



HAL
open science

Approche hybride de modélisation explicable du métabolisme des écosystèmes microbiens

Maxime Lecomte

► **To cite this version:**

Maxime Lecomte. Approche hybride de modélisation explicable du métabolisme des écosystèmes microbiens. Modélisation et simulation. Université de Bordeaux, 2024. Français. NNT : 2024BORD0094 . tel-04583312

HAL Id: tel-04583312

<https://theses.hal.science/tel-04583312>

Submitted on 22 May 2024

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

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



Distributed under a Creative Commons Attribution 4.0 International License

THÈSE PRÉSENTÉE
POUR OBTENIR LE GRADE DE
DOCTEUR
DE L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE MATHÉMATIQUES ET
INFORMATIQUE
SPÉCIALITÉ INFORMATIQUE

Par **Maxime LECOMTE**

Approche hybride de modélisation explicable du métabolisme
des écosystèmes microbiens

Hybrid approach for explainable metabolic modelling of
microbial ecosystems

Sous la direction de : **David SHERMAN** et
Hélène FALENTIN

Soutenue le 29/04/2024

Membres du jury :

Mme. BAROUKH, Caroline	Chargée de recherche HDR	INRAE	Rapportrice
Mme. COCAIGN-BOUSQUET, Muriel	Directrice de recherche	INRAE	Rapportrice
M. COTTRET, Ludovic	Ingénieur de recherche	INRAE	Examinateur
Mme. LAROCHE, Béatrice	Directrice de recherche	INRAE	Examinatrice
M. MARKOV, Gabriel	Chargé de recherche	CNRS	Examinateur
M. SIMON, Laurent	Professeur	LaBRI	Président
Mme. FALENTIN, Hélène	Ingénieure de Recherche HDR	INRAE	Co-Directrice de thèse
M. SHERMAN, David	Directeur de Recherche	Inria	Co-Directeur de thèse

Membre invitée :

Mme. FRIOUX, Clémence	Chargée de Recherche	Inria	Encadrante
-----------------------	----------------------	-------	------------

Table des matières

Résumé	5
Remerciements	8
Introduction	10
0.1 Émergence de la microbiologie <i>in silico</i>	11
0.2 La biologie des systèmes comme support de l'étude du métabolisme	12
0.3 Objectifs de la thèse	12
1 Modélisation du métabolisme : d'une cellule à un écosystème	15
1.1 Les modèles du métabolisme comme support de l'analyse microbienne	16
1.1.1 Reconstruction de réseaux métaboliques	16
1.1.2 Représentations mathématiques et informatiques du métabolisme	20
1.1.2.1 Matrice stœchiométrique	21
1.1.2.2 Graphe	21
1.1.3 Analyse structurelle de réseaux métaboliques	21
1.2 Modélisation métabolique à l'échelle d'un génome	24
1.2.1 Méthodes numériques d'un réseau métabolique	24
1.2.1.1 Constantes cinétiques des réactions	24
1.2.1.2 Contraintes numériques	24
1.2.1.3 Intégration de données multi-omiques	27
1.2.2 Méthodes qualitatives d'analyse d'un réseau métabolique	28
1.2.2.1 Graphes	28
1.2.2.2 Problème SAT	28
1.2.2.3 Application à des problèmes biologiques	29
1.2.3 Défis rencontrés lors de la modélisation qualitative et numérique d'une cellule	30
1.2.4 Méthodes hybrides de modélisation à l'échelle d'un modèle métabo- lique	30
1.2.4.1 Indépendants	31
1.2.4.2 Couplé	31
1.3 Modélisation du métabolisme des écosystèmes microbiens	31
1.3.1 Méthodes numériques appliquées à plusieurs réseaux métaboliques	32
1.3.1.1 Équations différentielles ordinaires (ODE)	32
1.3.1.2 Analyse statique sous contraintes	32
1.3.1.3 Analyse dynamique sous contraintes	34
1.3.2 Applications	36
1.3.3 Méthodes et modèles qualitatifs de l'étude d'une communauté bac- térienne	37

1.3.3.1	Approches basées sur les graphes	38
1.3.3.2	Modèles basés sur le raisonnement	38
1.3.4	Défis rencontrés lors de l'analyse de la modélisation qualitative et quantitative du métabolisme des communautés	40
2	Modèle numérique du métabolisme bactérien au cours de la production d'un fromage	42
2.1	Introduction	43
2.2	Méthode pour la construction d'un modèle numérique du métabolisme . . .	44
2.2.1	Présentation des données	44
2.2.1.1	Données omiques	44
2.2.1.2	Données de calibration des modèles individuelles	45
2.2.1.3	Données de validation du modèle de communauté	46
2.2.2	Du génome annoté au réseau métabolique à l'échelle du génome . .	48
2.2.3	Modélisation et implémentation numérique	48
2.2.4	Raffinement des modèles métaboliques individuels	51
2.2.5	Calibration des modèles métaboliques individuels	51
2.3	Application de la méthode numérique à l'écosystème fromager	52
2.3.1	A l'échelle de chaque souche	52
2.3.2	Simulations à l'échelle de la communauté	64
2.3.3	Interactions bactériennes et exploration du modèle de communauté	70
2.4	Discussion et Conclusion	76
3	Modèle discret pour le criblage haut-débit du potentiel d'interaction métabolique dans les écosystèmes microbiens complexes	82
3.1	Introduction	83
3.2	Méthode pour calculer un score d'interaction à partir d'une base de connaissance	85
3.2.1	Base de faits d'un réseau métabolique	85
3.2.2	Productibilité d'un métabolite	85
3.2.3	Inférence de règles logiques pour calculer les potentiels de coopération et de compétition	86
3.2.4	Calcul des potentiels de coopération et de compétition à partir des règles logiques	89
3.2.5	Expériences <i>in silico</i> pour tester les propriétés des scores	91
3.2.5.1	Implémentation	94
3.3	Application et description de notre approche sur des communautés jouets .	94
3.4	Application à des communautés synthétiques	101
3.4.1	Description des potentiels de coopération et de compétition dans différents écosystèmes	101
3.4.2	Comparaison avec des méthodes basées sur des contraintes et application à des données expérimentales	112
3.5	Discussion et conclusion	115
4	Enrichissement des modèles qualitatifs : modèle de logique linéaire temporelle et méthode de sélection de communauté à façon	116
4.1	Introduction	117
4.2	Critères d'optimisation à partir de contraintes biologiques <i>a priori</i>	118
4.2.1	Méthodologie	119

4.2.2	Résultats préliminaires	124
4.3	Discussion sur la sélection de communautés	126
4.4	Intégration d'une temporalité réactionnelle	128
4.4.0.1	Démarche suivie	129
4.4.1	Confrontation des cas identifiés par Cabalar et collègues aux besoins des expérimentalistes	130
4.4.2	Application de <i>telingo</i> sur des organismes didactiques	131
4.5	Application à la communauté fromagère du chapitre 2	135
4.5.1	Cas de la planification	136
4.5.2	Cas d'explication	137
4.6	Discussion sur l'intégration de la temporalité aux modèles discrets	138
5	Conclusion et perspectives	139
	Bilan des contributions	139
	Discussion	141
	Perspectives	146
	Références	147
	Annexe A Modélisation numérique du métabolisme de 3 bactéries du fro- mage	164
	A.1 Composition du milieu extracellulaire utilisé pour les simulations	164
	Annexe B Modélisation logique du métabolisme des communautés bacté- riennes	166
	B1 Taxonomie des génomes utilisés pour reconstruire les communautés simulées	166
	B2 Potentiels d'interactions graines spécifique	168
	B3 Plus value d'ajout ou de retirer une espèce dans le milieu graine spécifique .	170
	.1 Milieux générique et spécifique utilisés pour les simulations	171
	Annexe C Contributions scientifiques	173
	C.1 Articles scientifiques	173
	C.2 Posters	173
	C.3 Présentations orales	173
	Liste des acronymes	174
	Liste des Figures	177
	Liste des Tables	178

Résumé

Titre : Approche hybride de modélisation explicable du métabolisme des écosystèmes microbiens

Résumé : Les communautés microbiennes sont des systèmes complexes composés de diverses espèces de micro-organismes interagissant entre elles et avec leur environnement. La biologie des systèmes offre un cadre pour leur étude, alliant expérimentation, génération de données à haut-débit et leur intégration dans des modèles informatiques. La compréhension de ces communautés passe notamment par celle de leur métabolisme et des échanges de molécules entre espèces susceptibles d'impacter positivement ou négativement chacun des membres. Le métabolisme est un ensemble de réactions biochimiques et peut s'abstraire à l'échelle d'un génome par des réseaux faisant le lien entre les gènes et les réactions d'un organisme. Ces réseaux permettent de construire des modèles métaboliques, représentations informatiques ou mathématiques du comportement des organismes dans des conditions expérimentales. Le passage de l'individu à la communauté, composée de quelques espèces en conditions contrôlées, ou de plusieurs centaines en conditions environnementales, soulève des difficultés méthodologiques dans la construction des modèles. Ce manuscrit de thèse traite de la modélisation du métabolisme et des interactions métaboliques au sein des écosystèmes microbiens en mettant l'accent sur l'explication des mécanismes cellulaires qui justifient les interactions bactériennes. Des solutions numériques sont majoritairement utilisées - assurant la précision des résultats - mais sont confrontées à l'importante combinatoire engendrée par les interactions bactériennes pour des communautés de grande taille. Les réponses apportées par les approches discrètes surmontent la problématique du passage à l'échelle mais sont limitées à une analyse par paire d'organismes. Afin d'identifier un potentiel ajustement méthodologique - conciliant les avantages des deux démarches, *i.e.* trouver une approche hybride - une première contribution se focalise sur le développement d'un modèle numérique dynamique et précis d'une communauté fromagère composée de trois souches. Notre stratégie itérative a permis l'intégration de données hétérogènes au moyen d'étapes de raffinement et de calibration dynamique. Ces allers-retours entre la connaissance et le modèle ont assuré la bonne prédiction des concentrations des métabolites dosés en métabolomique ainsi que des densités bactériennes au cours de la cinétique de fabrication du fromage. Dans une seconde contribution, nous proposons un modèle par raisonnement permettant de cibler des potentiels de coopération et de compétition dans des communautés bactériennes. Ce modèle repose sur l'inférence de règles logiques inférées de la biologie pour évaluer et comparer les potentiels d'interaction de communautés. Des potentiels d'interaction spécifiques à des écosystèmes ont été révélés ainsi que la pertinence de son utilisation grâce à sa rapidité d'exécution. Enfin, la troisième contribution est une réflexion portant sur l'enrichissement du modèle logique. Nous proposons un prototype s'appuyant sur l'inférence de règles logiques et permettant de (i) sélectionner la meilleure communauté à partir de contraintes biologiques et

(ii) d'apporter une notion temporelle, pouvant influencer les potentiels d'interactions. Par cette thèse, nous avons montré que la construction d'un modèle de modélisation hybride du métabolisme n'est pas nécessaire, mais qu'une approche hybride, utilisant des modèles numériques, pour des communautés de petite taille et des modèles discrets, pour analyser rapidement les communautés de taille réelle semble être pertinente.

Mots clés : représentation des connaissances et raisonnement, modélisation numérique, réseaux métaboliques, communautés microbiennes, modélisation du métabolisme

Laboratoire d'accueil : Centre Inria de l'université de Bordeaux, 200 av. de la Vieille Tour, 33405 Talence.

Title : Hybrid approach for explainable metabolic modelling of microbial ecosystems'

Abstract : Microbial communities are complex systems composed of various species of microorganisms interacting with each other and with their environment. Systems biology offers a framework for their study, combining experimentation, high-throughput data generation and the integration of the latter into computer models. Understanding these ecosystems requires the analysis of their metabolism and of the exchange of molecules between members, which can have positive or negative impacts on each other. Metabolism is a set of biochemical reactions that can be represented as a genome-scale metabolic network through the association of genes and reactions in an organism. These networks can be used to build metabolic models, mathematical representations of organism behavior under given environmental conditions. Scaling up to the analysis of a community composed of a few species under controlled conditions, or several hundreds in natural environments, raises methodological difficulties in the construction of models. This thesis manuscript deals with the construction of computational models for the analysis of metabolism and metabolic interactions in microbial ecosystems, with a particular attention on explaining the cellular mechanisms underlying bacterial interactions. Numerical solutions are mainly used - ensuring the accuracy of results - but they face combinatorial issues generated by bacterial interactions in large-scale communities. Discrete approaches overcome this problem, but are limited to pairwise analysis. In order to identify a putative methodological tradeoff reconciling the advantages of both approaches, i.e. finding a hybrid approach, a first contribution focuses on the development of a dynamic and accurate numerical model of a cheese bacterial community composed of three strains. Our iterative strategy enables the integration of heterogeneous data through refinement and dynamic calibration steps. This back-and-forth between knowledge and model ensures the accurate prediction of metabolite concentrations and bacterial densities during cheese production. We propose as a second contribution a reasoning-based model for deciphering cooperative and competitive potentials in bacterial communities. This model relies on the inference of logical rules motivated by biology in order to evaluate and compare community interaction potentials. Ecosystem-specific interactions potentials are retrieved, and the fast execution of the reasoning-based approach facilitates the screening of collections of communities. Finally, the third contribution is a reflection on the enrichment of the logic model. We propose a prototype based on the inference of logical rules, enabling (i) the selection of the best community based on biological constraints and (ii) the inference of a temporal notion, which can impact interaction potentials. Through this thesis, we demonstrate that the construction of a hybrid model of metabolism is not required but that a hybrid approach, using numerical models for small communities and discrete models for rapid analysis of full-size communities, seems to be relevant.

Keywords : knowledge representation and reasoning, numerical modelling, metabolic networks, microbial communities, metabolic modelling

Remerciements

Je remercie Caroline Barouck et Muriel Cocaign-Bousquet d'avoir accepté d'être rapportrices de mon manuscrit de thèse. Je remercie également Ludovic Cottret, Béatrice Laroche, Gabriel Markov, Laurent Simon d'avoir accepté de participer à mon jury de soutenance.

Je souhaite avant tout remercier mon directeur, ma co-directrice et mon encadrante de thèse, David Sherman, Hélène Falentin et Clémence Frioux, qui m'ont guidé et soutenu tout au long de ce périple.

David, je te remercie, à travers ton expérience de directeur, pour les mots que tu as su trouver dans les périodes difficiles que j'ai traversées, me redonnant confiance en moi. Tes conseils, tes décisions et tes remarques m'ont permis de ne pas céder sous les moments de rush ainsi que, bien entendu, tes blagues lors des repas.

Hélène, je te remercie pour ton optimisme, ton enthousiasme à toute épreuve et ton énergie, qui ont largement contribué à nos échanges fructueux. De plus, lors des parties plus difficiles pour toi, tes remarques et questions pertinentes ont su marquer l'importance de l'interdisciplinarité de cette thèse.

Clémence, je te remercie pour tout. Toi qui m'as convaincu de faire une thèse, alors que je n'étais qu'un étudiant de master, naïf et résigné à ne pas en faire une. Toi qui as été disponible et patiente pendant ces trois années. Toi qui m'as fait grandir scientifiquement, grâce tes remarques et à nos nombreux échanges. Je voudrais également te remercier pour tes longues relectures et remarques minutieuses et pertinentes sur les papiers et le manuscrit. Enfin, aujourd'hui, je ne suis plus aussi naïf, mais je suis convaincu d'une chose, c'est de la chance que les futurs doctorants et doctorantes auront de t'avoir en tant que directrice de thèse.

Je tenais aussi à remercier Simon Labarthe, qui a été impliqué dans la première partie de ma thèse. Simon, tu as su me transmettre ta motivation pour les mathématiques appliquées, tes convictions scientifiques ainsi que ta gaieté au quotidien lors de notre collaboration.

Je souhaite par ailleurs remercier toutes les personnes qui ont contribué à cette thèse : l'école doctorale mathématiques et informatique ainsi que l'équipe PLEIADE du centre Inria de l'université de bordeaux et MICROBIO de l'UMR STLO Rennes.

Je remercie également les collègues devenus des amis. En commençant par Coralie, notre ingénieure de l'équipe, mais aussi l'organisatrice officielle d'événements. Chabname,

pour son éternel premier degré dévastateur, grumpy Pablo alias le chercheur migrateur. Mais également, Alexis (monsieur fanfare), Guillaume (guigui), Myriam (Sexy mimi), Gala, Elsa (la Mamie), Vincent, Raluca (Ralou), Isabeau (Covid-woman), Arie, Rafael et à tous les autres. Merci à tous.

Enfin, je remercie mes amis qui m'ont soutenu pendant cette épreuve : Lucas (Mon poney), Marion (Maillon), JC, Manon (la naine), JP (4ème pointu), Cams (Excessifix), Nico (Monsieur le papa), Maé, Méline, Célia, Valentine, Yannou et Cléa. Et surtout ma famille, qui a cru en moi jusqu'au bout : ma mère, mon père et ma sœur.

Introduction

Sommaire

0.1	Émergence de la microbiologie <i>in silico</i>	11
0.2	La biologie des systèmes comme support de l'étude du mé- tabolisme	12
0.3	Objectifs de la thèse	12

0.1 Émergence de la microbiologie *in silico*

La compréhension des mécanismes biologiques inhérents à un micro-organisme ou un écosystème microbien ne passe plus uniquement par des expérimentations à la paillasse mais aussi par des traitements de données et simulations informatiques. Au commencement de la microbiologie, les scientifiques se concentraient sur les maladies infectieuses, affectant à la fois l'humain et les animaux. Le rôle des bactéries dans certaines pathologies, mais aussi dans des processus de bio-transformation, comme la fermentation, a ensuite commencé à être élucidé. Aujourd'hui, aller plus loin dans l'étude de ces systèmes biologiques implique la production de données à haut-débit dites **omiques**. La première d'entre elles, permettant de caractériser le vivant à l'échelle moléculaire, et la plus utilisée, est la génomique. Les molécules d'ADN sont séquencées, générant des lectures et motivant le développement d'algorithmes informatiques d'assemblage de ces lectures en contigs. D'autres données omiques portant sur l'analyse des ARN (transcriptomique), des protéines (protéomique) et de la biochimie (métabolomique) ont émergé, accélérant le développement en parallèle de **modèles informatiques**. Ces derniers représentent de façon simplifiée un processus ou une réalité ; ils sont construits et alimentés par des données, ici les données omiques, et proposent des sorties dont l'objectif est de contribuer à répondre à des questions biologiques.

Au-delà de l'étude des bactéries cultivables, la microbiologie s'attelle à étudier des écosystèmes microbiens, aussi appelés microbiotes, composés de plusieurs centaines ou milliers d'espèces différentes. Difficilement contrôlables expérimentalement, composés d'espèces parfois inconnues, ces ensembles sont auditables par la génération de données omiques à haut-débit. Les modèles informatiques sont alors essentiels pour l'intégration de ces données massives et hétérogènes, et surtout pour générer des hypothèses testables sur le fonctionnement des écosystèmes. Au sein de ces écosystèmes microbiens, il existe des mécanismes complexes d'interactions nécessitant l'analyse du système dans sa globalité et rendant insuffisante l'analyse des membres individuels. Ces interactions ont des effets positifs, négatifs ou neutres sur les organismes qui composent la communauté (Faust and Raes, 2012). Leur rôle est d'importance et varie selon l'écosystème : protection contre des pathogènes de l'intestin (Zhang et al., 2015), recyclage des matières organiques dans l'écosystème du sol (Hoorman, 2011; Yadav et al., 2018) ou encore, libération de composés d'arômes (McSweeney and Sousa, 2000) responsables de la qualité du vin par exemple (Virdis et al., 2021). Révéler ces interactions en laboratoire est possible en élaborant des co-cultures par paires ou petits groupes d'espèces (Weiss et al., 2022) mais devient complexe pour des communautés de grande taille. Ainsi, il existe un réel besoin de créer des modèles informatiques permettant de pallier aux limitations expérimentales et proposer des hypothèses pour expliquer les interactions.

En somme, le développement de modèles informatiques pour la microbiologie est motivé par les données omiques, permettant de les analyser, mais également de générer de nouvelles hypothèses, testables en laboratoire, pour potentiellement influencer sur les systèmes biologiques. Lier les modèles aux données n'est pas un processus trivial, la **biologie des systèmes** est au cœur de ces problématiques.

0.2 La biologie des systèmes comme support de l'étude du métabolisme

La biologie des systèmes associe un organisme à un système et consiste à étudier ce dernier dans son ensemble, et non comme un assemblage de gènes et de protéines ou autres sous-systèmes individuels (Kitano, 2002). La biologie des systèmes englobe quatre aspects principaux : les données, essentiellement *omiques*, qui constituent la connaissance biologique ; la partie *computationnelle*, qui construit des modèles informatiques pour effectuer des simulations selon les hypothèses des biologistes et les données disponibles ; l'*analyse* de ces simulations, permettant de formuler des hypothèses et des prédictions sur le système ; et enfin une dimension *technologique*, qui consiste à établir des protocoles expérimentaux générant les données initiales et de vérifier les hypothèses du modèle. Tout comme la microbiologie a évolué vers l'étude des microbiotes, la biologie des systèmes est elle aussi passée à l'échelle de l'écosystème : le système d'étude comprend alors l'ensemble des membres de l'écosystème (Odum, 1983).

Une question concrète reste l'identification de méthodes pour caractériser les interactions bactériennes dans le cadre posé par la biologie des systèmes. Nous savons que les interactions métaboliques entre bactéries ou avec leur hôte consistent en l'échange de molécules impactant ou non le phénotype de l'organisme receveur (Faust and Raes, 2012). Or, la biologie des systèmes concentre son étude en intégrant plusieurs niveaux que sont les gènes, les protéines ou encore les métabolites. Dans la littérature, un réseau métabolique, qui est une abstraction informatique de l'ensemble des réactions biochimiques pouvant être catalysées par un organisme, c'est-à-dire son **métabolisme**, a comme concept initial des associations gène-protéine-réaction (GPR) (Thiele and Palsson, 2010). Ainsi, modéliser le métabolisme à l'aide de réseaux construits à l'échelle du génome (GEMs) répond aux contraintes de la biologie des systèmes et est une piste pour mieux caractériser les communautés bactériennes. A l'échelle d'un organisme, d'une espèce, il existe de nombreuses méthodes informatiques et mathématiques permettant l'analyse métabolique, comme par exemple des analyses numériques basées sur des flux (Orth et al., 2010) ou encore sur la topologie des réseaux (Kuffner et al., 2000). Si des challenges méthodologiques persistent encore sur la modélisation à l'échelle d'une espèce, ils sont bien plus nombreux à l'échelle de la communauté, qui génère des nouvelles problématiques de modélisation, et de gestion des coûts calculatoires et de la combinatoire des interactions microbiennes possibles. À ces difficultés s'ajoutent la moindre disponibilité de données et la plus faible connaissance biologique des acteurs individuels en présence. Pourtant, l'étude à haut débit des communautés bactériennes *in silico* est nécessaire pour proposer de nouvelles hypothèses biologiques dans le cercle vertueux de la biologie des systèmes. Cela nécessite le développement de nouvelles méthodes informatiques et mathématiques, discrètes ou numériques, mettant en avant les interactions bactériennes *via* l'étude du métabolisme.

0.3 Objectifs de la thèse

Du point de vue de la modélisation, le défi principal auquel tente de répondre cette thèse est l'identification d'un modèle avec le meilleur compromis entre : précision des résultats, versatilité, temps de calcul et pouvoir d'explicabilité. À cette problématique s'ajoute le nombre de mécanismes métaboliques que l'on souhaite modéliser qui augmente avec le nombre d'organismes que le système compte. Ainsi, analyser des communautés naturelles rapidement, avec une parfaite compréhension de l'ensemble du métabolisme de

chaque individu au sein de la communauté permettant la prédiction de nouveaux processus d'interactions, est utopique, et la simplification du système par un modèle pertinent est d'importance. Dans cette thèse, nous avons identifié deux grands formalismes permettant d'analyser des communautés microbiennes et de mettre en avant des interactions. Il existe des méthodes de modélisation numérique du métabolisme, apportant la précision requise, mais pouvant être limitées par la disponibilité des données, le besoin de connaissance biologique ainsi que par le temps de calcul. Ces limites sont contournées par des approches discrètes, dégradant la précision numérique mais apportant la capacité d'analyser des communautés naturelles rapidement. Ainsi, l'objectif principal de la thèse est de trouver un compromis entre ces méthodes numériques et discrètes en développant une approche hybride de modélisation explicable du métabolisme des communautés microbiennes. Les chapitres 2, 3, et 4 présentent les apports de la thèse pour répondre à cet objectif.

Le chapitre 1 est une revue de la littérature concernant les différents moyens de représenter le métabolisme ainsi que les méthodes d'analyses existantes.

Le chapitre 2 décrit la création d'un modèle numérique explicable du métabolisme d'une petite communauté à partir de données hétérogènes. Pour cela, nous avons à notre disposition trois souches modèles utilisées dans la production de fromage : *Lactococcus lactis*, *Lactiplantibacillus plantarum* et *Propionibacterium freudenreichii*, ainsi que des données de croissances bactériennes, de concentrations de métabolites dans des conditions de cultures pures et de co-culture. Ces données hétérogènes, couplées à une stratégie itérative basée sur l'analyse de l'équilibre des flux (Orth et al., 2010), permettent d'obtenir des résultats conformes aux données expérimentales. Ce modèle numérique permet aussi de générer des hypothèses testables sur l'expression des voies métaboliques ainsi que sur les interactions bactériennes mises en jeu durant la fabrication du fromage.

Le chapitre 3 présente une méthode pour pseudo-quantifier des interactions bactériennes et comparer des communautés de grande taille entre elles. À l'aide d'une approche par raisonnement, nous avons établi des règles logiques servant à calculer le potentiel de coopération et de compétition de communautés naturelles. Afin de tester le passage à l'échelle, nous utiliserons des génomes microbiens provenant de divers écosystèmes pour créer des communautés synthétiques de taille variable. Par ailleurs, nous comparerons cette approche par raisonnement aux modèles numériques existants.

Le chapitre 4 est une réflexion sur les pistes d'amélioration des modèles discrets en vue d'augmenter la précision qu'ils apportent au regard des apports des deux précédents chapitres. Nous présenterons deux prototypes permettant d'affiner les sorties du modèle discret pour la sélection "à façon" de communautés bactériennes et d'intégrer une dimension de temporalité dans les approches basées sur le raisonnement.

Enfin, le chapitre 5 présente les conclusions, discussions et perspectives de la thèse.

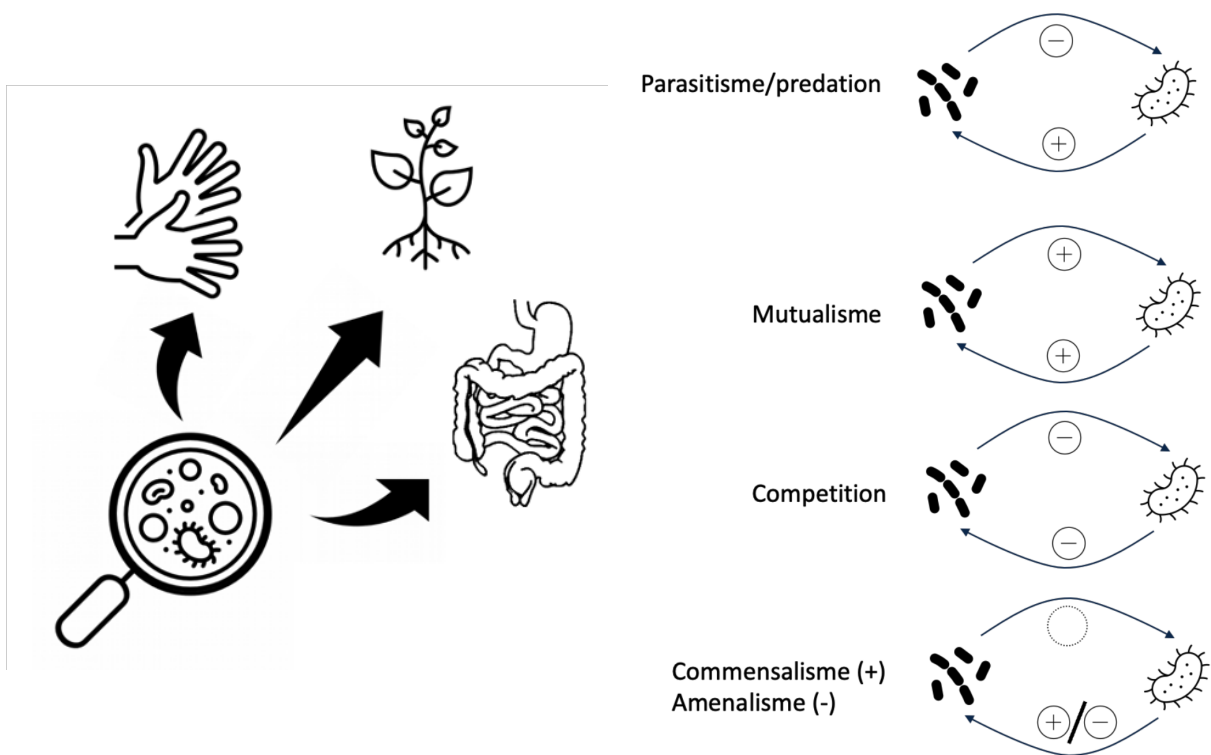


FIGURE 1 – **Présence de bactéries dans divers écosystèmes et illustrations des interactions bactériennes.** Un + (respectivement -) signifie une interaction positive (respectivement négative) entre les populations microbiennes. Un cercle en pointillé représente une interaction sans effet.

Chapitre 1

Modélisation du métabolisme : d'une cellule à un écosystème

Sommaire

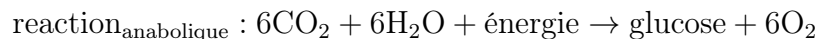
1.1 Les modèles du métabolisme comme support de l'analyse microbienne	16
1.1.1 Reconstruction de réseaux métaboliques	16
1.1.2 Représentations mathématiques et informatiques du métabolisme	20
1.1.3 Analyse structurale de réseaux métaboliques	21
1.2 Modélisation métabolique à l'échelle d'un génome	24
1.2.1 Méthodes numériques d'un réseau métabolique	24
1.2.2 Méthodes qualitatives d'analyse d'un réseau métabolique	28
1.2.3 Défis rencontrés lors de la modélisation qualitative et numérique d'une cellule	30
1.2.4 Méthodes hybrides de modélisation à l'échelle d'un modèle métabolique	30
1.3 Modélisation du métabolisme des écosystèmes microbiens . .	31
1.3.1 Méthodes numériques appliquées à plusieurs réseaux métaboliques	32
1.3.2 Applications	36
1.3.3 Méthodes et modèles qualitatifs de l'étude d'une communauté bactérienne	37
1.3.4 Défis rencontrés lors de l'analyse de la modélisation qualitative et quantitative du métabolisme des communautés	40

1.1 Les modèles du métabolisme comme support de l'analyse microbienne

1.1.1 Reconstruction de réseaux métaboliques

Le *métabolisme* décrit toutes les réactions biochimiques au sein d'un organisme permettant de produire de l'énergie, de dégrader et de produire du matériel métabolique (Sánchez López de Nava A, 2023). Un moyen de le représenter est d'utiliser le *réseau métabolique*. Un réseau métabolique décrit un ensemble de voies métaboliques et est composé d'une succession de réactions biochimiques, ainsi que des interconnexions entre ces voies, mises en avant par des métabolites en communs (Goldford and Segrè, 2018). Le réseau métabolique est souvent utilisé par les biologistes, leur permettant principalement d'identifier les voies métaboliques, caractérisées par un ensemble de réactions chimiques catalysées par des enzymes. Pour effectuer des simulations numériques ou étudier la topologie du réseau, le réseau métabolique est utilisé dans des modèles mathématiques ou informatiques dits *modèles métaboliques*.

Afin d'illustrer ce que sont que les réactions métaboliques, nous allons prendre l'exemple d'un processus connu : la photosynthèse. Les plantes utilisent l'énergie du soleil, *i.e.* la lumière, pour la convertir en sucre, comme le glucose, et en dioxygène (O_2), par la réaction anabolique suivante :



La réaction est dite anabolique lorsqu'elle produit de grosses molécules à partir de plus petites en consommant de l'énergie, sous forme de molécules d'ATP par exemple. Les réactions cataboliques sont responsables du processus inverse, et libèrent de l'énergie. Un exemple de réaction catabolique est celle représentant le processus de respiration :



Ainsi, les voies métaboliques sont constituées de multiples réactions anaboliques et cataboliques, comme illustrées par la figure 1.1 qui représente l'ensemble des voies métaboliques primaires de la bactérie *Escherichia coli*.

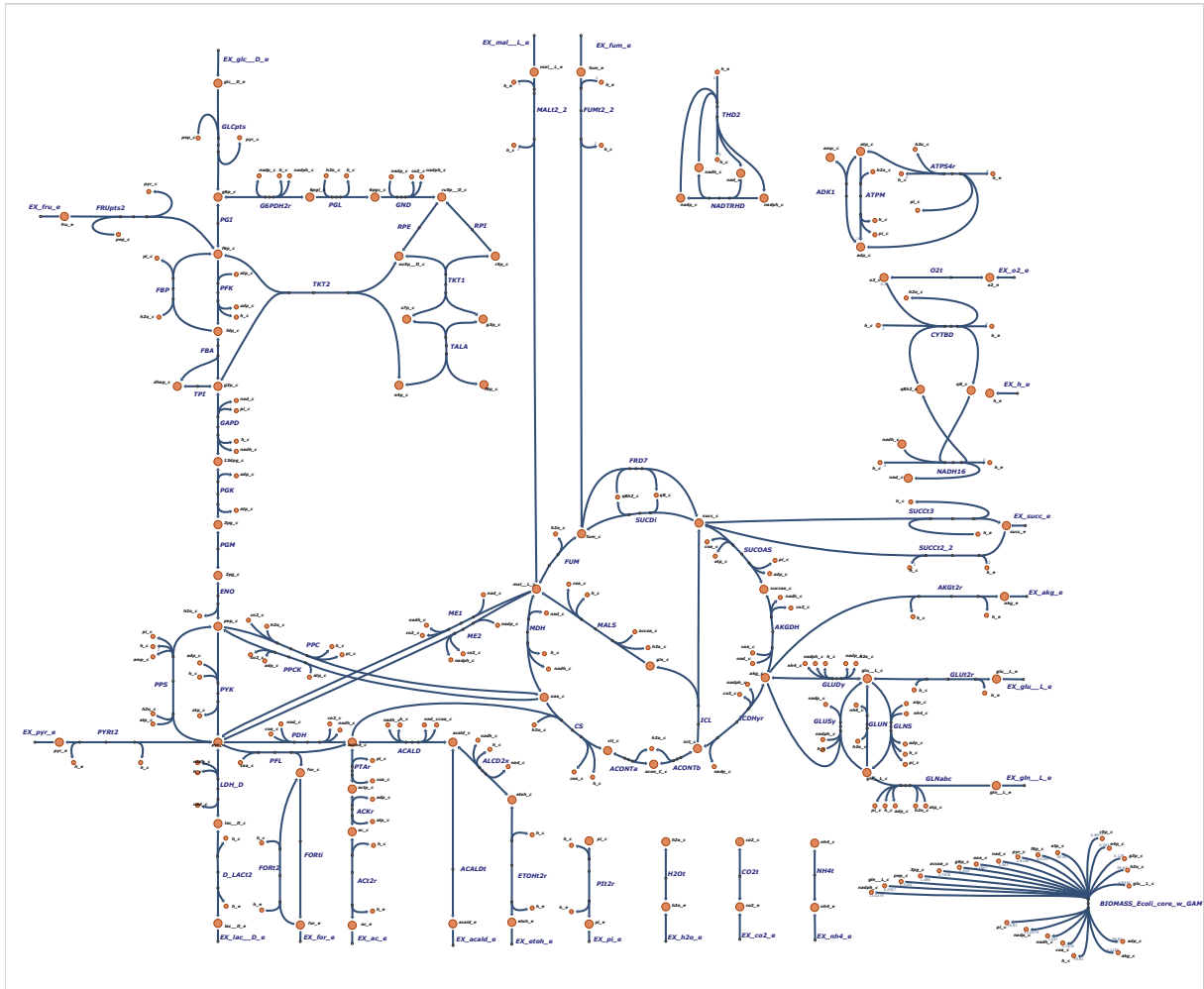


FIGURE 1.1 – Illustration d’un ensemble de voies métaboliques de *E. coli* dans l’application web Escher (King et al., 2015)

Le plus petit réseau métabolique sur lequel une analyse peut être faite est constitué d’une réaction de production et d’une réaction de consommation. Chez la bactérie, on retrouve des réseaux métaboliques allant de quelques réactions à plus de deux mille dans le cas des réseaux métaboliques à l’échelle du génome, et peut atteindre plus de dix mille réactions dans le cas du réseau métabolique humain (Robinson et al., 2020).

Un réseau métabolique se caractérise par l’association de gènes à des protéines qui catalysent des réactions (GPR) Thiele and Palsson (2010). Ces associations sont décrites par des règles d’association booléennes entre les gènes et les réactions. Considérons les réactions $r1$, $r2$ associées aux gènes $g1$, $g2$, $g3$ par les relations logiques suivantes :

$$r1 : (g1 \wedge g2) \vee (g1 \wedge g3) \quad (1.1)$$

$$r2 : g1 \wedge (g2 \vee g3) \quad (1.2)$$

La règle booléenne 1.1 stipule que, pour que la réaction $r1$ puisse être catalysée, il faut que le complexe $g1$ et $g2$ soit exprimé ou que le complexe $g1$ et $g3$ soit exprimé. Dans le cas de la règle d’association 1.2, la réaction $r2$ est activée si $g1$ est exprimé avec $g2$ ou $g3$. Dans ce cas précis, si seul $g2$ ou $g3$ est exprimé mais que $g1$ ne l’est pas, alors la réaction

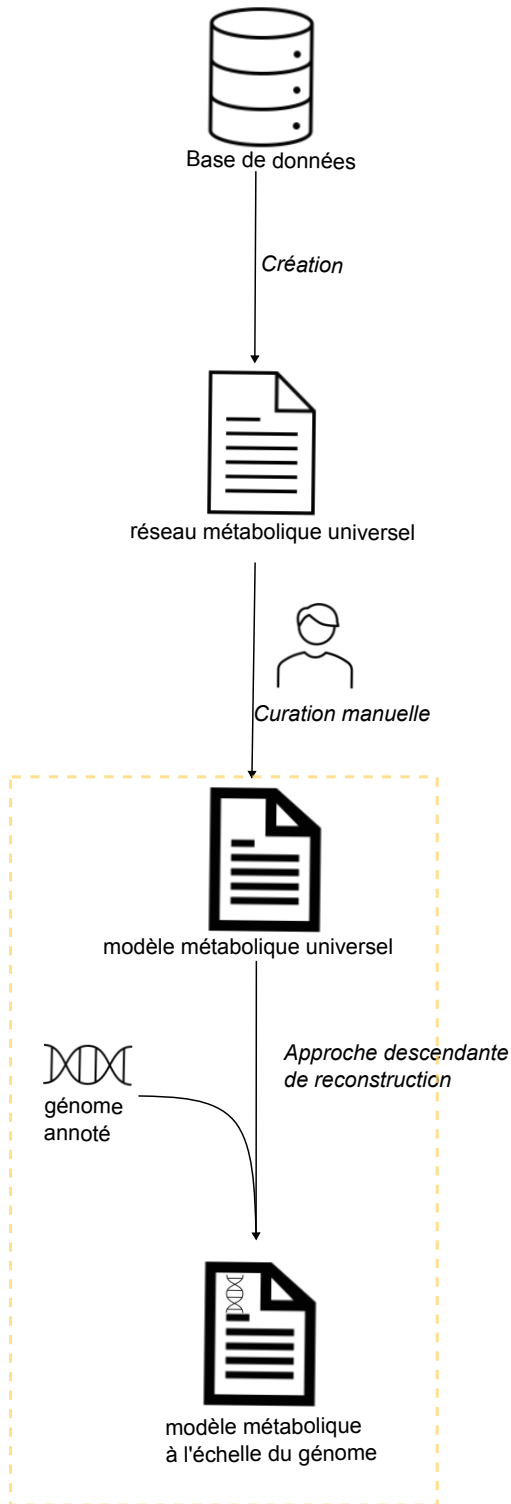
r_2 ne pourra pas être catalysée. Des associations plus complexes et non-triviales existent cependant, telles que décrites dans (Issa and Sherman, 2018), où une représentation logique du système 3-méthyl-2-oxopentanoate est formulée (équation 34).

Afin de reconstruire le réseau métabolique, un génome annoté est nécessaire. À partir d’une séquence génomique d’un organisme, deux annotations du génome se distinguent (Médigue et al., 2002) : structurale, permettant d’identifier les gènes (par exemple, avec les outils AUGUSTUS (Nachtweide and Stanke, 2019) et Glimmer (Delcher et al., 2007)), et fonctionnelle, pour retrouver la biochimie et les fonctions biologiques portées par une protéine (par exemple, Blast (Altschul et al., 1990), par un processus d’alignement). Un grand panel de bases de données existe permettant une annotation à plusieurs échelles du génome. Par exemple, BRENDA (Chang et al., 2021) se concentre sur les enzymes et les données liées au métabolisme, et EggNOG fournit une annotation fonctionnelle basée sur des relations d’orthologie Huerta-Cepas et al. (2019). Des outils utilisant les différentes bases de données existent comme par exemple EggNOG-mapper (Cantalapiedra et al., 2021), RAST (Aziz et al., 2008), Prokka (Seemann, 2014) ou encore MicroScope (Vallenet et al., 2020).

Le génome étant préalablement annoté, le réseau métabolique peut-être reconstruit manuellement et/ou automatiquement avec une approche ascendante (bottom-up) et/ou une approche descendante (top-down), en utilisant une large collection d’outils (Mendoza et al., 2019) (Figure 1.2). L’approche ascendante crée une ébauche de réseau métabolique à partir d’une séquence génomique, alors que l’approche descendante dérive le modèle métabolique à l’échelle d’un génome à partir d’un modèle universel déjà curé. Tous les outils de reconstruction utilisent une base de données leur permettant d’assigner une fonction métabolique portée par un gène. Les approches bottom-up se caractérisent par un processus itératif de génération d’ébauches de réseaux métaboliques, suivi d’étapes de curation, tandis que les approches descendantes sont définies par la production de réseaux métaboliques “prêts à l’emploi”. Dans ce deuxième cas, les fonctions assignées dans les modèles métaboliques gardent la curation d’un modèle universel fonctionnel, leur assurant ainsi une bonne qualité due à leur faible nombre d’assignation de faux-positif (Machado et al., 2018).

Parmi les deux approches, on retrouve GEMSiRV (Liao et al., 2012), CarveMe (Machado et al., 2018) et MetaDraft (Olivier, 2018) qui représentent les outils utilisant l’approche descendante. Les outils CarveMe et MetaDraft se servent de la base de données BIGG (King et al., 2016) pour construire leurs modèles universels. Une plus grande bibliographie existe pour l’approche ascendante avec FAME (Boele and Olivier, 2012) une application web, CoReCo (Pitkänen et al., 2014), Merlin (Dias et al., 2015), RAVEN (Wang et al., 2018) et autoKEGGRec (Karlsen et al., 2018) qui s’appuient sur les bases de données KEGG (Kanehisa and Goto, 2000) et/ou MetaCyc (Caspi et al., 2018). Pathway-tools (Karp et al., 2022) ou encore AuReMe (Aite et al., 2018), sont tous deux basés sur la base de données MetaCyc (Caspi et al., 2018). Nous pouvons également citer les outils ModelSEED (Seaver et al., 2021), Kbase (Arkin et al., 2018) ou encore gapseq (Zimmermann et al., 2021). L’outil rBioNet (Thorleifsson and Thiele, 2011), allie à la fois la reconstruction descendante et ascendante, permettant à l’utilisateur ou l’utilisatrice de créer soi-même son réseau métabolique ou d’en importer un sur lequel il ou elle peut se baser pour le reconstruire.

(a) Approche descendante de reconstruction de modèles métaboliques



(b) Approche ascendante de reconstruction de modèles métaboliques

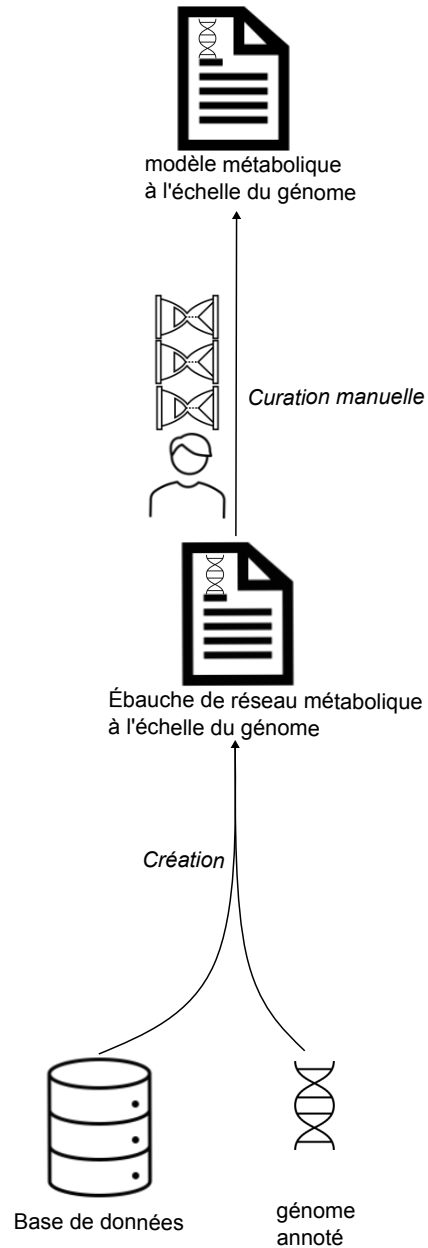


FIGURE 1.2 – Illustration des approches descendante (a) et ascendante (b) pour reconstruire des modèles métaboliques à l'échelle du génome fonctionnels. En (a), la partie de curation manuelle n'est pas prise en compte dans l'approche de reconstruction, ce qui n'est pas le cas de l'approche ascendante (b).

Diffusion et qualité des réseaux métaboliques

Il existe une grande diversité d'outils permettant de reconstruire le métabolisme, mais la question de la standardisation de sa représentation perdure (Stobbe et al., 2014). En effet, des formats hétérogènes sont associés aux différentes bases de données : BioCyc (Karp et al., 2018) utilise des fichiers PGDB (Pathway Genome DataBase), le format BioPAX (Demir et al., 2010) ou encore SBML (Hucka et al., 2019). La base de données KEGG (Kanehisa and Goto, 2000) utilise d'autres fichiers et sa propre façon de représenter le métabolisme : KEGG Markup Language. Même si le choix syntaxique utilisé n'est pas commun à tous, nous retrouvons majoritairement le format *Systems Biology Markup Language* (SBML) (Hucka et al., 2019), qui permet de représenter les voies de signalisation mais également, des paramètres mathématiques utiles pour la modélisation.

Des efforts motivés par la communauté scientifique ont été faits concernant l'évaluation de la qualité du réseau métabolique. MEMOTE (Lieven et al., 2020) évalue la cohérence du réseau métabolique en terme du nombre de gènes associés à des réactions, le nombre de réactions non associées à un gène ou encore le nombre de métabolites qui ne sont jamais consommés mais produits possiblement accumulés dans le système. Le score évaluant la qualité du réseau est borné par la méthode de reconstruction choisie (Lieven et al., 2020).

1.1.2 Représentations mathématiques et informatiques du métabolisme

Le métabolisme peut se représenter de deux grandes façons : sous forme d'un hypergraphe dirigé simple ou bipartite, ou sous forme d'une matrice stœchiométrique (Barabási and Oltvai, 2004; Belcour et al., 2020). La représentation sous forme d'un graphe informe sur les liens entre les réactions et les métabolites d'un système biologique plutôt que sur la quantité de métabolites produits et/ou consommés, ces dernières étant au cœur de la matrice stœchiométrique. Nous allons prendre l'exemple d'une voie de production connue, la glycolyse, chez une espèce lactique : *Lactiplantibacillus plantarum* pour illustrer ces deux représentations.

La glycolyse chez *Lactiplantibacillus plantarum* est composée des réactions suivantes :

- (HEX1) : 1 glucose \rightarrow 1 glucose-6-phosphate
- (PGI) : 1 glucose-6-phosphate \leftrightarrow 1 fructose-6-phosphate
- (PFK) : 1 fructose-6-phosphate \rightarrow 1 D-Fructose 1,6-bisphosphate
- (FBA) : 1 D-Fructose 1,6-bisphosphate \leftrightarrow 1 Dihydroxyacetone phosphate
+1 Glyceraldehyde 3-phosphate
- (GAPD) : 1 Glyceraldehyde 3-phosphate \leftrightarrow 1 3-Phospho-D-glyceroyl phosphate
- (PGK) : 1 3-Phospho-D-glycerate \leftrightarrow 1 3-Phospho-D-glyceroyl phosphate
- (PGM) : 1 3-Phospho-D-glycerate \leftrightarrow 1 2-Phospho-D-glycerate
- (ENO) : 1 2-Phospho-D-glycerate \leftrightarrow 1 Phosphoenolpyruvate
- (PYK) : 1 Phosphoenolpyruvate \rightarrow 1 pyruvate.

HEX1, PGI, PFK, FBA, GAPD, PGK, PGM, ENO, PYK sont les réactions métaboliques. Glucose, glucose-6-phosphate, fructose-6-phosphate, D-Fructose 1,6-bisphosphate, Dihydroxyacetone phosphate, Glyceraldehyde 3-phosphate, 3-Phospho-D-glyceroyl phosphate, 3-Phospho-D-glycerate, 2-Phospho-D-glycerate, Phosphoenolpyruvate et pyruvate sont les métabolites principaux, substrats et produits, impliqués dans la glycolyse de *Lactiplantibacillus plantarum*.

1.1.2.1 Matrice stœchiométrique

Le métabolisme peut être représenté sous forme d’une matrice stœchiométrique dans laquelle l’ensemble des réactions et des métabolites du système constituent respectivement les colonnes et les lignes. Le nombre à l’intérieur de la matrice représente le coefficient stœchiométrique de la réaction biochimique. Ainsi, la glycolyse chez *Lactiplantibacillus plantarum* est représentée par la matrice stœchiométrique suivante :

$$\begin{array}{l}
 \text{glucose} \\
 \text{glucose-6-phosphate} \\
 \text{fructose-6-phosphate} \\
 \text{D-Fructose 1,6-bisphosphate} \\
 \text{Dihydroxyacetone phosphate} \\
 \text{Glyceraldehyde 3-phosphate} \\
 \text{3-Phospho-D-glyceroyl phosphate} \\
 \text{3-Phospho-D-glycerate} \\
 \text{2-Phospho-D-glycerate} \\
 \text{Phosphoenolpyruvate} \\
 \text{pyruvate}
 \end{array}
 \begin{pmatrix}
 \text{HEX1} & \text{PGI} & \text{PFK} & \text{FBA} & \text{GAPD} & \text{PGK} & \text{PGM} & \text{ENO} & \text{PYK} \\
 -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
 \end{pmatrix}$$

1.1.2.2 Graphe

Pour représenter des réseaux métaboliques, un graphe bipartite dirigé pondéré est souvent privilégié (Barabási and Oltvai, 2004). Le graphe est alors composé de deux types de sommets, un pour les métabolites et un pour les réactions, et d’arcs symbolisant les relations “être produit par” ou “être consommé par” entre les métabolites et les réactions. Mathématiquement, un graphe bipartite G peut être défini comme un tuple $G = \langle M, R, E \rangle$ où R , M et E représentent les ensembles respectifs de réactions, de métabolites et des arcs reliant les métabolites aux réactions. Un réactant m consommé par la réaction r est défini par la relation $e = \langle m, r \rangle$, et un produit m de la réaction r est défini comme suit $e = \langle m, r \rangle$ où $m \in M$ et $r \in R$.

Une représentation de la glycolyse de *Lactiplantibacillus plantarum* sous forme d’un graphe bipartite est illustrée dans la figure 1.3, dans laquelle les sommets carrés représentent les réactions et les sommets en cercles représentent les métabolites.

1.1.3 Analyse structurelle de réseaux métaboliques

Avant d’aborder les processus de modélisation de réseaux métaboliques, une étude de la structure, *i.e.* topologie, des réseaux métaboliques peut-être faite au moyen d’algorithmes de la théorie des graphes. Dans un graphe non orienté, seule la structure du réseau peut être analysée et comparée. L’analyse de la topologie peut se faire à l’échelle des substrats uniquement (Jeong et al., 2011) ou des réactions uniquement (Wagner and Fell, 2001). À partir d’un graphe, des propriétés topologiques comme les plus courts chemins entre des sommets (Wagner and Fell, 2001), la modularité (Holme, 2011) ou la centralité (Kim et al., 2019) sont souvent utilisées en biologie des systèmes. La modularité d’un réseau est sa capacité à être décomposable en sous-graphes densément connectés mais faiblement reliés entre eux. La centralité d’un sommet est un indice qui capture l’importance de ce sommet dans le graphe. D’autres méthodes se basent sur la détection de motifs

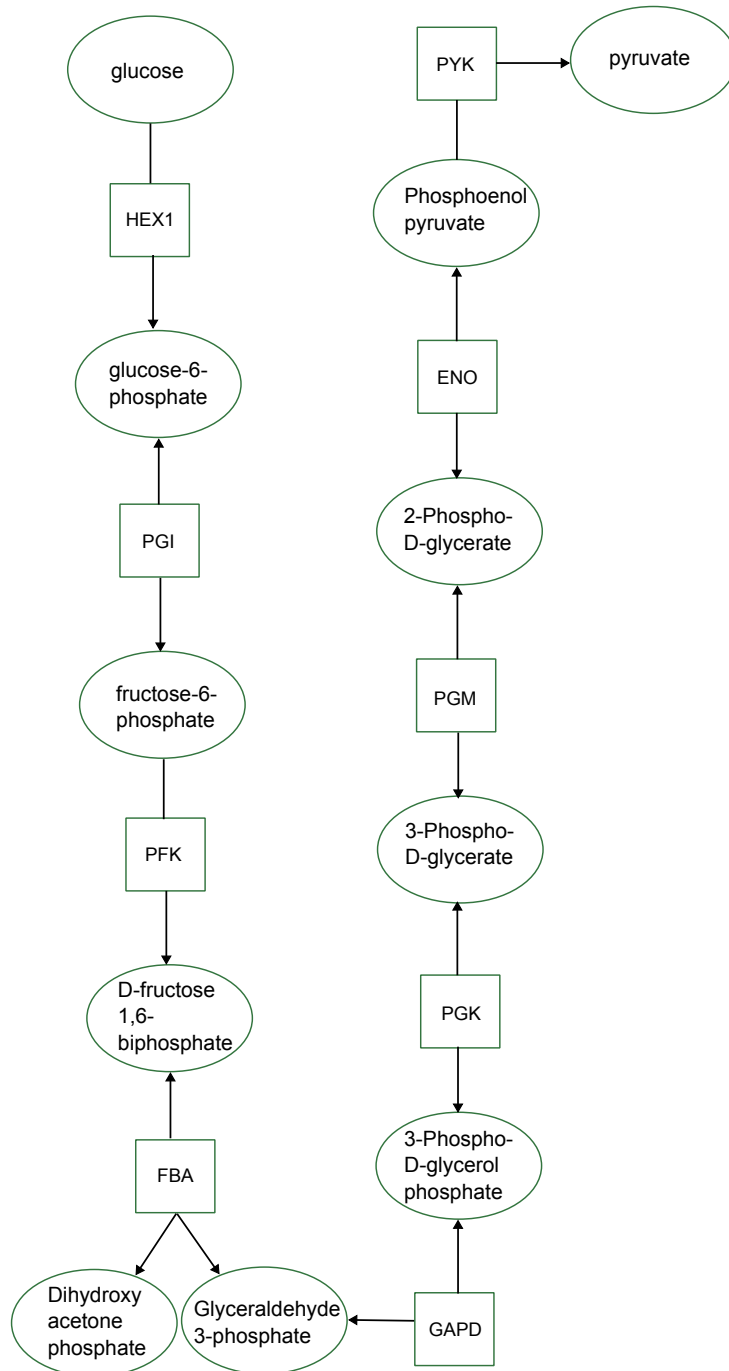


FIGURE 1.3 – Représentation du métabolisme sous forme d'un graphe bipartite selon la *Systems Biology Graphical Notation* (SBGN) (Bergmann et al., 2020). Les sommets carrés décrivent les réactions, les cercles les métabolites et les arcs représentent les relations de consommation et de production des métabolites.

où pour chaque occurrence du motif, un sous-graphe est généré (Lacroix et al., 2006). La comparaison de réseaux permet d'identifier des différences fonctionnelles, de détecter des variations entre souches d'une même espèce ou encore, de formuler des hypothèses de capacité biosynthétiques entre les organismes, sans prendre en compte les interactions entre elles (Biggs et al., 2015; Ma and Zeng, 2003). Dans (Ay et al., 2012), les auteurs et autrices utilisent une technique d'alignement de réseaux métaboliques pour identifier des motifs communs conservés entre plusieurs organismes. Dans un réseau complexe, étudier

la connectivité d'un substrat est massivement utilisé par la communauté scientifique (Ma and Zeng, 2003), cependant, les co-facteurs tels que ATP, ADP etc sont sur-représentés dans les molécules les plus centrales ce qui rend fastidieuse l'analyse par les biologistes (Sweetlove and Fernie, 2005). Néanmoins, identifier des composés fortement connectés permet de montrer des propriétés de robustesse d'un réseau métabolique face à des délétions enzymatiques. Une étude a montré que dans 91% des cas, la délétion d'une enzyme chez un organisme n'affectait pas sa viabilité (Lemke et al., 2004).

Comparer la structure de réseaux métaboliques est purement qualitatif mais permet de souligner des hypothèses sur ce que chaque réseau métabolique pourrait être capable de produire et de consommer. Pour vérifier ces hypothèses, on utilise la modélisation du métabolisme qui, à partir d'un environnement nutritionnel, vise à inférer les ensembles de réactions activées et métabolites produits. Au moyen d'approches numériques et qualitatives, des hypothèses testables en laboratoire sur différents systèmes biologiques peuvent être générées et de nouvelles connaissances peuvent être prédites. De plus, l'élaboration de tels modèles métaboliques permet l'intégration de diverses données omiques, affinant ainsi les prédictions (Passi et al., 2022).

1.2 Modélisation métabolique à l'échelle d'un génome

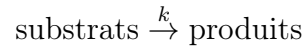
La modélisation du métabolisme permet d'effectuer des simulations en utilisant le réseau métabolique et un formalisme mathématique ou informatique adéquat. Les prédictions du modèle métabolique peuvent être affinées à l'aide de données multi-omiques. On peut trouver dans la littérature de nombreuses méthodes d'analyse du métabolisme, aussi bien à l'échelle d'une voie métabolique particulière qu'à l'échelle d'une communauté bactérienne composée de plusieurs centaines de bactéries, se basant sur divers formalismes mathématiques : numérique, logique ou hybride. Une revue des méthodes d'analyse des modèles métaboliques à l'échelle d'un génome est proposée dans les paragraphes suivants.

1.2.1 Méthodes numériques d'un réseau métabolique

La modélisation numérique du métabolisme informe quantitativement sur des mécanismes biologiques intracellulaires d'une espèce en se basant sur la matrice stœchiométrique ou bien sur les données cinétiques.

1.2.1.1 Constantes cinétiques des réactions

Les réactions biochimiques consomment des substrats et produisent des produits à une vitesse cinétique k selon la réaction globale suivante :



Il existe des modèles cinétiques se basant sur des constantes cinétiques et des équations du type Michaelis-Menten qui intègrent la constante d'affinité des enzymes pour leur substrat (Costa et al., 2016). Soit E , S , ES , P des substrats, k_1 , k_{-1} et k_{cat} des constantes cinétiques, le schéma réactionnel global peut être décrit comme suit :



De ce schéma, l'équation de Michaelis-Menten permet de calculer la vitesse d'une réaction (une concentration par une unité de temps) :

$$V = \frac{V_{max}[S]}{K_m + [S]} \quad (1.3)$$

dans laquelle $V_{max}[S]$ correspond à la vitesse maximale V_{max} de transformation d'un substrat (en unité de concentration par unité de temps) multipliée par la concentration du substrat $[S]$ (en mol/L⁻¹) et K_m représente la constante d'affinité (en mol/L⁻¹).

L'avantage des modèles cinétiques est qu'ils assurent un haut niveau de confiance sur les prédictions. En revanche, acquérir les données permettant d'inférer les paramètres cinétiques est coûteux en temps et limite leur utilisation à des réseaux métaboliques plus petits (généralement de l'ordre de la vingtaine de réactions) (van Rosmalen et al., 2021).

1.2.1.2 Contraintes numériques

Avec les avancées des technologies à haut débit, des méthodes computationnelles se sont développées pour analyser ces données telles que des modèles considèrent des contraintes liées au réseau métabolique : topologiques, thermodynamiques et enzymatiques (Rajvanshi and Venkatesh, 2013).

Parmi les méthodes sous contraintes, on retrouve les méthodes d’analyse par équilibre des flux (FBA, de l’anglais *Flux balance analysis*) (Orth et al., 2010). Leur objectif est de calculer les flux de consommation et de production de métabolites pour chaque réaction qui compose le système. Pour cela, le système doit satisfaire la loi de conservation de la masse qui définit l’équilibre d’un système réactionnel et se décrit de la façon suivante :

$$\frac{d[\mathbf{x}]}{dt} = \mathbf{S}.v \quad (1.4)$$

dans lequel, \mathbf{S} représente la matrice stœchiométrique, v le vecteur de flux des composés des réactions et $\frac{d[\mathbf{x}]}{dt}$ la variation de la concentration du composé x au cours du temps t .

À l’état stable du système, on n’observe aucune accumulation de métabolites à l’intérieur du système, c’est-à-dire, que pour un même métabolite m_i intracellulaire, la quantité de m_i produite est égale à la quantité de m_i consommée. Le flux de chaque réaction, i.e. l’approximation de la vitesse cinétique, est ajusté de façon à ce qu’il n’y ait pas d’accumulation dans le milieu intracellulaire. Ainsi, dans l’équation 1.4, la concentration du métabolite x ne varie pas au cours du temps et l’équation sous hypothèse d’état stable s’écrit de la façon suivante :

$$\mathbf{S}.v_{int} = 0 \quad (1.5)$$

Cette hypothèse forte est justifiable car la vitesse des réactions intracellulaires est plus rapide que les changements macroscopiques observables, par exemple, la croissance d’une plante.

Lors de la résolution d’un problème FBA, une fonction objective est optimisée, soit en la maximisant, soit en la minimisant. Généralement, la fonction objective est une réaction de biomasse simulant la croissance d’un organisme en fonction des substrats disponibles. Généralement, cette optimisation vise à rediriger les flux des métabolites à l’intérieur de la cellule pour permettre une production de biomasse maximale (Feist and Palsson, 2010; Orth et al., 2010). La conception d’une telle fonction n’est pas un processus trivial, d’autant plus que chaque espèce ne croît pas à la même vitesse, mais les fonctions de biomasses présentent souvent des familles de composés en commun (Feist and Palsson, 2010).

Dans un réseau métabolique à l’échelle du génome, il y a plus de réactions que de métabolites et donc, pour un même métabolite, il existe plusieurs chemins possibles de dégradation et production. Ainsi, lors de l’étape d’optimisation de la fonction objective, la solution donnée par la résolution d’un problème FBA n’est pas unique (voir Figure 1.4).

Afin de restreindre l’espace de recherche par le solveur et de respecter les contraintes thermodynamiques du système, les valeurs possibles de flux sont bornées par un flux minimal $v_{i_{min}}$ et maximal $v_{i_{max}}$, de la façon suivante :

$$v_{i_{min}} \leq v_i \leq v_{i_{max}} \quad (1.6)$$

Toutes ces contraintes se traduisent sous le programme linéaire global suivant :

$$\begin{aligned} &\text{maximiser/minimiser } f_{obj} \\ &\text{tel que } (\mathbf{S}.v)_{int} = 0 \\ &\text{et } v_{i_{min}} \leq v_i \leq v_{i_{max}} \end{aligned} \quad (1.7)$$

où f_{obj} est la fonction objective, $(S.v)_{int} = 0$ suppose l'état stable du système et $v_{i_{min}} \leq v_i \leq v_{i_{max}}$ les valeurs de flux inférieures et supérieures. (Orth et al., 2010)

Soit un réseau métabolique composé de 2 réactions dont celle de la réaction de biomasse. Après optimisation du flux de la réaction de biomasse, deux optima sont trouvés (un seul sera choisi après n itérations) (voir Figure 1.4). Biologiquement, cela signifie qu'il existe un ensemble de valeurs de flux possibles pour la seconde réaction qui garantit une croissance optimale. Ces ensembles de valeurs sont déterminables par une méthode décrite dans les paragraphes ci-dessous.

L'approche FBA constitue la base des méthodes de modélisation du métabolisme par contraintes. Dans les prochains paragraphes, une liste non exhaustive des variantes du FBA couramment utilisées dans la littérature sera présentée.

Analyse parcimonieuse par équilibre des flux Une des variantes de l'analyse par équilibre des flux permettant d'effectuer des analyses complémentaires est *l'analyse parcimonieuse par équilibre de flux* (pFBA) (Lewis et al., 2010). Elle vise à trouver une unique solution en minimisant la somme totale des flux impliqués dans l'optimisation de la fonction objective, tout en maintenant un flux optimal de croissance.

Analyse de la variabilité des flux *L'analyse de la variabilité des flux* (FVA) (Mahadevan and Schilling, 2003), évalue pour chaque réaction, les valeurs de flux possibles. En effet le FBA trouve une distribution de flux non unique satisfaisant les contraintes du problème linéaire (1.7). Le FVA permet ainsi d'identifier des *réactions bloquées*, ayant un flux nul dans toutes les solutions optimales, des *réactions essentielles*, ayant toujours un flux non nul, et des *réactions alternatives*, avec un flux nul ou non nul selon les solutions.

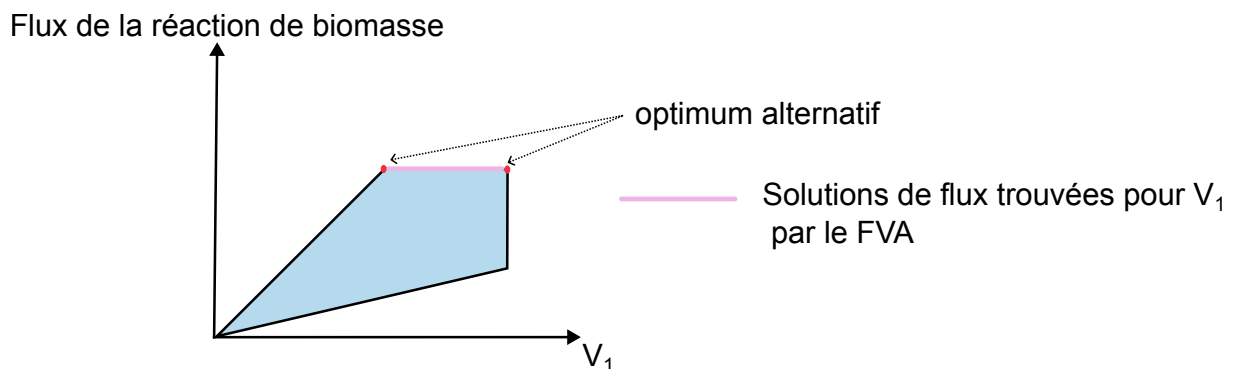


FIGURE 1.4 – Cône de solution obtenu lors de la résolution d'un FBA.

Analyse par équilibre des ressources Une extension du problème d'optimisation énoncé plus haut permettant de prendre en compte la machinerie de production des protéines se fait grâce à *l'analyse par équilibre des ressources* (RBA) (Goelzer et al., 2015). Dans ce dernier cas, chaque cellule utilise les ressources internes de façon économique.

Analyse par équilibre dynamique des flux Ces méthodes peuvent être utilisées dynamiquement grâce à un processus itératif : *analyse par équilibre dynamique de flux* (dFBA) (Mahadevan et al., 2002a). Le processus dynamique se modélise à l'aide d'un

système d'équations différentielles ordinaires (ODE). En plus de la dimension temporelle, la spatialisation peut également être modélisée avec des équations différentielles partielles (EDP) (Succurro and Ebenhöh, 2018) ou encore avec un dFBA avec l'outil COMETS (Harcombe et al., 2014). A chaque pas de temps δt , défini manuellement, le taux de croissance de chaque espèce et les concentrations des composés à suivre sont calculés grâce à la résolution d'un FBA.

Analyse des modes de flux élémentaires Les *modes de flux élémentaires* ou *elementary flux mode* en anglais (EFM), sont les ensembles minimaux de réactions activées qui garantissent l'état stationnaire du système (Schuster and Hilgetag, 1994; Trinh et al., 2009). Cette méthode permet principalement d'identifier des voies métaboliques optimales pour la production de composés ou encore d'évaluer la robustesse du réseau, c'est-à-dire, sa capacité à répondre à des changements de structure (Klamt et al., 2002). Une des limites de cette méthode est que le nombre de modes élémentaires augmente rapidement avec la taille du génome, ce qui complexifie leur calcul pour des réseaux à l'échelle génomique. Récemment, une méthode combinant la programmation par ensemble de réponse (*answer set programming*, ASP), un paradigme logique adapté à la résolution de problèmes combinatoires (Lifschitz, 2008) et un solveur linéaire pour calculer des EFM. L'application de la méthode au métabolisme central de *E. coli* a montré une réduction du coût calculatoire en intégrant différentes contraintes linéaires au programme logique (Mahout et al., 2020).

1.2.1.3 Intégration de données multi-omiques

Les méthodes d'optimisation sous contraintes peuvent être combinées à des données omiques, par exemple transcriptomiques, en ajoutant des contraintes linéaires afin que les réactions activées dans le modèle soient similaires à celles activées expérimentalement. Quelques outils permettant cette intégration sont présentés ci-dessous. GIMME (Becker and Palsson, 2008) crée un modèle métabolique cohérent avec les données biologiques et la fonction objective à maximiser, en minimisant les réactions trouvées inactives par l'outil. INIT (Agren et al., 2012) maximise l'utilisation des réactions basées sur un score de confiance et minimise les réactions sans protéine associée. INIT peut intégrer à la fois des données transcriptomiques ou protéomiques. Enfin, plus récemment, RIPTiDe (Jenior et al., 2020) utilise une approche parcimonieuse afin d'identifier les solutions les plus efficaces en termes de coûts, reflétant ainsi au mieux la transcription de la cellule.

1.2.2 Méthodes qualitatives d’analyse d’un réseau métabolique

La modélisation qualitative du métabolisme vise l’émergence de propriétés topologiques et fonctionnelles globales (Steuer, 2007). Dans les paragraphes suivants, des méthodes qualitatives pour analyser un réseau métabolique seront présentées.

1.2.2.1 Graphes

Les travaux de (Arabzadeh et al., 2018) présentent une comparaison d’approches numériques pour détecter des EFMs et une approche de détection basée sur l’analyse des graphes. En plus de ces analyses, certains auteurs et autrices se concentrent sur l’identification de voies métaboliques à partir d’un ensemble de métabolites sources et/ou de métabolites cibles (Heiden et al., 2002). Kuffner (Kuffner et al., 2000) et ses collègues démontrent que l’analyse des chemins métaboliques à partir d’une paire source-cible est un non-sens biologique. En énumérant l’ensemble des chemins possibles permettant d’aller du glucose au pyruvate, plus de 500 000 voies métaboliques ont été trouvées. Une autre approche consiste à utiliser les données *a priori* comme par exemple les données d’expression de gènes, pour contraindre les voies métaboliques (Zien et al., 2000). Enfin, dans (Simeonidis et al., 2003) les auteurs utilisent une analyse du plus court chemin entre deux molécules enzymatiques. Ils comparent avec les distances des deux gènes associés et démontrent qu’il n’existe pas de lien entre les deux. Cela suggère qu’il existe un mécanisme d’évolution de voies métaboliques. Dans une approche différente, NetSeed propose de trouver un ensemble de graines possibles en se basant à la fois sur le calcul de composantes fortement connexes (SCC) et sur le concept de l’écologie inversée (Carr and Borenstein, 2012).

1.2.2.2 Problème SAT

Un problème satisfiable (SAT) est utilisé pour déterminer si une formule booléenne peut être satisfiable par un ensemble de valeurs. Par exemple, pour que la proposition booléenne 1.8 soit vraie, il faut que la valeur que l’on assigne à x_1 soit différente d’elle-même. Ainsi, aucune valeur ne peut être assignée à x_1 qui puisse vérifier l’expression booléenne et donc, cette proposition logique n’est pas satisfiable.

$$x_1 \wedge \neg x_1 \tag{1.8}$$

En revanche, l’expression booléenne 1.9 stipule qu’il faut que la variable x_1 soit différente de x_2 . Il existe un nombre infini de solutions, comme $x_1 = 1$ et $x_2 = 0$, qui satisfont le problème logique ; ce problème est donc SAT.

$$x_1 \wedge \neg x_2 \tag{1.9}$$

Paradigme logique par ensemble de réponse (ASP) Ce paradigme logique est une forme de programmation déclarative utilisée pour résoudre des problèmes combinatoires (Lifschitz, 2008). L’ASP décrit les problèmes sous forme de règles, de contraintes et de faits logiques. Il est composé de 3 structures : une tête, un corps et un séparateur (\vdash écrit $:-$). Par exemple, la proposition 1.1 dit : “*X vole si X est un oiseau et X n’est pas une autruche*”. Après exécution du programme, l’ensemble réponse composé des atomes `oiseau(toto)`, `autruche(titi)` et `vole(toto)` satisfait l’ensemble des règles et de contraintes.

```

1 oiseau(toto). % faits
2 autruche(titi). % faits
3 vole(X) :- oiseau(X); not autruche(X).

```

Listing 1.1 – Exemple d’un code ASP générant une seule solution.

En revanche, aucun atome `vole` n’est déduit pour la proposition 1.2, puisque `toto` est à la fois un oiseau et une autruche.

```

1 oiseau(toto). % faits
2 autruche(toto). % faits
3 vole(X) :- oiseau(X); not autruche(X).

```

Listing 1.2 – Exemple d’un code ASP ne générant aucune solution.

Lors de sa résolution, le système `clingo` instancie les règles décrites en ASP sous forme de variables (étape de *grounding* avec `gringo`). Puis le solveur `clasp` applique les différentes contraintes aux modèles afin de sélectionner le ou les modèles stables, c’est-à-dire, qui satisfont l’ensemble des règles et de contraintes. Ces modèles stables sont exprimés avec le même vocabulaire des règles logiques, permettant aux utilisateurs et utilisatrices d’obtenir une explication du résultat. Le schéma global du processus de résolution de problèmes en ASP est récapitulé dans la figure 1.5.

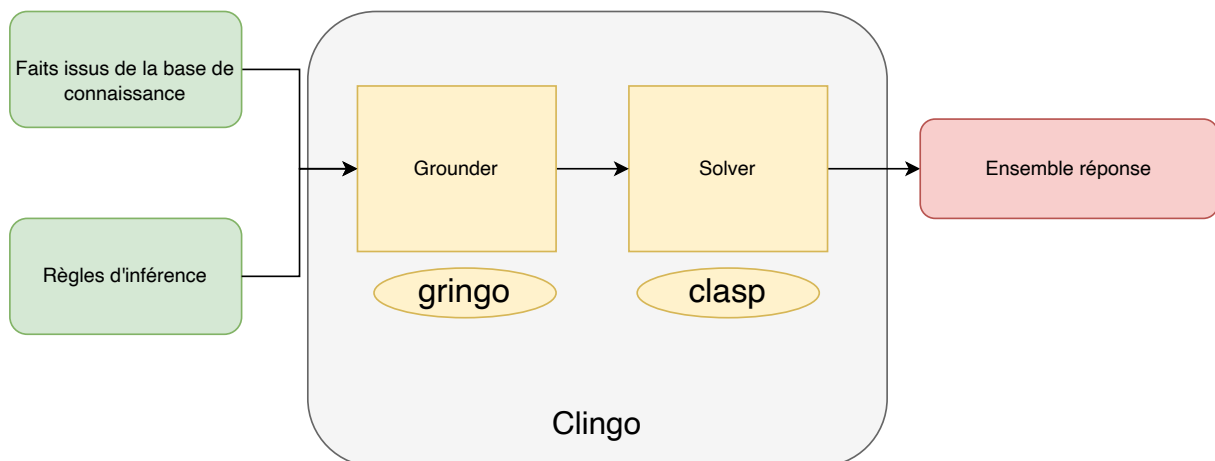


FIGURE 1.5 – Schéma récapitulant les étapes de résolution d’un problème ASP à partir des règles logiques.

1.2.2.3 Application à des problèmes biologiques

Les travaux de (Tiwari et al., 2007) proposent une méthode SAT pour analyser des réseaux de réactions ; (Corblin et al., 2006) appliquent une méthode basée sur SAT pour déchiffrer des réseaux de régulation de gènes.

Des travaux couplent l’ASP avec l’algorithme d’expansion de réseaux métaboliques (Ebenhöh et al., 2004) et définissent la notion de productibilité d’un métabolite sous forme de règles logiques. C’est le cas dans les travaux de (Collet et al., 2013), appliqués au réseau métabolique de *Ectocarpus siliculosus* et ceux de (Prigent et al., 2017) dont l’outil de complétion de réseaux métaboliques, Meneco, ajoute des réactions débloquent

la production de composés d'intérêt. Dans la lignée de ces travaux, des outils d'exploration de la productibilité métabolique dans les réseaux ont été développés (Aite et al., 2018) et l'élargissement de l'applicabilité au screening de communautés microbiennes a été modélisé. Dans tous ces travaux, l'algorithme d'expansion de réseau devient un raisonnement logique (Levesque, 1986) représenté sous forme de règles. Ainsi, un composé est productible par une réaction si l'ensemble des réactants de cette réaction est également productible. Cette règle récursive est implémentée en ASP par la proposition 1.3.

```

1 productibilite(Compose) :- graine(Compose).
2
3 productibilite(Compose) :- produit(Compose, Reaction);
4   reaction(Reaction);
5   productibilite(Compose2) : reactant(Compose2, Reaction).

```

Listing 1.3 – Code ASP permettant de calculer le scope métabolique

Au sein de la proposition 1.3, il y a tout d'abord une étape d'initialisation, informant que les graines, c'est-à-dire le milieu nutritionnel, sont déjà productibles, puis la partie récursive.

Afin de générer des règles, il faut une base de connaissances adéquate, c'est-à-dire, l'ensemble des faits décrivant, dans notre cas, le réseau métabolique. Dans le cas de la proposition 1.3 décrivant la productibilité d'un composé, la base de connaissances est constituée de l'ensemble du milieu de culture représenté par l'atome `graine`, des réactions (`reaction`), des réactants (`reactant`) et des produits (`produit`).

1.2.3 Défis rencontrés lors de la modélisation qualitative et numérique d'une cellule

Pour analyser une cellule du point de vue du métabolisme, des choix de modélisation doivent être pris en compte. Tout d'abord, au niveau des réactions de transport qui permettent le passage de métabolites du milieu intracellulaire vers le milieu extracellulaire. Pour les deux types de modélisation, des réactions de transports doivent être ajoutées aux modèles métaboliques prenant en compte les différents éléments du milieu de culture. Un second défi rencontré est la prise en compte des cycles de régénération des cofacteurs, *e.g.* NADH, NADPH, FADH, etc. Avec une approche par raisonnement, ils sont souvent considérés comme disponibles afin d'initier leurs cycles de régénération. Avec les méthodes FBA, les contraintes font converger les flux de métabolites vers les voies thermodynamiquement faisables.

1.2.4 Méthodes hybrides de modélisation à l'échelle d'un modèle métabolique

La modélisation est dite *hybride* lorsqu'une variable d'état peut rassembler des informations issues de formalismes mathématiques ou informatiques distincts, comme par exemple discret et continu (Cruz and Kemp, 2022). Il existe trois grands types de modèle hybride : (i) indépendant, (ii) couplé ou (iii) imbriqué (Stéphanou and Volpert, 2016). Le modèle indépendant consiste en au moins deux modèles décrivant le même phénomène à des échelles de précision différentes et où aucune interaction entre les modèles n'existe. Le

modèle couplé consiste en au moins deux modèles pouvant communiquer unidirectionnellement ou bidirectionnellement entre eux. Enfin, le modèle hybride imbriqué forme une seule entité composée de sous-parties sensibles au changement du contexte, déclenché par un seuillage par exemple. Il ne sera pas détaillé dans cette section.

1.2.4.1 Indépendants

Au sein des modèles indépendants, c'est l'approche qui est qualifiée d'hybride et non le modèle en lui-même. Les réseaux métaboliques possèdent la propriété d'être décomposables (Fisher and Piterman, 2010), c'est-à-dire qu'ils peuvent être décomposés en plusieurs petites structures qui sont plus simples et permettant une analyse facilitée. On peut imaginer une approche de reconstruction de réseaux métaboliques hybride, utilisant comme premier modèle une reconstruction bottom-up et de l'autre, une reconstruction top-down. Ce type d'approche est utile pour décrire un même phénomène à des échelles différentes, permettant ainsi d'identifier lequel des modèles utilisés est le plus précis et le moins gourmand en calcul.

1.2.4.2 Couplé

Dans le cadre des modèles couplés, les variables de sortie d'un modèle servent en entrée d'un autre. Ce type de modèle hybride sert principalement à traiter des problèmes à plusieurs échelles (Brook and Waters, 2008). Dans le cas des modèles hybrides appliqués à la biologie des systèmes, Frioux et al (Frioux et al., 2019) ont développé une méthode de complétion de réseaux de métabolique hybride combinant un formalisme logique d'expansion de réseau en ASP (Gebser et al., 2019) et la modélisation du FBA (Janhunen et al., 2017), résolvant respectivement les contraintes logiques et les contraintes linéaires du programme. Les modèles trouvés satisfont ainsi les deux types de contraintes. Dans le cas où les contraintes linéaires ne sont pas satisfaites, une communication bidirectionnelle entre les deux solveurs informe de la non-satisfiabilité.

Toutes les approches décrites ci-dessus permettent l'analyse d'un réseau métabolique à l'échelle d'un génome. Comme nous l'avons vu dans l'introduction, les données générées en grande quantité sur des communautés non cultivables motivent le développement de modèles de communauté pour générer des hypothèses. Dans le reste de ce chapitre, des méthodes à l'échelle de l'écosystème analysant des communautés de microorganismes seront présentées, avec leurs avantages, leurs limites et leurs applications.

1.3 Modélisation du métabolisme des écosystèmes microbiens

De nombreux nombreux outils, utilisant le formalisme numérique et ou discret, existent, comme illustrés par la Figure 1.10. Tout comme pour la section sur l'analyse du métabolisme à l'échelle du génome, les méthodes numériques seront d'abord présentées.

1.3.1 Méthodes numériques appliquées à plusieurs réseaux métaboliques

1.3.1.1 Équations différentielles ordinaires (ODE)

Un moyen pour décrire la dynamique spatio-temporelle d'une communauté microbienne est d'utiliser les équations différentielles (Succurro and Ebenhöh, 2018). En 1838, Verhulst a défini le premier modèle décrivant la croissance d'une bactérie dans une culture au cours du temps (F., 1838). Par la suite, les modèles proie-prédateurs ont été utilisés pour décrire une interaction, positive ou négative, d'une espèce prédatrice sur la croissance d'une autre en communauté (Nair et al., 2019). On peut également retrouver les modèles de Lokta et Volterra, largement utilisés pour étudier l'évolution dans le temps d'un écosystème et principalement, les modèles généralisés de Lokta-Volterra, prenant en compte les données d'abondance relative de bactéries lors d'une analyse en métagénomique (Berry and Widder, 2014). En revanche, ces modèles ne se basent que sur les interactions trophiques entre les bactéries et ne prennent pas en compte l'environnement nutritionnel. Il existe cependant une extension du modèle de Lokta-Volterra généralisé permettant de tenir compte d'une potentielle perturbation environnementale sur le comportement d'un individu (Stein et al., 2013). Toutes ces méthodes sont "data-driven", i.e. dépendantes des données et de la connaissance que l'on peut avoir sur le système biologique d'étude, et ne sont pas mécanistiques, c'est-à-dire qu'elles ne proposent pas d'hypothèses sur les mécanismes sous-jacents des interactions. Ainsi pour permettre une analyse détaillée du système biologique permettant de répondre aux questions, "qui produit quoi?" ou "comment une perturbation dans la composition bactérienne ou dans les milieux nutritifs impacte la communauté bactérienne?", il faut inférer cette connaissance sous forme de paramètres. En d'autres termes, chaque mécanisme de régulation que l'on souhaite modéliser doit être décrit par, *a minima*, une équation.

1.3.1.2 Analyse statique sous contraintes

Les modèles basés sur les flux s'appliquent également à l'échelle de la communauté. Le modèle de communauté peut se décrire de trois façons possibles : compartimentalisé, soupe ou encore indépendant (voir Figure 1.6) (Colarusso et al., 2021). Chaque modèle possède des propriétés propres (Succurro and Ebenhöh, 2018). Le modèle soupe (Figure 1.6b) permet d'explorer le potentiel métabolique de la communauté sans la nécessité d'obtenir une résolution de chaque modèle individuel. Cela se traduit généralement par une optimisation d'une fonction de biomasse communautaire créée en assemblant les compositions de chaque fonction de biomasse des organismes. Le modèle indépendant comme le modèle compartimentalisé permet de prendre en compte les abondances de chaque individu. Dans le modèle compartimentalisé 1.6a), chaque bactérie possède son propre espace intracellulaire et extracellulaire, permettant ainsi des échanges entre espèces. Un espace extracellulaire les englobe contenant le milieu nutritionnel. La fonction de biomasse communautaire correspond généralement à la somme des valeurs de flux de biomasse individuel. Dans un modèle indépendant (Figure 1.6c), chaque bactérie conserve son propre espace intracellulaire mais partage le même espace extracellulaire avec l'ensemble de la population. La fonction de biomasse communautaire peut être soit multi-objective soit à l'échelle de chaque individu, comme pour le modèle compartimentalisé.

En fonction du type de représentation du modèle de communauté il existe plusieurs outils de modélisation de communautés microbiennes (Colarusso et al., 2021). Dans Henry et al (Henry et al., 2016), les auteurs ont choisi de maximiser une fonction de biomasse

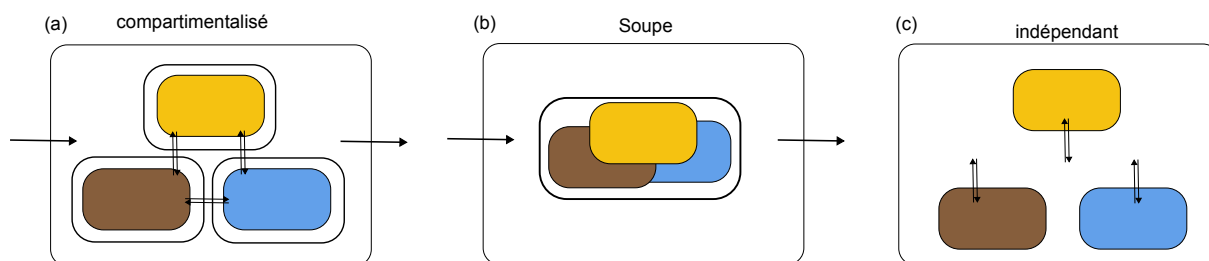


FIGURE 1.6 – **Différents types de stratégies pour construire un modèle de communauté.** (a) Dans le modèle compartimentalisé, chaque bactérie possède son propre compartiment cytosolique et extracellulaire. Un compartiment extracellulaire commun à l'ensemble des bactéries est également disponible. (b) Dans le modèle soupe, les bactéries sont assemblées en une seule pour former un supra-organisme. (c) Les bactéries ont chacune leur propre compartiment cytosolique mais partagent un milieu extracellulaire commun.

communautaire se basant sur la composition des réactions de biomasse de chaque taxon présent dans la communauté. Microbiome modeling toolbox (Baldini et al., 2019), MiCOM (Diener et al., 2020) et SteadyCom (Chan et al., 2017) maximisent tous la somme des flux de la fonction de biomasse de chaque organisme et permettent l'étude de plus grandes communautés. Le cFBA (Khandelwal et al., 2013) ou encore la méthode développée par Koch et al. (Koch et al., 2016) permettent une analyse sur de petites communautés. Enfin, dans la catégorie d'une optimisation à plusieurs niveaux, on peut retrouver les outils suivants : OptCom (Zomorodi and Maranas, 2012), NECom (Cai et al., 2020), maximisant les fonctions de biomasse de chaque modèle individuel puis les flux de production de biomasse à l'échelle de la communauté. Pour des raisons de coût calculatoire, appliquer ces outils sur des grandes communautés microbiennes est non-réalisable. La table 1.1 récapitule les caractéristiques des outils mentionnées ci-dessus.

Toutes ces méthodes basées sur des contraintes permettent d'étudier le métabolisme et par conséquent peuvent être utilisées pour mettre en évidence l'impact positif (coopération) ou négatif (compétition) d'une interaction bactérienne. Ainsi, dans (Freilich et al., 2011) les scores de potentiels de coopération et de compétition sont inférés en caractérisant les relations de donneurs-receveurs et de gagnants-perdants au sein de l'écosystème. Plus récemment, MMinte (Mendes-Soares et al., 2016) prédit l'impact positif ou négatif d'une espèce sur une autre en comparant les croissances des espèces seules et lorsqu'elles sont en interactions deux à deux. L'analyse des interactions deux à deux ne permet pas de déduire l'ensemble des comportements de la communauté bactérienne puisque de nouvelles interactions se produisent, disparaissent ou encore se conservent (Morin et al., 2022). SMETANA (Zelezniak et al., 2015) et MiCOM (Diener et al., 2020) contournent cette limite et proposent d'étudier les interactions bactériennes de l'ensemble des espèces de la communauté.

SMETANA et MiCOM définissent des interactions métaboliques à partir d'un ensemble de GEMs et des environnements nutritionnels, comme illustré par les Figures 1.81.7. Contrairement à SMETANA, MiCOM permet de prendre en compte des données d'abondance relative de chaque espèce présente dans la communauté. De plus, MiCOM ne suppose pas que toutes les espèces microbiennes se développent à un même taux et permet d'obtenir des valeurs de flux spécifiques à chaque espèce à partir de l'abondance

relative contrairement à SteadyCOM, qui prédit l'abondance relative en supposant que toutes les espèces microbiennes poussent à un même taux.

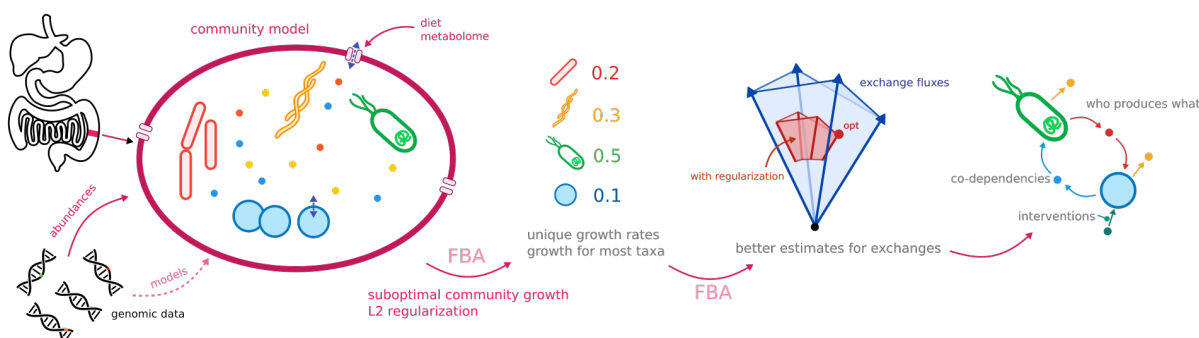


FIGURE 1.7 – Stratégie mise en place par l’outil MiCOM (Figure extraite de la documentation MiCOM). Deux optimisations sont faites : la première vise à obtenir une solution unique et la seconde vise à améliorer la prédiction des échanges métaboliques.

SMETANA définit, comme illustré par la Figure 1.8, la coopération via le potentiel d’interaction métabolique (MIP), représentant le nombre maximal de composés nutritifs essentiels que la communauté peut produire, au moyen des interactions bactériennes. La compétition est estimée par le maximum de recouvrement de métabolites (MRO) au sein du milieu minimum pour la croissance de toutes les espèces.

En plus de ces deux scores, SMETANA permet également d’obtenir d’autres indications sur la communauté sous la forme d’autres scores. Le premier, le *species coupling score* (SCS) qui mesure la dépendance d’une espèce à survivre en la présence des autres. Le second, le *metabolite uptake score* (MUS), qui mesure la nécessité d’une espèce à consommer un métabolite pour sa survie. Et enfin le *metabolite production score* (MPS) qui mesure la capacité d’une espèce à produire un métabolite.

MiCOM ne permet pas d’obtenir directement une mesure de la coopération et de la compétition mais, tout comme SMETANA, donne une liste de métabolites potentiellement échangés avec pour chaque, le flux d’export et d’import.

1.3.1.3 Analyse dynamique sous contraintes

Les méthodes précédentes permettent d’analyser de manière statique des communautés microbiennes de tailles variables. Or, un écosystème est un environnement dynamique dans lequel la composition taxonomique et la quantité de ressources disponibles fluctuent au cours du temps. Une liste de méthodes, prenant en compte la temporalité dans un écosystème, ainsi que leurs limites, est proposée dans les paragraphes suivants.

La méthode DMMM (Zhuang et al., 2011) utilise des équations différentielles ordinaires pour décrire la loi de conservation de la masse pour tous les métabolites échangés et la production de biomasse de chaque espèce au sein de la communauté. Le temps est alors discrétisé, et à chaque pas de temps, une optimisation est faite. La concentration des métabolites disponibles dans le milieu extra-cellulaire est ajustée après la consommation et la production des différents métabolites échangés ainsi que la densité cellulaire (méthode explicite d’Euler). Avec ce choix temporel d’ajuster la concentration des métabolites

Méthodes computationnelles	Données d'entrée	Taille de la communauté	Fonction objective	Prédictions	Type d'interactions bactériennes	Choix de modélisation de communauté
cFBA	GEMs	Petite	Maximisation de la croissance communautaire	Identification de métabolites échangeables	Interaction deux-à-deux (au minimum)	Compartmentalisé
MiCOM	GEMs	Large	2 étapes d'optimisation	Interactions métaboliques Direction des transferts métaboliques	Interaction communautaire	Compartmentalisé
OptCOM	GEMs	Petite	Multi-objectif Maximisation de la biomasse de chaque individu Maximisation de la croissance communautaire	Prédit la direction des transferts de métabolites	Interaction deux-à-deux	Compartmentalisé
SteadyCOM	GEMs	Large	Maximisation de la croissance communautaire	Composition microbienne (abondance) dans un environnement	-	Compartmentalisé
Koch et al	GEMs	Petite	Multi-objectif Maximisation de la croissance communautaire Maximisation de la biomasse de chaque individu	Composition microbienne (abondance)	-	Compartmentalisé avec autorisation d'accumulation dans le milieu
Henry et al	GEMs	Petite	Somme des coefficients stœchiométriques de chacune des réactions de biomasse de chaque individu	Interactions métaboliques inter espèces	Interactions deux-à-deux (au minimum)	Soupe compartmentalisé
Microbiome modeling toolbox (MMT)	GEMs	Large	Maximisation de la croissance communautaire	Profils métaboliques	Interaction deux-à-deux	Compartmentalisé
SMETANA	GEMs	Petite	Multi-objectif (du au calcul des différents scores)	Interactions métaboliques Direction des transferts métaboliques	Score d'interaction communautaire	Indépendant ou compartmentalisé
MMinte	OTU	Large	Maximisation de la croissance communautaire	Identification des impacts positifs et négatifs des interactions bactériennes	Interaction deux-à-deux	Compartmentalisé
NECom	GEMs	Petite	Multi-objectif Maximisation de la biomasse de chaque individu Maximisation de la croissance communautaire	Interactions métaboliques microbiennes	Interaction deux-à-deux	Indépendant

TABLE 1.1 – Table récapitulative d'outils de modélisation statique d'une communauté microbienne.

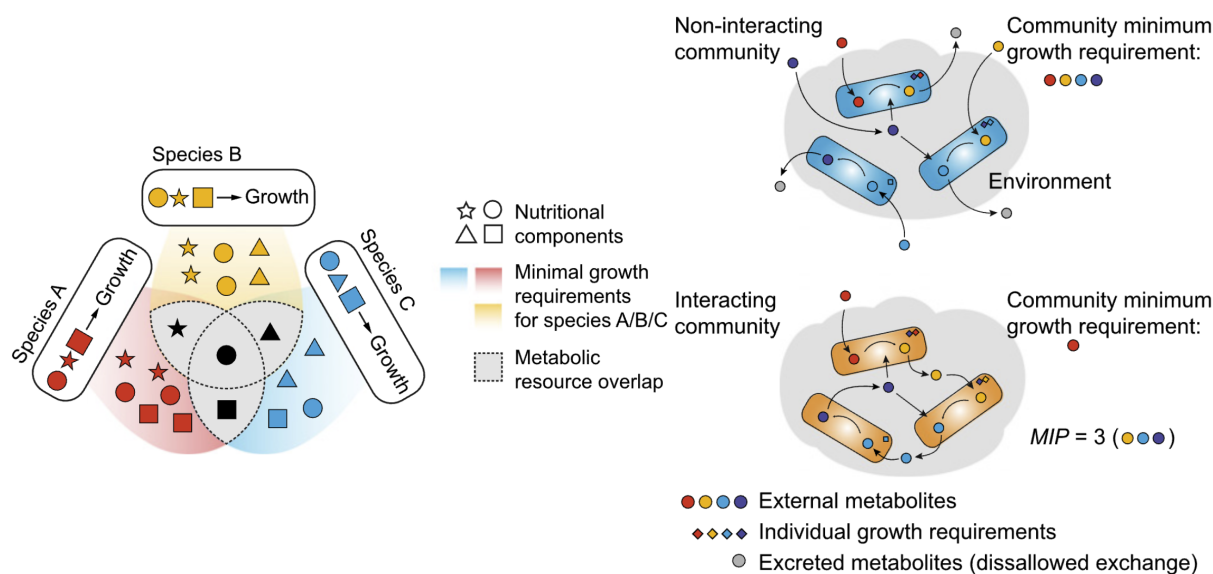


FIGURE 1.8 – Illustration indiquant comment le MRO, proxy de la compétition (à gauche), et comment le MIP, proxy de la coopération (à droite), sont calculés par SMETANA. Figures issues de (Zelezniak et al., 2015).

extracellulaires après l’importation des nutriments, un risque du modèle est de prédire que les espèces consomment un nutriment dont la concentration n’est pas suffisante par rapport à la densité cellulaire. Une première possibilité consiste à rendre indisponible tous les métabolites dont la concentration est négative. Une seconde possibilité revient à diminuer l’intervalle de temps entre deux calculs au risque d’augmenter le coût calculatoire. Ainsi, *μbialSim* (Popp and Centler, 2020) solutionne ce problème en réduisant le pas de temps seulement lorsque la concentration d’un métabolite échangé est négative.

Les modèles dynamiques de communautés doivent prendre en compte le cas où il n’y a pas de solution possible pour le système dynamique étudié ; on parle de *résolution non faisable*, autrement dit, lorsqu’il y a pas un flux unique de production ou de consommation pour un métabolite. La faisabilité d’un modèle est liée à la quantité de nutriments limitants disponibles pour le système. DFBALab (Gomez and Barton, 2018) tente de résoudre ces deux problèmes en priorisant le choix des métabolites d’intérêts à maximiser. Ainsi, une unique solution est trouvée par le solveur linéaire permettant d’outrepasser la seconde contrainte. En revanche, l’utilisateur ou l’utilisatrice doit donner une liste de métabolites échangés en amont pour que le résultat de la simulation reste cohérent. Cela présuppose une bonne connaissance du système biologique d’étude pour effectuer ce choix et est donc peu recommandé pour l’étude de communauté à grande échelle. D’autres méthodes existent utilisant une approche de résolution plus précise (Runge-Kutta explicite) mais plus coûteuse en temps de calcul que celle d’Euler explicite (Schroeder and Saha, 2020) ; ou encore, une optimisation à deux niveaux (Zomorodi, 2014), ne nécessitant pas l’intervention d’une connaissance *a priori* du système d’étude.

Une des limites fondamentales de ces utilisations est que tous les échanges métaboliques sont conditionnés par l’optimisation de la fonction de biomasse. Ainsi, un ensemble de métabolites pouvant être échangés au sein d’une communauté mais n’étant pas essentiel pour la biomasse d’un individu ne sera donc pas identifié. Dans les travaux de Zomorodi et al (Zomorodi and Segrè, 2017), ce problème est abordé en ajoutant des contraintes supplémentaires qui pré-traitent différentes stratégies comme la prise en compte de métabolites échangés coûteux. En revanche, identifier l’ensemble des stratégies à l’échelle d’une communauté ne semble pas être réalisable. Au sein de l’outil NECom (Cai et al., 2020), se basant sur une optimisation à deux niveaux, les auteurs relâchent les contraintes de flux acceptant plus de métabolites potentiellement échangés. La table 1.2 récapitule les caractéristiques des outils mentionnés ci-dessus.

1.3.2 Applications

En plus de l’étude des interactions au sein de communautés bactériennes naturelles ou contrôllées, la création de communautés synthétiques est une application pertinente aux approches de modélisation du métabolisme des écosystèmes microbiens. Elle permettent (i) d’identifier des communautés microbiennes d’intérêt à partir d’une liste de bactéries candidates, (ii) d’évaluer l’impact d’une modification génétique sur la synergie de la communauté, et enfin (iii) d’évaluer les perturbations dans le milieu de culture ou dans la composition de la biomasse sur la communauté. Dans (Chiu et al., 2014), l’utilisation d’un modèle dynamique a permis de mettre en avant une communauté composée de deux espèces qui produisent et échangent des composés uniquement lorsqu’elles sont en co-culture. Dans (Zuñiga et al., 2020), le modèle dynamique sélectionne des souches pho-

Méthodes computationnelles	Données d'entrée	Fonction objective	Traitement substrats limitants	Type d'interactions bactériennes	Choix de modélisation de communauté
DMMM	GEMs	Maximise le taux de croissance de chaque souche	Ajustement des bornes de consommation après les processus cellulaires	Interactions deux-à-deux (a minima)	Indépendant
μ bialSim	GEMs	Maximise le taux de croissance de chaque souche	Ajustement des bornes de consommation après les processus cellulaires	Interaction communautaire	Compartmentalisé
DFBALab	GEMs	Multi-objectif Maximisation de la biomasse Maximisation des flux d'échanges supplémentaire	-	-	Compartmentalisé
COMETS	GEMs	Maximisation de la biomasse individuelle	Équations de Michaelis-Menten	-	Compartmentalisé

TABLE 1.2 – **Table récapitulant les outils de modélisation dynamique d'une communauté microbienne.**

totrophes partenaires prometteuses pour la biotechnologie durable. D'autres méthodes, comme par exemple OptAux, s'intéressent à la création de souches auxotrophes, c'est-à-dire ne pouvant synthétiser un composé par elles-mêmes, qui sont vérifiables expérimentalement (Lloyd et al., 2019). Un autre point de vue d'analyse est celui du milieu de culture, comme par exemple dans (Klitgord and Segrè, 2010), où l'objectif est de sélectionner un milieu de culture permettant d'observer soit des interactions du type mutualisme ou du commensalisme entre des paires d'organismes, soit une production de métabolites cibles (Pacheco and Segrè, 2021). Dans le même esprit, FLYCOP (García-Jiménez et al., 2018) crée un consortium à partir d'un milieu de culture dans le but de trouver tous les échanges métaboliques permettant la production de bioplastique. Dans (Machado et al., 2021), l'impact *in silico* de facteurs abiotiques (introduction d'une nouvelle espèce) et biotiques (disponibilité des nutriments) sur la communauté a été analysé. Ces travaux ont montré que les communautés compétitives sont plus sensibles aux perturbations abiotiques, et que les communautés coopératives se sont révélées être sensibles à l'introduction d'une nouvelle espèce.

Toutes ces méthodes permettent d'analyser des communautés microbiennes sous différents points de vue : identification statique ou dynamique des interactions métaboliques au sein d'une communauté donnée, recherche d'un milieu de culture pour la création d'une communauté avec le comportement souhaité, compréhension détaillée de mécanismes métaboliques intra-cellulaires ou encore, impact d'une perturbation sur le potentiel métabolique de la communauté. Malgré la précision numérique que l'on obtient, l'utilisation de ces modèles ne permet pas une analyse sur des communautés naturelles, notamment en raison de la combinatoire exponentielle des échanges métaboliques. Des méthodes qualitatives, s'appliquant à l'échelle d'une communauté, sacrifient la précision numérique pour s'atteler aux limitations précédentes.

1.3.3 Méthodes et modèles qualitatifs de l'étude d'une communauté bactérienne

Comme pour les méthodes quantitatives, des approches qualitatives se sont focalisées sur l'étude des interactions bactériennes. Nous allons voir les stratégies d'approches basées sur l'analyse des graphes métaboliques d'approches basées sur le raisonnement permettant l'identification des potentiels de coopération et de compétition au sein de communautés microbiennes de tailles variables. Pour rappel, un échange métabolique se traduit par la

production d'un composé par un microorganisme dans l'environnement nutritionnel qui est lui-même consommé par un autre, impactant ou non sa croissance (Faust and Raes, 2012). La compétition se traduit par une co-consommation d'un nutriment limitant ayant des effets néfastes sur la croissance ou sur le potentiel métabolique. La coopération et la compétition sont des mécanismes biologiques centraux pour mieux comprendre comment une communauté bactérienne fonctionne.

1.3.3.1 Approches basées sur les graphes

Certaines approches étudient la structure de réseaux métaboliques, sous forme d'un graphe, en déterminant la relation de dépendance des éléments du graphe au milieu nutritif. Les éléments identifiés sont donc utilisés comme une approximation des métabolites échangés (Levy and Borenstein). Ainsi, en se basant sur le principe d'écologie inversée consistant à déduire des processus écologiques en utilisant les données génomiques massives, la détection de potentiels de coopération et de compétition est possible. Un moyen de calculer ces potentiels au sein d'une communauté bactérienne est de déterminer la composition nutritionnelle (graines) requise pour les membres de la communauté (Carr and Borenstein, 2012). Dans le but de quantifier la complémentarité entre toutes les paires d'espèces possibles, NetCooperate (Levy et al., 2015) mesure à la fois la capacité d'un organisme hôte à répondre aux besoins nutritionnels d'un parasite endosymbiotique (BSS score) et quantifie l'aide mutuelle de deux organismes en fournissant un indice de potentiel syntrophique (MCI score). En se basant sur une analyse du milieu nutritionnel théorique, on peut également déterminer un indice de compétition métabolique en analysant les composés connectés à plusieurs organismes (Levy and Borenstein, 2013). Ainsi, NetCompt estime le potentiel de compétition en se basant sur le chevauchement métabolique (Kreimer et al., 2012). Toutes ces méthodes sont limitées à une analyse par paire d'organismes ne reflétant pas le réel potentiel coopératif ou compétitif d'une communauté microbienne complexe mais informant l'impact d'un individu sur un autre.

1.3.3.2 Modèles basés sur le raisonnement

On a vu dans §1.2.2.3 ci-dessus que la méthode par raisonnement pouvait identifier des composés potentiellement productibles et consommables. Les travaux de (Frioux et al., 2018) ont décrit un modèle de communauté basé sur le raisonnement utilisant une représentation du potentiel de coopération à travers des échanges métaboliques potentiels.

De l'anglais *Microbiome Screening and COmmunity selection using TOpology*, MiS-CoTo sélectionne des communautés minimales en utilisant le formalisme logique par ensemble de réponses ASP, afin d'effectuer des optimisations sur la sélection (illustré par le Figure 1.9). Tout comme les modèles à l'échelle d'un génome, MiS-CoTo décrit la notion de productibilité d'un organisme telle que définie par l'algorithme d'expansion de réseaux métaboliques (Ebenhöh et al., 2004) sous deux règles récursives que sont :

- La productibilité est initiée par la présence de graines
- Les produits d'une réaction sont productibles si tous les réactants de cette réaction sont également productibles. La particularité ici est de prendre en compte qu'un autre réseau métabolique peut rendre disponibles les précurseurs de la réaction, et de tracer cette connaissance pour définir des interactions potentielles.

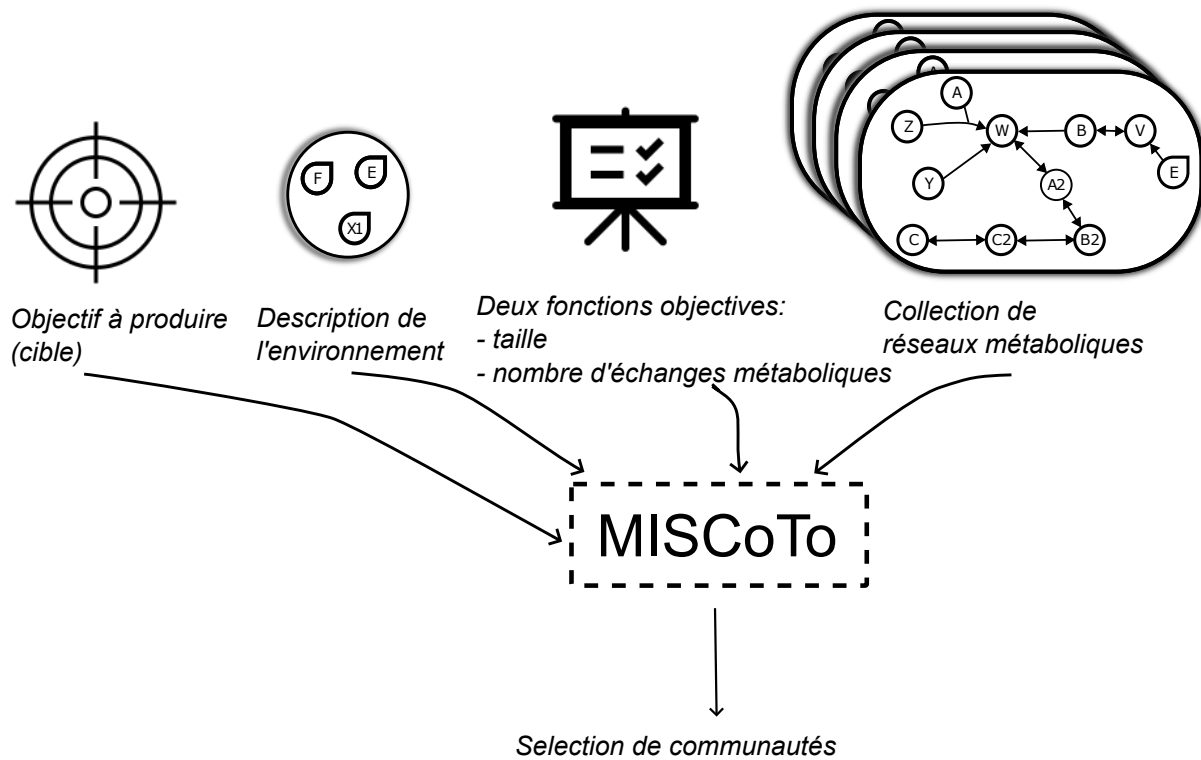
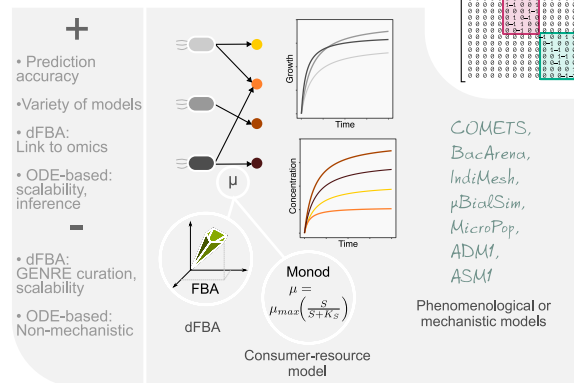
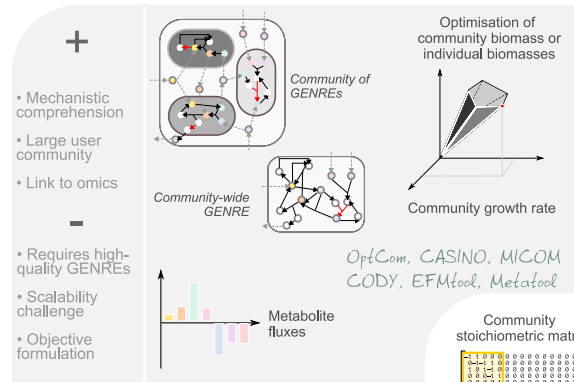


FIGURE 1.9 – **Illustration du fonctionnement de l’outil MiSCoTo.** A partir d’une collection de réseaux métaboliques, d’un ensemble de graines, d’une liste de métabolites à produire et d’une optimisation, MiSCoTo sélectionne des communautés minimales satisfaisant un objectif métabolique, c’est-à-dire la productibilité de composés d’intérêt.

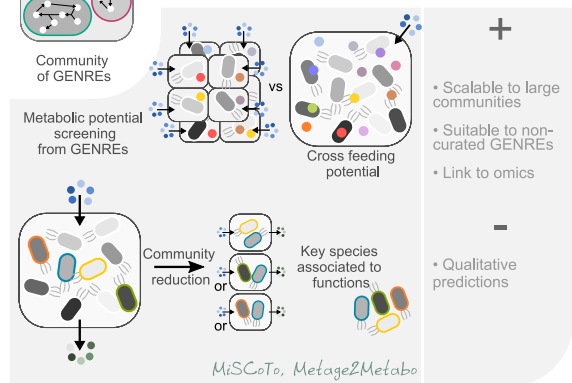
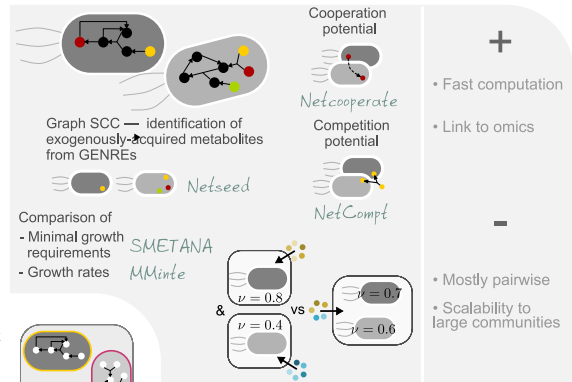
MiSCoTo est utilisé dans le pipeline Metage2Metabo (Belcour et al., 2020) qui combine reconstruction de réseaux métaboliques pour une collection de génomes, estimation des capacités métaboliques des populations microbiennes individuellement et en communauté, identification de la valeur ajoutée de la coopération métabolique, et criblage des espèces d’intérêt pour un objectif métabolique parmi tous les membres de la communauté.

Predict small-scale community metabolism



Assess community dynamics

Decipher cooperation and competition potentials



Characterize large-scale community metabolism

FIGURE 1.10 – Illustration des méthodes numériques et qualitatives de modélisation du métabolisme d’une communauté avec leurs avantages et inconvénients. Extrait de Cerk et al. *Community-scale models of microbiomes : articulating metabolic modelling and metagenome sequencing*, 2024, Environmental Microbiology (Cerk et al., 2024)

1.3.4 Défis rencontrés lors de l’analyse de la modélisation qualitative et quantitative du métabolisme des communautés

Des questions similaires à celles déjà rencontrées (voir §1.2.3) se posent également au niveau de la communauté. Une hypothèse faite par plusieurs approches méthodologiques est que l’ensemble des métabolites produits en intracellulaire par une population microbienne peut être disponible pour nourrir des interactions avec la communauté. Cela peut générer des faux positifs dans l’inférence des interactions cooperatives. Néanmoins, cette hypothèse forte peut se justifier par l’existence de phénomènes de lyse cellulaire, libérant ainsi tous les produits métaboliques dans le milieu extracellulaire (Fazzino et al., 2020) ou par les fuites métaboliques passives, c’est-à-dire sans coût énergétique pour la cellule, qui ont été montrées comme étant dominantes dans les interactions microbiennes (Pacheco et al., 2019). Une seconde notion à prendre en compte est celle de la frontière des compartiments métaboliques propres à chaque espèce de la communauté.

Conclusion

Dans ce chapitre, nous avons proposé une revue de la littérature sur la thématique de la modélisation du métabolisme, en présentant dans un premier temps une définition et différentes représentations d'un réseau métabolique puis une collection de méthodes quantitatives et qualitatives d'analyse du métabolisme à l'échelle d'une cellule et d'un écosystème bactérien.

Parmi les approches de modélisation du métabolisme, nous avons vu que les modèles numériques sont pertinents d'une part, pour leurs capacités à intégrer des types de données omiques au moyen de contraintes linéaires ou d'optimisation, et d'autre part, pour la précision numérique qu'ils offrent. Cependant, cette aptitude rend délicate leur utilisation à des communautés plus complexes pour des raisons de coût calculatoire et de curation des modèles d'une part, et de la disponibilité des données d'entrée d'autre part.

Les approches basées sur le raisonnement sont une alternative qui dépasse ces limites en faisant une abstraction booléenne du métabolisme qui fournissent de surcroît un potentiel d'explicabilité aux modèles lorsqu'elles utilisent des paradigmes comme ASP. En revanche, la précision apportée par ces modèles reste néanmoins limitante. *In fine*, il n'existe pas de méthode universelle permettant de répondre à notre objectif, et un compromis doit être fait. Dans le reste de ce manuscrit, nous essayerons de mieux comprendre les apports et les enjeux de chaque type de modélisation dans le but de créer une approche hybride répondant à notre problématique.

Nous verrons ainsi, dans le chapitre 3, une application de cette approche de raisonnement sur un ensemble d'écosystèmes et comment l'explicabilité d'un modèle permet de définir un score de coopération et de compétition pour une communauté donnée. L'intérêt d'obtenir des modèles explicables sera particulièrement exploité au cours du chapitre 4 consistant à (i) sélectionner un ensemble de communautés d'intérêt à partir d'un ensemble de contraintes, (ii) intégrer une temporalité dans les modèles qualitatifs. Enfin, les données nécessaires pour réaliser ces travaux se sont uniquement basées sur des données génomiques. Dans le cas idéal, plusieurs types de données omiques permettent d'apporter des informations à différentes échelles du métabolisme. Ainsi, nous verrons dans le chapitre 2, la plus-value des données multi-omiques dans l'analyse du métabolisme d'une communauté contrôlée, et le degré de précision que l'on peut obtenir dans la prédiction des mécanismes d'interaction à l'aide de modèles métaboliques.

Chapitre 2

Modèle numérique du métabolisme bactérien au cours de la production d'un fromage

Sommaire

2.1	Introduction	43
2.2	Méthode pour la construction d'un modèle numérique du métabolisme	44
2.2.1	Présentation des données	44
2.2.2	Du génome annoté au réseau métabolique à l'échelle du génome	48
2.2.3	Modélisation et implémentation numérique	48
2.2.4	Raffinement des modèles métaboliques individuels	51
2.2.5	Calibration des modèles métaboliques individuels	51
2.3	Application de la méthode numérique à l'écosystème fromager	52
2.3.1	A l'échelle de chaque souche	52
2.3.2	Simulations à l'échelle de la communauté	64
2.3.3	Interactions bactériennes et exploration du modèle de communauté	70
2.4	Discussion et Conclusion	76

Les travaux présentés dans ce chapitre ont été publiés dans le journal *Metabolic Engineering*¹ (Lecomte et al., 2024)

Une liste des abréviations propres au métabolisme du consortium est disponible à la fin de ce chapitre section 2.4.

1. <https://www.sciencedirect.com/journal/metabolic-engineering>

2.1 Introduction

Ce chapitre s'intéresse à la possibilité pour un modèle numérique, tel que nous l'avons défini dans le chapitre précédent, de répondre à plusieurs enjeux biologiques avec un critère d'explicabilité satisfaisant. Ce modèle numérique s'inscrit dans le projet TANGO coordonné par l'INRAE STLO (Cao et al., 2021). Différentes productions de fromage à pâte pressée non cuite de 2,5kg ont été réalisées en triplicat, avec un unique consortium composé de deux bactéries lactiques et une bactérie propionique, mais en suivant des itinéraires légèrement modifiés. Ces changements ont permis de mettre en évidence, de façon expérimentale, l'impact de ces itinéraires sur l'arôme du fromage. Nous proposons dans ce chapitre, d'identifier, à l'aide d'un modèle numérique, les voies métaboliques conduisant à la production des composés d'arômes ainsi que les métabolites échangés entre les espèces dans l'itinéraire standard de fabrication.

Il existe des modèles numériques du métabolisme, garantissant la fiabilité des résultats ainsi que la précision mais ces modèles sont souvent des "boîtes noires" ne permettant pas d'obtenir une explication mécanistique des résultats (Popp and Centler, 2020; Zomorodi, 2014) notamment à travers les voies métaboliques activées. Popp et al précisent d'ailleurs que les résultats de simulation ne doivent pas engendrer une interprétation biologique détaillée étant donnée l'inférence de paramètres génériques. Une conséquence est la difficulté à inférer des processus de régulation biologique connus dans la littérature ou encore, d'intégrer des données hétérogènes. Il existe des méthodes permettant l'intégration de données omiques ou cinétiques, mais séparément. Ainsi, les travaux de (Jenior et al., 2020) permettent d'intégrer des données méta-transcriptomiques au sein de modèles métaboliques à l'échelle du génome (GEM). Au moyen de ces données et d'une analyse parcimonieuse de la distribution des flux dans un réseau métabolique, le chemin le plus coûteux représentant l'investissement d'une cellule dans la transcription est identifié. Avec des données de croissance cinétique, de pH ou encore de métabolomique, Ozcan et al (Özcan et al., 2020) ont construit un modèle dynamique prédisant des concentrations de métabolites, et ont inféré des mécanismes biologiques appris à partir des données. Ayant à disposition des données hétérogènes similaires à (Özcan et al., 2020), nous avons développé une approche numérique prenant en compte l'hétérogénéité des données disponibles et garantissant une explicabilité des résultats obtenus.

En s'inspirant de la stratégie développée par (Özcan et al., 2020), optimisant les modèles individuels et de communautés avec les données, nous avons développé une stratégie itérative et explicative permettant d'intégrer un ensemble de données : génomique de chaque souche, métabolomique, pH, croissance de bactéries, et des dosages de métabolites. Cette stratégie a consisté à extrapoler les modèles individuels optimisés à partir des données et de la littérature pour prédire les comportements en communauté de chaque souche bactérienne, les concentrations de métabolites bactériens produits pendant la fermentation, la biomasse, ainsi que les interactions basées sur l'échange de métabolites entre espèces. Notre modèle numérique repose sur l'étude du métabolisme de trois souches bactériennes, que sont *Propionibacterium freudenreichii* CIRM-BIA122, *Lactiplantibacillus plantarum* CIRMB-BIA465 et *Lactococcus lactis* biovar *diacetyl lactis* CIRM-BIA1206. Tout d'abord, nous avons reconstruit le métabolisme de chaque souche microbienne en utilisant les données génomiques, puis procédé au raffinement manuel des réseaux métaboliques au moyen de la littérature et des données de dosage. Dans un second temps, une étape de calibration dynamique, utilisant les données de dosages, de pH, de croissance et la littérature, a permis de garantir la précision et la fiabilité de chaque modèle individuel

(voir Figure 2.1).

Les travaux présentés dans ce chapitre ont mené à la publication d'un article² de journal (Lecomte et al., 2024) et de plusieurs présentations scientifiques et posters (Voir Annexe .1).

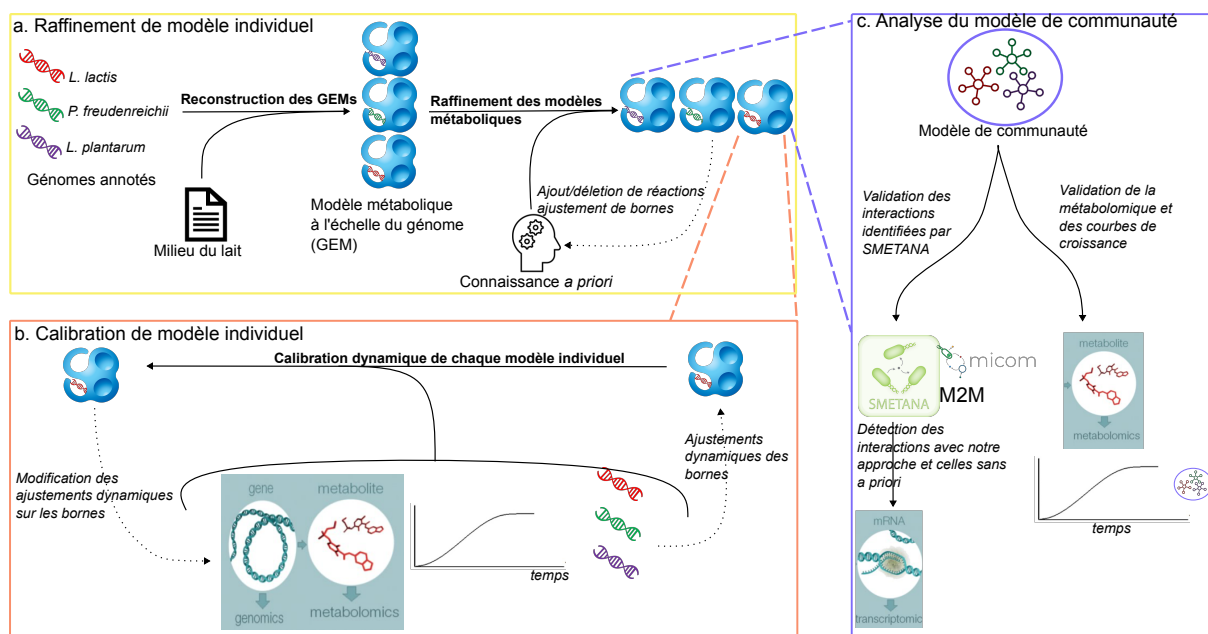


FIGURE 2.1 – Récapitulatif des étapes de raffinement et de calibration des modèles métaboliques, des simulations ainsi que des données utilisées. Le cadre en jaune représente l'étape de raffinement des modèles métaboliques individuels ; en rouge, l'étape de calibration dynamique des modèles individuels ; en violet, l'étape de validation du modèle de communauté. Les flèches en pointillés représentent les boucles itératives des deux processus.

2.2 Méthode pour la construction d'un modèle numérique du métabolisme

Afin de construire un modèle numérique du métabolisme, les réseaux métaboliques à l'échelle du génome doivent être reconstruits à partir d'un génome annoté, puis curés et idéalement calibrés avec ces données multi-omiques. Nous décrivons dans un premier temps l'ensemble des données nécessaires, puis l'étape de reconstruction d'ébauche de réseaux métaboliques. Dans un second temps, nous verrons en détail l'approche itérative développée au moyen de l'étape de curation et de calibration dynamique des modèles.

2.2.1 Présentation des données

2.2.1.1 Données omiques

Génomique Les souches utilisées sont décrites dans (Cao et al., 2021). Brièvement, la communauté bactérienne contrôlée est composée de deux bactéries lactiques (LAB),

2. <https://www.sciencedirect.com/science/article/pii/S1096717624000302>

L. lactis subsp. *lactis* biovar *diacetylactis* CIRMBIA1206, *L. plantarum* CIRMBIA465, et une bactérie propionique *P. freudenreichii* CIRMBIA122 provenant du Centre International de Ressources Microbiennes des Bactéries d'Intérêt Alimentaire (CIRM-BIA³). Leur génome a été séquencé, assemblé et a été intégré sur la plateforme MicroScope hébergée au Genoscope (CEA, Évry, France) pour une annotation automatique selon (Vallenet et al., 2019). Les génomes annotés sont disponibles à l'ENA (EBI, Cambridge) sous le numéro d'accèsion PRJEB54980.

Métatranscriptomique Un protocole détaillé est décrit dans les travaux de (Cao et al., 2021), nous ne décrivons ici que les étapes principales. Les données métatranscriptomiques ont été obtenues à cinq étapes de la fabrication du fromage : moulage, démoulage, après saumurage et affinage à 4 et 7 semaines. Les lectures sont disponibles à l'ENA (EBI, Cambridge) sous le numéro d'accèsion PRJEB42478. Ces lectures ont été alignées sur un génome de référence afin de déduire des informations relatives à l'expression des gènes, puis une étape de comptage a lieu. Les données brutes de comptage de RNASeq ont été normalisées en deux étapes. Tout d'abord, un facteur d'échelle spécifique à l'espèce a été appliqué pour éliminer les biais de composition entre les différentes bibliothèques utilisées en utilisant la méthode TMM (Trimmed Mean of M-Values) telle qu'implémentée dans le package edgeR version 3.32.1 (Robinson et al., 2009). Une étape supplémentaire de normalisation (de type RPKM) a été réalisée par échantillon pour corriger la longueur des gènes et permettre la comparaison. Enfin, la moyenne des réplicats a été calculée.

Métabolomique Les méthodes sont décrites dans (Cao et al., 2021). Brièvement, les sucres et les acides organiques dans les échantillons ont été quantifiés en utilisant la chromatographie liquide à haute performance (HPLC). Les composés volatiles ont été analysés en utilisant l'extraction par sorption de l'espace de la tête (HS) et couplés à la chromatographie en phase gazeuse-spectrométrie de masse (GC-MS).

2.2.1.2 Données de calibration des modèles individuelles

Un suivi cinétique de la croissance bactérienne et du pH a été réalisé sur lait pour chaque souche individuelle. Les concentrations en acides organiques dans la culture de *P. freudenreichii* ont été évaluées en conditions microaérophiles (Table 2.2). Le milieu de culture de *P. freudenreichii* a été supplémenté en lactate pour reproduire la croissance en co-culture et en hydrolysats de protéine du lait simulant sa protéolyse. En prenant en compte toutes ces conditions, chacune des souches a atteint un seuil de cultivabilité supérieure à $8 \log_{10}$ CFU/g. Le pH, d'une valeur de 6.7 lors de l'inoculation des deux bactéries lactiques, a atteint respectivement 5.7 et 5.1 pour *L. plantarum* et *L. lactis*. Le début de la phase plateau, correspondant à l'entrée en phase stationnaire, a été atteint en 48, 14 et 8 heures pour *P. freudenreichii*, *L. plantarum* et *L. lactis* respectivement. Les données de dosage des métabolites de *P. freudenreichii* montrent une production de propionate élevée (8 grammes par litre de lait), une forte consommation de lactate (9 grammes par litre de lait) et une faible production d'acétate et de succinate, respectivement 3,07 grammes par litre de lait et 0,371 gramme par litre de lait (2.1).

Données cinétiques de croissances et de pH Pour chaque souche nous avons collecté les données de croissance et de pH qui ont servi à calibrer dynamiquement notre

3. <https://collection-cirmbia.fr/>

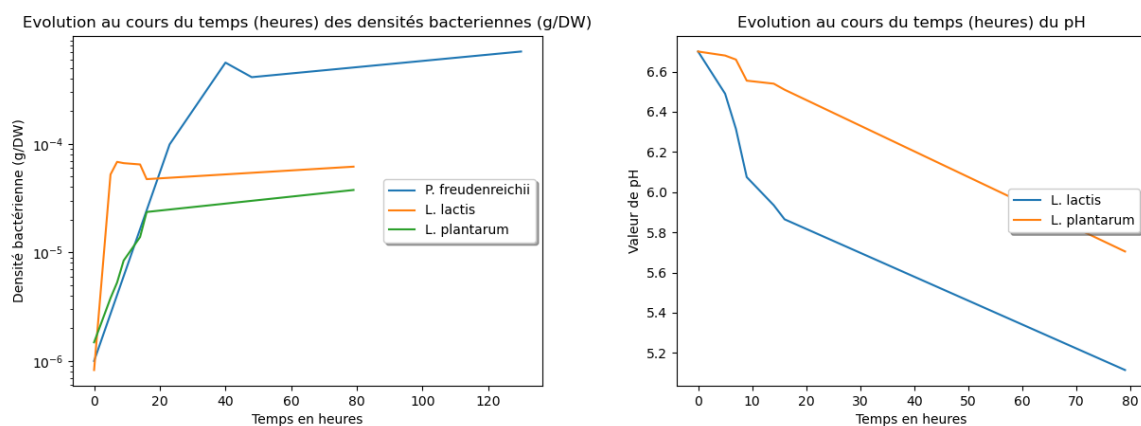


FIGURE 2.2 – **Données de cultures pures utilisées pour calibrer chaque souche.** Le temps est en heures et les valeurs de densités bactériennes représentent des moyennes sur les différents réplicats en g.DW^{-1} .

modèle. Elles sont représentées dans la table 2.2.

Les données de dosage d'acide Les données de dosages des acides lactiques, de l'acétate, du succinate et du propionate lors de l'inoculation de la souche et en fin de fermentation de *P. freudenreichii* sont représentées dans la table 2.1.

Acides	Concentration T_0	Concentration $T_f = 89 \text{ h}$	Erreur standard
Lactate	16.5	7.88	0.08
Acetate	0	3.07	0.02
Succinate	0	0.371	0.063
Propionate	0	8.31	0.02

TABLE 2.1 – **Données de dosage d'acide en g/L pour *P. freudenreichii*.**

2.2.1.3 Données de validation du modèle de communauté

Données de croissances et de pH en communauté Au regard de notre stratégie itérative de calibration des modèles individuels puis validation en communauté, les données représentées dans la table 2.3 ont servi de données tests pour évaluer la précision du modèle. Toutes ces données proviennent de l'article de (Cao et al., 2021). Un suivi de la croissance des bactéries ainsi que de la production des métabolites ont été réalisés durant la fabrication du fromage. Les seuils de cultivabilité de *L. lactis*, *L. plantarum* et *P. freudenreichii* ont atteint respectivement $8.45 \log_{10}\text{CFU/g}$, $8.47 \log_{10}\text{CFU/g}$ et $8.59 \log_{10}\text{CFU/g}$ en approximativement 1200 heures.

Conversion des unités Lors de la calibration, nous avons appliqué un coefficient multiplicateur de 0.33×10^{-12} transformant les CFU en grammes de matière sèche. En effet, d'après (Bakken and Olsen, 1983), 1cm^3 de bactéries équivaut à 0.33 de matière sèche de biomasse et en supposant que 1 CFU a pour volume $1\mu\text{m}^3$, nous avons donc : $1\text{CFU} = 10^{-12}\text{cm}^3$ soit $0.33 \times 10^{-12}\text{g.DW}^{-1}$. Nous avons également converti les unités des métabolites suivis en dynamique, de g/kg de lait en mmol/g de lait.

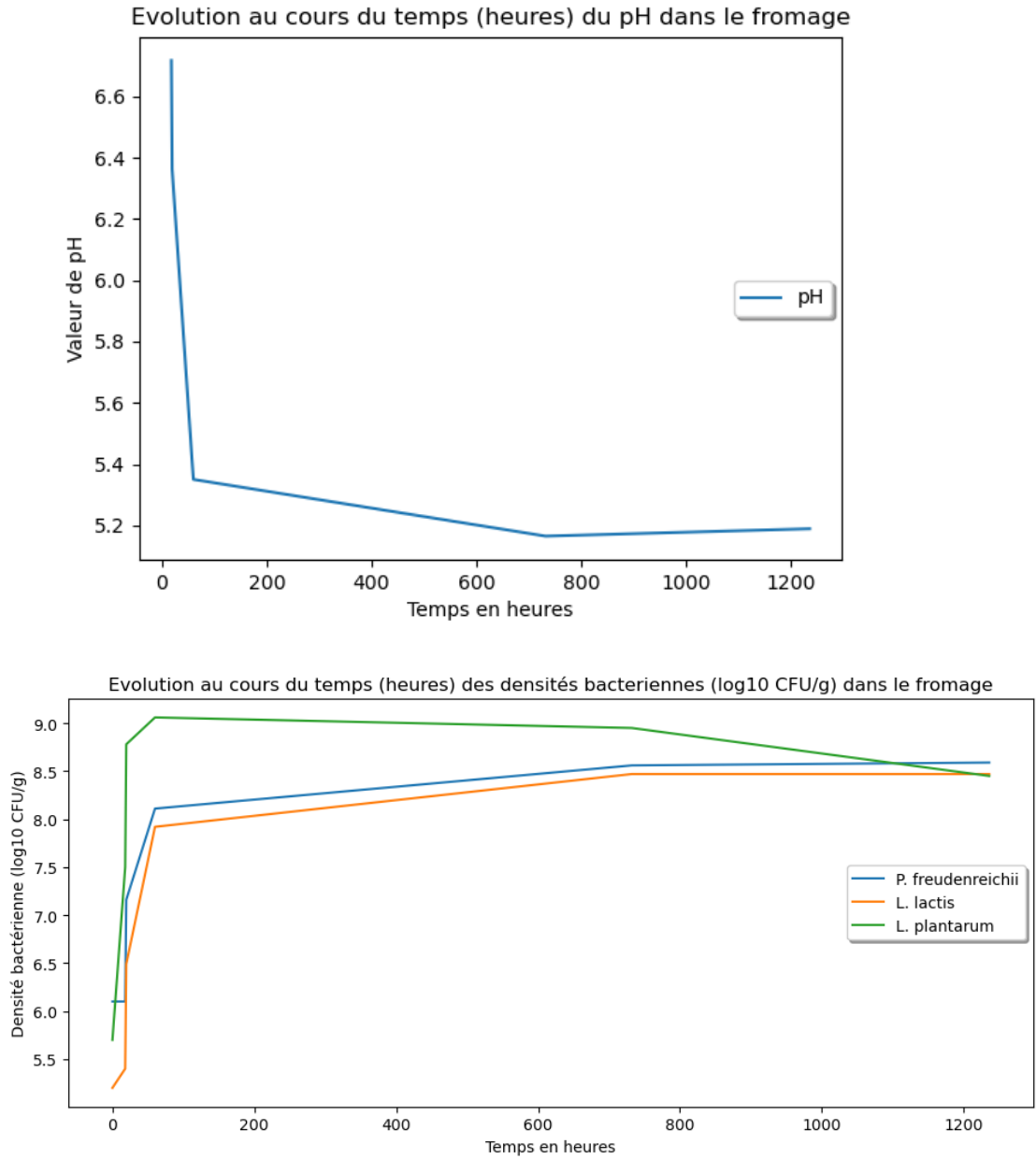


FIGURE 2.3 – Données de croissances (en bas) et de pH (en haut) en co-culture pour tester les prédictions de notre modèle de communauté.

2.2.2 Du génome annoté au réseau métabolique à l'échelle du génome

Pour construire un modèle numérique du métabolisme de la production du fromage, nous avons reconstruit les réseaux métaboliques à l'échelle du génome avec l'outil de reconstruction CarveMe (Machado et al., 2018) version 1.4.1 et enrichi l'annotation existante avec ModelPolisher (Römer et al., 2016) version 2.0 en ajoutant des métadonnées liées à la base de connaissance BIGG (King et al., 2016). Afin de vérifier la qualité des GEM, chaque modèle métabolique a été validé avec MEMOTE (Lieven et al., 2020) version 0.13.0.

2.2.3 Modélisation et implémentation numérique

Analyse statique

Pour permettre une analyse numérique des réseaux métaboliques, nous avons développé notre stratégie numérique itérative en nous basant sur l'analyse par équilibre des flux (FBA)(Orth et al., 2010). Brièvement, à partir de la composition nutritionnelle du lait, le FBA calcule une distribution de flux optimale satisfaisant une hypothèse d'état quasi-stationnaire, des contraintes thermodynamiques, et maximisant une fonction objective, ici la réaction de biomasse.

Plus formellement, l'objectif de la résolution du problème FBA est de :

$$\begin{aligned} \text{trouver } v^* \in \mathbb{R}^{N_r} \text{ tel que } v^* = & \quad \arg \max & \quad v_{\text{biomasse}} & \quad (2.1) \\ & \quad S.v = 0 \\ & \quad v_{\min}^{\text{in}} \leq v \leq v_{\max}^{\text{in}} \\ & \quad v_{\min}^{\text{ex}} \leq v \leq v_{\max}^{\text{ex}} \end{aligned}$$

où v_{biomasse} est le flux de la réaction de biomasse du modèle en $\text{mmol.h}^{-1}.\text{g}_{DM}^{-1}$, S est la matrice stœchiométrique dérivée du modèle métabolique, v_{\min}^{in} et v_{\max}^{in} (resp. v_{\min}^{ex} et v_{\max}^{ex}) définissant les bornes de flux intracellulaires ou d'échanges pour le vecteur de flux v dans \mathbb{R}^{N_r} .

L'équation (2.2) définit alors, pour chaque bactérie i , la correspondance μ_i entre les vecteurs de contraintes sur les réactions d'échange définissant l'environnement nutritionnel et le flux optimal v^* obtenu dans ce contexte nutritionnel.

$$\mu_i(v_{\min}^{\text{ex}}, v_{\max}^{\text{ex}}) = v^* \quad (2.2)$$

Dans un souci de garantir une solution unique, l'ordre lexicographique (Gomez et al., 2014) a été utilisé lors de l'optimisation. Nous avons défini la direction du flux des métabolites suivis dynamiquement à optimiser, définissant de nouvelles contraintes à satisfaire et réduisant l'espace de solution.

Modélisation dynamique

Le modèle dynamique de la communauté microbienne a été développé en s'inspirant de l'analyse dynamique par équilibre des flux (dFBA) de (Mahadevan et al., 2002a) et en ajoutant des mécanismes de dynamiques de population au modèle FBA.

Le premier mécanisme décrit dynamiquement est le *quorum sensing*, correspondant au processus de régulation de la croissance de la population bactérienne. Notons $b = (b_i)$

pour $i \in \mathcal{B} = \{L. lactis, L. plantarum, P. freudenreichii\}$ les densités de la population bactérienne (g/L de lait) et \mathcal{R}_i le vecteur modélisant ce processus, le *quorum sensing* a été modélisé de la façon suivante :

$$\mathcal{R}_i = \lambda_i \left(1 - \frac{b_i}{\beta_i}\right), \text{ for } i \in \mathcal{B} \quad (2.3)$$

où λ_i est un poids et β_i est la capacité de charge de la population, correspondant à la valeur de la phase plateau de chaque bactérie dans l'expérience (g/L de lait).

Nous avons par la suite défini des flux d'import de composés présents dans le modèle. Pour un substrat j , un modèle métabolique i et $m = (m_j)_{1 \leq j \leq N_m}$ la concentration m (mmol/g de lait) de composés métaboliques N_m qui sont expérimentalement suivis, le flux de la borne inférieure $c_{min,i,j}^{ex}$ et supérieure $c_{max,i,j}^{ex}$ a été défini comme suit :

$$c_{min,i,j}^{ex} = \max\left(-\frac{m_j}{\Delta t \sum_{i \in \mathcal{M}_j} b_i}, v_{i,j}^{int}\right) \quad c_{max,i,j}^{ex} = 0, \quad (2.4)$$

où \mathcal{M}_j est le sous-ensemble de bactéries métabolisant j , et Δt est un l'intervalle temporel d'importation (en heures). Ces équations reflètent le partage uniforme des ressources parmi les bactéries durant la fenêtre temporelle Δt . Quand les substrats sont en excès, le flux est contraint par la limite d'import intrinsèque $v_{i,j}^{int}$ ($mmol.g_{DM}^{-1}.h^{-1}$) et est partagé équitablement entre les bactéries le consommant : $-\frac{m_j}{\Delta t \sum_{i \in \mathcal{M}_j} b_i}$.

Nous avons par la suite modulé le flux des bornes inférieures $c_{min,i,j}^{ex}$ des composés que l'on suit en dynamique et pour lesquels nous disposons de données biochimiques. En s'inspirant de l'approche de (Özcan et al., 2020), nous avons régulé le métabolisme du lactose chez les bactéries lactiques. Il est modulé négativement par la forme non dissociée de l'acide lactique avec une décroissance exponentielle. Pour $j = lcts_e$ et $i \in \{L. lactis, L. plantarum\}$, les flux des bornes inférieures $c_{min,i,j}^{ex}$ sont définis par :

$$c_{min,i,j}^{ex} = \max\left(-\frac{m_{lcts_e}}{\Delta t * \sum_{i \in \mathcal{M}(lcts_e)} b_i}, -\mu_{max,lcts} * 10^{(-k_{lac} * \phi_{undiss})} - \mu_{min,lcts}\right)$$

où k_{lac} correspond à la décroissance exponentielle régulant la consommation du lactose, $\mu_{max,lcts}$ et $\mu_{min,lcts}$ sont les valeurs maximales et minimales respectivement du flux de consommation du lactose, et ϕ_{undiss} est une fonction calculant la concentration de la forme non dissociée de l'acide lactique (mmol/g de lait) à partir de celle du lactate. Cette fonction ϕ_{undiss} est dérivée de l'équation de Henserson-Hasselbalch comme dans (Özcan et al., 2020) et s'écrit

$$\phi_{undiss}(m_{lac_L_e}, m_{lac_D_e}) = \frac{m_{lac_L_e} + m_{lac_D_e}}{1 + 10^{c_1 * (m_{lac_L_e} + m_{lac_D_e}) + c_2}}. \quad (2.5)$$

Dans cette équation, le terme $c_1 * (m_{lac_L_e} + m_{lac_D_e}) + c_2$ est une approximation linéaire de $(pH - pKa)$, alors que les paramètres c_1 et c_2 sont inférés par une régression linéaire des données de co-culture. Cette approximation est déduite de (Özcan et al., 2020), où les auteurs expriment la concentration de la forme non dissociée de l'acide lactique, à partir des concentrations de l'acide lactique, du pH et du pKa, de la façon suivante :

$$[LacH] = \frac{[Lac]}{1 + exp^{pH - pKa}}$$

où $[LacH]$ est la concentration de la forme non dissociée de l'acide lactique et $[Lac]$ la concentration de l'acide lactique. En supposant que le pH est linéairement corrélé à la

concentration de l'acide lactique, la concentration de la forme non dissociée de l'acide lactique est calculé de la façon suivante :

$$[LacH] = \frac{[lac]}{1 + exp^{c1 \times [Lac] + c2}}$$

De ces deux équations, nous avons pu écrire un modèle de pH en fonction des concentration des acides lactiques comme suit :

$$pH = pKa + c1 \times [Lac] + c2$$

La production de lactate est régulée de la même manière, alors que la production d'acétate est régulée en fonction de la disponibilité en lactose.

Pour *P. freudenreichii*, les bornes supérieures de production sont calculées à partir des dosages de métabolites et des données de croissance dans les expériences de culture pure, en supposant un flux constant.

Pour $i \in \mathcal{B}$ et $1 \leq j \leq N_m$, nous avons donc le système final suivant, décrivant la densité de bactérie $\partial_t b_i$ et la concentration de métabolites (mmol/g de lait) $\partial_t m_j$ au cours du temps :

$$\partial_t b_i = \mathcal{R}_i(b_i) \mu_{i,i}((c_{min,i}^{ex}, c_{max,i}^{ex})(b, m)) b_i \quad (2.6)$$

$$\partial_t m_j = \sum_{i \in \mathcal{B}} \mu_{i,j}((c_{min,i}^{ex}, c_{max,i}^{ex})(b, m)) b_i \quad (2.7)$$

dans lequel le terme $\mu_{i,i}$ est la composante correspondant au métabolite j (ou à la biomasse i) de μ_i , calculée à partir de 2.2 compte tenu de l'ensemble des contraintes $(c_{min,i}^{ex}, c_{max,i}^{ex})(b, m)$. Ces contraintes dépendent dynamiquement de l'état de la variable b et m , c'est-à-dire, de la densité de b au cours du temps ($\partial_t b_i$) et de la quantité de m disponible calculée par $\partial_t m_i$.

Implémentation numérique

Le système dynamique est résolu avec le schéma semi-implicite d'Euler. À chaque étape n , après le calcul de F_j correspondant au flux total du composé j comme défini par l'ensemble des GEMs,

$$F_j = \sum_{i \in \mathcal{B}} \mu_{i,j}((c_{min,i}^{ex}, c_{max,i}^{ex})(b^n, m^n)) b_i,$$

nous calculons les densités bactériennes (b_i^{n+1}) et la concentration du composé j (m_j^{n+1}) au temps suivant :

$$b_i^{n+1} = b_i^n + \Delta t * F_{b_i}$$

$$m_j^{n+1} = \begin{cases} m_j^n + \Delta t * F_j & \text{si } F_j > 0 \text{ (cas explicite)} \\ m_j^n / (1 - \Delta t * F_j / m_j^n) & \text{sinon (cas implicite)} \end{cases}$$

Ce schéma semi-implicite d'Euler a été utilisé pour garantir la positivité de la solution. En effet, en utilisant uniquement le cas explicite, si F_j est négatif alors la densité de la bactérie b_i^{n+1} le sera également. Le cas implicite seul fonctionnerait mais augmenterait le temps de calcul. Nous avons établi un compromis en développant ce schéma semi-explicite.

2.2.4 Raffinement des modèles métaboliques individuels

Pour retrouver les comportements observés dans la littérature, les réactions à ajouter ou à contraindre dans les GEMs diffèrent mais le processus itératif de vérification reste le même. Au moyen d'une analyse par équilibre des flux (FBA) (Orth et al., 2010) maximisant la biomasse de chaque souche, les voies métaboliques activées, c'est-à-dire celles pour lesquelles un flux non nul est présent, peuvent être détectées. La première étape consiste à vérifier si toutes les voies métaboliques d'intérêt sont fonctionnelles de manière qualitative, *i.e.* observation de production, de consommation et activation ou non de voies métaboliques. Dans la cas d'une non-activation *in silico* de voies métaboliques dans un modèle, nous avons développé le protocole suivant : détection de fuite de flux symbolisée par des valeurs de flux extrêmes, identification d'activation de voies métaboliques incohérentes avec l'état des connaissances de l'organisme, ajustement des valeurs de bornes de régulation ou blocage de réactions. Pour chaque modification apportée, une vérification de la distribution des flux est faite jusqu'à obtenir les observations souhaitées.

2.2.5 Calibration des modèles métaboliques individuels

Afin de retrouver numériquement les valeurs expérimentales de chaque souche, nous avons calibré chacun de nos modèles métaboliques. Pour rappel, notre stratégie a consisté à obtenir des modèles individuels suffisamment précis pour déduire des comportements mécanistiques à l'échelle de la communauté. Nous avons donc estimé des paramètres propres à chaque souche, modifiant les flux d'import des métabolites d'intérêt, en résolvant le problème d'optimisation visant à minimiser la fonction coût J suivante, pour les bactéries lactiques (équations (2.8))

$$J(b_i, pH|\theta_i, b_{i,exp}, pH_{exp}) = \left\| \frac{\log_{10}(b_i) - \log_{10}(b_{i,exp})}{\sigma_{\log,i,exp}} \right\|^2 + \alpha \left\| \frac{pH - pH_{exp}}{\sigma_{pH,exp}} \right\|^2 \quad (2.8)$$

où σ (resp. $\sigma_{\log,i}$) est l'écart-type (resp. l'écart-type log transformé) des données correspondantes, et $\theta_i = (k_{lac}, \mu_{max,lcts})$ si $i = L. lactis$ et $\theta_i = (k_{lac}, \mu_{max,lcts}, \lambda_i)$ si $i = L. plantarum$, b_i correspond à la densité de la population bactérienne ($g_{DM}^{-1} \cdot g^{-1}$ de lait) de i , pour $i \in \{L. lactis, L. plantarum\}$ et α un poids.

Concernant *P. freudenreichii*, la fonction coût J cherchant à minimiser l'écart aux données est représentée par l'équation (2.9) :

$$J(b, m|\theta_i, b_{exp}, m_{exp}) = \left\| \frac{\log_{10}(b) - \log_{10}(b_{exp})}{\sigma_{\log,b,exp}} \right\|^2 + \alpha \left\| \frac{m - m_{exp}}{\sigma_{m,exp}} \right\|^2 \quad (2.9)$$

où m_{exp} est le dosage final d'acétate, lactate, propionate et succinate, et $\theta_i = (\lambda_i)$.

Afin d'éviter la sur-prédiction, nous avons délibérément optimisé un petit nombre de paramètres θ pour chaque souche. L'ensemble des paramètres ainsi que leurs rôles sont décrits dans la table 2.2 :

GEM	Paramètres	Explication
<i>P. freudenreichii</i>	λ_i	Modification des bornes d'import des nutriments suivis en dynamique
<i>L. plantarum</i>	k_{lac}	Paramètre de la régulation de la consommation du lactose en fonction de la concentration de lactate
	$\mu_{max,lcts}$	Paramètre de la régulation de la consommation du lactose en fonction de la concentration de lactate
	λ_i	Modification des bornes d'import des nutriments suivis en dynamique
<i>L. lactis</i>	k_{lac}	Paramètre de la régulation de la consommation du lactose en fonction de la concentration de lactate
	$\mu_{max,lcts}$	Paramètre de la régulation de la consommation du lactose en fonction de la concentration de lactate

TABLE 2.2 – Ensemble des paramètres optimisés pour chaque modèle individuel avec leur justification.

2.3 Application de la méthode numérique à l'écosystème fromager

Nous avons testé notre méthodologie itérative de raffinement et de calibration sur les 3 souches bactériennes présentées plus haut. L'objectif de cette application est triple : d'une part, montrer qu'à partir de modèles métaboliques individuels nous pouvons prédire numériquement des interactions, des concentrations de métabolites et de densités bactériennes, et d'autre part, être suffisamment "boîte-blanche" pour fournir des hypothèses testables sur des voies activées pour la production d'arômes dans le milieu extra-cellulaire. Enfin, ce modèle numérique doit être capable d'intégrer les données hétérogènes décrites plus haut. Dans la suite de ce chapitre, nous allons dans un premier temps nous focaliser sur les résultats des processus itératifs de raffinement et de calibration de modèles individuels. Une explication minutieuse des modifications réactionnelles et une validation souche par souche seront présentées. Puis, nous montrerons comment les données de cultures pures ont permis de calibrer dynamiquement les modèles. Dans un second temps, nous présenterons l'implémentation du modèle de communauté ainsi que son exploration sur le plan des croissances bactériennes, de la production des composés d'arômes suivie en dynamique ainsi que sur les potentielles interactions métaboliques. Enfin, nous explorerons les possibles interactions bactériennes du modèle avec des outils sans *a priori* : MICOM, SMETANA et Metage2Metabo.

2.3.1 A l'échelle de chaque souche

Raffinement des modèles FBA individuels

Les réseaux métaboliques à l'échelle du génome ont été reconstruits pour chaque souche individuelle à partir de leurs séquences génomiques annotées. Après une première étape automatique de reconstruction faite avec CarveMe, nous avons tout d'abord ajusté les flux des métabolites présents dans les réseaux métaboliques en cohérence avec la composition du lait. Les bornes inférieures des réactions d'échanges de composés qui n'apparaissent pas dans la composition du lait, sont ainsi mises à 0. Cette liste est constituée des réactions d'échange du glucose ($EX_glc_D_e$) et du coenzyme A (EX_coa_e) pour *L. lactis* et *L. plantarum*, de l'amidon ($EX_starch1200_e$) pour *L. lactis* et *P. freudenreichii*, du dihydroxyacétone (EX_dha_e) pour *L. plantarum* et *P. freudenreichii* et du galactose (EX_gal_e) pour *L. plantarum*.

Nous avons ensuite réalisé un raffinement manuel du métabolisme carboné central basé sur la littérature et la connaissance biologique dans le but de prendre en compte les spécificités de chaque souche bactérienne. Le métabolisme carboné central produit du NADH,

du NADPH et ses formes réduites correspondantes, de l'ATP, et les précurseurs nécessaires pour la biosynthèse des molécules essentielles pour la croissance ainsi que pour l'activité métabolique des cellules bactériennes, e.g. acides aminés, purine, pyrimidine, glycerol-3-phosphate, acides gras, N-acetyl-glucosamine, vitamines. Parmi les voies métaboliques du métabolisme carboné central, la glycolyse et la voie des pentoses phosphate sont communes aux trois espèces et sont presque identiques. Les sources de carbone telles que le lactose, l'acide citrique et l'acide lactique, sont converties en pyruvate, qui sera utilisé dans des voies spécifiques représentées dans les Figures 2.6,2.4,2.5. Dans la suite du chapitre, les caractéristiques métaboliques de chaque souche seront étudiées afin de justifier les raffinements réalisés, en se focalisant sur les voies métaboliques secondaires permettant la production de composés organoleptiques.

Bactéries lactiques Les bactéries lactiques *L. lactis* et *L. plantarum* utilisent le lactose et le citrate, les deux principales sources de carbone du lait (Widyastuti et al., 2014). Le lactose est dégradé par l'enzyme beta-galactosidase en galactose et glucose, et ce dernier alimentent la glycolyse. Concernant le galactose et l'acide citrique, leur métabolisme dépend de la souche étudiée (Palles et al., 1998). D'un côté, le lactose est converti en glucose-6-phosphate *via* la voie métabolique de Leloir avant d'entrer dans la glycolyse, et de l'autre l'acide citrique est converti en pyruvate par l'enzyme citrate lyase (Roissart, 1994). Du point de vue du métabolisme secondaire, le butanediol et le diacétyle sont responsables de la saveur beurrée du fromage, et sont produits par les deux modèles à partir du pyruvate grâce à la voie de l'acétolactate. L'ensemble des voies métaboliques décrites ci-dessus est récapitulé dans les Figures 2.4 et 2.5.

L. lactis Pour permettre la production de certains composés d'intérêt, comme le butanediol, le GEM a été manuellement curé (Fig. 2.4). En reprenant la méthode de raffinement de modèle décrite plus haut, le flux de l'acétoine-dehydrogenase a été bloqué (ACTD2) permettant la production du butanediol. Les bornes de l'acétolactate decarboxylase (ACLDC) et de l'acétolactate synthase (ACLS) ont été modifiées de façon à permettre l'activation de la voie de l'acétolactate, comme retrouvé dans la littérature (Carroll et al., 1999; Makhlof, 2006; Swindell et al., 1996). Enfin, la consommation de lactose a été régulée en modifiant les bornes de la réaction LACZ ce qui a permis un flux de production de lactate pour *L. lactis* (Fig. 2.4).

Pour valider ce modèle, nous avons considéré un exemple d'activation de voie métabolique durant la production du fromage, validable au regard de la littérature et des données de métatranscriptomiques. Il est montré que *L. lactis* consomme le lactose à la fois par la voie du tagatose, codée dans l'opéron plasmidique *lacABCD*, ou par la voie de Leloir, codée dans l'opéron du chromosome *galKTE*. D'un point de vue des données métatranscriptomiques, *L. lactis* exprime les gènes associés à la voie du tagatose au début de la production de fromage quand la concentration de lactose est grande, et ceux de la voie de Leloir durant l'affinage quand la concentration de lactose est faible. Cela suggère fortement que la concentration de lactose détermine quelle voie métabolique est employée, et prédit qu'une grande quantité de lactose dans le milieu nutritionnel devrait produire un flux important à travers la voie du tagatose. Le modèle FBA de *L. lactis* est en accord avec les données métatranscriptomiques. Nous avons observé un flux important de dégradation du lactose à travers la voie du tagatose (Fig. 2.7 a).

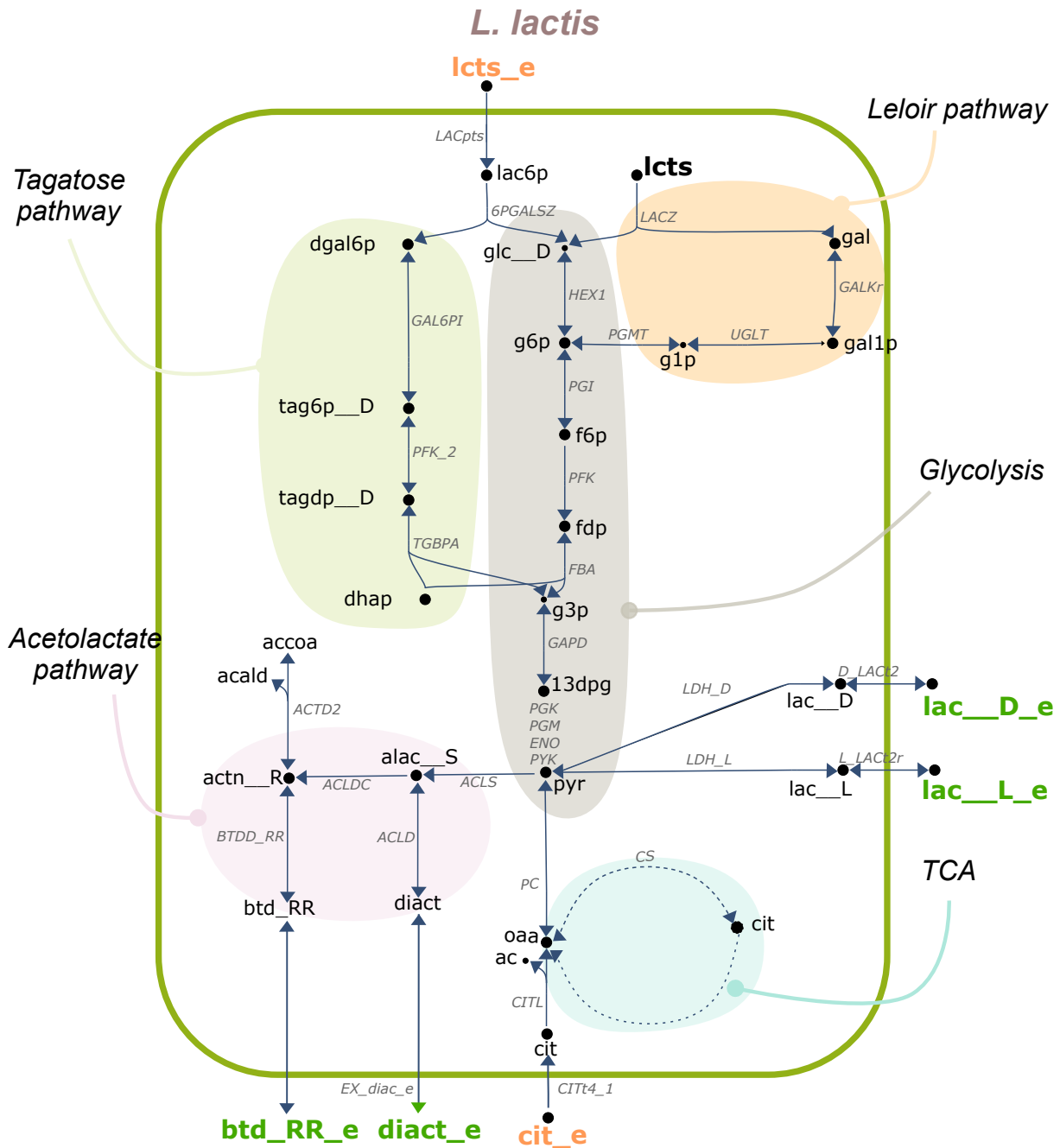


FIGURE 2.4 – Cartes des voies métaboliques du métabolisme carboné central d'intérêt de *L. lactis*. La voie métabolique de la glycolyse est représentée en gris, celle de Leloir en saumon, acétolactate en rose, le cycle du TCA en bleu clair, et le tagatose en vert. Les métabolites en orange et vert correspondent aux entrées et sorties du modèle métabolique. La liste des métabolites et des réactions est répertoriée à la fin du chapitre à la section 2.4

L. plantarum L'exemple pris pour *L. plantarum* concerne la dégradation du glucose et du galactose dans le cas des fermentations homolactique et hétérolactique (voir figure 2.5). Dans les deux configurations, le pyruvate est produit à partir de la dégradation du glucose et du galactose. Durant la fermentation homolactique, le pyruvate est associé au relargage de l'acide lactique dans le milieu extracellulaire. Concernant la fermentation

hétérolactique, on observe en plus une production d'éthanol et d'acétate. Nous savons que *L. plantarum* est une bactérie hétérolactique pouvant produire de l'acétate grâce à l'enzyme transketolase (numéro EC 4.1.2.9) à travers la voie métabolique de l'acétate kinase comme décrit dans (Abedi and Hashemi, 2020). Pour permettre cette production dans notre modèle, nous avons bloqué la pyruvate formate lyase (PFL) (Quatravaux et al., 2006). Concernant le butanediol, la même curation que pour *L. lactis* a été effectuée. Enfin, pour reproduire la dégradation du galactose et du glucose par la voie de Leloir et de la dégradation du lactose grâce à la β -galactosidase *LACZ*, la réaction *LCTSt3ipp* a été bloquée et la réaction UTP-glucose-1-phosphate uridylyltransferase (*GALU_i*) a été rendue réversible pour compléter la voie de dégradation du galactose (Fig. 2.8 a).

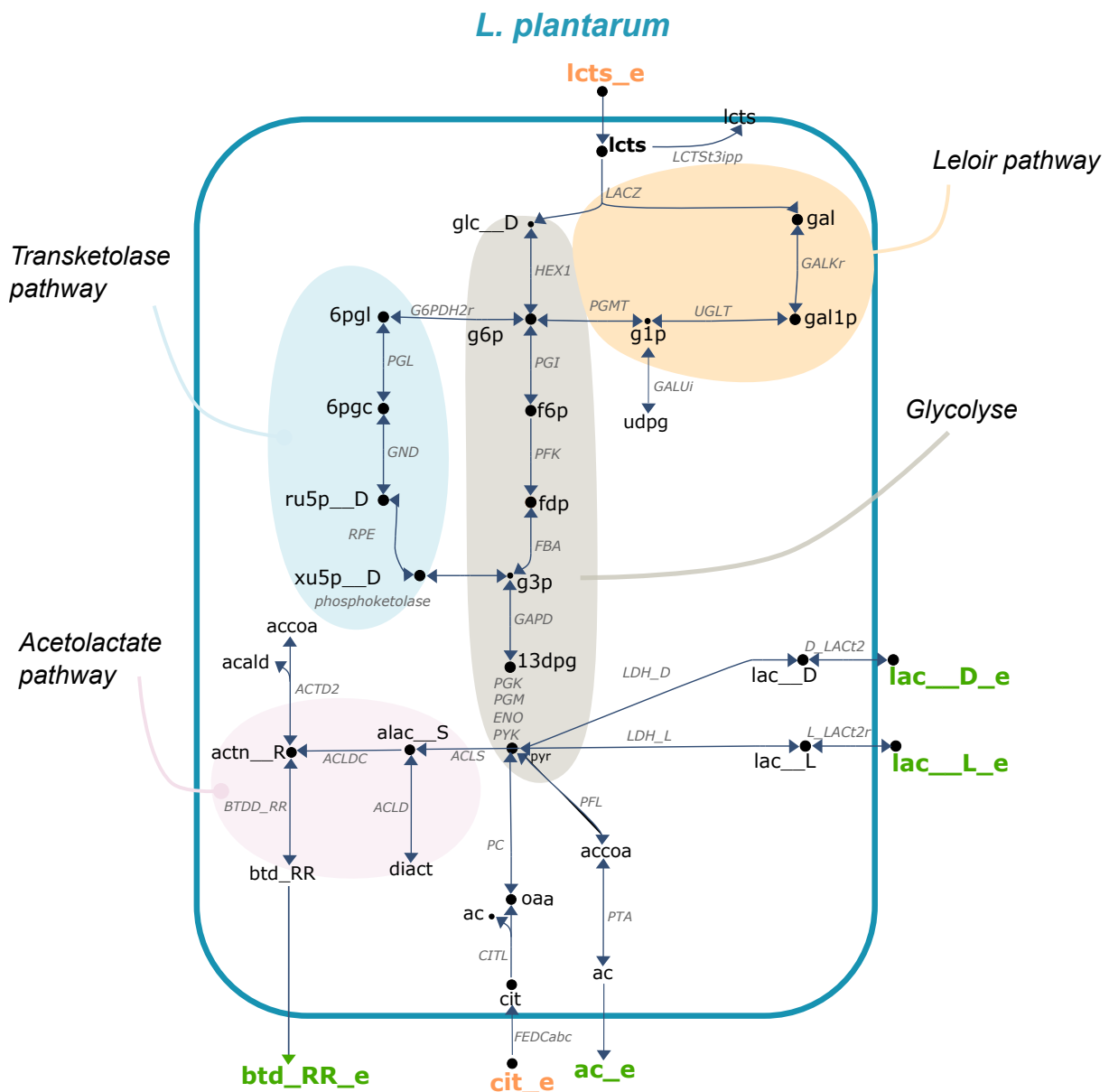


FIGURE 2.5 – Cartes des voies métaboliques du métabolisme carboné central d'intérêt de *L. plantarum*. La voie métabolique de la glycolyse est représentée en gris, celle de Leloir en saumon, acétolactate en rose, et la transketolase en bleue. Les métabolites en orange et vert correspondent aux entrées et sorties du modèle métabolique. La liste des métabolites et des réactions est répertoriée à la fin du chapitre à la section 2.4

P. freudenreichii La souche CIRM-BIA122 de *P. freudenreichii* consomme à la fois l'acide lactique, sa principale source de carbone (Dank et al., 2021; Thierry et al., 2011) et/ou le lactose (Loux et al., 2015). L'acide lactique et le lactose sont catabolisés par les voies de la glycolyse et des pentoses phosphate, menant à la production de pyruvate. Une part du pyruvate alimente le cycle des acides tricarboxyliques (TCA) et la voie de Wood-Werkman (Fig. 2.6) dans laquelle l'acide propionique peut être produit. Cette voie métabolique est spécifique au genre *Propionibacterium* et est assurée en redirigeant le succinate à partir du TCA. Les productions d'acide acétique et du CO₂ (Turgay et al., 2020) à partir du pyruvate sont permises grâce à l'activité de la pyruvate oxidase et de la pyruvate dehydrogenase respectivement. Pour assurer la proportion correcte de propionate/acétate, le GEM de *P. freudenreichii* a requis de plus amples curations. En effet, les bornes des réactions *PPAKr*, *PTAr*, *2131pyrpp* ont été modifiées pour orienter le métabolisme vers la voie de Wood-Werkman, et la réaction *POX2* a été bloquée pour réguler le flux d'acétate.

Un moyen de valider nos modifications est de calculer le ratio de concentration de propionate sur l'acétate afin de vérifier si ce dernier coïncide avec les observations biologiques. Nous avons obtenu la valeur de 2.19, proche du ratio de 2 mentionné dans les travaux de (Cao et al., 2021; Crow, 1986; Dank et al., 2021; Turgay et al., 2020).

L'ensemble des modifications portées à chaque modèle métabolique est résumé dans la table 2.4.

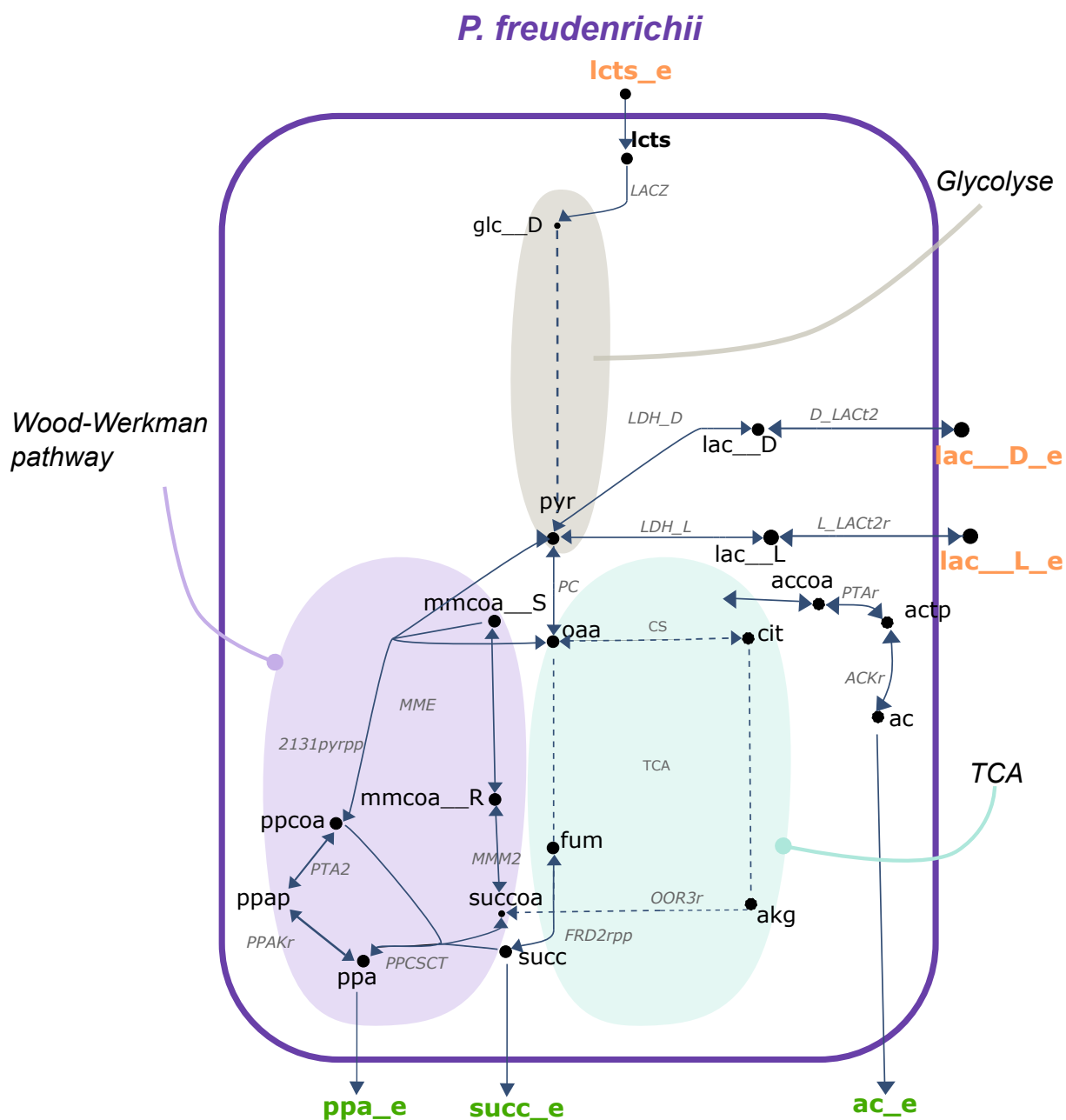


FIGURE 2.6 – Cartes des voies métaboliques du métabolisme carboné central d'intérêt de *P. freudenreichii*. La voie métabolique de la glycolyse est représentée en gris, le cycle du TCA en bleu clair, et la voie de Wood-Werkman en violet. Les métabolites en orange et vert correspondent aux entrées et sorties des modèles métaboliques. La liste des métabolites et des réactions est répertoriée à la fin du chapitre à la section 2.4

Nous avons par la suite vérifié la qualité de chaque réseau métabolique : une faible fraction de réactions universellement bloquées ($< 3\%$), et une fraction limitée de réactions sans association de gène-proteine-réactions (GPR) ($< 30\%$) (Lieven et al., 2020) ont été observées. Puis, nous les avons analysés en FBA (Orth et al., 2010; Palsson and Varma, 1994) vérifiant ainsi la production des molécules de la biomasse à partir de l'environnement nutritionnel du lait. Nous avons obtenu des taux de croissance allant de $0.187 \text{ mmol.gDW}^{-1}.\text{hr}^{-1}$ pour *P. freudenreichii* à $2.17 \text{ mmol.gDW}^{-1}.\text{hr}^{-1}$ pour *L. lactis*. Comme les conditions dans le fromage correspondent à des conditions de microaérobiose

(Miyoshi et al., 2003), une attention particulière a été donnée au rôle de l’oxygène sur le métabolisme des trois bactéries, en modifiant la borne supérieure de la réaction d’importation de l’oxygène dans le compartiment cytosolique. Les caractéristiques générales des trois modèles sont décrites dans la Table 2.3. Nous observons que *L. lactis* a un plus petit nombre de réactions que les deux autres souches, une différence qualitative qui est aussi observée pour les autres souches de la même espèce dans d’autres bases de données (Noronha et al., 2018).

	<i>P. freudenreichii</i>	<i>L. plantarum</i>	<i>L. lactis</i>
Nombre de gènes	1473	1433	1272
Nombre de métabolites	1284	1045	939
Nombre de réactions	1790	1523	1337
Pourcentage de réactions associées a au moins un gène	76.6	79.0	82.1
Taux de croissance en mmol.gDW ⁻¹ .hr ⁻¹	0.187	0.645	2.172

TABLE 2.3 – **Caractéristiques générales des trois modèles métaboliques**

Le raffinement des trois modèles FBA a permis la prédiction de production ou de consommation de plusieurs composés d’arômes tels que les acides lactique, acétique et propionique, le diacétyl et le butanediol (Cao et al., 2021; Smid and Kleerebezem, 2014) (Fig. 2.6,2.4,2.5). Les spécificités métaboliques de chaque souche en fonction de leur métabolisme primaire ont été reproduites fidèlement : la voie de Wood-Werkman pour *P. freudenreichii*, la voie du tagatose pour *L. lactis* et la voie de la transketolase pour *L. plantarum*.

La méthode itérative a permis de répondre à la première exigence que le modèle numérique doit satisfaire : "vérifier que les voies métaboliques d’intérêts permettant la production des composés d’arômes soient présentes et activées". Nous avons reproduit de manière qualitative, la production et la consommation de métabolites connues de la littérature. Nous avons également prédit de nouveaux comportements métaboliques, comme la dégradation du citrate chez *L. plantarum* par les enzymes de la famille CITL. Afin d’obtenir des prédictions numériquement proches des courbes de croissance et des concentrations de métabolites, nous avons ensuite calibré chaque modèle individuel.

Calibration dynamique

Nous avons utilisé les données cinétiques à disposition pour raffiner chaque modèle (voir 2.2.1.2). Cette étape de calibration consiste à minimiser l’écart entre les données expérimentales et la simulation de façon dynamique en estimant des paramètres (voir 2.4). Pour ce faire, le modèle numérique devait permettre : d’intégrer des données hétérogènes et de modéliser la dynamique de croissance et de production/consommation de métabolites pour chaque souche.

L’objectif à atteindre est de retrouver numériquement les observations biologiques que sont : la croissance bactérienne, l’évolution du pH pour *L. lactis* et *L. plantarum* ainsi

GEM	Reaction	Bornes	Valeur	Justification
<i>L. lactis</i>	ACTD2	borne inférieure borne supérieure	0 0	Permet l'activation de la voie métabolique de l'acetolactate
	ACLDC	borne inférieure borne supérieure	0 2	Permet l'activation de la voie métabolique de l'acetolactate
	ACLS	borne inférieure borne supérieure	2 1000	Permet l'activation de la voie métabolique de l'acetolactate
	LACZ	borne inférieure borne supérieure	0.0006 2	Oblige la consommation de lactose
<i>L. plantarum</i>	PFL	borne inférieure borne supérieure	0 0	Permet la production des deux isomères de lactate à partir du lactose
	LCTSt3ipp	borne inférieure borne supérieure	0 0	Redirige le flux de lactose dans la voie métabolique de la Glycolyse
	ACTD2	borne inférieure borne supérieure	0 0	Permet l'activation de la voie métabolique de l'acetolactate
	GALU _i	borne inférieure borne supérieure	-30 10	Permet la production des deux isomères de lactate à partir du lactose
	ACLS	borne inférieure borne supérieure	2 1000	Permet l'activation de la voie métabolique de l'acetolactate
<i>P. freudenreichii</i>	6PGALSZ	borne inférieure borne supérieure	0 0	Réaction appartenant à la voie du tagatose, non présente chez <i>P. freudenreichii</i>
	XYLI2	borne inférieure borne supérieure	0 0	A partir du glucose, le fructose était produit à la place du glucose-6-phosphate et la Glycolyse n'était pas activée.
	POX2	borne inférieure borne supérieure	0 0	Régule le flux d'acétate
	PPAK _r	borne inférieure borne supérieure	0 0	Inhibe la surproduction de l'acétate
	LACZ	borne inférieure borne supérieure	0 5	Autorise la consommation de lactose
	PTA _r	borne inférieure borne supérieure	-1000 8	Permet d'obtenir le bon ratio de production propionate/acetate
	2131pyrpp	borne inférieure borne supérieure	8 1000	Permet d'obtenir le bon ratio de production propionate/acetate

TABLE 2.4 – Raffinement manuel des GEMs avec les justifications correspondantes

que les concentrations de métabolites pour *P. freudenreichii*.

Pour estimer les paramètres, nous avons résolu le problème de minimisation décrit dans les équations 2.8 pour les bactéries lactiques et 2.9 pour *P. freudenreichii*. Brièvement, nous cherchons à minimiser l'écart aux données de croissance pour les 3 espèces, l'écart aux données de pH pour les deux bactéries lactiques et l'écart aux données de dosage de métabolites pour *P. freudenreichii*. Avec ce type de méthode, le sur-apprentissage est une problématique à prendre en compte. Afin de l'éviter, nous avons optimisé un petit nombre de paramètres pour chaque espèce : seulement deux paramètres ont été utilisés pour calibrer la croissance de *L. lactis* et la régulation du pH, trois pour *L. plantarum* régulant le pH ainsi que sa croissance, et enfin, un paramètre pour *P. freudenreichii*, ajustant sa croissance.

Ce principe de calibration des modèles se base également sur un processus itératif induit par les allers-retours entre les données simulées et les paramètres inférés. En effet, la solution informatique d'optimiser des paramètres pour ajuster les modèles n'est pas

suffisante. Nous avons dû établir des contraintes de modélisation dynamique en nous basant sur la connaissance biologique de l'espèce ainsi que sur les résultats des simulations. Expérimentalement, nous avons observé que le métabolisme des bactéries lactiques est toujours actif pendant la phase stationnaire au regard de la courbe de pH, indiquant que les bactéries continuent d'acidifier le milieu. Elles peuvent ainsi continuer de produire du lactate et de l'acétate à partir du lactose puisque ce dernier n'est pas considéré comme limitant dans les conditions expérimentales. La première contrainte dynamique a consisté à réguler le flux de production des acides lactique et acétique par le flux de consommation du lactose, source de carbone principale.

L. lactis* et *L. plantarum Pour moduler le flux de production des acides lactiques, les bornes supérieures des réactions d'échange du lactate EX_lac__L_e et EX_lac__D_e, sont modulées comme suit :

$$\min\left(\frac{m_{lcts_e}}{\Delta t * \sum_{i \in \mathcal{M}(lcts_e)} b_i}, \mu_{max,lcts} * 10^{(-k_{lac} * \phi_{undiss})} + \mu_{min,lcts}\right) * 4 \quad (2.10)$$

Toutes les constantes sont décrites dans la sous section 2.2.3. Pour rappel, m_{lcts_e} est la concentration du lactose (mmol/g de lait), $\mathcal{M}(lcts_e)$ le sous-ensemble des bactéries métabolisant le lactose, $\mu_{max,lcts}$ et $\mu_{min,lcts}$ valeurs maximales et minimales respectivement du flux de consommation du lactose, k_{lac} une décroissance exponentielle et ϕ_{undiss} la fonction qui calcule la concentration de la forme non dissociée de l'acide lactique à partir du lactate.

Comme la production de lactate est reliée à la disponibilité du lactose, nous avons modulé dynamiquement l'export de lactate en fonction de la quantité de lactose à importer, avec un coefficient stœchiométrique de 4 (voir eq. (2.11)). Ce coefficient permet de mieux ajuster la courbe de production de lactate aux données expérimentales.

Étant donné que l'acétate n'est pas relié dans notre modèle au calcul du pH, mais est produit à partir du lactose par l'intermédiaire du pyruvate, nous avons appliqué une simple régulation de la production de l'acétate par la quantité de lactose disponible. Soit EX_ac_e, la réaction d'échange de l'acétate, sa borne supérieure sera modulée dynamiquement par :

$$\max\left(-\frac{m_{lcts_e}}{\Delta t * \sum_{i \in \mathcal{M}(lcts_e)} b_i}, v_{i,ac_e}^{exp}\right)$$

La production d'acétate est régulée par la disponibilité de lactose quand celui ci est épuisé, représenté par $-\frac{m_{lcts_e}}{\Delta t * \sum_{i \in \mathcal{M}(lcts_e)} b_i}$, autrement sa production sera limitée par sa limite d'export physiologique intrinsèque (en $\text{mmol.h}^{-1}.\text{g}_{DM}^{-1}$) : $v_{i,ac_e}^{exp} = 0.5$ pour *L. lactis* et $v_{i,ac_e}^{exp} = 1.0$ pour *L. plantarum*.

Lorsque les concentrations extra-cellulaires de lactose sont suffisamment importantes, le flux de consommation du lactose par les bactéries lactiques est également régulé par l'acidité libérée par ces bactéries : $-\mu_{max,lcts} * 10^{(-k_{lac} * \phi_{undiss})} - \mu_{min,lcts}$. En effet, le métabolisme des bactéries lactiques continue même lorsqu'elles sont à l'état stationnaire. En revanche, lorsque la concentration de lactose devient insuffisante, un partage uniforme entre les bactéries consommatrices, composées des deux bactéries lactiques et de la bactérie propionique est mis en place : $-\frac{m_{lcts_e}}{\Delta t * \sum_{i \in \mathcal{M}(lcts_e)} b_i}$. L'approche modélisant ce processus est inspirée des travaux de (Özcan et al., 2020) dans lesquelles l'acidité libérée par les

bactéries lactiques est représentée par le forme non dissociée de l'acide lactique.

Soit EX_lcts_e, la réaction d'échange du lactose. Sa borne inférieure des bactéries lactiques sera modulée dynamiquement par :

$$\max\left(-\frac{m_{lcts_e}}{\Delta t * \sum_{i \in \mathcal{M}(lcts_e)} b_i}, -\mu_{max,lcts} * 10^{(-k_{lac} * \phi_{undiss})} - \mu_{min,lcts}\right) \quad (2.11)$$

Concernant la régulation par le pH, la quantité d'acide lactique non dissocié est calculée à partir de la concentration de lactate au moyen de la fonction ϕ_{undiss} (voir équation 2.5 pour l'expression de cette fonction); puis, avec une croissance exponentielle à partir de la valeur $-\mu_{max,lcts}^{lact} - \mu_{min,lcts}^{lact}$ (respectivement $-\mu_{max,lcts}^{plant} - \mu_{min,lcts}^{plant}$ pour *L. plantarum*) à la valeur $-\mu_{min,lcts}^{lact}$ ($-\mu_{min,lcts}^{plant}$) quand la concentration d'acide lactique non dissocié augmente, avec un taux exponentiel k_{lac} . Les valeurs de $\mu_{max,lcts}^{lact}$ (resp. $\mu_{max,lcts}^{plant}$), $\mu_{min,lcts}^{lact}$ (resp. $\mu_{min,lcts}^{plant}$) et k_{lac} sont inférées à partir des expériences de co-cultures.

Pour *P. freudenreichii*, lorsque le lactose est épuisé numériquement, nous avons appliqué le même partage uniforme des ressources que pour les bactéries lactiques : $-\frac{m_{lcts_e}}{\Delta t * \sum_{i \in \mathcal{M}(lcts_e)} b_i}$. Sinon, comme il n'est pas impacté par sa propre acidification, un flux constant F_{lcts} estimé des données de métabolomiques en culture pure est appliqué lorsque la quantité de lactose est suffisamment abondante. Ce flux constant est estimé de la façon suivante :

$$\mu_{i,j} = \alpha \frac{m_j(T) - m_j(0)}{\int_0^T b_i(t) dt} \quad (2.12)$$

où $\mu_{i,j}$ correspond à la valeur de flux (en $\text{mmol.h}^{-1}.\text{g}_{DM}^{-1}$), m_j la concentration du métabolite (mmol.g^{-1} de lait, T valeur finale, 0 valeur initiale, $b_i(t)$ la densité bactérienne (en $\text{g}_{DM}^{-1}.\text{g}^{-1}$ de lait), et α , un coefficient prenant en compte les erreurs d'approximation.

La borne supérieure du lactose pour chaque bactérie reste inchangée.

Pour rappel, le lactate, l'acétate, le succinate et le propionate sont dosés pour *P. freudenreichii* et les concentrations initiales et finales sont récupérées. Ainsi, les bornes des flux de production de ces composés sont estimées à partir des données dans le but de rapprocher la valeur de flux de production prédite par le modèle avec celle des données expérimentales. Dans ces conditions expérimentales, les nutriments ne sont pas considérés comme limitants, ainsi nous avons modélisé la phase plateau par une croissance logistique.

Pour *P. freudenreichii*, la borne de consommation du lactate est modélisée comme suit :

$$\max\left(-\frac{m_{lac_{L,e}}}{\Delta t * \sum_{i \in \mathcal{M}(lac_L)} b_i}, F_{lac} * K_{freud} * \frac{b_{freud}}{B_{p,freud}}\right) \quad (2.13)$$

Comme précédemment avec le lactose, lorsque la quantité de lactate est insuffisante, nous avons utilisé un partage uniforme des ressources $-\frac{m_{lac_{L,e}}}{\Delta t * \sum_{i \in \mathcal{M}(lac_L)} b_i}$ parmi les bactéries consommatrices du lactate $\mathcal{M}(lac_L)$. Dans le cas contraire, un flux F_{lac} estimé des données de métabolomiques en culture pure (voir l'équation 2.12) est appliqué après une régulation par la fraction $\frac{b_{freud}}{B_{p,freud}}$, où $B_{p,freud}$ est la valeur plateau de la concentration de la biomasse en pure culture, et b_{freud} et la densité de biomasse actuelle, et le facteur K_{freud} , qui est

inféré des données de croissance en culture pure. Cette modulation modélise un ordre de grandeur différent du rendement du métabolisme en culture pure ou en co-culture.

Comme l'acétate, le propionate et le succinate ne sont pas consommés par aucun des modèles métaboliques, seule la borne supérieure (de production) est modulée. Soit les réactions d'échanges d'acétate (EX_ac_e), de propionate (EX_ppa_e) et de succinate (EX_succ_e), la valeur de la borne supérieure est égale à :

$$F_i * K_{freud} * \frac{b_{freud}}{B_{p,freud}}, \quad \text{pour } i \in \{ EX_ac_e, EX_ppa_e, EX_succ_e \}$$

où le taux maximal de production est défini par F_i , pour $i \in \{ EX_ac_e, EX_ppa_e, EX_succ_e \}$ estimé à partir des données de métabolomique de culture pure (voir la section 2.12). Le même terme de modulation que dans l'équation 2.13 est appliqué, modélisant un ordre de grandeur différent du rendement du métabolisme en culture pure ou en co-culture.

Cette étape itérative de calibration dynamique est coûteuse en temps. En effet, pour chacune des nouvelles valeurs de paramètres, nous avons lancé une simulation dynamique de chaque souche optimisée.

Pour les autres métabolites suivis dynamiquement, c'est-à-dire le butanediol, le diacétyl et le citrate, la borne inférieure est modulée selon le principe de partage équitable décrit plus haut lorsque la quantité de substrat devient insuffisante.

Soient EX_diact_e, EX_btd_RR_e et EX_cit_e les réactions d'échanges correspondantes aux diacétyl, butanediol et citrate, leurs bornes inférieures sont ainsi modulées par :

$$\max\left(-\frac{m_j}{\Delta t * \sum_{i \in \mathcal{M}(j)} b_i}, v_{i,j}^{int}\right)$$

pour $i = L. lactis$ si $j \in \{ EX_diact_e, EX_btd_RR_e, EX_cit_e, EX_ac_e \}$
pour $i = L. plantarum$ si $j \in \{ EX_btd_RR_e \}$

Quand le substrat n'est pas limité, l'import est contraint par la limite physiologique intrinsèque modélisée par $v_{i,j}^{int} = -8$ pour $i = L. lactis$ et $j \in \{ EX_diact_e, EX_btd_RR_e \}$, $v_{i,j}^{int} = -1$ pour $j = EX_cit_e$ et $v_{i,j}^{int} = -5$ pour $i = L. plantarum$ et $j \in \{ EX_ac_e, EX_btd_RR_e \}$ ou sinon, par la disponibilité du substrat uniformément partagé par les micro-organismes consommateurs : $-\frac{m_j}{\Delta t * \sum_{i \in \mathcal{M}(j)} b_i}$.

En plus de ces régulations dynamiques, nous avons observé *via* la méthode de FVA un flux important de consommation du citrate, de la serine, du glutamate et de l'alanine, empêchant d'obtenir la précision numérique des composés dosés en culture pure. Ainsi, pour réorienter les flux dans les voies métaboliques attendues nous avons réduit chacune des bornes d'importation de ces composés à -12, améliorant les prédictions.

Toutes ces modifications faites, soit à partir des connaissances biologiques soit à partir des données biochimiques, rendent la prédiction des modèles plus précise et plus fidèle aux données.

Simulation des modèles calibrés

Nous avons ainsi curé manuellement et calibré dynamiquement chacun de nos modèles FBA individuels dans le but premier de retrouver les valeurs de concentration des expériences biologiques en culture pure. Les simulations statiques et dynamiques de nos modèles calibrés sont représentées dans les Figures 2.7, 2.8 et 2.9. Chaque sous figure (a) représente les flux des voies métaboliques d'intérêt calculés par le FBA lors de l'inoculation alors qu'aucun métabolite n'est limitant.

Le modèle FBA de *L. lactis* (figure 2.7a), nous montre une préférence d'utilisation de la voie métabolique du tagatose plutôt que de Leloir pour dégrader le lactose. Nous remarquons également que les deux isomères de l'acide lactique sont produits *via* la voie de la glycolyse avec un plus grand flux de production pour le D-lactate, expliquant ainsi l'acidification du lait et du caillé durant le processus de fabrication de fromage.(Fig. 2.7a). Concernant la voie de l'acétolactate, nous remarquons qu'elle est bien active grâce à la production du diacétyl et du butanediol. Enfin, nous observons également un flux de consommation du citrate, *via* la réaction *CITL* encodée par le gène *citL*.

Contrairement à *L. lactis*, le modèle métabolique de *L. plantarum* semble avoir une préférence de production pour le L-lactate, montrant ainsi sa contribution à l'acidification du fromage. La voie de la glycolyse est active, ainsi que la consommation du lactose jusqu'à la production des deux formes du lactate. Nous remarquons également un flux de production de l'acide acétique par la voie de l'acétatekinase, contribuant plus faiblement à l'acidification (Fig. 2.5). Les simulations dynamiques de *L. plantarum* ont montré une activation des voies fermentaires hétérolactique et homolactique. Cependant à notre connaissance, la voie métabolique de la transkétolase, représentant la fermentation hétérolactique, est observée pour la première fois dans un contexte laitier. La consommation du citrate est connue pour être dépendante de la souche au sein de l'espèce *L. plantarum* (Palles et al., 1998). Elle a été observée dans notre cas à travers la réaction de transport *FEDCabc*, conduisant à une augmentation de la quantité de pyruvate intracellulaire.

Enfin, le modèle FBA de *P. freudenreichii* consomme préférentiellement du lactose malgré la présence de l'acide lactique (Fig. 2.9a). Cependant, nous retrouvons une activation simultanée des cycles du TCA et de Wood-Werkman (Deborde and Boyaval, 2000), le premier permettant la régénération des sources de carbone nécessaires pour l'activation du second et ainsi, la production de propionate.

Les sous-figures (b) des Figures 2.7,2.8,2.9 montrent que les courbes de croissance simulées (en vert) sont fidèles aux données de croissances expérimentales de chaque souche microbienne. Concernant le pH, pour les sous-figures 2.7, 2.8 (c), nous avons observé que l'utilisation de la concentration du lactate dans une approximation du pH permet d'expliquer le pH observé lors des cultures pures de *L. lactis* et *L. plantarum*. Enfin, le modèle métabolique de *P. freudenreichii* retrouve les concentrations finales prédites par les expériences biologiques (Figure 2.9(c)). Afin de vérifier statistiquement que l'apprentissage de nos modèles est correct, nous avons représenté sur la figure Fig. 2.10 (a) les valeurs de croissance, de pH et de métabolites calculés par notre modèle et celles des données expérimentales. Nous observons une très bonne corrélation avec une valeur de $R^2 = 0.99$, montrant ainsi que nos modèles métaboliques individuels sont suffisamment bien ajustés.

Nous venons de voir tout le processus d'itération statique et dynamique nécessaire pour

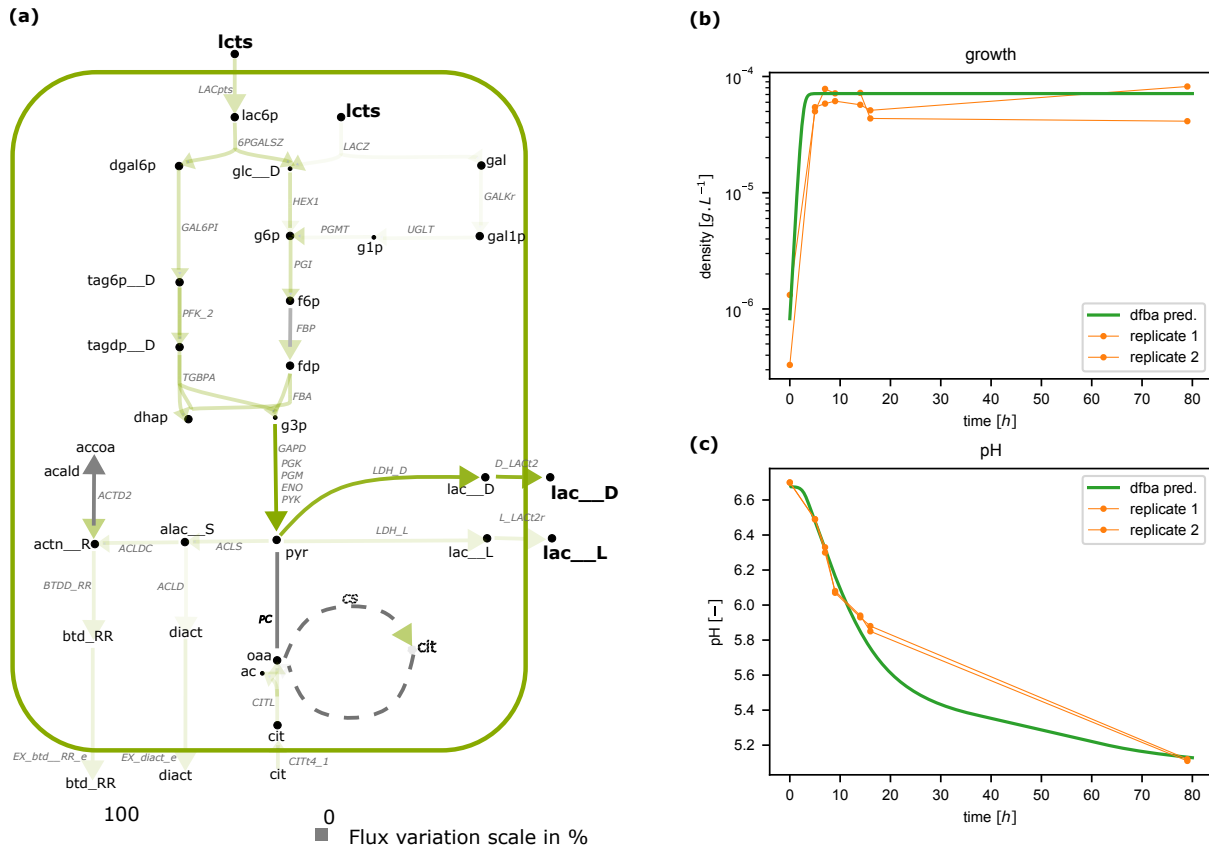


FIGURE 2.7 – Réseau métabolique de *L. lactis* après calibration sur les données de culture pure. (a) Représentation d’un FBA appliqué sur le métabolisme carboné central lors de l’inoculation de *L. lactis*. L’échelle de couleur représente les valeurs de flux prédites par le FBA et normalisées par la plus haute valeur de flux prédite sur les voies métaboliques illustrées. (b) Simulation de la croissance de *L. lactis* au cours du temps après inférence des paramètres dans le modèle (courbe verte) et d’après les données expérimentales. (c) Courbes de pH dans des conditions de culture pure simulées par le modèle (courbe verte) et issues des données expérimentales. La liste des métabolites et des réactions est répertoriée à la fin du chapitre à la section 2.4

obtenir des modèles métaboliques satisfaisants et des prédictions numériquement fiables. L’étape finale pour la construction d’un modèle numérique explicatif du fonctionnement de cette communauté fromagère est de passer à l’échelle de la communauté afin de prédire les croissances de chaque souche lors de la production du fromage, et dans un deuxième temps, d’explorer les interactions bactériennes potentielles et les contributions de chaque souche pour la production et la consommation de métabolites.

2.3.2 Simulations à l’échelle de la communauté

En premier lieu, nous avons créé un modèle de communauté dans lequel chaque espèce maximise sa propre biomasse et avons utilisé les modèles calibrés sur les données de cultures pures pour construire ce modèle. Nous avons ensuite ajusté la valeur du “quorum sensing”, correspondant à la valeur à laquelle nous observons une phase plateau de croissance, pour chaque modèle individuel en intégrant la valeur de “quorum sensing” des données expérimentales en co-culture (voir la Table 2.2.1.3 et l’équation 2.3 dans le but de

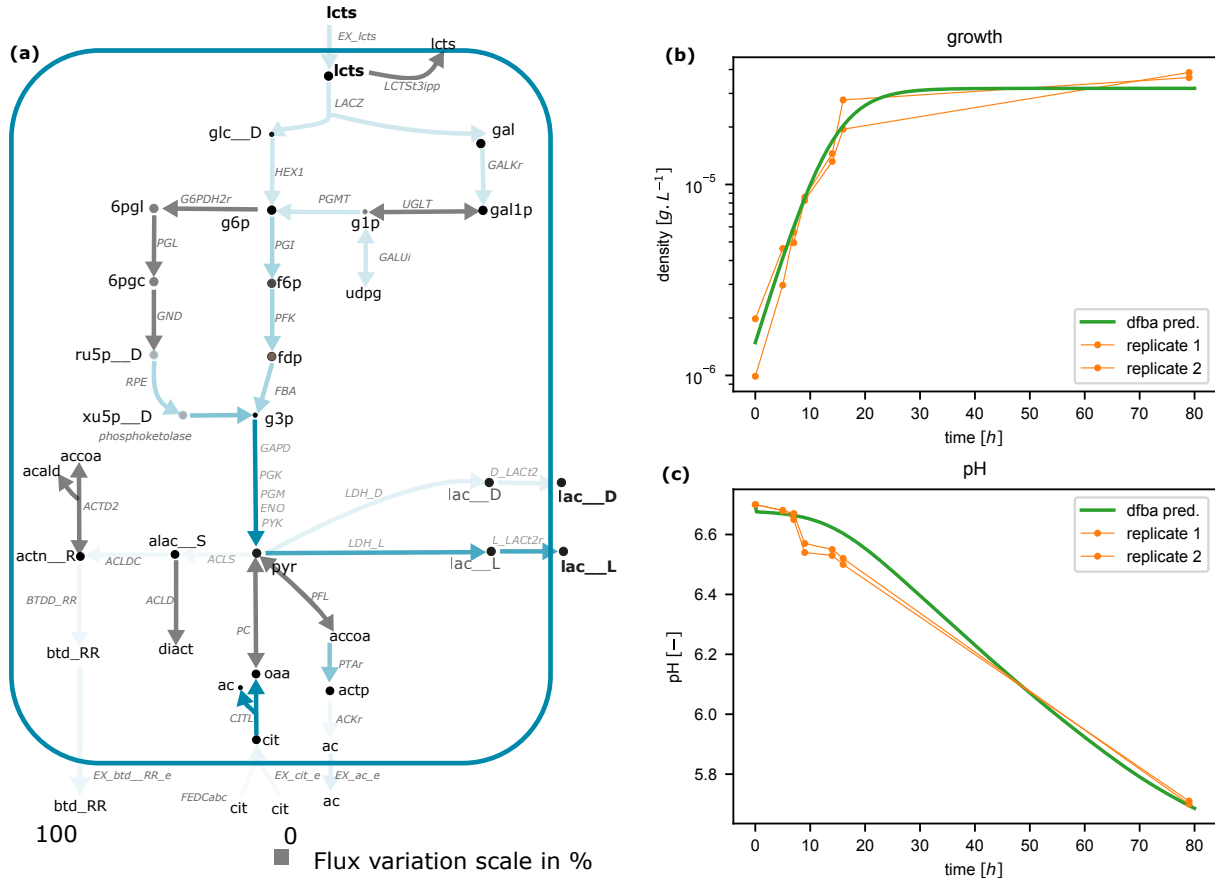


FIGURE 2.8 – Réseau métabolique de *L. plantarum* après calibration sur les données de culture pure. (a) Représentation d'un FBA appliqué sur le métabolisme carboné central lors de l'inoculation de *L. plantarum*. L'échelle de couleur représente les valeurs de flux prédites par le FBA et normalisées par la plus haute valeur de flux prédite sur les voies métaboliques illustrées. (b) Simulation de la croissance de *L. plantarum* au cours du temps après inférence des paramètres dans le modèle (courbe verte) et d'après les données expérimentales. (c) Courbes de pH dans des conditions de culture pure et simulées par le modèle (courbe verte) et issues des données expérimentales. La liste des métabolites et des réactions est répertoriée à la fin du chapitre à la section 2.4

reproduire le comportement métabolique de la communauté *in silico* (voir eq. 2.6-2.7). Ce modèle de communauté n'a subi aucun processus de calibration sur les nouvelles données, c'est-à-dire, aucun paramètre basé sur la métabolomique ou sur les données de croissance en co-culture n'a été inféré afin d'affiner les prédictions. Ces données vont en revanche être consacrées à la validation des prédictions du modèle, c'est-à-dire servir de données tests.

Les simulations de la croissance, du pH et de la métabolomique du modèle de communauté reproduisent fidèlement les valeurs expérimentales (Fig. 2.11(a)-(c), Table 2.3). Nous avons remarqué cependant une légère surestimation de la croissance de *P. freudenreichii* durant la phase exponentielle. Concernant la prédiction des concentrations de métabolites issus de la métabolomique, nous avons observé une consommation plus lente du lactose tandis que celle du citrate s'ajuste bien avec les données expérimentales. La production de lactate est légèrement sur-prédite durant la phase exponentielle résultant

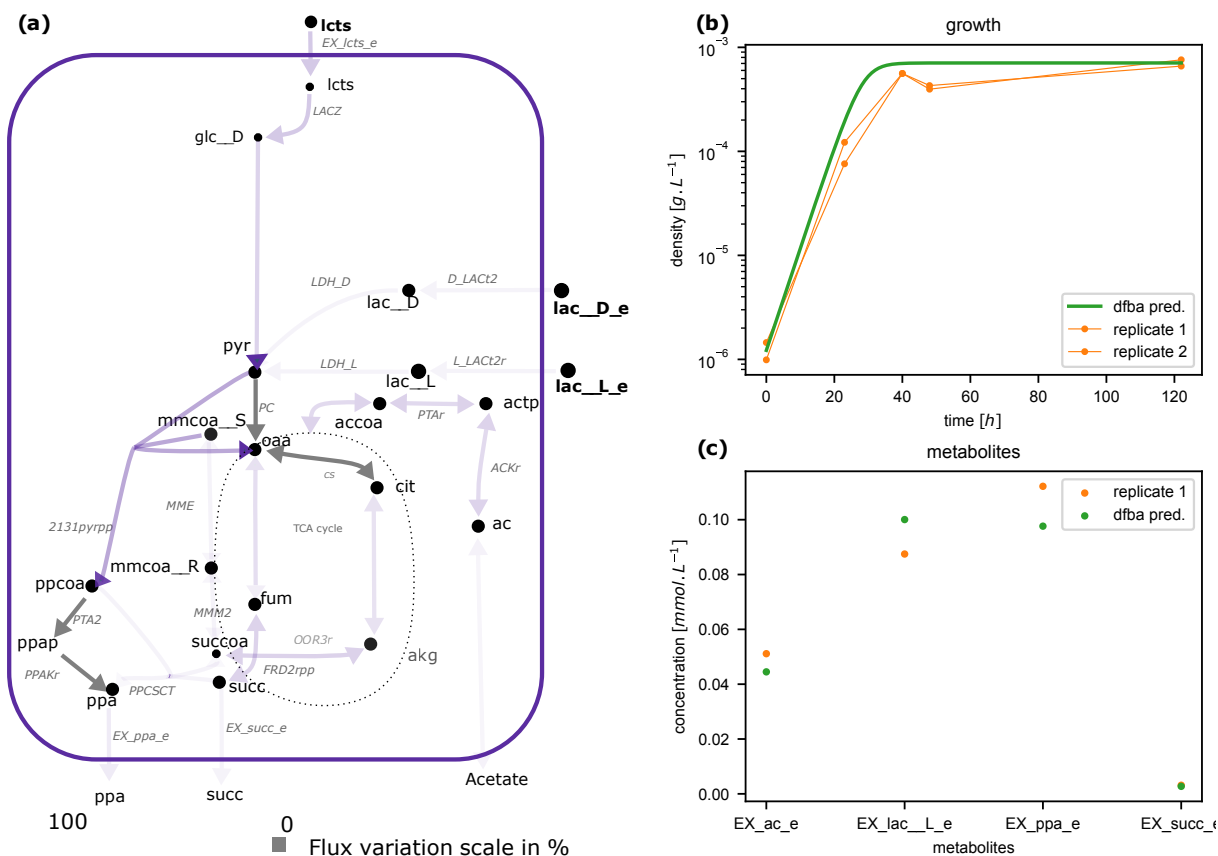


FIGURE 2.9 – Réseau métabolique de *P. freudenreichii* après calibration sur les données de culture pure. (a) Représentation d'un FBA appliqué sur le métabolisme carboné central lors de l'inoculation de *P. freudenreichii*. L'échelle de couleur représente les valeurs de flux prédites par le FBA et normalisées par la plus haute valeur de flux prédite sur les voies métaboliques illustrées. (b) Simulation de la croissance de *P. freudenreichii* au cours du temps après inférence des paramètres dans le modèle (courbe verte) et d'après les données expérimentales. (c) Concentrations de lactate, acetate, propionate et succinate simulées (points verts) observées (en orange) au temps final. La liste des métabolites et des réactions est répertoriée à la fin du chapitre à la section 2.4

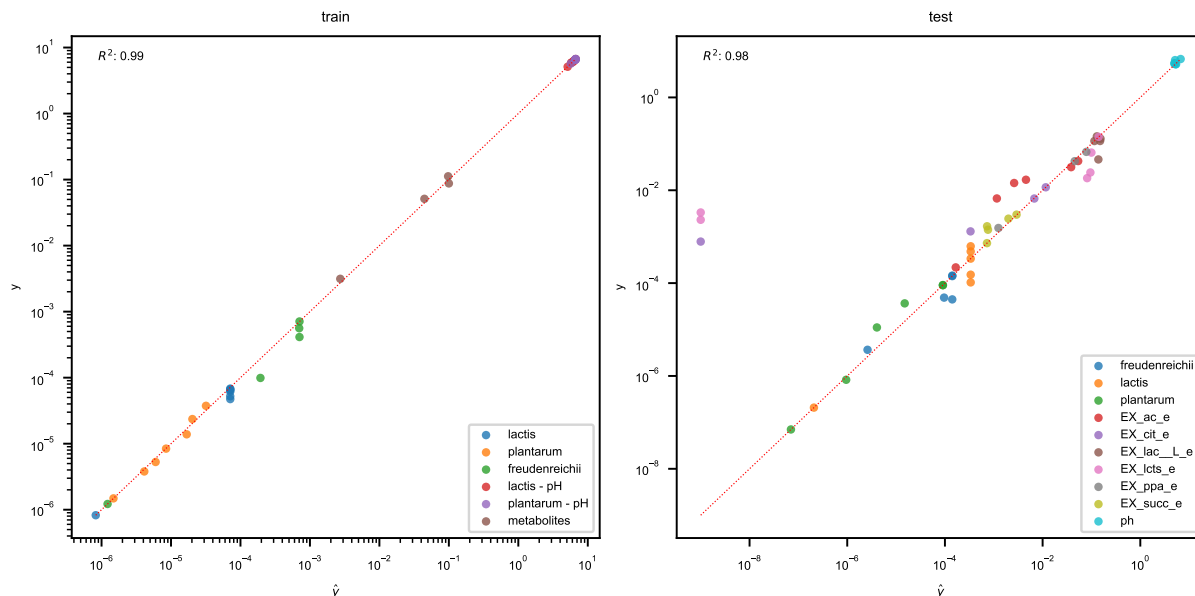


FIGURE 2.10 – **Validation des prédictions après apprentissage.** Les valeurs simulées issues des modèles (\hat{y}) et celles des données expérimentales respectives (y) lors de la phase d'apprentissage sur les données de cultures pures (a) et lors de la phase de test en communauté (b). La couleur représente le type de la donnée et la bissectrice est affichée par une ligne en pointillé rouge. Le R^2 est calculé selon la formule suivante : $R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$ où \bar{y} est la valeur moyenne de y . Nous pouvons observer une bonne capacité du modèle à prédire les concentrations.

d'une sur-acidification du fromage durant les étapes de moulage et de démoulage. En revanche, la production d'acétate est sous-prédite durant les premières phases de la fabrication du fromage, mais le taux de production pendant la maturation est correctement modélisé. Les courbes du propionate, et dans une moindre mesure, celle du succinate sont prédites avec précision par le modèle suggérant que la production de composés organoleptiques par *P. freudenreichii* est correctement capturée. De plus, les voies métaboliques activées en mono-cultures le sont également en co-culture (Figure 2.12). Enfin, afin de valider statistiquement les prédictions du modèle en communauté, nous avons fait une régression entre les valeurs prédites par le modèle et celles des données, et nous observons un R^2 global de 0.98 pour cette étape de validation (Figure 2.10, b).

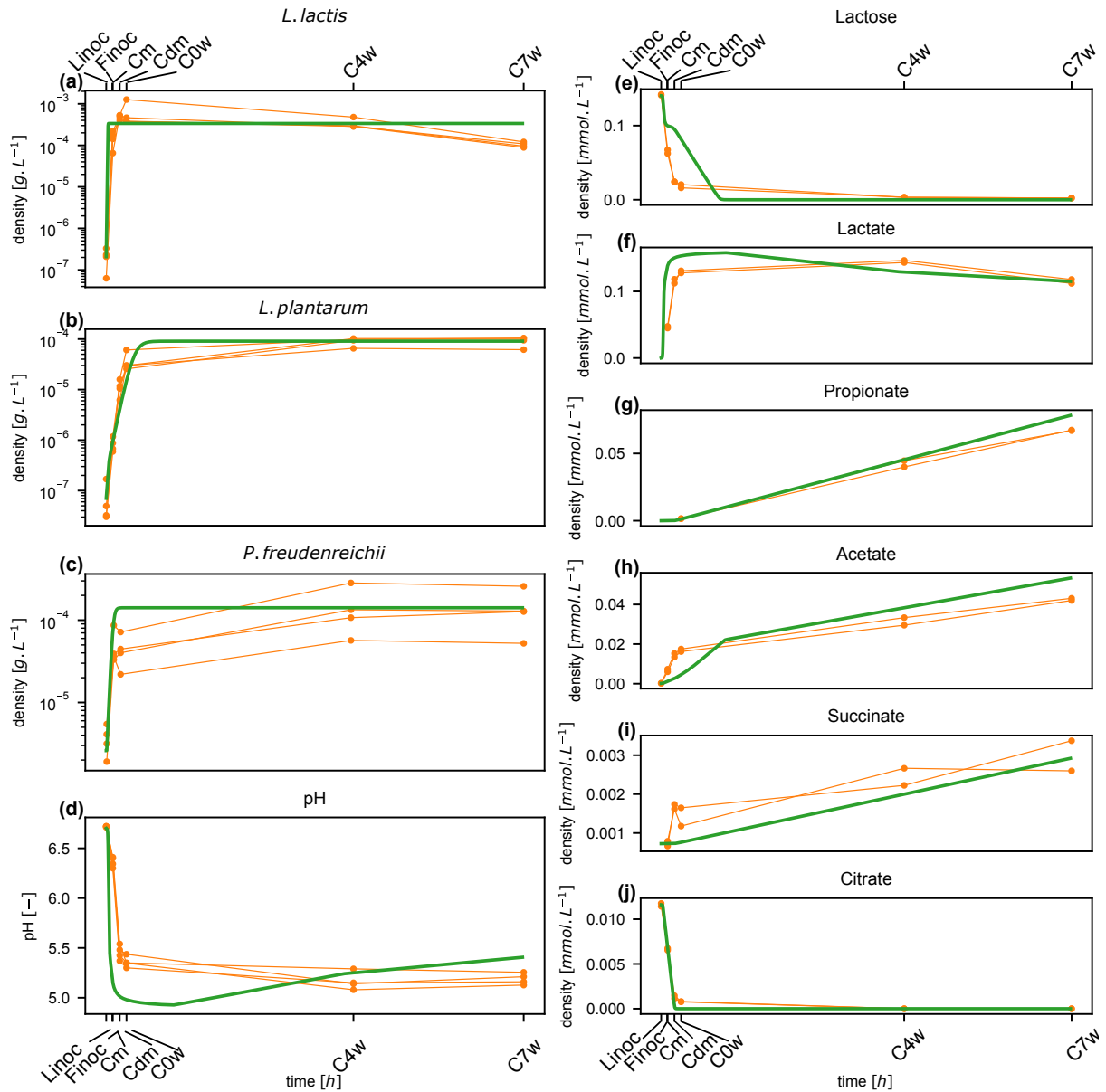


FIGURE 2.11 – **Prédiction de la croissance de chaque souche ainsi que des concentrations des métabolites**(a-c) Courbes de croissance de *L. lactis*, *P. freudenreichii* and *L. plantarum* calculées par le modèle (en vert) et celles observées expérimentalement (en orange). (d-j) Profils métabolique et du pH en condition de co-culture calculé par le modèle et expérimentalement observé. L'écart-type est représenté pour chaque point des données. Abréviations : Linoc, inoculation des LAB (t = 0) ; Finoc, inoculation de *P. freudenreichii* (t = 18 heures) ; Cm, moulage (t = 19.5 heures) ; Cdm, démoulage (t = 40 heures) ; C0w, début d'affinage (t = 60 heures) ; C4w, 4 semaines d'affinage (t = 732 heures) ; C7w, 7 semaines d'affinage (t = 1236 heures).

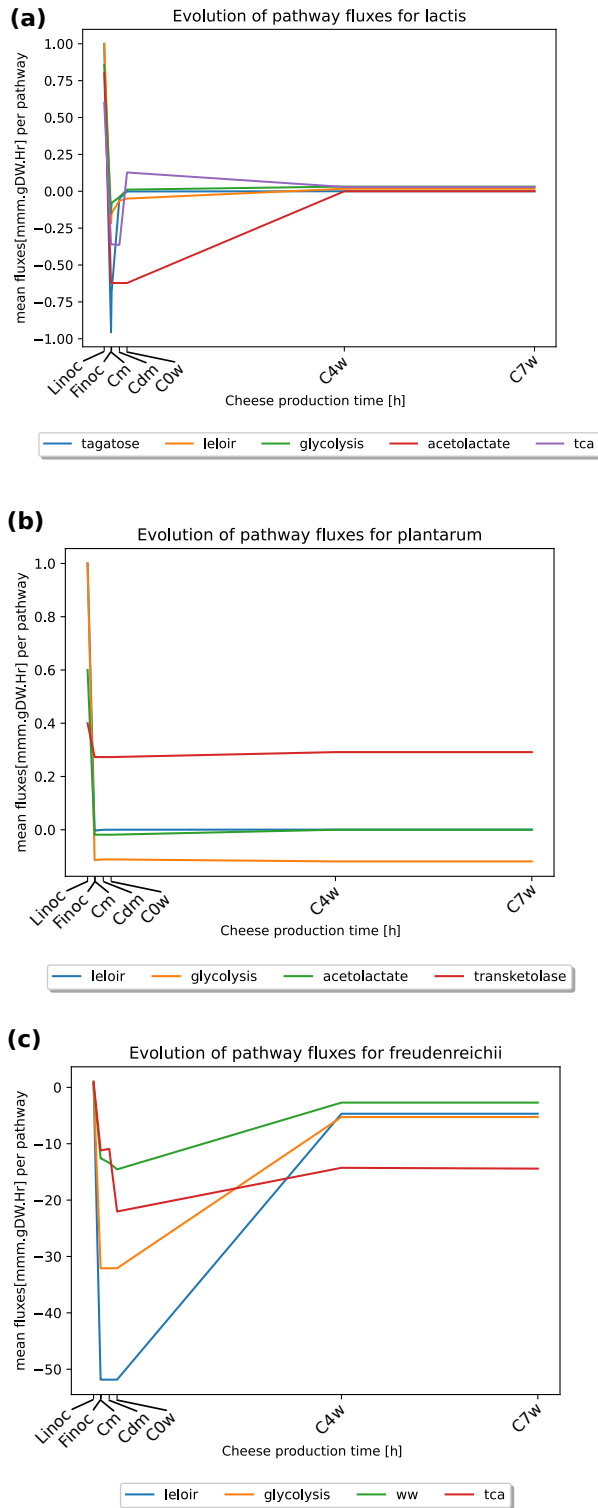


FIGURE 2.12 – **Illustration du changement de voies métaboliques durant le processus de production du fromage.** Pour chaque bactérie nous avons lancé une simulation dynamique (dFBA) et les flux réactionnels pour chaque voie métabolique d'intérêt des Figures 2.4,2.5,2.6 ont été retrouvés à l'inoculation de bactéries lactiques (Linoc, t=0 heure), de *P. freudenreichii* (Finoc, t=18 heures), au moulage (t=19.5 heures), démoulage (Cdm, t=40heures), début de l'affinage (C0w, t=60 heures), après 4 (C4w, t=732 heures) et 7 (C7w, t=1236 heures) semaines d'affinage. Nous avons normalisé les flux par la valeur de flux lors de l'inoculum.

En somme, nous avons montré qu'utiliser des modèles métaboliques curés et calibrés individuellement, ainsi qu'un contrôle dynamique du métabolisme, basé sur la connaissance de la littérature, permet de retrouver des observations biologiques dans cette communauté. En effet, des mécanismes connus de la communauté, comme par exemple, l'interaction de commensalisme entre les bactéries *P. freudenreichii* et *L. lactis*, ont provoqué des raffinements des modèles métaboliques. Le modèle métabolique de *L. lactis* a donc été curé pour permettre la production de lactate à partir du lactose. Un point d'importance est la composition du milieu de culture pour la bactérie propionique lors des expériences culture pure. Nous avons inféré des paramètres à l'échelle de *P. freudenreichii* mais en se basant sur des conditions nutritionnelles de co-culture puisque du lactate et un hydrolysate de peptides ont été ajoutés au lait. Malgré les deux étapes itératives coûteuses en temps, l'interactivité entre modèle numérique et la connaissance biologique *a priori* permet d'obtenir des prédictions à l'échelle d'une communauté avec une bonne précision. Grâce à cette assurance, nous avons pu exploiter davantage le modèle de communauté pour caractériser de possibles interactions bactériennes. Dans les paragraphes suivants, nous déduirons dans un premier temps des interactions que notre approche numérique a pu mettre en évidence en analysant les flux relatifs et absolus de production et de consommation de métabolite par espèce. Dans un second temps, nous verrons ce que des outils sans connaissance *a priori* peuvent découvrir.

2.3.3 Interactions bactériennes et exploration du modèle de communauté

Notre modèle numérique de communauté peut révéler des interactions microbiennes du type syntrophiques, dans lesquelles une espèce consomme des produits d'une autre, ou du type compétition pour un nutriment. Dans ce but, nous nous sommes focalisés sur la dynamique des flux des métabolites d'intérêt pour chaque espèce (Fig. 2.13, (a)) en calculant, à chaque pas de temps de la dynamique, et à l'échelle de la population, les flux d'échange $\mu_{i,j}b_i$ pour chaque métabolite j et bactérie i et les concentrations des métabolites associées. Nous observons que *L. lactis* consomme très rapidement le lactose et le citrate durant sa phase exponentielle. Nous pouvons observer que *L. plantarum* ne contribue pas ou peu à la consommation totale de ces métabolites suggérant ainsi que d'autres métabolites non suivis en dynamique sont utilisés pour la croissance de cette souche, comme par exemple les acides aminés. La croissance plus lente de *L. plantarum* dans les premières heures du processus de production du fromage pourrait expliquer cette observation. En effet, on remarque que la régulation sur les bornes de consommation du lactose et du citrate est mise en jeu lorsque la densité bactérienne de *L. plantarum* est encore faible. De manière surprenante, nous observons que *P. freudenreichii* termine la déplétion du lactose à la place de *L. plantarum* au tout début de la phase d'affinage, bien que le lactate, source préférentielle de cette espèce, soit déjà produit, indiquant une ségrégation temporelle de l'utilisation de lactose, et une absence de compétition directe pour ce substrat.

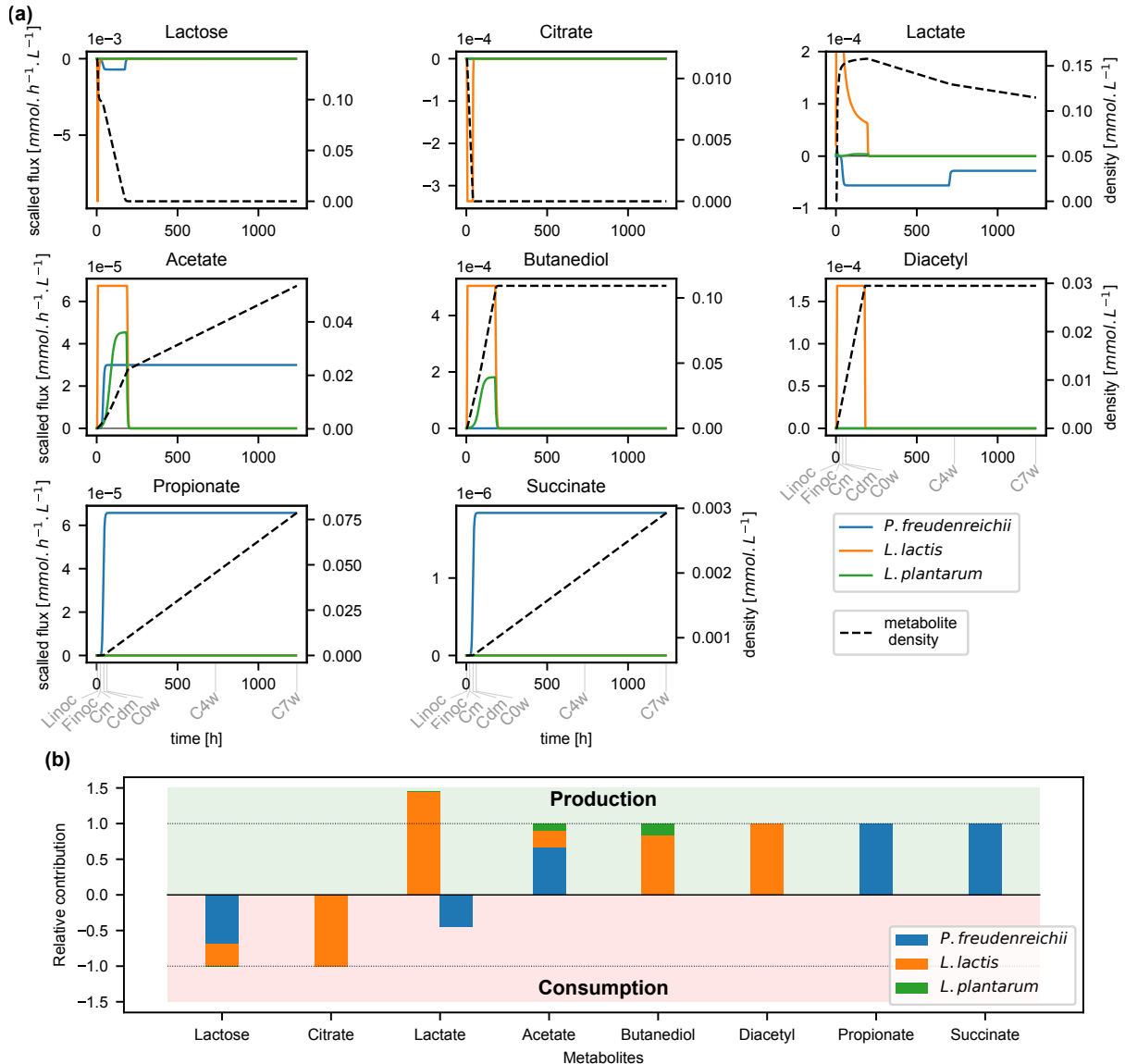


FIGURE 2.13 – (a) **Flux dynamiques de consommation et de production de métabolites d'intérêt.** Nous représentons pour chaque bactérie les flux dynamiques de consommation et de production de chaque métabolite que l'on suit en simulation (lignes colorées). La dynamique de la concentration de ces mêmes métabolites dans le fromage est aussi représentée par les traits en pointillés. (b) **Contribution relative des bactéries dans le devenir des métabolites.** Nous avons calculé pour chaque souche sa contribution totale à la consommation et la production de ces métabolites et intégré cette valeur en temps par $\int_0^t \mu_{i,j} b_i$ pour le métabolite j et la bactérie i , voir eq.2.7). Par la suite, nous avons normalisé le résultat par les échanges totaux au sein de la communauté. La valeur 1 (resp. -1) représente la production (consommation) nette, *i.e.*, la différence de concentration du métabolite à l'état final et initial (ligne en pointillé). Abbreviations : Linoc, inoculation des bactéries lactiques ($t = 0$) ; Finoc, *P. freudenreichii* inoculation ($t = 18$ heures) ; Cm, moulage ($t = 19.5$ heures) ; Cdm, Démoulage ($t = 40$ heures) ; C0w, début de l'affinage ($t = 60$ heures) ; C4w, 4 semaines d'affinage ($t = 732$ heures) ; C7w, 4 semaines d'affinage ($t = 1236$ heures).

La production de lactate étant corrélée avec la consommation de lactose, nous obser-

vons une très forte production par *L. lactis* pendant 250 heures, correspondant à l’affinage dans le processus de fabrication du fromage. En raison de la plus faible croissance de *L. plantarum*, ce dernier contribue peu à la production de lactate. Durant la phase d’affinage, le lactate est consommé par *P. freudenreichii* contribuant ainsi à l’augmentation du pH dans la communauté comme observé dans la Figure 2.11 d. De façon surprenante, toutes les souches sont capables de produire l’acétate dans le milieu extracellulaire. *L. lactis* est le contributeur majeur suivi par *L. plantarum* jusqu’au temps $t = 250$ heures, à partir duquel la phase plateau de sa croissance commence. Ce schéma est retrouvé chez les bactéries lactiques intervenant également dans la production du butanediol *via* la voie de l’acétolactate. Cependant, *P. freudenreichii* continue de garder un taux de production constant d’acétate durant la maturation du fromage malgré la phase plateau atteinte. Enfin, le propionate et le succinate sont produits à taux constant par *P. freudenreichii* durant tout le processus d’affinage et seule la souche de *L. lactis* semble être capable de produire le diacétyle jusqu’au début de la période de l’affinage.

Afin d’obtenir la contribution en flux net de chaque souche au sein de ce consortium, nous avons intégré dans le temps la prédiction du FBA $\int_0^T \mu_{i,j}(t)b_i(t)dt$. Après normalisation de cet échange net par les échanges totaux au sein de la communauté (*i.e.* la somme des échanges net de chaque individu), nous avons obtenu les contributions de chaque souche dans la production et la consommation de chaque métabolite (Fig. 2.13, (b)). Ces contributions confirment que le succinate et le propionate sont uniquement produits par *P. freudenreichii*, alors que le diacétyle (production) et le citrate (consommation) sont métabolisés par *L. lactis*. *L. plantarum* contribue à la production du butanediol, bien que le principal producteur semble être *L. lactis*. Dans une moindre mesure, *L. lactis* contribue plus que *L. plantarum* à la production de l’acétate dominée par *P. freudenreichii*. En terme d’interactions bactériennes, on observe une syntrophie pour le lactate entre *L. lactis* (producteur) et *P. freudenreichii* (consommateur).

Nous avons démontré ci-dessus que capturer la complexité des processus métaboliques dans une communauté microbienne durant la fabrication de fromage requiert des considérations de la dynamique sous-jacente du système. Nous nous demandons si l’application d’approches computationnelles *a priori* reposant sur les modèles métaboliques à l’échelle du génome peuvent mettre en avant de nouvelles interactions putatives entre les souches. À cette fin, nous avons utilisé deux méthodes basées sur les flux, SMETANA (Zelezniak et al., 2015) et MiCOM (Diener et al., 2020), et une approche qualitative basée sur le raisonnement, Metage2Metabo (Belcour et al., 2020) dans le but de suggérer de nouvelles complémentarités métaboliques au sein du consortium.

Les approches quantitative, MiCOM et SMETANA, ont mis en avant respectivement 14 et 25 métabolites échangeables (Table 2.5 et 2.6), et Metage2Metabo, identifie 11 métabolites qui ne peuvent pas être produits par une espèce sans interactions dans la communauté (Table 2.7).

Une première observation est qu’un nombre assez important (11) de métabolites est commun dans les prédictions des composés échangeables de SMETANA et MiCOM : lactate, qui est aussi prédit dans le modèle dynamique, phénylalanine, serine, malate, succinate, xanthine, H₂S, 2-ketoglutarate, glycolate et acetaldehyde. Les autres composés échangés prédits incluent principalement des acides aminés (isoleucine, proline, glycine, alanine). Pour chaque composé mis en avant par SMETANA nous avons utilisé les données méta-transcriptomiques dans le but d’évaluer la validité des interactions les plus plausibles selon le score SMETANA. Le H₂S est prédit comme produit de *L. lactis*

et *L. plantarum* au bénéfice de *P. freudenreichii*, le ribose est produit par *L. lactis* et *L. plantarum* et consommé par *P. freudenreichii*, le glycerol produit par les bactéries lactiques et consommé par *P. freudenreichii* et enfin la phénylalanine produite uniquement par *L. plantarum* et consommée par *P. freudenreichii*. Pour chacun de ces métabolites prédits comme échangés, nous avons vérifié l'expression des gènes associés à la production de ces métabolites chez les espèces donneuses, et à la consommation de ces métabolites chez les espèces receveuses (voir méthodes et Figure 2.14). Les résultats suggèrent que les interactions pour H₂S, ribose et le glycerol sont plausibles à plusieurs étapes de la fabrication du fromage. À l'inverse, alors que les données d'expression de gènes montrent que les voies métaboliques de la consommation de la phénylalanine sont fortement exprimées chez *P. freudenreichii*, leur production par *L. plantarum* ne l'est pas, suggérant que cette interaction semble être moins recevable.

ID Bigg	ID Metacyc	Nom	Ontologie Metacyc	fluxes d'export	fluxes d'import
acald_e	ACETALD	acétaldéhyde	Aldehydes-Or-Ketones, Aldehydes	Pf; Ll	Lp
akg_e	2-KETOGLUTARATE	2-oxoglutarate	Acids, Organic-Acids	Pf	Lp
ala__D_e	D-ALANINE	D-ALANINE	All-Amino-Acids, Acids, Amino-Acids, Organic-Acids	Pf; Lp	Ll
ala__L_e	L-ALPHA-ALANINE	L-ALPHA-ALANINE	All-Amino-Acids, Acids, Amino-Acids, Organic-Acids	Pf; Lp	Ll
co2_e	CARBON-DIOXIDE	CARBON-DIOXIDE	Others, Others	Pf; Ll	Lp
coa_e	CO-A	co enzyme A	Groups, Others	Pf; Ll	Lp
fe2_e	FE+2	Fe2+	Ions, Inorganic-Ions, Cations	Lp	Pf; Ll
gly_e	GLY	Glycine	All-Amino-Acids, Acids, Amino-Acids, Organic-Acids	Lp	Pf; Ll
glyclt_e	GLYCOLATE	GLYCOLATE	Acids, Organic-Acids	Pf	Lp
gthox_e	OXIDIZED-GLUTATHIONE	glutathione disulfide	ORGANOSULFUR, All-Glutathiones	Lp	Ll
gthrd_e	GLUTATHIONE	glutathione	ORGANOSULFUR, All-Glutathiones, Thiols	Ll	Lp
h2o2_e	HYDROGEN-PEROXIDE	hydrogen peroxide	Peroxides, Others	Ll	Lp
h2o_e	WATER	H2O	Pseudo-Compounds, Others	Pf; Ll	Lp
h2s_e	HS	hydrogen sulfide	Ions, Anions, Inorganic-Ions	Ll	Lp; Pf
ile__L_e	ILE	isoleucine	All-Amino-Acids, Acids, Amino-Acids, Organic-Acids	Lp	Pf; Ll
lac__D_e	D-LACTATE	d-lactate	Acids, Organic-Acids	Pf	Ll; Lp
lac__L_e	L-LACTATE	l-lactate	Acids, Organic-Acids	Ll	Lp; Pf
mal__L_e	MAL	malate	Acids, Organic-Acids	Pf; Lp	Ll
nh4_e	AMMONIUM	ammonium	Ions, Inorganic-Ions, Cations	Lp	Pf; Ll
phe__L_e	PHE	phenylalanine	Aromatics, All-Amino-Acids, Acids, Amino-Acids, Organic-Acids, Organic-aromatic-compounds	Pf	Lp
pro__L_e	PRO	proline	All-Amino-Acids, Acids, Amino-Acids, Organic-Acids	Ll	Pf
ser__D_e	D-SERINE	serine	All-Amino-Acids, Acids, Amino-Acids, Organic-Acids	Ll	Lp; Pf
ser__L_e	SER	serine	All-Amino-Acids, Acids, Amino-Acids, Organic-Acids	Ll	Lp; Pf
succ_e	SUC	succinate	Acids, Organic-Acids	Ll	Pf
xan_e	XANTHINE	xanthine	Others, Others	Ll	Lp; Pf

TABLE 2.5 – Métabolites échangeables sur la communauté du fromage identifiés par MICOM Abréviations : Ll, *L. lactis* ; Lp, *L. plantarum*, Pf, *P. freudenreichii*

ID Bigg	ID Metacyc	Nom	Ontologie Metacyc	fluxes d'export	fluxes d'import
acald_e	ACETALD	acetaldéhyde	Aldehydes-Or-Ketones, Aldehydes	Pf Lp Ll	Ll Lp Ll Lp
akg_e	2-KETOGLUTARATE	2-oxoglutarate	Acids, Organic-Acids	Lp Pf	Pf Lp
fum_e	FUM	fumarate	Acids, Organic-Acids	Pf	Lp
glyc_e	GLYCEROL	GLYCEROL	Alcohols, All-Carbohydrates, Sugar-alcohols, Carbohydrates	Lp Ll Ll	Pf Pf Lp
glyclt_e	GLYCOLATE	GLYCOLATE	Acids, Organic-Acids	Lp	Pf
h2s_e	HS	hydrogen sulfide	Ions, Anions, Inorganic-Ions	Lp Pf Ll	Pf Lp Lp
lac__D_e	D-LACTATE	d-lactate	Acids, Organic-Acids	Pf Ll	Lp Lp
lac__L_e	L-LACTATE	l-lactate	Acids, Organic-Acids	Ll	Lp
mal__L_e	MAL	malate	Acids, Organic-Acids	Pf Lp Pf Ll	Ll Ll Lp Lp
phe__L_e	PHE	phenylalanine	Aromatics, All-Amino-Acids, Acids, Amino-Acids Organic-Acids, Organic-aromatic-compounds	Lp	Pf
rib__D_e	D-Ribopyranose	D-Ribopyranose	Aldehydes-Or-Ketones, All-Carbohydrates, Aldehydes, Carbohydrates	Ll Lp Pf Ll	Pf Pf Lp Lp
ser__D_e	D-SERINE	serine	All-Amino-Acids, Acids, Amino-Acids, Organic-Acids	Ll Lp Pf Lp Pf Ll	Pf Pf Ll Ll Lp Lp
succ_e	SUC	succinate	Acids, Organic-Acids	Ll Lp	Pf Pf
xan_e	XANTHINE	xanthine	Others, Others	Ll Lp	Pf Pf

TABLE 2.6 – Métabolites échangeables sur la communauté du fromage identifiés par SMETANA. Abréviations : Ll, *L. lactis* ; Lp, *L. plantarum*, Pf, *P. freudenreichii*

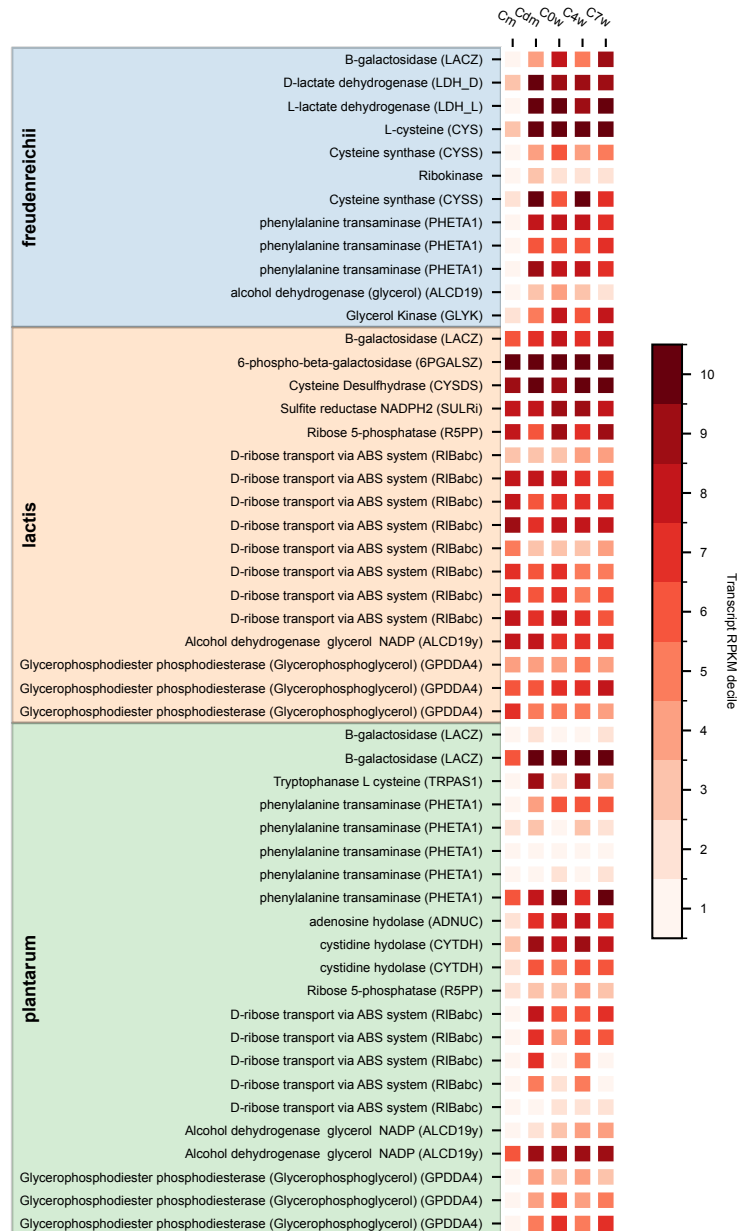


FIGURE 2.14 – **Expression des gènes des données métatranscriptomiques** Cette carte thermique affiche le décile du nombre de RPKM de transcrits de gènes à un temp donné (colonnes) et dans un micro-organisme donné (les zones colorées indiquent à quelle bactérie le gène appartient). Les déciles ont ensuite été associés à un gradient de couleurs, des déciles les plus élevés (rouge foncé, expression la plus élevée dans ce micro-organisme à ce moment précis) aux déciles les plus bas (rouge clair, faible expression). Les gènes affichés ont été sélectionnés manuellement à partir de la liste des interactions possibles établie par SMETANA : pour un métabolite d'interaction sélectionné, les gènes impliqués dans la production (chez le donneur) et dans la consommation (chez le receveur) ont été retenus pour l'analyse. Les métabolites d'interaction sélectionnés sont le *H2S*, le ribose et la phénylalanine. En outre, les gènes impliqués dans la consommation de lactose ont été ajoutés à cette liste. Abréviations : Cm, moulage (t = 19.5 hours) ; Cdm, démoulage (t = 40 hours) ; C0w, début de l'affinage (t = 60 hours) ; C4w, 4 semaines d'affinage (t = 732 hours) ; C7w, 7 semaines d'affinage (t = 1236 hours).

L'ensemble des métabolites dont la production est prédite par Metage2Metabo comme

nécessitant des interactions dans la communauté inclut divers acides gras, comme le benzyl-Coa, le glyceraldehyde et la xanthosine, indiquant une complémentarité métabolique potentielle entre les souches.

ID Bigg	ID Metacyc	Nom	Ontologie Metacyc
galt1p benzcoa xtsn	D-galactose-1-phosphate BENZOYLCOA XANTHOSINE	a D-galactopyranose 1-phosphate benzoyl-CoA XANTHOSINE	All-Carbohydrates, Carbohydrates Esters, Thioesters Organic-heterocyclic-compound, All-Nucleosides, Nitrogen-Molecular-Entities, Organic-heteromonocyclic-compounds, Organic-Heteropolycyclic-Compounds, Organonitrogen-Compounds, Nucleosides Aldehydes-Or-Ketones, All-Carbohydrates, Aldehydes, Carbohydrates
glyald 3hocoa	GLYCERALD ∅	glyceraldehyde ∅	∅
3hhdcoa	CPD0-2232	(S)-3-hydroxyhexadecanoyl-CoA	Thioesters, Esters
3hdcoa	CPD0-2244	(S)-3-hydroxydecanoyl-CoA	Thioesters, Esters
3odcoa	CPD0-2123	3-oxodecanoyl-CoA	Thioesters, Esters
3hhcoa	OH-HEXANOYL-COA	(S)-3-hydroxyhexanoyl-CoA	Thioesters, Esters
udcpp	UNDECAPRENYL-P	all-trans-undecaprenyl phosphate	Lipids, Polyisoprenoids
3oocoa	CPD0-2106	3-oxooctanoyl-CoA	Thioesters, Esters

TABLE 2.7 – Liste de métabolites prédits par Metage2Metabo comme étant productibles uniquement par coopération métabolique

2.4 Discussion et Conclusion

Nous avons construit un modèle numérique du métabolisme bactérien décrivant du point de vue métabolique, le processus de fabrication de fromage. En utilisant les modèles métaboliques de trois souches bactériennes et des données omiques, nous avons reproduit avec précision, les motifs de production de métabolites ainsi que la dynamique des populations bactériennes. En nous reposant sur les conditions expérimentales de cultures pures pour calibrer les modèles métaboliques individuels avec un petit nombre de paramètres pour chacun, nous avons ainsi limité le sur-apprentissage du modèle durant les simulations individuelles. Une originalité du modèle est sa capacité à prédire les dynamiques sur tout le procédé de fabrication du fromage (sept semaines). Notre travail a généré des hypothèses sur les voies métaboliques mises en jeu dans le lait, ainsi que la contribution de chaque souche bactérienne dans la consommation de nutriments et la productibilité de composés organoleptiques.

La connaissance des microbiologistes, la littérature scientifique et des données multi-omiques étaient nécessaires pour obtenir des modèles métaboliques à l'échelle du génome et des simulations dynamiques de qualité. Le métabolisme carboné central des bactéries lactiques produit principalement l'acide lactique à partir de la dégradation du lactose *via* les voies métaboliques de la glycolyse, du tagatose et de Leloir (Kleerebezem et al., 2003; Van Rooijen et al., 1991; Widyastuti et al., 2014). Pour la première fois, l'utilisation du citrate et la fermentation hétérolactique ont été observées dans le modèle curé de *L. plantarum* en lait. La vérification des modèles métaboliques individuels et de l'activation des différentes voies métaboliques décrites plus haut, a été une étape importante pour valider les modèles, permettant de reproduire des activités enzymatiques observées dans (Carroll et al., 1999; Quatravaux et al., 2006). La bactérie propionique *P. freudenreichii* convertit le lactose, et préférentiellement le lactate en propionate d'après (Borghesi et al., 2021; Loux et al., 2015; Thierry et al., 2011). Nous avons ajouté l'enzyme propionyl-CoA :succinate Coa transferase (2.8.3.-) pour compléter le cycle de Wood-Werkman, et ainsi assurer la production de propionate.

Dans (Özcan et al., 2020), les paramètres individuels de calibration pour chaque nutriment du milieu de culture sont imposés dans le but d'expliquer les données de métabolo-

mique à l'échelle de l'espèce, et de prédire les composés de la métabolomique à l'échelle de la communauté. Dans ces travaux, une nouvelle étape d'optimisation à l'échelle de la communauté a été faite afin d'affiner leur prédiction. Nous avons implémenté un modèle FBA dynamique en se basant sur (Mahadevan et al., 2002b), qui a utilisé des paramètres de calibration spécifiques aux souches pour prédire la croissance, le pH et les concentrations métaboliques. Pour éviter le sur-apprentissage du modèle, nous avons réduit le nombre de paramètres inférés pour chaque bactérie. Seulement deux paramètres ont été gardés pour la régulation du pH pour les bactéries lactiques et un paramètre se focalisant sur les croissances de *L. plantarum* et *P. freudenreichii*. Des calibrations additionnelles ont été menées pour *P. freudenreichii* sans aucune optimisation mathématique supplémentaire, en calculant la valeur de la borne supérieure des métabolites dosés en monoculture. Ainsi, nos paramètres ne sont pas propres à chaque nutriment et nous n'avons effectué aucune optimisation à l'échelle de la communauté.

Le modèle de co-culture a donné des indices sur le comportement de la communauté pendant la fabrication du fromage et a suggéré temporellement la contribution de chaque espèce impliquée dans la production des composés organoleptiques. (Cao et al., 2021) a montré que la production de propionate peut être attribuée uniquement à *P. freudenreichii*, et que le diacétyl semble être seulement produit par *L. lactis*. D'après le modèle de communauté, le butanediol a été produit pendant l'étape du moulage et au début de l'affinage par, à la fois *L. lactis* et *L. plantarum*, comme observé dans les modèles de monocultures respectifs. L'acétate est produit précocement jusqu'au début de l'affinage par les 3 espèces, puis, *P. freudenreichii*, responsable de la majeure partie de production de l'acétate, semble le produire à flux constant jusqu'à la fin de la fabrication de fromage. Concernant l'utilisation des sources de carbone, alors que *L. plantarum* possède les voies métaboliques de dégradation du lactose et du citrate, qui sont complètement activées dans le modèle en monoculture, il apparaît que ces voies métaboliques sont fortement sous-régulées en co-culture et *L. plantarum* ne semble plus être capable de métaboliser le citrate. Lors de la mise en culture avec *L. lactis*, *L. plantarum* atteint son efficacité métabolique maximale après la phase exponentielle de *L. lactis* comme ce dernier a la plus faible vitesse de croissance. Étant donné que *L. lactis* pousse principalement sur le lactose en produisant du lactate, sa phase exponentielle est associée à une forte acidification du milieu, réprimant probablement le métabolisme de *L. plantarum* (cf équation sur la régulation du pH sur les bactéries lactiques 2.10). *L. plantarum* peut utiliser des acides aminés ou d'autres métabolites non suivis par le modèle dynamique permettant sa croissance. Pour explorer cette piste, nous avons utilisé des outils dans le but d'identifier des interactions bactériennes en dehors de l'ensemble des composés suivis par le modèle dynamique en co-culture. Nous avons remarqué que parmi les possibles interactions, le groupe des acides aminés était bien représenté.

Dans le modèle de co-culture, *P. freudenreichii* garde son métabolisme activé dans la dernière phase de production d'un fromage. Ce métabolisme est reflété par la consommation de lactate durant tout l'affinage, et le relargage de composés organoleptiques à flux constant (Figure 2.13). Ce comportement est cohérent avec les capacités métaboliques connues de *P. freudenreichii* et avec les données métabolomiques obtenues expérimentalement. Dans le modèle, *P. freudenreichii* consomme aussi le lactose restant après la croissance de *L. lactis*. En effet, durant sa phase exponentielle, *L. lactis* métabolise principalement le lactose, produisant fortement du lactate et réduisant le pH. La réduction du pH inhibe l'absorption du lactose dans les deux bactéries lactiques, et donc stoppe la

consommation de lactose par les *Lactobacilli*. Comme le lactose reste dans le milieu de culture et que *P. freudenreichii* est l'unique bactérie capable de l'utiliser, il épuise donc le lactose encore présent. Dans les données métatranscriptomiques, le gène LACZ est fortement activé par *P. freudenreichii* durant l'affinage, confirmant la dégradation du lactose par cette souche après l'acidification (Figure 2.14)

Ce chapitre a révélé comment un processus itératif combinant expertise biologique, données hétérogènes et modélisation permettent d'obtenir des prédictions précises sur les mécanismes responsables du comportement dynamique de la communauté microbienne. D'autre part, cela souligne également que la quantité de données et les efforts nécessaires pour créer des modèles de haute qualité restent un prix élevé à payer malgré l'amélioration des approches de simulation. Des développements méthodologiques doivent encore être proposés afin d'automatiser la calibration des modèles avec des données, et de garantir à la fois l'exactitude des prédictions et l'extensibilité à des communautés plus importantes ou à des communautés de composition empirique.

Dans le chapitre suivant, nous présentons une approche permettant l'analyse haut débit de communautés bactériennes en utilisant une approche discrète du métabolisme. Nous nous focaliserons sur la mise en évidence de potentiels de coopération et compétition dans des communautés synthétiques et réelles. Du point de vue général de la thèse, le chapitre sur la modélisation numérique et celui sur la modélisation discrète du métabolisme constituent respectivement le cas riche et le cas pauvre pour l'analyse computationnelle du métabolisme. Un dernier chapitre se concentrera sur l'enrichissement de modèle logique pour se rapprocher du modèle numérique tout en conservant les propriétés de passage à l'échelle et d'explicabilité.

Abréviations métaboliques

13dpg 3-Phospho-D-glyceroyl phosphate
6pgc 6-Phospho-D-gluconate
6pgl 6-phospho-D-glucono-1,5-lactone
ac acétate
acald acetaldehyde
accoa acetyl-coa
actn__R acetoine
akg 2-Oxoglutarate
alac__S (S)-2-Acetolactate
btd__RR butanediol
cit citrate
dgal6p d-galactose-6-phosphate
dhap Dihydroxyacetone phosphate
diact diacetyl
f6p fructose-6-phosphate
fdp D-Fructose 1,6-bisphosphate
fum fumarate
g1p glucose-1-phosphate
g3p Glyceraldehyde 3-phosphate
g6p glucose-6-phosphate
gal galactose
gal1p galactose-1-phosphate
glc__D glucose
lac__D, lac__L citrate
lac6p lactose-6-phosphate
lcts lactose
mmcoa__R (R)-Methylmalonyl-CoA
mmcoa__S (S)-Methylmalonyl-CoA
oaa Oxaloacetate
ppa propionate
ppap Propanoyl phosphate
ppcoa propionyl-coa
pyr pyruvate
ru5p__D D-Ribulose 5-phosphate
succ succinate
succoa succinyl-coa
tag6p__D D-tagatose-6-phosphate
tagdp__D D-Tagatose 1,6-biphosphate
udpg UDPglucose
xu5p__D D-Xylulose 5-phosphate

Abréviations réactionnelles

2131pyrpp Methylmalonyl-CoA carboxyltransferase 5S subunit
6PGALSZ 6-phospho-beta-galactosidase
ACLD Acetolactate decarboxylase
ACTD2 Acetoin dehydrogenase
ACLDC Acetolactate decarboxylase
ACLS Acetolactate synthase
BTDD_RR R R butanediol dehydrogenase
CITL Citrate lyase
CITt4_1 Citrate transport via sodium symport
CS Citrate synthase
D_LACt2 D lactate transport via proton symport
ENO Enolase
FBA Fructose-bisphosphate aldolase
FEDCabc FEDCabc
FRD2rpp Fumarate reductase/succinate dehydrogenase
G6PDH2r Glucose 6-phosphate dehydrogenase
GAL6PI Galactose-6-phosphate isomerase
GALkr Galactokinase
GALUi UTP-glucose-1-phosphate uridylyltransferase (irreversible)
GAPD Glyceraldehyde-3-phosphate dehydrogenase
GND Phosphogluconate dehydrogenase
HEX1 Hexokinase (D-glucose :ATP)
L_LACt2r L lactate reversible transport via proton symport
LACpts Lactose transport via PEP :Pyr PTS
LACZ B-galactosidase
LCTSt3ipp Lactose transport via proton aniport (periplasm)
LDH_D D-lactate dehydrogenase ;
LDH_L L-lactate dehydrogenase
MME Methylmalonyl-CoA epimerase
MMM2 Methylmalonyl-CoA mutase
OOR3r 2-oxoglutarate synthase (rev)
PC Pyruvate carboxylase
PFK_2 Phosphofructokinase
PFL Pyruvate formate lyase ;
PGI Glucose-6-phosphate isomerase
PGK Phosphoglycerate kinase
PGL 6-phosphogluconolactonase
PGM Phosphoglycerate mutase
PGMT Phosphoglucomutase
phosphoketolase phosphoketolase
PPAKr Propionate kinase
PPCSCT Propanoyl-CoA : succinate CoA-transferase
PTA Phosphotransacetylase
PTA2 Phosphate acetyltransferase
PYK Pyruvate kinase
RPE Ribulose 5-phosphate 3-epimerase
TGBPA Tagatose-bisphosphate aldolase

UGLT UDPglucose–hexose-1-phosphate uridylyltransferase

Chapitre 3

Modèle discret pour le criblage haut-débit du potentiel d'interaction métabolique dans les écosystèmes microbiens complexes

Sommaire

3.1	Introduction	83
3.2	Méthode pour calculer un score d'interaction à partir d'une base de connaissance	85
3.2.1	Base de faits d'un réseau métabolique	85
3.2.2	Productibilité d'un métabolite	85
3.2.3	Inférence de règles logiques pour calculer les potentiels de coopération et de compétition	86
3.2.4	Calcul des potentiels de coopération et de compétition à partir des règles logiques	89
3.2.5	Expériences <i>in silico</i> pour tester les propriétés des scores	91
3.3	Application et description de notre approche sur des communautés jouets	94
3.4	Application à des communautés synthétiques	101
3.4.1	Description des potentiels de coopération et de compétition dans différents écosystèmes	101
3.4.2	Comparaison avec des méthodes basées sur des contraintes et application à des données expérimentales	112
3.5	Discussion et conclusion	115

Article en cours de rédaction

3.1 Introduction

Dans le chapitre précédent, nous avons développé une approche itérative permettant de prédire des concentrations de métabolites, la densité bactérienne ainsi que quelques interactions dans un écosystème composé de 3 souches bactériennes. Cette approche est numériquement trop coûteuse en temps et en données pour permettre une analyse à haut-débit d'écosystèmes microbiens plus complexes. De plus, la prédiction des interactions bactériennes a pu être faite avec confiance grâce à la connaissance *a priori* des biologistes, de la littérature et des données expérimentales. Dans le cadre d'un écosystème naturel, appliquer le processus itératif décrit dans le chapitre 2 est compromis par le manque de connaissances *a priori* du système. En effet, raffiner des modèles métaboliques appartenant à des organismes peu étudiés et générer des données de calibration pour des communautés composées de centaines d'espèces sont irréalisables.

La stabilité des communautés microbiennes repose en partie sur les interactions bactérie-bactérie, principalement caractérisées par la consommation et la production de métabolites. Ces interactions sont particulièrement importantes car elles sont responsables du fonctionnement du microbiome : protection de l'hôte dans le microbiote de l'intestin (Wilmes et al., 2022), recyclage de matières organique dans le sol (Rousk et al., 2016), participation à la croissance de la plante (Fournier et al., 2022) ou encore, développement d'arôme dans la fermentation du fromage (Cao et al., 2021). Pour rappel, il existe plusieurs types d'interaction comme illustré par la Figure 1. Des interactions coopératives peuvent être définies comme un échange métabolique positif, sans effet négatif pour la bactérie productrice et la receveuse. A l'opposé, la compétition peut être décrite comme une co-consommation d'un substrat limitant, impactant la croissance et /ou la production métabolique des bactéries impliquées.

Dans la littérature, nous avons observé que les méthodes pour identifier des potentiels de coopération et de compétition dans des communautés de grande taille, se positionnent à l'échelle du métabolisme (voir chapitre 1). En se basant sur les relations d'association gène-protéine-réaction d'un réseau métabolique à l'échelle du génome (GEM), une représentation sous forme d'une matrice stœchiométrique, d'un graphe ou à l'aide d'approches de raisonnement peut être utilisée pour mettre en avant ces interactions. Comme nous l'avons vu, les approches numériques sont principalement limitées par le coût calculatoire. Les approches basées sur les graphes ne sont pas non plus adaptées à la caractérisation de ces potentiels d'interactions lorsque que le nombre d'organisme dans la communauté est important : les analyses se font par paires d'organismes, rendant fastidieuse l'analyse à l'échelle de la communauté. Il existe cependant un compromis, permettant de caractériser la communauté dans son ensemble, comme pour les méthodes numériques, et permettant de passer à l'échelle de grandes communautés : **l'approche par raisonnement** (Belcour et al., 2020; Frioux et al., 2018). Dans ce cadre, des règles décrivant la notion de productibilité d'un composé et le comportement d'une communauté sont décrites au moyen d'un formalisme logique (voir chapitre 1).

Pour caractériser les potentiels d'interactions à l'aide d'une approche par raisonnement, une base de connaissances décrivant l'objet d'étude sous forme d'une liste de faits biologiques doit être créée. Dans notre cas, elle est constituée des attributs d'un réseau métabolique que sont : produit, réactant, réaction, métabolite. Cette base de connaissances va permettre d'analyser efficacement la productibilité des composés, la faisabilité

des réactions ou encore, d'identifier les échanges métaboliques etc. Toutes ces propriétés émergentes sont calculables à partir d'un ensemble de règles inspirées des définitions biologiques décrivant ces mêmes particularités. Ainsi, chaque modèle en sortie découlera explicitement de règles biologiques auditables. Le paradigme de programmation par ensemble de réponse (*Answer Set Programming* ou ASP) a été choisi pour ses qualités : explicabilité des modèles et résolution des problèmes combinatoires.

Dans ce chapitre, nous avons ainsi utilisé ASP et nous sommes inspirés de l'algorithme d'expansion du réseau métabolique d'Ebenhöh (Ebenhöh et al., 2004) et de son élargissement à la caractérisation de communauté (Frioux et al., 2018) en créant une nouvelle base de connaissances à partir d'un ensemble de GEMs. Pour exploiter cette base de faits, des nouvelles règles logiques sont créées permettant le calcul d'un *potentiel de coopération* et d'un *potentiel de compétition*. Dans la suite, nous présenterons la nouvelle base de connaissances et les règles d'inférence logique que nous avons développées, puis une série de tests montrant la pertinence des scores de coopération et de compétition. Enfin, une comparaison à des méthodes quantitatives sera effectuée.

Les travaux présentés dans ce chapitre ont mené à un article en cours de rédaction et à plusieurs posters scientifiques et présentations orales (Lecomte et al., 2022a,b,c)

3.2 Méthode pour calculer un score d’interaction à partir d’une base de connaissance

Le choix de la représentation de la connaissance est primordial pour toute approche par raisonnement. Nous détaillerons dans un premier temps, le formalisme utilisé pour représenter la base de faits, ainsi que les règles logiques pour inférer les potentiels d’interaction. Nous développerons ensuite notre méthode de calcul de potentiels d’interaction.

3.2.1 Base de faits d’un réseau métabolique

À partir de la donnée d’entrée, des GEMs, nous définissons un ensemble de *réactions* biochimiques qui produisent et consomment des biomolécules. Lors d’un échange de biomolécules entre taxons, nous insistons sur une notion centrale dans cette approche par raisonnement qui est la *provenance* de la biomolécule. La provenance peut s’appliquer à toute la chaîne des réactions à l’origine d’une molécule, mais par convenance nous ne considérons que le dernier taxon qui la produit. Ainsi, un *métabolite* est formellement une biomolécule dont la *provenance* est connue. Soit \mathcal{M} l’ensemble des métabolites définis sur un ensemble \mathcal{T} de taxa et un ensemble \mathcal{B} de biomolécules :

$$\begin{aligned} \text{taxon} : \mathcal{M} &\rightarrow \mathcal{T} \\ \text{biomolécule} : \mathcal{M} &\rightarrow \mathcal{B} \end{aligned} \tag{3.1}$$

En regardant les relations métabolite-réaction, nous pouvons retrouver les ensembles de métabolites qui peuvent être produits par les réactions, les taxa, ainsi que les ensembles de biomolécules servant de réactants à ces réactions. Dans le cas de la production d’un métabolite, l’ensemble des taxons produisant la biomolécule peut-être retrouvée.

Tout comme pour le modèle numérique, un choix concernant la prise en compte des réactions de transport présentes dans le réseau métabolique a été fait. Il a été montré que le phénomène de lyse cellulaire est une potentielle source d’interaction métabolique, entraînant la mort de la cellule et libérant son contenu métabolique dans le milieu extracellulaire (Fazzino et al., 2020). De plus, (Pacheco et al., 2019) ont montré que les sécrétions passives de métabolites étaient une source de mutualisme au sein de communautés bactériennes. De ce fait, lors de la conception du modèle de communauté, nous avons choisi de considérer les ensembles de métabolites produits en intracellulaire disponibles pour les autres taxons. Ainsi, un modèle de Communauté $\mathbf{G} = \{G_1, \dots, G_n\}$ peut-être décrit comme le produit cartésien de l’union des métabolites et réactions défini par le graphe bipartite, et se lit ainsi :

$$\text{merge}(\mathbf{G}) = \left\langle \bigcup_{i=1}^n M_i, \bigcup_{i=1}^n R_i, \bigcup_{i=1}^n E_i \right\rangle .$$

3.2.2 Productibilité d’un métabolite

Pour rappel, selon l’algorithme d’expansion de réseau Ebenhöh et al. (2004), un métabolite mp est productible si l’ensemble des réactants d’une réaction, permettant la production de ce métabolite mp , sont disponibles à partir d’un ensemble de graines, constituant l’ensemble des composés du milieu nutritif. Dans la suite de ce chapitre, nous qualifierons

de *scope* l'ensemble des métabolites productibles par un réseau à partir de graines. Dans la littérature plusieurs types de *scope* sont définis (Frioux et al., 2018) pour projeter ces ensembles sur l'individu ou sur la communauté. Mais cette déclinaison n'est pas nécessaire du moment où nous réfléchissons l'appartenance à une communauté sous forme de relation dans la base de connaissances. Ainsi, nous créons un ensemble de communautés, *a minima* une constituée de l'ensemble des taxons et une pour chaque taxon individuel. L'appartenance d'une biomolécule au *scope* d'une communauté donnée est enregistrée dans la base de connaissances par une relation `in_community`. Du fait de cette représentation, nous n'avons plus la nécessité de distinguer le *scope* associé à un individu de celui de la communauté comme dans (Frioux et al., 2018).

3.2.3 Inférence de règles logiques pour calculer les potentiels de coopération et de compétition

Cette sous-section présente l'implémentation en ASP des définitions de `scope`, de coopération et de compétition.

Inférence du *scope* métabolique. Le *scope* associe un objet métabolite, dans lequel un nom et un taxon de provenance sont connectés, (`metabolite(M,B)`) et une communauté (`Communaute`). le *scope* métabolique d'une communauté est défini par le code 3.1 et 3.2. Dans la suite de cette section, chaque listing sera expliqué ligne par ligne sous forme d'item, afin de comprendre les règles logiques.

```

1 scope(metabolite(M,B), Community) :-
2   taxon(B),
3   in_community(Community, B),
4   reaction(R,B),
5   product(M,R,B),
6   available(M2, Community) : reactant(M2,R,B).

```

Listing 3.1 – Code ASP permettant de calculer le *scope* métabolique

Au sein du listing 3.1,

ligne 1 : un métabolite M provenant d'un taxon B est dans le *scope* d'une communauté si :

ligne 2 : B est un taxon et

ligne 3 : que B est un membre de la communauté et

ligne 3-4 : qu'il existe une réaction R du taxon B qui produit M et

ligne 5 : que tous les métabolites réactants de la réaction R du taxon B sont disponibles pour la communauté.

La notion de *scope* individuel ou *iscope* dans (Frioux et al., 2018) par rapport à un taxon est simplement le *scope* du listing 3.1, calculé en fixant le nom `Community` à celui du taxon individuel.

De cette règle utilise le concept de *disponibilité* d'un métabolite (listing 3.2) :

```

1 available(M, Community) :- seed(M), community(Community).
2 available(M, Community) :- scope(metabolite(M,B), Community).

```

Listing 3.2 – Code ASP permettant de savoir si un métabolite est disponible

ligne 1 : Un métabolite M est disponible pour une communauté ($Community$) si le métabolite M est une graine ($seed$) et que $Community$ est une communauté de GEMs ou

ligne 2 : que le métabolite M appartenant à un taxon B est dans le **scope** métabolique de la communauté $Community$.

Il nous paraît pertinent de tracer la consommation et l'éventuelle production des graines par les membres de la communauté (listings 3.3 et 3.4), afin d'explorer l'interaction d'une communauté microbienne avec son environnement nutritionnel.

```
1 produced_seed(S, Community) :- seed(S), product(S,R,B),
2   activated(reaction(R,B), Community).
```

Listing 3.3 – Code ASP déterminant la production d'une graine par une communauté

```
1 consumed_seed(S, Community) :- seed(S), reactant(S,R,B),
2   activated(reaction(R,B), Community).
```

Listing 3.4 – Code ASP déterminant la consommation d'une graine par une communauté

Une graine S est produite (resp. consommée) si elle est le produit (resp. réactant) d'une réaction activée dans la communauté. La notion de *réaction activée*, est exprimée dans le listing 3.5.

```
1 activated(reaction(R,B), Community) :-
2   reaction(R,B),
3   community(Community),
4   in_community(Community, B),
5   available(M, Community) : reactant(M,R,B).
```

Listing 3.5 – Code ASP déterminant les réactions activées dans une communauté

Ainsi, une réaction R du taxon B , dans une communauté $Community$ est activée si :

ligne 2 : R est une réaction et

ligne 3 : $Community$ est une communauté et

ligne 4 : que B est un taxon de la communauté et

ligne 5 : que tous les réactants M de la réaction R du taxon B sont disponibles pour la communauté.

Ces notions de graines produites et consommées peuvent être des proxys du statut de nutriment en excès ou limitant. Par exemple, une graine présente initialement dans l'environnement, qui n'est jamais consommée mais est néanmoins produite suggère une graine potentiellement en excès.

Inférence des métabolites échangeables. Nous faisons le postulat que les métabolites échangés correspondent à l'approximation la plus directe de la coopération d'une communauté. Ainsi, un métabolite est échangé entre un taxon producteur P et un taxon consommateur C , en prenant en compte les graines S , si le producteur peut synthétiser la biomolécule mais le consommateur ne le peut pas. De manière plus formelle, le métabolite est dans le scope S de la communauté héritant d'une provenance de P , mais qu'il est pas productible à partir des graines par C tout seul.

Le code ASP correspondant est défini par le code 3.6 :

```
1 exchange(M,P,C) :- taxon(P),
2   taxon(C),
3   P != C,
```

```

4     reactant(M,_,C),
5     product(M,_,P),
6     scope(metabolite(M,P), all),
7     not scope(metabolite(M,C), self(C)).

```

Listing 3.6 – Code ASP permettant d’obtenir l’ensemble des métabolites potentiellement échangés.

où :

- ligne 1 : un métabolite M est échangeable entre un producteur P et un consommateur C si P est un taxon et
- ligne 2 : que C est un taxon et
- ligne 3 : P est différent de C et
- ligne 4 : que M est un réactant d’une réaction de C et
- ligne 5 : que M est un produit d’une réaction de P et
- ligne 6 : que M est dans le `scope` communautaire produit par P et
- ligne 7 : que M n’est pas dans le `scope` de C.

Nous avons décrit le *scope* métabolique pour un GEM ou une communauté de GEMs, et identifié les métabolites échangeables comme un proxy de la coopération. La compétition quant à elle va reposer sur la notion de *substrat limitant*, reliée à la quantité de ce substrat disponible pour la communauté. Ce formalisme discret ne permettant actuellement pas de prendre la notion de quantité, nous nous sommes inspirés d’un concept d’économie pour définir une compétition : la *polyopsonie*. D’origine grecque, l’étymologie de *monopsonie* appliqué à l’économie signifie un marché dans lequel il y a un seul acheteur (*mono* signifie un, et *opsonia* signifie achat). Un marché où il existe un petit nombre d’acheteurs est caractérisé par l’oligopsonie. En s’inspirant des deux concepts économiques ci-dessus, nous avons élaboré le concept de *polyopsonie* pour représenter un marché où il existe plusieurs acheteurs, dans notre cas plusieurs consommateurs d’un substrat limitant.

Inférence de la polyopsonie. Nous utilisons donc le concept de polyopsonie comme proxy de la compétition en faisant l’hypothèse que les substrats limitants, *i.e.* métabolites échangeables et les graines consommés par plus d’un organisme, sont susceptibles de faire l’objet de compétition.

Le facteur déterminant un substrat limitant est le nombre de taxons consommant le métabolite. Ainsi, dans le code ASP permettant d’obtenir l’ensemble des substrats limitants (voir listing 3.7), nous comptons le nombre de taxons consommant chaque substrat, échangé ou considéré comme un nutriment, et appliqué une condition sur le nombre de consommateurs :

```

1 % cas pour les metabolites échangeables
2 polyopsonist(M,N) :- N=#count{C,M:exchange(M,_,C)},
3                       exchange(M,_,_), N > 1.
4
5 % cas pour les graines
6 polyopsonist(S,N) :- N=#count{B:seed_consumed_by_taxon(S,B)},
7                       N > 1, seed(S).

```

Listing 3.7 – Code ASP déterminant les substrats limitants consommés par plusieurs taxons

où :

- ligne 2-3 : Un composé échangé M est limitant si lorsque le nombre de consommateurs C impliqués dans l'échange de ce métabolite M est strictement supérieur à 1.
- ligne 6-7 : Une graine M est considérée limitante lorsque le nombre de consommateurs C de cette graine M est strictement supérieur à 1.

Les codes ASP que nous avons décrit correspondent à des règles qui sont biologiquement explicables et génèrent un vocabulaire contrôlé (par cet ensemble de règles). Dans les sections suivantes nous montrerons comment à partir cette extension de la base de connaissances nous avons calculé les potentiels de coopération et de compétition.

3.2.4 Calcul des potentiels de coopération et de compétition à partir des règles logiques

À partir des atomes déduits dans les règles précédemment décrite, nous souhaitons calculer des potentiels de coopération et de compétition. Ces scores ont pour objectif de confronter des communautés d'une taille comparable entre elles et de pouvoir les classer de la plus coopérative à la plus compétitive.

Calcul du potentiel de coopération L'inférence du potentiel de coopération se base sur les taxons impliqués dans l'échange d'un métabolite (voir listing 3.6). Pour chaque métabolite échangeable, nous avons un ensemble non nul de paires de producteurs et consommateurs.

À partir de ces ensembles, identifier les taxa réellement producteurs et consommateurs nécessite l'ajout de connaissances *a priori*. Désirant de pouvoir appliquer notre démarche à n'importe quel écosystème et sans ajout de connaissances, nous avons supposé que chacun des taxa producteurs (resp. consommateurs) contribue à l'échange. Une contribution égale de chaque taxon producteur (resp. consommateur), c'est à dire que chaque taxon peut produire (resp. consommer) le métabolite échangé, surestimerait la coopération. À l'inverse, supposer qu'un seul taxon parmi les producteurs (resp. consommateurs) contribue à l'échange, pourrait sous-estimer la coopération.

Lors d'un échange métabolique, il est probable qu'un taxon producteur ne puisse pas fournir le composé échangé à l'ensemble des taxons consommateurs. Un compromis équitable a consisté à stipuler que chaque taxon producteur (resp. consommateur) peut produire (resp. consommer) le métabolite échangé, mais la contribution des taxa n'est pas égale. Ainsi, pour chaque métabolite échangeable, un poids w de 1 est attribué au premier producteur (resp. consommateur) de l'ensemble des paires de producteurs-consommateurs puis d'ajouter un bonus exponentiel dégressif pour chacun des producteurs (resp. consommateurs) de l'ensemble des paires de producteurs-consommateurs restantes. Formellement, ce poids w s'écrit :

$$w(k) = 2 - 0.5^{k-1}$$

dans lequel, k représente la taille de l'ensemble des producteurs (resp. consommateurs) pour un métabolite échangeable.

Il est important de noter que w n'a pas de sémantique en biologie et ne sert qu'à pondérer la redondance de la production et de la consommation des molécules. Nous ne prétendons pas que $2 - 0.5^{k-1}$ est le coefficient optimal mais il permet de ne pas surestimer

ou sous-estimer la coopération en attribuant tout de même une contribution pour chaque organisme. De plus, ce poids a permis à la fois de pénaliser la redondance de producteurs (resp. consommateurs) pour chaque nouveau métabolite échangeable, et à la fois suppose qu'un métabolite échangé n'est pas obligatoirement disponible pour l'ensemble des populations.

En appliquant ce poids aux ensembles de producteurs et de consommateurs, nous pouvons déterminer un score de coopération pour ces deux ensembles. Pour un ensemble de métabolites échangeables M , dans lesquels P_m et C_m sont les ensembles de producteurs et des consommateurs associés à un métabolite échangeable m , la contribution au score de coopération est la somme pondérée des tailles de ces ensembles. Le score final est la somme de ces contributions :

$$\text{CooP} = \sum_{m \in M} w(|P_m|) + \sum_{m \in M} w(|C_m|)$$

implémentée en Python comme décrit dans le listing 3.8 :

```
1 Coop = sum(weight(len(p for p, _ in pairs)) for pairs in exchanges.values())
2     + sum(weight(len(c for _, c in pairs)) for pairs in exchanges.values())
```

Listing 3.8 – Code Python permettant de calculer le score de coopération.

Calcul du score de compétition. L'inférence de la compétition se base sur le nombre de consommateurs de substrats limitants. Le potentiel de compétition se calcule en considérant l'ensemble des consommateurs de tous les substrats polyopsonistes, représenté par `sum(polyopsonist.values())`, divisé par la taille de la communauté, représenté par `len(community.taxa)` dans le listing 3.9.

```
1 Comp = sum(polyopsonist.values()) / len(community.taxa)
```

Listing 3.9 – Code Python permettant de calculer le score de compétition.

Calcul de scores complémentaires : rho (ρ) et delta (δ). La plus-value d'une communauté représente les capacités métaboliques qui émergent des interactions entre les membres de la communauté. Nous proposons deux métriques pour évaluer cette plus-value. Delta (δ) est la différence entre le nombre de métabolites productibles par la communauté et la taille de l'union des métabolites productibles individuellement. Rho (ρ) est la différence entre le nombre de tuples réactions activées×taxon dans la communauté et la taille de l'union des tuples réactions activées×taxon individuels.

Ces deux mesures rho et delta se traduisent par les codes Python 3.10 et 3.11

```
1 community_active = len(community.activated(seeds))
2 individual_active = sum(len(community.activated(seeds, choice=t)) for t in
3     community.taxa)
3 score = community_active - individual_active
```

Listing 3.10 – Code Python permettant de calculer le score rho

où :

ligne 1 : L'ensemble des réactions activées en communauté à partir du milieu nutritionnel (`seeds`) est calculé et la taille de l'ensemble est assignée à la variable `community_active`

ligne 2 : Pour chaque taxon présent dans la communauté, l'ensemble des réactions activées d'un taxon (symbolisé par `choice=t`) à partir du milieu nutritionnel est calculé. Puis une somme du nombre de réactions activées de chaque taxon est calculée et assignée à la variable `individual_active`.

ligne 3 : Différence entre le nombre total de réactions activées en communauté et individuellement.

```
1 community_scope = len(community.scope(seeds))
2 individual_scope = sum(len(community.scope(seeds, choice=t)) for t in
    community.taxa)
3 score = community_scope - individual_scope
```

Listing 3.11 – Code Python permettant de calculer le score delta.

où :

ligne 1 : L'ensemble des métabolites productibles en communauté à partir du milieu nutritionnel (`seeds`) est calculé et la taille de l'ensemble est assignée à la variable `community_scope`

ligne 2 : Pour chaque taxon présent dans la communauté, l'ensemble des métabolites productibles d'un taxon (symbolisé par `choice=t`) à partir du milieu nutritionnel est calculé. Puis une somme du nombre de métabolites productibles de chaque taxon est calculée et assignée à la variable `individual_scope`.

ligne 3 : Différence entre le nombre total de métabolites productibles en communauté et individuellement.

3.2.5 Expériences *in silico* pour tester les propriétés des scores

Afin de tester les scores précédents, nous avons généré des communautés artificielles à partir de collections de réseaux métaboliques et calculé les potentiels de coopération et de compétition pour chacune d'entre elles. Nous avons également comparé ces potentiels avec des scores similaires produits par d'autres outils sur différents jeux de données. La sous-section suivante explique la provenance des données ainsi que les différents tests effectués.

Données et génération de communautés synthétiques Nous avons collecté des génomes appartenant à différents écosystèmes : 857 du sol de la base de données Ref-Soil Choi et al. (2017), 1520 du microbiote humain Zou et al. (2019), 186 de la racine d'*Arabidopsis thaliana* (PRJNA297942) et 189 de la feuille (PRJNA297956) d'*Arabidopsis thaliana* Bai et al. (2015). Nous avons effectué une annotation structurale avec Prodigal Hyatt et al. (2010) v.2.6.3 et une annotation fonctionnelle avec eggNOG-mapper Cantalapiedra et al. (2021) v.2.1.6 sur la base de données eggNOG Huerta-Cepas et al. (2019) v.5.0.2. Nous avons reconstruit les réseaux métaboliques à l'échelle du génome en utilisant Pathway-Tools v.25.5 Karp et al. (2022) et mpwt v.0.7 Belcour et al. (2020). Nous avons ensuite définis les milieux nutritionnels (graines) à tester : un milieu générique contenant des macros et micro-nutriments standards, et un milieu spécifique à chaque écosystème. Ces milieux sont disponibles dans l'annexe 2.

À partir de ces données, nous avons créé des communautés artificielles pour chaque écosystème. Ainsi, pour l'intestin, la feuille, la racine et le sol, 50 communautés de taille 5, 10, 20, 30, 50, 75, 100 et 150 ont été générées aléatoirement. Nous avons en plus créé une communauté factice, noté *mixte* dans le reste du chapitre, composée de mélanges de réseaux métaboliques des écosystèmes pour les mêmes tailles de communautés et dans des

proportions égales.

Afin de vérifier si la structure du réseau a une importance dans les scores des potentiels d'interaction, nous avons testé la similarité des réseaux métaboliques deux à deux, vérifiant la corrélation entre les scores obtenus et la topologie réactionnelle du réseau métabolique. 50 communautés de taille 2 ont été générées aléatoirement pour chaque écosystème, puis nous avons calculé les scores pour chacune des communautés. La similarité des deux réseaux de chaque communauté a été calculé avec l'indice de similarité de Jaccard selon l'équation 3.2 :

$$Jaccard(G_1, G_2) = \frac{|R_1 \cap R_2|}{|R_1 \cup R_2|} \quad (3.2)$$

où G_1, G_2 correspondent à deux réseaux métaboliques, R_1 l'ensemble des réactions métaboliques du réseaux G_1 et R_2 , l'ensemble des réactions métaboliques du réseau G_2 .

Pouvoir prédictif des scores pour chaque écosystème Nous avons testé si un potentiel d'interaction d'une communauté peut prédire l'écosystème auquel elle est associée. Nous avons utilisé des SVM, *Support Vector Machine* et le package scikit-learn version 1.1.1. Il s'agit d'apprentissage machine supervisé permettant de classer les éléments d'un jeu de données en deux groupes. La Figure 3.1 présente la procédure suivie.

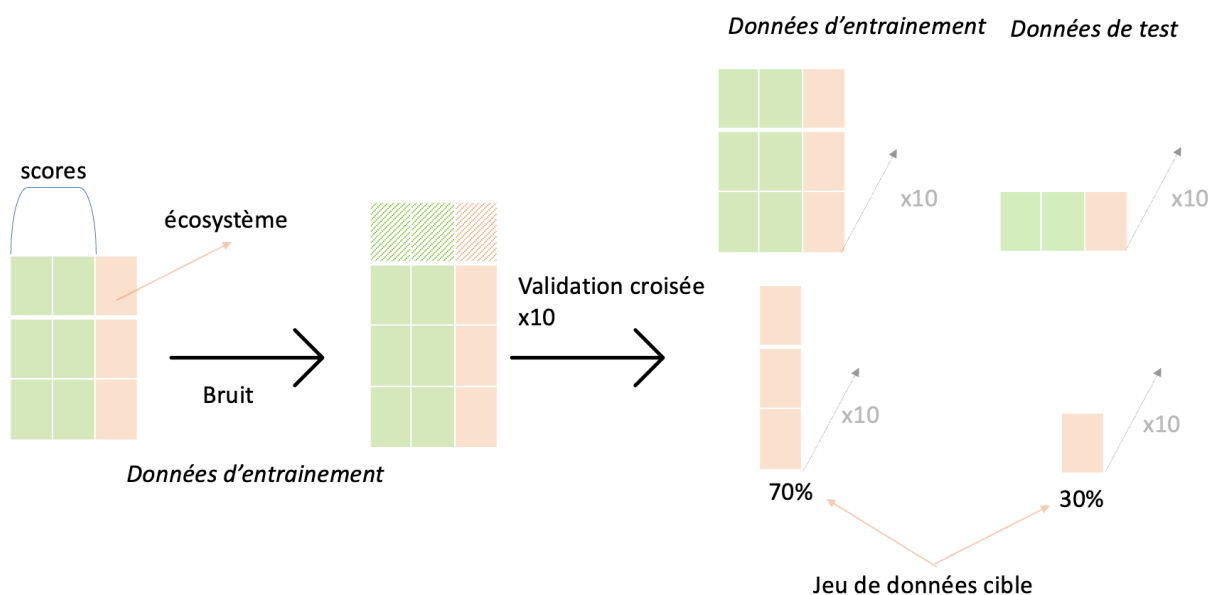


FIGURE 3.1 – Préparation des données afin d'être utilisé par le SVM. Le jeu de données d'entraînement est constitué du ou des scores d'interaction (en vert) et de l'écosystème associé à ces scores (saumon). Le jeu de données cible est constitué de l'écosystème que l'on souhaite prédire.

En entrée, deux jeux de données sont fournis : un comportant l'écosystème que l'on souhaite prédire (jeu de données cible) et un second comprenant le ou les scores d'interaction avec leur écosystème (données d'entraînement). Nous avons bruité les données d'entraînement puis effectué 10 validations croisées sur les deux jeux de données. Cela consiste à partitionner aléatoirement 10 fois les jeux de données d'entraînement et cibles de la façon suivante : 70% des jeux de données servent à entraîner le modèle et 30% servent à le tester. Par exemple, si nous voulons prédire l'écosystème de la feuille à partir du score de coopération, seul le score de coopération sera pris en compte pour l'ensemble des écosystèmes dans les données d'entraînement, et seul l'écosystème de la feuille sera présent dans le jeu de données cible.

Comparaison avec des outils numériques sur des données expérimentales Notre approche permet l'analyse de communauté de grande taille, et à ma connaissance, il n'existe pas de méthode qualitative permettant d'évaluer le potentiel de coopération et de compétition à l'échelle de la communauté. Nous avons donc comparé notre approche avec des méthodes numériques de référence basées l'analyse des flux : SMETANA (Zelezniak et al., 2015) et MiCOM (Diener et al., 2020).

Ces deux outils ont été précédemment décrits dans le chapitre 1. Brièvement, SMETANA propose deux scores pour calculer la coopération (MIP) et la compétition (MRO) dans une communauté. SMETANA calcule ses scores indépendamment du milieu de culture fourni en entrée et reflète les limites théoriques des interactions (Machado et al., 2021). Nous avons lancé SMETANA version 1.2.0 sur 250 communautés composées de 3 à 30 réseaux métaboliques reconstruits avec CarveMe (Machado et al., 2018) sur une collection de génomes bactériens de référence du NCBI RefSeq (release 84)¹. Les scores de coopération (MIP) et de compétition (MRO) calculés par SMETANA sont comparés à ceux obtenus avec notre méthode.

MICOM est originalement développé pour la modélisation du microbiote humain pour lequel l'outil prédit des taux de croissance et des flux d'échanges métaboliques. Nous avons utilisé MICOM version 0.32.0 sur 188 communautés issues d'une analyse métagénomique, de taille variant de 16 à 98 espèces (Diener et al., 2020). Les modèles métaboliques proviennent de la base de données VMH (Noronha et al., 2018) Une curation a été faite sur les réactions de transport de ces réseaux : afin de s'assurer que seules les graines du milieu nutritionnel sont utilisées comme substrat nutritif, nous avons retiré ces réactions des GEMs. Enfin, nous avons lancé les simulations sur le milieu de culture donné par MICOM : *vmh_high_fiber_agora.qza*. Cependant, MICOM ne donne pas d'indice direct de coopération et de compétition. En utilisant la fonction `knockout_taxa` nous avons mesuré l'impact de la perte d'un taxon sur la croissance d'un autre taxon. D'après la documentation de MICOM, une augmentation (resp. une diminution) du taux de croissance après l'étape de knockout indique une coopération (resp. compétition) entre ces deux espèces. En sommant les valeurs positives (resp. négatives), nous avons calculé un pseudo score de compétition (resp. coopération) à l'échelle de la communauté.

Nous avons ensuite comparé les scores calculés par notre approche, ceux de SMETANA et MICOM sur des données réelles. Nous avons récupéré des génomes de souches bactériennes du kéfir (Blasche et al., 2021) et leurs réseaux métaboliques ont été reconstruits avec gapseq version 1.2 (Zimmermann et al., 2021) et Pathway-Tools v.25.5 (Karp et al., 2022) à l'aide de mpwt v.0.7 (Belcour et al., 2020). Nous avons également comparé avec

1. https://github.com/cdanielmachado/embl_gems

les modèles déjà reconstruits par l’outil CarveMe². Ce jeu de données est composé de 2 communautés : une composée de 23 souches initialement cultivées sur milieu lait liquide et une de 31 souches initialement cultivées sur milieu lait solide ; toutes deux contiennent les mêmes nutriments. En utilisant SMETANA, MICOM et notre approche nous avons calculé les potentiels de coopération et de compétition, ainsi que le nombre de métabolites échangeables.

3.2.5.1 Implémentation

Nous avons développé un paquet Python nommé CoCoMiCo, de l’anglais *Cooperation and competition in microbial communities*. Le code source est disponible à l’adresse suivante : <https://gitlab.inria.fr/ccmc/CoCoMiCo> et est installable avec pypi <https://pypi.org/project>. Un tutoriel de notre approche est présenté sous forme de notebook jupyter³ dans lequel notre approche par raisonnement a été appliquée sur des communautés jouets décrites dans les Figures 3.2 et 3.3. Ce jupyter permet d’explorer les fonctionnalités intéressantes de notre approche, comme l’explicabilité de nos résultats.

3.3 Application et description de notre approche sur des communautés jouets

Afin de garantir la bonne compréhension de la méthode, nous avons illustré par la Figure 3.2, les concepts que nous avons décrits. CoCoMiCo permet d’identifier les potentiels de coopération et de compétition au moyen d’un modèle basé sur le raisonnement. Pour y parvenir, l’ensemble des réseaux métaboliques de la communauté ainsi qu’un descriptif du milieu nutritionnel (correspondant aux graines) ont été fournis, afin d’identifier les voies métaboliques activées. La Figure 3.2a illustre les dépendances nécessaires ainsi que le fonctionnement de CoCoMiCo. Toutes les notions métaboliques que sont : graines, produits, réactants, taxon, communauté et environnement ainsi que les relations entre elles, sont montrées dans la sous figure 3.2b.

2. https://github.com/cdanielmachado/kefir_paper

3. [https://gitlab.inria.fr/ccmc/CoCoMiCo/-/blob/dev-0-3/cocomico.ipynb?ref_type=](https://gitlab.inria.fr/ccmc/CoCoMiCo/-/blob/dev-0-3/cocomico.ipynb?ref_type=heads)
heads

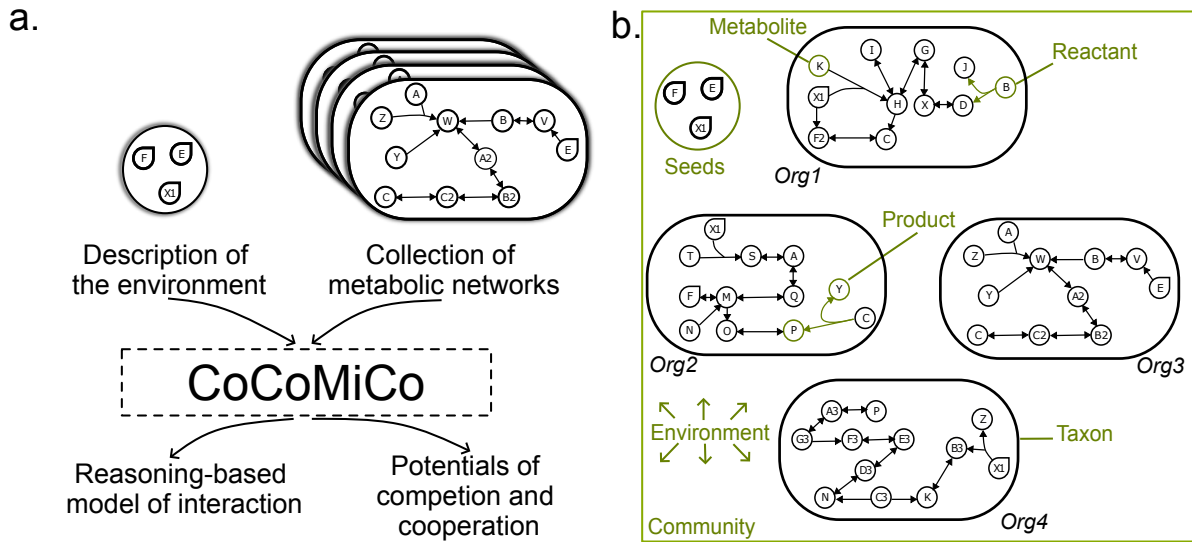


FIGURE 3.2 – **Modèle basé sur le raisonnement et son application sur des données jouets.** (a.) L’approche générale de CoCoMiCo est décrite. En entrée de l’outil, nous avons besoin d’une collection de GEMs décrivant les organismes présents dans la communauté, ainsi que l’ensemble des graines décrivant le milieu nutritionnel. En sortie, nous obtenons le vocabulaire contrôlé issu des règles logiques nous permettant de calculer les scores de coopération et de compétition. (b.) Description de la base de faits générée par l’approche par raisonnement et appliquée à cette communauté jouet de 4 GEMs.

Pour s’exempter de toute confusion avec un algorithme utilisant seulement la topologie des graphes métaboliques pour l’identification de métabolites d’intérêt, *i.e.* échangeables et limitants, les Figures 3.4 et 3.3a,b représentent la modélisation faite au moyen de l’algorithme d’expansion d’Ebenhöh (Ebenhöh et al., 2004). À partir d’un ensemble de graines, que sont X1, F et E, le **scope** métabolique de chaque espèce, *i.e.* l’ensemble de composés productibles, est indiqué en jaune dans les Figures 3.3a et b, montrant par la même occasion, les réactions activées par l’intermédiaire des échanges (en bleu) (Frioux et al., 2018).

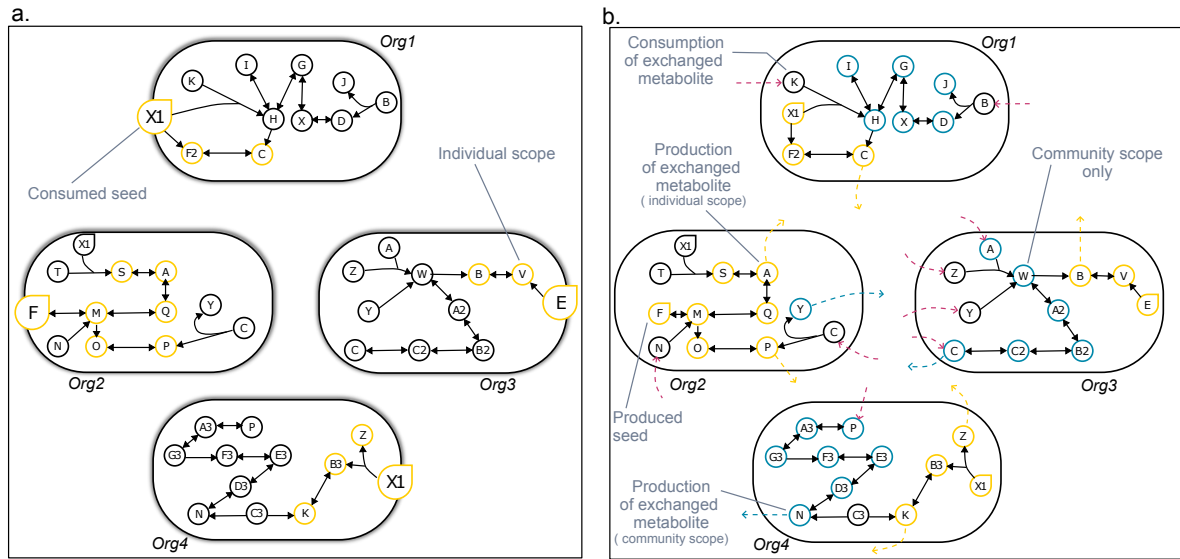


FIGURE 3.3 – **Potentiel métabolique de la communauté** (a.) Potentiel métabolique individuel de chaque GEM de la communauté. Les métabolites en jaune sont les nouveaux métabolites productibles à partir des graines X1, E et F. (b.) Illustration du potentiel métabolique de la communauté. Le potentiel individuel pour chaque GEM est représenté en jaune et le potentiel métabolique de la communauté en bleu, montrant ainsi la plus-value de la communauté. Les flèches sortantes en bleues et jaunes indiquent les exports de composés potentiellement échangeable : en bleu si la molécule est produite uniquement en communauté, en jaune sinon. Les flèches roses indiquent la consommation des métabolites échangés.

Pour cette communauté de taille 4, le **scope** métabolique de chaque taxon, représenté par le listing 3.14, est donc composé de :

```

1 Org4: {X1, Z, B3, K}
2 Org3: {E, V, B}
3 Org2: {F, M, O, P, Q, A, S}
4 Org1: {X1, F2, C}

```

Listing 3.12 – Résultat du scope de chaque taxon

En communauté, représenté par la sous figure 3.3b, le **scope** communautaire, explicité par le listing ?? est composé alors de :

```

1 % scope métabolique communautaire
2 Org4: {N, D3, E3, F3, G3, A3, P}
3 Org3: {C, C2, B2, A2, W, A}
4 Org2: {Y}
5 Org1: {J, D, X, G, H, I}

```

Listing 3.13 – Résultat du scope communautaire. Seules les métabolites productibles en communauté sont montrés.

Le base de connaissance que nous avons décrite nous permet d'identifier le taxon associé à la production du métabolite. À l'échelle de la communauté, le **scope** communautaire, démontre la plus-value métabolique des interactions métaboliques de la communauté. Pour

inférer les scores de coopération et de compétition, CoCoMiCo met en avant les métabolites polyopsonistes, *i.e.* les composés échangés et les graines qui sont co-consommés, et les métabolites monopsonistes, *i.e.* ceux qui sont consommés uniquement par un organisme. Ces concepts sont illustrés dans la Figure 3.4.

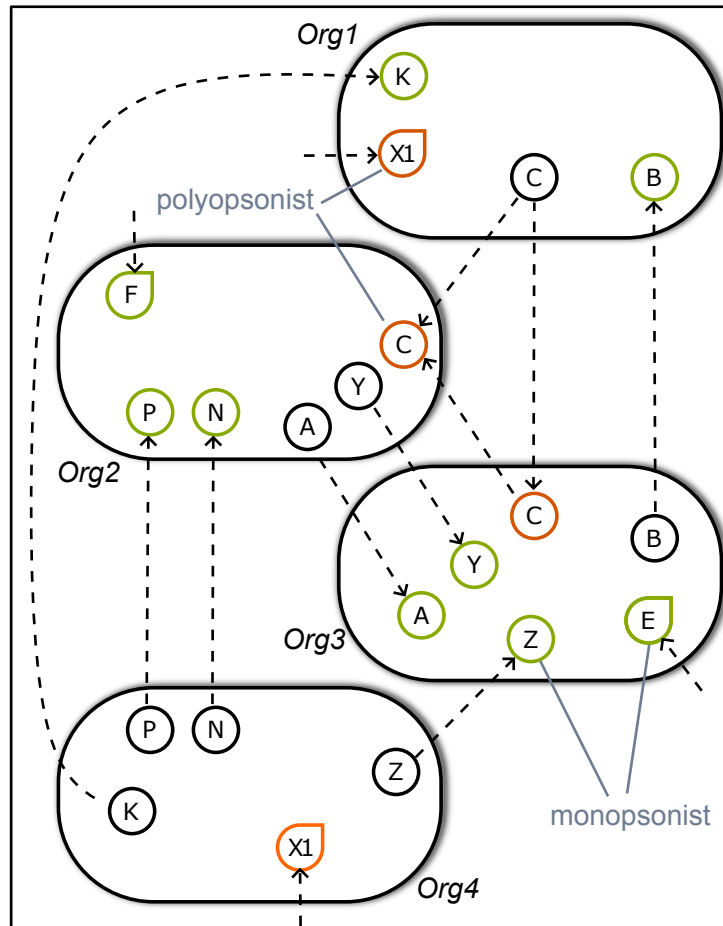


FIGURE 3.4 – **Modèle basé sur le raisonnement et son application sur des données jouets.** Les graines prédites comme consommées et les métabolites échangeables sont illustrés. Les éléments considérés comme polyopsonistes (orange) ont un nombre de GEMs les consommant strictement supérieur à 1. Les métabolites en vert traduisent la monopsonie : un seul GEM les consomme. Pour chaque cas, CoCoMiCo nous informe sur le producteur et le consommateur de chaque métabolite.

Ainsi, les métabolites polyopsonistes et monopsonistes sont :

```

1 % metabolites polyopsonistes et leurs consommateurs
2 X1: {Org1, Org4}
3 C: {Org2, Org3}
4
5 % producteurs des metabolites monopsonistes
6 Org3: {Z, A, Y, E}
7 Org2: {F, P, N}
8 Org1: {K}

```

Listing 3.14 – Métabolites considérés comme limitants (polyopsonie) et non limitants (monopsonie)

En comparaison avec les travaux précédents se basant sur la même approche de raisonnement (Frioux et al., 2018), CoCoMiCo décrit systématiquement les échanges métaboliques ainsi que les substrats limitants afin d’inférer des potentiels de coopération et de compétition à l’échelle de la communauté, que sont respectivement $CooP$ et $ComP$. En plus de cette inférence, deux autres métriques que sont δ et ρ sont calculées montrant respectivement la plus-value des échanges métaboliques sur le nombre de composés productibles et le nombre de réactions activées grâce aux échanges.

Les figures 3.5,3.6 et 3.7 montrent à travers des exemples de sous-communautés issues des Figures 3.4 et 3.3, les valeurs prises par $CooP$, $ComP$, δ et ρ . Dans toutes les communautés, le potentiel métabolique est estimé à partir d’un même milieu de culture (E,F et X1).

Dans les sous figures 3.5a et b, la plus-value de l’organisme 4 est illustrée par les scores calculés par CoCoMiCo. Le scores de coopération est constant, car le nombre de métabolites polyopsonistes et le nombre d’échanges restent identiques. Cependant, les deux métriques δ et ρ informent mécaniquement sur l’impact de l’organisme 4 par rapport à l’organisme 2. En effet, 5 fois plus de métabolites sont devenus productibles et 10 fois plus de réactions ont été nouvellement activées avec l’organisme 4.

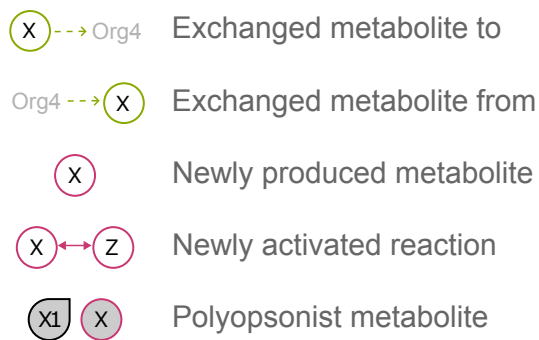
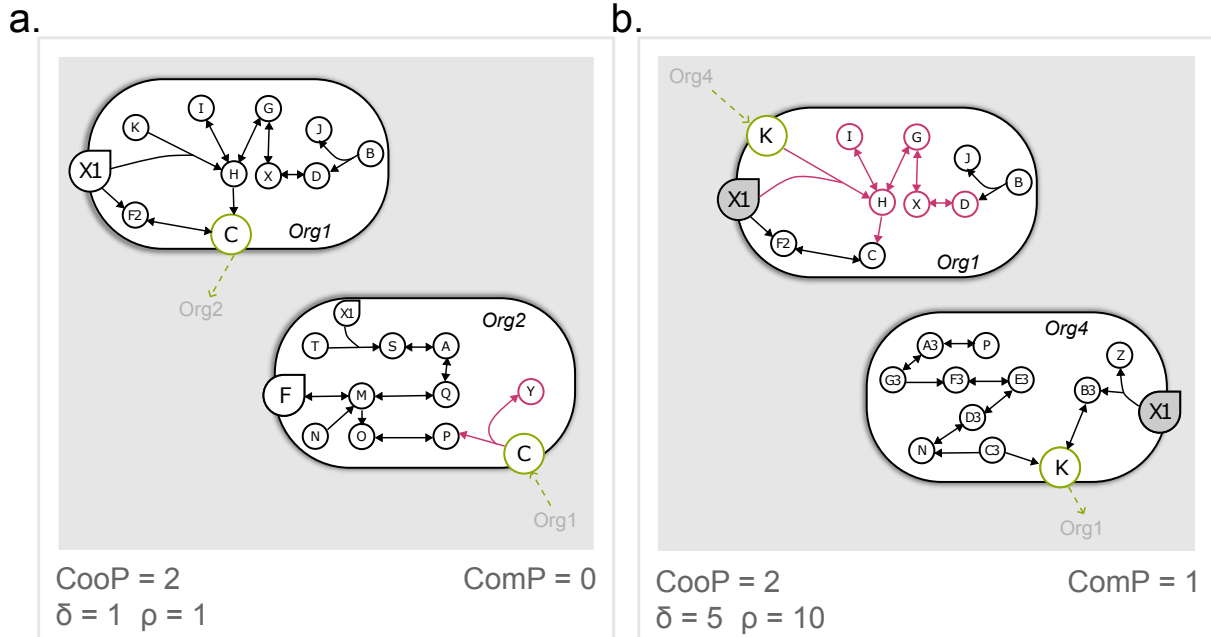


FIGURE 3.5 – **Application de CoCoMiCo sur les exemples jouets (a et b).** L'ensemble des concepts expliqués en amont sont représentés sur cette communauté de taille 2. Les scores de coopération et de compétition, respectivement $CooP$ et $ComP$, ainsi que les plus-values métaboliques et réactionnelles, respectivement $\bar{\delta}$ et ρ sont également représentées. Les composés en vert sont les composés échangés, en rose les composés produits grâce aux échanges.

Dans la Figure 3.6, la plus-value de l'organisme 4 est également montrée dans une communauté de taille 3, dans laquelle, les organismes 1 et 2 sont présents dans les deux cas. Le nombre de substrats polyopsonistes étant identique (2), le score de compétition est constant, cependant la communauté de la sous figure 3.6b semble moins coopérative (score de coopération passant de 9 à 8) alors que le nombre de composés échangés reste le même dans les deux cas : 4. Les sous métriques ρ et $\bar{\delta}$ sont proches. Nous observons que ces deux communautés de taille 3 ont un potentiel de compétition plus faible que les communautés de taille 2 de la Figure 3.5, malgré le nombre d'échanges plus important.

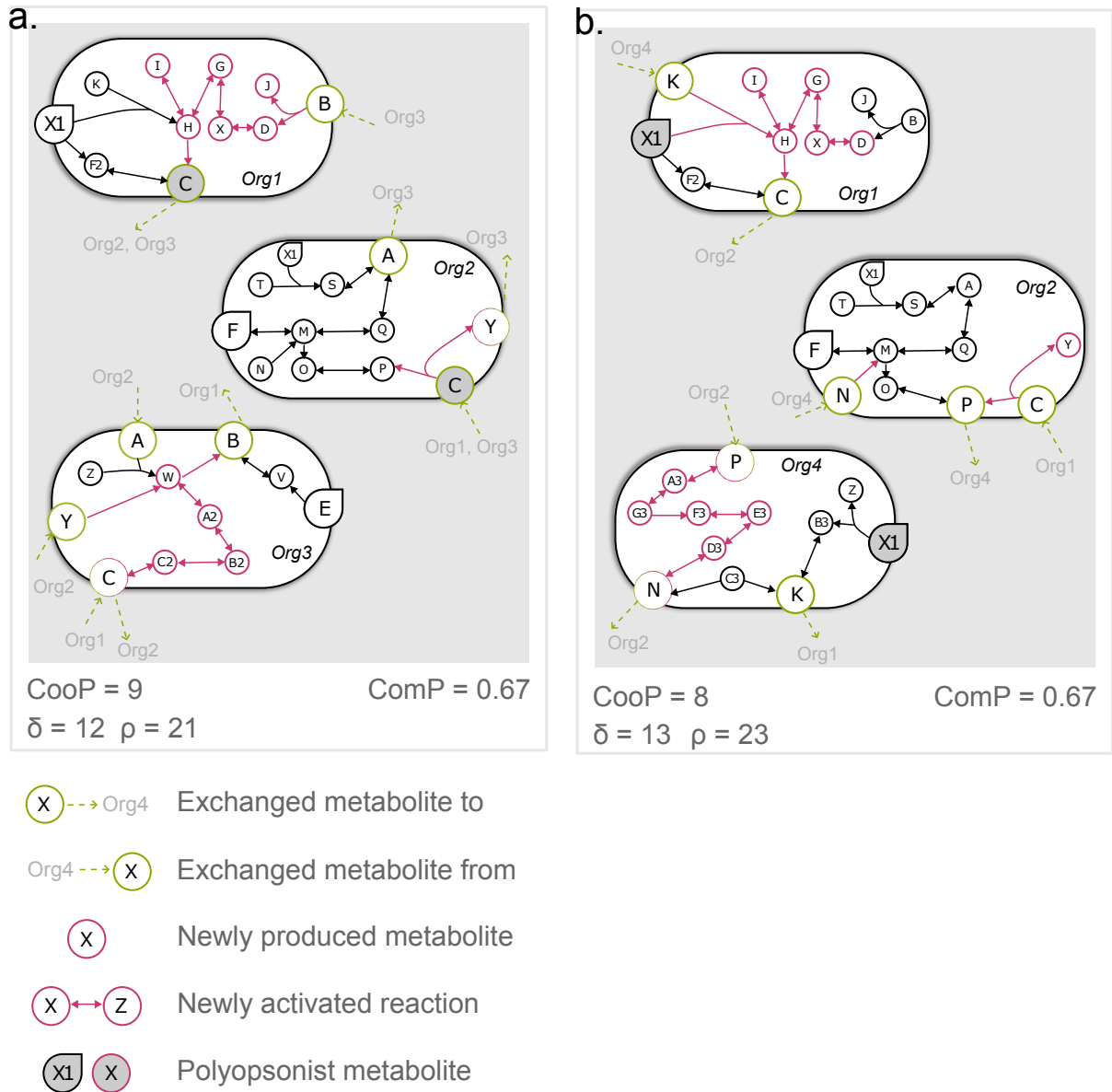


FIGURE 3.6 – **Application de CoCoMiCo sur les exemples jouets (a et b).** L'ensemble des concepts expliqués en amont sont représentés sur cette communauté de taille 3. Les scores de coopération et de compétition, respectivement $CooP$ et $ComP$, ainsi que les plus-values métaboliques et réactionnelles, respectivement δ et ρ sont également représentés. Les composés en vert sont les composés échangés, en rose les composés produits grâce aux échanges.

Enfin, la Figure 3.7 montre les scores finaux obtenus pour la communauté que nous avons décrites dans les Figures 3.3 et 3.4. Nous pouvons identifier que les organismes 3 et 4 doublent le score de coopération ($cooP = 17$) sans rentrer en compétition ($comP = 1$).

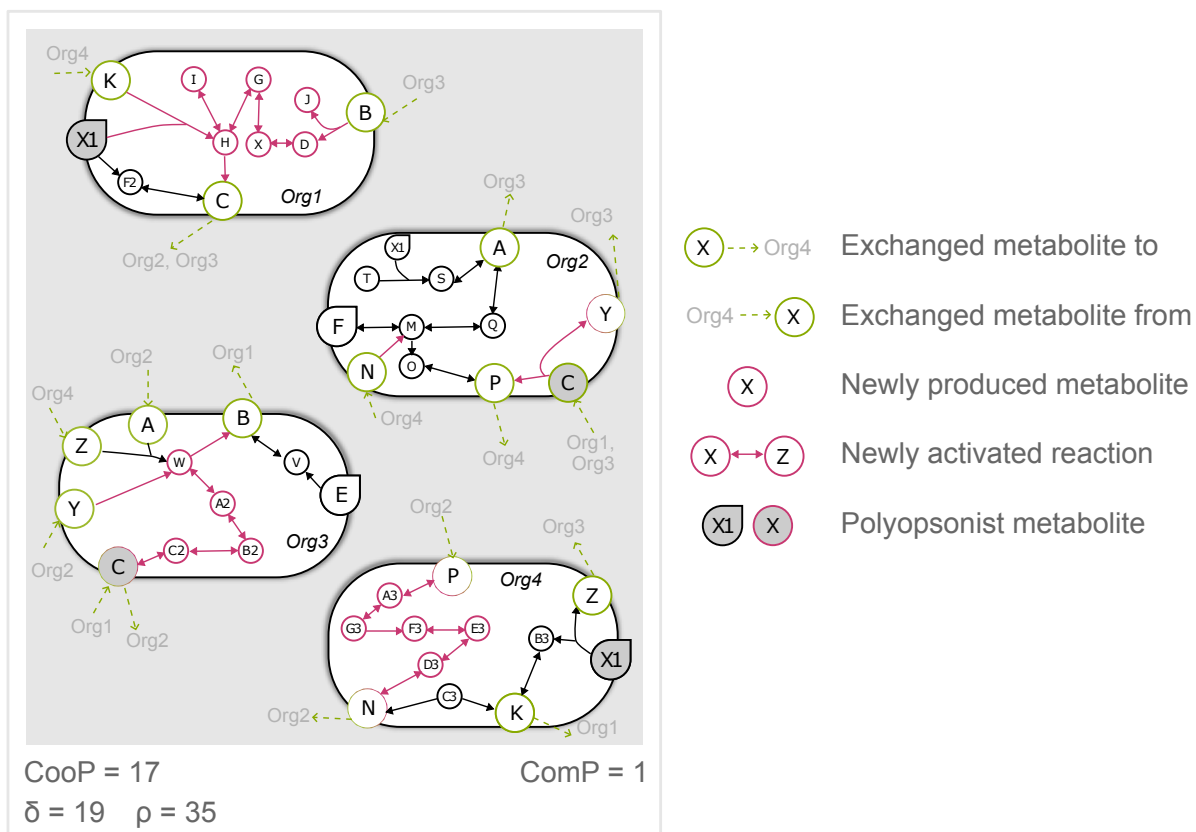


FIGURE 3.7 – **Application de CoCoMiCo sur un exemple jouet** L'ensemble des concepts expliqués en amont sont représentés sur cette communauté de taille 4. Les scores de coopération et de compétition, respectivement $CooP$ et $ComP$, ainsi que les plus-values métaboliques et réactionnelles, respectivement δ et ρ sont également représentées. Les composés en verts sont les composés échangés, ceux en rose les composés produits grâce aux échanges.

3.4 Application à des communautés synthétiques

Nous venons de décrire le fonctionnement de CoCoMiCo sur des exemples jouets. Des informations plus complètes sont disponibles dans le jupyter notebook ⁴.

3.4.1 Description des potentiels de coopération et de compétition dans différents écosystèmes

Afin de tester nos scores, nous avons appliqué CoCoMiCo sur 50 communautés de taille différentes pour 4 écosystèmes : l'intestin, la feuille, la racine, et le sol. Pour rappel, nous avons également fait les simulations avec deux conditions nutritionnelles : un ensemble de graines générique commun aux quatre environnements, et des simulations avec un milieu spécifique pour chaque écosystème. Dans les paragraphes suivants, nous étudierons le comportement des scores de compétition et de coopération dans les 4 écosystèmes.

4. https://gitlab.inria.fr/ccmc/CoCoMiCo/-/blob/dev-0-3/cocomico.ipynb?ref_type=heads

Application de CoCoMiCo à partir d'un milieu de culture générique

Une première analyse avec un milieu nutritionnel générique est faite afin de comparer les résultats entre eux, puis une seconde avec un milieu nutritionnel propre à chaque écosystème sera présentée, afin d'identifier l'impact d'un changement de milieu sur les scores d'interaction.

Lorsque l'on crée un modèle par raisonnement se basant sur la topologie d'un objet, ici le réseau métabolique, nous devons vérifier si les résultats que l'on obtient ne sont pas directement reliés à cette topologie. Pour cela, nous avons vérifié qu'il n'existe pas de corrélation entre la similarité des réseaux métaboliques et les potentiels de coopération et de compétition.

Similarité des réseaux métaboliques et potentiels d'interaction Notre approche par le raisonnement est une méthode d'inférence qualitative ne permettant pas de calculer les flux métaboliques, et se repose donc sur la structure du modèle métabolique. Nous avons testé la similarité des réseaux en terme de composition en réactions métaboliques calculée par l'indice de Jaccard (voir méthodes) et testé la corrélation de cet indice avec les potentiels de coopération et de compétition pour chaque écosystème. La figure 3.8 et la Table 3.1 montrent qu'il n'y a pas de corrélation entre la topologie des réseaux métaboliques et les potentiels d'interaction obtenus. Ainsi, l'utilisation d'un formalisme discret prenant en compte les conditions environnementales, apporte plus d'information que la seule structure des réseaux pour analyser et caractériser des communautés bactériennes.

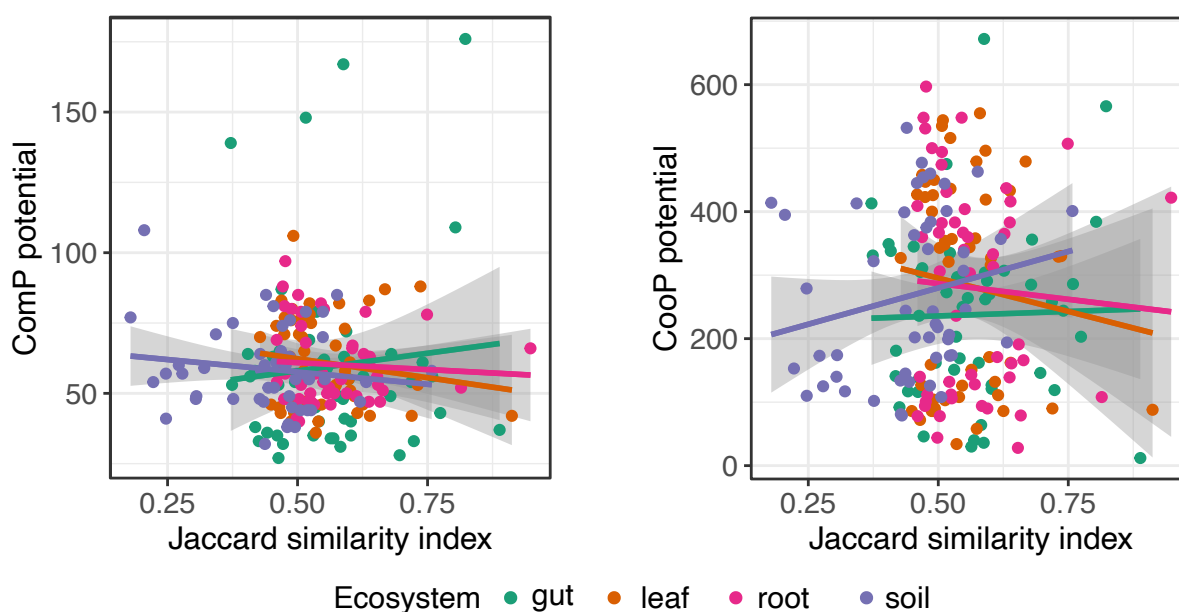


FIGURE 3.8 – Scores de coopération et de compétition calculés pour 50 communautés de taille 2 de chaque écosystème. Pour chaque paire, la similarité des réseaux métaboliques a été calculée par l'indice de Jaccard. Les tests statistiques sont montrés dans la Table 3.1

Description globale des potentiels Les scores de coopération et de compétition calculés pour chaque écosystème (intestin, sol, racine et feuille) sont représentés par la Figure

ecosystem	n	Competition			Cooperation		
		S	rho estimate	p. value	S	rho estimate	p. value
gut	50	21889.05	-0.051	0.724	21128.12	-0.014	0.920
leaf	50	21483.21	-0.031	0.827	23141.56	-0.111	0.441
root	50	22515.28	-0.081	0.575	23113.31	-0.109	0.447
soil	50	15744.51	0.243	0.087	22505.37	-0.080	0.577

TABLE 3.1 – **Résultat des tests de corrélation (Spearman) entre l'indice de similarité de Jaccard et les scores de potentiels d'interaction**

3.9. La sous figure 3.9a montre que les scores varient à la fois selon l'écosystème et à la fois selon la taille de la communauté. En effet, la taille de la communauté est fortement corrélée aux scores d'interaction calculés par CoCoMiCo : (ComP : Spearman $\rho = 0.74$, $P < 2,2e-16$, CoopP : Spearman $\rho = 0.80$, $P < 2,2e-16$). La forte association avec la coopération n'est pas surprenante, puisque le nombre de voies métaboliques augmente avec l'ajout de nouveaux membres dans la communauté, ainsi que le nombre de métabolites échangeables. Comme la compétition se repose en partie sur le nombre de métabolites échangés, l'augmentation du potentiel de compétition lorsque que la taille de la communauté augmente est attendue.

Nous avons ensuite testé si, indépendamment de la taille de la communauté, les potentiels de coopération et de compétition étaient discernables entre écosystèmes. Les résultats, illustrés par la sous-figure 3.9b, montrent que le potentiel de compétition distingue plus difficilement les communautés de la feuille, de la racine et du sol entre elles. Au contraire, le potentiel de coopération, différencie bien la racine et la feuille du sol mais peine à différencier la racine de la feuille.

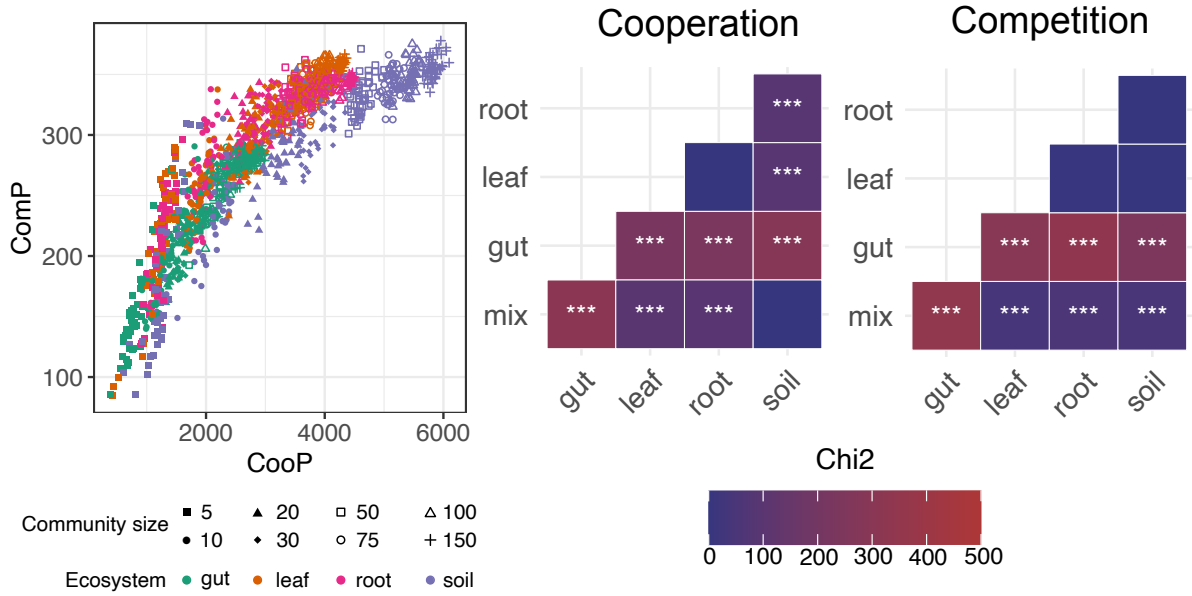


FIGURE 3.9 – (a) Distribution des scores de coopération et de compétition dans des communautés simulées de 5 à 150 bactéries isolées de l’intestin, de la racine, de la feuille et du sol. (b) Résultats d’un test du χ^2 de Kruskal-Wallis comparant les potentiels de coopération et de compétition obtenus en (a) entre écosystèmes. (*) P -values ajustées (Benjamini-Hochberg) : $P < 0.05$, (**) P -values ajustées (Benjamini-Hochberg) : $P < 0.01$, (***) P -values ajustées (Benjamini-Hochberg) : $P < 0.001$. La communauté *mixte* est un mélange composé de réseaux métaboliques associés à des génomes de la racine, de la feuille, du sol et de l’intestin en proportions équitables.

On observe que les communautés de l’intestin présentent des scores de compétition et de coopération moins élevés que les autres écosystèmes. Cette observation est explicable par le nombre de métabolites échangés comme indiqué dans la Table 3.2 dans laquelle le sol présente en moyenne deux fois plus de composés échangés que les communautés de l’intestin. Nous avons également remarqué que le nombre de métabolites échangés pour la feuille et la racine sont proches. Nous avons supposé que ces résultats puissent être reliés à la diversité taxonomique des GEMs utilisés dans la création des communautés synthétiques (Figure 1). Les communautés du sol et de l’intestin sont les plus diverses et présentent respectivement les plus grandes et les plus petites valeurs de compétitions, suggérant que la diversité taxonomique ne semble pas être le seul facteur déterminant les potentiels d’interaction. En revanche, la similitude des scores d’interaction pour les écosystèmes de la feuille et de la racine, est liée à leur grande proximité de diversité taxonomique.

	sol	intestin	racine	feuille	mixte
métabolites échangeables	1516 \pm 27	709 \pm 45	1171 \pm 64	1105 \pm 11	1525 \pm 53
CooP	5833 \pm 102	2739 \pm 175	4361 \pm 137	4273 \pm 40	5877 \pm 189
ComP	355 \pm 9	276 \pm 11	345 \pm 3	357 \pm 4	384 \pm 12

TABLE 3.2 – Moyenne des scores d’interactions ainsi que du nombre de métabolites échangeables pour tous les écosystèmes et pour une taille de communauté de 150.

Afin de vérifier statistiquement nos observations, nous avons comparé les scores de compétition et de coopération pour chaque taille et par écosystème (Figure 3.10)

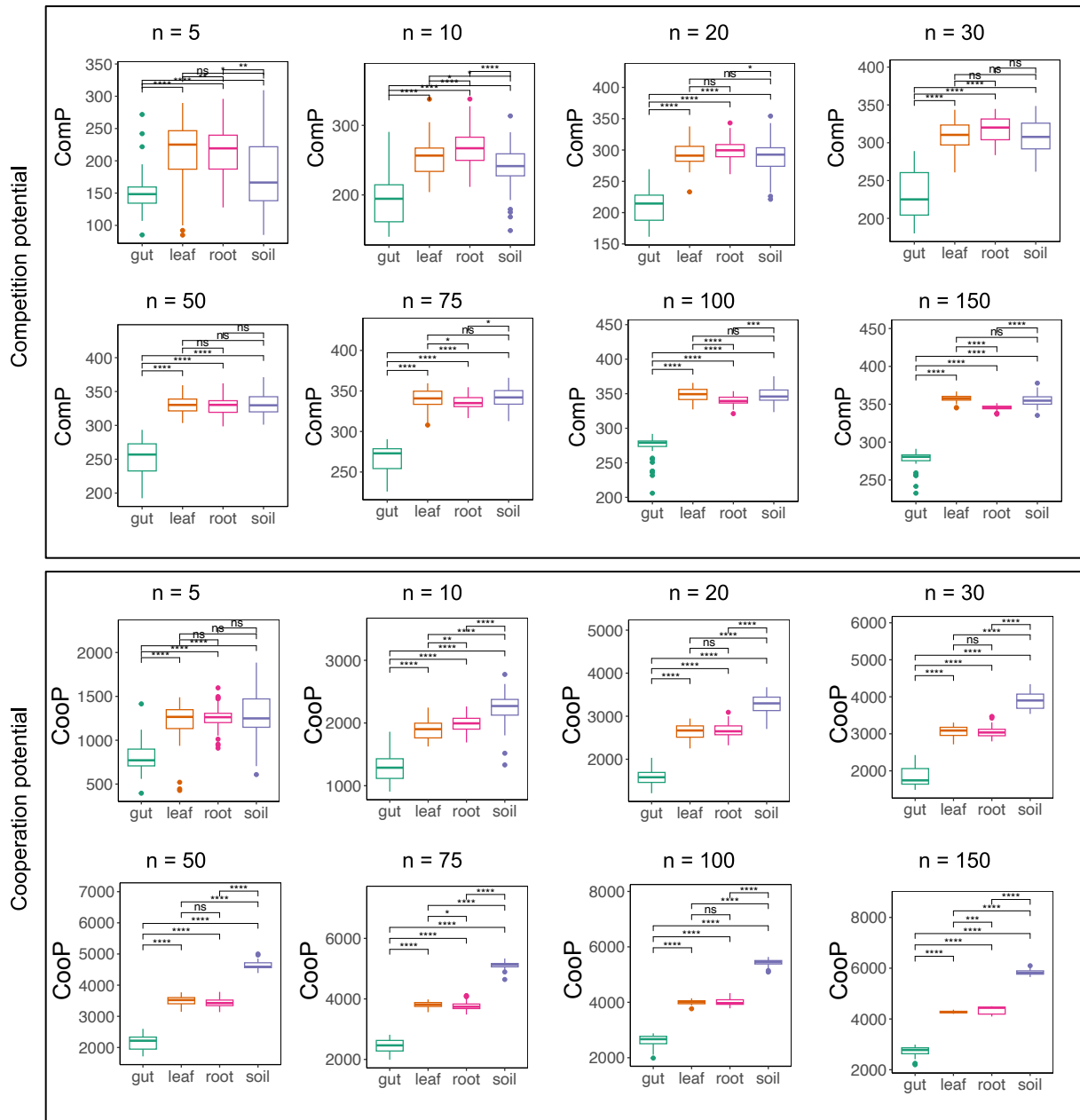


FIGURE 3.10 – Distribution des scores de coopération et de compétition dans des communautés simulées de 5 à 150 bactéries isolées de l'intestin, de la racine, de la feuille et du sol. Une comparaison de moyenne avec le test de Wilcoxon a été faite. (*) P -values ajustées (Benjamini-Hochberg) : $P < 0.05$, (**) P -values ajustées (Benjamini-Hochberg) : $P < 0.01$, (***) P -values ajustées (Benjamini-Hochberg) : $P < 0.001$.

Nous observons que le potentiel de compétition distingue globalement moins bien les communautés de petites tailles (entre 5 et 30), surtout entre les écosystèmes de la racine, de la feuille et du sol, ce qui est cohérent avec les observations décrites plus haut. Le score de coopération semble être plus apte à discerner des communautés de plus de dix taxons, sauf entre les écosystèmes de la feuille et de la racine. Les communautés de grandes tailles

étant plus réalistes, ces dernières sont distinguées par nos potentiels d'interactions, apportant de la confiance à nos métriques.

Un point nouveau, et peu documenté dans la littérature était la mesure de l'impact d'une modification de la composition bactérienne sur les scores de coopération et de compétition. Notre approche ne pouvant prendre en compte des abondances, nous avons modélisé l'impact d'un changement discret en ajoutant ou en enlevant une population bactériennes (i.e un réseau métabolique) dans chacune des communautés de tailles 3 à 150 de l'intestin (voir Figure 4). Par rapport à la communauté originale, les scores de coopération et de compétition ne permettent plus de distinguer la plus-value d'ajouter ou de retirer une espèce à partir d'une communauté de taille 10. Cela suggère donc que l'effet d'une modification d'un facteur biotique peut-être capturé par CoCoMiCo uniquement pour des écosystèmes de petites tailles. Les communautés de tailles réelles ne verront pas leurs potentiels d'interactions changés significativement.

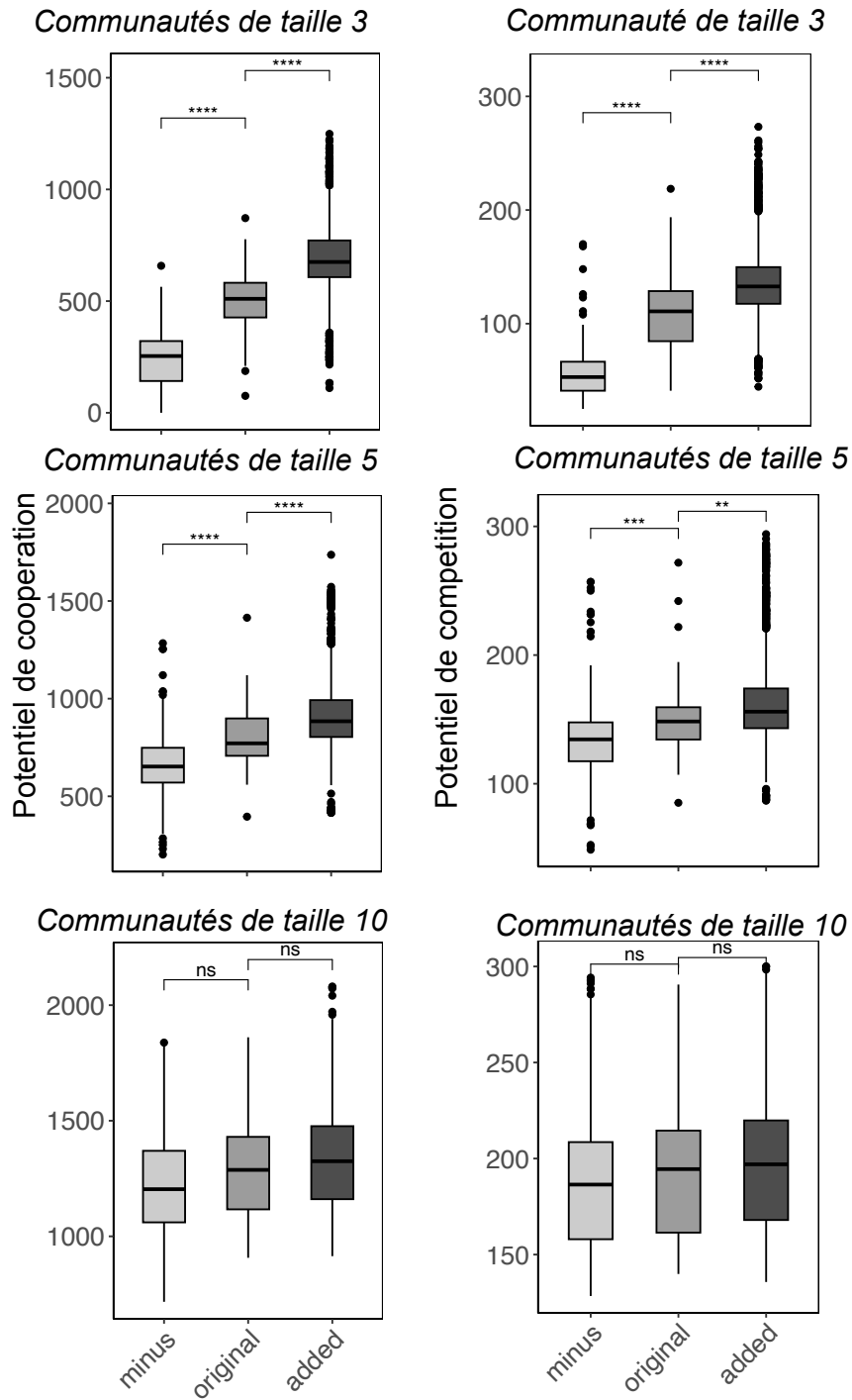


FIGURE 3.11 – Effet d'ajouter ("added") ou de retirer une espèce ("minus") sur les potentiels de coopération et de compétition pour chaque communauté. Le test de comparaison de moyenne Wilcoxon a été utilisé sur des communautés de taille 3 à 10 sur l'écosystème de l'intestin.

Comparaison des potentiels d'interaction propres aux écosystèmes avec ceux d'un écosystème non-réaliste. Nous avons en plus créé une cinquième catégorie de communauté synthétique composée d'une combinaison des réseaux métaboliques appartenant aux 4 autres écosystèmes, dénotée "mixte". Cette communauté non réaliste permet de vérifier que les potentiels de coopération et de compétition sont spécifiques aux écosys-

tèmes, et pourraient permettre de discerner des communautés non plausibles. Nous avons remarqué que le potentiel de compétition a permis de distinguer l'écosystème mixte des autres, ce qui n'est pas le cas pour le score de coopération où le sol et un écosystème factice seraient similaires du point de vue de ce potentiel d'interaction (voir Figures 3.9 et 3.10). Nous retrouvons également ce résultat dans la Table 3.2 où les valeurs de coopération, de compétition et du nombre de métabolites échangés sont particulièrement proches de l'écosystème du sol. Lorsque l'on compare par taille de communauté (voir Figure 3) les deux potentiels de coopération et de compétition arrivent à différencier l'écosystème factice. En revanche, comme illustré dans la Figure 3.9, l'écosystème du sol et le mixte semblent être identiques du point de vue de la coopération.

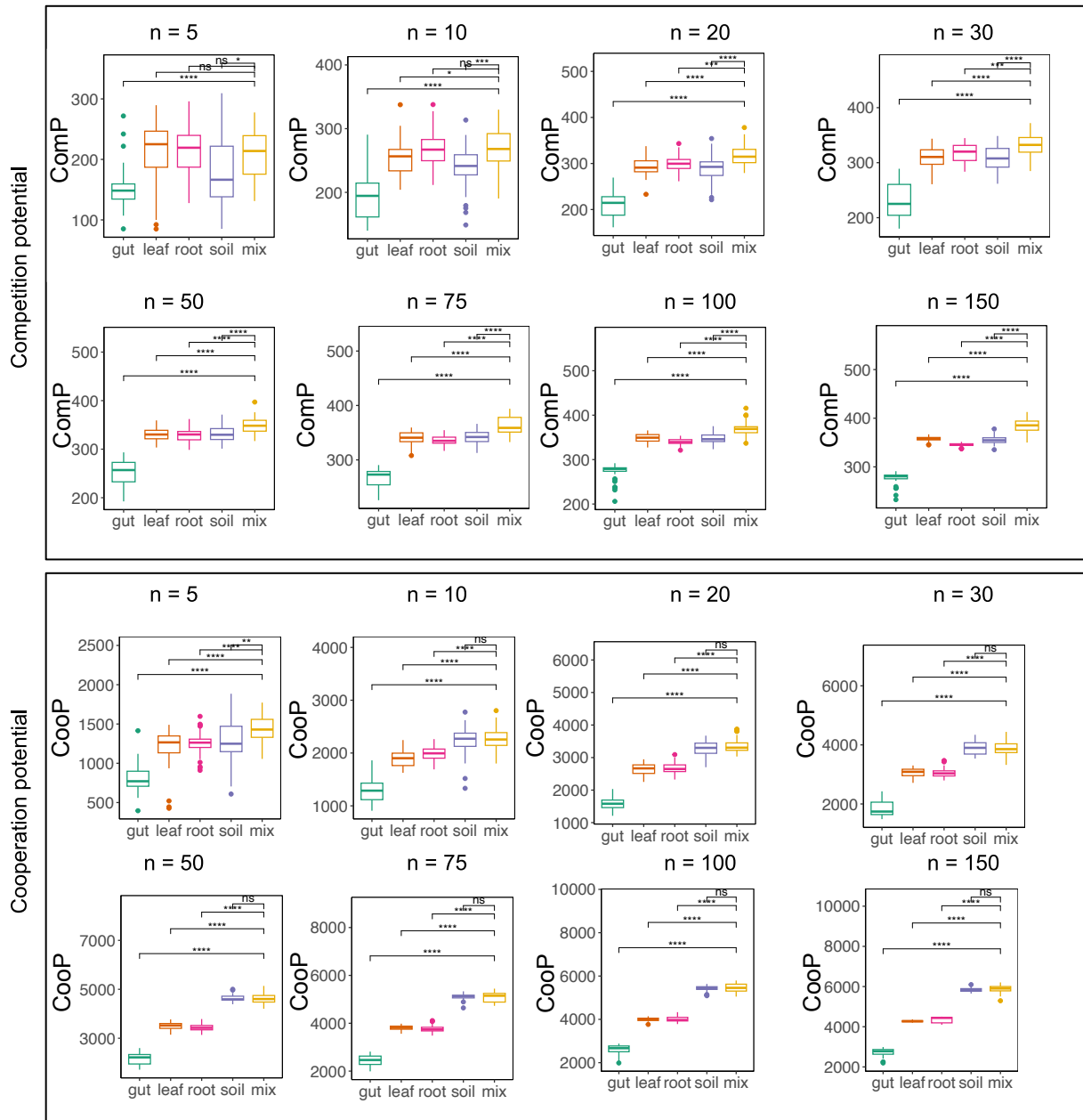


FIGURE 3.12 – Distribution des scores de coopération et de compétition dans des communautés simulées de 5 à 150 bactéries isolées de l’intestin, de la racine, de la feuille, sol et du mixte. Seule une comparaison deux à deux avec le mixte a été faite. La comparaison de moyenne est faite avec le test de Wilcoxon. (*) P -values ajustées (Benjamini-Hochberg) : $P < 0.05$, (**) P -values ajustées (Benjamini-Hochberg) : $P < 0.01$, (***) P -values ajustées (Benjamini-Hochberg) : $P < 0.001$

Une étape terminale d’exploitation des potentiels d’interaction serait de pouvoir prédire des écosystèmes à partir d’un score de coopération et de compétition. Les résultats de ce test sont présentés dans le paragraphe suivant.

Prédiction d’écosystème à partir de potentiel d’interaction Nous avons montré que les potentiels d’interaction semblent être en général spécifiques à des écosystèmes. Une hypothèse sous-jacente consiste à se demander si, à partir des potentiels d’interaction de

coopération et de compétition de chaque écosystème, il est possible de prédire l'écosystème correspondant. Afin de tester et vérifier cette hypothèse, nous avons utilisé un SVM (de l'*anglais*, Support Vector Machine) non linéaire (voir méthodes). Les résultats sont présentés dans la Table 3.3.

	sol	intestin	racine	feuille
CooP	0.67 ± 0.34	0.74 ± 0.32	0.48 ± 0.34	0.53 ± 0.35
ComP	0.40 ± 0.31	0.93 ± 0.12	0.44 ± 0.29	0.42 ± 0.23
ComP et CooP	0.71 ± 0.29	0.74 ± 0.32	0.49 ± 0.35	0.5 ± 0.39

TABLE 3.3 – **Résultat de prédiction des écosystèmes de la racine, de la feuille, du sol et de l'intestin, à partir des graines génériques.** Les colonnes correspondent aux écosystèmes que l'on souhaite prédire à partir d'uniquement le score de coopération (CooP), d'uniquement le score de compétition (Comp) ou une combinaison des deux (ComP et CooP).

Nous observons qu'à partir des graines génériques, seul l'intestin semble pouvoir être prédit correctement par les 3 combinaisons de scores. Le score de compétition semble plus se démarquer avec une moyenne de 0.93. Ce chiffre peut être interprété comme une probabilité d'obtenir le bon résultat. Ainsi, un score de 0.93 signifie que nous avons 93% de chance de prédire l'écosystème de l'intestin à partir du score de compétition. L'écosystème du sol semble être uniquement prédictible avec une combinaison du score de coopération et de compétition. En revanche, prédire les écosystèmes de la feuille et de la racine ne semble pas être faisable, ce qui est explicable par la proximité taxonomique de ces deux écosystèmes générant des potentiels métaboliques proches et des scores d'interaction similaires.

Application de CoCoMiCo à partir d'un milieu de culture spécifique

Nous avons utilisé un milieu de culture générique pour toutes les communautés pour comparer les potentiels d'interaction entre eux. Afin d'identifier une possible gamme des potentiels de coopération et de compétition propre à chaque écosystème, un milieu de culture spécifique de l'intestin et un milieu commun aux 3 autres écosystèmes ont été utilisés.

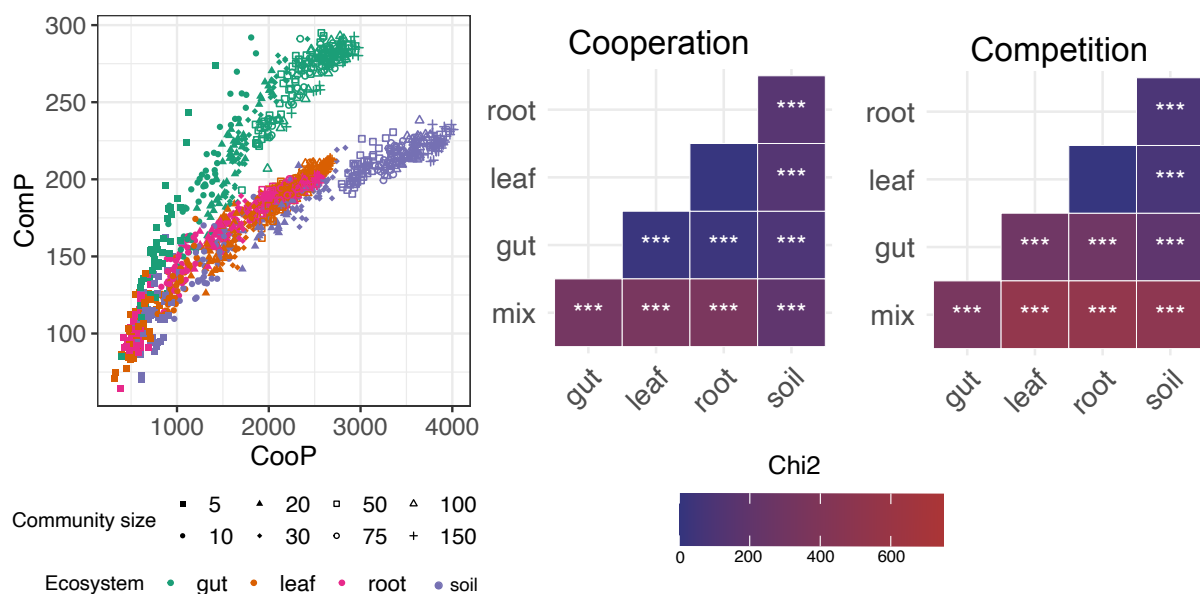


FIGURE 3.13 – (a) Distribution des scores de coopération et de compétition dans des communautés simulées de 5 à 150 bactéries isolées de l’intestin, de la racine, de la feuille et du sol. (b) Résultats d’un test du χ^2 de Kruskal-Wallis comparant les potentiels de coopération et de compétition obtenus en (a) avec ceux de la communauté mixte.

Il existe des différences notables entre la Figure 3.13, utilisant environnement nutritionnel spécifique et la Figure 3.9, utilisant un milieu générique. L’intestin semble avoir des scores de compétitions plus importants, et la valeur maximale de coopération est réduite à 4000 contre 6000 dans la Figure 3.9. Lorsque l’on compare les potentiels de coopération et de compétition sans tenir compte de la taille de la communauté, les deux potentiels distinguent presque tous les écosystèmes. Comme précédemment, la racine et la feuille ne sont pas significativement différents. Malgré cette amélioration, les potentiels d’interaction ne permettent toujours pas de prédire correctement des écosystèmes (voir Table 3.4)

	sol	intestin	racine	feuille
CooP	0.90 ± 0.15	0.48 ± 0.23	0.56 ± 0.37	0.47 ± 0.34
ComP	0.66 ± 0.25	0.96 ± 0.04	0.63 ± 0.39	0.49 ± 0.33
ComP et CooP	0.79 ± 0.29	0.87 ± 0.15	0.59 ± 0.40	0.44 ± 0.35

TABLE 3.4 – **Résultat de prédiction des écosystèmes de la racine, de la feuille, du sol et de l’intestin, à partir de graines spécifiques.** Les colonnes correspondent aux écosystèmes que l’on souhaite prédire à partir d’uniquement le score de coopération (CooP), d’uniquement le score de compétition (Comp) ou une combinaison des deux (ComP et CooP).

Le résultat de la Table 3.4 indique qu’il n’est toujours pas possible de déterminer le potentiel d’interaction permettant de prédire un écosystème. Cependant, la combinaison des deux scores ainsi que l’utilisation de graines spécifiques, semblent permettre la prédiction de l’intestin et du sol avec plus de confiance.

3.4.2 Comparaison avec des méthodes basées sur des contraintes et application à des données expérimentales

Les résultats précédents ont montré une application *in silico* des potentiels d'interaction. Dans la suite de ce chapitre, nous allons présenter deux cas d'étude pour comparer CoCoMiCo et d'autres outils de l'état de l'art. Dans le premier cas, nous avons calculé les potentiels d'interaction avec CoCoMiCo sur 250 communautés créées artificiellement, de taille 3 à 30 et comparé avec les scores obtenus par SMETANA. Puis, nous avons calculé les potentiels de coopération et compétition avec CoCoMiCo sur 188 communautés issues d'une analyse métagénomique, allant d'une taille 16 à 98 et comparé ces potentiels avec ceux calculés par MICOM. Dans le second cas d'étude, nous avons comparé les résultats de coopération et de compétition obtenus par MICOM, SMETANA et CoCoMiCo sur 2 communautés du kéfir provenant de (Blasche et al., 2021).

Comparaison de méthodes numériques dans le calcul de score d'interaction

Dans cette section, nous présenterons les résultats de la première étude de cas. L'objectif est de vérifier, pour une même communauté, si nos scores sont corrélés avec ceux d'outils reconnus de la littérature. La première remarque est que notre score de coopération présente une bonne corrélation avec les scores des méthodes numériques avec une valeur de $\rho = 0.8$ et une p-value de 2.2^{-16} pour SMETANA et $\rho = 0.69$ avec une p-value de 2.2^{-16} pour MICOM (Figure 3.14).

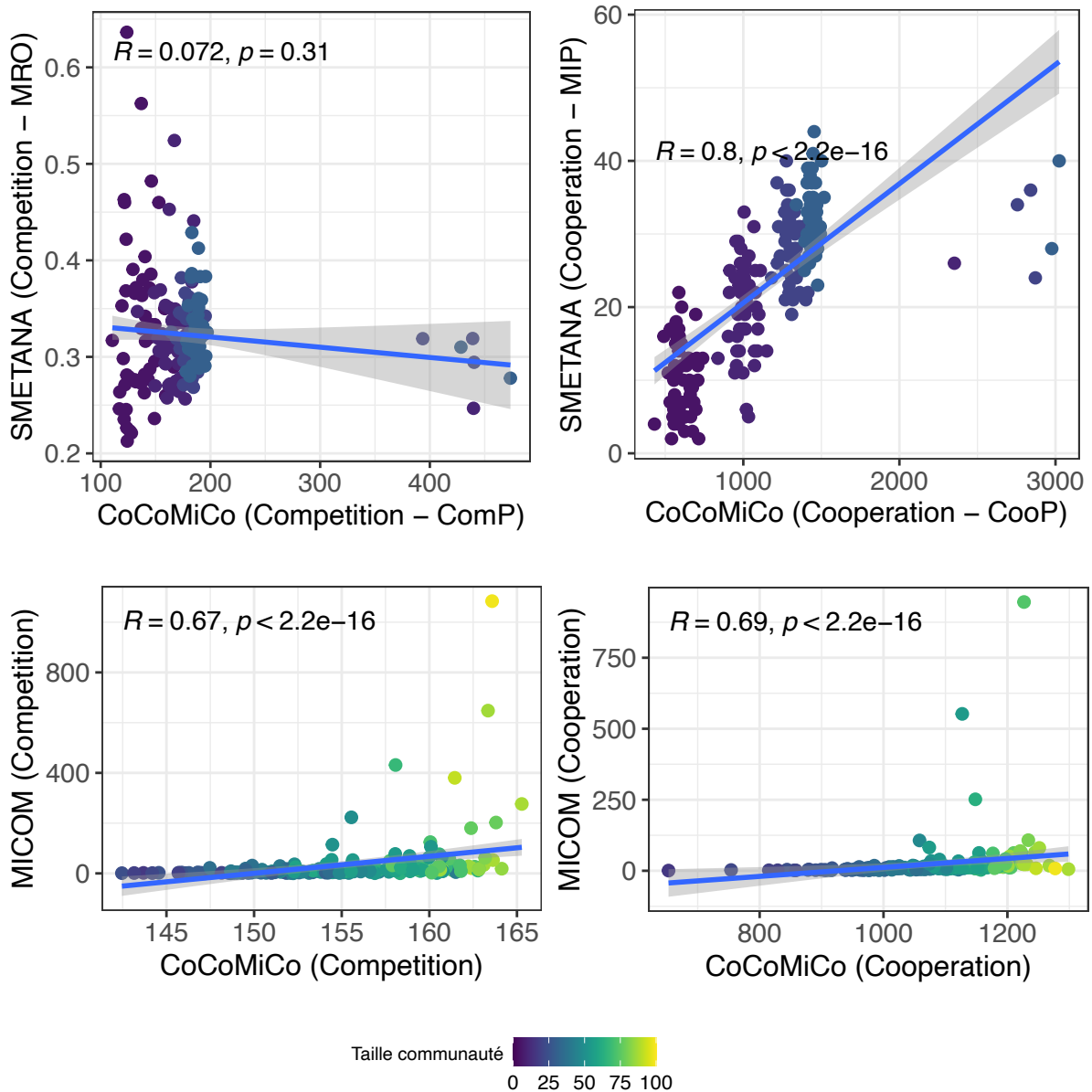


FIGURE 3.14 – **Comparaison des approches numériques MICOM et SMETANA avec notre approche.** (a) Comparaison des scores de coopération et de compétition développés par SMETANA, respectivement MIP et MRO. 250 communautés de taille 3 à 30 ont été générées aléatoirement. (b) Comparaison des scores de coopération et de compétition *ad hoc* de MICOM sur 188 communautés issues d’analyses métagénomiques de taille 16 à 98 GEMs. En (a) et en (b), les corrélations ont été calculées avec la corrélation de Spearman.

La compétition en revanche, est seulement corrélée avec le score MICOM avec un ρ de 0.67 et une p-value de $2.2 \cdot 10^{-16}$. L’absence de corrélation avec le score SMETANA (ρ de 0.072 avec une p-value de 0.31) peut être expliquée par la prise en compte d’un milieu de culture théorique minimum dans cet outil indiquant qu’il ne prend pas en considération l’ensemble des éléments nutritifs comme dans CoCoMiCo et MICOM.

Un point non négligeable à notifier est le temps de calcul des approches numériques.

SMETANA ne pouvait être utilisé sur des communautés de taille supérieure à 30 pour une allocation mémoire de 10Go. MICOM a permis en revanche l’analyse de grandes communautés (98) mais nécessite plusieurs heures par communauté avec 20Go de mémoire. En comparaison, CoCoMiCo a mis environ 3 minutes pour analyser une communauté de taille 150 avec 10Go de mémoire.

Application sur des données expérimentales de communauté microbienne du kéfir

Nous avons appliqué CoCoMiCo ainsi que les deux approches numériques MICOM et SMETANA sur 2 communautés du kéfir de tailles différentes, 23 et 31. Lors de l’analyse des communautés sur milieu lait, les auteurs et autrices montrent un réseau d’interaction plus dense et globalement négatif pour la plus grande communauté, cultivé sur du gélose d’agarose (noté lait-solide). L’ensemble des résultats est présenté dans la Table 3.5. Seul CoCoMiCo a pu être appliqué aux réseaux métaboliques reconstruits avec Pathways-Tools et dépourvus d’une fonction de biomasse. Cette dernière est nécessaire pour l’application d’une méthode numérique.

	Communauté	Pathway Tools		CarveMe		gapseq	
		Coopération	Compétition	Coopération	Compétition	Coopération	Compétition
CoCoMiCo	lait-liquide	1090.68 (287)	123.26	2183.06 (426)	149.65	2039.34 (520)	233.74
	lait-solide	1098.85 (278)	124.5	2651.19 (497)	152.74	2039.39 (499)	243.53
SMETANA	lait-liquide	-	-	22 (59)	0.329	NA	0.682
	lait-solide	-	-	22 (66)	0.348	NA	0.693
MICOM	lait-liquide	-	-	24.75 (99)	2.75	22.435 (95)	22.432
	lait-solide	-	-	34.80 (109)	3.8	31.476 (103)	31.454

TABLE 3.5 – **CoCoMiCo, Smetana et MICOM sont appliqués sur les données expérimentales du kéfir (Blasche et al., 2021)** Les potentiels de coopération et de compétition sont obtenus pour les deux communautés cultivées sur lait-liquide (n=23) ou lait-solide (n=31) à partir des réseaux métaboliques reconstruits avec Pathway Tools, CarveMe et gapseq, et analysés par MICOM, SMETANA et CoCoMiCo. Parce qu’il n’utilise pas de fonction objective, seul CoCoMiCo a pu calculer des potentiels d’interaction sur les réseaux métaboliques reconstruits par Pathway-Tool. SMETANA n’a pas pu déterminer certains scores de coopération, ce qui indique qu’au moins une bactérie ne pouvait croître dans dans le modèle de communauté.

Globalement, les scores de potentiels de coopération et de compétition sont tous plus importants dans la plus grande communauté, facteur sûrement déterminé par la grande corrélation des scores avec la taille de la communauté (cf Figure 3.9). Le nombre de métabolites échangés est moindre pour la plus petite communauté et pour les réseaux reconstruits avec Pathway-Tools et gapseq. Le réseau d’interaction positive 2 à 2 étant plus dense dans cette communauté, le nombre d’échanges métaboliques est attendu comme plus important. Les réseaux reconstruits avec CarveMe montrent l’inverse suggérant donc que la méthode de reconstruction peut impacter des propriétés intrinsèques à la communauté (ici, le nombre d’échanges). Enfin, dernier point, notre méthode basée sur le raisonnement

décrit un nombre de métabolites échangés nettement plus important que pour MICOM et SMETANA. Cela est également attendu puisque notre approche identifie toutes les interactions potentielles sans optimisation numérique, le potentiel est dans ce cas une borne supérieure des interactions réellement en place dans la communauté.

3.5 Discussion et conclusion

Dans ce chapitre nous avons proposé une méthode permettant d’analyser des communautés naturelles de façon rapide mettant en avant des potentiels de coopération et de compétition. L’une des forces de notre approche est son adaptabilité. En effet, l’obtention de réseaux métaboliques de bonne qualité n’est pas courante dans l’analyse haut-débit de communautés microbiennes. Cette qualité vient d’une curation manuelle approfondie issue de la connaissance extraite de la littérature ou d’expériences *in vitro*. Ainsi, les réseaux métaboliques sont au stade d’ébauche et déterminer une réaction de biomasse peut être compliquée. Cependant, une approche comme CoCoMiCo permet d’obtenir un aperçu des potentiels métaboliques de la communauté, ainsi que ses possibles interactions. En revanche, permettre à l’utilisateur ou l’utilisatrice d’effectuer ou non une optimisation sur une fonction objective de son choix serait une amélioration pertinente pour la méthode.

L’inférence à grande échelle de mécanismes biologiques à partir d’un modèle informatique de communautés bactériennes est un challenge non résolu. Une des difficultés est de cibler l’analyse de communautés de grande taille. Notre approche discrète est un élément de réponse, en proposant de calculer des potentiels de coopération et de compétition à faible coût et utilisable pour les grands jeux de données. Une seconde difficulté concerne la sélection d’une communauté qui serait intéressante pour un objectif donné. Notre approche procure des scores, susceptibles de servir de critère pour sélectionner des communautés candidates et ainsi effectuer prioritairement des opérations de curation manuelle sur les réseaux sélectionnés.

L’explicabilité des modèles à grande échelle est indispensable pour inférer des mécanismes biologiques. La création d’un modèle par raisonnement permet à la fois d’inférer des mécanismes biologiques mais également d’obtenir une traçabilité de l’explication du résultat. De ce fait, les scores de coopération et de compétition sont directement inférés de règles logiques auditables et explicables par et pour les expérimentalistes. Tout comme SMETANA et MICOM à plus petite échelle, notre approche informe l’utilisateur ou l’utilisatrice sur les bactéries intervenant dans les interactions, et propose une liste de composés impliqués.

Afin d’améliorer ce modèle logique et de se rapprocher d’un modèle numérique dynamique et explicable, nous explorerons dans la suite de cette thèse deux pistes. La première permet de sélectionner des communautés à partir des scores d’interaction et en intégrant des optimisations. La seconde, est une preuve de concept intégrant une notion de temporalité aux modèles discrets. Les résultats ainsi que la méthodologie sont présentés dans le chapitre 4.

Chapitre 4

Enrichissement des modèles qualitatifs : modèle de logique linéaire temporelle et méthode de sélection de communauté à façon

Sommaire

4.1	Introduction	117
4.2	Critères d'optimisation à partir de contraintes biologiques <i>a priori</i>	118
4.2.1	Méthodologie	119
4.2.2	Résultats préliminaires	124
4.3	Discussion sur la sélection de communautés	126
4.4	Intégration d'une temporalité réactionnelle	128
4.4.1	Confrontation des cas identifiés par Cabalar et collègues aux besoins des expérimentalistes	130
4.4.2	Application de <i>telingo</i> sur des organismes didactiques	131
4.5	Application à la communauté fromagère du chapitre 2	135
4.5.1	Cas de la planification	136
4.5.2	Cas d'explication	137
4.6	Discussion sur l'intégration de la temporalité aux modèles discrets	138

4.1 Introduction

Ce chapitre 4 présente une synthèse de l’approche hybride développée au cours de cette thèse. D’une part, nous avons construit dans le chapitre 2 un modèle numérique et dynamique du métabolisme de trois bactéries pendant le processus de production de fromage. D’autre part, nous avons développé dans le chapitre 3 un modèle discret par raisonnement, criblant les potentiels de coopération et de compétition d’écosystèmes microbiens de grande taille. Une synthèse des deux approches permettrait d’enrichir l’une avec les avantages de l’autre, ou vice-versa. Un enrichissement de l’approche numérique, conserverait les avantages de la connaissance précise du comportement d’un écosystème microbien, en ajoutant des avantages de passage à l’échelle. Un enrichissement de l’approche discrète conserverait les avantages, notamment, du passage à l’échelle en ajoutant plus de connaissances des comportements des écosystèmes microbiens. Nous avons choisi de privilégier l’amélioration de l’analyse de communautés complexes par la recherche de plus de précision et d’explicabilité.

Nous proposons dans ce chapitre un enrichissement du cadre logique en se basant sur : (i) le retour d’expérience des besoins des biologistes, (ii) sur les enseignements appris de la modélisation numérique. Ces travaux ont pour but de se rapprocher des conclusions qualitatives que les expérimentalistes tirent de l’analyse du modèle numérique. En effet, il n’est pas rare que l’analyse des résultats numériques soit qualitative et conclut, par exemple, qu’un métabolite est produit, en excès ou encore en compétition, sans que la valeur exacte de sa concentration dans le temps soit exploitée. Nous voudrions fournir de tels résultats discrets en épargnant le coût de la construction d’un modèle numérique précis.

Une première amélioration consiste à réduire les sorties du modèle par raisonnement en prenant en compte les besoins des biologistes, traduites sous forme de contraintes logiques. Cette réduction se traduit par une sélection de communautés cohérentes vis à vis des nouvelles règles logiques. En lien avec les travaux présentés du chapitre 3, nous faisons le postulat que définir des fonctions objectives en cohérence avec les besoins des biologistes enrichirait le cadre logique. Dans (Frioux et al., 2018), les auteurs et autrices définissent deux fonctions objectives : minimisation de la taille de la communauté et du nombre d’échanges métaboliques pour sélectionner des communautés candidates. Laisser à l’utilisateur ou l’utilisatrice le choix des optimisations apporterait de la souplesse d’utilisation.

Une seconde amélioration consiste à ajouter une notion de temporalité dans notre modèle discret, sous forme de contraintes logiques temporelles. Cet ajout permettrait en théorie d’obtenir les résultats qualitatifs du modèle numérique du chapitre 2, tels que, la prédiction qu’un métabolite serait en compétition ou qu’un substrat serait en excès. Dans la littérature, il existe un formalisme logique introduit par les travaux de Cabalar et collègues (Cabalar, 2019), adaptant la logique linéaire temporelle (LTL) (Pnueli, 1977) aux concepts de *raisonnement sur les actions et les changements* (RAC) (Prendinger and Schurz, 1996). Cette adaptation permet d’expliquer la cause d’un changement d’un système à partir de règles logiques temporelles et d’observations. Une implémentation efficace est assurée par une extension de solveur ASP : *telingo*. En somme, une inférence de la temporalité et de contraintes biologiques dans le cadre des modèles logiques aurait pour conséquence, de réduire le nombre de solutions possibles et d’affiner les résultats du modèle logique original.

Ainsi, nous verrons au cours de ce chapitre comment l'apport d'une souplesse et d'une facilité d'utilisation permet d'enrichir les possibilités de sélection d'une communauté candidate. Puis dans un second temps, nous verrons comment l'ajout de la temporalité en se basant sur le RAC permet d'apporter des connaissances temporelles supplémentaires par rapport au modèle dynamique du chapitre 2.

4.2 Critères d'optimisation à partir de contraintes biologiques *a priori*

L'une des questions que nous nous sommes posées est la définition de la "meilleure" communauté pour répondre à un objectif. Les travaux de Frioux et ses collègues (Frioux et al., 2018), proposent de minimiser la taille de la communauté ainsi que les échanges métaboliques conduisant à la production de métabolites cibles. Or, au cours de discussions avec des biologistes, nous avons identifié que les choix de minimisation faits dans MiSCoTo n'étaient pas toujours satisfaisants bien qu'ils permettent de limiter l'explosion combinatoire. Retirer l'optimisation n'est pas opportun mais se pose alors la question du choix de la fonction objective à optimiser. En complément de celles utilisées dans MiSCoTo, nous avons calculé des potentiels de coopération et de compétition, ainsi qu'étendu le modèle existant, avec les notions de substrats polyopsonistes et du nombre de contributeurs dans un échange. Ces options supplémentaires enrichissent le choix d'optimisation, contribuant à la définition d'une communauté à fort potentiel de coopération ou de compétition. Parmi l'ensemble de ces éléments (nombre d'échanges, substrats polyopsonistes etc.), il n'existe à notre connaissance, pas de consensus sur la description de la communauté avec de fortes propriétés coopératives/compétitives. Nous avons ainsi développé un prototype laissant à l'utilisateur ou l'utilisatrice la possibilité de sélectionner des communautés selon un ensemble de contraintes pertinent pour leur domaine d'application. Ces contraintes se catégorisent en deux parties. La première, consiste à filtrer la communauté selon la composition taxonomique de cette dernière, c'est-à-dire, la composition bactérienne de la communauté. La seconde, consiste à optimiser des concepts du modèle discret du chapitre 3, que sont les propriétés des interactions, afin d'affiner la sélection. Dans la suite du chapitre nous allons présenter les nouvelles contraintes à implémenter, le prototype réalisé et sa souplesse d'utilisation (voir Figure 4.1

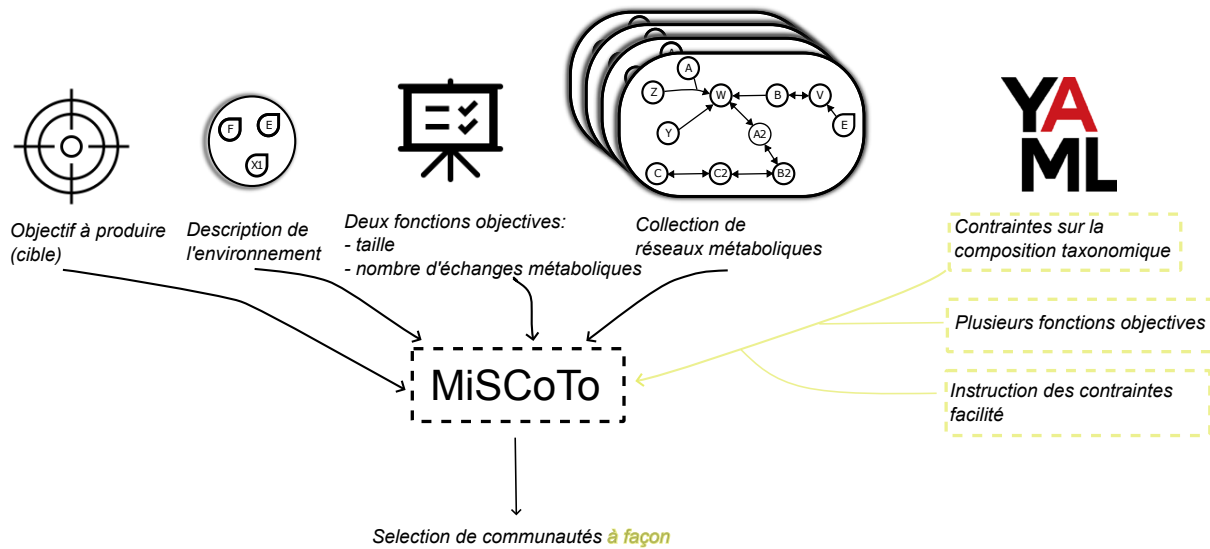


FIGURE 4.1 – Positionnement du développement de notre prototype vis-à-vis de l’outil existant MiSCoTo (Frioux et al., 2018). Les nouveautés sont montrées en jaune, et l’existant en noir.

4.2.1 Méthodologie

Lors du l’élaboration de ce prototype, nous avons réfléchi à la possibilité pour la personne utilisatrice de définir des contraintes diverses tout en respectant un cadre qui facilite leur interprétation par le programme. Nous allons présenter les contraintes supplémentaires, la base de faits créée à partir de ces contraintes et comment la ou le biologiste peut les utiliser. Dans un second temps, l’implémentation de ces nouvelles contraintes sous la représentation mathématique et sous le formalisme ASP seront décrites.

Contraintes de sélection et interface d’utilisation

Puisque les contraintes de modélisation discrète ont été définies dans le chapitre 3, nous allons uniquement décrire les nouvelles contraintes liées à la composition de la communauté. Afin de faciliter la lisibilité et l’accessibilité, nous avons utilisé le format yaml pour décrire les contraintes de façon lisible mais néanmoins encadrée.

Contrainte sur la taille des communautés. L’utilisateur ou l’utilisatrice peut choisir une sélection des tailles de communautés souhaitées. Le prototype propose de filtrer les communautés en sélectionnant une taille fixe, c’est à dire, parmi l’ensemble des combinaisons possibles ou bornée par le nombre minimum et maximum de populations au sein de la communauté. Ainsi, seules les communautés dont les tailles sont comprises entre la valeur minimale et la valeur maximale seront sélectionnées.

Par exemple, un ou une biologiste aimerait obtenir l’ensemble des communautés qui ont une taille comprise entre 6 et 8. Le fichier de configuration se structurerait ainsi :

```

1 # Size constraints
2 SIZE:
3   size:
4   lower bound: 6
5   upper bound: 8

```

À partir de la lecture de ce fichier YAML, l'outil génère la liste de fait ASP, permettant de représenter les contraintes sous forme de connaissance :

```
1 lower_bound_size("6").
2 upper_bound_size("8").
```

Dans le cas où une taille fixe est demandée par les biologistes, seul le mot clé **size** est concerné. Par exemple, la demande "je ne veux que des communautés de taille 4" se traduit par :

```
1 # Size constraints
2 SIZE:
3   size: 4
4   lower bound:
5   upper bound:
```

Avec la **liste de faits ASP** suivante :

```
1 lower_bound_size("4").
2 upper_bound_size("4").
```

Dans le cas où une taille fixe et une taille bornée sont renseignées, l'outil considère seulement la taille fixe.

Contraintes sur la composition taxonomique bactérienne En plus de la taille, il est possible de représenter des contraintes de présence ou d'absence de populations représentées par leur réseau métabolique. Chaque réseau métabolique étant associé à un taxon, par exemple au grain de la souche ou de l'espèce, l'ensemble représente la composition taxonomique de la collection. Cette contrainte répond à des demandes biologiques telles que : "Au sein de ma communauté, je souhaite que les souches **souche1** et **souche2** soient présentes, car elles permettent la libération de composés d'arômes d'intérêt, mais je ne veux surtout pas les souches **souche3** et **souche4**, car je sais qu'elles sont en compétition avec les précédentes souches". Nous pouvons noter que les raisons biologiques justifiantes les intérêts pour ces souches ne sont pas prises en compte au sein de l'outil. Ces contraintes de présence absolue et d'interdiction se traduisent respectivement dans le fichier de configuration par les clés d'entrée suivantes : **always** et **forbid** :

```
1 # Bacterial composition
2 constraints
3 BACTERIAL COMPOSITION:
4   always:
5     - souche1
6     - souche2
7   forbid:
8     - souche3
9     - souche4
```

La liste de faits ASP, à partir de ces contraintes, est la suivante :

```
1 always("souche1").
2 always("souche2").
3 forbid("souche3").
4 forbid("souche4").
```

Nous avons également défini des contraintes supplémentaires sur la composition taxonomique de la communauté : associations interdites et obligatoires, nommées respectivement `forbidden asso` et `mandatory asso`. Une demande biologique illustrant ces propos serait : "S'il existe une communauté dans laquelle une des souches parmi la **souche5**, la **souche6** ou la **souche7** est présente, alors l'ensemble de ces souches doit être présent également. Cependant, il ne faut pas que les souches **souche6**, **souche8** et **souche9** soient toutes les trois présentes ensemble dans la même communauté, même si des sous-ensembles sont autorisés." Ces contraintes diffèrent de `always` et de `forbid` du fait des notions d'ensemble, et l'ensemble nul est considéré comme valide du point de vue méthodologique : par exemple, une communauté où il y a ni la **souche5**, ni la **souche6** ou ni la **souche7**, est valide. Ces associations permettent également de définir des différents groupes de souches interdites ou obligatoires. Par exemple, en complétant la demande : "En plus de ces associations, si une des souches **souche10** et **souche11** est présente, alors, l'ensemble doit être présent également. Cependant, les souches **souche12** et **souche13** ne doit pas être dans la communauté ensemble". Grâce au fichier de configuration, ces contraintes sont formalisables comme suit :

```

1 # Bacterial composition
2 constraints
3 BACTERIAL COMPOSITION:
4 forbidden asso:
5     -
6     - souche6
7     - souche8
8     - souche9
9     -
10    - souche12
11    - souche13
12
13 mandatory asso:
14    -
15    - souche5
16    - souche6
17    - souche7
18    -
19    - souche10
20    - souche11

```

Les différents groupes d'ensemble interdits (resp. obligatoires) sont symbolisés par - à la ligne 5 et 9 (ligne 14 et 18). Ces groupes sont également visibles dans liste de faits ASP générée automatiquement à partir de ces contraintes :

1 forbidden_asso("0", "souche6").	1 mandatory_asso("0", "souche5").
2 forbidden_asso("0", "souche8").	2 mandatory_asso("0", "souche6").
3 forbidden_asso("0", "souche9").	3 mandatory_asso("0", "souche7").
4 forbidden_asso("1", "souche12").	4 mandatory_asso("1", "souche10").
5 forbidden_asso("1", "souche13").	5 mandatory_asso("1", "souche11").

La première variable de chaque atome est un identifiant unique de l'association obligatoire ou interdite considérée.

Flexibilité des fonctions objectives. Laisser le choix à l'utilisateur ou l'utilisatrice d'optimiser une ou plusieurs fonctions objectives correspond à l'enrichissement majeur

du modèle logique. Notre prototype d'outil prend actuellement les choix d'optimisation suivants : maximiser ou minimiser les échanges métaboliques, les métabolites cibles à produire, le scope métabolique et le nombre de substrats polyopsonistes. Les contraintes de maximisation ou de minimisation sont directement décrites dans le fichier de configuration :

```

1 # Optimisation
2 OPTIMISATION:
3   - max
4   - max
5
6 # Objectif
7 OBJ:
8   - target
9   - exchanged

```

Les lignes 3 et 4 indiquent le choix de l'optimisation que l'utilisateur ou l'utilisatrice souhaite faire, et les lignes 8 et 9 sont les objectifs possibles décrits plus haut. Dans cet exemple, le ou la biologiste veut en premier maximiser la production d'un ensemble de cibles, puis, dans un second temps, maximiser le nombre de métabolites échangeables. L'implémentation de cette contrainte se traduit de la façon suivante en ASP :

```

1 #maximize {1@2, M:target(M), pscope(M,_,_) }; 0@2}.
2 #maximize {1@1, S:exchanged(S); 0@1}.

```

Les priorités d'optimisation sont traduites par 1@2 et 1@1, dans lequel la ligne 1 maximise le nombre de métabolites cibles présents dans le scope communautaire (pscope).

Implémentation des nouvelles règles de sélection

Nous venons de voir comment nous avons construit notre base de faits utiles pour formaliser les règles et les contraintes pour la sélection des communautés. Dans la suite de ce chapitre, nous allons formaliser les différentes règles et contraintes en ASP et expliciter le formalisme mathématique.

Sélection par la taille de communauté Nous sélectionnons des communautés de taille satisfaisant les contraintes de la manière suivante :

```

1 lower_bound_size {chosen_bacteria(X):bacteria(X)} upper_bound_size.

```

Listing 4.1 – Contrainte de sélection par la taille

Le listing 4.1 sélectionne un sous ensemble de taille minimale `lower_bound_size` et de taille maximale `upper_bound_size` parmi l'ensemble `chosen_bacteria(X)`, où X est dans le domaine défini par l'ensemble des bactéries (`bacteria`). Cela revient à sélectionner au hasard des bactéries (`chosen_bacteria(X)`) parmi l'ensemble des bactéries initialement défini `bacteria`, tel que le nombre de bactéries sélectionnées inférieur à `upper_bound_size` et supérieur à `lower_bound_size`.

Soit t un sous-ensemble de taxons appartenant à l'ensemble de taxons \mathcal{T} pour lesquels un réseau métabolique est disponible, la cardinalité $|t|$ doit être comprise entre les valeurs lb et ub :

$$t \in \mathcal{T}$$

$$lb < |t| < ub$$

Sélection par des contraintes de composition de la communauté Nous avons laissé la possibilité à l'utilisateur et à l'utilisatrice de filtrer la composition de la communauté, avec les prédicats `always`, `forbid`, `mandatory_asso` et `forbidden_asso`.

Le prédicat `always`, représentant un taxon qui doit être présent dans toutes les communautés possibles, est implémenté en ASP comme suit :

```
1 :- always(X); not chosen_bacteria(X).
```

Cette règle ne dispose pas de tête, on parle alors de contrainte d'intégrité. Elle se lit comme suit : il ne doit pas exister de modèle dans lequel une population définie par le prédicat `always(X)` n'est pas présente. Cela implique de sélectionner systématiquement la population décrite dans `always`.

L'implémentation ASP de l'existence de populations qui ne doivent pas être présentes dans la communauté se lit :

```
1 :- chosen_bacteria(X); forbid(X).
```

Cela se traduit par : il ne peut y avoir de modèle dans lequel une population décrite par (`forbid(X)`) est sélectionnée.

Nous avons également défini des associations obligatoires (`mandatory_asso`) et interdites (`forbidden_asso`). Pour rappel, l'utilisateur ou l'utilisatrice pouvait définir des groupes d'association, nommés `group_forbidden(G)` et `group_mandatory(G)` dans le cas des associations interdites et obligatoires respectivement.

Le listing 4.2 présente le premier cas : il ne peut y avoir de modèle contenant l'ensemble des bactéries interdites (`chosen_bacteria(X) : forbidden_asso(G,X).`)

```
1 :- group_forbidden(G); chosen_bacteria(X) : forbidden_asso(G,X).
```

Listing 4.2 – Règle définissant les associations interdites

Enfin, les associations obligatoires dans une communauté sont définies par la règle décrite dans le listing 4.3.

```
1 chosen_bacteria(X) :- chosen_bacteria(X1);
2                       group_mandatory(G);
3                       mandatory_asso(G,X);
4                       mandatory_asso(G,X1).
```

Listing 4.3 – Règle définissant les associations obligatoires

Cette règle sélectionne une population `X` si une population `X1` est déjà sélectionnée et que les deux appartiennent à la même association obligatoire `G` (`mandatory_asso(G,X1)` et `mandatory_asso(G,X)`).

4.2.2 Résultats préliminaires

Nous venons de décrire et détailler la base de connaissances nécessaire pour sélectionner des communautés à façon. Dans cette sous-section, nous allons dans un premier temps, identifier des questions biologiques provenant d'un retour d'expériences sur les précédents chapitres, ainsi que de discussions avec les biologistes, et comment l'outil de sélection permet d'y répondre. Puis dans un second temps, nous allons appliquer notre méthode de sélection de communauté à façon sur le jeu de données issu de (Belcour et al., 2020), afin de montrer son efficacité à passer à l'échelle.

Exemples

Toutes les questions ci-dessous vont chercher à identifier des ensembles de communautés candidates pour des objectifs métaboliques supposés connus. Les questions seront présentées des plus larges aux plus précises. Nous utiliserons les réseaux métaboliques reconstruits à l'échelle du génome de (Weiss et al., 2022) représentant des populations microbiennes intestinales murines. Ils ont été reconstruits avec l'outil gapseq (Zimmermann et al., 2021) et forment un consortium bactérien composé de 12 GEMs.

Pour une communauté de taille 5, existe-t-il une communauté permettant la production d'un ensemble de métabolite cible ? Cette question nous permet d'identifier les deux contraintes à utiliser au sein de l'outil : la liste des composés que l'on doit produire, comprenant ici par exemple la liste des acides aminés, du propionate et du 6-Oxoheptanoate, ainsi que la taille de la communauté finale, *i.e.* 5.

Le fichier initial est donc constitué des informations suivantes :

```
1 # Size constraints
2 SIZE:
3   size: 5
```

N'ayant précisé aucune optimisation, l'ensemble des cibles, que sont les acides aminés, devront être dans le scope de la communauté. Le résultat pour cette situation est **unsatisfiable**. Cela signifie qu'il n'existe pas de modèle permettant de satisfaire les contraintes données, et donc aucune communauté de tailles 5 n'a été trouvée comme étant capable de produire l'ensemble des cibles à partir de l'environnement fourni. Le premier réflexe est de relâcher les contraintes. La contrainte la plus forte ici consiste à obtenir au moins une communauté permettant de **produire l'ensemble des éléments cibles**. Une reformulation possible serait de produire un **maximum** de ces métabolites. En appliquant cette modification nous obtenons une communauté composée des souches *I48*, *KB1m*, *YL44*, *YL45*, *YL27*. Cette communauté est capable d'en produire 20, seul le 6-Oxoheptanoate n'est pas productible à partir des graines fournies.

Pour une communauté composée de 6 taxons, laquelle permet la plus grande production métabolique en limitant les échanges entre les membres de la communauté ? Nous avons ici 2 types de contraintes, d'une part des optimisations : maximisation du scope communautaire et minimisation des échanges métaboliques ; et d'autre part une contrainte sur la taille de la communauté. Ces contraintes se traduisent comme suit dans le fichier de configuration :

```
1 # Size constraints
```

```

2 SIZE:
3   size: 6
4
5   # Optimisation
6 OPTIMISATION:
7   - max
8   - min
9
10  # Objectif
11 OBJ:
12   - scope
13   - exchanged

```

La meilleure communauté, selon le solveur, présente 794 échanges possibles et 990 métabolites productibles en communauté. La communauté sélectionnée est composée de *YL58*, *KB1f*, *I48*, *YL2* et *YL31*. Le formalisme développé dans le chapitre 3 est également accessible avec cet outil. Ainsi, le nombre de métabolites échangeables, les consommateurs et producteurs intervenant dans l'échange, les substrats polyopsonistes et les potentiels de coopération et de compétition sont accessibles pour chaque communauté.

À partir d'une communauté de 150 réseaux métaboliques de l'intestin, j'aimerais étudier une communauté composée de 5 taxons garantissant une productivité d'un maximum de composés d'intérêt parmi 23 En plus de ces contraintes, il y a des contraintes de composition taxonomique qui sont récapitulées dans le listing ci-dessous. Cette demande factice teste le passage à l'échelle de la méthode. Nous avons donc utilisé 150 réseaux métaboliques reconstruits dans le chapitre 3 :

contrainte d'entrée :

```
1 # Size constraints
2 SIZE:
3   size:
4     lower bound: 5
5     upper bound: 5
6
7 # Bacterial composition
8 constraints
9 BACTERIAL COMPOSITION:
10  in:
11    - GCA_003433685
12    - GCA_003433675
13    - GCA_003433665
14  out:
15    - GCA_003433745
16    - GCA_003433755
17    - GCA_003433765
18  forbidden asso:
19    -
20      - GCA_003433795
21      - GCA_003433825
22      - GCA_003433845
23    -
24      - GCA_003433855
25      - GCA_003433865
26
27  mandatory asso:
28    -
29      - GCA_003433905
30      - GCA_003433925
31      - GCA_003433945
32    -
33      - GCA_003433955
34      - GCA_003433985
35
36 # Optimisation
37 OPTIMISATION:
38   - max
39
40 # Objectif
41 OBJ:
42   - target
```

sortie :

```
1 Answer: 1
2 GCA_003433665
3 GCA_003433675
4 GCA_003434075
5 GCA_003436935
6 GCA_003433685
```

Performance :

Models : 1

Time : 752.543s

En maximisant une liste de composés d'intérêt (23) et à partir des paramètres ci-dessus, nous avons trouvé une communauté candidate en environ 12 minutes et 30 secondes. Sur les 23 composés devant être produits, 7 peuvent l'être dans le milieu décrit.

4.3 Discussion sur la sélection de communautés

Cette première exploration d'enrichissement d'un modèle logique a répondu aux remarques majeures que nous avons énoncées plus haut : filtrer le nombre de solutions d'un modèle logique et assouplir les fonctions objectives. L'outil développé dans le chapitre 3 permet de donner une indication des potentiels de coopération et de compétition de communautés microbiennes. Un travail supplémentaire reste néanmoins nécessaire pour intégrer les potentiels d'interaction comme critère d'optimisation. Avec ce prototype, nous avons proposé un moyen de sélectionner la meilleure communauté selon les critères de l'utilisateur ou de l'utilisatrice, et par extension, de définir une communauté avec un fort potentiel de coopération ou de compétition. Cela a été possible en ajoutant une optimi-

sation d'une ou de plusieurs fonctions objectives. La plus-value par rapport au logiciel MiSCoTo est que la personne utilisatrice choisit la fonction objective à optimiser.

Cependant, aucune validation expérimentale n'a été faite à ce jour. Au regard de la littérature concernant la création de communautés à façon, une validation expérimentale est nécessaire pour créer une bonne communauté synthétique (Sharma and Shukla, 2020; Vázquez-Castellanos et al., 2019). Par exemple, Macchi et collègues (Macchi et al., 2021) ont construit une communauté bactérienne à partir de 6 génomes isolés pour la dégradation du phénanthrène.

Vis à vis des interactions bactériennes, ce prototype permet de se concentrer essentiellement sur le potentiel de coopération d'une communauté bactérienne. Dans le chapitre 2, un des composés polyopsonistes était le lactose co-consommé par deux souches. Pourtant, cette co-consommation ne se déroule pas au même moment dans le temps, empêchant de définir un lien de compétition pour ce substrat. Afin d'essayer de prendre en compte une pseudo-temporalité au sein des modèles logiques, nous allons utiliser la logique temporelle de Cabalar et collègues (Cabalar, 2019).

4.4 Intégration d'une temporalité réactionnelle

Le raisonnement sur les actions et le changement (RAC) permet d'étendre notre cadre formel à la temporalité par l'usage de la logique linéaire temporelle (LTL). Cet enrichissement temporel propose des explications séquentielles de l'observation. Cinq RAC ont été identifiés dans (Cabalar, 2019) :

- La **simulation** : étant donné un état initial du système et une suite d'actions, la simulation permet de prédire *l'état du système à un temps futur*.
- L'**explication** : étant donné un état initial et un état futur du système, l'explication propose *une séquence d'états intermédiaires* du système menant de l'état initial à l'état final.
- La **planification** : étant donné un état initial du système et un objectif, la planification propose *une séquence d'actions*, éventuellement non-déterministe, garantissant l'objectif.
- La **diagnostic** : à partir d'une définition de transitions d'état normal à anormal, ainsi qu'un ensemble d'état observé (normaux et anormaux), le diagnostic propose un *ensemble minimal de transitions* expliquant les observations anormales.
- La **vérification** : vérifier une *propriété temporelle* du système (comme sûreté, vivacité ou équité) ou une propriété de représentation (comme l'équivalence de modèle).

Le premier défi est l'application de ces concepts RAC à la biologie des systèmes. Nous avons donc associé le **système** à une communauté de GEMs et les **actions** à des activations de réactions. Ainsi, chaque nouvelle activation d'une réaction, selon l'algorithme d'expansion du réseau métabolique d'Ebenhöh, constituera un nouvel état du système. D'après le modèle de la communauté fromagère du chapitre 2, nous savons que le propionate est produit à partir du lactate. Les états du système peuvent être décrits ainsi :

1. *état initial* : présence de lactose, absence de lactate
2. *état intermédiaire* : production et présence de lactate ; absence de propionate
3. *état futur* : production et présence de propionate

À partir de cette information nous pouvons déduire *a minima* une relation de dépendance logique temporelle : pour produire le propionate dans le futur, il a fallu du lactate dans le passé. La première exigence de la logique temporelle est de décrire au sein des règles logiques, ces notions temporelles au moyen d'opérateurs spécifiques (voir Figure 4.2). La seconde, consiste à traduire ces règles temporelles dans le cadre de la biologie : pour qu'à *un moment de la cinétique de fermentation* le propionate soit productible, il faut que le lactate soit productible *avant*.

En appliquant la LTL à des réseaux métaboliques, les notions temporelles *un moment* et *avant* ont été définies selon la longueur du chemin menant à la production du composé final, le propionate. À partir d'un ensemble de nutriments disponibles dans le milieu extracellulaire, et un ensemble fini de voies métaboliques menant à la production d'un composé cible, le solveur *telingo* priorisera le chemin le plus court. Il proposera l'ensemble des voies métaboliques avec le plus petit nombre de réactions. De ce fait, la temporalité n'est pas représentée comme une dimension, telle que dans le dFBA par exemple, mais par une séquence d'actions, activation ou non d'une réaction : on parle de *temporalité réactionnelle*. Dans la suite de ce chapitre nous allons décrire le vocabulaire temporel à

travers les cas d'utilisations identifiés par Cabalar (2019) et les appliquer à la communauté du chapitre 2.

Inférence temporelle dans la programmation logique

Ce modèle de logique par raisonnement se base sur le même formalisme que celui décrit dans le chapitre 3 : une base de faits, des règles et contraintes doivent être générées. *Telingo* permet d'intégrer des contraintes temporelles, et l'ensemble de ces règles et contraintes utilisent les opérateurs logiques temporels décrits dans la Figure 4.2.

<code>&initial</code>	I	<i>initial</i>	<code>&final</code>	F	<i>final</i>
<code>'p</code>	•p	<i>previous</i>	<code>p'</code>	◦p	<i>next</i>
<code><</code>	•	<i>previous</i>	<code>></code>	◦	<i>next</i>
<code><?</code>	S	<i>since</i>	<code>>?</code>	U	<i>until</i>
<code><*</code>	T	<i>trigger</i>	<code>>*</code>	R	<i>release</i>
<code><?</code>	◆	<i>eventually before</i>	<code>>?</code>	◇	<i>eventually afterward</i>
<code><*</code>	■	<i>always before</i>	<code>>*</code>	□	<i>always afterward</i>
<code><:</code>	∘	<i>weak previous</i>	<code>>:</code>	∘	<i>weak next</i>

FIGURE 4.2 – Opérateurs temporels du futur et du passé en *telingo* et logique temporelle issue de (Cabalar, 2019)

En reprenant l'exemple des items 2 et 3, nous avons défini une proposition qui est toujours vraie : le propionate est *productible* si le lactate est *productible* **au temps précédent** et que le lactate est *présent* **au temps précédent** :

```
1   producible(propionate) ⊢ □ (•producible(lactate);
2                                 •present(lactate)).
```

En ASP, la règle temporelle serait donc :

```
1   producible(propionate) :- &tel{>*(<producible(lactate);
2                                 <present(lactate))}.
```

À partir de cette nouvelle syntaxe, nous décrivons dans la suite une communauté didactique du chapitre 2 afin d'appliquer la LTL et d'illustrer la temporalité réactionnelle au moyen de trois concepts RAC : **planification**, **vérification** et **explication**.

4.4.0.1 Démarche suivie

Nous avons dans un premier temps représenté les réseaux métaboliques de *P. freudenreichii*, *L. plantarum* et *L. lactis* de manière simplifiée et minimaliste. Ce choix méthodologique permet d'une part d'éviter la problématique du passage à l'échelle, et d'autre part de résoudre les défis rencontrés lors de l'application des concepts des travaux de Cabalar et collègues. Parmi ces enjeux, nous avons dû gérer la disponibilité d'un composé pour une réaction ainsi que la disponibilité des produits jusqu'à la fin de la simulation. Dans les sous-sections qui suivent, nous présenterons les différents cas d'utilisation RAC ainsi que leur résolution à partir d'une communauté didactique, c'est-à-dire, répondant aux divers enjeux créés par ce formalisme, puis, l'application de ces concepts sur les réseaux métaboliques complets du chapitre 2.

4.4.1 Confrontation des cas identifiés par Cabalar et collègues aux besoins des expérimentalistes

Ensemble de règles en *telingo* pour la modélisation d'une communauté de bactéries

Scope Nous avons redéfini le modèle exprimé dans le chapitre 3 en intégrant le formalisme temporel de la LTL. Ainsi, les notions de réactions activées, explicitées par le listing 4.4 et de composés productibles, décrit par le listing 4.5 ont dû être modifiées. Une réaction *peut* être activée (représentée par le choix $0\{1\}$) si l'ensemble des réactants ont été produits *avant* (représenté par l'apostrophe '). Un composé est considéré comme productible si c'est un produit d'une réaction activée.

```
1 0{isActivated(Rea,Org)}1 :- product(Compound,Rea,Org,_);
2 'productible(Reactant,Org) : reactant(Reactant,Rea,Org,).
```

Listing 4.4 – Règle d'activation d'une réaction

```
1 productible(Compound,Org):- product(Compound,Rea,Org,_),
2 isActivated(Rea,Org).
```

Listing 4.5 – Règle de productibilité d'un composé

L'évolution de la concentration d'un métabolite au cours du temps n'étant pas inférée dans le modèle, nous ne pouvons déterminer le moment réactionnel où une réaction n'est plus activée. Nous faisons alors l'hypothèse que si une réaction est activée une fois, elle reste activée, laissant les produits de la réaction disponibles jusqu'à la fin de la simulation. Cette règle s'exprime dans le listing 4.6 dans lequel, l'opérateur temporel $<?$ signifie *tant que*. Cette règle peut se traduire de la façon suivante : Tant que la réaction a été activée dans le passé (par une voie métabolique), elle restera activée.

```
1 not isActivated(Rea,Org) :- not not &tel{ <? isActivated(Rea,Org)},
2 reaction(Rea,Org).
```

Listing 4.6 – Règle interdisant qu'une réaction soit activée deux fois

Disponibilité des nutriments nouvellement productibles Avec cette notion temporelle à l'échelle réactionnelle, un composé produit en amont doit être disponible pour une réaction qui se situerait en aval d'une chaîne réactionnelle. La Figure 4.3 illustre ce concept que nous avons développé. À partir du composé extracellulaire A_e , le composé C est productible au temps réactionnel 2 (représenté par t_2) dans la Figure 4.3. Il est ainsi disponible pour la réaction située trois temps réactionnels en aval, la réaction R4.

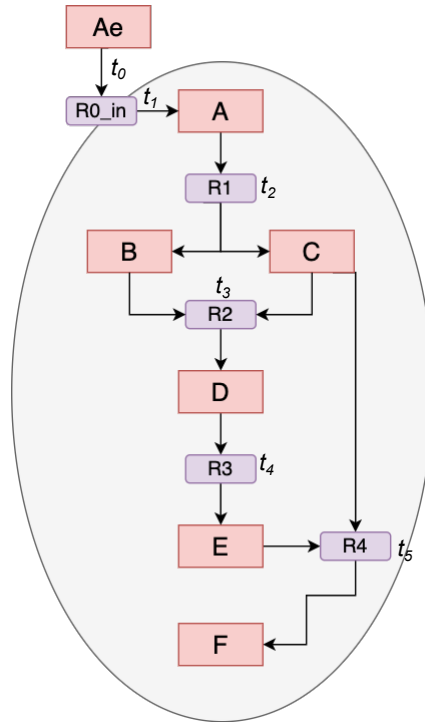


FIGURE 4.3 – Réutilisation du composé C par la réaction R4 dans le cas où C était productible trois temps réactionnels avant

Avec cette nouvelle propriété, la définition de réaction activée est redéfinie ainsi : Une réaction est activée si tous les réactants de la réaction ont été productibles avant, représentée par l’apostrophe à ligne 2 c’est à dire, **•disponible** en notation LTL :

```

1 disponible(Compound,Org) :- producible(Compound,Org) .
2 disponible(Compound,Org) :- 'disponible(Compound,Org) .

```

Au sein d’une communauté bactérienne, nous observons des échanges métaboliques entre bactéries. Les métabolites échangés ont été produits et ils deviennent alors disponibles pour l’ensemble des membres de la communauté.

4.4.2 Application de *telingo* sur des organismes didactiques

Nous venons de voir les règles de base permettant de modéliser le métabolisme d’une communauté bactérienne avec la syntaxe temporelle. Dans les sections suivantes, nous verrons l’application du formalisme *telingo* sur des communautés didactiques afin de vérifier si les concepts RAC s’adaptent à la biologie des systèmes. Les règles et les contraintes utilisées pour modéliser les concepts RAC ne seront pas détaillées, seul l’aspect temporel sera mis en avant.

Planification

Nous avons dès à présent les éléments nécessaires pour modéliser les concepts RAC. Pour cela nous avons utilisé les réseaux métaboliques illustrés par les Figures 4.4 et 4.5. Dans le cas de la planification, l’objectif à atteindre est de déterminer la communauté minimale pour la production du composé Ne. Dans la Figure 4.4, ce composé peut être

produit soit par l'organisme 1 seul soit par l'organisme 2 par l'intermédiaire de 2 échanges métaboliques. La sortie du modèle est indiquée par le chemin en vert.

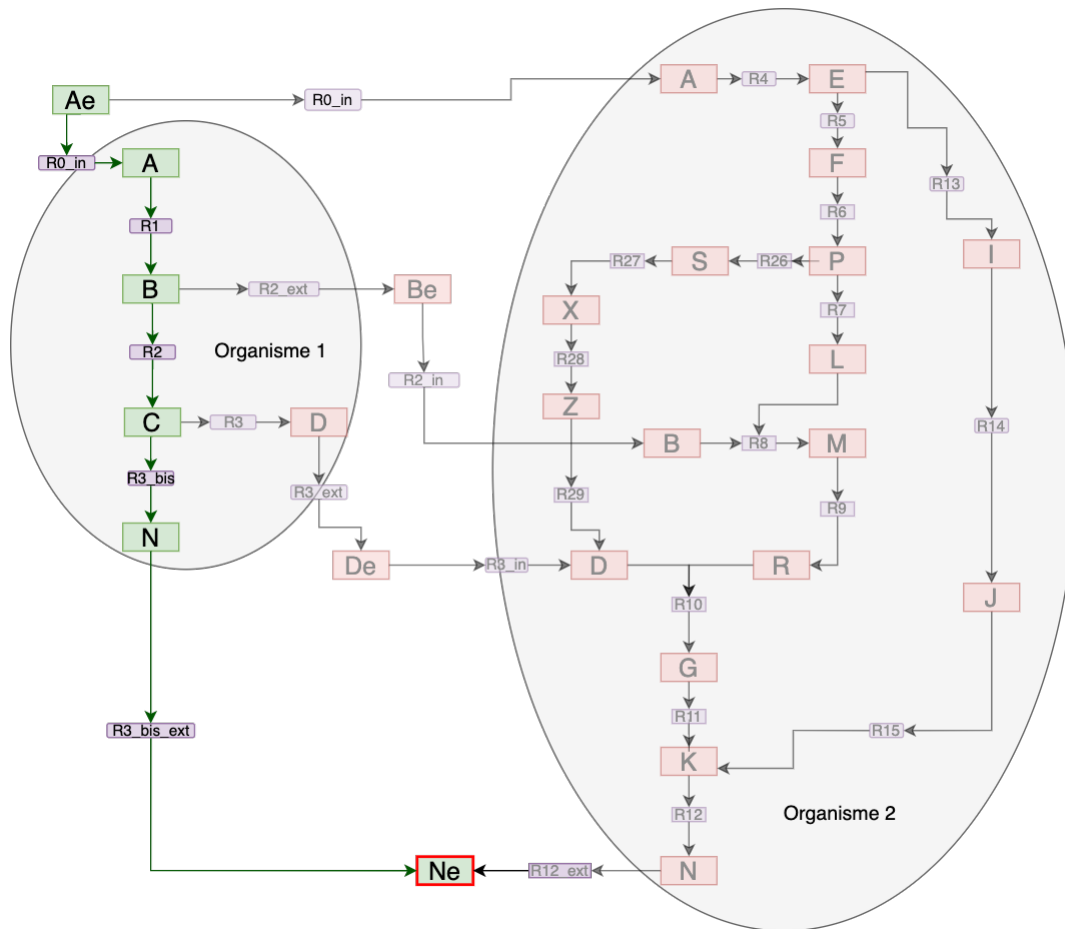


FIGURE 4.4 – Schéma didactique permettant de valider la solution donnée par le modèle dans le cas de la planification. Le composé Ne est productible à partir d'un seul organisme via le chemin vert.

telingo montre que l'organisme 1 seul peut produire le composé Ne. Biologiquement, une communauté composée d'une seule population peut être un non-sens, mais nous n'avons pas écrit de règles en ASP indiquant le contraire. Le schéma réactionnel trouvé par *telingo* est indiqué par le listing 4.7.

```

1 Answer: 1
2 State 0:
3 State 1:
4   isActivated("R0_in","Org1")
5   producible("A","Org1")
6 State 2:
7   isActivated("R1","Org1")
8   producible("B","Org1")
9 State 3:
10  isActivated("R2","Org1")
11  producible("C","Org1")
12 State 4:
13  isActivated("R3_bis","Org1")
14  producible("N","Org1")

```

```

15 State 5:
16   isActivated("R3_bis_ext","Org1")
17   producible("Ne","Org1")
18 Optimization: 1 5
19 OPTIMUM FOUND
20
21 Models      : 1
22   Optimum    : yes
23 Optimization: 1 5
24 Calls       : 6
25 Time        : 0.265s (Solving: 0.00s 1st Model: 0.00s Unsat: 0.00s)
26 CPU Time    : 0.197s

```

Listing 4.7 – Résultat de la planification obtenue après application des règles sur la base de faits

La séquence des actions montrant la production de Ne est représentée par **State** dans la sortie du modèle. Ici, Au temps réactionnel 0, aucune réaction n'est activée. Seul le composé extracellulaire Ae serait marqué comme étant **disponible**. Au temps réactionnel suivant, **State 1**, la réaction important le composé Ae dans le milieu intracellulaire est activée par la réaction R0_in (voir ligne 4 du listing), et le composé A est **productible** (ligne 5). Au temps réactionnel suivant, B est **productible** et la réaction R1 est activée. En procédant ainsi, le composé Ne est **productible** et **disponible** pour l'ensemble de la communauté en 5 étapes, correspondant au temps réactionnel 5. À partir de l'état initial du système, c'est-à-dire, deux organismes et un composé dans le milieu extracellulaire, nous avons proposé une séquence d'actions garantissant notre objectif, *i.e.* la chaîne réactionnelle minimale produisant le composé Ne. Nous répondons donc au cas de la planification.

Vérification

Le cas de vérification consiste simplement à vérifier qu'il existe une séquence de réactions activables permettant de produire un composé. Ainsi, nous avons vérifié la productibilité du composé Ne (voir listing 4.7) :

Nous pouvons voir qu'à chaque pas de temps décrit par **State**, il existe au moins une réaction qui est activée, permettant à terme, de produire le composé Ne.

Explication

Dans le cas de l'explication, nous cherchons à déterminer le chemin réactionnel permettant la production d'un composé sachant que l'on a observé la production d'un autre composé, à un moment dans le temps. L'objectif est de savoir s'il existe une relation entre une observation faite à un moment dans le temps et l'objectif final. Dans le cas de notre exemple de l'explication représentée par la Figure 4.5, nous avons cherché à expliquer la production du composé Ne sachant que le composé De a été observé. Le résultat du modèle est indiqué par le tracé en vert sur la Figure 4.5 et par le listing 4.8 :

```

1 State 0:
2 State 1:
3   isActivated("R0_in","Org1") isActivated("R0_in","Org2")
4   producible("A","Org1") producible("A","Org2")
5 State 2:
6   isActivated("R1","Org1") isActivated("R4","Org2")
7   producible("B","Org1") producible("E","Org2")
8 State 3:
9   isActivated("R2","Org1") isActivated("R2_ext","Org1") isActivated("R5","Org2")

```

```

10     producible("Be","Org1") producible("C","Org1") producible("F","Org2")
11 State 4:
12     isActivated("R2_in","Org2") isActivated("R3","Org1") isActivated("R6","Org2")
13     producible("B","Org2") producible("D","Org1") producible("P","Org2")
14 State 5:
15     isActivated("R7","Org2")
16     producible("L","Org2")
17 State 6:
18     isActivated("R3_ext","Org1") isActivated("R8","Org2")
19     producible("De","Org1") producible("M","Org2")
20 State 7:
21     isActivated("R3_in","Org2") isActivated("R9","Org2")
22     prod_explain("D","Org2")
23     producible("D","Org2") producible("R","Org2")
24 State 8:
25     isActivated("R10","Org2")
26     prod_activate("G","Org2")
27     prod_activate_cof("G","Org2")
28     producible("G","Org2")
29 State 9:
30     isActivated("R11","Org2")
31     prod_activate("K","Org2")
32     prod_activate_cof("K","Org2")
33     producible("K","Org2")
34 State 10:
35     isActivated("R12","Org2")
36     prod_activate("N","Org2")
37     prod_activate_cof("N","Org2")
38     producible("N","Org2")
39 Optimization: 3 17
40 OPTIMUM FOUND
41 Models      : 4
42   Optimum   : yes
43 Optimization : 3 17
44 Calls      : 11
45 Time       : 0.366s (Solving: 0.02s 1st Model: 0.00s Unsat: 0.00s)
46 CPU Time   : 0.350s

```

Listing 4.8 – Résultat obtenu du cas d’explication après application des règles sur la base de faits didactiques

Au temps réactionnel 6 (ligne 19), nous voyons que l’observation est expliquée (De disponible), mais qu’il faut attendre le temps réactionnel 10 pour que le composé final Ne soit productible. Nous pouvons également remarquer que le modèle explique en même temps la production du composé Be pour permettre la production de Ne : Nous avons défini Be comme étant essentiel à la production de Ne dans le cas de l’explication.

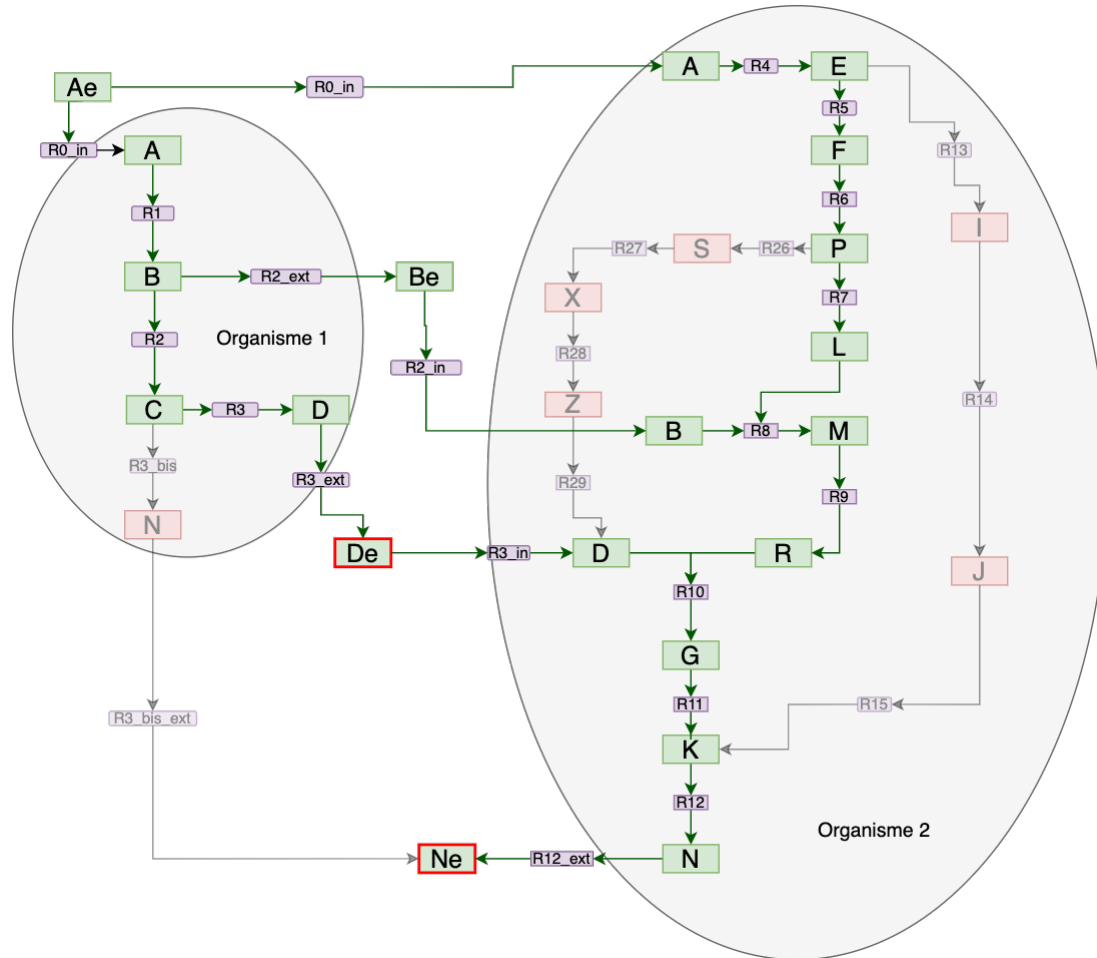


FIGURE 4.5 – Schéma didactique permettant de valider la solution donnée par le modèle dans le cas de l’explication en montrant que le composé Ne est productible à partir de l’élément observable De.

Cette implémentation temporelle, sur un schéma didactique, met en avant deux points intéressants par rapport à un dFBA. Le premier, est la possibilité d’avoir la suite réactionnelle conduisant à un ou plusieurs objectifs finaux de façon directe. Cette représentation en temps réactionnel, permet donc de générer des hypothèses testables sur l’activation ou non de voies métaboliques d’intérêt, à travers des méthodes de knockout enzymatique par exemple. La seconde est la mise en avant indirecte d’une coopération entre ces deux organismes. Pour rappel, la connaissance initiale montrait une observation dans le milieu extracellulaire de De et non un échange entre ces deux organismes. Nous avons également montré qu’un autre composé est échangé et essentiel, pour permettre de réaliser l’objectif final.

4.5 Application à la communauté fromagère du chapitre 2

Nous avons repris le milieu du lait ainsi que les réseaux métaboliques entiers du chapitre 2, et généré la nouvelle base de faits applicables à *telingo*. Nous ne présenterons que deux concepts RAC : la planification et l’explication.

4.5.1 Cas de la planification

Dans le cas de la planification, le système initial est composé des 3 bactéries et du milieu nutritionnel du lait. L'objectif à atteindre est la production de propionate, noté `M_ppa_e` dans la sortie du modèle représenté par 4.9 :

```
1 State 0:
2 State 1:
3   isActivated("R_ADK1_rev","lactis") isActivated("R_ARGDr","lactis")
4   producible("M_amp_c","lactis") producible("M_atp_c","lactis")
5     producible("M_citr__L_c","lactis") producible("M_nh4_c","lactis")
6 State 2:
7   isActivated("R_NTD7","lactis") isActivated("R_OCBT_rev","lactis")
8     isActivated("R_THRD_L","lactis")
9   producible("M_2obut_c","lactis") producible("M_adn_c","lactis")
10    producible("M_cbp_c","lactis") producible("M_nh4_c","lactis")
11    producible("M_orn_c","lactis") producible("M_pi_c","lactis")
12 ...
13 ...
14 ...
15 State 17:
16   isActivated("R_PPAKr","lactis")
17   producible("M_atp_c","lactis") producible("M_ppa_c","lactis")
18 State 18:
19   isActivated("R_DM_ppa_c","lactis")
20   producible("M_ppa_p","lactis")
21 State 19:
22   isActivated("R_DM_ppa_p","lactis")
23   producible("M_ppa_e","lactis")
24 Optimization: 1 26
25 OPTIMUM FOUND
26 Models      : 44
27   Optimum    : yes
28 Optimization : 1 26
29 Calls       : 20
30 Time        : 65.678s (Solving: 15.03s 1st Model: 0.08s Unsat: 0.10s)
31 CPU Time    : 57.789s
```

Listing 4.9 – Résultat de la planification obtenue après application des règles sur la base de faits issue des données réelles

Un premier résultat intéressant consiste à observer que la méthode semble passer à l'échelle d'une communauté bactérienne de taille 3 (65 secondes), composé de plusieurs milliers de réactions. Un second point surprenant concerne l'unique organisme prédit comme producteur du propionate : *L. lactis*. En effet, nous savons que seule la bactérie propionique (*P. freudenreichii*) est capable de produire ce composé expérimentalement. Ce résultat n'est néanmoins pas surprenant puisque nous nous basons ici seulement sur la topologie des réseaux métaboliques. Enfin, parmi les 44 modèles trouvés par `telingo`, aucun ne prédit d'échanges métaboliques au sein de la communauté, comparé aux résultats des méthodes numériques décrits dans le chapitre 2. Ce résultat montre que l'application du concept RAC de la planification à un jeu de données réel mérite une investigation supplémentaire afin d'être pleinement exploitée.

4.5.2 Cas d'explication

Nous avons également vu dans le chapitre 2, que le propionate est productible à partir du lactate, produit par les bactéries lactiques *L. lactis* et *L. plantarum*. Ainsi, expliquer la production de propionate sachant le lactate observé et consommé par *P. freudenreichii* constitue un exemple d'utilisation pour le cas d'explication. Dans un souci de visualisation des résultats, nous avons représenté dans la Figure 4.6, une simplification des voies métaboliques utilisées par le modèle pour produire le propionate sachant le lactate observé. Pour rappel, aucune simplification des réseaux métaboliques n'a été faite.

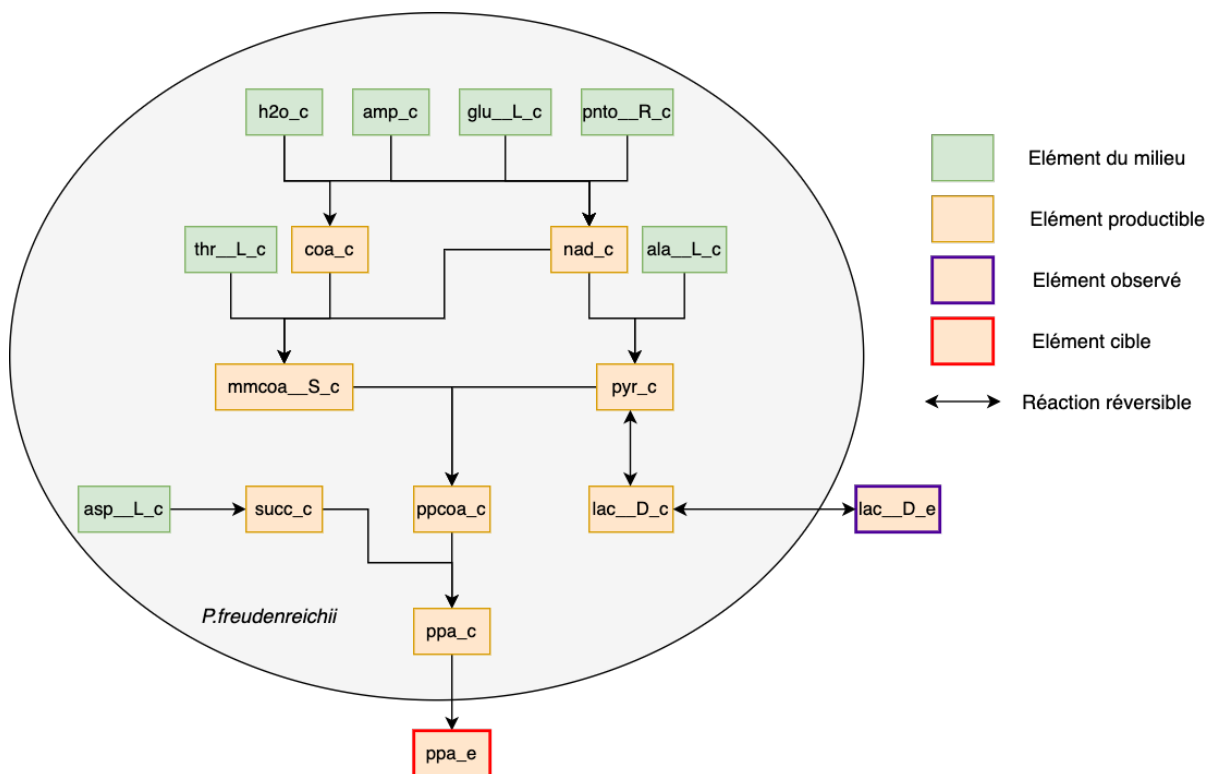


FIGURE 4.6 – Schéma permettant de visualiser le résultat simplifié du modèle d'explication sur les données réelles.

Contrairement au cas de planification, il a fallu 23 étapes pour expliquer la production de propionate à partir du lactate. Malgré l'ajout d'une connaissance *a priori* (consommation du lactate par *P. freudenreichii*) le modèle ne permet toujours pas d'identifier les interactions entre les bactéries lactiques et la bactérie propionique. *L. lactis* semble être capable de produire le propionate à partir du lactate. Tout comme la planification, un travail supplémentaire est nécessaire. Cependant, nous pouvons reconnaître un cycle important intervenant dans la production du propionate, identifié par l'une des solutions proposées par *telingo* : le cycle de Wood-Werkman.

4.6 Discussion sur l'intégration de la temporalité aux modèles discrets

Nous avons montré, à travers cette application de la logique temporelle, une piste d'enrichissement de notre approche hybride en utilisant le raisonnement sur les actions et le changement. Pour y arriver, nous avons dû tout d'abord traduire des besoins des biologistes en termes de RAC et LTL. Nous avons analysé leur demande cas par cas au moyen d'exemples didactiques, fournissant une explication détaillée et temporelle à des observations faites par des expérimentalistes. Nous avons ensuite appliqué ce formalisme à des vrais jeux de données, montrant des limites à cette approche.

En somme, ce formalisme logique temporel est applicable à la biologie des systèmes mais le manque de précision qualitatif empêche d'obtenir une réelle plus-value de la méthode. En effet, lors de l'application sur un jeu de données réelles, l'identification des voies métaboliques activables, à chaque temps réactionnel, est contrebalancée par les erreurs biologiques des prédictions. En revanche, sur les exemples didactiques, elle permet l'identification de nouvelles interactions bactériennes expliquant une observation biologique et générant des nouvelles hypothèses testables.

Dans le but d'affiner ce modèle, une inférence de nouvelles règles et contraintes à partir des observations biologiques contribuerait à son amélioration. À partir d'une base de connaissance à utiliser, provenant de l'expert ou l'experte, de données omiques ou de la biologie, quelles sont les règles pouvant être apprises ? Par exemple, en reprenant l'application aux données du chapitre 2, nous savons que le lactate ne peut venir que des bactéries lactiques. Ainsi, une proposition temporelle concernant le cas d'explication pourrait être : "Expliquer la production de propionate, sachant que le lactate est observé et produit par les bactéries lactiques et consommé par *P. freudenreichii*". Cette démarche aurait pour but d'éliminer des faux positifs : biologiquement non connus (production de lactate par *P. freudenreichii*), biologiquement impossibles (production de lactate par *P. freudenreichii*) ou encore, expérimentalement impossible à vérifier (l'enchaînement temporel d'un élément produit et/ou consommé en intracellulaire).

Chapitre 5

Conclusion et perspectives

Bilan des contributions

Dans cette thèse nous avons développé plusieurs méthodes d'analyse et modélisation de communautés microbiennes applicables à des communautés formant un écosystème simplifié, ou bien, à des communautés simulées provenant d'écosystèmes naturels complexes. Notre approche propose d'une part de caractériser la coopération et la compétition au sein d'une communauté et d'autre part, de proposer un modèle de sélection de communautés pour permettre l'analyse numérique de communautés d'intérêt.

Notre méthodologie numérique se base sur la connaissance *a priori* du système biologique d'étude ainsi que sur une grande disponibilité de données pour la construction des modèles. Ainsi, notre approche itérative devrait être applicable pour toute communauté pour laquelle ces contraintes sont respectées. En effet, le raffinement de chaque modèle est un processus universel consistant à ajouter, bloquer ou modifier des réactions des réseaux métaboliques. En se basant sur l'analyse de la variabilité des flux et de la connaissance avec *a priori*, les chemins métaboliques utilisés ont pu être déterminés et la cohérence de ces résultats vis à vis de la littérature a été vérifiée. De plus, l'ajustement des mécanismes biologiques, tels que la consommation du lactose ou du lactate, à travers l'optimisation des paramètres, a été déduit de l'étude des courbes de simulations dynamiques. Enfin, nous avons montré qu'il était possible, à partir d'un mécanisme équitable de partage des ressources et d'un calibrage à l'échelle de chaque bactérie, de mettre en avant des interactions bactériennes en communauté et d'obtenir des prédictions numériques satisfaisantes. Ces résultats ont fait l'objet d'un article publié dans le journal *Metabolic Engineering*.

Nous avons par la suite cherché à définir les limites de l'analyse des interactions microbiennes dans le cas le plus pauvre, c'est-à-dire, avec uniquement des données génomiques. En basant notre méthodologie développée sur le principe de raisonnement, les propriétés d'explicabilité des modèles et le passage à l'échelle de communautés de grande taille ont été conservées. Grâce à cette approche d'abstraction booléenne construite comme une extension de celle du logiciel MiSCoTo (Frioux et al., 2018), nous avons caractérisé des potentiels de coopération et de compétition permettant une comparaison de communautés entières et non deux-à-deux. Même si la méthode n'a pas la précision des approches numériques, elle compense par sa capacité de criblage haut-débit, permettant d'effectuer un premier filtre de caractérisation de l'écosystème microbien. Lors de la comparaison avec des méthodes quantitatives, nous avons démontré que nos scores, étaient corrélés avec les

scores des méthodes basées sur les contraintes de flux. De part le faible temps d'exécution de notre approche, ces résultats confortent l'utilisation d'une méthode moins précise mais peu coûteuse pour permettre d'identifier des ensembles considérés comme pertinents du point de vue de la compétition, de la coopération, des substrats limitants ou bien des métabolites échangeables. Ce travail est l'objet d'un article scientifique en cours de rédaction.

Nous avons enfin cherché à enrichir ces modèles logiques du point de vue de l'identification de communautés pertinentes en ajoutant des contraintes fournies par l'utilisateur ou l'utilisatrice. En effet, dans un monde idéal, un modèle hybride dans lequel les formalismes logique et numérique communiquent, permettrait à la fois une sélection de communautés pertinentes et une analyse numérique de ces dernières. La méthode de (Frioux et al., 2018) est une première étape dans la construction d'un tel modèle qui sélectionne uniquement des communautés minimales, en minimisant le nombre d'échanges métaboliques au sein de la communauté. En s'inspirant de cette méthode, nous avons construit un prototype qui, à partir d'un ensemble de réseaux métaboliques, permet de sélectionner une communauté avec des contraintes de taille de composition, ou encore sur le nombre de métabolites cibles devant être produits. Par ailleurs, l'ordre et le choix d'optimisation sont également laissés à l'utilisateur ou l'utilisatrice. Grâce à ce travail, le nombre de communautés trouvées, respectant les règles et les contraintes logiques définies en amont, est réduit et l'utilisation d'un modèle numérique précis et performant sur des communautés de petite taille peut être ainsi utilisé.

Discussion

Discussion informatique

Dans le domaine de la biologie, l'identification des mécanismes sous-jacents expliquant une observation est un défi majeur. D'après Lipton (2017), les modèles dits "boîtes-noire" permettent difficilement une interprétation du résultat du modèle, et ainsi ne correspondent pas à la demande des biologistes. A l'opposé, nous retrouvons des modèles "boîtes blanches" qui, en plus d'identifier des mécanismes responsables de l'observable et d'obtenir une bonne prédiction, permettent d'émettre des hypothèses testables en laboratoire. Nous avons développé deux modèles informatiques répondant aux doubles enjeux des biologistes : bonnes capacités prédictives et pouvoir d'explicabilité. En reprenant la terminologie de Lipton (2017), nos modèles assurent donc la *causalité*, *i.e.* l'explicabilité. Parmi les autres critères de Lipton (2017), nos modèles répondent également à la *transparence*, c'est-à-dire, qu'ils peuvent être auditables par les biologistes. Ce critère traite la question de la manière dont le modèle fonctionne, au niveau de l'algorithme, des paramètres et du modèle dans son ensemble.

Le modèle numérique présenté dans le chapitre 2 s'intéresse à un écosystème microbien étudié expérimentalement pour l'élaboration d'un fromage industriel. Pour rappel, notre stratégie itérative consistait à calibrer chaque souche métabolique pour inférer des potentiels d'interactions en communauté. Avec cette stratégie, nous avons montré que la prédiction du modèle numérique était de bonne qualité. Nos résultats de simulation, à l'échelle de chaque souche métabolique et en communauté, sont numériquement proches des données expérimentales. En raison de cette précision, de la généralité du modèle et de l'utilisation explicite de mécanismes biologiques, nous appelons ce modèle numérique un modèle numérique mécanistique. La *causalité* du modèle est assurée par l'identification de voies métaboliques activées, de flux de consommation et ou de production ainsi que par la contribution de chaque modèle métabolique. Concernant la souche de *L. plantarum*, nous avons observé pour la première fois en lait l'activation de la voie hétérolactique, représenté par un flux non-nul de la voie de la transketolase. De plus, nous avons observé une utilisation du citrate par *L. plantarum* et une préférence d'utilisation du lactose au regard du lactate par *P. freudenreichii*. Notre modèle numérique explique ces résultats par une valeur de flux de consommation de ces composés. Nous avons également implémenté un partage équilibré des ressources : lorsqu'elles deviennent limitantes, *i.e.* en quantité insuffisante pour l'ensemble des organismes, une modification des bornes est faite affinant la prédiction des concentrations des métabolites. Toutes ces explications fournies par le modèle permettent de générer des hypothèses testables expérimentalement, permettant ainsi de qualifier notre modèle numérique de modèle explicable. Enfin, nous assurons la *transparence* en explicitant les noms des métabolites, des réactions et des voies métaboliques ainsi que des gènes qui leur sont associés.

Le modèle discret du chapitre 3 modélise l'interaction métabolique entre organismes au sein d'un écosystème microbien et prédit le potentiel de compétition et de coopération entre ces organismes. Avec l'approche par raisonnement, nous avons formalisé des définitions biologiques sous forme de règles et contraintes logiques : à partir d'un ensemble nutritif nous calculons le potentiel métabolique pour chaque souche et en communauté (*scope*), les échanges métaboliques (*exchanged*) et les substrats en compétition (*polyopsonistic*). Cette formalisation explicite nous permet de calculer des potentiels d'interaction,

compétition et coopération, et garantit une bonne confiance dans les prédictions de notre modèle. Nos prédictions permettent d’obtenir des résultats qui corrélerent avec ceux fournis par les outils numériques existants qui font références, SMETANA et MICOM, pour les mêmes données. Notre logiciel réussit une batterie de tests induits par un benchmark très complet, que nous avons conçu pour valider les prédictions et scores spécifiques d’un écosystème, évaluer le passage à l’échelle et vérifier le criblage haut-débit des communautés.

Tout comme le modèle numérique mécanistique, ce modèle discret assure également la *causalité* : le potentiel de coopération est expliqué par le nombre de métabolites échangés ainsi que par les espèces intervenant dans l’échange. Le potentiel de compétition est quant à lui expliqué par le nombre de consommateurs associé à un substrat limitant. Le proxy de la coopération, représenté par la notion de métabolites échangeables, est expliqué en termes biologiques : un métabolite est échangeable entre deux organismes A et B, s’il est productible par l’organisme A, consommable par l’organisme B et bénéficie à B. De même, la compétition est expliquée en termes biologiques : un substrat nutritionnel est en compétition si ce dernier est limitant et co-consommé. Toutes ces explications issues de notre modèle permettent de générer des hypothèses testables en laboratoire. La production d’un métabolite échangé entre organismes coopérant A et B peut être détectée en métabolomique par chromatographie en phase liquide couplée à la spectrométrie de masse (LCMS) ou par phénotypage biochimique ciblé, et associée avec la présence ou l’absence de l’espèce A. La consommation d’un métabolite par deux organismes en compétition peut être testée de façon similaire.

Enfin, notre modèle assure également la *transparence* car chaque règle logique du modèle est lisible, vérifiable et critiquable par les biologistes. En reprenant la notion biologique de coopération ci-dessus, un métabolite est dit échangé s’il est dans le *scope* d’un producteur, qu’il est réactant d’une réaction d’un consommateur et qu’il n’est pas dans le *scope* individuel du consommateur.

Les modèles explicables développés dans cette thèse ont été élaborés pour répondre à des enjeux biologiques précis. Dans les deux cas, identifier la bonne méthode correspondante à ces enjeux n’a pas été trivial mais la conclusion d’une analyse menée en concertation avec les collègues biologistes. Les critères de prédiction et d’explicabilité des modèles ne permettent pas en soi de choisir entre un modèle numérique et discret.

L’approche hybride développée répond à de multiples enjeux que sont : la résolution de problèmes combinatoires permettant de passer à l’échelle de communautés naturelles, la génération d’hypothèses sur des potentiels interactions bactériennes, l’obtention d’ensembles de métabolites productibles et consommés par plusieurs organismes, et la précision des résultats. Chacun des modèles mécanistiques que nous avons développés permet de répondre à ces enjeux plus ou moins efficacement. Le modèle numérique a démontré sa capacité pour répondre à ces enjeux et est particulièrement adapté pour donner une dimension numérique de l’explicabilité, des *scope* métaboliques et des hypothèses testables. Le modèle par raisonnement est quant à lui pertinent pour sa capacité à déterminer des potentiels d’interaction pour des grandes communautés en résolvant des problèmes combinatoires et de passer ainsi à l’échelle. De plus, ce modèle procure une explication détaillée et rapide du fonctionnement de la communauté : par exemple, l’ensemble des métabolites échangés ainsi que les ensembles de producteurs et de consommateurs intervenant pour l’échange de ce métabolite. En somme, chaque enjeu peut être résolu par un modèle, et donc, un unique modèle hybride n’est pas forcément nécessaire.

Notre étude permet cependant d’évaluer l’opportunité que représente la construction des modèles hybrides pour des écosystèmes microbiens complexes. Quelles auraient été les propriétés d’un modèle hybride, combinant les avantages de la précision numérique

d'un modèle numérique avec celles du passage à l'échelle de communautés naturelles d'un modèle discret ? Dans un souci de satisfaire les contraintes explicitées dans le paragraphe ci-dessus, ce nouveau modèle hybride doit toujours être explicable, doit donner des prédictions numériquement proches des valeurs expérimentales, et doit pouvoir identifier des potentiels d'interaction à grande échelle dans le but de générer des hypothèses testables sur les voies métaboliques activées, les espèces impliquées dans un échange métabolique ou encore celles en réelle compétition.

Deux pistes peuvent être explorées : améliorer le passage à l'échelle du modèle numérique ou enrichir le modèle discret. Nous avons opté pour le second choix en proposant un premier enrichissement temporel. En utilisant la logique temporelle (Cabalar, 2019), il est possible, à l'aide de règles dédiées, d'identifier quand un métabolite est productible. Cet ajout temporel permettrait d'affiner la prédiction du potentiel de compétition : deux espèces sont en compétition si pour un même substrat limitant, au moins ces deux espèces le consomment *au même moment*. Nous avons par la suite permis la sélection à façon de communautés pour la construction d'un design expérimental. Cet ajout permet d'améliorer l'explicabilité du modèle puisque chacun des modèles proposés par l'approche discrète est expliqué par les critères de sélection en amont.

Le modèle numérique représente le but final à atteindre de part sa précision. Pour améliorer son passage à l'échelle, les paramètres inférés pour chaque souche métabolique peuvent être déduits à partir d'un méta-modèle pré-entraîné (Frioux et al., 2023). Dans un second temps, les flux de consommation et de production peuvent être calculés à partir de la concentration des métabolites d'intérêt à chaque point de temps. Enfin, nous pourrions discrétiser le temps et calculer un FBA dans un intervalle de temps plus important, réduisant ainsi le temps de calcul mais augmentant l'approximation entre deux pas de temps.

Discussion biologique

La fermentation bactérienne est au cœur de la fabrication fromagère et elle se caractérise par une acidification du milieu, une production de composés d'arômes et enfin par une croissance bactérienne. Les échanges de métabolites entre espèces favorisent le fonctionnement de l'écosystème.

L'approche discrète du chapitre 3 peut se révéler utile pour identifier des hypothèses biologiques, en lien avec la fermentation bactérienne par exemple. Comparer des ensembles de métabolites produits et consommés, par entité biologique au sein d'un écosystème est une solution. Parmi ces ensembles, nous avons calculé les ensembles de métabolites : (i) échangeables entre différentes souches, (ii) déjà présents dans le milieu extra-cellulaire, (iii) nouvellement disponibles, libérables au cours d'une fermentation bactérienne par exemple, (iv) pouvant être limitants. L'analyse comparée de ces ensembles a permis de prédire les espèces en compétition, les échanges et les souches impliquées dans les échanges. Ces échanges peuvent impacter la croissance des bactéries, la production de composés d'intérêts et le pH. En plus de l'identification d'interactions en comparant des ensembles, nous avons calculé des potentiels de coopération et de compétition, ce qui ajoute une dimension pseudo-quantitative à la caractérisation et permet de comparer des communautés de divers écosystèmes entre eux.

La dimension quantitative d'une interaction bactérienne est apportée par le modèle numérique mécanistique du chapitre 2. Cette dimension est caractérisée par la prédiction de la concentration des métabolites, et la quantité de biomasse produite correspondant au taux de croissance de l'espèce. Les interactions sont identifiées en analysant les flux de métabolites (i) entrant dans la cellule bactérienne et (ii) sortant vers le milieu extracel-

lulaire. Par exemple, à l'échelle de chaque cellule, tous les modèles métaboliques ont un flux de consommation du lactose non nul, suggérant une compétition pour ce dernier. Le lactate est produit par les bactéries lactiques et consommé par *P. freudenreichii*, ce qui est symbolisé par des flux de lactate sortant et entrant respectivement. Ces mécanismes révèlent une coopération entre les bactéries lactiques, *L. plantarum* et *L. lactis*, et la bactérie propionique *P. freudenreichii*. Cette analyse est faite grâce aux méthodes d'analyse de flux (FBA, FVA) (Mahadevan and Schilling, 2003; Orth et al., 2010) sous l'hypothèse d'état stationnaire et de maximisation de la biomasse. Il est possible de faire varier cette fonction objective, par exemple maximisation de la synthèse d'ATP ou de lactate, la distribution des flux changerait et donc, de nouvelles hypothèses d'interactions verraient le jour.

Ces interactions ont des impacts cinétiques sur la concentration des composés d'arômes et sur la densité bactérienne. Au cours du temps, la concentration et la densité bactérienne varient, et une croissance nulle ou ralentie d'une souche peut-être expliquée par une compétition pour un substrat ou par un précurseur limitant, au sens biologique. La production d'un composé d'arôme peut également être ralentie ou arrêtée, suggérant une compétition. Ces hypothèses ont été mises en avant par une analyse dynamique de l'équilibre des flux (dFBA) (Mahadevan et al., 2002a) et a permis de proposer des valeurs de contribution relative de chaque espèce à la consommation et la production de composés d'intérêt. Nous avons ainsi suggéré des producteurs et des consommateurs pour chaque composé pouvant être testé en laboratoire. La validation biologique pourrait être réalisée par un knockout des gènes correspondants. En plus de ces impacts, nous avons pu déduire quelles sont les espèces participant le plus à l'acidification du milieu grâce à notre modèle de pH : *i.e* *L. lactis* au début de la fabrication et *P. freudenreichii* pendant l'affinage.

Les deux approches, discrète et numérique que nous avons développées dans cette thèse, ont permis de générer des hypothèses testables, affinant la compréhension des interactions qui impactent la croissance bactérienne et/ou la production de métabolites. À partir de ces hypothèses, nous discuterons ci-dessous des expérimentations biologiques permettant de valider ou d'invalider les hypothèses afin d'améliorer les modèles explicatifs. Nous traiterons uniquement le cas où des génomes annotés et les souches correspondantes sont à notre disposition.

Les approches numérique et discrète que nous avons développées ont permis de détecter des interactions bactériennes. Dans le cas d'un échange de métabolites entre deux souches, nous avons, pour les deux approches, les bactéries productrices et consommatrices ainsi que le sens de l'échange. Par exemple, le lactate, mis en avant par notre implémentation du dFBA, est produit par les bactéries lactiques et consommé par *P. freudenreichii*, la phénylalanine, révélée par SMETANA (Zelezniak et al., 2015) et MiCOM (Diener et al., 2020), serait produite par *L. plantarum* et consommée par *P. freudenreichii*. Ces deux approches permettent également de nous indiquer des souches co-consommatrices de substrats qui peuvent conduire à une compétition entre souches si ces substrats devenaient limitants. Par exemple, le lactose est co-consommé par la communauté, ou encore la glycine, est produite par *L. plantarum* et co-consommée par *L. lactis* et *P. freudenreichii*. Toutes ces interactions révélées *in silico* avec des outils *a priori* ou *sans a priori* peuvent être validées expérimentalement au moyen d'une comparaison entre des co-cultures et des cultures pures. Afin de vérifier l'impact de ces interactions, la comparaison porterait sur la croissance des bactéries ainsi que sur le dosage de métabolites.

Dans le cas d'une hypothèse de co-consommation d'un substrat, une décroissance

bactérienne et/ou une diminution de la concentration de ce métabolite devraient être observées en co-culture. Afin de vérifier les souches impliquées, un marquage isotopique à froid de ce métabolite suivi d'une spectroscopie avec la méthode Raman (Cui et al., 2022; Wang et al., 2020) permettraient d'identifier les consommateurs de ce métabolite. Les activations et les inhibitions de productions métaboliques peuvent être mises en avant avec les données métatranscriptomiques ou métaprotéomiques. Pour révéler expérimentalement ces échanges d'intermédiaires métaboliques (1), il faut valider indépendamment la production de ce métabolite par l'espèce productrice en culture pure et la consommation par l'espèce consommatrice avec un marquage isotopique.

Retour d'expérience sur l'interdisciplinarité de la thèse

Au cours de la collaboration établie avec les biologistes dans cette thèse, nous avons pu mettre en évidence deux façons d'exploiter les modèles développés. Dans un premier temps, seul le *résultat* du modèle est important afin de vérifier expérimentalement les prédictions. Par exemple, dans le chapitre 2, nous avons mis en avant de nouvelles interactions métaboliques potentielles, à la fois avec notre modèle avec *a priori* et les outils *sans a priori*, pouvant être validées ou réfutées grâce aux expérimentations que nous avons mentionnées plus haut. Dans un second temps, pour une ré-exploitation des résultats de la thèse, la propriété de *généricité* des outils et du savoir-faire est importante. Cette généralité permet l'utilisation pour une autre question de recherche ou avec des données d'entrées différentes. Par exemple, les outils développés des chapitres 3 et 4 ont permis l'utilisation de réseaux métaboliques appartenant à n'importe quel écosystème, et permettent de s'intéresser à la compétition, coopération, sélection de communautés selon différents critères biologiques etc. Par ailleurs, le degré de *diffusion* des logiciels développés est un facteur à prendre en compte pour un travail interdisciplinaire. En effet, le déploiement de nos outils sur une plateforme, telle que Galaxy (Community, 2022) par exemple, permettrait de garantir un accès simplifié pour les biologistes et d'approfondir leurs recherches sur la compréhension des interactions bactériennes.

Perspectives

Avec ces trois chapitres, nous avons contribué à l’avancement des analyses du métabolisme des communautés microbiennes sur le plan numérique et discret. Cependant plusieurs points d’amélioration subsistent. À court terme, valoriser davantage l’intégration des données métatranscriptomiques serait intéressant. En effet, dans le chapitre 2, nous avons utilisé ces données hétérogènes pour valider des interactions mises en évidence par SMETANA en étudiant l’expression des gènes associés à la production et à la consommation des métabolites présumés échangés. Il serait pertinent de tester de contraindre les réseaux métaboliques par les données méta-transcriptomiques avec différentes méthodes (Agren et al., 2012; Becker and Palsson, 2008; Jenior et al., 2020) et de comparer les résultats à nos données simulées au niveau des voies métaboliques exprimées ainsi qu’aux données métabolomiques. De plus, nous pourrions nous demander si seules les régulations des expressions géniques suffisent pour obtenir un modèle numérique précis, ce qui aurait pour conséquence de réduire le temps passé au raffinement et à la calibration de modèles métaboliques. Ces données pourraient également servir de contrainte *a priori* pour les modèles discrets du chapitre 3 et 4. En effet, inférer des règles logiques sur l’expression de voies métaboliques connues peut changer les potentiels d’interactions de communautés microbiennes et réduire la combinatoire.

À moyen terme, la construction d’un modèle hybride des écosystèmes microbiens pourrait être faite. Même si nous avons conclu qu’un modèle hybride n’est pas nécessaire, ce type d’approche pourrait néanmoins apporter une plus-value. En effet, les métabolites suivis en dynamique dans le modèle numérique du chapitre 2 proviennent d’une question biologique. Dans Bellazzi et al. (2001), le modèle structurel alimente le modèle numérique en données permettant d’enrichir l’analyse. Dans notre cas, le couplage de notre modèle numérique à une approche de criblage, comme le modèle logique du chapitre 3, permettrait de fournir des métabolites d’intérêt en entrée. Une difficulté mise en évidence, est le grand nombre de composés échangeables détectés par une telle approche. Dans le but de filtrer ces solutions, un premier apport du modèle numérique serait de vérifier la faisabilité de production des composés identifiés par le modèle logique. Cette faisabilité pourrait être transmise au modèle logique comme une donnée d’entrée sous forme de règles et ou contraintes nouvelles. Nous aurions à terme, un modèle hybride couplé, dans lequel le formalisme numérique et le formalisme discret communiquent.

Enfin à plus long terme, la construction d’un modèle de jumeau numérique du processus de la fermentation bactérienne est une perspective attrayante. En effet, dans les travaux de Udugama et al. (2021), les auteurs construisent un modèle de jumeau numérique et utilisent la fermentation de l’éthanol pour démontrer l’efficacité de leur modèle. Selon la définition de L’INRAE, un jumeau numérique est défini comme “une représentation numérique de l’objet ou du système d’intérêt (physique ou biologique; e.g. cellule, tissu, plante, animal), incluant des variables décrivant son état et son évolution de façon dynamique”¹. Nous avons des données d’entrée dynamiques sur le processus de fermentation (métatranscriptomiques, métabolomiques et cinétiques), un algorithme de modélisation et de simulation. Une caractérisation manquante est la boucle de rétroaction du modèle sur l’état du processus modélisé : *i.e.* une prise de décision.

1. <https://digitbio.hub.inrae.fr/thematiques/vers-le-jumeau-numerique#Jumeaux>

Références

Bibliographie

- [1] Abedi, E. and Hashemi, S. M. B. (2020). Lactic acid production – producing microorganisms and substrates sources-state of art. *Heliyon*, 6(10) :e04974.
- [2] Agren, R., Bordel, S., Mardinoglu, A., Pornputtpong, N., Nookaew, I., and Nielsen, J. (2012). Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLoS Computational Biology*, 8(5).
- [3] Aite, M., Chevallier, M., Frioux, C., Trottier, C., Got, J., Cortés, M. P., Mendoza, S. N., Carrier, G., Dameron, O., Guillaudeux, N., Latorre, M., Loira, N., Markov, G. V., Maass, A., and Siegel, A. (2018). Traceability, reproducibility and wiki-exploration for “à-la-carte” reconstructions of genome-scale metabolic models. *PLoS Computational Biology*, 14(5) :1–25.
- [4] Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3) :403–410.
- [5] Arabzadeh, M., Saheb Zamani, M., Sedighi, M., and Marashi, S. A. (2018). A graph-based approach to analyze flux-balanced pathways in metabolic networks. *BioSystems*, 165 :40–51.
- [6] Arkin, A. P., Cottingham, R. W., Henry, C. S., Harris, N. L., Stevens, R. L., Maslov, S., Dehal, P., Ware, D., Perez, F., Canon, S., Sneddon, M. W., Henderson, M. L., Riehl, W. J., Murphy-Olson, D., Chan, S. Y., Kamimura, R. T., Kumari, S., Drake, M. M., Brettin, T. S., Glass, E. M., Chivian, D., Gunter, D., Weston, D. J., Allen, B. H., Baumohl, J., Best, A. A., Bowen, B., Brenner, S. E., Bun, C. C., Chandonia, J. M., Chia, J. M., Colasanti, R., Conrad, N., Davis, J. J., Davison, B. H., DeJongh, M., Devoid, S., Dietrich, E., Dubchak, I., Edirisinghe, J. N., Fang, G., Faria, J. P., Frybarger, P. M., Gerlach, W., Gerstein, M., Greiner, A., Gurtowski, J., Haun, H. L., He, F., Jain, R., Joachimiak, M. P., Keegan, K. P., Kondo, S., Kumar, V., Land, M. L., Meyer, F., Mills, M., Novichkov, P. S., Oh, T., Olsen, G. J., Olson, R., Parrello, B., Pasternak, S., Pearson, E., Poon, S. S., Price, G. A., Ramakrishnan, S., Ranjan, P., Ronald, P. C., Schatz, M. C., Seaver, S. M., Shukla, M., Sutormin, R. A., Syed, M. H., Thomason, J., Tintle, N. L., Wang, D., Xia, F., Yoo, H., Yoo, S., and Yu, D. (2018). KBase : The United States department of energy systems biology knowledgebase. *Nature Biotechnology*, 36(7) :566–569.
- [7] Ay, F., Dang, M., and Kahveci, T. (2012). Metabolic network alignment in large scale by network compression. *BMC Bioinformatics*, 13(SUPPL.3) :1–19.
- [8] Aziz, R. K., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., Pusch,

- G. D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., and Zagnitko, O. (2008). The RAST Server : Rapid annotations using subsystems technology. *BMC Genomics*, 9 :1–15.
- [9] Bai, Y., Müller, D. B., Srinivas, G., Garrido-Oter, R., Potthoff, E., Rott, M., Dombrowski, N., Münch, P. C., Spaepen, S., Remus-Emsermann, M., Hüttel, B., McHardy, A. C., Vorholt, J. A., and Schulze-Lefert, P. (2015). Functional overlap of the Arabidopsis leaf and root microbiota. *Nature*, 528(7582) :364–369.
- [10] Bakken, L. R. and Olsen, R. A. (1983). Buoyant densities and dry-matter contents of microorganisms : Conversion of a measured biovolume into biomass. *Applied and Environmental Microbiology*.
- [11] Baldini, F., Heinken, A., Heirendt, L., Magnusdottir, S., Fleming, R. M., and Thiele, I. (2019). The microbiome modeling toolbox : From microbial interactions to personalized microbial communities. *Bioinformatics*, 35(13) :2332–2334.
- [12] Barabási, A.-L. and Oltvai, Z. N. (2004). Network biology : understanding the cell’s functional organization. *Nature Reviews Genetics*, 5(2) :101–113.
- [13] Becker, S. A. and Palsson, B. O. (2008). Context-Specific Metabolic Networks Are Consistent with Experiments. *PLoS Computational Biology*, 4(5) :e1000082.
- [14] Belcour, A., Frioux, C., Aite, M., Bretaudeau, A., Hildebrand, F., and Siegel, A. (2020). Metage2metabo, microbiota-scale metabolic complementarity for the identification of key species. *eLife*.
- [15] Bellazzi, R., Guglielmann, R., Ironi, L., and Patrini, C. (2001). A hybrid input-output approach to model metabolic systems : An application to intracellular thiamine kinetics. *Journal of Biomedical Informatics*, 34(4) :221–248.
- [16] Bergmann, F. T., Czauderna, T., Dorusoz, U., Rougny, A., Dräger, A., Touré, V., Mazein, A., Blinov, M. L., and Luna, A. (2020). Systems biology graphical notation markup language (SBGNML) version 0.3. *Journal of Integrative Bioinformatics*, (0) :20200016. Place : Berlin, Boston Publisher : De Gruyter.
- [17] Berry, D. and Widder, S. (2014). Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Frontiers in Microbiology*, 5(MAY) :1–14.
- [18] Biggs, M. B., Medlock, G. L., Kolling, G. L., and Papin, J. A. (2015). Metabolic network modeling of microbial communities. *Wiley Interdisciplinary Reviews : Systems Biology and Medicine*, 7(5) :317–334.
- [19] Blasche, S., Kim, Y., Mars, R. A. T., Machado, D., Maansson, M., Kafkia, E., Milanesi, A., Zeller, G., Teusink, B., Nielsen, J., Benes, V., Neves, R., Sauer, U., and Patil, K. R. (2021). Metabolic cooperation and spatiotemporal niche partitioning in a kefir microbial community. *Nature Microbiology*, 6(2) :196–208.
- [20] Boele, J. and Olivier, B. G. (2012). MetabolicNetwork_Modeling_Server FAME , the Flux Analysis and Modeling Environment. *BMC Systems Biology*, 6(8).
- [21] Borghei, R., Mobasheri, M., and Sobati, T. (2021). Genome Scale Constraint-Based Metabolic Reconstruction of Propionibacterium Freudenreichii DSM 20271. *Preprint from Research square*, pages 1–29.

- [22] Brook, B. S. and Waters, S. L. (2008). Mathematical challenges in integrative physiology. *Journal of Mathematical Biology*, 56(6) :893–896.
- [23] Cabalar, P. (2019). `telingo = ASP + Time`. *LNCS*, pages 1–13.
- [24] Cai, J., Tan, T., and Chan, S. H. (2020). Predicting Nash equilibria for microbial metabolic interactions. *Bioinformatics*, 36(24) :5649–5655.
- [25] Cantalapiedra, C. P., Hernandez-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). `eggNOG-mapper v2` : Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution*, 38(12) :5825–5829.
- [26] Cao, W., Aubert, J., Maillard, M. B., Boissel, F., Leduc, A., Thomas, J. L., Deutsch, S. M., Camier, B., Kerjough, A., Parayre, S., Harel-Oger, M., Garric, G., Thierry, A., and Falentin, H. (2021). Fine-Tuning of Process Parameters Modulates Specific Metabolic Bacterial Activities and Aroma Compound Production in Semi-Hard Cheese. *Journal of Agricultural and Food Chemistry*, 69(30) :8511–8529.
- [27] Carr, R. and Borenstein, E. (2012). NetSeed : a network-based reverse-ecology tool for calculating the metabolic interface of an organism with its environment. 28(5) :734–735.
- [28] Carroll, N. M., Ross, R. P., Kelly, S. M., Price, N. C., Sheehan, D., and Cogan, T. M. (1999). Characterization of recombinant acetolactate synthase from *Leuconostoc lactis* NCW1. *Enzyme and Microbial Technology*, 25(1-2) :61–67.
- [29] Caspi, R., Billington, R., Fulcher, C. A., Keseler, I. M., Kothari, A., Krummenacker, M., Latendresse, M., Midford, P. E., Ong, Q., Ong, W. K., Paley, S., Subhraveti, P., and Karp, P. D. (2018). The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Research*, 46(D1) :D633–D639.
- [30] Cerk, K., Ugalde-Salas, P., Nedjad, C. G., Lecomte, M., Muller, C., Sherman, D. J., Hildebrand, F., Labarthe, S., and Frioux, C. (2024). Community-scale models of microbiomes : Articulating metabolic modelling and metagenome sequencing. *Microbial Biotechnology*, 17(1) :e14396. e14396 MICROBIO-2023-392.R1.
- [31] Chan, S. H. J., Simons, M. N., and Maranas, C. D. (2017). SteadyCom : Predicting microbial abundances while ensuring community stability. *PLoS Computational Biology*, 13(5) :1–25.
- [32] Chang, A., Jeske, L., Ulbrich, S., Hofmann, J., Koblit, J., Schomburg, I., Neumann-Schaal, M., Jahn, D., and Schomburg, D. (2021). BRENDA, the ELIXIR core data resource in 2021 : New developments and updates. *Nucleic Acids Research*, 49(D1) :D498–D508.
- [33] Chiu, H. C., Levy, R., and Borenstein, E. (2014). Emergent Biosynthetic Capacity in Simple Microbial Communities. *PLoS Computational Biology*, 10(7).
- [34] Choi, J., Yang, F., Stepanauskas, R., Cardenas, E., Garoutte, A., Williams, R., Flater, J., Tiedje, J. M., Hofmockel, K. S., Gelder, B., and Howe, A. (2017). Strategies to improve reference databases for soil microbiomes. *ISME Journal*, 11(4) :829–834.

- [35] Colarusso, A. V., Goodchild-Michelman, I., Rayle, M., and Zomorodi, A. R. (2021). Computational modeling of metabolism in microbial communities on a genome-scale. *Current Opinion in Systems Biology*, 26 :46–57.
- [36] Collet, G., Eveillard, D., Gebser, M., Prigent, S., Schaub, T., Siegel, A., and Thiele, S. (2013). Extending the metabolic network of *ectocarpus siliculosus* using answer set programming. In Cabalar, P. and Son, T. C., editors, *Logic Programming and Nonmonotonic Reasoning*, pages 245–256, Berlin, Heidelberg. Springer Berlin Heidelberg.
- [37] Community, T. G. (2022). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses : 2022 update. *Nucleic Acids Research*, 50(W1) :W345–W351.
- [38] Corblin, F., Bordeaux, L., Hamadi, Y., Fanchon, E., and Trilling, L. (2006). A SAT-based approach to decipher Gene Regulatory Networks. *Gene*, pages 1088–1091.
- [39] Costa, R. S., Hartmann, A., and Vinga, S. (2016). Kinetic modeling of cell metabolism for microbial production. *Journal of Biotechnology*, 219 :126–141.
- [40] Crow, V. L. (1986). Metabolism of aspartate by *propionibacterium freudenreichii* subsp. *shermanii* : Effect on lactate fermentation. *Applied and Environmental Microbiology*, 52(2) :359–365.
- [41] Cruz, D. A. and Kemp, M. L. (2022). Hybrid computational modeling methods for systems biology. *Progress in Biomedical Engineering*, 4(1).
- [42] Cui, D., Kong, L., Wang, Y., Zhu, Y., and Zhang, C. (2022). In situ identification of environmental microorganisms with raman spectroscopy. *Environmental Science and Ecotechnology*, 11 :100187.
- [43] Dank, A., van Mastrigt, O., Boeren, S., Lillevang, S. K., Abee, T., and Smid, E. J. (2021). *Propionibacterium freudenreichii* thrives in microaerobic conditions by complete oxidation of lactate to CO₂. *Environmental Microbiology*, 23(6) :3116–3129.
- [44] Deborde, C. and Boyaval, P. (2000). Interactions between pyruvate and lactate metabolism in *Propionibacterium freudenreichii* subsp. *shermanii* : In vivo ¹³C nuclear magnetic resonance studies. *Applied and Environmental Microbiology*, 66(5) :2012–2020.
- [45] Delcher, A. L., Bratke, K. A., Powers, E. C., and Salzberg, S. L. (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics*, 23(6) :673–679.
- [46] Demir, E., Cary, M. P., Paley, S., Fukuda, K., Lemer, C., Vastrik, I., Wu, G., D’Eustachio, P., Schaefer, C., Luciano, J., Schacherer, F., Martinez-Flores, I., Hu, Z., Jimenez-Jacinto, V., Joshi-Tope, G., Kandasamy, K., Lopez-Fuentes, A. C., Mi, H., Pichler, E., Rodchenkov, I., Splendiani, A., Tkachev, S., Zucker, J., Gopinath, G., Rajasimha, H., Ramakrishnan, R., Shah, I., Syed, M., Anwar, N., Babur, Ö., Blinov, M., Brauner, E., Corwin, D., Donaldson, S., Gibbons, F., Goldberg, R., Hornbeck, P., Luna, A., Murray-Rust, P., Neumann, E., Reubenacker, O., Samwald, M., Van Iersel, M., Wimalaratne, S., Allen, K., Braun, B., Whirl-Carrillo, M., Cheung, K. H., Dahlquist, K., Finney, A., Gillespie, M., Glass, E., Gong, L., Haw, R., Honig, M., Hubaut, O., Kane, D., Krupa, S., Kutmon, M., Leonard, J., Marks, D., Merberg, D., Petri, V., Pico, A., Ravenscroft,

- D., Ren, L., Shah, N., Sunshine, M., Tang, R., Whaley, R., Letovksy, S., Buetow, K. H., Rzhetsky, A., Schachter, V., Sobral, B. S., Dogrusoz, U., McWeeney, S., Aladjem, M., Birney, E., Collado-Vides, J., Goto, S., Hucka, M., Novère, N. L., Maltsev, N., Pandey, A., Thomas, P., Wingender, E., Karp, P. D., Sander, C., and Bader, G. D. (2010). The BioPAX community standard for pathway data sharing. *Nature Biotechnology*, 28(9) :935–942.
- [47] Dias, O., Rocha, M., Ferreira, E. C., and Rocha, I. (2015). Reconstructing genome-scale metabolic models with merlin. *Nucleic acids research*, 43(8) :3899–3910.
- [48] Diener, C., Gibbons, S. M., and Resendis-Antonio, O. (2020). MICOM : Metagenome-Scale Modeling To Infer Metabolic Interactions in the Gut Microbiota. *mSystems*, 5(1).
- [49] Ebenhöf, O., Handorf, T., and Heinrich, R. (2004). Structural analysis of expanding metabolic networks. *Genome informatics. International Conference on Genome Informatics*, 15(1) :35–45.
- [50] F., V. P. (1838). Notice sur la loi que la population suit dans son accroissement. *Correspondence mathématique et physique*, 10 :113–129.
- [51] Faust, K. and Raes, J. (2012). Microbial interactions : From networks to models. *Nature Reviews Microbiology*, 10(8) :538–550.
- [52] Fazzino, L., Anisman, J., Chacón, J. M., Heineman, R. H., and Harcombe, W. R. (2020). Lytic bacteriophage have diverse indirect effects in a synthetic cross-feeding community. *ISME Journal*, 14(1) :123–134.
- [53] Feist, A. M. and Palsson, B. O. (2010). The biomass objective function. *Current Opinion in Microbiology*, 13(3) :344–349. Ecology and industrial microbiology * Special section : Systems biology.
- [54] Fisher, J. and Piterman, N. (2010). The executable pathway to biological networks. *Briefings in Functional Genomics and Proteomics*, 9(1) :79–92.
- [55] Fournier, P., Pellan, L., Barroso-Bergadà, D., Bohan, D. A., Candresse, T., Delmotte, F., Dufour, M.-C., Lauvergeat, V., Le Marrec, C., Marais, A., Martins, G., Masneuf-Pomarède, I., Rey, P., Sherman, D., This, P., Frioux, C., Labarthe, S., and Vacher, C. (2022). Chapter two - the functional microbiome of grapevine throughout plant evolutionary history and lifetime. In Bohan, D. A. and Dumbrell, A., editors, *Functional Microbiomes*, volume 67 of *Advances in Ecological Research*, pages 27–99. Academic Press.
- [56] Freilich, S., Zarecki, R., Eilam, O., Segal, E. S., Henry, C. S., Kupiec, M., Gophna, U., Sharan, R., and Ruppin, E. (2011). Competitive and cooperative metabolic interactions in bacterial communities. *Nature Communications*, 2(1).
- [57] Frioux, C., Fremy, E., Trottier, C., and Siegel, A. (2018). Scalable and exhaustive screening of metabolic functions carried out by microbial consortia. *Bioinformatics*, 34(17) :i934–i943.
- [58] Frioux, C., Huet, S., Labarthe, S., Martinelli, J., Malou, T., Sherman, D., Taupin, M.-L., and Ugalde-Salas, P. (2023). Accelerating metabolic models evaluation with statistical metamodels : application to Salmonella infection models . *ESAIM : Proceedings and Surveys*, 73(September) :187–217.

- [59] Frioux, C., Schaub, T., Schellhorn, S., Siegel, A., and Wanko, P. (2019). Hybrid metabolic network completion. *Theory and Practice of Logic Programming*, 19(1) :83–108.
- [60] García-Jiménez, B., García, J. L., and Nogales, J. (2018). FLYCOP : Metabolic modeling-based analysis and engineering microbial communities. *Bioinformatics*, 34(17) :i954–i963.
- [61] Gebser, M., Kaminski, R., Kaufmann, B., and Schaub, T. (2019). Multi-shot ASP solving with clingo. *Theory and Practice of Logic Programming*, 19(1) :27–82.
- [62] Goelzer, A., Muntel, J., Chubukov, V., Jules, M., Prestel, E., Nölker, R., Mariadasou, M., Aymerich, S., Hecker, M., Noirot, P., Becher, D., and Fromion, V. (2015). Quantitative prediction of genome-wide resource allocation in bacteria. *Metabolic Engineering*, 32 :232–243.
- [63] Goldford, J. E. and Segrè, D. (2018). Modern views of ancient metabolic networks. *Current Opinion in Systems Biology*, 8 :117–124.
- [64] Gomez, J. A. and Barton, P. I. (2018). Dynamic flux balance analysis using DFBalab. *Methods in Molecular Biology*, 1716 :353–370.
- [65] Gomez, J. A., Höffner, K., and Barton, P. I. (2014). Dfbalab : a fast and reliable matlab code for dynamic flux balance analysis. *BMC bioinformatics*, 15(1) :1–10.
- [66] Harcombe, W. R., Riehl, W. J., Dukovski, I., Granger, B. R., Betts, A., Lang, A. H., Bonilla, G., Kar, A., Leiby, N., Mehta, P., Marx, C. J., and Segrè, D. (2014). Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial dynamics. *Cell Reports*, 7(4) :1104–1115.
- [67] Heiden, J. V., Wernisch, L., Gilbert, D., and Wodak, S. J. (2002). Graph-Based Analysis of Metabolic Networks.
- [68] Henry, C. S., Bernstein, H. C., Weisenhorn, P., Taylor, R. C., Lee, J. Y., Zucker, J., and Song, H. S. (2016). Microbial Community Metabolic Modeling : A Community Data-Driven Network Reconstruction. *Journal of Cellular Physiology*, 231(11) :2339–2345.
- [69] Holme, P. (2011). Metabolic robustness and network modularity : A model study. *PLoS ONE*, 6(2) :1–6.
- [70] Hoorman, J. J. (2011). The Role of Soil Bacteria. *The Ohio State University Extension*, pages 1–4.
- [71] Hucka, M., Bergmann, F. T., Chaouiya, C., Dräger, A., Hoops, S., Keating, S. M., König, M., Novère, N. L., Myers, C. J., Olivier, B. G., Sahle, S., Schaff, J. C., Sheriff, R., Smith, L. P., Waltemath, D., Wilkinson, D. J., and Zhang, F. (2019). *The Systems Biology Markup Language (SBML) : Language Specification for Level 3 Version 2 Core Release 2*, volume 16.
- [72] Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernandez-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., Von Mering, C., and Bork, P. (2019). EggNOG 5.0 : A hierarchical, functionally and phylogenetically annotated

- orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, 47(D1) :D309–D314.
- [73] Hyatt, D., Chen, G.-L., LoCasio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal : prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, 11(1) :119.
- [74] Issa, R. and Sherman, D. J. (2018). Deriving Gene Associations for Metabolic Models. Research Report RR-9246, Inria Bordeaux Sud-Ouest.
- [75] Janhunen, T., Kaminski, R., Ostrowski, M., Schellhorn, S., Wanko, P., and Schaub, T. (2017). Clingo goes linear constraints over reals and integers. *Theory and Practice of Logic Programming*, 17(5-6) :872–888.
- [76] Junior, M. L., Moutinho, T. J., Dougherty, B. V., and Papin, J. A. (2020). Transcriptome-guided parsimonious flux analysis improves predictions with metabolic networks in complex environments. *PLoS Computational Biology*, 16(4).
- [77] Jeong, H., Tombort, B., Albert, R., Oltvai, Z. N., and Barabasi, A. L. (2011). The large-scale organization of metabolic networks. *The Structure and Dynamics of Networks*, 9781400841356(1990) :211–214.
- [78] Kanehisa, M. and Goto, S. (2000). KEGG : Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28(1) :27–30.
- [79] Karlsen, E., Schulz, C., and Almaas, E. (2018). Automated generation of genome-scale metabolic draft reconstructions based on KEGG. *BMC Bioinformatics*, 19(1) :1–11.
- [80] Karp, P. D., Billington, R., Caspi, R., Fulcher, C. A., Latendresse, M., Kothari, A., Keseler, I. M., Krummenacker, M., Midford, P. E., Ong, Q., Ong, W. K., Paley, S. M., and Subhraveti, P. (2018). The BioCyc collection of microbial genomes and metabolic pathways. *Briefings in Bioinformatics*, 20(4) :1085–1093.
- [81] Karp, P. D., Paley, S., Krummenacker, M., Kothari, A., Wannemuehler, M. J., and Phillips, G. J. (2022). Pathway Tools Management of Pathway/Genome Data for Microbial Communities. *Frontiers in Bioinformatics*, 2(April) :1–11.
- [82] Khandelwal, R. A., Olivier, B. G., Röling, W. F., Teusink, B., and Bruggeman, F. J. (2013). Community Flux Balance Analysis for Microbial Consortia at Balanced Growth. *PLoS ONE*, 8(5).
- [83] Kim, E. Y., Ashlock, D., and Yoon, S. H. (2019). Identification of critical connectors in the directed reaction-centric graphs of microbial metabolic networks. *BMC Bioinformatics*, 20(1) :1–13.
- [84] King, Z. A., Dräger, A., Ebrahim, A., Sonnenschein, N., Lewis, N. E., and Palsson, B. O. (2015). Escher : A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways. *PLoS Computational Biology*.
- [85] King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., Ebrahim, A., Palsson, B. O., and Lewis, N. E. (2016). BiGG Models : A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research*.

- [86] Kitano, H. (2002). Systems biology : A brief overview. *Science*, 295(5560) :1662–1664.
- [87] Klamt, S., Bettenbrock, K., Schuster, S., and Gilles, E. D. (2002). 58345 151..202. 420(November) :1–4.
- [88] Kleerebezem, M., Boekhorst, J., Van Kranenburg, R., Molenaar, D., Kuipers, O. P., Leer, R., Turchini, R., Peters, S. A., Sandbrink, H. M., Fiers, M. W., Stiekema, W., Klein Lankhorst, R. M., Bron, P. A., Hoffer, S. M., Nierop Groot, M. N., Kerkhoven, R., De Vries, M., Ursing, B., De Vos, W. M., and Siezen, R. J. (2003). Complete genome sequence of *Lactobacillus plantarum* WCFS1. *Proceedings of the National Academy of Sciences of the United States of America*, 100(4) :1990–1995.
- [89] Klitgord, N. and Segrè, D. (2010). Environments that induce synthetic microbial ecosystems. *PLOS Computational Biology*, 6(11) :1–17.
- [90] Koch, S., Benndorf, D., Fronk, K., Reichl, U., and Klamt, S. (2016). Predicting compositions of microbial communities from stoichiometric models with applications for the biogas process. *Biotechnology for Biofuels*, 9(1) :1–16.
- [91] Kreimer, A., Doron-Faigenboim, A., Borenstein, E., and Freilich, S. (2012). NetCmpt : A network-based tool for calculating the metabolic competition between bacterial species. *Bioinformatics*.
- [92] Kuffner, R., R, Z., and T, L. (2000). Pathway analysis in metabolic databases via differential metabolic display (DMD). *Bioinformatics*, 16(9) :825–836.
- [93] Lacroix, V., Fernandes, C. G., and Sagot, M. F. (2006). Motif search in graphs : Application to metabolic networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 3(4) :360–368.
- [94] Lecomte, M., Cao, W., Aubert, J., Sherman, D. J., Falentin, H., Frioux, C., and Labarthe, S. (2024). Revealing the dynamics and mechanisms of bacterial interactions in cheese production with metabolic modelling. *Metabolic Engineering*, 83 :24–38.
- [95] Lecomte, M., Labarthe, S., Sherman, D. J., Falentin, H., and Frioux, C. (2022a). Characterizing microbial interactions in controlled and natural microbial communities. In *Workshop SymBioDiversity*, Santiago, Chile.
- [96] Lecomte, M., Sherman, D. J., Falentin, H., and Frioux, C. (2022b). Characterising competition and cooperation potentials in microbial communities using discrete models of metabolism. An exploratory analysis. JOBIM 2022 - JOURNÉES OUVERTES EN BIOLOGIE, INFORMATIQUE ET MATHÉMATIQUES. Poster.
- [97] Lecomte, M., Sherman, D. J., Falentin, H., and Frioux, C. (2022c). Reasoning approaches for the characterization of cooperation and competition in large-scale microbial communities. In *Journée GT Bioss 2022*, Nantes, France.
- [98] Lemke, N., Herédia, F., Barcellos, C. K., dos Reis, A. N., and Mombach, J. C. (2004). Essentiality and damage in metabolic networks. *Bioinformatics*, 20(1) :115–119.
- [99] Letunic, I. and Bork, P. (2007). Interactive Tree Of Life (iTOL) : An online tool for phylogenetic tree display and annotation. *Bioinformatics*, 23(1) :127–128.

- [100] Levesque, H. J. (1986). Knowledge representation and reasoning. *Annual Review of Computer Science*, 1(1) :255–287.
- [101] Levy, R. and Borenstein, E. Reverse Ecology : From Systems to Environments and Back. pages 329–345.
- [102] Levy, R. and Borenstein, E. (2013). Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. *Proceedings of the National Academy of Sciences*, 110(31) :12804 LP – 12809.
- [103] Levy, R., Carr, R., Kreimer, A., Freilich, S., and Borenstein, E. (2015). NetCooperate : A network-based tool for inferring host-microbe and microbe-microbe cooperation. *BMC Bioinformatics*.
- [104] Lewis, N. E., Hixson, K. K., Conrad, T. M., Lerman, J. A., Charusanti, P., Polpitiya, A. D., Adkins, J. N., Schramm, G., Purvine, S. O., Lopez-Ferrer, D., Weitz, K. K., Eils, R., König, R., Smith, R. D., and Palsson, B. (2010). Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Molecular Systems Biology*, 6(390).
- [105] Liao, Y. C., Tsai, M. H., Chen, F. C., and Hsiung, C. A. (2012). GEMSiRV : A software platform for GENome-scale metabolic model simulation, reconstruction and visualization. *Bioinformatics*, 28(13) :1752–1758.
- [106] Lieven, C., Beber, M. E., Olivier, B. G., Bergmann, F. T., Ataman, M., Babaei, P., Bartell, J. A., Blank, L. M., Chauhan, S., Correia, K., Diener, C., Dräger, A., Ebert, B. E., Edirisinghe, J. N., Faria, J. P., Feist, A. M., Fengos, G., Fleming, R. M. T., García-Jiménez, B., Hatzimanikatis, V., Helvoirt, W. v., Henry, C. S., Hermjakob, H., Herrgård, M. J., Kaafarani, A., Kim, H. U., King, Z., Klamt, S., Klipp, E., Koehorst, J. J., König, M., Lakshmanan, M., Lee, D.-Y., Lee, S. Y., Lee, S., Lewis, N. E., Liu, F., Ma, H., Machado, D., Mahadevan, R., Maia, P., Mardinoglu, A., Medlock, G. L., Monk, J. M., Nielsen, J., Nielsen, L. K., Nogales, J., Nookaew, I., Palsson, B. O., Papin, J. A., Patil, K. R., Poolman, M., Price, N. D., Resendis-Antonio, O., Richelle, A., Rocha, I., Sánchez, B. J., Schaap, P. J., Sheriff, R. S. M., Shoaie, S., Sonnenschein, N., Teusink, B., Vilaça, P., Vik, J. O., Wodke, J. A. H., Xavier, J. C., Yuan, Q., Zakhartsev, M., and Zhang, C. (2020). MEMOTE for standardized genome-scale metabolic model testing. *Nature Biotechnology*, 38(3) :272–276.
- [107] Lifschitz, V. (2008). What is answer set programming? *Proceedings of the National Conference on Artificial Intelligence*, 3 :1594–1597.
- [108] Lipton, Z. C. (2017). The mythos of model interpretability.
- [109] Lloyd, C. J., King, Z. A., Sandberg, T. E., Hefner, Y., Olson, C. A., Phaneuf, P. V., O’Brien, E. J., Sanders, J. G., Salido, R. A., Sanders, K., Brennan, C., Humphrey, G., Knight, R., and Feist, A. M. (2019). The genetic basis for adaptation of model-designed syntrophic co-cultures. *PLoS Computational Biology*, 15(3) :1–28.
- [110] Loux, V., Mariadassou, M., Almeida, S., Chiapello, H., Hammani, A., Buratti, J., Gendrault, A., Barbe, V., Aury, J. M., Deutsch, S. M., Parayre, S., Madec, M. N., Chuat, V., Jan, G., Peterlongo, P., Azevedo, V., Le Loir, Y., and Falentin, H. (2015). Mutations and genomic islands can explain the strain dependency of sugar utilization in 21 strains of *Propionibacterium freudenreichii*. *BMC Genomics*, 16(1) :1–19.

- [111] Ma, H. and Zeng, A.-p. (2003). Genome Data and Analysis of Their Global. *Bioinformatics*, 19(2) :270–277.
- [112] Macchi, M., Festa, S., Nieto, E., Irazoqui, J. M., Vega-Vela, N. E., Junca, H., Valacco, M. P., Amadio, A. F., Morelli, I. S., and Coppotelli, B. M. (2021). Design and evaluation of synthetic bacterial consortia for optimized phenanthrene degradation through the integration of genomics and shotgun proteomics. *Biotechnology Reports*, 29.
- [113] Machado, D., Andrejev, S., Tramontano, M., and Patil, K. R. (2018). Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Research*.
- [114] Machado, D., Maistrenko, O. M., Andrejev, S., Kim, Y., Bork, P., Patil, K. R., and Patil, K. R. (2021). *Polarization of microbial communities between competitive and cooperative metabolism*, volume 5.
- [115] Mahadevan, R., Edwards, J. S., and Doyle, F. J. (2002a). Dynamic Flux Balance Analysis of Diauxic Growth in Escherichia coli. *Biophysical Journal*, 83(3) :1331–1340.
- [116] Mahadevan, R., Edwards, J. S., and Doyle, F. J. (2002b). Dynamic Flux Balance Analysis of Diauxic Growth in Escherichia coli. *Biophysical Journal*, 83(3) :1331–1340.
- [117] Mahadevan, R. and Schilling, C. H. (2003). The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic Engineering*.
- [118] Mahout, M., Carlson, R. P., and Peres, S. (2020). Answer set programming for computing constraints-based elementary flux modes : Application to escherichia coli core metabolism. *Processes*, 8(12) :1–17.
- [119] Makhlof, A. (2006). *Méthodologie pour l’optimisation dynamique multicritère d’un procédé discontinu alimenté : application à la production bactérienne d’arômes laitiers*. PhD thesis, Institut National Polytechnique de Lorraine.
- [120] McSweeney, P. L. and Sousa, M. J. (2000). Biochemical pathways for the production of flavour compounds in cheeses during ripening : A review. *Lait*, 80(3) :293–324.
- [121] Mendes-Soares, H., Mundy, M., Soares, L. M., and Chia, N. (2016). MMinte : An application for predicting metabolic interactions among the microbial species in a community. *BMC Bioinformatics*, 17(1) :1–10.
- [122] Mendoza, S. N., Olivier, B. G., Molenaar, D., and Teusink, B. (2019). A systematic assessment of current genome-scale metabolic reconstruction tools. *Genome Biology*, 20(1) :1–20.
- [123] Miyoshi, A., Rochat, T., and Gratadoux, J.-j. (2003). Miyoshi. *Genetics and Molecular Research*, 2(4) :348–359.
- [124] Morin, M. A., Morrison, A. J., Harms, M. J., and Dutton, R. J. (2022). Higher-order interactions shape microbial interactions as microbial community complexity increases. *Scientific Reports*, 12(1) :1–16.

- [125] Médigue, C., Bocs, S., Labarre, L., Math??, C., and Vallenet, D. (2002). Bioinformatique (1) l’annotation in silico des s??quences g??nomiques. *Medecine/Sciences*, 18(2) :237–250.
- [126] Nachtweide, S. and Stanke, M. (2019). Chapter 8. 1962 :139–160.
- [127] Nair, R. R., Vasse, M., Wielgoss, S., Sun, L., Yu, Y. T. N., and Velicer, G. J. (2019). Bacterial predator-prey coevolution accelerates genome evolution and selects on virulence-associated prey defences. *Nature Communications*, 10(1) :1–10.
- [128] Noronha, A., Modamio, J., Jarosz, Y., Guerard, E., Sompairac, N., Preciat, G., Daniélsdóttir, A. D., Krecke, M., Merten, D., Haraldsdóttir, H. S., Heinken, A., Heirendt, L., Magnúsdóttir, S., Ravcheev, D. A., Sahoo, S., Gawron, P., Friscioni, L., Garcia, B., Prendergast, M., Puente, A., Rodrigues, M., Roy, A., Rouquaya, M., Wiltgen, L., Žagare, A., John, E., Krueger, M., Kuperstein, I., Zinovyev, A., Schneider, R., Fleming, R. M. T., and Thiele, I. (2018). The Virtual Metabolic Human database : integrating human and gut microbiome metabolism with nutrition and disease. *Nucleic Acids Research*, 47(D1) :D614–D624.
- [129] Odum, H. T. (1983). *Systems ecology ; an introduction*.
- [130] Olivier, B. (2018). `SystemsBioinformatics/cbmpy-metadraft` : MetaDraft is now available.
- [131] Orth, J. D., Thiele, I., and Palsson, B. O. (2010). What is flux balance analysis?
- [132] Pacheco, A. R., Moel, M., and Segrè, D. (2019). Costless metabolic secretions as drivers of interspecies interactions in microbial ecosystems. *Nature Communications*, 10(1) :103.
- [133] Pacheco, A. R. and Segrè, D. (2021). An evolutionary algorithm for designing microbial communities via environmental modification. *Journal of the Royal Society Interface*, 18(179).
- [134] Palles, T., Beresford, T., Condon, S., and Cogan, T. M. (1998). Citrate metabolism in *Lactobacillus casei* and *Lactobacillus plantarum*. *Journal of Applied Microbiology*, 85(1) :147–154.
- [135] Palsson, B. O. and Varma, A. (1994). Metabolic Flux Balancing : Basic Concepts, Scientific and Practical Use. *Biotechnology*, 12(October) :994–998.
- [136] Passi, A., Tibocho-Bonilla, J. D., Kumar, M., Tec-Campos, D., Zengler, K., and Zuniga, C. (2022). Genome-scale metabolic modeling enables in-depth understanding of big data. *Metabolites*, 12(1).
- [137] Pitkänen, E., Jouhten, P., Hou, J., Syed, M. F., Blomberg, P., Kludas, J., Oja, M., Holm, L., Penttilä, M., Rousu, J., and Arvas, M. (2014). Comparative Genome-Scale Reconstruction of Gapless Metabolic Networks for Present and Ancestral Species. *PLoS Computational Biology*, 10(2).
- [138] Pnueli, A. (1977). The temporal logic of programs. In *Proceedings of the 18th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 46—57.

- [139] Popp, D. and Centler, F. (2020). μ BialSim : Constraint-Based Dynamic Simulation of Complex Microbiomes. *Frontiers in Bioengineering and Biotechnology*, 8(June).
- [140] Prendinger, H. and Schurz, G. (1996). Reasoning about action and change : A dynamic logic approach. *Journal of Logic, Language, and Information*, 5(2) :209–245.
- [141] Prigent, S., Frioux, C., Dittami, S. M., Thiele, S., Larhlimi, A., Collet, G., Gutknecht, F., Got, J., Eveillard, D., Bourdon, J., Plewniak, F., Tonon, T., and Siegel, A. (2017). Meneco, a Topology-Based Gap-Filling Tool Applicable to Degraded Genome-Wide Metabolic Networks. *PLoS Computational Biology*, 13(1) :e1005276.
- [142] Quatravaux, S., Remize, F., Bryckaert, E., Colavizza, D., and Guzzo, J. (2006). Examination of *Lactobacillus plantarum* lactate metabolism side effects in relation to the modulation of aeration parameters. *Journal of Applied Microbiology*, 101(4) :903–912.
- [143] Rajvanshi, M. and Venkatesh, K. V. (2013). *Encyclopedia of Systems Biology*.
- [144] Robinson, J. L., Kocabaş, P., Wang, H., Cholley, P. E., Cook, D., Nilsson, A., Anton, M., Ferreira, R., Domenzain, I., Billa, V., Limeta, A., Hedin, A., Gustafsson, J., Kerkhoven, E. J., Svensson, L. T., Palsson, B. O., Mardinoglu, A., Hansson, L., Uhlén, M., and Nielsen, J. (2020). An atlas of human metabolism. *Science Signaling*, 13(624) :1–12.
- [145] Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2009). edgeR : a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1) :139–140.
- [146] Roissart, H. d. (DL 1994). *Bactéries lactiques : aspects fondamentaux et technologiques / H. de Roissart, F. M. Luquet, coordonnateurs*. Lorica, Uriage.
- [147] Römer, M., Eichner, J., Dräger, A., Wrzodek, C., Wrzodek, F., and Zell, A. (2016). ZBIT bioinformatics toolbox : A web-platform for systems biology and expression data analysis. *PLoS ONE*.
- [148] Rousk, K., Michelsen, A., and Rousk, J. (2016). Microbial control of soil organic matter mineralization responses to labile carbon in subarctic climate change treatments. *Global Change Biology*, 22(12) :4150–4161.
- [149] Schroeder, W. L. and Saha, R. (2020). Introducing an Optimization- and explicit Runge-Kutta- based Approach to Perform Dynamic Flux Balance Analysis. *Scientific Reports*, 10(1) :1–28.
- [150] Schuster, S. and Hilgetag, C. (1994). On elementary flux modes in biochemical reaction systems at steady state. *Journal of Biological Systems*, 02(02) :165–182.
- [151] Seaver, S. M., Liu, F., Zhang, Q., Jeffryes, J., Faria, J. P., Edirisinghe, J. N., Mundy, M., Chia, N., Noor, E., Beber, M. E., Best, A. A., DeJongh, M., Kimbrel, J. A., D’haeseleer, P., McCorkle, S. R., Bolton, J. R., Pearson, E., Canon, S., Wood-Charlson, E. M., Cottingham, R. W., Arkin, A. P., and Henry, C. S. (2021). The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. *Nucleic Acids Research*.

- [152] Seemann, T. (2014). Prokka : rapid prokaryotic genome annotation. *Bioinformatics*, 30(14) :2068–2069.
- [153] Sharma, B. and Shukla, P. (2020). Designing synthetic microbial communities for effectual bioremediation : A review. *Biocatalysis and Biotransformation*, 38(6) :405–414.
- [154] Simeonidis, E., Rison, S. C., Thornton, J. M., Bogle, I. D. L., and Papageorgiou, L. G. (2003). Analysis of metabolic networks using a pathway distance metric through linear programming. *Metabolic Engineering*, 5(3) :211–219.
- [155] Smid, E. J. and Kleerebezem, M. (2014). Production of aroma compounds in lactic fermentations. *Annual Review of Food Science and Technology*, 5(1) :313–326.
- [156] Stein, R. R., Bucci, V., Toussaint, N. C., Buffie, C. G., Räscht, G., Pamer, E. G., Sander, C., and Xavier, J. B. (2013). Ecological Modeling from Time-Series Inference : Insight into Dynamics and Stability of Intestinal Microbiota. *PLoS Computational Biology*, 9(12) :31–36.
- [157] Stéphanou, A. and Volpert, V. (2016). Hybrid Modelling in Biology : A Classification Review. *Mathematical Modelling of Natural Phenomena*, 11(1) :37–38.
- [158] Steuer, R. (2007). Computational approaches to the topology, stability and dynamics of metabolic networks. *Phytochemistry*, 68(16-18) :2139–2151.
- [159] Stobbe, M. D., Jansen, G. A., Moerland, P. D., and van Kampen, A. H. (2014). Knowledge representation in metabolic pathway databases. *Briefings in Bioinformatics*, 15(3) :455–470.
- [160] Succurro, A. and Ebenhöf, O. (2018). Review and perspective on mathematical modeling of microbial ecosystems. *Biochemical Society Transactions*, 46(2) :403–412.
- [161] Succurro, A. and Ebenhöf, O. (2018). Review and perspective on mathematical modeling of microbial ecosystems. *Biochemical Society Transactions*, 46(2) :403–412.
- [162] Sweetlove, L. J. and Fernie, A. R. (2005). Regulation of metabolic networks : Understanding metabolic complexity in the systems biology era. *New Phytologist*, 168(1) :9–24.
- [163] Swindell, S. R., Benson, K. H., Griffin, H. G., Renault, P., Ehrlich, S. D., and Gasson, M. J. (1996). Genetic manipulation of the pathway for diacetyl metabolism in *Lactococcus lactis*. *Applied and Environmental Microbiology*, 62(7) :2641–2643.
- [164] Sánchez López de Nava A, R. A. (2023). *Physiology, Metabolism*.
- [165] Thiele, I. and Palsson, B. O. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*, 5(1) :93–121.
- [166] Thierry, A., Deutsch, S. M., Falentin, H., Dalmasso, M., Cousin, F. J., and Jan, G. (2011). New insights into physiology and metabolism of *Propionibacterium freudenreichii*. *International Journal of Food Microbiology*, 149(1) :19–27.
- [167] Thorleifsson, S. G. and Thiele, I. (2011). rBioNet : A COBRA toolbox extension for reconstructing high-quality biochemical networks. *Bioinformatics*, 27(14) :2009–2010.

- [168] Tiwari, A., Talcott, C., Knapp, M., Lincoln, P., and Laderoute, K. (2007). Analyzing pathways using sat-based approaches. In Anai, H., Horimoto, K., and Kutsia, T., editors, *Algebraic Biology*, pages 155–169, Berlin, Heidelberg. Springer Berlin Heidelberg.
- [169] Trinh, C. T., Wlaschin, A., and Sreenc, F. (2009). Elementary mode analysis : A useful metabolic pathway analysis tool for characterizing cellular metabolism. *Applied Microbiology and Biotechnology*, 81(5) :813–826.
- [170] Turgay, M., Bachmann, H.-P., Irmeler, S., von Ah, U., Frö Hlich-Wyder, M.-T., Falentin, H., Deutsch, S.-M., Jan, G., and Thierry, A. (2020). Propionibacterium spp. and Acidipropionibacterium spp. In *Reference Module in Food Science*, Reference Module in Food Science. Elsevier.
- [171] Udugama, I. A., Lopez, P. C., Gargalo, C. L., Li, X., Bayer, C., and Gernaey, K. V. (2021). Digital twin in biomanufacturing : challenges and opportunities towards its implementation. *Systems Microbiology and Biomanufacturing*, 1(3) :257–274.
- [172] Vallenet, D., Calteau, A., Dubois, M., Amours, P., Bazin, A., Beuvin, M., Burlot, L., Bussell, X., Fouteau, S., Gautreau, G., Lajus, A., Langlois, J., Planel, R., Roche, D., Rollin, J., Rouy, Z., Sabatet, V., and Médigue, C. (2020). MicroScope : An integrated platform for the annotation and exploration of microbial gene functions through genomic, pangenomic and metabolic comparative analysis. *Nucleic Acids Research*, 48(D1) :D579–D589.
- [173] Vallenet, D., Calteau, A., Dubois, M., Amours, P., Bazin, A., Beuvin, M., Burlot, L., Bussell, X., Fouteau, S., Gautreau, G., Lajus, A., Langlois, J., Planel, R., Roche, D., Rollin, J., Rouy, Z., Sabatet, V., and Médigue, C. (2019). MicroScope : an integrated platform for the annotation and exploration of microbial gene functions through genomic, pangenomic and metabolic comparative analysis. *Nucleic Acids Research*.
- [174] Van Rooijen, R. J., Van Schalkwijk, S., and De Vos, W. M. (1991). Molecular cloning, characterization, and nucleotide sequence of the tagatose 6-phosphate pathway gene cluster of the lactose operon of *Lactococcus lactis*. *Journal of Biological Chemistry*, 266(11) :7176–7181.
- [175] van Rosmalen, R., Smith, R., Martins dos Santos, V., Fleck, C., and Suarez-Diez, M. (2021). Model reduction of genome-scale metabolic models as a basis for targeted kinetic models. *Metabolic Engineering*, 64 :74–84.
- [176] Vázquez-Castellanos, J. F., Biclôt, A., Vrancken, G., Huys, G. R., and Raes, J. (2019). Design of synthetic microbial consortia for gut microbiota modulation. *Current Opinion in Pharmacology*, 49 :52–59.
- [177] Viridis, C., Sumby, K., Bartowsky, E., and Jiranek, V. (2021). Lactic Acid Bacteria in Wine : Technological Advances and Evaluation of Their Functional Role. *Frontiers in Microbiology*, 11(January) :1–16.
- [178] Wagner, A. and Fell, D. A. (2001). The small world inside large metabolic networks. *Proceedings of the Royal Society B : Biological Sciences*, 268(1478) :1803–1810.

- [179] Wang, H., Marcišauskas, S., Sánchez, B. J., Domenzain, I., Hermansson, D., Agren, R., Nielsen, J., and Kerkhoven, E. J. (2018). RAVEN 2.0 : A versatile toolbox for metabolic network reconstruction and a case study on *Streptomyces coelicolor*. *PLoS Computational Biology*, 14(10) :1–17.
- [180] Wang, Y., Xu, J., Kong, L., Liu, T., Yi, L., Wang, H., Huang, W. E., and Zheng, C. (2020). Raman–deuterium isotope probing to study metabolic activities of single bacterial cells in human intestinal microbiota. *Microbial Biotechnology*, 13(2) :572–583.
- [181] Weiss, A. S., Burrichter, A. G., Durai Raj, A. C., von Stempel, A., Meng, C., Kleigrewe, K., Münch, P. C., Rössler, L., Huber, C., Eisenreich, W., Jochum, L. M., Göing, S., Jung, K., Lincetto, C., Hübner, J., Marinos, G., Zimmermann, J., Kaleta, C., Sanchez, A., and Stecher, B. (2022). In vitro interaction network of a synthetic gut bacterial community. *ISME Journal*, 16(4) :1095–1109.
- [182] Widyastuti, Y., , R., and Febrisiantosa, A. (2014). The Role of Lactic Acid Bacteria in Milk Fermentation. *Food and Nutrition Sciences*, 05(04) :435–442.
- [183] Wilmes, P., Martin-Gallausiaux, C., Ostaszewski, M., Aho, V. T., Novikova, P. V., Laczny, C. C., and Schneider, J. G. (2022). The gut microbiome molecular complex in human health and disease. *Cell Host & Microbe*, 30(9) :1201–1206.
- [184] Yadav, S. K., Soni, R., and Rajput, A. S. (2018). Role of Microbes in Organic Farming for Sustainable Agro-Ecosystem. pages 241–252.
- [185] Zelezniak, A., Andrejev, S., Ponomarova, O., Mende, D. R., Bork, P., and Patil, K. R. (2015). Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proceedings of the National Academy of Sciences of the United States of America*.
- [186] Zhang, Y. J., Li, S., Gan, R. Y., Zhou, T., Xu, D. P., and Li, H. B. (2015). Impacts of gut bacteria on human health and diseases. *International Journal of Molecular Sciences*, 16(4) :7493–7519.
- [187] Zhuang, K., Izallalen, M., Mouser, P., Richter, H., Risso, C., Mahadevan, R., and Lovley, D. R. (2011). Genome-scale dynamic modeling of the competition between *Rhodospirillum rubrum* and *Geobacter* in anoxic subsurface environments. *ISME Journal*, 5(2) :305–316.
- [188] Zien, A., Küffner, R., Zimmer, R., and Lengauer, T. (2000). Analysis of gene expression data with pathway scores. *Proceedings / ... International Conference on Intelligent Systems for Molecular Biology ; ISMB. International Conference on Intelligent Systems for Molecular Biology*, 8 :407–417.
- [189] Zimmermann, J., Kaleta, C., and Waschina, S. (2021). gapseq : informed prediction of bacterial metabolic pathways and reconstruction of accurate metabolic models. *Genome Biology*, 22(1) :81.
- [190] Zomorodi, A. R. (2014). d-OptCom : Dynamic Multi-level and Multi-objective Metabolic Modeling of Microbial Communities.

- [191] Zomorodi, A. R. and Maranas, C. D. (2012). OptCom : A multi-level optimization framework for the metabolic modeling and analysis of microbial communities. *PLoS Computational Biology*, 8(2).
- [192] Zomorodi, A. R. and Segrè, D. (2017). Genome-driven evolutionary game theory helps understand the rise of metabolic interdependencies in microbial communities. *Nature Communications*, 8(1) :1–11.
- [193] Zou, Y., Xue, W., Luo, G., Deng, Z., Qin, P., Guo, R., Sun, H., Xia, Y., Liang, S., Dai, Y., Wan, D., Jiang, R., Su, L., Feng, Q., Jie, Z., Guo, T., Xia, Z., Liu, C., Yu, J., Lin, Y., Tang, S., Huo, G., Xu, X., Hou, Y., Liu, X., Wang, J., Yang, H., Kristiansen, K., Li, J., Jia, H., and Xiao, L. (2019). 1,520 Reference Genomes From Cultivated Human Gut Bacteria Enable Functional Microbiome Analyses. *Nature Biotechnology*, 37(2) :179–185.
- [194] Zuñiga, C., Li, T., Guarnieri, M. T., Jenkins, J. P., Li, C. T., Bingol, K., Kim, Y. M., Betenbaugh, M. J., and Zengler, K. (2020). Synthetic microbial communities of heterotrophs and phototrophs facilitate sustainable growth. *Nature Communications*, 11(1).
- [195] Özcan, E., Seven, M., Şirin, B., Çakır, T., Nikerel, E., Teusink, B., and Öner, E. T. (2020). Dynamic co-culture metabolic models reveal the fermentation dynamics, metabolic capacities and interplays of cheese starter cultures. *Biotechnology and Bioengineering*, 118(1) :223–237.

Modélisation numérique du métabolisme de 3 bactéries du fromage

Composition du milieu extracellulaire utilisé pour les simulations

Pour lancer les simulations numériques, nous avons utilisé les composés que l'on peut trouver dans le lait et listés dans la Table 1 :

Nom du composé	Identifiant du composé	Composé du lait	Cible	Numéro CAS
beta-carotene	caro	✓		7235-40-7
retinol	retinol	✓		68-26-8
zinc	zn2	✓		7440-66-6
cobalt	cobalt2	✓		7440-48-4
sodium	na1	✓		7440-23-5
potassium	k	✓		7440-09-7
magnesium	mg2	✓		7439-95-4
Phosphate	pi	✓		14265-44-2
mn2+	mn2	✓		16397-91-4
fer3+	fe3	✓		20074-52-6
fer2+	fe2	✓		7439-89-6
cuiivre	cu2	✓		7440-50-8
chloride	cl	✓		16887-00-6
calcium	ca2	✓		7440-70-2
proton	h	✓		12408-02-5
molybdate	mobd	✓		14259-85-9
arachidonique	arachd	✓		506-32-1
alpha-linoleique	lnlnca	✓		60-33-3
linoleique	lnlc	✓		60-33-3
stearique	ocdca	✓		57-11-4
oleique	ocdcea	✓		112-81-1
palmitique	D-hdca	✓		57-10-3
myristique	ttdca	✓		544-63-8
laurique	ddca	✓		143-07-7
caprylique	octa	✓		124-07-2
lactose	lcts	✓	✓	63-42-3
citrate	cit	✓		126-44-3
L-aspartate	asp__L	✓		56-84-8
D-aspartate	asp__D	✓		1783-96-6
L-threonine	thr__L	✓		72-19-5
L-Serine	ser__L	✓		56-45-1
D-glutamate	glu__D	✓		6893-26-1
L-glutamate	glu__L	✓		56-86-0
l-proline	pro__L	✓		147-85-3
D-proline	pro__D	✓		344-25-2
glycine	gly	✓		56-40-6
d-alanine	ala__D	✓		338-69-2
l-alanine	ala__L	✓		56-41-7
cysteine	cys__L	✓		52-90-4
valine	val__L	✓		72-18-4
methionine	met__L	✓		63-68-3
isoleucine	ile__L	✓		73-32-5
leucine	leu__L	✓		61-90-5
tryptophan	trp__L	✓		73-22-3
Tyrosine	tyr__L	✓		60-18-4
phenylalanine	phe__L	✓		63-91-2
Histidine	his__L	✓		71-00-1
Lysine	lys__L	✓		56-87-1
Arginine	arg__L	✓		74-79-3
ammonia	nh3	✓		7664-41-7
ammonium	nh4	✓		14798-03-9
sulfate	so4	✓		14808-79-8
Ni2+	ni2	✓		14701-22-5
Glutamine	gln__L	✓		56-85-9
asparagine	asn__L	✓		70-47-3
functional_form_vitamin_b12	b12	✓		68-19-9
caproïque	hxa	✓		142-62-1
vitamine_b6	pydxn	✓		65-23-6
vitamine_b6_P	pydx5p	✓		54-47-7
Water	h2o	✓		7732-18-5
O2	O2	✓		7782-44-7
vitamine_b5	pnto__R	✓		79-83-4
Vitamine_b3	nac	✓		59-67-6
Vitamine_b2	ribflv	✓		83-88-5
Vitamin_b1	thm	✓		59-43-8
Vitamin_C	ascb__L	✓		50-81-7
L-LACTATE	lac__L		✓	50-21-5
D-LACTATE	lac__D		✓	10326-41-7
5,6,7,8-tetrahydrofolate(vitamin_b9)	thf	✓		135-16-0
Diacyl	diact		✓	431-03-8
butanediol	btd__RR		✓	513-85-9
propionic acid	ppa		✓	79-09-4

TABLE 1 – Composition du milieu extracellulaire

Modélisation logique du métabolisme des communautés bactériennes

Taxonomie des génomes utilisés pour reconstruire les communautés simulées

2752 génomes de références isolés des microbiotes de l'intestin humain, feuille et racine d'*A. thaliana*, et du sol ont été utilisés pour la reconstruction des réseaux métaboliques à l'échelle du génome (GEMs) et l'assemblage de communautés simulées pour chacun des quatre habitats. La figure supplémentaire 1 montre la diversité taxonomique de ces génomes.

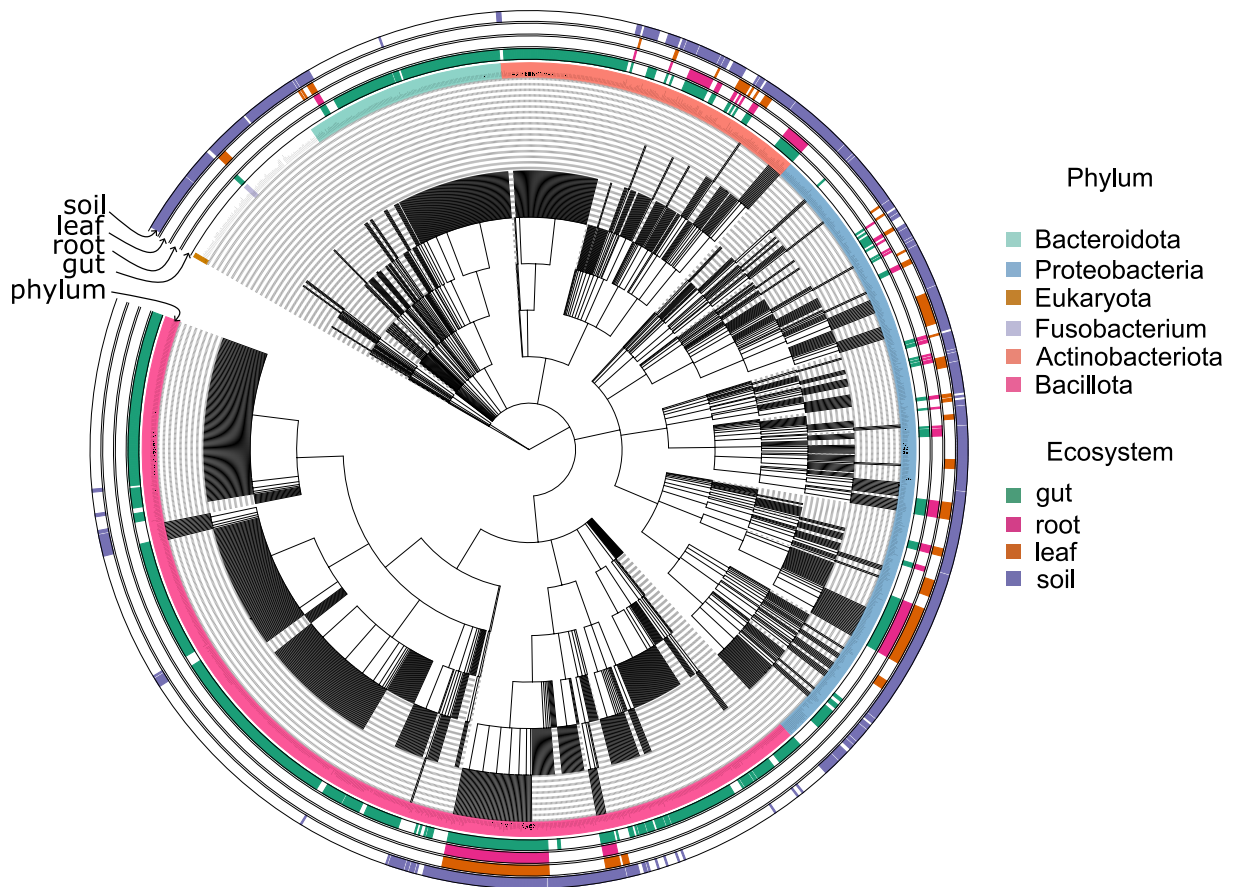


FIGURE 1 – **Taxonomie des génomes associés à chaque GEM des 4 écosystèmes.** Le cercle intérieur de couleur représente le phylum du génome. Les quatre cercles supplémentaires indiquent la présence du taxon parmi les génomes de l'écosystème correspondant. La figure a été créée avec iTOL (99).

Potentiels d'interactions graines spécifique

Sans le mixte

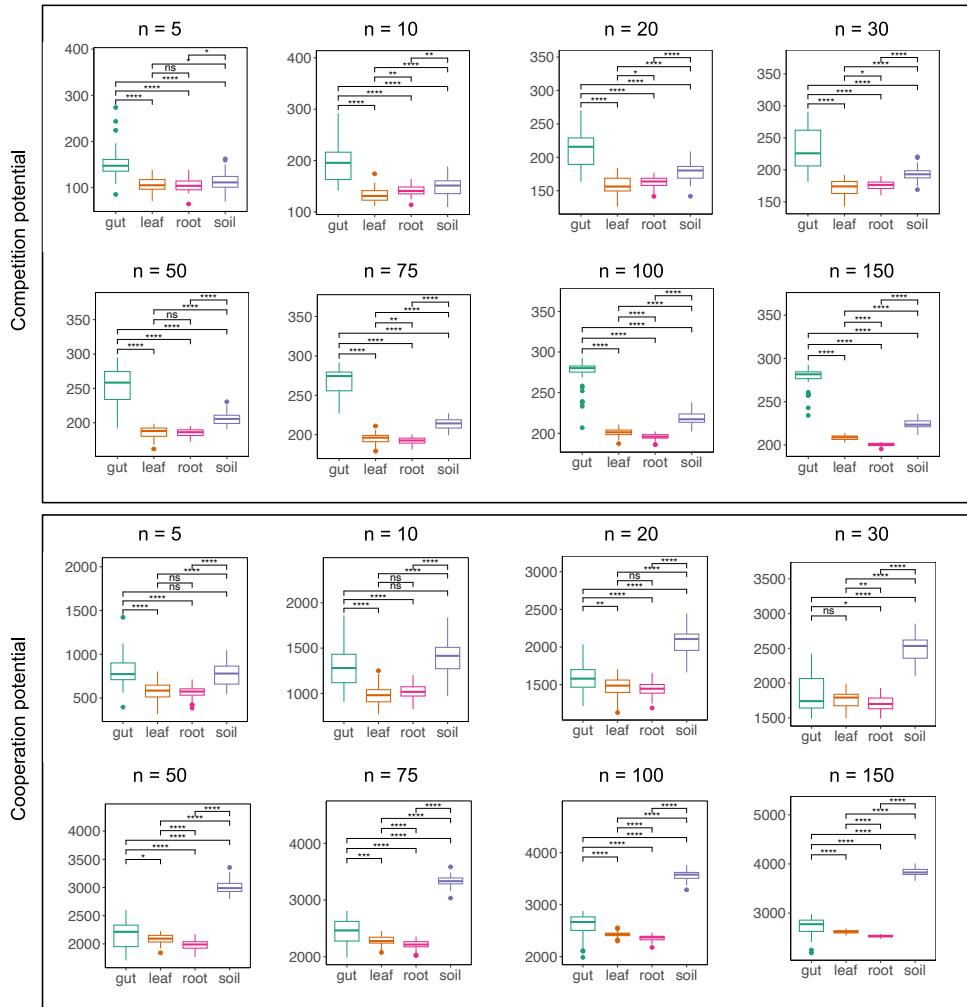


FIGURE 2 – Distribution des scores de coopération et de compétition dans des communautés simulées de 5 à 150 bactéries isolées de l'intestin, de la racine, de la feuille et du sol en utilisant les graines spécifiques. Une comparaison de moyennes avec le test de Wilcoxon a été faite.

Avec le mixte

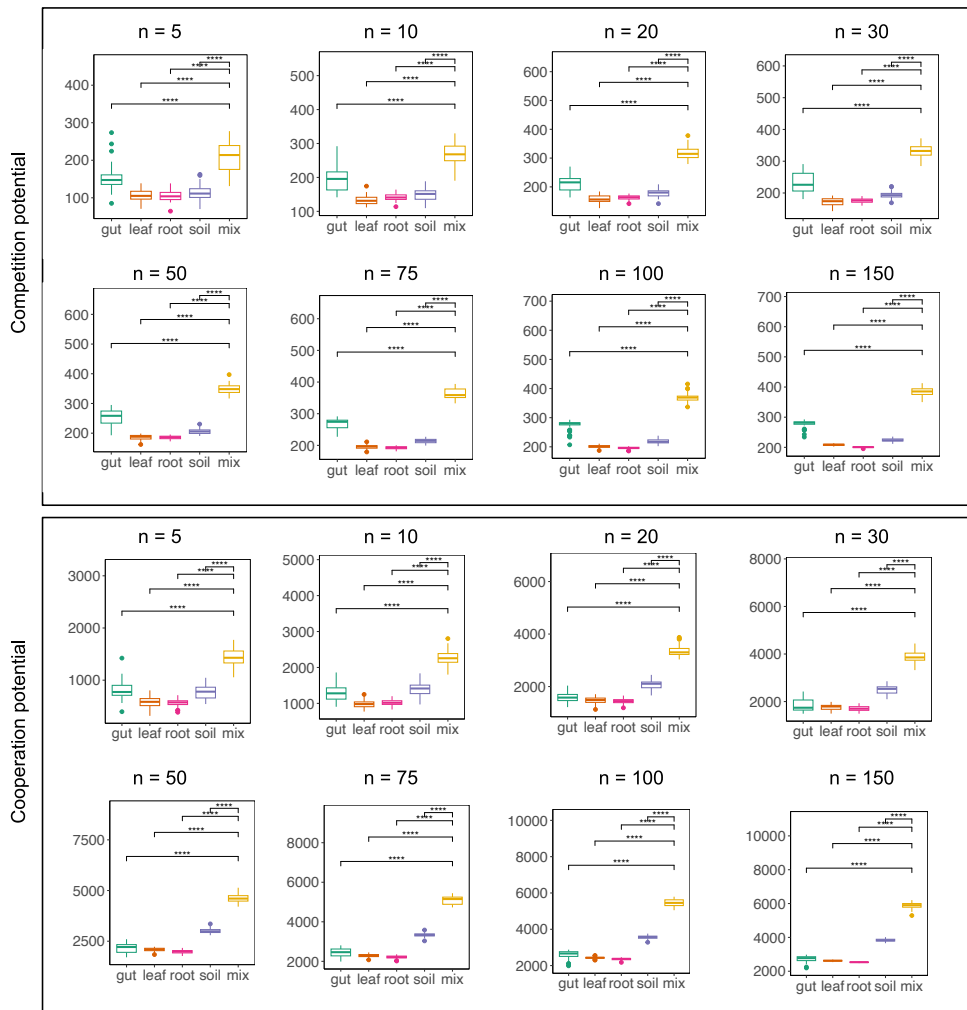


FIGURE 3 – Distribution des scores de coopération et de compétition dans des communautés simulées de 5 à 150 bactéries isolées de l'intestin, de la racine, de la feuille, sol et du mixte. Seul une comparaison avec deux à deux avec le mixte a été faite. La comparaison de moyennes avec le test de Wilcoxon a été fait.

Plus value d'ajout ou de retirer une espèce dans le milieu graine spécifique

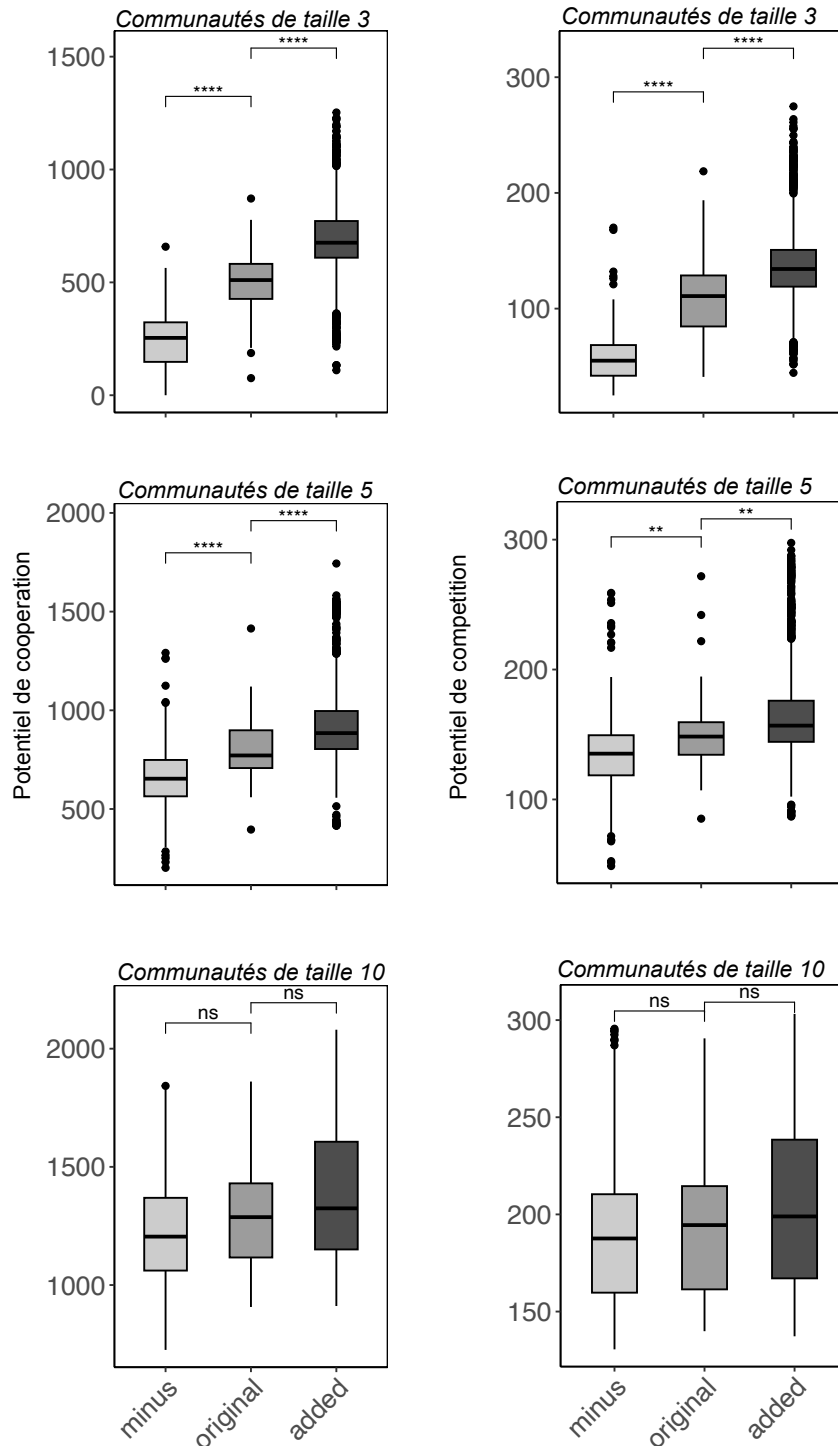


FIGURE 4 – Effet d'ajouter ("added") ou de retirer une espèce ("minus") sur les potentiels de coopération et de compétition pour chaque communauté avec les graines spécifiques. Le test de comparaison de moyennes Wilcoxon a été utilisé sur des communautés de taille 3 à 10 sur l'écosystème de l'intestin.

.1 Milieux générique et spécifique utilisés pour les simulations

Les deux milieux spécifiques pour analyser le microbiote de l'intestin et ceux de la feuille, racine et du sol ainsi que le milieu générique sont représentés dans la Table 2.

Composé du milieu générique	Composé du milieu spécifique pour la racine, feuille et le sol	Composé du milieu spécifique pour l'intestin
WATER	WATER	WATER
PROTON	PROTON	PROTON
CA+2	CA+2	CA+2
K+	K+	K+
MG+2	MG+2	MG+2
NA+	NA+	NA+
Pi	Pi	Pi
CU+2	CU+2	CU+2
FE+2	FE+2	FE+2
FE+3	FE+3	FE+3
MN+2	MN+2	MN+2
ZN+2	ZN+	2 ZN+2
CL-	CL-	CL-
CPD-387	CPD-387	CPD-387
AMMONIUM	AMMONIUM	AMMONIUM
NITRATE	NITRATE	NITRATE
OXYGEN-MOLECULE	OXYGEN-MOLECULE	OXYGEN-MOLECULE
SULFATE	SULFATE	SULFATE
SELENATE	SELENATE	SELENATE
ETOH	MANNITOL	ETOH
CELLULOSE	GLC	CELLULOSE
Starch	L-CITRULLINE	Starch
SORBITOL	FRU	SORBITOL
GLC	FUM	GLC
MANNITOL	MAL	MANNITOL
XYLITOL	D-LACTATE	XYLITOL
CPD-15972	L-LACTATE	CPD-15972
MALTOSE	METOH	MALTOSE
SUCROSE	SUC	SUCROSE
D-galactopyranose	SUCROSE	D-galactopyranose
BETA-D-FRUCTOSE	4-AMINO-BUTYRATE	BETA-D-FRUCTOSE
DODECANOATE	L-ALPHA-ALANINE	DODECANOATE
PALMITATE	ARG	PALMITATE
STEARIC_ACID	ASN	STEARIC_ACID
OLEATE-CPD	L-ASPARTATE	OLEATE-CPD
ARACHIDIC_ACID	CYS	ARACHIDIC_ACID
ARACHIDONIC_ACID	GLN	ARACHIDONIC_ACID
BUTYRIC_ACID	GLT	BUTYRIC_ACID
CHOLESTEROL	GLY	CHOLESTEROL
TETRACOSANOATE	HIS	TETRACOSANOATE
LINOLEIC_ACID	ILE	LINOLEIC_ACID
ASCORBATE	LEU	ASCORBATE
PANTOTHENATE	LYS	PANTOTHENATE
DOCOSANOATE	MET	DOCOSANOATE
LINOLENIC_ACID	PHE	LINOLENIC_ACID
L-ALPHA-ALANINE	PRO	CPD-3617
ASN	SER	CPD-7836
ARG	THR	CPD-8462
L-ASPARTATE	TRP	CPD-7830
CYS	TYR	CPD-17322
GLN	VAL	CPD-195
GLT	Vitamins-B12	CPD-9245
GLY	COB-I-ALAMIN	CPD-17188
HIS	CHOLINE	CPD-13792
ILE	Reduced-ferredoxins	CPD-12189
LEU	Acceptor	CPD-12653
LYS	Donor-H2	CPD-14292
MET	ACP	CPD0-2208
PHE	NAD	CPD-10244
PRO	NADP	L-ALPHA-ALANINE
SER	ADP	ASN
THR		ARG
TRP		L-ASPARTATE
TYR		CYS
VAL		GLN
Vitamins-B12		GLT
COB-I-ALAMIN		GLY
CHOLINE		HIS
THIAMINE		ILE
ADENOSYLCOBALAMIN		LEU
RIBOFLAVIN		LYS
NIACINE		MET
PYRIDOXAMINE		PHE
PYRIDOXINE		PRO
PYRIDOXAL		SER
BIOTIN		THR
VITAMIN_D3		TRP
ALPHA-TOCOPHEROL		TYR
5-METHYL-THF		VAL
THF		Retinols
Acceptor		THIAMINE
Donor-H2		ADENOSYLCOBALAMIN
Reduced-ferredoxins		RIBOFLAVIN
ACP		NIACINE
NAD		NIACINAMIDE
NADP		PYRIDOXAMINE
ADP		PYRIDOXINE
		PYRIDOXAL
		BIOTIN
		VITAMIN_D3
		ALPHA-TOCOPHEROL
		CPD-12826
		10-FORMYL-THF
		5-METHYL-THF
		THF
		Vitamins-B12
		COB-I-ALAMIN
		Acceptor
		Donor-H2
		Reduced-ferredoxins
		ACP
		URATE
		CPD1F-129
		NAD
		NADP
		ADP

TABLE 2 – Composition du milieu générique et spécifique

Contributions scientifiques

Articles scientifiques

- **Article de journal**, Maxime Lecomte, Wenfan Cao, Julie Aubert, David James Sherman, H el ene Falentin, Cl emence Frioux and Simon Labarthe, Revealing the dynamics and mechanisms of bacterial interactions in cheese production with metabolic modelling, **en r evision**², 2023
- **Article de journal**, Maxime Lecomte, Coralie Muller, Ariane Badoual, H el ene Falentin, David James Sherman and Cl emence Frioux, Reasoning-based metabolic modelling of cooperation and competition potentials in large-scale microbial communities, **en cours de r edaction**.

Posters

- **Conf erence internationale**, MPA, *Metabolic Pathway Analysis conference*, Metabolic modelling deciphers interactions in a cheese bacterial community, 2021
- **Conf erence internationale**, CMSB *19th International Conference on Computational Methods in Systems Biology*, Metabolic modelling deciphers interactions in a cheese bacterial community, 2021
- **Conf erence nationale**, JOBIM conference, *Journ ees Ouvertes en Biologie, Informatique et Math ematiques*, Characterising competition and cooperation potentials in microbial communities using discrete models of metabolism, 2022
- **Conf erence nationale**, JOBIM conference, *Journ ees Ouvertes en Biologie, Informatique et Math ematiques*, Metabolic modelling deciphers interactions in a cheese bacterial community, 2021

Pr esentations orales

- **S eminaire international**, GT SymBioDiversity, Characterizing microbial interactions in controlled and natural microbial communities³, 2022
- **Conf erence nationale**, Journ ee annuelle GT BIOSS, *Groupe de travail sur la Biologie syst emique symbolique*, Reasoning approaches for the characterization of cooperation and competition in large-scale microbial communities, 2022
- **Conf erence nationale** METABODAY, Metabolic modelling deciphers interactions in a cheese bacterial community, 2021
- **S eminaire BioNum**, *Sant e et bio-num erique*, Numerical and reasoning based metabolic modelling of bacterial communities, 2023

2. <https://www.sciencedirect.com/journal/metabolic-engineering>

3. <https://eventos.cmm.uchile.cl/symbiodiversity/>

Liste des acronymes

ASP	Answer Set Programming (anglais)	Programmation par ensemble de réponse
FBA	Flux balance analysis (anglais)	Analyse par équilibre des flux
GEM	Genome-scale metabolic network (anglais)	Réseau métabolique à l'échelle du gé- nome
LCMS	chromatographie en phase liquide couplé à la spectrométrie de masse	
LL	<i>L. lactis</i>	
LP	<i>L. plantarum</i>	
LTL	Logique linéaire temporelle	
PF	<i>P. freudenreichii</i>	
RAC	Raisonnement sur les actions et les changements	
SBML	Systems Biology Markup Language	

Table des figures

1	Présence de bactéries dans divers écosystèmes et illustrations des interactions bactériennes.	14
1.1	Illustration d'un ensemble de voies métaboliques de <i>E. coli</i> dans l'application web Escher (84)	17
1.2	Illustration des approches descendante (a) et ascendante (b) pour reconstruire des modèles métaboliques à l'échelle du génome fonctionnels.	19
1.3	Représentation du métabolisme sous forme d'un graphe bipartite selon la <i>Systems Biology Graphical Notation</i> (SBGN) (16).	22
1.4	Cône de solution obtenu lors de la résolution d'un FBA.	26
1.5	Schéma récapitulatif des étapes de résolution d'un problème ASP à partir des règles logiques	29
1.6	Différents types de stratégies pour construire un modèle de communauté.	33
1.7	Stratégie mise en place par l'outil MiCOM	34
1.8	Illustration du calcul du MRO (compétition) et du MIP (coopération) par SMETANA	35
1.9	Illustration du fonctionnement de l'outil MiSCoTo.	39
1.10	Illustration des méthodes numériques et qualitatives de modélisation du métabolisme d'une communauté avec leurs avantages et inconvénients.	40
2.1	Récapitulatif des étapes de raffinement et de calibration des modèles métaboliques, des simulations ainsi que des données utilisées.	44
2.2	Données de cultures pures utilisées pour calibrer chaque souche.	46
2.3	Données de croissances et de pH en co-culture pour tester les prédictions de notre modèle de communauté	47
2.4	Cartes des voies métaboliques du métabolisme carboné central d'intérêt de <i>L. lactis</i>	54
2.5	Cartes des voies métaboliques du métabolisme carboné central d'intérêt de <i>L. plantarum</i>	55
2.6	Cartes des voies métaboliques du métabolisme carboné central d'intérêt de <i>P. freudenreichii</i>	57
2.7	Réseau métabolique de <i>L. lactis</i> après calibration sur les données de culture pure.	64
2.8	Réseau métabolique de <i>L. plantarum</i> après calibration sur les données de culture pure.	65
2.9	Réseau métabolique de <i>P. freudenreichii</i> après calibration sur les données de culture pure.	66

2.10	Validation des prédictions après apprentissage.	67
2.11	Prédiction de la croissance de chaque souche ainsi que des concentrations des métabolites	68
2.12	Illustration du changement de voies métaboliques durant le processus de production du fromage.	69
2.13	Flux dynamiques de consommation et de production de métabolites d'intérêts.	71
2.14	Expression des gènes des données métatranscriptomiques	75
3.1	Préparation des données afin d'être utilisé par le SVM.	92
3.2	Modèle basé sur le raisonnement et son application sur des données jouets.	95
3.3	Potentiel métabolique de la communauté	96
3.4	Modèle basé sur le raisonnement et son application sur des données jouets.	97
3.5	Application de CoCoMiCo sur les exemples jouets (communauté de taille 2)	99
3.6	Application de CoCoMiCo sur les exemples jouets (communauté de taille 3)	100
3.7	Application de CoCoMiCo sur un exemple jouet (communauté de taille 4)	101
3.8	Scores de coopération et de compétition calculés pour 50 communautés de taille 2 de chaque écosystème.	102
3.9	Distribution des scores de coopération et de compétition dans des communautés simulées de 5 à 150 bactéries	104
3.10	Distribution des scores de coopération et de compétition dans des communautés simulées de 5 à 150 bactéries	105
3.11	Effet d'ajouter ("added") ou de retirer une espèce ("minus") sur les potentiels de coopération et de compétition pour chaque communauté.	107
3.12	Distribution des scores de coopération et de compétition dans des communautés simulées de 5 à 150 bactéries isolées	109
3.13	Distribution des scores de coopération et de compétition dans des communautés simulées de 5 à 150 bactéries isolées	111
3.14	Comparaison des approches numériques MICOM et SMETANA avec notre approche.	113
4.1	Positionnement du développement de notre prototype vis-à-vis de l'outil existant MiSCoTo (57)	119
4.2	Opérateurs temporels du futur et du passé en <i>telingo</i> et logique temporelle issue de (23)	129
4.3	Réutilisation du composé C par la réaction R4 dans le cas où C était productible trois temps réactionnels avant	131
4.4	Schéma didactique permettant de valider la solution donnée par le modèle dans le cas de la planification	132
4.5	Schéma didactique permettant de valider la solution donnée par le modèle dans le cas de l'explication	135
4.6	Schéma permettant de visualiser le résultat simplifié du modèle d'explication sur les données réelles.	137

1	Taxonomie des génomes associés à chaque GEM des 4 écosystèmes.	167
2	Distribution des scores de coopération et de compétition dans des communautés simulées de 5 à 150 bactéries en utilisant les graines spécifiques	168
3	Distribution des scores de coopération et de compétition dans des communautés simulées de 5 à 150 bactéries avec le mixte en utilisant les graines spécifiques	169
4	Effet d'ajouter ("added") ou de retirer une espèce ("minus") sur les potentiels de coopération et de compétition pour chaque communauté avec les graines spécifiques	170

Liste des tableaux

1.1	Table récapitulative d'outils de modélisation statique d'une communauté microbienne.	35
1.2	Table récapitulant les outils de modélisation dynamique d'une communauté microbienne.	37
2.1	Données de dosage d'acide en g/L pour <i>P. freudenreichii</i>	46
2.2	Ensemble des paramètres optimisés pour chaque modèle individuel avec leur justification.	52
2.3	Caractéristiques générales des trois modèles métaboliques	58
2.4	Raffinement manuel des GEMs avec les justifications correspondantes	59
2.5	Métabolites échangeables sur la communauté du fromage identifiés par MICOM	73
2.6	Métabolites échangeables sur la communauté du fromage identifiés par SMETANA.	74
2.7	Liste de métabolites prédits par Metage2Metabo comme étant productibles uniquement par coopération métabolique	76
3.1	Résultat des tests de corrélation (Spearman) entre l'indice de similarité de Jaccard et les scores de potentiels d'interaction . . .	103
3.2	Moyenne des scores d'interactions ainsi que du nombre de métabolites échangeables pour tous les écosystèmes et pour une taille de communauté de 150.	104
3.3	Résultat de prédiction des écosystèmes de la racine, de la feuille, du sol et de l'intestin, à partir des graines génériques.	110
3.4	Résultat de prédiction des écosystèmes de la racine, de la feuille, du sol et de l'intestin, à partir de graines spécifiques.	111
3.5	CoCoMiCo, Smetana et MICOM sont appliqués sur les données expérimentales du kéfir (19) et une comparaison des potentiels d'interaction a été faite.	114
1	Composition du milieu extracellulaire	165
2	Composition du milieu générique et spécifique	172