l'architecture moyenne du réseau et les interactions en jeu (voir figure I.2.2).

a. La bactérie Escherichia Coli

La bactérie *E. Coli* est un organisme unicellulaire en forme de bâtonnet qui mesure environ 4µm de long. C'est une cellule procaryote [1] qui vit dans l'intestin. Et elle est bien connue de la recherche scientifique : elle est un "outil" important pour la biologie moléculaire et un "sujet" central pour la bactériologie. Pour les études en laboratoire, elle présente l'avantage de se diviser rapidement. Avec un cycle cellulaire de l'ordre de la demi heure, elle permet en effet de réaliser des expériences sur des durées relativement courtes par rapport à d'autres organismes [2]. On note que les expériences qui mesurent le bruit d'expression génique s'étalent sur$\approx 17h$, et que sur ces durées, la variabilité induite par les mutations génétiques est négligeable.

1. A l'inverse des cellules eucaryotes qui possèdent un noyau, les cellules procaryotes possèdent un filament d'ADN qui n'est pas protégé par une membrane nucléaire.

2. Par exemple, les cyanobactéries qui se divisent (au mieux) toutes les ≈ 8 heures , requièrent des expériences qui s'étalent sur plusieurs journées, voire des semaines.

b. La réponse au stress nutrionnel

Lorsque la bactérie *E. Coli* est dans son environnement naturel, elle rencontre rarement des conditions de croissance constantes. En effet, l'acidité du milieu, la température, la concentration en oxygène ainsi que les conditions nutritionnelles sont susceptibles de varier au cours du temps sous l'effet de perturbations. Ainsi, pour croître de manière optimale, la bactérie doit en permanence s'adapter à cet environnement changeant. Ceci implique de détecter ces variations et de développer des réponses appropriées. Dans le cas du stress nutritionnel, lié par exemple à un manque de carbone, la bactérie cesse de croître exponentiellement et entre dans un état de croissance plus résistant, la phase stationnaire (cf. Figure I.2.1).

Pendant cette transition, chaque bactérie développe une habileté à survivre aux périodes prolongées de famine et devient plus résistante aux stress. Cela se fait à travers d'importantes modifications morphologiques et physiologiques qui peuvent être renversées dès que les conditions de croissance sont rétablies. Citons à titre d'exemple, le métabolisme cellulaire dont l'activité originelle de croissance maximale est réduite à une activité de maintenance. Ce processus d'adaptation est contrôlé par un réseau de régulation complexe qui repose sur des interactions entre les gènes, les protéines, les métabolites, et d'autres molécules. On distingue souvent les réseaux métabolites, des réseaux de transduction des signaux ou des réseaux de régulation génique. Les premiers assurent la connexion entre les petites molécules de la cellule à travers des réactions enzymatiques, tandis que la transduction des signaux passe en grande partie par la modification post-traductionnelle des protéines. Enfin, les réseaux de régulation génique, sur lesquels nous nous focaliserons, concernent surtout les interactions entre protéines et acides nucléiques, contrôlant la transcription et la traduction des gènes.

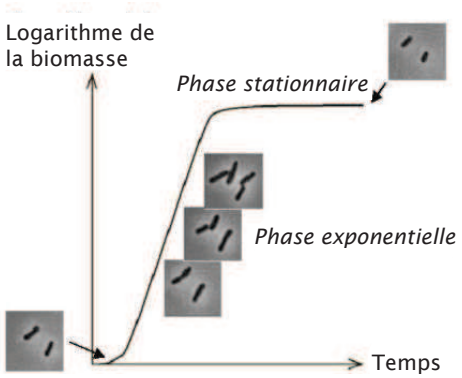


FIGURE I.2.1.: Etats de croissance d'une population de bactéries : phase exponentielle et stationnaire.

c. Le réseau de régulation qui gère la réponse au stress nutritionnel

Pour comprendre la réponse dynamique d'un réseau génétique complexe, nos collaborateurs bioinformaticiens ont développé un modèle qui décrit les interactions géniques

en jeu [38]. Et parallèlement à cette étude, nos collaborateurs biologistes ont développé des outils de génômiques [39] qui permettent de vérifier justement dans le cas du réseau de régulation qui gère la réponse au stress nutritionnel. Ces outils permettent de sonder la réponse des modules d'expression génique impliqués dans la réponse. Ils ont ainsi pu quantifier le signal en sortie des modules concernés. Le modèle proposé a pu être testé et réajusté. Dans ces deux études, la transition entre stress nutritionnel et croissance exponentielle est initiée par l'ajout de glucose au milieu de culture (environnement) défaillant en sources carbonnées.

Modèle

Un modèle simple (proposé par D.Ropers et al [38]) décrit l'adaptation de la croissance bactérienne au manque de carbone. Le réseau modélisé comprend 6 gènes codant des protéines dont l'activité dépend d'un signal de stress nutritionnel (le régulateur global CRP et l'adénylate cyclase Cya), des gènes impliqués dans le contrôle du métabolisme (le régulateur global Fis), la croissance cellulaire (les gènes *rrn* codant les ARN stables, considérés comme représentatifs de l'état de croissance), et le sur-enroulement de l'ADN, un modulateur important de l'expression génique (la gyrase GyrAB et la topoisomérase TopA). Pour décrire de manière plus détaillée notre réseau de régulation génique, nous le divisons en modules interdépendants (voir Figure I.2.2). On note que ce modèle a récemment été étendu pour intégrer les interactions entre les gènes et les métabolites ou molécules signalisatrices de la cellule [40]. Nous ne présentons ici que le premier modèle qui résume le réseau à six gènes en interaction.

Le module d'entrée : Ce module décrit la transition entre la détection du stimulus externe, un manque de carbone dans cette étude, et l'expression des gènes appropriés. L'étape de détection est simplifiée en considérant qu'un signal indicateur du manque de sources carbonées active directement l'enzyme Cya. L'enzyme activée catalyse alors la synthèse de l'adénosine monophosphate cyclique (cAMP) : ces petits métabolites, produits de l'adénosine triphosphate (ATP), viennent se lier aux protéines CRP. On a alors formation d'un complexe cAMP-CRP qui va pouvoir réguler l'expression de certains gènes. Le complexe ainsi activé est un facteur de transcription pour une multitude de gènes. Le résultat est que ce module s'autorégule en activant ou inhibant la transcription du gène *crp* ou en inhibant la transcription du gène *cya* (Figure I.2.2, module jaune).

Le module de régulation de Fis à travers CRP-cAMP : Ce module comprend deux gènes : *crp* et *fis*. Le complexe cAMP-CRP inhibe l'expression de *fis*. La protéine Fis à son tour contrôle l'expression de plusieurs gènes impliqués dans le métabolisme

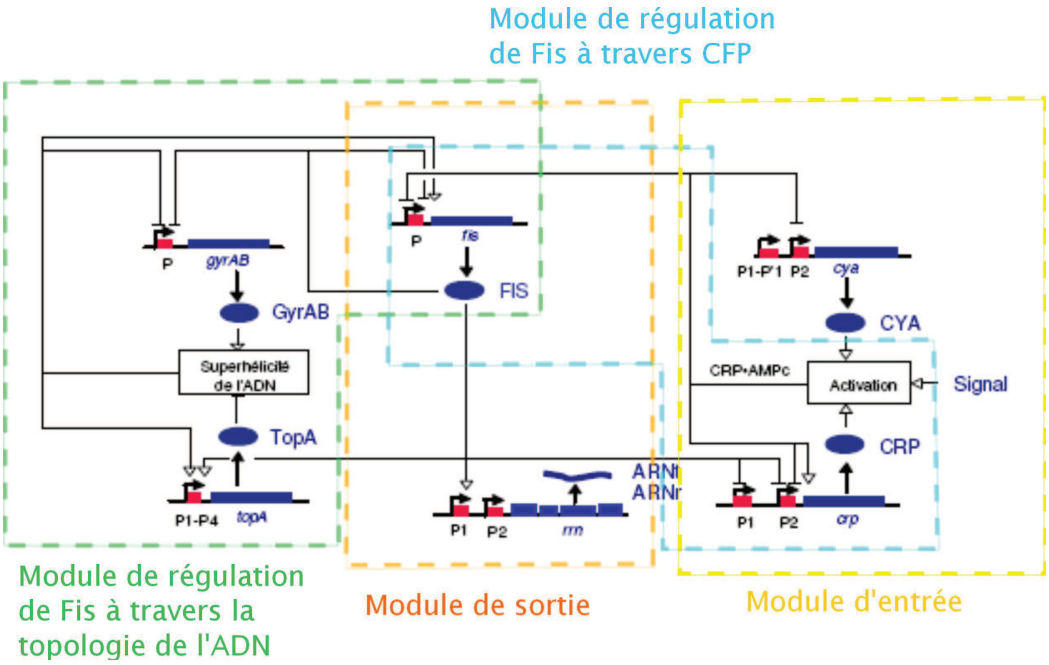


FIGURE I.2.2.: Réseau des gènes clés, protéines et interactions régulatrices impliqués dans la réponse au stress en carbone chez *E.Coli*. On distingue quatre modules : un module d'entrée (en jaune), un module de régulation de Fis à travers CRP (en cyan), un module de régulation de Fis par la topologie de l'ADN (en vert) et un module de sortie (en orange) qui relie la concentration de Fis au nombre d'ARN stables produits. D'après D. Ropers et al[38].

cellulaire : lorsque la protéine Fis est produite, elle inhibe la transcription du gène *crp* : les deux protéines s'inhibent mutuellement (Figure I.2.2, module cyan).

Le module de régulation de Fis à travers la topologie de l'ADN : En plus de contrôler l'expression du gène *crp*, la protéine Fis régule la topologie de l'ADN. Et la topologie de l'ADN influence la transcription de plusieurs gènes. En effet lorsque l'ADN circulaire, initialement sous forme relaxée, acquiert une superhélicité négative [3], la double hélice s'ouvre plus facilement. Réplication et transcription sont ainsi favorisées. Le contrôle de la superhélicité de l'ADN implique principalement deux topoisomérases : la topoisomérase GyrAB qui utilise l'énergie de l'ATP pour introduire un sur-enroulement négatif de l'ADN, et la topoisomérase TopA qui, à l'opposé, permet de relâcher les super tours négatifs de l'ADN. Ainsi, une augmentation des protéines GyrAB entraîne une augmentation de la superhélicité négative de l'ADN. Mais ceci a pour effet de favoriser la transcription des gènes *topA* et *fis*. On a donc fabrication de nombreuses protéines TopA et Fis qui vont de nouveau changer la superhélicité de l'ADN : TopA détruit le super enroulement négatif de l'ADN et Fis réduit la production de GyrAB. On a donc une diminution de la superhélicité de l'ADN. Inversement, cette diminution de la superhélicité annule l'inhibi-

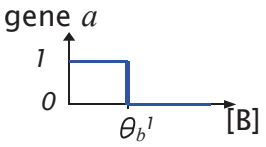
3. un enroulement de l'ADN inférieur à celui adopté par l'ADN sans contrainte.

tion des promoteurs gyrA et g*yrB*. La production de protéines GyrAB augmente et on a à nouveau une augmentation de la superhélicité de l'ADN. Le niveau de superhélicité de l'ADN est donc fortement autorégulé par les trois protéines Fis, GyrAB et TopA. Tout ceci est résumé dans le module vert de la Figure I.2.2.

Le module de sortie : Ce module contrôle la croissance cellulaire. Il comprend les gènes *fis* et l'ARN ribosomal, *rrn*. On considère que l'ARN ribosomal est représentatif de l'état de croissance de la cellule : la division et la croissance cellulaire requièrent un grand nombre de ribosomes pour traduire l'ARNm en protéines. Chez *E.Coli*, l'expression des gènes codant pour l'ARN stable est fortement stimulée par la protéine Fis (Figure I.2.2, module orange).

Simulations du modèle

Ce réseau, même simplifié à l'extrême, devient très complexe pour une description intuitive. Des formalismes mathématiques décrivant la dynamique des réseaux génétiques ont déjà été élaborés, notamment avec les équations différentielles ordinaires. Mais une telle approche requiert une parfaite connaissance des réactions biochimiques qui entrent en jeu, des informations quantitatives sur les concentrations moléculaires ainsi que des paramètres cinétiques. Ne possédant en général pas toutes ces connaissances dans les réseaux de transcription biologique, Hidde de Jong *et al* [41] ont élaboré un autre formalisme qui repose sur les équations différentielles linéaires par morceaux.
Dans ce formalisme, les interactions géniques sont modélisées par des fonctions échelon, comme représenté sur la figure ci-contre : sous l'effet d'une protéine régulatrice (B), le gène régulé (a) passe instantannée d'un état actif (1) à un état inactif (0). Les variables d'état correspondent alors aux concentrations des protéines ($[B]$ et $[A]$) codées par les gènes du réseau. Et les paramètres du modèle sont les seuils des fonctions échelon (θ_b^1).



Dans le cas de la réponse au stress nutritionnel d'*E. Coli*, le modèle comprend cinq équations différentielles qui décrivent les changements de concentration des cinq protéines impliquées : Cya, CRP, Fis, GyrAB et TopA. Une équation différentielle supplémentaire permet de décrire l'évolution de la concentration d'ARN stable. Et les paramètres (taux de synthèse ou concentrations des protéines régulatrices) sont modélisés par 48 inégalités contraintes (seuils des fonctions échelon) qui s'appuient sur les valeurs données dans des articles expérimentaux. A partir de ce modèle, des simulations sur le comportement de la bactérie en réponse à la présence ou l'absence de carbone ont pu être effectuées. Ces simulations fournissent une description qualitative du réseau, comme illustré sur la figure ci-dessous pour les protéines Fis et GyrA.

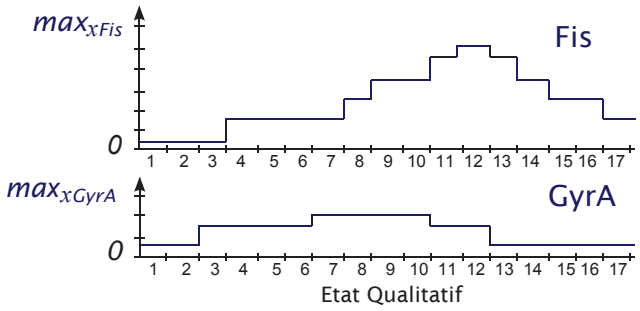


FIGURE I.2.3.: Simulation du réseau de régulation donné dans la figure I.2.2 : Evolution qualitative des concentrations de protéines Fis et GyrA lors de la réentrée en phase exponentielle des cellules *E. Coli*. L'échelle des abscisses donnent les états qualitatifs par lesquels les cellules passent avant d'atteindre la phase exponentielle : ces états sont caractérisés par les niveaux de concentration $(x_{Fis}, x_{GyrA}, x_{CRP}, x_{TopA}, x_{Cya}, rrn)$ des protéines du réseau etb d'ARN stable (rrn). Et l'échelle des ordonnées donnent la concentration qualitative des protéines. D'après D.Ropers et al [38].

Etude quantitative

Le modèle proposé par les bioinformaticiens repose sur une hypothèse forte : la régulation génique est instantanée avec une foncion échelon. Ce modèle nécessite donc d'être rigoureusement contrôlé par des expériences. Pour cela l'équipe du L.A.P.M dirigée par H.Geiselmann mesure le comportement dynamique de production des protéines impliquées dans le réseau. Pour ce faire, ils ont construit des rapporteurs fluorescents et luminescents dans *E. Coli*, et ce pour les différents noeuds du réseau. Ils suivent alors la fluorescence, la bioluminescence et l'absorbance sur microplaques, indépendamment pour chaque noeud du réseau. Ils déduisent ensuite le taux de synthèse des protéines rapportées ainsi que leurs concentrations dans les cellules[39]. Ils s'appuient ici sur un modèle qui décrit le module d'expression génique de la protéine sondée. Leur méthode a été validée par des mesures directes de Western blot et Nothern blot pour une protéine du réseau, Fis. En effet, sur la figure I.2.4, la concentration de GFP tirée des mesures de fluorescence (en bleu) affiche une même dynamique que la concentration de la protéine Fis rapportée (en noir).

d. Nos travaux de recherche

Ces études se sont focalisées sur la partie déterministe de la réponse. En effet, les bioinformaticiens négligent le bruit dans les différents modules d'expression génique, et les biologistes réalisent des expériences en population sans accéder à la variabilité inter-cellulaire. Nous nous proposons de compléter ces travaux, en étudiant la stochasticité d'expression génique dans le réseau de régulation. Nous nous sommes ici intéressées au bruit d'expression génique pour un noeud clé du réseau : GyrA. On désire, pour ce composant, suivre

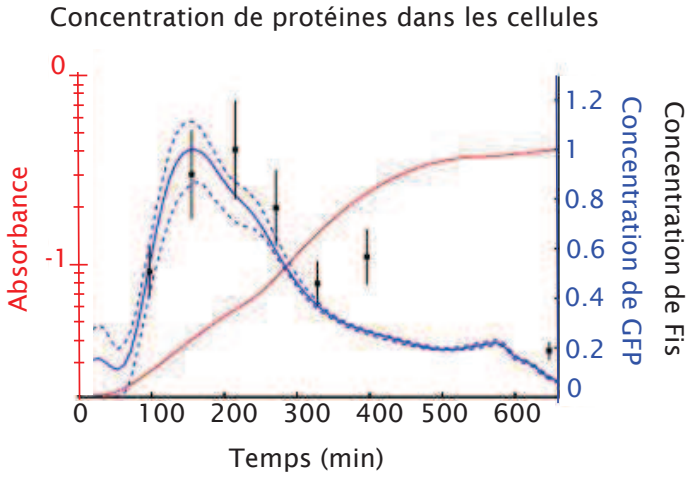


FIGURE I.2.4.: Comparaison entre la dynamique de la concentration des protéines Fis la concentration mesurée par Western blot (en noir) et la concentration des fluorophores GFP rapportant l'activité du gène *fis* (en bleu) : cette concentration est déduite des mesures de fluorescence. La concentration des rapporteurs et les valeurs retournées par le Western blot sont normalisées à l'intensité du pic en milieu de phase exponentielle. Et la courbe rouge donne l'absorbance de la population : cette grandeur reflète la croissance cellulaire. D'après H de Jong et al [39].

le bruit lors de la réponse au stress nutritionnel.

Comment évolue le bruit d'expression η^2_{total} lorsque les cellules passent d'un état stationnaire en stress nutrionnel (S_0) à un autre état stationnaire en croissance exponentielle (S_1) ? Est-ce que les mesures de bruit permettent d'extraire des informations quantitatives sur la dynamique du module d'expression génique étudié ? Toutes ces questions sont schématisées sur la figure I.2.5.

Pour répondre à ces questions, nous avons suivi l'évolution de l'expression de la protéine GyrA au niveau d'une bactérie individuelle : lorsque le milieu environnant est changé par ajout de glucose (signal excitateur), la cellule transite d'un état de stress nutritionnel (S_0) à un état de croissance exponentielle (S_1), et cette transition passe par l'activation du module GyrA. La microscopie de fluorescence, combinée à la microscopie de contraste de phase, nous permet alors de décrire la concentration des protéines rapporteuses (ayant le même promoteur que le gène *gyrA*) dans les cellules individuelles lors de cette transition d'état. A partir des données quantitatives ainsi recueillies, nous avons développé un modèle qui décrit la stochasticité en sortie du module GyrA. Ce modèle a été confronté aux résultats expérimentaux permettant ainsi d'extraire des informations sur les paramètres du module sondé.

Dans les parties qui suivent, nous présentons nos travaux de recherche avec d'abord le système expérimental utilisé pour mesurer le bruit : nous définissons le système rapporteur

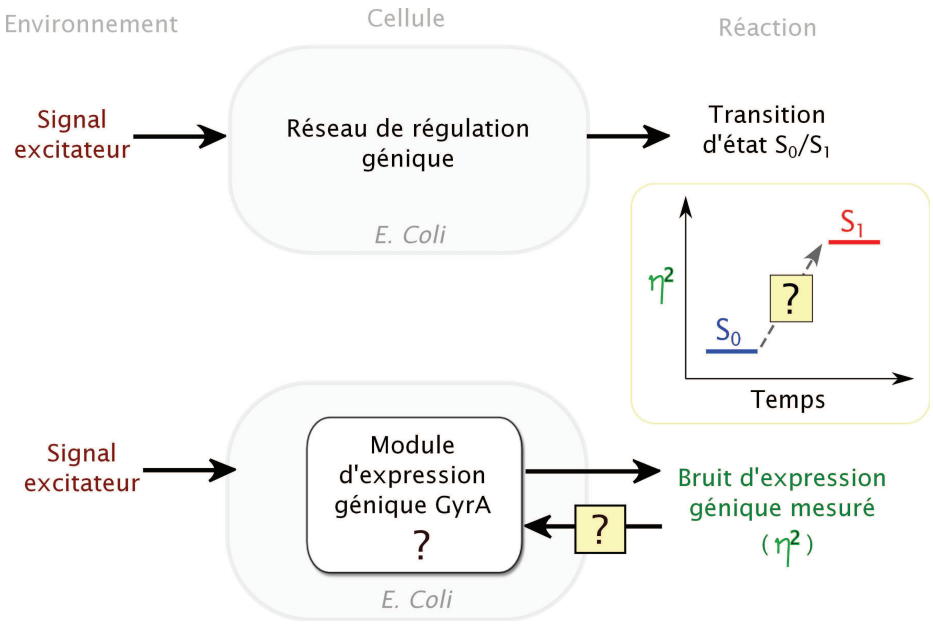


FIGURE I.2.5.: Questions posées dans la présente étude.

qui indique l'activité du module d'expression génique GyrA, et nous détaillons le protocole expérimental suivi pour initier la transition d'état. Nous caractérisons aussi le système de microscopie utilisé pour imager nos cellules individuelles.
Une fois ces conditions expérimentales bien définies, nous décrirons les méthodes utilisées pour analyser les images collectées : nous détaillerons la procédure suivie pour traiter les images et les données, et nous présenterons les outils de modélisation utilisés. A partir de là, nous pourrons présenter les résultats obtenus : nous donnerons la signature du bruit mesurée, ainsi que le modèle proposé pour décrire cette signature. Ce modèle sera confronté aux données expérimentales, et nous discuterons les informations extraites sur le module sondé.

Deuxième partie .

Dispositif expérimental

II.1. Biologie

II.1.1. Système rapporteur

Pour suivre l'expression génique *in vivo* et en temps réel, nous utilisons des rapporteurs fluorescents produits par la bactérie. Ce sont des protéines fluorescentes non toxiques (comme la GFP, Green Fluorescent Protein) : excitées à la bonne longueur d'onde, elles émettent une lumière proportionnelle à leur nombre ; et le nombre de protéines fluorescentes produites dépend de l'expression du gène d'étude.
Ainsi, le signal lumineux émis reflète l'expression génique étudiée, c'est à dire les mécanismes d'activation génique, de transcription et de traduction associés au gène d'étude. Ces mécanismes sont expliqués dans la section I.1.1 (voir figure I.1.1).

Mécanismes d'expression génique rapportés

Le nombre de fluorophores produits peut être corrélé à ces mécanismes de différentes manières : les rapporteurs traductionnels sont construits pour rapporter l'ensemble de ces mécanismes ; les rapporteurs transcriptionnels, eux, ne rapportent que les mécanismes d'activation génique, voir aussi de transcription.

Rapporteurs traductionnels : La synthèse du fluorophore FP (Fluorescent Protein) est couplée avec celle de la protéine étudiée. Le système de couplage est illustré dans la figure II.1.1 suivante : le gène rapporteur fp codant pour la protéine fluorescente est fusionné au gène x d'intérêt, et la fusion est telle qu'elle conduit à la synthèse d'une protéine hybride X-FP. Ce complexe protéique n'altère pas les propriétés fluorescentes de la FP. Ainsi, le signal de fluorescence émis reflète l'expression génique de l'ensemble {protéine d'intérêt, fluorophore}. Cet ensemble est caractéristique de la protéine étudiée, mais il peut présenter des propriétés différentes : la transcription et la traduction ne s'appliquent plus à un gène, mais

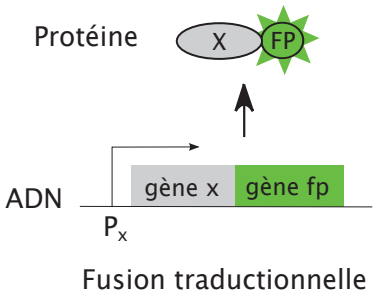


FIGURE II.1.1.: Rapporteur traductionnel.

à deux gènes fusionnés. De plus, le complexe protéique X-FP peut altérer les fonctions de la protéine X étudiée, perturbant ainsi toute la réponse cellulaire.

Rapporteurs transcriptionnels : Ils ne mesurent qu'une partie de l'expression du gène d'étude, mais la protéine d'intérêt conserve ses fonctions (elle n'est pas modifiée par une fusion protéique). On distingue deux types de rapporteurs transcriptionnels :

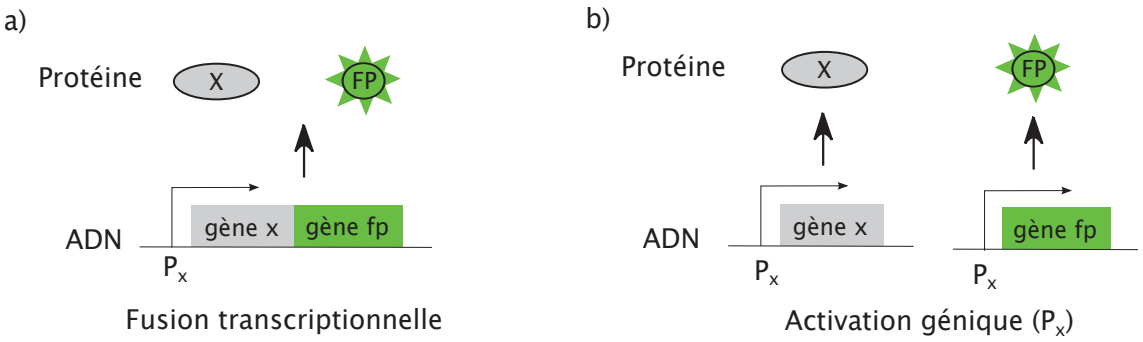


FIGURE II.1.2.: a) Fusion transcriptionnelle. b) Système rapporteur de l'activation génique. Ces deux systèmes sont expliqués dans ce qui suit.

- La fusion transcriptionnelle (figure II.1.2, a) mesure non seulement l'activation du gène étudié, mais aussi la transcription du système {gène d'étude, gène rapporteur}. Le gène fp codant pour le fluorophore FP est fusionné avec le gène x d'intérêt, et les deux protéines générées sont distinctes : les mécanismes précédents la traduction sont couplés, mais cette dernière étape se fait de manière indépendante.
- L'autre système rapporteur (figure II.1.2, b) mesure uniquement l'activation du gène étudié, c'est à dire l'activité du promoteur en terme de liaison avec l'ARN polymérase. Pour ce faire, on impose au gène rapporteur fp un mécanisme d'activation identique à celui du gène d'intérêt x : le gène codant pour la protéine fluorescente FP, est placé en aval du promoteur P_x du gène d'intérêt. Ainsi, le nombre de FP synthétisées reflète l'activité du promoteur d'intérêt.

Rapporteurs utilisés : Dans cette étude, nous travaillons avec des rapporteurs transcriptionnels qui indiquent le niveau d'activation du gène d'intérêt (figure II.1.2, b). Le système rapporteur est inséré dans la bactérie, sans altérer l'ADN du gène d'intérêt. Et lorsqu'on excite le fluorophore FP par une lumière de longueur d'onde adéquate ($\lambda_{excitation}$), il relaxe vers son état d'équilibre en émettant un rayonnement de fluorescence ($\lambda_{émission}$). Ainsi, le niveau de fluorescence de la cellule traduit l'activité du gène x, comme illustré sur la figure II.1.3.

Ces rapporteurs transcriptionnels ont été construits par nos collaborateurs biologistes du LAPM sur les souches sauvages *E. Coli* K-12 BW25113. Ce sont les souches utilisées par

Baba et al pour construire la collection Keio [42] : une librairie de mutants *E.Coli* (délétion d'un gène) largement distribuée dans le monde. Cette collection est bien adaptée pour les travaux de génie génétique réalisés par nos collaborateurs.

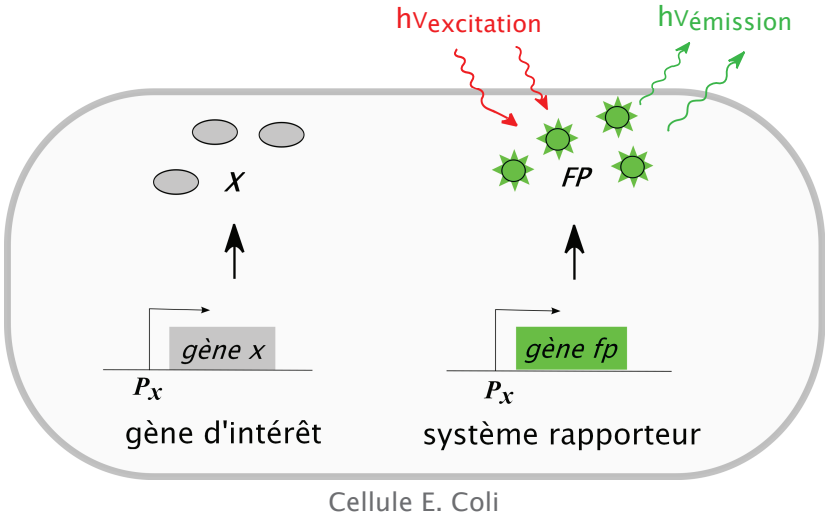


FIGURE II.1.3.: Rapporteur transcriptionnel révélant l'activité du gène d'intérêt x.

Protéines rapporteuses

Le système présenté précédemment peut être inséré dans la cellule de deux manières différentes, comme expliqué plus loin dans le paragraphe "Mode d'insertion des rapporteurs". Ces deux systèmes d'insertion utilisent tous deux des protéines fluorescentes pour rapporter l'activité du gène d'étude :

- Les rapporteurs plasmidiques synthétisent la GFPmut3*-ASV (Green Fluorescent Protein). Dans les prochaines sections, nous nous référerons à cette protéine sous le nom de GFP.
- Les rapporteurs chromosomaux synthétisent la E-CFP (Cyan Fluorescent Protein) et la E-YFP (Yellow Fluorescent Protein). Dans les prochaines parties, nous nous réfererons à ces protéines sous le nom de CFP et YFP.

Les fluorophores ainsi synthétisés sont soumis à l'éclairage de la lampe fluorescente : ils absorbent les rayons de longueur d'onde adéquate, et passent dans un état excité. Ils relaxent alors vers leur état fondamental en émettant un rayonnement de fluorescence (voir figure II.2.2 du chapitre II.2). Ce processus est quasi instantané.

Protéine fluorescente GFPmut3*-ASV

Propriétés spectrales : Une fois sous sa forme active, le fluorophore absorbe les longueurs d'onde bleue (maximum d'absorption pour $\lambda_{excitation} \approx 500$nm) et réémet un rayon-

nement de fluorescence dans le vert (maximum d'émission pour $\lambda_{émission} \approx 510nm$) [43]. Les spectres d'excitation et d'émission peuvent être visualisés dans la figure II.2.3.

Caractéristiques biologiques : La GFPmut3*-ASV est un variant de la GFP propre à la méduse *Aequorae victoria* : les mutations *mut3* la rendent plus lumineuse (signal $\approx$ 20 fois plus intense à 488nm) et lui confèrent un temps de maturation [1] plus court (de l'ordre de 7 minutes pour le variant contre $\approx$ 1 heure pour la GFP sauvage) [43, 44]. L'extension du gène par la séquence ASV permet, quant à elle, de réduire le temps de demi vie de la protéine ($\tau_{1/2} \approx 60$ min, alors qu'il est de l'ordre de la journée pour la protéine à l'état sauvage) [45]. Ainsi, l'échelle de temps du rapporteur est courte devant le phénomène cellulaire étudié (la transition stress nutritionnel/croissance exponentielle dure $\approx$ 4 heures) et on peut faire des mesures en dynamique. Les temps caractéristiques de la protéine sont donnés avec plus de précision dans le tableau II.1.1 suivant, et les méthodes de mesures sont détaillées à la suite.

Grandeur	Notation	Valeur (min^{-1})
Constante de repliement de la protéine	K_r	0.132
Taux de dégradation actif de la protéine	γ_p	$0,012 \pm 0.001$
Taux de dégradation de l'ARNm associé	γ_m	0.30 ± 0.13

TABLE II.1.1.: Temps caractéristiques de la protéine GFPmut3*-asv. La constante de repliement protéique K_r est l'inverse du temps de maturation (τ_r) : $K_r = \ln(2)/\tau_r$.

La constante de repliement protéique est tirée de mesures faites à 37°C par R.Iizuka et al [44] sur la GFPmut3 : les protéines fluorescentes ont été synthétisées *in vitro* en anaérobie pour bloquer la réaction d'oxidation nécessaire à la maturation du fluorophore. Les fluorophores ainsi synthétisés sont inactifs. Ils ont ensuite été dilués dans un milieu de maturation riche en oxygène : les protéines fluorescentes peuvent alors se replier pour adopter leur forme fonctionelle. Et le signal de fluorescence F émis augmente à mesure que les protéines arrivent à maturation :

$$F(t) = F_\infty(1 - e^{-K_r.t})$$

Les données collectées sur un spectrofluoromètre permettent alors de tracer $F(t)$, et un ajustement numérique de la courbe donne K_r. Cette constante est caractéristique du processus d'oxidation qui est le facteur limitant dans la cinétique de maturation [44, 48] : c'est donc la constante de repliement protéique.

1. Temps requis pour que la protéine se replie dans sa structure tridimensionnelle et adopte sa forme active (protéine fonctionnelle).

Le taux de dégradation de l'ARNm (γ_m) a été mesuré par l'équipe de biologistes du LAPM sur nos cellules *E. Coli* en pleine croissance exponentielle [39]. Il a été déterminé à partir de plasmides dans lesquels le gène de la GFPmut3*-ASV est activé plus fortement que dans nos plasmides. Ainsi, la production d'ARNm n'est limitée que par la transcription (et non plus aussi par l'activation génique) : beaucoup de messagers sont produits. Quand la quantité d'ARNm synthétisée est suffisante pour être détectée, la transcription est stoppée par ajout de Rifampicine au milieu de culture (150g/mL) : les ARN messagers (m) ne sont plus synthétisés, ils sont seulement dégradés. Dans ce cas, la quantité d'ARNm varie selon :

$$\frac{d\,m}{dt} = -\gamma_m \times m$$

Cette quantité a été mesurée par Nothern blot à intervalles de temps réguliers (toutes les minutes) : le nombre d'ARNm décroît exponentiellement, et un ajustement numérique de la courbe a permis d'extraire γ_m.

Le taux de dégradation protéique a également été mesuré par les biologistes sur nos cellules *E. Coli* en pleine croissance exponentielle [39]. Il a été mesuré sur les mêmes plasmides que nous utilisons : en croissance exponentielle, les cellules contiennent un nombre de fluorophores suffisant pour être détectés. Et lorsque la traduction est bloquée par du chloramphenicol (100µg/mL), les fluorophores FP ne sont plus synthétisés, ils sont seulement dégradés :

$$\frac{d\,FP}{dt} = -\gamma_p \times FP$$

Le signal de fluorescence émis par la population de cellules diminue donc exponentiellement : un ajustement numérique de la courbe de fluorescence obtenue par photométrie a permis d'extraire γ_p. Le taux de dégradation ainsi mesuré reflète les mécanismes de dégradation cellulaires, mais aussi le photo-blanchiment. Ce dernier est négligeable dans les expériences de population menées sur plaques multi puits par les biologistes, mais il n'est pas toujours négligeable dans nos expériences sur cellules uniques. En effet, pour pouvoir détecter la fluorescence émise par une cellule (et non une population de cellules), on utilise un rayonnement d'excitation plus intense. Pour avoir une valeur de γ_p plus caractéristique de nos expériences, nous avons donc mesuré le photoblanchiment dans nos conditions expérimentales : en début d'expérience, nos cellules stressées sont à l'état stationnaire, les fluorophores sont alors produits et dégradés en quantités égales. Dans ces conditions, le signal de fluorescence devrait rester constant. Toutefois, on observe pour certaines intensités d'exitation une atténuation du signal sous l'effet du photo-blanchiment (voir figure IV.2.3 du chapitre IV) : un ajustement numérique des courbes de fluorescence montre que la décroissance est exponentielle, de paramètre γ_{bleach}. Nous avons donc choisi d'utiliser les intensités d'excitation qui ne font pas apparaître le photo-blanchiment (voir section IV.2.1 du chapitre IV).

Protéines fluorescentes ECFP et EYFP

Propriétés spectrales : La ECFP absorbe la lumière UV (maximum d'absorption pour $\lambda_{excitation} \approx 434$nm) et réémet une lumière bleue (maximum d'émission pour $\lambda_{émission} \approx 477$nm) ; la EYFP, quant à elle, absorbe la lumière verte-jaune (maximum d'absorption pour $\lambda_{excitation} \approx 513$nm) et relaxe en émettant un rayonnement de fluorescence jaune (maximum d'émission pour $\lambda_{émission} \approx 527$nm) [46, 47]. Les spectres d'absorption et d'émission peuvent être visualisés dans le chapitre II.2 (voir section b, figure II.2.3). On note que les spectres d'émission des deux fluorophores sont disjoints : le signal provenant d'un fluorophore est aisément distinguable du signal émis par l'autre protéine fluorescente.

Caractéristiques biologiques : Les deux protéines fluorescentes sont des variants de la GFP produite par *Aequorae victoria* [46]. La ECFP (Enhanced CFP, Clontech) est dérivée de la GFP-mut1 (aussi appelée EGFP), un variant brillant de la GFP sauvage [43]. Des mutations sur ce variant permettent de décaler les spectres d'absorption/émission, et d'optimiser la brillance et la photostabilité de la CFP (Enhanced CFP) [46, 47]. La EYFP (Enhanced YFP-V68L/G69K) est dérivée directement de la GFP sauvage : la mutation de 3 acides aminés permet de décaler les spectres d'absorption/émission et d'augmenter la brillance du fluorophore (Enhanced YFP) [46, 47]. Le fluorochrome ainsi construit est très sensible au pH : deux mutations supplémentaires (V68L/G69K) permettent de réduire la sensibilité au pH cytosolique [49, 50].
Les temps caractéristiques des deux fluorophores sont donnés dans le tableau II.1.2 :

Grandeur	Notation	Protéine	Valeur (min^{-1})
Constante de repliement de la protéine	K_r	ECFP	0.0693
		EYFP	0.0426
Taux de dégradation de l'ARNm associé	γ_m	ECFP	0.30 ± 0.13
		EYFP	0.30 ± 0.13

TABLE II.1.2.: Temps caractéristiques des protéines ECFP et EYFP.

Les constantes de repliement sont tirées des mesures faites à 37°C par R.Iizuka et al [44] : la méthode utilisée est la même que celle vue précédemment pour la GFP. Nous remercions ici R.Iizuka pour nous avoir fourni les données non publiées concernant la ECFP.

Mode d'insertion des rapporteurs

Les rapporteurs ont été insérés dans les cellules de deux manières différentes, comme montré sur la figure II.1.4 et détaillé dans ce qui suit :

- Rapporteurs chromosomaux : le gène rapporteur est inséré directement dans l'ADN chromosomal de la cellule. Ce dernier porte l'information génétique de la bactérie et il se trouve en copie unique dans la cellule.
- Rapporteurs plasmidiques : le gène rapporteur est porté par un plasmide, morceau d'ADN circulaire qui possède sa propre origine de réplication, et qui cohabite avec l'ADN chromosomal de la bactérie. La cellule héberge plusieurs plasmides, ce qui permet de synthétiser beaucoup de fluorophores : on gagne ainsi en signal de fluorescence.

FIGURE II.1.4.: Rapporteurs chromosomaux (à gauche), ou rapporteurs plasmidiques (à droite).

Le système plasmidique est l'outil actuel de travail des biologistes du LAPM : il est plus facile à construire que le système chromosomal (les plasmides sont spécialement adaptés pour l'ingénierie génétique), il est aussi plus facile à sonder (gain de fluorescence élevé).

Par contre, c'est une sonde moins précise lorsqu'on s'intéresse aux cellules de manière individuelle. En effet, le nombre de rapporteurs plasmidiques peut varier entre deux cellules de mêmes âges à cause de la répartition binomiale lors de la division cellulaire ; le nombre de rapporteurs chromosomaux, au contraire, reste constant (les cellules filles héritent toutes deux exactement d'un ADN chromosomal). Par conséquent, la variabilité inter-cellulaire sondée par les plasmides n'est pas que le fruit de l'expression génique, c'est aussi le résultat des fluctuations du nombre de plasmides : les mesures sont bruitées par le système rapporteur.

De plus, ce système peut perturber la cellule : les plasmides utilisent l'énergie de la bactérie pour se répliquer, et cette demande d'énergie peut freiner la croissance cellulaire.

Qui plus est, ils introduisent une compétition entre la liaison des facteurs de transcription au gène chromosomal cible (x), et la liaison aux gènes rapporteurs (fp). Ceci peut avoir des conséquences non négligeables : supposons par exemple, que le facteur de transcription se lie plus facilement au gène rapporteur qu'au gène cible. Dans ce cas, la protéine cible n'interagit plus avec la protéine régulatrice. Et si elle est elle-même un facteur de transcription pour d'autres protéines (comme Fis et GyrA dans le réseau de régulation que nous étudions), la réponse de tout le réseau peut être modifiée. Cette compétition est aussi présente chez les rapporteurs chromosomaux, mais le risque de perturber le réseau

est limité : le gène cible concours avec seulement deux gènes rapporteurs, alors qu'il se mesure à une vingtaine de rapporteurs pour les plasmides (≈ 20 plasmides/cellule).

Les rapporteurs plasmidiques doivent donc être utilisés avec précaution, et les conclusions tirées des mesures doivent être vérifiées. C'est dans ce but qu'ont été construits les rapporteurs chromosomaux : obtenus en fin de thèse, ils fournissent un outil de contrôle. Les avantages et invonvénients des deux systèmes sont récapitulés dans le tableau II.1.3.

	Rapporteurs Plasmidiques	Rapporteurs Chromosomaux
Facilité de construction	Oui	Non
Gain de fluorescence	Elevé ≈ 20 rapporteurs/cellule	Faible 1 rapporteur/cellule
Information	Limitée : η_{total}	η_{total}, η_{int}, η_{ext}
Risque de perturber le réseau	Non négligeable le gène cible est en compétition avec ≈ 20 gènes rapporteurs	Faible le gène cible est en compétition avec 2 gènes rapporteurs
Risque de perturber toute la cellule	Non négligeable les plasmides pompent l'énergie de la cellule pour se répliquer	Négligeable pas de besoins supplémentaires en énergie

TABLE II.1.3.: Avantages et inconvénients des systèmes plasmidiques et chromosomaux.

Rapporteurs plasmidiques

Des plasmides contenant le gène de la *gfpmut3*-asv* ont été introduits dans la cellule. La transcription du rapporteur fluorescent y est induite par le promoteur du gène *gyrA*, ou *fis* selon le gène étudié :

- Plasmide pZE-P$_{fis}$gfp : la région promotrice de *fis* (P$_{fis}$) est clonée en amont du gène de la *gfp* sur le plasmide. Le système rapporte l'activité du gène *fis*.
- Plasmide pZE-P$_{gyrA}$gfp : la région promotrice de *gyrA* (P$_{gyrA}$) est fusionnée au gène de la *gfp* sur le plasmide. Le système rapporte l'activité du gène *gyrA*.

Le plasmide utilisé, pZE-gfp, a été construit par nos collaborateurs biologistes [39]. Il porte le système rapporteur ainsi qu'un gène amp^R de résistance à l'Ampicilline : on peut ainsi sélectionner les cellules rapporteuses sur leur tolérance à l'antibiotique . Il possède une origine de réplication ColE1, ce qui assure un nombre important de plasmides dans la cellule hôte [39] : $N_{plasmides} \approx 20$ copies/cellule. Cela permet, d'une part d'atténuer les effets de partitionnement induits par la division cellulaire ; d'autre part de gagner en intensité de fluorescence, sans pour autant noyer le bruit d'expression génique dans un

nombre trop important de molécules fluorescentes.
Tout ceci est résumé par la figure II.1.5 suivante :

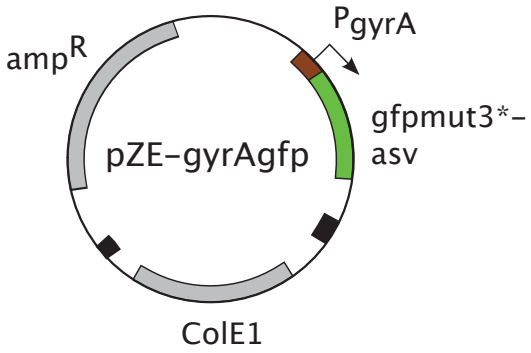


FIGURE II.1.5.: Plasmide pZE-P$_{gyrA}$gfp révélant l'activité du gène *gyrA*.

Ce plasmide ne perturbe pas le fonctionnement cellulaire : la croissance bactérienne n'est pas affectée, en terme de taux de croissance [39].

Rapporteurs chromosomaux

Ce nouveau système fournit un outil de contrôle pour nos rapporteurs plasmidiques. Qui plus est, il peut apporter un complément d'information sur la nature du bruit d'expression génique (intrinsèque/extrinsèque). Analogue à la construction de M.B.Elowitz et al [6], deux gènes rapporteurs codants pour deux protéines fluorescentes aux propriétés spectrales différentes (la YFP et la CFP) sont insérés dans le même chromosome, à deux loci différents : ils remplacent les gènes *galK* et *intS* (symbole <> sur la figure II.1.6). Les lieux d'insertion sont équidistants de l'origine de réplication *OriC*, et les deux gènes rapporteurs sont activés par le même promoteur du gène *gyrA* (P_{gyrA}) . On note que notre système présente l'avantage de ne pas comporter de résistance à un antibiotique, contrairement au système de M.B.Elowitz et al.
Nous laissons le mérite des constructions aux biologistes (en particulier C.Pinel, C.Ranquet, et H.Geiselmann), et ne donnons dans ce qui suit que les informations d'intérêt pour notre étude : les étapes de la construction sont simplifiées.

Cassette d'expression des gènes rapporteurs : Les deux gènes fluorophores ont été fournis par M.B.Elowitz ; la région promotrice du gène *gyrA* a été extraite des constructions plasmidiques réalisées par nos collaborateurs biologistes [39]. Ces deux fragments d'ADN ont ensuite été fusionnés pour donner les deux cassettes suivantes :

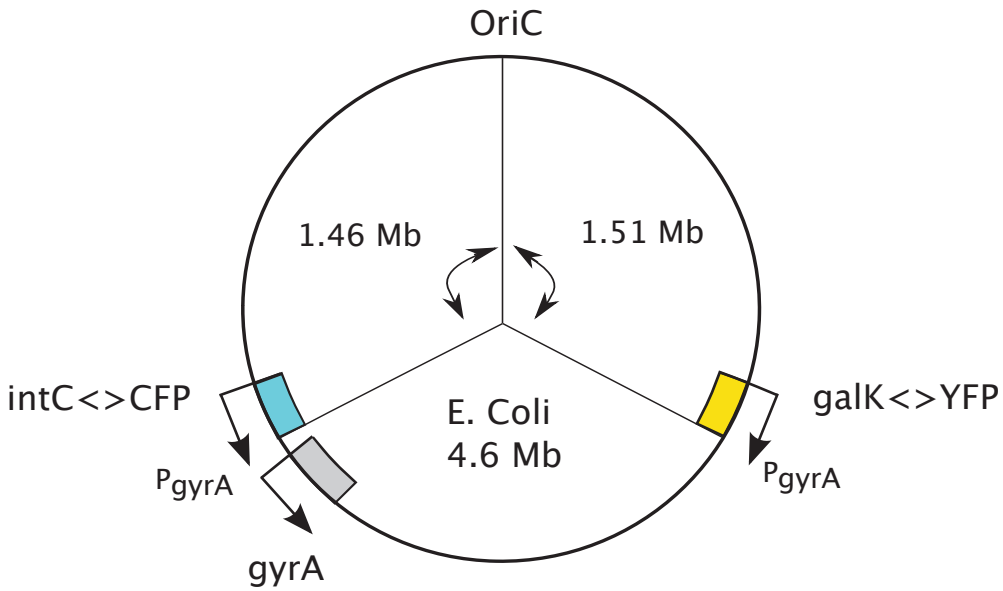


FIGURE II.1.6.: Doubles rapporteurs chromosomaux : carte du chromosome E.Coli avec l'origine de réplication (OriC), les lieux d'insertion des gènes rapporteurs *cfp* et *yfp*, et le loci du gène *gyrA*. Les deux gènes *cfp* et *yfp* sont tous deux activés par le promoteur du gène *gyrA*. Ils sont insérés directement sur le chromosome, à la place des gènes *galK* et *intS*. Le symbole <> indique le remplacement du gène *galK* par la cassette rapporteuse YFP (respectivement *intS* par CFP).

Intégration dans le chromosome Ces cassettes rapporteuses ont ensuite été intégrées dans le chromosome, à deux loci différents (cf figure II.1.6 précédente) :

- *intS* <> CFP : la cassette CFP est insérée dans la région occupée par le gène *intS* : elle remplace la séquence codante de *IntS* et elle est orientée dans le sens opposé du gène.
- *galK* <> YFPla cassette YFP est insérée dans la région occupée par le gène *galK*, remplaçant ainsi tout le gène. Elle est orientée dans le sens opposé de *galK*.

Les deux insertions ont été faites de manière indépendante, avec contre-sélection des cellules ayant bien intégré la cassette rapporteuse. Ce principe de sélection utilise les travaux de C.Ranquet [51] :

1. La cassette de contre-sélection ci-dessous est d'abord introduite à la place du gène d'accueil (*intS* ou *galK*). Sur cette cassette, le gène *Kan* procure aux bactéries une résistance à l'antibiotique Kanamicine : les cellules ayant bien intégré la cassette sont sélectionnées sur leur résistance à l'antibiotique.

Quant au gène *ccdb*, placé en aval du promoteur P_{bad}, il inhibe la croissance cellulaire

en présence d'Arabinose [2].

2. Une fois intégrée, la cassette de contre-sélection est remplacée par la cassette rapporteuse. Si le remplacement a bien été effectué, les cellules poussent de nouveau sous Arabinose. C'est ce dernier critère qui permet de sélectionner les cellules contenant la cassette rapporteuse, parmis celles où le remplacement de la cassette de contre-sélection a échoué.

Les deux constructions indépendantes, *intS* <> CFP et *galK* <> YFP ont ensuite été transduites dans le même chromosome, et vérifiées par séquençage [3].

Souches *E. Coli*

Au total, nous disposons de quatre souches différentes. Certaines ne contiennnent pas de système rapporteur, d'autres contiennent des plasmides qui rapportent soit l'activité du gène *fis*, soit l'activité du gène *gyrA* ; d'autres encore, contiennent des rapporteurs chromosomaux qui traduisent l'activité du gène *gyrA*. Par la suite, on nommera respectivement ces cellules souches : *wt*, *Fis*, *GyrA*, et *GyrA-K* :

- **Cellules *wt*** (Wild Type) : souches *E. Coli* BW25113 à l'état sauvage.
- **Cellules *Fis*** : souches *E. Coli* BW25113 avec plasmides contenant un gène de résistance à l'Ampicilline, et le gène de la GFP associé au promoteur de *fis*.
- **Cellules *GyrA*** : souches *E. Coli* BW25113 avec plasmides contenant un gène de résistance à l'Ampicilline, et le gène de la GFP associé au promoteur de *gyrA*.
- **Cellules *GyrA-K*** : souches *E. Coli* BW25113 contenant les gènes de la *cfp* et de la *yfp* associés au promoteur du gène *gyrA* sur le chromosome.

Ces cellules souches sont conservées à -80 °C dans du glycérol 20% . Cet antigel limite la formation de cristaux d'eau coupants à l'intérieur de la cellule : on évite ainsi la déterioration des cellules lors de la congélation.

2. Le sucre Arabinose induit le promoteur P_{bad} qui active le gène *ccdb*, générant ainsi la production de protéines *CcdB* : ces toxines inhibent la croissance cellulaire.

3. Le séquençage consiste à déterminer l'ordre d'enchaînement des nucléotides pour un fragment d'ADN donné.

II.1.2. Préparation des cellules

Les expériences effectuées consistent à placer les cellules dans des conditions de croissance optimale. Lorsqu'elles croissent de manière exponentielle, on les fait entrer dans une phase stationnaire en supprimant les sources de carbone. Puis, avant dégradation totale des bactéries, on réintroduit des sources carbonnées, ce qui marque la réentrée dans la phase exponentielle. C'est un peu avant cette réentrée qu'on commence à mesurer la fluorescence émise par chaque bactérie.
Ces différentes étapes sont résumées sur la figure II.1.7 suivante.

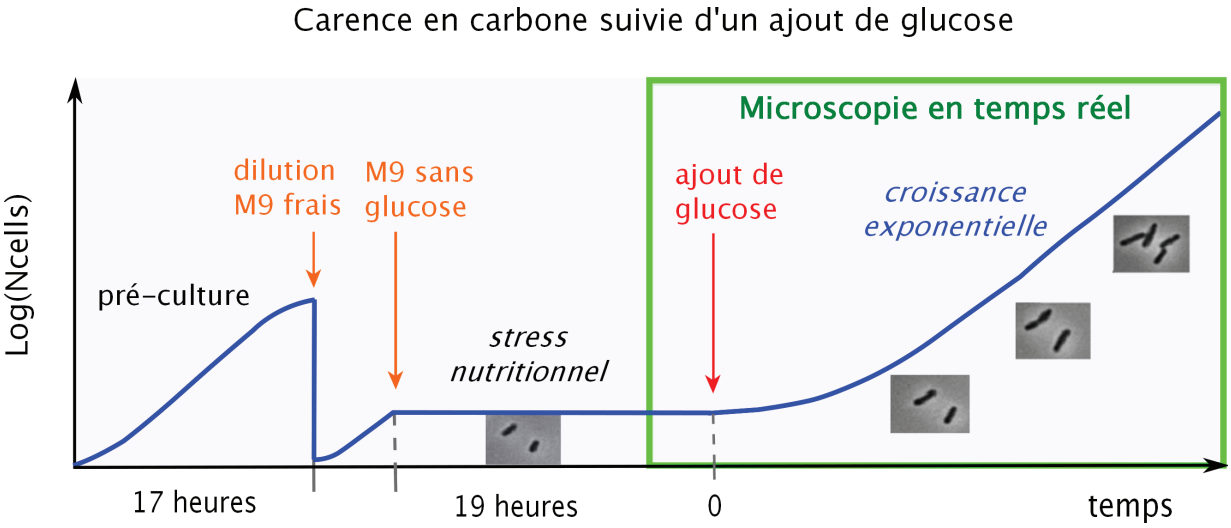


FIGURE II.1.7.: Protocole de préparation des cellules : les cellules sont d'abord cultivées dans du milieu M9 supplémenté en glucose $0,2\%$. Quand elles ont épuisé toutes les ressources du milieu, elles sont diluées dans du milieu M9 frais. On attend alors que les cellules soient en pleine phase de croissance exponentielle pour stopper leur croissance : elles sont transférées dans du milieu M9 ne contenant pas de glucose. Elles entrent alors dans une phase stationnaire de stress. Et avant qu'elles ne soient totalement dégradées, on commence à les observer au microscope : on sonde leur réponse à un ajout de glucose (qui correspond au temps 0).

Toutes ces étapes de préparation des cellules se font sous environnement stérile, exception faite de l'injection de glucose pendant l'acquisition d'images.

Préculture

Pour chaque souche, toutes les expériences sont réalisées à partir du même stock glycérol de cellules. Or ces stocks ne contiennent qu'une quantité limitée de cellules, et chaque expérience nécessite un grand nombre de cellules. De plus, les cellules stockées sont dans un état de sommeil fortement prononcé, voire mortes. Il faut donc réveiller ces bactéries et les mutliplier : c'est le rôle de la préculture.

Cette étape consiste à cultiver, pendant 17 heures, les souches originelles dans 4mL de milieu M9 (voir l'Appendice A) supplémenté en glucose (0,2%). Pour les souches contenant un rapporteur plasmidique, ce milieu est complété par 100µg/mL d'Ampicilline : cet antibiotique permet de sélectionner les cellules qui ont gardé le rapporteur. En effet, sans Ampicilline, les bactéries tendent à se débarasser de l'ADN plasmidique qui coûte en énergie. Par contre, en présence de l'antibiotique, la possession du plasmide devient un avantage puisqu'il confère à son hôte une résistance à l'Ampicilline : les cellules rapporteuses sont plus aptes à se développer que les souches sauvages, elles conservent le plasmide.

La pré-culture est réalisée dans des tubes à essai de 10mL qui sont agités à 200rpm dans un incubateur à 37°C. Les cellules se trouvent alors dans un milieu favorable à leur développement, elles croissent de manière exponentielle jusqu'à ce qu'elles aient consommé tout le glucose présent. Elle ne disposent alors plus de sources de carbone et d'énergie pour se développer.

Dilution et croissance à nouveau

Pour faire repartir la croissance cellulaire, il faut diluer nos cellules dans du milieu M9 frais supplémenté en glucose, et complété de 100µg/mL d'Ampicilline pour les cellules possédant des rapporteurs plasmidiques. Nous utilisons ici un fort taux de dilution, à savoir 500x. En effet, lors de la préculture, la composition du milieu est suceptible d'avoir évolué, avec des molécules relarguées par les bactéries ; ces molécules peuvent influencer le futur comportement des cellules [4]. En faisant le choix d'un fort taux de dilution, on s'assure donc que ces molécules sont en concentration négligeable dans le nouveau milieu de culture. Ce dernier est frais et la croissance à nouveau ne dépend pas de l'histoire du milieu de préculture : la dilution marque un nouveau départ pour les cellules.

Le milieu de dilution (20mL) est contenu dans une fiole, et il a aussi été incubé à 37°C pendant que les cellules étaient précultivées : on élimine ainsi toutes les réponses cellulaires de type choc thermique qui pourraient différencier nos expériences entre elles. Cette nouvelle culture est incubée à 37°C, sous 200rpm. Et les cellules disposent de nouveau de tous les nutriments nécessaires à leur croissance, elles réactivent leur machinerie cellulaire et recommencent à se multiplier exponentiellement comme le montre la figure II.1.8 ci-après.

4. Par exemple, les autoinducteurs sont des molécules crées par les bactéries qui sont exteriorisées dans le milieu de culture : elles peuvent ensuite être intériorisées par une autre bactérie, permettant ainsi la communication inter-cellulaire.

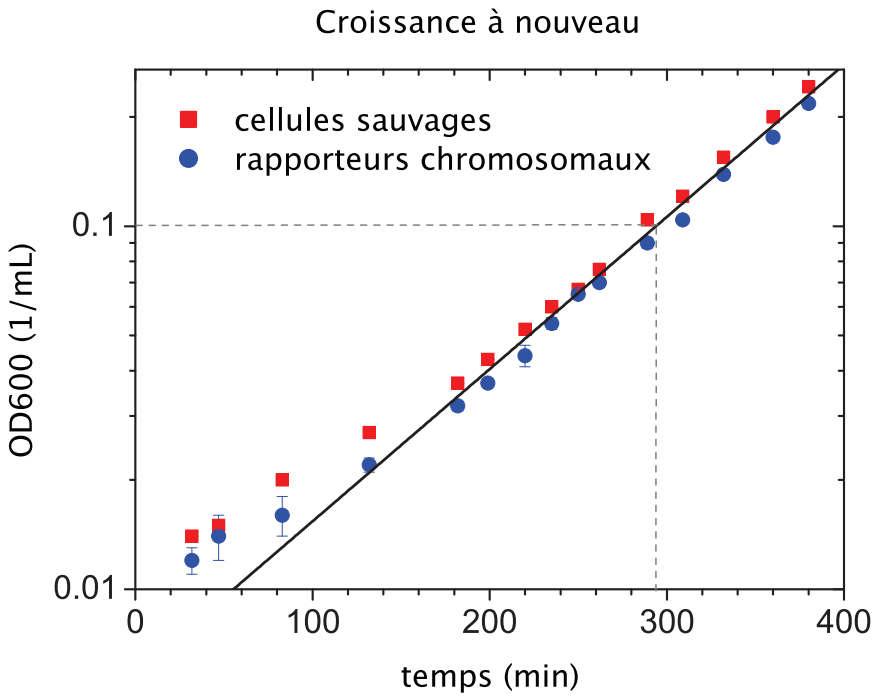


FIGURE II.1.8.: Croissance à nouveau des cellules après dilution 500x dans du milieu frais : l'absorbance reflète la concentration cellulaire de la culture (voir l'Appendice B). Elle est mesurée sur un photomètre à $\lambda = 600$nm, pour différents temps de la croissance à nouveaux. La courbe rouge représente les données d'une expérience menée sur des souches sauvages (wt) ; la courbe bleue correspond aux souches $GyrA$-K, contenant les rapporteurs chromosomaux. La droite correspond à une croissance exponnetielle ayant un temps de division $\tau_{1/2} = 70$min.

Stress nutritionnel

Lorsque les cellules sont en milieu de phase exponentielle, elles sont dans un état physiologique bien déterminé : c'est à ce moment qu'on les place en conditions de stress nutritionnel. D'après la courbe II.1.8 précédente, cela correspond à une OD_{600} de $0,1 - 0,2$.

Mise des cellules dans du M9 ne contenant pas de nutriments Le milieu de culture est changé par centrifugation-rinçage : la solution bactérienne est centrifugée (20min à 3000tr/min) et le culot des bactéries est remis en suspension dans 5mL de milieu M9 sans glucose, supplémenté en Ampicilline pour les cellules possédant des rapporteurs plasmidiques. Une deuxième centrifugation suivie d'une resuspension dans 5mL de milieu permet de s'assurer que tout le glucose résiduel est bien rincé. La solution finale est par ailleurs concentrée pour obtenir une solution de cellules à $OD \approx 0,3$. Cela nous permettra d'avoir un nombre acceptable de cellules imagées, comme nous le verrons dans la partie suivante de préparation des échantillons. Tous les milieux utilisés lors de cette étape ont subi une incubation préalable à 37°C pour s'affranchir des chocs thermiques non maîtrisés.

Phase de stress nutritionnel Nos cellules, en nombre suffisant, baignent maintenant dans un milieu défaillant en glucose. Elles sont remises dans l'incubateur 37°C-200rpm, et elles entrent dans la phase de stress nutritionnel. Elles y resteront affamées pendant 19 heures, jusqu'à qu'on réapprovisonne le milieu de culture en glucose. On s'assure ainsi qu'en début d'expérience la machinerie cellulaire a atteint le régime stationnaire (au sens physique) de la phase stationnaire (au sens biologique).

II.1.3. Préparation des échantillons pour la microscopie

On récupère les cellules en phase stationnaire pour les coller sur une boîte de Pétri de diamètre 50mm et équipée d'une lamelle $Nr\,0$ (épaisseur 100µm). La colle utilisée est un mélange de poly-L-Lysine $0,01\%$ (200µL) et de NaOH $0,1$M (200µL). Les polymères de Lysine sont chargés positivement, ce qui permet la liaison avec la surface de verre chargée négativement par les groupes SiO^-.

Après 20 minutes d'attente sur un bain à sec régulé avec précision à 37°C [5], on effectue deux rinçages de la colle avec 5mL d'H_2O stérile, suivi par un rinçage avec 1mL de M9 sans glucose. Les solutions utilisées pour le rinçage ont été préalablement incubées à 37°C pour éviter de bruiter la réponse étudiée par une réaction cellulaire aux chocs thermiques [6]. On peut alors rajouter les cellules précédemment préparées (200µL) : celles-ci se fixent sur le tapis de poly-L-lysine.

On laisse reposer cette préparation pendant 10 minutes, toujours sur le bain à sec à 37°C. Puis on rince le tout avec 5mL de M9 sans glucose (4 rinçages). Seules les bactéries qui ont bien adhéré subsistent ; on évite ainsi d'imager des bactéries partiellement collées (un exemple de cellule *comète* est pointé sur l'image de contraste de phase *III.1.1*). Lors de la visualisation sous microscopie, ces comètes ne seront pas visualisables sur toute leur longueur, ce qui présente un inconvénient pour nos mesures. On peut ensuite rajouter le milieu incubé à 37°C : 12mL de M9 sans glucose que l'on supplémente par de l'Ampicilline (100µg/mL) pour les cellules contenant des rapporteurs plasmidiques. Avant de refermer le couvercle, on perce la boîte de pétri en 3 endroits distincts pour permettre l'injection de glucose lors de l'acquisition d'images. L'échantillon est prêt, comme illustré sur la figure II.1.9 suivante, et il peut être imagé sous microscope.

5. bain à sec Stuart SBH130D
6. Le gène *gyrA* est régulé pour répondre aux chocs thermiques, en plus du stress nutritionnel.

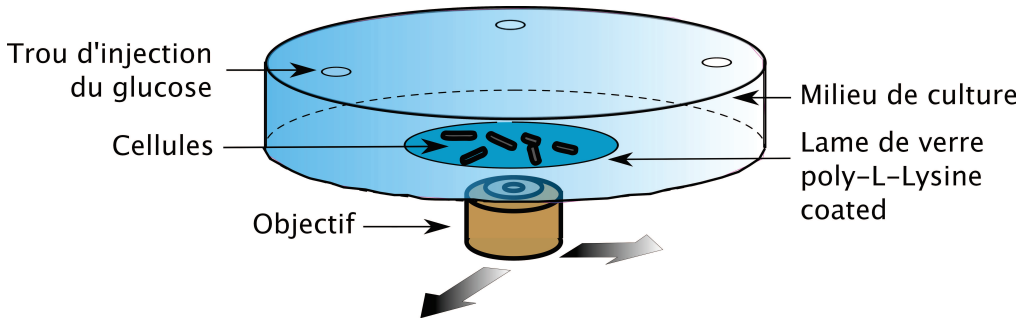
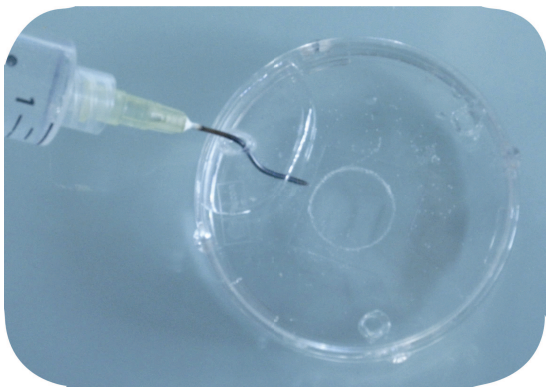


FIGURE II.1.9.: Echantillon imagé : le couvercle du pétri est troué en 3 endroits distincts pour permettre l'injection du glucose. Les cellules sont collées, via la poly-L-Lysine, sur la lamelle de verre du pétri, elles baignent dans le milieu de culture qui rempli la boîte de pétri. L'objectif du microscope vient alors se positionner sous la zone occupée par les cellules pour acquérir les images. Il est doté de mouvements de translation automatisé qui lui permettent de sonder plusieurs zones de l'échantillon.

II.1.4. Injection du glucose

Une fois l'acquisition d'image lancée, on peut rajouter le glucose au milieu de culture : 1mL de M9 supplémenté en glucose 2% est injecté dans le pétri.

Cette procédure est réalisée à l'aide d'une seringue Omnifix stérile de 10mL : une aiguille stérile de dimensions $1,1$x40mm permet d'injecter le glucose dans le pétri, via les trous percés. La pointe de l'aiguille est modelée à la flamme pour pouvoir s'insérer correctement dans le pétri, faisant fi des contraintes géométriques du microscope (boîte noire de contenance et condenseur du microscope inversé : voir la section II.2 pour plus de détails).



Cette opération manuelle est assez critique : si le glucose n'est pas injecté correctement, il atteindra la zone occupée par les cellules avec un laps de temps non négligeable. Les cellules mettent alors plus de temps à détecter la présence des nutriments après ajout de glucose. La réponse cellulaire est retardée.

Pour palier à ce problème, la procédure d'injection du glucose est suivie de 3 aspirations-réinjections de 10mL de milieu : on s'assure ainsi de bien mélanger le glucose au milieu, tout en prenant bien gare à ne pas bouger les cellules déjà imagées.

II.2. Microscopie

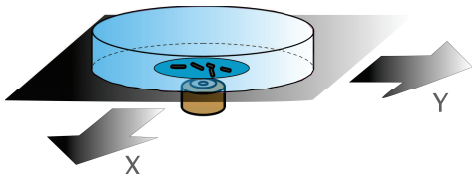
II.2.1. Le microscope et son environnement

Le système de vidéo-imagerie se compose de :

- Un microsocope inversé *Zeiss Axiovert 200M* entièrement motorisé.
- Une caméra *Roper scientific PHOTONMAX :512* couplée au système d'acquisition et d'analyse Winview.
- Une chambre thermostatée à 37°C.

Tous ces éléments sont représentés dans la figure II.2.1 ci-après et détaillés dans les paragraphes suivants.

Le microscope : Le microscope utilisé permet de travailler en contraste de phase et en fluorescence. Deux obturateurs externes *Uniblitz-VS35* permettent de commander les éclairages halogène et fluorescent avec des temps de réponse courts de l'ordre de 30 microsecondes.



Ce microscope est couplé à une platine qui permet de sélectionner les zones intéressantes de l'échantillon : l'échantillon est placé sur un porte échantillon situé sur la platine ; et la zone imagée peut être changée en translatant le plateau par rapport à l'objectif. Ces mouvements plans peuvent être transmis via un joysticks, ou contrôlés par ordinateur.

La caméra : Une caméra CCD ultra sensible permet de convertir le signal lumineux en signal électrique pour acquérir des images numériques. Pour minimiser le bruit électronique, un circuit de refroidissement à effet Peltier refroidit le capteur de la caméra à -80°C. L'élément Peltier est refroidi à son tour par un circuit externe d'eau.

La chambre thermostatée : Afin de maintenir les bactéries à 37°C, le microscope est couplé à une chambre thermostatée. Le régulateur de température utilisé est un double

module Peltier. Cette chambre thermostatée est contenue dans une boîte noire pour s'affranchir de toute lumière parasite.

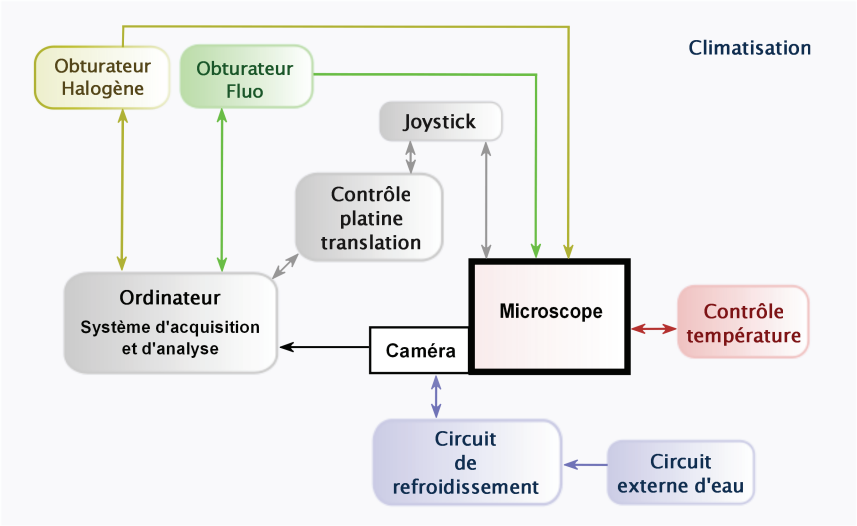


FIGURE II.2.1.: Microscope et son evironnement. En rouge, les éléments associés à la chambre thermostatée sur 37°C (boîte noire). Celle-ci contient le microscope, et elle laisse entrer la lumière des lampes halogène et fluorescente, via des obturateurs. En gris, les éléments associés au système d'acquisition et d'analyse. En bleu, les éléments associées au système de refroidissement de la caméra.

II.2.2. Acquisition des images

L'objectif utilisé pour visualiser nos cellules est un objectif 100x à immersion à huile, d'ouverture numérique 1.3 (Zeiss PlanNeofluar, Ph3 $100x/1.3$). Les images sont acquises en contraste de phase et en fluorescence. Pour acquérir chaque image, les échantillons sont exposés pendant 100ms à l'excitation de fluorescence ou à l'éclairage halogène, selon le type d'image acquise. Et la mise au point automatique se fait à partir d'une vingtaine d'images acquises en contraste de phase à différents plan focaux.

a. Source lumineuse

Pour les images acquises en contraste de phase, l'éclairage halogène est filtré par un filtre monochrome orange (Edmond Optics 540nm) : placé en amont du diaphragme, ce filtre permet de supprimer les rayons ultraviolets et bleus nocifs pour les cellules.
Pour les images acquises en fluorescence, la source d'excitation est une lampe à vapeur de Mercure (Osram, 1xHBO 103X/2). L'éclairage est réglé pour avoir un maximum d'homogénéité sur la zone imagée (on limite ainsi le bruit multiplicatif caractérisé dans la section III.1.1). L'intensité de la lampe est ajustée par des filtres atténuateurs : on éclaire suffi-

samment l'échantillon pour détecter un signal de fluorescence, sans pour autant dégrader nos cellules. La source lumineuse est ainsi atténuée d'un facteur $0,08$ pour les rapporteurs plasmidiques alors que pour les rapporteurs chromosomaux elle n'est pas atténuée.

b. Images de fluorescence

Les fluorophores soumis à l'éclairage de la lampe fluorescente sont excités. Et ils relaxent vers leur niveau d'énergie fondamental S_0 en émettant un rayonnement de fluorescence, comme illustré sur la figure II.2.2.a ci-après.

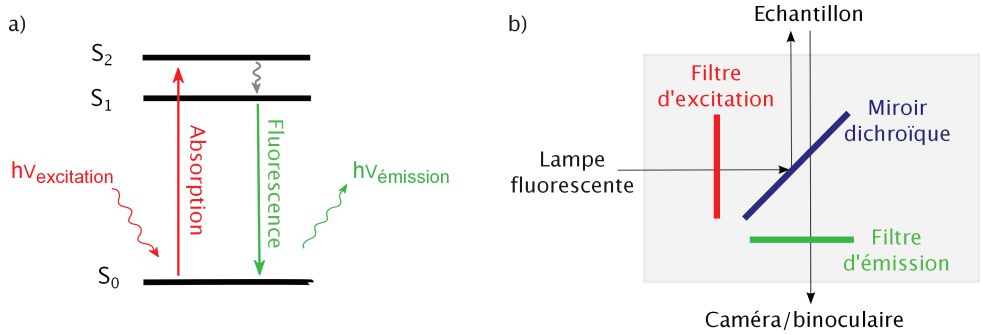


FIGURE II.2.2.: a) Principe de la fluorescence. b) Module réflecteur qui permet de sélectionner le rayonnement excitateur et le rayonnement émis. Le module est représenté dans la bulle grisé, avec ses principaux constituants : le filtre d'excitation, le miroir dicrhoïque et le filtre d'émission.

Les propriétés spectrales du rayonnement de fluorescence sont données pour les fluorophores d'étude dans la figure II.2.3 suivante.

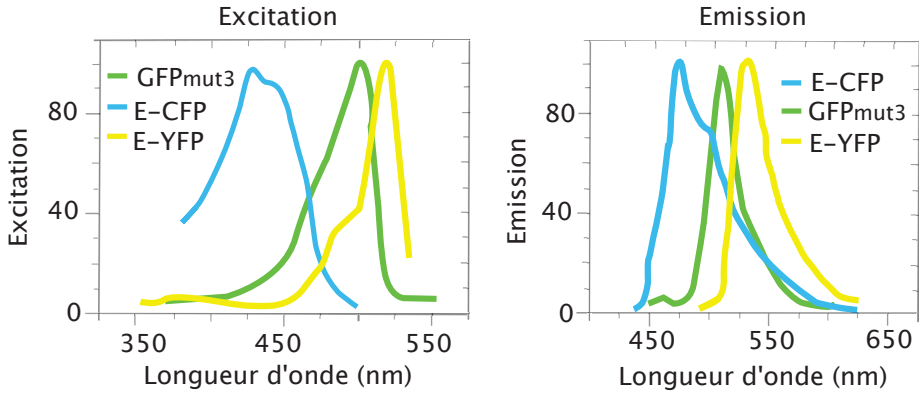


FIGURE II.2.3.: Propriétés spectrales des fluorophores E-CFP, E-YFP et GFPmut3. A gauche, les spectres d'excitation, et à droite, les spectres d'émission. Les données concernant la GFPmut3, la E-CFP et E-YFP sont tirées (resp.) des travaux de Z.Zhao et al [52] et de G.Patterson et al [47].

Pour chaque type de fluorophore sondé, l'éclairage de fluorescence est filtré pour ne récolter que les rayons d'énergie excitatrice $h\nu_{excitation}$. Et le rayonnement $h\nu_{émission}$ émis par

les rapporteurs fluorescents est sélectionné parmi toutes les longueurs d'onde présentes. Les modules réflecteurs, qui assurent ces fonctions (voir figure II.2.2.b) sont constitués d'un ensemble de trois composantes : deux filtres (appelés respectivement d'excitation et d'émission) et un miroir dicroïque. Les caractéristiques spectrales des modules utilisés sont données dans la figure II.2.4. Lorsqu'on les compare aux spectres de la figure II.2.3, on voit que les module réflecteurs *vert*, *cyan* et *jaune* sont adaptés pour imager (respectivement) la GFP, la E-CFP et la E-YFP. On note aussi que le spectre d'émission du cube *cyan* est disjoint de celui de la protéine E-YFP, et pour le cube *jaune*, le spectre ne laisse presque pas passer le signal émis par la E-CFP. Par conséquent, il n'y a pas de fuite de la E-YFP vers le canal *cyan* et les fuites de la E-CFP vers le canal *jaune* sont négligeables. Ainsi, les données de fluorescence extraites des deux rapporteurs chromosomaux ne sont pas corrélées par l'appareil de mesure.

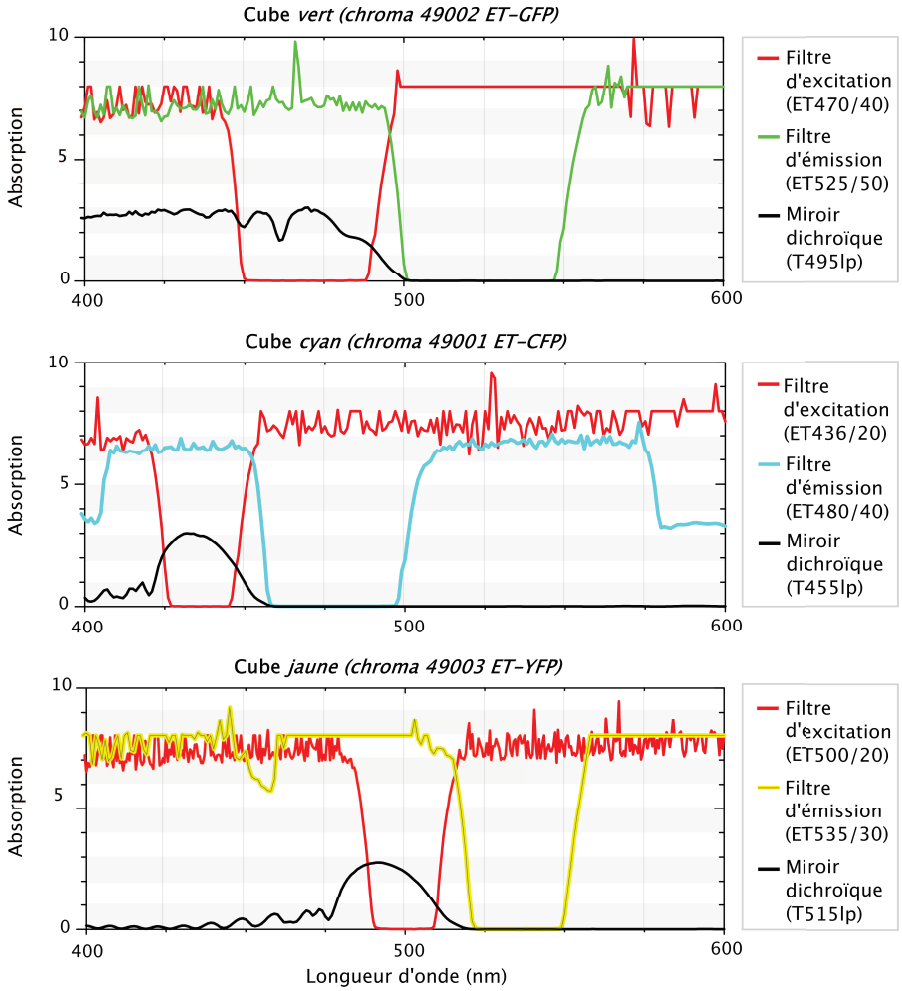


FIGURE II.2.4.: Propriétés spectrales des modules réflecteurs *vert*, *cyan* et *jaune* utilisés pour sonder les fluorophores GFP, E-CFP et E-YFP. En rouge, les filtres d'excitation qui ne laissent passer que les longueurs d'onde excitatrices pour le fluorophore sondé. En vert, cyan ou jaune, les filtres d'émission qui, eux, ne laissent passer que le rayonnement émis par le fluorophore. Et en noir, le miroir dichoïque qui sépare les rayons d'excitation de ceux émis. D'après [54].

c. Acquisition automatique

Un programme d'acquisition mis au point au laboratoire permet de contrôler tous les paramètres du microscope, de la technique d'imagerie utilisée aux caractéristiques d'illumination, et offre la possibilité de lancer des acquisitions automatiques. Les images sont ainsi acquises à intervalles de temps réguliers avec une mise au point automatique à chaque cycle d'aquisition.

Fréquence d'acquisition : Les rapporteurs plasmidiques sont imagés toutes les 5 minutes. Pour les rapporteurs chromosomaux, la fréquence d'acquisition est inférieure : les cellules sont sondées toutes les 10 minutes. En effet, les souches *gyrA-K* sont soumises à un rayonnement deux fois plus intense (la source lumineuse n'est pas atténuée alors qu'elle l'est d'un facteur $0,08$ pour les rapporteurs plasmidiques). De plus, elles sont éclairées deux fois plus longtemps : on acquiert deux images de fluorescence, une pour le rapporteur CFP et une pour le rapporteur YFP, alors qu'on en acquerrait qu'une pour les rapporteurs plasmidiques (un seul type de protéine rapporteuse, la GFP). Ces souches sont donc plus suceptibles d'être endommagées par l'excitation de fluorescence ! Pour palier à ce problème, on décide d'acquérir moins d'images lors de la réponse cellulaire.

Mise au point : La mise au point doit être préréglée manuellement en utilisant la lentille binoculaire ; un réglage plus fin peut être effectué via la caméra CCD couplée au système d'acquisition et d'analyse Winview. L'acquisition peut ensuite être lancée. Et à chaque cycle d'acquisition, la mise au point est refaite de manière automatique avec un réglage grossier qui utilise un pas de 1µm et un réglage plus fin qui utilise un pas de $0,25$µm : l'ordinateur sélectionne le plan focal qui fournit le maximum de contraste sur la zone de mise au point sélectionnée. Pour chaque cycle d'acquisition, on acquiert plusieurs champs de vue, avec pour chaque champ, une image en contraste de phase et une image en fluorescence (voire deux pour les rapporteurs chromosomaux). Et comme ces différents champs de vue ne sont pas toujours contenus dans le même plan (planéité non parfaite de la boîte de pétri ou du tapis de poly-L-lysine), la mise au point doit être réglée pour chaque champ de vue. Dans ce but, nous définissons en début d'acquisition l'écart de focale dz entre les champs imagés et le premier champ de vue. Ainsi, la mise au point se fait sur le premier champ de vue, et les autres champs sont défocalisés de dz pour être au point focus (voir figure II.2.5). Une fois que le focus est réglé sur l'image de contraste de phase, il est décalé de $0,8$µm pour les images de fluorescence. En effet, celles-ci possédent un léger décalage d'alignement : le plan focal des cellules fluorescentes est décalé de $0,8$µm par rapport au contraste de phase. Cette petite correction n'a été faite que sur les dernières expériences qui concernent les rapporteurs chromosomaux.

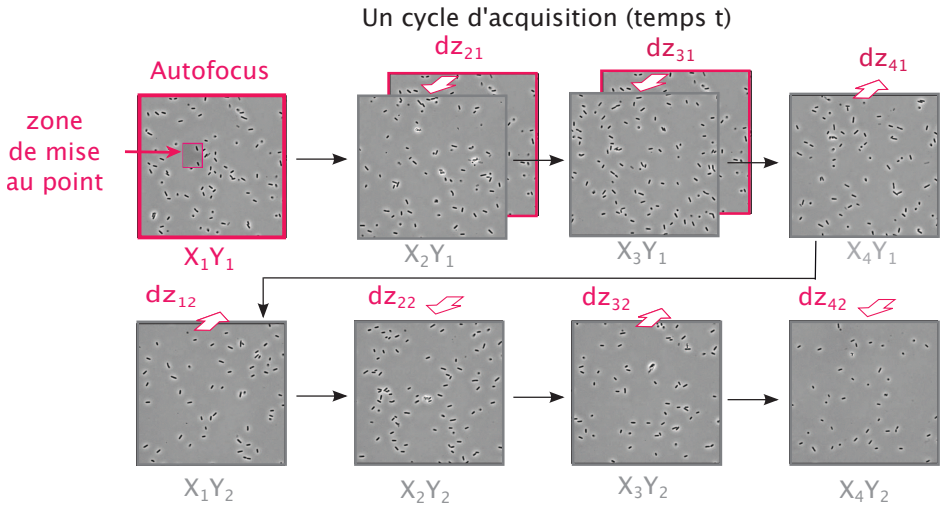


FIGURE II.2.5.: Mise au point lors d'un cycle d'acquisition d'images. Au début du cycle, la mise au point se fait automatiquement sur une zone prédéfinie du champs X_1Y_1. Puis l'acquisition d'images commence : le champ X_1Y_1 est imagé en contraste de phase et en fluorescence. La platine motorisée se déplace ensuite en X et en Y pour permettre d'imager les autres champs de vue. Pour chacun de ces nouveaux champs, la mise au point est corrigée de dz, la distance qui sépare le focus du champ considéré de celui du champ initial. Cette distance a été déterminée manuellement pour chaque champ de vue avant de lancer l'acquisition automatique.

II.2.3. Caractéristiques de la caméra

La caméra est composée d'un capteur CCD de 512×512 pixels de dimension $16 \times 16 \mu m^2$ chacun. Elle a deux modes de fonctionnement possibles :

- Un mode traditionnel où les photoélectrons détectés par le CCD sont d'abord lus, et ensuite amplifiés. Ce mode s'accompagne d'un bruit électronique faible, au détriment de la sensibilité.
- Un mode avec amplification sur puce (EMCCD) : les photoélectrons sont amplifiés directement au niveau du pixel de détection, puis il sont soumis à la chaîne traditionnelle de lecture et d'amplification finale. Ce mode a une sensibilité plus grande, mais il est accompagné d'un plus grand bruit électronique.

Compte tenu de nos conditions expérimentales (signal pas trop faible, mais nécessité de minimisation du bruit électronique), nous avons opté pour le mode de fonctionnement traditionnel.

En ce qui suit, nous donnons les caractéristiques de notre caméra déterminées pour nos conditions expérimentales : fréquence de lecture de 1MHz, temps d'acquisiton de 100ms, et refroidissement des capteurs CCD à $-80°C$. On s'intéresse principalement à deux grandeurs :

- Le bruit de lecture de la caméra, σ_{ro} : pour avoir des données exploitables, celui-ci doit être négligeable devant le bruit d'expression génique.

- Le décalage analogique de la caméra, bk : ce niveau de base s'ajoute au signal de sortie pour le décaler du zéro numérique. Et pour pouvoir traiter correctement nos images de fluorescence, ce signal additionnel doit être soustrait (voir chapitre III.1, section c).

a. Sources de bruit dans la caméra :

Dans son fonctionnement traditionnel, la caméra est principalement soumise à trois sources de bruit qui sont associées aux différents étages de la chaîne de détection (voir figure II.2.6 ci-dessous). Une description plus détaillée est donnée dans ce qui suit.

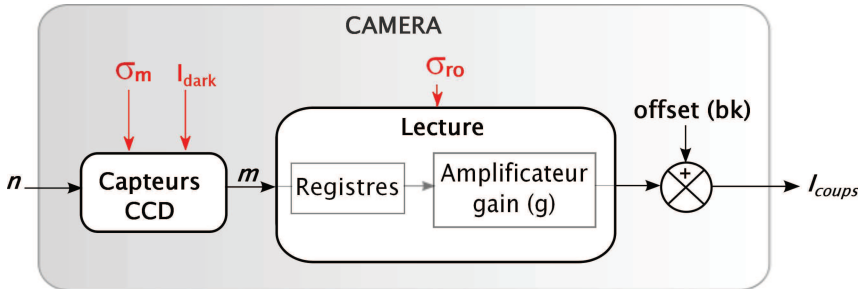


FIGURE II.2.6.: Caméra : principales étapes de conversion du signal de fluorescence en signal digital. La caméra est composée d'un capteur et d'un système de lecture. Chacun de ces étages est bruité.

On distingue ainsi :

- **Le bruit des photoélectrons σ_m** : les photons arrivant sur un pixel i (n_i) sont détectés par le photorécepteur et transformés en photoélectrons (m_i), proportionnel en moyenne au rendement quantique du photorécepteur (pour notre caméra, et dans la gamme spectrale d'intérêt, ce rendement atteint les 90%). Ce nombre m_i de photoélectrons/pixel est une variable de Poisson dont la variance est : $\sigma_m^2 = \langle m \rangle$
- **Le bruit des porteurs thermiques (courant d'obscurité I_{dark})** : les capteurs CCD (en présence ou non des photons) génèrent des porteurs thermiques qui s'accumulent dans les pixels au même titre que les photoélectrons. Ce bruit est une variable poissonienne, proportionnelle au temps d'acquisition (t) : $\sigma_{dark}^2 = I_{dark} \times t$. Le courant d'obscurité caractéristique de notre caméra CCD à -80°C est de l'ordre de :

$$I_{dark} \approx 0,008 e^-/\text{pixel/sec} \tag{II.2.1}$$

  Donc avec nos temps d'acquisition de 100ms, le bruit thermique vaut : $\sigma_{dark} \leq 0.3\sqrt{e-}$/pixel. Dans toutes nos expériences, ce bruit thermique sera négligé.
- **Le bruit de lecture σ_{ro} :** les photoélectrons collectés dans chaque pixel sont transférés dans un registre pour être lus. Ils seront ensuite amplifiés, et quantifiés

sur des unités analogique digitale (ADU). Le processus de lecture n'est pas parfait et il s'accompagne d'un bruit gaussien qui dépend de la fréquence de lecture, σ_{ro}.

Au final, la variance totale du nombre de photoélectrons au niveau d'un pixel est donnée par :

$$\sigma_{tot}^2 = \langle m \rangle + \sigma_{ro}^2 + I_{dark} \times t \approx \langle m \rangle + \sigma_{ro}^2 \tag{II.2.2}$$

Le signal électrique généré par les photoélectrons (m) est ensuite amplifié, supplémenté d'un offset (bk), et digitalisé sur 16 bits. Il est donc numérisé sur la gamme de lecture [bk, 65535] coups. Et le signal détecté par le pixel i est exprimé en coups : $(\text{coups})_i = g \times m_i + \text{bk}$, où le gain g est donné en coups/e-. Nous allons déterminer le bruit de lecture σ_{ro} de notre caméra, ainsi que son gain $G = \frac{1}{g}$ en e-/coups. Puis nous mesurerons la valeur bk du décalage analogique introduit par la caméra.

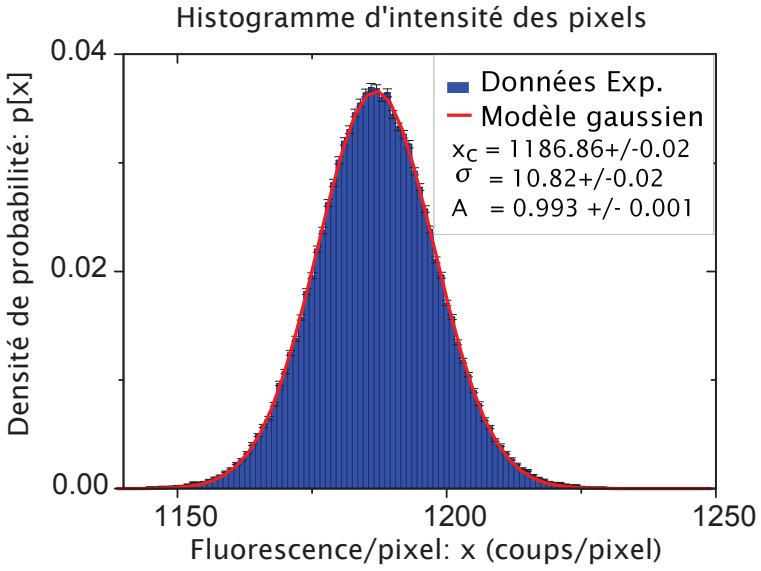


FIGURE II.2.7.: Histogramme d'une image de noir acquise pour un temps de pose court (100ms). Un échantillon homogène (milieu M9) est imagé avec une source d'exitation nulle (obturateur fluo fermé). On trace ensuite l'histogramme d'intensité des pixels (en bleu) et on ajuste numériquement l'histogramme par une gaussienne (courbe rouge) : $y = \frac{A}{\sigma\sqrt{2\pi}} e^{-\frac{(x-x_c)^2}{2\sigma^2}}$. La valeur moyenne correspond à l'offset de la caméra ; l'écart type donne le bruit de lecture de la caméra.

b. Bruit de lecture de la caméra

Le microscope étant contenu dans une boîte noire, le nombre de photoélectrons détectés (m) est quasi nul pour une exposition nulle. Dans ce cas, la relation II.2.2 s'écrit :$\sigma_{ro} \approx \sigma_{tot}$. Pour mesurer σ_{ro}, il suffit donc de mesurer le bruit total sur une image de noir (obturateur de la lampe fluorescente fermé) acquise avec un temps de pose

minimal. En ce sens, nous avons acquis une image de noir pour une exposition de 100ms. Puis nous avons tracé l'histogrammme des niveaux de gris de l'image : la statistique est très bien décrite par une gaussienne, comme le montre la figure II.2.7 ci-dessous. L'écart type de la gaussienne donne alors le bruit total de la caméra en coups/pixel ($\sigma_{tot,cps}$), qui est ici égal au bruit de lecture. D'après l'ajustement numérique de la figure II.2.7 on en déduit que le bruit de lecture en coups/pixel vaut : $\boldsymbol{\sigma_{ro} \approx 10,82 \pm 0,02}$ **coups/pixel .**

c. Gain de la caméra

Le nombre de charges collectées par un pixel photosensible i peut être exprimé en fonction du signal numérique correspondant ($I_{coups,i}$), de l'offset exprimé en coups (bk), et du gain g de la caméra :

$$m_i = \frac{1}{g} \times (I_{coups,i} - \text{bk})$$

Donc la variance dans le nombre de photoélectrons au niveau d'un pixel (σ^2_{tot}, en e-) est reliée à la variance détectée ($\sigma^2_{tot,det}$, en coups^2) par :

$$\sigma^2_{tot} = g^2 \times \sigma^2_{tot,det}$$

Des deux équations précédentes, on déduit alors que la relation II.2.2 peut se réécrire :

$$I_{coups} = G \times \sigma^2_{tot,det} + (\text{bk} - \frac{1}{G}\sigma^2_{ro}) \tag{II.2.3}$$

où I_{coups} représente ici le signal moyen détecté au niveau d'un pixel, et $G = \frac{1}{g}$ est le gain de la caméra est exprimé en e-/coups. On note que dans cette relation, le bruit de lecture σ_{ro} est exprimé en e-/pixel. Donc sur une image de fluorescence, le signal moyen détecté sur un pixel $\langle I_{coups} \rangle$, et la variance associée ($\sigma^2_{tot,det}$) sont reliés par une loi linéaire de coefficient directeur G. Par conséquent, pour déterminer G, il suffit d'acquérir des données expérimentales (I_{coups}, $\sigma^2_{tot,det}$), et de faire une régression linéaire des points expérimentaux. La pente de la droite ainsi obtenue donne G. Dans ce but, nous avons imagé plusieurs échantillons en fluorescence.

FIGURE II.2.8.: Lame/solution/lamelle.

Les échantillons sont préparés de manière très succincte : du milieu M9 est piégé entre une lame de verre épaisse ($e = 1,1$mm) et une lamelle ultrafine (épaisseur de l'ordre de 100µm). Le système *lame/milieu/lamelle* est fermé à la cire. Pour chaque échantillon imagé, le signal détecté I_{coups} est varié en changeant l'intensité d'excitation : la source lumineuse est attenuée par un filtre atténuateur de coefficient

0.05, 0.08, 0.5 ou 1. Sur chaque image acquise, on sélectionne alors une zone homogène et on détermine son niveau moyen de gris ainsi que la variance associée : $(I_{coups}, \sigma^2_{tot,det})$. On obtient ainsi la figure II.2.9.

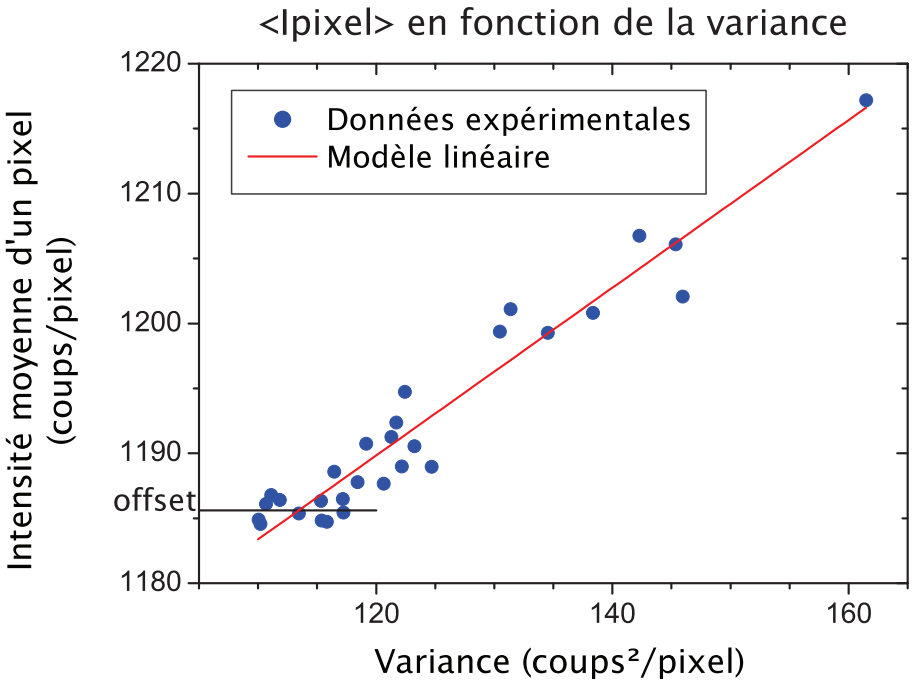


FIGURE II.2.9.: Intensité moyenne d'un pixel en fonction de la variance. En bleu, les données acquises pour 5 échantillons différents. Ces échantillons contiennent du M9 piégé entre une lampe épaisse et une lamelle ultrafine. Chaque échantillon est imagé en fluorescence (cube *vert*) avec une source d'excitation atténuée d'un facteur 0.05, 0.08, 0.5 ou 1. Les données sont ensuite collectées sur les zones homogènes des images. En rouge, la régression linéaire des données expérimentales ; la pente de la droite est de $G = 0,646$ et l'ordonnée à l'origine y_0 vaut $1112,35 \pm 6,82$.

D'après la courbe de cette figure, le gain de la caméra vaut :

$$\mathbf{G = 0,65 \pm 0,05 \text{ e-/coups}}$$

On déduit que le bruit de lecture σ_{ro} vaut :

$$\boldsymbol{\sigma_{ro} \approx 10,82 \pm 0,02 \text{ coups/pixel} = 7,03 \pm 0,55 \, e^-/\text{pixel}} \tag{II.2.4}$$

d. Décalage analogique de la caméra

Nous avons vu que pour une exposition nulle, le nombre de photoélectrons détectés est quasi nul. Dans ce cas, le signal détecté est donné par les électrons thermiques (courant d'obscurité I_{dark}) et le décalage analogique de la caméra (bk) :

$$I_{coups} = \text{bk} + I_{dark}$$

Notre caméra étant refroidie à $-80°C$, les effets thermiques sont minimisés, et le courant d'obscurité est faible devant les signaux détectés. En effet, une image de noir acquise pour un temps de pose de 100ms, fournit un signal de l'ordre de : $I_{coups} \approx 1185$ coups/pixel $\overset{G}{\longleftrightarrow}$ 773 e-/pixel . Soit, d'après la relation II.2.1 précédente, un signal détecté 10^6 fois plus grand que le courant d'obscurité. Dans ces conditions, le courant d'obscurité est négligeable devant le décalage analogique de la caméra, et on a : $I_{coups} = \text{bk}$. Pour avoir une mesure précise de ce décalage analogique, on utilise l'histogrammme des niveaux de gris tracé sur l'image de noir acquise pour un temps de pose quasi nul (100ms). Ce dernier est tracé dans la figure II.2.7 précédente. Et le centre de la gaussienne donne l'intensité moyenne des pixels, qui n'est autre que le décalage analogique de la caméra : bk= **1186.86 ± 0.02 coups.**

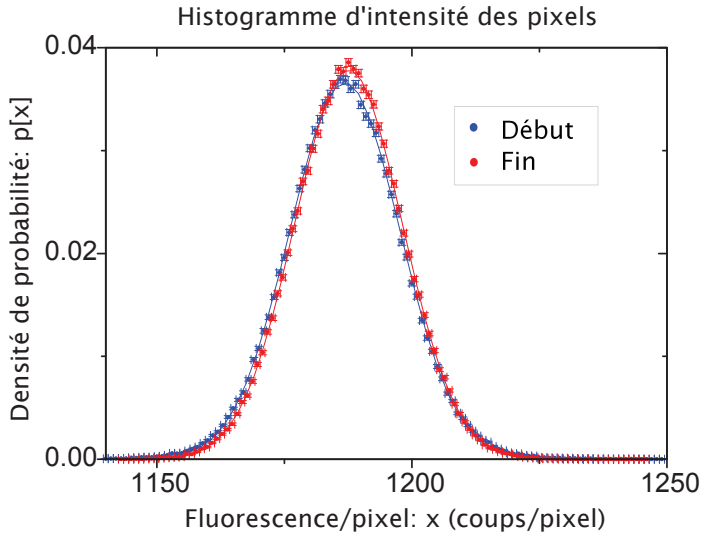


FIGURE II.2.10.: Distributions montrant les variations induites par la caméra. Une image issue d'une intégration dans le noir d'un temps de pose proche de 0 est acquise en début d'expérience (bleu), et en fin d'expérience (rouge). La distribution de l'intensité des pixels est ensuite tracée en début, puis de fin d'expérience (représentée par les points). Les deux distributions sont ensuite ajustées numériquement par une gaussienne (courbes continues rouges et bleues). Les paramètres de l'ajustement numérique sont en début d'expérience $x_c = 1186.86 \pm 0.02$, $\sigma = 10.82 \pm 0.02$ et $A = 0.993 \pm 0.002$, et en fin d'expérience $x_c = 1187.00 \pm 0.02$, $\sigma = 10.71 \pm 0.02$ et $A = 0.998 \pm 0.002$. La valeur moyenne donne le décalage analogique-numérique introduit par la caméra ; l'écart type, lui, donne le bruit de la caméra en coups/pixel.

e. Evolution des caractéristiques de la caméra pendant une expérience

Le décalage analogique et le bruit de lecture peuvent varier pendant la durée d'une expérience, et donc introduire des fluctuations instrumentales dans nos mesures. Pour

étudier ces variations, nous mesurons les caractéristiques de la caméra, avant de lancer l'acquisition d'images, et en fin d'expérience, quand toutes les images ont été acquises. Pour ce faire, nous acquérons une image de noir issue d'un temps d'intégration proche de zéro. Les caractéristiques de la caméra sont ensuite extraites comme expliqué dans le paragraphe précédent. Le résultat peut être visualisé sur la figure II.2.10 avec la distribution d'intensité des pixels en début d'expérience et en fin d'expérience. On constate que le décalage analogique introduit par la caméra reste constant pendant la durée d'une expérience, et il en va de même pour le bruit de lecture. Ce résultat est reproductible d'une expérience à l'autre.

II.2.4. Intensité d'éclairage de la lampe fluorescente

L'intensité de la lampe fluorescente peut varier d'une expérience à une autre, et au sein d'une même expérience, d'un temps d'acquisition à un autre. Ces fluctuations instrumentales créent ainsi des différences dans le signal de fluorescence détecté. Pour pouvoir comparer les données de différentes expériences, il faut donc normaliser nos mesures par l'intensité de la source lumineuse. En ce sens, nous devons calibrer la source lumineuse : pour chaque expérience, l'intensité I_{exc} de la lampe fluorescente est comparée à une intensité de référence I_0. Il suffit alors de normaliser nos mesures par $R = \frac{I_0}{I_{exc}}$ pour obtenir des données comparables. Pour accéder à l'intensité d'excitation, on exploitera la linéarite entre l'intensité d'excitation (I_{exc}) et l'intensité de fluorescence détectée (I_{det}).

a. Linéarité entre I_{det} et I_{exc} :

Soit un objet imagé en fluorescence par une source lumineuse de longueur d'onde λ_{exc} et d'intensité I_{exc}.
Le rayonnement de fluorescence détecté (I_{det}) s'obtient par les étapes de la figure II.2.11 ci-dessous : absorption du rayonnement d'excitation et émission d'un rayonnement de fluorescence, puis détection par la caméra CCD. La linéarité de chaque étape est discutée dans ce qui suit.

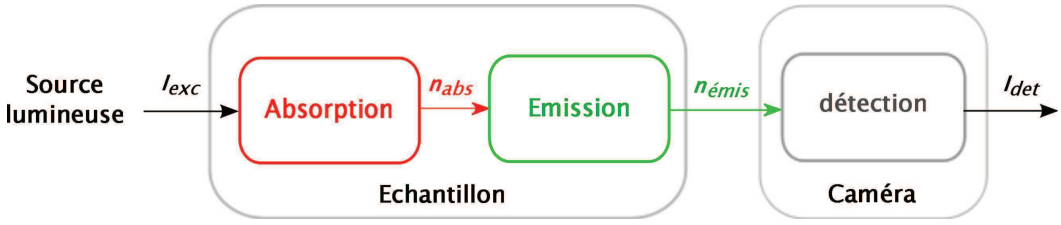


FIGURE II.2.11.: Etapes intermédiaires permettant de passer d'un signal lumineux excitateur (I_{exc}) au signal de fluorescence détecté (I_{det}).

Considérons une zone homogène de l'objet contenant des fluorophores en concentration c. Le nombre de photons absorbés par cette zone est relié à l'intensité de la source lumineuse par la loi de Beer-Lambert : $n_{abs} = I_{exc} \times \left(1 - e^{-\alpha\delta_{optique}c}\right)$, où α est le coefficient d'extinction molaire des fluorophores à la longueur d'onde utilisée, et $\delta_{optique}$ désigne le chemin optique du faisceau lumineux dans la zone traversée. Dans le cas de faibles concentrations en fluorophores, un développement limité au premier ordre permet d'approximer cette relation par une droite : $n_{abs} = I_{exc} \times (\alpha\delta_{optique}c)$. Les photons absorbés par les fluorophores sont ensuite convertis en rayonnement de fluorescence. Le nombre de photons réémis dépend du nombre de photons absorbés et du rendement quantique de fluorescence η : $n_{émis} = \eta\, n_{abs}$. Ce rayonnement de fluorescence est ensuite détecté par la caméra et converti en signal numérique avec un gain g et un décalage analogique (bk) donné par la caméra :$I_{det} = g \times n_{émis} + \mathrm{bk}$. Ces trois dernières relations permettent d'établir la linéarité entre l'intensité de fluorescence détectée et l'intensité d'excitation de la lampe fluorescente :

$$I_{det} = g \times \beta\, I_{exc} + \mathrm{bk} = a \times I_{exc} + \mathrm{bk} \tag{II.2.5}$$

où le facteur β ne dépend que de l'échantillon et des propriétés spectrales du rayonnement excitateur : $\beta = \eta(\alpha\delta_{optique})c$.

b. Objets rapportants l'intensité de la lampe fluorescente

On cherche à imager des objets homogènes pour lesquels cette relation est bien vérifiée expérimentalement. Ces objets seront utilisés pour déterminer le rapport R de normalisation dans les différentes expériences. On dispose de deux objets homogènes fluorescents :

- Le milieu de culture M9 qui est faiblement fluorescent. Il peut être mesuré directement sur l'échantillon de cellules, et permet donc de rapporter l'intensité d'excitation de fluorescence pendant toute l'expérience.
- Des billes fluorescentes (Fluoresbrite TM plan YG 1.0µm microsphères Polysciences, Inc.) contenant des fluorophores répartis de manière homogènes. Ces billes sont sondées sur des échantillons distincts de ceux contenant les cellules étudiées. Dans ces conditions, l'intensité d'excitation fluorescence ne peut être mesurée qu'en début et fin d'expérience.

On se propose d'étudier la linéarité de la relation II.2.5 pour chacun des deux objets : pour calibrer la lampe fluorescente, on utilisera l'objet pour lequel la linéarité est la plus robuste. Dans ce but, nous avons préparé des échantillons contenant les deux objets, et nous avons acquis des images de fluorescence sous différentes intensités d'excitation. Nous avons ensuite mesuré le signal détecté sur les zones homogènes occupées par les objets :

- milieu M9 : signal mesuré sur une zone homogène vide de billes, le *fond* de l'image (voir figure II.2.12).
- bille : signal mesuré sur les pixels de l'image occupés par les billes.

Puis nous avons tracé, pour chacun des deux objets, la dépendance en $I_{excitation}$ de l'intensité détectée. Une régression linéaire permet alors d'évaluer la linéarité de la relation II.2.5. Et la robustesse de cette linéarité est obtenue en sondant pour plusieurs échantillons, préparés dans des conditions différentes.

c. Préparation des échantillons

Nous avons préparé les échantillons de billes fluorescentes diluées dans du M9 de huit façons différentes. D'une part, les billes sont diluées à différents taux (dilution 1, 1/10 et 1/100) ; d'autre part, la préparation des solutions pour la microscopie varie. En effet, nous suivons 3 protocoles différents pour préparer les solutions à être imagées :

1. Le protocole 1 permet de fabriquer des échantillons de manière simple : 8µL de solution sont coincés entre une lame épaisse ($e = 1.1$mm) et une lamelle ultrafine (épaisseur de l'ordre de 100µm). Le système *lame/solution/lamelle* est fermé à la cire (voir figure II.2.8).
2. Le protocole 2 permet d'étudier l'effet des interactions avec la poly-L-lysine sur la linéarité de la relation II.2.5 (ces interactions pourraient affecter la fluorescence). Ici, la lamelle est préalablement poly-L-lysinée et les billes adhèrent à la lamelle : la technique de collage est la même que celle utilisée pour préparer nos échantillons de cellules (voir la section II.1.3). Les billes et le milieu M9 résiduel sont ensuite coincés entre une lame épaisse et une lamelle, comme dans le protocole 1. Le tout est fermé à la cire.
3. Le protocole de préparation 3 permet de se placer dans des conditions équivalentes à celles de nos expériences sur cellules : les billes sont fixées sur la lamelle collée au fond de la boîte de pétri, et elles baignent dans $\approx$ 13mL de milieu M9 (voir figure II.1.9). Ce troisième protocole est le même que celui suivi pour préparer nos échantillons de cellules (voir la section II.1.3) : dans ce cas on colle, non pas des cellules, mais des billes.

Au final, on dispose de 10 échantillons (voir tableau II.2.1). Ces échantillons ont ensuite été imagés dans les mêmes conditions : refroidissement de la caméra à -80°C et fréquence de lecture de 1MHz, temps d'acquisiton de 100ms, module réflecteur *jaune*. Pour chaque échantillon, les images de fluorescence sont acquises avec différentes intensités d'excitation : la source lumineuse est ajustée par des filtres neutres de facteur d'atténuation 1, 0.5, 0.08, 0.1 ou 0.05.

Nom	Dilution	Protocole
l1	1	1
l10	10	1
l100	100	1
pl1	1	2
pl10	10	2
pl100	100	2
pt10-1	100	3
p10-2	100	3
pt10-3	100	3
pt100	10	3

TABLE II.2.1.: Echantillons étudiés (le nom générique, le taux de dilution et le protocole de préparation suivi).

d. Etude de l'objet "milieu M9"

Sur chaque image de fluorescence ainsi acquise, on mesure le niveau moyen de gris du fond de l'image. Le fond de l'image est montré sur la figure II.2.12.

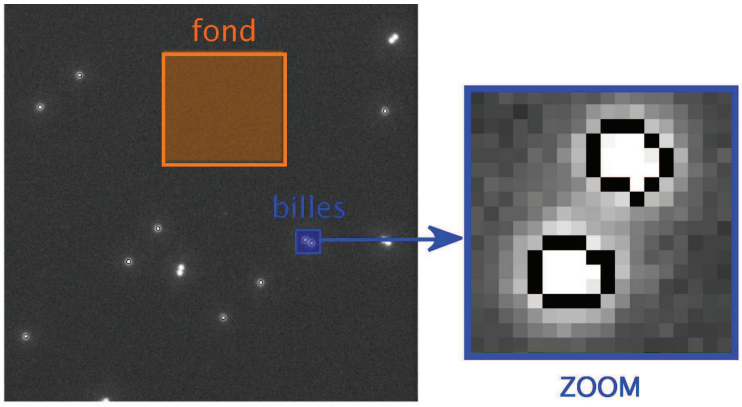


FIGURE II.2.12.: Image de fluorescence acquise pour l'échantillon l100 (100ms d'exposition avec un facteur d'atténuation 0,05 pour la lampe fluorescente, module réflecteur *jaune*). Les billes détectées par notre programme sont entourées en noir, et une partie du fond de l'image est surlignée en orange. A droite, l'image est zoomée sur deux billes détectées : celles-ci ont un diamètre $\phi \approx 6$ pixels.

Le niveau de gris I_{det} mesuré donne alors : $I_{det} = \mathrm{bk} + I_{fluo\ M9}$, où bk désigne le décalage analogique de la caméra et $I_{fluo\ M9}$ la densité de fluorescence relative à l'objet "M9". On trace alors, pour chaque échantillon, la dépendance de l'intensité détectée en fonction de l'intensité d'excitation (I_{exc}). Puis on fait une régression linéaire des points expérimentaux en imposant à l'ordonnée à l'origine la valeur de bk déterminée par la méthode de la section d précédente. On obtient alors le coefficient directeur "$a = g \times \beta$" de la relation II.2.5, comme montré pour l'échantillon l10 dans la figure II.2.13.

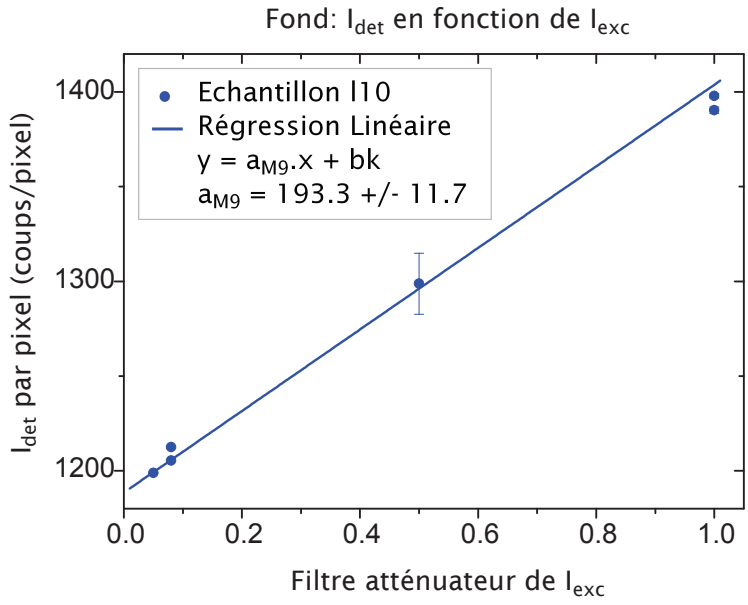


FIGURE II.2.13.: Objet milieu M9 : intensité détectée en fonction de l'intensité d'excitation. Les données (points) sont extraites de l'échantillon l10 imagé en fluorescence sous différentes excitations : la source lumineuse est atténuée d'un facteur 1, 0.5, 0.08, ou 0.05. L'intensité détectée est mesurée en prenant le niveaux moyen de gris du fond et les barres d'erreur sont données par l'écart type. Le fond est défini sur l'image de fluorescence.

e. Etude de l'objet "bille fluorescente"

Sur chaque image de fluorescence, on détecte les billes. Cette opération est réalisée par une routine Matlab que nous avons développée. Les billes détectées par la procédure sont montrées sur la figure II.2.12 (entourées en noir) pour l'image de fluorescence de l'échantillon l100. Comme on peut le voir sur le zoom de l'image, le diamètre des billes détectées vaut : $\phi \approx 6$ pixels. Et avec notre caméra et notre microscope, 1 pixel correspond à $\approx$ 0.156µm d'objet imagé. Donc le diamètre des billes détectées est de l'ordre de : $\phi \approx 0.94$µm. Ce résultat est cohérent avec les données du fournisseur, à savoir : $\phi = 1.00 \pm 0.03$µm.

On mesure ensuite la densité de fluorescence de chaque bille et on moyenne ces densités sur toutes les billes détectées. On obtient ainsi la densité moyenne de fluorescence I_{det} :

$$I_{det} = bk + I_{fluoM9} + I_{fluo\ bille} \tag{II.2.6}$$

où I_{fluoM9} désigne la densité de fluorescence relative à l'objet "M9", et $I_{fluo\ bille}$ celle relative à l'objet "billes".

Pour un même échantillon, on détermine ainsi l'intensité détectée pour des images acquises à différentes intensités d'excitation. Et comme pour l'objet "M9", on en déduit le coefficient directeur de la relation II.2.5 : la procédure est illustrée pour l'échantillon l10

dans la figure II.2.14.

Si on compare ce graphique avec la figure II.2.13 précédente qui caractérise l'objet "milieu M9", on constate que la densité de fluorescence (I_{det}) est ≈ 40 plus élevée lorsqu'elle est détectée sur les billes que lorsqu'elle est détectée sur le milieu M9. On peut donc négliger cette densité pour l'objet "M9" (I_{fluoM9}) dans la relation II.2.6 précédente. Dans ces conditions, l'intensité détectée sur les billes caractérise, non pas l'objet "bille+M9" comme indiqué dans la relation II.2.6 précédente, mais seulement l'objet "bille" :

$$I_{det} \approx bk + I_{fluo\ bille}$$

Par conséquent, le coefficient directeur a_{bille} de la figure II.2.14 ne dépend que de l'objet "bille" et des propriétés spectrales du rayonnement excitateur. Il caractérise donc la linéarité de la relation II.2.5 pour l'objet "bille".

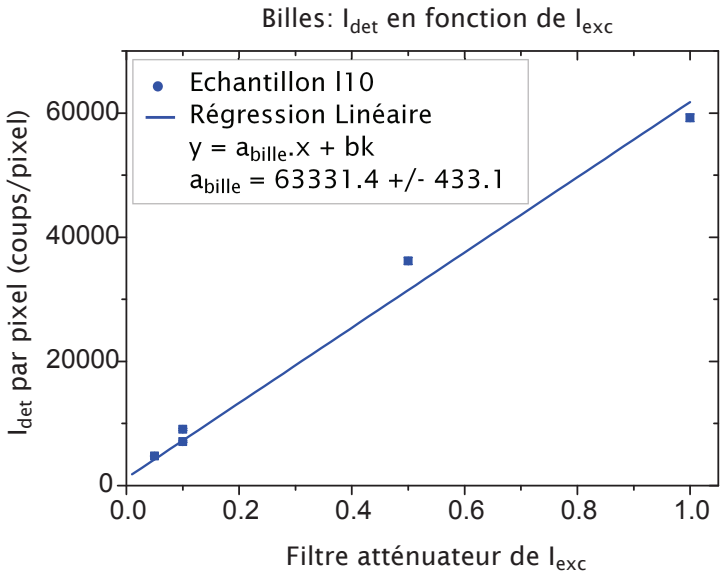


FIGURE II.2.14.: Objet bille : intensité détectée en fonction de l'intensité d'excitation. Les données (points) sont extraites de l'échantillon l10 imagé en fluorescence sous différentes excitations : la source lumineuse est atténuée d'un facteur 1, 0.5, 0.08, ou 0.05. L'intensité détectée est mesurée en prenant le niveau moyen de gris des billes ; les barres d'erreur sont données par l'écart type.

f. Robustesse de la linéarité pour les objets "M9" et "billes"

Nous avons ainsi mesuré le coefficient directeur de la relation II.2.5 pour l'objet "bille" et l'objet "M9". Les mesures ont été réalisées sur les 10 échantillons du tableau II.2.1. Pour chacun d'entre eux, ces mesures ont été normalisées par le coefficient directeur d'un des échantillons. Cela permet de comparer la robustesse de la linéarité pour les deux objets (normalisation par l'échantillon pl10 pour les billes et le milieu M9). Les résultats finaux

sont présentés sur la figure II.2.15, avec en abscisse l'échantillon considéré, et en ordonnée, le coefficient directeur normalisé de l'échantillon.

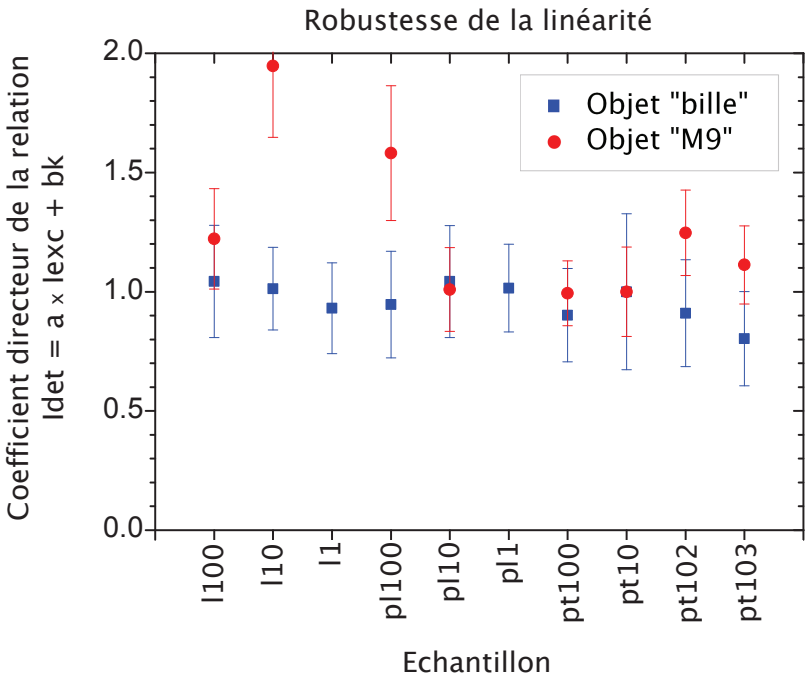


FIGURE II.2.15.: Robustesse de la linéarité pour l'objet "bille", et l'objet "M9" : les coefficients directeurs a normalisés par rapport à ceux l'échantillon pl10 (respectivement pour les billes et pour le milieu M9) sont donnés pour des échantillons préparés de 8 manières différentes (cf Tableau II.2.1). Ces coefficients directeurs ont été obtenus en imageant chaque échantillon sous différentes intensités d'excitation : pour image de fluorescence acquise, on mesure l'intensité détectée pour l'objet "bille" et l'objet "M9". La dépendance de l'intensité détectée en fonction de l'intensité d'exitation donne le coefficient directeur a de la relation II.2.5.

D'après la figure II.2.15, le coefficient directeur varie peu d'un échantillon à l'autre, et cette variation est moindre pour les billes. Ce résultat est compréhensible pour les échantillons *lame/solution/lamelle* : dans ces échantillons, la fluorescence de l'objet "M9" dépend de la quantité de milieu coincé entre la lame et la lamelle, et cette quantité n'est pas maîtrisée avec précision dans les protocoles 1 et 2. De plus, la fluorescence détectée peut aussi dépendre de la quantité de poly-L-lysine déposée : dans les échantillons préparés sur boîte de pétri ou sur lame polylysinée, cette quantité peut varier suite aux rinçages manuels. Toutefois, pour l'objet "bille", ces variations ont un effet négligeable puisque la fluorescence détectée est principalement générée par les billes (la fluorescence du milieu "M9" est négligeable devant celle des billes).

La linéarité est donc plus robuste pour les billes que pour le milieu M9. Par conséquent, nous utiliserons l'objet "bille" pour calibrer l'intensité de la lampe fluorescente.

g. Calibration de la lampe fluorescente

Pour chaque expérience, nous mesurons l'intensité détectée pour l'objet "bille", en début et en fin d'acquisition. Dans ce but, nous préparons un échantillon de billes diluées $100x$ dans du milieu M9 en suivant le protocole 1, qui est plus rapide que les autres. Cet échantillon est ensuite imagé en fluorescence sous module réflecteur *jaune*. Sur les images collectées, nous mesurons alors la densité de fluorescence de chaque bille. L'histogramme des densités de fluorescence mesurées sur les billes détectées est une gaussienne (comme montré dans la figure II.2.16) de centre donné par l'intensité détectée pour un pixel de l'objet bille. On en déduit l'intensité d'excitation (I_{exc}) au facteur de proportionnalité près (a) de l'équation II.2.5. Le facteur de proportionalité étant caractéristique de l'échantillon et des propriétés spectrales du rayonnement excitateur, il est le même pour tous les échantillons de calibration préparés (même protocole de préparation pour chaque expérience, et même module réflecteur utilisé).

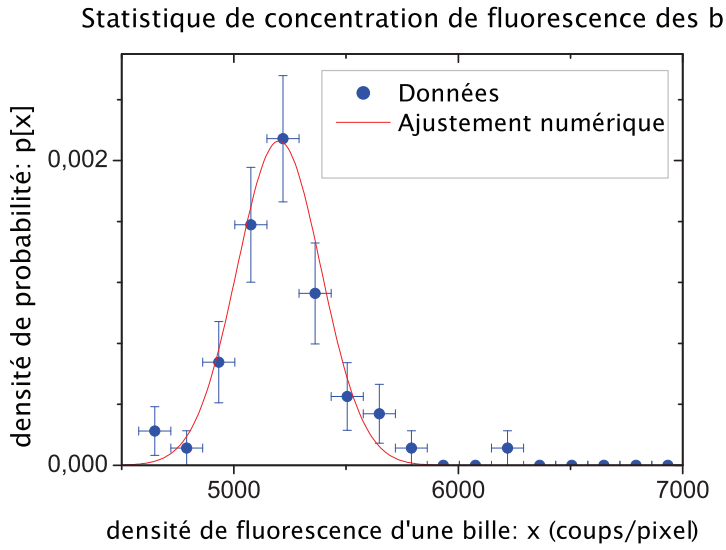


FIGURE II.2.16.: Statistique des concentrations de fluorescence pour les billes détectées sur les images de fluorescence. L'histogramme est tracé pour l'échantillon l100 imagé en utilisant le module réflecteur *jaune* sous une intensité d'excitation atténuée d'un facteur $0,05$. La ligne continue est un ajustement numérique de la distribution par une gaussienne : $y = \frac{A}{\sigma\sqrt{2\pi}} e^{-\frac{(x-x_C)^2}{2\sigma^2}}$. Le centre de la gaussienne donne la valeur de la densité moyenne de fluorescence pour l'objet "bille".

On peut donc corriger les données de fluorescence collectées sur les cellules : on calcule le rapport $R = \frac{I_{exc}}{I_0}$ où I_0 est l'intensité d'excitation référence, et on normalise les niveaux de fluorescence détectés par ce rapport. Nous verrons dans le chapitre IV.3.1 que cette calibration permet d'obtenir des données reproductibles d'une expérience à l'autre.

II.2.5. Autofluorescence cellulaire

Le signal de fluorescence détecté sur une cellule est produit par deux objets distincts:

- Les fluorophores rapporteurs "*fp*" introduits dans la cellule: une fois excités à la longueur d'onde adéquate, ils relaxent vers leur état fondamental en libérant l'énergie accumulée sous forme de photons.
- Les composants cellulaires naturels: certains sont excités à la même longueur d'onde que les fluorophores, et la relaxation radiative peut libérer la même énergie que les rapporteurs. On parle dans ce cas d'*autofluorescence*.

Pour extraire une information qui traduise le signal détecté en terme de quantité de protéines rapporteuses, il faut donc corriger le signal reçu de l'autofluorescence cellulaire. Cette correction est indispensable lorsqu'on s'intéresse à la variabilité inter-cellulaire: le bruit observé est l'effet combiné de l'aléat d'expression génique, et de la variabilité inter-cellulaire d'autofluorescence. Nos mesures brutes de stochasticité d'expression génique sont donc biaisées par le signal d'autofluorescence, et il va falloir appliquer des corrections. Ces corrections sont détaillées dans le chapitre III.3: elles supposent de connaître la loi de probabilité de l'autofluorescence.

Pour déterminer cette loi de probabilité, nous avons "*imagé*" des cellules *wt* sous module réflecteur *vert, cyan* et *jaune*. Ces cellules ont suivi le même traitement que nos cellules *Fis*, *GyrA*, ou *GyrA-K*, et elles sont imagées en phase stationnaire. A partir de là, nous avons tracé les distributions d'autofluorescence.

a. Distribution d'autofluorescence sous module réflecteur *vert*

Les résultats obtenus sont présentés sur la figure II.2.17 pour deux expériences indépendantes , toutes les deux effectuées à une plus forte intensité d'excitation . Les paramètres d'imagerie utilisés dans les deux expériences sont les mêmes. Et après normalisation de l'intensité de la lampe fluorescente dans l'expérience 3 (normalisation ici sur l'expérience 2, voir tableau II.2.2 ci-dessous), les deux distributions sont confondues. La distribution d'autofluorescence moyenne/cellule/pixel (voir section III.1.2e) est bien décrite par une gaussiène de paramètres approximatifs ici: $x_c = 6.2 \pm 0.1$coups/pixel et $\sigma = 2.2 \pm 0.1$ coups/pixel. On remarque que les distributions possèdent des valeurs négatives d'autofluorescence: ces valeurs négatives sont induites par le bruit gaussien de la caméra. En effet, la variabilité d'autofluorescence détectée est générée par deux systèmes distincts qui s'additionnent:

- les cellules, avec une autofluorescence intrinsèque (bien sur, toujours positive) qui diffère d'un individu à un autre.
- la caméra, avec une bruit de lecture dominant (voir equation II.2.4) et qui peut

prendre aussi des valeurs négatives.

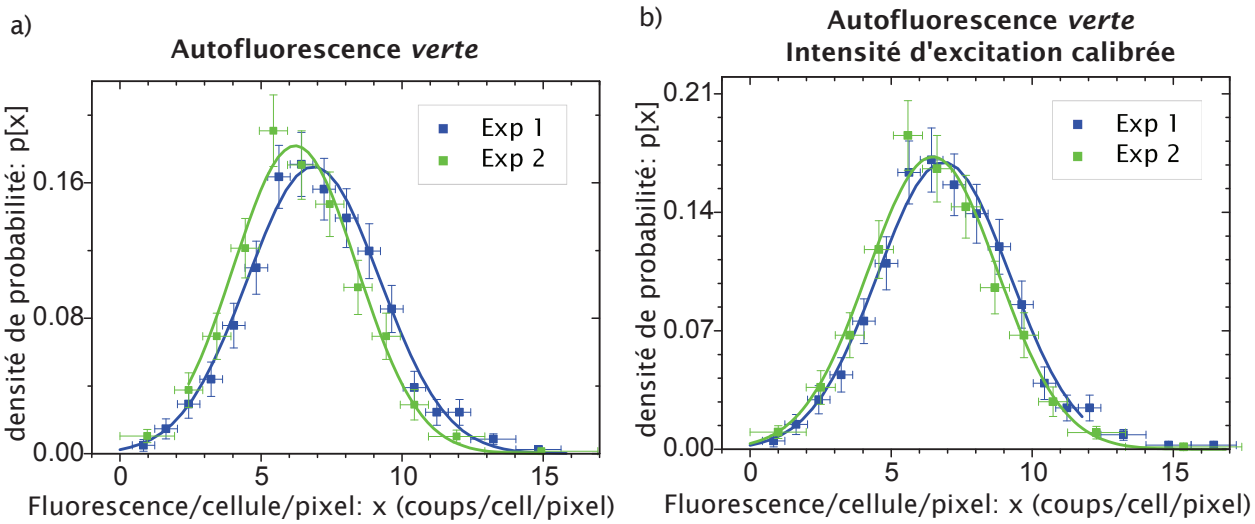


Figure II.2.17.: Distributions d'autofluorescence tracées sur des cellules sauvages *wt*. Celles-ci ont été imagées sous module reflécteur *vert*. Et les distributions présentées ont été obtenues dans deux expériences indépendantes, réalisées à 1 semaine d'intervalle. La statistique de l'expérience 1 est tracée à partir de 512 cellules *wt*, et celle de l'expérience 2 à partir de 344 cellules.
a) Les statistiques brutes (sans calibration de l'intensité d'excitation fluorescence) lissées par une gaussienne dont les paramètres sont donnés dans le tableau II.2.2. Les deux distributions sont légérement décalées à cause de l'intensité de la lampe fluorescente qui varie d'un jour à l'autre.
b) Les mêmes distributions après calibration de l'intensité de la lampe fluorescente.

Le tableau II.2.2 qui suit résume les paramètres de l'ajustement Gaussien pour 6 expériences indépendantes où l'intensité de la lampe fluorescente varie de 0.08 à 1. La méthode de calibration de l'intensité de la lampe fluorescente n'est pas appliquée.

Expérience n°	1	2	3	4	5	6
Nb de cellules	325	512	344	691	552	256
Attenuation Iex	0.08	0.5	0.5	0.5	1	1
Ajustement Gaussien						
· x_c (coups)	1.4 ± 0.1	6.9 ± 0.1	6.2 ± 0.1	7.5 ± 0.1	17.0 ± 0.1	21.1 ± 0.2
· σ (coups)	1.75 ± 0.1	2.4 ± 0.1	2.2 ± 0.1	2.2 ± 0.1	4.4 ± 0.1	5.5 ± 0.3

TABLE II.2.2.: Ajustement numérique de la distribution d'autofluorescence *verte* par une gaussienne (x_c, σ). La distribution d'autofluorescence est tracée à partir de cellules sauvages *wt* imagées sous module réflecteur *vert*. Nous présentons le résultat de 6 expériences indépendantes, réalisées dans des conditions identiques : seule l'intensité de la lampe fluorescente varie, notamment avec un filtre neutre de pouvoir d'atténuation 1 , 0.5 ou 0.08.

b. Distribution d'autofluorescence sous module réflecteur *cyan* et *jaune*

Les cellules utilisées pour tracer la distribution de l'expérience 1 du tableau II.2.2 ont aussi été imagées sous module réflecteur *jaune* et *cyan*. Et encore une fois, la distribution résultante peut être décrite par une gaussienne comme montré sur la figure II.2.18 suivante. Elles seront donc utilisées pour corriger les données de fluorescence collectées sur les cellules *GyrA-K*. L'intensité de la lampe d'excitation n'a pas été atténuée. L'ajustement numérique des statistiques montre que la loi de probabilité peut être décrite par une Gaussienne: le coefficient de détermination R^2 vaut 0.99 (resp. 0.95) pour l'autofluorescence *jaune* (resp. *cyan*).

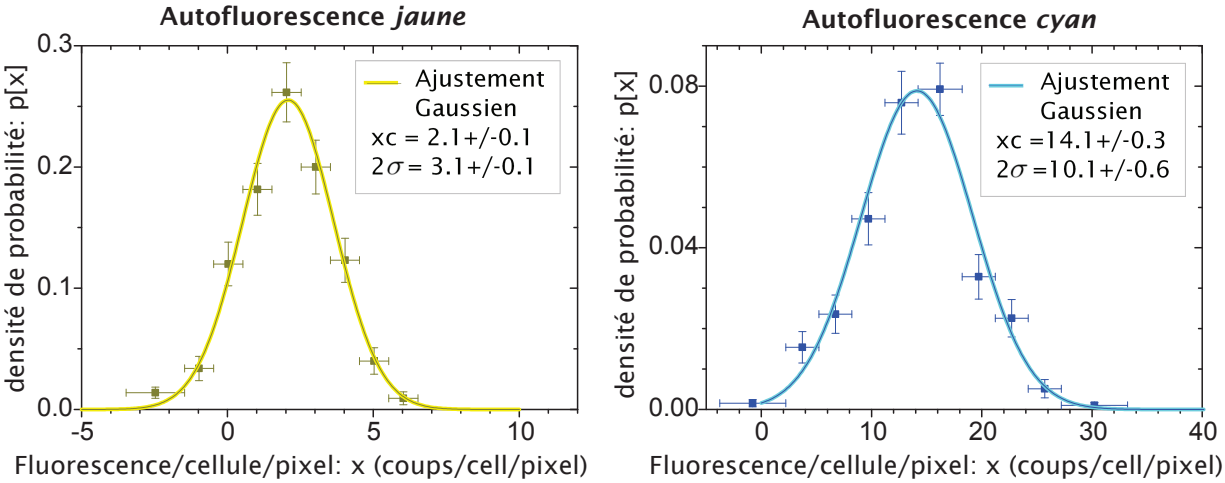


Figure II.2.18.: Distributions d'autofluorescence tracées sur 325 cellules sauvage *wt*. L'excitation de fluorescence procurée par la lampe fluorescente n'a pas été atténuée par un filtre neutre. A gauche, la distribution obtenue en YFP, et à droite la distribution obtenue lorsque les cellules sont imagées avec le module réflecteur *cyan*. Ces deux distributions ont été ajustées numériquement par une gaussienne (courbes lisses), et les paramètres de l'ajustement sont donnés dans l'encadré correspondant.

Part III.

Analyse des données

III.1. Traitement des images

Les images acquises peuvent être visualisées directement sur Winview en format binaire "spe"(propriétaire de Princeton Instruments), comme représenté sur la figure III.1.1 ci-dessous.

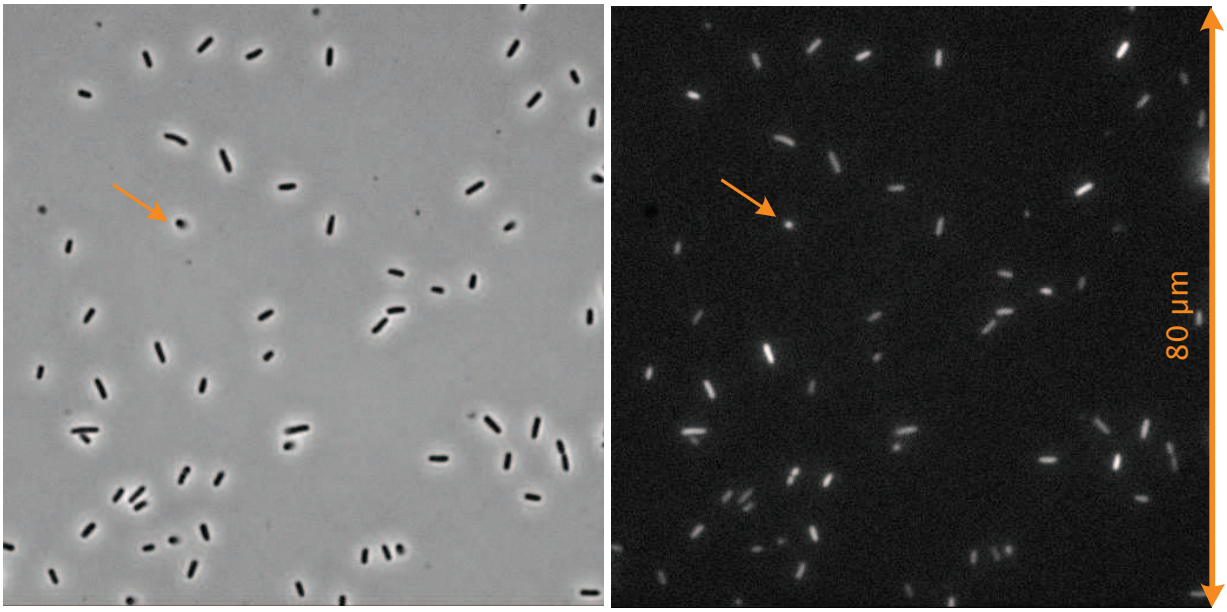


Figure III.1.1.: Cellules *E.Coli* ***GyrA*** (voir II.1.1) baignant dans du milieu M9 supplémenté en glucose 0,2% une heure plus tôt. A gauche, un champ de vue acquis en contraste de phase : les cellules apparaissent en noir sur fond gris. A droite, la même image obtenue par fluorescence : les cellules apparaissent en blanc sur fond noir. La cellule fléchée n'est pas exploitable : sa forme en comète indique qu'elle n'est pas collée parallèlement à la surface; ceci implique une absence d'information sur sa taille réelle et un signal de fluorescence diffus.

Pour en extraire les cellules avec l'information relative à leur taille et à leur niveau de fluorescence, il va falloir traiter ces images brutes. Ce traitement se fait sur Matlab à l'aide du programme « Schnitzcells » développé par M.Elowitz et ses collaborateurs, de l'université de Caltech (California Institute of Technology). En plus d'analyser chaque image obtenue, Schnitzcells permet de suivre les cellules et leur descendance dans le temps. Pour les images de fluorescence, ce traitement inclut une étape préalable de correction des artefacts d'imagerie.

III.1.1. Correction des artefacts d'imagerie

L'image de fluorescence brute doit être "nettoyée" des différents bruits qu'elle contient afin de n'avoir que le signal provenant des cellules observées. Pour extraire le signal utile de l'image brute, on détermine les signaux parasites et on applique des corrections lorsque nécessaire. Parmi les artefacts introduits par le montage expérimental, on distingue:

- *Les signaux parasites uniformes*: ils sont homogènes sur l'image de fluorescence et ils s'ajoutent au signal utile. Comme ils varient d'une image à une autre, on ne peut pas comparer les données de différentes images, sans appliquer de corrections.
- *Les signaux parasites inhomogènes*: ils varient au sein d'une même image. En conséquence, ils introduisent un biais dans la fluorescence détectée: pour s'assurer que la variabilité mesurée est bien le résultat de mécanismes cellulaires, et non pas de l'appareillage expérimental, il faut quantifier ces inhomogénéités artefactuelles, et les corriger lorsque nécessaire.

a. Signaux parasites uniformes

Ces signaux additifs sont les mêmes pour tous les pixels d'une même image. Ils proviennent de deux sources distinctes:

Signal parasite donné par l'offset analogique-digital de la caméra

L'offset analogique-digital se rajoute au signal de fluorescence émis par les cellules. Bien que ce dernier ne varie pas pendant la durée d'une expérience (cf chapitre II, section 1.0.3, figure 1.9), il peut varier d'une expérience à une autre. Pour comparer les données collectées sur différentes expériences, nous devons donc corriger l'offset sur le signal reçu.

Signal parasite induit par le milieu de culture

Le milieu dans lequel baignent nos cellules est légèrement fluorescent. Ce signal parasite se rajoute à la fluorescence cellulaire et varie dans une expérience:

- D'un temps d'acquisition à un autre: les cellules fluorescentes rayonnent, augmentant ainsi la fluorescence perçue du milieu (voir figure C.1 dans l'annexe C); ce rayonnement varie avec l'activation/l'inactivation des fluorophores et l'apparition de cellules filles surnageantes.
- D'un champ de vue à un autre: le rayonnement de fluorescence varie selon la densité de cellules du champs imagé et la présence de poussières fluorescentes.

Ainsi, pour extraire des informations de fluorescence comparables, il faut prendre en compte ces variations: pour chaque image, le signal reçu est corrigé de la fluorescence du

milieu environnant.
La fluorescence du milieu est mesurée sur l'image de fluorescence correspondante, dénuée de cellules. Cette dernière est obtenue en effaçant les cellules de l'image, comme l'explique la figure III.1.2 ci-dessous:

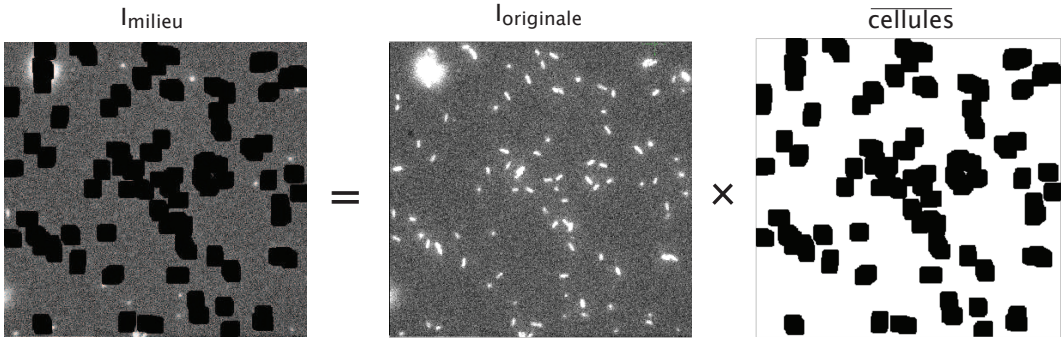


Figure III.1.2.: Image de fluorescence dénuée de cellules (I_{milieu}): celle-ci est reconstituée à partir de l'image de fluorescence brute ($I_{originale}$) et des cellules détectées sur l'image de contraste de phase. Les cellules détectées par l'algorithme de segmentation (expliqué plus loin) sont dilatées 10 fois pour s'affranchir du rayonnement de fluorescence cellulaire à proximité des cellules.

L'histogramme de l'image dénuée de cellules fournit alors le niveau de fluorescence du milieu: c'est le niveau de gris le plus fréquent (voir figure III.1.3 ci-après).
La fluorescence mesurée sur le milieu de culture est en fait une combinaison de la fluorescence générée par le milieu, et du décalage analogique-digital de la caméra:

$$F_{milieu\ mesurée} = F_{milieu} + \mathrm{bk}$$

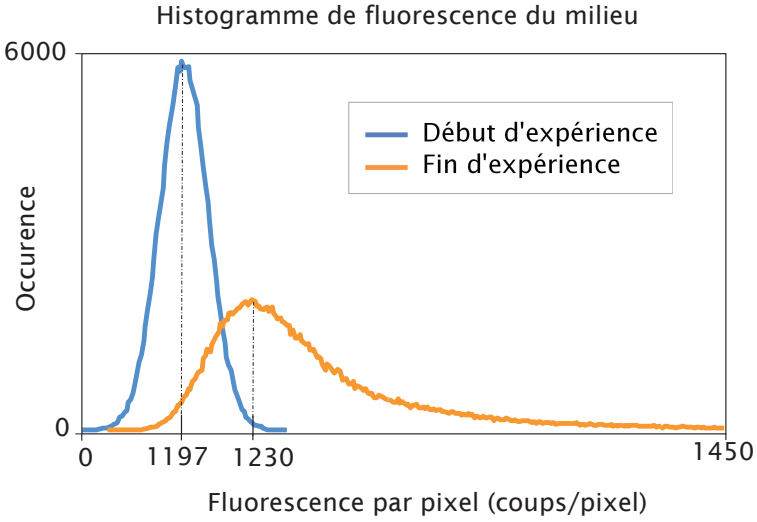


Figure III.1.3.: Histogramme de fluorescence du milieu. En bleu, histogramme en début d'expérience; En orange, histogramme en fin d'expérience.

b. Signaux parasites inhomogènes

Les inhomogénéités d'éclairage et les aberrations optiques (poussières, vignettage) se répercutent sur le signal de fluorescence avec des régions plus ou moins lumineuses sur l'image (voir figure III.1.4 plus loin). Par conséquent, la variabilité inter-cellulaire détectée n'est pas que le résultat de mécanismes cellulaires: c'est aussi le résultat des inhomogénéités introduites par l'appareillage. Il va donc falloir corriger nos images pour s'affranchir de ce bruit expérimental. Les inhomogénéités parasites peuvent être de deux natures:

- Additives: signaux parasites qui s'ajoutent au signal utile.
- Multiplicatives: inhomogénéités d'illumination et de transmission du signal de fluorescence.

Pour déterminer la nature des inhomogénéités, on étudie un échantillon fluorescent homogène. Pour chaque pixel (x, y) de l'image acquise, l'intensité de fluorescence est donnée par:

$$F(x,y) = I_{ex}(x,y)\Big(c.\eta + \eta_{parasite}(x,y)\Big) + \text{bk} \tag{III.1.1}$$

avec:

- η : rendement quantique de fluorescence pour les fluorophores étudiés. Ceux-ci sont répartis de façon uniforme sur la zone imagée (échantillon homogène).
- c: concentration en fluorophores de l'échantillon étudié.
- $\eta_{parasite}(x, y)$: rendement quantique de fluorescence pour les objets parasites présents dans le système optique.
- $I_{ex}(x, y)$: intensité d'illumination, pondérée par les amplificateurs/atténuateurs de signal présents dans le système optique.
- bk: offset analogique digital de la caméra.

La dépendance en (x, y) de $\eta_{parasite}$ traduit les inhomogénéités additives et $I_{ex}(x, y)$ represente les inhomogénéités multiplicatives (qui se propagent sur le signal d'intérêt). On peut donc déterminer la nature des inhomogénéités en faisant varier le rendement de fluorescence effectif de l'échantillon, "$c.\eta$" : les inhomogénéités additives sont inchangées, alors que celles de nature multiplicative sont accentuées lorsqu'on augmente ce rendement.
Dans ce but, nous avons préparé trois échantillons de Coumarine-500 ($1.3g.L^{-1}$) selon le protocole simplifié décrit au chapitre II.2 précédent (protocole 1 de la section c). Les molécules fluorescentes sont diluées dans l'éthanol à différentes concentrations: on contrôle ainsi le rendement de fluorescence effectif $c.\eta$.

- Echantillon 0: ethanol pur.
- Echantillon 1: Coumarine-500 diluée 10x dans l'éthanol (concentration c_1).

- Echantillon 2: Coumarine-500 diluée 100x dans l'éthanol (concentration c_2).

Les échantillons sont ensuite imagés sous module réflecteur *cyan* ou *vert*, et dans les mêmes conditions que nos cellules.

b.1. Caractérisation des inhomogénéités:

Les images acquises sont sujettes au bruit gaussien de la caméra. Par soucis de visibilité, on corrige cet artefact par un filtrage médian de voisinage 20 (voir figure III.1.4): on s'affranchit ainsi du bruit de lecture de la caméra. Les variations de gris résultantes mettent en valeur toutes les autres inhomogénéités du système optique.

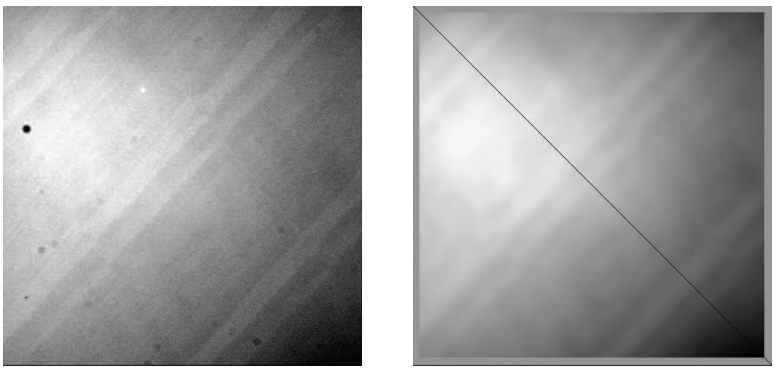


Figure III.1.4.: Echantillon 1 imagé sous cube *cyan*, avec un temps d'exposition de 100ms et une intensité d'illumination I_{ex}. A gauche, image brute. A droite, même image après application d'un filtre médian de voisinage 20. Les effets de bords sont inhibés: les points situés à ± 10 pixels des bords de l'image sont réassignés par le niveau de gris moyen de l'échantillon. La ligne noire indique la diagonale de l'image.

Sur l'image de droite de la figure III.1.4, on observe deux phénomènes: d'une part, le signal de fluorescence est accentué sur les coins gauches de l'image; d'autre part, la fluorescence présente des stries. Ce résultat est indépendant de la zone imagée. Au vu de l'image, on décide donc de caractériser les inhomogénéités par le profil de luminance selon la diagonale.

b.2. Linéarité de l'équation III.1.1

Dans un premier temps, on vérifie la linéarité de la relation III.1.1 pour chaque échantillon étudié: on acquiert des images à différentes intensités d'illumination (I_{ex}, $0,08\, I_{ex}$ et $0,05\, I_{ex}$). Si la relation III.1.1 est bien linéaire, la fluorescence détectée est proportionelle à l'intensité d'illumination. C'est bien le cas pour le niveau moyen de fluorescence détectée ($F_{détectée}$ - bk), comme le montre le tableau III.1.1.

Et pour s'assurer que cette linéarité est vérifiée pixel par pixel, on étudie la relation sur chacun des pixels de la diagonale de l'image. Pour ce faire, on trace le profil de luminance

$I_{\text{excitation}}$	$F_{\text{détectée}} - \text{offset}$
I_{ex}	I_{fluo}
$0,08\ I_{ex}$	$0,079\ I_{fluo} \pm 0.006\ I_{fluo}$
$0,05\ I_{ex}$	$0.049\ I_{fluo} \pm 0.004\ I_{fluo}$

Table III.1.1.: Echantillon 1. Facteur d'atténuation de la fluorescence détectée, selon l'intensité d'illumination. Les niveaux de fluorescence mésurés ont été moyennés sur tous les pixels de l'image correspondante, et la valeur moyenne a été soustraite de l'offset de la caméra. Les images de fluorescence sont acquises en gfp sous différentes illuminations (I_{ex}, $0,08\,I_{ex}$ et $0,05\,I_{ex}$). Elles sont ensuite corrigées du bruit de la caméra par un filtre médian de voisinage 20 (cf figure III.1.4).

de la diagonale, pour des mêmes images acquises à différentes intensités d'excitation. Et pour pouvoir aisément comparer ces profils de luminance entre eux, on les soustrait de l'offset de la caméra pour les normaliser par leur valeur moyenne. On obtient des courbes similaires, comme le montre la figure III.1.5 ci-après: la linéarité est vérifiée pixel par pixel, il n'y a pas de lumières parasites qui arrivent directement sur la caméra (l'entrée de la caméra est contenue dans la boîte noire du microscope: voir figure II.2.1). Ce résultat se retrouve pour les images acquises en module réflecteur *cyan*.

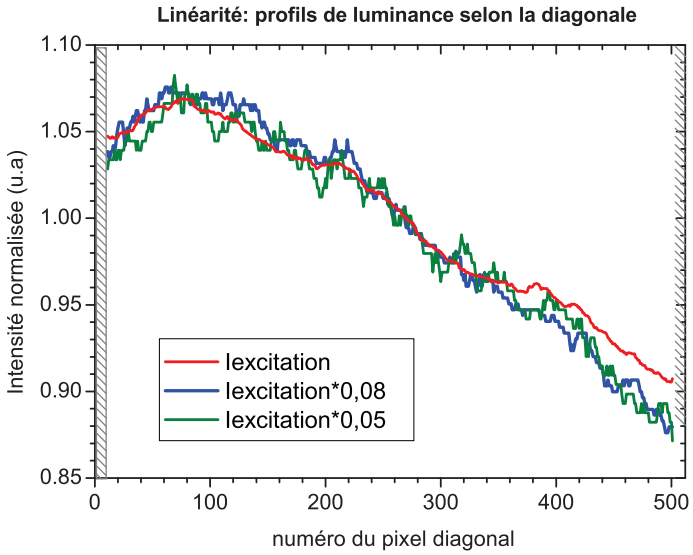


Figure III.1.5.: Linéarité de l'équation III.1.1: profils de luminance selon la diagonale pour un même échantillon de Coumarine. Les images de fluorescence sont acquises en module réflecteur *vert* avec une intensité d'illumination I_{ex} (rouge), $0,08\ I_{ex}$ (bleu) et $0,05\ I_{ex}$ (vert). Elles sont corrigées du bruit de la caméra par un filtre médian de voisinage 20: les pixels résultants affectés par les effets de bord du filtre ne sont pas représentés (quadrillage gris). Les profils de luminance ainsi obtenus sont soustraits de l'offset de la caméra. Puis ils sont normalisés par la valeur moyenne du profil éclairé sous I_{ex}, pondérée du facteur d'atténuation (1, $0,08$ ou $0,05$). Les fluctuations à forte fréquence (courbes vertes et bleues morcelées) s'expliquent par le bruit de lecture la caméra et le filtrage médian appliqué à l'image.

b.3. Nature des inhomogénéités

On détermine maintenant la nature des inhomogénéités en imagant trois échantillons de concentration différente en Coumarine (échantillons 0, 1 et 2). Les images sont acquises sous module réflecteur *cyan* car dans ce cas, les stries sont plus prononcées (cf figure III.1.4). Les profils de luminance obtenus sont soustraits du décalage analogique de la caméra. Et ils peuvent être visualisés sur la figure III.1.6 suivante.

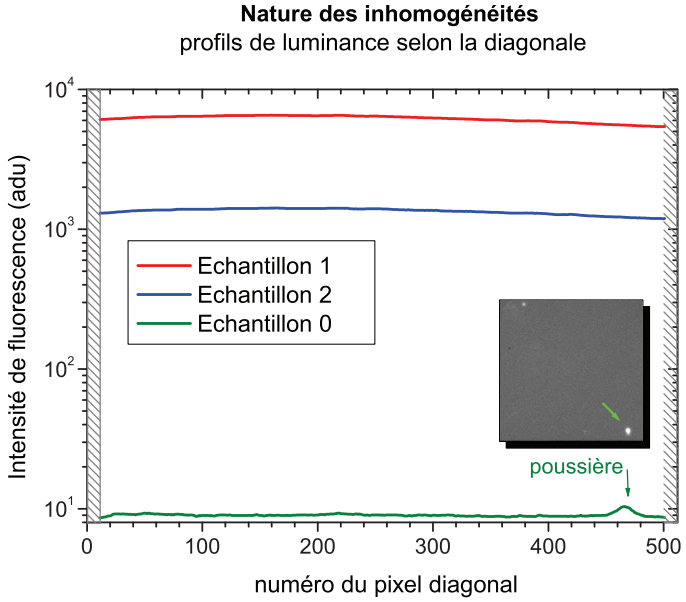


Figure III.1.6.: Profils de luminance selon la diagonale de trois échantillons distincts de Coumarine. L'échantillon 0 (vert) possède un rendement effectif $c.\eta$ négligeable (éthanol pur), l'échantillon 2 (bleu) un faible rendement (dilution 100x dans l'éthanol) et l'échantillon 1 (rouge) un fort rendement (dilution 10x dans l'éthanol). Les images acquises sous module réflecteur *cyan* sont corrigées du bruit de la caméra par un filtre médian de voisinage 20. Les pixels résultants affectés par les effets de bord ne sont pas représentés (quadrillage gris). Les profils de luminance calculés sont soustraits de l'offset de la caméra.

Les profils obtenus montrent que les échantillons 1 et 2 ont un rapport de fluorescence: $r = 4,6$. Ce rapport s'explique par les différents taux de dilution: selon la relation III.1.1, l'intensité mesurée varie d'un facteur 10 pour deux échantillons dilués 10x et 100x. Le rapport obtenu est deux fois moindre que la valeur escomptée, ce qui montre que dans la gamme de concentration de fluorophores utilisée on n'a pas une linéarité stricte du rendement de fluorescence. Toutefois en prenant un rapport relatif equivalent de concentration : $\frac{c_1}{c_2} = r^* = 4,6$, on peut écrire l'intensité de fluorescence mesurée sur chacun des deux échantillons par (cf relation III.1.1):

$$F_2(x,y) = I_{ex}(x,y)\Big(c_2.\eta + \eta_{parasite}(x,y)\Big) + \text{bk}$$
$$F_1(x,y) = I_{ex}(x,y)\Big((r^* \times c_2).\eta + \eta_{parasite}(x,y)\Big) + \text{bk}$$

On en déduit:

- les inhomogénéités multiplicatives normalisées:

$$\delta_{mult} = \frac{I_{ex}(x,y)}{\langle I_{ex} \rangle} = \frac{(F_1 - F_2)(x,y)}{\langle F_1 - F_2 \rangle} \tag{III.1.2}$$

- les inhomogénéités additives:

$$\begin{aligned} \delta_{add}\Big(\langle I_{ex} \rangle\Big) &= \eta_{parasite}(x,y).\langle I_{ex} \rangle \\ &= \frac{\big(r^* \times F_2 - F_1\big)(x,y) - (r^* - 1).\mathrm{bk}}{(r^* - 1).\delta_{mult}} \end{aligned} \tag{III.1.3}$$

Inhomogénéités multiplicatives:

Les inhomogénéités multiplicatives peuvent être visualisées sur le profil diagonal suivant:

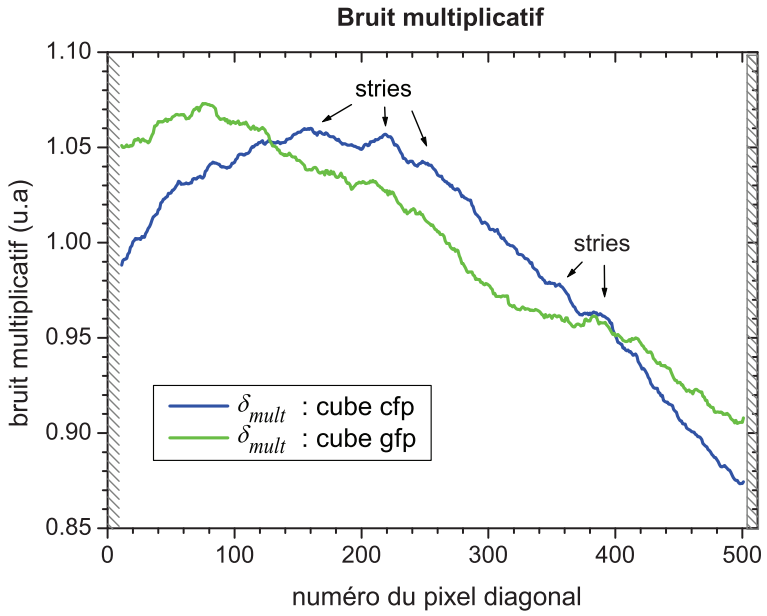


Figure III.1.7.: Inhomogénéités multiplicatives selon la diagonale de l'image. Celles-ci sont calculées à partir des échantillons 1 et 2 imagés sous module réflecteur *cyan* (courbe bleue) ou *vert* (courbe verte). Les images acquises sont préalablement corrigées du bruit de la caméra par un filtre médian de voisinage 20: les pixels résultants affectés par les effets de bord ne sont pas représentés (quadrillage gris). Les inhomogénéités "stries" très prononcées sont soulignées par des fléches.

On constate deux types d'inhomogénéités multiplicatives:

- Les stries: elles sont induites par le mouvement de l'échantillon sur l'huile d'immersion, avec une huile visqueuse qui n'est plus étalée de manière homogène sur l'objectif lors de l'acquisition d'image. Ces inhomogénéités induisent une variation de signal inférieure à $0,8\%$.

- L'éclairage de la lampe qui n'est pas uniforme: le profil est courbé. Le bruit résultant peut valoir jusqu'à 19% du signal cfp, et 17% du signal gfp.

Inhomogénéités additives:

Les inhomogénéités additives calculées sont fonction de l'intensité d'illumination I_{ex}. Nous les donnons pour l'intensité d'illumination utilisée dans nos images de cellules acquises sous module *cyan* (figure III.1.8).

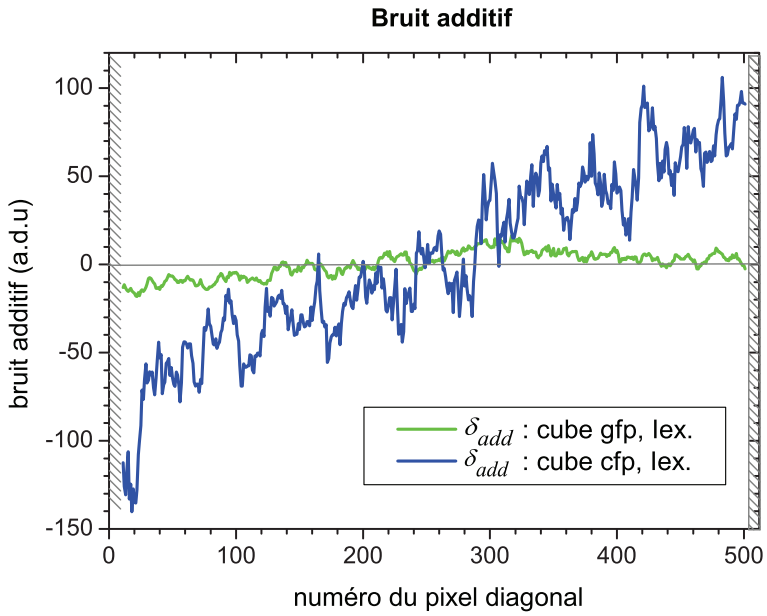


Figure III.1.8.: Inhomogénéités additives selon la diagonale de l'image. Celles-ci sont calculées à partir des échantillons 1 et 2 imagés sous module réflecteur *cyan* (courbe bleu) ou *vert* (courbe verte); l'intensité d'illumination utilisée est la même que dans nos expériences de cellules *gyrA-K* individuelles. Le bruit additif est déduit des images acquises par l'opération III.1.3. Les niveaux de gris ainsi obtenus sont corrigés du bruit de la caméra par un filtre médian de voisinage 20: les pixels résultants affectés par les effets de bord ne sont pas représentés (quadrillage gris).

La même étude a été menée pour les images acquises sous module réflecteur *vert*, puis *jaune* avec cette fois-ci des échantillons de Rhodamine-6G. Les résultats sont les mêmes que pour le module réflecteur *cyan*. On note toutefois que les inhomogénéités additives et les stries sont moins sensibles aux longueurs d'onde des modules réflecteurs cubes *jaune* et *vert*.

b.4. Carte des inhomogénéités

Le profil diagonal est un bon indicateur des inhomogénéités. Mais pour corriger les images d'intérêt, il faut obtenir la carte des inhomogénéités sur toute l'image: on applique les opérations III.1.2 et III.1.3 à tous les pixels de l'image, et non plus aux pixels diagonaux seulement. On obtient alors:

- Une image qui caractérise les inhomogénéités multiplicatives, l'image *flatfield* (figure III.1.9 ci-dessous).

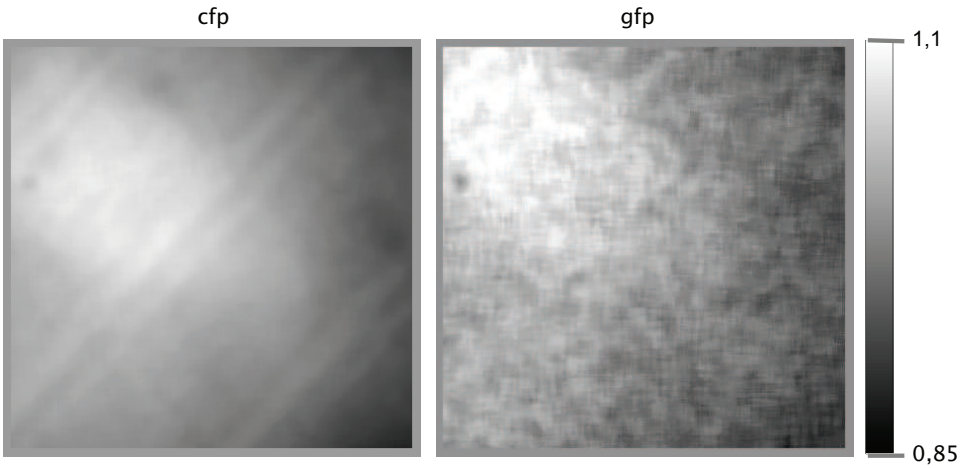


Figure III.1.9.: Image flatfield: carte des inhomogénéités multiplicatives sous module réflecteur *cyan* (image de gauche) et *vert* (image de droite). Ces images correctrices ont été obtenues à partir des échantillons 1 et 2 imagés par une intensité d'illumination $0,05I_{ex}$ (temps d'exposition de 100ms). Les cartes d'inhomogénéité sont ensuite calculées selon la relation III.1.2, puis elles sont filtrées par un filtre médian de voisinage 20 (cf figure III.1.4).

- Une image qui caractérise les inhomogénéités additives pour une intensité d'illumination donnée (voir figure III.1.10 ci-dessous).

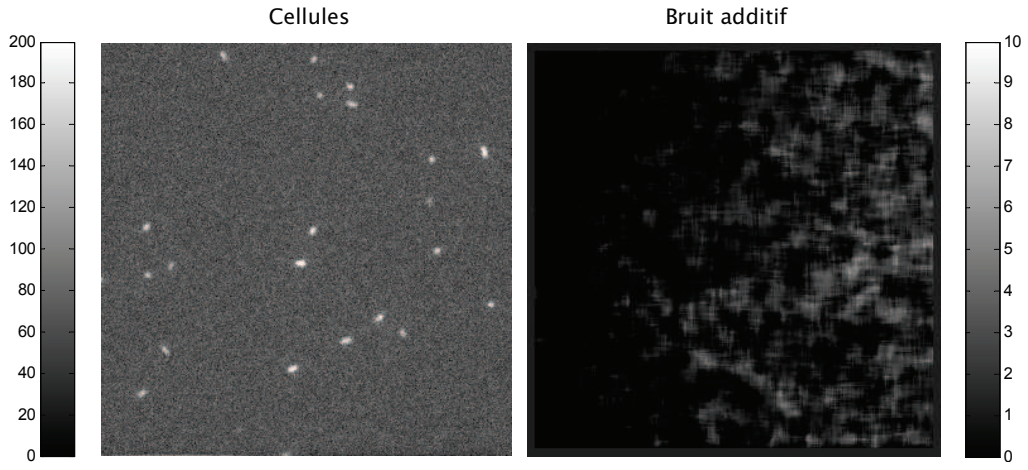


Figure III.1.10.: A gauche: cellules contenant des rapporteurs chromosomaux révélant l'activité du gène *gyrA*. Les cellules ont été imagées sous module réflecteur *cyan* par une intensité d'illumination $0,05I_{ex}$ (temps d'exposition de 100ms). A droite: carte des inhomogénéités additives pour l'intensité d'illumination utilisée sur l'image de gauche. Cette carte à été obtenue à partir des échantillons 1 et 2 imagés sous module *cyan* par une intensité d'illumination $0,05I_{ex}$ (temps d'exposition de 100ms). La carte d'inhomogénéité est ensuite calculée selon la relation III.1.3. Puis elle est filtrée par un filtre médian de voisinage 20. Par mesure de visibilité, les deux images sont représentées sur des échelles de niveaux de gris différentes.

c. Correction des signaux parasites

Au final, les signaux parasites sont corrigés sur l'image de fluorescence comme suit:

$$\text{Image utile} = \frac{\text{Image brute-signal additif uniforme}}{\text{carte flatfield}} \tag{III.1.4}$$

On note que la correction "flatfield" n'a pas été faite sur les images acquises en gfp avec les rapporteurs plasmidiques: comme le montre la figure III.1.11 ci-dessous, cette correction ne change pas nos analyses statistiques. Nos images sont maintenant prêtes à être analysées.

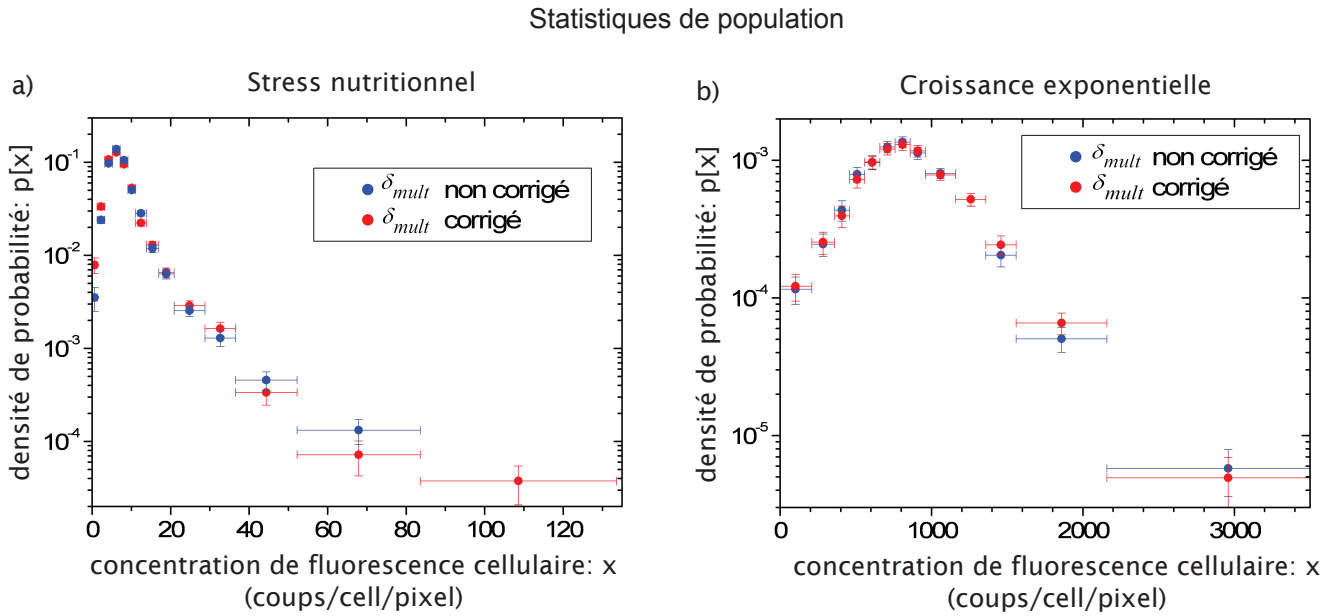


Figure III.1.11.: Données analysées sans correction du bruit multiplicatif (bleu) et avec correction (rouge). Les statistiques de population sont tracées sur des cellules contenant des rapporteurs plasmidiques qui traduisent l'activité du gène *gyrA*. Les cellules ont été imagées sous cube gfp par une intensité d'illumination $0,1I_{ex}$ (temps d'exposition de 100ms). Les données corrigées du bruit multiplicatif sont tirées des images de fluorescence corrigées par la carte flatfield. a) Statistique en début d'expérience calculée sur 2661 cellules (phase de stress nutritionnel). b) Statistique en fin d'expérience calculée sur 759 cellules (phase de croissance exponentielle).

III.1.2. Programme d'analyse d'images (Schnitzcells)

Le programme s'exécute en 3 étapes (schématisées sur la figure III.1.12) :

- Une étape de segmentation : les cellules sont détectées à partir des images à contraste de phase.

- Une étape de traçage cellulaire : le programme reconnaît les cellules à travers les différentes images. Ce traçage inclut la détection des événements de division cellulaire; il permet donc de dresser l'arbre généalogique de chaque cellule.
- Une étape d'extraction des informations associées aux cellules : on répertorie pour chaque lignée cellulaire [1] , l'intensité de fluorescence et la taille de la cellule aux différents temps d'acquisition.

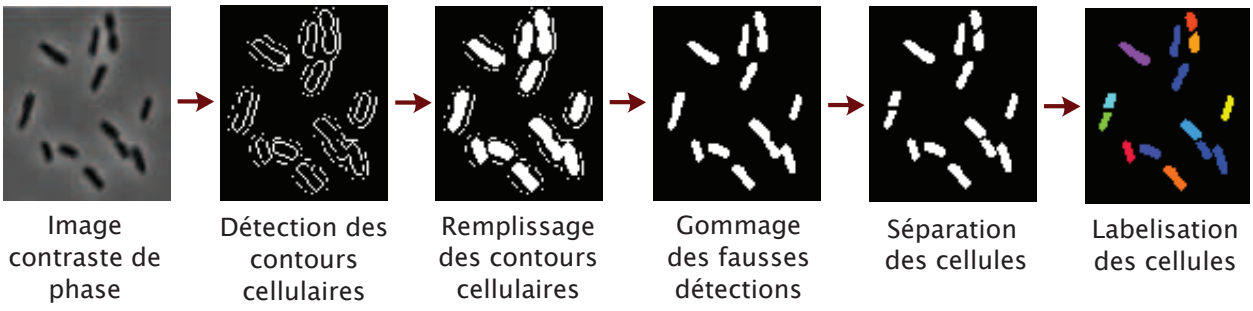


Figure III.1.12.: Principales étapes de la segmentation. Ces étapes sont détaillées dans l'annexe avec les étapes intermédiaires qui ne sont pas montrées ici.

a. Segmentation

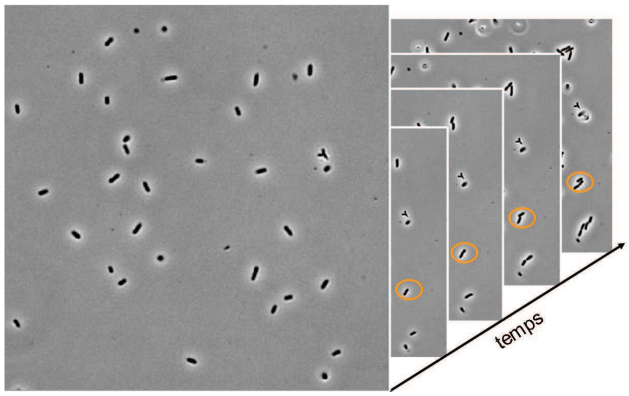
L'étape de segmentation inclut la détection des contours cellulaires, le remplissage de ces contours, le gommage des cellules faussement détectées, la distinction cellulaire avec séparation des cellules lorsqu'elles se touchent, et la labellisation des cellules (voir figure III.1.12). L'efficacité de la segmentation est commandée par un certain nombre de paramètres que nous détaillons dans l'annexe C. Citons, à titre d'exemple le nombre minimal de pixels constituant une cellule : si ce nombre est trop faible, peu de cellules sont détectées ; au contraire, s'il est trop élevé, trop de cellules apparaissent avec beaucoup de fausses détections.

Le programme de segmentation est initialement conçu pour détecter des cellules de type *E.Coli* ou bacille imagées sous objectif 100x par une caméra de résolution 12µm.pixel^{-1}. Or nos cellules *E.Coli* sont imagées par une caméra de résolution moindre, à savoir 16µm.pixel^{-1}. Les paramètres de segmentation ont donc été changés pour détecter correctement nos cellules: les valeurs utilisées sont données dans l'annexe. Par ailleurs, nous nous intéressons aussi à un type de cellules que le programme ne connaît pas: les cellules Vibrio Harveyi, imagées sous une résolution de 20µm.pixel^{-1}. Ces cellules ont la même forme en bâtonnet que eschericchia coli, mais leur taille est ≈ 2 fois plus petite. Le pro-

1. branche de l'arbre généalogique d'une cellule.

gramme de segmentation a donc été adapté à ces cellules (voir annexe C). Dans tous les cas, la segmentation n'est jamais parfaite et elle nécessite une vérification manuelle. Une sub-routine permet de vérifier les images segmentées une à une et de corriger les erreurs de segmentation. Cette étape manuelle permet par ailleurs de supprimer les cellules comètes (cf. Figure III.1.1) qui ne sont pas exploitables.

b. Traçage des cellules



Le traçage se fait sur une série d'images segmentées en cellules distinctes ; les images doivent correspondre au même champ de vue et doivent être acquises à des temps différents. L'opération consiste à identifier les points similaires sur les images successives pour reconnaître les cellules à travers les différentes photos (voir figure ci contre). Au final, les données ne sont plus représentées par une série d'image, mais par des lignées cellulaires qui contiennent les informations de la cellule aux différents temps d'acquisition. Des erreurs peuvent apparaître lors de ce traitement, notamment si deux images successives sont légèrement décalées suite à un mouvement micrométrique de l'échantillon par rapport à la caméra. Une vérification manuelle est donc encore nécessaire : les lignées sont vérifiées une à une. Elles peuvent être rejetées ou corrigées avant d'être approuvées.

c. Extraction des informations associées aux cellules

Maintenant qu'on connaît les arbres généalogiques, on peut s'intéresser aux caractéristiques des cellules, à savoir leur niveau de fluorescence et leur taille. Une sub-routine extrait pour chaque lignée, et chaque temps d'acquisition, les informations relatives à la cellule. A chaque étape du programme, les données sont enregistrées automatiquement dans les dossiers initialement préparés pour pouvoir être réutilisées : cela évite de reprendre le traitement du début lorsqu'il n'a pas abouti (chaque étape prend plusieurs heures). Au final, on obtient une structure Matlab qui contient toutes les lignées ; chaque lignée contient les caractéristiques de la cellule aux différents temps d'acquisition. Pour chaque expérience, on analyse ainsi un par un les champs de vue.

d. Rejection des cellules malades

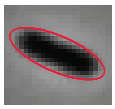
Certaines cellules détectées sont en train de mourir: elles ne doivent pas être considérées dans nos analyses. Nous avons donc rajouté une subroutine qui permet de rejeter ces cellules, sans pour autant effacer leurs informations relatives: cela permet de ne pas refaire le traitement d'image, si on décide un jour de s'intéresser au phénomène de mort cellulaire. Sur les images de contraste de phase, la mort cellulaire se traduit par une diminution dans le temps du niveau de gris de la cellule. Nous calculons donc pour chaque cellule, et chaque temps, le niveau de gris associé en contraste de phase. Les cellules suspectes sont ensuite sélectionnées sur les faibles niveaux de gris consistants à travers les différents frames. Pour chaque cellule ainsi sélectionnée, on vérifie sur les images de contraste de phase que la détection automatique n'est pas le produit d'un artefact; si nous sommes satisfaits de la sélection, nous rejetons la cellule malade. Par ailleurs, si nous remarquons des cellules malades non détectées, nous les rejetons également.

e. Post-traitement

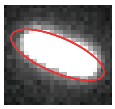
Afin de considérer un maximum de cellules lors des analyses statistiques, on regroupe pour chaque expérience, les branches mères et filles [2] des différents champs de vue. Ces branches sont ensuite sauvegardées. On peut alors en extraire les caractéristiques que l'on utilisera par la suite.

A savoir:

- la taille L des cellules aux différents temps de la lignée : elle est définie comme la longueur de l'axe majeur de la cellule, cette dernière étant modélisée par une ellipse.



- la fluorescence totale FG d'une cellule aux différents temps de la lignée : celle-ci est déterminée à partir des images de fluorescence corrigées des artefacts d'imagerie; elle est donnée par le niveau de gris de la cellule.



- le niveau moyen de fluorescence x d'une cellule aux différents temps de la lignée : c'est le niveau de gris moyen des pixels de la cellule.
- La surface SZ des cellules aux différents temps de la lignée: c'est le nombre total de pixels occupés par la cellule.

Une fois que les données d'intérêt ont été extraites, elles sont converties en format ASCII pour être analysées sous OriginPro 6.1.

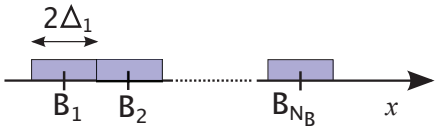
2. Equivalentes à une génération de la lignée cellulaire. Cela correspond à toutes les données relatives à une cellule, depuis sa naissance jusqu'à sa division ou sa mort.

III.2. Traitement des données

Pour quantifier la variabilité inter-cellulaire, les données reccueillies seront traitées sous forme de statistique. Pour chaque temps d'acquisition, on calcule ainsi la loi de probabilité associée au niveau moyen de fluorescence cellulaire x: Quelle est la probabilité, qu'à l'instant t, une cellule ait une niveau moyen de fluorescence x ?

III.2.1. Statistiques de population

Pour estimer cette loi de probabilité, on trace l'histogramme des niveaux de fluorescence mesurés à l'instant t. Pour ce faire, on discrétise l'espace continu des fluorescences cellulaires x, en N_B intervalles de classe (bins discrets):



Puis on compte pour chaque bin B_i, le nombre n_i de cellules imagées dont le niveau de fluorescence appartient au bin. On en déduit alors la probabilité P de mesurer parmi la population de cellules imagées, un niveau de fluorescence x qui appartienne au bin B_i:

$$P(B_i - \triangle_i \leq x \leq B_i + \triangle_i) = \frac{n_i}{N_{cell}}$$

où $2\triangle_i$ désigne la largeur du bin B_i, et N_{cell} la taille de la population (*i.e.* le nombre total de cellules imagées). Dans cet histogramme normalisé, la probabilité dépend de la taille du bin: plus le bin est large, plus on a de chance de mesurer un niveau de fluorescence x appartenant à $[B_i - \triangle_i,\ B_i + \triangle_i]$. Pour s'affranchir de cette dépendance, on travaillera avec les densités de probabilité p définies comme suit:

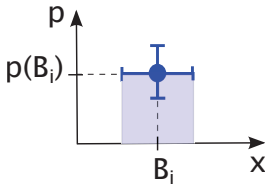
$$p(B_i - \triangle_i \leq x < B_i + \triangle_i) = \frac{n_i}{N_{cell}} \times \frac{1}{2\triangle_i}$$

Et à chaque valeur $x = B_i$, on associera la densité de probabilité $p(B_i)$ selon:

$$p(B_i) = p(B_i - \triangle_i \leq x < B_i + \triangle_i)$$

L'ensemble de ces densités de ces probabilités $\boldsymbol{p} = \{p(B_i)\}_i$ constitue la statistique de la population de cellules imagées.

a. Barres d'erreur



Cette statistique est tracée avec des incertitudes, comme illustré sur la figure ci-contre: d'une part, le lieu des bins est donné avec une imprécision $\triangle_i$; d'autre part, le contenu $p(B_i)$ des bins est mesuré avec des erreurs statistiques. En effet, comme la concentration de fluorescence cellulaire x est une variable aléatoire, le nombre n_i d'entrées dans un bin est incertain. Et on peut montrer qu'il suit une loi de probabilité binomiale de paramètre $P_i \approx n_i/N_{cell}$. L'écart type associé à l'occurrence n_i du bin B_i peut donc être évalué par [1]:

$$\sigma_{n_i} = \sqrt{n_i\Big(1 - \frac{n_i}{N_{cell}}\Big)}$$

avec N_{cell} qui correspond au nombre de cellules utilisées pour dresser la statistique. Dans ces conditions, l'écart type associé à la densité de probabilité $p(B_i)$ vaut [2]:

$$\sigma_{p(B_i)} = \sqrt{n_i\Big(1 - \frac{n_i}{N_{cell}}\Big)} \times \frac{1}{N_{cell} \times 2\triangle_i}$$

Donc au final, les densités de probabilité $p(B_i)$ sont mesurées avec les incertitudes:

$$e\big(p(B_i)\big) = \sigma_{p(B_i)}$$
$$e\big(B_i\big) = \triangle_i$$

b. Choix des bins

Pour que les distributions tracées dessinent bien la loi de probabilité, on essaiera de minimiser la barre d'erreur sur $p(B_i)$. Cela implique de choisir des bins larges, pour lesquels l'occurence n_i est grande [3]. Mais ces bins ne devront pas être trop larges, car sinon l'incertitude $e(B_i)$ devient grande, et on perd en précision sur la forme de la statistique. Le nombre de bins utilisés et la taille des bins doivent donc être choisis avec précaution. On choisit de tracer nos statistiques avec un nombre N_B de bins donné par la régle de Doane [55]:

$$N_B = 1 + log_2(N_{cell}) + log_2(1 + \gamma\sqrt{N_{cell}/6})$$

où le skewness γ mesure l'assymétrie de la distribution par rapport à sa moyenne μ:

1. L'écart type d'une distribution binomiale de paramètre P est donné par: $\sigma = \sqrt{N \times P(1-P)}$ où N désigne le nombre de réalisations étudiées.
2. Par définition de l'écart type et par monotonie de la fonction $p_i : n_i \to p(B_i)$, on a: $\sigma^2(p_i) = |\frac{dp_i}{dn_i}|^2 \times \sigma^2(n_i)$.
3. L'incertitude $e\big(p(B_i)\big)$ est proportionnelle à $(1 - n_i/N_{cell}) \times 1/(2\triangle_i)$.

$$\gamma = \frac{E[(x-\mu)^3]}{\sigma^3}$$

avec σ qui désigne l'écart type de la distribution [4].
A partir de là, on construit un premier histogramme où la taille $2h$ des bins est constante, fixée à:

$$2h = \frac{x_{max} - x_{min}}{N_B}$$

où x_{max} correspond au plus grand niveau de fluorescence mesuré sur la population de cellules imagées, et x_{min} au plus petit niveau de fluorescence mesuré.

Nos distributions sont assymétriques avec des densités de probabilité qui diminuent plus lentement pour les concentrations élevées de fluorescence cellulaire que pour les très faibles valeurs de fluorescence (voir figureIV.3.13 par exemple). On travaillera avec des bins de taille variable: la largeur des bins sera augmentée lorsqu'on se déplace vers les concentrations élevées de fluorescence cellulaire. On s'assure ainsi de tracer des statistiques lisses, où les bins qui caractérisent les concentrations de fluorescence cellulaire élevées, sont peuplés. En fait, on troque l'imprécision sur $p(B_i)$ contre la précision sur B_i : l'incertitude sur $p(B_i)$ est diminuée, et l'erreur $\triangle_i$ sur B_i est augmentée.
Pour ce faire, on utilise la taille constante $2h$ précédemment calculée. Et on double cette taille tous les q bins, comme expliqué sur la figure III.2.1 ci-dessous où $q = 3$.

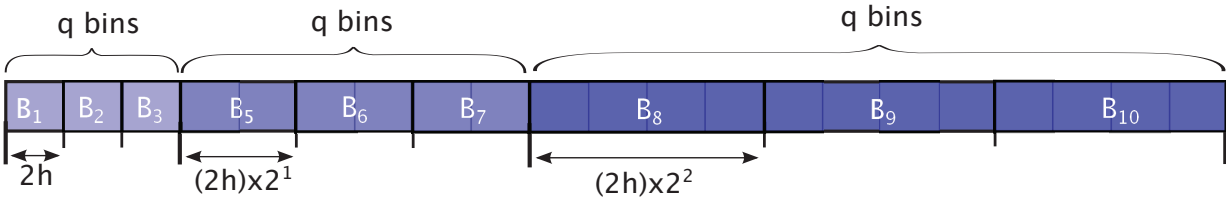


Figure III.2.1.: La taille des bins est variée pour s'assurer que, dans nos statistiques, tous les bins sont correctement remplis: elle est doublée tous les q bins (dans cet exemple, $q = 3$).

Le paramètre q est arbitraire, et il doit être fixé par l'utilisateur. Nous choisirons une valeur qui lisse la queue de la distribution, sans pour autant changer les bins proches de la moyenne. A titre d'exemple, nous présentons sur la figure III.2.2 suivante, une statistique tracée pour une taille de bin constante, et la même statistique tracée pour une taille de bin variable. Cette statistique a été calculée à partir de 195 cellules *GyrA*, et on voit très bien que le choix d'une taille de bin variable permet de lisser la courbe.

4. Celui-ci est estimé à partir de l'échantillon $\{x_i\}_i$ des niveaux de fluorescence mesurés:
$\sigma \approx \frac{1}{N_{cell}-1} \sum_{i=1}^{N_{cell}} (x_i - \mu)$

On travaillera donc avec les probabilités $\boldsymbol{p} = (p_1, p_2, ..., p_{N_B})$ mesurées sur les bins $\boldsymbol{B} = (B_1, B_2, ..., B_{N_B})$ de taille variable. On note que lorsque le binnage automatique ne lisse pas bien la statistique, il est réajusté manuellement pour les quelques bins qui génèrent de fortes "discontinuités" sur la distribution.

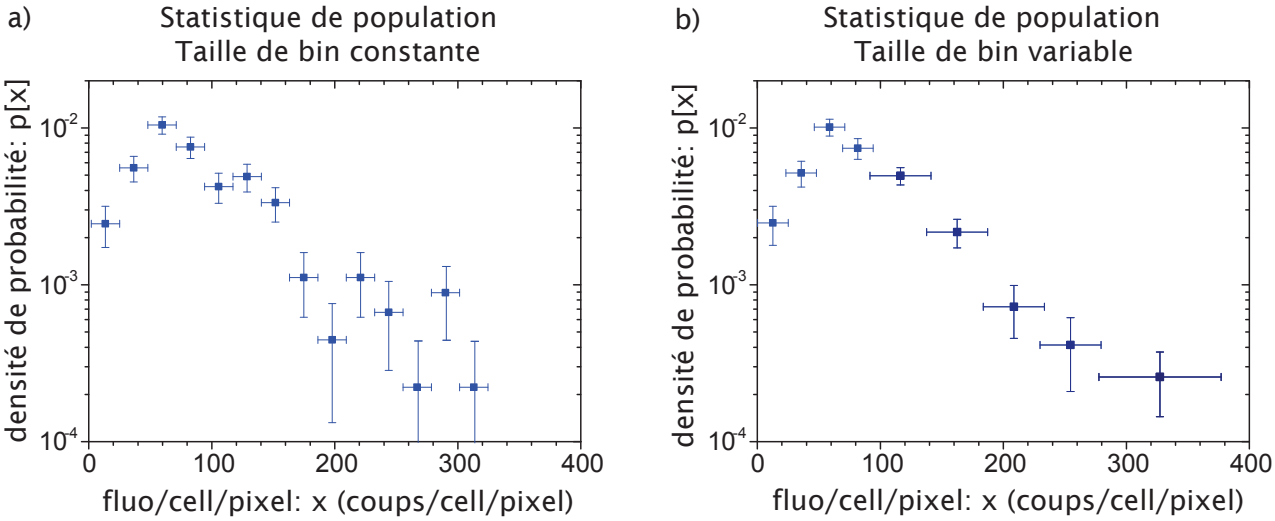


Figure III.2.2.: Statistique de population tracée sur 195 cellules contenant des rapporteurs plasmidiques qui indiquent l'activité du gène *gyrA*. Ces cellules ont été imagées 55 min après l'injection du glucose au milieu de culture. a) Distribution est tracée pour une taille de bin constante. b) La même distribution est tracée pour une taille de bin variable ($q = 4$).

III.2.2. Ajustement numérique des statistiques

Ces statistiques représentent les variations intercellulaires. Et pour accéder à la loi de probabilité f qui génère ces variations, on testera des modèles de loi de probabilité: le modèle $f : (x;\ a, b) \rightarrow p(x)$ reproduit-il le processus qui génère nos distributions expérimentales? Pour tester un modèle, on utilise l'ajustement numérique d'OriginPro 6.1 [5]. Ce dernier recherche les paramètres (a, b) du modèle qui décrivent le mieux nos statistiques expérimentales. Il indique pour les paramètres optimums, le niveau de ressemblance entre le modèle et la statistique expérimentale.

Recherche des paramètres optimums du modèle

Pour rechercher ces paramètres, OriginPro 6.1 utilise une méthode de régression non linéaire basée sur l'algorithme de Levenberg Marquadt: les paramètres (a, b) initiaux sont

5. Par soucis de notations, nous décrivons ici l'ajustement numérique pour un modèle à deux paramètres: a et b. Mais on note qu'OriginPro 6.1 est capable de réaliser des ajustements numériques sans restrictions sur le nombre de paramètres.

donnés par l'utilisateur, et l'algorithme itére les valeurs de ces paramètres (a, b) pour minimiser le χ^2.
Ce χ^2 quantifie la déviation entre la statistique expérimentale et la statistique générée par le modèle: il calcule, pour chaque bin B_i, l'écart entre les points expérimentaux p_i, et les densités de probabilité $f(b_i; a, b)$ générées par le modèle, et il somme le carré de ces écarts:

$$\chi^2(a,b) = \frac{1}{N_B - n_P} \sum_{i=1}^{N_B} \Big(w_i \times [p_i - f(B_i; a, b)]^2 \Big)$$

où $N_B - n_P$ défini le degré de liberté de l'ajustement: N_B est le nombre de bins utilisés pour réaliser l'ajustement, et n_P représente le nombre de paramètre à ajuster.
Quant à w_i, il donne le poids accordé à chaque point expérimental p_i. Pour privilégier les points expérimentaux qui sont précis, nous le choisissons comme suit:

$$w_i = \frac{1}{e(p_i)^2}$$

où $e(p_i)$ est la barre d'erreur sur p_i calculée dans la section a précédente.

A chaque pas d'itération, l'ajustement indique la valeur du χ^2. Et l'itération est arrétée manuellement lorsque le χ^2 n'est plus minimisé. L'ajustement retourne alors les valeurs optimales des paramètres du modèle (au sens du χ^2), et il indique les barres d'erreur associées.

Niveau de ressemblance entre le modèle et la statistique expérimentale

L'adéquation entre le modèle et les données observées est indiqué par deux quantités: le χ^2 et le coefficient de détermination R^2.
Ce coefficient de détermination est défini par la formule ci-dessous:

$$R^2 = \frac{\text{variation expliquée par le modèle de régression}}{\text{variation totale des données expérimentales}} = \frac{\sum_i \big(f(B_i; a, b) - \bar{\boldsymbol{p}}\big)^2}{\sum_i \big(p_i - \bar{\boldsymbol{p}}\big)^2}$$

où $\bar{\boldsymbol{p}}$ indique la moyenne de l'échantillon expérimental $\boldsymbol{p} = (p_1, p_2, ..., p_{N_B})$.
Ce coefficient peut se réécrire comme suit:

$$R^2 = 1 - \frac{\sum_i \big(p_i - f(B_i; a, b)\big)^2}{\sum_i \big(p_i - \bar{p}\big)^2}$$

Donc, plus l'ajustement $f(B_i; a, b)$ est proche des points expérimentaux p_i, plus ce coefficient se rapproche de 1.

Optimisation de l'ajustement numérique

Pour optimiser cet ajustement numérique, nous travaillerons avec une échelle d'ordonnée logarithmique lorsque le modèle testé sera une distribution gamma. Cela permet d'accorder le même poids aux valeurs faibles de la queue exponentielle et aux valeurs élevées de la bosse. Pour ce faire, nous traçons le logarithme népérien de la statistique expérimentale. Puis, nous utilisons comme fonction modèle, le logarithme de la distribution gamma:

$$f(x;\ a,b) = ln(b) - ln\big(\Gamma(a)\big) + (a-1) \times ln\left(\frac{x}{b}\right) - \frac{x}{b}$$

III.2.3. Equivalence de deux statistiques

Lorsqu'on veut comparer une statistique, non pas à un modèle continu, mais à une autre distribution expérimentale, on utilisera le test Kolmogorov-Smirnov qui est indépendant du binnage. Ce test évalue la véracité de l'hypothèse H_0 définie comme suit:

Hypothèse "H_0": les deux distributions sont équivalentes.

Test de Kolmogorov-Smirnov [56, 57]

Pour quantifier la véracité de cette hypothèse, on peut comparer les données de fluorescence relatives aux deux échantillons d'étude:

- Soit $[x_1^{(1)}, x_2^{(1)}, ..., x_{Ncells1}^{(1)}]$ l'échantillon E_1 mesuré, engendré par la distribution 1.
- Soit $[x_1^{(2)}, x_2^{(2)}, ..., x_{Ncells2}^{(2)}]$ l'échantillon E_2 mesuré, engendré par la distribution 2.

Pour déterminer si ces deux échantillons sont issues de la même loi de probabilité, on calculera la fonction F de répartition empirique. Celle-ci définit la probabilité que la concentration de fluorescence cellulaire mesurée soit inférieure à x:

$$F(x) = \frac{1}{N_{cells}} \times \sum_{i=1}^{Ncells} \delta_{x_i \leq x} \text{ avec } \delta_{x_i \leq x} = \begin{cases} 1 \text{ si } x_i \leq x \\ 0 \text{ sinon} \end{cases}$$

Le test de Kolmogorov-Smirnov calcule alors la valeur maximum $D_{Ncells1,Ncells2}$ de l'écart entre la fonction de répartition $F_1(x)$ relative à l'échantillon E_1, et celle $F_2(x)$ relative à l'échantillon E_2:

$$D_{Ncells1,Ncells2} = \sup_x |F_{1,Ncells1}(x) - F_{2,Ncells2}(x)|$$

où N_{cell} indique la taille de l'échantillon.
Si l'hypothèse H_0 est vraie, alors $D_{Ncells1,Ncells2}$ suit une loi de probabilité connue qui dépend des valeurs de N_{cells1} et N_{cells2}, et pas des distributions 1 et 2. Dans ce cas, pour une probabilité α donnée, on peut trouver (cf tableau III.2.1) la valeur $D_{critique}(\alpha)$ telle que:

$$P\Big(D_{Ncells1,Ncells2} > D_{critique}(\alpha)\Big) \Big\langle \alpha$$

Le " test de Kolmogorov avec un niveau de signification statistique $1-\alpha$ " consiste alors à rejeter H_0 si $D_{Ncells1,Ncells2} > D_{critique}(\alpha)$. Par exemple, dans le test avec un niveau de signification statistique de $1-\alpha = 95\%$, on rejetera H_0 si on mesure une valeur de $D_{Ncells1,Ncells2}$ supérieure à la valeur $D_{critique}(0,05)$. Si les distributions 1 et 2 sont identiques mais que la malchance nous a fourni des échantillons E_1 et E_2 non représentatifs de la vraie distribution dont ils sont issus, alors il est possible que nous rejettions H_0 à tord, mais nous avons l'assurance que cela ne se produira que dans $\alpha = 5\%$ des expériences.

Calcul de $D_{critique}$: Dans nos échantillons, le nombre N_{cell} de cellules considérées est supérieur à 100. Dans ces conditions, la valeur critique $D_{critique}$ s'obtient par le tableau III.2.1, où $s(n)$ est calculé à partir de la taille des échantillons étudiés [57] :

$$s(n) = \sqrt{\frac{N_{cells1} + N_{cells2}}{N_{cells1} \times N_{cells2}}}$$

Critical Values and Confidence Levels for K-S Two-Sample Test

α	$D_{crit}/s(n)$
0.10	1.22
0.05	1.36
0.01	1.63
0.005	1.73
0.001	1.95

TABLE III.2.1.: Valeurs critiques $D_{critique}$ du test de Kolmogorov-Smirnov. D'après I.T.Young [57] . Si $D_{Ncells1,Ncells2} > D_{crit}$, alors l'hypothèse H_0 est rejetée avec un niveau de signification $(1-\alpha)$ de confiance.

Calcul de D : Pour calculer D, on travaille sur les échantillons $[x_1, x_2, ..., x_{Ncell}]$ de concentrations de fluorescence cellulaire mesurées. Ces échantillons sont rangés par ordre croissant. Et pour chaque valeur x_i ainsi ordonnée, la probabilité d'observer une concentration de fluorescence cellulaire, inférieure ou égale à x_i, est donnée par:

$$F(x_i) = \frac{i}{N_{cell}}$$

où N_{cell} désigne le nombre de cellules qui composent l'échantillon (taille de l'échantillon). Pour pouvoir comparer $F_1(x)$ avec $F_2(x)$, on calcule les fonctions de répartition pour chacune des valeurs x_i mesurées dans l'ensemble {échantillon 1 + échantillon 2}.
D se déduit alors trivialement en prenant le maximum des écarts $|F_1(x) - F_2(x)|$.
Ces calculs sont effectués sous OriginPro 6.1.

III.3. Déconvolution de l'autofluorescence

Les statistiques obtenues traduisent les variations du signal moyen de fluorescence émis par chacune des cellules de la population imagée.
Or, le signal de fluorescence émis par une cellule est produit par deux objets distincts:

- Les fluorophores rapporteurs "fp" introduits dans la cellule: une fois excités à la longueur d'onde adéquate, ils relaxent vers leur état fondamental en libérant l'énergie accumulée sous forme de photons.
- Les composants cellulaires naturels: certains sont excités à la même longueur d'onde que les fluorophores, et la relaxation radiative peut libérer la même énergie que les rapporteurs. On parle dans ce cas d'*autofluorescence*.

La fluorescence détectée est donc la combinaison de deux signaux de fluorescence:

$$F_{cell} = F_a + F_{fp}$$

Ces deux quantités sont soumises à l'aléa cellulaire, et elles peuvent être considérées indépendantes. En terme de densité de probabilité, cela se traduit par:

$$p_{F_{cell}}(x) = \left(p_{F_{fp}} \otimes p_{F_a}\right)(x) = \int_{-\infty}^{+\infty} p_{F_{fp}}(u) \times p_{F_a}(u-x)du \qquad \text{(III.3.1)}$$

- $p_{F_{cell}}(x)$: densité de probabilité qu'une cellule ait un niveau de fluorescence totale x (voir figure III.3.1 ci-après). Elle est mesurée sur des cellules contenant les rapporteurs fluorescents.
- $p_{F_a}(x)$: densité de probabilité qu'une cellule ait un niveau d'autofluorescence x (voir figure III.3.1 ci-après). Elle est obtenue à partir de souches sauvages ne contenant pas de rapporteurs fluorescents (cellules wt).
- $p_{F_{fp}}(x)$: densité de probabilité que les fluorophores de la cellule génèrent un niveau de fluorescence x.

Nous nous intéressons à la variabilité inter-cellulaire dans la quantité de rapporteurs, c'est à dire: $p_{F_{fp}}$. Or nos expériences ne fournissent qu'une information sur la variabilité

dans la fluorescence cellulaire totale: $p_{F_{cell}}$, ou p_{F_a} pour les cellules ne contenant pas de fluorophores rapporteurs (voir figure III.3.1 ci-dessous). Pour accéder à l'information d'intérêt, il faudra donc déconvoluer l'autofluorescence du signal détecté sur les cellules contenant les rapporteurs.

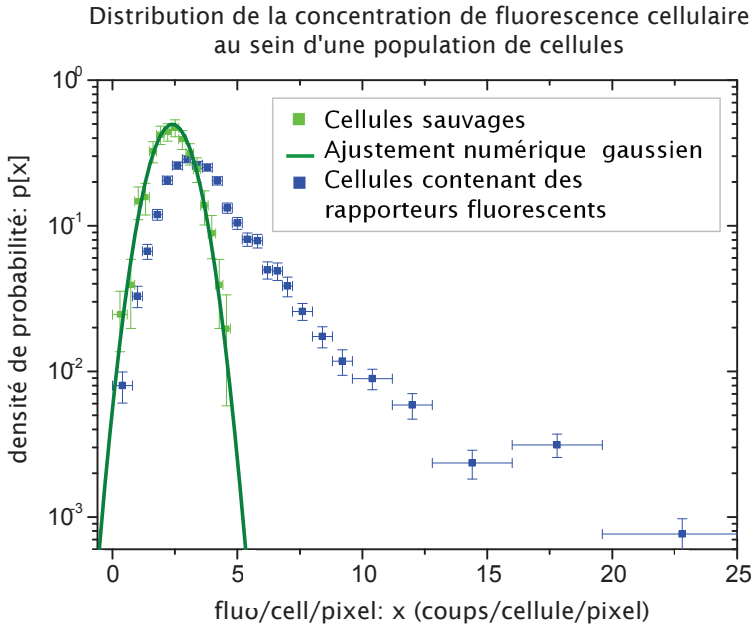


Figure III.3.1.: Distribution de la concentration de fluorescence cellulaire pour une population de cellules *GyrA* (1857 cellules, distribution bleue) et sauvages (691 cellules, distribution vert clair). Les deux types de cellules étudiées proviennent de deux échantillons distincts, fabriqués selon le même protocole. La courbe continue verte est l'ajustement numérique de la distribution d'autofluorescence par une gaussienne ($\chi^2 \approx 0.001$, $R^2 \approx 0.97$) de moyenne $x_c = 1.8$ et d'écart type $\sigma = 0.6$.

III.3.1. Déconvolution matricielle

a. Méthode

Soit $\boldsymbol{Y} = [y_1,\ y_2,\ \ldots\ ,\ y_k,\ \ldots\ ,\ y_N]$ la distribution observée ($p_{F_{cell}}$) tracée sur les N bins $\boldsymbol{B} = [B_1,\ B_2,\ \ldots\ ,\ B_k,\ \ldots\ ,\ B_N]$ de taille variable $2\triangle_{B_k}$ (voir figure III.3.2 ci-dessous). Cette distribution est tracée à partir de données expérimentales collectées sur des cellules contenant les rapporteurs fluorescents.

Soit $\boldsymbol{X} = [x_1,\ x_2,\ \ldots\ ,\ x_k,\ \ldots\ ,\ x_{Nb}]$ la distribution d'intérêt ($p_{F_{fp}}$) discrétisée sur les N_b bins $\boldsymbol{b} = [b_1,\ b_2,\ \ldots\ ,\ b_k,\ \ldots\ ,\ b_{Nb}]$ de taille variable $2\triangle_{b_k}$ (voir figure III.3.2 ci-dessous). C'est la distribution qu'on cherche à déterminer.

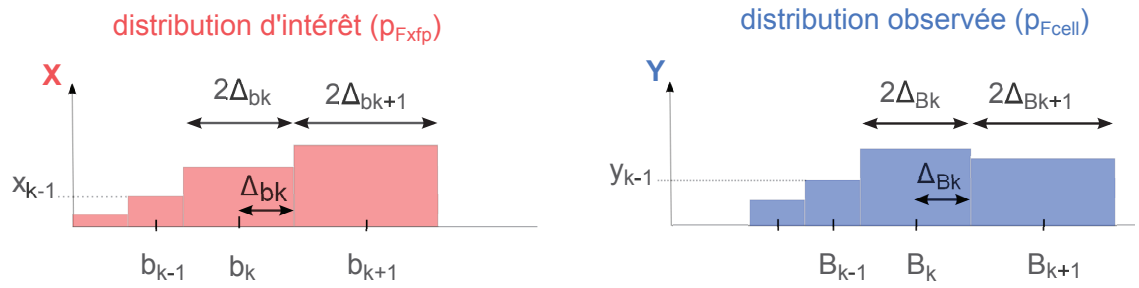


Figure III.3.2.: Binnage de la distribution d'intérêt, et de la distribution expérimentale (que l'on observe).La distribution observée $\boldsymbol{Y}$ porte l'information de la distribution d'intérêt: c'est le produit de convolution de la distribution $\boldsymbol{X}$ d'intérêt avec la distribution d'autofluorescence (p_{Fa}).

Afin de garder une résolution exploitable [1] sur $p_{F_{fp}}$, tout en restant compatible avec la précision expérimentale [2], nous prenons le même nombre de bins sur $\boldsymbol{X}$ que pour $\boldsymbol{Y}$: $N_b = N$. Avec ces notations discrètes, l'équation III.3.1 se réécrit:

$$\begin{aligned} y_i &= \sum_{k=1}^{N} x_k a_{i,k} \\ a_{i,k} &= \int_{B_i - b_k - \triangle b_k}^{B_i - b_k + \triangle b_k} p_{Fa}(u)du \end{aligned} \tag{III.3.2}$$

Dans l'algèbre matricielle, cela devient:

$$\boldsymbol{Y} = A \bullet \boldsymbol{X} \tag{III.3.3}$$

où A est la matrice de convolution qui porte l'information sur l'autofluorescence des cellules. C'est une matrice dont la taille $(N \times N_b)$ dépend du nombre de bins choisis pour $\boldsymbol{X}$ et $\boldsymbol{Y}$. Cette *matrice d'autofluorescence* est définie par:

$$A = (a_{i,k})_{i \leq N,\ k \leq N_b} \text{ avec : } \quad a_{i,k} = \int_{B_i - b_k - \triangle b_k}^{B_i - b_k + \triangle b_k} p_{Fa}(u)du$$

Et si A est inversible, alors on peut réécrire l'équation III.3.3 comme suit:

$$\boldsymbol{X} = A^{-1} \bullet \boldsymbol{Y} \tag{III.3.4}$$

Dans cette équation, la distribution observée $\boldsymbol{Y}$ est connue: elle est tirée de mesures réalisées sur des cellules contenant les rapporteurs fluorescents. Quant à la matrice A, elle est déduite de p_{Fa} (distribution d'autofluorescence).

1. Si $p_{F_{xfp}}$ n'est pas assez binnée, son ajustement numérique sera mauvais, car réalisé sur un échantillon de petite taille.
2. Si nous mesurions directement la fluorescence émise par les rapporteurs, le nombre limité de cellules imagées exigerait de tracer une distribution à faible fréquence d'échantillonage.

b. Calcul de la matrice A d'autofluorescence

Les expériences menées sur des cellules sauvages *wt* fournissent l'histogramme d'autofluorescence (voir figure III.3.1 précédente). Ce dernier retranscrit la loi de probabilité p_{F_a}: il permet donc de calculer les éléments $a_{i,k}$ de la matrice d'autofluorescence. Pour augmenter la précision sur ces éléments $a_{i,k}$ de convolution, nous décidons de travailler avec une loi de probabilité p_{F_a} continue, et non pas l'histogramme associé qui n'est qu'une estimée [3]. Pour ce faire, nous ajustons numériquement l'histogramme d'autofluorescence par une gaussienne: le résultat de l'ajustement fournit p_{Fa}, comme le montre la figure III.3.1 précédente. A partir de là, les éléments $a_{i,k}$ de A sont calculés numériquement selon:

$$a_{i,k} \approx \sum_{u \geq B_i - b_k - \triangle b_k}^{u \leq B_i - b_k + \triangle b_k} p_{Fa}(u) \times \triangle_u$$

où u est incrémenté par pas de $\triangle_u$, avec $\triangle_u$ qui est choisi 1000 fois plus petit que la largeur à mi hauteur de la gaussienne p_{Fa}. Ce calcul suppose de fixer le binnage $\boldsymbol{b}$ pour la distribution d'intérêt $\boldsymbol{X}$.

c. Choix des bins pour $\boldsymbol{X}$

Et comme ce binnage conditionne les éléments de la matrice de déconvolution, il est important de choisir un binnage adapté aux données expérimentales.

Si la fonction d'autofluorescence est un dirac centré en μ (*i.e.* $p_{F_a} = \delta_\mu$):
On a par propriété du produit de convolution:

$$p_{Fcell}(x) = (p_{xfp} \otimes \delta_\mu)(x) = p_{xfp}(x - \mu)$$

Dans ce cas, la loi de probabilité d'intérêt (p_{xfp}) se déduit de la loi observée (p_{Fcell}) par simple translation de μ, comme illustré sur la figure III.3.3 ci-contre. Un bon binnage $\boldsymbol{b}$ de l'histogramme d'intérêt ($\boldsymbol{X}$) s'appuie alors sur les bins $\boldsymbol{B}$ de l'histogramme observé ($\boldsymbol{Y}$) auxquels on soustrait μ:

$$\boldsymbol{b} = \boldsymbol{B} - \mu.$$

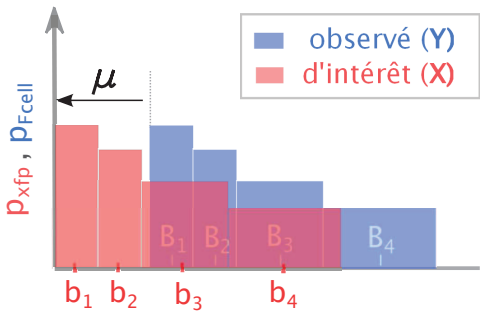


Figure III.3.3.: Convolution par δ_μ: Observée = Intérêt $\otimes$ δ_μ.

Pour une fonction d'autofluorescence gaussienne centrée en 0 (écart type σ):
Nous décidons de choisir le binnage $\boldsymbol{b}$ de la distribution d'intérêt $\boldsymbol{X}$ tel que:

3. La distribution d'autofluorescence est tracé à partir de ≈ 600 cellules, ce qui implique un nombre limité de bins (une dizaine) avec une imprécision non négligeable pour chaque bin.

- Pour le premier bin b_1:
  - Le contenu x_1 du bin b_1 (distribution $\boldsymbol{X}$) génére des entrées dans le bin d'observation B_1 (distribution $\boldsymbol{Y}$). D'après III.3.2, cela implique que: $a_{1,1} \neq 0$.
  - Tout bin b_0 inférieur à b_1 ne génére aucune entrée dans le premier bin d'observation B_1. D'après III.3.2, cela implique que: $\forall b_0 < b_1,\ a_{1,0} = 0$.
- Pour le dernier bin b_N:
  - Le contenu x_N du bin b_N (distribution $\boldsymbol{X}$) génére des entrées dans le bin d'observation B_N (distribution $\boldsymbol{Y}$). D'après III.3.2, cela implique que: $a_{N,N} \neq 0$.
  - Tout bin b_M supérieur à b_N ne génère aucune entrée dans le dernier bin d'observation B_N. D'après III.3.2, cela implique que: $\forall b_M > b_N,\ a_{N,M} = 0$.
- La taille des bins intermédiaires $(b_i)_{2 \leq i \leq N-1}$ (distribution $\boldsymbol{X}$) reste cohérente avec la précision expérimentale: nous leur imposerons une taille $(\triangle b_i)$ variable, qui reflète la taille $(\triangle B_i)$ du bin d'observation correspondant.

De plus, nous considérerons que les concentrations d'autofluorescence supérieures à 2σ, ou inférieures à -2σ, ont une probabilité d'apparition négligeable [4]. Dans ces conditions, les bins b_1 et b_N comme définis ci-dessous remplissent les deux premières contraintes:

$$b_1 = B_1 - 2\sigma + \triangle b_1$$

$$b_N = B_N + 2\sigma - \triangle b_N$$

Et pour la taille des bins, nous choisirons $\triangle_{B_N}$ tel que [5]:

$$\triangle_{b_N} = \triangle_{B_N}$$

Et pour le premier b_1, on décide de fixer la taille à:

$$\triangle b_1 = \sigma + \frac{\triangle B_1}{2}$$

Enfin, pour les bins intermédiaires, nous choisissons une taille qui respecte les contraintes précédemment définies (la dernière contrainte). Pour ce faire, nous choisissons la taille $\triangle_{bi}$ comme suit:

$$\triangle_{b_i} = \triangle_{B_i} \times \frac{I_b}{I_B}$$

où les longueurs I_b et I_B sont définies sur la figure III.3.4 suivante.

4. Pour une distribution gaussienne centrée en 0 et d'écart type σ, les événements ont seulement $4,5\%$ de chance de tomber en dehors de l'intervalle $[-2\sigma; 2\sigma]$.

5. Le dernier bin B_N de la distribution observée $\boldsymbol{Y}$ est du même ordre de grandeur, si ce n'est plus grand que la largeur de la gaussienne p_{F_a}. Dans, ces conditions, on peut considérer que la gaussienne de déconvolution est perçue par le bin B_N comme un dirac centré en 0. Un bon binnage de la distribution d'intérêt $\boldsymbol{X}$ satisfait alors: $\triangle_{b_N} = \triangle_{B_N}$.

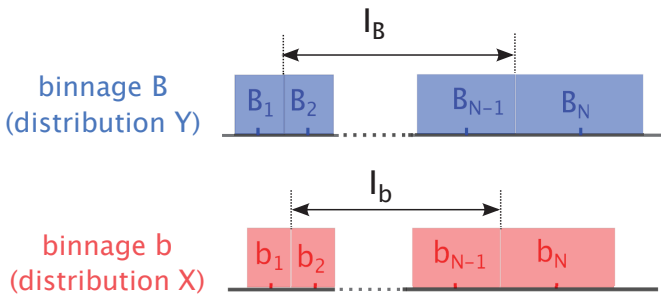


Figure III.3.4.: Les bins compris entre b_1 et b_N (distribution d'intérêt) ont une taille proportionnelle à celle des bins correspondants sur la distribution $\boldsymbol{Y}$ observée.

Le lieu des bins intermédiaires se déduit alors trivialement par récurrence:

$$b_i = b_{i-1} + \triangle b_{i-1} + \triangle b_i$$

Dans nos expériences, la fonction d'autofluorescence n'est pas une gaussienne centrée en 0: c'est une gaussienne centrée en μ (écart type σ).
Elle peut donc s'exprimer comme une combinaison de la gaussienne centrée en 0, et du dirac δ_μ vu au premier paragraphe:

$$\text{Gaussienne}(\mu, \sigma) = \text{Gaussienne}(0, \sigma) \otimes \delta_\mu$$

Dans ces conditions, l'opération de convolution définie dans l'équation III.3.1 peut se diviser en deux opérations: la distribution d'intérêt est d'abord convoluée à une gaussienne centrée en 0 d'écart type σ, puis le tout est convolué à un dirac centrée en μ.

$$p_{Fcell} = \Big(p_{Ffp} \otimes \text{Gaussienne}(0, \sigma)\Big) \otimes \delta_\mu$$

Pour déterminer le binnage $\boldsymbol{b}$ de $\boldsymbol{X}$, nous choisirons alors les mêmes contraintes que pour la gaussienne centrée en 0. Nous obtenons alors pour la distribution d'intérêt $\boldsymbol{X}$, le binnage $\boldsymbol{b}$ défini ci-dessous:

$$\begin{aligned}
\triangle b_1 &= \sigma + \triangle B_1 & b_1 &= \Big(B_1 - 2\sigma + \triangle b_1\Big) - \mu \\
\triangle_{b_i} &= \triangle_{B_i} \times \frac{I_b}{I_B} & b_i &= \Big(b_{i-1} + \triangle b_{i-1} + \triangle b_i\Big) - \mu \\
\triangle_{b_N} &= \triangle_{B_N} & b_N &= \Big(B_N + 2\sigma - \triangle b_N\Big) - \mu
\end{aligned}$$

d. Résultat

A présent la matrice d'autofluroescence A est entiérement déterminée: l'équation III.3.4 précédente ne contient donc qu'une inconnue, $\boldsymbol{X}$. Et du fait du nombre limité de cellules, la distribution observée $\boldsymbol{Y}$ est un échantillon de petite taille ($N \approx 30$ bins): l'équation matricielle peut donc être résolue par une simple opération Matlab.

Le résultat est aberrant, comme le montre la figure ci-dessous. La distribution obtenue présente des oscillations et elle comporte des valeurs négatives.

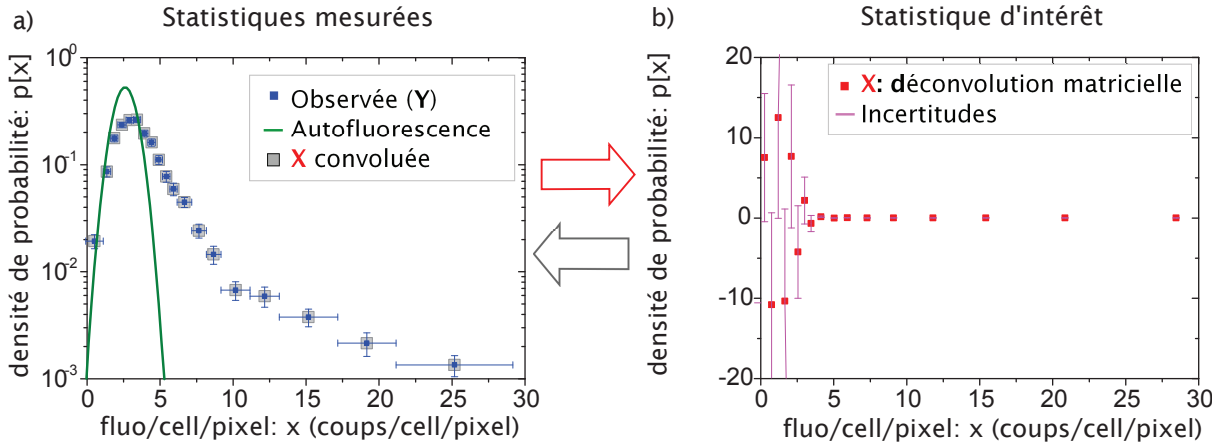


Figure III.3.5.: a) Statistiques mesurées. Distribution d'autofluorescence (courbe verte) et Distribution de la concentration cellulaire de fluorescence tracé à partir de 1857 cellules contenant des rapporteurs plasmidiques révélant l'activité du promoteur de gyrA (points bleus). b) Statistique d'intérêt calculée par déconvolution matricielle. Cette statistique génère par convolution (à l'autofluorescence) la distribution indiquée sur la figure (a) par les points grisés, celle-ci est identique à la distribution bleue..

Pourtant, la méthode matricielle donne la solution exacte de la déconvolution au sens mathématique. Mais par définition même, l'opération de déconvolution génère de grandes incertitudes sur $\vec{X}$: comme l'explique le paragraphe suivant, la déconvolution amplifie les erreurs statistiques.

e. La déconvolution, un problème mal conditionné

Incertitudes sur la distribution observée

L'histogramme observé $\boldsymbol{Y} = [y_1,\ y_2,\ \ldots\ ,\ y_N]$ ne correspond pas rigoureusement à $p_{F_{cell}}$: c'est en fait une estimée de $[p_{F_{cell}}1,\ p_{F_{cell}}2,\ \ldots\ ,\ p_{F_{cell}}N]$. Nous noterons $\boldsymbol{\delta_y}$ l'écart algébrique entre les valeurs estimées $(\boldsymbol{Y})$ et les valeurs exactes associées $(p_{F_{cell}})$:

$$\boldsymbol{\delta} = p_{F_{cell}} - \boldsymbol{Y}$$
$$\forall \text{ bin } B_k\text{: } \delta_k = p_{F_{cell}k} - y_k$$

Incertitudes sur la distribution d'intérêt

Comme l'exprime l'équation III.3.4 précédente, la distribution d'intérêt $\boldsymbol{X} = (x_i)_{1\leq x\leq N}$ se déduit de l'histogramme $\boldsymbol{Y}$ observé par:

$$\forall \text{ bin } b_i,\ x_i = f_i(\boldsymbol{Y}) = \sum_{k=1}^{N} (A^{-1})_{i,k} \times y_k$$

Si on s'intéresse maintenant, non pas à l'estimateur $\boldsymbol{X}$, mais à la loi de probabilité exacte $p_{F_{fp}}$, elle est fonction de $p_{F_{cell}}$ (loi de probabilité observée exacte):

$$\forall \text{ bin } b_i,\ \ p_{F_{fp}i} = f_i(p_{F_{cell}}) = \sum_{k=1}^{N} (A^{-1})_{i,k} \times p_{F_{cell}k}$$

Un développement de Taylor au 1^{er} ordre autour de $\boldsymbol{Y}$ permet alors d'écrire:

$$f_i(p_{F_{cell}}) = f_i(\boldsymbol{Y} + \boldsymbol{\delta}) \overset{Taylor}{=} f_i(\boldsymbol{Y}) + \sum_{k=1}^{N} \delta_k (A^{-1})_{i,k}$$

Cette équation peut être réécrite sous la forme:

$$p_{F_{xfp}i} - x_i = \sum_{k=1}^{N} (A^{-1})_{i,k} \times (p_{F_{cell}k} - y_k)$$

Propagation des erreurs statistiques

Les erreurs statistiques sur $\boldsymbol{Y}$ sont données par la déviation standarde σ_y :

$${\sigma_{y_k}}^2 = E\Big[(p_{F_{cell}k} - y_k)^2\Big]$$

Et les erreurs statistiques sur $\boldsymbol{X}$ sont définies par la covariance:

$$cov(x_i, x_j) = E\Big[(p_{F_{fp}i} - x_i)(p_{F_{fp}j} - x_j)\Big]$$
$$\sigma_{x_i}^2 = cov(x_i, x_i)$$

D'après la relation établie dans le paragraphe précédent, cette covariance vaut:

$$cov(x_i, x_j) = \sum_{k,l} \left(A_{i,k}^{-1} A_{j,l}^{-1} \times E\Big[(p_{F_{cell}k} - y_k)(p_{F_{cell}l} - y_l)\Big] \right)$$

Et comme pour la distribution observée $\boldsymbol{Y}$, le contenu de deux bins voisins n'est pas corrélé, on a:

$$\text{Pour } k \neq l : E\Big[(p_{F_{cell}k} - y_k)(p_{F_{cell}l} - y_l)\Big] = 0$$

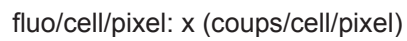
Par conséquent, les erreurs statistiques sur $\boldsymbol{X}$ sont données par:

$$cov(x_i, x_j) = \sum_{k} A_{i,k}^{-1} A_{j,k}^{-1} \sigma_{y_k}^2 \tag{III.3.5}$$

Les termes croisés de la covariance peuvent donc être significatifs, et même alterner de signe: les incertitudes sur les composantes calculées de $\boldsymbol{X}$ sont grandes, et le contenu de deux bins voisins peut être corrélé négativement (voir figure III.3.5 plus loin).
Ainsi, les erreurs statistiques se propagent très bien dans l'équation III.3.4: les valeurs calculées de $\boldsymbol{X}$ sont peu précises. Il s'en suit que la solution mathématique de l'équation III.3.1 n'est pas unique: deux distributions $\boldsymbol{X}$ différentes peuvent générer le même histogramme observé. C'est là tout le problème de la déconvolution!

f. Déconvolution pour une distribution $p_{F_{cell}}$ dénuée d'erreurs statistiques

Les aberrations de la figure III.3.5 sont donc induites par le manque d'information sur $\boldsymbol{Y}$ avec des erreurs statistiques qui sont amplifiées par l'opération de déconvolution.
On peut y remédier si on travaille avec des statistiques exactes, non entâchées par les erreurs statistiques. Dans ce cas, l'histogramme d'intérêt est retrouvé, aux erreurs de calculs près induites par Matlab (voir figure III.3.6 ci-dessous).



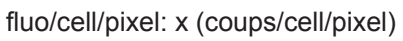


Figure III.3.6.: a) Distribution d'autofluorescence (vert) et distribution de concentration cellulaire de fluorescence (bleu). La distribution observée est simulée en convoluant (méthode matricielle) une distribution originale hypothétique par l'autofluorescence. La distribution hypothétique est ici prise comme une loi gamma de paramètres $a = 2,8$ et $b = 7,7$.
b) Statistique d'intérêt calculée par déconvolution matricielle (ronds magenta). Les erreurs de calculs de Matlab entraînent une dispersion des points pour les concentrations de fluorescence cellulaire élevées. La distribution originale des rapporteurs fluorescents est indiquée par la courbe rouge: la statistique observée (en bleu sur la figure de gauche) a été générée à partir de cette distribution, par convolution.

III.3.2. Régularisation du problème

Nous venons de voir que l'opération de déconvolution amplifie les erreurs, générant des solutions mathématiquement exactes, mais physiquement irrélévantes. Pour obtenir une solution physiquement acceptable, on décide de rajouter des contraintes sur la fonction de déconvolution au détriment du biais: on *régularise* le problème en faisant le choix d'une solution physiquement acceptable parmi celles mathématiquement possibles.

Dans un premier temps, on décide de rajouter les deux contraintes ci-dessous:

- la statistique d'intérêt ne doit pas comporter de valeurs négatives.
- La statistique d'intérêt ne doit pas comporter d'oscillations importantes.

a. Déconvolution itérative

La méthode matricielle ne permet pas de régulariser le problème comme souhaité. Nous nous penchons donc vers une autre méthode qui offre une plus grande liberté sur le choix du critère de régularisation: la déconvolution itérative.

Dans cette méthode, on cherche à estimer l'histogramme d'intérêt: on part d'une distribution supposée $\boldsymbol{X}^0 = (x_i^0)_{i=1:N}$ et on modifie les paramètres x_i^0 pour qu'ils génèrent, par convolution, une statistique qui corresponde mieux à la distribution observée. Les modifications sur les paramètres sont réitérées jusqu'à ce que la statistique calculée transcrive bien la distribution observée.

Cette méthode implique donc de définir:

- Une distribution initiale $\boldsymbol{X}^0$: celle-ci sera prise uniforme.
- Un critère d'erreur qui estime la véracité de la distribution calculée, en terme de correspondance avec la distribution observée: $e(\boldsymbol{Y}, A\boldsymbol{X})$.
- Un critère de régularisation qui contraint le problème à générer des solutions physiquement acceptables: $r(\boldsymbol{X})$.

Dans notre étude, la déconvolution itérative est réalisée avec l'algorithme de Lucy-Richardson [1, 2]: lorsque les erreurs statistiques sur $\boldsymbol{Y}$ sont poissonniennes, ce dernier converge vers la solution mathématique donnée par l'équation III.3.4. Dans nos distributions, les erreurs statistiques sur $\boldsymbol{Y}$ sont binomiales et elles peuvent être approximées comme poissonniennes: l'algorithme de Lucy-Richardson permet donc de déconvoluer l'autofluorescence du signal observé.

a.1. Algorithme de Lucy-Richardson [58, 59]

Dans la méthode présentée ici, on considère que l'histogramme d'intérêt correspond à l'histogramme observé lorsque lorsque la fonction de vraisemblance $L(A\boldsymbol{X} \mid \boldsymbol{Y})$ est maximale: la probabilité, que $\boldsymbol{X}$ génère par convolution la distribution observée, est maximale. Or, d'après le théorème de Bayes:

$$L(A\boldsymbol{X} \mid \boldsymbol{Y}) = \frac{L(\boldsymbol{Y} \mid A\boldsymbol{X}) \times L(A\boldsymbol{X})}{L(\boldsymbol{Y})}$$

Si $L(\boldsymbol{Y})$ est constante, maximiser $L(A\boldsymbol{X} \mid \boldsymbol{Y})$ revient à maximiser $L(\boldsymbol{Y} \mid A\boldsymbol{X})$. Notre critére d'erreur sera donc:

$$e(\boldsymbol{Y}, A\boldsymbol{X}) = L(\boldsymbol{Y} \mid A\boldsymbol{X})$$

Ainsi, à chaque itération r, on corrigera les x_i^r pour maximiser $e(\boldsymbol{Y}, A\boldsymbol{X}^r)$.
Pour ce faire, on utilise l'algorithme de Lucy-Richardson défini ci-dessous:

$$x_i^{r+1} = x_i^r \times \sum_{m=1}^{N} \left(\frac{y_m}{(A\boldsymbol{X}^r)_m} \times A_{m,i} \right)$$

L'algorithme est arrété lorsque la statistique générée $\boldsymbol{X}^{r+1}$ ressemble à la statistique $\boldsymbol{X}^r$ au pas précédent; on décide que ces deux statistiques se ressemblent lorsqu'elles diffèrent de moins de $0,1\%$. La différence est caractérisée par le χ^2:

$$\chi^2\left(\boldsymbol{X}^{r+1}, \boldsymbol{X}^r\right) = \sum_{i=1}^{N} \frac{(x_i^{r+1} - x_i^r)^2}{x_i^r}$$

Cette méthode présente l'avantage de ne pas générer de valeurs négatives pour l'histogramme d'intérêt, comme on peut le voir sur la courbe vert de la figure III.3.7ci-dessous.

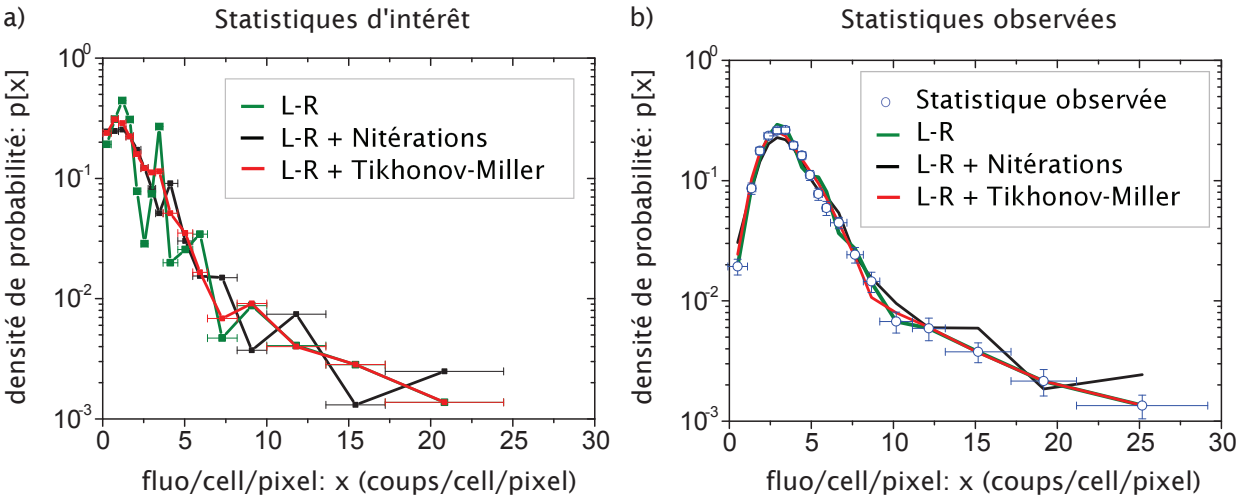


Figure III.3.7.: (a) Statistiques d'intérêt calculées par déconvolution itérative: algorithme de Lucy-Richardson. En rouge, l'algorithme est arrété lorsque $\vec{X}^{k+1}$ diffère de moins de 0.1% de $\vec{X}^k$ (L-R). En noir, l'algorithme est arrété lorsque lorsque le χ^2 entre $A\vec{X}$ et $\vec{Y}$ augmente entre deux itérations (L-R +Nitérations). La courbe verte est obtenue dans les même conditions que la courbe noire, avec un critère de régularisation supplémentaire sur la courbure: la régularisation de Tikhonov-Miller avec force de régularisarion $r = 0.01$ (L-R+Tikhonov-Miller). (b) Statistiques observées correspondantes: elles sont obtenues par convolution des statistiques d'intérêt donnés dans le graphique du dessus. La gaussienne de convolution utilisée peut être visualisée en début de chapitre, dans la figure III.3.1.

Mais nous retrouvons le même problème inhérent à la déconvolution: les incertitudes sur $\boldsymbol{X}$ sont amplifiées lorsque l'algorithme converge et se rapproche de la solution mathématique. Pour supprimer les oscillations résultantes, il va falloir rajouter des contraintes.

a.2. Régularisation sur le nombre d'itérations

Comme les incertitudes sont amplifiées à chaque itération, le problème peut être régularisé en stoppant l'algorithme avant que les oscillations apparaissent.
On décide de stopper l'algorithme dès que l'itération génère une distribution $A\boldsymbol{X}$ qui diffère de $\boldsymbol{Y}$ plus qu'elle ne le faisait au pas précédent. La différence est quantifiée par le χ^2:

$$\text{Si } \chi^2(A\boldsymbol{X}^{r+1}, \boldsymbol{Y}) > \chi^2(A\boldsymbol{X}^r, \boldsymbol{Y}) \text{ alors, l'algorithme est arrêté.}$$

Le χ^2 est défini ci-dessous: nous l'avons pondéré des incertitudes statistiques, pour accorder plus de poid aux bins précis de la distribution observée.

$$\chi^2(A\boldsymbol{X}^r, \boldsymbol{Y}) = \sum_{i=1}^{N} \left(\frac{\left(y_i - (A\boldsymbol{X}^r)_i\right)^2}{y_i} \times \frac{1}{\triangle_{Bi}\sigma_{yi}} \right)$$

Le résultat peut être visualisé sur la figure III.3.7 précédente. Les oscillations sont moins marquées, mais la distribution obtenue n'est toujours pas satisfaisante.

a.3. Régularisation de Tikhonov Miller

Une autre manière de régulariser le problème est d'agir directement sur la courbure de la statistique d'intérêt. Pour ce faire, on rajoute un terme de régularisation dans le critère d'erreur:

$$e(\boldsymbol{Y}, A\boldsymbol{X}) = L(\boldsymbol{Y} \mid A\boldsymbol{X}) \times \frac{1}{Regu(\boldsymbol{X}_i^r)}$$

Le terme de régularisation introduit par Tikhonov et Miller permet de filtrer les hautes fréquences, supprimant ainsi les courbures trop brutales:

$$Regu(\boldsymbol{X}) = 1 + r\left(\frac{d^2\boldsymbol{X}}{d^2 b_i}\right)$$

où la constante r définit la force de la régularisation, et la courbure est calculée comme suit:

$$\left(\frac{d^2\boldsymbol{X}}{d^2 b_i}\right) = \left(\frac{x_{i+1} - x_i}{\triangle_{b_{i+1}} + \triangle_{b_i}} - \frac{x_i - x_{i-1}}{\triangle_{b_i} + \triangle_{b_{i-1}}} \right) \times \frac{1}{\triangle_{b_{i+1}} + \triangle_{b_i}}$$

L'algorithme qui permet de réaliser cette contrainte est donné par:

$$x_i^{r+1} = x_i^r \times \sum_{m=1}^{N} \left(\frac{y_m}{(A\boldsymbol{X}^r)_m} \times A_{m,i} \times \frac{1}{Regu(\boldsymbol{X}_i^r)} \right)$$

III.3.3. Conclusion

Il existe d'autres méthodes de déconvolution qui ne s'appuient pas sur un "binage" présupposé de la statistique d'intérêt. Citons par exemple la méthode satellite qui paraît bien adaptée. Mais ces méthodes sont plus compliquées à mettre en oeuvre, et la déconvolution Lucy-Richardson avec régularisation de Tikhonov-Miller fournit déjà une idée sur l'allure de la statistique. A partir de là, on peut formuler des hypothèses sur la statistique d'intérêt: on suppose que celle-ci est décrite par une loi de probabilité adaptée. Et on teste la validité de l'hypothèse en convoluant la distribution d'intérêt hypothétique parc l'autofluorescence. Si l'hypothèse est valide, le résultat de la convolution doit décrire notre distribution observée. Nous verrons dans la section section IV.3.2, que l'hypothèse qui formule la statistique d'intérêt en terme de distribution gamma marche relativement bien.

On souligne toutefois que cette méthode ne fournit qu'une solution de l'équation de déconvolution: bien d'autres distributions peuvent être solution. De plus, notre méthode de déconvolution est biaisée par la régularisation, et elle perd l'information sur les barres d'erreurs statistiques. En conséquence, les paramètres de la distribution d'intérêt hypothétique sont fortement incertains. Qui plus est, pour tester une loi de probabilité d'intérêt, il faut la convoluer avec l'autofluorescence: cela suppose de calculer la matrice d'autofluorescence A, et donc de "biner" la loi de probabilité testée. Or le résultat de la déconvolution dépend fortement du binage choisi pour la loi de probabilité d'intérêt.

Pour toutes ces raisons, nous n'aurons recourt à la déconvolution que lorsque ceci sera absolument nécessaire.

Dans toutes nos expériences, le rendement quantique des rapporteurs fluorescents est plus important que celui des fluorophores naturels: à concentration égale, la fluorescence naturelle est négligeable devant celle des rapporteurs. Dans ces conditions, la déconvolution n'est nécessaire que lorsque la concentration des rapporteurs est faible devant celle des fluorophores naturels: dans ce cas, les deux signaux deviennent comparables et la fluorescence des rapporteurs est bruitée par l'autofluorescence.

- Pour les rapporteurs plasmidiques: les concentrations de rapporteurs sont élevées, et nous verrons que la déconvolution est rarement nécessaire. Elle ne l'est que lorsque les proteines fluorescentes se trouvent en faible quantité dans la cellule. Dans nos expériences, cela correspond à des cellules en état de stress nutritionnel.
- Pour les rapporteurs chromosomaux: les signaux détectés sont faibles. Et les fluorophores naturels sont particulièrement sensibles aux illuminations sous module réflecteur *cyan*. La déconvolution est donc nécessaire lorsqu'on s'intéresse aux rapporteurs *cfp*: seuls les histogrammes tracés pour des cellules qui s'approchent de la phase exponentielle ne nécessitent pas de déconvolution. En ce qui concerne le mod-

ule réflecteur *yellow*, l'autofluorescence est quasi nulle. Donc pour les rapporteurs *yfp*, il n'est pas nécessaire de déconvoluer l'autofluorescence. Cependant, lorsque le niveau de fluorescence est très faible, le bruit du fond pourra influencer la forme de notre distribution. Dans ce cas, il faudra déconvoluer non pas l'autofluorescence, mais le bruit du fond: les méthodes utilisées sont les mêmes que celles présentées pour l'autofluorescence.

III.4. Simulations

Nous avons vu en Introduction (chapitre I) que la stochasticité d'expression génique pouvait être modélisée par un ensemble de réactions probabilistes. Ces réactions formulent l'équation maîtresse. Celle-ci a été résolue dans des cas particuliers ([7, 10] : voir la section I.1.3). On obtient alors, pour chaque temps t, la loi de probabilité $P(n,t)$ associée au nombre n de protéines par cellule. Mais bien souvent, on ne sait pas résoudre l'équation maîtresse. Pour simplifier les calculs, on décide alors d'ajouter des contraintes : on se place à l'équilibre statistique, et on restreint l'équation aux moments du premier et du deuxième ordre, à savoir $\langle n \rangle$ et σ_n^2. Ainsi, J.Paulsson[4] montre qu'à l'équilibre statistique, les fluctuations statiques sont caractérisées par un niveau de bruit à :

$$\eta_{int}^2 = \frac{\sigma_n^2}{\langle n \rangle^2} = \frac{1}{\langle n \rangle} + \frac{1}{\langle m \rangle} \times \frac{1}{1 + (\gamma_m / \gamma_p)}$$

Le terme en $1/\langle n \rangle$ traduit les mécanismes stochastiques de naissance et de mort des protéines, et le terme en $1/\langle m \rangle$ reflète le bruit de transcription qui se propage sur les protéines, via un filtre passe-bas de fréquence de coupure γ_m/γ_p.

Lorsqu'on veut décrire la variabilité inter-cellulaire de manière plus complète, avec l'évolution dynamique de la distribution $P(n,t)$, on délaisse les équations pour lancer des simulations qui miment les réactions d'expression génique. Ces simulations présentent aussi l'avantage de pouvoir étudier des modèles plus complexes, où par exemple l'expression génique intègre le processus de maturation protéique.

Dans cette étude, les modèles proposés ont été étudiés à travers des simulations. Celles-ci sont effectuées avec l'algorithme de Gillespie [13], et son extension qui inclut les variations temporelles des taux de réaction [11].

III.4.1. L'algorithme de Gillespie

Cet algorithme s'appuie sur les hypothèses suivantes :

H1 : L'état d'une cellule est décrit à tout instant par ses quantités $(x_1, x_2, ..., x_M)$ de molécules X_1,X_2, ...,X_M.

H2 : L'une de ces quantités est modifiée par pas discret de 1 lorsqu'une réaction se produit. La cellule évolue alors vers un autre état $(x'_1, x'_2, ..., x'_M)$.

H3 : Chaque réaction Ri d'expression génique est instantanée lorsqu'elle se produit.

H4 : Les réactions Ri sont décrites par des taux c_i probabilistes constants dans le temps. Ces taux sont conditionnés par la quantité de réactants associés à la molécule X_i et par la force de l'interaction X_i-réactant. Par exemple, le taux de traduction k_p est conditionné par le nombre de ribosomes et d'ARNt, ainsi que par les propriétés physiques de l'interaction ribosome-ARNm. Ainsi, la probabilité qu'une molécule X_i réagisse pendant dt est $c_i dt$.

H5 : Ces taux de réaction sont homogènes dans l'espace. Donc si la cellule contient x_i molécules X_i à l'instant t, la probabilité que la réaction R_i se produise pendant l'intervalle de temps $[t;\ t+dt]$ vaut :

$$\omega_i dt = (c_i \times X_i)dt$$

H6 : Deux réactions successives R_i sont indépendantes. Donc, comme les taux de réaction sont constants, les réactions suivent des processus de Poisson de paramètres ω_i : la probabilité $P(\tau, R_i)$ que l'intervalle de temps τ sépare deux réactions R_i consécutives est donnée par :

$$P(\tau, R_i) = e^{-\omega_i \tau}$$

H7 : Les réactions $(R_1, R_2, ..., R_M)$ sont indépendantes. Par conséquent, la somme des réactions suit aussi un processus poissonnien, de paramètre $\omega = \sum_{i=1}^{M} \omega_i$. Ainsi, la probabilité $P(\tau)$ que l'intervalle de temps τ sépare deux réactions consécutives (quelles qu'elles soient) est donnée par :

$$P(\tau) = e^{-\omega \tau}$$

A partir de ces hypothèses, l'agorithme simule pour une cellule l'évolution temporelle des variables $(X_1, X_2, ..., X_M)$: il trace une trajectoire $(x_1, x_2, ..., x_M)(t)$.
Pour simuler une trajectoire, l'algorithme effectue 4 étapes en boucle : ces étapes miment l'occurrence aléatoire d'une réaction, et elles sont détaillées dans ce qui suit.

0. L'algorithme part d'un état initial $(x_1, x_2, ..., x_M)_0$ imposé par l'utilisateur.
1. Puis il calcule les probabilités ω_i de réaction par unité de temps, associées à cet état. Il en déduit ω :

$$\omega = \sum_{i=1}^{M} \omega_i$$

2. Il génère ensuite le temps τ au bout duquel une des réactions $(R_1, R_2, ..., R_M)$ se produit. Les hypothèses impliquent que ce temps est aléatoire et qu'il doit être généré selon une distribution exponentielle de paramètre ω. Il peut donc être facilement généré à partir d'un générateur uniforme standard : on génére aléatoirement un nombre r_1 selon une distribu-

tion uniforme U[0, 1], et on en déduit le temps τ par la relation :

$$\tau = \frac{1}{\omega}\ln(1/r_1)$$

3. Au temps τ, une réaction se produit : Laquelle ?
Pour déterminer de quelle réaction il s'agit, l'algorithme calcule, pour chaque réaction R_j, la probabilité que ce soit elle qui se produise : celle-ci est donnée par $p(R_i) = \omega_i/\omega$. L'algorithme divise alors l'intervalle $[0, 1]$ en M intervalles de longueur $p(R_i)$, et il génère un nombre aléatoire r_2 selon U[0, 1]. Il regarde ensuite dans quelle partie de l'intervalle r_2 tombe : cet intervalle détermine la réaction R_J qui se produit.
4. La réaction R_J se produit : le temps est implémenté de τ, et la variable X_J est incrémentée de 1 (ou de -1 pour une réaction de dégradation). L'algorithme définit ainsi un nouvel état cellulaire $(x_1, x_2, ..., x_M)(t + \tau)$.
Il recommence alors les étapes 1 à 4, et ce, jusqu'à ce qu'il atteigne le temps t_{max} fixé par l'utilisateur. Tout ceci est illustré sur la figure III.4.1 ci-contre.
L'algorithme a été codé en C++ sous l'éditeur Notepad+, et il a été compilé avec MingW.

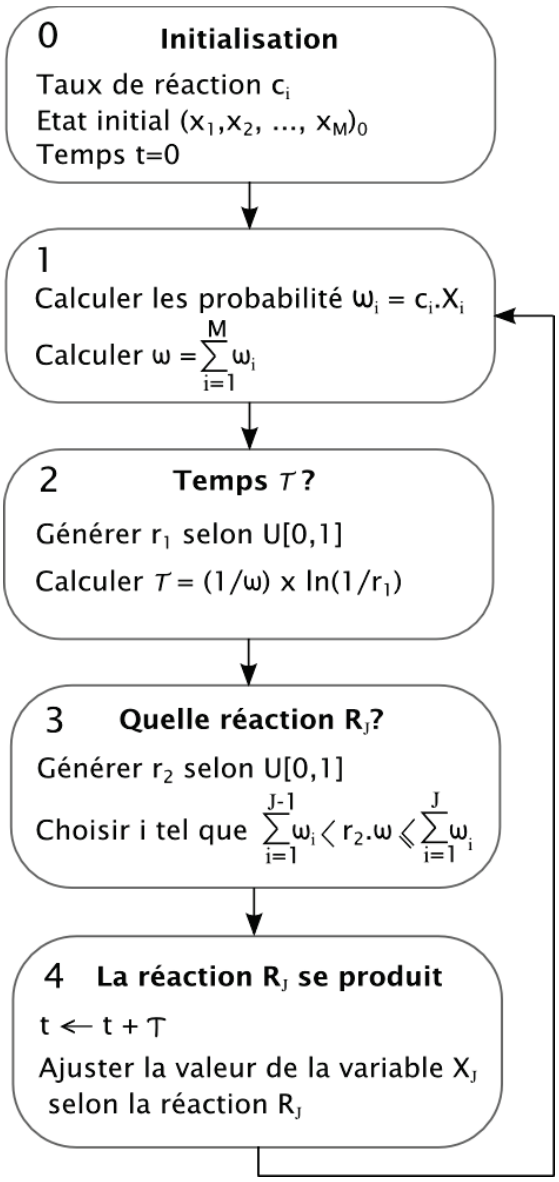


FIGURE III.4.1.: Algorithme de Gillespie.

Vérification du code

Pour vérifier la validité de notre code, nous avons étudié le modèle simple de Thattai et al., à travers les simulations, et à travers des calculs analytiques : les données simulées sont comparées avec les expressions analytiques.
Cette comparaison a été faite sur la variable n, avec la dynamique du moment d'ordre 1 et la statistique entière pour quelques temps.

Expression analytique de $\langle n \rangle(t)$ et de $P_n(t)$

Nous vous invitons à revoir le modèle de Thattai et al. présenté dans le Chapitre 1, et nous rappelons que pour ce modèle, l'équation maîtresse s'écrit :

$$\frac{dP_{m,n}}{dt} = \gamma_m\Big((m+1)P_{m+1,n} - mP_{m,n}\Big) + k_m\Big(P_{m-1,n} - P_{m,n}\Big) + \gamma_p\Big((n+1)P_{m,n+1} - nP_{m,n}\Big) + k_p\Big(mP_{m,n-1} - mP_{m,n}\Big)$$

Expression de $P_n(t)$: Pour simplifier, on considére qu'initialement (à $t = 0$) : $\langle m\rangle_0 = 0$, $\langle n\rangle_0 = 0$, et $\sigma^2_{n0} = 0$. De plus, on choisit $\gamma_m \gg \gamma_p$.
Dans ces conditions, nous pouvons utiliser les travaux de V.Shahrezaei et al. présentés dans l'introduction : ces théoritiens ont résolu l'équation maîtresse pour $t \times \gamma_m \gg 1$, et ils donnent l'expression mathématique de $P_n(t)$ [7].

Expression de $\langle n\rangle(t)$: Pour connaître l'expression de $\langle n\rangle$ pour tous les temps (et pas seulement pour $t \times \gamma_m \gg 1$), on résoud l'équation maîtresse pour :

$$\langle n\rangle(t) = \sum_{n\geq 0} n \sum_{m\geq 0} P_{m,n}(t)$$

Comme les réactions sont considérées indépendantes, on a :

$$P_{m,n}(t) = (P_m \times P_n)(t)$$

L'équation maîtresse donne alors :

$$\frac{d\langle n\rangle(t)}{dt} = \Big(k_p\langle m\rangle - \gamma_p\langle n\rangle\Big)(t)$$

où le nombre moyen d'ARNm par cellule $\langle m\rangle$ obéit à :

$$\frac{d\langle m\rangle(t)}{dt} = k_m - \gamma_m\langle m\rangle(t)$$
$$\Rightarrow \langle m\rangle(t) = \frac{k_m}{\gamma_m}(1 - e^{-\gamma_m t})$$

En résolvant l'équation différentielle sur $\langle n\rangle(t)$, on obtient un nombre moyen de protéines par cellule qui évolue selon :

$$\langle n\rangle(t) = \frac{k_p k_m}{\gamma_m}\left[\frac{1}{\gamma_p}\Big(1 - e^{-\gamma_p t}\Big) - \frac{1}{\gamma_m - \gamma_p}\Big(e^{-\gamma_m t} - e^{-\gamma_p t}\Big)\right]$$

Simulations

Nous avons simulé 100 000 trajectoires pour le modèle de Thattai et al. qui contient deux variables : l'état d'une cellule est décrit à tout instant par ses quantités n et m. Sur chacune des trajectoires k simulée, nous avons extrait $n_k(t)$. Puis nous avons moyenné cette quantité sur toutes les trajectoires : on obtient $\langle n\rangle(t)$. Nous avons aussi tracé, pour quelques temps, la distribution de $n_k(t)$ dans la population de trajectoires. Cette distribution est tracée sous Matlab à partir des données simulées : la taille de bin est fixée constante, égale à 3.

Comparaison des simulations avec les expressions analytiques

Les résultats obtenus sont présentés dans la figure suivante avec, en continu les courbes théoriques tirées de l'équation maîtresse, et en points les données simulées.

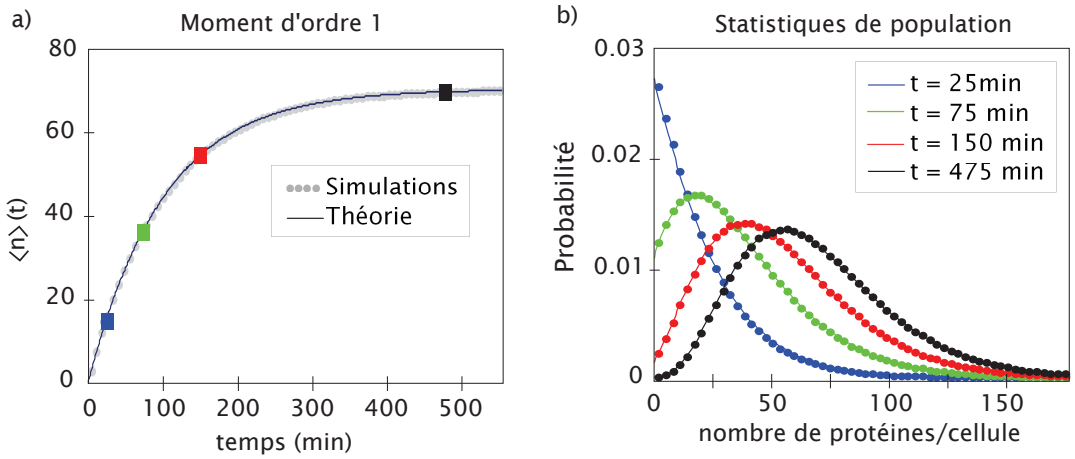


FIGURE III.4.2.: Comparaison sur le modèle de Thattai et al. : simulations (points) Versus formules mathématiques (lignes continues). La variable d'étude est la quantité de protéines par cellule (n) et les paramètres du modèle sont fixés à : $\gamma_m = 0.9$ min^{-1}, $\gamma_p = 0.01$ min^{-1}, $k_m = 0.05$ min^{-1}, $k_p = 12.6$ protéines.min^{-1}. a) Comparaison sur la dynamique du moment d'ordre 1. Les points colorés donnent les temps pour lesquels la statistique est tracée. b) Comparaison sur toute la statistique.

Les données simulées s'accordent très bien avec les formules analytiques : notre algorithme de simulation est correct, et il pourra être amélioré pour étudier des modèles plus complexes.

III.4.2. Extension : taux de réaction qui varient dans le temps

L'algorithme de Gillepsie est conçu pour décrire des phénomènes où les taux de réaction c_i restent constants dans le temps. Or, dans nos expériences en dynamique, ces taux sont suceptibles de varier au cours du temps [1]. Dans ce cas, les probabilités de réaction ω_i associées à un état dépendent du temps, et dans l'étape 2 de l'algorithme, τ ne peut plus être généré par la méthode de Gillespsie.
Pour intégrer les variations des taux c_i dans le calcul de τ, nous implémentons l'extension de l'algorithme proposée par V.Shahrezaei et al. [11].

1. Par exemple, les taux de dégradation varient lors du passage d'un état de stress nutritionnel à un état de croissance exponentielle : les composants cellulaires sont de plus en plus dégradés par la dilution cellulaire.

Hypothèses

Dans cette extension, les réactions sont décrites par des taux c_i probabilistes qui varient dans le temps. Ceci implique de reformuler les hypothèses H5, H6 et H7 :

H5 : Les taux de réaction c_i sont homogènes dans l'espace. Donc, si la cellule contient x_{i0} molécules X_i à l'instant t_0, la probabilité que la réaction R_i se produise pendant l'intervalle de temps $[t_0;\ t_0 + dt]$ vaut :

$$\omega_i(t_0)dt = (c_i(t_0) \times x_{i0})dt$$

H6 : Deux réactions successives sont indépendantes. Et comme les taux de réaction varient au cours du temps, la probabilité $P(\tau_i, R_i, t_0)$ que la prochaine réaction R_i se produise à $t_0 + \tau_i$ est maintenant donnée par :

$$P(\tau_i, R_i, t_0) = \omega_i(t_0 + \tau_i) \times P_{\bar{R}_i,t_0}(t_0 + \tau_i)$$

où $P_{\bar{R}_i,t_0}(t)$ est la probabilité qu'aucune réaction R_i ne se produise pendant $[t_0\ t]$.
Si au temps t_0 la cellule contient x_{i0} réactants X_i, alors cette probabilité $P_{\bar{R}_i}(t)$ vérifie la relation suivante :

$$P_{\bar{R}_i,t_0}(t + dt) = P_{\bar{R}_i,t_0}(t) \times \big(1 - \omega_i(t)dt\big)$$

On a donc :

$$P(\tau_i, R_i, t_0) = -\left.\frac{dP_{\bar{R}_i}(t)}{dt}\right|_{t_0+\tau_i}, \text{ avec } P_{\bar{R}_i}(t) = \exp\Big(-\int_{t_0}^{t}\omega_i(t')dt'\Big) \qquad \text{(III.4.1)}$$

H7 : les réactions $(R_1, R_2, ..., R_M)$ sont indépendantes. Donc la probabilité qu'une des réactions $(R_1, R_2, ..., R_M)$ se produise pendant l'intervalle de temps $[t_0;\ t_0 + dt]$ vaut :

$$\omega(t_0)dt = \sum_{i=1}^{M}\omega_i(t_0)dt$$

Si au temps t_0 la cellule est dans l'état $(x_{10}, x_{20}, ..., x_{M0})$, alors la probabilité $P(\tau, R, t_0)$ que la prochaine réaction (quelqu'elle soit) se produise à $t_0 + \tau$ est donnée par [2] :

$$P(\tau, R, t_0) = -\left.\frac{dP_{\bar{R}}(t)}{dt}\right|_{t_0+\tau}, \text{ avec } P_{\bar{R}}(t) = \exp\Big(-\int_{t_0}^{t}\omega(t')dt'\Big)$$

où $P_{\bar{R}}(t)$ est la probabilité qu'aucune des réactions $(R_1, R_2, ...R_M)$ ne se produise pendant $[t_0\ t]$.

2. Ce résultat s'obtient avec un raisonnement analogue à celui formulé dans l'hypothèse H6.

Algorithme

Le nouvel algorithme s'effectue toujours en 4 étapes qui gardent la même vocation que dans l'algorithme de la figure III.4.1 précédente. Mais la manière dont ces étapes sont implémentées change.

Implémentation de l'étape 4 dans la version améliorée de l'algorithme de Gillespie[11]

Soit $(x_{10}, x_{20}, ..., x_{M0})$ *l'état de la cellule au temps* t_0. *Quel est le temps* τ *au bout duquel une des réactions* $(R_1, R_2, ..., R_M)$ *se produit ?*

L'hypothèse H7 implique que ce temps doit être généré aléatoirement selon la distribution $P(\tau, R, t_0)$. D'après D.T.Gillespie cette opération revient à générer aléatoirement les temps τ_i associés à la prochaine occurence des réactions R_i , et à choisir le plus petit de ces temps [3] :

$$\tau = \min_i(\tau_i)$$

Pour chaque réaction R_i, l'hypothèse H6 implique que le temps τ_i doit être généré selon la distribution $P(\tau_i, R_i, t_0)$ de l'équation III.4.1. Ceci peut être réalisé à partir d'un générateur uniforme standard : on génére aléatoirement un nombre r selon une distribution uniforme U[0 1], et on en déduit τ_i selon [4] :

$$\int_0^{\tau_i} P(\tau_i', R_i, t_0) d\tau' = r$$

D'après la relation III.4.1, cela revient à générer aléatoirement un nombre r selon U[0 1], et à déterminer le temps τ_i qui vérifie :

$$\int_{t_0}^{t_0+\tau_i} \omega_i(t') dt' = ln(1/r)$$

Comme $\omega_i(t) = x_{i0} \times c_i(t)$, cette équation peut être résolue en résolvant :

$$\int_{\boldsymbol{t_0}}^{\boldsymbol{t_0+\tau_i}} \boldsymbol{c_i(t')dt'} = \frac{\boldsymbol{ln(1/r)}}{\boldsymbol{x_{i0}}} \qquad \text{(III.4.2)}$$

3. Parmi toutes les réactions $(R_1, R_2, ..., R_M)$ possibles, la prochaine des réactions à avoir lieu sera celle qui se produit en premier (*i.e.* la réaction pour laquelle τ_i est le plus petit.

4. La variable aléatoire T_i est distribuée selon $P(\tau_i, R_i, t_0)$.

Soit r_i une variable aléatoire qui s'exprime en fonction de τ_i selon : $r_i(T_i) = \int_0^{\tau_i} P(\tau_i', R_i, t_0) d\tau_i'$.

Avec cette définition de r_i, la méthode de transformation des variables aléatoires implique que r_i est distribuée uniformément sur [0 1].

Quelles sont les solutions de cette équation? Ces solutions se trouvent simplement lorsque le taux de réaction c_i reste constant dans le temps. Dans ce cas, on retrouve l'expression de Gillespie:

$$\tau_i = \frac{1}{x_{i0}c_i} \times ln(1/r)$$

Mais l'équation devient plus difficile à résoudre lorsque le taux de réaction c_i varie dans le temps...

Dans ce cas, la fonction $c_i(t)$ est approximée par une fonction affine par morceaux [8] : $\forall t_0$, sur $[t_0,\ t_0 + \triangle t]$ la fonction est approchée par son développement de Taylor au 1^{ier} ordre :

$$\forall t_0,\ \forall t \in [t_0,\ t_0 + \triangle t],\quad c_i(t) \simeq c_i(t_0) + (t - t_0) \times c_i'(t_0)$$

où $c_i'(t_0)$ est la dérivée première de la fonction $c_i(t)$ en t_0.

On considère que cette approximation est valide lorsque le prochain terme du développement de Taylor est 10 fois plus petit que le terme d'ordre 0. Ceci implique de correctement choisir l'intervalle de temps $\triangle t$. A titre d'exemple, si $c_i(t)$ est un polynôme P_4 d'ordre 4, $\triangle t$ est fixé à :

$$\triangle t = \min_t \left(\sqrt{\frac{1}{5} \frac{P_4(t)}{P_4''(t)}} \right)$$

Nous donnons une illustration de l'approximation affine par morceaux sur la figure III.4.3.

Dans l'exemple montré, le taux de réaction c_i varie dans le temps de manière polynômiale (polynôme d'ordre 4). L'approximation affine par morceaux est représentée en noir. La longueur des morceaux $\triangle t$ est fixe et elle choisie selon:

$$\triangle t = \min_t \left(\sqrt{\frac{1}{5} \frac{c_4(t)}{c_4''(t)}} \right)$$

Et comme le montre l'encadré orange (zoom des deux courbes sur le morceau où l'approximation affine est la plus critique), avec ce choix de $\triangle t$, l'approximation affine décrit relativement bien la fonction $c_i(t)$ (représentée en bleue).

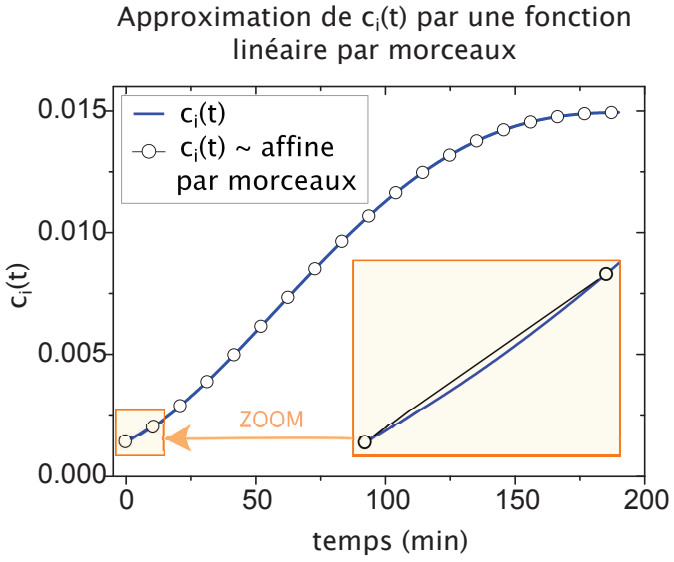


FIGURE III.4.3.: En bleu, les variations temporelles exactes du taux de réaction c_i: $c_i(t) = a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0$, où les coefficients a_0, a_1, a_2, a_3 et a_4 valent (resp.) $1,45.10^{-3}$, $4,5.10^{-5}$, $1,34.10^{-6}$, -1.10^{-8} et $1,93.10^{-11}$. En noir, ces variations sont approximées par une fonction affine par morceaux.

En approximant $c_i(t)$ par une fonction affine par morceaux, l'équation III.4.2 se simplifie :

Résolution de l'équation III.4.2 pour un taux de réaction dépendant du temps

- **SI** r est tel que $\left(\tau_i \leq \triangle t\right)$, **ALORS**:

$$\tau_i = \frac{c_i(t_0)}{c_i'(t_0)} \times \left(\sqrt{1 + \frac{2c_i'(t_0)}{x_{i0}c_i(t_0)^2} \times ln(1/r)} - 1 \right) \qquad \text{(III.4.3)}$$

avec $\tau_i = \infty$, si $\left(a < 0 \text{ et } r_i \geq e^{-\frac{x_{i0}(at_0+b)}{2a}} \right)$

- **SI** r_i est tel que $\left((n+1)\triangle t \geq \tau_i > n\triangle t\right)$ [5], **ALORS** l'équation III.4.2 se simplifie à [6] :

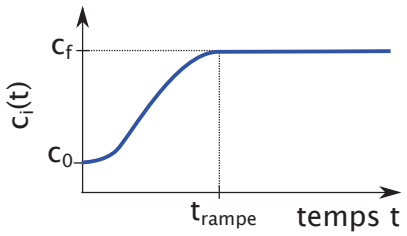
$$\int_{t_{0+n\triangle t}}^{t_0+n\triangle t+\tau} c_i(t')dt' = \frac{ln(1/r_3)}{x_{i0}}$$

où r_3 est un nouveau nombre aléatoire qui suit une distribution uniforme sur $[0,\ 1]$. Dans ces conditions, si on incrémente t_0 de $n\triangle t$, le laps de temps τ_i se calcule comme dans l'équation III.4.3 précédente.

Dans nos simulations, on considérera que les taux varient sur une durée limitée: lorsque l'équilibre cellulaire est atteint, ils deviennent constants.
Ainsi pour chaque réaction R_i, le taux c_i varie sur $[0\ \ t_{rampe,i}]$ pour atteindre sa valeur finale c_f en $t_{rampe,i}$.

$\forall t \in [t_0,\ t_0 + \triangle t]$, la fonction $c_i(t)$ s'écrira :

$$c_i(t) = \begin{cases} c_i(t_0) + (t - t_0) \times c_i'(t_0) & \text{si } t < t_{rampe,i} \\ c_f & \text{sinon} \end{cases}$$



Dans ces conditions, l'équation III.4.2 se résout comme indiqué dans l'encadré suivant.

Résolution de l'équation III.4.2 pour un taux de réaction qui varie sur $[0,\ t_{rampe}]$

- **SI** $(t_0 > t_{rampe})$, **ALORS** l'équation III.4.2 se réduit à:

$$\tau_i = \frac{1}{x_{i0}c_f} \times ln(1/r)$$

- **SI** $(t_0 < t_{rampe})$, **ALORS** l'équation III.4.2 nécessite de comparer τ_i à t_{rampe}. Dans ce but, on calcule la grandeur I_{rampe} définie ci-dessous :

$$I_{rampe} = \int_{t_0}^{t_{rampe}} c_i(t_0) + (t - t_0)c_i'(t_0)dt$$

  - **SI** $\left(x_{i0}I_{rampe} > ln(1/r)\right)$, **ALORS** τ_i est inférieur à t_{rampe} :
    - **SI** $\left(\tau_i \leq \triangle t\right)$. L'équation III.4.2 se réduit à l'équation III.4.3:

$$\tau_i = \frac{c_i(t_0)}{c_i'(t_0)} \times \left(\sqrt{1 + \frac{2c_i'(t_0)}{x_{i0}c_i(t_0)^2} \times ln(1/r)} - 1\right)$$

    - **SI** $\left((n+1)\triangle t > \tau_i \geq n\triangle t\right)$ [7] . On incrémente t_0 de $n\triangle t$, et on génére un nouveau nombre r_2 selon U[0 1]: le laps de temps τ_i se déduit de r_2 par la relation précédente.

  - **SI** $\left(x_{i0}I_{rampe} > ln(1/r_i)\right)$, **ALORS** τ_i est supérieur à t_{rampe}.
    Dans ce cas, l'équation III.4.3 se réduit à:

$$\tau_i = \frac{ln(1/r) - x_{i0}I_{rampe}}{x_{i0}c_f} + (t_{rampe} - t_0)$$

Implémentation de l'algorithme de Gillespie pour des taux qui dépendent du temps

Les différentes étapes de la version revue de l'algorithme de Gillespie sont récapitulées dans la figure III.4.4 qui suit.

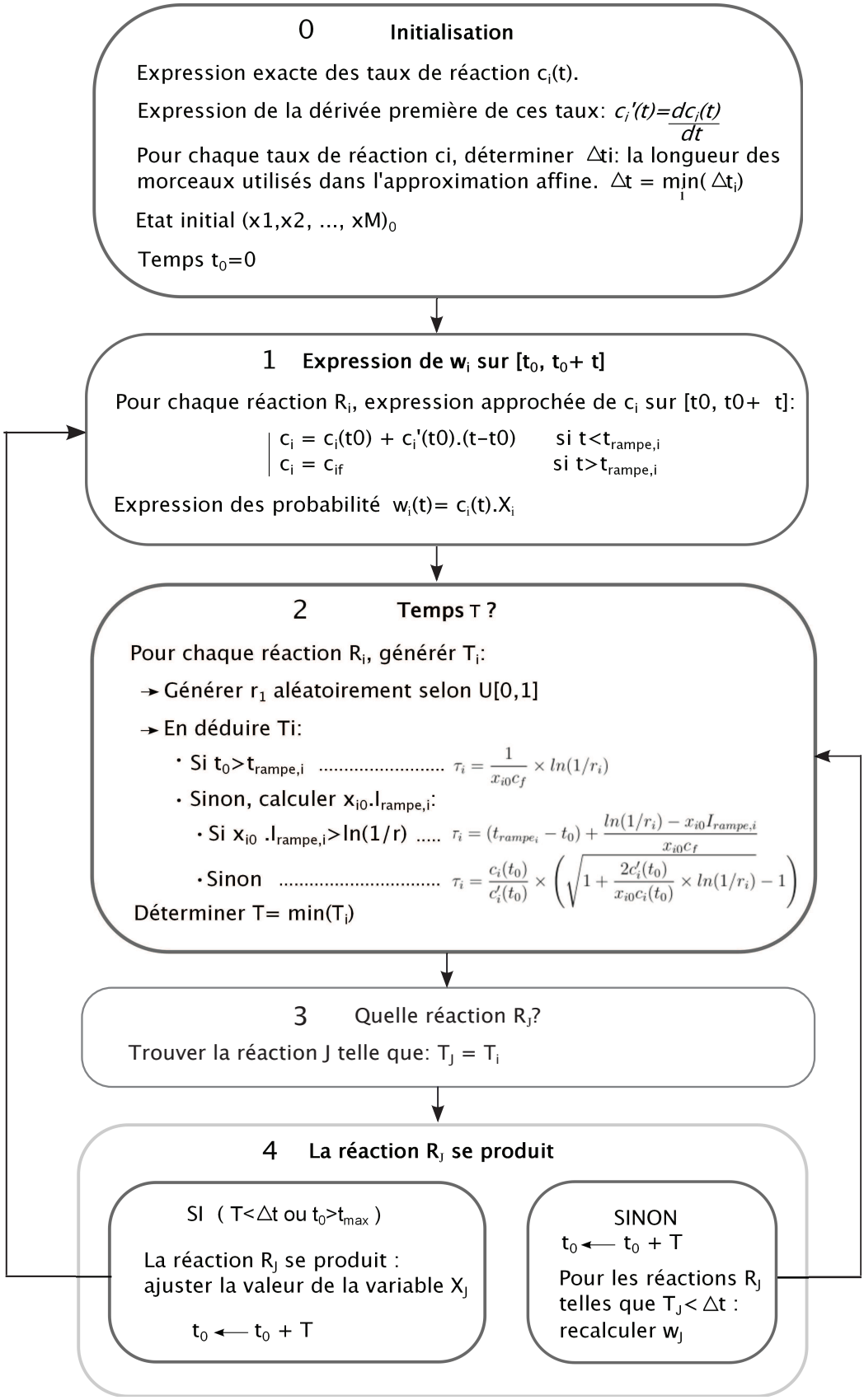


Figure III.4.4.: Algorithme de Gillespie adapté aux taux de réactions qui varient dans le temps.

Vérification du code

Pour vérifier la validité de notre code, on s'appuie encore une fois sur le modèle simple de Thattai et al.: on considère maintenant que les taux de réaction varient dans le temps. Pour simplifier, ces variations seront décrites par des rampes: les taux de réaction k_m, k_p et γ_p varient linéairement sur (respectivement) $[0 \quad t_{km}]$, $[0 \quad t_{kp}]$ et $[0 \quad t_{\gamma_m}]$. Le taux γ_m est, quant à lui, considéré constant. Tout ceci est résumé par la figure ci-dessous.

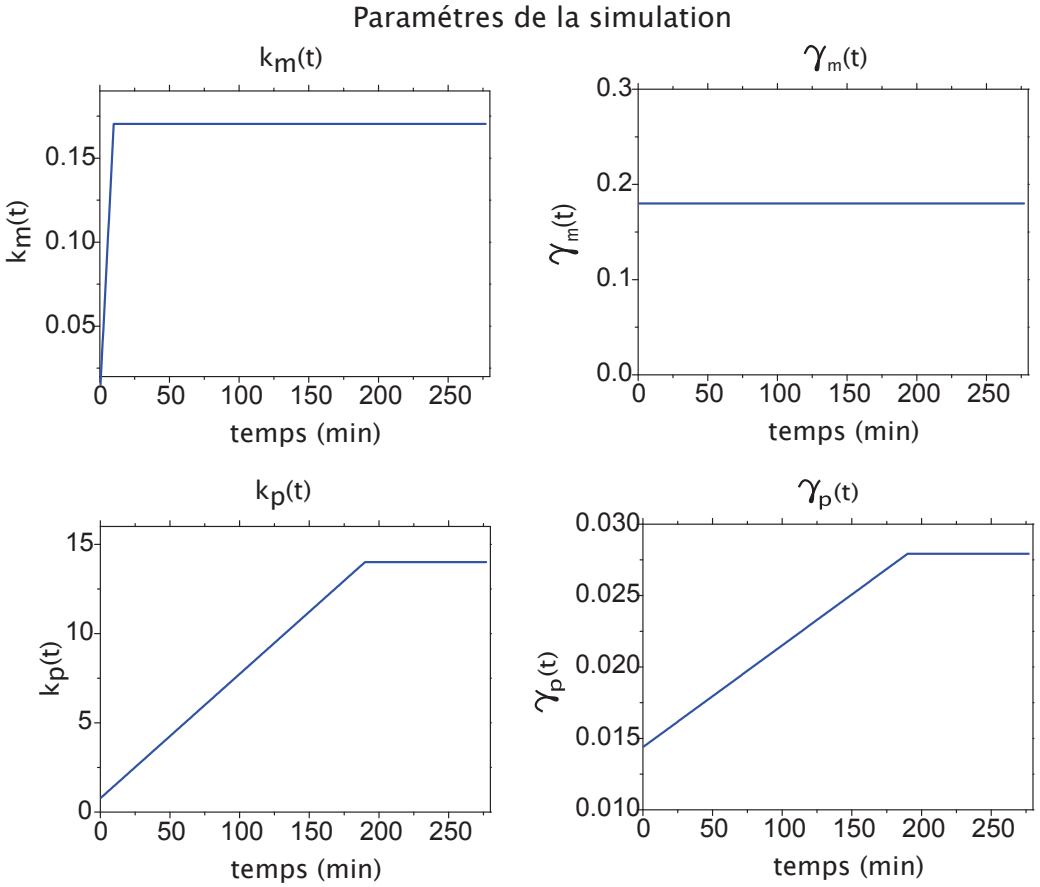


Figure III.4.5.: Variations des taux de réaction rentrés dans la simulation. Pour simplifier, ces variations sont ici décrites par des rampes.

On compare alors les données simulées avec les formules analytiques. Cette comparaison se fera sur les moments d'ordre 1 et 2 des variables n et m.

Expression analytique des moments d'ordre 1 et 2

Pour des taux de réactions qui varient dans le temps, l'équation maîtresse (voir section III.4.1 précédente) est difficile à résoudre.
Pour simplifier, nous ne nous intéresserons qu'aux moments du premier et du deuxième ordre de la statistique. A partir de l'équation maîtresse, on peut montrer que ces moments sont déterminés par le système d'équations IV.1.1 défini plus loin, dans le chapitre IV.1.

Pour nous épargner l'épreuve calculatoire de résolution du système, nous extraierons (des données simulées) les membres de droite et de gauche du système. Nous nous assurerons ensuite que les égalités sont vérifiées.

Simulations

Nous avons simulé 100 000 trajectoires pour le modèle de Thattai et al. qui contient deux variables : n et m. Sur chacune des trajectoires k simulée, nous avons extrait $n_k(t)$, $m_k(t)$ et $(n_k \times m_k)(t)$.

Puis nous avons moyenné ces quantités sur toutes les trajectoires : on obtient $\langle n \rangle(t)$, $\langle m \rangle(t)$ et $\langle n \times m \rangle(t)$. Nous avons également calculé l'écart type des ensembles de trajectoires $\{n_k(t)\}_k$, $\{m_k(t)\}_k$, et $\{n_k \times m_k(t)\}_k$. On obtient ainsi $\sigma^2_m(t)$, $\sigma^2_n(t)$ et $\sigma^2_{n,m}(t)$. Ces moments ont ensuite été dérivés sous OriginPro 6.1. La dérivée d'une grandeur x y est calculée selon :

$$\left.\frac{dx}{dt}\right|_{ti} \approx \frac{1}{2}\left(\frac{x_{i+1} - x_i}{t_{i+1} - t_i} + \frac{x_i - x_{i-1}}{t_i - t_{i-1}}\right)$$

Par ailleurs, nous avons déterminé le membre de droite du système d'équation : les variations des taux de réaction sont connues (ce sont les paramètres de la simulation), et les variables $\langle n \rangle$, $\langle m \rangle$, σ^2_m, σ^2_n et $\sigma^2_{n,m}$ sont retournées par la simulation.

Comparaison des simulations avec les expressions analytiques

Les résultats obtenus sont présentés dans la figure III.4.6 suivante avec, en rouge la dérivée des moments, et en bleu les membres de droite du système d'équation. Les deux grandeurs représentées sont tirées de données simulées.

On note que les équations sur les variances et le nombre d'ARNm sont bruitées : comme les taux de réaction choisis génèrent une faible quantité d'ARNm (valeur moyenne qui reste inférieure à 0.4), le bruit de transcription est important. En conséquence, la variable $\langle m \rangle(t)$ est bruitée, et ce bruit se retrouve dans les équations où le terme en $\langle m \rangle$ joue un rôle important.

Exeption faite de ce bruit, les deux grandeurs représentées sont égales (courbes confondues). Par conséquent, les équations analytiques sont vérifiées : notre algorithme de simulation est correct, et il pourra être amélioré pour étudier des modèles plus complexes.

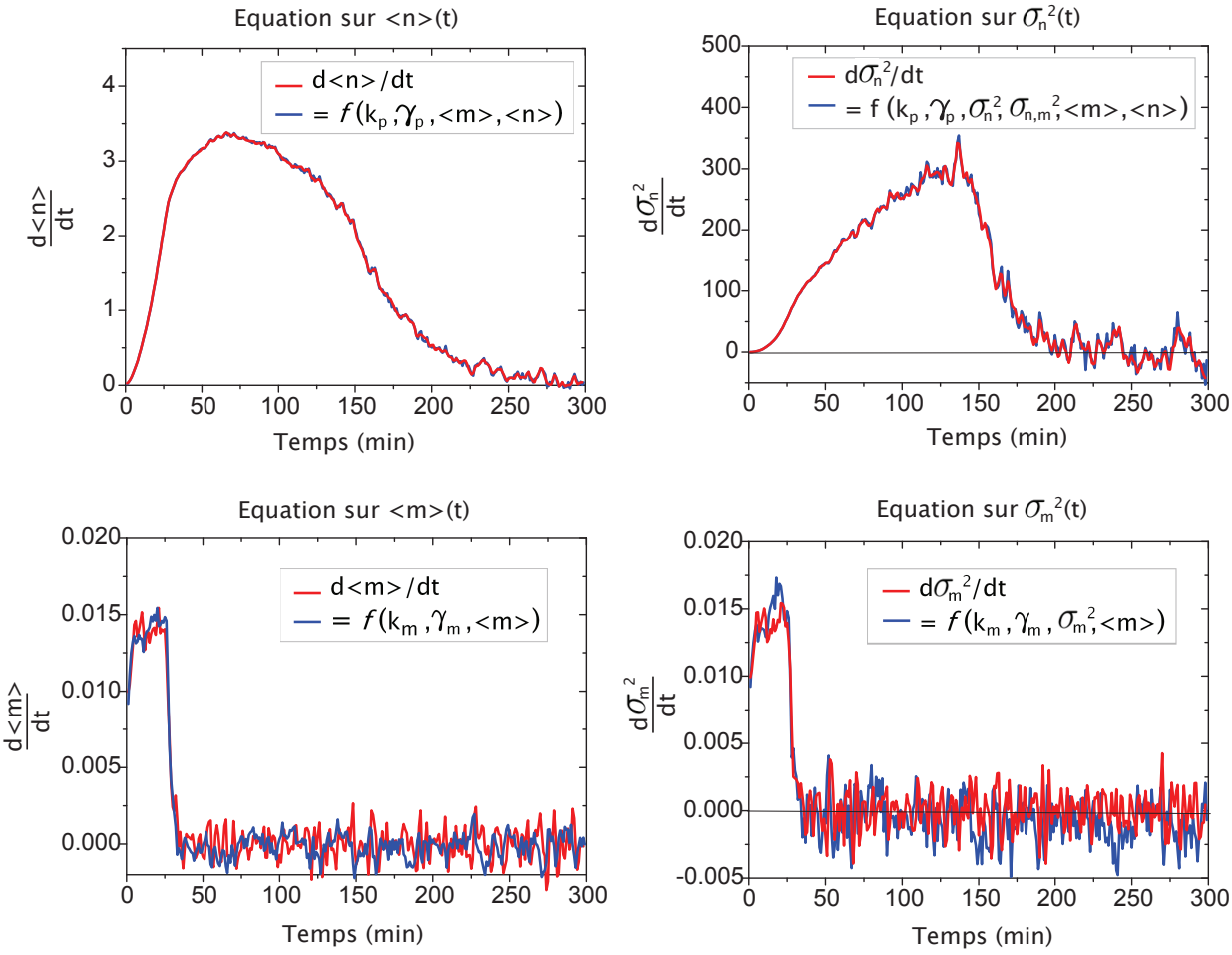


FIGURE III.4.6.: Comparaison des données simulées avec les formules analytiques : la comparaison se fait sur le modèle de Thattai et al. avec des taux de réactions qui varient de manière linéaire. Les variables d'étude sont la quantité de protéines par cellule (n) et la quantité d'ARNm cellulaire (m). Ces variables sont étudiées à travers les moments du premier et du deuxième ordre : On s'assure que ces moments vérifient les relations déterministes tirées de l'équation maîtresse. En rouge, la dérivée temporelle de ces moments ; En bleu, le terme qui définit ces dérivées selon l'équation maîtresse. Le nombre initial de molécules est uniformisé dans les cellules à : $n_0 = 0$ et $m_0 = 0$ et les taux de réaction utilisés sont donnés dans la figure III.4.5 précédente. Les données simulées sont acquises toutes les minutes, et elles comptent 100 000 trajectoires.

Quatrième partie .

Résultats

IV.1. Question posée et démarche suivie

Nous nous intéressons au module d'expression génique *gyrA*, et nous cherchons à caractériser le bruit en sortie du module. Pour ce faire, on bénéficie des travaux de confrères (voir la section I.1.3) : ceux-ci sont de plus en plus nombreux, et ils permettent d'ores et déjà de mieux comprendre la stochasticité d'expression génique. Toutefois, la plupart de ces travaux étudient le bruit d'expression génique lorsque l'environnement cellulaire n'évolue pas. Et les travaux qui se placent dans un environnement dynamique sont souvent restreints à des systèmes artificiels : les modules sondés interagissent avec des composants insérés par le scientifique, et les interactions sont réduites à un faible nombre [1].

Nous nous proposons ici de sonder la variabilité d'expression génique dans un système dynamique, et naturel : nos cellules évoluent d'un état de stress nutritionnel vers un état de croissance exponentielle, et le bruit est étudié sur un composant du réseau de régulation, naturel, qui gère cette transition d'état. On cherche alors à caractériser le bruit lors de la transition et à comprendre ses propriétés :

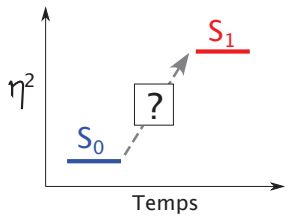


FIGURE IV.1.1.: Système d'étude : des cellules transitent entre deux états. Cette transition passe par un réseau de régulation naturel. L'état S_0 caractérise des cellules en condition de stress nutritionnel, et l'état S_1 décrit les mêmes cellules, en pleine croissance exponentielle. La transition est induite par l'ajout de glucose au milieu de culture jusqu'alors défaillant en sources carbonnées.

Quels sont les mécanismes en jeu ? Le bruit généré par ces mécanismes permet-il de remonter à leurs propriétés dynamiques ?

1. Par exemple, dans les travaux d'Ozbudak et al[12], le gène d'étude est régulé par une molécule insérée par l'expérimentateur. Et pour donner un autre exemple, les travaux de et J.M.Pedraza et al[24] étudient un gène qui fait partie d'un circuit d'interactions artificielles (*i.e.* construit par le scientifique).

Pour mener à bien cette étude, nous avons défini les *conditions expérimentales standard* : les conditions dans lesquelles on se place pour mesurer le bruit. Ces conditions sont déjà bien définies dans les chapitres précédents, et nous ne nous intéressons ici qu'aux paramètres expérimentaux qui modifient facilement le résultat. Ces paramètres critiques seront identifiés, puis réglés avec minutie.

A partir de là, les données collectées seront analysées pour déterminer la signature du bruit. Cette signature sera étudiée à travers les fluctuations statiques. Ces fluctuations reflètent les propriétés du module sondé, et nous voulons comprendre la relation qu'elles entretiennent avec les paramétres du module. Dans ce but, on recherche un modèle qui décrive les fluctuations observées ; ce modèle devra reproduire les proriétées aléatoires du module d'expression génique. Il comprendra donc deux composantes : celles-ci sont représentées sur la figure IV.1.2 et détaillées ci-dessous.

- Une *fonction de transfert* : elle relie les paramètres du module, au bruit d'expression génique observé. Elle représente donc l'ensemble des mécanismes cellulaires qui génèrent de l'aléa en sortie du module (*i.e.* sur la variable n, le nombre de protéines par cellule).
- Les *signaux d'entrée de la fonction de transfert* : ce sont les paramètres du module, ils quantifient les mécanismes cellulaires impliqués dans la fonction de transfert. Parmi ces signaux d'entrée, certains sont internes au module (taux des réactions d'expression génique), et d'autres sont externes au module (sources extrinsèques de bruit). Tous ces paramètres sont représentés dans la figure I.1.8.

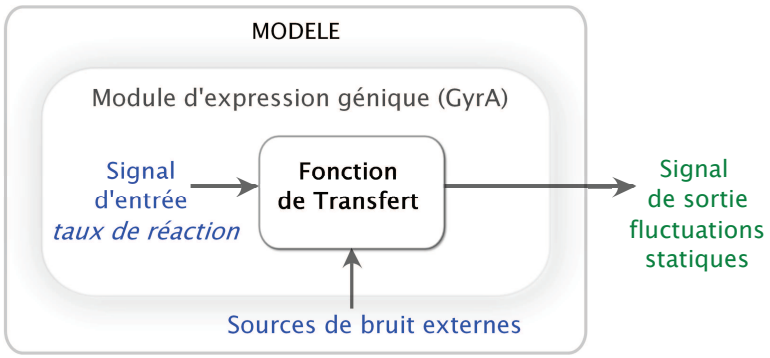


FIGURE IV.1.2.: Modélisation des mécanismes d'expression génique : le modèle inclut les paramètres du module d'expression génique, et la fonction de transfert qui intègre ces paramètres pour générer du bruit sur la variable en sortie du module (*i.e.* sur le nombre de protéines/cellule).

Pour réduire le nombre de variables du modèle, on réalisera une étude qui cherche à mesurer la valeur des signaux d'entrée : lorsque c'est possible, ces derniers seront extrapolés à partir de nos données expérimentales. On cherchera alors à ajuster les variables restantes sur les fluctuations observées : on veut ainsi déterminer la fonction de transfert et les paramètres inconnus du module sondé. Notre démarche est résumée sur la figure suivante.

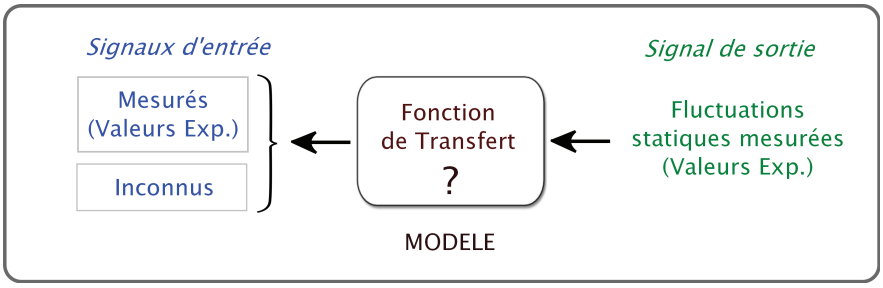


FIGURE IV.1.3.: Méthode d'identification des mécanismes impliqués dans le bruit d'expression génique, et utilisation du bruit mesuré pour quantifier ces mécanismes. Dans un premier temps, les fluctuations statiques en sortie du module d'expression génique seront mesurées et caractérisées. A partir de là, nous proposerons des modèles qui reproduisent la stochasticité d'expression génique. Ces modèles, une fois ajustés aux données expérimentales, permettront d'accéder à la fonction du transfert du module, et de quantifier les paramètres (signaux d'entrée) inconnus.

Modèle proposé

Nous essayerons de décrire nos données par un modèle simple : le nombre limité de cellules imagées fournit des statistiques dont la précision ne permet pas de différentier un modèle minutieu d'un modèle plus grossier. Nous commençons par tester le modèle simpliste de "transcription-traduction" où seules les sources intrinsèques interviennent : Un tel modèle permet-il d'expliquer les distributions observées ? Ce modèle à deux étages est présenté dans la section I.1.3), et nous le rappelons dans la figure IV.1.4 pour le cas étudié.

Comportement du modèle

Equation maîtresse

Ce modèle peut être décrit par l'équation maîtresse donnée dans la section I.1.3. Cette équation a été dérivée pour des taux de réaction qui restent invariants dans le temps. Or dans nos expériences où les cellules passent d'un état de stress nutritionnel à un état de croissance exponentielle, ces taux de réaction sont suceptibles d'évoluer. Et comme ces taux varient ici de manière continue, ils pourront être approximés constants sur des

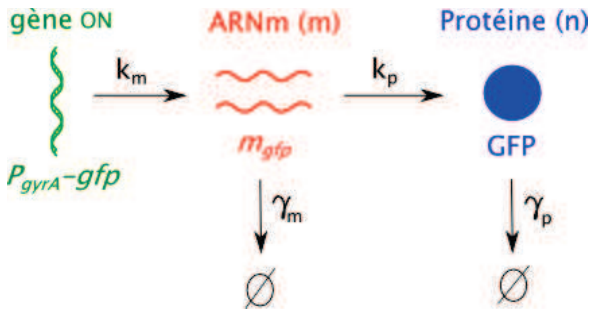


FIGURE IV.1.4.: Modèle à deux étages : transcription et traduction. k_m et k_p représentent (respectivement) les taux de transcription et de traduction. Et γ_m et γ_p sont les taux associés aux réactions de dégradation. Le module d'expression génique étudié ici se réfère au système rapporteur des cellules *GyrA* : le gène de la GFP rapporte l'activité promotrice du gène *gyrA*.

intervalles de temps dt infintésimales. Dans ces conditions, l'équation maîtresse reste inchangée : la probabilité $P_{m,n}$ d'avoir n proteines et m ARNm dans la cellule à l'instant t, doit juste s'écrire en fonction des taux instantanés de réaction.

$$\begin{aligned}\frac{dP_{m,n}(t)}{dt} &= \gamma_m(t)\Big((m+1)P_{m+1,n}(t) - mP_{m,n}(t)\Big) + k_m(t)\Big(P_{m-1,n}(t) - P_{m,n}(t)\Big) \\ &+ \gamma_p(t)\Big((n+1)P_{m,n+1}(t) - nP_{m,n}(t)\Big) + k_p(t)\Big(mP_{m,n-1}(t) - mP_{m,n}(t)\Big)\end{aligned}$$

Solutions

Cette équation a été résolue par V.Shahrezaei et al [16] pour des taux de réaction qui restent constants dans le temps. Comme évoqué précédemment, dans nos expériences l'environnement cellulaire change, et ces taux ne sont pas constants. Par conséquent, les résultats des confrères ne sont donc plus exploitables. On pourra toutefois les utiliser lorsque l'équilibre statistique sera atteint : dans ce cas, les taux de réaction n'évoluent plus, et les résultats de V.Shahrezaei et al restent valides. On sait ainsi qu'à l'équilibre statistique, si $\gamma_m \gg \gamma_p$, le modèle génère une distribution binomiale négative (I.1.1) de paramètres $a = k_{mEq} \setminus \gamma_{pEq}$ et $b = k_{pEq} \setminus \gamma_{mEq}$. Cette loi tend vers la distribution gamma pour n grand.

Lorsque les taux de réactions évoluent, les solutions analytiques ne sont plus connues. On peut cependant dériver des relations simples sur les moments du premier et du deuxième ordre. On peut ainsi montrer que les variables n (nombre de proteines) et m (nomnre d'ARNm) du modèle obéissent à :

$$\begin{aligned}
\frac{d\langle m\rangle(t)}{dt} &= k_m(t) - \gamma_m(t)\langle m\rangle(t) \\
\frac{d\sigma_m^2(t)}{dt} &= -2\gamma_{m(t)}\sigma_m^2(t) + k_m(t) + \gamma_m(t)\langle m\rangle(t) \\
\frac{d\langle n\rangle(t)}{dt} &= k_p(t)\langle m\rangle(t) - \gamma_p(t)\langle n\rangle(t) \\
\frac{d\sigma_n^2(t)}{dt} &= -2\gamma_p(t)\sigma_n^2(t) + 2k_p(t)\sigma_{n,m}^2 + k_p(t)\langle m\rangle(t) + \gamma_p(t)\langle n\rangle(t) \\
\frac{d\sigma_{n,m}^2(t)}{dt} &= -\big(\gamma_p(t) + \gamma_m(t)\big)\sigma_{n,m}^2(t) + k_p(t)\sigma_{n,m}^2(t)
\end{aligned} \tag{IV.1.1}$$

où $\langle m\rangle(t)$ et $\langle n\rangle(t)$ désignent les moyennes et $\sigma_m^2(t)$, $\sigma_m^2(t)$ et $\sigma_{n,m}^2(t)$ désignent respectivement les variances de m et n et le dernier la covariance de n et m. Ces équations analytiques permettent d'étudier le comportement moyen du modèle. Et pour une description plus complète qui donne la loi de probabilité entière, on se tournera vers les simulations. L'algorithme utilisé est décrit dans la section III.4, et il intègre les variations temporelles des taux de réaction.

Confrontation avec nos données expérimentales

Ce modèle sera confronté à nos données expérimentales. On dispose pour ce faire, des mesures de fluorescence qui reflètent la quantité de protéines GFP présentes dans une cellule. Le modèle sera donc ajusté sur la variable en sortie du module (*i.e.* la densité de GFP/cellule) : on jouera alors sur les paramètres inconnus du modèle pour reproduire, au mieux, les fluctuations statiques mesurées sur la GFP. Si au final le modèle s'accorde avec les données expérimentales, alors on aura trouvé une fonction de transfert (celle de la figure IV.1.4), et des valeurs pour ses entrées masquées (les taux de réaction qui ne sont pas connus) qui permettent d'expliquer nos résultats expérimentaux.

IV.2. Conditions expérimentales standard

Nous avons vu précédemment que nos expériences sur cellules uniques contiennent un grand nombre de paramétres expérimentaux. Ces paramètres sont autant de variables qui peuvent influencer les résultats. Il est donc important de bien définir les conditions expérimentales, et ce, d'autant plus qu'on s'intéresse au bruit dans les données collectées. En ce sens, nous avons mené (dans la section II.2) une étude qui caractérise les propriétés de notre système optique et celles du système d'acquisition utilisé.

Nous allons maintenant nous intéresser aux paramétres expérimentaux qui définissent *l'état et la réponse* de l'échantillon étudié. Ces paramétres sont donnés dans la section II.1 à travers les protocoles de préparation des échantillons et d'acquisitions. Ils agissent directement sur le vivant, et lorsque la relation est forte, de petites variations peuvent générer une variabilité non négligeable dans les données collectées. Pour s'affranchir de cette variabilité non désirée, nous avons mené une étude qui vise à identifier et contrôler ces paramétres critiques. A partir de là, nous avons défini les *conditions expérimentales standards* : les conditions dans lesquelles on se place pour sonder la variabilité d'expression génique.

IV.2.1. Paramètres expérimentaux

Pour identifier les paramètres critiques, nous avons mené plusieurs expériences. Ces dernières sont réalisées selon le même protocole, mais les paramètres sont réglés à des valeurs voisines : le temps de préculture, le taux de dilution, la durée du stress nutritionnel, l'application du signal excitateur (ajout de glucose), et l'intensité d'illumination sont légérement variées. Toutes ces différences sont résumées dans le prochain tableau IV.2.1, avec l'étiquette de l'expérience correspondante.
Pour étudier l'effet des différents paramètres, nous nous appuierons sur la densité de fluorescence cellulaire. Et pour faire ressortir les traits principaux de ce phénotype, les données individuelles seront moyennées. On travaille donc sur la population de cellules imagées.

Comportement global

Dans une première étude, nous nous intéresserons aux caractéristiques de la réponse fluorescente moyenne qui sont communes à toutes les expériences : cette étude permet de poser les bases de la réponse étudiée.

◦ Comportement au sein de nos expériences

La figure IV.2.1 regroupe les résultats des différentes expériences. Par mesure de clarté, les expériences sont regroupées selon le facteur d'atténuation du filtre neutre placé en aval de la lampe fluorescente.

Exp.	Préparation des échantillons				Signal excitateur	Imagerie
	Durée de la préculture	Taux de dilution	$O.D$ en fin de croissance à nouveau	Durée du stress nutrionnel	Homogénéisation du glucose	Filtre atténuateur
1	17h00	1/500	0.32	19h21	oui	0,08
2	17h05	1/500	0.33	19h08	oui	0,08
3	18h30	1/500	0.05	19h02	oui	0,5
4	18h30	1/500	0.04	20h13	oui	1
5	14h38	1/500	0.07	25h30	non	1
6	14h40	1/500	0.46	20h45	non	1
7	14h37	1/500	0.46	3h58	non	1
8	20h10	1/40	0.25	15h30	non	1
9	20h10	1/40	0.51	22h47	non	1
10	16h54	1/40	0.7	22h55	non	1
11	16h27	1/40	0.5	19h10	non	1

TABLE IV.2.1.: Paramètres expérimentaux utilisés dans différentes expériences : caractéristiques de préparation des échantillons, valeur du signal excitateur "ajout de glucose", et intensité d'illumination en fluorescence. Le signal excitateur est varié sur deux valeurs : le glucose est injecté dans le milieu de culture, sans les trois étapes d'aspiration-réinjection ("non"), ou avec ("oui") (cf. section II.1.4). Et l'intensité d'illumination est modulée via le filtre atténuateur placé en aval de la lampe fluorescente (facteur d'atténuation de 0.08, 0.5 ou 1).

On constate que toutes les expériences retournent le même trait de réponse: la densité moyenne de fluorescence cellulaire augmente suite à l'ajout de glucose, puis, à partir d'un certain moment, elle atteint un plateau. Nous verrons plus loin que ce plateau correspond aux conditions de croissance exponentielle. Ce comportement est cohérent avec les prédictions qualitatives des bioinformaticiens. En effet, dans un cadre déterministe, le passage de la phase stationnaire à la phase exponentielle s'accompagne d'une augmentation de l'activité du promoteur P_{gyrA}. Dans ces conditions, la production de GFP augmente dans

les cellules *GyrA*, et le signal de fluorescence est de plus en plus intense, comme observé sur les courbes de la figure IV.2.1.

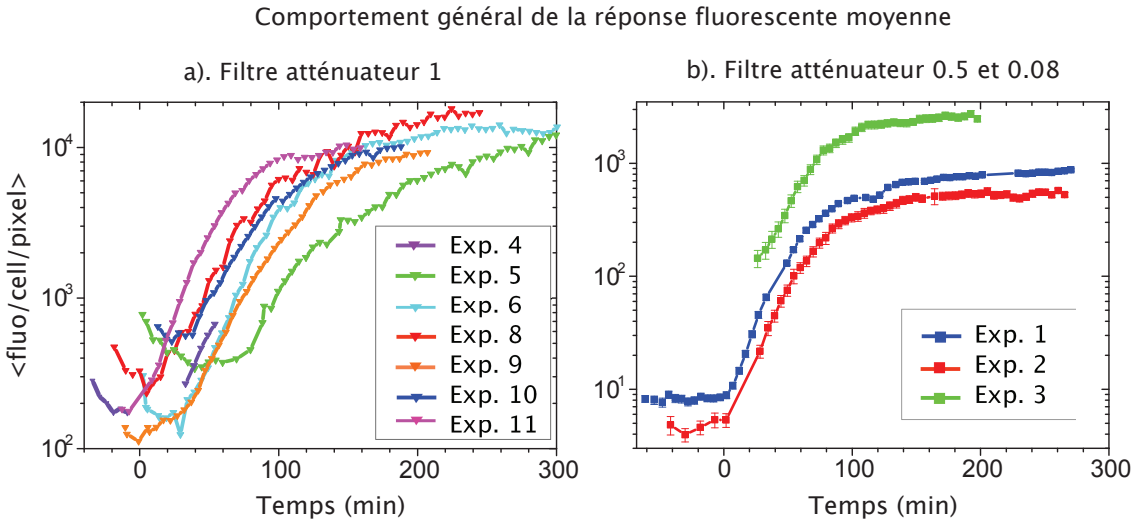


FIGURE IV.2.1.: Réponse fluorescente moyenne mesurée dans différentes expériences. Les valeurs sont brutes (non-calibrées par l'intensité d'excitation de la lampe de fluorescence). La population de cellules est placée en condition de stress nutritionnel. Puis à $t = 0$ min, des nutriments sont ajoutés au milieu de culture. a). Densité moyenne de fluorescence cellulaire pour une population de cellules pleinement éclairée en fluorescence. Par mesure de clarté, les barres d'erreur ne sont pas tracées. b). La même grandeur pour une population plus faiblement éclairée : l'illumination de fluorescence est atténuée d'un facteur 0,5 (Exp 3), ou 0,08 (Exp 1 et 2). On note qu'entre les courbes de la figure a) et celles de la figure b), la lampe fluorescente a été changée. Et entre les expériences n°4, 5, 6, et 8 verte, cyan, rouge et violet de la figure a), et les expériences n°9, 10 et 11, l'ampoule fluorescente a été changée.

○ Comparaison avec les mesures en population des biologistes

Les expériences en plaques multipuits de nos collaborateurs sont réalisées à partir des mêmes cellules souches que nous utilisons, et celles-ci ont subi un traitement similaire: seule l'étape de dilution varie, avec dans nos expériences une préculture diluée 500x, et dans les expériences en population une préculture diluée 40x [1]. De plus, dans nos analyses, si un grand nombre de cellules est imagé, le bruit stochastique ne devrait pas influencer le comportement de la population, puisqu'il est moyenné. Dans ces conditions, on devrait retrouver les mêmes traits de réponse dans nos expériences et dans celles des biologistes:

- La fluorescence émise par une bactérie (exprimée en unité de fluorescence FU) est donnée par l'absorbance et la fluorescence totale de l'échantillon en plaques multipuits :

1. On rappelle qu'avant d'être placée en conditions de stress nutritionnel, la préculture est diluée dans du milieu frais pour pouvoir travailler à partir de cellules qui, en milieu de croissance exponentielle, sont dans un état bien défini.

$$\langle \text{Fluorescence/cellule} \rangle \approx \frac{\text{Fluorescence de la solution}}{\text{Abs}}$$

où *Abs* est la mesure de la densité optique à une longueur d'onde de 600 nm. Elle est, pour des valeurs $< 0,5\text{cm}^{-1}$ proportionnelle à la concentration des cellules.

- Nos expériences permettent de connaître la densité de fluorescence cellulaire exprimée en coups (MG), et plus simplement, la fluorescence émise par les cellules individuelles (FG) en unité de coups/cellule. En moyennant cette grandeur FG sur toutes les cellules, on mesure donc la même quantité que les biologistes, via un facteur de proportionnalité qui reflète la variation d'intensité d'excitation et la sensibilité de la chaîne de détection.

On peut donc comparer nos mesures avec celles des biologistes. Les résultats sont présentés sur la figure IV.2.2. Et on obtient des résultats sur cellules uniques cohérents avec ceux obtenus en population.

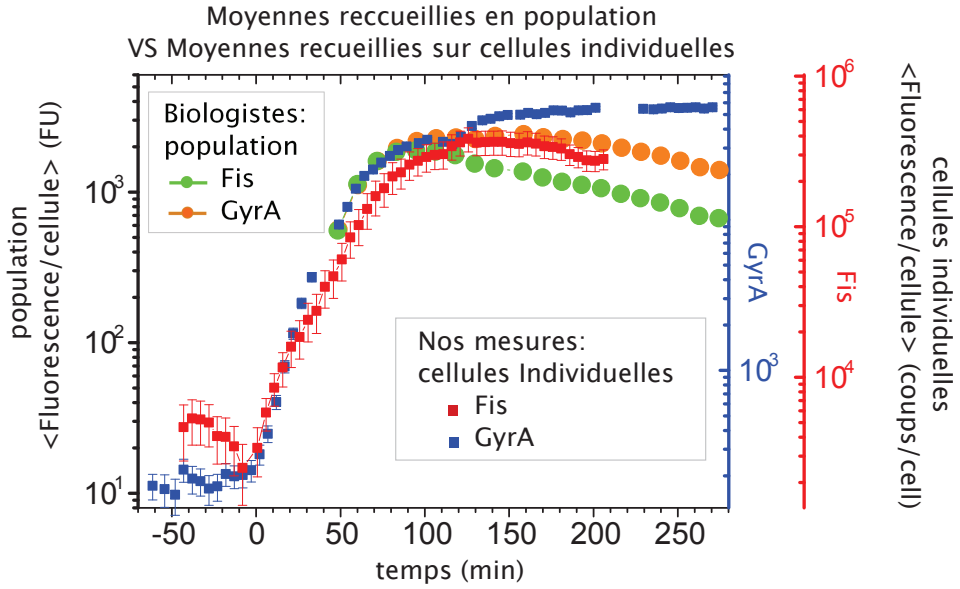


Figure IV.2.2.: Evolution temporelle de la fluorescence cellulaire lors de la réentrée dans la phase exponentielle de cellules *E.Coli*: comparaison des résultats obtenus par des biologistes avec des moyennes faites sur des cellules individuelles. En orange et vert, la fluorescence mesurée pour les cellules *GyrA* et *Fis* lors d'expériences en population. En bleu et rouge, la valeur moyennée sur $200-600$ cellules *GyrA* (coube bleu) et $80-170$ cellules *Fis* (courbe rouge). On note que les cellules Fis ont été pleinement éclairées en fluorescence, alors que pour les cellules GyrA, l'éclairement de fluorescence est atténué d'un facteur 0.08.

On observe cependant des détails qui diffèrent entre nos expériences et celles des biologistes. Dans les expériences réalisées en population, les cellules GyrA atteignent plus rapidement leur maximum de fluorescence cellulaire: le niveau de fluorescence est maximum à $t \approx 110$ min. Et après un plateau de ≈ 70 min, la fluorescence cellulaire commence à diminuer. Dans nos expériences sur cellules uniques, le maximum de fluorescence cel-

lulaire est atteint plus tard (à $t \approx 200$ min), et nous n'observons pas de décroissance pendant la durée de nos expériences.

Dans les expériences en population, la décroissance de la fluorescence cellulaire à partir de ≈ 70 min est exponentielle de paramètre $\tau = 165 \pm 3$ min, et ce, pour les deux types de cellules imagées (*Fis* et *GyrA*). On en déduit qu'elle traduit un phénomène global, indépendant du promoteur. De plus, elle n'est pas générée par des composants accumulés dans le milieu de culture [60].

Nous allons maintenant étudier les différences au sein de nos expériences. On cherchera à établir une corrélation entre les paramétres du tableau IV.2.1 et les différences observées sur les courbes de la figure IV.2.1. Cette étude nous permettra d'identifier les paramètres critiques dans nos expériences, et de réaliser des protocoles qui assurent un contrôle minutieux de ces paramètres.

Identification des paramètres critiques

○ Paramètre d'imagerie : "facteur d'atténuation de l'éclairage fluorescent"

Ce paramètre contrôle le niveau de fluorescence mesuré : lorsque l'intensité d'excitation de fluorescence est élevée, les cellules sont fortement excitées, et le signal de fluorencence émis par les individus est élevé. Ainsi, sur la figure IV.2.1-b, on trouve un rapport de l'ordre de $0,5/0,08$ entre les plateaux des courbes vertes (illumination de fluorescence atténuée d'un facteur $0,5$) et rouges (illumination de fluorescence atténuée d'un facteur $0,08$). Et sur la figure IV.2.1-a, on note une différence de comportement avant l'assimilation du glucose : pour les cellules pleinement éclairées en fluorescence, le niveau de fluorescence des cellules décroît, même si leur taille reste quasi inchangée (voir les figures IV.2.5 et IV.3.4 de la prochaine section IV.4). Pourtant, après $\approx$ 20h en phase stationnaire, la fluorescence par cellule devrait atteindre un régime de fonctionnement stationnaire, d'une valeur faible, mais constante. Dans notre cas, la diminution de la fluorescence apparaît avec le lancement

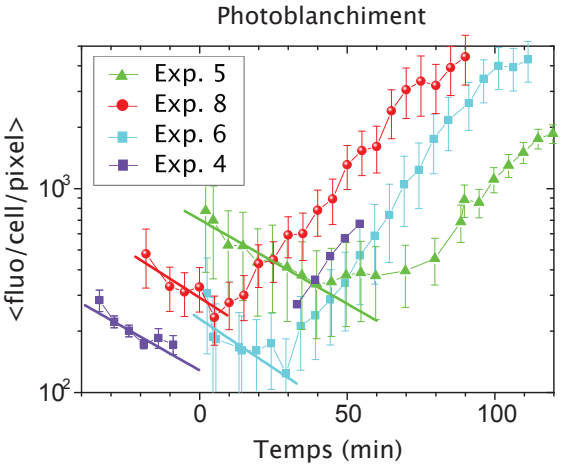


Figure IV.2.3.: Dégradation par photoblanchiment. Les points représentent les courbes de fluorescence de la figure IV.2.1-a avec les barres d'erreur associées. Et les droites donnent l'ajustement numérique des courbes par la fonction : $y = A.exp(-\frac{x}{\gamma_{bleach}})$.

des observations: lors des acquisitions, les cellules sont exposées à des éclairages intenses pendant des temps non négligeables (≈ 100 ms). Ainsi, les propriétés fluorescentes de la GFP seront perdues par photo blanchiment. Dans ce cas, on s'attend à un même type de décroissance pour toutes les expériences de la figure IV.2.1-a: c'est bien ce qu'on observe, comme souligné sur la figure IV.2.3.
Cette atténuation peut être caractérisée par un taux de dégradation de la GFP sous l'effet du photo blanchiment (γ_{bleach}). Un ajustement numérique des courbes de fluorescence montre que cette décroissance est exponentielle, de paramètre: $\gamma_{bleach} = 0,013 \pm 0,002$ min^{-1}. Lorsque les cellules sont éclairées par une lumière $\approx 12,5$ fois moins intense, il apparaît que la dégradation par photoblanchiment devient négligeable devant les processus naturels de dégradation de la GFP. Dans ce cas, le photoblanchiment n'a pas d'effet sur le signal détecté. Et le niveau moyen de fluorescence cellulaire reste constant pendant la phase stationnaire de stress nutritionnel (voir figure IV.2.1-b). Ainsi pour minimiser les perturbations introduites par notre système de mesure, l'intensité d'excitation en fluorescence sera atténuée d'un facteur $0,08$.

◦ Paramètre du signal excitateur : "procédure d'ajout du glucose"

L'ajout de glucose est un paramétre ultra-critique : il est très sensible à la gesture du scientifique, et lorsqu'il est mal homogénéisé dans le milieu de culture, il introduit un décalage temporel de la réponse cellulaire. Ainsi, sur la figure IV.2.1-a, la réponse cellulaire ne démarre pas au même moment dans les différentes expériences : les courbes de fluorescence commencent à augmenter à différents instants. Pour s'affranchir de cette variabilité, nous avons mis en place une procédure "d'aspiration-réinjection"(voir la section II.1.4) : avec cette nouvelle procédure, les expériences sont synchronisées en terme de réponse moyenne cellulaire, comme montré sur la figure IV.2.1-b.

◦ Paramètre de préparation des échantillons

Sur la figure IV.2.1-a, on note que le rapport entre les plateaux de début et fin d'acquisition varie d'une expérience à l'autre. Pourtant, les plateaux de stress nutritionnel prennent des valeurs très différentes, alors que les plateaux de fin d'expérience tendent vers la même valeur. Ce résultat dépend des paramétres de préparation des échantillons. Ces paramètres sont multiples (cf tableau IV.2.1), et chacun d'eux est réglé à des valeurs différentes dans les expériences. Dans ces conditions, il est difficile d'établir une corrélation avec le phénomène observé.
On pense toutefois que la différence dans le rapport des deux plateaux est fortement corrélée à l'étape de centrifugation-rinçage. On rappelle que cette étape permet d'enlever le glucose du milieu de culture. Cela permet de placer les cellules en condition de stress

nutritionnel. Dans nos premières expériences (courbes de la figure IV.2.1-a), la solution de "cellules+glucose" n'est centrifugée qu'une fois (pendant 5 ou 10 min) avant d'être rincée. Nous pensons que ce protocole de rinçage n'est pas suffisant pour enlever tout le glucose du milieu de culture. Dans ces conditions, les cellules ne sont pas dans le même état en début d'expérience : pour les courbes orange, rose, violette et cyan, elles sont en condition de stress avancé, alors que pour les courbes verte, rouge et bleu elles sont moins avancées dans le stade de stress nutritionnel.

◦ Conclusion : conditions expérimentales standard

Au final, les trois paramètres critiques sont ajustés aux valeurs données ci-dessous :

- Imagerie : facteur d'atténuation de l'éclairage fluorescent de $0,08$.
- Signal excitateur "ajout de glucose" : homogénéisation du sucre dans le milieu de culture.
- Préparation des échantillons : lorsque les cellules sont placées en condition de stress nutritionnel, le milieu de culture est changé par centrifugation-rinçage. Ce processus sera réalisé en trois étapes, avec trois centrifugation de 20 min, chacune suivie d'un rinçage.

Ces paramètres correspondent aux conditions des expériences 1 et 2. Et par la suite, nous travaillerons sur ces deux expériences.

IV.2.2. Cellules endommagées

Nos expériences de microscopie *in vivo* apportent des contraintes sur les cellules, comme par exemple l'opération de rinçage du glucose (3 centrifugation 3000rpm de 20 minutes chacune), l'étape d'attachement des cellules à la surface du pétridish, ou encore l'éclairage en contraste de phase et en fluorescence . Tout ceci peut perturber les cellules, et dans nos images en constraste de phase, on observe des cellules mortes. Ces cellules peuvent avoir été endommagées par notre protocole expérimental, mais elles peuvent aussi traduire des phénomènes biologiques. Par exemple, on peut imaginer qu'en conditions difficiles de stress nutritionnel, certains individus sont sacrifiés pour la survie de la communauté. Dans cette étude, nous ne nous intéressons pas aux causes du décès, mais aux conséquences: nous nous proposons d'étudier l'effet des cellules mortes sur nos analyses de bruit d'expression génique.

Pour ce faire, nous avons analysé des données où toutes les cellules imagées sont considérées : le programme de réjection des cellules malades vu au Chapitre 2 (section 2.2.5) n'est pas appliqué. Nous décrirons alors succinctement les *fluctuations statiques* mesurées, et nous verrons que dans nos analyses, ces fluctuations sont sensibles au nombre de cellules

mortes.

Etude statistique

Les fluctuations statiques dans l'activité du promoteur P_{gyrA} sont données par la variabilité intercellulaire de densité de fluorescence. Pour caractériser cette variabilité lors de la réponse au stress nutritionnel, nous avons tracé la distribution de population à différents instants : Comment les niveaux de fluorescence sont distribués dans la population de cellules ? Les résultats peuvent être visualisés sur la figure IV.2.4, avec la distribution à 3 temps d'acquisition différents.
Sur cette figure, on observe des distributions bimodales : une partie des cellules est caractérisée par une concentration de fluorescence cellulaire très faible, et une autre partie par une concentration de fluorescence cellulaire élevée. Dans nos premières recherches, nous avons attribué cette bimodalité à la régulation transcriptionnelle.

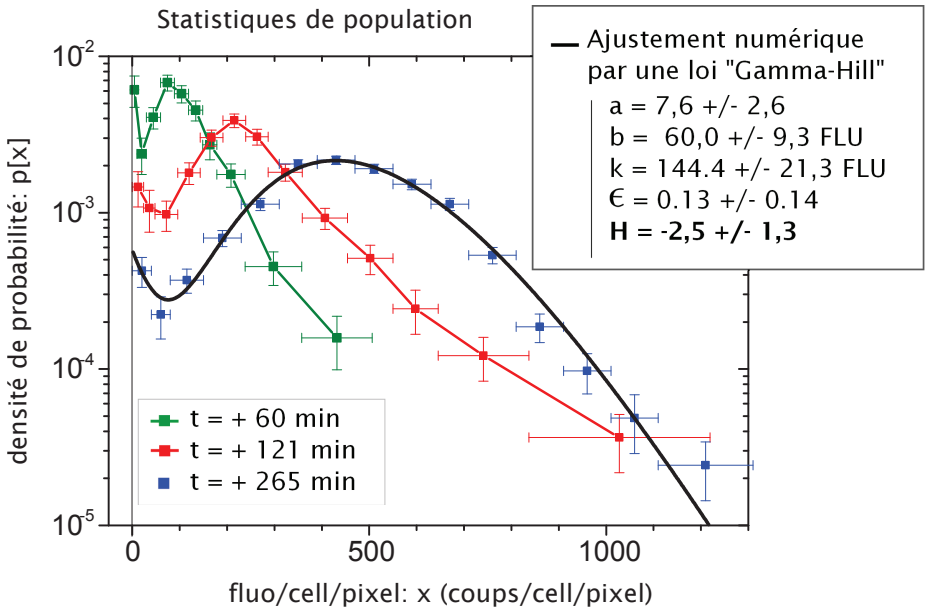


FIGURE IV.2.4.: Distribution de la fluorescence émise par pixel de cellules *GyrA* lors de la transition de phase entre état stationnaire et croissance exponentielle. Les trois distributions sont tracées sur la même population de cellules, imagée à différents moments : en vert, la distribution obtenue 60 min après injection de glucose (297 cellules), en rouge 121 min après l'ajout de glucose au milieu de culture (430 cellules), et en bleu la distribution en fin d'acquisition (265 min après l'ajout de glucose avec 1234 cellules imagées). Un ajustement numérique (courbe noire) permet de lisser la statistique en fin d'acquisition par la fonction définie dans l'équation IV.2.1 précédente. La valeur trouvée pour les paramètres a, b, k, H est donnée dans l'encadré, et le facteur de normalisation vaut : $A = 1,9.10^{-20} \pm 0,2.10^{-20}$. On note que l'échelle des ordonnées est logarithmique, et que la taille des bins a été ajustée manuellement pour obtenir une courbe continue, qui fait apparaître les détails près de $x = 0$ (pic de fluorescence en 0) sans pour autant génerer des valeurs trop imprécises (les bins sont remplis avec au moins 15 cellules près de $x = 0$).

Théorie

Dans leur papier théorique de 2006, N.Friedman *et al* ont montré qu'une production de protéines par bouffées génère une variabilité inter-cellulaire qui obéit à une distribution gamma. (voir la section I.1.3). Le modèle [17] est initialement affranchi de la régulation génique, mais ces scientifiques l'ont étendu aux cas des protéines qui s'autorégulent. Dans ce système plus complexe, l'autorégulation est modélisée par une dépendance en x de la fréquence de bouffée. Cette dernière n'est plus a, mais "$a \times c(x)$", où $c(x)$ désigne une fonction réponse. Dans le cas d'une protéine qui répond en forme de Hill, la fonction réponse s'écrit :

$$c(x) = \frac{k^H}{k^H + x^H} + \epsilon$$

où H désigne le coefficient de Hill [2] et $k = K_{ON}/K_{OFF}$ est la constante d'équilibre pour la liaison du facteur de transcription au promoteur. Quant à ϵ, il traduit le niveau basal de transcription du promoteur. Dans ce contexte, à l'équilibre statistique, la distribution de population est régie par une loi *gamma-Hill* :

$$p(x) = \frac{A \times x^{a(1+\epsilon)-1} e^{-x/b}}{\left[1 + \left(\frac{x}{k}\right)^H\right]^{a/H}} \tag{IV.2.1}$$

où A est un facteur de normalisation. Cette loi de probabilité est bimodale pour certaines valeurs des paramètres. Dans ce cas, la bimodalité est induite par la régulation transcriptionnelle, avec une fonction réponse qui module la fréquence des bouffées.

Une bimodalité induite par la régulation transcriptionnelle dans nos expériences ?

Inspirés par ces travaux théoriques, nous avons ajusté numériquement notre statistique expérimentale par la distribution gamma-Hill : en fin d'acquisition, on est à l'équilibre statistique (voir section IV.3.2 plus loin) et si les sources intrinsèques de bruit sont dominantes, la théorie peut s'appliquer. L'ajustement numérique obtenu est tracé en noir dans la figure IV.2.4 : il s'accorde très bien à nos données, et les ordres de grandeur pour les paramètres a et H semblent corrects.
Cependant, la valeur négative du coefficient de Hill suppose une autorégulation positive pour le gène *gyrA*, ce qui est contradictoire avec le réseau de régulation connu ([38]).
On peut imaginer plusieurs raisons à cette contradiction, mais en cherchant au plus simple, on peut l'expliquer par la présence de cellules mortes : avec un niveau de fluorescence

2. Le coefficient de Hill correspond au nombre de molécules inductrices nécessaires à l'activation du facteur de transcription : le facteur de transcription régulateur est rendu actif seulement si , par exemple plusieurs molécules inductrices s'y accrochent simultanément.

faible, celles-ci surestiment la densité de probabilité pour les concentrations de fluorescence cellulaire proches de zéro. Pour tester cette dernière hypothèse, nous nous sommes intéressés à la dynamique des cellules individuelles.

Etude dynamique

Dans cette étude dynamique, nous nous intéressons à deux grandeurs : la surface occupée par une cellule individuelle, et la densité de fluorescence cellulaire qu'elle renferme. Sur les trajectoires (dans le temps) des cellules individuelles, on observe :

- ◇ Des cellules qui se divisent (en bleu et rouge sur la figure IV.2.5), avec un niveau de fluorescence qui augmente pendant la durée de l'expérience. Cette augmentation traduit l'activation du gène *gyrA* suite à l'ajout de glucose.
- ◇ Des cellules qui ne se divisent pas (en vert sur la figure IV.2.5) : leur taille reste constante dans le temps, et le niveau de fluorescence est très faible pendant la durée de l'expérience. Nous attribuons ce phénomène à la mort cellulaire.

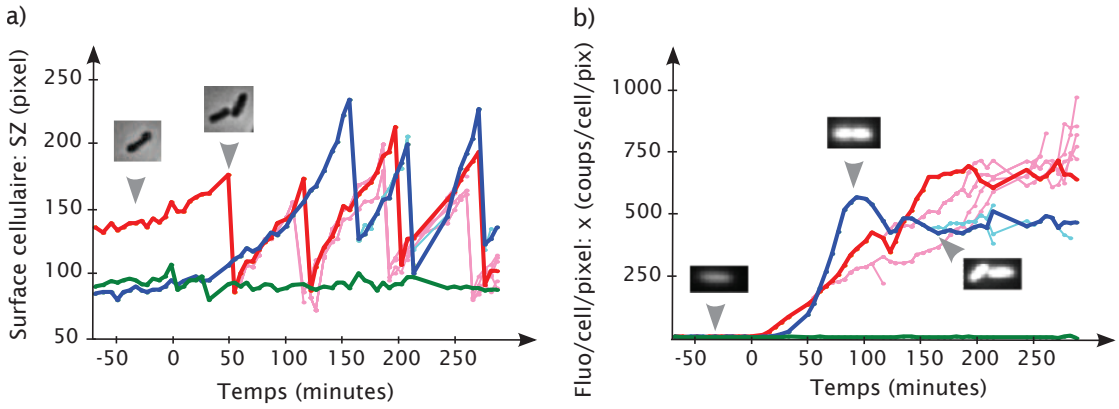


FIGURE IV.2.5.: Suivi individuel de trois cellules *GyrA* (courbes foncées) lors de la transition entre état stationnaire et croissance exponentielle : le temps $t = 0$ min indique l'ajout de glucose au milieu de culture. a) La dynamique de la surface cellulaire. b) La dynamique de la concentration de fluorescence cellulaire. Le lignage associé à la cellule suivie est indiqué par les courbes claires de même couleur.

La courbe verte de la figure précédente suggère que, dans nos échantillons, certaines cellules sont en train de mourir. Nous avons vu dans le chapitre 2 précédent que notre programme d'analyse d'images rejette ces cellules. Toutefois la méthode de "réjection" utilisée dépend de la sensibilité du scientifique [3], et les cellules en phase terminale sont parfois difficiles à distinguer. De plus, certaines cellules malades n'afficheront le phéno-

3. Les cellules malades qui sont détectées par le programme doivent être approuvées par le scientifique avant d'être rejetées.

type de mort cellulaire qu'en fin d'expérience [4]. Et nous ne rejetons que les cellules pour lesquelles le phénomène de mort cellulaire est déjà bien prononcé. Donc, malgré toutes nos précautions, nos données sont suceptibles de comporter des cellules mortes.

Ainsi sur la distribution de la figure IV.2.6 ci-dessous, les cellules suspectées mortes ont été rejetées par notre procédure de traitement d'image, mais le niveau de fluorescence cellulaire nul conserve une densité de probabilité qui ne suit pas la continuité de la distribution. Et nos données expérimentales sont encore relativement bien décrites par la loi de Gamma-Hill trouvée dans la figure IV.2.4 précédente.

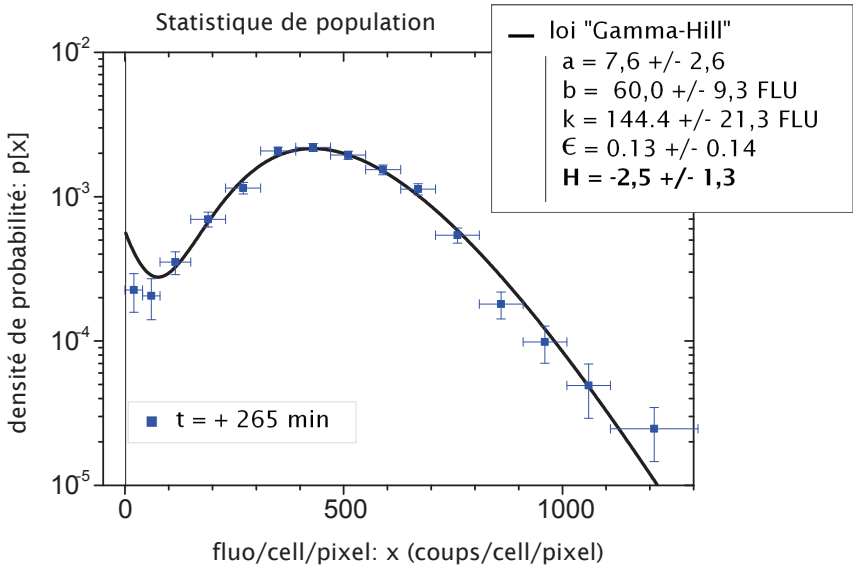


FIGURE IV.2.6.: Distribution de population en fin d'acquisition : les cellules *GyrA* sont imagées 265 min après l'injection du glucose, et celles qui sont suspectées d'être mortes sont éliminées lors du traitement d'images. La distribution (points bleus) est tracée sur 1218 cellules, et le binnage est le même que dans la figure IV.2.4 précédente. La courbe noire est la loi de Gamma-Hill retournée trouvée dans la figure IV.2.4 précédente.

Sélection des cellules qui se divisent

Pour s'assurer que notre densité de probabilité en 0 ne traduit pas le phénomène de mort cellulaire, nous avons rajouté un critère de réjection des cellules malades plus drastique. Ce dernier utilise le traçage des lignées cellulaires : toutes les cellules pour lesquelles aucun événement de division cellulaire est détecté, sont rejetées. Pour ce faire, nous avons mis au point une procédure Matlab qui permet de ne sélectionner, que les cellules, qui se divisent pendant la durée de l'expérience. Les données résultantes sont plus propres. Et

4. Pendant la durée de nos acquisition ($\approx$ 300 min), certaines cellules n'ont pas atteint un stade de dégradation cellulaire assez avancé pour afficher une diminution nette du niveau de gris sur les images en contraste de phase.

à présent, les distributions de population ne sont plus piquées en 0, comme illustré sur la figure IV.2.7 suivante.

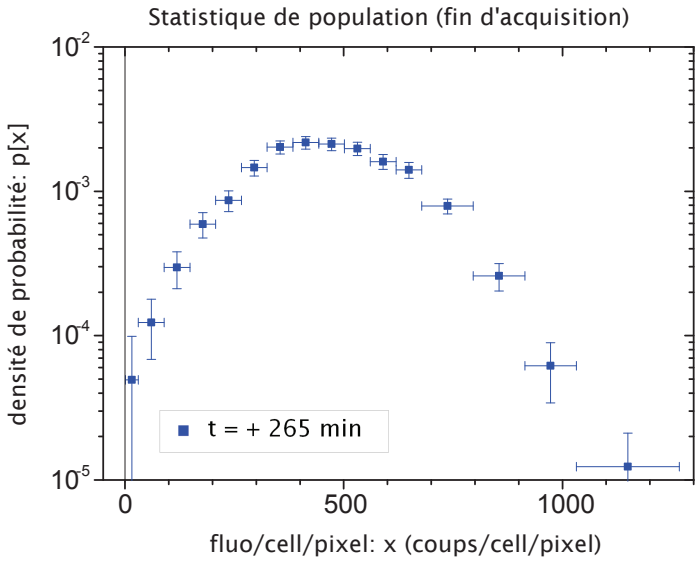


FIGURE IV.2.7.: Distribution de population en fin d'acquisition : les cellules *GyrA* sont imagées 265 min après l'injection du glucose, et celles qui sont suspectées d'être mortes sont éliminées lors du traitement d'images. La distribution (points bleus) est tracée sur 687 cellules, et le binnage est le même que dans la figure IV.2.4 précédente.

On note cependant que ce nouveau critère réduit considérablement le nombre de cellules saines. En effet, il rejette les cellules malades, mais aussi toutes les cellules en bonne santé qui se décollent du tapis de polylysine lors de la division cellulaire. De plus, les lignées cellulaires qui sont interrompues pour un temps d'acquisition (*i.e.* la cellule n'est pas considérée dans une image intermédiaire [5]) sont rejetées sur la tranche de temps qui ne contient pas l'événement de division cellulaire. En effet, notre programme d'analyse d'images caractérise les lignées interrompues par deux lignages différents [6] : l'un suit la cellule avant la discontinuité, et l'autre après. L'un des ces deux lignages contient l'événement de division cellulaire, alors que l'autre est traité comme une lignée à une seule génération. Ce dernier sera donc rejeté, même si il caractérise en réalité une cellule qui se divise. Ainsi, dans nos expériences sur cellules *gyrA*, le nombre de cellules analysées est réduit de $\approx 48\%$ pour les derniers temps d'acquisition, et de 20% pour les premiers temps d'acquisition. Dans des expériences où l'on part à la chasse aux cellules, ce n'est

5. Il arrive parfois qu'une cellule apparaîsse floue sur une image (léger décolement du tapis de poly-L-lysine), alors qu'elle est nettement imagée pour les autres temps d'acquisition (cellule collée de manière bien plane). Il arrive aussi que l'image de la cellule soit parasitée par un débris du milieu environnant, ce dernier apparaissant pour un seul temps d'acquisition. Dans ces cas, la cellule n'est pas analysée sur l'image défaillante, et la lignée cellulaire est interrompue.

6. Notre programme est conçu pour tracer les cellules à travers deux images directement consécutives : il n'est pas capable de pister une bactérie présente dans deux images, lorsque celles-ci sont séparées par une acquisition où la cellule n'est pas considérée.

pas négligeable !

On note également que ce nouveau critère de rejection ne distingue pas les cellules mourantes des cellules qui seraient trop lentes au redémarrage pour se diviser pendant la durée de l'expérience. Sous cette hypothèse, les cellules retardées traduisent un aspect important de la réponse cellulaire. Et pourtant, elles sont rejetées de l'analyse.

Dans nos analyses, **nous considérerons seulement les cellules qui se divisent** pendant la durée de notre expérience : bien que le nombre de cellules considérées soit réduit, il permet encore de donner l'allure de la statistique, comme montré sur la figure IV.2.7. Et nous supposerons que le pic en 0 de la figure IV.2.6 précédente (distribution tracée sur toutes les cellules) est l'effet de cellules malades qui n'ont pas été détectées lors du traitement d'images.

Cette sélection sur les cellules qui se divisent sera appliquée sur toutes les expériences menées en dynamique.

IV.3. Résultats expérimentaux

IV.3.1. Reproductibilité des expériences

Les conditions standard du chapitre précédent ont été définies pour réaliser des expériences dans les mêmes conditions. Nous allons maintenant vérifier que ces conditions génèrent des expériences reproductibles d'un jour à l'autre. Pour ce faire, nous nous appuierons sur les expériences 1 et 2 : elles ont été réalisées (et analysées) dans les mêmes conditions. On note cependant que d'un jour à l'autre, l'intensité de la lampe fluorescente varie. En conséquence, les cellules de l'expérience 2 ne sont pas soumises au même éclairage que celles de l'expérience 1. Cela se traduit sur la figure IV.2.1-b précédente, par un léger décalage des niveaux de fluorescence. Pour s'affranchir de ces variations entre les deux expériences, nous utilisons la méthode de calibration vue au chapitre 2 (section II.2). Si notre protocole expérimental est reproductible, les deux expériences doivent maintenant générer les mêmes données.

Comportement moyen de la population de cellules

Dans un premier temps, nous étudions la reproductibilité sur le comportement moyen de la population de cellules : on considère non pas les cellules individuelles, mais l'ensemble des cellules imagées. Si les deux expériences sont reproductibles, le comportement moyen doit être identique dans les deux expériences. Pour comparer ce comportement moyen entre les deux expériences, nous nous sommes appuyées sur la concentration de fluorescence cellulaire $\langle x \rangle$: cette grandeur est moyennée sur toute la population de cellules. La dynamique de la moyenne est présentée sur la figure IV.3.1 ci-dessous, avec en rouge le comportement de la population de cellules de l'expérience 1, et en bleu, celui de la population de l'expérience 2. Les deux courbes sont tracées pour des cellules *GyrA*, et on voit qu'après calibration de l'intensité de la lampe fluorescente, les deux courbes sont confondues. Ainsi après de nombreux problèmes de reproductibilité, nous disposons enfin d'un protocole expérimental qui, en moyenne, est reproductible d'une expérience à l'autre : Hourra !

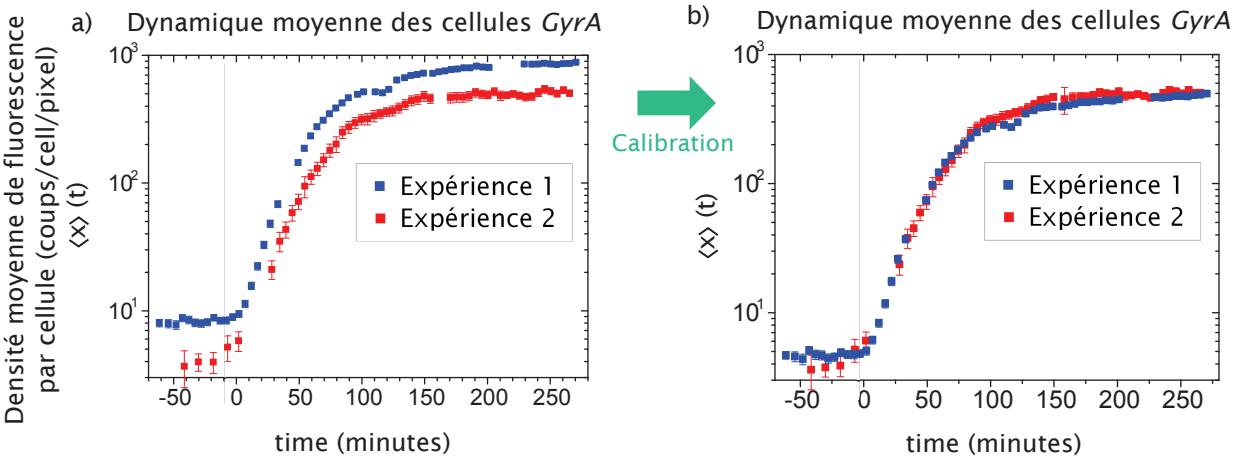


FIGURE IV.3.1.: Dynamique de la concentration moyenne de fluorescence cellulaire $\langle x \rangle$. Au temps $t = 0$ min, le glucose est injecté dans le milieu de culture et les cellules *GyrA* sont imagées toutes les ≈ 5 minutes. a) Données brutes. b) Données corrigées des variations d'excitation de fluorescence entre les deux expériences. On note que le temps d'injection du glucose est relevé à ± 2 min près. L'échantillon imagé dans l'expérience 1 comporte moins de cellules que celui de la deuxième expérience : ceci peut se voir sur les barres incertitudes qui sont plus larges dans l'expérience 1 que dans l'expérience 2. Les erreurs sur $\langle x \rangle$ sont données par l'erreur type de la moyenne.

Comportement individuel des cellules

Pour s'assurer que nos expériences sont reproductibles en terme de variabilité intercellulaire, on compare la statistique des deux populations de cellules imagées. Nous présentons sur la figure IV.3.2 ci-dessous la distribution obtenue pour les derniers temps d'acquisition. Et comme l'expérience 2 comporte peu de cellules [1], sa statistique possède de grandes barres d'incertitudes, avec des bins peu peuplés. Dans la limite de cette résolution, nous considérons que les deux distributions sont confondues : en fin d'acquisition, le comportement aléatoire des cellules de l'expérience 2 est similaire à celui des cellules de l'expérience 1. On note que ce résultat se retrouve sur les autres expériences, une fois que celles-ci ont été normalisées à l'intensité d'excitation de fluorescence utilisée dans l'expérience 1.

Pour comparer la statistique aux temps d'acquisition transitoires, on utilise le test de Kolmogorov-Smirnov vu au chapitre 3. Les résultats retournés par ce test sont présentés sur la figure IV.3.3 ci-après, avec à titre indicatif, le nombre de cellules imagées dans l'expérience 1. On note que ce nombre n'est pas correctement multiplié par la division cellulaire, car dans cette expérience, le nombre de champs de vue considérés diffère d'un temps d'acquisition à un autre [2]. Donc, entre deux temps d'acquisition, la taille de la

1. L'expérience 2 image au maximum 74 cellules (voir figure IV.3.3 plus loin), alors que dans l'expérience 1, on image ≈ 400 en fin d'acquisition.

2. Nous ne considérons pas les acquisitions qui sont défocalisées. Et dans cette expérience, il arrive que

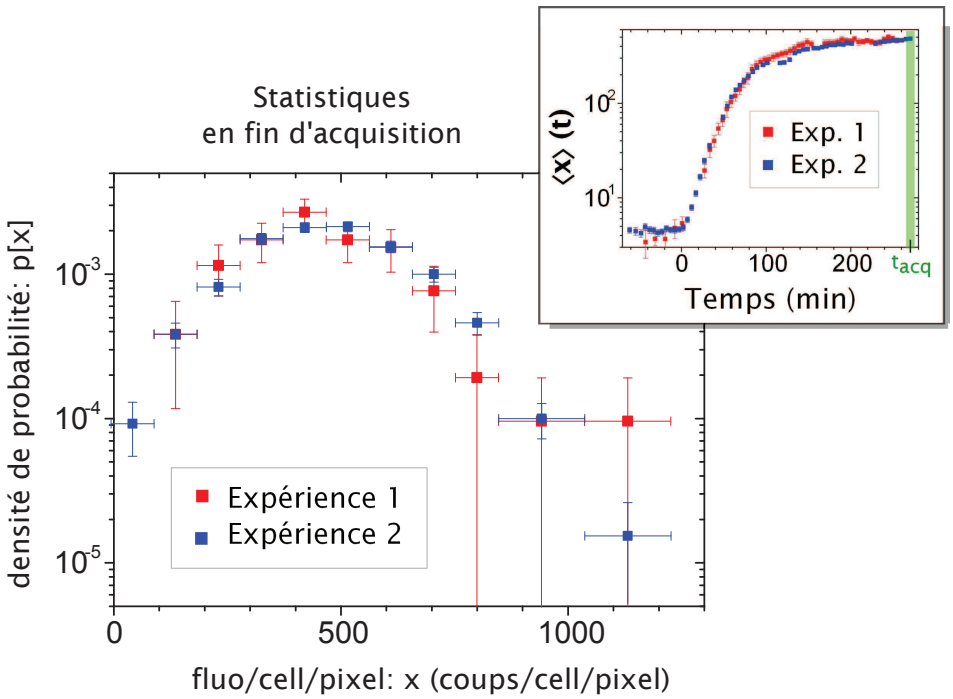


FIGURE IV.3.2.: Statistiques de population en fin d'acquisition ($t_{acq} = 265$ min) : le temps d'acquisition considéré est indiqué par le cadrillage vert sur les courbes moyennes de concentration de fluorescence cellulaire (courbe $\langle x(t) \rangle$ dans l'encadré du haut-droit). En bleu, la statistique tracée sur 377 cellules de l'expérience 2, et en rouge, la statistique tracée sur les 55 cellules de l'expérience 1. Le binnage est choisi sur l'expérience 2 selon la méthode décrite dans la section III.2.1 .

population diffère : d'une part, à cause de la multiplication cellulaire, d'autre part, à cause du nombre différent de champs de vu considérés. Lorsque le nombre de cellules est faible, le caractère statistique de la population est mal représenté. On obtient toutefois des valeurs de D qui restent inférieures à la valeur $D_{critique}$ (représentée par les points noirs sur la figure IV.3.3). Donc les données ne nous donnent aucune raison de rejeter l'hypothèse de distributions identiques : on considera que ces derniéres sont issues de la même loi de probabilité. Et comme les données de l'expérience 1 sont fortement erronées par les imprécisions statistiques, nous n'analyserons par la suite que les données de l'expérience 2.

Au final, nos expériences sur les cellules *GyrA* sont bien reproductibles. Et comme les cellules *Fis* et *GyrA-K* sont préparées et imagées de la même manière, ce résultat s'étend aux expériences menées sur les cellules *Fis* et *GyrA-K*.

pour une acquisition, les cellules d'un seul champs soit défocalisées. Dans ces conditions, la dynamique du champs de vue est interrompue, alors qu'elle reste inchangée pour les autres champs de vue.

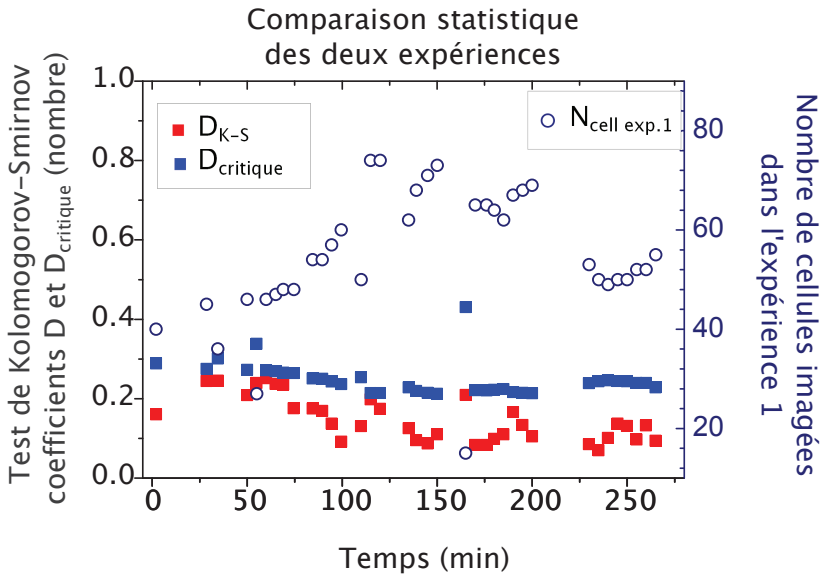


FIGURE IV.3.3.: Comparaison dynamique du comportement statistique des cellules de l'expérience 1, avec celles de l'expérience 2. Les similarités statistiques des deux populations sont données pour différents moments de la transition stress nutritionnel/croissance exponentielle. Elles sont indiquées par le coefficient D de Kolmogorov-Smirnov (en rouge) : celui-ci est calculé pour chaque temps d'acquisition sur les concentrations de fluorescence cellulaire. La valeur $D_{crititique}$ du test est calculé d'après le tableau III.2.1 pour un niveau de signification donné par : $\alpha = 0,01$. A titre indicatif, nous donnons pour chaque valeur de D calculée, le nombre de cellules considérées dans l'expérience 1 (ronds bleu-marine). Dans l'expérience 2, ce nombre est plus élevé : il est supérieur à 200 (non montré).

IV.3.2. Bruit en sortie du module : caractérisation

Maintenant que nos expériences sont reproductibles, on peut s'intéresser au bruit en sortie du module d'expression génique. Nos mesures de fluorescence cellulaire fournissent une indication sur la variable n en sortie du module (concentration de GFP/cellule), et elles permettent de mesurer la variabilité inter-cellulaire. Nous allons caractériser cette variabilité pour différents temps de la réponse au stress nutritionnel. Nous établirons ainsi la signature du bruit lors de la transition d'état. Par la suite, cette signature sera utilisée pour tester des modèles ; ces modèles détermineront la fonction de transfert du module, et ils permettront de remonter aux paramètres du module (voir figure IV.1.3 précédente).

Dans le modèle proposé dans la figure IV.1.4, le volume des cellules est uniforme dans la population et dans le temps. Dans nos expériences, ce n'est clairement pas le cas, comme le montre la figure IV.2.5 (voir section IV.3.1 précédente). Pour s'affranchir des différences de volume entre deux individus, et entre deux temps d'acquisition pour un même individu, on travaillera sur les concentrations de fluorescence cellulaire.

Dynamique des concentrations de fluorescence cellulaire individuelles

Notre programme d'analyse d'images permet de suivre les cellules dans le temps. On peut ainsi tracer la densité de fluorescence des individus lors de la réponse cellulaire. Les trajectoires ainsi obtenues sont représentées sur la figure IV.3.4 suivante, avec en plus, l'évolution temporelle de la surface cellulaire. Par mesure de clarté, nous n'avons représenté qu'une trentaine de trajectoires (notre expérience en contient ≈ 500).

Les courbes de surface montrent la croissance cellulaire induite par l'ajout de glucose, avec une variabilité qui apparaît dans la réponse cellulaire : les bactéries ne commencent pas à pousser au même moment.

Et si on s'intéresse maintenant aux courbes de fluorescence, on retrouve le comportement global prédit par l'étude de population (voir figure IV.3.1 de la section IV.3.1 précédente), mais on observe aussi la présence de fluctuations dans la réponse cellulaire. Et ces différences s'observent aussi sur les cellules qui suivent de manière globale le comportement de la population. En effet, les points de départ des différentes lignées cellulaires divergent, et cette diversité de comportement se retrouve en fin d'acquisition (voir les courbes bleues et rouges de la figure IV.3.4). A chaque instant, on observe des différences dans les niveaux de fluorescence étudiés. Une étude statistique sur l'intensité de fluorescence au sein de la population devrait permettre de caractériser ces fluctuations.

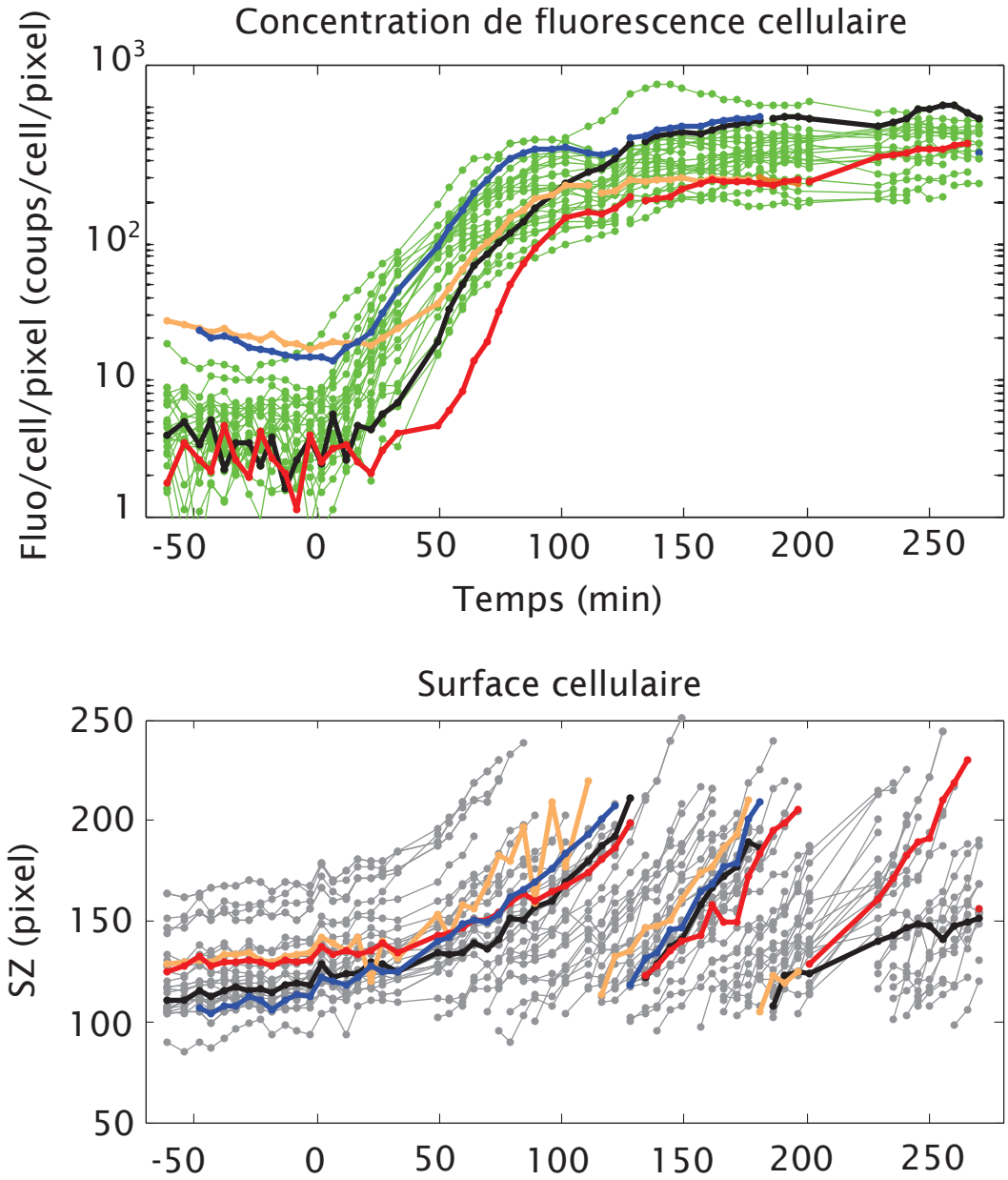


FIGURE IV.3.4.: Evolution temporelle de la concentration de fluorescence des cellules individuelles et de la surface qu'elles occupent, lors de la transition entre phase stationnaire et croissance exponentielle. Les cellules étudiées sont des bactéries *E. Coli* qui portent des rapporteurs plasmidiques révélant l'activité du gène *gyrA*. Et la transition d'état est initiée par l'apport de glucose ($t = 0$ min) aux cellules en pénurie de sources carbonnées. Chaque courbe correspond à une branche mère ou fille des lignées cellulaires. Et nous avons surligné quelques lignées en rouge, noir, bleu et beige. L'échelle des ordonnées est prise logarithmique pour les courbes de fluorescence, et linéaire pour la surface cellulaire.

Distributions en début d'acquisition

En début d'expérience, les contraintes expérimentales limitent le nombre de cellules imagées [3]. Ainsi, pour les temps qui précèdent l'injection de glucose, le nombre de cellules est faible (≈ 180 cellules pour chaque temps d'acquisition) et les statistiques générées sont imprécises (voir figure IV.3.5 suivante). Pour augmenter la résolution sur ces distributions, nous allons concaténer les données de plusieurs temps d'acquisition.

Concaténation des temps initiaux

Pour les temps précédant l'injection de glucose, les cellules sont à l'équilibre statistique : d'une part la concentration moyenne de fluorescence cellulaire ne varie pas (voir la figure IV.2.2 précédente), d'autre part la statistique reste inchangée. En effet, sur la figure IV.3.5 suivante, les trois distributions sont tracées pour des temps d'acquisition décalés de ≈ 20 min, et elles sont confondues. D'ailleurs, lorsqu'on compare les distributions deux à deux avec le test de Komogorov-Smirnov, on obtient des valeurs de D qui restent inférieures à la valeur critique $D_{critique}$ comme montré dans le tableau ci-dessous. L'hypothèse "H_0" qui suppose que les trois distributions sont équivalentes n'est donc pas rejetée, avec un niveau de signification de $99,9\%$. Par conséquent, nos données sont consistantes avec l'hypothèse d'équivalence statistique. Nous concluons de ces deux études que les deux échantillons ne diffèrent pas de manière significative : on considérera qu'ils sont issus de la même loi de probabilité.

Test de Kolmogorov-Smirnov : comparaison entre les statistiques mesurées à :	$D_{Ncells1,Ncells2}$	$D_{critique}$	Données consistantes avec l'hypothèse H_0 d'équivalence statistiques ?
$t = -8$ min et $t = -33$ min	0.09855	0,2212	oui
$t = -8$ min et $t = -54$ min	0.08067	0,2208	oui
$t = -33$ min et $t = -54$ min	0.09362	0,2226	oui

TABLE IV.3.1.: Valeurs du coefficient D retournées par le test de Kolmogorov-Smirnov pour les distributions de population mesurées à -54 minutes, -33 minutes, et -8 minutes. Ces distributions sont comparées deux à deux, et nous donnons la valeur critique $D_{critique}$ pour un niveau de signification correspondant à $\alpha = 0,001$: la valeur critique est calculée d'après la table III.2.1 de la section IV.3.2. Les valeurs de $D_{critique}$ restent supérieures aux valeurs de D retournées par le test : l'hypothèse H_0 qui stipule que les deux distributions suivent la même loi de probabilité est vérifiée.

3. Dans notre protocole expérimental, le suivi dynamique des cellules individuelles impose un nombre limité de cellules en début d'expérience. En effet, si la densité de cellules est trop élevée, les bactéries qui se multiplient envahissent rapidement le tapis de poly-L-lysine : elles deviennent vite indistinguables les unes des autres, et on ne peut pas extraire l'information d'une cellule individuelle pour les temps transitoires.

Pour les temps initiaux qui précédent l'injection de glucose, on est donc à l'équilibre statistique. On peut donc concaténer les données de plusieurs temps d'acquisition sans altérer les propriétés statistiques des cellules. En fusionnant ainsi les données relevées sur tous les temps qui correspondent au stress nutritionnel (12 temps d'acquisition), on augmente le nombre de cellules : on passe de ≈ 155 cellules pour un temps d'acquisition individuel, à 1857 cellules pour les temps d'acquisition fusionnés.

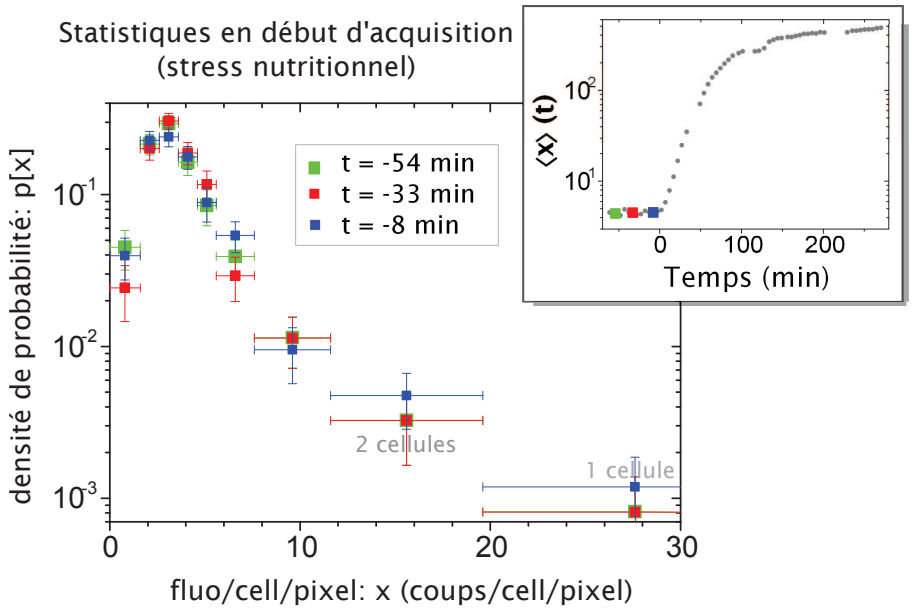


FIGURE IV.3.5.: Distributions tracées sur une population de cellules *gyrA*, en condition de stress nutritionnel, imagée à 3 temps différents indiqués par des points colorés sur la courbe moyenne de concentration de fluorescence cellulaire (courbe $\langle x(t) \rangle$ dans l'encadré du haut-droit) : ils sont compris entre -62 min et $t = 0$ min. Le binnage a été choisi manuellement (voir la section III.2.1) : la taille des bins proches de 0 est rétrécie pour faire ressortir les détails, sans pour autant perdre la résolution sur $p(B_i)$ (on a plus de 10 cellules par bins proches de 0). Les trois distributions sont confondues. On est donc à l'équilibre statistique avec $dp(x,t)/dt = 0$. On note que les derniers bins de la statistique sont très imprécis : ils ne sont peuplés que par une ou deux cellules, ce qui correspond à $\approx 0,6\%$ de la population imagée. Par conséquent, ces bins ne seront pas considérés dans les ajustements numériques.

Déconvolution de l'autofluorescence

On peut maintenant tracer une statistique plus précise. Celle-ci est présentée en bleu sur la figure IV.3.6 suivante. Elle traduit nos conditions initiales de stress nutritionnel et elle renferme 2 informations :

- ◇ La variabilité inter-cellulaire d'autofluorescence avec :
  - Le bruit d'autofluorescence brute.
  - Le bruit de la caméra.
- ◇ La variabilité inter-cellulaire dans la quantité de rapporteurs fluorescents.

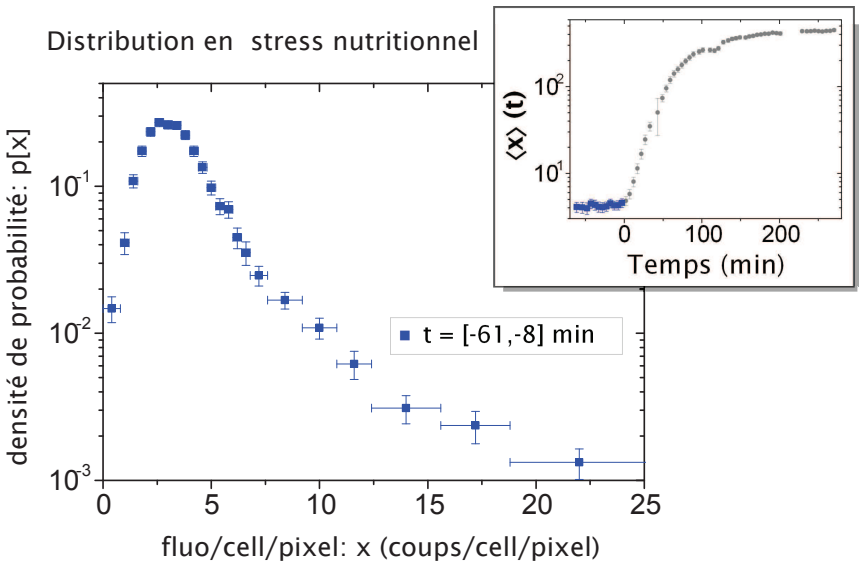


FIGURE IV.3.6.: Distribution tracée sur une population de cellules *gyrA* qui ne disposent pas encore de glucose (conditions de stress nutritionnel). Les données de fluorescence relative à un même champ de vue sont comptées plusieurs fois, chaque fois correspondant à un temps d'acquisition différent. Ces temps concaténés sont indiqués par les points bleus sur la courbe moyenne de concentration de fluorescence cellulaire (courbe $\langle x(t) \rangle$ dans l'encadré du haut-droit). Le binnage de la distribution a été choisi manuellement à partir du binnage donné dans la méthode du chapitre 2 : la taille des bins proches de 0 est rétrécie pour faire ressortir les détails, sans pour autant perdre la résolution sur $p(B_i)$ (on a plus de 38 cellules par bins proches de 0).

On ne s'intéresse qu'à la variabilité dans la quantité de rapporteurs : on veut sonder le bruit d'expression génique relatif au promoteur du gène *gyrA*. Pour accéder à l'information d'intérêt, il faut donc corriger la distribution observée du bruit d'autofluorescence. La correction se fait par déconvolution de la gaussienne d'autofluorescence. Cette opération se réalise en deux étapes :

1. Etape préliminaire :

   a) La statistique observée est corrigée de l'autofluorescence par déconvolution itérative avec l'algorithme de Lucy-Richardson qui utilise le critère de régularisation de Tikhonov (cf section III.3). La distribution résultante (statistique d'intérêt) est représentée en rouge sur la figure IV.3.7 ci-après, pour une gaussienne d'autofluorescence de paramètres $x_c = 2.0$ et $\sigma = 0.7$.

   b) On cherche à décrire notre statistique de population par une distribution gamma. Et les courbes de la figure IV.3.7 montrent que cette distribution (courbe bordeau) suffit à décrire la statistique d'intérêt (distribution rouge) : les deux courbes présentent des différences, mais comme la distribution rouge est erronée par la régularisation avec des incertitudes (non calculées) qui sont grandes, on peut considérer dans une première approximation que les deux

courbes décrivent la même loi de probabilité.

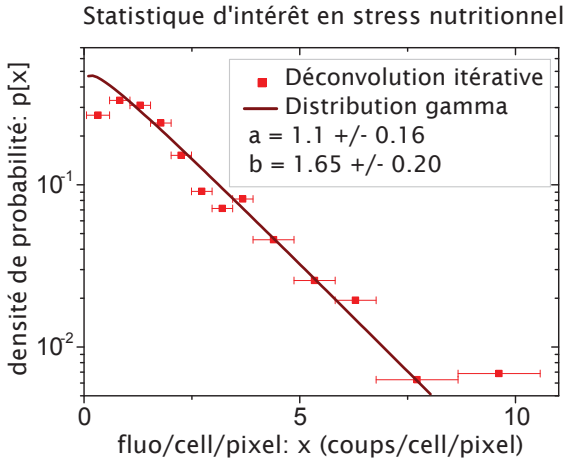


FIGURE IV.3.7.: Statistique d'intérêt en conditions de stress nutritionnel ($t < 0$min) : déconvolution itérative (algorithme de Lucy-Richardson) avec critère de régularisation de Tikhonov (points rouges). Seuls les 17 premiers points de la statistique observée sont considérés dans la déconvolution (les 4 autres points possèdent de grandes barres d'incertitudes) ; et la gaussienne d'autofluorescence utilisée pour la déconvolution a été prise avec les paramètres $x_c = 2.0$ et $\sigma = 0.7$. En bordeaux, l'ajustement numérique de la statistique rouge par une distribution gamma : $R^2 = 0.9666$ et $\chi^2 = 0.05694$ (les deux derniers points qui possèdent de grandes incertitudes en x ne sont pas considérés par l'ajustement).

c) Un ajustement numérique (en bordeau sur la figure IV.3.7) de la statistique d'intérêt donne alors un ordre de grandeur (grossier) pour les paramètres a et b du modèle gamma : $a_{grossier} \approx 1.1$ et $b_{grossier} \approx 1.65$.

2. Maintenant qu'on a une idée de la grandeur des paramètres a et b, on peut chercher à les affiner. En ce sens, nous avons développé une sous-routine Matlab qui teste les valeurs de a et b proches de $a_{grossier}$ et $b_{grossier}$. La sous routine explore l'espace $\vec{a} \times \vec{b} = [\ 0 \quad 3\ ] \times [\ 0 \quad 6\ ]$ par pas de 0.1. Pour chaque valeur de a et b testée :

   a) Elle trace la distribution gamma associée. Cette distribution est binnée sur des bins de taille fixe : $2\triangle_b = 0,2$.

   b) Elle convolue cette distribution discrète par l'autofluorescence. Pour ce faire, la sous routine calcule la matrice d'autofluorescence (voir III.3 [4]) et elle utilise l'équation de convolution matricielle III.3.3.

   c) Elle évalue la ressemblance entre la distribution convoluée ainsi obtenue, et la distribution observée. Cette ressemblance est évaluée par le $\chi 2$, avec le même poids accordé à chaque point expérimental [5].

4. Ici, la distribution d'intérêt est binnée sur des bins de taille 0.1. La distribution observée qu'elle génère (par convolution à l'autofluorescence) s'appuie sur le même binnage.

5. Le $\chi 2$ est ici normalisé au degré de liberté de l'ajustement (comme dans le logiciel OriginPro.6.1).

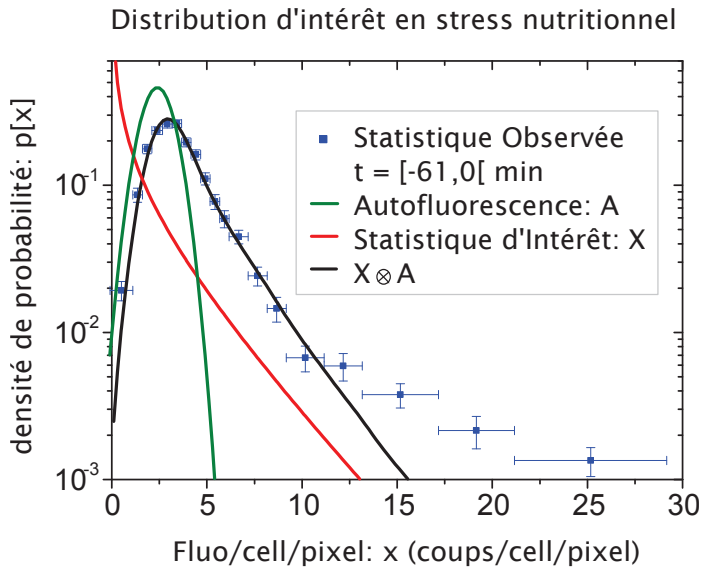


FIGURE IV.3.8.: Distribution de population (en bleu) tracée sur 1857 cellules *GyrA* en conditions de stress nutritionnel ($t < 0$min). En vert, l'autofluorescence cellulaire (gaussienne de paramètres $x_c = 2.4$ et $\sigma = 0.9$). En rouge, la statistique d'intérêt. Cette dernière est supposée suivre une distribution gamma de paramètres a et b. Les paramètres de cette distribution hypothétique ont été ajustés par une subroutine Matlab : ce sont ceux qui décrivent le mieux la statistique bleue après convolution par l'autofluorescence. La distribution rouge ainsi trouvée génère par convolution à l'autofluorescence la courbe noire : celle-ci correspond à la statistique observée et elle lisse très bien la distribution bleu observée ($\chi 2 = 0.02$). Les paramètres a et b utilisés pour tracer la courbe rouge valent ici : $a = 0.32$ et $b = 3.5$. Ce sont les valeurs qui minimisent le $\chi 2$. Le $\chi 2$ est calculé sur les 15 premiers points de la statistique observée sans tenir compte du 1^{er} point : les points qui ne sont pas considérées sont incertains (cf légende de la figure IV.3.5 précédente), et notre programme d'ajustement numérique ne considère pas les incertitudes en x.

On sélectionne alors, parmi toutes les valeurs testées , celles dont le $\chi 2$ est proche de 1 : $\chi 2 \in [0,9\,;\,1,1]$[6]. Sur ces couples (a, b) de valeurs sélectionnées, on calcule les moyennes a_m et b_m et les écarts type associés (σ_a et σ_b). Les paramètres de la distribution gamma (distribution d'intérêt) qui permettent de décrire la statistique observée seront pris comme suit : $a = a_m \pm \sigma_a$ et $b = b_m \pm \sigma_b$. Cette opération de déconvolution nécessite de fixer les paramètres de la gaussienne d'autofluorescence. Nous avons vu précédement, que ces paramètres sont déterminés avec des incertitudes non négligeables (incertitude de $\approx 30\%$ sur x_c). Dans ces barres d'incertitude, nous choisirons la valeur qui optimise la déconvolution dans l'espace discret $\vec{a} \times \vec{b} = [\ 0 \quad 3\ ] \times [\ 0 \quad 6\ ]$. Nous prendrons ainsi la valeur qui fournit le meilleur $\chi 2$ dans l'étape 2 de notre programme de déconvolution. Au final, la gaussienne d'autofluorescence utilisée dans la déconvolution sera donnée par les paramètres : $x_c = 2.4$ et $\sigma = 0.9$. Nous donnons sur la figure IV.3.8 suivante le résultat de

6. Pour le $\chi 2$ normalisé au degrè de liberté, un bon ajsutement sera caractérisé par un $\chi 2$ égal à 1

la déconvolution avec, en bleu, la distribution mesurée sur nos cellules *GyrA*, en vert, la gaussienne d'autofluorescence, et en rouge, la statistique d'intérêt. Cette dernière suit un modèle de distribution gamma ; les paramètres de ce modèle ont été déterminés par notre sous routine de déconvolution. Ces paramétres décrivent correctement le bruit mesuré, comme le montre la courbe noire : une fois convolué à l'autofluorescence (courbe noire), le modèle permet de retrouver la statistique observée (en bleu).

Conclusion Par la suite, nous considérons qu'en condition de stress nutritionnel, la variabilité inter-cellulaire des cellules *GyrA* suit une distribution gamma de paramètres :

$$\begin{cases} a_0 = 0.45 \pm 0.1 \\ b_0 = 3.5 \pm 0.8 \text{ coups/cell/pixel} \end{cases}$$

La même conclusion se retrouve lorsque, au lieu de regrouper les temps initiaux pour calculer la statistique d'intérêt (concaténation des temps initiaux), on calcule la statistique d'intérêt sur chacun des temps initiaux, et on moyenne ces résultats individuels. En effet, dans cette deuxième approche, on obtient des paramètres qui pour chacun des temps initiaux, avoisinent les valeurs trouvées par la première approche (voir figure IV.3.9 ci-dessous).

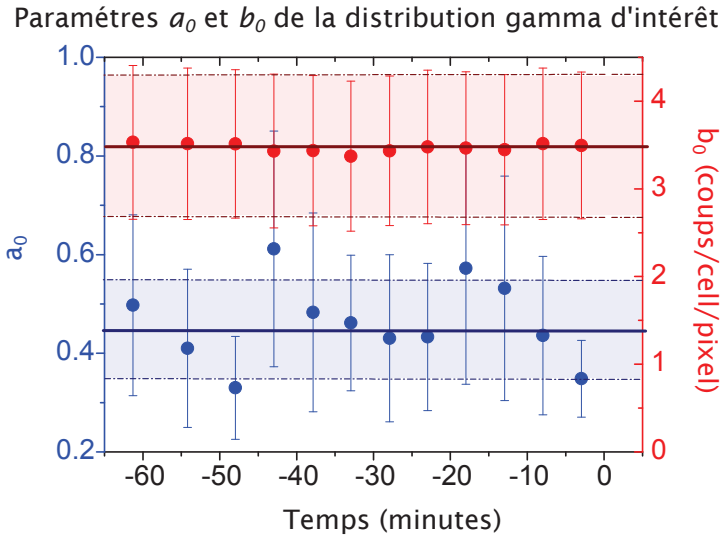


FIGURE IV.3.9.: Résultat de la déconvolution pour les temps précédent l'injection de glucose. L'ajout de sources carbonées est indiqué par le temps $t = 0$ min, et les cellules étudiées sont des cellules *GyrA*. En ronds rouges et bleus : les paramètres a et b de la distribution gamma d'intérêt calculés pour chacune des acquisitions antérieures à 0 minutes. En bordeau et bleu marine : les mêmes paramètres, mais calculés cette fois-ci sur les acquisitions initiales concaténées. Dans ce cas, les barres d'erreur sont indiquées par les lignes discontinues.

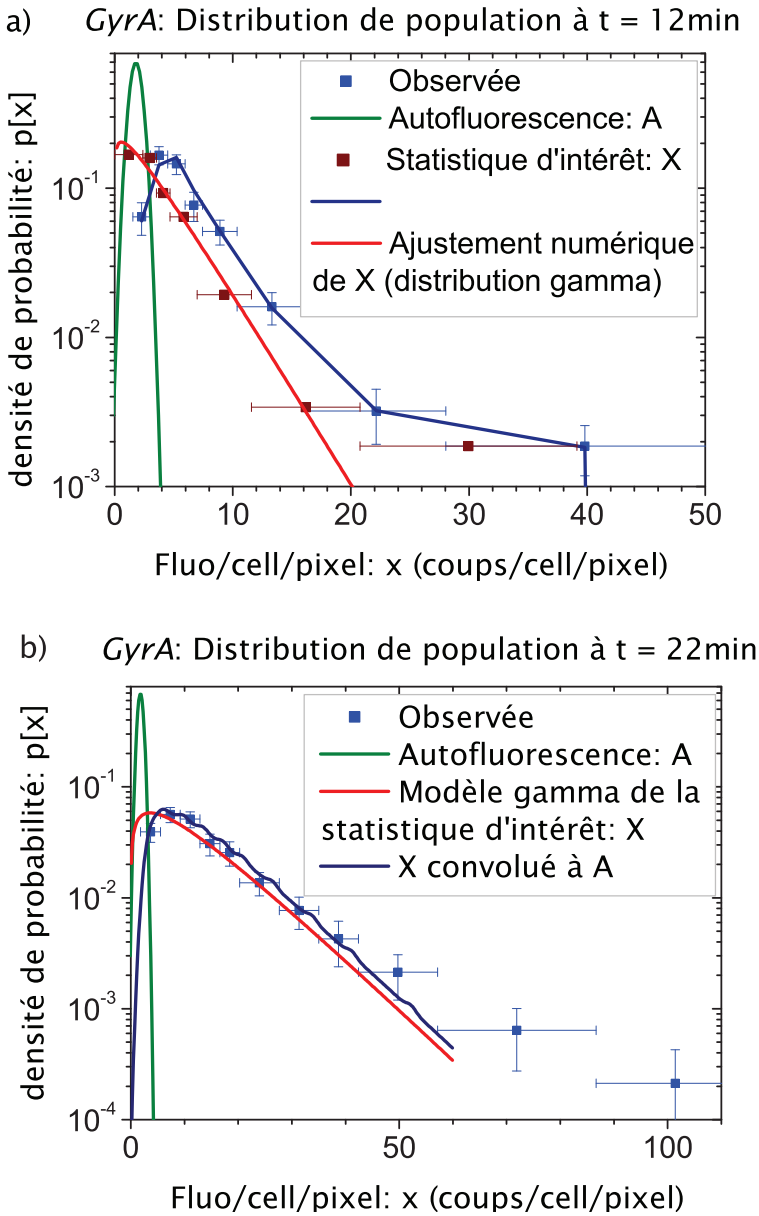


FIGURE IV.3.10.: Résultat de la déconvolution à $t = 12$ min (a) et $t = 22$ min (b) pour des distributions de populations observées sur 159 cellules *GyrA* (points bleus). La gaussienne d'autofluorescence utilisée est en vert. a)- Etape préliminaire de la déconvolution pour $t = 12$ min. En bordeaux, la statistique d'intérêt obtenue par déconvolution itérative de l'autofluorescence cellulaire. Et en bleu marine la même statistique, une fois convoluée à l'autofluorescence cellulaire. Enfin, en rouge, l'ajustement numérique de la statistique d'intérêt par une distribution gamma : $a = 1.17 \pm 0.16$, $b = 3.29 \pm 0.34$, $\chi 2 = 0.025$ et $R^2 = 0.99104$. b)- Résultat final de la déconvolution pour $t = 22$ min. En rouge, la statistique d'intérêt : une fois convoluée à l'autofluorescence cellulaire, elle génère la courbe bleu marine qui lisse la distribution observée. La statistique d'intérêt est modélisée par une distribution gamma de paramètres a et b. Et nous présentons les valeurs des paramètres a et b qui lissent le mieux la distribution observée (en terme de $\chi 2$), à savoir : $a = 1,4$ et $b = 9,0$. Ces paramètres optimums sont donnés à $\pm 0,1$ près.

Distributions en début de réponse

Une fois le glucose injecté dans le milieu de culture ($t = 0$ min), les cellules évoluent vers un nouvel état, et le signal de fluorescence cellulaire augmente (voir figure IV.2.2 précédente). Cette augmentation se fait graduellement, et pour les premiers temps d'acquisitions qui suivent l'ajout de glucose le signal de fluorescence reste faible. Dans ces conditions, la contribution d'autofluorescence n'est pas négligeable, et ici aussi, les statistiques mesurées doivent être corrigées.

Temps d'acquisition qui suivent l'ajout de glucose (t< 50 min)

Les corrections sont faites comme expliqué dans la section précédente : comme l'illustre la figure IV.3.10-a présentée sur la page précédente, la statistique d'intérêt obtenue par déconvolution itérative (algorithme de Lucy-Richardson avec régularisation de Tikhonov) peut être décrite par une distribution gamma. Nous présentons, sur la même figure IV.3.10 une illustration de la correction finale du bruit d'autofluorescence (sous figure b).

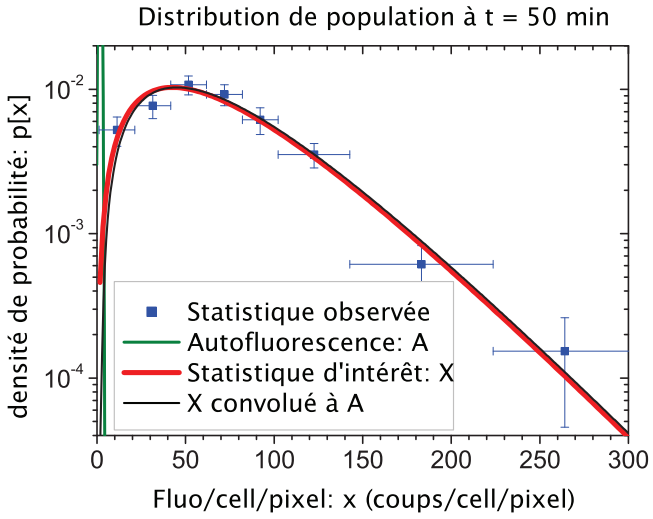


FIGURE IV.3.11.: Statistique de population mesurée sur 161 cellules *GyrA* à $t = 50$min (distribution bleue). En rouge, la statistique est ajustée numériquement par une distribution gamma. Puis en noir, l'ajustement (courbe rouge) est convolué à l'autofluorescence cellulaire (gaussienne verte de paramètre $x_c = 2.4$, $\sigma = 0.9$). Les courbes rouges et noires sont confondues : l'opération de convolution par l'autofluorescence ne déforme pas la statistique.

Temps d'acquisition postérieurs à 50 min

Pour les temps d'acquisition postérieurs à $t = 50$ min, la densité de fluorescence cellulaire est importante (voir figure IV.3.1 de la section IV.3.1 précédente). Par conséquent, les distributions sont maintenant discrétisées sur des bins larges, et le bruit d'autofluorescence est noyé par la largeur des bins (voir figure IV.3.11). Dans ces conditions, ce dernier ne dé-

forme pas la distribution d'intérêt : les statistiques observées et d'intérêt sont confondues, comme montré sur la figure précédente. Il n'est pas donc pas nécessaire de déconvoluer l'autofluorescence : la correction sera effectuée par simple soustraction du niveau moyen d'autofluorescence ($x_c = 2,4$).

Distributions en fin d'acquisition

Après un certain temps de réponse, la densité moyenne de fluorescence cellulaire n'évolue plus (voir figure IV.3.1 de la section IV.3.1 précédente) : le système atteint un nouvel état de phase stationnaire (au sens physique) qui est caractérisé par la croissance exponentielle. Les cellules sont dans un nouvel état, et la variabilité inter-cellulaire dans la quantité de rapporteurs fluorescents n'est plus la même que celle mesurée en début d'acquisition.

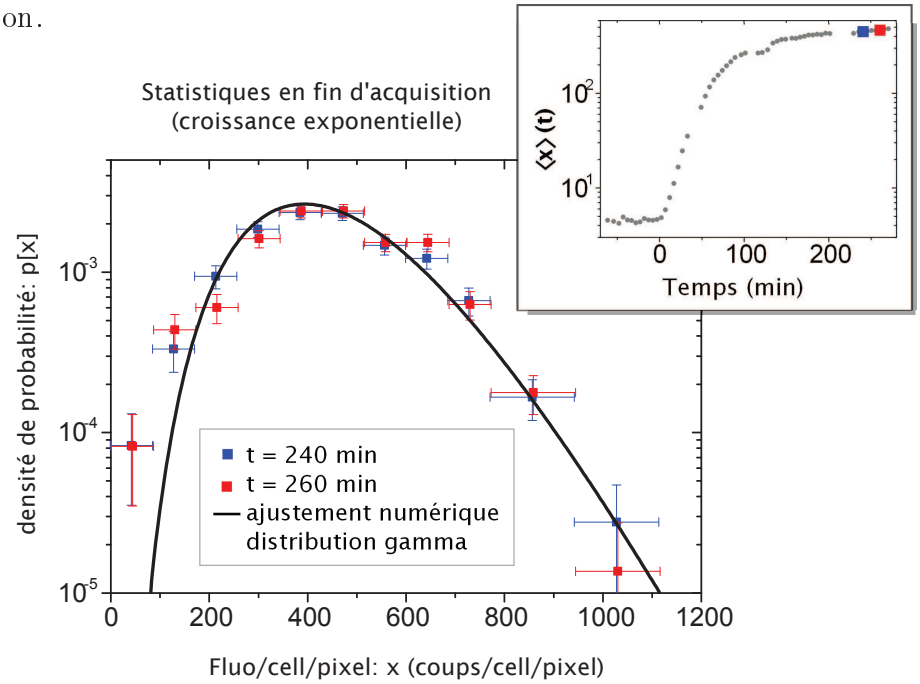


FIGURE IV.3.12.: Distributions de population tracées sur 421 (bleu) et 426 (rouge) cellules *GyrA*. En bleu, la distribution est tracée 240 minutes après l'injection de glucose, et en rouge, la distribution est tracée sur les mêmes cellules, imagées 20 minutes plus tard. Le binnage utilisé a été déterminé sur les données relatives au temps d'acquisition 240 min par la méthode décrite dans la section III.2.1 précédente.

Equilibre statistique de la phase exponentielle

En fin d'acquisition, les cellules atteignent l'équilibre statistique de la phase exponentielle. En effet, les distributions de populations tracées sur les cellules *GyrA* n'évoluent pas : les statistiques mesurées 240 min après l'injection de glucose, et 30 min plus tard sur les mêmes cellules, sont similaires, comme montré sur la figure IV.3.12.

Et ce résultat d'équivalence statistique se retrouve pour tous les temps d'acquisition compris entre 240 minutes et 260 minutes (non montré). D'ailleurs, le test de Kolmogorv-Smirnov qui compare les distributions mesurées deux à deux, retourne des valeurs de D qui restent inférieures à la valeur critique $D_{critique}$ (voir tableau IV.3.2).
On est donc bien à l'équilibre statistique à partir de 240 min.

Test de Kolmogorov-Smirnov : comparaison entre les statistiques mesurées à :	$D_{Ncells1,Ncells2}$	$D_{critique}$	Données consistantes avec l'hypothèse H_0 d'équivalence statistiques ?
$t = 240$ min et $t = 245$ min	0.02705	0,13539	oui
$t = 240$ min et $t = 250$ min	0.04976	0,13456	oui
$t = 240$ min et $t = 255$ min	0.05902	0,13416	oui
$t = 240$ min et $t = 260$ min	0.07029	0,13401	oui
$t = 245$ min et $t = 250$ min	0.0491	0,13554	oui
$t = 245$ min et $t = 255$ min	0.04802	0,13515	oui
$t = 245$ min et $t = 260$ min	0.06685	0,13499	oui
$t = 250$ min et $t = 255$ min	0.0328	0,13433	oui
$t = 250$ min et $t = 260$ min	0.03405	0,13417	oui
$t = 255$ min et $t = 260$ min	0.03056	0,13377	oui

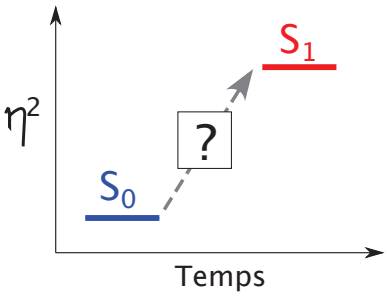
TABLE IV.3.2.: Valeurs du coefficient D retournées par le test de Kolmogorov-Smirnov pour les distributions de population mesurées entre 240 minutes et 260 minutes. Ces distributions sont comparées deux à deux, et nous donnons la valeur critique $D_{critique}$ pour un facteur de signification correspondant à $\alpha = 0,001$: la valeur critique est calculée d'après la table III.2.1 (voir chapitre IV.3.2 précédent). Les valeurs de $D_{critique}$ restent supérieures aux valeurs de D retournées par le test : les données sont consistantes avec l'hypothèse H_0 qui stipule que les deux distributions suivent la même loi de probabilité.

Loi de probabilité en fin d'acquisition

Nous proposons de décrire la variabilité intercellulaire en fin d'acquisition par une distribution gamma : c'est la distribution générée par le modèle théorique exposée au IV.1 en conditions d'équilibre statistique [7, 10]. Ce modèle décrit relativement bien nos données, comme le montre l'ajustement numérique de la figure IV.3.12 précédente : les valeurs de $\chi 2$ et R^2 retournées par l'ajustement sont de : $\chi 2 = 3.4$ et $R^2 = 0.82$. Cet ajustement a été réalisé sur toutes les statistiques mesurées entre $t = 240$ min et $t = 260$ min, et les paramètres a et b de l'ajustement valent :

$$\begin{cases} a_f = 8.02 \pm 0.35 \\ b_f = 56.00 \pm 2.70 \text{ coups/cell/(pixel)} \end{cases}$$

Dynamique de la distribution de population



Dans les paragraphes précédents, le bruit d'expression génique a été caractérisé pour des conditions d'équilibre statistique. Ces équilibres correspondent à deux états cellulaires différents : en début d'expérience, nos cellules sont en conditions de stress nutritionnel (S_0), et en fin d'acquisition, les cellules sont dans un stade avancé de croissance exponentielle (S_1) [7]. Pour ces deux états, la variabilité inter-cellulaire peut être décrite par une distribution gamma (voir la section I.1.3). Nous allons maintenant nous intéresser à la transition entre ces deux états stationnaires (au sens physique). Quelle est l'allure de la distribution ? Comment évolue la variabilité inter-cellulaire ?

Pour résoudre ces questions, nous ne pouvons plus nous appuyer sur des travaux théoriques présentés enI.1.3 : l'équation maîtresse de la section IV.1 n'est pas résolue pour un environnement cellulaire dynamique. Mais on peut déjà tirer des informations des courbes de la figure IV.3.4 précédente : ces courbes affichent une densité de fluorescence cellulaire qui augmente suite à l'ajout de glucose. Cela se traduira sur les distributions de population par une statistique qui se décale vers les niveaux élevés de fluorescence. C'est bien ce que l'on observe sur la figure IV.3.13 suivante. Cette figure représente des distributions de population pour différents temps de la réponse cellulaire. Les distributions sont tracées toutes les 30 min : ces temps sont indiqués dans l'encadré du coin droit, avec la densité moyenne de fluorescence cellulaire correspondante. Et plus on avance dans la transition d'état, plus le niveau moyen de fluorescence est élevé, plus la distribution se décale vers les valeurs élevées de x.

On constate aussi que les distributions restent monomodales, avec une bosse assymétrique qui s'écrase au fur à mesure qu'on avance dans la transition d'état. Et ce qui est surprenant, c'est que ces distributions peuvent être décrites par une distribution gamma : les courbes continues des figures IV.3.13 et sont des distributions gamma (convoluées à l'autofluorescence pour les temps inférieurs à 50 min [8]), et elles lissent très bien les statistiques mesurées sur nos les cellules. Ainsi, tout le long de la réponse, le bruit d'expression

7. Nous verrons dans la prochaine section qu'en fin d'acquisition, les cellules croissent exponentiellement, avec un taux de croissance qui reste constant.

8. Nous avons vu précédemment que pour ces temps, la distribution mesurée retranscrit la variabilité dans la quantité de rapporteurs (statistique d'intérêt), mais aussi le bruit d'autofluorescence (statistique parasite). Dans ce cas, la distribution gamma est ajustée sur la statistique d'intérêt pour retranscrire seulement le bruit d'expression génique (grandeur d'intérêt). Et pour pouvoir comparer cette distribution gamma avec la distribution mesurée, nous la présentons lorsqu'elle est, elle aussi, bruitée par l'autofluorescence. En d'autre terme, nous présentons la distribution gamma lorsqu'elle est convoluée par l'autofluorescence. De cette manière, on compare bien les mêmes grandeurs.

génique peut être décrit par une distribution gamma : la forme de la distribution est inchangée, seuls les paramètres a et b de la distribution gamma changent au cours du temps.

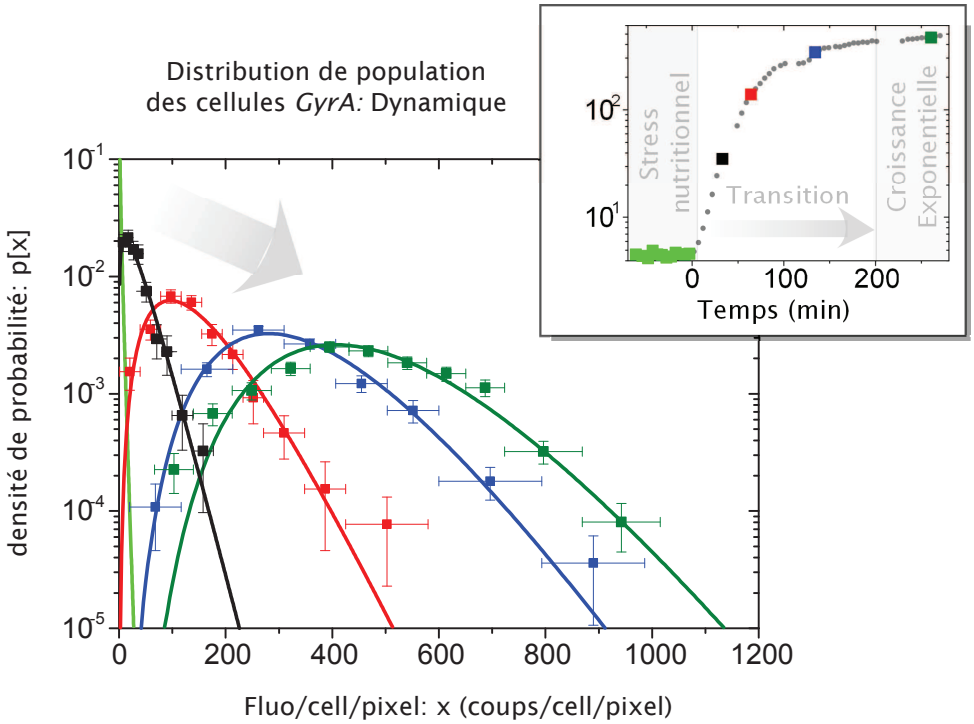


FIGURE IV.3.13.: Evolution de la variabilité inter-cellulaire lors de la transition d'état. Les cellules sondées sont des cellules *GyrA* : elles ont été imagées sous cube vert, avec une intensité d'excitation (lampe fluorescente) atténuée d'un facteur 0,08. Pendant nos acquisitions, elles passent d'un état de stress nutritionel à un état de croissance exponentielle, suite à l'ajout de glucose au milieu de culture ($t = 0$ min). Cette transition est indiquée dans l'encadré du coin droit, avec la densité moyenne de fluorescence cellulaire. Nous avons tracé la distribution de population pour pour différents instants de la réponse cellulaire : en vert clair, la distribution initiale de stress nutritionnel, en noir la distribution 33 minutes après l'injection de glucose, et en rouge, bleu et vert olive, les distributions (resp.) 60 min, 134 min et 260 min après l'injection de glucose. Ces temps sont indiqués dans l'encadré du coin droit par des points de même couleur que la statistique tracée. Les distributions sont ajustées numériquement par une distribution gamma (ligne continu).

Caractérisation du bruit en sortie du module

Au final, le bruit en sortie du module d'expression génique est entièrement décrit par les informations ci-dessous :

$$\left|\begin{array}{l} \text{forme} = \text{distribution gamma} \\ \text{paramètres de forme} = \begin{cases} a \\ b \end{cases} \end{array}\right.$$

EComme la forme de la statistique ne change pas au cours du temps, on pourra caractériser le bruit aux différents instants de la réponse par les paramètres a et b. Avec

cette description complète et à la fois simple, on peut facilement quantifier la dynamique du bruit lors de la réponse cellulaire. Leur dynamique est donnée sur la figure IV.3.14 ci-dessous, et elle sera utilisée pour tester le modèle proposé.

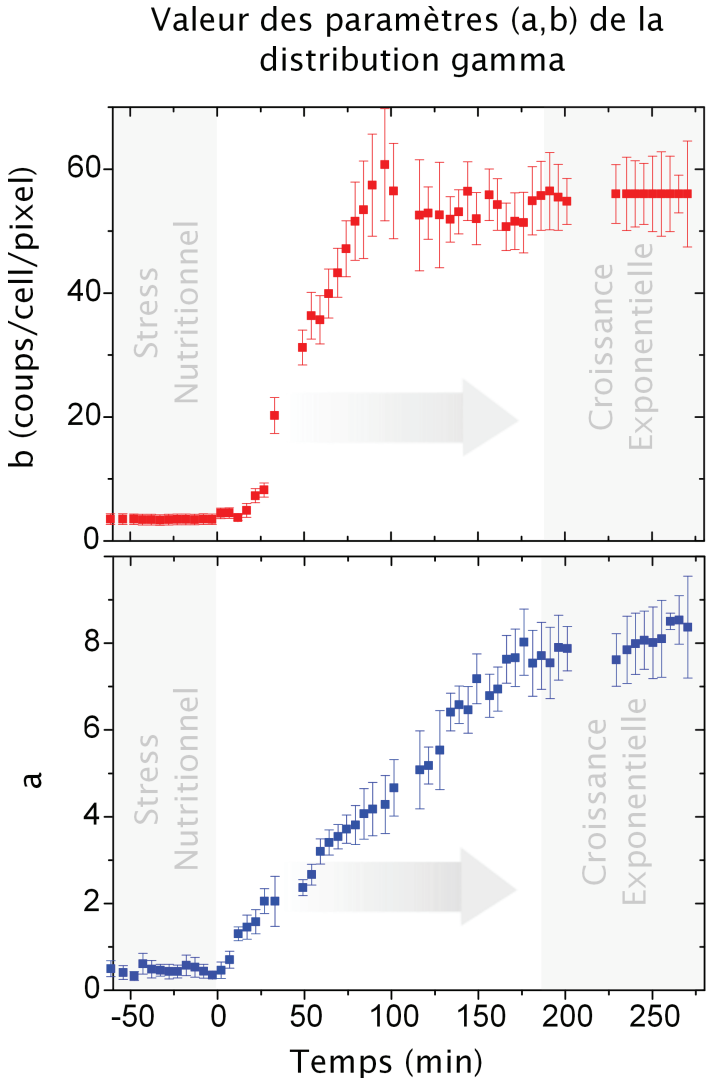


FIGURE IV.3.14.: Caractérisation de la variabilité inter-cellulaire lors de la transition d'état pour des cellules *GyrA*. Les cellules passent d'un état de stress nutritionel à un état de croissance exponentielle suite à l'apport de glucose ($t = 0$ min). Les fluctuations statiques générées en sortie du module sondé peuvent être décrites par une distribution gamma tout le long de la réponse. On caractérise alors ces fluctuations statiques par les paramètres a et b de la distribution : ceux-ci sont donnés par l'ajustement numérique de la statistique observée (modèle de distribution gamma utilisé par l'ajustement).

Par ailleurs le paramètre a de la distribution gamma est relié au niveau de bruit η^2_{total} par la relation suivante [9] : $\eta^2_{total} = \frac{1}{a}$. Nous donnons l'évolution de ce niveau de bruit

9. On rappelle que le niveau de bruit est défini par les moments du premier et deuxième ordre de la statistique : $\eta^2_{total} = \frac{\sigma_x^2}{\langle x \rangle}$. Et pour une distribution gamma, les paramètres $a$et b sont reliés par : $\langle x \rangle = ab$

dans la figure IV.3.15. On constate que le bruit en sortie du le module GyrA diminue considérablement lors de la transition. En effet, en condition de stress nutritionnel il vaut $\eta_{total} = (150 \pm 10)\%$, et après application du signal excitateur (ajout de glucose), il diminue pour atteindre une valeur stable vers 190 min, qui vaut $\eta_{total} = (35 \pm 2)\%$.

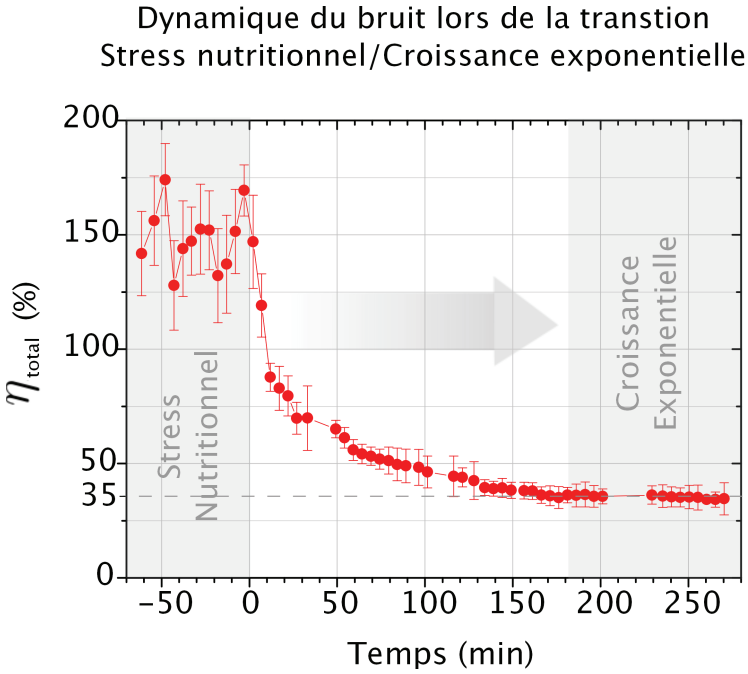


FIGURE IV.3.15.: Evolution temporelle du niveau de bruit pour les cellules *GyrA* lors de la transition stress nutritionnel/croissance exponentielle. Le temps 0 min indique l'application du signal excitateur : on ajoute le glucose au milieu de culture, et les cellules commencent à transiter vers l'état de croissance exponentielle. L'échelle des ordonnées est prise logarithmique.

Ce niveau de bruit η^2_{total} dépend de la concentration moyenne de fluorescence cellulaire $\langle x \rangle$. Pour un système qui ne comprend que des sources intrinsèques de bruit et qui est à l'équilibre statistique, la dépendance attendue est $\eta^2_{total} \sim \frac{1}{\langle x \rangle}$. Pour notre système qui transite entre deux points d'équilibre (donnés par les encadrés gris sur la figure IV.3.16), la dépendance est plus complexe même en présence de sources intrinsèques dominantes. Par exemple, nous présentons sur la figure IV.3.16, les courbes rouges et bleues simulées à partir d'un modèle à deux étages (transcription-traduction). On remarque que le niveau de bruit η^2_{total} est loin de suivre en régime dynamique une loi $\sim \frac{1}{\langle x \rangle}$. Les deux courbes sont générées à partir des taux de réactions optimisés dans le chapitre IV.5 : la courbe rouge est donnée par la solution $S_{m,p}$, et la bleue par la solution $S_{p,m}$ (voir tableau IV.5.1 et figure IV.5.5). Dans ces conditions, la dépendance non inverse observée dans nos expériences (courbe expérimentale en noir sur la figure IV.3.16) ne traduit pas forcément les sources

et $\sigma_x^2 = ab^2$.

extrinsèques, comme c'était le cas dans l'étude de Y.Taniguchi et al[21] où le bruit était mesuré à l'équilibre statistique (cf figure I.1.17).

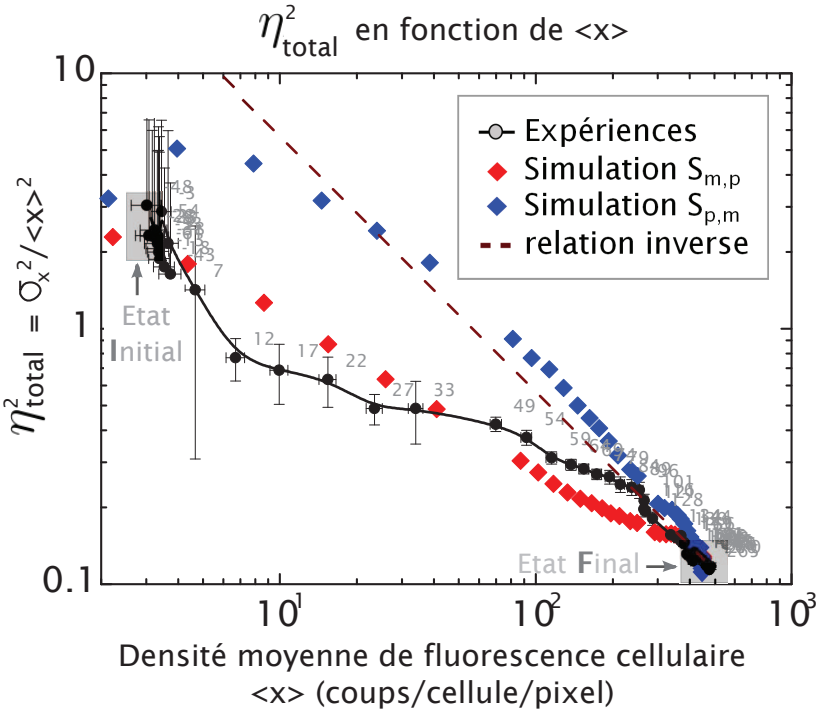


FIGURE IV.3.16.: Evolution du niveau de bruit dans des cellules *GyrA* selon la densité moyenne de fluorescence cellulaire. Cette densité augmente suite à la transition d'état stress nutritionnel/croissance exponentielle ; cette dernière est initiée par l'ajout de glucose au milieu de culture défaillant en sources carbonnées. En noir, les données expérimentales : chaque point est donné par l'état de la cellule à un instant de la transition d'état. Les temps ainsi sondés sont indiqués en minutes par les chiffres grisés. En rouge et bleu les données simulées à partir d'un modèle à deux étages (transcription-traduction) qui intégre les variations temporelles des taux de réaction. Les valeurs et variations temporelles des paramètres utilisés sont données par les solutions $S_{m,p}$ et $S_{p,m}$ du tableau IV.5.1. En pointillés bordeaux, un exemple de dépendance en $\frac{1}{\langle x \rangle}$ de η^2_{total}. L'échelle des ordonnées est prise logarithmique.

IV.4. Signaux d'entrée de la fonction de transfert : valeurs expérimentales

Maintenant qu'on a caractérisé le bruit en sortie du module d'expression génique, on peut tester le modèle proposé dans la section IV.1 : le bruit généré par le modèle sera comparé aux fluctuations statiques mesurées.
Le modèle proposé genère-t-il, lui aussi, une distribution gamma ? Si oui, les paramètres a et b de cette distribution évoluent-ils comme ceux de la figure IV.3.14 précédente ?

Comme cette étude comparative dépend de la valeur des signaux d'entrée, nous imposerons à la fonction de transfert des valeurs d'entrée compatibles avec nos données expérimentales. Dans le modèle de la figure IV.1, la fonction de transfert est donnée par les réactions d'expression génique considérées, et les signaux d'entrée sont les taux de réaction associés. Le modèle étudié contient donc quatres paramètres : les taux de production k_m et k_p, et les taux de dégradation γ_m et γ_p. Ces paramètres seront ajustés pour que la fonction de transfert proposée décrive, au mieux, nos données expérimentales. Comme nos expériences sont réalisées en dynamique pour des conditions différentes (stress nutritionnel/milieu de culture optimal), ces paramétres sont suceptibles d'évoluer dans le temps. Dans ces conditions, l'espace des paramètres devient vite important. Si le nombre de degrés de liberté du modèle est trop grand, l'ajustement aux données expérimentales est pauvre en conclusions. Dans cette étude, nous allons réduire cet espace pour restreindre le modèle aux conditions expérimentales. Nous nous appuierons ici sur des données expérimentales : nous verrons que nos expériences fournissent des informations sur les taux de réaction du module "GyrA".

IV.4.1. Taux de dégradation: composante induite par la dilution cellulaire

Les taux de dégradation γ_m et γ_p sont la combinaison de deux phénomènes:

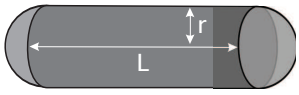
- ◇ La dégradation active de la protéine/ARNm: elle est caractérisée par le taux γ_{int} du tableau II.1.1.
- ◇ La dégradation par dilution cellulaire: lorsque les individus grandissent, le volume cellulaire augmente et la concentration de protéines/ARNm diminue dans les cellules. Cette dégradation sera caractérisée par un taux γ_{dil}.

Dans cette partie, nous ne nous intéressons qu'à la dégradation par dilution cellulaire. On dispose ici des images acquises en contraste de phase: celles-ci fournissent une information sur la taille des cellules, avec des variations de taille dans le temps qui traduisent la croissance cellulaire. Elles permettent donc de déterminer le taux de dilution qui caractérise la croissance de nos individus. Et comme dans le modèle d'expression génique proposé (voir section IV.1) les taux de réaction sont déterministes, nos valeurs expérimentales doivent être moyennées sur toute la population de cellules avant d'être inclues dans le modèle. On peut donc s'appuyer sur le comportement de la population pour déterminer γ_{dil}. Nous verrons cependant que les propriétés de notre population fournissent une information incomplète sur γ_{dil}. Pour palier à ce manque d'information, nous nous pencherons sur les données individuelles: nous verrons que celles-ci permettent de connaître le taux moyen de dilution à chaque instant.

Comportement de la population: Volume total occupé par les cellules

La croissance cellulaire peut être décrite, en moyenne, par la dynamique du volume total occupé par nos cellules.

Dans nos expériences, les bactéries sous forme de bâtonnets occupent chacune un volume V_i donné par:



$$V_i = \frac{4\pi r^3}{3} + \pi r^2\Big(L - 2 \times r\Big)$$

où r et L désignent respectivement le rayon et la longueur du cylindre-bâtonnet. On peut donc mesurer le volume total V occupé par nos cellules: $V = \sum_{\text{cellule i}} V_i$.

L'évolution temporelle de ce volume total est donnée dans la figure IV.4.1 suivante (courbe bleue), avec le nombre de cellules individuelles sur lesquelles ce volume est calculé (ronds bleu marine). On remarque que ce nombre n'est pas correctement multiplié par la division cellulaire: dans le traitement d'images, les cellules ne sont pas toutes considérées aux temps d'acquisitions suivants [1]. Par conséquent, la dynamique du volume total occupé nos cellules détectées reflète mal la croissance cellulaire. Ceci est particuliérement vrai

1. Ceci se produit par exemple lorsqu'une cellule se décolle du tapis de polyLysine après la division cellulaire, ou encore lorsque la cellule fille empiète sur une autre cellule.

au delà de $t \approx 100$ min (zone grisée sur la figure IV.4.1). Dans ces conditions, nous ne considérons que la partie de la courbe bleue antérieure à 90 min (zone non grisée sur la figure IV.4.1).

Ce morceau de courbe est comparé à l'absorbance mesurée dans les expériences en population des biologistes (courbe orange sur la figure IV.4.1): l'absorbance est mesurée sur une culture de bactéries en plaques multipuits[41], et elle est proportionnelle à la fraction de volume qu'occupent les cellules dans la solution. Comme les expériences des biologistes retranscrivent globalement le comportement de nos cellules (cf. figure IV.2.2 de la section IV.2.1 précédente), on pourra s'appuyer sur ces mesures d'absorbance pour tirer des premières conclusions sur le comportement de nos cellules. On constate ainsi qu'après un certain lapse de temps qui suit le rajout de glucose, nos cellules sont en régime de croissance exponentielle: la courbe bleue de notre population de cellules se superpose à la courbe orange des biologistes, qui est exponentielle.

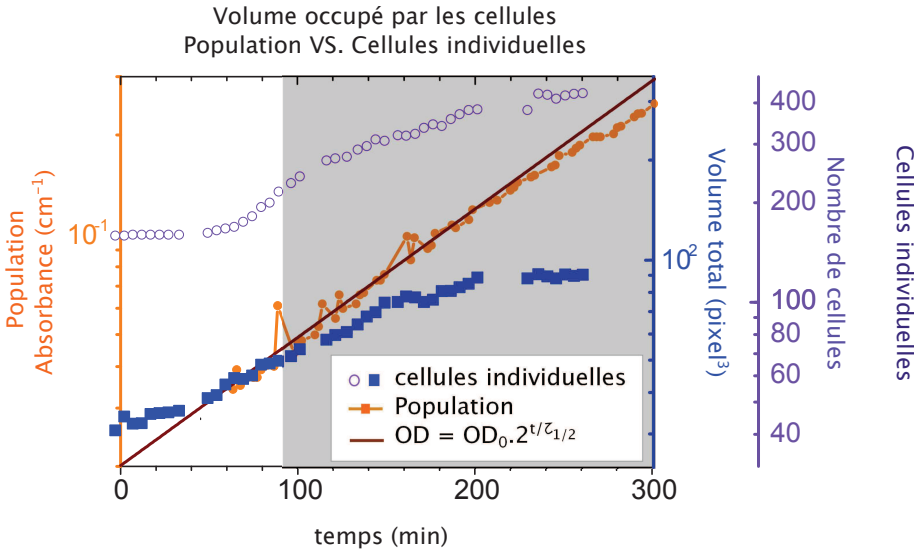


Figure IV.4.1.: Evolution temporelle de l'absorbance et du volume total lors de la réentrée dans la phase exponentielle de cellules E. Coli: comparaison des données reccueillies par les biologistes avec nos mesures. En orange, l'absorbance mesurée pour les cellules *GyrA* lors d'expériences en population. La courbe est lissée par la fonction $OD = OD_0 \times 2^{t/\tau_{1/2}}$ où OD_0 et $\tau_{1/2}$ représentent les deux paramètres de l'ajustement numérique: ils valent $OD_0 = 0.020 \pm 0.001$ cm^{-1} et $\tau_{1/2} = 77.7 \pm 2.9$ min. Les cercles violets indiquent le nombre de cellules individuelles considérées dans nos expériences, et la courbe bleue indique le volume total occupé par ces cellules. L'échelle des ordonnées est divisée par une valeur constante (dépendante du nombre de cellules initiales) pour faciliter la comparaison avec les biologistes. Sur l'échelle des ordonnées logarithmique, cette division correspond à un décalage de la courbe en ordonnée.

Cette figure permet de déterminer le temps de division ($\tau_{1/2}$) des cellules en pleine croissance exponentielle: dans ce régime, le volume total occupé par les cellules augmente selon:

$$V = V_0 \times 2^{\left(\frac{t}{\tau_{1/2}}\right)}.$$

Les courbes de la figure IV.4.1 permettent donc de déterminer le taux moyen de dilution cellulaire, $\langle \gamma_{dil} \rangle$. En effet, ce dernier est relié au temps de division par la relation ci-dessous:

$$\langle \gamma_{dil} \rangle = \frac{ln(2)}{\tau_{1/2}}$$

Taux de division collecté sur les expériences en population: Un ajustement numérique de la courbe d'absorbance mesurée en population (courbe orange) permet d'extraire la valeur de $\tau_{1/2}$. Et évidemment, les expériences en plaques multi-puits des biologistes présentent des variabilités d'une série d'expérience à une autre. Pour la courbe représentative de la figure IV.4.1 précédente, nous obtenons:$\tau_{1/2} = 77.7 \pm 2.9$ min.

Taux de division collecté sur nos expériences à cellules individuelles: Sur la courbe bleue de la figure IV.4.1 précédente, le nombre de points expérimentaux valides est faible, et on ne peut pas extraire d'information au delà de $t = 90$ min. De plus, le régime exponentiel n'est atteint qu'après $\approx$ 45min. Dans ces conditions, un ajustement numérique traduit mal le temps de division cellulaire. Qui plus est, on ne peut pas conclure pour les temps postérieurs à 90min. Ainsi cette étude de volume cellulaire ne permet pas de déterminer le temps de division dans nos expériences. Elle ne fournit qu'un ordre de grandeur pour $\tau_{1/2}$: la valeur obtenue dans les expériences en population des biologistes.

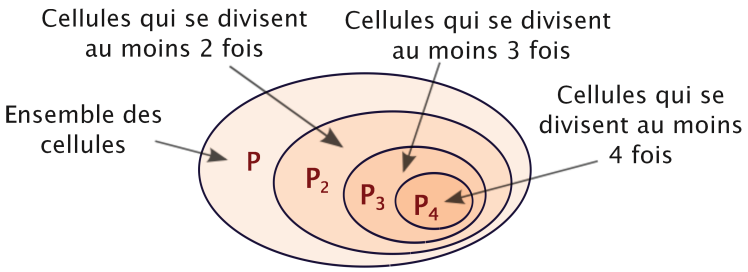
Taux de division calculé sur les évenements de division cellulaire

Pour déterminer notre temps de division avec plus de précision, nous nous penchons vers une méthode qui utilise, non pas les données moyennées, mais les informations collectées sur chaque individu. Le traçage des lignées cellulaires donne, pour chaque cellule, les changements de taille dans le temps et les évenement de division cellulaire. Ces informations individuelles, une fois moyennées, permettent d'extraire le temps moyen de division cellulaire.

Temps moyen de division cellulaire

Les temps $t_{division}$ relatifs à la $1^{ère}$, $2^{ième}$, $3^{ième}$ et $4^{ième}$ division cellulaire sont donnés par les événements de naissance détectés sur chaque lignée. A partir de ces données individuelles, on peut calculer pour chaque génération de la lignée, la durée τ_{cc} du cycle

cellulaire [2]. Ces durées, une fois moyennées sur toutes les cellules, correspondent au temps de division cellulaire $\tau_{1/2}$ déterminé dans les expériences en population. Les résultats obtenus sont présentés dans le tableau IV.4.1 pour l'ensemble des cellules, et des sous-populations d'individus. Ces sous-populations P2, P3 et P4 ne considérent que les cellules qui se divisent au moins (resp.) 2, 3 et 4 fois pendant l'expérience.



Sur l'ensemble des cellules sondées, seule une partie se divise plusieurs fois pendant la durée d'acquisition. C'est sur ces sous-ensembles de cellules P2, P3 et P4 qu'on peut extraire les temps de division pour les générations avancées.

Dans le tableau IV.4.1 qui mesure ces temps, on constate que la durée du cycle cellulaire diminue de génération en génération, suggérant que nos cellules sont de plus en plus rapides. Ce comportement traduit la réponse cellulaire étudiée : nos bactéries passent d'un état de croissance ralentie (stress nutritionnel) à un état de croissance exponentielle.
On est néammoins surpris de trouver des changements sur la durée du cycle cellulaire entre les nouvelles générations. En effet, comme l'indique la colonne $t_{division}$ du tableau IV.4.1, la première génération s'éteind vers 100 minutes. Les nouvelles générations se réfèrent donc à des ères postérieures à 100 min. Or, d'après les données des biologistes (courbe orange de la figure IV.4.1 précédente), la croissance cellulaire n'évolue plus au delà de ce temps. Il est donc inattendu de mesurer une accélération de la croissance pour les générations avancées de nos lignées cellulaires. On se demande alors si cette accélération est rééelle ou apparente. En effet, pour les générations avancées, les durées de cycle cellulaire sont calculées sur des sous-populations qui ne considèrent pas toutes les cellules. Cette perte de cellules dans les générations 2, 3 et 4 peut être induite par les erreurs de traçage cellulaire, mais elle peut aussi traduire des individus lents qui n'ont pas le temps de se diviser une nouvelle fois pendant l'expérience. Dans ce dernier cas, les sous population ne représentent que les cellules rapides, et on ne peut pas conclure sur le comportement de la population entière.

Pour savoir si les temps calculés dans le tableau IV.4.1 sont caractéristiques de la population entière de cellules (ou si au contraire, il décrivent une sous population plus rapide que les autres), nous avons étudié les changements de surface cellulaire. Comme illustré plus loin sur la figure IV.4.2, les changements de surface cellulaire permettent de

2. On note que p our la $1^{ère}$ génération, cette durée n'est pas connue : on manque ici d'information sur le début du cycle cellulaire. En effet, les cellules de la $1^{ère}$ génération "naissent" avant que l'acquisition d'images soit lancée.

différentier les cellules lentes des cellules perdues par le traçage cellulaire. En utilisant cette information de surface, on peut donc déterminer la cause des pertes cellulaires entre les générations.

durée du cycle cellulaire	$t_{division}$ (min)	$\langle\tau_{cc}\rangle$ (min)	N_{cells}
$1^{ère}$ génération (G_1)			
· population entière P	103.9 ±2.4	—	219
· sous-population P2	99.4±3.0	—	129
· sous-population P3	90.1±3.8	—	52
· sous-population P4	76.0±5.8	—	7
$2^{ième}$ génération (G_2)			
· sous-population P2	169.3 ±1.9	63.8 ±1.2	267
· sous-population P3	162.5 ±2.5	63.1± 1.8	94
· sous-population P4	135.9±5.1	59.8±3.6	12
$3^{ième}$ génération (G_3)			
· sous-population P3	220.5±2.0	49.6±1.4	185
· sous-population P4	192.2± 2.1	50.9± 2.5	20
$4^{ième}$ génération (G_4)			
· sous-population P4	250.5±1.8	43.7±2.3	38

Table IV.4.1.: Temps de division cellulaire déterminés à partir du traçage des cellules individuelles. Les cellules imagées sont des cellules *E. Coli*: nous cherchons pour chacune d'entre elles, les temps pour lesquels le 1^{er}, $2^{ième}$, $3^{ième}$ ou $4^{ième}$ événement de division cellulaire est détecté ($t_{division}$). On note que certaines cellules ne sont tracées qu'à partir des générations avancées; dans ce cas, le 1^{er} événement de division cellulaire détecté caractérise, non pas la $1^{ère}$ génération, mais une génération plus avancée. Pour s'assurer que, dans nos calculs, le premier temps de division cellulaire décrit bien la $1^{ère}$ génération, seules les lignées dont les ancêtres remontent (au moins) à l'ère 100min sont considérées (cf. figure IV.4.2). Et à partir des valeurs $t_{division}$ collectées sur ces lignées, nous déduisons, pour chaque cellule, la durée τ_{cc} des différentes générations. Nous moyennons ensuite ces durées sur toutes les cellules imagées, et nous obtenons la durée moyenne du cycle cellulaire, $\langle\tau_{cc}\rangle$. Nous indiquons aussi la durée moyenne lorsque celle-ci est moyennée, non pas sur toutes les cellules, mais sur les sous-population de cellules $P3$ et $P4$: ces sous populations ne considèrent que les cellules qui se divisent au moins (resp.) 3 et 4 fois pendant la durée de l'expérience. Les barres d'erreur sont données par la déviation standard de la moyenne, $sem = \sigma/\sqrt{(N_{cells})}$.

Dynamique moyenne de surface cellulaire

Dans cette étude, nous avons comparé la dynamique moyenne des sous population P2, P3 et P4 du tableau IV.4.1, avec celle de l'ensemble des cellules étudiées. Si les

sous populations caractérisent seulement les cellules rapides, la surface cellulaire devrait augmenter plus vite que pour l'ensemble des cellules.
Les dynamiques de surface cellulaire moyenne sont données dans la figure IV.4.3 suivante, avec les temps de division cellulaire calculés dans le tableau IV.4.1 précédent (indiqués par les fléches de même couleur que la sous-population étudiée). Par souci de clarté, nous n'avons représenté que la population entière et les sous-populations P2 et P3 : la dynamique de la sous-population P4 est présentée sur un graphique indépendant (figure IV.4.4), avec la même échelle.

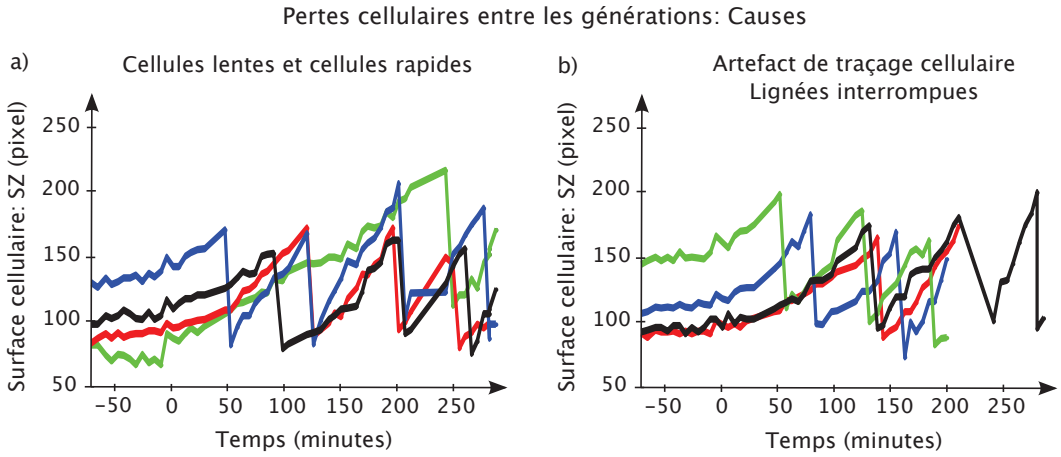


Figure IV.4.2.: Dynamique de la surface cellulaire pour quelques lignées de cellules *GyrA* lors de la transition stress nutritionnel/croissance exponentielle. Cette transition est initiée à $t = 0$ min par l'ajout de glucose au le milieu de culture. a) Quatres lignées suivies sur toute la durée d'acquisition. Ces lignées croissent à des vitesses différentes: la lignée verte est lente avec seulement deux générations, les lignées rouges et noires sont moyennement rapides (elles s'étalent sur 3 générations), et la lignée bleue est très rapide (elle a le temps d'enfanter 4 fois). b) Quatre lignées pour lesquelles le suivi temporel est interrompu: à un moment donné, le traçage cellulaire a perdu l'information de la cellule. Ainsi, les lignées rouges, bleues et vertes, n'ont (resp.) qu'un, deux et trois événements de division cellulaire détectés. Pourtant, elles continuent de se diviser pendant la durée de l'expérience. On note que le $2^{ième}$ événement de division cellulaire de la courbe noire est imprécis: entre $t = 200$ min et $t = 230$ min, il n'y a pas de point d'acquisition.

On constate que l'évolution de la sous-population P2 (courbe bleue de la figure IV.4.3) est similaire à celle de l'ensemble des cellules (courbe verte de la même figure). On s'attendait quelque peu à ce résultat : cette sous-population considère $\approx 61\%$ des cellules initiales[3], et les cellules manquantes sont principalement l'effet d'artefacts de

3. Le nombre initial de cellules (population P de la génération 1) est de 219. Donc, après division cellulaire (génération 2), on devrait détecter 438 cellules. Or, sur ces 438 cellules, seulement 267 (sous population P2) sont suivies jusqu'à la deuxième division cellulaire. Donc seulement 61% de la population totale peut être utilisée pour construire la sous-population P2.

traçage cellulaire. En effet, parmi les cellules qui ne se divisent pas une deuxième fois, seulement 19% sont suivies pendant toute la durée d'acquisition. Les autres sont perdues par le traçage cellulaire, et elles affichent des phénotypes de surface qui laissent supposer qu'elles continuent de se diviser pendant l'expérience (voir la courbe rouge de la figure IV.4.2.b). Dans ces conditions, les cellules initiales qui ne se retrouvent pas dans P2 sont majoritairement des cellules pour lesquelles le traçage cellulaire ne déclare pas les petits enfants. On en conclut que la sous-population P2 est représentative de l'ensemble des cellules étudiées: la durée $\langle \tau_{cc} \rangle$ calculée sur P2 caractérise la population entière de cellules.

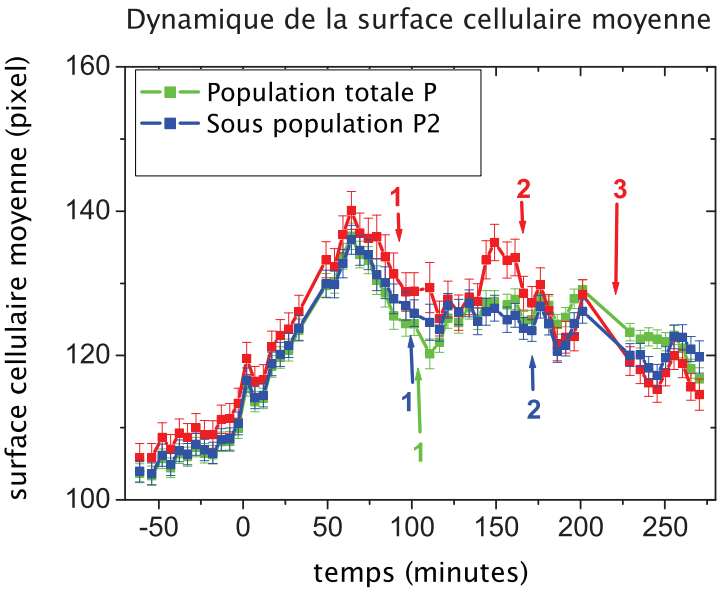


FIGURE IV.4.3.: Evolution temporelle de la surface cellulaire moyenne. En vert, la dynamique de l'ensemble des cellules considérées (population entière de cellules). En bleu, la dynamique de la sous-population P2 de cellules, et en rouge celle de la sous-population P3. Les surfaces cellulaires sont moyennées sur l'ensemble des cellules de la population sondée, et les erreurs sont données par la déviation standarde de la moyenne : $sem = \sigma/\sqrt{(N_{cells})}$. Les temps moyens de division cellulaire sont indiqués par les fléches (même couleur que la sous-population étudiée), avec le numéro de la division. Seules les lignées dont l'ancêtre remonte (au moins) à l'ère $t = 100$ min sont considérées (voir légende du tableau IV.4.1 précédent pour plus d'explications).

Et on considére que la sous-population P3 est aussi représentative de la population totale: elle n'a que 10 min d'avance sur cette dernière, et sa $2^{ième}$ génération s'étale sur la même durée que la population totale (même valeur de $\langle \tau_{cc} \rangle$ dans le tableau IV.4.1). De plus, elle affiche une dynamique de surface cellulaire (courbe rouge de la figure IV.4.3) qui retranscrit relativement bien celle de l'ensemble des cellules (courbe verte de la même figure).

Quant à la sous population P4, elle semble répondre plus rapidement à l'ajout de glucose que l'ensemble des cellules: d'après le tableau IV.4.1 précédent, la première division cellulaire a lieu à $t_{division} \approx 74$ min, alors que pour la population entière, elle se produit vers $t_{division} \approx 104$ min. Cependant, les durées $\langle \tau_{cc} \rangle$ du tableau IV.4.1 sont du même ordre de grandeur que celles de toute la population. De plus, sur le graphe de la prochaine figure IV.4.4, on remarque que la sous population P4 possède une surface cellulaire initiale plus importante que le reste des cellules. Et une fois cette surface normalisée à celle de l'ensemble des cellules, cette sous-population affiche une dynamique (courbe grise de la figure IV.4.4-b) semblable à celle des autres cellules (courbes verte, bleu et rouge de la figure IV.4.4-b). Donc les cellules de P4 possèdent un taux de croissance similaire à celui de la population totale. Dans ces conditions, elles ne sont pas plus rapides que l'ensemble des cellules: elles conservent simplement l'avance qu'elles ont au départ (surface cellulaire initiale plus importante). C'est cette avance qui leur permet de se diviser une $4^{ième}$ fois en fin d'acquisition.

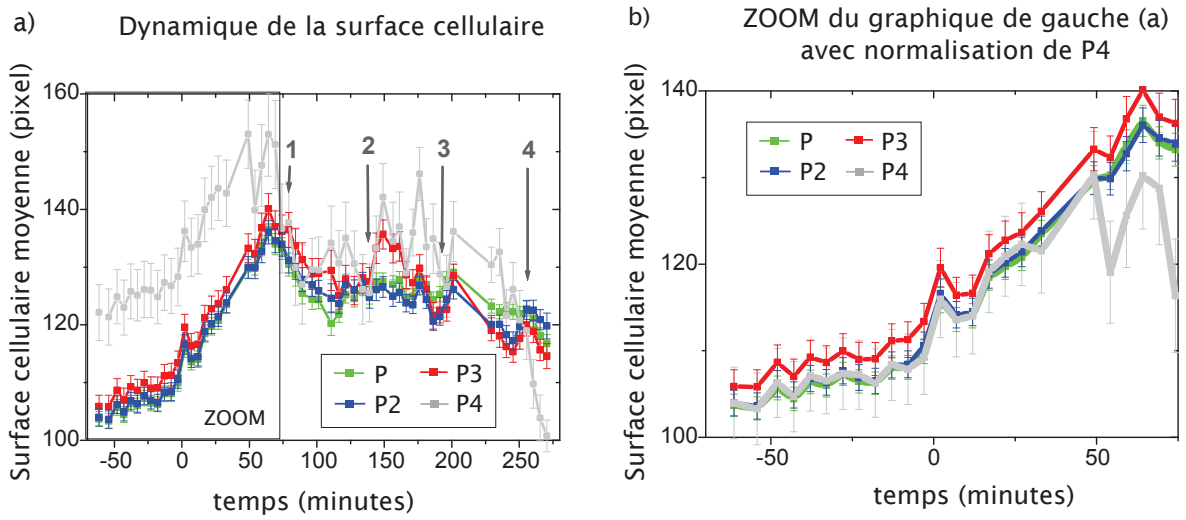


Figure IV.4.4.: Evolution temporelle de la surface cellulaire moyenne de cellules *GyrA*: le temps $t = 0$ min indique l'injection de glucose au milieu de culture défaillant en sources carbonnées. En gris, la dynamique de la sous population de cellules P4. En rouge, bleu et vert les dynamiques de (resp.) la sous-population de cellules P3, P2, et de l'ensemble des cellules considérées. Les surfaces cellulaires sont moyennées sur les cellules de la sous population considérée, et les erreurs sont données par la déviation standard de la moyenne : $sem = \sigma/\sqrt{(N_{cells})}$. Seules les lignées dont l'ancêtre remonte (au moins) à l'ère $t = 100$ min sont considérées (voir légende du tableau IV.4.1 précédent pour plus d'explications). a) Toute la dynamique. Les temps moyens de division cellulaire sont indiqués par les fléches grises pour la sous population P4, avec le numéro de la division associée. b) La dynamique sur les 135 premières minutes: la surface cellulaire de la sous population P4 est normalisée à celle de la la population entière de cellules (normalisation à partir du premier point d'acquisition).

Conclusion

On déduit de tout cela que les temps de division t_{div} calculés dans le tableau IV.4.1 sont représentatifs de l'ensemble des cellules étudiées, exception faite de la sous-population P4. Par contre, les durées de cycle cellulaire $\langle \tau_{cc} \rangle$ calculées sont valides pour l'ensemble de la population. Dans ces conditions, la durée moyenne du cycle cellulaire diminue entre la $1^{ière}$ et la $4^{ième}$ génération. Comme cette durée est inversement proportionelle au taux moyen de dilution [4], ce résultat indique que $\langle \gamma_{dil} \rangle$ augmente entre deux cycles cellulaires. Pour décrire ces variations de manière plus complète, nous avons calculé les taux instantanées de dilution: ceux-ci s'obtiennent à partir des variations de taille relatives à chaque cellule.

Taux de dilution calculé sur les variations individuelles de surface cellulaire

Dans nos expériences, la largeur des cellules varie peu, et les variations de volume cellulaire sont principalement induites par les changements de longueur cellulaire (L). Dans ce cas, le taux de dilution cellulaire (aussi appelé taux de croissance) est donné par le taux d'élongation d'une cellule. On a alors:

$$L_k(t_j) = L_k(t_0) \times 2^{\left(\gamma_{dil,k}(t_j) \times (t_j - t_0)\right)}$$

où γ_{dil} est le taux de dilution de la cellule k à l'instant t_j.
Le taux instantané de dilution de la cellule k s'exprime donc comme:

$$\gamma_{dil,k}(t_j) = \frac{ln\big(L_k(t_j)/L_k(t_0)\big)}{t_j - t_0} \times \frac{1}{ln(2)}$$

On note que la longueur cellulaire (L) est approximée par notre programme de traitement d'images : les cellules sont assimilées à des ellipses de longueur L, et cette description elliptique allonge les cellules (voir les images en fin de section e). Toutefois cette approximation permet de lisser nos courbes : les erreurs de focalisation qui élargissent légérement le résultat de la détection cellulaire, sont grossiérement filtrée par l'ellipse. La grandeur L, bien qu'approximée, permet donc de minimiser les erreurs de focalisation dans le calcul de γ_{dil} .Le taux moyen de dilution cellulaire se déduit des mesures sur cellules individuelles par:

$$\langle \gamma_{dil} \rangle (t_j) = \sum_{\text{cellules k}} \gamma_{dil,k}(t_j)$$

4. On rappelle que $\langle \gamma_{dil} \rangle = \dfrac{ln(2)}{\tau_{1/2}}$. Et on souligne que cette relation considère que le taux de dilution est constant au sein même d'un cycle cellulaire. Elle fournit donc une valeur de $\langle \gamma_{dil} \rangle$ qui, en plus d'être moyennée sur la population de cellules, est moyennée sur la durée du cycle cellulaire.

Fenêtre temporelle: Pour chaque cellule k, l'intervalle de temps $(t_j - t_0)$ détermine le nombre de points expérimentaux utilisés dans le calcul de $\gamma_{dil,k}(t_j)$. Comme notre taux de dilution varie au cours du temps, nos mesures instantanées impliquent de choisir une petite fenêtre, sur laquelle γ_{dil} peut être approximé constant. Nous prendrons la fenêtre donnée par la résolution temporelle (intervalle de temps qui sépare deux acquisitions):

$$\gamma_{dil,i}(t_j) = \frac{ln\big(L_j/L_{j-1}\big)}{t_j - t_{j-1}} \times \frac{1}{ln(2)}$$

Toutefois, dans cette fenêtre temporelle, les variations de surface cellulaire peuvent être noyées par l'imprécision sur l'espace occupé par la cellule [5]. Dans ce cas, le taux de dilution instantanée ne reflète pas la croissance cellulaire, mais l'incertitude dans la mesure de surface cellulaire. Lorsque cette imprécision affecte les cellules de manière indépendante, elle est moyennée dans le calcul de $\langle\gamma_{dil}\rangle$. Mais lorsqu'elle affecte toutes les cellules de manière analogue (notamment lorsqu'une acquisition est légérement défocalisée par rapport à la précédente), elle se retrouve sur la courbe moyenne de γ_{dil}, générant des oscillations qui peuvent cacher la dynamique globale du taux de dilution.
Ceci est particuliérement vrai pour les temps d'acquisition qui précèdent l'injection de glucose: la croissance cellulaire y est lente, et certaines imprécisions se retrouvent sur toutes les cellules (voir, à titre d'exemple, les variations de surface cellulaire autour 0 min sur les figures IV.4.2 et IV.4.3 précédentes). Pour remédier à cela, nous proposons de calculer les taux de dilution instantanés sur une fenêtre temporelle plus large: pour les acquisitions qui précédent l'injection de glucose, nous choisissons une fenêtre qui s'étale sur deux temps d'acquisition. En d'autres termes:

$$\text{Pour } t_j < 0 \text{ min: } \quad \gamma_{dil,i}(t_j) = \frac{ln\big(L_j/L_{j-2}\big)}{t_j - t_{j-2}} \times \frac{1}{ln(2)}$$

Taux de dilution calculé sur un événement de division cellulaire: Pour les cellules filles qui viennent de naître, la longueur cellulaire au temps prénatal (t_0) est extrapolée à partir de la taille de la cellule mère (L_m) et de la taille de la cellule soeur (L_{soeur}) [?]:

$$L_k(t_j) = \begin{cases} \dfrac{L_k(t_j)}{L_k(t_0) + L_{soeur}(t_0)} \times L_m(t_{end}) & \text{si } t_j \text{ précéde la naissance: } t_j \leq t_0 \\ L_k(t_j) & \text{si } t_j\text{suit la naissance: } t_j > t_0 \end{cases}$$

où t_0 et t_{end} sont (respectivement) la date de naissance des deux cellules filles, et la date à laquelle la division de la cellule mère commence. A partir de là, on peut on peut calculer le taux de dilution des cellules filles au moment de leur apparition.

5. Cette imprécision est donnée par la segmentation qui est très sensible aux petites erreurs de focalisation.

Résultats

Taux de dilution individuels: Nous donnons, sur la figure IV.4.5 le taux de dilution de trois cellules (en rouge, vert ou bleu), avec les changements de longueur cellulaire associés.

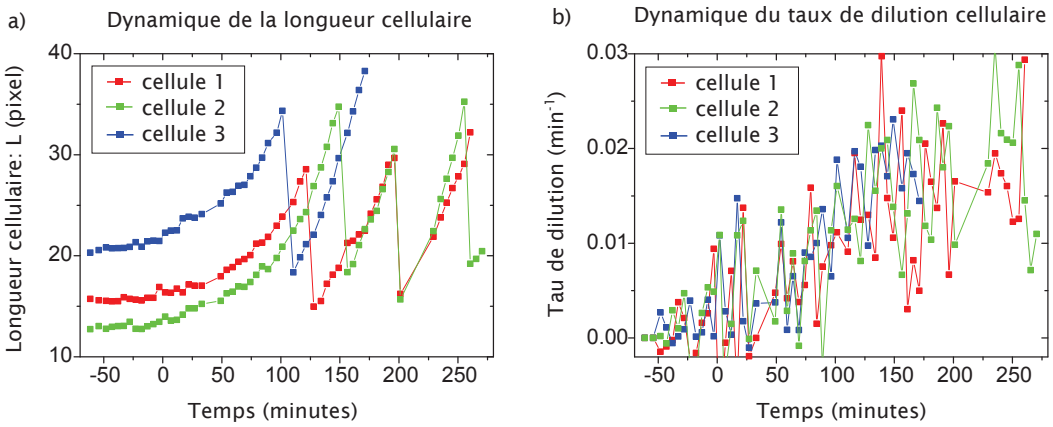


Figure IV.4.5.: a) Evolution temporelle de la longueur de 3 cellules *GyrA* lorsqu'on ajoute du glucose ($t = 0$ min) au milieu de culture défaillant en sources carbonnées. b) Taux de dilution correspondant calculé sur chacune des cellules individuelles.

On constate que les taux de dilution individuels sont fortement bruités par l'imprécision de mesure sur la longueur cellulaire. Et les oscillations générées par ces incertitudes de mesure cachent la dynamique de croissance d'un individu. Toutefois, l'ensemble des mesures individuelles permet de dessiner la dynamique globale du taux de dilution: celui-ci augmente suite l'ajout de glucose ($t = 0$ min), puis semble stagner à partir de ≈ 170 min.

Taux de dilution moyen: La dynamique moyenne de ce taux de dilution est donnée plus loin, sur la figure IV.4.6. Elle est moins bruitée par les incertitudes de mesure, et elle après lissage elle donne une information complète sur $\langle \gamma_{dil} \rangle(t)$. On retrouve le résultat du paragraphe précédent: le taux de croissance augmente suite à l'ajout de glucose, puis reste constant à partir de ≈ 170 min. Les deux plateaux de stress nutritionnel et de croissance exponentielle sont soulignés en noir. Et le plateau de croissance exponentielle retourne bien la durée de cycle cellulaire calculée dans le tableau IV.4.1:

$$\langle \gamma_{dil,0} \rangle = 1,39.10^{-3} \pm 0,06.10^{-3} \text{ min}^{-1} \longleftrightarrow \left\langle \tau_{cc,\text{ stress}} \right\rangle = 500 \pm 31 \text{ min}$$
$$\langle \gamma_{dil,f} \rangle = 0,0145 \pm 0,0040 \text{ min}^{-1} \longleftrightarrow \left\langle \tau_{cc,\text{ croissance exp}} \right\rangle = 47.8 \pm 0.2 \text{ min}$$

Quant à la partie croissante de la courbe, elle a été ajustée numeriquement par un polynôme de degré 4 (courbe noire). On note ici qu'en début d'expérience, le taux de dilution cellulaire $\langle \gamma_{dil,0} \rangle$ est légérement inférieur à la valeur initiale donnée par les

variations pôlynomiales de la figure IV.4.6. Par la suite, lorsque nous manipulerons la variable $\langle \gamma_{dil} \rangle(t)$, nous utiliserons le lissage noir de la figure IV.4.6.

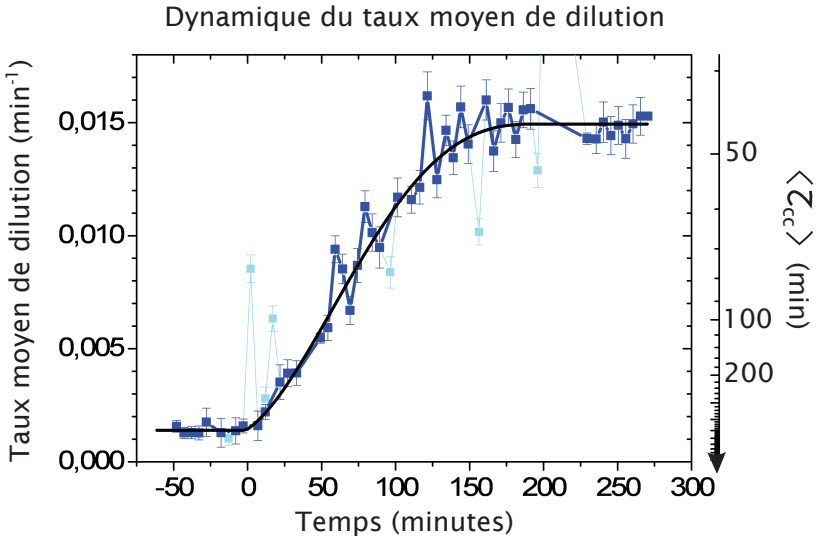


Figure IV.4.6.: Dynamique du taux moyen de dilution des cellules *GyrA* lorsqu'elles passent d'un état de stress nutritionnel à un état de croissance exponentielle : à $t = 0$ min, le glucose est ajouté dans le milieu de culture défaillant en sources carbonnées, et la croissance cellulaire repart. En cyan, le taux de dilution calculé sur tous les temps d'acquisition: seul le temps 201 min qui espacé de plus de 20 min de l'acquisition précédente n'est pas représenté. En bleu, le même taux de dilution, mais les points fortement oscillants qui sont induits par de légères erreurs de focus, ont été effacés. Enfin en noir, le lissage de la courbe bleue par un polynôme de degré 4 sur sa partie croissante ($t \in [0, 190]$ min), et par une constante pour la partie tracée sur les temps inférieure à 0 min, ou la partie qui concerne les acquisitions supérieures à 190 min. Le premier plateau de stress nutritionnel correspond à un taux de dilution de $\gamma_{dil,0} = 1,39.10^{-3} \pm 0,06.10^{-3}$ min^{-1}, et le plateau de croissance exponentielle est caractérisé par $\gamma_{dil,fin} = 0,0145 \pm 0,0042$ min^{-1}. Quant à la partie croissante, elle peut être décrite par une équation polynomiale de degré 4 ($y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4$): l'ajustement numérique (en noir) retranscrit bien la courbe bleue. Les coefficients du polynôme retournés par l'ajustement valent: $a_0 = 1,45.10^{-3}$, $a_1 = 0,45.10^{-4}$, $a_2 = 1,34.10^{-6}$, $a_3 = -1,00.10^{-8}$ et $a_4 = 1,93.10^{-11}$.

IV.4.2. Taux de dégradation

Maintenant qu'on connaît le taux de dilution cellulaire, on peut calculer les taux de dégradation de l'ARNm et des proteines: il suffit d'ajouter à γ_{dil} la composante active de la dégradation. Cette composante a été déterminée par nos collaborateurs biologistes lorsque nos cellules sont en pleine croissance exponentielle. Cela correspond dans nos expériences aux derniers temps d'acquisition. On sait donc que pour ces temps, la GFP

est caractérisée par les constantes de dégradation intrinséques ci-dessous (cf. tableau II.1.1 du chapitre II.1) :

$$\gamma_{mf,\text{ int}} = 0.30 \pm 0.13 \text{ min}^{-1}$$
$$\gamma_{pf,\text{int}} = 0.012 \pm 0.001 \text{ min}^{-1}$$

Nous considérerons que ces taux restent constants pendant la durée de l'expérience. Cela revient à supposer que les variations déterministes de $\gamma_{m,\text{int}}$ et $\gamma_{p,\text{int}}$ sont insignifiantes devant celles de k_m, k_p et γ_{dil}, et qu'elles ont un effet négligeable sur les fluctuations observées.

Taux de dégradation protéique : Sous l'hypothèse précédemment formulée, les variations de γ_p sont gouvernées par la dynamique de γ_{dil}. On a ainsi (d'après la figure IV.4.6) :

$$\gamma_p = \begin{cases} \gamma_{p0} & \text{si t<0 min} \\ a_0 + a_1 t + a_2 t^2 + a_3 t^3 + a_4 t^4 & \text{si } \in[0\ 190] \text{ min} \\ \gamma_{pf} & \text{si t>190 min} \end{cases}$$

où $a_0 = \gamma_{pf,\text{ int}} + 1,45.10^{-3}$, $\gamma_{p0} = \gamma_{pf,\text{ int}} + 1,39.10^{-3}$ et $\gamma_{pf} = \gamma_{pf,\text{ int}} + 0,0145$. Quant à a_1, a_2, a_3 et a_4, ils sont donnés dans la légende de la figure IV.4.6.

Taux de dégradation de l'ARNm : Pour l'ARNm, la dégradation active est beaucoup plus importante que la dégradation par dilution : $\gamma_{m,\text{ int}} \gg \gamma_{dil}$. Dans ces conditions, notre hypothèse de dégradation active constante implique que γ_m ne varie pas :$\gamma_m(t) \approx \gamma_{mf,\text{ int}}$

IV.4.3. Taux de production

On remarque dans les paragraphes précédents que $\gamma_m \gg \gamma_p$. Par conséquent, à l'équilibre statistique, si le nombre de protéines est important, le modèle proposé génère une distribution Gamma de paramètres [7, 10] :

$$a = \frac{k_m}{\gamma_p}, \text{ et } b = \frac{k_p}{\gamma_m} \tag{IV.4.1}$$

On rappelle que ce modèle se place dans les conditions suivantes :

$$\left\{ \begin{array}{c} \gamma_m \gg \gamma_p \\ \text{Equilibre statistique} \\ \text{Bruits de transcription et de traduction dominants} \\ \text{Nombre de protéines/cellule important} \end{array} \right\} \Rightarrow \left\{ \begin{array}{c} \text{Distribution gamma } (a,\ b): \\ \\ a = k_m/\gamma_p \\ b = k_p/\gamma_m \end{array} \right\}$$

Plusieurs de ces conditions sont vérifiées dans nos expériences. Et pour être applicable à nos mesures, le modèle implique de formuler seulement deux hypothèses:

1. Les taux de réaction sont déterministes : ils ne représentent pas une source extrinsèque de bruit.
2. Le bruit observé est principalement l'effet des mécanismes de transcription et de traduction.

Dans nos expériences, rien ne prouve pour l'instant que les deux hypothèses sont vérifiées. Toutefois, nous avons vu dans la section précédente, qu'en début et fin d'expérience on est à l'équilibre statistique, avec une variabilité inter-cellulaire qui peut être décrite par une distribution gamma. Dans ces conditions, le modèle (et ses hypothèses) reste cohérent avec nos résultats expérimentaux.

De plus, indépendemment du modèle proposé, la relation IV.4.1 a été vérifiée dans des expériences comparables aux nôtres [21] (voir section I.1.3). Ces expériences ont été réalisées à 30°C sur des cellules *E.Coli* qui ont atteint le régime stationnaire de la croissance exponentielle [6]. Ces cellules contiennent des rapporteurs chromosomaux fusionnés à la protéine d'intérêt. Et à l'instar de nos expériences qui retournent la densité de fluorescence cellulaire, leurs mesures de fluorescence ont été calibrées pour obtenir la densité de protéines par cellules. Pour la protéine GyrA, l'équipe de X.S.Xie mesure ainsi une statistique, qui comme nous, suit une distribution gamma. Les paramètres de cette statistique sont calculés à partir de la densité moyenne de protéines par cellule $\langle n_p \rangle$, et de la déviation standarde associée $\sigma^2_{n_p}$:

$$a = \frac{\langle n_p \rangle^2}{\sigma^2_{n_p}} \text{ et } b = \frac{\sigma^2_{n_p}}{\langle n_p \rangle^2}$$

Et avec un nombre moyen de protéines/cellule qui vaut 68.83±0.44, ils mesurent:

$$a_{f,Xie} = 4.93 \pm 0.11$$
$$b_{f,Xie} = 13.79 \pm 0.33 \text{ protéines/taille cellulaire}$$

Les rapports k_m/γ_p et k_p/γ_m ont été comparés aux paramètres $a_{f,Xie}$ et $b_{f,Xie}$ de la distribution gamma. Le taux moyen de transcription k_m de la protéine GyrA est déduit de la relation: $\langle m \rangle_{eq} = k_m/\gamma_m$ [7]: le nombre moyen d'ARNm par cellule $\langle m \rangle_{eq}$ est mesuré par séquençage, et le temps de demi-vie de l'ARNm γ_m est déterminé avec la même méthode que nos collaborateurs biologistes (voir le tableau II.1.1). Quant au taux moyen de dégradation protéique γ_p, il est principalement dominé par la dilution cellulaire: $\gamma_p \approx \gamma_{dil}$. Ils

6. Dans leurs expériences, la croissance exponentielle est caractérisée par un temps de division cellulaire de 150 minutes.

7. Cette relation s'obtient à partir de la première équation du système IV.1.1 présentée en début de chapitre): lorsque l'ARNm est à l'équilibre (i.e. $d\langle m \rangle/dt = 0$), cette relation devient $\langle m \rangle_{eq} = k_m/\gamma_m$.

mesurent ainsi (pour une fusion chromosomale de la protéine GyrA avec le rapporteur Venus):

$$\frac{k_m}{\gamma_p} = \frac{\left(\langle m \rangle \times \gamma_m\right)}{\gamma_{dil}} \approx 6.57$$
$$\frac{k_p}{\gamma_m} \approx 10.5$$

Ces deux rapports sont du même ordre de grandeur que les paramètres $a_{f,Xie}$ et $b_{f,Xie}$ précédents. On peut donc considérer qu'à l'équilibre statistique de la phase exponentielle, les paramètres de leur distribution gamma représentent:

$$a_{f,Xie} \approx \frac{k_m}{\gamma_p} \text{ et } b_{f,Xie} = \frac{k_p}{\gamma_m}$$

En extrapolant ce résultat à nos cellules GyrA, on peut donc utiliser la relation IV.4.1 précédente pour déterminer k_m et k_p (dans le cadre du modèle).

Taux de production en fin d'expérience

Ainsi les taux de transcription et de traduction peuvent être déterminés à partir des valeurs de a et b mesurées en fin d'expérience, où k_p sera exprimé en unité de fluorescence (coups/cellule/pixel). Pour notre GFP dans les cellules *GyrA* en fin d'expérience, ils valent :

$$\begin{cases} k_{mf} = a_f \times \gamma_{pf} = 0,22 \pm 0,09 \text{ molécules d'ARNm.min}^{-1} \\ k_{pf} = b_f \times \gamma_{mf} = 17,36 \pm 8,12 \text{ Unité de Fluorescence.min}^{-1} \end{cases}$$

Taux de réaction en début d'expérience

Et si on s'intéresse maintenant aux temps d'acquisition qui précédent l'injection de glucose, le nombre de protéines est faible. Par conséquent, le modèle ne génère pas strictement une distribution gamma : il génère en fait une distribution binomiale négative dont les paramètres a et b valent (respectivement) k_m/γ_p et k_p/γ_m [7]. Cette loi de probabilité tend vers la distribution gamma pour $n \gg 1$. Et pour nos valeurs de n de l'ordre de l'unité, les paramètres a_0 et b_0 trouvées dans la section IV.3.2 fournissent peu de différences entre les deux lois de probabilité. En effet, sur la figure IV.4.7 suivante, les deux distributions sont tracées pour des niveaux de fluorescence cellulaire qui correspondent aux valeurs observées (cf figure IV.3.5). Et sur ces valeurs, la distribution binomiale négative et la distribution gamma présentent des traits de caractère similaires : nos données expérimentales acquises en début d'expérience ne sont pas assez précises pour distinguer les deux distributions.
Ainsi, dans une première approximation, la distribution gamma tirée des données expérimentales peut être approchée par une distribution binomiale négative de mêmes paramètres. Dans ces conditions, les paramètres a et b mesurés sur nos données expérimentales

ont le même sens physique en début d'expérience qu'en fin d'expérience : on a toujours la relation IV.4.1 du paragraphe précédent. Donc en début d'expérience :

$$\begin{cases} k_{m0} = a_0 \times \gamma_{p0} = 5,36.10^{-3} \pm 1,8.10^{-3} \text{ molécules d'ARNm/min} \\ k_{p0} = b_0 \times \gamma_{m0} = 1,05 \pm 0.02 \text{ Unité de Fluorescence/min} \end{cases}$$

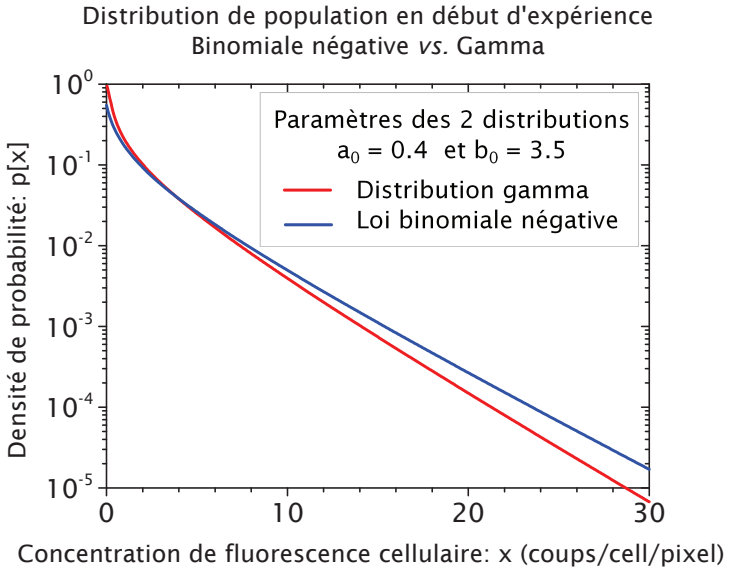


FIGURE IV.4.7.: Comparaison de la distribution gamma avec les distributions binomiales négative pour nos conditions expérimentales de début d'expérience. Les valeurs expérimentales des paramètres a_0 et b_0 des distributions ont été déterminés dans le paragraphe "Distribution en début d'acquistion" de la section IV.3.2 précédente. Et sur les données expérimentales (cf. figure IV.3.5), les concentrations de fluorescence observées étaient bornées entre 0 coups et 30 coups. Nous nous plaçons donc dans ces conditions : sur la présente figure, les deux lois de probabilité ne sont représentées que sur les niveaux [0 30] coups. Par ailleurs, dans les distributions expérimentales de début d'acquisition, les niveaux supérieurs a 11 coups étaient très faiblement peuplés : seulement $0,6\%$ de la population de cellules émet dans ces niveaux de fluorescence (voir figure IV.3.5). On rappelle aussi que les distributions expérimentales tracées en début d'expérience sont binnées sur des bins larges, à cause du nombre limité de cellules imagées. Dans la limite de ces résolutions expérimentales, on peut considérer que les deux distributions présentées ici (binomiale négative et gamma) sont équivalentes.

a. Variations des taux de réaction

Au final, notre modèle intégre des signaux d'entrée qui varient au cours du temps : seul le taux de dégradation de l'ARNm reste constant. Les autres taux de réaction évoluent, passant de k_{m0} à k_{mf}, de k_{p0} à k_{pf} et de γ_{p0} à γ_{pf}. Ces variations sont connues pour $\langle\gamma_p\rangle$, mais elles ne le sont pas pour les taux de production. Nous ferons ici des hypothèses, en

choisissant de décrire les variations de k_p et k_m de façon simple : elles seront modélisées par une fonction rampe. On considère donc que les taux de production augmente progressivement suite à l'ajout de glucose ; et dans la limite de notre résolution expérimentale, un modèle rampe nous paraît suffisant. Ainsi, les variations de k_p et k_m seront caractérisée par la durée des rampes : t_{km} et t_{kp}.
Tout ceci est illustré sur la figure IV.4.8 ci-dessous.

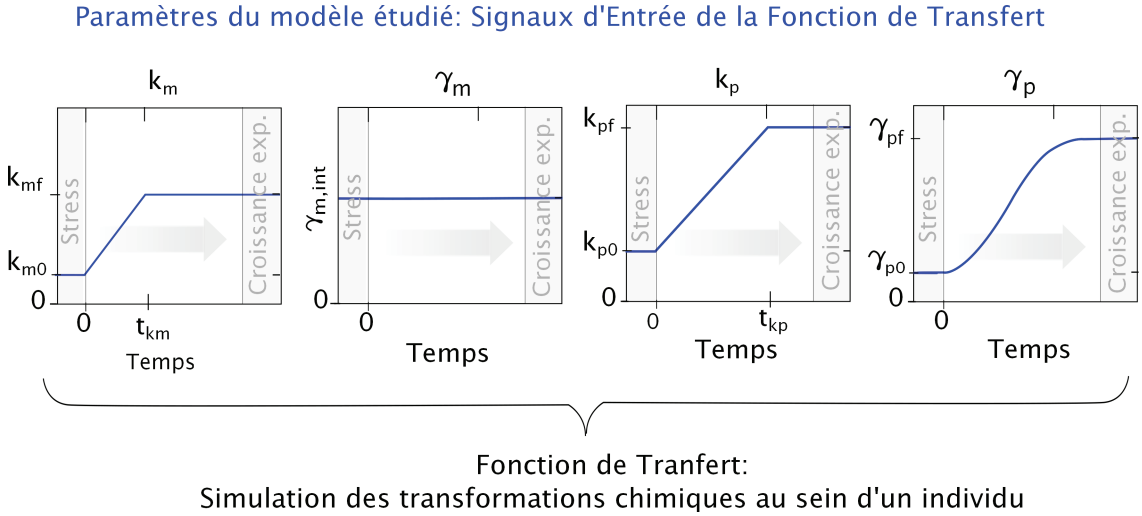


FIGURE IV.4.8.: Dans notre modèle qui décrit la transition entre stress nutritionnel et croissance exponentielle, les taux de réaction (signaux d'entrée de la fonction de transfert dans la figure IV.1.3) varient de manière déterministe dans le temps. Ces variations sont fournies de nos données expérimentales pour γ_p, et elles sont modélisées par une par une fonction rampe pour k_m et k_p. Les valeurs initiales et finales de la rampe données par nos résultats expérimentaux : seules les durées t_{km} et t_{kp} de la rampe sont inconnues. Le taux de dégradation actif des ARNm est, quant à lui, supposé constant pendant la durée de l'expérience, et on a : $\gamma_m(t) = \gamma_{m,\,\mathrm{int}}$.

IV.5. Résultats de la Modélisation

Dans les sections précédentes, nous avons posé les bases de notre modèle. Ce modèle a été construit pour comprendre la stochasticité générée par le module étudié (l'activité du gène *gyrA*) et en tirer des informations sur les paramètres cellulaires. Dans nos expériences, ce module évolue dans un micro-environnement dynamique auquel il est intimement lié : nos cellules transitent d'un état de stress nutritionnel à un état croissance exponentielle, et cette transition passe par l'activation du gène *gyrA*.

Pour contraindre ce modèle à nos conditions expérimentales, nous avons mesuré la valeur des signaux d'entrée dans nos cellules. On s'est appuyées ici sur trois hypothèses :

1. En début d'expérience, les valeurs de a et b mesurées (*i.e.* les paramètres de la distribution gamma expérimentale) gardent le même sens physique qu'en fin d'expérience : elles sont reliés aux taux de réaction par $a_0 = k_{m0}/\gamma_{p0}$ et $b_0 = k_{p0}/\gamma_{m0}$.
2. La dégradation active de la GFP et de l'ARNm associé ne varie pas pendant la durée de l'expérience.
3. Les taux de production $k_m(t)$ et $k_p(t)$ varient de manière linéaire pendant la transition : fonctions rampe de durées respectives t_{km} et t_{kp}.

Avec ces hypothèses, les inconnues du modèle se réduisent à deux paramètres : les durées t_{km} et t_{kp}. Nous disposons donc maintenant d'un modèle qui renferme un nombre restreint de paramètres, et nous pouvons le tester sans trop nous éloigner des conditions expérimentales. Comme il ne possède pas de solutions analytiques connues, son comportement stochastique sera étudié à travers des simulations.

IV.5.1. Simulations

Les simulations sont lancées pour le modèle à deux étages (transcription/traduction) avec les paramètres de la figure IV.4.8 précédente. La comparaison expérience/simulation se fera sur le nombre de protéines par cellule (n) en sortie du module d'expression génique. Et comme nos simulations considèrent que le volume cellulaire est fixe, la variable n du modèle sera traduite en terme de concentration de protéines/cellule. De cette manière, on se

place dans les mêmes conditions lorsqu'on compare le comportement des cellules simulées à celui des cellules réeles : dans les simulations comme dans les expériences, la grandeur concentration est affranchie des variations de volume cellulaire dans le temps. Et pour pouvoir comparer les données simulées à nos données expérimentales, cette concentration n de protéines/cellule sera assimilée à la concentration x de fluorescence cellulaire. On s'épargne ainsi l'imprécision de la conversion : Unité de Fluorescence (coups/cell/pixel) $\Leftrightarrow$ Nombre de protéines. Dans les simulations, le paramètre k_p reste donc exprimé en unité de fluorescence par minute (coups/cell/pixel. min^{-1}). Les deux étapes de la simulation (initialisation et transition) sont résumées sur la figure IV.5.1 et décrites par la suite.

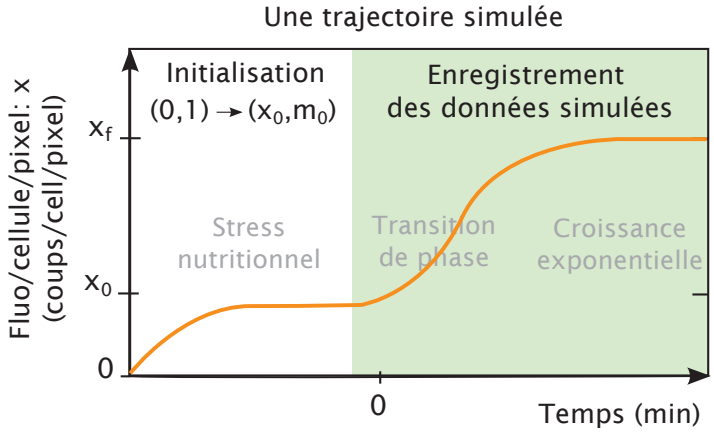


FIGURE IV.5.1.: Simulation d'une trajectoire. Dans une première étape, l'état de la cellule est initialisé : le modèle à deux étages est simulé avec les paramètres k_{m0}, γ_{m0}, k_{p0} et γ_{p0}. Il part ainsi de l'état $(0,1)$ et atteint l'état d'équilibre (x_0, m_0). De cette manière, à l'équilibre de stress nutritionnel, la variable x_0 est générée selon une distribution Gamma, comme observé dans nos expérinces. On commence alors à enregistrer les données simulées, à savoir $x(t)$, la concentration de fluorescence cellulaire. Puis à $t = 0$ min, le glucose est injecté : les paramètres de la simulation (k_m, k_p et γ_p) sont augmentés comme décrit dans la figure IV.4.8.

Initialisation des variables

Dans la section IV.3.2 précédente, nous avons décidé de caractériser la variabilité intercellulaire par une distribution Gamma, de paramètres a_0 et b_0 en début d'expérience. Notre modèle sera donc initialisé à ces conditions expérimentales : pour chaque trajectoire simulée, la variable initiale x_0 est générée selon la distribution expérimentale mesurée en début d'expérience.

Dans ce but, on impose aux cellules le même état $(x, m) = (0, 1)$, où m désigne le nombre d'ARNm/cellule et x la concentration de fluorescence cellulaire. Puis, on laisse chaque cellule évoluer avec les taux de réaction k_{m0}, γ_{mf0}, k_{p0} et γ_{pf0} : on simule pour chacune d'entre elles, le modèle à deux étages de la figure IV.1.4 (section IV.1). Dans ces conditions, lorsque les cellules atteignent l'équilibre statistique, les fluctuations inter-

cellulaires suivent une distribution binomiale négative [7]. Et nous avons vu dans la section IV.3.2 que cette loi de probabilité décrit bien la statistique expérimentale.
Le résultat de la simulation peut être visualisé sur la figure IV.5.2 : la distribution simulée (en bleu ou rouge) fournit relativement bien l'état initial souhaité (distribution binomiale négative en noir). La cellule simulée est alors dans son état initial (x_0, m_0) : cela correspond au premier temps d'acquisition sur nos données expérimentales.

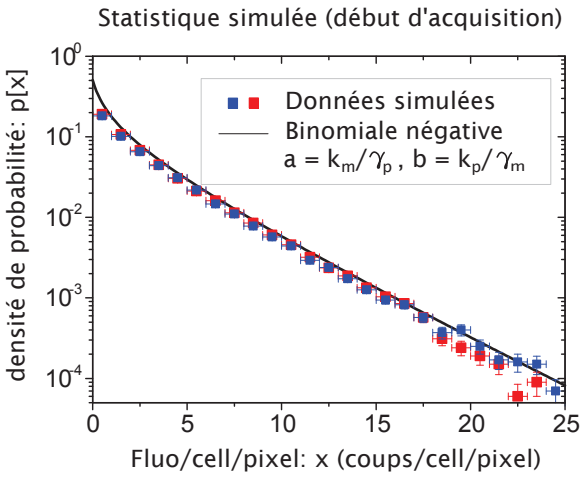


FIGURE IV.5.2.: Simulations : distributions à l'équilibre statistique de la phase de stress. Les distributions sont obtenues à partir de 100 000 trajectoires simulées avec le modèle à deux étages. Les paramètres du modèle utilisés sont : $\gamma_{m0} = 0.30$ min^{-1}, $\gamma_{p0} = 0.012$ min^{-1}, $k_{m0} = 0.0054$ min^{-1}, $k_{p0} = 1.05$ min^{-1}. La statistique rouge est tracée après 300 minutes d'évolution, et la bleu après 500 minutes d'évolution. Ces deux distributions sont confondues : on est à l'équilibre statistique. Et comme le montre la courbe noire, elles décrivent bien la distribution binomiale négative de paramètres : $a_0 = k_{m0}/\gamma_{mf}$ et $b_0 = k_{p0}/\gamma_{pf}$.

Simulation de la transition de phase Stress/Exponentielle

A partir de là, on commence à enregistrer la trajectoire $x(t)$ simulée. Cette trajectoire va mimer le passage du stress nutritionnel à la croissance exponentielle, suite à l'ajout de glucose. Dans nos simulations, le signal excitateur "ajout de glucose" est modélisé par une augmentation des taux de production : ces derniers ne valent plus k_{m0} et k_{p0}, mais ils augmentent maintenant de manière linéaire (cf. figure IV.4.8 précédente). La transition s'obtient alors en simulant le modèle à deux étages lorsque ces taux de production augmentent. On simule ainsi 1 000 trajectoires (*i.e.* cellules), voire 10 000 pour une information plus précise avec les statistiques de population. Puis on en extrait, pour chaque temps d'enregistrement :

- ◇ La concentration moyenne de fluorescence cellulaire : $\langle x(t) \rangle$.
- ◇ La statistique : $P_x(t)$.

IV.5.2. Ajustement du modèle aux données expérimentales

Nos simulations comptent deux inconnues : les durées des rampes associées aux taux de production (t_{km} et t_{kp}). Ces signaux d'entrée inconnus seront ajustés sur nos données expérimentales. Et pour ajuster plus finement la dynamique du modèle aux expériences, on pourra aussi jouer sur les valeurs initiales et finales des taux de réaction. En effet, ceux-ci sont tirés de mesures expérimentales, et ils sont donnés avec des incertitudes non négligeables (particulièrement k_p). On testera donc plusieurs valeurs pour ces paramètres, chacune étant contenue dans les intervalles d'erreur.

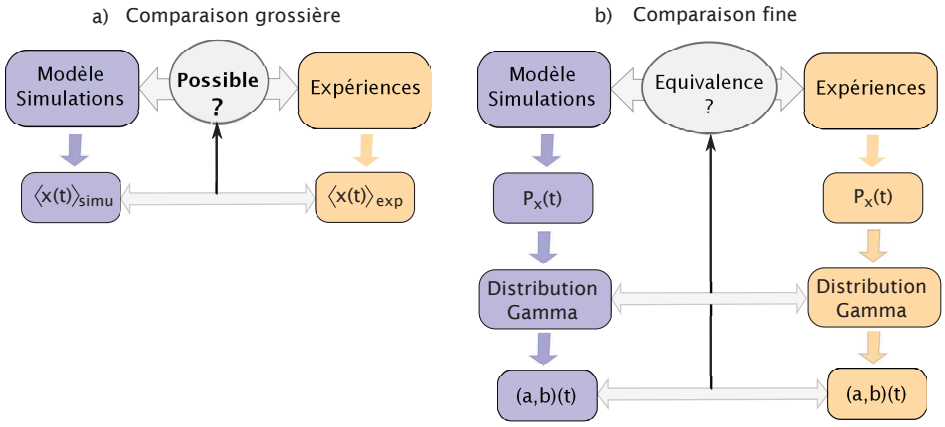


FIGURE IV.5.3.: Ajustement des simulations aux expériences. a) Une première comparaison sur $\langle x \rangle(t)$ permet de sélectionner les configurations/paramètres compatibles avec nos données expérimentales. b) Une comparaison plus fine est effectuée pour les valeurs sélectionnées. Cette comparaison se fait sur $(a, b)(t)$: pour chaque point temporel, la statistique $P_x(t)$ est ajustée numériquement par une distribution Gamma. Si les paramètres a et b de la distribution simulée évoluent comme ceux de la distribution expérimentale, alors le modèle simulé retranscrit bien nos expériences.

Espace des paramètres

En considérant les intervalles d'erreur dans nos mesures, les paramètres du modèle de la figure IV.1.4 sont contenus dans un espace E à 8 dimensions, à savoir : a_0, b_0, a_f, b_f, $\gamma_{m,\,\mathrm{int}}$, $\gamma_{p,\,\mathrm{int}}$, t_{km} et t_{kp}. Les quatres premières dimensions se réfèrent aux paramètres de la distribution gamma observée : ils fixent les valeurs initiales et finales des taux de production (cf. section IV.3.2 précédente). Quant aux dimensions $\gamma_{m,\,\mathrm{int}}$ et $\gamma_{p,\,\mathrm{int}}$, elles donnent la composante active de la dégradation de la GFP. On note ici que la composante induite par la dilution cellulaire [1] n'est pas considérée dans l'espace des paramètres :

1. On rappelle que les taux de dégradation s'expriment comme une combinaison du taux de dilution cellulaire et du taux de dégradation intrinsèque (dégradation active de la molécule).

on suppose que les incertitudes sur nos résultats de la figure IV.4.6 (cf. section IV.3.2 précédente) sont négligeables devant celles mesurées sur $\gamma_{m,\,\mathrm{int}}$ et $\gamma_{p,\,\mathrm{int}}$. Enfin, les deux derniéres dimension (t_{km} et t_{kp}) désignent les signaux d'entrée inconnus de la fonction de transfert : ils fixent la durée des variations sur les taux de production. Et comme ces variations ont lieu pendant la transition d'état qui s'étale sur ≈ 200 min, la valeur de ces signaux sera bornée à l'intervalle [0 200] min. Les bornes des autres dimensions ont été déterminées dans la section IV.3.2 précédente. Et au final, notre espace des paramètres est délimité par :

$$E = \begin{bmatrix} a_0 \\ b_0 \\ a_f \\ b_f \\ \gamma_{mf} \\ \gamma_{pf} \\ t_{km} \\ t_{kp} \end{bmatrix} = \begin{bmatrix} 0.4 \pm 0.1 \\ 3.5 \pm 0.8 \\ 8.02 \pm 0.35 \\ 56.00 \pm 2.7 \\ 0.30 \pm 0.13 \\ 0.027 \pm 0.002 \\ \in [0\ 200] \\ \in [0\ 200] \end{bmatrix}$$

Nous allons explorer cet espace pour trouver les valeurs qui décrivent le mieux nos données expérimentales. Nous réalisons cette opération en deux étapes. Ces deux étapes sont résumées dans la figure IV.5.3 :

1. On cherche d'abord à réduire les bornes des dimensions t_{km} et t_{kp}. Pour ce faire, on fixe les autres dimensions à leurs valeurs médianes, et on explore l'espace $t_{km} \times t_{kp}$. Pour simplifier cette étude, l'espace continu est réduit à un espace discret : on teste les valeurs de t_{km} par pas de 10 min, et on fait de même pour t_{kp}. Cela permet de réduire le nombre de simulations, avec seulement 21×21 valeurs testées. On sélectionne alors les couples de valeurs $(t_{km}; t_{kp})$ qui accordent le modèle aux données expérimentales : les nouvelles bornes de l'espace $t_{km} \times t_{kp}$ seront fixées autour de ces valeurs, avec une fenêtre de largeur 20 min. On note ici qu'une étude préliminaire (exploration aveugle de l'espace E) indique que les valeurs sélectionnées sont du même ordre de grandeur lorsqu'on fixe les autres dimensions à leurs extremums (au lieu d'utiliser les valeurs médianes).

2. Une fois que l'espace des paramètres a été réduit, il est exploré plus finement : les durées t_{km} et t_{kp} sont testées par pas de 5 min, et toutes les dimensions sont sondées. Pour les dimensions bornées par les incertitudes de mesure, on décide de ne tester que la valeur médiane et les extremums. Cela permet, encore une fois, de minimiser le nombre de simulations, avec $3^6 \times 4^2$ valeurs sondées.

Grandeur d'ajustement

Pour chaque simulation, on compare les données simulées avec nos données expérimentales. Cette comparaison se fait sur deux grandeurs :

1. L'évolution moyenne de la concentration de fluorescence cellulaire : $\langle x \rangle(t)$. Cette première comparaison permet d'éliminer les valeurs de l'espace qui s'éloignent des données expérimentales. Celles qui, au contraire, transcrivent bien nos expériences sont sélectionnées pour une étude plus fine sur la statistique.
2. La dynamique des fluctuations statiques : $P_x(t)$. Nous avons vu dans la section IV.3.2 que ces fluctuations peuvent être décrites à tout instant par une distribution Gamma. Pour simplifier, nous décidons donc de caractériser $P_x(t)$ par $(a, b)(t)$: les deux paramètres retournées par l'ajustement numérique gamma de $P_x(t)$. On s'assurera préalablement que les statistiques simulées peuvent, elles aussi, être décrites par une distribution gamma.

Evaluation de l'ajustement

Pour juger de l'adéquation entre les simulations et les expériences, on minimise le χ^2 (non normalisé au degrè de liberté). On calcule donc pour chaque point temporel, l'écart entre la moyenne expérimentale $\langle x(t) \rangle_{exp}$ et la moyenne simulée $\langle x(t) \rangle_{simu}$:

$$\chi^2 = \sum_{ti=1}^{tf} \frac{\Big(\langle x(t_i) \rangle_{simu} - \langle x(t_i) \rangle_{exp}\Big)^2}{\sigma^2_{exp}(t_i)}$$

où t_i désigne un temps d'acquisition de l'expérience, et $\sigma^2_{exp}(t_i)$ la variance associée aux concentrations de fluorescence cellulaire (x_{exp}) mesurées à l'instant t_i.
Lorsqu'on mène une étude plus fine sur la statistique, on compare la dynamique des paramètres $(a, b)(t)$ tirés des distributions expérimentales et simulées.

IV.5.3. Résultats

Réduction de l'espace $t_{km} \times t_{kp}$

Le niveau de ressemblance entre les données simulées et les données expérimentales (évalué par le χ^2 sur les courbes moyennes de densité de fluorescence/cellule) est donné dans la figure IV.5.4 pour toutes les valeurs de $(t_{km}; t_{kp})$ testées. Sur cette figure, on trouve un pli ayant des valeurs reduites, avec sur les bords du pli, deux couples de valeurs $(t_{km}; t_{kp})$ qui minimisent localement le χ^2. Ces deux couples optimums sont :

- $t_{kp} \approx 25$ min et $t_{km} \approx 145$ min.
- $t_{kp} \approx 155$ min et $t_{km} \approx 35$ min.

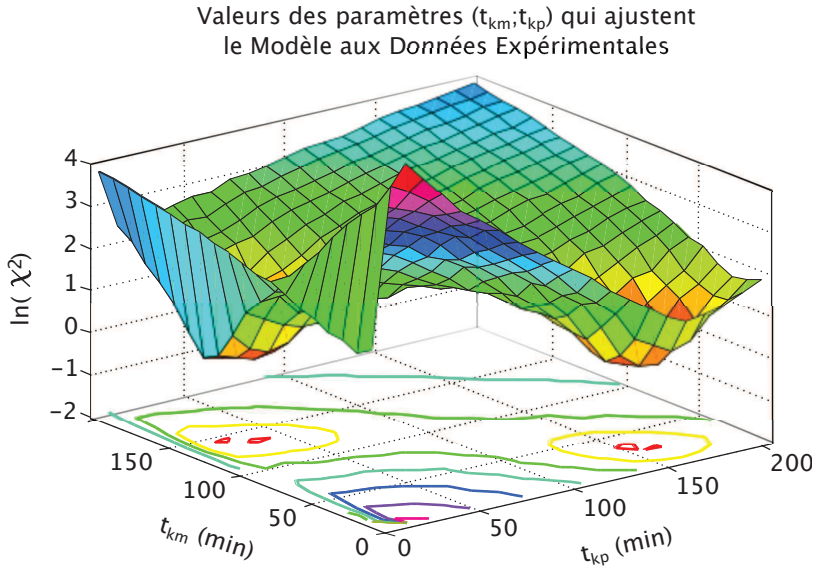


FIGURE IV.5.4.: Adéquation entre le modèle et les expériences en fonction des valeurs des paramètres t_{km} et t_{kp}. Ces durées de rampes décrivent l'augmentation de $k_m(t)$ et $k_p(t)$ suite à l'ajout de glucose au milieu de culture. Et elles sont variées entre 0 min et 200 min par pas de 10 min. Les autres paramètres du modèle sont fixés constants, égaux à leurs valeurs médianes. L'adéquation modèle/expérience est alors évaluée par la ressemblance entre les courbes $\langle x \rangle(t)$ expérimentales et simulées, où la courbe $\langle x \rangle(t)$ traduit l'évolution moyenne de la densité de fluorescence cellulaire au cours de la transition stress nutritionnel/croissance exponentielle. Et la ressemblance sur ces courbes simulée/expérimentale est évaluée par le $\chi 2$. Par mesure de clarté, l'échelle du $\chi 2$ est prise logarithmique.

Lorsqu'on décide de fixer les dimensions bornées par les incertitudes de mesures, non pas à leurs valeurs médianes, mais à leur valeurs extrêmes, on obtient des résultats similaires. On décide alors de réduire les dimensions t_{km} et t_{kp} à des intervalles de 40 min autour de ces valeurs optimales. Au final, notre espace des paramètres se réduit à l'espace :

$$E7 = \begin{bmatrix} a_0 \\ b_0 \\ a_f \\ b_f \\ \gamma_{mf} \\ \gamma_{pf} \\ t_{km} \times t_{kp} \end{bmatrix} = \begin{bmatrix} 0.4 \pm 0.1 \\ 3.5 \pm 0.8 \\ 8.02 \pm 0.35 \\ 56.00 \pm 2.7 \\ 0.30 \pm 0.13 \\ 0.027 \pm 0.002 \\ ([5\ 45] \times [125\ 165])\ \mathrm{U}\ ([135\ 175] \times [15\ 55]) \end{bmatrix}$$

Valeurs des paramètres qui ajustent le modèle aux données expérimentales

Dans cet espace réduit, on trouve deux sets de paramètres qui décrivent relativement bien nos concentrations moyennes de fluorescence/cellule expérimentales. Ceux-ci sont donnés dans le tableau IV.5.1.

set de paramètres	t_{km}	t_{kp}	$\gamma_{m,\text{ int}}$	$\gamma_{p,\text{ int}}$	a_0	b_0	a_f	b_f	$\chi 2$
$S_{p,m}$	135	25	0.31	0.013	0.45	3.5	8.02	56.0	0.94
$S_{m,p}$	25	145	0.18	0.012	0.45	3.5	7.67	58.7	0.85

Table IV.5.1.: Comparaison entre les expériences et le modèle à deux étages. Ce dernier est simulé avec les paramètres donnés dans les colonnes t_{km}, t_{kp} $\gamma_{m,\text{ int}}$, $\gamma_{p,\text{ int}}$, a_0, b_0, a_f, et b_f. La comparaison simulation/expérience se fait sur le moment d'ordre 1 $\langle x(t)\rangle$ de la statistique. Elle est quantifiée par le χ^2. Nous donnons ici les paramètres de la simulation qui minimisent ce χ^2.

Les courbes $\langle x(t)\rangle_{simu}$ générées pour ces deux sets de paramètres sont représentées dans la figure IV.5.5. Et elles décrivent toutes deux très bien nos expériences: la courbe noire qui donne la dynamique moyenne de $\langle x(t)\rangle$ dans les expériences est confondue avec la courbe simulée (en rouge). On note que sur la courbe expérimentale, les valeurs intiales de $\langle x\rangle_{exp}$ sont supérieures à la valeur moyenne donnée par les paramètres (a_0, b_0) de l'ajustement gamma: cet ajustement de la distribution de stress nutritionnel ne considére pas les niveaux de fluorescence élevés qui sont rares, mais qui tirent la moyenne $\langle x\rangle_{exp}$ vers le haut (voir figure IV.3.2 de la section IV.3.6 précédente).

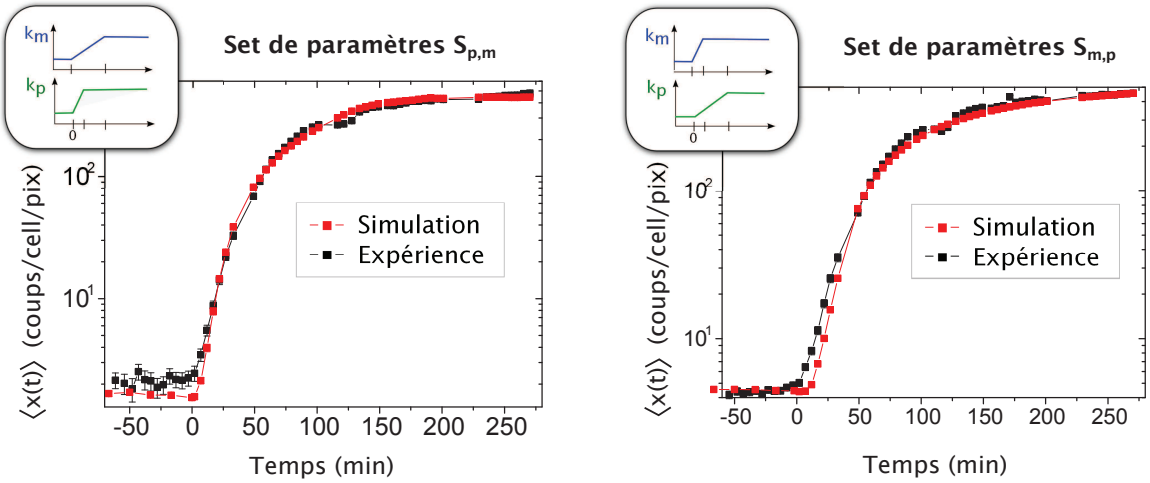


FIGURE IV.5.5.: Comparaison de $\langle x(t)\rangle_{simu}$ avec $\langle x(t)\rangle_{exp}$. Les courbes simulées sont tracées à partir de 10 000 trajectoires. Ces trajectoires sont obtenues pour le modèle à deux étages avec les taux de production qui varient de manière rampe, comme rappelé dans la bulle du coin gauche. Les paramètres utilisés dans la simulation sont donnés dans le tableau IV.5.1 précédent.

Dans ces conditions, il est normal qu'initialement les valeurs simulées (qui s'appuient sur les paramètres a_0 et b_0) soient légérement inférieures aux valeurs expérimentales de $\langle x\rangle(t)$. Exception faite de ces valeurs initiales, les courbes simulées décrivent parfaitement nos expériences, et les deux solutions ne sont pas distinguables l'une de l'autre. Le comportement déterministe de la population de cellules est donc retrouvé par notre modèle. Nous

allons maintenant mener une étude plus fine, sur la distribution de $x(t)$ dans la population de cellules simulées.

Les deux sets de paramètres sélectionnés permettent-ils de décrire toute la statistique ?

Etude statistique

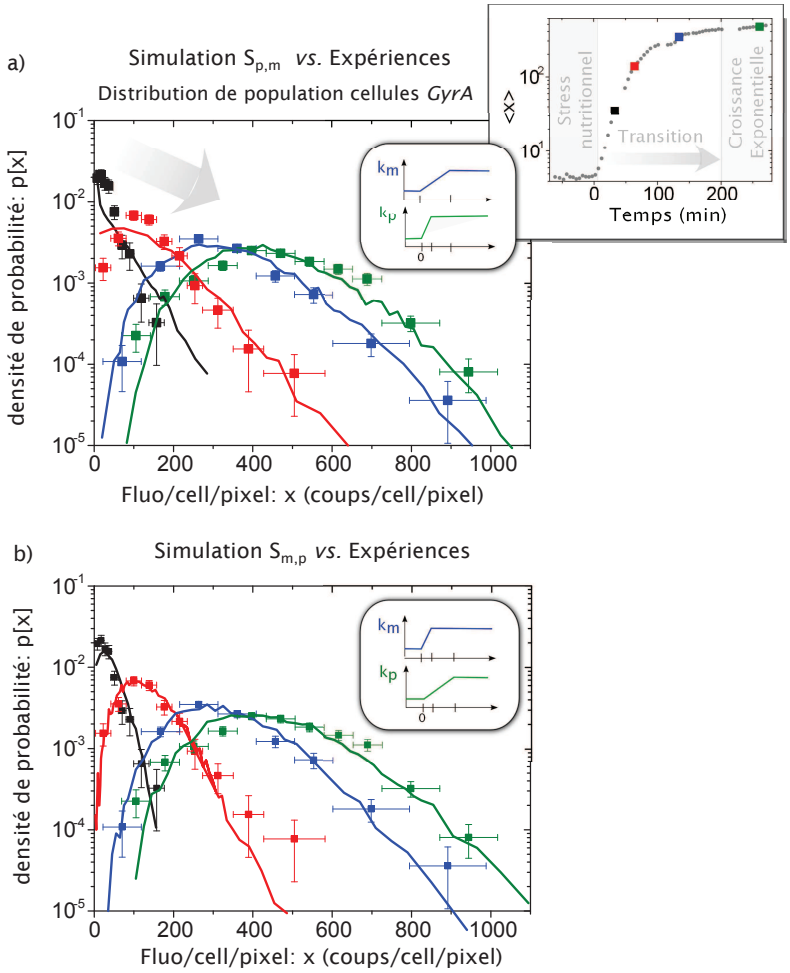


FIGURE IV.5.6.: Comparaison des statistiques simulée (ligne continue) et expérimentales (points de même couleur). Les distributions sont comparées pour 4 temps d'acquisition : ces temps sont indiqués sur la courbe de concentration moyenne de fluorescence cellulaire, en haut à droite (points de même couleur que les distributions). Les courbes simulées (selon les paramètres donnés tableau IV.5.1) sont tracées à partir de 10 000 trajectoires. Par mesure de clarté, les barres d'erreur sur les statistiques simulées ne sont pas représentées, mais on note que ce sont ces erreurs statistiques qui génèrent les petites oscillations sur les courbes simulées.

Nous espérons, avec cette étude plus fine pouvoir distinguer une solution de l'autre, et tirer des conclusions sur les paramètres t_{km} et t_{kp} du modèle. Les statistiques simulées sont représentées en continu sur les figures IV.5.6, et elles sont comparées aux distributions expérimentales (représentées par des points de même couleur). Par mesure de clarté, nous n'avons représenté que quelques d'acqusition : ces temps d'acquisition s'étalent sur toute

la transition d'état, et ils sont donnés dans l'encadré à droite de la sous-figure a. La figure IV.5.6-a donne les résultats pour la simulation lancée avec le set de paramètre $S_{p,m}$, et la figure IV.5.6-b donnent les résultats pour le set de paramètres $S_{m,p}$.

On constate que si d'un point de vue déterministe le modèle décrit bien les expériences, d'un point de vue statistique, la description n'est pas totale. En effet, la transition d'état (distributions noires) est mal décrite par les statistiques simulées. On retrouve néammoins les mêmes traits de caractére : les statistiques se décalent vers les niveaux élevés de fluorescence, avec une bosse qui s'écrase à mesure qu'on avance dans le temps. Qui plus est, toutes les statistiques simulées peuvent, elles aussi, être décrites par une distribution gamma. A titre illustratif, nous donnons sur la figure IV.5.6 le résultat de l'ajustement numérique gamma pour quatres distributions critiques. Celles-ci sont tracées à des temps transitoires, avec en noir la distribution tracée 17 min après l'injection de glucose, et en vert, rouge et bleu, les distributions tracées (resp.) 22 min, 33 min et 49 min après l'injection de glucose.

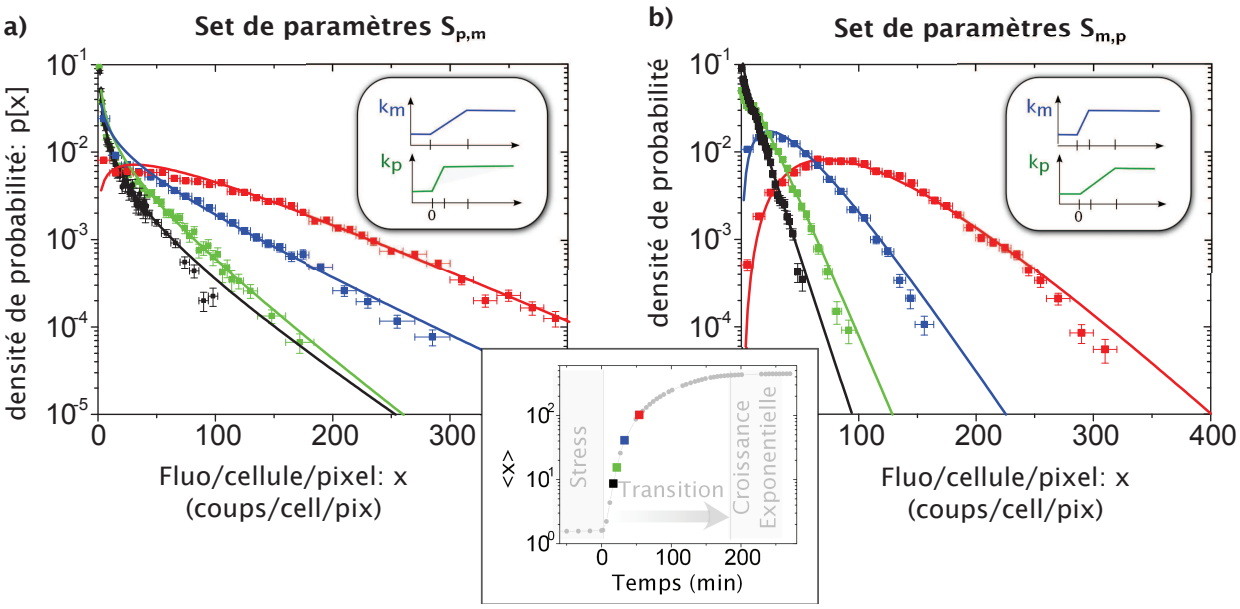


FIGURE IV.5.7.: Statistiques simulées et ajustement numérique par une distribution gamma. a). La statistique est simulée avec le set de paramètres n°1 du tableau IV.5.1. b). La statistique est simulée avec le set de paramètres n°2 du tableau IV.5.1. Les deux statistiques sont tracées à partir de 10 000 trajectoires pour des temps d'acquisition transitoires : en noir, la distribution simulée à $t = 17$ min, en vert, bleu et rouge, celles simulées à (resp.) $t = 22$ min $t = 33$ min, et $t = 49$ min. Le temps $t = 0$ min se réferre ici à l'injection de glucose dans le milieu de culture. Tous les temps d'acquisition considérés ici sont indiqués sur la courbe moyenne de concentration de fluorescence cellulaire (points colorés de même couleur que la distribution dans l'encadré du coin droit).

On constate que l'ajustement gamma est moins performant pour les temps qui suivent l'injection de glucose ($t < 20$ min) : pour ces temps, le nombre de protéines est faible et la

distribution décrit plutôt une loi de probabilité binomiale négative. On peut néammoins (globalement) décrire ces statistiques par une distribution gamma. On pourra donc faire une comparaison simulations/expériences sur les paramètres a et b de l'ajustement gamma. Cette comparaison est plus quantitative et facile à interpréter que la comparaison visuelle de la figure IV.5.6. Les résultats sont présentés sur la figure IV.5.8, avec en noir les données expérimentales, et en rouge et bleu, les données simulées pour les sets de paramètres (resp.) $S_{p,m}$ et $S_{m,p}$.

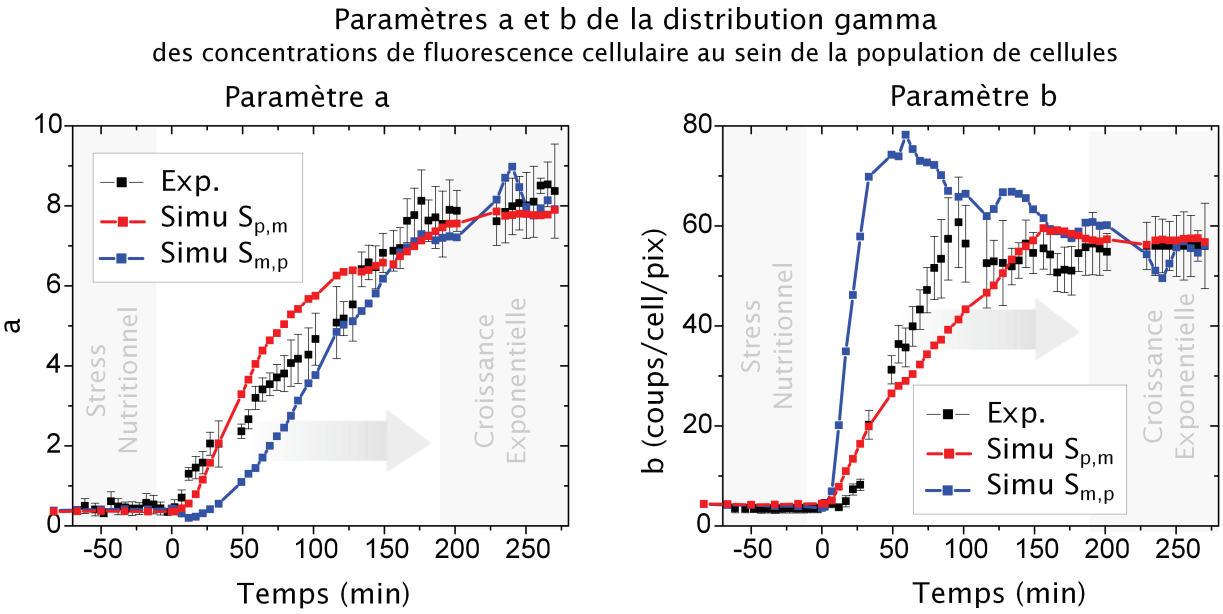


FIGURE IV.5.8.: Comparaison des simulations avec les expériences : paramètres a et b retournés par l'ajustement numérique gamma des distributions de population. En noir, les paramètres tirés des distributions expérimentales. En bleu et rouge, les paramètres tirés des distributions simulées. Ces distributions sont simulées pour un modèle à deux étages (transcription-traduction), avec les paramètres du tableau IV.5.1 : les courbes bleues sont simulées avec le set de paramètres $S_{p,m}$, et les courbes rouges avec le set de paramètres $S_{m,p}$.

On constate que les paramètres des distributions simulées présentent la même dynamique que ceux des distributions expérimentales. Mais encore une fois, la ressemblance n'est pas complète. Et les deux sets de paramètres semblent symétriques en terme d'écart par rapport aux valeurs expérimentales. Pour une simulation (courbes rouges), le paramètre a est en avance par rapport aux valeurs expérimentales et le paramètre b est en retard. Pour l'autre simulation, c'est le contraire (courbes bleues). Cela suggère que les valeurs optimales de t_{km} et t_{kp} trouvées sur la carte de la figure IV.5.4 ne sont les valeurs rééles de nos expériences : ces derniéres sont plutôt un compromis des deux couples de valeurs optimales trouvées. La symétrie observée sur nos valeurs optimales (en terme de χ^2 sur $\langle x \rangle(t)$) suggère que les durées des rampes t_{km} et t_{kp} sont fortement corrélées dans nos expériences.

Couplage de t_{km} et t_{kp}

Si on revient vers l'équation maîtresse et les premiers moments de cette équation IV.1.1, on voit que la valeur moyenne d'ARNm $\langle m \rangle(t)$ et celle de protéines $\langle n \rangle(t)$ sont reliées dynamiquement. Pour nos conditions expérimentales, la quantité d'ARNm/cellule simulée est en régime quasi-stationnaire. En effet, sur la figure IV.5.9 suivante, la courbe rouge qui représente $\langle m \rangle_{simu}(t)$ est confondue avec la courbe verte qui donne $k_m(t)/\gamma_m$. Les deux grandeurs sont donc égales, et dans le système d'équations IV.1.1, cela implique que $d\langle m \rangle/dt = 0$.

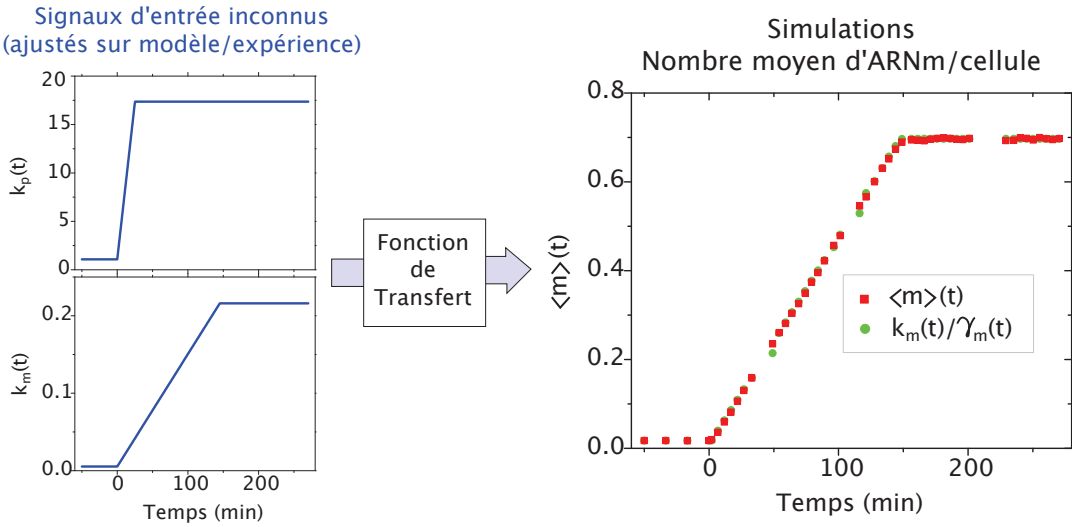


FIGURE IV.5.9.: Verification de l'hypothèse de quasi-stationnarité de l'ARNm. Résultats de la simulation pour des valeurs de $(t_{km}; t_{kp})$ optimales : évolution du nombre moyen d'ARNm (en rouge) dans la population de cellules simulées. Les simulations sont lancées pour 1000 cellules avec les valeurs médianes des paramètres déterminés expérimentalement. Les durées des rampes pour les taux de production k_m et k_p sont fixées à leurs valeurs optimales : $(t_{km}; t_{kp}) = (145; 25)$ minutes. Nous représentons en vert le rapport $k_m(t)/\gamma_m(t)$.

Ce résultat se retrouve pour les deux couples de solutions minimales, et aussi pour leurs valeurs voisines. Il se retrouve également lorsqu'on lance les simulations, non plus avec les valeurs médianes des paramètres expérimentaux, mais avec leur valeurs extrêmes.
Tout le long de la transition d'état, on a : $\langle m \rangle(t) = k_m(t)/\gamma_m$. A partir de là, le système d'équations IV.1.1 se simplifie et le produit des deux taux de production $k_p(t) \times k_m(t)$ est relié aux données expérimentales par :

$$(k_p(t) \times k_m(t))_{exp} = \gamma_m \cdot \left(\frac{d\langle n \rangle(t)}{dt} + \gamma_p \langle n \rangle(t) \right)$$

Toutes les grandeurs impliquées dans ce couplage sont mesurées dans nos expériences, et on a directement le produit des deux taux de production (voir la figure IV.5.10 suivante).

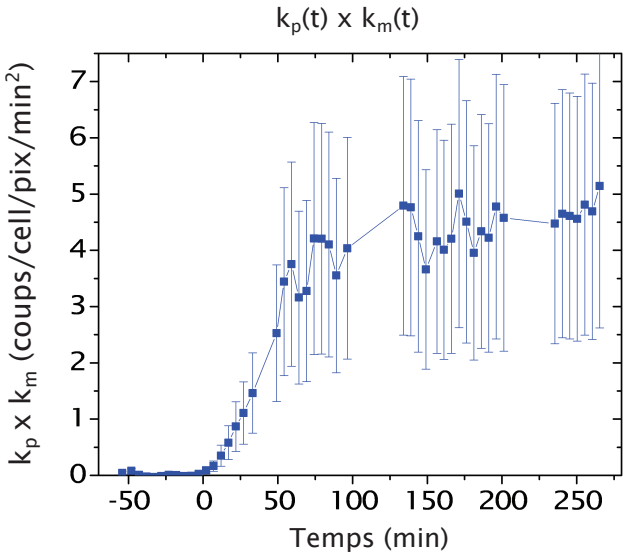


FIGURE IV.5.10.: Force de couplage exercée par nos expériences sur les paramètres t_{km} et t_{kp}.

On constate que le produit $\big(k_m(t) \times k_p(t)\big)$ n'évolue plus au delà de $\approx 100 - 120$ min. Donc les durées réeles des rampes pour k_m et k_p ne sont pas (resp.) $(25; 145)$ ou $(135, 25)$, mais plutôt de l'ordre de $(t_{km}, 120)$ ou $(120, t_{kp})$. Ce résultat reste cohérent avec la carte de la figure IV.5.4 : le χ^2 reste faible dans le pli qui relie les deux minimums, et la durée 120 min (que ce soit pour t_{km} ou t_{kp}) appartient à cette pli.

Conclusion

Pour affiner notre modèle, il faudrait donc lancer de nouvelles simulations en déterminant les paramètres t_{k_m} et t_{k_p} avec la carte de la figure IV.5.4, mais aussi avec la courbe de $(k_p(t) \times k_m(t))_{exp}$ de la figure IV.5.10. On peut toutefois d'ores et déjà tirer des conclusions sur notre modèle :

1. La plupart des raisonnements utilisés dans la littérature pour décrire le bruit intrinsèque se placent dans des conditions idéales (taux de réaction constants). Dans notre système où ces taux varient pendant la transition d'état, ces raisonnements ne sont plus applicables. Par exemple, la relation inverse entre le bruit et la concentration moyenne de fluorescence cellulaire n'est pas retrouvée dans un modèle qui considère seulement les sources intrinsèques (cf figure IV.3.16). Ainsi pour étudier des cellules qui répondent à une perturbation environnementale, on ne pourra pas s'appuyer sur les mêmes raisonnements. Dans notre cas, des études de simulations sont nécessaires pour comprendre le comportement du modèle.
2. D'un point de vue des courbes moyennes $\langle x \rangle$, le modèle décrit parfaitement les données expérimentales.

3. Une étude plus fine sur la statistique de population montre que cette équivalence moyenne est insuffisante pour conclure sur l'adéquation modèle/expérience :

   a) D'un point de vue statistique, le modèle retranscrit globalement la signature du bruit mesuré dans les expériences :

      i. Les distributions de population simulées peuvent être décrites par des distributions gamma, tout comme dans les expériences.

      ii. Les paramètres a et b de ces distributions affichent des dynamiques qui s'approchent globalement des conditions expérimentales.

   b) Cependant, la description n'est pas complète. Les valeurs des taux de réaction choisis pour le modèle (ajustés sur la courbe moyenne $\langle x \rangle(t)$) ne fournissent pas une adéquation totale entre les expériences et les simulations : les paramètres a et b de la distribution simulée diverge des valeurs expérimentales.

On conclut finalement de tout cela que notre modèle simple à deux étages retranscrit globablement le comportement du module d'expression GyrA dans nos cellules. Ce modèle donne une estimée pour les variations des paramètres k_m et k_p dans le module.

Nos **mesures de bruit d'expression génique, combinées aux outils de modélisation développés, permettent donc d'estimer les paramètres du module GyrA** lors de la réentrée en phase exponentielle des cellules *E. Coli*. Et **là où les études déterministes atteignent leurs limites** (les deux sets de paramètres sélectionnés décrivent tout deux parfaitement les données expérimentales d'un point de vue moyen), **les études de bruit, plus complètes, peuvent prendre le relais** (les valeurs réelles des paramètres t_{km} et t_{kp} apparaissent comme un compromis entre les deux sets de paramètres sélectionnés).

IV.6. Discussions

Dans la partie précédente, nous avons négligé un grand nombre de sources de bruit. En effet, notre modèle se réduit aux bruits de transcription et de traduction. Pourtant, la variabilité inter-cellulaire observée peut être générée par de multiples sources (voir figure I.1.8).

Nous nous proposons ici d'analyser briévement les sources de bruit négligées dans notre modèle. Nous étudierons l'importance des bruits extrinsèques, avec une première étude sur le bruit de croissance cellulaire, suivie du bruit d'activation génique et de partitionnement cellulaire. Enfin, nous déterminons l'importance du bruit introduit par notre système plasmidique en le comparant avec des mesures faites sur un autre rapporteur plasmidique (cellules *Fis*) et sur un rapporteur chromosomal (cellules *GyrA-K*).

Bruit d'état cellulaire

Dans nos expériences, les cellules individuelles sont dans des états cellulaires différents: pour un temps d'acquisition donné, la phase du cycle cellulaire varie d'une cellule à une autre, avec des individus sur le point de se diviser, d'autres qui viennent de se diviser, et d'autres encore, qui sont en milieu de cycle cellulaire (voir figure IV.4.2). De plus, les taux de croissance diffèrent d'un individu à l'autre, comme le montrent les différentes durées de cycle cellulaire (voir section IV.4.1 précédente). Or la capacité d'une bactérie à exprimer un gène peut dépendre de l'âge chronologique de la cellule, par exemple de l'état d'avancement de la duplication du chromosome. Toutes ces différences dans l'état cellulaire représentent donc des sources extrinsèques de variabilité [6] qui peuvent dicter les fluctuations intercellulaires observées.

Nous nous proposons ici d'étudier l'effet de ces sources potentielles de bruit sur la variabilité intercellulaire observée dans les cellules *GyrA*. Dans ce but, nous avons développé une routine Matlab qui synchronise artificiellement les cellules sur le troisième événement de division cellulaire: les cellules de la sous-population P3 définie dans le tableau IV.4.1 (voir section IV.4.1) sont décalées dans le temps pour que leur troisième événement de division cellulaire se produise exactement à $t = 225.2$ min. Les trajectoires des cellules individuelles ainsi corrigées peuvent être visualisées dans la figure IV.6.1.

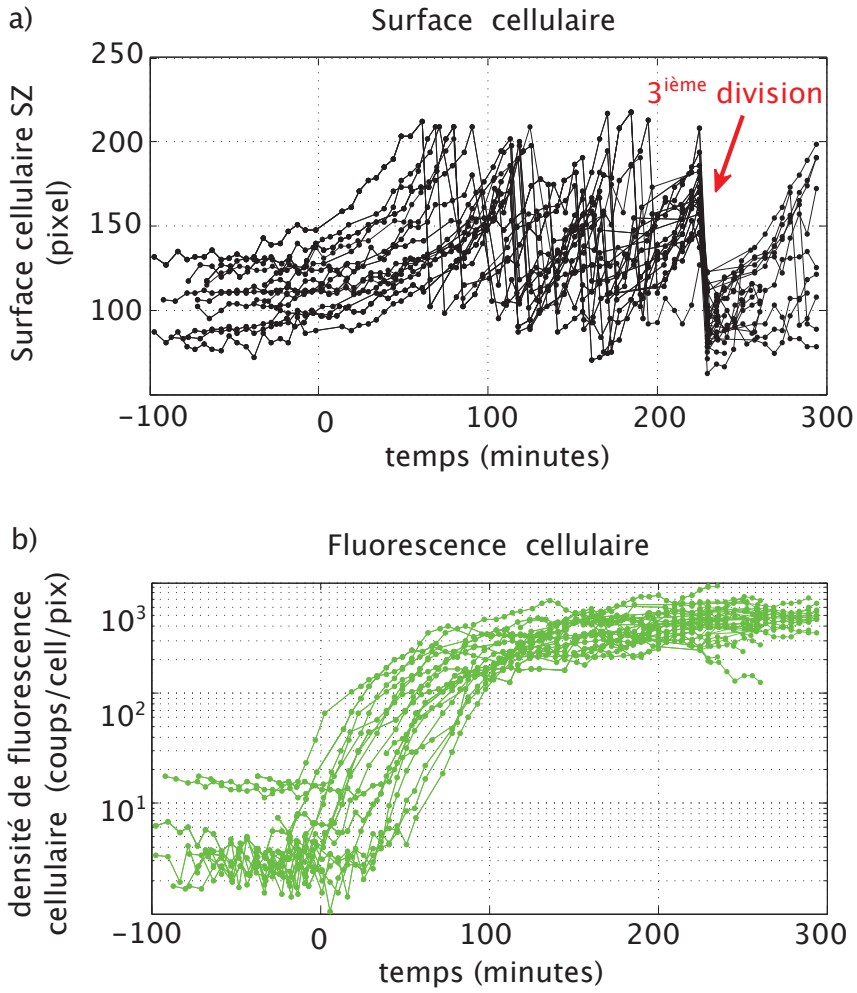


Figure IV.6.1.: Trajectoires de quelques cellules individuelles artifiellement synchronisées sur le troisième événement de division cellulaire. Les trajectoires de chaque individu sont décalées dans le temps pour le $3^{ième}$ division se produise exactement à $t = 225.4$ min. Les individus traités sont ici ceux qui se divisent au moins 3 fois pendant la durée de l'expérience (sous-population P3). a) Dynamique de la surface cellulaire d'individus synchronisés. b) Dynamique de la densité de fluorescence cellulaire correspondante.

Avec cette synchronisation artificielle, nos cellules *GyrA* sont maintenant dans le même état physiologique lors de la $3^{ième}$ division cellulaire. Dans ces conditions, la variabilité intercellulaire mesurée au temps de $3^{ième}$ division est affranchie du bruit de cycle cellulaire. Et comme la $3^{ième}$ division cellulaire apparaît $\approx$ 220 min après l'injection de glucose, les cellules ont déjà atteint l'équilibre statistique de la phase exponentielle (voir figures IV.3.14 et IV.4.6). Dans ces conditions, on peut comparer cette variabilité aux fluctuations statiques mesurées sur les cellules non synchronisées: en fin d'acquisition, on est à l'équilibre statistique pour les cellules non synchronisées, et après la $3^{ième}$ division cellulaire, les cellules synchronisées sont, elles aussi, à l'équilibre statistique. La variabilité générée par les deux types de cellules ne diffère alors que par le bruit d'état cellulaire. Et si ce dernier n'est pas dominant, les fluctuations statiques mesurées doivent être similaires.

Nous avons mesuré ces fluctuations pour les deux types de cellules (synchronisées/non synchronisées). Les résultats sont présentés sur la figure IV.6.2, avec en rouge la distribution de population tracée sur les cellules synchronisées (5 minutes après la $3^{ième}$ division cellulaire), et en bleu la distribution tracée sur les cellules non synchronisées (au temps d'acquisition $t = 260$ min).

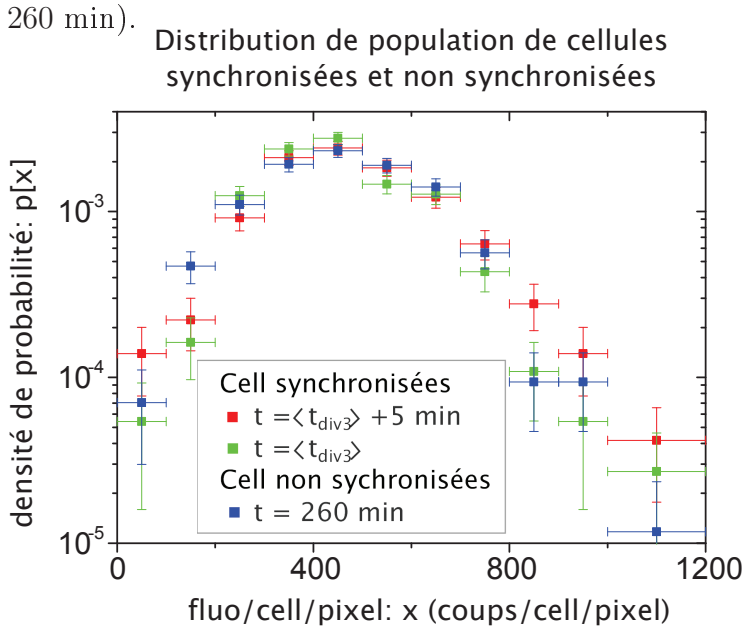


Figure IV.6.2.: Variabilité mesurée sur des cellules artificiellement synchronisées. En bleu, la distribution de population tracée sur 426 cellules *GyrA* non synchronisées. Ces cellules sont sondées 260 min après l'injection de glucose, à l'équilibre statistique de la croissance exponentielle (voir figure IV.3.12). En rouge et vert, les distributions tracées sur les même cellules, mais celles-ci ont été synchronisées sur le $3^{ième}$ événement de division cellulaire. La distribution verte est tracée sur les cellules imagées juste avant leur $3^{ième}$ division, et la rouge est tracée sur les cellules filles imagées 5 minutes plus tard. Le binnage utilisé est le même pour les trois distributions.

On constate que la distribution tracée sur les cellules synchronisées est confondue avec celle tracée sur les cellules non syncrhonisées. D'ailleurs le test de Kolmogorov-Smirnov fournit une valeur de D qui vaut 0.05, alors que la valeur critique calculée pour un niveau de confiance $\alpha = 0.1$ vaut : $D_{critique} = 0,09$. Ainsi, $D < D_{critique}$ et on peut considérer que les deux échantillons sont générés à partir de la même loi de probabilité. On conclut que la statistique n'est pas changée lorsqu'on supprime le bruit de cycle cellulaire : **le bruit observé dans nos cellules *GyrA* non synchronisées n'est pas dominé par le bruit extrinsèque généré par les différents états métaboliques entre les cellules, les différences de progression dans le cycle cellulaire, ou la variabilité dans l'âge de la cellule.**

Bruit de partitionnement sur les protéines

Lors de la division cellulaire, les protéines sont réparties aléatoirement entre les deux cellules filles. Les deux cellules filles diffèrent donc à cause du bruit d'expression génique,

mais aussi à cause de la quantité différente de protéines léguées par la cellules mère. L'étude précédente permet ici de mesurer la variabilité avant et après division cellulaire. En effet, la synchronisation artificielle des cellules fournit une population de cellules mères, et la population de cellules filles qui émerge de la $3^{ième}$ division cellulaire. En mesurant les fluctuations statiques avant et après cet événement de division cellulaire, on peut donc évaluer l'effet du partitionnement des protéines sur la variabilité observée. Et on retrouve deux distributions de population semblables (voir les distributions rouges et vertes de la figure IV.6.2), ce qui suggère que le bruit observé n'est pas dominé par ces effets de partitionnement. On s'attendait quelque peu à ce résultat puisqu'en croissance exponentielle, le nombre de protéines GyrA est important [1] : **la répartition binomiale des protéines [24] a donc peu d'effet sur la variabilité observée.**

Bruit d'activation génique et bruit extrinsèque sur l'étage de transcription

Le modèle proposé dans le chapitre précédent ne considére pas l'étage d'activation génique en amont de la transcription. Cette simplification se justifie par les expériences de Lok-hang So et al [22]. Ces expériences (voir section I.1.4) montrent que lorsque la concentration d'ARNm/cellule est inférieure 0.2, les mécanismes d'activation génique et de transcription se réduisent à un simple processus de Poisson. De plus, pour des concentrations plus grandes, mais qui restent inférieures à 1, ces deux mécanismes dévient faiblement d'un processus poissonnien: $\sigma_m^2/\langle m\rangle < 2$. Dans nos expériences, les simulations suggérent que la concentration d'ARNm reste inférieure à $0,7$ (voir figure IV.5.9). On peut donc, dans une bonne approximation, résumer les étages d'activation génique et de transcription à un processus poissonnien. Dans ces conditions, l'activation génique et la transcription se réduisent à un simple processus de naissance et de mort d'ARNm. Dans ce cas, le taux de transcription k_m décrit la synthése de l'ARNm et l'activation génique: en plus de la production d'ARNm, il renferme aussi l'information sur K_{ON} et K_{OFF}. Le taux de dégradation de l'ARNm, lui, garde le même sens. Avec cette nouvelle définition de k_m, **l'étage d'activation génique peut être négligé dans nos modèle**.

Par ailleurs, la même étude [22] suggére que le bruit extrinsèque est négligeable sur la variable m qui donne la quantité d'ARNm/cellule (cf section I.1.4). Donc dans notre modèle, le **bruit extrinsèque sera négligé sur l'étage de transcription.**

1. Y.Taniguchi et al[21] estiment ce nombre à ≈ 68 protéines/cellule pour une fusion chromosomale de la protéine rapporteuse (Vénus) avec la protéine GyrA. Dans nos expériences sur les rapporteurs plasmidiques, les cellules contiennent ≈ 20 systèmes rapporteurs. Elles devraient donc renfermer beaucoup plus de protéines rapporteuses.

Variabilité donnée par le plasmide

Comme évoqué dans la section I.1.4 le système rapporteur plasmidique introduit de la variabilité dans les cellules. Cette variabilité peut être non négligeable et cacher la signature du bruit d'expression génique. Pour s'assurer que ce n'est pas le cas dans nos expériences, nous avons mesuré le bruit sur deux modules d'expression génique différents: les modules relatifs aux noeuds Fis et GyrA du réseau de régulation. Ces deux modules sont sondés par le même système rapporteur plasmidique. Donc si la variabilité observée est dominée par le bruit du plasmide, les deux modules devraient afficher la même signature du bruit.

La signature du bruit pour le module GyrA est donnée dans le chapitre précédent, et nous présentons celle du module Fis dans la figure IV.6.3. Cette signature a été mesurée dans des cellules qui ont suivi le même traitement que les cellules *GyrA* précédentes: seule l'intensité d'excitation de fluorescence varie, avec pour les expériences présentées ici, un éclairage de fluorescence qui n'est pas atténué. On note que récemment des expériences ont été réalisées avec la même intensité de fluorescence que pour les cellules *GyrA*: elles sont actuellement en cours de traitement.

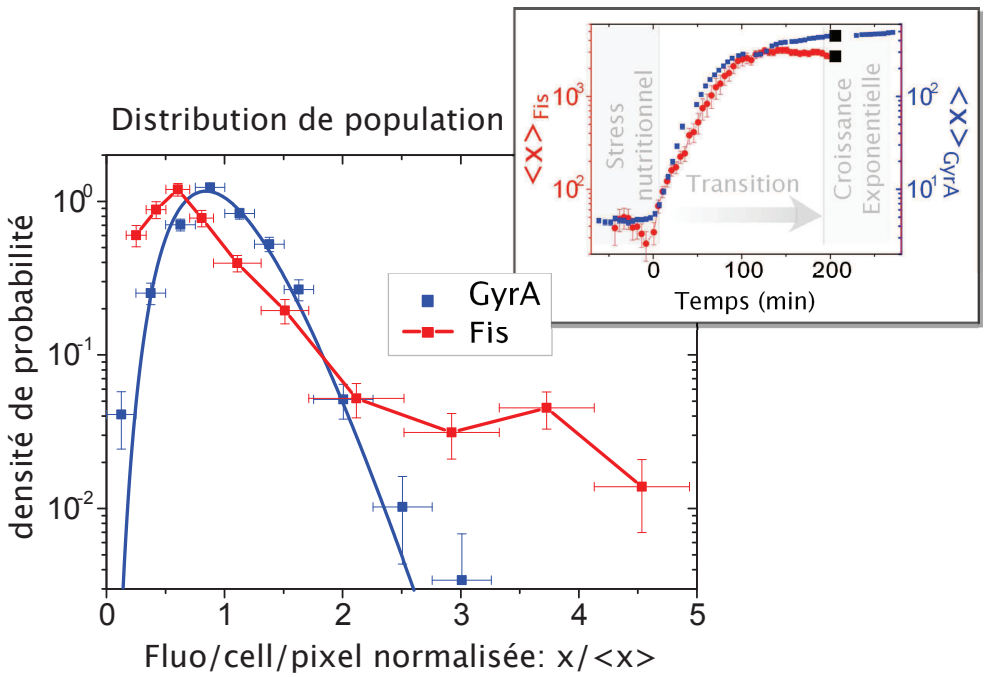


Figure IV.6.3.: Statistiques de population tracées sur des cellules Fis (en rouge) ou GyrA (en bleu) qui contiennent des rapporteurs plasmidiques révélateurs de l'activité du gène *fis* ou *gyrA*. Les données de fluorescence collectées sur les deux types de cellules ont été normalisées à la densité moyenne de fluorescence cellulaire. Et les deux statistiques sont tracées sur des cellules qui ont atteint le régime stationnaire de la phase exponentielle: le temps exact pour lequel les distributions sont tracées est indiqué sur la courbe moyenne de densité de fluorescence cellulaire (point noir dans l'encadré du coin droit).

Dans la figure IV.6.3, les densités de fluorescence cellulaire ont été normalisées à leurs valeurs moyennes: ceci permet de comparer la signature du bruit des deux modules sondés (GyrA et Fis).
On constate que pour les deux modules, la variabilité inter-cellulaire est décrite par des distributions différentes. Donc les modules Fis et GyrA ne génèrent pas les même fluctuations statiques. Ainsi, **le bruit observé ne semble pas être dominé par le bruit du plasmide**. On négligera ce dernier dans notre modèle.

Pour apporter plus de poids à cette conclusion, nous avons mesuré la variabilité inter-cellulaire avec un autre système rapporteur: le système de doubles rapporteurs chromosomaux construit par nos collaborateurs biologistes. Si la variabilité introduite par le plasmide est bien négligeable, ces deux systèmes devraient fournir les mêmes fluctuations statiques, puisqu'ils indiquent tous deux l'activité du gène *gyrA*. On note ici que cela suppose que les deux systèmes rapporteurs sont équivalents. Or nous avons vu dans la section II.1.1 que le système plasmidique pouvait perturber la cellule. Cette étude sur les rapporteurs chromosomaux permet donc aussi de contrôler la qualité du système rapporteur utilisé (plasmidique).

Expériences sur rapporteurs chromosomaux: les cellules *GyrA-K* (rapporteurs chromosomaux) ont subi le même traitement que les cellules *GyrA* (rapporteurs plasmidiques). On note toutefois que pour ces cellules *GyrA-K* l'intensité d'excitation de la lampe fluorescente n'est pas atténuée. Cela permet de compenser le nombre plus faible de rapporteurs: ces cellules contiennent moins de rapporteurs (un seul gène rapporteur par ADN chromosomal) que les cellules *GyrA*. Pour éviter le photo-blanchiment des cellules *GyrA-K* dans ces intensités d'excitation élevées, les images sont acquises moins souvent que dans les cellules *GyrA* (toutes les 10 min au lieu de toutes les 5 min), et nous sondons moins de champs de vue (le nombre de champs imagés est divisé de moitié par rapport aux expériences sur les cellules *GyrA*). Toutefois malgrè toutes ces précautions, le photo blanchiment reste non négligeable. Ne disposant pas encore d'un protocole qui permet de négliger ce photo blanchiment, nous avons mesuré les fluctuations statiques pour un seul instant de la transition. Les mesures ont été effectuées en fin d'expérience, lorsque les cellules ont atteint le régime stationnaire de la croissance exponentielle. Ces mesures ont été réalisées sur une zone de l'échantillon qui n'a pas été éclairée pendant la transition d'état. En conséquence, les cellules imagées n'ont pas encore été endommagées par l'éclairage de fluorescence.
On note que ces rapporteurs chromosomaux n'ont pas été vérifiés en tant qu'indicateurs fluorescents. En effet, la construction chromosomale a été vérifiée par séquençage et elle affiche des caractéristiques de croissance analogue aux cellules sauvages (cellules *wt*), mais

nos expériences préliminaires menées en dynamique montrent que le niveau de fluorescence augmente peu pendant la durée de l'expérience. Nous n'attribuons pas ce phénomène au photo blanchiment, et il reste encore incompris. Sur ce point, des expériences de contrôle menées sur une population de cellules cultivées en plaques multipuits devraient nous éclairer. Ces expériences sont en cours de traitement. On retiendra donc que, indépendamment du photo-blanchiment, ce système est à ce jour insuffisamment caractérisé pour autoriser des études dynamiques. Il permet néammoins de mesurer la variabilité inter-cellulaire pour un état bien défini: le régime stationnaire de la croissance exponentielle.

Résultats. Variabilité inter-cellulaire à l'équilibre statistique de la phase exponentielle: Nous présentons sur la figure IV.6.4 les distributions de population mesurées sur ce système rapporteur. Ces distributions ont été dressées sur les cellules *GyrA-K* qui ont atteint l'équilibre statistique de la phase exponentielle. Nous avons tracé deux distributions: l'une est mesurée sur les rapporteurs CFP (en bleu), et l'autre sur les rapporteurs YFP (en noir). Pour être comparables, ces deux distributions ont été normalisées à la densité de fluorescence cellulaire qui donne la bosse de la statistique (x_{max}). Ces deux distributions normalisées sont aussi comparées à la statistique observée sur les rapporteurs plasmidiques (en vert).

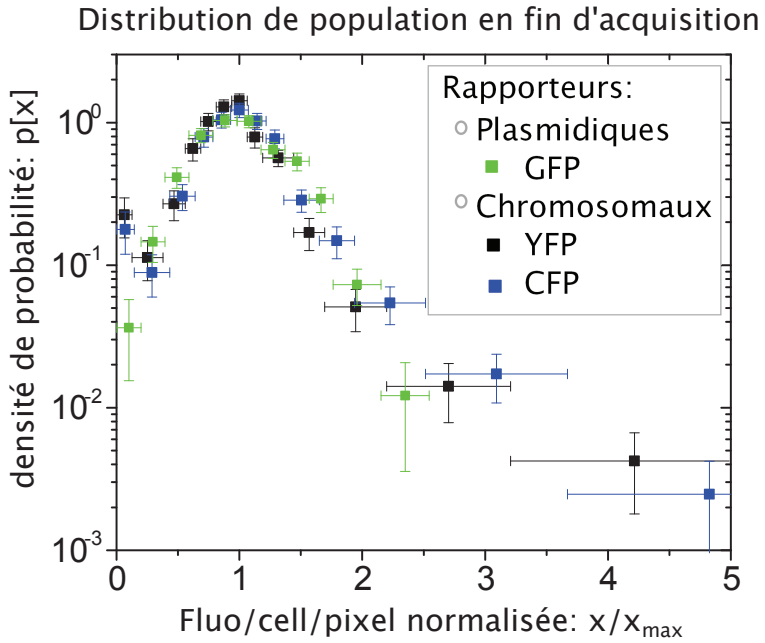


Figure IV.6.4.: Distributions de population tracées à l'équilibre statistique de la phase exponentielle. En bleu et noir, les distributions mesurées par le système rapporteur chromosomal. Celles-ci sont tracées sur 666 cellules *GyrA-K* : en bleu, le rapporteur CFP, et en noir le rapporteur YFP. Enfin en vert, la distribution mesurée par le système rapporteur plasmidique (protéine rapporteuse GFP).

On constate que les trois distributions affichent des signatures similaires : elles peuvent toutes les trois être décrites par une distribution gamma de paramètres similaires. Lorsqu'on s'intéresse aux distributions non normalisées, on trouve un paramètre a de l'ajustement gamma qui est du même ordre de grandeur pour les deux rapporteurs chromosomaux (YFP et CFP) et les rapporteurs plasmiques (voir tableau IV.6.1). Quant au paramètre b, il prend des valeurs différentes selon qu'il est mesuré sur la distribution "YFP", "CFP" ou "GFP" (voir tableau IV.6.1).

	a (sans unité)	b (coups/cell/pixel)
Rapporteurs plasmidiques	8.02 ± 0.35	56.00 ± 2.7
Rapporteurs chromosomaux :		
· CFP	10.2 ± 1.0	10.0 ± 1.0
· YFP	10.0 ± 0.8	3.0 ± 0.2

TABLE IV.6.1.: Paramètres a et b de l'ajustement gamma des distributions mesurées à l'équilibre statistique de la phase de croissance exponentielle. Ces paramètres sont mesurées sur les distributions brutes (non normalisées) tracées à partir des rapporteurs plasmidiques ou chromosomaux.

Les différences sur le paramètre b s'expliquent en partie par les différences d'intensité d'excitation : d'une part la source lumineuse n'émet pas avec la même intensité dans les longueurs d'ondes d'excitation de la GFP, YFP, et de la CFP, d'autre part, cette intensité d'excitation n'est pas atténuée avec le même facteur pour la GFP et les fluorophores YFP et CFP. Tout ceci est résumé dans le tableau IV.6.2. On note par ailleurs que l'efficacité quantique diffère d'un fluorophore à l'autre, ce qui introduit aussi des différences dans le paramètre b.

Intensité d'excitation fluorescence	GFP	YFP	CFP
Intensité de la source lumineuse dans les longueurs d'onde d'excitation du rapporteur	$\approx 410\mu W$	$\approx 200\ \mu W$	$\approx 1000\ \mu W$
Facteur d'atténuation du filtre neutre en aval de la lampe fluorescente	$0,08$	1	1

TABLE IV.6.2.: Différences dans l'intensité d'excitation de fluorescence utilisée pour imager les rapporteurs GFP, YFP et CFP.

On retiendra de cette étude que la signature du bruit n'est pas changée lorsqu'elle est mesurée sur les rapporteurs plasmidiques et chromosomaux. On en conclut que les plasmides ne perturbent pas les cellules. Ce résultat suggére que le système plasmidique introduit peu de variabilité dans les distributions mesurées, ce qui conforte nos conclusions précédantes sur le bruit du plasmide.

Bruit extrinsèque mesuré par le système rapporteur chromosomal

Ce système de doubles rapporteurs permet aussi de mesurer la composante du bruit extrinsèque définie par M.B.Elowitz [5]. On trouve ici :

$$\eta_{int}^2 = 22\% \text{ et } \eta_{ext}^2 = 26\%$$

Cela suggère que, contrairement à nos conclusions précédentes, la variabilité observée est dominée par les sources extrinsèques de bruit. Toutefois, comme souligné dans la section I.1.4, ce système sur-estime le bruit extrinsèque réél. Dans ces conditions, on ne peut pas conclure pas sur l'importance des sources extrinsèques par rapport aux sources intrinsèques, et ce d'autant plus que les deux protéines rapporteuses possèdent des temps de maturation différents (voir tableau II.1.2).

Processus de maturation protéique

Parmi les sources extrinsèques de bruit, on peut aussi trouver le processus de maturation protéique. En effet, pour pouvoir émettre un signal de fluorescence, les protéines n doivent adopter une structure spatiale qui les rend fonctionnelles. C'est ce nombre de protéines fonctionnelles que nous mesurons dans nos expériences. Or, le modèle à deux étages étudié dans la section précédente ne tient pas compte de ce processus de maturation : il fournit le nombre n de protéines par cellule, sans distinguer celles qui sont fonctionnelles de celles qui sont inactives. Pour valider les conclusions établies dans les sections précédentes, une étude supplémentaire est donc nécessaire, avec un modèle plus complet qui intégre le processus de maturation protéique. Ce modèle se construit en rajoutant un étage au modèle de la figure IV.1.4 : les protéines n peuvent maintenant être dégradées, ou se replier sur elles-même pour devenir fonctionnelles (taux de réaction K_r) [61]. Dans ce cas, nous les nommerons *chromophore*. Si ces chromophores ne sont pas à leur tour dégradés (taux de réaction γ_r), ils rempliront leur fonction qui est d'émettre un signal de fluorescence. Ce nouveaux modèle est présenté sur la figure IV.6.5.

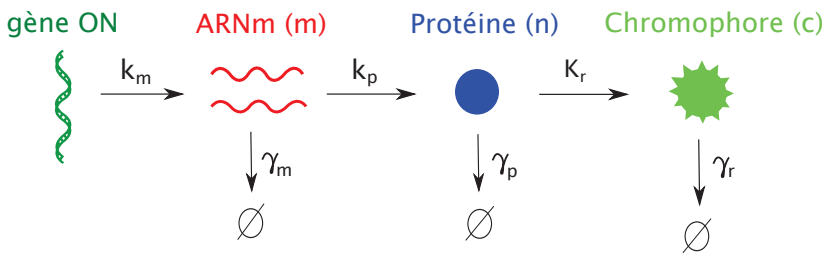


FIGURE IV.6.5.: Modèle à trois étages : transcription, traduction et maturation protéique.

Pour savoir si ce modèle plus complet change les résultats obtenus dans la section précédente, il faut le simuler. Les simulations comptent maintenant 3 variables et 7 paramètres :

- ◇ Les variables sont m, n et c, avec c qui désigne le nombre de chromophores dans la cellule.
- ◇ Les paramètres sont k_p, γ_m, γ_p, γ_r, K_r et t_{km}, t_{kp}. Ce sont les taux de réaction associés aux étages de transcription, traduction, et maturation (cf figure IV.6.5), avec les temps t_{km} et t_{kp} qui donnent la durée des variations sur les taux de production k_m et k_p.

Paramètres : Comme vu dans le modèle à deux étages, les taux de réaction k_m, k_p, γ_m, γ_p, γ_r, et K_r sont susceptibles de varier pendant la durée de l'expérience. On suppose que les taux de dégradation active ($\gamma_{m,\text{ int}}$ et $\gamma_{p,\text{ int}}$) restent constants pendant la durée de l'expérience : seuls le taux de dilution cellulaire (γ_{dil} : voir figure IV.4.6) et les taux de production varient (cf. section IV.5.2 précédente).

Taux de dégradation : $\gamma_{m,\,int}$, $\gamma_{p,\,int}$ et γ_{dil} ont été déterminés dans la section IV.5.2. Pour la dégradation γ_r du chromophore, on considère que le taux de réaction est identique à celui de la protéine non repliée : $\gamma_r = \gamma_p$.

Taux de production : La constante de repliement protéique de la GFP est connue. Elle vaut : $K_r = 0.132\ \text{min}^{-1}$ (cf tableau II.1.1). On propose de fixer ce taux constant pendant la durée de l'expérience. Dans ces conditions, les taux de transcription et de traduction sont les seuls paramètres qui varient pendant la simulation. On choisit de décrire ces variations par une fonction rampe, comme dans la section IV.5.2. Pour déterminer les valeurs initiales et finales de k_m et k_p (début et fin d'acquisition), on utilise un raisonnement analogue au modèle à deux étages. On considére que dans nos expériences :

1. Les sources intrinsèques sont les sources dominantes de bruit.
2. Ce bruit est principalement l'effet des mécanismes de transcription, de traduction, et de maturation protéique.
3. A l'équilibre statistique, les distributions expérimentales sont des distributions gamma dont le paramètre a vaut : $a = k_m/\gamma_p$. On suppose donc que le paramètre a garde le même sens physique que dans le modèle à deux étages.

Les hypothèses 1 et 2 explicitent le modèle à trois étages. Si elles sont valides, l'évolution moyenne des variables m, n et c satisfait :

$$\begin{aligned}
\frac{d\langle m(t)\rangle}{dt} &= \Big(k_m - \gamma_m\langle m\rangle\Big)(t) \\
\frac{d\langle n(t)\rangle}{dt} &= \Big(k_p\langle m\rangle - (\gamma_p + K_r)\langle n\rangle\Big)(t) \\
\frac{d\langle c(t)\rangle}{dt} &= \Big(K_r\langle n\rangle - \gamma_p\langle c\rangle\Big)(t)
\end{aligned}$$

Dans ces conditions, à l'équilibre statistique le nombre moyen de chromophores par cellule est donné par :

$$\langle c \rangle_\infty = \frac{k_p k_m}{\gamma_m \gamma_p} \times \left(\frac{K_r}{\gamma_p + K_r} \right)$$

Par ailleurs, nous avons vu qu'en début et fin d'expérience on est à l'équilibre, avec des distributions expérimentales qui peuvent être décrites par une distribution gamma de paramètres a et b. On a donc aussi [2] :

$$\langle c \rangle_\infty = a \times b$$

L'hypothèse 3 et les deux relations précédentes impliquent alors qu'à l'équilibre statistique :

$$a = \frac{k_m}{\gamma_p},\ b = \frac{k_p}{\gamma_m} \frac{K_r}{(\gamma_p + K_r)}$$

Ces équations sur a et b ne comportent que deux inconnues : k_m et k_p. Tous les autres paramètres sont tirés de nos données (γ_{dil}, a, b) ou des travaux de confrères expérimentateurs (K_r, $\gamma_{m,\,\mathrm{int}}$, $\gamma_{p,\,\mathrm{int}}$). On peut donc calculer les taux de production en début et fin d'expérience. Le modèle se réduit alors à deux inconnues qui sont les mêmes que dans le modèle à deux étages : ces inconnues sont les durées t_{km} et t_{kp} des rampes qui varient les taux de production k_m et k_p.

Simulations : Les simulations s'implémentent de manière similaire au modèle à deux étages (cf. sections III.4 et IV.5). Mais l'espace des paramètres contient maintenant une dimension supplémentaire : K_r . Les paramètres inconnus (t_{km} et t_{kp}) sont, encore une fois, ajustés sur les données expérimentales.

Résultats : Nous présentons sur la figure IV.6.6 la carte qui donne le niveau de ressemblance entre les données simulées et les données expérimentales, en fonction des valeurs de (t_{km}, t_{kp}) testées. Ce niveau ressemblance est évalué par le χ^2 : ce dernier est calculé sur la densité moyenne de fluorescence cellulaire $\langle x(t) \rangle$.

On constate que le relief de la carte n'est pas changé par rapport au modèle à deux étages (cf carte de la figure IV.5.4). On retrouve ainsi deux couples de valeurs $(t_{km}; t_{kp})$ qui minimisent le $\chi 2$. Ces valeurs optimales intégrent maintenant le processus de maturation protéique, et elles sont légérement inférieures aux valeurs retournées par le modèle à deux étages (différence d'environ 20 minutes sur un des deux paramètres). Elles restent néammoins du même ordre de grandeur (voir les tableaux IV.5.1 et IV.6.3).

2. La moyenne d'une quantité c répartie selon une distribution Gamma de paramètres a et b vaut : $\langle c \rangle = a \times b$.

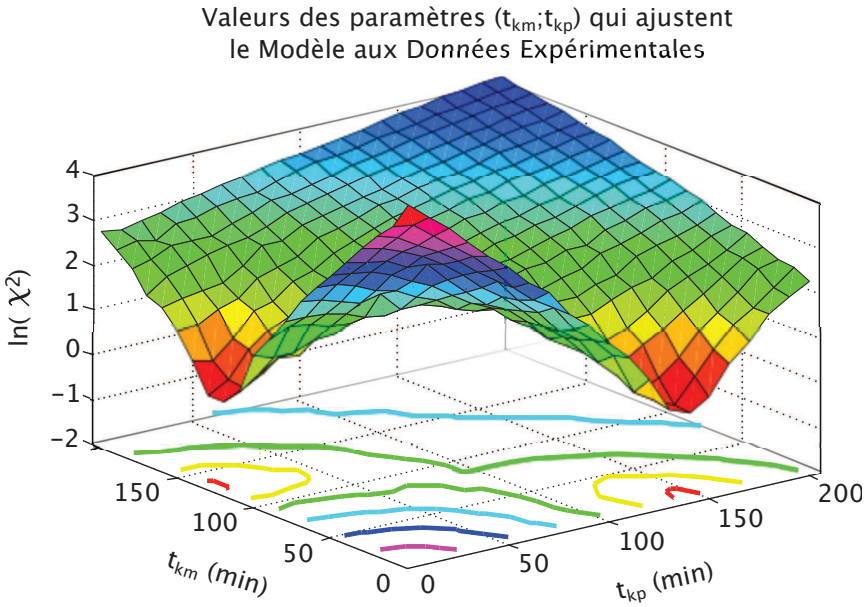


FIGURE IV.6.6.: Adéquation entre le modèle et les expériences pour différentes valeurs des paramètres t_{km} et t_{kp}. Ces temps sont les durées des rampes qui décrivent l'augmentation de $k_m(t)$ et $k_p(t)$ suite à l'ajout de glucose au milieu de culture. Ils sont variées entre 0 min et 200 min par pas de 10 min. Les autres paramètres du modèle à trois étages sont fixés constants, égaux à leurs valeurs médianes. L'adéquation modèle/expérience est alors évaluée par la ressemblance entre les courbes $\langle x \rangle(t)$ expérimentales et celles simulées. La grandeur$\langle x \rangle(t)$ traduit l'évolution de la densité moyenne de fluorescence cellulaire au cours de la transition stress nutritionnel/croissance exponentielle, et la ressemblance sur ces courbes simulée/expérimentale est évaluée par le $\chi 2$. Par mesure de clarté, l'échelle du $\chi 2$ est prise logarithmique.

L'espace $t_{km} \times t_{kp}$ est réduit autour de ces valeurs optimales, et les paramètres du modèle mesurés expérimentalement sont sondés. Pour ces paramètres connus, on teste les valeurs extrêmes et médianes données par les incertitudes de mesure (cf section IV.5). On trouve alors deux sets de paramètres qui décrivent de manière optimale les concentrations moyennes de fluorescence/cellule mesurées dans nos expériences. Ceux-ci sont donnés dans le tableau IV.6.3, et les courbes $\langle x(t) \rangle_{simu}$ générées par ces deux sets de paramètres sont représentées dans la figure IV.6.7. Encore une fois, les courbes moyenne ne permettent pas de distinguer les deux sets de paramètres optimums.

set de paramètres	t_{km}	t_{kp}	$\gamma_{m,\text{ int}}$	$\gamma_{p,\text{ int}}$	a_0	b_0	a_f	b_f	$\chi 2$
$S_{m,p}$	130	5	0.44	0.012	0.45	3.5	8.37	53.3	0.79
$S_{p,m}$	10	140	0.44	0.011	0.45	3.5	7.67	58.7	0.82

Table IV.6.3.: Comparaison entre les simulations du modèle à trois étages et les expériences. La comparaison est faite sur le moment d'ordre 1 $\langle x(t) \rangle$ de la statistique et elle est quantifiée par le χ^2. Les paramètres utilisés dans la simulation sont donnés par les colonnes t_{km}, t_{kp} $\gamma_{m,\text{ int}}$, $\gamma_{p,\text{ int}}$, a_0, b_0, a_f, et b_f : nous donnons les valeurs qui minimisent le χ^2.

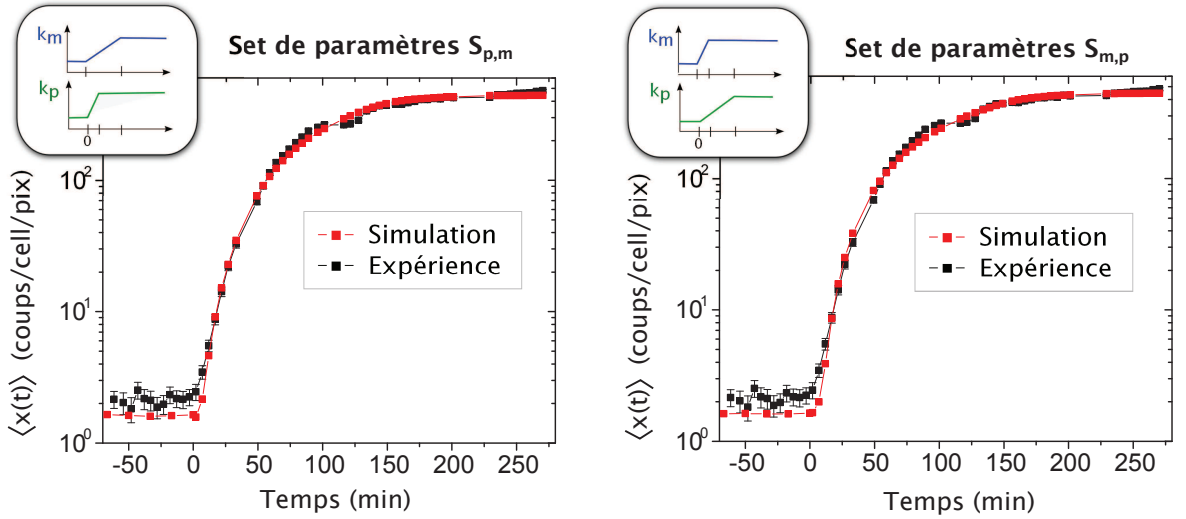


FIGURE IV.6.7.: Comparaison de $\langle x(t)\rangle_{simu}$ avec $\langle x(t)\rangle_{exp}$. Les courbes simulées sont tracées à partir de 10 000 trajectoires. Ces trajectoires sont obtenues pour le modèle à trois étages avec les taux de production qui varient selon une fonction rampe, comme rappelé dans la bulle du coin gauche. Les paramètres utilisés dans la simulation sont donnés dans le tableau IV.6.3 précédent.

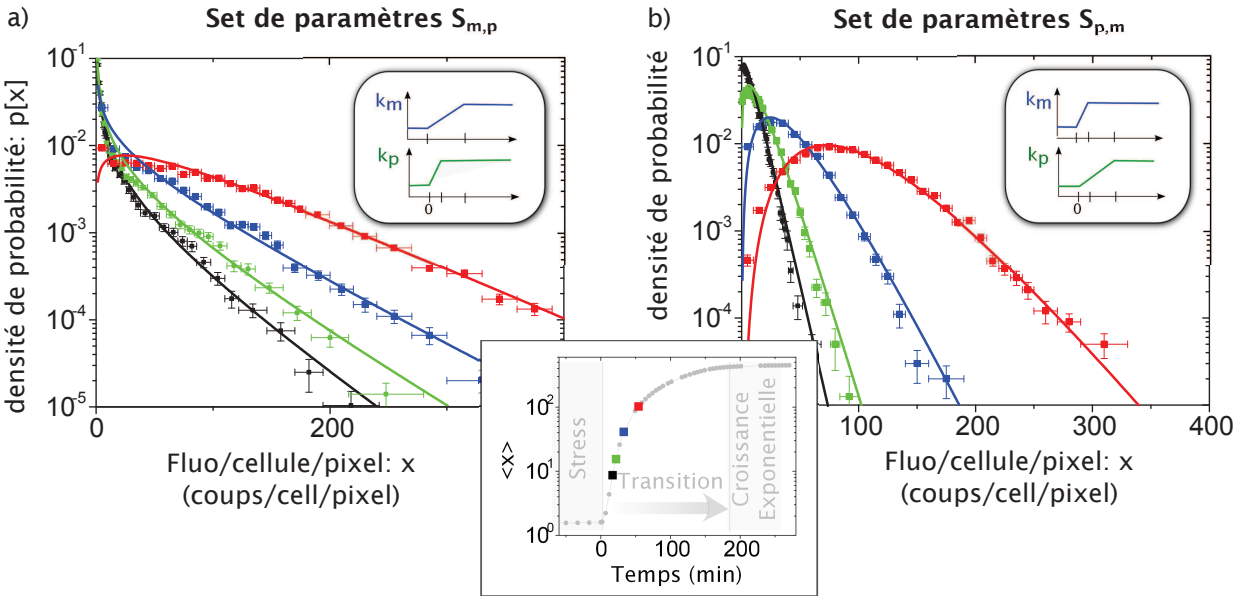


FIGURE IV.6.8.: Statistiques simulées avec le modèle à trois étage de la figure IV.6.5 et ajustement numérique par une distribution gamma. a). La statistique est simulée avec le set de paramètres $S_{m,p}$ du tableau IV.6.3. b). La statistique est simulée avec le set de paramètres $S_{p,m}$ du tableau IV.6.3. Les deux statistiques sont tracées à partir de 10 000 trajectoires pour des temps d'acquisition transitoires : en noir, la distribution simulée à $t = 17$ min, en vert, bleu et rouge, celles simulées à (resp.) $t = 22$ min $t = 33$ min, et $t = 49$ min. Le temps $t = 0$ min se réfère ici à l'injection de glucose dans le milieu de culture. Tous les temps d'acquisition considérés ici sont indiqués sur la courbe moyenne de concentration de fluorescence cellulaire (points colorés de même couleur que la distribution dans l'encadré du coin droit).

Si on réalise maintenant une étude plus fine sur la statistique entière (et non pas seulement sur le moment du 1^{er} ordre), on retrouve les résultats établis dans le cas du modèle à deux étages. En effet, sur la figure IV.6.8 les statistiques simulées (représentées par des points) peuvent encore une fois être décrites pas une distribution gamma (lignes continues). De plus, les distributions simulées ressemblent aux distributions expérimentales, comme montré sur la figure IV.6.9 pour quelques temps d'acquisition.

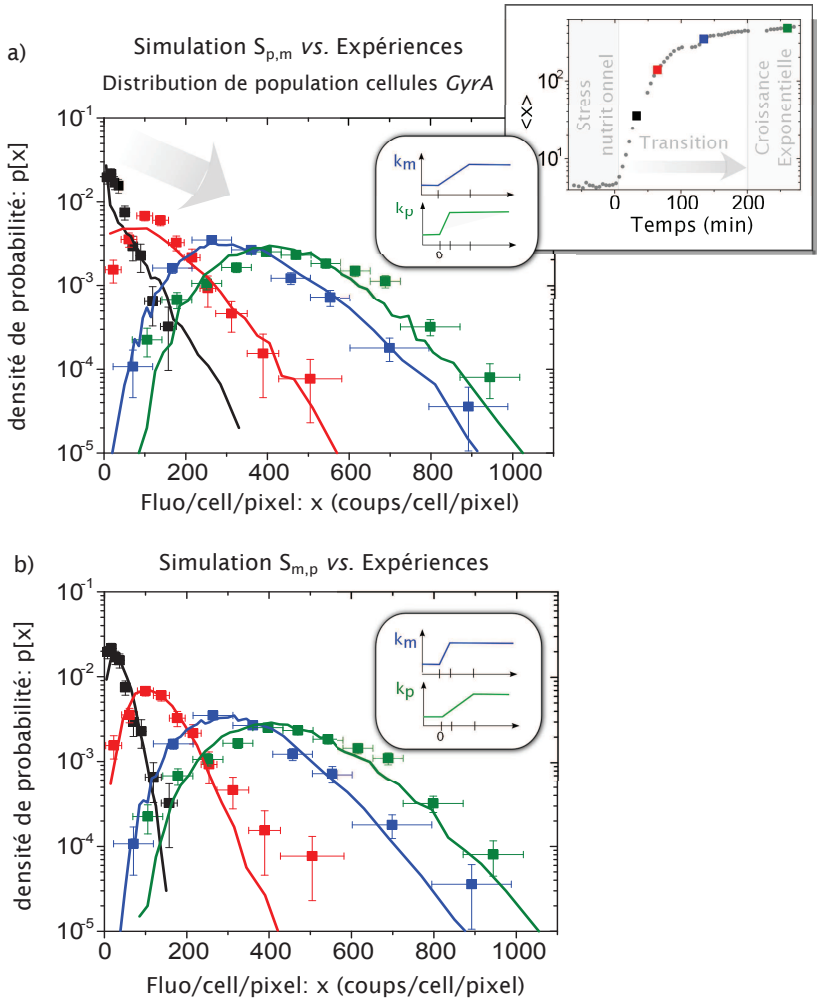


FIGURE IV.6.9.: Comparaison des statistiques simulée (ligne continue) et expérimentales (points de même couleur). Les distributions sont comparées pour 4 temps d'acquisition : ces temps sont indiqués sur la courbe de concentration moyenne de fluorescence cellulaire, en haut à droite (points de même couleur que les distributions). Les courbes simulées (selon les paramètres donnés tableau IV.6.3) sont tracées à partir de 10 000 trajectoires pour le modèle à trois étages de la figure IV.6.5. Par mesure de clarté, les barres d'erreur sur les statistiques simulées ne sont pas représentées. On note cependant que ce sont ces erreurs statistiques qui génèrent les petites oscillations sur les courbes simulées.

On décide alors de comparer plus finement les simulations aux données expérimentales en utilisant les paramètres a et b de l'ajustement gamma. Les résultats obtenus sont

présentés sur la figure IV.6.10 avec en noir les données expérimentales, et en rouge et bleu, les données simulées pour les sets de paramètres (resp.) $S_{m,p}$ et $S_{p,m}$ On constate que les résultats obtenus sont analogues à ceux fournis par le modèle à deux étages (cf. figure IV.5.8 de la section précédente).

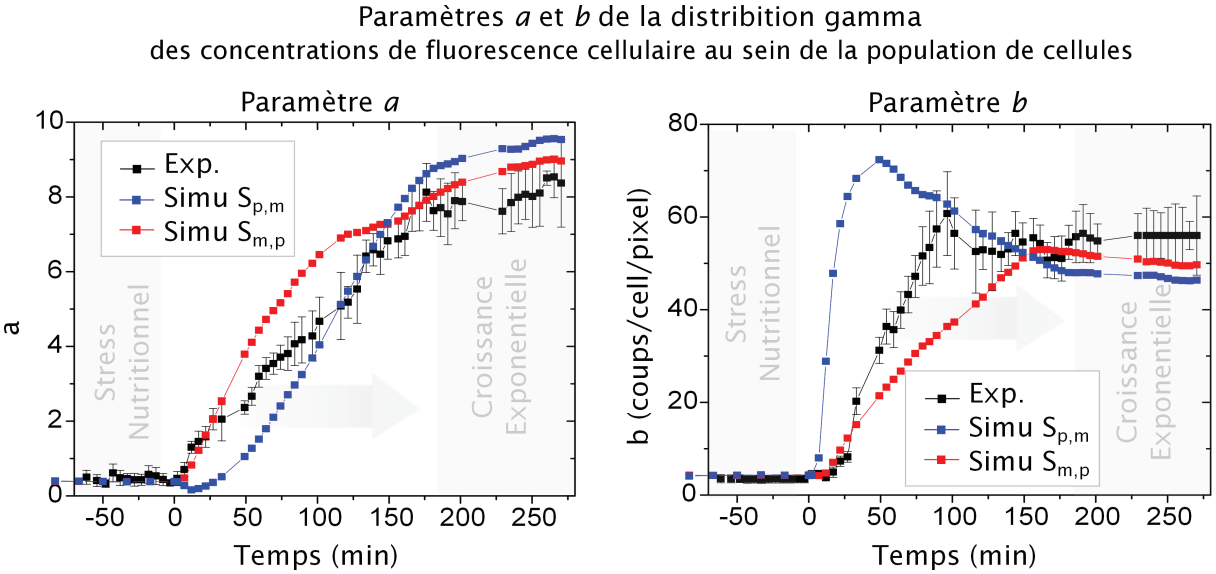


FIGURE IV.6.10.: Comparaison des simulations avec les expériences : paramètres a et b retournés par l'ajustement numérique gamma des distributions de population. En noir, les paramètres tirés des distributions expérimentales. En bleu et rouge, les paramètres tirés des distributions simulées. Ces distributions sont simulées pour un modèle à trois étages (transcription-traduction-maturation), avec les paramètres du tableau IV.6.3 : les courbes bleues sont simulées avec le set de paramètres $S_{p,m}$, et les courbes rouges avec le set de paramètres $S_{m,p}$.

Ces résultats préliminaires suggérent que **l'étage de maturation protéique ne change pas les conclusions de la section IV.5**.

Nous proposons par la suite d'étudier l'influence de cet étage sur le niveau de bruit extrinsèque mesuré par le système de double rapporteurs chromosomaux.

Le processus de maturation caractérisé par des taux différents pour les deux rapporteurs (CFP et YFP) introduit-il de la variabilité apparente dans les mesures qui comparent η^2_{ext} *et* η^2_{int} *?*

Si elle s'avère concluante, cette étude supplémentaire permettra de corriger une idée encore largement véhiculée par la communauté scientifique, à savoir que le système de doubles rapporteurs chromosomaux permet de mesurer l'importance réèle de η^2_{ext} par rapport à η^2_{int}.

Conclusions

Nous avons montré que de nombreuses sources de bruit extrinsèques peuvent être négligées dans notre modèle. Parmi les sources extrinsèques qui restent à investiguer, on trouve:

1. Le bruit sur le taux de production protéique (k_p). Ce dernier peut être étudié en lançant des simulations où le paramètre k_p varie de manière aléatoire autour de sa valeur moyenne. Nous proposons dans ce cas de décrire ces variations par une loi de probabilité gaussienne ou log-normale. Les données ainsi simulées pourront être comparées aux données obtenues lorsque k_p est considéré déterministe. Cette comparaison devraient mesurer l'influence du bruit de k_p sur les fluctuations en sortie du module d'expression génique. De plus, en comparant les données simulées aux fluctuations mesurées dans nos expériences, on devrait pouvoir déterminer si cette source extrinsèque de variabilité joue un rôle important dans nos mesures. On note ici que le taux de production protéique est déterminé par la quantité de ribosomes et d'ARNt dans la cellule. Les premières molécules se trouvent souvent en grand nombre dans la cellule, ce qui suggère que le bruit sur k_p est faible. Par contre, les ARNt peuvent se trouver en nombre plus réduit lors du stress nutritionnel. Ils peuvent donc introduire une variabilité non négligeable sur k_p.
2. Le bruit sur le taux de dégradation protéique (γ_p). On propose de lancer des simulations qui intégrent maintenant la variabilité sur γ_p. En comparant ces simulations aux résultats obtenus pour γ_p déterministe, on peut évaluer l'impact de cette variabilité extrinsèque sur le bruit en sortie du module d'expression génique.
3. L'étude sur le bruit introduit par l'étage de maturation doit être complétée avec des simulations qui mimeraient les expériences menées sur les doubles rapporteurs chromosomaux (et non sur les rapporteurs plasmidiques). Cette étude supplémentaire permettra d'étudier l'influence de l'étage de maturation sur les mesures de bruit extrinsèque. Quelle part occupe la différence du taux de maturation des doubles rapporteurs dans la valeur de η^2_{ext} mesurée? La variabilité extrinsèque mesurée par le système de doubles rapporteurs chromosomaux reflète-t-elle réelement les sources extrinsèques de bruit? Ou retranscrit-elle plutôt les différences dans le taux de maturation des deux rapporteurs?

Si les sources extrinsèques non étudiées sont effectivement négligeables dans nos expériences, alors le modèle proposé est valide. Une fois complété (test des valeurs de t_{km} et t_{kp} comprises entre les deux minimums de la carte de la figure IV.5.4), il permet de remonter aux informations dynamiques sur les paramètres du module d'expression génique sondé.

IV.7. Conclusions et perspectives

Dans ce travail de thèse, nous avons développé des méthodes qui permettent de sonder la variabilité en sortie d'un module d'expression génique. A l'instar de nombreux travaux qui se focalisent sur des conditions environementales statiques, cette variabilité est mesurée dans un environnement extracellulaire dynamique : nos cellules passent d'un état de stress nutritionnel à un état de croissance exponentielle. Cette transition d'état est implémentée par un réseau de régulation naturel qui implique le module étudié.
Pour sonder l'évolution du bruit dans le module, nous avons élaboré un protocole expérimental qui permet d'observer la transition d'état. Les paramètres critiques du protocole ont été identifiés pour être contrôlés avec plus de rigueur. Et comme le système d'imagerie est suceptible d'introduire des artefacts dans la variabilité intercellulaire mesurée, nous avons étudié les caractéristiques du dispositif microscopique et celles du systéme d'acquisition utilisé. Le signal mesuré a ainsi pu être corrigé des artefacts intrumentaux, et la variabilité inter-cellulaire observée a pu être imputée aux mécanismes cellulaires. A partir de là, nous avons développé des méthodes pour analyser les données collectées. Les cellules imagées ont été traitées par le programme fourni par M.B.Elowitz qui a été supplémenté de deux sous-routines Matlab. Ces opérations supplémentaires permettent d'éliminer les cellules "malades" parmis les individus détectés. Elles apparaissent essentielles. En effet, lorsque les cellules non saines sont inclues dans les mesures de bruit d'expression génique, la signature du bruit est changée. Suite à cette observation, nous avons décidé de nous placer dans des conditions idéales, où toutes les cellules sont saines. A partir de là, nous avons mis en place des outils d'analyse de données qui permettent de quantifier la signature du bruit (déconvolution, tracé des distributions avec une taille de bin variable). Pour modéliser les mécanismes responsables de cette signature, nous avons développé des outils de simulation qui reproduisent un module d'expression génique avec des taux de réaction qui varient dans le temps.

Une fois ces outils de mesure et d'analyse en place, nous avons proposé une méthode qui permet d'extraire les informations dynamiques du module d'expression génique à partir des mesures de variabilité inter-cellulaire. Cette méthode se résume à 4 étapes :

1. Caractérisation du bruit d'expression génique : dans nos expériences, la signature du bruit peut être caractérisée par une distribution gamma, et ce à chaque instant

de la transition d'état.

2. Proposition d'un modèle qui reproduit les mécanismes en jeu : le modèle proposé est simplifié aux étages de transcription et de traduction, avec des taux de production qui varient dans le temps.
3. Réduction du nombre de paramètres du modèle :
   a) Mesure des taux de dégradation du modèle : les mesures sur cellules uniques permettent de mesurer la dynamique du taux de dilution lors de la transition d'état. Les mesures en population des biologistes permettent de mesurer la dégradation active des variables (ARNm et protéines) du modèle.
   b) Extraction des taux de production en début et fin de transition d'état, à partir des mesures de bruit d'expression génique. Pour ces deux conditions d'équilibre statistique, la signature du bruit en forme de distribution gamma permet d'extraire les taux de production : ces derniers sont directement reliés aux paramètres de la distribution.
4. Ajustement du modèle sur les données expérimentales : les paramètres inconnues du modèle (2 paramètres) sont ajustés sur les expériences.

Les résultats fournis par cette méthode doivent bien sûr être complétés, mais ils sont déjà très prometteurs. En effet, ce modèle simplifié est ajusté sur seulement deux paramètres, et il fournit des données qui reproduisent globalement nos expériences. Pour apporter plus de poids à cette conclusion, des études supplémentaires doivent être réalisées, notamment sur les sources extrinsèques de bruit non étudiées. Si ces études complémentaires s'avérent concluantes, alors notre méthode permet d'accéder *in-vivo* aux propriétés dynamiques des paramètres internes du module. Ainsi, bien plus qu'une nuisance dans les expériences, le bruit renferme des informations utiles !

Module Fis du réseau de régulation étudié Nous proposons par la suite d'étudier le module Fis qui fournit une signature du bruit différente du module GyrA. Des expériences sur ce module ont déjà été réalisées, et les données collectées doivent être analysées et modélisées.

Microfluidique Nous proposons également d'étudier la transition d'état inverse chez *E. Coli* : les cellules passeraient d'un état de croissance exponentielle à un état de stress nutritionnel. Dans ce but, nous avons commencé à construire un circuit microfluidique qui permet de changer rapidement le milieu de culture. Ce système présente de nombreux avantages par rapport au système de collage sur boîte de pétri actuellement utilisé. Il permet d'une part de changer le milieu de culture optimal en un milieu défaillant en

sources carbonnées, d'autre part d'observer les cellules sur des temps plus longs. Dans ce système, les cellules sont piégées dans des canaux de dimensions micrométriques (quelques µm) qui sont alimentés en permanence par un flux continu de milieu. Lorsque les cellules se multiplient, les canaux se remplissent. Lorsque les canaux sont remplis, les nouvelles cellules sont évacuées par le flux continu de milieu vers la sortie (S). Tout ceci est illustré sur la figure IV.7.1.
Le circuit élaboré s'appuie sur les travaux de P.Wang et al [17], et nous y avons rajouté un composant qui permet de changer rapidement le milieu de culture (le "vent" sur la figure IV.7.1). Pour dessiner ce composant, nous nous sommes inspirées des travaux de A.Groissman et al [62].

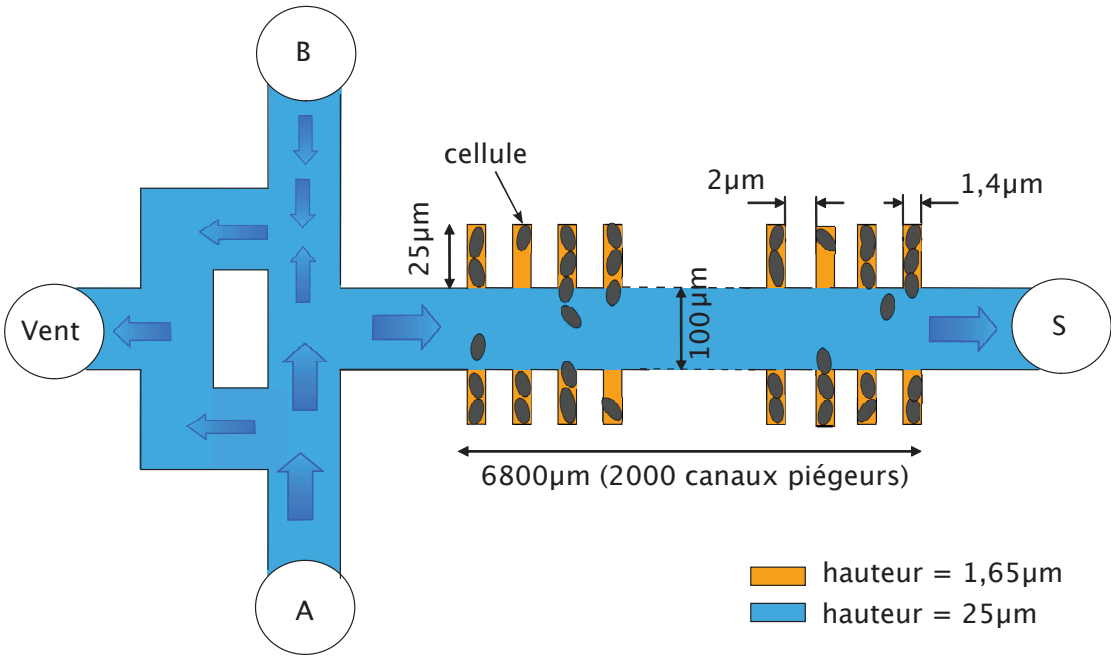


FIGURE IV.7.1.: Principe du système microfluidique : les cellules (ovales gris) sont piégées dans des canaux de dimensions proches des leurs (canaux oranges). Ces canaux sont alimentés en permanence par un flux de milieu continu (canaux bleus) : le milieu circule de l'entrée A (ou B) vers la sortie S. Ce milieu diffuse dans les canaux piégeurs, ce qui permet de rafraîchir le milieu dans lequel baignent les cellules : dans les canaux piégeurs, le temps de diffusion de molécules de taille nanométriques (4nm-38 nm) est de $\approx$ 1s-10s [17]. Le milieu de culture peut être changé en appliquant un différentiel de pression sur les entrées A et B : il peut provenir soit de l'entrée A, soit de l'entrée B. Ce changement de milieu est quasi-instantané ; il est effectué grâce au systéme de Vent (canaux qui relient les entrées A et B et la sortie "Vent") [62].

Les plans du circuit sont prêts, et nous avons réalisé une version "test" du circuit. Dans cette version, les échelles sont agrandies (dimensions minimales de l'ordre de 25 µm), et les cellules sont remplacées par des billes de taille mésoscopique ($\phi = 20$µm). Avec ces échelles agrandies, la construction du circuit est peu coûteuse. Ce circuit de taille supérieure est donc bien adapté pour les premiers tests.

Pour construire ce circuit "test", nous avons développé un protocole qui peut être réalisé dans la nouvelle salle blanche du laboratoire. La construction repose sur un processus de lithographie suivi d'un moulage sous hôte du circuit imprimé dans du PDMS. Le protocole élaboré implique notamment de construire un système à deux étages où deux motifs de différente hauteur (canaux oranges et bleus sur la figure IV.7.1) sont alignés. Il est actuellement au point, et des premiers tests ont pu être effectuées. Ces tests montrent que le composant rajouté pour changer le milieu de culture marche ! Il reste maintenant à construire la version réduite du circuit (adaptée à la taille des cellules), et à élaborer un protocole expérimentale qui permet d'utiliser ce circuit.

Processus de quorum sensing chez *V. Harvey* Nous proposons aussi d'étendre la méthodologie développée à un autre organisme : les cellules *Vibrio Harvey*. Ces bactéries sont capables de communiquer entre elles pour réguler collectivement l'expression génique en réponse aux changements de densité cellulaire. Ce type de réponse s'appelle le "*quorum sensing*". Il implique la production, le relarguage, et la redétection de molécules signalisatrices nommées autoinducteurs. Lorsque la densité cellulaire est importante, la concentration extra-cellulaire d'autoinducteurs augmente, et elle finit par atteindre un seuil critique au delà duquel le groupe bactérien répond en altérant sa propre expression génique. Cela permet aux bactéries d'adopter un comportement particulier lorsqu'elles vivent en communauté. Par exemple, le quorum sensing contrôle la sécretion de facteurs de virulence, la formation des biofilms, ou la bioluminescence [63]. Nous proposons ici d'étudier le quorum sensing qui contrôle la bioluminescence des bactéries *Vibrio Harvey*. Dans ce cas, le module d'expression génique étudié concerne les gènes de la bioluminescence et l'étage de maturation protéique est remplacé par la réaction de bioluminescence. Le signal en sortie du module est alors directement rapporté par la bioluminescence naturelle des cellules.

Des expériences de microscopie ont déjà été réalisées par I.Mihalcescu [64]. Le niveau de bioluminescence a été mesuré sur des mutants *V. Harvey* qui ne sont sensibles qu'à un seul autoinducteur [1]. Il apparaît alors à tout moment de l'induction du quorum sensing, ou à l'inverse, à tout moment de l'arrêt du quorum sensing, une distribution de la densité de bioluminescence cellulaire qui peut être décrite par une distribution gamma (voir figure IV.7.2).

On se demande alors si notre méthodologie peut être utilisée pour décrire cet autre phénoméne dynamique naturel. Si nos méthodes peuvent être appliquées à la bioluminescence de *V.Harvey*, elles permettent d'extraire des informations dynamiques sur les paramètres du module d'expression génique.

1. Ces mutants ont été fournis par B.Bassler.

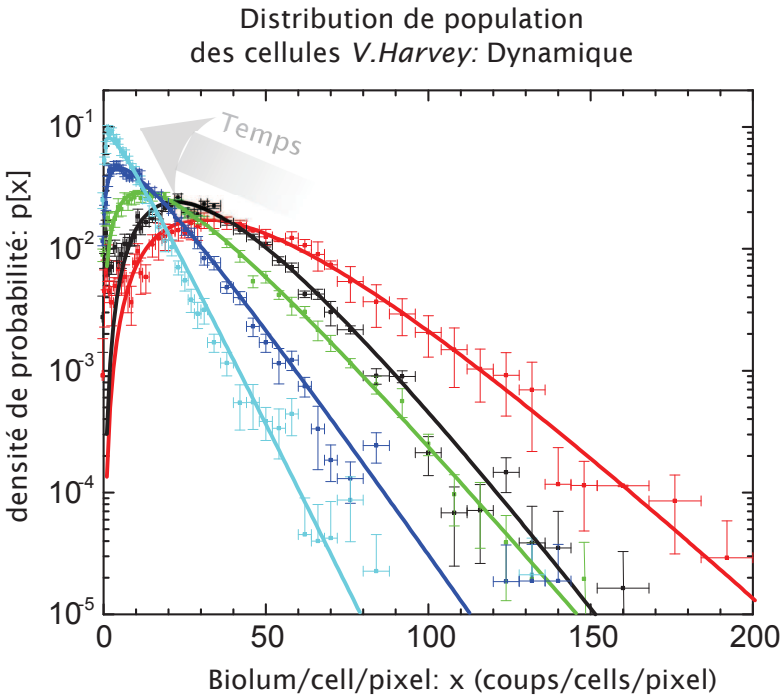


FIGURE IV.7.2.: Distributions de population tracées sur des cellules *Vibrio Harvey* qui passent d'un état de bioluminescence élevé (comportement de collectivité) à un état de bioluminescence faible (comportement d'individus isolés). La transition d'état est initiée par la suppression des molécules inductrices initialement présentes dans le milieu de culture en concentration supérieure au seuil de déclenchement du quorum sensing. Les distributions sont représentées par les points et leurs barres d'erreur, et les ajustements gamma sont donnés par les courbes continues de même couleur. D'après I.Mihalcescu [64].

Cinquième partie .

ANNEXES

A. Composition du milieu M9

Milieu M9

Milieu M9 sans glucose (pour 100mL de solution)	
H_2Omq	$78,4$mL
Sels×5	20mL
$MgSO_4$	200µL
$CaCl_2$	300µL
Sels×1000	100µL
Thiamine/Vit B11	1mL

TABLE A.1.: Composition du mileu M9 ne contenant pas de glucose.

Milieu M9 supplémenté en glucose $0,2$%

Milieu M9 avec glucose (pour 100mL de solution)	
H_2Omq	$77,4$mL
Sels×5	20mL
$MgSO_4$	200µL
$CaCl_2$	300µL
Sels×1000	100µL
Thiamine/Vit B11	1mL
Glucose 20%	1mL

TABLE A.2.: Composition du milieu M9 contenant du glucose.

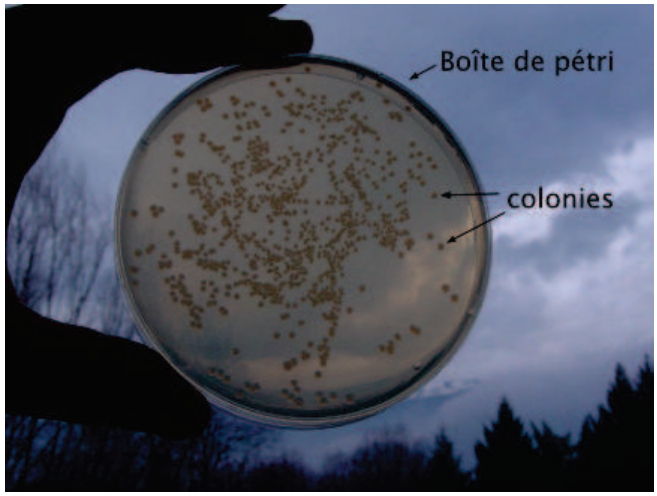
B. Correspondance entre absorbance et concentration cellulaire

Une solution initiale (S_0) de cellules a été préparée à partir de souches *GyrA* : ces dernières ont été cultivées pendant ≈ 16 heures, dans 6mL de M9 supplémenté en glucose (2%) et en Ampicilline (100µg/mL).

Cette préculture est ensuite diluée dans du milieu frais à différents taux : $\frac{8}{16}$, $\frac{6}{16}$, $\frac{4}{16}$, $\frac{3}{16}$, $\frac{2}{16}$, $\frac{1}{10}$, $\frac{1}{16}$, $\frac{1}{32}$, $\frac{1}{64}$, $\frac{1}{100}$et $\frac{1}{128}$. Pour chacune de ces solutions diluée, on mesure l'absorbance (photométrie à la longueur d'onde $\lambda = 600$nm). On obtient alors la correspondance entre Absorbance et Taux de dilution.

Et pour relier l'absorbance à la concentration cellulaire, il faut maintenant établir la relation entre la concentration cellulaire et le taux de dilution.
Pour ce faire, on détermine le nombre N_{cells} de cellules présentes dans la solution initiale : la concentration cellulaire des dilutions se déduit alors aisément lorsqu'on connaît le volume de S_0 dilué.

Pour déterminer le nombre de cellules présentes dans la solution initiale, on étale 3 solutions (préculture S_0 diluée $\frac{1}{10^6}$, $\frac{1}{10^7}$ et $\frac{1}{10^8}$) sur des boîtes de pétri distinctes. Les bactéries évoluent maintenant sur un gel d'agarose mélangé à du milieu LB. Et le LB procure tous les nutriments nécessaires à la croissance cellulaire : chaque cellule individuelle étalée forme une colonie, comme montré sur la photo suivante.



En comptant ces colonies sur chaque boîte de pétri préparée, on obtient une correspondance entre le taux de dilution et le nombre de cellules étalées. Cette relation est linéaire, comme montré dans la figure B.1 ci-dessous. La pente de la droite est de $5,5.10^9$cellules/mL : on en déduit que la solution initiale (dilution 1) comporte : $N_{cells} = 5,5.10^9$.

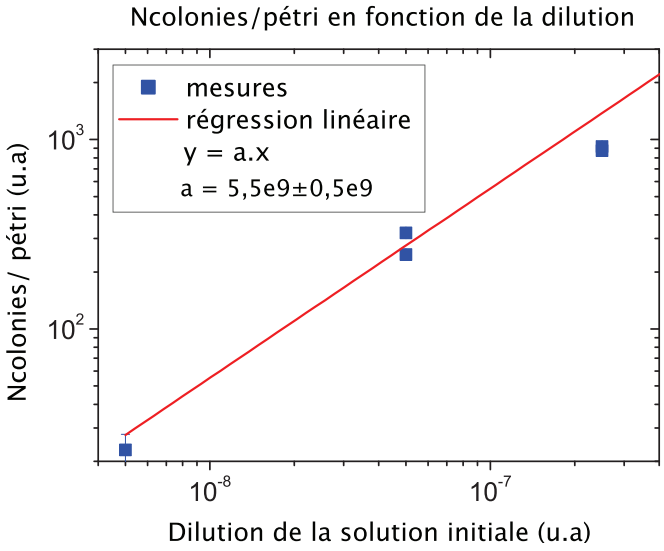


FIGURE B.1.: Nombre de colonies comptées dans le pétri, selon la solution étalée. Les solutions étalées sont obtenues à partir de la même solution initiale S_0, diluée à différents taux dans du milieu frais (dilution $\frac{1}{10^6}$, $\frac{1}{10^7}$ ou $\frac{1}{10^8}$).

On peut maintenant établir la correspondance entre l'Absorbance et la concentration cellulaire : on obtient la figure B.2 ci-dessous.

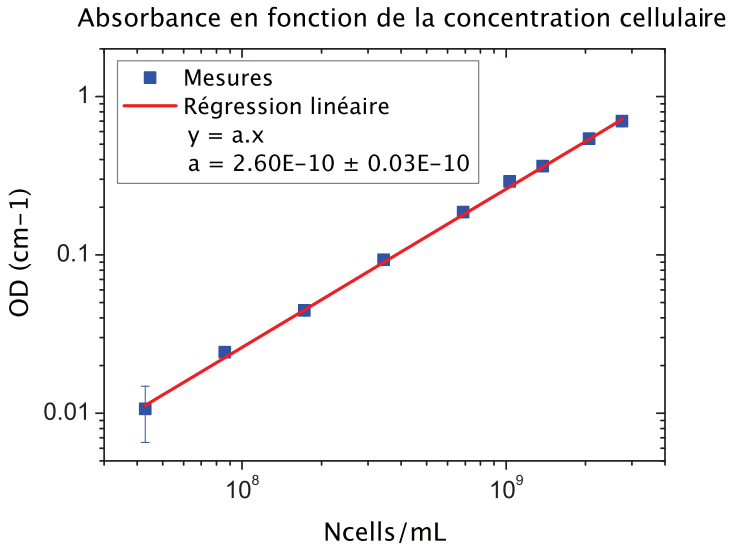


FIGURE B.2.: Correspondance entre l'absorbance mesurée (OD_{600}) et la concentration cellulaire : les points bleu représentent les données expérimentales avec des barres d'erreur statistiques poissonniennes ; La ligne continue rouge est une régression linéaire des données.

C. Programme de segmentation des cellules sur les images de contraste de phase

Nous décrivons ici les étapes principales du programme de segmentation développé par M.Elowitz sous Matlab, *Schnitzcells*. Ce programme initialement conçu pour les cellules *E. coli* ou les bacilles a été adapté aux cellules Vibrio Harvey: ces cellules ont la même forme en bâtonnet que Escherichia coli, mais leur taille est ≈ 2 fois plus petite. Les paramètres changés pour détecter correctement ces petites cellules sont donnés pour chaque étape de la segmentation. Ces étapes se résument à:

1. Pré-traitement: ajuster la taille des images acquises en bioluminescence et contraste de phase
2. Pré-traitement: élimination du bruit poivre et sel et ré-échelonnage des niveaux de gris.

3. Détection de contours.
4. Réduction des images à la zone d'intérêt contenant toutes les cellules.
5. Localisation des cellules.
6. Détection des cellules: remplissage des contours cellulaires localisés par une cellule.
7. Séparation des agrégats cellulaires.
8. Séparation des cellules filles fraîchement divisées.
9. Séparation des cellules accolées.
10. Elimination des petits objets artefactuels, labelisation et dilatation des cellules.

a. Pré-traitement-1

Une sous-routine formalise les images acquises sous forme de fichiers ".tif" et les dossiers nécessaires au rangement des données sont créés.
Dans le cas des cellules Vibrio Harveyi, cette subroutine contient une étape supplémentaire permettant d'ajuster la taille des images acquises en bioluminescence et de sélectionner la zone d'intérêt:

◦ Mise à l'échelle des images de bioluminescence:
Les images acquises en contraste de phase mesurent 1032×1062 pixels alors que celles acquises en bioluminescence sont de résolution plus faible: 516×531 pixels. Nous devons donc redimensionner les images pour avoir des matrices de même taille.
On décide d'ajuster la taille des images de fluorescence; on garde ainsi la résolution optimale pour la détection des cellules sur les images de contraste de phase. Les images de bioluminescence sont agrandies d'un facteur 2 par une interpolation au plus proches voisins.

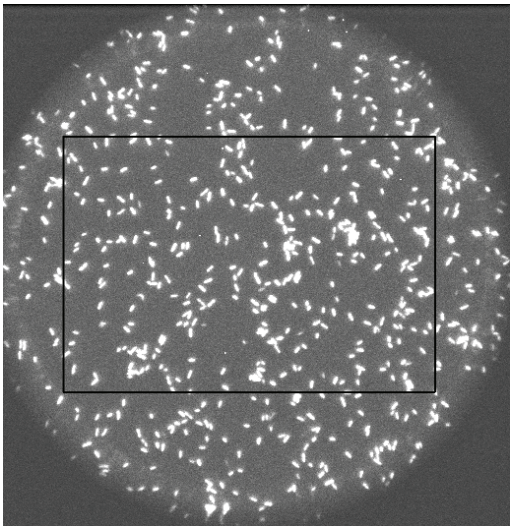


Figure C.1.: Image de bioluminescence. Les cellules Vibrio Harveyi sont confinées dans le disque donné par le tube du microscope.

◦ Sélection de la zone d'intérêt:
La zone exploitable des images est confinée dans un disque, donné par le tube de sortie du microscope. Pour permettre une segmentation efficace, on découpe les images: les variations de niveau de gris induites par le tube ne sont pas considérées.
Par soucis de simplicité, la découpe est rectangulaire. Les paramètres de découpe sont déterminés sur les images de bioluminescence (figure C.1), en prenant soin de ne pas sélectionner les cellules entâchées par les effets de bords du tube du microscope.

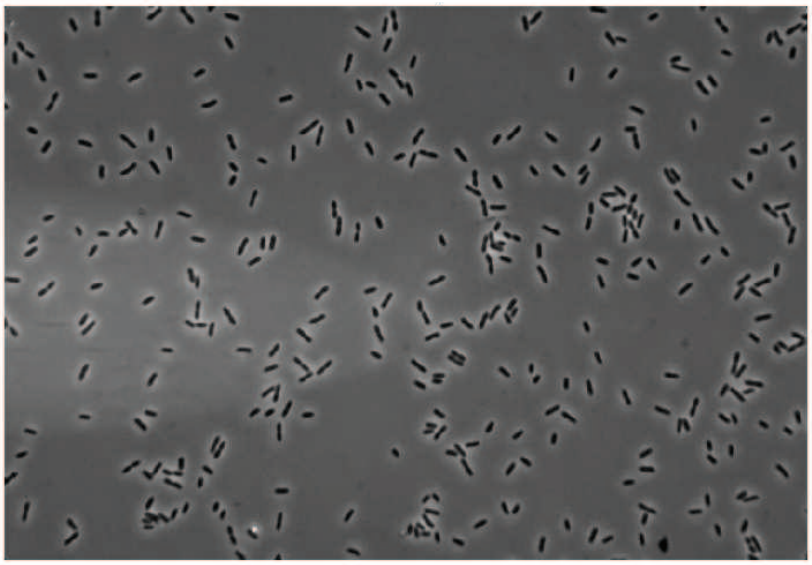
b. Pré-traitement -2

Les images acquises en contraste de phase sont filtrées par un filtre médian de voisinage 3×3 pour éliminer le bruit poivre et sel. Les niveaux de gris résultants sont échelonnés de 0 à 10000. Tous les traitements qui suivent se font à partir de cette image. Elle est représentée sur la figure C.2-a.

c. Détection de contours

Les contours cellulaires sont détectés par un filtre laplacien-gaussien (figure C.2-b). La taille du filtre est donnée par la déviation standarde σ de la gaussienne utilisée: on fixe $\sigma = 1,3$ pour V. Harvey ($\sigma = 2$ pour *E. Coli*).

a) Image Originale



b) Détection de contours par filtrage Laplacien-gaussien

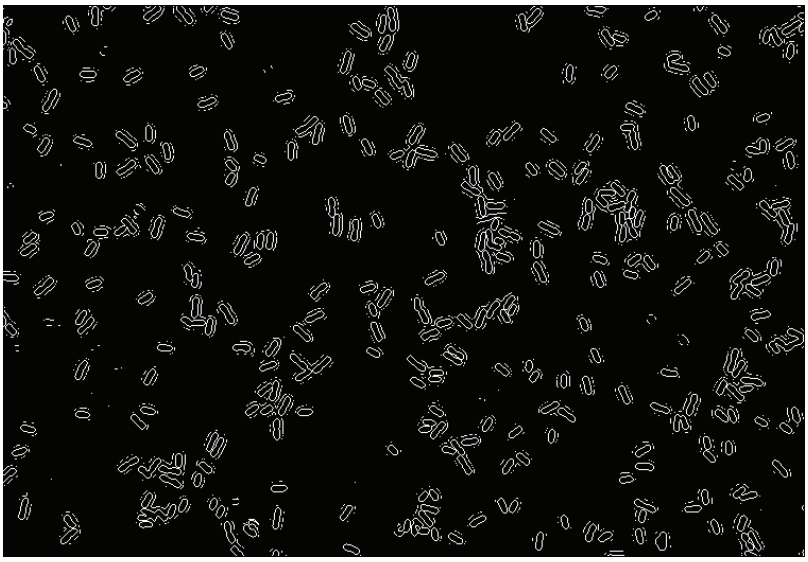


Figure C.2.: a). Image de contraste de phase après élimination du bruit poivre et sel et échelonnage des niveaux de gris sur [0 10000]. b). Détection de contours: image de contraste de phase après application d'un filtre laplacien-gaussien d'écart type $\sigma = 1,3$.

d. Réduction des images

Le programme développé par M.Elowitz et al travaille sur des images réduites, lestées des bords de l'image où les cellules sont absentes. Pour ce faire, schnitzcells détermine la *zone d'intérêt*, le plus petit rectangle de l'image contenant toutes les cellules, et la segmentation est limitée à cette zone.

Zone d'intérêt: Elle est déterminé à partir des contours cellulaires: les régions contenant des cellules sont sélectionnées sur la longueur des contours, équivalente au périmètre cellulaire. La zone d'intérêt est alors définie comme le rectangle minimal contenant toutes les régions sélectionnées.

Sélection des régions occupées par les cellules: Des boîtes de taille $30 pixels \times 30 pixels$ parcourent l'image de contours; si elles renferment un nombre de pixels-contours inférieur au périmètre minimal d'une cellule, les contours sont effacés, la zone est colorée en blanc. Dans le cas contraire, la région contient une cellule, elle est sélectionnée par coloration noire.
Pour les cellules *E. Coli*, le périmètre minimal d'une cellule est estimé à 60 pixels; pour les cellules V.Harvey, nous l'estimons à 40 pixels.
Par mesure de précaution, les régions sélectionnées sont dilatées, puis lissées par ouverture morphologique (figure C.3); l'élément structurant utilisé est un disque de rayon 5 pixels pour la dilatation et 25 pixels pour l'ouverture (respectivement 5 et 15 pixels pour les cellules V.Harvey).
On obtient une sélection grossière (figure C.3) qui permet de déduire la zone d'intérêt.

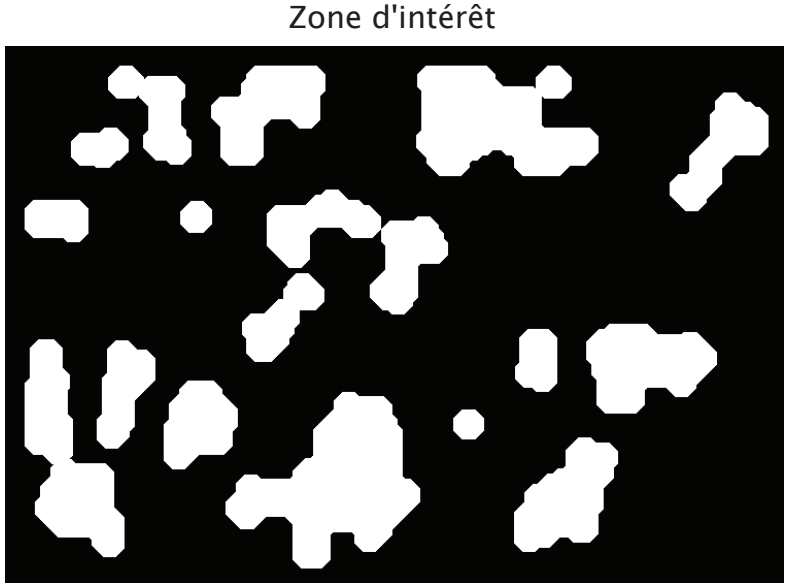


Figure C.3.: Région occupée par les cellules: les zones blanches sont considérées vides de cellules.

Images réduites: L'image de contour (figure C.2.a) et l'image originale (figure C.2.b) sont ensuite réduites aux dimensions du rectangle d'intérêt.

Par la suite, schnitzcells ne travaille que sur ces images réduites.

Image de contours lestée

Le programme développé par M.Elowitz simplifie davantage l'image de contours, en effaçant les régions vides de cellules.

Pour réaliser cette deuxième opération, schnitzcells construit de nouvelles régions d'intérêt. Ces régions apportent un complément d'information sur l'espace occupé par les cellules. Et tout objet détecté hors de cet espace sera effacé.

Construction de nouvelles zones d'intérêt

Cette fois-ci, les régions occupées par les cellules sont selectionnées sur leur niveau de gris. Ce dernier est particuliérement élevé pour les cellules par rapport au fond de l'image, et une binarisation de l'image originale [1] permet d'imposer le même niveau de gris à tout autre objet. Ainsi, après binarisation, les cellules se distinguent du reste de l'échantillon.

Elles sont alors lissées par fermeture morphologique (élement structurant circulaire de rayon 6 pixels, respectivement 4 pixels pour les cellules V.Harvey), et les trous intracellulaires sont remplis. Les cellules sont ensuites dilatées par un disque de rayon 3 pixels (respectivement 2 pixels pour V.Harvey).

Par mesure de précaution, les cellules sont dilatées une nouvelle fois et les trous intracellulaires résultants sont remplis; l'élément structurant utilisé pour cette dernière dilatation est un disque de rayon 10 pixels (respectivement 6 pixels pour les cellules V.Harvey). Les zones d'intérêt ainsi obtenues sont présentées sur la figure C.4-a.

Espace occupé par les cellules

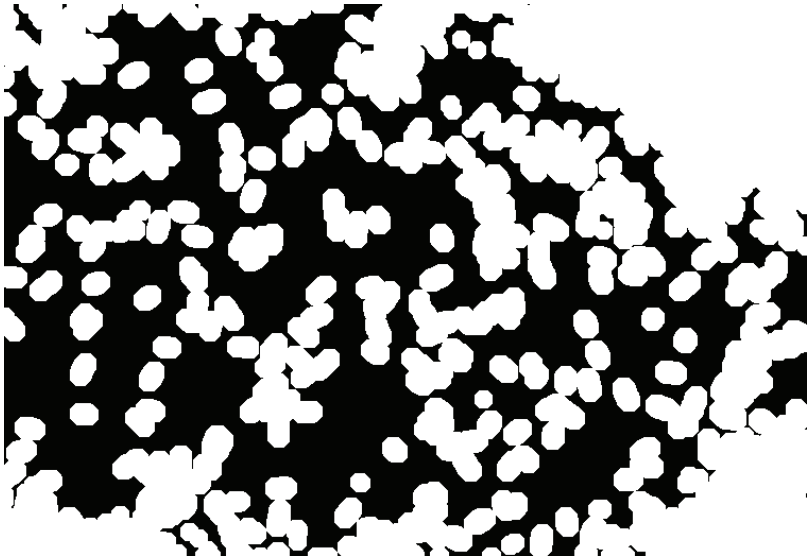
Les régions d'intérêt sélectionnées sur le niveau de gris sont combinées avec celles qui ont été déterminées à partir du contour cellulaire. Les régions résultantes représentent l'espace occupé par les cellules.

Image de contour lestée

Afin de ne considérer que les régions de l'image contenant des cellules, l'image de contour est filtrée par les régions d'intérêt: toutes les zones n'appartenant pas à l'espace occupé par les cellules sont effacées (voir figure C.4-b).

1. le seuillage d'histogramme utilisé s'inspire de la méthode d'Otsu ("A Threshold Selection Method from Gray Level Histograms", N.Otsu. IEEE 1979).

a) Régions d'intérêt



b) Image de contour lestée

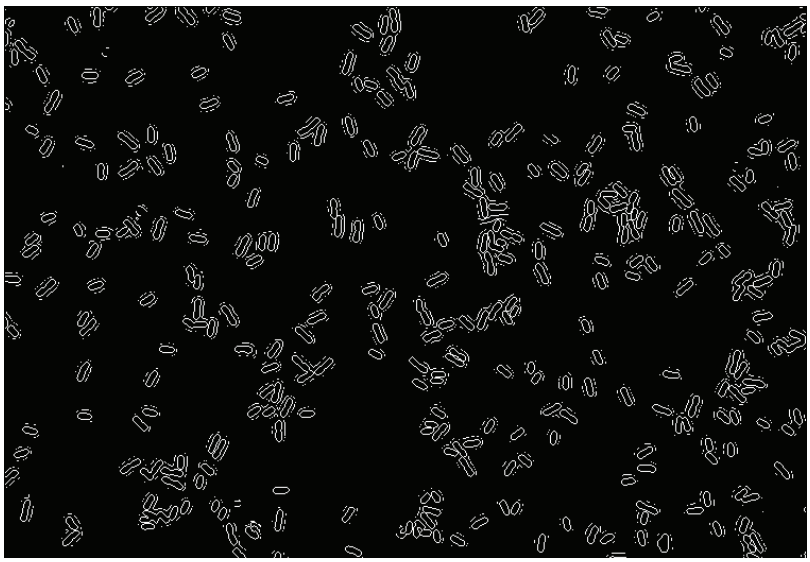


Figure C.4.: a). Régions d'intérêt déterminées sur le niveau de gris: les zones colorées en noir sont vides de cellules. b). Image de contour lestée des zones ne contenant pas de cellules.

e. Localisation des cellules

Remplissage des contours cellulaires

Sur l'image de contour lestée (figure C.4-b), les régions fermées comportant plus de 4 pixels correspondent à des cellules: elles seront coloriées en blanc, comme montré sur la figure C.5.

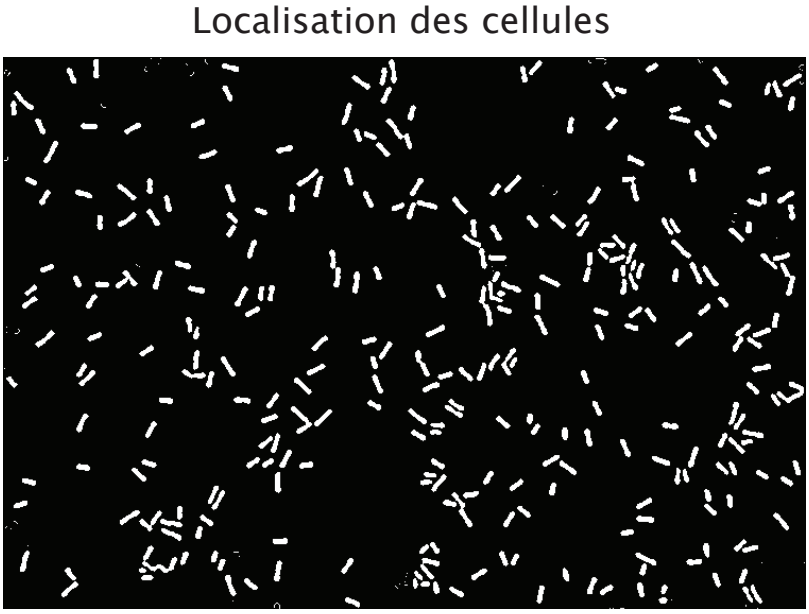


Figure C.5.: Image de contour lestée avec remplissage des cellules et élimination des faux contours cellulaires.

Elimination des faux contours cellulaires

La figure C.5 précédente comporte beaucoup de faux contours cellulaires. Ceux- ci sont effacés comme suit:

- Un filtre bottom hat [2] appliqué à l'image originale permet de ne faire ressortir que les cellules: presque tous les autres objets héritent d'un niveau de gris égal à zéro. L'élément structurant utilisé ici est un disque de rayon 5 pixels (respectivement 2 pixels pour les cellules V.Harvey).
- Une détection de contours basique, suivie d'un filtrage par les régions d'intérêt (cf figure C.6) fournit une nouvelle image de contours. Sur cette image, les faux contours cellulaires (traits blancs sur la figure C.6) ont quasiment disparu.
- Les nouveaux contours cellulaires ainsi obtenus sont remplis.

On obtient la figure C.6.

Nettoyage des petits objets artefactuels

Les cellules détectés sont assimilées à une ellipse; si la longueur de l'ellipse est inférieure à la longueur minimale de la cellule imagée, l'objet détecté correspond à une fausse détection. Dans ce cas, il est effacé. Pour les cellules E.Coli, la longueur minimale est fixée à 9 pixels; elle est fixée à 4 pixels pour les cellules V.Harvey. Le résultat peut être visualisé sur la figure C.7.

2. Image filtrée = Image originale avec fermeture morphologique - Image originale

Image de contour avec remplissage des cellules

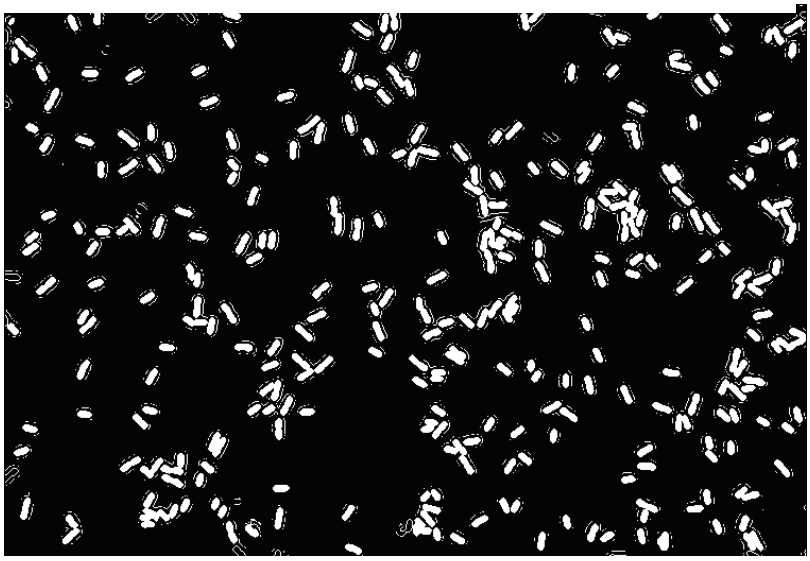


Figure C.6.: Image de contour lestée des zones ne contenant pas de cellules; coloriage des cellules.

Localisation des cellules, petits objets effacés

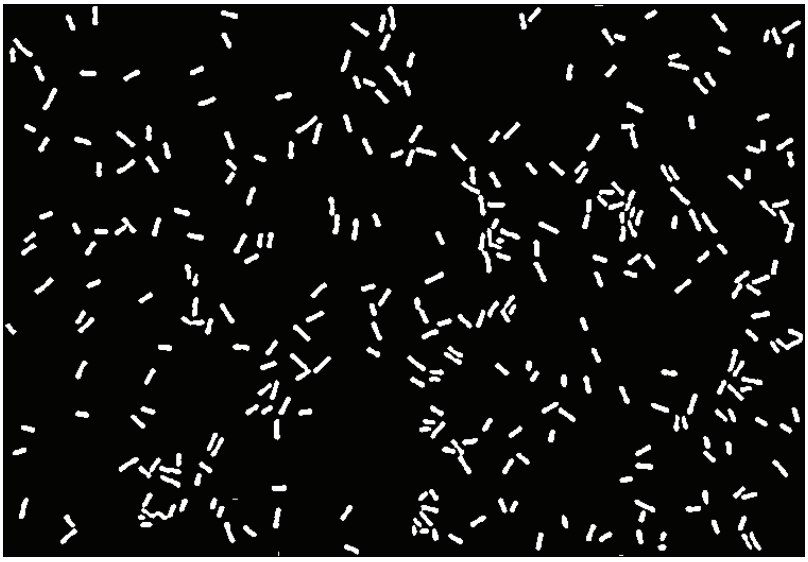


Figure C.7.: Image de contour lestée avec remplissage des cellules, élimination des faux contours cellulaires et nettoyage des petites objets artefactuels.

f. Détection de cellules

Les opérations précédentes fournissent un renseignement précis sur la localisation des cellules (figure C.7).

Les contours cellulaires peuvent maintenant être remplis plus rigoureusement: sur l'image de contours (figure C.6), tous les contours fermés pointés une cellule de la fig-

ure C.7 sont remplis par du blanc; tout le reste est coloré en noir.

On obtient une première segmentation de notre image (voir figure C.8):

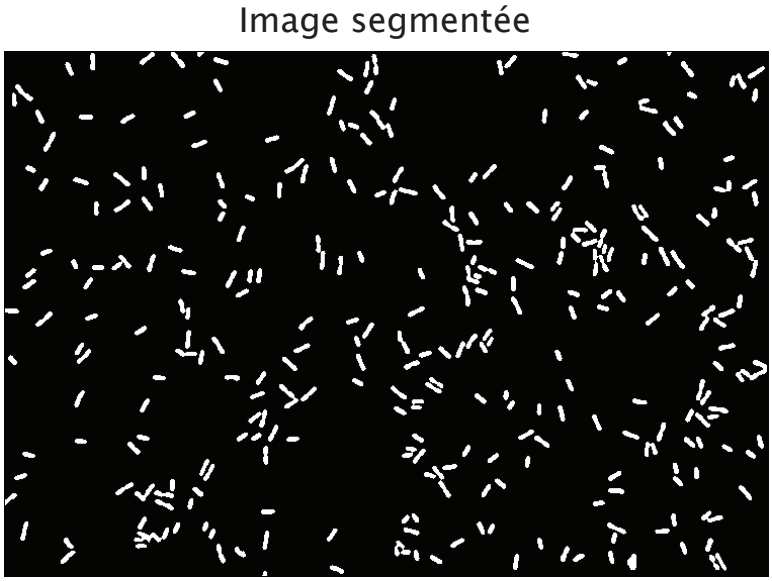


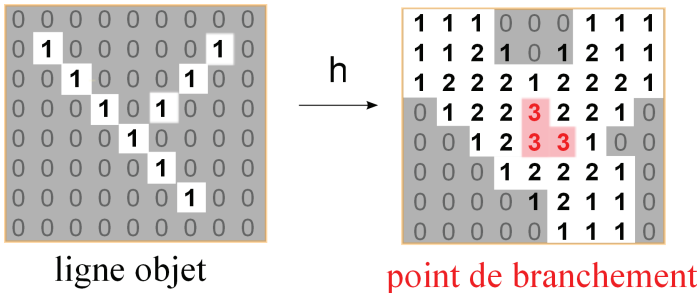
Figure C.8.: Détection des cellules.

g. Séparation des cellules agrégées

Les connections intercellulaires à 1 pixel sont effacées et les agrégats cellulaires sont séparés en plusieurs cellules distinctes, selon la courbure des contours.

Détection d'un agrégat cellulaire Pour chaque objet détecté dans la figure C.8, on cherche les *points de branchements*: les pixels appartenant à plusieurs cellules. Pour ce faire:

- Les objets détectés sont réduits à une ligne (*squelette*), comme schématisé sur la figure ci contre.

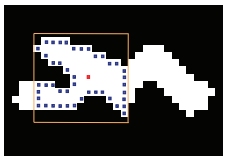


- Le squelette est élargi par un masque:

$$h = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

- Les pixels réassignés avec des valeurs supérieures à 3 sont des noeuds de branchement: l'objet détecté est un agrégat de plusieurs cellules.

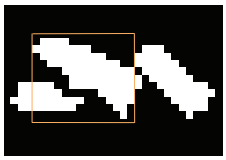
Détection des jonctions cellulaires: Schnitzcells étudie le périmètre de l'agrégat (en bleu marine sur la figure suivante) au voisinage des noeuds de branchements (en rouge sur la figure suivante): la courbure du périmètre indique les zones de jonction entre cellules. Nous prenons un voisinage d'ordre 9 (respectivement 6 pour les cellules V.Harvey), et avant de lancer cette étude, nous rajoutons un lissage des agrégats cellulaires pour les cellules V.Harvey.



- Pour chaque pixel du périmètre, la courbure est calculée à partir des 7 pixels-périmètre voisins chez *E. Coli* (respectivement 5 pour les cellules V.Harveyi). Lorsque le centre de courbure n'appartient pas à l'agrégat cellulaire, la zone étudiée est une zone de jonction entre cellules.
- Le maximum local des courbures ainsi sélectionnées donne le lieu exacte de jonction pour la zone étudiée. Le maximum local est calculé sur 6 courbures sélectionnées (respectivement 3 pour les cellules V.Harvey).
- Si deux maximums locaux sont espacés de moins de 3 pixels (respectivement 2 pour les cellules V.Harvey), ils appartiennent à la même zone de jonction. Dans ce cas, le lieu de jonction correspond au maximum local le plus grand.
- De tous les lieux de jonction ainsi détectés, seuls ceux dont la courbure est suffisamment convexe sont selectionnés. Pour les cellules *E. Coli*, la convexité est considérée suffisante lorsque la courbure varie de plus de 0.12 dans la zone de jonction (respectivement 0.001 pour les cellules V.Harveyi).

Séparation des cellules: Si l'objet détecté comporte plus d'un point de jonction, on cherche le point de jonction associé pour couper l'agrégat en deux.

Deux points de jonction sont dits associés si:



- La distance qui les sépare est inférieure à la largeur maximale de la cellule (fixée à 8 pixels pour les cellules *E. Coli* et 5 pixels pour les cellules V.Harvey).
- Lorsque deux points de jonction différents sont associés à un même point selon ce critère, c'est le point le plus proche du point d'étude qui est l'associé.

Pour chaque couple de points associés, on dessine une ligne noire qui les connecte: l'agrégat est sectionné par la ligne (voir la figure ci-contre).

Nettoyage de petits objets artefactuels: Tout les objets ainsi obtenus sont effacés si leur taille est inférieure à 5 $pixels^2$. Et sur l'image résultante, les connections intercellulaires à 1 pixel sont aussi effacées. On obtient la figure C.9.

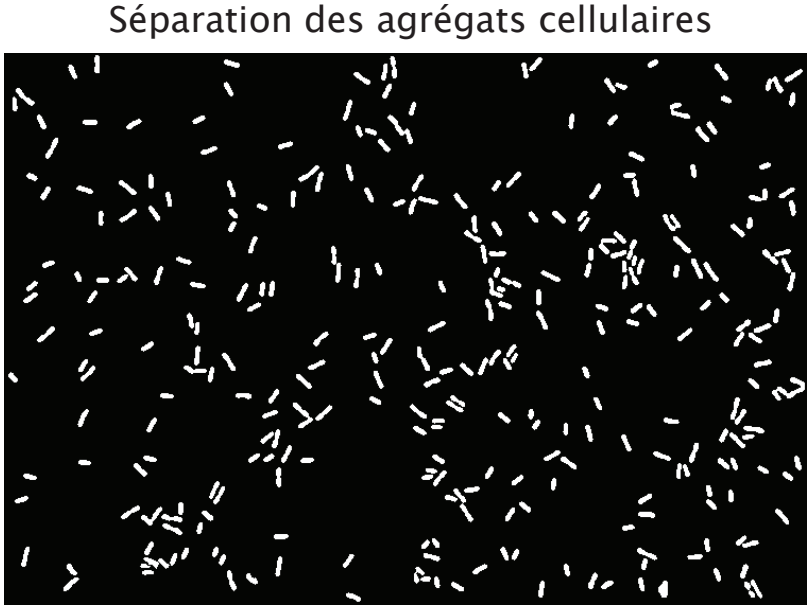


Figure C.9.: Image segmentée avec séparation des agrégats cellulaires en cellules distinctes.

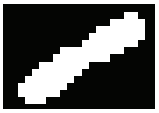
h. Séparation des cellules fraîchement divisée

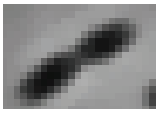
Les cellules filles fraîchement divisées ne se distinguent pas l'une de l'autre sur l'image C.9. Pour les séparer, Schnitzcells applique la même procédure décrite dans le paragraphe précédent. Mais il utilise cette fois-ci un critère moins exigeant sur la courbure: la convexité révèle un point de jonction lorsque la courbure dépasse $0,05$ (respectivement 10^{-3} pour les cellules V.Harvey). De plus, la séparation n'est validée que si les cellules résultantes ont une taille acceptable (longueur supérieure à 9 pixels pour les cellules E.Coli et 4 pixels pour les cellules V.Harvey).
Cette opération permet de séparer certaines cellules filles, mais beaucoup restent encore confondues comme le montre la figure C.10.

i. Scission des grosses cellules

Afin d'améliorer l'efficacité de séparation des cellules filles, schnitzcells applique une procédure de séparation supplémentaire. Cette fois-ci, les jonctions intercellulaires ne sont pas caractérisées par des contours fortmenent courbés, mais par un niveau de gris clair.

Figure C.10.: Image segmentée avec détection des événements de division cellulaire.





Détection des jonctions cellulaires: Les cellules qui ont une longueur supérieure à 30 pixels (respectivement 20 pour les cellules V.Harvey) sont de grosses cellules: elles peuvent englober deux cellules distinctes comme montré ci-contre.

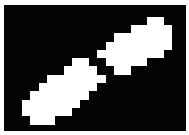
Pour déterminer si ces grosses cellules sont le fruit de deux cellules filles confondues, schnitzcells s'intéresse à l'intensité des pixels sur l'image de phase originale:

- Les grosses cellules sont squelettisées et chaque point du squelette est caractérisé par un niveau de gris; ce dernier est une moyenne de l'intensité des pixels voisins sur l'image de phase originale.
- Les zones de la cellule les plus claires sont recherchées: si un point-squelette est plus clair que ses 6 points-squelette voisins (respectivement 2 pour les cellules V.Harvey), il correspond à une zone claire.
- Si deux zones claires sont espacées de moins de 3 pixels (respectivement 2 pour les cellules V.Harvey), elles apartiennent à la même jonction cellulaire. Dans ce cas, le point squelette d'intérêt est celui qui a le niveau de gris le plus clair.
- Toute zone claire caractérisée par une intensité supérieure de 0.02 de celle de ses voisins (respectivement 0.001 pour les cellules V.Harvey), est une zone de branchement cellulaire: elle est sélectionnée.

Pour s'assurer de détecter tous les branchements cellulaires, on effectue la même opération

sur l'image segmentée: dans ce cas, les points squelettes de jonction sont caractérisés par un niveau de gris sombre.

Séparation des deux cellules filles: pour chaque zone de branchement cellulaire ainsi détectée, la cellule est coupée en deux. La coupe se fait par l'ajout d'un trait noir entre les deux points de jonction, les deux points de jonction étant définis comme les deux pixels du périmètre les plus proches de la zone de branchement.



Si les deux cellules résultantes ont une taille acceptable (longueur supérieure à 9 pixels pour les cellules E.Coli et 4 pixels pour les cellules V.Harvey), la séparation est validée. Au final, on obtient l'image C.11 ci dessous:

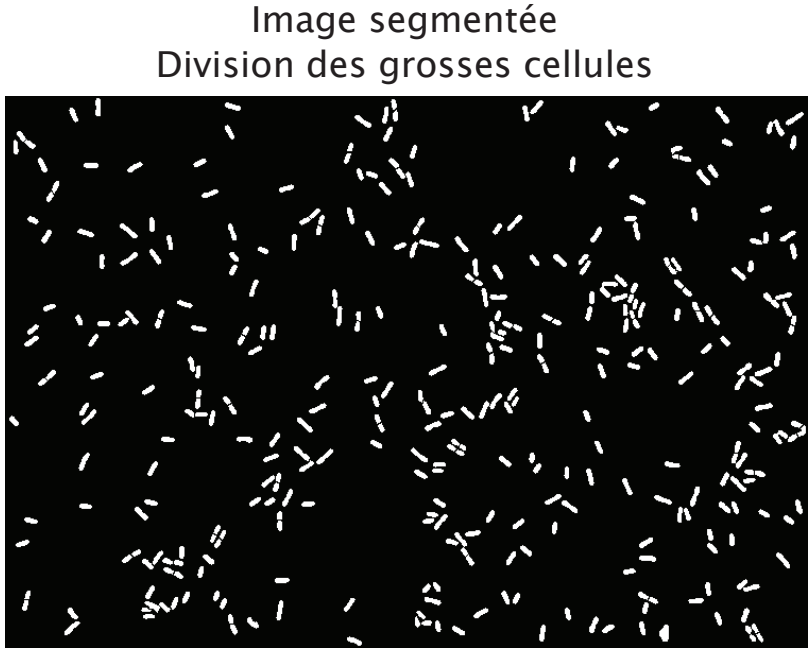


Figure C.11.: Image segmentée avec détection des événements de division cellulaire et division des grosses cellules.

j. Cloisonnement des cellules peu convexes

L'efficacité de séparation des cellules est encore améliorée en appliquant la même procédure, non pas aux grosses cellules, mais aux cellules qui comme sur la figure ci-contre sont peu convexes.

- Pour chaque cellules, Schnitzcells calcule le plus petit polygone convexe contenant la cellule.
- On définie alors la *solidité* comme la proportion de pixels du polygone convexe appartenant à la cellule.

- Si la solidité de la cellule est inférieure à 0.85 (voir figure ci contre), c'est qu'elle englobe deux cellules distinctes: celles ci sont alors séparées selon la procédure définie dans le paragraphe précédemment.
- Lorsque la procédure de séparation ne détecte pas de zone de branchement cellulaire, Schnitzcells applique une nouvelle procédure de séparation qui agit sur les points-squelette proches du centre de masse de la cellule.
  - Chacun de ces points est caractérisé par un angle: Schnitzcells calcule les angles formés avec les autres points squelette, et il considère l'angle maximal.
  - Les points-squelettes qui caractérisent la jonction sont ceux dont l'angle est plus petit que celui de leurs 6 points-squelette voisins (respectivement 2 pour les cellules V.Harvey).
  - Lorsque les jonctions ainsi caractérisées sont espacées de moins de 3 pixels (respectivement 2 pour les cellules V.Harvey), elles appartiennent à la même zone de branchement cellulaire. Dans ce cas, le point-squelette caractéristique est celui qui posséde l'angle le plus petit.
  - Toute jonction caractérisée par un angle inférieur à 2.7 est une zone de branchement cellulaire: elle est sélectionnée.
  - Lorsque l'espace séparant deux zones de branchement est inférieur à la longueur cellulaire minimale (9 pixels pour les cellules E.Coli et 4 pixels pour les cellules V.Harvey), les deux zones désignent le même noeud de branchement cellulaire: le point squelette caractéristique est moyenné sur les deux zones.
  - Pour chaque zone de jonction inter-cellulaire ainsi détectée, la cellule est coupée en deux. La coupe se fait par l'ajout d'un trait noir entre les deux points de jonction; ceux-ci étant définies comme les deux pixels du périmètre les plus proches du point-squelette de jonction.
  - Si les deux cellules résultantes ont une taille acceptable (longueur supérieure à 9 pixels pour les cellules E.Coli et 4 pixels pour les cellules V.Harvey), et une solidité acceptable (supérieure à 0.15), la séparation est acceptée.

Au final, on obtient l'image C.12.

k. Elimination des petites cellules

Une dernière opération est effectuée pour effacer les objets dont la surface est inférieure à la surface minimale d'une cellule (55 $pixels^2$ pour les cellules E.Coli et 36 $pixels^2$ pour les cellules V.Harvey).

L'image résultante est dilatée par un losange de rayon 1 pixel pour les cellules E.Coli. Pour les cellules V.Harvey, aucune dilatation n'est appliquée. Et les petites cellules dont la longueur est inférieure à 9 pixels (respectivement 4 pour les cellules V.Harvey) sont

effacés. Enfin, les cellules détectées sont labelisées. L'image segmentée finale est donnée dans la figure C.13.

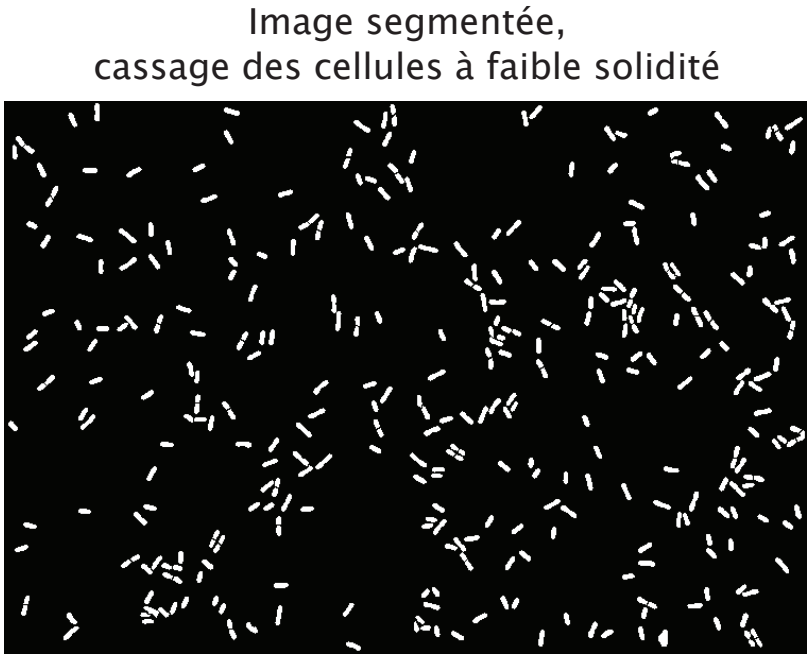


Figure C.12.: Image segmentée avec cassage des cellules à faible solidité.

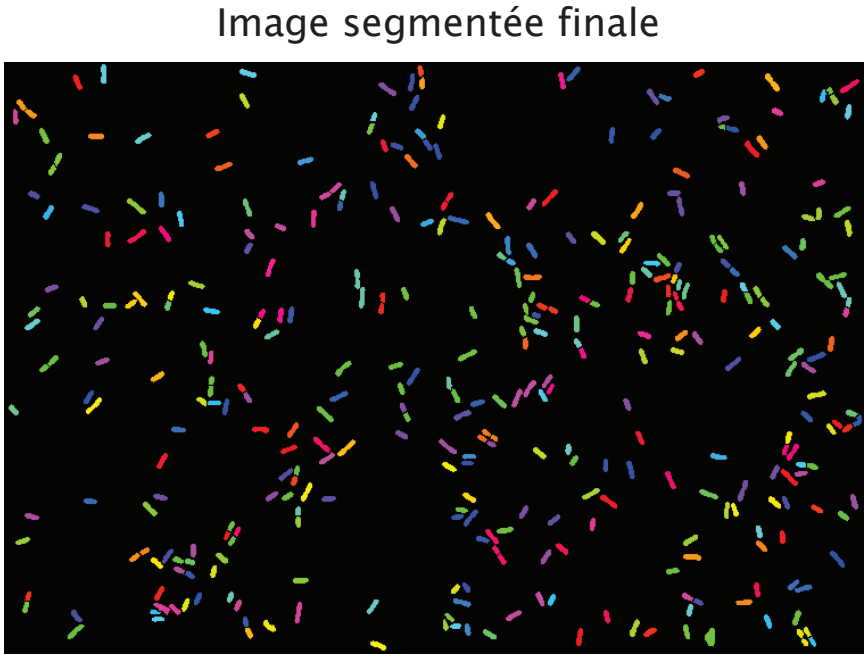


Figure C.13.: Image segmentée finale.

Bibliographie

[1] H.M.Mac Adams, A.Arkim. Stochastic mecanims in gene Expression. Proc Natl Acad Sci U S A. 1997 Feb 4 ;94(3) :814-9.

[2] A.Novick, M.Weiner. Enzyme induction as an all-or-none phenomenon. Proc Natl Acad Sci U S A. 1957 Jul 15 ;43(7) :553-66.

[3] M.B.Elowitz et S.Leibler. A synthetic Oscillatory network of transcriptional regulators. Nature. 2000 Jan 20 ;403(6767) :335-8.

[4] J.Paulsson. Models of stochastic gene expression. Phys. Life Rev. 2, 157-75 (2005).

[5] M.B.Elowitz, A.J.Levine, E.D.Siggia, P.S.Swain. Stochastic Gene expression in a Single Cell. Science 297 1183-1186,(2002).

[6] A.Hilfinger, J.Paulsson. Separating intrinsic from extrinsic fluctuations in dynamic biological systems. Proc Natl Acad Sci U S A. 2011 Jul 19 ;108(29) :12167-72. Epub 2011 Jul 5.

[7] V.Shahrezaei, P.S.Swain. Analytical distribution for stochastic gene expresssion. Proc Nat Acad Sci USA 105 (2008) 17256.

[8] M.Thattai, A.Oudenaarden. Intrinsic noise in gene regulatory network. Proc. Natl. Acad. Sci. U.S.A., Vol. 98, No. 15. (July 2001), pp. 8614-8619.

[9] A. Coulon, O.Gandrillon, G.Beslon. On the spontaneous stochastic dynamics of a single gene : complexity of the molecular interplay at the promoter. BMC Syst Biol. 2010 Jan 8 ;4 :2.

[10] N.Friedman, L.Cai, X.S.Xie. Linking Stochastic Dynamics to Population Distribution : An analytical Framework of Gene Expression. Phys Rev Lett. 2006 Oct 20 ;97(16) :168302. Epub 2006 Oct 19.

[11] V.Shahrezaei, J.F.Ollivier, P.S.Swain. Colored extrinsic fluctuation and stochastic gene expression. Molecular Systems. Biology, vol. 4, 2008.

[12] V.Shahrezaei, P.S.Swain. The stochastic nature of biochemical newtorks. Current Opinion in. Biotechnology, 19 :369–374, 2008.

[13] D.T.Gillespie. Exact Stochastic Simulation of Coupled Chemical Reactions. J. Phys. Chem., 81 (1977), pp. 2340–2361.

[14] I.Golding, J.Paulsson, S.M.Zawilski, E.C.Cox. Real-Time Kinetics of Gene Activity in Individual Bacteria. Cell. 2005 Dec 16 ;123(6) :1025-36.

[15] N.A.Cookson, S.W.Cookson, Lev.S.Tsimring, J.Hasty. Cell cycle-dependent variations in protein concentration. Nucleic Acids Res. 2010 May ;38(8) :2676-81. Epub 2009 Dec 17.

[16] C.Weiss-Schaber, I.Mihalcescu. Description dynamique de l'horloge circadienne des cyanobactéries sous entraînement par un modèle d'oscillateur de phase.Thèse Université de Grenoble 2010.

[17] P.Wang, L.Robert, J.Pelletier, W.L.Dang, F.Taddei, A.Wright, S.Jun. Robust growth of Eschericchia Coli. Current Biology 20, 1−5 (2010).

[18] L.Cai, N.Friedman, X.Sunney Xie. Stochastic protein expression in individual cells at the single molecule level. Nature. 2006 Mar 16 ;440(7082) :358-62.

[19] J.Yu, J.Xiao, X.Ren, K.Lao, X.S.Sunney. Probing Gene Expression in Live Celles, One Protein Molecule at a time. Science. 2006 Mar 17 ;311(5767) :1600-3.

[20] E.M.Ozbudak, M.Thattai, I.Kurster, A.D.Grossman, A.Oudenaarden. Regulation of noise in the expression of a single gene.bNature Genetics, Vol. 31 (May 2002), pp. 69-73.

[21] Y.Taniguchi, P.J.Choi, G.W.Li, H.Chen, M.Babu, J.Hearn, A.Emili, X.Sunney Xie. Quantifying *E. Coli* proteome and transcriptome with Single-Molecule Sensitivity in Single Cells. Science. 2010 Jul 30 ;329(5991) :533-8.

[22] Lok-hang So, A. Ghosh, C.Zong, L.A.Sepùlveda, R.Segev&I.Golding. General properties of transcriptional time series in *Escherichia coli*. Nat Genet. 2011 Jun ;43(6) :554-60. Epub 2011 May 1.

[23] P.S.Swain, M.B.Elowitz, E.D.Siggia. Intrinsic and Extrinsic contributions to stochasticity in gene expression. Proc Natl Acad Sci 99 12795-800 (2002).

[24] N.Rosenfeld, J.W.Young, U.Alon, PS.Swain, M.B.Elowitz. Gene Regulation at the Single-Cell level. Science 307(5717) : 1962-1965 (2005).

[25] D. Huh, J.Paulsson. Non-genetic heterogeneity from stochastic partitioning at cell division. Nat Genet. 2011 Feb ;43(2) :95-100. Epub 2010 Dec 26.

[26] Wong Ng J, Chatenay D, Robert J, Poirier MG. Plasmid copy number noise in monoclonal populations of bacteria. Phys Rev E Stat Nonlin Soft Matter Phys. 2010 Jan ;81(1 Pt 1) :011909. Epub 2010 Jan 14.

[27] B.Houchmandzadeh, I. Mihalcescu. Fluctuations importance and control in biological systems. Europhysicnews, vol 42/6 (2011).

[28] Süel GM, Kulkarni RP, Dworkin J, Garcia-Ojalvo J, Elowitz MB. Tunability and noise dependence in differentiation dynamics. Science. 2007 Mar 23 ;315(5819) :1716-9.
Maamar H, Raj A, Dubnau D. Noise in gene expression determines cell fate in Bacillus subtilis. Science. 2007 Jul 27 ;317(5837) :526-9. Epub 2007 Jun 14.EM.

[29] To TL, Maheshri N. Noise can induce bimodality in positive transcriptional feedback loops without bistability. Science. 2010 Feb 26 ;327(5969) :1142-5

[30] Ozbudak, M.Thattai, HN.Lim, BI.Shraiman, A.Van Oudenaarden. Multistability in the lactose utilization network of Escherichia coli. Nature. 2004 Feb 19 ;427(6976) :737-40.

[31] PJ.Choi, L.Cai, K.Frieda, XS.Xie. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. Science. 2008 Oct 17 ;322(5900) :442-6.

[32] E.Kussell, S.Leibler. Phenotypic diversity, population growth, and information in fluctuating environments. Science. 2005 Sep 23 ;309(5743) :2075-8. Epub 2005 Aug 25.
M.Acar, J.T Mettetal, A. van Oudenaarden. Stochastic switching as a survival strategy in fluctuating environments. Nature Genetics 40, 471 - 475 (2008) .
Blake WJ, Balázsi G, Kohanski MA, Isaacs FJ, Murphy KF, Kuang Y, Cantor CR, Walt DR, Collins JJ. Phenotypic consequences of promoter-mediated transcriptional noise. Mol Cell. 2006 Dec 28 ;24(6) :853-65.

[33] Thattai M, van Oudenaarden A. Stochastic gene expression in fluctuating environments. Genetics. 2004 May ;167(1) :523-30.

[34] Levin BR, Rozen DE. Non-inherited antibiotic resistance. Nat Rev Microbiol. 2006 Jul ;4(7) :556-62.

[35] Balaban NQ, Merrin J, Chait R, Kowalik L, Leibler S. Bacterial persistence as a phenotypic switch. Science. 2004 Sep 10 ;305(5690) :1622-5. Epub 2004 Aug 12.

[36] Levine JH, Fontes ME, Dworkin J, Elowitz MB. Pulsed feedback defers cellular differentiation. PLoS Biol. 2012 Jan ;10(1) :e1001252. Epub 2012 Jan 31.

[37] Raj A, van Oudenaarden A. Nature, nurture, or chance : stochastic gene expression and its consequences. Cell. 2008 Oct 17 ;135(2) :216-26.

[38] D.Ropers, H.de Jong, M.Page, D.Schneider, J.Geiselmann. Qualitative Simulation of the Carbon Starvation response in Escherichia Coli. Biosystems. 2006 May ;84(2) :124-52. Epub 2005 Dec 1.

[39] H.de Jong, C.Ranquet, D.Ropers, C.Pinel, and J.Geiselmann. Experimental and Computational validation of models of fluorescent and luminescent reporter genes in bacteria. BMC Syst Biol. 2010 Apr 29 ;4 :55.

[40] V.Baldazzi, D.Ropers, Y.Markowicz, D.Kahn, J.Geiselmann, H.de Jong. The Carbon Assimilation Network in *Echericchia Coli* is densely Connected and Largely Sign-Determined by Directions of Metabolic Fluxes. PLoS Comput Biol. 2010 Jun 10 ;6(6) :e1000812.

[41] H. de Jong, J. Geiselmann, C. Hernandez et M. Page (2003), Genetic Network Analyzer : Qualitative simulation of genetic regulatory networks, Bioinformatics, 19(3) :336-344.

[42] Baba T et al. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants : the Keio collection. Mol Syst Biol.2006 ;2 :2006.0008. Epub 2006 Feb 21.

[43] B.P.Cormack, R.H.Valdivia, S.Falkow. FACS-optimized mutants of the Green Fluorescent Protein (GFP).Gene. 1996 ;173(1 Spec No) :33-8.

[44] R.Iizuka, M.Yamagishi-Shirasaki, T.Funatsu. Kinetic study of de novo chromophore maturation of fluorescent proteins. Anal Biochem. 2011 Jul 15 ;414(2) :173-8. Epub 2011 Apr 1.

[45] J.B.Andersen, C.Sternberg, L.Kongsbak Poulsen, S.Petersen Bjørn, M.Givskov, S. Molin. New Unstable Variants of Green Fluorescent Protein for Studies of Transient Gene Expression in Bacteria. Appl Environ Microbiol. 1998 Jun ;64(6) :2240-6.

[46] J.M.Tarvaré, L.M. Fletcher, G.I.Welsh. Using green fluorescent protein to study intracellular signaling. The Journal of endocrinology (2001) Volume : 170, Issue : 2, Pages : 297-306.

[47] http ://www.clontech.com/

[48] Reid BG, Flynn GC. Chromophore formation in green fluorescent protein. Biochemistry. 1997 Jun 3 ;36(22) :6786-91.

T.D. Craggs. Green fluorescent protein : structure, folding and chromophore maturation. Chem. Soc. Rev., 2009, 38, 2865-2875.

[49] A.Miyawaki, O.Griesbeck, R.Heim, R.Y.Tsien . Dynamic and quantitative Ca2+ measurements using improved cameleons. Proc Natl Acad Sci U S A. 1999 Mar 2 ;96(5) :2135-40.
O.Griesbeck, S.Berg, R.E.Campbell, D.A.Zacharias, R.Y.Tsien. Reducing the environmental sensitivity of yellow fluorescent protein. J Biol Chem. 2001 Aug 3 ;276(31) :29188-94. Epub 2001 May 31.

[50] A.Miyawaki, R.Y.Tsien. Monitoring Protein Conformations and Interactions by Fluorescence Resonance Energy Transfert Between mutants of Green Fluorecent Protein. Methods Enzymol. 2000 ;327 :472-500.

[51] C.Ranquet et al, en cours de rédaction.

[52] Z.Zhao et al. Purification of green fluorescent protein using a two intein system. Appl Microbiol Biotechnol. 2008 Jan ;77(5) :1175-80. Epub 2007 Nov 1.

[53] G.Patterson, R.N.Day, and D.Piston. Fluorescent Protein Spectra. 2001 J Cell Sci 114, 837-838.

[54] http ://www.chroma.com/

[55] D.P.Doane. Aesthetic Frequency Classification. The American Statistician, vol. 30, pp. 181-183, 1976

[56] .F.J.Massey. The Kolmogorov-Smirnov test for goodness of fit. J. Am. Stat. Assoc. 46, 68 (1951).

[57] IT Young. Proof without prejudice : use of the Kolmogorov-Smirnov test for the analysis of histograms from flow systems and other sources. J Histochem Cytochem. 1977 Jul ;25(7) :935-41.

[58] W.H.Richardson. Bayesian-Based Iterative Method of Image Restoration. JOSA, Vol. 62, Issue 1, pp. 55-59 (1972) .

[59] L.B.Lucy. An iterative technique for the rectification of observed distribution. *T*he Astronomical Journal, Vol. 79 (June 1974), 745.

[60] C.A.Ball, R.Osuna, K.C.Ferguson, R.C.Johnson. Dramatic Changes in Fis upon Nutrient Upshift in *Escherichia coli*. J Bacteriol. 1992 Dec ;174(24) :8043-56.

[61] B.G.Reid, G.C.Flynn. Chromophore formation of Green Fluorescent Protein. Biochemistry (1997) *V*olume : 36, Issue : 22, Pages : 6786-6791.

[62] Groisman A, Lobo C, Cho H, Campbell JK, Dufour YS, Stevens AM, Levchenko A. A microfluidic chemostat for experiments with bacterial and yeast cells. Nat Methods. 2005 Sep ;2(9) :685-9.

[63] M.B. Miller et B.L Bassler. Quorum sensing in bacteria. Annu. Rev. Microbiol. 55, 265-199.

[64] I. Mihalcescu. Mémoire d'Habilitation à Diriger des Recherches. Novembre 2005.