



**HAL**  
open science

# Aspects temporel et spatial dans des systèmes de régulation génétique

Axel Cournac

► **To cite this version:**

Axel Cournac. Aspects temporel et spatial dans des systèmes de régulation génétique. Biophysique [physics.bio-ph]. Université Paris-Diderot - Paris VII, 2009. Français. NNT : 2009PA077255 . tel-00472602

**HAL Id: tel-00472602**

**<https://theses.hal.science/tel-00472602>**

Submitted on 12 Apr 2010

**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.



Distributed under a Creative Commons Attribution 4.0 International License

# THÈSE DE DOCTORAT DE L'UNIVERSITÉ PARIS VII

présentée par

**Axel Cournac**

pour obtenir le titre de Docteur de l'Université Paris VII

Spécialité : Interface Physique-Biologie

## **Aspects temporel et spatial dans des systèmes de régulation génétique**

Thèse soutenue le 16 décembre 2009 devant le jury composé de :

<b>Elisabeth Pécou</b> , <i>Laboratoire Jean-Alexandre Dieudonné</i>	Rapportrice
<b>Olivier Espéli</b> , <i>Centre de Génétique Moléculaire</i>	Rapporteur
<b>Jackie Plumbridge</b> , <i>Laboratoire de l'Expression Génétique Microbienne</i>	Examinatrice
<b>François Képès</b> , <i>Génopôle d'Evry</i>	Examineur
<b>Olivier Martin</b> , <i>Laboratoire de Physique Théorique et Modèles Statistiques</i>	Président du jury
<b>Jacques-Alexandre Sepulchre</b> , <i>Institut Non Linéaire de Nice</i>	Directeur de thèse

---

# Résumé

Cette thèse s'intéresse à plusieurs aspects concernant la régulation génétique sur les plans théorique et expérimental. Un premier travail théorique qui s'inscrit dans le cadre de la modélisation des systèmes biologiques se propose de trouver des réseaux de régulation simples qui puissent répondre de manière optimale lorsqu'ils sont soumis à une stimulation périodique. Le réseau appelé "Incoherent Feed Forward Loop" s'est avéré présenter la propriété intéressante de laisser passer des trains de pulses au profil temporel particulier. Des extensions de ce motif ("Diamond", "Double Diamond"...) ont été suggérées pouvant également présenter des propriétés intéressantes pour traiter des signaux plus complexes.

Un travail de revue et d'observations concernant les boucles d'ADN est ensuite présenté. Après avoir observé des points communs dans les systèmes de boucle d'ADN déjà connus, nous avons interrogé les bases de données pour savoir quelles étaient les régions de régulation présentant les mêmes caractéristiques. Nous proposons une liste de plusieurs opérons qui mériteraient une démarche expérimentale pour la mise en évidence d'une boucle d'ADN. Un travail expérimental de biologie moléculaire est ensuite présenté. Il s'est attaqué à tester une hypothèse présente dans plusieurs travaux de bioinformatique ou de physique statistique. La question testée est la suivante : est-ce que des facteurs de transcription pourraient se lier à des sites de liaison appartenant non pas à la même région de régulation d'un gène (comme dans le cas de l'opéron *lac*) mais à des sites de liaison appartenant à des gènes différents ? Nous avons testé cette hypothèse sur le système de régulation d'entrée en virulence de la bactérie *Erwinia chrysanthemi*. Les expériences reproduites et réalisées dans plusieurs conditions physiologiques différentes ont abouti à la conclusion que les régions de régulation supprimées ne semblent pas agir sur d'autres gènes par le mécanisme de boucle d'ADN. Le dernier chapitre propose une méthode expérimentale pour rechercher de telles interactions de façon large dans un génome de bactérie. La méthode est basée sur des techniques de biologie moléculaire couramment utilisées comme la mutagenèse aléatoire ou l' $\alpha$ -complémentation. Elle utilise le système enhanceur du promoteur de *glnA* d'*Escherichia coli* et vise à trouver des interactions 3D significatives entre segments d'ADN au sein d'un génome de procaryote.

**MOTS-CLÉS** : Réseaux génétiques - Modélisation- Stimulations périodiques - Boucle d'ADN - Organisation spatiale des génomes - Méthode expérimentale

# Abstract

This thesis deals with several aspects concerning genetic regulation on theoretical and experimental points of view. A first work in the field of modelisation of biological systems, proposes to find simple regulatory networks which have the ability to optimise their response when they are submitted to a periodic stimulation. The network called “Incoherent Feed Forward Loop” appeared to have the interesting property to let pass trains of pulses which have a particular temporal pattern. Some extensions of this motif (“Diamond”, “Double Diamond”...) are proposed to also have interesting properties to process more complex time-dependent signals.

Next, a work of review and observation concerning DNA looping is presented. After pointing out some common points in systems of DNA looping already known, we have examined the data bases to see if other regulatory regions have the same characteristics. We propose a list of several genes which might be good candidates to be regulated thanks to a DNA looping mechanism. An experimental work of molecular biology is then presented. It tackles a hypothesis found in several works from bioinformatics and statistical physics. The question is the following : can transcription factors bind to binding sites belonging not to a unique regulatory region of one gene (like in the *lac* operon) but to binding sites belonging to different genes ? We tested this hypothesis on the Pel regulatory system which rules the entry into virulence of the enterobacteria *Erwinia chrysanthemi*. Experiments reproduced and carried out in various physiological conditions lead to the conclusion that the deleted regulatory regions do not seem to act on other genes by a DNA looping mechanism. The last chapter proposes an experimental method to search for such interactions in a broader way inside a bacterial genome. The method is based on molecular biology techniques commonly used, like the random mutagenesis or  $\alpha$ -complementation. It uses the enhancer system of the promoter of *glnA* of *Escherichia coli* and aims at finding significant 3D interactions between DNA segments in a prokaryote genome.

**KEY-WORDS :** Genetic networks - Modelisation- Periodic stimulations - DNA looping - Spatial organisation of genomes - Experimental method

# Remerciements

J'adresse de vifs remerciements à Jacques-Alexandre Sepulchre pour avoir encadré le travail de cette thèse. Je lui suis particulièrement reconnaissant de m'avoir apporté avec bienveillance conseils, idées et fait part de son expérience quel que soit le chemin de recherche où nous nous engageons. Je remercie vivement Monsieur William Nasser du laboratoire Microbiologie, Adaptation et Pathogénie pour entre autre sa très grande générosité et sa passion pour le métier. Je remercie la biologiste Sylvie Reverchon pour notamment sa perspicacité et avec qui j'ai également eu la chance d'apprendre quelques unes des techniques de biologie moléculaire. Je remercie le biologiste François Képès de l'ATelier de Génomique Cognitive pour l'effet stimulant qu'a eu son travail sur le déroulement de cette thèse et pour avoir pris part au jury. Je remercie Elisabeth Pécou et Olivier Espéli pour avoir accepté la charge d'être rapporteurs de ce travail. Je remercie également Jackie Plumbridge et Olivier Martin pour avoir pris part au jury en tant qu'examinateurs.

Je remercie Yves Charon pour ses recommandations données avec bienveillance et dynamisme ainsi que le nouveau directeur de l'école doctorale Philippe Schwemling. Je suis reconnaissant à Roland Mastrippolito pour m'avoir donné la chance de poursuivre ma formation au sein du très intéressant M2 Physique des Systèmes Biologiques. Je remercie l'équipe enseignante du département de Physique de l'université Evry Val d'Essone notamment Véronica Beswick. Je suis reconnaissant à Samuel Bottani pour m'avoir suggéré l'apprentissage du langage informatique Perl, outil qui s'est avéré utile tout le long de ma thèse. Je remercie le chercheur et ami Laurent Le Guillou pour des discussions encourageantes au début de ma thèse. Je remercie Jean-Marc Di Meglio notamment pour son cours sur la matière molle dont je garde un merveilleux souvenir. Je salue toutes les personnes que j'ai eu la chance de rencontrer au MSC : Bérengère, Cyril, Delphine, François, Marion, Marc, Nicolas, Olivier, Pascal, Vivien. Je salue chaleureusement les étudiants et chercheurs rencontrés à l'Institut Non-Linéaire de Nice : Emilie, Florence, Fred, Freddy, 英俊, Hugo, Karim, Lauris, Marianne, Marta, Maryvonne, Pat, Stéphane et son directeur Jorge Tredicce pour m'avoir accueilli. Je remercie toutes les personnes que j'ai eu la joie de rencontrer au 4ème étage du bâtiment André Lwoff au sein de l'unité Microbiologie, Adaptation et Pathogénie : Arnaud, Camille, Camille, Claire, Denis, Julien, Géraldine, Honoré, 楠, Mathilde, وفاء, سناء, سوزان, 小慧, Yann, les chercheurs, Guy, Vladimir, le personnel technique Véronique, Yvette, Jean-Michel et sa directrice Nicole Cotte-Pattat pour m'avoir accueilli. Si vous tombez par hasard sur ces lignes : *Bonne chance pour la suite de la route, j'espère que l'on se reverra..* Je remercie mes amis : des potes du lycée à l'université d'Orsay en passant par la fac de Toulouse Paul Sabatier. Enfin, je remercie ma famille : ma mère Annick, mon père Maurice, ma sœur Anna et mon frère Jean-Marie pour leur amour et leur aide.



# Table des matières

Résumé . . . . .	iii
Abstract . . . . .	iv
Remerciements . . . . .	v
<b>Table des matières</b>	<b>1</b>
<b>1 Introduction générale</b>	<b>5</b>
1 L'information issue de la séquence . . . . .	5
2 L'organisation du vivant en réseaux de régulation . . . . .	6
3 Concepts importants concernant les réseaux de régulation biologique .	7
3.1 Conditions sur la topologie des réseaux de régulation . . . . .	7
3.2 Bruit dans les réseaux génétiques . . . . .	8
3.3 Relations topologie-fonction biologique . . . . .	9
4 Trois exemples de confrontation modélisation / expériences . . . . .	9
4.1 Meilleure stabilité par une boucle rétronégative . . . . .	9
4.2 Bistabilité pour une boucle rétropositive . . . . .	12
4.3 Oscillateur génétique . . . . .	15
5 Conclusion . . . . .	19
<b>2 Stimulations périodiques et régulation génétique</b>	<b>21</b>
1 Motivation et travaux antérieurs . . . . .	21
1.1 Travaux expérimentaux . . . . .	22
1.2 Travaux théoriques . . . . .	22
2 Stimulation périodique d'une simple activation . . . . .	23
3 Stimulation périodique d'une "Incoherent Feed Forward Loop" . . . . .	27
3.1 L'Incoherent Feedforward Loop . . . . .	27
3.2 Stimulation périodique par un signal créneau . . . . .	29
3.3 Étude plus détaillée des deux situations . . . . .	29
3.4 Réponse de l'Incoherent Feedforward Loop, modélisation avec des fonctions de Hill . . . . .	32
4 Motif de type "Diamond IFFL" . . . . .	35
5 Réponse optimale d'un système soumis à un signal type bursting . . .	38
5.1 Exemples d'application . . . . .	41
6 Résonance simple . . . . .	42
6.1 Motif de boucle rétro-négative simple . . . . .	43
6.2 Exemples d'applications possibles . . . . .	45
7 Perspectives de ce travail sur le plan théorique . . . . .	46
8 Conclusion . . . . .	47



<b>3</b>	<b>Boucles d'ADN et régulation génétique</b>	<b>63</b>
1	Cas connus et bien caractérisés de boucle d'ADN . . . . .	63
1.1	Boucle d'ADN à l'origine d'une activation transcriptionnelle .	64
1.2	Boucles d'ADN à l'origine d'une répression . . . . .	64
2	Techniques expérimentales pour mettre en évidence des boucles d'ADN	69
3	Avantages procurés par les boucles d'ADN . . . . .	73
4	Recherche de nouvelles boucles d'ADN . . . . .	77
4.1	Points communs entre les systèmes connus présentant des boucles d'ADN à l'origine d'une répression transcriptionnelle . . . . .	78
4.2	Interrogation de la base de donnée <i>RegulonDB</i> . . . . .	81
4.3	Corroboration bibliographique . . . . .	84
5	Liste de candidats obtenus . . . . .	84
5.1	Résultat des filtrages . . . . .	84
5.2	Cas d'intérêt . . . . .	84
6	Conclusion . . . . .	88
<b>4</b>	<b>Recherche d'interactions 3D sur le système Pel de la bactérie <i>Erwinia chrysanthemi</i></b>	<b>89</b>
1	Modèles d'organisation spatiale des génomes . . . . .	89
1.1	Organisation solénoïdale du génome . . . . .	90
1.2	Travaux bioinformatiques suggérant des interactions à dis- tance ou des corrélations à longue portée dans des génomes . .	91
1.3	Travaux issus de la physique statistique suggérant des organi- sations par boucle d'ADN . . . . .	92
1.4	Travaux de biologie expérimentale . . . . .	92
2	Choix du système d'étude . . . . .	93
2.1	Raison génétique . . . . .	93
2.2	Raison thermodynamique . . . . .	94
2.3	Raison physiologique . . . . .	95
2.4	Raison pratique . . . . .	96
3	Expériences réalisées sur la bactérie <i>Erwinia chrysanthemi</i> . . . . .	96
3.1	Idée générale des expériences réalisées . . . . .	96
3.2	Comparaison de deux mutants . . . . .	98
4	Matériels et méthodes . . . . .	99
4.1	Techniques d'ADN recombinant . . . . .	99
4.2	Souches bactériennes . . . . .	100
4.3	Milieus et conditions de culture . . . . .	100
4.4	Procédure suivie lors de cinétiques . . . . .	100
4.5	Construction du mutant "Inactivation" . . . . .	101
4.6	Electrofocalisation et vérification du phénotype du mutant "Inactivation" . . . . .	102
5	Résultats des expériences . . . . .	103
5.1	Comparaison mutant "Délétion" et mutant "Inactivation" dans milieu M63 + PGA + glucose . . . . .	103
5.2	Comparaison mutant "Délétion" et mutant "Inactivation" dans milieu M63 + glycérol avec ou sans PGA . . . . .	105
5.3	Effet de la délétion dans 3 mutants d' <i>Erwinia chrysanthemi</i> .	106
6	Conclusions . . . . .	107

<b>5</b>	<b>Méthode pour rechercher des interactions 3D dans un génome de bactérie</b>	<b>111</b>
1	Méthode pour tenter de trouver des interactions 3D . . . . .	111
1.1	Présentation du principe de la méthode . . . . .	111
1.2	Différentes étapes envisagées . . . . .	113
1.3	Points importants et difficultés anticipées de la méthode . . .	115
1.4	Comparaison avec d'autres méthodes déjà existantes . . . . .	117
1.5	Expériences ultérieurement envisageables . . . . .	118
2	Hypothèses pour le rapprochement 3D de gènes . . . . .	118
2.1	Conclusion . . . . .	119
	<b>Conclusions et Discussion</b>	<b>121</b>
	<b>Bibliographie</b>	<b>127</b>



# Chapitre 1

## Introduction générale

Nous donnons dans ce chapitre une introduction générale sur les processus de régulation génétique. Nous commençons par insister sur le fait que toute l'information d'un organisme n'est pas seulement ou formellement contenue dans la séquence de son génome mais également dans de nombreux processus "autour" d'elle. Par exemple, la dynamique des réseaux génétiques peut être très riche en comportements et propriétés qui ont un impact sur l'expression de l'information génétique. Nous essayons de mettre en relief des exemples de travaux qui, inspirés par des concepts issus de la physique théorique, ont donné des perspectives nouvelles ou des applications expérimentales dans l'étude de certains processus biologiques.

### 1 L'information issue de la séquence

Les séquences du génome d'un chimpanzé et de celui d'un être humain sont identiques à environ 95 % [155]. Ce chiffre frappant indique que toute l'information qui définit un organisme vivant n'est pas seulement ou directement contenue dans la séquence de son génome. Le nombre de gènes n'est pas non plus un indicateur direct de la complexité d'un organisme [31] : les génomes de nombreux végétaux possèdent plus de gènes que celui de l'homme, comme dans le cas du peuplier qui en possède presque deux fois plus [168] (ceci peut toutefois s'expliquer par la redondance de l'information car certaines parties ou parfois l'intégralité de génomes de végétaux se sont dupliquées). Il semble difficile d'établir une relation simple entre la quantité d'information contenue dans l'enchaînement des paires de bases et la complexité de certains processus biologiques ou phénotypes qui en résultent. Les SNP (pour "Single Nucleotide Polymorphisms") sont également un élément qui montre combien l'information contenue dans la séquence des nucléotides n'est pas une mesure directe de la complexité d'un phénotype. En effet, il suffit d'un changement d'une seule paire de base au sein d'un gène ou d'une région intergénique pour expliquer de très grandes différences dans les phénotypes au sein d'une même espèce, comme l'apparition de maladies.

Il y a donc comme de l'information manquante qu'il faut chercher dans tous les processus qui existent "autour" de la séquence, et qui sont chargés d'utiliser et d'exprimer l'information génétique [85]. Nous pouvons supposer que plus l'organisme est complexe, plus l'information contenue dans les processus autour de la séquence est importante. Les très grandes différences qui existent entre les procaryotes et

les eucaryotes pourraient en partie provenir du fait que ces derniers font intervenir des processus beaucoup plus riches et complexes pour traiter l'information de la séquence que les organismes procaryotes. Un axe de recherche, nommé l'*épigénétique* regroupe tous les changements héréditaires dans la fonction des gènes, ayant lieu sans altération de la séquence des nucléotides de l'ADN. Chez certains organismes, il y a par exemple les modifications sur les protéines qui enroulent l'ADN qui ont un impact sur l'expression génétique : la méthylation des histones (protéines associées à l'ADN génomique eucaryote) au niveau des résidus de lysines entraîne une fermeture de la chromatine. Au contraire, l'acétylation des lysines entraîne une ouverture de la chromatine permettant ainsi une transcription plus forte. D'autres modifications chimiques sur l'ADN peuvent influencer sur l'expression génétique : par exemple, une forte méthylation de cytosine en 5-methylcytosine dans les dimères C-G de l'ADN diminue l'expression d'un gène. Ces quelques processus n'expliquent cependant pas à eux seuls la grande complexité qui peut apparaître dans certains processus biologiques.

Il apparaît donc que la façon avec laquelle est exprimée l'information contenue dans la séquence est cruciale. D'autres mécanismes porteurs d'information sont donc à chercher dans la façon dont l'information est traitée notamment au niveau de la régulation transcriptionnelle et des interactions entre gènes et protéines. Nous présentons dans les sections suivantes un domaine de recherche qui permettrait d'apporter de la lumière sur ces questions en recherchant comment les gènes et les protéines interagissent au sein de réseaux de régulation. La dynamique de ces réseaux d'interaction peut en effet être très riche et constituer des éléments importants pour le monde vivant.

## 2 L'organisation du vivant en réseaux de régulation

Depuis une dizaine d'années, un nouveau champ de recherche semble émerger, qui est souvent appelé *Biologie Systémique*. Ce domaine de recherche tente à l'aide d'outils issus notamment de la physique théorique (physique non linéaire, physique statistique), des mathématiques (approche logique, théorie des jeux, processus de Markov) ou de l'informatique (algorithmes génétiques) de mieux comprendre certains processus du monde vivant notamment ceux impliqués dans les réseaux de régulation. Il est connu depuis les travaux de Jacob et Monod, que des protéines peuvent réguler l'expression d'autres gènes au niveau de la transcription. Il est apparu que très souvent, plusieurs gènes et protéines sont impliqués dans des interactions interconnectées que l'on peut représenter sous forme de réseaux. Chaque nœud correspond par exemple à un gène ou à une protéine et chaque branche correspond à une interaction : activation transcriptionnelle, inactivation d'une protéine par une autre, etc (voir Figures 1.2 ou 1.4 plus bas qui donnent des exemples de représentation de réseaux génétiques). Les outils mathématiques utilisés dans certaines branches de la physique théorique comme la théorie des systèmes dynamiques se sont révélés tout à fait appropriés pour étudier le comportement de plusieurs gènes et de leurs protéines impliqués dans un réseau de régulation. La concentration d'une protéine est décrite comme une variable dynamique dépendante du temps, chaque interaction pouvant être formulée à l'aide de fonctions mathématiques. Les propriétés du système obtenu peuvent alors s'étudier à l'aide d'outils analytiques et de la

résolution numérique. Nous fournissons plus de détails sur cette approche dans les sections suivantes.

## 3 Concepts importants concernant les réseaux de régulation biologique

De nombreux travaux théoriques et de modélisation se sont proposés de mieux comprendre les réseaux de régulation biologiques et de fournir des principes généraux concernant leur structure. Nous réunissons dans cette section des concepts qui ont été établis ces dernières années et qui nous semblent importants.

### 3.1 Conditions sur la topologie des réseaux de régulation

Les premiers travaux sur les réseaux génétiques se sont intéressés aux liens entre la structure du réseau et sa dynamique. En effet, pour qu'un réseau de régulation adopte une dynamique bien particulière (par exemple des oscillations temporelles dans la concentration des protéines), il faut des conditions précises sur la topologie du réseau c'est-à-dire sur l'agencement des interactions entre les gènes et protéines. Parmi ces études, deux types de dynamique ont particulièrement été étudiées : la première correspondant à la propriété de bistabilité et la seconde correspondant à l'apparition d'oscillations.

Pour qu'un réseau présente la propriété de bistabilité c'est-à-dire pour que le système puisse se comporter comme un interrupteur (sa réponse est de type "tout ou rien" suivant la valeur d'un stimulus), une condition nécessaire sur la topologie du réseau est qu'il doit contenir dans son graphe d'interactions au moins une boucle de rétroaction positive. Par exemple, une protéine A active une protéine B qui active A en retour. Il faut également des non-linéarités fortes dans les régulations qui entrent en jeu. Il est intéressant de constater que ce principe était intuitivement appréhendé par les biologistes (qui parlaient plutôt de réactions autocatalytiques [118]), a été conjecturé au début des années 80 par le théoricien René Thomas [166], puis a reçu une preuve mathématique en 2003 [161].

Beaucoup de travaux ont étudié les conditions nécessaires pour avoir des oscillations dans la concentration des protéines mises en jeu [119]. Ici, la condition nécessaire mais non suffisante est d'avoir une boucle de rétroaction négative dans le graphe d'interaction du réseau [166]. Il n'y a pas de critère suffisant concernant la topologie du réseau mais de fortes non-linéarités [53] dans les régulations donnent plus de chances pour que le réseau ait des oscillations. Un terme de délai [90, 100] permet également d'obtenir plus facilement des oscillations auto-entretenuës. Par exemple, si l'on tient compte de la durée nécessaire pour la synthèse d'une protéine et de son transport jusqu'à l'ADN et que ce terme est introduit dans les équations, il peut générer dans plusieurs modèles de régulation l'instabilité qui conduit aux oscillations.

Nous détaillons dans la section 4 des exemples pour illustrer et expliciter ces deux principes généraux.

### 3.2 Bruit dans les réseaux génétiques

Les réseaux génétiques font souvent intervenir un petit nombre de molécules interagissantes. En effet, une bactérie synthétise un nombre restreint de facteurs de transcription (par exemple de l'ordre de la dizaine de tétramères dans le cas du répresseur LacI de l'opéron *lac*) car leur synthèse présente un coût énergétique important. Ces petits nombres de molécules entraînent une composante aléatoire dans les processus de régulation qui peut avoir des conséquences sur l'expression des gènes ou même sur la survie d'une bactérie. De nombreux travaux se sont donc proposés de comprendre comment la nature pouvait atténuer ces fluctuations afin de rendre les processus de régulations plus sûrs. Nous pouvons donner l'exemple de l'auto-régulation négative qui permet une meilleure stabilité et résistance au bruit. Ce concept proposé théoriquement par MA Savageau en 1974 [149] a été montré de façon expérimentale par le travail de Becskei et Serrano en 2000 [16]. Nous le présentons comme exemple illustratif dans la section suivante. De façon analogue, Leibler et Barkai ont montré que pour avoir des oscillations robustes au bruit, un réseau construit avec une boucle rétro-négative et une boucle rétro-positive générant une situation d'hysteresis permettait la génération d'oscillations dont la période et l'amplitude présentaient une moins grande variabilité par rapport à un réseau composé d'une seule boucle négative [15]. De façon intéressante, nous pouvons donc retenir que la topologie d'un réseau peut avoir des conséquences sur d'autres aspects que sur sa dynamique propre, comme ici sur sa résistance aux fluctuations.

Une question intéressante est de savoir si cet aspect aléatoire des réactions biochimiques pourrait être à l'origine de dynamiques particulières. Il est alors important de se demander dans quels cas une modélisation déterministe est suffisante et dans quels cas les effets stochastiques ne pourront être omis. Cette question a par exemple été traitée par l'équipe de A. Golbeter dans [52]. Ces auteurs ont montré qu'ils pouvaient retrouver les mêmes dynamiques avec les deux approches, déterministes et stochastiques, pour un nombre assez restreint de molécules : ils observent qu'il y a encore concordance quand les ARN sont de l'ordre de plusieurs dizaines et les protéines de l'ordre de plusieurs centaines d'unités. Cependant, on peut remarquer que le bruit peut parfois changer la dynamique d'un système dans certains cas, notamment lorsque le système se trouve près d'une *bifurcation*. Par exemple, Vilar et al. ont montré dans [172] que pour un même jeu d'équations et de paramètres, la modélisation purement déterministe présentait un état stable alors que la modélisation stochastique montrait le système oscillant. Ceci pouvait s'expliquer par le fait que le système était proche d'une bifurcation de Hopf (c'est-à-dire à l'endroit où en changeant la valeur d'un paramètre, le système passe d'une dynamique stable à une dynamique oscillante<sup>1</sup>) et que les perturbations issues du bruit entraînaient le système dans un pseudo cycle limite. Dans ce cas particulier et dans quelques autres [75, 91], le bruit peut donc être à l'origine d'une dynamique particulière.

Par ailleurs, ce bruit peut aussi être exploité de façon positive par un organisme vivant. Il peut en tirer une plus grande variabilité d'expression génétique et donc de nombreux phénotypes différents dont l'un d'eux pourra peut-être permettre une meilleure survie [82].

---

<sup>1</sup>Mathématiquement, une bifurcation de Hopf correspond au changement d'état lorsque la matrice jacobienne du système possède une paire de valeurs propres complexes conjuguées imaginaires.

### 3.3 Relations topologie-fonction biologique

Pour réaliser une même dynamique (oscillations, interrupteur...), plusieurs réseaux aux topologies différentes sont en fait possibles. Par exemple, le travail de François et al. montre à l'aide d'algorithmes génétiques<sup>2</sup> qu'en tenant compte des interactions protéines régulatrices - gènes mais également des interactions entre protéines (dimérisation, inactivation...), plusieurs réseaux aux topologies différentes sont possibles pour construire par exemple un oscillateur [41, 42].

Cependant, si dans des modèles mathématiques la relation topologie-dynamique, et donc la relation topologie-fonction du réseau, est claire, il se pourrait que la correspondance ne soit pas aussi directe dans le monde vivant. En effet, si telle topologie de réseau confère telle fonction alors la pression de sélection devrait imposer que certaines topologies de réseaux sont préférentiellement gardées. Or certains travaux tendent à montrer que ce n'est pas toujours le cas [107]. Le travail de Mazurie et al. montre en effet que les topologies particulières de réseaux ne semblent pas se conserver entre différentes espèces de levures. La plupart de ces réseaux de régulation sont en fait reliés à une myriade d'autres interactions si bien que la notion de réseau simple agissant indépendamment et pour une fonction précise n'est peut-être qu'une vue simplifiée de la situation réelle. Toutefois, des approches synthétiques dans lesquelles de petits réseaux sont construits sur plasmides ont montré qu'ils pouvaient bien se comporter avec une fonction déterminée (oscillations [38] ou interrupteur [43]) prouvant une certaine réalité du concept de module fonctionnel associé à un réseau.

## 4 Trois exemples de confrontation modélisation / expériences

Nous donnons dans cette section trois exemples significatifs du domaine de la biologie systémique. Nous essayons de mettre en relief la confrontation théorie / expérimentation dans chacun des cas.

### 4.1 Meilleure stabilité par une boucle rétronégative

Dans un article de 1974 portant sur l'autorégulation, M.A Savageau compare les processus d'autoactivation et d'autorépression à des systèmes simplement régulés [149]. Il prédit que l'autorégulation négative pourrait conférer un avantage sur la stabilité de la régulation. Plus précisément, une boucle rétronégative conférerait une meilleure stabilité face aux fluctuations dans l'expression génétique dues au bruit inhérent à ces réactions biochimiques. Son travail a ensuite inspiré de nombreuses

---

<sup>2</sup>Un algorithme génétique reprend le principe de la sélection naturelle de Darwin : pour un problème donné, l'algorithme part de plusieurs mécanismes pouvant être des solutions au problème puis de manière aléatoire crée des changements (mutations) sur ceux-ci. Il ne garde que les mécanismes qui se rapprochent de la solution au problème et éliminent les autres. Puis il recommence le processus jusqu'à obtenir des mécanismes qui répondent au problème donné de manière optimale. Dans le travail cité, les auteurs partent de réseaux de régulation puis de manière aléatoire ajoutent ou suppriment des gènes ou changent des interactions. Au fur et à mesure du temps, ils ne gardent que les réseaux qui optimisent un score c'est-à-dire qui peuvent réaliser une fonction donnée au départ.



études théoriques ou de modélisation. Uri Alon, dans son livre “An Introduction to Systems Biology” [5] donne également une démonstration simple sur le fait qu’une autorégulation négative permet une meilleure robustesse face aux fluctuations. Une boucle rétronégative peut également produire une réponse plus rapide qu’une simple régulation (le temps de réponse qui est le temps pour que le système atteigne la moitié de son état stationnaire est plus court dans une boucle rétronégative que dans une simple régulation). Ceci avait également été prédit dans le travail de M.A. Savageau et a été démontré expérimentalement par l’équipe de Uri Alon [143]. Nous détaillons dans cette section un modèle de boucle rétronégative et la confrontation expérimentale concernant la stabilité face aux fluctuations.

### **Théorie et modèle**

Nous reprenons la démonstration théorique de Uri Alon [5] qui nous paraît intuitive et simple pour justifier la propriété de meilleure robustesse face aux fluctuations d’une boucle rétronégative par rapport à une simple régulation. Ceci nous permet d’introduire et de présenter le formalisme mathématique que nous utiliserons pour la suite de notre travail (Chapitre 2).

Pour une simple régulation d’un gène  $X$ , l’équation pour l’évolution temporelle de  $X$  s’écrit :

$$\frac{dX}{dt} = \beta - \alpha X \quad (1.1)$$

$\beta$  représente le taux de production de la protéine  $X$  et  $\alpha$  son taux de dégradation. L’état stationnaire du système c’est à dire l’état atteint où la concentration ne varie plus en fonction du temps est la valeur  $X_{st}$  pour laquelle l’équation 1.1 égale 0, nous avons :

$$X_{st} = \frac{\beta}{\alpha} \quad (1.2)$$

Maintenant, calculons l’état stationnaire pour un système autorégulé négativement. L’équation pour l’évolution temporelle de  $X$  s’écrit :

$$\frac{dX}{dt} = f(X) - \alpha X \quad (1.3)$$

$f(X)$  est une fonction qui rend compte de la répression. Il s’agit donc d’une fonction décroissante de  $X$ . Une bonne approximation pour beaucoup de promoteurs est une fonction de Hill décroissante :

$$f(X) = \frac{\beta}{1 + (X/\theta)^n} \quad (1.4)$$

L’expression de ces fonctions peut être justifiée et retrouvée en partant des lois d’action de masse des différentes réactions biochimiques (réaction de liaison de la protéine régulatrice sur l’ADN, synthèse de la protéine, etc) et en faisant l’hypothèse que la liaison sur l’ADN de la protéine régulatrice est rapide devant les autres réactions et peut donc être considérée à l’équilibre [112]. Le paramètre  $\theta$  est un paramètre de seuil, il représente la quantité de protéines régulatrices nécessaire pour que la régulation se fasse de moitié. Le coefficient de Hill  $n$  représente la coopérativité

réelle ou effective entre les protéines régulatrices. Par exemple, lorsque la régulation se fait par l'action de dimères, ce coefficient aura une valeur de deux<sup>3</sup>.

Pour plus de simplicité dans les calculs, il est possible d'utiliser l'*approximation logique* qui consiste à remplacer les fonctions de Hill par des fonctions Heaviside c'est à dire  $H(X) = 1$  si  $X > 0$  et  $H(X) = 0$  si  $X < 0$ . La régulation génétique est alors vue comme un processus de tout ou rien : si le facteur de transcription dépasse la valeur seuil alors le gène est régulé, dans le cas contraire, le gène n'est pas régulé. Dans ce cas, le système non-linéaire n'est pas approximé par un jeu d'équations linéaires mais par un jeu d'équations linéaires par morceaux [55]. Cette approche est en effet utilisée depuis plusieurs années dans la modélisation des réseaux génétiques et s'est avérée fructueuse [47]. Nous allons approximer l'équation 1.3 par l'équation suivante :

$$\frac{dX}{dt} = \beta H(\theta - X) - \alpha X \quad (1.5)$$

Ce qui équivaut à :

$$\frac{dX}{dt} = \beta - \alpha X \text{ si } X < \theta \quad (1.6)$$

$$\frac{dX}{dt} = -\alpha X \text{ si } X > \theta \quad (1.7)$$

Nous supposons que  $\theta < \beta/\alpha$ . Dans ce cas, si  $X < \theta$ , alors  $\frac{dX}{dt} > 0$  et  $X$  croît. Dès que  $X$  dépasse la valeur  $\theta$ ,  $\frac{dX}{dt} < 0$  et sa synthèse est arrêtée, sa concentration diminue. Il y a donc des oscillations transitoires et le système s'immobilise sur la valeur de  $\theta$ . Donc, nous avons comme état stationnaire pour ce système :

$$X_{st} = \theta \quad (1.8)$$

Nous pouvons alors comparer les expressions 1.2 et 1.8. L'état stationnaire d'un système simplement régulé dépend de  $\beta$  et de  $\alpha$ . Un changement sur l'un de ces deux paramètres aura un effet proportionnel sur l'état stationnaire. En revanche, pour le système auto-régulé négativement, l'état stationnaire ne dépend que de  $\theta$ . Ce dernier paramètre est lié à la constante de dissociation des protéines régulatrices sur l'ADN, il varie peu d'une cellule à l'autre contrairement aux paramètres  $\beta$  et  $\alpha$  qui sont beaucoup plus touchés par les fluctuations. Ainsi, le calcul présenté aboutit à la conclusion que l'autorégulation négative permet une meilleure robustesse aux fluctuations pour l'état stationnaire de la concentration des protéines.

## Expérimentation

Le travail expérimental de Becskei et Serrano [16] s'est attaqué sur le plan expérimental à vérifier la proposition précédente. Leur travail a consisté à construire un système autorégulé négativement et deux autres systèmes faisant intervenir les mêmes caractéristiques mais où l'autorégulation n'est pas possible. La répression de l'autorégulation négative se fait avec le répresseur fort TetR. Le taux d'expression

---

<sup>3</sup>Il est à noter que dans beaucoup de travaux qui se sont proposés de trouver une valeur pour le coefficient de Hill à partir de données expérimentales en fittant des courbes de régulation par exemple, celui-ci a très souvent une valeur non entière comme dans les travaux de [123] ou [43].

des protéines est mesuré pour une cellule unique grâce à un gène codant pour une GFP fusionné au gène *tetR*. La première construction correspond au système avec l'autorégulation négative. Les deux autres systèmes sont des systèmes faisant intervenir les mêmes ingrédients mais où l'autorégulation négative est impossible car le répresseur a subi une mutation et ne peut plus se lier sur l'ADN pour le deuxième et parce que les sites opérateurs sur l'ADN ont été changés pour le troisième (voir Figure 1.1).

Les distributions de la fluorescence et donc de la concentration en protéines sont montrées. Il apparaît clairement que la distribution pour la construction faisant intervenir l'autorégulation négative est plus resserrée que les distributions des deux autres constructions. L'autorégulation négative est bien un moyen pour lutter contre les fluctuations et permet d'avoir une population plus homogène. Nous mentionnerons dans le Chapitre 3 un autre mécanisme qui pourrait assurer une meilleure robustesse au bruit dans l'expression génétique, mécanisme qui a été prédit sur le plan théorique mais qui n'a pas encore été vérifié expérimentalement.

## 4.2 Bistabilité pour une boucle rétropositive

Il a été prédit également par des travaux théoriques que la propriété de bistabilité pouvait apparaître pour des réseaux qui possèdent une boucle de rétroaction positive. Ceci a donné lieu à de nombreux travaux de modélisation qui ont recherché en détail les conditions pour avoir la propriété de bistabilité [27]. Ceci a également eu des retombées sur le plan expérimental que ce soit sur des constructions synthétiques [43] ou sur des réseaux génétiques naturels [178]. En effet, la propriété de bistabilité est très liée à la fonction d'interrupteur et à celle de "mémoire génétique". Pour définir et illustrer cette propriété sur les plans théorique et expérimental, nous donnons comme exemple le travail réalisé par B. Kramer et M. Fussenegger [80] réalisé en 2005.

### Théorie et modèle

Le modèle est basé sur le même formalisme que le travail cité dans la section précédente : l'évolution temporelle de chaque protéine est décrite par une équation différentielle non linéaire. Le réseau d'interactions étudié analytiquement et construit expérimentalement fait intervenir une boucle de rétroaction positive exercée par la protéine TA (voir Figure 1.2). Le réseau peut être induit en faisant varier la concentration de l'antibiotique (EM) qui inactive le répresseur TR lequel agit sur le promoteur de TA. L'évolution temporelle de la concentration de TA peut s'écrire avec l'équation suivante :

$$\frac{dT A}{dt} = \beta \frac{T A^2}{1 + T A^2} \left(1 - \frac{T A}{2.5}\right) \frac{1}{1 + T R^2} - \alpha T A \quad (1.9)$$

$\beta$  représente le taux d'expression du promoteur quand celui-ci est pleinement activé par TA.  $TR$  est la concentration en répresseurs  $TR$  actifs (non inactivés par EM). Le terme  $\frac{T A^2}{1 + T A^2}$  représente l'autoactivation de TA. Comme le maximum de la concentration intracellulaire ne peut pas physiologiquement dépasser une certaine quantité, le terme  $\left(1 - \frac{T A}{2.5}\right)$  est donc introduit, ce qui limite TA à 2.5 en unités arbitraires. Le terme  $\frac{1}{1 + T R^2}$  rend compte de la répression de TR sur le promoteur de TA. Le

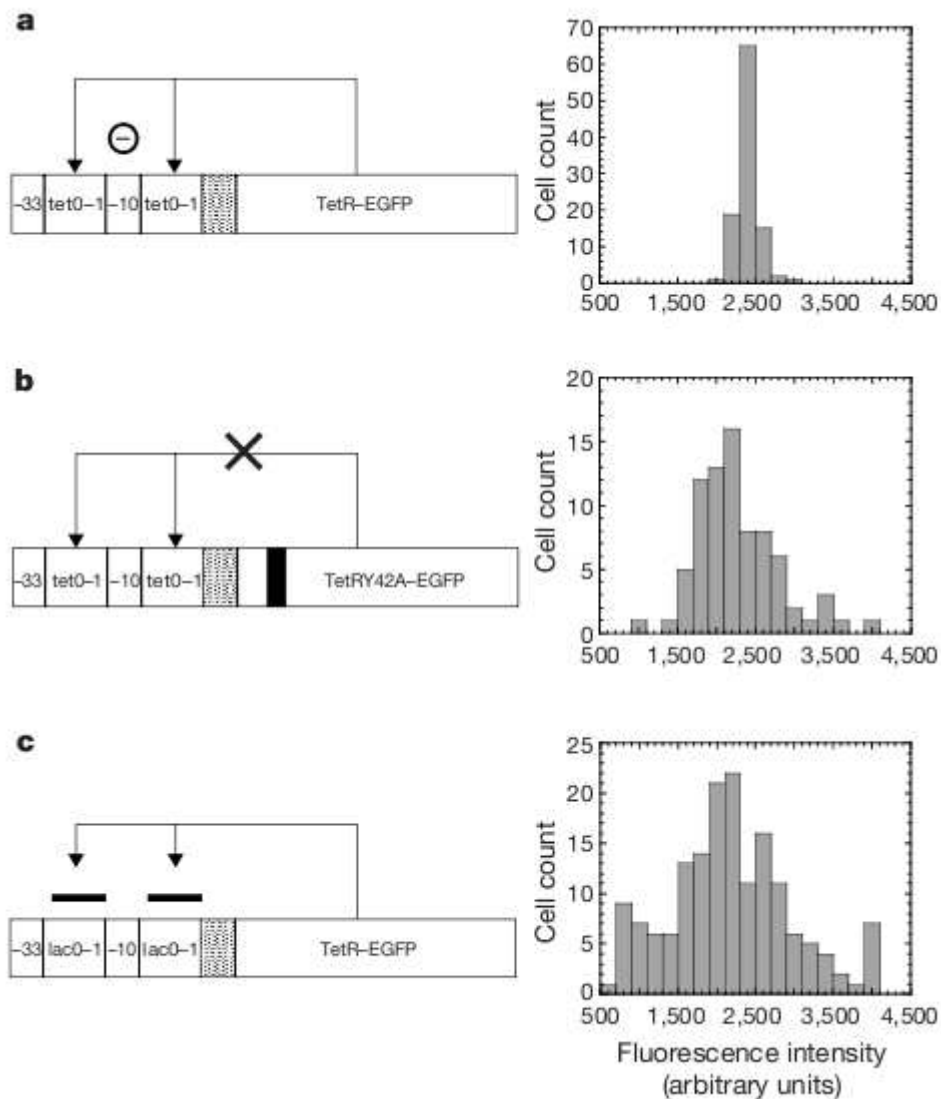


FIG. 1.1 – La figure représente trois constructions génétiques et la distribution de fluorescence de chacune d’elles. (a) Le premier système est une autorégulation négative exercée par TetR. (b) Ce système n’est pas autorégulé : le répresseur a été muté et ne peut plus se lier à l’ADN. (c) Ce système n’est pas autorégulé : les deux sites opérateurs ont été remplacés pour également obtenir un système non régulé. Les intensités moyennes des fluorescences des deuxième et troisième constructions ont été normalisées sur la fluorescence moyenne de la première d’un rapport de 38 et 14 respectivement. Figure issue de [16].

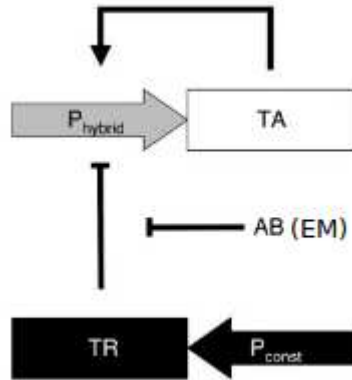


FIG. 1.2 – Réseau étudié analytiquement et expérimentalement dans le travail de B. Kramer et M. Fussenegger. Le réseau se compose d’une boucle rétropositive exercée par la protéine TA. Le réseau peut être allumé ou éteint en faisant varier la concentration de l’antibiotique EM qui inactive le répresseur TR qui agit sur le promoteur de TA. Figure issue de [80].

terme  $-\alpha TA$  représente la dégradation de  $TA$ . On peut montrer qu’en variant la concentration du répresseur actif TR (c’est-à-dire en changeant la concentration de l’antibiotique EM) l’équation 1.9 peut avoir une ou deux solutions, ce qui implique que le système peut avoir un ou deux états stationnaire(s) stable(s) (propriété de bi-stabilité). Cette propriété apparaît clairement sur la Figure 1.3 où l’on peut voir que pour certaines valeurs de EM, la concentration de TA peut converger vers une valeur haute ou vers une valeur basse. En fait, le système se dirigera vers l’une ou l’autre de ces deux valeurs suivant son histoire : si le système part de l’état haut (ON to OFF sur la Figure 1.3) il se dirigera vers la valeur haute, s’il part de l’état bas (OFF to ON sur la Figure 1.3), il convergera vers la valeur basse. C’est pour cette raison que la propriété de bistabilité est souvent associée à la notion de mémoire car le système peut “retenir” son état initial et se placer dans un état stationnaire et stable en fonction de celui-ci. Ceci a par exemple un impact en biologie du développement où un réseau génétique est à la base d’une décision pour qu’une cellule devienne tel ou tel type différencié [178].

## Expérimentation

Pour réaliser expérimentalement le réseau représenté Figure 1.2, B. Kramer et M. Fussenegger ont construit la partie contenant la boucle rétropositive de TA sur un plasmide et la partie contenant le répresseur TR sur un autre. Le rôle de l’activateur est joué par la tétracycline. La protéine qui sert de rapporteur est la protéine SEAP (secreted alkaline phosphatase) qui est fusionnée au gène de la tétracycline. Ils ont utilisé un répresseur (E-KRAB) sensible à l’antibiotique d’érythromycine (EM). C’est en variant la concentration de cette dernière qu’ils peuvent stimuler ou non le système. Ils ont ensuite inséré ces constructions dans des cellules de mammifères (ovaires de hamsters chinois). Le principe important de ces expériences est que les mesures des états stationnaires sont faites avec des cellules qui partent de l’état ON

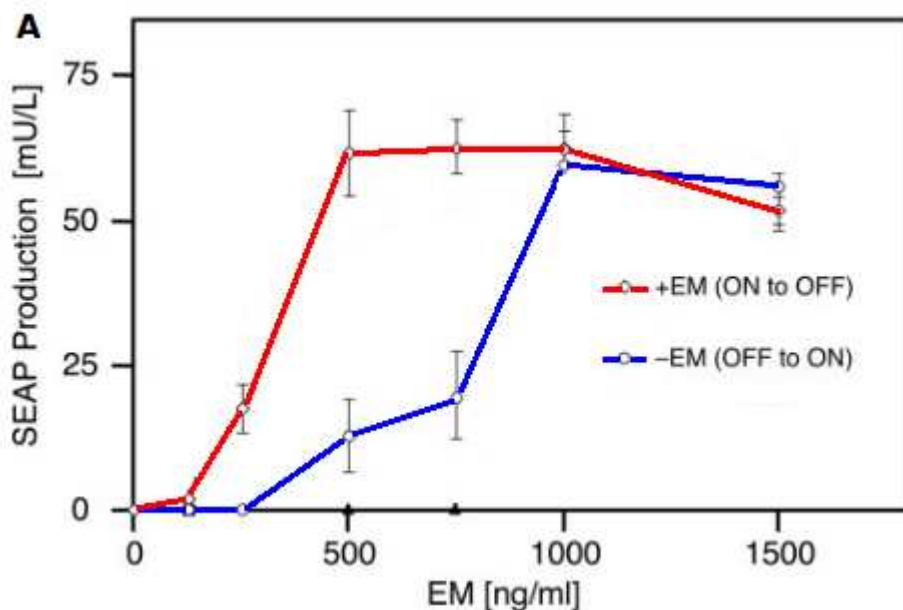


FIG. 1.3 – Réponse du système synthétique (représenté Figure 1.2). Pour différentes valeurs de l'antibiotique EM, la valeur stationnaire atteinte par le système est représentée. La courbe de couleur bleue donne l'état stationnaire atteint lorsque le système est parti d'un état de faible expression (état OFF) et la courbe rouge représente les états stationnaires atteints lorsque le système est parti de l'état haut (état ON). Le *cycle d'hystérésis* caractéristique d'un système bistable est clairement visible. Figure issue de [80].

et avec des cellules qui partent de l'état OFF. Pour faire partir les cellules de l'état ON, ils les ont induites en les incubant dans un milieu riche en EM pendant 3 jours ou dans un milieu sans antibiotique pour les faire partir de l'état OFF. Puis ils ont fait une cinétique de 48 heures pour mesurer la valeur de la concentration en SEAP pour situer l'état stationnaire.

### 4.3 Oscillateur génétique

Nous donnons ici un exemple de construction d'un oscillateur génétique basé sur le principe proposé par Barkai et Leibler [15]. Un des objectifs du travail de Atkinson et al. [10] est de construire un oscillateur à l'échelle de la population bactérienne et non sur une cellule unique comme dans le travail expérimental précédent du Repressilator de Elowitz et Leibler [38].

#### Théorie et modèle

Le module construit est un réseau minimal admettant des oscillations, il fait intervenir une boucle rétronégative et une boucle rétropositive (voir Figure 1.4). L'étude analytique de ce réseau montre qu'il est possible d'obtenir des oscillations dans les concentrations des protéines [58]. Un modèle théorique contenant les équations des différentes espèces ARN messagers et protéines peut être écrit avec le formalisme

présenté précédemment. Avec ces équations, il est possible de dessiner *le portrait de phase* du système c'est-à-dire un graphe qui donne les trajectoires dynamiques possibles que le système peut adopter pour différents jeux de paramètres cinétiques. Il est alors possible d'orienter les paramètres expérimentaux vers des valeurs pour que le système acquiert le comportement désiré. Par exemple, il est possible d'obtenir comme information quel paramètre (comme le temps de vie du répresseur) faut-il modifier pour être dans un régime d'oscillations auto-entretenu.

## Expérimentation

Les constructions se composent d'un module Activateur et d'un module Répresseur qui ont été placés en copie unique sur le chromosome principal d'*Escherichia coli* par recombinaison. La dynamique est observée en phase stationnaire. Pour maintenir pendant une longue durée les nutriments en quantité suffisante et une densité bactérienne constante, les auteurs ont utilisé un *chemostat*<sup>4</sup>. Le schéma des constructions est représenté sur la Figure 1.4. La boucle d'autoactivation est réalisée par le promoteur du gène *glnA* d'*Escherichia coli* (*glnAp2*) qui est fusionné au gène *glnG* qui code pour l'activateur NtrC (noté NR1 sur la figure). La transcription du promoteur *glnAp2* requiert la forme phosphorylée du facteur de transcription NtrC (noté NR1-P sur la figure). La fixation se fait sur deux sites opérateurs forts situés à distance du site de fixation de l'ARN polymérase. L'activation se fait par la formation d'une boucle d'ADN, les protéines d'NtrC vont aider à la formation du complexe ouvert (cf. Chapitre 3). Le gène *glnG* va donc s'autoactiver et aussi activer le répresseur LacI qui le réprime en retour formant ainsi la boucle rétronégative nécessaire pour l'apparition des oscillations.

Le promoteur *pglnA2* est réprimé par deux sites opérateurs forts (séquences idéales) pour le répresseur LacI en formant également une boucle d'ADN (cf. Chapitre 3). Le promoteur pour le répresseur est celui du gène de *glnK* d'*Escherichia coli* (*glnKp*) et requiert la forme phosphorylée de NtrC. Cependant les sites de liaison sont de plus faible affinité. Ce promoteur est fusionné au gène *lacI* qui va jouer le rôle de répresseur pour le gène *glnG*, ainsi que pour l'opéron *lacZYA* présent sur le chromosome principal.

Les deux modules ont été insérés dans des souches où les gènes *lacI*, *glnG* et *glnK* ont été inactivés de façon à ce que la seule source de NtrC provienne du module activateur et la seule source de LacI provienne du module répresseur.

Dans les cellules ainsi construites, l'opéron *lacZYA* et les opérons régulés par NtrC comme *glnA* (codant pour la synthétase de la glutamine) sont sous le seul contrôle des oscillateurs et ce sont eux qui vont servir de gènes rapporteurs.

Avec cette construction, les auteurs ont pu obtenir des oscillations amorties de période de 20h indépendamment du temps de division qui est de 2h (pour certaines

---

<sup>4</sup>Un chemostat est un appareil qui permet par un système de pompes et de filtres de maintenir un milieu de culture en quantité constante de nutriments et donc en concentration bactérienne constante en fonction du temps également. Il y a peu de travaux qui utilisent cet appareil dans le cadre d'un travail modélisation / expériences. Il nous paraît cependant tout à fait approprié pour les confrontations modélisation mathématique / expériences puisqu'une des premières informations acquises lors de travaux de modélisation est l'état stationnaire du système. Il est possible que beaucoup de réseaux génétiques aient besoin d'un chemostat pour pouvoir atteindre leur état stationnaire. Il semble que cela soit par exemple le cas pour le réseau de régulation d'entrée en virulence de la bactérie *Erwinia chrysanthemi* sur lequel nous avons travaillé (cf. Chapitre 4).

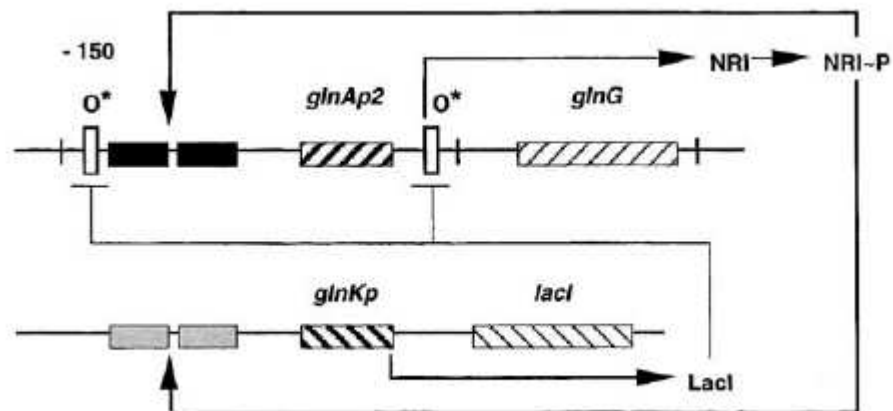


FIG. 1.4 – Construction génétique pour l'oscillateur artificiel du travail de Atkinson et al. [10]. Les promoteurs utilisés sont ceux des gènes *glnA* et *glnK* d'*Escherichia coli*. La transcription du promoteur *glnAp2* requiert la forme phosphorylée du facteur de transcription NtrC (noté NR1-P). Le promoteur est fusionné au gène *glnG* qui code pour le facteur de transcription NtrC (NR1). Les sites de liaison pour NtrC sont indiquées par des boîtes de couleur noire. Ce même promoteur est réprimé par les deux sites opérateurs pour le répresseur LacI (boîtes vides  $O^*$ ). Le promoteur de *glnK* requiert aussi la forme phosphorylée de NtrC. Cependant les sites de liaison pour ce facteur (boîtes grises) sont plus faibles. Ce promoteur est fusionné au gène *lacI* qui va coder pour le répresseur.



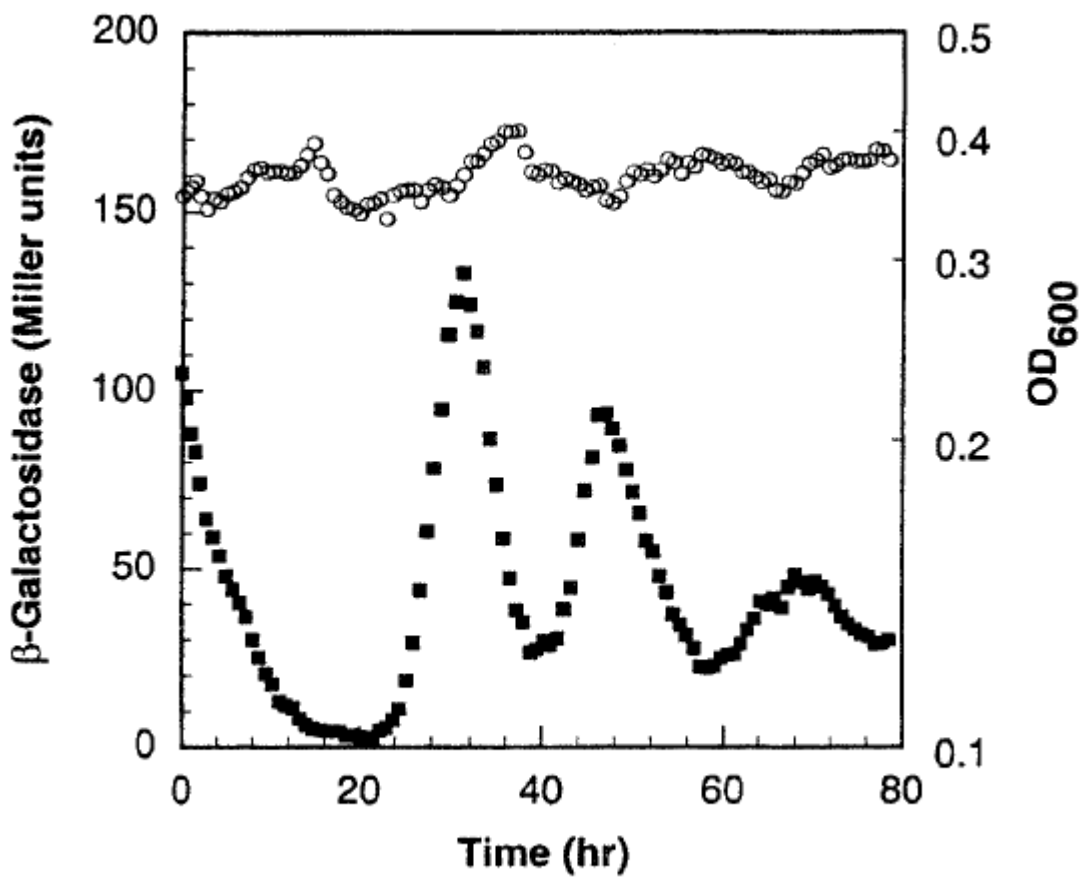


FIG. 1.5 – Oscillations pour le réseau synthétique représenté sur la figure 1.4. Trois oscillations amorties d’une période d’environ 20h sont visibles dans l’activité de la  $\beta$ -galactosidase. La densité optique du milieu qui correspond à la concentration bactérienne est également montrée, elle est maintenue constante grâce à un chemostat.

conditions, voir Figure 1.5). Comme pour le Repressilator, l’état génétique du réseau est donc transmis d’une génération à l’autre. Ils ont pu observer des oscillations sur des durées de 70h montrant la robustesse du réseau face au bruit ce qui constitue une amélioration par rapport au Repressilator.

En revanche, ils n’ont pas réussi à obtenir des oscillations auto-entretenues malgré une tentative de s’en approcher en augmentant notamment l’activation. Plus précisément, ils ont augmenté le nombre de copies du gène *glnG* de la construction en le plaçant plus près de l’origine de réplication du chromosome d’*Escherichia coli*.

Remarquons que pour ce système génétique synthétique, les deux processus de régulation font intervenir la formation d’une boucle d’ADN. En effet, d’une part, l’activation est exercée sur le promoteur *glnAp2* par la formation d’une boucle d’ADN grâce au facteur de transcription NtrC et aux deux sites enhancers placés à distance du site de fixation de l’ARN polymérase. Cette boucle aide à la fixation de l’ARN polymérase et à la formation du complexe ouvert. D’autre part, la répression

est exercée par une boucle d'ADN qui peut se former grâce aux deux sites opérateurs pour le répresseur LacI. En plus de la dynamique temporelle, la configuration géométrique de l'ADN joue donc aussi une part très importante dans l'expression génétique.

## 5 Conclusion

Au travers de ce chapitre d'introduction, nous avons vu combien des processus qui agissent "autour" de la séquence génétique à proprement dite peuvent être très riches et contenir en eux des informations qui participent à la complexité d'un processus biologique ou d'un organisme.

Les exemples utilisés dans ce chapitre ont illustré le fait que le formalisme et les concepts issus notamment de la physique non-linéaire ont donné des prédictions ou des éclaircissements dans la compréhension du vivant et des concrétisations sur le plan expérimental.

Nous pouvons remarquer que certains concepts issus de la physique théorique ou des mathématiques appliquées ont mis un certain temps pour être appréhendés sur le plan expérimental. La première raison est d'ordre technique : il a fallu dans certains cas que de nouvelles techniques expérimentales apparaissent pour pouvoir mettre en évidence de façon claire un phénomène notamment à l'échelle d'une cellule unique. Cependant, cela peut-être aussi dû à la durée nécessaire pour que ces idées se diffusent d'une communauté à l'autre.

Cette thèse s'est intéressée sur les plans théorique mais aussi expérimental à deux champs de recherche qui nous paraissent prometteurs. Le premier champ dans lequel nous avons travaillé est celui de la modélisation des réseaux de régulation que nous avons introduit dans ce chapitre. Dans ce cadre, nous nous sommes intéressés à l'aspect temporel qui peut intervenir dans les processus de régulation et que nous développons dans le Chapitre 2. Le second champ de recherche dans lequel nous avons travaillé considère l'impact de la configuration spatiale du polymère d'ADN sur les processus de régulation. Nous présentons le travail associé à ce second thème dans les Chapitres 3, 4 et 5.



# Chapitre 2

## Stimulations périodiques et régulation génétique

Dans le cadre de la modélisation des systèmes biologiques, le fait de s'intéresser au comportement de plusieurs gènes et protéines sous forme de réseaux a permis l'élaboration de concepts importants et novateurs qui ont eu des répercussions et des applications expérimentales comme vu dans le chapitre 1. Beaucoup de systèmes biologiques sont soumis à des rythmes réguliers ou à des stimulations périodiques. Nous nous sommes alors posés la question de savoir quels pourraient être les réseaux simples, existant ou déjà connus qui pourraient répondre de manière optimale à un signal périodique. En utilisant l'approche par équations différentielles non-linéaires et la résolution numérique, nous avons recherché de tels réseaux. Un des réseaux répertoriés chez la bactérie modèle *Escherichia coli* appelé "Incoherent Feed Forward Loop" présente des propriétés intéressantes concernant la réponse de signaux dépendant du temps. Des extensions de ce réseau sont proposées et des applications expérimentales sont évoquées.

### 1 Motivation et travaux antérieurs

Nous motivons dans cette section notre sujet d'étude et présentons les travaux antérieurs qui se sont également penchés sur ces questions. Jusqu'à présent, beaucoup de travaux expérimentaux ou de modélisation se sont intéressés aux états stationnaires des réseaux de régulation c'est-à-dire aux états où les concentrations en ARN ou en protéines sont constantes en fonction du temps. Dans des données de puces à ADN par exemple, qui donnent une valeur qui quantifie l'expression en ARN d'un ensemble de gènes, l'information sur l'évolution dans le temps des concentrations n'est pas connue. Dans certaines expériences ou dans certains contextes de la nature, il est cependant important de suivre dans le temps la concentration notamment en protéines. En effet, il est fréquent que des cellules ou des bactéries soient soumises à une stimulation périodique ou à des rythmes réguliers qui influent sur le comportement ou sur le cycle cellulaire [50]. Ceci se retrouve sur des échelles de temps très variées (voir tableau 2.1).

Par exemple, une cellule neuronale est soumise à des signaux dépendant du temps. Lors de l'apprentissage, les neurones reçoivent des stimulations régulières espacées dans le temps. Cette propriété est essentielle pour la formation de souve-

Système	Période
Oscillations des ions $Ca^{2+}$ dans le cytoplasme	1 min
Oscillations de P53 et Mdm2	5.5 h
Oscillations circadiennes	1 jour
Oscillation des hormones LH ou FSH	1 mois

TAB. 2.1 – Exemples de rythmes biologiques.

nirs [184]. D'autres exemples peuvent être trouvés dans des contextes différents : des cellules soumises à des variations de température ou à des changements périodiques de pression osmotique [109]. De nombreux organismes, des bactéries jusqu'aux mammifères les plus complexes, subissent l'alternance du jour et de la nuit, un rythme régulier qui peut influencer sur de nombreux processus. Il est intéressant de comprendre si une réponse optimale peut apparaître lors de stimulation de ce type et est-ce que la forme du signal temporel a une importance pour la réponse du système. Par exemple, est-ce que la production moyenne d'une protéine ou l'amplitude de la réponse peuvent être optimisées suivant le type du signal temporel d'entrée ou suivant le réseau de régulation sous-jacent. Ce chapitre se propose d'apporter quelques éléments de réponse à ces questions.

## 1.1 Travaux expérimentaux

De récents travaux expérimentaux ont été réalisés sur des systèmes biologiques pour comprendre leur réponse et leur propriétés lorsqu'ils sont soumis à une stimulation périodique. Par exemple, deux travaux parus en 2008 se sont tous deux intéressés à la réponse de la cascade MAPK Hog [63, 109] qui est chargée de l'adaptation de *Saccharomyces cerevisiae* aux variations de pression osmotique. Ces travaux semblent converger vers la conclusion qu'un tel système répond comme un filtre passe-bas aux variations de pression osmotique c'est-à-dire que pour des changements très rapides le système ne voit plus les variations. Les deux travaux sont basés sur des techniques de microfluidique.

Un travail antérieur [32], paru en 1998 a montré l'effet de la fréquence des oscillations des ions  $Ca^{2+}$  du cytosol sur l'expression de certains gènes. Des oscillations rapides stimulaient trois facteurs de transcription (NF-AT, Oct/OAP and NF- $\kappa$ B) alors que des oscillations plus lentes n'activaient seulement le facteur NF- $\kappa$ B. Enfin, le travail de Lipan et Wong [96] propose une méthode expérimentale basée sur l'utilisation de signaux périodiques pour étudier les réseaux génétiques. La méthode envisagée se base sur l'utilisation de champs électromagnétiques qui, à différentes fréquences, activent ou inactivent des protéines régulatrices ce qui permet de moduler périodiquement l'expression d'un gène.

## 1.2 Travaux théoriques

Ces questions n'ont pas encore reçu beaucoup d'attention sur le plan théorique hormis quelques travaux dont la plupart considèrent la réponse linéaire à des perturbations sinusoïdales. Par exemple, plusieurs auteurs démontrent la propriété de filtre passe-bas pour des réseaux moléculaires [51]. Pratiquement toutes ces études

considèrent des perturbations de faibles amplitudes, ce qui permet l'utilisation de la transformée de Fourier. Un point intéressant traité est par exemple comment les paramètres du système modifient la valeur de la fréquence de coupure. Au-dessus de la fréquence de coupure, le système n'a plus le temps de voir les oscillations et son amplitude de réponse devient nulle. Une grande majorité de systèmes physiques ou biologiques se comportent naturellement de cette façon. Il faut des ingrédients précis pour faire apparaître des propriétés de réponse en fréquence qui ne soient pas celles d'un filtre passe-bas. Peu de travaux de modélisation se sont déjà dirigés dans cette direction. Cependant, parmi eux, une récente étude [98] utilisant les outils de linéarisation et d'analyse de Fourier, montre qu'une cascade de signalisation avec une autorégulation négative peut se comporter comme un filtre passe-bande avec une fréquence caractéristique qui maximise l'amplitude de réponse. Cependant lorsque les non-linéarités du système sont importantes, comme cela est souvent le cas pour des modèles de réseaux de régulation, ou lorsque le système est linéaire mais avec un forçage périodique multiplicatif, alors la décomposition de Fourier du signal d'entrée n'est plus possible. De plus, les stimulations périodiques exercées sur des réseaux de régulation ne sont pas toujours harmoniques (forme sinusoïdale) mais sont plus proches de signaux types créneaux (alternance de phases "on" et de phases "off"). Parmi les auteurs qui se sont penchés sur ce problème, Smolen et al. ont étudié un réseau moléculaire stimulé par un train périodique de pulses [158], qui montre la capacité de maximiser sa réponse à un signal dont l'intervalle de temps entre les pulses doit être précisément choisi. Cependant ce modèle fait intervenir de nombreuses équations et paramètres et le mécanisme à l'origine de la propriété de filtre passe-bande n'apparaît pas de façon simple dans les équations. Un autre exemple de stimulations périodiques par un signal de pulses carrés a été étudié en détails par Li et Goldbeter dans le contexte des systèmes de récepteurs cellulaires [94]. Grâce à des calculs analytiques et des simulations numériques d'un modèle linéaire [93], ils montrent qu'il existe un schéma optimal pour un signal de pulses carrés qui maximise une observable du système définie par les auteurs comme l'*activité du récepteur*. Cependant, le modèle développé par Li et Goldbeter a également beaucoup de paramètres et l'identification des ingrédients essentiels qui permettent d'avoir un maximum dans l'activité du récepteur n'est pas évidente.

Notre objectif est de mettre en évidence des réseaux les plus simples possibles qui appartiennent par exemple à des motifs de régulation déjà connus ou répertoriés par la biologie systémique, et de faire ressortir quelques principes qui puissent garantir l'existence d'une réponse optimale lors d'une stimulation périodique. Nous allons rechercher des types de réseaux qui puissent s'appliquer au contexte de la régulation génétique mais également au contexte des voies de signalisation intracellulaire.

## 2 Stimulation périodique d'une simple activation

Nous commençons par étudier le cas d'une stimulation périodique d'une simple activation. Gène et protéine synthétisée sont considérés à l'aide d'une seule variable. Comme montré dans la Figure 2.1 (a), le gène  $Y$  est activé par le seul facteur de transcription  $X$ , l'équation 2.1 rend compte de cette simple régulation en utilisant le formalisme employé notamment par Alon [5] (cf. Chapitre 1) :

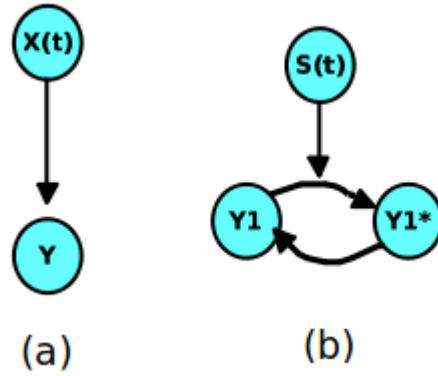


FIG. 2.1 – (a) Simple activation du facteur de transcription  $X$  sur un gène  $Y$  (b) Activation d’un cycle de protéines  $Y$  et  $Y^*$  par un signal extérieur  $S(t)$ .

$$\dot{Y} = \beta R(X) - \alpha Y \quad (2.1)$$

Dans cette équation,  $R(X)$  est la fonction de régulation, nous la choisissons bornée entre 0 et 1. Comme nous l’avons vu au chapitre 1, une fonction de Hill peut être utilisée :

$$R(X) = \frac{X^n}{\theta^n + X^n}. \quad (2.2)$$

Les valeurs numériques des différents paramètres ne sont pas encore connues de façon certaine et varient d’un système à l’autre. Nous essayons dans la suite de notre travail de trouver des valeurs numériques les plus représentatives que nous avons tirées de travaux expérimentaux. Pour la valeur du paramètre  $\theta$ , une valeur de 100 nM est raisonnable [144].  $\beta$  est le taux de production des protéines dans l’état non réprimé,  $10 \text{ nM} \cdot \text{min}^{-1}$  est une valeur raisonnable [144]. Enfin le taux de dégradation des protéines, qui rend compte du temps de demi-vie des protéines est de l’ordre de  $0.01 \text{ min}^{-1}$ . Nous utilisons dans un premier temps, l’approximation logique qui consiste à utiliser des fonctions Heaviside à la place des fonctions de Hill (formellement ceci revient à considérer  $n \rightarrow \infty$ ). Nous avons donc :

$$R(X) = H(X - \theta) \quad (2.3)$$

Lorsque  $R(x) = 1$ , le système converge vers un état stationnaire  $\kappa = \frac{\beta}{\alpha}$ .

Dans la suite, nous allons faire dépendre du temps la concentration  $X(t)$  qui prendra la forme d’un signal créneau d’amplitude 1 comme montré sur la Figure 2.2.

Le paramètre  $\tau$  représente la durée de la phase “on” correspondant à l’activation par le facteur de transcription  $X$ . L’intervalle entre deux pulses (ou encore la phase de “silence”) est notée  $\sigma$ . La période du signal de stimulation est donc  $T = \tau + \sigma$ . En réponse au signal d’entrée  $X(t)$ , nous pouvons calculer trois variables intéressantes à savoir les valeurs extrêmes : le minimum et le maximum atteints par la protéine stimulée  $Y(t)$  ainsi que sa valeur moyenne notées respectivement  $Y_{min}$ ,  $Y_{max}$  et  $\langle Y \rangle_T$ . Les valeurs extrêmes ont une importance quand la protéine  $Y$  est elle-même une

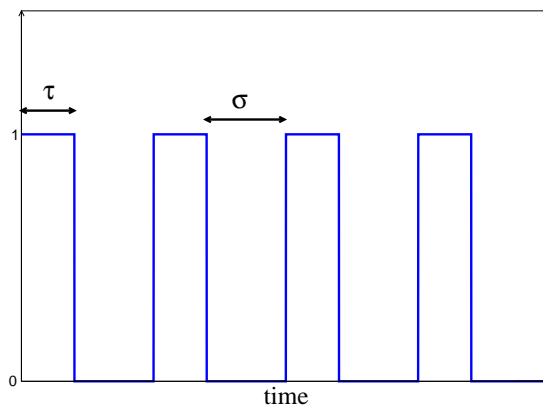


FIG. 2.2 – Signal créneau composé de phases “on” de durée  $\tau$  et de phase “off” d’une durée  $\sigma$  que nous utilisons comme signal de stimulation.

protéine régulatrice pour d’autres gènes puisque l’activation ou la répression des gènes dépend d’une valeur seuil qui peut ou non être franchie par  $Y$ . D’un autre côté, la concentration moyenne sur une période est aussi une quantité pertinente puisqu’elle donne la quantité de protéines produites et disponibles sur une période. Beaucoup de processus biologiques dépendent de cette quantité. Par exemple, la formation des connexions synaptiques dans les processus d’apprentissage dépend de la quantité moyenne de protéines produites. Dans le cas d’une simple activation avec un signal de pulses carrés pour l’équation 2.1, on peut facilement calculer analytiquement ces observables. Nous obtenons :

$$\begin{aligned}
 \langle Y \rangle_T &= \kappa \frac{\tau}{\tau + \sigma} \\
 Y_{\max} &= \kappa \frac{1 - e^{-\alpha\tau}}{1 - e^{-\alpha(\tau+\sigma)}} \\
 Y_{\min} &= \kappa \frac{e^{\alpha\tau} - 1}{e^{\alpha(\tau+\sigma)} - 1}
 \end{aligned} \tag{2.4}$$

La Figure 2.3 montre un exemple de ces fonctions quand l’intervalle entre les pulses  $\sigma$  est varié, pour une durée du pulse d’activation  $\tau$  fixée. Lorsqu’on augmente  $\sigma$ , les trois observables diminuent. Le maximum  $Y_{\max}$  reste dans l’intervalle  $[Y_L, \kappa]$ , où  $Y_L$  est la valeur asymptotique atteinte par  $Y_{\max}$  (ligne pointillée sur la Figure 2.3, obtenu quand le dénominateur de  $Y_{\max}$  égale 1). Ce niveau peut être contrôlé en choisissant la durée du pulse  $\tau$  ce qui peut s’avérer utile dans les processus de régulation transcriptionnelle.

Une autre classe importante de mécanismes d’activation en biologie moléculaire est la modification covalente de protéines (par exemple : la phosphorylation / déphosphorylation de protéines par des enzymes spécifiques) qui constitue la brique de base de nombreuses voies de signalisation cellulaire. Dans certains cas, l’enzyme qui permet la phosphorylation de la protéine, et donc son activation, est soumise à des



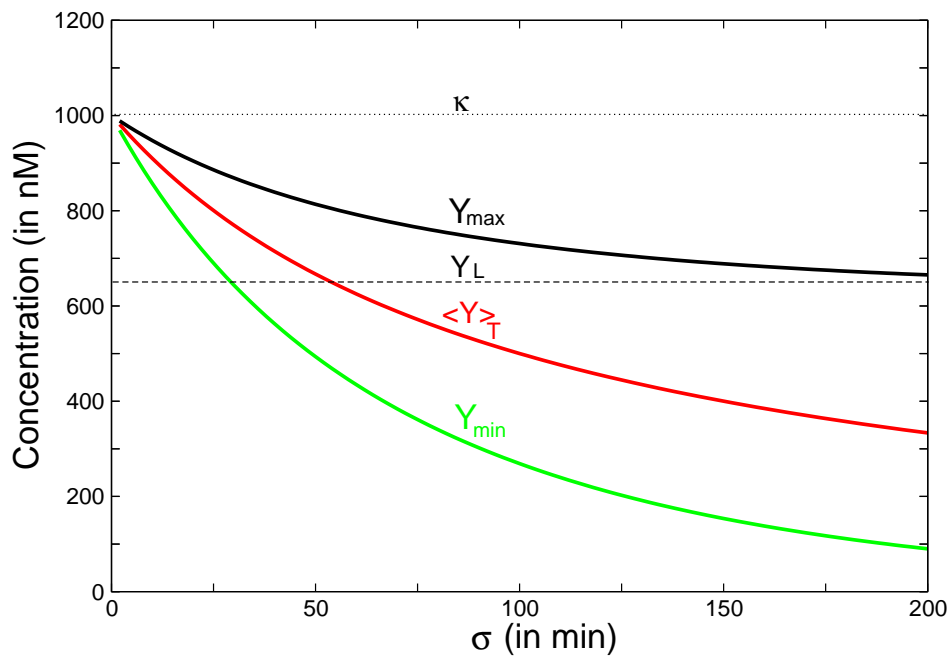


FIG. 2.3 –  $Y_{max}$  est le maximum atteint par la protéine,  $\langle Y \rangle_T$  est la concentration moyennée sur une période  $T$ ,  $Y_{min}$  est la concentration minimale atteinte,  $\kappa = \frac{\beta}{\alpha}$  (ligne pointillée) représente l'état stationnaire que la protéine atteint si la stimulation est constante et  $Y_L$  (ligne tiret) est la valeur asymptotique atteinte par  $Y_{max}$  quand l'intervalle entre les pulses  $\sigma$  devient très grand. Les paramètres sont ici :  $\alpha = 0.01 \text{ min}^{-1}$ ,  $\beta = 10 \text{ nM} \cdot \text{min}^{-1}$ ,  $\tau = 100 \text{ min}$ .

variations périodiques dans le temps. De nombreux exemples de cette situation ont été étudiés dans le contexte des oscillations des ions  $Ca^{2+}$  notamment de la phosphatase Calmoduline dont l'activité est modulée par des pulses spécifiques d'ions  $Ca^{2+}$  [35, 36, 48, 78]. Le motif de phosphorylation/déphosphorylation est habituellement représenté comme sur la Figure 2.1 (b) en dessinant deux flèches formant un cycle entre les différents états moléculaires, la troisième flèche indique la présence d'un signal externe stimulateur. Ici,  $Y$  correspond à la concentration de protéines non actives et  $Y^*$  à la concentration des protéines activées, la concentration totale de protéines est généralement une quantité conservée sur les échelles de temps :  $Y + Y^* = Y_{tot}$ . Plusieurs modèles théoriques peuvent être utilisés pour décrire la dynamique de l'interconversion entre  $Y$  et  $Y^*$  (on pourra trouver une revue récente dans [171]). Une forme simple d'équation pour décrire la dynamique d'un cycle entre protéines est donnée par l'équation suivante (d'après [62]) :

$$\dot{Y}^* = kS(t)(Y_{tot} - Y^*) - k'Y^* \quad (2.5)$$

où  $k$  et  $k'$  sont respectivement les constantes cinétiques des réactions d'activation et de désactivation. Ces équations sont obtenues en utilisant les lois d'action des masses et en faisant l'hypothèse que la concentration du complexe enzyme-substrat est petite devant la concentration totale des autres partenaires interagissant. Quand  $S(t)$  est un train de pulses périodiques, les observables  $Y_{max}$ ,  $Y_{min}$  et  $\langle Y \rangle_T$  peuvent être calculées de façon analytique en fonction de  $(\tau, \sigma)$ . Cependant, leurs expressions sont un peu plus compliquées que celles des équations 2.4 ce qui est dû aux différences dans les échelles de temps de la phase "on" et de la phase "off" de  $Y(t)$ . Pour une valeur fixée de  $\tau$ , le graphe de ces fonctions de  $\sigma$  est qualitativement le même que celui représenté Figure 2.3. Les valeurs extrêmes et la valeur moyenne de  $Y(t)$  diminuent lorsque l'intervalle entre les pulses  $\sigma$  est augmenté. En outre, des simulations que nous avons effectuées en utilisant le modèle Goldbeter-Koshland [48, 49] montrent que le comportement obtenu est le même dans le cas général d'équation non-linéaires représentant la dynamique du cycle.

Nous concluons de ces simples calculs qu'une activation sous forme de pulses de systèmes simples de régulation (2.1) ou (2.5) fournit une façon d'ajuster les concentrations de  $Y$  dans une plage de valeurs qui peut être contrôlée en réglant le schéma temporel  $(\tau, \sigma)$  des pulses.

## 3 Stimulation périodique d'une "Incoherent Feed Forward Loop"

### 3.1 L'Incoherent Feedforward Loop

Parmi les motifs<sup>1</sup> répertoriés par l'équipe de Uri Alon (voir par exemple la revue [6]), figure l'Incoherent Feed Forward Loop (en français, il faudrait probablement traduire par "circuit fermé incohérent" mais l'abréviation IFFL plus commode sera utilisée par la suite). Ce motif de régulation génétique se compose de 3 nœuds (voir

---

<sup>1</sup>Le terme *motif* désigne un réseau de régulation qui apparaît de manière plus importante dans un génome en comparaison avec des réseaux qui seraient générés aléatoirement (soulignant peut-être par là qu'ils confèreraient un avantage sélectif).

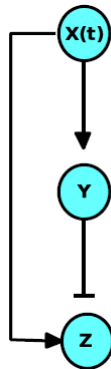


FIG. 2.4 – Le motif d’Incoherent Feed Forward Loop formé de 3 noeuds. Dans notre travail, l’expression de  $X$  est dépendante du temps avec un comportement périodique.

Figure 2.4) : un premier gène  $X$  qui active un second  $Y$  et le troisième  $Z$ , le gène  $Y$  réprime le gène  $Z$ . Ce motif intervient par exemple chez la bactérie *Escherichia coli* pour le système en charge de traiter le galactose [5]. Deux propriétés intéressantes de ce motif qui avaient été proposées de manière théorique ont été mises en évidence expérimentalement [102]. La première est que ce système peut générer une réponse sous forme d’un pulse lorsqu’il est soumis à une stimulation continue. La deuxième propriété est qu’il permet une réponse accélérée par rapport à une simple régulation. Nous montrons dans cette partie que ce motif présenterait également une propriété intéressante de traitement du signal lorsqu’il est soumis par un stimulus dépendant du temps.

Afin de simplifier le formalisme et pour obtenir des résultats simples analytiques, nous considérons à nouveau l’approximation logique, où les fonctions de régulation sont représentées par des fonctions de Heaviside qui prennent seulement la valeur 1 ou 0. Nous notons  $S(t) = R(X(t))$  un signal d’entrée provenant de  $X$ . La dynamique du système est donc décrite par les équations suivantes :

$$\dot{Y} = \beta S(t) - \alpha Y \quad (2.6)$$

$$\dot{Z} = \beta S(t) H(\theta - Y) - \alpha Z \quad (2.7)$$

La première équation est une simple régulation comme introduite dans la section précédente. La seconde équation gouverne la variable  $Z$  qui est activée seulement quand à la fois le signal  $S(t)$  est en phase “on” et quand le niveau du répresseur  $Y$  est en-dessous de la valeur seuil  $\theta$ . Pour des raisons de simplicité, nous avons considéré le même taux de dégradation ( $\alpha$ ) et le même taux de production maximal ( $\beta$ ) pour  $Y$  et  $Z$  mais ce choix n’est pas essentiel pour les résultats présentés dans cette section.

Quand le signal associé à  $X$  est en phase “on”, la variable  $Z$  commence par augmenter dû à l’activation par  $X$  mais cette activation est stoppée par le répresseur  $Y$  qui est également activé par  $X$ . La concentration de  $Z$  commence donc à décroître dès que  $Y$  dépasse la valeur seuil  $\theta$ . Ceci explique pourquoi ce motif répond sous la forme d’un pulse lorsqu’il est soumis à une stimulation continue comme montré par Mangan et al. [102]. Une valeur de référence pour la durée du pulse peut donc être

calculée par le temps  $\tau_\theta$  nécessaire pour que  $Y$  atteigne le seuil de répression  $\theta$  après l'établissement du signal. Pour le modèle simple de 2.7, la durée du pulse peut être simplement calculée :

$$\tau_\theta = -\frac{1}{\alpha} \log(1 - \theta/\kappa) \quad (2.8)$$

Si l'activation dure plus longtemps que  $\tau_\theta$ , le système peut retrouver son état initial. En d'autres termes, le motif IFFL peut seulement détecter des changements de stimulations, il *s'adapte* à un stimulus constant. Donc, comme mentionné par Li et Goldbeter [93] dans le contexte de la désensibilisation des récepteurs, pour atteindre un niveau donné de synthèse de  $Z$  dans un tel système, un schéma de stimulation sous forme pulsative est plus approprié qu'une stimulation continue. Nous considérons dans les sections suivantes un stimulus périodique  $S(t)$ , qui a la forme d'un signal créneau représenté Figure 2.2.

### 3.2 Stimulation périodique par un signal créneau

Nous nous proposons de stimuler le motif d'IFFL par un signal créneau périodique composé d'une phase "on" de durée  $\tau$  et d'une phase "off" d'une durée  $\sigma$  (comme montré sur la Figure 2.2). Dans un premier temps, nous faisons varier les durées  $\tau$  et  $\sigma$  et nous regardons la réponse de la quantité de la protéine  $Z$  moyennée sur une période. Nous représentons la réponse de  $\langle Z \rangle_T$  sur le graphe de la Figure 2.5. Premièrement, nous pouvons voir que le système va répondre de façon maximale pour certains couples de valeurs  $(\sigma, \tau)$ . En effet, nous pouvons observer que les valeurs optimales forment une arête sur le graphe de la fonction  $\langle Z \rangle_T$ , qui se projette dans le plan  $(\sigma, \tau)$  sur un segment quasi-rectiligne. Il s'avère que ce segment de droite a une pente d'environ  $\sigma/\tau \simeq \kappa/\theta - 1$ . De plus, nous pouvons voir sur une coupe de ce graphe que pour une durée  $\tau$  fixée, le système atteint un maximum pour une valeur précise de  $\sigma$ .

Lorsque la valeur  $\tau$  est fixée, nous pouvons voir également que la réponse devient non nulle qu'à partir d'une valeur particulière pour l'intervalle entre les pulses  $\sigma$  qui doit dépasser une valeur minimale (voir également les Figures 2.6 et 2.7 plus bas). Au vu de ces simulations, il apparaît que le motif d'IFFL a la propriété de ne laisser passer les trains périodiques qui ont une durée minimale de repos entre les phases d'activation. Plus précisément, pour la réponse de la protéine de sortie  $Z$ , deux cas peuvent être distingués selon que la durée du pulse est supérieure ou non à  $\tau_\theta$  de la IFFL. Nous allons décrire plus en détail ces deux cas et calculer les valeurs pour lesquelles le motif d'IFFL commence à répondre et donne une valeur optimale.

### 3.3 Étude plus détaillée des deux situations

Nous nous proposons dans cette section d'étudier plus en détails la réponse d'une IFFL à une stimulation périodique. L'activation de ce système par un train de pulses  $S(t)$  engendre des réponses temporelles périodiques pour  $Y(t)$  et  $Z(t)$  qui ont la forme de fonctions exponentielles par morceau. L'évolution temporelle de  $Y(t)$  peut être calculé analytiquement. En particulier, les équations 2.4 donnent les expressions analytiques pour les valeurs maximales et minimales de  $Y$ , noté  $Y_{\max}$  et  $Y_{\min}$ , ainsi que pour  $\langle Y \rangle_T$  la valeur de  $Y$  moyennée sur une période. Le taux de synthèse de  $Z$  est régi par la fonction de régulation  $S(t)H(\theta - Y(t))$ . Cette fonction a également

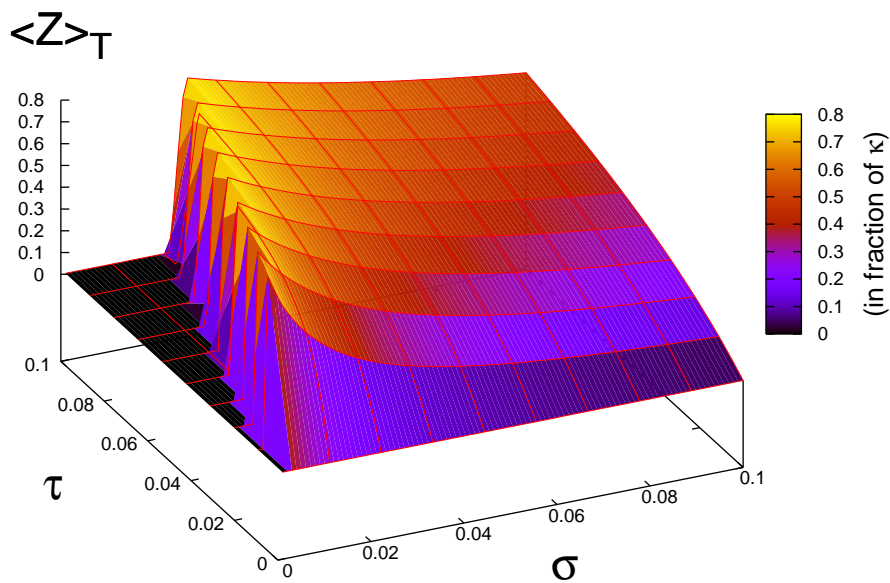


FIG. 2.5 – Réponse moyenne  $\langle Z \rangle_T$  en fonction de  $\sigma$  and  $\tau$  pour un motif d'IFFL (Figure 2.4) stimulé par un signal créneau (Figure 2.2). Les paramètres  $\sigma$  et  $\tau$  sont exprimés en unités d' $1/\alpha$  et  $\langle Z \rangle_T$  en fraction de  $\kappa = \frac{\beta}{\alpha}$ . Le paramètre de seuil  $\theta$  a été choisi égal à  $0.8 \kappa$ . La moyenne  $\langle Z \rangle_T$  peut être calculée analytiquement et est donnée par la formule de l'équation 2.13.

la forme d'un signal créneau de période  $T$  mais avec une plus petite durée pour la phase "on" notée  $\tau'$ . Le comportement temporel de  $Z$  a donc la même évolution qu'une simple activation mais stimulé par un train de pulses aux caractéristiques temporelles ( $\tau = \tau', \sigma = T - \tau'$ ), pour respectivement la durée du pulse et la durée entre les pulses. Nous pouvons alors avoir une expression pour les observables  $Z_{\max}$ ,  $Z_{\min}$  et  $\langle Z \rangle_T$  qui peuvent être calculés de la même façon que pour  $Y(t)$  mais en remplaçant  $\tau$  par  $\tau'$  et  $\sigma$  par  $T - \tau'$  dans les équations 2.4.

Deux cas doivent être distingués, selon la taille de la durée du pulse  $\tau$  en comparaison avec la valeur de  $\tau_\theta$  qui est le temps nécessaire pour que  $Y$  atteigne le seuil de répression  $\theta$  une fois la stimulation enclenchée.

Considérons un train de pulses périodiques avec la durée du pulse fixée à  $\tau$ , et regardons le comportement de  $\langle Z \rangle_T$  en fonction de la durée entre les pulses  $\sigma$ . Sur les Figures 2.6 et 2.7, nous pouvons voir que lorsque  $\sigma$  est proche de zéro, la valeur minimale du répresseur  $Y_{\min}$  reste au-dessus de  $\theta$  et donc  $Y$  réprime en permanence  $Z$ . Cette situation s'applique pour des valeurs de  $\sigma$  tant que  $Y_{\min}$  reste au-dessus de  $\theta$ . Notons  $\sigma_0$  le premier intervalle entre les pulses pour lequel  $Y_{\min} = \theta$ . Il peut être explicitement calculé :

$$\sigma_0 = \frac{1}{\alpha} \log \left( 1 + \frac{\kappa}{\theta} (e^{\alpha\tau} - 1) \right) - \tau \quad (2.9)$$

Pour des intervalles entre les pulses plus longs,  $\sigma > \sigma_0$ , le répresseur  $Y$  oscille avec une valeur minimum plus petite que  $\theta$  ce qui donne un bref intervalle de temps,  $\tau'$ , où l'expression de  $Z$  est activée. Donc  $Z$  commence à osciller, avec un maximum, un minimum et une valeur moyenne qui diminuent en fonction de  $\sigma$ . La durée de la phase "on" pendant laquelle  $Z$  est activé peut être calculée en considérant le temps mis par  $Y$  pour passer de  $Y_{\min}$  à  $\theta$ , c'est-à-dire :

$$\tau' = \frac{1}{\alpha} \log \left( \frac{\kappa - Y_{\min}}{\kappa - \theta} \right) \quad (\text{avec } Y_{\min} < \theta) \quad (2.10)$$

Ce résultat n'a du sens que si  $\tau' > 0$ , c'est-à-dire si  $Y_{\min} < \theta$  sinon nous définissons  $\tau' = 0$ . Par ailleurs, de l'équation 2.7, il apparaît clairement que l'activation de  $Z$  dure au plus  $\tau$ , donc  $\tau' \leq \tau$ . En utilisant cette condition, l'expression de  $Y_{\min}$  est une fonction de  $(\tau, \sigma)$  et peut être substituée dans la dernière équation pour avoir :

$$\tau' = \min \left( \tau, \frac{1}{\alpha} \log \left( \frac{1 - e^{-\alpha\sigma}}{(1 - e^{-\alpha(\tau+\sigma)})(1 - \theta/\kappa)} \right) \right) \quad (2.11)$$

### Cas des pulses de longue durée, $\tau > \tau_\theta$

La Figure 2.6 illustre ce cas. Ici  $Y_{\max}$  reste toujours au-dessus du seuil de répression  $\theta$  et  $\tau' < \tau$  est toujours valable. Dans ce cas  $Z_{\max}$  et  $\langle Z \rangle_T$  sont calculés en fonction de  $(\tau, \sigma)$  en utilisant l'équation 2.4 ce qui donne :

$$Z_{\max} = \kappa \left( \frac{1}{1 - e^{-\alpha(\tau+\sigma)}} - \left( 1 - \frac{\theta}{\kappa} \right) \frac{1}{1 - e^{-\alpha\sigma}} \right) \quad (2.12)$$

$$\langle Z \rangle_T = \frac{\kappa}{\alpha(\tau + \sigma)} \log \left( \frac{1 - e^{-\alpha\sigma}}{(1 - e^{-\alpha(\tau+\sigma)})(1 - \theta/\kappa)} \right) \quad (2.13)$$

en supposant toujours que  $\sigma \geq \sigma_0$  (eq 2.9) sinon  $Z_{\max} = \langle Z \rangle_T = 0$ . Dans ce cas,  $Z_{\max}$  est une fonction monotone et croissante de  $\sigma$ . En revanche,  $\langle Z \rangle_T$  a un maximum qui peut en principe se calculer en annulant la dérivée de  $\langle Z \rangle_T$  par rapport à  $\sigma$ .

### Cas des pulses de courte durée, $\tau < \tau_\theta$

Le second cas, illustré sur la Figure 2.7 concerne des durées de pulses inférieures à  $\tau_\theta$ . Il y a alors un lapse de temps  $\sigma_1$  entre les pulses pour lequel la valeur maximale atteinte par le répresseur  $Y_{\max}$  est en-dessous du seuil de répression  $\theta$  pour tout  $\sigma > \sigma_1$ . Cet intervalle entre les pulses peut être calculé grâce à la condition  $Y_{\max}(\sigma_1) = \theta$ , qui équivaut à :

$$\sigma_1 = -\frac{1}{\alpha} \log \left( 1 - \frac{\kappa}{\theta} (1 - e^{-\alpha\tau}) \right) - \tau \quad (2.14)$$

Donc, pour  $\sigma > \sigma_1$ , le répresseur n'est plus fonctionnel et seulement la régulation par le signal activateur  $S(t)$  a un effet. Pour  $\sigma > \sigma_1$ , on atteint la valeur  $\tau' = \tau$  dans l'équation 2.10 et  $Z_{\max} = Y_{\max}$ , et  $\langle Z \rangle_T = \langle Y \rangle_T$ , comme illustré sur la Figure 2.7. Dans ce cas, un maximum apparaît aussi pour la valeur maximale  $Z_{\max}$  quand l'intervalle entre les pulses prend la valeur  $\sigma = \sigma_1$ . Le même intervalle de temps donne la valeur optimale pour  $\langle Z \rangle_T$ .  $\langle Z \rangle_T$  est calculée en fonction de  $(\tau, \sigma)$  en utilisant l'équation 2.4 :

$$\langle Z \rangle_T = \frac{\min \left( \max \left( 0, \log \left( \frac{1 - e^{-\alpha\sigma}}{(1 - e^{-\alpha(\tau+\sigma)})(1 - \theta/\kappa)} \right) \right), \sigma \right)}{(\tau + \sigma)} \quad (2.15)$$

Le graphe de cette fonction est représenté sur la figure 2.5.

## 3.4 Réponse de l'Incoherent Feedforward Loop, modélisation avec des fonctions de Hill

Dans les sections précédentes, nous avons utilisé l'approximation logique pour rendre compte des régulations génétiques. Cela nous a permis d'obtenir des expressions analytiques pour  $Z_{\max}$  et  $\langle Z \rangle_T$ . Cependant, du point de vue quantitatif, cette approximation n'est pas appropriée pour beaucoup de systèmes biologiques qui ont une régulation plus graduelle, qui ne peut pas être modélisée par une fonction type "tout-ou-rien". Nous montrons dans cette section, le résultat de simulations de l'IFFL avec l'utilisation de fonctions de Hill classiques. Les équations du modèle peuvent s'écrire :

$$\dot{Y} = \beta_1 \frac{X^n}{\theta_1^n + X^n} - \alpha Y \quad (2.16)$$

$$\dot{Z} = \beta_2 \frac{X^n}{\theta_1^n + X^n} \frac{\theta_2^n}{\theta_2^n + Y^n} - \alpha Z \quad (2.17)$$

où  $X(t)$  est un signal externe. L'équation 2.17 modélise une régulation où les facteurs de transcription  $X$  et  $Y$  agissent de façon indépendante. Nous aurions pu également

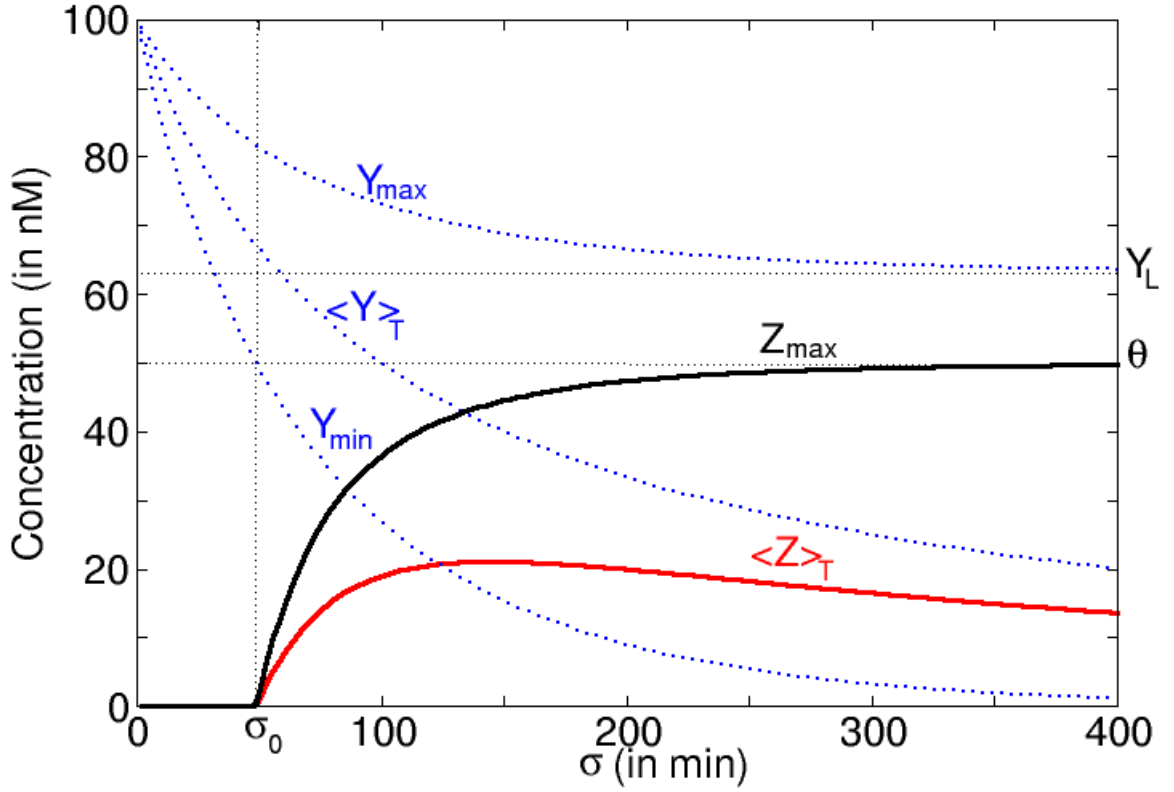


FIG. 2.6 – Réponse d'une IFFL à un signal périodique de pulses. La concentration moyennée sur une période  $T$   $\langle Z \rangle_T$  et  $Z_{max}$  la valeur maximum atteinte par  $Z$  sont représentées en fonction de l'intervalle de temps entre les pulses d'activation  $\sigma$  pour une durée de pulse fixée à  $\tau$ . Les valeurs extrêmes de la concentration du répresseur  $Y$  sont également représentées :  $Y_{max}$  est la concentration maximale atteinte,  $\langle Y \rangle_T$  est la concentration moyennée sur une période  $T$  et  $Y_{min}$  est la valeur minimale atteinte par  $Y$ .  $\kappa = \frac{\beta}{\alpha}$  (ligne pointillée) représente l'état stationnaire atteint par le système si la stimulation était continue et  $Y_L$  (ligne tiret) est la valeur asymptotique atteinte par  $Y_{max}$  quand l'intervalle entre les pulses  $\sigma$  devient très grand.  $\sigma_0$  est l'intervalle entre les pulses minimum pour que le système réponde et  $\sigma_1$  est l'intervalle pour lequel le système répond de façon optimale. La simulation numérique a été faite avec les paramètres suivants :  $\alpha = 0.01 \text{ min}^{-1}$ ,  $\beta = 1 \text{ nM} \cdot \text{min}^{-1}$ ,  $\theta = 50 \text{ nM}$ ,  $\tau_\theta = 70 \text{ min}$ . Avec  $\tau > \tau_\theta$ ,  $\tau = 100 \text{ min}$ , nous obtenons  $\sigma_1 \sim 145 \text{ min}$ .



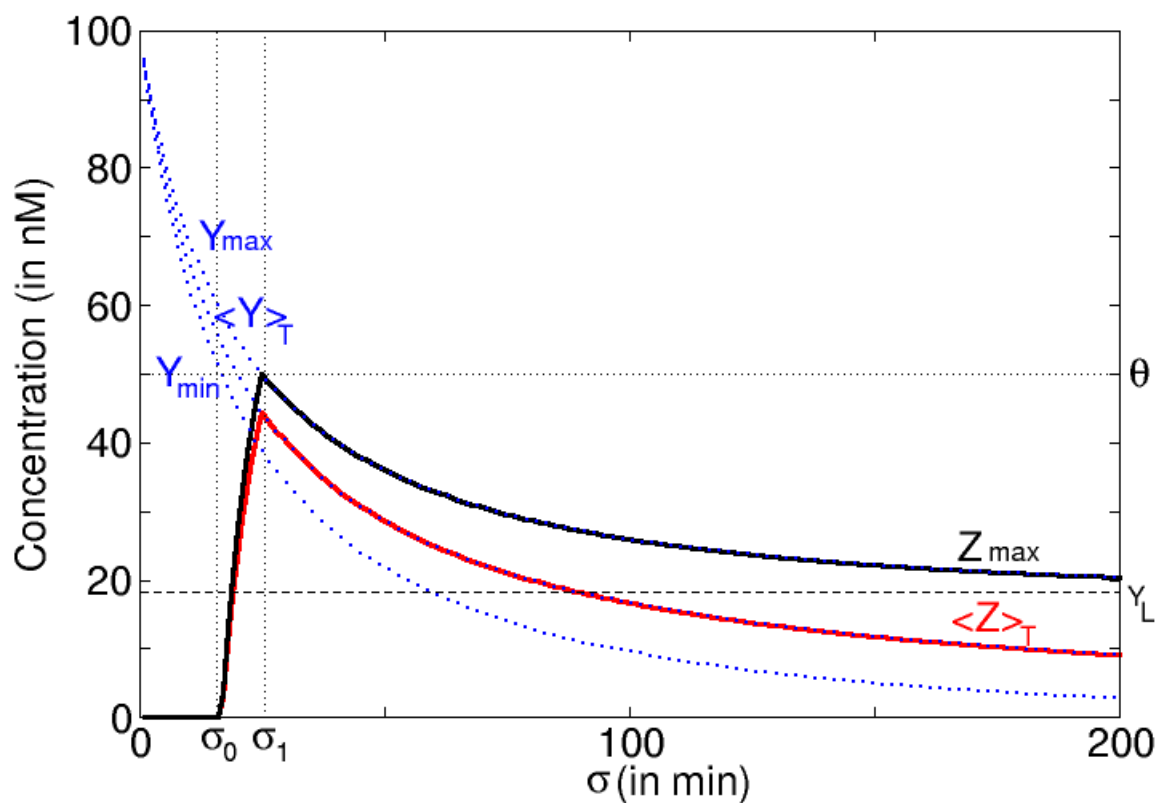


FIG. 2.7 – La concentration moyennée sur une période  $T$   $\langle Z \rangle_T$  et  $Z_{max}$  la valeur maximum atteinte par  $Z$  sont représentées en fonction de l'intervalle de temps entre les pulses d'activation  $\sigma$  pour une durée de pulse fixée à  $\tau$ . La simulation numérique a été faite avec les paramètres suivants :  $\alpha = 0.01 \text{ min}^{-1}$ ,  $\beta = 1 \text{ nM.min}^{-1}$ ,  $\theta = 50 \text{ nM}$ ,  $\tau_\theta = 70 \text{ min}$ . Avec  $\tau < \tau_\theta$ ,  $\tau = 20 \text{ min}$ , nous obtenons  $\sigma_1 \sim 25 \text{ min}$ .

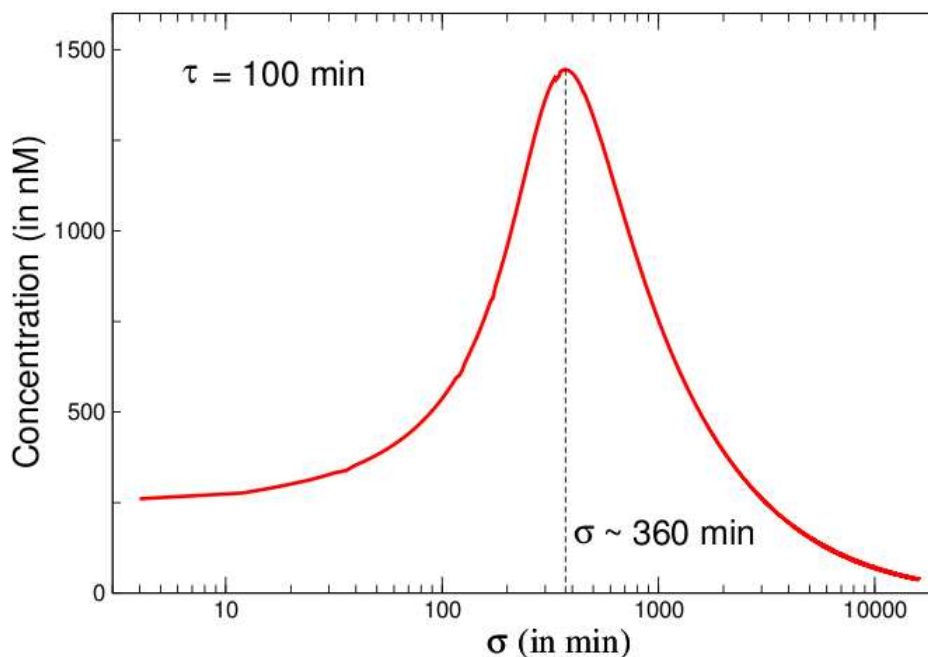


FIG. 2.8 – Réponse du motif d’IFFL à une stimulation périodique en fonction de  $\sigma$  simulée avec l’utilisation des fonctions de Hill.  $\langle Z \rangle_T$  la concentration moyennée sur une période  $T$  est représentée en fonction de l’intervalle entre les pulses  $\sigma$  pour une durée de pulse fixée à  $\tau$ . La simulation numérique a été réalisée en utilisant les équations 2.16 et 2.17 et avec les paramètres suivants :  $\alpha = 0.01 \text{ min}^{-1}$ ,  $\beta_1 = 10 \text{ nM.min}^{-1}$ ,  $\beta_2 = 1000 \text{ nM.min}^{-1}$ ,  $\theta_1 = 50 \text{ nM}$ ,  $\theta_2 = 50 \text{ nM}$ ,  $\tau = 100 \text{ min}$ . Le coefficient de Hill est  $n = 2$ . Pour ces valeurs de paramètres, nous obtenons  $\sigma_1 \sim 360 \text{ min}$ .

modéliser une régulation dans laquelle  $X$  et  $Y$  entrent en compétition. Pour cela, nous aurions pu utiliser l’équation pour l’évolution de  $Z$  :

$$\dot{Z} = \beta_2 \frac{X^n}{\theta_1^n + X^n + Y^n/\theta_2^n} - \alpha Z \quad (2.18)$$

Le principe d’optimisation de la réponse moyenne se retrouve également pour ce type de régulation. La Figure 2.8 montre la réponse pour une IFFL faisant intervenir les équations 2.16 et 2.17 avec un exposant  $n = 2$  ce qui est le plus courant dans les régulations génétiques puisque les protéines régulatrices se lient très souvent sous forme de dimères sur l’ADN. Nous retrouvons bien sur le graphe représenté en échelle logarithmique la propriété selon laquelle l’IFFL répond de manière significative pour une durée de repos entre les pulses bien définie.

## 4 Motif de type “Diamond IFFL”

La stimulation périodique d’un motif IFFL avec un signal créneau de pulses a révélé une propriété intéressante de réponse optimale selon le schéma temporel du signal d’entrée. Nous pouvons améliorer le précédent motif pour rendre sa fonction

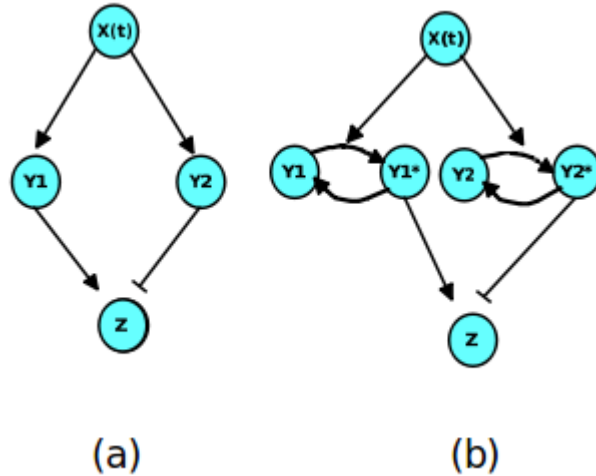


FIG. 2.9 – Exemples de réseaux de régulation faisant un motif type Diamond IFFL : (a) chaque noeud correspond à un gène/protéine (b) un réseau type Diamond faisant intervenir deux cycles de signalisation.

de filtre passe-bande plus sélective, c'est à dire lui conférer une capacité de répondre seulement sur une plage précise de périodes de pulsations. Nous considérons un motif type "Diamond IFFL" (que nous abrégeons par DIFFL) représenté sur la Figure 2.9. Le principe est très analogue à celui de l'IFFL mais l'activation sur le gène de sortie se fait en deux nœuds au lieu d'un. On peut décliner cette extension de la IFFL sous plusieurs formes notamment dans un réseau de régulation qui mélange voies de signalisation et régulation transcriptionnelle dans le même graphe d'interaction. Ce type de construction est en fait analogue aux réseaux qui ont été utilisés dans des études sur la formation de la mémoire à long terme au niveau moléculaire par l'équipe de Smolen [158].

Nous utilisons la description simplifiée de l'équation 2.5 pour la dynamique de modification covalente des cycles, avec les variables  $Y_1^*$  and  $Y_2^*$ , et l'approximation logique pour les fonctions de régulation sur le promoteur de  $Z$ . La dynamique du système représenté sur la Figure 2.9 (b) peut s'écrire comme précédemment avec les équations suivantes pour les protéines  $Y_1^*$ ,  $Y_2^*$  et  $Z$  :

$$\dot{Y}_1^* = kS(t)(Y_{1tot} - Y_1^*) - k'Y_1^* \quad (2.19)$$

$$\dot{Y}_2^* = kS(t)(Y_{2tot} - Y_2^*) - k'Y_2^* \quad (2.20)$$

$$\dot{Z} = \beta H(Y_1^* - \theta_1) H(\theta_2 - Y_2^*) - \alpha Z \quad (2.21)$$

Dans ce cas simple, on a choisi des valeurs identiques pour les paramètres cinétiques pour les cycles de modification covalente 1 et 2. Les évolutions dans le temps de  $Y_1^*$  et de  $Y_2^*$  sont synchronisées par le signal périodique  $S(t)$ . L'équation (2.21) montre que l'expression du gène  $Z$  est activée seulement lorsque  $\theta_1 < Y_i^* < \theta_2$  ( $i = 1, 2$ ), ce qui implique que la relation  $\theta_1 < \theta_2$  doit être satisfaite pour avoir une production de  $Z$ . Cette condition signifie que l'affinité de l'activateur pour le promoteur du gène  $Z$  est plus élevée que l'affinité du répresseur pour le même promoteur.

Nous considérons des stimulations périodiques pour ce système avec un signal

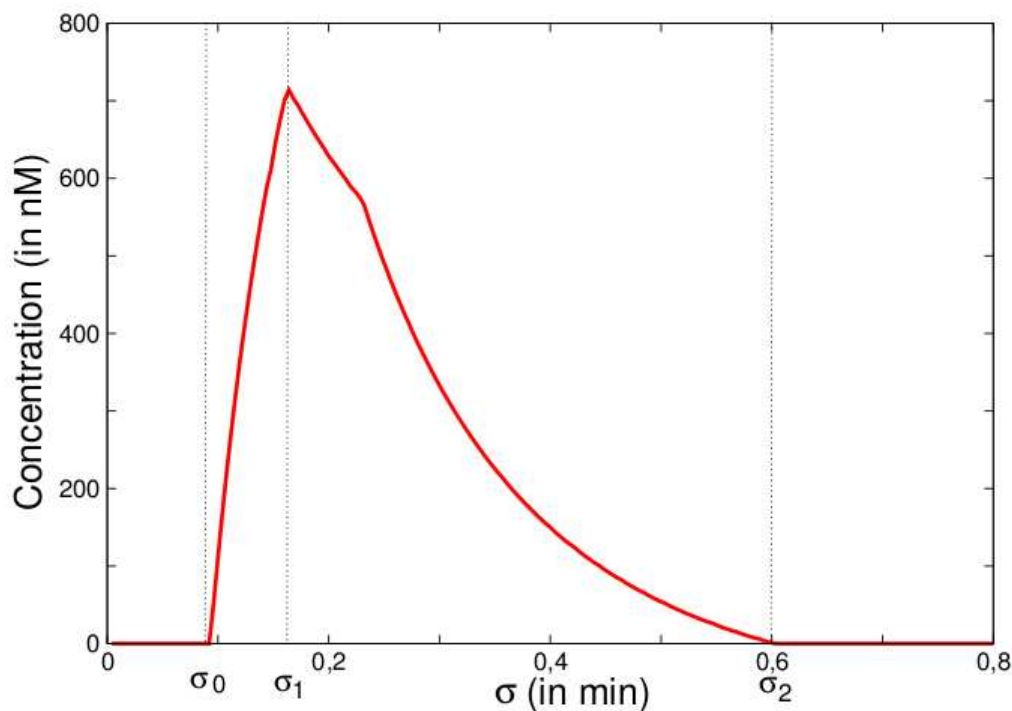


FIG. 2.10 – Réponse moyenne  $\langle Z \rangle_T$  de la protéine Z du motif Diamond IFFL avec des cycles de signalisation (représenté sur la Figure 2.9 (b)) pour une stimulation périodique.  $\sigma_0$  est la durée minimale entre les pulses pour que le système réponde. La durée entre les pulses  $\sigma_1$  donne la réponse moyenne la plus importante et  $\sigma_2$  est la durée maximale entre les pulses jusqu'à laquelle le système répond. La simulation numérique a été réalisée en utilisant les équations 2.21 et avec les paramètres suivants :  $k = 4 \text{ min}^{-1}$ ,  $k' = 2 \text{ min}^{-1}$ ,  $Y_{1Tot} = Y_{2Tot} = 500 \text{ nM}$ ,  $\alpha = 0.01 \text{ min}^{-1}$ ,  $\beta = 10 \text{ nM} \cdot \text{min}^{-1}$ ,  $\theta_1 = 180 \text{ nM}$ ,  $\theta_2 = 230 \text{ nM}$ ,  $\tau = 0.1 \text{ min}$ .

toujours de type créneau tel que représenté sur la Figure 2.2, en faisant varier  $\sigma$  la période de repos et en fixant la durée de l'activation  $\tau < \tau_{\theta_1}$  (cf. eq.(2.8)). La Figure 2.10 montre la réponse de la valeur moyenne de  $Z$  notée  $\langle Z \rangle_T$  en fonction de  $\sigma$ . Comme pour l'IFFL, il y a une valeur minimale de période de repos  $\sigma_0$  pour que le système donne une réponse non nulle dans l'expression du gène  $Z$ . Une fois que la valeur de  $\sigma$  est au-dessus,  $\langle Z \rangle_T$  augmente pour atteindre une valeur maximale pour une valeur  $\sigma = \sigma_1$  qui peut être calculée de la même façon que précédemment (eq.(2.14)). Si  $\sigma$  est encore augmenté,  $\langle Z \rangle_T$  diminue pour atteindre la valeur 0 dès que  $\sigma > \sigma_2$ . Cette caractéristique est due au nœud supplémentaire (activateur  $Y_1$ ) qui cesse d'activer le gène cible  $Z$  dès que la valeur maximale atteinte par  $Y_1$  devient plus petite que le seuil d'activation pour  $Z$ . La propriété nouvelle qui apparaît dans ce système, par rapport au motif IFFL simple, est donc la possibilité d'utiliser ce réseau comme un filtre passe-bande permettant au gène de sortie de ne répondre que pour un train de pulses relativement spécifique qui a une période de repos appartenant à une plage limitée  $[\sigma_0, \sigma_2]$ .

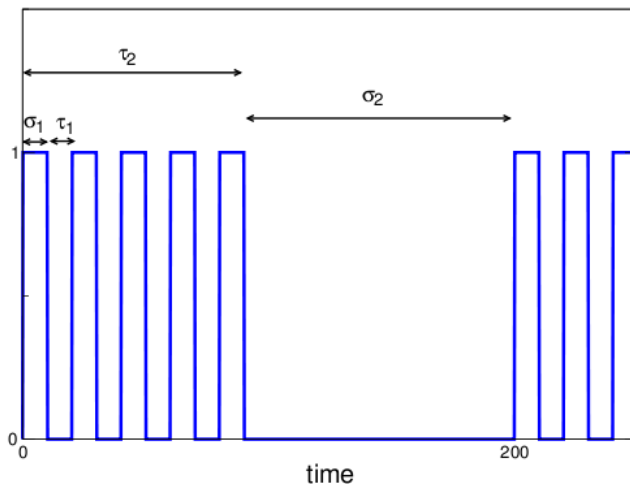


FIG. 2.11 – Signal type bursting composé de deux signaux d’échelles de temps différentes. Le premier signal est composé d’un train de pulses rapides avec une phase “on” notée  $\tau_1$  et une phase de repos entre les pulses de durée  $\sigma_1$ . Le deuxième signal correspond à l’échelle de temps lente (deux ordres de grandeur au-dessus). La durée de la phase “on” est notée  $\tau_2$  et celle de la phase “off” entre les pulses est notée  $\sigma_2$ . Pour des raisons pratiques, la figure n’est pas à échelle.

## 5 Réponse optimale d’un système soumis à un signal type bursting

Nous nous proposons dans cette section d’étudier une extension des modules présentés dans les sections précédentes. Nous pouvons combiner plusieurs réseaux ensemble dans des réseaux plus larges pour traiter des signaux plus complexes. Nous montrons dans cette section qu’en juxtaposant deux modules types “IFFL Diamond”, nous pouvons obtenir un réseau de régulation capable de répondre de manière sélective à un signal type “bursting”. Des oscillations sous forme de bursting sont des signaux qui font intervenir plusieurs échelles de temps qui sont par exemple émis par des cellules neuronales [69]. Le signal est formé par des trains de pulses qui alternent avec des périodes de repos<sup>2</sup>

Par conséquent, un signal de type bursting peut avoir un impact sur l’expression génétique notamment sur les cellules neuronales. Il est donc intéressant d’identifier quelles structures simples de réseaux de régulation pourraient traiter et sélectionner des signaux types bursting. Pour explorer cette question, nous considérons le réseau représenté sur la Figure 2.12 qui combine deux motifs DIFFL en série.

Chaque nœud de la Figure 2.12 est représenté par une seule espèce mais en pratique les noeuds correspondant aux protéines  $Y_1$  et  $Y_2$  dans le graphe peuvent être remplacés par des cycles de modification covalente comme montré sur la Figure

<sup>2</sup>Le signal électrique des neurones est une dépolarisation transitoire créée par la dynamique des canaux ioniques dans la membrane du neurone. Au niveau de la synapse, le signal de nature électrique est converti en un signal chimique qui affectent les récepteurs du neurone post-synaptique. Ensuite, par l’intermédiaire de réactions biochimiques en cascade, le signal peut être transmis jusqu’aux réseaux génétiques moléculaires présents dans le neurone.

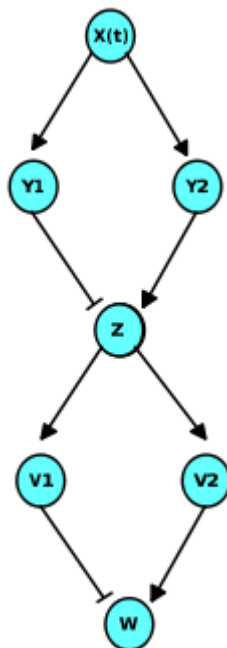


FIG. 2.12 – Réseau de régulation faisant intervenir deux motifs types “IFFL Diamond” en série.

2.9. Dans ce cas, le schéma donne un réseau de régulation où le signal d’entrée peut influencer l’expression du gène cible  $W$  par l’intermédiaire de voies de signalisation qui sont deux cycles de phosphorylation / déphosphorylation qui respectivement active et réprime le promoteur du gène intermédiaire  $Z$ . La protéine correspondante active le répresseur  $V_1$  et l’activateur  $V_2$  du gène cible  $W$ .

En employant des notations similaires à la section précédente, la dynamique du réseau représenté sur la Figure 2.12 peut être décrite par le système d’équations suivant :

$$\dot{Y}_1 = \beta_1 S(t) - \alpha_1 Y_1 \quad (2.22)$$

$$\dot{Y}_2 = \beta_1 S(t) - \alpha_1 Y_2 \quad (2.23)$$

$$\dot{Z} = \beta_1 H(Y_1 - \theta_1) H(\theta_2 - Y_2) - \alpha_1 Z \quad (2.24)$$

$$\dot{V}_1 = \beta_2 H(Z - \theta_3) - \alpha_2 V_1 \quad (2.25)$$

$$\dot{V}_2 = \beta_2 H(Z - \theta_4) - \alpha_2 V_2 \quad (2.26)$$

$$\dot{W} = \beta_2 H(V_1 - \theta_5) H(\theta_6 - V_2) - \alpha_2 W \quad (2.27)$$

Pour modéliser le signal d’entrée  $S(t)$ , nous considérons un signal temporel caractérisé par 4 constantes de temps  $(\tau_1, \sigma_1, \tau_2, \sigma_2)$  (voir la Figure 2.11). La période du signal est la somme  $T = \tau_2 + \sigma_2$ .  $\tau_2$  est la durée d’un train de pulses et  $\sigma_2$  est la période de repos. La phase du train de pulses est définie par une série de pulses carrés dans lesquels  $\tau_1$  est la durée du pulse et  $\sigma_1$  est l’intervalle entre les pulses. La période de repos qui sépare deux trains de pulses est  $\sigma_2$ . Donc le signal est caractérisé par deux différentes échelles de temps avec  $\tau_1 + \sigma_1 \ll \tau_2 + \sigma_2$ . Dans les

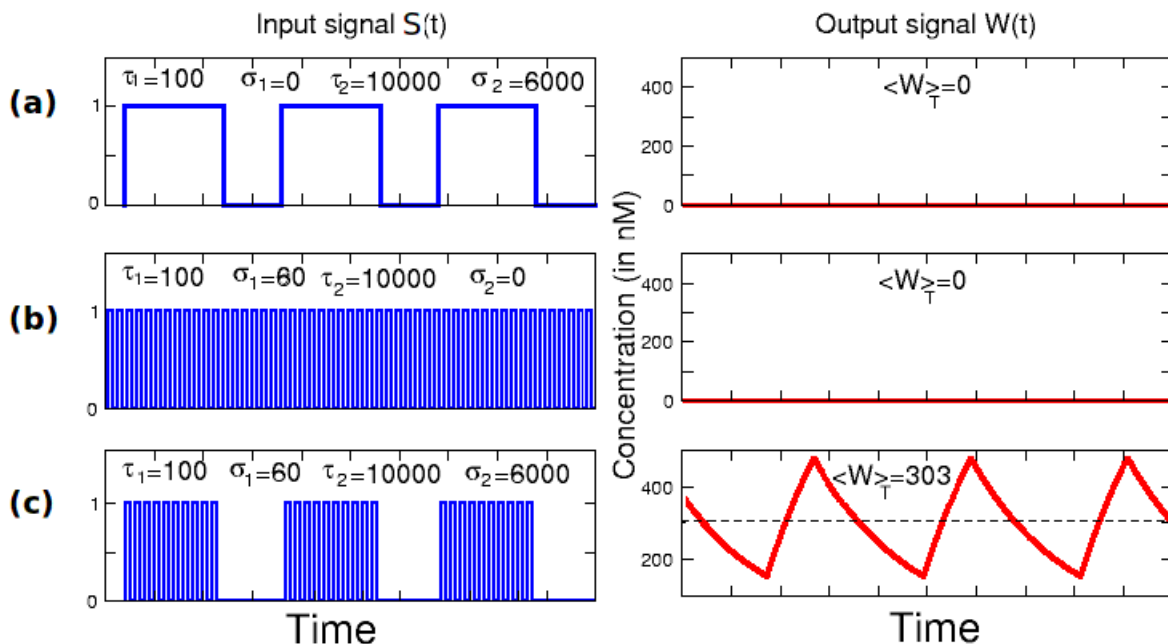


FIG. 2.13 – Le motif de double DIIF (Fig 2.12) est stimulé avec trois différents signaux d’entrée périodiques  $S(t)$  (sur la gauche). Les trois graphes de droite montre la réponse de la protéine de sortie  $W(t)$ . Les simulations numériques ont été réalisées avec les paramètres suivants :  $\tau_1 = 100$  s et  $\tau_2 = 10000$  s sont fixés. Les autres paramètres sont :  $\alpha_1 = 0.01 \text{ sec}^{-1}$ ,  $\beta_1 = 10 \text{ nM} \cdot \text{sec}^{-1}$ ,  $\alpha_2 = 0,006 \text{ min}^{-1}$ ,  $\beta_2 = 6 \text{ min}^{-1}$ ,  $\theta_1 = \theta_5 = 700 \text{ nM}$ ,  $\theta_2 = \theta_6 = 800 \text{ nM}$ ,  $\theta_3 = \theta_4 = 50 \text{ nM}$ . (a) avec  $\sigma_1 = 0$  s,  $\sigma_2 = 100 \text{ min}$  (b) avec  $\sigma_1 = 60$  s,  $\sigma_2 = 0 \text{ min}$  (c)  $\sigma_1 = 60$  s,  $\sigma_2 = 100 \text{ min}$ . Pour des raisons pratiques, les représentations des signaux bursting ne sont pas à l’échelle.

simulations reportées plus bas, ces échelles de temps sont séparées par deux ordres de grandeur. En effet, c’est typiquement l’ordre de différence entre les échelles de temps des dynamiques de cascades de signalisation et de régulation génétique (cf. Tableau 2.2). Quand le réseau représenté sur la Figure 2.12 est stimulé avec le signal  $S(t)$  décrit au-dessus, nous pouvons faire apparaître une propriété de traitement du signal. En effet, le gène de sortie  $Z$  ne s’exprime que selon un certain type de signal d’entrée. La Figure 2.13 illustre ce point : la réponse temporelle de  $W$  est montrée pour différents signaux de stimulation périodiques  $S(t)$ .

Lorsque le système est soumis à un train périodique de pulses carrés sans bursting (Figure 2.13 (a),  $\sigma_1 = 0$ ), le gène  $W$  n’est pas exprimé. De même si la stimulation consiste en un long train de pulses sans période de repos (Figure 2.13 (b),  $\sigma_2 = 0$ ), le gène  $W$  ne s’exprime pas non plus. Cependant, si l’on stimule le motif avec un signal type bursting particulier (Figure 2.13 (c)), le système répond de manière significative. Le réseau présenté Figure 2.12 permet donc de ne réagir que pour un type de bursting particulier. Si les intervalles de temps  $\tau_1$  et  $\tau_2$  sont fixés, le système possède des valeurs pour  $\sigma_1$  et  $\sigma_2$  pour lesquelles, la réponse moyenne de la protéine de sortie  $W$  est maximale. En effet, en regardant la Figure 2.13 (b), le système se comporte comme s’il filtrait les hautes fréquences. En revanche, lorsque l’on introduit des périodes de repos entre les trains de pulses sous la forme d’un signal bursting alors le système peut répondre de façon significative et la concentration de

la protéine  $W$  n'est pas nulle.

La raison pour laquelle le réseau représenté sur la Figure 2.12 peut sélectionner un signal type bursting est une conséquence directe des principes présentés dans les sections précédentes. Nous fixons  $\tau_2$  la phase active du signal lent et  $\sigma_2$  est l'intervalle de repos entre les trains de pulses. Durant cette phase "on", les pulses ont une durée fixée à  $\tau_1$ . Nous avons donc une réponse optimale pour la protéine intermédiaire  $Z$  si l'intervalle entre les pulses  $\sigma_1$  est choisi pour pouvoir déclencher une réponse dans le premier sous-réseau DIFFL. Par exemple, en utilisant les paramètres de la Figure 2.10, pour avoir une réponse maximale dans la concentration moyenne de  $Z$ , nous choisissons  $\sigma_1 \simeq 1/(4\alpha_1)$ . Ensuite durant la période de repos  $\sigma_2$ , l'expression de  $Z$  retourne à 0. Donc l'évolution temporelle de la protéine intermédiaire  $Z$  est reliée au signal rapide dans les trains de pulses caractérisé par les durées  $(\tau_2, \sigma_2)$ . La variable  $Z$  sert ensuite de stimulation pour le second sous-réseau de DIFFL composé des noeuds  $V_1, V_2$  et  $W$ . Pour un  $\tau_2$  fixé, le second sous-réseau va également répondre de façon optimale selon que la durée de repos  $\sigma_2$  a une durée précise. Par exemple, en utilisant les mêmes paramètres que la Figure 2.10, nous avons  $\sigma_2$  qui doit être égal à  $1/(4\alpha_2)$  pour avoir une réponse moyenne de  $W$  la plus importante. Donc, un signal temporel qui maximise la protéine de sortie est celui composé des durées  $\sigma_1 \ll \sigma_2$  c'est-à-dire un signal type bursting. Le premier sous-réseau de réponse rapide, pourrait être composé de cascades de signalisation, le second sous-réseau plus lent, pourrait faire intervenir des interactions génétiques.

## 5.1 Exemples d'application

Nous pensons que l'étude fournie dans les sections précédentes sur la réponse de l'Incoherent Feed Forward permet d'éclaircir les principes derrière le phénomène remarqué dans le travail de Smolen et al [158]. Dans cette référence, les auteurs proposent un modèle de réseau génétique pour la formation de la mémoire à long terme qui est basée sur le phénomène de potentiation à long terme et le renforcement synaptique des cellules neuronales. La potentiation à long terme apparaît dans une synapse par exemple lorsque la cellule est soumise à un signal externe comme des pulses de glutamate de forte amplitude. Ensuite le signal induit la phosphorylation de certains facteurs de transcription qui activent la transcription pour la synthèse de nouvelles protéines qui vont renforcer des connexions synaptiques déjà existantes ou en créer de nouvelles. Il a été montré que pour apprendre de manière efficace, la plupart des organismes, du simple mollusque aux mammifères supérieurs doivent recevoir des stimulations répétées (pour une revue sur l'organisme modèle de l'aplysie on pourra consulter [73]). Les sessions d'apprentissage doivent être espacées d'une durée minimale pour que l'apprentissage soit optimal. Un apprentissage où les stimuli seraient trop serrés ne peut produire une mémoire à long terme comme montré expérimentalement chez l'aplysie [22]. Donc le réseau génétique sous-jacent se met en route uniquement si les sessions d'apprentissage sont suffisamment espacées.

Le modèle de Smolen et al. (Figure 5 de [158]) a une structure équivalente au motif de DIFFL représenté Figure 2.9. L'analyse détaillée dans ce chapitre de ce motif nous permet de clarifier le phénomène observé par ces auteurs. L'IFFL nous paraît le réseau le plus simple pour avoir une telle maximisation et appréhender le phénomène. Les ingrédients nécessaires sont effectivement une compétition entre



deux processus de régulation. Pour que le mécanisme fonctionne, il faut également une différence de sensibilité au signal stimulateur par les deux processus qui entrent en compétition. Plus précisément, il est nécessaire que la branche liée à l'activation (celle reliant directement les noeuds  $X$  et  $Z$  dans l'IFFL) soit plus "efficace" pour transmettre le signal venant de  $X$  que la branche du réseau qui fait la répression (celle qui relie les noeuds  $X - Y - Z$  dans l'IFFL). Lorsque l'on va stimuler le système avec des périodes de repos de plus en plus longues, le signal d'entrée va être de moins en moins important. La branche de la répression va commencer à ne plus agir en premier et le système va commencer à synthétiser la protéine de sortie. Cette différence d'efficacité peut être obtenue par exemple en jouant sur les différentes constantes d'affinité des régulateurs qui entrent en jeu mais également en ajoutant des noeuds au réseau. Dans l'IFFL, le noeud  $Y$  permet en quelque sorte de ralentir la transmission du signal dans la branche de répression et donc il rend la transmission du signal par cette branche plus difficile. Si  $X$  et  $Y$  ont la même affinité pour le promoteur, le processus peut donc toujours marcher étant donné que la branche assurant la répression aura toujours un désavantage par rapport à la branche de l'activation. C'est pour cette raison également, que le fait d'ajouter un quatrième noeud dans le motif de DIFFL a permis une régulation plus fine. En effet, dans ce motif, les deux branches d'activation et de répression peuvent être à égalité dans l'efficacité de la transmission du signal. Il suffit seulement de jouer sur les constantes d'affinité des régulateurs pour avoir un système qui ne répond que pour une plage d'intervalles entre les pulses précise.

Les résultats de Li et Goldbeter mentionnés dans la section 1 [93] sont également l'un des rares exemples trouvés dans la littérature où il apparaît un processus de maximisation en réponse à un signal périodique. Bien que cette étude soit basée sur des calculs analytiques, il est difficile de trouver les ingrédients qui sont à l'origine du phénomène observé. Il serait intéressant de rechercher une généralisation du principe d'optimisation qui englobe ces différents modèles.

Le tableau 2.2 donne des valeurs typiques de paramètres que l'on peut rencontrer dans différents contextes (génétique ou de signalisation). Des estimations pour la durée d'un pulse d'activation ainsi que pour la valeur de l'intervalle entre les pulses qui donnerait une réponse optimale sont données. Les paramètres cinétiques donnés sont réalistes puisqu'issus de travaux expérimentaux, les valeurs calculées analytiquement pour  $\sigma_1$  donnent un ordre de grandeur qu'il faudrait envisager pour une implémentation expérimentale. Par exemple, plusieurs motifs d'IFFL ont été identifiés dans le génome d'*Escherichia coli*, il nous semblerait intéressant de stimuler l'un d'entre eux par un signal type créneau et voir si le principe d'optimisation de la réponse est faisable en pratique. L'utilisation d'un promoteur thermosensible nous semble une possibilité abordable pour stimuler un gène périodiquement. Une alternance de températures peut être réalisée sans trop de difficulté ce qui permettrait d'avoir un promoteur dont l'expression subit une alternance de phases "on" et phases "off".

## 6 Résonance simple

Les motifs que nous avons étudiés dans les sections précédentes donnaient un maximum dans la concentration *moyenne* des protéines de sortie pour un type de

Système				$\sigma$	$\tau$
Simple IFFL chez les bactéries [144]	$\beta = 100 \text{ nM} \cdot \text{min}^{-1}$	$k' = 1 \text{ s}^{-1}$	$\theta = 500 \text{ nM}$	1 to 10 min	25 to 390 min
IFFL avec cycles de signalisation [77]	$k' = 1 \text{ s}^{-1}$	$k = 1 \text{ s}^{-1}$	$\theta = 500 \text{ nM}$	1 s	2 s

TAB. 2.2 – Tableau donnant des échelles de temps qui pourraient entrer en jeu pour le processus d’optimisation de réponse moyenne pour un motif type IFFL.

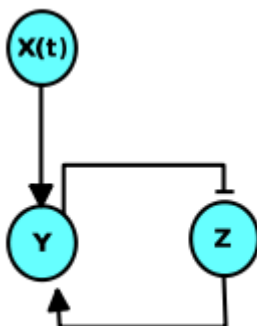


FIG. 2.14 – Le motif d’une boucle rétro-négative. Y active Z qui réprime Y, Y est stimulé par un signal extérieur périodique  $X(t)$ .

signal particulier. Il s’avère également intéressant d’un point de vue physiologique de trouver des réseaux de régulation qui pour un signal temporel donné, maximisent l’amplitude de réponse dans la concentration des protéines de sortie. C’est pourquoi nous nous proposons dans cette section d’étudier un réseau génétique simple et réaliste qui peut donner une *résonance* au sens physique. Ce phénomène est très bien connu et trouve des applications dans de nombreuses branches de la physique (mécanique, acoustique, etc).

## 6.1 Motif de boucle rétro-négative simple

Nous considérons dans cette section, le réseau présenté sur la Figure 2.14, il se compose de deux gènes/protéines Y et Z qui interagissent en une formant boucle rétro-négative. Le troisième noeud est un signal externe  $X(t)$  périodique qui active le gène Y.

Nous pouvons remarquer que ce type de réseau peut ne pas être seulement génétique mais faire intervenir des interactions de différente nature. Par exemple, le contrôle négatif peut être exercé par une régulation post-transcriptionnelle [58] (la protéine Y inactive la protéine Z en se dimérisant à elle). Par ailleurs, si la protéine Z, en plus d’activer Y, s’autoactive, le système peut se comporter comme un oscillateur biologique. Notons que cette auto-régulation positive est essentielle pour l’existence

Modèle	Cycle Cellulaire [169]	P53 [28]	Système synthétique [10]	Rythmes circadiens (Neurospora) [33]
Y (Répresseur)	CLB	Mdm2	LacI	FRQ
Z (Activateur)	CLN	P53	NRI	WC-1,WC-2

TAB. 2.3 – Tableau donnant des exemples de systèmes faisant intervenir une boucle rétro-négative à la base d’oscillateurs biologiques.

d’oscillations auto-entretenues [52] (à moins que la boucle rétro-négative possède un mécanisme de retard, cf. Chapitre 1). Ce réseau se retrouve en effet à la base de plusieurs oscillateurs connus (voir tableau 2.3). Par exemple, il a été utilisé pour des modèles simples de rythmes circadiens [33] ou pour étudier le cycle cellulaire [169] ou encore pour étudier la dynamique des oscillations de la protéine P53 [28]. Ce motif a également été étudié dans une perspective de biologie synthétique [10]. D’autres exemples existent comme le modèle de Song et al. qui considère également une boucle rétro-négative pour un modèle de formation de la mémoire [160].

Dans la configuration *sans* la boucle rétronégative, il apparaît néanmoins des propriétés intéressantes concernant la réponse à une stimulation périodique. Nous considérons donc ce réseau soumis à une stimulation périodique représentée par la variable  $X$ . Cette activation périodique peut être causée par exemple par le couplage du réseau avec un oscillateur autonome. Pour modéliser ce réseau génétique, nous utilisons les équations suivantes :

$$\begin{cases} \dot{Y} = (\beta_0 + \beta_1 S(t)) \frac{Z^n}{\theta^n + Z^n} - \alpha Y \\ \dot{Z} = \beta \frac{\theta^n}{\theta^n + Y^n} - \alpha Z \end{cases} \quad (2.28)$$

Nous utilisons dans cette section les fonctions de Hill pour rendre compte des interactions entre les gènes. Le phénomène de résonance est beaucoup clair avec un signal harmonique qu’avec un signal créneau, nous considérons donc dans la suite la réponse à un signal harmonique type sinusoïdal :

$$S(t) = \sin(\omega t) \quad (2.29)$$

Lorsque la stimulation périodique est absente c’est à dire  $\beta_1 = 0$ , le système possède un état stationnaire  $(Y_S, Z_S)$ . Il peut être montré que c’est toujours un foyer stable. La fréquence correspondante  $\omega_0$  peut être déterminée en calculant la partie imaginaire des valeurs propres du jacobien du système. Si le système est éloigné de son point équilibre, il y retournera avec des oscillations transitoires de fréquence  $\omega_0$ . Cette fréquence qui est propre au système et à la valeur de ses différents paramètres, peut être révélée en modulant l’expression de  $Y$ . En effet si nous stimulons le réseau représenté Figure 2.14 avec différentes fréquences, il répondra avec une plus forte amplitude pour la fréquence  $\omega_0$ . Des simulations numériques des équations 2.28 montrent clairement le phénomène de résonance qui pourrait apparaître avec des valeurs numériques plausibles pour les différents paramètres (voir Figure 2.15). Nous obtenons avec ces paramètres une valeur typique  $T_0 \sim 6$  heures.

Par ailleurs, si l’on change la coopérativité  $n$  qui entre en jeu dans les interactions, on peut voir que la résonance se fait de façon plus fine c’est à dire que la

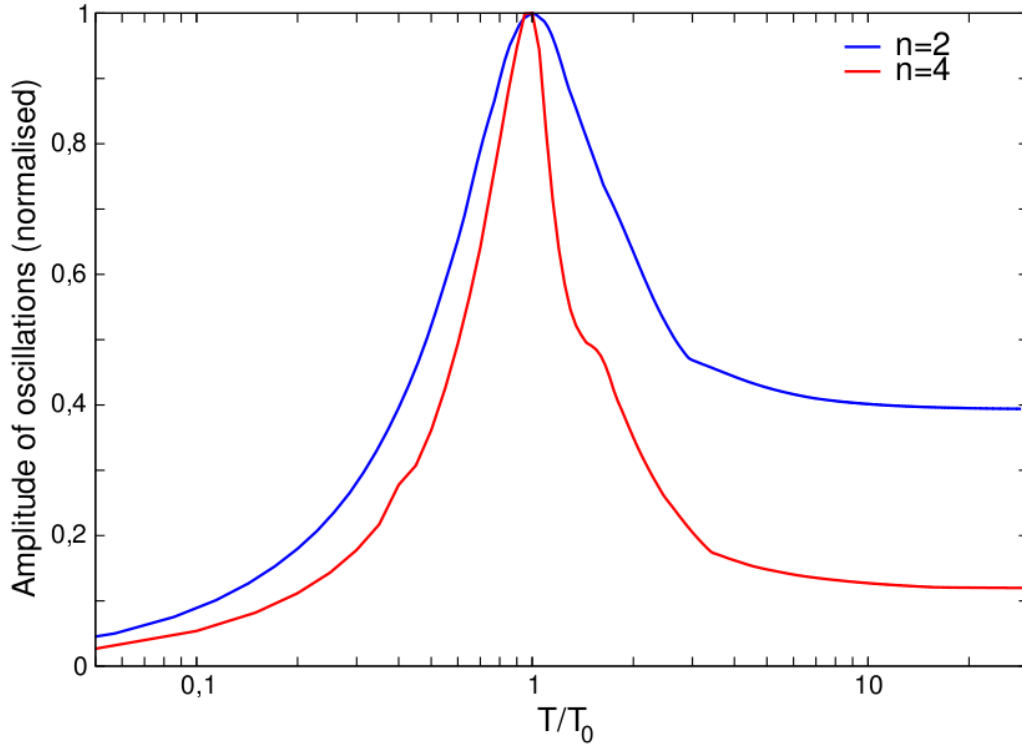


FIG. 2.15 – Réponse en *amplitude* du motif boucle de régulation rétro-négative présenté sur la Figure 2.14 à une stimulation harmonique en fonction de la période de stimulation  $T$ . La simulation numérique a été réalisée avec les équations 2.28 et les paramètres suivants :  $\beta_0 = 150 \text{ nM.min}^{-1}$ ,  $\beta_1 = 100 \text{ nM.min}^{-1}$ ,  $\beta_2 = 100 \text{ nM.min}^{-1}$ ,  $\theta_1 = 50 \text{ nM}$ ,  $\theta_2 = 50 \text{ nM}$ ,  $\alpha = 0.01 \text{ min}^{-1}$ ,  $n = 2$  ou  $4$ . Avec ces paramètres, nous obtenons  $T_0 \sim 6$  heures.

plage de fréquences qui font réagir plus fortement le système est plus étroite. Pour expliciter cette propriété, nous pouvons nous placer dans une situation simple pour pouvoir faire des calculs analytiques : nous considérons des valeurs identiques pour les paramètres de seuil qui entrent dans les fonctions de régulation et nous supposons aussi que  $\beta = \beta_0$ . L'état stationnaire pour  $Y$  et  $Z$  est symétrique c'est à dire  $Y_S = Z_S = \frac{\beta}{2\alpha}$  et l'expression de la fréquence de résonance est  $\omega_0 = n\alpha/2$  où  $n$  est le coefficient de coopérativité. Puisqu'en absence de stimulation périodique, l'amortissement des oscillations transitoires est caractérisé par le coefficient de dégradation  $\alpha$ , nous pouvons estimer le *facteur de qualité* de la résonance par le rapport  $\omega_0/\alpha = n/2$ . Il ne dépend uniquement du coefficient de coopérativité. Nous donnons sur la Figure 2.15, deux exemples de résonance avec deux coefficients  $n=2$  et  $n=4$ .

La qualité de la résonance et donc sa capacité à répondre précisément à une fréquence donnée ne dépend, pour ce système simple, que de la coopérativité des protéines régulatrices.

## 6.2 Exemples d'applications possibles

Nous avons donc remarqué dans cette section que le petit motif de boucle rétro-négative présente la propriété intéressante de maximiser son amplitude de réponse

selon la fréquence d'un signal stimulateur. Sans la stimulation périodique, le système converge vers un état stationnaire mais il possède cependant la propriété de résonance qui peut être "réveillée" lorsqu'il est stimulé périodiquement. Nous pensons que ce phénomène pourrait entrer en jeu dans plusieurs situations. Un premier exemple serait celui de la protéine P53 qui est engagée dans de nombreux processus de régulation notamment liés à l'intégrité du génome (elle est souvent appelée "gardienne du génome"). Dans certaines conditions, la concentration de cette protéine oscille avec une période de 5.5 heures. Dans le travail expérimental de Geva-Zatorsky et al. [45], les auteurs montrent que l'amplitude des oscillations présente une forte variabilité alors que la période des oscillations présente une étonnante régularité, ce qui suggérerait son importance dans la régulation. La protéine P53 est un facteur de transcription qui agit sur plusieurs gènes. L'étude réalisée dans cette section permet de spéculer que le comportement oscillant de P53 pourrait déclencher des résonances dans certains réseaux génétiques impliqués. Cette remarque peut s'appliquer pour d'autres systèmes de régulation liés aux réponses à des stress comme le système du régulateur LexA d'*Escherichia coli* ou pour le système NF- $\kappa$ B.

## 7 Perspectives de ce travail sur le plan théorique

Nous proposons dans cette section des perspectives à ce travail sur le plan théorique.

- Une poursuite possible serait de voir si les mécanismes présentés ici résistent à l'aspect stochastique inhérent à ces processus. Comme évoqué dans le chapitre 1, les régulations transcriptionnelles font souvent intervenir un faible nombre de protéines régulatrices ce qui confère un aspect aléatoire à ces processus. En utilisant notamment l'algorithme de Gillespie [46] (cf. Chapitre 1), il serait possible d'appréhender si les mécanismes de sélectivité du signal par rapport à son contenu temporel sont toujours possibles. Est-ce qu'un signal s'il est de nature périodique, est-il plus facilement transmissible à travers les réseaux de régulation et plus robuste au bruit qu'un simple signal continu ? Par ailleurs, une approche stochastique permettrait également d'explorer la question de l'utilisation de *résonance stochastique* dans ces réseaux de régulation. A notre connaissance, peu de travaux ont exploré cette question. Nous pouvons cependant citer le travail de Wang et al. [174] qui montre que le phénomène de résonance stochastique pourrait être utilisé et qu'une certaine valeur du bruit (qui correspond en fait au nombre de molécules) pourrait aider le signal à être transmis.
- Nous avons répertorié dans ce travail plusieurs modèles issus de différents travaux théoriques et de modélisation qui présentent un maximum en réponse à un signal périodique. Nous avons remarqué qu'un dénominateur commun de ces systèmes est qu'ils répondent en faisant un pulse lorsqu'ils sont soumis à un signal d'activation continu en forme d'échelon, (c'est le cas de l'IFFL ou du modèle de Li et Golbeter par exemple). Il est tentant de conjecturer que de façon générale, tout système qui répond de cette manière possède la propriété d'optimiser sa réponse moyenne selon un type de signal temporel périodique. Il serait intéressant de le démontrer rigoureusement.
- Enfin, nous n'avons considéré dans ce travail que des signaux qui avaient une

seule périodicité. Il serait également intéressant d'étudier si des mécanismes d'optimisation de réponse des réseaux de régulation peuvent être obtenus pour des signaux plus complexes qui possèdent plusieurs périodes ou même qui sont a périodiques.

## 8 Conclusion

Les mécanismes présentés dans ce chapitre identifient des stratégies possibles qui peuvent être employées par une cellule ou une bactérie pour intégrer des signaux dépendant du temps provenant notamment de l'environnement. Les motifs étudiés offrent la capacité intéressante de sélectionner un signal selon sa structure temporelle ou son contenu fréquentiel. Pour l'Incoherent Feed Forward Loop, seulement les trains de pulses avec une durée minimale entre les pulses d'activation peuvent être transmis. Nous pensons qu'un tel phénomène peut se produire soit dans un réseau génétique existant d'un organisme vivant ou bien qu'il pourrait être construit de manière synthétique grâce aux techniques de biologie moléculaire de construction génétique [38, 43]. La difficulté majeure sur le plan expérimental nous semble de produire la stimulation périodique. Certaines techniques récentes sont cependant prometteuses pour reproduire de tels signaux.

Comme abordé dans le chapitre 1, une question concernant le concept de motifs comme l'IFFL est de savoir si ces réseaux agissent de façon indépendante étant donné que dans un génome comme celui d'*Escherichia coli*, ils sont en fait imbriqués dans des réseaux beaucoup plus larges et reliés à d'autres interactions. Le fait d'avoir des réseaux qui ne répondent qu'à un signal au profil temporel particulier donnerait la possibilité intéressante de mettre en marche un ou des réseaux de façon précise au sein des myriades d'interactions d'un génome.

Enfin, de façon plus générale, les principes d'optimisation présentés ici peuvent être pensés dans d'autres domaines. Certains traitements médicaux font intervenir une périodicité : la prise à intervalle régulier de médicaments est un exemple ou bien le traitement de certaines pathologies neurologiques font intervenir des stimulations électriques dans le cerveau également à intervalle entre les décharges électriques précis. Dans ces deux cas, la durée de l'intervalle a son importance pour l'efficacité du traitement, et une recherche de son optimisation présente un grand intérêt.

Nous avons présenté ce travail dans un article publié dans la revue BMC Systems Biology, (revue librement accessible), nous reproduisons cette publication ci-après.

Research article

Open Access

## Simple molecular networks that respond optimally to time-periodic stimulation

Axel Cournac<sup>1,2</sup> and Jacques-Alexandre Sepulchre\*<sup>1</sup>

Address: <sup>1</sup>Institut Non Linéaire de Nice, Université de Nice Sophia-Antipolis, CNRS, Valbonne, France and <sup>2</sup>Laboratoire Matière et Systèmes Complexes, Université Diderot, Paris, France

Email: Axel Cournac - axel.cournac@inln.cnrs.fr; Jacques-Alexandre Sepulchre\* - jacques-alexandre.sepulchre@inln.cnrs.fr

\* Corresponding author

Published: 3 March 2009

Received: 11 August 2008

*BMC Systems Biology* 2009, **3**:29 doi:10.1186/1752-0509-3-29

Accepted: 3 March 2009

This article is available from: <http://www.biomedcentral.com/1752-0509/3/29>

© 2009 Cournac and Sepulchre; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Bacteria or cells receive many signals from their environment and from other organisms. In order to process this large amount of information, Systems Biology shows that a central role is played by regulatory networks composed of genes and proteins. The objective of this paper is to present and to discuss simple regulatory network motifs having the property to maximize their responses under time-periodic stimulations. In elucidating the mechanisms underlying these responses through simple networks the goal is to pinpoint general principles which optimize the oscillatory responses of molecular networks.

**Results:** We took a look at basic network motifs studied in the literature such as the Incoherent Feedforward Loop (IFFL) or the interlocked negative feedback loop. The former is also generalized to a diamond pattern, with network components being either purely genetic or combining genetic and signaling pathways. Using standard mathematics and numerical simulations, we explain the types of responses exhibited by the IFFL with respect to a train of periodic pulses. We show that this system has a non-vanishing response only if the inter-pulse interval is above a threshold. A slight generalisation of the IFFL (the diamond) is shown to work as an ideal pass-band filter. We next show a mechanism by which average of oscillatory response can be maximized by bursting temporal patterns. Finally we study the interlocked negative feedback loop, i.e. a 2-gene motif forming a loop where the nodes respectively activate and repress each other, and show situations where this system possesses a resonance under periodic stimulation.

**Conclusion:** We present several simple motif designs of molecular networks producing optimal output in response to periodic stimulations of the system. The identified mechanisms are simple and based on known network motifs in the literature, so that that they could be embodied in existing organisms, or easily implementable by means of synthetic biology. Moreover we show that these designs can be studied in different contexts of molecular biology, as for example in genetic networks or in signaling pathways.

### Background

For past several decades, the experimental data concerning protein concentrations in biological regulatory net-

works have mostly been obtained in conditions related to stationary states. In some natural contexts, however, or in some experiments, the relevant cellular response to meas-

ure is the one the cell displays when it is exposed to time-dependent signals. Such information can be essential to unravelling the regulatory principles of those molecular networks which by nature are repeatedly stimulated by time-varying inputs [1]. For example the gene expression within neuronal cells can be substantially affected by the time-dependent signals received from its afferent neurons, and this property is essential for the formation of memories [2]. Other examples can be met with in various contexts, like the frequency encoding phenomenon associated with  $\text{Ca}^{2+}$  oscillations [3-6], or the response to time-dependent osmolarity shocks [7]. We note that, in these systems an optimal-response may exist with respect to periodic stimulations, meaning that for example, the mean production of some activated transcription factors, or of some proteins of interest, would be maximized if the external periodic signal follows some specific time-course and shape. This question has received little attention in the context of biological networks. This research topic is nevertheless timely, as recent developments of experimental techniques in molecular biology enables one to access time-dependent concentrations, and investigate new problems about the time-response of molecular networks.

From an experimental point of view, some interesting experiments were recently performed to reveal such time-response properties in biological systems [7-9]. In order to anticipate more experimental studies in this direction, the objective of this paper is to identify basic network topologies which allow for the property of optimal oscillatory responses in molecular biology. In the current literature only a small number of studies have tackled this issue and most of the articles addressing this question principally considers the linear response to sinusoidal perturbations. For example several papers demonstrate the low-pass filter property of small molecular networks [10-13]. Nearly all of these studies consider small amplitude perturbations, and use Fourier transform, discussing how the system parameters modify the cut-off frequency. As a matter of fact, when a system behaves like a low-pass filter, the frequency which maximizes the output response is 0. So in this case the system does not display a genuine temporal specificity. Very few studies have looked for the possibility of maximizing some variables at non-zero frequency in biological networks. A recent study [14], using again the tools of linearization and of Fourier analysis, shows that a signaling cascade with a negative feedback can behave as a band-pass filter, characterized by a frequency which maximizes the amplitude of the response. But when the system's nonlinearity is important, as is often the case in models of transcriptional regulatory networks, or when the system is linear but the periodic forcing is multiplicative (thus not additive), then Fourier decomposition of the input signal is no longer of general interest. Furthermore, in many instances, the time-dependent signals

directed to regulatory networks are far from harmonic, they are often pulsatile. This is why a time-periodic input to molecular networks is often assumed as a periodic train of pulses. Smolen et al. studied a molecular network which was periodically stimulated by a train of spikes and which exhibited the ability to respond in a maximal fashion to a particular inter-spike interval [15]. Their model equations were nonlinear and associated with a lot of parameters, so the origin of the bandpass filter property is not straightforward. Another instance of periodic stimulations with a train of square pulses was thoroughly studied by Li and Goldbeter in the context of cellular receptor systems [16]. By using analytical calculations and numerical simulations of a linear model [17], they show that can exist an optimum stimulus pattern of periodic square pulses which maximizes a lumped observable. This is defined by the authors as the receptor activity. However, the Li and Goldbeter model has again many parameters, and the identification of the essential ingredients which enable one to induce a maximum amount of receptor activity is not obvious.

In the present paper we study simpler models, which pertain to basic network motifs found in biological regulations. We point out some principles which guarantee the existence of an optimal response in the production rates of output molecules, when these systems are activated by a periodic signal. The identified mechanisms are based on known network motifs in the literature and include, for example, the negative feedback loop and the incoherent feedforward loop (IFFL) motif studied by Alon and coworkers in the context of gene regulatory networks of bacteria [18]. Let us recall that in this framework a network motif is a small pattern of molecular interactions which recurs repeatedly in comparison with what would occur in random networks. Here we will make use of this concept of motif in a broader sense, without requiring that its ubiquity be statistically proved, but requesting that it represents a basic network pattern associated with specific information-processing properties [19]. Moreover we will study these designs in at least two different contexts of molecular biology, namely in genetic regulatory networks or in signaling pathways. In the context of genetic regulation, the mathematical models are usually strongly nonlinear, involving sigmoidal regulation functions. In order to use analytical tools to study the behavior of these systems under time-dependent signals, we will employ the "logic" approximation, where regulation functions are replaced by all-or-none functions. In this case the nonlinear system is not approximated by a set of linear equations but by a set of piecewise-linear equations [20]. This approach has indeed been developed for many years in the study of genetic networks and has revealed to be quite fruitful [21].



## Results

### Simple models of biomolecule activations

In this Section, we give a brief summary of the simplest models used in the literature to describe activation processes, either at the gene level, or at the level of post-translational modification of proteins. We then report elementary but useful results which are obtained when the activation is implemented by means of a train of periodic square pulses.

Transcriptional regulatory networks can be described with genes being represented by letters  $X, Y, \dots$ . The same notation is used for the concentrations of the corresponding proteins synthesized by these genes. We first introduce what is called the "simple regulation" by Alon [18], a situation which is depicted in Fig. 1(a) by an arrow from  $X$  to  $Y$ . In this case the gene  $Y$  is positively regulated by only one transcription factor  $X$ , and the basic process of synthesis of protein  $Y$  from its gene is described by a single differential equation:

$$\dot{Y} = \beta R(X) - \alpha Y \quad (1)$$

where  $R(X)$  is called the regulatory function. We consider regulatory functions bounded by 1. Typically  $R(X)$  has the form of a Hill function  $X^n/(X^n + \theta^n)$  with some cooperativity  $n$  and an activation threshold  $\theta$  but we will often use the *logic* approximation for  $R(X)$ , where the latter is replaced by the unit step function  $H(X - \theta)$  (with only two values  $H(X) = 0$  if  $X < \theta$ , and  $H(X) = 1$  otherwise). The parameters  $\beta$  and  $\alpha$  are respectively the maximum synthesis rate of protein  $Y$  and the degradation parameter, including the possible dilution effect from cell growth. Whenever  $R(X) = 1$  the system converges to a steady state  $\kappa = \beta/\alpha$ . In the sequel  $X(t)$  is considered as a function of time, and one defines  $S(t) = R(X(t))$ . A class of signals  $S(t)$  which will be considered below is a periodic train of square pulses of amplitude 1, whose temporal pattern is characterized by the numbers  $(\tau, \sigma)$  (cf. Fig. 2). The parameter  $\tau$  describes the duration of the "on-phase" corresponding to the activation of transcription factor  $X$  binding gene  $Y$ . The inter-pulse interval, or the silent phase between the pulses, is denoted by  $\sigma$ . Thus the period of  $S(t)$  is given by  $T = \tau + \sigma$ .

In response to the input signal  $S(t)$ , we will focus on three observables in this system, namely the extrema (minimum and maximum), and the average concentrations reached by  $Y(t)$ , denoted respectively by  $Y_{\min}$ ,  $Y_{\max}$  and  $\langle Y \rangle_T$ . The extreme values of  $Y(t)$  can be the most relevant quantities in systems where  $Y$  is itself a transcription factor for other genes, because then what matters is the comparison of  $Y$  with a threshold of activation or of repression. On the other hand, if  $Y$  is consumed by some downstream process, then the mean concentration of  $Y(t)$  averaged

over a period  $T$  of  $S(t)$ , denoted by  $\langle Y \rangle_T$ , is the relevant quantity since it quantifies the available protein synthesized over one period.

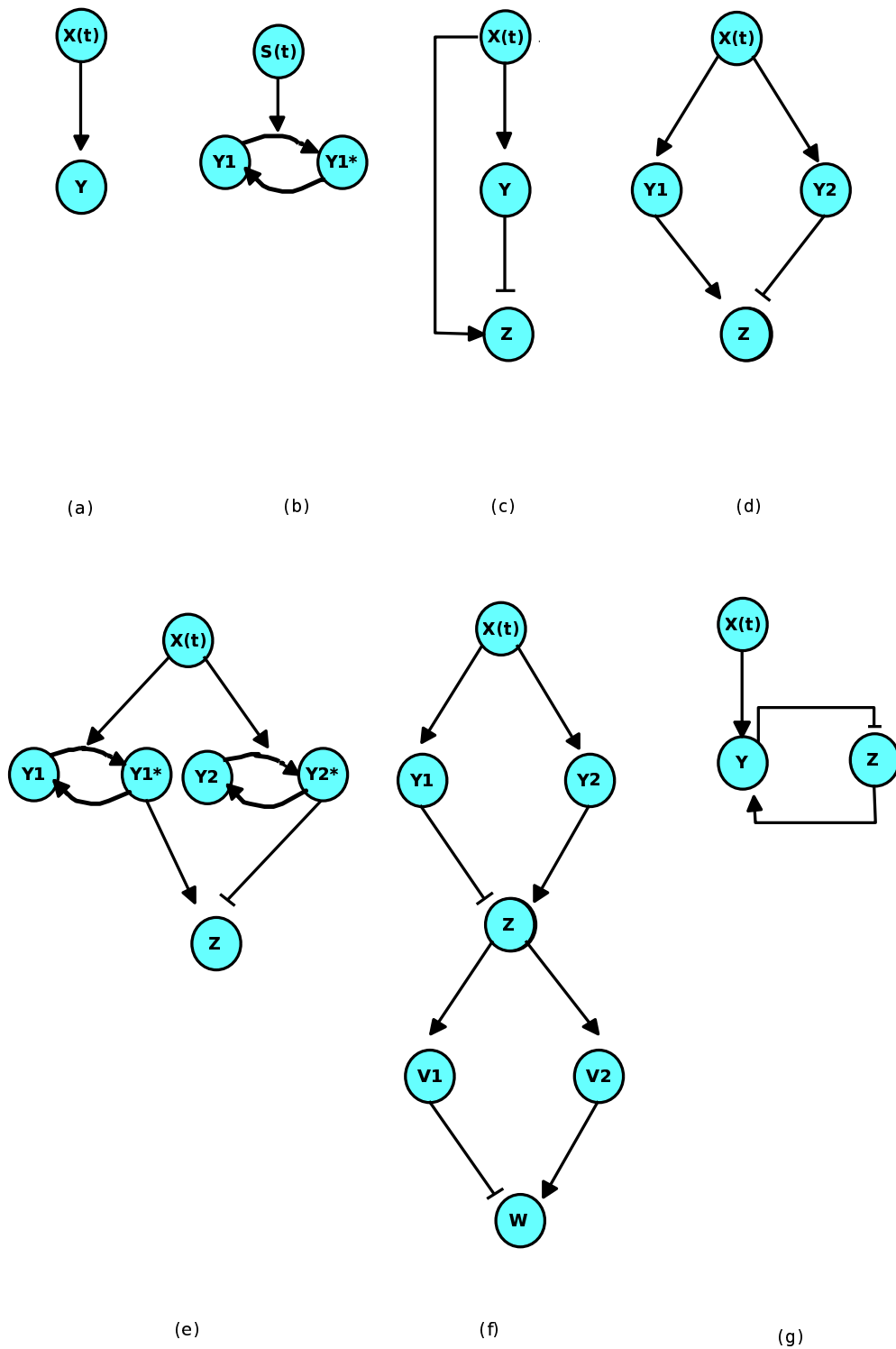
In the case of a simple regulation with periodic square pulses, the model equation (1) is easily worked out and from that calculation one deduces the observables  $\langle Y \rangle_T$ ,  $Y_{\max}$  and  $Y_{\min}$  as:

$$\begin{aligned} \langle Y \rangle_T &= \kappa \frac{\tau}{\tau + \sigma} \\ Y_{\max} &= \kappa \frac{1 - e^{-\alpha\tau}}{1 - e^{-\alpha(\tau + \sigma)}} \\ Y_{\min} &= \kappa \frac{e^{\alpha\tau} - 1}{e^{\alpha(\tau + \sigma)} - 1} \end{aligned} \quad (2)$$

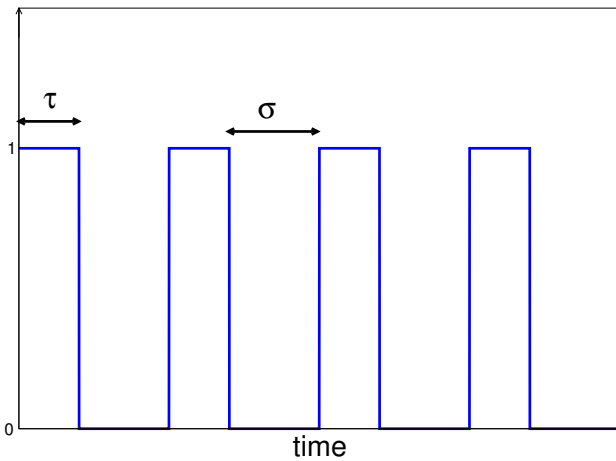
Fig. 3 shows an example of these functions when the inter-pulse  $\sigma$  is varied, for a fixed pulse duration  $\tau$ . When  $\sigma$  increases from 0 the observables are all decreasing. The maximum  $Y_{\max}$  stays within the interval  $[Y_L, \kappa]$ , where  $Y_L$  is the asymptotic value reached by  $Y_{\max}$  (dotted line on Fig. 3, obtained when the denominator of  $Y_{\max}$ , eq.(2), equals 1). We note that this level can be controlled by choosing the pulse duration  $\tau$ . This can be useful if  $Y$  is itself a transcription factor with respect to a target gene  $Z$  (see next Section).

Another important class of basic activation in molecular biology is the covalent modification of proteins (e.g. the phosphorylation/dephosphorylation), which constitutes the building block of signaling transduction pathways. In some cases, the enzyme promoting the activated form of the protein is submitted to time-periodic variations. Outstanding examples of this situation have been studied in the context of  $Ca^{2+}$  oscillations, whose temporal spiking patterns have been showed to significantly modulate the activity of CaM kinase II [3-6]. The activation/deactivation motif is usually represented as in Fig. 1(b) by drawing two arrows forming a cycle between two molecular states. A third arrow indicates the presence of the external signal  $S(t)$ . Here  $Y$  stands for the concentration of deactivated proteins, and  $Y^*$  the concentration of activated ones, assuming that the total protein concentration is a conserved quantity,  $Y + Y^* = Y_{tot}$ . Several theoretical models with different levels of complexity can be used to describe the dynamics of the interconversion between  $Y$  and  $Y^*$ . For a recent review, see e.g. [22]. The simplest form of equations which describe the dynamics of the cycle between these proteins is given in the following [23]:

$$\dot{Y}^* = kS(t)(Y_{tot} - Y^*) - k^*Y^* \quad (3)$$



**Figure 1**  
**Motifs of activations and regulations of small biomolecular networks.** (a) A simple activation (b) activation of a signaling cycle (c) the Incoherent Feed Forward Loop (d) the Diamond motif (e) Diamond motif with signaling cycles (f) two Diamond motifs associated in series (g) the interlocked negative feedback loop.



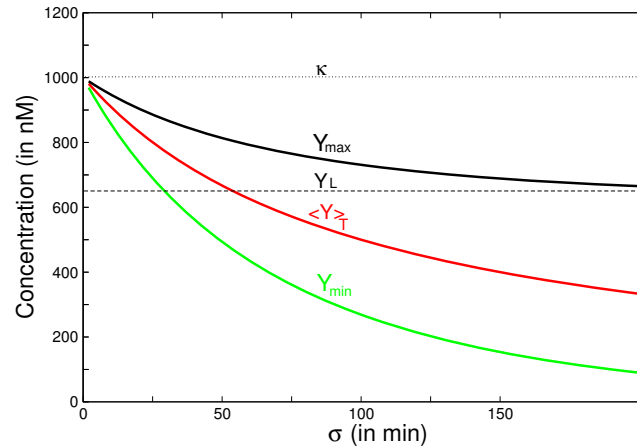
**Figure 2**  
**Input stimulation signal.** The pulsatile signal is shown, the duration of the "on-phase" is denoted by  $\tau$ , the inter-pulse interval or the silent phase between the pulses, is denoted by  $\sigma$ . Thus the period of  $S(t)$  is given by  $T = \tau + \sigma$ .

where  $k$  and  $k'$  denote kinetic constants of the respectively activating and deactivating reactions. When  $S(t)$  is a periodic train of square pulses as in the above observables  $Y_{max}$ ,  $Y_{min}$  and  $\langle Y \rangle_T$  can again be analytically calculated as a function of  $(\tau, \sigma)$ . However, their expressions are slightly more complicated than those of eqs.(2) due to the difference in time scales for the "on-phase" and "off-phase" of  $Y(t)$ . For a fixed value of  $\tau$  the graph of these functions of  $\sigma$  is qualitatively similar as seen in Fig. 3. The extrema and the mean values of  $Y(t)$  decrease when the inter-pulse interval  $\sigma$  is increased. Let us note that the same behavior is also achieved in the general case of nonlinear equations representing the dynamics of the covalent modifications by mean of the Goldbeter-Koshland model [3,24] (cf. additional file 1).

We conclude from elementary calculations that a pulsatile activation of the simple regulation schemes (1) or (3) provides a way to adjust the variations of  $Y$  in a range of values which can be controlled by tuning the temporal pattern  $(\tau, \sigma)$  of the pulses.

**Periodic activation of the IFFL**

In this Section we show that one basic design which enables one to obtain an optimal output in response to a periodic train of pulses, is the *incoherent feedforward loop* (IFFL) motif studied by Alon and co-workers in the context of transcriptional networks of bacteria [25]. We first give the definition of this motif and then study new properties which appear in this system under periodic stimulations.



**Figure 3**  
**The graph of extremum values of  $Y$  in function of  $\sigma$  for a simple activation process.** The extremum values of  $Y$  in function of  $\sigma$  for a simple activation process (Fig 1. (a)) are plotted thanks to the equations (2) of the main text.  $Y_{max}$  is the maximum reached by the protein,  $\langle Y \rangle_T$  is the mean concentration of the protein averaged over one period  $T$ ,  $Y_{min}$  is the minimal value reached by the protein,  $\kappa = \frac{\beta}{\alpha}$  (dotted line) represents the stationary state that the protein would attain if the stimulation was constant and  $Y_L$  (dashed line) is the asymptotic value reached by  $Y_{max}$  when the inter-pulse interval  $\sigma$  becomes very large. The parameters are:  $\alpha = 0.01 \text{ min}^{-1}$ ,  $\beta = 10 \text{ nM} \cdot \text{min}^{-1}$ ,  $\tau = 100 \text{ min}$ .

*The type-1 Incoherent Feedforward Loop*

According to the studies of [18], one of the most recurrent motif arising in a 3-nodes regulatory sub-network is the (type-1) incoherent feedforward loop seen in Fig. 1(c) (the abbreviation is IFFL in the following). Here the transcription factor  $X$  promotes the expression of both genes  $Y$  and  $Z$  but the expression of the latter is repressed by  $Y$ . In order to simplify the formalism so as to obtain simple analytical estimates, we again consider the logic approximation, where the regulation functions are described by step functions which take only binary values 0 or 1. As in the above,  $S(t) = R(X(t))$  denotes an input signal emanating from  $X$  and the dynamics of the system is described by the following equations:

$$\dot{Y} = \beta S(t) - \alpha Y \tag{4}$$

$$\dot{Z} = \beta S(t)H(\theta - Y) - \alpha Z \tag{5}$$

Here the first equation is a simple regulation as introduced in the previous section, and the second equation governs variable  $Z$  which is activated only when both the

signal  $S(t)$  is "on" and when the repressor level  $Y$  is smaller than the threshold  $\theta$ . We note that for simplicity's sake we have considered same degradation ( $\alpha$ ) and maximal production rates ( $\beta$ ), for  $Y$  and  $Z$ , but this choice is not essential for the results reported below.

The main dynamical property of the IFFL motif which was studied in [25] is its ability to create a pulse in  $Z$  concentration when it is submitted to a constant stimulus. Once the signal associated to  $X$  is switched "on", the variable  $Z$  rises due to its positive regulation by  $X$ , but this activation is terminated by the repressor  $Y$  which is also activated by  $X$ . The duration of the pulse can be quantified by the time  $\tau_\theta$  needed by  $Y$  to reach its repression threshold  $\theta$  after the onset of the signal. For the simple model above, the pulse duration is merely computed as:

$$\tau_\theta = -\frac{1}{\alpha} \log(1 - \theta / \kappa) \quad (6)$$

This will be called the proper pulse duration of the IFFL system. Thus if the activation lasts longer than the proper pulse  $\tau_\theta$  the system can recover its primary state. In other words the IFFL system can only detect changes of stimulation, meaning that it *adapts* to a constant stimulus. So, as mentioned by Li and Goldbeter [16] in regarding the context of receptor desensitization, in order to reach a given level of synthesis of  $Z$  in such system, a pulsatile pattern of stimulation must be considered instead of a continuous stimulation. Therefore in the following we consider a periodic stimulus  $S(t)$ , which has the form of a square-wave, similar to that studied in the previous Section.

*Pulsatile periodic activation of the IFFL*

In this Section we show that when the IFFL network is periodically stimulated by a train of pulses, a new property appears regarding the optimization of observables  $\langle Z \rangle_T$  and  $Z_{\max}$ . As illustrated in Fig. 4 and Fig. 5, the average value  $\langle Z \rangle_T$ , as well as  $Z_{\max}$  can reach maximal values for specific choices of the pulse pattern ( $\tau, \sigma$ ). We will describe this phenomenon in more detail.

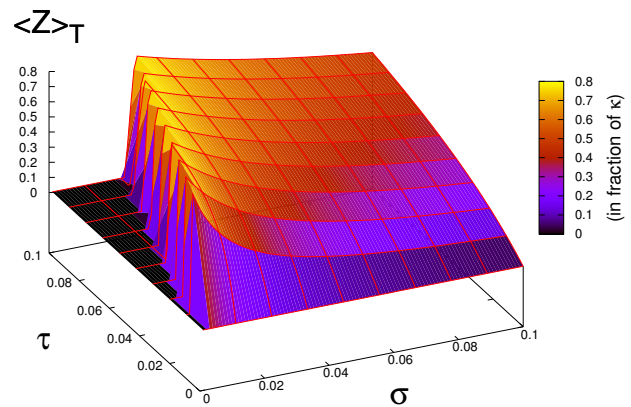
First, the evolution of the repressor  $Y(t)$  is identical to the one obtained in the previous section for a simple regulation, because eq.(4) coincides with eq.(1). However, in what concerns the evolution of the output protein  $Z$ , a detailed analysis of this system (see additional file 1) shows that two cases should be distinguished according to the size of the pulse duration  $\tau$  compared with the proper pulse  $\tau_\theta$  of the IFFL [cf. eq. (6)]. Here we detail only the case of short pulses, corresponding to the condition  $\tau < \tau_\theta$ . Fig. 4 shows a 3D-plot of  $\langle Z \rangle_T$  computed in function of the pulse parameters ( $\tau, \sigma$ ). One observes that the extrema of  $\langle Z \rangle_T$  form a ridge with a roughly constant slope  $\tau/\sigma$  in the parameter space ( $\tau, \sigma$ ). By fixing the pulse duration  $\tau$  Figs.

5 show the existence of maxima in  $\langle Z \rangle_T$ , as well as in  $Z_{\max}$  in Fig. 5b, reached for a critical inter-pulse interval  $\sigma_1$ . Another feature of these observables shown in Figs. 5 is that they start to increase from zero only if the inter-pulse interval  $\sigma$  is above a minimum value  $\sigma_0$ . When  $\tau < \tau_\theta$  the values of  $\sigma_0$  and of  $\sigma_1$  can be analytically computed as follows (cf. Suppl. Info.):

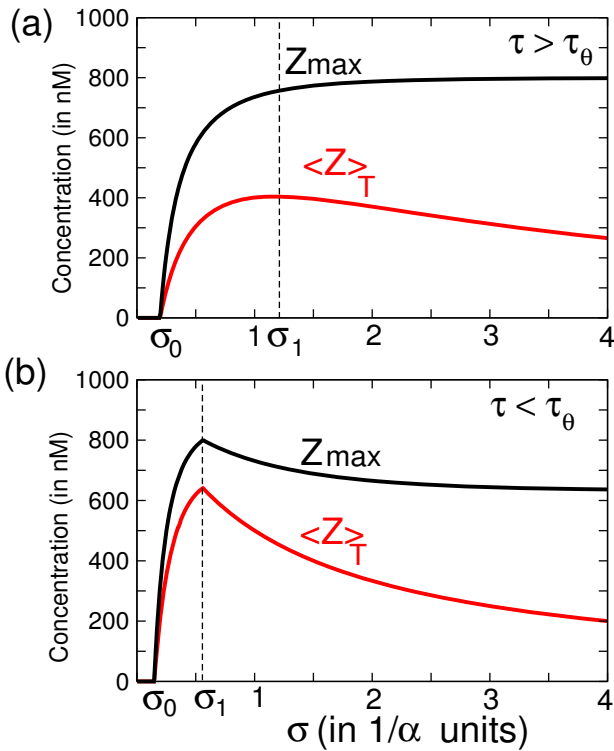
$$\sigma_0 = \frac{1}{\alpha} \log \left( 1 + \frac{\kappa}{\theta} (e^{\alpha\tau} - 1) \right) - \tau \quad (7)$$

$$\sigma_1 = -\frac{1}{\alpha} \log \left( 1 - \frac{\kappa}{\theta} (1 - e^{-\alpha\tau}) \right) - \tau \quad (8)$$

The latter expression can serve to estimate the optimal interspike interval  $\sigma_1$  which should be waited for between the pulses to maximize the response of the IFFL network. We note that when  $\alpha\tau$  is small, one can use the first-order approximation  $\sigma_0 \sim \sigma_1 \sim \tau (\kappa/\theta - 1)$ , where  $\kappa$  is the equilibrium concentration of  $Y$  proteins, and  $\theta < \kappa$  is the repression threshold of  $Y$  with respect to gene  $Z$ . Our estimate means that provided  $\alpha\tau$  is small enough, the optimal pulse pattern is given by only one condition on the ratio between the on/off phases, i.e.  $\sigma_1/\tau \sim \kappa/\theta - 1$ . It does not impose a unique value for the pulse duration  $\tau$ . Furthermore, it can be seen that the height of the optimal  $\langle Z \rangle_T$  and  $Z_{\max}$  is of the order of  $\theta$ . Thus the average and maximum productions of  $Z$  are optimized when  $\theta$  is close to its max-



**Figure 4**  
**Response of the IFFL to periodic stimulation in function of  $\sigma$  and  $\tau$ .** Average response  $\langle Z \rangle_T$  is plotted in function of  $\sigma$  and  $\tau$  for an IFFL motif (Fig 1. (c)) stimulated by a pulsatile signal. Parameters  $\sigma$  and  $\tau$  are expressed in  $1/\alpha$  units and  $\langle Z \rangle_T$  in fraction of  $\kappa = \frac{\beta}{\alpha}$ . Threshold parameter is chosen as  $\theta = 0.8 \kappa$ . This surface plot of  $\langle Z \rangle_T$  is analytically computed by using eq.(10) given in the Additional File 1.



**Figure 5**  
**Response of the IFFL to periodic stimulation in function of  $\sigma$ .** The average response  $\langle Z \rangle_T$  and the maximum of response  $Z_{max}$  are represented in function of the inter-pulse interval  $\sigma$  with a fixed  $\tau$  for an IFFL motif (Fig 1.(c)) stimulated by a pulsatile signal.  $\sigma$  is in  $1/\alpha$  units, the concentration is in nM.  $\sigma_0$  is the minimal inter-pulse interval for the system to respond. The inter-pulse interval  $\sigma_1$  gives the optimum average response for the system. The numerical simulation was done with the equations (4–5) of the main text and with the following parameters: (a) for the case  $\tau > \tau_\theta$   $\tau = 2$   $1/\alpha$  unit. For example, the parameters:  $\alpha = 0.01 \text{ min}^{-1}$ ,  $\beta = 10 \text{ nM} \cdot \text{min}^{-1}$ ,  $\theta = 800 \text{ nM}$  and  $\tau = 200 \text{ min}$  give  $\sigma_0 \sim 20 \text{ min}$  and  $\sigma_1 \sim 2 \text{ hours}$ . (b) for the case  $\tau < \tau_\theta$   $\tau = 1$   $1/\alpha$  unit. For example, the parameters:  $\alpha = 0.01 \text{ min}^{-1}$ ,  $\beta = 10 \text{ nM} \cdot \text{min}^{-1}$ ,  $\theta = 800 \text{ nM}$  and  $\tau = 100 \text{ min}$  give  $\sigma_0 \sim 15 \text{ min}$  and  $\sigma_1 \sim 1 \text{ hour}$ .

imal possible value, which is bounded by  $\kappa$ . But then  $\sigma_1 \ll \tau$ , meaning that in this case the optimal rest interval would be smaller than the pulse duration. Therefore for the IFFL with a high repressor threshold, optimal trains of square pulses would seem to be alike in a constant stimulation. However, for this system, there is a crucial difference between the situation of constant stimulus, where  $Z$

falls to zero, because it is completely repressed by  $Y$ , and the other situation where the activation of  $Y$  is periodically interrupted during for a small lapse of time. In real biochemical systems, however, the value of the threshold  $\theta$  cannot be arbitrarily chosen, and the implementation of our proposed principle depends on the context, being for instance the one of genetic regulation or the one of intracellular signaling. Based on plausible parameter values found in the literature, Table 1 gives examples of optimal pulse patterns predicted by our simple model in the case of genetic expression, eqs.(4)-(5), or in the case of signaling cycles, eqs.(3)-(5). The first line gives optimal inter-pulse durations  $\sigma_1$  for a simple IFFL composed of three genes with parameters in the range given in the experimental work of [26]. The second line gives an estimation of  $\sigma_1$  for an IFFL whose node  $Y$  is a signaling cycle. The numerical values for the parameters of the signaling cycle, ( $k' = k = 1 \text{ s}^{-1}$ ,  $Y_{tot} = 2 \text{ } \mu\text{M}$ ), are based on values estimated in [27].

In conclusion one sees that the IFFL network motif (Fig. 1(c)) studied by Alon *et al.* has a nice ability to provide one of the simplest mechanisms which maximizes the average synthesized protein  $\langle Z \rangle_T$  when periodically stimulated with a train of pulses.

**Mixed regulatory networks with a "Diamond" IFFL**

The periodic stimulation of the IFFL network with train of pulses reveals an interesting property of optimal response with respect to the pulse pattern. In this section we slightly extend the IFFL motif to obtain a design which improves its ability of bandpass filtering, i.e. the property to induce a non zero response only for a finite range of pulse periods. We consider the "diamond IFFL" motif depicted in Fig. 1(d). The scheme is analog to the IFFL but the activation of the target gene takes two nodes instead of only one. In fact, this extension of the IFFL can be studied in the gene network seen in Fig. 1(d). It can also be implemented in a mixed regulatory network which blends signaling and gene nodes in the same interaction graph (cf. Fig. 1(e)). This graph may schematically represent a transduction network consisting of a crosstalk between two pathways converging with opposite interactions in the same genetic system. A similar scheme has been used in the study of the formation of long term memory at the molecular level [15]. The comparison between the models of the literature and the outcome of our analysis will be provided in the Discussion section.

**Table 1: Examples of plausible numerical values for the different parameters**

Simple IFFL in bacteria [26]	$\alpha = 0.01 \text{ min}^{-1}$	$\beta = 100 \text{ nM} \cdot \text{min}^{-1}$	$\theta = 500 \text{ nM}$	1 to 10 min	25 to 390 min
IFFL with a signaling cycle [27]	$K' = 1 \text{ s}^{-1}$	$k = 1 \text{ s}^{-1}$	$\theta = 500 \text{ nM}$	1 s	2 s

We again use the simplified description of eq.(3) for the dynamics of the covalent modification cycles, with variable  $Y_1^*$  and  $Y_2^*$ , and the logic regulatory functions for the activation of the promoter of Z. The equations of the system depicted on Fig. 1(e) can be written as previously with the following dynamics for the proteins  $Y_1^*$ ,  $Y_2^*$  and Z:

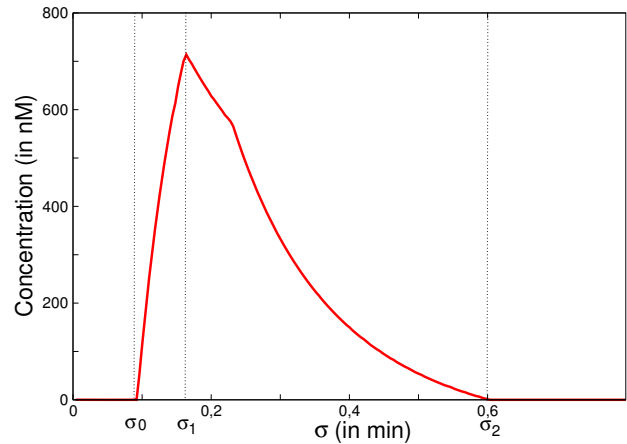
$$\dot{Y}_1^* = kS(t)(Y_{1tot} - Y_1^*) - k'Y_1^* \quad (9)$$

$$\dot{Y}_2^* = kS(t)(Y_{2tot} - Y_2^*) - k'Y_2^* \quad (10)$$

$$\dot{Z} = \beta H(Y_1^* - \theta_1)H(\theta_2 - Y_2^*) - \alpha Z \quad (11)$$

In this simple setting where identical kinetic parameters are assumed for the covalent modification cycles 1 and 2, the time-evolutions of  $Y_1^*$  and  $Y_2^*$  are synchronized by the periodic signal  $S(t)$ . In this case, eq. (11) shows that the expression of Z gene is activated only when  $\theta_1 < Y_1^* < \theta_2$  ( $i = 1, 2$ ), which entails that the relation  $\theta_1 < \theta_2$  must be assumed in order to have a non zero production of Z. This assumption means that the affinity of the activator for binding the promoter of Z gene is higher than the affinity of the repressor for the same promoter.

We consider periodic stimulations of this system with a pulse pattern  $(\tau, \sigma)$  such that  $\sigma$  is varying and the pulse duration  $\tau < \tau_{\theta_1}$  is fixed (cf. eq.(6)). Fig. 6 shows the variation of the mean value of synthesized Z, denoted by  $\langle Z \rangle_T$ . As for the IFFL, there is a lower limit of the inter-spike interval  $\sigma_0$  in which no production of the target gene Z occurs. But once  $\sigma$  is increased above  $\sigma_0$  there is a fast growth of  $\langle Z \rangle_T$  which culminates at  $\sigma_1$  which can be computed in the same way as before (eq.(8)). If  $\sigma$  is further augmented,  $\langle Z \rangle_T$  registers a decrease which reaches 0 when  $\sigma > \sigma_2$  on the graph. This property is due to the existence of a new node (activator Y1) in the "diamond" IFFL, which ceases to activate the target gene Z once  $Y_1$  max becomes smaller than its activation threshold of Z. Thus the new feature of the pulse-response of  $\langle Z \rangle_T$  is that (for a fixed  $\tau$ ) one obtains only a finite interval  $[\sigma_0, \sigma_2]$  of inter-pulse intervals in which the protein Z is produced. Therefore, the new property appearing in this system, as compared with the standard IFFL, gives us the possibility to use this architecture as a bandpass filtering, allowing the gene to respond only to specific periodic trains of pulses,



**Figure 6**  
**Response of the Diamond IFFL motif with signaling cycles to periodic stimulation in function of  $\sigma$ .** The average response  $\langle Z \rangle_T$  of the protein Z of the Diamond IFFL motif with signaling cycles (Fig. 1 (e)) is shown for a pulsatile stimulation.  $\sigma_0$  is the minimal inter-pulse interval for the system to respond. The inter-pulse interval  $\sigma_1$  gives the optimum average response for the system and  $\sigma_2$  is the maximal inter-pulse interval for the system to respond. The numerical simulation was done with the equations (9–11) and with the following parameters:  $k = 4 \text{ min}^{-1}$ ,  $k' = 2 \text{ min}^{-1}$ ,  $Y_{1tot} = Y_{2tot} = 500 \text{ nM}$ ,  $\alpha = 0.01 \text{ min}^{-1}$ ,  $\beta = 10 \text{ nM} \cdot \text{min}^{-1}$ ,  $\theta_1 = 180 \text{ nM}$ ,  $\theta_2 = 230 \text{ nM}$ ,  $\tau = 0.1 \text{ min}$ .

with the inter-pulse  $\sigma$  belonging to a limited range  $[\sigma_0, \sigma_2]$ .

#### Optimal output in response to bursting oscillations

The network motifs studied in the previous Sections can be combined in modules that can be linked in a larger network to process complex signals. We show in this Section that by linking two diamond IFFL motifs one obtains a class of regulatory networks which are able to selectively respond to bursting oscillations.

Bursting oscillations are time-dependent signals typically emitted by neuron cells [28]. Their temporal pattern is formed by trains of spikes alternating with refractory periods. The neuronal signal is a depolarization wave created by the dynamics of ionic channels in the neuron membranes. However, in the vicinity of a synapse, this electrical signal transforms into a chemical signal affecting the receptors of the post-synaptic neuron. It thereby propagates its influence in the molecular regulatory networks of this neuron. Consequently, a bursting neuronal signal can potentially impact on the gene expression of the post-synaptic neuron. Therefore it is worthwhile to identify simple topologies of regulatory networks with the ability to be sensitive and selective to bursting oscillations.



In order to explore this question we consider the network represented in Fig. 1(f), which combines in series two diamond IFFL motifs. Each node of Fig. 1(f) is represented by single species but in practice the proteins  $Y_1$  and  $Y_2$  on this graph could be replaced by covalent modification cycles as the ones depicted in Fig. 1(e). In this case Fig. 1(f) corresponds to a regulatory system where an input signal influences the expression of a target gene  $W$  by the intermediary of a signaling pathway, in which 2 phosphorylation/dephosphorylation cycles respectively activate and repress the promoter of the intermediate gene  $Z$ . The corresponding protein activates in turn the repressor  $V_1$  and the activator  $V_2$  of a target gene  $W$ .

Employing similar notations as in the previous Sections, the dynamics of the network on Fig. 1(f) can be described by the following system:

$$\dot{Y}_1 = \beta_1 S(t) - \alpha_1 Y_1 \quad (12)$$

$$\dot{Y}_2 = \beta_1 S(t) - \alpha_1 Y_2 \quad (13)$$

$$\dot{Z} = \beta_1 H(Y_1 - \theta_1) H(\theta_2 - Y_2) - \alpha_1 Z \quad (14)$$

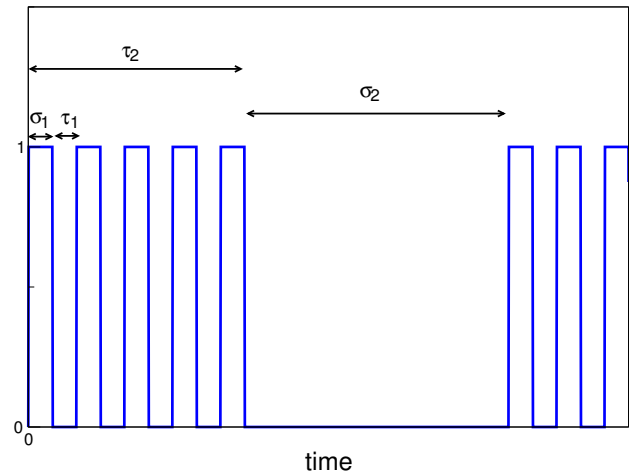
$$\dot{V}_1 = \beta_2 H(Z - \theta_3) - \alpha_2 V_1 \quad (15)$$

$$\dot{V}_2 = \beta_2 H(Z - \theta_4) - \alpha_2 V_2 \quad (16)$$

$$\dot{W} = \beta_2 H(V_1 - \theta_5) H(\theta_6 - V_2) - \alpha_2 W \quad (17)$$

To model the input bursting signal  $S(t)$  we consider a time-periodic pattern characterized by four time intervals ( $\tau_1, \sigma_1, \tau_2, \sigma_2$ ) (cf. Fig. 7). The period of the signal is given by  $T = \tau_2 + \sigma_2$ , where  $\tau_2$  is the duration of the bursting phase and  $\sigma_2$  is the quiescent period. The bursting phase is defined by a series of square spikes in which  $\tau_1$  is the width of the spike and  $\sigma_1$  is the interspike interval. The quiescent period which separates the bursting phases is given by  $\sigma_2$ . Thus the signal is characterized by two different time-scales  $\tau_1 + \sigma_1 \ll \tau_2 + \sigma_2$ . In the numerical simulations reported below these time scales are separated by two orders of magnitude, reflecting the typical temporal differences between the dynamics in signaling cascades and in genetic regulations (cf. Table 1).

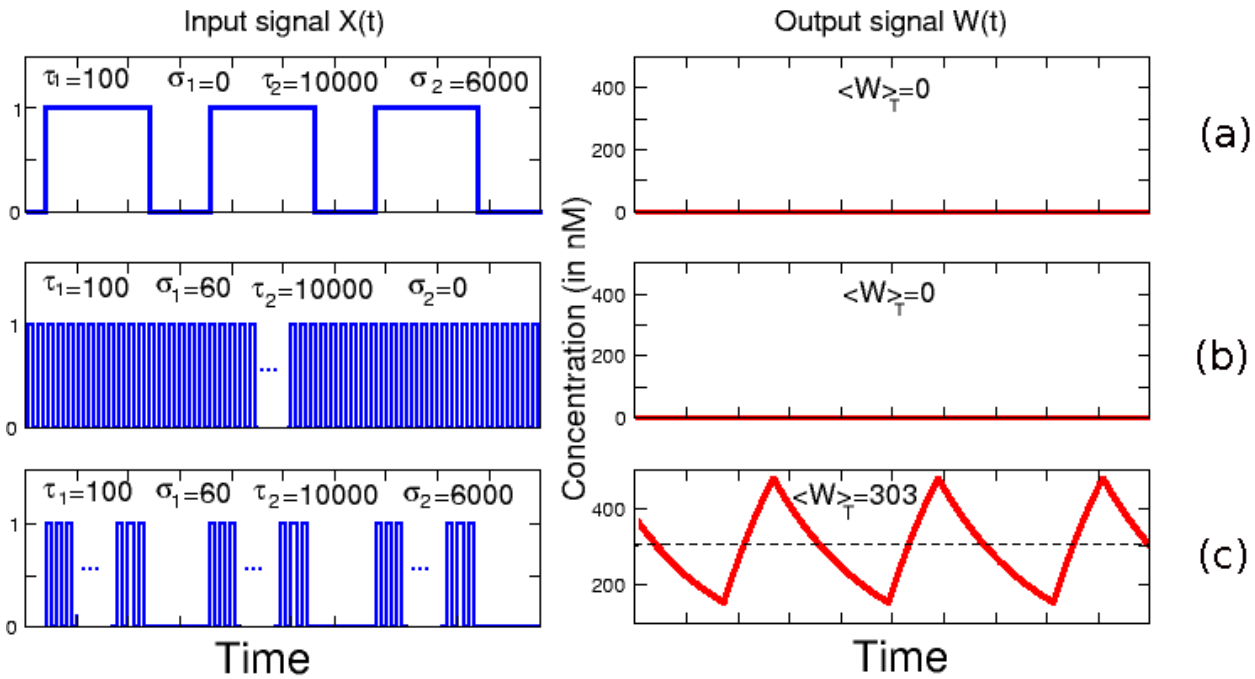
When the network represented in Fig. 1(f) is stimulated by the signal  $S(t)$  described above, it has the property to selectively recognize temporal patterns of bursting oscillations, allowing the target gene to be maximally transcribed in some conditions. To illustrate this point Fig. 8 shows the time-evolution of  $W$  in response to various periodic stimulations  $S(t)$ . When the system is submitted to periodic trains of square waves without bursting (Fig.



**Figure 7**  
**Bursting signal.** The signal is composed of 2 signals of different time scale. The first is a rapid pulsatile signal with the duration of the "on-phase" denoted by  $\tau_1$  and the silent phase between the pulses denoted by  $\sigma_1$ . The second signal is two orders of magnitude slower. The duration of the "on-phase" is denoted by  $\tau_2$  and the "off" phase between the pulses, is denoted by  $\sigma_2$ . For practical reasons, the graph is not at real scale.

8(a),  $\sigma_1 = 0$ ), the gene  $W$  is not expressed. Likewise if the stimulation consists of a long train of spikes without any quiescent period (Fig. 8(b)), the average level of  $W$  remains negligible. However, if we stimulate the motif with a specific bursting signal (Fig. 8(c)), the system gives a non-zero response. More generally, the striking feature of the network of Fig. 1(f) is to exhibit a non vanishing response only in a given range of pulse patterns. Moreover, if the time intervals  $\tau_1$  and  $\tau_2$  are fixed, the system possesses a set of maxima for some optimal values of  $(\sigma_1, \sigma_2)$ . In view of of Fig. 8(a-c), the system behaves as it filtered out low as well as high frequencies. But this conclusion is misleading since when high and low frequencies are mixed in the same input signal in the form of bursting oscillations, the system displays a non zero response in the evolution of  $W$ , with the possibility of optimizing the average level of  $W$  over one period.

The reason why the network in Fig. 1(f) manifests a large response to bursting oscillations is a direct consequence of the results discussed in the previous Sections. Let us again suppose that the active phase of the bursting lasts  $\tau_2$  and the quiescent interval is  $\sigma_2$ . During this active phase  $\tau_2$ , and assuming the spike duration  $\tau_1$  is fixed, an optimal output in the level of the intermediate protein  $Z$  is produced if the interspike interval  $\sigma_1$  is chosen as the one which maximizes the response curve of Fig. 6. For example, using the parameter of that figure, this corresponds to  $\sigma_1 \approx 1/(4\alpha_1)$ . On the other hand, during the refractory



**Figure 8**

**Response of the double DIFFL to a bursting signal.** The motif of double DIFFL (Fig. 1(f)) is stimulated with 3 different periodic input signals  $X(t)$  (on the left). The 3 graphs on the right give the response of the output  $W(t)$ . The numerical simulations were done with the following parameters  $\tau_1 = 100$  s and  $\tau_2 = 10000$  s are fixed. The other parameters are:  $\alpha_1 = 0.01$  sec<sup>-1</sup>,  $\beta_1 = 10$  nM.sec<sup>-1</sup>,  $\alpha_2 = 0.006$  min<sup>-1</sup>,  $\beta_2 = 6$  min<sup>-1</sup>,  $\theta_1 = \theta_5 = 700$  nM,  $\theta_2 = \theta_6 = 800$  nM,  $\theta_3 = \theta_4 = 50$  nM. (a) with  $\sigma_1 = 0$  s,  $\sigma_2 = 100$  min (b) with  $\sigma_1 = 60$  s,  $\sigma_2 = 0$  min (c)  $\sigma_1 = 60$  s,  $\sigma_2 = 100$  min. For practical reasons, the representations of the pulsatile signals are not at real scale.

period  $\sigma_2$  of the bursting oscillations, the activity of  $Z$  returns to zero. Thus the time-evolution of the intermediate protein  $Z$  is close to a periodic train of square pulses with time pattern  $(\tau_2, \sigma_2)$  characterizing respectively its "on" and "off" phases. Now this variable feeds the regulation functions of the second subnetwork  $V_1 - V_2 - W$ . But for a fixed  $\tau_2$  this subnetwork responds in an optimal way if the quiescent phase is chosen again as the critical  $\sigma_2$  which maximizes a response curve like the one shown in Fig. 6. For example, using the same parameter as for this figure,  $\sigma_2$  would correspond to  $1/(4\alpha_2)$ . Therefore, since the signaling time scale  $1/\alpha_1$  is much smaller than the one of gene regulation  $1/\alpha_2$ , we conclude that the temporal pattern which maximizes the response in  $W$  is  $\sigma_1 \ll \sigma_2$ . This agrees with the concept of bursting oscillations.

#### The interlocked negative feedback loop

The network motifs analyzed in the previous Sections can give rise to an optimum period of stimulation corresponding to a maximum in the *average* concentration of the output proteins. It is also interesting from a physiological point of view, to find biological networks that give an optimum frequency maximizing the oscillatory *amplitude*

of the output. So, we consider another basic network motif which provides a maximal response in function of a periodic stimulation. The motif network, which is shown in Fig. 1(g), is similar to the IFFL, but the repressor  $Y$  in addition to being activated by  $X$  is activated by  $Z$  in a feedback loop. Furthermore,  $Z$  is not directly activated by  $X$ . This negative loop is not a gene regulatory motif in the sense defined by Alon et al., but we still call it a motif as it is the building block of several molecular biology oscillators [29]. We note that often this regulatory motif is not purely genetic, but the negative control is induced by a post-transcriptional regulation [30]. Furthermore, when the species  $Z$  not only activates  $Y$  but also exerts an auto-activation, the system can become a biological oscillator. In fact this positive auto-regulation is essential in order to display self-sustained oscillations. For example, minimal models of circadian oscillators have been proposed, involving the pattern of activation seen in Fig. 1(g)[31]. Likewise, minimal models for the cellular cycle have also been proposed in this way [32]. This network motif also supports a model for the oscillatory dynamics of  $p53$  [33], oscillations in *Neurospora* [34]. It has also been studied in a synthetic biology perspective [35]. Other examples



**Table 2: Different models of the interlocked negative feedback loop. This table shows various regulators which are involved in an interlocked negative feedback loop.**

Model	Cellular cycle [32]	P53 [33]	Synthetic system [35]	Circadian rhythms (Neurospora) [34]
Y (Repressor)	CLB	Mdm2	LacI	FRQ
Z (Activator)	CLN	P53	NRI	WC-1, WC-2

exist (cf. Table 2), like for instance a model given by Song et al. [36] who also use the interlocked negative feedback loop to describe a model of memory formation.

All these examples have been proposed as models of biological clocks, because they admit autonomous oscillations of the produced protein concentrations. In absence of the positive auto-regulation of the activator, however, it can be shown that no autonomous oscillations are possible. This is the case we consider for the network motif depicted in Fig. 1(g). But even in this simpler network, interesting oscillatory behaviors can appear if the variable Z (or Y) is influenced by periodic variations coming from X. This periodic activation can be caused by a signal associated with an autonomous oscillator, or by the outcome of a signaling pathway which is periodically stimulated, as considered in the previous Section. In all the cases we represent the periodic activation of unit X in Fig. 1(g) by a signal  $S(t)$ , and the dynamics of the protein concentrations Y and Z are determined by the following equations:

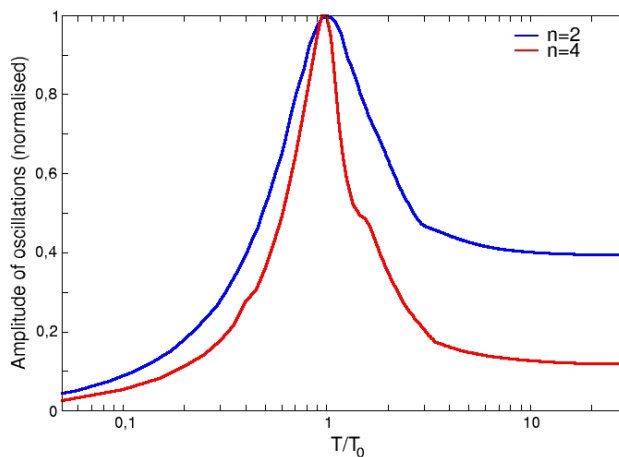
$$\begin{aligned} \dot{Y} &= (\beta_0 + \beta_1 S(t))R^+(Z) - \alpha Y \\ \dot{Z} &= \beta_2 R^-(Y) - \alpha Z \end{aligned} \tag{18}$$

In this system, the signal  $S(t)$  associated with  $X(t)$  can be a periodic square pulse of amplitude 1,  $R^+(Z)$  is an activating regulation function, like the Hill function introduced above, and  $R^-(Y) = 1 - R^+(Y)$  is a repressing function.

The resonance response is much clearer in this case if the input signal is sinusoidal rather than a square signal. The work performed by Lipan and Wong [37] proposes the use of oscillatory signals for studying genetic networks. These authors have devised a promising experimental procedure which consists of activating and deactivating the promoter efficiency with the use of electromagnetic fields. The net effect of their procedure amounts to periodically modulating the expression rate of genes.

When the periodic activation is absent,  $\beta_1 = 0$ , the system possesses a non-zero steady state  $(Y_s, Z_s)$  which can be shown to always be a stable focus. The corresponding frequency  $\omega_0$  is easily determined by computing the imaginary part of the Jacobian eigenvalues. Thus, if the system is drawn away from equilibrium, it will return to its steady state by exhibiting transient oscillations with the fre-

quency  $\omega_0$ . This frequency can also be revealed by periodically modulating the expression rate of Y. In this case we suppose that the modulation is operated with a frequency varying in a range around  $\omega_0$ . Numerical simulations of eqs. (18) show that in this range the network dynamics give rise to a resonance which coincides with  $\omega_0$ . In Fig. 9, we can see a maximum in the amplitude response of protein Y. A typical order or magnitude for the resonant period obtained from plausible parameters is  $T_0 \sim 300$  min. Moreover we observe that the resonance width mainly depends on the stiffness of the regulatory functions, quantified here by the cooperativity coefficient  $n$  of the Hill function. To explain this property, an easy but instructive example which can be analytically dealt with, considers identical thresholds  $\theta$  for the regulatory Hill functions  $R^+(Z)$  and  $R^-(Y)$ , and also assume  $\beta_0 = \beta$ . The steady state for Y and Z is symmetric, i.e.  $Y_s = Z_s = \kappa/2$  (we recall that  $\kappa = \beta/\alpha$ ). The natural frequency can be written as  $\omega_0 = n\alpha/2$ , where  $n$  is the Hill coefficient of functions  $R^+(Z)$  and  $R^-(Y)$ . Since, in absence of periodic stimulations, the damping of the oscillations is characterized by



**Figure 9**  
**Response of the interlocked negative feedback loop to a harmonic stimulation.** The response in amplitude of the interlocked negative feedback loop (Fig 1.(g)) to a harmonic stimulation in function of the period of stimulation T is shown. The numerical simulation was done with the equations (18) and with the following parameters:  $\beta_0 = 150$  nM.min<sup>-1</sup>,  $\beta_1 = 100$  nM.min<sup>-1</sup>,  $\beta_2 = 100$  nM.min<sup>-1</sup>,  $\theta_1 = 50$  nM,  $\theta_2 = 50$  nM,  $\alpha = 0.01$  min<sup>-1</sup>,  $n = 2$  or  $4$ . With these parameters, we have  $T_0 \sim 6$  hours.

$\alpha$ , we can estimate the "quality" factor of this resonance with the ratio  $\omega_0/\alpha = n/2$ . This indeed depends only on the Hill coefficient  $n$ . Examples of resonances with  $n = 2$  and  $n = 4$  are illustrated in Fig. 9.

## Discussion

In this study we have looked at simple molecular networks showing the ability to maximize their responses to periodic stimulations. Using analytical and numerical tools, we have identified two types of basic network motifs, and some of their straightforward extensions, which possess this property (Fig. 1). The first of these motifs is the Incoherent Feedforward Loop network (IFFL), for which we have shown that periodic activations with a train of square pulses lead to maximizing the time-average production of the output protein over one period provided that the pulse and the inter-pulse intervals are adequately chosen (Fig. 4). When the pulse duration is short enough, this optimization also applies to the maximal concentration of the output protein reached over one period (Fig. 5b). The mathematical analysis of this optimization property is simple, although it does not reckon as do most current studies which focus on this topic, with the assumption of small amplitude stimulations, which allows to work in the framework of linear filter analysis. We have also pointed out another network motif, namely the interlocked negative feedback loop which has the property to exhibit a maximum in the amplitude of oscillations when it is submitted to periodic stimulations. In this case the maximization is observed in the linear regime (i.e. for small amplitude of the stimulations) and it can be described as a standard phenomenon of resonance.

The interlocked negative loop is a well-studied motif in the context of oscillating regulatory networks (cf. examples given on Table 2). In all these examples the auto-activation of one of the two nodes of the network is imperative in order to produce sustained oscillations. Here we pinpoint a property of the interlocked feedback loop that exists without or with only a weak auto-activation. In this case, the system relaxes towards a stationary state in absence of periodic stimuli. It nevertheless possesses the ability to show up a resonant behavior which can be waked by periodic stimulations. This phenomenon might be important in some situations. For example the protein P53 is engaged in many regulation processes and it is known to oscillate with a period of 5.5 hours in some conditions. In [38] the authors show that the amplitude of oscillations of P53 is variable whereas the period of oscillations has a spectacular regularity. We speculate that even in absence of autonomous oscillations this system might show up an amplified response to external periodic stimulations (e.g. a periodic exposure to UV rays) because of the resonance phenomenon. A similar remark can be made regarding regulatory systems related to stress

responses as in the SOS regulator LexA of *E. coli* or in the NF- $\kappa$ B system.

As an illustrative application of the periodic activation of the diamond IFFL motif we believe that the feature studied in the present paper reveals the principle behind an intriguing phenomenon studied by Smolen et al. [15]. In this reference, the authors proposed a model for the formation of long term memory (LTM), a process which is based on the LTP phenomenon (Long Term Potentiation) in neurons and on the subsequent strengthening of synapses. The LTP appears in a synapse when the postsynaptic cell is submitted to some external signal, for instance in the form of glutamate pulses of significant amplitude. Then this signal induces the phosphorylation of transcription factors which genetically activate the production of new proteins in order to strengthen connections between synapses or to create new synapses. On the other hand, it has been shown that in order to learn, several animals or organisms need to receive stimulations in repeated sessions (a well-studied example is the aplysia, see [39] for a review). Interestingly these sessions must be spaced with a minimal interval. Massed training produced significantly shorter lasting memory as experimentally shown in [40]. Thus the genetic network underlying this regulation must respond only if the learning sessions are sufficiently spaced.

The model of Smolen et al. (Fig. 5 of [15]) had the same structure as the diamond IFFL shown in Fig. 1(e). It postulated the existence of two transcriptional factors, respectively an activator and a repressor, whose phosphorylated forms are activated by an external signal, allowing them to act on the promoter of a gene involved in the formation of LTM (e.g. the *creb* genes). The purpose of this modeling was to show the marked differences in the transcription rate of the LTM genes between the case of "massed stimuli" and the case of "spaced stimuli", as was experimentally observed e.g. by Yin et al. [2]. In fact an outstanding property of their model was the existence of a maximum in the transcription rate of the LTM genes, in function of the interspike interval (here corresponding to the rest time between the learning sessions). However the definite mechanism underlying this property is not easy to identify in that system, as the model counted a half dozen of nonlinear equations, with nearly 20 parameters to be chosen. Despite this complexity the authors did conclude by suggesting that a general mechanism enabling the tuning of the response of a gene promoter to an optimal stimulus frequency could be the presence of two competing processes with different sensitivities acting on this gene. In this article we have achieved a minimal model which puts this intuition on firm grounds. Our analysis shows that the presence of two competing processes related by the IFFL mechanism is indeed a key ingredient, a higher affinity of

the activator compared with the one of the repressor for the target gene promoter is another one.

The results of Li and Goldbeter [16] mentioned in the introduction are another rare example found in the literature where a maximization process of the response in a molecular network is shown to critically depend on some temporal patterns of activation. Although this study is based on analytical calculations the basic ingredient which explains the observed property is not transparent. Although the link of this model with the IFFL motif is not established we are currently working on a generalization of the optimization principle identified in the IFFL responses in order to place the two studies in the same framework.

Based on typical parameter values met in different systems (genetic, signaling), Table 1 provides some estimates of the pulse and of the inter-pulse intervals which allow for maximization of the average production of target proteins. Since these estimates seem compatible with plausible experimental values, they prompt experimental implementations of the scheme. For example, several IFFL regulatory motifs have been recognized in a genome map of *E. coli* [18]). It would be interesting to stimulate one of these loops with trains of periodic pulses in order to discover whether the predicted maximization is feasible in practice. Utilization of thermosensitive promoters may be one possibility for the implementation of the periodical stimulations [41]. Nevertheless we note that the phenomena presented here may be sensitive to molecular noise and to environmental perturbations. Presumably, the envisaged experiments might be less difficult to achieve in eukaryote organisms where the genetic regulations involve more proteins than in prokaryotes, in this way, molecular noise could be reduced.

A general issue concerning the concept of network motifs, like the IFFL, is that such a module is in fact embedded in larger and more complex regulatory networks [42]. This weakens the actual meaning of modular subnetworks. In this regard, however, by considering time dependent signals we propose that frequency selectivity mechanisms like the ones studied here, can give more sense to the modular approach. Finally we note that from a biomedical perspective the question of optimal response in molecular networks could be especially relevant in situations where a maximal protein synthesis would be beneficial in periodic stimulation-based therapies. For example some innovative treatments of neural based diseases like Parkinson's, or of severe depressions, use a novel technique consisting of deep brain stimulations [43]. The patient wears a pacemaker linked to electrodes implanted in some brain nuclei. Electrical stimulations are then transmitted at a given frequency. Although this type of clinical treat-

ment has been reported to lead to spectacular relief, the question of fine tuning the frequency of the pacemaker to achieve an optimal result is empirically solved for the time being [44]. A better understanding of the time-dependent responses of intracellular signaling pathways could be very relevant in these situations. In a different context, the traditional periodic intake of medicines is another example where optimization will emerge from rational thinking in terms of optimal pulsatile stimulations.

## Conclusion

The mechanisms presented in this article identify new possible strategies which can be employed by a cell to integrate time-dependent information provided by the environment. The studied network motifs offer the attractive possibility of selecting a signal according to its temporal structure or to its frequency content. Only messages which for instance, which have a minimal inter-pulse duration are allowed to pass. The identified mechanisms are simple and based on known network motifs in the literature, they can be embodied in existing organisms. They could also be implemented by means of synthetic biology [41,45].

## Methods

### Numerical simulations

Numerical simulations were made with programs written in C language and the Runge-Kutta routines from the GNU Scientific Library. The simulation codes of the models which were used to obtain the numerical results reported in the paper (e.g. Figs. 5, 6, 8 and 9) are provided in the additional file 2 (with the name "archive-codes").

### Authors' contributions

Conceived and designed the experiments: AC, JAS. Analyzed the data: AC, J-AS. Wrote the paper: AC, JAS. All authors read and approved the final manuscript.

## Additional material

### Additional File 1

*Supplementary Information. Additional details are given about the mathematical analysis of the periodic activation of the IFFL network by a train of square pulses and about a comparison of our study with the frequency coding analysis performed by Goldbeter et al. in [3].*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1752-0509-3-29-S1.pdf>]

### Additional File 2

*archive-codes. Computer codes in C language used for the numerical simulations.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1752-0509-3-29-S2.zip>]

## Acknowledgements

A.C. benefits from a PhD Grant from the French Minister of Research.

## References

- Goldbeter A: *Biochemical Oscillations and Cellular Rhythms: The molecular bases of periodic and chaotic behaviour* Cambridge University Press; 1996.
- Yin JC, Vecchio MD, Zhou H, Tully T: **CREB as a memory modulator: induced expression of a dCREB2 activator isoform enhances long-term memory in Drosophila.** *Cell* 1995, **81**:107-115.
- Goldbeter A, Dupont G, Berridge MJ: **Minimal model for signal-induced Ca<sup>2+</sup> oscillations and for their frequency encoding through protein phosphorylation.** *Proc Natl Acad Sci USA* 1990, **87**(4):1461-1465.
- Dupont G, Goldbeter A: **Protein phosphorylation driven by intracellular calcium oscillations: a kinetic analysis.** *Biophys Chem* 1992, **42**(3):257-270.
- Koninck PD, Schulman H: **Sensitivity of CaM kinase II to the frequency of Ca<sup>2+</sup> oscillations.** *Science* 1998, **279**(5348):227-230.
- Dupont G, Houart G, Koninck PD: **Sensitivity of CaM kinase II to the frequency of Ca<sup>2+</sup> oscillations: a simple model.** *Cell Calcium* 2003, **34**(6):485-497.
- Mettetal JT, Muzzey D, Gómez-Urbe C, van Oudenaarden A: **The frequency dependence of osmo-adaptation in Saccharomyces cerevisiae.** *Science* 2008, **319**(5862):482-484.
- Dolmetsch RE, Xu K, Lewis RS: **Calcium oscillations increase the efficiency and specificity of gene expression.** *Nature* 1998, **392**(6679):933-936.
- Hersen P, McClean MN, Mahadevan L, Ramanathan S: **Signal processing by the HOG MAP kinase pathway.** *Proc Natl Acad Sci USA* 2008, **105**(20):7165-7170.
- Gomez-Urbe C, Verghese GC, Mirny LA: **Operating Regimes of Signaling Cycles: Statics, Dynamics, and Noise Filtering.** *PLoS Comput Biol* 2007, **3**(12):e246.
- Cox CD, McCollum JM, Austin DW, Allen MS, Dar RD, Simpson ML: **Frequency domain analysis of noise in simple gene circuits.** *Chaos* 2006, **16**(2):026102.
- Shankaran H, Wiley HS, Resat H: **Receptor downregulation and desensitization enhance the information processing ability of signalling receptors.** *BMC Syst Biol* 2007, **1**:48.
- Shankaran H, Resat H, Wiley HS: **Cell surface receptors for signal transduction and ligand transport: a design principles study.** *PLoS Comput Biol* 2007, **3**(6):e101.
- Locasale JW: **Signal duration and the time scale dependence of signal integration in biochemical pathways.** *BMC Syst Biol* 2008, **2**(1):108. ArXiv:0802.2683v1 [q-bio.MN]
- Smolen P, Baxter DA, Byrne JH: **Frequency selectivity, multistability, and oscillations emerge from models of genetic regulatory systems.** *Am J Physiol* 1998, **274**(2 Pt 1):C531-C542.
- Li Y, Goldbeter A: **Frequency specificity in intercellular communication. Influence of patterns of periodic signaling on target cell responsiveness.** *Biophys J* 1989, **55**:125-145.
- Segel LA, Goldbeter A, Devreotes PN, Knox BE: **A mechanism for exact sensory adaptation based on receptor modification.** *J Theor Biol* 1986, **120**(2):151-179.
- Alon U: *An Introduction to Systems Biology* Chapman & Hall/CRC; 2006.
- Wolf DM, Arkin AP: **Motifs, modules and games in bacteria.** *Curr Opin Microbiol* 2003, **6**(2):125-134.
- Gouzé J, Sari T: **A class of piecewise linear differential equations arising in biological models.** *Dyn Syst* 2002, **17**:299-316.
- Glass L, Kauffman S: **The logical analysis of continuous, non-linear biochemical control networks.** *J Theor Biol* 1973, **39**:103-129.
- Ventura AC, Sepulchre JA, Merajver SD: **A hidden feedback in signaling cascades is revealed.** *PLoS Comput Biol* 2008, **4**(3):e1000041.
- Heinrich R, Neel BG, Rapoport TA: **Mathematical models of protein kinase signal transduction.** *Mol Cell* 2002, **9**(5):957-970.
- Goldbeter A, Koshland DE: **An amplified sensitivity arising from covalent modification in biological systems.** *Proc Natl Acad Sci USA* 1981, **78**(11):6840-6844.
- Mangan S, Itzkovitz S, Zaslaver A, Alon U: **The incoherent feed-forward loop accelerates the response-time of the gal system of Escherichia coli.** *J Mol Biol* 2006, **356**(5):1073-1081.
- Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB: **Gene regulation at the single-cell level.** *Science* 2005, **307**(5717):1962-1965.
- Kholodenko BN: **Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades.** *Eur J Biochem* 2000, **267**(6):1583-1588.
- Izhikevich EM: **Bursting.** *Scholarpedia* 2006, **1**(3):1300.
- Barkai N, Leibler S: **Circadian clocks limited by noise.** *Nature* 2000, **403**(6767):267-268.
- Guantes R, Poyatos JF: **Dynamical principles of two-component genetic oscillators.** *PLoS Comput Biol* 2006, **2**(3):e30.
- Gonze D, Halloy J, Goldbeter A: **Robustness of circadian rhythms with respect to molecular noise.** *Proc Natl Acad Sci USA* 2002, **99**(2):673-678.
- Tyson JJ, Csikasz-Nagy A, Novak B: **The dynamics of cell cycle regulation.** *Bioessays* 2002, **24**(12):1095-1109.
- Ciliberto A, Novak B, Tyson JJ: **Steady states and oscillations in the p53/Mdm2 network.** *Cell Cycle* 2005, **4**(3):488-493.
- Dunlap JC, Loros JJ: **The neurospora circadian system.** *J Biol Rhythms* 2004, **19**(5):414-424.
- Atkinson MR, Savageau MA, Myers JT, Ninfa AJ: **Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in Escherichia coli.** *Cell* 2003, **113**(5):597-607.
- Song H, Smolen P, Av-Ron E, Baxter DA, Byrne JH: **Dynamics of a minimal model of interlocked positive and negative feedback loops of transcriptional regulation by cAMP-response element binding proteins.** *Biophys J* 2007, **92**(10):3407-3424.
- Lipan O, Wong WH: **The use of oscillatory signals in the study of genetic networks.** *Proc Natl Acad Sci USA* 2005, **102**(20):7063-7068.
- Geva-Zatorsky N, Rosenfeld N, Itzkovitz S, Milo R, Sigal A, Dekel E, Yarnitzky T, Liron Y, Polak P, Lahav G, Alon U: **Oscillations and variability in the p53 system.** *Mol Syst Biol* 2006, **2**:2006.0033.
- Kandel ER: **The molecular biology of memory storage: a dialog between genes and synapses.** *Biosci Rep* 2004, **24**(4-5):475-522.
- Carew TJ, Pinsker HM, Kandel ER: **Long-Term Habituation of a Defensive Withdrawal Reflex in Aplysia.** *Science* 1972, **175**(4020):451-454.
- Gardner TS, Cantor CR, Collins JJ: **Construction of a genetic toggle switch in Escherichia coli.** *Nature* 2000, **403**(6767):339-342.
- Mazurie A, Bottani S, Vergassola M: **An evolutionary and functional assessment of regulatory network motifs.** *Genome Biol* 2005, **6**(4):R35.
- Perlmutter JS, Mink JW: **Deep brain stimulation.** *Annu Rev Neurosci* 2006, **29**:229-257.
- Henning J, Koczan D, Glass A, Karopka T, Pahnke J, Rolfs A, Benecke R, Gimsa U: **Deep brain stimulation in a rat model modulates TH, CaMKII $\alpha$  and Homer1 gene expression.** *Eur J Neurosci* 2007, **25**:239-250.
- Elowitz M, Leibler S: **A synthetic oscillatory network of transcriptional regulators.** *Nature* 2000, **403**:335-338.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
http://www.biomedcentral.com/info/publishing\_adv.asp



# Chapitre 3

## Boucles d'ADN et régulation génétique

Dans ce chapitre, nous faisons une revue des différents systèmes de régulation qui font intervenir des boucles d'ADN au sein de régions de régulation chez la bactérie *Escherichia coli*. Nous commençons par rappeler les systèmes déjà connus et bien étudiés d'opérons où intervient la formation de boucle d'ADN. Nous rassemblons les différentes méthodes expérimentales qui ont été utilisées pour mettre en évidence de telles boucles. Nous réunissons ensuite les différents avantages que pourrait procurer la formation de boucle sur la régulation transcriptionnelle. Certains d'entre eux ayant été proposés sur le plan théorique, nous faisons le point sur ce qui a été montré expérimentalement et sur ce qu'il reste à prouver en proposant des expériences si nécessaire. Nous pensons que d'autres boucles sont à trouver dans un génome comme celui de la bactérie *Escherichia coli*. Nous avons par conséquent interrogé les bases de données (*RegulonDB*) pour identifier quelles seraient les régions de régulation qui auraient de grandes chances de présenter de telles boucles d'ADN, compte tenu de plusieurs caractéristiques attendues portant par exemple sur le type de régulation ou sur la disposition des sites. Après plusieurs filtrages, une liste d'une trentaine de gènes est alors proposée dont nous pensons qu'elle contient des candidats les plus probables pour comporter une boucle d'ADN au sein de leur région de régulation et qui mériteraient une démarche expérimentale.

### 1 Cas connus et bien caractérisés de boucle d'ADN

Nous commençons dans cette section par rappeler brièvement les systèmes connus de boucles d'ADN intervenant dans des régions de régulation et nous décrivons de quelle façon les boucles sont agencées. Les cinq systèmes des opérons *lac*, *gal*, *deo*, *ara* et *glnA* ont déjà été répertoriés dans trois travaux de revue [1, 106, 151] parus au début des années 1990. Nous allons répertorier les cas qui ont depuis été découverts et étudiés. Nous distinguons deux types de configuration : les boucles d'ADN à l'origine d'une activation transcriptionnelle et les boucles d'ADN à l'origine d'une répression transcriptionnelle.

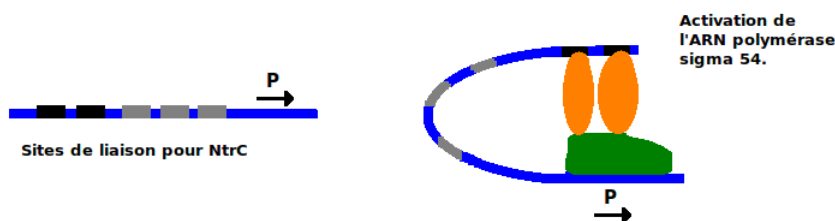


FIG. 3.1 – Mécanisme de boucle d’ADN qui permet l’activation de l’opéron *glnA* d’*Escherichia coli*. Les sites en noir sont des sites forts pour la forme phosphorylée du facteur de transcription NtrC. La boucle facilite la formation du complexe ouvert de l’ARN polymérase qui est de type  $\sigma^{54}$ .

### 1.1 Boucle d’ADN à l’origine d’une activation transcriptionnelle

Les systèmes faisant intervenir une boucle d’ADN pour une activation transcriptionnelle sont assez bien caractérisés. Ils font intervenir une ARN polymérase à la sous-unité  $\sigma^{54}$ . Comme cette ARN polymérase se lie faiblement sur son site promoteur, des protéines supplémentaires sont nécessaires pour faciliter la fixation de l’ARN polymérase et la formation du complexe ouvert. Ces protéines sont appelées EBP pour “Enhancer Binding Proteins”. En effet, ces protéines agissent à distance du site promoteur, en général à une centaine de paires de base et indépendamment du sens du site de liaison, ce qui rappelle les enhancers des génomes eucaryotes [136]. Nous illustrons le mécanisme sur l’exemple de l’opéron *glnA* qui a été le premier système de ce type à avoir été mis en évidence et le plus étudié.

#### Opéron *glnA*

L’opéron *glnA* est activé lorsque le milieu est déficient en azote. Les gènes de cet opéron codent pour des protéines impliquées dans le métabolisme de la glutamine qui peut être utilisée comme donneuse d’azote dans les biosynthèses de nombreux composés. Le facteur de transcription NtrC qui entre en jeu dans ce système interagit avec deux sites de liaison forts situés à 110 pb et 140 pb en amont du site promoteur, ce qui active la transcription en facilitant la fixation de l’ARN polymérase et la formation du complexe ouvert (voir Figure 3.1). L’affinité de l’ARN polymérase de type  $\sigma^{54}$  pour son site promoteur est faible et elle a donc besoin des protéines NtrC phosphorylées qui vont l’aider à se fixer.

Il existe plusieurs autres systèmes de ce type chez la bactérie *Escherichia coli*, basés sur le même mécanisme. Nous les regroupons dans le tableau 3.1 en donnant les informations importantes de chaque système.

### 1.2 Boucles d’ADN à l’origine d’une répression

Plusieurs cas d’opérons réprimés grâce à la formation d’une boucle d’ADN sont aujourd’hui bien connus et étudiés. Le mécanisme est le suivant : des facteurs de transcription bivalents (ex : tétramères) se lient sur deux sites de liaison espacés de

Nom opéron	nom EBP	disposition des sites	IHF
astC	NtrC	3 sites NtrC à -233, -253, -275 pb	non
atoD	AtoC	2 sites AtoC à -118.5, -139.5 pb	oui
fdhF	FlhA	1 site FlhA à -107.5 pb	non
glnH	NtrC	5 sites NtrC entre -79 et -152 pb	oui
glnK	NtrC	1 site NtrC à -87 pb	non
hycA	FhlA	1 site FhlA à -98.5	oui
hyfA	FhlA	1 site FhlA à -130 pb	non
hypA	FhlA	1 site FhlA à -730 pb	oui
nac	NtrC	1 site NtrC à -152 pb	non
norV	NorR	3 sites NorR à -132, -111 -87 pb	oui
prpB	PrpP	2 sites PrpR à -147.5 et -113 pb, présence d'un site CRP au milieu	non
pspA	PspF	2 sites PspF à -99, -119 pb	oui
pspG	PspF	1 site PspF à - 87 pb	oui
rutA	PhoP	1 site PhoP à -104 pb	non
ygjG	NtrC	2 sites NtrC à -193 et -171 pb, présence de Fis au milieu	oui
zraP	ZraR	1 site ZraR à -100.5 pb	non
zraS	ZraR	1 site ZraR à -92.5 pb	non

TAB. 3.1 – Systèmes de type enhancer chez *Escherichia coli* à l'origine d'une activation transcriptionnelle par boucle d'ADN. La première colonne donne le nom de l'opéron régulé par ce type de boucle. La deuxième colonne donne le nom de la protéine responsable de la boucle. La troisième colonne donne des informations concernant la position des sites de liaison impliqués par rapport au site promoteur. La quatrième colonne indique si la protéine IHF est présente dans la région de régulation car il a été montré un rôle de première importance de cette protéine dans ces systèmes [37].

plusieurs dizaines voire centaines de paires de bases en bouclant la molécule d'ADN comprise entre les sites de liaison. La formation de la boucle bloque l'accès de l'ARN polymérase à son site promoteur et inhibe la transcription.

### Opéron *lac*

L'opéron *lac*, est depuis le travail pionnier de Jacob et Monod [70], l'un des systèmes biologiques modèles pour la régulation génétique. Il est intéressant de remarquer que la formation d'une boucle d'ADN est au cœur du processus de régulation de ce système. La boucle de cet opéron a été l'une des plus étudiées notamment par l'équipe du biologiste Müller-Hill. La région de régulation se compose de 3 sites de liaison pour le répresseur LacI [114]. Un site principal qui a la plus forte affinité pour le répresseur est entouré de deux sites, dits auxiliaires (dont l'affinité est plus faible), avec lesquels une boucle d'ADN de respectivement 401 pb et 92 pb peut se former [121]. Les différentes configurations de boucles d'ADN de ce système sont représentées sur la Figure 3.2.

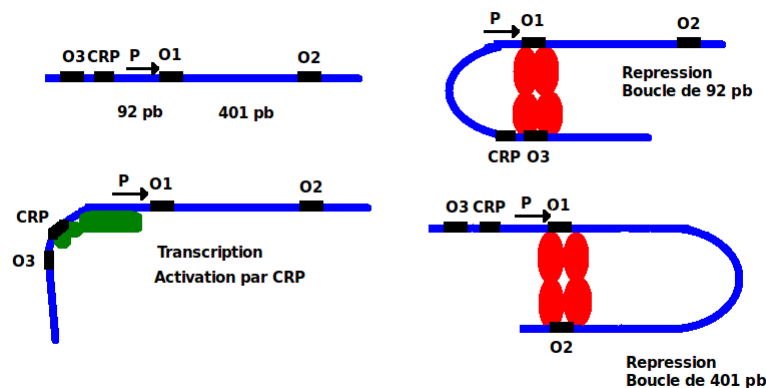


FIG. 3.2 – Système de boucle de l'opéron *lac*.

L'un des sites auxiliaires (O3 sur la Figure 3.2) entre en compétition avec un site pour CRP, ce qui permet une désactivation. Il est en effet nécessaire que CRP se lie sur ce site de liaison pour qu'il y ait activation transcriptionnelle de l'opéron (les séquences en -10 et -35 du site de fixation de l'ARN polymérase étant faibles, la protéine CRP intervient en incurvant l'ADN pour augmenter le niveau de fixation de l'ARN polymérase sur ce promoteur).

Il est également intéressant de noter la présence de 6 sites de liaison pour la protéine type histone H-NS à l'intérieur du segment d'ADN qui peut boucler (non représentés sur la figure). La protéine H-NS (pour histone-like nucleoid structuring<sup>1</sup>) est capable de condenser l'ADN in vitro et in vivo et se lie préférentiellement sur des séquences riches en AT ou des séquences d'ADN courbées [11].

### Opéron *gal*

La région de régulation de l'opéron *gal* comporte deux sites de liaison pour le répresseur galR séparés par 114 paires de bases (voir Figure 1.2). La boucle d'ADN

<sup>1</sup>bien qu'il n'y ait pas d'homologie de séquences avec les protéines d'histones.



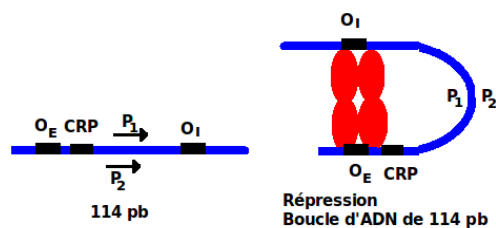


FIG. 3.3 – Boucle d'ADN dans l'opéron gal.

bloque l'action de l'ARN polymérase probablement en la gardant piégée et inhibe ainsi la formation du complexe ouvert. L'opéron *gal* se compose de trois gènes structuraux : les gènes *galE*, *galT* et *galK*. Ils codent tous trois pour des enzymes qui catalysent des réactions qui entrent en jeu dans le métabolisme du galactose. Lorsque du galactose est présent dans le milieu et le glucose absent, l'opéron *gal* est induit. En effet, en présence de galactose, GalR subit un changement conformationnel qui l'empêche de se fixer sur ses sites de liaison présents sur l'ADN. Il faut noter également la présence de 4 sites de liaison pour le facteur de transcription H-NS ainsi que deux sites de liaison pour la protéine HU (également une histone-like protéine qui peut notamment agir sur le sur-enroulement de l'ADN [14]) au sein de la séquence qui boucle (sites de liaison non représentés sur la figure). En effet, le répresseur GalR a des interactions dimère-dimère faibles et il a besoin de la liaison de la protéine HU qui joue un rôle architectural dans la formation de la boucle [153]. La faible interaction dimère-dimère dans la boucle à l'origine de la répression rend le système très réactif.

Enfin, les deux sites pour GalR peuvent également être liés par le facteur de transcription GalS. En effet, GalS est fortement homologue dans sa séquence d'acides aminés au facteur de transcription GalR. Il semblerait qu'ils agissent ensemble et soient capables d'interférences pour réguler l'expression des gènes du régulon *gal* [44]. Ce système présente beaucoup de similarités avec l'opéron *lac*.

### Opéron *ara*

L'opéron *ara* est à notre connaissance le premier opéron où a été mise en évidence la formation d'une boucle d'ADN [34].

En absence d'arabinose, une boucle d'ADN de 210 paires de bases se forme entre les sites  $O_2$  et  $I_1$ . La répression s'exerce sur les deux promoteurs  $P_{BAD}$  et  $P_C$ . Lorsque l'arabinose est présent dans le milieu, il rompt la boucle précédente et la protéine AraC vient se fixer sur les demi sites  $I_1$  et  $I_2$ , et active le promoteur  $P_{BAD}$  (voir Figure 3.4). AraC peut donc être activateur ou répresseur pour ce système. Des expériences réalisées *in vitro* ont montré que le facteur de transcription CRP déstabilise la boucle d'ADN entre les sites de liaison  $O_2$  et  $I_1$  [97]. L'obstruction à la formation de la boucle d'ADN serait causée par la courbure de l'ADN induite par la liaison de CRP. Enfin, des expériences [95] ont montré que l'état de surenroulement de l'ADN était nécessaire pour la formation de la boucle chez ce système.

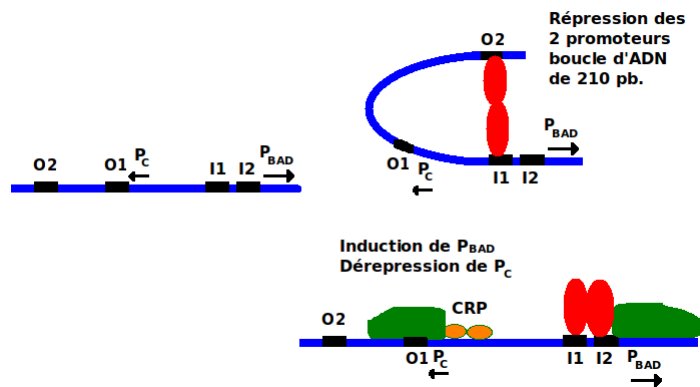


FIG. 3.4 – Système de boucle d’ADN au sein de la région de régulation de l’opéron *ara*.

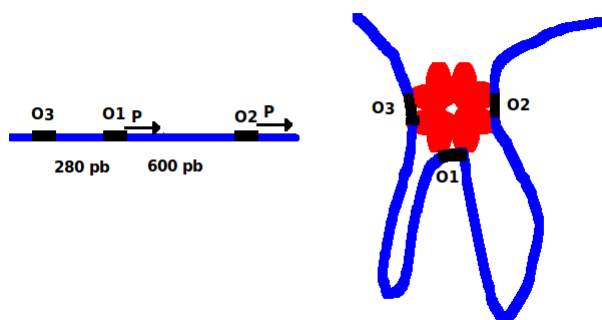


FIG. 3.5 – Système de double-boucles qui peut se former au sein de la région de régulation de l’opéron *deo*.

### Opéron *deo*

La région de régulation de l’opéron *deo* se compose de trois sites de liaison pour le répresseur DeoR. Deux d’entre eux chevauchent les deux sites promoteurs présents dans la région, ils sont espacés de 600 pb. Le troisième site est situé à 280 pb du premier promoteur (voir Figure 3.5). Ce système a la particularité de pouvoir former une double-boucles qui exerce la répression la plus forte sur le système. En effet, DeoR pouvant être sous forme d’un octamère, une seule protéine peut se lier à trois sites opérateurs. Cette double-boucles a bien été observée par microscopie électronique [8]. Enfin, nous retrouvons également dans ce système deux sites de liaison pour le facteur de transcription CRP qui joue un rôle d’activateur pour le deuxième promoteur (sites non montrés sur la figure).

### Opérons *nagE* et *nagB*

Les deux opérons divergents *nag* (*nagE* et *nagB*) sont également réprimés grâce à un mécanisme de formation d’une boucle d’ADN longue de 94 pb [127] entre deux sites opérateurs pour le facteur de transcription NagC (figure non montrée). Les deux sites de liaison pour NagC chevauchent les sites de liaison pour l’ARN

polymérase ce qui empêche la fixation de celle-ci. Les gènes de l'opéron interviennent dans le métabolisme du N-acetylglucosamine-6-phosphate (GlcNAc-6P) qui peut être utilisé comme source de carbone par la bactérie. Il est à noter que le facteur de transcription CRP a un rôle double dans la régulation : il est nécessaire pour l'activation de l'opéron mais il peut aussi avoir un rôle de répresseur pour ce système. En effet, la liaison de CRP stabilise la boucle d'ADN *in vitro*. L'ADN de la région intergénique *nagE-B* est une région intrinsèquement courbée avec la courbure centrée sur le site pour CRP. Il a été montré que le déplacement de 6 pb de ce site de liaison engendre une dérépression complète des deux opérons [128]. Étant donné qu'il n'y a pas d'interaction détectée entre CRP et NagC, une fonction architecturale pour CRP a été proposée où la protéine stabiliserait la courbure intrinsèque présente et donc la boucle.

### Opérons *glpA* et *glpT*

Le facteur de transcription GlpR fait partie de la même famille que le facteur DeoR. Il a été montré qu'il était à l'origine de boucle d'ADN dans au moins deux opérons qu'il régule. Les deux opérons *glpA* et *glpT* sont placés de manière divergente sur le génome d'*Escherichia coli* et sont régulés par une même boucle d'ADN [87, 180]. La boucle d'ADN à l'origine de la répression se forme à partir de plusieurs sites de liaison, certains placés loin en aval du site promoteur (configuration similaire à celle de l'opéron *gal*, figure non montrée). Les opérons sont induits par la présence du sn-glycérol 3-phosphate (glycérol-P) et jouent un rôle dans son métabolisme. Enfin, nous pouvons remarquer qu'il a été montré que la protéine IHF joue un rôle important dans la formation de la boucle d'ADN. En effet, la répression est diminuée de moitié dans les souches déficientes de IHF. Cet effet n'est apparent que lorsque les opérateurs loin en aval sont présents [180].

## 2 Techniques expérimentales pour mettre en évidence des boucles d'ADN

Une première expérience qui est réalisée pour mettre en évidence une boucle d'ADN est de muter l'un ou les deux sites de liaison impliqués afin d'observer un éventuel effet sur la régulation exercée. Si les effets de mutation ne sont pas simplement additifs, alors il y a coopérativité entre les sites. Ceci est visible dans les expériences réalisées par l'équipe de Müller-Hill où le fait de muter un site de liaison auxiliaire entraîne que la répression se fait encore à moitié ou au tiers. Le fait d'enlever les deux sites auxiliaires supprime presque entièrement la répression. Par ailleurs, des expériences d'empreintes à la DNase I permettent également de mettre en évidence une coopérativité entre sites de liaison : l'expérience consiste à regarder si un site de liaison est plus facilement lié si un autre site de liaison est présent sur le brin d'ADN, qui pourrait favoriser sa liaison (il est parfois possible aussi que la séquence d'ADN qui boucle entre les sites de liaison soit partiellement protégée de la DNase I c'est-à-dire que l'on peut observer un pattern de segmentation tous les 10-11 pb avec des zones plus protégées entre [126]). Toutefois, ces expériences peuvent seulement montrer qu'il y a une coopérativité entre les sites de liaison. C'est une condition nécessaire mais pas suffisante pour l'existence d'une boucle d'ADN.

Nous rassemblons dans cette section toutes les méthodes expérimentales qui à notre connaissance, permettent la mise en évidence de boucle d'ADN au sein de régions de régulation. Ces approches proviennent de plusieurs domaines différents faisant appel à des méthodes de génétique moléculaire (mutagénèse dirigée, PCR ...) ou à des méthodes plus orientées vers la biophysique (techniques de microscopie et fluorescence). Nous donnons ces différentes méthodes par ordre croissant de difficulté de mise en œuvre pour un laboratoire de biologie moléculaire.

### ① Déphaser les sites de liaison sur le tour de l'hélice d'ADN

Une première méthode qui fait appel à des techniques de biologie moléculaire est de rajouter des paires bases dans la séquence susceptible de boucler pour déphaser, sur le tour de la double hélice (qui fait environ 10.5 pb pour les bactéries<sup>2</sup>), les sites de liaison qui seraient impliqués et empêcher ainsi l'éventuelle formation de boucle. En effet, si les sites se retrouvent sur des côtés opposés sur l'hélice d'ADN, une énergie de torsion importante doit être fournie pour la formation du complexe (in vitro, 4 kcal/mol sont requis pour tordre d'un demi-tour une molécule linéaire d'ADN [64]). Si la boucle est au cœur du processus de régulation, un effet doit être visible lorsque l'on introduit la moitié du tour de la double hélice soit 5 paires de base. C'est la méthode qui avait été utilisée par l'équipe de Robert Schleif pour mettre en évidence la boucle d'ADN au sein de la région de régulation de l'opéron *ara* [34]. Dans ce système, la répression est diminuée de moitié dans les cas où un nombre impaire de demi-tours de l'hélice d'ADN est introduit mais la répression est normale pour des insertions de 0, 11 ou 31 paires de bases.

### ② Expériences de band shift

Une technique issue également de la biologie moléculaire et relativement ancienne est celle appelée "band shift". Elle permet d'une façon générale de déterminer si une protéine a une affinité forte pour une séquence d'ADN. En faisant migrer dans un gel de polyacrylamide des mélanges préalablement incubés d'ADN et de protéines régulatrices, on peut comparer les vitesses de migration des brins d'ADN éventuellement complexés avec les protéines en regardant les positions sur le gel. Les complexes protéines-ADN migrent dans le gel de polyacrylamide plus lentement que des d'ADN libres. De plus, les complexes protéines-ADN qui font intervenir une boucle d'ADN migrent encore plus lentement qu'un complexe sans boucle. L'équipe de Müller-Hill a effectué ce test dans [121] pour ajouter un argument en faveur de la boucle d'ADN au sein de l'opéron *lac*. Cette technique permet seulement d'apporter une vérification supplémentaire et il faut que la protéine soit purifiée.

### ③ Remplacer les sites impliqués par des sites du répresseur LacI

Une autre méthode pour prouver que deux sites de liaison interviennent dans une formation de boucle d'ADN est de les remplacer par deux sites de liaison du répresseur LacI (site principal O1 par exemple qui a la plus forte affinité). Puis l'on regarde si la répression sur le système testé en présence du répresseur LacI peut se

---

<sup>2</sup>cette valeur peut cependant varier notamment grâce à la présence d'histone like protéines.

faire tout aussi fortement. Si le système agit bien par la formation d'une boucle, on doit retrouver une valeur importante pour la répression, et similaire au cas sauvage avec les sites inchangés et le répresseur associé. Cette démarche a par exemple, été utilisée pour démontrer l'existence de la boucle de l'opéron *gal* [59]. Dans ce travail, les auteurs ont construit plusieurs régions de régulation différentes en remplaçant un ou les deux sites de liaison pour le répresseur GalR par un site pour le répresseur LacI. De mêmes valeurs dans l'expression génétique ont été trouvées pour la région de régulation du système sauvage en présence de GalR et pour la région de régulation où les deux sites avaient été remplacés par des sites de liaison pour LacI, en présence de ce dernier.

#### ④ Muter les acides aminés de la protéine régulatrice pour empêcher l'oligomérisation

Un autre moyen pour déceler la formation d'une boucle d'ADN qui se forme par une oligomérisation des protéines régulatrices est d'empêcher celle-ci puis de regarder si la régulation est affectée. C'est notamment l'approche menée par l'équipe de Müller-Hill dans [121] pour la boucle de l'opéron *lac*. Ces auteurs empêchent la tétramérisation du répresseur LacI et peuvent observer par exemple une différence d'un facteur 60 entre les valeurs de répression du système pouvant tétramériser et du système ne le pouvant pas (à des concentrations de 200 sous-unités de LacI par cellule). Plus précisément, dans cet exemple, la mutation s'était faite en introduisant un codon stop à la position 331 de la séquence codante du gène *lacI*. Cette modification enlève les résidus C-terminal du répresseur LacI qui permettent la tétramérisation [3].

Cette méthode nécessite donc une bonne connaissance de la protéine régulatrice pour savoir où se trouvent les sites qui permettent l'oligomérisation et ainsi faire les mutations au bon endroit sans modifier le pouvoir régulateur de la protéine ou sa capacité à se lier sur l'ADN. Les sites d'interaction entre protéines ne sont pas encore bien connus et il est difficile de trouver des motifs d'interactions qui pourraient se retrouver d'une protéine à l'autre même au moyen par exemple de la bioinformatique.

Il est intéressant de remarquer que l'on peut cumuler les deux approches précédentes. En effet, sur un système dont on voudrait mettre en évidence une boucle d'ADN entre deux sites, on peut les remplacer par deux sites de de liaison pour le répresseur LacI et comparer les valeurs de répression dans une souche où le répresseur LacI peut se tétramériser et une souche où le répresseur LacI reste sous forme dimérique. C'est une des méthodes employées pour valider la présence de boucle pour l'opéron *gal* [101].

#### ⑤ Expériences d'encaténation de plasmides

Une autre méthode basée sur des techniques de biologie moléculaire utilise l'encaténation de plasmides. Elle consiste à placer les parties d'ADN qui interagiraient par boucle d'ADN sur deux plasmides différents (par exemple les sites de fixation pour la protéine régulatrice sur un plasmide et le site du promoteur sur un autre plasmide). Ces deux plasmides sont enchaînés à l'aide du système de l'enzyme Résolvase issue du transposon Tn3 (plus précisément, cette opération se fait en partant

d'un seul plasmide initial contenant les deux parties d'ADN à tester et des sites de reconnaissance de la Résolvase). Puis l'on compare avec un mélange où les deux plasmides sont non enchaînés. Si les parties d'ADN interagissent par boucle d'ADN, alors une différence entre les situations plasmides enchaînés et plasmides non enchaînés doit être visible. C'est l'approche développée par l'équipe de Kustu pour démontrer la formation de boucle d'ADN dans l'opéron *glnA* [175] (l'on pourra également consulter la Figure 5.1 du chapitre 5). Cette approche nécessite donc des outils génétiques un peu moins répandus que les méthodes précédentes. A notre connaissance, cette technique n'a pas été appliquée pour mettre en évidence une boucle d'ADN à l'origine d'une répression transcriptionnelle.

### ⑥ Microscopie électronique pour l'observation directe de boucle d'ADN

Une preuve directe de la formation de boucle est la réalisation d'un cliché par microscopie électronique de la région d'ADN en présence des protéines régulatrices. Cette technique ne nous semble pas hors de portée de la majorité des laboratoires s'intéressant à la régulation génétique. En effet dès 1986, des clichés de la boucle qui se forme dans le système du phage  $\lambda$  ont été pris par l'équipe de Griffith [57]. La technique a depuis bénéficié de plusieurs améliorations et a par exemple été utilisée pour visualiser la boucle au sein de l'opéron *glnA* [163]. Cette technique se passe *in vitro* et nécessite que la protéine régulatrice soit purifiée. Par ailleurs, elle nécessite des expériences de contrôle et une statistique précise. En effet, des boucles d'ADN peuvent toujours se former avec une probabilité non nulle mais sans une signification biologique [151].

### ⑦ Techniques de TPM

Une technique, mise en place et développée beaucoup plus récemment [150], et faisant appel à des concepts issus de la physique statistique, est la technique dite TPM pour "Tethered Particle Motion". Une molécule d'ADN est fixée par l'une de ses extrémités sur un socle (en général une plaque de verre). A son autre extrémité qui est libre, une bille (par exemple une microbille de polystyrène de 0.23 microns de diamètre ou une bille fluorescente) est accrochée. Le mouvement brownien de la bille est alors suivi au microscope pendant une certaine durée. Si la molécule d'ADN se boucle, le mouvement de la bille n'est plus le même : l'espace exploré (dans le plan transverse à l'ADN) par la bille est plus restreint et ses propriétés statistiques (amplitude des fluctuations) ont changé [13]. Cette technique permet par ailleurs de différencier différentes configurations de boucles d'ADN [167]. En effet pour une même disposition des sites de liaison sur l'ADN, plusieurs géométries de boucles d'ADN sont possibles [186], faisant intervenir des énergies de courbure ou de torsion différentes, mais également des conformations particulières de la protéine régulatrice à l'origine de la boucle. La méthode TPM peut également fournir des informations sur la durée des états bouclé et non bouclé [131]. Cette technique est donc par les concepts utilisés et les résultats qu'elle peut produire beaucoup plus orientée vers la thermodynamique de la boucle d'ADN que sur la régulation génétique. La technique comme la précédente se passe *in vitro* et nécessite que la protéine régulatrice soit purifiée.

### ⑧ Techniques d'AFM

La technique de microscopie à force atomique (AFM) a été utilisée pour la première fois par l'équipe de Sankar Adhya pour visualiser la boucle d'ADN de l'opéron *gal* [99]. Cette technique permet de visualiser la topographie de la surface d'un échantillon. Elle est basée sur des interactions physiques (répulsion des nuages électroniques par exemple) entre l'échantillon et une pointe dont la hauteur peut être ajustée. La pointe balaie la surface à analyser, la distance entre la pointe et la surface est modifiée par un système de rétroaction. Un ordinateur enregistre cette hauteur et peut ainsi reconstituer une image de la surface [177]. Le travail de Lyubchenko et al. [99] a étudié des échantillons de plasmides contenant la région de régulation de *gal*. Ces auteurs retrouvent les asymétries sur les figures en forme de huit des plasmides qui correspondent à celles attendues pour la formation de la boucle entre les deux sites opérateurs.

Nous avons donc réuni dans cette section plusieurs méthodes pour mettre en évidence une boucle d'ADN au sein de régions de régulation. Il est important de garder à l'esprit que pour une mise en évidence claire de la boucle, il faut cumuler plusieurs méthodes pour valider l'interaction physique et l'impact sur la régulation.

## 3 Avantages procurés par les boucles d'ADN

Le mécanisme de régulation par formation de boucle d'ADN pourrait conférer d'intéressants avantages à la bactérie ou à la cellule. Nous réunissons les plus importants dans cette section. Certains de ces avantages ont été présentés et compris par les biologistes et d'autres ont été prédits par la physique statistique mais n'ont pas encore obtenu de vérification expérimentale. Nous faisons le point sur ce qui a été prouvé expérimentalement et proposons des expériences si nécessaire. Les avantages présentés ici concernent prioritairement les boucles d'ADN à l'origine d'une répression. En effet, le cas de la boucle d'ADN à l'origine d'une activation transcriptionnelle a été moins étudié sur le plan théorique, la plupart des travaux actuels s'étant basés sur le paradigme de la boucle d'ADN de l'opéron *lac*.

### Concentration locale accrue

Le premier avantage que procure la formation de boucle d'ADN est celui de *concentration locale accrue*, concept développé notamment par le biologiste Müller-Hill [113,115]. La boucle permet de créer une concentration effective plus importante au niveau de la région de régulation ce qui permet une répression très importante avec un nombre restreint de protéines régulatrices. La boucle d'ADN permet en quelque sorte de garder piégée une de ces protéines au voisinage de la région de régulation. Si  $p$  est la probabilité que la protéine se détache d'un site de liaison, alors grâce à la formation d'une boucle d'ADN faisant intervenir 2 sites de liaison, la probabilité pour que la protéine se détache devient  $p^2$ . Il est donc possible de diminuer d'un ordre de grandeur la constante de dissociation de la protéine sur le promoteur.

On peut dire que ce premier avantage a été montré expérimentalement par les expériences de l'équipe de Müller-Hill. En effet, lorsque les deux sites de liaison

auxiliaires sont inactivés et que toute boucle d'ADN est impossible, la répression est diminuée d'un facteur d'environ 10 par rapport à une configuration où une boucle d'ADN est encore possible (un site auxiliaire encore actif), ceci pour une concentration de 50 tétramères par cellule. Il n'y a pratiquement plus de différences entre les deux configurations précédentes à des concentrations de 900 tétramères par cellule car à de telles concentrations l'effet de concentration locale ne se fait plus sentir [120].

Ce point a également été démontré par la suite sur le plan théorique par le travail de J. Vilar et S. Leibler [173]. En se plaçant dans le formalisme de la physique statistique (ensemble canonique), ils calculent les probabilités des différents états du système (bouclé ou non bouclé). Les équations font apparaître que l'ajout d'un site auxiliaire est équivalent à une concentration effective plus forte du répresseur.

### Moins de fluctuations dans le temps dans la production des protéines

Dans le travail précédemment cité [173], il est également prédit un avantage important qui est celui de présenter moins de fluctuations temporelles dans l'expression génétique du système régulé. Ceci proviendrait du fait que la boucle d'ADN introduit naturellement une échelle de temps rapide dans les changements d'états de la transcription : elle permet des changements beaucoup plus rapides entre les états de forte et faible activité. Si ce changement entre les niveaux de transcription forts et faibles se faisait lentement, il y aurait de longues périodes d'état inactif ou actif, ce qui serait à l'origine d'importantes fluctuations temporelles. Le changement rapide vient du fait que la durée pour qu'un répresseur soit recapturé par l'opérateur principal est beaucoup plus courte que la durée nécessaire pour qu'un nouveau répresseur en solution se lie sur l'opérateur [145].

Ceci a été montré numériquement mais non expérimentalement. Ce point s'appuie sur des simulations numériques stochastiques avec l'algorithme de Gillespie [46] qui est couramment utilisé dans la modélisation de réseaux de régulation (cf. Chapitre 1). La comparaison entre les simulations stochastiques d'un système pouvant agir par boucle d'ADN et celles d'un système agissant grâce à un unique opérateur montrent des distributions du nombre de protéines produites plus resserrées dans le premier système.

Cette prédiction importante n'a pas encore été directement testée sur le plan expérimental, à notre connaissance. En effet, il faudrait quantifier l'expression protéinique en cellule unique au cours du temps d'un système faisant intervenir une boucle et d'un système constitué d'un simple opérateur puis comparer leur expression. Ces techniques nécessitent des moyens expérimentaux assez lourds. Cependant, les expériences réalisées sur l'opéron *lac* dans le travail de Yu et al. [185] apportent des résultats qui iraient dans le sens de cette prédiction théorique. En effet, les auteurs ont utilisé des techniques de protéines fluorescentes et de microscopie à cellule unique pour étudier, molécule par molécule, l'expression de l'opéron *lac* à l'état réprimé. Ils montrent que l'expression protéinique de l'opéron se déroule sous forme pulsatile comprenant des pulses de protéines suivis de longues périodes où le système n'est pas exprimé. Plus précisément, il apparaît en moyenne 1.2 pulse de protéines par cycle cellulaire et chaque pulse comporte en moyenne 4.2 protéines. Ces résultats semblent s'accorder avec la vision proposée dans l'article de Vilar et Leibler.



### Moins de variabilité sur le niveau d'expression protéique entre les cellules

Le mécanisme de boucle d'ADN permettrait de diminuer la sensibilité de la transcription aux variations dans le nombre de protéines régulatrices. Ceci a été proposé, toujours d'après le travail de Vilar et Leibler, grâce à la mise en équation et au formalisme de la physique statistique du système thermodynamique de la boucle d'ADN. En effet, la boucle d'ADN engendre une relation entre le nombre de protéines régulatrices et le niveau de répression qui n'est pas linéaire et qui atteint un plateau. Dans cette zone, les fluctuations ont moins d'effet sur le niveau de répression qu'un opérateur unique pour lequel la relation entre la valeur de répression et le nombre de répresseurs est linéaire. Le mécanisme de boucle d'ADN intervenant dans les processus de régulation pourrait ainsi produire des populations de cellules plus homogènes dans le nombre de protéines.

Ceci a été proposé théoriquement mais n'a pas été à ce jour confronté sur le plan expérimental.

Nous proposons ici des démarches expérimentales qui permettraient de tester cette prédiction. Une première possibilité est d'employer les moyens techniques de microscopie à fluorescence comme ceux employés par Besckei et al. [16] pour prouver qu'une boucle rétro-négative permet une meilleure stabilité dans l'expression génétique (cf. Chapitre 1). Le principe de la démarche serait le même, il s'agirait de construire des systèmes génétiques qui exercent une répression de même intensité mais dont l'un est régi par une boucle d'ADN et l'autre par un opérateur unique de plus forte affinité. Grâce au comptage des cellules et à la mesure de leur fluorescence, il serait possible d'obtenir des histogrammes donnant le nombre de cellules qui ont telle intensité de fluorescence et d'observer si le système qui est régulé par une boucle d'ADN a bien une distribution plus serrée. Par ailleurs, pour tester cette hypothèse, un cytomètre de flux pourrait être suffisant. D'après les prédictions théoriques, un système avec un opérateur unique entraîne des répressions pouvant aller du simple au triple alors qu'un système bouclé présente des répressions qui ne varieraient que de 25 %. Une telle différence pourrait être décelée sur des histogrammes de distribution de fluorescence issus d'un cytomètre de flux, technique plus accessible que la microscopie utilisée dans le travail de Besckei et al.

### Recherche plus rapide d'une cible par un facteur de transcription

Cette proposition théorique récente est issue du travail de Li et al. [92]. Dans ce travail, les auteurs réétudient le modèle de Berg et von Hippel qui propose un scénario pour la recherche d'une cible sur l'ADN par un facteur de transcription. Cette recherche se compose d'une diffusion 3D, puis d'une diffusion 1D sur l'ADN par des interactions non spécifiques ("sliding"). Le facteur de transcription peut aussi passer d'un endroit proche sur le brin d'ADN en faisant des sauts ("hopping") ou bien passer d'un segment d'ADN à un autre ("transfert inter-segmental") [17]. Ils tiennent compte dans ce modèle d'une hypothèse supplémentaire selon laquelle les autres protéines qui se lient sur l'ADN chromosomique, peuvent alors agir comme des barrages aux protéines chercheuses. Dans le cas où deux sites sont séparés par une distance supérieure à la distance de diffusion 1D sur l'ADN, ceux-ci sont vus comme deux cibles indépendantes et le facteur de transcription trouve deux fois plus vite une de ces deux cibles. Par exemple, pour l'opéron *lac*, lorsque le facteur

de transcription arrive sur un site auxiliaire, il sera amené par boucle d'ADN au site principal (qui est responsable de l'obstruction stérique de l'ARN polymérase). Les sites auxiliaires peuvent donc être vus comme des points d'attache qui amènent les facteurs de transcription aux sites principaux sans les faire passer par les processus de diffusion 3D ou 1D où ils peuvent être bloqués par les barrages des autres protéines. Le mécanisme de boucle permettrait donc une plus localisation rapide de la cible par un facteur de transcription.

Ceci n'a pas eu de confrontation sur le plan expérimental car ces résultats sont très récents.

### **Isolation des autres facteurs de transcription**

Un autre avantage des boucles d'ADN, qui n'a pas encore été très développé, est qu'elles pourraient isoler l'expression du gène des effets génomiques du milieu environnant [9,145]. Ce point a été évoqué par Michèle Amouyal dans [9].

Nous pouvons en effet prendre en compte le fait que beaucoup de facteurs de transcription vont se lier de manière non spécifique sur l'ADN ou également sur des séquences qui sont proches de leur séquence consensus mais qui ne leur sont pas destinées (par exemple des facteurs de transcription de la même famille vont se lier sur des sites d'un autre membre du fait qu'ils utilisent les mêmes motifs pour se lier à l'ADN). En bouclant la région d'ADN, il est possible que plusieurs sites deviennent alors inaccessibles. Les interférences dans la régulation transcriptionnelle venant d'autres protéines seraient ainsi diminuées par rapport à un système de régulation simple sans boucle d'ADN plus "ouvert". Ce point n'a cependant pas encore été suffisamment exploré sur le plan théorique. De plus, en formant une boucle d'ADN, cela permettrait également de réprimer plusieurs promoteurs en même temps (la présence de plusieurs promoteurs est en effet très fréquente dans les régions de régulation de procaryotes).

### **Implémentation aisée dans des constructions de Biologie Synthétique**

Par ailleurs, des promoteurs faisant intervenir des boucles d'ADN apparaissent comme de très bons ingrédients pour des constructions de Biologie Synthétique. En effet, le fait d'exercer une répression par boucle d'ADN ne nécessite pas de conditions trop contraignantes dans la construction d'une région de régulation synthétique. Par exemple, pour une répression par l'intermédiaire d'une boucle d'ADN avec le répresseur LacI, il suffit de deux sites opérateurs de bonne affinité et la boucle peut se former. Il n'y a pas de contraintes importantes par exemple sur l'agencement des sites de liaison au sein de la région de la régulation mais il est seulement nécessaire que les sites soient assez espacés. C'est pour cette raison que dans le travail de Atkinson et al. [10] (illustré dans le Chapitre 1), sont utilisés des mécanismes de boucle d'ADN, que ce soit pour la répression ou pour l'activation, dans la construction d'un réseau synthétique.

### **Organisation du polymère d'ADN à l'échelle génomique**

Les boucles d'ADN pourraient se former entre parties éloignées d'un génome et participaient ainsi à sa compaction et à son architecture. Plusieurs biologistes

ou physiciens ont proposé que les facteurs de transcription organiseraient le génome par boucle d'ADN et que cela pourrait être une propriété émergente des interactions moléculaires entre les protéines et l'ADN. [86, 164]. Les boucles pourraient réunir notamment des gènes régulés par les mêmes protéines et ainsi optimiser un génome aux processus transcriptionnels.

Ceci n'est pas démontré sur le plan expérimental. Il y a cependant quelques indices encourageants provenant notamment de travaux issus de la bioinformatique ou du traitement du signal. Cette thèse s'est attaquée à tester cette hypothèse sur le plan expérimental sur un système génétique précis (cf. Chapitre 4).

## 4 Recherche de nouvelles boucles d'ADN

Le but de ce chapitre est de fournir une liste de nouveaux gènes dont leur région de régulation aurait de grandes chances d'abriter une boucle d'ADN à l'origine de répression. Les boucles d'ADN à l'origine d'activation identifiées dans la section 1 sont en effet beaucoup plus aisément décelables puisqu'elles font intervenir une ARN polymérase à la sous-unité  $\sigma^{54}$ , ce qui est assez particulier chez les régions de régulation d'un génome procaryote comme celui d'*Escherichia coli*. On en compte en effet 38 (d'après la base de données RegulonDB) dans tout le génome d'*Escherichia coli*. Parmi ces 38 régions, beaucoup font intervenir une boucle d'ADN à l'origine d'une activation transcriptionnelle que nous avons répertoriées dans la section 1. De plus, les protéines dites Enhancer Binding Proteins (EPB) responsables de ce mécanisme de boucle d'ADN sont assez bien caractérisées [23]. Toutefois, certaines régions de régulation d'*Escherichia coli* ne sont pas encore bien connues. Par exemple, plusieurs régions de régulation font intervenir une ARN polymérase de type  $\sigma^{54}$  mais la connaissance des sites de liaison des facteurs de transcription n'est pas encore complète (par exemple, les régions de régulation des opérons *rtcB* et *ibpB* fonctionnent avec une ARN polymérase  $\sigma^{54}$  et plusieurs sites IHF ont été détectés mais il n'y a pas encore d'information concernant la présence de sites de liaison pour d'autres facteurs de transcription). Il est donc probable que d'autres boucles d'ADN responsables d'activation soient mises en évidence parmi ces régions de régulation.

Les boucles à l'origine de répression chez *Escherichia coli* s'appliquent à des systèmes très variés dont les caractéristiques sont moins aisément identifiables que les boucles à l'origine d'activation transcriptionnelle. Nous allons donc entreprendre la démarche représentée sur la Figure 3.6 pour trouver de nouvelles boucles d'ADN à l'origine de répression. Nous allons dans un premier temps faire ressortir les points communs que nous avons observés dans les différents systèmes de régulation déjà connus et bien étudiés (que nous avons répertoriés dans la section 1.2). Partant de ces points communs, nous allons demander aux bases de données à l'aide d'outils bioinformatiques quelles sont les régions de régulation qui ont des caractéristiques semblables. Nous filtrons les données selon plusieurs critères concernant notamment la disposition des sites de liaison, le type de régulation. Pour chacune de ces régions de régulation ainsi triées, nous regardons d'autres informations qui pourraient jouer dans la formation de boucle d'ADN comme la présence de la protéine CRP ou d'Histone-Like protéines ou bien le contenu en GC de la séquence. Enfin, nous confrontons cette liste de candidats et ces données avec des informations contenues dans les bases bibliographiques.

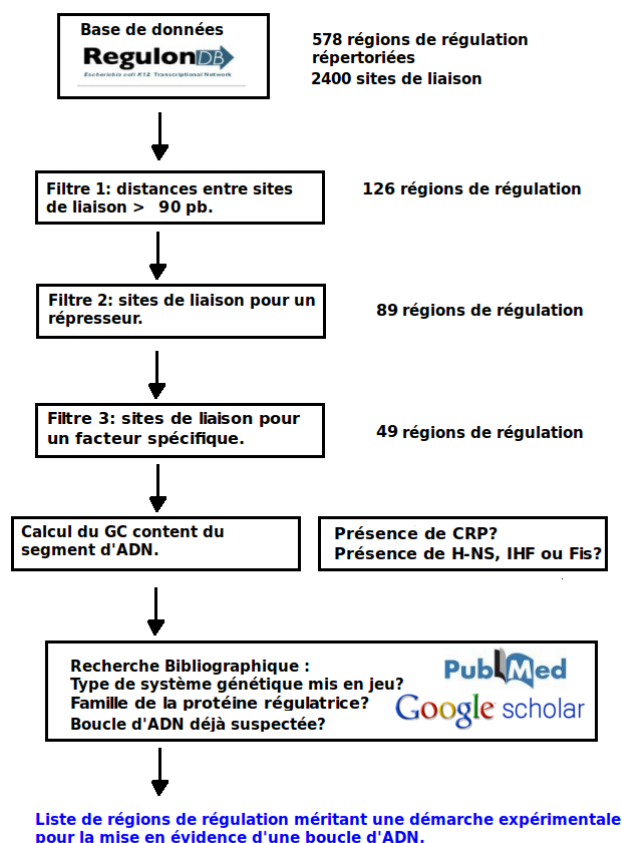


FIG. 3.6 – Schéma de la démarche suivie pour proposer une liste de régions de régulation ayant de grandes chances de comporter une boucle d'ADN.

#### 4.1 Points communs entre les systèmes connus présentant des boucles d'ADN à l'origine d'une répression transcriptionnelle

Si l'on regarde en détail les différents systèmes de régulation qui font intervenir la formation d'une boucle d'ADN à l'origine de répression, nous pouvons repérer des caractéristiques communes. Nous donnons dans cette section les différents points communs que nous avons remarqués. Nous pensons qu'ils peuvent servir pour donner des indices dans la recherche de nouvelles boucles d'ADN au sein de génomes comme celui d'*Escherichia coli*.

##### Point commun sur la géométrie

Pour ce type de boucles, compte tenu de la *longueur de persistance*<sup>3</sup> de l'ADN bactérien, la distance entre les sites de liaison est en général supérieure à 90 pb. En effet, il a été montré, que la longueur de persistance de l'ADN bactérien est

<sup>3</sup>La longueur de persistance d'un polymère est la longueur à partir de laquelle les corrélations dans la direction de la tangente sont perdues. Elle traduit donc la rigidité d'une macromolécule.

autour de 150 pb. En dessous de cette longueur, il est nécessaire de fournir une énergie importante pour boucler la molécule car les effets de rigidité du polymère se font sentir. Pour les différents systèmes connus, les boucles sont majoritairement au-dessus de 90 pb. Par ailleurs, nous pouvons également remarquer que plusieurs de ces systèmes font intervenir un site de liaison qui est placé loin en aval du site promoteur (par exemple, le site auxiliaire 2 de l'opéron *lac* est placé à 411 pb en aval du site promoteur). Ceci peut donc également constituer un indice d'une interaction par boucle d'ADN. Nous soulignons le fait que l'insertion de gène rapporteur au sein d'un gène pour évaluer sa régulation, démarche très courante en microbiologie, pourrait alors empêcher l'action de sites de liaison placés loin en aval du promoteur (et dans l'ORF) et interagissant par boucle d'ADN. Ceci pourrait donc expliquer des contradictions dans les résultats d'expériences.

### Point commun sur la présence de la protéine CRP

La présence d'un ou de plusieurs sites de liaison pour CRP au sein des régions est toujours remarquée pour les différents systèmes de boucles d'ADN déjà connus. Dans une grande majorité de ces systèmes, il joue un rôle d'activateur et l'on peut dire que, d'une manière générale, il s'oppose à la répression exercée par la boucle d'ADN, soit en faisant compétition avec un site de liaison où vient se fixer le répresseur par chevauchement des sites (comme avec le site opérateur O3 de l'opéron *lac*), ou soit en déstabilisant la boucle d'ADN (comme dans le cas de la boucle qui se forme entre les sites  $O_2$  et  $I_1$  dans la région de l'opéron *ara*). Cependant CRP peut également jouer le rôle de stabilisateur comme dans le cas de l'opéron *nag* [128].

### Nature de la séquence

L'un des paramètres importants pour la formation de boucle est le contenu en séquence de la partie d'ADN qui fait la boucle [106]. En effet, les séquences riches en AT sont plus souples que des séquences riches en liaisons GC. Ceci vient du fait que les liaisons entre l'adémine et la thymine font intervenir 2 liaisons hydrogène alors que celles entre la cytosine et la guanine en font intervenir trois. Pour ce type de boucle qui approche la longueur de persistance du polymère d'ADN, l'énergie de courbure entre en jeu et l'influence des séquences sur cette courbure est à prendre en compte. Nous avons déterminé le contenu en paires GC des différentes séquences entrant dans les boucles d'ADN déjà connues, nous les donnons dans le tableau 3.2. Les régions de régulation d'*Escherichia coli* sont déjà plus riches en liaisons AT (GC content : 41.3%) par rapport à l'ensemble du génome (50.8%). Parmi les séquences connues pour boucler, il apparaît que seules celles de l'opéron *galE* (qui fait 114 pb) et de l'opéron *glpA* ont un contenu en GC plus faible que la moyenne des régions de régulation d'*Escherichia coli*. Bien que le contenu de la séquence ne semble pas être un facteur déterminant pour la formation de boucle, cette propriété reste intéressante.

### Présence d'Histone Like Protéines

Les Histones Like Protéines peuvent jouer un rôle sur la topologie du polymère d'ADN notamment à l'échelle d'une région de régulation. Elles peuvent donc avoir

Système	GC content
Opéron <i>lac</i>	53.4% (55.2% pour la boucle de 401pb et 44.4% pour celle de 92pb.)
Opéron <i>gal</i>	38.8%
Opéron <i>ara</i>	44.3%
Opéron <i>deo</i>	49.6% (52.8% pour la boucle de 280pb et 48.0% pour celle de 600pb.)
Opéron <i>nag</i>	45.1%
Opéron <i>glpA</i>	32.8%
Régions de régulation d' <i>Escherichia coli</i>	41.3%
Génome d' <i>Escherichia coli</i>	50.8%

TAB. 3.2 – Tableau du contenu en GC des séquences impliquées dans les boucles d'ADN connues.

un impact sur la formation de boucle d'ADN. Par exemple, nous retrouvons la protéine H-NS dans les systèmes de l'opéron *lac* et de l'opéron *gal*, qui est connue pour jouer un rôle dans la structure du nucléoïde. Plus précisément, cette protéine se lie sur des séquences intrinsèquement courbes et des séquences riches en AT [11]. D'autres protéines sont connues pour jouer sur la topologie de l'ADN. IHF est connue pour pouvoir courber l'ADN jusqu'à des angles de  $160^\circ$  [139]. C'est pour cette raison que nous trouvons des sites de liaison pour cette protéine dans quasiment une région de régulation sur deux pour les systèmes enhancers à l'origine d'activation transcriptionnelle chez *Escherichia coli* répertoriés dans le tableau 3.1. La protéine HU peut également courber de petits segments d'ADN [165]. Elle apparaît au sein de la boucle pour l'opéron *gal* où il a été montré son rôle dans la stabilisation du complexe.

### Point commun sur le rôle physiologique du système régulé

En regardant de plus près quelles sont les protéines codées par les gènes qui sont régulées par un mécanisme de formation de boucle, on peut également remarquer des similitudes. Ce sont en général des systèmes qui doivent réagir à un changement environnemental et passer d'un état de très faible expression à un état de très forte activité. Le système de régulation joue en quelque sorte un rôle d'interrupteur. Par exemple, dans l'état induit, l'opéron *lac* a une activité enzymatique 1400 fois plus forte que dans l'état non induit.

Par ailleurs, il est possible que ces systèmes présentent la propriété de bistabilité au sens de la physique non-linéaire que nous avons définie dans le Chapitre 1. Il a été montré expérimentalement la propriété de multi-stabilité pour l'opéron *lac* dans le travail de Ozbudak et al. [123] grâce à une approche par cellule unique et des techniques de protéines fluorescentes de pointe<sup>4</sup>. Il nous semble tout à fait possible que les autres systèmes de boucle connus ou restant à découvrir possèdent

<sup>4</sup>La difficulté pour mettre en évidence la propriété de bistabilité vient du fait que l'on travaille souvent sur des populations de bactéries ce qui moyenne l'expression génétique. Il est alors difficile de distinguer une population de bactéries dont la moitié serait dans l'état de faible activité et l'autre moitié dans l'état de haute activité d'une

Système	Signal environnemental qui rompt la boucle et induit le système	Fonction des protéines régulées
Opéron <i>lac</i>	Présence du lactose	Protéines impliquées dans le métabolisme et le transport du lactose.
Opéron <i>gal</i>	Présence du galactose	Protéines impliquées dans le métabolisme du galactose.
Opéron <i>ara</i>	Présence d'arabinose	Protéines impliquées dans le métabolisme et le transport de l'arabinose.
Opéron <i>deo</i>	Présence de deoxyribonucleoside nucleotides	Protéines impliquées dans le catabolisme des deoxyribonucleoside nucleotides.
Opéron <i>nag</i>	Présence du GlcNAc	Protéines impliquées dans le métabolisme du GlcNAc-6P.
Opéron <i>glpA</i>	Présence du glycérol-P	Protéines impliquées dans le métabolisme du glycérol-Phosphate.

TAB. 3.3 – Point commun sur la physiologie et la fonction des systèmes régulés par une boucle d'ADN.

également cette propriété dynamique. Nous savons grâce à la physique des systèmes dynamiques qu'une interaction par boucle rétro-positive est une condition nécessaire pour qu'il y ait bistabilité (cf. Chapitre 1). Le graphe d'interactions peut donc être un indice supplémentaire pour suspecter un système d'abriter une boucle d'ADN. Par ailleurs, le mécanisme de boucle d'ADN responsable de la répression de l'opéron *lac*, même s'il est mentionné dans le travail de Ozbudak et al., n'apparaît pas de manière explicite dans les équations différentielles non-linéaires du modèle. La non-linéarité du modèle qu'ils utilisent provient surtout de la relation entre la concentration de répresseurs actifs et la concentration de l'inducteur qui est décrite par une fonction de Hill avec un coefficient égal à 2. Il serait intéressant de déterminer sur le plan expérimental si la bistabilité existe encore au sein de l'opéron *lac* lorsqu'il n'y a plus formation boucle d'ADN (par exemple avec un seul opérateur de plus forte affinité). La coopérativité introduite par le mécanisme de boucle d'ADN pourrait aussi contribuer à la non-linéarité du système et donc à sa capacité à être bistable.

## 4.2 Interrogation de la base de donnée *RegulonDB*

Compte tenu des points communs mentionnés précédemment, nous pouvons demander aux bases de données (*RegulonDB*) grâce aux outils de la bioinformatique (scripts en langage Perl) si d'autres régions de régulation présentent ces mêmes points particuliers. Sur les 578 régions de régulation d'*Escherichia coli* issues de la base de données *RegulonDB* (fichier "TF binding sites" mise à jour Janvier 2009), population homogène où toutes les bactéries auraient un même niveau d'expression génétique intermédiaire.

Système		avec CRP	avec H-NS	avec IHF	avec Fis
Régions d'E.Coli	578	174	14	67	87
Rep- <sup>&gt;90pb</sup> -Rep	89	55	7	21	28
Act- <sup>&gt;90pb</sup> -Act	59	32	3	14	22
Rep- <sup>&gt;90pb</sup> -Rep ≠ TF global	49	35	2	8	8
Act- <sup>&gt;90pb</sup> -Act ≠ TF global	19	10	2	4	6

TAB. 3.4 – Tableau donnant le nombre de régions de régulation ayant une configuration précise. La première ligne correspond à toutes les régions d'*Escherichia coli* répertoriées et les lignes suivantes correspondent à une configuration précise : sites espacés de plus de 90 pb, pour un répresseur ou activateur puis pour des facteurs spécifiques seulement (≠ TF global). Les colonnes 3,4,5 et 6 donnent le nombre de régions de régulations comportant au moins un site de liaison pour CRP, H-NS, IHF et Fis susceptibles d'intervenir dans des systèmes de boucle d'ADN.

nous passons trois filtres pour trier les données. Le premier filtre concerne la disposition des sites de liaison au sein des régions. Il sélectionne les régions de régulation qui ont plusieurs sites de liaison pour le même facteur de transcription, espacés d'au moins 90 pb (les interdistances sont calculées en prenant la position du milieu du site de liaison). Ce premier filtre donne 128 candidats. Nous avons choisi 90 pb compte tenu de la longueur de persistance de l'ADN et que la grande majorité des boucles déjà connues ont une longueur supérieure à 90 pb.

Le deuxième filtre sélectionne dans les régions de régulation précédentes celles dont le facteur de transcription en question exerce une fonction de répression (ou éventuellement la double fonction répression et activation). Nous obtenons 89 candidats, il est à noter que nous obtenons moins de régions de régulation avec des sites se liant à un activateur (59) (certaines régions ont les deux types de configurations).

Un grand nombre des régions de régulation font intervenir des sites de liaison pour les 7 facteurs de transcription globaux que sont CRP, IHF, FNR, Fis, ArcA, Lrp et H-NS [105]. Nous recherchons des systèmes faisant intervenir des facteurs de transcription spécifiques comme dans les cas de boucle déjà connus. Nous filtrons donc parmi les régions de régulation précédentes celles dont les sites de liaison ne soient pas des sites pour les 7 facteurs globaux.

Nous trouvons alors beaucoup plus de systèmes qui font intervenir des sites espacés de plus de 90 pb pour des répresseurs spécifiques (49 occurrences) que des systèmes présentant la même configuration pour des activateurs spécifiques (19 occurrences). Le nombre de sites de liaison pour des répresseurs est un peu plus élevé que celui pour des activateurs dans le génome d'*Escherichia coli* (Répresseur : 1263 sites, Activateur : 1099 sites et Double fonction : 52 sites) mais il est seulement de l'ordre de 10 % et ne peut expliquer l'écart important dans le nombre de régions trouvées. La configuration sites de liaison Rep-<sup>>90pb</sup>-Rep avec un répresseur spécifique apparaît donc de manière importante, signifiant peut être qu'il s'agirait d'une configuration ayant une fonction bien précise.



Partant de cette liste de gènes, nous regardons également les autres éléments qui pourraient jouer un rôle dans la formation de boucle d'ADN. Nous calculons pour chaque séquence susceptible de boucler le contenu en liaisons GC qui peut donner une indication sur la souplesse de la séquence. Cependant comme vu plus haut, nous savons que certaines des boucles existantes ont des contenus en GC plutôt élevés (plus élevés que la moyenne des régions de régulation d'*Escherichia coli*). Nous n'en faisons pas un filtre mais calculons tout de même cette quantité pour chaque segment d'ADN susceptible de boucler car cela reste une information intéressante notamment pour de petites boucles proches de 100 pb. Celui-ci est calculé sur la séquence des bases situées entre les milieux des sites de liaison pouvant former la boucle.

Nous regardons ensuite parmi les régions de régulation qui sont sorties des filtres s'il y a un ou plusieurs sites pour le facteur de transcription CRP qui comme vu plus haut, intervient dans tous les systèmes de boucle déjà connus (majoritairement comme compétiteur de la boucle d'ADN).

Parmi les 49 régions de régulation comportant des sites associés à un répresseur spécifique et espacés d'au moins 90 pb, nous en avons identifié 35 où le facteur de transcription CRP intervient, ce qui correspond à un pourcentage de 71% alors qu'il n'est que de 30% pour l'ensemble des régions de régulation d'*Escherichia coli*. Pour le même type de configuration mais pour des activateurs, nous avons identifié 10 régions contenant au moins un site pour CRP sur les 19 identifiées précédemment, soit un pourcentage de 53 %. La configuration Rep-<sup>>90pb</sup>-Rep avec un répresseur spécifique et la présence du facteur CRP semble donc ressortir de manière importante par rapport aux autres configurations. Ces chiffres soulignent que la présence du facteur de transcription CRP parmi des systèmes qui ont des sites espacés d'au moins 90 pb et se liant à un répresseur est significative.

Nous avons ensuite compté le nombre de régions de régulation où interviennent les histones-like protéines H-NS, IHF et Fis. Toutes les trois sont connues pour jouer sur la topologie de l'ADN et sont donc susceptibles d'avoir un rôle dans la formation de boucle d'ADN. De manière un peu surprenante, il n'est apparu que 2 uniques cas ayant la configuration Rep-<sup>>90pb</sup>-Rep  $\neq$  TF global, avec un site pour la protéine H-NS. Ce sont les régions de régulation de l'opéron *lac* et de l'opéron *gal*. Il semble donc que cette protéine ne joue un rôle dans la formation de boucle d'ADN au sein de région de régulation que dans ces deux cas déjà connus.

Nous avons également regardé le nombre d'occurrences pour les protéines IHF et Fis. Leur répartition suit à peu près celles des régions d'*Escherichia coli*, il ne semble pas ressortir de configurations particulières. Elles apparaissent cependant un peu plus dans les régions sélectionnées (interdistance > 90 pb), que ce soit pour les répresseurs ou les activateurs, mais ceci peut provenir du fait que cette sélection garde des régions riches en sites de liaison, là où interviennent fréquemment les facteurs globaux comme IHF et Fis. Enfin, la protéine HU n'apparaît que pour l'opéron *gal* ceci s'explique par le fait que peu de sites de liaison de cette protéine sont encore connus, (seulement 4 occurrences dans la base de données) nous ne l'indiquons pas dans le tableau 3.4.

### 4.3 Corroboration bibliographique

Une fois ces premiers filtres passés sur les bases de données, nous commençons une recherche dans les bases bibliographiques. L'objectif est d'une part, de déterminer si la région sélectionnée régule un système génétique qui a un comportement et des propriétés physiologiques proches de celles évoquées précédemment (réaction à un changement environnemental, système type interrupteur). Nous regardons à quelle famille appartient la protéine et si celle-ci est connue pour pouvoir former des oligomères. D'autre part, nous regardons s'il a déjà été entrepris une mise en évidence expérimentale de la présence de boucle au sein de ces régions de régulation ou si elle a déjà été suspectée.

## 5 Liste de candidats obtenus

### 5.1 Résultat des filtrages

Nous présentons dans cette section les opérons qui sont ressortis des trois filtrages de données. La liste des gènes candidats est présentée sous la forme d'un tableau : la première colonne indique le nom de l'opéron, la deuxième le nom du facteur de transcription qui serait à l'origine de la boucle, les colonnes suivantes indiquent la présence ou non des protéines CRP, H-NS, IHF, Fis. La septième colonne donne le pourcentage du contenu en GC de la séquence d'ADN qui pourrait boucler, la huitième colonne donne les positions des sites de liaison impliqués (par rapport à un site promoteur). Enfin, la dernière colonne donne le résultat d'une recherche de références bibliographiques mentionnant une boucle d'ADN pour l'opéron candidat. Nous avons retiré de cette liste les cas déjà connus de boucles d'ADN ainsi que d'autres cas redondants (certaines régions de régulation apparaissaient plusieurs fois car elles régulent des ARN non codant inclus dans des opérons). Donc le nombre de régions de régulation obtenues ici n'est pas 49 qui est le nombre de candidats brut, sélectionnés par les filtres bioinformatiques.

### 5.2 Cas d'intérêt

Nous décrivons dans cette section plus en détails plusieurs systèmes qui nous paraissent de très bons candidats pour comporter une boucle d'ADN parmi les opérons qui sont sortis des filtrages. Nous donnons des informations supplémentaires comme la famille de la protéine qui pourrait être à l'origine de la boucle ou la fonction des gènes régulés.

#### Opéron *agaS*

Plusieurs indices suggèrent à penser que l'opéron d'*agaS* est régulé par une boucle d'ADN : il réagit à un signal environnemental et les gènes de l'opéron ont des fonctions métaboliques (transport et catabolisme du N-acetylgalactosamine et d-galactosamine). La protéine AgaR fait partie la famille DeoR. Compte tenu du grand nombre de sites de liaison présents dans cette région (six), l'existence de multi-boucle (comme dans l'opéron *deo*, Figure 3.5) ne nous paraît pas impossible. L'existence de boucle d'ADN a été suggérée dans le travail de W. Ray [135]. A notre connaissance,

Nom opéron	TF en jeu	CRP	H-NS	IHF	Fis	GC content	Configuration des sites	Corroboration bibliographique.
agaS	AgaR	non	non	non	non	43,33%	6 sites de -276.5 pb à +24.5 pb	Boucle d'ADN suggérée dans [135].
gdhA	Nac	oui	non	non	non	33,06%	4 sites de -81 à +45 pb	Boucle fortement suggérée, méthodes ①, ② et ④ appliquées dans [54] et [142].
gntT	GntR	oui	non	non	non	42,14%	2 sites à -58.5 et +82.5 pb	Boucle d'ADN suggérée dans [124] et [130].
eda	GntR	non	non	non	non	39,82%	2 sites à -93.5 et +128.5 pb	Boucle d'ADN suggérée dans [130].
ftsQ	SdiA	non	non	non	non	43,92%	2 sites à - 163.5 et -15.5 pb	Pas de référence mentionnant une boucle d'ADN.
focA	NarL	oui	non	oui	non	31,09%	11 sites de -348 pb à +302 pb	Pas de référence mentionnant une boucle d'ADN.
dcuB	NarL	oui	non	non	non	39,33 %	6 sites de -109 à +43 pb	Pas de référence mentionnant une boucle d'ADN.
hybO	NarL	non	non	non	non	33,68 %	3 sites de -76 à -51 et +21 pb	Pas de référence mentionnant une boucle d'ADN.
purA	PurR	non	non	non	non	47,47%	2 sites à -90.5 et +9.5 pb	Boucle d'ADN infirmée? pas de coopérativité entre les deux sites montrée par la méthode ② dans [61].
phr	LexA	non	non	non	non	51,65%	2 sites à -66.6 et +25.5 pb	Pas de référence mentionnant une boucle d'ADN.
hipB	HipB	non	non	oui	non	43,16%	4 sites de -85.5 à +10 pb	Pas de référence mentionnant une boucle d'ADN.
lldP	LldR	non	non	non	non	38,40%	2 sites à -97 et +30 pb	Boucle d'ADN fortement suggérée, méthode ① appliquée dans [2].
uxuA	UxuR	oui	non	non	non	39,62%	2 sites à -159.5 et +53.5 pb.	Pas de référence mentionnant une boucle d'ADN.
uspA	FadR	non	non	oui	non	61,54%	2 sites à +5 et +97 pb	Pas de référence mentionnant une boucle d'ADN.
entC	Fur	oui	non	non	non	37,82%	7 sites de -100 à +58 pb	Pas de référence mentionnant une boucle d'ADN.
csgD	CpxR	oui	non	non	non	23,76%	6 sites de -100.5 à +2.5 pb	Pas de référence mentionnant une boucle d'ADN.
carA	aminopeptidase	non	non	oui	oui	37,15%	2 sites à -173 et -79 pb	Pas de référence mentionnant une boucle d'ADN.
ompF	OmpR	oui	non	oui	non	31,78%	3 sites à -370.5, -70.5 et -50.5 pb	Boucle d'ADN suggérée dans [18,65,134].
fadL	OmpR	oui	non	non	non	41,54%	2 paires à -147.4,-60 et +67.5, +113.5 pb	Pas de référence mentionnant une boucle d'ADN.
flhDC	OmpR	oui	non	oui	non	39,77%	2 sites à -145.5 et +18.5 pb	Boucle d'ADN suggérée dans [132,157].
gcvTHP	GcvA	non	non	non	non	31,20%	3 sites à -257,-228 et-52 pb	Boucle d'ADN suggérée, méthode ① appliquée dans [162].
metE	MetJ	non	non	non	non	38,26%	2 sites à +1.5 et 116.5 pb	Pas de référence mentionnant une boucle d'ADN.
metR	MetJ	non	non	non	non	39,77%	2 sites à -162.5 et -47.5 pb	Pas de référence mentionnant une boucle d'ADN.
aceB	IclR	oui	non	oui	non	32,46%	8 sites de -141 à -26 pb	Boucle d'ADN suggérée dans [179].
argG	ArgR	oui	non	non	non	35,00%	3 sites de -125.5 à +15.5 pb	Boucle d'ADN suggérée dans [26] et [81].
galS	GalR	oui	non	non	non	48,37%	2 sites à -61.5 et +91.5 pb	Pas de référence mentionnant une boucle d'ADN.
galP	GalS	oui	non	non	non	44,16%	4 sites de -207.5 à +136.5 pb	Pas de référence mentionnant une boucle d'ADN.
chbB	NagC	oui	non	non	non	39,82%	2 sites à -112 et +3 pb	Boucle d'ADN suggérée dans [129].
nanC	NagC	oui	non	non	non	38,43%	2 sites à -277 et -65 pb	Boucle d'ADN suggérée dans [159].
fimB	NagC	oui	non	oui	non	38,43%	2 sites à -674 et -462 pb	Boucle d'ADN suggérée dans [159].
glmU	NagC	non	non	non	non	40,79%	2 sites à -96 et +58 pb	Boucle d'ADN suggérée dans [126].
melA	MelR	oui	non	non	non	53,14%	4 sites de -237.5 à +62.5 pb	Boucle d'ADN suggérée [72] et méthodes ① et ② appliquées dans [146].
glpD	GlpR	oui	non	non	non	52,97%	2 paires de sites à -5, +19.5 et +414.5,+455.5 pb	Boucle suggérée, méthode ① appliquée dans [181].
glpF	GlpR	oui	non	non	non	49,29%	7 sites de -79.5 à +1043.5 pb	Boucle d'ADN suggérée dans [176].
ptsG	DgsA	oui	non	non	oui	36,00%	2 sites espacés de 169 pb	Boucle d'ADN suggérée dans [133].

aucune expérience faisant intervenir les méthodes répertoriées dans la section 2 n'a encore été appliquée sur cette région de régulation.

### Opérons *focA*, *dcuB* et *hybO*

Les opérons du régulon de NarL répondent à la présence de nitrate ou de nitrite. NarL réprime de nombreux opérons dont ceux de *focA* et *dcuB* impliqués dans des voies respiratoires alternatives comme la réduction du fumarate ou la fermentation de simples sucres. NarL fait partie de la famille LuxR/UhpA or LuxR est suspectée de faire une boucle d'ADN chez l'opéron *luxD* [156]. A notre connaissance, aucune étude n'a été entreprise pour mettre en évidence une boucle d'ADN entre deux des sites NarL de ces régions de régulation.

### Opéron *gdhA*

L'opéron *gdhA* présente de nombreux indices forts pour être régulé par une boucle d'ADN. Sa régulation a surtout été étudiée chez la bactérie *Klebsiella pneumoniae* (dont la région de régulation a une structure très proche de celle d'*Escherichia coli*). Le système est fortement réprimé (d'un facteur 20), rappelant le point commun sur la physiologie des systèmes réprimés par boucle d'ADN. L'ajout de 5 ou 15 pb diminue la répression [54]. Le fait d'empêcher la tétramérisation diminue également la répression [142]. Cependant, des expériences de band shift in vitro ne montrant pas de coopérativité, les auteurs n'ont pas encore pu conclure sur la formation de boucle d'ADN pour ce système [54] (le fait de ne pas observer de coopérativité in vitro signifie peut-être que la présence d'un autre facteur présent in vivo est nécessaire pour la coopérativité comme la protéine HU pour la boucle d'ADN de l'opéron *gal*).

### Opérons *gntT* et *eda*

Les deux opérons *gntT* et *eda* sont également d'excellents candidats pour être réprimés grâce à une boucle d'ADN faisant intervenir le facteur de transcription GntR comme cela était suggéré dans plusieurs travaux de l'équipe de T. Conway [124, 130]. GntR appartient à la famille des facteurs de transcription LacI/GalR. Il réprime les opérons impliqués dans le catabolisme du d-gluconate. Cependant, à notre connaissance, il n'a pas été entrepris de vérifier l'hypothèse de boucle d'ADN par l'une des méthodes présentées en section 2.

### Opéron *ftsQ*

Le facteur de transcription Sdia est très homologue aux facteurs de transcription du quorum-sensing qui appartiennent à la famille LuxR. Le système réagit à la présence de molécules liées au quorum sensing [183]. Nous n'avons trouvé aucune référence suggérant la présence de boucle d'ADN pour cette région de régulation.

### Opérons *uxuAB* et *uidA*

Le système de l'opéron *uxuAB* semble avoir beaucoup d'indices pour être réprimé par une boucle d'ADN. Le système peut être induit par le glucuronate ou le

fructuronate et les gènes de l'opéron sont impliqués dans son métabolisme. De plus, on remarque la présence de plusieurs sites pour la protéine CRP, l'opéron est en effet sensible à la répression catabolique. UxuR est une protéine qui appartient à la famille GntR (qui a également des cas suspectés de boucle d'ADN). Il serait donc très intéressant de tenter la mise en évidence d'une boucle d'ADN sur cet opéron.

Par ailleurs, il a été remarqué que les facteurs de transcription UidR and UxuR agissent par coopérativité au sein de la région de régulation de l'opéron uidA qui fait également partie du régulon de uxuR. Cependant un mécanisme de boucle d'ADN ne semble avoir été envisagé lors de ce travail [19]. La région de régulation de uidA contient un site pour UxuR et un site pour UidR, espacés de 213 pb (cette région de régulation n'est pas ressortie de nos filtrages car les deux sites de liaison font intervenir des facteurs de transcription différents). Il nous semble possible que la coopérativité qui avait été observée dans le travail de Blanco et al. soit due à la formation d'une boucle d'ADN. De plus, un site de liaison pour le facteur de transcription CRP est présent entre les deux sites. Ce cas présenterait alors une caractéristique importante car cela pourrait être la mise en évidence d'une boucle d'ADN formée par l'association de deux facteurs de transcription *différents* ce qui, à notre connaissance, n'a pas encore été beaucoup observé (cependant, d'après le travail de Kornacker et al. [79], il y aurait le cas pour le système de boucle de l'opéron *ara* qui pourrait faire intervenir une boucle grâce aux facteurs de transcription AraC et LexA).

### Opéron *uspA*

Cet opéron est régulé par le répresseur FadR qui fait partie de la famille de GntR qui possède également des cas suspectés de boucle d'ADN. Les deux sites de liaison au sein de la région de régulation sont espacés de 92 pb. Il est intéressant de noter que le deuxième site se trouve à 97 pb en aval du site promoteur. Ceci rappelle la boucle de l'opéron *lac* où l'un des sites auxiliaires est également placé loin en aval du site promoteur. La fonction physiologique de *uspA* n'est pas encore très bien cernée mais il est cependant connu que ce gène répond à une variété de signaux environnementaux ou de stress.

### Opérons *ompF*, *flhD* et *fadL*

Les gènes *ompF*, *flhD* et *fadL* sont d'excellents candidats pour être régulés grâce à une boucle d'ADN. Ce mécanisme a été suggéré pour les gènes *ompF* et *flhD* dans de nombreux travaux [18, 65, 132, 134, 157]. De plus, il a été montré que la forme phosphorylée de OmpR pouvait former des multimères stables [116]. OmpF est une porine de la membrane externe; FlhD est un facteur de transcription qui dirige la synthèse des flagelles bactériens et FadL est une protéine qui joue un rôle dans le métabolisme du glycolate (qui peut être utilisé comme source de carbone). Étant donné leur importance dans de nombreux processus de la vie d'*Escherichia coli*, les deux gènes *ompF* et *flhD* ont particulièrement été étudiés. Cependant, à notre connaissance, aucune des méthodes répertoriées dans la section 2 ne semble avoir été mise en œuvre pour confirmer la formation de boucle d'ADN pour ces systèmes.

### Opérons *metE* et *metR*

*MetJ* réprime l'expression des gènes impliqués dans la synthèse et le transport de la méthionine. Des dimères de dimères peuvent se former et les sites de liaison pour *MetJ* peuvent coopérer d'après le travail de Phillips et al. [125]. Comme les deux gènes *metE* et *metR* sont placés de façon divergente, la boucle d'ADN permettrait de réprimer les deux opérons simultanément. Les deux sites de liaison pour *MetJ* sont séparés de 115 paires. Ce système nous paraît donc être un très bon candidat pour une mise en évidence de boucle d'ADN.

### Opéron *melAB*

Il a été suspecté la formation d'une boucle d'ADN à l'origine de la forte répression de l'opéron *melAB* compte tenu de nombreux indices en ce sens. Par exemple, la protéine *MelR* fait partie de la famille *AraC* connue pour être à l'origine d'une régulation par boucle d'ADN. En 2008, Samarasinghe et al. ont donc cherché en mettre en évidence le mécanisme de boucle d'ADN. Ils ont appliqué la méthode de déphasage des sites de liaison qui s'est avérée en faveur d'un mécanisme de formation de boucle d'ADN. Cependant, ils ont également entrepris de visualiser la formation de boucle d'ADN par microscopie à force atomique et n'ont pas pu observer de boucle d'ADN [146]. Ils concluent que le mécanisme de régulation est probablement plus complexe. Il fait intervenir les 4 sites de liaison pour la protéine *MelR* qui serait à l'origine d'un complexe protéique-ADN appelé "Répressome".

## 6 Conclusion

Dans ce chapitre, nous avons réuni plusieurs informations dans le but de motiver une recherche systématique de boucle d'ADN dans les régions de régulation du génome d'*Escherichia coli*. Il apparaît clairement que ce mécanisme ne revêt pas seulement un aspect géométrique mais qu'il a un impact et des propriétés de première importance pour la régulation transcriptionnelle. Nous avons comptabilisé environ une trentaine de régions de régulation connues qui font intervenir une boucle d'ADN qui entre en jeu dans la régulation transcriptionnelle pour une activation ou pour une répression. Cela représente une proportion d'environ 5% de régions bouclées pour l'ensemble des régions du génome d'*Escherichia coli*. Ce pourcentage est sûrement très sous-estimé et d'autres boucles d'ADN sont à mettre en évidence dans le génome d'*Escherichia coli*. En réunissant plusieurs types d'informations et d'observations sur le mécanisme de boucle d'ADN, ce chapitre met en évidence un groupe d'une trentaine de gènes qui nous paraissent de très bons candidats pour allonger la liste existante des boucles d'ADN connues et bien caractérisées.

Plusieurs de ces systèmes ont été suspectés d'agir par boucle d'ADN compte tenu d'arguments en leur faveur. Chez certains, une forte coopérativité a déjà été démontrée. Si un mécanisme de boucle n'est pas à l'origine de la coopérativité observée chez ces systèmes alors peut être y-a-t-il à découvrir d'autres mécanismes de coopérativité entre sites de liaison placés à distance.

# Chapitre 4

## Recherche d'interactions 3D sur le système Pel de la bactérie *Erwinia chrysanthemi*

Nous avons vu dans le chapitre précédent que les boucles d'ADN pouvaient avoir un impact direct sur la régulation et conférer d'intéressants avantages. Nous présentons dans ce chapitre un travail expérimental où nous avons entrepris de rechercher des interactions par boucle d'ADN non pas entre sites de liaison appartenant à un même gène mais entre des sites de liaison de gènes différents. Après avoir présenté différents travaux qui ont stimulé ce travail et justifié le choix de notre système d'étude, nous détaillons les expériences réalisées, présentons les résultats de celles-ci et les conclusions.

### 1 Modèles d'organisation spatiale des génomes

De nombreux génomes aussi bien eucaryotes que procaryotes sont aujourd'hui séquencés et au moins en partie annotés. Un axe de recherche qui s'est développé ces dernières années est celui de comprendre si les génomes sont organisés dans leur conformation tri-dimensionnelle au sein de la cellule. Un premier argument en faveur de ce champ d'investigation est que d'une manière générale, l'on peut dire que la nature adopte des structures tridimensionnelles qui contiennent de l'information. La structure tertiaire des protéines est directement liée à leur fonction, les molécules d'ARN messagers au départ représentées comme des molécules linéaires acquièrent aussi des structures tridimensionnelles très particulières (boucles, têtes d'épingle...) qui ont une importance dans la régulation. Il est donc naturel d'imaginer que la molécule d'ADN acquière à l'échelle d'un génome une structure particulière, non aléatoire, peut-être au moins sur certaines parties, et que cette structure découle de contraintes précises. Nous présentons brièvement dans la section suivante des travaux qui proposent des modèles d'organisation spatiale de génome ou qui suggèrent qu'il pourrait y avoir des interactions ou des corrélations entre parties pourtant éloignées sur une représentation 1D d'un génome.

## 1.1 Organisation solénoïdale du génome

Le biologiste François Képès a proposé une organisation dite solénoïdale pour les génomes de la bactérie *Escherichia coli* [84] et de la levure *Saccharomyces cerevisiae* [83]. En se basant sur les données de génomique et de transcription, il remarque une périodicité sur les positions des différents gènes co-régulés par un même facteur de transcription. Par exemple, chez la bactérie E.Coli K12 MG1655, les gènes co-régulés par le facteur de transcription global CRP tendraient à se positionner tous les 92.8 kpb. Cette organisation conférerait d'intéressants avantages sur le plan biologique : concentration locale des facteurs de transcription au niveau des régions d'intérêt, réponse plus rapide et coordonnée d'un ensemble de gènes (ce qui est souvent le cas quand par exemple, une bactérie est soumise à un stress ou évolue dans un nouveau milieu aux sources de carbone différentes). Bien que la justification de la périodicité soit toujours en cours d'étude et qu'elle ait été remise en cause notamment par le travail de Lercher et al. [89], l'hypothèse "intergénique" développée dans ce travail nous a paru très intéressante et surtout pouvant être testée à la paillasse.

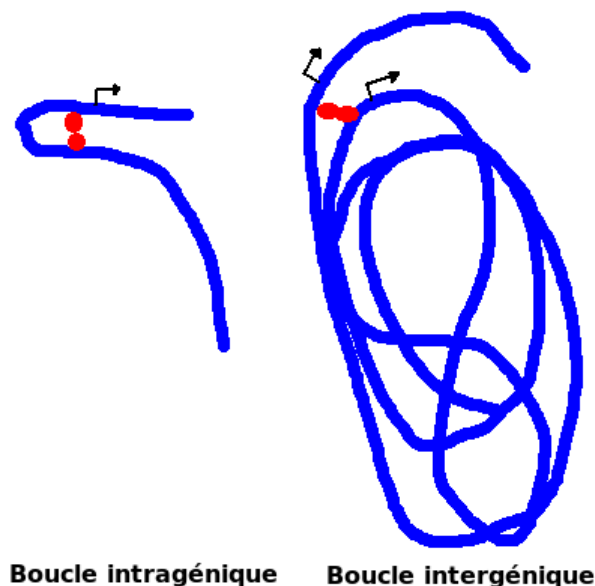


FIG. 4.1 – Boucle intragénique : un facteur de transcription se lie sur deux sites de liaison appartenant à la région de régulation d'un même gène ; boucle intergénique : un facteur de transcription pourrait se fixer sur deux sites de liaison appartenant à deux gènes distincts.

Cette hypothèse propose qu'un facteur de transcription pourrait se lier à deux sites de liaison appartenant non pas au même gène (comme dans les cas déjà connus de l'opéron *lac* et des quelques autres, voir chapitre 3) mais à deux gènes distincts (Figure 4.1). Il est intéressant de noter qu'elle avait également été formulée mais d'une autre façon dans le travail de Buchler et al. [21]. Ces auteurs se sont intéressés à la façon dont les régions de régulation pourraient implémenter des fonctions logiques (par exemple une porte OU faisant intervenir deux facteurs de transcription différents), ils voyaient comme limite à ces concepts le fait que, le génome bactéri-



rien étant très confiné, certains facteurs de transcription pourraient se lier à deux régions de régulations appartenant à deux gènes différents. Ils voyaient donc ceci comme conférant plutôt un désavantage et parlaient d’"interférences inter-géniques".

## 1.2 Travaux bioinformatiques suggérant des interactions à distance ou des corrélations à longue portée dans des génomes

Plusieurs travaux bioinformatiques dans le même esprit que [83, 84] ont également cherché des régularités ou des corrélations dans plusieurs types de données génomiques en utilisant notamment des outils de traitement du signal. Nous présentons ici quelques exemples significatifs.

Le travail de Khodursky et al. [71] utilise des outils d’analyse spectrale comme l’algorithme de Lomb<sup>1</sup> et les transformées en ondelette. Ces auteurs appliquent ces outils sur le signal 1D correspondant à l’activité transcriptionnelle sur la position dans le génome. Plusieurs périodicités apparaissent et sont classées en trois catégories : petite portée 16 kpb, moyenne portée 100-125 kpb et longue portée 600-800 kpb. Ils changent les conditions de culture ou inhibent la transcription (au moyen de la rifampicine) pour s’assurer que les fréquences sont bien stables et ne découlent pas d’un artefact introduit par exemple par la technique des puces à ADN.

Le travail de Palsson et al. [4] utilise également les outils des transformées en ondelette mais ici sur des signaux liés à la séquence des génomes, tels que la densité en gènes par kilo bases, le contenu en GC/AT et l’indice d’adaptation de codon. Ils trouvent des "patterns" dans de nombreux génomes procaryotes. Ils corrélient ensuite le degré de similitude avec des paramètres globaux tels que la taille du génome ou le pourcentage AT total du génome. Ils trouvent par exemple que le degré de pattern du signal du contenu en GC est positivement corrélé avec la taille du génome. Ces résultats d’ordre général semblent appuyer la thèse de contraintes notamment des processus transcriptionnels sur la structure globale des génomes.

Le travail de Carpentier et al. [24] cherche des corrélations dans des données de puce à ADN de la bactérie *Bacillus subtilis* et chez *Escherichia coli*. Ils regardent les corrélations sur le signal d’expression ARN entre deux gènes en fonction de leur inter-distance et utilisent une fonction de corrélation dite Kendall tau<sup>2</sup>. Ils trouvent chez *Bacillus Subtilis* des oscillations dans ce signal de période  $600 \pm 55$ ,  $240 \pm 21$ ,  $113 \pm 21$  et  $60 \pm 6$  gènes et chez *Escherichia coli* deux périodicités :  $557 \pm 30$  and  $100 \pm 18$  gènes. Ces deux périodicités sont assez proches du travail de Khodursky si on prend comme taille moyenne d’un gène 1kpb. Ils soulignent que ces corrélations apparaissent pour n’importe quel gène et indépendamment de la position. Cela ne dépendrait pas des processus impliquant des facteurs de transcription comme par exemple dans le modèle solénoïdal.

---

<sup>1</sup>Cet algorithme a la particularité de trouver des contenus fréquentiels dans des signaux qui peuvent être échantillonnés à pas irrégulier comme c’est le cas des positions des gènes.

<sup>2</sup>La fonction Kendall tau est une mesure de corrélation, elle quantifie le degré de relation entre deux variables.

### 1.3 Travaux issus de la physique statistique suggérant des organisations par boucle d'ADN

Le travail de Sumedha et al. [164] propose un modèle utilisant le formalisme de la physique statistique pour la formation de boucles d'ADN qui organiseraient le génome. Ils calculent les énergies libres de formation de boucle d'ADN en tenant compte de deux contributions : l'énergie d'affinité des protéines sur l'ADN et la contribution entropique. Plus précisément, ils montrent que dans le mécanisme d'agglomération des protéines se liant à l'ADN, la liaison d'un facteur de transcription sur un système déjà bouclé permet dans certains cas de minimiser l'énergie libre sans pertes entropiques majeures. Ce mécanisme introduit une attraction effective entre des domaines de liaison de différentes boucles. Il est important de retenir que pour qu'il y ait les agglomérats de protéines avec la formation de boucles d'ADN, deux ingrédients sont nécessaires : des protéines bivalentes c'est à dire qui puissent se lier à deux endroits distincts sur l'ADN (ce qui semble être souvent le cas des facteurs de transcription qui peuvent agir en oligomères) et deuxièmement des domaines de liaison contenant plus de deux sites de liaison chacun. Ce dernier point guidera notre choix du système d'étude.

Si dans le travail précédent, l'entropie était un obstacle à l'organisation globale des génomes, dans l'approche théorique développée par l'équipe de Peter Cook [103, 104], l'entropie est le moteur de l'organisation spatiale génomique via les *forces de déplétion*. En effet, en tenant compte de la taille des molécules d'ARN polymérasés et de l'environnement chargé en molécules d'une bactérie ou d'une cellule, l'entropie du système entier peut être minimisée en rapprochant les grosses molécules donnant au système total un nombre plus grand de micro-états accessibles. Les deux travaux ici cités ne sont donc pas en contradiction, une des perspectives du travail de Sumedha et al. étant justement de tenir compte de l'environnement chargé en molécules du milieu intra-cellulaire et de retrouver les propositions de Cook.

### 1.4 Travaux de biologie expérimentale

Le travail expérimental de l'équipe de Peter Cook mené depuis plusieurs années montre à l'aide notamment de sondes fluorescentes que la transcription aurait lieu en des endroits précis (appelés foci) au sein de la cellule ou du noyau. Par exemple, les expériences montrées dans [30] révèlent des clichés collectés par microscope confocal où l'on peut voir que chez des cellules humaines de type Hela, les transcrits nouvellement synthétisés sont concentrés en des endroits ponctuels au sein du noyau. Il propose un modèle d'organisation des génomes où les ARN polymérasés se regrouperaient en des endroits précis et transcriraient plusieurs gènes qui seraient réunis en bouclant l'ADN [103].

Un travail autre expérimental qui s'est attaqué à élucider l'organisation spatiale des génomes bactériens est celui de l'équipe de Frédéric Boccard, qui fait appel à des outils de génétique [170]. En utilisant un système de séquences spécifiques recombinantes, l'équipe a pu révéler quelles parties d'ADN du génome d'*Escherichia coli* pouvaient entrer en contact. En analysant les taux de recombinaison de sites éparpillés sur le génome, ils ont mis en évidence une organisation du génome en quatre macrodomaines et deux moins structurés. Deux de ces macrodomaines (appelés Ori et Ter) font presque 1 Mpb suggérant que des gènes espacés par de telles distances

pourraient tout de même avoir une probabilité non négligeable de se rencontrer.

## 2 Choix du système d'étude

Nous avons choisi de travailler sur un système génétique codant pour plusieurs enzymes, les pectate lyases, qui jouent un rôle dans la dégradation de la pectine, composant principal de la paroi cellulaire des plantes, chez la bactérie *Erwinia chrysanthemi* (nouvellement appelée *Dickeya dadantii* [148]). *Erwinia chrysanthemi* est une entérobactérie phytopathogène à gram négatif au génome d'environ 4700 gènes, proche de celui d'*Escherichia coli*. Plus précisément, nous avons choisi de travailler sur 3 gènes : les gènes *pelA*, *pelE* et *pelD* qui codent pour 3 des 5 enzymes majeures du système de virulence d'*Erwinia chrysanthemi* [67] et qui sont situés au même endroit sur le génome d'*Erwinia chrysanthemi*, à la position 3 573 000 pb (voir Figure 4.5). Ces enzymes jouent un rôle de première importance dans la virulence de cette bactérie, en clivant les liaisons osidiques  $\alpha$  1-4 du polygalacturonate (PGA) (un des premiers produits de la dégradation de la pectine). Le PGA pouvant être utilisé comme source de carbone, ces gènes sont au coeur du métabolisme et du cycle cellulaire de cette bactérie phytopathogène [66]. Les avantages mentionnés dans le chapitre 3 des boucles d'ADN comme une meilleure robustesse aux fluctuations temporelles dans l'expression génétique seraient tout à fait appropriés pour ce système génétique de première importance.

Nous avons choisi de travailler sur ce système pour plusieurs raisons dont nous donnons ici les quatre principales.

### 2.1 Raison génétique

Parmi les différents facteurs de transcription qui régulent les trois gènes du système Pel, certains sont suspectés de pouvoir former des boucles intragéniques. Par exemple, d'après des expériences réalisées en 1998 au laboratoire, le facteur de transcription PecT peut très probablement se dimériser ou se tétramériser. Il est suspecté de faire des boucles entre deux sites espacés sur l'ADN notamment dans sa propre région de régulation (expériences de "band shift" [25]). De plus, il exerce une répression, toutes ces caractéristiques lui font partager beaucoup de points communs avec le système de boucle de l'opéron *lac* (cf. Chapitre 3). PecT régule de nombreux autres gènes qui influent sur l'activité enzymatique Pel (voir Figure 4.2) avec lesquels il pourrait boucler. C'est pourquoi le fait d'éliminer des régions de régulation contenant des sites de liaison pour ce facteur donne de plus grandes chances de trouver un phénotype altéré.

Le facteur KdgR, également répresseur, fait partie de la famille des facteurs de transcription IclR [138], cette famille contient chez *Escherichia coli* un cas suspecté de formation de boucle d'ADN (opéron *aceB* [179], cf. Chapitre 3). Le regulon de KdgR se compose d'une trentaine de gènes dont la plupart sont impliqués dans le catabolisme de la pectine [140]. Par exemple, nous pouvons citer les gènes *out* codant pour les systèmes de sécrétion (de type 2) des enzymes Pel [29] (sorte de canaux qui permettent la sécrétion des enzymes dans le milieu extérieur), le gène *kdgM* qui code pour une porine qui peut charger les oligogalacturonides [20] ou encore le gène *kdgK* qui code pour une kinase qui intervient dans le métabolisme de la pectine [68],

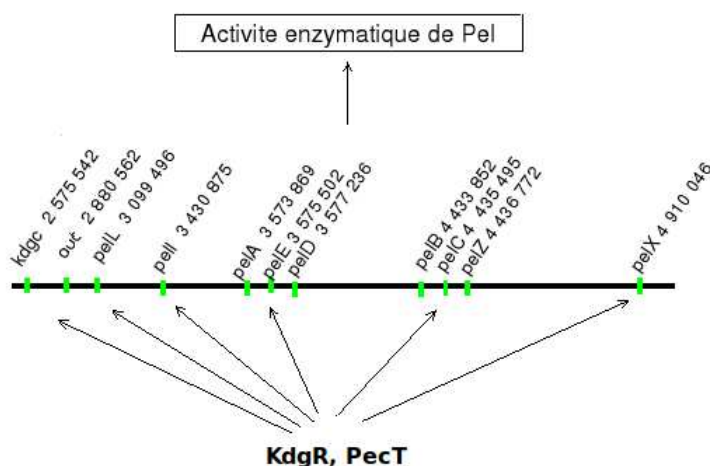


FIG. 4.2 – Représentation 1D de quelques gènes faisant partie des régulons des facteurs de transcription PecT et KdgR qui pourraient boucler avec les régions de régulations des gènes *pelA*, *pelE* et *pelD* et dont la dérégulation aurait un impact sur l'activité enzymatique Pel mesurée.

etc. Si les régions de régulation des gènes *pel* choisies peuvent influencer ces autres gènes par boucle d'ADN alors un effet sur le phénotype c'est à dire sur la croissance bactérienne et surtout sur la cinétique de l'activité enzymatique Pel devrait être visible.

Le système Pel d'Erwinia est également régulé par le facteur de transcription global CRP que l'on retrouve chez de nombreuses bactéries. Si à notre connaissance, aucune boucle formée par cette protéine n'a été mise en évidence, du fait de son très grand nombre de cibles il a souvent été suspecté de jouer un rôle dans l'organisation spatiale à l'échelle des génomes [56]. Le système Pel est également régulé par Fis ou H-NS, des facteurs de transcription qui jouent un rôle dans le surenroulement de l'ADN et que l'on retrouve dans de nombreuses entérobactéries. Ces trois facteurs de transcription CRP, Fis et H-NS sont des facteurs globaux et font partie des protéines régulatrices qui ont le plus grand nombre de cibles. Agir sur des régions de régulation qui contiennent des sites de liaison pour ces facteurs donne de plus grandes chances de trouver une interaction entre régions éloignées.

## 2.2 Raison thermodynamique

Des travaux issus de la physique statistique, comme celui mentionné précédemment de Sumedha et al. [164] ou [60] mais aussi ceux issus de la physique des polymères montrent que le fait de boucler un polymère génère un coût entropique. En effet, le système bouclé possède moins de micro-états accessibles, coût d'autant plus élevé que la boucle est grande si on considère un polymère libre [60]. Une façon de payer ce coût est d'augmenter l'énergie de liaison qui intervient. Or l'énergie de liaison entre un facteur de transcription et son site cible est plutôt faible (de l'ordre de 10-15 kcal/mol par exemple pour l'opéron *lac*). Une solution qui est en accord avec ce qui est observé dans les génomes de procaryotes, pour augmenter

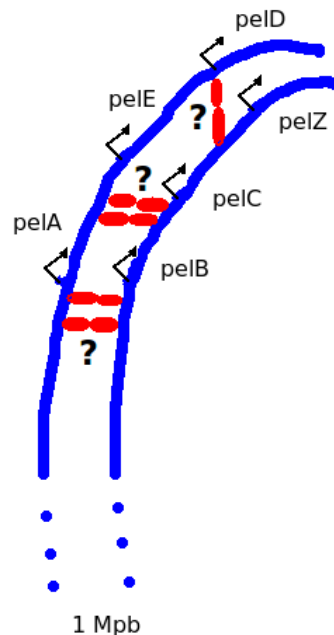


FIG. 4.3 – Le système Pel se compose notamment de deux clusters de 3 gènes séparés de 1Mpb qui pourraient entrer en contact et interagir. La présence de nombreux sites de fixation dans chaque région de régulation permettrait de payer un coût entropique important.

l'énergie de liaison est d'augmenter le nombre de sites de liaison. Le système que nous avons choisi se compose de deux clusters de 3 gènes qui pourraient interagir (voir Figure 4.3). Chacun des gènes est régulé par plusieurs facteurs de transcription différents. Par conséquent, ce système laisse la possibilité que la liaison se fasse en utilisant plusieurs sites de liaison des différents gènes : des facteurs de transcription pourraient faire des pont entre les régions de régulation de *pelA* et *pelB*, entre les régions de *pelE* et *pelB* et entre *pelC* et *pelZ*, par exemple. Ce système aurait donc la capacité à payer un coût entropique important susceptible d'entrer en jeu pour ce type d'interaction contrairement à un système simple faisant juste intervenir deux gènes corégulés par un même facteur de transcription.

### 2.3 Raison physiologique

L'entérobactérie *Erwinia chrysanthemi* est une bactérie à gram négatif qui peut provoquer ce qui est appelé la maladie de pourriture molle chez une large gamme de plantes [66]. Pour cela, elle produit plusieurs enzymes extracellulaires qui attaquent les composants des cellules des parois végétales. Parmi elles, figurent les cinq enzymes pectate lyases PelA, PelB, PelC, PelD et PelE codées par cinq gènes réunis au sein de deux clusters séparés par 1 Mpb. Elles jouent un rôle de première importance dans l'infection puisque ce sont elles qui entrent en jeu dans la dégradation des polymères de pectine composants essentiels de la paroi végétale [67]. Plus précisément, lors de l'induction d'une culture bactérienne, le PGA commence à être

dégradé par les pectate lyases et l'un des produits de cette dégradation, le KDG (2-keto-3-deoxygluconate) peut se lier au répresseur KdgR et l'inactiver. Tous les gènes du régulon de KdgR dont les gènes *pel* sont alors dérèprimés et fortement exprimés. Nous avons donc un système qui passe d'une sorte d'état latent où la production enzymatique est très faible à un état de très forte activité grâce à un signal environnemental inducteur. Ceci nous rappelle une caractéristique de la configuration de boucle d'ADN à l'origine de répression transcriptionnelle que nous avons identifiée dans le Chapitre 3. La physiologie du système et sa réponse est très proche des deux cas les plus étudiés où a été mise en évidence la formation de boucle d'ADN : l'opéron *lac* et également le phage  $\lambda$  où il y a également transition très nette entre deux états distincts (non induit à induit pour l'opéron *lac* et phase lysogène à phase lytique pour le phage  $\lambda$ ). La formation de boucle pourrait jouer un rôle pour faciliter ce type de transition (cf. Chapitre 3). Les caractéristiques concernant la physiologie et la dynamique de transition du système Pel en font donc également un bon candidat pour comporter une régulation génétique gérée par boucle d'ADN.

## 2.4 Raison pratique

Le modèle d'étude que nous avons choisi est celui de l'équipe "Facteurs de virulence chez la bactérie phytopathogène *Erwinia chrysanthemi*"<sup>3</sup> qui possède des années d'expériences sur ce modèle et notamment sur la régulation du système Pel. Le laboratoire possède en outre une bibliothèque de souches modifiées (répertoriées et conservées à -80°C) ainsi que de nombreux plasmides sur lesquels il est possible de faire des premières mesures sans les étapes de construction génétique. La bactérie est de plus assez proche phylogénétiquement de la bactérie modèle *Escherichia coli*.

# 3 Expériences réalisées sur la bactérie *Erwinia chrysanthemi*

## 3.1 Idée générale des expériences réalisées

Le but des expériences présentées dans ce chapitre est de chercher à mettre en évidence la possibilité d'une interaction à longue portée entre gènes corégulés (boucle intergénique). Par longue portée, nous voulons donc dire deux (ou plusieurs) gènes qui seraient éloignés sur une représentation 1D de l'ADN (quelques dizaines de milliers de paires de bases voire plus) mais qui se retrouveraient très proches dans une représentation réelle en 3D.

L'idée est de faire des expériences du type de celles réalisées par le biologiste Müller-Hill dans les années 90 [120, 121] où il avait mis en évidence le mécanisme de coopérativité par la formation d'une boucle d'ADN au sein de la région de régulation de l'opéron *lac*. Dans ces expériences, il faisait une mutation sur un ou deux des trois sites de liaison et il regardait l'influence sur la répression exercée par le facteur de transcription LacI sur l'expression de l'opéron *lac*. Il mesurait la répression en quantifiant l'activité de l'enzyme  $\beta$ -galactosidase dans les cas induit et non

---

<sup>3</sup>au sein de l'équipe de William Nasser, UMR Microbiologie, Adaptation et Pathogénie, Université Lyon 1 Campus de la Doua Bâtiment Lwoff, 10 rue Dubois, 69622 Villeurbanne.

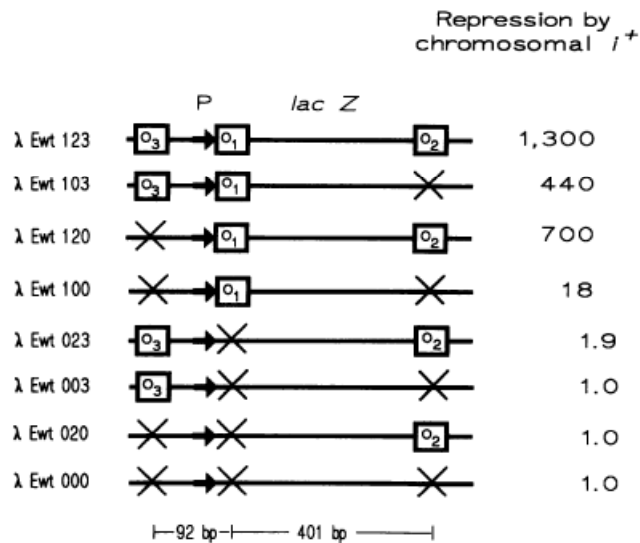


FIG. 4.4 – Résultats des expériences de Müller-Hill issus de [121] qui étudiaient l'influence de mutations d'un ou de deux des 3 sites de liaison de l'opéron lactose sur la répression. L'élimination d'un des 2 sites de liaison auxiliaires divise par 2 ou 3 la répression et l'élimination des deux sites auxiliaires divise la répression d'un facteur 70. Lorsque la boucle d'ADN n'est plus possible, l'effet de dérégulation est très fort.

induit. Les résultats importants de ces expériences sont rappelés dans la Figure 4.4. Il convient de retenir comme résultat que l'élimination d'un des 2 sites de liaison auxiliaires divise par 2 ou 3 la répression et que la mutation des deux sites auxiliaires en éliminant toute possibilité de boucle divise la répression d'un facteur 70, c'est ce type d'effet de dérégulation sur l'expression génétique que nous allons rechercher. Pour notre étude, nous recherchons un effet de coopérativité entre des sites de liaison de gènes différents et éloignés sur une représentation 1D parmi ceux du système Pel (Figure 4.5). Nous allons éliminer des sites de liaison d'une région de régulation d'un ou plusieurs gène(s) et regarder un éventuel effet sur l'activité enzymatique globale Pel qui résulte de l'expression d'autres gènes régulés par les mêmes facteurs de transcription. L'équipe de Müller-Hill utilisait pour quantifier la régulation génétique, la mesure de l'activité de l'enzyme  $\beta$ -galactosidase. Notre indicateur est la mesure de l'activité enzymatique Pel qui rend compte de l'expression de plusieurs gènes (voir Figure 4.2) notamment les 5 gènes codant pour les 5 enzymes majeures *pelA*, *pelB*, *pelC*, *pelD*, *pelE* mais également d'autres gènes (gènes du système de sécrétion *out*, gènes qui interviennent dans le catabolisme de la pectine, etc).

Le système Pel est plus complexe que celui de l'opéron *lac* car il fait intervenir un réseau de régulation aux nombreuses interactions. Ses propriétés dynamiques sont toujours en cours d'étude [76]. Il s'avère donc intéressant et nécessaire de suivre l'évolution dans le temps de cette activité étant donné que d'après les résultats récents obtenus par l'équipe grâce au travail de modélisation, le système n'atteint pas d'état stationnaire dans les conditions habituelles d'expérimentation (contrairement aux expériences de Müller-Hill sur l'opéron *lac*). En effet, il s'est avéré récemment que le substrat (PGA) était très vite consommé par les bactéries, ce qui implique que

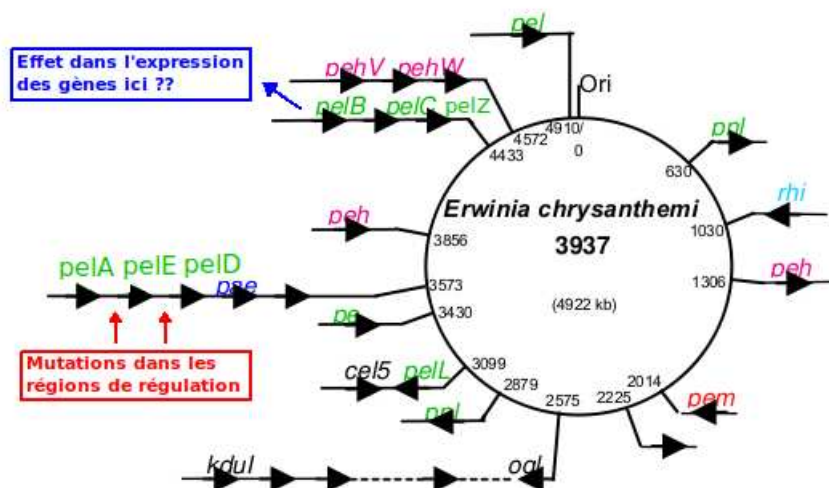


FIG. 4.5 – Représentation 1D du génome d'*Erwinia chrysanthemi*. Les principaux gènes entrant dans la virulence de cette bactérie phytopathogène sont représentés en couleur. L'idée générale des expériences est d'éliminer des régions de régulation d'un locus (gènes *pelA*, *pelE*, *pelD*) et de déceler un éventuel effet sur l'activité enzymatique Pel qui dépend de plusieurs gènes notamment des gènes *pelB* et *pelC* situés en un locus éloigné sur une représentation 1D du génome d'*Erwinia chrysanthemi*. Figure initiale issue de [88].

le système repasse après quelques heures dans un état non-induit. Il est donc utile de garder l'information sur la dynamique car il n'est pas impossible d'imaginer qu'une interaction par boucle d'ADN ait lieu à partir d'un moment précis de l'infection ou au retour vers l'état non induit.

### 3.2 Comparaison de deux mutants

Nous allons comparer deux mutants : un mutant appelé "Délétion" où les 3 gènes *pelA*, *pelE* et *pelD* ainsi que deux de leur région de régulation sont éliminées et un mutant appelé "Inactivation" où seuls les gènes ont été inactivés, laissant les régions de régulation intactes (voir Figure 4.6). Nous obtenons ainsi deux génomes comparables qui diffèrent seulement par la présence de deux régions de régulation. Nous pouvons alors tenter de voir un éventuel effet des régions de régulation sur les autres gènes *pel* ou sur les autres gènes régulés par les mêmes facteurs de transcription dont la dérégulation aurait un effet visible sur l'activité enzymatique Pel. En enlevant toute la région de régulation, nous pouvons en fait tester deux types d'interactions : notre première hypothèse est une interaction qui proviendrait d'une boucle intergénique faisant intervenir des sites de liaison particuliers (sites de KdgR, PecT, PecS...). La deuxième hypothèse serait une interaction des sites de fixation pour l'ARN polymérase. En effet, selon le concept proposé par Peter Cook mentionné plus haut d'*usines de transcription* [30], plusieurs ARN polymérases transcrivent des gènes à la chaîne au même endroit au sein de la cellule. Le fait d'éliminer un ou des sites de liaison pour l'ARN polymérase pourrait perturber le fonctionnement des usines et avoir des répercussions sur la transcription des autres gènes du groupe



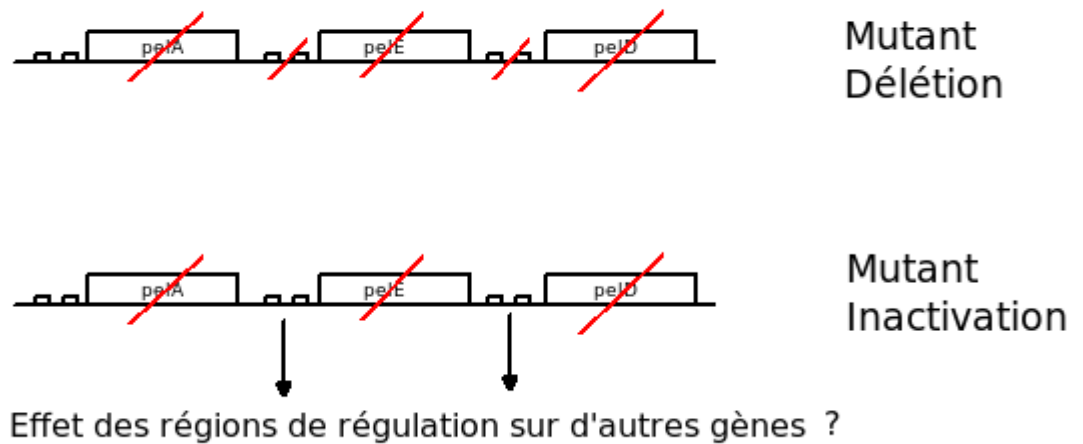


FIG. 4.6 – Comparaison de deux mutants : le mutant “Délétion” où les gènes *pelA*, *pelE* et *pelD* ainsi que deux de leur région de régulation ont été éliminées et le mutant “Inactivation” où les gènes *pelA*, *pelE* et *pelD* ont été inactivés mais les régions de régulation (sites de liaison pour différents facteurs de transcription et sites de fixation pour l’ARN polymérase) ont été laissées intactes pour détecter une éventuelle interaction de celles-ci avec d’autres gènes.

Pel. La comparaison de ces deux mutants permettrait donc de chercher d’une façon large, d’éventuels effets à distance.

## 4 Matériels et méthodes

### 4.1 Techniques d’ADN recombinant

Les préparations de plasmides, les digestions par enzymes de restriction, les ligations intra-moléculaires ou intermoléculaires, les électrophorèses sur gel d’agarose, les extractions de plasmides (“miniprep”) ont été réalisées en suivant les protocoles adaptés de [147]. Les transformations de bactéries se sont faites par la méthode chimique au  $CaCl_2$  et par choc thermique pour les bactéries *Escherichia coli* (méthode de Mandel et Higa). Des bactéries traitées avec une solution de  $CaCl_2$  froide puis soumises à un bref choc thermique à 42°C sont dans un état transitoire qui leur permet d’ingérer de l’ADN étranger (phage ou plasmide). Les bactéries *Erwinia chrysanthemi* ont été transformées par la méthode d’électroporation (une suspension de cellules et l’ADN à introduire sont soumis à une décharge électrique de haut voltage qui créent des pores au sein de la membrane cellulaire). La recombinaison homologue sur *Erwinia chrysanthemi* s’est faite en milieu bas phosphate comme décrit dans [141]. En cultivant les bactéries en milieu bas phosphate, celles-ci ont tendance à se débarrasser de leurs plasmides (il leur coûterait plus de les garder) mais en maintenant une pression de sélection forte par antibiotique, les bactéries ont comme

Souches d' <i>Erwinia Chrysanthemi</i>		
N° souche	Génotype ou description	Source
A4922	Type sauvage d' <i>Erwinia Chrysanthemi</i> 3937 issue de <i>Saintpaulia ionantha</i>	Collection laboratoire
A5089	Délétion des gènes <i>pelA</i> , <i>pelE</i> et <i>pelD</i> ainsi que deux de leur région de régulation (mutant "Délétion")	Collection laboratoire
A4476	Mutant <i>PecS</i> , gène inactivé <i>pecS</i> : :uidA Kan	Collection laboratoire
A4478	Mutant <i>PecT</i> , gène inactivé <i>pecT</i> : :uidA Kan	Collection laboratoire
A5107	Transduction de $\Delta pelAED$ : :Cm dans le mutant <i>pecT</i> : :uidA Kan	Ce travail
A5108	Transduction de $\Delta pelAED$ : :Cm dans le mutant <i>pecS</i> : :uidA Kan	Ce travail
A5109	Mutant <i>kdgR</i> , gène inactivé <i>kdgR</i> : :Sm (transduction WT A350 vers WT 3937)	Ce travail
A5160	gènes <i>pelA</i> , <i>pelE</i> , <i>pelD</i> inactivés en laissant les régions de régulation intactes (mutant "Inactivation")	Ce travail
A5161	Transduction de $\Delta pelAED$ : :Cm dans le mutant <i>kdgR</i> : :Sm	Ce travail

TAB. 4.1 – Différentes souches d'*Erwinia chrysanthemi* utilisées dans ce travail.

compromis d'insérer les gènes présents sur le plasmide par recombinaison homologue sur leur chromosome principal. Les fabrications de mutants par transduction ont été réalisées avec le phage  $\phi EC2$  en suivant [137] (la transduction est un transfert de matériel génétique par des bactériophages qui peuvent transmettre une mutation d'une souche de bactérie à une autre).

## 4.2 Souches bactériennes

Les souches utilisées et construites sont données dans le tableau 4.1.

## 4.3 Milieux et conditions de culture

Les cellules ont été cultivées en milieu riche LB pour *Escherichia coli* ou en milieu synthétique minimum M63 additionné d'une source de carbone pour *Erwinia chrysanthemi*. La température de culture a été de 30°C pour les bactéries *Erwinia chrysanthemi* et de 37°C pour les bactéries *Escherichia coli*, ce qui correspond à leur température optimale de croissance. Les cultures pour les deux bactéries se sont faites sous agitation mécanique (en général autour de 120 rpm), ceci pour faciliter la circulation de l'air dans le milieu et donc la croissance bactérienne. Les antibiotiques ont été utilisés avec les concentrations suivantes (en microgrammes par millilitre) : kanamycine 20, ampicilline 50, tétracycline 10, chloramphenicol 20.

## 4.4 Procédure suivie lors de cinétiques

Une cinétique commence par l'inoculation d'une préculture de nuit avec une (voire plusieurs) colonie(s) issue(s) de boîtes fraîches de Pétri. Celle-ci est en général réalisée dans les mêmes conditions (source de carbone, température, agitation...) que la culture. Le but de la préculture est de permettre aux bactéries de s'adapter au milieu liquide ce qui peut prendre un certain temps. Ce temps d'adaptation peut modifier le déroulement de la cinétique. Ensuite, un lavage est réalisé pour éliminer

les enzymes issues de la préculture (en centrifugeant pour garder les bactéries dans un culot et en renouvelant le milieu). Ceci est important pour ne pas commencer la cinétique dans des conditions initiales différentes, ce qui peut engendrer une grande variabilité dans les expressions même entre deux clones identiques. En effet, il est possible que le réseau génétique possède la propriété de bistabilité. Si le système dynamique démarre avec des conditions initiales différentes, nous pouvons obtenir deux cinétiques très différentes (expériences réalisées, figures non montrées).

L'inoculation du milieu pour la culture se fait en ajustant le volume d'inoculation pour partir sur des concentrations bactériennes identiques, ceci est important pour que les cultures soient les plus synchronisées possible au niveau des différentes phases du cycle de croissance et pour pouvoir mieux comparer ensuite. Le système de virulence Pel est en effet très sensible à la phase de croissance (notamment par la régulation de Fis [88] ou des processus de Quorum Sensing [117]). Le maximum d'activité enzymatique est couramment observé dans l'entrée de la phase stationnaire, quand la concentration bactérienne se stabilise.

Des échantillons du milieu de culture bactérienne ont été prélevés stérilement en différents temps et la densité optique (DO) a été mesurée par spectrophotométrie à 600 nm pour évaluer la turbidité du milieu et donc la concentration bactérienne. Après ajout de toluène et passage en hotte (pour éliminer un peu le toluène très toxique) les échantillons sont conservés à 4°C pour une mesure dans les deux jours suivants. Le toluène arrête le métabolisme et perméabilise les parois membranaires ce qui permet la libération d'éventuelles pectate lyases retenues dans l'espace périplasmique (c'est à dire entre les membranes interne et externe de la bactérie).

L'activité enzymatique Pel a été ensuite déterminée par spectrophotométrie en suivant la dégradation du PGA en des produits insaturés qui absorbent à la longueur d'onde 230 nm [111] (les pectate lyases clivent les liaisons osidiques  $\alpha$  du PGA et libèrent des oligogalacturonides insaturés). Plus précisément, les coefficients directeurs des courbes de production des oligogalacturonides insaturés sont calculés par régression linéaire et l'activité spécifique Pel est calculée suivant la formule<sup>4</sup> :

$$AS = 0.412 \times \frac{\Delta DO_{230}}{\Delta t} \times dilution\ enzyme / DO_{600} \quad (4.1)$$

où  $\frac{\Delta DO_{230}}{\Delta t}$  correspond à la pente de la production des oligogalacturonides insaturés, le terme *dilution enzyme* correspond à la dilution du milieu contenant les enzymes dans le tampon de réaction, le terme  $DO_{600}$  est la densité optique à 600 nm correspondant à la concentration bactérienne du milieu. La mesure s'est effectuée dans un tampon de réaction contenant Tris-HCl 0.1M,  $CaCl_2$  0.1 mM et PGA 1 g.L<sup>-1</sup>, pH 8 dans un bain à 37°C.

## 4.5 Construction du mutant "Inactivation"

Plusieurs constructions étaient possibles pour inactiver les trois gènes en laissant les régions de régulation intactes. Cependant, notre première hypothèse était que

---

<sup>4</sup>Cette formule provient de la relation :  $DO = \epsilon l c$  où  $\epsilon$  est le coefficient d'extinction molaire,  $l$  est la longueur de la cuve (1 cm) et  $c$  est la concentration en moles par litre. Le coefficient d'extinction molaire pour ce dosage est de 5200 c'est-à-dire que 1  $\mu mole/ml$  d'oligogalacturonides insaturés libérés par clivage du PGA donne une  $DO_{230}$  de 5.2. Sachant que 0.28 mg de poids sec bactérien (PSB) correspondent à une  $DO_{600}$  de 0.6, on peut calculer l'activité spécifique exprimée en  $\mu moles$  de produits libérés/min/mg de PSB :  $AS = \frac{\Delta DO_{230}}{\Delta t} \times dilution\ enzyme / 5.2 / (\frac{DO_{600}}{0.6} \times 0.28)$ .

le cluster composé des 3 gènes *pelA*, *pelE* et *pelD* pouvait interagir par boucle d'ADN avec le cluster des 3 gènes *pelB*, *pelC* et *pelZ*. Nous avons souhaité ne pas perturber la symétrie entre les différentes régions de régulations des deux différents clusters (voir Figure 4.3). Nous avons pour cela choisi d'insérer le gène de résistance (qui fait un peu plus de 1000 pb) au sein du dernier gène du cluster c'est-à-dire *pelD*. Nous avons alors opéré en plusieurs étapes. Nous sommes partis du plasmide PBGS 131 contenant les gènes *pelA*, *pelE*, *pelD* (fragment entre les sites SphI et FspI) insérés dans le plasmide entre les sites SphI et EcorV. La partie comprise entre les deux sites SalI du vecteur a été délétée. Le gène *pelA* a été inactivé en introduisant une séquence de nucléotides issue du site de polyclonage du plasmide pBC KS +/- (séquence comprise entre les sites HincII et EcoRI rempli avec l'enzyme d'ARN polymérase fragment Klenow) au niveau du site de coupure à bords francs Bst1107I. Le nombre de paires de bases introduites, non multiple de trois, décale le cadre de lecture du ribosome qui ne traduit plus les mêmes acides aminés. La séquence résultante ne permet plus en général, d'avoir une protéine fonctionnelle. Nous avons inactivé le gène *pelE* en agissant au niveau du site de coupure à bouts cohésifs de 4 paires de bases NotI, c'est à dire en coupant sur ce site de restriction, en remplissant les extrémités protubérantes grâce à l'enzyme Klenow puis en religuant le gène sur lui même. Le dernier gène *pelD* a été inactivé en insérant le gène de résistance au chloranphenicol issu du plasmide PCK17 (digéré par SmaI) au niveau du site de coupure à bouts cohésifs AsuII et traité à la Klenow pour le rendre à bouts francs. La construction ainsi réalisée sur plasmide a été insérée sur le chromosome principal par recombinaison homologue en milieu bas phosphate.

#### 4.6 Electrofocalisation et vérification du phénotype du mutant "Inactivation"

L'électrofocalisation nous a permis de trouver et de vérifier le mutants souhaité avec les 3 gènes *pelA*, *pelE* et *pelD* inactivés. L'électrofocalisation permet une séparation de protéines en fonction de leur point isoélectrique c'est à dire la valeur du pH pour lequel leur charge électrique résultante est nulle. L'électrofocalisation a été réalisée sur un gel épais de 4 mm de polyacrylamide contenant 35 % acrylamide (Pharmacia), 15 % Pharmalyte porteur en ampholytes dans la plage de pH 3-10. La focalisation s'est faite à puissance constante de 30 W pendant 90 min avec 0.1M NaOH comme cations et 0.1M  $H_2SO_4$  comme anions. Après l'électrofocalisation, le gel a été rincé dans un tampon contenant 0.1 M Tris-HCl, pH 8 pendant 15 min. L'activité pectinolytique a ensuite été révélée dans un tampon contenant 0.1 M Tris-HCl, pH 8, 1 mM  $CaCl_2$ , 15  $g.l^{-1}$  PGA en incubant le gel pendant 1h15 à 37 °C. Le gel a été abondamment rincé à l'eau. Les activités enzymatiques ont été révélées en immergeant le gel dans une solution de rouge de ruthenium 0.05% pendant une nuit.

Pour vérifier que nous avons bien obtenu un mutant "Inactivation" au phénotype souhaité avec les trois gènes inactivés, nous avons réalisé plusieurs électrofocalisations de protéines et testé plusieurs clones jusqu'à en trouver un où les trois gènes étaient inactivés.

Nous montrons ici le résultat du gel révélant le mutant "Inactivation" souhaité (Figure 4.7), les deux extrémités correspondent au type sauvage (les 5 enzymes

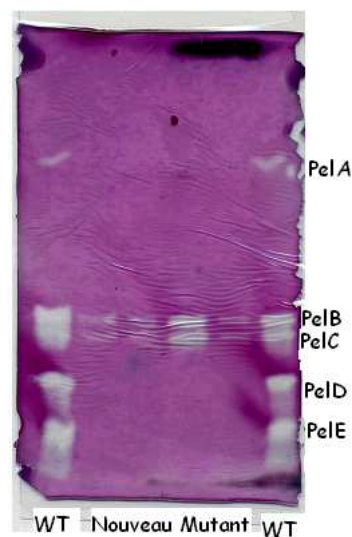


FIG. 4.7 – Résultat du gel de l'électrofocalisation montrant la bonne réalisation du mutant “Inactivation” souhaité.

majeures sont présentes), les couloirs du milieu correspondent au nouveau mutant “Inactivation” (extraits enzymatiques pris à plusieurs temps différents au cours du cycle bactérien). La difficulté à obtenir le mutant “Inactivation” exactement voulu est venue du fait que le gène de résistance étant placé à l'extrémité, la pression de sélection était relâchée pour l'autre partie pouvant recombiner (au niveau du gène *pelA*), ainsi à plusieurs reprises nous n'avions obtenu que des mutants avec seulement *pelE* et *pelD* d'inactivés. Le mutant “Délétion” a également été vérifié de la même façon (figure non montrée).

## 5 Résultats des expériences

Nous avons réalisé plusieurs cinétiques d'activité enzymatique Pel totale en testant plusieurs conditions différentes et en mesurant en différents temps : en regardant les différentes phases de croissance, en considérant un milieu inducteur ou non (c'est-à-dire avec ou sans PGA) et en changeant la source de carbone (glucose ou glycérol).

### 5.1 Comparaison mutant “Délétion” et mutant “Inactivation” dans milieu M63 + PGA + glucose

Les premières conditions de culture dans lesquelles nous avons réalisé la comparaison entre les deux différents mutants “Délétion” et “Inactivation” ont été les conditions habituellement utilisées par notre équipe. Ces conditions rendent compte d'un milieu inducteur c'est à dire un milieu minimum M63 (contenant des extraits de levure), du chlorure de calcium à une concentration de  $0,1mM$  (les enzymes Pel ont en effet besoin de  $Ca^{2+}$  comme cofacteur) et du glucose à 0.2% utilisé comme source de carbone. Nous reproduisons un exemple de cinétique montrant la crois-

sance bactérienne (mesure de densité optique) et montrant l'activité spécifique Pel totale pour les deux mutants "Déletion" et "Inactivation". Chaque valeur représente la moyenne d'au moins deux mesures indépendantes c'est à dire effectuées à partir d'enlenmeyers distincts.

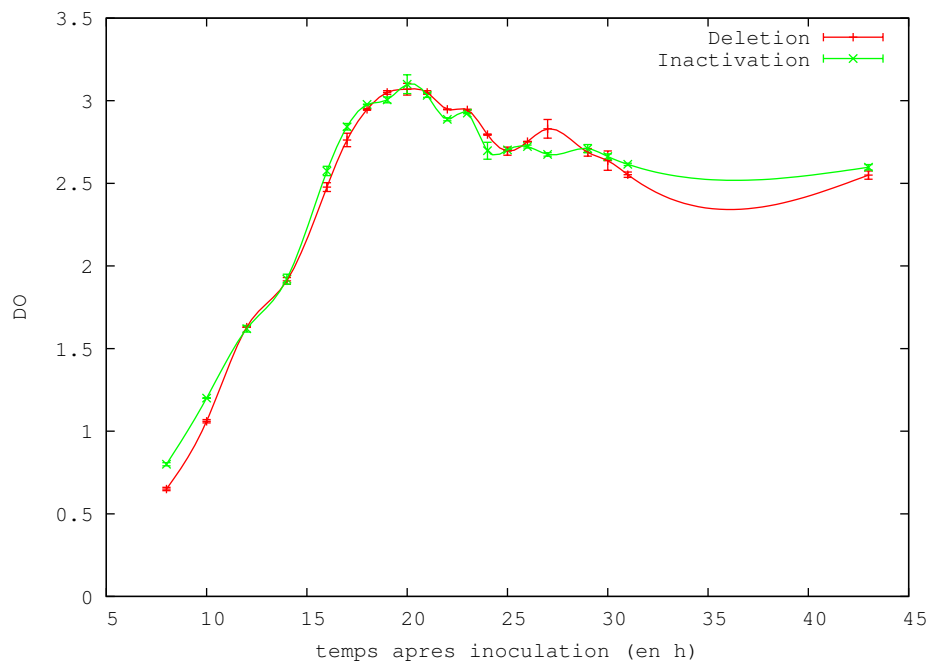


FIG. 4.8 – Exemple de cinétique : comparaison des densités optiques des milieux bactériens des deux mutants Déletion et Inactivation donnant l'évolution de la croissance bactérienne. Expérience réalisée dans un milieu minimum M63 + glucose 0.2% + PGA 0.2%. Chaque valeur représente la moyenne de deux expériences indépendantes. Les barres d'erreur indiquent l'écart type.

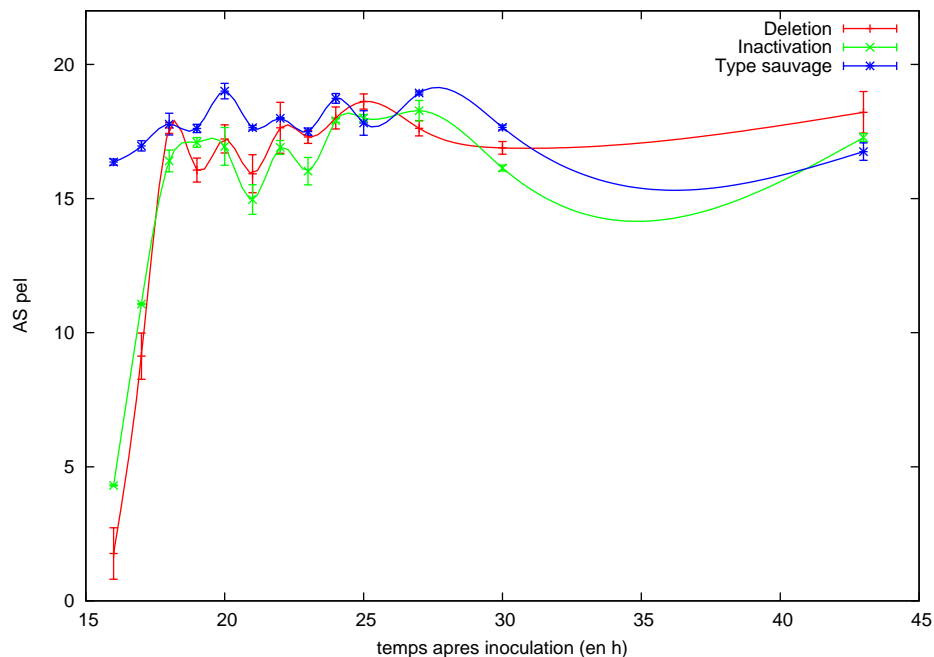


FIG. 4.9 – Exemple de cinétique : comparaison de l'activité spécifique Pel des deux mutants Délétion et Inactivation en milieu M63 + glucose 0.2% + PGA 0.2%. L'activité spécifique (AS) est exprimée en  $\mu$ moles de produits insaturés libérés par minute par milligramme de poids sec bactérien. Chaque valeur représente la moyenne de deux expériences indépendantes. Les barres d'erreur indiquent l'écart type.

Nous pouvons voir que les mutants Délétion et Inactivation se comportent de la même façon au niveau de la croissance et de la cinétique de l'activité enzymatique Pel. Par ailleurs, nous avons pu remarquer qu'un mutant ayant perdu les 3 gènes *pelA*, *pelE* et *pelD* entre en virulence environ 15 heures après l'inoculation soit plus tardivement que le type sauvage qui entre en virulence 10 heures après l'inoculation. Cependant, si l'on compare les niveaux activités enzymatiques Pel atteints par le type sauvage et les deux mutants auxquels les 3 gènes responsables de la virulence *pelA*, *pelE* et *pelD* ont été délétés ou inactivés, on peut voir qu'ils atteignent une activité spécifique semblable c'est à dire autour d'une valeur de 15. Ceci n'était pas attendu puisqu'en éteignant 3 des 5 gènes majeurs impliqués dans l'activité de virulence de cette bactérie, nous aurions plutôt envisagé pour ces mutants une activité spécifique Pel diminuée de moitié voire plus. Le fait d'enlever ces 3 gènes et de retarder l'entrée en virulence semble donc à un certain niveau entraîner un changement dans la régulation puisque les gènes majeurs restant *pelB* et *pelC* sont plus exprimés par rapport au type sauvage.

## 5.2 Comparaison mutant "Délétion" et mutant "Inactivation" dans milieu M63 + glycérol avec ou sans PGA

Le facteur de transcription global CRP réagit à la présence de glucose dans le milieu. Or nous avons vu (cf. chapitre 3) que CRP intervenait très souvent dans les mécanismes liés à la formation de boucle. En effet, il est possible que le facteur de transcription CRP entre en compte dans la formation ou non formation de boucle :

des expériences de cristallographie ont montré que CRP pouvait courber l'ADN [152]. CRP peut rompre par exemple la boucle de l'opéron Arabinose [97]. C'est pour cette raison qu'il nous a paru nécessaire de réaliser la comparaison entre les deux différents mutants dans des cultures avec une autre source de carbone que le glucose. Nous présentons ici deux exemples de cinétiques réalisées dans du milieu M63 contenant du glycérol à 0.4% seul puis en mode induit, en présence de PGA.

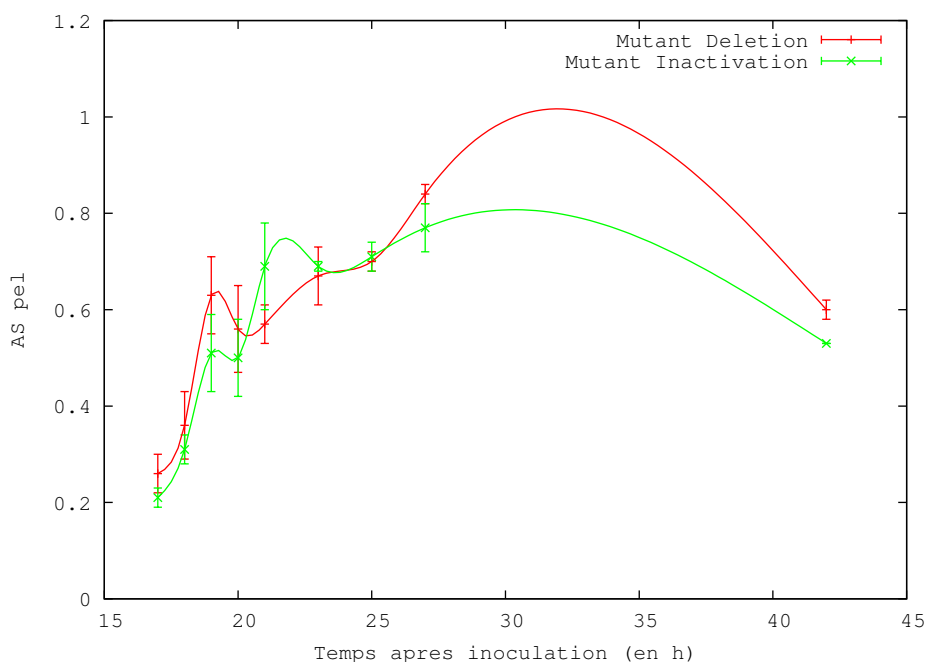


FIG. 4.10 – Exemple de cinétique : comparaison de l'activité spécifique Pel des deux mutants "Délétion" et "Inactivation" en milieu M63 + glycérol 0.4%. L'activité spécifique (AS) est exprimée en  $\mu$ moles de produits insaturés libérés par minute par milligramme de poids sec bactérien.

Nous n'avons donc pas observé de différences significatives entre les deux mutants dans une culture ayant le glycérol comme source de carbone.

### 5.3 Effet de la délétion dans 3 mutants d'*Erwinia chrysanthemi*

Pour essayer de mieux comprendre l'origine de l'effet de compensation par la souche qui a perdu 3 des 5 protéines majeures qui ont une activité enzymatique de dégradation de la pectine, nous avons étudié l'effet de cette délétion dans les principaux régulateurs du système Pel d'*Erwinia chrysanthemi*. L'effet de compensation semble se retrouver pour les mutants PecT et PecS mais pas pour KdgR.

Comme montré sur la figure 4.12, nous pouvons observer l'effet de compensation sur les mutants PecS et PecT. Pour ce dernier, la souche ayant perdu les trois gènes dépasse même la souche parentale ayant les 5 gènes. Pour le mutant KdgR, le niveau de la souche délétée reste à la moitié de la souche parentale ayant les 5 gènes. Ce dernier facteur de transcription pourrait donc être à l'origine de l'effet de dérégulation qui entraîne des niveaux d'activité spécifique Pel similaires entre les deux types de mutants.



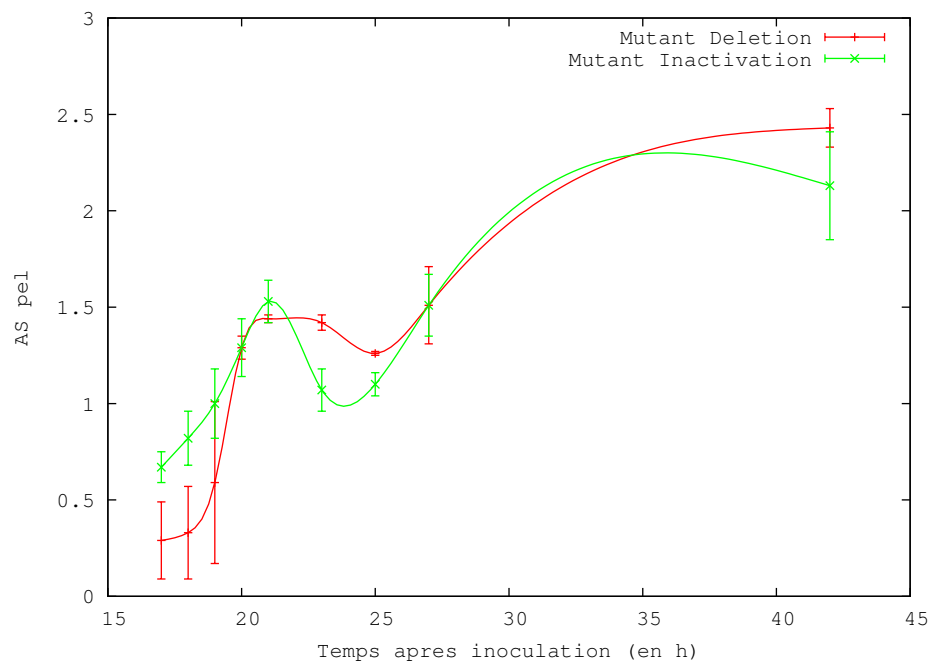


FIG. 4.11 – Exemple de cinétique : comparaison de l'activité spécifique Pel des deux mutants “Délétion” et “Inactivation” en milieu M63 + glycérol 0.4% + PGA 0.2%. L'activité spécifique (AS) est exprimée en  $\mu$ moles de produits insaturés libérés par minute par milligramme de poids sec bactérien.

## 6 Conclusions

Au vu des expériences réalisées et reproduites dans plusieurs conditions de cultures différentes, nous nous sommes dirigés vers la conclusion qu'il n'y a pas de différences phénotypiques entre les deux mutants. Donc, il n'y aurait pas d'effets des deux régions de régulation éliminées du premier cluster de gènes (gènes *pel A-E-D*) sur le second (gènes *pel B-C-Z*) ou éventuellement sur d'autres gènes *pel* ou gènes corégulés du génome d'*Erwinia chrysanthemi* (notamment les gènes des régulateurs de KdgR ou PecT) qui auraient entraîné des modifications dans l'expression génétique. Les deux mutants se comportent de la même façon au niveau de la croissance et au niveau de l'activité enzymatique Pel dans les différentes conditions physiologiques testées. Aucune différence dans les phénotypes des 2 mutants n'a pu être détectée de manière reproductible.

Pour ce système qui nous paraissait un très bon candidat pour une interaction à longue portée par boucle d'ADN et malgré les différents arguments apportés dans ce chapitre, il semble donc que ce mécanisme ne soit pas utilisé ici. Cette information sera discutée de façon plus large dans la partie Conclusion et Discussion. Nous pouvons par ailleurs toujours émettre l'hypothèse que l'effet n'est pas assez fort pour être visible de façon nette et reproductible. Il serait également nécessaire d'éliminer la région de régulation de *pelA* (expériences en cours). Nous pouvons aussi suggérer que le système puisse boucler d'une autre façon : la région de régulation de *pelB* entrerait en contact avec celle de *pelC* (ou celle de *pelA* avec *pelE*). La boucle ne serait longue que de 1200 pb (la longueur du gène *pelB*), nous serions dans une

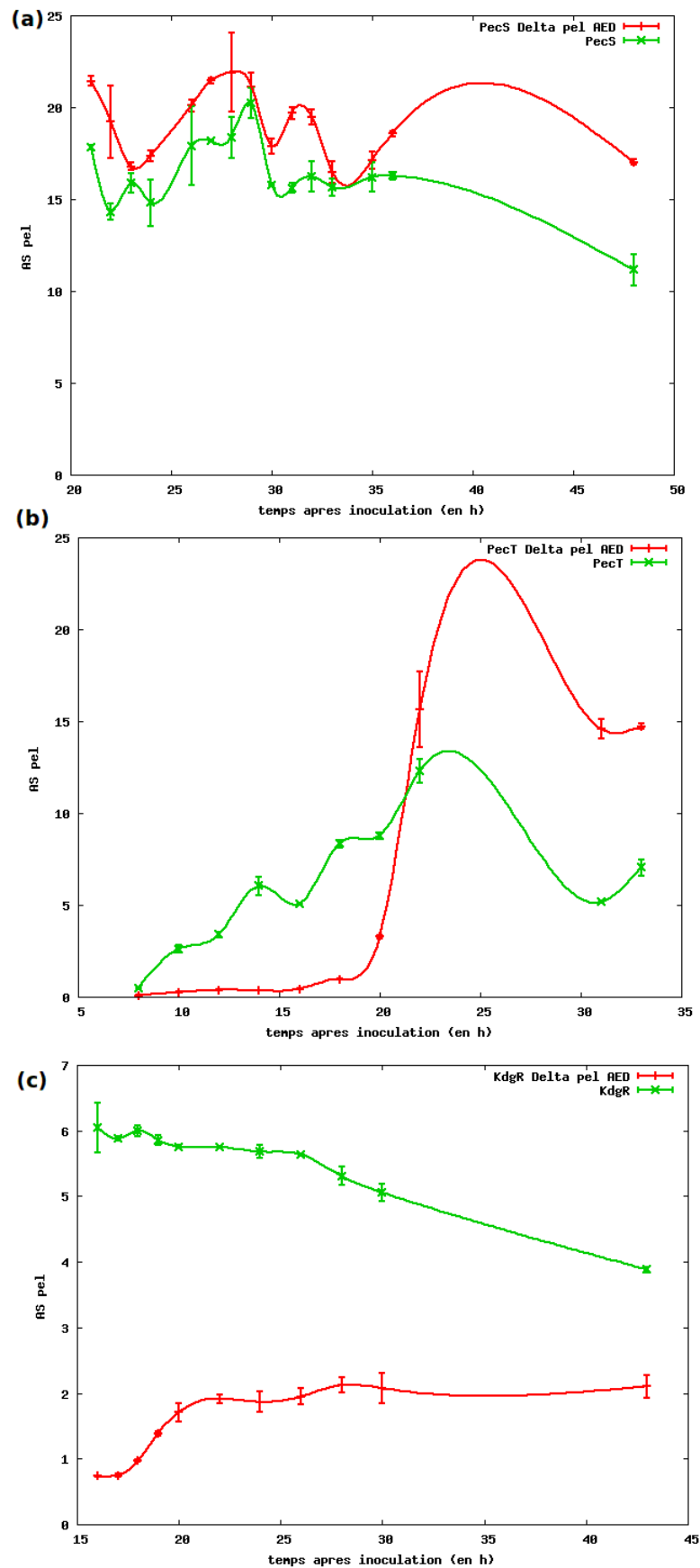


FIG. 4.12 – Cinétiques où l'on regarde l'effet de la délétion  $\Delta pelAED$  dans 3 différents mutants d'*Erwinia chrysanthemi* : (a) PecS, (b) PecT, (c) KdgR.

configuration assez proche de la boucle du phage  $\lambda$  (longue de 2400 pb) [187]. Nous espérons tester cette hypothèse prochainement.

Ces expériences ont par ailleurs révélé un phénotype intéressant : un mutant ayant perdu trois des cinq pectate lyases majeures restitue une activité enzymatique semblable au type sauvage. Ceci n'était pas attendu. Un effet de retard dans le maximum d'activité enzymatique Pel est toutefois observé entre le type sauvage et le mutant délété. L'origine de ce phénotype serait à éclaircir. Plusieurs mutants contenant la délétion des trois gènes et également délétés d'un gène codant l'un des principaux facteurs de transcription d'*Erwinia chrysanthemi* (PecS, PecT, KdgR) ont été construits (par transduction) pour voir si l'effet de compensation se retrouvait. Les premières cinétiques semblent montrer qu'il se retrouve pour 2 des principaux facteurs de transcription testés PecS et PecT mais pas pour KdgR. Cet effet est peut-être dû au retard dans l'induction du système. Du fait que la souche délétée parte avec un handicap dans le système Pel, l'entrée en virulence (le maximum d'activité enzymatique) se déclenche plus tard. A ce moment, l'environnement transcriptionnel est peut-être plus favorable (présence de plus d'activateurs ou quantité diminuée de répresseurs). Par conséquent, les gènes *pel* restant sont transcrits en plus forte quantité et la souche délétée rattrape la souche sauvage. Un exemple d'expérience pour tester cette hypothèse serait de comparer deux souches sauvages dont l'une est induite plus tardivement (en ajoutant le PGA dans le milieu au cours de la cinétique après 15h par exemple), peut-être que la souche induite plus tardivement aurait alors une plus forte activité enzymatique Pel. Ces expériences pourront ajouter des informations utiles pour la confrontation modélisation / expériences sur le système de virulence Pel qui a été initiée depuis plusieurs années au sein de l'équipe [154].

Enfin, nous pensons que la démarche présentée dans ce chapitre pourrait être testée chez d'autres systèmes de régulation et chez d'autres bactéries ou organismes plus complexes comme la Levure. Nous proposons dans le chapitre suivant une méthode qui vise à trouver de telles interactions en élargissant le champ d'investigation au génome complet de la bactérie.



# Chapitre 5

## Méthode pour rechercher des interactions 3D dans un génome de bactérie

Dans le chapitre précédent, nous avons cherché à savoir si une interaction à longue portée par boucle d'ADN était possible sur le système Pel de l'entérobactérie *Erwinia chrysanthemi*. Bien que la réponse à la question posée s'est dirigée vers la négative, nous pensons qu'il serait utile de poser la question différemment et de façon plus large : avec quels gènes ou régions de régulation un point donné d'un génome pourrait-il entrer en contact 3D proche et éventuellement interagir par boucle d'ADN ? Pour cela, nous proposons une méthode basée uniquement sur des techniques de biologie moléculaire traditionnelles couramment utilisées. Partant d'un point donné d'un génome, la méthode vise à trouver les endroits de celui-ci qui pourraient se retrouver proches de manière significative, assez longtemps pour pouvoir interagir.

### 1 Méthode pour tenter de trouver des interactions 3D

#### 1.1 Présentation du principe de la méthode

La méthode s'inspire de techniques couramment utilisées dans un laboratoire de biologie moléculaire comme le criblage de gènes avec l' $\alpha$  complémentation (clone bleu/blanc) ou le double hybride bactérien [74] et est basée sur l'utilisation du enhancer bactérien de l'opéron *glnA* d'*Escherichia coli* [175]. Partant d'un point choisi du génome, elle consiste à rechercher avec quelles autres parties du génome ce point pourrait entrer en contact proche.

Des expériences sur des plasmides réalisées par l'équipe de Kustu [175] montrent que lorsque les sites enhancer NtrC sont proches à 3D du promoteur de *glnA*, il y a activation transcriptionnelle : ces expériences sont basées sur des encaténations de plasmides (voir Figure 5.1). Les résultats montrent que l'activation transcriptionnelle se fait lorsque les plasmides sont liés et pas lorsqu'ils sont séparés pour une plage de concentrations en NtrC. Ces expériences montrent donc que quand deux segments d'ADN sont en contact proche 3D, il peut y avoir interaction et activation, ce système

nous paraît par conséquent approprié pour tester si deux segments d'ADN peuvent interagir bien que éloignés sur une représentation 1D dans un génome de bactérie. Il est à noter qu'il serait également possible d'utiliser l'un des autres systèmes enhancer d'*Escherichia coli* que nous avons répertoriés dans le chapitre 3 (cf. Tableau 3.1) puisqu'ils fonctionnent sur le même mécanisme.

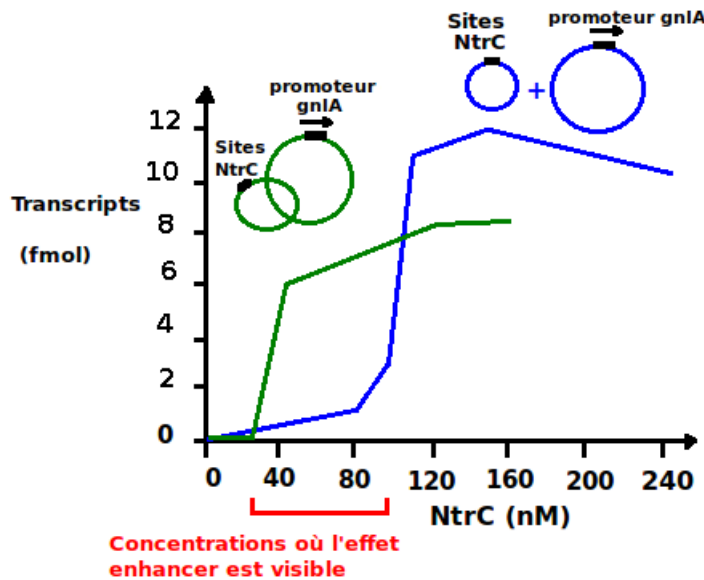


FIG. 5.1 – Les expériences réalisées par l'équipe de Kustu en 1990 sur le système enhancer du gène *glnA* d'*E.coli* montraient clairement que lorsque les parties sites NtrC enhancer et promoteur étaient proches à 3D du fait de l'encaténation (courbe verte), ils pouvaient interagir beaucoup plus que lorsque les mêmes plasmides étaient non-enchaînés (courbe bleue) pour une plage de concentrations en NtrC précise. Ce système nous paraît donc très intéressant pour rechercher des interactions entre segments d'ADN notamment à distance au sein d'un génome. Figure adaptée de la référence [175].

Principe général de la méthode (voir schéma Figure 5.4) :

On insère par recombinaison homologe, la construction *A* contenant le promoteur de *glnA*, le gène *lacZ* et un gène de résistance à un antibiotique, dans un gène ou dans une région intergénique où l'on désire connaître avec quelles autres parties du génome cet endroit pourrait entrer en contact 3D proche. Ensuite comme dans une mutagenèse aléatoire, on infecte la précédente souche avec la construction *B* qui est un phage Mu ou un transposon contenant les sites enhancers NtrC ainsi qu'une cassette de résistance à un antibiotique. Le phage Mu (ou le transposon) va s'insérer un peu partout dans le génome, peut être dans des endroits qui sont proches à 3D de la construction *A* introduite. Plusieurs centaines de colonies sont alors cultivées sur boîtes de Pétri. Si le phage s'insère dans un endroit proche à 3D, il y a activation transcriptionnelle de la fusion, le gène *lacZ* codant pour la  $\beta$ -galactosidase est exprimé et le ou les clones apparaissent bleu(s). On dose ensuite l'activité enzymatique par spectrophotométrie et on la compare aux témoins négatifs et positifs (voir Figure

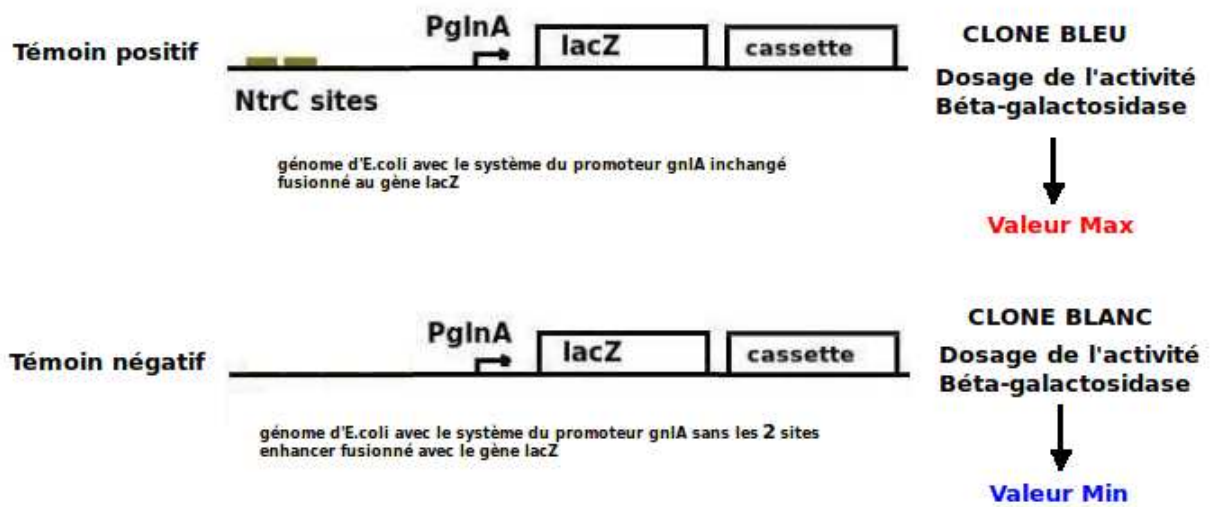


FIG. 5.2 – Le témoin positif et le témoin négatif de la méthode proposée. On doit avoir un clone bleu et une forte activité enzymatique de la bêta-galactosidase pour le témoin positif et un clone blanc et une faible activité  $\beta$ -galactosidase pour le témoin négatif.

1.1). Notre témoin positif sera le génome d'*Escherichia coli* avec le système du promoteur *glnA* inchangé et fusionné au gène *lacZ*. Il y a donc dans ce cas, interaction par boucle (cf. chapitre 3) et la machinerie transcriptionnelle se met en marche. Dans ce cas, on doit avoir un clone bleu et une activité de la  $\beta$ -galactosidase forte. Notre témoin négatif sera le génome d'E.coli avec le promoteur de *glnA* fusionné au gène *lacZ* et sans les sites enhancers. Ici, on doit avoir un clone blanc (après étalonnage si besoin) et une activité enzymatique de la  $\beta$ -galactosidase faible.

Il est à noter que notre témoin positif n'en est pas vraiment un (la boucle du système de *glnA* est intragénique) car il faudrait un système connu de boucle d'ADN entre deux gènes éloignés. Or à notre connaissance, aucune boucle de ce type n'a été détectée chez des génomes de procaryotes.

Parmi le ou les clones bleu(s) trouvés, on retrouve où s'était inséré le phage ou le transposon (comme en mutagénèse aléatoire) en séquençant la partie d'ADN bactérien qui est à côté du phage ou du transposon inséré puis on repère sur le génome les parties interagissantes.

## 1.2 Différentes étapes envisagées

1. La première étape consiste à construire les témoins positif et négatif pour s'assurer qu'il est possible de bien les distinguer par leur phénotype (couleur sur boîtes de Pétri et activité enzymatique  $\beta$ -galactosidase en milieu liquide). Il s'agit d'extraire le gène *glnA* et sa région de régulation par PCR pour le cloner dans un plasmide. Il faut ensuite insérer une fusion *lacZ* + gène de résistance à un antibiotique dans le gène *glnA*, faire une délétion des sites enhancer pour le témoin négatif et finalement insérer les constructions par recombinaison homologue dans le chromosome principal. Cette première étape

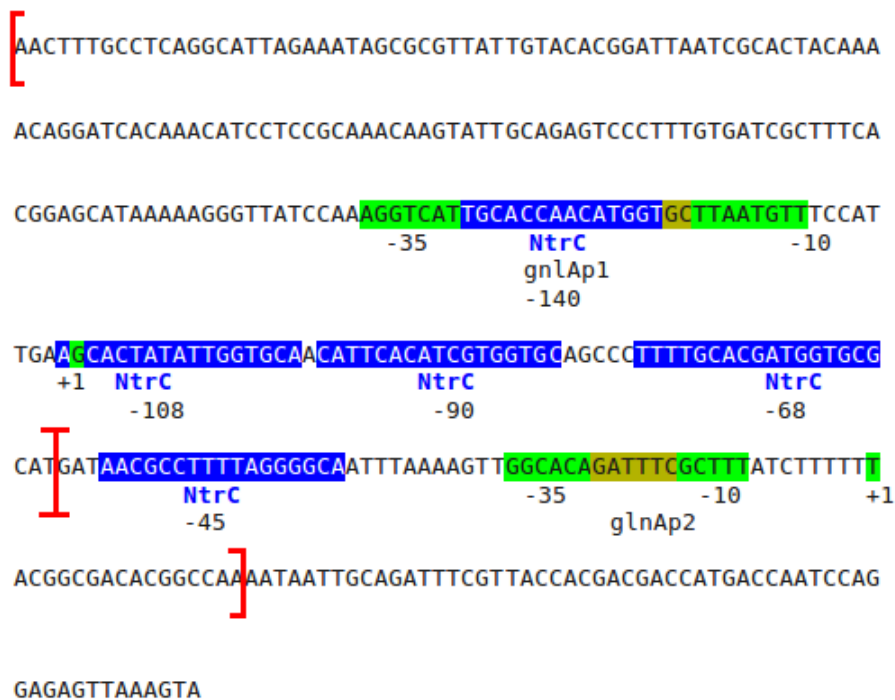


FIG. 5.3 – Séquence de la région de régulation du gène *glnA* chez *Escherichia coli* (issue de la base de données RSAT <http://rsat.ulb.ac.be/rsat/>), les différents sites pour NtrC sont indiqués en bleu, les deux promoteurs en vert. Les parties qu'il faut extraire sont indiquées entre crochets rouges.

permettrait de tester la faisabilité de la démarche et de voir s'il est possible de distinguer les deux témoins par leur phénotype.

2. Ensuite pour réaliser les constructions A et B, il s'agit par PCR d'obtenir les deux parties du système enhancer de l'opéron *glnA* d'*E.coli* : les sites enhancers et le promoteur de l'opéron *glnA*. La Figure 5.3 donne la séquence ainsi que la position des différents sites de liaison pour NtrC et les deux promoteurs, et indique les régions à extraire : une partie contenant les 2 sites enhancers d'NtrC (ou les 4 sites d'NtrC) et une partie contenant le promoteur 2 et un site NtrC. Plusieurs configurations sont possibles : l'effet enhancer montré sur la figure 5.1 fonctionne avec seulement 2 sites enhancer et la partie promoteur et un site d'NtrC.
3. Construction génétique A de la fusion : promoteur de *glnA* + gène *lacZ* de la  $\beta$ -galactosidase + un gène de résistance à un antibiotique. Cette construction ferait intervenir les techniques de biologie moléculaire propres au clonage de gène. Il est à noter que la fusion gène *lacZ* de la  $\beta$ -galactosidase + un gène de résistance à un antibiotique se trouve déjà dans de nombreux plasmides commerciaux ou dans des plasmides que l'on trouve facilement en laboratoire.
4. Clonage du gène (ou d'une région intergénique) dont on voudrait savoir les possibles points de contact 3D dans un plasmide (ou utilisation d'un tel plasmide déjà existant en laboratoire). On insère la construction A au sein du gène



- contenu sur le plasmide.
5. Puis par recombinaison homologue, on insère la construction précédente sur le chromosome principal.
  6. Construction génétique B d'un phage Mu ou d'un transposon contenant la partie enhancer de l'opéron *glnA* (c'est à dire les 4 ou 2 sites de liaison enhancer de l'activateur NtrC) ainsi qu'un gène de résistance à un antibiotique.
  7. Infection par le phage Mu ou transposon de la culture bactérienne qui contient la construction A dans son génome. Cette étape est la même qu'une mutagenèse aléatoire. Le phage ou transposon va s'insérer un peu partout dans le génome et peut-être à des endroits qui peuvent entrer en contact 3D avec l'appât.
  8. Culture sur milieu X-gal : s'il y a interaction, les colonies sont bleues.
  9. Dosage spectrométrique de l'activité  $\beta$ -galactosidase pour quantifier l'intensité de l'interaction (comme réalisé dans la technique du double hybride bactérien). Si l'interaction est très significative, on doit avoir une valeur proche de notre témoin positif.
  10. Pour savoir où s'est inséré le phage Mu ou le transposon : comme dans une mutagenèse aléatoire, on extrait le chromosome principal puis digestion par une enzyme de restriction qui coupe fréquemment / différenciation sur gel d'agarose / clonage des bouts / transformation / étalement sur milieux doublement sélectifs / extraction plasmidique.
  11. Séquençage et repérage sur le génome des points qui seraient entrés en contact 3D.

### 1.3 Points importants et difficultés anticipées de la méthode

L'avantage d'utiliser la machinerie génétique de la cellule permettrait de ne "filtrer" que des interactions qui ont une durée significative et d'éliminer des interactions transitoires. En effet, comme les sites enhancers permettent un accrochage plus efficace de l'ARN polymérase, celle-ci entre alors en jeu pour transcrire le gène de la  $\beta$ -galactosidase. Si l'interaction est juste transitoire, le mécanisme n'aura pas le temps de se mettre en route. Nous espérons donc ne pas avoir trop de faux-positifs avec la méthode proposée et trouver des gènes qui sont à côté de manière significative dans la cellule. Comme dans la méthode du double hybride bactérien, les clones de couleur bleue sont dosés par spectrophotométrie et comparés aux témoins négatif et positif. L'intensité de l'activité enzymatique donne une information sur la qualité de l'interaction.

La difficulté majeure nous semble que le système puisse bien répondre en bleu ou blanc lorsqu'il y a interaction. Plusieurs tests au préalable avec les deux témoins sont donc nécessaires car on doit pouvoir bien les différencier sur le phénotype visible et sur l'intensité de l'activité enzymatique. Il est peut-être nécessaire d'avoir une concentration intracellulaire précise de NtrC pour bien distinguer les deux cas. Pour cela, il faut peut-être inactiver le gène qui code pour le régulateur NtrC et le mettre sur un plasmide avec un promoteur plus faible ou plus fort ou bien un promoteur contrôlable comme un promoteur thermosensible pour pouvoir se situer dans la zone de concentration où l'effet enhancer est bien visible (voir Figure 5.1).

**METHODE RECHERCHE D'INTERACTIONS 3D PAR BOUCLE D'ADN DANS UN GENOME DE BACTÉRIE**

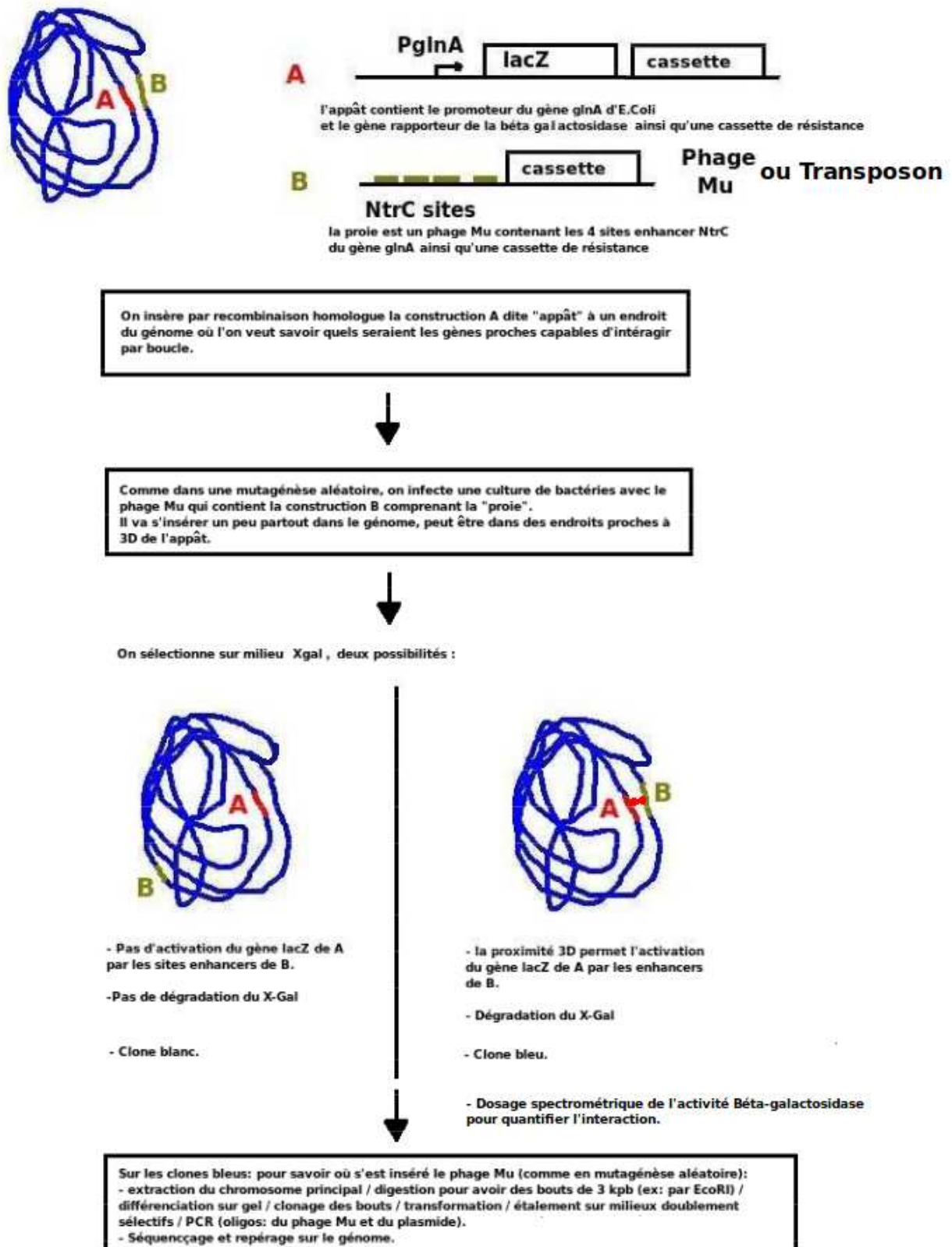


FIG. 5.4 – Schéma explicatif de la méthode proposée.

Une autre possibilité serait de modifier la quantité d'azote dans le milieu de culture, par exemple il doit être possible de modifier la quantité en sulfate d'ammonium dans le milieu de culture minimum M63. La concentration de NtrC intracellulaire dépendant de la concentration en azote du milieu de culture, cela permettrait de l'ajuster jusqu'à obtenir un milieu pour lequel les deux témoins sont parfaitement différenciables. Cette seconde possibilité serait plus rapide que la première, à la condition que l'on ne soit pas dans des conditions trop stressantes pour la bactérie.

Une autre difficulté à prendre en compte est celle que l'on retrouve dans une mutagenèse aléatoire : le phage Mu ou le transposon ne s'insère pas exactement partout dans le génome de la bactérie donc cela peut introduire un petit biais et des endroits non explorés.

## 1.4 Comparaison avec d'autres méthodes déjà existantes

La méthode nous paraît plus accessible avec les techniques dont dispose un laboratoire classique de biologie moléculaire que les méthodes dite 3C, 4C [122]... (qui ont été développées surtout pour les génomes eucaryotes jusqu'à présent). Par rapport à la technique 4C, la méthode proposée dans ce chapitre, puisqu'elle est basée sur des interactions protéine-ADN qui doivent durer assez longtemps pour enclencher la machinerie transcriptionnelle, permettrait de ne garder que des interactions assez significatives et donnerait a priori moins de faux positifs. La technique 4C étant basée sur les méthodes biochimiques du cross-linking, fait une sorte de cliché des interactions entre parties d'ADN et donne beaucoup d'interactions qui n'ont peut être pas de sens du point de vue biologique.

L'autre méthode qui, à notre connaissance, a été développée pour trouver des interactions à longue portée au sein de génome est celle développée par l'équipe de Frédéric Boccard [170]. Cette méthode est basée sur la recombinaison de séquences d'ADN spécifiques (sites att). La différence que nous voyons avec celle-ci est que l'approche proposée dans ce chapitre permet de filtrer les segments d'ADN pouvant interagir par boucle d'ADN, mécanisme qui est suggéré par plusieurs auteurs d'être à l'origine de la condensation et de l'organisation des génomes procaryotes et aussi eucaryotes [86, 103]. De plus, elle nous paraîtrait complémentaire et permettrait de vérifier l'existence et la fonction des macrodomaines. A priori, il serait logique d'affirmer que des boucles entre gènes éloignés peuvent se former au sein d'un macrodomaine mais pas entre gènes de macrodomaines différents.

De façon générale, nous savons qu'il a fallu plusieurs méthodes et approches différentes (biochimiques ou génétiques) pour prouver que deux protéines interagissent ensemble (double hybride, pull down, co-immunoprécipitation, etc) et ainsi pouvoir dessiner des cartes d'interaction entre protéines. Nous pensons que la méthode que nous proposons dans ce chapitre permettrait un regard différent et serait ainsi complémentaire aux techniques déjà existantes pour trouver et valider des interactions à longue portée entre segments d'ADN au sein d'un génome. Le recouvrement des résultats issus des différentes approches permettrait des validations et les confrontations nécessaires.

## 1.5 Expériences ultérieurement envisageables

Si cette méthode donnait des clones bleus et des gènes qui semblent entrer en contact et interagir, il serait alors possible de voir s'il y a effectivement interaction par boucle d'ADN entre les gènes révélés proches. Ceci pourrait être établi en faisant par exemple une délétion des régions de régulation concernées (comme dans le chapitre 4) ou par une mutation dirigée sur des sites de liaison et voir d'éventuels effets dans l'expression des gènes concernés. Nous pourrions également réaliser la méthode sur des mutants de protéines connues ou suspectées de jouer un rôle dans l'organisation du polymère d'ADN comme les histone like protéines (H-NS, Fis...) ou des facteurs globaux comme CRP soupçonnés d'intervenir dans l'organisation globale du génome comme celui d'*Escherichia coli* [56].

## 2 Hypothèses pour le rapprochement 3D de gènes

Nous pensons que l'approche expérimentale proposée dans ce chapitre pourrait peut-être permettre de trouver des interactions par boucle d'ADN à longue portée. D'une manière générale, elle viserait à apporter des éléments de réponse sur la façon dont un polymère d'ADN s'enroule dans l'espace à l'échelle du génome : quels gènes ou quelles régions inter-géniques se retrouvent proches à 3D et y a-t-il une raison particulière à ces rapprochements 3D ? Nous donnons ici des exemples d'hypothèses et de contraintes qui pourraient agir sur la conformation 3D des génomes bactériens et positionner deux gènes ou plusieurs côtes à côtes au sein de la bactérie :

1. Hypothèse de contraintes dues aux processus liés à la transcription : des gènes co-régulés par un même facteur de transcription se retrouvent proches à 3D, hypothèse présente sous des formes différentes dans plusieurs travaux (cf. Chapitre 4).
2. Hypothèse de contraintes post-traductionnelles : des gènes codant des protéines qui doivent s'agglomérer pour former des complexes hétéroprotéiniques pourraient se retrouver proches à 3D. Par exemple, chez *Escherichia coli*, la protéine IHF se compose de 2 sous protéines codées par deux gènes *ihfA* et *ihfB* qui sont positionnés à des endroits différents sur une représentation 1D (séparés d'environ 700 000 pb). Cela serait intuitivement plus facile pour la formation du complexe que les 2 gènes se retrouvent proches à 3D si on fait l'hypothèse que la transcription et la traduction se font au même endroit chez les procaryotes. Cette dernière hypothèse n'est cependant pas bien établie et des travaux expérimentaux semblent montrer que même si les procaryotes ne possèdent pas d'enveloppe nucléaire, la transcription et la traduction se font en des endroits différents au sein de la bactérie.
3. Hypothèse évolutive : dans l'histoire des génomes, des gènes se sont souvent dupliqués et ont ensuite évolué différemment. Les duplications se font souvent de façon proche sur la molécule d'ADN comme par exemple les gènes *pelB* et *pelC* d'*Erwinia chrysanthemi* issus d'un même gène ancestral et qui ont co-évolué tout à côté, nous pourrions donc également voir si parmi les duplications de gènes, certaines se sont faites sur des endroits proches à 3D.
4. Hypothèse des séquences répétitives : de façon intéressante, il existe chez *Escherichia coli* mais aussi chez d'autres organismes procaryotes des séquences

souvent palindromiques qui se répètent et que l'on retrouve éparpillées un peu partout dans le génome (pour une revue, on pourra consulter [12]). Il existe par exemple les BIMES (pour Bacterial Interspersed Mosaic Elements) chez plusieurs entérobactéries. Ce sont des séquences composées de plusieurs motifs qui se répètent : on en trouve environ 500 exemplaires chez *Escherichia coli*. Même si une fonction d'atténuation pour la transcription a été proposée [40] pour ces séquences, leur fonction reste mystérieuse. Nous savons qu'ils sont répartis dans tout le génome et aussi que certains d'entre eux contiennent un site de liaison pour la protéine IHF [39] (qui peut courber l'ADN). Enfin, il a également été montré que l'enzyme gyrase (qui peut séparer le double brin, agir sur le surenroulement de l'ADN...) peut aussi se lier à ces BIME [39]. Il est donc tout à fait tentant de proposer un rôle dans l'organisation spatiale des génomes pour ces séquences. Il nous semble par conséquent qu'il est important de tenir compte des informations concernant par exemple la position de ces séquences (<http://www.pasteur.fr/recherche/unites/pmtg/repet/>) dans l'étude des relations entre les positions tri-dimensionnelles des gènes. Nous donnerons plus de détails sur ce point dans la partie Conclusion et Discussion.

5. Autres hypothèses envisageables : d'une façon générale, on peut dire que les contraintes qui régissent l'organisation 1D des génomes bactériens (organisation sous forme d'opérons, gène d'un régulateur proche de son gène régulé par exemple) pourraient également se retrouver dans l'organisation 3D. On peut aussi garder la possibilité d'absence totale de contraintes : le génome adopterait finalement une structure libre où les positions 3D des gènes n'ont pas d'influence.

## 2.1 Conclusion

D'une manière générale, le but de la démarche expérimentale présentée dans ce chapitre serait de voir si des boucles d'ADN sont possibles entre régions éloignées sur une représentation 1D d'un génome. Il serait important d'en mettre clairement une en évidence, puisqu'à notre connaissance aucune interaction à longue portée de ce type n'a encore été mise au jour. La méthode pourrait permettre dans tous les cas (génome organisé ou pas) d'obtenir des informations sur la façon avec laquelle se replie le polymère d'ADN à l'échelle du génome : quelle serait la taille moyenne des boucles d'enroulement, si toutes les parties ou aucune peuvent interagir. Ceci pourrait apporter des éléments nouveaux ou des confirmations sur les connaissances acquises concernant l'organisation spatiale des génomes bactériens sur le plan expérimental.

Bien que nécessitant du temps à mettre en place, il nous a paru intéressant de proposer et de présenter cette approche car elle ne fait appel qu'à des techniques de biologie moléculaire couramment utilisées dans un laboratoire de microbiologie. De plus, une fois les deux premières constructions sont réalisées, la méthode pourrait être appliquée sur d'autres organismes procaryotes.



# Conclusions et Discussion

Cette thèse s'est intéressée à plusieurs aspects concernant la régulation génétique sur les plans théorique et expérimental. Un premier travail a été consacré à l'aspect temporel qui peut entrer en jeu dans des systèmes de régulation. Dans le cadre de la modélisation mathématique de systèmes biologiques, nous nous sommes proposés de trouver des réseaux de régulation simples qui répondent de manière optimale lorsqu'ils sont soumis à une stimulation périodique comme cela est souvent le cas dans la nature. Le réseau appelé "Incoherent Feed Forward Loop" s'est avéré présenter la propriété intéressante de ne laisser passer que des trains de pulses au profil temporel particulier. Des extensions de ce motif ("Diamond", "Double Diamond"... ) ont été suggérées comme pouvant présenter des propriétés intéressantes pour traiter des signaux dépendant du temps, plus complexes. Ces résultats théoriques doivent être replacés dans le cadre difficile de l'expérimentation et doivent tenir compte de l'aspect stochastique inhérent à ces processus génétiques. Cependant, nous pensons que les propositions théoriques abordées dans ce travail peuvent stimuler un travail expérimental.

Nous avons ensuite présenté plusieurs travaux liés aux questions concernant l'impact que la configuration du polymère d'ADN à petite ou grande échelle peut ou pourrait avoir sur la régulation génétique. Un travail de revue sur les boucles d'ADN a réuni les différents cas connus aujourd'hui, de petites boucles d'ADN présentes au sein de région de régulation chez la bactérie *Escherichia coli* et a répertorié les différentes méthodes expérimentales qui ont permis de mettre en évidence de telles boucles. Il nous a également semblé intéressant de répertorier les avantages que pourraient procurer la formation de boucle d'ADN pour la régulation génétique. La plupart d'entre eux sont pour l'instant au stade de propositions théoriques provenant notamment de travaux de physique statistique. Parmi ces avantages, la proposition que la boucle d'ADN assurerait une meilleure résistance au bruit et une population plus homogène nous semble une hypothèse très intéressante qui serait possible de confronter à l'expérimentation. Les techniques expérimentales, bien que lourdes, existent et ont déjà été utilisées pour mettre en évidence de telles propriétés. Nous avons remarqué plusieurs points communs sur les systèmes de boucle d'ADN connus concernant le type de régulation ou la présence de certains facteurs de transcription, notamment celle de la protéine CRP qui semble avoir un rôle prépondérant dans les systèmes de boucle d'ADN. Partant de ces points communs, nous avons recherché dans les bases de données des régions de régulation qui présenteraient des caractéristiques semblables. Nous pensons que des boucles d'ADN existent au sein d'autres régions du génome d'*Escherichia coli*. Le travail exposé dans le chapitre 3, en présentant des candidats probables de boucles d'ADN, vise à motiver une recherche et une mise en évidence expérimentale de ces boucles.

Dans le quatrième chapitre, un travail expérimental de biologie moléculaire est présenté. Ce travail s'est attaqué à tester une hypothèse présente dans la littérature depuis 2003, dans plusieurs travaux issus de la bio-informatique et de la physique statistique mais peu connue par la communauté des microbiologistes. La question testée est la suivante : est-ce que des facteurs de transcription pourraient se lier à des sites de liaison appartenant non pas à la même région de régulation d'un gène (comme dans le cas de l'opéron *lac*) mais à des sites de liaison appartenant à des gènes différents ? Après avoir réfléchi sur le choix de notre système d'étude, nous avons décidé de réaliser les expériences sur le système Pel de la bactérie phytopathogène *Erwinia chrysanthemi*. Ce système fait intervenir de nombreux facteurs de transcription aussi bien globaux (CRP, H-NS, Fis...), qui interviennent chez de nombreuses autres bactéries, ainsi que des facteurs de transcription plus particuliers à *Erwinia chrysanthemi* comme KdgR, PecT ou PecS. L'expérience a consisté à comparer deux génomes : un génome auquel ont été enlevés 3 gènes ainsi que deux de leur région de régulation (*pelA*, *pelE* et *pelD* faisant partie des 5 gènes majeurs responsables de la virulence chez cette bactérie) et un second génome où les mêmes 3 gènes ont été inactivés mais en laissant les régions de régulation intactes. Les deux génomes ainsi construits ne différaient donc que par la présence dans l'un, de deux régions de régulation qui pouvaient interagir avec d'autres gènes régulés par les mêmes facteurs de transcription. Nous avons recherché un effet de dérégulation qu'aurait pu entraîner la perte de ces deux régions de régulation sur deux signaux : la croissance bactérienne et l'activité enzymatique Pel (qui dépend de nombreux processus et donc de nombreux autres gènes). Nous avons réalisé et reproduit plusieurs cinétiques dans différentes conditions physiologiques (induit ou non induit : présence ou non du PGA qui peut être utilisé comme source de carbone) et dans des milieux aux sources de carbone différentes (glucose, glycérol...). Nous avons conclu qu'il n'y avait pas de différences entre les deux mutants et qu'ils se comportaient de la même manière tant au niveau de la croissance qu'au niveau de l'activité Pel. Autrement dit, nous n'avons pas observé de différence reproductible sur l'ensemble des cinétiques réalisées.

La possibilité que les régions de régulation éliminées bouclent avec des régions de régulation de gènes dont la dérégulation entraînerait un effet trop faible pour être détecté n'est pas totalement à exclure ou également que la région de régulation de *pelA* suffise à maintenir une certaine régulation (expériences en cours). Il reste également la possibilité que des boucles d'ADN se fassent à un niveau plus local : par exemple, la région de régulation de *pelA* bouclerait avec celle de *pelE* ou de *pelD*.

Nous pensons cependant que l'hypothèse d'interactions à longue portée par boucle d'ADN via des facteurs de transcription reste largement ouverte sur le plan expérimental. En effet, il semble que ces idées, développées depuis le début des années 2000, ne soient pas encore l'un des objets d'études majeurs de la communauté des microbiologistes expérimentateurs. La lecture du soucier<sup>1</sup> du laboratoire nous a montré que la très grande majorité des constructions qui ont été réalisées dans le laboratoire où nous avons travaillé (lequel nous semble tout à fait représentatif d'un laboratoire de biologie moléculaire) se répartit en deux types d'opérations : inacti-

---

<sup>1</sup>Un soucier est un document gardé au laboratoire et écrit par tous ses membres, qui regroupe tous les mutants qui ont été construits puis qui ont été mis en conservation pour des années. La modification génétique apportée au mutant y est décrite succinctement.



vation de gène par insertion d'un gène de résistance à un antibiotique et insertion de gène rapporteur dans un gène. Dans ces deux types de constructions, les régions de régulation des gènes sont laissées intactes. Les mutants du souche pour lesquels des régions de régulation ont été éliminées sont très rares (nous n'en avons trouvé qu'un seul exemple, il s'agissait d'un mutant qui avait subi une délétion de plusieurs gènes contigus et par conséquent, les régions de régulation entre les gènes avaient aussi été enlevées comme dans notre mutant Délétion). Il y a donc visiblement très peu de mutants où les biologistes auraient pu observer des phénotypes inattendus dus à l'élimination d'une région de régulation. La possibilité que des régions de régulation puissent agir sur d'autres gènes que ceux dont elles sont en amont nous semble donc une hypothèse toujours très ouverte sur le plan expérimental.

Nous avons proposé dans le dernier chapitre de cette thèse une méthode qui viserait à rechercher de façon large des interactions par boucle d'ADN. Plus généralement, l'approche tenterait de mettre en évidence quels seraient les segments d'ADN qui se retrouveraient proches à 3D de manière significative au sein d'un génome de bactérie. La méthode proposée nous paraît accessible pour un laboratoire de biologie moléculaire générique. En effet, elle ne fait appel qu'à des techniques qui sont utilisées de manière courante dans ce type de laboratoire. Elle est basée sur l'utilisation du système "enhancer" présent sur le promoteur du gène *glnA* d'*Escherichia coli* et s'inspire d'approches de biologie moléculaire comme la mutagenèse aléatoire ou la technique de l' $\alpha$ -complémentation. La méthode proposée implique que les segments d'ADN doivent se retrouver proches assez longtemps pour enclencher la machinerie transcriptionnelle et produire l'enzyme de la  $\beta$ -galactosidase, ce qui engendrera un clone de couleur bleue après la dégradation du X-Gal. Comme dans la méthode du double hybride bactérien, l'intensité de l'interaction pourrait être quantifiée en mesurant l'activité enzymatique de la  $\beta$ -galactosidase. Elle permettrait donc un filtrage des interactions qui auraient de plus grandes chances d'avoir du sens sur le plan de l'organisation 3D des génomes bactériens.

Dans les sections suivantes, nous discutons et ouvrons à d'autres hypothèses, les questions concernant les organisations spatiales de génomes bactériens abordées dans cette thèse.

### Un génome bactérien est-il organisé par les facteurs de transcription ?

Même si le système sur lequel nous avons réalisé les expériences reste particulier à la bactérie *Erwinia chrysanthemi*, celui-ci n'en demeure pas moins un système génétique aux caractéristiques très courantes dans la régulation chez les organismes procaryotes. En effet, il répond à un signal de l'environnement (présence ou non d'un inducteur) et il est régulé par les facteurs de transcription conventionnels (CRP, HNS...). De plus, par son rôle physiologique et par rapport aux autres arguments que nous avons développés dans le chapitre 4, ce système nous paraissait être parmi les gènes d'*Erwinia chrysanthemi*, l'un des systèmes les plus pertinents pour être régulé par une boucle intergénique.

Les expériences que nous avons réalisées sont peut-être un premier indice que les facteurs de transcription ne sont pas directement utilisés pour organiser l'ADN à l'échelle du génome. En effet, un argument qui irait dans ce sens serait de dire que la régulation transcriptionnelle est trop sujette à la stochasticité du fait du faible nombre de molécules, et trop sensible aux aléas de l'environnement pour que les

facteurs de transcription jouent également un rôle de “charpente” pour un génome bactérien. Des protéines plus spécifiques, dont le seul rôle serait d’organiser et de compacter le génome, pourraient plutôt tenir cette fonction.

Ce point rejoindrait le travail expérimental récent [108] où a été découverte une protéine MatP qui jouerait un rôle dans la compaction d’un des macrodomaines d’*Escherichia coli*. MatP n’est pas un facteur de transcription et sa fonction principale est d’organiser et de compacter le macrodomaine Ter. MatP se lie sur des sites de liaison qui sont au nombre de 23 copies dans le macrodomaine Ter (long de 800 kpb). Un des mécanismes proposés pour l’action de cette protéine est la formation de grandes boucles d’ADN entre les sites de liaison MatS. Il semble naturel d’extrapoler l’existence d’autres protéines qui pourraient jouer ce rôle, notamment pour organiser les autres macrodomaines d’*Escherichia coli*, sans pour autant être des protéines régulatrices.

Un autre argument serait de dire que les génomes de procaryotes, comme celui d’*Escherichia coli* sont déjà bien organisés à 1D grâce notamment aux opérons, pour assurer le bon fonctionnement des processus de régulation transcriptionnelle. De plus, ce sont des génomes “concis” dans le sens où presque la totalité de la séquence génomique est utile. Cette séquence contient dans sa grande majorité soit des séquences pour coder des protéines soit des régions régulatrices. L’organisation 1D pourrait donc suffire pour assurer les processus transcriptionnels efficacement et une organisation à 3D ne serait pas nécessaire pour ce point. Cependant, cet argument ne s’applique pas aux génomes eucaryotes, beaucoup plus compliqués et “brouillons” et qui ne sont pas dotés d’opérons. Ces génomes contiennent également de grandes quantités d’ADN non transcrites et dont on n’a pas trouvé d’utilité à ce jour (“ADN poubelle” ou “Junk DNA”). Il se pourrait donc que la nécessité d’organiser les molécules d’ADN à 3D à l’échelle génomique pour optimiser les processus transcriptionnels n’intervient que pour les organismes eucaryotes ou supérieurs.

### **Autres architectures génomiques possibles ?**

Nous avons présenté dans la section précédente des arguments qui soutiendraient la thèse qu’un génome bactérien ne serait pas organisé par les facteurs de transcription. Dans cette section, nous posons la question suivante : est-ce qu’un polymère d’ADN ne pourrait-il pas s’organiser sans l’intervention directe de protéines ? En d’autres termes, est-il possible que la longue molécule d’ADN puisse acquérir une structure 3D particulière seulement grâce à l’information de la séquence et aux propriétés physiques de l’ADN ? Nous savons que le contenu de la séquence peut avoir un effet sur la courbure de l’ADN. Peut-être que la séquence sur quelques endroits précis du génome dicte des courbures et une organisation à plus grande échelle. Il se pourrait alors que certaines séquences spécifiques aient un rôle précis dans l’organisation spatiale à l’échelle génomique. Comme mentionné dans le chapitre 5, il nous semble en effet important de garder dans les hypothèses de travail concernant les organisations spatiales de génomes, toutes les données récoltées concernant les séquences répétitives d’un génome comme celui d’*Escherichia coli* [12]. Il existe plusieurs familles de séquences qui se répètent, beaucoup d’entre elles sont palindromiques et leur longueur varie de la dizaine à la centaine de paires de base. Par exemple, les PU (Palyndromic Units) sont des séquences d’une quarantaine de paires de base. Elles se structurent en groupe pouvant compter jusqu’à une dizaine de PU.

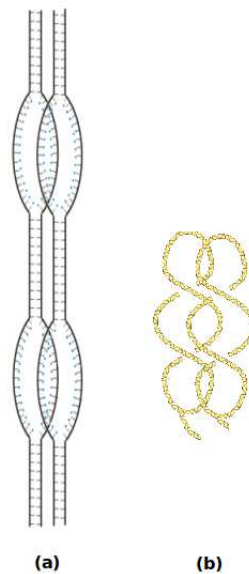


FIG. 5.5 – Mécanismes spéculatifs qui pourraient réunir des segments d’ADN sans l’intervention directe de protéines. (a) Des segments d’ADN pourraient échanger leurs brins sur des séquences identiques (qui pourraient par exemple être jouées par les séquences répétitives comme les séquences dites BIMEs). Les liaisons hydrogène entre les paires de bases de différents segments pourraient faire des attaches au sein d’un polymère d’ADN. (b) Les structures de surenroulement contrôlées notamment par l’enzyme gyrase pourraient être à l’origine d’enchevêtrements qui pourraient également faire des points d’attache entre parties d’une molécule d’ADN.

Ces clusters de PU sont appelés des BIMEs (pour Bacterial Interspersed Mosaic Element [12]).

Un premier point intéressant à noter est que par exemple certains BIMEs ont aussi des sites de liaison pour la protéine IHF [39] laquelle peut agir sur la conformation de l’ADN aidant notamment à courber l’ADN (cf. Chapitre 3).

Deuxièmement, ces séquences sont éparpillées un peu partout dans le génome d’*Escherichia coli* et en général aucune fonction globale ne leur a été attribuée (pour certains BIMEs, il a toutefois été montré un rôle dans l’atténuation de la transcription [40]). La nature de leur fonction semble toujours une question ouverte.

Nous pouvons alors poser la question si des segments d’ADN qui ont des séquences très proches ne pourraient pas s’hybrider en échangeant leurs simples brins et générer des points d’attache entre parties de génome proches à 3D (voir (a) de la Figure 5.5). L’énergie de liaison fournie par les liaisons hydrogènes entre les différents brins hybridés pourrait être suffisante pour faire des points d’attache au sein d’une molécule d’ADN. Par ailleurs, nous savons que la double hélice d’un segment d’ADN, même à température physiologique, peut s’ouvrir sur de petites zones (appelées “bulles d’ADN”). Il a récemment été montré sur le plan expérimental que la durée d’existence de ces bulles est de l’ordre de 20 à 100 microsecondes [7], ce qui en fait une durée assez longue pour les phénomènes de cette échelle. Peut-être que ces bulles d’ADN pourraient jouer un rôle dans la formation des liaisons et des hybridations entre segments d’ADN. Cette voie ne nous semble pas avoir été assez

abordée sur le plan théorique.

Nous pouvons avancer un second mécanisme spéculatif qui pourrait faire des points d'attache au sein d'un génome bactérien sans l'intervention directe de protéines qui lieraient des segments d'ADN entre eux. Il pourrait également faire intervenir des séquences répétitives présentes dans le génome d'*Escherichia coli*, comme par exemple les séquences dites REP (pour Repetitive Extragenic Palindromique).

L'article de Yang et Ames [182] donne des arguments très intéressants sur l'importance que pourrait avoir ces séquences REP dans l'organisation spatiale des génomes sans toutefois donner de mécanismes bien particuliers. Les auteurs montrent en effet dans ce travail que l'enzyme gyrase peut se lier sur ces séquences REP. Or la gyrase joue sur le surenroulement de l'ADN et peut faciliter le croisement de brins. La gyrase peut aussi se lier à deux sites séparés sur l'ADN [110] et donc rapprocher des brins. De plus, il a été montré que des mutants à la gyrase forment des nucléoides défectueux. Pourquoi ne pas imaginer une organisation génomique basée sur des structures formées par les surenroulements et des enchevêtrements de brins d'ADN qui attacheraient entre elles des parties d'ADN éloignées (voir (b) de la Figure 5.5). Enfin, ces mécanismes ne seraient pas incompatibles avec des mécanismes de formation de boucles d'ADN faisant intervenir des protéines mais ils pourraient être complémentaires à ceux-ci. Ces voies nous paraissent peu abordées ou discutées dans la littérature.

C'est en combinant ces informations, et en les confrontant à la réalité biologique grâce à de nouvelles méthodes expérimentales, que des éléments de réponse pourront être apportés aux différentes hypothèses concernant l'organisation spatiale des génomes bactériens.

# Bibliographie

- [1] S. Adhya. Multipartite genetic control elements : communication by DNA loop. *Annu Rev Genet*, 23 :227–250, 1989.
- [2] Laura Aguilera, Evangelina Campos, Rosa Giménez, Josefa Badía, Juan Aguilar, and Laura Baldoma. Dual role of LldR in regulation of the lldPRD operon, involved in L-lactate metabolism in *Escherichia coli*. *J Bacteriol*, 190(8) :2997–3005, Apr 2008.
- [3] S. Alberti, S. Oehler, B. von Wilcken-Bergmann, and B. Müller-Hill. Genetic analysis of the leucine heptad repeats of Lac repressor : evidence for a 4-helical bundle. *EMBO J*, 12(8) :3227–3236, Aug 1993.
- [4] Timothy E Allen, Nathan D Price, Andrew R Joyce, and Bernhard Ø Palsson. Long-range periodic patterns in microbial genomes indicate significant multi-scale chromosomal organization. *PLoS Comput Biol*, 2(1) :e2, Jan 2006.
- [5] Uri Alon. *An Introduction to Systems Biology*. Chapman & Hall/CRC, 2006.
- [6] Uri Alon. Network motifs : theory and experimental approaches. *Nat Rev Genet*, 8(6) :450–461, Jun 2007.
- [7] Grégoire Altan-Bonnet, Albert Libchaber, and Oleg Kravchenko. Bubble dynamics in double-stranded DNA. *Phys Rev Lett*, 90(13) :138101, Apr 2003.
- [8] M. Amouyal, L. Mortensen, H. Buc, and K. Hammer. Single and double loop formation when deoR repressor binds to its natural operator sites. *Cell*, 58(3) :545–551, Aug 1989.
- [9] Michèle Amouyal. Gene regulation at-a-distance in *E. coli* : new insights. *C R Biol*, 328(1) :1–9, Jan 2005.
- [10] Mariette R Atkinson, Michael A Savageau, Jesse T Myers, and Alexander J Ninfa. Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*. *Cell*, 113(5) :597–607, May 2003.
- [11] T. Atlung and H. Ingmer. H-NS : a modulator of environmentally regulated gene expression. *Mol Microbiol*, 24(1) :7–17, Apr 1997.
- [12] S. Bachellier, J. M. Clément, and M. Hofnung. Short palindromic repetitive DNA elements in enterobacteria : a survey. *Res Microbiol*, 150(9-10) :627–639, 1999.
- [13] Julien Baglio. Dynamique du biopolymère ADN en suivi de particule unique : simulations et expériences dans le cadre du mécanisme de transposition. Master’s thesis, ENS, 2008.
- [14] A. Bahloul, F. Boubrik, and J. Rouvière-Yaniv. Roles of *Escherichia coli* histone-like protein HU in DNA replication : HU-beta suppresses the thermosensitivity of dnaA46ts. *Biochimie*, 83(2) :219–229, Feb 2001.
- [15] N. Barkai and S. Leibler. Circadian clocks limited by noise. *Nature*, 403(6767) :267–268, Jan 2000.
- [16] A. Becskei and L. Serrano. Engineering stability in gene networks by autoregulation. *Nature*, 405(6786) :590–593, Jun 2000.
- [17] O. G. Berg, R. B. Winter, and P. H. von Hippel. Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry*, 20(24) :6929–6948, Nov 1981.
- [18] L. C. Bergstrom, L. Qin, S. L. Harlocker, L. A. Egger, and M. Inoué. Hierarchical and cooperative binding of OmpR to a fusion construct containing the ompC and ompF upstream regulatory sequences of *Escherichia coli*. *Genes Cells*, 3(12) :777–788, Dec 1998.
- [19] C. Blanco, P. Ritzenthaler, and M. Mata-Gilsinger. Negative dominant mutations of the uidR gene in *Escherichia coli* : genetic proof for a cooperative regulation of uidA expression. *Genetics*, 112(2) :173–182, Feb 1986.
- [20] Nicolas Blot, Catherine Berrier, Nicole Hugouvieux-Cotte-Pattat, Alexandre Ghazi, and Guy Condemine. The oligogalacturonate-specific porin KdgM of *Erwinia chrysanthemi* belongs to a new porin family. *J Biol Chem*, 277(10) :7936–7944, Mar 2002.
- [21] Nicolas E Buchler, Ulrich Gerland, and Terence Hwa. On schemes of combinatorial transcription logic. *Proc Natl Acad Sci U S A*, 100(9) :5136–5141, Apr 2003.
- [22] Thomas J Carew, Harold M Pinsker, and Eric R Kandel. Long-Term Habituation of a Defensive Withdrawal Reflex in *Aplysia*. *Science*, 175(4020) :451–454, Jan 1972.
- [23] Sacha De Carlo, Baoyu Chen, Timothy R Hoover, Elena Kondrashkina, Eva Nogales, and B. Tracy Nixon. The structural basis for regulated assembly and function of the transcriptional activator NtrC. *Genes Dev*, 20(11) :1485–1495, Jun 2006.
- [24] Anne-Sophie Carpentier, Bruno Torrèsani, Alex Grossmann, and Alain Hénaut. Decoding the nucleoid organisation of *Bacillus subtilis* and *Escherichia coli* through gene expression data. *BMC Genomics*, 6(1) :84, 2005.
- [25] A. Castillo, W. Nasser, G. Condemine, and S. Reverchon. The PecT repressor interacts with regulatory regions of pectate lyase genes in *Erwinia chrysanthemi*. *Biochim Biophys Acta*, 1442(2-3) :148–160, Nov 1998.
- [26] D. Charlier, M. Roovers, F. Van Vliet, A. Boyen, R. Cunin, Y. Nakamura, N. Glansdorff, and A. Piérard. Arginine regulon of *Escherichia coli* K-12. A study of repressor-operator interactions and of in vitro binding affinities versus in vivo repression. *J Mol Biol*, 226(2) :367–386, Jul 1992.
- [27] J. L. Cherry and F. R. Adler. How to make a biological switch. *J Theor Biol*, 203(2) :117–133, Mar 2000.

- [28] Andrea Ciliberto, Béla Novak, and John J Tyson. Steady states and oscillations in the p53/Mdm2 network. *Cell Cycle*, 4(3) :488–493, Mar 2005.
- [29] G. Condemine, C. Dorel, N. Hugouvieux-Cotte-Pattat, and J. Robert-Baudouy. Some of the out genes involved in the secretion of pectate lyases in *Erwinia chrysanthemi* are regulated by kdgR. *Mol Microbiol*, 6(21) :3199–3211, Nov 1992.
- [30] Peter R Cook. Predicting three-dimensional genome structure from transcriptional activity. *Nat Genet*, 32(3) :347–352, Nov 2002.
- [31] Frédéric Dardel and François Képès. *Bioinforma-tique Génomique et post-génomique*. 2002.
- [32] R. E. Dolmetsch, K. Xu, and R. S. Lewis. Calcium oscillations increase the efficiency and specificity of gene expression. *Nature*, 392(6679) :933–936, Apr 1998.
- [33] Jay C Dunlap and Jennifer J Loros. The neurospora circadian system. *J Biol Rhythms*, 19(5) :414–424, Oct 2004.
- [34] T. M. Dunn, S. Hahn, S. Ogdén, and R. F. Schleif. An operator at -280 base pairs that is required for repression of araBAD operon promoter : addition of DNA helical turns between the operator and promoter cyclically hinders repression. *Proc Natl Acad Sci U S A*, 81(16) :5017–5020, Aug 1984.
- [35] G. Dupont and A. Goldbeter. Protein phosphorylation driven by intracellular calcium oscillations : a kinetic analysis. *Biophys Chem*, 42(3) :257–270, Apr 1992.
- [36] Geneviève Dupont, Gérald Houart, and Paul De Koninck. Sensitivity of CaM kinase II to the frequency of Ca<sup>2+</sup> oscillations : a simple model. *Cell Calcium*, 34(6) :485–497, Dec 2003.
- [37] J. Dworkin, A. J. Ninfa, and P. Model. A protein-induced DNA bend increases the specificity of a prokaryotic enhancer-binding protein. *Genes Dev*, 12(6) :894–900, Mar 1998.
- [38] M.B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403 :335–338, 2000.
- [39] O. Espéli and F. Boccard. In vivo cleavage of *Escherichia coli* BIME-2 repeats by DNA gyrase : genetic characterization of the target and identification of the cut site. *Mol Microbiol*, 26(4) :767–777, Nov 1997.
- [40] O. Espéli, L. Moulin, and F. Boccard. Transcription attenuation associated with bacterial repetitive extragenic BIME elements. *J Mol Biol*, 314(3) :375–386, Nov 2001.
- [41] Paul François. *Réseaux génétiques conception, modélisation et dynamique*. PhD thesis, Ecole Normale Supérieure, 2005.
- [42] Paul François and Vincent Hakim. Design of genetic networks with specified functions by evolution in silico. *Proc Natl Acad Sci U S A*, 101(2) :580–585, Jan 2004.
- [43] T. S. Gardner, C. R. Cantor, and J. J. Collins. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 403(6767) :339–342, Jan 2000.
- [44] M. Geanakopoulou and S. Adhya. Functional characterization of roles of GalR and GalS as regulators of the gal regulon. *J Bacteriol*, 179(1) :228–234, Jan 1997.
- [45] Naama Geva-Zatorsky, Nitzan Rosenfeld, Shalev Itzkovitz, Ron Milo, Alex Sigal, Erez Dekel, Talia Yarnitzky, Yuvalal Liron, Paz Polak, Galit Lahav, and Uri Alon. Oscillations and variability in the p53 system. *Mol Syst Biol*, 2 :2006.0033, 2006.
- [46] D.T Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81 :2340–2361, 1977.
- [47] L. Glass and S.A Kauffman. The logical analysis of continuous, non-linear biochemical control networks. *J Theor Bio*, 39 :103–129, 1973.
- [48] A. Goldbeter, G. Dupont, and M. J. Berridge. Minimal model for signal-induced Ca<sup>2+</sup> oscillations and for their frequency encoding through protein phosphorylation. *Proc Natl Acad Sci U S A*, 87(4) :1461–1465, Feb 1990.
- [49] A. Goldbeter and D. E. Koshland. An amplified sensitivity arising from covalent modification in biological systems. *Proc Natl Acad Sci U S A*, 78(11) :6840–6844, Nov 1981.
- [50] Albert Goldbeter. *Biochemical Oscillations and Cellular Rhythms : The molecular bases of periodic and chaotic behaviour*. Cambridge University Press, 1996.
- [51] Carlos Gomez-Urbe, George C Verghese, and Leonid A Mirny. Operating Regimes of Signaling Cycles : Statics, Dynamics, and Noise Filtering. *PLoS Comput Biol*, 3(12) :e246, Dec 2007.
- [52] Didier Gonze, José Halloy, and Albert Goldbeter. Robustness of circadian rhythms with respect to molecular noise. *Proc Natl Acad Sci U S A*, 99(2) :673–678, Jan 2002.
- [53] B. C. Goodwin. Oscillatory behavior in enzymatic control processes. *Adv Enzyme Regul*, 3 :425–438, 1965.
- [54] Thomas J Goss, Brian K Janes, and Robert A Bender. Repression of glutamate dehydrogenase formation in *Klebsiella aerogenes* requires two binding sites for the nitrogen assimilation control protein, NAC. *J Bacteriol*, 184(24) :6966–6975, Dec 2002.
- [55] J.L Gouzé and T. Sari. A class of piecewise linear differential equations arising in biological models. *Dyn. Syst.*, 17 :299–316., 2002.
- [56] David C Grainger, Douglas Hurd, Marcus Harrison, Jolyon Holdstock, and Stephen J W Busby. Studies of the distribution of *Escherichia coli* cAMP-receptor protein and RNA polymerase along the *E. coli* chromosome. *Proc Natl Acad Sci U S A*, 102(49) :17693–17698, Dec 2005.
- [57] J. Griffith, A. Hochschild, and M. Ptashne. DNA loops induced by cooperative binding of lambda repressor. *Nature*, 322(6081) :750–752, 1986.
- [58] Raúl Guantes and Juan F Poyatos. Dynamical principles of two-component genetic oscillators. *PLoS Comput Biol*, 2(3) :e30, Mar 2006.
- [59] R. Haber and S. Adhya. Interaction of spatially separated protein-DNA complexes for control of gene expression : operator conversions. *Proc Natl Acad Sci U S A*, 85(24) :9683–9687, Dec 1988.
- [60] Andreas Hanke and Ralf Metzler. Entropy loss in long-distance DNA looping. *Biophys J*, 85(1) :167–173, Jul 2003.
- [61] B. He and H. Zalkin. Regulation of *Escherichia coli* purA by purine repressor, one component of a dual control mechanism. *J Bacteriol*, 176(4) :1009–1013, Feb 1994.
- [62] Reinhart Heinrich, Benjamin G Neel, and Tom A Rapoport. Mathematical models of protein kinase signal transduction. *Mol Cell*, 9(5) :957–970, May 2002.
- [63] Pascal Hersen, Megan N McClean, L. Mahadevan, and Sharad Ramanathan. Signal processing by the

- HOG MAP kinase pathway. *Proc Natl Acad Sci U S A*, 105(20) :7165–7170, May 2008.
- [64] D. S. Horowitz and J. C. Wang. Torsional rigidity of DNA and length dependence of the free energy of DNA supercoiling. *J Mol Biol*, 173(1) :75–91, Feb 1984.
- [65] K. J. Huang, J. L. Schieberl, and M. M. Igo. A distant upstream site involved in the negative regulation of the *Escherichia coli* ompF gene. *J Bacteriol*, 176(5) :1309–1315, Mar 1994.
- [66] N. Hugouvieux-Cotte-Pattat, G. Condemine, W. Nasser, and S. Reverchon. Regulation of pectinolysis in *Erwinia chrysanthemi*. *Annu Rev Microbiol*, 50 :213–257, 1996.
- [67] N. Hugouvieux-Cotte-Pattat, H. Dominguez, and J. Robert-Baudouy. Environmental conditions affect transcription of the pectinase genes of *Erwinia chrysanthemi* 3937. *J Bacteriol*, 174(23) :7807–7818, Dec 1992.
- [68] N. Hugouvieux-Cotte-Pattat, W. Nasser, and J. Robert-Baudouy. Molecular characterization of the *Erwinia chrysanthemi* kdgK gene involved in pectin degradation. *J Bacteriol*, 176(8) :2386–2392, Apr 1994.
- [69] Eugene M. Izhikevich. Bursting. *Scholarpedia*, 1(3) :1300, 2006.
- [70] F. JACOB and J. MONOD. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol*, 3 :318–356, Jun 1961.
- [71] Kyeong Soo Jeong, Jaeyong Ahn, and Arkady B Khodursky. Spatial patterns of transcriptional activity in the chromosome of *Escherichia coli*. *Genome Biol*, 5(11) :R86, 2004.
- [72] Christina Kahramanoglou, Christine L Webster, Mohamed Samir El-Robh, Tamara A Belyaeva, and Stephen J W Busby. Mutational analysis of the *Escherichia coli* melR gene suggests a two-state concerted model to explain transcriptional activation and repression in the melibiose operon. *J Bacteriol*, 188(9) :3199–3207, May 2006.
- [73] Eric R Kandel. The molecular biology of memory storage : a dialog between genes and synapses. *Biosci Rep*, 24(4-5) :475–522, 2004.
- [74] G. Karimova, J. Pidoux, A. Ullmann, and D. Laland. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proc Natl Acad Sci U S A*, 95(10) :5752–5756, May 1998.
- [75] T. B. Kepler and T. C. Elston. Stochasticity in transcriptional regulation : origins, consequences, and mathematical representations. *Biophys J*, 81(6) :3116–3136, Dec 2001.
- [76] Woaf P. Kepsu W.D. and Sepulchre J-A. Dynamics of the transition to pathogenicity in *Erwinia chrysanthemi*. to be published in Journal of Biological Systems.
- [77] B. N. Kholodenko. Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. *Eur J Biochem*, 267(6) :1583–1588, Mar 2000.
- [78] P. De Koninck and H. Schulman. Sensitivity of CaM kinase II to the frequency of Ca<sup>2+</sup> oscillations. *Science*, 279(5348) :227–230, Jan 1998.
- [79] M. G. Kornacker, B. Remsburg, and R. Menzel. Gene activation by the AraC protein can be inhibited by DNA looping between AraC and a LexA repressor that interacts with AraC : possible applications as a two-hybrid system. *Mol Microbiol*, 30(3) :615–624, Nov 1998.
- [80] Beat P Kramer and Martin Fussenegger. Hysteresis in a synthetic mammalian gene network. *Proc Natl Acad Sci U S A*, 102(27) :9517–9522, Jul 2005.
- [81] Evelyne Krin, Christine Laurent-Winter, Philippe N Bertin, Antoine Danchin, and Annie Kolb. Transcription regulation coupling of the divergent argG and metY promoters in *Escherichia coli* K-12. *J Bacteriol*, 185(10) :3139–3146, May 2003.
- [82] Edo Kussell and Stanislas Leibler. Phenotypic diversity, population growth, and information in fluctuating environments. *Science*, 309(5743) :2075–2078, Sep 2005.
- [83] François Képès. Periodic epi-organization of the yeast genome revealed by the distribution of promoter sites. *J Mol Biol*, 329(5) :859–865, Jun 2003.
- [84] François Képès. Periodic transcriptional organization of the *E.coli* genome. *J Mol Biol*, 340(5) :957–964, Jul 2004.
- [85] François Képès. [Epigenetics as an aspect of post-genomics]. *Med Sci (Paris)*, 21(4) :371–376, Apr 2005.
- [86] François Képès and Cédric Vaillant. Transcription-Based Solenoidal Model of Chromosomes. *Complexus*, 1-No 4 :171–180, 2003.
- [87] T. J. Larson, J. S. Cantwell, and A. T. van Loo-Bhattacharya. Interaction at a distance between multiple operators controls the adjacent, divergently transcribed glpTQ-glpACB operons of *Escherichia coli* K-12. *J Biol Chem*, 267(9) :6114–6121, Mar 1992.
- [88] Thomas LAUTIER. *Rôle de la protéine associée au nucléoïde Fis dans le contrôle de la virulence chez la bactérie phytopathogène Erwinia chrysanthemi*. PhD thesis, L’Institut National des Sciences Appliquées de Lyon, 2007.
- [89] Martin J Lercher and Laurence D Hurst. Co-expressed yeast genes cluster over a long range but are not regularly spaced. *J Mol Biol*, 359(3) :825–831, Jun 2006.
- [90] Julian Lewis. Autoinhibition with transcriptional delay : a simple mechanism for the zebrafish somitogenesis oscillator. *Curr Biol*, 13(16) :1398–1408, Aug 2003.
- [91] Difei Li and Chunguang Li. Noise-induced dynamics in the mixed-feedback-loop network motif. *Phys Rev E Stat Nonlin Soft Matter Phys*, 77(1 Pt 1) :011903, Jan 2008.
- [92] Gene-Wei Li, Otto G. Berg, and Johan Elf. Target location by DNA-Binding Proteins : Effects of Roadblocks and DNA looping. arXiv :0809.1063v1 [q-bio.QM], 2008.
- [93] Y. Li and A. Goldbeter. Frequency specificity in intercellular communication. Influence of patterns of periodic signaling on target cell responsiveness. *Biophys J*, 55(1) :125–145, Jan 1989.
- [94] Y. Li and A. Goldbeter. Frequency encoding of pulsatile signals of cAMP based on receptor desensitization in *Dictyostelium* cells. *J Theor Biol*, 146(3) :355–367, Oct 1990.
- [95] Giuseppe Lia, David Bensimon, Vincent Croquette, Jean-Francois Allemand, David Dunlap, Dale E A Lewis, Sankar Adhya, and Laura Finzi. Supercoiling and denaturation in Gal repressor/heat unstable nucleoid protein (HU)-mediated DNA looping. *Proc Natl Acad Sci U S A*, 100(20) :11373–11377, Sep 2003.

## BIBLIOGRAPHIE

---

- [96] Ovidiu Lipan and Wing H Wong. The use of oscillatory signals in the study of genetic networks. *Proc Natl Acad Sci U S A*, 102(20) :7063–7068, May 2005.
- [97] R. B. Lobell and R. F. Schleif. AraC-DNA looping : orientation and distance-dependent loop breaking by the cyclic AMP receptor protein. *J Mol Biol*, 218(1) :45–54, Mar 1991.
- [98] Jason W. Locasale. Signal duration and the time scale dependence of signal integration in biochemical pathways. arXiv :0802.2683v1 [q-bio.MN], 2008.
- [99] Y. L. Lyubchenko, L. S. Shlyakhtenko, T. Aki, and S. Adhya. Atomic force microscopic demonstration of DNA looping by GalR and HU. *Nucleic Acids Res*, 25(4) :873–876, Feb 1997.
- [100] M. C. Mackey and L. Glass. Oscillation and chaos in physiological control systems. *Science*, 197(4300) :287–289, Jul 1977.
- [101] N. Mandal, W. Su, R. Haber, S. Adhya, and H. Echols. DNA looping in cellular repression of transcription of the galactose operon. *Genes Dev*, 4(3) :410–418, Mar 1990.
- [102] S. Mangan, S. Itzkovitz, A. Zaslaver, and U. Alon. The incoherent feed-forward loop accelerates the response-time of the gal system of Escherichia coli. *J Mol Biol*, 356(5) :1073–1081, Mar 2006.
- [103] Davide Marenduzzo, Inês Faro-Trindade, and Peter R Cook. What are the molecular ties that maintain genomic loops? *Trends Genet*, 23(3) :126–133, Mar 2007.
- [104] Davide Marenduzzo, Cristian Micheletti, and Peter R Cook. Entropy-driven genome organization. *Biophys J*, 90(10) :3712–3721, May 2006.
- [105] Agustino Martínez-Antonio and Julio Collado-Vides. Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr Opin Microbiol*, 6(5) :482–489, Oct 2003.
- [106] K. S. Matthews. DNA looping. *Microbiol Rev*, 56(1) :123–136, Mar 1992.
- [107] Aurélien Mazurie, Samuel Bottani, and Massimo Vergassola. An evolutionary and functional assessment of regulatory network motifs. *Genome Biol*, 6(4) :R35, 2005.
- [108] Romain Mercier, Marie-Agnès Petit, Sophie Schbath, Stéphanie Robin, Meriem El Karoui, Frédéric Boccard, and Olivier Espéli. The MatP/matS site-specific system organizes the terminus region of the E. coli chromosome into a macrodomain. *Cell*, 135(3) :475–485, Oct 2008.
- [109] Jerome T Mettetal, Dale Muzzey, Carlos Gómez-Uribe, and Alexander van Oudenaarden. The frequency dependence of osmo-adaptation in *Saccharomyces cerevisiae*. *Science*, 319(5862) :482–484, Jan 2008.
- [110] C. L. Moore, L. Klevan, J. C. Wang, and J. D. Griffith. Gyrase . DNA complexes visualized as looped structures by electron microscopy. *J Biol Chem*, 258(7) :4612–4617, Apr 1983.
- [111] F. Moran, S. Nasuno, and M. P. Starr. Extracellular and intracellular polyglacturonic acid transeliminases of *Erwinia carotovora*. *Arch Biochem Biophys*, 123(2) :298–306, Feb 1968.
- [112] J.D Murray. *Mathematical Biology*. Springer, 1989.
- [113] B. Müller-Hill. The function of auxiliary operators. *Mol Microbiol*, 29(1) :13–18, Jul 1998.
- [114] Benno Müller-Hill. *The lac Operon : a short history of a genetic paradigm*. 1996.
- [115] Benno Müller-Hill. What is life? The paradigm of DNA and protein cooperation at high local concentrations. *Mol Microbiol*, 60(2) :253–255, Apr 2006.
- [116] K. Nakashima, K. Kanamaru, H. Aiba, and T. Mizuno. Signal transduction and osmoregulation in Escherichia coli. A novel type of mutation in the phosphorylation domain of the activator protein, OmpR, results in a defect in its phosphorylation-dependent DNA binding. *J Biol Chem*, 266(17) :10775–10780, Jun 1991.
- [117] William Nasser and Sylvie Reverchon. New insights into the regulatory mechanisms of the LuxR family of quorum sensing regulators. *Anal Bioanal Chem*, 387(2) :381–390, Jan 2007.
- [118] A. Novick and M. Weiner. ENZYME INDUCTION AS AN ALL-OR-NONE PHENOMENON. *Proc Natl Acad Sci U S A*, 43(7) :553–566, Jul 1957.
- [119] Béla Novák and John J Tyson. Design principles of biochemical oscillators. *Nat Rev Mol Cell Biol*, 9(12) :981–991, Dec 2008.
- [120] S. Oehler, M. Amouyal, P. Kolkhof, B. von Wilcken-Bergmann, and B. Müller-Hill. Quality and position of the three lac operators of E. coli define efficiency of repression. *EMBO J*, 13(14) :3348–3355, Jul 1994.
- [121] S. Oehler, E. R. Eismann, H. Krämer, and B. Müller-Hill. The three operators of the lac operon cooperate in repression. *EMBO J*, 9(4) :973–979, Apr 1990.
- [122] Rolf Ohlsson and Anita Göndör. The 4C technique : the 'Rosetta stone' for genome biology in 3D? *Curr Opin Cell Biol*, 19(3) :321–325, Jun 2007.
- [123] Ertugrul M Ozbudak, Mukund Thattai, Han N Lim, Boris I Shraiman, and Alexander Van Oudenaarden. Multistability in the lactose utilization network of Escherichia coli. *Nature*, 427(6976) :737–740, Feb 2004.
- [124] N. Peekhaus and T. Conway. Positive and negative transcriptional regulation of the Escherichia coli gluconate regulon gene gntT by GntR and the cyclic AMP (cAMP)-cAMP receptor protein complex. *J Bacteriol*, 180(7) :1777–1785, Apr 1998.
- [125] S. E. Phillips, I. Manfield, I. Parsons, B. E. Davidson, J. B. Rafferty, W. S. Somers, D. Margarita, G. N. Cohen, I. Saint-Girons, and P. G. Stockley. Cooperative tandem binding of met repressor of Escherichia coli. *Nature*, 341(6244) :711–715, Oct 1989.
- [126] J. Plumbridge. Co-ordinated regulation of amino sugar biosynthesis and degradation : the NagC repressor acts as both an activator and a repressor for the transcription of the glmUS operon and requires two separated NagC binding sites. *EMBO J*, 14(16) :3958–3965, Aug 1995.
- [127] J. Plumbridge and A. Kolb. DNA loop formation between Nag repressor molecules bound to its two operator sites is necessary for repression of the nag regulon of Escherichia coli in vivo. *Mol Microbiol*, 10(5) :973–981, Dec 1993.
- [128] J. Plumbridge and A. Kolb. DNA bending and expression of the divergent nagE-B operons. *Nucleic Acids Res*, 26(5) :1254–1260, Mar 1998.
- [129] Jacqueline Plumbridge and Olivier Pellegrini. Expression of the chitobiose operon of Escherichia coli



- is regulated by three transcription factors : NagC, ChbR and CAP. *Mol Microbiol*, 52(2) :437-449, Apr 2004.
- [130] A. Porco, N. Peekhaus, C. Bausch, S. Tong, T. Isturiz, and T. Conway. Molecular genetic characterization of the *Escherichia coli* gntT gene of GntI, the main system for gluconate metabolism. *J Bacteriol*, 179(5) :1584-1590, Mar 1997.
- [131] N. Pouget, C. Turlan, N. Destainville, L. Salomé, and M. Chandler. IS911 transpososome assembly as analysed by tethered particle motion. *Nucleic Acids Res*, 34(16) :4313-4323, 2006.
- [132] Birgit M Prüss, Christopher Besemann, Anne Denton, and Alan J Wolfe. A complex transcription network controls the early stages of biofilm development by *Escherichia coli*. *J Bacteriol*, 188(11) :3731-3739, Jun 2006.
- [133] Samir El Qaidi and Jacqueline Plumbridge. Switching control of expression of ptsG from the Mlc regulon to the NagC regulon. *J Bacteriol*, 190(13) :4677-4686, Jul 2008.
- [134] A. Rampersaud, S. L. Harlocker, and M. Inouye. The OmpR protein of *Escherichia coli* binds to sites in the ompF promoter region in a hierarchical manner determined by its degree of phosphorylation. *J Biol Chem*, 269(17) :12559-12566, Apr 1994.
- [135] William Keith Ray. *Characterization of AgaR and YihW, Members of the DeoR Family of Transcriptional Regulators, and GlpE, a Rhodanese Belonging to the GlpR Regulon, Also a Member of the DeoR Family*. PhD thesis, Faculty of the Virginia Polytechnic Institute and State, 1999.
- [136] L. J. Reitzer and B. Magasanik. Transcription of glnA in *E. coli* is stimulated by activator bound to sites far from the promoter. *Cell*, 45(6) :785-792, Jun 1986.
- [137] A. Resibois, M. Colet, M. Faelen, E. Schoonejans, and A. Toussaint. phiEC2, a new generalized transducing phage of *Erwinia chrysanthemi*. *Virology*, 137(1) :102-112, Aug 1984.
- [138] S. Reverchon, W. Nasser, and J. Robert-Baudouy. Characterization of kdgR, a gene of *Erwinia chrysanthemi* that regulates pectin degradation. *Mol Microbiol*, 5(9) :2203-2216, Sep 1991.
- [139] P. A. Rice, S. Yang, K. Mizuuchi, and H. A. Nash. Crystal structure of an IHF-DNA complex : a protein-induced DNA U-turn. *Cell*, 87(7) :1295-1306, Dec 1996.
- [140] Dmitry A Rodionov, Mikhail S Gelfand, and Nicole Hugouvieux-Cotte-Pattat. Comparative genomics of the KdgR regulon in *Erwinia chrysanthemi* 3937 and other gamma-proteobacteria. *Microbiology*, 150(Pt 11) :3571-3590, Nov 2004.
- [141] D. L. Roeder and A. Collmer. Marker-exchange mutagenesis of a pectate lyase isozyme gene in *Erwinia chrysanthemi*. *J Bacteriol*, 164(1) :51-56, Oct 1985.
- [142] Christopher J Rosario and Robert A Bender. Importance of tetramer formation by the nitrogen assimilation control protein for strong repression of glutamate dehydrogenase formation in *Klebsiella pneumoniae*. *J Bacteriol*, 187(24) :8291-8299, Dec 2005.
- [143] Nitzan Rosenfeld, Michael B Elowitz, and Uri Alon. Negative autoregulation speeds the response times of transcription networks. *J Mol Biol*, 323(5) :785-793, Nov 2002.
- [144] Nitzan Rosenfeld, Jonathan W Young, Uri Alon, Peter S Swain, and Michael B Elowitz. Gene regulation at the single-cell level. *Science*, 307(5717) :1962-1965, Mar 2005.
- [145] Leonor Saiz and Jose M G Vilar. DNA looping : the consequences and its control. *Curr Opin Struct Biol*, 16(3) :344-350, Jun 2006.
- [146] Shivanthi Samarasinghe, Mohamed Samir El-Robh, David C Grainger, Wenke Zhang, Panos Soultanas, and Stephen J W Busby. Autoregulation of the *Escherichia coli* melR promoter : repression involves four molecules of MelR. *Nucleic Acids Res*, 36(8) :2667-2676, May 2008.
- [147] Maniatis T Sambrook J, Fritsch E. F. *Molecular cloning : a laboratory manual*. Cold Spring Harbor Laboratory, 1989.
- [148] Régine Samson, Jean Bernard Legendre, Richard Christen, Marion Fischer-Le Saux, Wafa Achouak, and Louis Gardan. Transfer of *Pectobacterium chrysanthemi* (Burkholder et al. 1953) Brenner et al. 1973 and *Brenneria paradisiaca* to the genus *Dickeya* gen. nov. as *Dickeya chrysanthemi* comb. nov. and *Dickeya paradisiaca* comb. nov. and delineation of four novel species, *Dickeya dadantii* sp. nov., *Dickeya dianthicola* sp. nov., *Dickeya dieffenbachiae* sp. nov. and *Dickeya zeae* sp. nov. *Int J Syst Evol Microbiol*, 55(Pt 4) :1415-1427, Jul 2005.
- [149] M. A. Savageau. Comparison of classical and auto-genous systems of regulation in inducible operons. *Nature*, 252(5484) :546-549, Dec 1974.
- [150] D. A. Schafer, J. Gelles, M. P. Sheetz, and R. Landick. Transcription by single molecules of RNA polymerase observed by light microscopy. *Nature*, 352(6334) :444-448, Aug 1991.
- [151] Robert Schleif. DNA LOOPING. *Annu. Rev. Biochem.*, 61 :199-223, 1992.
- [152] S. C. Schultz, G. C. Shields, and T. A. Steitz. Crystal structure of a CAP-DNA complex : the DNA is bent by 90 degrees. *Science*, 253(5023) :1001-1007, Aug 1991.
- [153] Szabolcs Semsey, Mark Geanakopoulos, Dale E A Lewis, and Sankar Adhya. Operator-bound GalR dimers close DNA loops by direct interaction : tetramerization and inducer binding. *EMBO J*, 21(16) :4349-4356, Aug 2002.
- [154] Jacques-A. Sepulchre, Sylvie Reverchon, and William Nasser. Modeling the onset of virulence in a pectinolytic bacterium. *J Theor Biol*, 244(2) :239-257, Jan 2007.
- [155] Chimpanzee Sequencing and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, 437(7055) :69-87, Sep 2005.
- [156] G. S. Shadel and T. O. Baldwin. Identification of a distantly located regulatory element in the luxD gene required for negative autoregulation of the *Vibrio fischeri* luxR gene. *J Biol Chem*, 267(11) :7690-7695, Apr 1992.
- [157] S. Shin and C. Park. Modulation of flagellar expression in *Escherichia coli* by acetyl phosphate and the osmoregulator OmpR. *J Bacteriol*, 177(16) :4696-4702, Aug 1995.
- [158] P. Smolen, D. A. Baxter, and J. H. Byrne. Frequency selectivity, multistability, and oscillations emerge from models of genetic regulatory systems. *Am J Physiol*, 274(2 Pt 1) :C531-C542, Feb 1998.

- [159] Baljinder K Sohanpal, Sammia El-Labany, Maryam Lahooti, Jacqueline A Plumbridge, and Ian C Blomfield. Integrated regulatory responses of *fimB* to N-acetylneuraminic (sialic) acid and GlcNAc in *Escherichia coli* K-12. *Proc Natl Acad Sci U S A*, 101(46) :16322–16327, Nov 2004.
- [160] Hao Song, Paul Smolen, Evyatar Av-Ron, Douglas A Baxter, and John H Byrne. Dynamics of a minimal model of interlocked positive and negative feedback loops of transcriptional regulation by cAMP-response element binding proteins. *Biophys J*, 92(10) :3407–3424, May 2007.
- [161] Christophe Soulé. Graphic requirements for multistationarity. *Complexus*, 1 :123–133, 2003.
- [162] L. T. Stauffer and G. V. Stauffer. Spacing and orientation requirements of GcvA-binding sites 3 and 2 and the Lrp-binding region for *gcvT* :*lacZ* expression in *Escherichia coli*. *Microbiology*, 144 (Pt 5) :1417–1422, May 1998.
- [163] W. Su, S. Porter, S. Kustu, and H. Echols. DNA-looping and enhancer activity : association between DNA-bound NtrC activator and RNA polymerase at the bacterial *glnA* promoter. *Proc Natl Acad Sci U S A*, 87(14) :5504–5508, Jul 1990.
- [164] Sumedha and Martin Weigt. Agglomeration of DNA-looping proteins. arXiv :0801.1480v1 [q-bio.SC], 2008.
- [165] Kerren K Swinger and Phoebe A Rice. IHF and HU : flexible architects of bent DNA. *Curr Opin Struct Biol*, 14(1) :28–35, Feb 2004.
- [166] R. Thomas. [The biological role of positive and negative feedback loops]. *Bull Mem Acad R Med Belg*, 135(10) :664–673, 1980.
- [167] Kevin B Towles, John F Beausang, Hernan G Garcia, Rob Phillips, and Philip C Nelson. First-principles calculation of DNA looping in tethered particle experiments. *Phys Biol*, 6(2) :25001, 2009.
- [168] G. A. Tuskan, S. Difazio, S. Jansson, J. Bohlmann, I. Grigoriev, U. Hellsten, N. Putnam, S. Ralph, S. Rombauts, A. Salamov, J. Schein, L. Sterck, A. Aerts, R. R. Bhallerao, R. P. Bhallerao, D. Blautz, W. Boerjan, A. Brun, A. Brunner, V. Busov, M. Campbell, J. Carlson, M. Chalot, J. Chapman, G-L. Chen, D. Cooper, P. M. Coutinho, J. Couturier, S. Covert, Q. Cronk, R. Cunningham, J. Davis, S. Degroove, A. Déjardin, C. Depamphilis, J. Detter, B. Dirks, I. Dubchak, S. Duplessis, J. Ehrling, B. Ellis, K. Gendler, D. Goodstein, M. Gribskov, J. Grimwood, A. Groover, L. Gunter, B. Hamberger, B. Heinze, Y. Helariutta, B. Hensrissat, D. Holligan, R. Holt, W. Huang, N. Islam-Faridi, S. Jones, M. Jones-Rhoades, R. Jorgensen, C. Joshi, J. Kangasjärvi, J. Karlsson, C. Kelleher, R. Kirkpatrick, M. Kirst, A. Kohler, U. Kalluri, F. Larimer, J. Leebens-Mack, J-C. Leplé, P. Locascio, Y. Lou, S. Lucas, F. Martin, B. Montanini, C. Napoli, D. R. Nelson, C. Nelson, K. Nieminen, O. Nilsson, V. Pereda, G. Peter, R. Philippe, G. Pilate, A. Poliakov, J. Razumovskaya, P. Richardson, C. Rinaldi, K. Ritland, P. Rouzé, D. Ryabov, J. Schmutz, J. Schrader, B. Segerman, H. Shin, A. Siddiqui, F. Sterky, A. Terry, C-J. Tsai, E. Uberbacher, P. Unneberg, J. Vahala, K. Wall, S. Wessler, G. Yang, T. Yin, C. Douglas, M. Marra, G. Sandberg, Y. Van de Peer, and D. Rokhsar. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*, 313(5793) :1596–1604, Sep 2006.
- [169] John J Tyson, Attila Csikasz-Nagy, and Bela Novak. The dynamics of cell cycle regulation. *Bioessays*, 24(12) :1095–1109, Dec 2002.
- [170] Michèle Valens, Stéphanie Penaud, Michèle Rosignol, François Cornet, and Frédéric Boccard. Macrodome organization of the *Escherichia coli* chromosome. *EMBO J*, 23(21) :4330–4341, Oct 2004.
- [171] Alejandra C Ventura, Jacques-A. Sepulchre, and Sofia D Merajver. A hidden feedback in signaling cascades is revealed. *PLoS Comput Biol*, 4(3) :e1000041, Mar 2008.
- [172] José M G Vilar, Hao Yuan Kueh, Naama Barkai, and Stanislas Leibler. Mechanisms of noise-resistance in genetic oscillators. *Proc Natl Acad Sci U S A*, 99(9) :5988–5992, Apr 2002.
- [173] José M G Vilar and Stanislas Leibler. DNA looping and physical constraints on transcription regulation. *J Mol Biol*, 331(5) :981–989, Aug 2003.
- [174] Zhiwei Wang, Zhonghui Hou, Houwen Xin, and Zhizhou Zhang. Engineered internal noise stochastic resonator in gene network : a model study. *Biophys Chem*, 125(2-3) :281–285, Feb 2007.
- [175] A. Wedel, D. S. Weiss, D. Popham, P. Dröge, and S. Kustu. A bacterial enhancer functions to tether a transcriptional activator near a promoter. *Science*, 248(4954) :486–490, Apr 1990.
- [176] D. L. Weissenborn, N. Wittekindt, and T. J. Larson. Structure and regulation of the *glpFK* operon encoding glycerol diffusion facilitator and glycerol kinase of *Escherichia coli* K-12. *J Biol Chem*, 267(9) :6122–6131, Mar 1992.
- [177] Wikipedia. Microscope à force atomique.
- [178] Wen Xiong and James E Ferrell. A positive-feedback-based bistable 'memory module' that governs a cell fate decision. *Nature*, 426(6965) :460–465, Nov 2003.
- [179] Kaneyoshi Yamamoto and Akira Ishihama. Two different modes of transcription repression of the *Escherichia coli* acetate operon by IclR. *Mol Microbiol*, 47(1) :183–194, Jan 2003.
- [180] B. Yang, S. G. Gerhardt, and T. J. Larson. Action at a distance for *glp* repressor control of *glpTQ* transcription in *Escherichia coli* K-12. *Mol Microbiol*, 24(3) :511–521, May 1997.
- [181] B. Yang and T. J. Larson. Action at a distance for negative control of transcription of the *glpD* gene encoding sn-glycerol 3-phosphate dehydrogenase of *Escherichia coli* K-12. *J Bacteriol*, 178(24) :7090–7098, Dec 1996.
- [182] Y. Yang and G. F. Ames. DNA gyrase binds to the family of prokaryotic repetitive extragenic palindromic sequences. *Proc Natl Acad Sci U S A*, 85(23) :8850–8854, Dec 1988.
- [183] Yong Yao, Maria A Martinez-Yamout, Tobin J Dickerson, Andrew P Brogan, Peter E Wright, and H. Jane Dyson. Structure of the *Escherichia coli* quorum sensing protein SdiA : activation of the folding switch by acyl homoserine lactones. *J Mol Biol*, 355(2) :262–273, Jan 2006.
- [184] J. C. Yin, M. Del Vecchio, H. Zhou, and T. Tully. CREB as a memory modulator : induced expression of a dCREB2 activator isoform enhances long-term memory in *Drosophila*. *Cell*, 81(1) :107–115, Apr 1995.
- [185] Ji Yu, Jie Xiao, Xiaojia Ren, Kaiqin Lao, and X. Sunney Xie. Probing gene expression in live cells, one protein molecule at a time. *Science*, 311(5767) :1600–1603, Mar 2006.
- [186] Yongli Zhang, Abbye E McEwen, Donald M Crothers, and Stephen D Levene. Analysis of in-vivo

- LacR-mediated gene repression based on the mechanics of DNA looping. *PLoS One*, 1 :e136, 2006.
- [187] Chiara Zurla, Carlo Manzo, David Dunlap, Dale E A Lewis, Sankar Adhya, and Laura Finzi. Direct demonstration and quantification of long-range DNA looping by the lambda bacteriophage repressor. *Nucleic Acids Res*, 37(9) :2789–2795, May 2009.