



**HAL**  
open science

# Evolution et modélisation de processus biologiques : application à la régulation de la compétence naturelle pour la transformation génétique bactérienne chez les streptocoques

Mathias Weyder

► **To cite this version:**

Mathias Weyder. Evolution et modélisation de processus biologiques : application à la régulation de la compétence naturelle pour la transformation génétique bactérienne chez les streptocoques. Microbiologie et Parasitologie. Université Paul Sabatier - Toulouse III, 2017. Français. NNT : 2017TOU30021 . tel-01743844

**HAL Id: tel-01743844**

**<https://theses.hal.science/tel-01743844>**

Submitted on 26 Mar 2018

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

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

Université Fédérale



Toulouse Midi-Pyrénées

# THÈSE

En vue de l'obtention du

## DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

---

**Présentée et soutenue par :**

**Mathias WEYDER**

le 29 mars 2017

**Titre :**

Évolution et modélisation de processus biologiques :  
application à la régulation de la compétence naturelle pour la transformation  
génétique bactérienne chez les Streptocoques

---

**École doctorale et discipline ou spécialité :**

ED BSB : Microbiologie

**Unité de recherche :**

Laboratoire de Microbiologie et Génétique Moléculaires – UMR 5100

**Directeur/trice(s) de Thèse :**

Dr. Fabien JOURDAN

**Jury :**

Matthieu ARLAT – Professeur, Université Toulouse 3 – Président du Jury

Fabien JOURDAN – DR2, INRA – Directeur de Thèse

Marie BEURTON-AIMAR – Maître de Conférences, Université Bordeaux 2 – Rapporteur

Hidde DE JONG – DR, INRIA – Rapporteur

Vincent FROMION – DR2, INRA – Rapporteur

Christine GASPIN – DR, INRA – Examinatrice

## Remerciements

Je tiens à exprimer ici toute ma reconnaissance à Fabien Jourdan qui a accepté de suivre la fin de mes travaux de recherche et n'a jamais ménagé son temps pour me guider : ses observations fines et pertinentes m'ont été d'une aide particulièrement précieuse. Je ne saurais trop le remercier pour l'intérêt qu'il a porté à ma thèse, pour ses encouragements et son optimisme.

J'adresse bien sûr tous mes remerciements aux rapporteurs de ce travail, pour leur aimable participation à mon jury de soutenance de thèse.

Je remercie également Patrice Polard pour son écoute bienveillante et sa réactivité à gérer les questions administratives qui se sont posées, ainsi que les membres de son équipe et tout particulièrement Marc Prudhomme et Mathieu Berger pour avoir partagé toute leur expertise sur le pneumocoque.

J'associe à ces remerciements Philippe Valet, Directeur de l'École Doctorale de BSB : son intervention m'a permis de prolonger mon contrat de doctorant afin de soutenir cette thèse.

Mes remerciements aussi à l'ensemble de l'équipe de Gwennaëlle Fichant qui m'a accueilli durant trois ans.

Je dois enfin beaucoup à mes parents et à Soukaina : ils m'ont toujours soutenu avec confiance et dévouement. Qu'ils en soient chaleureusement remerciés !

# Évolution et modélisation de processus biologiques : application à la régulation de la compétence naturelle pour la transformation génétique

## Résumé

Afin de faire face à différents types de stress et s'adapter à de nouveaux environnements, les bactéries ont développé de nombreux mécanismes génétiquement régulés. La compétence pour la transformation naturelle est un processus qui favorise le transfert horizontal de gènes. Si les espèces phylogénétiquement éloignées partagent des mécanismes conservés d'intégration et de remaniement de l'ADN, les circuits de régulation de la compétence ne sont toutefois pas universels mais adaptés au mode de vie de chaque espèce. Chez les bactéries Gram-positives, les cascades de régulation de *Streptococcus pneumoniae* et *Bacillus subtilis* sont les mieux documentées. Si de nombreux modèles mathématiques ont été établis pour étudier différents aspects de la régulation des compétences chez *B. subtilis*, un seul modèle à échelle de population a été développé pour *S. pneumoniae*, il y a plus de dix ans, sur la base d'hypothèses contestées par de nouvelles données expérimentales.

Nous avons développé, chez *S. pneumoniae*, un modèle fondé sur la connaissance de la régulation de la compétence qui intègre les éléments biologiques essentiels connus à ce jour. La cohérence structurelle de la topologie du réseau est confirmée par le formalisme des réseaux de Petri. Le réseau est ensuite transformé en un ensemble d'équations différentielles ordinaires pour étudier son comportement dynamique. La cinétique des protéines a été estimée en utilisant des données de luminescence et l'estimation des paramètres a été contrainte à partir des connaissances disponibles. Après avoir testé des modèles alternatifs, nous avons proposé l'existence d'un produit de gène tardif supplémentaire pouvant inhiber l'action de ComW, l'activateur du facteur  $\sigma^X$ . Nous apportons également un nouvel éclairage sur cette cascade de régulation en prédisant la cinétique de composantes du système qui pourraient être impliquées dans des comportements spécifiques.

Ce modèle consolide les connaissances expérimentales acquises sur la régulation de la compétence chez *S. pneumoniae*. De plus, il peut être appliqué aux autres espèces de streptocoques appartenant aux groupes *mitis* et *anginosus* puisqu'ils partagent le même circuit régulateur. À l'échelle populationnelle, la transition vers l'état de compétence se produit d'abord dans une sous-population de cellules et se propage ensuite dans toute la population par contact physique cellule à cellule. En permettant la simulation du comportement d'une cellule individuelle, le modèle pourra servir de module dans la conception d'un modèle d'une population bactérienne composée de cellules hétérogènes.

## Table des matières

I.	Introduction .....	5
1.	La biologie des systèmes .....	5
2.	État de l'art sur les différents formalismes de modélisation .....	12
a.	Les réseaux booléens.....	14
b.	Les réseaux booléens probabilistes .....	17
c.	Les réseaux de Petri.....	18
d.	Les équations différentielles ordinaires.....	22
e.	Les modèles stochastiques.....	27
f.	Limites des approches quantitatives et qualitatives.....	28
g.	Les modèles hybrides .....	29
3.	Description du modèle biologique : la régulation de la transformation génétique.....	30
a.	Le transformasome .....	31
b.	Description de la régulation .....	33
c.	Chez <i>Bacillus subtilis</i> .....	33
d.	Chez les streptocoques .....	34
4.	État de l'art des modèles développés de régulation de la compétence .....	38
a.	Chez <i>Bacillus subtilis</i> .....	38
b.	Chez les streptocoques .....	39
II.	Construction et analyse statique du réseau de régulation ComCDE .....	43
1.	Simplifications réalisées pour concevoir le modèle .....	44
2.	Initiation de l'état de compétence : le système à deux composantes ComD/ComE.....	44
3.	Le régulateur central ComX .....	47
4.	L'arrêt de l'état de compétence .....	47
5.	Construction du réseau de Petri.....	48
6.	Analyse des invariants structuraux.....	50
III.	Analyse dynamique du réseau de régulation ComCDE .....	55
1.	Construction du système d'équations différentielles ordinaires.....	55
2.	Utilisation des données de fusion transcriptionnelle .....	57
3.	Traitement des données de luminescence.....	60
4.	Évidences expérimentales aidant à l'estimation des paramètres .....	64
5.	Optimisation des paramètres du modèle.....	65
6.	Analyse de la dynamique du système.....	67
7.	Comparaison des modèles alternatifs .....	68
8.	Étude de l'effet du pH du milieu sur le développement de la compétence .....	84
9.	Effet du niveau basal d'expression de <i>comCDE</i> .....	87

10.	La cinétique de ComE~P et analyse de l'espace des phases .....	89
11.	Étude de la période de non réponse au CSP .....	91
IV.	Discussion .....	94
V.	Conclusion et perspectives .....	100
VI.	Liste des abréviations .....	103
VII.	Bibliographie .....	104

## I. Introduction

### 1. La biologie des systèmes

Les systèmes biologiques comme les organismes, les cellules ou les biomolécules sont des systèmes organisés au travers de leurs structures et leurs fonctions. Ces systèmes ont été façonnés par l'évolution. Le dogme central de la biologie moléculaire peut être résumé de la façon suivante : les gènes codent pour des ARN messagers, les ARN messagers servent de modèles de synthèse pour les protéines et les protéines effectuent le travail de maintien et de développement cellulaire. Bien que l'information soit stockée dans les gènes sous la forme de séquence ADN, l'expression est rendue possible au travers de la machinerie cellulaire qui décode cette séquence, la traduit, donnant ainsi naissance à une structure et à une fonction.

La description des entités biologiques et de leurs propriétés peut se penser sur plusieurs niveaux d'organisation et sur différentes échelles de temps : nous pouvons étudier des phénomènes biologiques à l'échelle de populations, d'individus, de tissus, d'organes, de cellules, de compartiments et jusqu'à l'échelle moléculaire. Ces systèmes, ou entités, peuvent avoir des dimensions allant de l'ordre du mètre (pour l'homme) au micromètre pour de nombreux types cellulaires. Les échelles de temps, au cours desquelles ces systèmes évoluent, vont de millions d'années de processus évolutifs à la seconde pour de nombreuses réactions biochimiques. De plus, chaque action dans la cellule implique différents niveaux d'organisation, incluant les voies de signalisation, les réseaux de régulation ainsi que le métabolisme de la cellule. Ces différents niveaux organisationnels interagissent entre eux et chaque réseau individuel doit ainsi être pensé, sinon intégré, dans un cadre plus large.

En étudiant ces réseaux biologiques, nous pouvons être confrontés à des processus complexes qui ne peuvent pas être *a priori* expliqués, et dont l'issue ne peut être prédite par la seule intuition. En effet, des expériences permettent de générer des hypothèses sur le fonctionnement de processus biologiques, mais il demeure difficile de déterminer si ces hypothèses peuvent être combinées car il est souvent délicat de prévoir le comportement global d'un système complexe à partir de la seule connaissance de ses parties. Des modèles mathématiques et des simulations numériques peuvent alors aider à appréhender la dynamique de processus biologiques dans leurs globalités, à prédire leurs comportements et à comprendre l'effet de leurs interactions avec l'environnement.

Cette approche holistique contraste avec l'approche plus réductionniste de la biologie moléculaire. Le basculement vers une approche systémique s'effectue de façon progressive depuis la fin du 20<sup>e</sup> siècle, notamment avec l'émergence de nouvelles techniques de séquençage ayant permis une observation, à l'échelle d'un système, de l'organisation des

réseaux cellulaires. Ces observations ont révélé la grande complexité de ces réseaux qu'il était difficile d'appréhender en étudiant individuellement leurs composants. L'approche traditionnelle en biologie moléculaire consiste à scinder un système en plusieurs parties, d'en analyser chaque élément pour ensuite réassembler les parties en un système entier. En complément, l'intention de la biologie des systèmes consiste à modéliser et à comprendre le système entier en se fondant sur le paradigme suivant : le système entier contient plus d'informations que la somme de ses parties et ne peut être compris que comme un tout (Kitano, 2002).

Au sens large, un modèle est une représentation mathématique d'objets ou de processus qui explique les caractéristiques de ces objets ou de ces processus. Il existe cependant plusieurs paradigmes de modélisation : un ensemble de réactions biochimiques peut être représenté par un graphe avec des sommets correspondant aux métabolites et où les arrêtes spécifient les réactions entre métabolites ; ou bien, ce même réseau peut aussi être décrit par un système d'équations différentielles qui permet de prédire sa dynamique au cours du temps. Au-delà des concepts physiques généraux, un modèle doit également tenir compte des spécificités du système d'intérêt qui peuvent être justifiées par des travaux expérimentaux, par les connaissances biochimiques acquises sur plusieurs acteurs ainsi que par l'extrapolation de propriétés de systèmes proches dans l'évolution : un modèle doit être capable de résumer les connaissances acquises sur un système dans un cadre mathématique cohérent.

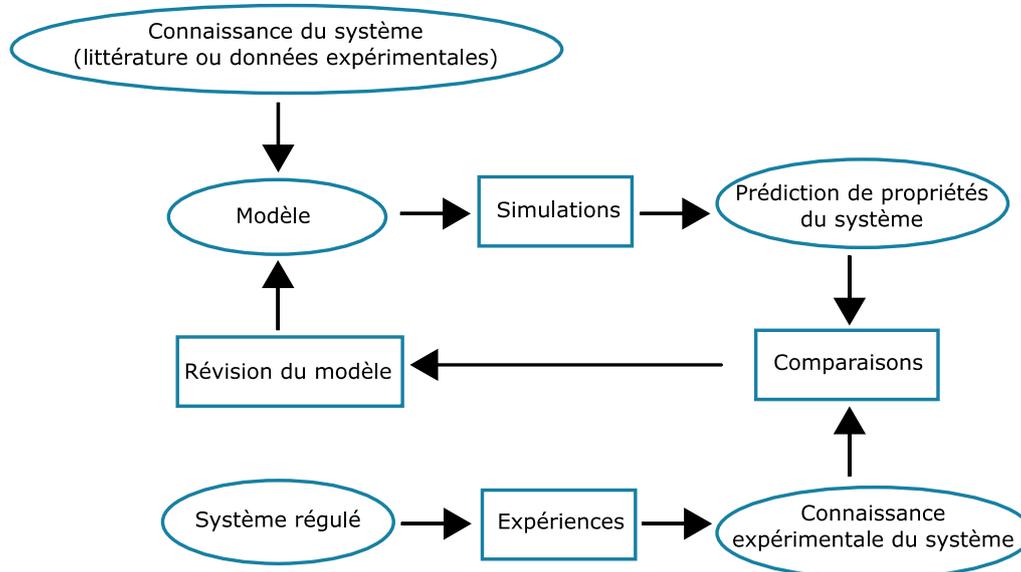
Aussi, un modèle peut se définir comme la description d'un système en utilisant un langage et des concepts mathématiques. Les modèles sont toujours composés d'un ensemble de variables, de liens établissant des relations entre les variables, et d'un ensemble de contraintes (Le Novère, 2015). Le modèle mathématique d'un système sera constitué de ces trois structures, quel que soit le formalisme de modélisation utilisé. Premièrement, les variables représentent des constituants physiques (comme les molécules) ou des paramètres (comme les vitesses de synthèse et de dégradation, les constantes d'association) qui peuvent varier pendant la simulation. Ces variables correspondent aux entités biologiques, entités connues dont nous souhaitons prédire la cinétique. Notons aussi qu'elles sont utilisées pour comparer les prédictions du modèle et les résultats expérimentaux. Deuxièmement, les liens mathématiques unissent les variables et représentent ce qui est déjà connu ou ce que nous souhaitons tester au travers du modèle. Ces relations mathématiques peuvent prendre plusieurs formes. Sans être exhaustif, il peut s'agir d'un changement de vitesse de synthèse d'une variable en fonction de la valeur d'une autre variable, d'une règle logique dans le cadre des modèles booléens ou l'assignation d'une valeur à une variable en fonction de la valeur d'une autre variable pour les formalismes basés sur des règles d'assignement. Troisièmement,

les contraintes représentent le contexte de l'analyse mais aussi les processus ignorés lors de celle-ci. Parmi les contraintes, nous pouvons citer les conditions initiales du système (les valeurs prises par l'ensemble des variables au début de la simulation), les quantités qui ne peuvent pas être négatives ou, par exemple, ne pas être supérieures à un certain seuil pour certaines variables.

Même des processus cellulaires simples impliquent de nombreux composants moléculaires et présentent des comportements non linéaires qui peuvent, de surcroît, interagir de manière complexe. Des modèles mathématiques d'interactions pertinentes à l'intérieur d'un système peuvent alors aider à décrire la structure du système de façon non ambiguë, à fournir une compréhension de sa fonction et à prédire le bon (ou le mauvais) comportement du réseau construit en regard des résultats expérimentaux. Grâce, entre autres, aux modèles mathématiques de réseaux de régulation, la biologie des systèmes est donc devenue une part importante des méthodologies utilisées en recherche en biologie. En complément du travail expérimental, les simulations numériques ont comme objectif de comprendre les bases moléculaires des fonctions cellulaires et de prédire les propriétés émergentes de systèmes complexes. Par exemple, les travaux sur la chimiotaxie bactérienne (Bray *et al.*, 1993, Tindall *et al.*, 2012) fournissent un exemple de recherche collaborative entre approche expérimentale et modélisation. Dans ce travail pionnier de modélisation du chimiotactisme bactérien, Bray et collaborateurs ont regroupé dans un réseau de réactions biochimiques l'ensemble des mécanismes connus jusqu'alors. Le réseau simulé reproduit précisément le phénotype de plus de 30 mutants, parmi lesquels des composés sont supprimés et/ou surexprimés, fournissant des cinétiques de réponses similaires à celles observées *in vivo*. Par ailleurs, notons que les modèles développés ces dernières années ont augmenté en taille comme le montre la reconstruction récente du métabolisme humain (Thiele *et al.*, 2013). De plus, la modélisation complète d'un microorganisme (Karr *et al.*, 2012) réalisée chez *Mycoplasma genitalium* ou encore les modèles multi-échelles d'organismes chez *Arabidopsis thaliana* (Chew *et al.*, 2014) prouvent qu'il est maintenant possible de modéliser, généralement sous forme modulaire, des réseaux de régulation avec des niveaux de précision et de complexité de plus en plus élevés.

Les modèles mathématiques fournissent donc un moyen de tester, *in silico*, des hypothèses biologiques (Nelander *et al.*, 2008, Ogilvie *et al.*, 2015). Ils permettent également de résumer de façon cohérente l'ensemble des résultats publiés et d'assembler l'information disséminée dans un cadre mathématique cohérent en fonction des données disponibles et des questions auxquelles l'on souhaite répondre. La modélisation permet aussi de comparer le modèle généré avec des données publiées disponibles ou de le mettre en regard de nouveaux résultats

expérimentaux (mutants ou traitement médicamenteux par exemple (Flobak *et al.*, 2015). Après plusieurs cycles de raffinement du modèle, ce dernier est particulièrement utile lorsqu'il permet de confirmer (ou d'infirmer) des hypothèses par rapport à l'intuition initiale, ou bien lorsqu'il permet de suggérer de nouvelles expériences (voir **Figure 1** pour une vue schématique du processus de reconstruction et d'analyse de modèle).



**Figure 1. Processus de construction et d'analyse d'un réseau de régulation** (adapté de de Jong, 2002). En partant d'un modèle initial construit à partir des connaissances acquises sur le système ou directement à partir de données expérimentales, le comportement du système est testé en simulant plusieurs conditions expérimentales. La comparaison des prédictions avec les profils expérimentaux d'expression donne une idée de l'adéquation du modèle aux données. Si les comportements prédits et observés ne sont pas congruents, le modèle doit être repensé. Les étapes de construction du modèle, de révision de celui-ci, de simulation du comportement du système et de comparaison du modèle aux données sont réitérées jusqu'à ce qu'un modèle acceptable soit obtenu.

Les comportements complexes des systèmes biologiques ne peuvent souvent pas être réduits aux propriétés élémentaires de chacun de leurs composants. Ainsi, une analyse mathématique d'un processus régulé à l'intérieur de la cellule aide à la compréhension de la dynamique de ce processus. Des modèles mathématiques explicatifs et prédictifs sont utiles pour comprendre un système dans sa globalité. Ils peuvent en effet mettre en évidence des événements cellulaires anormaux et apporter de l'information sur la perturbation des voies de signalisation comme celles impliquées dans le développement de cancers (Laubenbacher *et al.*, 2009). La modélisation de ces voies de signalisation peut améliorer la compréhension du développement de maladies et mieux discriminer les processus engagés dans leur progression. Ces analyses aident alors à suggérer de nouvelles approches thérapeutiques, comme par

exemple en pharmacocinétique et en pharmacodynamique, lors du développement de médicaments (Li *et al.*, 2016).

Un modèle ne reprend que certains aspects de la réalité mais cela peut être suffisant si l'intention du modèle est de répondre à des questions spécifiques. Par exemple, si le seul objectif d'un modèle est de prédire la sortie du système en fonction d'un signal d'entrée, ce modèle ne présentera qu'une relation entrée-sortie simplifiée et l'intérieur du modèle peut être considéré comme une boîte noire. En revanche, si un mécanisme biologique doit être élucidé précisément, alors la structure d'un système et les relations entre les parties doivent être décrites de façon réaliste. Mais le modèle mathématique doit rester aussi simple que possible pour permettre, non seulement une implémentation simple, mais aussi pour obtenir des résultats simples à interpréter. Les caractéristiques mentionnées ci-dessus ne définissent pas un modèle comme étant juste ou faux, mais elles déterminent s'il est approprié pour résoudre un problème. La phrase « Essentiellement tous les modèles sont faux mais certains sont utiles » formulée par le statisticien Georges Box est de ce fait une ligne de conduite appropriée pour construire un modèle. Le développement et l'analyse des modèles requièrent donc d'être adaptés en fonction du problème à résoudre, de la quantité mais aussi de la nature des données expérimentales disponibles.

Les modèles mathématiques sont généralement élaborés et utilisés de façon itérative (**Figure 1**). La première étape consiste à construire la structure du modèle en déterminant, à partir de la littérature ou de résultats expérimentaux, les entités biologiques pertinentes à inclure dans le modèle. Le nombre d'entités dépend à la fois de la question biologique posée et de la quantité de données disponibles pour paramétrer et valider le modèle. En effet, un modèle n'a pas nécessairement besoin d'inclure tous les composants connus d'un système, et les processus biologiques peuvent être décrits à différents niveaux de granularité. Choisir entre la possibilité d'une approximation d'un processus et l'exhaustivité de la description de ce processus est une étape clé lors de la construction d'un modèle mathématique. La seconde étape est la recherche de la nature des interactions entre composants et le choix des expressions mathématiques permettant de caractériser les relations entre ceux-ci. Cela peut aussi être réalisé à partir de la littérature scientifique ou par inférence à partir de résultats d'expériences de génomique fonctionnelle. Enfin, il est nécessaire de déterminer les valeurs des paramètres et les conditions initiales du système dans le cas des modèles quantitatifs, d'effectuer les simulations et les analyses mathématiques permettant de reproduire les observations ou menant à des prédictions. Dans chaque cycle, le modèle peut être incrémenté afin d'inclure de nouvelles variables, mais aussi être élagué si la complexité devient trop importante : considérer un nombre trop important d'espèces moléculaires peut en effet mener

à des difficultés lors du choix des expressions mathématiques ou lors de l'estimation des paramètres.

Il existe deux approches principales dans le processus de construction d'un modèle (Le Novère *et al.*, 2015). La première, appelée *bottom-up*, est basée sur les connaissances *a priori* : les relations entre entités sont obtenues à partir de la littérature scientifique, de bases de données publiques ou même à partir de modèles antérieurs. De nombreux modèles mathématiques de phénomènes cellulaires ont déjà été publiés dans la littérature scientifique selon cette approche. Ces modèles peuvent être entreposés dans des dépôts de modèles tels que les bases de données BioModels (Le Chelliah *et al.*, 2013), CellML (Cuellar *et al.*, 2003) ou JWS (Olivier *et al.*, 2004). Ces bases de données fournissent alors une collection de modèles pouvant servir de points de départ pour incrémenter un modèle antérieur ou composer un module dans la construction de modèles plus larges. Par exemple, le modèle de la voie de signalisation du facteur EGFR (Epidermal Growth Factor Receptor) lors de la progression tumorale (Bidkhorji. *et al.*, 2012) a été construit en utilisant des modèles précédents de voies de transduction de signal activées par le facteur EGF : la voie PI3K (phosphoinositide 3-kinase) (Kiyatkn *et al.*, 2006), la voie MAPK (mitogen-activated protein kinase) (Huang et Ferrell, 1996) et la voie de signalisation JAK/STAT (Yamada *et al.*, 2003). Outre les bases de données contenant des modèles, il existe de nombreuses bases de données d'interactions protéiques issues d'expériences d'immuno-précipitation, pouvant alors servir d'ébauches dans l'élaboration de la topologie du modèle. Nous pouvons citer les bases de données IntAct (Orchard *et al.*, 2014) ou STRING (Franceschini *et al.*, 2013) qui regroupent des données de plusieurs natures comme celles issues d'expériences d'interactions protéines-protéines ou de corrélations d'expressions de gènes. Cependant, ces bases de données ne contiennent que des informations de nature qualitative entre plusieurs entités et ne peuvent donc pas être directement utilisées dans le cadre d'un modèle dynamique. Les bases de données référençant des voies métaboliques comme KEGG pathway (Kanehisa *et al.*, 2014) et MetaCyc (Caspi *et al.*, 2016) fournissent des graphs orientés reliant les métabolites au travers de réactions. Les valeurs des paramètres des réactions biochimiques sous-jacentes à un réseau peuvent être retrouvées à partir de la littérature scientifique, mais aussi à partir de bases de données dédiées comme SABIO-RK (Witting *et al.*, 2012) ou BRENDA (Schomburg *et al.*, 2013). Aussi, au-delà de l'analyse du modèle, l'inférence des paramètres d'un réseau de régulation de gènes à partir des données expérimentales est devenue une problématique centrale en biologie computationnelle. Comme le nombre de paramètres dépasse souvent le nombre de mesures disponibles, les modélisateurs utilisent généralement des données issues de différentes expériences pour paramétrer des modèles et peuvent aussi inclure des

connaissances biologiques supplémentaires lors de l'estimation des paramètres (Bachmann *et al.*, 2011). Dans ce travail, les auteurs utilisent simultanément des données de RT-PCR quantitative, de spectrométrie de masse et des données de type dose-réponse afin de paramétrer leur modèle

En regard de l'approche *bottom-up* utilisant les interactions moléculaires connues, nous pouvons rapidement mentionner les approches *top-down* qui consistent à inférer la topologie d'un réseau biologique, inconnue, seulement à partir de données expérimentales, et souvent à partir d'expériences de génomique fonctionnelle. De nombreuses méthodes implémentées existent pour traiter ce problème d'inférence de réseau (Gardner *et al.*, 2003, Margolin *et al.*, 2006, Bonneau *et al.*, 2006, Wang *et al.*, 2007, Gitter *et al.*, 2011). Nous pouvons, sans être exhaustif, citer par exemple les méthodes d'inférence bayésienne (Yu *et al.*, 2004) qui permettent de calculer la probabilité qu'un ensemble de données soit produit par différentes topologies de réseau ou les méthodes statistiques qui d'établissent des corrélations entre variables (Langfelder *et al.*, 2008) et permettent alors de déterminer si les variables sont indépendantes ou non à partir du concept suivant : si deux gènes sont systématiquement co-régulés dans plusieurs conditions, il est probable qu'ils partagent des mécanismes de régulation communs.

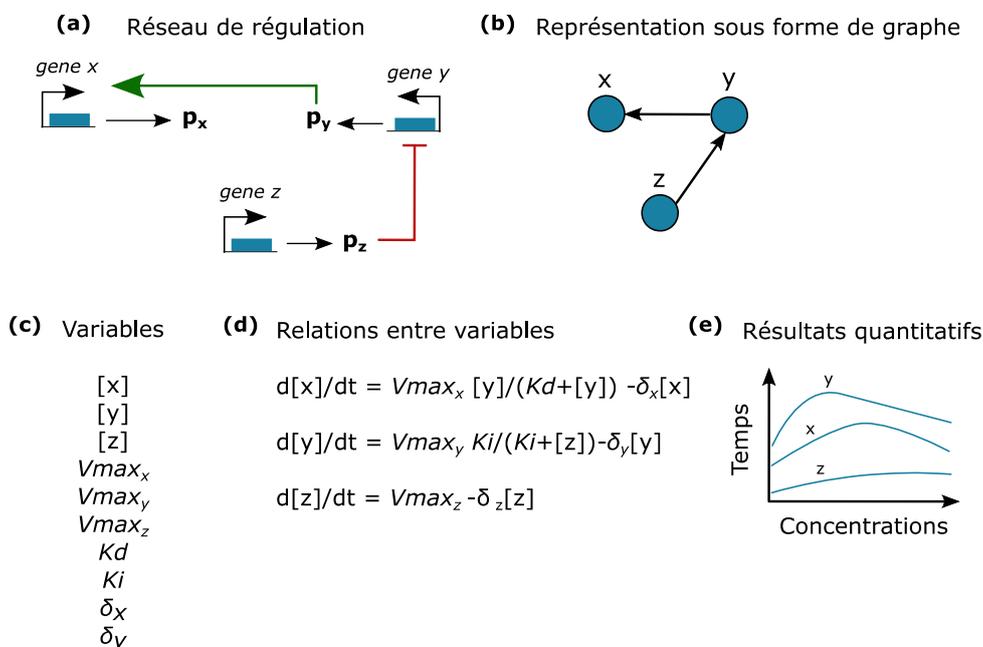
Les réseaux de régulation de gènes déterminent la manière dont les cellules contrôlent l'expression de leurs gènes, qui, outre les régulations post-transcriptionnelles et post-traductionnelles, déterminent le niveau de production des protéines essentielles au maintien des fonctions cellulaires. Les protéines synthétisées à partir des gènes peuvent ainsi fonctionner comme facteurs de transcription en liant les sites régulateurs d'autres gènes, comme catalyseurs de réactions métaboliques ou comme intermédiaires de voie de signalisation.

Les réseaux de régulation de gènes contrôlent la cellule à l'échelle génomique, en déterminant les gènes et la force avec laquelle ils sont transcrits en ARN qui servent eux-mêmes de modèles pour la synthèse protéique. Protéines et gènes ne fonctionnent pas de façon indépendante mais interagissent ensemble pour composer un réseau de régulation complexe : les protéines qui fonctionnent comme facteurs de transcription peuvent influencer positivement ou négativement l'expression d'autres gènes en liant leurs sites régulateurs, et ainsi influencer la production de protéines. Comme une majorité de réseaux de régulation de gènes implique de nombreux composés, interconnectés par des boucles de rétrocontrôle positives et négatives, dont la compréhension intuitive est difficile à appréhender, plusieurs

formalismes mathématiques ont été employés en biologie, pour mieux comprendre les systèmes régulés génétiquement (de Jong, 2002).

## 2. État de l'art sur les différents formalismes de modélisation

À un niveau abstrait, les composants d'un réseau peuvent être réduits à un ensemble de nœuds. Ces nœuds sont connectés les uns aux autres par des arêtes, chaque arête représentant une interaction entre deux composants. Les nœuds et les arêtes forment ensemble un réseau ou, selon un langage mathématique plus formel, un graphe. Habituellement, les interactions entre gènes sont représentées par des réseaux dans lesquels les gènes sont des sommets et les arêtes représentent les connexions régulatrices. Ces réseaux sont connus sous le nom de réseaux de régulation (**Figure 2**).



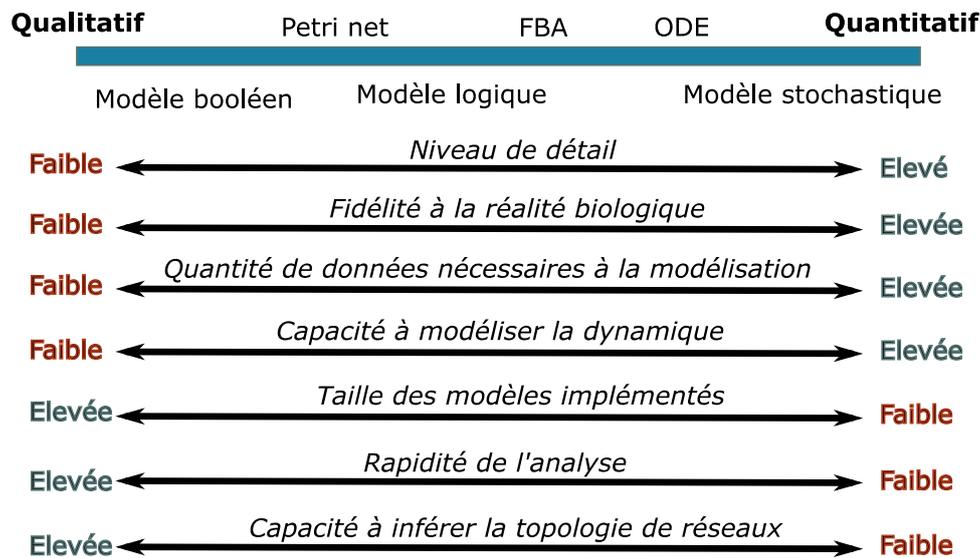
**Figure 2 Exemple d'un réseau de régulation de gènes constitué de trois gènes.** (a) La protéine p<sub>z</sub> agit comme inhibiteur de l'expression du gène y alors que p<sub>y</sub> agit comme activateur de l'expression de du gène x. (b) Les interactions entre les entités du réseau peuvent être représentées sous forme de graphe. (c) Variables du modèle. (d) Relation mathématique entre les variables sous la forme d'équations différentielles ordinaires. (e) En résolvant les équations du modèle, nous pouvons obtenir des prédictions pour les concentrations des variables d'état au cours du temps.

L'augmentation des ressources computationnelles associée à la génération de grandes quantités de données sous forme numérique a permis le développement d'un certain nombre de méthodes afin de modéliser et de simuler ces réseaux de régulation. Cette diversité

d'approches a parallèlement conduit à la naissance d'une incertitude quant au choix de la méthode la plus adaptée lors de la mise en œuvre d'une approche de modélisation (Le Novère, 2015). Plusieurs paradigmes de modélisation ont été développés dans le cadre de l'analyse des réseaux de régulation. Ces modèles peuvent globalement être divisés en deux classes de modèles : les modèles qualitatifs, et les modèles quantitatifs. Nous pouvons définir les modèles qualitatifs comme des modèles mathématiques dans lesquels les valeurs discrètes des variables sont déterminées par les combinaisons logiques des valeurs des autres variables. En complément, les modèles quantitatifs sont des modèles mathématiques dans lesquels les valeurs des variables sont déterminées par les paramètres du système et par l'analyse numérique des variables.

Chaque approche possède ses avantages et ses faiblesses. Suite à des développements récents, il est aussi possible que les approches dites hybrides, mélangeant à la fois paradigmes quantitatifs et qualitatifs, prennent une place de plus en plus importante en modélisation. Les modèles quantitatifs qui sont souvent basés sur l'application des cinétiques chimiques ont été utilisés pour simuler pléthore de réseaux métaboliques, de voies de signalisation et de réseaux de régulation de gènes (voir l'article de revue Chen *et al.*, 2010). Ces modèles permettent d'obtenir des concentrations exactes de molécules et d'examiner de façon fine la temporalité des acteurs du système. Parallèlement, l'apport d'expérimentations conduisant à une compréhension qualitative du système, comme des criblages phénotypiques, ont mené au développement de méthodes pour modéliser des réseaux de régulation de gènes sur la base de règles logiques (modélisation logique) (Thomas, 1973), permettant de répondre à des questions de nature plus qualitative.

Dépendant à la fois de la taille du modèle, de la nature et de la quantité de données disponibles et de la granularité des réponses attendues des simulations, différentes approches peuvent être utilisées pour représenter des mécanismes de régulation. Utilisés avec des données expérimentales quantitatives, les modèles quantitatifs sont des outils puissants pour décrire et comprendre des systèmes biologiques. Leurs simulations fournissent des prédictions quantitatives dans le temps qui peuvent être déterminantes dans la compréhension de processus biologiques. Cependant, bien que la modélisation quantitative soit un formalisme naturel pour représenter des réseaux de molécules et qu'elle fournisse des prédictions précises, le manque de données cinétiques empêche son utilisation dans un certain nombre de cas, rendant alors l'utilisation des formalismes qualitatifs préférable (**Figure 3**).



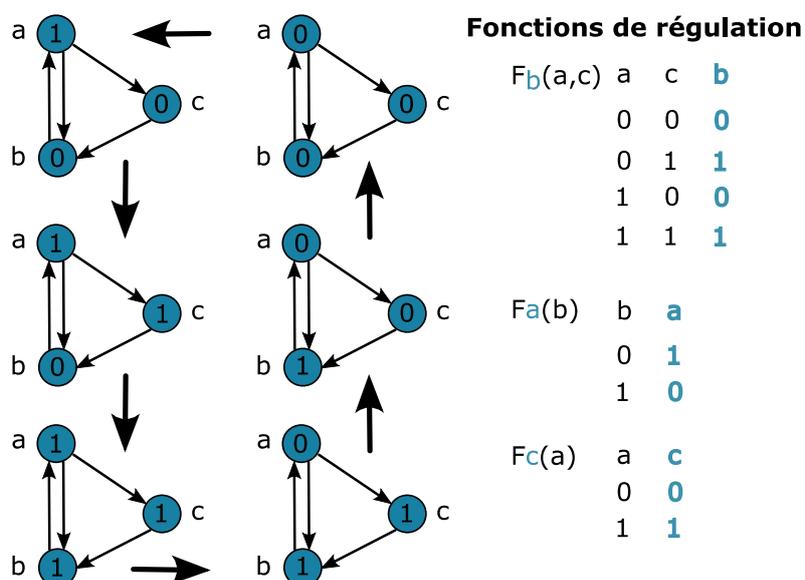
**Figure 3. Comparaison schématique des modèles de réseau de régulation** (adapté d'après Karlebach et Shamir, 2008). Les formalismes de modélisation sont listés le long de l'échelle, allant des modèles qualitatifs aux modèles quantitatifs.

Nous présenterons ici, sans être exhaustif, certains formalismes de modélisation (réseaux booléens, réseaux de Petri, les équations différentielles ordinaires ainsi que les modèles stochastiques). Ainsi, les équations différentielles linéaires par morceaux, formalisme proche des modèles logiques par leur niveau d'abstraction ne seront pas présentées (de Jong *et al.*, 2004, Casey *et al.*, 2006). De même, les analyses de flux, très largement utilisées pour les modèles de métabolismes ne seront pas traitées ici (Dandekar *et al.*, 2014).

#### a. Les réseaux booléens

L'utilisation de formalisme de modélisation se basant sur le modèle booléen a été initiée par Stuart Kauffman (Glass et Kauffman, 1973) et René Thomas (Thomas, 1973). Dans un modèle booléen, les variables peuvent aussi bien représenter l'activité d'un gène, la présence d'une protéine ou son absence dans la cellule, ou encore un phénotype (un individu malade ou sain par exemple). Dans un réseau booléen, la forme la plus simple des modèles logiques, une entité peut avoir deux états alternatifs : un état présent/actif (1) ou un état absent/inactif (0). Un gène peut par exemple être exprimé ou non, une protéine être présente ou non. Ainsi, un vecteur booléen décrit les états de toutes les entités et correspond à l'état global du système. L'état de chaque entité est ensuite mis à jour selon la présence des autres entités via une fonction booléenne : celle-ci représente les interactions entre les variables. Le système est mis à jour de façon synchrone, dès lors qu'à chaque incrémentation l'état de chaque entité est recalculé selon l'état de ses régulateurs au temps précédent et selon la fonction booléenne

(Figure 4). Ainsi, dans cet exemple, si la variable « a » est dans l'état 0 et que la variable « c » est dans l'état 1, alors « b » sera dans l'état 1. Les transitions entre les états dans un tel réseau sont déterministes. Une fois qu'un modèle logique est établi, les changements d'états consécutifs du système, appelés trajectoires, peuvent être reconstruits. Les changements d'état du système prennent place à chaque incrémentation du temps où le pas d'incrémentación peut avoir des valeurs différentes selon les transitions. En effet, dans les modèles logiques, le temps n'est pas représenté explicitement mais de façon abstraite. Au travers de ce formalisme, il est possible d'étudier les attracteurs possibles d'un système, c'est-à-dire lorsque un état du système ou une suite d'états sont reproduits à l'infini et peuvent être représentés par un état stable ou par un ensemble d'états traduisant un comportement oscillatoire.



**Figure 4. Exemple de réseau booléen** (adapté d'après Karlebach et Shamir, 2008). Chacune des entités a, b ou c du réseau peut être dans l'état 0 ou 1. Les transitions obéissent aux fonctions de régulation montrées à droite. Les flèches indiquent les régulateurs de chaque nœud du réseau. Les pas de temps sont représentés par les flèches épaisses. L'état global du système est constitué, à un temps donné, par la combinaison des états des 3 entités. Ainsi, le système boucle entre 6 états globaux du système et une suite d'états globaux consécutifs est appelée une trajectoire.

Les modèles booléens constituent la forme la plus simple des modèles logiques. Dans les modèles logiques, les variables peuvent prendre plusieurs valeurs, permettant alors de coder, de modéliser des comportements plus complexes avec des influences différentes d'une variable sur d'autres variables en fonction de son état. Les modèles logiques sont souvent utilisés avec des variables représentant des états qualitatifs qui prennent généralement comme valeurs des entiers discrets. Cela permet au modèle d'acquérir un aspect semi-quantitatif avec des quantités pouvant, par exemple, être considérées nulles, faibles, moyennes ou élevées, et

ainsi d'implémenter différents niveaux d'activation sur les partenaires du système (aucun effet, un effet stimulateur ou inhibiteur), comme l'illustre la modélisation du réseau de régulation contrôlant la morphogénèse florale chez *Arabidopsis thaliana* (Mendoza *et al.*, 1999). Dans ce travail, au travers de la modélisation de 10 gènes impliqués dans le développement floral, les auteurs proposent une analyse des états stables de ce réseau et retrouvent les motifs d'expressions géniques présents dans les organes floraux d'*Arabidopsis*. De plus, le modèle développé prédit l'existence d'un régulateur supplémentaire, permettant la transition de l'état végétatif à l'entrée de la voie de signalisation déterminant la floraison.

Initialement développés il y a plus de 40 ans (Thomas, 1973), les modèles booléens sont restés relativement théoriques jusqu'à la fin des années 1990 où ils ont appuyé la modélisation de réseaux de régulation de gènes essentiellement impliqués dans la régulation du développement (Sánchez *et al.*, 1997, Yuh *et al.*, 1998). Les réseaux booléens ont aussi été utilisés pour analyser la relation entre les fonctions de régulation et la stabilité du réseau transcriptionnel chez la levure (Kauffman *et al.*, 2003). Dans ce travail, les auteurs ont notamment montré que le réseau était stable lorsque des fonctions aléatoires de régulation étaient utilisées mais aussi que le nombre d'états stables augmente lorsque les fonctions de régulation utilisées avaient une réalité biologique.

Afin d'étudier la dynamique de la régulation du cycle cellulaire chez la levure (Li *et al.*, 2004), des chercheurs ont construit un réseau de régulation à partir de la littérature, réseau dans lequel les fonctions de régulation sont des fonctions seuils, c'est-à-dire que la sortie du modèle est déterminée par la somme de ses entrées et par la comparaison de cette somme à un seuil. Par exemple, un gène qui est surexprimé si au moins deux des trois facteurs de transcription le régulant sont actifs, peut être modélisé par une telle fonction. Le modèle génère un ensemble de trajectoires et les auteurs montrent que ces trajectoires empruntent une suite d'états qui correspond aux phases du cycle cellulaire de la levure. De plus, la plupart des changements topologiques du modèle n'ont pas impacté de façon dynamique son comportement, indiquant que le modèle généré est robuste. Cependant, étant donné que l'analyse repose sur une énumération exhaustive de toutes les trajectoires possibles, cette méthode demeure utilisable seulement sur les petits réseaux.

Depuis, les modèles logiques ont pris une part grandissante dans la compréhension de la différenciation cellulaire et ont été notamment utilisés dans l'étude de l'hématopoïèse (Bonzanni *et al.*, 2013) et du développement des cellules souches embryonnaires (Xu *et al.*, 2014). De même que pour les réseaux de régulation de gènes, la modélisation des voies de signalisation cellulaire souffre également du manque de données pour réaliser des modèles

quantitatifs. Dans ces cas, les modèles qualitatifs ont permis de réaliser des avancées, par exemple dans la compréhension de plusieurs voies de signalisation (Wittmann *et al.*, 2008, Calzone *et al.*, 2010, Grieco *et al.*, 2013).

### **b. Les réseaux booléens probabilistes**

Il existe plusieurs paradigmes de modélisations logiques et de nombreuses contributions méthodologiques ont été déployées depuis les premiers modèles (voir la revue Morris *et al.*, 2010). En particulier, l'intégration du temps dans le modèle est sujette à plusieurs variations. Un schéma de mise à jour est en effet nécessaire lorsque des modèles qualitatifs sont simulés. Les variables dans des modèles logiques peuvent être actualisées de façon synchrone (où toutes les valeurs de toutes les variables sont calculées après chaque transition) ou de façon asynchrone quand une seule valeur de variable est mise à jour après chaque transition (Garg *et al.*, 2008). Le schéma asynchrone nécessite l'ajout de fonctions de priorisation afin de déterminer quelle variable sera la première à être mise à jour.

Souvent, à cause du manque de données expérimentales ou de la mauvaise compréhension d'un système, plusieurs fonctions de régulation peuvent être possibles pour décrire une réaction. Shmulevich et collaborateurs (2003) ont intégré cette incertitude dans les modèles logiques. Les auteurs ont travaillé sur cet aspect en modifiant un modèle booléen de telle façon qu'une entité puisse avoir plusieurs fonctions régulatrices, chacune d'entre elles étant probabilisée *a priori* par rapport aux données expérimentales. Ainsi, à chaque pas de temps, chaque entité est soumise à une fonction de régulation qui est sélectionnée de façon aléatoire selon les probabilités définies. Le modèle peut alors être considéré comme stochastique car un état initial donné du système peut mener à plusieurs trajectoires. Ce modèle génère une séquence d'états du système constituant une chaîne de Markov (processus stochastique dans lequel l'état suivant dépend seulement de l'état présent, sans tenir compte de la suite d'états ayant mené à l'état présent). Par exemple, un réseau probabiliste booléen a été utilisé pour modéliser un sous-réseau de 15 gènes inféré à partir de données d'expression de gliome humain (Shmulevich *et al.*, 2003). Cette analyse démontre que la distribution des variables à l'état stationnaire peut indiquer des relations de régulation possibles entre celles-ci : les entités qui ont les mêmes états dans une proportion significative des états globaux sont susceptibles d'être liées, comme proposé ici entre le facteur de transcription NFκB et le récepteur Tie-2 régulant l'angiogenèse et le développement tumoral. Ainsi, les réseaux booléens probabilistes offrent une grande flexibilité, en intégrant notamment différentes fonctions de probabilité pour mettre à jour chaque sommet. Dans ce modèle stochastique, un seul état initial peut alors

évoluer vers différentes trajectoires, chacune associée à une probabilité. Les simulations stochastiques de ces modèles permettent de modéliser la variabilité des réponses aux perturbations environnementales (Liang *et al.*, 2012).

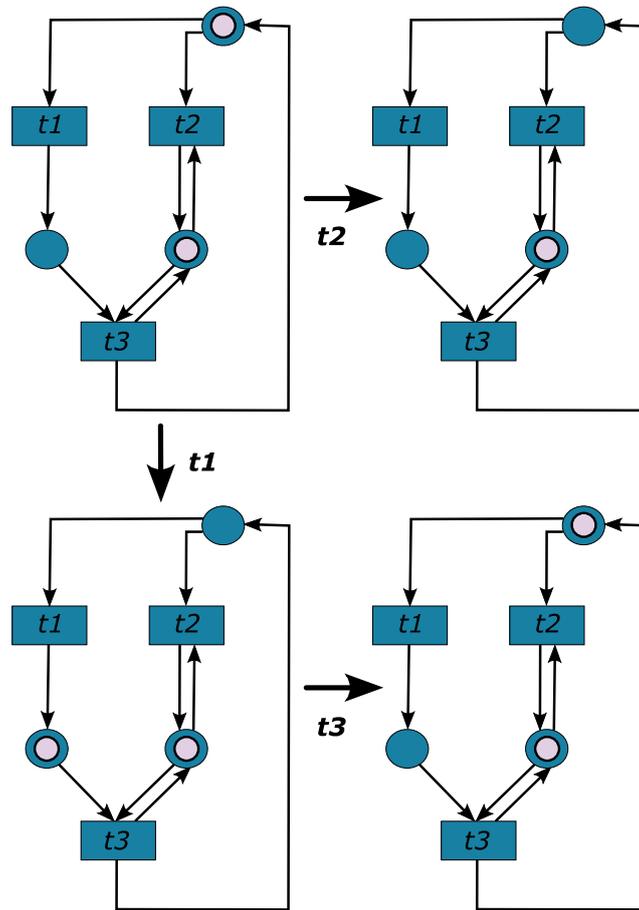
De récentes avancées méthodologiques portant sur les réseaux booléens ont permis d'intégrer un temps continu pendant les simulations en utilisant un formalisme discret (Stoll *et al.*, 2012). En effet, l'inconvénient majeur dans les modèles qualitatifs demeure l'absence d'états transitoires entre états du système où le temps est représenté par des pas discrets. Dans leurs travaux, Stoll et collaborateurs proposent une modélisation qualitative en temps continu, basée sur un processus de Markov en temps continu et appliquée dans un espace booléen. Par ailleurs, des régulations continues ont aussi été intégrées dans les modèles logiques grâce à la logique floue (Aldridge *et al.*, 2009) qui permet d'encoder les valeurs intermédiaires prises par les variables.

Comme pour les modèles quantitatifs, une part importante dans la construction des modèles logiques réside dans le paramétrage du modèle qui consiste ici, non pas à choisir des valeurs numériques de paramètres, mais à déterminer les fonctions logiques permettant de mettre à jour les valeurs des variables. Afin de faciliter le processus de construction du modèle, des méthodes (Terfve *et al.*, 2012) permettent d'adapter automatiquement le modèle logique à des données expérimentales. Enfin, les modèles logiques permettent également l'utilisation de techniques de *model checking* qui évaluent si un modèle est capable de produire un comportement donné (Carrillo *et al.*, 2012) mais que nous ne détaillerons pas ici.

### c. Les réseaux de Petri

Les réseaux de Petri, nommés d'après Carl Petri (Petri, 1964), offrent une représentation graphique et un formalisme mathématique applicable à la modélisation de systèmes concurrents (Petri, 1962, Murata, 1989, Chaouiya, 2007, Koch et Heiner, 2008). Un réseau de Petri standard est un graphe orienté biparti (composé de deux classes de sommets tels que deux sommets de la même classe ne sont pas connectés) : les places et les transitions (voir **Figure 5**). En biologie, les places représentent généralement les ressources du système comme, par exemple des protéines, des métabolites ou même des gènes. Les transitions représentent les événements pouvant changer l'état des ressources. Les places et les transitions sont alors interconnectées par des arcs dirigés et orientés qui décrivent les relations entre ressources et événements. Ainsi, les arcs ne connectent seulement que des sommets de nature différente. Dans la représentation graphique, les places sont représentées par des cercles, les transitions par des carrés et les arcs dirigés par des flèches. Un arc connecte ainsi

une réaction (une transition) avec ses préconditions, les substrats de la réaction d'une réaction (les pré-places), avec les post-conditions, les produits d'une réaction (post-places). Le poids des arcs définit le nombre d'éléments d'une ressource qui est consommé ou produit par une transition. Ce sont les coefficients stœchiométriques des réactions. Enfin, les jetons sont les éléments dynamiques du réseau et peuvent se déplacer d'une place à une autre selon la transition connectant les deux places. Pour une place représentant une ressource, le nombre de jetons de cette place indique le nombre de ressources disponibles pour cette place (par exemple le nombre de protéines). À chaque instant dans un réseau de Petri, une place peut contenir 0 ou un nombre entier positif de jetons. La distribution du nombre de jetons sur l'ensemble des places décrit un état du système, appelé aussi marquage du réseau de Petri. Le marquage initial exprime l'état initial du système. Pour démarrer le réseau, les transitions doivent être réalisables. Une transition est dite réalisable si ses pré-places contiennent au moins autant de jetons que le nombre défini par le poids des arcs connectés aux pré-places de la transition. À tout moment, chaque transition qui a suffisamment de jetons dans ses pré-places peut être déclenchée. Si une transition est réalisée, le nombre requis de jetons est retiré des pré-places et le nombre requis de jetons est ajouté aux post-places, selon le poids de chaque arc. Ce mouvement de jetons représente alors l'évolution dynamique du système. Dans l'exemple présenté, les transitions  $t_1$  et  $t_3$  peuvent être déclenchées en alternance indéfiniment, alors qu'aucune autre transition ne peut être déclenchée après le déclenchement de  $t_2$  (**Figure 5**).



**Figure 5. Exemple de réseau de Petri** (adapté d'après Karlebach et Shamir, 2008). Le réseau contient des places (cercles bleu) qui sont les entités du modèle et des « transitions » (rectangles) qui constituent les fonctions de régulation et définissent la dynamique du modèle. Les arcs connectent les pré-places aux transitions et les transitions vers leurs post-places. Les jetons sont représentés par des points associés aux places. Les labels des flèches indiquent quelle transition est déclenchée.

Les réseaux de Petri permettent de réaliser des analyses de structure du réseau parmi lesquelles les recherches de T-invariants et de P-invariants. Les P-invariants représentent un ensemble de places dont la somme des jetons est constante et met en évidence une conservation de matière dans une sous-partie du réseau. Les T-invariants représentent un ensemble de transitions dont les réalisations permettront au système de retourner à son état initial. Ces T-invariants permettent par exemple de mettre en évidence des comportements oscillatoires ou cycliques du réseau de Petri étudié. Les analyses des P et des T-invariants peuvent donc être utilisées pour vérifier, qualitativement, la cohérence d'un modèle, et pour tester la validité des hypothèses biologiques incorporées dans le modèle.

Afin d'obtenir les invariants du réseau, il est nécessaire de calculer la matrice d'incidence  $C$  qui contient pour chaque transition (colonne) et pour chaque place (ligne) du réseau la différence entre le nombre de jetons et le nombre de jetons consommés, donné par :

$$C = Post - Pre$$

où  $Post$  est la matrice de post-condition et  $Pre$  la matrice de pré-condition. La matrice de pré-condition contient le poids des arcs allant des pré-places aux transitions et la matrice de post-condition contient le poids des arcs allant d'une transition aux post-places. Un vecteur de places  $x$  est appelé P-invariant s'il s'agit d'un vecteur d'entiers non-négatifs solution du système linéaire d'équation  $x^T \cdot C = 0$  et un vecteur de transitions  $y$  est appelé T-invariant s'il s'agit d'un vecteur d'entiers non-négatifs solution du système linéaire d'équation  $C \cdot y = 0$  (voir **Figure 6**). Dans cet exemple, comme chaque place appartient à un P-invariant et que chaque transition appartient à un T-invariant, on dira que le modèle est couvert à la fois par des P et des T-invariants.

Représentation graphique du réseau de Petri	Matrice de pré-condition (Pre)	Matrice de post-condition (Post)	Matrice d'incidence C
	$\begin{bmatrix} -1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} -1 & 1 \\ 1 & -1 \\ 1 & -1 \end{bmatrix}$
P-invariants $x^T \cdot C = 0$ $\begin{cases} -P1 + P2 + P3 = 0 \\ P1 - P2 - P3 = 0 \end{cases}$ Solutions $[1,1,0]^T$ $[1,0,1]^T$		T-invariants $C \cdot y = 0$ $\begin{cases} -T1 + T2 = 0 \\ T1 - T2 = 0 \\ T1 - T2 = 0 \end{cases}$ Solution $[1,1]$	

**Figure 6. Exemple de calcul d'invariants d'un réseau de Petri.** Les places (P1, P2 et P3) sont représentées par des ronds et les transitions T1 et T2 sont représentées par des carrés. Seules les places P2 et P3 sont marquées avec des jetons. Comme les arcs ne sont pas pondérés, cela signifie que les poids sont de 1, *i.e.*, un jeton est consommé ou produit par chaque transition. Les lignes dans les matrices de pré-condition, de post-condition et de la matrice d'incidence correspondent aux trois places dans l'ordre suivant P1, P2 et P3, et les colonnes aux deux transitions T1 et T2. La matrice d'incidence est utilisée pour calculer les P et les T-invariants.

L'ensemble de transitions (ou de places) correspondant aux valeurs non nulles dans un T-invariant  $x$  (ou P-invariant) est appelé support de cet invariant  $x$  et s'écrit  $supp(x)$ . Un

invariant  $x$  est appelé minimal s'il ne contient pas d'autre invariant  $z$ , *i.e.*, il n'existe aucun invariant  $z$  tel que son support est inclus dans le support de l'invariant  $x$  ( $\nexists$  invariant  $z$   $supp(z) \subset supp(x)$ ) (Murata, 1989).

La description de réactions biochimiques, principalement très qualitative, utilisant les réseaux de Petri, est simple à assimiler, et les réseaux de Petri ont en majorité été très utilisés pour décrire des réseaux métaboliques (Marwan *et al.*, 2005, Sackmann *et al.*, 2006, Grunwald *et al.*, 2008, Formanowicz *et al.*, 2011, Cordero *et al.*, 2011, Formanowicz *et al.*, 2013, Geistlinger *et al.*, 2013). Voici des exemples de questions auxquelles il est possible de répondre en utilisant les réseaux de Petri : combien de transitions sont nécessaires pour que le système transite d'un état A à un état B ? Toutes les réactions sont-elles essentielles au comportement du système ?

Chaouiya et collaborateurs (2004) ont montré que les réseaux de Petri pouvaient représenter non seulement des voies de biosynthèse mais aussi des réseaux de régulation, en y ajoutant des fonctions de régulation. Cela a ainsi mené à la publication d'un travail exposant que les couches métaboliques et régulatrices des réseaux pouvaient être intégrées dans un réseau de Petri, afin de modéliser par exemple la voie de biosynthèse du tryptophane chez *E. coli* (Simao *et al.*, 2005). Steggles et collègues ont modélisé le réseau de régulation de la sporulation de *Bacillus subtilis* (Steggles *et al.*, 2007) en utilisant les réseaux de Petri. Leurs résultats génèrent un comportement qui est en accord avec les résultats expérimentaux ; par exemple, lorsque le réseau est initialisé dans un état qui correspond à l'état végétatif, en activant le signal de sporulation la dynamique du système mène à un état correspondant à la sporulation. Le modèle prédit par ailleurs correctement les capacités de sporulation des différents mutants.

#### **d. Les équations différentielles ordinaires**

Sans aucun doute, l'approche la plus utilisée pour modéliser des réseaux moléculaires dynamiques est celle basée sur l'application des cinétiques chimiques au travers des équations différentielles ordinaires. Une réaction chimique est la transformation d'un ensemble de substances appelées réactants en un autre ensemble de substances appelées produits. Les quantités consommées ou produites par une réaction sont appelées coefficients stœchiométriques. Ce coefficient est positif pour les produits et négatif pour les réactants. Si une substance n'est ni consommée ni produite par une réaction, alors son coefficient stœchiométrique est nul. À l'échelle microscopique, pour qu'une réaction chimique ait lieu, les acteurs de cette réaction doivent entrer en collision. La probabilité d'une telle collision

dépend notamment de la densité locale des réactants et, dans des environnements homogènes, de leur concentration. La relation mathématique décrivant ce principe fut découverte par Guldberg et Waage (1864) qui supposèrent qu'à l'équilibre, les vitesses des réactions *forward* et *backward* étaient égales et que ces vitesses étaient proportionnelles à la concentration des réactants exposant le coefficient de stœchiométrie. Cette relation, liant vitesse d'une réaction et concentration des réactants, est appelée loi d'action de masse. Les vitesses de réactions sont proportionnelles au produit de la concentration des réactants exposant l'ordre partiel des réactions. La somme de tous les ordres partiels est appelée l'ordre d'une réaction et le facteur de proportionnalité est la constante de vitesse  $k$ , d'où la relation suivante :

$$v = k * [A]^a * [B]^b$$

La loi d'action de masse permet de modéliser toutes les réactions élémentaires de liaisons, de dissociations, de catalyses et d'états transitoires d'un système. Cette approche est la plus précise si une étude quantitative, bien souvent pour de petits réseaux moléculaires, doit être entreprise. Cependant, dans certaines conditions, des lois de vitesse simplifiées peuvent être utilisées. La simplification la plus fréquente des réactions de synthèse est la cinétique d'Henry-Michaelis-Menten. En effet, si la formation du complexe enzyme-substrat (ou ligand-récepteur ou promoteur-facteur de transcription) est plus rapide que la synthèse du produit enzymatique, il n'est alors pas nécessaire de modéliser le complexe enzyme-substrat. La vitesse de formation du produit enzymatique est alors :

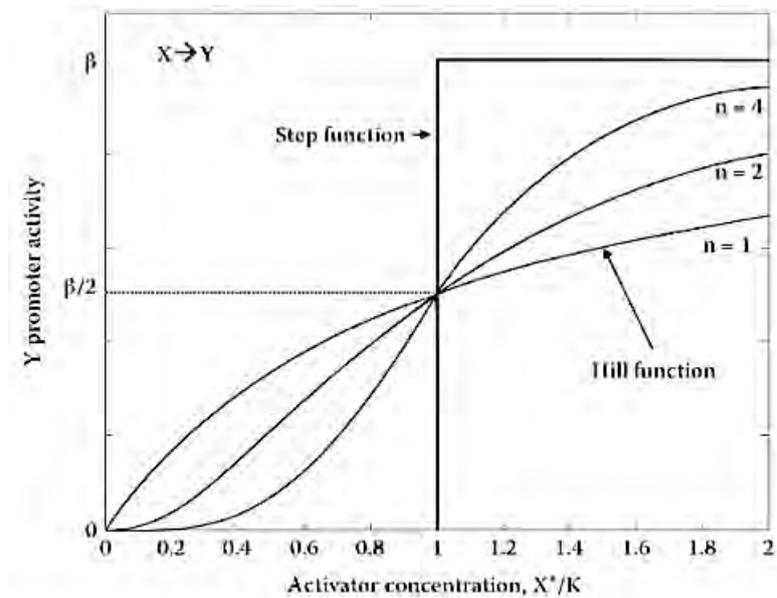
$$v = \frac{v_{max} * [S]}{Km + [S]}$$

où  $v_{max}$  est la vitesse maximale de la réaction,  $[S]$  la concentration en substrat et  $Km$  la concentration de substrat nécessaire pour atteindre la moitié de la vitesse maximale de la réaction. Souvent, le détail des mécanismes moléculaires sous-jacents à un processus est inconnu, ou bien le but est de réduire la complexité du processus et de se concentrer sur l'effet principal d'un composant. Une façon générique de représenter des modulations de synthèse si les mécanismes biochimiques sous-jacents sont inconnus est de multiplier l'activité d'un gène ou la vitesse d'un processus par une fonction de Hill de la forme :

$$\frac{X^n}{(X^n + K^n)}$$

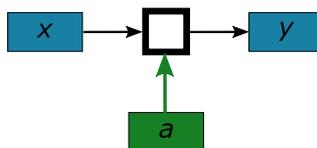
où  $K$  représente la concentration de  $X$  pour laquelle son effet est de 50%. Ainsi, si  $X$  tend vers 0, la fonction (et donc l'activité du promoteur) tendra vers 0 et, inversement, si  $X$  est très grand la fonction tendra vers 1. La valeur de  $K$  représente donc la sensibilité de la réponse.

Ainsi, plus  $K$  est faible, plus rapide sera la réponse pour une valeur de  $X$  donnée. L'exposant  $n$  contrôle la pente de la réponse à  $X$  (correspondant à la coopérativité). Avec  $n$  égal à 1, le système répond linéairement à  $X$  et lorsque la valeur  $n$  augmente, la réponse sera de moins en moins linéaire et deviendra alors une réponse de type seuil (**Figure 7**).



**Figure 7. Activité d'un promoteur décrit par une fonction de Hill** (extrait de « An Introduction to Systems Biology: Design Principles of Biological Circuits » de U. Alon). L'activité du promoteur  $Y$  est exprimée en fonction de la quantité de l'activateur  $X$  pour différentes valeurs du coefficient de Hill  $n$ . L'activité maximale du promoteur est  $\beta$ .

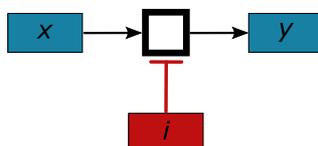
Prenons l'exemple de la réaction suivante, où le réactant  $x$  est transformé en produit  $y$  grâce à l'action de l'enzyme  $a$  :



En utilisant la fonction de Hill, l'évolution de  $y$  au cours du temps peut alors s'écrire :

$$\frac{d[y]}{dt} = v_{max} * [x] * \frac{[a]^n}{Ka^n + [a]^n}$$

Et inversement, si un composé  $i$  est inhibiteur de la même réaction :

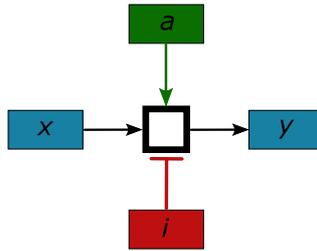


L'on peut décrire l'évolution de  $y$  au cours du temps par l'équation suivante :

$$\frac{d[y]}{dt} = v_{max} * [x] * \frac{Ki^m}{Ki^m + [i]^m}$$

où  $Ki$  représente la concentration de l'inhibiteur  $X$  pour laquelle son effet est de réduire de moitié la vitesse maximale de la réaction.

Finalement, il est possible de combiner plusieurs fonctions de Hill et nous pouvons alors modéliser la coexistence d'un activateur et d'un inhibiteur agissant sur un même processus :



Dans le cas d'une inhibition allostérique, nous obtenons alors l'équation différentielle suivante:

$$\frac{d[y]}{dt} = v_{max} * [x] * \frac{[a]^n}{Ka^n + [a]^n} * \frac{Ki^m}{Ki^m + [i]^m}$$

ou dans le cas d'une inhibition compétitive entre deux régulateurs liant un seul site :

$$\frac{d[y]}{dt} = v_{max} * [x] * \frac{[a]^n}{Ka^n * \left(1 + \frac{[i]^m}{Ki^m}\right) + [a]^n}$$

De la même façon, il sera possible de combiner plusieurs fonctions de régulation pour le même régulateur afin de représenter différents modes de régulation selon la concentration du régulateur. Des réactions plus complexes peuvent aussi être modélisées en prenant par exemple en compte des niveaux basaux d'expressions en l'absence d'activateurs, une non-indépendance entre plusieurs régulateurs, etc.

Sans être exhaustif, cette formulation a été utilisée, au travers des équations différentielles ordinaires, dans la modélisation de très nombreux systèmes biochimiques comme, par exemple, pour la voie de signalisation du calcium dans des cellules musculaires (Parthimos *et al.*, 2007), dans des modèles du cycle cellulaire chez les mammifères (Yao *et al.*, 2011) et dans des modèles de réseaux de régulation contrôlant le cycle cellulaire chez les bactéries (Li *et al.*, 2008, Lin *et al.*, 2010). Les fonctions de Hill ont aussi été employées pour comprendre

la segmentation cellulaire en biologie du développement (Von Dassow *et al.*, 2000, Goldbeter *et al.*, 2008, Ozbudak *et al.*, 2008).

Bien que les fonctions de Hill soient très souvent utilisées, il faut être conscient que cette relation repose sur des hypothèses qui peuvent ne pas toujours être valides en pratique : l'hypothèse centrale de ce modèle est que la liaison ainsi que la dissociation du régulateur sont des processus extrêmement rapides en regard des processus régulés et peuvent ainsi être ignorées parce que la fraction de régulateur liée est à l'équilibre. Bien que cette hypothèse ne soit pas toujours valide pour des cascades de signalisation dans lesquelles les perturbations et les réponses sont sur la même échelle de temps que les associations et les dissociations, elle demeure généralement valable pour les réseaux de régulation de gènes grâce à la différence importante entre la dynamique de diffusion du facteur de transcription et de sa liaison au promoteur par rapport au processus de transcription et de traduction.

Ainsi, les équations différentielles ordinaires permettent de décrire le changement d'une variable du système au cours du temps en fonction des valeurs des autres variables et des paramètres du système. Les équations différentielles ordinaires sont déduites de la combinaison de différentes réactions dans lesquelles une entité, représentée par une variable, est impliquée. L'abondance de chaque espèce moléculaire (la concentration d'un métabolite, une quantité d'ARNm, etc.) est assignée à une variable d'état dans le modèle. Cette quantité est contrôlée de façon dynamique par combinaison de tous les processus qui contribuent à augmenter la quantité d'une espèce moléculaire (par exemple la synthèse, l'import ou l'activation) et de tous les processus qui contribuent à la diminution de la quantité de cette espèce (par exemple dégradation, export et inhibition). Chaque processus est caractérisé par une vitesse, modélisée par une réaction (loi d'action de masse, fonction de Hill, etc.) modulée par plusieurs paramètres, y compris la quantité ou l'activité d'autres espèces moléculaires. Le système peut ensuite être simulé, en utilisant une méthode d'intégration numérique, ce qui permet de calculer les changements des valeurs de variables sur de courts intervalles de temps. Finalement, le comportement dynamique obtenu avec le modèle est l'évolution de l'ensemble des variables d'état au cours du temps. Les équations de vitesse ont la forme mathématique suivante :

$$\frac{dx_i}{dt} = f_i(x), \quad 1 \leq i \leq n,$$

où  $x = [x_1, \dots, x_n] \geq 0$  est un vecteur de concentrations de protéines, d'ARN messagers ou encore de métabolites du système, et  $f_i$  est une fonction généralement non linéaire comme par

exemple la fonction de Hill mentionnée précédemment. La vitesse de synthèse de  $x_i$  est ainsi dépendante des concentrations du vecteur  $x$ .

Outre les variables, le modèle comporte également des paramètres dont les valeurs sont généralement fixées. Ces paramètres du modèle caractérisent l'interaction du système avec l'environnement. Par exemple, nous pouvons citer les constantes d'association, les vitesses maximales de synthèse ou encore les constantes de dégradation. Un changement dans les valeurs des paramètres du modèle correspond à un changement des conditions environnementales ou du système lui-même. Ainsi, les paramètres du modèle sont la plupart du temps tenus constants pendant les simulations, mais leurs valeurs peuvent varier pour explorer le comportement du système sous certaines perturbations ou dans un environnement différent, correspondant par exemple à différentes conditions expérimentales.

Enfin, les équations différentielles ordinaires constituent un paradigme déterministe. Un modèle mathématique est dit déterministe si son comportement est exactement reproductible. Bien que le comportement d'un modèle déterministe soit dépendant d'un ensemble de conditions spécifiques, aucune force extérieure n'influence le système et plusieurs simulations répétées sous une même condition sont toujours parfaitement identiques (ce sont des répliques parfaits). Les formalismes déterministes comme les équations différentielles ordinaires sont très bien adaptés à la modélisation dynamique du fonctionnement d'un réseau de régulation, mais ils ne considèrent uniquement que le comportement moyen d'une population de cellules génétiquement identiques, contrairement aux modèles stochastiques.

#### **e. Les modèles stochastiques**

En complément des modèles déterministes, les modèles stochastiques permettent d'introduire du bruit, de l'aléatoire, dans le comportement des systèmes. En effet, de nombreuses preuves expérimentales suggèrent que l'expression génique, à la fois chez les eucaryotes et les procaryotes, est un processus intrinsèquement stochastique (Elowitz *et al.*, 2002, Rao *et al.*, 2002, Longo *et al.*, 2006). Cela est particulièrement vrai lorsque les réactions biochimiques cellulaires (e.g. liaison/dissociation de facteurs de transcription aux promoteurs, traduction, etc.) sont réalisées par des molécules présentes en faible quantité, rendant les formalismes déterministes peu précis pour décrire le comportement d'une cellule (Paulsson 2002, Wilkinson 2006). En effet, il existe une variabilité inter-cellulaire, dite « intrinsèque », liée au déplacement des molécules de façon aléatoire à l'intérieur de la cellule entraînant une fluctuation de l'expression génique et une variabilité, dite « extrinsèque », liée à la fluctuation des autres composants cellulaires (Elowitz *et al.*, 2002). La stochasticité des processus

biologiques n'est pas considérée lors de la modélisation par équations différentielles ordinaires en moyennant les comportements divergents au sein d'une population cellulaire. Tout réseau biologique est composé de réactions de nature stochastiques et peut alors avoir des issues différentes à partir d'une même condition initiale. Lorsque le nombre de molécules impliquées dans un processus est important, la loi d'action de masse peut être utilisée pour calculer la variation de concentration et quasi aucun effet stochastique n'est observable. Cependant, lorsque le nombre de molécules impliquées dans la régulation d'un processus est faible, il est alors possible d'observer des effets stochastiques comme dans le processus décisionnel du phage lambda d'entrée en phase lytique ou en phase lysogénique (Arkin *et al.*, 1998) ou lors du déclenchement de la compétence chez *Bacillus subtilis* (Süel *et al.*, 2006). Ainsi, des expériences menées à l'échelle de la cellule unique ont montré l'aspect stochastique des processus de transcription et de traduction (Ozbudak *et al.*, 2002, Raj *et al.*, 2006, Cai *et al.*, 2006).

L'algorithme utilisé pour simuler des modèles stochastiques, différents des modèles déterministes, est l'algorithme de Gillespie (Gillespie, 1976) qui utilise comme paramètres le nombre initial de molécules de toutes les espèces moléculaires ainsi que les probabilités de chacune des réactions. La dynamique du système est ensuite simulée réaction par réaction. La probabilité d'une réaction est alors la probabilité qu'une combinaison donnée de molécules puisse participer à ces réactions dans un intervalle de temps infinitésimal.

#### **f. Limites des approches quantitatives et qualitatives**

Bien que les modèles quantitatifs constituent un formalisme permettant de décrire de façon réaliste les mécanismes biologiques, et bien que cette approche fournisse les prédictions très précises, le manque de données quantitatives, entraînant des difficultés à optimiser de façon satisfaisante les paramètres d'un modèle, empêche bien souvent l'utilisation de ce paradigme de modélisation. De plus, la description des réactions, en raison de sa précision, est extrêmement sensible à la multiplicité d'états que peut prendre une protéine. Par exemple, une protéine peut posséder plusieurs conformations, différents états de méthylations, ou bien être intégrée à plusieurs complexes protéiques. Des promoteurs peuvent avoir différents niveaux d'occupation. Ces états, lorsqu'ils sont inconnus, peuvent alors impacter la nature des prédictions faites par le modèle et lorsqu'ils sont *a priori* connus, ils peuvent mener à une explosion combinatoire du nombre de variables et de réactions, rendant le modèle difficile à interpréter (Le Novère, 2015). Enfin, le cadre théorique des réactions chimiques suppose une distribution homogène des réactants dans le système. Or, beaucoup de réactions de

signalisation sont distribuées de façon hétérogène dans les compartiments cellulaires, prenant place, par exemple aux pôles des cellules ou au sein des membranes cellulaires. De même, les régulateurs de gènes sont présents en faible nombre et les réactions prennent place de façon localisée dans des régions spécifiques de l'ADN.

Malgré leur rapidité d'implémentation, les modèles logiques présentent certains inconvénients. D'abord, le nombre d'états du graphe d'états de transition augmente de façon exponentielle avec le nombre de variables. Si le calcul du graphe d'états de transition et des attracteurs (c'est-à-dire les solutions stables du modèle) n'est pas un problème complexe pour les ordinateurs modernes, en revanche, extraire de l'information de ce graphe peut s'avérer compliqué. De plus, l'absence de représentation du temps dans les modèles booléens simples rend difficile la prise en compte des processus rapides et des délais dans le cas de réactions lentes. Enfin, à cause de la nature purement qualitative de ce paradigme, il est souvent difficile de choisir entre des comportements alternatifs proposés par le modèle logique : par exemple, une boucle de rétrocontrôle négative devrait toujours donner lieu à un comportement périodique, alors que, dans un modèle quantitatif, le modèle peut converger vers un état d'équilibre ou avoir un comportement oscillatoire dépendant de la force du rétrocontrôle négatif. Ainsi, pour pallier ces problèmes, des modèles hybrides, mélangeant paradigmes qualitatifs et quantitatifs, ont été développés (Le Novère, 2015).

#### **g. Les modèles hybrides**

En raison de l'étendue des processus biologiques à représenter et à cause de l'hétérogénéité des mesures expérimentales disponibles pour construire et valider le modèle, il est possible d'utiliser des approches hybrides de modélisation dans lesquelles des représentations qualitatives et quantitatives sont utilisées dans un même modèle. Cette approche a été proposée dans le cadre de la modélisation de réseaux de régulation de gènes (McAdams et Shapiro, 1995). Une possibilité, par exemple, est d'utiliser à la fois les équations différentielles ordinaires pour représenter les processus lents et les équations différentielles par morceaux pour représenter sous forme booléenne l'état des composés qui changent rapidement de forme (Singhania *et al.*, 2011). Des modèles logiques de voies de signalisation peuvent également être combinés avec des modèles quantitatifs du métabolisme, pour représenter les cinétiques quantitatives de parties du réseau lorsque cela est nécessaire (Ryll *et al.*, 2014).

Les processus biologiques prennent place sur différentes échelles de temps ; cela a mené au développement de modèles où le métabolisme est représenté à l'état stationnaire et au sein

duquel est intégrée la modélisation logique de la régulation du réseau de gènes (Shlomi *et al.*, 2007). Ces approches ont connu un succès notamment avec la modélisation entière de *Mycoplasma genitalium* qui, sous une forme modulaire, associe différents paradigmes de modélisation faisant de l'ensemble du modèle un système hybride. Ce modèle, basé sur la synthèse de plus de 900 publications, utilise les équations différentielles ordinaires, des processus stochastiques et des analyses de flux pour simuler 28 modules représentant l'ensemble des réseaux de gènes, les voies de signalisation et les voies métaboliques de ce microorganisme (Karr *et al.*, 2012). Les modules ont été construits, paramétrés et testés indépendamment. Certains processus biologiques ont été étudiés de façon quantitative alors que d'autres processus, moins bien caractérisés, sont décrits à partir de formalismes qualitatifs. Ainsi, chaque module a été décrit à l'aide du formalisme mathématique le plus approprié. Par exemple, le métabolisme a été modélisé en utilisant une analyse de balance de flux alors que la dégradation des ARN et des protéines a été modélisée grâce à des processus de Poisson.

### 3. Description du modèle biologique : la régulation de la transformation génétique

*Streptococcus pneumoniae*, également appelé pneumocoque fait partie, sous forme commensale, de la flore naturelle des muqueuses, la sphère oropharyngée. Sous l'influence de certains facteurs comme une immunodépression, il peut devenir pathogène et être alors responsable de pathologies respiratoires (pneumonies, bronchites, pleurésies, sinusites). Le pneumocoque est notamment la cause la plus commune de méningites bactériennes communautaires chez l'adulte (Catterall, 1999). Plus d'un million de décès lui sont attribués chaque année dans le monde, parmi lesquels une majorité d'enfants, particulièrement dans les pays en voie de développement.

Découverte chez le pathogène humain *Streptococcus pneumoniae* (Griffith, 1928), la transformation génétique bactérienne est un mécanisme génétique programmé de transfert horizontal de gènes qui est largement distribué dans différents groupes taxonomiques. L'état physiologique transitoire, et par conséquent régulé, pendant lequel une bactérie est capable d'assimiler l'ADN extracellulaire est appelé état de compétence. Ce mécanisme peut permettre aux bactéries d'acquérir de nouveaux gènes ou de nouveaux allèles et, ainsi, de s'adapter à des changements de conditions environnementales, favorisant par exemple l'apparition de résistances aux antibiotiques ou permettant d'échapper aux vaccins (Croucher *et al.*, 2011). En effet, *S. pneumoniae* possède de nombreux sérotypes capsulaires. La capsule est une enveloppe polysaccharidique recouvrant la bactérie et qui, chez *S. pneumoniae*, est

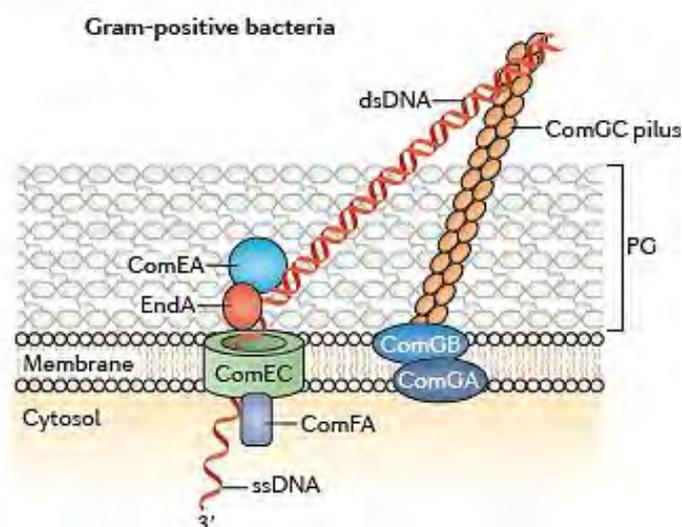
indispensable à la virulence car elle permet d'inhiber la phagocytose des macrophages. La capsule est aussi l'un des principaux déterminants antigéniques ciblés du système immunitaire. C'est pourquoi elle est utilisée dans la préparation de vaccins. Cependant, à cause de la transformation génétique naturelle, il a été constaté une importante variabilité capsulaire, rendant problématique l'élaboration d'un vaccin efficace (Croucher *et al.*, 2011).

La transformation naturelle est un programme génétique qui a été référencé chez environ 80 espèces (Johnston *et al.*, 2014). Dans cette étude, les auteurs ont montré que même des espèces éloignées dans l'évolution partagent une machinerie protéique conservée permettant l'internalisation et le remaniement de l'ADN. Ce mécanisme repose sur un ensemble de gènes conservés, exprimés de façon simultanée. Contrairement aux autres mécanismes de transfert de gènes comme la transduction et la conjugaison, la transformation génétique est entièrement dirigée par la cellule receveuse et toutes les protéines essentielles à la transformation génétique sont codées par le génome cœur. En revanche, les mécanismes régulant la mise en place de cette machinerie protéique semblent diverger entre les espèces (Fontaine *et al.*, 2015, Johnston *et al.*, 2014). En effet, la plupart des bactéries transformables n'expriment pas de façon constitutive les protéines impliquées dans ce processus et nécessitent des conditions spécifiques pour développer la compétence pour la transformation génétique. La période d'opportunité de la compétence apparaît être régulée en réponse à des signaux environnementaux espèce-spécifique (Claverys *et al.*, 2006). La transformation génétique est un processus impliquant deux partenaires : l'ADN exogène et une cellule receveuse. L'internalisation de l'ADN exogène et son intégration dans le génome hôte sont réalisées par recombinaison homologue grâce à une machinerie protéique appelée transformasome.

#### a. Le transformasome

L'internalisation de l'ADN nécessite l'assemblage d'un complexe multi-protéique (Chen et Dubnau, 2004) impliqué dans le remaniement de l'ADN entrant (**Figure 8**). Les bactéries Gram-positives et les bactéries Gram-négatives partagent des mécanismes hautement similaires pour l'intégration de l'ADN exogène avec, pour différence majeure, la nécessité chez les bactéries Gram-négatives de transporter l'ADN au travers de la membrane externe grâce au canal formé par la protéine PilQ. Chez les bactéries Gram-positives, le pseudopilus participant au transport de l'ADN exogène vers l'intérieur de la cellule est codé par l'opéron ComG, et ComGC est la piline majoritaire composant le pseudopilus (Chen *et al.*, 2006). Cet appendice lie directement l'ADN double brin exogène et est essentiel pour la transformation. Malgré de nombreuses années d'investigation, le rôle du pseudopilus n'a pas été clairement

établi mais il semble très probable qu'il transmette l'ADN exogène à son récepteur ComEA (Provvedi et Dubnau, 1999) localisé à la membrane cytoplasmique chez les bactéries Grampositives et dans l'espace péri-plasmique chez les bactéries Gram-négatives. ComEA délivre probablement l'ADN double-brin à une protéine qui le transforme en ADN simple brin pour son internalisation. Cette nucléase a été identifiée chez *S. pneumoniae* : il s'agit de la protéine EndA (Puyet *et al.*, 1990) qui dégrade un brin de l'ADN et permet l'import du brin complémentaire dans l'orientation 3'→5' au travers du pore membranaire formé par la protéine ComEC (Draskovic et Dubnau, 2007). Chez les Firmicutes, la protéine ComFA associée au pore ComEC joue probablement le rôle de translocase entraînant l'internalisation de l'ADN simple brin (Londoño-Vallejo et Dubnau, 1994, Chen et Dubnau, 2004).



**Figure 8. Représentation schématique du transformasome chez les bactéries Gram positives** (d'après Johnston *et al.*, 2014).

Dans la cellule, l'ADN simple brin est protégé des nucléases intra-cellulaires par liaison à des protéines liant l'ADN simple brin (SsbB, DprA, RecA, CoiA). L'ADN simple brin internalisé est lié à la protéine DprA qui aide au chargement de la protéine recombinase RecA sur l'ADN simple brin (Mortier-Barrière *et al.*, 2007). RecA polymérise sur l'ADN simple brin et permet la recherche d'homologie sur le chromosome de la cellule hôte. Lorsqu'une séquence homologue est trouvée, une double recombinaison homologue permet la formation d'un complexe associant ADN simple brin avec une séquence homologue de l'ADN.

Le fait que les protéines constituant le transformasome appartiennent au génome cœur des bactéries identifiées comme transformables et que, de plus, la phylogénie des deux protéines clés de la transformation que sont DprA et ComEC soit congruente avec celle des espèces

(Johnston *et al.*, 2014) suggèrent que la transformation est un processus ancien qui a été hérité d'un ancêtre commun.

### **b. Description de la régulation**

Si les machineries d'internalisation et de remaniement de l'ADN sont conservées au cours de l'évolution, les circuits de régulation de la compétence sont quant à eux très variables parmi les espèces transformables et représentent plutôt des adaptations au mode de vie de chaque espèce (Johnston *et al.*, 2014).

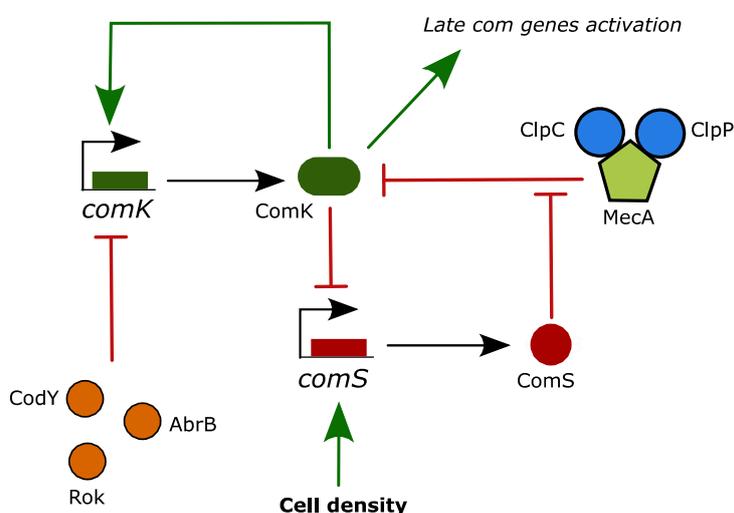
Cette régulation peut prendre place à différents moments lors de la croissance d'une population bactérienne selon l'espèce considérée : par exemple, en phase exponentielle de croissance chez *S. pneumoniae* ou en phase stationnaire de croissance chez *B. subtilis*. De plus, les régulateurs centraux qui contrôlent l'expression des effecteurs de la transformation peuvent être des facteurs sigma alternatifs liant l'ARN polymérase comme le facteur  $\sigma^x$  de *S. pneumoniae* (Lee et Morrison, 1999) retrouvé chez tous les streptocoques, ou des activateurs de transcription comme la protéine ComK chez *B. subtilis* (Mohan et Dubnau, 1990) ou encore cofacteurs de transcription (voir Johnston *et al.*, 2014 pour des exemples).

Outre la nature différente des régulateurs centraux de la compétence dans les espèces transformables, les cascades de régulation qui gouvernent l'activité de ces régulateurs sont également diverses. Nous restreindrons la description de ces cascades aux bactéries Gram-positives. Parmi ces bactéries, ces cascades ont été étudiées en détails chez *B. subtilis* et chez les streptocoques.

### **c. Chez *Bacillus subtilis***

Un exemple de facteur de transcription contrôlant le développement de la compétence est le facteur ComK de *B. subtilis* (Mohan et Dubnau, 1990). Chez cette bactérie, la compétence est induite en phase stationnaire de croissance lorsque ComK lie son propre promoteur et active sa propre expression, générant alors une boucle d'auto-activation. Cependant, à l'échelle d'une population, seulement 15% des cellules deviennent compétentes lors de la phase stationnaire de croissance alors que la population est homogène lors de la phase exponentielle de croissance. Ce phénomène est attribué à la stochasticité de l'expression de ComK, permettant la coexistence de deux états cellulaires différents dans la population (il y a bistabilité de la population) : une fraction de cellules, dépassant un seuil de concentration de ComK, peut entrer dans la boucle d'activation de ComK, alors que l'autre fraction de la

population ne dispose pas d'une quantité suffisante de ComK pour activer cette boucle (Maamar et Dubnau, 2005). Lorsque ComK a été synthétisé, il permet l'expression d'une centaine de gènes, dont ceux nécessaires à la transformation. Les gènes de la compétence sont directement régulés par la liaison de ComK à leurs sites promoteurs (Berka *et al.*, 2002).



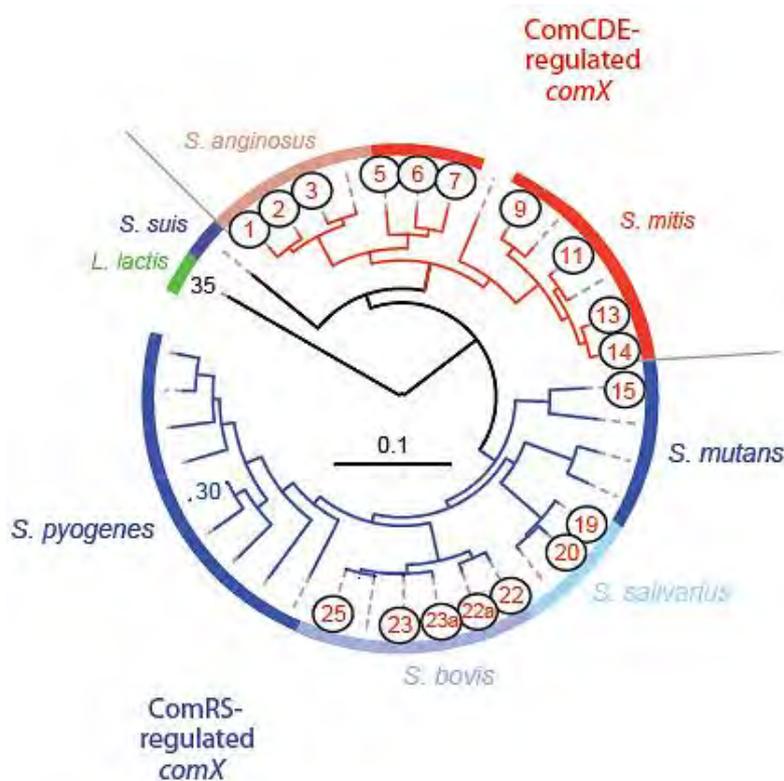
**Figure 9. Représentation schématique de la régulation de la compétence pour la transformation génétique chez *B. subtilis*.**

La quantité du régulateur central de compétence chez *B. subtilis*, ComK, est régulée au niveau transcriptionnel mais aussi au niveau post-traductionnel (**Figure 9**). Cinq protéines contrôlent la transcription de ComK : ComK lui-même et le régulateur de réponse DegU activent la transcription alors que les protéines AbrB, CodY et Rok inhibent la transcription de ComK (voir l'article de revue Claverys *et al.*, 2006). En phase exponentielle de croissance, la concentration de ComK est conservée à un niveau basal à cause de sa dégradation par le complexe protéique MecA-ClpC-ClpP. Grâce à une voie de signalisation, la concentration de la protéine ComS s'accroît avec la densité cellulaire. Le rôle de ComS est alors d'augmenter la stabilité de ComK en le protégeant du complexe de dégradation MecA-ClpC-ClpP (Turgay *et al.*, 1997). L'expression de *comS* dépend du régulateur de réponse ComA, induit par la voie de signalisation du *quorum-sensing*, mais qui est également un acteur du réseau régulant la sporulation.

#### d. Chez les streptocoques

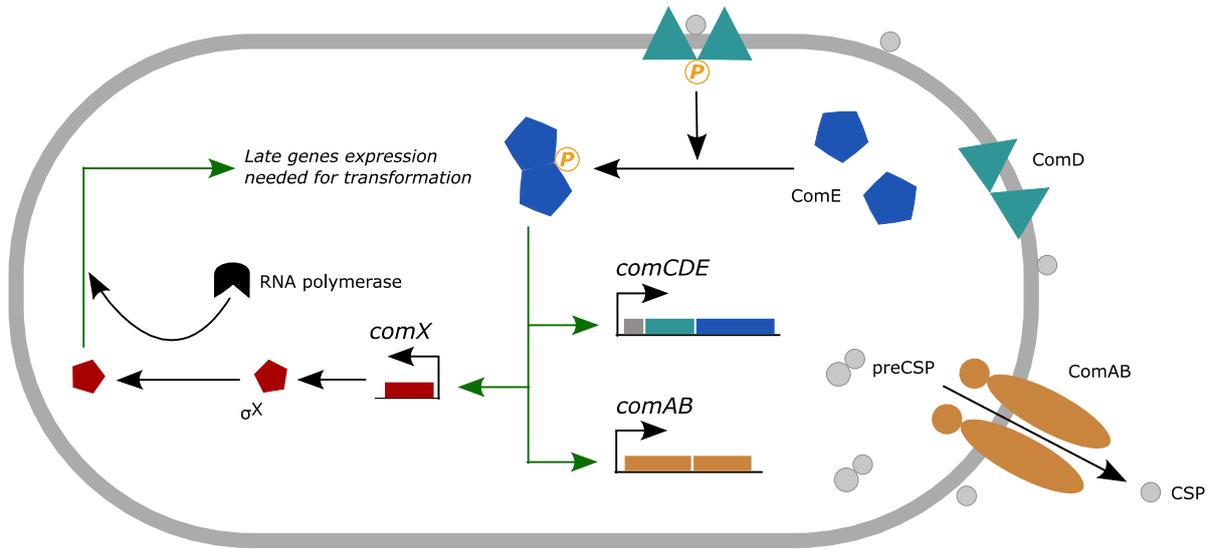
Deux cascades de régulation différentes du régulateur central de la compétence  $\sigma^x$  ont été identifiées chez les différentes espèces de streptocoques (**Figure 10**) (Martin *et al.*, 2006, Johnston *et al.*, 2014, Fontaine *et al.*, 2015). L'une dépend du système à deux composants

ComDE et est rencontrée dans les génomes des streptocoques appartenant aux groupes taxonomiques *S. mitis* et *S. anginosus* ; l'autre, plus récemment découverte, implique le système ComRS et est présente dans les espèces de streptocoques appartenant aux cinq autres groupes taxonomiques (*S. bovis*, *S. mutans*, *S. pyogenes*, *S. salivarius* et *S. suis*).



**Figure 10. Arbre phylogénétique des streptocoques** (d'après Johnston *et al.*, 2014). Les espèces régulées par le paradigme ComCDE sont colorées en rouge et les espèces régulées par le paradigme ComRS sont colorées en bleu.

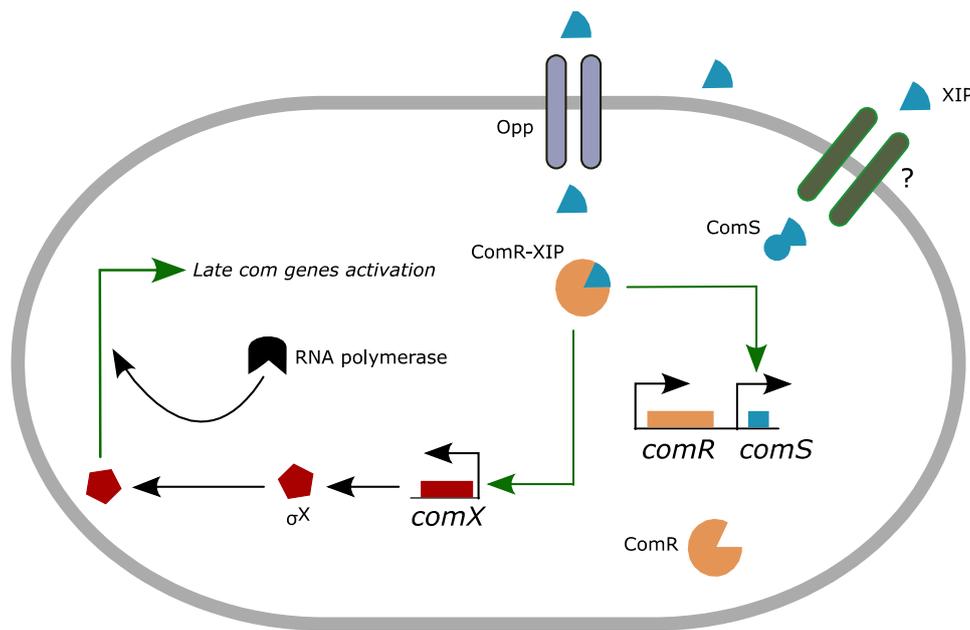
Néanmoins, chez toutes les espèces de streptocoques montrées comme étant naturellement transformables, l'entrée dans l'état de compétence est toujours dépendante de la synthèse du facteur sigma alternatif  $\sigma^x$ . Ainsi, dans ce genre, la compétence se développe en deux phases distinctes et successives dans le temps : l'expression des gènes dits précoces de la compétence composant essentiellement les circuits de régulation, suivie de l'expression des gènes dits tardifs de la compétence codant notamment pour les effecteurs de la transformation génétique.



**Figure 11. Représentation simplifiée de la régulation de la compétence chez *S. pneumoniae*.**

La cascade de régulation sous le contrôle du système à deux composants ComDE a été intensivement étudiée chez *S. pneumoniae*, appartenant au phylogroupe *S. mitis*. L'expression des gènes précoces de la compétence est induite par le CSP (Competence-Stimulating Peptide), codé par le gène *comC* (Håvarstein *et al.*, 1995), mûri et exporté vers le milieu extracellulaire par le transporteur ComAB (Håvarstein, *et al.*, 1991) (**Figure 11**). À l'extérieur de la cellule, le CSP active un système à deux composants (TCS) : le système ComDE (Pestova *et al.*, 1996). L'interaction entre le CSP et l'histidine kinase transmembranaire ComD permet la phosphorylation de ComD qui peut alors phosphoryler et activer le régulateur de réponse ComE (Martin *et al.*, 2013). ComE sous forme phosphorylée active en retour l'expression des opérons *comCDE* et *comAB* en liant les motifs répétés en tandems localisés dans la région promoteur de ces opérons (Martin *et al.*, 2013, Boudes *et al.*, 2014), formant alors une boucle de rétroaction positive amplifiant le signal et coordonnant l'ensemble de la population. Par ailleurs ComE, sous sa forme active, accroît l'expression du gène *comX* dont le produit  $\sigma^x$  permet l'expression des gènes nécessaires à la transformation génétique. Cela provoque une reprogrammation transcriptionnelle permettant l'entrée dans l'état de compétence, par ailleurs nommé « X-state ». Remarquons que les espèces de streptocoques diffèrent dans le nombre de gènes *comX* par souche : la plupart des souches possèdent deux copies identiques de *comX* alors que *S. anginosus* en possède trois. Durant la phase tardive de la compétence, le facteur de transcription  $\sigma^x$  s'associe de façon transitoire avec l'ARN polymérase, ce qui permet de cibler spécifiquement les motifs de liaison de  $\sigma^x$  à l'ADN, appelés aussi CIN-box, dont la séquence consensus chez *S. pneumoniae* est TACGAATA.

Le second circuit de régulation du facteur  $\sigma^x$  impliquant le système ComRS a été découvert beaucoup plus récemment chez *S. thermophilus* et *S. salivarius* (Gardan *et al.*, 2009, Fontaine *et al.*, 2010). Il implique le régulon ComRS (**Figure 12**). Le système est induit par l'expression du peptide ComS qui est ensuite clivé et exporté par un facteur encore inconnu. Le produit mature XIP (*comX*-induced peptide) s'accumule dans le milieu extracellulaire, puis ce peptide est réimporté dans le cytoplasme par un transporteur Opp (aussi appelé Ami), et interagit directement avec ComR (Fontaine *et al.*, 2010, Fontaine *et al.*, 2013) qui appartient à la famille des régulateurs transcriptionnels Rgg. Le complexe ComR-XIP permet alors l'expression de *comX* qui autorise l'entrée de la cellule dans l'état de compétence.



**Figure 12. Représentation schématique de la régulation de la compétence par le système ComRS** (modifié d'après Fontaine *et al.*, 2010).

Ainsi, jusqu'à présent, deux modes de signalisation intercellulaire gouvernant l'expression de  $\sigma^x$  ont été mis en évidence chez les streptocoques : le système ComCDE et le système ComRS. Ces deux processus diffèrent essentiellement par le mécanisme de détection du phéropéptide extracellulaire et par la voie de signalisation menant à l'activation de la transcription de *comX*. Deux cascades de régulation différentes ont donc évolué, les deux régulant l'entrée et la sortie de l'état de compétence au travers de la synthèse de  $\sigma^x$ .

Des modèles mathématiques explicatifs et prédictifs ont été développés pour comprendre les paramètres qui modulent l'entrée dans l'état de compétence pour ces différents organismes, venant ainsi compléter les résultats expérimentaux.

## 4. État de l'art des modèles développés de régulation de la compétence

### a. Chez *Bacillus subtilis*

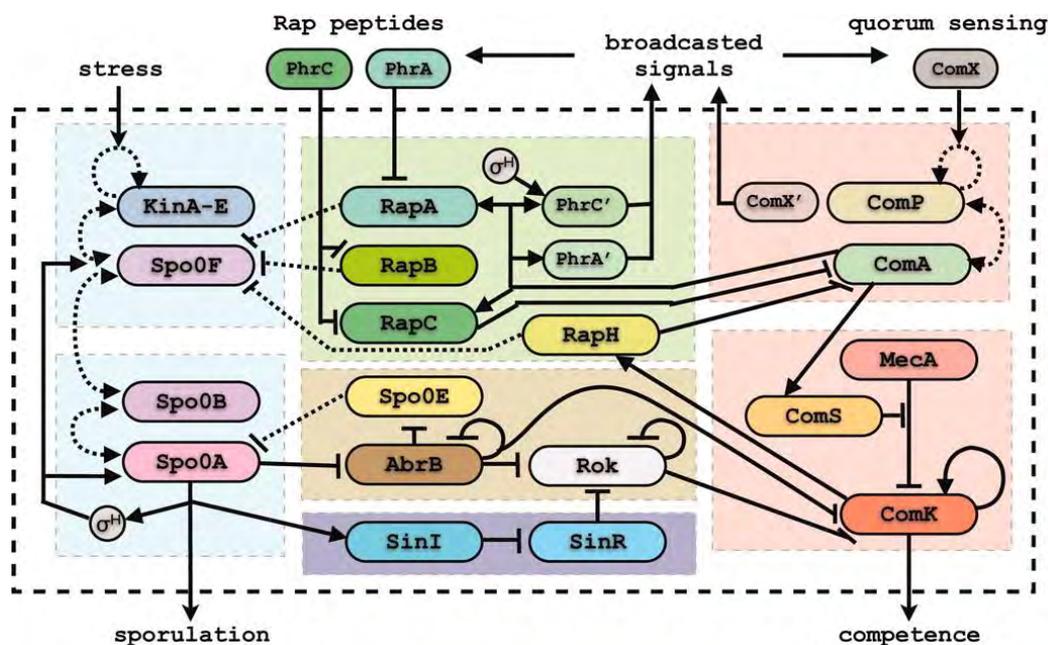
De nombreux aspects de la régulation de la compétence ont été analysés au travers de modèles mathématiques en utilisant *Bacillus subtilis* comme modèle d'étude biologique.

L'entrée dans l'état de compétence a été étudiée afin de déterminer sous quelles conditions la bistabilité du système pouvait survenir (Leisner *et al.*, 2009). Dans ce modèle, les auteurs ont privilégié l'analyse de l'entrée dans l'état de compétence et ont modélisé ComS comme un paramètre externe qui représente le signal de *quorum-sensing*, mais ils n'ont pas représenté le rétrocontrôle négatif de ComK sur *comS*, et n'ont donc pas étudié l'arrêt de l'état de compétence. Leurs résultats ont montré que pendant la phase exponentielle de croissance, lorsque la concentration de ComS est basse et que la dégradation de ComK est élevée, le système était monostable, ce qui signifie que la concentration de ComK n'était pas suffisante pour activer la compétence. En revanche, si la quantité de ComK augmente, par exemple grâce à une augmentation de ComS protégeant ainsi ComK de la dégradation par les protéases, le système peut entrer dans un autre état stable, autorisant la bistabilité du système comme réponse au bruit dans l'expression génique.

La bistabilité entre état végétatif et état de compétence a été étudiée en associant analyses expérimentales et simulations stochastiques (Süel *et al.*, 2006, Dubnau *et al.*, 2006, Maamar *et al.*, 2007). Les auteurs ont montré que des variations stochastiques de la transcription de *comK* étaient responsables de l'entrée dans l'état de compétence d'une fraction d'une population de cellules.

Le mécanisme d'entrée dans des programmes génétiques concurrents a également été modélisé chez *B. subtilis* (Schultz *et al.*, 2009) en représentant chaque processus (la compétence et la sporulation) sous forme modulaire. La sporulation survient lorsque les conditions environnementales, telle une carence en nutriments ou des dommages subis par l'ADN, deviennent défavorables à la croissance de la cellule. La sporulation se termine par la lyse de la « cellule mère » et la formation d'une cellule fille, la spore, pouvant résister à des dangers comme la chaleur, les radiations et les produits chimiques toxiques. La sporulation n'est pas initiée automatiquement lors d'une limitation des nutriments, mais elle est le résultat final d'une série d'étapes qui pourraient être décrites comme une suite de décisions cellulaires permettant de répondre au stress environnemental de façon adaptée (Stephens, 1998). Dans les travaux de Schultz et collaborateurs (2009), le module régulant la compétence (au travers de la dynamique de ComK) est modélisé grâce à un modèle stochastique et le module de la sporulation (au travers de la dynamique de SpoA, l'activateur central de la sporulation) est

modélisé au travers d'un minuteur dont l'avancement est ajusté par le niveau de stress impactant le système.



**Figure 13. Le réseau de transduction du signal de la sporulation et de la compétence** (d'après Schultz *et al.*, 2009). Le module de sporulation est présenté à gauche (bleu), et le module de compétence est présenté à droite (rouge).

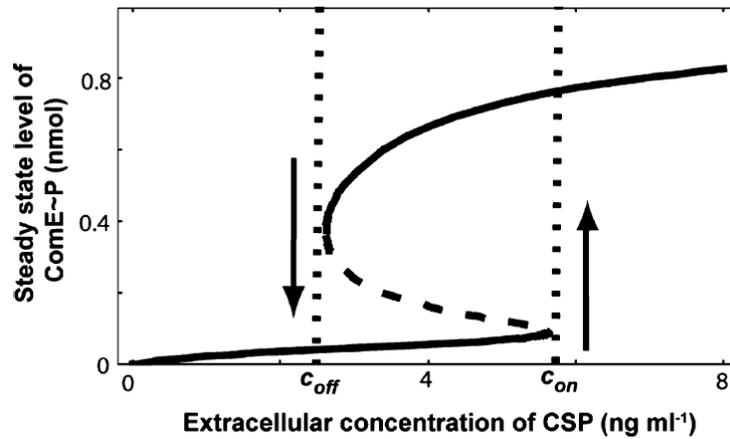
Chaque module est composé d'une unité de détection (KinA-E pour la sporulation et ComP pour la compétence) et d'une unité de régulation qui comprend un régulateur central (SpoA ou ComK) dont le fonctionnement est modulé par l'unité de détection (**Figure 13**). Les unités de détection des deux modules sont aussi déclenchées par le système Rap. L'interaction entre les deux modules est médiée par un module de décision, comprenant les protéines AbrB et Rok, qui permet une transition vers l'état de compétence pendant une fenêtre de temps où la sporulation est elle aussi possible. Le modèle développé a permis une meilleure compréhension de l'intégration des signaux extracellulaires menant au déclenchement de la sporulation ou de la compétence.

### b. Chez les streptocoques

Bien qu'un grand nombre de résultats expérimentaux soient disponibles sur la régulation de la compétence chez *Streptococcus pneumoniae*, une unique approche de modélisation a été entreprise à ce jour pour modéliser ce circuit de régulation (Karlsson *et al.*, 2006). Les auteurs y décrivent, au travers d'un modèle déterministe, le déclenchement de la compétence à l'échelle d'une population homogène, en représentant les mécanismes de régulation à l'échelle moléculaire.

Au travers du modèle, les auteurs ont proposé un mécanisme possible expliquant la fermeture rapide de l'état de compétence, car les mécanismes identifiés expérimentalement depuis (Mirouze *et al.*, 2013) étaient inconnus au moment de la publication. Pour cela, les auteurs ont effectué une analyse d'élasticité permettant d'évaluer l'impact des valeurs de différents paramètres du modèle sur les courbes de bifurcation. En effet, les systèmes d'équations différentielles paramétrées peuvent avoir différents comportements asymptotiques (tendre vers un équilibre, avoir un comportement oscillatoire, etc.) en fonction des valeurs de leurs paramètres. Il peut donc exister certaines valeurs de paramètres pour lesquelles le comportement du système passe d'un état qualitatif à un autre (par exemple, l'attracteur du système était un point équilibre et devient un cycle). Ce changement d'état qualitatif est une bifurcation et la valeur du paramètre associée est appelée valeur de bifurcation. L'analyse de bifurcations a donc pour objectif d'étudier l'influence des valeurs de paramètres sur les états stables d'un système. Cette analyse a conduit les auteurs à proposer qu'un répresseur, sous le contrôle de  $\sigma^x$ , peut être capable de lier les promoteurs des opérons *comCDE* et *comX*, inhibant alors leurs expressions. *A posteriori*, ce répresseur peut être considéré comme correspondant à la protéine DprA dont le gène est sous le contrôle de  $\sigma^x$ . Cependant, son action sur les promoteurs *comCDE* et *comX* est indirecte et a lieu via la séquestration de ComE~P par DprA.

Dans ce même travail, les auteurs font varier le paramètre du modèle assimilable à la densité cellulaire et montrent, à partir d'une courbe de bifurcation, que le système modélisé est bistable, avec deux états correspondant à l'état végétatif et à l'état de compétence. Ils exposent la nécessité de dépasser un seuil de densité cellulaire, assimilé à une concentration de CSP, pour que la population entre en état de compétence, ainsi que la nécessité de descendre sous un second seuil de densité cellulaire pour sortir de l'état de compétence (**Figure 14**). Les auteurs ont ainsi développé un modèle de régulation dans une population homogène en se basant sur l'hypothèse que l'initiation de la compétence était contrôlée par un système de détection du quorum (ComABCDE), ce qui signifie qu'une densité cellulaire suffisante doit être atteinte par la population pour pouvoir détecter le CSP diffusable et pour que celle-ci puisse entrer en masse dans l'état de compétence.



**Figure 14. État d'équilibre de ComE~P en fonction de la concentration extracellulaire de CSP** (d'après Karlsson *et al.*, 2006). Le paramètre CSP permet un basculement entre les deux états stables (lignes continues). À faible concentration de CSP, ComE~P a un état d'équilibre bas. Lorsque le CSP a atteint une valeur seuil (point  $c_{on}$  sur le diagramme), l'état d'équilibre de ComE~P passe à une valeur plus élevée. La valeur d'équilibre de ComE~P ne retournera à l'état stationnaire initial que si la concentration de CSP diminue suffisamment et passe un autre point de commutation (point  $c_{off}$  dans le diagramme). La ligne pointillée indique des états stationnaires instables.

Cependant, plusieurs observations sont contraires à l'idée que la concentration de CSP dans le milieu de croissance serait proportionnelle à la densité cellulaire. En effet, lorsque les cellules de *S. pneumoniae* de type sauvage sont déplacées vers un milieu plus alcalin, la compétence se développe à un temps fixe après inoculation quelle que soit la taille de l'inoculum (Claverys *et al.*, 2006), démontrant que le processus peut se produire indépendamment de la densité cellulaire. De plus, des résultats très récents (Prudhomme *et al.*, 2016) montrent que la transition vers l'état de compétence repose sur un mécanisme dépendant du temps de croissance qui permet alors l'émergence d'une sous-population de cellules initiateuses de la compétence. La propagation de la compétence à l'ensemble de la population est ensuite réalisée par contact intercellulaire puisque le CSP est retenu à la surface des cellules (Prudhomme *et al.*, 2016).

Nous avons choisi *S. pneumoniae* comme organisme modèle car, parmi les streptocoques, il s'agit de l'organisme dont la régulation de l'état de compétence a été la plus largement étudiée : beaucoup de résultats expérimentaux ont donc été accumulés sur sa régulation. En effet, comme le système de régulation ComRS a été découvert plus récemment (Gardan *et al.*, 2009, Fontaine *et al.*, 2010), les connaissances disponibles sur cette cascade de régulation sont moins abondantes et de nombreuses questions restent encore ouvertes, comme celle de l'arrêt de la boucle autocatalytique d'activation de la compétence. Cependant, deux modèles mathématiques du paradigme ComRS, chez *S. mutans* (Son *et al.*, 2012) et chez *S.*

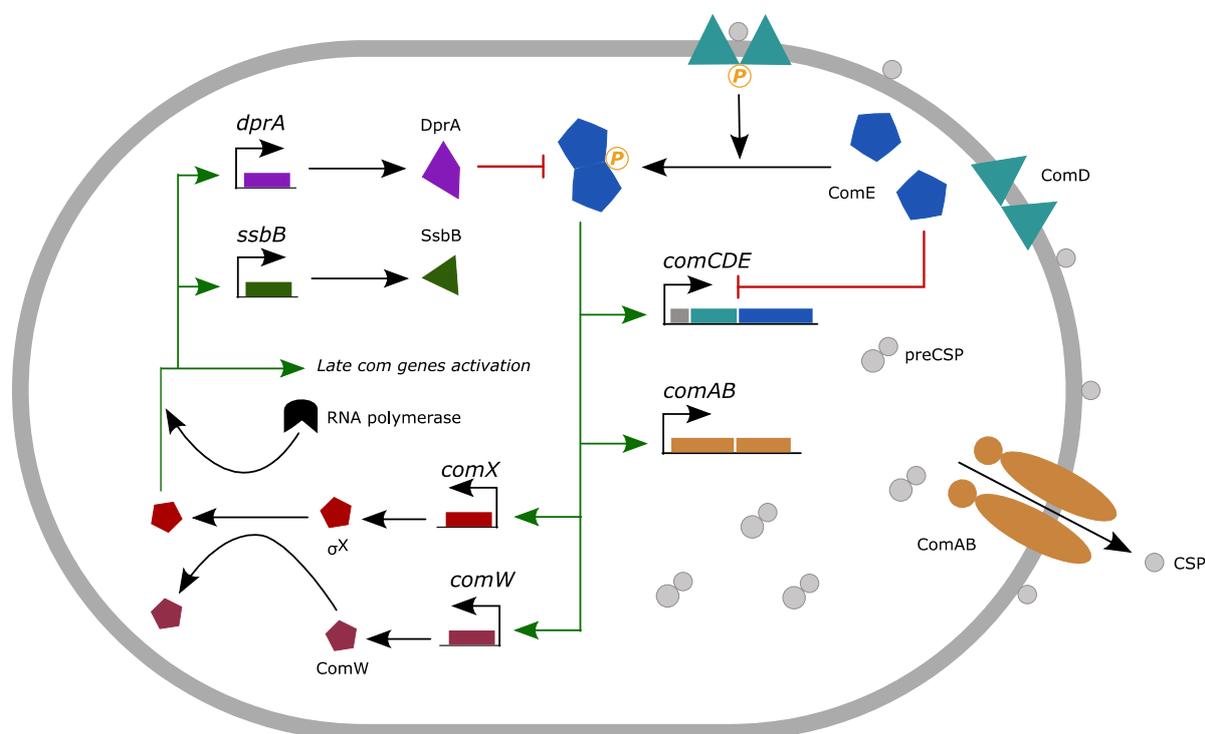
*thermophilus* (Haustenne *et al.*, 2015) ont été publiés pour aider à répondre à certaines interrogations : les deux modèles ont été établis à l'échelle de la population afin d'étudier notamment les composantes du système ComRS critiques pour la production de ComX ; en combinant à la fois approche théorique et expérimentale, les auteurs démontrent par exemple que la concentration de ComR est le facteur limitant majeur chez *S. suis* pour entrer en état de compétence (Haustenne *et al.*, 2015).

Comme les hypothèses antérieures utilisées pour développer le modèle mathématique précédent de la régulation des compétences chez *S. pneumoniae* sont contestées par de nouvelles données expérimentales, nous abordons la modélisation dynamique de la régulation de la compétence à l'échelle cellulaire qui intégrera les nouvelles connaissances acquises et notamment celles concernant les mécanismes de fermeture de l'état de compétence. Pour cela, *S. pneumoniae* est un organisme de choix car, dans les expériences où la compétence est induite par ajout de CSP exogène, la majorité des cellules dans la population déclenche alors la compétence de manière synchrone. Par conséquent, les mesures quantitatives effectuées à l'échelle d'une population de cellules synchrones permettent d'acquérir des valeurs moyennes d'expression génique au niveau des cellules individuelles, en appliquant des procédures de normalisation appropriées (de Jong *et al.*, 2010, Stefan *et al.*, 2015).

Nous avons adopté une approche en deux étapes. Une modélisation graphique, discrète et qualitative, au travers des réseaux de Petri, nous a préalablement permis de visualiser la topologie du réseau de régulation et d'en vérifier la cohérence structurale (partie II). Le choix du paradigme de modélisation a été guidé par la nature des données qui étaient à notre disposition pour concevoir le modèle. La disponibilité des données quantitatives obtenues à partir de mesures d'expression de gènes par fusion transcriptionnelle a porté notre choix vers l'utilisation d'un formalisme quantitatif tel que les équations différentielles ordinaires (partie III). Le modèle nous permettra alors de vérifier si tous les éléments du système ont été pris en compte, après avoir comparé les résultats des simulations aux données expérimentales. Le modèle sera ensuite utilisé pour explorer les comportements du système dans plusieurs conditions, chez différents mutants génétiques, ou pour étudier l'influence de paramètres particuliers.

## II. Construction et analyse statique du réseau de régulation ComCDE

Un modèle étant une représentation abstraite de la réalité, il ne peut pas toujours expliquer tous les détails d'un système biologique. Mais il peut aider à comprendre les relations structurales entre les entités ainsi que le comportement dynamique du système au cours du temps. Il a donc été nécessaire de créer un modèle dont le comportement des simulations reste cohérent avec les résultats expérimentaux acquis, tout en restant aussi simple que possible. Ci-dessous sont données en détail les connaissances sur le réseau de régulation afin de motiver les choix de modélisations effectuées, ainsi que les hypothèses posées pour simplifier le modèle. Une vue globale du réseau est présentée en **Figure 15** et les réactions ont été résumées dans le **Tableau 1**.



**Figure 15. Vue d'ensemble du réseau de régulation contrôlant l'état de compétence chez *S. pneumoniae*.** La compétence est induite par le CSP. Le pré-CSP est exporté et mûri par le transporteur ComAB. Le CSP extracellulaire peut alors lier l'histidine kinase ComD qui s'auto-phosphoryle et transfère le groupe phosphoryl au régulateur de réponse ComE. Ensuite, le dimère ComE~P active la transcription des gènes précoces de la compétence, dont *comCDE* et *comAB*, générant une boucle de régulation positive. Les gènes précoces de la compétence incluent aussi *comX* et *comW*. *comX* code pour le facteur  $\sigma^X$  qui nécessite ComW pour lier l'ARN polymérase.  $\sigma^X$  contrôle l'activation d'un grand ensemble de gènes tardifs qui sont représentés ici par *ssbB*, couramment utilisé expérimentalement comme rapporteur de l'expression des gènes tardifs, et par *dprA* qui est impliqué dans la fermeture de l'état de compétence en séquestrant ComE~P. La fermeture de la compétence implique également ComE qui inhibe la transcription *comCDE* en rentrant en compétition avec ComE~P pour lier  $P_{comC}$ . Les flèches vertes représentent les réactions d'activation, les lignes rouges représentent les réactions d'inhibition et les flèches noires représentent les autres réactions (synthèse, liaison ou d'export).

## 1. Simplifications réalisées pour concevoir le modèle

D'abord, pour conserver un modèle simple, les deux étapes du processus d'expression de gènes ne sont pas explicitement dissociées, *i.e.* la transcription et la traduction. Nous considérons ce processus comme une seule étape et le réseau a donc été directement représenté à l'échelle protéique. Cette simplification est basée sur l'hypothèse que la concentration d'ARNm est à l'état de quasi équilibre, dans le sens où la concentration des ARNm est supposée s'adapter de façon quasi instantanée en réponse au changement de l'activité des promoteurs. Ainsi, il est possible de négliger la variation de concentration d'ARNm et d'écrire directement la variation de la quantité d'une protéine en fonction de l'activité d'un promoteur. Cette simplification est aussi connue comme approximation de quasi-équilibre. Lorsque les processus lents dominent un système, les réactions rapides sont supposées être en continu à l'état de quasi-équilibre (Chen *et al.*, 2010).

D'autres simplifications, propres au modèle, ont été faites pour réduire le nombre d'acteurs du réseau : i) la dégradation du CSP par la protéase membranaire HtrA (Cassone *et al.*, 2012) n'a pas été explicitement modélisée mais est supposée incluse dans la constante de dégradation du CSP ; ii) nous n'avons pas non plus explicitement inclus dans le réseau l'ATP et l'ADP, cofacteurs des réactions de phosphorylation. Il est supposé que la concentration de l'ATP n'est pas un facteur limitant ; iii) parmi les produits des gènes tardifs de la compétence, *i.e.* le produit des gènes sous le contrôle de ComX, nous avons pris en compte seulement DprA dont l'action dans la fermeture de l'état de compétence a été récemment décrite et SsbB dont la cinétique est couramment utilisée expérimentalement comme rapporteur de l'état de compétence ; iv) étant un facteur sigma, ComX agit sur la régulation de ses gènes cibles en formant un complexe avec l'ARN polymérase. Cependant, comme seulement l'expression de ComX est régulée par le réseau de la compétence, nous n'avons pas inclus explicitement l'ARN polymérase dans le modèle et nous avons considéré que seule la concentration de ComX pouvait affecter l'expression des gènes tardifs de la compétence ; v) il n'existe aucune donnée quantitative sur les constantes d'interactions protéiques-protéiques, l'interaction DprA-ComE~P ou DprA-ComE. Il est seulement connu que DprA a plus d'affinité pour ComE~P que pour ComE (Mirouze *et al.*, 2013) : donc, afin de conserver un réseau simple, la liaison DprA-ComE a été ignorée.

## 2. Initiation de l'état de compétence : le système à deux composantes ComD/ComE

Lorsque la concentration seuil de CSP est atteinte, le CSP active le système à deux composantes ComD/ComE (**Figure 15**). Cependant, le mécanisme de transduction du signal

de ce système à deux composantes n'est pas complètement élucidé. Comme ce système appartient à la famille AgrA/AlgR/LytR (Lange *et al.*, 1999), il est apparu intéressant de transposer les propriétés acquises sur le système AgrC/AgrA de *Staphylococcus aureus* au système ComD/ComE. Il a été montré que l'histidine kinase AgrC formait un dimère avant son autophosphorylation (George Cisar *et al.*, 2009). Ce dimère possède une activité d'autophosphorylation basale qui est indépendante de la liaison du ligand. Cette propriété a été aussi observée pour ComD (Martin *et al.*, 2010). De plus, des résultats ont aussi suggéré que le dimère AgrC possède deux sites de liaison indépendants pour le ligand (Wang *et al.*, 2014). Ainsi, par analogie, nous avons modélisé le processus de phosphorylation de ComD de la même manière : formation d'un dimère de ComD (**Tableau 1**, réaction (2) et réaction inverse (3)) dont l'autophosphorylation peut être indépendante de la liaison du CSP (**Tableau 1**, réaction (4)) ou activée par l'interaction avec le CSP (**Tableau 1**, réaction (5)).

**Tableau 1. Description des réactions contrôlant la régulation de compétence.**

Description	Réaction	Nom de la transition dans le réseau de Petri
<i>Maturation et export du pré-CSP par l'ABC transporteur ComAB</i>	pre-CSP + ComAB $\rightarrow$ CSP + ComAB <b>(1)</b>	CSP_export
<i>Dimérisation d l'histidine kinase ComD</i>	2 ComD $\rightarrow$ (ComD) <sub>D</sub> <b>(2)</b>	dimerization
<i>Dissociation de (ComD)<sub>D</sub></i>	(ComD) <sub>D</sub> $\rightarrow$ 2 ComD <b>(3)</b>	monomerization
<i>Auto activation de (ComD)<sub>D</sub></i>	(ComD) <sub>D</sub> $\rightarrow$ (ComD <sub>act</sub> ) <sub>D</sub> <b>(4)</b>	auto_phospho
<i>La liaison du CSP à (ComD)<sub>D</sub> active au travers de l'autophosphorylation</i>	2 CSP + (ComD) <sub>D</sub> + phosphates $\rightarrow$ (ComD~P-CSP) <sub>D</sub> = (ComD <sub>act</sub> ) <sub>D</sub> <b>(5)</b>	ComD_phosphorylation
<i>Activation de ComE dimérique par transfert du groupe phosphate de (ComD<sub>act</sub>)<sub>D</sub></i>	(ComD <sub>act</sub> ) <sub>D</sub> + 2 ComE $\rightarrow$ (ComE~P) <sub>D</sub> + (ComD-CSP) <sub>D</sub> <b>(6)</b>	transphosphorylation
<i>Déphosphorylation spontanée de (ComE~P)<sub>D</sub></i>	(ComE~P) <sub>D</sub> $\rightarrow$ 2 ComE + phosphates <b>(7)</b>	ComEP_dephosphorylation
<i>Synthèse des gènes précoces de la compétence par (ComE~P)<sub>D</sub></i>	(ComE~P) <sub>D</sub> $\rightarrow$ ComE + ComD + ComAB + pre-CSP + ComW + ComX <sub>ina</sub> + (ComE~P) <sub>D</sub> <b>(8)</b>	ComD_synth; ComE_synth; ComAB_synth; preCSP_synth; ComW_synth; ComX_synth
<i>Activation de ComX</i>	ComX <sub>ina</sub> + ComW $\rightarrow$ ComX <sub>act</sub> + ComW <b>(9)</b>	ComX_activation
<i>Synthèse des gènes tardifs de la compétence grâce au facteur ComX sous sa forme active</i>	ComX <sub>act</sub> $\rightarrow$ ComX <sub>ina</sub> + DprA + SsbB <b>(10)</b>	DprA_synth; SsbB_synth
<i>Inactivation de ComE~P par séquestration dans un complexe avec DprA</i>	2 DprA + (ComE~P) <sub>D</sub> $\rightarrow$ (DprA-ComE~P) <sub>D</sub> <b>(11)</b>	DprA_ComEP_binding
<i>Dégradation des composés du réseau</i>	pre-CSP $\rightarrow$ $\emptyset$ ; ComAB $\rightarrow$ $\emptyset$ ; CSP $\rightarrow$ $\emptyset$ ; ComD $\rightarrow$ $\emptyset$ ; (ComD-CSP) <sub>D</sub> $\rightarrow$ $\emptyset$ ; ComE $\rightarrow$ $\emptyset$ ; ComX <sub>act</sub> $\rightarrow$ $\emptyset$ ; ComX <sub>ina</sub> $\rightarrow$ $\emptyset$ ; DprA $\rightarrow$ $\emptyset$ ; (DprA-ComE~P) <sub>D</sub> $\rightarrow$ $\emptyset$ SsbB $\rightarrow$ $\emptyset$ ; ComY $\rightarrow$ $\emptyset$ or ComZ $\rightarrow$ $\emptyset^{\&}$ <b>(12)</b>	preCSP_decay; ComAB_decay; CSP_decay; ComD_decay; ComD_consumed_decay; ComE_decay; ComX_ina_decay; ComXact_decay; ComW_decay; DprA_decay; SsbB_decay

Malgré plusieurs tentatives *in vivo*, la transphosphorylation de ComE n'a pas encore pu être directement mise en évidence expérimentalement (Martin *et al.*, 2013). Cependant, il a démontré que le mutant phospho-mimétique ComE<sup>D58E</sup> était majoritairement à l'état dimérique en solution, alors que le monomère ComE existait principalement sous la forme monomérique (Martin *et al.*, 2013). De plus, les données structurales réalisées sur les complexes ComD/ComE d'une part, et ComE/*comCDE* d'autre part, ont laissé présumer que le transfert du groupe phosphoryle de l'histidine kinase à son régulateur de réponse ComE pouvait permettre la dimérisation de ComE au travers de la liaison d'un dimère ComD phosphorylé à deux monomères de ComE (Sanchez *et al.*, 2015). Ainsi, il a été supposé dans le modèle qu'un dimère de ComD~P puisse, par transphosphorylation, activer un dimère de ComE (**Tableau 1**, réaction (6)). Suite à la transphosphorylation, le CSP lié à un dimère de ComD ((ComD-CSP)<sub>D</sub>) peut entrer dans un cycle de phosphorylation-transphosphorylation et ainsi être capable d'activer plusieurs molécules de ComE, ou bien devenir inactif. L'observation du fait que le nombre de molécules de ComD augmente de 1500 par cellule avant l'induction de la compétence à 39 000 par cellule compétence (Martin *et al.*, 2013) ainsi que les récents résultats expérimentaux montrant que ComD pouvait retenir le CSP à la surface des cellules (Prudhomme *et al.*, 2016) ont permis de privilégier la seconde hypothèse. Il a donc été supposé que lorsqu'un dimère de ComD a phosphorylé un dimère de ComE suite à la liaison du CSP, le dimère de ComD devient inactif et ne peut être réutilisé. Le devenir de la forme inactive du complexe CSP-ComD est ensuite simplement modélisé grâce à une réaction de dégradation (**Tableau 1**, réaction (12)).

La déphosphorylation de ComE~P n'a pas encore été documentée. Actuellement, aucune activité de phosphatase impliquant l'histidine kinase ComD n'a été décrite (alors que certains systèmes à deux composantes l'ont parfois été), ni aucune autre protéine pouvant remplir cette fonction. De plus, aucune activité phosphatase de l'homologue histidine kinase AgrC de *S. aureus* sur son régulateur de réponse AgrA n'a été détectée : il a été conclu que la diminution du niveau de phosphorylation de AgrA était due à son activité intrinsèque de déphosphorylation (Wang *et al.*, 2014). Il a donc été supposé dans le modèle que l'état phosphorylé de ComE était extrêmement labile et que ComE~P pouvait s'auto-déphosphoryler spontanément (**Tableau 1**, réaction (7)).

Les vitesses de dégradation de (ComD~P-CSP)<sub>D</sub> et (ComE~P) étant très faibles en regard de la stabilité de leurs formes phosphorylées, les réactions de dégradation n'ont pas été ajoutées dans le modèle, les composants étant consommés, respectivement, seulement par les réactions de transphosphorylation et de déphosphorylation.

### 3. Le régulateur central ComX

ComX est un facteur sigma responsable de l'activation des gènes tardifs de la compétence. Il s'agit d'une protéine qui apparaît de façon transitoire environ 5 minutes après l'addition de CSP, atteint un pic à environ 15 minutes, et disparaît ensuite avec une demi-vie approximative de 5 minutes (Luo et Morrison, 2003). Il a été montré que ComW, le produit d'un autre gène précoce de la compétence (Luo *et al.*, 2004), est nécessaire à la fois à l'accumulation de ComX (Sung et Morrison, 2005) en antagonisant l'action des protéase ClpE/ClpP (Piotrowski *et al.*, 2009) et à l'activation de ComX (Sung et Morrison, 2005). ComW pourrait améliorer la liaison de ComX au cœur de l'ARN polymérase (Tovpeko et Morisson, 2014, Tovpeko *et al.*, 2016). Cependant, les mécanismes d'action de ComW ne sont pas connus. La temporalité de ComW est proche de celle de ComX (d'abord détectable 5 minutes après l'ajout de CSP, atteignant un pic à 15 minutes, et presque complètement disparu après 30 minutes) (Piotrowski *et al.*, 2009). Dans un premier temps, il a été présumé que la compétence et la transcription des gènes tardifs pouvaient se terminer à cause de la disparition de ComX. Cependant, des observations ont montré que les cellules *clpP*, où les protéines ComX et ComW sont stabilisées, quittaient l'état de compétence aussi rapidement que des cellules sauvages, suggérant qu'un autre mécanisme pouvait être impliqué dans la terminaison de transcription des gènes tardifs (Piotrowski *et al.*, 2009). Pour intégrer ces connaissances dans notre modèle, nous découplons les deux rôles de ComW. La stabilisation ComX a été incluse dans le paramètre de demi-vie de ComX pour éviter l'incorporation de la protéase ClpE-ClpP dans notre modèle. L'activation de ComX par ComW a été prise en compte en créant deux formes de ComX : une forme inactive et une forme active. ComX est d'abord synthétisé sous forme inactive qui devient active sous l'action de ComW (**Tableau 1**, réactions (8)-(9)).

### 4. L'arrêt de l'état de compétence

Il a été récemment avancé que l'arrêt de l'expression du promoteur *comCDE* ( $P_{comC}$ ) était intrinsèque à ComD/ComE car la diminution de transcription de *comCDE* survient en l'absence de synthèse de tout gène tardif (Martin *et al.*, 2013, Mirouze *et al.*, 2013). À partir de l'observation que la forme non phosphorylée de ComE liait de façon forte  $P_{comC}$  *in vitro* et que sa surexpression empêchait la compétence spontanée, il a été exposé que l'accumulation de ComE en réponse au CSP entraînait une compétition entre ComE~P et ComE pour lier  $P_{comC}$ , prévenant ainsi la transcription de ce dernier (Martin *et al.*, 2013). Cependant, il n'existe quasi aucune antagonisation de la forme non phosphorylée de ComE sur le promoteur du gène *comX* ( $P_{comX}$ ) car l'affinité de ComE pour ce promoteur est très faible (Martin *et al.*, 2013), ce qui suggère que ComE est peu efficace pour inhiber l'expression de  $P_{comX}$ . En

réalité, la fermeture de l'expression de  $P_{\text{comX}}$  nécessite l'action de DprA, un gène tardif de la compétence (Mirouze *et al.*, 2013, Weng *et al.*, 2013). Des expériences de doubles hybrides ont montré une interaction très forte entre DprA et ComE~P, suggérant une interaction physique directe entre les deux protéines mais aussi une interaction plus faible avec ComE (Mirouze *et al.*, 2013, Weng *et al.*, 2013). Il a aussi été montré que la dimérisation de DprA était nécessaire à l'arrêt de l'expression des gènes précoces de la compétence (Mirouze *et al.*, 2013). Ainsi, l'effet de la protéine DprA serait de changer le ratio ComE/ComE~P en faveur de la répression des promoteurs de gènes précoces. Comme le mécanisme d'action de DprA sur ce ratio n'a pas encore été élucidé, deux hypothèses ont été proposées : DprA pourrait former un complexe avec ComE~P pour bloquer l'action de ComE~P en séquestrant ce dernier, ou pour permettre la déphosphorylation de ComE~P. Il a été choisi d'implémenter dans le réseau le scénario de séquestration, bien que l'interaction DprA-ComE ait été montrée, mais aucune évidence expérimentale n'a encore permis de mettre en évidence le couplage d'une réaction de déphosphorylation à cette réaction (**Tableau 1**, réaction (11)). Comme DprA et ComE partagent des demi-vies proches (Martin *et al.*, 2013, Mirouze *et al.*, 2013), l'on a considéré que la réaction issue de ce complexe est sa dégradation (**Tableau 1**, réaction (12)).

## 5. Construction du réseau de Petri

Le modèle à l'échelle cellulaire a d'abord été construit en utilisant le formalisme des réseaux de Petri. Ce formalisme offre une représentation graphique simple de la structure du réseau, ce qui facilite les échanges entre les experts du domaine biologique représenté et les modélisateurs pendant le processus de construction du réseau. De plus, ce formalisme permet de tester, qualitativement, la validité de la structure du réseau étudié.



La compétition entre ComE et  $(ComE\sim P)_D$  pour la liaison à  $P_{comC}$  et  $P_{comAB}$  est à l'origine de l'arrêt de l'état de compétence (Martin *et al.*, 2013). La dissociation de  $(ComE\sim P)_D$  des promoteurs permet de ramener le promoteur à l'état libre (transitions *dissociation\_PcomC\_ComEP* ou *dissociation\_PcomC\_ComAB*) ; elle est suivie de la liaison de ComE qui permet de transformer les promoteurs en état inactif ( $P_{comC\_ina}$ ,  $P_{comAB\_ina}$ ) et d'inhiber ainsi l'expression des promoteurs  $P_{comC}$  et  $P_{comAB}$  (transitions *association\_PcomC\_ComE*, *association\_PcomAB\_ComE*). Comme ComE est un inhibiteur compétitif pour les deux réactions, les réactions réverses ont aussi été modélisées : en partant d'une forme inactive de  $P_{comC}$  ( $P_{comAB}$ ), la dissociation de ComE génère un promoteur à l'état libre de tout régulateur, capable à nouveau de lier ComE ou  $(ComE\sim P)_D$ .

Dans les cas de la régulation de *comX* et de *comW*, comme l'effet inhibiteur de ComE n'est pas modélisé, le modèle a été simplifié en omettant la représentation du promoteur. Ainsi,  $(ComE\sim P)_D$  active directement la synthèse de ComX et de ComW.

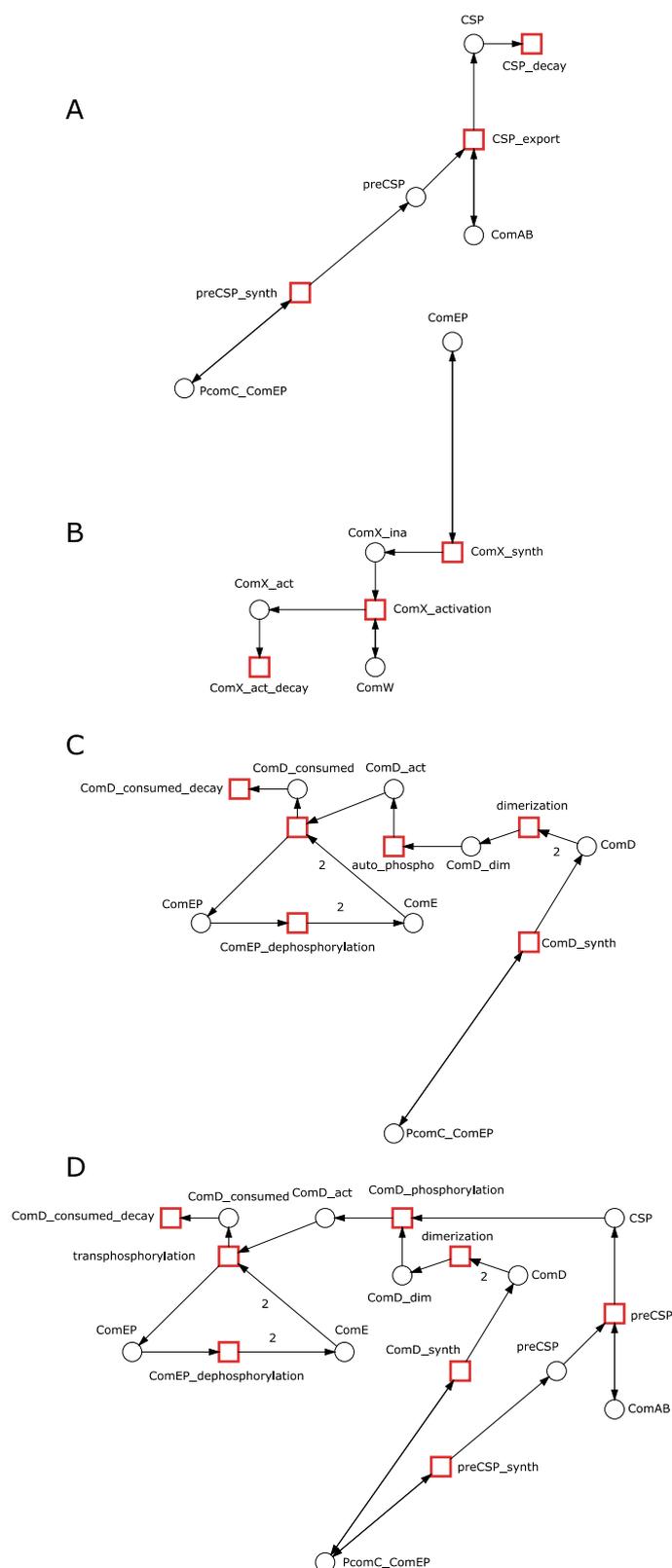
La plateforme de modélisation Snoopy (Marwan *et al.*, 2012) a été utilisée pour implémenter le réseau.

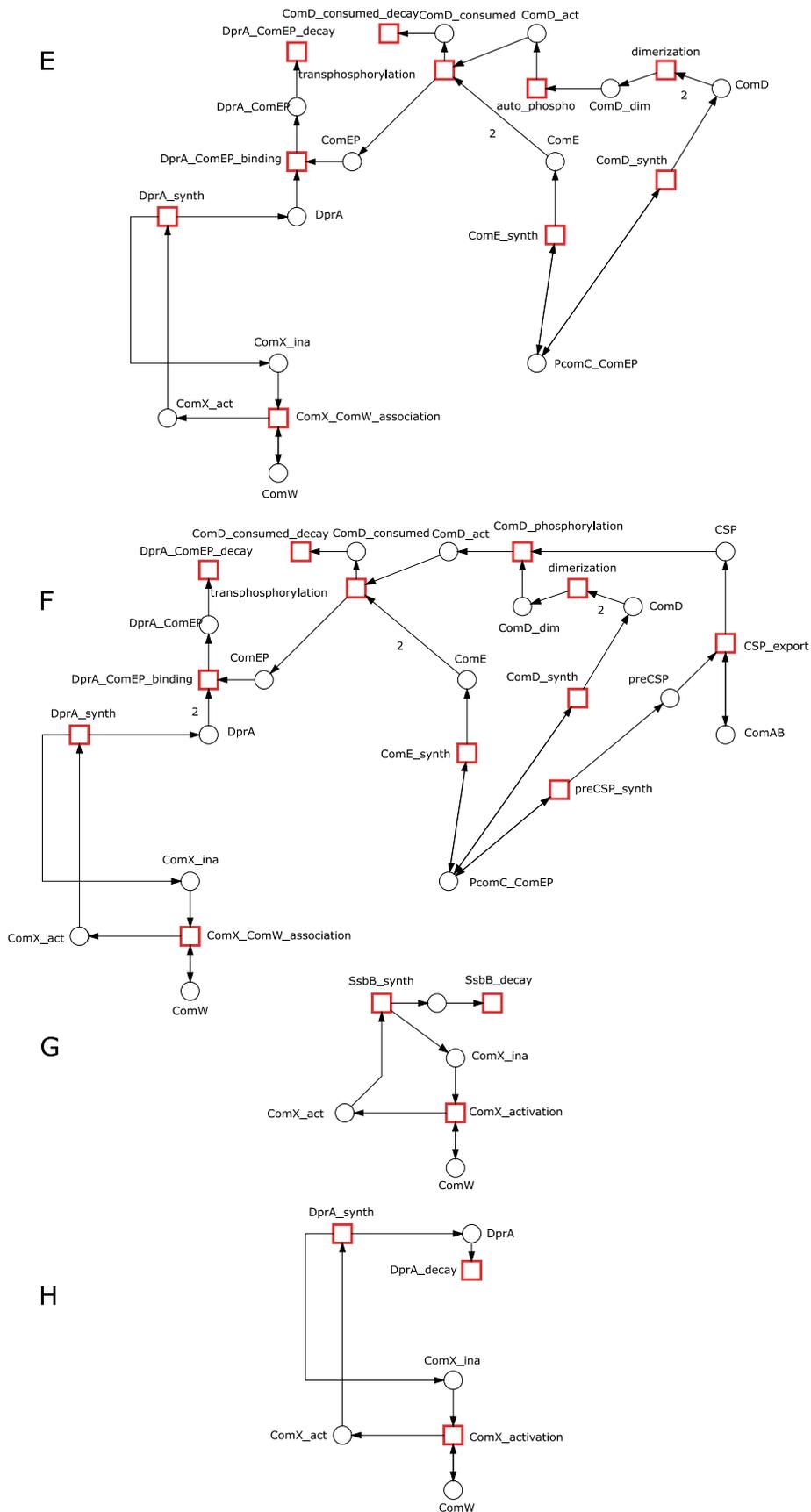
## 6. Analyse des invariants structuraux

Avant d'étudier la dynamique du réseau en utilisant les équations différentielles ordinaires, nous effectuons ici, préalablement, une validation qualitative du modèle qui dépend seulement de la structure du graphe et qui ne nécessite aucun marquage initial du réseau de Petri. Les invariants structuraux qui montrent la cohérence structurale du réseau ont été calculés. Les invariants des réseaux de Petri peuvent être divisés en deux classes : les invariants de places (P-invariants) et les invariants de transitions (T-invariants). Les P-invariants correspondent à la conservation de matière dans le réseau. Les T-invariants décrivent une séquence de transition permettant de ramener le système à son état initial. En partant d'un état donné du système et si toutes les transitions composant un T-invariant sont activées, cette suite de transitions permettra au système de retourner à son état initial. Pour obtenir les invariants, il est nécessaire de calculer la matrice d'incidence. Le logiciel Charlie (Heiner *et al.*, 2015) a été utilisé pour calculer les P-invariants et les T-invariants minimaux.

19 T-invariants minimaux ont été obtenus parmi lesquels 8 sont des T-invariants non triviaux. Les T-invariants triviaux recouvrent 6 invariants composés chacun par deux réactions – une synthèse et une dégradation (des produits des gènes précoces pré-CSP, ComD, ComE, ComAB, ComW et  $ComX_{inact}$ ) –, et 5 invariants composés de deux transitions représentant les

réactions réversibles (dimérisation de ComD et monomérisation du dimère de ComD, association et dissociation de ComE et de (ComE~P)<sub>D</sub> des promoteurs P<sub>comC</sub> et P<sub>comAB</sub>).





**Figure 17. Représentation des T-invariants non-triviaux extraits du réseau de Petri.** Tous les T-invariants non triviaux ont été extraits de la figure 16. Les transitions formant chaque T-invariant sont colorées en rouge. Comme pour la figure 16, les coefficients stœchiométriques de réaction ne sont indiqués que lorsqu'ils diffèrent de 1.

Les interprétations biologiques des 8 T-invariants sont les suivantes :

- Le pré-CSP est synthétisé, mûri, exporté et, enfin, le CSP est dégradé (**Figure 17A**).
- ComX est synthétisé dans sa forme inactive, puis est activé suite à son interaction avec ComW ; puis la forme active est dégradée (**Figure 17B**).
- ComD est synthétisé, se dimérise puis s'auto-phosphoryle spontanément. Le dimère phosphorylé de ComD phosphoryle ensuite le régulateur de réponse ComE puis est ensuite dégradé. Enfin, grâce à son auto-déphosphorylation, ComE~P retourne sous sa forme initiale ComE (**Figure 17C**).
- La seule différence avec la description précédente est que la phosphorylation du dimère de ComD nécessite la liaison du CSP (**Figure 17D**). Ainsi le pré-CSP doit être synthétisé, mûri puis exporté. La transition *auto\_phospho* est remplacée par la transition *ComD\_phosphorylation*.
- ComE~P est produit par le transfert d'un groupe phosphate à ComE à partir d'un dimère de ComD qui s'auto-phosphoryle. Le dimère de ComD consommé est dégradé. ComE~P est séquestré par DprA qui est synthétisé après la transition de ComX de sa forme inactive à sa forme active, suite à son interaction avec ComW. Le complexe DprA-ComE~P est ensuite dégradé (**Figure 17E**).
- Ce T-invariant est composé de la même suite de réactions que le précédent, sauf que la phosphorylation du dimère de ComD nécessite la liaison du CSP et ainsi la synthèse du pré-CSP, sa maturation et son export (**Figure 17F**).
- SsbB est synthétisé puis dégradé. Sa synthèse nécessite la transition de l'état inactif à la forme active de ComX via l'interaction de ce dernier avec ComW (**Figure 17G**).
- DprA est synthétisé puis dégradé. Comme pour SsbB, sa synthèse nécessite la transition de l'état inactif à la forme active de ComX par interaction avec ComW (**Figure 17H**).

Tous les T-invariants cités ci-dessus ont un sens biologique. Comme chaque transition du réseau compose un T-invariant, notre modèle est donc couvert de T-invariants. Ainsi, chaque réaction du système participe à un comportement élémentaire du système (Koch et Heiner, 2008). Cela constitue une propriété importante car l'état de compétence est un état physiologique transitoire et le système doit retourner à son état initial. Le modèle permet à un système qui est dans un état végétatif, d'entrer de façon transitoire en état de compétence, puis de retourner à son état initial. En d'autres termes, dans le réseau implémenté, si une séquence de transition menant à l'état de compétence existe, cela signifie qu'il existe aussi une séquence de transitions qui permet au système de retourner à l'état initial.

Un P-invariant correspond à un ensemble de places parmi lesquelles la somme de jetons est constante, indépendamment de toute transition pouvant moduler le nombre de jetons sur ces places. Dans le modèle, hormis deux P-invariants qui correspondent à la conservation de jetons entre les différentes formes (libres, actives, inactives) des deux promoteurs  $P_{comC}$  et  $P_{comAB}$ , aucun P-invariant n'est observé. Ce résultat est cohérent car l'état de compétence est un état transitoire, contrôlé génétiquement, permettant une reprogrammation cellulaire lorsque la cellule se trouve face à certains stress environnementaux. Ainsi, les protéines nécessaires à la mise en place de cet état physiologique doivent être synthétisées lorsque la cellule entre en état de compétence, puis dégradées lorsque l'état de compétence disparaît pour permettre alors à la cellule de retourner à son état physiologique initial. Il n'y a donc pas de conservation de masse dans l'ensemble du modèle, propriété reflétée par l'absence de couverture de P-invariants sur le réseau. Comme chaque transition représentant la synthèse de chaque composé a été associée à une transition de dégradation, il n'y a pas d'accumulation infinie d'une molécule donnée.

Comme les invariants structuraux calculés valident la cohérence structurale du modèle, le comportement dynamique du réseau a été étudiée dans un second temps. Pour bénéficier des données expérimentales de luminométrie, le réseau a été transformé en équations différentielles ordinaires, en utilisant les cinétiques chimiques.

### III. Analyse dynamique du réseau de régulation ComCDE

#### 1. Construction du système d'équations différentielles ordinaires

Les équations différentielles ordinaires permettent d'exprimer la variation d'une entité en fonction de la concentration des autres éléments d'un système. Chaque place du réseau de Petri correspond donc à une espèce moléculaire qui est modélisée par une variable d'état  $x_i(t)$  (par exemple le régulateur de réponse ComE correspondra à  $x_1(t)$ , le CSP à  $x_2(t)$ , etc.). Les valeurs de toutes les variables d'état, décrivant le système à un temps  $t$ , constituent l'état du système à ce temps  $t$ . L'évolution au cours du temps de  $x_i(t)$  peut être exprimée par une équation différentielle dont la forme générale est la suivante :

$$\frac{dx_i}{dt} = \text{synthèse}(x_i) - \text{dégradation}(x_i) \pm \text{liaison}(x_i) \pm \text{dissociation}(x_i) \\ \pm \text{modification post traductionnelle}(x_i)$$

L'ensemble des termes positifs correspond à des vitesses de gain et l'ensemble des termes négatifs correspond à des vitesses de perte. Lorsque les deux vitesses sont égales pour toutes les variables, le système a atteint un point d'équilibre aussi appelé état d'équilibre.

Pour une variable d'état donnée (une place donnée du réseau de Petri), toutes les transitions orientées vers la place contribuent à l'augmentation de la quantité de cette variable, et tous les arcs sortant d'une place donnée contribuent à la diminution de celle-ci. Les cinétiques des 15 espèces moléculaires du réseau, correspondant aux 15 places du réseau de Petri, ont ainsi été traduites par 15 équations différentielles qui sont reportées dans le **Tableau 2**.

**Tableau 2. Équations différentielles pour le modèle de régulation décrit dans la Figure 16.**

$$\frac{d[\text{preCSP}]}{dt} = \beta_{\text{ComCDE}} + v_{\text{maxComCDE}} \frac{[(\text{ComE} \sim \text{P})_D]}{[(\text{ComE} \sim \text{P})_D] + K_{\text{ComE} \sim \text{P}} * \left(1 + \frac{[\text{ComE}^e]}{K_{i_{\text{ComE}}^e}}\right)} - \varepsilon * [\text{ComAB}] * [\text{preCSP}] \\ - \gamma_{\text{preCSP}} * [\text{preCSP}]$$

$$\frac{d[\text{CSP}]}{dt} = \varepsilon * [\text{ComAB}] * [\text{preCSP}] - v_{\text{maxact}} * \frac{[\text{CSP}]}{[\text{CSP}] + K_{\text{act}}} * [(\text{ComD})_D] - \gamma_{\text{CSP}} * [\text{CSP}]$$

$$\frac{d[\text{ComAB}]}{dt} = \beta_{\text{ComAB}} + v_{\text{maxComAB}} \frac{[(\text{ComE} \sim \text{P})_D]}{[(\text{ComE} \sim \text{P})_D] + K_{\text{ComE} \sim \text{P}_{AB}} * \left(1 + \frac{[\text{ComE}^e]}{K_{i_{\text{ComE}_{AB}}^e}}\right)} - \gamma_{\text{ComAB}} * [\text{ComAB}]$$

$$\frac{d[\text{ComD}]}{dt} = \beta_{\text{ComCDE}} + v_{\text{maxComCDE}} \frac{[(\text{ComE} \sim \text{P})_D]}{[(\text{ComE} \sim \text{P})_D] + K_{\text{ComE} \sim \text{P}} * \left(1 + \frac{[\text{ComE}^e]}{K_{i_{\text{ComE}}^e}}\right)} - 2 * k_{\text{on}_D} * [\text{ComD}]^2 \\ + 2 * k_{\text{off}_D} * [(\text{ComD})_D] - \gamma_{\text{ComD}} * [\text{ComD}]$$

$$\frac{d[(\text{ComD})_D]}{dt} = k_{\text{on}_D} * [\text{ComD}]^2 - k_{\text{off}_D} * [(\text{ComD})_D] - v_{\text{maxact}} * \frac{[\text{CSP}]}{[\text{CSP}] + K_{\text{act}}} * [(\text{ComD})_D] - \alpha_{\text{auto}} \\ * [(\text{ComD})_D]$$

$$\frac{d[(ComD_{act})_D]}{dt} = v_{max_{act}} * \frac{[CSP]}{[CSP] + K_{act}} * [(ComD)_D] + \alpha_{auto} * [(ComD)_D] - \lambda * [ComE]^2 * [(ComD_{act})_D]$$

$$\frac{d[(ComD_{consumed})_D]}{dt} = \lambda * [ComE]^2 * [(ComD_{act})_D] - \gamma_{ComD} * [(ComD_{consumed})_D]$$

$$\frac{d[ComE]}{dt} = \beta_{ComCDE} + v_{max_{ComCDE}} * \frac{[(ComE \sim P)_D]}{[(ComE \sim P)_D] + K_{ComE \sim P} * \left(1 + \frac{[ComE]^e}{K_{ComE}^e}\right)} + 2 * \rho * [(ComE \sim P)_D] - 2 * \lambda * [ComE]^2 * [(ComD_{act})_D] - \gamma_{ComE} * [ComE]$$

$$\frac{d[(ComE \sim P)_D]}{dt} = \lambda * [ComE]^2 * [(ComD_{act})_D] - k_{on_{DprA_{EP}}} * [DprA]^2 * [(ComE \sim P)_D] - \rho * [(ComE \sim P)_D]$$

$$\frac{d[ComX_{ina}]}{dt} = v_{max_{ComX}} * \frac{[(ComE \sim P)_D]^x}{[(ComE \sim P)_D]^x + K_{ComX}^x} + v_{max_{DprA}} * \frac{[ComX_{act}]^d}{[ComX_{act}]^d + K_{DprA}^d} + v_{max_{SsbB}} * \frac{[ComX_{act}]^s}{[ComX_{act}]^s + K_{SsbB}^s} - \omega_1 * [ComW] * [ComX_{ina}] - \gamma_{ComX} * [ComX_{ina}]$$

$$\frac{d[ComX_{act}]}{dt} = \omega_1 * [ComW] * [ComX_{ina}] - v_{max_{DprA}} * \frac{[ComX_{act}]^d}{[ComX_{act}]^d + K_{DprA}^d} - v_{max_{SsbB}} * \frac{[ComX_{act}]^s}{[ComX_{act}]^s + K_{SsbB}^s} - \gamma_{ComX} * [ComX_{act}]$$

$$\frac{d[ComW]}{dt} = v_{max_{ComW}} * \frac{[(ComE \sim P)_D]^w}{[(ComE \sim P)_D]^w + K_{ComW}^w} - \gamma_{ComW} * [ComW]$$

$$\frac{d[DprA]}{dt} = v_{max_{DprA}} * \frac{[ComX_{act}]^d}{[ComX_{act}]^d + K_{DprA}^d} - 2 * k_{on_{DprA_{EP}}} * [DprA]^2 * [(ComE \sim P)_D] - \gamma_{DprA} * [DprA]$$

$$\frac{d[(DprA_{ComE \sim P})_D]}{dt} = k_{on_{DprA_{EP}}} * [DprA]^2 * [(ComE \sim P)_D] - \gamma_{DprA} * [(DprA_{ComE \sim P})_D]$$

$$\frac{d[SsbB]}{dt} = v_{max_{SsbB}} * \frac{[ComX_{act}]^s}{[ComX_{act}]^s + K_{SsbB}^s} - \gamma_{SsbB} * [SsbB]$$

$$[ComD]_{total} = [ComD] + 2 * [(ComD)_D] + 2 * [(ComD_{act})_D] + 2 * [(ComD_{consumed})_D]$$

$$[ComE]_{total} = [ComE] + 2 * [(ComE \sim P)_D] + 2 * [(DprA_{ComE \sim P})_D]$$

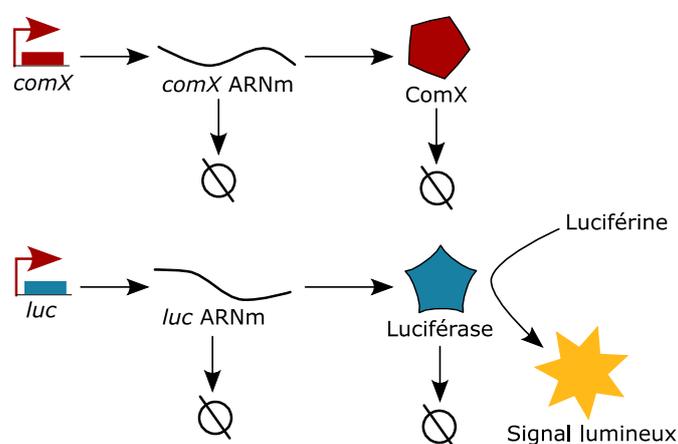
$$[ComX]_{total} = [ComX_{ina}] + [ComX_{act}]$$

$$[DprA]_{total} = [DprA] + 2 * [(DprA_{ComE \sim P})_D]$$

Les approches utilisant les équations différentielles ordinaires peuvent fournir un niveau de détail élevé sur les dynamiques des composants du système, mais nécessitent cependant des données quantitatives, par exemple les cinétiques de composés, afin d'estimer les paramètres du modèle. Toutefois, les mesures quantitatives sont souvent partielles et disponibles seulement pour une fraction des entités du système. Ainsi, plusieurs paramètres des modèles quantitatifs doivent être inférés. Ici, les paramètres du réseau de régulation ont été ajustés et estimés à partir d'évidences expérimentales, mais aussi grâce à l'utilisation de mesures de fusion transcriptionnelle.

## 2. Utilisation des données de fusion transcriptionnelle

Les données de fusion transcriptionnelle seront utilisées pour paramétrer le modèle. En effet, les expérimentalistes de l'équipe « Transformation du pneumocoque » travaillent depuis des années avec des données de luminescence pour comprendre les mécanismes de régulation de la compétence (Prudhomme et Claverys, 2007). Ces données serviront à définir les paramètres du système d'équations différentielles ordinaires afin d'en étudier le comportement et le comparer aux données expérimentales. Cependant, les données brutes issues des mesures de luminescence reflètent la quantité de luciférase et non pas la quantité de la protéine d'intérêt. En effet, la demi-vie de la protéine d'intérêt et celle de la luciférase peuvent différer, et il en est de même pour les ARNm (**Figure 18**). Il est donc nécessaire de rechercher une méthode mathématique afin de tirer bénéfice des données disponibles et les utiliser comme apprentissage des cinétiques des protéines régulant la compétence.



**Figure 18. Représentation schématique de l'expression d'un gène rapporteur et d'un gène de l'hôte.** Expression du gène *comX* et du gène codant pour la luciférase impliquant les réactions de transcription, de traduction et de dégradation des produits de l'expression génique (modifié d'après de Jong *et al.*, 2010).

Les données de luminescence et de fluorescence de gènes rapporteurs constituent des outils très populaires pour quantifier *in vivo* l'expression de gènes. Le principe de cette technologie est de fusionner la région promotrice, et parfois même une partie de la région codante d'un gène d'intérêt à un gène rapporteur (**Figure 18**). Le gène rapporteur est exprimé à partir d'un plasmide ou peut être intégré au sein du chromosome hôte. Une bactérie portant la construction émet alors un signal lumineux qui est proportionnel à la quantité de protéine rapportrice dans la cellule. Ce signal de luminescence, ou de fluorescence, est mesurable et reflète ainsi le niveau d'expression du gène d'intérêt. Les données qui résultent de ces expériences sont utilisables, notamment dans le contexte de l'inférence ou de l'analyse de

réseaux de régulation dans les cellules bactériennes (Ronen *et al.*, 2002, Berthoumieux *et al.*, 2013, Stefan *et al.*, 2015).

L'utilisation de fusions transcriptionnelles permet de mesurer l'expression de gènes à l'échelle de la population, mais aussi à l'échelle de la cellule unique. À l'échelle de la cellule unique, grâce à l'utilisation à la fois de la microscopie et de la fluorescence, il a notamment été démontré que la fluctuation de l'expression génique était due à la combinaison du bruit interne et du bruit externe du système étudié (Elowitz *et al.*, 2002). En revanche, les lectures automatiques à partir de microplaques permettent, elles, la mesure de l'expression de gènes dans une population cellulaire plutôt que l'expression à l'échelle de cellules individuelles. Ce sont ces données qui sont à notre disposition. La résolution plus faible de ces données est cependant compensée par un débit plus élevé car plusieurs gènes peuvent être mesurés en parallèle, avec un ensemble de mesures obtenues dans le temps ce qui constitue un atout par rapport aux puces à ADN. De nombreux exemples de quantification de l'expression de gènes à l'échelle populationnelle sont retrouvés dans la littérature, associés à une approche de modélisation. Nous pouvons citer comme exemples l'étude du processus décisionnaire régulant l'entrée en phase lytique ou lysogène chez le phage lambda (Kobiler *et al.*, 2005), l'étude du stress oxydatif (Lu *et al.*, 2005) ou bien l'étude du système SOS en réponse à des dommages à l'ADN (Ronen *et al.*, 2002).

Dans une expérience de microplaque, jusqu'à 96 cultures peuvent être suivies en parallèle, et cela sur plusieurs heures. Cela génère une quantité importante de données : afin de les interpréter de façon adéquate, il est nécessaire de comprendre les facteurs agissant sur ces mesures d'expressions de gènes. Une utilisation des données de fusion transcriptionnelles consiste à inférer les activités de promoteurs à partir des données brutes d'expressions de gènes rapporteurs. Des modèles mathématiques ont été développés pour inférer des quantités interprétables à partir des mesures brutes (Ronen *et al.*, 2002, Finkenstädt *et al.*, 2008, Huang *et al.*, 2008, Wang *et al.*, 2008, Aïchaoui *et al.*, 2012). Ces méthodes permettent de reconstruire la concentration, relative, de l'ARNm rapporteur ainsi que la concentration de la protéine rapportrice à partir des données. Ces valeurs sont considérées comme une estimation de la cinétique d'expression du gène hôte correspondant. Toutefois, la cinétique de la protéine rapportrice et celle exprimée par le gène hôte peuvent différer, notamment en raison de la différence de demi-vies entre la protéine rapportrice et la protéine hôte. Une méthode mathématique a été développée pour analyser et corriger ces biais (de Jong *et al.*, 2010, Stefan *et al.*, 2015). Cette méthode, présentée ci-dessous, a été utilisée dans le travail présenté ici.

Si  $A(t)$  correspond à la valeur d'absorbance corrigée de l'absorbance de fond au temps  $t$ , et que  $I(t)$  représente l'intensité brute mesurée de luminescence (en unité relative de luminescence RLU) au temps  $t$ , alors le ratio  $r(t) = I(t)/A(t)$  correspond à la quantité de luminescence moyenne par cellule au temps  $t$ . Le signal de luminescence peut alors être transformé en activité du promoteur en calculant sa dérivée selon la formule suivante :

$$f(t) = \frac{d}{dt}r(t) + (\gamma_r + \mu(t))r(t) = \frac{\frac{d}{dt}I(t)}{A(t)} + \gamma_r \frac{I(t)}{A(t)}$$

où  $\gamma_r$  [ $\text{min}^{-1}$ ] est la constante de dégradation de la luciférase et où  $\mu(t)$  [ $\text{min}^{-1}$ ] est la vitesse de croissance de la population bactérienne. La relation entre la demi-vie et la constante de dégradation d'une protéine est donnée par  $t_{1/2} = \ln 2 / \gamma_r$ . Le temps de demi-vie de la luciférase utilisée dans les expériences a été estimé à 15 minutes (Prudhomme et Claverys, 2007), donnant  $\gamma_r = 0,04621$ . L'activité du promoteur  $f(t)$  est exprimée en  $\text{RLU min}^{-1}$ .

L'évolution au cours du temps de la protéine d'intérêt  $p(t)$  [ $\text{RLU.min}^{-1}$ ] a été calculée en utilisant l'équation suivante établie par Stefan et collaborateurs (2015). Ce modèle opère une correction de la différence de demi-vies entre le reporteur luciférase et la protéine dont l'activité du gène est mesurée :

$$\frac{d}{dt}p(t) = f(t) - (\gamma_p + \mu(t))p(t), \quad \text{avec } p(0) = \frac{\gamma_r}{\gamma_p} r(t)$$

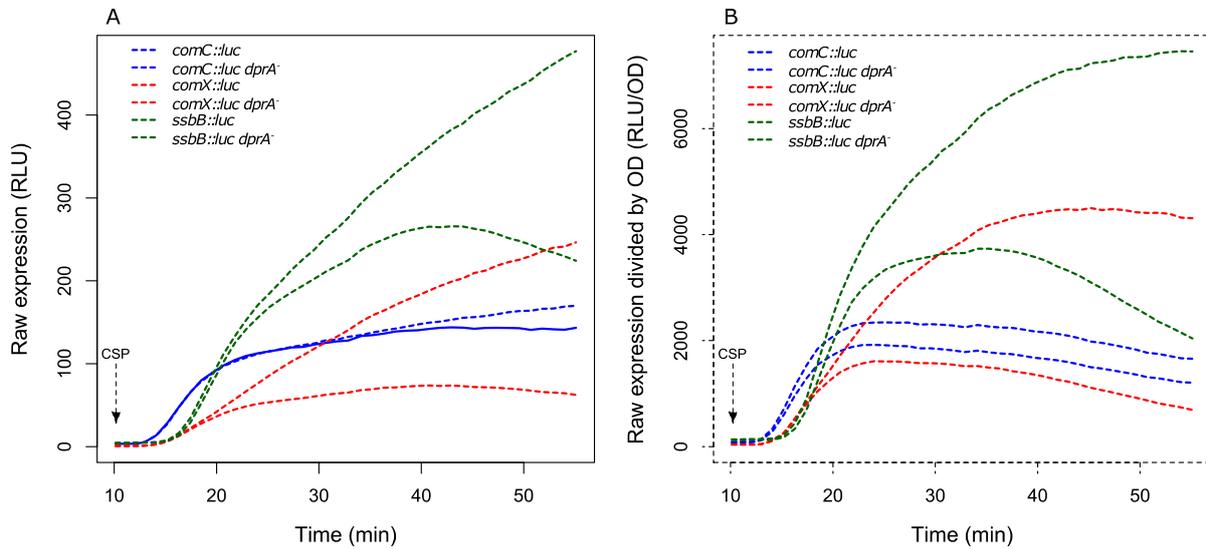
où  $\gamma_p$  [ $\text{min}^{-1}$ ] est la constante de dégradation de la protéine. À partir des résultats publiés et par mesure de simplification, nous avons considéré seulement deux durées de vie différentes ; une pour ComD, ComE, DprA et ComAB qui a été estimée à 80 minutes ( $t_{1/2} = 55,45$  min et  $\gamma_p = 0,0125$ ) et l'autre pour ComX, ComW et SsbB qui a été estimée à 8 minutes ( $t_{1/2} = 5,45$  min et  $\gamma_p = 0,125$ ) (Piotrowski *et al.*, 2009 pour ComX et ComW ; Mirouze *et al.*, 2013 pour DprA et SsbB ; Martin *et al.*, 2013 pour ComE et ComD). La concentration initiale de protéines  $p(0)$  a été obtenue à partir de données de gènes rapporteurs à  $t = 0$  et en supposant que l'expression des gènes impliqués dans l'état de compétence était à l'état d'équilibre au début de l'expérience.

Les équations ont été résolues par intégration numérique en utilisant la méthode Euler implémentée dans le package *deSolve* du logiciel R. Finalement, les cinétiques de protéines obtenues ont été normalisées par rapport à la concentration protéique maximale obtenue sur l'ensemble des jeux de données utilisés.

### 3. Traitement des données de luminescence

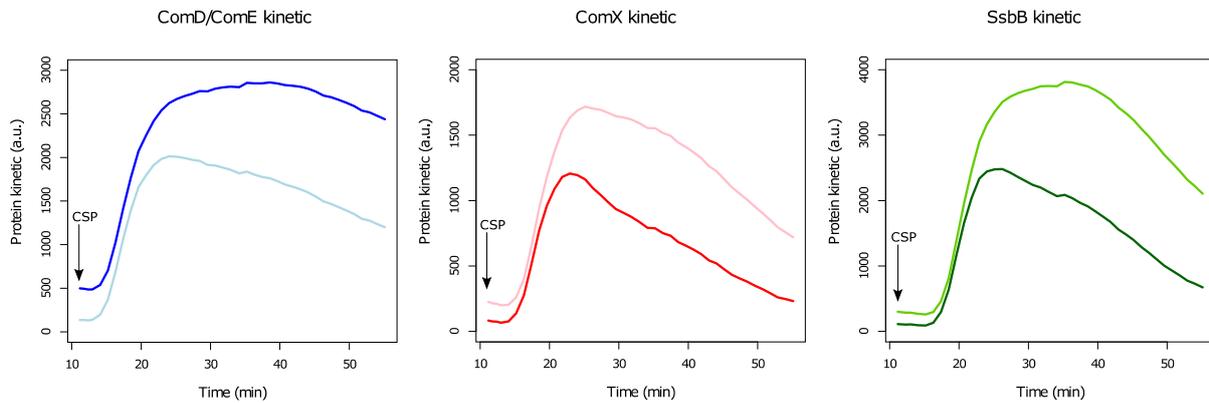
Afin d'estimer les paramètres du modèle et pour obtenir une cinétique de référence d'expression des gènes de la compétence, les données recueillies par fusion transcriptionnelle avec la luciférase ont été utilisées afin d'obtenir une mesure en temps réel de l'expression des gènes. Dans ces expériences, la compétence est induite par l'ajout de CSP exogène à une concentration saturante afin de synchroniser la réponse populationnelle. Les données utilisées de luminescence sont celles produites par Mirouze et collaborateurs (Mirouze *et al.*, 2013) où les fusions du gène rapporteur luciférase aux promoteurs des gènes d'intérêts (*comC::luc*, *comX::luc* et *ssbB::luc*) ont été introduites dans une souche sauvage ainsi que dans une souche *dprA*<sup>-</sup>. La luminescence (exprimée en unité relative de luminescence, RLU) et l'absorbance (OD<sub>492nm</sub>) sont enregistrées toutes les minutes après l'ajout de CSP sur une période de 45 minutes. Les signaux de luminescence brute ont été transformés en activités de promoteurs et en cinétique protéique par cellule en appliquant l'approche présentée précédemment. La quantité de luminescence par cellule en fonction du temps a été calculée à partir du ratio  $I(t)/A(t)$ , où  $I(t)$  représente l'intensité de luminescence (en RLU) et  $A(t)$  les valeurs d'absorbance corrigées par soustraction de l'absorbance de fond mesurée dans des puits contenant seulement du milieu de croissance. Il est raisonnable de supposer que ce ratio permet d'estimer la concentration moyenne de protéine rapportrice par cellule. Ainsi, la dynamique du système peut être décrite par l'évolution dans le temps de la concentration de luciférase.

Si l'on compare la variation au cours du temps des données brutes de luminescence (**Figure 19A**) à l'évolution du ratio  $r(t)$  (**Figure 19B**), nous remarquons que la normalisation par l'absorbance permet la correction des effets causés à l'augmentation différentielle de la taille des populations. En réalité, il a été montré qu'une souche *dprA*<sup>-</sup> possède une vitesse de croissance plus faible qu'une souche sauvage (Mirouze *et al.*, 2013). Ici, nous observons que les données de luminescence brutes sous-estiment l'accroissement de l'expression des gènes entre une souche *dprA*<sup>-</sup> et une souche sauvage par rapport aux données normalisées de luminescence par cellule. Cela est particulièrement vrai dans le cas de l'opéron *comCDE* où l'inactivation de *dprA* semble n'avoir quasi aucun effet sur son expression lorsque les données brutes sont analysées (**Figure 19A**), mais apparaît avoir un effet lorsque les valeurs normalisées de luminescence sont utilisées (**Figure 19B**). La correction de l'effet de la taille de la population est donc cruciale pour une analyse rigoureuse des données de luminescence en termes de niveau d'expression génique.



**Figure 19. Comparaison des données brutes de luminescence aux données normalisées par la taille de la population pour les différents promoteurs.** Pour les panneaux de gauche et de droite, les lignes de couleurs continues représentent les données expérimentales acquises pour la souche WT et les lignes en pointillés correspondent aux données expérimentales acquises pour la souche *dprA*. Le panneau de gauche représente les intensités lumineuses brutes et le panneau de droite représente les intensités lumineuses brutes divisées par l'absorbance de la population. Le CSP est ajouté à t = 10 minutes.

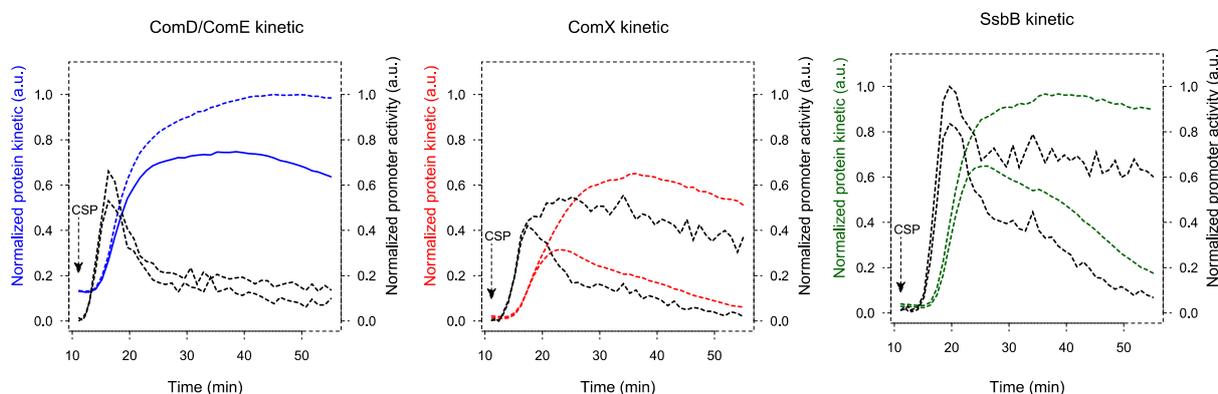
L'étape suivante est la transformation du signal normalisé de luminescence par cellule en activité de promoteur et en cinétiques protéiques en suivant toujours le modèle mathématique développé par Stefan et collaborateurs (2015). Pour calculer l'évolution au cours du temps de la concentration de la protéine d'intérêt, une correction est réalisée afin de prendre en compte la différence de demi-vie entre la luciférase et la protéine dont l'expression est mesurée. L'effet de cette correction sur nos données est montré en **Figure 20**. Pour ComX et SsbB, qui ont une durée de vie plus courte (estimée à 8 minutes) que celle de la luciférase mesurée à 21,6 minutes, nous notons que les valeurs non corrigées de luminescence surestiment clairement les quantités protéiques. À l'inverse, pour ComE et ComD dont les durées de vie ont été estimées à 80 minutes, les valeurs non corrigées de luminescence sous-estiment la concentration protéique au cours du temps.



**Figure 20. Effet de la correction entre la durée de vie de la luciférase et celle des autres protéines.** Les lignes de couleur plus foncées représentent les concentrations de protéines calculées avec la durée de vie de la protéine (respectivement 80 minutes pour ComD/ComE et 8 minutes pour ComX et SsbB). Les couleurs plus claires représentent les concentrations de protéines calculées en utilisant la durée de vie de la luciférase mesurée (21,6 minutes).

Les cinétiques des activités des promoteurs et des concentrations protéiques déduites des données de luminescence des fusions *comC::luc*, *comX::luc* et *ssbB::luc* obtenues par Mirouze et collaborateurs (2013) dans des cellules sauvages et *dprA*<sup>-</sup> sont reportées en **Figure 21**. Pour les cellules sauvages, après l'ajout de CSP dans le milieu, un pulse d'activité pour chaque promoteur est observé, en accord avec le fait que la compétence est un état physiologique transitoire. Les courbes représentant les cinétiques protéiques ont des allures similaires aux courbes d'activités des promoteurs, avec un décalage dû au temps nécessaire à la traduction des ARNm en protéines et à une diminution de la pente qui est cohérente avec ce qui est connu quant à la demi-vie des protéines (*i.e.* plus élevée pour ComX et SsbB qui ont des demi-vies courtes que pour ComD et ComE qui ont des demi-vies longues). Dans les cellules *dprA*<sup>-</sup>, l'altération de l'arrêt de la transcription de *comX* est clairement observée, ainsi que son incidence sur le maintien de l'expression de *ssbB*. De plus, les cinétiques protéiques obtenues dans les cellules sauvages sont en accord avec les résultats expérimentaux précédemment publiés, confortant le choix effectué de demi-vie protéique utilisée. Pour ComD et ComE, le pic de concentration est atteint 15 minutes après l'ajout de CSP et la concentration reste stable sur une période d'au moins 25 minutes, ce qui concorde avec les résultats de *western blot* obtenus par Martin et collaborateurs (2013). ComX apparaît entre 5 et 10 minutes après l'ajout de CSP, atteignant un pic de concentration après 15 minutes, avant d'être dégradé. La cinétique suit également les résultats expérimentaux obtenus par *western blot* avec un anticorps anti-ComX (Piotrowski *et al.*, 2009). La cinétique de SsbB est aussi cohérente avec les précédents résultats expérimentaux (Mirouze *et al.*, 2013) : la protéine est détectée environ 5 minutes après l'ajout de CSP, la concentration maximale apparaît après 15-

20 minutes, et la protéine disparaît quasi totalement après 60 minutes. Enfin, la présence de ComX dans un contexte *dprA*<sup>-</sup> est prolongée avec la même temporalité, comme dans le travail de Weng et collaborateurs (2013).



**Figure 21. Estimation de l'activité des promoteurs et des cinétiques protéiques déduites des données de luminescence.** Les activités des promoteurs calculées à partir des signaux de luminescence sont représentées par des lignes noires (échelle de droite). Les cinétiques protéiques reconstituées et corrigées des différences de demi-vie entre la luciférase et la protéine dont l'activité génétique est mesurée sont illustrées par des lignes colorées (échelle de gauche). Les lignes pleines et les pointillés correspondent respectivement aux valeurs calculées obtenues à partir des données brutes de luminescence dans les souches WT et *dprA*<sup>-</sup>. Les temps de vie de la protéine utilisés pour le calcul sont : 8 min pour ComX et SsbB, 80 min pour ComD et ComE et 21,6 min pour luciférase. Les gènes *comD* et *comE* étant en opéron, la ligne bleue symbolise à la fois la cinétique de ComD et celle de ComE. Les lignes rouges et vertes représentent respectivement les cinétiques ComX et SsbB. Les activités des promoteurs et les concentrations des protéines ont été normalisées par rapport aux valeurs maximales obtenues dans les ensembles de données globales calculées. Par conséquent, les valeurs sont données en unités arbitraires (a.u.).

Trois jeux de données ont été utilisés simultanément pour estimer les valeurs des paramètres du modèle : les deux précédemment décrits contenant les cinétiques de luminescence transformées en cinétiques protéiques obtenues à partir de souches sauvages et *dprA*<sup>-</sup> (Mirouze *et al.*, 2013) et un troisième correspondant à la cinétique des protéines dans une souche *clpP*<sup>-</sup>. Pour ce dernier, il n'existe pas de données de luminescence directement disponibles, mais il a été observé que dans ce mutant les cellules quittent l'état de compétence aussi rapidement que dans des cellules sauvages (Piotrowski *et al.*, 2009). Ainsi, les données cinétiques d'un mutant *clpP*<sup>-</sup> où ComX devient stable ont été générées en utilisant le jeu de données sauvage, tout en choisissant une constante de dégradation  $\gamma_p$  de ComX égale à 0 afin d'inférer la cinétique de ComX dans ce contexte génétique.

Après avoir validé les cinétiques des concentrations protéiques, celles-ci ont été utilisées pour estimer les paramètres des équations différentielles et pour comparer la dynamique de la prédiction du réseau à ces données d'apprentissage. Les valeurs des paramètres ont été estimées en utilisant les données de luminométrie, mais aussi en exploitant les connaissances publiées disponibles.

#### 4. Évidences expérimentales aidant à l'estimation des paramètres

Les connaissances qualitatives acquises par expériences de gels retards ont montré que ComE~P avait une plus forte affinité que ComE pour P<sub>comC</sub> mais qu'il avait beaucoup moins d'affinité pour P<sub>comX</sub> (Martin *et al.*, 2013). Ces informations ont permis d'ordonner les affinités des constantes régulatrices pour leurs séquences cibles. Ainsi, les contraintes suivantes ont été ajoutées lors de l'estimation des paramètres :  $K_{ComX} > K_{ComE} > K_{ComE\sim P}$ .

Comme aucune mesure de durée de vie précise n'a été réalisée, des *westerns blots* issus de diverses publications ont été utilisés pour déterminer un intervalle de constantes de dégradation qui soit le plus cohérent possible avec les observations. L'intervalle [4-12] minutes a été utilisé pour ComX et SsbB qui sont des protéines instables et l'intervalle [40-120] minutes pour ComE, ComD et DprA qui sont des protéines plus stables (Weng *et al.*, 2013, Martin *et al.*, 2013, Mirouze *et al.*, 2013).

Les résultats expérimentaux obtenus sur le système AgrC/AgrA de *Staphylococcus aureus* ont été également exploités. En effet, comme ComD possède 50% de similarité avec AgrC et que ComE possède 52% de similarité avec AgrA, l'on peut supposer que ces deux systèmes à deux composants partagent des propriétés biochimiques similaires. Un temps de demi-vie de 3,9 minutes de la forme phosphorylée du régulateur de réponse AgrA a été mesuré, correspondant à une durée de vie moyenne de 5,26 minutes et à une vitesse d'auto-déphosphorylation de  $0.18 \text{ min}^{-1}$  (Wang *et al.*, 2014). Ainsi, la recherche dans l'espace des paramètres pour estimer la vitesse  $\rho$  de déphosphorylation de ComE~P a été restreinte dans l'intervalle [0,1-0,4]. De plus, la vitesse de transfert du groupe phosphorylé entre la forme active de l'histidine kinase AgrC et le régulateur de réponse AgrA a également été mesurée avec un  $t_{1/2}$  égale à 66 secondes, correspondant à une vitesse de transphosphorylation de  $0.63 \text{ min}^{-1}$  (Wang *et al.*, 2014). Ainsi, la recherche de la constante de transphosphorylation  $\lambda$  a été restreinte dans l'intervalle [0,4-1].

Les concentrations initiales de protéines ont été choisies telles que le système soit à l'état d'équilibre au début de la simulation, état qui correspond à l'état végétatif/non-compétent, et de sorte que l'ajout de CSP exogène puisse initier le développement de la compétence.

## 5. Optimisation des paramètres du modèle

L'optimisation est au cœur de nombreux problèmes rencontrés en modélisation et dans la conception de systèmes biologiques (Banga, 2008). Cela implique que soient définis quelques concepts fondamentaux : les variables décisionnelles (ce sont les variables qui peuvent être modifiées lors de la recherche de la meilleure solution), une fonction objectif (l'indice de performance qui quantifie la qualité d'une solution définie par un ensemble de variables de décision et qui peut être minimisée ou maximisée) et des contraintes à l'optimisation (exigences qui doivent être satisfaites lors de l'estimation, généralement exprimées sous la forme d'égalités et d'inégalités). Les variables décisionnelles peuvent être continues (générant des problèmes d'optimisation dits continus) ou discrètes (représentées par des nombres entiers). Ces dernières, également appelées optimisations combinatoires, sont souvent utilisées lors de la conception *in silico* de réseaux de régulation de gènes ou lors d'optimisation de modèles par rapport à des données expérimentales (Saez-Rodriguez *et al.*, 2009, Schreiber *et al.*, 2016).

Ainsi, l'inférence des paramètres des équations différentielles ordinaires peut être formulée comme un problème d'optimisation avec une fonction objective qui doit être minimisée par l'ajustement des valeurs des paramètres du modèle. Un choix couramment fait pour calculer cette fonction objective est de calculer la somme des carrés des résidus entre mesures expérimentales et simulations du modèle. Pour un vecteur de paramètre  $P$ , la valeur de la fonction objective est donnée par la formule suivante (Hoops *et al.*, 2006) :

$$E(P) = \sum_{i=1}^n \sum_{j=1}^k \omega_j (x_{i,j} - y_{i,j}(P))^2$$

où  $n$  est le nombre de points de mesures expérimentales,  $k$  le nombre de variables,  $y_{i,j}(P)$  sont les données simulées correspondants aux données expérimentales  $x_{i,j}$ .  $\omega_j$  est un poids permettant de normaliser l'importance de chaque variable afin que toutes aient la même importance dans l'ajustement des paramètres. En effet, les variables ayant les valeurs moyennes les plus élevées contribuent plus fortement à l'accroissement de la fonction objective et peuvent ainsi être mieux ajustées par l'algorithme d'optimisation au détriment des autres variables du modèle. Afin que toutes les variables possèdent une même importance, il

est nécessaire de pondérer chacune d'elles. La méthode *mean square*, implémentée dans COPASI (Hoops *et al.*, 2006), a été utilisée et consiste à diviser chaque résidu par la moyenne de la variable considérée élevée au carré ( $\omega_j$ ).

La fonction d'optimisation appliquée est la méthode de *Particle Swarm Optimization* (PSO), implémentée dans le logiciel COPASI, qui a été originalement proposée par Kennedy et Eberhart (1995). Cette méthode d'optimisation stochastique est inspirée des comportements sociaux et des mouvements dynamiques des essaims comme ceux observés chez les insectes, les oiseaux et les poissons. Chaque particule de l'essaim correspond à un vecteur de paramètres et parcourt l'espace des paramètres afin d'identifier le vecteur de paramètres reproduisant au mieux les résultats expérimentaux. Une particule est décrite par une position dans l'espace des paramètres, par une vitesse et enfin par une direction. Chaque particule possède une mémoire de ses précédentes valeurs de fonctions objectives et garde aussi en mémoire sa meilleure valeur de fonction objective. La particule conserve également la meilleure solution atteinte par les particules voisines. Chaque particule de l'essaim a une position initiale fixée de façon aléatoire dans l'espace des paramètres. Ces particules parcourent alors l'espace des paramètres et ajustent leurs positions et leurs vitesses à partir de leur propre meilleure valeur de fonction objective et à partir de celles atteintes par leurs voisins. La taille de l'essaim est fixée à 50 particules, valeur par défaut de COPASI.

Ainsi, chaque itération de l'algorithme nécessite pour chaque particule de l'essaim une procédure de mise à jour en trois étapes. D'abord, le positionnement des particules dans l'espace des paramètres se fait par rapport aux valeurs courantes des paramètres du modèle. Puis, le modèle est simulé avec les valeurs de paramètres du modèle. Enfin, les simulations sont comparées avec les données expérimentales en utilisant la fonction objective décrite précédemment et de nouveaux vecteurs de paramètres sont générés grâce à l'algorithme d'optimisation. Le processus itératif s'arrête si le changement de la valeur de la fonction objective devient plus faible qu'une valeur seuil ou si un nombre spécifié d'itération est atteint. Des informations supplémentaires sur cette méthode d'optimisation peuvent être trouvées dans l'article de Poli et collaborateurs (2007).

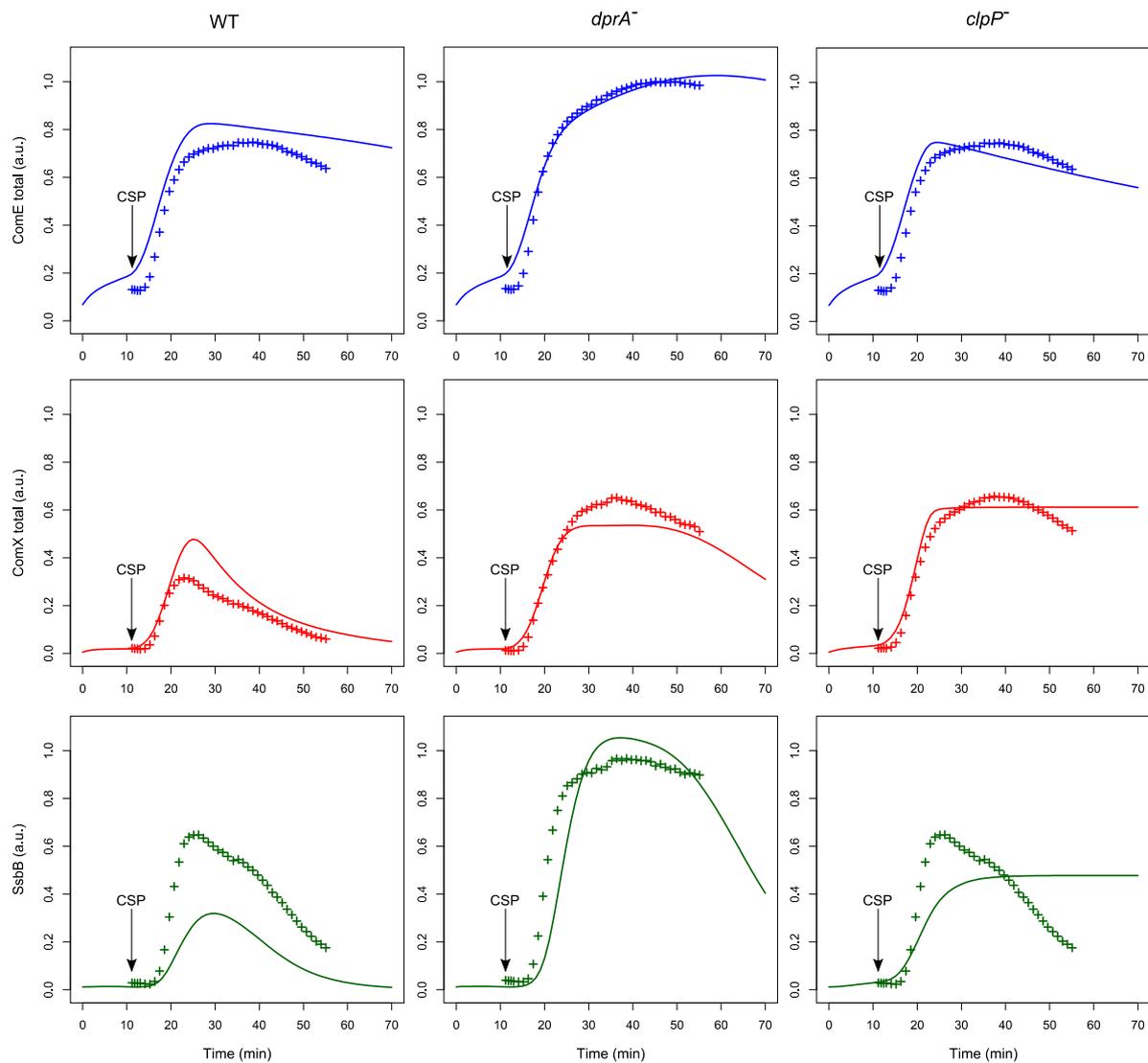
La méthode d'optimisation *Particle Swarm* (PSO) a été utilisée avec les paramètres par défaut proposés dans COPASI, à l'exception du nombre d'itérations qui a été augmenté de 2000 à 10 000 itérations afin d'obtenir le vecteur de paramètres le plus optimisé possible. Les performances des méthodes d'optimisation les plus populaires (programmation évolutive, algorithme génétique, recuit simulé et PSO) ont été en effet comparées par Baker et collaborateur pour retrouver les paramètres cinétiques des enzymes impliquées dans la

glycolyse (Baker *et al.*, 2010). Ce travail a montré que la méthode PSO était la méthode la plus performante pour minimiser la fonction objective.

L'algorithme *Particle Swarm*, implémenté dans le logiciel COPASI (Hoops *et al.*, 2006), a donc été utilisé comme algorithme d'optimisation pour minimiser les résidus entre les cinétiques des concentrations protéiques calculées à partir des données de luminescence de ComD/ComE, ComX et SsbB et les cinétiques de  $[\text{ComD}]_{\text{total}}$ ,  $[\text{ComE}]_{\text{total}}$ ,  $[\text{ComX}]_{\text{total}}$  et  $[\text{SsbB}]$  obtenues à partir du modèle (voir **Tableau 2**). Les simulations du réseau sont réalisées par intégration numérique des équations différentielles ordinaires avec la méthode LSODA (Petzold, 1983). Pendant l'estimation des paramètres du modèle, pour générer les cinétiques protéiques dans un mutant  $dprA^-$ , le paramètre  $vmax_{dprA}$  correspondant à la vitesse maximale de synthèse de  $dprA$  prend la valeur 0 et, dans le mutant  $clpP^-$ , les constantes de dégradation de ComX et ComW ( $\gamma_{comX}$  et  $\gamma_{comW}$ ) sont respectivement fixées à 0.

## 6. Analyse de la dynamique du système

La comparaison des simulations du modèle avec les mesures expérimentales a été réalisée à partir de trois souches (WT,  $dprA^-$  et  $clpP^-$ ) : elle montre une relative bonne cohérence entre simulations et données (**Figure 22**). Cependant, un certain écart est observé entre simulations et données, plus particulièrement pour la cinétique de SsbB : alors que, dans un mutant  $clpP^-$ , la cinétique de SsbB décroît comme dans une souche sauvage, cette dynamique n'est pas retrouvée dans la simulation où la concentration de SsbB reste stable pendant plus de 50 minutes. Nous pouvons conclure qu'il existe probablement un acteur supplémentaire inconnu qui doit être impliqué dans l'arrêt de transcription des gènes tardifs, comme l'interprètent Piotrowski et collaborateurs (2009) ainsi que Weng et collaborateurs (2013). Ce réseau a été par conséquent modifié pour y intégrer un nouvel acteur agissant sur l'activité de ComX.

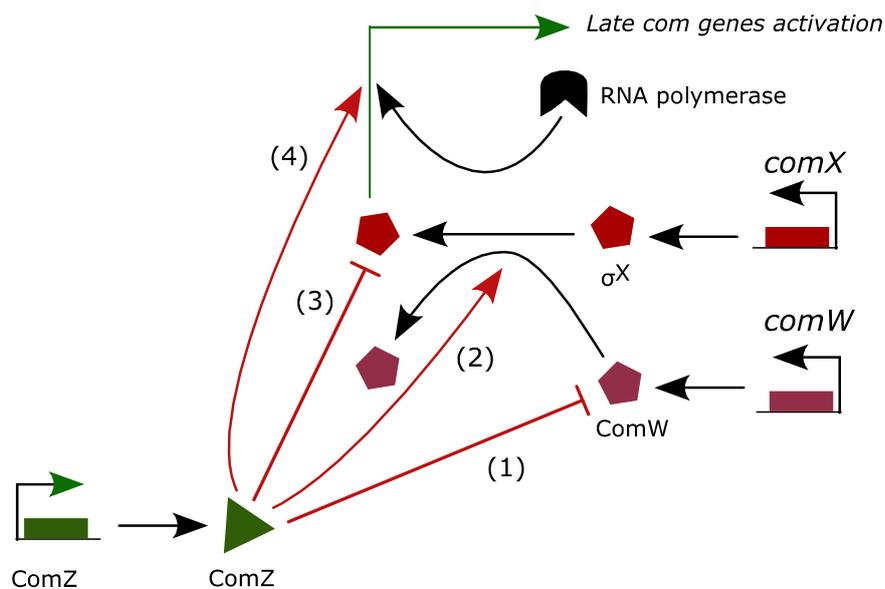


**Figure 22. Comparaison des cinétiques protéiques expérimentales et simulées pour le modèle initial.** La comparaison des données simulées avec les mesures expérimentales est montrée pour la souche WT, la souche mutante *dprA* et la souche mutante *clpP*. Puisqu’aucune donnée de luminescence n’était disponible pour le mutant *clpP*, elles ont été générées en utilisant le jeu de données WT et en fixant la constante de dégradation de ComX à 0 pour inférer la cinétique ComX. Pour les deux autres souches, la cinétique protéique reconstruite correspond à celle de la Figure 19. Les cinétiques expérimentales et simulées des protéines sont respectivement représentées par des croix et des lignes pleines. Des simulations pour le mutant *dprA* sont réalisées en définissant  $v_{max_{dprA}} = 0$  et des simulations pour le mutant *clpP* sont réalisées en définissant les paramètres  $\gamma_{ComX}$  et  $\gamma_{ComW}$  égaux à 0. Le développement de la compétence est induit en ajoutant une unité arbitraire (a.u) de CSP à  $t = 10$  minutes afin de reproduire le protocole expérimental – 100 ng / mL de CSP ajouté après 10 minutes d’incubation (Mirouze *et al.*, 2013) –. On utilise le même code de couleurs que pour la figure 21.

## 7. Comparaison des modèles alternatifs

Un nouveau gène, appelé *comZ*, a été ajouté et sa synthèse peut être sous le contrôle de ComE~P en admettant qu’il s’agit d’un gène précoce de la compétence, ou sous le contrôle de

ComX s'il s'agit un gène tardif de la compétence. Plusieurs mécanismes d'actions alternatifs de ComZ ont été considérés (**Figure 23**) : i) ComZ inhibe directement la forme active de ComX, ii) ComZ interagit avec ComW et empêche ainsi l'activation de ComX par ComW, iii) ComZ est en compétition avec ComW pour interagir avec la forme inactive de ComX et empêche la formation d'une forme active de ComX et enfin, iv) ComZ et la forme active de ComX sont en compétition pour l'ARN polymérase ; ComZ correspondrait ici au facteur  $\sigma$  de ménage ( $\sigma^A$ ).



**Figure 23. Modèles alternatifs possibles d'action de ComZ sur la fermeture de l'état de compétence. (1)** ComZ interagit avec ComW et ainsi empêche ainsi l'activation de ComX par ComW. **(2)** ComZ est en compétition avec ComW pour interagir avec la forme inactive de ComX et empêche la formation d'une forme active de ComX. **(3)** ComZ inhibe la forme active de ComX. **(4)** ComZ et la forme active de ComX sont en compétition pour l'ARN polymérase. Dans chacun des modèles, ComZ peut être régulé, soit par ComE~P, soit par ComX

**Tableau 3. Réactions modifiées lors d'une interaction entre ComW et un produit de gène tardif ComZ inhibant l'activité de ComW (voir Figure 23 (1)).**

Nouvelles réactions

**Synthèse de ComZ**



**Inhibition of active form of ComW**



Nouvelles équations

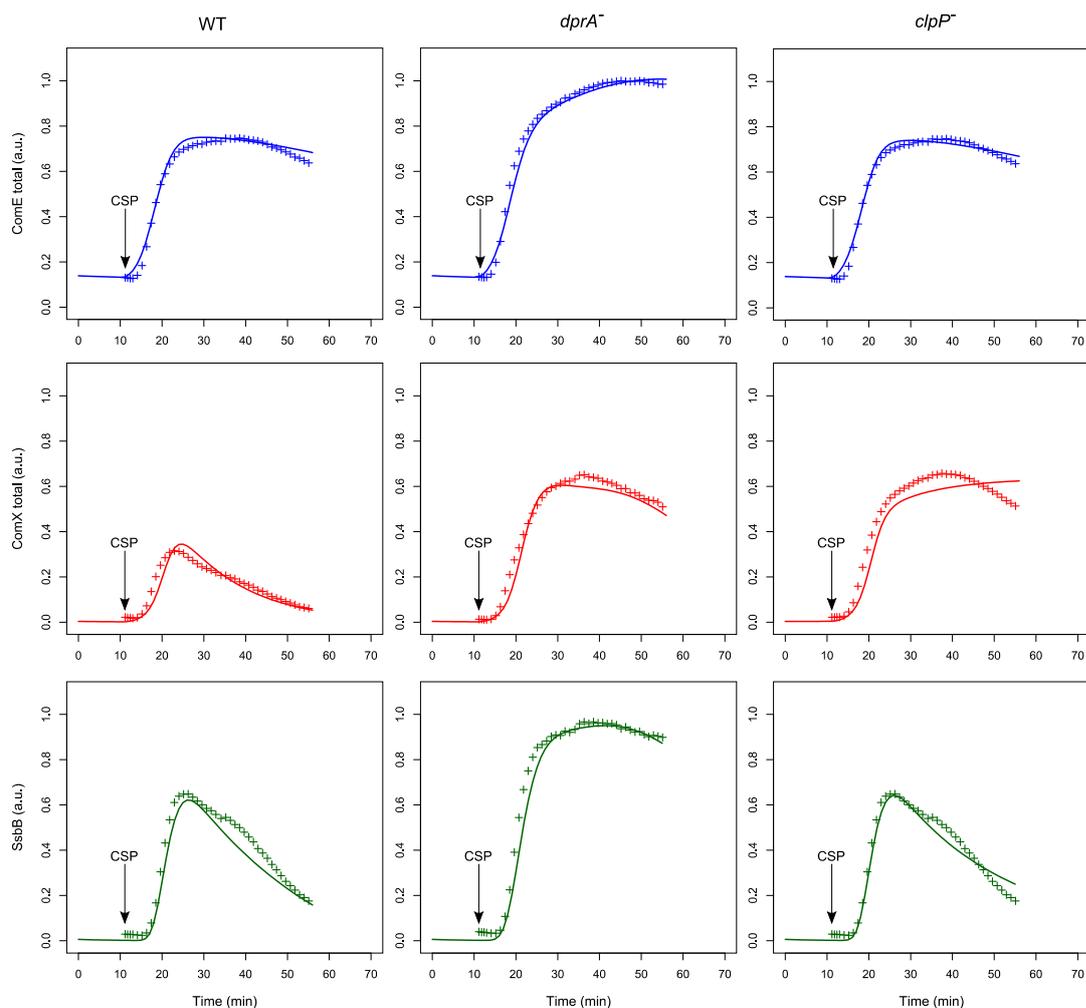
$$\frac{d[\text{ComZ}]}{dt} = v_{\text{max}_{\text{ComZ}}} * \frac{[\text{ComX}_{\text{act}}]^z}{[\text{ComX}_{\text{act}}]^z + K_{\text{ComZ}}^z} - \omega_2 * [\text{ComZ}] * [\text{ComW}_{\text{act}}] - \gamma_{\text{ComZ}} * [\text{ComZ}]$$

$$\frac{d[\text{ComW}_{\text{act}}]}{dt} = v_{\text{max}_{\text{ComW}}} \frac{[(\text{ComE} \sim \text{P})_D]^w}{[(\text{ComE} \sim \text{P})_D]^w + K_{\text{ComW}}^w} - \omega_2 * [\text{ComZ}] * [\text{ComW}_{\text{act}}] - \gamma_{\text{ComW}} * [\text{ComW}_{\text{ina}}]$$

$$\frac{d[\text{ComW}_{\text{ina}}]}{dt} = \omega_2 * [\text{ComZ}] * [\text{ComW}_{\text{act}}] - \gamma_{\text{ComW}} * [\text{ComW}_{\text{ina}}]$$

$$[\text{ComW}]_{\text{total}} = [\text{ComW}_{\text{ina}}] + [\text{ComW}_{\text{act}}]$$

Nouvelles simulations

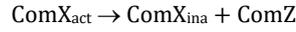


**Tableau 4. Réactions modifiées lors d'une compétition entre ComW et un produit de gène tardif de la compétence ComZ empêchant la formation d'une forme active de ComX (Figure 23 (2))**

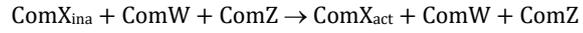
---

Nouvelles réactions

**Synthèse de ComZ**



**Inhibition de l'activation de ComX**



Nouvelles équations

$$\frac{d[\text{ComZ}]}{dt} = v_{\text{max}_{\text{comZ}}} * \frac{[\text{ComX}_{\text{act}}]^z}{[\text{ComX}_{\text{act}}]^z + K_{\text{comZ}}^z} - \gamma_{\text{ComZ}} * [\text{ComZ}]$$

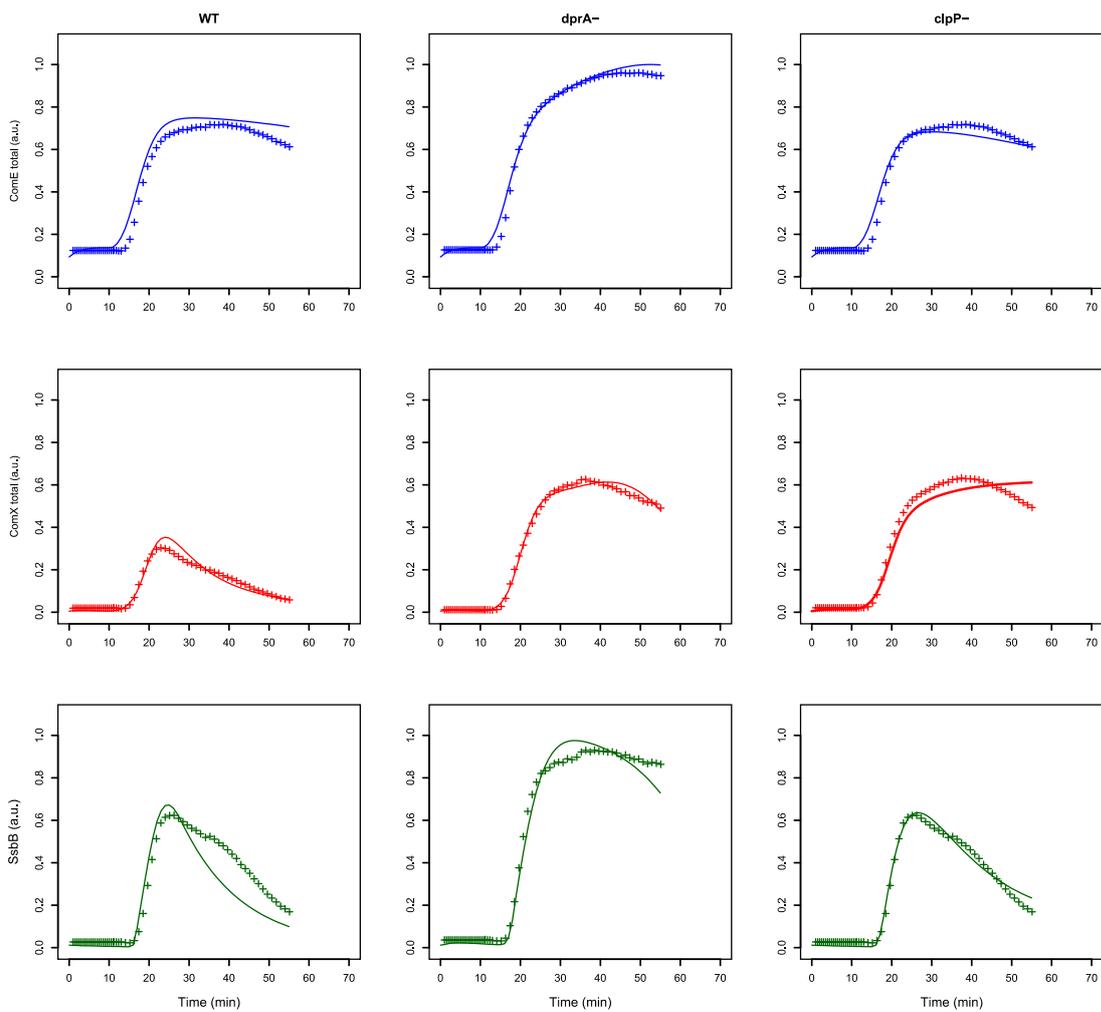
$$\begin{aligned} \frac{d[\text{ComX}_{\text{act}}]}{dt} = & v_{\text{max}_{\text{act}}} * \frac{[\text{ComW}_{\text{act}}]}{[\text{ComW}_{\text{act}}] + K_{\text{ComX}_{\text{act}}}} * \frac{K_{\text{ComX}_{\text{ini}}}}{[\text{ComZ}] + K_{\text{ComX}_{\text{ini}}}} * [\text{ComX}_{\text{ina}}] \\ & - v_{\text{max}_{\text{DprA}}} * \frac{[\text{ComX}_{\text{act}}]^d}{[\text{ComX}_{\text{act}}]^d + K_{\text{DprA}}^d} - v_{\text{max}_{\text{SsbB}}} * \frac{[\text{ComX}_{\text{act}}]^s}{[\text{ComX}_{\text{act}}]^s + K_{\text{SsbB}}^s} \\ & - v_{\text{max}_{\text{comZ}}} * \frac{[\text{ComX}_{\text{act}}]^z}{[\text{ComX}_{\text{act}}]^z + K_{\text{comZ}}^z} - \gamma_{\text{ComX}} * [\text{ComX}_{\text{act}}] \end{aligned}$$

$$\begin{aligned} \frac{d[\text{ComX}_{\text{ina}}]}{dt} = & v_{\text{max}_{\text{ComX}}} * \frac{[(\text{ComE} \sim \text{P})_D]^x}{[(\text{ComE} \sim \text{P})_D]^x + K_{\text{ComX}}^x} + v_{\text{max}_{\text{DprA}}} * \frac{[\text{ComX}_{\text{act}}]^d}{[\text{ComX}_{\text{act}}]^d + K_{\text{DprA}}^d} \\ & + v_{\text{max}_{\text{SsbB}}} * \frac{[\text{ComX}_{\text{act}}]^s}{[\text{ComX}_{\text{act}}]^s + K_{\text{SsbB}}^s} + v_{\text{max}_{\text{comZ}}} * \frac{[\text{ComX}_{\text{act}}]^z}{[\text{ComX}_{\text{act}}]^z + K_{\text{comZ}}^z} \\ & - v_{\text{max}_{\text{act}}} * \frac{[\text{ComW}_{\text{act}}]}{[\text{ComW}_{\text{act}}] + K_{\text{ComX}_{\text{act}}}} * \frac{K_{\text{ComX}_{\text{ini}}}}{[\text{ComZ}] + K_{\text{ComX}_{\text{ini}}}} * [\text{ComX}_{\text{ina}}] \\ & - \gamma_{\text{ComX}} * [\text{ComX}_{\text{ina}}] \end{aligned}$$


---

---

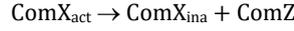
Nouvelles simulations



**Tableau 5. Réactions modifiées lors d'une compétition entre la forme active de ComX et un produit de gène tardif ComZ pour l'ARN polymérase (voir Figure 23 (4)).**

Nouvelles réactions

**Synthèse de ComZ**



**Synthèse de SsbB**



**Synthèse de DprA**



Nouvelles équations

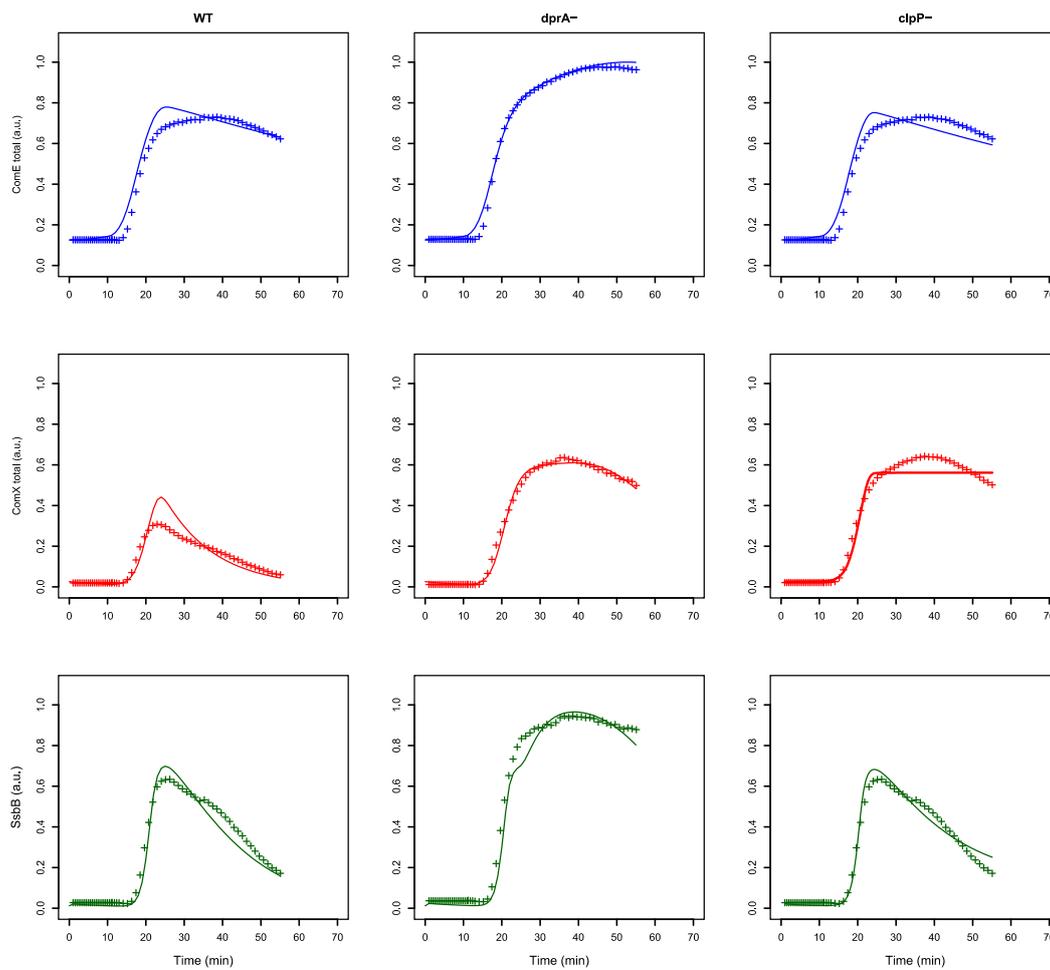
$$\frac{d[\text{ComZ}]}{dt} = \beta_{\text{ComZ}} + v_{\text{max}_{\text{ComZ}}} * \frac{[\text{ComX}_{\text{act}}]^z}{[\text{ComX}_{\text{act}}]^z + K_{\text{ComZ}}^z} \left( 1 - \frac{[\text{ComZ}]^i}{[\text{ComZ}]^i + K_{i_{\text{ComZ}}}} \right) - \gamma_{\text{ComZ}} * [\text{ComZ}]$$

$$\frac{d[\text{SsbB}]}{dt} = v_{\text{max}_{\text{SsbB}}} * \frac{[\text{ComX}_{\text{act}}]^s}{[\text{ComX}_{\text{act}}]^s + K_{\text{SsbB}}^s} * \left( 1 - \frac{[\text{ComZ}]^i}{[\text{ComZ}]^i + K_{i_{\text{SsbB}}}} \right) - \gamma_{\text{SsbB}} * [\text{SsbB}]$$

$$\begin{aligned} \frac{d[\text{DprA}]}{dt} = & v_{\text{max}_{\text{DprA}}} * \frac{[\text{ComX}_{\text{act}}]^d}{[\text{ComX}_{\text{act}}]^d + K_{\text{DprA}}^d} * \left( 1 - \frac{[\text{ComZ}]^u}{[\text{ComZ}]^u + K_{i_{\text{DprA}}}} \right) \\ & - 2 * k_{\text{on}_{\text{DprA}_{\text{EP}}}} * [\text{DprA}]^2 * [(\text{ComE} \sim \text{P})_D] - \gamma_{\text{DprA}} * [\text{DprA}] \end{aligned}$$

$$\begin{aligned} \frac{d[\text{ComX}_{\text{act}}]}{dt} = & \omega_1 * [\text{ComW}] * [\text{ComX}_{\text{ina}}] - v_{\text{max}_{\text{DprA}}} * \frac{[\text{ComX}_{\text{act}}]^d}{[\text{ComX}_{\text{act}}]^d + K_{\text{DprA}}^d} \left( 1 - \frac{[\text{ComZ}]^u}{[\text{ComZ}]^u + K_{i_{\text{DprA}}}} \right) \\ & - v_{\text{max}_{\text{SsbB}}} * \frac{[\text{ComX}_{\text{act}}]^s}{[\text{ComX}_{\text{act}}]^s + K_{\text{SsbB}}^s} \left( 1 - \frac{[\text{ComZ}]^i}{[\text{ComZ}]^i + K_{i_{\text{SsbB}}}} \right) \\ & - v_{\text{max}_{\text{ComZ}}} * \frac{[\text{ComX}_{\text{act}}]^z}{[\text{ComX}_{\text{act}}]^z + K_{\text{ComZ}}^z} \left( 1 - \frac{[\text{ComZ}]^i}{[\text{ComZ}]^i + K_{i_{\text{ComZ}}}} \right) - \gamma_{\text{ComX}} * [\text{ComX}_{\text{act}}] \end{aligned}$$

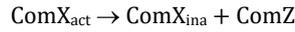
$$\begin{aligned} \frac{d[\text{ComX}_{\text{ina}}]}{dt} = & v_{\text{max}_{\text{ComX}}} * \frac{[(\text{ComE} \sim \text{P})_D]^x}{[(\text{ComE} \sim \text{P})_D]^x + K_{\text{ComX}}^x} - \omega_1 * [\text{ComW}] * [\text{ComX}_{\text{ina}}] \\ & + v_{\text{max}_{\text{DprA}}} * \frac{[\text{ComX}_{\text{act}}]^d}{[\text{ComX}_{\text{act}}]^d + K_{\text{DprA}}^d} \left( 1 - \frac{[\text{ComZ}]^u}{[\text{ComZ}]^u + K_{i_{\text{DprA}}}} \right) \\ & + v_{\text{max}_{\text{SsbB}}} * \frac{[\text{ComX}_{\text{act}}]^s}{[\text{ComX}_{\text{act}}]^s + K_{\text{SsbB}}^s} \left( 1 - \frac{[\text{ComZ}]^i}{[\text{ComZ}]^i + K_{i_{\text{SsbB}}}} \right) \\ & + v_{\text{max}_{\text{ComZ}}} * \frac{[\text{ComX}_{\text{act}}]^z}{[\text{ComX}_{\text{act}}]^z + K_{\text{ComZ}}^z} \left( 1 - \frac{[\text{ComZ}]^i}{[\text{ComZ}]^i + K_{i_{\text{ComZ}}}} \right) - \gamma_{\text{ComX}} * [\text{ComX}_{\text{ina}}] \end{aligned}$$



**Tableau 6. Réactions modifiées lors d'une inhibition de la forme active de ComX par un gène tardif de la compétence ComZ (voir Figure 23 (3)).**

Nouvelles réactions

**Synthèse de ComZ**



**Inhibition de ComX**



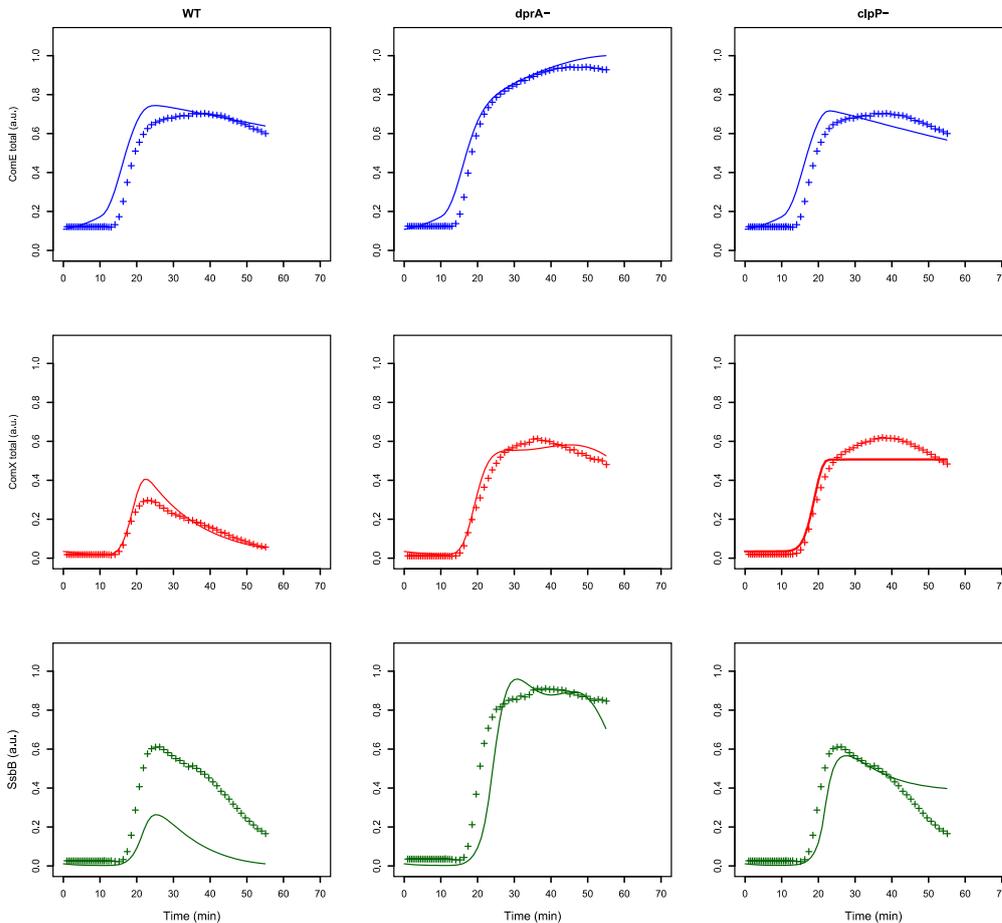
Nouvelles équations

$$\frac{d[\text{ComZ}]}{dt} = v_{\text{max}_{\text{ComZ}}} * \frac{[\text{ComX}_{\text{act}}]^z}{[\text{ComX}_{\text{act}}]^z + K_{\text{ComZ}}^z} - \omega_2 * [\text{ComZ}] * [\text{ComW}_{\text{act}}] - \gamma_{\text{ComZ}} * [\text{ComZ}]$$

$$\begin{aligned} \frac{d[\text{ComX}_{\text{act}}]}{dt} = & \omega_1 * [\text{ComW}] * [\text{ComX}_{\text{ina}}] - \omega_2 * [\text{ComX}_{\text{act}}] * [\text{ComZ}] \\ & - v_{\text{max}_{\text{DprA}}} * \frac{[\text{ComX}_{\text{act}}]^d}{[\text{ComX}_{\text{act}}]^d + K_{\text{DprA}}^d} - v_{\text{max}_{\text{SsbB}}} * \frac{[\text{ComX}_{\text{act}}]^s}{[\text{ComX}_{\text{act}}]^s + K_{\text{SsbB}}^s} \\ & - v_{\text{max}_{\text{ComZ}}} * \frac{[\text{ComX}_{\text{act}}]^z}{[\text{ComX}_{\text{act}}]^z + K_{\text{ComZ}}^z} - \gamma_{\text{ComX}} * [\text{ComX}_{\text{act}}] \end{aligned}$$

$$\begin{aligned} \frac{d[\text{ComX}_{\text{ina}}]}{dt} = & v_{\text{max}_{\text{ComX}}} * \frac{[(\text{ComE} \sim \text{P})_D]^x}{[(\text{ComE} \sim \text{P})_D]^x + K_{\text{ComX}}^x} + v_{\text{max}_{\text{DprA}}} * \frac{[\text{ComX}_{\text{act}}]^d}{[\text{ComX}_{\text{act}}]^d + K_{\text{DprA}}^d} \\ & + v_{\text{max}_{\text{SsbB}}} * \frac{[\text{ComX}_{\text{act}}]^s}{[\text{ComX}_{\text{act}}]^s + K_{\text{SsbB}}^s} + v_{\text{max}_{\text{ComZ}}} * \frac{[\text{ComX}_{\text{act}}]^z}{[\text{ComX}_{\text{act}}]^z + K_{\text{ComZ}}^z} \\ & + \omega_2 * [\text{ComX}_{\text{act}}] * [\text{ComZ}] - \omega_1 * [\text{ComW}] * [\text{ComX}_{\text{ina}}] - \gamma_{\text{ComX}} * [\text{ComX}_{\text{ina}}] \end{aligned}$$

Nouvelles simulations



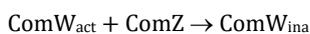
**Tableau 7. Réactions modifiées lors d'une interaction entre ComW et un produit de gène précoce ComZ inhibant l'activité de ComW (voir Figure 23 (1)).**

Nouvelles réactions

**Synthèse de ComZ**



**Inhibition de ComW**



Nouvelles équations

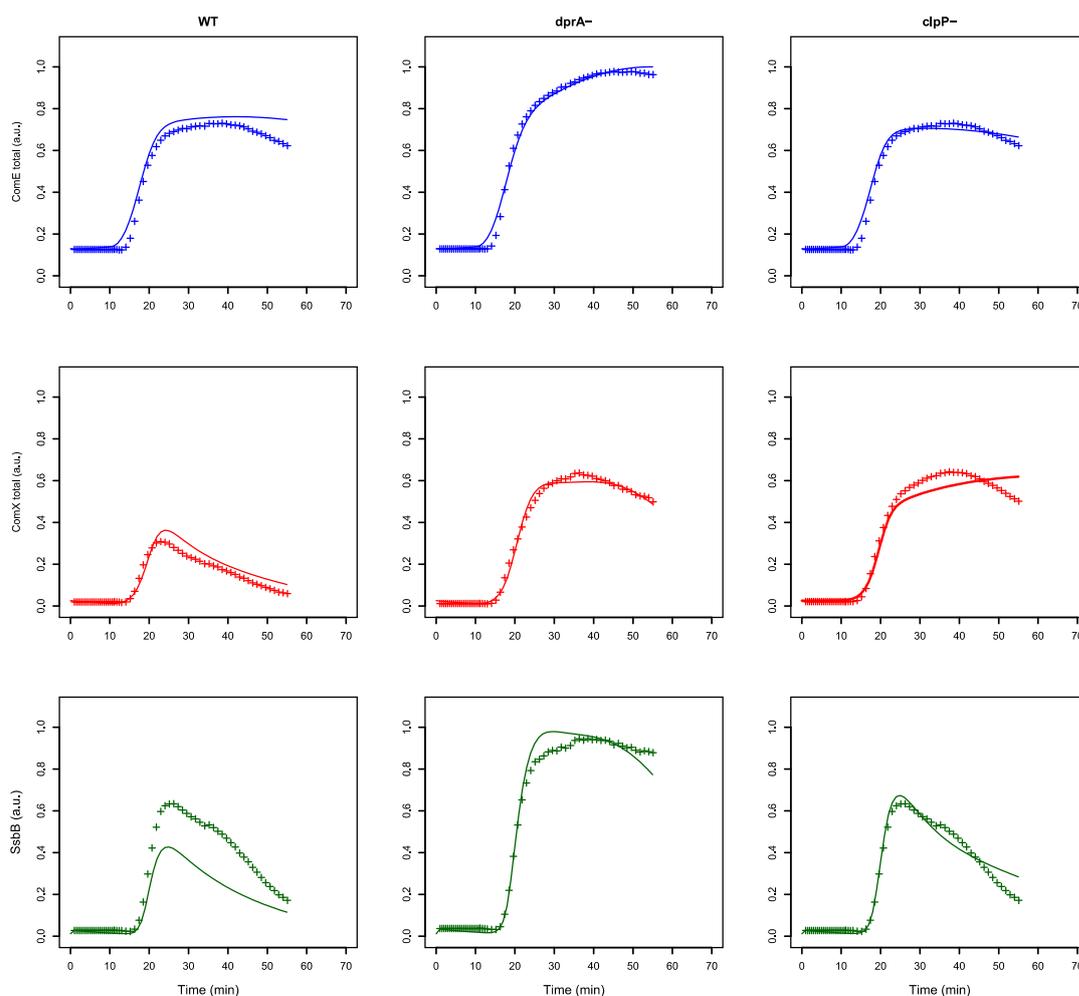
$$\frac{d[\text{ComZ}]}{dt} = v_{max_{ComZ}} * \frac{[\text{ComE} \sim \text{P}]^Z}{[\text{ComE} \sim \text{P}]^Z + K_{ComZ}^Z} - \omega_2 * [\text{ComZ}] * [\text{ComW}_{act}] - \gamma_{ComZ} * [\text{ComZ}]$$

$$\frac{d[\text{ComW}_{act}]}{dt} = v_{max_{ComW}} * \frac{[(\text{ComE} \sim \text{P})_D]^w}{[(\text{ComE} \sim \text{P})_D]^w + K_{ComW}^w} - \omega_2 * [\text{ComZ}] * [\text{ComW}_{act}] - \gamma_{ComW} * [\text{ComW}_{act}]$$

$$\frac{d[\text{ComW}_{ina}]}{dt} = \omega_2 * [\text{ComZ}] * [\text{ComW}_{act}] - \gamma_{ComW} * [\text{ComW}_{ina}]$$

$$[\text{ComW}]_{total} = [\text{ComW}_{ina}] + [\text{ComW}_{act}]$$

Nouvelles simulations



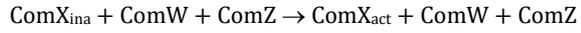
**Tableau 8. Réactions modifiées lors d'une compétition entre ComW et un produit de gène précoce de la compétence ComZ empêchant la formation d'une forme active de ComX (voir Figure 23 (2)).**

Nouvelles réactions

Synthèse de ComZ



Inhibition de l'activation de ComX



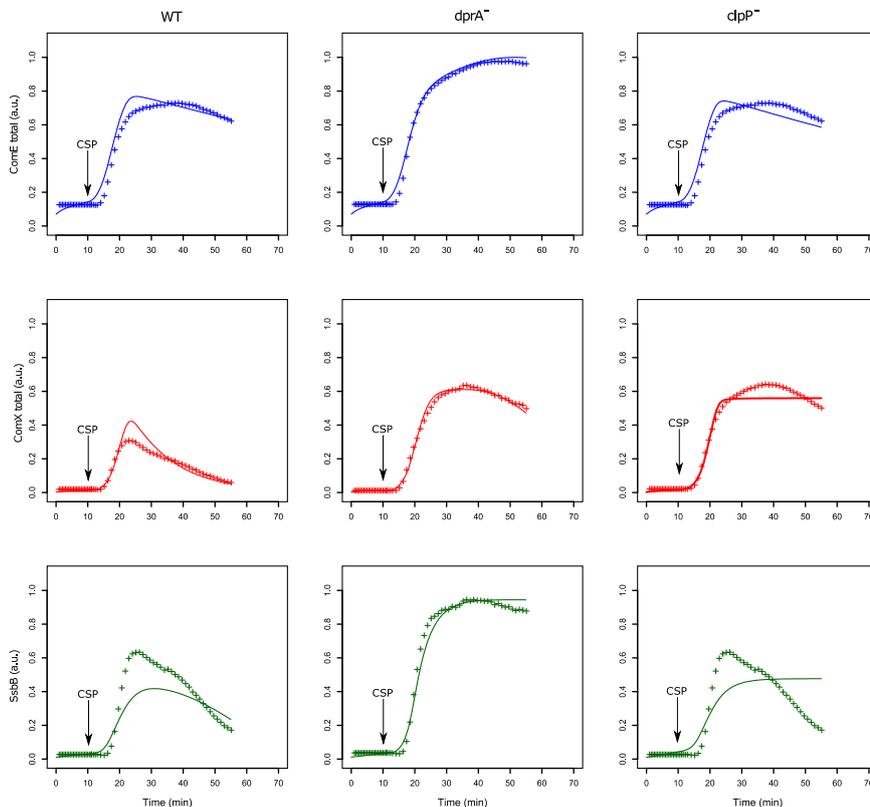
Nouvelles équations

$$\frac{d[ComZ]}{dt} = v_{max_{ComZ}} * \frac{[ComE\sim P]^Z}{[ComE\sim P]^Z + K_{ComZ}^Z} - \gamma_{ComZ} * [ComZ]$$

$$\begin{aligned} \frac{d[ComX_{act}]}{dt} = & v_{max_{act}} * \frac{[ComW_{act}]}{[ComW_{act}] + K_{ComX_{act}}} * \frac{K_{ComX_{ini}}}{[ComZ] + K_{ComX_{ini}}} * [ComX_{ina}] \\ & - v_{max_{DprA}} * \frac{[ComX_{act}]^d}{[ComX_{act}]^d + K_{DprA}^d} - v_{max_{SsbB}} * \frac{[ComX_{act}]^s}{[ComX_{act}]^s + K_{SsbB}^s} \\ & - v_{max_{ComZ}} * \frac{[ComX_{act}]^z}{[ComX_{act}]^z + K_{ComZ}^z} - \gamma_{ComX} * [ComX_{act}] \end{aligned}$$

$$\begin{aligned} \frac{d[ComX_{ina}]}{dt} = & v_{max_{ComX}} * \frac{[(ComE\sim P)_D]^x}{[(ComE\sim P)_D]^x + K_{ComX}^x} + v_{max_{DprA}} * \frac{[ComX_{act}]^d}{[ComX_{act}]^d + K_{DprA}^d} \\ & + v_{max_{SsbB}} * \frac{[ComX_{act}]^s}{[ComX_{act}]^s + K_{SsbB}^s} + v_{max_{ComZ}} * \frac{[ComX_{act}]^z}{[ComX_{act}]^z + K_{ComZ}^z} \\ & - v_{max_{act}} * \frac{[ComW_{act}]}{[ComW_{act}] + K_{ComX_{act}}} * \frac{K_{ComX_{ini}}}{[ComZ] + K_{ComX_{ini}}} * [ComX_{ina}] \\ & - \gamma_{ComX} * [ComX_{ina}] \end{aligned}$$

Nouvelles simulations



**Tableau 9. Réactions modifiées lors d'une compétition entre la forme active de ComX et un produit de gène précoce ComZ pour l'ARN polymérase (voir Figure 23 (4)).**

---

Nouvelles réactions



Nouvelles équations

$$\frac{d[ComZ]}{dt} = \beta_{ComZ} + v_{max_{ComZ}} * \frac{[ComE \sim P]^Z}{[ComE \sim P]^Z + K_{ComZ}^Z} - \gamma_{ComZ} * [ComZ]$$

$$\frac{d[SsbB]}{dt} = v_{max_{SsbB}} * \frac{[ComX_{act}]^s}{[ComX_{act}]^s + K_{SsbB}^s} * \left(1 - \frac{[ComZ]^i}{[ComZ]^i + K_{i_{SsbB}}^i}\right) - \gamma_{SsbB} * [SsbB]$$

$$\begin{aligned} \frac{d[DprA]}{dt} = & v_{max_{DprA}} * \frac{[ComX_{act}]^d}{[ComX_{act}]^d + K_{DprA}^d} * \left(1 - \frac{[ComZ]^u}{[ComZ]^u + K_{i_{DprA}}^u}\right) \\ & - 2 * k_{on_{DprA_{EP}}} * [DprA]^2 * [(ComE \sim P)_D] - \gamma_{DprA} * [DprA] \end{aligned}$$

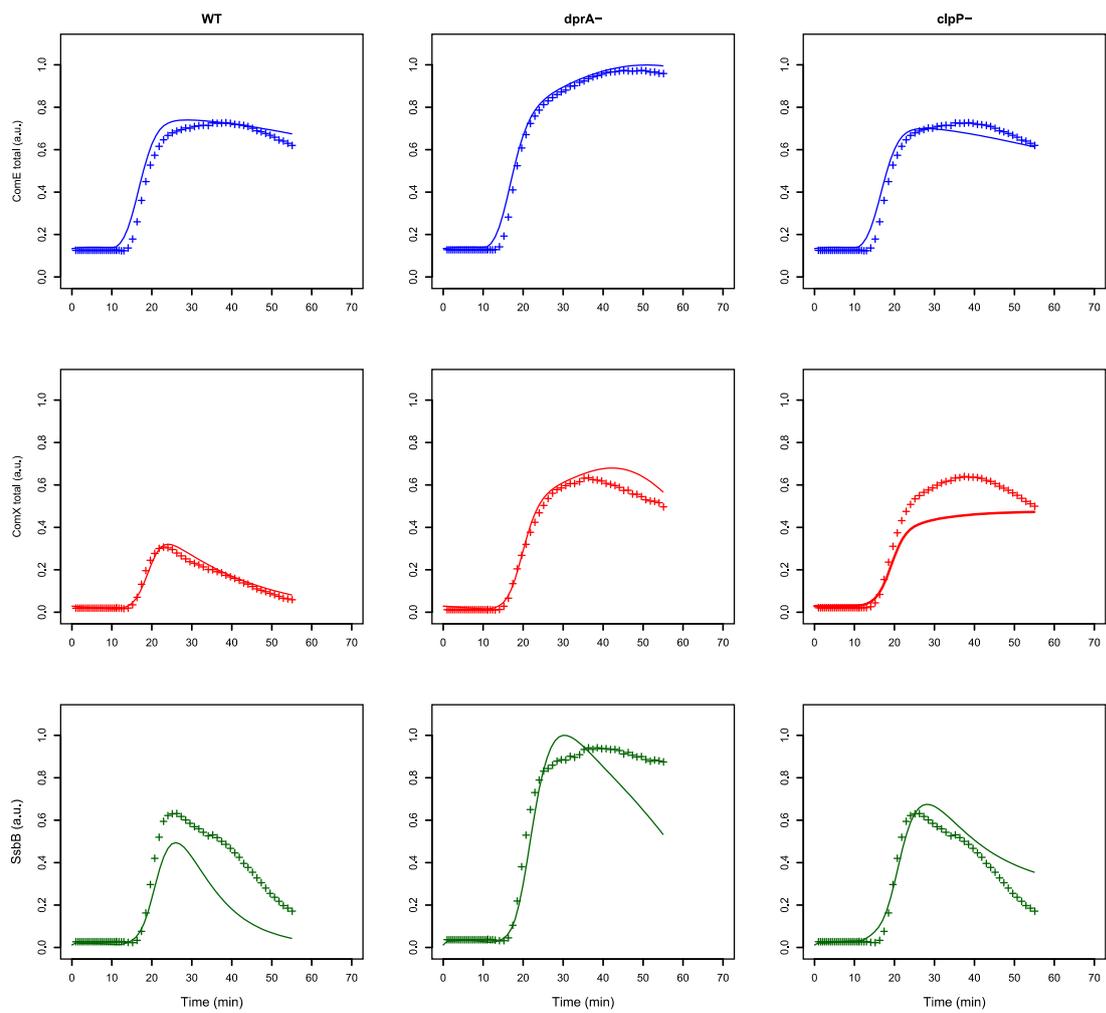
$$\begin{aligned} \frac{d[ComX_{act}]}{dt} = & \omega_1 * [ComW] * [ComX_{ina}] \\ & - v_{max_{DprA}} * \frac{[ComX_{act}]^d}{[ComX_{act}]^d + K_{DprA}^d} * \left(1 - \frac{[ComZ]^u}{[ComZ]^u + K_{i_{DprA}}^u}\right) \\ & - v_{max_{SsbB}} * \frac{[ComX_{act}]^s}{[ComX_{act}]^s + K_{SsbB}^s} * \left(1 - \frac{[ComZ]^i}{[ComZ]^i + K_{i_{SsbB}}^i}\right) - \gamma_{ComX} * [ComX_{act}] \end{aligned}$$

$$\begin{aligned} \frac{d[ComX_{ina}]}{dt} = & v_{max_{ComX}} * \frac{[(ComE \sim P)_D]^x}{[(ComE \sim P)_D]^x + K_{ComX}^x} - \omega_1 * [ComW] * [ComX_{ina}] \\ & + v_{max_{DprA}} * \frac{[ComX_{act}]^d}{[ComX_{act}]^d + K_{DprA}^d} * \left(1 - \frac{[ComZ]^u}{[ComZ]^u + K_{i_{DprA}}^u}\right) \\ & + v_{max_{SsbB}} * \frac{[ComX_{act}]^s}{[ComX_{act}]^s + K_{SsbB}^s} * \left(1 - \frac{[ComZ]^i}{[ComZ]^i + K_{i_{SsbB}}^i}\right) - \gamma_{ComX} * [ComX_{ina}] \end{aligned}$$


---

---

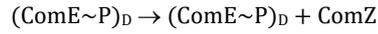
Nouvelles simulations



**Tableau 10. Réactions modifiées lors d'une inhibition de la forme active de ComX par un gène précoce de la compétence ComZ (voir Figure 23 (3)).**

Nouvelles réactions

**Synthèse de ComZ**



**Inhibition de ComX**



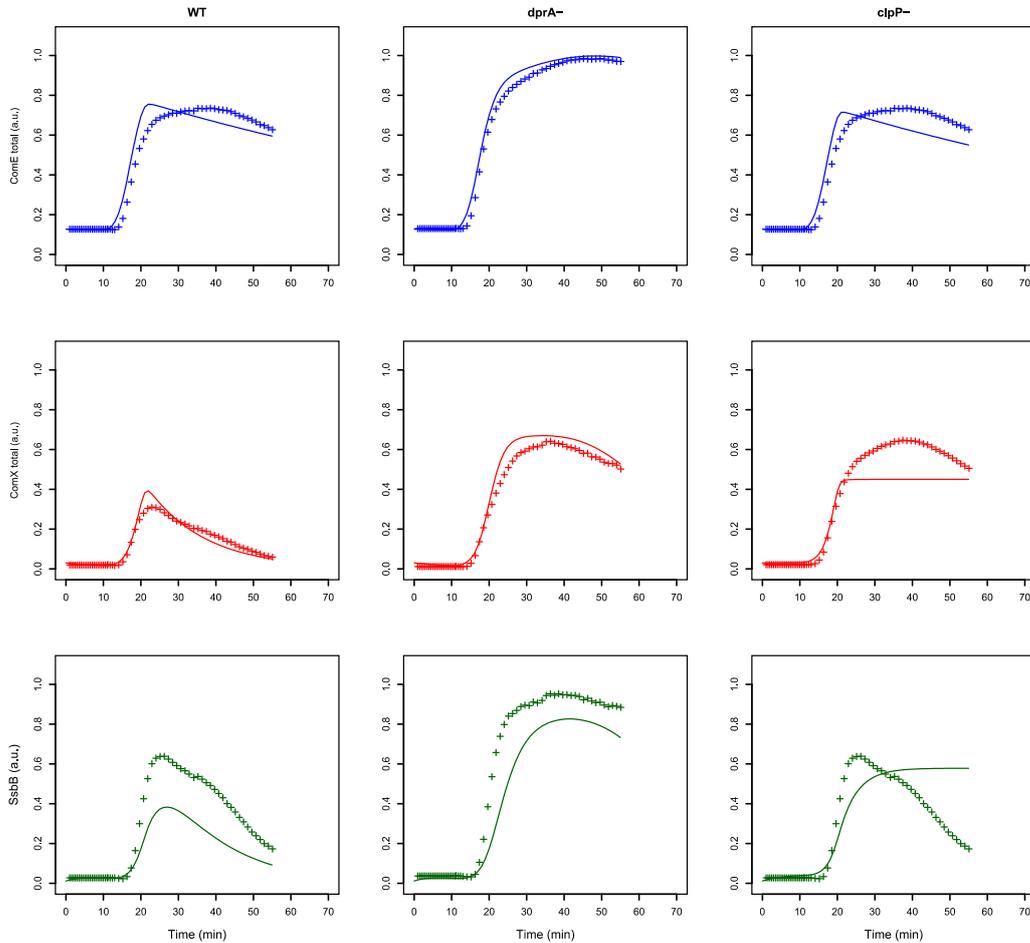
Nouvelles équations

$$\frac{d[\text{ComZ}]}{dt} = v_{max_{comZ}} * \frac{[\text{ComE}\sim\text{P}]^Z}{[\text{ComE}\sim\text{P}]^Z + K_{ComZ}^Z} - \omega_2 * [\text{ComX}_{act}] * [\text{ComZ}] - \gamma_{ComZ} * [\text{ComZ}]$$

$$\begin{aligned} \frac{d[\text{ComX}_{act}]}{dt} = & \omega_1 * [\text{ComW}] * [\text{ComX}_{ina}] - \omega_2 * [\text{ComX}_{act}] * [\text{ComZ}] \\ & - v_{max_{DprA}} * \frac{[\text{ComX}_{act}]^d}{[\text{ComX}_{act}]^d + K_{KprA}^d} - v_{max_{SsbB}} * \frac{[\text{ComX}_{act}]^s}{[\text{ComX}_{act}]^s + K_{KsbB}^s} \\ & - \gamma_{ComX} * [\text{ComX}_{act}] \end{aligned}$$

$$\begin{aligned} \frac{d[\text{ComX}_{ina}]}{dt} = & v_{max_{ComX}} * \frac{[(\text{ComE}\sim\text{P})_D]^x}{[(\text{ComE}\sim\text{P})_D]^x + K_{ComX}^x} + v_{max_{DprA}} * \frac{[\text{ComX}_{act}]^d}{[\text{ComX}_{act}]^d + K_{DprA}^d} \\ & + v_{max_{SsbB}} * \frac{[\text{ComX}_{act}]^s}{[\text{ComX}_{act}]^s + K_{SsbB}^s} + \omega_2 * [\text{ComX}_{act}] * [\text{ComZ}] \\ & - \omega_1 * [\text{ComW}] * [\text{ComX}_{ina}] - \gamma_{ComX} * [\text{ComX}_{ina}] \end{aligned}$$

Nouvelles simulations



Pour chacun des modèles, les changements dans les équations différentielles sont rapportés dans les **tableaux 3 à 10**. L'estimation des paramètres a été effectuée et les simulations sont réalisées avec le même protocole que celui décrit précédemment. Comme les modèles ne peuvent pas être directement comparés sur la base de leurs fonctions objectives (mesurant la somme des carrés des résidus entre mesures et simulations) car ces derniers diffèrent en nombre de paramètres, le calcul de l'AIC (Akaike's Information Criterion) a été considéré (**Tableau 11**). La valeur de l'AIC a été calculée pour chaque modèle candidat, et le modèle sélectionné a été celui ayant la valeur d'AIC la plus faible. Le modèle sélectionné correspond à celui où ComZ est le produit d'un gène tardif interagissant directement avec ComW et empêchant donc la formation d'une forme active de ComX (**Tableau 3**). Ce modèle, décrit plus en détail *infra*, est également testé dans d'autres conditions spécifiques qui ont été explorées expérimentalement. Il est apparu qu'une inhibition directe de la forme active de ComX par ComZ, que ce dernier soit régulé par ComE~P ou par ComX, pouvait être éliminée car les simulations reproduisent très mal les cinétiques expérimentales. Nous pouvons également remarquer que, pour un même mécanisme d'action de ComZ, les plus faibles valeurs d'AIC sont obtenues avec les modèles basés sur une expression de ComZ comme un produit de gène tardif de la compétence, ce qui signifie que ces modèles reproduisent mieux les données expérimentales que ceux où ComZ est régulé en tant que gène précoce de la compétence.

**Tableau 11. Calcul de l'AIC et adéquation des modèles aux données expérimentales.**

Modèle alternatif	RSS	Nombre de paramètres	AIC <sup>&amp;</sup>
ComZ (tardif) interagit avec ComW	0.33	44	-5760.948
ComZ (tardif) est en compétition avec ComW pour interagir avec la forme inactive de ComX	0.45	47	-5520.471
ComZ (tardif) et la forme active de ComX sont en compétition pour l'ARN polymérase	0.516	48	-5411.006
ComZ (précoce) interagit avec ComW	0.63	44	-5272.098
ComZ (précoce) est en compétition avec ComW pour interagir avec la forme inactive de ComX	1.55	47	-4585.479
ComZ (précoce) et la forme active de ComX sont en compétition pour l'ARN polymérase	1.69	48	-4518.105
ComZ (tardif) inhibe la forme active de ComX	2.46	44	-4242.277
Premier modèle (Figure 22)	2.78	40	-4157.826
ComZ (précoce) inhibe la forme active de ComX	3.32	44	-4015.626

<sup>&</sup>  $AIC = n \log\left(\frac{RSS}{n}\right) + 2K$  où  $K$  est le nombre de paramètres à estimer, dépendant du modèle ;  $n$  est le nombre d'observations (756).

Les valeurs des paramètres estimés (**Tableau 12**) sont en accord avec les mesures expérimentales. Les valeurs estimées de constantes de dégradations donnent une demi-vie

moyenne de ComD, ComE et DprA de 83 minutes qui ont été constatées comme stables sur une période d'au moins 80 minutes (Weng *et al.*, 2013, Mirouze *et al.*, 2013, Martin *et al.*, 2013) et une demi-vie moyenne de 8 minutes de ComX et ComW et 6,1 minutes pour SsbB dont les demi-vies se situent autour de 5 minutes (Piotrowsky *et al.*, 2009, Mirouze *et al.*, 2013). Les valeurs estimées de constante de liaison de ComE~P à  $P_{comC}$  (0.154554) montrent une réduction d'un facteur 3 par rapport à la valeur estimée de liaison de ComE à  $P_{comC}$  (0.437349) qui est du même ordre que la magnitude de réduction du  $K_d$  apparent pour  $P_{comC}$  (environ d'un facteur 4) montré pour le mutant ComE phospho-mimétique (Martin *et al.*, 2013) par rapport au  $K_d$  de ComE. La constante de liaison estimée de ComE~P à  $P_{comX}$  (4.48531) indique une augmentation d'un facteur 29 par rapport à son affinité pour  $P_{comC}$  ce qui est supérieur au facteur 10 rapporté par Martin et collaborateurs (2013). Enfin, la demi-vie de ComE~P est estimée à 2,23 minutes, proches des 3,9 minutes mesurées pour la forme phosphorylée de AgrA (Wang *et al.*, 2014) et la valeur de la constante de transphosphorylation  $\lambda$  a été estimée à  $1 \text{ min}^{-1}$ , proche de  $0,63 \text{ min}^{-1}$  mesuré pour le transfert du groupe phosphoryle du système à deux composantes AgrA/AgrC (Wang *et al.*, 2014).

**Tableau 12. Description des paramètres du modèle.**

Symbole	Description	Valeur <sup>&amp;</sup>	Note/Source
$\beta_{comCDE}$	Niveau basal d'expression de l'opéron <i>comCDE</i>	1.00009e-05 (a.u.min <sup>-1</sup> )	-
$\beta_{comAB}$	Niveau basal d'expression de l'opéron <i>comAB</i>	9.99959e-06 (a.u.min <sup>-1</sup> )	-
$\alpha_{auto}$	Constante d'autophosphorylation de ComD	9.99972e-07 (a.u.min <sup>-1</sup> )	-
$vmax_{comCDE}$	Vitesse maximale d'expression de l'opéron <i>comCDE</i> à partir du promoteur $P_{comC}$	7.18016 (a.u.min <sup>-1</sup> )	-
$vmax_{comAB}$	Vitesse maximale d'expression de l'opéron <i>comAB</i> à partir du promoteur $P_{comAB}$	1.62831 (a.u.min <sup>-1</sup> )	-
$vmax_{comX}$	Vitesse maximale d'expression de <i>comX</i> à partir du promoteur $P_{comX}$	5.46e+02 (a.u.min <sup>-1</sup> )	-
$vmax_{comW}$	Vitesse maximale d'expression de <i>comW</i> à partir du promoteur $P_{comW}$	3.07e+03 (a.u.min <sup>-1</sup> )	-
$vmax_{dprA}$	Vitesse maximale d'expression de <i>dprA</i> à partir du promoteur $P_{dprA}$	2.62e+01 (a.u.min <sup>-1</sup> )	-
$vmax_{ssbB}$	Vitesse maximale d'expression de <i>ssbB</i> à partir du promoteur $P_{ssbB}$	2.30e+02 (a.u.min <sup>-1</sup> )	-
$vmax_{comZ}$	Vitesse maximale d'expression de <i>comZ</i> à partir du promoteur $P_{comZ}$	4.48e+03 (a.u.min <sup>-1</sup> )	-
$vmax_{act}$	Vitesse maximale de conversion de (ComD) <sub>D</sub> en (ComD <sub>act</sub> ) <sub>D</sub> en utilisant le CSP comme substrat	10 (a.u.min <sup>-1</sup> )	-
$Ki_{ComE}$	Constante de liaison entre ComE et $P_{comC}$	0.437349 (a.u.)	Ordonné avec les autres constantes de liaison à partir des gels-retards de Martin <i>et al.</i> , 2013

$K_{ComE\sim P}$	Constante de liaison entre ComE~P et $P_{comC}$	0.154554 (a.u.)	Ordonné avec les autres constantes de liaison à partir des gels-retards de Martin <i>et al.</i> , 2013
$Ki_{ComE\_AB}$	Constante de liaison entre ComE et $P_{comAB}$	0.599997 (a.u.)	-
$K_{ComE\sim P\_AB}$	Constante de liaison entre ComE~P et $P_{comAB}$	0.11596 (a.u.)	-
$K_{ComX}$	Constante de liaison entre ComE~P et $P_{comX}$	4.48531 (a.u.)	Ordonné avec les autres constantes de liaison à partir des gels-retards de Martin <i>et al.</i> , 2013
$K_{ComW}$	Constante de liaison entre ComE~P et $P_{comW}$	7.43985 (a.u.)	-
$K_{DprA}$	Constante de liaison entre $ComX_{act}$ et $P_{dprA}$	5.75922 (a.u.)	-
$K_{SsbB}$	Constante de liaison entre $ComX_{act}$ et $P_{ssbB}$	1.08504 (a.u.)	-
$K_{ComZ}$	Constante de liaison entre $ComX_{act}$ et $P_{comZ}$	0.19036 (a.u.)	-
$K_{act}$	Quantité de CSP nécessaire pour atteindre la moitié de la vitesse maximale d'activation de $(ComD)_D$	4.42425 (a.u.)	-
$k_{on\_D}$	Constante d'association de l'homodimère $(ComD)_D$	0.00191809 (a.u.)	-
$k_{off\_D}$	Constante de dissociation de l'homodimère $(ComD)_D$	0.00165586 (a.u.)	-
$k_{on\_DprA\_EP}$	Constante de liaison entre ComE~P et DprA	0.999995	-
$\epsilon$	Constante d'export du pré-CSP (ComC)	0.000399909 (a.u.min <sup>-1</sup> )	-
$\lambda$	Constante de transphosphorylation	1 (min <sup>-1</sup> )	Transfert du groupement phosphorylé entre AgrC et AgrA avec $t_{1/2} = 66$ s (Wang <i>et al.</i> , 2014)
$\rho$	Vitesse de dephosphorylation spontanée de ComE~P	0.310407 (min <sup>-1</sup> )	La demi-vie de la protéine homologue AgrA sous la forme phosphorylée est environ 3,9 minutes (Wang <i>et al.</i> , 2014)
$\omega_1$	Constante de liaison entre $ComX_{ina}$ et ComW	8.9263 (a.u.)	-
$\omega_2$	Constante de liaison entre $ComW_{act}$ et ComZ	2.42e+02 (a.u.)	-
$e$	Coefficient de Hill	2.9127 (a.u.)	-
$x$	Coefficient de Hill	1.28427 (a.u.)	-
$d$	Coefficient de Hill	1 (a.u.)	-
$s$	Coefficient de Hill	1.7740 (a.u.) <sup>9</sup>	-
$w$	Coefficient de Hill	1.27205 (a.u.)	-
$z$	Coefficient de Hill	3.99964 (a.u.)	-
$\gamma_{preCSP}$	Constante de dégradation du pré-CSP	0.00833326 (min <sup>-1</sup> )	-
$\gamma_{CSP}$	Constante de dégradation du CSP	0.08 (min <sup>-1</sup> )	-
$\gamma_{ComAB}$	Constante de dégradation de ComAB	0.00806448 (min <sup>-1</sup> )	-
$\gamma_{ComD}$	Constante de dégradation pour ComD et $(ComD_{used})_D$	0.0083333 (min <sup>-1</sup> )	Pas de dégradation significative de ComD sur une longue période (Martin <i>et al.</i> , 2013)

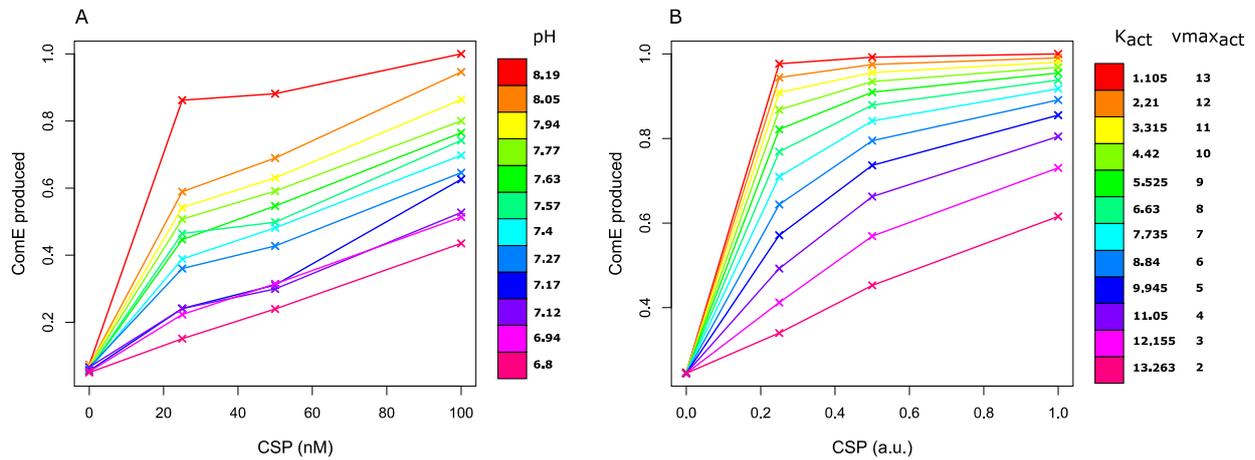
$\gamma_{ComE}$	Constante de dégradation de ComE	0.00833318 (min <sup>-1</sup> )	Pas de dégradation significative de ComE sur une longue (Ween <i>et al.</i> , 1999, Martin <i>et al.</i> , 2013)
$\gamma_{ComX}$	Constante de dégradation de ComX <sub>ina</sub> et ComX <sub>act</sub>	0.0859111 (min <sup>-1</sup> )	La demi-vie de ComX est estimée à environ 5 minutes (Weng <i>et al.</i> , 2013, Luo <i>et al.</i> , 2004)
$\gamma_{ComW}$	Constante de dégradation de ComW <sub>ina</sub> et ComW <sub>act</sub>	0.0859111 (min <sup>-1</sup> )	On a supposé la demi-vie de ComW égale à celle de ComX
$\gamma_{DprA}$	Constante de dégradation de DprA	0.00833343 (min <sup>-1</sup> )	Comme ComD et ComE, DprA est une protéine stable (Mirouze <i>et al.</i> , 2013)
$\gamma_{SsbB}$	Constante de dégradation de SsbB	0.113757 (min <sup>-1</sup> )	La demi-vie de SsbB est relativement courte (environ 5 minutes) (Mirouze <i>et al.</i> , 2013)
$\gamma_{ComZ}$	Constante de dégradation de ComZ	0.00833335 (min <sup>-1</sup> )	-

<sup>&</sup>a.u. = arbitrary units

Dans le cas du système sauvage mais aussi dans les mutants *dprA*<sup>-</sup> et *clpP*<sup>-</sup>, une très bonne adéquation est retrouvée entre les simulations et les cinétiques mesurées (**Tableau 3**). Avec ce modèle, la cinétique de la protéine SsbB montre qu'il est possible sortir de l'état de compétence aussi rapidement que dans un modèle sauvage malgré la stabilisation de ComX dans un mutant *clpP*<sup>-</sup>. Ces observations permettent de valider la topologie du modèle envisagé, composé de 17 équations et de 44 paramètres.

## 8. Étude de l'effet du pH du milieu sur le développement de la compétence

Pour valider le modèle, nous étudions ici l'impact du pH du milieu de croissance sur le déclenchement de la compétence car ce paramètre est un déterminant important du contrôle de la compétence (Chen et Morrison, 1987). Cependant, cet effet n'a jamais été analysé expérimentalement en détail. Il a été réalisé un ensemble d'expériences où l'activité du promoteur *comCDE* a été mesurée dans un contexte génétique qui ne permet pas l'export ou l'expression de CSP (souche *comA*<sup>-</sup>). Les populations cellulaires ont d'abord été cultivées dans les mêmes conditions pour permettre une expression basale du ComDE afin d'atteindre le même état physiologique, non compétent. Les échantillons ont ensuite été séparés et exposés à différents pH de milieu de croissance (allant de 6,8 à 8,19) et à différentes concentrations de CSP (0, 25, 50 et 100 ng / mL).



**Figure 24. La valeur de pH initiale du milieu de croissance affecte le développement de la compétence. (A)** Chaque point des courbes correspond à la quantité de ComE synthétisée sur une période de 20 minutes après addition de CSP pour chaque combinaison CSP-pH testée expérimentalement. Cette quantité est obtenue en calculant la surface sous la courbe de la cinétique de ComE et est normalisée par rapport à la valeur maximale de ComE synthétisée sur l'ensemble des expériences. Concentrations CSP utilisées dans les expériences : 0, 25, 50 et 100 ng / mL. Souche utilisée : *comC::luc* R1205 (*comAB*). **(B)** Chaque point des courbes correspond à la quantité simulée de  $ComE_{total}$  synthétisée sur une période de 20 minutes après addition de CSP. Les quantités de protéines sont calculées comme dans (A) et normalisées par rapport à la valeur maximale de ComE synthétisée sur l'ensemble des simulations. Les simulations sont effectuées en modifiant les deux paramètres correspondant à l'affinité d'un dimère de ComD pour le CSP ( $K_{act}$ ) et le taux d'autophosphorylation induit par CSP de ComD ( $vmax_{act}$ ). Quantité de CSP utilisée dans les simulations : 0, 0,25, 0,5, 1 a.u. Le mutant *comA<sup>-</sup>* est simulé en fixant  $\beta_{comAB} = 0$  (taux de synthèse basal de ComAB) et  $vmax_{comAB}$  (vitesse de synthèse maximale de ComAB).

Pour chaque combinaison de CSP-pH, les données de luminescence ont été traitées comme précédemment pour obtenir une cinétique moyenne de ComE par cellule. Ensuite, la quantité de ComE produite pendant les 20 premières minutes a été obtenue en calculant l'aire sous la courbe des cinétiques de ComE. Nous supposons que cette valeur est proportionnelle à la vitesse initiale d'accumulation de ComE, ainsi qu'à la force d'expression de *comCDE*. Les résultats (**Figure 24A**) montrent clairement que le pH initial du milieu de culture agit sur la vitesse initiale de développement de la compétence. Pour les valeurs les plus acides de pH ( $\leq 7.12$ ) et pour une valeur de CSP utilisée de 25 ng/mL pour l'induction du système, une très faible quantité de ComE est produite et, même à des concentrations plus importantes (100 ng/mL), la quantité produite de ComE atteint seulement la moitié de celle produite à pH 8,19. Plus particulièrement, une quantité encore plus élevée de ComE est obtenue pour des cellules induites avec 25 ng/mL de CSP à pH 7,57 que pour des cellules induites avec 100 ng/mL de CSP à pH 6,8. Ainsi pour une concentration de CSP donnée, plus le milieu de culture est basique, plus importante est la quantité de ComE produite. Cet effet est le plus prononcé à pH

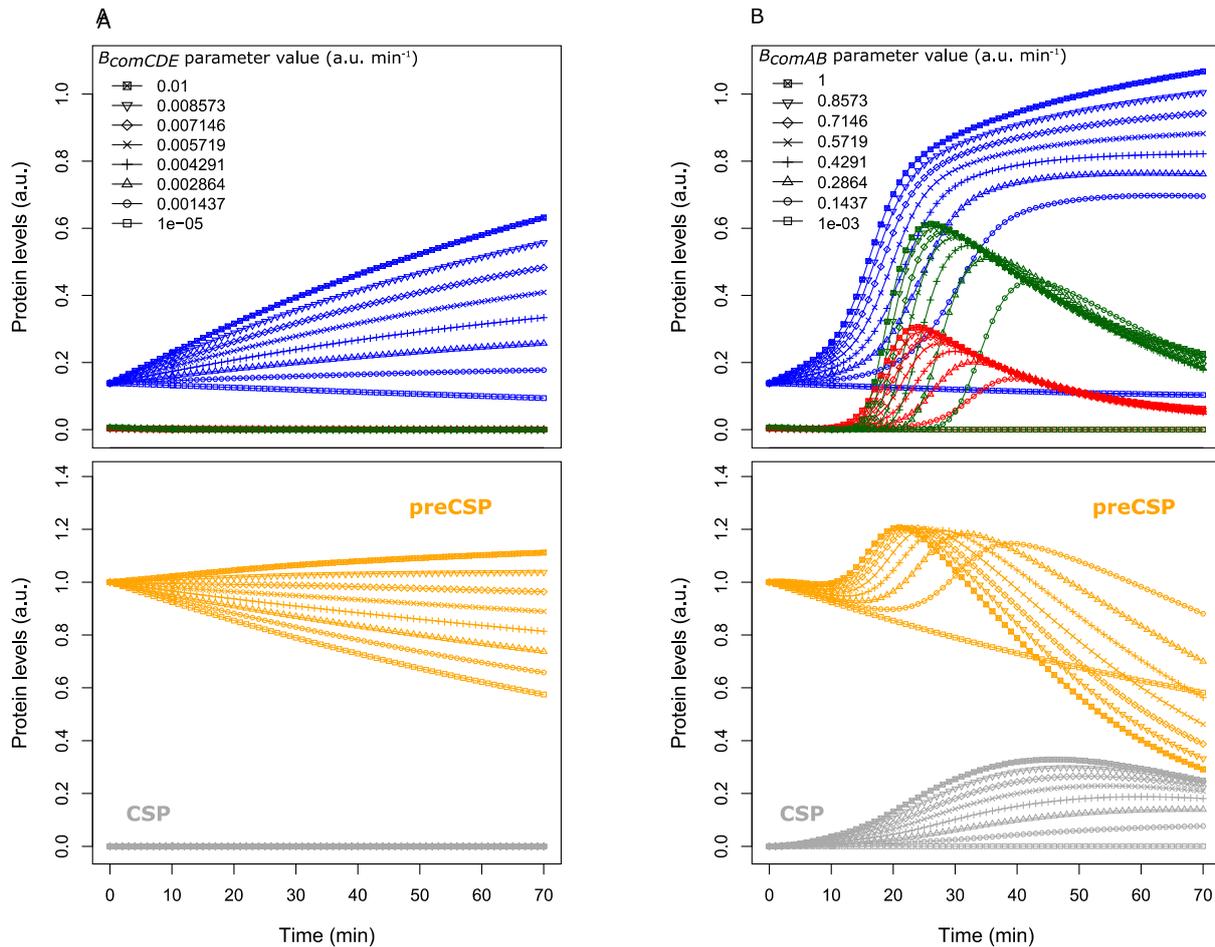
8.19 pour lequel la quantité de ComE produite atteint rapidement un plateau à une concentration de CSP de 25 ng/mL. Ces résultats confirment ainsi que la valeur du pH du milieu de culture est un paramètre important pour contrôler le déclenchement de la compétence.

Pour vérifier si le modèle pouvait reproduire et expliquer ces résultats expérimentaux, il a été supposé que le pH du milieu pouvait affecter les réactions qui prennent place dans l'environnement extracellulaire, respectivement : i) l'interaction entre le CSP et ComD, et donc avoir un effet sur la vitesse de phosphorylation de ComD (**Tableau 1**, réaction (5)), ou ii) la vitesse de dégradation du CSP (**Tableau 1**, réaction (12)). Comme le pH n'affecte pas l'activité de la protéase HtrA impliquée dans la dégradation du CSP (Cassone *et al.*, 2012), il a été supposé que la vitesse de dégradation du CSP n'était pas impactée par une variation du pH. Ainsi, pour mimer la variation du pH, un balayage des paramètres a été réalisé à la fois sur l'affinité du dimère de ComD pour le CSP ( $K_{act}$ ) et sur la vitesse maximale de transformation du dimère de ComD dans sa forme phosphorylée suite à son interaction avec le CSP ( $vmax_{act}$ ). Afin d'établir une échelle de paramètres, nous avons supposé que les valeurs  $K_{act}$  et  $vmax_{act}$  (respectivement 4.42 et 10) sont celles valables à pH 7,8, car les mesures réalisées pour les populations sauvages et *dprA*<sup>-</sup> ont été acquises à cette valeur de pH du milieu de culture (Mirouze *et al.*, 2013). Le balayage des paramètres a été réalisé pour 4 concentrations de CSP différentes pour induire le système. Après l'ajustement des paramètres, nous pouvons observer une bonne adéquation entre les courbes expérimentales et les courbes simulées (**Figure 24B**). Plus particulièrement, les simulations reproduisent correctement la forme des courbes expérimentales, notamment pour le pH le plus acide, avec une quantité de ComE qui augmente linéairement et atteint un maximum qui est environ la moitié de celui obtenu avec les valeurs de pH du milieu le plus alcalin. Dans ce cas, les valeurs de ComE atteignent rapidement un plateau pour une faible concentration de CSP, comme cela a été observé expérimentalement à pH 8,19. Ainsi, en adaptant les valeurs de deux paramètres du modèle modulant la phosphorylation de ComD par le CSP, il est possible de reproduire les résultats expérimentaux. Ces derniers viennent donc appuyer la validation de notre modèle. De plus, le modèle nous permet d'émettre l'hypothèse que la valeur du pH du milieu de croissance affecte l'activation de ComD par le CSP. ComD pourrait alors être considéré comme un senseur de pH.

Ce modèle validé en regard de nouvelles données expérimentales, plusieurs autres caractéristiques ont été explorées : l'effet du niveau basal d'expression des gènes précoces de la compétence, la cinétique du régulateur central ComE~P et enfin la période de non-réponse au CSP suivant l'induction de la compétence.

### 9. Effet du niveau basal d'expression de *comCDE*

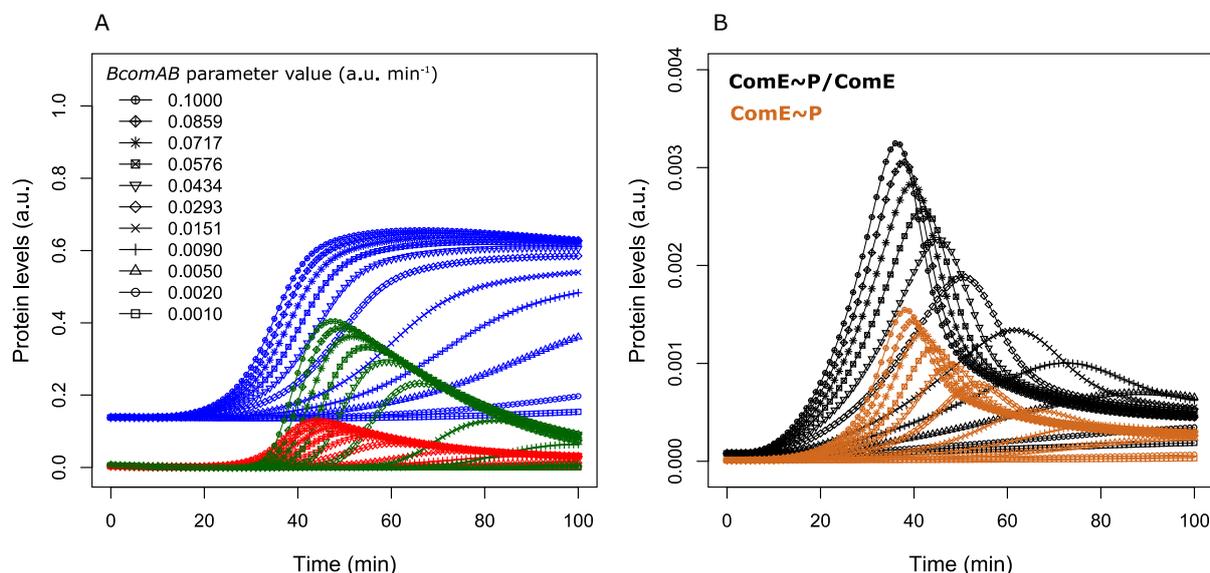
L'effet de l'augmentation du niveau basal de l'expression de l'opéron *comCDE* sur la régulation de la compétence a été testé avec, initialement, l'idée que la surexpression de *comCDE* pourrait permettre de dé-réprimer l'état de compétence. La dynamique du réseau a été suivie, sans ajout de CSP, pour induire le système. Un balayage des paramètres a été réalisé sur le niveau basal de transcription de l'opéron *comCDE* (de  $10^{-5}$  to  $10^{-2}$  a.u. min<sup>-1</sup>). Contrairement à ce qui pouvait être attendu, cette augmentation du niveau basal n'active pas la compétence (**Figure 25A**). Nous observons que la quantité intracellulaire de ComE et celle de pré-CSP intracellulaire s'accroissent avec l'augmentation du niveau basal d'expression de l'opéron, mais le niveau de CSP mûré et exporté reste très faible ( $\sim 10^{-5}$  a.u.). Cependant, lorsque l'on augmente simultanément le niveau basal d'expression des opérons *comCDE* et *comAB*, l'état de compétence est induit (**Figure 25B**). Nous pouvons déduire de ce résultat que l'export de CSP grâce à ComAB est limitant, comme déjà observé expérimentalement (Martin *et al.*, 2000), ce qui mène alors à un excès de ComE par rapport à ComE~P et empêche ComE~P d'entrer en compétition avec ComE. Ces résultats sont en outre cohérents avec les résultats expérimentaux qui montrent qu'une mutation dans le terminateur du tARN<sub>Arg</sub> localisé juste en amont de l'opéron *comCDE* mène à une augmentation de la transcription de cet opéron et inhibe l'activation de l'état de la compétence (Guiral *et al.*, 2006). L'effet de la surexpression des gènes *comAB* a aussi été démontré : il peut supprimer cet effet inhibiteur, ce qui suggère que ce phénotype résulte d'un déséquilibre entre la concentration intracellulaire de pré-CSP et la quantité de CSP mûre produit (Guiral *et al.*, 2006).



**Figure 25. Les vitesses basales d'expression de *comCDE* et de *comAB* influencent l'entrée dans l'état de compétence. (A)** La dynamique du réseau est suivie sans ajout de CSP avec balayage des paramètres sur le niveau basal d'expression de l'opéron *comCDE* de  $10^{-5}$  à  $10^{-2}$  a.u.  $\text{min}^{-1}$  sans modification des autres paramètres du modèle. Les cinétiques  $\text{ComE}_{\text{total}}$ ,  $\text{ComX}_{\text{total}}$  et SsbB sont représentées en utilisant le même code de couleurs que celui de la Figure 22. Les cinétiques pré-CSP (jaune) et CSP (gris) sont également rapportées. **(B)** Les mêmes simulations que dans (A) sont effectuées en faisant varier simultanément la vitesse de synthèse basale de ComAB de  $10^{-3}$  à 1 a.u.  $\text{min}^{-1}$ .

Ensuite, pour une valeur donnée du niveau basal de transcription de *comCDE* ( $0.000576$  a.u.  $\text{min}^{-1}$ ), le niveau basal d'expression de *comAB* a été augmenté de façon graduelle de  $10^{-3}$   $\text{min}^{-1}$  à  $10^{-1}$   $\text{min}^{-1}$ . La synthèse de ComE~P, ComX et SsbB est observée dès que le niveau basal de transcription de *comAB* atteint  $9 \cdot 10^{-3}$  a.u.  $\text{min}^{-1}$ , indiquant que l'activation de l'état de compétence peut-être restauré (**Figure 26A**). Cependant, plus le niveau basal d'expression de *comAB* est faible, plus un décalage et un retard du déclenchement de la compétence sont observés par rapport aux valeurs les plus élevées. Par exemple, l'état de compétence, correspondant à l'apparition de SsbB, se déclenche après 40 minutes avec un niveau basal d'expression égale à  $0,0434$  et seulement après 60 minutes avec un niveau basal égal à  $0,0151$ . Donc, ce retard décroît avec l'augmentation du niveau basal de transcription de l'opéron *comAB*. Cela est cohérent avec un export efficace de pré-CSP, permettant au CSP

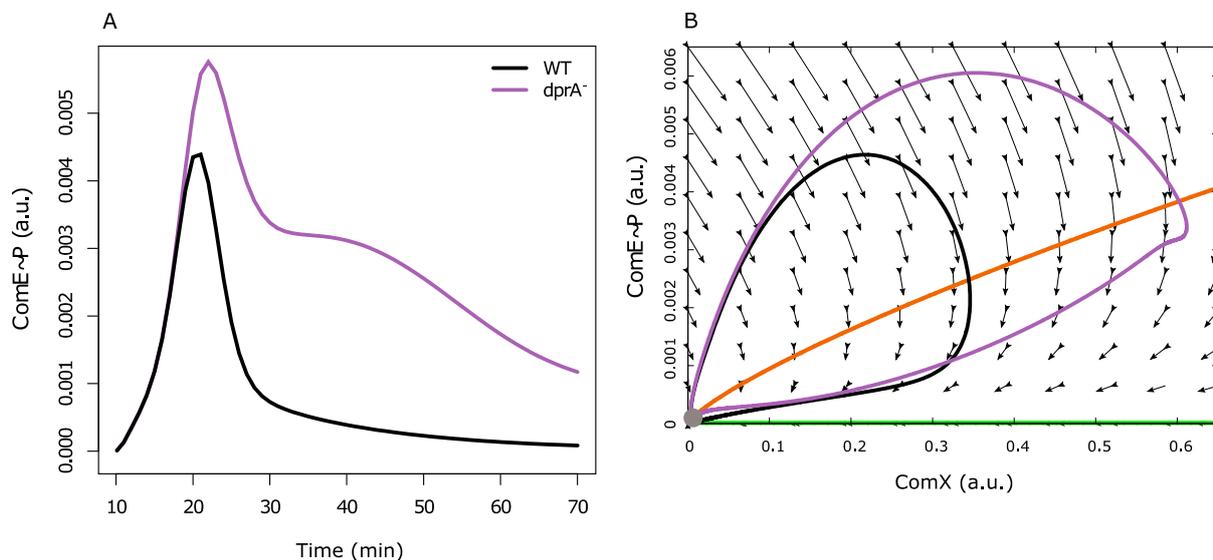
extracellulaire d'atteindre plus rapidement la concentration critique de CSP nécessaire pour activer le système ComD/ComE et pour mettre alors en place la boucle de rétroaction positive (**Figure 26B**). Pour chaque simulation, le ratio ComE~P/ComE a été calculé au moment où ComE~P commence à être synthétisé. Ce ratio varie entre  $10^{-4}$  et  $1,5 \cdot 10^{-4}$ . Nous pouvons conclure que, sous cette valeur, ComE~P n'est pas capable d'entrer en compétition avec ComE pour lier le promoteur  $P_{\text{comC}}$ .



**Figure 26. Le rapport ComE~P/ComE est important pour le développement spontané de la compétence.** (A) La dynamique du réseau, où le taux de synthèse basale du ComCDE est fixé à  $5.76 \cdot 10^{-3}$  a.u. min<sup>-1</sup>, est représentée pour un balayage de paramètres de  $10^{-3}$  à  $10^{-1}$  a.u. min<sup>-1</sup> du taux de synthèse basale de ComAB. Les cinétiques ComEtotal, ComXtotal et SsbB sont représentées en utilisant le même code de couleurs que la Figure 22. (B) Les courbes ComE~P (marron) et ComE~P/ComE (noir) sont présentées.

### 10. La cinétique de ComE~P et analyse de l'espace des phases

Comme ComE~P n'a jamais été purifié, sa cinétique ainsi que sa concentration relative par rapport à la forme non phosphorylée de ComE sont inconnues. ComE~P étant une variable d'état du modèle, nous accédons à sa cinétique dans les simulations (**Figure 27A**). D'abord, nous pouvons remarquer qu'une très faible quantité de ComE~P (environ 0,004 a.u. dans le modèle sauvage) est produite par rapport à la quantité de ComE<sub>total</sub> (presque 0,8 a.u. dans le modèle sauvage, voir **Tableau 3**). Cela est dû à une très faible durée de vie de la forme phosphorylée de ComE qui empêche son accumulation, ce qui peut aussi expliquer pourquoi la protéine n'a encore jamais été expérimentalement purifiée. Plus précisément, dans le modèle sauvage, la cinétique simulée de ComE~P présente un pic très étroit, avec un maximum centré 10 minutes après l'addition de CSP. Cette cinétique est étendue dans un mutant *dprA*<sup>-</sup>, ce qui explique la prolongation de la cinétique des gènes de la compétence dans ce mutant.



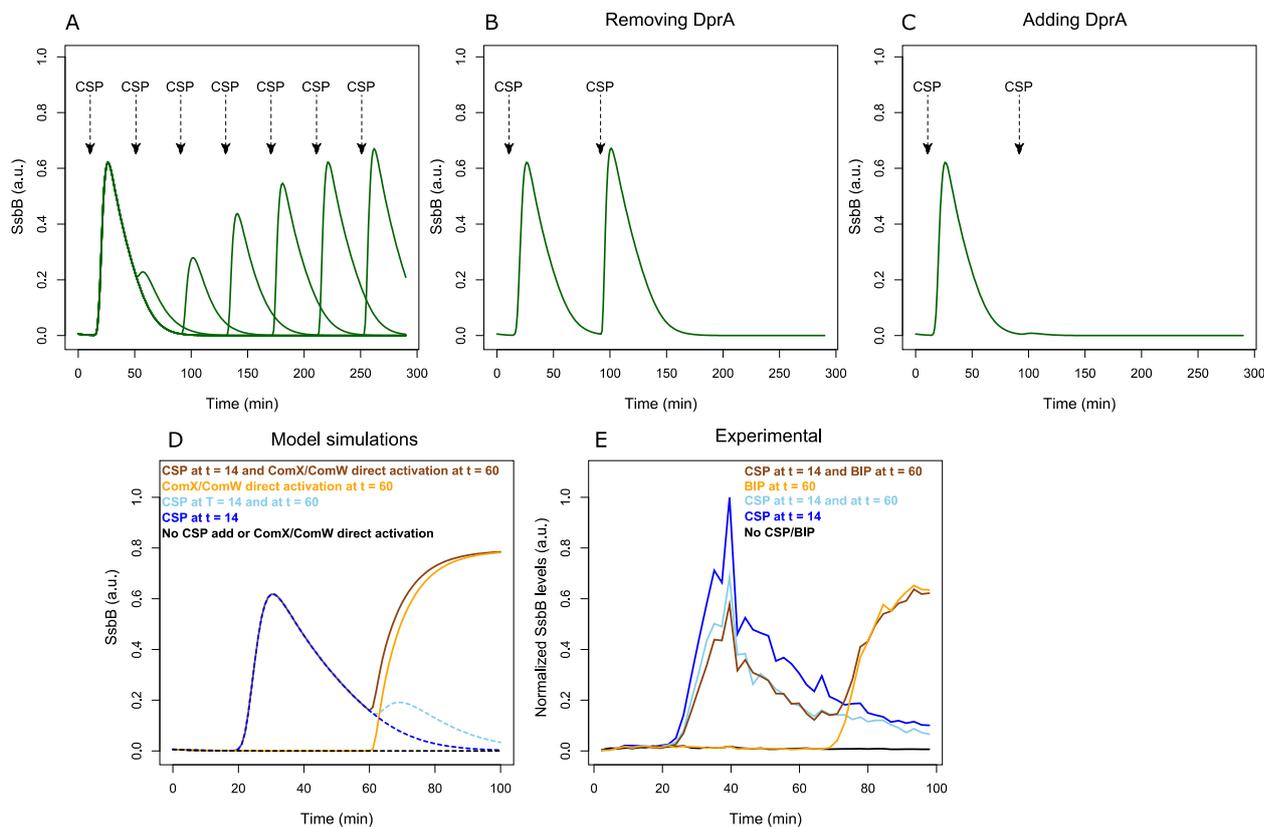
**Figure 27. Cinétique de ComE~P dans les modèles sauvages et *dprA*<sup>-</sup>.** (A) Les cinétiques de ComE~P dans des conditions sauvage et *dprA*<sup>-</sup> sont représentées suite à l'ajout de CSP et jusqu'à 70 minutes après l'addition de CSP. (B) Le code couleur est le même que celui de (A) : la trajectoire du système sauvage est symbolisée par une ligne noire, et la trajectoire dans le système *dprA*<sup>-</sup> par une ligne mauve. La trajectoire dans l'espace des phases a été calculée pendant 100 minutes pour le système sauvage et pendant 210 pour le système *dprA*<sup>-</sup> pour permettre aux systèmes de retourner à leurs états initiaux dans chacune des conditions. L'excursion dans l'espace des phases est initiée par l'ajout d'une unité arbitraire de CSP. Les flèches noires représentent le champ de vecteur du système dynamique. Les nullclines sont dessinées en orange et en vert et représentent respectivement les états du système où  $\frac{dComX}{dt} = 0$  et  $\frac{dComE\sim P}{dt} = 0$ . L'état d'équilibre du système correspond au point d'intersection des nullclines et est représenté par un rond gris dans l'espace des phases et correspond à l'état végétatif/non-compétent. Le champ de vecteur a été calculé grâce au logiciel XPPAUT.

L'étude de la dynamique de ComE~P en fonction de la dynamique de ComX<sub>total</sub> a été réalisée pour un modèle intact et pour un modèle *dprA*<sup>-</sup>. Sur le diagramme de phase (**Figure 27B**), nous observons un unique point d'équilibre, correspondant à l'état végétatif cellulaire. Le champ de vecteur du système dynamique, déterminant la trajectoire du système dans l'espace des phases, permet de ramener le système à son état stationnaire. L'ajout de CSP induit une excursion du système dans l'espace des phases, correspondant à l'état de compétence, plus large dans le cas du mutant *dprA*<sup>-</sup>. Plus précisément, dans le modèle sauvage, après l'addition de CSP, la concentration de ComE~P augmente, suivie par l'augmentation de la concentration de ComX<sub>total</sub>. Puis, après avoir atteint un pic à environ 0,0045 a.u., la concentration de ComE~P libre commence à chuter à cause de sa propre auto-déphosphorylation et de la séquestration par DprA. Enfin, lorsque la quantité de ComE~P passe sous un seuil (0,0025 a.u.), la vitesse de dégradation de ComX devient plus élevée que sa vitesse de synthèse et le système peut alors retourner à son état végétatif initial. Dans un mutant *dprA*<sup>-</sup>, une excursion plus large dans l'espace des phases reflète la prolongation dans l'état de compétence. En effet,

le système retourne totalement à l'état initial après 100 minutes pour le modèle intact, et après plus de 200 minutes pour le modèle *dprA*<sup>-</sup>.

### 11. Étude de la période de non réponse au CSP

Il a été observé que, suite à la fermeture de l'état de compétence, les cellules ne répondent plus à la présence d'une concentration élevée de CSP sur une période de 60-80 minutes, appelée période réfractaire au CSP (Hotchkiss, 1954, Chen et Morrison, 1987). La stabilité de DprA pourrait expliquer la période réfractaire au CSP (Mirouze *et al.*, 2013). Cette hypothèse a été testée en simulant l'ajout de CSP dans le milieu à la fin de la phase de fermeture de l'état de compétence (**Figure 28A**). Il est apparu que sur une période de 100-120 minutes, la réponse du modèle à l'ajout de CSP était moins efficace que celle obtenue après le premier ajout. Nous pouvons supposer que cette observation résulte d'une quantité résiduelle libre de protéines DprA (non complexée avec ComE~P). Ainsi, après un second ajout de CSP, les protéines DprA résiduelles et néo-synthétisées pourraient lier les molécules de ComE~P néo-phosphorylées, empêchant ainsi le déclenchement d'une nouvelle vague de compétence. Cependant, dans le modèle, la quantité de DprA résiduelle est insuffisante pour séquestrer l'ensemble du ComE~P synthétisé, et une seconde vague de compétence peut être initiée. Le niveau atteint par le pic est néanmoins deux fois plus faible que celui observé suite au premier ajout de CSP. Pour corroborer l'hypothèse du rôle de DprA dans la période réfractaire au CSP, deux simulations supplémentaires ont été réalisées. Dans la première, toutes les protéines DprA résiduelles sont supprimées avant le second ajout de CSP pendant la fermeture de l'état de compétence. Cela a pour effet de supprimer la période réfractaire : le second pic de compétence est alors identique à celui observé lors du premier ajout de CSP (**Figure 28B**). Dans la seconde expérience, des protéines DprA sont ajoutées avant le second ajout de CSP pendant la fermeture de la compétence. Dans ce cas, le système ne répond plus au CSP (**Figure 28C**). Les simulations renforcent donc l'hypothèse que DprA est l'acteur principal de la période de non réponse au CSP.



**Figure 28. Etude de la période réfractaire au CSP.** (A) Superposition de six cinétiques de SsbB obtenues par six simulations indépendantes. Dans chaque simulation, le CSP externe est ajouté deux fois : d'abord au temps  $t = 10$  min dans toutes les expériences et à différents moments après l'arrêt de la compétence ( $t = 50, 90, 130, 170, 210, 250$ ) (B) Simulation de la cinétique de SsbB où la compétence est d'abord induite par addition de CSP externe à 10 min et où le système est nettoyé de toute protéine DprA résiduelle une minute avant la seconde addition de CSP à  $t = 90$  min, donc après coupure de compétence. L'élimination des protéines DprA résiduelles est obtenue en substituant sa quantité courante (1,35 a.u.) par celle du modèle au début de l'expérience (0,25 a.u.). (C) Cinétique simulée de SsbB dans une même expérience que dans (B), mais où la quantité de DprA est augmentée une minute avant la deuxième addition de CSP externe en changeant sa valeur courante (1,35 a.u.) à 2,7 a.u. (D) Simulation de la cinétique de SsbB dans le modèle modifié où les synthèses de ComX et de ComW sont indépendantes de la voie de régulation ComCDE et contournent cette voie par une induction directe de leurs gènes. L'induction directe est simulée en introduisant un nouveau terme de synthèse constitutif de ComX et ComW. La valeur de ce paramètre change de 0 à 0,1 a.u. au moment de l'induction. (E) La cinétique expérimentale de SsbB obtenue à partir de données de luminescence où le CSP externe est ajouté à  $t = 14$  et / ou le peptide BIP est ajoutée à  $t = 60$ .

Pour confirmer l'hypothèse que la période réfractaire est dépendante de la séquestration de ComE~P par la protéine DprA résiduelle, le modèle a été modifié de telle façon que l'activation de *comX* et de *comW* par ComE~P puisse être contournée. Pour cela, il a été ajouté dans les équations différentielles, décrivant la cinétique de ComX et celle de ComW, un terme correspondant à une synthèse constitutive de ComX et de ComW. Ainsi, dans ce nouveau modèle, ComX et ComW peuvent être produits à travers deux circuits différents : le

premier sous le contrôle de CSP, impliquant le système ComD/ComE ou sous l'induction directe de ComX et ComW, indépendamment du régulon *comCDE*. Les simulations sont réalisées, soit par ajout directe de CSP, soit par activation directe de *comX* et *comW* à la fin de la fermeture de la compétence (**Figure 28D**). Alors que le modèle ne répond quasi pas au second ajout de CSP, il peut développer une seconde vague de compétence après activation directe de *comX/comW* malgré la présence de la protéine DprA. Pour valider expérimentalement cette prédiction, la plateforme d'expression CEPr (informations supplémentaires dans Johnston *et al.*, 2016) a été utilisée. Cette plateforme est contrôlée par le système à deux composantes BlpR/BlpH (de Saizieu *et al.*, 2000). Ce système à deux composantes peut être activé par la bactériocine BIP (Bacteriocin Inducing Peptide). Les gènes *comX* et *comW* ont été clonés comme un opéron sous la régulation de cette plateforme. La construction a été introduite dans une souche R3584 portant le gène luciférase sous la régulation du promoteur de *ssbB*. Cette souche R3932 est ainsi capable de synthétiser ComX et ComW par la voie ComD/ComE ou par la voie BlpR/BlpH. Pendant la fermeture de la compétence, seulement l'ajout de BIP permet une réexpression du gène tardif de la compétence (**Figure 28E**) comme prédit par le modèle. Ces résultats viennent appuyer l'hypothèse que la période de non-réponse au CSP est liée à un verrouillage de la voie ComD/ComE, probablement due à l'accumulation de DprA qui empêche la formation d'un second réservoir de ComE~P disponible.

#### IV. Discussion

La transformation génétique est un mécanisme permettant le transfert horizontal de gènes qui est l'aboutissement d'un changement d'état physiologique de la cellule appelé compétence. La compétence est un état transitoire nécessitant des conditions environnementales spécifiques pour la mise en œuvre de la transformation. Chez le pathogène humain *S.pneumoniae*, ce processus contribue de façon importante à la plasticité du génome de cet organisme ; il est notamment responsable de l'acquisition de résistances aux antibiotiques (Tomasz, 1964), mais aussi de la résistance aux vaccins au travers de la variations des peptides capsulaires (Golubchik *et al.*, 2012, Croucher *et al.*, 2013). Ce processus étant une force majeure conduisant l'évolution des génomes bactériens, il est important de comprendre comment cet état transitoire se déclenche et quelle est la dynamique de ce comportement.

En intégrant les connaissances expérimentales acquises et en se reposant sur des jeux de données adéquats, la modélisation mathématique permet d'appréhender la dynamique du comportement d'un système, d'acquérir de nouvelles connaissances qui ne sont pas accessibles expérimentalement et de proposer de nouveaux protocoles car la dynamique du système peut être simulée dans de nouvelles conditions n'ayant pas été testées expérimentalement. Cependant, les circuits de régulation de l'état de compétence ne sont pas universels mais composent des adaptations des modes de vies des différentes espèces. Ainsi, différents modèles mathématiques doivent être développés pour tenir compte des spécificités de chaque réseau de régulation.

Chez les bactéries Gram positives, les cascades de régulation les mieux documentées sont celles des deux organismes modèles *S. pneumoniae* et *B. subtilis*. Si plusieurs études mathématiques ont été établies pour étudier différents aspects de la régulation chez *B. subtilis* (Maamar et Dubnau, 2005, Süel *et al.*, 2006, Maamar *et al.*, 2007, Shultz *et al.*, 2007, Leisner *et al.*, 2009, Shultz *et al.*, 2009, Shultz *et al.*, 2013), une seule étude a été entreprise pour modéliser ce processus chez *S. pneumoniae* (Karlsson *et al.*, 2006). Ce travail a eu pour but de suggérer un mécanisme pour la fermeture rapide de l'état de compétence qui n'était alors pas élucidé au moment de la publication. Les auteurs ont proposé qu'un répresseur, sous le contrôle de  $\sigma^X$ , pouvait lier les promoteurs des opérons *comCDE* et *comX*. *A posteriori*, ce répresseur peut être considéré comme correspondant à la protéine DprA dont l'expression est sous le contrôle de  $\sigma^X$ . Cependant, son action sur les promoteurs *comCDE* et *comX* apparaît maintenant être indirecte et agir via la séquestration de ComE~P.

Comme les hypothèses précédemment formulées pour développer le premier modèle de régulation de la compétence chez *S. pneumoniae* sont remises en cause par de nouvelles

données expérimentales, et que la compréhension du réseau de régulation de la compétence s'est élargie ces dernières années, nous avons développé un modèle dynamique de la compétence induite par le CSP à l'échelle cellulaire qui récapitule les connaissances biologiques acquises jusqu'à maintenant. L'apprentissage de la topologie du modèle ne se base pas ici sur une méthode de reconstruction automatique de réseau à partir d'expérience de génomique fonctionnel mais, essentiellement, à partir d'une synthèse des connaissances acquises à ce jour. Certes, le modèle développé corrobore les découvertes expérimentales réalisées, mais il a également mis en évidence un manque dans nos connaissances dans la fermeture de l'état de compétence et nous a permis de proposer de nouvelles hypothèses. La population cellulaire est induite en état de compétence dans les expériences grâce à l'ajout de CSP ce qui permet un développement synchrone de la compétence dans la population cellulaire. Ainsi, une fois normalisées par une méthode appropriée, les mesures quantitatives en temps réel de l'expression des gènes réalisées à l'échelle populationnelle peuvent être utilisées comme mesure de l'expression de gène à l'échelle d'un individu moyen de la population.

Le modèle a été construit en utilisant, dans un premier temps, les réseaux de Petri comme formalisme de modélisation afin de valider la topologie du réseau. Le réseau a ensuite été transformé en équations différentielles ordinaires pour simuler son comportement et pour comparer sa dynamique aux résultats expérimentaux. Afin d'estimer les cinétiques d'expression des gènes, des données obtenues par fusion transcriptionnelle du gène codant pour la luciférase ont été utilisées pour les gènes d'intérêts (*comC::luc*, *comX::luc* et *ssbB::luc*), insérés dans une souche sauvage (WT) ainsi que dans une souche *dprA*<sup>-</sup> (Mirouze *et al.*, 2013). Les données de luminescence ont été normalisées de façon à obtenir un signal de luminescence moyen par cellule et ces valeurs ont été transformées en activité de promoteurs et en cinétiques protéiques. Plutôt que d'émettre l'hypothèse, comme cela a été fait pour la modélisation du circuit de régulation ComRS chez *S. thermophilus* (Haustenne *et al.*, 2015), que la vitesse de dégradation de la luciférase était identique à la vitesse de dégradation de la protéine d'intérêt, il a été réalisé une correction de la différence des demi-vies entre la protéine luciférase et celle de la protéine dont l'activité du gène est mesurée (de Jong *et al.*, 2010; Stefan *et al.*, 2015). Cette correction est en effet importante pour obtenir une estimation précise des cinétiques protéiques, sans quoi les concentrations obtenues des protéines qui ont une durée de vie plus courte que celle de la luciférase sont surestimées et, à l'inverse, les cinétiques déduites des protéines ayant une durée de vie plus longue sont sous-estimées.

Les paramètres des équations différentielles ont été estimés en utilisant à la fois trois jeux de données expérimentaux (acquis pour une souche sauvage et pour des mutants *dprA*<sup>-</sup> ou *clpP*<sup>-</sup>)

et, lorsque des connaissances biologiques sont disponibles sur certaines réactions, en appliquant des contraintes sur les valeurs des paramètres afin de restreindre la recherche dans l'espace des paramètres. Dans une souche *clpP*<sup>-</sup> où ComX et ComW sont stables, il a été démontré que les cellules quittaient l'état de compétence aussi rapidement que des cellules sauvages. Bien qu'aucune donnée de luminescence ne soit disponible pour ce mutant, il est apparu important de tenir compte de cette observation pour évaluer les connaissances actuelles sur la fermeture de l'état de compétence. Ainsi, pour le mutant *clpP*<sup>-</sup>, les données cinétiques des protéines ont été générées en utilisant les données de luminescence de la souche sauvage mais en ajustant la constante de dégradation de ComX à 0 pour reconstruire la cinétique de cette protéine.

Un premier modèle a été réalisé et intègre la fermeture de la compétence via la séquestration de ComE~P par DprA, permettant alors à ComE d'inhiber l'expression de l'opéron *comCDE*. Cependant, nos simulations n'ont pas pu reproduire l'arrêt de l'état de compétence pourtant observé dans un mutant *clpP*<sup>-</sup> avec une cinétique de SsbB (couramment utilisée expérimentalement comme rapporteur des gènes tardifs de la compétence) qui ne décroît pas. Aussi, comme le suggèrent Weng et collaborateurs (2013) dans la discussion de leurs résultats expérimentaux portant sur l'arrêt de l'état de compétence chez *S. pneumoniae*, il a été ajouté dans le modèle un acteur supplémentaire, en plus de DprA, appelé ComZ, pouvant être soit un gène précoce soit un gène tardif de la compétence. Quatre hypothèses alternatives du mode d'action de ComZ ont été testées : ComZ peut cibler ComW, ComX ou directement la forme active de ComX ou encore ComZ pourrait correspondre au facteur  $\sigma^A$  et alors entrer en compétition avec la forme active ComX pour l'ARN polymérase. Parmi les 8 modèles testés, celui reproduisant le mieux les résultats expérimentaux (ayant la plus faible valeur d'AIC), correspond au modèle où ComZ est le produit d'un gène tardif qui interagit directement avec ComW, empêchant alors la formation d'une forme active de ComX. Ce modèle a montré une bonne adéquation entre cinétiques simulées et mesurées expérimentalement, pour une souche sauvage mais aussi pour les souches *clpP*<sup>-</sup> et *dprA*<sup>-</sup>. Afin de valider ces résultats pour le système *clpP*<sup>-</sup>, des mesures expérimentales de suivi en temps réel de l'expression des gènes de la compétence dans une souche *clpP*<sup>-</sup> pourraient être réalisées. De plus, des résultats expérimentaux récents (Tovpeko *et al.*, 2016) nous permettraient de conforter ces prédictions car les auteurs montrent dans leur travail que le mécanisme de fermeture correcte de l'état de compétence n'est pas possible en l'absence de ComW. Les auteurs expliquent en effet que ComW pourrait jouer un rôle dans la fermeture des gènes tardifs de la compétence. Nos simulations prédisent que ComW n'est pas un acteur direct de la fermeture de la compétence mais qu'il serait la cible d'un inhibiteur encore inconnu. L'hypothèse selon laquelle l'arrêt de

la compétence impliquerait une compétition pour la liaison à l'ARN polymérase entre la forme active de ComX et le facteur  $\sigma^A$  a été proposée, car des mutants spontanés suppresseurs du phénotype *comW* sont retrouvés dans le gène *rpoD* codant pour le facteur  $\sigma^A$  (Tovpeko et Morrison, 2014), ce qui pourrait alors affaiblir l'affinité de  $\sigma^A$  pour l'ARN polymérase en faveur de la liaison de  $\sigma^X$ . Cependant, cette hypothèse n'a pas été retenue car la valeur d'AIC de ce modèle est classée quatrième parmi les 8 modèles testés, après 3 modèles où ComZ empêche l'action de ComW sur ComX.

Pour prendre en compte le rôle de ComW dans l'activation de ComX, l'hypothèse d'une interaction directe entre ComW et ComX permettant à ComX de passer de l'état inactif à l'état actif, a été corroborée par l'interaction directe observée entre ComX et ComW (Tovpeko *et al.*, 2016).

Ainsi, la modélisation confirme l'hypothèse selon laquelle ComW fonctionne comme un activateur de facteur sigma en améliorant la liaison de  $\sigma^X$  au cœur de l'ARN polymérase (Tovpeko et Morrison, 2014; Tovpeko *et al.*, 2016) et prédit que l'arrêt de l'état de compétence nécessite, outre l'action de DprA, l'inhibition de ComW par le produit d'un gène tardif.

Pour valider le modèle, nous avons démontré que celui-ci pouvait reproduire qualitativement les données de luminescence obtenues pour l'expression de l'opéron *comCDE* lors du développement de la compétence à différentes valeurs de pH du milieu de croissance, et pour différentes concentrations de CSP ajoutées dans des populations cellulaires de même densité initiales. Les résultats expérimentaux ont montré que le pH du milieu de croissance n'affectait pas directement l'activité de ComD car le niveau d'expression de *comCDE* demeure faible en absence d'ajout de CSP exogène. Ainsi, il a été supposé que le pH du milieu de croissance pourrait influencer les réactions prenant place dans l'environnement extracellulaire. En modulant, de façon conjointe, les valeurs des paramètres représentant l'affinité du dimère de ComD pour le CSP ( $K_{act}$ ) ainsi que la vitesse de phosphorylation de ComD par le CSP ( $v_{max_{act}}$ ), nos simulations ont pu reproduire la tendance des données expérimentales. Les résultats montrent que l'acidité initiale du milieu de croissance empêche le développement de la compétence. Le modèle mathématique suggère que le pH du milieu pourrait produire deux effets, l'un affectant l'affinité de ComD pour le CSP qui est améliorée par un pH basique, l'autre affectant la vitesse de phosphorylation de ComD qui est aussi plus efficace à pH basique. La variation d'un seul de ces paramètres ne permet en effet pas de reproduire les résultats expérimentaux. Nous pouvons supposer que le pH alcalin, en agissant sur la liaison

de ComD au CSP, induit une stabilisation de la forme active du domaine cytoplasmique ComD, favorisant alors la phosphorylation de ce dernier et inversement pour un pH acide.

Le modèle mathématique une fois validé, certains comportements du réseau de régulation décrits expérimentalement ont été étudiés afin d'offrir une explication des phénomènes et de prédire les composants pouvant être impliqués. Les simulations ont montré que la période de non-réponse au CSP pouvait être due à la présence résiduelle de protéine DprA, bloquant alors la voie ComD/ComE grâce à la séquestration de ComE~P, et empêchant donc le développement d'une seconde vague de compétence. L'hypothèse selon laquelle la stabilité de DprA expliquerait la période de non-réponse au CSP a été d'abord proposée lorsque le rôle de DprA dans la fermeture de l'état de compétence a été établi (Mirouze *et al.*, 2013), sans pour autant avoir été expérimentalement démontrée. Il a été réalisé la construction d'une souche pouvant produire ComX et ComW au travers de deux circuits de régulation indépendants, impliquant ou non la voie ComE~P. Nous avons montré qu'une seconde vague de compétence pouvait être observée seulement lorsque le système était induit de façon indépendante du système à deux composantes ComD/ComE. Ainsi, malgré la présence résiduelle de DprA, le contournement de la voie de régulation ComE~P permet de supprimer la période de non-réponse au CSP, ce qui appuie notre prédiction.

Nos simulations ont également montré que *comCDE* et *comAB* étaient tout aussi importants pour le développement de l'état de compétence. L'augmentation seule de l'expression basale de *comCDE* ne permettait pas un développement spontané de la compétence. En augmentant simultanément le niveau basal d'expression de *comCDE* et celui de *comAB* permettait en revanche un déclenchement de la compétence, montrant ainsi que l'export du pré-CSP par ComAB était limitant comme cela a déjà été démontré expérimentalement (Martin *et al.*, 2000, Guiral *et al.*, 2006). Pour une valeur fixée d'expression basale de *comCDE*, le niveau d'expression de *comAB* a été graduellement augmenté et, pour chaque simulation, le ratio ComE~P/ComE a été calculé au temps où ComE~P commence à être synthétisé. Ce ratio reste quasi constant, entre  $10^{-4}$  et  $1.5 \cdot 10^{-4}$ , révélant que, sous cette valeur, ComE~P ne peut pas rentrer en compétition avec ComE pour lier P<sub>comC</sub> et ne peut pas enclencher la boucle de rétroaction positive menant à l'état de compétence.

Finalement, en suivant la cinétique de ComE~P dans une souche sauvage, il a été possible d'estimer que la quantité produite de cette espèce ne comptait, à son maximum, que seulement pour 0,5% de la quantité totale produite de ComE. Cette faible quantité peut être expliquée par la très faible durée de vie estimée de ComE~P (3,22 minutes), empêchant toute

accumulation de la forme phosphorylée de ComE. Cela pourrait en outre fournir une explication quant à l'absence de purification de ComE~P jusqu'à présent.

Ce modèle mathématique a été développé pour consolider les connaissances expérimentales acquises sur la voie ComDE du réseau de régulation de la compétence, mais aussi pour tenter d'acquérir de nouvelles connaissances sur la régulation de cette voie. Le modèle a été élaboré pour la cellule bactérienne *S. pneumoniae* car ce processus a été particulièrement étudié chez cet organisme. Cependant, le modèle développé peut aussi être adapté à d'autres espèces de streptocoques appartenant aux phylogroupes *Streptococcus mitis* et *Streptococcus anginosus* tant que les génomes d'intérêt possèdent un orthologue du gène codant pour ComW de *Streptococcus pneumoniae*. En effet, si les opérons *comCDE*, *comX* et *dprA* sont conservés chez toutes les espèces de ces deux phylogroupes, il n'a pas été identifié d'orthologue de *comW* dans les séquences génomiques de *S. gordonii*, *S. intermedius*, *S. constellatus*, *S. sanguinis*, *S. parasanguinis* ni dans la majorité des souches de *S. anginosus* (à l'exception des souches C1051 et J4211). Pour ces espèces, une connaissance plus approfondie serait nécessaire afin d'établir si une autre protéine peut avoir le même rôle que celui de ComW dans l'activation et dans la stabilisation de ComX, ou bien si ComX n'a pas besoin de cofacteurs pour lier le cœur de l'ARN polymérase. Dans ce cas, notre modèle pourrait être modifié en supprimant les réactions où ComW est impliqué, et l'activation du gène *comX* par ComE~P mènerait alors directement à la synthèse d'une forme active de ComX. Ainsi, pour les autres espèces bactériennes, le modèle peut être *a priori* utilisé sans ajustement majeur de la topologie du réseau, mais une nouvelle estimation des paramètres du réseau devrait cependant être réalisée à partir de données d'expression propres à l'espèce bactérienne considérée. En effet, cela permettrait de représenter plus précisément la dynamique du réseau qui est susceptible de varier en fonction de l'espèce considérée, ou même entre souches d'une même espèce comme cela a été déjà observé entre les souches R800 et CP1250 de *S. pneumoniae* où le délai entre l'expression des gènes précoces et tardifs de la compétence ainsi que les vitesses maximales d'expression des gènes sont différents (Martin *et al.*, 2013). Une mauvaise adéquation entre simulations et données expérimentales permettrait alors de supposer l'existence de nouveaux régulateurs encore non identifiés.

## V. Conclusion et perspectives

Un modèle fondé sur la connaissance du système a été développé et permet d'étudier de façon holistique la régulation de la compétence, après induction par le CSP, à l'échelle de la cellule individuelle de *S. pneumoniae*. Ce modèle consolide les connaissances acquises expérimentalement sur la voie ComDE de la régulation de la compétence en unifiant l'ensemble des connaissances recueillies jusqu'à présent, et offre un cadre permettant d'analyser plusieurs aspects de cette cascade de régulation. Le modèle prédit que la fermeture correcte de la compétence nécessite, outre l'action de DprA, l'inhibition par un gène tardif de ComW qui a pour rôle d'augmenter l'affinité de liaison de  $\sigma^X$  au cœur de l'ARN polymérase. Nous avons aussi corroboré l'hypothèse selon laquelle le niveau basal d'expression de *comCDE* et celui de *comAB* seraient importants pour provoquer un déclenchement spontané de l'état de compétence en agissant sur le ratio ComE~P/ComE. En combinant approches expérimentales et modélisation mathématique, il a été possible de confirmer que le pH du milieu de croissance affectait le développement de la compétence en agissant à la fois sur l'affinité du dimère de ComD pour le CSP et sur la vitesse maximale de la forme non phosphorylée de ComD en forme phosphorylée suite à l'interaction avec le CSP. Les pH alcalins semblent faciliter ces réactions alors que les pH acides les défavorisent. Nous avons enfin confirmé l'hypothèse que DprA est responsable de la période de non-réponse au CSP en bloquant la voie ComDE suite à la séquestration de ComE~P, et empêchant alors le développement d'une seconde vague de compétence.

Le modèle mathématique pourrait être appliqué sans ajustement majeur de la topologie du réseau aux autres espèces de streptocoques appartenant aux groupes *Streptococcus mitis* et *Streptococcus anginosus* dans la mesure où leurs génomes possèdent un orthologue du gène codant pour ComW chez *S. pneumoniae*. Dans le cas contraire, des modifications mineures peuvent être appliquées de telle façon que ComE~P permette directement la synthèse d'une forme active de ComX.

La suite de ce travail pourrait consister à modéliser la synchronisation de la compétence non plus à l'échelle de la cellule unique mais à l'échelle de la population. Le modèle pourrait alors être considéré comme un module d'un modèle plus complexe pouvant utiliser, par exemple, les modèles multi-agents comme formalisme. En effet, le modèle actuel permet de représenter le comportement d'une cellule individuelle car plusieurs observations expérimentales contredisent l'hypothèse selon laquelle l'initiation de la compétence serait contrôlée par un mécanisme de *quorum-sensing* (Claverys *et al.*, 2006). En réalité, le déclenchement de la compétence arriverait dans une sous-population cellulaire et la propagation du signal à l'ensemble de la population s'effectuerait grâce au contact physique cellule à cellule

permettant la transmission du CSP (Prudhomme *et al.*, 2016). Ainsi, afin de créer un modèle dynamique à l'échelle populationnelle, l'hétérogénéité de la population et les interactions intercellulaires devraient être prises en compte. La population d'un tel modèle pourrait être composée de trois types cellulaires : les cellules initiatrices de la compétence pouvant développer l'état de compétence de façon spontanée (activation autocrine), les cellules nécessitant l'activation de la compétence par liaison du CSP aux récepteurs ComD (activation paracrine) et les cellules non compétentes ne répondant pas au CSP. En effet, chez *S. pneumoniae*, il existe deux allèles *comC/comD* pouvant produire deux peptides CSP1 et CSP2, chacun étant reconnu respectivement par son histidine kinase associée ComD1 et ComD2 (Havarstein *et al.*, 1997, Reichmann *et al.*, 2000; Iannelli *et al.*, 2005). Ainsi, si une cellule réceptrice possède un allèle ComD2, elle ne peut pas reconnaître le CSP1 externe ni entrer en compétence. Ces cellules, non compétentes, seront la cible de protéines de lyse produites par les cellules compétentes et codées par les gènes tardifs. L'ADN relâché par les cellules lysées peut alors être intégré par les cellules compétentes (Claverys *et al.*, 2007, Johnsborg *et al.*, 2008). Si le modèle actuel reflète l'induction de la compétence de façon CSP-dépendante, quelques ajustements seulement doivent être réalisés pour modéliser les deux autres types cellulaires. Pour les cellules initiatrices de la compétence, la compétence spontanée peut être induite en ajustant les niveaux d'expressions basales des opérons *comCDE* et *comAB* (voir la partie résultat). Pour les cellules non compétentes, l'on peut faire la distinction entre deux phénotypes, CSP1 et CSP2, de telle façon que, si la quantité externe de CSP est de type CSP1 et si la cellule possède un récepteur de type ComD2, l'état de compétence ne peut être initié. Le modèle pourrait donc être intégré dans une modélisation plus complexe afin de permettre une meilleure compréhension de la synchronisation de la compétence à l'intérieur d'une population de *S. pneumoniae*. Les approches telles que les systèmes multi-agents ou les modèles basés sur les individus (ABMs, IBMs) (Gorochowski *et al.*, 2012, Hellweger *et al.*, 2016) pourraient être des paradigmes de modélisation adaptés. Ces formalismes considèrent une population d'agents ou d'individus autonomes, chacun suivant un ensemble de règles et pouvant interagir avec les autres individus à l'intérieur d'un même environnement. Le modèle actuel peut être utilisé pour décrire chaque agent et son évolution (déclenchement ou non de la compétence) au cours du temps, en fonction de la capture du CSP par l'agent grâce au contact avec une cellule compétente. Cet événement pourrait survenir de façon différente selon que le mode de vie de la bactérie est sous forme planctonique ou en biofilm. De telles approches ont été utilisées pour modéliser plusieurs processus biologiques dans des populations de micro-organismes, et des plateformes de modélisation existent, telles que INDISIM pour simuler la croissance et le comportement de colonies bactériennes (Ginovart *et al.*, 2002), BSim pour la régulation génétique

(Goroehowski *et al.*, 2012), Framework pour des modèles de biofilms (Xavier *et al.*, 2005) ou AgentCell pour la signalisation chimiotactique (Emonet et Cluzel, 2008).

Enfin, le modèle pourrait aussi être utilisé dans la modélisation de processus biologiques plus complexes où il pourrait être considéré comme un module interconnecté avec d'autres modules comme, par exemple, celui contrôlant le cycle cellulaire, ce qui permettrait alors d'intégrer la compétence dans le cycle de vie du pneumocoque.

## **VI. Liste des abréviations**

ADN : Acide désoxyribonucléique

ARN : Acide ribonucléique

SsbB : Single-stranded DNA-binding protein B

RecA : Recombinase A

CSP : Competence-Stimulating Peptide

XIP : *comX*-induced peptide

TCS : Two-Component System

PSO : Particle Swarm Optimization

RLU : Relative Light Unit

AIC : Akaike's Information Criterion

RSS : Root Sum Square

## VII. Bibliographie

- Kitano H. Systems Biology: A Brief Overview. *Science* 2002; 295(5560):1662-4.
- Le Novère N. Quantitative and logic modelling of molecular and gene networks. *Nat Rev Genet.* 2015; 16(3):146-58.
- Bray D, Bourret R, Simon M. Computer simulation of the phosphorylation cascade controlling bacterial chemotaxis. *Mol. Biol. Cell.* 1993; 4:482-469.
- Tindall MJ., Gaffney EA, Maini PK, Armitage JP. Theoretical insights into bacterial chemotaxis. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 2012; 4:247-259.
- Thiele I *et al.* A community-driven global reconstruction of human metabolism: *Nature Biotech.* 2013; 31:419-425.
- Karr JR *et al.* A whole-cell computational model predicts phenotype from genotype. *Cell.* 2012; 150:389-401.
- Chew YH *et al.* Multiscale digital *Arabidopsis* predicts individual organ and whole-organism growth. *Proc Natl Acad Sci USA.* 2014; 111: E4127-E4136.
- Nelander S *et al.* Models from experiments: combinatorial drug perturbations of cancer cells. *Mol Syst Biol.* 2008; 4: 216.
- Ogilvie, Wierling C, Kessler T, Lehrach H and Lange B. Predictive Modeling of Drug Treatment in the Area of Personalized Medicine. *Cancer Inform.* 2015; 14(Suppl 4): 95–103.
- Flobak A, Baudot A, Remy E, Thommesen L, Thieffry D, Kuiper M and Lægreid A. Discovery of Drug Synergies in Gastric Cancer Cells Predicted by Logical Modeling. *PLoS Comput Biol.* 2015; 11(8): e1004426.
- Laubenbacher R, Hower V, Jarrah A, Torti SV, ShualaeV, Mendes P, Torti FM, AkmanS. A systems biology of cancer. *Biochim Biophys Acta.* 2009; 1796(2):129-39.
- Li XL, Oduola WO, Qian L, Dougherty ER. Integrating Multiscale Modeling with Drug Effects for Cancer Treatment. *Cancer Inform.* 2016; 14:21-31.
- Chelliah V, Laibe C, Le Novère N. BioModels Database: a repository of mathematical models of biological processes. *Methods Mol Biol.* 2013; 1021:189-99.

- Cuellar AA, Lloyd CM, Nielsen PF, Bullivant DP, Nickerson DP, Hunter PJ. An overview of CellML 1.1, a biological model description language SIMULATION: Transactions of The Society for Modeling and Simulation International. 2003; 79(12):740-747.
- Olivier BG, Snoep J. Web-based kinetic modelling using JWS Online. *Bioinformatics*. 2004; 20:2143-2144.
- Bidkhorji G, Moeini A, Masoudi-Nejad A. Modeling of tumor progression in NSCLC and intrinsic resistance to TKI in loss of PTEN expression. *PLoS ONE*. 2012; 7, e48004.
- Kiyatkn A *et al.* Scaffolding protein Grb2-associated binder 1 sustains epidermal growth factor-induced mitogenic and survival signaling by multiple positive feedback loops. *J Biol Chem*. 2006; 281 (29):19925-38.
- Huang CY, Ferrel JE. Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proc Natl Acad Sci USA*. 1996; 93(19):10078-83.
- Yamada S, Shiono S, Joo, A, Yoshimura A. Control mechanism of JAK/STAT signal transduction pathway. *FEBS Lett*. 2003; 534:190-196.
- Orchard S *et al.* The MinAct project – IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res*. 2014; 42:D358-D363.
- Franceschini A *et al.* STRING V9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res*. 2014; 41:D808-D815.
- Kanehisa M, Goto S, Sato Y, *et al.* Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res*. 2014; 42:D199-205
- Caspi R, Billington R, Ferrer L, *et al.* The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res*. 2016; 44:D471-80
- Witting U, *et al.* Sabio-RK – database for biochemical reactions kinetics. *Nucleic Acids Res*. 2012; 40:D790-D796.
- Schomburg I, Chang A, Placzek S, *et al.* BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic Acids Res*. 2013; 41:D764-72.

Bachmann J, Raue A, Schilling M, Böhm ME, Kreutz C, Kaschek D, Busch H, Gretz N, Lehmann WD, Timmer J, Klingmüller U. Division of labor by dual feedback regulators controls JAK2/STAT5 signaling over broad ligand range. *Mol Syst Biol.* 2011; 7:516.

Gardner, T. S., di Bernardo, D., Lorenz, D. & Collins, J. J. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science.* 2003; 301, 102-105.

Margolin, A. a *et al.* ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinform.* 2006: 7 (Suppl. 1), 7.

Bonneau, R. *et al.* The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets *de novo*. *Genome Biol.* 2006; 7(5), R36.

Wang RS, Wang Y, Zhang XS, Chen L. Inferring transcriptional regulatory networks from high-throughput data. *Bioinformatics.* 2007; 23(22):3056-64.

Gitter, A., Klein-Seetharaman, J., Gupta, A. & Bar-Joseph, Z. Discovering pathways by orienting edges in protein interaction networks. *Nucleic Acids Res.* 2011; 39, e22.

Yu J, Smith VA, Wang PP, Hartemink AJ, Jarvis ED. Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinformatics.* 2004; 20:3594-3603.

Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008; 9:559.

de Jong H. Modeling and Simulation of Genetic Regulatory Sstems: A Literature Review. *J Comp Biol.* 2002; 9:67-103.

Chen WW, Niepel M, Sorger PK. Classic and contemporary approaches to modeling biochemical reactions. *Genes Dev.* 2010; 24(17):1861-75.

Thomas, R. Boolean formalization of genetic control circuits: *J Theor Biol.* 1973; 42(3):563-585.

De Jong H, Gouzé JL, Hernandez C, Page M, Sari T, Geiselman J. Qualitative simulation of genetic regulatory networks using piecewise-linear models. *Bull Math Biol.* 2004; 66(2):301-40.

Casey R, de Jong H, Gouzé JL. Piecewise-linear models of genetic regulatory networks: equilibria and their stability. *J Math Biol.* 2006; 52(1):27-56.

- Dandekar T, Fieselman A, Majeed S, Ahmed Z. Software applications toward quantitative metabolic flux analysis and modeling. *Brief Bioinform.* 2014; 15(1):91-107.
- Glass L, Kauffman SA: The logical analysis of continuous, non-linear biochemical control networks. *J Theor Biol.* 1973; 39:103-129.
- Mendoza L, Thieffry D, Alvarez-Buylla ER. Genetic control of flower morphogenesis in *Arabidopsis thaliana*: a logical analysis. *Bioinformatics.* 1999; 15:593-606.
- Sánchez L, Van Helden J, Thieffry, D. Establishment of the dorso-ventral pattern during embryonic development of *Drosophila melanogaster*: a logical analysis. *J Theor Biol.* 1997; 189: 377-389.
- Yuh CH, Bolouri H, Davidson EH. Genomic cis-regulatory logic: experimental and computational analysis of a sea urchin gene. *Science.* 1998; 279: 1896-1902.
- Kauffman S, Peterson C, Samuelsson B, Troein C. Random Boolean network models and the yeast transcriptional network. *Proc Natl Acad Sci USA* 2003; 100: 14796-14799.
- Li F, Long T, Lu Y, Ouyang Q, Tang C. The yeast cell-cycle network is robustly designed. *Proc. Natl Acad. Sci. USA.* 2004; 101:4781-4786.
- Bonzanni N, *et al.* Hard-wired heterogeneity in blood stem cells revealed using a dynamic regulatory network model. *Bioinformatics.* 2013; 29: i80-i88.
- Xu H, Ang YS, Sevilla A, Lemischka IR, Ma'ayan A. Construction and validation of a regulatory network for pluripotency and self-renewal of mouse embryonic stem cells. *PLoS Comput Biol.* 2014; 10:e1003777.
- Wittmann DM. *et al.* Transforming Boolean models to continuous models: methodology and application to T-cell receptor signaling. *BMC Syst. Biol.* 2009; 3, 98.
- Calzone L. *et al.* Mathematical modelling of cell-fate decision in response to death receptor engagement. *PLoS Comput. Biol.* 2010; 6:e1000702.
- Grieco L. *et al.* Integrative modelling of the influence of MAPK network on cancer cell fate decision. *PLoS Comput Biol.* 2013; 9:e1003286.
- Morris MK, Saez-Rodriguez J, Sorger PK, Lauffenburger DA. Logic-based models for the analysis of cell signaling networks. *Biochemistry.* 2010; 49(15):3216-24.

- Garg A, Di Cara A, Xenarios I, Mendoza L, De Micheli G. Synchronous versus asynchronous modeling of gene regulatory networks. *Bioinformatics*. 2008; 24: 1917-1925.
- Shmulevich I, Gluhovsky I, Hashimoto RF, Dougherty ER, Zhan W. Steady-state analysis of genetic regulatory networks modelled by probabilistic Boolean networks. *Comp. Funct. Genomics* 2003; 4: 601-608.
- Liang J, Han J. Stochastic Boolean networks: an efficient approach to modeling gene regulatory networks. *BMC Syst Biol*. 2012; 6, 113.
- Stoll G, Viara Z, Barillot E, Calzone L. Continuous time Boolean modeling for biological signaling: application of Gillespie algorithm. *BMC Syst Biol*. 2012; 6: 116.
- Aldridge BB, Saez-Rodriguez J, Muhlich JL, Sorger PK, Lauffenburger DA. Fuzzy logic analysis of kinase pathway crosstalk in TNF/EGF/ insulin-induced signaling. *PLoS Comput Biol*. 2009; 5, e1000340.
- Terfve C *et al.* CellNOptR: a flexible toolkit to train protein signaling networks to data using multiple logic formalisms. *BMC Syst Biol*. 2012; 6:133.
- Carrillo M, Góngora PA, Rosenblueth DA. An overview of existing modeling tools making use of model checking in the analysis of biochemical networks. *Front Plant Sci*. 2012; 3, 155.
- Petri CA: Kommunikation mit Automaten. *Schriften des Instituts für Instrumentelle Mathematik* (1962)
- Marwan W, Sujathab A, Starostzick C. Reconstructing the regulatory network controlling commitment and sporulation in *Physarum polycephalum* based on hierarchical Petri Net modelling and simulation. *J Theor Biol*. 2005; 236: 349-365.
- Sackmann A, Heiner M, Koch I. Application of Petri net based analysis techniques to signal transduction pathways. *BMC Bioinformatics*. 2006; 7:482.
- Grunwald S, Speer A, Ackermann J, Koch I. Petri net modelling of gene regulation of the Duchenne muscular dystrophy. *BioSystems*. 2008; 92:189-205.
- Formanowicz D, Sackmann A, Kozak A, Błazewicz J, Formanowicz P. Some aspects of the anemia of chronic disorders modeled and analyzed by petri net based approach. *Bioprocess Biosyst Eng*. 2011; 34:581-595.

Cordero F, Beccuti M, Fornari C, Lanzardo S, Conti L, Cavallo F, Balbo G, Calogero R. Multi-level model for the investigation of oncoantigen-driven vaccination effect. *BMC Bioinformatics*. 2013; 14(Suppl 6):S11.

Formanowicz D, Kozak A, Głowacki T, Radom M, Formanowicz P. Hemojuvelin-hepcidin axis modeled and analyzed using Petri nets. *J Biomed Inform*. 2013; 46(6):1030-43.

Geistlinger L, Csaba G, Dirmeier S, Kuffner R, Zimmer R. A comprehensive gene regulatory network for the diauxic shift in *Saccharomyces cerevisiae*. *Nucleic Acids Research*. 2013; 1-12.

Chaouiya C, Remy E, Ruet P. & Thiéffry, D. *Proceedings of the 25th International Conference on Applications and Theory of Petri Nets* (eds Cortadella, J. & Reisig, W.) (Springer, Berlin, 2004).

Simao E, Remy E, Thiéffry D, Chaouiya C. Qualitative modelling of regulated metabolic pathways: application to the tryptophan biosynthesis in *E. coli*. *Bioinformatics*. 2005; 21: 190-196.

Steggles LJ, Banks R, Shaw O, Wipat A. Qualitatively modelling and analyzing genetic regulatory networks: a Petri net approach. *Bioinformatics*. 2007; 23:336–343.

Alon U. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman & Hall/CRC, 2007.

Parthimos D, Haddock RE, Hill CE, Griffith TM. Dynamics of a three-variable nonlinear model of vasomotion: comparison of theory and experiment. *Biophys J*. 2007; 93:1534-1556.

Yao G, Tan C, West M, Nevins JR, You L. Origin of bistability underlying mammalian cell cycle entry. *Mol Syst Biol*. 2011; 7:485.

Li S, Brazhnik P, Sobral B, Tyson JJ. A quantitative study of the division cycle of *Caulobacter crescentus* stalked cells. *PLoS Comput Biol*. 2008; 4(1): e9.

Lin Y, Crosson S, Scherer NF. Single-gene tuning of *Caulobacter* cell cycle period and noise, swarming motility, and surface adhesion. *Mol Syst Biol*. 2010; 6:445.

Von Dassow G, Meir E, Munro E, Odell G. The segment polarity network is a robust developmental module. *Nature*. 2000; 406:188-192.

Goldbeter A, Pourquié O. Modeling the segmentation clock as a network of coupled oscillations in the Notch, Wnt and FGF signaling pathways. *J Theor Biol.* 2008; 252: 574-585.

Ozbudak EM, Lewis J. Notch signalling synchronizes the zebrafish segmentation clock but is not needed to create somite boundaries. *PLoS Genet.* 2008; 4: e15.

Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. *Science.* 2002; 297(5584):1183-6.

Rao CV, Wolf DM, Arkin AP. Control, exploitation and tolerance of intracellular noise. *Nature*, 2002 ; 420(6912):231-7.

Longo D, Hasty J. Dynamics of single-cell gene expression. *Mol Syst Biol.* 2006; 2:64.

Paulsson J. Models of stochastic gene expression. *Physics of Life Reviews.* 2005 2(2):157–175.

Wilkinson D. Stochastic Modelling for Systems Biology. Chapman & Hall/CRC, 2006

Süel GM, Garcia-Ojalvo J, Liberman LM and Elowitz MB. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature.* 2006; 440:545-550.

Arkin A, Ross J, McAdams HH. Stochastic kinetic analysis of developmental pathway bifurcation in phage  $\lambda$ -infected *Escherichia coli* cells. *Genetics.* 1998; 149:1633-1648.

Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A. Regulation of noise in the expression of a single gene. *Nature Genet.* 2002; 31:69-73.

Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* 2006; 4:e309.

Cai L, Friedman N, Xie XS. Stochastic protein expression in individual cells at the single molecule level. *Nature* 2006; 440:358-362.

Gillespie D. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comp Phys.* 1976; 22:403-434.

McAdams H, Shapiro L. Circuit simulation of genetic networks. *Science* 1995; 269:650-656.

Singhania R, Sramkoski RM, Jacobberger JW, Tyson JJ. A hybrid model of mammalian cell cycle regulation. *PLoS Comput. Biol.* 2011; 7, e1001077.

Ryll A *et al.* A model integration approach linking signaling and gene-regulatory logic with kinetic metabolic models. *Biosystems*. 2014; 124:26-38.

Shlomi T, Eisenberg Y, Sharan R, Ruppin E. A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Mol Syst Biol*. 2007; 3:101.

Catterall JR. Streptococcus pneumoniae. *Thorax* 1999; 54: 929-937.

Griffith F. The Significance of Pneumococcal Types. *J. Hyg. (Lond.)*. 1928;27:113–59.

Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, McGee L, von Gottberg A, Song JH, Ko KS, Pichon B, Baker S, Parry CM, Lambertsen LM, Shahinas D, Pillai DR, Mitchell TJ, Dougan G, Tomasz A, Klugman KP, Parkhill J, Hanage WP, Bentley SD: Rapid pneumococcal evolution in response to clinical interventions. *Science*. 2011; 331(6016):430-4.

Johnston C, Martin B, Fichant G, Polard P, Claverys JP. Bacterial transformation: distribution, shared mechanisms and divergent control. *Nat Rev Microbiol*. 2014; 12(3):181-96.

Fontaine L, Wahl A, Flécharde M, Mignolet J, Hols P. Regulation of competence for natural transformation in streptococci. *Infect Genet Evol*. 2015; 33:343-60.

Claverys JP, Prudhomme M, Martin B. Induction of competence regulons as a general response to stress in gram-positive bacteria. *Annu Rev Microbiol*. 2006; 60:451-75.

Chen I, Dubnau D. DNA uptake during bacterial transformation. *Nat Rev Microbiol*. 2004; 2(3):241-9.

Chen I, Provvedi R, Dubnau D. A macromolecular complex formed by a pilin-like protein in competent *Bacillus subtilis*. *J. Biol. Chem*. 2006; 28(31):21720-7.

Provvedi R and Dubnau D: ComEA is a DNA receptor for transformation of competent *Bacillus subtilis*. *Mol Microbiol*. 1999; 31:271-280.

Puyet A, Greenberg B, Lacks SA. Genetic and structural characterization of EndA. A membrane bound nuclease required for transformation of Streptococcus pneumoniae *J Mol Biol*. 1990; 231:727-738.

Draskovic I, Dubnau D. Biogenesis of a putative channel protein, ComEC, required for DNA uptake: membrane topology, oligomerization and formation, of disulphide bonds. *Mol Microbiol*. 2007; 55:881-836.

Londoño-Vallejo JA et Dubnau D. Mutation of the putative nucleotide binding site of the *Bacillus subtilis* membrane protein ComFA abolishes the uptake of DNA during transformation. *J Bacteriol.* 1994; 176(15): 4642-4645.

Mortier-Barrière I, Velten M, Dupaigne P, Mirouze N, Piétrement O, McGovern S, Fichant G, Martin B, Noirot P, Le Cam E, Polard P, Claverys JP. A key presynaptic role in transformation for a widespread bacterial protein: DprA conveys incoming ssDNA to RecA. *Cell.* 2007; 130(5):824-836.

Lee MS, Morrison DA. Identification of a new regulator in *Streptococcus pneumoniae* linking quorum sensing to competence for genetic transformation *J.Bacteriol.* 1999; 181: 5004-5016.

Mohan S, Dubnau D. Transcriptional regulation of *comC*: evidence for a competence-specific transcription factor in *Bacillus subtilis*. *J. Bacteriol.* 1990; 172:4064-4071.

Campbell EA, Choi SY, Masure HR. A competence regulon in *Streptococcus pneumoniae* revealed by genomic analysis. *Mol Microbiol* 1998; 27(5):929-39.

Maamar H, Dubnau D. Bistability in the *Bacillus subtilis* K-state (competence) system requires a positive feedback loop. *Mol Microbiol.* 2005; 56:615-624.

Berka RM *et al.* Microarray analysis of the *Bacillus subtilis* K-state: genome wide expression changes dependent on ComK. *Mol Microbiol.* 2002; 43:1331-1345.

Turgay K, Hamoen LW, Venema G, Dubnau D. Biochemical characterization of a molecular switch involving the heat shock protein Clpc, which controls the activity of ComK, the transcription factor of *Bacillus subtilis*. *Genes Dev.* 1997; 11:119-128.

Havarstein LS, Coomaraswamy G and Morrison DA. An unmodified heptadecapeptide pheromone induces competence for genetic transformation in *Streptococcus pneumoniae*. *Proc Natl Acad Sci USA.* 1995; 92:11140-11144.

Pestova EV, Havarstein LS, Morrison DA. Regulation of competence for genetic transformation in *Streptococcus pneumoniae* by an auto-induced peptide pheromone and a two-component regulatory system. *Mol Microbiol.* 1996; 21(4):853-864.

Martin B, Soulet AL, Mirouze N, Prudhomme M, Mortier-Barrière I, Granadel C, Noirot-Gros MF, Noirot P, Polard P, Claverys JP. ComE/ComE~P interplay dictates activation or extinction status of pneumococcal X-state (competence). *Mol Microbiol.* 2013; 87(2):394-411.

Boudes M, Sanchez D, Graille M, van Tilbeurgh H, Durand D, Quevillon-Cheruel S. Structural insights into the dimerization of the response regulator ComE from *Streptococcus pneumoniae*. *Nucleic Acids Res.* 2014; 42(8):5302-5313.

Martin B, Quentin Y, Fichant G, Claverys JP. Independent evolution of competence regulatory cascades in streptococci? *Trends Microbiol.* 2006; 14(8):339-345.

Gardan R, Besset C, Guillot A, Gitton C, Monnet V: The oligopeptide transport system is essential for the development of natural competence in *Streptococcus thermophilus* strain LMD-9. *J Bacteriol.* 2009; 191(14):4647-4655.

Fontaine L, Boutry C, de Frahan MH, Delplace B, Fremaux C, Horvath P, Boyaval P, Hols P: A novel pheromone quorum-sensing system controls the development of natural competence in *Streptococcus thermophilus* and *Streptococcus salivarius*. *J Bacteriol.* 2010; 192(5):1444-1454.

Fontaine L, Goffin P, Dubout H, Delplace B, Baulard A, Lecat-Guillet N, Chambellon E, Gardan R, Hols P. Mechanism of competence activation by the ComRS signalling system in streptococci. *Mol Microbiol.* 2013; 87(6):1113-32.

Leisner M, Khur JT, Rädler JO, Frey E and Maier B. Kinetics of genetic switching into the state of bacterial competence. *Biophys J.* 2009; 96:1178-1188.

Liebal UW, Millat R, de Jong IG, Kuipers OP, Völker U and Wolkenhauer O. How mathematical modeling elucidates signaling in *Bacillus subtilis*. *Mol. Microbiol.* 2010; 77:1083-1095.

Dubnau D, Losick R. Bistability in bacteria. *Mol Microbiol.* 2006; 61(3):564-72.

Maamar H, Raj A, Dubnau D. Noise in gene expression determines cell fate in *Bacillus subtilis*. *Science.* 2007; 317:526-529.

Schultz D, Wolynes PG, Ben JE, and Onuchic JN. Deciding fate in adverse times: sporulation and competence in *Bacillus subtilis*. *Proc Natl Acad Sci USA.* 2009; 106:21027–21034.

Stephens C. Bacterial sporulation: A question of commitment? *Curr Biol.* 1998; 8:R45–R48.

Karlsson D, Karlsson S, Gustafsson E, Normark BH, Nilsson P. Modeling the regulation of the competence-evoking quorum sensing network in *Streptococcus pneumoniae*. *BioSystems.* 2006; 90(1):211-223.

Mirouze N, Bergé MA, Soulet AL, Mortier-Barrière I, Quentin Y, Fichant G, Granadel C, Noirot-Gros MF, Noirot P, Polard P, Martin B, Claverys JP. Direct involvement of DprA, the transformation-dedicated RecA loader, in the shut-off of pneumococcal competence. *Proc Natl Acad Sci USA*. 2013; 110(11):E1035-E1044.

Prudhomme M, Berge M, Martin B, Polard P. Pneumococcal Competence Coordination Relies on a Cell-Contact Sensing Mechanism. Pneumococcal Competence Coordination Relies on a Cell-Contact Sensing Mechanism. *PLoS Genet*. 2016; 29;12(6):e1006113.

Son M, Ahn SJ, Guo Q, Burne RA, Hagen SJ. Microfluidic study of competence regulation in *Streptococcus mutans*: environmental inputs modulate bimodal and unimodal expression of *comX*. *Mol Microbiol*. 2012; 86(2):258-72.

Haustenne L, Bastin G, Hols P, Fontaine L. Modeling of the ComRS Signaling Pathway Reveals the Limiting Factors Controlling Competence in *Streptococcus thermophilus*. *Front Microbiol*. 2015; 6:1413.

de Jong H, Ranquet C, Ropers D, Pinel C, Geiselmann J. Experimental and computational validation of models of fluorescent and luminescent reporter genes in bacteria. *BMC Syst Biol*. 2010; 4:55.

Stefan D, Pinel C, Pinhal S, Cinquemani E, Geiselmann J, de Jong H. Inference of quantitative models of bacterial promoters from time-series reporter gene data. *PLoS Comput Biol*. 2015; 11(1): e1004028.

Cassone M, Gagne AL, Spruce LA, Seeholzer SH, Sebert ME. The HtrA protease from *Streptococcus pneumoniae* digests both denatured proteins and the competence-stimulating peptide. *J Biol Chem*. 2012; 287(46):38449-59.

Murata T. Petri Nets: properties, analysis and applications. *Proc IEEE*; 1989. 77:541-80.

Chaouiya C. Petri net modelling of biological networks. *Brief Bioinformatics*. 2007; 8(4):210-219.

Koch, I. et Heiner, M. Analysis of Biological Networks, John Wiley & Sons, Inc (Björn H. Junker and Falk Schreiber, editors); 2008. p. 139-179.

Ronen M, Rosenberg R, Shraiman B, Alon U. Assigning numbers to the arrows: Parameterizing a gene regulation network by using accurate expression kinetics. *Proc Natl Acad Sci USA*. 2002; 99(16):10555-10560.

- Berthoumieux S. *et al.* Shared control of gene expression in bacteria by transcription factors and global physiology of the cell. *Mol Syst Biol.* 2013; 9: 634.
- Kobiler O, Rokney A, Friedman N, Court D, Stavans J, Oppenheim A. Quantitative kinetic analysis of the bacteriophage  $\lambda$  genetic network. *Proc Natl Acad Sci USA.* 2005; 102(12):4470-4475.
- Lu C, Albano C, Bentley W, Rao G. Quantitative and kinetic study of oxidative stress regulons using green fluorescent protein. *Biotechnol Bioeng.* 2005; 89(5):574-587.
- Finkenstädt B, Heron E, Komorowski M, Edwards K, Tang S, Harper C, Davis J, White M, Millar A, Rand D. Reconstruction of transcriptional dynamics from gene reporter data using differential equations. *Bioinformatics.* 2008; 24(24):2901-2907.
- Huang Z, Senocak F, Jayaraman A, Hahn J. Integrated modeling and experimental approach for determining transcription factor profiles from fluorescent reporter data. *BMC Syst Biol.* 2008; 2:64.
- Wang X, Errede B, Elston T. Mathematical analysis and quantification of fluorescent proteins as transcriptional reporters. *Biophys J.* 2008; 94(6):2017-2026.
- Aïchaoui L, Jules M, Le Chat L, Aymerich S, Formion V, Goelzer A: BasyLiCA. a tool for automatic processing of a Bacterial Live Cell Array. *Bioinformatics.* 2012; 28(20):2705-6.
- Prudhomme M et Claverys JP. There will be a light: the use of luc transcriptional fusions in living pneumococcal cells. Hakenbeck R and Chhatwal GS, editors. *The Molecular Biology of Streptococci.* (Norfolk, UK: Horizon Scientific Press), 2007. p. 519-524.
- Banga JR. Optimization in computational systems biology *BMC Syst Biol.* 2008; 2: 47.
- Saez-Rodriguez J, Alexopoulos LG, Epperlein J, Samaga R, Lauffenburger DA, Klamt S, Sorger PK. Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. *Mol Syst Biol.* 2009; 5:331.
- Schreiber J, Arter M, Lapique N, Haefliger B, Benenson Y. Model guided combinatorial optimization of complex synthetic gene networks. *Mol Syst Biol.* 2016; 12(12):899.
- Hoops S, Sahle S, Gauges R, Lee C, Pahle J, Simus N, Singhal M, Xu L, Mendes P, Kummer U. COPASI: a COMplex PATHway Simulator. *Bioinformatics.* 2006; 22(24):3067-74.
- Kennedy, J. and Eberhart, R. Particle Swarm Optimization. In *Proceedings of IEEE International Conference on Neural Networks IV.* p. 1942-1948. Piscataway: IEEE, 1995.

- Poli R, Kennedy J, Blackwell T. Particle swarm optimization: An overview. *Swarm Intell.* 2007, 1:33-57.
- Baker SM, Schallau K, Junker BH. Comparison of different algorithms for simultaneous estimation of multiple parameters in kinetic metabolic models. *J Integr Bioinform.* 2010; 7(3):133.
- Petzold L. Automatic selection of methods for solving stiff and nonstiff systems of ordinary differential equations. *SIAM J. Sci and Stat. Comput.* 1983; 4(1):136-148.
- Lange R., Wagner C, de Saizieu A, Flint N, Molnos J, Stieger M, Caspers P, Kamber M, Keck W, Amrein KE. Domain organization and molecular characterization of 13 two-component systems identified by genome sequencing of *Streptococcus pneumoniae*. *Gene.* 1999; 237(1):223-234.
- George Cisar EA, Geisinger E, Muir TW, Richard RP. Symmetric signaling within asymmetric dimers of the *Staphylococcus aureus* receptor histidine kinase AgrC. *Mol Microbiol.* 2009; 74(1):44-57.
- Martin B, Granadel C, Campo N, Hénard V, Prudhomme M, Claverys JP. Expression and maintenance of ComD-ComE, the two-component signal-transduction system that controls competence of *Streptococcus pneumoniae*. *Mol Microbiol.* 2010; 75(6):1513-28.
- Wang B, Zhao A, Novick RP, Muir TW. Activation and inhibition of the receptor histidine kinase AgrC occurs through opposite helical transduction motions. *Mol Cell.* 2014; 53(6):929-40.
- Sanchez D, Boudes M, van Tilbeurgh H, Durand D, Quevillon-Cheruel S. Modeling the ComD/ComE/comcde interaction network using small angle X-ray scattering. *FEBS J.* 2015; 282(8):1538-53.
- Luo P, Morrison DA. Transient association of an alternative sigma factor, ComX, with RNA polymerase during the period of competence for genetic transformation in *Streptococcus pneumoniae*. *J. Bacteriol.* 2003; 185:349–58.
- Luo P, Li H, Morrison DA. Identification of ComW as a new component in the regulation of genetic transformation in *Streptococcus pneumoniae*. *Mol Microbiol.* 2004; 54(1):172-83.
- Sung CK, Morrison DA. Two distinct functions of ComW in stabilization and activation of the alternative sigma factor ComX in *Streptococcus pneumoniae*. *J Bacteriol.* 2005; 187(9):3052-61.

Piotrowski A, Luo P, Morrison DA. Competence for genetic transformation in *Streptococcus pneumoniae*: termination of activity of the alternative sigma factor ComX is independent of proteolysis of ComX and ComW. *J Bacteriol.* 2009; 191(10):3359-66.

Tovpeko Y, Morrison DA. Competence for genetic transformation in *Streptococcus pneumoniae*: mutations in  $\sigma A$  bypass the comW requirement. *J Bacteriol.* 2014; 196(21):3724-34.

Tovpeko Y, Bai J, Morrison DA. Competence for Genetic Transformation in *Streptococcus pneumoniae*: Mutations in  $\sigma A$  Bypass the ComW Requirement for Late Gene Expression. *J Bacteriol.* 2016; 198(17):2370-8.

Weng L, Piotrowski A, Morrison DA. Exit from competence for genetic transformation in *Streptococcus pneumoniae* is regulated at multiple levels. *PLoS One.* 2013; 8(5):e64197.

Marwan W, Rohr C, Heiner M. Petri nets in Snoopy: A unifying framework for the graphical display, computational modelling, and simulation of bacterial regulatory networks; *Methods in Molecular Biology – Bacterial Molecular Networks* (van Helden J, Toussaint A and Thiéffry D, editors), Humana Press; 2007. p. 409-437.

Heiner M, Schwarick M, Wegener J. Charlie--an extensible Petri net analysis tool. In Proc. PETRI NETS, Brussels, Springer, LNCS. 2015; 9115:200-211.

Chen JD, Morrison DA. Modulation of competence for genetic transformation in *Streptococcus pneumoniae*. *J Gen Microbiol.* 1987; 133(7):1959-67.

Martin B, Prudhomme M, Alloing G, Granadel C and Claverys JP. Cross-regulation of competence pheromone production and export in the early control of transformation in *Streptococcus pneumoniae*. *Mol Microbiol.* 2000; 38(4):867-878

Guiral S, Hénard V, Granadel C, Martin B, Claverys JP. Inhibition of competence development in *Streptococcus pneumoniae* by increased basal level expression of the ComDE two-component regulatory system. *Microbiology.* 2006; 152:323-331.

Hotchkiss RD. CYCLICAL BEHAVIOR IN PNEUMOCOCCAL GROWTH AND TRANSFORMABILITY OCCASIONED BY ENVIRONMENTAL CHANGES. *Proc Natl Acad Sci U S A.* 1954; 40:49–55.

de Saizieu A, Gardès C, Flint N, Wagner C, Kamber M, Mitchell TJ, Keck W, Amrein KE, Lange R. Microarray-based identification of a novel *Streptococcus pneumoniae* regulon controlled by an autoinduced peptide. *J Bacteriol.* 2000; 182(17):4696-4703.

Golubchik T, Brueggemann AB, Street T, Gertz RE, Spencer CCA, Ho T, et al. Pneumococcal genome sequencing tracks a vaccine escape variant formed through a multi-fragment recombination event. *Nat. Genet.* 2012; 44:352–5.

Havarstein LS, Hakenbeck R., Gaustad P. Natural competence in the genus *Streptococcus*: evidence that streptococci can change phenotype by interspecies recombinational exchanges. *J Bacteriol.* 1997; 179:6589-6594.

Reichmann, P et Hakenbeck, R. Allelic variation in a peptide-inducible two-component system of *Streptococcus pneumoniae*. *FEMS Microbiol Lett.* 2000; 190:231-236.

Iannelli, F, Oggioni MR, Pozzi G. Sensor domain of histidine kinase ComD confers competence phenotype specificity in *Streptococcus pneumoniae*. *FEMS Microbiol Lett.* 2005; 252:321-326.

Claverys JP, Martin B, Havarstein LS. Competence-induced fratricide in streptococci. *Mol Microbiol.* 2007; 64:1423-1433.

Johnsborg, O, Eldholm V, Bjornstad ML, Havarstein LS. A predatory mechanism dramatically increases the efficiency of lateral gene transfer in *Streptococcus pneumoniae* and related commensal species. *Mol Microbiol.* 2008; 69:245-253.

Gorochoowski TE, Matyjaszkiewicz A., Todd T, Oak N, Kowalska K, Reid S, Tsaneva-Atanasova KT, Savery NJ, Grierson CS, di Bernardo M. BSim: An Agent-Based Tool for Modeling Bacterial Populations in Systems and Synthetic Biology. *PLoS ONE.* 2012; 7:e42790.

Hellweger FL, Clegg RJ, Clark JR, Plugge CM, Kreft J-U. Advancing microbial sciences by individual-based modelling. *Nat. Rev. Microbiol.* 2016; 14:461–71.

Ginovart M, López D, Valls J. INDISIM, An Individual-based Discrete Simulation Model to Study Bacterial Cultures. *J Theor Biol.* 2002; 214:305-319.

Xavier JB, Picioreanu C, van Loosdrecht MC. A framework for multidimensional modelling of activity and structure of multispecies biofilms. *Environ Microbiol.* 2005; 7:1085-1103.

Emonet T, Cluzel P. Relationship between cellular response and behavioral variability in bacterial chemotaxis. *Proc Natl Acad Sci USA.* 2008; 105:3304-3309.

# Evolution and modeling of biological processes : application to the regulation of natural competence for bacterial genetic transformation in Streptococci

## Abstract

Bacteria have evolved many types of genetically induced mechanisms to face different types of stresses and to adapt to new environments. Competence for natural transformation is one such process that promotes horizontal gene transfer. If phylogenetically distant species share conserved uptake and processing apparatus, competence regulatory circuits are not universal but adapted to every species' lifestyle. In Gram-positive bacteria, *Streptococcus pneumoniae* and *Bacillus subtilis* regulatory cascades are the best documented. If many mathematical models have been established to study different aspects of competence regulation in *B. subtilis*, only one population-scaled model has been developed for *S. pneumoniae*, a decade ago, based on hypotheses that are challenged by new experimental data.

We develop, in *S. pneumoniae*, a knowledge-based model of the competence regulation at cell level that integrates the enriched biological knowledge acquired to date. The structural consistency of the network topology is confirmed using Petri net formalism. The network is further turned into a set of ordinary differential equations to study its dynamics behavior. Protein kinetics are estimated using time-series luminescence data and other parameter estimations are constrained according to available knowledge. We point out some gap in competence shut-off knowledge, and, after testing alternative models, we predict the requirement of a yet unknown late *com* gene product inhibiting the action of ComW, the  $\sigma^x$  factor activator. We also bring new insights into this regulatory cascade by predicting the system components that might be involved in specific experimental behavior.

Our model consolidates the experimental knowledge acquired on competence regulation in *S. pneumoniae*. Moreover, it can be applied to the other streptococci species belonging to the *mitis* and *anginosus* groups since they shared the same regulatory circuit. In the population, the competence shift happens first in a subpopulation of cells and spreads into the whole population through cell to cell contact. Allowing simulation of individual cell behavior, our model will provide a brick for the design of a population-scale model composed of heterogeneous cells.

**AUTEUR : WEYDER Mathias**

**TITRE : Évolution et modélisation de processus biologiques : application à la régulation de la compétence naturelle pour la transformation génétique bactérienne chez les Streptocoques**

**DIRECTEUR DE THÈSE : Dr. Fabien JOURDAN**

**LIEU ET DATE DE SOUTENANCE : Université Toulouse 3 Paul Sabatier – 29 mars 2017**

---

**RÉSUMÉ : Afin de faire face à différents types de stress et s'adapter à de nouveaux environnements, les bactéries ont développé de nombreux mécanismes génétiquement régulés. La compétence pour la transformation naturelle est un processus qui favorise le transfert horizontal de gènes. Parmi les nombreuses cascades de régulation, celle régulant la compétence chez *Streptococcus pneumoniae* est une des mieux caractérisées.**

**Nous avons développé, chez *S. pneumoniae*, un modèle qui intègre les éléments biologiques essentiels connus à ce jour. La cohérence structurelle du réseau est confirmée par le formalisme des réseaux de Petri puis le réseau est transformé en un système d'équations différentielles ordinaires pour étudier son comportement dynamique.**

**Ainsi, ce modèle mathématique apporte un nouvel éclairage sur cette cascade de régulation de la compétence chez *S. pneumoniae*. Le modèle souligne également que la fermeture abrupte de la compétence serait multifactorielle, nécessitant un autre acteur encore inconnu.**

---

**MOTS CLÉS : *Streptococcus pneumoniae* – Modélisation mathématique – Réseau de régulation – Compétence bactérienne – Réseaux de Petri – Équations différentielles ordinaires**

---

**DISCIPLINE ADMINISTRATIVE : ED BSB : Microbiologie**

---

**INTITULÉ ET ADRESSE DU LABORATOIRE : Laboratoire de Microbiologie et Génétique Moléculaires – UMR 5100 – Université Paul Sabatier – 118, route de Narbonne – 31062 TOULOUSE CEDEX 9**