HAL
open science

Emergence de structures modulaires dans les régulations des systèmes biologiques : théorie et applications à Bacillus subtilis

Anne Goelzer

To cite this version:

Anne Goelzer. Emergence de structures modulaires dans les régulations des systèmes biologiques : théorie et applications à Bacillus subtilis. Sciences de l'ingénieur [physics]. Ecole Centrale de Lyon, 2010. Français. NNT : 2010ECDL0030 . tel-00597796

HAL Id: tel-00597796

https://theses.hal.science/tel-00597796

Submitted on 2 Jun 2011

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

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

N°d'ordre 2010-30 Année 2010

ÉCOLE CENTRALE DE LYON

THÈSE

pour obtenir le grade de

DOCTEUR

Spécialité : "Informatique Automatique"

préparée dans le laboratoire Ampère et dans l'unité de Mathématique, Informatique et Génome de l'Institut National de la Recherche Agronomique

dans le cadre de **l'École Doctorale "Électonique, Électrotechnique et Automatique"**

par

Anne Goelzer

Émergence de structures modulaires dans les régulations des systèmes biologiques

Théorie et applications à *Bacillus subtilis*

soutenue publiquement le 4 novembre 2010 devant la commission d'examen

JURY

Président du jury	M. François BURET	Professeur des Universités, École Centrale de Lyon
Rapporteur	M. Claude GAILLARDIN	Professeur des Universités, AgroParisTech
Rapporteur	M. Salvatore MONACO	Professeur des Universités, Université de Rome
Examinateur	M. Philippe NOIROT	Directeur de Recherche, Inra
Directeur de thèse	M. Vincent FROMION	Directeur de Recherche, Inra
Directeur de thèse	M. Gérard SCORLETTI	Professeur des Universités, École Centrale de Lyon

ECOLE CENTRALE DE LYON

DIRECTION DE LA RECHERCHE

Liste des personnes Habilitées à Diriger des Recherches en poste à l'Ecole Centrale de Lyon

Nom-Prénom	*Corps grade*	Laboratoire ou à défaut département ECL	Etablissement
AURIOL Philippe	*professeur*	AMPERE	ECL
BEROUAL Abderrahmane	*professeur*	AMPERE	ECL
BURET François	*maître de conférences*	AMPERE	ECL
JAFFREZIC-RENAULT Nicole	*directeur de recherche*	AMPERE	CNRS/ECL
KRÄHENBÜHL Laurent	*directeur de recherche*	AMPERE	CNRS/ECL
MARTELET Claude	*professeur*	AMPERE	ECL
NICOLAS Alain	*professeur*	AMPERE	ECL
NICOLAS Laurent	*directeur de recherche*	AMPERE	CNRS/ECL
SCORLETTI Gérard	*professeur*	AMPERE	ECL
SIMONET Pascal	*chargé de recherche*	AMPERE	CNRS/ECL
THOMAS Gérard	*professeur*	AMPERE	ECL
VOLLAIRE Christian	*maître de conférences*	AMPERE	ECL
		Nbre Ampère 12	
HELLOUIN Yves	*maître de conférences*	DER EEA	ECL
LE HELLEY Michel	*professeur*	DER EEA	ECL
		Nbre DER EEA 2	
GUIRALDENQ Pierre	*professeur émérite*	DER STMS	ECL
VINCENT Léo	*professeur*	DER STMS	ECL
		Nbre DER STMS 2	
LOHEAC Jean-Pierre	*maître de conférences*	ICJ	ECL
MAITRE Jean-François	*professeur émérite*	ICJ	ECL
MARION Martine	*professeur*	ICJ	ECL
MOUSSAOUI Mohand	*professeur*	ICJ	ECL
MUSY François	*maître de conférences*	ICJ	ECL
ROUY MIRONESCU Elisabeth	*professeur*	ICJ	ECL
ZINE Abdel-Malek	*maître de conférences*	ICJ	ECL
		Nbre ICJ 7	
DAVID Bertrand	*professeur*	ICTT	ECL
		Nbre ICTT 1	
CALLARD Anne-Ségolène	*maître de conférences*	INL	ECL
CLOAREC Jean-Pierre	*maître de conférences*	INL	ECL
GAFFIOT Frédéric	*professeur*	INL	ECL
GAGNAIRE Alain	*maître de conférences*	INL	ECL
GARRIGUES Michel	*directeur de recherche*	INL	CNRS/ECL
GENDRY Michel	*directeur de recherche*	INL	CNRS/ECL
GRENET Geneviève	*directeur de recherche*	INL	CNRS/ECL

HOLLINGER Guy	*directeur de recherche*	INL	CNRS/ECL
JOSEPH Jacques	*professeur*	INL	ECL
KRAWCZYK Stanislas	*directeur de recherche*	INL	CNRS/ECL
LETARTRE Xavier	*chargé de recherche*	INL	CNRS/ECL
MARTIN Jean-René	*professeur émérite*	INL	ECL
O'CONNOR Ian	*maître de conférences*	INL	ECL
PHANER-GOUTORBE Magali	*professeur*	INL	ECL
ROBACH Yves	*professeur*	INL	ECL
SAINT-GIRONS Guillaume	*chargé de recherche*	INL	CNRS/ECL
SEASSAL Christian	*chargé de recherche*	INL	CNRS/ECL
SOUTEYRAND Eliane	*directeur de recherche*	INL	CNRS/ECL
TARDY Jacques	*directeur de recherche*	INL	CNRS/ECL
VIKTOROVITCH Pierre	*directeur de recherche*	INL	CNRS/ECL

Nbre INL 20

CHEN Liming	*professeur*	LIRIS	ECL

Nbre LIRIS 1

BAILLY Christophe	*professeur*	LMFA	ECL
BERTOGLIO Jean-Pierre	*directeur de recherche*	LMFA	CNRS/ECL
BLANC-BENON Philippe	*directeur de recherche*	LMFA	CNRS/ECL
BOGEY Christophe	*chargé de recherche*	LMFA	CNRS/ECL
CAMBON Claude	*directeur de recherche*	LMFA	CNRS/ECL
CARRIERE Philippe	*chargé de recherche*	LMFA	CNRS/ECL
CHAMPOUSSIN J-Claude	*professeur émérite*	LMFA	ECL
COMTE-BELLOT geneviévre	*professeur émérite*	LMFA	ECL
FERRAND Pascal	*directeur de recherche*	LMFA	CNRS/ECL
GALLAND Marie-Annick	*maître de conférences*	LMFA	ECL
GODEFERD Fabien	*chargé de recherche*	LMFA	CNRS/ECL
GOROKHOVSKI Mikhail	*professeur*	LMFA	ECL
HENRY Daniel	*directeur de recherche*	LMFA	CNRS/ECL
JEANDEL Denis	*professeur*	LMFA	ECL
JUVE Daniel	*professeur*	LMFA	ECL
LE RIBAULT Catherine	*chargée de recherche*	LMFA	CNRS/ECL
LEBOEUF Francis	*professeur*	LMFA	ECL
PERKINS Richard	*professeur*	LMFA	ECL
ROGER Michel	*professeur*	LMFA	ECL
SCOTT Julian	*professeur*	LMFA	ECL
SHAO Liang	*chargé de recherche*	LMFA	CNRS/ECL
SIMOENS Serge	*chargé de recherche*	LMFA	CNRS/ECL
TREBINJAC Isabelle	*maître de conférences*	LMFA	ECL

Nbre LMFA 23

BENAYOUN Stéphane	*professeur*	LTDS	ECL
CAMBOU Bernard	*professeur*	LTDS	ECL
COQUILLET Bernard	*maître de conférences*	LTDS	ECL
DANESCU Alexandre	*maître de conférences*	LTDS	ECL
FOUVRY Siegfrid	*chargé de recherche*	LTDS	CNRS/ECL
GEORGES Jean-Marie	*professeur émérite*	LTDS	ECL
GUERRET Chrystelle	*chargé de recherche*	LTDS	CNRS/ECL
HERTZ Dominique	*past*	LTDS	ECL
ICHCHOU Mohamed	*maître de conférences*	LTDS	ECL
JEZEQUEL Louis	*professeur*	LTDS	ECL
JUVE Denyse	*ingénieur de recherche*	LTDS	ECL
KAPSA Philippe	*directeur de recherche*	LTDS	CNRS/ECL

LE BOT Alain	*chargé de recherche*	LTDS	CNRS/ECL
LOUBET Jean-Luc	*directeur de recherche*	LTDS	CNRS/ECL
MARTIN Jean-Michel	*professeur*	LTDS	ECL
MATHIA Thomas	*directeur de recherche*	LTDS	CNRS/ECL
MAZUYER Denis	*professeur*	LTDS	ECL
PERRET-LIAUDET Joël	*maître de conférences*	LTDS	ECL
SALVIA Michelle	*maître de conférences*	LTDS	ECL
SIDOROFF François	*professeur*	LTDS	ECL
SINOU Jean-Jacques	*maître de conférences*	LTDS	ECL
STREMSDOERFER Guy	*professeur*	LTDS	ECL
THOUVEREZ Fabrice	*professeur*	LTDS	ECL
TREHEUX Daniel	*professeur*	LTDS	ECL
VANNES André-Bernard	*professeur émérite*	LTDS	ECL
		Nbre LTDS 25	
	Total HdR ECL	91	

A ma famille.
A Richard et Lucie.

Remerciements

Les travaux présentés dans ce mémoire de thèse ont été effectués dans le cadre de ma formation continue au sein de l'Institut National de la Recherche en Agronomie, en parallèle de mes activités d'ingénierie. A ce propos, je remercie le directeur du département Mathématiques et Informatique Appliquées Monsieur B. Goffinet, ainsi que les deux directeurs successifs de l'unité Mathématique Informatique et Génome (MIG) Messieurs F. Rodolphe et J.F. Gibrat pour m'avoir donné l'autorisation d'effectuer ce travail de thèse. Cette thèse a été co-dirigée par Messieurs V. Fromion de l'unité MIG et G. Scorletti du laboratoire Ampère de l'École Centrale de Lyon. Je remercie également Monsieur L. Nicolas, directeur du laboratoire Ampère, pour m'avoir accueillie au sein de son laboratoire.

Je remercie Messieurs S. Monaco, Professeur à l'Université «La Sapienza» de Rome et C. Gaillardin, Professeur des Universités de l'AgroParisTech pour m'avoir fait l'honneur de rapporter cette thèse, d'y avoir consacré du temps et pour leurs remarques qui ont substantiellement amélioré la clarté du document. Je suis très honorée que Monsieur F. Buret, Professeur des Universités de l'École Centrale de Lyon, ait accepté de présider le jury de thèse. Enfin, je remercie Monsieur P. Noirot, Directeur de Recherche à l'Inra, d'avoir accepté d'être membre du jury.

Je tiens à remercier mes deux directeurs de thèse Messieurs Vincent Fromion et Gérard Scorletti pour leur encadrement, leur enseignement scientifique et leurs conseils dispensés non seulement au cours de cette thèse, mais aussi depuis ces dix dernières années. Merci Gérard pour tes excellents cours d'Automatique à l'Ensicaen et pour m'avoir orientée en 2002 vers le poste d'ingénieur d'études à l'Inra. Merci Vincent pour ton soutien, ta confiance depuis maintenant sept ans de collaboration continue, et pour m'avoir tenue à l'écart de l'avancement des projets lors de mon congé maternité.

Je remercie l'ensemble des personnes ayant contribué à ce travail de thèse et tout particulièrement les deux post-doctorantes de MIG, Fadia Bekkal Brikci et Catherine Tanous. Je remercie également Madame I. Martin-Verstraete ainsi que Messieurs P. Noirot, S. Aymerich et P. Bessières pour les discussions qui m'ont aidée à élaborer le modèle de connaissance de *Bacillus subtilis*. Je tiens enfin à remercier l'ensemble des partenaires du projet européen BaSysBio pour les données expérimentales de transcriptome et de métabolome que nous avons utilisées dans cette thèse.

Pour tous leurs efforts et leur disponibilité, je remercie chaleureusement les secrétaires de MIG, Mesdames S. Cally et J. Degrouard, et plus généralement toutes les personnes de MIG en particulier C. Caron, E. Montaudon et V. Martin pour le support informatique ainsi que Aude, Élodie, Laurent, Leslie et Safta.

Enfin un grand merci à mes proches pour leur soutien inconditionnel tout au long de ces années, et surtout à toi Richard pour ton aide et ta disponibilité ces derniers mois pour Lucie et moi.

Résumé

Cette thèse consiste à étudier l'organisation du système de contrôle des voies métaboliques des bactéries afin de dégager des propriétés systémiques révélant son fonctionnement. Dans un premier temps, nous montrons que le contrôle des voies métaboliques est hautement structuré et peut se décomposer en modules fortement découplés en régime stationnaire. Ces modules possèdent des propriétés mathématiques remarquables ayant des conséquences importantes en biologie. Cette décomposition, basée intrinsèquement sur la vision système de l'Automatique, offre un cadre théorique formel général d'analyse du contrôle des voies métaboliques qui s'est révélé effectif pour analyser des données expérimentales.
Dans un deuxième temps, nous nous intéressons aux raisons possibles de l'émergence de cette structure de contrôle modulaire. Nous identifions un ensemble de contraintes structurelles agissant au niveau de la répartition d'une ressource commune, les protéines, entre les processus cellulaires. Respecter ces contraintes pour un taux de croissance donné conduit à formaliser et résoudre un problème d'optimisation convexe non différentiable, que nous appelons Resource Balance Analysis. Ce problème d'optimisation se résout numériquement à l'échelle de la bactérie grâce à un problème de Programmation Linéaire équivalent. Plusieurs propriétés sont déduites de l'analyse théorique du critère obtenu. Tout d'abord, le taux de croissance est structurellement limité par la répartition d'une quantité finie de protéines entre les voies métaboliques et les ribosomes. Ensuite, l'émergence des modules dans les voies métaboliques provient d'une politique générale d'économie en protéines chez la bactérie pour gagner du taux de croissance. Certaines stratégies de transport bien connues comme la répression catabolique ou la substitution de transporteurs haute/basse affinité sont prédites par notre méthode et peuvent alors être interprétées comme le moyen de maximiser la croissance tout en minimisant l'investissement en protéines.
Mots-clés : Biologie des systèmes, réseaux de régulation génétique et métabolique, modularité, limitation du taux de croissance, gestion des ressources, Resource Balance Analysis, optimisation convexe, prédiction des modules.

Abstract

This thesis consists in studying the organization of the control system of metabolic pathways of bacteria to identify systemic properties revealing its operation. At first, we show that control of metabolic pathways is highly structured and can be decomposed into modules strongly decoupled in steady-state. These modules are defined by their singular mathematical properties having important implications in biology. This decomposition, based inherently on the system outlook of automatic control, offers a formal theoretical analysis of general control of metabolic pathways, which has been effective in analyzing experimental data.
In a second step, we consider the possible reasons for the emergence of this modular control structure. We identify a set of structural constraints acting at the distribution of a common resource, the proteins between cellular processes. Satisfying these constraints for a given growth rate leads to formalize and to solve a non-differentiable convex optimization problem, that we call Resource Balance Analysis. This optimization problem is solved numerically at the scale of the bacteria through an equivalent linear programming problem. Several properties are derived from theoretical analysis of the obtained criterion. First, the growth rate is structurally limited by the distribution of a finite amount of proteins between the metabolic pathways and the ribosomes. Second, the emergence of modules in metabolic pathways arises from a policy of economy in proteins in the bacterium to increase the growth rate. Some well known transport strategies such as catabolite repression or the substitution between low/high affinity transporters are predicted by our method and could consequently be interpreted as ways to maximize growth while minimizing investment in proteins.

Keywords : Systems biology, genetic and metabolic regulatory networks, modularity, growth rate limitation, resource management, Resource Balance Analysis, convex optimization, module prediction.

Avertissement au lecteur

Cette thèse s'inscrit pleinement entre deux disciplines très différentes que sont l'automatique et la biologie. Même si l'objet central de cette thèse est la bactérie, les outils et méthodes développés ici trouvent leur origine dans le champs de l'automatique. Nous avons toutefois essayé de rédiger ce manuscrit afin qu'il puisse être lu par les deux communautés sachant que certaines exigences des deux communautés sont difficilement conciliables.

Pour les biologistes : *L'automatique est la science du contrôle des systèmes, et par définition même, considère et manipule des concepts et des objets variés dans des disciplines différentes, telles l'aéronautique ou l'automobile et plus récemment la biologie. L'une des spécificités de l'automatique est d'utiliser des modèles abstraits intégrant uniquement les phénomènes essentiels au problème considéré. Ainsi nous avertissons le lecteur biologiste que nous avons délibérément réduit les détails de biologie dans l'écriture pour permettre, espérons-le, aux automaticiens de suivre le fil du raisonnement. Nous avons néanmoins considéré ces (très nombreux) détails dans les approches proposées dans cette thèse.*

Pour les automaticiens : *Malgré nos efforts pour réduire la complexité nécessaire à nos développements, la bactérie demeure un système d'une grande complexité composé d'un très grand nombre d'éléments différents. Nous avons décrit les mécanismes principaux le plus succinctement possible dans une perspective systémique.*

Conventions utilisées

On utilisera les conventions classiques de biologie pour nommer les gènes et les protéines. Le nom de gène s'écrit en italique et ne commence pas par une majuscule, alors que la protéine associée ne s'écrit pas en italique et commence par une majuscule. Par exemple, le gène *purR* désigne le gène codant pour la protéine PurR. Les métabolites sont notés en format de texte normal et sont nommés explicitement ou à travers un acronyme usuellement utilisé par les biologistes. Par exemple, le métabolite adénosine-triphosphate n'est jamais désigné en tant que tel dans les articles de biologie. Ce métabolite est toujours nommé par son acronyme ATP, et ce sera également le cas dans cette thèse. Pour plus de clarté pour le lecteur non biologiste, nous nous sommes efforcés de préciser la nature de l'acronyme. Toutefois, en cas de doute, le lecteur se reportera à la table des abréviations.

Table des matières

Introduction **1**

I Les modules dans les voies métaboliques **9**

1 Modules dans les systèmes biologiques **11**
1.1 Le réseau de régulation génétique 12
1.2 Motifs statistiquement sur-représentés 14
1.2.1 Les *Feedforward loops* en Biologie 15
1.2.2 Une possible hiérarchisation du contrôle 17
1.2.3 Les boucles négatives et positives 19
1.2.4 Conséquences 21
1.3 Structure du réseau métabolique 22
1.3.1 Identification de métabolites «hubs» 22
1.3.2 Les modes élémentaires 24
1.4 Les notions élémentaires d'Automatique dans les systèmes biologiques . . 25

2 Structure modulaire du réseau métabolique des bactéries **29**
2.1 Un modèle de connaissance exhaustif 30
2.1.1 Les modèles globaux du métabolisme existant 30
2.1.2 Un modèle pour *B. subtilis* intégrant les régulations 31
2.2 Topologie du réseau de régulation des voies métaboliques de *B. subtilis* . 32
2.2.1 Le rôle clé des métabolites 33
2.2.2 Organisation du réseau de régulation génétique 34
2.2.3 Deux structures spécifiques de contrôle 36
2.2.4 Vers l'élaboration d'un cadre formel pour l'analyse du réseau métabolique 38
2.3 Structure de contrôle par le produit final (SCPF) 39
2.3.1 Définition du système étudié 39
2.3.2 Caractérisation du régime d'équilibre 45
2.3.3 Intégration de différentes composantes biologiques 54
2.4 Vers une notion de module élémentaire 57
2.4.1 Structure de contrôle par le produit initial (SCPI) 58
2.4.2 Impact d'un régulateur global 62

2.4.3 Interconnexion entre les modules 63
2.4.4 Exemples de connexions présentes dans le réseau métabolique de *Bacillus subtilis* 69
2.4.5 Comparaison avec *Escherichia coli* 79
2.5 Conclusion 79

3 Identification et validation des modules chez *Bacillus subtilis* 81
3.1 Décomposition modulaire du réseau métabolique 82
3.1.1 Les modules élémentaires identifiés chez *B. subtilis* 82
3.1.2 Les régulateurs globaux 82
3.1.3 Une représentation modulaire de la fonction métabolique 84
3.2 Des questions biologiques ouvertes 84
3.2.1 Redondance des gènes 87
3.2.2 Isoenzyme et redondance 87
3.2.3 Le rôle critique de la première enzyme 88
3.3 Confrontation à des données expérimentales 91
3.3.1 Exemple d'une structure de contrôle par le produit final : la synthèse de la lysine 92
3.3.2 Exemple d'une structure de contrôle par le produit initial : la voie basse de la glycolyse 99
3.3.3 Adaptation des voies métaboliques à faible taux de croissance . . 102
3.4 Conclusion 112
3.A Identification des modules 115

II Vers une prédiction des modules 119

4 La méthode Resource Balance Analysis 121
4.1 Identification des contraintes structurelles en phase exponentielle de croissance 122
4.1.1 Une vision systémique de la cellule 122
4.1.2 Répartition d'une ressource commune, les protéines, entre différents sous-systèmes 125
4.1.3 La densité de composants intracellulaires est limitée 126
4.1.4 La loi de conservation de la masse 126
4.2 Formalisation mathématique des contraintes structurelles 127
4.2.1 Notations 127
4.2.2 De la répartition des protéines à la formalisation de trois contraintes structurelles 128
4.2.3 Un problème de faisabilité à μ donné : $P_f(\mu)$ 131
4.3 Un problème sympatique à résoudre 132
4.3.1 Toute solution faisable est non nulle 132
4.3.2 Un résultat technique intéressant 133
4.3.3 Convexité de $P_f(\mu)$ 133

4.3.4 L'ensemble des μ faisables est compact 134
4.4 Équivalence entre $P_f(\mu)$ et un problème de Programmation Linéaire . . 136
4.5 Différents mécanismes pour économiser des protéines 140
4.5.1 Prédiction théorique des structures de régulation dans le réseau métabolique 142
4.5.2 Sensibilité de la solution à des variations de P_G, k_{E_i} et k_T 143
4.6 Comparaison de la méthode RBA avec la littérature 147
4.7 Vers une méthode plus prédictive 150
4.8 Conclusion 154
4.A Démonstration du lemme 4.2 155
4.B La proposition 4.10 et sa démonstration 156
4.C Dépendance des paramètres k_T et k_{E_i} avec μ 159
4.C.1 Extension des propositions 4.1, 4.3 et 4.4 pour k_T dépendant de μ 159
4.C.2 Extension des propositions 4.1, 4.3 et 4.4 pour k_E dépendant de μ 160

5 Application de la méthode RBA à *Bacillus subtilis* 163
5.1 Construction de $P_f^{lp}(\mu)$ pour *Bacillus subtilis* 164
5.1.1 Calcul des cœfficients 165
5.1.2 Identification des paramètres 167
5.2 Répartition des ressources vis-à-vis du taux de croissance 169
5.2.1 Répartition des protéines entre les ribosomes et les voies métaboliques 169
5.3 Conséquences pour les voies métaboliques 173
5.3.1 Impact sur la biosynthèse des acides aminés 173
5.3.2 Prédiction de l'adaption des voies métaboliques pour différents milieux 178
5.4 Prédiction des modules élémentaires 182
5.5 Transport des ressources extracellulaires 184
5.5.1 Substitution de transporteurs haute et basse affinité 185
5.5.2 Prédiction de la hiérarchie des sources de carbone 186
5.5.3 Prédiction de la hiérarchie des sources d'azote 188
5.5.4 Vers la prédiction des régulations globales 189
5.6 Conclusion 189

Conclusion générale 191

Annexes 196

A Bases de Biologie 197
A.1 Les différents modes de vie de *Bacillus subtilis* 197
A.2 Les processus cellulaires nécessaires à la croissance exponentielle 198
A.3 La régulation de l'expression des gènes 201

B Résultats associés au réseau métabolique de *Bacillus subtilis* 205
B.1 Modèle de connaissance 205
B.2 La répartition des flux métaboliques de *Bacillus subtilis* pour différents milieux 205
B.2.1 Variations des acides aminés 205
B.2.2 Variations des sources de carbone 205

Bibliographie 206

Abréviations 224

Introduction générale

Décrypter le fonctionnement d'une cellule vivante reste à ce jour un enjeu majeur en Biologie tant les potentiels dans le milieu médical, pharmaceutique ou agroalimentaire sont importants. Lorsque la cellule s'adapte à des stimuli, il s'agit de comprendre quels sont les mécanismes biologiques en jeu, d'identifier précisément leur réponse et enfin de caractériser le système de contrôle coordonnant ces mécanismes.
En tant que science expérimentale basée sur l'observation de phénomènes biologiques, les avancées majeures sont intrinsèquement liées au progrès technologique. Au début du 20ème siècle, il était acquis que les organismes comme l'homme étaient composés de différents modules, les organes, connectés et coordonnés entre eux à travers d'une part le système nerveux pour l'information et le contrôle, et d'autre part le réseau vasculaire pour apporter entre autres les nutriments et l'énergie nécessaire à leur fonctionnement. L'avènement de la biologie moléculaire au cours de ce siècle associé à des découvertes majeures, comme l'existence de l'ADN, ont permis de passer de l'échelle macroscopique à l'échelle cellulaire. De façon remarquable, les bactéries tout comme les cellules humaines sont composées des mêmes processus cellulaires élémentaires (réplication de l'ADN, traduction des protéines par les ribosomes par exemple), même si bien entendu les cellules humaines (eucaryotes) sont beaucoup plus complexes et organisées que les bactéries. Ce changement d'échelle a profondément modifié la manière d'appréhender les systèmes biologiques par les chercheurs. Face à la complexité des mécanismes biologiques d'une part, et à la diversité des techniques nécessaires à leur étude d'autre part, les biologistes se sont spécialisés non seulement dans des processus biologiques particuliers mais aussi dans des disciplines scientifiques (biochimie, biophysique, génétique, ...). Cette approche dite réductionniste de la Biologie a permis des avancées scientifiques sans précédent. On citera par exemple la révolution associée à la découverte de l'ADN ou au premier séquençage d'un génome. Mais au fil du temps, la spécialisation de la Biologie a aussi été accompagnée par une certaine perte de vue du fonctionnement général de la cellule et *in fine* par une difficulté à intégrer et à réconcilier des données de différente nature. Or la fin du 20ème siècle a justement été marquée par l'explosion de technologie haut-débit permettant d'acquérir pour les bactéries des données d'excellente qualité à grande échelle pour tous les niveaux de la cellule : transcriptome, protéome, fluxome, intéractome, métabolome...

Face à cette masse de données, le biologiste a besoin maintenant de méthodes et d'outils intégratifs mais plus encore d'une vision «système» de la cellule, ce qui a donné naissance à une nouvelle thématique de recherche, la «Biologie des Systèmes» au début

du 21ème siècle [111] :

> *System Biology aims to develop a system-level understanding of biological systems. System-level understanding requires a set of principles and methodologies that links the behaviors of molecules to system characteristics and functions. Ultimately, cells, organisms, and human beings will be described and understood at the system level grounded on a consistent framework of knowledge that is underpinned by the basic principles of physics.*

Le but de la Biologie des Systèmes est ainsi non seulement de réconcilier les données entre elles, mais surtout de dégager des principes génériques de fonctionnement sur les cellules vivantes. Dans [111], l'auteur postule que la vision système, la théorie du contrôle et les outils méthodologiques associés issus de l'Automatique sont les clés qui permettront de révéler les grands principes de fonctionnement des cellules vivantes. Un des objectifs de cette thèse consiste à évaluer le potentiel et l'intérêt de la vision système et des outils méthodologiques développés dans le cadre de l'Automatique pour analyser le fonctionnement des systèmes biologiques à l'échelle globale. Si des concepts élémentaires comme la boucle de rétroaction ont été identifiés et appliqués avec succès pour analyser le fonctionnement de sous-parties spécifiques de la cellule comme la mobilité chez les bactéries [13], leur application systématique à l'échelle de la cellule entière afin de dégager des principes généraux de fonctionnement reste inexplorée. La question est tout simplement de savoir si une telle analyse est possible à l'échelle de la cellule entière.

Les systèmes biologiques sont en effet d'une incroyable complexité comparés aux systèmes conçus par les hommes. Les systèmes biologiques sont composés de milliers d'entités différentes correspondant ainsi à des systèmes qui peuvent être décrits par plus de 10000 variables d'états et ce déjà pour les organismes les plus simples comme les bactéries. A titre de comparaison, les systèmes physiques construits par les hommes sont représentés par des modèles simplifiés de «petite» dimension : ordre 2 pour un missile [67], ordre 16 pour un avion [58], ordre 2 pour un bief de canal d'irrigation [128], ordre 9 pour une transmission automatique intégrant moteurs thermiques et électriques pour un véhicule [180]. Développer un modèle dynamique de la cellule dans sa totalité reste donc utopique à ce jour compte tenu de la seule dimension du système. A cela s'ajoute la difficulté de décrire les dynamiques par des équations non linéaires ou encore le nombre de paramètres du modèle à identifier. Il est donc indispensable de simplifier ce système, tout en identifiant et en intégrant les éléments clés nécessaires au fonctionnement global de la cellule. C'est ici que le savoir accumulé par les biologistes ces dernières années doit être intégré et reste plus que jamais indispensable.
La vision systémique de l'Automatique semble donc bien pertinente pour analyser les systèmes biologiques car intrinsèquement celle-ci consiste d'une part à décomposer un système très compliqué en un ensemble de sous-systèmes possédant de bonnes propriétés locales, et d'autre part à étudier *a posteriori* les propriétés globales résultant de l'interconnexion de ces sous-systèmes. Dans les systèmes conçus par les hommes, les propriétés du système global ne sont pas données par les propriétés locales de chaque sous-système. Dans cette thèse, nous nous efforcerons donc de décomposer une cellule vivante en un ensemble de sous-systèmes fonctionnels permettant d'une part de réduire la complexité

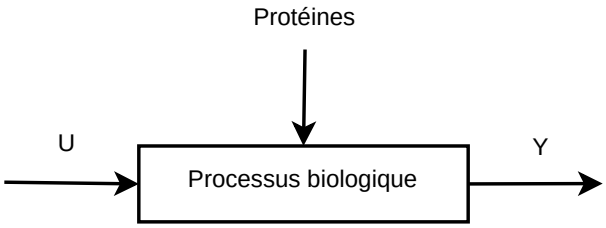


FIG. 1 – Un processus biologique représenté comme un module

intrinsèque du système global, et d'autre part de dégager des principes généraux (et génériques !) de fonctionnement et d'organisation de la cellule. L'angle d'attaque d'une telle démarche est toutefois difficile et nécessite de s'investir pleinement dans la Biologie afin d'analyser le fonctionnement global de la cellule. Bien que délicate et coûteuse en temps, cette démarche mêlant étroitement Automatique et Biologie constitue le coeur et l'originalité de ce travail de thèse, ce qui, de fait, est complémentaire des nombreuses études basées sur de l'analyse de données présentées au cours du premier chapitre dont les objectifs ne sont pas de remettre en perspective les résultats dans le contexte du fonctionnement global du système.

Dans ce contexte, nous pouvons dès à présent présenter une première analyse systémique gros grain du fonctionnement global d'une cellule. Toute cellule vivante est organisée en différents processus biologiques élémentaires effectuant des tâches spécifiques [86], représentées sur la figure 2 et décrites dans l'annexe A.2. Si on représente ces processus biologiques d'un point de vue entrée/sortie (voir la figure 1), chaque processus est composé entre autres de protéines transformant des entrées U en sorties Y. Dans la majorité des processus biologiques composant la cellule, les entrées U correspondent à de l'énergie et certains métabolites particuliers, appelés précurseurs métaboliques. Les sorties Y correspondent aux différents composants cellulaires (ADN, ARN, protéines, etc.). A ce titre, deux processus biologiques sont remarquables :

- *l'appareil de traduction des protéines* : qui prend en entrée U de l'énergie et des ARNt chargés pour produire en sortie Y les protéines composant tous les processus cellulaires ;
- *le réseau métabolique* : qui prend en entrée U des nutriments du milieu extracellulaire et qui produit en sortie Y les différents précurseurs métaboliques et l'énergie nécessaire à tous les processus cellulaires.

Bien qu'absente dans la littérature, cette décomposition systémique de la cellule constitue le fondement des résultats de cette thèse. Compte tenu de la diversité des milieux et de la compétition féroce existante entre toutes formes de vie, les cellules ont développé des mécanismes de régulation leur permettant non seulement de s'adapter, mais aussi pour les bactéries de tirer leur épingle du jeu. Les cellules possèdent un système de contrôle adéquat afin d'assurer à chaque instant d'une part la régulation de chaque sous-système et d'autre part leur coordination pour réaliser des stratégies complexes issues de l'évolution et de la sélection naturelle. Dans ce contexte, l'étude des régulations des pro-

cessus biologiques d'un point de vue expérimental et théorique est en pleine expansion. Les biologistes ont identifié de nombreux mécanismes de régulation en particulier au niveau des voies métaboliques pour certains organismes modèles comme *Escherichia coli*, *Bacillus subtilis* ou *Saccharomyces cerevisiae*. L'objectif de ces études expérimentales et théoriques est de dégager des propriétés suffisamment générales pour être conservées potentiellement chez d'autres organismes. Comme nous le montrerons dans le premier chapitre, il existe des régularités dans une sous-partie du système de contrôle global. Ces régularités sont souvent appelées «modules» par les biologistes [4, 12], mais elles ne correspondent pas en réalité à des sous-systèmes que définiraient les automaticiens, c'est-à-dire définis à travers leurs bonnes propriétés locales.

L'objectif central de cette thèse consiste ainsi à étudier l'organisation du système de contrôle des processus cellulaires afin de dégager des propriétés systémiques révélant son fonctionnement. Cette question est très ambitieuse et nécessite une très bonne connaissance du système de contrôle agissant à tous les niveaux de la cellule au moins pour un organisme spécifique. Compte tenu du niveau de connaissance actuel du système de contrôle pour les différents processus cellulaires, nous nous sommes focalisés sur le système de contrôle des voies métaboliques de la bactérie modèle de la famille des Gram + *Bacillus subtilis*[1]. Il convient néanmoins de veiller à la généricité de nos résultats en les comparant à ceux obtenus pour un autre organisme comme par exemple *Escherichia coli*, la bactérie modèle de la famille des Gram -. Nos résultats n'en seront alors que renforcés surtout si l'on considère que ces deux bactéries ont divergé il y a plus de trois milliards d'années [14]. Au cours des différents chapitres de cette thèse, nous montrerons qu'effectivement la structure du contrôle des voies métaboliques peut se décomposer en modules possédant de bonnes propriétés locales tant au niveau théorique qu'au niveau biologique. Cette décomposition, basée intrinsèquement sur la vision système de l'Automatique, offre un cadre théorique formel général d'analyse du contrôle des voies métaboliques et représente ainsi une contribution majeure de ce travail de thèse. Mais l'intérêt de l'application des concepts de l'Automatique en Biologie ne s'arrêtent pas là. En effet, il est bien connu en Automatique que par défaut la structure d'un correcteur obtenu par des algorithmes classiques comme le H_∞ n'est pas décentralisée. Lors de la conception d'un correcteur décentralisé, l'automaticien doit ajouter des contraintes sur le correcteur afin de lui imposer une structure spécifique, tout en imposant des critères de performance requis par le cahier des charges [200]. Sur cette base, nous avons exploré dans un deuxième temps un ensemble de contraintes structurelles possibles agissant sur la bactérie permettant d'expliquer l'émergence de cette structure de contrôle modulaire.

[1] En tant qu'organisme modèle, le niveau de connaissance, en particulier concernant les mécanismes de régulation est en effet important. Sans être complète, la structure de contrôle est néanmoins suffisamment représentative pour dégager des résultats génériques. Par ailleurs, *Bacillus subtilis* est une bactérie d'intérêt à la fois pour l'agro-alimentaire (production d'amylase ou d'antibiotiques), et comme organisme modèle pour des bactéries pathogènes phylogénétiquement proches telles que *Staphylococcus aureus*, *Streptococcus pneumoniae*, *Bacillus anthracis* (responsable de l'anthrax ou maladie du charbon), *Bacillus cereus* et *Listeria monocytogenes* (à l'origine d'infections alimentaires). Une meilleure compréhension du contrôle des voies métaboliques chez *Bacillus subtilis* est une première étape vers la compréhension des coordinations des processus cellulaires.

L'identification et la formalisation mathématique de ces contraintes en un problème d'optimisation constituent une autre contribution majeure de cette thèse.

Ce mémoire de thèse se décompose en deux parties, suivies d'un chapitre d'annexes rappelant les principales notions de Biologie nécessaires à la compréhension de cette thèse (annexe A). La première partie est consacrée à l'existence, l'étude et la validation des modules dans les voies métaboliques, bien identifiés du point de vue entrée/sortie, autour de trois chapitres.

Modules dans les systèmes biologiques (chapitre 1) Au cours de chapitre, nous introduisons la problématique de cette première partie à travers une étude bibliographique. L'objectif est de présenter les principaux résultats obtenus dans la littérature sur l'organisation du contrôle des processus biologiques à l'échelle de la cellule. Nous montrerons qu'il existe des régularités statistiquement sur-représentées dans la structure de contrôle, qui ont été appelées «modules» par les biologistes. Néanmoins, une vision systémique de la cellule est nécessaire pour obtenir une définition bien posée de ces modules d'un point de vue entrée/sortie. Au regard de cette étude, nous donnons les éléments à intégrer, et en particulier le couplage entre les métabolites et les régulations génétiques.

Structure modulaire du réseau métabolique de *Bacillus subtilis* (chapitre 2) Dans ce chapitre, nous identifions deux structures de contrôle aux propriétés mathématiques remarquables dans les voies métaboliques de *Bacillus subtilis*, et qui à ce titre, correspondent à des modules bien identifiés du point de vue entrées/sorties. Nous verrons que les propriétés obtenues offrent un cadre théorique pour expliquer non seulement le contrôle local d'un module, mais aussi les coordinations entre ces modules.

Identification et validation des modules chez *Bacillus subtilis* (chapitre 3) Ce chapitre confronte les résultats théoriques présentés au chapitre précédent avec des données issues de la littérature et des données expérimentales obtenues dans le cadre du projet européen BaSysBio[2]. Nous identifions les modules dans les voies métaboliques de *Bacillus subtilis* et illustrons leur comportement face à des données de transcriptome et de métabolome.

[2]Ce projet de Biologie des Systèmes (2007-2010) a pour but d'améliorer la compréhension théorique des différents mécanismes de régulation intervenant à tous les niveaux de la bactérie dans deux conditions physiologiques différentes : le passage d'une source privilégiée de carbone (le glucose) à une autre (le malate), et le stress oxydatif (présence de peroxyde). Dans ce contexte, nous sommes intervenus pleinement dans les différentes phases du projet : construction et analyse de modèles dynamiques centrés sur les questions biologiques d'intérêt, traitement et analyse des données expérimentales, validation des modèles. Dans cette thèse, nous présenterons uniquement les aspects liés au traitement et à l'analyse des données pour notre propre intérêt, à savoir la validation des modules et de notre cadre formel d'analyse des voies métaboliques.

La deuxième partie de cette thèse est consacrée à l'identification, la formalisation et la validation des contraintes structurelles permettant d'expliquer l'émergence d'une structure de contrôle modulaire dans les voies métaboliques à travers deux chapitres.

La méthode Resource Balance Analysis (chapitre 4) Ce chapitre regroupe les résultats majeurs de ce travail de thèse. A travers une analyse systémique de la bactérie, nous identifions trois contraintes structurelles actives à chaque instant basées sur la répartition des protéines entre les processus cellulaires. Leur formalisation mathématique en phase exponentielle conduit à un problème d'optimisation non différentiable dont nous étudierons les propriétés mathématiques. Ce cadre formel est appelé **Resource Balance Analysis**.

Application de la méthode RBA à *Bacillus subtilis* (chapitre 5) Ce chapitre présente la mise en œuvre de la méthode RBA sur le réseau métabolique de *B. subtilis*. Nous illustrons les résultats théoriques obtenus au chapitre précédent.

Ces travaux de thèse ont conduit à deux publications :

- Goelzer, A. and Bekkal Brikci, F. and Martin-Verstraete, I. and Noirot, P. and Bessières, P. and Aymerich, S. and Fromion, V. *Reconstruction and analysis of the genetic and metabolic regulatory networks of the central metabolism of* Bacillus subtilis. BMC System Biology, 2008. 2 :20.
- Goelzer, A. and Fromion, V. and Scorletti, G. *Cell design in bacteria as a convex optimization problem.* Proceedings of the 48th IEEE Conference on Decision and Control, 2009. pp :4517-4522.

Dans la suite de cette thèse, nous parlerons indifféremment de cellule ou de bactérie.

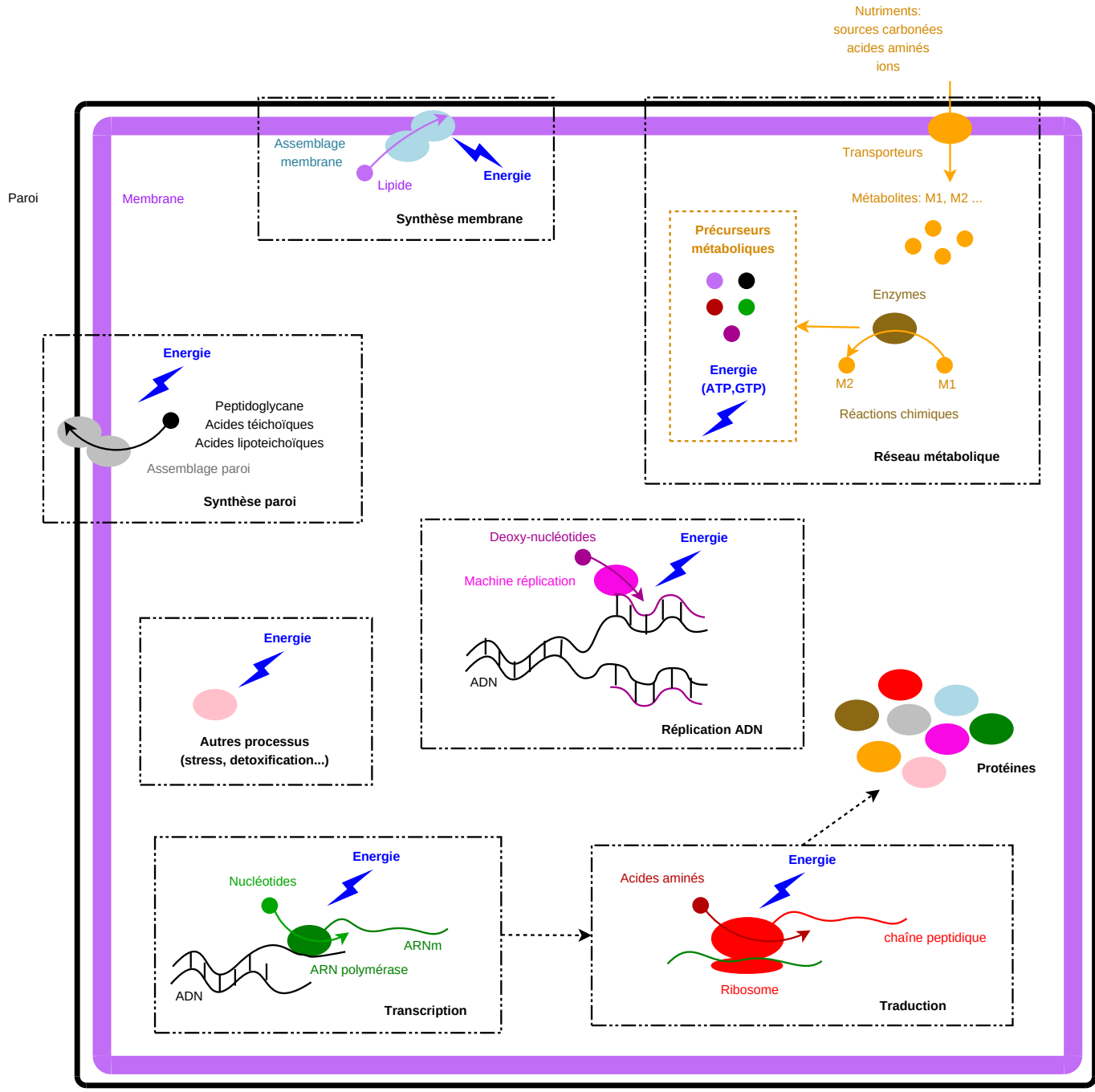


FIG. 2 – Les processus cellulaires majeurs de la cellule en phase de croissance exponentielle

Première partie

Les modules dans les voies métaboliques

Chapitre 1

Modules dans les systèmes biologiques

Comme nous l'avons montré en introduction de cette thèse, tout être vivant, en partant de l'homme pour finir aux microorganismes comme les bactéries apparaissent généralement composés de sous-systèmes effectuant des tâches spécifiques différentes de façon coordonnée. Au niveau macroscopique pour les hommes par exemple, ces sous-systèmes correspondent aux organes, alors qu'à l'échelle microscopique, pour les bactéries par exemple, ces sous-systèmes correspondent aux différents processus cellulaires comme la réplication de l'ADN, la traduction des protéines, etc. Au cours du 20ème siècle, ces différents sous-systèmes ont été particulièrement étudiés au niveau moléculaire si bien que leur principe de fonctionnement est actuellement bien connu. En revanche formaliser et rationnaliser l'ensemble de ces sous-systèmes afin de dégager des principes de fonctionnement généraux reste actuellement un défi majeur.
Le fonctionnement d'un tel système nécessite en effet un contrôle adéquat pour répondre à des perturbations variées (un stress oxydatif, des carences en minéraux ou en carbone, des mutations génétiques...) dans un contexte de compétition perpétuelle avec les autres organismes vivants. Dans une cellule, il existe différents mécanismes de régulation identifiés expérimentalement pour contrôler ces processus cellulaires. Ils peuvent agir au niveau de l'expression des gènes (niveau transcriptionnel), de la traduction de la protéine (niveau traductionnel), de la modification de l'activité de la protéine (niveau post-traductionnel et allostérique). Jusqu'à la fin des années 90, ces mécanismes de régulation étaient étudiés au cas par cas pour un processus cellulaire donné. L'explosion technologique récente permet maintenant d'explorer de manière systématique et à l'échelle de la cellule certains mécanismes de régulation agissant au niveau de l'expression des gènes (transcriptome, Chip-Chip), de la traduction des protéines (protéome) ou de la phosphorylation des protéines (phospho-protéome).

Dans ce chapitre, nous allons montrer comment l'étude de la structure et du fonctionnement du réseau de régulation génétique d'une bactérie a été appréhendée dans la littérature afin de situer ce travail de thèse. Ces études sont principalement basées sur de l'analyse systématique de données à grande échelle par des méthodes statistiques. Il s'agit de définir l'objet que ces études ont considéré et de clarifier la démarche suivie pour savoir si leurs résultats peuvent s'interpréter par une démarche d'analyse des systèmes

propre à l'Automatique. La vision système propre à l'Automatique consiste à isoler des sous-blocs possédant des bonnes propriétés du point de vue entrées-sorties pour ensuite les interconnecter. Généralement, de bonnes propriétés locales pour chaque sous-système ne garantissent pas de bonnes propriétés globales au système interconnecté[1]. On cherchera également à évaluer dans quelle mesure les concepts élémentaires d'Automatique comme la boucle de rétroaction ou la boucle ouverte peuvent être projetés sur la totalité de la cellule dans toute sa complexité.

1.1 Le réseau de régulation génétique

Le réseau de régulation génétique peut se définir comme le graphe engendré par l'ensemble des régulateurs agissant au niveau de l'*expression des gènes*, c'est-à-dire au niveau transcriptionnel uniquement. Traditionnellement, les régulateurs considérés correspondent aux facteurs de transcription (FT) et aux facteurs sigma qui modulent l'initiation de la transcription des gènes (voir Annexe A.3 pour des détails complémentaires sur le sujet). Le graphe du réseau de régulation génétique est donc composé de nœuds et d'arêtes, où les nœuds correspondent aux régulateurs génétiques (FTs et facteurs sigma). Une arête relie un nœud A à un nœud B si A régule la transcription de B. B appartient au régulon de A, c'est-à-dire à l'ensemble des gènes sous le contrôle direct de A. Par ailleurs, le graphe est signé si à une arête est associé un signe (positif ou négatif) selon l'effet de la régulation sur la transcription. Un exemple de graphe signé et non signé est donné sur la figure 1.1 dans le cas de trois gènes A, B, C, où A régule B, qui régule C, qui régule A. Le régulon associé à un régulateur correspond à l'ensemble des gènes sous le contrôle direct de ce régulateur.
A ce stade, on notera déjà un premier problème associé à cette représentation, à savoir l'oubli d'un autre type de modulateurs transcriptionnels : les effecteurs. Les effecteurs sont des molécules qui modulent l'activité du facteur de transcription en se liant directement à celui-ci. Par exemple, lorsque l'effecteur est lié au facteur de transcription, le complexe résultant n'est alors plus capable de se fixer sur l'ADN, ou possède au contraire une plus grande affinité pour son site de fixation sur l'ADN. Généralement, cet effecteur correspond à un métabolite. Le résultat final sur la transcription du gène dépend donc d'une part de l'effet positif ou négatif du facteur de transcription, et d'autre part de l'effet positif ou négatif de l'effecteur. Généralement, le graphe signé du réseau de régulation génétique n'intègre pas les effets des effecteurs.

Remarque 1.1 *Nous ne nous intéresserons pas dans cette thèse à l'inférence du réseau de régulation à partir des données expérimentales. Cet aspect est néanmoins crucial lorsque le réseau de régulation génétique n'est pas bien connu, et constitue un champs de recherche important en Biologie des Systèmes. On citera juste quelques références permettant d'identifier le réseau de régulation à partir de :*

[1] De façon tout à fait remarquable, nous verrons au cours du chapitre 4 de cette thèse, que pour notre système biologique pourtant très complexe, l'optimisation de la performance de chaque sous-système contribuent à la performance globale du système.

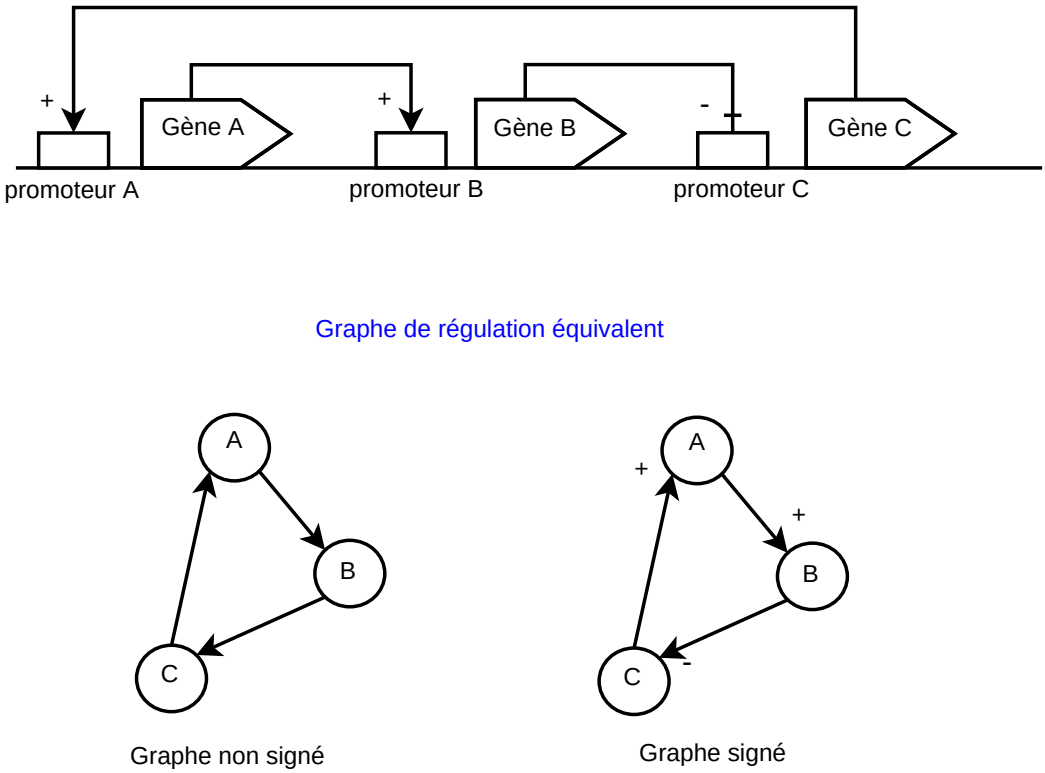


FIG. 1.1 – Un réseau de régulation où le gène A active le gène B, qui réprime le gène C, qui active le gène A

- *données de transcriptome par des algorithmes de clustering [53], bi-clustering [99, 98, 220], décomposition en valeurs singulières (SVD) [244] ;*
- *données Chip-Chip permettant de localiser des sites d'accrochage des FTs sur l'ADN [124, 246, 11] ;*
- *la combinaison de données Chip-Chip et de transcriptome [11, 246, 130] ;*
- *la combinaison de données de transcriptome et de données in silico de détection de promoteurs ou de sites d'accrochage de FT [90, 40].*

Les différentes études sur la structure du réseau de régulation génétique mentionnées dans les sections suivantes ont tenté de dégager des principes de fonctionnement des régulations, suffisamment généraux pour être pertinents sur différents organismes procaryotes ou eucaryotes. L'idée ici est que si la structure du réseau de régulation est conservée entre les différents organismes, alors les résultats obtenus sur un organisme spécifique pourront être appliqués à d'autres. C'est l'un des fondements (et le pari) de la Biologie des Systèmes.

1.2 Motifs statistiquement sur-représentés

Dans la plupart des travaux centrés sur ce sujet, la recherche d'une structure dans le réseau de régulation a été appréhendée à travers la recherche de motifs statistiquement sur-représentés, nommés *feedforward*, *feedback*, boucles positives, boucles négatives, dans le graphe des régulations signé ou non pour l'organisme modèle des bactéries Gram-*Escherichia coli* [12, 5, 78, 214, 225, 203, 147, 140] ou des eucaryotes *Saccharomyces cerevisiae* [124]. Dans cette étude bibliographique, nous nous centrerons sur celui d'*E. coli*. En effet, son réseau de régulation a été relativement bien identifié comparé à d'autres organismes et est disponible dans des bases de données [107, 181]. Nous allons décrire succinctement ces motifs, en précisant leurs propriétés mathématiques lorsqu'elles ont été étudiées, tout en les projetant dans le contexte de l'Automatique. L'existence de ces motifs est intéressante car ils indiquent que certaines régularités existent dans les réseaux génétiques, au moins pour ces deux organismes modèles eucaryote et procaryote.

On distinguera tout d'abord les études menées sur le graphe non signé des régulations selon deux axes : la recherche de motifs élémentaires statistiquement sur-représentés d'une part et la classification des facteurs des régulateurs génétiques pour révéler des possibles niveaux de contrôle différents d'autre part. Trois motifs différents appelés *Feed-forward loop* (FFL), *Single Input Module* (SIM) et *Dense Overlapping Regulons* (DOR) ont ainsi été identifiés dans le graphe non signé des régulations [203, 147, 140]. Dans un deuxième temps, on étudiera les boucles négatives et positives identifiées dans le graphe signé des régulations pour lesquels des résultats de stabilité (donc proche de l'Automatique) ont été obtenus [78, 214, 225].

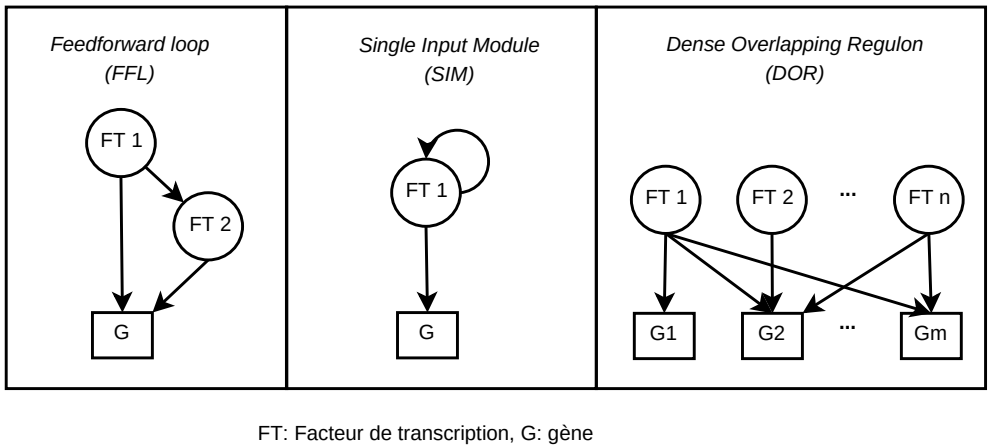


FIG. 1.2 – Les motifs récurrents dans le réseau de régulation génétique d'*E. coli* [203]

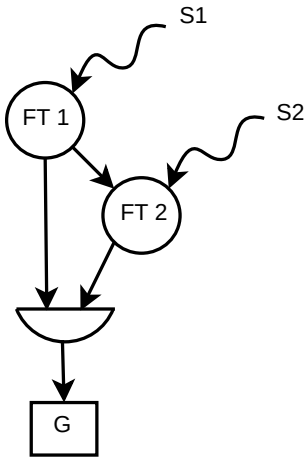


FIG. 1.3 – La *feedforward loop* [140]

1.2.1 Les *Feedforward loops* en Biologie

Les trois types de motifs affichés sur la figure 1.2 ont été identifiés dans le réseau de régulation génétique d'*E. coli* en explorant la matrice de connectivité[2] associée au graphe non signé du réseau de régulation [203] : la FFL (*Feedforward loop*), le SIM (*Single Input Module*) et le DOR (*Dense Overlapping Regulons*). Les deux premiers motifs sont clairs d'après la figure 1.2. Le dernier a été obtenu par classification (*clustering*) sur les régions denses du réseau, c'est-à-dire pour lesquelles un même groupe de gènes ou d'opérons sont régulés par un même ensemble de facteurs de transcription. On remarquera que ces motifs ne donnent pas le signe de la régulation du gène G_i. Pour un motif SIM, si la

[2]définie comme la matrice M de taille $n \times n$, où n est le nombre de facteurs de transcription intégrés dans le graphe, avec $M_{ij} = 1$ si le i-ème facteur de transcription régule le j-ème facteur de transcription, et $M_{ij} = 0$ sinon.

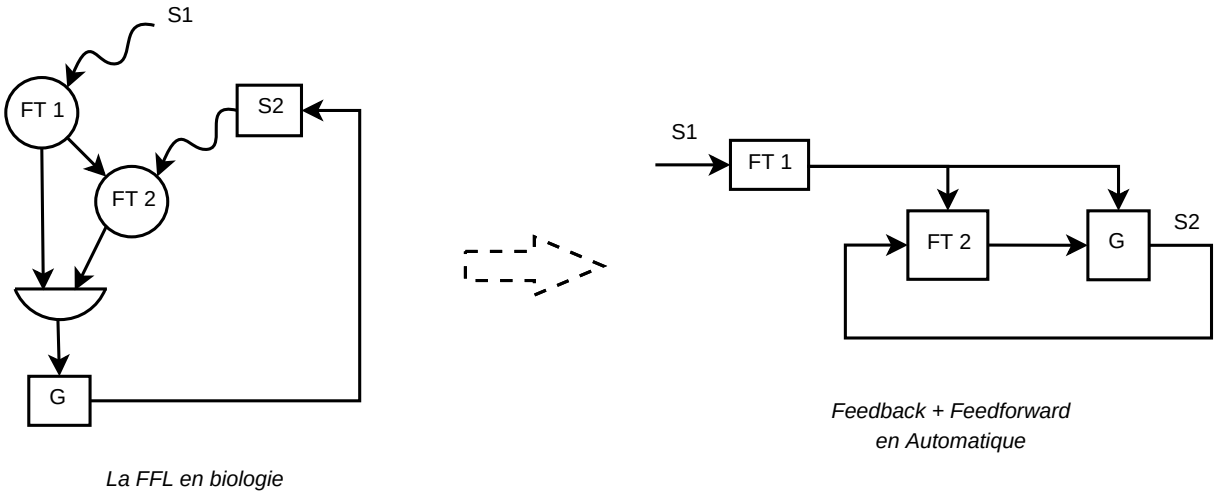


FIG. 1.4 – Le parallèle en Automatique

régulation est négative, ce motif correspond alors à une auto-régulation, ce qui représente 70% des motifs SIM. Le signe de la FFL dépend de l'intégration des effets des facteurs de transcription impliqués, eux-mêmes sensibles à leurs effecteurs S_1 et S_2 (voir la figure 1.3 et les). On distinguera deux grands types de FFLs : les FFLs «cohérentes» lorsque les signes des régulations issues des deux branches (directe et indirecte) sont les mêmes, et les FFLs «incohérentes» dans le cas contraire [203, 140]. Il semblerait, en simulation pour un jeu de paramètres donné, que les FFL incohérentes permettent d'accélérer le temps de réponse de la transcription du gène G. En revanche, les FFLs cohérentes permettraient de filtrer le signal S_1. Seul un long signal persistant $S_1 \neq 0$ permettrait d'avoir la transcription de G pour $S_2 \neq 0$ [140]. Cependant ces résultats ont été simplement déduits de simulations effectuées pour un jeu de paramètres donné. Aucune étude théorique sur l'obtention et la stabilité des points d'équilibre n'a été menée de manière systématique pour tous les cas de FFL. Toutefois, une étude théorique et expérimentale a montré qu'une FFL incohérente permet de générer un pulse au niveau de la transcription [141, 105].

Du point de vue de l'Automatique, la régulation associée aux facteurs de transcription FT_1 et FT_2 correspond bien à un feedforward au sens de l'Automatique, si on les effecteurs S_1 et S_2 ne sont pas considérés. Lorsque les effecteurs sont intégrés dans l'analyse, la représentation de la régulation dépend de l'impact du gène G sur ces effecteurs. Si, par exemple, G modifie pour une raison quelconque S_2, on obtient le type de représentation de la figure 1.4, correspondant à une boucle de rétroaction (*feedback*) et une boucle de pré-commande (*feedforward*) en Automatique. Cette projection des motifs dans le contexte de l'Automatique montre le rôle central des effecteurs dans la définition même des motifs. Une seconde approche de structure des régulations consiste à essayer de hiérarchiser les régulateurs en fonction du nombre de connexions dans le graphe et du nombre de régulations induites.

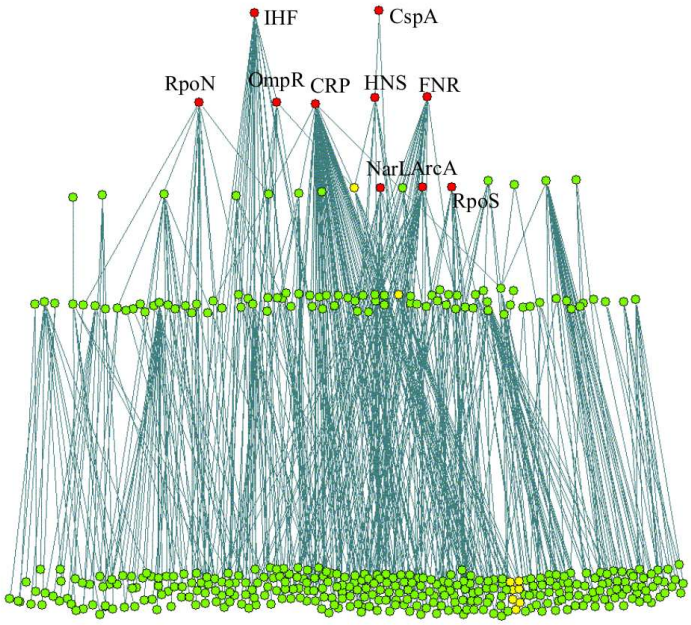


FIG. 1.5 – Structure multi-couche hiérarchique du réseau de régulation génétique d'*E.coli* (extrait de [132])

1.2.2 Une possible hiérarchisation du contrôle

L'organisation topologique du réseau de régulation génétique est ici explorée d'une part à travers le nombre de connexions pour un régulateur, et d'autre part en les ordonnant selon différents critères [132, 143, 66]. Sur les figures 1.5 et 1.6, les régulateurs génétiques d'*E. coli* sont hiérarchisés respectivement en fonction du nombre de régulateurs génétiques sous leur contrôle direct et indirect, et en fonction de leurs liens fonctionnels. Il apparaît une hiérarchisation du contrôle, avec quelques régulateurs «haut niveau» et fortement connectés comme le facteur de transcription CRP représenté sur les figures 1.5 et 1.6, qui sont impliqués dans la régulation de nombreux autres régulateurs. En revanche, la majorité des régulateurs n'induit aucune régulation, hormis les autorégulations (le régulateur se régule lui-même). Les régulateurs peuvent ensuite être regroupés suivant la fonction des gènes de leur régulon (figure 1.6). On obtient ainsi une représentation fonctionnelle et hiérarchique du réseau de régulation génétique [66]. Là encore, certains régulateurs interviennent dans plusieurs fonctions biologiques et ont ainsi un rôle pléiotrope. Malheureusement, ce type d'analyse ne permet pas d'identifier clairement le rôle d'un régulateur. Ici encore une classification intégrant les effecteurs métaboliques pourrait changer la structure. L'avantage de cette représentation est d'identifier les régulateurs fortement connectés qui sont potentiellement des bons candidats pour des régulations globales de haut niveau.

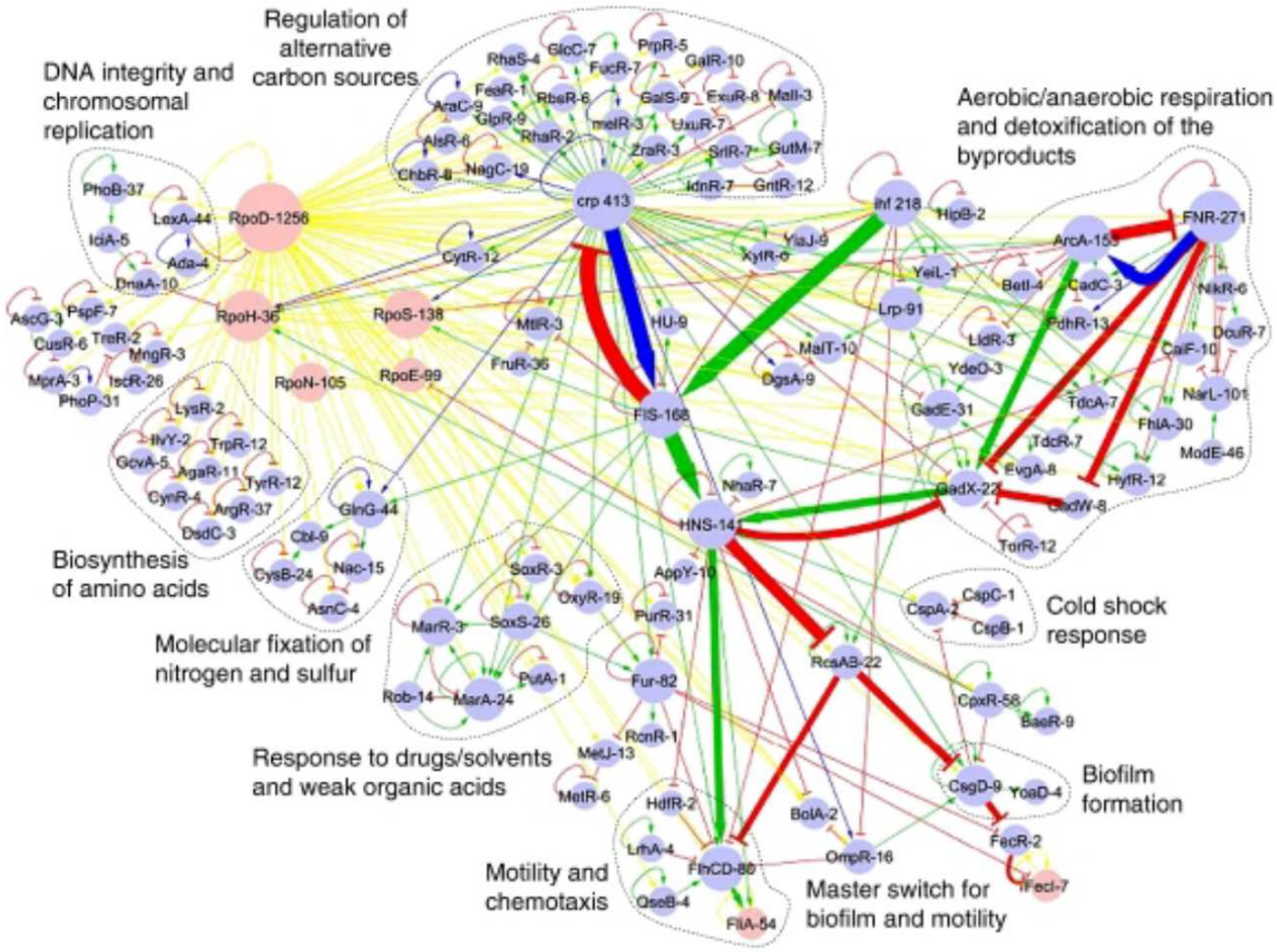


FIG. 1.6 – Organisation modulaire et fonctionnelle du réseau de régulation génétique d'*E.coli* (extrait de [143])

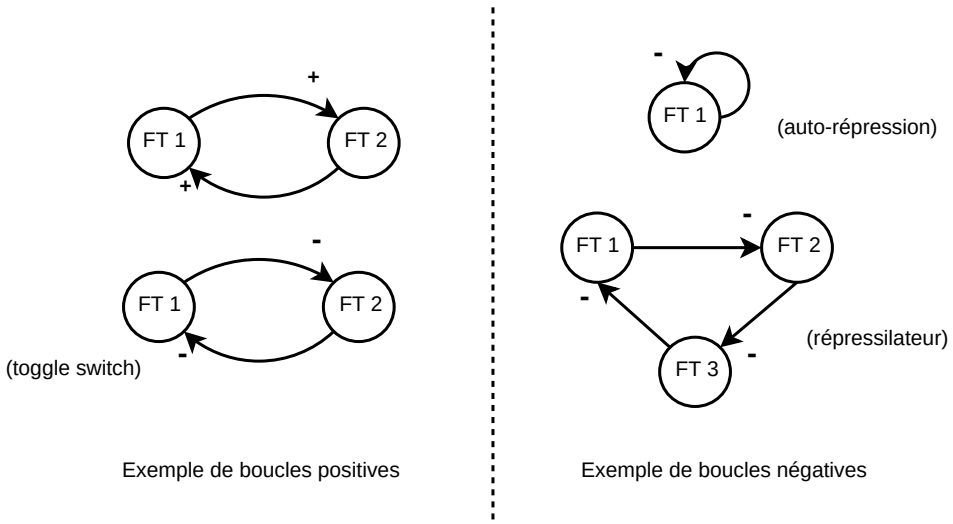


FIG. 1.7 – Exemples de boucles négatives et positives dans le réseau de régulation génétique

1.2.3 Les boucles négatives et positives

Étudions maintenant l'autre grand type de motifs statistiquement significatifs obtenus en considérant le graphe signé du réseau de régulation, les boucles négatives et positives. Quelques exemples parmi ceux étudiés dans la littérature sont indiqués sur la figure 1.7. Une boucle dans le réseau de régulation est ici définie comme un cycle impliquant plusieurs facteurs de transcription. Une boucle est dite négative (respectivement positive) si le produit des signes de toutes les arêtes la composant est négatif (respectivement positif).
De nombreux mathématiciens et automaticiens se sont intéressés aux propriétés associées à ces deux types de boucles, notamment en terme de dynamique et de stabilité. Le biologiste René Thomas a postulé il y a 20 ans, les deux conjectures suivantes [223] :

1. un système ayant plusieurs états stationnaires distincts non dégénérés possède au moins une boucle positive ;
2. un système présentant un comportement périodique stable possède au moins une boucle négative.

Ces deux conjectures ont des implications importantes en Biologie, car de tels comportements existent effectivement. On citera comme exemple de comportements stables l'horloge circadienne, ou l'homéostasie des ions (comme le pH), et comme exemple de multistationarité la différenciation cellulaire, ou l'entrée en sporulation pour *B. subtilis*. Ces conjectures ont été démontrées dans [78] sous des hypothèses réalistes pour les systèmes biologiques. La démonstration la plus générale de la première conjecture est donnée dans [214].
La recherche systématique de boucles positives (ou négatives) dans les réseaux de régulation génétique permet ainsi d'isoler des sous-systèmes présentant potentiellement

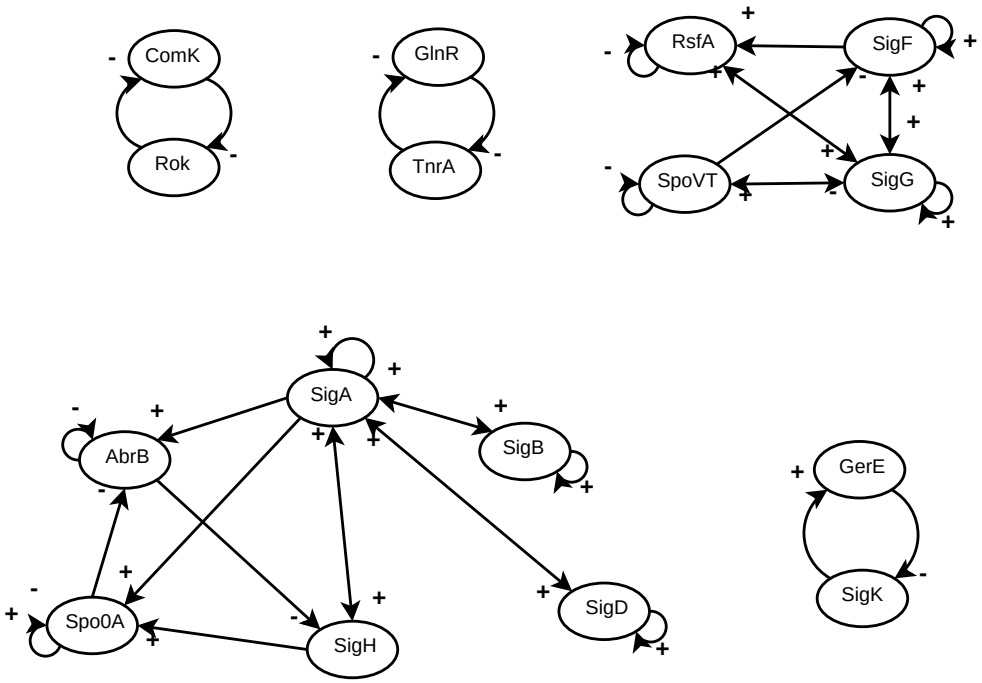


FIG. 1.8 – Les boucles identifiées dans le réseau de régulation de *Bacillus subtilis*

des propriétés dynamiques particulières. Les conjectures de Thomas étant des conditions nécessaires, il convient d'étudier au cas par cas les boucles observées. Si on recherche les boucles dans le réseau de régulation de *Bacillus subtilis*, (par des algorithmes calculant les composantes fortement connexes[3] par exemple [221]), on obtient les boucles représentées sur la figure 1.8. Parmi les 143 régulateurs composant le graphe, seuls 12 interviennent dans des boucles. Les boucles en tant que telles ne permettent donc pas de définir une structure dans la régulation. En revanche, elles peuvent révéler la présence d'états physiologiques mutuellement exclusifs. Par exemple, la boucle positive de type «*toggle switch*» (voir la figure 1.7 en bas à gauche) impliquant les facteurs de transcription GlnR et TnrA engendre deux états stables physiologiques distincts, correspondant effectivement à la présence (GlnR présent, TnrA absent) ou à l'absence (GlnR absent, TnrA présent) d'une source d'azote privilégiée comme la glutamine.
Les boucles négatives, et en particulier le cas de l'auto-régulation ont été étudiées de manière théorique [238], notamment pour leur parallèle en Automatique. La boucle de rétroaction en Automatique permet en effet de désensibiliser le système contrôlé vis-à-vis de perturbations non mesurées [96]. Dans [238], les auteurs montrent qu'une boucle de type auto-régulation ne possède qu'un seul point d'équilibre, qui de plus est stable.
Expérimentalement, les propriétés théoriques des boucles négatives/positives ont été observées à travers la construction et l'implémentation de circuits synthétiques dans *E. coli* [87, 89, 88], notamment le répressilateur [55] (figure 1.7 en bas à droite), le *toggle switch*

[3]Une composante fortement connexe d'un graphe est un sous-ensemble maximal de sommets tels que deux quelconques d'entre eux soient reliés par un chemin.

[71] (figure 1.7 en bas à gauche), l'auto-régulation [17] (figure 1.7 en haut à gauche). Selon ces études, le répressilateur présente un comportement oscillatoire, le *toggle switch* un comportement bi-stable et l'auto-régulation permet de diminuer l'effet du bruit issu de l'expression des gènes sur la production de la protéine.

La recherche de boucles positives dans le réseau de régulation chez les bactéries est donc pertinente pour identifier les possibles circuits décisionnels associés aux différents modes de vie. En revanche, elle ne permet pas de comprendre le rôle de la régulation sans la projeter dans son contexte fonctionnel (comme GlnR/TnrA). De plus, il existe peu de cycles dans les régulations. Par contre, le nombre de boucles augmenterait sensiblement si d'autres types de signaux, comme les effecteurs métaboliques étaient intégrés dans le graphe. Si on reprend l'exemple de la FFL de la figure 1.4 où le gène G agit aussi sur S_2, aucune boucle n'est présente si on considère uniquement les facteurs de transcription. Au contraire, si on inclut l'effecteur S_2 dans le graphe sous forme de nœud, on obtient alors la boucle représentée sur la figure 1.9 dont le signe dépend de l'effet de FT_2 sur S_2 et de S_2 sur FT_2. Étant donné que 88% des régulateurs impliquées dans les voies métaboliques

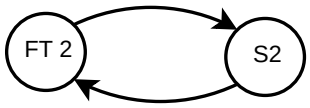


FIG. 1.9 – Existence d'une boucle dans la FFL si les liens métaboliques sont intégrés

répondent à un signal provenant du métabolisme (ion, métabolite, ou protéine impliquée dans le métabolisme) [75], le nombre de boucles augmente certainement si les effecteurs sont intégrés.

1.2.4 Conséquences

Ces différentes approches montrent que le réseau de régulation génétique comporte des structures statistiquement significatives [4, 12]. Cependant, on ne peut pas pour autant conclure à l'existence de sous-systèmes du fait de l'incomplétude du réseau de régulation génétique. En effet, certaines entrées du réseau de régulation ont été négligées comme par exemple les effecteurs métaboliques et tous les autres mécanismes de régulation génétique comme les *riboswitches* par exemple (voir Annexe A.3). De plus, les motifs obtenus par ces méthodes ne permettent pas d'expliquer le rôle et la fonction de la régulation dans la physiologie de la bactérie.
Au regard de cette première étude bibliographique, nous avons identifié la boucle de rétroaction de la figure 1.4 comme un premier élément important pour le fonctionnement des régulations et qui a souvent été négligée dans la littérature. Le couplage entre le réseau génétique et le réseau métabolique doit être intégré, et ce à deux niveaux :

- impact du réseau génétique sur le réseau métabolique à travers la synthèse des enzymes ;

– impact du réseau métabolique sur le réseau génétique via les effecteurs métaboliques des mécanismes de régulation génétique (c'est-à-dire les facteurs de transcription et tous les autres).

De plus, la plupart des effecteurs métaboliques identifiés expérimentalement sont associés au contrôle des voies métaboliques. Le réseau métabolique est par conséquent le sous-système cellulaire idéal pour comprendre le fonctionnement et la coordination des régulations.

1.3 Structure du réseau métabolique

L'organisation des voies métaboliques (et leur contrôle enzymatique) a été également particulièrement étudiée dans la littérature [103, 199]. Là encore le réseau métabolique peut être représenté par un graphe, dans lequel des motifs statistiquement significatifs peuvent être identifiés. Toutefois, la recherche de motifs dans le graphe associé au réseau métabolique a été généralement menée en le découplant du réseau génétique. Là encore l'impact du métabolisme sur la réponse du réseau génétique n'a pas été intégré, limitant ainsi l'intérêt de cette approche pour identifier une structure de contrôle pour les voies métaboliques. Toutefois, la recherche de motifs dans le réseau métabolique seul a l'avantage d'identifier de manière systématique :

– les métabolites fortement connectés (métabolites «hub») ayant potentiellement un fort impact sur le fonctionnement du système global, et donc un impact sur la régulation des voies métaboliques ;
– les «modes élémentaires» définis comme l'ensemble des chemins possibles pour aller d'un point A à un point B et permettant par combinaison linéaire de recouvrir toutes les voies métaboliques.

1.3.1 Identification de métabolites «hubs»

L'organisation topologique du réseau métabolique a été généralement abordée en explorant la connectivité des métabolites dans les réactions chimiques [103, 16, 145]. Le réseau métabolique est alors considéré comme un graphe où les nœuds correspondent aux métabolites. Deux métabolites sont liés par une arête s'ils participent à la même réaction enzymatique. La connectivité d'un métabolite correspond au nombre d'arêtes associées. Les métabolites sont ensuite classés en fonction du nombre de connexions [16] ou par le nombre de métabolites produits grâce au métabolite en question [145].

Des études portant sur la connectivité des métabolites pour 43 organismes montrent que la probabilité d'avoir un nœud du réseau métabolique fortement connecté suit une loi en puissance, caractéristique d'une architecture de type *scale-free* typique d'une organisation en petits mondes [103, 235] (voir la figure 1.10 gauche), et modulaire hiérarchique [175] (voir la figure 1.10 droite). Les modules topologiques obtenus recouvrent les fonctions métaboliques principales. Ainsi peu de métabolites sont fortement connectés. Plus récemment, ces métabolites fortement connectés, appelés métabolites «hubs», ont été identifiés pour *E. coli* [235, 16, 145]. Il s'agit des métabolites riches en énergie et donc nécessaires

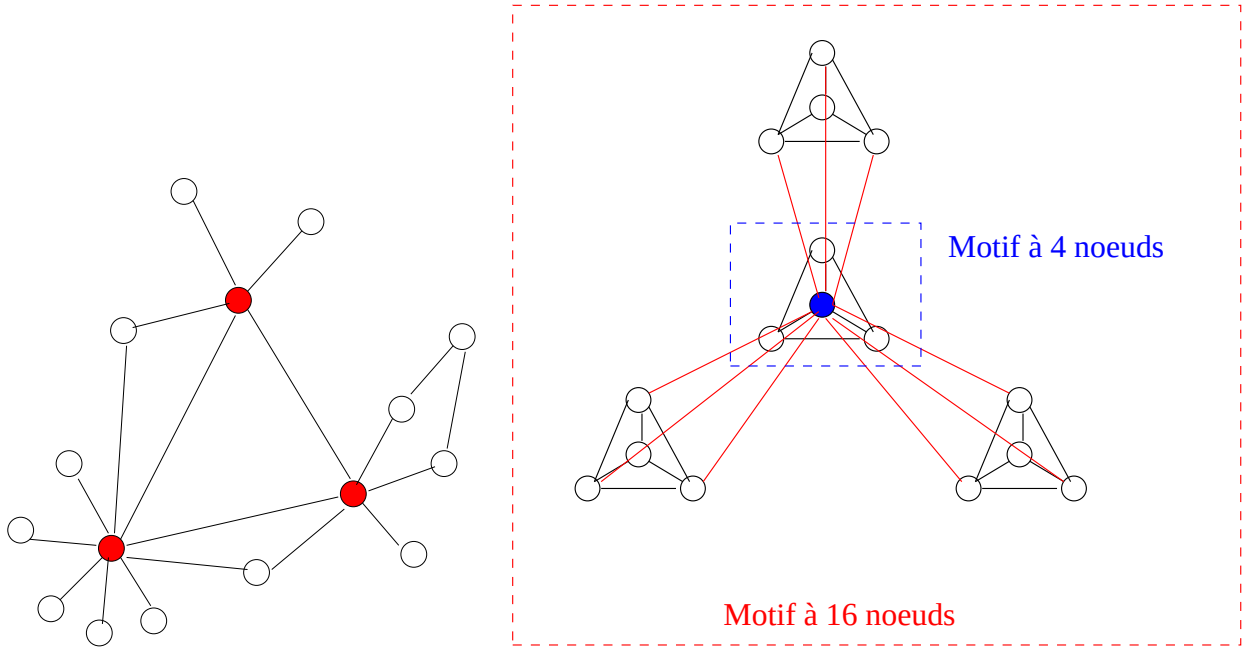


FIG. 1.10 – Réseau de type *scale-free* (gauche) et Hiérarchie modulaire d'un réseau *scale-free* (droite)

à de nombreuses réactions chimiques [16] : H^+, H_2O, ATP, ADP, P_i, PP_i, CO_2, NAD^+, $NADP^+$, NADH, NADPH, CoA, AMP, glutamate, NH_4. Ces métabolites, à l'exception du glutamate) sont considérés comme des co-facteurs, et sont souvent couplés et régulés à travers d'autres facteurs physiologiques comme la gestion du pH, la force proton motrice ou l'état redox de la cellule, ou avec d'autres processus biologiques intracellulaires. Afin d'identifier les métabolites centraux impliqués uniquement dans le métabolisme, certaines études ont exclu ces métabolites de leur analyse et obtiennent alors comme métabolites hubs le glutamate, le pyruvate, la coenzyme-A, le 2-oxoglutarate, la glutamine et l'aspartate principalement [235]. On notera également que le GTP ne fait pas partie des métabolites fortement connectés au sein du réseau métabolique. En revanche, le GTP est utilisé comme source d'énergie pour des processus comme la traduction des protéines (voir la figure 2 page 7).

Intérêt des métabolites hubs ? Selon ces auteurs, les métabolites hubs jouent un rôle central dans le fonctionnement du réseau métabolique car la variation de l'un entre eux affecterait plusieurs motifs et donc un réajustement global de la fonction métabolique. Au contraire, la variation de métabolites appartenant à un seul motif n'affecterait probablement que celui-ci. De façon remarquable, nous montrerons dans le prochain chapitre de cette thèse qu'il n'en est rien. Ces métabolites fortement connectés interviennent généralement dans des étapes intermédiaires des voies métaboliques. Nous montrerons qu'ils n'ont alors en général aucun impact sur le régime d'équilibre.
Ces auteurs ont également souligné l'intérêt d'une structure topologique «*scale-free*»

vis-à-vis des perturbations. Une décomposition en sous-systèmes permettrait de limiter l'impact de perturbations sur la cellule. Si une perturbation affecte un motif en particulier, la perturbation resterait confinée à l'intérieur de celui-ci et engendrerait peu d'impacts sur les autres, et ainsi un réajustement minime à l'échelle de la cellule [216]. Les processus cellulaires ayant une structure de type «*scale-free*», ou *scale-free* hiérarchique résisteraient mieux aux perturbations aléatoires [27, 12] car la probabilité pour la perturbation d'affecter un nœud «hub» est faible [103, 175]. Cependant, cet argument topologique ne tient pas face à la réalité. Prenons l'exemple de la souche *trp-* de *B. subtilis* présentant une mutation délétère dans le gène *trpC*. Cette souche ne peut pas pousser en absence de tryptophane dans le milieu extérieur. La mutation a lieu typiquement au niveau d'un nœud faiblement connecté du réseau métabolique, et pourtant cette mutation est létale en absence de tryptophane.

Si cette analyse identifie le rôle pléiotrope de certains métabolites, elle ne révèle en aucun cas le fonctionnement des voies métaboliques dans la cellule, ni ses coordinations avec les autres processus cellulaires. En revanche, l'approche dite des modes élémentaires identifie de manière systématique tous les chemins possibles dans les voies métaboliques pour aller d'un point A à un point B, intégrant la notion de lien fonctionnel entre les composés impliqués [197, 198].

1.3.2 Les modes élémentaires

L'intérêt d'une telle approche est de révéler pour un métabolite d'intérêt tous les chemins possibles de production et de consommation. Si les chemins sont intuitivement identifiables pour les métabolites faiblement connectés, ils sont en revanche plus interconnectés pour les métabolites hubs. Chez différents organismes, les voies de synthèse des principaux co-facteurs comme l'ATP ou le NADPH sont peu nombreuses. La recherche des modes élémentaires pour ces co-facteurs permet d'identifier les voies possibles de synthèse et donc de révéler potentiellement des stratégies de coordination fine pour produire ces métabolites d'intérêt. Les modes élémentaires d'un réseau métabolique révèlent également la fragilité de certaines voies vis-à-vis des mutations génétiques [215]. Malheureusement, le calcul des modes élémentaires ne se résoud pas en temps polynômial et ne peut s'appliquer qu'à des sous-parties du réseau métabolique [199]. Par ailleurs, rien n'indique qu'un mode élémentaire soit effectivement fonctionnel. En définissant les chemins extrêmes comme les voies topologiques minimales permettant d'obtenir tous les modes élémentaires par combinaison linéaire [190, 164], la prise en compte des régulations génétiques réduit le nombre de chemins extrêmes de 67.5% à 97.5% [41]. Les régulations génétiques sont ainsi indispensables pour comprendre le fonctionnement du réseau métabolique et ses coordinations avec les processus cellulaires.

Au-delà du rôle central du couplage entre le réseau génétique et le réseau métabolique dans le fonctionnement des voies métaboliques, cette seconde étude bibliographique montre aussi qu'il existe une structure dans le réseau métabolique avec des métabolites fortement connectés et faiblement connectés. Les biologistes auraient ainsi pu choisir de

représenter le réseau métabolique sous forme d'arbre tels ceux représentés sur la figure 1.10. Or ce n'est pas le cas. Ils ont choisi de représenter le réseau métabolique suivant les chemins de production et de dégradation des précurseurs métaboliques utilisés par les autres processus cellulaires comme les acides aminés, les nucléotides, etc. Le réseau métabolique est ainsi structuré selon la fonction des voies métaboliques, ce qui peut refléter les objectifs de contrôle de la voie. Par exemple, la voie de synthèse d'un acide aminé pourrait être régulée de manière à toujours garantir la disponibilité de cet acide aminé. Dans la littérature, cette notion d'objectif de contrôle dans les régulations n'est généralement pas abordée explicitement, alors qu'en Automatique elle est essentielle à la synthèse de tout régulateur. Néanmoins, cette notion existe implicitement dans la représentation même du réseau métabolique.

1.4 Les notions élémentaires d'Automatique dans les systèmes biologiques

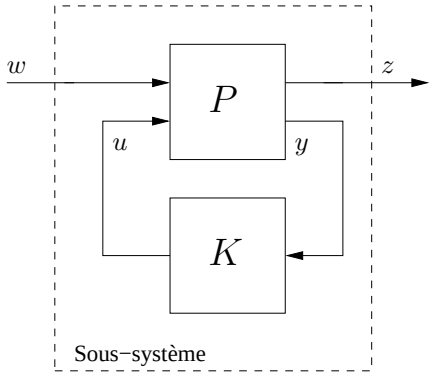


FIG. 1.11 – Un exemple de sous-système défini classiquement en Automatique

On résumera les différents travaux mentionnés dans les sections précédentes autour des deux points suivants :

- il existe des régularités statistiquement sur-représentées dans le réseau de régulation génétique et dans le réseau métabolique ;
- les concepts élémentaires de l'Automatique comme les sous-systèmes (ou modules), le «feedback» et le «feedforward» sont mentionnés [4, 12] mais sont mal définis du point de vue entrée/sortie de l'Automatique [186] comme nous l'avons montré dans le cas de la FFL de la figure 1.9 page 21.

La question consiste maintenant à savoir si ces régularités correspondent à des sous-systèmes opérationnels qui vont nous permettrent de mieux comprendre le fonctionnement global du système. La réponse à cette question est malheureusement négative, et ce pour plusieurs raisons.

Tout d'abord, l'objet considéré dans les différentes études est incomplet, notamment à cause du découplage artificiel du réseau de régulation et du réseau métabolique. Comme

nous l'avons montré, il est indispensable d'intégrer tous les niveaux de contrôle, et en particulier les effecteurs des facteurs de transcription. La réduction du réseau de régulation génétique aux facteurs de transcription et au facteurs sigma dans les différentes études se justifiait par l'existence de technologies bien adaptées pour observer l'impact de ces régulateurs, mais aussi tout simplement parce que le rôle et l'impact des autres types de régulations n'étaient pas connus. Les régulations par les petits ARNms ou les ARN non codant par exemple viennent tout juste d'être découvertes grâce au progrès technologique en parallèle. Il va de soi que les biologistes découvriront dans le futur d'autres mécanismes de régulation complexes et élégants. Les sous-systèmes recherchés doivent pouvoir aider à la compréhension des principes de fonctionnement du système global qui, potentiellement, ne seront pas remis en cause à chaque découverte de nouveaux mécanismes de régulation.

Un deuxième point fondamental concernant la construction d'un sous-système en Automatique relève de son utilité dans le contexte du système global. Il doit servir à quelque chose. Cette vision finaliste des systèmes correspond à la vision systémique propre à l'Automatique qui consiste à isoler des sous-blocs possédant des bonnes propriétés du point de vue entrées-sorties pour ensuite les interconnecter. En ce sens, la décomposition en sous-systèmes est plus un art qu'une science et reste profondément empirique. Un sous-système n'a d'intérêt que s'il permet de simplifier l'analyse du système global. Généralement en Automatique, un sous-système est choisi d'après ses bonnes propriétés locales du point de vue entrées/sorties. Par exemple, considérons le système linéaire P bouclé par un correcteur K de la figure 1.11 où le régulateur a été construit de manière à garantir des propriétés de stabilité et de performance pour le système en boucle fermée vis-à-vis d'un cahier des charges donné. Le principe de la boucle de rétroaction consiste à agir sur le système à travers les signaux u sur la base de l'information disponible du système, c'est-à-dire sur la base des signaux mesurés y. La boucle de rétroaction permet de désensibiliser le système bouclé vis-à-vis de perturbations non mesurées à partir des signaux mesurés. En ce sens, le sous-système défini du point de vue entrées/sorties par le couple (w,z) possède de bonnes propriétés locales. A première vue, pour les systèmes biologiques, nous pouvons interpréter les signaux w, z, u et y comme suit. Les signaux w et z sont les caractéristiques de l'utilité du sous-système dans le contexte du système global. La boucle de régulation des métabolites sur le réseau génétique correspond aux signaux mesurés y alors que l'activation (ou la répression) de la synthèse des protéines est le signal de commande u. Selon cette représentation, les différents mécanismes de régulations intervenant à différents niveaux dans la cellule ne sont «que» la réalisation et l'implémentation du correcteur. Les principes de fonctionnement sous-jacents du sous-système sont alors bien indépendants de la réalisation biologique de la régulation.
Dans le chapitre suivant, nous allons chercher à identifier des sous-systèmes dans le réseau métabolique d'après leurs bonnes propriétés locales. Il reste à savoir sur quelles propriétés seront basées l'identification des sous-systèmes. Comme nous le verrons, le système que nous allons manipuler est composé de centaines de variable d'états régies par des dynamiques fortement non linéaires. Rechercher des sous-systèmes sur la base de leur stabilité n'est actuellement pas envisageable du fait qu'il n'existe pas d'outils adap-

tés applicables à des systèmes non linéaires d'une telle complexité de part leur grande dimension et leur forte interconnexion. Cette question reste d'ailleurs largement ouverte depuis les années 70 et ce malgré les nombreux résultats obtenus sur des linéarisations de sous-systèmes autour d'un régime d'équilibre [227, 9, 6, 7]. Nous chercherons ainsi à identifier des sous-systèmes sur la base :

- des bonnes propriétés biologiques associées ;
- de la finalité du sous-système (à quoi il sert) ;
- de l'existence d'un régime permanent unique pour un sous-système ;
- de la capacité à réduire la complexité du système global en considérant ces sous-systèmes en particulier.

Compte tenu de ces caractéristiques recherchées, la recherche de la structure du réseau métabolique (intégrant les régulations) ne peut être réalisée que manuellement et nécessite pour cela de disposer d'un réseau métabolique dont les différents niveaux de contrôle ont été bien caractérisés expérimentalement par les biologistes. Au début de cette thèse, un tel objet d'études n'existait pas et ce quel que soit l'organisme considéré. La première tâche de notre analyse, présentée également dans le chapitre suivant, a donc consisté à reconstruire le réseau de régulation des voies métaboliques intégrant tous les niveaux de contrôle pour un organisme spécifique, la bactérie *Bacillus subtilis*.

Chapitre 2

Structure modulaire du réseau métabolique des bactéries

L'objectif de ce chapitre consiste à répondre à la question suivante : peut-on identifier des sous-systèmes et plus largement identifier une quelconque structure dans le contrôle du réseau métabolique ?
Cette question est très ambitieuse car, au début de cette thèse, d'une part nous ne savions pas si les outils d'Automatique pourraient être utilisés sur un système aussi complexe et d'autre part nous ne disposions pas d'une vision claire de notre objet d'étude. Notre démarche s'est donc effectuée en deux temps.
Tout d'abord, il a fallu clarifier cet objet d'étude en se focalisant sur le réseau métabolique d'un organisme spécifique qui a été bien caractérisé expérimentalement par les biologistes. Il s'agit en effet de disposer d'un réseau métabolique possédant un éventail suffisamment exhaustif des régulations pour identifier sa structure et ses règles de fonctionnement si elles existent. Nous avons ainsi «reconstruit» le réseau métabolique ainsi que son contrôle pour la bactérie modèle de la famille des Gram + *Bacillus subtilis*. Ce travail est déjà en soit considérable car si les réactions chimiques composant le réseau métabolique sont accessibles dans des bases de données, le réseau de régulation a été reconstruit manuellement à travers la lecture de nombreux articles de Biologie. Ceci étant, il convient de veiller alors à la généricité de nos résultats, car rien n'indique que les règles de fonctionnement que nous obtiendrons puissent s'appliquer pour un autre organisme comme par exemple *Escherichia coli*, la bactérie modèle de la famille des Gram -. Le réseau métabolique d'*Escherichia coli* et ses régulations sont certainement les systèmes biologiques les mieux caractérisés du point de vue expérimental (après l'opéron lactose) [107]. Nous allons nous centrer sur *Bacillus subtilis* tout en veillant à confronter nos résultats avec le réseau métabolique d'*Escherichia coli* afin de vérifier la généricité de nos résultats. Nos résultats n'en seront alors que renforcés surtout si l'on considère que ces deux bactéries ont divergé il y a plus de trois milliards d'années [14].
Dans un deuxième temps, nous nous sommes attachés à mettre à jour des sous-systèmes spécifiques dans le réseau métabolique ainsi reconstruit. Ce travail s'est révélé difficile car leur détermination relève plus d'un art que d'une science. Notre objectif est de faire émerger une ou plusieurs structures possédant des propriétés spécifiques, qui restent faci-

lement manipulables et qui puissent être interconnectées. Cette décomposition doit nous permettre de simplifier l'objet. Sous ces contraintes, nous avons défini deux structures de contrôle particulières (cf. la section 2.2) possédant des propriétés singulières en régime permanent du point de vue entrées/sorties (voir les sections 2.3 et 2.4.1). Cette identification de deux formes de sous-systèmes est à la base de la construction d'un cadre formel théorique d'étude des réseaux métaboliques bactériens en régime permanent dans leur globalité (cf. section 2.4) en considérant des configurations biologiques réalistes comme

- l'organisation des gènes en opérons,
- l'irréversibilité/réversibilité des enzymes,
- la présence de co-facteurs, co-métabolites et d'isoenzymes,
- la présence de point de branchements dans les voies métaboliques.

Le modèle complet du réseau métabolique de *Bacillus subtilis*, la définition des deux structures de contrôle et leurs propriétés en régime permanent ainsi que le cadre formel d'études des voies métaboliques sont les contributions majeures de ce chapitre de thèse. Ce modèle, la définition et l'interprétation biologique des deux structures de contrôle ainsi que les premiers résultats théoriques sur lesquels s'est basée l'élaboration du cadre formel d'études des voies métaboliques ont également donné lieu à une publication [75].

2.1 Un modèle de connaissance exhaustif

2.1.1 Les modèles globaux du métabolisme existant

Ces dernières années, les reconstructions de réseaux métaboliques ont explosé grâce notamment au séquençage de nombreuses espèces, aux plateformes d'annotation des génomes de plus en plus performantes [30], et à certains outils de reconstruction automatique comparant l'annotation du génome avec les voies métaboliques existantes dans le vivant [106]. Une validation manuelle du modèle reste néanmoins nécessaire pour prévenir les erreurs intrinsèques dûes à l'annotation automatique du génome.

Ces modèles, donnés sous forme de fichiers ou de bases de données, décrivent généralement l'ensemble des réactions chimiques du réseau métabolique ainsi que les gènes associés à ces réactions. Développés tout d'abord sur les micro-organismes comme *Helicobacter pylori* [189], *Haemophilus influenzae* [191], *Lactococcus lactis* [163], *Escherichia coli* [176], *Saccharomyces cerevisiae* [68], des modèles sont maintenant obtenus pour les organismes supérieurs : *Corynebacterium glutamicum* [113], *Aspergillus oryzae* [234], l'*Homo Sapiens* [49] à titre d'exemples.

En revanche, peu de modèles intègrent en plus les régulations (au sens large). Seules les régulations transcriptionnelles par facteur de transcription ont été intégrées pour certains organismes modèles : *Saccharomyces cerevisiae* [94, 95] et *Escherichia coli* [182]. Cependant, les mécanismes de régulation par terminaison précoce de la transcription ou de la traduction sont généralement occultés. De plus, aucun modèle n'intègre explicitement les effecteurs métaboliques des facteurs de transcription à l'exception de la version la plus récente de la base de données EcoCyc centrée sur *Escherichia coli* [182]. Les métabolites modulent en effet l'activité des facteurs de transcription, et sont par conséquent des acteurs clés dans les régulations génétiques. Enfin, les régulations traductionnelles et

post-traductionnelles ne sont pas prises en compte. Il est cependant nécessaire d'intégrer tous les acteurs impliqués dans la régulation de la fonction métabolique pour comprendre la coordination du réseau génétique et du réseau métabolique en réponse à un changement de conditions environnementales [117, 151] ou l'organisation générale du réseau de régulation [12, 5].

Comme nous l'avons souligné en introduction de ce chapitre, notre objectif est de reconstruire le réseau de régulation génétique et métabolique de la bactérie modèle *Bacillus subtilis* intégrant tous les niveaux de régulation : transcriptionnelles, traductionnelles, post-traductionnelles (phosphorylation ou protéolyse par exemple) ainsi que la modulation de l'activité des enzymes. Le réseau de régulation des voies métaboliques pour cet organisme a en effet été relativement bien identifié expérimentalement par les biologistes [212]. A ce stade, deux approches possibles s'offraient à nous. La première consiste à reconstruire le réseau manuellement, contrairement à la deuxième approche qui se base sur la reconstruction automatique du réseau métabolique et de son réseau de régulation par des outils bioinformatiques. La reconstruction automatique du réseau de régulation génétique à partir de la séquence ADN est un problème difficile. Les algorithmes de détection de motifs permettraient de prédire des sites de fixation des facteurs de transcription à l'ADN [157], mais la plupart des motifs de fixation sont inconnus. Par ailleurs, ces prédictions nécessitent ensuite d'être validées par des expériences spécifiques. La reconstruction du réseau de régulation et des voies métaboliques a donc été effectuée manuellement à travers la lecture de centaines d'articles de Biologie (voir l'ensemble des références données dans le modèle) afin d'identifier précisément tous les acteurs en jeu [93]. Chaque régulation et chaque voie métabolique ont ainsi été manuellement validées. En parallèle, nous nous sommes appuyés sur différentes bases de données référençant pour *Bacillus subtilis* des informations pertinentes pour notre modèle :

- SubtiList qui décrit l'annotation de l'ensemble des gènes de *Bacillus subtilis* [150] ;
- KEGG [159] et PAREO (une version relationnelle de KEGG développée dans l'unité MIG) référençant l'ensemble des voies métaboliques existant dans la nature et qui projette l'annotation des gènes de n'importe quel organisme (et par conséquent de *Bacillus subtilis* en particulier) sur ces voies métaboliques ;
- DBTBS qui contient les facteurs de transcription et les facteurs sigma ainsi que les régulons associés de *Bacillus subtilis* [139] ;
- BRENDA qui regroupe les caractéristiques biochimiques des enzymes disponibles dans la littérature pour n'importe quel organisme [193].

Sur la base de ces éléments, nous avons construit un modèle de connaissance du réseau métabolique et de sa régulation décrivant les liens existants entre ses composants. Ce modèle constitue une contribution majeure de cette thèse et est à la base des développements théoriques obtenus dans la suite de ce chapitre.

2.1.2 Un modèle pour *B. subtilis* intégrant les régulations

Afin d'étudier l'organisation fonctionnelle des régulations des voies métaboliques, nous nous sommes focalisés sur les voies métaboliques principales de *B. subtilis* pour les-

quelles l'information sur les régulations était disponible. Les voies métaboliques peuvent être considérées comme «mineures» ou «majeures» en fonction du flux présent dans ces voies. Les voies métaboliques majeures correspondent aux voies de production d'énergie (ATP, NADH, NADPH) et des précurseurs métaboliques pour les protéines, l'ARN, l'ADN, la paroi et la membrane. Les voies mineures regroupent les voies de synthèse des cofacteurs et des métabolites secondaires (vitamines, co-enzymes, antibiotiques, ...). Les voies métaboliques mineures sont généralement linéaires et peu interconnectées, alors que les voies majeures sont au contraire fortement interconnectées entre elles, impliquant des mécanismes de régulation fins pour assurer leur coordination. Par conséquent, nous avons considéré pour cette étude uniquement les voies métaboliques suivantes, dont le fonctionnement est rappelé en annexe A.2 : le métabolisme du carbone central, la respiration aérobie et anaérobie, la fermentation, toutes les voies de synthèse, dégradation et conversion des acides aminés, des nucléotides, des deoxy-nucléotides, des phospholipides, du peptidoglycane, des acides teichoïques et teichuroniques.
Le modèle associé est un modèle de connaissance, décrivant les réactions chimiques, les régulations transcriptionnelles, traductionnelles, post-traductionnelles et enzymatiques connues. Contrairement aux représentations classiques des réseaux de régulation génétique incluant uniquement les facteurs de transcription [94], nous avons intégré les régulations par terminaison précoce de la transcription ainsi que les régulations traductionnelles et post-traductionnelles suivantes :

- la séquestration de la séquence d'accrochage du ribosome (RBS) ;
- les petits ARNs non-codant ;
- des interactions protéines-protéines ;
- la phosphorylation des protéines ;
- protéolyse.

Nous avons aussi systématiquement inclus la description des complexes enzymatiques en tant que somme de plusieurs gènes, l'organisation des gènes en opéron ainsi que le nombre de promoteurs pour chaque gène. Les conditions de transcription, traduction de chaque enzyme sont ainsi déduites et représentées sous forme booléenne. Le modèle final que nous avons construit intègre 534 gènes, 563 réactions, 65 facteurs de transcription (FT), 9 facteurs sigma et 21 autres mécanismes de régulation génétique (transcriptionnel ou traductionnel)(cf. le tableau 2.1), soit en totalité 95 mécanismes de régulation différents. Ce modèle correspond au fichier *modele_Bsub.xls* donné en annexe B.1. Une version antérieure de ce modèle a été publié sans les voies de synthèse des acides teichoïques et teichuroniques, du peptidoglycane et des phospholipides pour lesquelles quasiment aucune information sur les régulations n'est disponible [75].

2.2 Topologie du réseau de régulation des voies métaboliques de *B. subtilis*

Le réseau de régulation (multi-niveaux) des voies métaboliques ayant été reconstruit d'après la littérature, nous pouvons maintenant caractériser sa structure à travers une approche d'analyse des systèmes propre à l'Automatique, c'est-à-dire la recherche des ob-

Gènes	534	organisés en 186 opérons
Métabolites	456	
Réactions	563	voies métaboliques, complexes
Régulations enzymatiques	79	par cofacteurs, ions et métabolites
Régulations génétiques	65 TFs 9 facteurs sigma 21 autres	avec les effecteurs métaboliques distinction entre activation et modulation
Fonctions intégrées	10	métabolisme du carbone central, fermentation, phosphorylation oxydative, transport et dégradation des carbohydrates, synthèse/dégradation des acides aminés, synthèse/dégradation des (deoxy)-nucléotides, synthèse/dégradation des acides gras, synthèse/dégradation des phospholipides, synthèse du peptidoglycane, synthèse des acides teichoïques/teichuroniques,

TAB. 2.1 – Intégration de l'information. Les 21 autres mécanismes de régulation incluent RtpA, S-box, L-box, A-box, G-box, Plead, SR1 et 14 T-boxes.

jectifs et le rôle de chaque régulation impliquée. Notre analyse de la topologie du réseau de régulation s'effectuera en trois temps. Tout d'abord l'impact du réseau métabolique sur le réseau génétique à travers les effecteurs métaboliques va être évalué de manière quantitative. Puis, nous nous intéresserons à l'organisation hiérarchique du réseau de régulation du métabolisme en utilisant la classification proposée dans [132] et rappelée dans la section 1.2.2 page 17 du chapitre précédent. Enfin, en combinant les deux approches, nous en déduirons la présence de deux grands types de structure de contrôle «bas niveau» dans les voies métaboliques, coordonnés par des régulateurs «haut niveau».

2.2.1 Le rôle clé des métabolites

Régulateurs génétiques	FTs	Facteurs sigma	Autres	Total
Effecteurs métaboliques	40	0	19	59
Effecteurs liés au métabolisme (ion, enzyme)	8	0	0	8
Sans effecteur	2	9	2	13
Effecteurs métaboliques potentiels	8	0	0	8
Effecteurs inconnus	7	0	0	7
Total	65	9	21	95

TAB. 2.2 – Classification des régulations génétiques en fonction de la nature de leur effecteur métabolique

Parmi les 79 régulations génétiques dont l'effecteur est connu, 70% (59) de ces effecteurs correspondent à un métabolite (cf. le tableau 2.2). De plus, 8 autres régulations répondent à la concentration d'un ion ou sont phosphorylées par une sous-unité des transporteurs de sucre. Ces effecteurs sont donc également fortement couplés au métabolisme. Treize régulateurs n'ont pas d'effecteur : AbrB, RtpA, YtlI et SR1, qui sont tous sous le contrôle direct d'autres régulateurs génétiques (respectivement Spo0A ; T-box et Plead ; CymR et Spx ; CcpN et CcpA), et 9 facteurs sigma. Le(s) effecteur(s) de 15 régulateurs génétiques est (sont) encore inconnu(s) à ce jour. Cependant, on peut supposer d'après la littérature qu'au moins 8 d'entre eux sont sensibles à la concentration d'un métabolite puisque soit ils appartiennent à des systèmes à deux composantes (YycF, ResD), soit ils sont impliqués dans le catabolisme de carbohydrates (YtrA, FruR, RbsR), soit encore ils régulent la synthèse d'enzymes spécifiques en réponse à la disponibilité de certains nutriments (CysL, CcpN, GlnR). Les 7 régulateurs restants sont ArfM, Spo0A, SpoVT, SpoIIID, GerE, CcpB et Spx.
Par conséquent, le couplage entre les voies métaboliques et les régulations génétiques est très important : au moins 13% des 456 métabolites considérés sont impliqués dans une régulation génétique et 53% des gènes des voies métaboliques sont sous le contrôle direct d'un ou plusieurs métabolites via une régulation génétique. De plus, les métabolites participent aussi au contrôle de l'activité des enzymes. Plus précisément, chaque voie métabolique du modèle pour laquelle une régulation est connue, est sous le contrôle direct d'un métabolite, soit à travers une régulation enzymatique, soit à travers un effecteur métabolique d'un régulateur génétique.
Comme nous l'avions pressenti au cours du chapitre précédent, nous constatons ici que les métabolites jouent bien un rôle critique dans la régulation générale des voies métaboliques à travers ces deux niveaux de contrôle (génétique et enzymatique).

2.2.2 Organisation du réseau de régulation génétique

Nous avons exploré l'organisation du réseau de régulation génétique de *Bacillus subtilis* en appliquant la représentation hiérarchique de [203, 133, 132], présentée sur la figure 1.5 page 17 du chapitre précédent et rappelée ici sur la figure 2.1. La cascade de régulations induites par chaque régulateur génétique est tout d'abord identifiée. Le nombre de régulations induites définit alors le niveau du régulateur de la manière suivante. Un régulateur dit de niveau 1 n'induit aucun autre régulateur. Un régulateur est dit de niveau n si les autres régulateurs sous son contrôle sont au plus de niveau $n-1$. Chez *E. coli*, seulement 23 régulateurs génétiques sont de niveau 3 ou plus, et 2 régulateurs génétiques appartiennent au niveau le plus élevé, le niveau 5 (voir la figure 1.5 et [132]). Comme nous l'avons souligné dans le chapitre précédent, l'objectif pour nous est d'identifier les régulateurs pouvant impacter différentes voies métaboliques, et donc de jouer un rôle «pléiotrope». L'organisation hiérarchique du réseau de régulation génétique de *B. subtilis* est représentée sur la figure 2.2, et inclut non seulement les facteurs de transcription, mais aussi tous les autres mécanismes de régulation transcriptionnelle et traductionnelle contrairement à l'étude proposée dans [132] qui intègre uniquement les facteurs de transcription et les facteurs sigma : terminaison précoce de la transcription

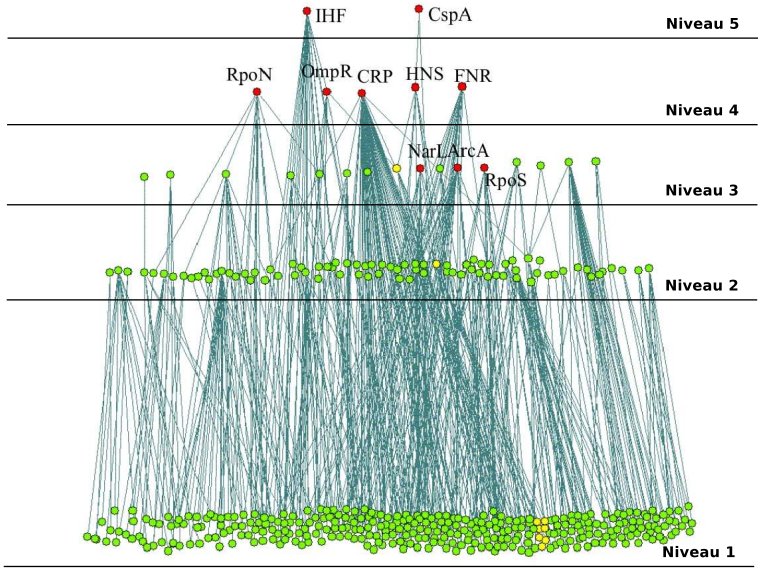


FIG. 2.1 – Cinq niveaux de contrôle dans le réseau de régulation génétique d'*E.coli* (extrait de [132])

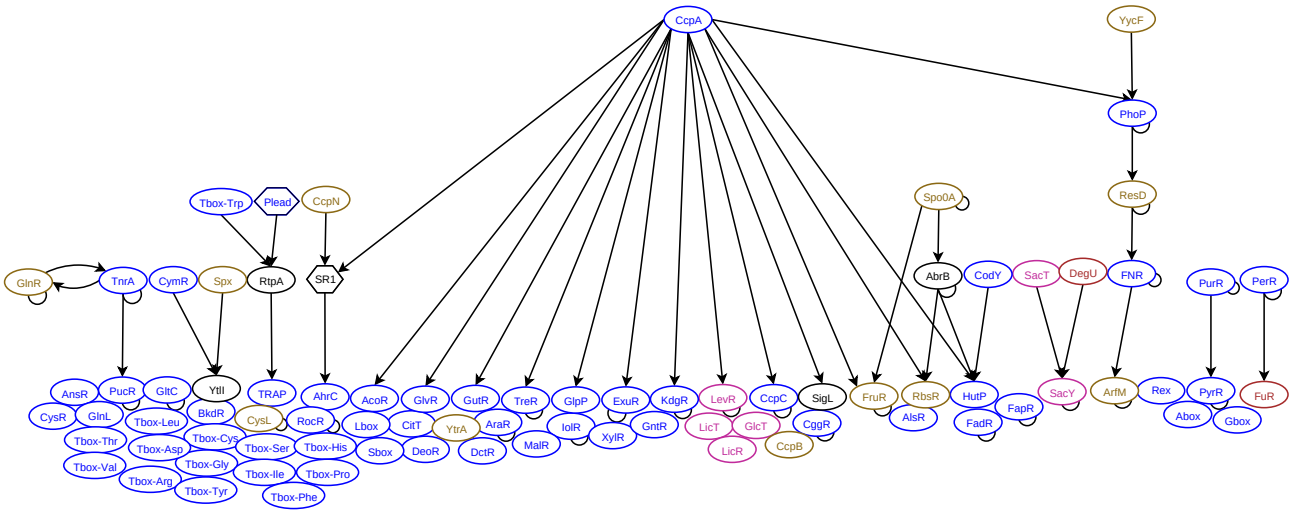


FIG. 2.2 – Une organisation hiérarchique du réseau de régulation génétique du réseau métabolique de *Bacillus subtilis*. Le code couleur correspond au type d'effecteur du régulateur : bleu (métabolite), violet (transporteur de sucre), rouge (ion), noir (rien), marron (inconnu).

(T-box, G-box, A-box, L-box, S-box), prévention de la traduction par de petits ARN non codant (SR1) ou par peptide leader (Plead). En revanche, les facteurs sigma ne sont pas représentés. En considérant uniquement les voies métaboliques majeures de *B. subtilis*, on observe déjà cinq niveaux de régulations.

On constate immédiatement une structure pyramidale, comme celle observée chez *E. coli*, où la majorité des régulateurs se concentrent sur le niveau le plus faible. Plus le niveau est élevé, moins de régulateurs sont présents. Plus précisément, 75% des régulateurs génétiques impliqués dans les voies métaboliques sont de niveau 1. Ils régulent uniquement des gènes directement impliqués dans des voies métaboliques. De plus, 88% de ces régulateurs ont un métabolite, un ion ou une sous-unité des transporteurs de sucre comme effecteur métabolique, ce qui les relie de façon directe au réseau métabolique. Caractérisons maintenant d'un point de vue fonctionnel les régulateurs haut niveau.
CcpA et YycF sont les régulateurs de plus haut niveau (niveau 5). YycF est indirectement impliqué dans la régulation des voies métaboliques fermentaires et respiratoires via la cascade de régulation PhoP, ResD, FNR and ArfM. CcpA est également connecté à cette même cascade de régulation. 17% des régulateurs génétiques sont sous le contrôle direct de CcpA. D'autres régulateurs, comme DegU, AbrB et Spo0A, connus pour être pléiotropes chez *B. subtilis* apparaissent seulement au niveau 2 ou 3. Si d'autres processus biologiques comme la sporulation ou la compétence étaient intégrés au modèle, alors le niveau de ces régulateurs augmenterait certainement. Cependant, DegU et Spo0A modulent uniquement quelques enzymes spécifiques impliquées dans d'autres processus biologiques. Par exemple, l'enzyme CitB est nécessaire au cours de la sporulation [43] et sa synthèse est controlée par CodY et AbrB [109]. Ces régulateurs ont ainsi un faible impact sur la régulation directe des voies métaboliques, suggérant ainsi un découplage partiel entre le contrôle des voies métaboliques et les autres processus biologiques.

La figure 2.2 suggère deux grands niveaux de contrôle des voies métaboliques. Un contrôle «bas niveau» intégrant les régulateurs de niveau 1, et un contrôle plus «haut niveau» intégrant les régulateurs de niveau 2 et plus. Il s'agit maintenant de préciser cette structure lorsque le couplage entre la voie métabolique et les différents régulateurs la contrôlant est intégré.

2.2.3 Deux structures spécifiques de contrôle

Deux structures générales de contrôle des voies métaboliques peuvent être identifiées d'après le modèle et la représentation hiérarchique de la figure 2.2 et sont représentées sur la figure 2.3.
La première, que nous appellerons «structure de contrôle par le produit final» (SCPF), a comme acteur principal le dernier métabolite. Celui-ci réprime la première enzyme à la fois au niveau de son activité et de sa synthèse grâce à un régulateur génétique. La seconde, nommée «structure de contrôle par le produit initial» (SCPI), implique le premier métabolite de la voie. Une concentration croissante du premier métabolite induit la synthèse des enzymes de la voie métabolique via un régulateur génétique.
Typiquement, les régulateurs génétiques impliqués dans ces deux structures de contrôle sont des régulateurs de niveau 1 sur la figure 2.2. Des régulateurs de plus haut niveau peuvent également moduler la synthèse des enzymes d'une voie métabolique (cf la figure 2.4). Toutefois, le signal métabolique (ou autre) modulant l'activité de ces régulateurs

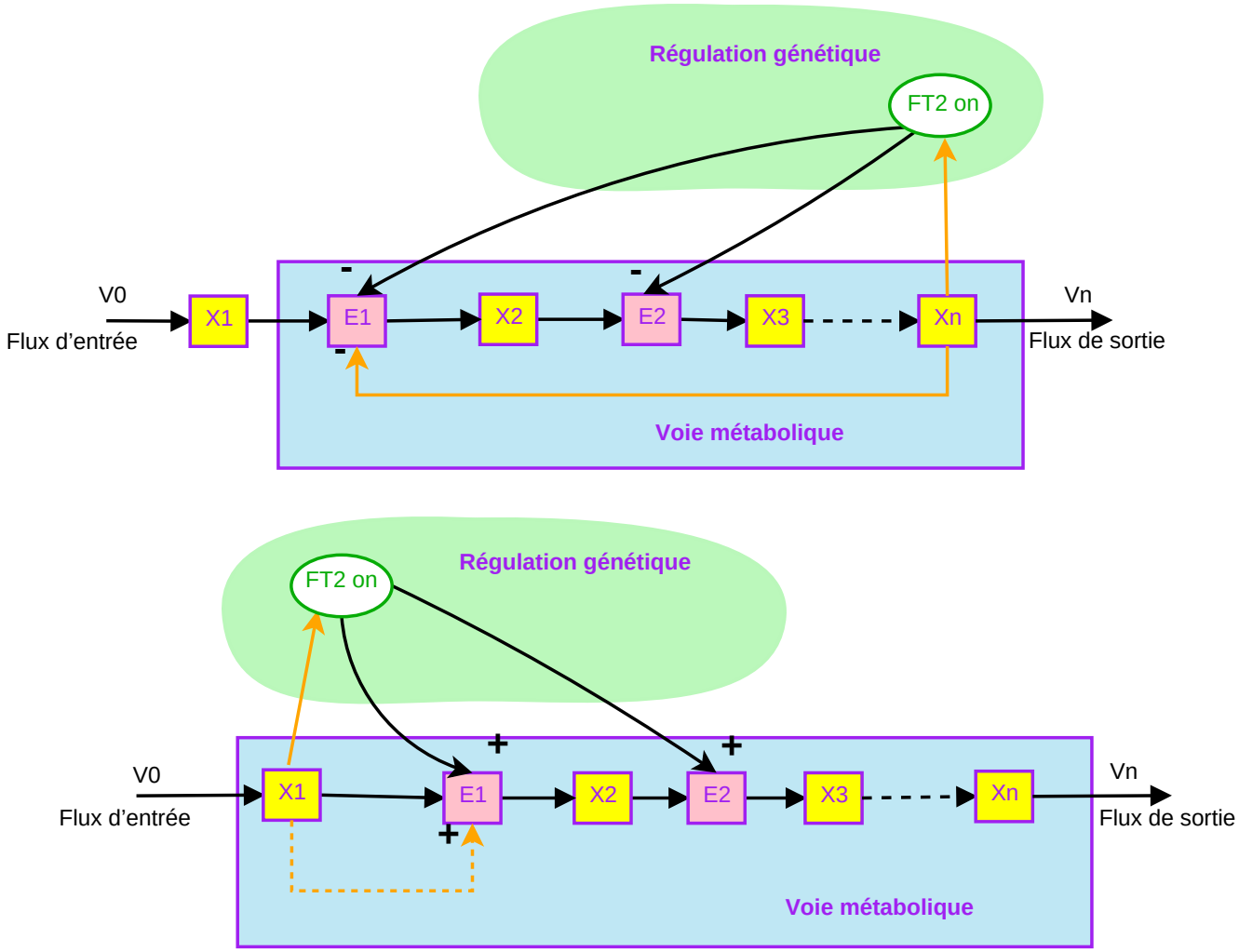


FIG. 2.3 – Deux structures de contrôle : par le produit final, ou par le premier métabolite. Les enzymes (resp. métabolites) sont en rose (resp. jaune), et le régulateur est représenté par une ellipse. Les flèches en orange désignent le contrôle par les métabolites (effecteur du régulateur ou modulation de l'activité de l'enzyme).

haut niveau n'est pas un métabolite intermédiaire de la voie métabolique contrôlée.
Ces deux structures de contrôle nous amènent à définir deux niveaux de contrôle dans les voies métaboliques : un contrôle dit **local** correspondant aux structures de la figure 2.3, et un contrôle dit **global** représenté sur la figure 2.4.

Définition de la régulation locale/globale des voies métaboliques

La régulation locale d'une voie métabolique correspond à tout type de régulation génétique (transcriptionnelle, traductionnelle, post-traductionnelle) qui implique un métabolite de la voie contrôlée. La régulation globale d'une voie métabolique est définie comme l'ensemble des régulations non locales. D'un point de vue opérationnel, la régu-

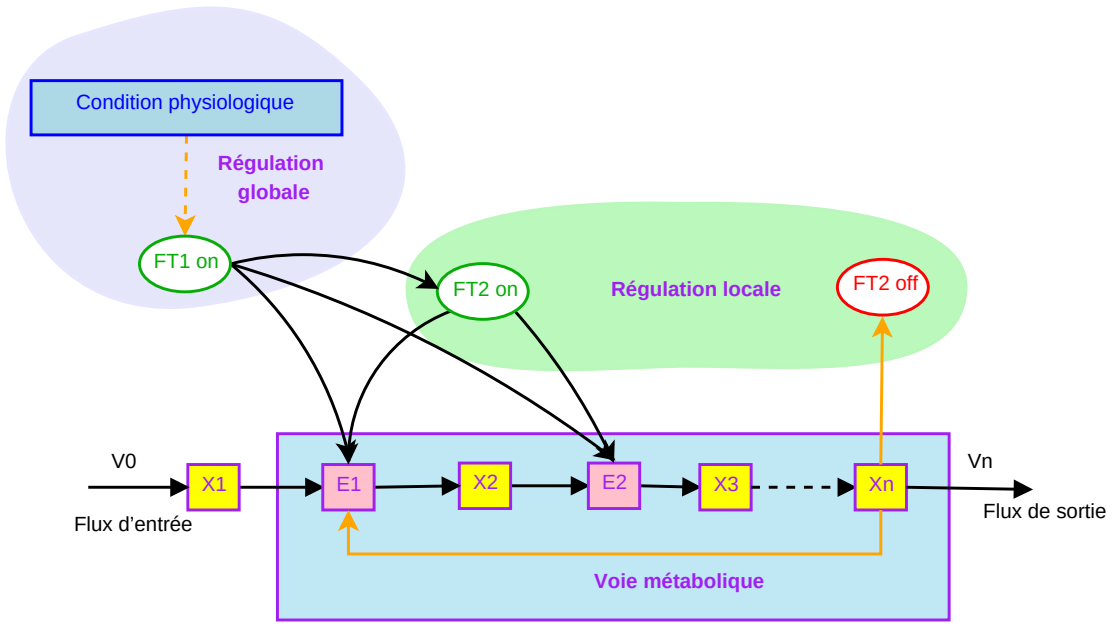


FIG. 2.4 – Deux niveaux de contrôle pour une même voie métabolique

lation locale garantit l'induction ou la répression des enzymes de la voie métabolique en fonction de la concentration d'un métabolite intermédiaire de la voie métabolique. La régulation globale permet de modifier, voire de prendre le pas sur la régulation locale.

Le choix de ces structures comme sous-systèmes élémentaires, même s'il paraît simple à première vue, s'est basé sur leurs bonnes propriétés mathématiques intrinsèques que nous allons présenter dans les sections suivantes. Ces structures de contrôle correspondent à des sous-systèmes d'un point de vue entrée/sortie : ce sont donc des modules et seront traités comme tels dans la suite de cette thèse. De plus, ils permettent de découper le réseau métabolique dans son ensemble en sous-parties fonctionnelles qui correspondent souvent au découpage empirique du réseau métabolique fait par les biologistes.

2.2.4 Vers l'élaboration d'un cadre formel pour l'analyse du réseau métabolique

Nous allons proposer dans la suite de ce chapitre un cadre théorique formel permettant :

- de découper fonctionnellement le réseau métabolique en sous-systèmes autonomes ;
- de donner des conditions d'existence et d'unicité du point de fonctionnement en régime permanent associé au réseau métabolique ;
- de préciser le rôle et l'impact de configurations métaboliques classiques identifiées d'après notre modèle, comme l'organisation des gènes en opéron, la réversibilité des enzymes, la présence d'isoenzymes, de co-métabolites et co-facteurs, ou de points de branchement.

La majorité des études existantes se sont focalisées sur le comportement d'une voie métabolique (ou d'une voie de signalisation) spécifique pour une configuration métabolique donnée [227, 183, 9, 213]. Les résultats obtenus se focalisent essentiellement sur l'obtention de condition de stabilité, ce qui explique le choix d'un modèle simplifié compte tenu de la nonlinéarité forte des équations manipulées. De plus, leur résultats sont rarement interprétés d'un point de vue biologique. Certains travaux ont traité des configurations métaboliques plus réalistes comme [6, 7] sans intégrer toutefois les régulations génétiques. A la différence de ces études, l'originalité de notre approche consiste bien à attaquer la question de l'existence et de l'unicité d'un équilibre structurel pour l'ensemble des voies métaboliques quelle que soit leur configuration et leur régulation génétique et enzymatique. Grâce à la phase de reconstruction du réseau de régulation du métabolisme, nous avons ainsi dégagé quelles étaient les configurations biologiques pertinentes à considérer, à savoir les deux structures de contrôle ainsi que les configurations métaboliques classiques précisées ci-dessus. Tout comme le modèle de connaissance de *Bacillus subtilis*, ce cadre formel constitue une contribution majeure de ce travail de thèse.
La section suivante sera ainsi consacrée à la structure de contrôle par le produit final, que nous étudierons de façon approfondie. Nous allons tout d'abord considérer un cas simplifié de voie métabolique pour présenter les propriétés générales, qui sera étendu pour étudier l'impact des configurations métaboliques classiques. On donnera dans la section 2.4.1 les résultats pour la structure de contrôle par le produit initial, uniquement dans un cas simplifié. Les cas étendus se dérivent facilement à partir des résultats obtenus pour le produit final. Enfin dans les sections 2.4.2 et 2.4.3, nous étudierons les coordinations et connexions entre ces structures de contrôle.

2.3 Structure de contrôle par le produit final (SCPF)

Les résultats mathématiques de cette section ont été obtenus en collaboration avec Fadia Bekkal Brikci (post-doctorante au laboratoire) non seulement dans le cas de ces structures de contrôle, mais aussi dans le cas général du contrôle des voies métaboliques.

2.3.1 Définition du système étudié

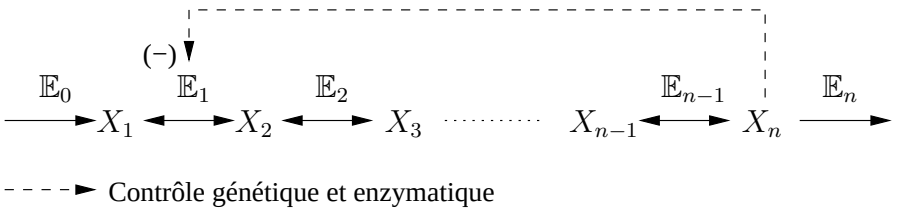


FIG. 2.5 – Voie métabolique contrôlée par le produit final

Considérons la voie métabolique linéaire représentée sur la figure 2.5, composée de n métabolites $(X_1, \ldots, X_n)$ et $n-1$ enzymes $(\mathbb{E}_1, \ldots, \mathbb{E}_{n-1})$, contrôlée par le produit final

à la fois au niveau génétique à travers un facteur de transcription par exemple, et à la fois au niveau métabolique à travers la modulation de l'activité enzymatique de $\mathbb{E}_1$.

a. Rappels sur la cinétique des enzymes

Concrètement, chaque enzyme catalyse une réaction chimique de la façon suivante (en prenant l'exemple de l'enzyme $\mathbb{E}_1$:

$$X_1 + \mathbb{E}_1 \overset{k_1/k_{-1}}{\rightleftharpoons} [X_1\mathbb{E}_1] \overset{k_3/k_{-3}}{\rightleftharpoons} [X_2\mathbb{E}_1] \overset{k_2/k_{-2}}{\rightleftharpoons} X_2 + \mathbb{E}_1$$

où les différents k_i et k_{-i} désignent respectivement les vitesses de réaction dans le sens X_1 vers X_2 (sens *forward*) et dans le sens X_2 vers X_1 (sens *reverse*). Le principe est le suivant pour le sens *forward*. Le métabolite substrat X_1 se fixe sur l'enzyme $\mathbb{E}_1$ et forme ainsi le complexe $[X_1\mathbb{E}_1]$. L'enzyme transforme ensuite le substrat X_1 en produit X_2 et forme alors le complexe $[X_2\mathbb{E}_1]$. Enfin le métabolite produit X_2 est relâché. L'enzyme n'est donc pas transformée au cours de la réaction chimique. Il en va de même pour le sens inverse. Une enzyme est dite réversible lorsqu'elle est capable de catalyser les deux sens de la réaction chimique, contrairement à une enzyme dite irréversible qui ne catalyse qu'un seul sens.
La vitesse de formation du produit d'une enzyme dépend des concentrations des substrats, des produits et de l'enzyme elle-même. Classiquement, la vitesse de la réaction $[X_1\mathbb{E}_1] \rightleftharpoons [X_2\mathbb{E}_1]$ est supposée être très rapide par rapport aux autres réactions, de sorte que nous avons $[X_1\mathbb{E}_1] = [X_2\mathbb{E}_1]$ [233]. On considère finalement la réaction enzymatique suivante :

$$X_1 + \mathbb{E}_1 \overset{k_1/k_{-1}}{\rightleftharpoons} [X_1\mathbb{E}_1] \overset{k_2/k_{-2}}{\rightleftharpoons} X_2 + \mathbb{E}_1$$

On peut alors rappeler les quelques étapes mathématiques pour obtenir les équations de cinétique enzymatique classique afin de mettre en avant clairement certaines hypothèses. Pour le cas où l'enzyme $\mathbb{E}_1$ est irréversible, c'est-à-dire $k_{-2} \approx 0$, on obtient le système d'équations différentielles suivant [233] :

$$\left\{ \begin{array}{rcl} \dot{x}_1(t) & = & -k_1 x_1(t) E_1(t) + k_{-1}[x_1 E_1](t) \\ \dot{x}_2(t) & = & k_2[x_1 E_1](t) \\ \dot{[x_1 E_1]}(t) & = & k_1 x_1(t) E_1(t) - (k_2 + k_{-1})[x_1 E_1](t) \end{array} \right.$$

où x_1, x_2, E_1 et $[x_1 E_1]$ désignent respectivement les concentrations de X_1, X_2, $\mathbb{E}_1$ et $[X_1\mathbb{E}_1]$. De plus, nous avons à tout instant la conservation de la masse pour l'enzyme $\mathbb{E}_1$:

$$[x_1 E_1](t) + E_1(t) = E_T$$

où E_T désigne la concentration totale de l'enzyme $\mathbb{E}_1$ supposée constante. Nous substituons donc le terme $E_1(t)$ par le terme $E_T - [x_1 E_1](t)$ dans la dernière équation du système d'équations différentielles pour obtenir :

$$\dot{[x_1 E_1]}(t) = k_1 x_1(t) E_T - (k_2 + k_{-1} + k_1 x_1(t))[x_1 E_1](t)$$

Lorsque la concentration x_1 du métabolite X_1 est en excès, ce qui est généralement toujours le cas [102], nous avons donc $k_1 x_1(t) >> k_2 + k_{-1} >> 0$. La dynamique de cette équation est très rapide de sorte qu'elle peut être considérée dans un état stationnaire vis-à-vis de l'équation associée à $\dot{x}_1$. Le régime permanent $[x_1\dot{E}_1](t) = 0$ implique que les différentes formes de l'enzyme ne varient plus dans le temps : l'*équilibre biochimique* est atteint. Nous avons alors :

$$[x_1\dot{E}_1](t) = 0 \quad \Longrightarrow \quad \overline{[x_1 E_1]} = \frac{k_1 x_1}{(k_2 + k_{-1} + k_1 x_1)} E_T$$

Ainsi on obtient le système d'équations différentielles suivant pour la réaction enzymatique catalysée par $\mathbb{E}_1$ à l'équilibre biochimique :

$$\begin{cases} \dot{x}_1(t) &= -k_2 \dfrac{x_1(t)}{K_1 + x_1(t)} E_T \\ \dot{x}_2(t) &= k_2 \dfrac{x_1(t)}{K_1 + x_1(t)} E_T \end{cases}$$

avec $K_1 = \dfrac{k_{-1} + k_2}{k_1}$. Ce paramètre correspond à la constante de Michaelis-Menten, bien connue en cinétique enzymatique. Le terme $k_2 \dfrac{x_1(t)}{K_1 + x_1(t)} E_T$ désigne par définition le **flux métabolique** $\nu_E(t)$ à travers l'enzyme $\mathbb{E}_1$ à l'instant t. Un flux métabolique correspond à un débit de matière instantané à travers une enzyme. Ce flux ν_E s'écrit donc comme une fonction non linéaire du substrat de l'enzyme x_1 multipliée par la concentration totale de l'enzyme E_T.

$$\nu_E = f_E(x_1) E_T \tag{2.1}$$

où $f_E(x_1)$ est par définition une fonction continue et strictement croissante de x_1, telle que :

- $f_E(0) = 0$ et pour tout $x_1 > 0$, on a $f_E(x_1) > 0$.
- $\lim_{x_1 \longrightarrow \infty} = k_2 > 0$.

En outre, l'activité d'une enzyme peut être modifiée par différents facteurs comme la température ou le pH, mais aussi par accrochage direct d'un métabolite sur l'enzyme. Typiquement ces métabolites correspondent aux régulations enzymatiques que nous avons identifiées lors de la construction du modèle de connaissance. Nous n'allons pas dériver mathématiquement tous les cas possibles de modulations enzymatiques. Le principe consiste (i) à revenir à la description biochimique de l'enzyme afin de représenter un autre complexe enzymatique où le modulateur est lié, puis (ii) à écrire l'équilibre biochimique associé à la conservation de la masse de l'enzyme.

Cette description des cinétiques enzymatiques est riche mais elle nécessite l'identification des différents paramètres cinétiques et donc des expériences biologiques spécifiques. Si cette identification est possible lors de l'étude d'une enzyme particulière, il semble impensable d'identifier ces paramètres pour l'ensemble des enzymes inclues dans notre modèle. Ainsi nous allons représenter les cinétiques enzymatiques à travers des fonctions

correspondantes à l'équation (2.1) dont les hypothèses associées seront consistantes par rapport aux équations réelles de cinétique enzymatique.

b. Modélisation de la voie métabolique

Les concentrations du i-ème métabolite et de la i-ème enzyme sont notées respectivement x_i et E_i. Le flux ν_0 associé à l'enzyme $\mathbb{E}_0$ produit le métabolite X_1 en entrée de la voie. Le flux de sortie ν_n correspond au flux de métabolite X_n à travers l'enzyme $\mathbb{E}_n$ qui peut par exemple être consommé par la cellule, ou correspondre à un flux d'entrée pour une autre voie métabolique. L'évolution au cours du temps de ces différentes concentrations obéit au système d'équations différentielles suivant :

$$\left\{\begin{array}{lcl} \dot{x}_1(t) & = & \nu_0(t) - E_1(t)f_1(x_1(t), x_2(t), x_n(t)) \\ \dot{x}_2(t) & = & E_1(t)f_1(x_1(t), x_2(t), x_n(t)) - E_2(t)f_2(x_2(t), x_3(t)) \\ \vdots & \vdots & \vdots \qquad\qquad \vdots \\ \dot{x}_n(t) & = & E_{n-1}(t)f_{n-1}(x_{n-1}(t), x_n(t)) - \nu_n(t) \end{array}\right. \tag{2.2}$$

où $\nu_n(t) \triangleq E_n(t)f_n(x_n(t))$ et les caractéristiques des activités enzymatiques f_i sont telles que :

(a) **Enzymes réversibles :**

- pour les enzymes intermédiaires $\mathbb{E}_i$ pour $i \in \{2, \ldots, n-1\}$: f_i est continue, strictement croissante en x_i et strictement décroissante en x_{i+1} telle que $f_i(0,0) = 0$, $f_i(x_i, 0) > 0$ pour tout $x_i > 0$, $f_i(0, x_{i+1}) < 0$ pour tout $x_{i+1} > 0$. De plus, il existe $M_i > 0$, $M_i' \geq 0$ tels que $f_i(x_i, x_{i+1}) \in]-M_i', M_i[$ pour tout $x_i > 0$ et $x_{i+1} \geq 0$. Enfin, on suppose que pour tout $x_i > 0$ il existe toujours $x_{i+1} > 0$ tel que $f_i(x_i, x_{i+1}) = 0$.

- pour la première enzyme $\mathbb{E}_1$:$f_1(0, 0, x_n) = 0$ pour tout $x_n \geq 0$, $f_1(x_1, 0, x_n) > 0$ pour tout $x_1 > 0$ et tout $x_n \geq 0$ et $f_1(0, x_2, x_n) < 0$ pour tout $x_2 > 0$ et tout $x_n \geq 0$. De plus, il existe $M_1 > 0$, $M_1' \geq 0$ tels que $f_1(x_1, x_2, x_n) \in]-M_1', M_1[$, pour tout $x_1 > 0$, $x_2 \geq 0$ et $x_n \geq 0$. On suppose également que pour tout $x_1 > 0$ et $x_n \geq 0$ il existe $x_2 > 0$ tel que $f_1(x_1, x_2, x_n) = 0$. Enfin, $f_1(x_1, x_2, x_n)$ est une fonction strictement croissante (resp. strictement décroissante) en fonction de son 1er argument (resp. son deuxième argument). De plus, si pour tout $x_1 > 0$ et $x_2 > 0$, $f_1(x_1, x_2, 0) < 0$ alors f_1 est une fonction strictement croissante en fonction de son troisième argument et pour tout $x_1 > 0$ et $x_2 > 0$, $f_1(x_1, x_2, 0) > 0$ alors f_1 est une fonction strictement décroissante en fonction de son troisième argument et

$$\text{pour tout } x_1 > 0 \text{ et } x_2 > 0, \lim_{x_n \to +\infty} f_1(x_1, x_2, x_n) = 0.$$

- pour la dernière enzyme $\mathbb{E}_n$: $\mathbb{E}_n$ décrit l'ensemble des réactions qui utilisent le produit final et résume ainsi la relation entre le flux fourni par la voie métabolique et la concentration finale. Ainsi, les propriétés de f_n dépendent principalement des propriétés des autres modules. f_n est une fonction continue strictement croissante de x_n telle que $f_n(0) = 0$.

(b) **Enzymes irréversibles :**

- pour les enzymes intermédiaires $\mathbb{E}_i$ pour $i \in \{2, \ldots, n-1\}$: f_i est continue, strictement croissante en x_i telle que $f_i(0) = 0$. De plus, il existe $M_i > 0$ tel que $f_i(x_i) \in]0, M_i[$ pour tout $x_i > 0$ et

$$\lim_{x_i \to +\infty} f_i(x_i) = M_i.$$

- pour la première enzyme $\mathbb{E}_1$: $f_1(0, x_n) = 0$, pour tout $x_n \geq 0$ et il existe $M_1 > 0$ tel que $f_1(x_1, x_n) \in]0, M_1[$, pour tout $x_1 > 0$ et $x_n \geq 0$. De plus, f_1 est une fonction continue, strictement croissante de son premier argument, et strictement décroissante de son second argument telle que

$$\text{pour tout } x_1 > 0, \lim_{x_n \to +\infty} f_1(x_1, x_n) = 0.$$

c. Modélisation du contrôle génétique

L'originalité de notre démarche par rapport aux études existantes [6, 7] consiste non seulement à considérer les réactions enzymatiques, mais aussi à intégrer le contrôle génétique, c'est-à-dire la synthèse des enzymes. Peu de modèles dynamiques ont considéré les régulations génétiques dans les voies métaboliques [227, 183]. Notre modèle de la régulation génétique s'inscrit pleinement dans le cadre de ces papiers.
La synthèse d'une enzyme se décompose en deux étapes. L'ARN polymérase transcrit le gène codant pour cette enzyme en un ARN messager, qui est ensuite traduit par un ribosome pour former l'enzyme finale. En notant m, Y_L et R_L respectivement les concentrations de l'ARN messager, des ARN polymérases et des ribosomes libres (c'est-à-dire non séquestré par un ARN messager), un modèle dynamique simplifié associé à la synthèse d'une enzyme $\mathbb{E}$ peut s'écrire :

$$\begin{cases} \dot{m}(t) &= k_m Y_L(t - \tau_m) - k_d m(t) \\ \dot{E}(t) &= k_e m(t) R_L(t - \tau_e) - \mu E(t) \end{cases} \tag{2.3}$$

où

(a) k_m, k_d et k_e désignent respectivement l'affinité du promoteur du gène pour l'ARN polymérase, la durée de vie de l'ARN messager et l'affinité du ribosome pour l'ARN messager ;

(b) μ est le taux de croissance de la bactérie supposé être dans une phase de croissance exponentielle (μ est alors tel que $\dot{N}(t) = \mu N(t)$ où $N(t)$ est la concentration de la population) ;

(c) le retard τ_m correspond au retard nécessaire à la transcription de l'ARN messager ;

(d) le retard τ_e correspond au retard nécessaire à la traduction de l'ARN messager.

De plus, si la synthèse de l'ARN messager ou de la protéine est inhibée par un facteur (par exemple un métabolite), alors ces équations sont modulées par des fonctions de ce facteur. Par exemple, si la transcription de l'ARN messager est inhibée par un métabolite X de concentration x, alors l'équation de la synthèse de l'ARN messager devient :

$$\dot{m}(t) = k_m f_I(x(t)) Y_L(t - \tau_m) - k_d m(t)$$

où $f_I(x(t))$ est une fonction continue, positive et strictement décroissante de x.

Tout comme dans le modèle de la réaction enzymatique, les dynamiques d'évolution des ARN messagers et des enzymes sont très différentes. La constante de dégradation de l'ARN messager est de l'ordre de 2-3 min [56] alors que le taux de croissance μ correspond à des temps de division allant de 30min à plusieurs heures suivant le milieu extracellulaire [184]. Ici encore nous allons considérer que l'équation différentielle de l'ARN messager est en régime permanent vis-à-vis de celle de la production de l'enzyme

$$\bar{m} = \frac{k_m}{k_d} f_I(\bar{x}) \bar{Y}_L,$$

que nous substituons dans l'équation de la production de l'enzyme pour donner :

$$\dot{E}(t) = k_e \frac{k_m}{k_d} f_I(\bar{x}) \bar{Y}_L R_L(t - \tau_e) - \mu E(t)$$

En outre, le retard de traduction τ_e est également très petit devant μ (de l'ordre de quelques minutes [28]) et sera donc négligé par la suite. Le terme de production de l'enzyme $k_e \frac{k_m}{k_d} f_I(\bar{x}) \bar{Y}_L R_L(t)$ résultant est une fonction croissante des concentrations des ARN polymérases et des ribosomes libres, et une fonction strictement décroissante de x pour notre exemple de régulation choisi.

Ainsi, nous allons considérer dans la suite de ce chapitre le système dynamique suivant pour la synthèse des enzymes suivant les cas de régulation connus :

$$\begin{cases} \dot{E}(t) &= g(x(t)) - \mu E(t) \\ \dot{E}(t) &= h(t) - \mu E(t) \end{cases} \tag{2.4}$$

La première équation sera utilisée pour décrire la synthèse d'enzymes régulée par un métabolite de concentration x, et la seconde équation correspondra à la synthèse d'enzymes non régulées. Les fonctions g et h sont des fonctions continues, positives, strictement monotones par rapport à leurs arguments et bornées car il existe intrinsèquement une valeur maximale à $\bar{Y}_L$ et $\bar{R}_L$ (ainsi qu'à tout autre paramètre physiologique dont le modèle pourrait dépendre) car le volume de la bactérie est fini.

Reprenons maintenant notre exemple de voie métabolique donnée sur la figure 2.5 où la synthèse de la première enzyme $\mathbb{E}_1$ est inhibée par le métabolite final x_n. On obtient

le système d'équations différentielles suivant en reprenant les conventions du système (2.4) :

$$\begin{cases} \dot{E}_1(t) &= g(x_n(t)) - \mu E_1(t) \\ \dot{E}_i(t) &= h_i(t) - \mu E_i(t) \quad \text{pour } i \in \{2, \ldots, n\} \end{cases} \tag{2.5}$$

où

(a) la caractéristique de la fonction génétique g est telle que g est une fonction continue positive, strictement décroissante et continue en x_n. De plus, nous supposons que $g(0) = P_{max}$ avec $P_{max} > 0$ et $\lim_{x \to +\infty} g(x) = 0$;

(b) les fonctions h_i sont des fonctions continues, positives, strictement croissantes et bornées.

Remarque 2.1 *Dans la suite de cette thèse, nous allons considérer que, lorsque la synthèse des concentrations des enzymes intermédiaires n'est pas régulée par un métabolite appartenant à la voie métabolique, le régime permanent existe toujours et qu'il est déterminé à travers les régimes d'équilibres des autres composants de la cellule, comme la concentration des ribosomes en régime permanent. Dans ces cas-là, nous ne ferons plus apparaître les équations relatives à la synthèse des enzymes intermédiaires* $\mathbb{E}_i$ *dans les systèmes d'équations différentielles.*

d. Description finale du modèle

On obtient donc le système d'équations différentielles final associé à la figure 2.5 :

$$\begin{cases} \dot{x}_1(t) &= \nu_0(t) - E_1(t) f_1(x_1(t), x_2(t), x_n(t)) \\ \dot{x}_2(t) &= E_1(t) f_1(x_1(t), x_2(t), x_n(t)) - E_2(t) f_2(x_2(t), x_3(t)) \\ \vdots & \vdots \qquad\qquad \vdots \qquad\qquad\qquad \vdots \\ \dot{x}_n(t) &= E_{n-1}(t) f_{n-1}(x_{n-1}(t), x_n(t)) - E_n(t) f_n(x_n(t)) \\ \dot{E}_1(t) &= g(x_n(t)) - \mu E_1(t) \end{cases} \tag{2.6}$$

La dépendance des équations par rapport au temps sera omise dans les notations pour des raisons de simplicité et de lisibilité dans la suite de ce chapitre.

2.3.2 Caractérisation du régime d'équilibre

Nous allons étudier les propriétés en régime permanent de la structure de contrôle par le produit final pour une configuration biologique de voie métabolique simple : toutes les enzymes sont irréversibles, les gènes codant pour les enzymes sont organisés en un seul opéron sur l'ADN. Nous étendrons ensuite les résultats à d'autres cas plus généraux : enzymes réversibles, gènes indépendants mais aussi lorsque des isoenzymes, des co-métabolites ou des points de branchement avec d'autres voies métaboliques sont présents. Ces configurations, identifiées sur la base de notre modèle, sont suffisamment réalistes pour représenter par combinaison, l'ensemble des voies métaboliques de *Bacillus subtilis*. Nous avons toutefois choisi de mettre en avant les conséquences de la réversibilité

des enzymes et de l'organisation des gènes car elles touchent respectivement le niveau métabolique et génétique.

a. Les enzymes sont irréversibles et les gènes sont en opéron

Toutes les enzymes sont supposées ici irréversibles, et les gènes codant pour les enzymes $(\mathbb{E}_1, \ldots, \mathbb{E}_{n-1})$ sont organisés en opéron. Les gènes organisés en un même opéron sont transcrits en un seul ARN messager. En conséquence, nous supposons que les concentrations des enzymes sont proportionnelles, ce qui conduit à l'hypothèse suivante.

Hypothèse 2.1 (Hypothèse de l'opéron) *Pour tout $i \in \{2, \ldots, n-1\}$ il existe $\alpha_i \in]0, 1]$ tel que $E_i(t) = \alpha_i E_1(t)$.*

Sous ces hypothèses, le système d'équations (2.6) devient :

$$\left\{\begin{array}{lcl} \dot{x}_1 & = & \nu_0 - E_1 f_1(x_1, x_n) \\ \dot{x}_2 & = & E_1 f_1(x_1, x_n) - E_2 f_2(x_2) \\ \vdots & = & \vdots \qquad\qquad \vdots \\ \dot{x}_n & = & E_{n-1} f_{n-1}(x_{n-1}) - E_n f_n(x_n) \\ \dot{E}_1 & = & g(x_n) - \mu E_1 \\ E_i & = & \alpha_i E_1 \quad \text{pour } i \in \{2, \ldots, n-1\} \end{array}\right. \tag{2.7}$$

dont le régime d'équilibre est caractérisé par :

Proposition 2.2 *Pour tout $\mu > 0$, $E_n > 0$ et $\bar{x}_1 > 0$ il existe un unique régime d'équilibre $(\bar{E}_1, \ldots, \bar{E}_{n-1})$ et $(\bar{x}_2, \ldots, \bar{x}_n)$ pour le système (2.7) tel que*

$$\left\{\begin{array}{l} \bar{E}_1 = \dfrac{g(\bar{x}_n)}{\mu} \\ f_1(\bar{x}_1, \bar{x}_n) g(\bar{x}_n) = \mu E_n f_n(\bar{x}_n) \\ \nu_0 = E_n f_n(\bar{x}_n) \end{array}\right. \tag{2.8}$$

et tel que pour $i \in \{2, \ldots, n-1\}$,

$$\bar{E}_i = \alpha_i \frac{g(\bar{x}_n)}{\mu} \qquad \text{et} \qquad \bar{x}_i = f_i^{-1}\left(\frac{\mu E_n f_n(\bar{x}_n)}{\alpha_i g(\bar{x}_n)}\right), \tag{2.9}$$

si et seulement si pour tout $i \in \{2, \ldots, n-1\}$, on a

$$\frac{\mu E_n f_n(\bar{x}_n)}{g(\bar{x}_n)} < \alpha_i M_i. \tag{2.10}$$

La démonstration de cette proposition est un cas particulier de la démonstration de la proposition 2.8 donnée page 50. Cette proposition indique que le système (2.7) a un unique point d'équilibre si toutes les enzymes de la voie métabolique ne saturent pas (la condition (2.10) est vraie), c'est-à-dire si le dernier point de la proposition est satisfait.

De plus, x_n et implicitement f_1 et g jouent un rôle central dans la définition du point d'équilibre. La monotonicité de f_1 et g par rapport à x_n permet de conclure sur l'unicité de x_n. De manière surprenante, les caractéristiques des fonctions enzymatiques f_i ainsi que le niveau des enzymes intermédiaires E_i n'ont aucun impact sur le point d'équilibre si aucune de ces enzymes intermédiaires ne sature.
Donc la sensibilité du point d'équilibre à une perturbation constante de concentration d'enzyme E_n (ou à une demande de flux ν_n) dépend uniquement de la caractéristique génétique g et enzymatique f_1 de la première enzyme. La prédiction du comportement de la voie métabolique à l'équilibre est ainsi simple même si de nombreuses réactions composent cette voie.

Remarque 2.2 *Par définition du régime d'équilibre, nous avons :*

$$\frac{\mu E_n f_n(\bar{x}_n)}{g(\bar{x}_n)} = f_1(\bar{x}_1, \bar{x}_n)$$

Donc la condition (2.10) peut être réécrite comme :

$$\frac{\mu E_n f_n(\bar{x}_n)}{g(\bar{x}_n)} < \alpha_i M_i \Longleftrightarrow f_1(\bar{x}_1, \bar{x}_n) < \alpha_i M_i.$$

À partir de cette expression, nous pouvons en déduire que si $M_1 < \alpha_i M_i$ pour tout $i \in \{2, \cdots, n-1\}$ alors la condition (2.10) reste vraie. L'activité maximale de la première enzyme doit ainsi rester inférieure aux activités maximales des enzymes intermédiaires multipliées par le paramètre α_i.

a.1 Comportement des composants de la voie métabolique

Étudions maintenant la variation de la demande de flux en fonction de la variation de la concentration de l'enzyme $\mathbb{E}_n$. Selon la définition de l'équilibre, nous avons :

$$E_n = \frac{f_1(\bar{x}_1, \bar{x}_n) g(\bar{x}_n)}{\mu f_n(\bar{x}_n)},$$

Donc $\bar{x}_n$ décroît strictement lorsque E_n augmente. En outre, par définition de l'équilibre, la demande de flux finale $\bar{\nu}_n \triangleq E_n f_n(\bar{x}_n)$ est par définition de l'équilibre égale à

$$f_1(\bar{x}_1, \bar{x}_n) g(\bar{x}_n) = \bar{\nu}_n,$$

alors puisque la partie gauche de l'équation est une fonction décroissante de $\bar{x}_n$, alors lorsque E_n augmente on en déduit que $\bar{\nu}_n$ augmente également (et aussi longtemps qu'aucune enzyme ne sature). Cette analyse nous permet de montrer que si E_n tend vers l'infini, alors $\bar{x}_n$ tend vers 0. Ce résultat implique que la voie métabolique a une capacité maximale de flux, donnée par :

Corollaire 2.3 *Soient les hypothèses de la proposition 2.2 satisfaites ; en conséquence, la demande de flux est bornée supérieurement par*

$$\bar{\nu}_n < \bar{\nu}_{n,max} \triangleq \frac{g(0)}{\mu} f_1(\bar{x}_1, 0). \tag{2.11}$$

Le flux de sortie est donc borné. Là encore cette valeur maximale dépend uniquement des caractéristiques de f_1 et de g de la première enzyme, sous l'hypothèse que les enzymes intermédiaires de la voie ne saturent pas. En revanche, la borne supérieure de la demande de flux n'est pas atteignable car elle correspond à une concentration infinie de l'enzyme finale, *i.e* $E_n = \infty$. Si la capacité maximale de flux de chaque enzyme intermédiaire reste supérieure ou égale à $\bar{\nu}_{n,max}$, ce qui est équivalent à $\alpha_i M_i \geq f_1(\bar{x}_1, 0)$ pour tout $i \in \{2, \ldots, n-1\}$, alors pour tout $\epsilon > 0$ (assez petit), il existe E_n tel que $\bar{\nu}_n = \bar{\nu}_{n,max} - \epsilon$.

Considérons maintenant l'impact de variations dans la demande de flux sur les différentes concentrations des métabolites de la voie.

Corollaire 2.4 *Soient les hypothèses de la proposition 2.2 satisfaites. Alors, pour tout* $i \in \{2, \ldots, n-1\}$, $\bar{x}_i = \bar{x}_i(E_n)$ *est une fonction strictement croissante de* E_n *alors que* $\bar{x}_n = \bar{x}_n(E_n)$ *est une fonction strictement décroissante de* E_n.

Lorsque la demande de flux de sortie augmente lors d'une augmentation de E_n, les concentrations des métabolites intermédiaires de la voie augmentent alors que celle du produit final diminue. De même, les effets dus aux variations de la concentration x_1 conduisent à un comportement intéressant de la voie :

Corollaire 2.5 *Soient les hypothèses de la proposition 2.2 satisfaites. Alors, pour tout* $i \in \{2, \cdots, n\}$, $\bar{x}_i(\bar{x}_1)$ *et* $\bar{\nu}_n(\bar{x}_1)$ *sont des fonctions strictement croissantes de* $\bar{x}_1$.

Remarque 2.3 *Il est possible d'exprimer* $\bar{x}_n$ *en fonction de* $\bar{x}_1$ *ainsi :* $\bar{x}_n \triangleq H(\bar{x}_1)$. *Par conséquent, en régime d'équilibre, le flux d'entrée (et le flux de sortie) et la concentration du premier métabolite* $\bar{x}_1$ *sont liés par la relation monotone strictement croissante* $\nu_0 = E_n f_n(H(\bar{x}_1))$. *On obtient alors à l'équilibre une description entrée/sortie correspondant à une enzyme fictive reliant* ν_0 *à* $\bar{x}_1$ *et intégrant toutes les propriétés du module à travers les fonctions* H *et* f_n.

Remarque 2.4 *Un flux métabolique correspond à un débit de matière à travers une enzyme :* $\nu = E.f_E(x)$. *Un flux métabolique est donc une grandeur intensive, alors que les concentrations des métabolites sont des grandeurs extensives. Ceci explique pourquoi dans la plupart des mécanismes de régulation génétique, seule la concentration d'un métabolite est utilisée (et non le flux). Tout comme dans la loi d'Ohm* ($U = RI$) *où l'intensité du courant* I *est mesurée à travers une mesure de la tension* U *pour une résistance* R, *la cellule mesure son flux* ν *à travers la mesure de la concentration* x *via un mécanisme particulier comme par exemple un facteur de transcription.*

a.2 Saturation d'une enzyme de la voie

Tous les résultats précédents ont été obtenus sur la base de la non saturation des enzymes intermédiaires de la voie métabolique. Examinons maintenant les conséquences de la saturation d'une enzyme. La saturation peut être causée par une concentration

d'enzymes insuffisante ou, comme nous allons le voir ci-après, à une limitation en co-métabolites ou en co-facteurs comme par exemple dans le cas d'un appauvrissement du milieu en ions (voir la section 2.3.3).

Corollaire 2.6 *Soient les hypothèses de la proposition 2.2 satisfaites et définissons*

$$\psi_{i^*,sat} = \min_{i\in\{2,\cdots,n-1\}} \alpha_i M_i$$

et i^*, la valeur de l'indice i pour lequel le minimum est atteint (s'il y a plusieurs valeurs, alors i^* correspond à l'ensemble des valeurs possibles). Si $\psi_{i^*,sat}$ est tel que $\psi_{i^*,sat} < f_1(\bar{x}_1, 0)$ alors il existe E_n^* et $\bar{x}_n^*$ tels que

$$\frac{\mu E_n^* f_n(\bar{x}_n^*)}{g(\bar{x}_n^*)} = \psi_{i^*,sat}$$

et

$$\lim_{\substack{< \\ E_n \to E_n^*}} \bar{x}_{i^*} = +\infty.$$

En outre, pour $E_n \geq E_n^*$ la voie métabolique a un régime saturé.

Le flux de sortie est ainsi fixé à travers la saturation de l'enzyme $\psi_{i^*,sat}$ et par la caractéristique g de la régulation génétique de la première enzyme. De plus, la concentration du métabolite $\bar{x}_{i^*}$, substrat de l'enzyme i^* tend vers l'infini lorsque E_n tend vers E_n^*.

a.3 Interprétation biologique de ces résultats théoriques

Les voies de synthèse des acides aminés sont généralement régulées par le produit final. L'enzyme $\mathbb{E}_n$ correspond alors à l'ARNt synthase (voir l'annexe A.2 pour un rappel sur le rôle de l'ARNt synthase) et le flux de sortie ν_n correspond au flux d'ARNt chargés consommé par les ribosomes pour la production des protéines en régime permanent. Ainsi une augmentation de la demande des ribosomes dûe par exemple à une augmentation de leur activité se traduira par une augmentation de la concentration E_n de l'ARNt synthase grâce à une régulation génétique appropriée, qui induira à son tour une diminution de la concentration x_n de l'acide aminé. La diminution de x_n entraînera alors un réajustement de la voie toute entière (enzyme et métabolites) de manière à fournir le flux demandé sous l'hypothèse qu'aucune enzyme intermédiaire ne sature. En d'autres termes à $\bar{x}_1$ fixé, la concentration de l'acide aminé x_n doit diminuer pour augmenter la capacité de la voie de synthèse et satisfaire ainsi la demande en flux, dans la limite maximale définie par les caractéristiques de la première enzyme (corollaire 2.3).

b. Les gènes ne sont plus organisés en opéron

Étudions maintenant les conséquences de la relaxation de la première hypothèse sur ces différents résultats, à savoir l'organisation en opéron des gènes. Les gènes codant pour une voie métabolique n'appartiennent pas toujours au même opéron. On citera comme exemple chez *Bacillus subtilis* la voie de synthèse de la lysine, étudiée en détail

dans la section 3.3.1 du prochain chapitre. Le résultat obtenu dans la proposition 2.2 est légèrement modifié, car seule la condition de saturation change. On supposera ici que le régime permanent des enzymes intermédiaires existe et donné par $(E_i)_{i\in\{2,\cdots n\}} > 0$.

Proposition 2.7 *Pour tout $\mu > 0$, $\bar{x}_1 > 0$ et $E_i > 0$ pour $i \in \{2, \cdots, n\}$, il existe un régime d'équilibre unique $\bar{E}_1$ et $(\bar{x}_2, \ldots, \bar{x}_n)$ pour (2.7) tel que*

$$\begin{cases} \bar{E}_1 = \dfrac{g(\bar{x}_n)}{\mu}, \\ f_1(\bar{x}_1, \bar{x}_n)g(\bar{x}_n) = \mu E_n f_n(\bar{x}_n), \\ \nu_0 = E_n f_n(\bar{x}_n) \end{cases} \tag{2.12}$$

et tel que pour tout $i = \{2, \ldots, n-1\}$,

$$\bar{x}_i = f_i^{-1}\left(\frac{E_n f_n(\bar{x}_n)}{E_i}\right) \tag{2.13}$$

si et seulement si $E_n f_n(\bar{x}_n) < E_i M_i$.

Ces résultats sont proches de ceux obtenus lorsque les gènes appartiennent à un même opéron. En effet, aussi longtemps qu'aucune enzyme intermédiaire n'est saturée, le lien entre la demande en flux et les concentrations des métabolites initiaux et finaux, obtenu dans la proposition 2.2, reste inchangé. Comme dans le cas précédent, si nous supposons qu'il existe $i^* \in \{2, \cdots n-1\}$ tel que $E_{i^*}M_{i^*} < \bar{\nu}_{n,max}$, alors il existe $E_n^* > 0$ et $x_n^* > 0$ tels que $E_n^* f_n(x_n^*) = E_i^* M_i^*$ et $\lim\limits_{\bar{E}_n \lneq E_n^*} \bar{x}_{i^*} = +\infty$.
Ainsi tous les résultats présentés dans le cas précédent peuvent être facilement étendus lorsque l'hypothèse de l'organisation des gènes en opéron est supprimée.

c. Toutes les enzymes sont réversibles

Intéressons nous maintenant à la suppression de la dernière hypothèse concernant l'irréversibilité des enzymes de la voie. Nous présentons tout d'abord un résultat général pour les voies métaboliques dans lesquelles toutes les enzymes de la voie, y compris la première enzyme sont réversibles. Cette représentation change les propriétés obtenues à l'équilibre dans la proposition 2.2. Nous montrerons ensuite qu'une seule étape irréversible dans la voie permet une restauration partielle des propriétés de la proposition 2.2.

Considérons tout d'abord le cas où toutes les enzymes de la voie sont réversibles (avec une organisation en opéron pour les gènes).

Proposition 2.8 *Si l'hypothèse 2.1 est satisfaite et si l'ensemble des enzymes de la voie $\mathbb{E}_i$ pour tout $i \in \{1, \cdots, n-1\}$ sont réversibles, alors pour tout $\mu > 0$, $E_n > 0$ et $\bar{x}_1 > 0$, il existe un unique régime d'équilibre pour le système (2.6), $(\bar{E}_1, \ldots, \bar{E}_{n-1})$ et $(\bar{x}_2, \ldots, \bar{x}_n)$*

tel que :

$$\begin{cases} \bar{x}_n = H_n(H_2(...(H_{n-1}(\bar{x}_n, \bar{x}_1), \bar{x}_1) \ldots, \bar{x}_1), \bar{x}_1), \\ \bar{x}_i = H_i(H_{i+1}(...(H_{n-1}(\bar{x}_n, \bar{x}_1), \bar{x}_1) \ldots, \bar{x}_1), \bar{x}_1), \ \textit{pour } i \in \{2, \cdots, n-1\}, \\ E_1 = \dfrac{g(\bar{x}_n)}{\mu}, \\ E_i = \alpha_i \dfrac{g(\bar{x}_n)}{\mu} \ \textit{pour } i \in \{2, \cdots, n-1\} \\ \nu_0 = E_n f_n(\bar{x}_n) \end{cases} \tag{2.14}$$

où H_i sont des fonctions strictement croissantes de leur argument pour tout $i \in \{2, \cdots, n-1\}$ et H_n est une fonction strictement décroissante de son premier argument, et strictement croissante de son second argument.

Démonstration

La démonstration est faite par récurrence.

Étape 1 : Tout d'abord, prouvons qu'il existe $x_2^* > 0$ tel que pour tout $\bar{x}_2 \in [0, x_2^*]$, il existe un unique $\bar{x}_n \geq 0$ tel que

$$\frac{g(\bar{x}_n)}{\mu} f_1(\bar{x}_1, \bar{x}_2, \bar{x}_n) = E_n f_n(\bar{x}_n). \tag{2.15}$$

La monotonicité des fonctions de part et d'autre de l'équation en fonction de $\bar{x}_n$ permet de noter que pour tout $\bar{x}_1 > 0$ il existe $x_2^* > 0$ tel que $f_1(\bar{x}_1, \bar{x}_2, 0) > 0$ pour tout $\bar{x}_2 \in [0, x_2^*[$ avec $f_1(\bar{x}_1, x_2^*, 0) = 0$. Alors, pour tout $\bar{x}_2 \in [0, x_2^*[$, la partie gauche de l'équation (2.15) est une fonction strictement décroissante de $\bar{x}_n$, strictement positive pour $\bar{x}_n = 0$ et qui tend vers 0 lorsque $\bar{x}_n$ tend vers l'infini. En outre, le côté droit de l'équation (2.15) est une fonction strictement croissante de $\bar{x}_n$, égale à 0 pour $\bar{x}_n = 0$. Par conséquent, pour tout $\bar{x}_2 \in [0, x_2^*[$, les deux courbes en fonction de $\bar{x}_n$ ont nécessairement une intersection unique et pour $\bar{x}_2 = x_2^*$, $\bar{x}_n = 0$ est la seule solution de l'équation (2.15) (voir la figure 2.6 de gauche). L'étape 1 est démontrée.

Nous sommes donc en mesure de définir une fonction $\bar{x}_n \triangleq H_n(\bar{x}_2, \bar{x}_1)$ bien définie, continue et strictement décroissante en fonction de $\bar{x}_2$ pour $\bar{x}_2 \in [0, x_2^*]$ et telle que $H_n(0, \bar{x}_1) > 0$ et $H_n(x_2^*, \bar{x}_1) = 0$. Dans la suite de cette démonstration, nous avons omis la dépendance des équations en $\bar{x}_1$ pour alléger les notations.

Étape 2 : Le reste de la démonstration est faite par récurrence. Si le régime d'équilibre existe, alors $\bar{x}_2$ et $\bar{x}_3$ sont liés comme suit :

$$\alpha_2 \frac{g(H_n(\bar{x}_2))}{\mu} f_2(\bar{x}_2, \bar{x}_3) = E_n f_n(H_n(\bar{x}_2)), \tag{2.16}$$

où $\bar{x}_n$ a été remplacé par son expression et $\bar{x}_1$ n'est plus noté pour alléger les notations. Comme pour la première étape, nous pouvons prouver qu'il existe[1] $\bar{x}_3^* > 0$ tel que pour tout $\bar{x}_3 \in [0, x_3^*]$, il existe $x_2 \in [0, x_2^*]$ tel que l'équation (2.16) est vraie. Nous sommes

[1] En fait, x_3^* est tel que $f_2(x_2^*, x_3^*) = 0$, ce qui garantit que $f_2(x_2^*, \bar{x}_3) > 0$ pour tout $[0, x_3^*[$.

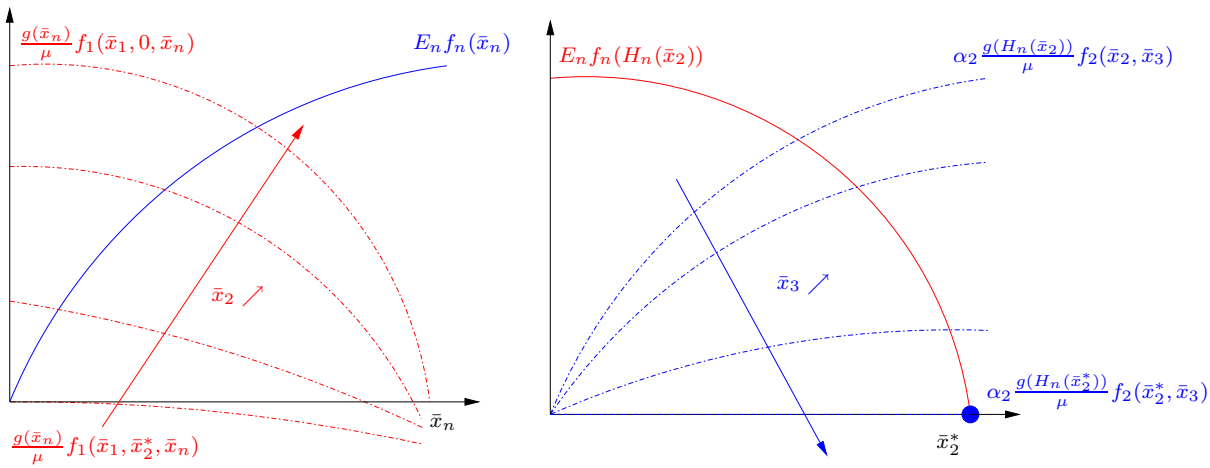


FIG. 2.6 – Intersection des courbes $E_n f_n(\bar{x}_n)$ et $f_1(\bar{x}_1, \bar{x}_2, \bar{x}_n)$ pour tout $\bar{x}_2 \in [0, x_2^*]$ (figure de gauche) et des courbes $E_n f_n(H_n(\bar{x}_2))$ et $\alpha_2 \dfrac{g(H_n(\bar{x}_2))}{\mu} f_2(\bar{x}_2, \bar{x}_3)$ (figure de droite)

alors en mesure de définir $\bar{x}_2 \triangleq H_2(\bar{x}_3)$, qui est une fonction bien définie continue et strictement croissante en $\bar{x}_3$ pour tout $\bar{x}_3 \in [0, x_3^*]$ et telle que $H_2(0) > 0$ et $H_2(x_3^*) = x_2^*$ (voir la figure 2.6 de droite).

Étape 3 : La démonstration de l'étape 2 peut être répétée pour tout $i \in \{3, \cdots, n-1\}$. En effet, par définition, $\bar{x}_i$ doit être la solution de l'équation suivante :

$$\alpha_i \frac{g(H_n(H_2(...(H_{i-1}(\bar{x}_i)))))}{\mu} f_i(\bar{x}_i, \bar{x}_{i+1}) = E_n f_n(H_n(H_2(...(H_{i-1}(\bar{x}_i))))). \tag{2.17}$$

Puis, comme pour l'étape précédente, il est facile de prouver que H_i existe tel que $x_i \triangleq H_i(x_{i+1})$ est une fonction bien définie, continue et strictement croissante en $\bar{x}_{i+1}$ pour $\bar{x}_{i+1} \in [0, x_{i+1}^*]$ et telle que $H_i(0) > 0$ et $H_i(x_{i+1}^*) = x_i^*$.

Étape 4 : En combinant les résultats des étapes précédentes, nous déduisons que $\bar{x}_n$ existe si l'équation suivante a une solution :

$$\bar{x}_n = H_n(H_2(...(H_{n-1}(\bar{x}_n)))). \tag{2.18}$$

Par définition, H_{n-1} est définie sur $[0, x_n^*]$ telle que $H_{n-1}(0) > 0$ et $H_{n-1}(x_n^*) = x_{n-1}^*$. Notons que $H_n(H_2(...(H_{n-1}(0))) > 0$ et $H_n(H_2(...(H_{n-1}(x_n^*))) = H_n(x_2^*) = 0$ et parce que la partie droite (respectivement gauche) de l'équation (2.18) est une fonction strictement décroissante (respectivement strictement croissante) de $\bar{x}_n$, on en déduit qu'il existe un unique $\bar{x}_n \in [0, x_n^*[$, solution de l'équation (2.18) ce qui conclut cette proposition.

■

Ainsi lorsque toutes les enzymes de la voie sont réversibles, la voie a toujours un régime d'équilibre. Nous n'allons pas développer davantage l'analyse de cette structure. En effet,

l'analyse systématique des réseaux métaboliques des bactéries indique que presque toutes les voies métaboliques linéaires possèdent au moins une étape irréversible [107, 75]. Dans la plupart des cas, l'étape irréversible correspond à la première ou la deuxième enzyme de la voie. La présence d'une étape irréversible en début de voie conduit aux résultats de la proposition 2.2 comme suit :

Corollaire 2.9 *Si l'hypothèse 2.1 est satisfaite et si l'enzyme $\mathbb{E}_1$ est irréversible, alors pour tout $\mu > 0$, $E_n > 0$ et $\bar{x}_1 > 0$ il existe un unique régime d'équilibre $(\bar{E}_1, \ldots, \bar{E}_{n-1})$ et $(\bar{x}_2, \ldots, \bar{x}_n)$ pour le système (2.6) tel que*

$$\begin{cases} \bar{E}_1 = \dfrac{g(\bar{x}_n)}{\mu} \\ f_1(\bar{x}_1, \bar{x}_n) g(\bar{x}_n) = \mu E_n f_n(\bar{x}_n) \\ \nu_0 = E_n f_n(\bar{x}_n), \end{cases} \tag{2.19}$$

si et seulement si $(\bar{x}_2, \cdots, \bar{x}_{n-1})$ existe tel que

$$\alpha_i \frac{g(\bar{x}_n)}{\mu} f_i(\bar{x}_i, \bar{x}_{i+1}) = E_n f_n(\bar{x}_n)$$

pour $i \in \{2, \ldots, n-1\}$.

Comme dans la proposition 2.2, tant qu'aucune enzyme ne sature, le régime d'équilibre de la voie métabolique est défini seulement par f_1, g et f_n. Contrairement à la proposition 2.2, la condition de saturation est plus difficile à décrire et dépend explicitement du régime d'équilibre.

Remarque 2.5 *Le raisonnement précédent peut être facilement appliqué à une voie métabolique linéaire où la première étape irréversible correspond à la i-ème réaction. En effet, les arguments utilisés dans la démonstration de la proposition 2.8 nous permettent également de déduire que $\bar{x}_i$ est uniquement déterminé par*

$$\alpha_i \frac{g(H_n(H_2(\ldots H_{i-1}(\bar{x}_i)\ldots)))}{\mu} f_i(\bar{x}_i) = E_n f_n(H_n(H_2(\ldots H_{i-1}(\bar{x}_i)\ldots))),$$

où les différentes fonctions H_k pour $k \in \{2, \ldots, i-1\}$ sont des fonctions strictement croissantes de $\bar{x}_i$, et H_n est une fonction strictement décroissantes de son argument. Ces fonctions sont obtenues en suivant strictement la démonstration de la proposition 2.8. On en déduit que $\bar{x}_n$ est indépendant de toutes les fonctions $f_{i+1}, ..., f_{n-1}$ (s'il existe un régime d'équilibre). Il reste à s'assurer que la partie basse de la voie (après la i-ème enzyme) n'est pas saturée. Par exemple, il est facile de montrer que si toutes les enzymes après la i-ème étape sont réversibles, il existe un régime d'équilibre si et seulement s'il existe $(\bar{x}_{i+1}, \cdots, \bar{x}_{n-1})$ tel que $\alpha_i \dfrac{g(\bar{x}_n)}{\mu} f_i(\bar{x}_i, \bar{x}_{i+1}) = E_n f_n(\bar{x}_n)$ pour $i \in \{i+1, \ldots, n-1\}$.

2.3.3 Intégration de différentes composantes biologiques

Nous allons maintenant étudier plus précisément l'impact des différentes configurations biologiques suivantes sur le régime permanent : la présence d'isoenzymes, de cofacteurs et de co-métabolites, ou de points de branchement avec d'autres voies métaboliques. Les résultats mathématiques sont montrés dans le cas le plus général de la structure de contrôle par le produit final où seule la première enzyme est irréversible et où les gènes codant pour les enzymes de la voie métabolique ne sont pas organisés en opéron.

a. Rôle d'une isoenzyme

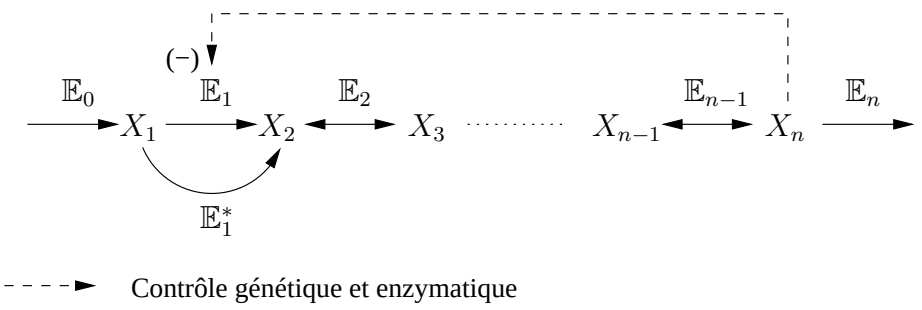


FIG. 2.7 – Présence d'une isoenzyme dans la voie métabolique

Deux isoenzymes correspondent à deux enzymes catalysant la même réaction chimique. Soit l'isoenzyme E_1^* représentée sur la figure 2.7 catalysant la même réaction que l'enzyme $\mathbb{E}_1$ (la première réaction) et non contrôlée par aucun intermédiaire de la voie métabolique en question (ni au niveau génétique, ni au niveau enzymatique). Le comportement de la voie métabolique au cours du temps devient :

$$\begin{cases} \dot{x}_1 = \nu_0 - E_1 f_1(x_1, x_n) - E_1^* f_1^*(x_1) \\ \dot{x}_2 = E_1 f_1(x_1, x_n) + E_1^* f_1^*(x_1) - E_2 f_2(x_2, x_3) \\ \vdots \\ \dot{x}_n = E_{n-1} f_{n-1}(x_{n-1}, x_n) - E_n f_n(x_n) \\ \dot{E}_1 = g(x_n) - \mu E_1 \end{cases} \tag{2.20}$$

Le point d'équilibre du précédent système satisfait l'équation suivante :

$$f_1(\bar{x}_1, \bar{x}_n)\frac{g(\bar{x}_n)}{\mu} + E_1^* f_1^*(\bar{x}_1) = E_n f_n(\bar{x}_n). \tag{2.21}$$

On peut déduire de (2.21) que l'augmentation du flux $\bar{\nu}_n$ peut être obtenue soit par la diminution du produit final $\bar{x}_n$, soit par l'augmentation de la concentration de l'isoenzyme E_1^* tant que les enzymes intermédiaires de la voie ne saturent pas. La capacité maximale

du flux dans la voie métabolique correspond maintenant à :

$$\bar{\nu}_{n,max} = f_1(\bar{x}_1, 0)\frac{g(0)}{\mu} + E_1^* f_1^*(\bar{x}_1). \tag{2.22}$$

Ainsi une façon de satisfaire la demande en flux ν_n est d'augmenter la concentration de l'isoenzyme E_1^*.

b. Impact des co-métabolites et des co-facteurs

Des facteurs externes sont souvent impliqués dans les réactions des voies métaboliques et correspondent soit à des métabolites comme par exemple l'ATP, NAD, le glutamate et la glutamine, soit à des co-facteurs tels que les ions (par exemple Mg^{2+}, Zn^{2+}) ou les vitamines. Ces facteurs modifient la vitesse de réaction d'enzymes spécifiques bien qu'ils soient de nature très différente : les ions et les vitamines (les co-facteurs) sont généralement associés à leurs enzymes et sont donc considérés comme une partie de l'enzyme alors que les co-métabolites sont transformés par l'enzyme en d'autres produits. Dans tous les cas, ces facteurs peuvent facilement être inclus dans notre analyse en introduisant leur effet sur la vitesse de réaction f_i de l'enzyme d'intérêt. Si l'on suppose que la i-ème réaction nécessite un co-métabolite, noté X_P de concentration p, alors la vitesse de réaction pour une réaction irréversible (resp. réversible) est donnée par $f_i(x_i, p)$ (resp. réversible : $f_i(x_i, x_{i+1}, p)$ telle que $f_i(x_i, 0) = 0$ (resp. réversible : $f_i(x_i, x_{i+1}, 0) = 0$) et où la fonction f_i est supposée croissante en p. Dans ces conditions, trois cas sont à considérer :

1. le co-métabolite/co-facteur agit sur la première enzyme $\mathbb{E}_1$ de la voie ;
2. le co-métabolite/co-facteur agit sur la dernière enzyme $\mathbb{E}_n$ de la voie ;
3. le co-métabolite/co-facteur agit sur une des enzymes intermédiaires $\mathbb{E}_2, \ldots, \mathbb{E}_{n-1}$ de la voie.

Les deux premières situations ont un effet direct sur le régime d'équilibre, la troisième n'a d'effet que s'il conduit à la saturation d'une étape intermédiaire de la voie. Il s'agit d'une propriété remarquable de la structure de contrôle par le produit final. Malgré la modification de la concentration d'un co-métabolite nécessaire pour produire le produit final, l'équilibre des principales variables de la voie ($\bar{\nu}_n, \bar{x}_n$ et $\bar{E}_1$) demeure inchangé, tant qu'aucune enzyme de la voie ne sature.

Si l'activité de la première enzyme dépend de p, alors le flux maximal de la voie est donné par

$$\bar{\nu}_{n,max}(\bar{p}) \triangleq \frac{g(0)}{\mu} f_1(\bar{x}_1, 0, \bar{p}),$$

où le co-métabolite/co-facteur p atteint son régime propre équilibre $\bar{p}$. Si le co-facteur diminue l'activité de la première enzyme, alors la concentration du produit final x_n diminue également et la concentration de la première enzyme augmente. Si le co-métabolite/co-facteur agit au niveau de l'enzyme finale $\mathbb{E}_n$ de la voie, alors par définition on a :

$$f_1(\bar{x}_1, \bar{x}_n) g(\bar{x}_n) = \mu E_n f_n(\bar{x}_n, \bar{p}).$$

La limitation de la concentration du co-métabolite/co-facteur conduit à diminuer la demande de flux. Par conséquent $\bar{x}_n$ est une fonction croissante de $\bar{p}$ et la concentration de la première enzyme est une fonction décroissante de $\bar{p}$.

c. Impact d'un point de branchement

Flux de sortie au niveau d'une enzyme intermédiaire

Considérons maintenant qu'il existe un flux de sortie ν_i au niveau de la i-ème enzyme correspondant à la demande en flux pour la voie métabolique de la figure 2.5, dans le cas où les gènes n'appartiennent pas au même opéron, et seule la première enzyme est irréversible. La voie métabolique est alors décrite par le système d'équations différentielles suivant :

$$\begin{cases} \dot{x}_1 = \nu_0 - E_1 f_1(x_1, x_n) \\ \dot{x}_2 = E_1 f_1(x_1, x_n) - E_2 f_2(x_2, x_3) \\ \quad \vdots \\ \dot{x}_i = E_{i-1} f_{i-1}(x_{i-1}, x_i) - E_i f_i(x_i, x_{i+1}) - E_i^* f_i^*(x_i) \\ \dot{x}_{i+1} = E_i f_i(x_i, x_{i+1}) - E_{i+1} f_{i+1}(x_{i+1}, x_{i+2}) \\ \quad \vdots \\ \dot{x}_n = E_{n-1} f_{n-1}(x_{n-1}, x_n) - E_n f_n(x_n) \\ \dot{E}_1 = g(x_n) - \mu E_1 \end{cases} \tag{2.23}$$

où $\nu_i = E_i^* f_i^*(x_i)$. On obtient alors le résultat suivant pour le régime d'équilibre :

Proposition 2.10 *Pour tout $\mu > 0$, $\bar{x}_1 > 0$ et $(E_i > 0)_{i \in 2,\ldots,n}$, il existe un unique régime d'équilibre pour le système (2.23) tel que :*

$$\frac{g(\bar{x}_n)}{\mu} f_1(\bar{x}_1, \bar{x}_n) = E_i^* f_i^*(H_i(\bar{x}_n)) + E_n f_n(\bar{x}_n), \tag{2.24}$$

où H_i est une fonction strictement croissante de son argument si et seulement si

$$\begin{cases} E_j f_j(\bar{x}_j, \bar{x}_{j+1}) = E_i^* f_i^*(H_i(\bar{x}_n)) + E_n f_n(\bar{x}_n) \ \textit{pour } 2 \leq j < i \\ E_j f_j(\bar{x}_j, \bar{x}_{j+1}) = E_n f_n(\bar{x}_n) \ \textit{pour } i \leq j < n \end{cases} \tag{2.25}$$

Démonstration

La démonstration se déduit de celle de la proposition 2.8 en remarquant que les enzymes $(\mathbb{E}_i, \ldots, \mathbb{E}_{n-1})$ sont toutes réversibles. D'après la proposition 2.8, on en déduit que $\bar{x}_i$ s'écrit comme une fonction H_i strictement croissante de $\bar{x}_n$. Les conditions de saturation des enzymes correspondent à celles du corollaire 2.9 modifiées pour le cas où les gènes n'appartiennent pas au même opéron.

■

On déduit de la proposition 2.10 que lorsque le flux de sortie ν_i augmente, $\bar{x}_n$ doit diminuer pour augmenter le flux dans la voie métabolique. Plus précisément, pour tout

$\nu_0 > 0$, $\nu_0 - \bar{\nu}_i = E_j f_j(\bar{x}_j)$ pour $j \geq i$ où $\bar{\nu}_i = E_i^* f_i^* (H_i(\bar{x}_n))$. Par conséquent, une augmentation de la demande de flux $\bar{\nu}_i$ (à travers l'augmentation de E_i^*) conduit à diminuer $\bar{x}_{j\geq i}$ et donc à diminuer $\bar{x}_n$ et le flux de sortie $\bar{\nu}_n$ aussi longtemps que les enzymes satisfont la condition de saturation (2.25).

Flux d'entrée au niveau d'une enzyme intermédiaire

Considérons maintenant qu'il existe un flux d'entrée ν_i au niveau de la i-ème enzyme pour les mêmes conditions de voie métabolique. Ce flux de métabolite x_i peut être importé du milieu via un transporteur, ou bien produit par une autre voie métabolique. La voie métabolique est alors décrite par le système d'équations différentielles suivant :

$$\begin{cases} \dot{x}_1 = \nu_0 - E_1 f_1(x_1, x_n) \\ \dot{x}_2 = E_1 f_1(x_1, x_n) - E_2 f_2(x_2, x_3) \\ \quad \vdots \\ \dot{x}_i = E_{i-1} f_{i-1}(x_{i-1}, x_i) - E_i f_i(x_i, x_{i+1}) + \nu_i \\ \dot{x}_{i+1} = E_i f_i(x_i, x_{i+1}) - E_{i+1} f_{i+1}(x_{i+1}, x_{i+2}) \\ \quad \vdots \\ \dot{x}_n = E_{n-1} f_{n-1}(x_{n-1}, x_n) - E_n f_n(x_n) \\ \dot{E}_1 = g(x_n) - \mu E_1 \end{cases} \tag{2.26}$$

et est caractérisé par le régime d'équilibre suivant :

Proposition 2.11 *Pour tout $\mu > 0$, $\bar{x}_1 > 0$ et $(E_i > 0)_{i\in 2,\ldots,n}$ il existe un unique régime d'équilibre pour le système (2.26) tel que :*

$$\frac{g(\bar{x}_n)}{\mu} f_1(\bar{x}_1, \bar{x}_n) + \nu_i = E_n f_n(\bar{x}_n), \tag{2.27}$$

si et seulement si

$$\begin{cases} E_j f_j(\bar{x}_j, \bar{x}_{j+1}) + \nu_i = E_n f_n(\bar{x}_n) \text{ pour } 2 \leq j < i \\ E_j f_j(\bar{x}_j, \bar{x}_{j+1}) = E_n f_n(\bar{x}_n) \text{ pour } i \leq j < n. \end{cases} \tag{2.28}$$

Contrairement au cas d'un flux de sortie, si le flux d'entrée ν_i augmente, alors le flux de sortie $\bar{\nu}_n$ augmente. La condition de saturation (2.28) à partir de la i-ème enzyme est donc plus forte.

De plus, on a $E_n f_n(\bar{x}_n) - \nu_i = \dfrac{g(\bar{x}_n)}{\mu} f_1(\bar{x}_1, \bar{x}_n)$. Si le flux ν_i augmente, alors $\dfrac{g(\bar{x}_n)}{\mu} f_1(\bar{x}_1, \bar{x}_n)$ diminue. Par conséquent, le flux issu de la voie de synthèse contrôlée par x_n est substitué par le flux d'entrée extérieur ν_i pour satisfaire la demande de flux finale fixée par $\mathbb{E}_n$.

2.4 Vers une notion de module élémentaire

Comme nous l'avons montré dans la section précédente, la structure de contrôle par le produit final possède de bonnes propriétés en régime permanent du point de vue

entrée/sortie $(x_1, \nu_0) \longrightarrow (x_n, \nu_n)$ et caractérise simplement l'équilibre pour des configurations biologiques réalistes nous permettant de représenter les caractéristiques des voies métaboliques. En outre, cette structure possède une bonne propriété de découplage vis-à-vis des co-métabolites, même s'il s'agit de métabolites dit hubs (voir la section 1.3.1 page 22 du chapitre précédent). L'ensemble de ces éléments s'étend pour la seconde structure de contrôle par le produit initial renforçant ainsi les notions introduites précédemment pour la caractérisation du régime permanent et posant les bases d'un cadre formel théorique pour l'analyse des réseaux métaboliques.

2.4.1 Structure de contrôle par le produit initial (SCPI)

La structure de contrôle par le produit initial (SCPI) sera ici étudiée, mais de façon moins approfondie que la structure de contrôle par le produit final. L'objectif est uniquement de déterminer les conditions d'existence du régime d'équilibre ainsi que le comportement qualitatif du point d'équilibre pour pouvoir d'une part étudier les règles de connexions entre modules et d'autre part confronter nos prédictions avec les données expérimentales du prochain chapitre.

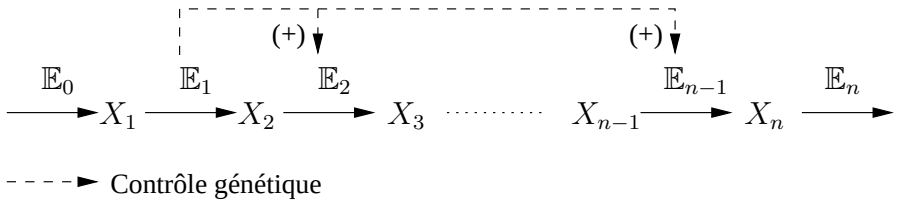


FIG. 2.8 – Structure de contrôle par le produit initial

Nous considérons la voie métabolique linéaire représentée sur la figure 2.8 composée de n métabolites $(X_1, \ldots, X_n)$ et $n-1$ enzymes irréversibles $(\mathbb{E}_1, \ldots, \mathbb{E}_{n-1})$ organisées en opéron, dont la synthèse est contrôlée par le produit initial par l'intermédiaire d'un facteur de transcription par exemple. La synthèse des enzymes est induite lorsque la concentration du métabolite initial augmente. Le comportement de cette voie métabolique obéit au système d'équations différentielles suivant :

$$\begin{cases} \dot{x}_1 = \nu_0 - E_1 f_1(x_1) \\ \vdots \\ \dot{x}_i = E_{i-1} f_{i-1}(x_{i-1}) - E_i f_i(x_i) \\ \vdots \\ \dot{x}_n(t) = E_{n-1} f_{n-1}(x_{n-1}) - E_n f_n(x_n) \\ \dot{E}_1 = g(x_1) - \mu E_1, \\ E_i = \alpha_i E_1 \quad pour \quad i = 2, \ldots, n-1 \end{cases} \tag{2.29}$$

où les f_i ont les mêmes caractéristiques que dans la section précédente (voir la section 2.3) et la fonction g est une fonction positive, strictement croissante et continue de x_1 et telle

que

$$\forall x_1 > 0, \quad g(x_1) > 0, \quad \text{et} \quad g(0) = 0$$

On suppose de plus qu'il existe $P_{max} > 0$ tel que $\lim_{x \to +\infty} g(x) = P_{max}$.

Proposition 2.12 *Pour tout $\mu > 0$, $\bar{x}_1 > 0$ et $E_n > 0$, il existe un unique point d'équilibre $(\bar{x}_2, \ldots, \bar{x}_n)$ et $(\bar{E}_1, \ldots, \bar{E}_{n-1})$ au système (2.29) tel que*

$$\begin{cases} \bar{E}_1 = \dfrac{g(\bar{x}_1)}{\mu} \\ \nu_0 = \dfrac{g(\bar{x}_1)}{\mu} f_1(\bar{x}_1) \\ \bar{x}_i = f_i^{-1}\left(\frac{\alpha_{i-1}}{\alpha_i} f_{i-1}(\bar{x}_{i-1})\right) \quad \textit{pour } i = \{2, \ldots, n\} \end{cases} \tag{2.30}$$

si et seulement si pour tout $i \in \{2, \ldots, n-1\}$, on a

$$M_1 < \alpha_i M_i. \tag{2.31}$$

De plus, $\bar{x}_i = \bar{x}_i(\nu_0)$ pour $i = 1, \ldots, n$ sont des fonctions strictement croissantes de ν_0.

Démonstration

La démonstration de cette proposition est directe en écrivant le régime d'équilibre, qui par définition correspond à $\nu_0 = \dfrac{g(\bar{x}_1)}{\mu} f_1(\bar{x}_1)$ et en utilisant la monotonicité des différentes fonctions. L'existence du régime d'équilibre est obtenu si et seulement si les enzymes de la voie métabolique ne sont pas saturées. Cela signifie que la capacité maximale de chaque enzyme doit être supérieure à ν_0.

■

Ainsi, lorsque $\bar{x}_1$ augmente, le flux ν_0 et les concentrations des métabolites en aval augmentent. Toutes les propriétés obtenues dans le cas du produit final peuvent être écrites pour le cas du produit initial. Les démonstrations de tous ces résultats ne présentent aucune difficulté technique et se déduisent aisément des démonstrations précédentes. Nous rappelons juste ci-dessous les conséquences des hypothèses de l'organisation des gènes en opéron et de la réversibilité des enzymes sur cette structure de contrôle.

Remarque 2.6 *L'hypothèse de l'opéron peut être relaxée sans problème en modifiant la condition de saturation pour les enzymes $(\mathbb{E}_i)_{i \in \{2,\ldots,n\}}$ par $\nu_0 < M_i E_i$.*
Comme dans le cas de la structure de contrôle par le produit final, l'hypothèse de réversibilité des enzymes change les propriétés du régime d'équilibre. Dans le cas des enzymes réversibles, il s'agit également d'écrire le i-ème métabolite en fonction des $i-1$ précédents. En suivant la démonstration de la proposition 2.8 page 50, on peut écrire qu'il existe $\bar{x}_2^*$ tel que pour tout $\bar{x}_2 \in [0, \bar{x}_2^*]$, on a $\bar{x}_2 = H_2(\bar{x}_1)$ où H_2 est une fonction strictement croissante en $\bar{x}_1$, grâce à la stricte monotonicité des fonctions f_1 et g et ainsi de suite pour tous les métabolites intermédiaires, à savoir $\bar{x}_i = H_i(\bar{x}_1)$ pour $i \in \{2, \ldots, n\}$. On en déduit donc que les concentrations des métabolites $\bar{x}_2, \ldots, \bar{x}_n$ augmentent quand $\bar{x}_1$ augmente. En écrivant $\nu_0 = E_n f_n(\bar{x}_n) = E_n f_n(H_n(\bar{x}_1, \nu_0))$, on en déduit $\bar{x}_1$.

Le flux maximal capable de passer dans la voie métabolique dépend aussi du premier métabolite de la façon suivante.

Proposition 2.13 *Pour tout $\mu > 0$ le flux d'entrée est borné et a comme valeur maximale pour $\bar{x}_1 > 0$,*

$$\nu_{0,max} \triangleq \frac{P_{max}}{\mu} M_1$$

si cette valeur est compatible avec les conditions de saturation des autres enzymes de la voie métabolique.

Comparaison du fonctionnement des deux structures de contrôle

Les deux structures de contrôle par le produit final et initial possèdent des caractéristiques communes, obtenues sur la base qu'aucune enzyme de la voie métabolique ne sature :

- le régime d'équilibre est déterminé d'après les caractéristiques de la première enzyme et de son contrôle génétique ;
- la capacité maximale de la voie métabolique est limitée ;
- les co-métabolites intervenant sur les enzymes intermédiaires n'ont aucun impact sur le flux entrée/sortie et sur le contrôle génétique ;
- la présence d'une enzyme irréversible empêche la propagation directe de l'information véhiculée par les concentrations des métabolites de l'aval vers l'amont.

Il existe toutefois des différences notables. La structure de contrôle par le produit final est intrinsèquement pilotée par la demande de flux aval à travers $\bar{x}_n$, contrairement à la structure de contrôle par le produit initial qui est pilotée par le flux amont à travers $\bar{x}_1$. De plus, la caractéristique f_n de l'enzyme $\mathbb{E}_n$ n'a aucune incidence sur l'existence du régime d'équilibre de la structure de contrôle par le produit initial, pourvu qu'elle soit monotone. La fonction f_n peut être strictement croissante ou décroissante en x_n. Autrement dit, une voie métabolique pilotée par une telle structure de contrôle ne peut pas s'adapter à une demande de flux final ν_n.

Sur la base du modèle que nous avons développé pour *Bacillus subtilis* (voir la section 2.1.2), nous avons identifié cette structure de contrôle par le produit initial dans

- toutes les voies de dégradation dont la régulation génétique a été caractérisée expérimentalement : les voies de dégradation des carbohydrates, des acides aminés, des purines et des acides gras.
- quelques voies de biosynthèse, à savoir celle des purines et des acides gras.

Pour toutes ces voies métaboliques, la première enzyme (ou la seconde) est irréversible, ce qui empêche la propagation de l'information de l'aval vers l'amont de la voie métabolique. La concentration du métabolite final en régime permanent $\bar{x}_n$ n'a aucun impact sur le régime d'équilibre déterminé par $\bar{x}_1$.
Ceci est très surprenant pour les cas des voies de biosynthèse pilotées par le produit initial telles que la synthèse des purines ou des acides gras. Les synthèses des purines ou des acides gras produisent des précurseurs métaboliques essentiels à la synthèse de l'ADN ou

de la membrane. Ces précurseurs correspondent typiquement aux métabolites X_n. On imagine mal dans ce cas que des variations de demande en X_n n'aient aucun impact sur le flux en amont. Dans le cas de la voie de synthèse des purines qui a été particulièrement étudiée dans la littérature et pour différents organismes [236, 174, 187], la SCPI est en fait associée à une inhibition enzymatique du métabolite final X_n sur l'enzyme $\mathbb{E}_0$ produisant X_1. Schématiquement, on obtient la structure de contrôle étendue de la figure 2.9.

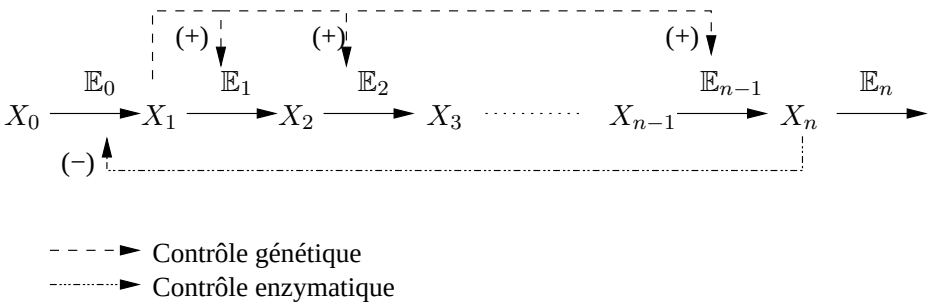


FIG. 2.9 – La structure de contrôle par le produit initial associée à l'inhibition enzymatique de $\mathbb{E}_0$ par le produit final x_n

Le flux ν_0 du système (2.29) s'écrit maintenant :

$$\nu_0 = E_0 f_0(x_0, x_n) \tag{2.32}$$

avec $E_0 > 0$ fixé et f_0 vérifie les caractéristiques d'une enzyme irréversible inhibée par un métabolite. f_0 est une fonction strictement décroissante en x_n, et croissante en x_0.

On obtient alors la proposition suivante dans le cas où toutes les enzymes sont irréversibles et les gènes en opéron :

Proposition 2.14 *Pour tout $\mu > 0$, $E_0 > 0$, $E_n > 0$ et $x_0 > 0$, il existe un unique point d'équilibre $(\bar{x}_1, \ldots, \bar{x}_n)$ et $(\bar{E}_1, \ldots, \bar{E}_{n-1})$ au système (2.29) augmenté de l'équation (2.32) tel que*

$$\begin{cases} E_0 f_0(x_0, \bar{x}_n) = E_n f_n(\bar{x}_n) \\ \nu_0 = E_0 f_0(x_0, \bar{x}_n) \\ \nu_n = E_n f_n(\bar{x}_n) \end{cases} \tag{2.33}$$

si et seulement si $\nu_0 < E_1 M_1$ et pour tout $i \in \{2, \ldots, n-1\}$, on a

$$\nu_0 < \alpha_i E_1 M_i. \tag{2.34}$$

De plus, on a :

- *$\bar{x}_n = \bar{x}_n(x_0)$ est une fonction strictement croissante de x_0.*
- *$\bar{x}_n = \bar{x}_n(E_n)$ est une fonction strictement décroissante de E_n et $\bar{x}_i = \bar{x}_i(E_n)$ pour $i = 1, \ldots, n-1$ sont des fonctions strictement croissantes de E_n.*

– $\bar{x}_i = \bar{x}_i(E_0)$ *pour* $i = 1, \ldots, n$ *sont des fonctions strictement croissantes de* E_0.

Démonstration
L'élément clé de la démonstration de cette proposition est donné par la relation $E_0 f_0(x_0, \bar{x}_n) = E_n f_n(\bar{x}_n)$ où $x_0 > 0$ fixé, $\bar{x}_n$ est déterminé en utilisant les propriétés de monotonicité des fonctions f_0 et f_n. Les flux d'entrée et de sortie ν_0 et ν_n sont alors fixés. Les régimes d'équilibre des autres composantes du module sont déterminés en utilisant les mêmes arguments obtenus pour la proposition 2.12.

■

De façon remarquable, la concentration $\bar{x}_n$ du métabolite final à l'équilibre est complètement déterminée par les concentrations et les caractéristiques des enzymes en entrée et en sortie du module $\mathbb{E}_0$ et $\mathbb{E}_n$. A E_0 et x_0 fixés, le flux en entrée ν_0 est déterminé directement en fonction du flux demandé ν_n à travers $\bar{x}_n$. Les autres composantes du module $(\bar{E}_i, \bar{x}_i)$ pour $i \in \{1, \ldots, n-1\}$ sont déterminées pour satisfaire la demande de flux. Le régime d'équilibre du module dépend uniquement des caractéristiques de la première enzyme $\mathbb{E}_1$.

2.4.2 Impact d'un régulateur global

Les propriétés mathématiques associées aux deux types de modules ayant été caractérisées dans les sections précédentes, il convient maintenant d'étudier les moyens de coordination entre ces modules, à savoir les régulations globales et la connexion directe entre deux modules. Il s'agit de montrer comment les propriétés mathématiques intrinsèques aux modules sont modifiées lorsque les régulations globales sont prises en compte ou lorsque deux modules sont interconnectés. Dans un premier temps, nous nous intéressons à l'impact d'un régulateur global sur la structure de contrôle par le produit final. Les résultats pour l'autre structure se déduisent aisément.
Considérons que la synthèse de la première enzyme de la voie métabolique est également contrôlée par un régulateur dit global. Son effet est intégré dans la dynamique de $\mathbb{E}_1$ de la façon suivante :

$$\dot{E}_1(t) = g(x_n(t), q(t)) - \mu E_1(t)$$

où $q(t)$ est l'effet du régulateur global. Ce paramètre peut également représenter tout acteur susceptible d'impacter la synthèse de l'enzyme $\mathbb{E}_1$.
En supposant que le facteur externe atteigne son propre régime d'équilibre $\bar{q}$, on déduit des résultats précédents que le régulateur global modifie la relation entre la concentration du produit final $\bar{x}_n$, la demande en flux et la concentration d'enzyme :

$$f_1(\bar{x}_1, \bar{x}_n) g(\bar{x}_n, \bar{q}) = \mu E_n f_n(\bar{x}_n).$$

Tant que les enzymes intermédiaires ne saturent pas, ce régulateur global modifie le régime d'équilibre au niveau de

– la concentration des enzymes

$$\bar{E}_1 = \frac{g(\bar{x}_n, \bar{q})}{\mu} \quad \text{et} \quad \bar{E}_i = \alpha_i \frac{g(\bar{x}_n, \bar{q})}{\mu} \text{ pour } i \in \{2, \ldots, n-1\}$$

si les gènes sont en opéron.

– la concentration du produit final

$$f_1(\bar{x}_1, \bar{x}_n) g(\bar{x}_n, \bar{q}) = \mu E_n f_n(\bar{x}_n).$$

– la capacité maximale de la voie métabolique

$$\bar{\nu}_{n,max}(\bar{q}) \triangleq \frac{g(0, \bar{q})}{\mu} f_1(\bar{x}_1, 0)$$

– les concentrations des métabolites intermédiaires

$$\bar{x}_i = f_i^{-1}\left(\frac{\mu E_n f_n(\bar{x}_n)}{\alpha_i g(\bar{x}_n, \bar{q})}\right)$$

Un régulateur global modifie ainsi la capacité maximale de la voie métabolique en modulant directement la concentration des enzymes de la voie. Si l'effet du régulateur global q est de diminuer la fonction de production g pour une valeur donnée de x_n alors pour une concentration croissante de q, le flux de demande ν_n va diminuer. En revanche, si son effet est d'augmenter g, alors pour des concentrations croissantes de q, la demande de flux augmente également.

Remarque 2.7 *La régulation globale correspond à une régulation de type feedforward en Automatique.*

2.4.3 Interconnexion entre les modules

Dans cette section, nous nous intéressons aux conditions d'existence et d'unicité d'un régime d'équilibre structurel suivant les configurations de connexion entre deux modules. Nous avons déjà étudié l'impact de points de branchement à l'intérieur d'un module dans la section 2.3.3. Nous allons donc nous concentrer sur la connexion entre deux modules en série et en parallèle. Nous allons tout d'abord introduire un résultat générique sur l'interconnexion de modules en série, avant de donner les règles de connexion des modules sous forme de tableaux récapitulatifs. On étudiera plus en détails dans la prochaine section les principaux exemples de connexion de modules que nous avons identifiés grâce au modèle de connaissance de *Bacillus subtilis*. Nous utiliserons les abréviations suivantes pour des raisons de lisibilité dans les différents schémas :

– SCPF pour structure de contrôle par le produit final ;
– SCPI pour structure de contrôle par le produit initial.

De plus, nous considérons que pour un module (i) les gènes ne sont pas en opéron et que (ii) seule la première enzyme est irréversible.

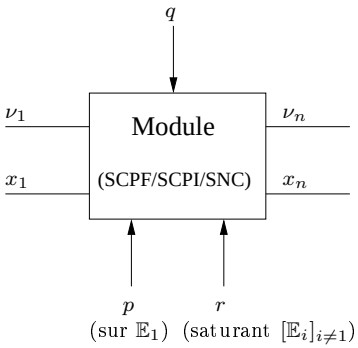


FIG. 2.10 – La représentation entrée/sortie d'un module

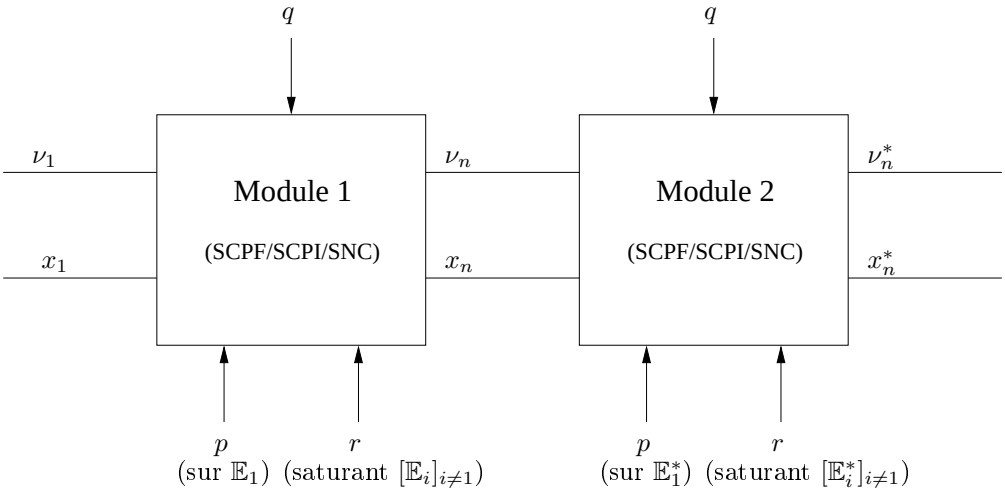


FIG. 2.11 – Connexion en série de deux modules

a. La représentation entrée/sortie d'un module

Sur la base des résultats présentés au cours de ce chapitre, nous avons caractérisé deux types de module (la SCPF et la SCPI) définis par une relation monotone en entrée/sortie. En parallèle de ces deux modules, nous introduisons également un troisième module composé d'enzymes non contrôlées par un métabolite de la voie aussi bien au niveau génétique qu'au niveau enzymatique, que nous appellerons SNC pour Structure Non Contrôlée. La caractéristique entrée/sortie de ce dernier module à l'équilibre, obtenue sous l'hypothèse qu'aucune enzyme du module ne sature et que la première enzyme est irréversible s'écrit :

$$E_1 f_1(\bar{x}_1) = E_n f_n(\bar{x}_n).$$

Il vient que le régime d'équilibre est fixé par la concentration du métabolite initial $\bar{x}_1$ et les caractéristiques des enzymes f_1 et f_n. La figure 2.10 donne une représentation entrée/sortie générique pour les trois types de modules.

En régime d'équilibre, un module est donc caractérisé par sa relation entrée/sortie réduite à la première enzyme, dont l'existence est conditionnée au fait qu'aucune enzyme ne sature. Dans la suite de cette section, nous supposerons que cette condition d'existence est toujours satisfaite. On rappelle ci-dessous les relations entrées/sorties obtenues pour les trois types de module :

- pour le module SCPF : $\nu_1 = E_n f_n(H(\bar{x}_1)) \triangleq H_{pf}(\bar{x}_1)$ (conséquences des corollaires 2.4 et 2.5 étendus dans le cas où seule la première enzyme est irréversible et les gènes ne sont pas en opéron), où H_{pf} est une fonction strictement croissante de son argument ;
- pour le module SCPI : $\nu_1 = H_{pi}(\bar{x}_1)$ définie dans la proposition 2.12 étendue pour les mêmes conditions, où H_{pi} est une fonction strictement croissante de son argument ;
- pour le module SNC : $\nu_1 = E_1 f_1(\bar{x}_1) \triangleq H_{snc}(\bar{x}_1)$, où H_{snc} est une fonction strictement croissante de son argument.

On peut en déduire les conséquences suivantes pour les connexions des modules en série (voir la figure 2.11) :

- la connexion en série de modules SCPF mène à un système ayant un régime permanent unique. Pour un module SCPF numéro i donné, tous les modules SCPF situés en aval sont réduits à travers la relation strictement croissante $\nu_{i+1} = H_{pf}(\bar{x}_{i+1})$. Le régime permanent obtenu est celui d'une enzyme fictive $\hat{\mathbb{E}}_{i+1}$ telle que $\nu_{i+1} = \hat{E}_{i+1} f_{i+1}(\bar{x}_{i+1})$, où f_{i+1} est strictement croissante.
- la connexion en série de modules SCPI mène à un système ayant un régime permanent unique. Pour un module SCPI numéro i donné, tous les modules SCPI situés en amont sont réduits à travers la relation strictement croissante $\nu_{i-1} = H_{pi}(\bar{x}_{i-1})$. Le régime permanent obtenu est celui d'une enzyme fictive $\hat{\mathbb{E}}_{i-1}$ telle que $\nu_{i-1} = \hat{E}_{i-1} f_{i-1}(\bar{x}_{i-1})$, où f_{i-1} est strictement croissante.

b. Les règles d'interconnexion des modules

Les règles d'interconnexion des modules peuvent se déduire aisément à partir des démonstrations des propositions 2.8, 2.12, 2.14 données respectivement pages 50, 59 et 61 pour des connexions de modules en série (voir la figure 2.11) ou en parallèle (voir la figure 2.12) sous l'hypothèse qu'aucune enzyme ne sature. Les tableaux 2.3 et 2.4 donnent les règles d'interconnexion en série et en parallèle entre les modules respectivement. Plus précisément, ces tableaux indiquent pour les différentes connexions s'il existe directement un régime équilibre non nul, et comment évoluent x_n, x_n^*, E_1 et E_1^* et x_1, x_1^* suivant les configurations pour des variations en ν_1, ν_n et ν_n^*. L'existence du régime d'équilibre se déduit toujours à travers la monotonicité des différentes fonctions. Il peut arriver, dans le cas par exemple de la connexion SNC/SCPI/SCPF en série associée à une boucle de rétroaction positive (voir le tableau 2.3), qu'on ne puisse pas conclure directement sur l'existence de l'équilibre. Typiquement, on obtient une condition nécessaire d'intersection entre deux fonctions croissantes comme suit :

$$E_0 f_0(x_0, \bar{x}_n) = E_n f_n(\bar{x}_n)$$

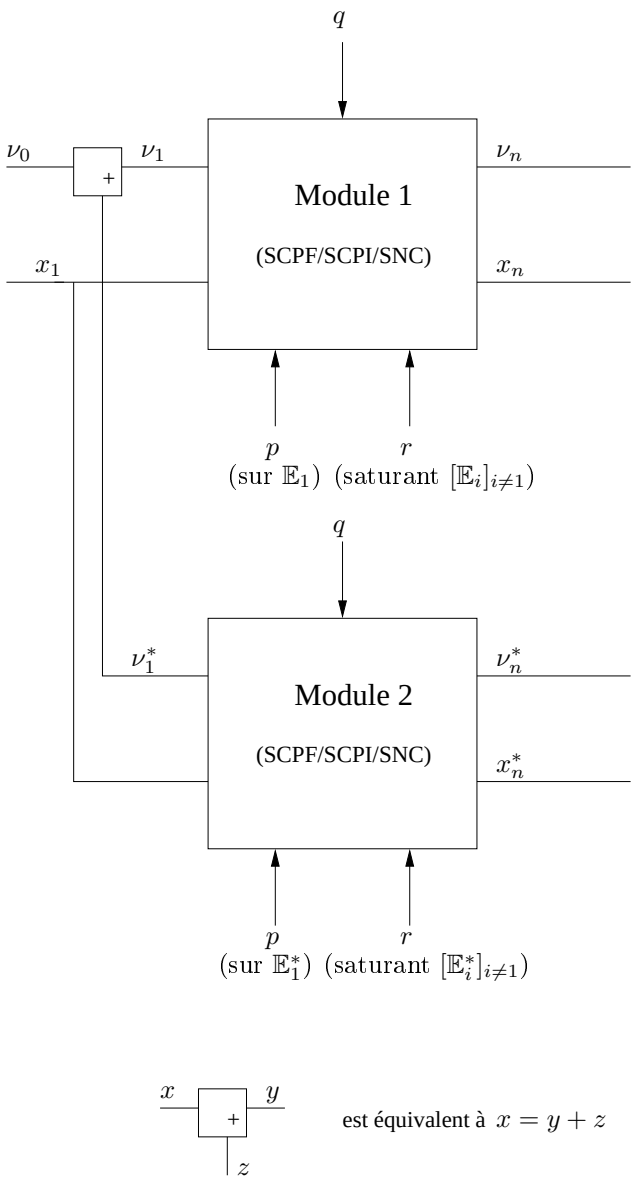
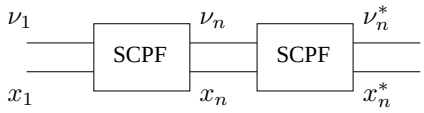


FIG. 2.12 – Connexion en parallèle de deux modules

où f_0 et f_n sont toutes les deux des fonctions strictement croissantes de $\bar{x}_n$. Par convention, on dira alors que la condition d'existence de l'équilibre est conditionnée, contrairement aux cas où l'existence de l'équilibre est obtenue structurellement. On utilisera les notations suivantes pour des raisons de lisibilité : f_c pour fonction strictement croissante et f_d pour fonction strictement décroissante dans les deux tableaux.



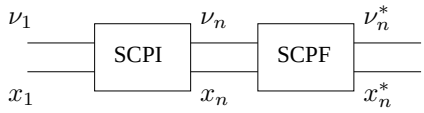
Existence d'un équilibre non nul

Adaptation possible pour des variations de ν_n, ν_n^*

$x_n = f_d(\nu_n)$, $E_1 = f_c(\nu_n)$, $x_n^* = f_d(\nu_n)$, $E_1^* = f_c(\nu_n)$,

$x_n = f_d(\nu_n^*)$, $E_1 = f_c(\nu_n^*)$, $x_n^* = f_d(\nu_n^*)$, $E_1^* = f_c(\nu_n^*)$,

$x_n = f_c(\nu_1)$, $E_1 = f_d(\nu_1)$, $x_n^* = f_c(\nu_1)$, $E_1^* = f_d(\nu_1)$.



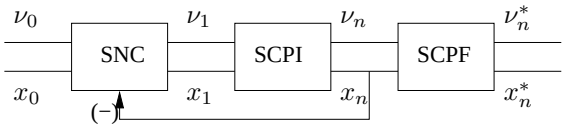
Existence d'un équilibre non nul

Adaptation SCPI impossible pour des variations de ν_n, ν_n^*

$x_n = f_d(\nu_n)$, $x_n^* = f_d(\nu_n)$, $E_1^* = f_c(\nu_n)$,

$x_n = f_d(\nu_n^*)$, $x_n^* = f_d(\nu_n^*)$, $E_1^* = f_c(\nu_n^*)$,

$x_n = f_c(\nu_1)$, $E_1 = f_c(\nu_1)$, $x_n^* = f_c(\nu_1)$, $E_1^* = f_d(\nu_1)$.



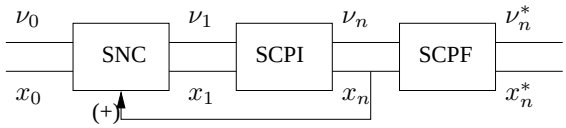
Existence d'un équilibre non nul

Adaptation possible pour des variations de ν_n, ν_n^*

$x_n = f_d(\nu_n)$, $x_1 = f_c(\nu_n)$, $E_1 = f_c(\nu_n)$, $x_n^* = f_d(\nu_n)$, $E_1^* = f_c(\nu_n)$,

$x_n = f_d(\nu_n^*)$, $x_1 = f_c(\nu_n^*)$, $E_1 = f_c(\nu_n^*)$, $x_n^* = f_d(\nu_n^*)$, $E_1^* = f_c(\nu_n^*)$,

$x_n = f_c(\nu_0)$, $E_1 = f_c(\nu_0)$, $x_n^* = f_c(\nu_0)$, $E_1^* = f_d(\nu_0)$.



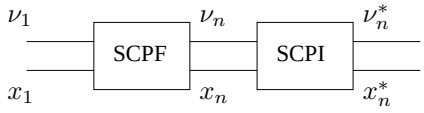
Équilibre conditionné à f_0, f_n, f_n^*

Adaptation SCPI contraire pour des variations de ν_n, ν_n^*

$x_n = f_d(\nu_n)$, $x_1 = f_d(\nu_n)$, $E_1 = f_d(\nu_n)$, $x_n^* = f_d(\nu_n)$, $E_1^* = f_c(\nu_n)$,

$x_n = f_d(\nu_n^*)$, $x_1 = f_d(\nu_n^*)$, $E_1 = f_d(\nu_n^*)$, $x_n^* = f_d(\nu_n^*)$, $E_1^* = f_c(\nu_n^*)$,

$x_n = f_c(\nu_0)$, $x_1 = f_c(\nu_0)$, $E_1 = f_c(\nu_0)$, $x_n^* = f_c(\nu_0)$, $E_1^* = f_d(\nu_0)$.



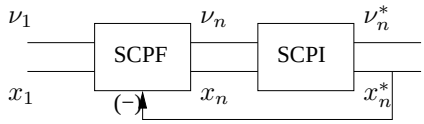
Existence d'un équilibre non nul

Adaptation SCPI impossible pour des variations de ν_n^*

$x_n = f_d(\nu_n)$, $E_1 = f_c(\nu_n)$, $x_n^* = f_d(\nu_n)$, $E_1^* = f_d(\nu_n)$,

$x_n^* = f_d(\nu_n^*)$,

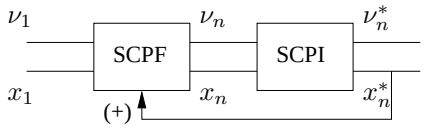
$x_n = f_c(\nu_1)$, $E_1 = f_d(\nu_1)$, $x_n^* = f_c(\nu_1)$, $E_1^* = f_c(\nu_1)$.



Pas d'équilibre

Adaptation SCPI contraire pour des variations de ν_n, ν_n^*

$x_n = f_d(\nu_n)$, $E_1 = f_c(\nu_n)$, $x_n^* = f_d(\nu_n)$, $E_1^* = f_d(\nu_n)$.



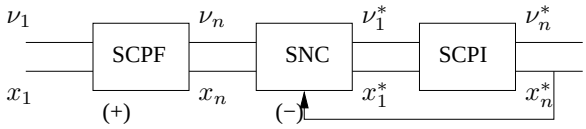
Existence d'un équilibre non nul

Adaptation possible pour des variations de ν_n, ν_n^*

$x_n = f_d(\nu_n)$, $E_1 = f_c(\nu_n)$, $x_n^* = f_d(\nu_n)$, $E_1^* = f_c(\nu_n)$,

$x_n = f_d(\nu_n^*)$, $E_1 = f_c(\nu_n^*)$, $x_n^* = f_d(\nu_n^*)$, $E_1^* = f_c(\nu_n^*)$,

$x_n = f_c(\nu_1)$, $E_1 = f_c(\nu_1)$, $x_n^* = f_c(\nu_1)$, $E_1^* = f_d(\nu_1)$.



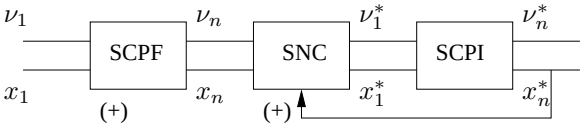
Existence d'un équilibre non nul

Adaptation possible pour des variations de ν_n, ν_n^*

$x_n = f_d(\nu_n)$, $E_1 = f_d(\nu_n)$, $x_1^* = f_c(\nu_n)$, $x_n^* = f_d(\nu_n)$, $E_1^* = f_c(\nu_n)$,

$x_n = f_d(\nu_n^*)$, $E_1 = f_d(\nu_n^*)$, $x_n^* = f_d(\nu_n^*)$, $x_1^* = f_c(\nu_n^*)$, $E_1^* = f_c(\nu_n^*)$,

$x_n = f_c(\nu_1)$, $E_1 = f_d(\nu_1)$, $x_n^* = f_c(\nu_1)$, $x_1^* = f_d(\nu_1)$, $E_1^* = f_d(\nu_1)$.



Équilibre conditionné à f_0^* et f_n^*

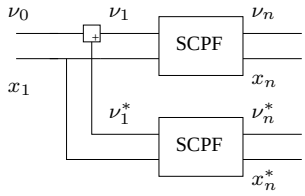
Adaptation SCPI impossible pour des variations de ν_n, ν_n^*

$x_n = f_d(\nu_n)$, $E_1 = f_c(\nu_n)$, $x_n^* = f_d(\nu_n)$, $x_1^* = f_d(\nu_n)$, $E_1^* = f_d(\nu_n)$,

$x_n = f_d(\nu_n^*)$, $E_1 = f_c(\nu_n^*)$, $x_n^* = f_d(\nu_n^*)$, $x_1^* = f_d(\nu_n^*)$, $E_1^* = f_d(\nu_n^*)$,

$x_n = f_c(\nu_1)$, $E_1 = f_d(\nu_1)$, $x_n^* = f_c(\nu_1)$, $x_1^* = f_d(\nu_1)$, $E_1^* = f_d(\nu_1)$.

TAB. 2.3 – Règles d'interconnexion en série entre plusieurs modules. On suppose que (i) le flux d'entrée ν_1 (ou ν_0) est capable de maintenir le métabolite initial x_1 (ou x_0) fixé ; (ii) les enzymes des modules ne saturent pas.



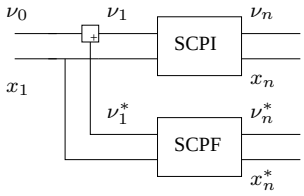
Existence d'un équilibre non nul

Adaptation possible pour des variations de ν_n, ν_n^*

$x_n = f_d(\nu_n)$, $E_1 = f_c(\nu_n)$, pas d'effet sur ν_n^*,

$x_n^* = f_d(\nu_n^*)$, $E_1^* = f_c(\nu_n^*)$, pas d'effet sur ν_n,

$x_n = f_c(\nu_0)$, $E_1 = f_d(\nu_0)$, $x_n^* = f_c(\nu_0)$, $E_1^* = f_d(\nu_0)$.



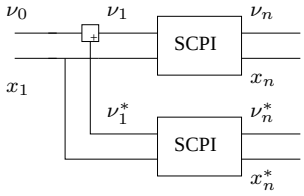
Existence d'un équilibre non nul

Adaptation SCPI impossible pour des variations de ν_n

$x_n = f_d(\nu_n)$,

$x_n^* = f_d(\nu_n^*)$, $E_1^* = f_c(\nu_n^*)$, pas d'effet sur ν_n,

$x_n = f_c(\nu_0)$, $E_1 = f_c(\nu_0)$, $x_n^* = f_c(\nu_0)$, $E_1^* = f_d(\nu_0)$.

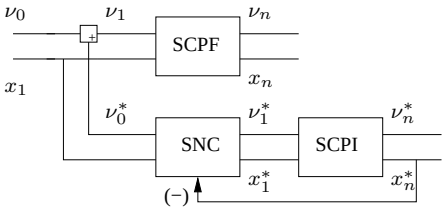


Existence d'un équilibre non nul

Adaptation SCPI impossible pour des variations de ν_n^*

$x_n = f_d(\nu_n)$, $x_n^* = f_d(\nu_n^*)$,

$x_n = f_c(\nu_0)$, $E_1 = f_c(\nu_0)$, $x_n^* = f_c(\nu_0)$, $E_1^* = f_c(\nu_0)$.



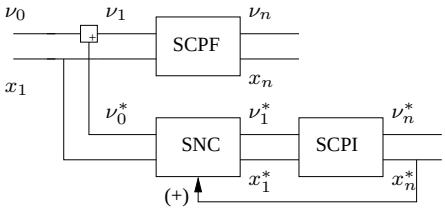
Existence d'un équilibre non nul

Adaptation possible pour des variations de ν_n, ν_n^*

$x_n = f_d(\nu_n)$, $E_1 = f_c(\nu_n)$, pas d'effet sur ν_n^*,

$x_n^* = f_d(\nu_n^*)$, $x_1^* = f_c(\nu_n^*)$, $E_1^* = f_c(\nu_n^*)$, pas d'effet sur ν_n,

$x_n = f_c(\nu_0)$, $E_1 = f_d(\nu_0)$, $x_n^* = f_c(\nu_0)$, $x_1^* = f_c(\nu_0)$, $E_1^* = f_c(\nu_0)$.



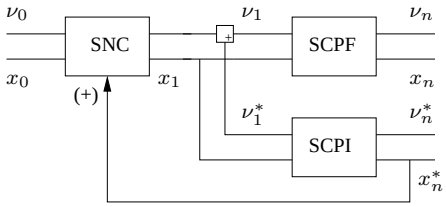
Équilibre conditionné à f_0^*, f_n^*

Adaptation SCPI impossible pour des variations de ν_n^*

$x_n = f_d(\nu_n)$, $E_1 = f_c(\nu_n)$, pas d'effet sur ν_n^*,

$x_n^* = f_d(\nu_n^*)$, $x_1^* = f_d(\nu_n^*)$, $E_1^* = f_d(\nu_n^*)$, pas d'effet sur ν_n,

$x_n = f_c(\nu_0)$, $E_1 = f_d(\nu_0)$, $x_n^* = f_c(\nu_0)$, $x_1^* = f_c(\nu_0)$, $E_1^* = f_c(\nu_0)$.



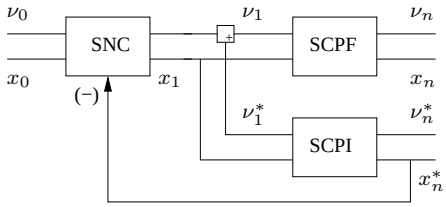
Équilibre conditionné à f_0, f_n^*

Adaptation SCPI impossible pour des variations de ν_n, ν_n^*

$x_n = f_d(\nu_n)$, $E_1 = f_c(\nu_n)$, $x_1 = f_d(\nu_n)$, $x_n^* = f_d(\nu_n)$, $E_1^* = f_d(\nu_n)$,

$x_n = f_d(\nu_n^*)$, $E_1 = f_c(\nu_n^*)$, $x_1 = f_d(\nu_n^*)$, $x_n^* = f_d(\nu_n^*)$, $E_1^* = f_d(\nu_n^*)$,

$x_1 = f_c(\nu_0)$, $x_n = f_c(\nu_0)$, $E_1 = f_d(\nu_0)$, $x_n^* = f_c(\nu_0)$, $E_1^* = f_c(\nu_0)$.



Existence d'un équilibre non nul

Adaptation possible pour des variations de ν_n, ν_n^*

$x_n = f_d(\nu_n)$, $E_1 = f_c(\nu_n)$, $x_1 = f_d(\nu_n)$, $x_n^* = f_d(\nu_n)$, $E_1^* = f_d(\nu_n)$,

$x_n = f_c(\nu_n^*)$, $E_1 = f_d(\nu_n^*)$, $x_1 = f_c(\nu_n^*)$, $x_n^* = f_d(\nu_n^*)$, $E_1^* = f_c(\nu_n^*)$,

$x_1 = f_c(\nu_0)$, $x_n = f_c(\nu_0)$, $E_1 = f_d(\nu_0)$, $x_n^* = f_c(\nu_0)$, $E_1^* = f_c(\nu_0)$.

TAB. 2.4 – Règles d'interconnexion en parallèle entre plusieurs modules. On suppose que (i) ν_0 est capable de maintenir le métabolite initial x_1 (ou x_0) fixé ; (ii) les enzymes des modules ne saturent pas.

2.4.4 Exemples de connexions présentes dans le réseau métabolique de *Bacillus subtilis*

Les tableaux 2.3 et 2.4 nous donnent les règles d'interconnexion des modules sans se soucier de leur présence effective dans un organisme. Dans cette section, nous allons détailler les régimes d'équilibre correspondant aux différentes combinaisons de modules que nous avons identifiées chez *Bacillus subtilis* sur la base du modèle de connaissance reconstruit, et qui sont donc pertinentes d'un point de vue biologique. Ces différentes combinaisons sont également présentes dans le contrôle du réseau métabolique d'*Escherichia coli* ce qui renforce encore l'intérêt de les étudier de façon plus approfondie. Ces connexions correspondent à des connexions en série de modules SCPF/SCPF, SCPI/SCPF, SCPF/SCPI, SCPI/SCPI avec un point de branchement supplémentaire entre les deux modules représentant une demande de métabolite intermédiaire. On étudiera également les cas où la SCPI correspond à une SCPI associée à une inhibition enzymatique de la première enzyme par le métabolite X_n introduite dans la proposition 2.14 (et appelée SCPIb par la suite). Pour ces différentes connexions, nous préciserons leur localisation dans le réseau métabolique, et nous donnerons le régime d'équilibre associé ainsi que l'évolution des différentes composantes du module sous l'hypothèse que (i) les gènes ne sont pas en opéron et que (ii) seule la première enzyme du module est irréversible.

a. Connexion SCPF/SCPF

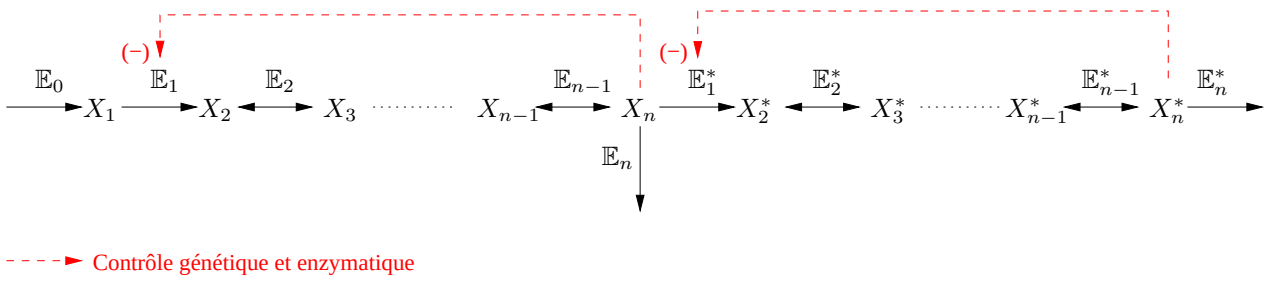


FIG. 2.13 – Connexion entre deux modules SCPF

Nous avons identifié dans le modèle de connaissance de *B. subtilis* le motif représenté sur la figure 2.13 correspondant à la mise en série de modules SCPF avec un point de branchement intermédiaire entre :

- les voies de synthèse du glutamate et de la glutamine [39, 167, 243, 241, 65] ;
- les voies de synthèse du glutamate et de la proline [39, 167, 38] ;
- les voies de synthèse du S-adénosyl-méthionine et de la cystéine [10, 166].

Chez la bactérie *E. coli* pour laquelle les régulations des voies des acides aminés ont été particulièrement bien identifiées, on retrouve également cette connexion entre les voies de synthèse de la thréonine et de l'isoleucine [107].

La combinaison de ces deux modules obéit au système d'équations différentielles suivant :

$$\begin{cases} \dot{x}_1 = \nu_0 - E_1 f_1(x_1, x_n) \\ \dot{x}_2 = E_1 f_1(x_1, x_n) - E_2 f_2(x_2, x_3) \\ \quad \vdots \\ \dot{x}_n = E_{n-1} f_{n-1}(x_{n-1}, x_n) - E_n f_n(x_n) - E_1^* f_1^*(x_n, x_n^*) \\ \dot{x^*}_2 = E_1^* f_1^*(x_n, x_n^*) - E_2^* f_2^*(x_2^*, x_3^*) \\ \quad \vdots \\ \dot{x^*}_n = E_{n-1}^* f_{n-1}^*(x_{n-1}^*, x_n^*) - E_n^* f_n^*(x_n^*) \\ \dot{E}_1 = g_1(x_n) - \mu E_1 \\ \dot{E^*}_1 = g_1^*(x_n^*) - \mu E_1^* \end{cases} \tag{2.35}$$

où g_1 et g_1^* sont des fonctions strictement décroissantes de leurs arguments. On obtient alors le résultat suivant pour le régime d'équilibre :

Proposition 2.15 *Pour tout $\mu > 0$, $E_i > 0$, $E_i^* > 0$ pour $i \in \{2, \cdots, n\}$ et $\bar{x}_1 > 0$, il existe un unique régime d'équilibre $(\bar{E}_1, \bar{E}_1^*)$ au système (2.35) et $(\bar{x}_2, \ldots, \bar{x}_n, \bar{x}_2^*, \ldots, \bar{x}_n^*)$ pour (2.35) tel que*

$$\begin{cases} \bar{E}_1 = \dfrac{g_1(\bar{x}_n)}{\mu} \\ \bar{E}_1^* = \dfrac{g_1^*(\bar{x}_n^*)}{\mu}, \\ f_1(\bar{x}_1, \bar{x}_n) g_1(\bar{x}_n) = \mu(E_n f_n(\bar{x}_n) + E_n^* f_n^*(\bar{x}_n^*)) \\ f_1^*(\bar{x}_n, \bar{x}_n^*) g_1^*(\bar{x}_n^*) = \mu E_n^* f_n^*(\bar{x}_n^*) \end{cases} \tag{2.36}$$

si et seulement si $(\bar{x}_2, \cdots, \bar{x}_{n-1}, \bar{x}_2^*, \cdots, \bar{x}_{n-1}^*)$ existe tel que

$$\begin{cases} E_i f_i(\bar{x}_i, \bar{x}_{i+1}) = E_n f_n(\bar{x}_n) + E_n^* f_n^*(\bar{x}_n^*) \\ E_i^* f_i^*(\bar{x}_i^*, \bar{x}_{i+1}^*) = E_n^* f_n^*(\bar{x}_n^*) \end{cases} \tag{2.37}$$

pour $i \in \{2, \ldots, n-1\}$.

Démonstration

La démonstration de cette proposition est immédiate en utilisant les arguments de la démonstration de la proposition 2.8. Sous l'hypothèse que les enzymes du deuxième module SCPF ne saturent pas (cf. la deuxième équation de la condition (2.37)), on déduit que $\bar{x}_n^*$ s'écrit comme une fonction croissante H_n^* de $\bar{x}_n$ ($\bar{x}_n^* = H_n^*(\bar{x}_n)$) en utilisant la monotonicité des fonctions de part et d'autre de l'équation $f_1^*(\bar{x}_n, \bar{x}_n^*) g_1^*(\bar{x}_n^*) = \mu E_n^* f_n^*(\bar{x}_n^*)$. Puis, sous l'hypothèse que les enzymes du premier module SCPF ne saturent pas (cf. la première équation de la condition (2.37)), on utilise le même argument pour conclure sur l'unicité de $\bar{x}_n$ en utilisant l'équation $f_1(\bar{x}_1, \bar{x}_n) g_1(\bar{x}_n) = \mu(E_n f_n(\bar{x}_n) + E_n^* f_n^*(H_n^*(\bar{x}_n)))$.

■

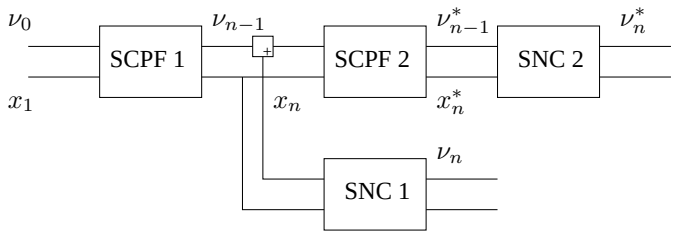


FIG. 2.14 – Représentation modulaire de la figure 2.13

La figure 2.13 peut également se représenter sous forme de modules connectés en reprenant les notations des tableaux 2.3 et 2.4 (voir la figure 2.14). La démonstration de la proposition 2.15 se réduit alors à vérifier que la connexion SNC 1, SCPF 2 et SNC 2 est équivalente à une fonction strictement croissante de $\bar{x}_n$, ce qui permet d'obtenir la démonstration alternative suivante de la proposition 2.15.

Démonstration

De la théorie des modules, nous pouvons déduire sous l'hypothèse que les enzymes des modules SNC 1, SCPF 2 et SNC 2 ne saturent pas, que :

- la connexion entre les modules SCPF 2 et SNC 2 est décrite en régime permanent par la relation entrée/sortie $\bar{\nu}^*_n = H_{pf_2}(\bar{x}_n)$ où H_{pf_2} une fonction strictement croissante de $\bar{x}_n$ (conséquence des corollaires 2.4 et 2.5 étendus dans le cas où seule la première enzyme est irréversible et les gènes ne sont pas en opéron) ;
- le module SNC 1 est décrit par la relation entrée/sortie $\bar{\nu}_n = H_{snc_1}(\bar{x}_n)$ où H_{snc_1} une fonction strictement croissante de $\bar{x}_n$.

La connexion en parallèle des modules SCPF 2, SNC 2 et SNC 1 est équivalente à :

$$\nu_{n-1} = H_{snc_1}(\bar{x}_n) + H_{pf_2}(\bar{x}_n) \triangleq H_{con}(\bar{x}_n)$$

où H_{con} est une fonction strictement croissante de $\bar{x}_n$. On en déduit finalement sous l'hypothèse que les enzymes du module SCPF 1 ne sature pas, que le module SCPF 1 est décrit par la relation entrée/sortie $\bar{\nu}_0 = H_{pf_1}(\bar{x}_1)$ où H_{pf_1} est une fonction strictement croissante de $\bar{x}_1$.

■

De plus, pour des variations des flux de demande $\bar{\nu}_n = E_n f_n(\bar{x}_n)$ et $\bar{\nu}^*_n = E^*_n f_n(\bar{x}^*_n)$ causées respectivement par une augmentation des concentrations des enzymes $\mathbb{E}_n$ et $\mathbb{E}^*_n$, on peut déduire le corollaire suivant.

Corollaire 2.16 *Supposons que les hypothèses de la proposition 2.15 sont satisfaites. Alors, on a :*

- $\bar{x}_n = \bar{x}_n(E^*_n)$ *et* $\bar{x}^*_n = \bar{x}^*_n(E^*_n)$ *sont des fonctions strictement décroissantes de* E^*_n *et donc de* $\bar{\nu}^*_n$ *;*
- $\bar{x}_n = \bar{x}_n(E_n)$ *et* $\bar{x}^*_n = \bar{x}^*_n(E_n)$ *sont des fonctions strictement décroissantes de* E_n *et donc de* $\bar{\nu}_n$*.*

Ces résultats indiquent que la voie métabolique est pilotée par les demandes de flux $\bar{\nu}_n$ (associée à $\mathbb{E}_n$) et $\bar{\nu}_n^*$ (associée à $\mathbb{E}_n^*$), et qu'il existe toujours un régime d'équilibre sous l'hypothèse qu'aucune enzyme ne sature. Une augmentation du flux de demande $\bar{\nu}_n^*$ par une augmentation de E_n^* fait diminuer la concentration $\bar{x}_n^*$. Comme $\bar{x}_n^*$ diminue, la concentration E_1^* augmente. Donc le flux de demande pour $\bar{x}_n$ augmente, et finalement $\bar{x}_n$ décroît. La diminution de $\bar{x}_n$ entraîne l'augmentation de la concentration E_1.
De même, une augmentation du flux de demande $\bar{\nu}_n$ par une augmentation de E_n fait diminuer la concentration $\bar{x}_n$. Donc E_1 augmente. De plus, comme $\bar{x}_n^*$ est une fonction croissante de $\bar{x}_n$, $\bar{x}_n^*$ diminue et la concentration E_1^* augmente pour satisfaire la demande de flux ν_n^*.

Pour les autres cas de connexion identifiés chez *Bacillus subtilis*, nous donnerons uniquement les propositions détaillant les régimes d'équilibre ainsi que l'évolution des composantes principales des modules en omettant les démonstrations. Elles ne présentent aucune difficulté technique et elles se déduisent aisément en écrivant la représentation de la connexion sous forme de modules entrées/sorties. Il va de soi que nous avons toutefois rédigé précisément l'ensemble de ces démonstrations.

b. Connexion SCPI/SCPF

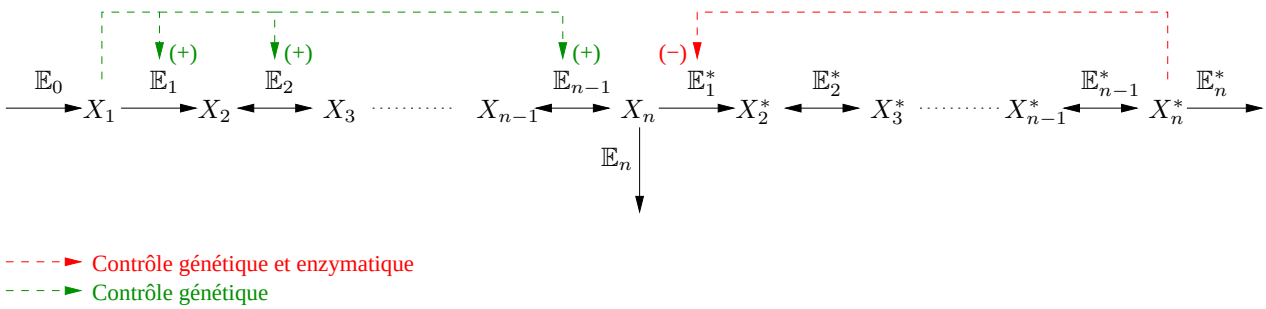


FIG. 2.15 – Connexion entre un module SCPI et un module SCPF

Sur la base de l'actuel réseau de régulation connu de *Bacillus subtilis*, nous n'avons pas identifié de connexion de type SCPI/SCPF représentée sur la figure 2.15. En revanche, la connexion SCPIb/SCPF se retrouve au niveau de la glycolyse et la voie de synthèse de l'isoleucine, la leucine et la valine [228, 47, 224, 207, 208]. Nous allons détailler ci-dessous les résultats pour ces deux types de connexion.

La combinaison de ces deux modules obéit au système d'équations différentielles sui-

vant :

$$\begin{cases} \dot{x}_1 = \nu_0 - E_1 f_1(x_1) \\ \dot{x}_2 = E_1 f_1(x_1) - E_2 f_2(x_2, x_3) \\ \quad \vdots \\ \dot{x}_n = E_{n-1} f_{n-1}(x_{n-1}, x_n) - E_n f_n(x_n) - E_1^* f_1^*(x_n, x_n^*) \\ \dot{x^*}_2 = E_1^* f_1^*(x_n, x_n^*) - E_2^* f_2^*(x_2^*, x_3^*) \\ \quad \vdots \\ \dot{x^*}_n = E_{n-1}^* f_{n-1}^*(x_{n-1}^*, x_n^*) - E_n^* f_n^*(x_n^*) \\ \dot{E}_1 = g_1(x_1) - \mu E_1 \\ \dot{E^*}_1 = g_1^*(x_n^*) - \mu E_1^* \end{cases} \tag{2.38}$$

où g_1 (respectivement g_1^*) est une fonction strictement croissante (respectivement décroissante) de son argument. On obtient alors le résultat suivant pour le régime d'équilibre :

Proposition 2.17 *Pour tout $\mu > 0$, $E_i > 0$, $E_i^* > 0$ pour $i \in \{2, \cdots, n\}$ et $\bar{x}_1 > 0$, il existe un unique régime d'équilibre $(\bar{E}_1, \bar{E}_1^*)$ et $(\bar{x}_2, \ldots, \bar{x}_n, \bar{x}_2^*, \ldots, \bar{x}_n^*)$ pour le système (2.38) tel que*

$$\begin{cases} \bar{E}_1 = \dfrac{g_1(\bar{x}_1)}{\mu} \\ \bar{E}_1^* = \dfrac{g_1^*(\bar{x}_n^*)}{\mu} \\ f_1(\bar{x}_1) g_1(\bar{x}_1) = \mu(E_n f_n(\bar{x}_n) + E_n^* f_n^*(\bar{x}_n^*)) \\ f_1^*(\bar{x}_n, \bar{x}_n^*) g_1^*(\bar{x}_n^*) = \mu E_n^* f_n^*(\bar{x}_n^*) \\ \nu_0 = \dfrac{g_1(\bar{x}_1)}{\mu} f_1(\bar{x}_1) \end{cases} \tag{2.39}$$

si et seulement si $(\bar{x}_2, \cdots, \bar{x}_{n-1}, \bar{x}_2^*, \cdots, \bar{x}_{n-1}^*)$ existe tel que

$$\begin{cases} \nu_0 = E_i f_i(\bar{x}_i, \bar{x}_{i+1}) \\ E_i^* f_i^*(\bar{x}_i^*, \bar{x}_{i+1}^*) = E_n^* f_n^*(\bar{x}_n^*) \end{cases} \tag{2.40}$$

pour $i \in \{2, \ldots, n-1\}$. De plus, on a

- *$\bar{x}_n = \bar{x}_n(E_n^*)$ et $\bar{x}_n^* = \bar{x}_n^*(E_n^*)$ sont des fonctions strictement décroissantes de E_n^* et donc de $\bar{\nu}_n^*$;*
- *$\bar{x}_n = \bar{x}_n(E_n)$ et $\bar{x}_n^* = \bar{x}_n^*(E_n)$ sont des fonctions strictement décroissantes de E_n et donc de $\bar{\nu}_n$;*
- *$\bar{x}_n = \bar{x}_n(\nu_0)$ et $\bar{x}_n^* = \bar{x}_n^*(\nu_0)$ sont des fonctions strictement croissantes de ν_0.*

Dans cette configuration, les produits finaux des deux modules $\bar{x}_n^*$ et $\bar{x}_n$ diminuent avec les demandes de flux $\bar{\nu}_n$ et $\bar{\nu}_n^*$. La capacité du module SCPF s'adapte en fonction de la demande. Si $\bar{\nu}_n^*$ augmente, alors $\bar{x}_n^*$ diminue et E_1^* augmente pour satisfaire cette demande. Par conséquent $\bar{x}_n$ diminue. L'irréversibilité de l'enzyme E_1 bloque la remontée d'information jusqu'à $\bar{x}_1$ pour adapter le module SCPI au flux $\bar{\nu}_n^*$ demandé. Il en va de

même pour une demande en flux $\bar{\nu}_n$.

Dans le cas de la SCPI modifiée (SCPIb) où le flux ν_0 est une fonction décroissante de $\bar{x}_n$ (équation 2.32 page 61) :

$$\nu_0 = E_0 f_0(x_0, \bar{x}_n)$$

alors on obtient le résultat suivant.

Proposition 2.18 *Pour tout $\mu > 0$, $x_0 > 0$, $E_0 > 0$, $E_i > 0$, $E_i^* > 0$ pour $i \in \{2, \cdots, n\}$, il existe un unique régime d'équilibre $(\bar{E}_1, \bar{E}_1^*)$ et $(\bar{x}_1, \ldots, \bar{x}_n, \bar{x}_2^*, \ldots, \bar{x}_n^*)$ au système (2.38) augmenté de l'équation (2.32) tel que*

$$\begin{cases} \bar{E}_1 = \dfrac{g_1(\bar{x}_1)}{\mu} \\ \bar{E}_1^* = \dfrac{g_1^*(\bar{x}_n^*)}{\mu} \\ E_0 f_0(x_0, \bar{x}_n) = E_n f_n(\bar{x}_n) + E_n^* f_n^*(\bar{x}_n^*) \\ \mu E_0 f_0(x_0, \bar{x}_n) = f_1(\bar{x}_1) g_1(\bar{x}_1) \\ f_1^*(\bar{x}_n, \bar{x}_n^*) g_1^*(\bar{x}_n^*) = \mu E_n^* f_n^*(\bar{x}_n^*) \end{cases} \tag{2.41}$$

si et seulement si $(\bar{x}_2, \cdots, \bar{x}_{n-1}, \bar{x}_2^*, \cdots, \bar{x}_{n-1}^*)$ existe tel que

$$E_0 f_0(x_0, \bar{x}_n) = E_i f_i(\bar{x}_i, \bar{x}_{i+1})$$
$$E_i^* f_i^*(\bar{x}_i^*, \bar{x}_{i+1}^*) = E_n^* f_n^*(\bar{x}_n^*)$$

pour $i \in \{2, \ldots, n-1\}$. De plus, on a :

- *$\bar{x}_n^* = \bar{x}_n^*(E_n^*)$, $\bar{x}_n = \bar{x}_n(E_n^*)$ sont des fonctions strictement décroissantes de E_n^* et $\bar{x}_1 = \bar{x}_1(E_n^*)$, $\nu_0 = \nu_0(E_n^*)$ sont des fonctions strictement croissantes de E_n^* et donc de ν_n^* ;*
- *$\bar{x}_n^* = \bar{x}_n^*(E_n)$, $\bar{x}_n = \bar{x}_n(E_n)$ sont des fonctions strictement décroissantes de E_n et $\bar{x}_1 = \bar{x}_1(E_n)$, $\nu_0 = \nu_0(E_n)$ sont des fonctions strictement croissantes de E_n et donc de ν_n.*

Sous cette configuration, les composants des deux modules s'adaptent à des variations de flux de sortie ν_n et ν_n^*.

c. Connexion SCPF/SCPI

Sur la base du modèle de connaissance de *Bacillus subtilis* reconstruit, nous avons constaté que la connexion SCPF/SCPI représentée sur la figure 2.16 correspond à la connexion habituelle entre les voies de synthèse et de dégradation des acides aminés comme l'arginine [146, 70]. Les voies de synthèse et de dégradation de l'histidine sont aussi probablement connectées à travers une SCPF/SCPI [240, 242, 61, 62]. Malheureusement, le contrôle de la voie de synthèse de l'histidine est inconnu. Chez *E. coli*, cette

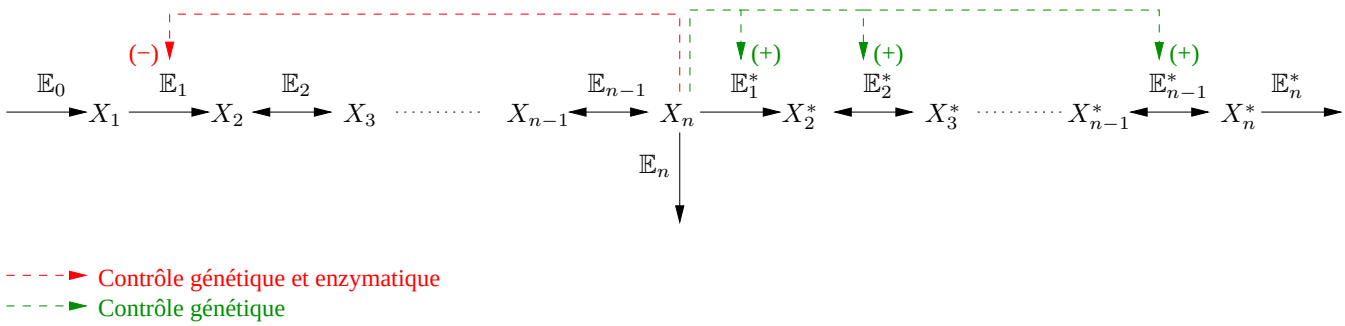


FIG. 2.16 – Connexion entre un module SCPF et un module SCPI

voie métabolique est régulée par l'histidine à travers l'ARNt chargé correspondant, donc par une structure SCPF [107]. En revanche, un régulateur global est souvent présent sur le module SCPI pour éviter que les voies de synthèse et de dégradation soient actives en même temps [36, 242, 61, 62].

La combinaison de ces deux modules obéit au système d'équations différentielles suivant :

$$\begin{cases} \dot{x}_1 = \nu_0 - E_1 f_1(x_1, x_n) \\ \dot{x}_2 = E_1 f_1(x_1, x_n) - E_2 f_2(x_2, x_3) \\ \quad \vdots \\ \dot{x}_n = E_{n-1} f_{n-1}(x_{n-1}, x_n) - E_n f_n(x_n) - E_1^* f_1^*(x_n) \\ \dot{x^*}_2 = E_1^* f_1^*(x_n) - E_2^* f_2^*(x_2^*, x_3^*) \\ \quad \vdots \\ \dot{x^*}_n = E_{n-1}^* f_{n-1}^*(x_{n-1}^*, x_n^*) - E_n^* f_n^*(x_n^*) \\ \dot{E}_1 = g_1(x_n) - \mu E_1 \\ \dot{E^*}_1 = g_1^*(x_n) - \mu E_1^* \end{cases} \tag{2.42}$$

où g_1 est une fonction strictement décroissante de x_n, g_1^* est une fonction strictement croissante de x_n. On obtient alors le résultat suivant pour le régime d'équilibre :

Proposition 2.19 *Pour tout $\mu > 0$, $E_i > 0$, $E_i^* > 0$ pour $i \in \{2, \cdots, n\}$ et $\bar{x}_1 > 0$, il existe un unique régime d'équilibre $(\bar{E}_1, \bar{E}_1^*)$ et $(\bar{x}_2, \ldots, \bar{x}_n, \bar{x}_2^*, \ldots, \bar{x}_n^*)$ au système (2.42) tel que*

$$\begin{cases} \bar{E}_1 = \dfrac{g_1(\bar{x}_n)}{\mu} \\ \bar{E}_1^* = \dfrac{g_1^*(\bar{x}_n)}{\mu} \\ f_1(\bar{x}_1, \bar{x}_n) g_1(\bar{x}_n) = \mu(E_n f_n(\bar{x}_n) + E_n^* f_n^*(\bar{x}_n^*)) \\ f_1^*(\bar{x}_n) g_1^*(\bar{x}_n) = \mu E_n^* f_n^*(\bar{x}_n^*) \end{cases} \tag{2.43}$$

si et seulement si $(\bar{x}_2, \cdots, \bar{x}_{n-1}, \bar{x}_2^*, \cdots, \bar{x}_{n-1}^*)$ *existe tel que*

$$\begin{cases} E_i f_i(\bar{x}_i, \bar{x}_{i+1}) = E_n f_n(\bar{x}_n) + E_n^* f_n^*(\bar{x}_n^*) \\ E_i^* f_i^*(\bar{x}_i^*, \bar{x}_{i+1}^*) = E_n^* f_n^*(\bar{x}_n^*) \end{cases} \tag{2.44}$$

pour $i \in \{2, \ldots, n-1\}$. *De plus, on a :*

- $\bar{x}_n^* = \bar{x}_n^*(E_n^*)$ *est une fonction strictement décroissante de* E_n^* *et donc de* $\bar{\nu}_n^*$ *;*
- $\bar{x}_n = \bar{x}_n(E_n)$ *et* $\bar{x}_n^* = \bar{x}_n^*(E_n)$ *sont des fonctions strictement décroissantes de* E_n *et donc de* $\bar{\nu}_n$.

Face à une demande de flux $\bar{\nu}_n$ associée à une augmentation de la concentration E_n, $\bar{x}_n$ diminue. La concentration E_1 augmente en conséquence pour satisfaire la demande de flux $\bar{\nu}_n$. En revanche, comme $\bar{x}_n$ diminue, la concentration E_1^* diminue également, ce qui provoque la diminution de $\bar{x}_n^*$ et donc du flux de sortie $\bar{\nu}_n^*$. Une variation de flux $\bar{\nu}_n^*$ n'a aucun impact sur $\bar{x}_n$ (et donc sur le module SCPF) à cause de l'irréversibilité de l'enzyme $\mathbb{E}_1^*$. Il vient alors que pour une augmentation en E_n^*, toutes les concentrations de métabolites $(\bar{x}_2^*, \cdots, \bar{x}_n^*)$ diminuent jusqu'à 0. Le flux $\bar{\nu}_n^*$ devient nul.

d. Connexion SCPI/SCPI

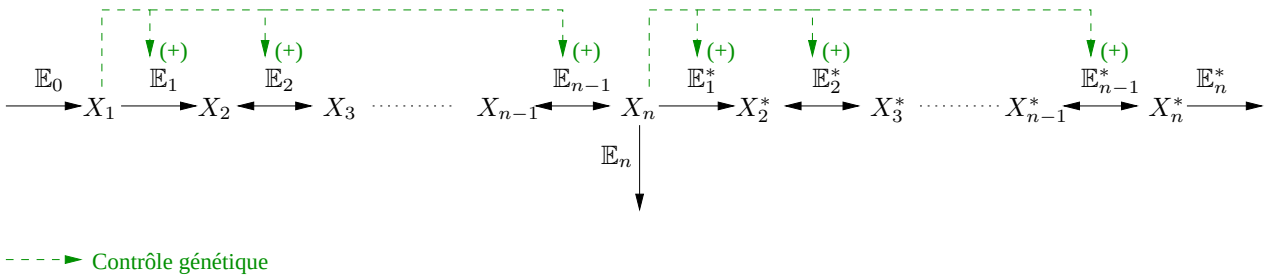


FIG. 2.17 – Connexion entre un module SCPI et un module SCPI

Enfin, nous avons identifié la présence de la connexion SCPI/SCPI représentée sur la figure 2.17 au niveau des voies de synthèse et de dégradation des acides gras [195, 144]. Ici encore un régulateur global (CcpA) empêche la dégradation des acides gras en conditions glycolytiques [144, 148]. De même, la connexion SCPI/SCPIb représentée sur la figure 2.18 se retrouve au niveau de la connexion entre les voies d'assimilation des différents carbohydrates et la glycolyse (voir toutes les références dans le modèle associées au transport et à la dégradation des carbohydrates et [47]). La connexion SCPIb/SCPIb n'est pas présente dans l'actuel réseau de régulation de *B. subtilis*. On ne peut en revanche pas conclure de manière systématique sur la présence de tel type de connexion par rapport à une autre à cause de l'incomplétude du réseau de régulation.

La combinaison de ces deux modules obéit au système d'équations différentielles suivant :

$$\begin{cases} \dot{x}_1 = \nu_0 - E_1 f_1(x_1) \\ \dot{x}_2 = E_1 f_1(x_1) - E_2 f_2(x_2, x_3) \\ \quad \vdots \\ \dot{x}_n = E_{n-1} f_{n-1}(x_{n-1}, x_n) - E_n f_n(x_n) - E_1^* f_1^*(x_n) \\ \dot{x^*}_2 = E_1^* f_1^*(x_n) - E_2^* f_2^*(x_2^*, x_3^*) \\ \quad \vdots \\ \dot{x^*}_n = E_{n-1}^* f_{n-1}^*(x_{n-1}^*, x_n^*) - E_n^* f_n^*(x_n^*) \\ \dot{E}_1 = g_1(x_1) - \mu E_1 \\ \dot{E^*}_1 = g_1^*(x_n) - \mu E_1^* \end{cases} \tag{2.45}$$

où g_1 et g_1^* sont des fonctions strictement croissantes de x_1 et x_n respectivement. On obtient alors le résultat suivant pour le régime d'équilibre :

Proposition 2.20 *Pour tout $\mu > 0$, $E_i > 0$, $E_i^* > 0$ pour $i \in \{2, \cdots, n\}$ et $\bar{x}_1 > 0$, il existe un unique régime d'équilibre $(\bar{E}_1, \bar{E}_1^*)$ et $(\bar{x}_2, \ldots, \bar{x}_n, \bar{x}_2^*, \ldots, \bar{x}_n^*)$ au système (2.45) tel que*

$$\begin{cases} \bar{E}_1 = \dfrac{g_1(\bar{x}_1)}{\mu} \\ \bar{E}_1^* = \dfrac{g_1^*(\bar{x}_n)}{\mu} \\ \nu_0 = \dfrac{g_1(\bar{x}_1)}{\mu} f_1(\bar{x}_1) \\ \bar{x}_i = H_i(H_{i-1}(\ldots H_1(\nu_0))) \\ \bar{x}_2^* = H_1^*(\nu_0) \\ \bar{x}_i^* = H_i^*(H_{i-1}^*(\ldots H_1^*(\nu_0))) \end{cases} \tag{2.46}$$

si et seulement si $(\bar{x}_2, \cdots, \bar{x}_{n-1}, \bar{x}_2^*, \cdots, \bar{x}_{n-1}^*)$ existe tel que

$$\begin{cases} \nu_0 = E_i f_i(\bar{x}_i, \bar{x}_{i+1}) \\ \nu_0 - E_n f_n(\bar{x}_n) = E_i^* f_i^*(\bar{x}_i^*, \bar{x}_{i+1}^*) \end{cases}$$

pour $i \in \{2, \ldots, n-1\}$. De plus, on a :

- *$\bar{x}_n = \bar{x}_n(\nu_0)$ et $\bar{x}_n^* = \bar{x}_n^*(\nu_0)$ sont des fonctions strictement croissantes de ν_0 ;*
- *$\bar{x}_n^* = \bar{x}_n^*(E_n)$ est une fonction strictement décroissante de E_n et donc de $\bar{\nu}_n$.*

Les deux modules sont incapables de compenser des variations de demande de flux $\bar{\nu}_n$ et $\bar{\nu}_n^*$.

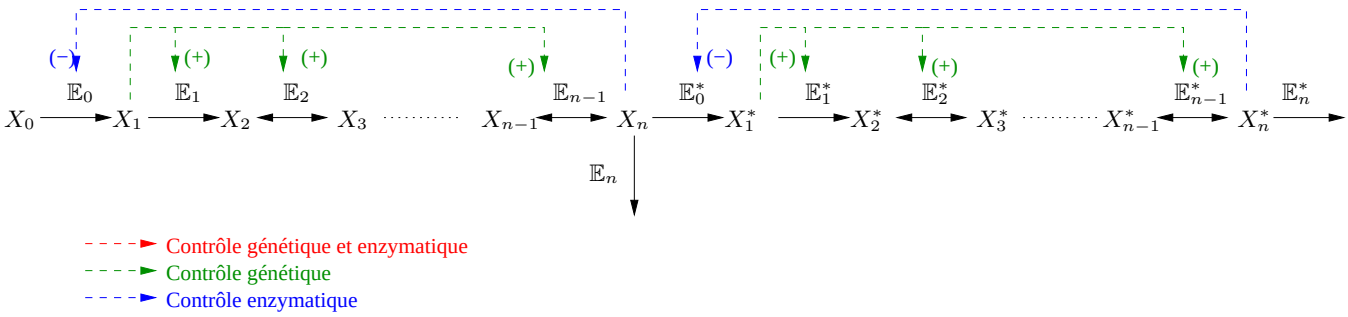


FIG. 2.18 – Connexion entre un module SCPIb et un module SCPIb

La combinaison de deux modules SCPIb/SCPIb représentée sur la figure 2.18 obéit au système d'équations différentielles suivant :

$$\begin{cases} \dot{x}_1 = E_0 f_0(x_0, x_n) - E_1 f_1(x_1) \\ \dot{x}_2 = E_1 f_1(x_1) - E_2 f_2(x_2, x_3) \\ \quad \vdots \\ \dot{x}_n = E_{n-1} f_{n-1}(x_{n-1}, x_n) - E_n f_n(x_n) - E_0^* f_0^*(x_n, x_n^*) \\ \dot{x^*}_1 = E_0^* f_0^*(x_n, x_n^*) - E_1^* f_1^*(x_1^*) \\ \dot{x^*}_2 = E_1^* f_1^*(x_1^*) - E_2^* f_2^*(x_2^*, x_3^*) \\ \quad \vdots \\ \dot{x^*}_n = E_{n-1}^* f_{n-1}^*(x_{n-1}^*, x_n^*) - E_n^* f_n^*(x_n^*) \\ \dot{E}_1 = g_1(x_1) - \mu E_1 \\ \dot{E^*}_1 = g_1^*(x_1^*) - \mu E_1^* \end{cases} \tag{2.47}$$

où g_1 et g_1^* sont des fonctions strictement croissantes de x_1 et x_1^* respectivement. On obtient alors le résultat suivant pour le régime d'équilibre :

Proposition 2.21 *Pour tout $\mu > 0$, $E_i > 0$, $E_i^* > 0$ pour $i \in \{0, 2, \cdots, n\}$ et $x_0 > 0$, il existe un unique régime d'équilibre $(\bar{E}_1, \bar{E}_1^*)$ et $(\bar{x}_1, \ldots, \bar{x}_n, \bar{x}_1^*, \ldots, \bar{x}_n^*)$ au système (2.47)*

tel que

$$\begin{cases} \bar{E}_1 = \dfrac{g_1(\bar{x}_1)}{\mu} \\ \bar{E}_1^* = \dfrac{g_1^*(\bar{x}_n)}{\mu} \\ E_0 f_0(x_0, \bar{x}_n) = E_n f_n(\bar{x}_n) + E_0^* f_0^*(\bar{x}_n, \bar{x}_n^*) \\ \bar{x}_1 = H_1(\bar{x}_n) \\ \bar{x}_i = H_i(H_{i-1}(\dots H_1(\bar{x}_n))) \\ E_0^* f_0^*(\bar{x}_n, \bar{x}_n^*) = E_n^* f_n^*(\bar{x}_n^*) \\ \bar{x}_1^* = H_1^*(\bar{x}_n^*) \\ \bar{x}_i^* = H_i^*(H_{i-1}^*(\dots H_1^*(\bar{x}_n^*))) \end{cases} \tag{2.48}$$

si et seulement si $(\bar{x}_1, \cdots, \bar{x}_{n-1}, \bar{x}_1^*, \cdots, \bar{x}_{n-1}^*)$ *existe tel que*

$$\begin{cases} E_0 f_0(x_0, \bar{x}_n) = E_i f_i(\bar{x}_i, \bar{x}_{i+1}) \\ E_0^* f_0^*(\bar{x}_n, \bar{x}_n^*) = E_i^* f_i^*(\bar{x}_i^*, \bar{x}_{i+1}^*) \end{cases}$$

pour $i \in \{2, \dots, n-1\}$. *De plus, on a :*

- $\bar{x}_n = \bar{x}_n(E_n)$ *est une fonction strictement décroissante de* E_n *et donc de* ν_n *;*
- $\bar{x}_n^* = \bar{x}_n^*(E_n^*)$ *et* $\bar{x}_n = \bar{x}_n(E_n^*)$ *sont des fonctions strictement décroissantes de* E_n^* *et donc de* $\bar{\nu}_n^*$.

2.4.5 Comparaison avec *Escherichia coli*

Contrairement à *Bacillus subtilis*, nous n'avons pas reconstruit manuellement le réseau de régulation des voies métaboliques pour *Escherichia coli*. Nous avons utilisé la base de données EcoCyc [107] (http ://ecocyc.org/) qui intègre les régulations transcriptionnelles ainsi que les effecteurs métaboliques connus. On dispose bien d'une image du réseau de régulation de *Escherichia coli* où la boucle de rétroaction globale des métabolites sur la transcription des gènes est présente.
Pour les mêmes voies métaboliques que nous avions reconstruites chez *Bacillus subtilis*, nous avons systématiquement vérifié que la structure de la régulation pouvait se mettre sous les formes SCPF, SCPI ou SCPIb. Nous pouvons ainsi en déduire que l'ensemble de nos résultats sont valides pour un organisme aussi phylogénétiquement éloigné de *Bacillus subtilis* qu'est *Escherichia coli*.

2.5 Conclusion

La question posée en introduction de ce chapitre était la suivante : existe-t-il des sous-systèmes autonomes (au sens de l'Automatique) dans le réseau métabolique lorsque les différents niveaux de contrôle sont intégrés et peut-on formellement définir ces sous-systèmes ?

La réponse à cette question en régime permanent est bien évidemment positive au regard des développements théoriques donnés dans les différentes sections. Sur la base d'un modèle de connaissance réaliste d'un réseau métabolique décrivant tous les niveaux de contrôle connus, nous avons identifié deux structures de contrôle dans les voies métaboliques remarquables de par leurs propriétés, grâce à la prise en compte du couplage du réseau métabolique avec le réseau de régulation génétique via les effecteurs métaboliques. De l'étude de ces structures de contrôle en régime permanent, nous avons en particulier montré, sous l'hypothèse qu'aucune enzyme ne sature, que :

- la première enzyme irréversible est critique pour la définition du régime permanent et correspond généralement à la première enzyme de la voie métabolique ;
- les co-facteurs comme l'ATP n'ont aucun impact sur l'équilibre s'ils interviennent après la première enzyme irréversible (pour la structure de contrôle par le produit final) ;
- les régulations globales modifient les propriétés des modules élémentaires en régime permanent. En particulier, les régulations globales permettent d'allumer ou d'éteindre complètement certains modules ou simplement de moduler leur capacité maximale de flux métabolique ;
- les isoenzymes augmentent la capacité maximale de la voie métabolique.

De plus, l'analyse à l'équilibre de ces voies métaboliques aussi longues soient-elles est grandement simplifiée puisque le régime d'équilibre est uniquement déterminé par la première enzyme irréversible. Nous avons ainsi isolé des sous-systèmes bien identifiés du point de vue entrée/sortie, possédant un unique régime d'équilibre et pour lesquels on peut prédire l'évolution qualitative de tous ses composants. En ce sens, ces sous-systèmes correspondent à des **modules élémentaires fonctionnels**.

Le cadre formel associé aux développements théoriques des modules est suffisamment général pour étudier l'existence et l'unicité d'un équilibre structurel des voies métaboliques, et même du réseau métabolique dans son ensemble. Nos travaux sont les premiers de ce genre réalisés à une telle échelle globale, et ce en explorant de manière systématique toutes les configurations biologiques identifiées dans notre modèle. Une perspective intéressante à ce travail serait d'étudier la stabilité des modules élémentaires et de leur interconnexion. L'étude de la stabilité des voies métaboliques est un problème largement ouvert compte tenu de la très grande dimension du système et de la nonlinéarité des équations. Certains résultats ont été obtenus pour le système linéarisé autour d'un voisinage pour une voie métabolique spécifique [227, 9, 6, 7]. Néanmoins l'obtention de résultats sur la stabilité globale du système non linéaire même dans le cas d'un seul module est une question plus que difficile et délicate.

Dans le prochain chapitre, nous allons caractériser ces modules élémentaires dans le réseau métabolique de *Bacillus subtilis* et confronter nos prédictions théoriques concernant l'évolution qualitative des composants des modules à des données expérimentales produites dans le cadre du projet européen BaSysBio.

Chapitre 3

Identification et validation des modules chez *Bacillus subtilis*

D'après le chapitre précédent, le réseau métabolique peut être décomposé en sous-systèmes définis de part leurs propriétés entrées/sorties en régime permanent, que nous avons nommés **modules élémentaires**. Ces modules sont contrôlés par une régulation dite «locale» garantissant l'induction ou l'inhibition des enzymes de la voie métabolique en fonction d'un métabolite clé appartenant à la voie métabolique. Les propriétés théoriques associées à ces modules sont singulières et ont des implications biologiques fortes. Par exemple, le rôle des enzymes irréversibles est critique car elles découpent le réseau métabolique et empêche la propagation de l'information vers l'amont de ces enzymes. Des régulations génétiques et enzymatiques sont alors nécessaires pour assurer la remontée de l'information. En outre, malgré la très forte connectivité des voies métaboliques, ces enzymes irréversibles permettent de découpler le régime permanent du module des co-métabolites et des co-facteurs intervenant en aval de cette enzyme. Ainsi la variation de l'un de ces co-métabolites n'entraîne pas un réajustement global de l'ensemble des modules où ce co-métabolite intervient. Les modules élémentaires sont coordonnés entre eux par des régulateurs globaux qui modifient les propriétés des modules en régime permament. La réponse théorique du module à une variation du métabolite clé est ainsi différente. En particulier, les régulations globales permettent d'allumer ou d'éteindre complètement des modules lorsque ceux-ci ne sont pas nécessaires ou tout simplement de moduler la capacité maximale de flux métabolique du module.

L'objectif de ce chapitre consiste à valider dans une certaine mesure notre découpage en modules élémentaires du réseau métabolique de *Bacillus subtilis* en le confrontant à des données expérimentales de transcriptome et de métabolome obtenues dans le cadre du projet européen BaSysBio. Pour cela, en nous basant sur les résultats théoriques développés au cours du chapitre précédent, nous allons prédire de manière qualitative le comportement de tous les composants d'une voie métabolique régulée localement (flux métaboliques, concentration des ARNm, des protéines, des métabolites) en réponse à des variations du flux demandé ou des métabolites clés. Ces prédictions seront alors comparées aux données expérimentales.

3.1 Décomposition modulaire du réseau métabolique

3.1.1 Les modules élémentaires identifiés chez *B. subtilis*

Les modules élémentaires ont été identifiés en appliquant strictement les définitions des régulations locales et globales sur les voies métaboliques, et résumés sur les tableaux 3.13, 3.14 et 3.15 donnés en annexes 3.A de ce chapitre de thèse page 115.
Dans certains cas, l'identification de certains modules a été délicate car :

- le module ne correspond pas à une voie métabolique «classique». Par exemple, la glycine est produite à partir de la sérine par l'enzyme GlyA, dont la synthèse est réprimée par PurR, le régulateur principal de la synthèse des purines [187]. La glycine étant nécessaire à la synthèse des purines, le gène *glyA* appartient au module de la synthèse des purines, et ne constitue pas un module autonome ;
- les isoenzymes appartiennent généralement à des modules différents. Dans la voie de synthèse de la proline, plusieurs isoenzymes sont présentes : ProB et ProJ d'un part ainsi que ProG, ProH et ProI d'autre part. Un premier module est formé par les gènes *proA*, *proB* et *proI*, qui sont contrôlés par une T-box spécifique de la proline située en amont du gène *proI* et de l'opéron *proBA*. Ce module est requis pour l'anabolisme[1] ; le second module est composé des gènes *proJ et proH* et est induit pendant un stress osmotique ; la régulation de la troisième isoenzyme, ProG, étant inconnue, on ne peut l'assigner à un module ;
- un régulateur local dans une voie métabolique peut correspondre à un régulateur global pour d'autres voies. Le régulateur global CodY dont nous discuterons dans la section suivante est aussi un régulateur local de la voie de synthèse des acides aminés branchés.

Cette définition est également applicable pour d'autres organismes, en particulier chez la bactérie *E. coli* pour laquelle l'identification des modules est possible grâce aux connaissances disponibles dans les bases de données EcoCyc [182] et RegulonDB [181].

3.1.2 Les régulateurs globaux

D'après la figure 2.2 page 35 et le modèle de connaissance de *Bacillus subtilis* développé au chapitre précédent, les voies métaboliques sont principalement sous le contrôle global de trois facteurs de transcription : CcpA, CodY et TnrA. L'activité de ces régulateurs est également modulée par des métabolites directement par accrochage du métabolite sur le facteur de transcription, ou indirectement par l'intermédiaire d'une enzyme :

- **CcpA :** principal régulateur responsable de la répression catabolique chez *Bacillus subtilis*, il gère principalement le transport et le catabolisme des sources de carbone [46], ainsi que la dégradation des acides gras et de certains acides aminés. Pour

[1] Dans le réseau métabolique, l'anabolisme correspond à l'ensemble des voies de biosynthèse des précurseurs métaboliques. Par opposition, le catabolisme correspond à l'ensemble des voies de dégradation des composants cellulaires (acides aminés, carbohydrates, acides gras...) dont le but est de produire de l'énergie sous forme d'ATP, de NADH ou de GTP et aussi des squelettes carbonés élémentaires requis pour l'anabolisme.

être capable de se fixer sur l'ADN, CcpA a besoin d'être en complexe avec la forme phosphorylée sur la sérine 46 de la protéine HPr (HPr-Ser46-P). Le fructose-1,6-biphosphate (FBP) est nécessaire à la formation de HPr-Ser46-P d'une part [173], et stimule la formation du complexe CcpA/HPr-Ser46-P d'autre part [202]. Nous considérons ainsi le FBP comme étant le principal effecteur métabolique de CcpA. Comme le FBP n'est pas un métabolite intermédiaire des voies directement contrôlées par CcpA, CcpA est bien un régulateur global au sens de notre définition des voies métaboliques qu'il régule ;

- **TnrA :** il s'agit du principal régulateur responsable de la gestion des sources d'azote [239]. Lorsque la glutamine est en excès, celle-ci se fixe sur la glutamine synthase GlnA et inhibe son activité. Ce complexe enzymatique séquestre alors TnrA [243]. Ainsi, nous considérons la glutamine comme l'effecteur métabolique de TnrA. TnrA agit sur diverses voies métaboliques, incluant la dégradation de l'asparagine et des purines conduisant à la production de NH3 [64, 19]. La glutamine n'étant pas directement impliquée dans ces voies, TnrA est un régulateur global ;
- **CodY :** CodY agit sur différentes voies métaboliques, incluant la voie de synthèse des acides aminés branchés, ainsi que les voies de dégradation d'histidine et d'asparagine produisant du glutamate. De plus, CodY est aussi impliqué dans le contrôle de la compétence, de la mobilité et de la sporulation [149]. CodY est donc un régulateur global. Les acides aminés branchés et le GTP sont les effecteurs métaboliques de CodY [126]. Cependant, CodY peut également être considéré comme un régulateur local de la voie de synthèse des acides aminés branchés, car la modulation de la fixation de CodY sur l'ADN par le GTP est faible [207]. Dans ces conditions, les acides aminés branchés sont considérés comme les principaux effecteurs de CodY.

Les autres régulateurs globaux identifiés selon notre définition sont PhoR, Spx, Fur, SpoIIID, SpoVT, Spo0A, AbrB et DegU, que l'on peut décomposer en deux sous-ensembles. Le premier se compose de régulateurs globaux qui

- gèrent la fonction métabolique en fonction de la disponibilité d'une ressource particulière critique, comme par exemple la présence de phosphate, mesurée par le régulateur global à deux composantes PhoP/R [3] ;
- adaptent la capacité du réseau métabolique pour augmenter/diminuer la production de certains métabolites, comme par exemple la stimulation de la conversion de méthionine en cystéine par Spx quand la méthionine est la seule source de soufre [37] ou la répression d'enzymes possédant un centre fer-soufre en absence de fer [69].

Ces trois régulateurs (PhoR, Spx, Fur) sont sous le contrôle de signaux spécifiques qui sont respectivement la concentration en phosphate, la présence de ponts disulfides et la disponibilité d'ions Fe^{2+} libres dans la cellule.
Le second sous-ensemble est caractérisé par des facteurs de transcription impliqués dans le contrôle des fonctions biologiques autre que le métabolisme, comme la sporulation (SpoIIID, SpoVT, Spo0A) et la transition entre la phase exponentielle et la phase stationnaire (AbrB and DegU).

En appliquant notre définition, notre ensemble de régulateurs globaux coïncide avec les régulateurs globaux définis par les biologistes [151, 212]. Notre définition des régulations

locales/globales s'en trouve ainsi renforcée.

3.1.3 Une représentation modulaire de la fonction métabolique

Sur la base de nos résultats énoncés au chapitre précédent, le réseau de régulation du métabolisme chez *Bacillus subtilis* est hautement structuré et organisé en modules élémentaires contrôlés localement et coordonnés entre eux à travers des régulations globales. Un module élémentaire est régulé localement par un métabolite interne au module. Les régulateurs globaux impliqués dans la coordination des modules répondent à des signaux métaboliques caractérisant un état physiologique spécifique. Si nous reprenons nos trois régulateurs globaux principaux chez *Bacillus subtilis*, CcpA, TnrA et CodY, le FBP, principal effecteur de CcpA, représente la disponibilité de sources glycolytiques favorables à la croissance [209]. La glutamine, effecteur de TnrA, représente la disponibilité de sources d'azote préférentielles [239]. Concernant les effecteurs de CodY, le GTP est impliqué dans la synthèse des ribosomes [119] et dans la sporulation [158]. La concentration des acides aminés branchés peut représenter à la fois leur disponibilité dans le milieu extérieur, mais aussi un rôle potentiel dans la gestion globale de la cellule. Les signaux métaboliques mesurés par les régulateurs globaux sont ainsi des signaux pléiotropes car ils représentent un état physiologique spécifique. Il serait alors intéressant d'identifier pour chaque régulateur global de *B. subtilis*, la/les condition(s) physiologique(s) associée(s) et représentée(s) par le signal métabolique.
Enfin, on peut déduire des régulations locales les règles on/off d'adaptation de tous les modules élémentaires en fonction des conditions environnementales (voir la figure 3.1, les nutriments extracellulaires sont en bleu). On obtient ainsi une représentation de l'adaptation du réseau de régulation génétique pour le réseau métabolique.

3.2 Des questions biologiques ouvertes

La décomposition modulaire du réseau métabolique de *Bacillus subtilis* nous permet ici de donner des éléments de réponse à certaines questions ouvertes en Biologie comme la présence de gènes dupliqués, correspondant alors aux isoenzymes pour les voies métaboliques. La présence de gènes dupliqués est souvent interprétée dans la littérature comme un moyen de robustifier le système biologique (voir [216] et les références mentionnées). La robustesse des systèmes biologiques est une notion fondamentale qui a émergé au début des années 2000, issue à l'origine d'un parallèle avec la robustesse des systèmes en ingénierie [111, 112]. Bien que clairement définie en Automatique, l'extension de cette notion pour les systèmes biologiques est compliquée [216]. En Automatique, on parle de robustesse en stabilité et de robustesse en performance de la loi de commande d'un système vis-à-vis d'incertitudes (bornées en norme). Le calcul de la borne supérieure de la valeur singulière structurée nous donne même une condition suffisante pour la robustesse de la loi de commande [58].
En Biologie, les notions de perturbation et d'incertitude sont floues, tout comme la définition même de la robustesse. Les perturbations et incertitudes sont de nature diverse

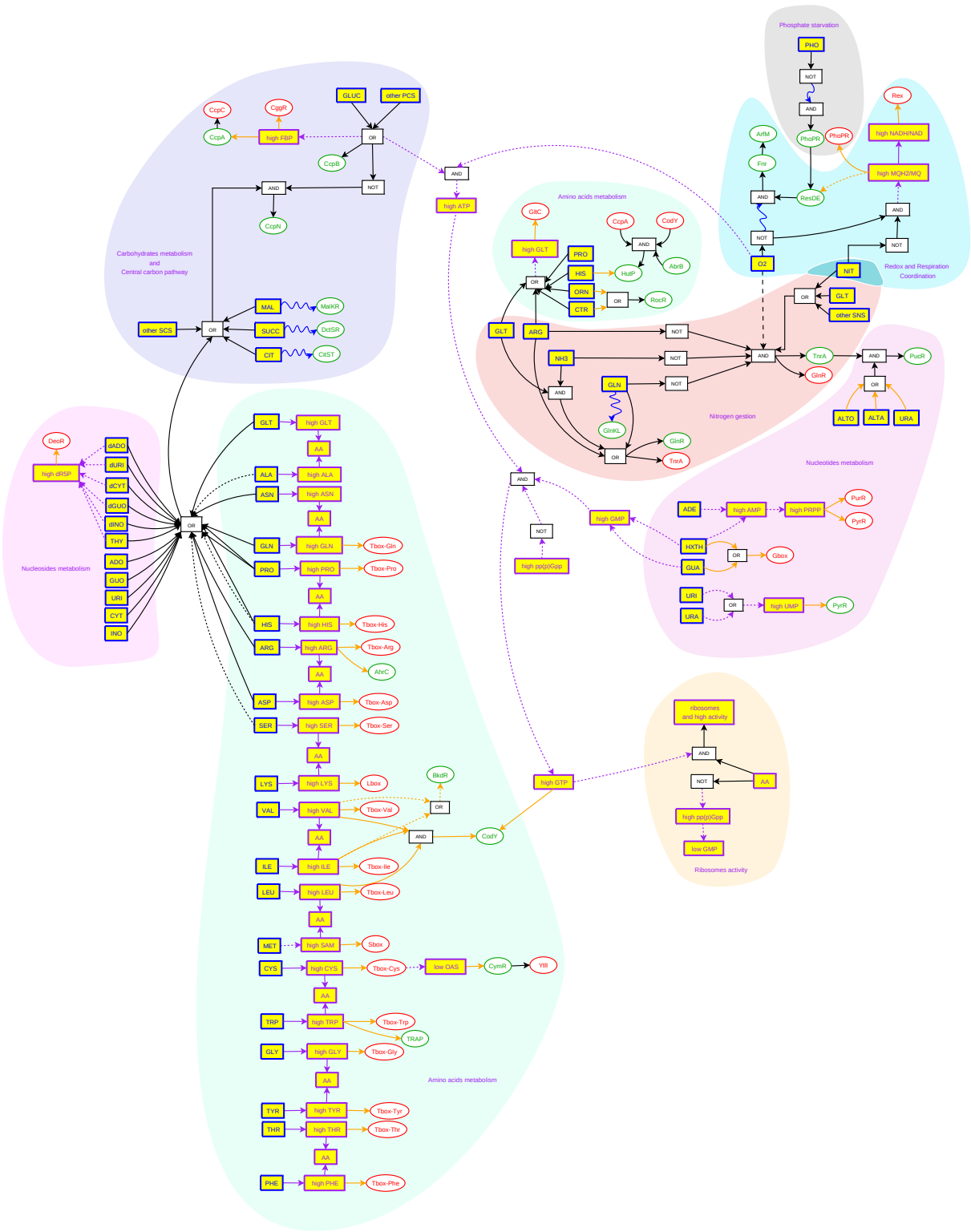


FIG. 3.1 – Une vue modulaire du réseau métabolique de *B. subtilis*

et peuvent correspondre aussi bien à des mutations génétiques, à des modifications environnementales qu'à du bruit intracellulaire. Classiquement, la robustesse en Biologie se définit comme la capacité pour l'organisme (ici la bactérie) à conserver un phénotype stable vis-à-vis des perturbations [216]. Or que recouvre exactement le terme phénotype ? Comment évaluer ses changements vues la complexité et la diversité des mécanismes le composant pour évaluer la robustesse ? Pour obtenir une quantification précise de la robustesse en Biologie comme on pourrait le faire en Automatique, il faudrait d'une part écrire un modèle dynamique global de la cellule, puis isoler les perturbations au niveau entrées/sorties par représentation LFT, et d'autre part accéder à de nombreuses mesures à tous les niveaux de la cellule pour observer les variations du phénotype. Ceci semble clairement impossible actuellement vu le nombre d'états du système, de paramètres à identifier et de mesures à effectuer. La mathématisation de la robustesse au sens de l'Automatique en Biologie reste donc un problème ouvert à l'échelle de la cellule.

Remarque 3.1 *A l'échelle d'un sous-système néanmoins, des modèles dynamiques ont été obtenus et ont permis d'étudier la robustesse du système vis-à-vis de perturbations [13, 125, 205]. On notera par exemple la robustesse d'un modèle dynamique de l'horloge circadienne vis-à-vis de perturbations des paramètres cinétiques [125] ou la «robustesse» du système de mobilité bactérien vis-à-vis de perturbations de concentration de la source de nutriment [13]. Dans ce dernier cas, on parlerait plutôt de rejet de perturbation et non pas de robustesse d'un point de vue de l'Automatique pour décrire cette propriété.*

Par conséquent, la majorité des études centrées sur la robustesse des systèmes biologiques se sont focalisées sur le couple perturbation/phénotype le plus simple à évaluer, à savoir la délétion de gène comparée à la croissance. Le principe est d'effectuer des délétions de gènes et d'observer expérimentalement si la cellule est toujours capable d'assurer sa croissance [116, 215]. Généralement, ce sont toujours des effets délétères qui ont été considérés de manière systématique. Jamais une mutation qui augmenterait le gain d'une régulation par exemple n'a été considérée, alors que son effet serait potentiellement aussi néfaste (déstabilisation de la boucle ...). Face à des délétions de gènes, certains auteurs ont postulé que la robustesse des cellules était obtenue grâce à la redondance de certains composants. En effet, pour se protéger contre la perte d'un certain composant cellulaire, il suffit de proposer une voie alternative pour réaliser la fonction du composant. Dans le cas d'un gène, par exemple, il suffit de dupliquer ce gène pour se prévenir d'une mutation génétique. Dans le cas d'une voie métabolique de synthèse d'un métabolite d'intérêt, il suffit d'avoir une voie de synthèse alternative. La démonstration expérimentale de la redondance a ainsi généré de nombreuses études sur l'essentialité des gènes [116] d'une part, et des voies métaboliques chez les bactéries d'autre part [215]. Or ce concept de redondance peut être mis en défaut dans les deux cas, comme nous allons le voir ci-dessous.

3.2.1 Redondance des gènes

La recherche des gènes dit «essentiels» dans une bactérie comme *Bacillus subtilis* a consisté à détruire systématiquement chacun des 4106 gènes et de regarder son phénotype (croissance ou non) en milieu riche [116]. Pour un milieu riche, seuls 192 gènes ne peuvent être retirés sans obtenir la mort de la bactérie et sont donc considérés comme essentiels. Or certains gènes qualifiés de non essentiels le deviennent lorsque la composition du milieu change comme nous l'avons montré dans [75]. Par exemple, la souche dite *trp*- de *Bacillus subtilis* qui ne peut plus produire de tryptophane à cause d'une mutation sur un des gènes de la voie de biosynthèse (*trpC*), ne peut pas pousser en absence de tryptophane dans le milieu extérieur. En ce sens, le gène *trpC* est essentiel. La notion d'essentialité/redondance des gènes est ainsi intrinsèquement liée aux conditions de test. En outre, si le but des gènes dupliqués est d'accroître la robustesse, pourquoi tous les gènes ne sont-ils pas dupliqués ? Un exemple concret de redondance de gènes dans le réseau métabolique est la présence d'isoenzymes.

3.2.2 Isoenzyme et redondance

Dans la section 2.3.3, nous avons montré que le rôle d'une isoenzyme va bien au-delà de la simple redondance. Elle permet d'augmenter la capacité maximale du flux dans la voie métabolique. Ainsi une mutation sur une des deux isoenzymes lorsqu'elles sont actives n'est certes pas létale mais devrait affecter le taux de croissance de la bactérie. La présence d'une isoenzyme indique donc que dans certaines conditions, le flux dans la voie métabolique doit pouvoir augmenter de manière importante. Intrinsèquement, cela implique également que les isoenzymes sont régulées différemment (soit au niveau génétique, soit au niveau de la spécificité des substrats). Considérons maintenant la liste des isoenzymes présentes chez *Bacillus subtilis* et leur régulation résumée dans les tableaux 3.1 et 3.2. Parmi les isoenzymes dont les régulations sont connues, la plupart sont :

- actives dans différentes conditions physiologiques comme la croissance, la sporulation, ou le stress osmotique (ex : *asnH*, *asnB* et *asnO*) ;
- impliquées dans différentes voies métaboliques et régulées différemment soit au niveau enzymatique, soit au niveau transcriptionnel (ex : *carA*/*carB* et *pyrAA*/*pyrAB*).

Seules quelques isoenzymes possèdent la même régulation. Il s'agit des couples d'aminoacyl ARNt synthases (*tyrS*/*tyrZ*, *thrS*/*thrZ*), qui sont régulées par les mêmes mécanismes de T-box spécifiques respectivement en tyrosine et en thréonine. Elles sont pourtant actives dans des conditions physiologiques différentes [172, 92]. D'autres isoenzymes du tableau 3.2 comme *cysK*/*mccA*, *cysK*/*mccB*, et *fabHA*/*fabHB*/*fabR*, *fadE*/*acdA* et *fadN*/*fadB* sont régulées de la même manière et présentes dans les mêmes conditions physiologiques. Elles présentent néanmoins des spécificités de substrats complémentaires et catalysent donc des réactions chimiques différentes [97, 35, 29, 144, 117].

Nous venons ainsi de montrer que sur la base de la connaissance actuelle des régulations de *Bacillus subtilis*, il n'existe aucune isoenzyme strictement redondante (au sens où les deux enzymes catalysent la même réaction chimique et sont contrôlées de manière

identique). Ce résultat va dans le sens des deux chapitres suivants de cette thèse, où on montrera que toute protéine produite pour les voies métaboliques diminue le taux de croissance. Dans ce contexte, produire des isoenzymes «pour rien» dans le but de se prévenir de mutations génétiques est une perte de taux de croissance.

3.2.3 Le rôle critique de la première enzyme

L'étude théorique de la structure de contrôle par le produit final a révélé le rôle fondamental de la première enzyme irréversible (correspondant le plus souvent à la première enzyme d'une voie métabolique) dans la définition du point d'équilibre et sur la capacité maximale de la voie métabolique (voir la proposition 2.2 page 46 et le corollaire 2.3 page 47 du chapitre précédent).
Dans ce contexte, il est intéressant d'examiner les propriétés biologiques des premières enzymes des voies métaboliques de *Bacillus subtilis*, en particulier lors de modifications brutales des conditions extérieures comme l'épuisement d'une ressource. La bactérie doit avoir la capacité d'augmenter ou au contraire de diminuer brutalement les flux métaboliques dans certaines voies, en particulier celles consommant de nombreux co-métabolites, de manière à économiser la ressource.
Dans le cas de la structure de contrôle par le produit final par exemple où la première enzyme de la voie métabolique est souvent irréversible, la diminution brutale de la demande finale de flux entraîne l'augmentation de la concentration du métabolite final, ce qui a alors pour conséquence de (*i*) diminuer le flux métabolique passant dans la voie métabolique grâce au contrôle enzymatique du produit final sur la première enzyme et de (*ii*) bloquer la synthèse de la première enzyme. La concentration enzymatique diminue par un simple effet de dilution. Or on peut imaginer que dans certains cas, la bactérie veuille avoir la possibilité de couper complètement le flux métabolique sans conserver une concentration élevée du produit final, en inactivant l'enzyme par phosphorylation par exemple ou en dégradant spécifiquement la ou les enzymes de la voie. Notre étude théorique montre qu'inactiver la première enzyme irréversible suffit à l'adaptation du flux métabolique. Une étude expérimentale récente sur *Bacillus subtilis* a montré que les premières enzymes, qui sont d'ailleurs irréversibles, des voies de synthèse des principaux constituants de la cellule sont dégradées par la protéase ClpCP en condition de carence en glucose : les purines (*purF*), les pyrimidines (*pyrB*), les acides aminés branchés (*ilvB*), le peptidoglycane (*glmS*, *murAA*) [72]. Par ailleurs, les premières enzymes des voies de synthèse des acides aminés aromatiques (*aroA*), de l'arginine (*carA*), du glutamate (*gltA*) sont des substrats potentiels de ClpCP, de même que l'enzyme *pycA* qui catalyse la première étape du cycle de Krebs nécessaire à la synthèse d'aspartate, de glutamate et de leurs dérivés [72].
De même les premières enzymes des voies de synthèse des acides aminés aromatiques, de la glycine et du GMP correspondant respectivement aux enzymes AroA, GlyA, GuaB peuvent être phosphorylées [57, 134] et pourraient donc être inactivées de façon transitoire.

Voie métabolique	Gène	Régulation génétique
cycle de Krebs	*citA*	CitR
	citZ	CcpA, CcpC
	mmgD	CcpA, σ_E
cycle de Krebs	*malS*	inconnu
	ytsJ	inconnu
	maeA	MalR
voie des pentoses phosphates	*gntZ*	GntR, CcpA, CcpB
	yqjI	inconnu
	yqeC	inconnu
conversion des pyrimidines	*cmk*	inconnu
	pyrH	inconnu
conversion des pyrimidines	*thyA*	inconnu
	thyB	inconnu
conversion des pyrimidines	*dut*	inconnu
	yosS	inconnu
conversion des pyrimidines	*upp*	inconnu
	pyrR	PurR, PyrR
thiorédoxines	*trxB*	Spx
	yumC	inconnu
	ycgT	Fur
synthèse des acides gras	*fabHA*	FapR
	fabHB	FapR
	fabF	FapR
synthèse des acides gras	*fabI*	FapR
	fabL	inconnu
dégradation des acides gras	*lcfA*	CcpA
	lcfB	FadR
	yhfT	inconnu
dégradation des acides gras	*fadE*	FadR
	acdA	FadR
	mmgC	σ_E
dégradation des acides gras	*fadN*	FadR
	fadB	FadR
	mmgB	σ_E
dégradation des acides gras	*fadA*	FadR
	yhfS	inconnu
	mmgA	σ_E

TAB. 3.1 – Isoenzymes dans la voie centrale des carbones, les nucléotides et les acides gras

Voie métabolique	Gène	Régulation génétique
dégradation du glutamate	*rocG*	RocR, CcpA
	gudB	inconnu
synthèse de l'asparagine	*asnB*	inconnu
	asnH	AbrB, CodY
	asnO	SpoIIID
dégradation de l'asparagine	*ansA*	AnsR
	ansZ	TnrA
synthèse d'acétoine et d'isoleucine, leucine, valine	*alsS*	AlsR
	ilvB-ilvH	T-box leucine CcpA, TnrA, CodY
synthèse d'isoleucine de leucine et de valine	*ywaA*	inconnu
	ybgE	CodY
synthèse de la lysine/thréonine	*lysC*	L-box
	dapG	constitutif
	yclM	inconnu
synthèse de la thréonine	*thrS*	T-box thréonine
	thrZ	T-box thréonine
synthèse de la méthionine	*metC*	S-box
	patB	constitutive
assimilation du sulfate	*sat*	CysR
	yitA	σ_K
assimilation du sulfate	*cysH*	CysR
	yitB	σ_K
assimilation du sulfate	*cysC*	CysR
	yisZ	σ_K
synthèse de la cystéine	*cysK*	CymR
	mccA	CymR
synthèse de la cystéine	*cysK*	CymR
	mccB	CymR
	patB	constitutif
	metC	S-box
synthèse de l'arginine et des pyrimidine	*carA-carB*	AhrC, AbrB
	pyrAA-pyrAB	PyrR, PurR
dégradation de l'arginine et de la proline	*rocA*	RocR, AhrC
	putC	CcpA
synthèse de la proline	*proB*	T-box proline
	proJ	stress osmotique
synthèse de la proline	*proG*	inconnu
	proH	stress osmotique
	proI	T-box proline
synthèse de l'histidine	*hisS*	T-box histidine
	hisZ	inconnu
synthèse de la tyrosine et de la phenylalanine	*aroA*	inconnu
	aroH	inconnu
synthèse de la tyrosine et de la phenylalanine	*aroJ*	TRAP,inconnu
	X	inconnu
synthèse de la tyrosine	*tyrS*	T-box tyrosine
	tyrZ	T-box tyrosine

TAB. 3.2 – Isoenzymes dans les voies des acides aminés

3.3 Confrontation à des données expérimentales

L'objectif est maintenant de valider dans une certaine mesure nos résultats théoriques obtenus au chapitre précédent. Nos résultats offrent une interprétation rationnelle forte des spécificités biologiques existant dans le réseau métabolique. On ne reviendra pas sur le rôle clé des enzymes irréversibles et du découplage vis-à-vis des co-métabolites et des co-facteurs résultant car nous en avons déjà longuement discuté. Il s'agit maintenant d'exploiter un autre résultat théorique à savoir la prédiction du comportement qualitatif de certains composants du réseau métaboliques (voir les corollaires 2.4 et 2.5 page 48 ainsi que les règles d'interconnexion des modules page 67) : les concentrations des enzymes, des métabolites et des flux métaboliques.

Nous allons confronter nos prédictions sur le fonctionnement des modules à des données expérimentales sur deux exemples de voies métaboliques :
- la synthèse de la lysine (exemple de contrôle par le produit final) ;
- la voie basse de la glycolyse (exemple de contrôle par le produit initial).

Les données expérimentales ont été obtenues dans le cadre du projet européen BaSysBio et correspondent à des données de transcriptome et de métabolome dans trois conditions de croissance différentes : glucose, malate et malate plus glucose (noté malate+glucose par la suite). Pour valider le comportement qualitatif des concentrations des enzymes, on supposera que celles-ci varient comme les concentrations des ARN messagers. Sous l'hypothèse qu'aucune régulation traductionnelle ou post-traductionnelle n'intervienne, nous avions en effet montré au chapitre précédent que la dynamique des concentrations des enzymes correspond à celle des concentrations des ARN messagers filtrés par un premier ordre (voir le système (2.3) page 43) :

$$\dot{E}(t) = k_e m(t) R_L(t) - \mu E(t)$$

Le retard τ_e du système (2.3) a ici été négligé car nous nous intéressons uniquement au régime permanent, et plus particulièrement à la comparaison entre trois régimes permanents de croissance différents (glucose, malate et malate+glucose). Le régime d'équilibre de l'équation précédente dépend de la régulation génétique s'il y en a une, mais aussi de variables dépendant directement de la valeur du taux de croissance comme les concentrations des ribosomes (voir le système (2.4) de la section 2.3.1). Les taux de croissance dans les conditions malate et glucose sont quasi identiques (0.57h^{-1} et 0.59h^{-1} respectivement) alors que le taux de croissance en malate+glucose est plus élevé (0.75h^{-1}). L'impact du taux de croissance sur l'adaptation générale de la cellule sera ainsi limité lorsque les conditions malate et glucose sont comparées. En revanche, les comparaisons malate et malate+glucose, tout comme glucose et malate+glucose permettront de révéler la part du taux de croissance dans l'adaptation des voies métaboliques.

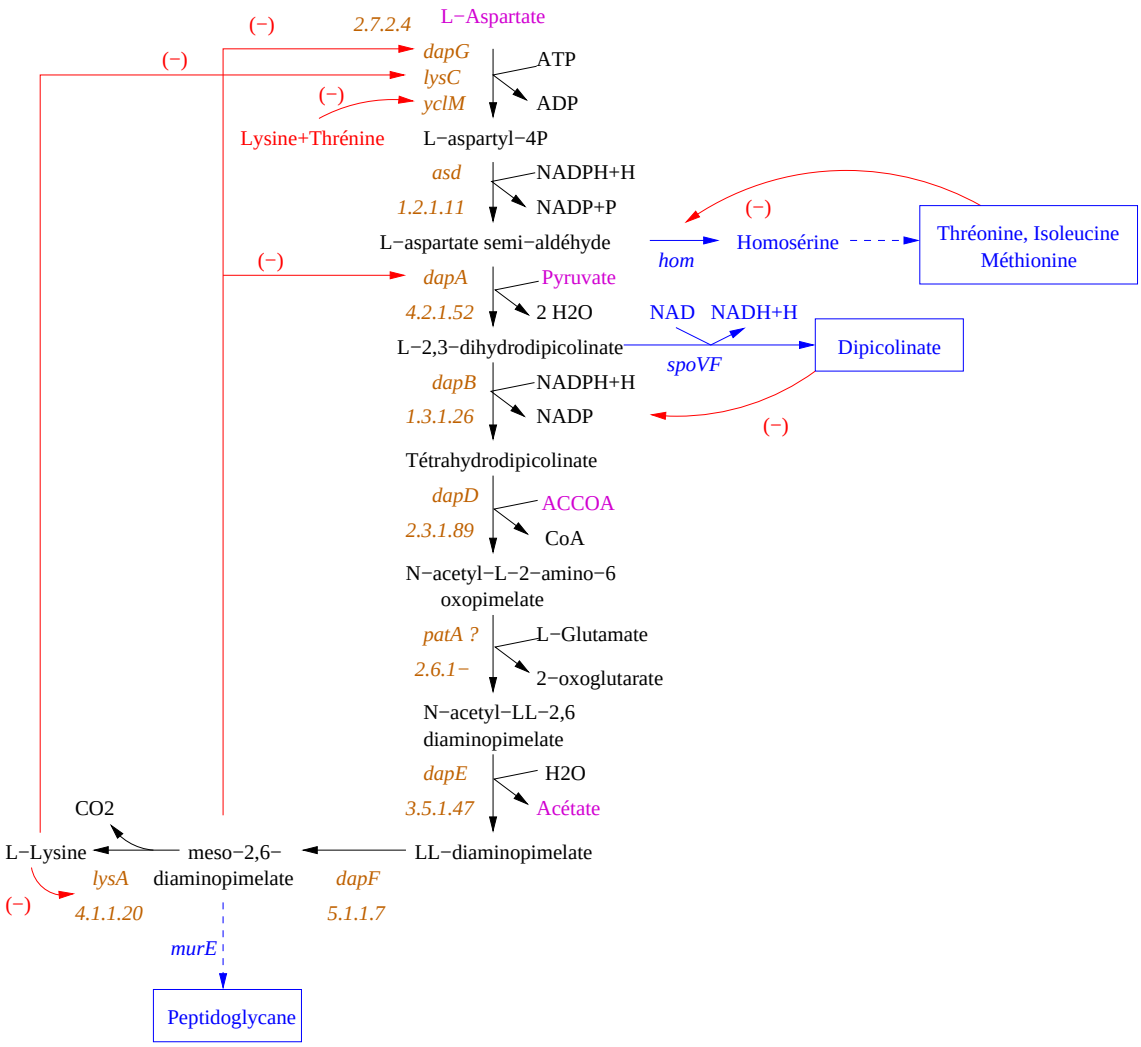


FIG. 3.2 – La voie de synthèse de la lysine et ses régulations enzymatiques connues

3.3.1 Exemple d'une structure de contrôle par le produit final : la synthèse de la lysine

a. Description de la voie métabolique

Les caractéristiques de la voie de synthèse de la lysine sont représentées sur les figures 3.2 et 3.3. La figure 3.2 décrit les réactions chimiques composant la voie de synthèse de la lysine et toutes ses régulations enzymatiques identifiées expérimentalement :

- trois isoenzymes composent la première étape de la voie métabolique [179, 178, 80, 81] ;
- trois points de branchement au niveau du L-aspartate-semi-aldéhyde, du L-2,3-dihydrodipicolinate et du meso-2,6-diaminopimelate pour synthétiser respectivement la thréonine/méthionine, le dipicolinate et le peptidoglycane ;
- l'activité de la première isoenzyme DapG, codée par le gène *dapG*, est inhibée par

le meso-2,6-diaminopimelate [179] ;
- l'activité de la deuxième isoenzyme LysC, codée par le gène *lysC*, est inhibée par la lysine [179, 80, 81] ;
- l'activité de la troisième isoenzyme YclM, codée par le gène *yclM*, est inhibée par la lysine et la thréonine de façon conjointe [81] ;
- la plupart des enzymes utilise des co-métabolites pour leur réaction enzymatique.

La synthèse de dipicolinate n'a lieu qu'en sporulation et non pas en phase exponentielle de croissance [44]. Par conséquent la transcription du gène *spoVF* n'a pas lieu dans les conditions expérimentales considérées. Dans la suite, nous allons négliger le point de branchement de synthèse du dipicolinate.

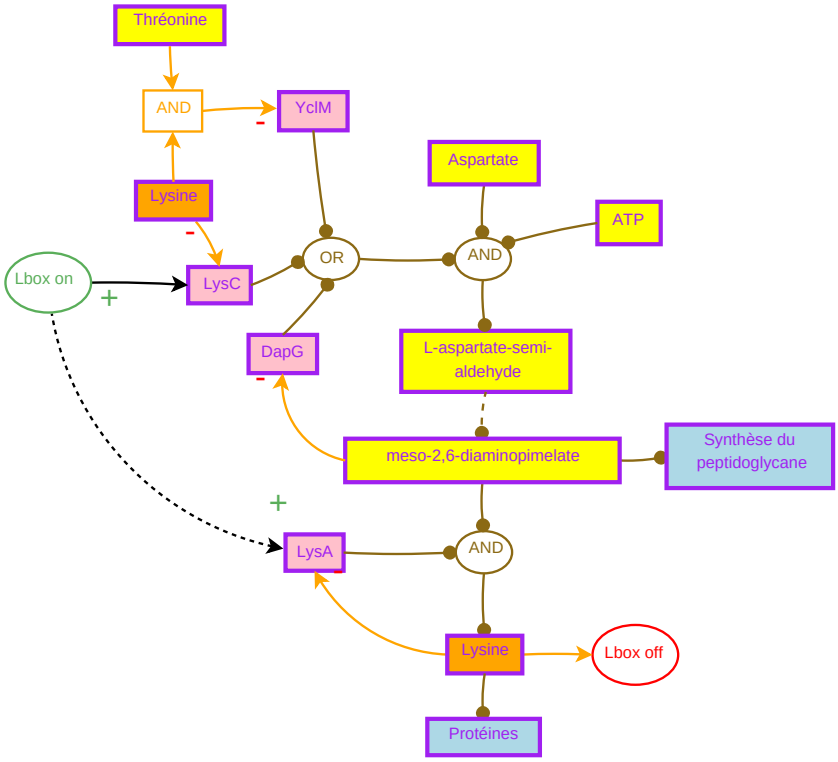


FIG. 3.3 – La régulation de la voie de la lysine

La figure 3.3 souligne les éléments clés de la régulation génétique et enzymatique de cette voie, ainsi que ses connexions avec les autres voies métaboliques :
- la transcription du gène *dapG* est constitutive, c'est-à-dire non régulée [179] ;
- la régulation précise de la transcription du gène *yclM* est inconnue, mais semble être induite par la lysine et réprimée par la thréonine [81] ;
- les gènes *dapG, asd* et *dapA* sont organisés en opéron [34] ;
- la transcription du gène *lysC* est réprimée par terminaison précoce de l'ARN messager grâce à un mécanisme de type L-box spécifique de la lysine [83]. De plus, la transcription de ce gène semble induit en présence de méthionine même si le

mécanisme de régulation n'est pas connu [81] ;
- les autres gènes composant la voie métabolique excepté *lysC* ne sont pas organisés en opéron et la régulation de leur transcription est inconnue.

La L-box correspond à un riboswitch impliquant la lysine. La lysine se fixe directement sur l'ARNm naissant de *lysC* et cause ainsi un changement structurel du messager qui termine la transcription. Le gène *lysA* pourrait être régulé de la même façon, même si la régulation n'a pas été clairement caractérisée [83]. Nous n'allons donc pas intégrer ce mécanisme dans notre analyse, de même que tout autre mécanisme de régulation qui n'a pas été identifié expérimentalement. Ainsi les mécanismes de régulation génétique inconnus régulant *yclM* et *lysC* ne seront pas considérés dans notre analyse. Au regard de la complexité des régulations dans la Biologie, ces effets de régulation observés en présence de lysine, méthionine ou de thréonine pourraient être des effets indirects.

b. Formalisation mathématique de la voie métabolique

La structure de contrôle de cette voie métabolique correspond donc bien à une régulation par le produit final avec :
- les trois isoenzymes LysC, DapG et YclM sont supposées irréversibles car la réaction enzymatique consomme un ATP ;
- l'inhibition de l'activité de la première enzyme par le produit final (la lysine) ;
- la répression de la synthèse de la première enzyme par le produit final (mécanisme L-box).

Les autres enzymes de la voie métabolique sont supposées réversibles, à l'exception de l'enzyme LysA. En effet, cette réaction libère un CO2, ce qui rend difficile la réaction inverse. La voie métabolique est alors décrite par le système d'équations différentielles suivant :

$$\left\{\begin{array}{rcl}
\dot{x}_1 & = & \nu_0 - E_1 f_1(x_1, x_{10}, p_1) - E_1^* f_1^*(x_1, x_9, p_1) - E_1^{**} f_1^{**}(x_1, x_{10}p_2, p_1) \\
\dot{x}_2 & = & E_1 f_1(x_1, x_{10}, p_1) + E_1^* f_1^*(x_1, x_9, p_1) + E_1^{**} f_1^{**}(x_1, x_{10}p_2, p_1) - \ldots \\
 & & E_2 f_2(x_2, x_3, p_3, p_4, p_5, p_6) \\
\dot{x}_3 & = & E_2 f_2(x_2, x_3, p_3, p_4, p_5, p_6) - E_3 f_3(x_3, x_4, p_7) - \nu_{hom} \\
\dot{x}_4 & = & E_3 f_3(x_3, x_4, p_7) - E_4 f_4(x_4, x_5, p_3, p_4, p_5) \\
\dot{x}_5 & = & E_4 f_4(x_4, x_5, p_3, p_4, p_5) - E_5 f_5(x_5, x_6, p_8, p_9) \\
\dot{x}_6 & = & E_5 f_5(x_5, x_6, p_8, p_9) - E_6 f_6(x_6, x_7, p_{10}, p_{11}) \\
\dot{x}_7 & = & E_6 f_6(x_6, x_7, p_{10}, p_{11}) - E_7 f_7(x_7, x_8, p_{12}) \\
\dot{x}_8 & = & E_7 f_7(x_7, x_8, p_{12}) - E_8 f_8(x_8, x_9) \\
\dot{x}_9 & = & E_8 f_8(x_8, x_9) - E_9 f_9(x_9) - \nu_{mur} \\
\dot{x}_{10} & = & E_9 f_9(x_9) - E_{10} f_{10}(x_{10}) \\
\dot{E}_1 & = & g_1(x_{10}) - \mu E_1
\end{array}\right. \tag{3.1}$$

où les notations des différents métabolites, co-métabolites et enzymes sont donnés dans le tableau 3.3, et les propriétés de monotonicité des différentes fonctions f_i en fonction de leurs arguments sont données par les caractéristiques des enzymes décrites précédemment page 42. Le point essentiel étant que la fonction f_1 (respectivement f_1^*, f_1^{**}) soit strictement décroissante en x_{10} (respectivement x_9, $x_{10}p_2$). Le terme $x_{10}p_2$ de la fonction

f_1^{**} décrit l'inhibition conjointe de l'enzyme YclM par la thréonine et la lysine. La fonction g_1 est continue, bornée et strictement décroissante en x_{10}. L'eau (H2O) n'est pas intégrée dans les équations différentielles car ce métabolite est toujours en excès dans la cellule (sauf dans le cas d'un stress osmotique non considéré ici). La vitesse des réactions associées n'est donc pas déterminée par ce métabolite.

Notation	Métabolite	Notation	Enzyme
x_1 :	aspartate	E_1 :	LysC
x_2 :	L-aspartyl-4P	E_1^* :	DapG
x_3 :	L-asparte-semi-aldéhyde	E_1^{**} :	YclM
x_4 :	L-2,3-dihydrodipicolinate	E_2 :	Asd
x_5 :	tetrahydrodipicolinate	E_3 :	DapA
x_6 :	N-acétyl-L-2-amino-6-oxopimelate	E_4 :	DapB
x_7 :	N-acétyl-LL-2,6-diaminopimelate	E_5 :	DapD
x_8 :	LL-diaminopimelate	E_6 :	PatA
x_9 :	meso-2,6-diaminopimelate	E_7 :	DapE
x_{10} :	lysine	E_8 :	DapF
p_1 :	ATP	E_9 :	LysA
p_2 :	thréonine	E_{10} :	LysS
p_3 :	NADPH	E_{hom}	Hom
p_4 :	H	E_{mur}	MurE
p_5 :	NADP		
p_6 :	P		
p_7 :	pyruvate		
p_8 :	ACCOA		
p_9 :	CoA		
p_{10} :	glutamate		
p_{11} :	2-oxoglutarate		
p_{12} :	acétate		

TAB. 3.3 – Notations pour les métabolites, les co-métabolites et les enzymes dans le système (3.1)

Le métabolite initial x_1 de la section 2.3 correspond à l'aspartate (voir la figure 3.2) et l'enzyme final $\mathbb{E}_n$ est la ARNt synthase spécifique de la lysine, codée par le gène *lysS*. De plus, les gènes impliqués dans la voie métabolique (*asd, dapA, dapB, dapD, patA, dapE, dapF* et *lysA*) ne sont pas en opéron avec le gène *lysC*. Les flux ν_{hom} et ν_{mur} correspondent respectivement aux flux de synthèse de la thréonine/méthionine et du peptidoglycane. Ces flux sont supposés déterminés par ces différentes voies de synthèse. Sur la base de leur régulation actuellement connue, le rôle des trois isoenzymes semble donc bien différent :

- le flux de synthèse à travers l'enzyme DapG répond à la demande de meso-2,6-diaminopimelate qui correspond à la fois aux besoins en lysine et en peptidoglycane ;

- le flux de synthèse à travers l'enzyme LysC répond à la demande de lysine ;
- le rôle de l'enzyme YclM est moins clair. L'inhibition conjointe de cette enzyme par la lysine et la thréonine suggère un rôle de coordination entre la voie de synthèse de la lysine et de la thréonine.

c. Caractérisation du régime d'équilibre

D'après les résultats de la section 2.3, sous l'hypothèse qu'aucune enzyme ne sature, il existe un unique régime d'équilibre au système (3.1) donné par :

$$\begin{cases} \bar{E}_1 = \dfrac{g(\bar{x}_{10})}{\mu} \\ \dfrac{g(\bar{x}_{10})}{\mu} f_1(\bar{x}_1, \bar{x}_{10}, \bar{p}_1) + \bar{E}_1^* f_1^*(\bar{x}_1, \bar{x}_9, \bar{p}_1) + \bar{E}_1^{**} f_1^{**}(\bar{x}_1, \bar{x}_{10}\bar{p}_2, \bar{p}_1) = \dots \\ \qquad\qquad \bar{E}_{10} f_{10}(\bar{x}_{10}) + E_{mur} f_{mur}(\bar{x}_9) + E_{hom} f_{hom}(\bar{x}_3) \\ \nu_{mur} = E_{mur} f_{mur}(\bar{x}_9) \\ \nu_{hom} = E_{hom} f_{hom}(\bar{x}_3) \\ \bar{x}_i = H_i(\bar{x}_{i+1}) \quad \text{pour } i \in \{2, \dots, 8\} \\ \bar{x}_9 = H_9(\bar{x}_1, \bar{p}_1, \bar{p}_2) \\ \bar{x}_{10} = H_{10}(\bar{x}_1, \bar{p}_1, \bar{p}_2) \end{cases}$$

où les fonctions H_i pour $i \in \{2, \dots, 8\}$ sont des fonctions strictement croissantes de leur argument, et H_9 et H_{10} sont des fonctions strictement croissantes de $\bar{x}_1$ et de $\bar{p}_1$, et strictement décroissante en fonction de $\bar{p}_2$. Ce régime permanent associé à ce module spécifique dépend donc de :

- des métabolites différents de la lysine et agissant au niveau de LysC correspondant ici à l'aspartate x_1, à l'ATP p_1 et à la thréonine p_2 ;
- du flux final demandé $\nu_{10} \triangleq E_{10} f_{10}(x_{10})$, défini par l'activité de l'enzyme LysS[2] et des flux intermédiaires ν_{mur} et ν_{hom} ;
- des facteurs extérieurs modulant la transcription/traduction de la première enzyme intégré dans la fonction g_1 comme l'activité/le nombre des ARN polymérases et/ou des ribosomes,

et ne dépend ni des co-métabolites (p_3, p_4, p_5, p_6, p_7, p_8, p_9, p_{10}, p_{11}, p_{12}), ni des concentrations des métabolites intermédiaires (x_2, x_3, x_4, x_5, x_6, x_7, x_8).

Dans ce contexte, la prédiction qualitative des variations des composants clés de ce module, obtenue sur la base des corollaires 2.4 et 2.5 pages 48 étendus pour la configuration enzymatique et génétique propre à la voie de la lysine, est indiquée dans le

[2]La fonction f_{10} devrait aussi être croissante en p_1 (voir l'annexe A.2), ce qui modifierait le régime d'équilibre. Le comportement du régime d'équilibre vis-à-vis de variations en p_1 dépend alors des caractéristiques cinétiques f_1 et f_{10} des enzymes E_1 et E_{10}. Dans le cadre de cette analyse, nous avons considéré que l'activité de LysS était quasi saturée par l'ATP (f_{10} ne dépend pas de p_1).

tableau 3.4 en supposant que les activités des ARN polymérases et des ribosomes restent constantes. De même, les prédictions obtenues pour des variations d'ATP ou de Thréonine-Lysine sont obtenues sous les hypothèses des corollaires 2.4 et 2.5, c'est-à-dire pour une concentration x_1 correspondant à l'aspartate fixée.

LysS	Aspartate	ATP	Thréonine-Lysine	Lysine	Flux	ARNm-*lysC*
+				-	+	+
-				+	-	-
	+			+	+	-
	-			-	-	+
		+		+	+	-
		-		-	-	+
			+		-	-
			-		+	+

TAB. 3.4 – Prédiction des variations des composants clés de la voie de la lysine

d. Comparaison avec les données expérimentales

Comparons maintenant ces prédictions avec les résultats expérimentaux présentés dans le tableau 3.5 et 3.6 pour les mêmes éléments. La concentration d'ATP n'a malheureusement pas été mesurée. Néanmoins, certaines expériences biologiques ont montré que la concentration d'ATP augmente avec le taux de croissance [26]. Nous supposerons donc par la suite que la concentration d'ATP est plus élevée en condition de malate+glucose qu'en conditions de malate ou de glucose.

Éléments du module	malate	glucose	malate+glucose
Aspartate (μmol/gdwc)	10.47	1.42	8
Lysine (μmol/gdwc)	0.24	0.13	0.56
Thréonine (μmol/gdwc)	1.35	1.33	3.6
ARNm-*lysC* (log)	12.3	14.3	12.6
ARNm-*lysS* (log)	13.1	13.7	13.4

TAB. 3.5 – Résultats expérimentaux pour les éléments clés de la voie de la lysine pour les trois conditions (Glucose, Malate et Malate+Glucose). L'unité μmol/gdwc signifie micromoles par gramme de poids sec.

Éléments du module	glc/mal	mal/mal+glc	glc/mal+glc
Aspartate	+	-	+
Lysine	+	+	+
Thréonine	0	+	+
ATP	0	+	+
ARNm-*lysC*	-	0	-
ARNm-*lysS*	0	0	0

TAB. 3.6 – Comparaison de l'évolution des différents composants entre les conditions en glucose versus malate (glc/mal), malate versus malate+glucose (mal/mal+glc) et en glucose versus malate+glucose (glc/mal+glc). L'évolution de l'ATP est basée sur [26]. Sur ce tableau, la notation «0» signifie que la variation relative du composant reste inférieure à 5%.

Les taux de croissance dans les conditions malate et glucose sont quasi identiques (0.57h^{-1} et 0.59h^{-1} respectivement), ce qui limite l'impact du taux de croissance sur (*i*) la synthèse des enzymes (via les concentrations des ribosomes et des ARN polymérases), sur (*ii*) la demande en flux de lysine par les ribosomes et sur (*iii*) la concentration d'ATP sous l'hypothèse que la concentration d'ATP varie comme le taux de croissance [26]. Nous sommes alors complètement dans le cadre des résultats de la section 2.3 page 39 lorsque les conditions en malate et en glucose sont comparées. On en déduit que la variation de la concentration de lysine résulte de la variation de la concentration d'aspartate, le premier métabolite de la voie lorsqu'on compare les conditions en malate et en glucose.

En malate+glucose, nous verrons dans la section 3.3.3 que l'effet du taux de croissance est également limité sur les concentrations des ribosomes et des ARN polymérases. Lorsque les conditions de glucose et malate+glucose sont comparées, on peut donc déduire que la variation de la concentration de lysine résulte de la variation de la concentration d'aspartate et potentiellement aussi d'ATP car le taux de croissance a augmenté [26]. Dans ce cas, l'effet de l'ATP va dans le même sens que l'aspartate, à savoir l'augmentation de la lysine. En revanche, lorsque les conditions malate et malate+glucose sont comparées, les variations de l'aspartate et de l'ATP vont dans le sens inverse. Toutefois la lysine augmente, ce qui renforce l'idée que l'ATP augmente. En revanche, même si la lysine augmente, l'impact sur la transcription du gène *lysC* codant pour la première enzyme de la voie est quasi nul. Ceci semble indiquer que le seuil de saturation de la L-box pour la lysine est déjà atteint.

3.3.2 Exemple d'une structure de contrôle par le produit initial : la voie basse de la glycolyse

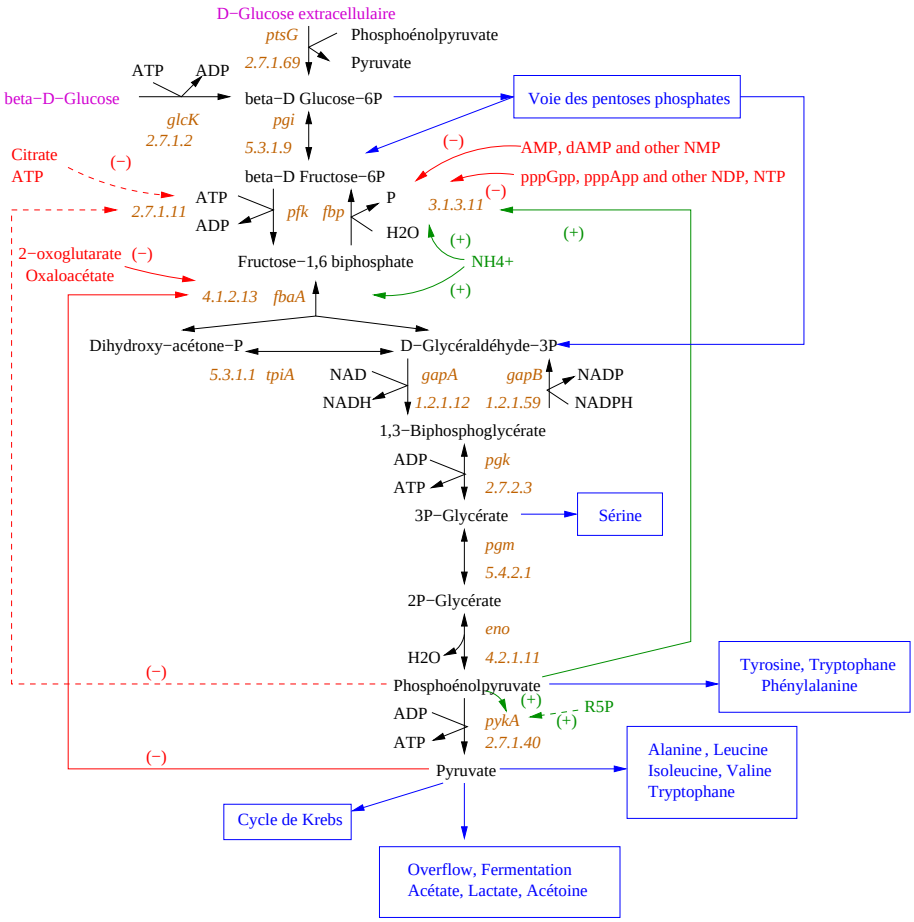


FIG. 3.4 – La glycolyse et ses régulations enzymatiques connues

La glycolyse est la voie métabolique commune de dégradation des sources carbonées permettant de produire d'une part de l'énergie sous forme d'ATP et de NADH, et d'autre part des précurseurs métaboliques nécessaires à la croissance. Les réactions chimiques composant la glycolyse ainsi que les gènes codant pour les enzymes associées sont représentées sur la figure 3.4. Les régulations génétiques caractérisées expérimentalement sont localisées au niveau de la partie basse de la glycolyse. Les gènes *tpiA*, *gapA*, *pgk*, *pgm* et *eno* sont organisés en opéron, et régulés par le facteur de transcription CggR, sensible à la concentration de fructose-1,6-biphosphate (FBP) situé en amont de ces gènes [47] (cf. la figure 3.5). CggR réprime la transcription de cet opéron pour de faibles concentrations de FBP [47].

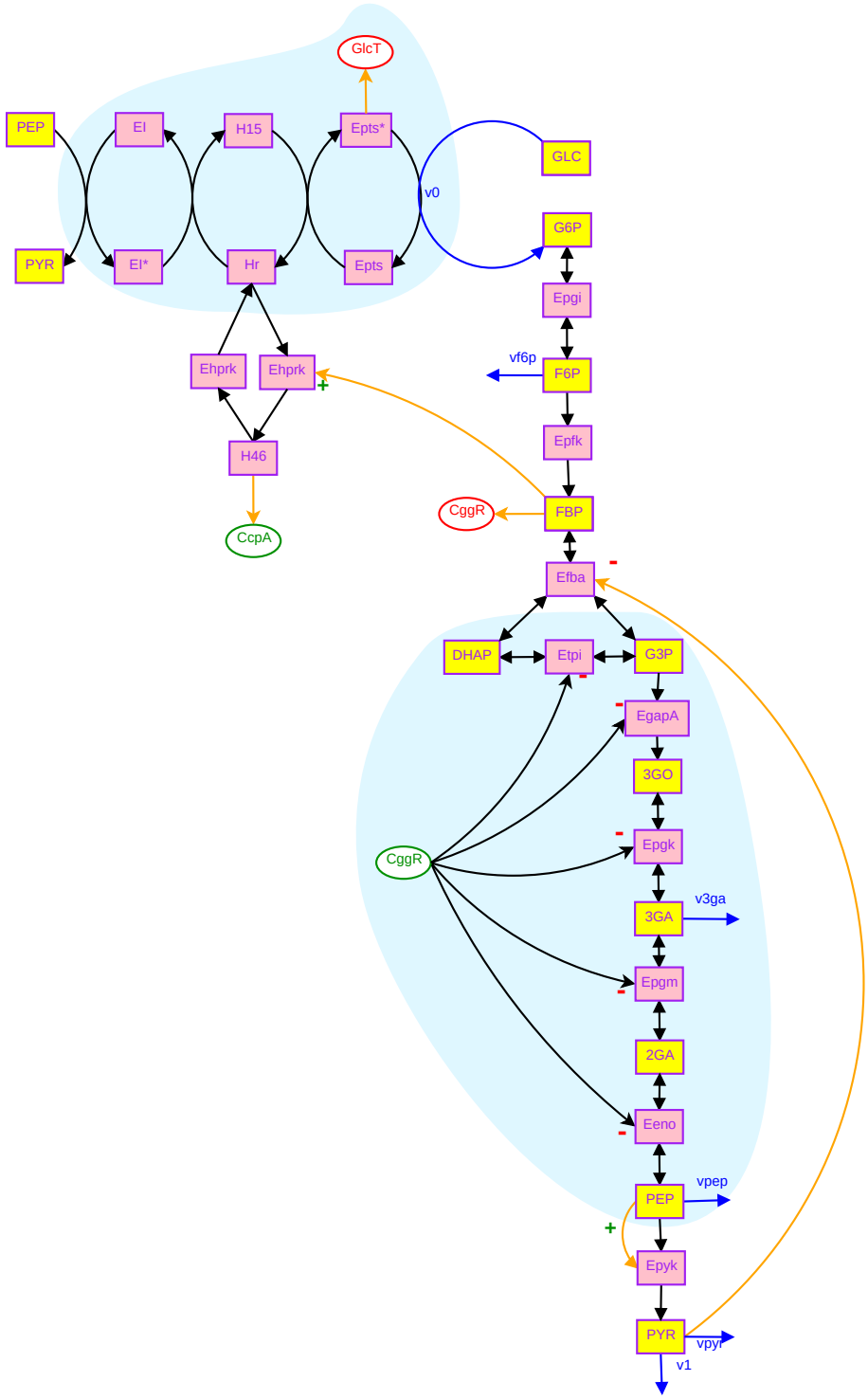


FIG. 3.5 – La régulation de la glycolyse

Considérons le module composé des gènes *fbaA*, *tpiA*, *gapA*, *pgk*, *pgm* et *eno*. L'enzyme GapA est supposée irréversible compte tenu du ratio NADH/NAD très faible dans nos trois conditions de culture aérobie [217]. Cette structure de contrôle par le produit initial (FBP) ne correspond pas exactement à celle étudiée dans la section 2.4.1 page 58 car la première enzyme n'est pas contrôlée au niveau génétique par le métabolite initial FBP. Néanmoins nos résultats théoriques peuvent être étendus sans difficulté, car la caractéristique enzymatique de la première enzyme FbaA est strictement croissante avec le FBP.

D'après la proposition 2.12 page 59 étendue à notre cas, le régime permanent de ce module dépend :
- de la concentration de fructose-1,6-biphosphate ;
- des concentrations des métabolites agissant sur la première enzyme : le pyruvate, l'oxaloacétate et le 2-oxoglutarate ;
- des caractéristiques de l'activité enzymatique de l'enzyme FbaA ;
- des flux arrivant de la voie des pentoses phosphates et du flux de synthèse de sérine.

Nous pouvons alors déduire les prédictions du tableau 3.7 concernant l'évolution des différents composants du module.

Flux_{in}	pyruvate oxaloacétate 2-oxoglutarate	FBP	ARNm-*tpiA* ARNm-*gapA* ARNm-*pgm*	Flux	glycéraldéhyde-3P 3-phosphoglycérate 2-phosphoglycérate
+		+	+	+	+
-		-	-	-	-
	+	+	+	-	-
	-	-	-	+	+

TAB. 3.7 – Prédiction des variations des composants clés de la voie basse de la glycolyse. Flux_{in} correspond au flux de synthèse du FBP provenant du haut de la glycolyse.

En particulier, si la concentration de FBP augmente :
- la transcription des gènes *tpiA*, *gapA*, *pgk*, *pgm* et *eno* augmente ;
- le flux à travers le module augmente ;
- les concentrations des métabolites intermédiaires comme le 3-phosphoglycérate, le 3-phosphoglycérate et le 2-phosphoglycérate augmentent également.

Les résultats expérimentaux associés à différents composants du module sont résumés sur les tableaux 3.8 et 3.9, et confirment nos prédictions concernant la transcription des gènes de l'opéron. L'expression de l'opéron est parfaitement en accord avec la concentration de fructose-1,6-biphosphate, à savoir une expression plus élevée en glucose qu'en malate+glucose, elle-même supérieure à l'expression en malate. Les concentrations des métabolites intermédiaires sont effectivement plus élevées en glucose. En revanche, elles sont du même ordre de grandeur en malate et malate+glucose.

Éléments du module	malate	glucose	malate+glucose
fructose-1,6-biphosphate (μmol/gdwc)	2.51	12.43	10.15
glycéraldéhyde-3P(μmol/gdwc)	0.11	0.19	0.13
3-phosphoglycérate (μmol/gdwc)	52.6	147	47.2
2-phosphoglycérate (μmol/gdwc)	0.11	0.41	0.16
ARNm-*gapA* (log)	13.86	15.86	15.6
ARNm-*tpiA* (log)	12.96	15.13	14.55
ARNm-*pgm* (log)	13.9	15.6	15.11

TAB. 3.8 – Résultats expérimentaux pour les clés de la voie de la glycolyse pour les trois conditions (Glucose, Malate et Malate+Glucose)

Éléments du module	glc/mal	mal/mal+glc	glc/mal+glc
fructose-1,6-biphosphate	-	+	-
glycéraldéhyde-3P	-	+	-
3-phosphoglycérate	-	-	-
2-phosphoglycérate	-	+	-
ARNm-*gapA*	-	+	-
ARNm-*tpiA*	-	+	-
ARNm-*pgm*	-	+	-

TAB. 3.9 – Comparaison de l'évolution des différents composants entre les conditions en glucose versus malate (glc/mal), malate versus malate+glucose (mal/mal+glc) et en glucose versus malate+glucose (glc/mal+glc).

3.3.3 Adaptation des voies métaboliques à faible taux de croissance

Les exemples des deux sections précédentes ont illustré l'intérêt et l'utilité de notre définition des modules non seulement dans la prédiction du comportement des voies métaboliques, mais aussi pour déterminer les causes à des effets biologiques observés. Nous avons ainsi montré le rôle central de certains métabolites dans la réadaptation génétique des voies métaboliques. Plus généralement, cette analyse montre aussi l'utilité des modèles mathématiques pour réconcilier des données expérimentales de nature différente. Pour aller plus loin, notre analyse peut être réalisée sur l'ensemble des modules que nous avons identifiés et montrer que dans nos conditions expérimentales spécifiques la réadaptation des voies métaboliques est essentiellement causée par les variations des concentrations des métabolites.

L'objectif est tout d'abord d'illustrer que l'analyse effectuée pour les deux structures de contrôle présentée dans les sections précédentes peut être réalisée à l'échelle du réseau métabolique, ou du moins pour certaines sous-parties où les régulations génétiques sont connues et les concentrations des métabolites sont mesurées. De façon plus spécifique, il

s'agit également de déterminer le poids des différents composants du réseau métabolique ainsi que des principaux facteurs connus agissant sur son adaptation dans nos trois conditions expérimentales. La part de Biologie pure composant cette section est donc très importante. Nous avons inclus de nombreuses figures pour faciliter la lecture à des non spécialistes du métabolisme des bactéries, et nous avons résumé à la fin de chaque sous-section les principaux résultats obtenus. L'exercise de style reste néanmoins délicat au regard de la dimension et de la forte connectivité du réseau métabolique.

a. Impact des métabolites sur la synthèse des enzymes

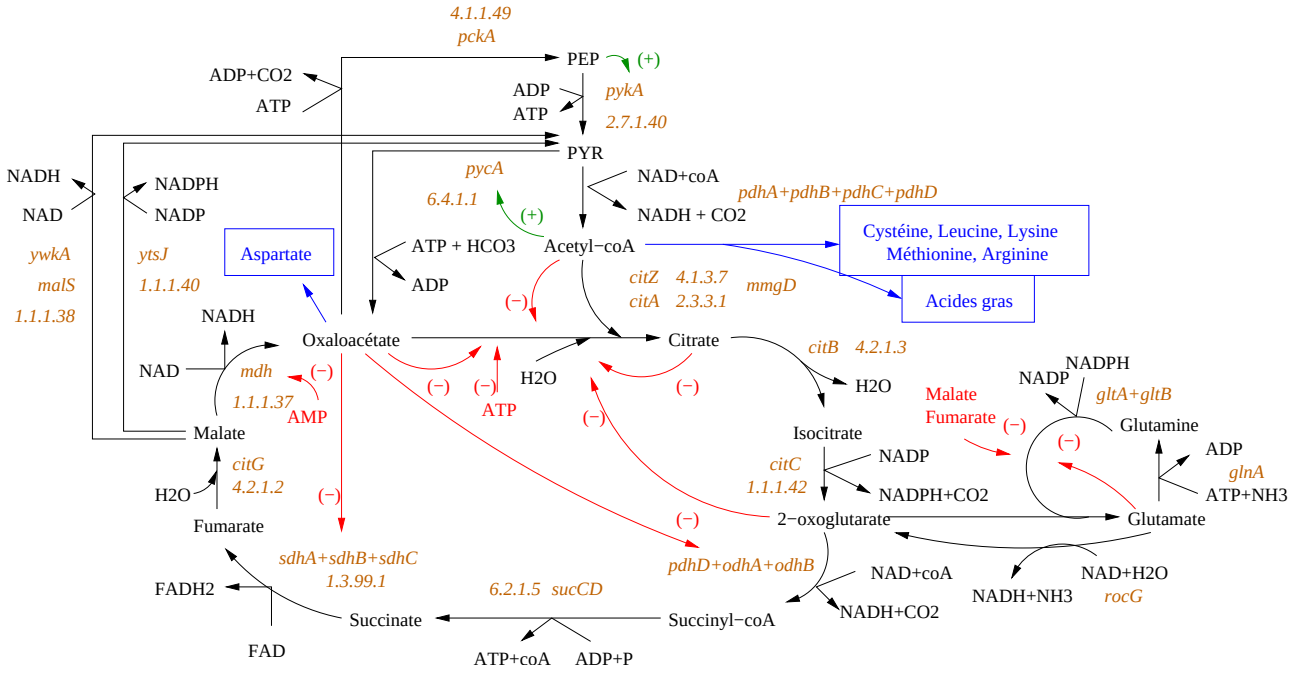


FIG. 3.6 – Les régulations enzymatiques du cycle de Krebs

L'étude de la synthèse de la voie de la lysine a révélé le rôle central de la concentration d'aspartate sur le régime permanent de cette voie métabolique pour nos trois conditions de croissance. L'utilisation judicieuse d'une concentration élevée d'aspartate permet d'obtenir un même flux de synthèse de lysine avec une concentration d'enzyme plus faible en malate comparée à la condition en glucose. Comme nous allons le voir ci-dessous, la concentration élevée d'aspartate en malate et malate+glucose résulte directement d'une concentration intracellulaire de malate élevée. A ce titre, nous allons nous centrer uniquement sur la comparaison des conditions glucose/malate, et glucose/malate+glucose, soit encore lorsque la variation de la concentration de malate est importante et entraîne un réajustement global des voies métaboliques.

La figure 3.6 représente le cycle de Krebs correspondant à une voie métabolique essentielle pour la bactérie. Le cycle de Krebs fournit d'une part de l'énergie à la cellule

sous forme d'ATP et de NADH, et d'autre part produit des précurseurs métaboliques indispensables à la synthèse d'acides aminés ou d'acides gras comme l'oxaloacétate (nécessaire à la synthèse d'aspartate), l'acétyl-coA (pour la synthèse de cystéine, d'acides gras, d'arginine) et le 2-oxoglutarate (pour la synthèse de glutamate). Le malate est un métabolite intermédiaire du cycle de Krebs, qui, chez *Bacillus subtilis*, est transporté en présence de glucose [48]. La présence de malate extracellulaire a donc des conséquences sur le régime permanent de cette voie métabolique en malate et malate+glucose.
Sur la figure 3.6, on constate que le malate est :

- produit par l'enzyme CitG (et également transporté via les transporteurs MaeN et YflS) ;
- consommé par les enzymes maliques MalS, MaeA et YtsJ produisant du pyruvate, et par l'enzyme Mdh qui produit de l'oxaloacétate.

Afin de déterminer les conséquences du malate extracellulaire, nous identifions tout d'abord les structures de contrôle présentes dans le cycle de Krebs et plus particulièrement autour du malate. La synthèse des transporteurs MaeN et YflS et de l'enzyme MaeA est induite en présence de malate extracellulaire par un régulateur dit à doubles composantes YufL/YufM. La partie YufL est localisée dans la membrane et s'active en présence de malate extracellulaire. La forme activée de YufL active alors à son tour YufM qui induit la transcription des gènes d'intérêt, à savoir *maeN*, *maeA* et *yflS*. La structure de contrôle correspond à une structure de contrôle par le produit initial. La synthèse des enzymes CitG et Mdh est régulée par le régulateur global CcpA. De plus, sur la base de la régulation connue, l'enzyme Mdh appartient à un module impliquant le régulateur local CcpC sensible au citrate (voir le tableau 3.13 page 116). Les régulations de MalS et YtsJ sont inconnues. On peut donc en déduire que les enzymes CitG, MalS et YtsJ correspondent à des structures de module non contrôlées localement.

Prédiction des conséquences d'une concentration intracellulaire de malate élevée sur le cycle de Krebs : L'ensemble des prédictions qualitatives des composants est donné dans le tableau 3.10 page 105. En présence de malate, c'est-à-dire en conditions de malate et malate+glucose comparées à la condition en glucose, la régulation génétique par le système à deux composantes YufL/YufM est activée et conduit alors à la synthèse des transporteurs MaeN et YflS, ainsi que de l'enzyme malique MaeA. Nos résultats théoriques de la section 2.4.1 indiquent qu'en présence de malate par rapport au glucose seul :

- le flux de malate transporté est plus élevé,
- la concentration de malate est plus élevé,

ce qui est effectivement constaté expérimentalement (voir le tableau 3.11). Nous prédisons alors que l'augmentation de la concentration de malate a des conséquences sur :

- les structures non contrôlées et contrôlées directement connectées au malate, c'est-à-dire MalS et YtsJ d'une part et Mdh d'autre part ;
- les enzymes dont l'activité est régulée par le malate, c'est-à-dire l'enzyme GltAB qui produit du glutamate à partir du 2-oxoglutarate (voir la figure 3.6).

Sur la base de nos résultats théoriques des sections 2.4.3 page 63, nous en déduisons qu'en malate et malate+glucose par rapport au glucose seul

Composants cellulaires	glc/mal	glc/mal+glc
malate	+	+
$\text{flux}_{\text{MalS}}$	+	+
$\text{flux}_{\text{YtsJ}}$	+	+
flux_{Mdh}	+	+
pyruvate	+	+
oxaloacétate	+	+
$\text{flux}_{\text{CitZ}}$	+	+
flux_{Odh}	-	-
flux_{Sdh}	-	-
succinate	+	+
2-oxoglutarate	+	+
aspartate	+	+
asparagine	+	+
lysine	+	+
ARNm-*lysC*	-	-
$\text{flux}_{\text{GltAB}}$	-	-
glutamine	+	+
2-oxoglutarate	+	+
$\text{flux}_{\text{GlnA}}$	-	-
glutamate	+	+
arginine	+	+
ornithine	+	+
ARNm-*argJ*	-	-
ARNm-*argB*	-	-
ARNm-*argC*	-	-
ARNm-*argD*	-	-
ARNm-*argF*	-	-
ARNm-*carA*	-	-
ARNm-*carB*	-	-
proline	+	+
ARNm-*proA*	-	-
ARNm-*proB*	-	-
tryptophane	+	+
ARNm-*trpA*	-	-
ARNm-*trpB*	-	-
ARNm-*trpC*	-	-
ARNm-*trpD*	-	-
ARNm-*trpE*	-	-
ARNm-*trpF*	-	-
ARNm-*trpG*	-	-

TAB. 3.10 – Prédiction de l'évolution des différents composants en présence de malate, c'est-à-dire entre les conditions en glucose versus malate (glc/mal) et glucose versus malate+glucose (glc/mal+glc).

- les concentrations de pyruvate et d'oxaloacétate sont plus élevées ;
- les concentrations de 2-oxoglutarate et de glutamine sont plus élevées car le flux à travers l'enzyme GltAB diminue.

L'aspartate est produite à partir d'oxaloacétate par l'enzyme AspB dont la synthèse est *a priori* non contrôlée [100]. La structure de contrôle correspond à une structure non contrôlée. Sous l'hypothèse que l'oxaloacétate est plus élevée en malate et malate+glucose, on en déduit donc que l'aspartate est également plus élevée dans les mêmes conditions.

Métabolites	malate	glucose	malate+glucose	glc/mal	glc/mal+glc
malate	178.4	0.94	73.21	+	+
citrate	6.34	1.61	4	+	+
isocitrate	0.43	0.42	0.33	0	-
2-oxoglutarate	10	1.3	9.36	+	+
succinate	125.12	9.53	84.62	+	+
fumarate	0.57	0.45	0.38	+	-
pyruvate	236.37	14.5	301.5	+	+
alanine	16.17	6.83	16.9	+	+
aspartate	10.47	1.42	8	+	+
glutamate	80.16	65.7	80.17	+	+
glutamine	60	6.3	15.3	+	+
histidine	0.22	0.3	0.22	-	-
isoleucine	0.67	0.6	1.61	+	+
leucine	1.5	2	3.15	-	+
lysine	0.24	0.13	0.56	+	+
méthionine	0.4	0.23	0.44	+	+
phénylalanine	0.07	0.21	0.15	-	-
proline	4.13	2.9	4.32	+	+
sérine	0.94	1.25	1.1	-	-
thréonine	1.35	1.33	3.6	0	+
tryptophane	4.61	1.71	3.8	+	+
tyrosine	0.27	0.24	0.33	+	+
valine	4.51	9.8	17.76	-	+
ornithine	5.57	2.72	5.32	+	+

TAB. 3.11 – Concentration de métabolites d'intérêt pour les trois conditions en μmol/gdwc, et leur évolution en glucose versus malate (glc/mal) et en glucose versus malate+glucose (glc/mal+glc). Sur ce tableau, la notation «0» signifie que la variation relative du composant reste inférieure à 5%.

Comparaison avec les données expérimentales : Le tableau 3.11 résume les concentrations des métabolites mesurés dans nos trois conditions expérimentales (malate, malate+glucose et glucose). Les concentrations de pyruvate, d'aspartate de 2-

oxoglutarate et de glutamine sont effectivement plus élevées en présence de malate (correspondant aux conditions de malate et malate+glucose). La concentration d'oxaloacétate n'a pas été mesurée car ce composé est très volatile et donc difficile à mesurer. Néanmoins, compte tenu des concentrations élevées de malate et d'aspartate, la concentration d'oxaloacétate est certainement plus élevée dans les mêmes conditions. On supposera donc par la suite que la concentration d'oxaloacétate est effectivement plus élevée en présence de malate, ce qui nous amène à réaliser une deuxième série de prédictions liées à la concentration élevée d'oxaloacétate d'une part, et de glutamine d'autre part.

Prédiction de l'impact des concentrations élevées d'oxaloacétate et de glutamine en présence de malate : Sur la base des régulations enzymatiques connues représentées sur la figure 3.6, les concentrations élevées d'oxaloacétate et de glutamine en présence de malate vont inhiber fortement l'activité de plusieurs enzymes impliquées dans le cycle de Krebs et dans le noeud glutamate/glutamine, à savoir :

- la succinate deshydrogénase, codée par les gènes *sdhABC* est inhibée par l'oxaloacétate [162]. Par conséquent, nos résultats théoriques montrent que la concentration de succinate est plus élevée en présence de malate ;
- la 2-oxoglutarate deshydrogénase, codée par les gènes *odhAB-pdhD* est inhibée par l'oxaloacétate [162]. Par conséquent, nos résultats théoriques montrent que la concentration de 2-oxoglutarate est plus élevée en présence de malate, ce qui bien dans le même sens que l'inhibition de l'enzyme GltAB par le malate ;
- la glutamine synthase, codée par le gène *glnA*, est fortement inhibée par la glutamine [243]. Par conséquent, nos résultats théoriques montrent que la concentration de glutamate est plus élevée en présence de malate.

Le tableau 3.10 résume l'ensemble des prédictions, qui sont de plus validées par les expériences biologiques (voir le tableau 3.11).

Résumé : A ce stade, la structure de la régulation enzymatique connue est consistente pour expliquer les concentrations élevées des métabolites intermédiaires du cycle de Krebs ainsi que de certains acides aminés comme l'aspartate, la glutamine et le glutamate. Or ces acides aminés correspondent à des acides aminés précurseurs de nombreuses autres voies métaboliques. De fait, ce sont des métabolites fortement connectés. Compte tenu des résultats mathématiques de la section 2.3.3 page 55, nous pouvons déduire que ces acides aminés

- participent à la définition du régime permanent s'ils agissent sur la première enzyme irréversible du module (voir la section 3.2.3 page 88) ;

- n'ont aucun impact sur le régime permanent des autres modules auxquels ils appartiennent si (*i*) l'enzyme en question est une enzyme intermédiaire du module et si (*ii*) cette enzyme ne sature pas.

Dans ce contexte, nous avons considéré tous les modules (ou les modules potentiels) auxquels l'aspartate (cf. la figure 3.7), la glutamine et le glutamate appartiennent (cf. la figure 3.8). L'aspartate et le glutamate jouent le rôle de donneur de groupe NH3 dans de nombreuses voies métaboliques, et participent donc à des réactions intermédiaires. Ces

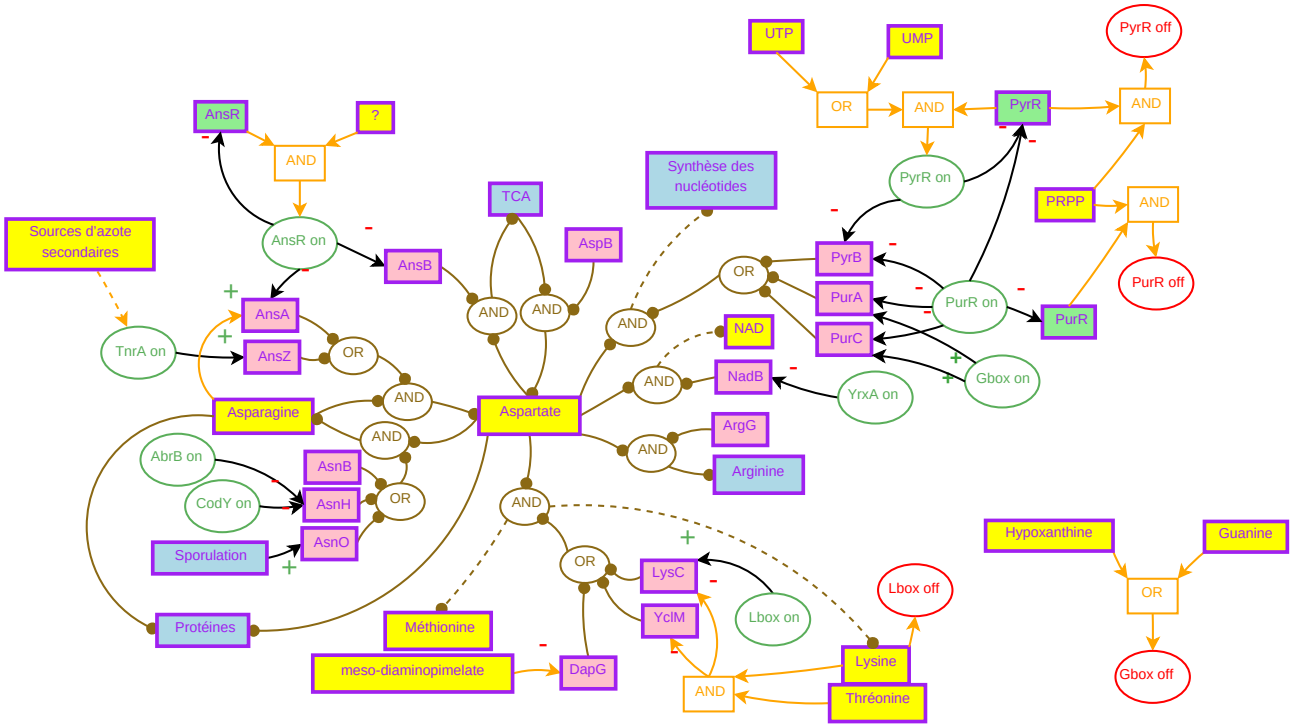


FIG. 3.7 – Les modules élémentaires impliquant l'aspartate

acides aminés correspondent néanmoins aux métabolites initiaux dans les voies métaboliques suivantes :

- **aspartate** : la synthèse de l'asparagine (*asnH*, *asnB*, *asnO*), de la lysine/thréonine (*lysC*, *dapG*, *yclM*), du NAD (*nadB*) ;
- **glutamine** : la synthèse de l'arginine (*carA*, *carB*), de l'asparagine (*asnH*, *asnB*, *asnO*) et du tryptophane (*trpE*+*trpG*) ;
- **glutamate** : la synthèse de l'arginine (*argJ*) et la synthèse de la proline (*proB*).

Nous allons maintenant prédire les conséquences des concentrations élevées d'aspartate, de glutamine et de glutamate en présence de malate sur ces voies métaboliques.

Effet de l'aspartate : Parmi les voies de synthèse associées à l'aspartate, seule la voie de la lysine (*lysC*) étudiée dans la section 3.3.1 permet de conclure. En effet, les régulations des autres gènes ne sont pas connues ou ne correspondent pas à des régulations locales, de sorte que les résultats mathématiques ne peuvent être appliqués.

Effet du glutamate et de la glutamine : Les voies de synthèse d'arginine et de proline correspondent à des de structures de contrôle par le produit final. La voie de synthèse de l'arginine est en effet composée des gènes *argJ*, *argB*, *argC*, *argD*, *argF*, *carA*, *carB* dont la transcription est inhibée par des concentrations élevées d'arginine via le facteur de transcription AhrC [146]. De même, la transcription des gènes *proB*, *proA* impliqués dans la voie de synthèse de la proline est réprimée par des hautes concentrations

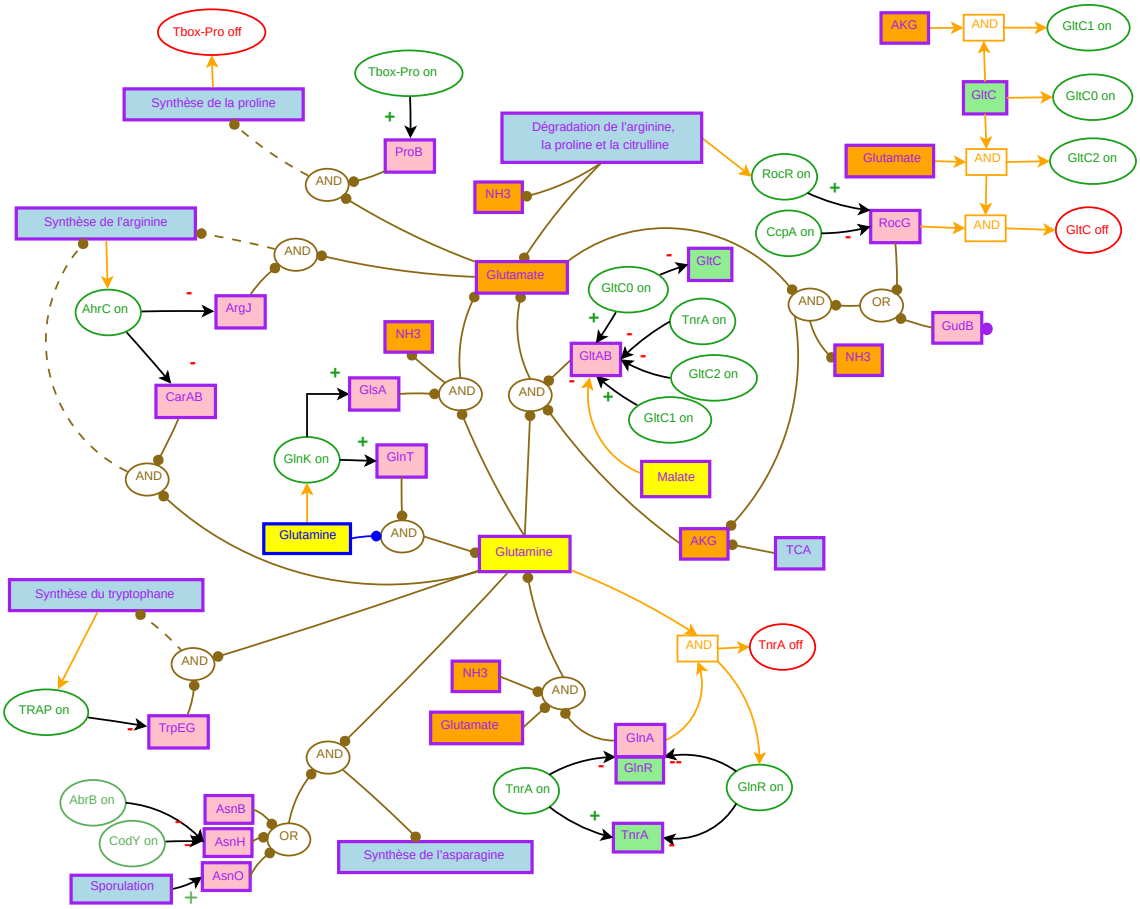


FIG. 3.8 – Les modules élémentaires impliquant le glutamate et la glutamine

d'ARNt chargés en proline via un mécanisme de T-box spécifique à la proline [38]. En tant que métabolites initiaux de structures de contrôle par le produit final, les concentrations élevées de glutamine et glutamate en présence de malate conduisent à des concentrations d'arginine et de proline d'après les résultats théoriques du corollaire 2.5. La concentration élevée de proline conduit également à augmenter la concentration d'ARNt chargés en proline. On prédit alors (voir le tableau 3.10 page 105) en présence de malate (c'est-à-dire pour des conditions de malate et malate+glucose) :

- la répression des gènes *argJ*, *argB*, *argC*, *argD*, *argF*, *carA*, *carB* ;
- la répression des gènes *proB*, *proA*.

Les résultats expérimentaux résumés dans les tableaux 3.11 et 3.12 indiquent que la concentration de proline est effectivement plus élevée en malate, malate+glucose qu'en glucose, ce qui implique une répression plus importante des gènes *proB*, *proA* en malate. La concentration d'arginine n'est malheureusement pas mesurée, mais la concentration en ornithine, un intermédiaire de la voie de l'arginine, est plus élevée en malate et malate+glucose, suggérant que la concentration en arginine évolue de même. L'expression de l'opéron *arg-carAB* est plus faible dans ces conditions, ce qui va dans le sens d'une

concentration plus élevée d'arginine.

Éléments du module	malate	glucose	malate+glucose	glc/mal	glc/mal+glc
glutamate (μmol/gdwc)	80.16	65.7	80.17	+	+
proline (μmol/gdwc)	4.13	2.9	4.32	+	+
ARNm-*proA* (log)	12.10	12.74	12.22	-	-
ARNm-*proB* (log)	12.33	13.18	12.47	-	-
glutamine (μmol/gdwc)	60	6.3	15.3	+	+
ornithine (μmol/gdwc)	5.57	2.72	5.32	+	+
ARNm-*argJ* (log)	13.45	14.01	13.53	-	-
ARNm-*argB* (log)	13.16	13.7	13.18	-	-
ARNm-*argC* (log)	12.34	13.43	12.86	-	-
ARNm-*argD* (log)	13.18	13.71	13.2	-	-
ARNm-*argF* (log)	13.3	13.9	13.47	-	-
ARNm-*carA* (log)	13.32	13.83	13.42	-	-
ARNm-*carB* (log)	13.35	13.88	13.35	-	-
tryptophane (μmol/gdwc)	4.61	1.71	3.8	+	+
ARNm-*trpE* (log)	11.21	12.3	11.55	-	-
ARNm-*trpD* (log)	11.5	12	11.6	-	-
ARNm-*trpA* (log)	13.76	14.55	14.14	-	-

TAB. 3.12 – Variation de la transcription des gènes impliqués dans (a) la synthèse de l'arginine et contrôlés par AhrC ; (b) la synthèse de la proline et régulés par une T-box spécifique proline ; (c) la synthèse du tryptophane régulés par TRAP.

Effet de la glutamine : La voie de synthèse du tryptophane correspond également à une structure de contrôle par le produit final, composée des gènes *trpE*, *trpF*, *trpG*, *trpD*, *trpB*, *trpC* et *trpA* dont la transcription est régulée par la protéine TRAP sensible à la concentration de tryptophane [77]. La concentration élevée de glutamine en présence de malate devrait conduire à une augmentation de la concentration de tryptophane, et donc à la répression de la transcription de ces gènes (voir le tableau 3.10 page 105). Si la concentration de tryptophane augmente effectivement en malate et malate+glucose (voir le tableau 3.12), son impact sur la transcription des gènes est faible (voir le tableau 3.12). La répression est toutefois légèrement plus forte en malate. La variation de la concentration en tryptophane peut être en dessous du seuil de détection de la régulation par TRAP.

Résumé : Cette section confirme le rôle central et judicieux de certaines concentrations de métabolites dans l'adaptation des flux métaboliques et du réseau génétique dans nos conditions. En conditions de malate et de glucose, les taux de croissance sont quasi identiques (0.57h^{-1} et 0.59h^{-1} respectivement). Des taux de croissance similaires impliquent un flux de production de biomasse similaire, soit encore des flux de consommation d'acides aminés pour la croissance similaires sous l'hypothèse que la composition

de la cellule n'a changé, ce qui est le cas ici. Pour les voies de synthèse de la lysine ou de l'arginine par exemple, nos résultats indiquent que pour le même flux d'acide aminé demandé pour la biomasse, le flux de synthèse correspondant est réalisé avec moins de protéines en malate qu'en glucose, grâce à une concentration de métabolite initial élevé. Ces résultats expérimentaux confirment bien le rôle central des métabolites dans l'adaptation du réseau métabolique par le biai du contrôle génétique directement par les effecteurs métaboliques des régulateurs génétiques, et indirectement via les premiers métabolites. Nous allons également constater dans la section suivante que l'adaptation générale de la cellule s'effectue principalement à un niveau métabolique.

b. Impact des métabolites sur la coordination entre les ribosomes et les voies métaboliques

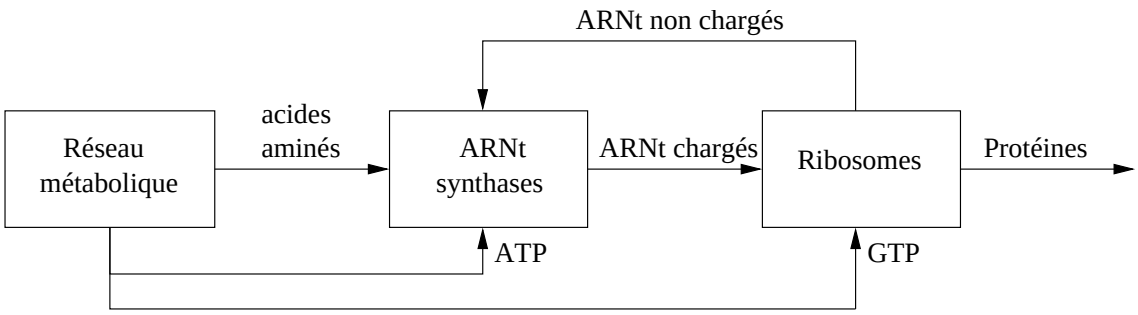


FIG. 3.9 – Une représentation systémique et simplifiée de la traduction des protéines

Le taux de croissance en malate+glucose est plus élevé ($0.75h^{-1}$) qu'en condition de malate ($0.57h^{-1}$) ou glucose seul ($0.59h^{-1}$). Le taux de croissance plus élevé en malate+glucose traduit une production de biomasse plus importante. Sachant que plus de la moitié de la biomasse est composée de protéines [153], l'augmentation du taux de croissance se traduit nécessairement par une augmentation de l'activité des ribosomes. Le flux d'ARNt chargés consommé par les ribosomes augmente. Nous allons montrer dans cette section à quel niveau s'effectue l'adaptation de l'activité des ribosomes d'une part et l'adaptation du réseau métabolique d'autre part.
Le réseau métabolique doit s'adapter pour répondre à cette demande de flux d'ARNt chargés plus élevée, à travers (*i*) l'augmentation des concentrations des enzymes des voies métaboliques, (*ii*) l'augmentation des concentrations de certains métabolites (*iii*) ou les deux. Comme indiqué dans le paragraphe précédent, la cellule adapte à la fois les concentrations enzymatiques des voies tout en utilisant au mieux l'augmentation de certains métabolites en malate+glucose.
Etudions maintenant l'augmentation de l'activité des ribosomes. Comme dans le cas du réseau métabolique, l'augmentation de l'activité des ribosomes (représentée sur la figure 3.9) peut être réalisée (*i*) en augmentant la concentration des ribosomes ; (*ii*) en augmentant l'efficacité de la traduction par les ribosomes, par exemple en augmentant les concentrations des ARNt chargés ou du GTP ; (*iii*) ou les deux. Classiquement, une aug-

mentation du taux de croissance s'accompagne par une augmentation des concentrations des ribosomes et des ARNt synthases dans la cellule [28], sauf à faible taux de croissance où les ribosomes et les ARNt synthases sont en excès dans la cellule [142, 127]. Dans ce cas, l'augmentation de l'activité des ribosomes serait réalisée par une augmentation de la concentration des ARNt chargés [142]. Dans nos conditions, l'expression des ARNt synthases et des protéines ribosomales est similaire dans les trois conditions, suggérant que nous sommes bien dans des conditions de «faible taux de croissance». De plus, les gènes codant pour les ARNt étant en opéron avec les protéines ribosomales, nous pouvons supposer que les concentrations des ARNts sont restées identiques dans nos conditions. Or le flux de synthèse des ARNt chargés obéit à l'expression suivante :

$$\nu_{aa} = E_{aa} f_{aa}(A, ATP, ARNt)$$

où E_{aa} désigne la concentration de la ARNt synthase, f_{aa} l'activité de cette enzyme, et A (resp. $ARNt$) est la concentration de l'acide aminé d'intérêt (resp. de l'ARNt correspondant). La fonction f_{aa} est une fonction croissante de ses arguments. Comme E_{aa}, et $ARNt$ sont similaires dans les trois conditions, l'augmentation du flux ν_{aa} est donc lié à l'augmentation des concentrations A, ATP ou des deux. Le tableau 3.11 indique que les concentrations des acides aminés a en effet augmenté en malate+glucose. Certaines expériences ont montré que la concentration en ATP augmente avec le taux de croissance [26]. Ainsi le flux de ARNt chargés pour les acides aminés augmente grâce aux concentrations croissantes d'acides aminés et aussi probablement d'ATP.

Les propriétés mathématiques associées aux modules se sont révélées effectives pour analyser l'adaptation des voies métaboliques dans des conditions de culture différentes. A faible taux de croissance, l'adaptation générale du métabolisme s'effectue à la fois au niveau génétique et au niveau métabolique, grâce à l'utilisation judicieuse de certaines concentrations métaboliques élevées d'une part, et aux régulations génétiques d'autre part. Ces métabolites ont un rôle stratégique certain dans l'adaptation du métabolisme car ils vont permettre d'obtenir un même flux de synthèse de précurseurs métaboliques mais avec moins d'enzymes. L'économie de protéines est réalisée grâce à la régulation génétique. Nous montrerons dans le chapitre suivant que cette stratégie d'économie en protéines est cruciale pour la bactérie car c'est un moyen effectif d'augmenter son taux de croissance.

3.4 Conclusion

La structure du contrôle des voies métaboliques de *Bacillus subtilis* apparaît donc modulaire, hautement coordonnée et hiérarchisée en fonction de quelques signaux pléiotropes bien choisis (voir la figure 3.1) comme le FBP et le GTP. De plus, le cadre formel d'analyse des voies métaboliques introduit au chapitre précédent s'est révélé effectif pour comprendre le comportement du réseau métabolique de cette bactérie dans des conditions expérimentales différentes et ce à plusieurs niveaux.
La prédiction du comportement qualitatif des différents composants d'une voie métabolique associée à leur confrontation avec des données expérimentales nous a permis de

montrer l'intérêt de notre approche, et ce à l'échelle du réseau métabolique entier. La définition et la manipulation des modules permet en régime permanent d'une part de réduire la complexité du réseau métabolique, et d'autre part d'assurer un découplage vis-à-vis des métabolites fortement connectés dont les variations pourraient entraîner un réajustement global des modules. La connectivité des voies métaboliques entre elles est donc également réduite et seuls certains métabolites spécifiques semblent jouer un rôle dans l'adaptation globale de la cellule. En pratique, il suffirait donc de mesurer uniquement ces métaboliques spécifiques ainsi que certaines grandeurs physiologiques comme les concentrations des ribosomes pour comprendre la stratégie d'adaptation du réseau métabolique face à une variation de la ressource. Par ailleurs, ce découplage du régime permanent des modules vis-à-vis des métabolites intermédiaires peut s'interpréter du point de vue de l'Automatique en terme d'une meilleure contrôlabilité des voies métaboliques.

La question est maintenant de comprendre comment une structure modulaire de contrôle a émergé. En Automatique, une telle structure de contrôle est appelée commande décentralisée. Classiquement, les méthodes efficaces de synthèse de correcteurs comme la commande H_∞ concernent les correcteurs centralisés, c'est-à-dire sans structure particulière. L'obtention d'un correcteur décentralisé «optimal» suivant un certain critère mêlant à la fois les contraintes de performance que doit remplir le correcteur et la contrainte de structure du correcteur résulte en un problème d'optimisation de complexité importante qui dans le cas général n'est pas résolu de façon efficace [200, 23, 204]. En pratique, pour obtenir un correcteur décentralisé par des algorithmes efficaces de synthèse de correcteurs classiques, il est nécessaire de mettre en place une procédure qui n'assure pas l'optimalité du correcteur synthétisé. La structure du correcteur est imposée en parallèle des contraintes de performance associées au cahier des charges[3]. Deux approches classiques sont possibles [210] : la synthèse séquentielle et la synthèse indépendante de lois de commande locale qui consiste à synthétiser indépendamment des lois de commandes locales et à vérifier ensuite les propriétés du système bouclé. Une adaptation de la seconde méthode développée dans [200] permet de synthétiser directement une loi de commande décentralisée pour un système multivariable en contraignant chaque sous-système à avoir un comportement entrée/sortie cohérent avec l'objectif global de performance [200].
Or comment une telle structure de contrôle décentralisée a-t-elle pu émerger «naturellement» à travers l'évolution ? Contrairement aux ingénieurs, la solution «optimale» de contrôle selon un certain critère résultant de l'évolution est donc structurée et satisfait aussi un critère de performance imposé par la compétition féroce entre bactéries vis-à-vis de l'accès aux ressources extracellulaires pour conquérir une niche écologique. Ainsi en régime permanent pour une condition environnementale donnée, on peut attendre que la structure du réseau de régulation est compatible avec la configuration du réseau métabolique assurant la meilleure performance.

[3]Le cahier des charges d'un correcteur correspond à l'ensemble des spécifications en termes de stabilité et de performance que doit assurer le correcteur, comme par exemple un temps de réponse limité ou encore la désensibilisation vis-à-vis de perturbations non mesurées.

En poursuivant ce parallèle avec l'Automatique, la bactérie en tant que système est forcément soumise à des contraintes structurelles qui ont permis l'émergence d'une structure de contrôle décentralisé. Le prochain chapitre de cette thèse est ainsi consacré à l'identification de ces contraintes structurelles, à leur formalisation en un problème d'optimisation et à leurs conséquences sur la cellule en tant que système. En particulier, nous nous attacherons à prédire les configurations du réseau métabolique pour différentes conditions environnementales et à confronter ces prédictions avec le réseau de régulation génétique expérimentalement identifié afin de valider le critère d'optimisation obtenu.

Annexes

3.A Identification des modules

Les modules sont décrits sur les trois tableaux donnés à partir de la page suivante.

Modules	Gènes	Régulation Locale	Régulation Globale
PTS-glucose	*ptsG, ptsH, ptsI*	GlcT	
PTS-fructose1	*fruA, ptsH, ptsI, fruK*	FruR	CcpA, Spo0A
PTS-fructose2	*levD, levE, levF, levG, ptsH, ptsI*	LevR	CcpA
PTS-sucrose	*sacP, ptsH, ptsI, sacA*	SacT, SacX	
PTS-maltose	*malP, ptsH, ptsI, malA*	GlvR	CcpA
PTS-mannose	*manP, ptsH, ptsI, manA*	ManR	CcpA
PTS-trehalose	*treP, ptsH, ptsI, treA*	TreR	CcpA
PTS-salicin-arbutin	*bglP, ptsH, ptsI, bglA*	LicT	CcpA
PTS-lichenan	*licA, licB, licC, ptsH, ptsI, licH*	LicR	CcpA
PTS-acetyl-glucosamine	*nagP, ptsH, ptsI, nagA, nagBA*	YvoA	
T-inositol	*iolF, iolT, iolG, iolE, iolD, iolC, iolJ, mmsA*	IolR	CcpA
T-sorbitol	*gutP, gutB*	GutR	CcpA
T-deoxygluconate	*kdgT, kdgK, kdgA, kdgI, kdgD*	KdgR	CcpA
T-glucu/galacturonate	*exuT, uxaC, uxaB, uxaA, uxuB, uxuA*	ExuR	CcpA
T-gluconate	*gntP, gntK, gntZ*	GntR	CcpA, CcpB
T-ribose	*rbsA, rbsB, rbsC, rbsK*	RbsR	CcpA, AbrB
T-arabinose	*araP, araQ, araE, araA, araB, araDK*	AraR	CcpA
T-xylose	*xylA, xylB*	XylR	CcpA
T-glycerol	*glpT, glpF, glpD, glpK*	GlpP	CcpA
T-malate	*maeN, yflS, maeA*	MalR	
T-succinate-fumarate	*dctP*	DctR	CcpA
T-citrate	*citM*	CitT	CcpA
T-acetoin	*ytrB, ytrC, ytrD, ytrE, ytrF*	YtrA	CcpA
Glycolyse	*fbaA, tpiA, gapA, pgk, pgm, eno*	CggR	
TCA-citrate	*mdh, citZ, citB, citC*	CcpC	CcpA
NADH-gestion	*lctP, ldh, alsS, alsD, ndh, ywcJ*	Rex	ResD, AlsR, ArfM
	cydA, cydB,cydC,cydD		
Glucosamine-6P-syn	*glmS*	Rbox-GlcN6P	

TAB. 3.13 – Décomposition du réseau métabolique de *Bacillus subtilis* en modules élémentaires

Modules	Gènes	Régulation Locale	Régulation Globale
Purine-synthesis	*pbuX, pbuO, purF, purD,purN,purL,purQ*	PurR	
	purM, purE, purK, purC, purB, purH, purA		
	glyA, guaC, pyrP, pyrAA, pyrAB, pyrB, pyrC		
	pyrK, pyrD, pyrE, pyrF, pyrR, xpt		
Gua/Hxth-syn	*pbuX, pbuG, yxjA, purF, purD,purN,purL*	Gbox	
	purQ, purM, purE, purK, purC, purB		
	purH, hprT, guaA, guaB		
Adenine-syn	*pbuE*	Abox	
UMP-synthesis	*pyrAA, pyrAB, pyrB, pyrC, pyrK*	PyrR	
	pyrD, pyrE, pyrF, pyrR		
CTP-synthesis	*pyrG*	Rbox-CTP	
Nucleoside-deg	*nupC, pdp, deoC*	DeoR	CcpA
Purine-deg	*pucK, pucJ, pucI, guaD, pucA, pucB, pucC*	PucR	GlnR, TnrA
	pucD,pucE, pucL, pucM, pucH, pucF, pucG		
	ureA, ureB, ureC		
Acides-gras-syn	*fabD, fabHA, fabHB, fabF, fabG*	FapR	
	fabZ, ycsD, fabI, plsX, plsC		
Acides-gras-deg	*lcfB, fadE, acdA, fadN, fadB, fadA*	FadR	CcpA
Arginine-tRNA	*argS*	Tbox-Arg	
Aspartate-tRNA	*aspS*	Tbox-Asp	
Glutamine-tRNA	*gatA, gatB, gatC*	Tbox-Gln	
Glycine-tRNA	*glyQ, glyS*	Tbox-Gly	
Histidine-tRNA	*hisS*	Tbox-His	
Isoleucine-tRNA	*ileS*	Tbox-Ile	
Leucine-tRNA	*leuS*	Tbox-Leu	
Phenylalanine-tRNA	*pheS, pheT*	Tbox-Phe	
Sérine-tRNA	*serS*	Tbox-Ser	
Threonine-tRNA	*thrS, thrZ*	Tbox-Thr	
Thryptophane-tRNA	*trpS*	Tbox-Trp	
Tyrosine-tRNA	*tyrS, tyrZ*	Tbox-Tyr	
Valine-tRNA	*valS*	Tbox-Val	

TAB. 3.14 – Décomposition du réseau métabolique de *Bacillus subtilis* en modules élémentaires (suite)

Modules	Gènes	Régulation Locale	Régulation Globale
Aspx-deg	*ansA, ansB*	AnsR	
Arginine-synthesis	*argJ, argB, argC, argD, argF, carA, carB*	AhrC	AbrB, CcpA, RocR
	argG, argH, argI, rocD, rocA		
Arginine-deg	*rocC, rocE, argI, rocD, rocA, rocG*	RocR	CcpA, AhrC
T-glutamine-deg	*glnT, glsA*	GlnL	
Glutamine-syn	*glnA*	GlnR	TnrA
Glutamate-syn	*gltA, gltB*	GltC	TnrA, FsrA
Glycine-deg	*gcvPA, gcvPB, gcvT*	Rbox-Gly	
Histidine-deg	*hutH, hutU, hutI, hutG*	HutP	CcpA, CodY, AbrB
Proline-syn	*proB, proA, proI*	Tbox-Pro	
Lysine-syn	*lysC, asd, dapA, dapB, dapH, patA, dapL, dapF, lysA*	Lbox	
T-cystine	*tcyJ, tcyK, tcyL, tcyM, tcyN*	YtlI	
Sulfide-syn	*cysJ, cysI*	CysL	
Cysteine-syn-tRNA	*cysE, cysK, cysS*	Tbox-Cys, CymR	
Cysteine-syn	*tcyP, tcyJ, tcyK, tcyL, tcyM, tcyN, ytlI, ssuA, ssuB, ssuC*	CymR	
	ssuD, cysK, mtnN, luxS, mccA, mccB, sat, cysC, cysH		
SAM-metabolism	*metN, metP, metQ, metI, metC, metE, metK*	Sbox	
	speD, speE, mtnN, mtnK, mtnA, mtnB, mtnW, mtnX, mtnD, mtnE		
BCAA-deg	*bcd, bkdAA, bkdAB, bkdB, lpdV*	BkdR	CodY, TnrA, CcpA
Isoleucine-syn	*ctrA, ilvA, ilvB, ilvH, ilvC, ilvD, ilvK, ilvE*	CodY	Tbox-Leu, TnrA, CcpA
Leucine-syn	*ctrA, ilvB, ilvH, ilvC, ilvD, ilvK, ilvE*	Tbox-Leu	CodY, TnrA, CcpA
	leuA, leuB, leuC, leuD		
Valine-syn	*ctrA, ilvB, ilvH, ilvC, ilvD, ilvK, ilvE*	CodY	Tbox-Leu, TnrA, CcpA
Tryptophan-syn	*trpE, pabA, trpD, trpF, trpC, trpA, trpB, trpF, hisC*	TRAP	

TAB. 3.15 – Décomposition du réseau métabolique de *Bacillus subtilis* en modules élémentaires (suite)

Deuxième partie

Vers une prédiction des modules

Chapitre 4

La méthode Resource Balance Analysis

Au regard des résultats présentés dans les deux chapitres précédents, l'évolution et la pression de sélection des systèmes biologiques ont permis l'émergence d'organismes possédant une structure de contrôle modulaire. Notre étude ayant portée uniquement sur les deux organismes modèles que sont *Bacillus subtilis* et *Escherichia coli*, nous ne pouvons bien évidemment pas conclure sur la modularité systématique du contrôle des voies métaboliques pour tous les êtres vivants. Néanmoins, cette organisation modulaire reste remarquable si on raisonne par rapport aux contrôleurs que les automaticiens savent réaliser pour les systèmes physiques. Seul l'ajout de contraintes spécifiques lors de la synthèse du correcteur par des méthodes efficaces permet d'obtenir un contrôleur structuré modulaire pour le système à contrôler [200]. En poursuivant ce parallèle entre Automatique et Biologie, nous pouvons alors imaginer que les bactéries *Bacillus subtilis* et *Escherichia coli* sont soumises à des contraintes structurelles permettant d'expliquer l'émergence d'une structure de contrôle modulaire. Si les mécanismes moléculaires précis réalisant le contrôle des voies métaboliques, c'est-à-dire l'implémentation biologique de la régulation, diffère fortement entre ces deux organismes, les configurations de la structure des régulations et la configuration du réseau métabolique associée sont relativement similaires [75, 107]. On peut alors supposer que ces deux bactéries sont soumises à des contraintes structurelles de même nature.
A ce stade, la question centrale de la suite de cette thèse consiste à identifier ces contraintes structurelles et à les formaliser mathématiquement en un problème d'optimisation dont nous explorerons les propriétés théoriques au cours de ce chapitre. Notamment, la complexité intrinsèque du problème d'optimisation sera particulièrement étudiée. De plus, nous montrerons qu'effectivement ce problème d'optimisation permet de retrouver les différentes configurations du réseau métabolique observées par les biologistes pour des conditions environnementales variées. Il ne s'agit pas de retrouver les schémas de régulations explicites, c'est-à-dire l'implémentation physique des régulations au niveau moléculaire, mais bien leur résultat final sur la configuration et le dimensionnement des voies métaboliques. En d'autres termes, les configurations des voies métaboliques obtenues par ce problème d'optimisation correspondent au dimensionnement du réseau métabolique que devrait réaliser la bactérie pour être optimale suivant le critère et les contraintes choisis. La validation du problème d'optimisation sera réalisée au cours

du chapitre suivant en comparant entre autres les configuration des voies métaboliques prédites avec le réseau de régulation expérimentalement identifié, c'est-à-dire avec les modules élémentaires identifiés au chapitre 3. Ce problème d'optimisation et ses propriétés théoriques constituent l'une des contributions principales de cette thèse et ont été publiés dans un article de conférence [76].

4.1 Identification des contraintes structurelles en phase exponentielle de croissance

Cette section est dédiée à l'identification des contraintes structurelles pouvant agir sur une bactérie comme *Bacillus subtilis* ou *Escherichia coli* grâce à l'analyse de leur fonctionnement dans une phase de vie importante : la phase exponentielle de croissance.

L'analyse du fonctionnement d'une bactérie comme *Bacillus subtilis* en phase exponentielle consiste tout d'abord à identifier les éléments nécessaires à la croissance. En phase exponentielle de croissance, le volume de la population de bactéries croît de façon exponentielle :

$$V(t) = V_0 e^{\mu t}.$$

V_0 correspond au volume initial de la population à $t = 0$. Un certain nombre de processus biologiques sont à l'œuvre pour assurer la croissance exponentielle. Par exemple, une bactérie en forme de bâtonnet comme *Bacillus subtilis* ou *Escherichia coli* doit

- répliquer son ADN ;
- produire et assembler de la paroi cellulaire et de la membrane pour s'allonger ;
- former le septum de division[1].

Bien entendu, la bactérie doit en parallèle assurer la synthèse de tous les composants nécessaires à l'ensemble de ces processus. Compte tenu de la complexité et la grande sophistication de l'ensemble des mécanismes en jeu, nous devons réduire et simplifier notre objet d'étude afin de ne saisir que l'essentiel des mécanismes à travers une analyse systémique de la cellule propre à l'Automatique.

4.1.1 Une vision systémique de la cellule

Le figure 4.1 et la figure 2 page 7 décrivent des représentations systémiques et simplifiées d'une cellule, intégrant les principales fonctions cellulaires en phase de croissance : le réseau métabolique, les appareils de transcription et de réplication des gènes, de traduction des protéines, ainsi que les machineries d'assemblage de la paroi et de la membrane cellulaire. La vision système consiste à considérer chaque sous-système cellulaire comme un bloc entrée-sortie transformant des entrées U en sorties Y. Par exemple, le sous-système «réseau métabolique» a comme entrées

[1] Le septum de division désigne la paroi transversale produite dans une bactérie lors de la division cellulaire pour séparer la bactérie en deux.

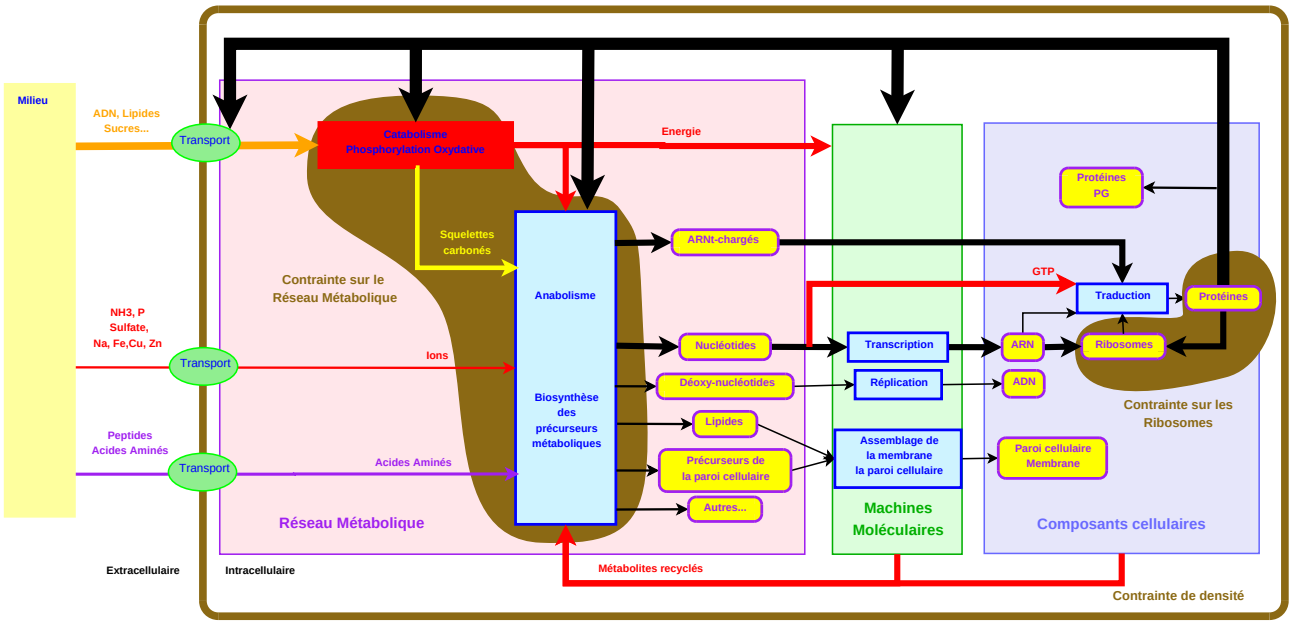


FIG. 4.1 – Une représentation systémique de la cellule

– des chaînes carbonées, des ions, des acides aminés importés depuis le milieu extérieur,
– des métabolites produits au cours de la synthèse de composants cellulaires,

et comme sorties Y tous les précurseurs métaboliques requis pour la synthèse des composés cellulaires tels que les ARN de transfert chargés (ARNtc) ou les nucléotides utilisés lors de la synthèse des protéines. Toutefois, si on s'arrête à cette description entrée/sortie du réseau métabolique, on oublie des acteurs importants, les protéines, correspondant aux enzymes et aux transporteurs. Les protéines font partie des entrées de ce sous-système cellulaire. En poursuivant cette décomposition pour chaque processus, la figure 4.1 et la figure 2 page 7 illustrent le fait remarquable suivant : tous les sous-systèmes cellulaires sont composés de protéines (et prennent donc en entrée des protéines). Les protéines constituent un couplage fort entre tous les processus cellulaires. Un processus cellulaire peut donc se représenter comme le bloc entrée/sortie de la figure 4.2 composé entre autres de protéines transformant des entrées U en sorties Y. Dans la majorité des processus biologiques composant la cellule, les entrées U correspondent à de l'énergie et aux précurseurs métaboliques utilisés. Les sorties Y correspondent aux différents composants cellulaires (ADN, ARN, protéines, etc.). A ce titre, deux processus biologiques sont remarquables :

– *l'appareil de traduction des protéines* : qui prend en entrée U de l'énergie et des ARNt chargés pour produire en sortie Y les protéines composant tous les processus cellulaires : les enzymes/transporteurs du réseau métabolique, les protéines de l'appareil de traduction (protéines ribosomales, les facteurs d'élongation, etc), et finalement toutes les protéines impliquées dans les processus d'assemblage moléculaires, ou dans les processus biologiques comme la prévention du stress ou la

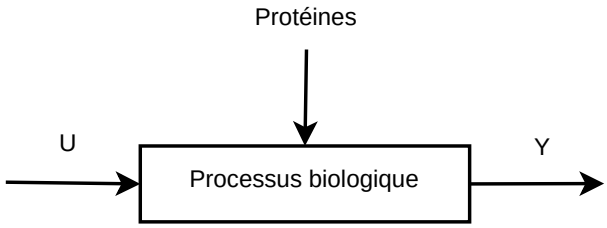


FIG. 4.2 – La représentation entrée/sortie de chaque processus cellulaire

préparation à la phase stationnaire ;
- *le réseau métabolique* : qui prend en entrée U des nutriments du milieu extracellulaire et qui produit en sortie Y les différents précurseurs métaboliques et l'énergie nécessaire à tous les processus cellulaires.

Nous venons d'isoler ici un couplage structurel agissant sur l'ensemble du système : une ressource commune, les protéines, doit être répartie entre tous les processus cellulaires de façon «intelligente» pour assurer la croissance et la survie de la bactérie.

Conséquences de la phase de croissance sur le couplage par les protéines : En phase exponentielle de croissance, la population se divise à un taux de croissance μ constant et les concentrations des différents composants cellulaires sont constantes, par définition du régime établi de la phase exponentielle. Par exemple, pour une protéine $\mathcal{P}_i$ de concentration $P_i(t)$, on a donc

$$\frac{dP_i(t)}{dt} = 0.$$

Il en va de même pour les concentrations de tous les composants cellulaires comme les métabolites, les ribosomes, etc. Ainsi en régime établi, les bactéries sont capables de maintenir constante la concentration de tous les composants cellulaires grâce à des mécanismes complexes malgré la variation de volume. Puisque par définition d'une concentration $P_i(t) = \frac{n_i(t)}{V(t)}$ où $n_i(t)$ est le nombre de protéines $\mathcal{P}_i$, on en déduit que

$$\frac{dP_i(t)}{dt} = \underbrace{\frac{dn_i(t)}{dt}\frac{1}{V(t)}}_{\text{Production}\stackrel{\Delta}{=}p_i(t)} - \underbrace{\frac{dV(t)}{dt}\frac{n_i(t)}{V^2(t)}}_{\text{Dilution}} = p_i(t) - \mu P_i(t)$$

et donc $p_i = \mu P_i$ en régime établi afin de maintenir constante la concentration P_i pour chaque bactérie malgré la variation de volume.

Remarque 4.1 *On peut rebondir ici sur les résultats théoriques présentés au cours du chapitre 2. Le terme de production instantanné $p_i(t)$ correspond pour le cas d'une enzyme aux caractéristiques $g(x(t))$ et $h(t)$ du contrôle génétique du système (2.4) page 44. La configuration du réseau métabolique et en particulier des métabolites impliqués dans le contrôle génétique doivent être tels que le régime établi existe pour un taux de croissance donné μ c'est-à-dire tels que $p_i = \mu P_i$.*

La conséquence de cette remarque pour le réseau métabolique est fondamentale. En effet, prenons le cas de n'importe quelle protéine $\mathcal{P}_i$. La synthèse d'une protéine consomme un certain nombre de métabolites comme des acides aminés sous forme d'ARNt chargés et de l'énergie sous forme de GTP principalement. Par conséquent, si une quantité α_{ki} de précurseur métabolique spécifique X_k est consommée lors de la synthèse de la protéine $\mathcal{P}_i$, alors un flux de X_k égal à $\mu\alpha_{ki}P_i$ est consommé pour maintenir la concentration P_i constante au taux de croissance μ. Concrètement, cela signifie que pour maintenir constante la concentration de la protéine $\mathcal{P}_i$, on a

- un flux de consommation associé à chaque acide aminé composant la protéine égal à $\mu\alpha_{aa,i}P_i$ où $\alpha_{aa,i}$ est le nombre de cet acide aminé composant la protéine ;
- un flux de consommation d'énergie (sous forme de GTP) correspondant à $\mu\alpha_{gtp,i}P_i$ où $\alpha_{gtp,i}$ est le nombre de GTP consommés pour produire cette protéine.

Ainsi au niveau d'un acide aminé par exemple, le flux consommé pour produire toutes les protéines de la cellule correspond à la somme de tous les flux nécessaires à la production de chaque protéine individuelle : $\mu\sum_i \alpha_{aa,i}P_i$. Si la concentration des protéines restait constante quel que soit le taux de croissance, alors ce flux d'acides aminés consommé devrait augmenter avec μ. En régime établi, le réseau métabolique doit pouvoir fournir l'ensemble des précurseurs métaboliques nécessaire à la synthèse des protéines (et plus généralement à l'ensemble de la biomasse[2]) pour un taux de croissance donné.

La remarque 4.1 a également une conséquence plus générale à l'échelle de la cellule. En effet, maintenir la concentration P_i de chaque protéine constante a non seulement un impact sur la capacité du réseau métabolique pour fournir les précurseurs métaboliques comme nous venons de le montrer, mais aussi sur tous les processus impliqués dans la synthèse même des protéines comme l'appareil de transcription et surtout l'appareil de traduction des protéines, qui rappelons le est aussi composé de protéines. Autrement dit, pour produire des protéines, il faut des protéines ! La capacité de l'appareil de traduction des protéines doit ainsi être capable de produire l'ensemble des protéines à un taux de croissance donné, et en particulier de s'auto-reproduire.

4.1.2 Répartition d'une ressource commune, les protéines, entre différents sous-systèmes

Le couplage des processus cellulaires à travers les protéines va impliquer pour la cellule une gestion rigoureuse de sa ressource en protéines. La bactérie doit en effet être considérée comme un système autonome qui s'auto-reproduit dans un environnement très compétitif vis-à-vis des nutriments extracellulaires. Il ne faut pas oublier que la bactérie est en conccurence avec ses «soeurs» pour l'accès aux ressources. De plus, le coût de production d'une protéine est élevé. Environ 85% des ressources extracellulaires sont utilisées par la bactérie pour produire des protéines. L'affectation d'une protéine à un processus cellulaire doit être «réfléchie» au regard de l'investissement requis. Rappelons

[2]La biomasse correspond à la masse sèche des cellules. Par définition, elle est composée de protéines, de lipides, de chaînes carbonées, de nucléotides, de métabolites, d'ions...

ici les aspects identifiés précédemment sur la capacité de certains processus cellulaires :
- la capacité du réseau métabolique doit pouvoir fournir l'ensemble des précurseurs métaboliques nécessaire à la biomasse pour un taux de croissance donné ;
- la capacité de l'appareil de traduction des protéines doit être capable de produire l'ensemble des protéines à un taux de croissance donné ;

Dans la section 4.2, nous formaliserons mathématiquement ces deux contraintes structurelles par des contraintes inégalités centrées sur les flux de synthèse des précurseurs métaboliques (traduisant la contrainte de capacité du réseau métabolique) et sur le flux de synthèse global d'acides aminés (traduisant la contrainte de capacité de l'appareil de traduction) pour un taux de croissance donné.
Par ailleurs, le volume de la cellule étant limité, il s'agit d'affecter entre les processus biologiques une ressource qui est non seulement commune, mais aussi limitée. Nous exprimerons cette limitation par une contrainte de densité intracellulaire, qui nous semble pertinente d'un point de vue biologique.

4.1.3 La densité de composants intracellulaires est limitée

La cellule doit également s'assurer que la densité intracellulaire reste limitée afin d'assurer la diffusion adéquate des composés cellulaires. En effet, comme tout processus chimique, les temps de réalisation des processus biologiques dépendent de la probabilité de rencontre entre plusieurs entités dans un volume donné [230]. Si ce temps moyen augmentait trop à cause d'une densité intracellulaire trop élevée, la cellule pourrait être amenée à prendre de mauvaises décisions. Dans le cas de *Escherichia coli*, la densité de la cellule reste même constante quel que soit le taux de croissance [120]. Plus généralement, les auteurs proposent que la densité intracellulaire serait constante chez les bactéries se divisant grâce à la formation d'un septum [120]. En revanche, la densité pour les bactéries en forme de coque n'est pas constante durant le cycle cellulaire. Dans tous les cas, limiter la densité permet de maintenir une diffusion adéquate des composants de la cellule [118].

4.1.4 La loi de conservation de la masse

La loi de conservation de la masse est une contrainte structurelle très forte pour le réseau métabolique et s'exprime à travers la matrice de stœchiométrie S de la façon suivante :

$$S\nu = 0,$$

où $\nu \triangleq (\nu_1, \ldots, \nu_m)^T$ désigne le vecteur des flux métaboliques composant l'ensemble du réseau métabolique en régime établi. La matrice S se déduit en remarquant que l'évolution du i-ème métabolite c_i en régime établi s'écrit :

$$\frac{dc_i}{dt} = \nu_i - \nu_{i+1} = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} \nu_i \\ \nu_{i+1} \end{bmatrix} = 0,$$

et en décrivant l'évolution de l'ensemble des métabolites, on obtient $S\nu = 0$.

4.2 Formalisation mathématique des contraintes structurelles

Les contraintes structurelles introduites dans les sections 4.1.2, 4.1.3 et 4.1.4 sont intrinsèquement liées à la capacité de croissance de la bactérie en phase exponentielle. Nous allons donc maintenant formaliser de façon mathématique les contraintes structurelles identifiées précédemment :

- le compromis sur la répartition des protéines entre les voies métaboliques, les ribosomes et les autres processus cellulaires ;
- la conservation de la masse dans la cellule ;
- la limitation de la densité intracellulaire.

4.2.1 Notations

Au cours de la section 4.1, nous avons réparti les protéines en trois classes de processus biologiques : le réseau métabolique, l'appareil de traduction des protéines et les autres processus. Ce dernier ensemble de protéines est noté $\mathcal{P}_G$ sur la figure 4.1. Introduisons tout d'abord les notations pour ces trois ensembles.
Le réseau métabolique est composé de :

- N_i métabolites internes $X_i \triangleq (X_{i_1}, \ldots, X_{i_{N_i}})$;
- N_p précurseurs métaboliques $X_p \triangleq (X_{p_1}, \ldots, X_{p_{N_p}})$ consommés au cours de la synthèse des composants cellulaires ;
- N_r métaboliques «recyclés» $X_r \triangleq (X_{r_1}, \ldots, X_{r_{N_r}})$ produits durant la synthèse de ces mêmes composants cellulaires ;
- N_c macro-composants $X_c \triangleq (X_{c_1}, \ldots, X_{c_{N_c}})$ comme la paroi cellulaire ou la membrane dont la concentration intracellulaire est indépendante du taux de croissance ;
- N_m enzymes $\mathbb{E} \triangleq (\mathbb{E}_1, \ldots, \mathbb{E}_{N_m})$ de concentration $E \triangleq (E_1, \ldots, E_{N_m})$, dont les flux métaboliques associés correspondent à $\nu \triangleq (\nu_1, \ldots, \nu_{N_m})$. Par simplicité, on notera m la dimension du vecteur ν à la place de N_m dans la suite de ce chapitre.

Le réseau métabolique sera décrit par sa matrice de stœchiométrie S de dimension $(N_i + N_p + N_r + N_c) \times m$.
La concentration de l'appareil de traduction est assimilée à la concentration du principal acteur dans la traduction des protéines, les ribosomes, et sera notée R.
L'ensemble de dimension N_G des protéines n'appartenant ni au réseau métabolique, ni à l'appareil de traduction sera désigné par $\mathcal{P}_G$ et la concentration de la protéine $\mathcal{P}_{G_i}$ appartenant à $\mathcal{P}_G$ sera notée P_{G_i}. Par définition, cet ensemble est composé de protéines essentielles en phase de croissance comme par exemple les protéines impliquées dans la réplication de l'ADN, mais aussi de protéines potentiellement dispensables mais qui sont nécessaires pour la survie de la bactérie sur un long terme. On citera par exemple les protéines impliquées dans la prévention du stress.

Couplage entre la concentration d'une enzyme et le flux métabolique

Les deux chapitres précédents ont montré les liens existants entre flux et concentration de l'enzyme. Nous supposons dans la suite que $E_j(t)$, la concentration de l'enzyme $\mathbb{E}_j$ à l'instant t est reliée au flux instantané $\nu_j(t)$ par la relation suivante

$$\nu_j(t) = \pm k_{E_j} E_j(t) \tag{4.1}$$

où $k_{E_j} > 0$ correspond à l'efficacité de l'enzyme. A travers cette relation, nous supposons que les enzymes fonctionnent en régime quasi saturé. Cette hypothèse est supportée par de récents résultats montrés dans [156] pour une enzyme spécifique de la voie des pentoses phosphate Zwf chez *E. coli*.
Evaluons maintenant l'impact de la variation de volume pour cette enzyme sous cette hypothèse. En régime établi, la bactérie doit maintenir la concentration de l'enzyme constante, ce qui au niveau du réseau métabolique, se traduit par la consommation de différents précurseurs métaboliques. Si une quantité α_{ij} de précurseur métabolique X_{p_i} est consommée lors de la synthèse de l'enzyme $\mathbb{E}_j$, alors le flux ν_{ij} de métabolite X_{p_i} nécessaire pour maintenir la concentration E_j constante à l'équilibre est donné par

$$\nu_{ij} = \mu \alpha_{ij} E_j = \mu \alpha_{ij} \frac{|\nu_j|}{k_{E_j}}. \tag{4.2}$$

Sous cette hypothèse, augmenter les flux métaboliques dans une voie métabolique est équivalent à augmenter la concentration des enzymes. Cette dépendance par un paramètre fixe k_E suppose que l'adaptation des flux métaboliques s'effectue essentiellement au niveau génétique, par la modulation des concentrations des enzymes. Cette hypothèse semble licite pour des taux de croissance élevés, et nécessite une validation expérimentale (actuellement en cours de réalisation). Pour des taux de croissance plus faibles comme pour les conditions expérimentales du chapitre précédent, l'adaptation du réseau métabolique se fait aussi en partie par les métabolites. Nous montrerons dans ce chapitre et dans le chapitre suivant que malgré cette représentation simplifiée de la cellule, les résultats obtenus sont tout à fait pertinents lorsqu'ils sont comparés aux données issues de la littérature.

4.2.2 De la répartition des protéines à la formalisation de trois contraintes structurelles

Pour augmenter son taux de croissance, la bactérie doit augmenter globalement le flux de synthèse de biomasse, et donc à ce titre les flux de synthèse des voies métaboliques. Le partage de la ressource en protéines dans la cellule doit alors garantir que :

- la capacité du réseau métabolique augmente pour augmenter le flux de production des précurseurs métaboliques. Sous l'hypothèse donnée par l'équation 4.1, cela signifie que la concentration des enzymes augmente ;
- la capacité totale de production des protéines, c'est-à-dire la concentration des protéines de l'appareil de traduction augmente également.

Ainsi, un compromis existe quant à l'utilisation des précurseurs métaboliques suivant s'ils participent à l'élaboration de protéines dédiées au réseau métabolique, ou dédiées à l'appareil de traduction des protéines (principalement les ribosomes).
Ce compromis est traduit en deux contraintes, (C_1) et (C_2). L'une, (C_1), porte sur la capacité du réseau métabolique et sur son fonctionnement, et l'autre (C_2) sur la capacité de l'appareil de traduction à produire les protéines. La contrainte (C_1) est également éclatée en 4 sous-contraintes, chacune caractérisant une contrainte de capacité spécifique à laquelle est soumis le réseau métabolique. Enfin la contrainte (C_3) décrit la limitation de la densité intracellulaire.

Contrainte (C_1) : «Capacité du réseau métabolique»

En régime établi, c'est-à-dire à μ fixé, différents points doivent être assurés :

(C_{1a}) : la capacité du réseau métabolique doit être suffisante pour produire tous les précurseurs métaboliques requis pour la croissance de la cellule, y compris ceux utilisés dans la traduction des protéines, ou par les machineries moléculaires. Formellement, cette contrainte signifie que le flux de synthèse des N_p précurseurs métaboliques produits par le réseau métabolique doit être supérieur au flux consommé par la synthèse des composants cellulaires.
Pour tout $i \in \{1, \ldots, N_p\}$,

$$-\sum_{j=1}^{m} S_{p_{ij}} \nu_j + \mu \left(\sum_{j=1}^{m} C_{M_{ij}}^{M_p} |\nu_j| + C_{R_i}^{M_p} R + \sum_{j=1}^{N_G} C_{G_{ij}}^{M_p} P_{G_j} \right) - \nu_Y \leq 0$$

où S_p est la sous-matrice de S de dimension $N_p \times m$ spécifique des métabolites X_p, et ν_Y correspond à un flux «gratuit» de métabolites échangé avec l'environnement par diffusion. Les cœfficients $C_{R_i}^{M_p}$ et $C_{G_{ij}}^{M_p}$ sont positifs et correspondent respectivement au nombre de précurseurs métaboliques numéro i (X_{p_i}) requis pour la synthèse d'un ribosome et de la j-ème protéine appartenant à $\mathcal{P}_G$. Le cœfficient $C_{M_{ij}}^{M_p}$ est positif et défini par $C_{M_{ij}}^{M_p} = \frac{\alpha_{ij}^p}{k_{E_j}}$ où α_{ij}^p est le nombre de précurseurs métaboliques numéro i (X_{p_i}) requis pour la synthèse de la j-ème protéine appartenant au réseau métabolique.

(C_{1b}) : la capacité du réseau métabolique doit être suffisante pour maintenir constante la concentration d'un ensemble de macro-composants X_c (membrane, paroi cellulaire) à la valeur $\bar{X}_c \triangleq (\bar{X}_{c_1}, \ldots, \bar{X}_{c_{N_c}})$, où $\bar{X}_{c_i}$ désigne la concentration cible du i-ème macro-composant X_{c_i}.
Pour tout $i \in \{1, \ldots, N_c\}$,

$$-\sum_{j=1}^{m} S_{c_{ij}} \nu_j + \mu \bar{X}_{c_i} \leq 0$$

où S_c est la sous-matrice de S de dimension $N_c \times m$ spécifique des macro-composants X_c.

(C_{1c}) : la capacité du réseau métabolique doit être suffisante pour absorber tous les N_r métabolites recyclés produits lors de la synthèse des composants cellulaires.
Pour tout $i \in \{1, \ldots, N_r\}$,

$$\sum_{j=1}^{m} S_{r_{ij}} \nu_j + \mu \left(\sum_{j=1}^{m} C_{M_{ij}}^{M_r} |\nu_j| + C_{R_i}^{M_r} R + \sum_{j=1}^{N_G} C_{G_{ij}}^{M_r} P_{G_j} \right) \leq 0$$

où S_r est la sous-matrice de S de dimension $N_r \times m$ spécifique des métabolites X_r. Les cœfficients $C_{R_i}^{M_r}$, $C_{G_{ij}}^{M_r}$ sont également positifs et correspondent respectivement au nombre de métabolites recyclés numéro i (X_{r_i}) produits durant la synthèse d'un ribosome et de la j-ème protéine appartenant à $\mathcal{P}_G$. Le cœfficient $C_{M_{ij}}^{M_r}$ est positif et défini par $C_{M_{ij}}^{M_r} = \frac{\alpha_{ij}^r}{k_{E_j}}$ où α_{ij}^r est le nombre de métabolites recyclés numéro i (X_{r_i}) produits durant la synthèse de la j-ème protéine appartenant au réseau métabolique.

(C_{1d}) : les flux métaboliques doivent satisfaire la loi de conservation de la masse en régime établi.
Pour tout $i \in \{1, \ldots, N_i\}$,

$$\sum_{j=1}^{m} S_{I_{ij}} \nu_j = 0$$

où S_I est la sous-matrice de S de dimension $N_i \times m$ spécifique des métabolites internes X_i.

Contrainte (C_2) : «Capacité de l'appareil de traduction»

La capacité de l'appareil de traduction doit être suffisante pour maintenir constante la concentration de toutes les protéines de la cellule en régime établi au taux de croissance μ :

$$\mu \left(\sum_{j=1}^{m} C_{M_j}^{R} |\nu_j| + C_R^R R + \sum_{j=1}^{N_G} C_{G_j}^{R} P_{G_j} \right) - k_T R \leq 0$$

où k_T désigne l'efficacité de la traduction (autour de 12 à 20 acides aminés par seconde à 37°C [28]). Cette efficacité de traduction a été déterminée sur les ribosomes actifs de la cellule. Or environ 20% des ribosomes de la cellule en phase exponentielle sont en cours de synthèse et donc non fonctionnels. Nous verrons dans la section 4.7 comment prendre en compte ce phénomène : cela revient à modifier k_T et ne change pas la nature de la contrainte. Les cœfficients C_R^R et $C_{G_j}^R$ sont strictement positifs et correspondent respectivement au nombre total d'acides aminés par ribosome et pour la j-ème protéine appartenant à $\mathcal{P}_G$. Le cœfficient $C_{M_j}^R$ est strictement positif et défini par $C_{M_j}^R = \frac{n_j^R}{k_{E_j}}$ où n_j^R est le nombre total d'acides aminés de la j-ème protéine appartenant au réseau métabolique.

Contrainte (C_3) : «Limitation de la densité intracellulaire»

La densité de la cellule est limitée pour assurer la diffusion adéquate de tous les composants cellulaires (protéines, métabolites, ions, etc) [18, 232].

$$\sum_{j=1}^{m} C_{M_j}^D |\nu_j| + C_R^D R + \sum_{j=1}^{N_G} C_{G_j}^D P_{G_j} - \bar{D} \leq 0$$

où $\bar{D}$ est la densité moyenne équivalente en protéines de la cellule. Les cœfficients $C_{G_j}^D$ et $C_{M_j}^D$ sont respectivement égaux à $C_{G_j}^R$ et $C_{M_j}^R$ et C_D^R correspond à la densité équivalente en protéines d'un ribosome.

Des notations vectorielles introduites pour décrire l'ensemble $\mathcal{P}_G$: Dans la suite, pour alléger les notations, on utilisera les vecteurs suivants pour manipuler les cœfficients associés à P_G dans les différentes contraintes :

- $C_{G_i}^x \triangleq (C_{G_{i1}}^x, \ldots, C_{G_{iN_G}}^x)$ pour $x \in \{M_p, M_r\}$;
- $C_G^x \triangleq (C_{G_1}^x, \ldots, C_{G_{N_G}}^x)$ pour $x \in \{R, D\}$.

De plus, on définit $P_G \triangleq (P_{G_1}, \ldots, P_{G_{N_G}})$ et on désignera par $P_G \geq 0$, l'ensemble des inégalités $P_{G_j} \geq 0$ pour $j \in \{1, \ldots, N_G\}$.

4.2.3 Un problème de faisabilité à μ donné : $P_f(\mu)$

Conséquence de la phase de croissance équilibrée pour les nutriments extracellulaires

Nous supposons que les ressources extracellulaires sont constantes : les ressources ne s'épuisent pas, et le flux d'entrée de ces nutriments est supposé dépendre uniquement de la concentration des transporteurs.

Formalisation du problème d'optimisation : «Resource Balance Analysis»

Si la cellule peut croître avec un taux de croissance μ, alors les trois contraintes structurelles (C_1), (C_2) et (C_3) sont nécessairement satisfaites. Ainsi pour un taux de croissance μ donné, on va chercher s'il existe une répartition des ressources (ν, R) qui satisfasse les trois contraintes, ce qui conduit au problème d'optimisation $P_f(\mu)$ suivant :

Pour $P_G \geq 0$, $\mu \geq 0$ fixés,

$$\begin{array}{ll} \text{trouver} & R \geq 0, \nu \in \mathcal{R}^m \\ \text{soumis à} & \\ & (C_{1a}), (C_{1b}), (C_{1c}), (C_{1d}), (C_2), (C_3). \end{array}$$

Définissons également l'ensemble des solutions faisables associées à $P_f(\mu)$:

$$\mathcal{C}_{\mu, P_{G_1}, \ldots, P_{G_{N_G}}, k_{E_1}, \ldots, k_{E_m}, k_T} = \{(R, \nu) \in \mathcal{R}^+ \times \mathcal{R}^m | (C_{1a}),\ (C_{1b}),\ (C_{1c}),\ (C_{1d}),\ (C_2),\ (C_3)\}.$$

Dans la suite de ce chapitre, nous noterons $\mathcal{C}_{\mu,P_G}$ à la place de $\mathcal{C}_{\mu,P_{G_1},\ldots,P_{G_{N_G}},k_{E_1},\ldots,k_{E_m},k_T}$ pour alléger les notations.

Puisque notre méthode intègre de manière explicite la gestion des ressources, nous l'appellerons par la suite la méthode «**Resource Balance Analysis**» (**RBA**).

$P_f(\mu)$ peut se mettre sous la forme suivante correspondant d'après [154] à un problème de faisabilité :

$$\begin{array}{ll} \text{trouver} & \xi \in \mathcal{R}^{m+1} \\ \text{tel que} & \\ & \xi \in \mathcal{C} \end{array}$$

où $\mathcal{C}$ désigne un ensemble de n contraintes telles que :

$$\mathcal{C} \triangleq \{\xi \in \mathcal{R}^{m+1} \quad | \quad f_i(\xi) \leq 0 \text{ pour } i \in \{1, \ldots, n\}\}.$$

Par définition, $P_f(\mu)$ est dit faisable s'il a une solution et dans ce cas $\mathcal{C}_{\mu,P_G} \neq \emptyset$.
De plus, $P_f(\mu)$ est un problème de faisabilité non-différentiable à cause de la présence de la valeur absolue dans les différentes contraintes. Heureusement, cette valeur absolue ne constitue pas un obstacle à une analyse de $P_f(\mu)$, qui va nous permettre de déduire des propriétés remarquables dans la suite de ce chapitre.

4.3 Un problème sympatique à résoudre

Nous étudions dans cette section les propriétés de $P_f(\mu)$ vis-à-vis de μ intéressantes aussi bien au niveau théorique que pour leur implication au niveau biologique :

1. si $P_f(\mu)$ est faisable, alors la solution faisable $(\bar{\nu}, \bar{R})$ est non nulle ;
2. $P_f(\mu)$ est convexe ;
3. si $P_f(\mu^+)$ est faisable alors $P_f(\mu)$ est également faisable pour tout $\mu \in [0, \mu^+)$;
4. il existe $\mu^* > 0$ tel que pour $\mu > \mu^*$, $P_f(\mu)$ est infaisable.

A chaque étape, le résultat théorique obtenu sera discuté d'un point de vue biologique de manière à souligner explicitement les aspects biologiques garantis par les mathématiques.

Soient les 6 ensembles suivants : $I_m \triangleq \{1, \ldots, m\}$, $I_p \triangleq \{1, \ldots, N_p\}$, $I_r \triangleq \{1, \ldots, N_r\}$, $I_i \triangleq \{1, \ldots, N_i\}$ $I_c \triangleq \{1, \ldots, N_c\}$, et $I_g \triangleq \{1, \ldots, N_G\}$.

4.3.1 Toute solution faisable est non nulle

Proposition 4.1 *Si pour $\mu > 0$ et pour $P_G \geq 0$, $P_f(\mu)$ est faisable alors tout couple de solution faisable $(\bar{R}, \bar{\nu}) \in \mathcal{C}_{\mu,P_G}$ est tel que $\bar{R} > 0$ et $\bar{\nu} \neq 0$, c'est-à-dire qu'il existe un sous-ensemble non vide U de I_m, tel que pour tout $j \in U$, $\bar{\nu}_j \neq 0$, et pour tout $k \in I_m/U$, $\bar{\nu}_k = 0$.*

Démonstration
La démonstration est faite par l'absurde. En effet, la contrainte (C_{1a}), pour $\mu > 0$ fixé, $\bar{\nu} = 0$ et pour tout $\bar{R} \geq 0$ conduit à $\mu \bar{X}_{c_i} \leq 0$, donc une contradiction est obtenue. Nous venons donc de prouver qu'il existe un sous-ensemble non vide U de I_m, tel que pour tout $j \in U$, $\bar{\nu}_j \neq 0$.
Pour conclure, si nous supposons que $\bar{R} = 0$ alors (C_2) correspond à $\mu \sum_{j \in U_1} C^R_{M_j} |\bar{\nu}_j| \leq 0$, ce qui conduit à une nouvelle contradiction.

■

La proposition 4.1 souligne explicitement que si $P_f(\mu)$ est faisable alors toute solution faisable est non nulle. D'un point de vue biologique, cela signifie que si la bactérie est capable de croître au taux de croissance μ, alors il existe une répartition de ressources non nulle pour permettre la croissance cellulaire, et plus particulièrement une concentration non nulle de ribosomes et d'un sous-ensemble d'enzymes et de transporteurs du réseau métabolique. Malgré l'évidence biologique d'un tel résultat, il corrobore la formalisation mathématique choisie et surtout les contraintes structurelles intégrées dans le problème d'optimisation.

4.3.2 Un résultat technique intéressant

Lemme 4.2 *Si pour $P_G \geq 0$ et $\mu > 0$ fixés, $\mathcal{C}_{\mu,P_G} \neq \emptyset$ et pour $(\bar{R}, \bar{\nu}) \in \mathcal{C}_{\mu,P_G}$, les contraintes (C_{1a}), (C_{1c}), (C_2) et (C_3) sont strictes, alors il existe $\delta\mu > 0$ tel que $\mathcal{C}_{\mu+\delta\mu,P_G} \neq \emptyset$.*

La démonstration du lemme 4.2 est reportée en annexe 4.A de ce chapitre page 155. Ce lemme indique que si toutes les inégalités (C_{1a}), (C_{1c}), (C_2) et (C_3) sont strictes, alors il existe une valeur de μ supérieure pour laquelle $P_f(\mu)$ reste faisable. En revanche, si l'une de ces inégalités n'est pas stricte, on ne peut pas conclure. Ce résultat sera utilisé par la suite dans plusieurs démonstrations.

4.3.3 Convexité de $P_f(\mu)$

Prouvons maintenant la convexité de l'ensemble $\mathcal{C}_{\mu,P_G}$. Dans ce cas, le problème de faisabilité associé $P_f(\mu)$ est aussi convexe.

Proposition 4.3 *Pour tout $P_G \geq 0$ et pour tout $\mu \geq 0$, $\mathcal{C}_{\mu,P_G}$ est convexe.*

Démonstration
Soient $(\bar{R}', \bar{\nu}'), (\bar{R}'', \bar{\nu}'') \in \mathcal{C}_{\mu,P_G}$. Pour conclure que $\mathcal{C}_{\mu,P_G}$ est convexe, il suffit de montrer que pour tout $\lambda \in [0,1]$, $(\bar{R}^\lambda, \bar{\nu}^\lambda) \triangleq \left(\lambda \bar{R}' + (1-\lambda)\bar{R}'', \lambda \bar{\nu}' + (1-\lambda)\bar{\nu}''\right) \in \mathcal{C}_{\mu,P_G}$.
Soit $i \in I_p$, et montrons que $(\bar{R}^\lambda, \bar{\nu}^\lambda)$ satisfait la i-ème contrainte de (C_{1a}). Puisque pour tout $j \in I_m$ les cœfficients $C^{M_p}_{M_{ij}}$ sont positifs, nous avons pour tout $j \in I_m$ et $\lambda \in [0,1]$,

$$C^{M_p}_{M_{ij}} |\lambda \bar{\nu}'_j + (1-\lambda)\bar{\nu}''_j| \leq \lambda C^{M_p}_{M_{ij}} |\bar{\nu}'_j| + (1-\lambda) C^{M_p}_{M_{ij}} |\bar{\nu}''_j|.$$

On en déduit donc que pour $i \in I_p$, la contrainte (C_{1a}) obtenue pour $(\bar{R}^\lambda, \bar{\nu}^\lambda)$ est bornée supérieurement par la somme des deux contraintes (C_{1a}) obtenues respectivement pour $\lambda(\bar{R}', \bar{\nu}')$ et pour $(1-\lambda)(\bar{R}'', \bar{\nu}'')$. Comme $(\bar{R}', \bar{\nu}'), (\bar{R}'', \bar{\nu}'') \in \mathcal{C}_{\mu,P_G}$, les contraintes (C_{1a}) obtenues pour $\lambda(\bar{R}', \bar{\nu}')$ d'une part, et pour $(1-\lambda)(\bar{R}'', \bar{\nu}'')$ d'autre part sont faisables, on en déduit que la contrainte (C_{1a}) obtenue pour $(\bar{R}^\lambda, \bar{\nu}^\lambda)$ est aussi faisable.
Pour tout $i \in I_c$, les contraintes (C_{1b}) sont affines en ν_j. Donc si ces contraintes sont satisfaites pour $\nu_j = \bar{\nu}_j'$ et $\nu_j = \bar{\nu}_j''$, alors elles sont aussi satisfaites pour $\nu_j^\lambda = \lambda\bar{\nu}_j' + (1-\lambda)\bar{\nu}_j''$.
On peut prouver de la même façon que $(\bar{R}^\lambda, \bar{\nu}^\lambda)$ satisfait aussi les contraintes C_{1c}, C_{1d}, C_2 and C_3.
Nous pouvons donc conclure que $(\bar{R}^\lambda, \bar{\nu}^\lambda) \in \mathcal{C}_{\mu,P_G}$ et donc que $\mathcal{C}_{\mu,P_G}$ est un ensemble convexe.

■

Cette proposition est fondamentale dans le contexte de l'optimisation (voir la section 4.4). Elle indique que $P_f(\mu)$ se résout en temps polynomial même à l'échelle de la cellule [154]. La présence de la valeur absolue dans les différentes contraintes n'a aucun impact sur la convexité de $P_f(\mu)$. Montrons maintenant qu'il existe une borne supérieure atteignable pour μ vis à vis d'un ensemble de ressources donné.

4.3.4 L'ensemble des μ faisables est compact

Nous allons maintenant montrer que l'ensemble des μ tels que $P_f(\mu)$ soit faisable est un ensemble compact.

Proposition 4.4 *$P_f(\mu)$ a les propriétés suivantes :*

1. *Si pour $P_G \geq 0$ et pour $\mu^+ > 0$, $P_f(\mu^+)$ est faisable alors pour tout $\mu \in [0, \mu^+]$, $P_f(\mu)$ est aussi faisable et $\mathcal{C}_{\mu^+,P_G} \subseteq \mathcal{C}_{\mu,P_G}$;*
2. *Pour tout $P_G \geq 0$, il existe $\mu^* \geq 0$ fini tel que $P_f(\mu^*)$ est faisable et pour tout $\mu > \mu^*$, $P_f(\mu)$ n'est pas faisable.*

Démonstration

Démonstration du point 1 : Soit $\delta\mu > 0$ et $\mu > 0$ tel que $\mu^+ = \mu + \delta\mu$. La démonstration de cette proposition est directe en substituant μ^+ par $\mu + \delta\mu$ dans toutes les contraintes et en utilisant les signes positifs et strictement positifs des différents cœfficients dans $P_f(\mu)$.

Démonstration du point 2 : La démonstration de cette proposition est faite en deux étapes. On montre d'abord que la borne supérieure existe, puis qu'elle est atteinte.

Existence d'une borne supérieure pour μ. La démonstration est faite par l'absurde en montrant que s'il existe $\mu^* > 0$, tel que pour tout $\mu > \mu^*$, $P_f(\mu)$ n'est pas faisable.
Grâce à la contrainte (C_3), puisque C_G^D, C_R^D et pour tout $j \in I_m$ les cœfficients $C_{M_j}^D$ sont tous strictement positifs, on peut déduire qu'il existe des bornes supérieures finies $\bar{\nu}_j^{max}$, $\bar{R}^{max}$ pour chaque $\bar{\nu}_j$ et $\bar{R}$ telles que $\bar{R} \leq \bar{R}^{max}$ et pour tout $j \in I_m$, $|\bar{\nu}_j| \leq \bar{\nu}_j^{max}$.
La contrainte (C_2) devient ainsi $\mu(\sum_{j=1}^m C_{M_j}^R|\bar{\nu}_j| + C_R^R\bar{R} + C_G^R P_G^T) \leq k_T\bar{R}^{max}$. Pour conclure, nous devons montrer que $\mu(\sum_{j=1}^m C_{M_j}^R|\bar{\nu}_j| + C_R^R\bar{R} + C_G^R P_G^T)$ est une fonction non bornée de μ. Ce résultat est évident si $P_G > 0$. Si $P_G = 0$, la contrainte (C_{1b}) implique qu'au moins certains flux ν_j sont nécessairement des fonctions croissantes de μ et la contradiction est obtenue. Donc il existe une valeur $\mu^* \geq 0$ finie telle que pour tout $\mu > \mu^*$, $\mathcal{C}_{\mu,P_G} = \emptyset$.

$P_f(\mu^*)$ est faisable. Si $\mu^* = 0$, le résultat est obtenu puisque $(0,0)$ appartient à $\mathcal{C}_{0,P_G}$. Supposons maintenant que $\mu^* > 0$ et que pour tout $\mu > \mu^*$, $P_f(\mu)$ n'est pas faisable. La démonstration est faite par l'absurde. Si $P_f(\mu^*)$ n'est pas faisable, alors il existe $\epsilon > 0$ tel qu'au moins une des inégalités $(C_{1a}), (C_{1b}), (C_{1c})$ ou (C_2) est supérieure ou égale à ϵ pour tout (R, ν). Pour tout $\delta\mu > 0$ suffisamment petit, il existe $(\bar{R}, \bar{\nu}) \in \mathcal{C}_{\mu^*-\delta\mu,P_G}$ tel que les contraintes $(C_{1a}), (C_{1b}), (C_{1c})$ or (C_2) sont satisfaites pour $\mu = \mu^* - \delta\mu$.
De plus, $\bar{R}$ and $\bar{\nu}$ sont uniformément bornées vis-à-vis de μ puisque grâce au point 1 de cette proposition, l'ensemble des solutions faisables associé à des valeurs croissantes de μ est inclus dans les ensembles de solutions faisables associés à des valeurs de μ inférieures. Donc il existe $M > 0$ (indépendant de μ) tel que pour tout $\delta\mu > 0$ il existe $(\bar{R}, \bar{\nu}) \in \mathcal{C}_{\mu^*-\delta\mu,P_G}$, tel que

$$
\begin{array}{ll}
(C_{1a}) & \text{pour tout } i \in I_p, \\
 & -\sum_{j=1}^m S_{p_{ij}}\bar{\nu}_j + \mu^*\left(\sum_{j=1}^m C_{M_{ij}}^{M_p}|\bar{\nu}_j| + C_{R_i}^{M_p}\bar{R} + C_{G_i}^{M_p}P_G^T\right) - \nu_Y \leq M\delta\mu \\
(C_{1b}) & \text{pour tout } i \in I_c, \\
 & -\sum_{j=1}^m S_{c_{ij}}\bar{\nu}_j + \mu^*\bar{X}_{c_i} \leq M\delta\mu \\
(C_{1c}) & \text{pour tout } i \in I_r, \\
 & \sum_{j=1}^m S_{r_{ij}}\bar{\nu}_j + \mu^*\left(\sum_{j=1}^m C_{M_{ij}}^{M_r}|\bar{\nu}_j| + C_{R_i}^{M_r}\bar{R} + C_{G_i}^{M_r}P_G^T\right) \leq M\delta\mu \\
(C_2) & \mu^*(\sum_{j=1}^m C_{M_j}^R|\bar{\nu}_j| + C_R^R\bar{R} + C_G^R P_G^T) - k_T\bar{R} \leq M\delta\mu
\end{array}
$$

Quel que soit $\epsilon > 0$, en choisissant $\delta\mu$ tel que $\delta\mu < \epsilon/M$, on obtient une contradiction ce qui conclut cette démonstration.

■

La traduction biologique du premier point de cette proposition signifie que s'il existe une répartition de ressources pour un taux de croissance donné, alors elle existe aussi pour des valeurs de taux de croissance inférieures. Concrètement, on prédit ici que la bactérie est capable de croître avec un taux de croissance μ et aussi pour toute valeur de taux de croissance inférieure. Là encore, ce résultat est trivial d'un point de vue biologique, mais il renforce bien la pertinence des contraintes choisies.

Le deuxième point de la proposition indique qu'il existe un taux de croissance maximal atteignable pour un ensemble de conditions environnementales donné. La prise en compte des trois contraintes structurelles (C_1), (C_2) et (C_3) impose un compromis sur le devenir des protéines, ce qui intrinsèquement limite (et fixe) le taux de croissance. Grâce aux propositions 4.3 et 4.4, on en déduit la valeur maximale du taux de croissance μ^* peut être calculée par dichotomie. Pour un ensemble de ressources environnementales donné, le taux de croissance maximal que peut atteindre la bactérie peut être approché par la résolution itérative de problèmes d'optimisation convexe non différentiables. On obtient également intrinsèquement la répartition en protéines de la cellule, c'est-à-dire les concentrations des ribosomes, des enzymes et des transporteurs.

Un cadre formel révélant des principes fondamentaux de fonctionnement pour les bactéries

Ces résultats sont particulièrement intéressants car ils offrent un cadre rationnel d'explication à la limitation structurelle du taux de croissance. Les biologistes ont en effet constaté que le taux de croissance des bactéries varie en fonction du milieu extracellulaire, et notamment vis-à-vis de la composition de celui-ci [184]. Néanmoins, la question n'avait jamais été abordée sous l'angle formel, ni du point de vue de l'existence de contraintes structurelles limitant le taux de croissance. La méthode RBA s'inscrit ainsi pleinement dans un des objectifs de la Biologie des Systèmes correspondant à la recherche de principes fondamentaux de fonctionnement des cellules vivantes. Nous ajouterons ici que la vision système propre à l'Automatique a joué un rôle central pour l'identification des différentes contraintes et notamment pour le partage de la ressource en protéines. Il est clair que la vision système, les méthodes et outils associés à l'Automatique apporte un avantage certain pour l'analyse des systèmes biologiques.

4.4 Équivalence entre $P_f(\mu)$ et un problème de Programmation Linéaire

Dans les sections 4.2.3 et 4.3.3, nous avons souligné que $P_f(\mu)$ correspond à un problème de faisabilité non-différentiable et convexe. La convexité est une propriété cruciale pour la résolution des problèmes d'optimisation [154]. Afin de bien saisir l'apport de la convexité dans le contexte de l'optimisation et par conséquent pour $P_f(\mu)$, nous allons rapidement introduire les principales notions d'optimisation nécessaires à la compréhension des résultats théoriques de cette section.
Trouver le minimum d'une fonction peut s'écrire comme le problème d'optimisation suivant :

$$\begin{array}{ll} \text{minimiser} & h(\xi) \\ \text{tel que} & \\ & \xi \in \mathcal{R}^m \end{array}$$

où h désigne la fonction à minimiser, encore appelée fonction de coût. La fonction h est continue et différentiable de $\mathcal{R}^m$ dans $\mathcal{R}$, et m correspond à la dimension du problème

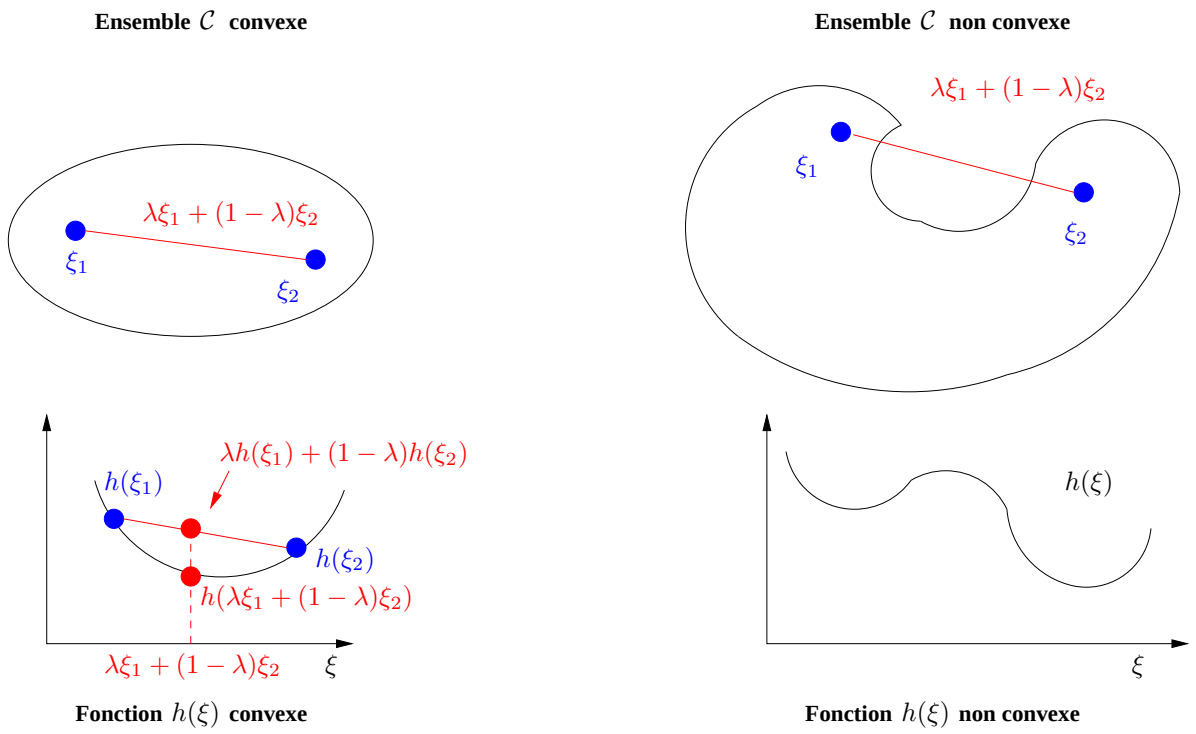


FIG. 4.3 – Exemple de deux ensembles $\mathcal{C}$ convexe (en haut à gauche) et non convexe (en haut à droite) et de deux fonctions $h(x)$ convexe (en bas à gauche) et non convexe (en bas à droite)

d'optimisation. Si de plus, l'ensemble des solutions faisables est contraint par des inégalités, par exemple par l'ensemble $\mathcal{C} \triangleq \{\xi \in \mathcal{R}^m \quad | \quad f_i(\xi) \leq 0 \text{ pour } i \in \{1, \ldots, n\}\}$ où les fonctions f_i sont également des fonctions continues et différentiables de $\mathcal{R}^m$ dans $\mathcal{R}$, on parle alors de problème d'optimisation sous contraintes :

$$\begin{array}{ll} \text{minimiser} & h(\xi) \\ \text{tel que} & \\ & \xi \in \mathcal{C}. \end{array}$$

Généralement, la résolution d'un tel problème d'optimisation par la méthode du gradient ou de newton par exemple ne conduisent pas forcément à approcher l'optimum global si l'ensemble $\mathcal{C}$ et la fonction h ne possèdent pas de bonnes propriétés [22, 154]. S'il existe des minima locaux à la fonction h, en fonction du point initial choisi, on peut tomber sur un minimum local (voir la figure 4.3 en bas à droite). Trouver alors le minimum global pour ce problème d'optimisation nécessite par exemple de tirer aléatoirement des conditions initiales qui vont paver l'ensemble $\mathcal{C}$, ou de pouvoir sortir d'un minimum local [154]. Le temps de résolution associé à ces méthodes croît de façon exponentielle avec la taille du problème d'optimisation [154].

En revanche si le problème d'optimisation est convexe, c'est-à-dire si l'ensemble $\mathcal{C}$ et la fonction de coût h sont convexes, alors certaines propriétés sont garanties. Par définition,

la fonction $h(\xi)$ est convexe si pour tout ξ_1, ξ_2 appartenant à $\mathcal{C}$ et pour tout $\lambda \in]0,1[$, $h(\lambda\xi_1 + (1-\lambda)\xi_2) \leq \lambda h(\xi_1) + (1-\lambda)h(\xi_2)$. Tout minimum local correspond alors à un minimum global. De même, l'ensemble $\mathcal{C}$ est convexe si pour tout ξ_1, ξ_2 appartenant à $\mathcal{C}$ et pour tout $\lambda \in]0,1[$, $\xi_\lambda \triangleq \lambda\xi_1 + (1-\lambda)\xi_2$ appartient aussi à $\mathcal{C}$. Concrètement, cela signifie que le segment reliant deux points appartenant à un ensemble convexe appartient aussi à cet ensemble (voir la figure 4.3 en haut à gauche), soit encore que pour tout $i \in \{1, \ldots, n\}$

$$f_i(\xi_\lambda) \leq 0$$

Cette propriété est très forte : l'ensemble des solutions faisables peut être parcouru dans son ensemble pour toute condition initiale appartenant à cet ensemble. Un problème d'optimisation convexe se résoud numériquement en un temps raisonnable, c'est-à-dire que le temps de résolution est une fonction polynômiale de la taille du problème d'optimisation [21, 24, 154]. Malheureusement, il n'existe pas forcément d'algorithmes efficaces pour résoudre n'importe quel problème d'optimisation convexe. La convexité nous garantit uniquement que si le minimum existe dans l'ensemble convexe, il est atteint en temps polynômial. Le degré de performance de l'algorithme de résolution peut être très élevé et conduit alors à des temps effectifs de résolution trop longs.

Parmi l'ensemble des problèmes d'optimisation convexe, les problèmes de programmation linéaire (LP) sont ceux que l'on peut le plus efficacement résoudre. Un problème de programmation linéaire correspond au problème d'optimisation sous contraintes suivant :

$$\begin{array}{ll} \text{minimiser} & c^T\xi \\ \text{tel que} & \\ & a_i^T\xi \leq b_i, \quad i = 1, \ldots, p \end{array}$$

où ξ désigne toujours le vecteur de variables de décision appartenant à $\mathcal{R}^m$, les vecteurs c, $a_1, \ldots, a_p$ appartiennent à $\mathcal{R}^m$, et les scalaires $b_1, \ldots, b_p$ appartiennent à $\mathcal{R}$. Généralement, on peut noter l'ensemble des inégalités sous forme matricielle par $A\xi \leq b$ où $A \in \mathcal{R}^{p\times m}$ et $b \in \mathcal{R}^p$. Par définition du problème de programmation linéaire, la fonction objectif et les contraintes inégalité intégrées dans un problème de programmation linéaire sont affines vis-à-vis des variables de décision ξ. Cette classe de problèmes d'optimisation convexe ayant été particulièrement étudiée [24, 154], des algorithmes efficaces en temps de résolution polynômial basés sur la méthode du point intérieur ont été développés [21, 154, 24, 74]. Ces algorithmes restent efficaces en particulier pour des problèmes de très grande dimension. A ce titre, ces algorithmes sont donc particulièrement bien appropriés pour résoudre des problèmes d'optimisation centrés sur les systèmes biologiques dont la dimension peut rapidement être très grande.

Dans cette section, nous montrons qu'au delà de ses propriétés remarquables vis-à-vis de μ (convexité, unicité et atteignabilité d'une valeur maximale μ^*), $P_f(\mu)$ est aussi équivalent à un problème de programmation linéaire $P_f^{lp}(\mu)$, que l'on peut résoudre efficacement. D'un point de vue pratique, cette équivalence est fondamentale car nous allons disposer d'un outil prédictif effectif que nous exploiterons au cours du prochain chapitre.

Soit $P_f^{lp}(\mu)$ le problème de faisabilité LP suivant :

trouver	$R \geq 0, \nu \in \mathcal{R}^m, \nu^{max} \in \mathcal{R}_+^m$
soumis à	
(C_{1a}^{lp})	pour tout $i \in I_p$, $-\sum_{j=1}^m S_{p_{ij}}\nu_j + \mu\left(\sum_{j=1}^m C_{M_{ij}}^{M_p}\nu_j^{max} + C_{R_i}^{M_p}R + C_{G_i}^{M_p}P_G^T\right) - \nu_Y \leq 0$
(C_{1b}^{lp})	pour tout $i \in I_c$, $-\sum_{j=1}^m S_{c_{ij}}\nu_j + \mu\bar{X}_{c_i} \leq 0$
(C_{1c}^{lp})	pour tout $i \in I_r$, $\sum_{j=1}^m S_{r_{ij}}\nu_j + \mu\left(\sum_{j=1}^m C_{M_{ij}}^{M_r}\nu_j^{max} + C_{R_i}^{M_r}R + C_{G_i}^{M_r}P_G^T\right) \leq 0$
(C_{1d}^{lp})	pour tout $i \in I_i$, $\sum_{j=1}^m S_{I_{ij}}\nu_j = 0$
(C_2^{lp})	$\mu(\sum_{j=1}^m C_{M_j}^R\nu_j^{max} + C_R^R R + C_G^R P_G^T) - k_T R \leq 0$
(C_3^{lp})	$\sum_{j=1}^m C_{M_j}^D\nu_j^{max} + C_R^D R + C_G^D P_G^T - \bar{D} \leq 0$
(C_4^{lp})	pour tout $j \in I_m$, $\nu_j - \nu_j^{max} \leq 0$ et $-(\nu_j + \nu_j^{max}) \leq 0$

et l'ensemble constitué d'égalités et d'inégalités suivant :

$$\mathcal{C}_{\mu,P_G}^{lp} = \{(R,\nu) \in \mathcal{R}^+ \times \mathcal{R}^m | (C_{1a}^{lp}),\ (C_{1b}^{lp}),\ (C_{1c}^{lp}),\ (C_{1d}^{lp}),\ (C_2^{lp}),\ (C_3^{lp})\}.$$

Proposition 4.5 *Pour $P_G \geq 0$ et $\mu \geq 0$ fixés, $\mathcal{C}_{\mu,P_G} = \mathcal{C}_{\mu,P_G}^{lp}$.*

Démonstration

L'équivalence est montrée en deux temps.

(si) Montrons d'abord que $\mathcal{C}_{\mu,P_G}^{lp} \subseteq \mathcal{C}_{\mu,P_G}$.

Supposons que pour $\mu \geq 0$, $P_G \geq 0$ fixés, $\mathcal{C}_{\mu,P_G}^{lp} \neq \emptyset$. Soit $(\bar{R},\bar{\nu}) \in \mathcal{C}_{\mu,P_G}^{lp}$. Alors il existe $\bar{\nu}^{max}$ tel que (C_{1a}^{lp}), (C_{1b}^{lp}), (C_{1c}^{lp}), (C_{1d}^{lp}), (C_2^{lp}), (C_3^{lp}) et (C_4^{lp}) sont satisfaites. Les contraintes (C_4^{lp}) impliquent que pour tout $j \in I_m$, $|\bar{\nu}_j| \leq \bar{\nu}_j^{max}$.

Puisque dans les contraintes (C_{1a}^{lp}), (C_{1c}^{lp}), (C_2^{lp}) et (C_3^{lp}), (i) les cœfficients multipliant $\bar{\nu}_j^{max}$ sont positifs ; et (ii) $\bar{R}$, $\bar{\nu}$, $\bar{\nu}^{max}$ satisfont ces contraintes, on en déduit que $\bar{R}$, $\bar{\nu}$ satisfont aussi(C_{1a}), (C_{1c}), (C_2) and (C_3). De plus, les contraintes (C_{1b}) et (C_{1b}^{lp}) ainsi que (C_{1d}) et (C_{1d}^{lp}) sont les mêmes. On en déduit donc que (C_{1b}) et (C_{1d}) sont aussi satisfaites pour $\bar{\nu}$.

Donc $(\bar{R},\bar{\nu}) \in \mathcal{C}_{\mu,P_G}$. On en conclut que $\mathcal{C}_{\mu,P_G}^{lp} \subseteq \mathcal{C}_{\mu,P_G}$.

(seulement si) Montrons maintenant que $\mathcal{C}_{\mu,P_G} \subseteq \mathcal{C}_{\mu,P_G}^{lp}$.

Supposons que pour $\mu \geq 0$ et $P_G \geq 0$ fixés, $\mathcal{C}_{\mu,P_G} \neq \emptyset$. Soit $(\bar{R},\bar{\nu}) \in \mathcal{C}_{\mu,P_G}$.

Il suffit d'introduire pour chaque $j \in I_m$, une variable positive ν_j^{max} telle que $\nu_j^{max} = |\nu_j|$. Cette équation satisfait de manière évidente les contraintes (C_4^{lp}) de $P_f^{lp}(\mu)$. Les autres

contraintes définissant $\mathcal{C}^{lp}_{\mu,P_G}$ sont obtenues par la substitution directe de $|\nu_j|$ par ν_j^{max}. Donc $(\bar{R}, \bar{\nu}) \in \mathcal{C}^{lp}_{\mu,P_G}$ ce qui conclut cette démonstration.

■

La proposition 4.5 démontre que $P_f(\mu)$ est équivalent au problème de programmation linéaire $P_f^{lp}(\mu)$ [21, 154, 24]. Les propositions 4.1, 4.3 et 4.4 obtenues pour $P_f(\mu)$ peuvent être étendues à $P_f^{lp}(\mu)$. On en déduit donc que pour un ensemble de ressources extérieures donné, le taux de croissance maximal atteignable μ^* peut être calculé par dichotomie à travers la résolution itérative de $P_f^{lp}(\mu)$ pour chaque valeur de μ ainsi que les concentrations des protéines du réseau métabolique et des ribosomes pour μ^*.
Par rapport au problème d'optimisation initial $P_f(\mu)$ qui comprenait N_m+1 variables de décision, $N_p+N_r+N_c+2$ contraintes inégalités et N_i contraintes égalités, $P_f^{lp}(\mu)$ possède $2N_m+1$ variables de décision, $N_p+N_r+N_c+2+2*N_m$ contraintes inégalités et N_i contraintes égalités. La dimension du problème de programmation linéaire croît de façon polynômiale avec la dimension du problème initial. Nous sommes donc dans un cadre d'équivalence sympathique puisque le nombre de variables de décision et de contraintes du problème de programmation linéaire n'augmente pas de façon exponentielle.

Remarque 4.2 *$P_f(\mu)$ calcule une répartition de ressources faisable pour un taux de croissance donné μ, c'est-à-dire la concentration des ribosomes et des protéines des voies métaboliques. Pour les voies métaboliques principales de Bacillus subtilis, cette répartition des ressources correspond à 359 variables de décision (voir le chapitre suivant), ce qui ne représente pas en-soi un problème de grande dimension. Néanmoins, la représentation de la cellule considérée est simplifiée, et nous verrons dans la section 4.7 comment intégrer certains phénomènes biologiques qui augmentent la dimension du problème d'optimisation. A ce titre, l'équivalence de $P_f(\mu)$ avec un problème de programmation linéaire reste de notre point de vue le meilleur choix possible en termes d'outils associés à l'optimisation, compte tenu de la très grande dimension théorique des systèmes biologiques. Cela nous laisse ainsi une grande marge de manœuvre pour complexifier notre modèle.*

4.5 Différents mécanismes pour économiser des protéines

Grâce à notre méthode RBA, nous disposons à la fois (*i*) d'un outil de prédiction permettant pour la première fois de prédire non seulement pour la valeur maximale du taux de croissance pour un ensemble de ressource donné, mais aussi pour les concentrations des enzymes et des ribosomes, et (*ii*) d'un cadre formel révélant des principes fondamentaux du fonctionnement des bactéries en phase exponentielle de croissance comme la limitation structurelle du taux de croissance à travers un compromis sur l'allocation des protéines produites pour les différents processus cellulaires. La gestion des protéines est par conséquent critique pour la cellule. Dans ce contexte, on peut imaginer que tout mécanisme modifiant l'allocation des protéines à un processus biologique comme par

exemple les régulations génétiques, a un impact sur le taux de croissance. Dans cette section, nous allons effectivement montrer qu'au delà de la limitation structurelle du taux de croissance, la méthode RBA permet aussi d'expliquer l'existence de certains mécanismes qui permettent de gérer au mieux la disponibilité de la ressource en protéines, et dont l'impact sur le taux de croissance est critique.

Notre analyse s'effectue en deux temps. Tout d'abord, nous montrons grâce à l'analyse théorique de $P_f(\mu)$ que la répression de sous-systèmes métaboliques permet d'augmenter le taux de croissance. D'un point de vue biologique, ceci signifie que la bactérie développe des stratégies comme les régulations génétiques pour éteindre des voies métaboliques lorqu'elles ne sont pas indispensables afin d'économiser des protéines et donc d'augmenter le taux de croissance. Notre méthode RBA permet ainsi de retrouver les configurations des voies métaboliques connues pour différents milieux. Ces configurations prédites coïncident bien, comme nous le verrons par la suite, aux configurations métaboliques observées grâce au réseau de régulation génétique connu [75]. Par abus de language, on parlera par la suite de prédiction du réseau de régulation ou de prédiction des modules. Néanmoins, et ce point est crucial, il ne s'agit pas ici de prédire l'implémentation biologique de la régulation en tant que telle (facteur de transcription, riboswitch, etc.), mais bien du régime établi final obtenu correspondant à un ensemble de sous-systèmes métaboliques allumés ou éteints. Ces différentes configurations sont obtenues grâce au coût intrinsèque en protéines différent entre les voies métaboliques intégré dans le problème d'optimisation.

Remarque 4.3 *Pour la cellule, une régulation génétique réalise une économie de protéines à deux niveaux. Non seulement les protéines en tant que telles sont économisées, et donc des précurseurs métaboliques et des ribosomes sont économisés et peuvent être affectés à la synthèse d'autres protéines (contraintes (C_1) et (C_2)), mais aussi elles ne contribuent plus à l'augmentation de la densité intracellulaire. Dans la suite de cette thèse, on utilisera le terme «économie en protéines» pour désigner cette double économie (précurseurs et densité). De même, le terme «coût d'induction» d'une protéine (ou d'une voie métabolique) fera référence à la quantité de précurseurs métaboliques et de ribosomes nécessaires à la synthèse de cette protéine (ou d'une voie métabolique) ainsi qu'à son impact sur la densité intracellulaire.*

Mais les stratégies générales d'économie en protéines ne sont pas réalisées uniquement au niveau des régulations génétiques. De façon remarquable, la structure du critère d'optimisation montre aussi que l'efficacité des ribosomes et l'efficacité des enzymes ont été optimisées par l'évolution afin de minimiser la quantité de protéines nécessaire à la réalisation de la fonction [52, 91, 245, 45] (voir la section 4.5.2). Si ces différents aspects sont intégrés dans $P_f(\mu)$, les échelles de temps nécessaires à l'émergence d'une régulation génétique et à l'optimisation des efficacités des enzymes et des ribosomes ne sont sans doute pas comparables. Bien entendu, il existe aussi d'autres contraintes agissant sur le système non intégrées dans notre problème d'optimisation impliquant notamment la survie sur un long terme ou les stratégies de coordination entre les bactéries. Néanmoins, la structure actuelle du critère permet déjà de révéler des principes généraux de

fonctionnement.

4.5.1 Prédiction théorique des structures de régulation dans le réseau métabolique

Le problème de faisabilité $P_f(\mu)$ permet également de gérer la priorité d'entrée des ressources extérieures en fonction du coût de leur voie d'assimilation et de leur voie de synthèse *de novo* respectivement. Une voie métabolique est effectivement composée de plusieurs enzymes, chacune ayant des caractéristiques distinctes (la composition en acides aminés, la longueur, etc.). Plusieurs voies métaboliques peuvent conduire à la production du même métabolite. Par conséquent, le choix entre deux voies métaboliques est crucial pour la cellule si le taux de croissance est affecté.
Ce problème peut être facilement résolu numériquement grâce à la proposition 4.5 à travers la résolution de $P_f^{lp}(\mu)$. Le problème est beaucoup plus complexe d'un point de vue analytique. En effet, chaque voie métabolique comprend généralement de nombreux co-métabolites comme l'ATP ou le NADPH. Ainsi, le choix d'une voie de synthèse optimale implique également l'identification des voies optimales pour ces co-métabolites, ce qui reste difficile analytiquement. Cependant, un résultat théorique peut être obtenu comme une preuve de principe dans un cas simplifié.

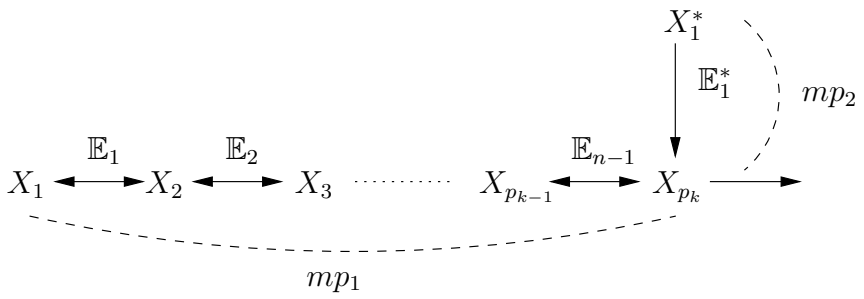


FIG. 4.4 – Exemple de voies métaboliques linéaires mp_1 et mp_2

Considérons deux voies métaboliques mp_1 et mp_2, représentées sur la figure 4.4, linéaires distinctes de synthèse du k-ème précurseur métabolique X_{p_k} avec $k \in I_p$ et supposons que mp_2 est plus courte que mp_1, *i.e* les deux voies ont la même composition moyenne en acides aminés, mais moins de protéines sont impliquées dans mp_2 que dans mp_1 pour produire le même flux de X_{p_k}. On supposera que mp_1 et mp_2 sont composées respectivement de N_1 et N_2 enzymes telles que $N_2 < N_1$. Par exemple, dans le cas où X_{p_k} serait un acide aminé, mp_1 et mp_2 pourraient correspondre respectivement à une voie de synthèse *de novo* et à un transporteur.

Corollaire 4.6 *Supposons $P_G \geq 0$ et $\mu > 0$ fixés et $P_f(\mu)$ est faisable avec la voie mp_1 active, et la voie mp_2 inactive, alors il existe $\delta\mu > 0$ tel que $P_f(\mu + \delta\mu)$ est faisable, mp_2 est active et mp_1 inactive.*

Le corollaire 4.6 est une conséquence de la proposition reportée en annexe 4.B de ce chapitre page 156 obtenue dans le cas plus général où la composition en protéines entre les deux voies est différente. Le corollaire 4.6 indique que choisir une voie métabolique «économe» en terme de précurseurs métaboliques comparée à une voie «chère» permet d'augmenter le taux de croissance. Concrètement, si on reprend notre exemple de deux voies de synthèse d'un acide aminé soit par un unique transporteur, soit par une voie de synthèse *de novo*, la résolution de $P_f(\mu)$ conduira à éteindre complètement la voie de synthèse *de novo* et à allumer le transporteur si l'acide aminé est présent dans le milieu extracellulaire. Toutefois, la présence des co-facteurs dans les voies métaboliques complexifie la solution analytique. Une solution de $P_f(\mu)$ pourrait induire à la fois les chemins «économe» et «cher» si le coût de production d'un co-métabolite, produit par la voie «chère» est plus faible que celui obtenu par toute autre voie de production alternative. La distribution globale des flux métaboliques obtenue au cours de la résolution de $P_f(\mu)$ est ainsi fortement contrainte à la fois par la stœchiométrie du réseau métabolique et par le coût en précurseurs métaboliques de l'induction de chaque voie métabolique. En conclusion, nous prédisons à travers ce résultat que la bactérie peut développer des stratégies telles que les régulations génétiques pour moduler la distribution des flux métaboliques à travers les voies métaboliques en fonction de leur coût de production.

4.5.2 Sensibilité de la solution à des variations de P_G, k_{E_i} et k_T

Pour des raisons de lisibilité de la proposition 4.8, nous utilisons les notations spécifiques suivantes pour les variations de k_{E_i} et k_T : $\mathcal{C}_{\mu,\bar{k}_{E_i}} \triangleq \mathcal{C}_{\mu,P_G,k_{E_1},\ldots,\bar{k}_{E_i},\ldots,k_{E_m},k_T}$ et $\mathcal{C}_{\mu,\bar{k}_T} \triangleq \mathcal{C}_{\mu,P_G,k_{E_1},\ldots,k_{E_m},\bar{k}_T}$.
Par ailleurs, nous introduisons l'hypothèse suivante :

Hypothèse 4.7 *Pour tout $i \in I_p$, les cœfficients $C^{M_p}_{M_{ij}}$ pour tout $j \in I_m$, $C^{M_p}_{R_i}$ et $C^{M_p}_{G_{ij}}$ pour tout $j \in I_g$ sont strictement positifs. Pour tout $i \in I_r$, les cœfficients $C^{M_r}_{M_{ij}}$ pour tout $j \in I_m$, $C^{M_r}_{R_i}$ et $C^{M_r}_{G_{ij}}$ pour tout $j \in I_g$ sont strictement positifs.*

Dans notre cas, cette hypothèse implique que toutes les protéines sont composées de tous les acides aminés, ce qui est généralement le cas.

Proposition 4.8 *Si l'hypothèse 4.7 est satisfaite et si pour $\mu > 0$, $P_G \geq 0$, $k_T > 0$ et $k_{E_i} > 0$ pour $i \in I_m$, $P_f(\mu)$ est faisable, alors :*

1. *Pour tout $0 \leq \delta P_G \leq P_G$[3], il existe $\delta\mu > 0$ tel que $\mathcal{C}_{\mu+\delta\mu,P_G-\delta P_G} \neq \emptyset$;*
2. *Pour tout $\bar{k}_{E_i} > k_{E_i}$, il existe $\delta\mu > 0$ tel que $\mathcal{C}_{\mu+\delta\mu,\bar{k}_{E_i}} \neq \emptyset$;*
3. *Pour tout $\bar{k}_T > k_T$, il existe $\delta\mu > 0$ tel que $\mathcal{C}_{\mu+\delta\mu,\bar{k}_T} \neq \emptyset$ et $\delta R > 0$ tel que $(\bar{R} - \delta R, \bar{\nu}) \in \mathcal{C}_{\mu,\bar{k}_T}$ où $(\bar{R}, \bar{\nu}) \in \mathcal{C}_{\mu,k_T}$.*

[3]Comme pour le vecteur de concentrations P_G, la notation $\delta P_G \geq 0$ signifie que pour tout $i \in I_g$, $\delta P_{G_i} \geq 0$.

Démonstration

Démonstration du point 1 : Pour démontrer ce résultat, nous allons construire un point faisable appartenant à l'ensemble $\mathcal{C}_{\mu+\delta\mu,P_G-\delta P_G}$.
Supposons que pour $P_G > 0$ et $\mu > 0$ fixés, $\mathcal{C}_{\mu,P_G} \neq \emptyset$ et soit $(\bar{R}, \bar{\nu}) \in \mathcal{C}_{\mu,P_G}$. Soit $\delta P_G > 0$ tel que $\bar{P}_G \triangleq P_G - \delta P_G \geq 0$. En substituant P_G par $\bar{P}_G$ dans les différentes contraintes, on obtient :

(C_{1a}^{δ}) pour tout $i \in I_p$
$-\sum_{j=1}^{m} S_{p_{ij}}\bar{\nu}_j + \mu\left(\sum_{j=1}^{m} C_{M_{ij}}^{M_p}|\bar{\nu}_j| + C_{R_i}^{M_p}\bar{R} + C_{G_i}^{M_p}\bar{P}_G^T\right) - \nu_Y \leq -\mu C_{G_i}^{M_p}\delta P_G^T < 0$

(C_{1b}^{δ}) et pour tout $i \in I_c$,
$-\sum_{j=1}^{m} S_{c_{ij}}\bar{\nu}_j + \mu\bar{X}_{c_i} \leq 0$

(C_{1c}^{δ}) pour tout $i \in I_r$
$\sum_{j=1}^{m} S_{r_{ij}}\bar{\nu}_j + \mu\left(\sum_{j=1}^{m} C_{M_{ij}}^{M_r}|\bar{\nu}_j| + C_{R_i}^{M_r}\bar{R} + C_{G_i}^{M_r}\bar{P}_G^T\right) \leq -\mu C_{G_i}^{M_r}\delta P_G^T < 0$

(C_{1d}^{δ}) pour tout $i \in I_i$
$\sum_{j=1}^{m} S_{I_{ij}}\bar{\nu}_j = 0$

(C_2^{δ}) $\mu(\sum_{j=1}^{m} C_{M_j}^{R}|\bar{\nu}_j| + C_R^R\bar{R} + C_G^R\bar{P}_G^T) - k_T\bar{R} \leq -\mu C_G^R\delta P_G^T < 0$

(C_3^{δ}) $\sum_{j=1}^{m} C_{M_j}^{D}|\bar{\nu}_j| + C_R^D\bar{R} + C_G^D\bar{P}_G^T - \bar{D} \leq -C_G^D\delta P_G^T < 0$

Si l'hypothèse 4.7 est vraie, alors les contraintes (C_{1a}^{δ}), (C_{1c}^{δ}), (C_2^{δ}) et (C_3^{δ}) sont strictes. Nous pouvons donc déduire grâce au lemme 4.2, qu'il existe $\delta\mu > 0$ tel que $\mathcal{C}_{\mu+\delta\mu,\bar{P}_G} \neq \emptyset$, ce qui conclut la démonstration du point 1 de cette proposition.

Démonstration du point 2 : Soit $l \in I_m$ et $\bar{k}_{E_l} > k_{E_l}$. En notant que $\dfrac{1}{\bar{k}_{E_l}} < \dfrac{1}{k_{E_l}}$ et sous l'hypothèse 4.7, les contraintes (C_{1a}), (C_{1c}), (C_2), (C_3) écrites pour $\bar{k}_{E_l}$ sont strictes.

pour tout $i \in I_p$
$-\sum_{j=1}^{m} S_{p_{ij}}\bar{\nu}_j + \mu\left(\dfrac{\alpha_{il}^p}{\bar{k}_{E_l}}|\bar{\nu}_l| + \sum_{j=1,j\neq l}^{m} C_{M_{ij}}^{M_p}|\bar{\nu}_j| + C_{R_i}^{M_p}\bar{R} + C_{G_i}^{M_p}P_G^T\right) - \nu_Y < \ldots$
$\quad -\sum_{j=1}^{m} S_{p_{ij}}\bar{\nu}_j + \mu\left(\dfrac{\alpha_{il}^p}{k_{E_l}}|\bar{\nu}_l| + \sum_{j=1,j\neq l}^{m} C_{M_{ij}}^{M_p}|\bar{\nu}_j| + C_{R_i}^{M_p}\bar{R} + C_{G_i}^{M_p}P_G^T\right) - \nu_Y \leq 0$
et pour tout $i \in I_c$,
$-\sum_{j=1}^{m} S_{c_{ij}}\bar{\nu}_j + \mu\bar{X}_{c_i} \leq 0$
pour tout $i \in I_r$
$\sum_{j=1}^{m} S_{r_{ij}}\bar{\nu}_j + \mu\left(\dfrac{\alpha_{il}^r}{\bar{k}_{E_l}}|\bar{\nu}_l| + \sum_{j=1,j\neq l}^{m} C_{M_{ij}}^{M_r}|\bar{\nu}_j| + C_{R_i}^{M_r}\bar{R} + C_{G_i}^{M_r}P_G^T\right) < \ldots$
$\quad \sum_{j=1}^{m} S_{r_{ij}}\bar{\nu}_j + \mu\left(\dfrac{\alpha_{il}^r}{k_{E_l}}|\bar{\nu}_l| + \sum_{j=1,j\neq l}^{m} C_{M_{ij}}^{M_r}|\bar{\nu}_j| + C_{R_i}^{M_r}\bar{R} + C_{G_i}^{M_r}P_G^T\right) \leq 0$
pour tout $i \in I_i$
$\sum_{j=1}^{m} S_{I_{ij}}\bar{\nu}_j = 0$

$$\mu(\frac{n_j^R}{\bar{k}_{E_l}}|\bar{\nu}_l| + \sum_{j=1,j\neq l}^m C_{M_j}^R|\bar{\nu}_j| + C_R^R\bar{R} + C_G^R P_G^T) - k_T\bar{R} < \dots$$

$$\mu(\frac{n_j^R}{k_{E_l}}|\bar{\nu}_l| + \sum_{j=1,j\neq l}^m C_{M_j}^R|\bar{\nu}_j| + C_R^R\bar{R} + C_G^R P_G^T) - k_T\bar{R} \leq 0$$

$$\frac{n_j^D}{\bar{k}_{E_l}}|\bar{\nu}_l| + \sum_{j=1,j\neq l}^m C_{M_j}^D|\bar{\nu}_j| + C_R^D\bar{R} + C_G^D P_G^T - \bar{D} < \dots$$

$$\frac{n_j^D}{k_{E_l}}|\bar{\nu}_l| + \sum_{j=1,j\neq l}^m C_{M_j}^D|\bar{\nu}_j| + C_R^D\bar{R} + C_G^D P_G^T - \bar{D} \leq 0$$

Comme l'hypothèse 4.7 est vraie, les inégalités (C_{1a}), (C_{1c}), (C_2) et (C_3) avec $\bar{k}_{E_l}$ sont strictes. Nous pouvons donc déduire grâce au lemme 4.2, qu'il existe $\delta\mu > 0$ tel que $\mathcal{C}_{\mu+\delta\mu,P_G} \neq \emptyset$, ce qui conclut la démonstration du point 2 de cette proposition.

Démonstration du point 3 : Soient $P_G \geq 0$, $\mu > 0$ et k_T fixés, $(\bar{R},\bar{\nu}) \in \mathcal{C}_{\mu,k_T}$ et $\delta k_T > 0$ tel que $\bar{k}_T \triangleq k_T + \delta k_T > 0$.
Remarquons tout d'abord que comme $(\bar{R},\bar{\nu}) \in \mathcal{C}_{\mu,P_G}$, (C_2) est satisfaite, donc $k_T - \mu C_R^R > 0$. Donc $\bar{k}_T - \mu C_R^R > 0$.
En soustrayant $\delta k_T\bar{R}$ de part et d'autre dans la contrainte (C_2), on obtient :

$$\mu(\sum_{j=1}^m C_{M_j}^R|\bar{\nu}_j| + C_R^R\bar{R} + C_G^R\bar{P}_G^T) - \bar{k}_T\bar{R} \leq -\delta k_T\bar{R} < 0.$$

En choisissant $\delta R > 0$ tel que $\delta R < \dfrac{\delta k_T\bar{R}}{\bar{k}_T - \mu C_R^R}$, l'inégalité précédente devient :

$$\mu(\sum_{j=1}^m C_{M_j}^R|\bar{\nu}_j| + C_R^R(\bar{R}-\delta R) + C_G^R P_G^T) - \bar{k}_T(\bar{R}-\delta R) \leq -\delta k_T\bar{R} + (\bar{k}_T - \mu C_R^R)\delta R < 0.$$

On en déduit que (C_2) est stricte. En soustrayant respectivement $\mu C_{R_i}^{M_p}\delta R$, $\mu C_{R_i}^{M_r}\delta R$, $C_R^D\delta R$ de part et d'autre des inégalités (C_{1a}), (C_{1c}) et (C_3), les inégalités (C_{1a}), (C_{1c}) et (C_3) sont strictes grâce à l'hypothèse 4.7. Nous pouvons ainsi déduire grâce au lemme 4.2, qu'il existe $\delta\mu > 0$ tel que $\mathcal{C}_{\mu+\delta\mu,P_G} \neq \emptyset$.
Par ailleurs, le point $(\bar{R}-\delta R,\bar{\nu})$ satisfait les contraintes (C_{1a}), (C_{1c}), (C_{1b}), (C_{1d}), (C_2), (C_3). Donc $(\bar{R}-\delta R,\bar{\nu}) \in \mathcal{C}_{\mu,P_G}$, ce qui conclut la démonstration de cette proposition.

■

Ainsi toute variation de paramètres permettant d'économiser des protéines contribue à l'augmentation du taux de croissance.

$\mathcal{P}_G$ reflète un compromis entre la croissance et la survie de la bactérie

Le premier point de la proposition 4.8 indique que le taux de croissance augmente lorsque l'ensemble des protéines $\mathcal{P}_G$ décroît. Concrètement, toute protéine non dédiée à la production de ribosomes et du réseau métabolique pénalise la bactérie en terme de

taux de croissance pur. En poursuivant ce raisonnement jusqu'au bout, une stratégie «optimale» maximisant la croissance consisterait donc à avoir un ensemble $\mathcal{P}_G$ vide.
Or, comme nous l'avons souligné dans la section 4.2, l'ensemble $\mathcal{P}_G$ est composé entre autres de protéines non dédiées au réseau métabolique mais essentielles en phase de croissance. Typiquement les protéines impliquées dans la réplication de l'ADN, dans l'assemblage de la paroi et de la membrane sont essentielles pour assurer la croissance et la division cellulaire. Par définition, l'ensemble $\mathcal{P}_G$ n'est pas vide.
En outre, il est bien évident que la bactérie est confrontée à un environnement complexe et potentiellement néfaste (stress, carence en nutriments, etc.). Certaines protéines de $\mathcal{P}_G$, produites en phase de croissance, sont présentes afin d'anticiper ces conditions néfastes ou tout simplement anticiper la phase stationnaire [59]. Les protéines dans $\mathcal{P}_G$, même si elles coûtent du taux de croissance, vont donner à la bactérie un avantage compétitif sur un plus long terme en augmentant son potentiel de survie. Plus précisément, les protéines $\mathcal{P}_G$ dépendent de l'état physiologique de la bactérie et leur fonction est directement liée à l'adaptation de la bactérie à sa niche écologique. De nombreuses expériences biologiques disponibles dans la littérature permettent de confirmer l'impact des protéines $\mathcal{P}_G$ sur le taux de croissance notamment en réponse à différents stress [237, 59, 219]. Par exemple pour *B. subtilis*, la synthèse de protéines impliquées dans la mobilité est active en phase exponentielle. Ce type de protéines typiquement appartient à notre ensemble $\mathcal{P}_G$. Si le principal inducteur de la mobilité, (le facteur sigma σ_D) est désactivé, la souche mutante croît plus vite que la souche sauvage [59].
Cependant, le poids de $\mathcal{P}_G$ par rapport aux autres protéines est difficile à évaluer pour des taux de croissance faibles comparés à des taux de croissance plus rapides. En effet, les stratégies développées par la bactérie peuvent être très complexes pour des taux de croissance faibles ou durant la transition entre la phase de croissance et la phase stationnaire [84, 168, 211, 196, 33]. Chez *B. subtilis*, ces stratégies incluent par exemple la synthèse de protéines de protection à différents stress [237, 219], d'antibiotiques [131] ou des stratégies haut niveau de décision entre différents modes de vie alternatifs (voir l'annexe A.1 page 197 et [196, 33]). D'ailleurs, les bactéries sont plus sensibles à un stress oxydatif en phase de croissance qu'en phase stationnaire. Ainsi, la composition de l'ensemble $\mathcal{P}_G$ peut varier énormément et surtout la proportion de $\mathcal{P}_G$ par rapport à l'ensemble des autres protéines peut devenir très importante, voire majoritaire. Une remarque importante concerne les protéines sécrétées dans le milieu extracellulaire, appartenant typiquement à $\mathcal{P}_G$ aussi. Elles ont théoriquement moins d'impact sur le taux de croissance que les autres protéines de $\mathcal{P}_G$ car elles ne contribuent pas à la limitation de la densité intracellulaire.
La composition de $\mathcal{P}_G$ résulte donc d'un compromis entre la croissance et la survie de la bactérie dans un contexte compétitif avec les autres bactéries. Par rapport aux régulations génétiques qui peuvent émerger à court terme dans les voies métaboliques, des stratégies plus complexes modifiant $\mathcal{P}_G$ apparaissent à moyen terme car elles sont le résultat de la compétition entre les bactéries, du type et de la fréquence de l'environnement rencontré.

Remarque 4.4 *Un argument souvent suggéré dans les articles de biologie est que l'éco-*

nomie de protéines permettrait d'augmenter le taux de croissance. Ici nous avons formellement démontré dans les propositions 4.10, 4.8 et le corollaire 4.6 qu'effectivement toute économie de protéines conduit à augmenter le taux de croissance. Les régulations génétiques sont un mécanisme parmi d'autres pour obtenir ce résultat.

Pour aller encore plus loin, on peut déduire de ces propositions la conséquence suivante sur l'optimalité des composants cellulaires.

Optimisation locale de chaque composant cellulaire

Les points 2 et 3 de la proposition 4.8 montrent que les efficacités des enzymes et des ribosomes ont aussi été optimisées pour réaliser leur tâche avec le moins de protéines possible. L'échelle de temps pour améliorer ces efficacités est plus longue. La nature des contraintes agissant en phase exponentielle sur la bactérie impose que structurellement, tout mécanisme permettant d'économiser des protéines conduit à augmenter le taux de croissance. L'optimisation locale de chaque élément contribue à l'optimum global associé à la maximisation du taux de croissance. Cette propriété est remarquable et nous pouvons imaginer que la pression de sélection a conduit à sélectionner l'efficacité des voies métaboliques [91, 245, 45] et des ribosomes [52] et à l'émergence des régulations génétiques pour éviter la production de protéines non essentielles dans des conditions environnementales données. A l'inverse de ce qui est généralement observé en ingénierie, l'optimisation du dimensionnement de la cellule dans sa globalité contribue à optimiser localement chaque élément du réseau métabolique, de l'appareil de traduction, et plus généralement de tous les processus cellulaires.

4.6 Comparaison de la méthode RBA avec la littérature

A ce jour, il n'existe pas de méthodes mathématiques permettant de prédire les concentrations des enzymes et des ribosomes. En revanche, de nombreuses méthodes de prédiction de flux métaboliques ont été développées autour d'une méthode commune appelée «Flux Balance Analysis». Notre méthode s'inscrit d'ailleurs dans ce contexte de croissance et de flux dits balancés (ou équilibrés).

La méthode «Flux Balance Analysis» (FBA)

La méthode FBA pose le problème de prédiction des flux métaboliques en régime établi comme un problème de programmation linéaire sous contraintes à travers la maximisation d'une fonction objectif [231]

$$
\begin{array}{rcll}
(P_{fba}) & = & \text{maximiser} & c^T\nu \\
 & & \text{soumis à} & \\
 & & & S\nu = 0 \\
 & & & \alpha_i \leq \nu_i \leq \beta_i
\end{array}
$$

où $\nu \triangleq (\nu_1, \ldots, \nu_m)$ désigne le vecteur des flux métaboliques en régime établi, S est la matrice de stœchiométrie du réseau métabolique, α_i et β_i définissent l'intervalle de valeurs admissible pour le i-ème flux, et c est la fonction objectif. La contrainte structurelle $S\nu = 0$ intégrée dans P_{fba} correspond également à la loi de conservation de la masse intégrée dans la méthode RBA. De même que la méthode RBA se résoud par un problème LP, la méthode FBA se formule également comme un problème d'optimisation convexe de type LP. A titre d'exemple, la résolution de P_{fba} ne prend que quelques secondes. Les deux problèmes d'optimisation P_{fba} et $P_f^{lp}(\mu)$ sont de complexité similaire.

Ce problème d'optimisation P_{fba} provient en fait de l'indétermination du problème suivant :

$$S.\nu = b$$

où b est un vecteur de flux mesurés expérimentalement et bien entendu le nombre de flux mesurés est inférieur au nombre total de flux métaboliques m. La matrice S n'étant pas de rang plein, l'ajout de contraintes inégalités du type $\alpha_i \leq \nu_i \leq \beta_i$ permet de réduire l'espace des solutions faisables. Les bornes α_i et β_i de P_{fba} sont ainsi déterminées soit expérimentalement, soit en imposant une contrainte de thermodynamique. Dans le cas d'enzymes irréversibles par exemple, le flux métabolique est strictement positif ($\alpha_i = 0$). Pour un flux mesuré, on impose alors $\nu_i = b_i$ où b_i désigne la mesure du flux ν_i. Toutefois, peu de flux métaboliques peuvent être contraints sur un intervalle de valeurs déterminé expérimentalement. Il en résulte que l'espace des solutions généré par l'ensemble de contraintes $S\nu = 0$ et $\alpha_i \leq \nu_i \leq \beta_i$ est trop grand pour espérer calculer la distribution des flux métaboliques à l'échelle de la cellule. Ainsi le choix de la fonction objectif c est primordial.

La maximisation du taux de croissance μ a souvent été choisie comme fonction objectif. En effet, la sélection naturelle a souvent encouragé les populations bactériennes ayant maximisé leur performance de croissance dans une niche écologique donnée [201]. Cette fonction objectif est traduite en une demande de flux pour des précurseurs métaboliques comme les acides aminés ou l'ATP, à travers la composition moyenne de la cellule à un taux de croissance donné. Le taux de croissance devient un flux μ tel que :

$$c_1 m_1 + \ldots + c_k m_k \xrightarrow{\triangleq \mu} \text{biomasse}$$

où $m_1, \ldots, m_k$ sont les précurseurs métaboliques et $c_1, \ldots, c_k$ désignent les quantités en millimoles par gramme de poids sec de précurseurs métaboliques nécessaires à la formation d'un gramme de biomasse sèche à un taux de croissance donné.
Maximiser la croissance s'est révélé être une hypothèse pertinente dans certaines applications FBA comme la prédiction de gènes létaux [51, 165] ou d'états métaboliques optimaux [50].

Les extensions des FBA

Cependant, certaines prédictions ne correspondent pas aux phénotypes observés expérimentalement, et ce pour différentes raisons. Tout d'abord, la composition moyenne

de la cellule varie avec le taux de croissance. Les prédictions sont sensiblement améliorées quand la composition de la cellule est adaptée en fonction du taux de croissance [170]. Ensuite, la solution ν de P_{fba} n'est pas nécessairement unique [123, 138]. De plus, la matrice de stœchiométrie S n'intègre pas traditionnellement les différents coûts en métabolites dûs à la synthèse *de novo* des composants cellulaires, comme par exemple l'hydrolyse de 2 GTP par acide aminé inséré durant l'élongation de la chaîne peptidique lors de la traduction des protéines. Cependant, des travaux théoriques et expérimentaux récents ont permis d'intégrer ces coûts pour *Escherichia coli* [2, 1, 222]. Enfin, les régulations génétiques ne sont pas prises en compte implicitement en tant que contraintes structurelles.
Ainsi, des modifications au problème initial P_{fba}, résumées dans [108], ont été apportées pour réduire l'espace des solutions faisables et améliorer potentiellement les prédictions :

- la méthode rFBA [42, 137, 41] qui intègre les régulations génétiques à travers les bornes α_i et β_i grâce à l'utilisation d'opérateurs logiques Booléens pour déterminer si un flux est «on» ($\alpha_i = -\infty$ et $\beta_i = +\infty$) ou «off» ($\alpha_i = 0$ et $\beta_i = 0$). Ici encore les régulations génétiques ne sont pas intégrées en tant que contrainte structurelle mais comme un ensemble de contraintes spécifiques sur certains flux métaboliques ;
- la méthode FBAwMC qui intègre des contraintes spatiales [18, 232]. Contrairement à la méthode rFBA, la méthode FBAwMC intègre une vraie contrainte structurelle complémentaire à savoir un volume intracellulaire limité. La concentration maximale des enzymes est donc intrinsèquement limitée. Cette contrainte est de même nature que la contrainte de densité de notre méthode RBA ;
- l'intégration de la thermodynamique via une adaptation des cœfficients α_i et β_i [15] ;
- l'exploration de fonctions objectif alternatives [115, 194] ou l'inférence de fonctions objectif à partir de données expérimentales sans connaissance *a priori* [201, 31, 73].

La méthode RBA étend la méthode FBA en proposant et intégrant des contraintes structurelles supplémentaires, en plus de la loi de conservation de la masse (contrainte C_{1d}). L'espace des solutions faisables dans la méthode RBA est donc intrinsèquement plus petit que celui des FBA. Nous avons conservé les avantages des FBA soulignés à savoir la représentation simple du réseau métabolique sous forme de matrice de stœchiométrie ainsi que la formulation en un problème de programmation linéaire. Par contre, nous nous sommes affranchis des principaux inconvénients de la méthode FBA :

- imposer une composition *a priori* de la biomasse n'est plus nécessaire. La composition de la biomasse s'adapte d'elle-même à travers la répartition des protéines entre ribosomes et voies métaboliques ;
- spécifier une fonction objectif n'est plus indispensable non plus pour obtenir une solution réaliste d'un point de vue biologique. Toutefois, en imposer une reste toujours possible. La méthode RBA permet donc de résoudre facilement un critère multi-objectif comme maximiser à la fois la biomasse et l'ATP produit par exemple ;
- imposer des bornes α_i, β_i sur les flux et en particulier sur les flux d'entrée n'est plus indispensable à l'obtention d'une solution. En effet, la contrainte de densité (C_3) limite intrinsèquement la concentration des protéines et limite ainsi les flux

métaboliques. Cependant, le formalisme choisi permet d'ajouter dans les RBA des contraintes sur les valeurs de flux si on le désire.

En outre, la méthode RBA fournit directement la concentration des protéines (ribosomes, enzymes, transporteurs, etc.) et non uniquement les flux métaboliques comme les FBA. Le cœfficient de conversion k_{E_i} entre flux et concentration enzymatique a été principalement introduit pour pouvoir manipuler la matrice de stœchiométrie S. La valeur k_{E_i} pour chaque enzyme doit néanmoins être identifiée à partir des données expérimentales. Enfin, nous avons montré que l'intégration des nouvelles contraintes structurelles permet de prédire et de recouvrir des modules fonctionnels connus dans les voies métaboliques. La méthode RBA intègre en effet des caractéristiques des voies métaboliques comme la composition en acides aminés de chaque enzyme contrairement à la méthode FBA. Si nous reprenons notre exemple de deux voies de synthèse différentes mp_1 et mp_2, dans les conditions considérées, la méthode FBA nous donnerait l'ensemble de relations algébriques suivantes

$$\begin{aligned}
&\nu_1 = \nu_2 \\
&\quad\vdots \\
&\nu_{N_1-1} = \nu_{N_1} \\
&\nu_1^* = \nu_2^* \\
&\quad\vdots \\
&\nu_{N_2-1}^* = \nu_{N_2}^* \\
&\nu_{N_1} + \nu_{N_2}^* = c_k\mu
\end{aligned}$$

où c_k désigne la quantité en mmol/gdwc de précurseur X_{p_k} nécessaire pour faire 1 gramme de biomasse. A μ fixé, l'ensemble des solutions se situe dans un hyperplan défini par $(\nu_1, \nu_1^*) \in \mathcal{R}^m \times \mathcal{R}^m$ tel que $\nu_1 + \nu_1^* = c_k\mu$. On obtient une infinité de solution alors que notre méthode RBA va privilégier la voie courte de synthèse. Nous avons démontré explicitement à travers le corollaire 4.6 d'une part que l'espace des solutions des RBA pour l'exemple considéré était restreint par rapport à celui des FBA, et d'autre part que les voies métaboliques courtes et peu chères en protéines sont privilégiées.

4.7 Vers une méthode plus prédictive

Afin d'identifier les contraintes structurelles agissant sur la bactérie, nous avons simplifié le comportement de la cellule à travers une vision d'analyse des systèmes propre à l'Automatique. En particulier, nous avons négligé, tout comme la méthode FBA, la dépendance de certains paramètres avec le taux de croissance, l'usage de différents codons pour un même acide aminé, la dégradation des protéines, les ribosomes inactifs ou en formation, le sur-dimensionnement de la glycolyse, l'affinité des transporteurs pour leur substrat ou encore l'impact du stress oxydatif sur le réseau métabolique. Nous allons montrer dans cette section qu'en restant dans le même cadre mathématique, il est possible de modifier le problème d'optimisation afin de prendre en compte ces différents phénomènes pour améliorer le pouvoir de prédiction de notre méthode.

a. Dépendance des paramètres k_T et k_{E_i} avec μ

Certaines expériences biologiques montrent que les paramètres k_T et k_{E_i} pourraient augmenter avec le taux de croissance, grâce à l'augmentation des concentrations des ARNt chargés pour k_T, ou l'augmentation de certaines concentrations métaboliques pour k_{E_i}. Les expériences biologiques présentées dans la section 3.3 du chapitre précédent ont typiquement montré qu'à faible taux de croissance, l'adaptation des flux métaboliques est effectuée en grande partie par les variations de certains métabolites. Dans ces conditions, le problème d'optimisation initial est modifié car des contraintes additionnelles sont nécessaires pour garantir la monotonicité de la solution (point 1 de la proposition 4.4).

Efficacité des ribosomes k_T : Supposons que k_T est une fonction continue, positive croissante et bornée du taux de croissance μ [28]. telle que pour tout $\delta\mu$ suffisamment petit, $k_T(\mu+\delta\mu) = k_T(\mu) + \dfrac{\partial k_T(\mu)}{\partial \mu}\delta\mu$. A μ fixé, le problème d'optimisation $P_f(\mu)$ reste convexe. En revanche, l'extension de la proposition 4.4 nécessite l'ajout de la contrainte suivante

$$\sum_{j=1}^{m} C_{M_j}^{R}|\nu_j| + \left(C_R^R - \frac{\partial k_T(\mu)}{\partial \mu}\right) R + C_G^R P_G^T \geq 0$$

pour démontrer le point 1 de cette proposition (voir l'annexe 4.C de ce chapitre page 159). Cette inégalité signifie que le gain en protéines associé à la diminution de μ est plus grand que la perte d'efficacité des ribosomes.

Efficacité des enzymes k_E : Supposons que pour $l \in I_m$, l'efficacité k_{E_l} est une fonction continue, positive croissante et bornée du taux de croissance μ telle que pour tout $\delta\mu$ suffisamment petit, $k_{E_l}(\mu+\delta\mu) = k_{E_l}(\mu) + \dfrac{\partial k_{E_l}(\mu)}{\partial \mu}\delta\mu$.
A μ fixé, le problème d'optimisation $P_f(\mu)$ reste convexe. En revanche, le point 1 de la proposition 4.4 nécessite l'ajout d'un certain nombre de contraintes énoncées ci-dessous (voir l'annexe 4.C de ce chapitre page 159). Elles traduisent le fait que pour une variation $\delta\mu < 0$, le gain en protéines et en précurseurs métaboliques associé à la diminution de μ compense la perte d'efficacité de l'enzyme E_l.

– (C_{1a}^*). Pour tout $i \in I_p$,

$$-k_{E_l}(\mu+\delta\mu)\left[\alpha_{il}^p \frac{|\bar{\nu}_l|}{k_{E_l}(\mu+\delta\mu)} + \textstyle\sum_{j=1,j\neq l}^{m} C_{M_{ij}}^{M_p}|\bar{\nu}_j| + C_{R_i}^{M_p}\bar{R} + C_{G_i}^{M_p}P_G^T\right]\ldots$$
$$+\frac{\partial k_{E_l}(\mu)}{\partial \mu}\left[\textstyle\sum_{j=1}^{m} S_{p_{ij}}\bar{\nu}_j - \mu\left(\sum_{j=1,j\neq l}^{m} C_{M_{ij}}^{M_p}|\bar{\nu}_j| + C_{R_i}^{M_p}\bar{R} + C_{G_i}^{M_p}P_G^T\right) - \nu_Y\right] \leq 0$$

(C_{1a}^*) signifie que le flux de précurseurs économisé grâce à la diminution de μ est plus grand que celui consommé pour compenser la baisse d'efficacité de l'enzyme E_l.

– Pour (C^*_{1c}). Pour tout $i \in I_r$,

$$k_{E_l}(\mu+\delta\mu)\left[\alpha^r_{il}\frac{|\bar{\nu}_l|}{k_{E_l}(\mu+\delta\mu)} + \sum_{j=1,j\neq l}^m C^{M_r}_{M_{ij}}|\bar{\nu}_j| + C^{M_r}_{R_i}\bar{R} + C^{M_r}_{G_i}P^T_G\right]\dots$$
$$+\frac{\partial k_{E_l}(\mu)}{\partial\mu}\left[-\sum_{j=1}^m S_{r_{ij}}\bar{\nu}_j - \mu\left(\sum_{j=1,j\neq l}^m C^{M_r}_{M_{ij}}|\bar{\nu}_j| + C^{M_r}_{R_i}\bar{R} + C^{M_r}_{G_i}P^T_G\right)\right] \leq 0$$

(C^*_{1c}) signifie que le flux de métabolites recyclés économisé grâce à la diminution de μ est plus petit que celui associé à la baisse d'efficacité de l'enzyme E_l.

– Pour (C^*_2).

$$-k_{E_l}(\mu+\delta\mu)\left[n^R_l\frac{|\bar{\nu}_l|}{k_{E_l}(\mu+\delta\mu)} + \sum_{j=1,j\neq l}^m C^R_{M_j}|\bar{\nu}_j| + C^R_R\bar{R} + C^R_G P^T_G\right]\dots$$
$$-\frac{\partial k_{E_l}(\mu)}{\partial\mu}\left[\mu\left(\sum_{j=1,j\neq l}^m C^R_{M_j}|\bar{\nu}_j| + C^R_R\bar{R} + C^R_G P^T_G\right) - k_T\bar{R}\right] \leq 0$$

(C^*_2) signifie que le flux d'acides aminés économisé grâce à la diminution de μ est plus grand que celui nécessaire pour compenser la baisse d'efficacité de l'enzyme E_l.

– Pour (C^*_3).

$$-\frac{\partial k_{E_l}(\mu)}{\partial\mu}\left(\sum_{j=1,j\neq l}^m C^D_{M_j}|\bar{\nu}_j| + C^D_R\bar{R} + C^D_G P^T_G - \bar{D}\right) < 0$$

(C^*_3) signifie que les acides aminés nécessaires à la compensation de la baisse d'efficacité de l'enzyme E_l satisfont la contrainte de densité.

b. L'usage de codons

La composition en acides aminés des protéines est intrinsèquement prise en compte à travers les cœfficients $C^{M_p}_{M_{ij}}$, $C^{M_p}_{R_i}$ et $C^{M_p}_{G_i}$ associés aux acides aminés. L'usage de codons peut être pris en compte de la même façon. Au lieu de considérer seulement 20 acides aminés, tous les différents codons des acides aminés sont inclus dans le vecteur X_p des précurseurs métaboliques, soit au maximum 61 codons différents. On obtient alors pour une même protéine 61 cœfficients $C^{M_p}_{M_{ij}}$ différents, correspondant cette fois au nombre de codons spécifique codant pour un certain acide aminé divisé par k_E. Il est aussi possible d'introduire des efficacités de traduction différentes pour les ribosomes, traduisant la plus faible affinité des ARNt chargés associés aux codons rares [54].

c. La dégradation des protéines

La dégradation naturelle de chaque protéine peut être incluse dans $P_f(\mu)$ en ajoutant aux contraintes (C_{1a}), (C_{1c}), et (C_2) les termes suivants :

$$\begin{array}{ll} (C_{1a}) & \text{pour tout } i \in \{1, \ldots, N_p\} \\ & \gamma_j C_{M_{ij}}^{M_p} |\nu_j| \\ (C_{1c}) & \text{pour tout } i \in \{1, \ldots, N_r\} \\ & \gamma_j C_{M_{ij}}^{M_r} |\nu_j| \\ (C_2) & \gamma_j C_{M_j}^{R} |\nu_j| \end{array}$$

où γ_j in h^{-1} désigne la vitesse de dégradation de la j-ème protéine.

d. Les ribosomes inactifs

Environ 20% des ribosomes sont inactifs dans la cellule [28]. Dans la contrainte (C_2), seulement 80% des ribosomes produisent réellement des protéines, soit :

$$\mu \left(\sum_{j=1}^{m} C_{M_j}^{R} |\nu_j| + C_R^R R + C_G^R P_G^T \right) - k_T R \leq 0$$

En substituant R par $0.8R$, on intègre ainsi la fraction inactive des ribosomes dans la cellule.

e. Des réserves en co-facteurs

Il est généralement admis que la glycolyse ne fonctionne jamais au maximum de sa capacité. Ainsi, elle peut répondre à une demande instantanée d'ATP en augmentant légèrement le flux dans cette voie. Pour intégrer ce phénomène, il suffit de demander que le flux de synthèse d'ATP reste supérieur à un certain seuil :

$$-\sum_{j=1}^{m} S_{atp,j} \nu_j + b_{atp} \leq 0$$

où $-\sum_{j=1}^{m} S_{atp,j} \nu_j$ désigne la somme des flux de production et de consommation d'ATP, et b_{atp} est le flux minimal d'ATP à fournir. On pourrait imposer ce type de contraintes supplémentaires pour tous les co-facteurs (NADH, NADPH, etc.).

f. L'affinité différente des transporteurs

Plutôt que de manipuler des flux d'entrée de nutriments, le formalisme choisi permet de manipuler directement leur concentration extracellulaire dans le milieu. Le cœfficient k_E liant le flux à la concentration du transporteur est remplacé par une fonction non linéaire de la concentration extracellulaire du nutriment S_{ext} :

$$k_{E_i} = k_{max} \frac{S_{ext}}{K_{m_i} + S_{ext}}$$

où k_{max} correspond à la vitesse maximale du transporteur, et K_{m_i} la constante de Michaëlis-Menten. Ainsi l'impact de faibles concentrations en nutriments est intégré dans le problème d'optimisation. Étudier la substitution de transporteurs entre un basse affinité (une perméase généralement peu coûteuse) et un haute affinité (généralement un transporteur de type ABC coûteux) en fonction de la concentration extérieure du substrat devient possible.

g. L'impact du stress oxydatif sur le réseau métabolique

En stress oxydatif, l'activité de certaines enzymes, notamment celles possédant un centre fer-soufre, diminue ou est complètement stoppée [101]. Comme précédemment, l'efficacité k_{E_i} de certaines enzymes connues pour être abîmées en stress oxydatif peut être ajustée :

$$k_{E_i} = k_{max} \frac{K_i}{K_i + H_2O_2}$$

Dans ces conditions, l'impact d'un stress oxydatif sur la distribution des flux métaboliques ou sur la substitution d'enzymes avec/sans centre fer-soufre peut également être appréhendé.

4.8 Conclusion

Le cadre formel présenté dans ce chapitre, appelé méthode Resource Balance Analysis (RBA), permet donc d'expliquer entre autres l'émergence des modules dans les voies métabolique, à travers la formalisation de la gestion des ressources de la cellule en régime établi en un problème d'optimisation convexe $P_f(\mu)$. Outre ses propriétés remarquables comme la convexité, l'étude de $P_f(\mu)$ montre d'une part que le taux de croissance est structurellement limité et d'autre part que tout mécanisme permettant d'économiser des protéines conduit à augmenter le taux de croissance. La limitation intrinsèque de la ressource (contrainte C_3) est ainsi critique. Pour augmenter le taux de croissance, la cellule doit économiser des protéines à un endroit, comme par exemple éteindre certaines voies métaboliques, et allouer ces protéines ainsi économisées aux ribosomes et aux voies métaboliques encore actives. L'économie en protéines est réalisée non seulement au niveau des précurseurs métaboliques et des ribosomes libres, mais aussi au niveau de la densité intracellulaire. L'émergence des modules dans les voies métaboliques est par conséquent une conséquence directe de la limitation de la ressource.

Au delà d'un cadre formel, la méthode RBA est équivalente à un problème de programmation linéaire, et peut donc se résoudre efficacement pour toute bactérie. Au cours du chapitre suivant, nous allons appliquer cette méthode à *Bacillus subtilis* afin d'illustrer la capacité de prédiction de notre méthode : évolution des concentrations des ribosomes et des enzymes pour différents milieux, prédiction des modules dans les voies métaboliques et comparaison avec les modules fonctionnels identifiés au chapitre précédent...

Annexes

4.A Démonstration du lemme 4.2

Soit $(\bar{R}, \bar{\nu}) \in \mathcal{C}_{\mu,P_G}$ tel qu'il existe $\epsilon > 0$ tel que :

(C_{1a}^{ϵ}) pour tout $i \in I_p$
$-\sum_{j=1}^{m} S_{p_{ij}} \bar{\nu}_j + \mu \left(\sum_{j=1}^{m} C_{M_{ij}}^{M_p} |\bar{\nu}_j| + C_{R_i}^{M_p} \bar{R} + C_{G_i}^{M_p} P_G^T \right) - \nu_Y \leq -\epsilon$

(C_{1b}^{ϵ}) and for all $i \in I_c$,
$-\sum_{j=1}^{m} S_{c_{ij}} \bar{\nu}_j + \mu \bar{X}_{c_i} \leq 0$

(C_{1c}^{ϵ}) pour tout $i \in I_r$
$\sum_{j=1}^{m} S_{r_{ij}} \bar{\nu}_j + \mu \left(\sum_{j=1}^{m} C_{M_{ij}}^{M_r} |\bar{\nu}_j| + C_{R_i}^{M_r} \bar{R} + C_{G_i}^{M_r} P_G^T \right) \leq -\epsilon$

(C_{1d}^{ϵ}) pour tout $i \in I_i$
$\sum_{j=1}^{m} S_{I_{ij}} \bar{\nu}_j = 0$

(C_2^{ϵ}) $\mu(\sum_{j=1}^{m} C_{M_j}^{R} |\bar{\nu}_j| + C_R^R \bar{R} + C_G^R P_G^T) - k_T \bar{R} \leq -\epsilon$

(C_3^{ϵ}) $\sum_{j=1}^{m} C_{M_j}^{D} |\bar{\nu}_j| + C_R^D \bar{R} + C_G^D P_G^T - \bar{D} \leq -\epsilon$

On a pour tout $\frac{\delta\mu}{\mu} \geq 0$, et pour tout $i \in I_c$,

$$-\sum_{j=1}^{m} S_{c_{ij}} \bar{\nu}_j \frac{\delta\mu}{\mu} + \delta\mu \bar{X}_{c_i} \leq 0.$$

Introduisons maintenant les grandeurs suivantes : $\delta\nu \triangleq (\delta\nu_1, \ldots, \delta\nu_m)^T$ avec $\delta\nu_j \triangleq \bar{\nu}_j \frac{\delta\mu}{\mu}$ et $\delta R \triangleq \bar{R} \frac{\delta\mu}{\mu}$. En multipliant (C_{1a}^{ϵ}), (C_{1b}^{ϵ}), (C_{1c}^{ϵ}), (C_{1d}^{ϵ}) et (C_2^{ϵ}) par $\frac{\delta\mu}{\mu}$ et en additionnant

les inégalités obtenues à (C^{ϵ}_{1a}), (C^{ϵ}_{1b}), (C^{ϵ}_{1c}), (C^{ϵ}_{1d}) et (C^{ϵ}_{2}), on obtient :

(C^{ϵ}_{1a}) pour tout $i \in I_p$
$$-\sum_{j=1}^{m} S_{p_{ij}}(\bar{\nu}_j + \delta\nu_j) + (\mu + \delta\mu)\left(\sum_{j=1}^{m} C^{M_p}_{M_{ij}}|\bar{\nu}_j| + C^{M_p}_{R_i}\bar{R} + C^{M_p}_{G_i}P^T_G\right) - \nu_Y(1 + \tfrac{\delta\mu}{\mu}) \leq \ldots$$
$$-\epsilon(1 + \tfrac{\delta\mu}{\mu})$$

(C^{ϵ}_{1b}) et pour tout $i \in I_c$,
$$-\sum_{j=1}^{m} S_{c_{ij}}(\bar{\nu}_j + \delta\nu_j) + (\mu + \delta\mu)\bar{X}_{c_i} \leq 0$$

(C^{ϵ}_{1c}) pour tout $i \in I_r$
$$\sum_{j=1}^{m} S_{r_{ij}}(\bar{\nu}_j + \delta\nu_j) + (\mu + \delta\mu)\left(\sum_{j=1}^{m} C^{M_r}_{M_{ij}}|\bar{\nu}_j| + C^{M_r}_{R_i}\bar{R} + C^{M_r}_{G_i}P^T_G\right) \leq -\epsilon(1 + \tfrac{\delta\mu}{\mu})$$

(C^{ϵ}_{1d}) pour tout $i \in I_i$
$$\sum_{j=1}^{m} S_{I_{ij}}(\bar{\nu}_j + \delta\nu_j) = 0$$

(C^{ϵ}_{2}) $(\mu + \delta\mu)(\sum_{j=1}^{m} C^{R}_{M_j}|\bar{\nu}_j| + C^{R}_{R}\bar{R} + C^{R}_{G}P^T_G) - k_T(\bar{R} + \delta R) \leq -\epsilon(1 + \frac{\delta\mu}{\mu})$

Finalement, en ajoutant les contributions de $|\delta\nu_j|$ et δR à (C^{ϵ}_{1a}), (C^{ϵ}_{1c}), (C^{ϵ}_{2}) et (C^{ϵ}_{3}), on obtient :

pour tout $i \in I_p$
$$-\sum_{j=1}^{m} S_{p_{ij}}(\bar{\nu}_j + \delta\nu_j) + (\mu + \delta\mu)\left(\sum_{j=1}^{m} C^{M_p}_{M_{ij}}(|\bar{\nu}_j| + |\delta\nu_j|) + C^{M_p}_{R_i}(\bar{R} + \delta R) + C^{M_p}_{G_i}P^T_G\right) - \ldots$$
$$\nu_Y(1 + \tfrac{\delta\mu}{\mu}) \leq -\epsilon(1 + \tfrac{\delta\mu}{\mu}) + (\mu + \delta\mu)(\sum_{j=1}^{m} C^{M_p}_{M_{ij}}|\delta\nu_j| + C^{M_p}_{R_i}\delta R)$$

et pour tout $i \in I_c$,
$$-\sum_{j=1}^{m} S_{c_{ij}}(\bar{\nu}_j + \delta\nu_j) + (\mu + \delta\mu)\bar{X}_{c_i} \leq 0$$

pour tout $i \in I_r$
$$\sum_{j=1}^{m} S_{r_{ij}}(\bar{\nu}_j + \delta\nu_j) + (\mu + \delta\mu)\left(\sum_{j=1}^{m} C^{M_r}_{M_{ij}}(|\bar{\nu}_j| + |\delta\nu_j|) + C^{M_r}_{R_i}(\bar{R} + \delta R) + C^{M_r}_{G_i}P^T_G\right) \leq \ldots$$
$$-\epsilon(1 + \tfrac{\delta\mu}{\mu}) + (\mu + \delta\mu)(\sum_{j=1}^{m} C^{M_r}_{M_{ij}}|\delta\nu_j| + C^{M_r}_{R_i}\delta R)$$

pour tout $i \in I_i$
$$\sum_{j=1}^{m} S_{I_{ij}}(\bar{\nu}_j + \delta\nu_j) = 0$$

$$(\mu + \delta\mu)(\sum_{j=1}^{m} C^{R}_{M_j}(|\bar{\nu}_j| + |\delta\nu_j|) + C^{R}_{R}(\bar{R} + \delta R) + C^{R}_{G}P^T_G) - k_T(\bar{R} + \delta R) \leq \ldots$$
$$-\epsilon(1 + \tfrac{\delta\mu}{\mu}) + (\mu + \delta\mu)(\sum_{j=1}^{m} C^{R}_{M_j}|\delta\nu_j| + C^{R}_{R}\delta R)$$

$$\sum_{j=1}^{m} C^{D}_{M_j}(|\bar{\nu}_j| + |\delta\nu_j|) + C^{D}_{R}(\bar{R} + \delta R) + C^{D}_{G}P^T_G - \bar{D} \leq \ldots$$
$$-\epsilon + \sum_{j=1}^{m} C^{D}_{M_j}|\delta\nu_j| + C^{D}_{R}\delta R$$

Puisque $\epsilon > 0$ est fixé et que $\delta\mu$ (et donc $\delta\nu_j$, δR) peut être choisi aussi petit que l'on veut, nous en déduisons qu'il existe $\delta\mu > 0$ tel que $(\bar{R} + \delta R, \bar{\nu} + \delta\nu) \in \mathcal{C}_{\mu+\delta\mu,P_G}$. Donc $\mathcal{C}_{\mu+\delta\mu,P_G} \neq \emptyset$ ce qui conclut cette démonstration.

4.B La proposition 4.10 et sa démonstration

On suppose que les voies métaboliques mp_1 et mp_2 sont composées respectivement de N_1 et N_2 enzymes, telles qu'elles ne produisent ni ne consomment aucun co-métabolite.

Concrètement, cela signifie que le produit de la i-ème enzyme correspond au substrat de la $(i+1)$-ème enzyme (comme indiqué sur la figure 4.4). Soient $I_1 \subseteq I_m$ et $I_2 \subseteq I_m$ l'ensemble des indices des flux métaboliques associés aux enzymes appartenant à ces voies métaboliques. Introduisons maintenant l'hypothèse suivante qui caractérise les conséquences sur les différents cœfficients d'un chemin «court» par rapport à un chemin «long».

Hypothèse 4.9 *Pour les deux voies métaboliques mp_1 et mp_2 introduites précédemment :*

(i) Le nombre de chaque précurseur métabolique nécessaires pour la synthèse des enzymes de mp_2 est inférieur à celui associé à mp_1 : pour tout $i \in I_p$,

$$\sum_{j \in I_2} C_{M_{ij}}^{M_p} < \sum_{j \in I_1} C_{M_{ij}}^{M_p},$$

(ii) Le nombre de chaque métabolite recyclé produit durant la synthèse des enzymes de mp_2 est inférieur à celui associé à mp_1 : pour tout $i \in I_r$,

$$\sum_{j \in I_2} C_{M_{ij}}^{M_r} < \sum_{j \in I_1} C_{M_{ij}}^{M_r},$$

(iii) La longueur totale en acides aminés de mp_2 obtenue en sommant les longueurs individuelles en acides aminés de chaque enzyme constituant mp_2 est inférieure à celle de mp_1 :

$$\sum_{j \in I_2} C_{M_j}^{R} < \sum_{j \in I_1} C_{M_j}^{R},$$

(iv) L'espace intracellulaire occupé par les enzymes de mp_2 est inférieur à celui occupé par les enzymes de mp_1 :

$$\sum_{j \in I_2} C_{M_j}^{D} < \sum_{j \in I_1} C_{M_j}^{D}$$

.

Cette hypothèse indique simplement que le nombres de précurseurs métaboliques, le nombre de métabolites recyclés et la quantité totale d'acides aminés de la voie métabolique mp_2 sont inférieurs à ceux de mp_1, ce qui est généralement le cas lorsqu'on compare deux voies métaboliques [75].
Sous cette hypothèse, on en déduit la proposition suivante.

Proposition 4.10 *Soit l'hypothèse 4.9 vraie. Pour tout $P_G \geq 0$ et tout $\mu > 0$, si $P_f(\mu)$ est faisable et qu'une solution faisable $(\bar{R}, \bar{\nu})$ est telle que $\bar{\nu}_j \neq 0$ for $j \in I_1 \cup I_2$, alors il existe $\delta\mu > 0$ tel que $P_f(\mu + \delta\mu)$ est également faisable.*

Démonstration
Supposons que pour $P_G \geq 0$ et $\mu > 0$ fixés, $\mathcal{C}_{\mu,P_G} \neq \emptyset$ et qu'il existe une solution faisable $(\bar{R}, \bar{\nu}) \in \mathcal{C}_{\mu,P_G}$ telle que $\bar{\nu}_j = \bar{\nu} > 0$ pour $j \in I_1$ et $\bar{\nu}_j = 0$ pour $j \in I_2$. On a $|\bar{\nu}| = \bar{\nu}$.

L'ensemble des contraintes de $P_f(\mu)$ s'écrivent alors de la façon suivante

$$\begin{array}{l}
\text{pour tout } i \in I_p, i = k \\
-\sum_{j\in I_1} S_{p_{ij}}\bar{\nu} + \mu\left(\sum_{j\in I_1} C^{M_p}_{M_{ij}}\bar{\nu} + \sum_{j\notin I_1} C^{M_p}_{M_{ij}}|\bar{\nu}_j| + C^{M_p}_{R_i}\bar{R} + C^{M_p}_{G_i}P^T_G\right) - \nu_Y \leq 0 \\
\text{pour tout } i \in I_p, i \neq k \\
-\sum_{j\notin I_1} S_{p_{ij}}\bar{\nu}_j + \mu\left(\sum_{j\in I_1} C^{M_p}_{M_{ij}}\bar{\nu} + \sum_{j\notin I_1} C^{M_p}_{M_{ij}}|\bar{\nu}_j| + C^{M_p}_{R_i}\bar{R} + C^{M_p}_{G_i}P^T_G\right) - \nu_Y \leq 0
\end{array}$$

$$\begin{array}{l}
\text{pout tout } i \in I_c, \\
-\sum_{j\notin I_1} S_{c_{ij}}\bar{\nu}_j + \mu\bar{X}_{c_i} \leq 0 \\
\text{pour tout } i \in I_r \\
\sum_{j\notin I_1} S_{r_{ij}}\bar{\nu}_j + \mu\left(\sum_{j\in I_1} C^{M_r}_{M_{ij}}\bar{\nu} + \sum_{j\notin I_1} C^{M_r}_{M_{ij}}|\bar{\nu}_j| + C^{M_r}_{R_i}\bar{R} + C^{M_r}_{G_i}P^T_G\right) \leq 0 \\
\text{pour tout } i \in I_i \\
\sum_{j\in I_1} S_{I_{ij}}\bar{\nu} + \sum_{j\notin I_1} S_{I_{ij}}\bar{\nu}_j = 0
\end{array}$$

$$\begin{array}{l}
\mu\left(\sum_{j\in I_1} C^R_{M_j}\bar{\nu} + \sum_{j\notin I_1} C^R_{M_j}|\bar{\nu}_j| + C^R_R\bar{R} + C^R_G P^T_G\right) - k_T\bar{R} \leq 0 \\
\sum_{j\in I_1} C^D_{M_j}\bar{\nu} + \sum_{j\notin I_1} C^D_{M_j}|\bar{\nu}_j| + C^D_R\bar{R} + C^D_G P^T_G - \bar{D} \leq 0
\end{array}$$

Substituons maintenant la synthèse de X_{p_k} obtenue par mp_1 dans le système d'inégalités et d'égalités précédent, par mp_2 en additionnant et en soustrayant le même flux $\bar{\nu}$ à travers mp_2 dans toutes les contraintes où $\bar{\nu}$ apparaît. On obtient :

$$\begin{array}{l}
\text{pour tout } i \in I_p, i = k \\
-\sum_{j\in I_2} S_{p_{ij}}\bar{\nu} + \mu\left(\sum_{j\in I_2} C^{M_p}_{M_{ij}}\bar{\nu} + \sum_{j\notin I_1\cup I_2} C^{M_p}_{M_{ij}}|\bar{\nu}_j| + C^{M_p}_{R_i}\bar{R} + C^{M_p}_{G_i}P^T_G\right) - \nu_Y \leq \ldots \\
\qquad \bar{\nu}\mu\left(\sum_{j\in I_2} C^{M_p}_{M_{ij}} - \sum_{j\in I_1} C^{M_p}_{M_{ij}}\right) \\
\text{pour tout } i \in I_p, i \neq k \\
-\sum_{j\in I_1\cup I_2} S_{p_{ij}}\bar{\nu}_j + \mu\left(\sum_{j\in I_2} C^{M_p}_{M_{ij}}\bar{\nu} + \sum_{j\notin I_1\cup I_2} C^{M_p}_{M_{ij}}|\bar{\nu}_j| + C^{M_p}_{R_i}\bar{R} + C^{M_p}_{G_i}P^T_G\right) - \nu_Y \leq \ldots \\
\qquad \bar{\nu}\mu\left(\sum_{j\in I_2} C^{M_p}_{M_{ij}} - \sum_{j\in I_1} C^{M_p}_{M_{ij}}\right) < 0
\end{array}$$

pour tout $i \in I_c$,

$$-\sum_{j \notin I_1} S_{c_{ij}} \bar{\nu}_j + \mu \bar{X}_{c_i} \leq 0$$

pour tout $i \in I_r$

$$\sum_{j \notin I_1 \cup I_2} S_{r_{ij}} \bar{\nu}_j + \mu \left(\sum_{j \in I_2} C_{M_{ij}}^{M_r} \bar{\nu} + \sum_{j \notin I_1 \cup I_2} C_{M_{ij}}^{M_r} |\bar{\nu}_j| + C_{R_i}^{M_r} \bar{R} + C_{G_i}^{M_r} P_G^T \right) \leq \dots$$

$$\bar{\nu}\mu \left(\sum_{j \in I_2} C_{M_{ij}}^{M_r} - \sum_{j \in I_1} C_{M_{ij}}^{M_r} \right) < 0$$

pour tout $i \in I_i$

$$\sum_{j \in I_1} S_{I_{ij}} \bar{\nu} + \left(\sum_{j \in I_2} S_{I_{ij}} \bar{\nu} - \sum_{j \in I_2} S_{I_{ij}} \bar{\nu} \right) + \sum_{j \notin I_1 \cup I_2} S_{I_{ij}} \bar{\nu}_j = 0$$

$$\mu \left(\sum_{j \in I_2} C_{M_j}^R \bar{\nu} + \sum_{j \notin I_1 \cup I_2} C_{M_j}^R |\bar{\nu}_j| + C_R^R \bar{R} + C_G^R P_G^T \right) - k_T \bar{R} \leq \dots$$

$$\bar{\nu}\mu \left(\sum_{j \in I_2} C_{M_j}^R - \sum_{j \in I_1} C_{M_j}^R \right) < 0$$

$$\sum_{j \in I_2} C_{M_j}^D \bar{\nu} + \sum_{j \notin I_1 \cup I_2} C_{M_j}^D |\bar{\nu}_j| + C_R^D \bar{R} + C_G^D P_G^T - \bar{D} \leq \bar{\nu} \left(\sum_{j \in I_2} C_{M_j}^D - \sum_{j \in I_1} C_{M_j}^D \right) < 0$$

Grâce à l'hypothèse 4.9, on en déduit que les inégalités (C_{1a}), (C_{1c}), (C_2) et (C_3) sont strictes. D'après le lemme 4.2, on peut conclure qu'il existe un $\delta\mu > 0$ tel que $\mathcal{C}_{\mu+\delta\mu,P_G} \neq \emptyset$.

■

4.C Dépendance des paramètres k_T et k_{E_i} avec μ

4.C.1 Extension des propositions 4.1, 4.3 et 4.4 pour k_T dépendant de μ

Supposons que k_T est une fonction continue, positive croissante et bornée du taux de croissance μ [28]. telle que pour tout $\delta\mu$ suffisamment petit, $k_T(\mu + \delta\mu) = k_T(\mu) + \dfrac{\partial k_T(\mu)}{\partial \mu}\delta\mu$.

La propositions 4.1 et 4.3 peuvent être étendues directement dans ce cas en reprenant les mêmes démonstrations car μ est fixé. La proposition 4.4 peut aussi être étendue. La démonstration du point 1 est obtenue en suivant le même raisonnement que la démonstration initiale. Supposons que pour $\mu^+ > 0$, le problème $P_f(\mu^+)$ étendu pour k_T dépendent de μ est faisable, alors en substituant μ^+ par $\mu + \delta\mu$ avec $\delta\mu > 0$ on obtient pour la contrainte (C_2) :

$$\mu \left(\sum_{j=1}^m C_{M_j}^R |\bar{\nu}_j| + C_R^R \bar{R} + C_G^R P_G^T \right) - k_T \bar{R} \leq \dots$$
$$-\delta\mu \left(\sum_{j=1}^m C_{M_j}^R |\bar{\nu}_j| + C_R^R \bar{R} + C_G^R P_G^T \right) + \frac{\partial k_T(\mu)}{\partial \mu} \delta\mu \bar{R}$$

Une condition suffisante pour que la partie gauche de l'inégalité soit négative est que la solution $(\bar{R}, \bar{\nu})$ satisfasse l'inégalité suivante :

$$-\delta\mu \left(\sum_{j=1}^{m} C_{M_j}^{R} |\bar{\nu}_j| + (C_R^R - \frac{\partial k_T(\mu)}{\partial \mu})\bar{R} + C_G^R P_G^T \right) \leq 0$$

Le point 2 ne pose aucun problème sous l'hypothèse que le point 1 est démontré (c'est-à-dire que l'inégalité précédente est satisfaite) : la démonstration est identique puisque pour tout $\mu \geq 0$, la fonction $k_T(\mu)$ est bornée.

4.C.2 Extension des propositions 4.1, 4.3 et 4.4 pour k_E dépendant de μ

Supposons que pour $l \in I_m$, l'efficacité k_{E_l} est une fonction continue, positive croissante et bornée du taux de croissance μ telle que pour tout $\delta\mu$ suffisamment petit, $k_{E_l}(\mu + \delta\mu) = k_{E_l}(\mu) + \dfrac{\partial k_{E_l}(\mu)}{\partial \mu}\delta\mu$.
La propositions 4.1 et 4.3 peuvent être étendues directement dans ce cas en reprenant exactement les mêmes démonstrations car μ est fixé. La proposition 4.4 peut aussi être étendue. La démonstration du point 1 est obtenue en suivant le même raisonnement que la démonstration initiale. Supposons que pour $\mu^+ > 0$, le problème $P_f(\mu^+)$ étendu pour k_{E_l} dépendent de μ est faisable, alors en substituant μ^+ par $\mu + \delta\mu$ avec $\delta\mu > 0$ on obtient pour la contrainte (C_{1a})
Pour tout $i \in I_p$,

$$\begin{aligned}
&-\textstyle\sum_{j=1}^{m} S_{p_{ij}}\nu_j + \mu \left(\alpha_{il}^{p} \dfrac{|\bar{\nu}_l|}{k_{E_l}(\mu)} \sum_{j=1, j\neq l}^{m} C_{M_{ij}}^{M_p} |\nu_j| + C_{R_i}^{M_p} R + C_{G_i}^{M_p} P_G^T \right) - \nu_Y \leq \dots \\
&\quad + \frac{\delta\mu}{k_{E_l}(\mu)} \frac{\partial k_{E_l}(\mu)}{\partial \mu} \left[\textstyle\sum_{j=1}^{m} S_{p_{ij}}\bar{\nu}_j - \mu \left(\sum_{j=1, j\neq l}^{m} C_{M_{ij}}^{M_p} |\bar{\nu}_j| + C_{R_i}^{M_p} \bar{R} + C_{G_i}^{M_p} P_G^T \right) - \nu_Y \right] \dots \\
&\quad - \frac{\delta\mu}{k_{E_l}(\mu)} k_{E_l}(\mu + \delta\mu) \left[\alpha_{il}^{p} \frac{|\bar{\nu}_l|}{k_{E_l}(\mu + \delta\mu)} + \left(\textstyle\sum_{j=1, j\neq l}^{m} C_{M_{ij}}^{M_p} |\bar{\nu}_j| + C_{R_i}^{M_p} \bar{R} + C_{G_i}^{M_p} P_G^T \right) \right]
\end{aligned}$$

Une condition suffisante pour que la partie gauche de l'inégalité soit négative est que la solution $(\bar{R}, \bar{\nu})$ satisfasse l'inégalité suivante :

$$\begin{aligned}
&\frac{\partial k_{E_l}(\mu)}{\partial \mu} \left[\textstyle\sum_{j=1}^{m} S_{p_{ij}}\bar{\nu}_j - \mu \left(\sum_{j=1, j\neq l}^{m} C_{M_{ij}}^{M_p} |\bar{\nu}_j| + C_{R_i}^{M_p} \bar{R} + C_{G_i}^{M_p} P_G^T \right) - \nu_Y \right] \dots \\
&\quad - k_{E_l}(\mu + \delta\mu) \left[\alpha_{il}^{p} \frac{|\bar{\nu}_l|}{k_{E_l}(\mu + \delta\mu)} + \left(\textstyle\sum_{j=1, j\neq l}^{m} C_{M_{ij}}^{M_p} |\bar{\nu}_j| + C_{R_i}^{M_p} \bar{R} + C_{G_i}^{M_p} P_G^T \right) \right] \leq 0
\end{aligned}$$

Le même raisonnement est appliqué pour les contraintes (C_{1c}), (C_2) et (C_3), et on aboutit aux inégalités suivantes

– Pour (C_{1c}^*). Pour tout $i \in I_r$,

$$\left[\alpha_{il}^r \frac{|\bar{\nu}_l|}{k_{E_l}(\mu+\delta\mu)} + \sum_{j=1,j\neq l}^m C_{M_{ij}}^{M_r}|\bar{\nu}_j| + C_{R_i}^{M_r}\bar{R} + C_{G_i}^{M_r}P_G^T\right]\dots$$
$$+\frac{1}{k_{E_l}(\mu+\delta\mu)}\frac{\partial k_{E_l}(\mu)}{\partial\mu}\left[-\sum_{j=1}^m S_{r_{ij}}\bar{\nu}_j - \mu\left(\sum_{j=1,j\neq l}^m C_{M_{ij}}^{M_r}|\bar{\nu}_j| + C_{R_i}^{M_r}\bar{R} + C_{G_i}^{M_r}P_G^T\right)\right] \leq 0$$

– Pour (C_2^*).

$$-\left[n_j^R \frac{|\bar{\nu}_l|}{k_{E_l}(\mu+\delta\mu)} + \sum_{j=1,j\neq l}^m C_{M_j}^R|\bar{\nu}_j| + C_R^R\bar{R} + C_G^R P_G^T\right]\dots$$
$$-\frac{1}{k_{E_l}(\mu+\delta\mu)}\frac{\partial k_{E_l}(\mu)}{\partial\mu}\left[\mu\left(\sum_{j=1,j\neq l}^m C_{M_j}^R|\bar{\nu}_j| + C_R^R\bar{R} + C_G^R P_G^T\right) - k_T\bar{R}\right] \leq 0$$

– Pour (C_3^*).

$$-\frac{\partial k_{E_l}(\mu)}{\partial\mu}\left(\sum_{j=1,j\neq l}^m C_{M_j}^D|\bar{\nu}_j| + C_R^D\bar{R} + C_G^D P_G^T - \bar{D}\right) < 0$$

Le point 2 de la proposition 4.4 étendue au cas où k_E dépend de μ ne pose aucun problème sous l'hypothèse que le point 1 est démontré (c'est-à-dire les inégalités précédentes sont satisfaites) : la démonstration est identique puisque pour tout $\mu \geq 0$, la fonction $k_E(\mu)$ est bornée.

Chapitre 5

Application de la méthode RBA à *Bacillus subtilis*

Dans ce chapitre, nous appliquons la méthode RBA à *Bacillus subtilis* afin de répondre à différentes questions biologiques, et notamment concernant la question centrale de cette thèse : Peut-on effectivement prédire les modules élémentaires dans le réseau métabolique ?
Nous allons résoudre $P_f^{lp}(\mu)$ pour le réseau métabolique et les caractéristiques des protéines de la bactérie modèle *Bacillus subtilis*, pour différents milieux extracellulaires. Compte tenu de la diversité d'applications soulignées dans la section 4.7 du chapitre précédent, nous avons choisi d'illustrer trois résultats théoriques majeurs qui permettent d'éclairer le fonctionnement de la cellule, à savoir :

1. l'existence d'un compromis sur les protéines qui détermine le taux de croissance maximal μ^* (Section 4.3) ;
2. la prédiction des modules fonctionnels élémentaires dans les voies métaboliques (section 4.5.1) ;
3. le transport des nutriments (section 4.7).

Pour chaque résultat théorique, on étudiera différents aspects :

1. **le compromis sur les protéines :** pour différents milieux extracellulaires
   - le taux de croissance maximal μ^* ;
   - la concentration des transporteurs, des enzymes, des ribosomes et les flux métaboliques (section 5.2) ;
   - la répartition des protéines dans la cellule en fonction du taux de croissance (section 5.2.1).
2. **les modules fonctionnels élémentaires** prédits (section 5.4) :
   - la distribution des flux métaboliques selon les différents milieux ;
   - la comparaison avec ceux identifiés dans le chapitre 3 ;
3. **le transport des nutriments :**
   - la substitution de transporteurs haute et basse affinité (section 5.5.1) ;
   - la hiérarchie des sources de sucre (section 5.5.2) ;

– la hiérarchie des sources d'azote (section 5.5.3) ;

L'ensemble de ces prédictions sera confronté à des données expérimentales ou issues de la littérature. Mais identifions tout d'abord les différentes variables et paramètres inclus dans le problème d'optimisation $P_f(\mu)$ (et donc aussi $P_f^{lp}(\mu)$).

5.1 Construction de $P_f^{lp}(\mu)$ pour *Bacillus subtilis*

Le problème de faisabilité $P_f^{lp}(\mu)$ a été implémenté en Matlab 7.2 en utilisant les valeurs numériques données dans la section 5.1.2 et dans les tableaux 5.1, 5.2, 5.3, et en considérant les principales voies métaboliques de *Bacillus subtilis* intégrées dans le modèle de connaissance du chapitre 2 : la voie centrale des carbones, la respiration aérobie, le métabolisme des acides aminés (synthèse, dégradation et transport), les synthèses des nucléotides, de l'ADN, des acides gras, des phospholipides, du peptidoglycane, ainsi que des acides teichoiques [160, 75]. Ces voies métaboliques comprennent 342 gènes, codant pour 277 enzymes et 54 transporteurs, représentant en totalité 358 réactions. Les transporteurs suivants ont notamment été inclus :

- 23 transporteurs pour les acides aminés, dont deux transporteurs haute et basse affinité pour la cystine, la glutamine et la méthionine ;
- 28 transporteurs pour différentes sources de carbone : glucose, fructose, sucrose, mannose, mannitol, tréhalose, lichenan, cellobiose, glucosamine, n-acétylglucosamine, glycérol, maltose, sorbitol, inositol, gluconate, arabinose, ribose, malate, pyruvate, succinate, fumarate, citrate, 2-oxoglutarate et acétate ;
- 2 transporteurs haute et basse affinité pour le phosphate, un transporteur pour le sulfate ;
- l'azote assimilable et l'oxygène sont supposés être transportés par diffusion à travers la membrane.

A partir de ce réseau métabolique, la matrice de stœchiométrie S est calculée. Les précurseurs métaboliques, les métabolites recyclés et les macro-composants sont ensuite identifiés :

- pour les précurseurs métaboliques X_p : les 20 ARNt chargés (pour inclure dans le bilan stœchiométrique l'ATP consommé par les ARNt synthases) ainsi que les nucléotides triphosphates (ATP, GTP, UTP et CTP) ;
- pour les métabolites recyclés X_r : le GDP résultant de l'hydrolyse du GTP lors de la traduction des protéines ;
- pour les macro-composants X_c : l'ADN, le peptidoglycane, les acides teichoiques ainsi que les principaux phospholipides composant la membrane de *B. subtilis* (mono-, di- et triglucosyldiacylglycérol, cardiolipine, phosphatidylglycérol, lysyl-phosphatidylglycérol, phosphatidyléthanolamine) [188, 229] ;
- les métabolites internes X_i correspondant à l'ensemble des métabolites du réseau métabolique non inclus dans X_p, X_r et X_c.

Les sous-matrices S_p, S_r, S_c et S_I de $P_f^{lp}(\mu)$ sont ainsi également déduites de la matrice S.

Nous avons ensuite supposé, pour simplifier, que l'appareil de traduction des protéines

Acides aminés	Proportion (%)
alanine	4.5
arginine	6.4
asparagine	3.7
aspartate	3.7
cystéine	1.3
glutamine	7.2
glutamate	7.2
glycine	5.8
histidine	2.4
isoleucine	6.7
leucine	8.6
lysine	9.0
méthionine	3.2
proline	5.5
phénylalanine	3.5
sérine	4.3
thréonine	4.2
tryptophane	2.1
tyrosine	3.8
valine	6.8

TAB. 5.1 – Composition moyenne en acides aminés des protéines de *Bacillus subtilis* [8]

est composé de ribosomes, et des deux facteurs d'élongation EF-G et EF-Tu. Un ribosome est composé d'un ARN ribosomal (ARNr) et de 52 protéines ribosomales [122]. Dix opérons codent pour des ARNr dans *B. subtilis*. Nous en avons donc choisi un connu pour être exprimé en phase exponentielle. Il est composé des gènes *rrnO-16S*, *rrnO-23S* et *rrnO-5S*, d'une longueur totale de 4593 nucléotides.

On aboutit alors à 359 variables de décision pour $P_f(\mu)$ correspondant à 717 variables de décision pour $P_f^{lp}(\mu)$, 753 contraintes inégalités et 247 contraintes égalités.

5.1.1 Calcul des cœfficients

Les différentes variables et contraintes ayant été caractérisées, déterminons maintenant l'ensemble des cœfficients inclus dans les contraintes de $P_f^{lp}(\mu)$.

Pour chaque protéine (enzyme, transporteur, protéine ribosomale ou protéine dans $\mathcal{P}_G$), nous avons considéré sa longueur exacte en acides aminés et la composition moyenne en acides aminés proposée dans [8] et rappelée dans le tableau 5.1 pour calculer les différents cœfficients (C_M^M, C_M^R, C_M^D, C_R^M, C_R^R, C_R^D, C_G^M, C_G^R, C_G^D) de $P_f^{lp}(\mu)$. Lorsque le gène codant pour une enzyme ou un transporteur était inconnu dans l'annotation officielle de *Bacillus subtilis*, nous avons considéré une protéine de 200 acides aminés. Les

Cœfficient	Contrainte	X_p/X_r	Valeur
$C^{M_p}_{M_{ij}}$	C^{lp}_{1a}	acide aminé	s_{ij}/k_{E_j}
$C^{M_p}_{M_{atp,j}}$	C^{lp}_{1a}	ATP	0
$C^{M_p}_{M_{gtp,j}}$	C^{lp}_{1a}	GTP	$(1+2m_j)/k_{E_j}$
$C^{M_p}_{M_{ctp,j}}$	C^{lp}_{1a}	CTP	0
$C^{M_p}_{M_{utp,j}}$	C^{lp}_{1a}	UTP	0
$C^{M_p}_{R_i}$	C^{lp}_{1a}	acide aminé	$\sum_{j=1}^{52} s_{ij}n_j$
$C^{M_p}_{R_{atp}}$	C^{lp}_{1a}	ATP	r_{atp}
$C^{M_p}_{R_{gtp}}$	C^{lp}_{1a}	GTP	$r_{gtp} + n_{R,gtp} + \sum_{j=1}^{52}(1+2m_j)n_j$
$C^{M_p}_{R_{ctp}}$	C^{lp}_{1a}	CTP	r_{ctp}
$C^{M_p}_{R_{utp}}$	C^{lp}_{1a}	UTP	r_{utp}
$C^{M_p}_{G_i}$	C^{lp}_{1a}	acide aminé	$\sum_{j=1}^{N_G} s_{ij}$
$C^{M_p}_{G_{atp}}$	C^{lp}_{1a}	ATP	0
$C^{M_p}_{G_{gtp}}$	C^{lp}_{1a}	GTP	$\sum_{j=1}^{N_G}(1+2m_j)$
$C^{M_p}_{G_{ctp}}$	C^{lp}_{1a}	CTP	0
$C^{M_p}_{G_{utp}}$	C^{lp}_{1a}	UTP	0
$C^{M_r}_{M_{gdp,j}}$	C^{lp}_{1c}	GDP	$(1+2m_j)/k_{E_j}$
$C^{M_r}_{R_{gdp}}$	C^{lp}_{1c}	GDP	$n_{R,gtp} + \sum_{j=1}^{52}(1+2m_j)n_j$
$C^{M_r}_{G_{gdp}}$	C^{lp}_{1c}	GDP	$\sum_{j=1}^{N_G}(1+2m_j)$
$C^{R}_{M_j}$	C^{lp}_{2}	-	m_j/k_{E_j}
C^{R}_{R}	C^{lp}_{2}	-	$\sum_{j=1}^{52} m_j n_j$
C^{R}_{G}	C^{lp}_{2}	-	$\sum_{j=1}^{N_G} m_j$
$C^{D}_{M_j}$	C^{lp}_{3}	-	m_j/k_{E_j}
C^{D}_{R}	C^{lp}_{3}	-	$\sum_{j=1}^{52} m_j n_j$
C^{D}_{G}	C^{lp}_{3}	-	$\sum_{j=1}^{N_G} m_j$

TAB. 5.2 – Cœfficients de $P^{lp}_f(\mu)$ (et $P_f(\mu)$)

Macro-composants	Concentration (μmol/gdwc)
ADN	80.7
peptidoglycane	101.82
acides teichoïques	10.9
monoglucosyldiacylglycérol	0.085750
diglucosyldiacylglycérol	0.110292
triglucosyldiacylglycérol	0.065833
cardiolipine	0.004642
phosphatidylglycérol	0.175859
lysylphosphatidylglycérol	0.022057
phosphatidyléthanolamine	0.559509

TAB. 5.3 – Concentration des macro-composants $\bar{X}_c$ en micromoles par gramme de poids sec (μmol/gdwc) intégrés dans $P_f(\mu)$ pour $\mu = 0.1h^{-1}$ [160]

cœfficients du problème d'optimisation sont donc déduits (cf. tableau 5.2) en explicitant pour chaque protéine et pour chaque ribosome le nombre de métabolites consommés ou produits permettant de maintenir constante la concentration de ces protéines et ribosomes en régime établi (cf. équation (4.2)), et avec pour une protéine $\mathcal{P}_j$ les notations suivantes :

- m_j la longueur totale en acides aminés de la protéine $\mathcal{P}_j$;
- n_j le nombre de protéines $\mathcal{P}_j$ si un ribosome contient plusieurs copies de $\mathcal{P}_j$;
- s_{kj} le nombre du k-ème acide aminé composant $\mathcal{P}_j$, pour $k \in \{1, \ldots, 20\}$

Par définition, on a $\sum_{k=1}^{20} s_{kj} = m_j$. Pour le calcul des cœfficients C_R^M associés aux nucléotides, nous avons considéré la composition exacte de l'ARNr en nucléotides avec r_k le nombre du k-ème nucléotide pour $k \in \{atp, gtp, ctp, utp\}$ et $n_{R,gtp}$ le nombre de GTP consommé durant l'assemblage du ribosome par les GTPases.

5.1.2 Identification des paramètres

Les paramètres physiologiques $\bar{X}_c$ et $\bar{D}$ de *B. subtilis* sont obtenus à partir de la littérature. Pour $\bar{X}_c$, défini comme les concentrations cibles de l'ADN, des phospholipides, du peptidoglycane et des acides teichoiques durant la phase exponentielle, nous avons utilisé les concentrations moyennes mesurées en micromoles par gramme de poids sec (μmol/gdwc) dans [160] et rappelées dans le tableau 5.3.

Nous avons ensuite considéré une densité moyenne $\bar{D}$ de 5.411 millimoles par gramme de poids sec (noté mmol/gdwc par la suite). Cette valeur de densité intégre à la fois la densité des protéines et la densité des ARNr convertie en densité équivalente de protéines pour *B. subtilis*. Pour cela, nous avons utilisé :

- la densité en protéines mesurée dans [185], égale à 4.910 millimoles par gramme de poids sec correspondant à 52% de la masse sèche ;
- le poucentage d'ARNr par cellule égal à 5.3% déduit du pourcentage d'ARN total mesuré par cellule (6.55% [185]) et de la fraction d'ARNr composant l'ensemble des

ARNs (81% [153]) ;
- la masse moyenne en gramme par mole d'un nucléotide (337.45) et d'un acide aminé (136.9) déterminées dans [160].

On déduit donc que la densité équivalente en protéines des ARNr vaut 0.5 mmol/gdwc en calculant la masse totale de la cellule ($4.910 * 136.9/0.52 = 1292.7$mg), puis la masse des ARNr ($1292.7 * 0.053 = 68.6$mg) et enfin la densité équivalente en protéines ($68.6/(337.45 * 136.9) = 0.5011$mmol/gdwc). La densité moyenne totale $\bar{D}$ correspond donc à 5.411mmol/gdwc ($4.910 + 0.5011$).

Dans le cadre de cette thèse, nous avons identifié les paramètres k_E, k_T d'après la littérature :
- pour k_T, une efficacité moyenne de traduction de 15 acides aminés par seconde [28] ;
- pour k_E, le cœfficient de conversion entre le flux et la concentration de protéine, une même vitesse de 20 s^{-1} pour toutes les enzymes, déterminée comme la vitesse moyenne des différentes enzymes disponibles dans la base de donnée BRENDA [193].

En outre, au lieu de considérer directement les flux d'import des nutriments extracellulaire, nous introduisons directement la concentration extracellulaire des nutriments dans $P_f(\mu)$. Le cœfficient k_E est ainsi remplacé par une fonction non-linéaire de type Michaëlis-Menten (voir la section 2.3.1 page 39) de la concentration externe S_i du nutriment pour chaque transporteur :

$$k_E = k_{max} \frac{S_i}{K_{m_i} + S_i} \tag{5.1}$$

avec K_{m_i} la constante d'affinité du transporteur, et le turnover k_{max} dont les valeurs dépendent de l'affinité du transporteur pour son substrat. Dans le cas d'un transporteur basse affinité, $K_{m_i} = 0.8$mM et $k_{max} = 20s^{-1}$ alors que $K_{m_i} = 0.008$mM et $k_{max} = 10s^{-1}$ pour un transporteur haute affinité.
Néanmoins, ces paramètres k_E et k_T pourraient être identifiés à partir de données expérimentales, notamment à partir de mesures de flux et de protéines quantitatives, c'est-à-dire des concentrations absolues de protéines.

Enfin, comme nous l'avons souligné dans le chapitre précédent, la composition de l'ensemble $\mathcal{P}_G$ est difficile à déterminer car elle dépend des conditions environnementales considérées. De plus, même dans le cas où la composition de cet ensemble serait déterminée, nous ne disposons pas à ce jour des concentrations absolues des protéines P_{G_i} pour chaque protéine de cet ensemble. Nous avons donc réduit notre ensemble à une seule protéine fictive de 200 acides aminés, de concentration notée P_G. Nous avons fixé sa concentration à 40% de la totalité des protéines, c'est-à-dire à 40% de la densité en protéines 4.910 mmol/gdwc mesurée [185]. Ainsi la concentration P_G est égale à 0.0098mmol/gdwc, déterminée par $4.910 * 0.4/200 = 0.0098$ afin d'obtenir des valeurs de taux de croissance prédites en accord avec la littérature [184].

5.2 Répartition des ressources vis-à-vis du taux de croissance

Dans un premier temps, nous illustrons le compromis existant entre les protéines et le taux de croissance à travers le calcul de la répartition en protéines entre le réseau métabolique et les ribosomes pour différents milieux, correspondant à des variations théoriques (et expérimentales) importantes de taux de croissance. Pour cela, nous considérons 21 milieux, en partant d'un milieu minimum pauvre composé de glucose, azote assimilable, sulfate, phosphate et oxygène, auquel on ajoute chacun des 20 acides aminés l'un après l'autre. On obtient finalement un milieu riche comprenant tous les acides aminés : alanine (ala), arginine (arg), asparagine (asn), aspartate (asp), cystéine (cys), glutamine (gln), glutamate (glu), glycine (gly), histidine (his), isoleucine (ile), leucine (leu), lysine (lys), méthionine (met), phénylalanine (phe), proline (pro), sérine (ser), thréonine (thr), tryptophane (trp), tyrosine (tyr) et valine (val).

5.2.1 Répartition des protéines entre les ribosomes et les voies métaboliques

Sur la figure 5.1, nous avons représenté les prédictions de la répartition des protéines entre les ribosomes et les voies métaboliques obtenues pour différents milieux extracellulaires et donc différents taux de croissance. La concentration en protéines dédiées aux ribosomes augmente linéairement avec le taux de croissance alors que celle dédiée aux voies métaboliques diminue. Nous obtenons ici qualitativement le même type de répartition de ressources que pour *E. coli* [142] (voir figure 5.2) et *S. cerevisiae* [25] : trois ensembles de gènes peuvent être distingués. L'expression de deux groupes est μ-dépendant, l'un induit, l'autre réprimé alors que l'expression du troisième est μ-indépendant et correspond à notre ensemble $\mathcal{P}_G$ [25].
Nos résultats et ceux de [142, 25] impliquent que pour un ensemble de conditions environnementales donné, toute protéine économisée, par la répression d'une voie métabolique par exemple, réduit la quantité de précurseurs métaboliques allouée au réseau métabolique[1]. Ces précurseurs économisés peuvent être ensuite répartis entre les trois ensembles de protéines ($\mathcal{P}_G$, métabolique et ribosomale) pour augmenter la concentration des ribosomes et donc augmenter le taux de croissance. Réciproquement, tout précurseur métabolique non dédié à accroître la concentration des ribosomes conduit à diminuer le taux de croissance. Pour renforcer cette idée, C. Tanous (post-doctorante dans notre unité) a conduit des expériences biologiques déséquilibrant cette répartition en protéines.

Validation *in vivo* du principe de la répartition des protéines

Si la quantité de précurseurs métaboliques dédiée à la maintenance de la concentration des ribosomes en régime établi augmente alors le taux de croissance doit diminuer.

[1] et la densité intracellulaire.

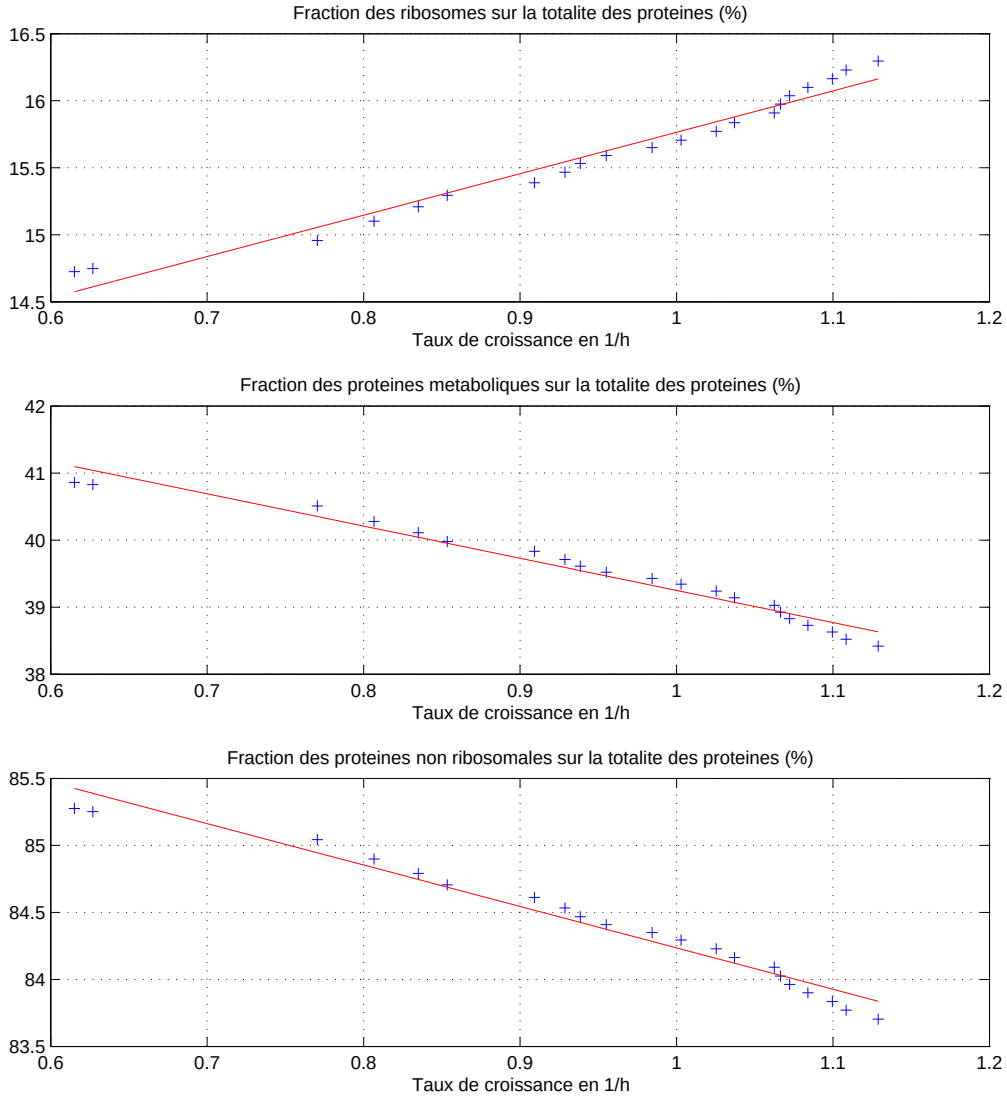


FIG. 5.1 – Répartition des protéines entre les ribosomes, les voies métaboliques et toutes les protéines non ribosomales ($\mathcal{P}_M$ et $\mathcal{P}_G$)

Pour cela, il suffit de mimer un ribosome «plus gros», intégrant une protéine suppplémentaire par exemple. Nous avons donc choisi d'exprimer le gène *lacZ* codant pour la β-galactosidase sous le contrôle du promoteur d'ARN ribosomal *rrnO*. La β-galactosidase est une protéine longue de 687 acides aminés, ce qui représente environ 10% de la quantité totale d'acides aminés composant un ribosome. C. Tanous a réalisé une fusion *lacZ* (*PrrnO-lacZ*) du promoteur P1 de *rrnO*, qui a ensuite été intégrée en simple copie au

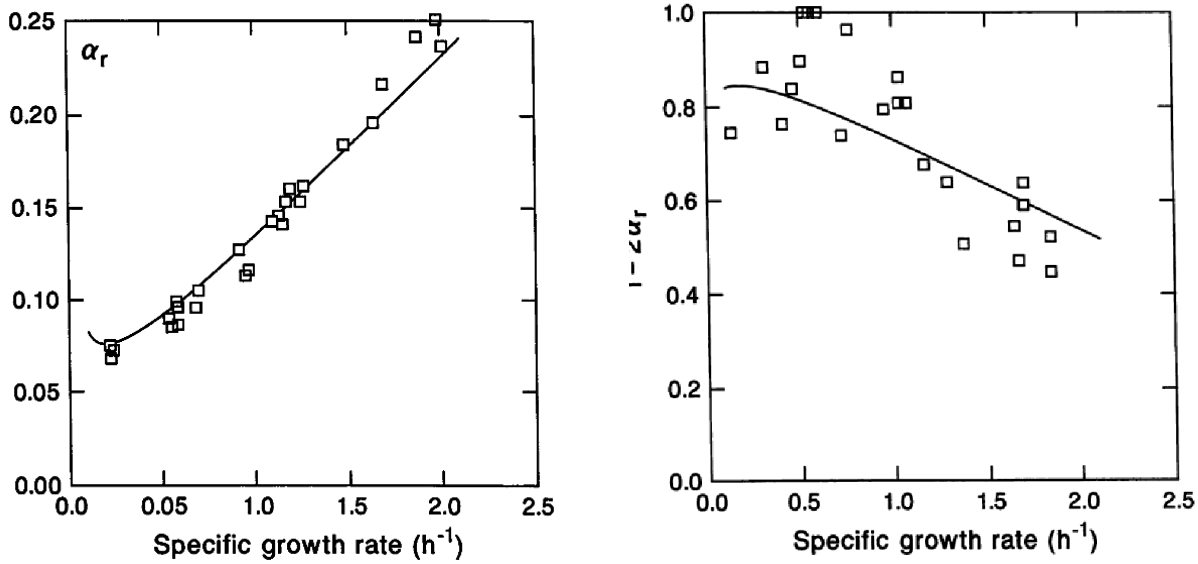


FIG. 5.2 – Evolution des ribosomes et des protéines métaboliques avec le taux de croissance pour *E. coli* (extrait de [142])

locus *amyE* de *B. subtilis*.
Nous avons comparé la croissance de la souche sauvage et de la souche *PrrnO-lacZ* dans un milieu minimum et dans un milieu riche comprenant les 20 acides aminés (cf. la figure 5.3). Le taux de croissance est réduit de 20% (resp. 10%) dans le cas du milieu riche (resp. minimum). De plus, la période de latence augmente dans les deux milieux, mais elle est beaucoup plus importante en milieu minimum.
L'impact sur le taux de croissance est plus fort en milieu riche. En effet, la cellule produit majoritairement des ribosomes à des taux de croissance très élevés [28]. Le coût de production d'un ribosome en terme de précurseurs métaboliques ayant sensiblement augmenté dans la souche *PrrnO-lacZ*, il est naturel que son impact soit plus fort à taux de croissance élévé qu'à taux de croissance faible.
Cette expérience renforce donc l'idée que la bactérie a tout intérêt à développer des stratégies complexes comme les régulations génétiques pour réaliser des économies de protéines afin d'augmenter le taux de croissance. Pour le métabolisme, il s'agit d'éteindre des voies métaboliques lorsqu'elles ne sont pas nécessaires. A cette fin, nous allons maintenant détailler la répartition des protéines entre les différentes voies métaboliques afin d'identifier précisément dans quelles voies métaboliques sont réalisées les économies de protéines.

Répartition détaillée entre les ribosomes, le carbone central, la respiration et les voies de synthèse des acides aminés

La figure 5.4 illustre pour ces 21 milieux la répartition des protéines entre les ribosomes, et les protéines impliquées dans les voies métaboliques majoritaires (en terme de

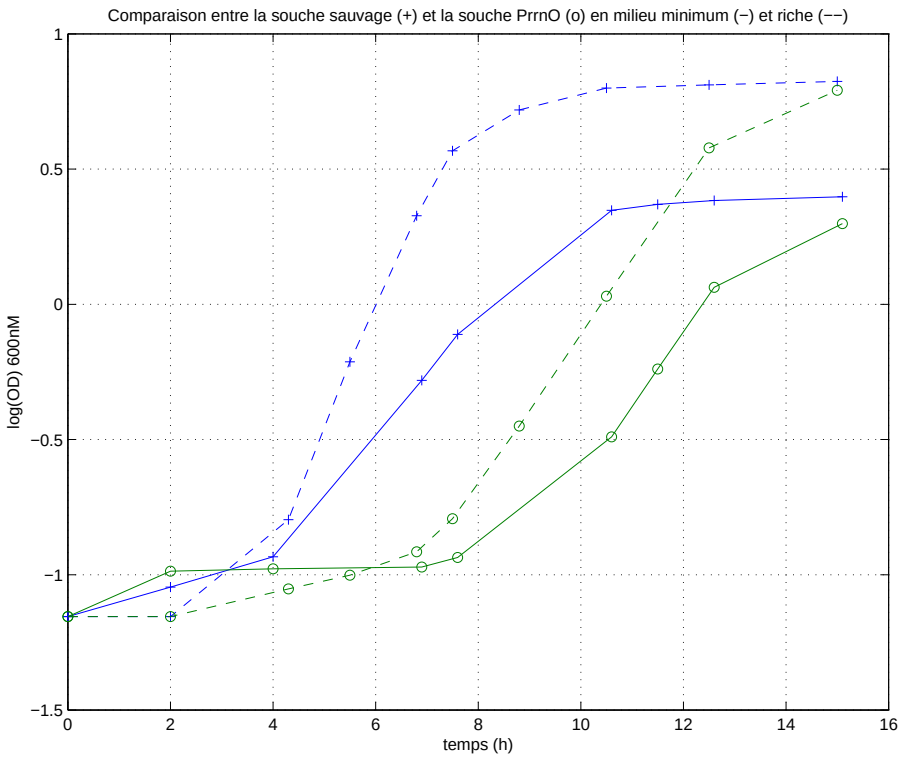


FIG. 5.3 – Croissance de la souche sauvage et de la souche *PrrnO-lacZ* en milieu riche et en milieu minimum

flux) à savoir la voie centrale des carbones, la respiration aérobie et les voies de synthèse des acides aminés (transporteurs et voie de synthèse *de novo*). La proportion de protéines dédiées aux ribosomes augmente nettement avec le taux de croissance, alors que celle dédiée aux de synthèses des acides aminés diminue. La proportion des protéines impliquées dans la voie centrale des carbones reste constante alors que celle associée à la respiration augmente légèrement. Ainsi ce sont les économies en protéines réalisées au niveau des voies de biosynthèse des acides aminés qui permettent d'augmenter le taux de croissance. Les protéines économisées sont alors réparties entre les ribosomes et les autres voies métaboliques encore actives pour augmenter le taux de croissance. Nous allons maintenant détailler plus particulièrement l'adaptation des voies métaboliques (et en particulier les voies de synthèse des acides aminés).

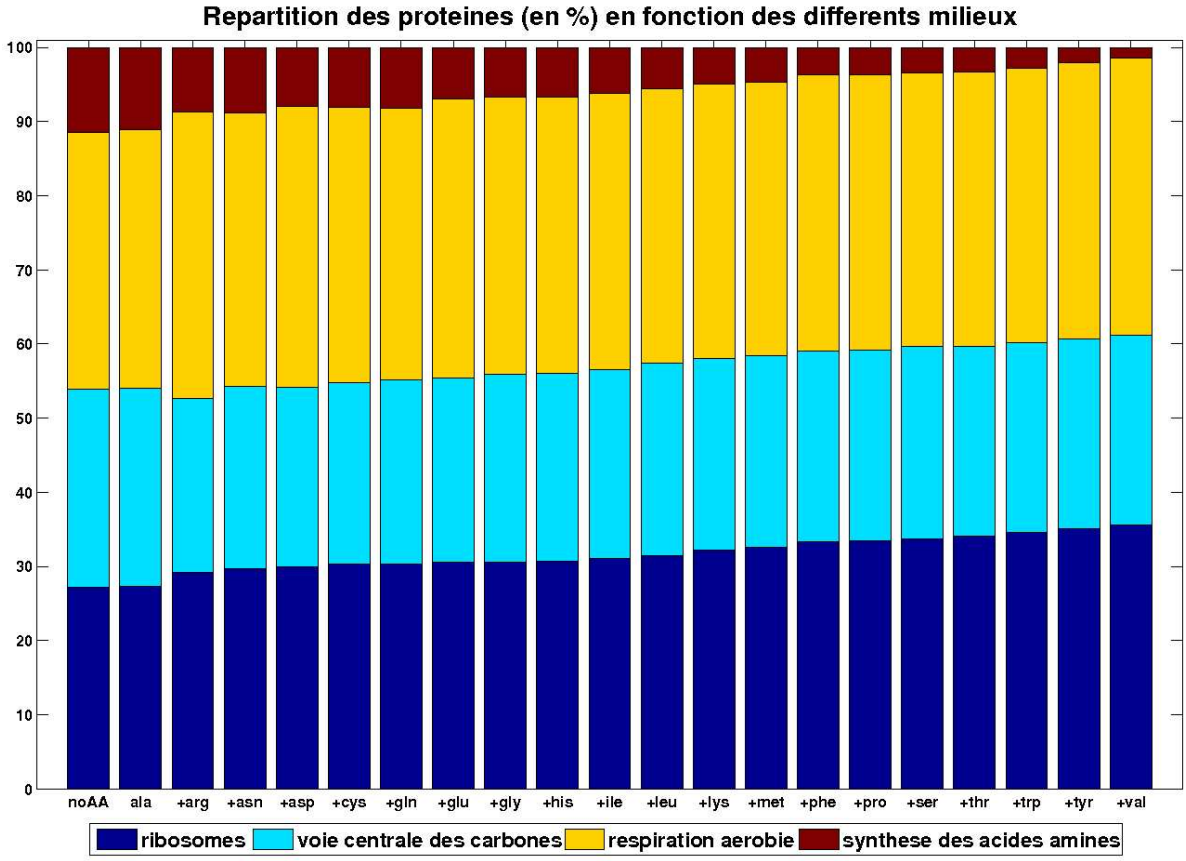


FIG. 5.4 – Répartition des protéines prédite entre les ribosomes et les voies métaboliques

5.3 Conséquences pour les voies métaboliques

5.3.1 Impact sur la biosynthèse des acides aminés

Les figures 5.5 et 5.6 représentent l'évolution des protéines appartenant respectivement aux voies de biosynthèse et des transporteurs des acides aminés en fonction des différents milieux (et donc en fonction du taux de croissance). En abscisse se trouvent les différents milieux, où chaque acide aminé est ajouté l'un après l'autre. Par exemple, l'abscisse «+his» signifie que l'histidine est ajoutée. En ordonnée, se trouvent les voies de synthèse des acides aminés. Par exemple, sur la figure 5.5, l'ordonnée «his» correspond à la voie de synthèse *de novo* de l'histidine, composée des gènes *hisG*, *hisI*, *hisA*, *hisH*, *hisF*, *hisB*, *aroJ*, *hisC* et *hisD*. Les concentrations des protéines appartenant à chaque voie métabolique ont été additionnées pour obtenir la concentration totale de protéines pour un milieu donné pour chaque voie métabolique. De plus, les deux premières lignes correspondent respectivement aux variations prédites du taux de croissance et des concentrations des ribosomes.
Chaque concentration totale de protéines, de ribosomes et de taux de croissance sont ensuite normalisés sur les différents milieux pour obtenir une variation entre 0 et 1, où

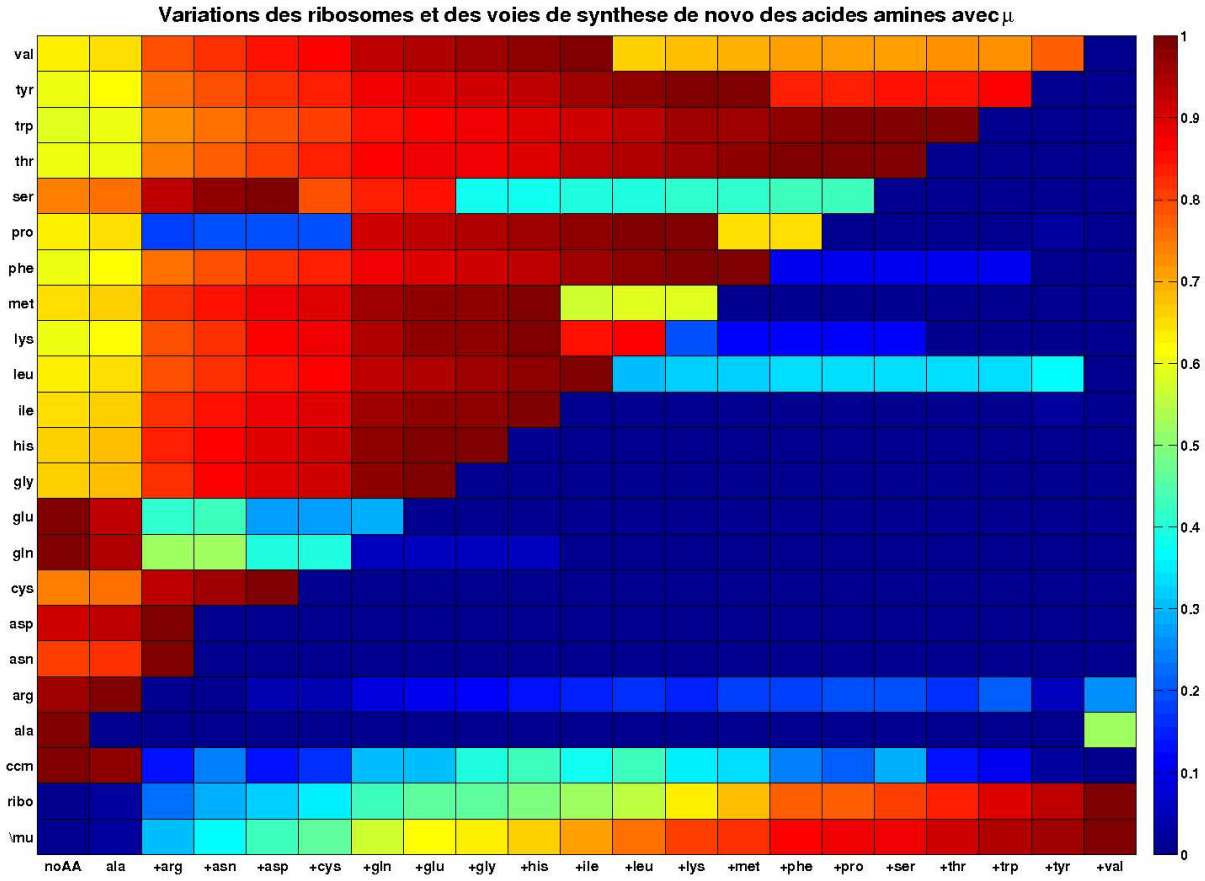


FIG. 5.5 – Evolution des voies de synthèse *de novo* des acides aminés : bleu (réprimé), rouge (fort)

le niveau 0 correspond à la valeur la plus faible (en bleu foncé sur les deux figures) et le niveau 1 la valeur la plus forte (en rouge foncé). Pour le j-ème milieu, la concentration normalisée $\bar{C}_j$ est donnée par

$$\bar{C}_j = \frac{C_j - C_m}{\max_{i\in\{1,\ldots,21\}}(C_i) - C_m}$$

où $C_m = \min_{i\in\{1,\ldots,21\}} C_i$. En outre, pour les voies de synthèse *de novo* d'acides aminés ou pour leurs transporteurs, le niveau 0 (en bleu foncé) correspond à la répression totale de la voie, c'est-à-dire que les concentrations des protéines sont nulles.

Au fur et à mesure que les acides aminés sont ajoutés sur les deux figures 5.5 et 5.6, le taux de croissance prédit augmente, tout comme les concentrations prédites des ribosomes. De plus, la comparaison de ces deux figures montre que lors de l'ajout d'un acide aminé, sa voie de synthèse *de novo* est réprimée (en bleu foncé) et son transporteur est induit (en rouge). Prenons l'exemple de la voie de biosynthèse de l'histidine (ligne et colonne «his/+his» sur les figures 5.5 et 5.6) afin de bien caractériser le principe général

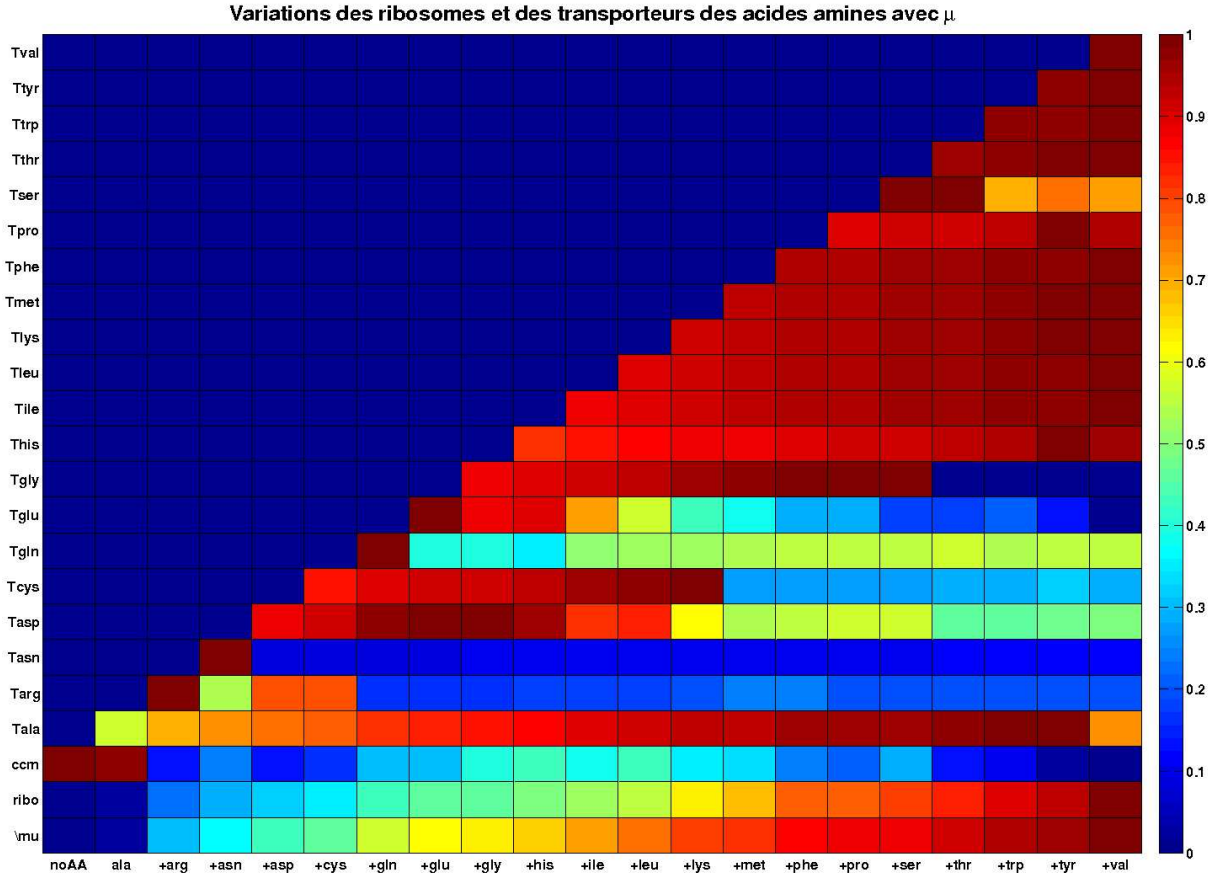


FIG. 5.6 – Evolution du transport des acides aminés :bleu (réprimé), rouge (fort)

de fonctionnement.

La voie de synthèse de l'histidine comme principe de fonctionnement

Lorsque le taux de croissance augmente, la quantité de biomasse produite augmente, de sorte que le flux d'acides aminés consommé nécessaire à la production de l'ensemble de la biomasse augmente. Pour la voie de l'histidine en particulier, le flux de synthèse de l'histidine augmente avec le taux de croissance, quelle que soit la voie de synthèse (synthèse *de novo* ou transporteur).
Or nous avons supposé dans le problème d'optimisation que le flux métabolique était directement proportionnel à la concentration de l'enzyme à travers le cœfficient k_E constant. Par conséquent, nos résultats indiquent qu'avant l'ajout d'histidine, la concentration des protéines impliquées dans la voie de biosynthèse *de novo* augmente pour satisfaire la demande d'histidine pour la croissance (contrainte C_{1a} de $P_f(\mu)$). Après l'ajout d'histidine, la voie de biosynthèse est complètement réprimée, alors qu'en même temps, le transporteur est induit. Ensuite, la concentration du transporteur augmente avec le taux de croissance, pour satisfaire la demande de flux en histidine. Le transporteur

se substitue à la voie de biosynthèse.

Remarque 5.1 *Le fonctionnement réel est bien évidemment plus complexe. Le cœfficient k_E correspond à l'activité de l'enzyme et est par conséquent une fonction non linéaire des substrats et des produits. Des variations de ces métabolites peuvent modifier k_E et donc modifier le flux. L'adaptation au flux demandé est obtenue conjointement en adaptant la concentration des enzymes de la voie et les métabolites. On peut facilement imaginer que si k_E augmente suffisamment, la demande en flux puisse être satisfaite uniquement par des variations de métabolites. Nous avons d'ailleurs présenté dans la section 3.3 un tel exemple d'adaptation des flux par des variations de métabolites, pour des valeurs faible du taux de croissance. On pourrait alors remplacer k_E par une fonction croissante et bornée de μ comme nous l'avons souligné dans la section 4.7. Néanmoins, nous obtenons un comportement cohérent avec la physiologie connue de B. subtilis malgré ces simplifications.*

La voie de synthèse de l'histidine illustre d'une part le principe général du compromis protéines/ribosomes, et d'autre part la substitution de deux voies métaboliques ayant des coûts d'induction en protéines différent[2]. Or une analyse de toutes les voies de synthèse des acides aminés de la figure 5.5 montre que le comportement de certaines voies en fonction des différents milieux diffère de celui de l'histidine correspondant à (*i*) l'augmentation du flux dans la voie de synthèse, puis (*ii*) à la répression de la voie de synthèse *de novo* et à l'induction du transporteur lors de l'ajout de l'acide aminé. Les voies de synthèse de la proline et de la sérine présentent par exemple des variations importantes alors même que la proline et la sérine ne sont pas encore ajoutées dans le milieu. En effet, il ne faut pas oublier que le réseau métabolique est un réseau fortement interconnecté. Certains acides aminés, comme la sérine, sont par exemple impliqués dans d'autres voies de biosynthèse d'acides aminés. D'autres, comme l'arginine, peuvent être dégradés et utilisés comme source préférentielle d'azote. Or la solution optimale correspond à celle qui économisera le plus de protéines dans le réseau métabolique. Par conséquent, la répartition globale des protéines dépend à la fois de la stœchiométrie du réseau métabolique et du coût en protéines nécessaire à l'induction de chaque voie métabolique. Afin de bien comprendre l'impact de cette double contrainte (coûts de production et stœchiométrie), nous allons détailler le cas de la synthèse de la sérine.

La synthèse de la sérine

La sérine est un acide aminé impliqué dans les voies de synthèse de la cystéine, de la glycine et du tryptophane [75]. De plus, la cystéine est également un précurseur de la voie de synthèse de la méthionine chez *Bacillus subtilis* [97] (voir la figure 5.7). Par conséquent, la sérine l'est aussi indirectement. En milieu minimum sans acide aminé, la capacité de la voie de biosynthèse de la sérine doit être suffisante pour couvrir non seulement le flux de synthèse de sérine nécessaire pour la synthèse des protéines, mais

[2]On rappelle ici que le coût en protéines d'induction d'une voie métabolique réprésente à la fois le coût en précurseurs métaboliques et en ribosomes nécessaires à la synthèse de la protéine, et le coût équivalent en densité intracellulaire.

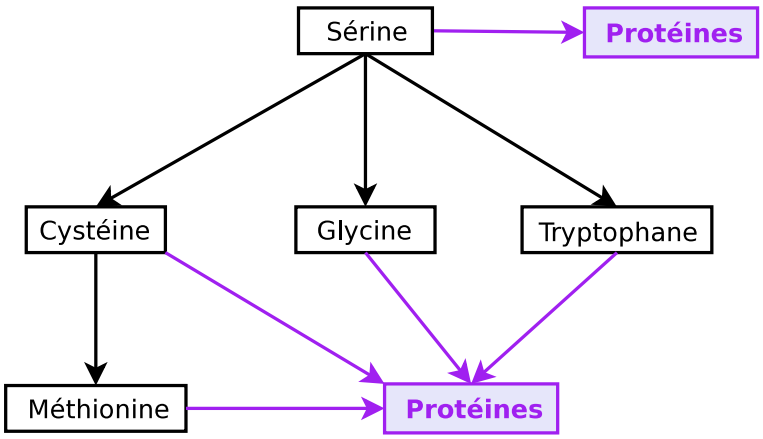


FIG. 5.7 – La sérine comme précurseur de la cystéine, la glycine, le tryptophane et la méthionine

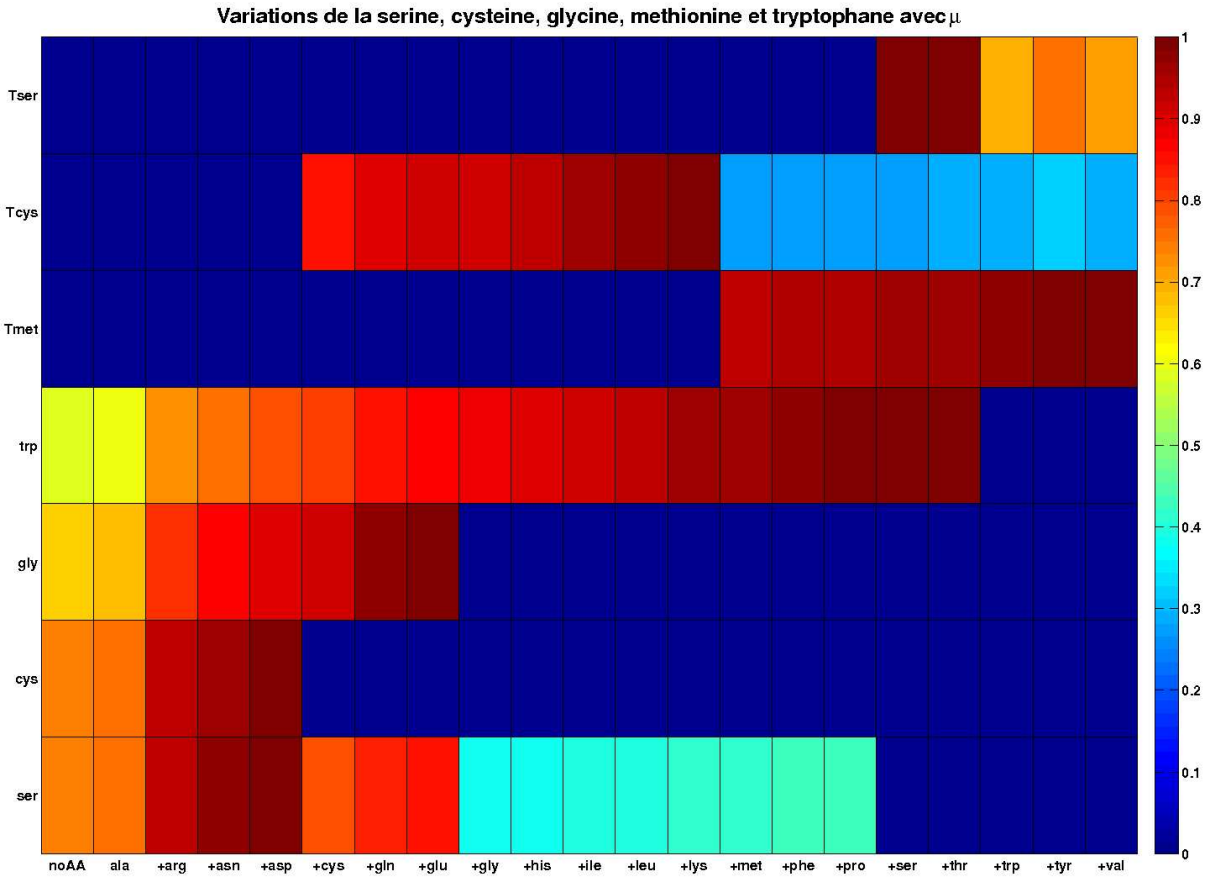


FIG. 5.8 – Evolution de la cystéine, la glycine, le tryptophane et la méthionine au regard de la sérine

aussi les flux de synthèse de cystéine, tryptophane, de glycine et de méthionine. Ainsi sur la figure 5.8, on constate que la concentration des protéines de la voie de synthèse *de novo* de la sérine (ligne «ser») diminue lors de l'ajout de cystéine (colonne «+cys») et de glycine (colonne «+gly»). En effet, les synthèses de cystéine et glycine sont alors réalisées via les transporteurs qui ne consomment pas de sérine. Par conséquent, le flux de synthèse de sérine doit alors uniquement couvrir les besoins en sérine et en tryptophane, donc la concentration des protéines impliquées dans la voie de synthèse *de novo* de la sérine diminue. Lors de l'ajout de sérine (colonne «+ser»), le transport de sérine est induit (ligne «Tser») et fournit alors le flux de synthèse de sérine et de tryptophane. Après l'ajout de tryptophane (colonne «+trp»), la concentration des transporteurs de sérine diminue en conséquence (voir la figure 5.8). On notera également la diminution de la concentration des transporteurs de cystéine (ligne «Tcys») lors de l'ajout de méthionine (colonne «+met») reflétant l'économie de protéines réalisées au niveau de la synthèse de la cystéine.

5.3.2 Prédiction de l'adaption des voies métaboliques pour différents milieux

Pour expliquer le comportement de certaines voies métaboliques comme celle de la proline lors de l'ajout d'arginine, il est donc nécessaire de considérer (et de représenter) le réseau métabolique dans son ensemble afin de repérer les connexions critiques. Compte tenu du nombre important de réactions composant le réseau métabolique (351), il faut concevoir une représentation concise mais pertinente du réseau métabolique et des résultats associés, sans perte d'information. Nous avons ainsi choisi une visualisation graphique de la voie centrale des carbone ainsi que des voies de synthèse et de dégradation des acides aminés qui sont particulièrement impactées par les variations du milieu que nous considérons. Nous projetons sur cette représentation les prédictions des flux métabolique pour différents milieux et ce pour chaque gène impliqué dans les voies métaboliques. Les deux figures en annexe B.2 page 205 représentent la répartition des flux[3] pour,

- les 21 milieux où les acides aminés sont incorporés l'un après l'autre (figure B.2.1 page 205) ;
- 22 milieux différents pour les 22 sources de carbone suivantes (figure B.2.2 page 205) : glucose (glc), fructose (fru), mannose (man), sucrose (suc), mannitol (mtl), tréhalose (tre), lichenan (lic), cellobiose (cel), glycérol (glo), maltose (malt), ribose (rib), sorbitol (sor), inositol (ino), gluconate (glcn), arabinose (ara), malate (mal), pyruvate (pyr), succinate (succ), fumarate (fum), citrate (cit), 2-oxoglutarate (akg) et lactate (lct).

Cette représentation visuelle permet de comparer l'adaptation des voies métaboliques en fonction des différents milieux. Pour chaque enzyme appartenant à une voie métabolique, la valeur du flux est indiquée par une couleur, dont l'échelle est systématiquement représentée sur la sous-figure d'intérêt. L'échelle de couleur est centrée autour de 0 (vert),

[3]et donc des concentrations des enzymes et des transporteurs au cœfficient de conversion k_E près.

où les valeurs négatives sont indiquées en bleu et les valeurs positives en rouge. Lors de la section précédente, nous soulignions que le comportement de la synthèse de la sérine différait de celui de l'histidine. On constate sur la figure B.2.1 que la synthèse de la proline s'effectue via la dégradation de l'arginine par les enzymes EargI, ErocD et EproI correspondant aux gènes *argI*, *rocD* et *proI* de *Bacillus subtilis* pour les milieux contenant de l'arginine, de l'asparagine, de l'aspartate et de la cystéine (voir la figure nommée *arginine/proline degradation*). Dans la suite de notre analyse et sur les figures B.2.1 et B.2.2 pages 205, nous écrirons pour simplifier que la concentration des enzymes augmente lorsque le flux métabolique dans une voie métabolique augmente du fait de notre coefficient de conversion k_E. De même, nous noterons les concentrations prédites des enzymes avec un «E» majuscule, pour les différencier des noms des enzymes classiques. Par exemple, la prédiction de la concentration de l'enzyme ArgI, codée par le gène *argI*, est noté EargI.

Une première analyse «gros grain» de la distribution des flux indique que nous recouvrons la physiologie connue de *B. subtilis*, et ce à différents niveaux :

- en présence d'un acide aminé donné, les voies de biosynthèse sont réprimées et les transporteurs sont induits (annexe B.2.1) ;
- les concentrations des enzymes associées aux voies de synthèse d'un acide aminé (synthèse *de novo* ou transporteur) et de la ARNt synthase augmentent avec le taux de croissance (annexe B.2.1) car la quantité de biomasse à produire augmente ;
- en présence d'une source de carbone donnée, les transporteurs et les voies d'assimilation associées sont induits (annexe B.2.2) ;
- les enzymes EgapA et Epfk (correspondant aux gènes *gapA* et *pfkA*) sont induites en conditions glycolytiques et réprimées en conditions néoglucogéniques (voir la sous-figure *EMP* de l'annexe B.2.2). Le flux glycolytique est positif, indiquant que le flux part du «haut», c'est-à-dire d'une source de carbone glycolytique, vers le «bas» (vers le pyruvate) ;
- les enzymes EgapB et Epck (correspondant aux gènes *gapB* et *pckA*) sont induites en conditions néoglucogéniques et réprimées en conditions glycolytiques (voir la sous-figure *EMP* de l'annexe B.2.2). Le flux glycolytique est ici négatif, indiquant que le flux remonte la glycolyse pour alimenter les différentes voies de biosynthèse branchées au niveau du glucose-6-phosphate, du fructose-6-phosphate,etc.
- la synthèse des enzymes EcitZ, EcitB et EcitC, correspondant aux gènes *citZ*, *citB* et *citC*, est réprimée en présence d'arginine (voir la sous-figure *TCA-full* de l'annexe B.2.1). Or la partie tricarboxylique du cycle de Krebs de *Bacillus subtilis* est effectivement réprimée quand une source d'azote privilégiée comme la glutamine ou l'arginine est présente [85, 161, 162, 60] ;
- le flux d'acétate excrété est plus important en malate qu'en glucose (voir la figure *T-acetate* de l'annexe B.2.2), en accord avec une récente étude expérimentale [114].

Néanmoins, certaines configurations prédites ne sont pas en accord avec la physiologie connue (ou attendue) de *Bacillus subtilis* :

- l'enzyme EcarAB appartenant à la voie de synthèse de l'arginine n'est pas réprimée

par l'ajout d'arginine [146] (voir la sous-figure *Arginine synthesis* de l'annexe B.2.1) ;
- la première enzyme (EilvBH2) de la voie de synthèse de la valine et de la leucine est réprimée en absence de valine et leucine dans le milieu [135, 224] (voir la sous-figure *2-keto-isovalerate synthesis* de l'annexe B.2.1) ;
- la voie de synthèse de l'alpha-keto-butyrate (EilvA), un intermédiaire dans la voie de synthèse de l'isoleucine, est réprimée en absence d'isoleucine [136] (voir la sous-figure *AKB synthesis* de l'annexe B.2.1).

Nous cherchons donc à expliquer ces différences. Les deux premiers points s'expliquent par la présence d'isoenzymes dans le réseau métabolique ayant des poids en acides aminés différents. La répression de EilvA s'explique par l'existence d'une voie de synthèse d'alpha-keto-butyrate alternative moins coûteuse en acides aminés.

Le poids des isoenzymes dans la répartition des flux métaboliques

La première enzyme de la voie de synthèse de la valine et de la leucine (EilvBH2) correspond à une acétolactate synthase catalysant la réaction

$$2 \text{ pyruvate} \longrightarrow 2\text{-acétolactate} + CO_2$$

Cette réaction est non seulement impliquée dans la synthèse de la valine et de la leucine, mais aussi dans la synthèse de l'acétoine (voir la sous-figure *T-acetoin-syn* de l'annexe B.2.1). Chez *Bacillus subtilis*, l'acétolactate synthase est codée soit par les gènes *ilvB* et *ilvH*, soit par le gène *alsS*. Or l'enzyme IlvBH contient 748 acides aminés alors que AlsS n'en contient que 570. En toute logique, $P_f(\mu)$ privilégie de produire l'acétolactate par AlsS et non par IlvBH. Toutefois, les gènes *ilvB* et *ilvH* sont en opéron avec *ilvC*, *leuA*, *leuB*, *leuC* et *leuD* impliqués dans la voie de synthèse de la valine et de la leucine et régulés par la disponibilité en leucine, valine, et isoleucine à travers CodY et une T-box spécifique pour la leucine [79, 208]. Le gène *alsS* est en opéron avec *alsD* impliqué dans la synthèse d'acétoine et régulés *a priori* par l'acétate via AlsR [177]. Il semble donc acquis, au regard de ces articles, que les synthèses des enzymes AlsS et IlvBH sont régulées différement et ne peuvent pas en général se substituer l'une à l'autre. Les efficacités k_{E_i} de ces deux enzymes sont probablement différentes, contrairement à nos conditions de test. De plus, l'activité de cette enzyme est potentiellement régulée par la valine à travers un site de fixation spécifique. Cette enzyme est en effet très proche de celle d'Escherichia coli régulée spécifiquement par la valine [107]. La présence de ce site de fixation augmente probablement la longueur totale en acides aminés. En revanche, une régulation enzymatique par la valine permet d'adapter le flux métabolique à des variations de demande comme nous l'avons vu au chapitre 2, et donc de mieux contrôler la voie métabolique.

Le même raisonnement s'applique pour l'isoenzyme EcarAB. Cette enzyme est impliquée dans la voie de synthèse de l'arginine et catalyse la réaction suivante

$$\text{glutamine} + \text{ATP} + CO_2 \longrightarrow \text{ADP} + \text{P} + \text{glutamate} + \text{carbamoyl-P}$$

Cette réaction est également présente dans la voie de synthèse des pyrimidines (non représentée sur les figures en annexe, mais intégrée pour la résolution de problème d'optimisation). L'enzyme CarAB, codée par les gènes *carA* et *carB*, de la voie de synthèse de l'arginine contient 1383 acides aminés, alors que l'enzyme impliquée dans la voie de synthèse des pyrimidines et codée par les gènes *pyrAA* et *pyrAB* en contient 1435. Notre problème d'optimisation prédit l'utilisation systématique de l'enzyme EcarAB pour la synthèse des pyrimidines, même lors de l'ajout d'arginine dans le milieu. Or l'enzyme PyrAAB correspond à la première étape de la voie des synthèse des pyrimidines, dont la synthèse est régulée par le produit final de la voie métabolique [226, 129]. A ce titre, l'activité de cette enzyme est aussi probablement régulée par le produit final, ce qui expliquerait une longueur en acides aminés plus grande. Dans ce cas encore, les mécanismes de régulation génétique sont différents et empêchent la substitution des deux isoenzymes [146, 226].

Pour $P_f(\mu)$, le coût en acides aminés est donc déterminant non seulement pour le choix entre deux isoenzymes, mais aussi pour le choix entre différentes voies de biosynthèse pour un métabolite. Dans la section précédente, nous avons illustré le choix entre les transporteurs et les voies de biosynthèse classique pour un acide aminé. Ce coût en protéines s'exerce également sur le choix de synthèse des co-facteurs et des métabolites intermédiaires, comme l'alpha-keto-butyrate présenté ci-après.

Synthèse de l'alpha-keto-butyrate

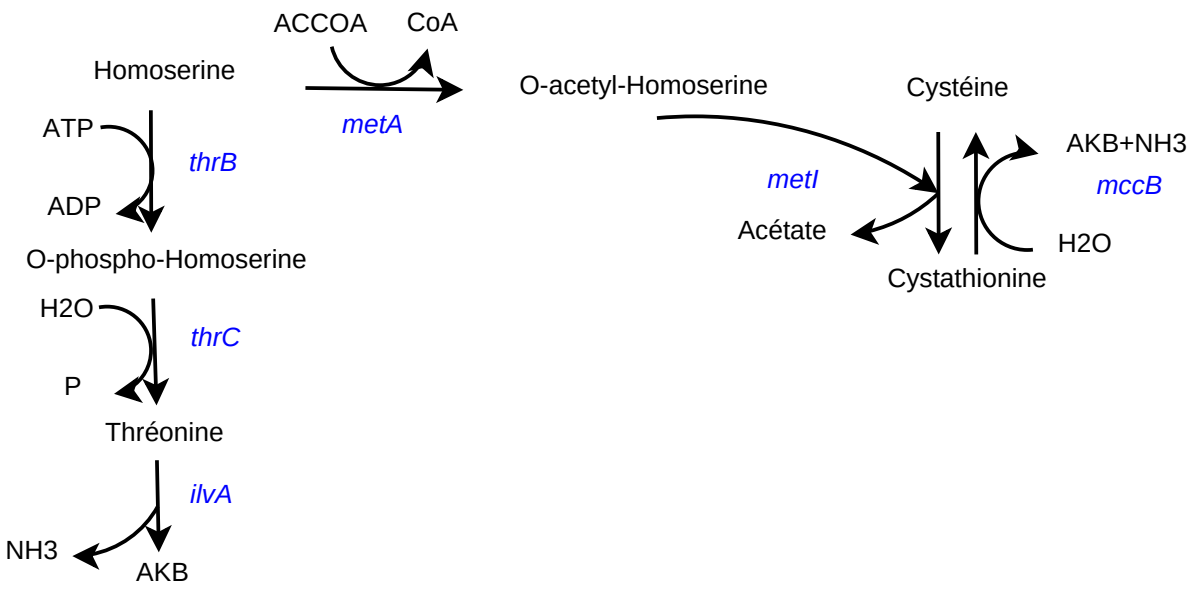


FIG. 5.9 – Deux voies alternatives de synthèse de l'AKB

L'alpha-keto-butyrate (AKB) est un métabolite intermédiaire dans la voie de l'isoleucine, issu de la dégradation de la thréonine par l'enzyme IlvA. Chez *Bacillus subtilis*, on

peut dénombrer deux voies métaboliques alternatives de synthèse de l'AKB , représentée sur la figure 5.9. La première voie métabolique est composée des enzymes ThrB, ThrC et IlvA, codées respectivement par les gènes *thrB*, *thrC* et *ilvA*. La réaction équivalente de cette voie métabolique s'écrit :

$$\text{homosérine} + \text{ATP} + \text{H}_2\text{O} \longrightarrow \text{ADP} + \text{P} + \text{NH}_3 + \text{AKB}$$

La deuxième voie métabolique est composée des enzymes MetA, MetI et MccB, codées respectivement par les gènes *metA*, *metI* et *mccB*. La réaction équivalente correspond alors à

$$\text{homosérine} + \text{ACCOA} + \text{H}_2\text{O} \longrightarrow \text{CoA} + \text{acétate} + \text{NH}_3 + \text{AKB}$$

Le coût en protéines d'induction de la voie 1 obtenu en additionant le nombre d'acides aminés pour ThrB, ThrC et IlvA s'élève à 1083 acides aminés alors que celui de la voie 2 est de 1054. Les co-facteurs impliqués dans ces deux voies alternatives sont différents. Dans un cas, un ATP est consommé, alors que dans l'autre cas, il s'agit d'un ACCOA. Or il existe une voie de synthèse dont les enzymes sont codées par les gènes *pta* et *ackA* [171, 206], dont la réaction réduite est équivalente à :

$$\text{ACCOA} + \text{P} + \text{ADP} \longrightarrow \text{ATP} + \text{acétate}$$

Non seulement la deuxième voie métabolique permet de produire un AKB, mais en consommant un ACCOA, une économie de 718 acides aminés est potentiellement réalisée au niveau des enzymes Pta et AckA (ayant respectivement 323 et 395 acides aminés). On obtient ainsi une troisième voie alternative à la synthèse de l'AKB dont la réaction réduite est équivalente à celle de la voie 1 mais dont le coût d'induction en acides aminés s'élève à 336 acides aminés (1054-718). Cette différence en coût d'induction explique le choix de cette voie de synthèse pour l'AKB par $P_f(\mu)$ plutôt que la traditionnelle voie 1. Certaines expériences biologiques suggèrent que ce mode élémentaire puisse dans certains cas être fonctionnel. En effet, une isoenzyme équivalente à MccB existe mais n'a pas été encore identifiée [97] (cf. figure 5.10).

5.4 Prédiction des modules élémentaires

La résolution de $P_f^{lp}(\mu)$ pour tous ces différents milieux nous permet ainsi de prédire les configurations du réseau métabolique pour différents milieux, et donc d'identifier des groupes d'enzymes induites dans les mêmes conditions. Ces groupes d'enzymes correspondent ainsi à des potentiels modules élémentaires que nous pouvons comparer aux modules identifiés au chapitre 3. La majorité des modules élémentaires ont été retrouvés, à l'exception des cas mentionnés précédemment. La méthode RBA prédit les configurations des régulations locales selon la définition du chapitre 2. En outre, nous prédisons l'existence de nouveaux modules potentiels dans les voies métaboliques de *Bacillus subtilis*.

Remarque 5.2 *Par abus de notation, on écrira par la suite que notre méthode prédit les régulations locales. Néanmoins, on rappelle bien que notre méthode ne prédit pas la réalisation biologique d'une régulation (facteur de transcription, ou T-box par exemple), mais bien les conséquences de celle-ci sur le régime établi du réseau métabolique.*

De nouveaux modules fonctionnels prédits ...

Il s'agit des modules suivants :
- la synthèse du 6-phosphogluconate : *zwf*;
- la synthèse de l'alpha-keto-glutarate : *citZ, citB, citC*;
- la synthèse d'aspartate : *pycA, aspB*;
- la synthèse d'asparagine : *asnB*;
- la synthèse d'alanine : *alaT*;
- la synthèse de sérine : *serA*, *serC*;
- la synthèse de thréonine et la synthèse de la glycine : *thrB*, *thrC*, *tdh*, *kbl*, *glyA*;
- la synthèse de l'histidine : *hisG, hisI, hisA, hisF, hisH, hisC, hisB, hisJ, hisD*;
- la synthèse du chorismate : *aroA, aroB, aroC, aroD, aroK, aroE, aroF*;
- la synthèse du phénylalanine : *pheA, hisC*;
- la synthèse de la tyrosine : *tyrA, hisC*;
- la dégradation de l'alanine : *ald*;
- la dégradation de l'asparagine seul : *ansA*;
- la dégradation de l'aspartate seul : *ansB*;
- la dégradation du glutamate seul : *rocG*;
- la dégradation de la sérine : *sdaAA*, *sdaAB*.

Ces modules correspondent majoritairement à des voies courtes de synthèse composées d'une ou de deux enzymes, à l'exception de la synthèse de la thréonine, de l'histidine et du chorismate. Chez *E. coli*, des modules fonctionnels existent pour les synthèses de la thréonine, de l'histidine et du chorismate [107]. La transcription des gènes impliqués dans les synthèses de thréonine et d'histidine est en effet régulée par des peptides leaders riches respectivement en thréonine et histidine [152]. Trois isoenzymes catalysent la première étape de la synthèse du chorismate. La transcription de chaque enzyme est réprimée par un des acides aminés aromatiques à travers les facteurs de transcription TyrR et TrpR [152]. Chez *B. subtilis*, les régulations génétiques des voies de synthèse de l'histidine, de la thréonine et du chorismate n'ont pas été étudiées, expliquant ainsi l'absence de modules connus. Cependant, comme pour *B. subtilis*, les voies de synthèse courtes impliquant uniquement une ou deux enzymes chez *E. coli* ne semblent pas régulées au niveau génétique [100, 169, 20].

Les seules vraies différences avec les modules identifiés dans le chapitre 3 concernent la répression de l'opéron *citZCH* par CcpC dans le cycle de Krebs (module TCA-citrate du tableau 3.13) et la dégradation de l'aspartate et de l'asparagine régulée par AnsR (module Aspx-deg du tableau 3.15). Le citrate semble être le signal métabolique mesuré par CcpC [110]. Cependant, l'opéron *citZCH* semble également régulé par un autre mécanisme [104], ce qui remettrait en cause le module défini. La dégradation de l'aspartate est induite en présence d'asparagine par AnsR [218, 63]. Or nos prédictions suggèrent

qu'elle pourrait également être induite par l'aspartate.

Présence effective d'une régulation locale : vers une explication

Les modules élémentaires identifiés au chapitre 3 sont ainsi correctement prédits par notre problème d'optimisation. Les modules élémentaires étant définis par une régulation génétique locale, nous pouvons en déduire que $P_f(\mu)$ permet donc bien de retrouver les configurations associées à ce type de régulation dans les voies métaboliques, et ce quel que soit le type de régulateur (facteur de transcription, riboswitch,etc.). En revance, certaines régulations locales prédites pour les voies métaboliques courtes n'existent probablement pas en réalité. Dans notre méthode, nous n'avons pas intégré le coût de la régulation génétique elle-même. Or un facteur de transcription ou un peptide leader sont aussi composés d'acides aminés. Pour les voies de synthèse courtes comme les synthèse d'aspartate ou d'alanine, le coût associé à l'implémentation de la régulation pourrait être supérieur au gain en acides aminés apporté par la répression de l'enzyme. La présence effective d'une régulation locale pourrait bien résulter d'un compromis entre le coût d'induction de la voie métabolique et le coût de l'implémentation d'un régulateur, tout en notant bien que le coût d'induction de la voie métabolique dépend aussi directement de l'efficacité des enzymes.
Un deuxième point concerne l'utilisation privilégiée de concentrations métaboliques élevées dans l'adaptation des voies métaboliques, comme nous l'avons montré au cours de la section 3.3.3 page 102. Des concentrations élevées de métabolites judicieusement choisis, comme les premiers métabolites d'une longue voie de synthèse, permettent de diminuer la quantité totale de protéines investies dans la voie métabolique. Or les voies de synthèse courtes non régulées correspondent justement aux voies de synthèse de ces premiers métabolites comme l'aspartate ou le pyruvate. L'intérêt serait alors de produire ces métabolites le plus possible dans tous les cas afin d'économiser des protéines sur les voies métaboliques situées en aval de ces métabolites clés.

5.5 Transport des ressources extracellulaires

Si la prédiction des régulations locales est correcte, la prédiction des régulations globales n'est en revanche pas évidente. Or, comme nous l'avons souligné dans les chapitres 2 et 3, les régulations globales dans les voies métaboliques correspondent à des stratégies «haut niveau» soit pour gérer les transitions entre la phase de croissance et la phase stationnaire ou le stress, soit pour gérer les priorités au niveau des ressources extracellulaires. Compte tenu du cadre de notre méthode (en phase exponentielle de croissance sur les voies métaboliques), nous allons nous centrer sur la prédiction des stratégies haut niveau vis-à-vis du transport des ressources extérieures, et ce pour trois types de stratégies de transport bien connues :

- la substitution de transporteurs haute et basse affinité en fonction de la concentration extracellulaire du nutriment ;
- la hiérarchie des sources de carbone connue sous le nom de «répression catabolique» ;
- la hiérarchie des sources d'azote.

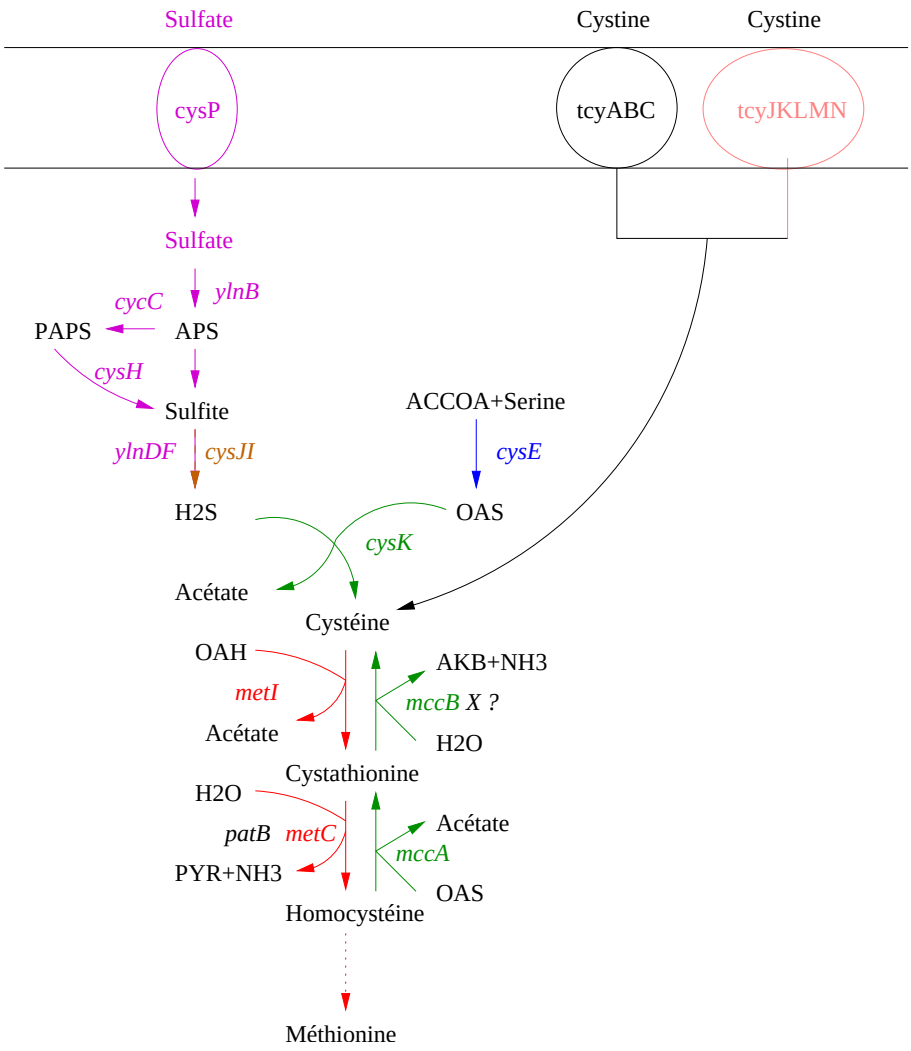


FIG. 5.10 – La voie de synthèse de la cystéine

5.5.1 Substitution de transporteurs haute et basse affinité

Prenons l'exemple de la voie de synthèse de la cystéine de la figure 5.10, avec 2 transporteurs de cystine : le transporteur TcyABC basse affinité composé de 749 acides aminés, codé par les gènes *tcyA*, *tcyB* et *tcyC*, et le transporteur TcyJKLMN haute affinité composé de 1272 acides aminés codé par les gènes *tcyJ*, *tcyK*, *tcyL*, *tcyM* et *tcyN*. La cystéine n'existe en effet que sous sa forme réduite, la cystine, dans le milieu extracellulaire. Dans un environnement oxydant comme le sol, un pont dissulfure se forme entre les groupements thild (-SH) libres de deux cystéines, formant la cystine. La cystine importée est ensuite convertie en deux cystéines dans la cellule. Ces deux transporteurs sont des transporteurs dit actifs, consommant de l'énergie sous forme d'ATP pour importer leur substrat selon la réaction :

$$S_{ext} + ATP \longrightarrow S_{int} + ADP + P$$

Le transporteur haute affinité se caractérise de plus par une affinité très faible, permettant ainsi le transport actif de faibles concentrations de substrat. Nous allons traduire cette affinité différente entre les transporteurs en faisant varier la constante de Michaëlis-Menten K_{m_i} et le turnover k_{max} de l'équation (5.1) décrivant l'activité du transporteur

$$k_E = k_{max}\frac{S_i}{K_{m_i} + S_i}.$$

Dans le cas du transporteur basse affinité TcyABC, $K_{m_i} = 0.8$mM et $k_{max} = 20s^{-1}$ alors que $K_{m_i} = 0.008$mM et $k_{max} = 10s^{-1}$ pour le transporteur haute affinité TcyJKLMN.

Remarque 5.3 *Contrairement à l'exemple proposé, un transporteur faible affinité chez les bactéries est généralement de type symport et coûte un H^+ de la façon suivante :*

$$S_{ext} + H^+_{ext} \longrightarrow S_{int} + H^+_{int}$$

En considérant que

$$ADP + P + 4H^+_{ext} \longrightarrow ATP + 4H^+_{int},$$

il vient que le transport faible affinité de substrat coûte également 1/4 d'ATP.

Dans ces conditions, nous obtenons pour une gamme de cystine allant de 0.05mM à 0.6mM la substitution entre les transporteurs TcyABC et TcyJKLMN donnée sur la figure 5.11 avec un milieu contenant du glucose, de l'azote assimilable, du sulfate, du phosphate et de l'oxygène. Ces prédictions sont en accord avec la littérature [32]. Ainsi, les contraintes et le formalisme mathématique choisis permettent de prédire la substitution de transporteurs de faible/haute affinité. Sur la figure 5.11, la zone de basculement calculée est progressive. $P_f^{lp}(\mu)$ a été résolu par dichotomie avec un seuil de test d'égalité entre la borne inférieure et supérieure pour μ égal à 1e-3. Si ce seuil est diminué à 1e-10, la zone de basculement rétrécit (voir la figure 5.12). Une prédiction fine pour la valeur du seuil de basculement entre les transporteurs demanderait néanmoins des expériences biologiques complémentaires pour identifier précisément les paramètres K_m et k_{max} pour ces transporteurs.

5.5.2 Prédiction de la hiérarchie des sources de carbone

Chez les bactéries, l'utilisation des sources de carbone secondaires (transporteurs et enzymes), est réprimée par la présence de glucose. Ce phénomène biologique est appelé «*répression catabolique*». Chez *Bacillus subtilis*, toutes les sources de carbone à l'exception du malate et du saccharose sont connus pour être réprimées par le glucose [75].
Pour tester si $P_f(\mu)$ permet de prédire cette hiérarchie des sources carbonées, nous avons résolu $P_f^{lp}(\mu)$ avec 10mM de glucose ajouté à 10mM d'une autre source carbonée parmi les 22 introduites dans la section 5.3.2, avec pour chaque transporteur $K_m = 0.8$mM et $k_{max} = 20s^{-1}$. Nous obtenons les prédictions suivantes :

- répression des sources carbonées suivantes : gluconate, ribose, arabinose, inositol, glycérol, N-acétylglucosamine et mannitol ;

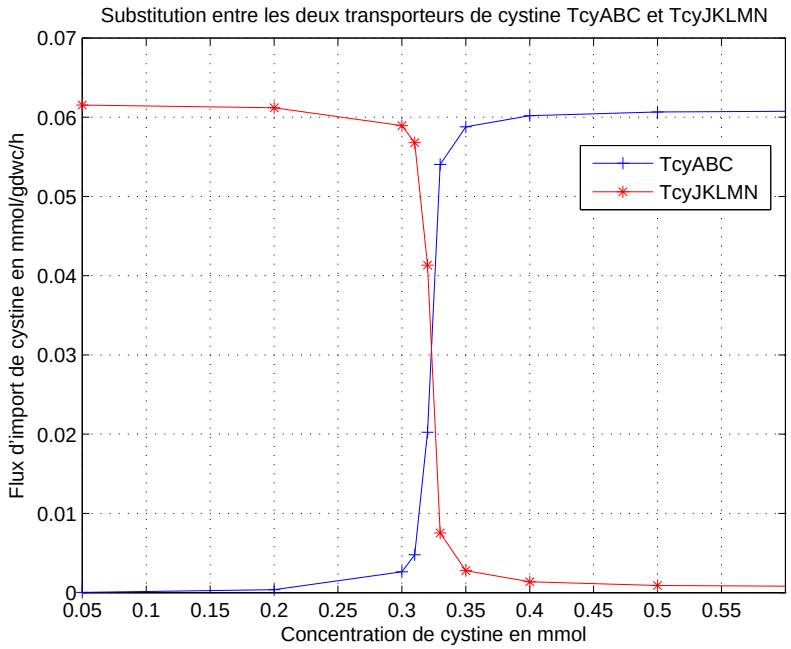


FIG. 5.11 – Substitution entre les deux transporteurs haute (TcyJKLMN) et basse affinité (TcyABC) de cystine calculée avec un seuil de 10^{-3} pour la dichotomie

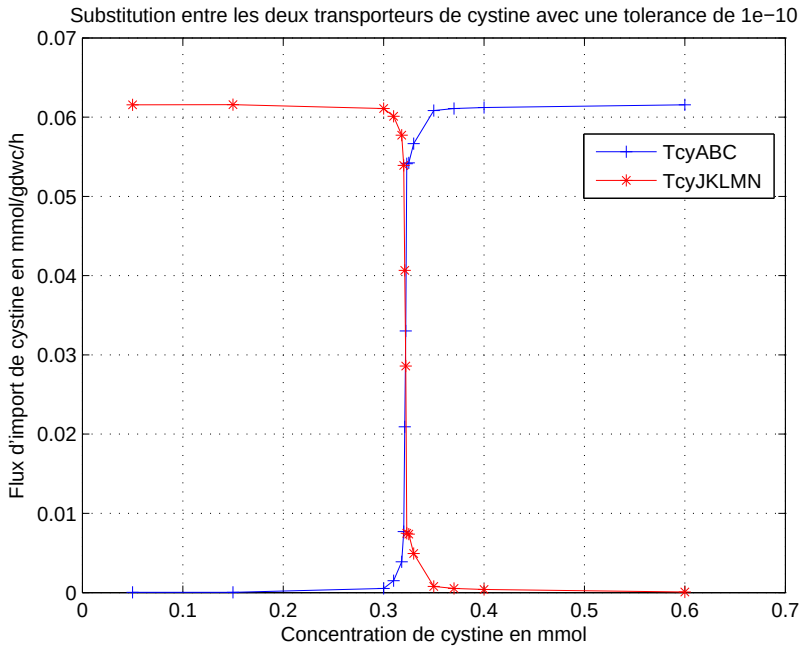


FIG. 5.12 – Substitution entre les deux transporteurs haute (TcyJKLMN) et basse affinité (TcyABC) de cystine calculée avec un seuil de 10^{-10} pour la dichotomie

- utilisation conjointe avec le glucose pour : fructose, mannose, saccharose, malate, succinate, fumarate, citrate, 2-oxoglutarate et pyruvate ;
- répression de l'utilisation du glucose pour les disaccharides suivant : tréhalose, lichenan, cellobiose et maltose.

A première vue, nos résultats ne sont pas (ou peu) en accord avec la hiérarchie connue des sources de carbone. Or le choix d'utilisation des transporteurs est dépend de leur efficacité comme l'étude de la substitution de transporteurs haute et basse affinité l'a montrée. La prédiction de la hiérarchie des sources de carbone a été réalisée avec des efficacités identiques pour les transporteurs. Or on peut supposer que l'efficacité de transport d'un disaccharide est plus faible qu'un monosaccharide vu que la molécule à transporter est deux fois plus grosse. Un disaccharide est en effet composé de 12 carbones, comparé à 6 carbones pour un monosaccharide.
Pour une efficacité deux fois plus faible $k_{max} = 10s^{-1}$, nous prédisons maintenant la répression de l'utilisation du tréhalose, du lichenan, du cellobiose et du maltose par le glucose, et que le saccharose est toujours utilisé conjointement avec le glucose. Ces prédictions sont en accord avec les régulations génétiques connues pour les voies de transport et d'assimilation de ces sources carbonées [75]. En revanche, l'utilisation conjointe du glucose d'une part avec le fructose et le mannose, et d'autre part avec le succinate, le fumarate et le citrate ne semble pas *a priori* en accord avec la littérature. Nos résultats indiquent que les sources de carbone néoglucogéniques (malate, succinate, fumarate, citrate et 2-oxoglutarate) sont utilisées préférentiellement pour la synthèse de l'aspartate et du glutamate (et leurs dérivés). Certaines expériences biologiques récentes vont dans le sens de ces prédictions [192, 114]. De même, lors d'un ajout de glucose dans un milieu contenant initialement du malate, le transporteur de fructose (et potentiellement de mannose) FruA est induit, ce qui suggère une utilisation conjointe du fructose (et du mannose) et du glucose. Cette expérience a été conduite dans le cadre du projet BaSysBio.
Grâce à la méthode RBA, nous pouvons conclure de cette étude que la répression catabolique des sources carbonés pour un organisme dépend à la fois du coût de l'induction de la voie d'assimilation (transporteur et enzymes) et de l'efficacité du transporteur. Notre méthode permettrait là encore de prédire la hiérarchie des sources carbonées sous l'hypothèse que les paramètres cinétiques des transporteurs sont identifiés expérimentalement.

5.5.3 Prédiction de la hiérarchie des sources d'azote

Comme pour les sources de carbone, il existe des sources d'azote privilégiées pour les bactéries. Dans le cas de *Bacillus subtilis*, les sources d'azote privilégiées correspondent par ordre d'importance (taux de croissance maximal) à la glutamine, l'arginine et de l'ammonium (NH3) [60, 239]. Viennent ensuite toujours par ordre d'importance la proline, l'aspartate, le glutamate et l'urée.
Afin de prédire cette hiérarchie des sources azotées, nous avons résolu $P_f^{lp}(\mu)$ avec 10mM d'azote ajoutés à 10mM d'une autre source azotée, avec pour chaque transporteur $K_m = 0.8$mM et $k_{max} = 20s^{-1}$. Nous obtenons les résultats suivant :

– répression du transport de NH3 pour les sources azotées suivantes : arginine, glutamine et histidine ;
– répression partielle du transport de NH3 pour la proline, l'asparagine et l'aspartate.

La répression partielle du transport de NH3 pour l'asparagine ou l'aspartate s'explique aisément car (*i*) l'aspartate est utilisé comme précurseur par de nombreuses voies métaboliques ; (*ii*) l'asparagine se dégrade en aspartate. L'aspartate est ici utilisé en tant que précurseur métabolique et non comme source d'azote.
$P_f(\mu)$ permet bien de prédire que (*i*) l'arginine et la glutamine sont des sources d'azote privilégiées ; (*ii*) la proline et l'aspartate sont de bonnes sources d'azote intermédiaires. En revanche, chez *Bacillus subtilis*, l'histidine est une faible source d'azote [60]. De plus, sa voie de dégradation est réprimée par CcpA et CodY [242, 61]. Donc en conditions glycolytiques et/ou de forte croissance, l'histidine extracellulaire ne peut être utilisée. A première vue, ce phénotype est surprenant car l'histidine est dégradée en glutamate et NH3 (par l'intermédiaire de formamide). Comme la méthode RBA le prédit, l'histidine devrait ainsi avoir le même statut que l'arginine en tant que source d'azote, et devrait donc être dégradée en conditions glycolytiques.

5.5.4 Vers la prédiction des régulations globales

L'environnement naturel de *Bacillus subtilis* pourrait expliquer cette différence entre la prédiction et le phénotype réel. *Bacillus subtilis* vit en général dans le sol, près des racines ou sur les feuilles des plantes. Par ailleurs, les plantes sont connues pour excréter de la glutamine, de l'aspartate, de la glycine et de la sérine [82, 155]. La présence d'histidine dans l'habitat naturel de cette bactérie semble donc bien occasionnelle, ce qui expliquerait la répression de la voie de dégradation. Notre problème d'optimisation n'intègre en effet aucune référence temporelle. Les prédictions sont obtenues sous l'hypothèse que les concentrations extracellulaires des nutriments dans le milieu restent constantes et permettent ainsi la maintenance du même régime stationnaire pendant un moment (voir la section 4.2.3 du chapitre précédent). La bactérie a ainsi très bien pu empêcher la dégradation de l'histidine en conditions glycolytiques si le coût associé à l'induction de la voie est supérieur au gain apporté par la dégradation de l'histidine sur un long terme. En effet, induire une telle voie pour un court moment si l'histidine s'épuise, va permettre de gagner du taux de croissance à court terme. Mais à plus long terme, les enzymes, si elles ne sont pas dégradées, vont occuper de la place dans la cellule et avoir un effet inverse sur le taux de croissance (contrainte de densité). L'hypothèse de disponibilité de la ressource se révèle critique pour l'émergence des régulations globales. Il sera nécessaire par la suite d'intégrer les aspects temporels et l'évolution du milieu extracellulaire en fonction du temps.

5.6 Conclusion

Les prédictions obtenues dans ce chapitre sont prometteuses et qualitativement correctes lorsqu'elles sont comparées aux données de la littérature. Les contraintes structu-

relles inclues dans le problème d'optimisation convexe sont ainsi globalement validées. L'obtention de prédictions quantitatives précises pour les concentrations de protéines nécessite en revanche l'identification des paramètres k_{E_i} pour chaque enzyme. Nous avons d'ailleurs souligné à plusieurs reprises dans ce chapitre l'impact critique de ce paramètre dans le cas de l'étude de la répression catabolique en particulier. Une façon d'identifier ces paramètres serait de disposer de mesures de concentrations absolues de protéines et de flux métaboliques pour différents taux de croissance. Les technologies associées à la mesure de concentration absolue de protéines sont actuellement en cours de développement et devraient aboutir dans les années à venir à des données précises en haut débit et à l'échelle de la cellule. Nous devrions ainsi disposer des données nécessaires à la calibration globale du modèle de *Bacillus subtilis* pour la méthode RBA d'ici peu de temps.

Au delà du côté purement prédictif, la méthode RBA constitue un cadre formel pour une meilleure compréhension du fonctionnement des bactéries et pour appréhender des questions biologiques fondamentales diverses. Dans ce chapitre, nous avons notamment montré que l'émergence des régulations locales dans les voies métaboliques résulte directement du compromis sur les protéines entre les ribosomes et les voies métaboliques. La présence effective d'une régulation locale obéit à des contraintes supplémentaires que nous n'avons pas intégrées dans notre méthode à ce jour. Les régulations globales vis-à-vis de la gestion des ressources extracellulaires reflètent des stratégies d'adaptation sur un plus long terme et requièrent une bonne connaissance de l'environnement de vie des bactéries. L'hypothèse sur la disponibilité infinie de la ressource extracellulaire pourrait être une limitation de la méthode RBA pour la prédiction des régulations globales. L'intégration des aspects temporels dans notre problème d'optimisation initial et notamment la diminution de la ressource au cours du temps permettront probablement d'améliorer les prédictions des structures de régulation globale. La présence des régulations globales dans le réseau métabolique résulterait aussi d'un compromis entre la fréquence du stimuli extérieur et le coût d'induction des voies métaboliques sur un long terme.

Conclusion générale

Au cours de cette thèse fortement ancrée dans la discipline émergente qu'est la Biologie des Systèmes, nous nous sommes attachés à considérer la bactérie en tant que système afin de dégager des propriétés génériques sur le contrôle des voies métaboliques. L'intérêt ici est double. Il s'agit d'une part de dégager des principes de fonctionnement généraux applicables pour d'autres organismes bactériens ou non, et d'autre part d'évaluer la pertinence des méthodes et outils développés en Automatique pour les systèmes biologiques.

Au cours de la première partie, nous avons proposé un découpage du réseau métabolique en modules bien définis du point de vue entrée/sortie de part leur propriétés systémiques grâce à la prise en compte de tous les niveaux de régulation des voies métaboliques (transcriptionnel, traductionnel, post-traductionnel et enzymatique). Le cadre d'études proposé ici n'est pas spécifique à un organisme. Nous nous sommes certes centrés sur *Bacillus subtilis* pour des raisons pratiques, mais une rapide étude comparative nous a montré que nos résultats s'appliquent sans difficulté pour d'autres procaryotes comme *Escherichia coli*. La comparaison des régulations des voies métaboliques entre ces deux organismes est d'ailleurs édifiante. La structure de contrôle de certaines voies métaboliques comme la voie de synthèse des purines est très conservée même entre ces deux organismes ayant divergé il y a plus de trois milliards d'années [14]. On peut imaginer que ces voies métaboliques sont soumises aux mêmes objectifs de contrôle et subissent les mêmes contraintes. D'autres voies métaboliques comme la glycolyse sont par contre régulées de manière très différente. Ici les objectifs de contrôle sont certainement différents. Si la structure de la régulation peut être très conservée, les mécanismes moléculaires réalisant ces régulations varient énormément entre ces organismes [107, 75]. Ainsi les propriétés fondamentales associées aux modules démontrées dans cette thèse sont valides pour ces deux organismes et certainement pour une grande diversité d'organismes partageant les mêmes spécificités.
Les modules possèdent des propriétés singulières, tant au niveau mathématique qu'au niveau biologique. En effet, la première enzyme irréversible joue un rôle clé dans la définition du point d'équilibre puisque seules les caractérisques de cette enzyme et de son contrôle génétique déterminent le régime d'équilibre. De plus, cette enzyme correspond généralement à la première enzyme du module. Ainsi la première conséquence de l'existence des modules consiste à réduire fortement la complexité du réseau métabolique. D'un point de vue fonctionnel, l'irréversibilité des enzymes est critique pour l'adaptation du régime d'équilibre en réponse à des demandes de flux. Les enzymes irréversibles découpent le réseau métabolique en empêchant la propagation d'information

de l'aval vers l'amont des voies métaboliques. Une régulation enzymatique (ou génétique par l'intermédiaire d'un effecteur métabolique) est généralement nécessaire pour assurer la transmission de l'information. Un deuxième point essentiel de notre étude concerne le découplage du régime d'équilibre des modules vis-à-vis des co-métabolites. Dès lors que les enzymes du module ne saturent pas, les métabolites dit hubs (comme le glutamate) intervenant souvent dans des étapes intermédiaires de plusieurs modules différents ne modifient pas la caractéristique entrée/sortie du module. Leurs variations n'induisent pas un réajustement global de tous les régimes d'équilibre entrée/sortie des modules. La connectivité des voies métaboliques entre elles est donc également réduite et seuls certains métabolites spécifiques semblent jouer un rôle dans l'adaptation globale de la cellule. Ces différentes propriétés théoriques fondamentales nous ont amenés à construire un cadre formel d'étude des voies métaboliques centré sur les propriétés des modules et leur interconnexion en régime établi. Ce cadre formel permet d'étudier l'existence d'un équilibre pour une interconnexion de modules donnée, ainsi que l'évolution qualitative de tous ses composants vis-à-vis de perturbations en entrée/sortie de l'interconnexion. Ce cadre s'est également révélé effectif pour analyser des données expérimentales. Ceci implique que les hypothèses retenues d'une part lors de l'élaboration du modèle décrivant les composants du module, et d'autre part lors de la mise à jour des propriétés théoriques des modules sont licites et validées.

Le réseau métabolique se décompose ainsi en modules fortement découplés à l'équilibre, ce qui intuitivement semble être une bonne propriété en terme de contrôlabilité et de robustesse si on effectue un parallèle avec les systèmes contrôlés en Ingénierie. De façon empirique, les ingénieurs vont préférer un correcteur décentralisé structuré pour un système multivariable de grande dimension plutôt qu'un correcteur multivariable classique plein où toutes les mesures contribueraient aux signaux de commande. Le mélange d'information rend l'identification des causes d'un dysfonctionnement difficile, le système est moins bien maîtrisé par les hommes. Un correcteur modulaire structuré permet en Ingénierie de se prévenir des évènements «hasardeux», et donc de robustifier le système à commander de grande dimension vis-à-vis d'incertitudes ou de pannes. De plus, les systèmes construits par les hommes doivent être opérationnels immédiatement, contrairement à l'évolution qui a éliminé les solutions inadaptées. On imagine mal une sélection réelle sur le contrôle d'une centrale nucléaire ! Si les hommes ont construit des régulateurs structurés car ils n'avaient pas le choix compte tenu des contraintes (prévention des pannes et des évènements hasardeux, délais de mise en service des systèmes, etc.), la sélection naturelle a de façon surprenante fait émerger aussi une solution modulaire à son problème de contrôle. En poursuivant le parallèle entre Ingénierie et Biologie, on est alors tenté de conclure que la réalisation modulaire des réseaux de régulation obtenue par l'évolution est le résultat d'une contrainte de robustesse comme le disait Kitano dans [111]. Or les résultats obtenus dans la deuxième partie de cette thèse indiquent en première analyse que la répartition parcimonieuse entre les processus cellulaires d'une ressource commune, les protéines, explique l'émergence d'une structure modulaire de contrôle.

Nous avons identifié des contraintes structurelles agissant sur la bactérie à tout instant. L'analyse de ces contraintes pour la phase exponentielle de croissance nous a conduit à formaliser des contraintes sur la gestion d'une ressource critique, les protéines. Satisfaire cet ensemble de contraintes pour un taux de croissance donné signifie alors qu'il existe une répartition en protéines entre les différents sous-systèmes cellulaires. Le cadre formel obtenu, appelé Resource Balance Analysis, est un problème d'optimisation convexe non différentiable. A travers la résolution d'un problème de programmation linéaire équivalent, nous obtenons pour la première fois une méthode permettant de prédire non seulement les flux métaboliques, mais aussi les concentrations des ribosomes et de toutes les protéines impliquées dans le réseau métabolique. L'analyse du critère a également dégagé deux propriétés fondamentales, en plus de la convexité. Tout d'abord, nous avons montré numériquement pour le réseau métabolique complet, et analytiquement sur un exemple simplifié, que l'émergence des modules dans les voies métaboliques provient d'une politique générale d'économie en protéines chez la bactérie pour gagner du taux de croissance. La modularité provient alors de la gestion de la ressource et non pas comme une contrainte de robustesse. Du fait de la limitation en protéines totales imposée par la densité intracellulaire, la bactérie doit éteindre certains sous-systèmes afin d'économiser des protéines pour ensuite les répartir sur les sous-systèmes encore actifs. Cette politique d'économie va encore plus loin. Nous avons aussi montré que l'optimisation de chaque sous-système individuel, afin de diminuer la quantité de protéines nécessaire pour achever sa tâche, contribue à l'optimum global à savoir l'augmentation du taux de croissance. Il vient que toute protéine non dédiée à l'augmentation du taux de croissance pénalise la bactérie en terme de croissance pure. La stratégie optimale pour la croissance serait d'éteindre toutes les protéines inutiles pour la croissance pour ne conserver que les protéines des voies métaboliques, les ribosomes, et plus généralement uniquement les processus cellulaires nécessaires à la duplication de la bactérie.
Or le problème d'optimisation global que résout «en temps réel» la bactérie est plus compliqué que celui proposé, car non seulement la bactérie doit se diviser lorsque les conditions le permettent, mais elle doit également survivre. La bactérie doit s'adapter et gérer les transitions entre ressource abondante et limitée, mais aussi se prémunir de différents stress ou d'antibiotiques à tout moment. Ces protéines coûtent du taux de croissance à court terme, mais sur un plus long terme elles donnent un avantage compétitif à la bactérie. On peut voir ici un compromis s'apparentant au compromis performance/robustesse en Ingénierie où la bactérie doit pouvoir croître le plus rapidement possible tout en étant capable de survivre en cas de problèmes. De même, la modularité dans les voies métaboliques ayant été conservée par l'évolution, on peut imaginer qu'elle permet non seulement de gagner en performance, mais aussi en robustesse sans quoi une autre politique de contrôle aurait certainement émergée. Néanmoins, même s'il est incomplet et simplifié, le cadre théorique de la méthode RBA propose un rationnel cohérent sur le fonctionnement de la bactérie en phase exponentielle unique à ce jour.

Ces travaux de thèse ouvrent de nombreuses perspectives tant sur le plan théorique que sur le plan expérimental. Concernant la première partie sur la définition des modules, nous nous sommes intéressés uniquement aux conditions d'existence du régime

d'équilibre. Une extension à cette première analyse théorique serait d'étudier la stabilité et la dynamique des modules, en regardant par exemple :

- la stabilité exponentielle locale autour d'un régime d'équilibre pour un module spécifique, et ensuite pour les différents types de connexion de modules considérés ;
- la stabilité du système non linéaire.

Comme pour l'existence du point d'équilibre, il conviendrait de définir un cadre générique pour analyser la stabilité des modules et de leur interconnexion. Néanmoins, comme nous l'avons souligné au chapitre 1, l'étude de la stabilité même dans le cas de voies métaboliques très simplifiées est un problème difficile du fait qu'il n'existe pas d'outils adaptés applicables à des systèmes non linéaires d'une telle dimension [227, 9, 6, 7]. Une approche serait déjà de caractériser la stabilité locale pour les deux structures de contrôle et d'évaluer le poids des différentes configurations biologiques comme l'irréversibilité des enzymes sur la stabilité du module afin de dégager des propriétés générales si elles existent.
La dynamique d'un module serait étudiée en prenant explicitement en compte les échelles de temps différentes entre la réponse du réseau métabolique (rapide de l'ordre de la milliseconde) par rapport à celle du réseau génétique (plus lente de l'ordre de la dizaine de minutes [184, 56]). Notamment les résultats théoriques associées aux perturbations singulières pourraient être appliqués aux systèmes biologiques afin de découpler le réseau génétique du réseau métabolique [121].

Les perspectives principales associées à la méthode RBA consistent tout d'abord à améliorer le pouvoir de prédiction de la méthode en affinant l'identification des paramètres avec des données expérimentales. L'identification des paramètres devrait être possible à court terme grâce à un autre projet européen, BaSynTech, dans lequel nous sommes impliqués. L'un des objectifs explicite de ce projet européen consiste à valider notre méthode RBA en produisant notamment des données de protéines quantitatives. Ces données nous permettront de mieux calibrer les différents paramètres intégrés dans la méthode. Un deuxième point concerne la validation globale du principe du compromis sur les protéines en perturbant artificiellement leur répartition. Nous avons déjà présenté au cours du chapitre 4 une expérience biologique où l'on explorait l'impact d'un ribosome plus gros sur le taux de croissance. L'idée est de perturber également les ensembles $\mathcal{P}_G$ et $\mathcal{P}_M$ pour différents taux de croissance. Ces expériences biologiques font partie du travail d'une thèse de Biologie des Systèmes actuellement en cours.
D'un point de vue biologique, la structure particulière générique du problème d'optimisation de notre méthode permet d'appréhender de nombreuses questions fondamentales comme la composition en acides aminés des enzymes, l'usage des codons, ou l'évolution des réseaux de régulation des organismes par rapport à leur niche écologique. De même, la méthode RBA définit un cadre formel permettant la modification rationnelle des bactéries. Les perspectives d'application de notre méthode dans la nouvelle discipline émergente qu'est la Biologie Synthétique pour la production de métabolites d'intérêt par exemple sont nombreuses.

D'un point de vue théorique, une extension immédiate serait d'intégrer les aspects tem-

porels dans la méthode RBA, que l'on appellerait la méthode RBA dynamique (dRBA). Notamment on intégrerait la diminution des concentrations extracellulaires en fonction de temps. Cette méthode nous permettrait hypothétiquement de révéler des stratégies sur le long terme vis-à-vis de la ressource. Si une protéine est produite pour rien, elle occupe de la place dans la cellule jusqu'à ce que par effet de dilution, sa concentration redevienne nulle. L'intégration des aspects temporels inclut intrinsèquement cet effet «mémoire». Une deuxième extension dans le prolongement de la première, consiste à intégrer l'évolution de tous les métabolites dans le problème d'optimisation. Il s'agit d'intégrer les concentrations des métabolites dans la description des flux métaboliques. Ce problème est difficile car même en considérant que les métabolites jouent linéairement sur le flux, on obtient un ensemble de contraintes qui, sans plus de précisions, sont bilinéaires.

Au regard de l'ensemble de ces résultats et perspectives, il apparaît que la vision systémique intrinsèque à l'Automatique est pertinente pour appréhender les systèmes biologiques, de même que les concepts fondamentaux associés comme les modules (considérés en tant que sous-systèmes autonomes). Il est clair que les deux disciplines sont complémentaires et permettront dans l'avenir des avancées majeures en Biologie, du moins peut-on l'espérer...

Annexe A

Bases de Biologie

A.1 Les différents modes de vie de *Bacillus subtilis*

Une bactérie comme *Bacillus subtilis* est soumise à différents modes de vie. On distinguera tout d'abord sur la figure A.1 les deux premiers grands modes de vie, à savoir la phase de croissance et la phase stationnaire, avec une troisième phase dite de transition entre les deux.
Lorsque tous les nutriments nécessaires à la croissance sont réunis, la bactérie grossit (son volume augmente) et se divise. C'est la phase exponentielle de croissance. Dès qu'un (ou plusieurs) nutriment(s) devient (deviennent) limitant, la bactérie entre en phase de croissance transitoire, où la vitesse de division est ralentie, et ce jusqu'à épuisement de la ressource. La bactérie entre alors en phase stationnaire où des modes de vie radicalement différents peuvent avoir lieu comme la sporulation/cannibalisme ou la mort. D'autres modes de vies peuvent avoir lieu pendant ces trois phases : la compétence, la formation de biofilm ou la mobilité. La sporulation traduit la capacité pour la bactérie à empaqueter son ADN en une spore extrêmement résistante. Dans cette phase, une partie de la population bactérienne se suicide pour fournir à l'autre partie les ressources nécessaires à la sporulation (cannibalisme). La compétence permet d'assimiler de l'ADN extracellulaire et de l'insérer par recombinaison dans le chromosome. La mobilité se traduit par la construction d'un flagelle qui permet à la bactérie de se déplacer vers un endroit plus favorable.

La variété des modes de vie est complètement dépendant de la bactérie considérée. Certains sont mutuellement exclusifs (sporulation et mobilité), d'autres non (compétence et mobilité). Néanmoins, le fait même de pouvoir les distinguer traduit une organisation modulaire des processus impliqués, module qui peut être «éteint» (pas de flagelle) ou «allumé» (présence du flagelle).

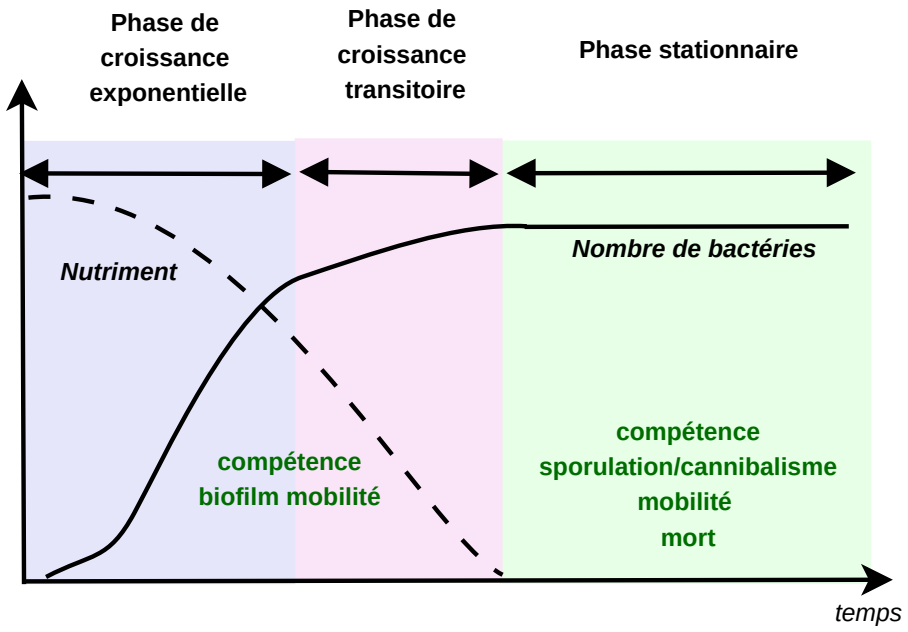


FIG. A.1 – Les différents modes de vie de *Bacillus subtilis* au cours du temps

A.2 Les processus cellulaires nécessaires à la croissance exponentielle

La cellule est composée de différents processus biologiques élémentaires coordonnés assurant sa croissance et sa survie (cf. la figure A.2). Chaque processus effectue une tâche «élémentaire» que nous allons détailler rapidement.
La cellule bactérienne se compose (pour simplifier) d'une paroi extérieure, d'une membrane interne, d'un chromosome d'ADN, et d'ARN messager (ARNm), d'ARN ribosomal (ARNr), d'ARN de transfert (ARNt), de lipides, de protéines, de métabolites et d'ions.

Synthèse de la paroi/membrane : La paroi est composée principalement de peptidoglycane, composé de chaînes carbonées et azotées, d'acides teichoïques et lipoteichoïques. Elle protège la cellule en maintenant d'une part la charge de la cellule, et d'autre part en permettant le passage des nutriments vers la membrane. La paroi est assemblée au cours de la phase de croissance par des complexes protéiques ancrés dans la membrane et permettant ainsi le passage des composants à travers celle-ci.
La membrane est composée de deux feuillets lipidiques, assurant l'imperméabilité de la membrane et isolant le milieu intracellulaire de l'extérieur. De nombreuses protéines sont intégrées à la membrane pour assurer les échanges avec l'extérieur. On citera par exemple les transporteurs de nutriments, les oxydases et ATPases impliquées dans la respiration aérobie ou encore divers complexes protéines impliqués dans la sécrétion de protéines. Parmi ces protéines se trouvent également celles assemblant la membrane elle-même.

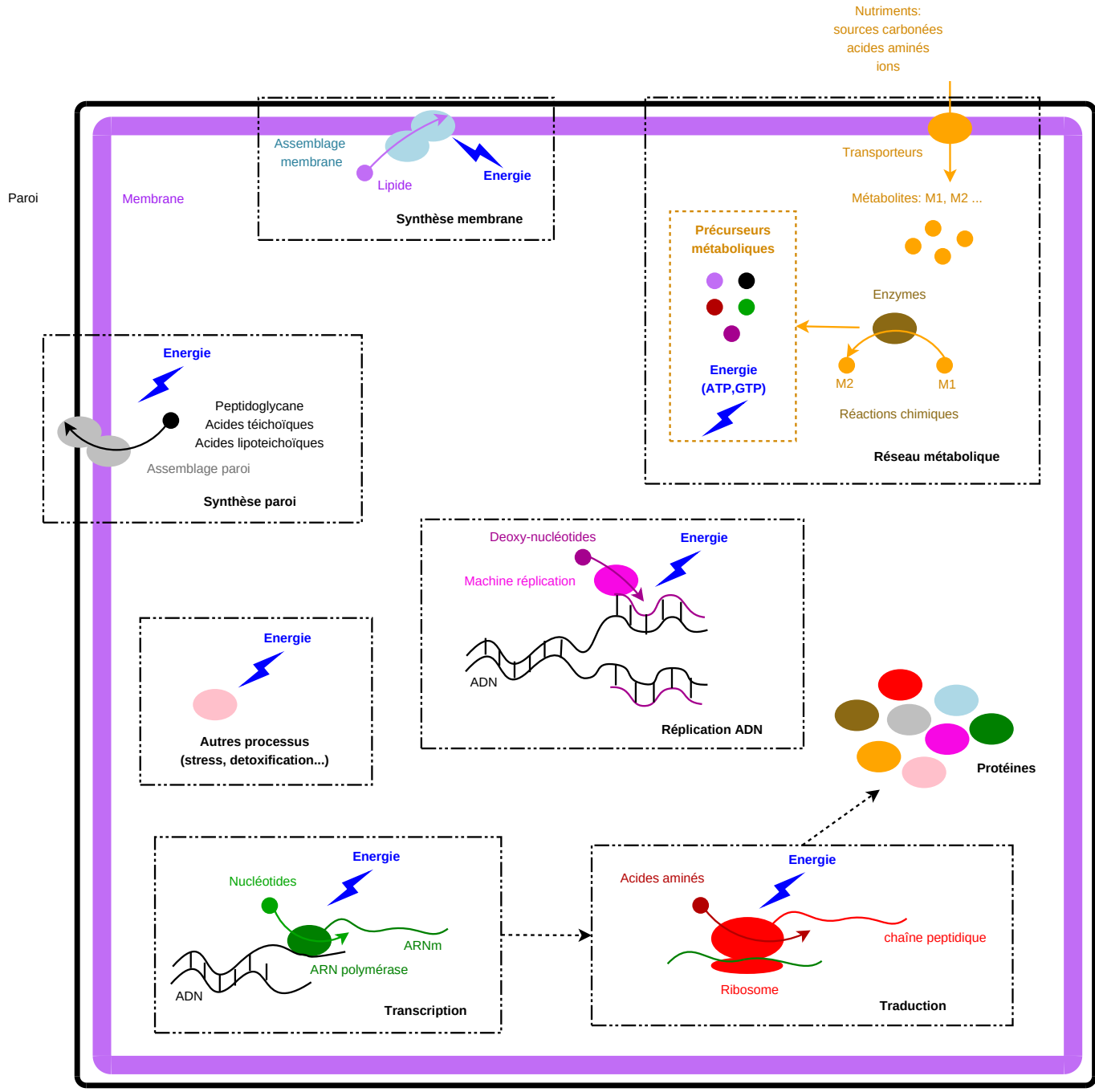


FIG. A.2 – Les processus cellulaires majeurs de la cellule en phase de croissance exponentielle

Réplication de l'ADN : L'ADN est composé de deux brins formant une hélice et reliés par des liaisons hydrogènes. Chaque brin est composé de nucléotides correspondant aux 4 bases azotées connues : adenine (A), thymine (T), guanine (G) et cytosine (C). Lors de la croissance, l'ADN est répliqué pour former deux chromosomes distincts, dont l'un des deux ira dans la cellule fille. Chaque brin est dupliqué par un complexe protéique appelé ADN polymérase par insertion de deoxy-nucléotides.

Synthèse de protéines : La synthèse d'une protéine s'effectue en deux temps. Tout d'abord le gêne d'intérêt est transcrit en ARNm au cours du processus de **transcription** par un complexe protéique appelé ARN polymérase. L'ARN polymérase se fixe sur une séquence particulière appelée promoteur et située en amont du gène (phase d'initiation). L'allongement de l'ARNm (phase d'élongation) s'effectue par insertion de nucléotides (A,U,C,G) par lecture directe du code génétique (resp. T,A,G,C). La transcription s'arrête lorsque l'ARN polymérase rencontre un codon stop. L'ARN polymérase se décroche alors de l'ADN. Dans un deuxième temps, l'ARNm est traduit en protéine au cours de la phase de **traduction**. Le principal acteur impliqué est un énorme complexe composé de 50 protéines environ et d'ARNr appelé ribosome. Le ribosome se fixe sur l'ARNm sur une séquence spécifique appelée *Ribosome Binding Sequence* (RBS) (phase d'initiation). Le ribosome se déplace le long de l'ARNm tout en produisant une chaîne peptidique. A chaque pas, le ribosome lit la séquence codante et insère l'acide aminé correspondant à la fin de la chaîne peptidique grâce à des ARNt chargés (phase d'élongation). L'ARNt se compose d'une partie où l'acide aminé (AA) se fixe (ARNt chargé), et d'une partie appelée anticodon assurant la correspondance entre le code génétique porté par l'ARNm et l'acide aminé selon le tableau A.3. Le chargement de l'acide aminé sur le ARNt est réalisé par des enzymes spécifiques, les ARNt synthases, selon la réaction

$$\text{ARNt} + \text{AA} + ATP \longrightarrow AMP + PP + \text{AA-ARNt}$$

Le ribosome s'arrête lorsqu'il rencontre un codon stop. La chaîne peptidique est ensuite repliée par des protéines chaperones pour former la protéine finale.

Synthèse des précurseurs métaboliques et de l'énergie : Tous les processus décrits précédemment ont des points communs. Ils insèrent des métabolites particuliers (lipides, acides aminés, nucléotides, deoxy-nucléotides ...) tout en consommant de l'énergie (principalement sous forme d'ATP et de GTP) pour produire les différents macro-composants cellulaires. La synthèse de ces précurseurs métaboliques/énergie est réalisée par le réseau métabolique, grâce à une succession de réactions chimiques transformant les nutriments importés à l'intérieur de la cellule. Le réseau métabolique est composé de protéines particulières (les enzymes et les transporteurs), chacune catalysant une réaction chimique transformant les métabolites substrats en produits. Certaines de ses réactions produisent également de l'énergie (ATP, GTP), nécessaire au fonctionnement de tous les processus cellulaires.

Autres processus : Il est bien évident que la cellule ne se limite pas aux seuls processus cités ci-dessus. On citera notamment les processus impliqués dans la maintenance

	U		C		A		G		
U	UUU	phénylalanine	UCU	sérine	UAU	tyrosine	UGU	cystéine	U
	UUC		UCC		UAC		UGC		C
	UUA	leucine	UCA		UAA	stop	UGA	stop/sélénocystéine	A
	UUG		UCG		UAG		UGG	tryptophane	G
C	CUU		CCU	proline	CAU	histidine	CGU	arginine	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	glutamine	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	isoleucine	ACU	thréonine	AAU	asparagine	AGU	sérine	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	lysine	AGA	arginine	A
	AUG	méthionine/start	ACG		AAG		AGG		G
G	GUU	valine	GCU	alanine	GAU	acide aspartique	GGU	glycine	U
	GUC		GCC		GAC		GGC		C
	GUA		GCA		GAA	acide glutamique	GGA		A
	GUG		GCG		GAG		GGG		G

FIG. A.3 – Code génétique

de la pression osmotique, ou dans la prévention et de la réponse au stress (oxydatif, thermique,...).

Pour résumer : Schématiquement, tous les processus cellulaires sont composés de protéines. L'objectif du réseau métabolique est de produire les précurseurs métaboliques et l'énergie nécessaires à la croissance. Les autres processus cellulaires consomment des précurseurs métaboliques et de l'énergie pour synthétiser les différents composants cellulaires. Parmi ces processus, on notera le statut particulier de l'appareil de traduction (ribosomes) dont l'objectif est de produire des protéines.

A.3 La régulation de l'expression des gènes

Il existe différents mécanismes de régulation permettant de moduler l'expression d'un gène. On distinguera les mécanismes modulant l'initiation de l'élongation de la transcription.
Les régulations agissant au niveau de l'initiation de la transcription sont généralement les plus courantes et impliquent des protéines particulières, les facteurs de transcription (FT) et les facteurs sigma (voir la figure A.4) :

- *facteur sigma* : sous-unité de l'ARN polymérase qui reconnaît la séquence promotrice du gène et permet l'accrochage de l'ARN polymérase sur le promoteur. Il existe différents facteurs sigma dans les bactéries qui sont en compétition pour former la

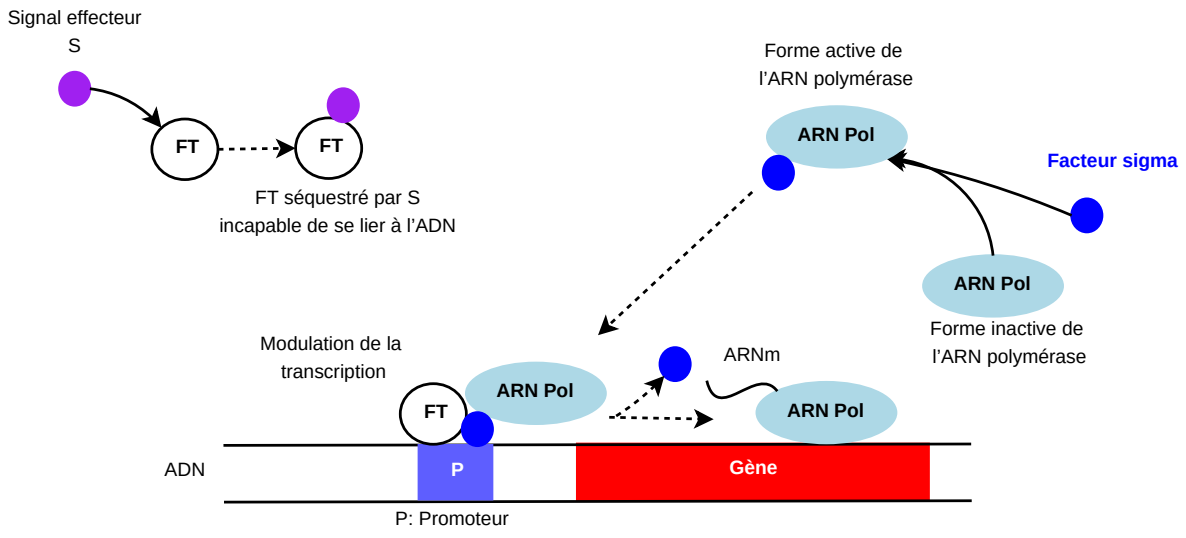


FIG. A.4 – Régulation de la transcription des gènes par les facteurs de transcription et les facteurs sigma

forme active de l'ARN polymérase capable de s'accrocher sur l'ADN.
- *facteur de transcription* : protéine qui augmente (activateur) ou diminue (répresseur) l'affinité de l'ARN polymérase pour son promoteur en se fixant sur des séquences spécifiques de l'ADN proches du promoteur.

Ainsi l'initiation de la transcription est doublement régulée. Seule l'ARN polymérase associée au bon facteur sigma pourra s'accrocher, l'accrochage étant par ailleurs modulé par les FT.
De plus, l'activité du facteur de transcription peut être modulée par un effecteur (généralement un métabolite). L'effecteur module l'accrochage du FT sur l'ADN. Dans certains cas, il peut stimuler ou au contraire empêcher l'accrochage du FT sur l'ADN, comme représenté sur la figure A.4. Le résultat final sur la transcription du gène dépend donc d'une part de l'effet positif ou négatif du FT, et d'autre part de l'effet positif ou négatif de l'effecteur sur le FT. Le régulon associé à un régulateur correspond à l'ensemble des gènes sous le contrôle direct de ce régulateur.

Les mécanismes modulant l'élongation de la transcription, autrefois minoritaires et peu étudiés, jouent en fait un rôle central notamment dans la régulation des voies métaboliques. Le principe général de ces régulations est le suivant. Après l'initiation, l'ARNm naissant possède deux motifs particulier, un terminateur précoce, et un motif anti-terminateur (ou anti-anti-terminateur) sur lequel l'élément régulateur va se fixer. L'effet de l'élément régulateur dépend de la stabilité de l'ARNm naissant vis à vis de la boucle de terminaison. Soit le terminateur est stable, et dans ce cas l'accrochage de l'élément régulateur va permettre de le déstabiliser et ainsi d'autoriser la transcription. Soit le terminateur est instable, et la transcription a lieu en l'absence d'élément régulateur. Les principaux éléments régulateurs identifiés chez *Bacillus subtilis* sont :

- les ARNt non chargés : la transcription a lieu lorsque l'ARNt non chargé s'accroche sur l'ARNm ;
- les métabolites : la transcription s'arrête généralement lorsque le métabolite est fixé sur l'ARNm ;
- des protéines : ici il n'y a pas de règles générales concernant l'effet.

Les régulations par terminaison précoce de la transcription par des métabolites sont appelés *riboswitches* en biologie.

Annexe B

Résultats associés au réseau métabolique de *Bacillus subtilis*

B.1 Modèle de connaissance

Le modèle de connaissance correspond au fichier *modele_Bsub.xls*, disponible à l'URL suivante :

`http://genome.jouy.inra.fr/~agoelzer/These/modele_Bsub.xls`

B.2 La répartition des flux métaboliques de *Bacillus subtilis* pour différents milieux

B.2.1 Variations des acides aminés

Cette figure correspond à la figure *RBA_acides_amines.jpeg*, disponible à l'URL suivante :

`http://genome.jouy.inra.fr/~agoelzer/These/RBA_acides_amines.jpeg`

B.2.2 Variations des sources de carbone

Cette figure correspond à la figure *RBA_sources_carbone.jpeg*, disponible à l'URL suivante :

`http://genome.jouy.inra.fr/~agoelzer/These/RBA_sources_carbone.jpeg`

Bibliographie

[1] T.E. Allen, M.J. Herrgard, M. Liu, Y. Qiu, J.D. Glasner, F.R. Blattner, and B.O. Palsson. Genome-scale analysis of the uses of the *Escherichia coli* genome : model-driven analysis of heterogeneous data sets. *J Bacteriol*, 185(1) :6392–6399, November 2003.

[2] T.E. Allen and B.O. Palsson. Sequence-based analysis of metabolic demands for protein synthesis in prokaryotes. *J Theor Biol*, 220(1) :1–18, January 2003.

[3] N.E. Allenby, N. O'Connor, Z. Pragai, A.C. Ward, A. Wipat, and C.R. Harwood. Genome-wide transcriptional analysis of the phosphate starvation stimulon of *Bacillus subtilis*. *J Bacteriol*, 187(23) :8063–8080, December 2005.

[4] E. Alm and A.P. Arkin. Biological networks. *Curr Opin Struct Biol*, 13(2) :193–202, April 2003.

[5] U. Alon. Network motifs : theory and experimental approaches. *Nat Rev Genet*, 8 :450–461, 2007.

[6] R. Alves and M.A. Savageau. Effect of overall feedback inhibition in unbranched biosynthetic pathways. *Biophys J*, 79 :2290–2304, November 2000.

[7] R. Alves and M.A. Savageau. Irreversibility in unbranched pathways : preferred positions based on regulatory considerations. *Biophys J*, 80 :1174–1185, March 2001.

[8] R. Alves and M.A. Savageau. Evidence of selection for low cognate amino acid bias in amino acid biosynthetic enzymes. *Mol Microbiol*, 56(4) :1017–1034, May 2005.

[9] M. Arcak and E.D. Sontag. Diagonal stability of a class of cyclic systems and its connection with the secant criterion. *Automatica*, pages 1531–1537, 2006.

[10] S. Auger, W.H. Yuen, A. Danchin, and I. Martin-Verstraete. The metic operon involved in methionine biosynthesis in *Bacillus subtilis* is controlled by transcription antitermination. *Microbiology*, 148(Pt2) :507–518, February 2002.

[11] Z. Bar-Joseph, G.K. Gerber, T.I. Lee, N.J. Rinaldi, J.Y. Yoo, F. Robert, D.B. Gordon, E. Fraenkel, T.S. Jaakkola, R.A. Young, and D.K. Gifford. Computational discovery of gene modules and regulatory networks. *Nat Biotechnol*, 21(11) :1337–1342, November 2003.

[12] A.L. Barabasi and Z.N. Oltvai. Network biology : understanding the cell's functional organization. *Nat Rev Genet*, 5 :101–113, February 2004.

[13] N. Barkai and S. Leibler. Robustness in simple biochemical networks. *Nature*, 387 :913–917, June 1997.

[14] F.U. Battistuzzi, A. Feijao, and B.S. Hedges. A genomic timescale of prokaryote evolution : insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BMC Evol Biol*, 4 :44, November 2004.

[15] D. Beard, S. Liang, and H. Qian. Energy balance for analysis of complex metabolic networks. *Biophys J*, 83(1) :79–86, July 2002.

[16] S.A. Becker, N.D. Price, and B.O. Palsson. Metabolite coupling in genome-scale metabolic networks. *BMC Bioinformatics*, 7 :111, March 2006.

[17] A. Becskei and L. Serrano. Engineering stability in gene networks by autoregulation. *Nature*, 405, June 2000.

[18] Q.K. Beg, A. Vazquez, J. Ernst, M.A. Demenezes, Z. Bar-Joseph, A.L. Barabasi, and Z.N. Oltvai. Intracellular crowding defines the mode and sequence of substrate uptake by *Escherichia coli* and contrains its metabolic activity. *Proc Natl Acad Sci USA*, 104(31) :12663–12668, July 2007.

[19] L. Beier, P. Nygaard, H. Jarmer, and H.H. Saxild. Transcription analysis of the *Bacillus subtilis* PucR regulon and identification of a cis-acting sequence required for PucR-regulated expression of genes involved in purine catabolism. *J Bacteriol*, 184(12) :3232–3241, June 2002.

[20] B.R. Belitsky. Biosynthesis of amino acids of the glutamate ans aspartate families, alanine and polyamines. In A.L. Sonenshein, editor, *Bacillus subtilis and its closest relatives : from genes to cells*, pages 203–231. American Society of Microbiology Press, Washington D.C., USA, 2001.

[21] A. Ben-Tal and A. Nemirovski. *Lectures on modern convex optimization : analysis, algorithms, and engineering applications.* MPS/SIAM Series on Optimization, 2001.

[22] D.P. Bertsekas. *Nonlinear programming.* Athena Scientific, 2nd edition, 2003.

[23] V. Blondel and J.N. Tsitsiklis. NP-hardness of some linear control design problems. *European Jour. Control*, 1, 1995.

[24] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.

[25] M.J. Brauer, C. Huttenhower, E.M. Airoldi, R. Rosenstein, J.C. Matese, D. Gresham, V.M. Boer, O.G. Troyanskaya, and D. Botstein. Coordination of growth rate, cell cycle, stress response, and metabolic activity in yeast. *Mol Biol Cell*, 19(1) :352–367, January 2008.

[26] M.J. Brauer, J. Yuan, B.D. Bennett, W. Lu, E. Kimball, D. Botstein, and J.D. Rabinowitz. Conservation of the metabolomic response to starvation across two divergent microbes. *Proc Natl Acad Sci USA*, 103(51) :19302–19307, December 2006.

[27] D. Bray. Molecular networks : the top-down view. *Science*, 301(5641) :1864–1865, September 2003.

[28] H. Bremer and P.P. Dennis. Modulation of chemical composition and other parameters of the cell by growth rate. In F.C. Neidhart, editor, *Escherichia coli and salmonella : cellular and molecular biology*, pages 1553–1569. American Society of Microbiology Press, Washington D.C., USA, 2nd edition, 1996.

[29] E.M. Bryan, B.W. Beall, and C.P. Jr. Moran. A sigma E dependent operon subject to catabolite repression during sporulation in *Bacillus subtilis*. *J Bacteriol*, 178(16) :4778–4786, August 1996.

[30] K. Bryson, V. Loux, R. Bossy, P. Nicolas, S. Chaillou, M. van de Guchte, S. Penaud, E. Maguin, M. Hoebeke, P. Bessières, and J.F. Gibrat. AGMIAL : implementing an annotation strategy for prokaryote genomes as a distributed system. *Nucleic Acids Res*, 34 :3533–3545, 2006.

[31] A.P. Burgard and C.D. Maranas. Optimization-based framework for inferring and testing hypothesized metabolic objective functions. *Biotechnol Bioeng*, 82(6) :670–677, June 2003.

[32] P. Burguiere, S. Auger, M.F. Hullo, A. Danchin, and I. Martin-Verstraete. Three different systems participate in L-cystine uptake in *Bacillus subtilis*. *J Bacteriol*, 186(15) :4875–4884, August 2004.

[33] Y. Chai, T. Norman, R. Kolter, and R. Losick. An epigenetic switch governing daughter cell separation in *Bacillus subtilis*. *Genes Dev*, 24 :754–765, March 2010.

[34] N.Y. Chen, S.Q. Jiang, D.A. Klein, and H. Paulus. Organization and nucleotide sequence of the *Bacillus subtilis* diaminopimelate operon, a cluster of genes encoding the first three enzymes of diaminopimelate synthesis and dipicolinate synthase. *J Biol Chem*, 268(13) :9448–9465, May 1993.

[35] K.H. Choi, R.J. Heath, and C.O. Rock. β-ketoacyl-acyl carrier protein synthase III (FabH) is a determining factor in branched-chain fatty acid biosynthesis. *J Bacteriol*, 182(2) :365–370, January 2000.

[36] S.K. Choi and M.H. Jr. Saier. Regulation of sigL expression by the catabolite control protein CcpA involves a roadblock mechanism in *Bacillus subtilis* : potential connection between carbon and nitrogen metabolism. *J Bacteriol*, 187 :6856–6861, 2005.

[37] S.Y. Choi, D. Reyes, M. Leelakriangsak, and P. Zuber. The global regulator Spx functions in the control of organosulfur metabolism in *Bacillus subtilis*. *J Bacteriol*, 188(16) :5741–5751, August 2006.

[38] A. Chopin, V. Biaudet, and D. Ehrlich. Analysis of the *Bacillus subtilis* genome sequence reveals nine new T-box leaders. *Mol Microbiol*, 29(2) :662, July 1998.

[39] F.M. Commichau, C. Herzberg, P. Tripal, O. Valerius, and J. Stülke. A regulatory protein-protein interaction governs glutamate biosynthesis in *Bacillus subtilis* : the glutamate dehydrogenase RocG moonlights in controlling the transcription factor GltC. *Mol Microbiol*, 65(3) :642–654, August 2007.

[40] E.M. Conlon, X.S. Liu, J.D. Lieb, and J.S. Liu. Integrating regulatory motif discovery and genome-wide expression analysis. *Proc Natl Acad Sci USA*, 100(6) :3339–3344, March 2003.

[41] M.W. Covert and B.O. Palsson. Constraints-based models : regulation of gene expression reduces the steady-state solution space. *J Theor Biol*, 221(3) :309–325, April 2003.

[42] M.W. Covert, C.H. Schilling, and B.O. Palsson. Regulation of gene expression in flux balance models of metabolism. *J Theor Biol*, 213(1) :73–88, November 2001.

[43] J.E. Craig, M.J. Ford, D.C. Blaydon, and A.L. Sonenshein. A null mutation in the *Bacillus subtilis* aconitase gene causes a block in Spo0A-phosphate-dependent gene expression. *J Bacteriol*, 179 :7351–7359, 1997.

[44] R.A. Daniel and J. Errington. Cloning, DNA sequence, functional analysis and transcriptional regulation of the genes encoding dipicolinic acid synthetase required for sporulation in *Bacillus subtilis*. *J Mol Biol*, 232(2) :468–483, July 1993.

[45] E. Dekel and U. Alon. Optimality and evolutionary tuning of the expression level of a protein. *Nature*, 436 :588–592, July 2005.

[46] J. Deutscher, A. Galinier, and I. Martin-Verstraete. Carbohydrate uptake and metabolism. In A.L. Sonenshein, editor, *Bacillus subtilis and its closest relatives : from genes to cells*, pages 129–150. American Society of Microbiology Press, Washington D.C., USA, 2001.

[47] T. Doan and S. Aymerich. Regulation of the central glycolytic genes in *Bacillus subtilis* : binding of the repressor CggR to its single DNA target sequence is modulated by fructose-1,6-bisphosphate. *Mol Microbiol*, 47(6) :1709–1721, March 2003.

[48] T. Doan, P. Servant, S. Tojo, H. Yamaguchi, G. Lerondel, K. Yoshida, Y. Fujita, and S. Aymerich. The *Bacillus subtilis* ywkA gene encodes a malic enzyme and its transcription is activated by the YufL/YufM two-component system in response to malate. *Microbiology*, 149(9) :2331–2343, September 2003.

[49] N.C. Duarte, S.A. Becker, N. Jamshidi, I. Thiele, M.L. Mo, T.D. Vo, R. Srivas, and B.O. Palsson. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci USA*, 104(6) :1777–1782, February 2007.

[50] J.S. Edwards, R.U. Ibarra, and B.O. Palsson. In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat Biotechnol*, 19(2) :125–130, February 2001.

[51] J.S. Edwards and B.O. Palsson. The *Escherichia coli* MG1655 in silico metabolic genotype : its definition, characteristics, and capabilities. *Proc Natl Acad Sci USA*, 97(10) :5528–5533, May 2000.

[52] M. Ehrenberg and C.G. Kurland. Costs of accuracy determined by a maximal growth rate constraint. *Quaterly Review of Biophysics*, 1 :45–82, 1984.

[53] M.B. Eisen, P.T. Spellman, P.O. Brown, and D. Botstein. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA*, 95 :14863–14868, December 1998.

[54] J. Elf, D. Nilsson, T. Tenson, and M. Ehrenberg. Selective charging of tRNA isoacceptors explains patterns of codon usage. *Science*, 300 :1718–1724, June 2003.

[55] M. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403 :335–338, January 2000.

[56] S. Even, O. Pellegrini, L. Zig, V. Labas, J. Vinh, D. Bréchemmier-Baey, and H. Putzer. Ribonucleases J1 and J2 : two novel endoribonucleases in *B.subtilis* with functional homology to *E.coli* RNase E. *Nucleic Acids Res*, 33(7) :2141–2152, April 2005.

[57] C. Eymann, D. Becher, J. Bernhardt, K. Gronau, A. Klutzny, and M. Hecker. Dynamics of protein phosphorylation on ser/thr/tyr in *Bacillus subtilis*. *Proteomics*, 7(19) :3509–3526, October 2007.

[58] G. Ferreres. *A practical approach to robustness analysis with aeronautical applications*. Kluwer Academic Publishers, 1999.

[59] E. Fischer and U. Sauer. Large-scale *in vivo* flux analysis shows rigidity and suboptimal performance of *Bacillius subtilis* metabolism. *Nat Genet*, 37(6) :636–640, 2005.

[60] S.H. Fisher. Utilization of amino acids and other nitrogen-containing compounds. In A.L. Sonenshein, editor, *Bacillus subtilis and other gram-positive bacteria. Biochemistry, physiology and molecular genetics*, pages 221–228. American Society of Microbiology Press, Washington D.C., USA, 1993.

[61] S.H. Fisher, K. Rohrer, and A.E. Ferson. Role of CodY in regulation of the *Bacillus subtilis* hut operon. *J Bacteriol*, 178(13) :3779–3784, July 1996.

[62] S.H. Fisher, M.A. Strauch, M.R. Atkinson, and L.V. Jr. Wray. Modulation of *Bacillus subtilis* catabolite repression by transition state regulatory protein AbrB. *J Bacteriol*, 176(7) :1903–1912, April 1994.

[63] S.H. Fisher and L.V. Wray. *Bacillus subtilis* 168 contains two differentially regulated genes encoding L-asparaginase. *J Bacteriol*, 184(8) :2148–2154, April 2002.

[64] S.H. Fisher and L.V. Jr. Wray. Mutations in the *Bacillus subtilis* glnRA operon that cause nitrogen source-dependent defects in regulation of TnrA activity. *J Bacteriol*, 184(16) :4636–4639, August 2002.

[65] S.H. Fisher and L.V. Jr. Wray. *Bacillus subtilis* glutamine synthetase regulates its own synthesis by acting as a chaperone to stabilize GlnR-DNA complexes. *Proc Natl Acad Sci USA*, 105(3) :1014–1019, January 2008.

[66] J.A. Freyre-Gonzalez, J.A. Alonso-Pavon, L.G. Trevino-Quintanilla, and J. Collado-Vides. Functional architecture of *Escherichia coli* : new insights provided by a natural decomposition approach. *Genome Biol*, 9(10) :R154, October 2008.

[67] V. Fromion, G. Scorletti, and G. Ferreres. Nonlinear performance of a pi controlled missile : en explanation. In *36th Conference on Decision and Control*, 1997.

[68] J. Förster, I. Famili, P. Fu, B.O. Palsson, and J. Nielsen. Genome-scale reconstruction of *Saccharomyces cerevisiae* metabolic network. *Genome Research*, 13 :244–253, 2003.

[69] A. Gaballa, H. Antelmann, C. Aguilar, S.K. Khakh, K.B. Song, G.T. Smaldone, and J.D. Helmann. The *Bacillus subtilis* iron-sparing response is mediated by a Fur-regulated small RNA and three small, basic proteins. *Proc Natl Acad Sci USA*, 105(33) :11927–11932, August 2008.

[70] R. Gardan, G. Rapoport, and M. Debarbouille. Expression of the rocDEF operon involved in arginine catabolism in *Bacillus subtilis*. *J Mol Biol*, 249(5) :843–856, June 1995.

[71] T.S. Gardner, C.R. Cantor, and J.J. Collins. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 403 :339–342, January 2000.

[72] U. Gerth, H. Kock, I. Kusters, S. Michalik, R.L. Switzer, and M. Hecker. Clp-dependent proteolysis down-regulates central metabolic pathways in glucose-starved *Bacillus subtilis*. *J Bacteriol*, 190(1) :321–331, January 2008.

[73] E.P. Gianchandani, M.A. Oberhardt, A.P. Burgard, C.D. Maranas, and J.A. Papin. Predicting biological system objectives de novo from internal state measurements. *BMC Bioinformatics*, 9 :43, January 2008.

[74] J.C. Gilbert, C. Lemaréchal, and C. Sagastizábal. *Numerical Optimization : Theoretical and Practical Aspects*. Springer-Verlag, 2nd edition, 2006.

[75] A. Goelzer, F. Bekkal Brikci, I. Martin-Verstraete, P. Noirot, P. Bessières, S. Aymerich, and V. Fromion. Reconstruction and analysis of the genetic and metabolic regulatory networks of the central metabolism of *Bacillus subtilis*. *BMC Syst Biol*, 2 :20, February 2008.

[76] A. Goelzer, V. Fromion, and G. Scorletti. Cell design in bacteria as a convex optimization problem. *Proceedings of the 48th IEEE Conference on Decision and Control*, pages 4517–4522, December 2009.

[77] P. Gollnick, P. Babitzke, A. Antson, and C. Yanofsky. Complexity in regulation of tryptophan biosynthesis in *Bacillus subtilis*. *Annual Review Genetics*, 39 :47–68, 2005.

[78] J.L. Gouzé. Positive and negative circuts in dynamical systems. *Journal of Biological Systems*, 6 :11–15, 1998.

[79] J.A. Grandoni, S.B. Fulmer, V. Brizzio, S.A. Zahler, and J.M. Calvo. Regions of the *Bacillus subtilis* ilv-leu operon involved in regulation by leucine. *J Bacteriol*, 175(23) :7581–7593, December 1993.

[80] L.M. Graves and R.L. Switzer. Aspartokinase II from *Bacillus subtilis* is degraded in response to nutrient limitation. *J Biol Chem*, 65(25) :14947–14955, September 1990.

[81] L.M. Graves and R.L. Switzer. Aspartokinase III, a new isozyme in *Bacillus subtilis* 168. *J Bacteriol*, 172(1) :218–223, January 1990.

[82] J.S. Grayston, D. Vaughan, and D. Jones. Rhizosphere carbon flow in trees, in comparison with annual plants : the importance of root exudation and its impact on microbial activity and nutrient availability. *Appl Soil Ecol*, 5 :29–56, July 1996.

[83] F.J. Grundy, S.C. Lehman, and T.M. Henkin. The L box regulon : lysine sensing by leader RNAs of bacterial lysine biosynthesis genes. *Proc Natl Acad Sci USA*, 100(21) :12057–12062, October 2003.

[84] L.W. Hamoen, G. Venema, and O.P. Kuipers. Controlling competence in *Bacillus subtilis* : shared use of regulators. *Microbiology*, 149(Pt1) :9–17, January 2003.

[85] S.H. Hanson and D.P. Cox. Effect of different nutritional conditions on the synthesis of tricarboxylic acid cycle enzymes. *J Bacteriol*, 93(6) :1777–1787, June 1967.

[86] L.H. Hartwell, J.J. Hopfield, S. Leibler, and A.W. Murray. From molecular to modular cell biology. *Nature*, 402 Supp :C47–C52, December 1999.

[87] J. Hasty, F. Isaacs, M. Dolnik, D. McMillen, and J.J. Collins. Designer gene networks : Towards fundamental cellular control. *Chaos*, 11(1) :207–220, March 2001.

[88] J. Hasty, D. McMillen, and J.J. Collins. Engineered gene circuits. *Nature*, 420 :224–230, November 2002.

[89] J. Hasty, D. McMillen, F. Isaacs, and J.J. Collins. Computational studies of gene regulatory networks : in numero molecular biology. *Nat Rev Genet*, 2 :268–279, 2001.

[90] P.M. Haverty, U. Hansen, and Z. Weng. Computational inference of transcriptional regulatory networks from expression profiling and transcription factor binding site identification. *Nucleic Acids Res*, 32(1) :179–188, January 2004.

[91] R. Heinrich, F. Montero, E. Klipp, T.G. Waddell, and E. Meléndez-Hevia. Theoretical approaches to the evolutionary optimization of glycolysis : thermodynamic and kinetic constraints. *Eur J Biochem*, 243(1-2) :191–201, January 1997.

[92] T.M. Henkin, B.L. Glass, and F.J. Grundy. Analysis of the *Bacillus subtilis* tyrS gene : conservation of a regulatory sequence in multiple tRNA synthetase genes. *J Bacteriol*, 174(4) :1299–1306, February 1992.

[93] M.J. Herrgard, M. Covert, and B.O. Palsson. Reconstruction of microbial transcriptional regulatory networks. *Curr Opin Biotechnol*, 15(1) :70–77, February 2004.

[94] M.J. Herrgard, B.S. Lee, V. Portnoy, and B.O. Palsson. Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. *Genome Res*, 16(5) :627–635, May 2006.

[95] M.J. Herrgard, N. Swainston, P. Dobson, W.B. Dunn, and K.Y. et al. Arga. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat Biotechnol*, 26(10) :1155–1160, October 2008.

[96] I.M. Horowitz. *Synthesis of Feedback Systems*. Academic Press, New York, 1963.

[97] M.F. Hullo, S. Auger, O. Soutourina, O. Barzu, M. Yvon, A. Danchin, and I. Martin-Verstraete. Conversion of methionine to cysteine in *Bacillus subtilis* and its regulation. *J Bacteriol*, 189(1) :187–197, January 2007.

[98] J. Ihmels, S. Bergmann, and N. Barkai. Defining transcription modules using large-scale gene expression data. *Bioinformatics*, 20(13) :1993–2003, September 2004.

[99] J. Ihmels, G. Friedlander, S. Bergmann, O. Sarig, Y. Ziv, and N. Barkai. Revealing modular organization in the yeast transcriptional network. *Nat Genet*, 31 :370–377, August 2002.

[100] T. Iijima, M.D. Diesterhaft, and E. Freese. Sodium effect of growth on aspartate and genetic analysis of a *Bacillus subtilis* mutant with high aspartase activity. *J Bacteriol*, 129(3) :1440–1447, March 1977.

[101] J.A. Imlay. Pathways of oxidative damage. *Annu Rev Microbiol*, 57 :395–418, 2003.

[102] N. Ishii, K. Nakahigashi, T. Baba, M. Robert, T. Soga, A. Kanai, T. Hirasawa, M. Naba, K. Hirai, A. Hoque, P.Y. Ho, Y. Kakazu, K. Sugawara, S. Igarashi, S. Harada, T. Masuda, N. Sugiyama, T. Togashi, M. Hasegawa, Y. Takai, K. Yugi, K. Arakawa, N. Iwata, Y. Toya, Y. Nakayama, T. Nishioka, K. Shimizu, H. Mori, and M. Tomita. Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. *Science*, 316(5824) :593–597, April 2007.

[103] H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, and A.L. Barabasi. The large-scale organization of metabolic networks. *Nature*, 407 :651–654, October 2000.

[104] C. Jourlin-Castelli, N. Mani, M.M. Nakano, and A.L. Sonenshein. CcpC, a novel regulator of the LysR family required for glucose repression of the citB gene in *Bacillus subtilis*. *J Mol Biol*, 295(4) :865–878, January 2000.

[105] S. Kaplan, A. Bren, E. Dekel, and U. Alon. The incoherent feed-forward loop can generate non-monotonic input functions for genes. *Mol Syst Biol*, 4 :203, 2008.

[106] P.D. Karp, S. Paley, and P. Romero. The pathway tools software. *Bioinformatics*, 18 :S225–S232, 2002.

[107] P.D. Karp, M. Riley, M. Saier, I.T. Paulsen, S.M. Paley, and A. Pellegrini-Toole. The ecocyc and metacyc databases. *Nucleic Acids Res*, 28(1) :56–59, January 2000.

[108] K.J. Kauffman, P. Prakash, and J.S. Edwards. Advances in flux balance analysis. *Curr Opin Biotechnol*, 14(5) :491–496, October 2003.

[109] H.J. Kim, S.I. Kim, M. Ratnayake-Lecamwasam, K. Tachikawa, A.L. Sonenshein, and M. Strauch. Complex regulation of the *Bacillus subtilis* aconitase gene. *J Bacteriol*, 185(5) :1672–1680, March 2003.

[110] S.I. Kim, C. Jourlin-Castelli, S.R. Wellington, and A.L. Sonenshein. Mechanism of repression by *Bacillus subtilis* CcpC, a LysR family regulator. *J Mol Biol*, 334(4) :609–624, December 2003.

[111] H. Kitano. Systems biology : toward system-level understanding of biological systems. In *Foundations of Systems Biology*, pages 1–29. MIT Press, 2001.

[112] H. Kitano. Towards a theory of biological robustness. *Mol Syst Biol*, 3 :137, 2007.

[113] K.R. Kjeldsen and J. Nielsen. In silico genome-scale reconstruction and validation of the *Corynebacterium glutamicum* metabolic network. *Biotechnol Bioeng*, 102(2) :583–597, February 2009.

[114] R.J. Kleijn, J.M. Buescher, L. Le Chat, M. Jules, S. Aymerich, and U. Sauer. Metabolic fluxes during strong carbon catabolite repression by malate in *Bacillus subtilis*. *J Biol Chem*, 285(3) :1587–1596, January 2010.

[115] A.L. Knorr, R. Jain, and R. Srivastavan. Bayesian-based selection of metabolic objective functions. *Bioinformatics*, 23(3) :351–357, February 2007.

[116] K. Kobayashi, S.D. Ehrlich, A. Albertini, G. Amati, K.K. Andersen, M. Arnaud, K. Asai, S. Ashikaga, S. Aymerich, P. Bessieres, F. Boland, S.C. Brignell, S. Bron, K. Bunai, J. Chapuis, L.C. Christiansen, A. Danchin, M. Debarbouille, E. Dervyn, E. Deuerling, K. Devine, S.K. Devine, O. Dreesen, J. Errington, S. Fillinger, S.J. Foster, Y. Fujita, A. Galizzi, R. Gardan, C. Eschevins, T. Fukushima, K. Haga, C.R. Harwood, M. Hecker, D. Hosoya, M.F. Hullo, H. Kakeshita, D. Karamata, Y. Kasahara, F. Kawamura, K. Koga, P. Koski, R. Kuwana, D. Imamura, M. Ishimaru, S. Ishikawa, I. Ishio, D. Le Coq, A. Masson, C. Mauel, R. Meima, R.P. Mellado, A. Moir, S. Moriya, E. Nagakawa, H. Nanamiya, S. Nakai, P. Nygaard, M. Ogura, T. Ohanan, M. O'Reilly, M. O'Rourke, Z. Pragai, H.M. Pooley, G. Rapoport, J.P. Rawlins, L.A. Rivas, C. Rivolta, A. Sadaie, Y. Sadaie, M. Sarvas, T. Sato, H.H. Saxild, E. Scanlan, W. Schumann, J.F. Seegers, J. Sekiguchi, A. Sekowska, S.J. Seror, M. Simon, P. Stragier, R. Studer, H. Takamatsu, T. Tanaka, M. Takeuchi, H.B. Thomaides, V. Vagner, J.M. van Dijl, K. Watabe, A. Wipat, H. Yamamoto, M. Yamamoto, Y. Yamamoto, K. Yamane, K. Yata, K. Yoshida, H. Yoshikawa, U. Zuber, and N. Ogasawara. Essential *Bacillus subtilis* genes. *Proc Natl Acad Sci USA*, 100(8) :4678–4683, April 2003.

[117] T. Koburger, J. Weibezahn, J. Bernhardt, G. Homuth, and M. Hecker. Genome-wide mRNA profiling in glucose starved *Bacillus subtilis* cells. *Mol Genet Genomics*, 274(1) :1–12, August 2005.

[118] M.C. Konopka, I.A. Shkel, S. Cayley, M.T. Record, and J.C. Weisshaar. Crowding and confinement effects on protein diffusion in vivo. *J Bacteriol*, 188(17) :6115–6123, September 2006.

[119] L. Krásný and R.L. Gourse. An alternative strategy for bacterial ribosome synthesis : *Bacillus subtilis* rRNA transcription regulation. *EMBO J*, 33(22) :4473–4483, 2004.

[120] H.E. Kubitschek, W.W. Baldwin, S.J. Schroeter, and R. Graetzer. Independence of buoyant cell density and growth rate in *Escherichia coli*. *J Bacteriol*, 158(1) :296–299, April 1984.

[121] A. Kumar, P.D. Christofides, and P. Daoutidis. Singular perturbation modeling of nonlinear processes with nonexplicit time-scale multiplicity. *Chem Eng Sci*, 53(8) :1491–1504, 1998.

[122] F. Kunst, N. Ogasawara, I. Moszer, A.M. Albertini, G. Alloni, V. Azevedo, M.G. Bertero, P. Bessieres, A. Bolotin, S. Borchert, R. Borriss, L. Boursier, A. Brans, M. Braun, S.C. Brignell, S. Bron, S. Brouillet, C.V. Bruschi, B. Caldwell, V. Capuano, N.M. Carter, S.K. Choi, J.J. Codani, I.F. Connerton, and A. et al. Danchin. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature*, 390(6657) :249–256, November 1997.

[123] S. Lee, C. Palakornkule, M.M. Domach, and I.E. Grossmann. Recursive MILP model for finding all the alternate optima in LP models for metabolic networks. *Comput Chem Eng*, 24 :711–716, July 2000.

[124] T.I. Lee, N.J. Rinaldi, F. Robert, D.T. Odom, Z. Bar-Joseph, G.K. Gerber, N.M. Hannett, C.T. Harbison, C.M. Thompson, I. Simon, J. Zeitlinger, E.G. Jennings, H.L. Murray, D.B. Gordon, B. Ren, J.J. Wyrick, J.B. Tagne, T.L. Volkert, E. Fraenkel, D.K. Gifford, and R.A. Young. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science*, 298 :799–804, October 2002.

[125] J.C. Leloup and A. Goldbeter. Modeling the mammalian circadian clock : sensitivity analysis and multiplicity of oscillatory mechanisms. *J Theor Biol*, 230(4) :541–562, October 2004.

[126] V.M. Levdikov, E. Blagova, P. Joseph, A.L. Sonenshein, and A.J. Wilkinson. The structure of CodY, a GTP- and isoleucine-responsive regulator of stationary phase and virulence in gram-positive bacteria. *J Biol Chem*, 281(16) :11366–11373, April 2006.

[127] S. Levy and N. Barkai. Coordination of gene expression with growth rate : a feedback or a feed-forward strategy ? *FEBS Lett*, 583(24) :3974–3978, December 2009.

[128] X. Litrico and V. Fromion. *Modeling and Control of Hydrosystems*. Springer-Verlag, 2009.

[129] Y. Lu and R.L. Switzer. Transcriptional attenuation of *Bacillus subtilis* pyr operon by the PyrR regulatory protein and uridine nucleotides in vitro. *J Bacteriol*, 178(24) :7206–7211, December 1996.

[130] N.M. Luscombe, M.M. Babu, H. Yu, M. Snyder, S.A. Teichmann, and M. Gerstein. Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature*, 431(7006) :308–312, September 2004.

[131] D. López, H. Vlamakis, R. Losick, and R. Kolter. Cannibalism enhances biofilm development in *Bacillus subtilis*. *Mol Microbiol*, 74(3) :609–618, November 2009.

[132] H.W. Ma, J. Buer, and A.P. Zeng. Hierarchical structure and modules in the *Escherichia coli* transcriptional regulatory network revealed by a new top-down approach. *BMC Bioinformatics*, 5 :199, 2004.

[133] H.W. Ma, B. Kumar, U. Ditges, F. Gunzer, J. Buer, and A.P. Zeng. An extended transcriptional regulatory network of *Escherichia coli* and analysis of its hierarchical structure and network motifs. *Nucleic Acids Res*, 32(22) :6643–6649, December 2004.

[134] B. Macek, I. Mijakovic, J.V. Olsen, F. Gnad, C. Kumar, P.R. Jensen, and M. Mann. The serine/threonine/tyrosine phosphoproteome of the model bacterium *Bacillus subtilis*. *Mol Cell Proteomics*, 6(4) :697–707, April 2007.

[135] U. Mader, S. Hennig, M. Hecker, and G. Homuth. Transcriptional organization and posttranscriptional regulation of the *Bacillius subtilis* branched-chain amino acid biosynthesis genes. *J Bacteriol*, 186(8) :2240–2252, April 2004.

[136] U. Mader, G. Homuth, C. Scharf, K. Buttner, R. Bode, and M. Hecker. Transcriptome and proteome analysis of *Bacillius subtilis* gene expression modulated by amino acid availability. *J Bacteriol*, 184(15) :4288–4295, August 2002.

[137] R. Mahadevan, J.S. Edwards, and F. 3rd. Doyle. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys J*, 83 :1331–1340, September 2002.

[138] R. Mahadevan and C.H. Schilling. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab Eng*, 5(4) :264–276, October 2003.

[139] Y. Makita, M. Nakao, N. Ogasawara, and K. Nakai. DBTBS : database of transcriptional regulation in *Bacillus subtilis* and its contribution to comparative genomics. *Nucleic Acids Res*, 32 :D75–D77, 2004.

[140] S. Mangan and U. Alon. Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci USA*, 100(21) :11980–11985, October 2003.

[141] S. Mangan, S. Itzkovitz, A. Zaslaver, and U. Alon. The incoherent feed-forward loop accelerates the response-time of the gal system of *Escherichia coli*. *J Mol Biol*, 356(5) :1073–1081, March 2006.

[142] A.G. Marr. Growth rate of *Escherichia coli*. *Microbiol Rev*, 55(2) :316–333, June 1991.

[143] A. Martínez-Antonio, S. Janga, and D. Thieffry. Functional organisation of *Escherichia coli* transcriptional regulatory network. *J Mol Biol*, 381(1) :238–247, August 2008.

[144] H. Matsuoka, K. Hirooka, and Y. Fujita. Organization and function of the YsiA regulon of *Bacillus subtilis* involved in fatty acid degradation. *J Biol Chem*, 282(8) :5180–5194, February 2007.

[145] F. Matthäus, C. Salazar, and O. Ebenhöh. Biosynthetic potentials of metabolites and their hierarchical organization. *PLoS Comput Biol*, 4(4) :e1000049, April 2008.

[146] C.M. Miller, S. Baumberg, and P.G. Stockley. Operator interactions by the *Bacillus subtilis* arginine repressor/activator, AhrC : novel positioning and DNA-mediated assembly of a transcriptional activator at catabolic sites. *Mol Microbiol*, 26(1) :37–48, October 1997.

[147] R. Milo, S.S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs : Simple building blocks of complex networks. *Science*, 298 :824–827, October 2002.

[148] Y. Miwa, A. Nakata, A. Ogiwara, M. Yamamoto, and Y. Fujita. Evaluation and characterization of catabolite-responsive elements (cre) of *Bacillus subtilis*. *Nucleic Acids Res*, 28(5) :1206–1210, March 2000.

[149] V. Molle, Y. Nakaura, R.P. Shivers, H. Yamaguchi, R. Losick, Y. Fujita, and A.L. Sonenshein. Additional targets of the *Bacillus subtilis* global regulator CodY identified by chromatin immunoprecipitation and genome-wide transcript analysis. *J Bacteriol*, 185(6) :1911–1922, March 2003.

[150] I. Moszer, L.M. Jones, S. Moreira, C. Fabry, and A. Danchin. SubtiList : the reference database for the *Bacillus subtilis* genome. *Nucleic Acids Res*, 30(1) :62–65, January 2002.

[151] U. Mäder, G. Homuth, U. Völker, and M. Hecker. Global regulatory systems mediating general starvation responses in *Bacillus subtilis*. In Y. Fujita, editor, *Global regulatory networks in Bacillus subtilis*, pages 205–228. Transworld Research Network, Trivandrum, Kerala, India, 1st edition, 2007.

[152] F.C. Neidhart. *Escherichia coli and salmonella : cellular and molecular biology*. American Society of Microbiology Press, Washington D.C., USA, 2nd edition, 1996.

[153] F.C. Neidhart and H.E. Umbarger. Chemical composition of *Escherichia coli*. In F.C. Neidhart, editor, *Escherichia coli and salmonella : cellular and molecular biology*, pages 13–16. American Society of Microbiology Press, Washington D.C., USA, 2nd edition, 1996.

[154] Y. Nesterov. *Introductory lectures on convex optimization : a basic course*. Kluwer Academic Publishers, 2004.

[155] C. Nguyen. Rhizodeposition of organic C by plants : mechanisms and controls. *Agronomie*, 23 :375–396, 2003.

[156] C. Nicolas. *Analyse globale de la réponse adaptative du métabolisme central d'*E. Coli *à une modulation contrôlée de la glucose-6-phosphate déshydrogénase : étude par fluxomique et corrélation transcriptome / fluxome*. PhD thesis, Institut National des Sciences Appliquées de Toulouse, Toulouse, France, 2007. In French.

[157] P. Nicolas, A.S. Tocquet, V. Miele, and F. Muri. A reversible jump Markov chain Monte Carlo algorithm for bacterial promoter motifs discovery. *J Comput Biol*, 13(3) :651–667, April 2006.

[158] K. Ochi, J.C. Kandala, and E. Freese. Initiation of *Bacillus subtilis* sporulation by the stringent response to partial amino acid deprivation. *J Biol Chem*, 256(13) :6866–6875, June 1981.

[159] H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, and M. Kanehisa. KEGG : Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*, 27(1) :29–34, January 1999.

[160] Y.K. Oh, B.O. Palsson, S.M. Park, C.H. Schilling, and R. Mahadevan. Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. *J Biol Chem*, 282(39) :28791–28799, September 2007.

[161] M. Ohne. Regulation of aconitase synthesis in *Bacillus subtilis* : induction, feedback repression and catabolite repression. *J Bacteriol*, 117(3) :1295–1305, March 1974.

[162] M. Ohne. Regulation of the dicarboxylic acid part of the citric acid cycle in *Bacillus subtilis*. *J Bacteriol*, 122(1) :224–234, April 1975.

[163] A.P. Oliveira, J. Nielsen, and J. Förster. Modeling *Lactococcus lactis* using a genome-scale flux model. *BMC Microbiology*, 5(39) :1–15, June 2005.

[164] J.A. Papin, J. Stelling, N.D. Price, S. Klamt, S. Schuster, and B.O. Palsson. Comparison of network-based pathway analysis methods. *Trends Biotechnol*, 22(8) :400–405, August 2004.

[165] B. Papp, C. Pal, and L.D. Hurst. Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature*, 429(6945) :661–664, June 2004.

[166] M. Pelchat and J. Lapointe. In vivo and in vitro processing of the *Bacillus subtilis* transcript coding for glutamyl-tRNA synthetase, serine acetyltransferase, and cysteinyl-tRNA synthetase. *RNA*, 5(2) :281–289, February 1999.

[167] S. Picossia, B.R. Belitskya, and A.L. Sonenshein. Molecular Mechanism of the Regulation of *Bacillus subtilis* gltAB Expression by GltC. *J Mol Biol*, 365(5) :1298–1313, February 2007.

[168] P.J. Piggot and D.W. Hilbert. Sporulation of *Bacillus subtilis*. *Curr Opin Microbiol*, 7(6) :579–586, December 2004.

[169] M.M. Ponce-de leon and L.I. Pizer. Serine biosynthesis and its regulation in *Bacillus subtilis*. *J Bacteriol*, 110(3) :895–904, June 1972.

[170] J. Pramanik and J.D. Keasling. Stoichiometric model of *Escherichia coli* metabolism : incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. *Biotechnol Bioeng*, 56(4) :398–421, November 1997.

[171] E. Presecan-Siedel, A. Galinier, R. Longin, J. Deutscher, A. Danchin, P. Glaser, and I. Martin-Verstraete. Catabolite regulation of the pta gene as part of carbon flow pathways in *Bacillus subtilis*. *J Bacteriol*, 181(22) :6889–6897, November 1999.

[172] H. Putzer, N. Gendron, and M. Grunberg-Manago. Co-ordinate expression of the two threonyl-tRNA synthetase genes in *Bacillus subtilis* : control by transcriptional antitermination involving a conserved regulatory sequence. *EMBO J*, 11(8) :3117–3127, August 1992.

[173] H. Ramstrom, S. Sanglier, E. Leize-Wagner, C. Philippe, A. Van Dorsselaer, and J. Haiech. Properties and regulation of the bifunctional enzyme HPr kinase/phosphatase in *Bacillus subtilis*. *J Biol Chem*, 278(2) :1174–1185, January 2003.

[174] P Rappu, T Pullinen, and P Mantsala. In vivo effect of mutations at the prpp binding site of the bacillus subtilis purine repressor. *J Bacteriol*, 185(22) :6728–6731, November 2003.

[175] E. Ravasz, A.L. Somera, D.A. Mongru, Z.N. Oltvai, and A.L. Barabasi. Hierarchical organization of modularity in metabolic networks. *Science*, 297 :1551–1555, August 2002.

[176] J. Reed, T. Vo, C. Schilling, and B.O. Palsson. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol*, 4(9) :R54, 2003.

[177] M.C. Renna, N. Najimudin, L.R. Winik, and S.A. Zahler. Regulation of the *Bacillus subtilis* alsS, alsD, and alsR genes involved in post-exponential-phase production of acetoin. *J Bacteriol*, 175(12) :3863–3875, June 1993.

[178] A. Rosner. Control of lysine biosynthesis in *Bacillus subtilis* : inhibition of diaminopimelate decarboxylase by lysine. *J Bacteriol*, 121(1) :20–28, January 1975.

[179] A. Rosner and H. Paulus. Regulation of aspartokinase in *Bacillus subtilis*. The separation and properties of two isofunctional enzymes. *J Biol Chem*, 246(9) :2965–2971, May 1971.

[180] L. Royer, A. Goelzer, P. De Larminat, and Pichon Y. Procédé et dispositif de commande d'un groupe motopropulseur avec une transmission infiniment variante. Technical Report WO 2004/106097 A2, Organisation Mondiale de la Propriété Intellectuelle, 2004.

[181] H. Salgado, S. Gama-Castro, A. Martinez-Antonio, E. Diaz-Peredo, F. Sanchez-Solano, M. Peralta-Gil, D. Garcia-Alonso, V. Jimenez-Jacinto, A. Santos-Zavaleta, C. Bonavides-Martinez, and J. Collado-Vides. RegulonDB (version 4.0) : transcriptional regulation, operon organization and growth conditions in *Escherichia coli* K-12. *Nucleic Acids Res*, 32 :D303–D306, 2004.

[182] H. Salgado, A. Santos-Zavaleta, S. Gama-Castro, M. Peralta-Gil, M.I. Penaloza-Spinola, A. Martinez-Antonio, P.D. Karp, and J. Collado-Vides. The comprehensive updated regulatory network of *Escherichia coli* K-12. *BMC Bioinformatics*, 7(1) :5, January 2006.

[183] M. Santillan and M.C. Mackey. Dynamic regulation of the tryptophan operon : A modeling study and comparison with experimental data. *Proc Natl Acad Sci USA*, 98(4) :1364–1369, 2001.

[184] M.G. Sargent. Control of cell length. *J Bacteriol*, 123(1) :7–19, July 1975.

[185] U. Sauer, V. Hatzimanikatis, H.P. Hohmann, M. Manneberg, A.P. van Loon, and J.E. Bailey. Physiology and metabolic fluxes of wild-type and riboflavin-producing *Bacillus subtilis*. *Appl Environ Microbiol*, 62(10) :3687–3696, October 1996.

[186] H.M. Sauro. Modularity defined. *Mol Syst Biol*, 4 :166, 2008.

[187] H.H. Saxild, K. Brunstedt, K.I. Nielsen, H. Jarmer, and P. Nygaard. Definition of the *Bacillus subtilis* PurR operator using genetic and bioinformatic tools and expansion of the PurR regulon with glyA, guaC, pbuG, xpt-pbuX, yqhZ-folD, and pbuO. *J Bacteriol*, 183(21) :6175–6183, November 2001.

[188] D.J. Scheffers. The cell wall of *Bacillus subtilis*. In P. Graumann, editor, *Bacillus : Cellular and Molecular Biology*, pages 331–374. Caister Academic Press, Norfolk, UK, 1st edition, 2007.

[189] C.H. Schilling, M. Covert, I. Famili, G. Church, J. Edwards, and B.O. Palsson. Genome-scale metabolic model of *Helicobacter pylori* 26695. *J Bacteriol*, 184(16) :4582–4593, August 2002.

[190] C.H. Schilling, D. Letscher, and B.O. Palsson. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J Theor Biol*, 203(3) :229–248, April 2000.

[191] C.H. Schilling and B.O. Palsson. Assessment of the metabolic capabilities of *Haemophilus influenzae Rd* through a genome-scale pathway analysis. *J Theor Biol*, 203(3) :249–283, April 2000.

[192] O. Schilling, O. Frick, C. Herzberg, A. Ehrenreich, E. Heinzle, C. Wittmann, and J. Stülke. Transcriptional and metabolic responses of *Bacillus subtilis* to the availability of organic acids : transcription regulation is important but not sufficient to account for metabolic adaptation. *Appl Environ Microbiol*, 73(2) :499–507, January 2007.

[193] I. Schomburg, A. Chang, C. Ebeling, M. Gremse, C. Heldt, G. Huhn, and D. Schomburg. BRENDA, the enzyme database : updates and major new developments. *Nucleic Acids Res*, 32 :D431–433, January 2004.

[194] R. Schuetz, L. Kuepfer, and U. Sauer. Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Mol Syst Biol*, 3 :119, 2007.

[195] G.E. Schujman, L. Paoletti, A.D. Grossman, and D. de Mendoza. FapR, a bacterial transcription factor involved in global regulation of membrane lipid biosynthesis. *Dev Cell*, 4(5) :663–672, May 2003.

[196] D. Schultz, P.G. Wolynes, E. Ben Jacob, and J.N. Onuchic. Deciding fate in adverse times : sporulation and competence in *Bacillus subtilis*. *Proc Natl Acad Sci USA*, 106(50) :21027–21034, December 2009.

[197] S. Schuster, T. Dandekar, and D. Fell. Detection of elementary flux modes in biochemical networks : a promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol*, 17(2) :53–60, February 1999.

[198] S. Schuster, D.A. Fell, and T. Dandekar. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat Biotechnol*, 18(3) :326–332, March 2000.

[199] J.M. Schwartz, C. Gaugain, J.C. Nacher, A. de Daruvar, and M. Kanehisa. Observing metabolic functions at the genome scale. *Genome Biol*, 8(6) :R123, 2007.

[200] G. Scorletti. *Approche Unifiée de l'analyse et la commande des systèmes par formulation LMI*. PhD thesis, Université d'Orsay, Paris, France, 1997. In French.

[201] D. Segre, D. Vitkup, and G.M. Church. Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci USA*, 99(23) :15112–15117, November 2002.

[202] G. Seidel, M. Diel, N. Fuchsbauer, and W. Hillen. Quantitative interdependence of coeffectors, CcpA and cre in carbon catabolite regulation of *Bacillus subtilis*. *FEBS J*, 272(10) :2566–2577, May 2005.

[203] S.S. Shen-Orr, R. Milo, S. Mangan, and U. Alon. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat Genet*, 31(1) :64–68, May 2002.

[204] J.K. Shiau and J.H. Chow. Robust decentralized state feedback control design using an iterative linear matrix inequality algorithm. In *13th Triennal World Congres*, pages 203–208, July 1996.

[205] G. Shinar, R. Milo, M.R. Martinez, and U. Alon. Input output robustness in simple bacterial signaling systems. *Proc Natl Acad Sci USA*, 104(50) :19931–19935, December 2007.

[206] R.P. Shivers, S.S. Dineen, and A.L. Sonenshein. Positive regulation of *Bacillus subtilis* ackA by CodY and CcpA : establishing a potential hierarchy in carbon flow. *Mol Microbiol*, 62(3) :811–822, November 2006.

[207] R.P. Shivers and A.L. Sonenshein. Activation of the *Bacillus subtilis* global regulator CodY by direct interaction with branched-chain amino acids. *Mol Microbiol*, 53(2) :599–611, July 2004.

[208] R.P. Shivers and A.L. Sonenshein. *Bacillus subtilis* ilvB operon : an intersection of global regulons. *Mol Microbiol*, 56(6) :1549–1559, June 2005.

[209] K.D. Singh, M.H. Schmalisch, J. Stülke, and B. Görke. Carbon catabolite repression in *Bacillus subtilis* : quantitative analysis of repression exerted by different carbon sources. *J Bacteriol*, 190(21) :7275–7284, November 2008.

[210] S. Skogestad and M. Morari. Robust performance of decentralized control systems by independent design. 25(1) :119–125, 1989.

[211] W.K. Smits, C. Bongiorni, J.W. Veening, L.W. Hamoen, O.P. Kuipers, and M. Perego. Temporal separation of distinct differentiation pathways by a dual specificity Rap-Phr system in *Bacillus subtilis*. *Mol Microbiol*, 65(1) :103–120, July 2007.

[212] A.L. Sonenshein. Control of key metabolic intersections in *Bacillus subtilis*. *Nat Rev Microbiol*, 5(12) :917–927, 2007.

[213] E.D. Sontag. Asymptotic amplitudes and Cauchy gains : a small-gain principle and an application to inhibitory biological feedback. *Systems and Control Letters*, 47 :167–179, 2002.

[214] C. Soulé. Graphic requirements for multistationarity. *ComPlexUs*, 1 :123–133, 2003.

[215] J. Stelling, S. Klamt, K. Bettenbrock, S. Schuster, and E.D. Gilles. Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 420 :190–193, November 2002.

[216] J. Stelling, U. Sauer, Z. Szallasi, F.J. 3rd Doyle, and J. Doyle. Robustness of cellular functions. *Cell*, 118(6) :675–685, September 2004.

[217] G.N. Stephanopoulos, A.A. Aristidou, and J. Nielsen. *Metab Eng.* Academic Press, 1998.

[218] D. Sun and P. Setlow. Cloning and nucleotide sequence of the *Bacillus subtilis* ansR gene, which encodes a repressor of the ans operon coding for L-asparaginase and L-aspartase. *J Bacteriol*, 175(9) :2501–2506, May 1993.

[219] le T. Tam, H. Antelmann, C. Eymann, D. Albrecht, J. Bernhardt, and M. Hecker. Proteome signatures for stress and starvation in *Bacillus subtilis* as revealed by a 2-D gel image color coding approach. *Proteomics*, 6(16) :4565–4585, August 2006.

[220] A. Tanay, R. Sharan, M. Kupiec, and R. Shamir. Revealing modularity and organization in the yeast molecular network by integrated analysis of highly heterogeneous genomewide data. *Proc Natl Acad Sci USA*, 101(9) :2981–2986, March 2004.

[221] Cormen T.H., Leiserson C.E., Rivest R.L., and Stein C. *Introduction to algorithms.* MIT Press, 2001. (2nd edition).

[222] I. Thiele, N. Jamshidi, R.M. Fleming, and B.O. Palsson. Genome-scale reconstruction of *Escherichia coli*'s transcriptional and translational machinery : a knowledge base, its mathematical formulation, and its functional characterization. *PLoS Comput Biol*, 5(3) :e1000312, March 2009.

[223] R. Thomas. On the relation between the logical structrure of systems and their ability to generate multiple steady states or sustained oscillations. In J. Della-Dora, J. Demongeot and B. Lacolle, editors, *Numerical methods in the study of critical phenomena*, pages 180–193. Springer-Verlag, 1981.

[224] S. Tojo, T. Satomura, K. Morisaki, J. Deutscher, K. Hirooka, and Y. Fujita. Elaborate transcription regulation of the *Bacillus subtilis* ilv-leu operon involved in the biosynthesis of branched-chain amino acids through global regulators of CcpA, CodY and TnrA. *Mol Microbiol*, 56(6) :1560–1573, June 2005.

[225] L. Tournier and M. Chaves. Uncovering operational interactions in genetic networks using asynchronous Boolean dynamics. *J Theor Biol*, 260(2) :196–209, September 2009.

[226] R.J. Turner, Y. Lu, and R.L. Switzer. Regulation of the *Bacillus subtilis* pyrimidine biosynthesis (pyr) gene cluster by an autogenous transcriptional attenuation mechanism. *J Bacteriol*, 176(12) :3708–3722, June 1994.

[227] J.J. Tyson and H.G. Othmer. The dynamics of feedback control circuits in biochemical pathways. *J Theor Biol*, 5, 1978.

[228] S. Ujita and K. Kimura. Fructose-1,6-biphosphate aldolase from *Bacillus subtilis. Methods Enzymol*, 90(Pt 5) :235–241, November 1982.

[229] J.M. Van Dijl, G. Buist, M. Sibbald, J. Zweers, J.Y. Dubois, and H. Tjalsma. Ins and outs of the *Bacillus subtilis* membrane proteome. In P. Graumann, editor, *Bacillus : Cellular and Molecular Biology*, pages 331–374. Caister Academic Press, Norfolk, UK, 1st edition, 2007.

[230] N.D. Van Kampen. *Stochastic processes in physics and chemistry : third edition.* Elsevier, 2007.

[231] A. Varma and B.O. Palsson. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl Environ Microbiol*, 60(10) :3724–3731, October 1994.

[232] A. Vazquez, Q.K. Beg, M.A. Demenezes, J. Ernst, Z. Bar-Joseph, A.L. Barabasi, L.G. Boros, and Z.N. Oltvai. Impact of the solvent capacity constraint on *E. coli* metabolism. *BMC Syst Biol*, 2 :7, January 2008.

[233] M. Volkenstein. *Biophysique.* Edition Mir, 1985.

[234] W. Vongsangnak, P. Olsen, K. Hansen, S. Krogsgaard, and J. Nielsen. Improved annotation through genome-scale metabolic modeling of *Aspergillus oryzae. BMC Genomics*, 9 :245, May 2008.

[235] A. Wagner and D.A. Fell. The small world inside large metabolic networks. *Proceedings in Biological sciences*, 268(1478) :1803–1810, September 2001.

[236] M. Weng, P.L. Nagy, and H. Zalkin. Identification of the *Bacillus subtilis* pur operon repressor. *Proc Natl Acad Sci USA*, 92(16) :7455–7459, August 1995.

[237] A.M. Whatmore, J.A. Chudek, and R.H. Reed. The effects of osmotic upshock on the intracellular solute pools of *Bacillus subtilis*. *J Gen Microbiol*, 136(12) :2527–2535, December 1990.

[238] D.M. Wolf and F.H. Eeckman. On the relationship between genomic regulatory element organization and gene regulatory dynamics. *J Theor Biol*, 195 :167–186, November 1998.

[239] L.V. Jr. Wray, A.E. Ferson, K. Rohrer, and S.H. Fisher. TnrA, a transcription factor required for global nitrogen regulation in *Bacillus subtilis*. *Proc Natl Acad Sci USA*, 93(17) :8841–8845, August 1996.

[240] L.V. Jr. Wray and S.H. Fisher. Analysis of *Bacillus subtilis* hut operon expression indicates that histidine-dependent induction is mediated primarily by transcriptional antitermination and that amino acid repression is mediated by two mechanisms : regulation of transcription initiation and inhibition of histidine transport. *J Bacteriol*, 176(17) :5466–5473, September 1994.

[241] L.V. Jr. Wray and S.H. Fisher. A feedback-resistant mutant of *Bacillus subtilis* glutamine synthetase with pleiotropic defects in nitrogen-regulated gene expression. *J Biol Chem*, 280(39) :33298–33304, September 2005.

[242] L.V. Jr. Wray, F.K. Pettengill, and S.H. Fisher. Catabolite repression of the *Bacillus subtilis* hut operon requires a cis-acting site located downstream of the transcription initiation site. *J Bacteriol*, 176(7) :1894–1902, April 1994.

[243] L.V. Jr. Wray, J.M. Zalieckas, and S.H. Fisher. *Bacillus subtilis* glutamine synthetase controls gene expression through a protein-protein interaction with transcription factor TnrA. *Cell*, 107(4) :427–435, November 2001.

[244] M.K. Yeung, J. Tegner, and J.J. Collins. Reverse engineering gene networks using singular value decomposition and robust regression. *Proc Natl Acad Sci USA*, 99(9) :6163–6168, April 2002.

[245] A. Zaslaver, A.E. Mayo, R. Rosenberg, P. Bashkin, H. Sberro, M. Tsalyuk, M.G. Surette, and U. Alon. Just-in-time transcription program in metabolic pathways. *Nat Genet*, 36(5) :486–491, May 2004.

[246] Z. Zhang and M. Gerstein. Reconstructing genetic networks in yeast. *Nat Biotechnol*, 21(11) :1295–1297, November 2003.

Abréviations

Termes utilisés

Nom abrégé	Nom complet
DOR	Dense Overlapping Regulon
EDO	Equations différentielles ordinaires
FBA	Flux Balance Analysis
FBAwMC	méthode FBA alternative
rFBA	méthode FBA alternative
FFL	Feedforward loop
LP	Programmation linéaire
LFT	Linear fractional transformation
RBA	Resource Balance Analysis
SCPI	Structure de contrôle par le produit initial
SCPIb	SCPI combinée à une inhibition enzymatique de la première enzyme
SCPF	Structure de contrôle par le produit final
SIM	Single Input Module
SNC	Structure non contrôlée
SVD	Singular Value Decomposition

Glossaire des abréviations biologiques

Gènes et transcription

Nom abrégé	Nom complet	Type de composant cellulaire
ADN	Acide désoxyribonucléique	
ARN	Acide ribonucléique	issu de la transcription de l'ADN
ARNm	ARN messager	ARN codant pour une protéine
ARNr	ARN ribosomal	ARN composant un ribosome
ARNt	ARN de transfert	ARN nécessaire à la traduction d'un ARNm par les ribosomes
ARNtc	ARN de transfert chargé	ARNt chargé avec un acide aminé
RBS	Ribosome Binding Sequence	site de fixation du ribosome sur l'ARNm
A-box	riboswitch	régule la transcription des gènes par l'adénine
G-box	riboswitch	régule la transcription des gènes par la guanine
L-box	riboswitch	régule la transcription des gènes par la lysine
S-box	riboswitch	régule la transcription des gènes par le S-adénosyl-méthionine
T-box	mécanisme de régulation	régule la transcription des gènes par un ARNt
SR1	ARN non codant	régule la traduction de certains gènes
P-lead	peptide leader	régule la transcription de certains gènes

Protéines mentionnées

Facteurs sigma

Facteur sigma	Gène codant	Fonction dans la cellule
SigA (σ_A)	*sigA*	croissance exponentielle
SigB (σ_B)	*sigB*	réponse au stress
SigD (σ_D)	*sigD*	mobilité
SigE (σ_E)	*sigE*	sporulation
SigF (σ_F)	*sigF*	sporulation
SigG (σ_G)	*sigG*	sporulation
SigH (σ_H)	*sigH*	sporulation
SigK (σ_K)	*sigK*	sporulation

Facteurs de transcription

Facteur de transcription	Gène codant	Fonction dans la cellule
AbrB	*abrB*	transition vers la phase stationnaire
AnsR	*ansR*	dégradation de l'asparagine
ArfM	*arfM*	fermentation
CcpA	*ccpA*	gestion des sources de carbone
CcpB	*ccpB*	gestion des sources de carbone
CcpC	*ccpC*	cycle de Krebs
CcpN	*ccpN*	gluconéogenèse
CggR	*cggR*	glycolyse
CodY	*codY*	régulateur global
ComK	*comK*	compétence
CRP	*crp*	gestion des sources de carbone
CymR	*cymR*	synthèse de la cystéine
CysL	*cysL*	dégradation du sulfite
DegU	*degU*	transition vers la phase stationnaire
FNR	*fnr*	respiration anaérobie
FruR	*fruR*	transport et assimilation du fructose
Fur	*fur*	gestion du fer
GerE	*gerE*	germination
GlnR	*glnR*	synthèse de la glutamine
PhoP	*phoP*	gestion du phosphate
PurR	*purR*	synthèse des nucléotides
RbsR	*rbsR*	transport et assimilation du ribose
ResD	*resD*	respiration aérobie
Rok	*rok*	formation des biofilms
RsfA	*rsfA*	sporulation
Spo0A	*spo0A*	sporulation
SpoIIID	*spoIIID*	sporulation
SpoVT	*spoVT*	sporulation
Spx	*spx*	gestion des ponts disulfides
TnrA	*tnrA*	gestion des sources d'azote
TrpR	*trpR*	gestion du tryptophane
TyrR	*tyrR*	gestion de la tyrosine
YtlI	*ytlI*	transport de la cystine
YtrA	*ytrA*	transport de l'acétoine
YycF	*yycF*	gestion de la paroi

Autres protéines

Nom	Gène codant	Type	Fonction dans la cellule
ARNt synthase	-	protéine	charge un ARNt avec un acide aminé
GTPase	-	protéine	hydrolyse un GTP pour catalyser sa réaction
FT	-	facteur de transcription	régule la transcription des gènes
ABC	-	transporteur	3 sous-unités (ABC) et ATP dépendant
PTS	-	transporteur	phosphotransferase transport system
AckA	*ackA*	enzyme	synthèse de l'acétate
AlsS	*alsS*	enzyme	synthèse de l'acétoine
ArgI	*argI*	enzyme	dégradation de l'arginine
AroA	*aroA*	enzyme	synthèse du chorismate
AspB	*aspB*	enzyme	synthèse de l'aspartate
CarAB	*carA-carB*	complexe enzymatique	synthèse de l'arginine
CitB	*citB*	enzyme	cycle de Krebs
ClpCP	*clpC-clpP*	protéase	dégradation des protéines
FruA	*fruA*	transporteur	transport du fructose
EF-G	*fusA*	facteur d'élongation	traduction des protéines
EF-Tu	*tuf*	facteur d'élongation	traduction des protéines
GapA	*glnA*	enzyme	glycolyse
GlnA	*glnA*	enzyme	synthèse de la glutamine
GlyA	*glyA*	enzyme	synthèse de la glycine
GuaB	*guaB*	enzyme	synthèse du GMP
IlvA	*ilvA*	enzyme	synthèse de l'isoleucine
IlvBH	*ilvB-ilvH*	complexe enzymatique	synthèse de l'isoleucine, de la valine, de la leucine
HPr	*ptsH*	protéine	transporteur des sucres PTS
HPr-Ser46-P	*ptsH*	protéine	forme phosphorylée de HPr sur la sérine 46
Mdh	*mdh*	enzyme	cycle de Krebs
MccB	*mccB*	enzyme	synthèse de la cystéine
MetA	*metA*	enzyme	synthèse de la méthionine
MetI	*metI*	enzyme	synthèse de la méthionine
ProB	*proB*	enzyme	synthèse de la proline
ProG	*proG*	enzyme	synthèse de la proline
ProH	*proH*	enzyme	synthèse de la proline
ProI	*proI*	enzyme	synthèse de la proline
ProJ	*proJ*	enzyme	synthèse de la proline
Pta	*pta*	enzyme	synthèse de l'acétate
PyrAAB	*pyrAA-pyrAB*	complexe enzymatique	synthèse des pyrimidines
RtpA	*rtpA*	protéine/régulateur	synthèse du tryptophane
TcyABC	*tcyA-tcyB-tcyC*	transporteur	transport de la cystine
TcyJKLMN	*tcyJ-tcyK-tcyL* *tcyM-tcyN*	transporteur	transport de la cystine
ThrB	*thrB*	enzyme	synthèse de la thréonine
ThrC	*thrC*	enzyme	synthèse de la thréonine
TRAP	*mtrB*	protéine/régulateur	synthèse du tryptophane
Zwf	*zwf*	enzyme	voie des pentoses phosphates

Métabolites mentionnés

Nom	Type de métabolite	Fonction dans la cellule
dATP	deoxynucléotide	synthèse de l'ADN
dCTP	deoxynucléotide	synthèse de l'ADN
dGTP	deoxynucléotide	synthèse de l'ADN
dTTP	deoxynucléotide	synthèse de l'ADN
ATP	nucléotide	synthèse des ARNs, source d'énergie
ADP	nucléotide	précurseur de l'ATP, métabolite recyclé
AMP	nucléotide	précurseur de l'ADP, métabolite recyclé
CTP	nucléotide	synthèse des ARNs
GTP	nucléotide	synthèse des ARNs, source d'énergie
GDP	nucléotide	précurseur du GTP, métabolite recyclé
GMP	nucléotide	précurseur du GDP, métabolite recyclé
UTP	nucléotide	synthèse des ARNs
NADH/NAD	état redox	source d'énergie pour la synthèse d'ATP par respiration
NADPH/NADP	état redox	source d'énergie utilisée dans l'anabolisme
alanine	acide aminé	synthèse des protéines
arginine	acide aminé	synthèse des protéines
asparagine	acide aminé	synthèse des protéines
aspartate	acide aminé	synthèse des protéines
cystéine	acide aminé	synthèse des protéines
glutamine	acide aminé	synthèse des protéines
glutamate	acide aminé	synthèse des protéines
glycine	acide aminé	synthèse des protéines
histidine	acide aminé	synthèse des protéines
isoleucine	acide aminé	synthèse des protéines
leucine	acide aminé	synthèse des protéines
lysine	acide aminé	synthèse des protéines
méthionine	acide aminé	synthèse des protéines
phénylalanine	acide aminé	synthèse des protéines
proline	acide aminé	synthèse des protéines
sérine	acide aminé	synthèse des protéines
thréonine	acide aminé	synthèse des protéines
tryptophane	acide aminé	synthèse des protéines
tyrosine	acide aminé	synthèse des protéines
valine	acide aminé	synthèse des protéines
chorismate	métabolite	précurseur du tryptophane, de la phénylalanine et de la tyrosine
ornithine	métabolite	synthèse de l'arginine

Nom	Type de métabolite	Fonction dans la cellule
arabinose	chaîne carbonée	source de carbone glycolytique
cellobiose	chaîne carbonée	source de carbone glycolytique
fructose	chaîne carbonée	source de carbone glycolytique
gluconate	chaîne carbonée	source de carbone glycolytique
glucosamine	chaîne carbonée	source de carbone glycolytique
glucose	chaîne carbonée	source de carbone glycolytique
glycérol	chaîne carbonée	source de carbone glycolytique
inositol	chaîne carbonée	source de carbone glycolytique
lichénan	chaîne carbonée	source de carbone glycolytique
maltose	chaîne carbonée	source de carbone glycolytique
mannitol	chaîne carbonée	source de carbone glycolytique
mannose	chaîne carbonée	source de carbone glycolytique
n-acétylglucosamine	chaîne carbonée	source de carbone glycolytique
ribose	chaîne carbonée	source de carbone glycolytique
sorbitol	chaîne carbonée	source de carbone glycolytique
sucrose	chaîne carbonée	source de carbone glycolytique
tréhalose	chaîne carbonée	source de carbone glycolytique
acétate	chaîne carbonée	source de carbone néoglucogénique, produit excrété
citrate	chaîne carbonée	source de carbone néoglucogénique, cycle de Krebs
fumarate	chaîne carbonée	source de carbone néoglucogénique, cycle de Krebs
lactate	chaîne carbonée	source de carbone néoglucogénique, fermentation
malate	chaîne carbonée	source de carbone néoglucogénique, cycle de Krebs
pyruvate	chaîne carbonée	source de carbone néoglucogénique, glycolyse
succinate	chaîne carbonée	source de carbone néoglucogénique, cycle de Krebs
2-oxoglutarate	chaîne carbonée	source de carbone néoglucogénique, cycle de Krebs
AKB	alpha-kéto-butyrate	synthèse de l'isoleucine
ACCOA	acétyl-coA	cycle de Krebs
CoA	co-enzyme A	co-facteur
FBP	fructose-1,6-biphosphate	glycolyse
formamide	métabolite	dégradation de l'histidine
glycéraldéhyde-3P	métabolite	glycolyse
oxaloacétate	métabolite	cycle de Krebs
3-phosphoglycérate	métabolite	glycolyse
2-phosphoglycérate	métabolite	glycolyse
6-phosphogluconate	métabolite	voie des pentoses phosphates
NH3	ammonium	source d'azote
$NH4^+$	azote assimilable	source d'azote
H2O	eau	
CO2	dioxyde de carbone	
sulfate	ion	source de soufre
P_i	phosphate	source de phosphate
PP_i	diphosphate	source de phosphate
H^+	proton	transport des composés, pH
Fe^{2+}	fer	transfert d'électrons
Mn^{2+}	manganèse	transfert d'électrons
Mg^{2+}	magnésium	transfert d'électrons
Zn^{2+}	zinc	transfert d'électrons

Composants cellulaires

Nom	Type de composant cellulaire
peptidoglycane	paroi cellulaire
acides teichoïques	paroi cellulaire
acides teichuroniques	paroi cellulaire
acides gras	précurseurs de tous les lipides
monoglucosyldiacylglycérol	glycolypide composant la membrane cellulaire
diglucosyldiacylglycérol	glycolypide composant la membrane cellulaire
triglucosyldiacylglycérol	glycolypide composant la membrane cellulaire
cardiolipine	phospholipide composant la membrane cellulaire
phosphatidylglycérol	phospholipide composant la membrane cellulaire
lysylphosphatidylglycérol	phospholipide composant la membrane cellulaire
phosphatidyléthanolamine	phospholipide composant la membrane cellulaire

Table des figures

1 Un processus biologique représenté comme un module 3
2 Les processus cellulaires majeurs de la cellule en phase de croissance exponentielle 7

1.1 Un réseau de régulation où le gène A active le gène B, qui réprime le gène C, qui active le gène A 13
1.2 Les motifs récurrents dans le réseau de régulation génétique d'*E. coli* [203] 15
1.3 La *feedforward loop* [140] 15
1.4 Le parallèle en Automatique 16
1.5 Structure multi-couche hiérarchique du réseau de régulation génétique d'*E.coli* (extrait de [132]) 17
1.6 Organisation modulaire et fonctionnelle du réseau de régulation génétique d'*E.coli* (extrait de [143]) 18
1.7 Exemples de boucles négatives et positives dans le réseau de régulation génétique 19
1.8 Les boucles identifiées dans le réseau de régulation de *Bacillus subtilis* . . 20
1.9 Existence d'une boucle dans la FFL si les liens métaboliques sont intégrés 21
1.10 Réseau de type *scale-free* (gauche) et Hiérarchie modulaire d'un réseau *scale-free* (droite) 23
1.11 Un exemple de sous-système défini classiquement en Automatique 25

2.1 Cinq niveaux de contrôle dans le réseau de régulation génétique d'*E.coli* (extrait de [132]) 35
2.2 Une organisation hiérarchique du réseau de régulation génétique du réseau métabolique de *Bacillus subtilis*. Le code couleur correspond au type d'effecteur du régulateur : bleu (métabolite), violet (transporteur de sucre), rouge (ion), noir (rien), marron (inconnu). 35
2.3 Deux structures de contrôle : par le produit final, ou par le premier métabolite. Les enzymes (resp. métabolites) sont en rose (resp. jaune), et le régulateur est représenté par une ellipse. Les flèches en orange désignent le contrôle par les métabolites (effecteur du régulateur ou modulation de l'activité de l'enzyme). 37
2.4 Deux niveaux de contrôle pour une même voie métabolique 38
2.5 Voie métabolique contrôlée par le produit final 39

2.6 Intersection des courbes $E_n f_n(\bar{x}_n)$ et $f_1(\bar{x}_1, \bar{x}_2, \bar{x}_n)$ pour tout $\bar{x}_2 \in [0, x_2^*]$ (figure de gauche) et des courbes $E_n f_n(H_n(\bar{x}_2))$ et $\alpha_2 \frac{g(H_n(\bar{x}_2))}{\mu} f_2(\bar{x}_2, \bar{x}_3)$ (figure de droite) 52
2.7 Présence d'une isoenzyme dans la voie métabolique 54
2.8 Structure de contrôle par le produit initial 58
2.9 La structure de contrôle par le produit initial associée à l'inhibition enzymatique de $\mathbb{E}_0$ par le produit final x_n 61
2.10 La représentation entrée/sortie d'un module 64
2.11 Connexion en série de deux modules 64
2.12 Connexion en parallèle de deux modules 66
2.13 Connexion entre deux modules SCPF 69
2.14 Représentation modulaire de la figure 2.13 71
2.15 Connexion entre un module SCPI et un module SCPF 72
2.16 Connexion entre un module SCPF et un module SCPI 75
2.17 Connexion entre un module SCPI et un module SCPI 76
2.18 Connexion entre un module SCPIb et un module SCPIb 78

3.1 Une vue modulaire du réseau métabolique de *B. subtilis* 85
3.2 La voie de synthèse de la lysine et ses régulations enzymatiques connues . 92
3.3 La régulation de la voie de la lysine 93
3.4 La glycolyse et ses régulations enzymatiques connues 99
3.5 La régulation de la glycolyse 100
3.6 Les régulations enzymatiques du cycle de Krebs 103
3.7 Les modules élémentaires impliquant l'aspartate 108
3.8 Les modules élémentaires impliquant le glutamate et la glutamine 109
3.9 Une représentation systémique et simplifiée de la traduction des protéines 111

4.1 Une représentation systémique de la cellule 123
4.2 La représentation entrée/sortie de chaque processus cellulaire 124
4.3 Exemple de deux ensembles $\mathcal{C}$ convexe (en haut à gauche) et non convexe (en haut à droite) et de deux fonctions $h(x)$ convexe (en bas à gauche) et non convexe (en bas à droite) 137
4.4 Exemple de voies métaboliques linéaires mp_1 et mp_2 142

5.1 Répartition des protéines entre les ribosomes, les voies métaboliques et toutes les protéines non ribosomales ($\mathcal{P}_M$ et $\mathcal{P}_G$) 170
5.2 Evolution des ribosomes et des protéines métaboliques avec le taux de croissance pour *E. coli* (extrait de [142]) 171
5.3 Croissance de la souche sauvage et de la souche *PrrnO-lacZ* en milieu riche et en milieu minimum 172
5.4 Répartition des protéines prédite entre les ribosomes et les voies métaboliques 173
5.5 Evolution des voies de synthèse *de novo* des acides aminés : bleu (réprimé), rouge (fort) 174

5.6 Evolution du transport des acides aminés :bleu (réprimé), rouge (fort) . . 175
5.7 La sérine comme précurseur de la cystéine, la glycine, le tryptophane et la méthionine 177
5.8 Evolution de la cystéine, la glycine, le tryptophane et la méthionine au regard de la sérine 177
5.9 Deux voies alternatives de synthèse de l'AKB 181
5.10 La voie de synthèse de la cystéine 185
5.11 Substitution entre les deux transporteurs haute (TcyJKLMN) et basse affinité (TcyABC) de cystine calculée avec un seuil de 10^{-3} pour la dichotomie 187
5.12 Substitution entre les deux transporteurs haute (TcyJKLMN) et basse affinité (TcyABC) de cystine calculée avec un seuil de 10^{-10} pour la dichotomie 187

A.1 Les différents modes de vie de *Bacillus subtilis* au cours du temps 198
A.2 Les processus cellulaires majeurs de la cellule en phase de croissance exponentielle 199
A.3 Code génétique 201
A.4 Régulation de la transcription des gènes par les facteurs de transcription et les facteurs sigma 202

Liste des tableaux

2.1 Intégration de l'information. Les 21 autres mécanismes de régulation incluent RtpA, S-box, L-box, A-box, G-box, Plead, SR1 et 14 T-boxes. . . 33
2.2 Classification des régulations génétiques en fonction de la nature de leur effecteur métabolique 33
2.3 Règles d'interconnexion en série entre plusieurs modules. On suppose que (i) le flux d'entrée ν_1 (ou ν_0) est capable de maintenir le métabolite initial x_1 (ou x_0) fixé ; (ii) les enzymes des modules ne saturent pas. 67
2.4 Règles d'interconnexion en parallèle entre plusieurs modules. On suppose que (i) ν_0 est capable de maintenir le métabolite initial x_1 (ou x_0) fixé ; (ii) les enzymes des modules ne saturent pas. 68

3.1 Isoenzymes dans la voie centrale des carbones, les nucléotides et les acides gras 89
3.2 Isoenzymes dans les voies des acides aminés 90
3.3 Notations pour les métabolites, les co-métabolites et les enzymes dans le système (3.1) 95
3.4 Prédiction des variations des composants clés de la voie de la lysine . . . 97
3.5 Résultats expérimentaux pour les éléments clés de la voie de la lysine pour les trois conditions (Glucose, Malate et Malate+Glucose). L'unité μmol/gdwc signifie micromoles par gramme de poids sec. 97
3.6 Comparaison de l'évolution des différents composants entre les conditions en glucose versus malate (glc/mal), malate versus malate+glucose (mal/mal+glc) et en glucose versus malate+glucose (glc/mal+glc). L'évolution de l'ATP est basée sur [26]. Sur ce tableau, la notation «0» signifie que la variation relative du composant reste inférieure à 5%. 97
3.7 Prédiction des variations des composants clés de la voie basse de la glycolyse. Flux_{in} correspond au flux de synthèse du FBP provenant du haut de la glycolyse. 101
3.8 Résultats expérimentaux pour les clés de la voie de la glycolyse pour les trois conditions (Glucose, Malate et Malate+Glucose) 102
3.9 Comparaison de l'évolution des différents composants entre les conditions en glucose versus malate (glc/mal), malate versus malate+glucose (mal/mal+glc) et en glucose versus malate+glucose (glc/mal+glc). . . . 102

3.10 Prédiction de l'évolution des différents composants en présence de malate, c'est-à-dire entre les conditions en glucose versus malate (glc/mal) et glucose versus malate+glucose (glc/mal+glc). 105
3.11 Concentration de métabolites d'intérêt pour les trois conditions en μmol/gdwc, et leur évolution en glucose versus malate (glc/mal) et en glucose versus malate+glucose (glc/mal+glc). Sur ce tableau, la notation «0» signifie que la variation relative du composant reste inférieure à 5%. 106
3.12 Variation de la transcription des gènes impliqués dans (a) la synthèse de l'arginine et contrôlés par AhrC ; (b) la synthèse de la proline et régulés par une T-box spécifique proline ; (c) la synthèse du tryptophane régulés par TRAP. 110
3.13 Décomposition du réseau métabolique de *Bacillus subtilis* en modules élémentaires 116
3.14 Décomposition du réseau métabolique de *Bacillus subtilis* en modules élémentaires (suite) 117
3.15 Décomposition du réseau métabolique de *Bacillus subtilis* en modules élémentaires (suite) 118

5.1 Composition moyenne en acides aminés des protéines de *Bacillus subtilis* [8] 165
5.2 Cœfficients de $P_f^{lp}(\mu)$ (et $P_f(\mu)$) 166
5.3 Concentration des macro-composants $\bar{X}_c$ en micromoles par gramme de poids sec (μmol/gdwc) intégrés dans $P_f(\mu)$ pour $\mu = 0.1h^{-1}$ [160] 167