



**HAL**  
open science

## Structure de réseaux biologiques : rôle des nœuds internes vis-à-vis de la production de composés

Julie Laniau

► **To cite this version:**

Julie Laniau. Structure de réseaux biologiques : rôle des nœuds internes vis-à-vis de la production de composés. Recherche opérationnelle [math.OC]. Inria Rennes - Bretagne Atlantique, 2017. Français. NNT: . tel-01656474

**HAL Id: tel-01656474**

**<https://hal.science/tel-01656474v1>**

Submitted on 5 Dec 2017

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

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

**THÈSE / UNIVERSITÉ DE RENNES 1**  
*sous le sceau de l'Université Bretagne Loire*

pour le grade de  
**DOCTEUR DE L'UNIVERSITÉ DE RENNES 1**

*Mention : Informatique*

**Ecole doctorale Matisse**

présentée par

**Julie LANIAU**

préparée à l'unité de recherche IRISA - UMR6074  
Dyliss (DYNAMics, Logics and Inference for biological Systems and  
Sequences)  
IRISA (Institut de Recherche en Informatique et Système  
Aléatoire)

---

**Structure de réseaux  
biologiques :  
rôle des nœuds internes  
vis-à-vis de la  
production de composés**

**Thèse soutenue à Rennes  
le 23 octobre 2017**

devant le jury composé de :

**Philippe DAGUE**

Professeur, Université Paris-Sud / rapporteur

**David SHERMAN**

Directeur de recherche, INRIA, Bordeaux 1 / rapporteur

**Rumen ANDONOV**

Professeur, Rennes 1 / examinateur

**Bruno SAINT-JEAN**

Chercheur, Ifremer Nantes / examinateur

**Anne SIEGEL**

Directrice de recherche, IRISA - CNRS, Rennes 1 /  
directeur de thèse

**Damien EVEILLARD**

Enseignant-chercheur, Nantes / co-directeur de thèse



Il y a trois stades de connaissance : ce qu'on croit, ce qu'on sait et ce qu'on peut prouver.

Straczynski, Joe Michaël



## Remerciements

Je tiens à remercier avant toute chose mes deux directeurs de thèse Damien Eveillard et Anne Siegel pour leur accompagnement tout au long de cette thèse malgré les rebondissements. Merci de m'avoir fait confiance et pour votre soutien professionnel, scientifique et personnel qui m'a permis de vivre ces quatre années agréables bien que pleines en émotions et très enrichissantes.

Je voudrais également remercier les personnes ayant accepté la charge de rapporteur de cette thèse, Philippe Dague, David Sherman. Merci également à Rumen Andonov et Bruno Saint-Jean pour avoir accepté d'être dans le jury de ce travail.

Pour continuer, je tiens à remercier l'ensemble des membres, qu'ils soient anciens ou actuels, des trois équipes Dyliss, GenScale et de la plateforme GenOuest pour l'ambiance studieuse mais détendue de travail m'ayant permise de travailler dans des conditions excellentes. Merci pour les discussions et l'aide scientifique qui m'ont aidée à mieux comprendre la biologie et la symbiose des organismes. Je voudrais aussi remercier les personnes du projet *Algae in silico* avec lesquelles ce fut un plaisir de discuter pour faire avancer la science et découvrir plein de domaines en interaction les uns avec les autres dans le même but, et particulièrement Caroline Baroukh qui m'a apporté son expertise biologique tout le long de ma thèse.

Mes remerciements s'adressent également à ceux qui m'ont aidée dans mon travail mais aussi à m'aérer l'esprit. Merci d'abord à Mathilde, Gaëlle, Vincent et Sylvain, les vieux doctorants de l'équipe Symbiose, pour l'accueil et l'ambiance amicale parmi les Symbiotes et pour la découverte du patrimoine rennais puis français. Je voudrais ensuite remercier mes collègues de bureau d'avoir su supporter ma geek-titude : ceux du premier bureau, Aymeric et Charles pour les discussions silencieuses mais animées, et ceux du second bureau, Camille et Patrick, pour celles plus vocales mais tout aussi sympathiques. Merci aussi à tous ceux avec qui j'ai passé de bons moments que ce soit au cinéma, à la piscine ou à la cuisine, tous vous énumérer serait long mais j'espère que vous vous reconnaîtrez. Enfin merci à mes amis de Nantes et d'Angers pour leur soutien et leurs encouragements durant ces 4 années d'exil à Rennes.

Je ne peux terminer mes remerciements sans penser à ma famille, merci pour votre soutien et votre patience tout au long de ces quatre années.

Et enfin, merci à l'ensemble des personnes auxquelles je ne pense pas en ce moment mais qui m'ont apporté leur aide et ont permis de faire de ces années de thèse une expérience enrichissante et inoubliable.



# Table des matières

<b>Introduction</b>	<b>9</b>
<b>1 Modélisation des réseaux métaboliques</b>	<b>13</b>
1.1 Représentation des réseaux métaboliques	13
1.2 Protocole de reconstruction	15
1.3 Les réseaux métaboliques existants	19
1.4 Six exemples caractéristiques	22
<b>2 Un concept de productibilité au sein d'un graphe métabolique</b>	<b>27</b>
2.1 Des flux dans le réseau	27
2.2 Flux Balance Analysis	28
2.3 Flux Variability Analysis	30
2.4 Productibilité stoechiométrique de composés métaboliques	32
2.5 Productibilité topologique de composés métaboliques	34
2.6 Complémentarité des études de flux et topologique	37
<b>3 Des éléments particuliers dans un réseau métabolique</b>	<b>43</b>
3.1 Des gènes essentiels	43
3.2 Des réactions essentielles	44
3.3 Des ensembles essentiels de réactions	45
3.4 Des métabolites essentiels	47
<b>4 Raffiner et étendre le concept de métabolite essentiel</b>	<b>49</b>
4.1 Métabolites essentiels du point de vue de la productibilité	49
4.2 Métabolites essentiels du point de vue de l'efficacité	52
4.3 Métabolites essentiels du point de vue de la persistance	53
4.4 Construction de classes de composés métaboliques en fonction du critère d'essentialité vis-à-vis de la persistance	58
<b>5 Analyses des métabolites essentiels dans des réseaux à échelle génomique</b>	<b>63</b>
5.1 Etude quantitative	63
5.2 Identification de redondances dans le réseau	68
5.3 Impact de l'initiation des cycles métaboliques	69
5.4 Identification de co-produits critiques	74
<b>6 Étude des classes de métabolites ME-persistance</b>	<b>79</b>
6.1 Propriétés des classes de six réseaux à échelle génomique	79
6.2 Etude des clusters d'un ensemble de 3600 réseaux métaboliques	84
<b>Conclusion et perspectives</b>	<b>91</b>
<b>A Annexe : Différentes approches de programmation par contraintes, appliquées à l'exemple du gapfilling</b>	<b>95</b>
A.1 Gapfilling avec programmation linéaire par contraintes	95
A.2 Gapfilling avec programmation logique	98



<b>B Annexe : Données supplémentaires des résultats obtenus sur la clusterisation des six cas de réseaux métaboliques étudiés</b>	<b>100</b>
B.1 Données supplémentaire des figures représentant les six cas étudiés de réseaux métaboliques . . . . .	100
B.2 Clusters d'intérêt dans les six cas étudiés de réseaux métaboliques . . . . .	102
<b>Bibliography</b>	<b>112</b>

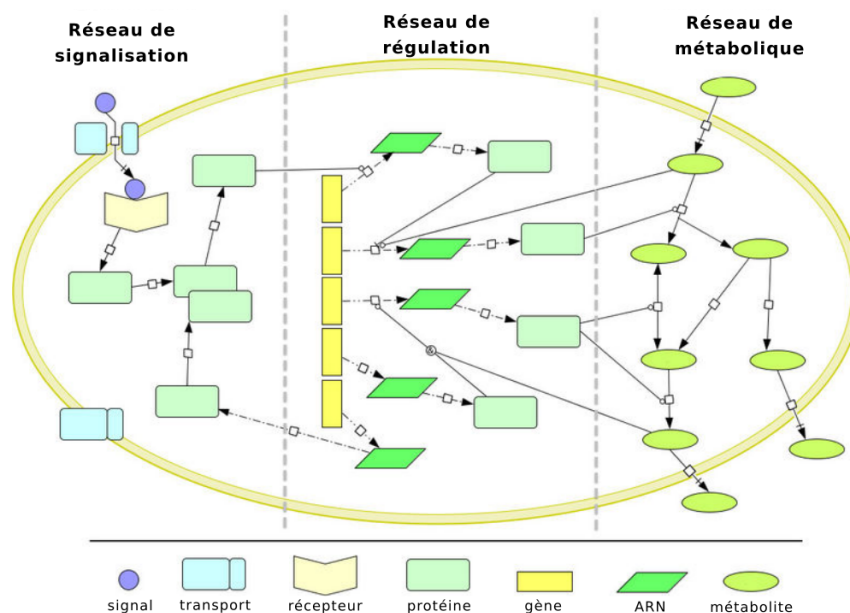
# Introduction : biologie des systèmes et modélisation des mécanismes moléculaires

L'étude du vivant nécessite depuis toujours des interactions entre scientifiques de divers domaines, à la fois pour faire progresser les technologies d'observation et de mesure, et pour permettre l'interprétation des données et leur intégration dans des modèles de connaissance. Ainsi fut créé le réseau européen ERASysBio mettant en relation différentes agences venant de treize pays afin de promouvoir le développement coordonné de la biologie des systèmes. Comme le présente ERASysBio, "*La biologie des systèmes a pour but de comprendre les interactions dynamiques entre composants d'un système vivant, et entre systèmes vivants en interaction avec l'environnement. La biologie systémique est une approche par laquelle des questions biologiques sont abordées en intégrant les expérimentations avec la modélisation, la simulation et la théorie dans des cycles itératifs. La modélisation n'est pas le but final mais un outil pour faire avancer la compréhension du système, développer des expérimentations plus ciblées et finalement aboutir à des prédictions.*".

Les prémisses de la biologie des systèmes trouvent leur origine dans les travaux de Jacob et Monod pour modéliser, étudier et comprendre le fonctionnement de l'opéron lactose et les mécanismes géniques [44]. Cependant, le développement de la biologie des systèmes a véritablement commencé au XX<sup>ème</sup> siècle avec les avancées technologiques relatives liées à la génomique et la "post-génomique", c'est-à-dire le développement des techniques de mesure à l'échelle d'une cellule entière, des transcrits, flux métaboliques, ou activités protéiques [80, 53]. Ainsi, pour les espèces les plus étudiées (dites "modèles") comme *Escherichia coli* chez les procaryotes ou l'homme chez les eucaryotes, de nombreuses données sont désormais accessibles : génome, transcriptome, protéome et métabolome [66]. Ces dix dernières années, avec la démocratisation de nouvelles technologies de séquençage ("NGS"), ces informations sont aussi disponibles pour de nombreux organismes "non-modèles", aux propriétés biologiques spécifiques, amplifiant les besoins en intégration et en modélisation.

La biologie des systèmes vise l'intégration de l'ensemble des connaissances biologiques en utilisant des approches automatiques (combinant mathématique et informatique) ayant pour but la modélisation *in silico* de la réponse d'un organisme à des perturbations, qu'elles soient de nature génétique ou environnementale. La modélisation vise ainsi l'étude du comportement de la cellule, des organes ou de l'organisme entier. Selon un schéma proposé par Hiroaki Kitano [58] qui ne s'est pas démenti depuis 15 ans, l'étude des systèmes biologiques repose sur quatre points essentiels : il s'agit de commencer par la conception du système étudié, puis de procéder à l'analyse de sa structure, suivie par celle de sa dynamique et enfin à l'analyse du contrôle du système. Chacun de ces points est nécessaire à une compréhension fine du comportement de l'organisme en réponse à des stimuli du milieu environnemental ou à des perturbations génétiques. Ainsi la conception et l'analyse structurelle du système permettent non seulement de représenter la connaissance biologique sur l'organisme étudié mais déjà de mettre en avant de nouvelles connaissances. En analysant la dynamique du système et son contrôle, il devient possible de prédire les actions du système pour répondre à une situation donnée.

La conception d'un système biologique s'appuie sur une représentation des mécanismes moléculaires sous la forme d'un "réseau" pour décrire les interactions entre les différentes entités d'une



**FIGURE 1 – Représentation des trois principaux types de réseaux existant en biologie des système** Les mécanismes moléculaires au sein d'un organisme peuvent être représentés sous la forme de réseaux, de trois types différents, qui peuvent cependant interagir. Un organisme reçoit des signaux de l'environnement, déclenchant une cascade de signaux internes pouvant être représentée par un réseau de signalisation. Ces signaux provoquent des réponses au niveau de la régulation des gènes pouvant être modélisée par un réseau de régulation génique ou transcriptionnelle qui vise à réguler la production d'ARN puis de protéines. Les protéines ainsi obtenues peuvent catalyser des réactions métaboliques. Ces dernières permettent la transformation de molécules (en particulier des nutriments) en d'autres et en énergie, modélisée au sein de réseaux métaboliques.

cellule (ADN, ARN, protéines, ...). On distingue trois grands types d'interactions biologiques, comme illustré dans la Figure 1 [71].

- \* Les réseaux de signalisation [47] représentent les voies de signalisation présentes dans un organisme. Elles permettent de modéliser les associations et dissociations entre des molécules, formant ainsi des complexes qui jouent un rôle de transmission de signal entre l'environnement et les molécules centrales de l'organisme. Les chaînes de formation de complexe ou de modification moléculaire (phosphorylation...) permettent au final de déclencher une réaction de l'organisme appropriée au stimulus ayant causé la création du signal initial.
- \* Les réseaux de régulation génétique [18] modélisent les interactions liées à la régulation entre des gènes, des protéines et/ou de petites molécules et permettent à l'organisme de réguler la production de protéines en fonction de stimuli. Ainsi selon le besoin, la production de protéines ciblées augmente ou diminue afin d'adapter au mieux la réponse de l'organisme.
- \* Les réseaux métaboliques représentent l'ensemble des transformations chimiques de molécules en d'autres molécules se produisant au sein d'un organisme, sous le contrôle d'enzymes (protéines codées par des gènes). Ce type de réseau permet d'étudier la chimie au sein d'un organisme, d'explorer la capacité de ce dernier à utiliser les ressources présentes dans le milieu naturel pour croître et produire les éléments nécessaires à sa survie, en particulier d'un point de vue énergétique [41, 72].

Ainsi, la conception d'un réseau consiste principalement à représenter et à structurer les connaissances biologiques, souvent sous la forme d'un graphe. Ce dernier peut être comparé à un réseau routier dont les villes ou les molécules sont les nœuds et les arcs, les routes ou les interactions. Suivant le schéma de Kitano, la phase suivante de l'analyse ("analyse structurelle") s'appuie sur des méthodes de théorie des graphes qui sont utilisées pour étudier la structure de l'organisme étudié. De nombreuses analyses de graphes biologiques, qu'ils décrivent des interactions métabo-

liques, géniques ou de signalisation, ont ainsi montré qu'ils contiennent des motifs et des structures caractéristiques. Il est ainsi démontré dans les travaux de H. Jeong et al [45] que les réseaux métaboliques semblent suivre une structure dite de réseau "petit monde". De même, U. Alon [4] a mis en évidence des motifs récurrents dans les réseaux de régulation génique[46].

Pour comprendre le rôle de ces motifs et plus généralement pour modéliser le comportement dynamique du système, une dynamique peut ensuite être ajoutée au graphe. Cette dynamique permet de représenter le comportement de l'organisme de la même façon qu'une carte routière peut-être enrichie par des informations sur le trafic routier. La dynamique permet de modéliser les flux de matière existant entre les différentes entités du graphe, l'activité des interactions pouvant être très variable. L'étude de la dynamique au sein du graphe permet ainsi de comprendre les interactions au cours du temps entre les différents acteurs du réseau. On pourra ainsi prédire les quantités de flux alimentant la production et la consommation des éléments du réseau et identifier les corrélations entre les évolutions des molécules, malgré parfois leur distance dans le graphe. On pourra aussi comprendre l'impact de modifications, de mutations sur la structure et la dynamique de l'organisme. Enfin, ces approches dynamiques permettent aussi de mettre en lumière des points de contrôle et de régulation des flux, en particulier pour optimiser le comportement de l'organisme étudié dans une situation donnée.

L'étude de l'ensemble de la dynamique du graphe peut se réaliser à travers différentes méthodes, selon le point de vue voulu et les données disponibles, pouvant être quantitatives, qualitatives voire stochastiques. Au niveau quantitatif, par analogie avec les systèmes physiques, les méthodes s'appuient principalement sur des modélisations par des équations différentielles ordinaires et s'appuient sur l'identification de paramètres cinétiques. L'identification des paramètres dépend fortement des données disponibles sur le système et de sa taille. Différentes stratégies [76] ont été mises en place pour faciliter cette identification ou abstraire le système lorsque l'identification est impossible comme l'utilisation d'étude de réactions similaires [5],[68]. Lorsque les quantités de molécules sont très faibles et que le système est de petite taille, des modélisations stochastiques doivent être mises en oeuvre pour capturer la variabilité des réponses du systèmes. Lorsque les quantités de molécules ne peuvent pas être estimées (par exemple pour les réseaux de signalisation), des modélisations discrètes (réseaux booléens) sont utilisées pour capturer les propriétés de la dynamique. Enfin, lorsque les quantités de molécules sont importantes, et que le système est de grande taille, les paramètres cinétiques ne peuvent pas être appris et la littérature s'appuie sur une hypothèse de quasi-stationnarité. Nous nous ramènon alors à l'étude d'un système linéaire sous contraintes [78]. Ces approches ont en particulier été développées dans le cadre de l'étude des réseaux métaboliques à l'échelle génomique durant les 15 dernières années. En alternative, les contraintes linéaires peuvent être relaxées sous la forme de contraintes qualitatives sur les chemins du graphe [38], avec un impact fort en biotechnologie.

Dans cette thèse, nous nous sommes focalisés sur l'étude des réseaux métaboliques à grande échelle dont nous présentons la construction et modélisation dans le chapitre 1. En particulier, nous avons cherché à comprendre si les approches (plutôt de nature dynamique) relatives à la modélisation des propagations de flux dans le système permettaient d'identifier des éléments structuraux d'importance dans un réseau métabolique. Pour cela, nous allons confronter et comparer dans le chapitre 2 des approches par contraintes linéaires, liées aux flux dans le réseau, et par contraintes qualitatives, dépendant de la topologie, permettant de définir dans chacun des cas une notion de productibilité dans un réseau métabolique. Suite à cette définition, nous pouvons montrer dans le chapitre 3 que le concept d'essentiel dans la production d'une biomasse peut avoir plusieurs définitions dépendantes du point de vue choisi, bien que toujours liées aux flux des réactions. Nous présentons alors, dans le chapitre 4, trois concepts de métabolites essentiels pour la production d'une biomasse, définis respectivement dans le cas d'une étude de production d'une biomasse, d'une étude d'une production optimale d'une biomasse et enfin d'une étude d'une accessibilité topologique des composants d'une biomasse. Nous définissons également le concept de clusters de métabolites essentiels dans le cas topologique. Les résultats obtenus dans l'étude de ces composants essentiels sont présentés dans le chapitre 5 et ceux pour les clusters dans le chapitre 6.



# Chapitre 1

## Modélisation des réseaux métaboliques

Dans ce premier chapitre de bibliographie, nous introduisons les principaux éléments relatifs à la modélisation des réseaux métaboliques, sous la forme d'un graphe et sous leur version enrichie par les informations stœchiométriques. Nous décrivons ensuite rapidement les différentes étapes de la reconstruction d'un réseau métabolique à partir de données génomiques, pour mettre en évidence les biais induits par ces protocoles de reconstruction. Enfin, nous détaillerons les propriétés de six réseaux métaboliques que nous avons sélectionnés comme étant caractéristiques d'espèces procaryotes et eucaryotes simples, et qui seront les réseaux sur lesquels toutes les méthodes développées dans cette thèse seront illustrées.

### 1.1 Représentation des réseaux métaboliques

#### 1.1.1 Introduction

Les réseaux métaboliques représentent une partie des connaissances biologiques d'un organisme correspondant à la production et à la consommation de molécules contrôlées par des protéines appelées enzymes. Les informations relatives aux réseaux métaboliques sont majoritairement extraites des annotations de génomes, qui permettent de lier un gène à des réactions métaboliques via des associations gène-protéine ou GPR (Gene protein Relation), plus précisément grâce aux enzymes catalysant chacune des réactions obtenues en étudiant la transcription et la traduction de ces mêmes gènes. Notons cependant que les enzymes ne sont que peu représentées dans les réseaux métaboliques, qui souvent se focalisent sur les relations directes entre métabolites et réactions.

Ainsi, un réseau métabolique représente l'ensemble des processus chimiques, c'est-à-dire les réactions métaboliques, se déroulant au sein d'un organisme. Il modélise la transformation de métabolites substrats (ou nutriments) en métabolites produits (souvent les constituants de la biomasse), permettant à l'organisme de croître, se développer et interagir avec son environnement. L'enchaînement de plusieurs réactions induisant la production de molécules, immédiatement suivie par leur consommation, met en avant des chemins appelés "voies métaboliques". Un des exemples de voie métabolique les plus connus car se retrouvant dans de nombreux organismes est celui de la glycolyse, transformant le glucose initial en Acétyl-CoA, comme illustré Figure 1.1. Un réseau métabolique est donc composé de nombreuses réactions métaboliques pouvant être regroupées en voies métaboliques. Par exemple, un des réseaux métaboliques les plus étudiés est celui d'*Escherichia coli*, qui s'est développé et amélioré au fil des ans [9] et dont le plus récent, nommé *iJO1366*, est composé de 1136 métabolites et 2251 réactions métaboliques.

#### 1.1.2 Graphe orienté biparti

Un réseau métabolique peut être représenté sous la forme de deux types de graphes orientés, l'un hypergraphe uniparti et l'autre biparti.

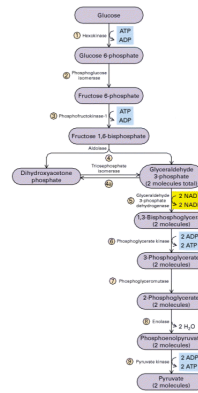


FIGURE 1.1 – Exemple de voie métabolique : la glycolyse

Nous nous intéressons plus particulièrement à la modélisation en graphe orienté biparti, qui sera noté  $G = (M \cup R, E)$ . Ce dernier est qualifié de biparti car les nœuds le composant sont divisibles en deux ensembles distincts  $M$  (métabolites) et  $R$ , et chaque arc dans l'ensemble  $E$  possède une extrémité appartenant à chacun de ces ensembles.

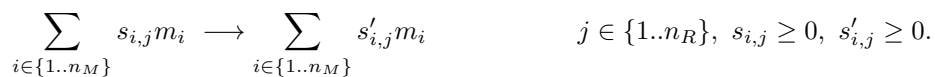
Pour représenter un réseau métabolique, le graphe est défini comme suit.

- L'ensemble de nœuds  $M$  correspond aux métabolites présents au sein de l'organisme.
- L'ensemble de nœuds  $R$  symbolise les réactions se produisant dans le réseau, transformant les métabolites définis précédemment. Ces réactions peuvent être irréversibles (unidirectionnelles) ou réversibles (bidirectionnelles).
- L'ensemble  $E \subset M \times R \cup R \times M$  est constitué d'arcs orientés liant des nœuds substrats à des nœuds réactions ou des nœuds réactions à des nœuds produits.

Un graphe réseau métabolique est donc caractérisé par un nombre  $n_M$  de métabolites que nous notons  $m_1, m_2, \dots, m_{n_M}$  et un nombre  $n_R$  de réactions notées  $r_1, r_2, \dots, r_{n_R}$ , se composant ainsi de  $n_M + n_R$  nœuds.

### 1.1.3 Stoechiométrie

Comme vu précédemment, une réaction est une transformation chimique de métabolites substrats en métabolites produits. On décrit les quantités respectives de métabolites substrats et produits impliqués dans une réaction sous la forme de coefficients. Une réaction métabolique se décrit alors de la manière suivante :

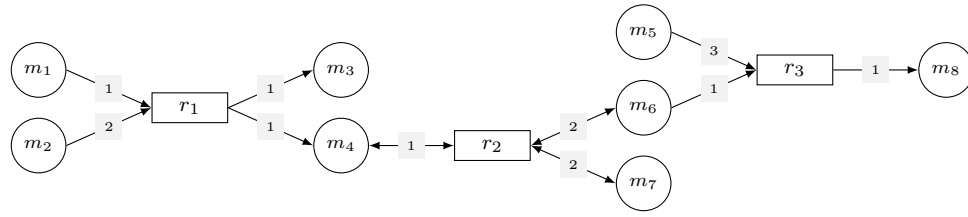


Les coefficients  $s_{i,j}$  et  $s'_{i,j}$  sont des nombres réels positifs dit de stoechiométrie et correspondent aux quantités de chaque élément consommé ou produit lors de la  $j^{\text{ème}}$  réaction.

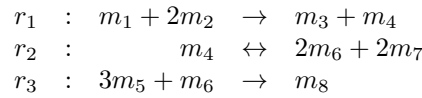
Les coefficients stoechiométriques permettent d'enrichir le graphe représentant un réseau métabolique. Dans ce contexte, on décrira un réseau par l'ensemble  $G = (M \cup R, E, s, s' : E \rightarrow \mathbb{R})$  qui vérifie les contraintes suivantes :

$$\begin{aligned} (m_i, r_j) \in E &\Leftrightarrow \exists i \in \{1..n_M\}, j \in \{1..n_R\} \mid s_{i,j} > 0 \\ (r_j, m_i) \in E &\Leftrightarrow \exists i \in \{1..n_M\}, j \in \{1..n_R\} \mid s'_{i,j} > 0 \end{aligned}$$

Il est aussi possible de représenter le réseau métabolique à l'aide d'une matrice  $\mathcal{M}_G$  dite stoechiométrique à partir de ces données, dont les dimensions sont  $n_M \times n_R$ , et dont les coefficients sont définis par  $\mathcal{M}_{Gij} = [s'_{i,j}] - [s_{i,j}]$ . Dans cette matrice, une ligne correspond à un métabolite du réseau et une colonne à une de ses réactions. Un exemple de ces différentes représentations est présenté dans la Figure 1.2.



(a) Graphe orienté biparti



(b) Réactions du réseau métabolique

$$\begin{pmatrix}
 -1 & 0 & 0 \\
 -2 & 0 & 0 \\
 1 & 0 & 0 \\
 1 & -1 & 0 \\
 0 & 0 & -3 \\
 0 & 2 & -1 \\
 0 & 2 & 0 \\
 0 & 0 & 1
 \end{pmatrix}$$

(c) Matrice stœchiométrique

FIGURE 1.2 – **Exemple de représentations d'un réseau métabolique** Le réseau métabolique proposé dans cet exemple est composé de deux réactions irréversibles  $r_1$  et  $r_3$  et d'une réaction réversible  $r_2$ .

### 1.1.4 Imports (sources) et exports (cibles) de métabolites

Afin de se rapprocher au plus près de la réalité biologique, on distingue assez naturellement deux sous-ensembles de métabolites de  $M$ .

Le premier ensemble de métabolites, noté  $S$ , correspond aux éléments nutritifs du milieu entrant dans le réseau métabolique qui sont appelés *seeds* (sources) tandis que le second ensemble de métabolites, noté  $T$ , regroupe les métabolites exportés du réseau métabolique qui sont appelés *targets* (cibles). Il est possible de définir ces ensembles en étudiant les représentations du réseau, que ce soit le graphe ou la matrice.

En effet, les métabolites sources sont, en ce qui concerne le graphe, les nœuds métabolites n'ayant pas d'arc entrant alors que, dans la matrice stœchiométrique, ces derniers sont détectés si l'ensemble des coefficients d'une ligne sont négatifs ou nuls :

$$S = \underbrace{\{m_i \in M \mid \nexists j \in \{1..n_R\}, (r_j, m_i) \in E\}}_{\text{réseau représenté par le graphe } G=(M \cup R, E)} = \underbrace{\{m_i, i \in \{1..n_M\} \mid \forall j \in \{1..n_R\}, s_{i,j} \leq 0\}}_{\text{réseau représenté par la matrice } \mathcal{M}_G}$$

Les métabolites cibles peuvent être détectés au niveau du graphe car ils correspondent aux nœuds métabolites n'ayant aucun arc sortant et, en ce qui concerne la matrice stœchiométrique, aux lignes ne possédant aucun coefficient négatif.

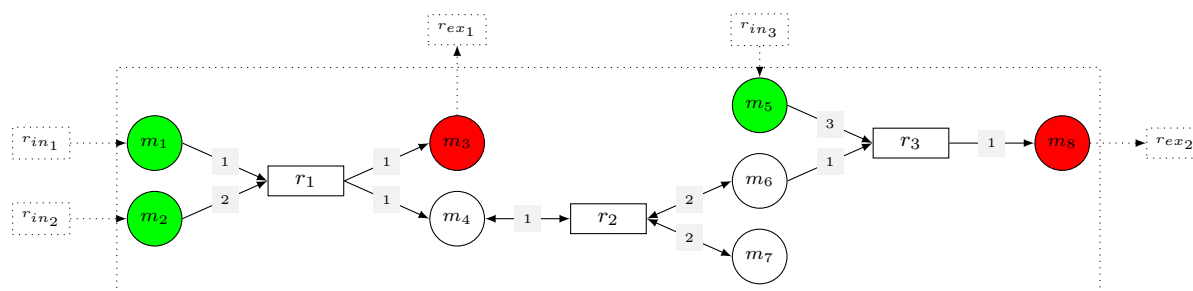
$$T = \underbrace{\{m_i \in M \mid \nexists j \in \{1..n_R\}, (m_i, r_j) \in E\}}_{\text{réseau représenté par le graphe } G=(M \cup R, E)} = \underbrace{\{m_i, i \in \{1..n_M\} \mid \forall j \in \{1..n_R\}, s_{i,j} \geq 0\}}_{\text{réseau représenté par la matrice } \mathcal{M}_G}$$

Pour représenter ces apports et exports d'éléments dans le réseau métabolique, on a l'habitude d'ajouter des réactions au réseau métabolique. Si un métabolite est incorporé dans le réseau via une réaction dite d'import, ce fait est matérialisé par " $m \rightarrow$ ". D'un autre côté, lors de l'export d'un métabolite, une réaction d'export est ajoutée de la façon suivante " $m \rightarrow$ ". La Figure 1.3 présente l'exemple décrit dans la Figure 1.2 avec ces réactions d'import et d'export.

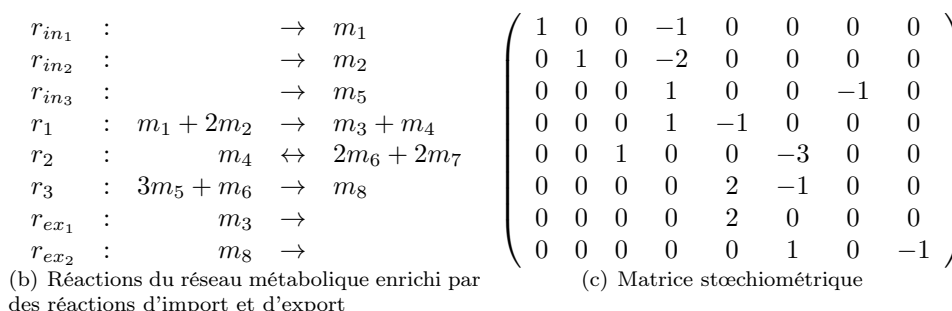
## 1.2 Protocole de reconstruction

La modélisation d'un réseau métabolique (que ce soit sous forme d'un graphe ou d'une matrice stœchiométrique) s'est longtemps appuyée sur les connaissances chimiques qui ont permis d'iden-





(a) Graphe orienté biparti



(b) Réactions du réseau métabolique enrichi par des réactions d'import et d'export

(c) Matrice stœchiométrique

FIGURE 1.3 – Exemple de représentation d'un réseau métabolique avec ajout des réactions de transport Des réactions d'import et d'export sont ajoutées au graphe du réseau afin de modéliser les transports d'éléments entre le milieu externe et le milieu interne de l'organisme.

tifier le rôle de différentes enzymes pour catalyser les transformations de molécules en d'autres molécules. Dans les années 2000, cette approche a été révolutionnée et enrichie avec le développement des techniques de séquençage de haut débit et des méthodes omiques, qui permettent d'obtenir de très nombreuses données sur la composition des réseaux métaboliques. En particulier, ces approches permettent d'identifier pour un organisme donné un catalogue de protéines potentiellement produites par cet organisme.

Cependant, l'information génétique étant fortement incomplète, et les connaissances chimiques ne suffisant pas à caractériser les spécificités d'une réaction biochimique dans un organisme donné (en particulier sur l'équilibre des co-facteurs), il est impossible d'obtenir un réseau métabolique directement à partir d'informations génomiques et différents protocoles peuvent être utilisés. Thiele et Palsson présentent dans leurs travaux [95] une méthode globale de reconstruction de réseau métabolique en essayant d'homogénéiser les différents protocoles pour la création de ces réseaux. La Figure 1.4, tirée de cet article, montre l'ensemble des étapes nécessaires.

Une méthodologie basée sur les travaux précédents est mise en avant dans [37], où les auteurs étudient les différents outils de reconstruction notamment appliqués à des réseaux bactériens en automatisant certaines étapes. Celle-ci est divisée en quatre étapes :

1. construction d'une ébauche du réseau métabolique
2. amélioration de l'ébauche
3. conversion du réseau dans un format utilisable par des méthodes automatiques
4. évaluation de la qualité du réseau final

Les étapes 2 et 3 sont souvent réalisées en parallèle car certaines méthodes d'amélioration peuvent être automatisées. De plus, l'évaluation finale permet de déterminer si de nouvelles améliorations doivent être faites.

### 1.2.1 Création d'une ébauche de réseau métabolique

La première étape de reconstruction d'un réseau métabolique consiste en l'élaboration d'une première ébauche ou *draft* du réseau métabolique brut, réalisée en extrayant les informations à partir de deux sources biologiques. D'une part, il est possible de s'appuyer sur les annotations

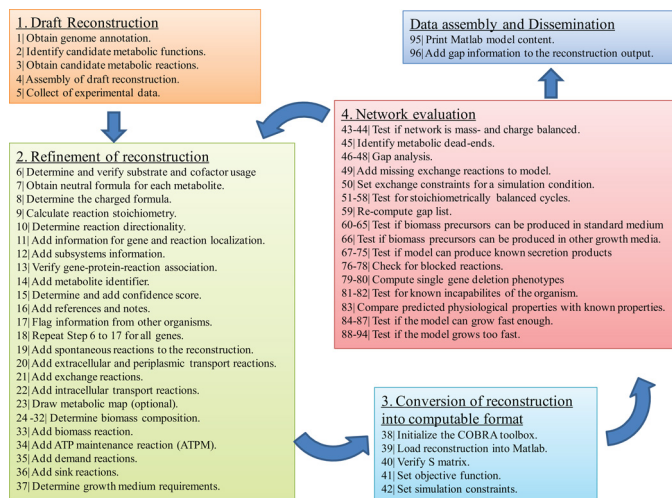


FIGURE 1.4 – Ensemble des étapes présentées par Thiele et Palsson pour la reconstruction d'un réseau métabolique [95] Les étapes 2 à 4 sont recommencées jusqu'à obtenir un réseau final satisfaisant, répondant aux attentes biologiques, dans l'étape 5.

fonctionnelles du génome de l'espèce étudiée, de son transcriptome ou de son protéome en utilisant différentes données telles que :

- le numéro EC, permettant de classifier les enzymes selon la réaction qu'elles catalysent,
- le terme GO, lié à l'ontologie des gènes dans les espèces,
- certains mots clés
- les noms génériques des réactions

Comme vu précédemment, l'ensemble de ces données permettent de savoir quelles enzymes peuvent être créées dans l'organisme, déterminant ainsi les réactions pouvant se dérouler, mais aussi quelles protéines sont présentes en son sein.

D'autre part, des informations pour créer cette ébauche peuvent être récupérées par étude d'espèces voisines. Il est notamment possible de rechercher des homologies entre gènes à partir de profils HMMs (*Hidden Markov Models* ou modèles de Markov cachés) permettant de retrouver des enzymes pour des gènes non annotés. Pour ce faire, ces données sont récupérables à partir de bases de données comme KEGG [50] ou Metacyc [15]. Ces ajouts sont faits selon des scores d'alignement afin de déterminer si l'espèce de référence utilisée est suffisamment proche de celle étudiée pour valider les changements. Il existe pour cela plusieurs outils d'alignement pour les homologies entre espèces comme décrits dans les travaux de Palsson [80] ou de Li et al [69].

Mais de nombreuses erreurs peuvent se produire lors de l'élaboration de l'ébauche du réseau métabolique. En effet, les annotations du génome de l'espèce étudiée peuvent présenter des erreurs entraînant alors des ajouts de réactions métaboliques finalement non présentes dans l'organisme. De plus, l'utilisation des logiciels d'alignement peut changer le résultat final selon l'outil utilisé. Ce problème a été mis en avant par Loira et al [70], qui ont créé un score de réconciliation afin de pallier ce problème. L'ébauche doit donc être vérifiée avant d'être considérée comme un réseau métabolique final.

### 1.2.2 Raffinement du réseau métabolique et conversion en format analysable

L'ébauche du réseau métabolique doit être vérifiée et complétée afin de pouvoir utiliser ce dernier dans des analyses plus poussées de l'organisme modélisé. Des connaissances biologiques sur l'espèce étudiée sont nécessaires à ce processus de nettoyage des réactions du réseau métabolique. Ainsi différentes études sont faites pour déterminer la présence des réactions comme les profils métaboliques mettant en avant les métabolites présents à un instant dans l'organisme, des études de productibilité et des preuves expérimentales. Les deux dernières méthodes impliquent une connaissance non seulement de l'espèce mais aussi de l'environnement de croissance de celle-ci.

Des preuves expérimentales peuvent être obtenues en modifiant, grâce à des techniques géniques, le génome de l'organisme étudié, créant ainsi des mutants et permettant de comprendre l'impact de ces gènes modifiés au sein de l'espèce. Une autre méthode est la modification du milieu de croissance que ce soit au niveau des métabolites présents ou des quantités de ces derniers.

Des comparaisons entre l'évolution expérimentale des quantités de molécules et les études de productibilité *in silico* du réseau métabolique peuvent alors être réalisées, ces études pouvant être de deux types selon la simulation de production étudiée.

D'une part, d'étudier seulement la topologie du réseau en regardant s'il existe dans le graphe créé un chemin des métabolites sources correspondant aux nutriments du milieu jusqu'aux métabolites cibles définis comme éléments de croissance de l'organisme [87]. D'autre part, l'étude du réseau métabolique via un graphe peut être réalisée d'un point de vue numérique en prenant les quantités produites de métabolites notamment en utilisant des analyses de flux (FBA) [28] que nous expliquerons dans le chapitre 2.2.

Si le comportement *in vivo* de l'organisme dans un milieu donné correspond aux simulations effectuées *in silico* sur le réseau métabolique alors cette ébauche est considérée comme valable. Mais souvent, le réseau est incomplet, empêchant la productibilité topologique ou numérique des métabolites cibles. Il est alors nécessaire d'ajouter des réactions afin d'obtenir la production de l'ensemble des cibles, de manière manuelle [25] ou automatique. Différents logiciels de *gapfilling* existent utilisant des algorithmes divers dont trois exemples sont présentés en Annexe A.

Des outils automatiques pour réaliser une partie de ces analyses peuvent être utilisés dès que le réseau et ses connaissances associées sont formatés de manière standardisée. L'un des formats les plus utilisés est le SBML pour *Systems Biology Markup Language* [42] listant dans un premier temps l'ensemble des métabolites intervenant dans le réseau puis, dans un second temps, la liste des réactions avec leurs substrats et leurs produits. Ce format reste cependant relativement permissif pouvant entraîner des difficultés lors de l'élaboration d'outils génériques, mais permet d'évaluer aussi le réseau métabolique obtenu après les différents ajouts opérés pour compléter ce dernier.

### 1.2.3 Évaluation du réseau métabolique final

L'ajout des réactions métaboliques manuellement ou avec des outils automatiques permet de rendre le réseau valide au sens où il permet généralement alors de modéliser *in silico* la croissance de l'organisme. Cependant, les modifications introduites dans le réseau doivent être validées. Un moyen d'étudier la cohérence des réactions métaboliques ajoutées dans le réseau est la recherche de gènes pouvant coder les enzymes liées à ces réactions. Pour ce faire, il est possible de rechercher des orthologues [69] ou la signature de famille de protéines liées à la réaction [29]. Les outils automatiques vus précédemment pour le raffinement de l'ébauche métabolique peuvent aussi être utilisés pour évaluer le réseau métabolique notamment au niveau fonctionnel. Pour cela, des méthodes numériques telle que l'analyse de flux (FBA expliquée chapitre 2.2) permettent de modéliser la croissance de l'organisme en utilisant des algorithmes d'optimisation linéaire ou de programmation par contraintes. Différentes approches permettant à la fois la complétion du réseau et son évaluation sont présentées dans l'Annexe A.

Le point important à retenir de cette description du protocole de reconstruction d'un réseau métabolique est que de nombreux biais peuvent être introduits. D'abord, la qualité d'un réseau dépend fortement de la qualité des annotations du génome de l'espèce considérée, qui peuvent introduire des réactions artéfactuelles. Ensuite, la validité des réseaux, au sens où ils sont capables de prédire *in silico* la croissance de l'organisme, est obtenue via l'utilisation de différents algorithmes et méthodes pour rajouter des interactions (*gap-filling*), introduire des imports et des exports artéfactuels, ou corriger les équilibres de co-facteurs. Dans l'ensemble, les différentes méthodes utilisées sont susceptibles de se répercuter sur la structure et les prédictions des différents réseaux, et il peut être important de détecter les différents biais introduits par ces méthodes.

## 1.3 Les réseaux métaboliques existants

Le protocole de reconstruction présenté précédemment, composé de quatre étapes commençant par la réalisation d'une ébauche de réseau suivie par une complétion puis un formatage, et enfin une vérification, a été de nombreuses fois utilisé afin de reconstruire les réseaux métaboliques d'espèces connues et ainsi essayer de comprendre le fonctionnement métabolique de ces organismes.

Plus précisément, les réseaux métaboliques sont devenus cruciaux depuis 50 ans pour comprendre le fonctionnement des organismes. La compréhension notamment du corps humain est un des enjeux majeurs de la réalisation de réseau métabolique. Ce dernier fut donc un des premiers modèles reconstruits [21] en 2007.

Cependant, les réseaux métaboliques de différentes espèces ont été modélisés avant l'obtention du modèle humain, tel qu'il est possible de le voir dans la Figure 1.5 tirée des travaux de Kim *et al* [56], notamment des organismes bactériens (procaryotes). Ces derniers ayant un système biologique plus facilement accessible et surtout un génome plus simple à séquencer, assembler et à annoter, ils sont plus nombreux à avoir un réseau métabolique modélisé.

Ainsi, dans la figure présentée, sur 62 espèces modélisées au cours du temps, 74% sont des bactéries, 19% des Eucaryotes et 7% des Archées.

De plus, alors que certaines espèces n'ont qu'une modélisation apparaissant dans la figure, d'autres ont été l'objet de plusieurs représentations métaboliques sur plusieurs années. De cette façon, pour les 62 organismes montrés dans la figure étudiée, nous pouvons compter 95 réseaux modélisés différents (71% Bactéries, 4% Archées et 25% Eucaryotes).

Pour chaque espèce représentée dans la Figure 1.5, nous pouvons aussi voir le nombre de gènes identifiés ainsi que le nombre de réactions et de métabolites présents dans le réseau modélisé. Une évolution de la taille des réseaux est mise en avant car, alors que les plus anciennes modélisations comptaient une centaine de gènes et réactions, de nos jours, les réseaux métaboliques sont dit d'échelle génomique avec un millier de réactions métaboliques en leur sein.

Le choix d'étudier ces espèces est souvent lié à un besoin de compréhension de l'organisme soit en tant que modèle tel que *Saccharomyces cerevisiae*, soit dans le but d'une production contrôlée de métabolites d'intérêt notamment en agroalimentaire ou dans l'industrie pétrolière [11, 84], base de la biologie synthétique. La Figure 1.6, obtenue à partir des travaux de Monk et al [77], montre la couverture phylogénétique des espèces dont le réseau métabolique a été modélisé.

Comme dans la figure précédente, nous pouvons voir qu'un grand nombre de réseaux reconstruits appartient aux Bactéries (69%, 7% pour les Archées et 24% pour les Eucaryotes). Il est aussi possible de voir que les espèces les plus représentées sont les Protéobactéries, dont *Escherichia coli* est l'un des organismes modèles les plus étudiés. Les Eucaryotes et les Archées sont sous-représentés avec de nombreux phyla totalement non représentés dans cette figure datant de février 2013 mais, de par l'avancement des techniques omiques, de nouvelles espèces sont séquencées et de nouveaux réseaux de ces phyla, alors non étudiés, sont modélisés.

Les différents modèles créés sont regroupés dans des bases de données telles que *BIGG Models* [57]. Cette dernière contient 75 réseaux métabolites de taille génomique, hautement nettoyés manuellement et permet à l'utilisateur d'étudier, chercher ou visualiser des modélisations de réseaux métaboliques. Celles-ci possèdent des précisions sur les gènes annotés et sont liées à des bases de données externes confirmant les informations utilisées pour créer le réseau métabolique dont les identifiants de métabolites et de réactions sont standardisés pour faciliter les échanges entre bases de données. De plus, *BIGG Models* propose des outils de modélisation de réseau métabolique et d'analyse de ces derniers notamment en permettant leur exportation en format analysable automatiquement comme SBML.

Plus généralement, la base de données Biocyc[52] est, avec la base KEGG, une des banque de données de référence concernant les réseaux métaboliques procaryotes et eukaryotes. La version 19.5 de cette base contenait 7 296 réseaux bactériens reconstruits avec des protocoles automatiques (Tier 3) et 27 réseaux bactériens curés manuellement (Tier 1 et Tier 2). On peut noter que les réseaux reconstruits automatiquement contenaient 10% de réactions en plus que les réseaux ayant fait l'objet d'une curation manuelle.

Cela nous donne une indication sur les biais induits par les protocoles de reconstruction. Pour

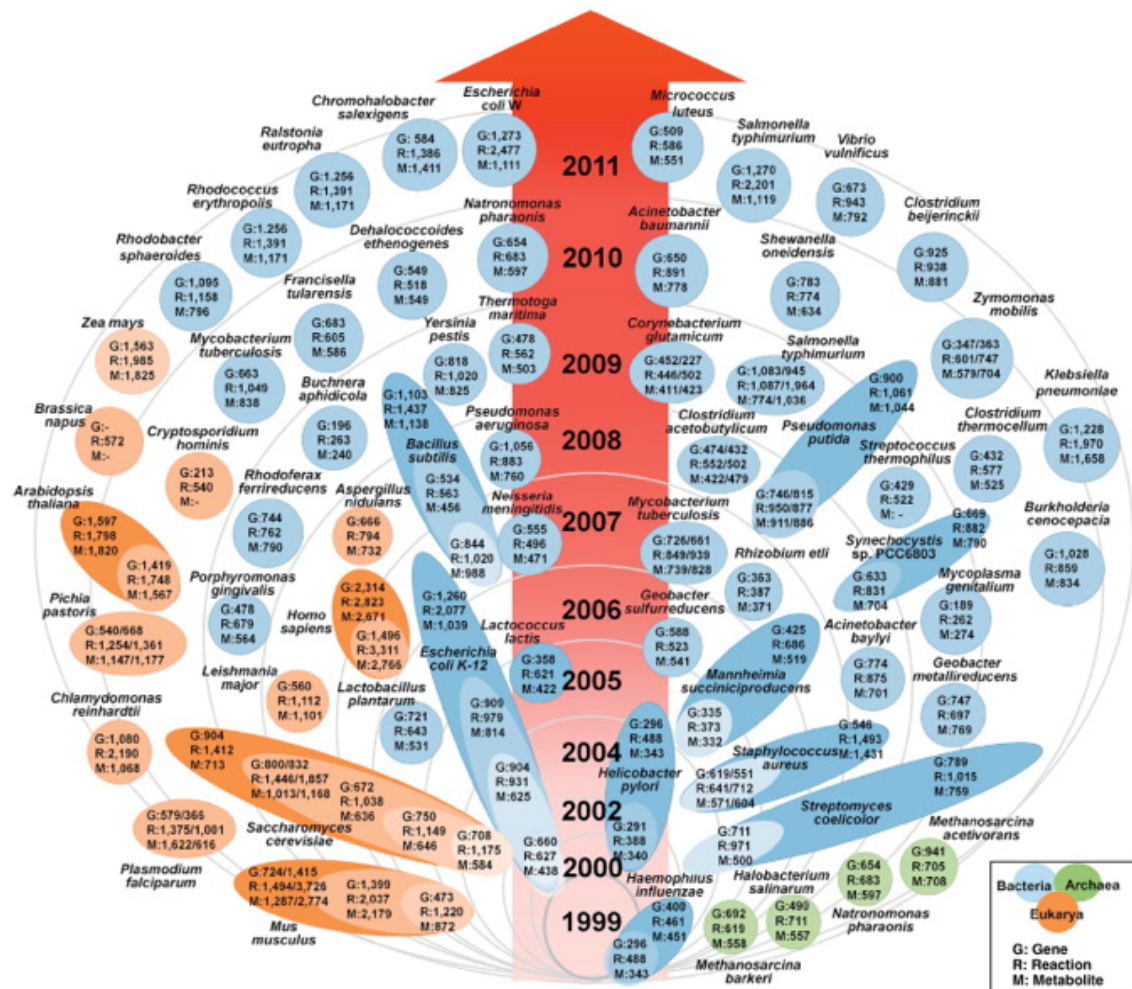


FIGURE 1.5 – Reconstruction des réseaux métaboliques des espèces entre 1999 et 2011  
 Figure tirée de [56] paru en 2012.

Metabolic reconstructions for 78 species across the tree of life

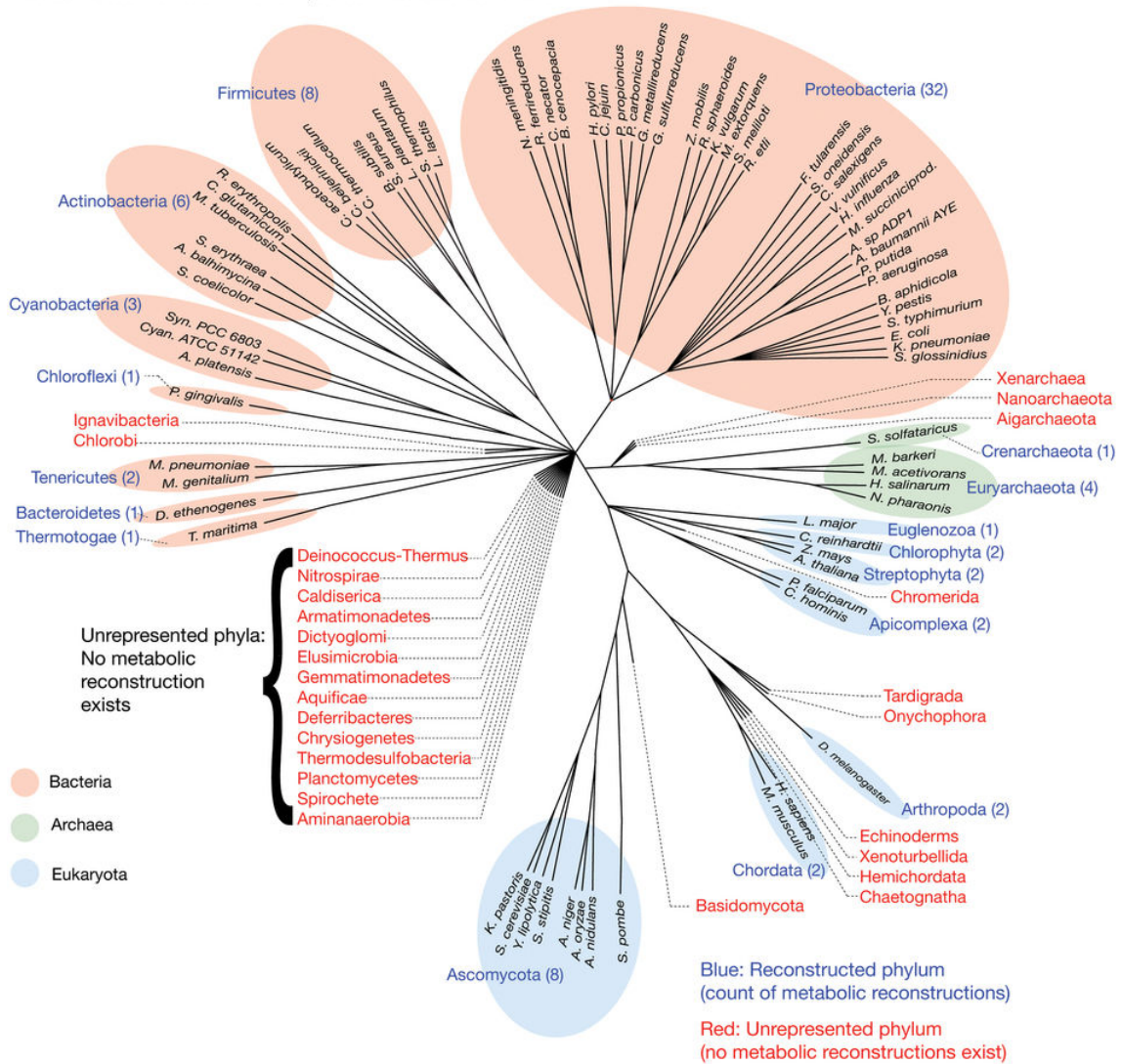


FIGURE 1.6 – Couverture des reconstructions de réseaux métaboliques tirée de [77], février 2013

limiter ces biais, nous prendrons le parti, dans cette thèse, de développer des approches qui croisent et combinent différentes approches par contraintes (linéaires et combinatoires) de manière à rendre les résultats d'analyses plus robustes.

## 1.4 Six exemples caractéristiques

Nous avons vu dans la section précédente qu'il est de nos jours relativement aisé d'obtenir le réseau métabolique d'une espèce ou bien d'une espèce proche ou modèle de par la mise en circulation des données biologiques de ces derniers via des bases de données.

Tout au long de cette thèse, nous avons considéré plusieurs réseaux métaboliques d'échelle génomique appartenant à différent phylum, certains considérés comme espèce modèle tandis que d'autres sont plus spécialisés.

### 1.4.1 Réseaux modèles

Les réseaux modèles les plus étudiés sont à ce jour les réseaux relatifs à la bactérie *Escherichia coli*. Cette bactérie se retrouve dans la flore intestinale de l'homme et des animaux à sang chaud. Elle est très étudiée car, bien que la plupart des souches de *Escherichia coli* soient inoffensives, quelques-unes sont pathogènes. De plus, les capacités expérimentales pour activer ou inhiber ses gènes sont extrêmement développées.

Nous nous intéressons à la souche *Escherichia coli* str. K-12 substr. MG1655. Plusieurs réseaux métaboliques ont vu le jour à partir de multiples données et de l'avancement des technologies pour cette dernière. Nous avons utilisés trois de ces réseaux métaboliques présentés ici chronologiquement et publiquement accessibles : *iJR904*[85], *iAF1260*[26] and *iJO1360*[78]. Ces derniers représentent l'évolution du réseau métabolique d'un organisme isolé et permettent de reproduire les améliorations proposées par de nombreux modèles de l'organisme.

#### Réseau E. Coli *iJR904* premier réseau à échelle génomique

Ce réseau fut publié par Reed et al [85] en 2003. Il correspond à une expansion du premier réseau métabolique établi à l'échelle génomique, *iJE660*, créé à partir de données génomiques, biochimiques et physiologiques. Cette amélioration a été possible grâce à l'apport de nouvelles données génomiques et techniques comme l'association Gène-Protéine-Réaction mettant en avant les relations entre ces trois entités. Les auteurs de ce réseau ont ajouté les réactions liées à la découverte de nouveaux gènes et pris en compte, pour chaque réaction, l'équilibre des éléments et des charges. Le réseau métabolique *iJR904* est constitué de 931 réactions biochimiques (627 dans *iJE660*) obtenues à partir de 904 gènes (660 pour *iJE660*).

Il a ensuite été utilisé pour étudier les métabolites notamment appelés *deadends* car étant soit produits mais non consommés, soit consommés mais non produits, ainsi que pour étudier et comparer les réactions et les voies métaboliques nouvellement ouvertes par rapport au réseau précédent par des analyses de flux. Cela a permis d'orienter de futures recherches biologiques en mettant en avant des éléments possibles pouvant être ajoutés au réseau mais devant être confirmés par des expériences et données biologiques.

#### Réseau E. Coli *iAF1260* introduction de compartiments

Le réseau E. Coli *iAF1260* a été proposé en 2007 lors de travaux réalisés par Feist et al [26] et est dans la continuité du réseau précédent *iJR904*.

De nouvelles données génomiques ainsi que le développement de nouvelles bases de données ont permis aux auteurs de ce réseau de le développer en incluant de nouveaux métabolites, au nombre de 414, et 1146 réactions retrouvées mais aussi de nettoyer le réseau précédent en supprimant 289 réactions dont certaines ont été remplacées par des réactions équivalentes (254).

Une deuxième amélioration du réseau a été de diviser ce dernier en sous-réseaux correspondant à des compartiments ayant une réalité biologique : le cytoplasme, le périplasme et l'espace extracellulaire. Ce dernier modélise le milieu environnemental dans lequel la bactérie E. Coli évolue. Chaque métabolite est alors spécifié dans un ou plusieurs compartiments et des réactions de transport entre compartiments sont ajoutées au sein du réseau. Le réseau possède alors plusieurs

copies du même métabolite et/ou de la même réaction réparties dans les compartiments. Les auteurs de [26] posent alors comme condition de construction du réseau métabolique de minimiser le nombre de groupements de réactions. De plus, une vérification de l'équilibre thermodynamique des réactions ainsi qu'une comparaison avec des bases de données sont réalisées pour valider les changements effectués.

Le réseau *iAF1260* ainsi créé est ensuite transformé en un modèle mathématique afin d'étudier les flux en son sein. Des métabolites sont spécifiés comme appartenant au système externe au métabolisme (environnement) et des réactions d'import ou d'export sont ajoutées pour permettre à ces derniers d'entrer dans le réseau métabolique pour les nutriments ou de sortir pour des métabolites cibles. Le réseau ayant des incomplétudes l'empêchant de prédire la production de biomasse, une complétion de ces derniers est effectuée en se basant sur l'analyse de voies métaboliques.

Pour finir, une réaction de biomasse est créée pour modéliser la croissance de l'organisme. Le modèle ainsi obtenu est utilisé pour étudier la thermodynamique du réseau ainsi que la robustesse et la sensibilité du réseau métabolique *iAF1260* en jouant sur le blocage des flux dans certaines réactions.

### Réseau *E. coli iJO1360* introduction de réponses à différents stress

Un nouveau réseau pour *E. coli* a été proposé en 2011 par Orth et al [78] et nommé *iJO1360*. Il correspond à une nouvelle mise à jour du réseau *iAF1260* vu précédemment au vu de l'avancée des techniques de recherche biologique et l'obtention de nouvelles données. En effet, les méthodes de séquençage du génome ayant été améliorées, de nouveaux gènes et ainsi de nouvelles réactions, dont l'équilibre des masses et des charges a été vérifié, ont pu être caractérisés au sein de l'organisme.

Un autre moyen ayant permis notamment de compléter les incomplétudes dans le réseau précédent fut l'étude de la réaction de l'organisme à quatre conditions environnementales différentes. De plus, la comparaison de la réponse phénotypique à ces dernières a aussi permis de corriger certaines réactions ayant été ajoutées auparavant. La complétion des trous du réseau métabolique restant fut étudiée en recherchant les métabolites bloquants car n'ayant pas, au sein du réseau, soit une réaction le produisant, soit une réaction le consommant, puis en recherchant dans la littérature et les bases de données des réactions pouvant être omises dans le réseau et permettant de combler ces manquements.

Le nouveau réseau *iJO1360* ainsi obtenu contient alors 2251 réactions métaboliques et 1136 métabolites, répartis comme dans le réseau précédent *iAF1260* dans trois compartiments biologiques : le cytoplasme, le périplasme et l'espace extracellulaire. La plupart des réactions ajoutées entraîne la création de nouvelles voies métaboliques mais certaines permettent la complétion de trou ou relie des réactions isolées au reste du réseau métabolique. De plus, les réactions de biomasse modélisant la croissance de *E. coli* dans plusieurs milieux ont été adaptées aux nouvelles données obtenues expérimentalement. Afin de valider ces changements, notamment au niveau des ajouts sans confirmation biologique, des comparaisons homologues ainsi que génétiques ont été faites avec d'autres génomes de *E. coli*. Comme précédemment, le réseau *iJO1360* a été interprété en modèle mathématique afin de pouvoir étudier les flux présents en son sein et a pour but d'être un modèle de réseau métabolique afin de pouvoir étudier des espèces moins connues.

### Réseau *Synechocystis* métabolisme central d'une cyanobactérie marine

Avec *E. coli*, la cyanobactérie *Synechocystis* est aussi considérée comme une espèce modèle. En effet, cet organisme pouvant se retrouver dans différents milieux, il est intéressant d'étudier sa réponse à un stress environnemental et son adaptation phénotypique. Pour cela, de nombreuses études de son réseau métabolique ont été faites. Nous nous intéressons plus particulièrement à la souche *Synechocystis* sp. PCC 6803 [61] notamment pour la complémentarité de photosynthèse avec *E. coli* permettant de cibler des caractéristiques d'oxydo-réduction sous des conditions d'hétérotrophie. En effet, cet organisme est considéré comme un modèle pour l'étude de la photosynthèse oxygénique des plantes hautes, ces dernières ayant un génome plus compliqué à analyser par rapport à *Synechocystis* étant donnée leur compartimentation cellulaire.

Le séquençage complet de cette cyanobactérie a été réalisé en 1996 et publié dans [51]. Cela a permis de créer le réseau métabolique correspondant à ce génome en plusieurs étapes. Dans un premier temps, les auteurs ont construit un premier réseau allégé pour estimer la taille du génome



de l'organisme puis ont effectué plusieurs analyses des séquences obtenues via séquençage génique afin de limiter les pertes de connaissances. Les gènes ainsi obtenus sont comparés dans des bases de données pour retrouver les protéines liées à ces gènes. Certains correspondent entièrement à des éléments de la base de données alors que d'autres sont des homologues de gènes existant. Le réseau métabolique ainsi reconstruit comprend 758 réactions métaboliques et 600 métabolites et a été utilisé pour étudier le génome d'un micro-organisme.

### 1.4.2 Organismes exotiques

Quelques autres organismes plus exotiques peuvent être aussi étudiés afin de compléter le plus large panel d'organismes possibles. Nous nous sommes concentrés sur deux organismes étudiés dans le cadre de collaborations de mon équipe de thèse.

#### Réseau *A. ferrooxidans* bactérie extrémophile

Un organisme intéressant à étudier de par son environnement mais non modèle car très spécifique est la bactérie *Acidithiobacillus ferrooxidans*, str. Welenen [14]. En effet, cette dernière peut être trouvée dans des environnements contenant du soufre et du fer qu'elle utilise comme sources d'énergie pour croître autotrophiquement. De plus, elle joue un rôle important pour les environnements acides dans le cycle géologique et biochimique des nutriments et métaux du milieu. Elle peut donc se retrouver très utile pour l'exploitation minière, permettant une extraction moins polluante.

Le réseau métabolique présenté par [14] est le premier réseau à échelle génomique reconstruit de *Acidithiobacillus ferrooxidans*. Les auteurs ont utilisé pour cela les nouvelles technologies de séquençage de génome d'un organisme mais aussi des bases de données pour retrouver les gènes et réactions omis ainsi que des outils de reconstruction informatiques pour compléter les manques restés dans le réseau métabolique. Des comparaisons avec des espèces plus ou moins proches ont aussi été faites pour confirmer certains ajouts, de même que des recoupements avec des données expérimentales obtenues sous diverses conditions environnementales. Le réseau métabolique ainsi obtenu est composé de 587 réactions métaboliques et de transport ou échanges, et 573 métabolites. Chacune des réactions du réseau est équilibrée au niveau des masses et des charges ainsi que thermodynamiquement orientée, permettant de vérifier que le réseau métabolique est possible au niveau des flux de matières pour produire une biomasse.

Le réseau métabolique de *Acidithiobacillus ferrooxidans* a été créé dans le but d'étudier des phénotypes particuliers se déroulant dans des environnements extrêmes et permettre ainsi de comprendre l'adaptation de certaines espèces à ces milieux hostiles. Dans [14], les auteurs se sont surtout penchés sur des voies métaboliques peu connues en modélisant la réponse du système à des conditions particulières de façon automatique et en la comparant aux données expérimentales. De plus, ce travail permettrait aussi de développer les réseaux métaboliques correspondant à des espèces proches de *Acidithiobacillus ferrooxidans* ou présentant des similitudes avec cette dernière.

#### Réseau *T. lutea* métabolisme central d'une micro-algue eucaryote

Un comportement intéressant à étudier est la réaction d'un organisme au cycle jour/nuit. De nombreux organismes voient leur phénotype évoluer au cours de la journée et de la nuit, l'un d'eux est la micro-algue eucaryote *Tisochrysis lutea*. Cette dernière est intéressante car elle est capable d'accumuler de grandes quantités de lipides et donc un candidat de choix pour le développement notamment de biocarburants à partir de ces mêmes lipides. Pourtant son séquençage est récent.

Un réseau métabolique centré sur le métabolisme central présenté dans [8] a été reconstruit à partir du réseau de Kliphuis et al. dans [60] représentant le réseau cœur commun à toutes micro-algues eucaryotes. Des modifications ont été apportées en ajoutant des réactions de transport et de synthèse de macromolécules vérifiées expérimentalement. De plus, le réseau métabolique a été complété par des réactions prises dans des organismes proches de *Tisochrysis lutea*. La biomasse a été modifiée de façon à prendre en compte l'accumulation de matière au sein du réseau. Le réseau métabolique de *Tisochrysis lutea* final est composé de 157 métabolites internes et 162 réactions dont 13 réactions d'échanges avec le milieu externe et 1 réaction d'échange interne entre chloroplaste et cytosol.

L'étude de ce réseau métabolique a nécessité le développement de nouvelles techniques autre que celles d'analyse en flux (FBA expliquée chapitre 2.2) car l'hypothèse d'état stationnaire néces-

saire à ces études impliquant qu'aucune accumulation de matière ne peut être utilisée, est invalide. Pour cela, le réseau métabolique est décomposé en sous-réseaux selon les fonctions métaboliques prenant en compte la compartimentation cellulaire et les voies métaboliques. Six ensembles sont ainsi créés : la photosynthèse, la partie supérieure de la glycolyse, la synthèse de carbohydrates, la partie inférieure de la glycolyse, la synthèse de lipides et la synthèse de la biomasse. Ces derniers sont alors étudiés sous l'hypothèse d'état stationnaire puis réduits pour créer un modèle composé de 16 métabolites et 8 réactions macroscopiques permettant alors d'étudier l'accumulation des métabolites du réseau métabolique et de mieux comprendre les mécanismes intracellulaires mis en place pour répondre aux changements environnementaux.

Ce dernier travail est particulièrement novateur d'un point de vue méthodologique puisque la phase de décomposition en sous-réseaux a permis en particulier de diminuer drastiquement le nombre de comportements possibles du réseau en limitant le nombre de flux au sein de chaque sous-réseau et en recentrant la variabilité sur les quelques réactions d'échanges entre les différents sous-réseaux. Il faut cependant noter que l'identification des métabolites et réactions clés reliant les sous-réseaux a été faite manuellement à partir de l'expertise biologique. Comme nous le verrons dans nos résultats, les approches que nous développons dans cette thèse pour comprendre le rôle des métabolites au sein d'un réseau permettront d'identifier de manière automatique des sous-réseaux qui recouvrent globalement la décomposition manuelle introduite dans [8].

## Conclusion

La conclusion à retenir de ce chapitre de bibliographie est que les réseaux métaboliques peuvent être vus comme des graphes biparti dont les arêtes sont enrichies par des attributs correspondant à des coefficients stoechiométriques. Les technologies de séquençage développées les vingt dernières années ont permis d'accéder à une information majeure pour la reconstruction de réseaux métaboliques à grande-échelle. Cependant, il faut bien prendre conscience que malgré l'abondance de réseaux métaboliques reconstruits à ce jour, leur qualité est très variable et de nombreux biais méthodologiques peuvent se répercuter sur la robustesse des prédictions réalisées à partir d'un réseau métabolique. Dans cette optique, un des objectifs de cette thèse sera d'utiliser et d'exploiter différentes méthodes d'analyse pour mettre en évidence le rôle de certains composés clés et de gagner en finesse et robustesse dans l'étude d'un réseau métabolique.



## Chapitre 2

# Un concept de productibilité au sein d'un graphe métabolique

Un réseau métabolique représente l'ensemble des réactions se produisant dans un organisme obtenues via de multiples techniques de biotechnologie permettant de modéliser ce dernier sous forme notamment de graphe biparti orienté pondéré. Ce dernier peut alors être analysé sous différents aspects mais une notion récurrente est la capacité du réseau à produire des éléments. Cependant le concept de productibilité peut différer selon le point de vue adopté dans l'étude du réseau métabolique. Dans ce chapitre de bibliographie, nous montrerons que la notion d'élément productible dans le graphe métabolique est essentiellement associée à des études de flux par la méthode dite de *Flux Balance Analysis* ou celle de *Flux Variability Analysis* mais qu'une nouvelle notion de productibilité mise en avant notamment par Handorf et al [39] se focalisant sur une étude topologique du réseau métabolique est nécessaire pour décrire la capacité d'une cellule à croître sur du long terme.

### 2.1 Des flux dans le réseau

#### 2.1.1 Un graphe de flux

Comme nous l'avons défini dans le chapitre 1, un réseau métabolique peut être représenté sous forme de graphe biparti orienté et étiqueté par la stœchiométrie des réactions. Cela signifie que les nœuds du graphe sont répartis en deux ensembles, l'un symbolisant des métabolites et l'autre des réactions. Les arcs du graphe possèdent une extrémité dans chacun des ensembles et sont étiquetés par la stœchiométrie de la réaction correspondante.

**Définition 1 (Graphe métabolique avec stœchiométrie)** *On note par  $G = (R \cup M, E, s)$  la modélisation d'un réseau métabolique qui possède les caractéristiques suivantes :*

- $R$  ensemble des réactions
- $M$  ensemble des métabolites
- $m \in M, r \in R, \begin{cases} (m, r) \in E \Leftrightarrow \text{le métabolite } m \text{ est substrat de la réaction } r \\ (r, m) \in E \Leftrightarrow \text{le métabolite } m \text{ est produit de la réaction } r \end{cases}$
- $s : E \rightarrow \mathbb{R}$  fonction de stœchiométrie.

L'étude du réseau correspond alors à une étude de graphe de flux qui décrivent les transferts de matière au sein de l'organisme. Cependant tout organisme possède des flux de matières variables dans le temps. En effet, par exemple, certaines espèces changent de comportement selon le moment de la journée, produisant des métabolites la nuit, utilisés ensuite de jour, entraînant des accumulations d'éléments qui ne sont pas modélisables par une étude du graphe de flux.

#### 2.1.2 Des hypothèses pour modéliser des flux

Afin d'étudier les flux au sein d'un réseau métabolique, un formalisme de balance de flux est proposé notamment par Orth et al dans [79] ou Edward et Palsson [24]. Deux hypothèses sont posées

afin de modéliser les flux dans un réseau métabolique, afin de se mettre en situation de flux constants dans le temps. La première dite de conservation de masse implique que la concentration d'un métabolite dans le réseau est uniquement dépendante des flux de production et de consommation du réseau. La deuxième hypothèse sous-entend que le système modélisé est dans un état quasi-stationnaire [92], supposant aucune accumulation de matière. Dans ce cas, la somme des flux entrants dans un nœud métabolite est strictement égale à la somme des flux sortants. L'ensemble du réseau métabolique et de ces hypothèses sont la base d'une méthode d'étude des flux dans le réseau appelée *Flux Balance Analysis* ou FBA [79] utilisant la programmation linéaire pour résoudre le problème de flux ainsi modélisé.

### 2.1.3 La programmation linéaire par contraintes

La programmation linéaire par contraintes appartient au domaine de la recherche opérationnelle modélisant les problèmes sous forme mathématique afin de les résoudre [17]. Pour cela, le problème est décomposé en différentes parties identifiées comme d'une part les variables inconnues que nous cherchons à résoudre et d'autre part les contraintes auxquelles les variables sont soumises. Nous parlons dans notre cas de programmation linéaire car les contraintes sont modélisées sous forme de fonction linéaire des variables. Par exemple, si nous nommons les variables  $x_i$ , la forme d'une contrainte peut être de la forme  $\alpha x_i + \beta x_j < \gamma$  dans le cas d'une contrainte d'infériorité, celle-ci pouvant également être d'égalité ou de supériorité.

La modélisation en programmation linéaire par contraintes peut avoir deux buts selon la résolution du problème étudié. Ainsi ce dernier peut seulement chercher à calculer s'il existe une solution, nous parlons alors de problème de décision, et, lorsque c'est possible, retourne une solution possible. Un autre but de la programmation linéaire est la recherche d'un optimum par une fonction d'optimisation ajoutée au problème. Cette dernière est une fonction linéaire dépendante des variables du réseau pouvant être maximisée ou minimisée.

## 2.2 Flux Balance Analysis

Le problème de FBA [78] correspond à la modélisation mathématique permettant de calculer la valeur des flux dans un graphe métabolique. Il appartient aux problèmes de décision définis en programmation linéaire par contraintes avec recherche d'une solution.

### 2.2.1 Une modélisation de la fonctionnalité du réseau métabolique

La méthode de *Flux Balance Analysis* est une technique d'analyse de réseau métabolique d'un point de vue quantitatif prenant en compte la stoechiométrie des réactions du réseau. Elle permet d'étudier la distribution des flux passant dans les différentes réactions au sein de ce dernier, simulant ainsi le métabolisme d'un organisme lorsque celui-ci est à l'état stable. Le FBA permet de faire de l'ingénierie et étudier la réponse d'un organisme à différentes conditions de croissance ou à des modifications du réseau dues à des suppressions de gènes par exemple. De plus, le calcul des flux dans le réseau est très rapide même pour les réseaux de très grande taille avec cette méthode. Grâce à cette rapidité de calcul et le peu d'informations nécessaires pour effectuer ces simulations, le *Flux Balance Analysis* se différencie des autres méthodes d'étude stoechiométrique de réseau métabolique utilisant des équations différentielles car il n'est pas nécessaire de connaître les constantes cinétiques des enzymes liées aux réactions ou la concentration des métabolites internes au réseau.

### 2.2.2 Modélisation mathématiques du problème de flux dans un réseau métabolique

Le problème de FBA est décomposé en deux parties en tant que problème de programmation linéaire : les variables et les contraintes.

**Variabes du problème de Flux Balance Analysis** Les variables du problème, nommées  $v_r$ , pour chacune des réactions  $r$  du réseau, modélisent la valeur des flux dans les réactions sous hypothèse d'état quasi-stationnaire. Elles peuvent prendre pour valeur un nombre réel correspondant à la quantité de flux traversant la réaction. Dans un cadre général (non stationnaire), la concentration d'un métabolite suit une cinétique qui est modélisée par l'équation différentielle suivante :

$$\frac{dC_i}{dt} = \sum_{j \in J} \mathcal{M}_{G_{ij}} \cdot v_j \quad \forall i \in I$$

avec  $\mathcal{M}_G$  matrice stœchiométrique définie par la [Définition 2](#) ci-dessous.

**Définition 2 (Matrice stœchiométrique d'un réseau métabolique)** La matrice stœchiométrique de  $G = (R \cup M, E, s)$  est notée  $\mathcal{M}_G \in n_M \times n_R$ , et est définie comme suit :

$$\begin{aligned} - I &= \{1..n_M\} \\ - J &= \{1..n_R\} \\ - \mathcal{M}_{G_{ij}} &= \begin{cases} s_{ij} < 0 & \text{si le métabolite } i \text{ réactant de la réaction } j \\ s_{ij} = 0 & \text{si le métabolite } i \text{ absent de la réaction } j \\ s_{ij} > 0 & \text{si le métabolite } i \text{ produit de la réaction } j \end{cases} \end{aligned}$$

Le problème de flux est ensuite contraint afin de modéliser au plus près la biologie.

**Contraintes du problème de Flux Balance Analysis** Un problème de flux dans un graphe présente des contraintes sur la valeur des variables. Dans le cas d'un réseau métabolique, ces dernières, représentant les flux dans les réactions, sont contraintes par des limites inférieures et supérieures réelles :

$$\begin{aligned} LB_j &\leq v_j \leq UB_j & \forall j \in J, \\ v_j &\in \mathbb{R} & \forall j \in J \end{aligned}$$

Ces contraintes permettent de bloquer éventuellement le sens d'une réaction selon la valeur de la borne inférieure. Ainsi si la réaction est biologiquement réversible, cette dernière est strictement négative tandis que, si elle est irréversible, la borne inférieure est nulle :

$$LB_j \begin{cases} < 0 & \text{si la réaction } j \text{ est réversible} \\ = 0 & \text{si la réaction } j \text{ est irréversible.} \end{cases}$$

En ce qui concerne les réactions d'import et d'export de métabolites du réseau métabolique, non réversibles, les flux sont considérés comme illimités car non restreints par le milieu environnemental. Ceci entraîne soit la suppression des contraintes correspondantes dans le problème linéaire, soit l'utilisation d'une constante  $C$  dont la valeur n'est pas limitative telle que :  $0 \leq v_j \leq C$ .

La contrainte suivante est typique du problème de Flux Balance Analysis lié au réseaux métaboliques car elle correspond aux hypothèses posées précédemment relatives à la loi de conservation de masse forçant le flux d'entrée d'un métabolite tel qu'il soit égal au flux de sortie, et à l'état stationnaire interdisant l'accumulation d'élément dans le réseau. Pour cela, les quantités de métabolites nécessaires dans une réaction, la stœchiométrie, sont représentées avec la matrice stœchiométrique donnée dans la [Définition 2](#). Ainsi la loi de conservation de masse et l'état stationnaire du réseau crée la contrainte que la somme des flux entrants et sortants d'un nœud métabolique est nulle :

$$\sum_{j \in J} \mathcal{M}_{G_{ij}} \cdot v_j = 0 \quad \forall i \in I.$$

**Fonction objectif du problème de Flux Balance Analysis** Afin de diriger l'étude des flux dans un réseau métabolique, une fonction objectif est ajoutée au programme linéaire de FBA. Cette dernière peut modéliser plusieurs comportements biologiques possibles dépendant de l'étude suivie, par exemple, la minimisation ou maximisation de la production d'énergie (métabolites d'ATP) [10]. Dans notre thèse, nous nous intéressons essentiellement à la production de métabolites cibles. Dans le cas de l'étude en flux, ces derniers sont regroupés dans une réaction dite de biomasse [27] représentant le poids sec d'une cellule de l'organisme, c'est-à-dire la somme des quantités de ses

composants les plus importants. Pour déterminer sa composition, des techniques de biologie peuvent être utilisées permettant de mesurer la concentration des différents métabolites dans l'organisme ou il est possible de se baser sur des biomasses définies pour des espèces proches de celle étudiée. L'étude de flux a donc pour but de maximiser le flux d'une réaction particulière, la réaction de biomasse, modélisée comme suit :

$$\text{Maximiser } z = v_b.$$

### 2.2.3 Le problème complet de Flux Balance Analysis

Le problème complet de Flux Balance Analysis est donc uniquement composé des contraintes de la conservation de masse, de l'état stationnaire, et des limites de flux dans le réseau métabolique et d'une fonction d'optimisation guidant la résolution. Nous cherchons alors à vérifier que le problème est satisfiable c'est-à-dire qu'il existe au moins une solution à ce dernier.

**Définition 3 (Flux Balance Analysis)** *Un réseau métabolique  $G = (R \cup M, E, s)$  vérifie le problème de Flux Balance Analysis associé à la biomasse  $v_b$  s'il existe au moins une solution à l'ensemble des équations suivantes, dont les variables sont les  $v_j$ , pour  $j \in R$ .*

$$\begin{aligned} & \text{Maximiser } z = v_b \\ & \text{subject to :} \\ & \sum_{j \in J} \mathcal{M}_{G_{ij}} v_j = 0 \quad \forall i \in I \\ & LB_j \leq v_j \leq UB_j \quad \forall j \in J \\ & v_j \in \mathbb{R} \quad \forall j \in J \end{aligned}$$

La solution du problème de *Flux Balance Analysis* peut être obtenue de façon automatique grâce à des logiciels, dont le nombre ne cesse d'augmenter [65], développés pour des études en FBA sur des réseaux métaboliques tels que OptFlux [86] ou MetaNetX [32]. L'un des moyens les plus utilisés est le package python *cobraPy*, développé par Ebrahim et al [23], permettant de récupérer la valeur du flux passant dans la réaction de biomasse  $v_b$  en utilisant des solveurs tels que glpk [73] ou cplex [43].

Le méthode de *Flux Balance Analysis* est habituellement utilisée pour vérifier si un flux de biomasse est possible dans les conditions imposées dans le réseau métabolique modélisé. Ainsi, cette technique peut être un moyen de valider un réseau métabolique [22] ou de vérifier la robustesse de ce dernier lorsque des réactions métaboliques sont bloquées ou le milieu de croissance modifié [82].

## 2.3 Flux Variability Analysis

Si la méthode de *Flux Balance Analysis* permet de calculer des flux dans un réseau pour obtenir une solution de production maximale des métabolites de la biomasse, il est possible par une méthode proche d'étudier ce réseau lorsqu'il produit une biomasse optimale. En effet, si le problème de FBA retourne une solution de production optimale de biomasse, des expériences *in vivo* montrent que les organismes possèdent de multiples voies de production pouvant être plus ou moins utilisées pour produire une biomasse optimale [40]. Il est alors intéressant de définir les limites effectives des flux des réactions dans le réseau permettant une conservation du flux optimal de biomasse. Ce nouveau problème est appelé *Flux Variability Analysis*, développé par Gudmundsson et Thiele dans [36], dont le concept est l'étude de l'ensemble des solutions optimales possibles au problème de FBA défini par la [Définition 3](#).

### 2.3.1 Le problème de Flux Variability Analysis : une variante du problème de *Flux Balance Analysis*

Le but de cette étude en *Flux Variability Analysis* est de définir la variation de flux possible dans chacune des réactions du réseau métabolique tout en prenant en compte la production maximale de la biomasse. Pour cela, dans un premier temps, la valeur maximale du flux présent dans la réaction de biomasse est calculée en utilisant le problème de *Flux Balance Analysis* de la [Définition 3](#) sur

le réseau entier étudié.

### Contraintes du problème de Flux Variability Analysis

Dans un second temps, le problème de *Flux Variability Analysis* étudie chaque flux  $v_r$  de réaction  $r \in J$  qui fait alors l'objet de deux programmes linéaires proches afin de calculer les limites inférieure et supérieure de ce dernier. Pour cela, les contraintes du problème en FBA sont réutilisées pour contraindre le système au niveau de l'état stationnaire et la conservation des masses ainsi que les limites des autres flux du réseau.

$$\begin{aligned} \sum_{j \in J} \mathcal{M}_{G_{ij}} \cdot v_j &= 0 & \forall i \in I. \\ LB_j \leq v_j \leq UB_j & & \forall j \in J \setminus r, \\ v_j \in \mathbb{R} & & \forall j \in J \end{aligned}$$

Une contrainte sur le flux de la réaction de biomasse  $v_b$  est ajoutée afin de fixer ce dernier à la valeur du flux optimal  $v^{max}$  de la biomasse calculée précédemment par le problème de FBA ([Définition 3](#)) sur l'ensemble du réseau.

$$v_b = v^{max}$$

### Fonction objectif du problème de Flux Variability Analysis

Le problème de FVA fait intervenir deux problèmes linéaires afin de calculer respectivement la limite inférieure et celle supérieure du flux  $v_r$  de chacune des réactions. Si ces problèmes possèdent les mêmes contraintes, ils diffèrent au niveau de l'optimisation car, pour calculer la limite inférieure, le premier problème minimise le flux de la réaction sous contraintes définies précédemment tandis que le second maximise ce flux dans le but d'obtenir la limite supérieure.

$$\text{minimiser/maximiser } z = v_r$$

### 2.3.2 Le problème complet de Flux Variability Analysis

Le problème de *Flux Variability Analysis* est un problème dont la résolution se déroule en plusieurs étapes. En effet, afin de le résoudre, la valeur optimale du flux de biomasse dans le réseau complet étudié doit dans un premier temps être calculée notamment grâce au FBA ([Définition 3](#)). Puis pour chacune des réactions  $r$  du réseau, les limites inférieure et supérieure sont calculées grâce au programme linéaire défini dans la [Définition 4](#).

**Définition 4 (Flux Variability Analysis)** Soit  $v^{max}$ , la valeur optimale de la biomasse calculée par la [Définition 3](#) dans le réseau complet,  $f \in [0;1]$  coefficient de biomasse. Le problème de FVA permettant de calculer les limites inférieure et supérieure pour le flux  $v_r$  de la réaction  $r$  appartenant au graphe  $G = (R \cup M, R, s)$  est défini par :

$$\begin{aligned} &\text{minimiser/maximiser } z = v_r \\ &\text{subject to :} \\ &\sum_{j \in J} \mathcal{M}_{G_{ij}} \cdot v_j = 0, & \forall i \in I \\ &LB_j \leq v_j \leq UB_j, & \forall j \in J \setminus r \\ &v_b = f \cdot v^{max} \\ &v_j \in \mathbb{R}, & \forall j \in J \end{aligned}$$

Il est possible d'utiliser la méthode de *Flux Variability Analysis* via plusieurs implémentations, utilisées soit sur des plateformes comme Matlab soit par elles-mêmes. Un exemple d'implémentation est fastFVA [35] utilisant une compilation Matlab et les solveurs linéaires GLPK ou CPLEX.

Que ce soit la méthode de *Flux Balance Analysis* ou celle de *Flux Variability Analysis*, le calcul des flux dans un réseau métabolique permet de comprendre les mécanismes de l'organisme étudié en mettant en avant des voies métaboliques de production de métabolites cibles définies dans la réaction de biomasse.



## 2.4 Productibilité stoechiométrique de composés métaboliques

### 2.4.1 Une notion de productibilité stoechiométrique

La résolution des problèmes de flux de FVA et FBA permet de définir l'ensemble des réactions utilisables par l'organisme permettant d'obtenir un flux dans la réaction de biomasse et les métabolites productibles lors de cette production de biomasse. Ainsi une réaction est utilisable si il existe une solution en FBA telle que son flux ne soit pas nul ou, en FVA, qu'une des limites du flux correspondant à cette réaction ne soit pas égale à zéro. Ces analyses permettent aussi de déduire le sens d'utilisation de la réaction dans le cas des réactions réversibles, pouvant se dérouler dans les deux sens : si le flux de la réaction est positif, alors elle est utilisée dans son sens original tandis que si le flux est négatif, la réaction est prise dans son sens inverse (les réactants deviennent produits et vice versa).

Les spécifications des problèmes de FBA et FVA permettent de définir précisément dans quel cadre on considère qu'une réaction a la capacité d'être activée dans un réseau métabolique. Nous désignons cela par le terme d'"accessibilité".

**Définition 5 (Accessibilité stoechiométrique d'une réaction)** Soit  $G = (R \cup M, E, s)$  un réseau métabolique de matrice stoechiométrique  $M_G$ . Une réaction  $r \in R$  est dite accessible stoechiométriquement si et seulement s'il existe un ensemble de flux de réactions  $\{v_j, j \in R\}$  réponse du programme linéaire défini dans la [Définition 3](#) tel que  $v_r \neq 0$ .

En ce qui concerne l'accessibilité des métabolites en étude de flux, un élément est dit productible s'il est le produit d'une réaction accessible, donc dont le flux n'est pas nul pour au moins une solution au problème de FBA.

**Définition 6 (Accessibilité stoechiométrique d'un métabolite)** Soit  $G = (R \cup M, E, s)$  un réseau métabolique de matrice stoechiométrique  $M_G$ . Un métabolite  $m \in M$  est dit accessible stoechiométriquement s'il existe une réaction  $r \in R$  qui est elle-même accessible stoechiométriquement et tel que  $m \in prds(r)$ .

L'étude des flux au sein du réseau métabolique permet de définir quelles réactions peuvent être utilisées dans le réseau via la notion d'accessibilité et en calculant les métabolites productibles de ce dernier. Cependant, la prise en compte de la stoechiométrie et l'utilisation des flux avec les hypothèses d'état stationnaire et de conservation de masse posent des limites à cette notion de productibilité dans le réseau métabolique.

### 2.4.2 Exemple et limites d'une étude stoechiométrique

Une étude des flux d'un réseau métabolique permet de comprendre la dynamique de production des métaboliques dans l'organisme. Cependant, il est nécessaire de poser deux hypothèses afin de pouvoir résoudre le problème modélisant cette étude, celles-ci empêchant l'accumulation des métabolites dans le réseau, impliquant que le flux entrant dans un nœud métabolite est égal au flux sortant de ce dernier. Cette contrainte limite alors les réactions pouvant être utilisées dans le réseau d'un point de vue stoechiométrique.

#### Des voies métaboliques inaccessibles par manque d'information

L'accessibilité d'une réaction et la productibilité d'un métabolite sont dépendantes de la stoechiométrie du réseau métabolique et sont donc sensibles aux informations erronées ou insuffisantes liées aux quantités de matière nécessaires aux réactions. En effet, certaines réactions constituant ces voies métaboliques n'ont pas de stoechiométrie précisée par manque de connaissances ou cette dernière ne permet pas l'activation de la réaction car les besoins en métabolites substrats ne sont pas satisfaits. Une autre raison de l'absence de flux dans une voie métabolique peut aussi être la production de métabolite ne pouvant pas être ensuite consommé que ce soit par une réaction de l'organisme ou par une réaction d'export, entraînant une accumulation de cet élément si la voie est utilisée, allant à l'encontre des hypothèses de conservation de masse et d'état stationnaire posées

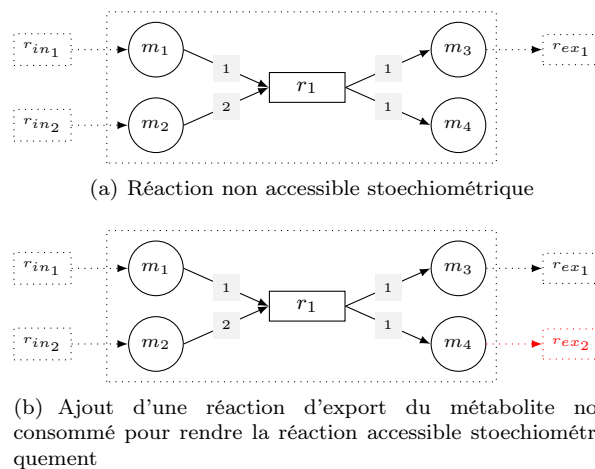


FIGURE 2.1 – **Exemple de réaction non accessible d'un point de vue stoechiométrique**  
 Dans la sous-figure (a), la réaction  $r_1$  est inutilisable car l'un de ses produits n'est pas consommé par une autre réaction. Une correction habituellement utilisée est présentée dans la sous-figure (b) avec la création d'une réaction d'export

dans le cas d'une étude de flux.

L'exemple proposé dans la Figure 2.1 montre une réaction non accessible stoechiométriquement car le produit  $m_4$  n'est pas consommé que ce soit par une réaction métabolique ou une réaction d'export. L'activation de la réaction  $r_1$  entraînerait alors une accumulation de ce composant, non permise par les hypothèses d'état stationnaire et de conservation de masse. La correction simple de ce réseau permettant d'utiliser la réaction  $r_1$ , montrée dans la Figure 2.1, est l'utilisation d'une réaction d'export consommant le métabolite faisant défaut  $m_4$ . Certains logiciels d'analyse présentés dans la section 2.2.3 comme *cobraPy* [23] permettent de rechercher les réactions présentant ce problème de produit non consommé.

### L'accessibilité au niveau d'un cycle métabolique

Un cas particulier dans le réseau métabolique qui est intéressant à étudier est le cas où les flux forment un cycle dans le réseau métabolique. Dans le cas d'une étude de flux, un cycle est accessible stoechiométriquement si la stoechiométrie des réactions constituant ce dernier est finement calculée, n'entraînant pas d'accumulation de métabolites : les quantités créées de métabolites internes dans le cycle sont égales à celles consommées de chacun d'eux.

Un exemple de cycle est présenté dans la Figure 2.2 où nous pouvons voir que la stoechiométrie du cycle ne permet pas à la fois de continuer ce dernier et de produire un élément ciblé  $t$ . En effet, le cycle est composé de 3 réactions métaboliques dont la deuxième  $r_2$  produit un métabolite  $m_2$  devant être consommé par la troisième réaction  $r_3$  afin de produire un substrat  $m_3$  de la première  $r_1$ . Cependant une autre réaction,  $r_4$ , nécessite aussi un métabolite  $m_2$  pour produire un métabolite cible  $t$ . La compétition entre les deux réactions entraîne une impossibilité de produire l'élément ciblé. La sous-figure (b) de la Figure 2.2 montre que le changement de pondération de l'arc liant  $r_2$  à  $m_2$  et signifiant que la réaction  $r_2$  produit alors 2 métabolites  $m_2$  permet ainsi d'accéder aux deux réactions  $r_3$  et  $r_4$ , l'une continuant le cycle et l'autre produisant la cible. Nous pouvons donc voir par cet exemple que la stoechiométrie dans un cycle métabolique doit être finement étudiée.

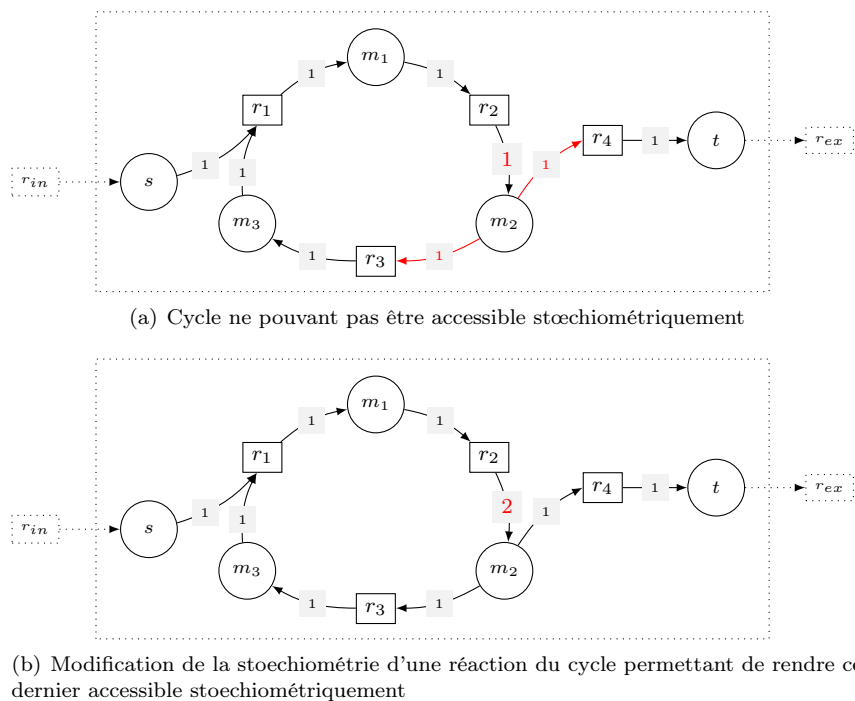


FIGURE 2.2 – **Exemple de cycle inaccessible en étude de flux** Le cycle de la sous-figure (a) nécessite un métabolite  $m_2$  pour continuer son déroulement mais la réaction  $r_4$  a également besoin de ce composant pour produire la cible  $t$ . Or un seul métabolite  $m_2$  est produit obligatoirement utilisé pour continuer le cycle. La sous-figure (b) montre que le simple changement dans la pondération de la réaction  $r_2$  permet une utilisation du cycle et la production du métabolite  $t$

## 2.5 Productibilité topologique de composés métaboliques

### 2.5.1 Un point de vue topologique d'un réseau métabolique

Les réseaux métaboliques sont représentés par des graphes orientés bipartis et essentiellement modélisés afin d'étudier la production de métabolites cibles. Nous avons précédemment montré les méthodes d'étude de flux dans un réseau métabolique et leurs limites liées à la stœchiométrie des réactions. Ces dernières peuvent être contournées en faisant le choix d'une étude topologique où nous nous concentrons sur les chemins entre les métabolites sources et ceux cibles correspondant à des voies métaboliques de production.

Dans une étude topologique d'un réseau métabolique, la stœchiométrie n'étant pas prise en compte, les arcs du graphe correspondant au réseau ne sont pas étiquetés comme nous pouvons le voir dans la [Définition 7](#).

**Définition 7 (Graphe métabolique sans stœchiométrie)** On désigne par "graphe métabolique sans stœchiométrie" l'ensemble  $G = (R \cup M, E)$  qui vérifie les spécifications suivantes :

- $R$  ensemble des réactions
- $M$  ensemble des métabolites
- $m \in M, r \in R, \begin{cases} (m, r) \in E \Leftrightarrow \text{le métabolite } m \text{ est substrat de la réaction } r \\ (r, m) \in E \Leftrightarrow \text{le métabolite } m \text{ est produit de la réaction } r \end{cases}$

Comme pour une étude de flux, le but d'une étude topologique est l'analyse des éléments productibles dans le réseau métabolique, ces derniers sont regroupés dans la littérature sous le terme d'ensemble scope.

### 2.5.2 Scope de métabolites sources dans un réseau métabolique

Le scope d'un ensemble de métabolites sources dans un réseau métabolique, noté  $\sum(G, S)$ , correspond à l'ensemble des métabolites productibles à partir des nutriments du milieu dans le réseau d'un point de vue topologique. Ce concept est lié à la notion d'expansion de réseau métabolique développée par Handorf et al [38] ainsi que Kruse et Ebenhöh [62] se reposant sur des notions biologiques pour définir les éléments productibles appartenant au scope des nutriments du réseau. Ainsi, un métabolite va être considéré comme productible s'il existe une réaction possible dans le réseau métabolique produisant ce métabolite tandis qu'une réaction est possible si l'ensemble de ses métabolites substrats est productible. D'un point de vue topologique, en se basant sur les notions biologiques, un nœud métabolite est accessible si un des nœuds prédécesseurs, réactions, est accessible topologiquement alors que un nœud réaction est accessible dans le graphe si l'ensemble des nœuds prédécesseurs, métabolites, est accessible. Afin de se rapprocher des termes biologiques, les prédécesseurs d'un nœud réaction sont appelés réactants et notés  $rcts(r)$  tandis que ses successeurs sont nommés produits :  $prds(r)$ .

**Définition 8 (Réactants et produits d'une réaction)** Soit  $G = (R \cup M, E)$ , le graphe d'un réseau métabolique,

$$\forall r \in R, \begin{cases} rcts(r) = m \in M \mid (m, r) \in E \\ prds(r) = m \in M \mid (r, m) \in E \end{cases}$$

La définition du scope d'un ensemble de métabolites sources dans un réseau métabolique [38],[62] rassemble les métabolites accessibles dans le réseau à partir de l'ensemble source par récursion. En effet, le scope initial est composé des métabolites sources puis les produits des réactions dont l'ensemble des réactants appartiennent à l'ensemble initial sont ajoutés au scope. La boucle de récursivité ajoute alors les métabolites produits des réactions dont les substrats sont dans le scope mis à jour à chaque tour jusqu'à l'impossibilité d'ajouter de nouveaux éléments à l'ensemble scope (Définition 9).

**Définition 9 (Scope de métabolites sources)** Soit  $G = (R \cup M, E)$ , le graphe d'un réseau métabolique. Soit  $S \subset M$ , un ensemble de nœuds source. L'ensemble scope, noté  $\sum(G, S)$ , des métabolites accessibles dans le réseau à partir de  $S$  est défini par

$$\sum(G, S) = \cup_i M_i \text{ avec } \begin{cases} M_0 = S \\ M_0^r = \{r \in R \mid rcts(r) \subset M_0\} \\ M_{i+1} = \{m \in M \mid \exists r \in M_i^r \ \& \ m \in prds(r)\} \cup M_i \\ M_{i+1}^r = \{r \in R \mid rcts(r) \subset M_{i+1}\} \setminus M_i^r \end{cases}$$

Cette définition du scope d'un ensemble source dans un graphe métabolique suit le concept défini par Handorf et al dans [39] qui présenté également un algorithme calculant un ensemble minimal de nœuds métabolites sources permettant d'accéder à l'ensemble des nœuds métabolites du graphe. Ceci correspond à trouver un ensemble minimal de nutriments pour produire l'ensemble des métabolites du réseau métabolique, d'un point de vue topologique.

À partir de ce concept d'ensemble des métabolites accessibles, il est possible de définir une notion de productibilité topologique au sein d'un graphe métabolique, différente dans le cas d'un métabolite ou celui d'une réaction. D'une part, dans une étude topologique d'un réseau métabolique, une réaction est considérée comme accessible si l'ensemble des substrats de cette dernière est accessible donc dans le scope des métabolites sources, définition proposée dans [30].

**Définition 10 (Accessibilité topologique d'une réaction)** Soit  $G = (R \cup M, E)$  un réseau métabolique. Soit  $S \subset M$  ensemble de métabolites sources. Une réaction  $r \in R$  est dite accessible topologiquement si et seulement  $rcts(r) \subset \sum(G, S)$ .

D'autre part, un métabolite est dit accessible topologiquement si il appartient au scope de l'ensemble source dans le réseau étudié ou produit d'une réaction accessible topologiquement.

**Définition 11 (Accessibilité topologique d'un métabolite)** Soit  $G = (R \cup M, E)$  un réseau métabolique. Soit  $S \subset M$  ensemble de métabolites sources. Un métabolite  $m \in M$  est dit accessible topologiquement si  $m \in \sum(G, S)$ .

### 2.5.3 Calcul automatique du scope d'un réseau métabolique par programmation par ensemble réponse

Schaub et Thiele ont proposé dans [89] une méthode de calcul automatique permettant de récupérer l'ensemble scope regroupant les métabolites accessibles à partir de nutriments dans un réseau métabolique mais également de calculer l'ensemble minimal de sources pour produire un ensemble de métabolites cibles. Ils utilisent pour cela une technologie de résolution déclarative par ensemble réponse appelé ASP, *Answer Set Programming* [7],[34]. Cette dernière propose un langage de haut niveau permettant de modéliser des problèmes combinatoires dans un cadre déclaratif et possède des solveurs dont la haute performance permet une résolution rapide. Ainsi ASP permet aux utilisateurs de se concentrer sur la modélisation du problème au lieu de la résolution de ce dernier, l'idée principale étant d'exprimer le problème sous forme de modèle logique inspiré par les approches SAT, liées aux problèmes de satisfaisabilité booléenne, et les bases de données. La résolution de ces problèmes logiques avec ASP permet de déterminer s'il existe une solution et ainsi obtenir un des ensembles-réponses ou Answer Set mais également, avec la combinaison des modes de résolution supportés, l'union et l'intersection de ces ensembles ainsi que l'ensemble des ensembles-réponses d'un problème modélisé grâce à ASP.

Comme nous venons de le décrire, la modélisation avec *Answer Set Programming* repose sur de la programmation logique définissant des règles proches de celles utilisées en Prolog de la forme :

$$\mathbf{h} \text{ :- } a_1, \dots, a_m, \text{not } a_{m+1}, \dots, \text{not } a_n .$$

dans laquelle les termes  $\mathbf{h}$  et  $a_i$  sont des littéraux pouvant prendre une valeur booléenne et **not** est la négation par défaut en programmation ASP. Les littéraux sont des prédicats, se présentant ainsi  $\text{nom\_predicat}(a_1, \dots, a_n)$ , dont les arguments sont des atomes constants (entier, chaîne de caractère,...), commençant par une lettre minuscule, ou des variables d'un domaine fini, dont le nom commence par une majuscule ou est un underscore (variable non spécifiée car non décisive pour la résolution). Une règle permet de définir l'affectation de la tête en fonction du corps ainsi  $\mathbf{h}$  est avéré vrai si  $a_1, \dots, a_m, \text{not } a_{m+1}, \dots, \text{not } a_n$  est satisfait, c'est-à-dire, si les littéraux  $a_1, \dots, a_m$  sont vrais et que rien ne prouve que les  $a_{m+1}, \dots, a_n$  sont vrais. L'affectation, vrai ou faux, de  $\mathbf{h}$  constitue la solution du problème et est donc un ensemble-réponse. Si le corps de la règle est vide :

$$\mathbf{h} .$$

alors le prédicat est un fait car toujours vrai représentant une donnée du problème combinatoire étudié. A l'inverse, si la tête de la règle est vide :

$$\text{: - } a_1, \dots, a_m, \text{not } a_{m+1}, \dots, \text{not } a_n .$$

alors celle-ci décrit une situation impossible dans l'ensemble solution. Par exemple, si la règle est de la forme **: - a, not b**. alors, dans la solution, si le littéral **a** est vrai, **b** ne peut pas être faux donc est vrai. L'ensemble des règles permet de modéliser les faits et contraintes du problème combinatoire étudié afin d'utiliser un solveur.

La technologie *Answer Set Programming* est une méthode récente de résolution de problèmes combinatoires à grande échelle. Nous avons choisi de travailler avec la suite logicielle Potassco, notamment le solveur *Clasp*, qui ne cesse d'évoluer depuis le début de cette thèse en présentant de nouvelles heuristiques de résolution, de nouveaux solveurs, de nouvelles syntaxes.

Un réseau métabolique représenté sous forme de graphe modélise les chemins des voies métaboliques permettant par différentes méthodes, notamment par programmation d'ensemble-réponse, d'analyser les possibilités de production et d'utilisation des métabolites et réactions. Mais ces méthodes bien que permettant de contourner les limites des études de flux, possèdent leurs propres limites.

### 2.5.4 Exemple et limites d'une étude topologique

Une étude de productibilité au niveau topologique d'un réseau métabolique a pour but de calculer le scope d'un ensemble de métabolites sources. Le résultat est donc dépendant de l'ensemble de nœuds sources définis initialement.

### Des éléments accessibles topologiquement selon l'ensemble source

L'ensemble source d'un réseau métabolique défini pour étudier topologiquement ce dernier est créé à partir des nutriments du milieu de croissance de l'organisme modélisé par le réseau et a un impact direct sur les résultats obtenus par cette étude. En effet, l'ajout d'un métabolite dans l'ensemble source peut débloquent une partie plus ou moins importante du réseau métabolique tandis que la suppression d'un élément peut empêcher l'utilisation de certains chemins. Le calcul d'un ensemble minimal de sources est donc important dans le but de produire un ensemble de métabolites cibles.

Dans l'exemple présenté dans la Figure 2.3, l'algorithme de calcul du scope de l'ensemble source  $S$  constitué du singleton  $m_2$  est déroulé étape par étape. L'initialisation  $\sum(G, S)_0$  regroupe le métabolite constituant la source,  $m_2$ , ainsi que les réactions accessibles topologiquement à partir de cet ensemble initial (dont le substrat unique est  $m_2$ ) :  $r_2$  et  $r_3$  définissant  $\sum(G, S)_0 = \{m_2, r_2, r_3\}$ . L'étape suivante,  $\sum(G, S)_1$ , ajoute les produits des réactions appartenant au scope :  $m_4, m_5, m_6$  puis les réactions dont les réactants sont alors dans  $\sum(G, S)_1$  :  $r_4, r_5$ , mettant à jour  $\sum(G, S)_1 = \{m_2, m_4, m_5, m_6, r_2, r_3, r_4, r_5\}$ . Une nouvelle étape permet d'étendre le scope avec les métabolites  $m_8, m_9$  et  $m_{10}$  et l'algorithme de calcul se finit alors car aucune nouvelle réaction n'est accessible topologiquement, fixant le scope final à  $\sum(G, S)_1 = \{m_2, m_4, m_5, m_6, m_8, m_9, m_{10}, r_2, r_3, r_4, r_5\}$ . Dans cet exemple, nous pouvons voir que des éléments restent inaccessibles topologiquement si l'ensemble source  $S$  est le singleton  $m_2$  mais l'ajout, d'une part, du métabolite  $m_1$  à  $S$  permettant d'accéder à la réaction  $r_1$  puis au métabolite  $m_3$  et, d'autre part, du métabolite  $m_7$ , lié à la réaction  $r_6$  produisant le métabolite  $m_{11}$ , rend possible l'accessibilité totale du réseau métabolique.

### L'accessibilité topologique d'un cycle métabolique

Un réseau métabolique comportant habituellement un cycle, il est intéressant d'étudier le calcul du scope au sein de ce réseau. Dans le cas d'une étude topologique, un cycle est utilisable si une des réactions le composant est accessible donc si les substrats d'une des réactions sont dans le scope de l'ensemble des métabolites sources du réseau.

Nous présentons un exemple de cycle non accessible topologiquement dans la Figure 2.4, dans lequel l'ensemble source  $s$  est constitué du singleton  $\{s\}$ . Nous pouvons voir dans la sous-figure (a) que le cycle n'est pas utilisable car aucune réaction n'est accessible topologiquement à partir de  $S$ , réduisant le scope  $\sum(G, S)$  au singleton  $\{s\}$  mais l'ajout de  $m_3$  dans l'ensemble source  $S$ , Figure (b), permet d'accéder à la réaction  $r_1$ , initialisant ainsi le cycle et permettant d'ajouter l'ensemble des métabolites et des réactions au scope du nouvel ensemble de sources :  $\sum(G, S) = \{s, m_1, m_2, m_3, t, r_1, r_2, r_3, r_4\}$ . L'accessibilité topologique dans un cycle nécessite donc parfois l'ajout de métabolites sources afin d'initialiser ces derniers montrant les limites de l'étude topologique [19, 90, 1].

Nous avons mis en avant une notion de productibilité pouvant être stœchiométrique ou topologique présentant des caractéristiques différentes et ayant des limites respectives dans un réseau métabolique. Il est donc intéressant de comparer les deux approches dans l'étude d'un réseau métabolique.

## 2.6 Complémentarité des études de flux et topologique

Nous avons vu dans les sections 2.4 et 2.5 qu'un graphe modélisant un réseau métabolique peut être étudié de différents points de vue, que ce soit au niveau stœchiométrique via des études de flux (FBA ou FVA), ou au niveau topologique avec la notion de scope d'ensemble source. Les méthodes utilisées ont chacune leurs atouts et leurs limites pouvant être comparées.

Kruse and Ebenhöh [62] ainsi que Prigent et al [83] comparent dans leurs travaux les deux approches de productibilité en étudiant les éléments productibles dans le cas d'une étude stœchiométrique et celui d'une approche topologique dans des graphes modélisant les mêmes réseaux métaboliques.

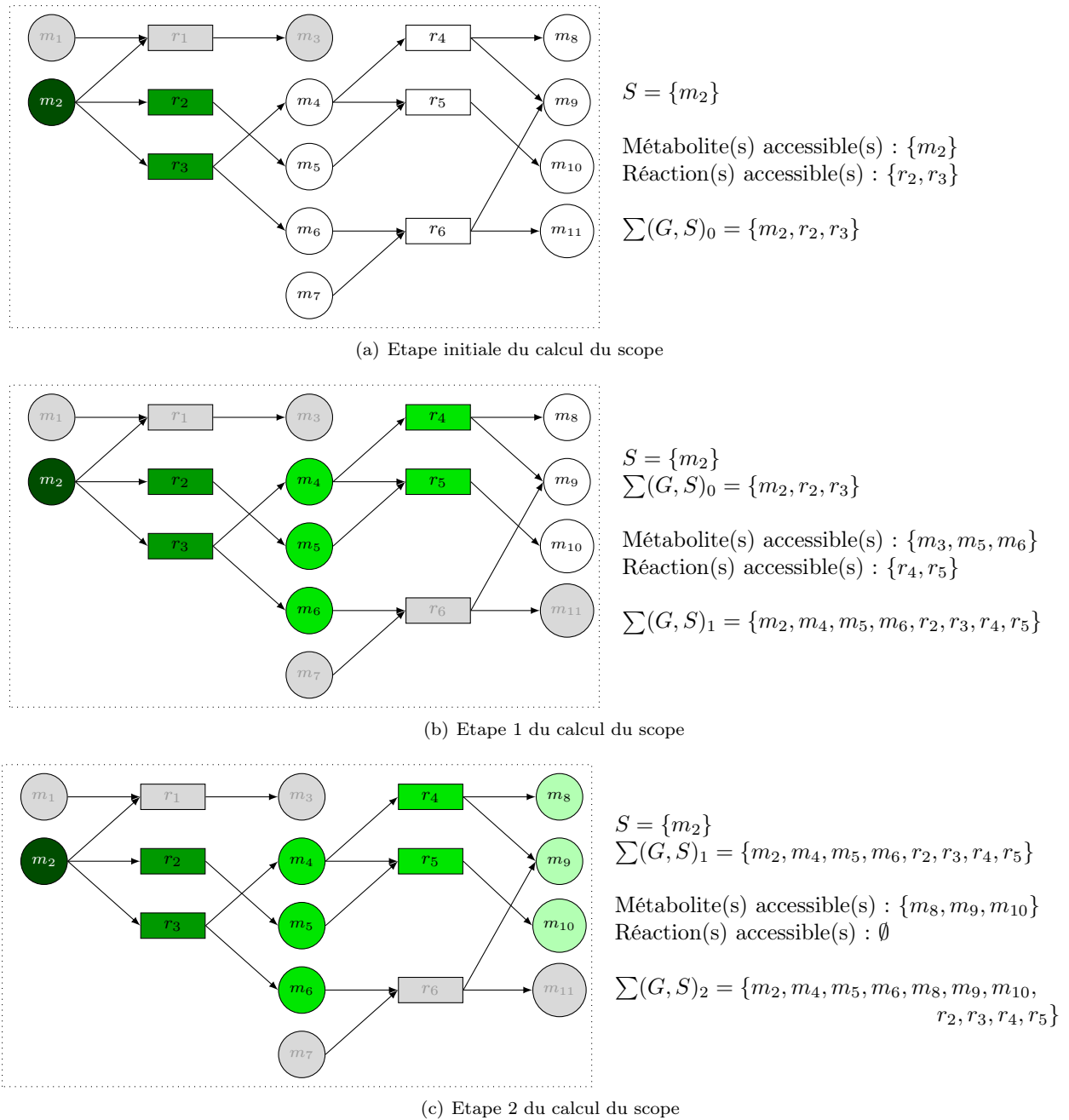


FIGURE 2.3 – Exemple de calcul de scope dans un réseau métabolique

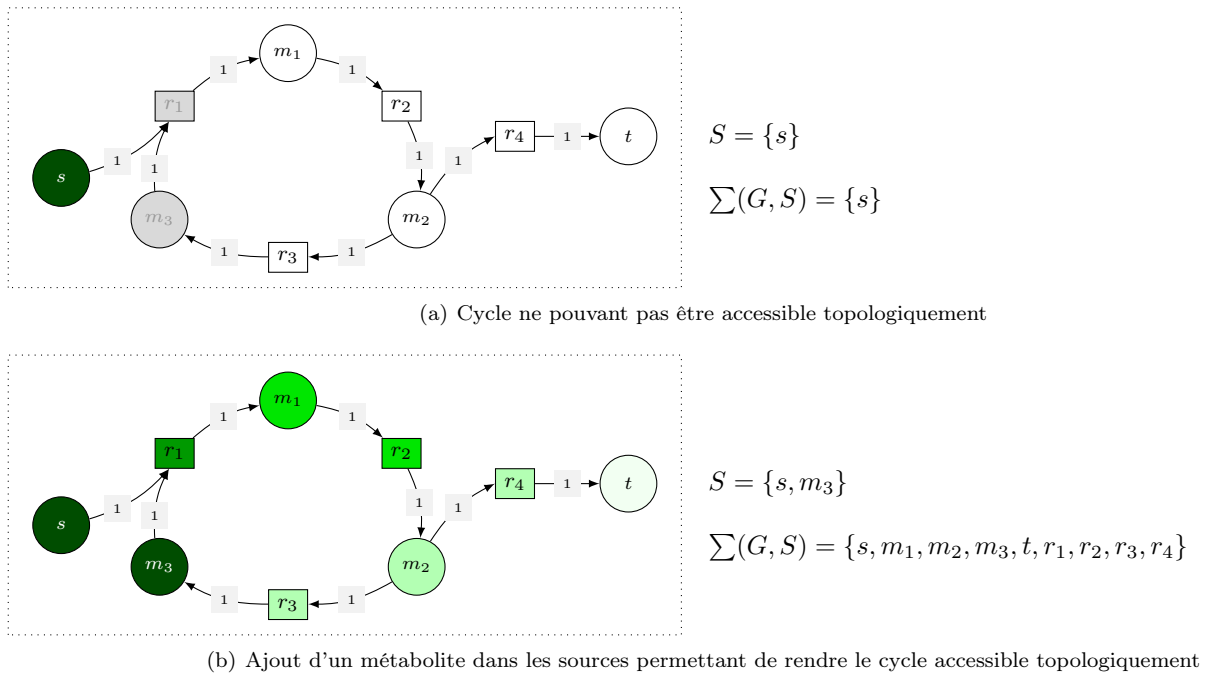


FIGURE 2.4 – **Exemple de cycle inaccessible en étude topologique** Le cycle de la sous-figure (a) nécessite, d'un point de vue stœchiométrique, un métabolite  $m_2$  pour continuer son déroulement mais la réaction  $r_4$  a également besoin de ce composant pour produire la cible  $t$ . Or un seul métabolite  $m_2$  est produit obligatoirement utilisé pour continuer le cycle. La sous-figure (b) montre que, d'un point de vue topologique, le simple ajout du métabolite  $m_3$  dans l'ensemble des sources permet une utilisation du cycle et la production du métabolite  $t$

La Figure 2.5 est tirée des travaux de Prigent et al [83] déterminant les métabolites productibles dans les trois réseaux métaboliques présentés selon la méthode d'étude topologique ou stœchiométrique. Chacun des réseaux présente un cas particulier mettant en évidence les différences entre productibilité topologique et productibilité stœchiométrique.

Le premier réseau, composé d'un métabolite source  $S$ , de quatre métabolites intermédiaires ( $a, b, c, d$ ) et du métabolite cible  $T_1$ , montre l'influence de la stœchiométrie dans la productibilité des métabolites en posant une variable  $n$  pouvant prendre plusieurs valeurs comme stœchiométrie d'une réaction. D'un point de vue topologique, l'ensemble des métabolites est accessible quelle que soit la valeur de  $n$  car la stœchiométrie n'est pas prise en compte. D'un point de vue des flux, lié à la stœchiométrie, nous pouvons voir que la valeur de  $n$  modifie l'ensemble de métabolites productibles. En effet, l'équilibre des masses devant être vérifié, pour pouvoir produire  $d$  et  $c$  puis  $T_1$ , deux métabolites  $a$  doivent être consommés donc produits dans exactement la même quantité, expliquant la productibilité stœchiométrique des métabolites intermédiaires  $a, b, c, d$  dans uniquement le cas où  $n=2$ . Nous voyons dans cet exemple les contraintes imposées par les hypothèses liées aux flux (Définition 3) pouvant être contournées par une étude en topologie.

Le deuxième réseau métabolique de la Figure 2.5 présente le cas d'un métabolite hors cible non consommé par une réaction ( $f$ ) dans la production du métabolite cible  $T_2$  à partir de  $S$  via trois réactions impliquant les métabolites  $e, f$  et  $g$ . Dans ce réseau comme dans le précédent, tous les métabolites sont productibles topologiquement mais le passage en stœchiométrie entraîne une impossibilité de production dans ce dernier à cause de la non consommation du métabolite  $f$ . En effet, un métabolite ne pouvant pas être accumulé dans une étude stœchiométrique, le métabolite  $f$  ne peut pas être produit donc la réaction consommant  $e$  est inactive, entraînant une inactivation de la réaction produisant ce dernier. Aucun métabolite n'est donc productible stœchiométriquement. Cet exemple montre l'importance de la gestion des métabolites secondaires dans une étude de flux pour la production de métabolites cibles, ne posant aucun problème dans une étude topologique.



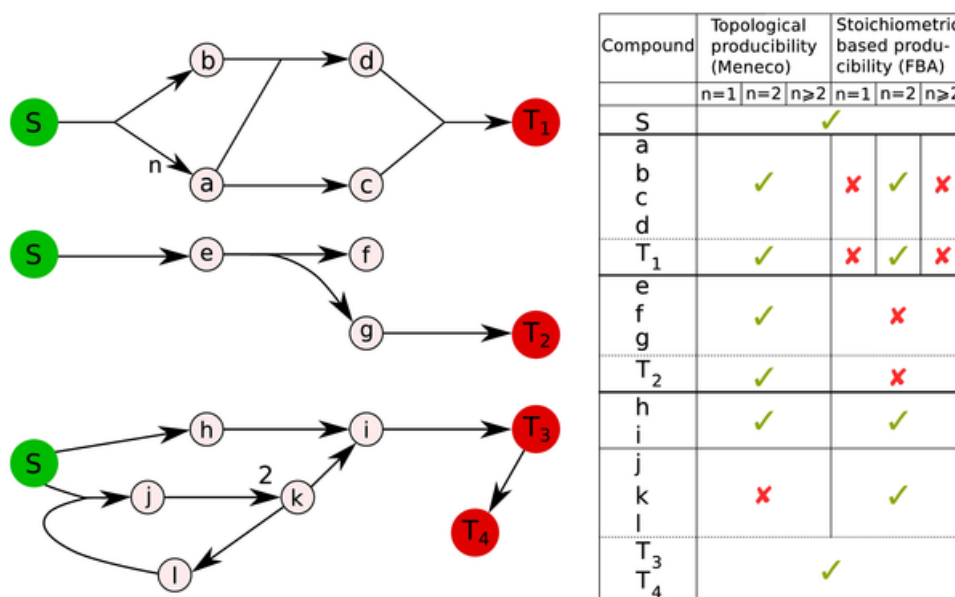


FIGURE 2.5 – **Comparaison de la productivité stœchiométrique et topologique dans trois réseaux métaboliques** Figure tirée de [83]. Les réseaux métaboliques sont représentés sous forme de graphe uniparti dans lequel un nœud représente un métabolite et un arc, une réaction. Chaque métabolite est testé s'il est productible topologiquement puis stœchiométriquement. Le premier réseau montre l'impact de la stœchiométrie selon la valeur de la pondération  $n$  correspondant à la stœchiométrie de la réaction produisant  $a$  et  $b$  à partir de  $S$  dans le cas du premier réseau. Le deuxième réseau montre un cas de non consommation de métabolite tandis que le troisième possède un cycle.

Le troisième et dernier exemple de réseau métabolique dans la Figure 2.5 présente le cas de deux voies de production dont l'une est constituée d'un cycle. Dans ce cas, des éléments ne sont pas accessibles topologiquement à cause du cycle. Comme nous l'avons observé dans l'exemple de la Figure 2.4, dans une étude topologique, aucune des réactions du cycle n'étant accessible, ce dernier n'est pas initialisé. Cependant la seconde voie de production permet la production topologique des métabolites  $h, i$  et des cibles  $T_3, T_4$ . Dans le cas d'une étude de flux, comme présentée dans l'exemple de la Figure 2.2, le cycle est utilisable car le fait de produire deux métabolites  $k$  permet d'activer la réaction continuant le cycle et produisant  $l$  ainsi que la réaction produisant  $i$ . L'ensemble des métabolites du réseau est donc productible stœchiométriquement. Cet exemple montre les limites de la topologie dans l'étude de la productibilité au niveau d'un cycle, non bloqué dans le cas d'une étude stœchiométrique.

Les exemples présentés dans la Figure 2.5 montrent que l'ensemble des éléments productibles en topologie n'est pas systématiquement égal à celui calculé en étude stœchiométrique. Kruse et Ebenhöh [62] proposent alors une différenciation de notion liée à la productibilité pour les métabolites dans chacun des cas. En effet, dans le cas d'une étude de flux, un métabolite est productible dans l'hypothèse d'un état stationnaire avec possibilité d'auto-production de métabolites internes mais, dans le cas d'un état non stationnaire donc avec une croissance persistante, Kruse et Ebenhöh considèrent un métabolite comme productible s'il n'est accessible qu'à partir de nutriments et sans la consommation d'intermédiaires non accessibles, revenant ainsi à la définition du scope de Handorf et al [38]. Si Kruse et Ebenhöh continuent de définir les métabolites accessibles stœchiométriques comme productibles (*productivity*), ils définissent le scope comme l'ensemble des métabolites persistants (*sustainability*).

La comparaison d'étude de productibilité topologique et stœchiométrique montre que si la to-

pologie permet, dans la majorité des cas, d'étendre l'ensemble des métabolites productibles par rapport à une étude de flux, limitée par les hypothèse de conservation de masse et d'état stationnaire, elle ne permet pas l'activation de cycle si ce dernier n'est pas initialisé. Cependant les deux méthodes ont des moyens de correction permettant d'améliorer l'ensemble des éléments productibles. Ainsi, pour une étude topologique, la recherche de métabolites sources secondaires ou co-facteurs peut être nécessaires pour initialiser les cycles, méthode proposée notamment dans [83], tandis que la vérification de la stœchiométrie et l'ajout de réactions d'export dans le réseau permet pour une étude stœchiométrique de contourner les limites imposées par les hypothèses de la méthode. Ces exemples montrent également que les deux études sont complémentaires pour comprendre les mécanismes de production des métabolites d'un réseau métabolique.

## Conclusion

Nous avons montré dans ce chapitre qu'un réseau métabolique pouvait être étudié sous deux aspects selon la prise en compte ou non de la stœchiométrie des réactions. La première méthode, la plus utilisée, consiste à étudier le réseau comme un graphe de flux, avec deux techniques possibles : le *Flux Balance Analysis*, recherchant une solution optimale de production de biomasse, et le *Flux Variability Analysis*, calculant les limites inférieure et supérieure de chacun des flux des réactions du réseau dans le cas d'un flux maximal de croissance. Les deux analyses de flux reposent sur deux hypothèses, la conservation de masse impliquant que, pour tout nœud métabolite, les flux entrants sont égaux aux flux sortants, et l'état quasi-stationnaire, interdisant toute accumulation de métabolites dans le réseau. L'autre méthode, plus récente, étudie le réseau métabolique d'un point de vue topologique, n'utilisant pas les données de stœchiométrie des réactions. Une étude du réseau dans les deux cas a pour but d'analyser la productibilité au sein de l'organisme mais cette notion est différente selon la méthode utilisée. Ainsi, d'une part, un métabolite est productible stœchiométriquement si une réaction le produisant possède un flux positif dans la résolution du problème de flux correspondant et, d'autre part, un métabolite est productible topologiquement s'il est produit d'une réaction dont les substrats appartiennent au scope de l'ensemble des métabolites sources, concept défini notamment par Kruse et Ebenhöh [62]. Les deux approches retournent des résultats différents mais complémentaires car des voies accessibles d'un point de vue ne le sont pas de l'autre. L'analyse des mécanismes de production dans un réseau métabolique notamment dans la croissance de l'organisme met en avant des éléments particuliers intervenant à divers niveaux du réseau et associés à la production d'une biomasse. Nous nous intéressons alors aux éléments dit *essentiels* à la production de la biomasse dans le chapitre 3.



## Chapitre 3

# Des éléments particuliers dans un réseau métabolique

Un réseau métabolique modélisé sous forme de graphe biparti orienté pondéré est essentiellement utilisé afin d'étudier la productibilité au sein de l'organisme. Cette analyse des éléments productibles dans le réseau met en avant des composants particuliers dans la structure du réseau métabolique et plus précisément pour la production de la biomasse dans un graphe de flux. En effet, nous montrerons dans ce chapitre de bibliographie que, dans la littérature, plusieurs éléments sont qualifiés de "essentiels" dans la production de la biomasse, que ce soit des gènes, des réactions, des ensembles de réactions ou des métabolites. De plus, ces notions d'essentialité pour les différents composants sont uniquement liées à une étude de flux, la notion d'élément essentiel n'apparaissant pas dans la littérature au niveau topologique.

### 3.1 Des gènes essentiels

Une des premières notions d'essentialité mise en avant dans la littérature est celle de gène essentiel principalement démontré par des expériences *in vivo* via la création puis la comparaison de mutants d'espèces connues pour lesquelles le gène étudié est inactivé par différentes techniques géniques. Plusieurs études de réseau métabolique montrent que des gènes sont essentiels pour permettre à l'organisme de se développer. Deux définitions de gènes essentiels sont alors posées selon les conditions d'étude.

La première définit un gène comme essentiel si la délétion ou l'inactivation de ce dernier est létal pour l'organisme. En effet, sans la présence de ce gène, l'organisme ne peut se développer. Ainsi, Edwards et Palsson étudient, dans leurs travaux [24] l'espèce *Escherichia coli* et son réseau métabolique pour déterminer les gènes nécessaires à la croissance de l'organisme notamment en utilisant la méthode de *Flux Balance Analysis* (Définition 3). Dans le même but, Winzeler *et al*, dans [98], regardent l'effet de la délétion d'un ou plusieurs gènes dans le réseau de *S. cerevisiae*. Mais ces études de la croissance *in vivo* d'un organisme se font souvent dans des conditions particulières posées comme présentées dans [22] où les auteurs utilisent la définition des travaux précédents pour rechercher des gènes essentiels chez *S. cerevisiae* dans le cas d'une croissance en milieu aérobie.

La deuxième définition proposée dans la littérature prend en compte le milieu de croissance de l'organisme et/ou les métabolites cibles dans la réaction de biomasse, définissant ainsi une essentialité conditionnelle. Dans ce cas, une inactivation du gène étudié entraîne une non croissance de l'organisme dans des conditions environnementales précises. Ce dernier peut cependant se développer si le milieu de culture est modifié ou la biomasse redéfinie. Timmermans et Van Melderer utilisent ce concept dans [97] afin de montrer que certains gènes de *Escherichia coli* sont nécessaires à la croissance de cette espèce sous certaines conditions environnementales par analyse de mutants. D'autres travaux tels que [74] reposent sur une étude topologique du réseau métabolique ou sur une étude de flux de ce dernier comme [81].

Mais les gènes sont liés aux réactions du réseau métabolique donc une autre notion d'essentialité peut aussi se retrouver au niveau des réactions au niveau d'une étude de flux.

## 3.2 Des réactions essentielles

Le concept d'essentialité peut aussi être appliqué à des réactions qui sont dites essentielles si leur inactivation entraîne la mort de l'organisme. Un moyen de déterminer le statut d'une réaction est l'utilisation d'une étude de flux en *Flux Balance Analysis* (Définition 3) cherchant une affectation des variables de flux du réseau telle que le flux de biomasse soit positif. En effet, en forçant les limites supérieure et inférieure du flux de la réaction étudiée à zéro dans le programme linéaire modélisant le réseau métabolique, il est possible de simuler l'inactivation de cette réaction et d'analyser l'effet sur la production de la biomasse. Si, lors de la résolution du problème linéaire, le flux de la réaction de biomasse est nul suite à la modification des limites, alors la réaction est considérée comme essentielle pour la production de la biomasse définie. Cependant, une étude en FBA pour le calcul des réactions essentielles doit être faite pour chacune des réactions du réseau métabolique.

Une étude en *Flux Variability Analysis* permet d'un autre côté de définir l'ensemble des réactions essentielles avec un unique calcul en analysant les valeurs de flux résultats de la résolution du programme linéaire (Définition 4). Il est possible de caractériser les réactions selon les valeurs des limites des flux calculés, celles-ci pouvant être dans trois états : bloquées, alternatives ou essentielles. Dans le cas d'une réaction bloquée, aucun flux ne passe jamais par celle-ci. Les limites minimum et maximum sont donc toutes les deux nulles ou négligeables. Pour une réaction alternative, un flux passe par cette dernière dans certaines conditions de production optimale de la biomasse tandis que dans d'autres situations, le flux est nul. Les limites de la réaction métabolique sont donc différentes avec une limite inférieure nulle et une limite supérieure non égale à zéro. Dans la situation d'une réaction essentielle, le flux de cette dernière reste non nul quelles que soient les conditions de production. Ainsi, les valeurs des deux limites de flux, minimale et maximale, sont supérieures à zéro. Pour résumer, les valeurs des limites des flux des réactions, résultats de l'étude en FVA, permettent de caractériser l'importance d'une réaction au sein d'un réseau métabolique comme nous le définissons dans la Définition 12.

**Définition 12 (Caractérisation des réactions selon leur flux en FBA)** Soit  $G = (RUM, E, s)$ , le graphe métabolique d'un réseau métabolique. Soit, pour chaque réaction  $r \in R$ , les limites inférieure  $LB_r$  et supérieure  $UB_r$  calculées par le problème linéaire correspondant à  $G$  en FVA (Définition 4). Les réactions, inversibles, peuvent être caractérisées comme suit :

- $r$  bloquée si et seulement si  $LB_r = UB_r = 0$
- $r$  alternative si et seulement si  $LB_r = 0$  &  $UB_r > 0$
- $r$  essentielle si et seulement si  $LB_r > 0$

L'exemple présenté dans la Figure 3.1 présente les trois types de réactions dans un réseau métabolique. La réaction  $r_1$  est essentielle car si son flux est nul, la rendant ainsi inactive, le métabolite  $m_2$ , nécessaire à l'activation de la réaction  $r_4$  produisant le métabolite  $t$  cible, n'est plus accessible. Deux exemples de réactions alternatives sont les réactions  $r_2$  et  $r_3$  pouvant toutes les deux produire le métabolite  $m_3$ . Le flux de l'une peut donc être nul si celui de l'autre ne l'est pas, le métabolite  $m_3$  continuant alors d'être accessible. Enfin la réaction  $r_5$  est bloquée car elle produit le métabolite  $m_5$  non consommé par la suite que ce soit par une réaction métabolique ou par une réaction d'export. L'activation de cette dernière entraînerait donc une accumulation d'un métabolite, impossible selon les hypothèses posées, elle est ainsi bloquée quelle que soit la solution trouvée.

La notion de réaction essentielle est présente dans la littérature dans des travaux pour différentes raisons. Ainsi, Winzeler et al dans [98] mettent en avant des réactions essentielles grâce à la méthode de FBA afin de trouver les gènes essentiels présentés précédemment tandis que Duarte et al dans [22] calculent les réactions essentielles afin de valider un nouveau réseau de *Saccharomyces cerevisiae*. Mais certains auteurs comme Timmermans et al [97] ou Manimaram et [74] utilisent

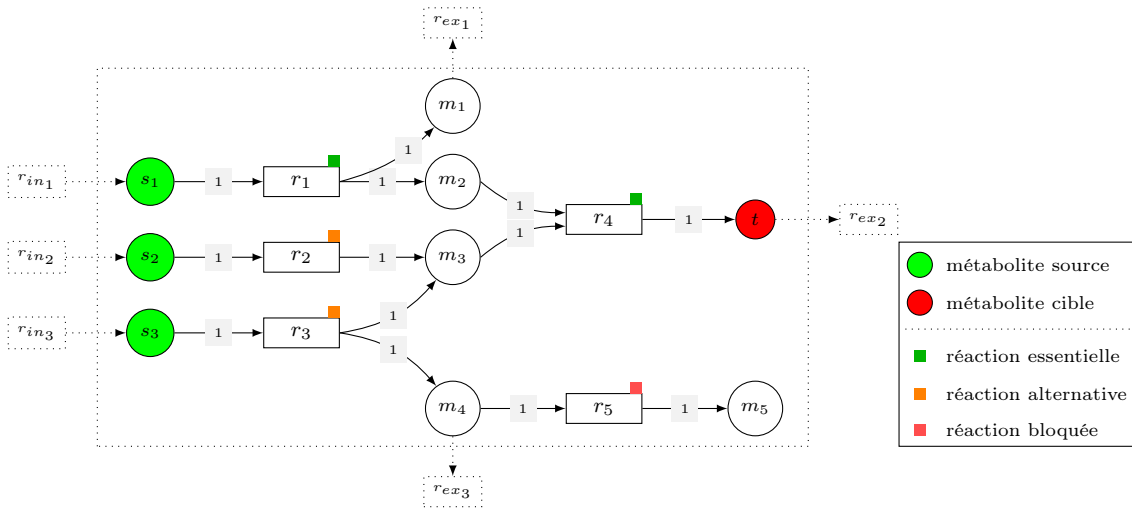


FIGURE 3.1 – Exemple de réseau métabolique présentant les trois statuts de réactions

la notion d'essentialité conditionnelle car le calcul des réactions essentielles est dépendant de la définition de la biomasse.

Si l'étude de flux permet principalement dans la littérature de définir les réactions essentielles du réseau métabolique, Samal *et al* ont montré que ces dernières étaient liées topologiquement à des nœuds métabolites dans leurs travaux [88]. Ils mettent ainsi en avant que, dans le/un graphe modélisant le réseau métabolique, les réactions dites essentielles ont pour prédécesseurs ou successeurs au moins un nœud métabolite de faible degré entrant ou sortant. Dans ce but, les auteurs calculent les nœuds métabolites caractérisés par un successeur ou un prédécesseur unique dans le graphe et montrent qu'ils interviennent dans les réactions déterminées comme essentielles par les études de flux. Cela met aussi en avant l'importance des métabolites dans la notion d'essentialité.

### 3.3 Des ensembles essentiels de réactions

Plusieurs méthodes d'analyse de flux prennent en compte des ensembles de réactions à la place de réactions isolées et mettent en avant des relations entre les réactions pour la production de la biomasse. Ainsi, si précédemment les notions d'essentialité étaient liées à une réaction et définies selon son inactivation dans le réseau, les deux méthodes présentées dans cette section mettent en avant des ensembles de réactions dont les interactions sont essentielles pour la production stœchiométrique de la biomasse

#### 3.3.1 Modes élémentaires de flux (EFM : Elementary Flux Mode)

Une méthode d'analyse des relations entre les flux d'un réseau métabolique est la technique des modes élémentaires de flux ou *Elementary Flux Modes* (EFM) décrite par Schuster et Hilgetag[91]. Dans ce cas, le but est de décomposer le réseau en unités fonctionnelles minimales pour comprendre l'interaction des réactions et plusieurs moyens sont possibles [99].

Tout comme la méthode de Flux Balance Analysis, la recherche des EFM utilise l'hypothèse d'état stationnaire pour les métabolites au sein d'un réseau métabolique dont l'ensemble des réactions métaboliques est équilibré au niveau des masses et ayant un modèle linéaire fonctionnel. Un mode de flux est représenté par un vecteur de flux  $v$  non nul respectant les contraintes d'état stationnaire ( $\mathcal{M}.v = 0$  avec  $\mathcal{M}$  matrice stœchiométrique) ainsi que les contraintes de réactions réversibles et irréversibles.

**Définition 13 (Support d'un mode de flux)** Soit  $\mathcal{M}_G$  la matrice stœchiométrique de  $G = (R \cup M, E, s)$ . Le support d'un mode de flux  $v \in \mathcal{R}^{\#R}$  est défini comme l'ensemble des indices

des flux de réactions non nuls du vecteur de flux représentant le mode :

$$\text{supp}(v) = \{i \mid v_i \neq 0\}$$

Un mode élémentaire de flux est un ensemble minimal de réactions actives opérant sous hypothèse d'état quasi-stationnaire. Ainsi, si une des réactions appartenant à un EFM est inactivée alors ce dernier n'est plus opérationnel et l'état quasi-stationnaire n'est plus assuré.

**Définition 14 (Mode élémentaire de flux (EFM))** Soit  $\mathcal{M}_G$  la matrice stœchiométrique de  $G = (R \cup M, E, s)$ . Un mode élémentaire de flux ( $\text{EFM}_G \subseteq R$ ) est défini par un vecteur de flux  $v$  tel que :

$$\begin{aligned} v &\neq 0 \\ \mathcal{M}_G \cdot v &= 0 \\ \nexists v' \text{ mode de flux} \mid \text{supp}(v') \subset \text{supp}(v) \end{aligned}$$

Les modes élémentaires de flux peuvent donc présenter des intersections mais un ensemble complet de modes élémentaires permet de modéliser l'ensemble du réseau métabolique étudié. Chaque EFM correspond à un bloc fonctionnel minimal du réseau.

Avant l'utilisation de la programmation linéaire, un moyen de trouver un mode élémentaire de flux est l'utilisation d'un algorithme glouton comme présenté dans [99]. À partir de l'ensemble des réactions métaboliques actives du système biologique, une réaction est enlevée et la présence d'un flux non nul est testée. Si le flux est nul, la réaction est remise dans l'ensemble, sinon elle est définitivement ôtée. Un mode élémentaire minimal (EFM) est obtenu lorsqu'aucune réaction métabolique ne peut être inactivée sans entraîner une inactivation de l'ensemble. Leur calcul est difficile et passe difficilement l'échelle biologique car ils doivent prendre en compte l'ensemble des ensembles de réactions biologiques présentes dans le réseau biologique. L'utilisation de la programmation linéaire a permis non seulement d'accélérer le processus, de calculer l'ensemble des modes élémentaires de flux avec une unique résolution mais surtout d'appliquer la méthode des EFM à des réseaux métaboliques de plus grande taille bien que le calcul des modes élémentaires de flux reste long et demandeur en ressources [2].

Plusieurs logiciels existent permettant de calculer les modes élémentaires de flux dans un réseau, qu'ils soient spécialisés (efmtool [93], Metatool [49]) ou intégrés dans des outils plus généralisés. Ces logiciels présentent souvent une variante de la méthode de la double description [31] générant des EFM par combinaison de deux modes élémentaires de flux vérifiés, et utilisant différents moyens pour réduire le temps de calcul ou le nombre de modes calculé. Ainsi, l'utilisation de nouvelles données comme la thermodynamique des réactions [64] ou le réseau de régulation transcriptionnel booléen correspondant [48] permettent de diminuer le nombre de réactions utiles dans le réseau et donc le nombre de EFM possibles. De plus, la décomposition en sous-réseaux du réseau initial permet de paralléliser le calcul des modes élémentaires de flux dans chaque partie mais un EFM d'un sous-réseau ne correspond pas nécessairement à un EFM du réseau total.

Grâce aux modes élémentaires de flux, il est alors possible de déterminer les ensembles minimaux de réactions actives nécessaires à la production de métabolites cibles en récupérant les EFM contenant la réaction cible mais aussi les cycles autoactivés dans le réseau [99].

### 3.3.2 Ensemble minimal de coupes (MCS : Minimal Cut Set)

Les ensembles minimaux de coupes ou *minimal cut sets* (MCS) sont une méthode issue de la théorie des graphes de flux, adaptée au réseau métabolique [59, 13]. Ce dernier est représenté sous forme d'un hypergraphe uniparti dont les nœuds sont les métabolites et les arcs les réactions. Les réactions d'import et d'export sont ajoutées de la même façon que pour la FBA pour rendre le réseau fonctionnel pour produire une biomasse, selon les spécifications introduites dans la Définition 1. Un ensemble de coupe ou *cut set* (CS) est un ensemble d'arc-réactions dont l'inactivation complète entraîne celle de la réaction objectif.

**Définition 15 (Ensemble de coupe (CS : Cut Set))** Soit  $G = (R \cup M, E, s)$  le graphe d'un réseau métabolique. Soit  $CS_G$  un sous-ensemble de réactions tel que  $CS_G \subseteq R$ . Soit  $v_b$  le flux

maximal de biomasse possible dans  $G$  et  $v'_b$  celui dans  $G$  avec la contrainte suivante ajoutée,  $v_r = 0, \forall r \in CS_G$  (inactivation des réactions).  $CS_G$  est un ensemble de coupe si et seulement si  $v_b > 0$  et  $v'_b = 0$

À partir de la [Définition 15](#), nous pouvons définir un ensemble minimal de coupes comme un ensemble minimal de réactions dont l'inactivation entraîne une inactivation de la réaction de biomasse. Un CS est considéré comme minimal s'il n'existe pas de CS inclus dans ce dernier.

**Définition 16 (Ensemble de coupe minimal (MCS : Minimal Cut Set))** Soit  $G = (R \cup M, E, s)$  le graphe d'un réseau métabolique. Un ensemble de coupe  $CS_G \subset R$  est un ensemble de coupe minimal de  $G$  si et seulement si il n'existe pas d'ensemble de coupe  $CS'_G \subset R$  tel que  $CS'_G \subset CS_G$ .

L'étude de l'ensemble des minimal cut sets permet de mettre en avant des structures dans le réseau liées à la robustesse et la fragilité de ce dernier [20]. Ainsi un MCS avec peu de réactions montre une fragilité du système car l'inactivité de ce petit nombre de réactions entraîne de graves conséquences dans le réseau, tandis qu'à l'inverse, un MCS avec un grand nombre de réactions est le résultat d'une grande robustesse du réseau métabolique. De plus, un grand nombre de MCS dans un graphe représentant un système biologique est le signe d'une fragilité dans les voies métaboliques car il y a plus de possibilités d'inactivation de la réaction de biomasse. Dans le cas contraire, un petit nombre de MCS entraîne une structure robuste car seule une inactivation précise de réactions est la cause de celle de la réaction objectif.

Les deux méthodes présentées dans cette section permettent le calcul d'un part des ensembles minimaux de coupes et d'autre part des modes élémentaires de flux, regroupant des réactions afin de trouver des ensembles minimaux essentiels à la production de la biomasse définie. Ces éléments sont également liés à la robustesse du réseau métabolique car le nombre d'ensembles de chacune des méthodes montre la capacité de l'organisme à s'adapter à des modifications du réseau métabolique. De plus, Ballerstein et al montrent dans leur travaux [6] que les deux méthodes sont des problèmes duals l'un de l'autre signifiant que la résolution de l'un permet d'obtenir la solution de l'autre.

### 3.4 Des métabolites essentiels

Une dernière notion d'essentialité peut être définie au niveau des nœuds métabolites dans le réseau métabolique. En effet, en agissant sur les réactions utilisant le métabolite étudié, il est possible d'analyser les conséquences d'une non-accessibilité du métabolite pour la production d'une biomasse. Le concept de métabolites essentiels apparaît notamment dans les études de flux faites dans différents travaux mais avec des notions d'essentialité différentes.

Une première définition des métabolites essentiels est proche de la définition de réaction essentielle. Un métabolite est dit essentiel si le fait d'empêcher sa production est létal pour l'organisme étudié [54]. Au niveau des flux, ce cas peut être modélisé en imposant les flux des réactions produisant le métabolite étudié à zéro puis en calculant la valeur du flux de la biomasse. Si ce dernier est nul alors le métabolite est essentiel. Cette notion est donc liée aux réactions essentielles du réseau métabolique. Dans l'exemple de la [Figure 3.1](#), nous pouvons définir comme essentiels les métabolites  $m_2$  et  $m_3$  car ils sont les substrats de la réaction produisant le métabolite cible.

Une autre définition d'essentialité pour les métabolites est proposée par Kim *et al* dans leurs travaux [55]. Ils définissent ainsi un métabolite essentiel comme un élément du réseau métabolique permettant de garder un flux positif pour la réaction de biomasse malgré de fortes perturbations dans le réseau même. L'étude est comme précédemment au niveau des flux en créant pour chaque nœud métabolite du graphe deux programmes linéaires modélisant les flux dans deux sous-graphes. D'une part, le premier de ces sous-graphes est constitué de l'ensemble des éléments du réseau métabolique auquel sont enlevées ou inactivées toutes les réactions consommant le métabolite étudié à l'exception d'une d'entre elles. D'autre part, le deuxième problème étudie l'ensemble des réactions du réseau métabolique auquel est enlevé ou inactivé l'ensemble complet des réactions utilisant le métabolite voulu comme substrat. Pour chacune des deux études de flux, la valeur du



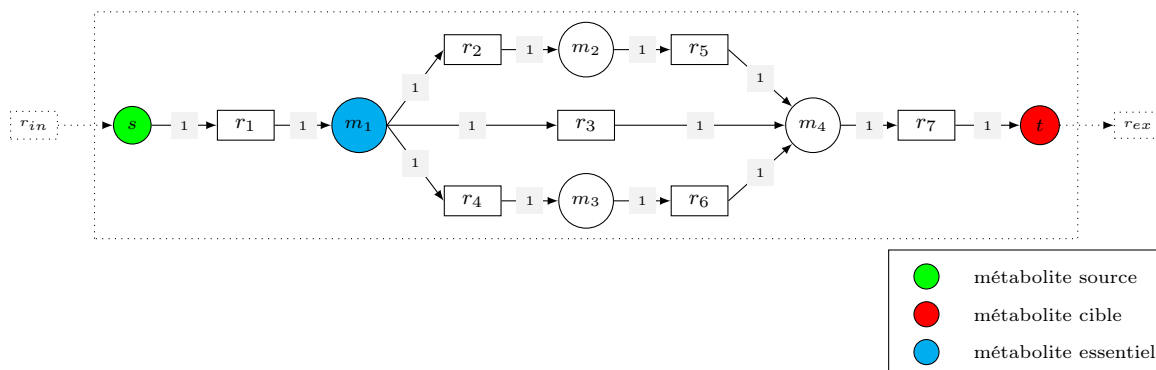


FIGURE 3.2 – Exemple de métabolite essentiel permettant au réseau métabolique de résister à de fortes perturbations

flux de la réaction de biomasse est relevée et si cette dernière est positive pour le problème avec une unique réaction consommatrice et nulle pour le problème avec aucune réaction utilisatrice du métabolite alors celui-ci est considéré comme essentiel. Ainsi, les auteurs de [55] modélisent une forte perturbation dans le réseau métabolique et mettent en avant les métabolites permettant à l'organisme de survivre malgré ces modifications.

Dans l'exemple présenté à la Figure 3.2, nous pouvons voir trois voies métaboliques commençant par le métabolite  $m_1$  et finissant par  $m_4$ . Les trois chemins sont équivalents ou alternatifs comme définis dans la section précédente. Le fait donc d'inactiver deux des trois voies n'entraîne pas de changement dans le flux de production du métabolite  $t$  cible alors que si les trois voies de production sont non accessibles, l'organisme n'est plus capable de produire l'élément cible. Dans ce cas, selon la définition précédente, le métabolite  $m_1$  est défini comme essentiel.

## Conclusion

Des études de flux au niveau de réseaux métaboliques permettent d'identifier des éléments essentiels à la production de métabolites cibles définis dans une réaction de biomasse. Ces derniers peuvent être des gènes, des réactions, des ensembles de réactions ou encore des métabolites. Cependant ces analyses, uniquement stœchiométriques, mettent en avant des notions d'essentialité différentes bien que proches dépendant du point de vue utilisé par les auteurs. Il est donc délicat d'obtenir un concept généraliste d'essentialité.

De plus, de manière intéressante, aucune notion d'essentialité n'est présentée dans la littérature en utilisant les concepts topologiques d'accessibilité et de scope. Or, comme nous l'avons montré, les études de productibilité stœchiométrique et topologique se complètent, mettant chacune en lumière des éléments différents. Il est donc intéressant d'étudier la notion d'essentialité d'un point de stœchiométrie mais aussi au niveau topologique. Ce sera l'objet du chapitre suivant.

## Chapitre 4

# Raffiner et étendre le concept de métabolite essentiel

Dans les chapitres bibliographiques, nous avons mis en évidence des éléments clés d'un réseau métabolique tels que des métabolites essentiels ou ensembles de réactions structurales comme les ensembles de coupure ou les modes élémentaires. Cependant, il faut noter que ces ensembles sont principalement basés sur un formalisme de flux pour décrire la capacité d'une cellule à produire des éléments cibles. Or, nous avons aussi vu précédemment que les approches par flux sont limitées pour décrire la capacité d'une cellule à croître sur le long-terme, qui est mieux modélisée par des études d'accessibilité dans un graphe. Dans ce contexte, notre première contribution consiste à unifier et étendre le concept de métabolite essentiel à trois sémantiques de productibilité de métabolites. Les différences entre ces sémantiques seront exploitées dans le chapitre suivant pour identifier des métabolites clés du système.

### 4.1 Métabolites essentiels du point de vue de la productibilité

Comme nous l'avons vu dans le chapitre 3, la notion d'essentialité est principalement modélisée à l'aide de flux de matières dans l'organisme modélisé, en s'appuyant sur des analyses de flux, notamment le *Flux Balance Analysis* introduit au chapitre 2. Une première contribution a consisté à formaliser une définition de métabolites essentiels (ME) à la production de métabolites cibles d'un point de vue de la production de biomasse à l'état stationnaire. L'inactivation de toutes les voies qui consomment un tel métabolite entraîne une impossibilité de l'organisme à produire de la biomasse à partir de nutriments du milieu défini. Nous nommons alors un tel métabolite, un métabolite essentiel du point de vue de la productibilité du réseau (ME-productibilité).

#### 4.1.1 Élagage d'un réseau par rapport à un métabolite

Un préliminaire pour introduire la définition d'un métabolite essentiel du point de vue de la productibilité du réseau est d'introduire une opération d'élagage pour modéliser le réseau métabolique privé de l'ensemble des réactions utilisant le composé étudié en tant que substrat. Ainsi le métabolite peut être produit mais non consommé. En modélisation sous forme de graphe, ce sous-réseau lié à un métabolite correspond au sous-graphe du graphe métabolique (Définition 7) dont l'ensemble des successeurs du nœud métabolite étudié est enlevé ou, selon les définitions données précédemment, auquel les nœuds réactions dont le nœud métabolite étudié est un substrat sont enlevés. Nous devons aussi prendre en compte les réactions réversibles dont le sens peut s'inverser et regroupées dans l'ensemble *Rev*. Les réactions enlevées du réseau pour créer le sous-réseau du métabolite étudié sont donc les réactions ayant le métabolite comme réactants et celles réversibles ayant ce dernier comme produits.

**Définition 17 (Sous-réseau élagué par rapport à un métabolite)** *Soit un réseau métabolique  $G = (R \cup M, E, s)$ . Le sous-réseau élagué par rapport un métabolite  $m \in M$  est noté*

$prune(G, m) = (R' \cup M, E', s)$  et est défini comme suit.

$$\begin{cases} R' = R \setminus (\{r \in R \mid m \in rcts(r)\} \cup \{r \in Rev(G) \mid m \in prds(r)\}) \\ E' = (M \times R' \cup R' \times M) \cap E. \end{cases}$$

#### 4.1.2 Définition d'un métabolite essentiel du point de vue de la productibilité du réseau

Nous nous reposons sur le calcul du sous-réseau élagué pour définir si un métabolite est un métabolite essentiel du point de vue de la productibilité du réseau. En effet, un tel nœud métabolite ne peut être ni source ni cible et est nécessaire à la production de la biomasse dans le réseau métabolique. Dans le cas de l'étude des flux, cela signifie que, si le métabolite étudié est retiré du réseau alors le flux de la réaction de biomasse devient égal à zéro. Nous rappelons que, dans le graphe  $G = (R \cup M, E, s)$ , comme nous l'avons expliqué dans le chapitre 2, les flux sont contraints tels que  $LB_r \leq v_r \leq UB_r, \forall r \in R$  et que l'hypothèse d'état stationnaire est vérifiée signifiant qu'aucun métabolite ne peut être accumulé dans le réseau :  $\sum_{(r,m) \in E} s(r, m) \cdot v_r - \sum_{(m,r) \in E} s(m, r) \cdot v_r = 0, \forall m \in M$ .

**Définition 18 (Métabolite ME-productibilité ou carrefour stoechiométrique)** Soit un réseau métabolique  $G = (R \cup M, E, s)$ , soit un ensemble de métabolites sources  $S \subset M$  et soit une réaction cible  $r_b \in R$  dont le flux est  $v_b$  telle que  $T = rcts(r_b)$ . On note  $prune(G, m)$  le graphe obtenu par l'opération d'élagage de  $G$  vis-à-vis de  $m$ .

Un métabolite  $m \in M$  est appelé un métabolite essentiel du point de vue de la productibilité du réseau ou métabolite ME-productibilité si :

$$\begin{cases} m \notin S \cup T \\ v_b > 0, \text{ dans } G = (R \cup M, E, s) \\ v'_b = 0 \text{ dans } prune(G, m). \end{cases}$$

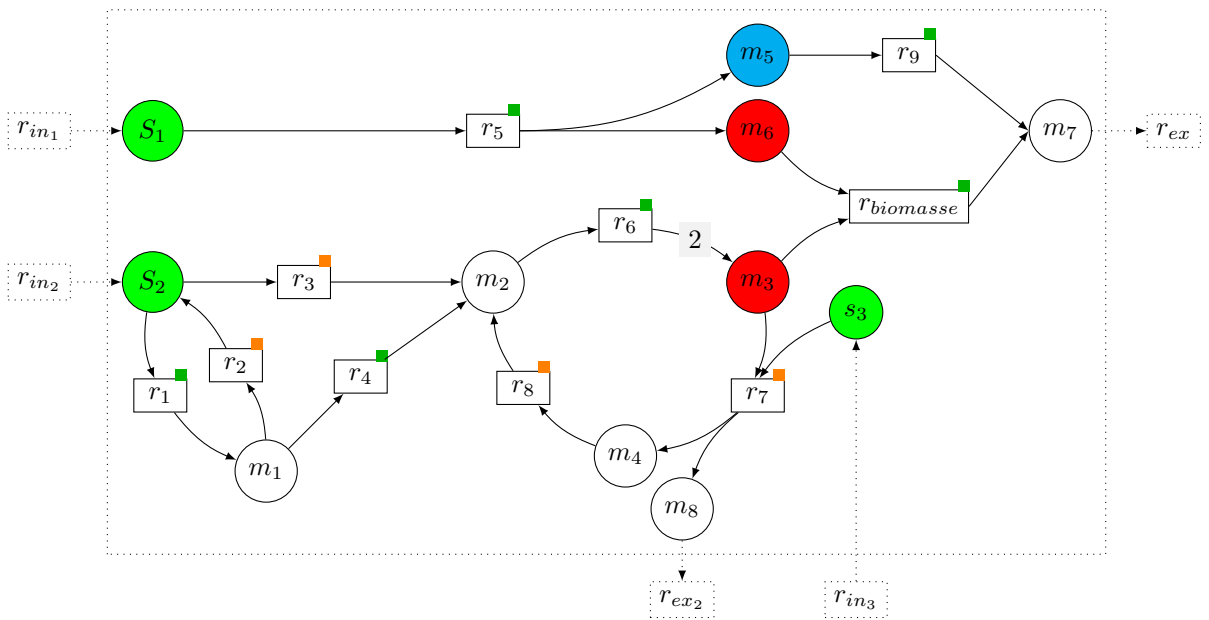
D'un point de vue phénotypique, ces métabolites sont donc essentiels vis-à-vis de la production de biomasse à l'état quasi-stationnaire. D'un point de vue méthodologique, leur définition s'appuyant sur la matrice stoechiométrique du réseau, nous les appellerons *carrefours stoechiométriques*.

#### 4.1.3 Exemple de ME-productibilité

Plusieurs situations dans un réseau métabolique peuvent répondre à la définition de métabolite essentiel du point de vue de la productibilité du réseau. Un métabolite ME-productibilité peut se trouver dans la voie unique de production d'un métabolite substrat de la réaction cible, voie accessible stoechiométriquement, ou au commencement de plusieurs voies métabolique alternatives de production d'un réactant de la réaction cible. Dans un second temps, un métabolite ME-productibilité peut être mis en avant au niveau d'une réaction produisant un métabolite non directement nécessaire pour la production des métabolites cibles mais dont l'équilibre des masses doit être assuré. Un exemple de ce cas particulier est présenté dans la Figure 4.1.

Dans cet exemple, nous nous intéressons au métabolite  $m_5$ , produit secondaire d'une réaction créant un métabolite cible ( $m_6$ ). Nous créons le sous-réseau élagué par rapport à  $m_5$  dans la sous-figure (b) en enlevant la réaction  $r_9$  qui utilise le métabolite comme substrat. Dans le réseau élagué par rapport à ce métabolite secondaire, nous voyons que le fait d'enlever l'ensemble des réactions consommant ce dernier bloque la réaction le produisant car, si celle-ci n'était pas inactivée alors  $m_5$  serait accumulé dans le réseau métabolique, violant la loi d'état stationnaire posée lors d'une étude de *Flux Balance Analysis* (Définition 3). La réaction  $r_5$  étant donc bloquée dans le réseau élagué, la réaction de biomasse, n'ayant pas l'ensemble de ses substrats produits, est inactivée : aucun flux ne peut passer par elle. Le métabolite  $m_5$  répond donc à la définition de métabolite essentiel du point de vue de la productibilité du réseau car, dans son sous-réseau, la réaction de biomasse possède un flux nul.

Nous pouvons faire le lien en observant l'exemple présenté dans la Figure 4.1 avec les définitions de métabolite essentiel posées dans [55],[54] et des métabolites liés à la productibilité dans les travaux de Kruse and Ebenhöh [62] que nous avons présentés dans le chapitre 3 car le fait d'enlever



(a) Exemple d'un ME-productibilité dans un réseau métabolique

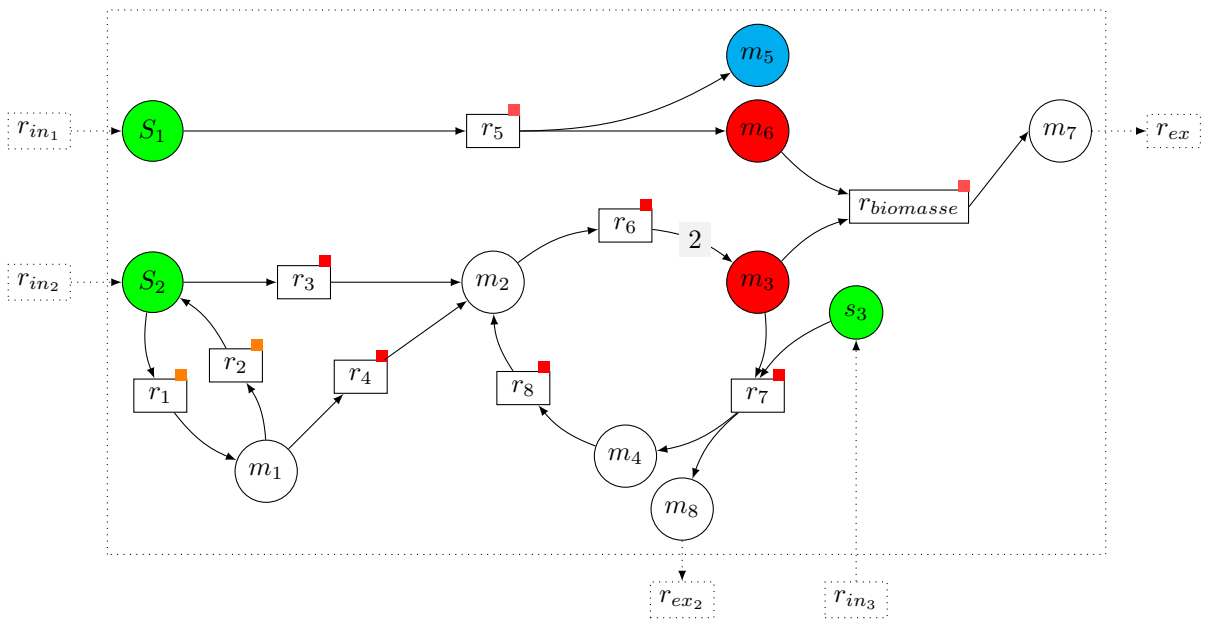
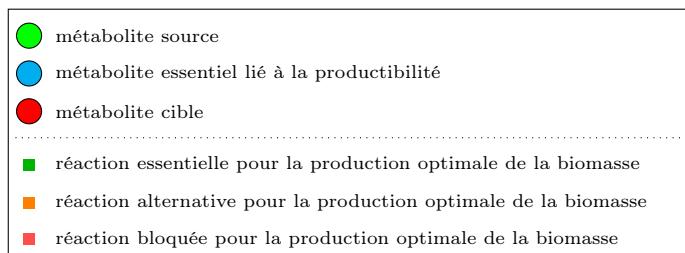
(b) Sous-réseau élagué par rapport à  $m_5$ 

FIGURE 4.1 – Dans l'exemple présenté dans cette figure, nous considérons la voie composée des réactions  $r_1$  et  $r_4$  comme plus efficace que la voie composée de la réaction  $r_3$ . (a) Le métabolite  $m_5$  est un métabolite essentiel du point de vue de la productibilité du réseau (métabolite ME-productibilité). (b) L'élagage du réseau vis-à-vis de  $m_5$  entraîne l'inactivation de  $m_6$  et l'impossibilité de produire de la biomasse.

un métabolite ME-productibilité du réseau métabolique entraîne une absence de flux dans la réaction de biomasse définie. De plus, nous pouvons faire le lien avec la notion d'ensemble minimal de coupes dans un cas particulier. En effet, comme décrit dans le chapitre 3, un ensemble minimal de coupes (MCS) est un ensemble de réactions dont l'inactivation entraîne une absence de flux dans la réaction de biomasse (Définition 16). Or si la totalité des réactions d'un MCS ont un substrat en commun, le fait de bloquer ce métabolite entraîne l'inactivation de l'ensemble de ces réactions et donc l'absence de flux de croissance. Le métabolite en commun répond de ce fait à la définition d'un métabolite essentiel du point de vue de la productibilité du réseau (Définition 18).

#### 4.1.4 Calcul des métabolites ME-productibilité

Le réseau exemple proposé est de petite taille et l'objectif est de pouvoir rechercher les métabolites répondant à la définition des métabolites ME-productibilité dans un réseau réel de taille génomique. Afin de calculer les métabolites ME-productibilité dans un réseau métabolique, nous avons utilisé un package Python créé afin de pouvoir faire une étude de *Flux Balance Analysis* dans un réseau métabolique appelé *cobrapy* [23]. Ce package développé par Ebrahim *et al* prend en entrée un fichier sbml correspondant au réseau que nous voulons étudier et calcule le flux de la réaction de biomasse défini dans le fichier ainsi que les flux de chacune des réactions du réseau métabolique. Pour chacun des métabolites du réseau, nous créons donc le fichier modélisant le réseau élagué lié à ce dernier en enlevant les réactions où il apparaît en tant que substrat et nous utilisons le package *cobrapy* pour calculer la valeur du flux de biomasse. Si ce dernier est nul, négligeable ou négatif, nous considérons que le métabolite est ME-productibilité.

## 4.2 Métabolites essentiels du point de vue de l'efficacité

Le rôle d'un métabolite dans un réseau peut aussi être lié à la production optimale de la biomasse, étudiée par des analyses de flux comme nous l'avons vu précédemment. Nous nous intéressons donc en dernier lieu aux métabolites nécessaires à la production maximale des éléments ciblés, regroupés dans la biomasse. Nous appelons ces composés des métabolites essentiels du point de vue de l'efficacité du réseau.

### 4.2.1 Définition d'un métabolite essentiel du point de vue de l'efficacité du réseau

Un métabolite essentiel du point de vue de l'efficacité du réseau (ou ME-efficacité) est essentiel à la production optimale de la biomasse définie dans le réseau métabolique calculée dans une étude de flux dans le graphe associé à ce réseau. Il est donc nécessaire pour produire une quantité maximale de métabolites ciblés mais, l'absence du métabolite ME-efficacité n'empêche pas une production de biomasse réduite. Nous nous reposons sur une étude en *Flux Variability Analysis* (Définition 4) afin de déterminer les réactions essentielles (Définition 12) à la production maximale de la biomasse correspondant au flux optimal pouvant être dans la réaction de biomasse définie. Nous définissons en tant que métabolite ME-efficacité les substrats de ces réactions essentielles calculées. En effet, le fait d'enlever ces métabolites du réseau entraîne une absence de flux dans une réaction essentielle pour obtenir un flux maximal dans la réaction de biomasse. Comme vu précédemment dans le chapitre 2, les flux sont contraints :

$$LB_r \leq v_r \leq UB_r \quad \forall r \in R. \quad (4.1)$$

et l'hypothèse d'état stationnaire dans le réseau métabolique  $G = (R \cup M, E, s)$  qui implique que tout flux entrant est égal au flux sortant pour tous les nœuds du réseau et donc aucun métabolite n'est accumulé dans le réseau, est vérifiée :

$$\sum_{(r,m) \in E} s(r,m) \cdot v_r - \sum_{(m,r) \in E} s(m,r) \cdot v_r = 0 \quad \forall m \in M. \quad (4.2)$$

Ces deux hypothèses sont prises en compte lors de la définition du problème linéaire en *Flux Variability Analysis* et nous les prenons en compte pour définir les métabolites essentiels du point de vue de l'efficacité du réseau.

**Définition 19 (Métabolite ME-efficacité ou carrefour optimal)** Soit un réseau métabolique  $G = (R \cup M, E, s)$ , soit un ensemble de métabolites sources  $S \subset M$  et soit une réaction cible  $r_b \in R$  dont le flux est  $v_b$  telle que  $T = rcts(r_b)$ . Un métabolite  $m \in M$  est appelé métabolite essentiel du point de vue de l'efficacité du réseau ou métabolite ME-efficacité si :  $\exists r_0 \in R, m \in rcts(r_0)$  tel que

$$\begin{aligned} & \max\{v_b \mid \exists\{v_r, r \in R\} \subset \mathbb{R}^+ \text{ tel que Eqs. (4.1), (4.2) vérifiées}\} \\ & > \max\{v_b \mid \exists\{v_r, r \in R\} \subset \mathbb{R}^+ \text{ tel que } v_{r_0} = 0 \text{ et Eqs. (4.1), (4.2) vérifiées}\}. \end{aligned}$$

D'un point de vue phénotypique, ces métabolites sont donc essentiels vis-à-vis de la production optimale de biomasse à l'état quasi-stationnaire. D'un point de vue méthodologique, leur définition s'appuyant sur des critères d'optimisation de systèmes linéaires, nous les appellerons *carrefours optimaux*.

### 4.2.2 Exemple de métabolite ME-efficacité

La recherche des métabolites ME-efficacité est liée à la recherche des réactions essentielles pour la maximisation du flux de la réaction de biomasse. L'ensemble des métabolites ME-efficacité est donc l'ensemble des substrats de ces dernières dans le réseau métabolique. Un exemple de métabolite essentiel du point de vue de l'efficacité du réseau peut être observé au niveau d'une voie métabolique pouvant être alternative mais présentant un avantage dans la quantité d'éléments produits, entraînant une augmentation de la quantité de biomasse produite. Ainsi la présence d'autres voies alternatives permet de garder une production de biomasse bien que moindre.

Dans l'exemple présenté dans la Figure 4.2, deux métabolites ont les propriétés d'un métabolite essentiel du point de vue de l'efficacité du réseau :  $m_1$  et  $m_5$ . Le métabolite  $m_5$  est ME-efficacité car la réaction  $r_9$  est essentielle pour la production de la biomasse pour les mêmes raisons que précédemment évoquées pour les métabolites ME-productibilité : le métabolite  $m_5$  devant être consommé afin d'éviter une accumulation en violation avec l'hypothèse d'état stationnaire.

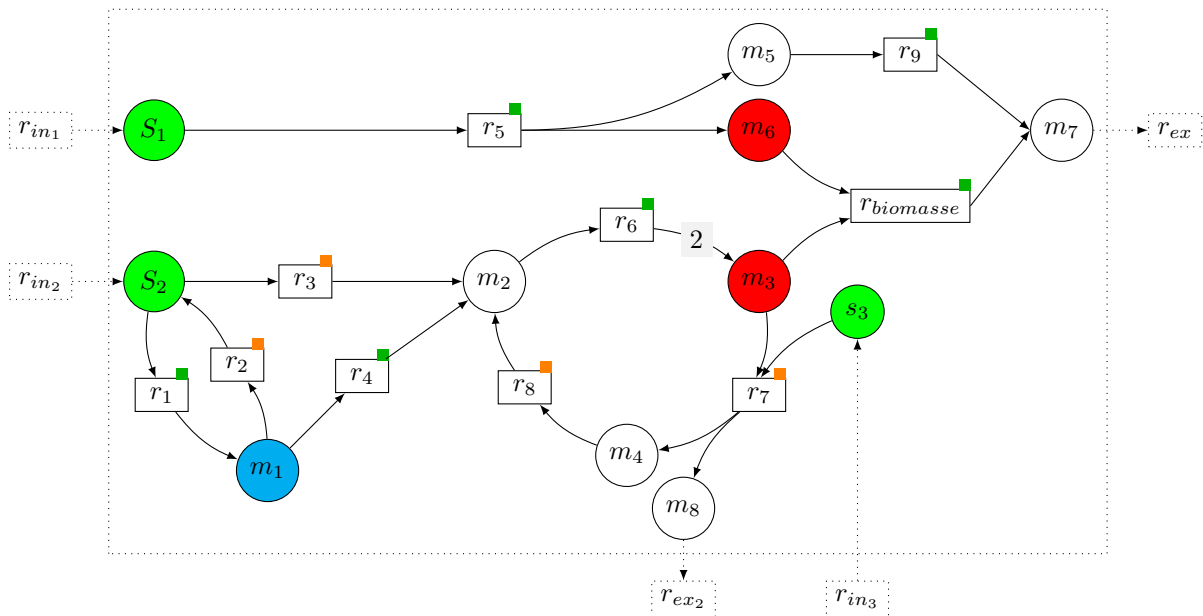
De manière plus subtile, le métabolite  $m_1$  se positionne dans la situation expliquée précédemment d'impact sur l'efficacité du système. En effet, pour produire une biomasse optimale, le flux de matière utilise les réactions  $r_1$  et  $r_4$  afin de produire le métabolite  $m_2$  mais la suppression de la réaction  $r_2$  empêchant ainsi la production de  $m_1$ , entraîne l'utilisation de la voie secondaire alternative composée de la réaction  $r_3$  et ne bloque pas la production d'une biomasse dans le graphe élagué par rapport à  $m_1$ .

### 4.2.3 Calcul des métabolites ME-efficacité

Les métabolites essentiels du point de vue de l'efficacité du réseau dans un réseau métabolique sont liés aux réactions essentielles à la production optimale de la biomasse définie. Pour calculer ces derniers, nous devons donc rechercher ces réactions essentielles. Comme pour les métabolites essentiels du point de vue de la productibilité du réseau, nous utilisons le package *cobrapy* afin de réaliser une étude en *Flux Variability Analysis* (Définition 4) et ainsi obtenir les limites inférieure et supérieure de chacun des flux des réactions et d'en déterminer le statut : essentielle, alternative ou bloquée, tel que nous l'avons spécifié dans la Définition 12. Nous récupérons alors l'ensemble des substrats de chaque réaction désignée comme essentielle, créant ainsi l'ensemble des métabolites ME-efficacité.

## 4.3 Métabolites essentiels du point de vue de la persistance

Comme vu dans le chapitre 2, les approches par flux ont des limites pour modéliser la croissance d'un organisme. Puisqu'elles s'appuient sur une hypothèse d'état quasi-stationnaire, elles autorisent la production interne et auto-équilibrée de co-facteurs sans prendre en compte le fait que la cellule peut croître et avoir besoin d'apports extérieurs de nutriments [38, 62]. Ainsi, dans le chapitre 2, nous nous sommes plus précisément penchés sur les chemins existant entre les nœuds



(a) Exemple d'un métabolite essentiel du point de vue de l'efficacité du réseau dans un réseau métabolique

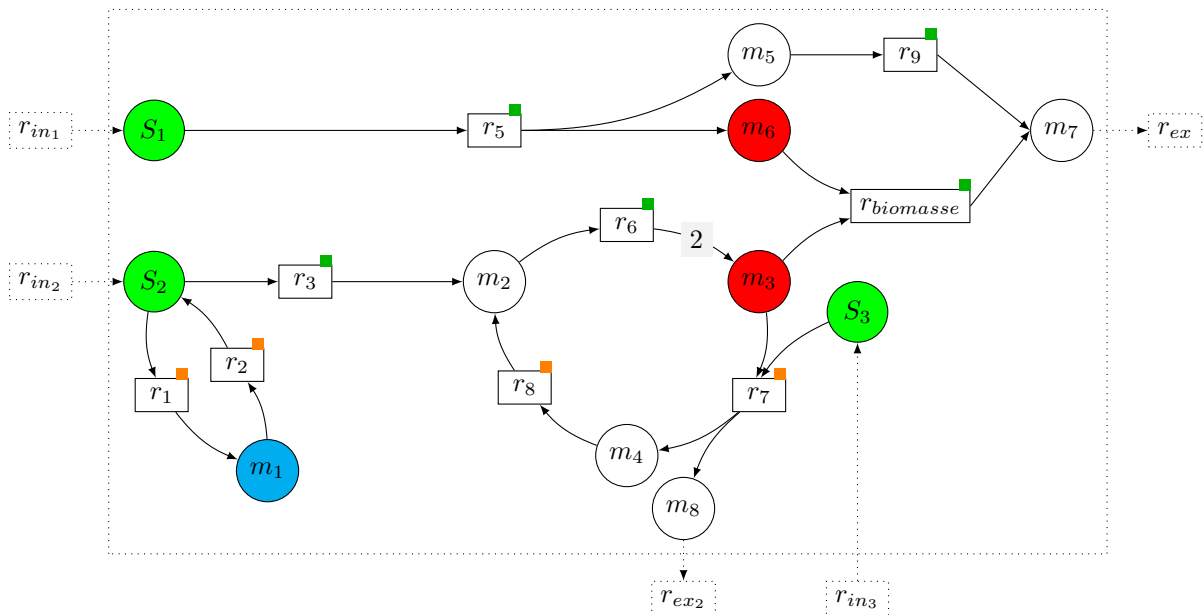
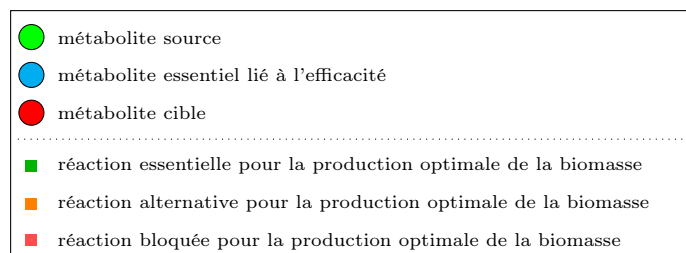
(b) Sous-réseau élagué par rapport au métabolite  $m_1$ 

FIGURE 4.2 – (a) Le métabolite  $m_1$  est un métabolite essentiel du point de vue de l'efficacité du réseau (ME-efficacité). (b) L'élagage d'une réaction,  $r_4$ , vis-à-vis de  $m_1$  dans le réseau entraîne l'inactivation du flux optimal qui permet de produire de la biomasse. Cependant, la production de biomasse globale n'est pas nécessairement bloquée.

métabolites sources du réseau et les nœuds métabolites cibles que nous voulons atteindre, ces derniers devant être dans le *scope*  $\Sigma(G, S)$  de l'ensemble source (Définition 9). Nous allons exploiter cette sémantique pour introduire un concept d'essentialité vis-à-vis de la persistance d'un réseau.

### 4.3.1 Définition d'un métabolite essentiel du point de vue de la persistance du réseau

La définition d'un métabolite essentiel du point de vue de la persistance du réseau repose sur des parcours de graphe appliqués sur un sous-réseau auquel le métabolite étudié est enlevé. Nous utilisons donc de nouveau, comme pour les ME-productibilité, la définition d'un réseau élagué par rapport à un métabolite (Définition 17). Comme précédemment, nous nous reposons sur le calcul du réseau élagué pour définir si un composé est un ME-persistance. Ces derniers ne peuvent être ni source ni cible et sont nécessaires à l'accessibilité topologique d'au moins un nœuds cible, signifiant que, dans le sous-réseau élagué par rapport au métabolite, au moins un nœud métabolite cible n'appartient pas au *scope* des nœuds métabolites sources (Définition 9).

**Définition 20 (métabolite ME-persistance ou carrefour topologique)** *Soit un réseau métabolique  $G = (R \cup M, E)$ , soit un ensemble de métabolites sources  $S$  et soit un ensemble de métabolites cibles  $T$ . On note  $\text{prune}(G, m)$  le graphe obtenu par l'opération d'élagage de  $G$  vis-à-vis de  $m$ . Un métabolite  $m \in M$  est appelé un métabolite essentiel du point de vue de la persistance du réseau (ou métabolite ME-persistance) si*

$$\begin{cases} m \notin S \cup T \\ \forall t \in T, t \in \Sigma(G, S) \\ \exists t \in T \mid t \notin \Sigma(\text{prune}(G, m), S). \end{cases}$$

D'un point de vue phénotypique, ces métabolites sont donc essentiels vis-à-vis de la persistance de la croissance du système. D'un point de vue méthodologique, leur définition s'appuyant sur des critères de parcours de graphes, nous les appellerons *carrefours topologiques*.

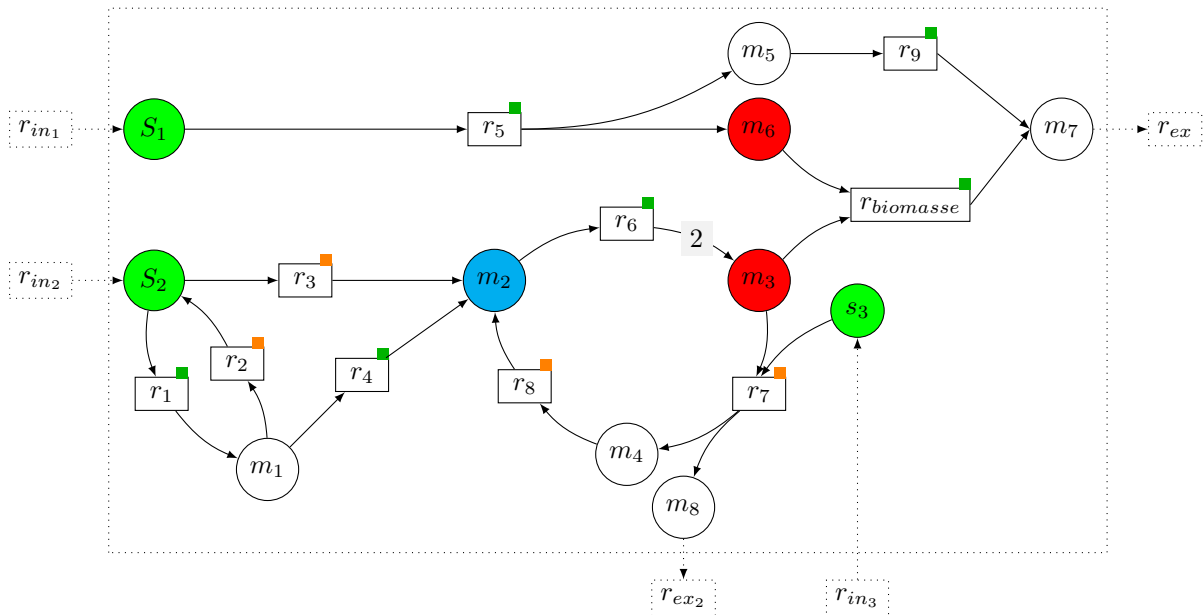
### 4.3.2 Exemple de métabolite ME-persistance

La Définition 20 permet d'introduire un concept de ce que serait un nœud métabolite essentiel à l'accessibilité topologique (Définition 11) des nœuds ciblés c'est-à-dire nécessaire à la production des métabolites modélisés par ces nœuds. En nous basant sur les définitions de la notion d'essentialité vues précédemment dans le chapitre 2, nous avons défini ces métabolites éventuels comme les nœuds métabolites entraînant une impossibilité d'accessibilité des nœuds cibles s'ils sont enlevés du graphe représentant le réseau métabolique et les avons nommés métabolites essentiels du point de vue de la persistance du réseau (métabolite ME-persistance) car, dans ce cas, seule la topologie du réseau est prise en compte sans intervention de la stœchiométrie. D'un point de vue graphe, le fait d'enlever un des nœuds ME-persistance entraîne une absence de chemin entre les nœuds sources et au moins un nœud cible.

Nous pouvons donner en exemple d'un métabolite ME-persistance un nœud se trouvant dans le chemin unique d'accessibilité d'un métabolite cible correspondant à une voie unique de production du métabolite modélisé par ce dernier, ou un nœud se trouvant à l'origine de plusieurs chemins permettant d'accéder à un même nœud cible, modélisant plusieurs voies de production qui sont topologiquement équivalentes. Un exemple est présenté dans la Figure 4.3. Dans un premier temps, les cas sont semblables à ceux présentés pour les métabolites ME-productibilité.

Dans cet exemple, nous nous intéressons au métabolite  $m_2$ . Nous créons le sous-réseau élagué par rapport à  $m_2$  dans la sous-figure (b) en enlevant la réaction  $r_6$  qui utilise le métabolite comme substrat. Le nœud métabolite  $m_3$  n'est alors plus accessible topologiquement dans ce sous-réseau expliquant le statut de métabolite essentiel du point de vue de la persistance du réseau du métabolite  $m_2$ . On peut noter que  $m_2$  est aussi un métabolite ME-productibilité et ME-efficacité. Par contre,  $m_5$ , qui est un métabolite ME-productibilité, n'est pas ME-persistance. Dans le chapitre suivant, nous verrons cependant qu'il existe des cas où des éléments sont des métabolites ME-persistance sans être ni ME-productibilité ni ME-efficacité.





(a) Exemple d'un métabolite essentiel du point de vue de la persistance du réseau dans un réseau métabolique

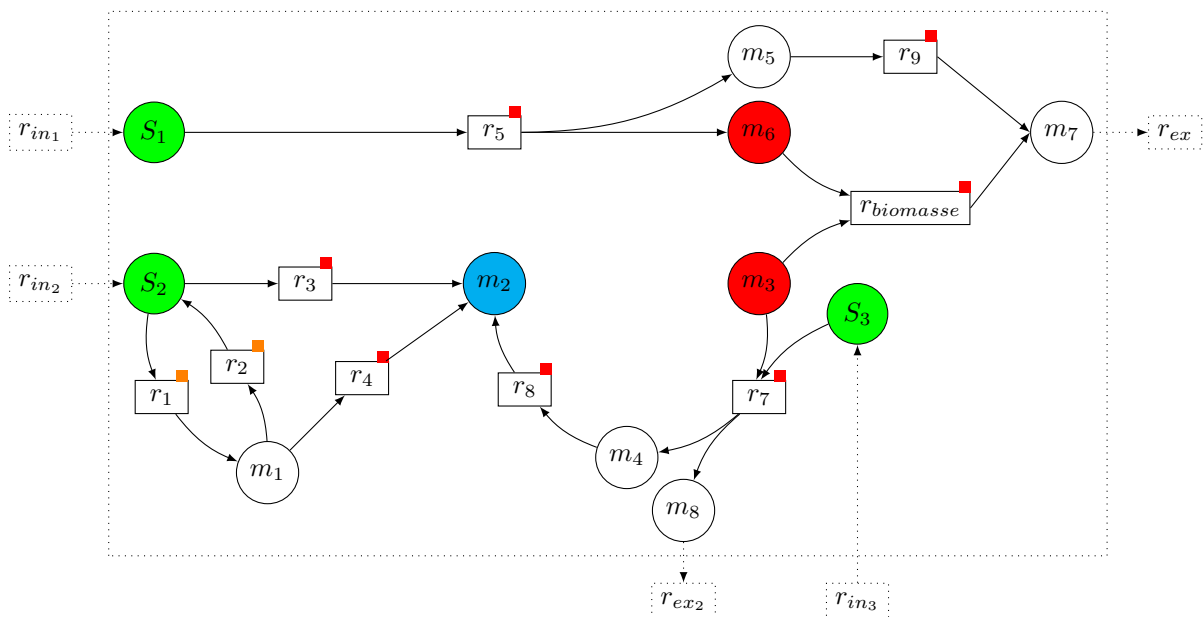
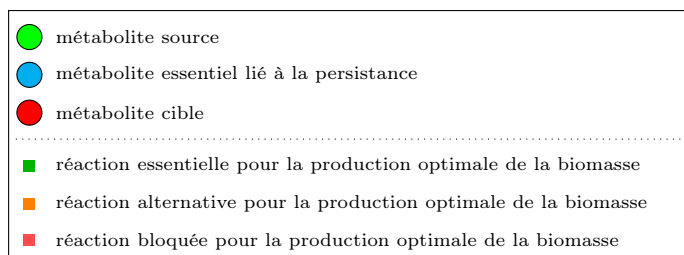
(b) Réseau élagué par rapport au métabolite  $m_2$ 

FIGURE 4.3 – (a) Le métabolite  $m_2$  est un métabolite essentiel du point de vue de la persistance du réseau (ME-persistance). (b) L'élagage du réseau vis-à-vis de  $m_2$  entraîne l'inactivation de  $m_3$  et l'impossibilité de produire de la biomasse d'un point de vue topologique.

### 4.3.3 Calcul de métabolite ME-persistance

La propriété des métabolites essentiels du point de vue de la persistance du réseau étant combinatoire, la principale difficulté est l'identification des métabolites répondant à la propriété d'accessibilité au sein d'un graphe métabolique à partir d'un ensemble de métabolites sources ([Définition 11](#)) car cette dernière est définie de façon récursive. Dans l'optique de gagner en efficacité et éviter le calcul pour chacun des métabolites dans le réseau élagué par rapport à un métabolite étudié, nous utilisons une approche de programmation logique modélisant l'ensemble complet des contraintes satisfaisant l'ensemble des carrefours topologiques et permettant leur identification avec un unique lancement du programme. Cette approche avait été introduite par [89] pour une application à la complétion topologique de réseau métabolique (voir [Annexe A](#)), implémentée en programmation par ensemble-réponse [33] nommée ASP que nous avons présentée dans le chapitre 2. Nous l'étendons ici pour prendre en compte la notion d'élagage.

Le programme logique permettant de définir récursivement les métabolites du *scope* des métabolites sources dans un réseau métabolique est décrit dans le Listing 4.1. Un élément du *scope* est soit un métabolite source (ligne 1) soit le produit d'une réaction dont chacun des métabolites réactants sont dans le *scope* du réseau (ligne 2). La seconde ligne utilise une notation  $p:q$  qui est satisfaite si  $p$  est vrai pour tout  $q$  possible.

Listing 4.1 – Calcul des métabolites présents dans le *scope* d'un réseau

```

1 scope(M):- seed(M).
2 scope(M2):- product(M2,R), scope(M1):reactant(M1,R).

```

La définition de sous-graphe élagué par rapport à un métabolite est décrite par le programme du Listing 4.2. Dans un premier temps, les métabolites utilisés pour élaguer le réseau sont définis comme limités aux métabolites réactants d'une réaction et présents dans le *scope*. De plus, ils ne doivent être ni métabolites sources, ni métabolites cibles (ligne 1). Ensuite, une réaction  $R$  appartenant au sous-réseau élagué par rapport à un de ces métabolites  $M$  est représenté par le prédicat `pruned(R,M)`. Ce sont toutes les réactions du réseau métabolique qui n'utilisent pas  $M$  comme réactant.

Listing 4.2 – Calcul des graphes élagués

```

1 prune(M):- scope(M), not seed(M), not target(R), reactant(M,R).
2
3 pruned(R,M) :- prune(M), reactant(_,R), not reactant(M,R).

```

Finalement, selon la [Définition 20](#), un métabolite essentiel du point de vue de la persistance du réseau doit être nécessaire pour produire au moins un des métabolites cibles. En d'autres mots, il existe un des métabolites cibles n'appartenant pas au *scope* des métabolites sources dans le sous-réseau élagué vis-à-vis du métabolite ME-persistance considéré. Pour modéliser cette propriété dans un programme ASP (Listing 4.3), premièrement, le *scope* du sous-réseau élagué par rapport au métabolite  $PM$  est calculé (ligne 1-3). Un métabolite ME-persistance  $C$  (`sustainableEM`) est un métabolite qui est nécessaire pour produire un substrat  $TM$ , métabolite cible, ce dernier n'étant pas dans le *scope* des métabolites sources dans le réseau élagué vis-à-vis de  $C$  (lignes 5-6). L'ensemble des métabolites ME-persistance est calculé en prenant en compte tous les substrats possibles (ligne 7).

Listing 4.3 – Calcul des ME-persistance

```

1 scope(M,PM):- seed(M).
2 scope(M2,PM):- product(M2,R), pruned(R,PM),
3                 scope(M1,PM):reactant(M1,R).
4
5 sustainableEM(C,TM):- target(R), reactant(TM,R), scope(TM),
6                       prune(C), not scope(TM,C)
7 sustainableEM(C):- sustainableEM(C,_).

```

## 4.4 Construction de classes de composés métaboliques en fonction du critère d'essentialité vis-à-vis de la persistance

Afin de mieux comprendre le rôle des métabolites essentiels dans la structure d'un réseau métabolique, nous nous sommes intéressés aux relations entre ces derniers. Les études de la structure des réseaux ayant pour but de comprendre la production de métabolites ciblés, nous avons voulu mettre en relation des carrefours ayant le même impact sur la production d'éléments dans le réseau mais, ne voulant pas dépendre de la stœchiométrie de la réaction cible étudiée, nous avons décidé de nous focaliser sur les métabolites essentiels du point de vue de la persistance du réseau (ME-persistance) qui, contrairement aux métabolites essentiels du point de vue de la productibilité du réseau (ME-productibilité) et métabolites essentiels du point de vue de l'efficacité du réseau (ME-efficacité), dépendent directement des métabolites cibles individuellement et pas de la stœchiométrie du système ou de relations stœchiométriques entre les coefficients de la réaction cible.

### 4.4.1 Définition de classes de métabolites ME-persistance

Pour mettre mieux en avant la structure d'un réseau métabolique représenté par un graphe  $G = (R \cup M, E)$ , les métabolites ME-persistance (ou carrefours topologiques) sont regroupés selon leur impact sur l'ensemble des métabolites cibles  $T$  à partir de l'ensemble des métabolites sources  $S$ . Nous définissons l'ensemble  $\mathcal{T}(c)$  des *métabolites cibles duales associées à un carrefour  $c$*  correspondant à l'ensemble des métabolites cibles qui ne sont plus topologiquement accessibles dans les sous-réseaux liés au carrefour  $c$ .

**Définition 21 (Ensemble de cibles duales à un métabolite)** Soit  $G = (R \cup M, E)$  un réseau métabolique,  $S$  un ensemble de métabolites sources et  $r_T$  une réaction cible. L'ensemble des cibles duales au métabolite  $c$  est défini comme suit :

$$\mathcal{T}(c, r_T) = \{t \in rcts(r_T), t \notin \Sigma(\text{prune}(G, c), S)\} \quad (\neq 0 \text{ par défaut})$$

À partir de la définition précédente, nous considérons la relation  $\mathcal{R}_C$  définie sur l'ensemble des métabolites ME-persistance qui regroupe les métabolites ME-persistance liés par une réaction, l'un étant substrat et l'autre produit, et ayant le même ensemble de métabolites cibles duales.

**Définition 22 (Relation d'équivalence entre métabolites ME-persistance)** Soit  $G = (R \cup M, E)$  un réseau métabolique,  $S$  un ensemble de métabolites sources et  $r_T$  une réaction cible. Soient  $c_1$  et  $c_2$  deux métabolites ME-persistance. Les deux métabolites sont en relation  $c_1 \mathcal{R}_C c_2$  si et seulement si :

$$c_1 \mathcal{R}_C c_2 \Leftrightarrow \begin{cases} \exists r \in R \mid c_1 \in rcts(r) \text{ et } c_2 \in prds(r); \\ \mathcal{T}(c_1, r_T) = \mathcal{T}(c_2, r_T). \end{cases}$$

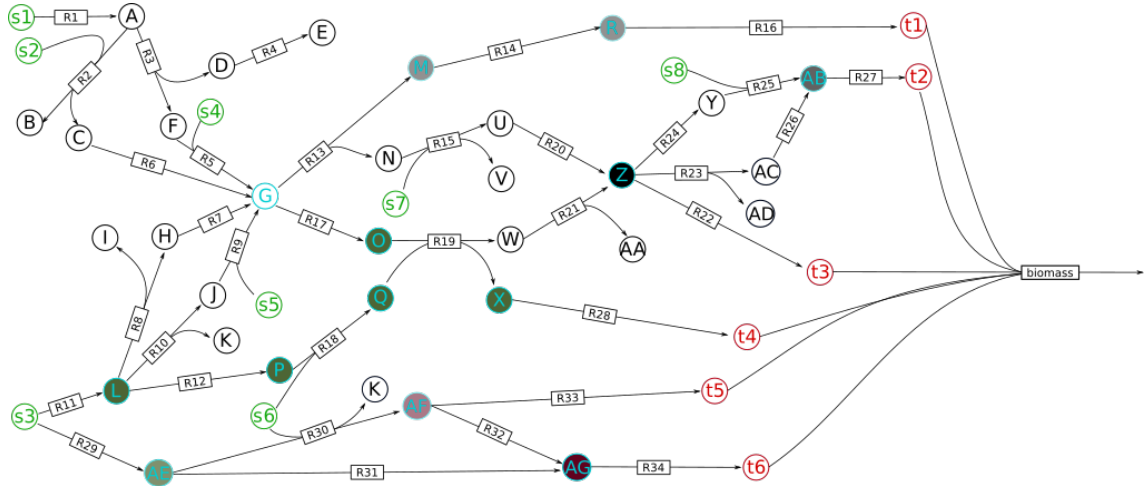
Nous étendons cette relation  $\mathcal{R}_C$  par symétrie, transitivité et réflexivité en une relation d'équivalence  $\overline{\mathcal{R}}_C$ .

**Définition 23 (Cluster de métabolites ME-persistance)** Les clusters d'un réseau métabolique  $G = (R \cup M, E)$  associé à un ensemble de sources  $S$  et une réaction cible  $r_T$  sont les classes d'équivalence de la relation  $\overline{\mathcal{R}}_C$ .

Chaque cluster regroupe les carrefours topologiques (métabolites ME-persistance) liés par une chaîne de réactions (sous-graphe connexe) et dont la suppression a exactement le même effet sur la production des composants de la réaction cible.

### 4.4.2 Exemple de clusters

Un exemple de clusterisation est donné Figure 4.4. Dans cette figure présentant un réseau métabolique de 34 réactions, nous étudions l'accessibilité topologique de six métabolites cibles :  $t1, t2, t3, t4, t5$  et  $t6$ , substrats de la réaction de biomasse, à partir des métabolites sources  $s1, s2, s3, s4, s5, s6, s7$  et  $s8$  symbolisant les nutriments du milieu de croissance. Dans ce réseau métabolique, la réaction de



(a) Réseau métabolique présentant plusieurs métabolites essentiels du point de vue de la persistance du réseau

- $\mathcal{T}(G) = \{t1, t2, t3, t4\}$
- $\mathcal{T}(L) = \{t4\}$
- $\mathcal{T}(M) = \{t1\}$
- $\mathcal{T}(O) = \{t4\}$
- $\mathcal{T}(P) = \{t4\}$
- $\mathcal{T}(Q) = \{t4\}$
- $\mathcal{T}(R) = \{t1\}$
- $\mathcal{T}(Z) = \{t2, t3\}$
- $\mathcal{T}(X) = \{t4\}$
- $\mathcal{T}(AB) = \{t2\}$
- $\mathcal{T}(AE) = \{t5, t6\}$
- $\mathcal{T}(AF) = \{t5\}$
- $\mathcal{T}(AG) = \{t6\}$

(b) Ensembles de cibles duales des métabolites ME-persistance

- $\overline{\mathcal{R}}_{C1} = \{G\}$
- $\overline{\mathcal{R}}_{C2} = \{M, R\}$
- $\overline{\mathcal{R}}_{C3} = \{L, O, P, Q, X\}$
- $\overline{\mathcal{R}}_{C4} = \{Z\}$
- $\overline{\mathcal{R}}_{C5} = \{AB\}$
- $\overline{\mathcal{R}}_{C6} = \{AE\}$
- $\overline{\mathcal{R}}_{C7} = \{AF\}$
- $\overline{\mathcal{R}}_{C8} = \{AG\}$

(c) Clusters de métabolites ME-persistance

FIGURE 4.4 – Exemple de clusterisation des métabolites ME-persistance d'un réseau métabolique

biomasse est topologiquement accessible car l'ensemble des nœuds cibles (substrats de la biomasse) appartiennent au scope des nœuds sources selon la [Définition 11](#).

Nous recherchons, dans le réseau exemple, l'ensemble des métabolites ME-persistence répondant à la [Définition 20](#), nécessaire à l'accessibilité topologique des cibles définies. Plusieurs nœuds métabolites répondent aux critères de cette dernière :

* G	* O	* R	* AB	* AF
* L	* P	* X	* AE	* AG
* M	* Q	* Z		

Le réseau métabolique de l'exemple [4.4](#) présente donc 13 métabolites ME-persistence pouvant être regroupés selon la [Définition 22](#). Ainsi, pour chacun des métabolites ME-persistence présentés ci-dessus, nous définissons son ensemble de cibles duales. Le métabolite ME-persistence  $G$  est associé à l'ensemble de cibles duales  $\{t1, t2, t3, t4\}$  tandis que les métabolites  $M$  et  $R$  ont pour ensemble dual le singleton  $\{t1\}$ , et que le ME-persistence  $AB$  a pour métabolite dual  $\{t2\}$ . Les métabolites  $\{t2, t3\}$  sont les cibles duales associées au métabolite ME-persistence  $Z$  alors que  $\{t4\}$  est le singleton dual lié à  $L, O, P, Q$  et  $X$ . Pour finir, le métabolite ME-persistence  $AE$  est associé à l'ensemble des cibles duales  $\{t5, t6\}$ , le métabolite  $AF$ , au singleton  $\{t5\}$  et le métabolite essentiel du point de vue de la persistance du réseau  $AG$  au singleton également  $\{t6\}$ .

Nous nous intéressons alors à la clusterisation de ces différents métabolites ME-persistence avec la relation d'équivalence introduite précédemment ([Définition 22](#)). Pour cela, les métabolites doivent non seulement avoir le même ensemble de cibles duales mais également être liés par une réaction. Nous pouvons ainsi voir que les métabolites ME-persistence  $M$  et  $R$  ont pour cibles duales  $\{t1\}$  et sont respectivement substrat et produit de la réaction  $r14$ , ils constituent un premier ensemble de métabolites essentiels du point de vue de la persistance du réseau. Les métabolites ME-persistence  $G$  ainsi que  $Z$  ont tous les deux des ensembles de cibles duales uniques, le premier associé à  $\{t1, t2, t3, t4\}$  et le deuxième à  $\{t2, t3\}$ , chacun constituant un cluster singleton dans la répartition des carrefours topologiques. De la même façon, le métabolite  $AB$  est associé au métabolite cible  $\{t2\}$  et est isolé dans le réseau métabolique, il constitue donc lui-même un cluster.

En ce qui concerne les métabolites ME-persistence  $L, O, P, Q$  et  $X$ , ils sont tous associés au même singleton de cibles duales,  $\{t4\}$ , et sont liés deux à deux par des réactions métaboliques permettant de définir les relations suivantes :  $L \mathcal{R}_C P, P \mathcal{R}_C Q, O \mathcal{R}_C X$  et  $Q \mathcal{R}_C X$ . Par transitivité et réflexivité, ces métabolites appartiennent à la même classe d'équivalence de  $\overline{\mathcal{R}}_C$  signifiant qu'ils constituent un même cluster. Finalement, bien que  $AE, AF$  et  $AG$  soient connectés par une chaîne de réactions, ils appartiennent à des clusters différents car ils ne sont pas associés aux mêmes cibles duales :  $\mathcal{T}(AE) = \{t5, t6\}, \mathcal{T}(AF) = \{t5\}$  et  $\mathcal{T}(AG) = \{t6\}$ .

L'exemple présenté dans la [Figure 4.4](#), constitué de 34 réactions métaboliques, permet l'accessibilité topologique de six métabolites cibles à partir de huit métabolites sources en mettant en avant treize métabolites ME-persistence regroupés dans huit clusters.

### 4.4.3 Calcul des clusters de métabolites ME-persistence

Comme précédemment pour le calcul des métabolites ME-persistence, nous utilisons la programmation par ensembles-réponses pour calculer l'ensemble des clusters de ME-persistence dans un réseau métabolique. Le programme ASP permettant de calculer les clusters est divisé en deux parties. Dans un premier temps, la relation entre métabolites ME-persistence ([Définition 22](#)) est décrite par le prédicat `relation(C1,C2)`, calculé pour toutes paires de carrefours  $C1$  et  $C2$ .

Listing 4.4 – Calcul des relations entre métabolites ME-persistence

```

1 relation(C1,C2):- sustainableEM(C1), sustainableEM(C2),
2   1{ R : reactant(C2,R), product(C1,R);
3     R : reactant(C1,R), product(C2,R) },
4   0{ TM : sustainableEM(C1,TM), not sustainableEM(C2,TM) ;
5     TM : sustainableEM(C2,TM), not sustainableEM(C1,TM) } 0.
```

La relation présentée dans le [Listing 4.4](#) est satisfaite s'il existe au moins une réaction  $R$  pour laquelle  $C1$  et  $C2$  sont respectivement réactant et produit (ou l'inverse), et si aucun métabolite cible

TM associé à C1 avec le prédicat `sustainableEM(C, TM)` n'est un métabolite cible associé à C2 et réciproque..

Dans un second temps, cette relation est utilisée dans le prédicat `cluster(C1, C2)` pour définir les clusters de métabolites ME-persistance dans le Listing 4.5. Nous incluons dans la définition un ordre lexicographique afin d'éviter les répétitions, le premier élément du prédicat correspondant au métabolite ME-persistance de référence du cluster et les autres métabolites sont ajoutés par extension de la transitivité. La dernière ligne du programme présenté dans le Listing 4.5 vérifie que l'intersection des clusters de ME-persistance est vide en assurant qu'il n'existe pas de métabolite présent dans deux clusters différents.

Listing 4.5 – Calcul des clusters de métabolites ME-persistance

```

1 cluster(C1, C2):- relation(C1, C2), C1<C2,
2                   not cluster(C3, C1) : sustainableEM(C3), C3<C1;
3                   not cluster(C3, C2) : sustainableEM(C3), C3<C1.
4 cluster(C1, C3):- cluster(C1, C2), relation(C2, C3), C1<C3.
5 :- cluster(X, _), cluster(_, X).
```

Ce programme en ASP permet alors de calculer l'ensemble des clusters de ME-persistance dans un réseau métabolique. Il a été encapsulé dans un logiciel appelé **Spaceman** permettant de calculer dans un premier temps les ME-persistance puis, dans un second temps, les clusters de métabolites, retournant un ensemble de fichiers correspondant chacun à un cluster de métabolites désignés par leur carrefour de référence et contenant la liste des métabolites ME-persistance appartenant au cluster et celle des cibles duales associées. Nous avons pu étudier les résultats de la recherche de ces différents carrefours sur les réseaux métaboliques présentés précédemment.

## Conclusion

Nous avons mis en avant dans un réseau métabolique grâce aux définitions précédentes, plusieurs rôles de métabolites essentiels vis-à-vis de la production d'une biomasse définie. De manière originale, nous nous sommes appuyés sur différentes interprétations de la productibilité d'une réaction cible (ou biomasse) pour en déduire différents concepts de métabolites essentiels. Les métabolites ME-productibilité sont nécessaires à la production de biomasse à l'état quasi-stationnaire; leur définition s'appuie principalement sur la stoechiométrie du réseau métabolique. Les métabolites ME-efficacité sont nécessaires à la production optimale de biomasse; leur définition s'appuie principalement sur un critère d'optimisation MILP (*Mixed integer linear programming* ou optimisation linéaire en nombres entiers). Les métabolites ME-persistance sont nécessaires à la production persistante et pérenne de biomasse, leur définition s'appuyant sur un critère topologique de parcours de graphe. Ce dernier critère permet de définir facilement un concept de cluster de métabolites en fonction des composants cibles touchés par l'élagage d'un réseau vis-à-vis d'un métabolite ME-persistance. De manière intéressante, nous montrons sur des exemples jouets que ces définitions ne sont pas équivalentes. Ce point sera développé plus amplement dans le chapitre suivant via l'étude de six réseaux à échelle génomique. L'application du concept de cluster sera étudié dans le chapitre 6.

Concernant le positionnement bibliographique, ces concepts englobent et étendent les différentes définitions de métabolites essentiels qui peuvent être trouvées dans la littérature. Dans la littérature, le concept le plus proche que nous avons identifié est le concept de *réaction essentielle* pouvant correspondre à deux situations, soit le fait de retirer cette réaction a un effet létal sur la croissance de l'organisme (basé sur le formalisme de production stoechiométrique), soit la réaction porte un flux optimal pour la production optimale de la biomasse (basé sur le formalisme de l'optimisation de la biomasse.) Selon notre formalisme, les substrats des réactions essentielles de la première classe sont des métabolites ME-productibilité tandis que les réactants des réactions essentielles de la seconde classe sont ME-efficacité.

Un autre concept se trouvant dans la littérature est le concept de *métabolite essentiel* qui décrit des métabolites dont la délétion de l'ensemble des réactions consommant ce dernier est létale alors que la suppression des réactions une à une est non létale. Ces métabolites sont également ME-productibilité en accord avec notre formalisme, s'ajoutant aux substrats des réactions essentielles

discutés précédemment. En particulier, les métabolites ME-productibilité sont soit des réactants de réactions essentielles (basées sur la létalité/stoechiométrie), soit des métabolites essentiels, alors que les métabolites ME-efficacité sont des réactants de réactions essentielles (basées sur la FVA). Le concept de métabolite ME-persistance correspond à l'extension du concept de métabolite ME-productibilité au cas d'une sémantique de productibilité décrite par des parcours de graphe.

Pour résumer, notre concept de métabolite essentiel phénotypique ou carrefour décrit un composé tel que le fait d'enlever soit une réaction, soit toutes les réactions consommant ce métabolite affecte la croissance phénotypique (persistance, productibilité ou efficacité) d'une réaction cible correspondant . Cela concerne les notions concernant les réactions essentielles liées au FVA, les réactions essentielles/létales et les métabolites essentiels.

Un autre concept proche est donné par les Ensembles Minimaux de Coupe (*Minimal Cut Set*) consistant en un ensemble de réactions dont la suppression est létale (selon le formalisme basé sur la stoechiométrie). Cependant, la méthode des Ensembles Minimaux de Coupe n'impose aucune contrainte sur les substrats de ces réactions particulières. Notre analyse montre que la situation des métabolites ME-productibilité est un cas très spécifique de cette méthode puisque qu'un métabolite est un ME-productibilité si et seulement si il existe un ensemble minimal de coupe dont toutes les réactions partagent le même composant métabolique en tant que substrat.

Du point de vue logiciel, nous avons développé des méthodes et implémenté un logiciel nommé *Conquests* permettant de calculer efficacement l'ensemble des métabolites ME-productibilité, ME-efficacité et ME-persistance d'un réseau métabolique sous format sbml. Ce logiciel retourne un fichier pour chaque type de métabolite essentiel. De plus, nous avons créé un second logiciel appelé *Spaceman* permettant de calculer dans un premier temps les métabolites essentiels du point de vue de la persistance du réseau puis, dans un second temps, les clusters de métabolites, retournant un ensemble de fichiers correspondant chacun à un cluster de métabolites désignés par leur carrefour de référence et contenant la liste des métabolites ME-persistance appartenant au cluster et celle des cibles duales associées. Nous avons pu étudier les résultats de la recherche de ces différents carrefours sur les réseaux métaboliques présentés précédemment.

## Chapitre 5

# Analyses des métabolites essentiels dans des réseaux à échelle génomique

Dans le chapitre précédent, nous avons mis en avant dans un réseau métabolique plusieurs rôles phénotypiques (production, optimisation, persistance) de métabolites essentiels vis-à-vis de la production d'une biomasse définie. Nous avons montré sur des exemples jouets que ces trois définitions n'étaient pas équivalentes.

Afin de mieux comprendre la nature de ces trois types de métabolites essentiels, dans ce chapitre, nous étudions les métabolites essentiels pour les six réseaux métaboliques que nous avons présentés dans la section 1.4 : les trois réseaux de *E. coli* *iAF1260*, *iJO1360* et *iJR904* ainsi que les réseaux métaboliques de *Synechocystis*, *A. ferrooxidans* et *T. lutea*. Comme nous le verrons, nous mettrons en avant que la comparaison fine des différentes propriétés permet de pointer le rôle subtil de composés du réseau, de manière à mieux comprendre le fonctionnement du système. L'ensemble des résultats présentés dans ce chapitre a fait l'objet d'une publication dans le journal *PeerJ* [67].

Dans la suite, pour ne pas porter à confusion avec les métabolites essentiels de la littérature, nous désignerons par "carrefours" l'ensemble des métabolites essentiels (ME-persistance, ME-productibilité et ME-efficacité).

### 5.1 Etude quantitative

#### 5.1.1 Un nombre important de carrefours

Nous avons calculé l'ensemble des métabolites carrefours pour les six réseaux considérés. Les résultats sont décrits dans la Table 5.1. Notre étude des six réseaux métaboliques suggère qu'un réseau métabolique contient quelques petites centaines de métabolites carrefours, qu'ils soient ME-persistance, ME-productibilité ou ME-efficacité. Plus précisément, notre analyse montre que le nombre de carrefours se trouve entre 113 et 423 dans les exemples étudiés, ce qui représente 7% à 70,5% du réseau métabolique pour les espèces étudiées. De manière plus importante, le nombre de carrefours apparaît comme indépendant de la taille du réseau. Ainsi le réseau métabolique ayant le plus grand nombre de métabolites carrefours est celui de *Synechocystis*, le quatrième réseau par la taille tandis que celui avec le moins de carrefours est le réseau *iJR904* de *E. coli*, le troisième plus grand réseau.

Les observations que nous avons mises en avant suggèrent qu'un réseau métabolique ayant un faible pourcentage de carrefours correspond à un nombre relativement élevé de voies métaboliques au sein du corps principal de l'organisme qui est lié à la modélisation de la biomasse. Dans le cas contraire, les réseaux les plus récents dont les organismes sont moins étudiés tel que *T. lutea* se concentrent sur la description du métabolisme primaire, dû au petit nombre d'expérimentations réalisées sur ces organismes en comparaison aux réseaux modèles. Cela peut suggérer que le pourcentage de carrefours, quel que soit leur type, est un indicateur à la fois de la variété de phénotypes rencontrés pour l'organisme et du degré de curation réalisé sur le réseau métabolique.



	Nombre de réactions	Nombre de métabolites	Nombre total de métabolites essentiels (ou carrefours)	ME-persistance	ME-productibilité	ME-efficacité
<i>iJR904</i>	1075	904	113 (12.5%)	70 (7.7%)	65 (7.2%)	87 (9.6%)
<i>iAF1260</i>	2382	1967	177 (9.0%)	98 (5%)	105 (5.3%)	166 (8.4%)
<i>iJO1360</i>	2582	2129	150 (7.1%)	66 (3.1%)	67 (3.1%)	143 (6.7%)
<i>Synechocystis</i>	759	600	423 (70.5%)	325 (54.2%)	380 (63.3%)	375 (62.5%)
<i>A. ferrooxidans</i> st. Wenelen	620	579	376 (64.9%)	316 (54.6%)	340 (58.7%)	372 (64.2%)
<i>T. lutea</i>	316	324	210 (64.8%)	198 (61.1%)	208 (64.2%)	208 (64.2%)

TABLE 5.1 – Nombre total de carrefours pour les six réseaux métaboliques des espèces procaryotes et celle unicellulaire eucaryote. Le pourcentage de chaque cellule correspond au ratio de métabolites ayant la propriété correspondante (désignée par le nom de la colonne) dans le réseau par rapport au nombre total de métabolites dans ce dernier.

### 5.1.2 Tous les éléments hautement connectés sont des métabolites essentiels, des cibles ou des sources

Nous nous sommes ensuite intéressés aux liens entre les métabolites carrefours ME-persistance, ME-productibilité et ME-efficacité et les caractéristiques des nœuds métabolites dans le graphe modélisant le réseau, notamment au niveau de la connectivité de ces derniers. Dans la Figure 5.1, les composants de chaque réseau métabolique sont ordonnés selon leur degré de connectivité et leur rôle dans le fonctionnement du réseau, i.e. carrefour, métabolite cible, métabolite source ou composant générique (ceux qui n'ont aucun rôle identifié dans le fonctionnement en analysant les carrefours).

Nous pouvons voir que dans les réseaux *iJR904*, *iJO1360* et *iAF1260*, les 14 métabolites dont le degré est le plus haut (avec une connectivité entre 70 et 1040) ont un rôle important dans le fonctionnement de ces derniers. L'unique exception est la molécule *H2O* du périplasma dans le réseau *iJO1360*, un composé qui paraît jouer un rôle redondant dans le réseau avec la molécule *H2O* cytoplasmique. Dans les réseaux *A. ferrooxidans* et *T. lutea*, les 31 composés dont le degré est le plus élevé (avec une connectivité entre 10 et 310) ont un rôle dans le fonctionnement du réseau.

Notre analyse suggère alors que, quel que soit le réseau considéré, la totalité des composants fortement connectés se répartit entre carrefours, métabolites cibles, et métabolites sources et a donc un rôle fonctionnel dans le réseau.

### 5.1.3 Complémentarité

Les différents types de carrefours, ME-persistance, ME-productibilité et ME-efficacité, ne s'excluent pas, au contraire, de nombreux carrefours se retrouvant à l'intersection de plusieurs définitions. La Table 5.2 décrit le nombre de carrefours de chaque type dans les différents réseaux métaboliques choisis. En étudiant cette dernière, nous pouvons voir que le type de carrefour le plus fréquent est ME-efficacité : en se référant aux divers réseaux métaboliques étudiés, de 77% à 99% des carrefours sont ME-efficacité.

Les nombres de carrefours ME-persistance et ME-productibilité sont relativement comparables : le ratio de carrefours ME-productibilité est compris entre 44.7% (réseau *iJO1360*) et 99% (réseau *A. ferrooxidans*), tandis que le ratio de carrefours ME-persistance se trouve entre 46% (réseau

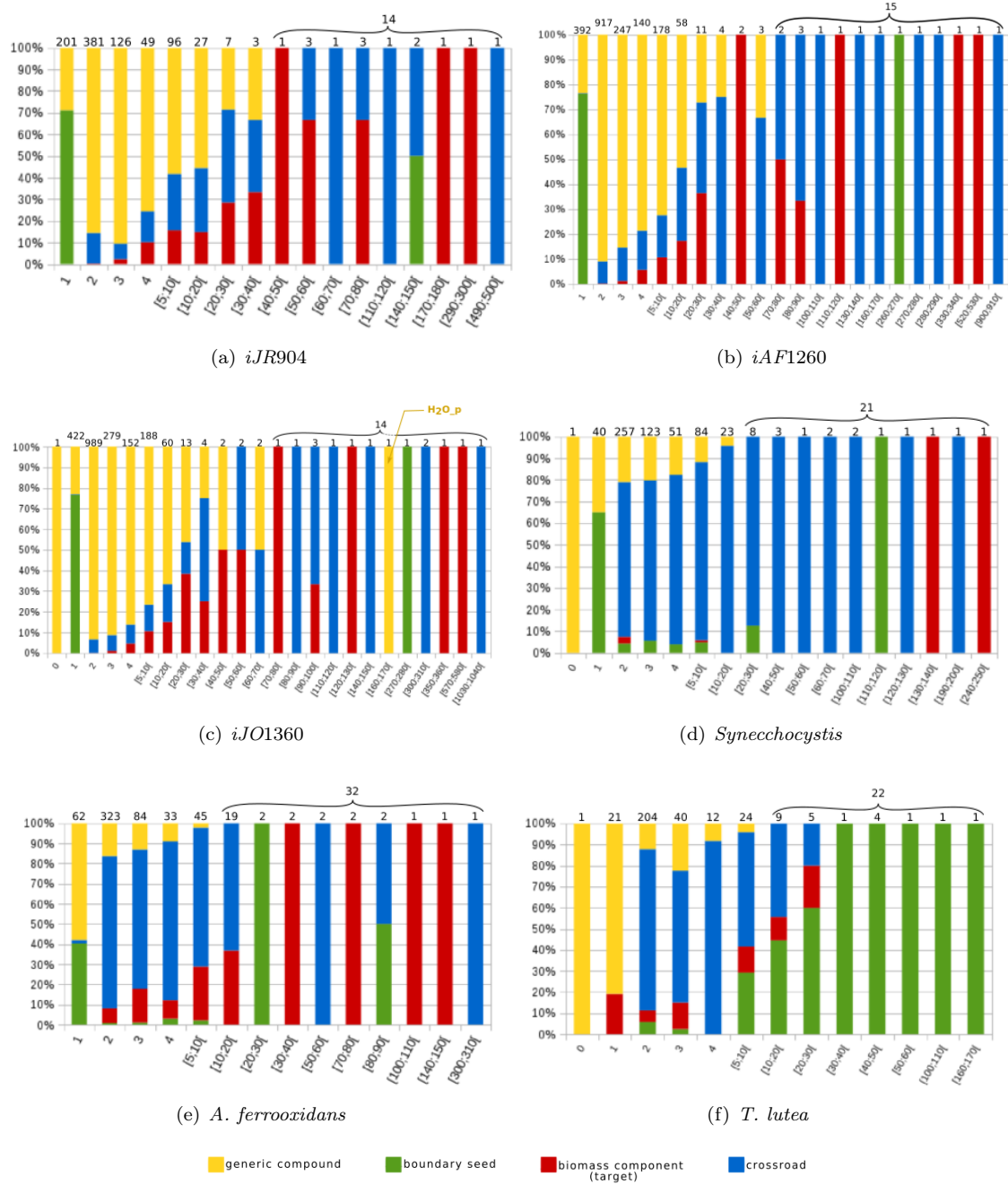


FIGURE 5.1 – Classification des composants métaboliques selon leur degré de connectivité (axe des abscisses) et leur rôle dans le fonctionnement du réseau. Pour chaque intervalle  $[a, b]$  représenté sur l'axe des  $x$ , la totalité des composés dont le degré de connectivité appartient à  $[a, b]$  sont regroupés dans les quatre classes de métabolites répartis selon leur rôle dans le fonctionnement du réseau : carrefour (bleu), nutriment limite (vert), composé de la biomasse (rouge), autre composant générique sans aucun rôle fonctionnel spécifique (jaune). Le pourcentage relatif à chacune de ces classes est représenté sur l'axe des ordonnées.

	Nombre de carrefours	& ME-persistance & ME-productibilité & ME-efficacité	ME-persistance	ME-productibilité	ME-efficacité	& ME-persistance & ME-productibilité	& ME-persistance & ME-efficacité	& ME-productibilité & ME-efficacité
<i>iJR904</i>	113	47 (41%)	11	4	36 (32%)	11	1	3
<i>iAF1260</i>	177	89 (44%)	1	2	71 (35%)	8	0	6
<i>iJO1360</i>	150	56 (37%)	1	3	76 (50%)	3	6	5
<i>Synechocystis</i>	423	292 (69%)	9	17	32 (8%)	22	2	49
<i>A. ferrooxidans</i>	376	306 (81%)	1	1	28 (8%)	2	7	31
<i>T. lutea</i>	210	196 (93%)	0	1	1 (0.1%)	1	1	10

TABLE 5.2 – Nombre des différents types de carrefours isolés et dans les intersections

*iJO1360*) et 86% (réseau *T. lutea*). De façon similaire, pour tous les réseaux, certains carrefours ne sont ni ME-persistance, ni ME-productibilité. Cette analyse confirme que les trois classes de carrefours sont complémentaires et doivent être considérées ensemble pour entrevoir la complexité du fonctionnement du réseau.

#### 5.1.4 Intersection des types de carrefours : un squelette du réseau métabolique

Les intersections entre les différentes classes de composés dans les six réseaux métaboliques sont montrée dans la Figure 5.2. Au moins 26% (et jusqu'à 93%) des métabolites essentiels sont liés à la fois à la productivité, à l'efficacité et à la persistance du réseau. En comparant les composants présentant les trois caractéristiques dans les réseaux *iJR904*, *iAF1260* et *iJO1360*, nous avons constaté le fait que 20 métabolites ME-productibilité, ME-efficacité et ME-persistance sont présents dans les trois réseaux, dont les nœuds très connectés *udp*, *h*, *phosphate* et *phosphoenolpyruvate*. Les quatre composés peuvent en conséquence être considérés comme des régulateurs majeurs de la production de biomasse de *E. coli* dans le graphe, selon des formalismes stœchiométrique et topologique. Étonnamment, les 16 autres métabolites essentiels dans le squelette commun des réseaux *iJO1360*, *iAF1260* et *iJR904* possèdent un degré de connectivité faible dans le réseau (de 2 à 7). Ces composants sont *Chorismate*, *DTDP*, *ahdt*, *Dihydroneopterin*, *Dephospho-CoA*, *1,2-Dihydronaphthalene-1,2-diol*, *6-hydroxymethyl-dihydropterin pyrophosphate*, *4-amino-4-deoxychorismate*, *Dihydroneopterin monophosphate*, *6-hydroxymethyl dihydropterin*, *N-((R)-4-Phosphopantothenoyl)-L-cysteine*, *6-hydroxymethyldihydropterin*, *5-O-(1-Carboxyvinyl)-3-phosphoshikimate*, *D-4'-Phosphopantothenate*, *Pantetheine 4'-phosphate*. Une interprétation est que ces métabolites font référence à une structure des composés du réseau nécessaire pour produire la biomasse et peuvent constituer un *squelette* de la structure du réseau concernant la production de la biomasse.

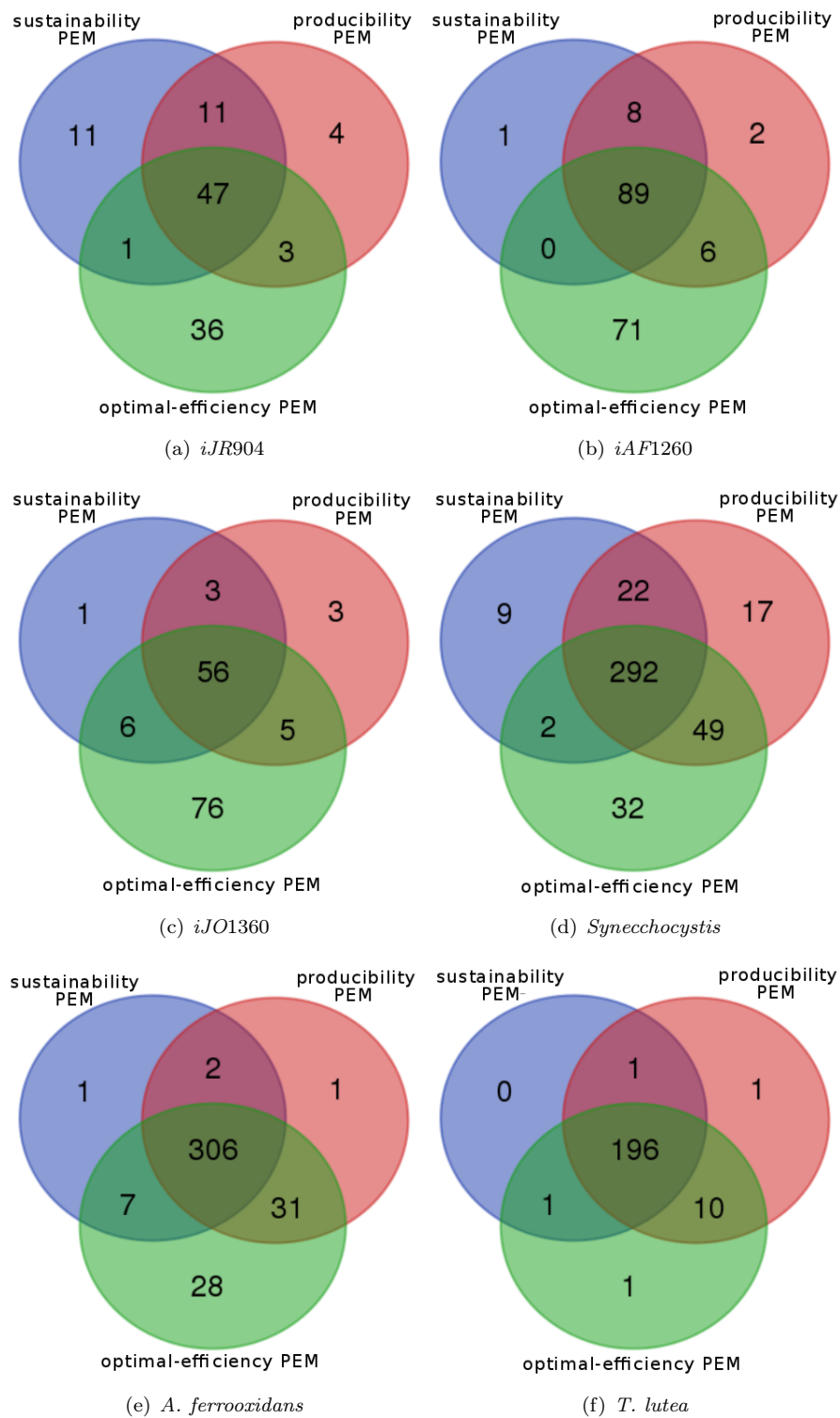


FIGURE 5.2 – Intersections des ensembles de métabolites essentiels (ME-persistance, ME-productibilité et ME-efficacité) pour six réseaux métaboliques à échelle génomique.

### 5.1.5 Intersections vides ou peu représentées : indice sur les protocoles de reconstruction

Nous pouvons voir dans la Figure 5.2 et la Table 5.2 que certaines intersections sont soit vides, soit contiennent très peu de carrefours dans certains des réseaux métaboliques étudiés. Ainsi, si nous nous intéressons par exemple au réseau *iAF1260* de *E. coli*, nous pouvons remarquer que tous les métabolites ME-persistance et ME-efficacité sont aussi des métabolites ME-productibilité. Une interprétation de ce phénomène est que la productibilité dans ce réseau s'appuie principalement sur des équilibres optimaux de co-facteurs. Ce point sera détaillé plus tard dans la section 5.4. La faible présence de métabolites ME-persistance en dehors du squelette (9 éléments) suggère également que le réseau *iAF1260* est essentiellement modélisé pour des études stœchiométriques.

Cette analyse est valable de la même façon dans le cas du réseau métabolique de *T. lutea* dont les carrefours du squelette représente 93% de l'ensemble des carrefours, mettant ainsi en avant un réseau métabolique cœur.

## 5.2 Identification de redondances dans le réseau

Après avoir évalué les propriétés génériques des carrefours dans les six réseaux à grande-échelle que nous étudions, nous nous sommes concentrés sur l'interprétation biologique des métabolites qui vérifient une ou deux des trois propriétés d'essentialité. Il s'agit donc des métabolites qui sont dans les feuilles des diagrammes de Wenn présentés dans la Figure 5.2. L'intérêt de cette étude est que ces métabolites sont relativement peu nombreux dans chacun des exemples étudiés. Dans un cadre applicatif, ils peuvent donc faire l'objet d'analyses bibliographiques poussées. Le but des sections suivantes est donc d'interpréter ces propriétés pour guider de futures analyses.

Dans cette première partie, nous nous concentrons sur les propriétés liées aux carrefours qui retranscrivent une redondance du système biologique. Selon notre interprétation, une telle redondance apparaît pour les métabolites qui vérifient les propriétés de ME-persistance et ME-productibilité (sans être ME-efficacité) d'une part, et les métabolites qui sont exclusivement ME-efficacité d'autre part.

### 5.2.1 Propriétés des métabolites ME-persistance et ME-productibilité qui ne sont pas ME-efficacité.

Une telle situation se produit lorsqu'un composé est nécessaire pour produire un métabolite cible des points de vue topologique et stœchiométrique mais pas au niveau de l'optimalité. Ainsi, chacune des réactions consommant ce composé peut être enlevée indépendamment sans impacter sur la production optimale de la biomasse. La principale explication est qu'un tel métabolite est le substrat d'au moins deux voies métaboliques équivalentes, alternatives topologiquement et stœchiométriquement, pour produire un métabolite cible.

Un exemple tiré des réseaux étudiés est représenté dans la Figure 5.3. Dans le réseau *A. ferrooxidans*, le composé  $\alpha$ -D-glucose 1-phosphate qui est un métabolite ME-persistance, ME-productibilité et ME-efficacité, est un précurseur de deux métabolites cibles : *glycogen* et *lps\_AFE*. Il est essentiel pour leur production topologique, stœchiométrique et optimale.

La production de ce composé est assurée par une voie métabolique commençant par le métabolite  $\beta$ -D-fructofranose 6-phosphate (qui est également ME-persistance, ME-productibilité et ME-efficacité), transformé ensuite en  $\beta$ -D-glucose 6-phosphate, qui, lui, est un métabolite ME-persistance et ME-productibilité mais n'est pas un métabolite ME-efficacité car la voie se divise alors en deux voies possibles. Ainsi ce dernier peut être transformé soit en  $\alpha$ -D-glucose 6-phosphate, soit en  $\beta$ -D-glucose 1-phosphate. Aucune autre voie ne permet la production du  $\alpha$ -D-glucose 1-phosphate et aucune des deux n'est privilégiée par rapport à l'autre, n'étant essentielle pour aucun type de production, topologique, stœchiométrique ou optimale, expliquant le statut de carrefour topologique et stœchiométrique mais non optimal du  $\beta$ -D-glucose 6-phosphate.

Dans les six réseaux métabolites, cette propriété n'est pas fréquente dans les réseaux *iJO1360*, *A. ferrooxidans* et *T. lutea*, où nous pouvons compter respectivement 3, 2 et 1 métabolites ME-

persistance et ME-productibilité mais non ME-efficacité. Dans les réseaux *Synechocystis*, *iAF1260* et *iJR904*, cette situation est plus fréquente et se produit de 8 à 22 fois. En particulier, le nombre relativement élevé de tels métabolites ME-persistance et ME-productibilité seulement, dans le réseau *Synechocystis* suggère que le fonctionnement optimal est peu contraint. Cela peut être expliqué par la présence d'un nombre important de flux d'export, que nous avons noté lors de l'étude des métabolites qui ne sont que ME-productibilité, ce qui génère des redondances dans les voies métaboliques intracellulaires.

### 5.2.2 Propriété des métabolites qui sont uniquement ME-efficacité

Un carrefour exclusivement ME-efficacité est, selon la [Définition 19](#), nécessaire pour produire de la biomasse de façon optimale. Cependant, puisqu'il n'est pas un métabolite ME-persistance ou ME-productibilité, le fait d'enlever toutes les réactions utilisant le carrefour comme réactant n'empêche pas l'accessibilité topologique et stœchiométrique des métabolites cibles constituant les substrats de la biomasse.

Cette situation apparaît essentiellement lorsque les métabolites cibles de la biomasse peuvent être produits par au moins deux voies métaboliques distinctes mais que l'un des chemins est privilégié par le système lors de l'optimisation des flux pour la production maximale de la biomasse.

Un exemple est représenté dans la [Figure 5.4](#) modélisant une partie du réseau compartimenté *Synechocystis*. Dans ce dernier, le métabolite *putrescine* cytosolique est un précurseur des métabolites cibles. Il s'agit d'un métabolite ME-persistance, ME-productibilité et ME-efficacité. Dans ce réseau, deux voies métaboliques peuvent produire ce composé : l'une d'elles utilise un flux d'importation de *putrescine* venant du compartiment extracellulaire et, l'autre, via des voies internes, est caractérisée par un flux de production faisant intervenir en premier lieu le métabolite *L-arginine* qui vérifie les trois propriétés d'essentialité. Ce métabolite est utilisé pour produire l'*agmathine*, qui est un métabolite ME-efficacité mais non ME-persistance ni ME-productibilité, ainsi qu'un co-produit de la voie métabolique interne à l'organisme qui est un métabolite ME-productibilité et ME-efficacité mais pas ME-persistance. Ceci indique l'importance du métabolite dans une production stœchiométrique mais pas topologique.

Ce lien de coproduction a pour conséquence dans le réseau de *Synechocystis* l'utilisation favorisée de la voie métabolique comprenant la *L-arginine* pour produire le métabolite *putrescine* dans le but de produire de façon optimale la biomasse. Cela est mis en avant par la caractérisation des réactions de cette voie comme essentielles dans l'étude de l'optimalité des flux. Cette sélection d'une voie métabolique explique le fait que l'*agmathine* est un métabolite ME-efficacité, de même que l'*urée*, car ils ne sont pas nécessaires pour l'accessibilité topologique et stœchiométrique de la *putrescine* mais permettent d'optimiser les flux dans le réseau de *Synechocystis*. En effet, en ce qui concerne le carrefour *putrescine*, la suppression des réactions ayant ce dernier comme substrat entraîne une diminution de la croissance de la biomasse de 47.5 à 41.1 mmol/gDW/h.

Dans les six réseaux étudiés, le cas des métabolites ME-efficacité qui ne sont ni ME-persistance ni ME-productibilité est la situation la plus fréquente après le cas des métabolites carrefours présentant les trois propriétés, ME-persistance, ME-productibilité et ME-efficacité. Dans les réseaux *iJR904*, *iAF1260* et *iJO1360*, nous retrouvons entre 32% et 50% de carrefours dans cette situation tandis que les réseaux les plus récents *Synechocystis* et *A. ferrooxidans* possèdent 10% de carrefours ME-efficacité. Le réseau *T. lutea* fait exception avec peu de différences entre les critères topologiques, stœchiométriques et optimaux.

## 5.3 Impact de l'initiation des cycles métaboliques

Nous avons identifié un second ensemble de propriétés qui sont, de leur côté, relatives à la différence entre les critères topologique et stœchiométrique. Comme nous allons le voir, ces situations semblent suggérer un impact de l'accessibilité du milieu de culture sur la capacité de la cellule à produire de la biomasse.

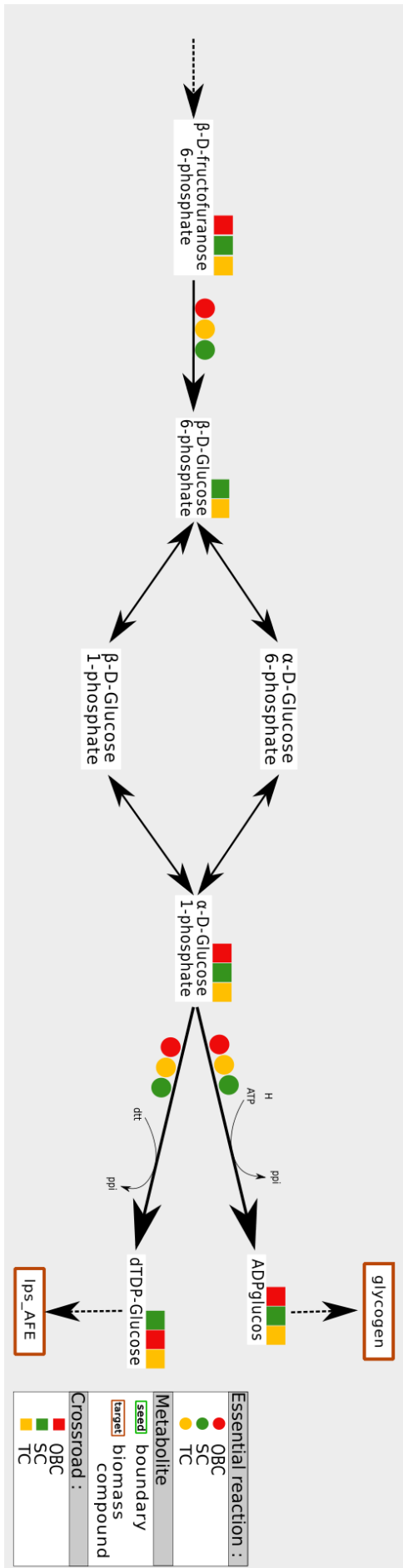


FIGURE 5.3 – Redondance du réseau : exemple d'un métabolite à la fois un ME-persistance et ME-productibilité mais qui n'est pas un ME-efficacité. Dans le réseau de *A. ferrooxidans*, la production de *glycogen* et de *lps\_AFE* peut se dérouler via deux paires de réactions alternatives, en utilisant soit  $\alpha$ -D-glucose 6-phosphate soit  $\beta$ -D-glucose 1-phosphate. De plus,  $\beta$ -D-glucose 6-phosphate n'est pas un métabolite ME-efficacité mais est un métabolite ME-persistance et ME-productibilité.

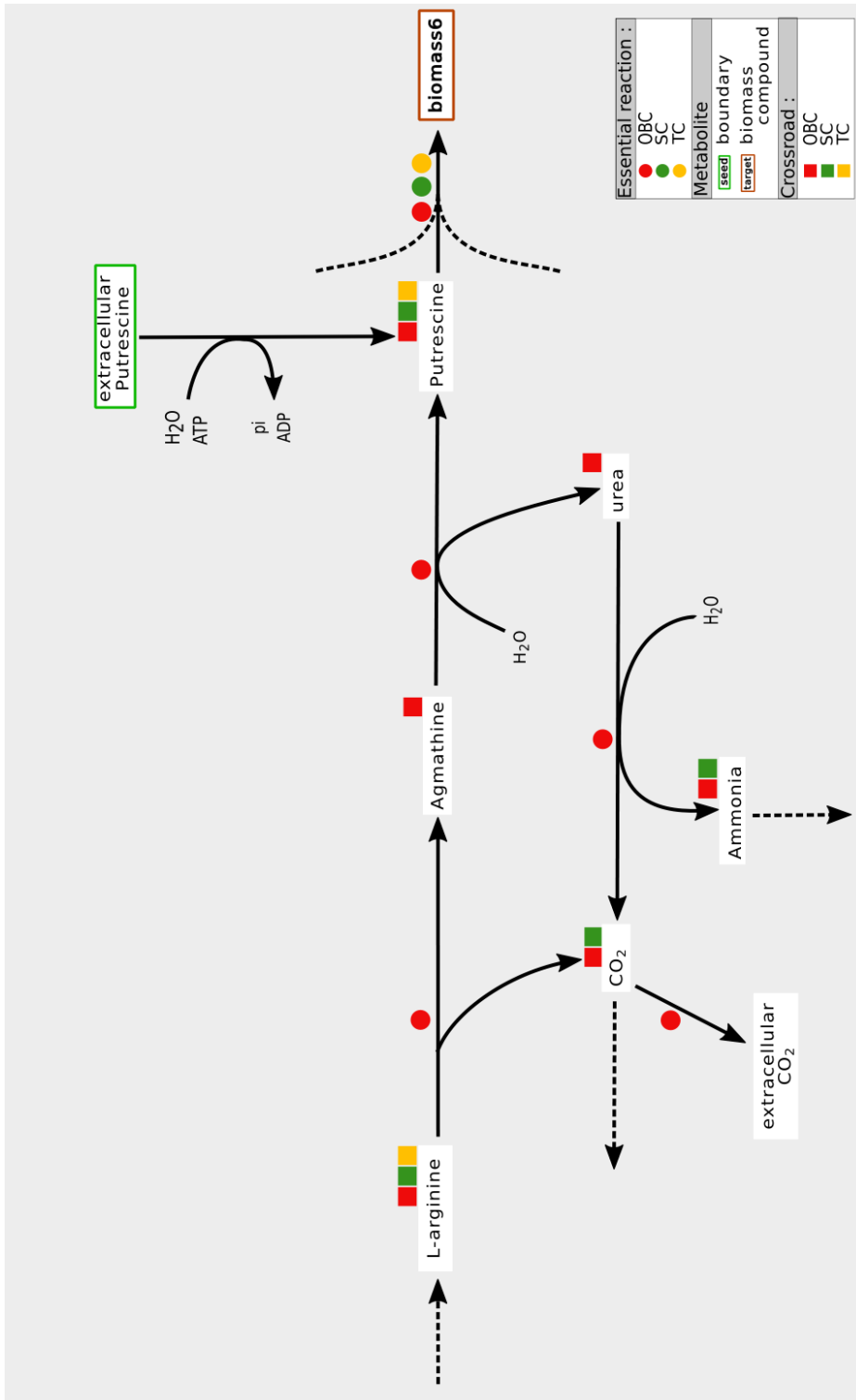


FIGURE 5.4 – Un choix optimal de voies métaboliques : exemple d'un métabolite ME-efficacité qui n'est ni ME-persistence ni ME-productivité. Dans le réseau de *Synechocystis*, la production de *Putrescine* est assurée par une voie métabolique initiée par *L-arginine* au détriment d'une voie métabolique alternative qui consiste à importer de la *Putrescine* depuis le milieu extracellulaire. La raison principale à cela est que la voie métabolique utilisant *L-arginine* permet aussi la production de *Ammonia*. Dans ce scénario, *urea* et *Agmatine* sont des métabolites ME-efficacité mais ne sont ni des métabolites ME-persistence ni des métabolites ME-productivité.



### 5.3.1 Propriété des métabolites exclusivement ME-persistence

Un carrefour ne satisfaisant que le critère topologique est un métabolite nécessaire pour l'accessibilité topologique d'un métabolite cible. Cependant ce composé n'est pas nécessaire pour produire ce même métabolite cible d'un point de vue stœchiométrique, optimal ou non. Par exemple, en se référant au critère lié au graphe, dans le réseau *iJO1360*, le métabolite *thiamine* du cytosol est nécessaire pour produire la cible *thiamine diphosphate*. En effet, si nous étudions le sous-réseau lié à ce métabolite, en enlevant toutes les réactions dont le substrat est la *thiamine* cytosolique, le métabolite cible *thiamine diphosphate* n'est plus accessible topologiquement. Par contre, dans le sous-réseau ainsi créé, la biomasse est toujours stœchiométriquement productible de telle sorte que le ratio de croissance maximale, lié à la réaction de biomasse, reste inchangé par rapport à ce même ratio dans le réseau initial (11,747 mmol/gDW/h).

Ce paradoxe apparent est principalement expliqué par les hypothèses d'état stationnaire et de conservation de masse liées à la dynamique du réseau et sous-entendues par les sémantiques d'accessibilité topologique et stœchiométrique. Nous supposons que ces différentes sémantiques ont un grand impact sur l'interprétation du fonctionnement des cycles au sein du réseau et de voies secondaires non analysées lors d'études stœchiométriques pour diverses raisons et pouvant mettre en évidence des voies d'initiation de ces cycles métaboliques.

Un exemple est représenté dans la Figure 5.5. Dans le réseau *iJR904*, le composé *Peptidoglycan subunit* (Peptido\_EC) cytosolique, un métabolite cible, est produit à partir du métabolite *diphosphoundecaprenyl* intervenant dans un cycle dont l'ensemble des métabolites vérifient chacun les trois propriétés ME-persistence, ME-productibilité et ME-efficacité.

Dans le cas de l'hypothèse d'état stationnaire, l'analyse stœchiométrique montre que la production de *Peptidoglycan subunit* est assurée par le cycle, qui est auto-activé, puisque toutes les réactions du cycle sont essentielles pour la production topologique, stœchiométrique et optimale de la biomasse. Cependant ce type d'analyse assume implicitement qu'au moins un des métabolites du cycle est présent lors de l'activation de ce système et qu'aucun d'entre eux n'est dégradé durant le fonctionnement de ce dernier. D'un autre côté, d'un point de vue topologique, l'analyse montre que les cycles doivent d'être initialisés à partir d'une voie métabolique externe pour être considérés comme accessibles.

Dans l'exemple que nous présentons dans la Figure 5.5, le cycle est initialisé par une voie métabolique linéaire de dix réactions commençant par le métabolite *pyruvate* (qui est un métabolite ME-persistence, ME-productibilité et ME-efficacité), et du composant *3-phospho-D-glycerate* qui est un métabolite ME-productibilité et ME-persistence mais pas ME-efficacité. Notre analyse montre que les neuf composés de cette voie métabolique linéaire sont nécessaires pour l'accessibilité topologique du métabolite cible *Peptidoglycan subunit* mais qu'ils ne sont pas nécessaires pour produire ce composé de la biomasse. En effet, *Peptido\_EC* étant produit à partir d'un cycle auto-alimenté dans un état stationnaire (accessibilité stœchiométrique), les réactions composant cette voie ont la valeur de leur flux égale à zéro dans l'analyse de flux.

Cette analyse suggère que la production de *Peptidoglycan subunit* est possible via un cycle auto-activé, initialisé dans un premier temps par le *pyruvate*.

Pour résumer, les métabolites ME-persistence qui ne sont ni ME-productibilité ni ME-efficacité suggèrent l'existence de cycles auto-alimentés dans l'analyse des flux du réseau et proposent des candidats pour l'initialisation de ces derniers lorsque la cellule ou l'organisme n'est pas dans un état stationnaire. Cependant, dans les six réseaux métabolites étudiés, ce phénomène est observé dans peu de cas : une unique fois dans les réseaux *iAF1260*, *iJO1360* et *A. ferrooxidans*, neuf fois dans le réseau *Synechocystis* et onze fois dans le réseau *iJR904*. Cela suggère qu'un nombre restreint de métabolites ME-persistence qui ne sont ni ME-productibilité ni ME-efficacité peut être considéré comme une métrique d'un réseau métabolique de qualité.

### 5.3.2 Propriété des métabolites ME-persistence et ME-efficacité qui ne sont pas ME-productibilité

Cette situation est relativement délicate. En effet, lorsqu'un composé a cette propriété, cela signifie qu'il est le substrat d'une réaction dont le flux est toujours positif quand la cellule produit

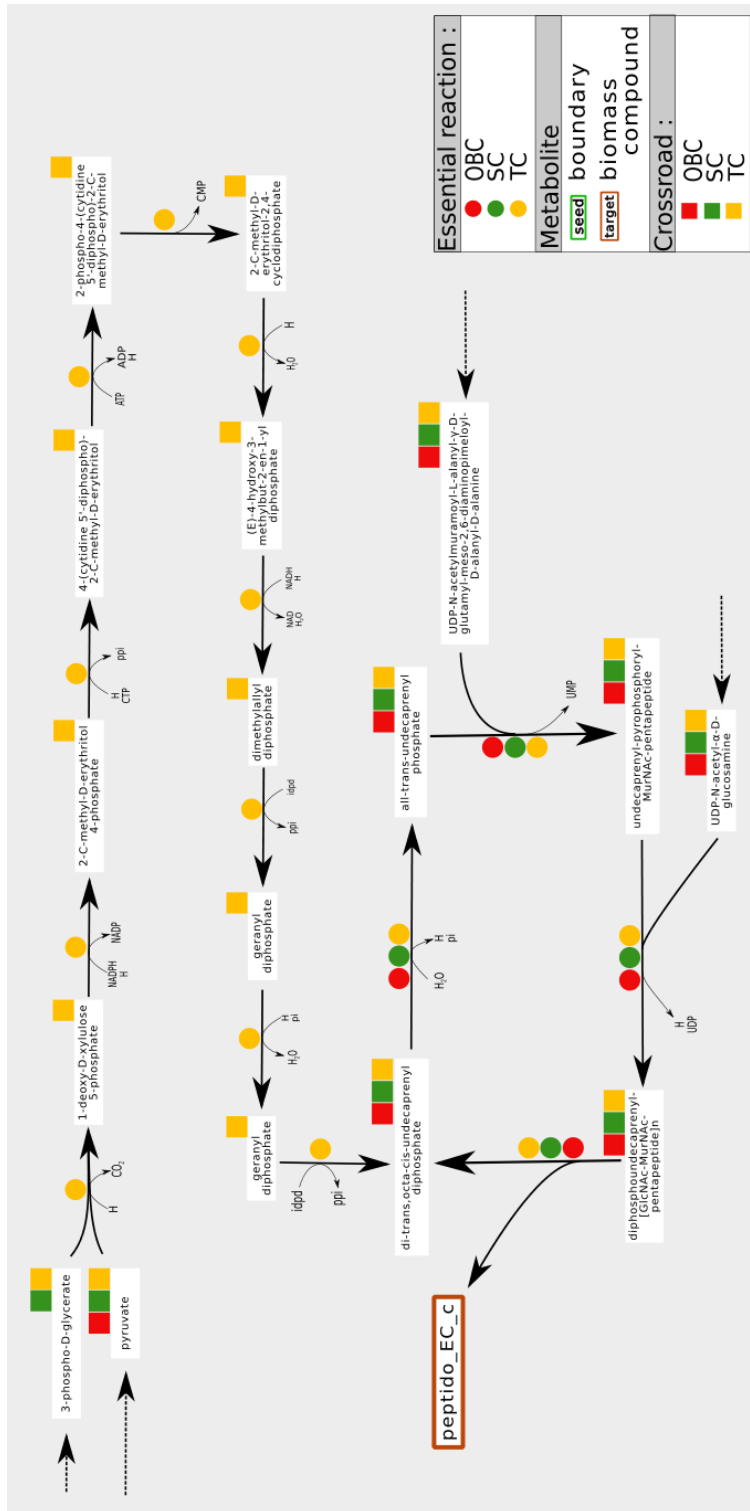


FIGURE 5.5 – Voies métaboliques dont l'activation est dépendante d'un état initial de la cellule. Exemple d'un métabolite ME-persistence qui n'est ni ME-productibilité ni ME-efficacité. Dans le réseau *iJR904*, la production d'un composant de la biomasse *peptido\_EC* (Peptidoglycan subunit) est assurée par un cycle partant de *Undecaprenyl diphosphate* jusqu'à *diphosphoundecaprenyl*. L'ensemble des éléments impliqués dans ce cycle sont des métabolites ME-persistence, ME-productibilité et ME-efficacité (carrés rouge, vert & jaune). Le cycle est initialisé par une voie métabolique commençant de *pyruvate* et composé de neuf métabolites ME-persistence qui ne sont ni ME-productibilité ni ME-efficacité. En effet, à l'état stationnaire, l'activation de la voie métabolique d'initiation n'est pas requise pour la production d'une biomasse optimale. Alors que cette voie métabolique peut être nécessaire pour initier le cycle hors état stationnaire. Ce rôle lié au fonctionnement du réseau métabolique est reflété par la propriété des métabolites ME-persistence.

de la biomasse avec un ratio optimal. La propriété liée au métabolite ME-persistance signifie que la suppression de cette réaction du réseau implique que l'un des métabolites cibles, substrats de la réaction de biomasse, n'est plus accessible selon des critères topologiques. Cependant, le métabolite cible garde sa capacité d'être accessible au niveau stœchiométrique, bien que le ratio de croissance soit diminué.

Une autre manière d'expliquer ce cas est qu'il existe une voie alternative de production du métabolite cible mais que ce chemin alternatif n'est pas accessible à partir des métabolites sources du réseau métabolique d'un point de vue topologique. Par exemple, dans le réseau *A. ferrooxidans*, le fait d'enlever les réactions consommant le métabolite *Alpha-D-Ribose\_1-phosphate* entraîne une diminution du ratio de production de la biomasse optimale de 3,6 à 3,4 mmol/gDW/h. De plus, bien qu'il y ait toujours production de biomasse, il n'existe plus de chemin partant de l'ensemble des métabolites sources et atteignant les métabolites cibles. L'explication de ce paradoxe est similaire au cas présenté dans la Figure 5.5, en relation avec un cycle auto-alimenté.

Pour illustrer le cas d'un métabolite ME-persistance et ME-efficacité qui n'est pas ME-productibilité, la Figure 5.6 représente un sous-réseau du réseau *iJO1360*. Dans le compartiment cytosolique, le métabolite *thiamine phosphate*, un carrefour avec les trois propriétés d'essentialité (ME-persistance, ME-productibilité et ME-efficacité), est le précurseur unique du *Thiamine diphosphate*, un métabolite cible. Sa production optimale est assurée grâce à une voie métabolique, initialisée à partir des métabolites sources, qui importe le *Thiamine* du compartiment extracellulaire au périplasma puis de ce dernier au cytosol où le *Thiamine* est transformé en *Thiamine phosphate*.

L'ensemble des réactions présentes dans cette voie métabolique est essentiel pour accéder topologiquement au *Thiamine diphosphate* et pour produire de façon optimale la biomasse mais non pour produire stœchiométriquement cette dernière. En effet, lorsque la réaction transformant le *Thiamine* en *Thiamine phosphate* est supprimée du réseau métabolique, ce dernier a la possibilité d'activer un cycle permettant de produire le *4-methyl-5-(2-phosphoethyl)-thiazole* qui, avec le *2-methyl-4-amino-5-hydroxymethylpyrimidine diphosphate*, permet également la création de *Thiamine phosphate*.

Nous pouvons cependant noter que ce cycle n'est pas utilisable à partir des métabolites sources car aucun de ses composants n'est accessible topologiquement, alors que, dans une étude stœchiométrique, avec les hypothèses d'état stationnaire et de conservation de masse, le cycle est accessible. De plus, l'activation du cycle est très dépendante de la présence d'au moins un de ses composants à l'initiation de la dynamique de la cellule.

En étudiant les six réseaux de taille génomique choisis comme exemple, cette situation se présente plus souvent dans les réseaux *iJO1360* (6 fois) et *A. ferrooxidans* (7 fois) que dans les réseaux *iJR904* et *T. lutea* (1 cas chacun) et *Synechocystis* (2 cas) alors qu'aucun carrefour ME-persistance, ME-efficacité et non ME-productibilité n'est présent dans le réseau *iAF1260*. Le nombre de métabolites présentant ces caractéristiques montre la capacité des réseaux métaboliques à adapter eux-mêmes leurs flux internes en activant ou non des cycles internes équilibrés.

## 5.4 Identification de co-produits critiques

La dernière propriété des réseaux métaboliques que nous avons identifiée comme liée aux propriétés d'essentialité est relative au rôle des co-produits dans les réactions du système.

### 5.4.1 Rôle des métabolites exclusivement ME-productibilité.

De tels métabolites ont la capacité de réduire à zéro le ratio de croissance de la biomasse lorsque les réactions dont ils sont les réactants sont enlevées du réseau métabolique. De plus, aucun des flux supprimés n'est nécessaire pour produire de façon optimale la biomasse dans le réseau initial et la suppression n'enlève pas l'accessibilité topologique des métabolites cibles. Un exemple dans le réseau *iAF1260* est la molécule de *carbon dioxide* cytosolique. En enlevant ce métabolite du réseau, avec l'ensemble des réactions utilisant ce dernier comme substrat, nous n'affectons pas l'accessibilité topologique des métabolites cibles. Cependant le taux de croissance est réduit de 9021 à zéro mmol/gDW/h car les composants de la biomasse ne peuvent pas être produits à cause de l'hypothèse des équilibres de masse au sein du réseau qui n'est alors plus assurée.

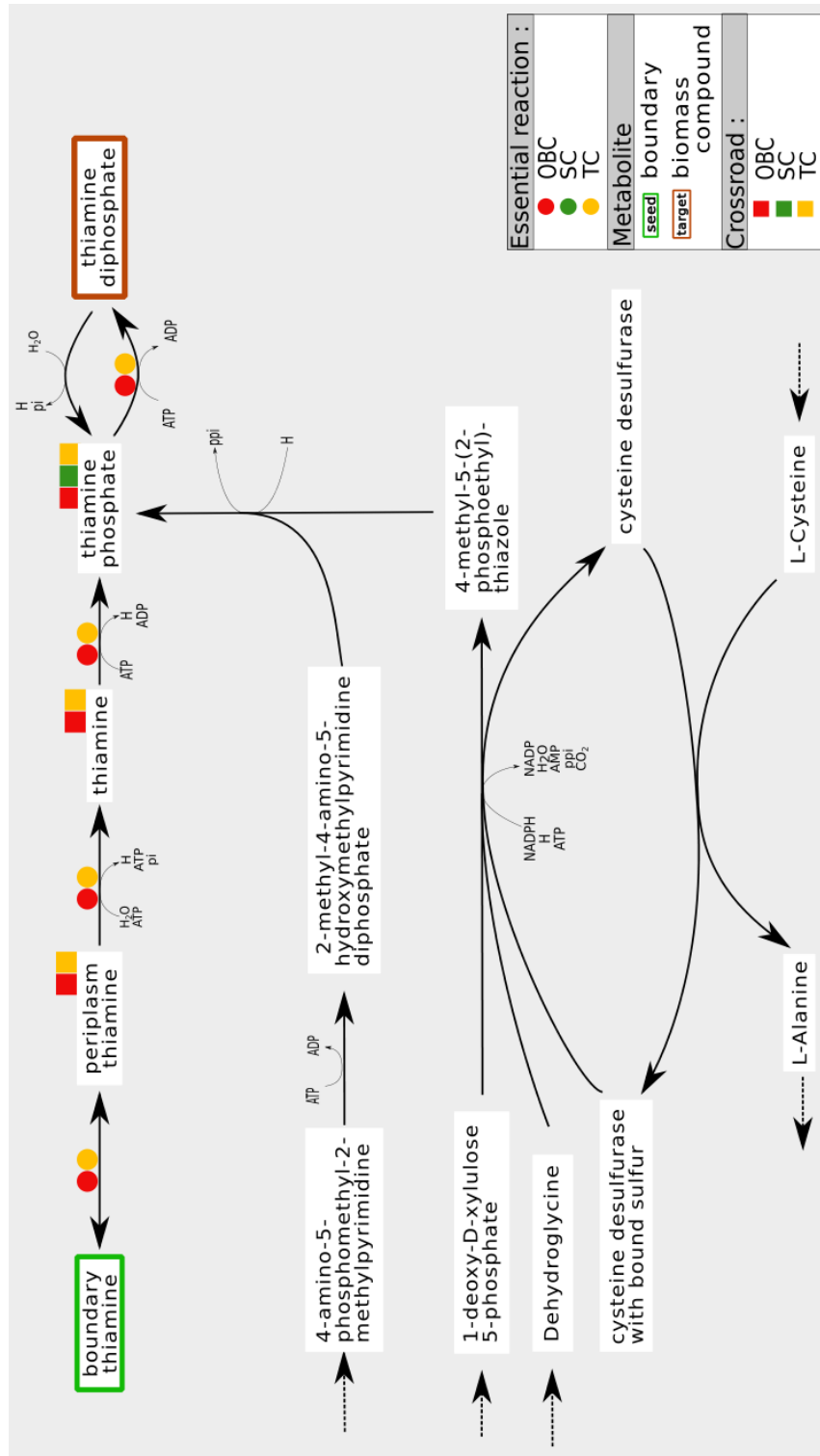


FIGURE 5.6 – Des cycles internes permettant une production non-optimale de la biomasse : exemple d'un métabolite ME-persistence et ME-efficacité qui n'est pas un métabolite ME-productibilité. Dans le réseau *iJO1360*, la délétion des réactions consommant *Thiamine* pour produire *Thiamine phosphate* permet au système d'activer un cycle interne. Ce dernier implique des composants métaboliques qui ne sont pas accessibles à partir des nutriments du milieu extracellulaire selon des critères topologiques liés au graphe.

Comme illustré dans la Figure 5.7, un tel composant correspond principalement à des co-produits de réactions essentielles qui sont donc nécessairement produits pour assurer la production de la biomasse. Dans le périplasma du réseau représenté, le composé *murein5px4p* (un métabolite cible) est le produit d'une réaction dont le substrat est *murein5p5p*. Cette dernière réaction est nécessaire pour l'optimalité des flux dans le réseau et le métabolite substrat est un métabolite ME-persistance, ME-productibilité et ME-efficacité. Un autre produit de cette réaction est la *D-alanine* qui est un carrefour ME-productibilité uniquement, signifiant qu'il est nécessaire stœchiométriquement mais non topologiquement ni pour l'optimalité des flux.

Nous pouvons voir ainsi que, d'une part, alors que le *murein5p5p* peut être utilisé pour produire un autre métabolite cible, la *L-alanine*, une autre voie permet également la production de ce composant et, d'autre part, il n'est pas substrat d'une réaction essentielle pour produire de façon optimale la biomasse. Cependant, lorsque les réactions d'export de ce co-produit dans les autres compartiments sont supprimées du réseau, le composé *D-alanine* s'accumule et la loi de conservation de masse n'est plus assurée. Ainsi, la production de biomasse, dans le sous-réseau correspondant à ce métabolite, est impossible : l'analyse des flux indique que le ratio de production de *murein5px4p* est nul. La biomasse est donc productible sous la condition que le co-produit *D-alanine* dans le périplasma soit consommé par une réaction métabolique ou une réaction d'export.

En d'autres termes, les métabolites ME-productibilité qui ne sont ni ME-persistance ni ME-efficacité se réfèrent à des co-produits de réactions essentielles dans une étude stœchiométrique. Ils peuvent avoir un impact important sur le taux de croissance de l'organisme puisque leur ratio de consommation est un contrôleur direct de la production de biomasse dans le cas d'un état stationnaire. Dans la pratique, une telle situation apparaît peu dans les réseaux métaboliques étudiés sauf pour celui de *Synechocystis* : une seule fois dans les réseaux *A. ferrooxidans* et *T. lutea*, deux, trois et quatre fois respectivement pour les réseaux *iJR904*, *iAF1260* et *iJO1360*. Cette observation suggère que les co-produits doivent être méticuleusement étudiés dans l'ensemble de ces réseaux. Au contraire, le réseau *Synechocystis* contient 17 métabolites ME-productibilité qui ne sont ni ME-persistance ni ME-efficacité. Cela laisse à penser que la production de la biomasse est sous le contrôle de nombreux flux d'export de co-produits. Une analyse de sensibilité et de curation de ce réseau basé sur ces critères pourrait être utile.

#### 5.4.2 Rôle des métabolites ME-productibilité et ME-efficacité qui ne sont pas ME-persistance

Un tel métabolite est lié aux propriétés stœchiométriques du réseau métabolique mais n'intervient pas d'un point de vue topologique. En effet, un composant répondant aux caractéristiques d'un métabolite ME-productibilité et ME-efficacité est non seulement nécessaire à l'accessibilité stœchiométrique de métabolites cibles mais est également substrat d'une réaction essentielle pour maximiser les flux dans le réseau. Nous pouvons avancer deux explications pour ce cas particulier de carrefour. En effet, d'un premier abord, nous pouvons nous reposer sur l'exemple précédent car si le co-produit d'une réaction essentielle stœchiométriquement n'est consommé que par une unique réaction métabolique ou d'export, cette dernière devient alors aussi essentielle et le co-produit est non seulement ME-productibilité mais aussi ME-efficacité. D'un second abord, le fait que le métabolite ne soit pas ME-persistance nous permet d'avancer une autre hypothèse pour laquelle il existe une voie métabolique secondaire topologiquement accessible mais non utilisable dans une étude de flux à cause de diverses raisons telle que une non conservation de masse pour un nœud métabolite.

Dans la Figure 5.4, nous pouvons voir un exemple de la première explication avancée avec les métabolites *CO2* et *Ammonia* qui sont tous deux des métabolites ME-productibilité et ME-efficacité mais pas ME-persistance. En effet, de tels carrefours sont des co-produits de réactions essentielles entraînant la production de métabolites cibles. Enlever les réactions qui consomment ces métabolites entraîne une accumulation de ces derniers dans la cellule et annule toute production de biomasse (la valeur de flux est égale à zéro), bien que l'accessibilité topologique ne soit pas affectée.

Cette dernière situation est très fréquente dans les réseaux *Synechocystis*, *A. ferrooxidans* et *T. lutea*. En effet, la suppression de toutes les réactions consommant un métabolite avec de telles

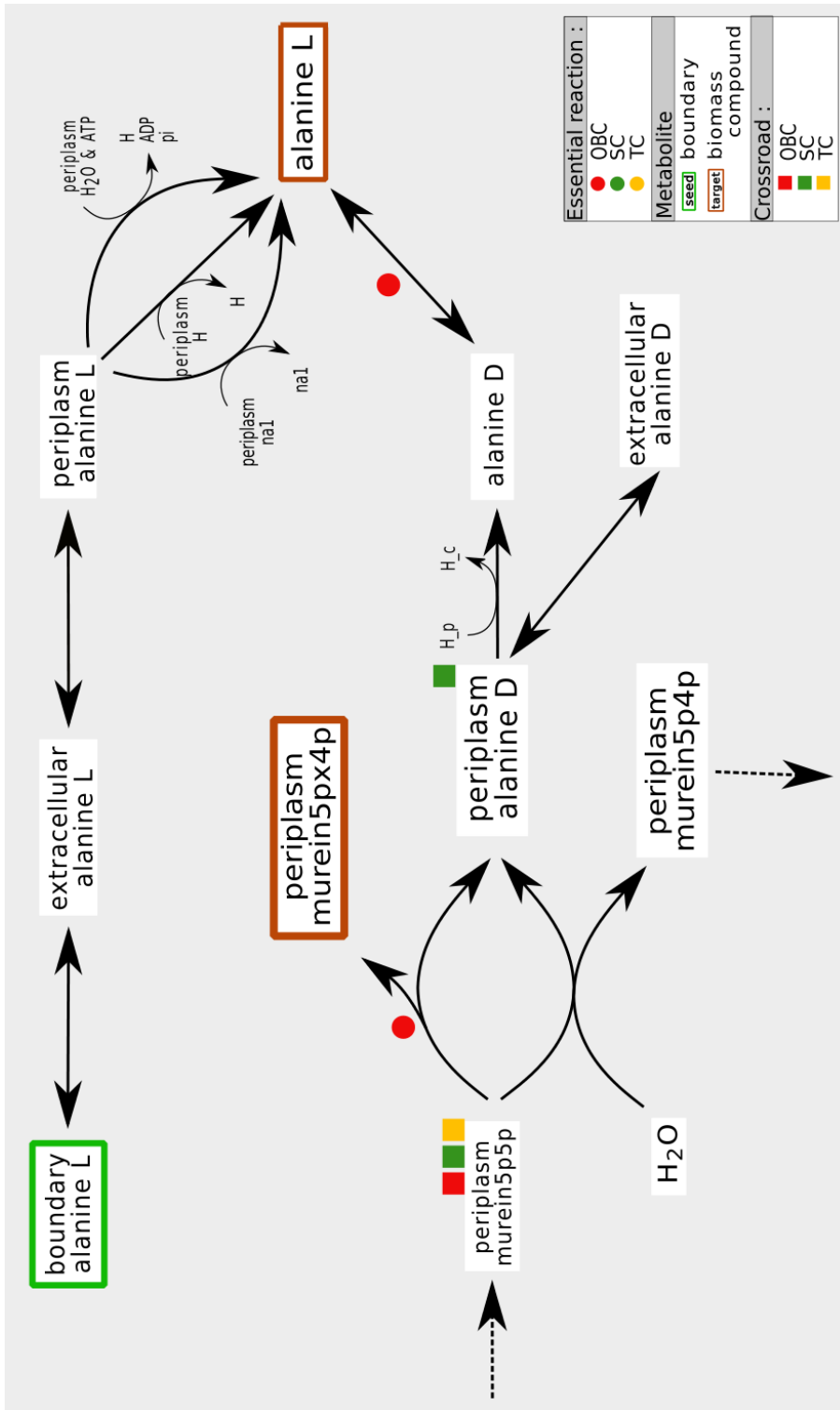


FIGURE 5.7 – Un réglage fin de l'équilibre des masses : exemple d'un métabolite ME-productibilité qui n'est ni un métabolite ME-persistence ni un métabolite ME-efficacité. Dans le réseau *iAF1260*, la production de *murein5px4p* périplasmique (un composant de la biomasse) est sous le contrôle de l'export d'un de ses co-produits *D-alanine* vers le cytoplasme ou le compartiment externe. D'un point de vue graphe, stœchiométrique et optimale pour les flux, le composé *murein5p5p* est requis pour la production de *murein5px4p*. Au contraire, la présence de *D-alanine* a seulement un impact sur la production optimale de la biomasse selon les flux de matière. En effet, le fait d'enlever la réaction d'export de la *D-alanine* implique son accumulation, un mécanisme non autorisé dans un état stationnaire.

propriétés ne modifie pas l'accessibilité topologique des métabolites cibles mais, au contraire, réduit drastiquement le taux de croissance de la biomasse car enlève des réactions essentielles.

## Conclusion

Dans ce chapitre, nous avons étudié les métabolites essentiels (ME-persistance, ME-productibilité et ME-efficacité) de six réseaux métaboliques d'échelle génomiques. L'étude de ces métabolites montrent qu'il existe un nombre variable de métabolites essentiels (pour au moins une des trois propriétés) dans un réseau métabolique (de 7% à 70% des métabolites sur nos exemples), les réseaux les plus aboutis en terme de curation manuelle et de descriptions de phénotypes étant ceux qui contiennent le moins de métabolites essentiels. Ceci suggère que le pourcentage de métabolites essentiels, quel que soit leur type, est un indicateur à la fois de la variabilité de réponses phénotypiques décrites par le réseau et de son degré de curation. De plus, au sein des métabolites essentiels, plus de 40% des métabolites partagent les trois propriétés d'être à la fois ME-persistance, ME-productibilité et ME-efficacité. Notre interprétation est que ces composés peuvent être vus comme les constituants principaux pour la production de biomasse, sans doute comme un squelette de la structure du réseau.

De manière plus fine, nous avons montré que les métabolites essentiels qui ne vérifient pas l'ensemble des trois propriétés sont très utiles pour comprendre le fonctionnement fin du système et éventuellement suggérer des analyses plus poussées en termes de curation. Ainsi, les métabolites ME-persistance et ME-productibilité qui ne sont pas ME-efficacité sont des indicateurs de voies redondantes. Les métabolites qui sont exclusivement ME-efficacité pointent aussi du doigt des redondances de voies qui n'ont pas la même efficacité. Par ailleurs, les métabolites qui sont exclusivement ME-persistance suggèrent l'existence de cycles auto-alimentés dans l'analyse des flux du réseau et permettent de proposer des candidats pour l'initialisation de ces derniers lorsque la cellule ou l'organisme n'est pas dans un état stationnaire. De même, les métabolites ME-persistance et ME-efficacité qui ne sont pas ME-productibilité pointent du doigt des dépendances fortes entre l'activation d'un cycle et l'état initial de la cellule. Enfin, les métabolites qui sont exclusivement ME-productibilité semblent souligner le rôle de co-produits qui peuvent avoir un impact important sur le taux de croissance de l'organisme puisque leur ratio de consommation est un contrôleur direct de la production de biomasse. Un rôle similaire apparaît pour les métabolites ME-productibilité et ME-efficacité qui ne sont pas ME-persistance.

Dans l'ensemble, cette étude suggère que la recherche de carrefours (ou métabolites essentiels phénotypiques) peut avoir de nombreuses applications dans l'étude du métabolisme d'organisme. D'une part pour la curation fine de réseau, mais aussi pour suggérer des voies métaboliques permettant une efficacité importante de la production en utilisant des composants clés.

## Chapitre 6

# Étude des classes de métabolites ME-persistance

Dans le chapitre 4, nous avons introduit trois définitions pour des métabolites phénotypiques essentiels (carrefours) en s'appuyant sur trois sémantiques distinctes pour modéliser la production de molécules dans un réseau métabolique. Dans le chapitre 5, nous avons montré que la comparaison des propriétés de ces métabolites carrefours permettait d'identifier dans un réseau des composés et des sous-réseaux ayant un rôle dynamique subtil : redondance de voies métaboliques, cycles dont l'activation est dépendante de l'état initial du système, rôle des co-produits dans la production de biomasse.

Cependant, rappelons que dans le chapitre 4, nous avons indiqué que la définition de métabolite essentiel à l'aide de critères topologiques permettait de s'abstraire de la composition quantitative de la biomasse et d'établir des relations directes entre un composé et l'ensemble des composants de la biomasse sur la production desquels il a un impact. En particulier, à l'aide de cette propriété, il est devenu possible d'identifier des classes de métabolites ME-persistance relatives aux ensembles de composés qu'ils affectent.

Le but de ce chapitre est d'explorer plus précisément la signification biologique et l'intérêt de cette classification des métabolites ME-persistance. Dans une première partie, nous étudierons les classes de métabolites obtenus pour les six réseaux à échelle génomique que nous avons étudiés. Dans une seconde partie, nous montrerons que ces classes de métabolites permettent de discriminer les protocoles de reconstruction de réseaux métaboliques.

### 6.1 Propriétés des classes de six réseaux à échelle génomique

Dans un premier temps, nous avons appliqué notre approche de calcul des classes de métabolites ME-persistance implémentée en programmation par ensembles réponses (logiciel *Spaceman*, voir sous-section 4.4.3) aux six réseaux métaboliques qui sont étudiés dans ce manuscrit.

#### 6.1.1 Validation experte des clusters dans le réseau *T. lutea*

Une première analyse que nous avons réalisée a été d'étudier précisément les classes du réseau *T. lutea*. En effet, comme indiqué dans la section 1.4, ce réseau avait été extensivement et précisément étudié dans [8] pour sortir de l'hypothèse d'état stationnaire impliquant qu'aucune accumulation de matière ne peut être faite. Pour cela, le réseau métabolique avait été décomposé en sous-réseaux selon les fonctions métaboliques prenant en compte la compartimentation cellulaire et les voies métaboliques. Concrètement, ces composés ont été répartis en six classes : le cycle de Calvin, la glycolyse haute et la glycolyse basse, la synthèse de carbohydrates, la synthèse de lipides et la synthèse de biomasses. Ces classes sont validées ensuite par des expérimentations biologiques et la décomposition en modes élémentaires (voir [8] pour plus de détails).

Nous avons ainsi pu comparer le résultat de la décomposition des métabolites ME-persistance à l'aide de *Spaceman* avec la décomposition manuelle réalisée dans [8] pour montrer des réponses



phénotypiques au phénomène jour-nuit. Notre analyse suggère que 196 métabolites parmi les 210 de composés du réseau (soit 93%) sont des métabolites ME-persistance. Ils se répartissent en 39 clusters, dont un certain nombre de singletons (12) et 9 clusters prépondérants qui contiennent de 13 à 29 métabolites (3 clusters ayant respectivement 5, 6 et 7 carrefours, 3 clusters à 8 composants et enfin 3 autres clusters ayant respectivement 10, 31 et 52 métabolites).

De manière intéressante, les clusters 30 et 3 sont pratiquement similaires aux classes nommées "haute glycolyse" et "synthèse des lipides". De plus, la plupart des clusters de petite taille obtenus par *Spaceman* correspondent à la classe la plus grande appelée "synthèse de la biomasse". Une telle division de la classe peut s'expliquer par l'ensemble des métabolites utilisés dans [8] pour réduire la complexité du réseau métabolique de telle sorte que le nombre de paramètres cinétiques est aussi réduit. En effet, un tel ensemble augmente le degré de liberté du réseau (i.e. illustré par 24 modes élémentaires au lieu de un ou deux pour les autres classes) ce qui est infaisable pour les clusters (i.e., la suppression d'un métabolite implique la non-production d'une cible).

Cette analyse qualitative (réalisée "à l'aveugle") est donc apparue plutôt encourageante en terme d'utilisation de *Spaceman* pour faciliter la décomposition de réseau. En particulier, les clusters de grande taille semblent avoir un rôle fonctionnel tandis que les clusters de petite taille pourraient être regroupés dans un unique cluster fonctionnel.

### 6.1.2 Distribution des métabolites ME-persistance dans les clusters

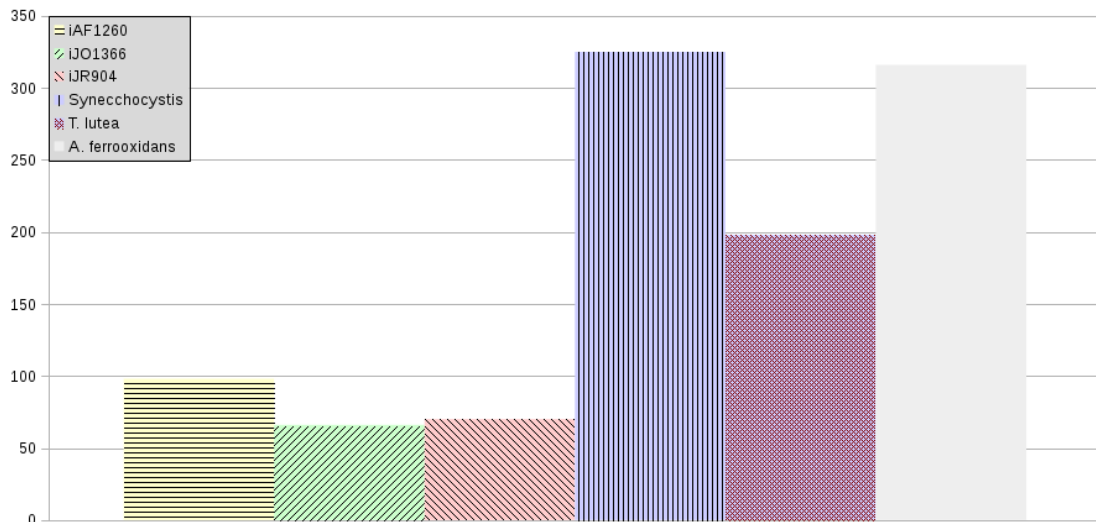
De manière exhaustive, la Figure 6.1 décrit le nombre de clusters dans les six réseaux métaboliques étudiés, *iAF1260*, *iJR904*, *iJO1360*, *Synechocystis*, *A. ferrooxidans* et *T. lutea*. Comme indiqué au chapitre 4, le nombre de métabolites ME-persistance varie entre 70 et 325, ce qui représente 2% à 65% des métabolites des réseaux. Le nombre de clusters se situe entre 13 et 77, ce qui représente de 1% à 13% des métabolites du réseau. Le ratio entre le nombre de clusters et celui de métabolites ME-persistance varie entre 18% et 33% confirmant que les métabolites ME-persistance sont loin d'être indépendants et qu'ils doivent être considérés ensemble selon leur relation via la topologie du réseau.

La Figure 6.2 représente la distribution des métabolites ME-persistance au sein des différents clusters. Cette étude confirme le phénomène identifié pour *T. lutea*. Plus précisément, pour chaque réseau métabolique étudié, la plupart des clusters contiennent seulement un ou deux métabolites ME-persistance, impliquant que le nombre médian des métabolites ME-persistance par cluster est toujours égal à 2 pour l'ensemble des six réseaux, et le troisième quartile (75%) est plus ou moins égal à 5. De plus, chaque réseau métabolique possède un ou un petit nombre de clusters irréguliers contenant plus de cinq métabolites ME-persistance. Dans l'ensemble, l'analyse suggère que la distribution des métabolites ME-persistance au sein des clusters est très homogène.

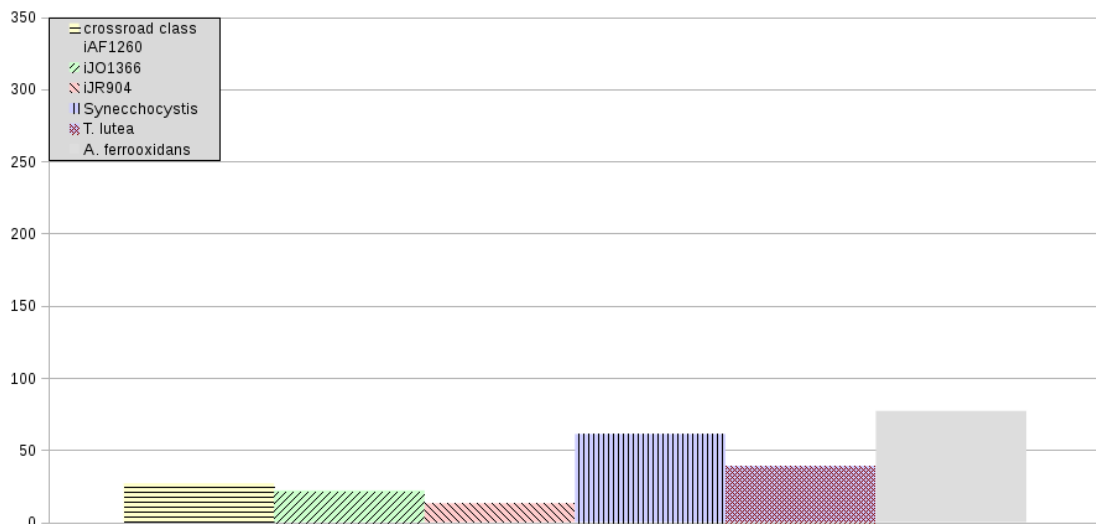
### 6.1.3 Chaque étude de cas montre un petit nombre d'ensembles dominants duals de cibles

Par définition, un métabolite ME-persistance supprime des blocs de production d'au moins un composant de la biomasse (e.g., un *métabolite cible*). De façon similaire, les clusters de métabolites ME-persistance regroupe ceux qui sont à la fois connectés par une chaîne de réactions métaboliques et qui altèrent la production du même ensemble de cibles duales (voir Définition 21). Cet ensemble de cibles, que nous avons formellement défini par  $\mathcal{T}(c) = \{t \in rcts(r_T), t \notin \Sigma(prune(G, c), S)\}$ , où  $c$  est l'un des métabolites ME-persistance du cluster, est un attribut de chaque cluster que nous appelons *ensemble cibles duales*.

La Figure 6.3 représente la distribution du nombre d'éléments dans chacun de ces ensembles de cibles duales. Pour tous les réseaux métaboliques étudiés, la médiane est inférieure ou égale à 2 et le troisième quartile (75%) est très bas ce qui suggère que la plupart des clusters (et métabolites ME-persistance associés) représentent de petits ensembles de composants métaboliques mais ayant un impact critique sur la production d'une seule cible. À l'inverse, les clusters de grande taille présents dans la Figure 6.3 montrent que, dans chaque réseau métabolique, il existe entre un et cinq clusters de métabolites ME-persistance liés à un nombre relativement large de métabolites cibles. Ces clusters regroupent des métabolites ME-persistance intervenant dans le contrôle global de la production de la biomasse.



(a) Nombre de métabolites ME-persistance



(b) Nombre de classe

FIGURE 6.1 – **Vue d'ensemble des métabolites ME-persistance et de leurs clusters appliquée à six réseaux métaboliques distincts.** Le nombre de métabolites ME-persistance (a) est très variable pour les six réseaux tandis que leur distribution dans les clusters (b) est homogène.

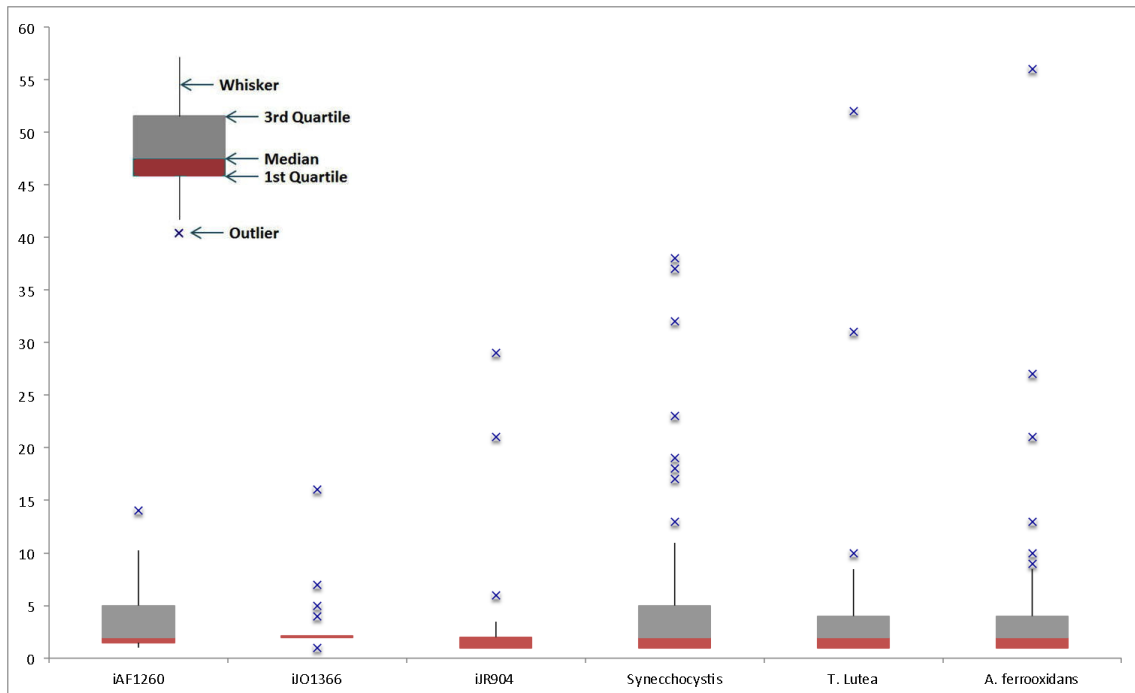


FIGURE 6.2 – **Distribution des métabolites ME-persistence dans les clusters de six réseaux métaboliques** La distribution du nombre de métabolites ME-persistence appartenant à chaque cluster indique que plusieurs clusters ont un unique métabolite ME-persistence mais que très peu de clusters regroupent un large nombre de métabolites ME-persistence.

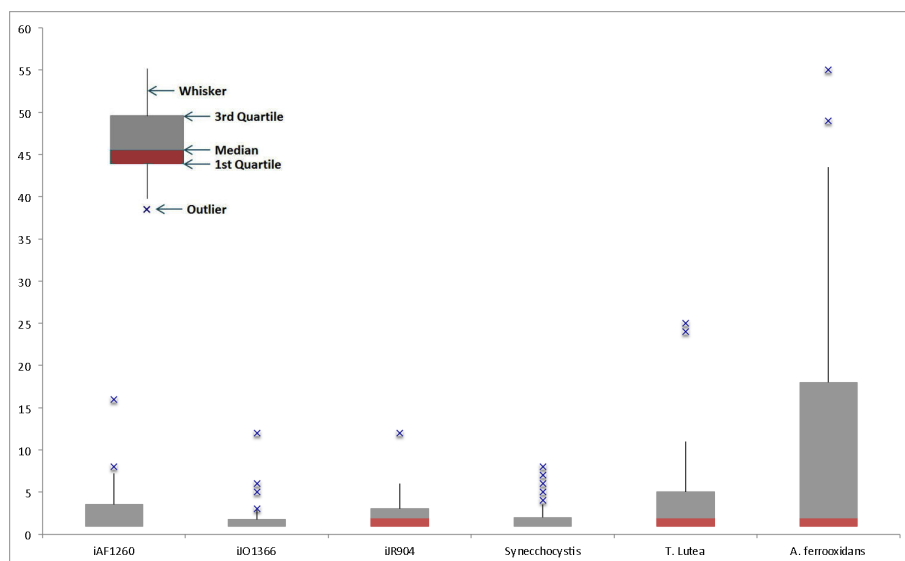


FIGURE 6.3 – **Représentation sous la forme de boîte à moustache de la distribution du nombre de cibles duales associées à chaque cluster.** Pour tous les cas d'études, l'ensemble des cibles duales pour chaque cluster a été calculé. La Figure représente la distribution des cardinalités, en incluant les cas extrêmes qui sont des points ayant une valeur plus importante que celle du troisième quartile plus 1.5 fois la différence entre le premier et le troisième quartile.

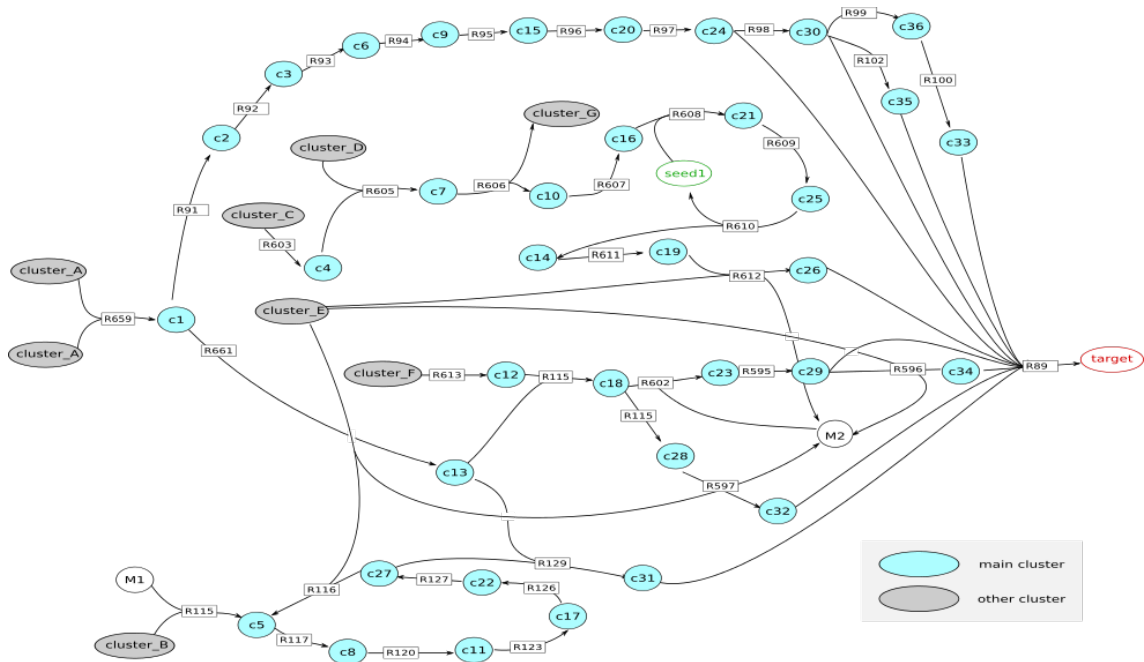


FIGURE 6.4 – Un cluster de 36 métabolites ME-persistance (en bleu) dans le réseau *Synechocystis* contrôlant chacun la production de la même métabolite cible

Nous avons plus précisément analysé le cluster associés à chacun des plus grands ensembles de cibles duales dans les réseaux *iAF1260*, *iJO1360*, *iJR904* et *A. ferrooxidans*. De manière intéressante, il s'avère que dans ces quatre cas, le cluster qui a le plus grand ensemble de cibles duales contient un unique métabolite ME-persistance, à savoir, l'hydrogène. Ceci montre le rôle majeur de ce composé interne dans la production de biomasse chez de nombreuses espèces microbiales.

Dans le réseau *Synechocystis*, l'étude des clusters associés aux ensembles de cibles duales les plus importants met en avant aussi le rôle dominant des métabolites Phosphate et GTP, ce qui est cohérent avec le rôle clé de ces derniers dans les métabolismes autotrophiques. Notre analyse retourne aussi deux clusters de 10 et 23 métabolites ME-persistance qui contrôlent la production des mêmes ensembles de six et sept composants de la biomasse. Ainsi, pour l'exemple représenté dans la Figure 6.4, les 36 métabolites ME-persistance travaillent ensemble pour faciliter la production de la biomasse. Ils sont absolument nécessaires pour produire un réactant de la biomasse, ce qui montre des dépendances internes au sein du réseau.

L'analyse du réseau de *T. lutea* met l'accent sur un ensemble de 24 métabolites, incluant les sommets *Amino Acids*, *Cholorophyl A* et *L-Aspartate* qui appartiennent à l'ensemble de cibles duales de deux clusters : le premier rassemble les métabolites *Nitrite* et *Ammonium*; le second cluster contient le métabolite *Oxalacetic acid*. Un autre métabolite ME-persistance mis en lumière par l'analyse est  $HCO_3$ , dont l'ensemble de cibles duales rassemble 25 métabolites comme précédemment avec le *Phosphatidate*. Dans l'ensemble, cette analyse dessine les connexions entre les caractéristiques fonctionnelles du métabolisme principal.

#### 6.1.4 Correspondance entre cluster et ensemble de cibles duales

Pour mieux comprendre la relation entre les clusters et leur ensemble dual de cible, la Figure 6.5 décrit les relations entre le nombre de métabolites ME-persistance dans un cluster et le nombre de métabolites dans son ensemble de cibles duales. Chaque cercle représente un cluster dont l'abscisse décrit le nombre de ME-persistance qu'il contient et l'ordonnée le nombre de cibles duales. La surface du cercle représente le nombre de clusters qui ont les mêmes abscisses/ordonnées.

Un premier enseignement, donné par l'existence d'un cercle de large surface proche du point (0,0) dans chaque figure, est que la majorité des clusters rassemble un petit nombre de méta-

bolites ME-persistence et sont associés à un petit nombre de métabolites cibles. Ils représentent des associations locales entre les métabolites internes et les cibles exportées, associations qui sont essentielles pour étudier les structures présentes dans le réseau métabolique.

Il est utile de noter que les clusters ayant les plus larges classes de métabolites ME-persistence (leur abscisse est maximale) sont toujours associés à une cible unique (ordonnée égale à 1). En particulier, le plus grand cluster chez *A. ferrooxidans* (56 carrefours) représente un sous-réseau contrôlant la production de la cible *Phospholipids*. Dans le réseau *T. lutea*, le cluster ayant la taille maximale (52 carrefours) est lié à la production de *Phosphatidate*. Pour *Synechocystis*, le cluster ayant le plus de métabolites ME-persistence (38 métabolites) contrôle la production du composé *Lipid* de la biomasse tandis qu'un autre grand cluster (37 carrefours) est lié au métabolite *Pigment*. Le plus grand cluster de *iJR904* représente un sous-réseau nécessaire à la production du métabolite *Peptidoglycan subunit* de *E. coli*. Le réseau *iJO1360* met en avant la combinaison de composés qui contrôlent la production de *Tetrahydropholate* tandis que le réseau de *iAF1260* contient un sous-réseau lié à la production de *Murein units*.

Ces analyses suggèrent ainsi que les réseaux métaboliques pourraient contenir des sous-ensembles de réactions spécifiques qui méritent une étude attentive, en lien avec la production topologique des composés d'une biomasse. Il s'agit d'abord des plus grands clusters de métabolites ME-persistence d'un réseau, qui a priori (au moins dans nos exemples) seraient en petit nombre. Il s'agit ensuite des clusters de métabolites ME-persistence dont l'ensemble de cibles duales a une taille optimale dans le réseau. Là encore, dans nos six exemples étudiés, ces ensembles sont peu nombreux et se distinguent par leur cohérence biologique.

## 6.2 Etude des clusters d'un ensemble de 3600 réseaux métaboliques

Nous avons réalisé une étude à plus grande échelle pour essayer de comprendre si la distribution des clusters et des classes de cibles duales observée dans la Figure 6.5 est spécifique aux six réseaux métaboliques étudiés ou si elle est générique.

### 6.2.1 Ensemble-test de réseaux métaboliques

Pour cela, nous avons exploité un ensemble de réseaux métaboliques tests construits par Prigent et al dans [83]. Il s'agit d'un ensemble-test basé sur 90 fonctions de biomasse différentes pour le réseau *iJR904*. Les fonctions de biomasse ont été calculées à l'aide du logiciel Gurobi (i) d'une part pour que le réseau *iJR904* soit capable de produire chacune des biomasses mais (ii) d'autre part pour qu'il existe 40 réseaux dégradés de *iJR904* qui ne sont plus capables de produire aucune de ces 90 biomasses.

Ainsi, nous avons à notre disposition 90 réseaux métaboliques fonctionnels, ainsi que 3600 réseaux (90 biomasses  $\times$  40 réseaux dégradés) dégradés non fonctionnels. Il faut noter que les 40 réseaux dégradés ont été obtenus de manière à varier les types de dégradation : certains réseaux ont été dégradés en enlevant de nombreuses réactions essentielles, d'autres en se focalisant sur des réactions alternatives. De plus, le taux de dégradation a été choisi pour varier entre 10% et 40%.

Comme nous l'avons présenté dans la section 1.2, en 2010, Thiele et Palsson ont décrit un protocole général permettant de reconstruire un réseau métabolique de taille génomique comme vu précédemment. Dans un premier temps, la reconstruction automatique de réseau, appelée graphe de réseau métabolique, est obtenue à partir des annotations du génome. On obtient ainsi des réseaux de structure semblable à celle des 3600 réseaux dégradés que nous avons étudiés dans notre ensemble test. Dans un second temps, les bases de données de référence des réactions métaboliques sont utilisées pour vérifier si l'ajout de réactions au réseau permet à des métabolites d'intérêt d'être produits dans un milieu de croissance donné. L'identification de ces réactions manquantes constitue la tâche de complétion (ou gapfilling) du réseau métabolite.

Plusieurs approches existent pour résoudre ces problèmes de complétion. Les réactions peuvent être choisies pour optimiser un score linéaire modélisant la quantité produite de biomasse dans un

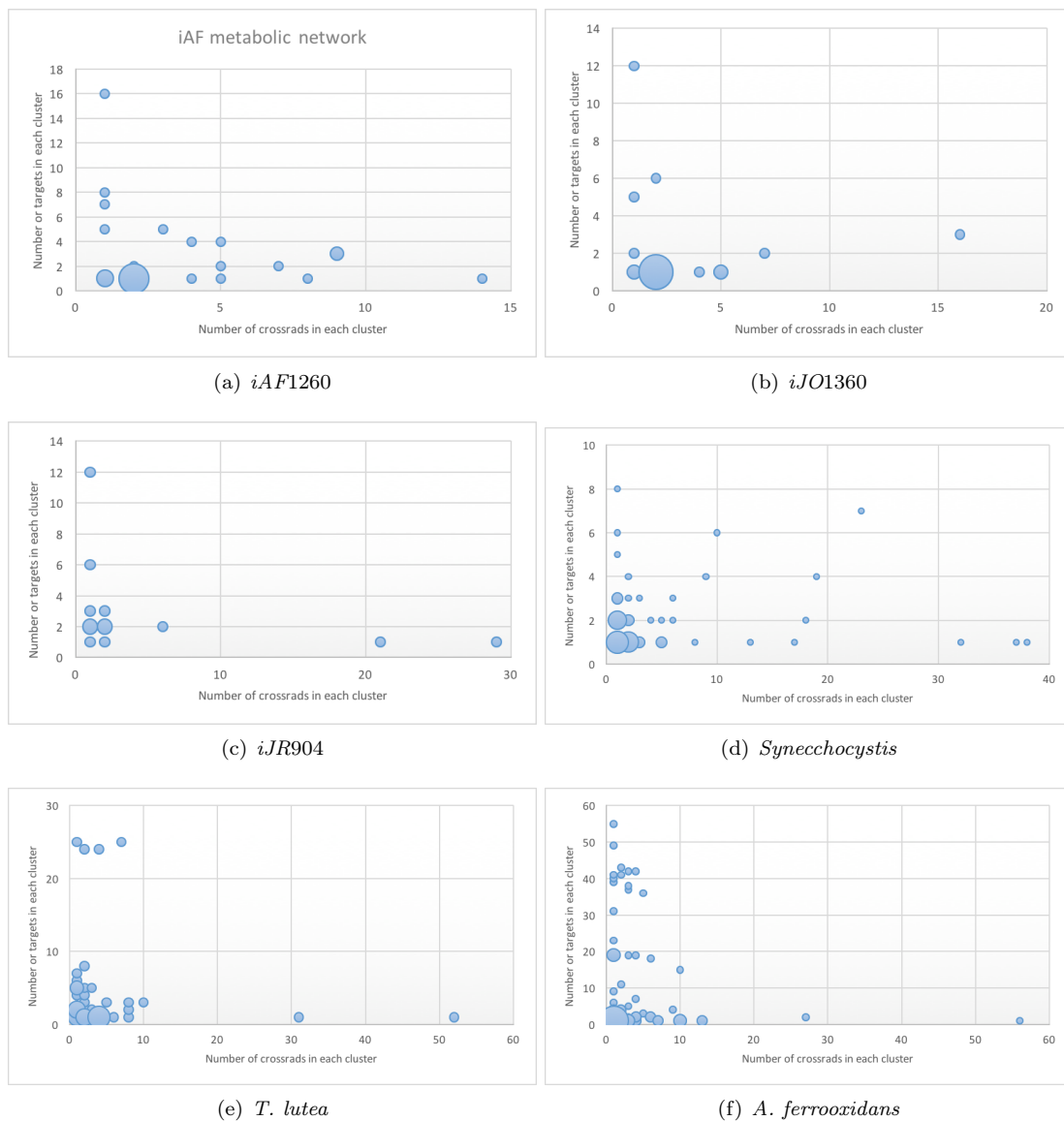


FIGURE 6.5 – Correspondance entre le nombre de métabolites ME-persistence dans un cluster et le nombre de composants dans les ensembles des cibles duales associés La surface du cercle correspond au nombre de paires (cluster, ensemble cibles duales) ayant la même cardinalité. Par exemple, dans le réseau *Synechocystis*, il y a un cluster avec 39 métabolites ME-persistence associé à un ensemble singleton de cible duale, tandis que deux clusters avec un unique métabolite ME-persistence sont associés chacun avec trois cibles duales.

système, comme utilisé dans le logiciel *Gapfill* [63] (voir Annexe A.1.2). Il s'agit ainsi d'un dérivé de l'approche FBA introduite à la section 2.2. Son dérivé, l'approche *FastGapFill* [96] (voir Annexe A.1.3), a pour but de restaurer la production des tous les composés non internes du système et d'optimiser la production de biomasse. Un critère basé sur les graphes, s'appuyant sur le concept de scope (voir la section 2.5.4) peut aussi être utilisé, par exemple dans l'outil *Meneco* [89, 83]. Dans [83], les auteurs ont réalisé une complétion des 3600 réseaux dégradés avec les méthodes *Gapfill*, *Meneco* et *FastGapFill*. Leurs analyses mettent donc à disposition 11800 réseaux reconstruits pour différentes biomasses et différents taux de dégradation.

Nous allons nous appuyer sur ces différents réseaux pour étudier finement la robustesse du concept de métabolite ME-persistance et de cluster associé.

## 6.2.2 Classification des fonctions de biomasse en fonction de leurs clusters

La Figure 6.6 décrit les nombres correspondant aux métabolites ME-persistance et aux clusters du réseau *iJR904* associé à chacune des 90 biomasses. On note d'une part une grande variabilité du nombre de métabolites ME-persistance (min=7, moyenne=63, médiane=71, max=81), vis-à-vis du fait que le réseau *iJR904* initial contenait 70 métabolites ME-persistance.

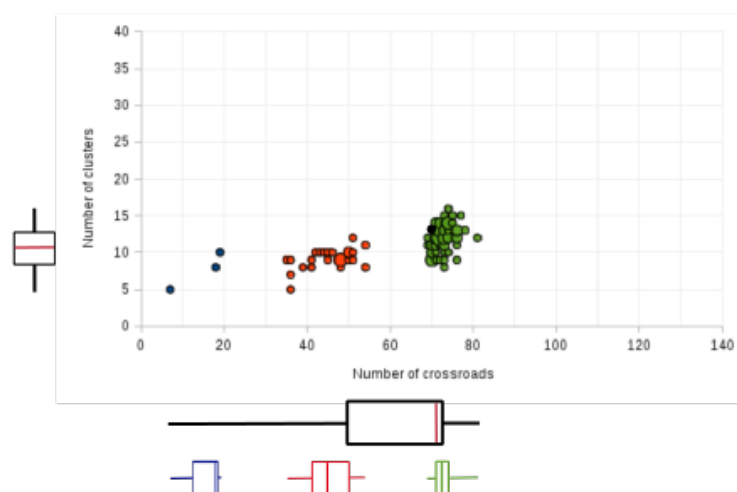


FIGURE 6.6 – Nombre de métabolites ME-persistance et clusters dans le réseau *iJR904* associé à 90 biomasses différentes. 90 réactions de biomasse ont été dérivées à partir des réactions de biomasse de *iJR904*. Le nombre de carrefours et clusters dans le réseau est calculé. Les nœuds bleus correspondent aux biomasses ayant un fort impact sur la classification, les nœuds rouges correspondant aux biomasses ayant un impact moyen, les nœuds verts correspondent aux biomasses à faible impact.

Nous remarquons en particulier que les fonctions de biomasse peuvent être réparties en trois groupes selon leur distribution métabolites ME-persistance/cluster : pour 62 fonctions de biomasse (points verts), les nombres de carrefours et clusters sont proches des caractéristiques du réseau initial *iJR904*. Nous les appelons *fonctions de biomasse avec un impact mineur*. 25 fonctions de biomasse (points rouges) entraînent une diminution de moitié du nombre de métabolites ME-persistance mais gardent un nombre de clusters similaires. Nous les nommons *fonctions de biomasse à impact intermédiaire*.

Finalement, pour trois fonctions de biomasse (points bleus), le réseau montre peu de métabolites ME-persistance (entre cinq et dix) bien que le nombre de clusters ne décroisse pas. Nous les appelons *fonctions de biomasse à fort impact*. La raison du petit nombre de clusters dans ce dernier groupe est que les composants de la réaction de biomasse ont été sélectionnés de telle sorte que, avec une analyse de la variabilité des flux (Flux Variability Analysis, *FVA*), il n'y a que peu de réactions

essentielles à la production de la biomasse, c'est-à-dire, peu de réactions toujours utilisées pour produire cette dernière. Ainsi, ces réseaux sont très redondants pour la production de la biomasse et ont un faible nombre de carrefours.

Contrairement aux métabolites ME-persistence, le nombre de clusters est relativement robuste (min=5, moyenne=11, médiane=11, max=16), bien que légèrement plus faible que le nombre de clusters dans le réseau *iJR904* original (13). En moyenne, chaque réseau contient cinq clusters de métabolites ME-persistence isolés (min=3, max=8), ce qui correspond à une large partie des clusters pour les fonctions de biomasse à fort impact (entre 22% et 42% des clusters sont à carrefour unique) et une faible part des réactions de biomasse à faible et moyen impact (entre 8% et 17% des clusters sont à carrefour unique). En plus de ces clusters à métabolites ME-persistence isolé, l'analyse met en avant des clusters additionnels qui montrent des voies métaboliques individuelles associées à une unique cible duale, réactant de la biomasse. Ainsi, la composition de la réaction de biomasse a un faible impact sur le nombre de clusters et presque aucun impact sur le nombre de clusters à métabolites ME-persistence isolés bien que cela entraîne une diminution nette du nombre de métabolites ME-persistence.

### 6.2.3 Dépendance de la clusterisation à la topologie du réseau

Pour simuler l'impact de la qualité d'un réseau métabolique sur la distribution des carrefours, le réseau *iJR904* fut dégradé en supprimant entre 10% et 40% des réactions dans l'optique de créer 40 réseaux, chacun ne pouvant produire aucune des 90 biomasses décrites précédemment. Chacun de ces réseaux dégradés produit seulement un sous-ensemble des métabolites composant les 90 fonctions de biomasse. Pour chaque paire consistant en un réseau dégradé associé à une fonction de biomasse (3 600 réseaux métaboliques), nous avons calculé et représenté les nombres de métabolites ME-persistence et de clusters dans la Figure 6.7.

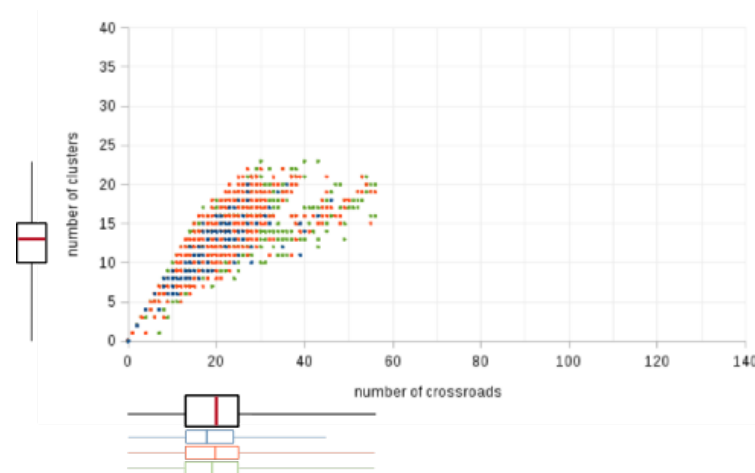


FIGURE 6.7 – Nombre de métabolites ME-persistence et clusters dans 3600 réseaux métaboliques construits à partir de 40 réseaux métaboliques *iJR904* dégradés, chacun associé à 90 biomasses différentes. De 10% à 40% des réactions du réseau de *iJR904* sont retirées de ce dernier pour créer 40 réseaux dégradés. Les nombres de métabolites ME-persistence et de clusters sont représentés pour l'ensemble des réseaux dégradés associés à toutes les 90 réactions de biomasse.

On note que la dégradation du réseau métabolique a un fort impact sur le nombre de métabolites ME-persistence, dont la valeur moyenne diminue de 63 dans les réseaux non dégradés (Fig. 6.6) à 21. Une explication est la procédure de dégradation qui bloque l'accessibilité de nombreux chemins métaboliques, et supprime des voies entre les métabolites et leurs cibles associées. Par exemple, 84 réseaux sur les 3600 étudiés n'ont pas de métabolites ME-persistence. Pour un très petit nombre de cibles possiblement productibles, la redondance dans le réseau ne crée aucun métabolites ME-persistence.



Inversement, le nombre de clusters est moins affectés par la procédure de dégradation : la valeur moyenne est de 13 par rapport aux 10 clusters du réseau non dégradé. Dans les réseaux dégradés, les clusters contiennent en moyenne 1,53 métabolites ME-persistence contre 5,75 dans le réseau non-dégradé, montrant que les métabolites ME-persistence dans les réseaux dégradés sont presque tous isolés et que l'opération de clusterisation a peu d'effet, au contraire, du réseau non-dégradé.

La dégradation du réseau métabolique entraîne plusieurs conséquences contradictoires pour le nombre de métabolites ME-persistence. La suppression de certaines réactions diminue naturellement le nombre de métabolites cibles accessibles et donc de métabolites ME-persistence. Cette situation se retrouve dans la plupart des réseaux associés à des fonctions de biomasse à faible impact (points verts) : un relativement grand nombre de métabolites ME-persistence isolés perdent leur fonction durant le processus de dégradation puisque les voies métaboliques entre les carrefours et leur(s) cible(s) associée(s) sont supprimées.

D'un autre côté et de manière contradictoire, la suppression de réactions peut aussi diminuer la redondance des voies métaboliques permettant la production des mêmes cibles. Lorsque l'accessibilité à travers une voie redondante est modifiée, sa voie alternative devient une voie métabolique obligatoire et l'ensemble des métabolites intervenant au sein de celle-ci deviennent des métabolites ME-persistence et sont regroupés dans un même cluster. Ce phénomène tend à favoriser une augmentation du nombre de métabolites ME-persistence. On l'observe dans le cas des réseaux associés à une fonction de biomasse à fort impact (points bleus). Ce phénomène confirme que les composés impliqués dans les réactions de biomasse correspondantes sont principalement produits par des voies métaboliques hautement redondantes dans le réseau *iJR904*. La procédure de dégradation supprime quelques redondances et exhibe des métabolites qui deviennent nécessaires à la production de la biomasse.

L'impact de la procédure de dégradation sur les clusters est moins importante. Une baisse de diversité des voies métaboliques favorise une fusion des clusters et donc diminue leur quantité (points verts). De manière contradictoire, si on supprime des réactions correspondant aux connexions entre métabolites ME-persistence, la relation de transitivité n'est plus assurée, entraînant une division de ces mêmes clusters. Cependant, nos expériences montrent que ce phénomène se produit peu fréquemment.

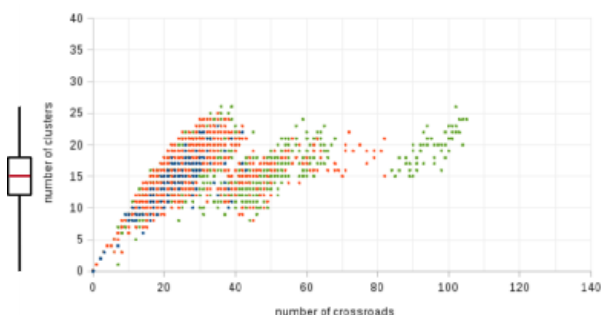
Ainsi, ces expérimentations suggèrent que des réseaux non fonctionnels, dégradés conservent un nombre de clusters relativement proche du nombre de clusters du réseau initial, tandis que le nombre de métabolites ME-persistence est fortement affecté par la dégradation.

#### 6.2.4 Impact de procédures de reconstruction fonctionnelle sur la clusterisation des carrefours

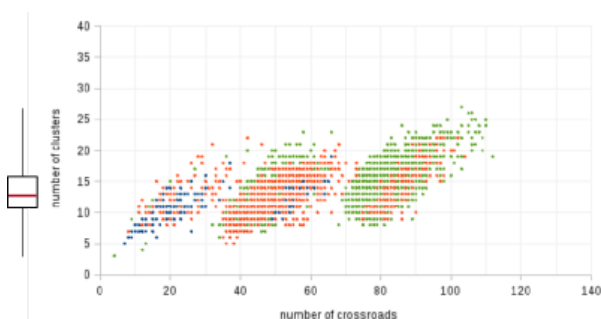
Pour mieux comprendre l'impact de la dégradation puis de la reconstruction des réseaux métaboliques, nous avons utilisé notre approche pour calculer les métabolites ME-persistence et leurs clusters sur ces 10 800 réseaux métaboliques complétés avec les méthodes *fastgapfill*, *gapfill* et *meneco*. Le nombre de métabolites ME-persistence et de clusters sont représentés dans la Figure 6.8.

Les résultats montrent que les réseaux complétés avec la méthode *Gapfill* ne retrouvent aucune des caractéristiques des réseaux initiaux. La valeur moyenne du nombre de métabolites ME-persistence est de 30, plus proche de la valeur moyenne des réseaux dégradés (i.e. 20) que de celle des réseaux initiaux (i.e. 63). Cela peut être expliqué par le fait que, comme montré dans [83], *Gapfill* ne parvient pas à identifier beaucoup de réactions obligatoires du réseau *iJR904* lorsque se déroule la reconstruction. À l'inverse, notre analyse met en avant que *Meneco* surestime le nombre de métabolites ME-persistence (valeur moyenne 83) et le nombre de clusters (valeur moyenne 20). Finalement, *FastGapFill* montre de très bonnes performances dans la restauration des nombres moyens des métabolites ME-persistence (qui est de 64) et de clusters (de 13) par rapport aux caractéristiques générales des réseaux initiaux.

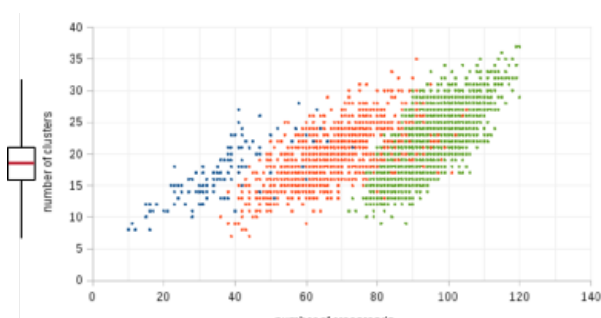
Si nous nous intéressons aux trois groupes des réactions de biomasse identifiés dans la Figure 6.6, il apparaît cependant que les réseaux complétés par *FastGapFill* possèdent un nombre de métabolites ME-persistence beaucoup plus variable pour les fonctions de biomasse ayant un impact



(a) Gapfill



(b) FastGapfill



(c) Meneco

FIGURE 6.8 – Métabolites ME-persistence et clusters des 10 800 réseaux obtenus après complétion de 3600 réseaux dégradés. Chacun des 3 600 réseaux métaboliques dégradés de *iJR904* est complété avec trois méthodes de complétion dans l'optique de restaurer la production d'une biomasse correspondante. Les nombres de métabolites ME-persistence et de clusters obtenus pour chacune des méthodes sont représentés dans les sous-figures correspondantes. Les différentes couleurs correspondant à la classification des réactions de biomasse donnée dans la Figure 6.6.

moyen (points rouges) et faible (points verts). Cela suggère que *FastFapFill* ne parvient à reconstituer que quelques caractéristiques des nœuds internes selon la réaction de biomasse : bien que l’algorithme de complétion implémenté dans *FastGapFill* tend à optimiser la production de biomasse, il autorise aussi l’export de l’ensemble des composés frontières du réseau, possiblement en ajoutant de nouvelles réactions d’import et d’export. Nos résultats suggèrent que la combinaison de ces deux objectifs introduit un biais dépendant du rôle joué par les métabolites internes dans le réseau.

Inversement, *Meneco* retrouve ces caractéristiques en regroupant ensemble les points bleus, rouges et verts. Notre principale hypothèse est que *Meneco* surestime le nombre moyen de métabolites ME-persistance et de clusters en introduisant des métabolites ME-persistance dans les réseaux complétés : le nombre moyen de clusters à métabolites ME-persistance isolé ajoutés par rapport au réseau non-dégradé est de 7 pour *Meneco* tandis que, pour *FastGapFill*, il est égal à 1. Ce biais peut être dû à la méthode de résolution de *Meneco* qui fait l’union de la totalité des ensembles minimaux de complétion du réseau métabolique, et en plus crée des redondances locales, éventuellement artificielles, dans le réseau complété.

## Conclusion

Les analyses présentées dans ce chapitre portent principalement sur la notion de cluster de métabolites ME-persistance introduite dans le chapitre 4. Dans une première étude qualitative de six réseaux métaboliques de taille génomique, nous avons étudié la nature et la distribution du nombre de composants des clusters. Notre étude suggère qu’un certain nombre de clusters correspondent à des métabolites isolés, mais qu’un petit nombre de clusters sont particulièrement intéressants vis-à-vis de la production de la biomasse. Ces clusters de grande taille ont été en particulier validés comme étant biologiquement importants dans le réseau *Synechocystis*. Ainsi, notre étude suggère que les réseaux métaboliques pourraient contenir des sous-ensembles de réactions spécifiques qui méritent une étude attentive, en lien avec la production topologique des composés d’une biomasse. Il s’agit d’abord des (petits) clusters de métabolites ME-persistance dont l’ensemble de cibles duales est relativement important, qui décrivent des composés intervenant dans un grand nombre de voies métaboliques et sont liées aux principales fonctions biologiques du réseau. Il s’agit ensuite des plus grands clusters de métabolites ME-persistance d’un réseau, associé à un faible nombre de cibles duales, qui pointent les composés internes nécessaires à la production de cibles individuelles.

Dans un second temps, nous avons réalisé une étude à grande échelle relative à la robustesse des métabolites ME-persistance et de leurs clusters. Cette étude visait à confirmer l’intérêt des clusters suggéré par l’étude des six réseaux exemples. Pour cela, nous avons analysés les métabolites ME-persistance d’un ensemble de test de 3600 réseaux dégradés (non fonctionnels). Notre étude a confirmé que le concept de métabolite ME-persistance est très sensible à la réaction de biomasse et à la topologie du réseau. En particulier, nous en concluons que le rôle fonctionnel joué par les métabolites ME-persistance dans un réseau métabolique de faible qualité (non curé manuellement) peut drastiquement changer par rapport à un réseau réel. L’impact de cette sensibilité peut, en particulier, être important sur l’étude des réseaux reconstruits manuellement.

De manière intéressante, notre analyse suggère que l’utilisation des clusters permet de limiter fortement cette incertitude. En particulier, les réseaux reconstruits avec les méthodes *meneco* (approche topologique) et *FastGapfill* (approche FBA) parviennent à retrouver une partie des caractéristiques initiales du système. *Meneco* est la meilleure approche pour retrouver des caractéristiques liées à la réaction de biomasse mais il surestime le nombre de métabolites ME-persistance en introduisant des éléments artefactuels. *FastGapFill* est excellent en terme de nombre de métabolites ME-persistance, mais il ne retrouve pas spécifiquement les caractéristiques des biomasses. Enfin, de manière cruciale, le nombre de clusters semble relativement stable entre les réseaux initiaux, dégradés et complétés. Ceci suggère que la clusterisation des métabolites ME-persistance est une opération robuste à la perte d’information dans des réseaux métaboliques.

# Conclusion et perspectives

## Conclusion

L'ensemble des travaux réalisés dans cette thèse appartiennent au domaine de la bio-informatique, c'est-à-dire, de l'informatique appliquée à des problèmes biologiques afin d'aider à la compréhension des systèmes biologiques via des méthodes automatiques. Nos travaux s'intéressent plus précisément à l'étude des réseaux métaboliques des systèmes biologiques composés des réactions métaboliques pouvant se produire au sein de ces derniers. Cela est possible grâce à la modélisation mathématique des réseaux métaboliques notamment avec des graphes bipartis orientés pondérés composés d'un ensemble de nœuds métabolites et d'un autre de nœuds réactions dont les arcs représentent les liens entre la réaction et ses composants, substrats et produits, et sont pondérés par la stœchiométrie de la réaction.

L'étude d'un réseau métabolique peut être réalisée selon deux formalismes dont le but est souvent l'évaluation de la production de métabolites modélisant une croissance de biomasse. La majorité des travaux étudie un réseau métabolique en tant que graphe de flux, développant ainsi la méthode de *FBA* et ses dérivées (*FVA*), qui utilise la programmation linéaire par contraintes et ses solveurs afin de déterminer une quantité de matière pouvant être produite. Une autre méthode d'étude de réseaux métaboliques relativement récente, se focalise sur la topologie de ces derniers en utilisant l'expansion de réseau et le scope défini comme l'ensemble des métabolites productibles à partir d'un ensemble source de métabolites. Ce procédé a été implémenté via une technique de programmation par ensemble-réponse nommée ASP (*Answer Set Programming*). Les deux formalismes permettent non seulement de modéliser un réseau métabolique, d'étudier ce dernier mais aussi de le compléter si nécessaire, dans le but de pouvoir prédire une production d'éléments du réseau métabolique.

La littérature sur l'étude des réseaux métaboliques a mis en évidence notamment des métabolites considérés comme essentiels pour produire une biomasse. Cette notion d'essentialité peut différer selon les travaux réalisés mais est toujours liée à la productibilité stœchiométrique, en étude de flux. Or, si cette dernière permet de définir une production effective de composants métaboliques, une étude topologique met en avant une persistance du réseau dans cette production. La première partie de cette thèse a consisté à définir trois notions d'essentialité pour les métabolites selon le formalisme utilisé que nous généralisons sous le terme de *carrefour*. Le premier cas définit le concept de métabolite essentiel du point de vue de la productibilité du réseau, lié à une production stœchiométrique de la biomasse. Le second concept, les métabolites essentiels du point de vue de l'efficacité du réseau, correspond aux composants essentiels calculés pour la production stœchiométrique optimale de la biomasse. Le dernier concept, que nous nommons métabolite essentiel du point de vue de la persistance du réseau, repose sur la persistance du réseau et l'accessibilité topologique des composants de la biomasse. Ces trois définitions de métabolites essentiels pour la production de la biomasse ne s'excluent pas mutuellement. De plus, le fait que les ME-persistance prennent en compte les composants de la biomasse séparément nous a permis de définir une clusterisation de ces derniers en fonction notamment des métabolites cibles pour lesquels ils sont nécessaires.

Les résultats, que nous avons obtenus dans la seconde partie de cette thèse, nous ont permis de mettre en évidence le rôle des métabolites essentiels dans les réseaux métaboliques, selon les

caractéristiques qu'ils présentent. Nous avons utilisé pour cela six réseaux métaboliques de taille génomique pris dans la littérature : trois réseaux de *E. coli* nommés *iJO1360*, *iAF1260* et *iJR904*, un réseau de *Synechocystis*, un réseau de *A. ferrooxidans* et un réseau *T. lutea*. De plus, nous avons regardé l'impact du gapfilling sur la recherche de carrefours mais surtout sur la clusterisation des métabolites essentiels du point de vue de la persistance du réseau avec trois méthodes de complétion basées soit sur les flux dans le réseau (Gapfill et Fastgapfill), soit sur l'expansion du réseau (Meneco), en dégradant un réseau métabolique puis le complétant avec ces trois méthodes.

Nous avons pu ainsi voir que les composants répondant soit seulement à la définition des ME-persistance et celle des ME-productibilité, soit uniquement à la définition des ME-efficacité, permettent de déterminer la présence de redondance de voies métaboliques dans le réseau. De plus, l'impact de l'initiation des cycles métaboliques peut être mis en avant par les métabolites présentant soit seulement les propriétés d'un ME-persistance, soit uniquement celles d'un ME-persistance et d'un ME-efficacité. Enfin, nous pouvons, en étudiant les éléments soit exclusivement ME-productibilité, soit seulement ME-productibilité et ME-efficacité, mettre en évidence le rôle de métabolites en tant que co-produits de réactions. L'ensemble de ces premiers résultats, présentés dans les chapitres 4 et 5, font l'objet d'une publication en révision mineure pour la revue PeerJ. Mais les carrefours peuvent être la cause d'une incertitude montrée en comparant les réseaux obtenus par gapfilling, incertitude limitée par la clusterisation des ME-persistance, dont la répartition reste stable malgré la perte d'informations. La clusterisation nous a ainsi permis de regrouper non seulement les métabolites ayant le même impact sur la production des composants de la biomasse mais également, par ces derniers, de regrouper des réactions importantes liés à cette dernière. Elle a aussi mis en avant une nouvelle analyse des réseaux obtenus par gapfilling par rapport au réseau initial étudié.

L'ensemble de ces résultats ont pu être obtenus grâce à deux packages python que nous avons développés permettant à partir d'un fichier modélisant un réseau métabolique sous le format sbml de calculer d'une part les métabolites essentiels et d'autre part les clusters des métabolites essentiels du point de vue de la persistance du réseau. Le premier package *Conquests* utilise dans un premier temps la programmation linéaire par contraintes utilisée en *FVA* et *FBA* via le package *cobrapy* afin de calculer les métabolites essentiels du point de vue de la productibilité du réseau et les ME-efficacité, puis la programmation par ensemble-réponse avec *ASP* pour chercher les métabolites essentiels du point de vue de la persistance du réseau du réseau métaboliques donné en entrée. Le second package python nommé *Spaceman* recherche les clusters du réseau métabolique liés aux ME-persistance avec la technologie ASP également.

## Perspectives

Ce travail a défini plusieurs notions pour l'étude d'un réseau métabolique au niveau de l'essentialité des métabolites dans la production d'une biomasse définie. Plusieurs perspectives à plus ou moins long terme ressortent alors.

### Application de la recherche des différents types de carrefour à d'autres types de réseaux

Dans le travail que nous présentons ici, nous nous sommes intéressés aux réseaux métaboliques d'organismes isolés. Or, de nos jours, l'intérêt de la communauté biologique se porte non seulement sur les organismes mais aussi sur les communautés d'organismes et notamment les interactions entre les différentes espèces présentes dans un même milieu. En effet, il a été montré, notamment pour des bactéries [12], qu'un organisme a besoin, pour croître dans des conditions particulières, de la présence d'autres espèces lui apportant des métabolites nécessaires. Il serait donc intéressant de rechercher les trois types de carrefours dans un réseau correspondant à un écosystème afin de voir les différences avec un organisme individuel.

De plus, nous avons montré dans l'introduction qu'il existait plusieurs types de réseaux autres que métaboliques (réseau de signalisation et de régulations 1). Ces réseaux peuvent également

être représentés avec la modélisation sous forme de graphe biparti, nous permettant d'appliquer notre méthode. Cette application pourrait montrer de nouveaux résultats selon le type de réseau et apporter de nouvelles connaissances.

Enfin, nous avons étudié deux formalismes de réseau métabolique : le *Flux Balance Analysis* et ses dérivées (FVA) ainsi que l'expansion topologique. Mais de nouveaux formalismes se sont développés notamment le *regulatory FBA* (rFBA)[16] combinant le FBA, la logique booléenne et des équations différentielles. La recherche des carrefours dans les réseaux modélisés grâce à ces nouvelles méthodes mettrait en avant de nouvelles données.

## Une nouvelle méthode d'analyse d'un réseau métabolique

Nous avons conclu que la comparaison systémique de plusieurs approches de modélisation (topologique, stœchiométrique et basé sur les flux optimaux) du même réseau peut mettre en avant des composants ayant différents rôles selon les différents formalismes. Notre hypothèse est que ces métabolites souvent sont porteurs d'informations pertinentes pouvant être utilisées pour comprendre la dynamique du système. Cela encourage l'utilisation du package *Conquests* dans deux tâches complémentaires. Soit les informations sur la dynamique du système sont biologiquement non pertinentes car non vérifiées *in vivo* et ces nouvelles informations peuvent être utilisées pour finaliser la curation du réseau métabolique. Soit les informations sur la dynamique du système sont biologiquement pertinentes et permettent de mettre en avant le rôle de cycle interne et l'équilibre des masses dans le but de produire une réaction de biomasse ciblée devant être déterminée. De façon intéressante, le nombre de métabolites essentiels appartenant à une ou deux des trois classes que nous décrivons est particulièrement petit dans les modèles de taille génomique qui contiennent habituellement plus d'une centaine de métabolites. Cela rend plus facile la curation et l'analyse du modèle car ces composants peuvent être facilement étudiés manuellement.

## Clusterisation des métabolites essentiels stœchiométriques

Dans ce travail, nous avons introduit, dans le chapitre 4, le concept de classe de métabolites essentiels pour la production d'une biomasse définie dépendant des composants cibles de la biomasse accessible. Cependant cette notion est calculée sur les métabolite essentiel du point du vue de la persistance du réseau car la topologie permet de prendre en compte chacun des métabolites de la biomasses indépendamment contrairement à l'analyse de flux qui ne distinguent pas les différents composants de la biomasse. Il serait donc intéressant de définir la notion de classe pour les métabolites essentiels stœchiométriques malgré cette non distinction des éléments.

## Compression du réseau métabolique en fonction des clusters de carrefours

Une autre perspective au niveau de la clusterisation des métabolites essentiels du point du vue de la persistance du réseau est liée à la compression du réseau. Comme nous l'avons montré dans le chapitre 5, les métabolites essentiels permettent d'avoir un aperçu des principales interactions entre les voies métaboliques associées à la production de la biomasse dans le réseau métabolique. Notre processus de clusterisation permet de regrouper les métabolites essentiels du point du vue de la persistance du réseau selon la topologie du réseau et les métabolites cibles qu'ils permettent de produire. L'étape suivante sera d'utiliser ces clusters pour compresser le réseau en un squelette robuste du fonctionnement du modèle. La difficulté est cependant de définir une relation propre entre métabolites non-carrefours et les clusters du réseau dans le but de les compresser. En faisant l'analogie avec les réseaux métaboliques compartimentés, les clusters pourraient jouer le rôle des composés qui peuvent être transportés entre les compartiments abstraits composés de métabolites non-carrefours.

## Une nouvelle méthode de gapfilling avec le package *Menecowi*

La dernière perspective concerne la reconstruction de réseaux métaboliques. En effet, comme nous l'avons mis en avant dans le chapitre 6, la méthode de *gapfilling* utilisée dans la reconstruction d'un réseau métabolique a un impact sur la structure finale de ce dernier. Les informations obtenues

par le calcul des métabolites essentiels peuvent être ainsi utilisées de deux façons possibles dans une reconstruction. D'une part, si la connaissance biologique le permet, l'utilisation d'une liste de métabolites devant avoir les caractéristiques de métabolites essentiels au sein du réseau reconstruit permet d'orienter la complétion de celui-ci. D'autre part, si aucune liste de métabolites n'est donnée, il serait intéressant de compléter le réseau métabolique étudié en prenant en compte une optimisation pouvant porter sur le nombre de carrefours ou de clusters, la composition des clusters en carrefours ou des cibles duales. Nous avons commencé ce travail, notamment sur l'utilisation d'une liste de métabolites essentiels, via le design d'un package python nommé *Menecowi* utilisant la notion de métabolite essentiel du point de vue de la persistance du réseau avec le langage *ASP* que nous avons présenté dans la section 2.5. Il reste à valider cette approche sur des benchmarks pour en montrer la valeur ajoutée.

## Annexe A

# Annexe : Différentes approches de programmation par contraintes, appliquées à l'exemple du gapfilling

Comme nous l'avons vu dans le chapitre 1, la construction d'un réseau métabolique n'est pas chose aisée car les données obtenues expérimentalement ne permettent pas de créer un réseau complet. Certaines parties du réseau restent non connues, entraînant le développement de différents moyens pour identifier les réactions manquantes dans un réseau. On parle de "complétion de réseaux métaboliques". La modélisation mathématique d'un réseau métabolique permet d'automatiser ces complétions en se basant essentiellement sur la théorie des graphes. Le réseau métabolique étant sous forme d'un graphe enrichi par des flux, le but de ces complétions est l'ajout si possible au réseau d'arcs représentant les réactions prises dans une base de données de réactions métaboliques.

La complétion peut avoir plusieurs buts bien que deux principaux objectifs sont essentiellement utilisés dans la littérature. Le premier but est de rendre possible la production d'une biomasse donnée à partir d'un milieu donc de nutriments posés. Des réactions métaboliques sont ajoutées afin d'obtenir un flux positif au niveau de la réaction de biomasse en créant un chemin à partir des nœuds métabolites représentant les nutriments jusqu'à cette réaction objectif. Un deuxième objectif de la complétion peut simplement être la reconstruction de voies métaboliques. Ainsi si une partie d'une voie existe dans le réseau métabolique avec un pourcentage supérieur à un taux donné, cette dernière est complétée à partir de bases de données de voies métaboliques en ajoutant les réactions omises dans la première construction.

L'utilisation de la théorie des graphes afin de créer le réseau métabolique permet l'utilisation des techniques de recherche opérationnelle telle que la programmation linéaire ou la programmation par contraintes afin de compléter le réseau métabolique. Dans cette annexe, nous détaillons trois approches qui ont été utilisées pour créer les jeux de tests que nous exploitons au chapitre 6.

### A.1 Gapfilling avec programmation linéaire par contraintes

Une des méthodes pour compléter un réseau métabolique est la programmation linéaire par contraintes via l'utilisation des flux dans ce dernier que nous avons présentés précédemment et qui est utilisée dans de nombreux logiciels de reconstruction et analyse de réseaux métaboliques.

#### A.1.1 Principe de la programmation linéaire par contraintes appliqué à la reconstruction métabolique

De la même façon que pour le problème de *Flux Balance Analysis* expliqué dans le chapitre 2, la reconstruction d'un réseau métabolique utilisant la programmation linéaire par contraintes est modélisée sous forme d'un programme linéaire composé de variables, contraintes et d'une fonction d'optimisation et créé à partir du réseau métabolique étudié et d'une base de données de référence



$BdD$  de réactions métaboliques. Chaque réaction du réseau modélisé et de la base de référence est représentée par une variable modélisant le flux de la réaction et les contraintes du problème de  $FBA$  sont retrouvées dans ce problème de reconstruction telles que la contrainte d'état stationnaire et d'équilibre des masses.

$$\begin{aligned} \forall r \in R \cup BdD, \quad v_r &\in \mathbb{R} \\ \forall j \in \{1..n_R\}, \quad v_r = v_j, LB_j &\leq v_r \leq UB_j \\ \forall i \in \{1..n_M\} \quad \sum_{j \in J} \mathcal{M}_{G_{ij}} \cdot v_j &= 0 \end{aligned}$$

Une autre variable binaire est ajoutée pour chaque réaction de la base de données afin de symboliser l'ajout ou non de cette dernière dans le réseau métabolique pour compléter celui-ci.

$$\forall j \in BdD, y_j = \begin{cases} 1 & \text{si la réaction } j \text{ est ajoutée au réseau} \\ 0 & \text{sinon} \end{cases}$$

De plus, une contrainte est ajoutée pour chacune de ces dernières afin que toute réaction ajoutée ait son flux non nul tandis que les réactions de la base de données non sélectionnées ont leur flux égal à zéro.

$$\forall j \in BdD, LB_j \cdot y_j \leq v_j \leq UB_j \cdot y_j$$

Ces éléments constituent la base de la reconstruction d'un réseau métabolique par complétion utilisant la programmation linéaire par contraintes. D'autres contraintes peuvent être ajoutées selon la méthode de complétion choisie et la fonction objectif du problème linéaire varie également. Nous présentons par la suite deux méthodes de complétion proches utilisant la programmation linéaire par contraintes.

### A.1.2 Gapfill : une complétion du réseau métabolique pour permettre la production de biomasse

GapFill est un outil de complétion de réseau métabolique en lien avec GapFind, un outil de recherche d'incomplétudes dans un réseau métabolique. Tous deux développés en 2007 par Kumar et ses collaborateurs [63], ils se basent sur de la programmation linéaire pour identifier des trous et rechercher des solutions à partir de bases de données pour combler ces derniers sous hypothèse d'un état stationnaire. Le réseau métabolique est considéré sous forme de graphe enrichi par des flux avec la possibilité de réactions réversibles signifiant que certains arc-réactions peuvent être à double sens.

GapFind est l'approche permettant de retrouver les trous présents dans le réseau métabolique en se basant sur l'accessibilité des métabolites dans ce dernier ([Définition 6](#)). En effet, l'outil recherche les métabolites bloquants dans le réseau, c'est-à-dire des métabolites n'ayant aucune réaction active le produisant ou le consommant au sein du réseau métabolique. Ces métabolites sont importants à étudier car ils peuvent être la raison d'un blocage de réactions nécessaires à la production de la biomasse. Par exemple, comme nous l'avons vu au chapitre 2 si une réaction nécessite un métabolite qui n'est pas produit, elle est considérée comme non activable car un de ses substrats est manquant, et si une réaction produit un métabolite non consommé, elle est considérée comme non activable car un de ses produits est accumulé, violant ainsi la loi d'équilibre des masses et l'hypothèse d'état stationnaire.

Des réactions peuvent ainsi être bloquées à cause de métabolites mais aussi si le flux passant dans ces dernières n'est pas assez important. Les auteurs utilisent un ensemble de contraintes afin de déterminer si un métabolite est bloquant pour des réactions du réseau :

- un métabolite  $j$  est considéré comme produit s'il existe une réaction ayant un flux non nul (strictement positif pour les réactions irréversibles) entraînant la production d'une quantité non nulle de ce métabolite  $j$
- le flux de chaque réaction doit être compris dans les limites imposées à la réaction par le modèle

- un minimum de matière devant toujours être produit en état stationnaire de par la division des cellules, les auteurs fixent la contrainte suivante  $\sum_{j \in M} S_{ij} v_j \geq 0, \forall i = 1..n_M$
- la fonction objectif correspond à une maximisation des métabolites productibles

GapFind permet de par ces contraintes de savoir si des trous existent dans le réseau métabolique étudié et donc s'il est nécessaire d'utiliser le second outil GapFill.

GapFill permet donc de reconstruire un réseau métaboliques dans lequel les auteurs ont mis en place plusieurs moyens de reconstruction :

- inversion du sens d'une réaction irréversible
- ajout de réactions provenant d'une base de données de réactions métaboliques
- ajout de réactions de transport pour importer des métabolites non produits ou exporter des métabolites non consommés

Le but de cette complétion des manques du réseau métabolique est de rendre l'ensemble des métabolites produits et consommés, permettant ainsi une analyse des flux permettant la production d'une biomasse. Ce problème est modélisé grâce à de la programmation linéaire en nombres entiers dont l'objectif est de minimiser le nombre de réactions ajoutées au réseau pour le compléter sous les contraintes que :

- une réaction est ajoutée si elle intervient dans la production minimale d'un métabolite initialement non produit
- le flux de chaque réaction doit être compris dans les limites imposées à la réaction par le modèle
- une réaction ajoutée doit être active signifiant que le flux la traversant ne peut être nul.

L'objectif de ce problème est donc d'ajouter un minimum de réactions actives dans le réseau métabolique afin que chaque métabolite ciblé soient produit ou consommé.

L'utilisation de ces deux outils GapFind et GapFill a été faite sur plusieurs organismes comme *Gordonia alkanivorans* [3], *Salmonella Typhimurium LT2* [94] ou *Clostridium beijerinckii* [75]. Ils permettent une complétion et une modélisation de réseau métabolique permettant de futures analyses mais, n'ayant pas de but de production particulier, ces outils tendent à combler l'ensemble des trous dans le réseau, pouvant ainsi créer des voies métaboliques non nécessaires et donc de faux positifs. Une amélioration possible et utilisée par la méthode de reconstruction Fastgapfill présentée dans la suite est l'orientation de la complétion dans le but de produire un ensemble de métabolites cibles formant la biomasse de l'organisme.

### A.1.3 Fastgapfill : une complétion pour produire l'ensemble des variables du modèle

FastGapFill est un logiciel développé par I. Thiele et ses collaborateurs [96] dont le but est de compléter un réseau métabolique compartimenté afin que ce dernier soit cohérent au niveau des flux et que chacune des réactions métaboliques présentes au sein du réseau soit active dans au moins un modèle de distribution de flux. La reconstruction du réseau métabolique avec FastGapfill porte sur plusieurs étapes qui se suivent méthodiquement.

La première étape commence par la formulation du problème de complétion en lui-même. Le réseau étudié est d'abord analysé afin de déterminer si des réactions sont bloquées car ne pouvant pas avoir de flux non nul mettant en avant des manques dans le réseau métabolique. A partir d'une base de données, des réactions métaboliques sont ajoutées au réseau avec pour contrainte qu'une réaction peut être ajoutée si cela entraîne l'apparition d'un flux au niveau d'une réaction précédemment bloquée.

La deuxième étape de la reconstruction proposée par FastGapFill consiste à l'analyse du réseau métabolique dans le but de vérifier la constance des flux au sein de ce dernier. Pour cela, l'outil commence à partir d'un sous-réseau du réseau étudié de petite taille et constant au niveau des flux. Des ensembles de réactions présentes dans le reste du réseau sont ajoutés par un algorithme glouton si cet ajout ne rompt pas la propriété de constance des flux au sein du nouveau sous-réseau créé. Cette méthode permet d'identifier de façon efficace les réactions bloquées dans le réseau métabolique.

L'étape suivante analyse le réseau dans sa globalité avec sa compartimentation auquel les réactions bloquées découvertes précédemment sont enlevées dans un premier temps. Pour chaque

compartiment du réseau, est ajouté l'ensemble des réactions d'une base de données. De plus, des réactions réversibles de transport sont ajoutées pour les métabolites présent dans un compartiment non cytosolique tandis que des réactions d'échanges sont ajoutées pour les métabolites extra-cellulaires. Certaines réactions initialement bloquées deviennent alors possibles et sont alors remises dans le réseau métabolique. L'ensemble des réactions métaboliques du réseau sans celles de la base de données constitue le réseau cœur du modèle.

La quatrième étape de la reconstruction porte de nouveau sur la constance des flux au sein du réseau cœur défini dans l'étape précédente. Des réactions de la base de données utilisée précédemment sont ajoutés à ce réseau dans cette optique. Une optimisation est utilisée afin de diriger l'ajout des réactions choisies tout en minimisant l'ensemble de ces dernières. Ainsi les réactions se voient attribuer des poids selon leur action, si celle-ci est une réaction métabolique ou une réaction de transfert. La fonction objectif est faite par défaut pour donner la priorité aux réactions métaboliques par rapport aux réactions de transfert mais cela dépend des poids donnés par l'utilisateur permettant ainsi de calculer plusieurs reconstructions possibles.

La cinquième étape est optionnelle et propose une double optimisation au sein du réseau :

- une maximisation des flux passant dans chaque réaction bloquée dans le réseau métabolique initial et devenue active par les étapes précédentes
- une minimisation la norme Euclidienne des flux du sous-réseau calculé précédemment

Cette étape propose un sous-réseau associé à un flux répondant aux critères d'optimisation.

La dernière étape de la complétion effectuée par FastGapFill consiste en une analyse de la stœchiométrie des réactions métaboliques du réseau créé. En effet, des inconsistances au niveau des stœchiométries dans les bases de données peuvent provoquer des erreurs dans l'équilibre des masses du réseau. FastGapfill est capable de détecter ces problèmes en utilisant une méthode scalaire. De plus, il es possible de récupérer un ensemble maximal de métabolites impliqués dans des réactions respectant la loi de conservation de masse.

FastGapFill a été testé sur cinq réseaux métaboliques afin de montrer ses performances sur des réseaux métaboliques ayant différentes tailles. Pour chacun de ces réseaux, plusieurs reconstructions sont effectuées en modifiant les poids des réactions entre celles métaboliques et celles de transport. De plus, l'outil est développé pour pouvoir être utilisé avec différentes bases de données ayant des identifiants en corrélation avec les données obtenues initialement.

L'outil que nous présentons ici repose sur l'analyse et l'utilisation des flux du réseau métabolique pour ajouter des réactions métaboliques. De plus, une hypothèse d'état stationnaire contraint le système afin que tout métabolite produit soit consommé. Ces contraintes entraînent une curation du réseau initial et une limitation des réactions utilisables pour reconstruire le réseau. De plus, ces méthodes rendent la complétion très sensible aux erreurs de stœchiométrie pouvant être présentes dans les bases de données. Un relâchement de ces contraintes et donc une étude plus topologique du réseau métabolique permet, comme le montre la méthode de reconstruction suivante, un plus grand choix de réactions métaboliques pour compléter le réseau.

## A.2 Gapfilling avec programmation logique

Meneco pour *Metabolic Network Construction* est une méthode de complétion de réseau métabolique se différenciant des deux autres méthodes présentées précédemment. En effet, alors que les méthodes GapFill et FastGapFill utilisent des méthodes de flux pour compléter les manques du réseau métabolique, Meneco se focalise sur l'accessibilité topologique des éléments de la biomasse à partir de métabolites sources que nous avons définie dans le chapitre 2. De plus, Meneco utilise également, pour la complétion de réseau métabolique, le paradigme Answer Set Programming (ASP) qui permet de calculer un ensemble minimal de réactions métaboliques à ajouter mais aussi l'ensemble de ces ensembles minimaux de réactions, l'union et l'intersection de ces derniers.

Dans un premier temps, Meneco calcule le scope des métabolites sources donnés en entrée de l'outil correspondant donc à l'ensemble des métabolites accessibles. Il vérifie alors si des métabolites cibles de la biomasse n'appartiennent pas à ce scope pour se concentrer sur la production de ces derniers. Pour chacun de ces métabolites non accessibles, trois actions sont effectuées :

- une recherche d'ensembles de réactions prises dans une base de données et permettant son accessibilité
- l'intersection de ces ensembles donnant l'ensemble des réactions essentielles à sa production
- l'ajout de l'intersection au réseau métabolique

Des premiers ajouts de réactions sont donc faits mais cela ne permet pas de façon certaine l'accessibilité des métabolites cibles.

A partir du nouveau réseau complété par l'étape précédente, une nouvelle complétion est mise en œuvre en prenant en compte l'ensemble des métabolites cibles non accessibles initialement. Cette reconstruction repose sur une minimisation du nombre de réactions métaboliques ajoutées pour rendre accessible les métabolites cibles. La technologie utilisée, ASP, permet de calculer l'ensemble des ensembles répondant aux critères d'optimisation. L'union et l'intersection de ces ensembles réponses sont donc possibles permettant une meilleure compréhension des voies métaboliques minimales permettant l'accessibilité des métabolites cibles.

Meneco est un outil de reconstruction de réseau métabolique et a été notamment utilisé pour compléter le réseau d'*Ectocarpus siliculosus* mais aussi *Candidatus Phaeoamrinobacter ectocarpus* [83]. Il permet aussi de mettre en avant des propriétés au niveau des réactions du réseau métabolique. En effet, les réactions ont pu être divisées en trois ensembles :

- réaction bloquée : réaction qui n'intervient pas dans l'accessibilité d'un métabolite cible
- réaction essentielle : réaction nécessaire à l'accessibilité d'un métabolite cible (si la réaction est absente, un métabolite cible est inaccessible)
- réaction alternative : réaction appartenant à une voie métabolique alternative pour l'accessibilité d'un métabolite cible (si la réaction est absente, une autre voie métabolique prend le relais)

Meneco est donc aussi intéressant pour analyser un réseau métabolique.

Meneco se reposant sur des concepts topologiques, il permet de mettre en évidence des voies métaboliques oubliées par des analyses de flux. Par contre, les solutions présentées par cet outil doivent être ensuite vérifiées biologiquement.

## Annexe B

# Annexe : Données supplémentaires des résultats obtenus sur la clusterisation des six cas de réseaux métaboliques étudiés

Dans le chapitre 6, nous avons mis en avant des clusters dans les six réseaux que nous utilisons comme références : *iJO1360*, *iAF1260*, *iJR904*, *Synechocystis*, *A. ferrooxidans* et *T. lutea*. Nous présentons dans cette annexe les différentes données ayant permis de réaliser les figures du chapitre de résultats sur les classes de carrefours ainsi que les clusters que nous avons plus spécifiquement analysés dans les réseaux avec pour chacun l'ensemble des métabolites essentiels du point de vue de la persistance du réseau présents dans le cluster et l'ensemble des métabolites cibles duals.

### B.1 Données supplémentaire des figures représentant les six cas étudiés de réseaux métaboliques

	<i>iAF1260</i>	<i>iJO1360</i>	<i>iJR904</i>	<i>Synechocystis</i>	<i>T. lutea</i>	<i>A. ferrooxidans</i>	
Nombre de réactions dans le réseau	1967	2129	904	599	299	2456	lié aux Figures 6.1 et 6.2
Nombre de métabolites dans le réseau	2382	2582	1075	759	286	1781	lié aux Figures 6.1 et 6.2
Nombre de carrefours dans le réseau	98	66	70	325	198	316	lié aux Figures 6.1, 6.2 et 6.5
Ratio #carrefours/#métabolites	4%	2%	7%	43%	69%	18%	
Nombre de clusters	27	22	13	61	39	77	lié aux Figures 6.1 et 6.2
Nombre de clusters de carrefours singleton	7	5	6	25	12	37	lié aux Figures 6.3 et 6.4
Ratio #clusters de carrefours singleton/#carrefours	7%	7%	8.5%	19%	19.7%	24%	
Ratio #clusters de carrefours singleton/#clusters	26%	23%	38.5%	41%	31%	48%	
Nombre moyen de carrefours par cluster	3,6	3	5,4	5,3	5,1	4,1	lié aux Figures 6.1 et 6.2
Nombre de clusters avec une unique cible duale	15	16	4	32	16	30	lié aux Figures 6.3, 6.4 and Fig. 6.5
Ratio #clusters avec une unique cible duale/#clusters	55%	73%	31%	52.5%	41%	39%	
Nombre moyen de cibles duales par clusters	2,8	2,1	2,9	2,1	4,8	10,5	lié aux Figures 6.5

TABLE B.1 – Données supplémentaire des Figures 6.1, 6.2, 6.3, 6.4 and 6.5. Distribution des carrefours et des cibles dans les clusters des six cas de réseaux métaboliques étudiés.

	<i>iAF1260</i>	<i>iJO1360</i>	<i>iJR904</i>	<i>Synechocystis</i>	<i>T. lutea</i>	<i>A. ferrooxidans</i>
Minimum	1	1	1	1	1	1
Quartile 25%	1,5	2	1	1	1	1
Moyenne	3,6	3	5,4	5,3	5,1	4,1
Médiane	2	2	2	2	2	2
Quartile 75%	5	2	2	5	4	4
Maximum	14	16	29	38	52	56
Nombre de clusters irréguliers	1	5	3	8	3	6

(a) Distribution des carrefours

	<i>iAF1260</i>	<i>iJO1360</i>	<i>iJR904</i>	<i>Synechocystis</i>	<i>T. lutea</i>	<i>A. ferrooxidans</i>
Minimum	1	1	1	1	1	1
Quartile 25%	1	1	1	1	1	1
Moyenne	2,8	2,1	2,9	2	4,8	10,5
Médiane	1	1	2	1	2	2
Quartile 75%	3,5	1,75	3	2	5	18
Maximum	16	12	12	8	25	55
Nombre de clusters irréguliers	2	4	1	5	2	3

(b) Distribution des métabolites cibles duals

TABLE B.2 – **Données supplémentaires des Figures 6.6 et 6.8.** Distribution des carrefours et des métabolites cibles duals dans les clusters des six cas d'études de réseaux métaboliques.

## B.2 Clusters d'intérêt dans les six cas étudiés de réseaux métaboliques

### B.2.1 Cluster du réseau *iJO1360*

Cluster avec le nombre maximal de métabolites cibles duales

- \* **Cluster.** Le cluster contient un unique carrefour.
  - \* Hydrogen
- \* **Ensemble des cibles dual.** L'ensemble des cibles dual contient 12 métabolites cibles.
 

* Flavin adenine dinucleotide oxidized	* 10-Formyltetrahydrofolate
* L-Tyrosine	* Chloride
* L-Phenylalanine	* 5,10-Methylenetetrahydrofolate
* Coenzyme A	* L-Tryptophan
* NAD	* Riboflavin
* NADP	* 5,6,7,8-Tetrahydrofolate

### B.2.2 Cluster du réseau *iAF1260*

Cluster avec le nombre maximal de métabolites cibles duales

- \* **Cluster.** Le cluster contient un unique carrefour.
  - \* Hydrogen
- \* **Ensemble des cibles dual.** L'ensemble des cibles dual contient 16 métabolites cibles.
 

* Flavin adenine dinucleotide oxidized	* Pyridoxal 5'-phosphate
* L-Tyrosine	* 10-Formyltetrahydrofolate
* 2-Octaprenyl-6-hydroxyphenol	* Chloride
* L-Phenylalanine	* 5,10-Methylenetetrahydrofolate
* Coenzyme A	* Two disaccharide linked murein units
* Undecaprenyl diphosphate	* L-Tryptophan
* NAD	* Riboflavin
* NADP	* 5,6,7,8-Tetrahydrofolate

### B.2.3 Cluster du réseau *iJR904*

Cluster avec le nombre maximal de métabolites cibles duales

- \* **Cluster.** Le cluster contient un unique carrefour.
  - \* Hydrogen
- \* **Ensemble des cibles dual.** L'ensemble des cibles dual contient 12 métabolites cibles.
 

* NAD	* Acetyl-CoA
* L-Tryptophan	* 5-Methyltetrahydrofolate
* L-Tyrosine	* NADPH
* Peptidoglycan subunit	* Glycogen
* UDPglucose	* NADP
* NADH	* L-Phenylalanine

### B.2.4 Clusters du réseau *Synechocystis*

Cluster avec le nombre maximal de métabolites cibles duales

- \* **Cluster.** Le cluster contient un unique carrefour.
  - \* Hydrogen
- \* **Ensemble des cibles dual.** L'ensemble des cibles dual contient 12 métabolites cibles.
 

* RNA component	* Pigment component
* inorganic Ion component	* Lipid component
* DNA component	* Cell wall component
* Soluble pool component	* Protein component

**Cluster avec le nombre maximal de carrefours**

\* **Cluster.** Le cluster contient 38 carrefours.

- |   |   |
|---|---|
| * CDP-diacylglycerol_PG                               | * Octadecanoyl-[acyl-carrier protein]           |
| * 3-D-Glucosyl-1,2-<br>diacylglycerol_MGDG            | * sn-Glycerol 3-phosphate                       |
| * 1,2-Diacyl-3-beta-D-galactosyl-sn-<br>glycerol_MGDG | * trans-Hexadec-2-enoyl                         |
| * Oleoyl-[acyl-carrier protein]                       | * Phosphatidate_PG                              |
| * UDP-6-sulfoquinovose                                | * Stearidonoyl-[acyl-carrier protein]           |
| * Hexadecenoyl-[acyl-carrier protein]                 | * 3-D-Glucosyl-1,2-<br>diacylglycerol_DGDG      |
| * Linoleoyl-[acyl-carrier protein]                    | * Adenylyl sulfate                              |
| * Phosphatidate_DGDG                                  | * 1,2-Diacyl-sn-glycerol_SQDG                   |
| * 1,2-Diacyl-3-beta-D-galactosyl-sn-<br>glycerol_DGDG | * 1,2-Diacyl-sn-glycerol_MGDG                   |
| * Phosphatidylglycerophosphate_PG                     | * (3R)-3-Hydroxypalmitoyl                       |
| * gamma-Linolenoyl-ACP                                | * Phosphatidate_MGDG                            |
| * 3-Oxostearoyl-[acyl-carrier protein]                | * Phosphatidate_SQDG                            |
| * Sulfoquinovosyldiacylglycerol_SQDG                  | * 1,2-Diacyl-sn-glycerol_DGDG                   |
| * Malonyl-[acyl-carrier protein]                      | * Digalactosyl-diacylglycerol_DGDG              |
| * 3-Hydroxyoctadecanoyl-[acyl-carrier<br>protein]     | * 3'-Phosphoadenylyl sulfate                    |
| * alpha-Linolenoyl-ACP                                | * (2E)-Octadecenoyl-[acyl-carrier pro-<br>tein] |
| * (9Z)-Hexadecanoyl-[acyl-carrier pro-<br>tein]       | * Sulfite                                       |
| * Acyl-carrier protein                                | * Phosphatidylglycerol_PG                       |
|   | * UDP-glucose                                   |
|   | * 3-Oxohexadecanoyl-[acyl-carrier pro-<br>tein] |

\* **Ensemble des cibles dual.** L'ensemble des cibles dual contient un unique métabolite cible.

- \* Lipid component

**B.2.5 ClusterS du réseau *A. ferrooxidans*****Cluster avec le nombre maximal de métabolites cibles duales**

\* **Cluster.** Le cluster contient un unique carrefour.

- \* AMP

\* **Ensemble des cibles dual.** L'ensemble des cibles dual contient 55 métabolites cibles.



- \* NADP
- \* L-Glutamate
- \* Glutathione
- \* Riboflavin
- \* L-Phenylalanine
- \* NAD
- \* Chorismate
- \* Spermidine
- \* L-Glutamine
- \* 5,10-Methylenetetrahydrofolate
- \* S-Adenosyl-L-methionine
- \* L-Tyrosine
- \* Putrescine
- \* ATP
- \* CoA
- \* L-Aspartate
- \* L-Methionine
- \* 5-Methyltetrahydrofolate
- \* L-Alanine
- \* UTP
- \* Pyridoxal\_phosphate
- \* L-Threonine
- \* L-Proline
- \* L-Isoleucine
- \* dGTP
- \* Acetyl-CoA
- \* Thiamine\_diphosphate
- \* CoQH2
- \* L-Histidine
- \* Fe3
- \* GTP
- \* Glycogen
- \* Heme\_O
- \* L-Leucine
- \* Lipopolysaccharide-AFE
- \* L-Lysine
- \* Tetrahydrofolate
- \* FAD
- \* NADPH
- \* dATP
- \* Undecaprenyl\_diphosphate
- \* Phospholipid-AFE
- \* dCTP
- \* L-Valine
- \* dTTP
- \* Succinyl-CoA
- \* Glycine
- \* L-Arginine
- \* Mg
- \* L-Serine
- \* Siroheme
- \* L-Asparagine
- \* murein5px4p
- \* 10-Formyltetrahydrofolate
- \* CTP

### Cluster avec le nombre maximal de carrefours

- \* **Cluster.** Le cluster contient 55 carrefours.

- \* FA\_PL-sn-glycerol\_3-phosphate-AFE
- \* V-C15-hydroxyacyl-ACP
- \* Pentanoyl-ACP
- \* 3-oxopentanoyl-ACP
- \* cis-hydroxytetradecenoyl-ACP
- \* cis-oxooctadecenoyl-ACP
- \* 3-cis-Decenoyl-ACP
- \* cis-Hexadecenoyl-ACP
- \* 2-trans-Heptadecenoyl-ACP
- \* 1,2-Diacyl-sn-glycerol\_3-phosphate-AFE
- \* cis-2-trans-Octadidecenoyl-ACP
- \* 3-oxonadecanoyl-ACP
- \* R-3-hydroxypentanoyl-ACP
- \* cis-oxododecenoyl-ACP
- \* Glycerone\_phosphate
- \* FA\_PL-ACP
- \* 3-cis-Dodecenoyl-ACP
- \* V-C15-oxoacyl-ACP
- \* CDP-di-acyl-glycerol-AFE
- \* cis-oxohexadecenoyl-ACP
- \* C19\_cyclopropane-ACP
- \* cis-Tetradecenoyl-ACP
- \* V-C15-trans-enoyl-ACP
- \* 3-oxooctadecanoyl-ACP

- \* cis-oxotetradecenoyl-ACP
  - \* R-3-hydroxyoctadecanoyl-ACP
  - \* Heptadecanoyl-ACP
  - \* cis-2-trans-Tetradidecenoyl-ACP
  - \* L-1-phosphatidylglycerol-phosphate-AFE
  - \* cis-hydroxyoctadecenoyl-ACP
  - \* Propionyl-CoA
  - \* 2-trans-Octadecenoyl-ACP
  - \* Cardiolipin-AFE
  - \* R-3-hydroxynonadecanoyl-ACP
  - \* R-3-hydroxyhexadecanoyl-ACP
  - \* Hexadecanoyl-ACP
  - \* cis-Octadecenoyl-ACP
  - \* R-3-hydroxyheptadecanoyl-ACP
  - \* cis-2-trans-Hexadidecenoyl-ACP
  - \* cis-hydroxyhexadecenoyl-ACP
  - \* 2-trans-Hexadecenoyl-ACP
  - \* 3-oxoheptadecanoyl-ACP
  - \* sn-Glycerol\_3-phosphate
  - \* 3-oxohexadecanoyl-ACP
  - \* Phosphatidyl-N-methylethanolamine-AFE
  - \* Phosphatidylethanolamine-AFE
  - \* Octadecanoyl-ACP
  - \* 2-trans-Pentenoyl-ACP
  - \* Methylmalonyl-CoA
  - \* Glycerol
  - \* cis-2-trans-Dodidecenoyl-ACP
  - \* cis-hydroxydodecenoyl-ACP
  - \* Phosphatidylserine-AFE
  - \* 2-trans-Nonadecenoyl-ACP
  - \* Phosphatidylglycerol-AFE
  - \* Pentadecanoyl-ACP
- \* **Ensemble des cibles dual.** L'ensemble des cibles dual contient un unique métabolite cible.
- \* Phospholipid-AFE

### B.2.6 Cluster du réseau *T. lutea*

#### Cluster avec le nombre maximal de métabolites cibles duales

- \* **Cluster.** Le cluster contient un unique carrefour.
  - \* hydrogencarbonate
- \* **Ensemble des cibles dual.** L'ensemble des cibles dual contient 25 métabolites cibles.
  - \* L-glutamate (GLT)
  - \* L-glutamine (GLN)
  - \* L-cysteine (CYS)
  - \* L-asparagine (ASN)
  - \* L-arginine (ARG)
  - \* glycine (GLY)
  - \* L-valine (VAL)
  - \* L-isoleucine (ILE)
  - \* L-histidine (HIS)
  - \* L-phosphatidate
  - \* L-aspartate
  - \* Chlorophyll-A
  - \* RNAs
  - \* DNAs
  - \* L-tryptophan (TRP)
  - \* L-tyrosine (TYR)
  - \* L-serine (SER)
  - \* L-threonine (THR)
  - \* L-phenylalanine (PHE)
  - \* L-proline (PRO)
  - \* L-lysine (LYS)
  - \* L-methionine (MET)
  - \* L-alpha-alanine
  - \* 4-hydroxy-L-proline
  - \* L-leucine (LEU)



# Bibliographie

- [1] V. Acuna, P. V. Milreu, L. Cottret, A. Marchetti-Spaccamela, L. Stougie, and M. F. Sagot. Algorithms and complexity of enumerating minimal precursor sets in genome-wide metabolic networks. *Bioinformatics*, 28(19) :2474–2483, Oct 2012.
- [2] Vicente Acuña, Alberto Marchetti-Spaccamela, Marie-France Sagot, and Leen Stougie. A note on the complexity of finding and enumerating elementary modes. *Biosystems*, 99(3) :210–214, 2010.
- [3] Shilpi Aggarwal, IA Karimi, and Gregorius Reinaldi Ivan. In silico modeling and evaluation of gordonia alkanivorans for biodesulfurization. *Molecular BioSystems*, 9(10) :2530–2540, 2013.
- [4] Uri Alon. Network motifs : theory and experimental approaches. *Nature Reviews Genetics*, 8(6) :450–461, 2007.
- [5] Geoffroy Andrieux, Laurent Fattet, Michel Le Borgne, Ruth Rimokh, and Nathalie Th  ret. Dynamic regulation of tgf- $\beta$  signaling by *tif1 $\gamma$*  : a computational approach. *PloS one*, 7(3) :e33761, 2012.
- [6] Kathrin Ballerstein, Axel von Kamp, Steffen Klamt, and Utz-Uwe Haus. Minimal cut sets in a metabolic network are elementary modes in a dual network. *Bioinformatics*, 28(3) :381–387, 2011.
- [7] Chitta Baral. *Knowledge representation, reasoning and declarative problem solving*. Cambridge university press, 2003.
- [8] C. Baroukh, R. Munoz-Tamayo, J. P. Steyer, and O. Bernard. DRUM : a new framework for metabolic modeling under non-balanced growth. Application to the carbon metabolism of unicellular microalgae. *PLoS ONE*, 9(8) :e104499, 2014.
- [9] David J Baumler, Roman G Peplinski, Jennifer L Reed, Jeremy D Glasner, and Nicole T Perna. The evolution of metabolic networks of e. coli. *BMC systems biology*, 5(1) :182, 2011.
- [10] Daniel A Beard, Shou-dan Liang, and Hong Qian. Energy balance for analysis of complex metabolic networks. *Biophysical journal*, 83(1) :79–86, 2002.
- [11] Steven A Benner and A Michael Sismour. Synthetic biology. *Nature reviews. Genetics*, 6(7) :533, 2005.
- [12] Philippe N Bertin, Audrey Heinrich-Salmeron, Eric Pelletier, Florence Goulhen-Chollet, Florence Ars  ne-Ploetze, S  bastien Gallien, B  atrice Lauga, Corinne Casiot, Alexandra Calteau, David Vallenet, et al. Metabolic diversity among main microorganisms inside an arsenic-rich ecosystem revealed by meta-and proteo-genomics. *The ISME journal*, 5(11) :1735, 2011.
- [13] M. Beurton-Aimar, T. V. Nguyen, and S. Colombie. Metabolic network reconstruction and their topological analysis. *Methods Mol. Biol.*, 1090 :19–38, 2014.
- [14] Miguel A. Campodonico, Daniela Vaisman, Jean F. Castro, Valeria Razmilic, Francesca Mercado, Barbara A. Andrews, Adam M. Feist, and Juan A. Asenjo. Acidithiobacillus ferrooxidans’s comprehensive model driven analysis of the electron transfer metabolism and synthetic strain design for biomining applications. *Metabolic Engineering Communications*, 3 :84 – 96, 2016.
- [15] Ron Caspi, Hartmut Foerster, Carol A Fulcher, Pallavi Kaipa, Markus Krummenacker, Mario Latendresse, Suzanne Paley, Seung Y Rhee, Alexander G Shearer, Christophe Tissier, et al. The metacyc database of metabolic pathways and enzymes and the biocyc collection of pathway/genome databases. *Nucleic acids research*, 36(suppl\_1) :D623–D631, 2007.

- [16] Markus W Covert, Nan Xiao, Tiffany J Chen, and Jonathan R Karr. Integrating metabolic, transcriptional regulatory and signal transduction models in escherichia coli. *Bioinformatics*, 24(18) :2044–2050, 2008.
- [17] George Dantzig. *Linear programming and extensions*. Princeton university press, 2016.
- [18] Eric Davidson and Michael Levin. Gene regulatory networks, 2005.
- [19] L. F. de Figueiredo, S. Schuster, C. Kaleta, and D. A. Fell. Can sugars be produced from fatty acids? A test case for pathway analysis tools. *Bioinformatics*, 25(1) :152–158, Jan 2009.
- [20] David Deutscher, Isaac Meilijson, Martin Kupiec, and Eytan Ruppin. Multiple knockout analysis of genetic robustness in the yeast metabolic network. *Nature genetics*, 38(9) :993, 2006.
- [21] Natalie C Duarte, Scott A Becker, Neema Jamshidi, Ines Thiele, Monica L Mo, Thuy D Vo, Rohith Srivas, and Bernhard Ø Palsson. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proceedings of the National Academy of Sciences*, 104(6) :1777–1782, 2007.
- [22] Natalie C Duarte, Markus J Herrgård, and Bernhard Ø Palsson. Reconstruction and validation of *saccharomyces cerevisiae* ind750, a fully compartmentalized genome-scale metabolic model. *Genome research*, 14(7) :1298–1309, 2004.
- [23] Ali Ebrahim, Joshua A Lerman, Bernhard O Palsson, and Daniel R Hyduke. Cobrapy : Constraints-based reconstruction and analysis for python. *BMC systems biology*, 7(1) :74, 2013.
- [24] JS Edwards and BO Palsson. The escherichia coli mg1655 in silico metabolic genotype : its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences*, 97(10) :5528–5533, 2000.
- [25] Michele Fabris, Michiel Matthijs, Stephane Rombauts, Wim Vyverman, Alain Goossens, and Gino JE Baart. The metabolic blueprint of *phaeodactylum tricornutum* reveals a eukaryotic entner–doudoroff glycolytic pathway. *The Plant Journal*, 70(6) :1004–1014, 2012.
- [26] Adam M Feist, Christopher S Henry, Jennifer L Reed, Markus Krummenacker, Andrew R Joyce, Peter D Karp, Linda J Broadbelt, Vassily Hatzimanikatis, and Bernhard Ø Palsson. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular systems biology*, 3 :121, 2007.
- [27] Adam M Feist and Bernhard O Palsson. The biomass objective function. *Current opinion in microbiology*, 13(3) :344–349, 2010.
- [28] David A Fell and J Rankin Small. Fat synthesis in adipose tissue. an examination of stoichiometric constraints. *Biochemical Journal*, 238(3) :781–786, 1986.
- [29] Robert D Finn, Alex Bateman, Jody Clements, Penelope Coggill, Ruth Y Eberhardt, Sean R Eddy, Andreas Heger, Kirstie Hetherington, Liisa Holm, Jaina Mistry, et al. Pfam : the protein families database. *Nucleic acids research*, 42(D1) :D222–D230, 2013.
- [30] Clémence Frioux, Torsten Schaub, Sebastian Schellhorn, Anne Siegel, and Philipp Wanko. *Hybrid Metabolic Network Completion*, pages 308–321. Springer International Publishing, Cham, 2017.
- [31] Komei Fukuda and Alain Prodon. Double description method revisited. *Combinatorics and computer science*, pages 91–111, 1996.
- [32] Mathias Ganter, Thomas Bernard, Sébastien Moretti, Joerg Stelling, and Marco Pagni. Metanetx. org : a website and repository for accessing, analysing and manipulating metabolic networks. *Bioinformatics*, 29(6) :815–816, 2013.
- [33] M. Gebser, R. Kaminski, B. Kaufmann, and T. Schaub. *Answer Set Solving in Practice*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan and Claypool Publishers, 2012.
- [34] Martin Gebser, Roland Kaminski, Benjamin Kaufmann, and Torsten Schaub. Answer set solving in practice. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 6(3) :1–238, 2012.

- [35] Matthieu Graindorge, Cécile Giustini, Anne Claire Jacomin, Alexandra Kraut, Gilles Curien, and Michel Matringe. Identification of a plant gene encoding glutamate/aspartate-prephenate aminotransferase : The last homeless enzyme of aromatic amino acids biosynthesis. *FEBS letters*, 584(20) :4357–4360, 2010.
- [36] Steinn Gudmundsson and Ines Thiele. Computationally efficient flux variability analysis. *BMC bioinformatics*, 11(1) :489, 2010.
- [37] Joshua J Hamilton and Jennifer L Reed. Software platforms to facilitate reconstructing genome-scale metabolic networks. *Environmental microbiology*, 16(1) :49–59, 2014.
- [38] T. Handorf, O. Ebenhoh, and R. Heinrich. Expanding metabolic networks : scopes of compounds, robustness, and evolution. *J. Mol. Evol.*, 61(4) :498–512, Oct 2005.
- [39] Thomas Handorf, Nils Christian, Oliver Ebenhöh, and Daniel Kahn. An environmental perspective on metabolism. *Journal of theoretical biology*, 252(3) :530–537, 2008.
- [40] Jordan Hay and Jörg Schwender. Metabolic network reconstruction and flux variability analysis of storage synthesis in developing oilseed rape (*brassica napus* l.) embryos. *The Plant Journal*, 67(3) :526–541, 2011.
- [41] Christopher S Henry, Matthew DeJongh, Aaron A Best, Paul M Frybarger, Ben Lindsay, and Rick L Stevens. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nature biotechnology*, 28(9) :977–982, 2010.
- [42] Michael Hucka, Andrew Finney, Herbert M Sauro, Hamid Bolouri, John C Doyle, Hiroaki Kitano, Adam P Arkin, Benjamin J Bornstein, Dennis Bray, Athel Cornish-Bowden, et al. The systems biology markup language (sbml) : a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4) :524–531, 2003.
- [43] IBM ILOG. Cplex optimizer. *En ligne*. Available : <http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer>, 2012.
- [44] François Jacob and Jacques Monod. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of molecular biology*, 3(3) :318–356, 1961.
- [45] Hawoong Jeong, Bálint Tombor, Réka Albert, Zoltan N Oltvai, and A-L Barabási. The large-scale organization of metabolic networks. *arXiv preprint cond-mat/0010278*, 2000.
- [46] Tao Jia, Yang-Yu Liu, Endre Csóka, Márton Pósfai, Jean-Jacques Slotine, and Albert-László Barabási. Emergence of bimodality in controlling complex networks. *arXiv preprint arXiv :1505.06476*, 2015.
- [47] J Dedrick Jordan, Emmanuel M Landau, and Ravi Iyengar. Signaling networks : the origins of cellular multitasking. *Cell*, 103(2) :193–200, 2000.
- [48] Christian Jungreuthmayer, David E Ruckerbauer, and Jürgen Zanghellini. regfntool : Speeding up elementary flux mode calculation using transcriptional regulatory rules in the form of three-state logic. *Biosystems*, 113(1) :37–39, 2013.
- [49] Axel von Kamp and Stefan Schuster. Metatool 5.0 : fast and flexible elementary modes analysis. *Bioinformatics*, 22(15) :1930–1931, 2006.
- [50] Minoru Kanehisa, Susumu Goto, Yoko Sato, Masayuki Kawashima, Miho Furumichi, and Mao Tanabe. Data, information, knowledge and principle : back to metabolism in kegg. *Nucleic acids research*, 42(D1) :D199–D205, 2013.
- [51] Takakazu Kaneko, Shusei Sato, Hirokazu Kotani, Ayako Tanaka, Erika Asamizu, Yasukazu Nakamura, Nobuyuki Miyajima, Makoto Hirosawa, Masahiro Sugiura, Shigemi Sasamoto, et al. Sequence analysis of the genome of the unicellular cyanobacterium *synechocystis* sp. strain pcc6803. ii. sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA research*, 3(3) :109–136, 1996.
- [52] Peter D Karp, Christos A Ouzounis, Caroline Moore-Kochlacs, Leon Goldovsky, Pallavi Kaipa, Dag Ahrén, Sophia Tsoka, Nikos Darzentas, Victor Kunin, and Núria López-Bigas. Expansion of the biocyc collection of pathway/genome databases to 160 genomes. *Nucleic acids research*, 33(19) :6083–6089, 2005.
- [53] Eric Karsenti. Self-organization in cell biology : a brief history. *Nature Reviews Molecular Cell Biology*, 9(3) :255–262, 2008.

- [54] Hyun Uk Kim, Tae Yong Kim, and Sang Yup Lee. Genome-scale metabolic network analysis and drug targeting of multi-drug resistant pathogen *acinetobacter baumannii* aye. *Molecular BioSystems*, 6(2) :339–348, 2010.
- [55] Pan-Jun Kim, Dong-Yup Lee, Tae Yong Kim, Kwang Ho Lee, Hawoong Jeong, Sang Yup Lee, and Sunwon Park. Metabolite essentiality elucidates robustness of *escherichia coli* metabolism. *Proceedings of the National Academy of Sciences*, 104(34) :13638–13642, 2007.
- [56] Tae Yong Kim, Seung Bum Sohn, Yu Bin Kim, Won Jun Kim, and Sang Yup Lee. Recent advances in reconstruction and applications of genome-scale metabolic models. *Current opinion in biotechnology*, 23(4) :617–623, 2012.
- [57] Zachary A King, Justin Lu, Andreas Dräger, Philip Miller, Stephen Federowicz, Joshua A Lerman, Ali Ebrahim, Bernhard O Palsson, and Nathan E Lewis. Bigg models : A platform for integrating, standardizing and sharing genome-scale models. *Nucleic acids research*, 44(D1) :D515–D522, 2015.
- [58] Hiroaki Kitano et al. *Foundations of systems biology*. MIT press Cambridge, 2001.
- [59] Steffen Klamt and Ernst Dieter Gilles. Minimal cut sets in biochemical reaction networks. *Bioinformatics*, 20(2) :226–234, 2004.
- [60] Anna MJ Kliphuis, Anne J Klok, Dirk E Martens, Packo P Lamers, Marcel Janssen, and René H Wijffels. Metabolic modeling of *chlamydomonas reinhardtii* : energy requirements for photoautotrophic growth and maintenance. *Journal of applied phycology*, 24(2) :253, 2012.
- [61] Henning Knoop, Marianne Gründel, Yvonne Zilliges, Robert Lehmann, Sabrina Hoffmann, Wolfgang Lockau, and Ralf Steuer. Flux Balance Analysis of Cyanobacterial Metabolism : The Metabolic Network of *Synechocystis* sp. PCC 6803. *PLoS Computational Biology*, 9(6) :e1003081, jun 2013.
- [62] Kai Kruse and Oliver Ebenhöf. Comparing flux balance analysis to network expansion. *Genome Informatics*, 20 :91–101, 2008.
- [63] Vinay Satish Kumar, Madhukar S Dasika, and Costas D Maranas. Optimization based automated curation of metabolic reconstructions. *BMC bioinformatics*, 8(1) :212, 2007.
- [64] Anne Kümmel, Sven Panke, and Matthias Heinemann. Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. *Molecular systems biology*, 2(1), 2006.
- [65] Meiyappan Lakshmanan, Geoffrey Koh, Bevan K.S. Chung, and Dong-Yup Lee. Software applications for flux balance analysis. *Briefings in Bioinformatics*, 15(1) :108–122, 2014.
- [66] Eric S Lander, Lauren M Linton, Bruce Birren, Chad Nusbaum, Michael C Zody, Jennifer Baldwin, Keri Devon, Ken Dewar, Michael Doyle, William Fitzhugh, et al. Initial sequencing and analysis of the human genome. *Nature*, 409(6822) :860–921, 2001.
- [67] Julie Laniau, Clémence Firoux, Jacques Nicolas, Caroline Baroukh, Maria-Paz Cortes, Jeanne Got, Camille Trottier, Damien Eveillard, and Anne Siegel. Combining graph and flux-based structures to decipher phenotypic essential metabolites within metabolic networks. *PeerJ*, 2017.
- [68] Sébastien Laurent, Adrien Richard, Odile Mulner-Lorillon, Julia Morales, Didier Flament, Virginie Glippa, Jérémie Bourdon, Pauline Gosselin, Anne Siegel, Patrick Cormier, et al. Modelization of the regulation of protein synthesis following fertilization in sea urchin shows requirement of two processes : a destabilization of eif4e : 4e-bp complex and a great stimulation of the 4e-bp-degradation mechanism, both rapamycin-sensitive. *Frontiers in genetics*, 5, 2014.
- [69] Li Li, Christian J Stoeckert, and David S Roos. Orthomcl : identification of ortholog groups for eukaryotic genomes. *Genome research*, 13(9) :2178–2189, 2003.
- [70] Nicolas Loira, Thierry Dulermo, Jean-Marc Nicaud, and David James Sherman. A genome-scale metabolic model of the lipid-accumulating yeast *yarrowia lipolytica*. *BMC systems biology*, 6(1) :35, 2012.
- [71] Daniel Machado, Rafael S Costa, Miguel Rocha, Eugénio C Ferreira, Bruce Tidor, and Isabel Rocha. Modeling formalisms in systems biology. *AMB express*, 1(1) :45, 2011.

- [72] Stefania Magnúsdóttir, Almut Heinken, Laura Kutt, Dmitry A Ravcheev, Eugen Bauer, Alberto Noronha, Kacy Greenhalgh, Christian Jäger, Joanna Baginska, Paul Wilmes, Ronan M T Fleming, and Ines Thiele. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nature Biotechnology*, 35(1) :81–89, January 2017.
- [73] Andrew Makhorin et al. Glpk (gnu linear programming kit), 2008.
- [74] P Manimaran, Shubhada R Hegde, and Shekhar C Mande. Prediction of conditional gene essentiality through graph theoretical analysis of genome-wide functional linkages. *Molecular Biosystems*, 5(12) :1936–1942, 2009.
- [75] Caroline B Milne, James A Eddy, Ravali Raju, Soroush Ardekani, Pan-Jun Kim, Ryan S Senger, Yong-Su Jin, Hans P Blaschek, and Nathan D Price. Metabolic network reconstruction and genome-scale model of butanol-producing strain *Clostridium beijerinckii* ncimb 8052. *BMC systems biology*, 5(1) :130, 2011.
- [76] Carmen G Moles, Pedro Mendes, and Julio R Banga. Parameter estimation in biochemical pathways : a comparison of global optimization methods. *Genome research*, 13(11) :2467–2474, 2003.
- [77] Jonathan Monk, Juan Nogales, and Bernhard O Palsson. Optimizing genome-scale network reconstructions. *Nature biotechnology*, 32(5) :447–452, 2014.
- [78] Jeffrey D Orth, Tom M Conrad, Jessica Na, Joshua A Lerman, Hojung Nam, Adam M Feist, and Bernhard Ø Palsson. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism–2011. *Molecular systems biology*, 7 :535, oct 2011.
- [79] Jeffrey D Orth, Ines Thiele, and Bernhard Ø Palsson. What is flux balance analysis? *Nature biotechnology*, 28(3) :245–248, 2010.
- [80] Bernhard Palsson. In silico biology through “omics”. *Nature biotechnology*, 20(7) :649–650, 2002.
- [81] Maria Concetta Palumbo, Alfredo Colosimo, Alessandro Giuliani, and Lorenzo Farina. Essentiality is an emergent property of metabolic network wiring. *FEBS Letters*, 581(13) :2485–2489, 2007.
- [82] Nathan D Price, Jennifer L Reed, and Bernhard Ø Palsson. Genome-scale models of microbial cells : evaluating the consequences of constraints. *Nature reviews. Microbiology*, 2(11) :886, 2004.
- [83] Sylvain Prigent, Clémence Frioux, Simon M Dittami, Sven Thiele, Abdelhalim Larhlimi, Guillaume Collet, Fabien Gutknecht, Jeanne Got, Damien Eveillard, Jérémie Bourdon, et al. Meneco, a topology-based gap-filling tool applicable to degraded genome-wide metabolic networks. *PLoS computational biology*, 13(1) :e1005276, 2017.
- [84] Priscilla EM Purnick and Ron Weiss. The second wave of synthetic biology : from modules to systems. *Nature reviews Molecular cell biology*, 10(6) :410–422, 2009.
- [85] Jennifer L Reed, Thuy D Vo, Christophe H Schilling, and Bernhard O Palsson. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome biology*, 4(9) :R54, 2003.
- [86] Isabel Rocha, Paulo Maia, Pedro Evangelista, Paulo Vilaça, Simão Soares, José P Pinto, Jens Nielsen, Kiran R Patil, Eugénio C Ferreira, and Miguel Rocha. Optflux : an open-source software platform for in silico metabolic engineering. *BMC systems biology*, 4(1) :45, 2010.
- [87] PR Romero and P Karp. Nutrient-related analysis of pathway/genome databases. In *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, pages 471–482, 2001.
- [88] Areejit Samal, Shalini Singh, Varun Giri, Sandeep Krishna, Nandula Raghuram, and Sanjay Jain. Low degree metabolites explain essential reactions and enhance modularity in biological networks. *BMC bioinformatics*, 7(1) :118, 2006.
- [89] Torsten Schaub and Sven Thiele. Metabolic network expansion with answer set programming. In *International Conference on Logic Programming*, pages 312–326. Springer, 2009.
- [90] Torsten Schaub and Sven Thiele. *Metabolic Network Expansion with Answer Set Programming*, pages 312–326. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.



- 
- [91] Stefan Schuster and Claus Hilgetag. On elementary flux modes in biochemical reaction systems at steady state. *Journal of Biological Systems*, 2(02) :165–182, 1994.
- [92] Jörg Stelling, Steffen Klamt, Katja Bettenbrock, Stefan Schuster, and Ernst Dieter Gilles. Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 420(6912) :190–193, 2002.
- [93] Marco Terzer and Jörg Stelling. Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics*, 24(19) :2229–2235, 2008.
- [94] Ines Thiele, Daniel R Hyduke, Benjamin Steeb, Guy Fankam, Douglas K Allen, Susanna Bazzani, Pep Charusanti, Feng-Chi Chen, Ronan MT Fleming, Chao A Hsiung, et al. A community effort towards a knowledge-base and mathematical model of the human pathogen salmonella typhimurium lt2. *BMC systems biology*, 5(1) :8, 2011.
- [95] Ines Thiele and Bernhard Ø Palsson. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature protocols*, 5(1) :93, 2010.
- [96] Ines Thiele, Nikos Vlassis, and Ronan MT Fleming. fastgapfill : efficient gap filling in metabolic networks. *Bioinformatics*, 30(17) :2529–2531, 2014.
- [97] Johan Timmermans and Laurence Van Melderen. Conditional essentiality of the csra gene in escherichia coli. *Journal of bacteriology*, 191(5) :1722–1724, 2009.
- [98] Elizabeth A Winzeler, Daniel D Shoemaker, Anna Astromoff, Hong Liang, Keith Anderson, Bruno Andre, Rhonda Bangham, Rocio Benito, Jef D Boeke, Howard Bussey, et al. Functional characterization of the s. cerevisiae genome by gene deletion and parallel analysis. *science*, 285(5429) :901–906, 1999.
- [99] Jürgen Zanghellini, David E Ruckerbauer, Michael Hanscho, and Christian Jungreuthmayer. Elementary flux modes in a nutshell : properties, calculation and applications. *Biotechnology journal*, 8(9) :1009–1016, 2013.

## Résumé

Durant cette thèse nous nous sommes intéressés aux réseaux métaboliques et notamment leur modélisation sous forme d'un graphe biparti orienté pondéré. Ce dernier permet d'étudier la production d'éléments cibles métaboliques regroupés dans une biomasse à partir de composants provenant du milieu de croissance de l'organisme. Nous nous sommes plus particulièrement penchés sur le rôle des métabolites internes au réseau et la notion d'essentialité de ces derniers pour la production d'une biomasse dont nous avons raffiné la définition dans le cas d'une étude de flux (métabolite essentiel du point de vue de la productibilité du réseau et métabolite essentiel du point de vue de l'efficacité du réseau) puis étendu cette dernière dans le cas d'une étude topologique (métabolite essentiel du point de vue de la persistance du réseau). Nous nous sommes pour cela reposés sur le formalisme d'un part de *Flux Balance Analysis* et ses dérivés, et d'autre part d'expansion de réseau, afin de définir un métabolite essentiel (ou carrefour), nous permettant de mettre au point un package python (*Conquests*) cherchant les carrefours dans un réseau métabolite. Nous avons appliqué ce dernier à six réseaux métaboliques dont quatre provenant d'espèces modèles (*iJO1360*, *iAF1260* et *iJR904* d'*E. coli* et *Synechocystis*) et les deux autres d'espèces plus spécifiques (*A. ferrooxidans* et *T. lutea*). Nous avons aussi défini le concept de cluster de métabolites essentiels du point de vue de la persistance du réseau lié aux composants de la biomasse auxquels ils sont nécessaires et que nous avons appliqué sur les six réseaux métaboliques précédents et sur 3600 réseaux dégradés du réseau *iJR904* de *E. coli* puis reconstruits selon trois méthodes de gapfilling (Gapfill, Fastgapfill et Meneco) afin de comparer ces dernières. Ces études nous ont permis de mettre en avant l'importance de métabolites internes dans la production de composés cibles.

## Abstract

In this thesis we are interested in metabolic networks and, in particular, their modelling with a weighted directed bipartite graph. This representation makes it possible to study the production of target metabolic elements, constituting a biomass, from components coming from the growth medium of the organism. We focused on the role of metabolites inside the network and the notion of essentiality for these elements for the production of a biomass whose definition we have refined in the case of a flow study (metabolite essential for biomass producibility and metabolite essential for biomass efficiency) and extended this notion in the case of a topological study (metabolite essential for biomass sustainability). We rely on the formalism of *Flux Balance Analysis* and its derivatives, and of network expansion, in order to define an essential metabolite (ME or crossroad), allowing us to develop a python package (*Conquests*) looking for crossroads in a metabolic network. We applied our concept to six metabolic networks, four of which came from model species (*iJO1360*, *iAF1260* and *iJR904* of *E. coli* and *Synechocystis*) and the other two from more specific species (*A. ferrooxidans* and *T. lutea*). We have also defined the concept of cluster of ME-sustainability, related to the biomass components to which they are required and which we have applied over the six previous metabolic networks and over 3600 degraded networks of *iJR904* of *E. coli* and reconstructed according to three methods of gapfilling (Gapfill, Fastgapfill and Meneco) to compare the results. These studies have allowed us to highlight the importance of internal metabolites in the production of target compounds.