

Chapitre 1

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

Soumis au journal Metabolic Engineering ¹

Une listes des abréviations propre au métabolisme est disponible à la fin de ce chapitre section 1.4.

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

1.1 Introduction

Au sein de ce chapitre nous allons aborder la question de comment un modèle numérique, tel que nous l'avons défini, permet de répondre à plusieurs enjeux biologiques avec un critère d'explicabilité satisfaisant. Ce modèle numérique s'inscrit dans un projet, appelé TANGO (?) où les objectifs sont multiples. Différentes productions de fromage à pâte pressée non cuites de 2,5kg ont été réalisées en triplicat avec les mêmes bactéries lactiques et propionique mais en suivant des itinéraires légèrement modifiés. Ces changements ont permis de mettre en évidence expérimentalement 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, les métabolites échangés entre les espèces dans l'itinéraire standard de fabrication.

À notre connaissance il existe des modèles numériques du métabolisme, garantissant la fiabilité des résultats ainsi que la précision, cependant, ces modèles sont souvent "boîtes noires" et ne permettent pas d'obtenir une explication mécanistique des résultats (?). Ainsi, ces méthodes numériques fournissent des résultats de simulations mais ne permettent pas d'expliquer les voies métaboliques activées. De plus, Popp et al précisent également que ces résultats ne doivent pas engendrer une interprétation biologique détaillée étant donné l'inférence de paramètres génériques. Une conséquence de ces modèles est la difficulté à inférer des processus de régulation biologique trouvés 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 (?) 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, ils identifient le chemin le plus coûteux représentant l'investissement d'une cellule dans la transcription. Avec les données de croissance cinétique, du pH ou encore de la métabolomique (?) ont construit un modèle dynamique afin de prédire des concentrations de métabolites et ont inféré des mécanismes biologiques appris à partir des données. Ayant à disposition des données hétérogènes similaires à (?), 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 (?), optimisant les modèles individuels et de communautés avec les données, nous avons alors développé une stratégie itérative 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 bactériens 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 de la fiabilité de chaque modèle individuel.

Les travaux présentés dans ce chapitre ont mené à la soumission d'un article² de journal (?) en cours de publication et de plusieurs présentations scientifiques et posters [ML – lien vers HAL].

1.2 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, curés et idéalement calibrés avec ces données multi-omiques. Nous décrirons 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 détails l'approche itérative développée au moyen de l'étape de curation et de calibration dynamique des modèles.

1.2.1 Présentation des données

1.2.1.1 Données omiques

Génomique Les souches utilisées sont décrites dans (?). Brièvement, la communauté bactérienne contrôlée est composée de deux bactéries lactiques (LAB), *L. lactis* subsp. *lactis* biovar *diacetylactis* CIRMBIA1206 (nommée *L. lactis*), *L. plantarum* CIRMBIA465 (nommée *L. plantarum*), et une bactérie propionique *P. freudenreichii* CIRMBIA122 (nommée *P. freudenreichii*) 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 (?). 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 (?), nous 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 relative à l'expression des gènes, puis une étape de comptage à 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 (?). Une étape supplémentaire de normalisation (du type RPKM) au sein de l'échantillon a été réalisée pour corriger la longueur des gènes et permettre la comparaison. Enfin, la moyenne des réplicats a été calculée.

Métabolomique Les méthodes sont décrites dans (?). Brièvement, les sucres et les acides organiques dans les échantillons ont été quantifiés en utilisant la chromatographie

2. <https://www.biorxiv.org/content/10.1101/2023.05.05.539417v1>

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

liquide à haute performance (HPLC). Les composés volatils 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).

1.2.1.2 Données de calibration

Un suivi cinétique de la croissance bactérienne et du pH a été réalisé sur lait pour chaque souche individuelle. De plus, les concentrations en acides organiques pour *P. freudenreichii* ont été évaluées en condition microaéroophile (Table 1.1). Le milieu de culture de *P. freudenreichii* a été supplémenté en lactate pour reproduire la croissance en co-culture et en hydrolysate 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*. Pour *P. freudenreichii*, *L. plantarum* et *L. lactis*, le début de la phase plateau, correspondant à l'entrée en phase stationnaire, a mis respectivement 48, 14 et 8 heures. 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 1.2).

Données cinétique de croissance et de pH Pour chaque souche nous avons collecté les données de croissance et de pH qui ont servi à calibrer dynamiquement notre modèle. Elles sont représentées dans la table 1.1.

GEM	Temps	Densité bactériennes	pH
<i>P. freudenreichii</i>	0	0.000001	∅
	23	0.000099	∅
	40	0.000561	∅
	48	0.000412	∅
	130	0.000710	∅
<i>L. plantarum</i>	0	1.485e-06	6.7
	5	3.795e-06	6.68
	7	5.28e-06	6.66
	9	8.415e-06	6.555
	14	1.386e-05	6.54
	16	2.3595e-05	6.51
	79	3.762e-05	5.705
<i>L. lactis</i>	0	8.25e-07	6.7
	5	5.230500e-05	6.49
	7	6.831e-05	6.315
	9	6.649500e-05	6.075
	14	6.468e-05	5.935
	16	4.735e-05	5.865
	79	6.171e-05	5.115

TABLE 1.1 – **Données de cultures pures utilisés pour calibrer chaque souche.** Temps en heures et les valeurs de densités bactériennes représentent des moyennes sur les différents réplicats en g.DW⁻¹.

Les données de dosage d'acide Les 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 1.2.

Acides	Concentration T_0	Concentration $T_f = 89$ 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 1.2 – **Données de dosage d'acide en g/L pour *P. freudenreichii*.**

Données de croissance et de pH en communauté Au regard de notre stratégie itérative, les données représentées par la table 1.3 ont servi de données tests pour évaluer la précision du modèle. Toutes ces données proviennent de l'article de (?). 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 log₁₀CFU/g, 8.47 log₁₀CFU/g et 8.59

$\log_{10}\text{CFU/g}$ en approximativement 1200 heures.

GEM	Etapas	Temps (h)	Population bactérienne (\log_{10} CFU/g)
<i>P. freudenreichii</i>	Linoc	0	6.1
	Finoc	18	6.1
	Cm	19.5	7.16
	C0w	60	8.11
	C4w	732	8.56
	C7w	1236	8.59
<i>L. plantarum</i>	Linoc	0	5.2
	Finoc	18	5.4
	Cm	19.5	6.49
	C0w	60	7.92
	C4w	732	8.47
	C7w	1236	8.47
<i>L. lactis</i>	Linoc	0	5.7
	Finoc	18	7.5
	Cm	19.5	8.78
	C0w	60	9.06
	C4w	732	8.95
	C7w	1236	8.45

TABLE 1.3 – **Donnes de croissance en co-culture pour tester les prédictions de notre modèle de communauté**

Etapas	pH
Finoc	6.717
Cm	6.362
C0w	5.35
C4w	5.165
C7w	5.189

TABLE 1.4 – **Données de pH.**

Conversion des unités Lors de la calibration, nous avons appliqué un coefficient multiplicateur de 0.33×10^{-12} transformant les CFU en gramme de matière sèche. En effet, d'après (?), 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 dû également convertir les unités des métabolites que l'on suit en dynamique, de g/kg de lait en mmol/g de lait.

1.2.2 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 (?) version 1.4.1 et enrichi l'annotation existante avec Model-
Polisher (?) version 2.0 en ajoutant des métadonnées liées aux identifiants BIGG depuis
la base de connaissance BIGG (?). Afin de vérifier la qualité des GEM, chaque modèle
métabolique a été validé avec MEMOTE (?) version 0.13.0.

1.2.3 Modélisation et implémentation numérique

Analyse statique

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

$$\begin{aligned} \text{trouver } v^* \in \mathbb{R}^{N_r} \text{ tel que } v^* = & \arg \max_{\substack{S.v = 0 \\ v_{min}^{in} \leq v \leq v_{max}^{in} \\ v_{min}^{ex} \leq v \leq v_{max}^{ex}}} v_{biomasse} \end{aligned} \quad (1.1)$$

ou $v_{biomasse}$ est le flux de la réaction de biomasse de modèle, 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 contenant \mathbb{R}^{N_r} dimension.

L'équation (1.1) 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.

$$\mu_i(v_{min}^{ex}, v_{max}^{ex}) = v^* \quad (1.2)$$

Dans un souci de garantir une solution unique, l'ordre lexicographique (?) a été utilisé, c'est à dire, nous avons défini la direction du flux d'échange des métabolites suivis dynamiquement, définissant de nouvelles contraintes à satisfaire et réduisant l'espace de solution.

Analyse dynamique

Ayant à notre disposition les données cinétiques de chaque souche individuelle ainsi qu'en communauté, nous avons implémenté le dynamisme de cette communauté bactérienne en s'inspirant de l'analyse par équilibre des flux dynamique (dFBA) de (?) et en ajoutant ainsi des mécanismes dynamiques de population au modèle FBA.

Le premier mécanisme que l'on a décrit dynamiquement est le *quorum sensing*, correspondant au processus de régulation de la population sur la croissance de la population. Notons $b = (b_i)$ pour $i \in \mathcal{B} = \{L. lactis, L. plantarum, P. freudenreichii\}$ les densités de la population bactérienne et \mathcal{R}_i le vecteur modélisant ce processus, le *quorum sensing* a été modélisé de la façon suivante :

$$\mathcal{R}_i = \lambda_i \left(1 - \frac{b_i}{\beta_i}\right), \text{ for } i \in \mathcal{B} \quad (1.3)$$

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.

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 de composés métaboliques N_m qui sont expérimentalement suivis, le flux de la borne inférieure $c_{min,i,j}^{ex}$ et supérieure $c_{max,i,j}^{ex}$ ont été définis comme suit :

$$c_{min,i,j}^{ex} = \max\left(-\frac{m_j}{\Delta t \sum_{i \in \mathcal{M}_j} b_i}, v_{i,j}^{int}\right) \quad c_{max,i,j}^{ex} = 0, \quad (1.4)$$

où \mathcal{M}_j est le sous-ensemble de bactéries métabolisant j et Δt est un l'intervalle temporel d'importation. 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 limité par la limite d'import intrinsèque $v_{i,j}^{int}$, il 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 correspondant aux composés des données biochimiques. En s'inspirant de l'approche de (?), nous avons régulé le métabolisme du lactose chez les bactéries lactiques. Elle est ainsi modulée 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, $\mu_{max,lcts}$ et $\mu_{min,lcts}$ sont les valeurs maximales et minimales et ϕ_{undiss} est une fonction calculant la concentration de la forme non dissociée de l'acide lactique à partir du lactate. Cette fonction ϕ_{undiss} est dérivée de l'équation de Henserson-Hasselbalch comme dans (?) et s'écrit

$$\phi_{undiss}(m_{lac_L_e}, m_{lac_D_e}) = \frac{m_{lac_L_e} + m_{lac_D_e}}{1 + 10^{c_1 * (m_{lac_L_e} + m_{lac_D_e}) + c_2}}. \quad (1.5)$$

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-cultures. 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 $\partial_t m_j$ au cours du temps :

$$\partial_t b_i = \mathcal{R}_i(b_i) \mu_{i,i}((c_{min,i}^{ex}, c_{max,i}^{ex})(b, m)) b_i \quad (1.6)$$

$$\partial_t m_j = \sum_{i \in \mathcal{B}} \mu_{i,j}((c_{min,i}^{ex}, c_{max,i}^{ex})(b, m)) b_i \quad (1.7)$$

dans lequel le terme $\mu_{i,j}$ est la composante correspondant au métabolite j (ou à la biomasse i) de μ_i , calculé à partir de 1.2 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. A 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 :

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

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

1.2.4 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) (?) 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. Lors d'une inactivation *in silico* dans le modèle de voies partielles de voies métaboliques, nous avons développé le protocole suivant : détection de fuite de flux symbolisée par des valeurs de flux extrêmes, identification d'activation de voies métaboliques incohérente 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.

1.2.5 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 de minimisation J suivant pour les bactéries lactiques (équations (1.8))

$$J(b_i, pH|\theta_i, b_{i,exp}, pH_{exp}) = \left\| \frac{\log_{10}(b_i) - \log_{10}(b_{i,exp})}{\sigma_{log,i,exp}} \right\|^2 + \alpha \left\| \frac{pH - pH_{exp}}{\sigma_{pH,exp}} \right\|^2 \quad (1.8)$$

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 de i , pour $i \in \{L. lactis, L. plantarum\}$ et α un poids.

Concernant *P. freudenreichii*, la fonction coût cherchant à minimiser l'écart aux données est représentée par l'équation (1.9) :

$$J(b, m|\theta_i, b_{exp}, m_{exp}) = \left\| \frac{\log_{10}(b) - \log_{10}(b_{exp})}{\sigma_{log,b,exp}} \right\|^2 + \alpha \left\| \frac{m - m_{exp}}{\sigma_{m,exp}} \right\|^2 \quad (1.9)$$

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 1.5 :

GEM	Paramètres	Explication
<i>P. freudenreichii</i>	λ_i	Modification des bornes d'import des nutriments suivi 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 suivi 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 1.5 – Ensemble des paramètres optimisés pour chaque modèle individuel avec leur fonction.

1.3 Application de la méthode numérique sur 3 souches du fromage

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 hétérogènes nous ont permis de calibrer dynamiquement nos 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 M2M.

1.3.1 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 depuis leurs séquences génomiques annotées respectives. 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 de réactions d’échanges de composés qui n’apparaissent pas dans la composition du lait, sont ainsi mises à 0. La 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 dihydroxyacetone (*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 and 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 pentose phosphate sont communes aux trois espèces et 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 Figure 1.4,1.2,1.3. Dans la suite du chapitre, les caractéristiques métaboliques de chaque souche seront étudiées afin de justifier les raffinements faits, 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 (?). 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 (?). 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 (?). 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 1.2 et 1.3.

L. lactis Pour permettre la production de certains composés d’intérêt, comme le butanediol, le GEM a été manuellement curé (Fig. 1.2). En reprenant la méthode de

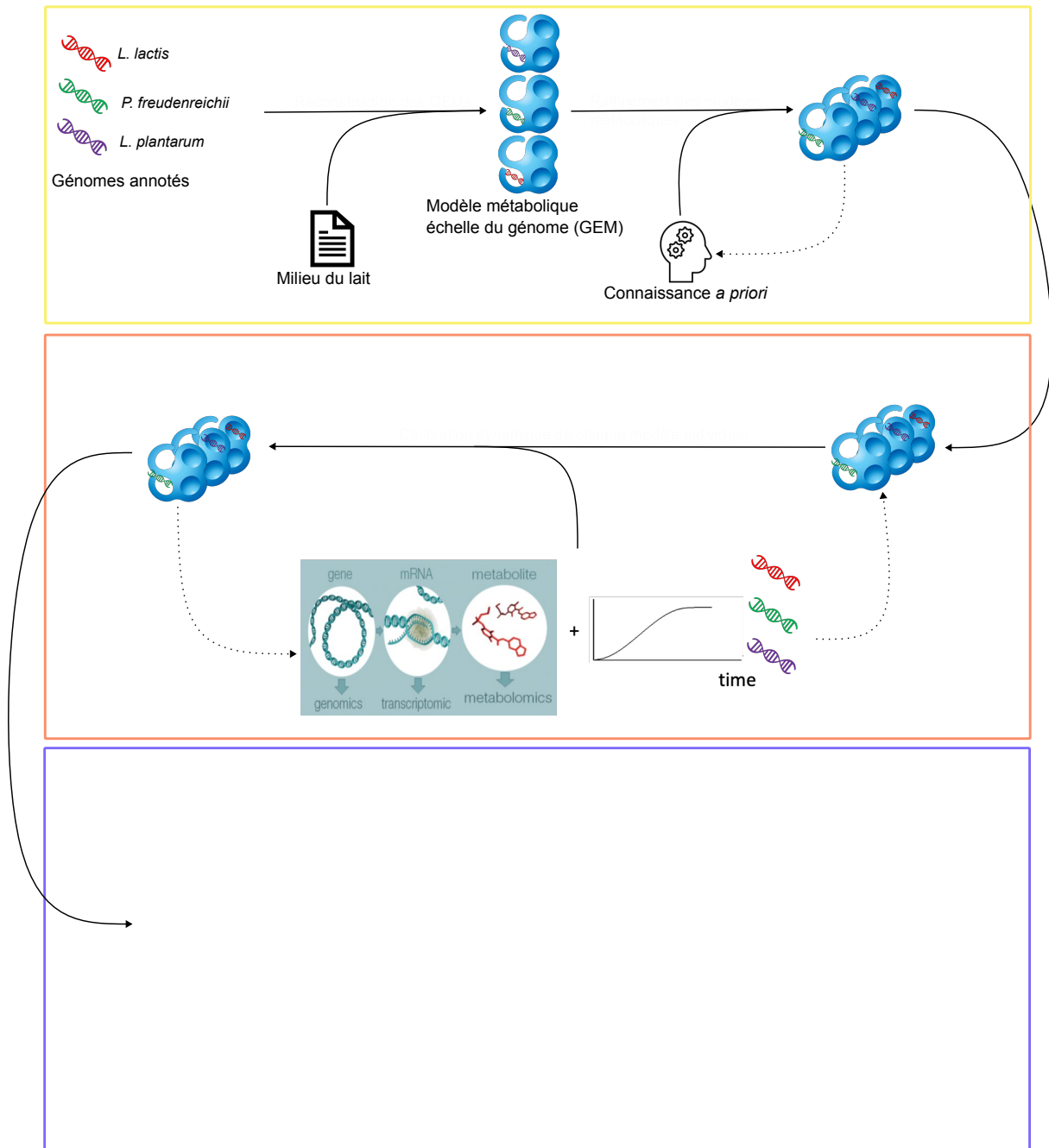


FIGURE 1.1 – Récapitulatif des étapes de raffinement et de calibration des modèles métaboliques, des simulations ainsi que des données utilisées

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 (???). 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. 1.2).

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 metatranscriptomiques, *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. 1.5 a).

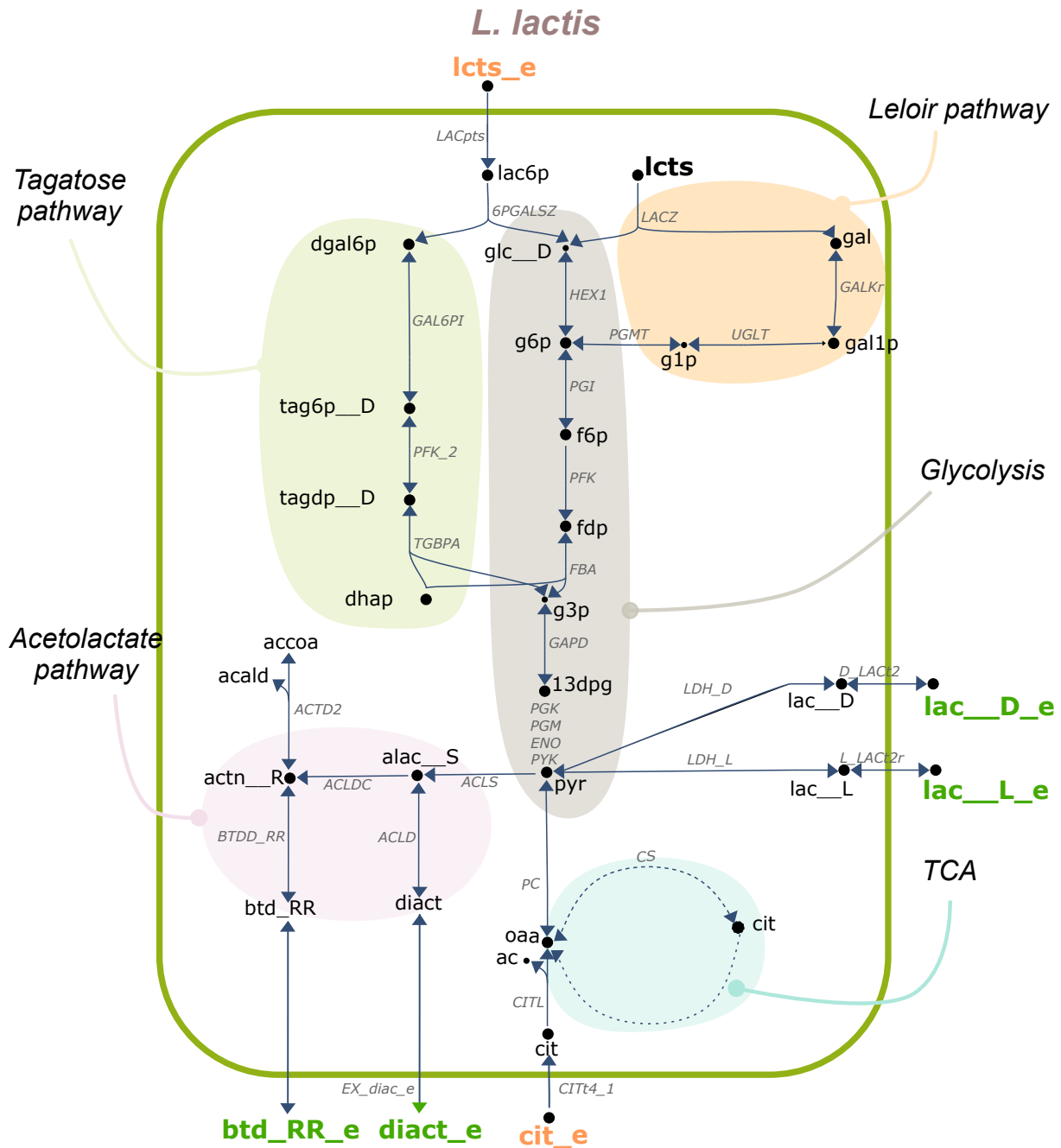


FIGURE 1.2 – Cartes des voies métaboliques du métabolisme carboné central d'intérêt de *L. lactis*. La voie métabolique de la glycolyse est représentée en gris, celle de Leloir en saumon, acétolactate en rose, le cycle du TCA en bleu clair, et le tagatose en vert. Les métabolites en orange et vert correspondent aux entrées et sorties du modèle métabolique. La liste des métabolites et des réactions est répertoriée à la fin du chapitre à la section 1.4

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 1.3). 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

P. freudenreichii La souche CIRM-BIA122 de *P. freudenreichii* consomme à la fois l'acide lactique, sa principale source de carbone (??) et/ou le lactose (?). L'acide lactique et le lactose sont catabolisés par les voies de la glycolyse et des pentose-phosphate, menant à la production de pyruvate. Une part du pyruvate alimente le cycle des acides tricarboxylique (TCA) et la voie de Wood-Werkman (Fig. 1.4) 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 (?) 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, ce qui est confirmé par ratio de 2 trouvé dans les travaux de (????).

L'ensemble des modifications portées à chaque modèle métabolique est résumé dans la table 1.7.

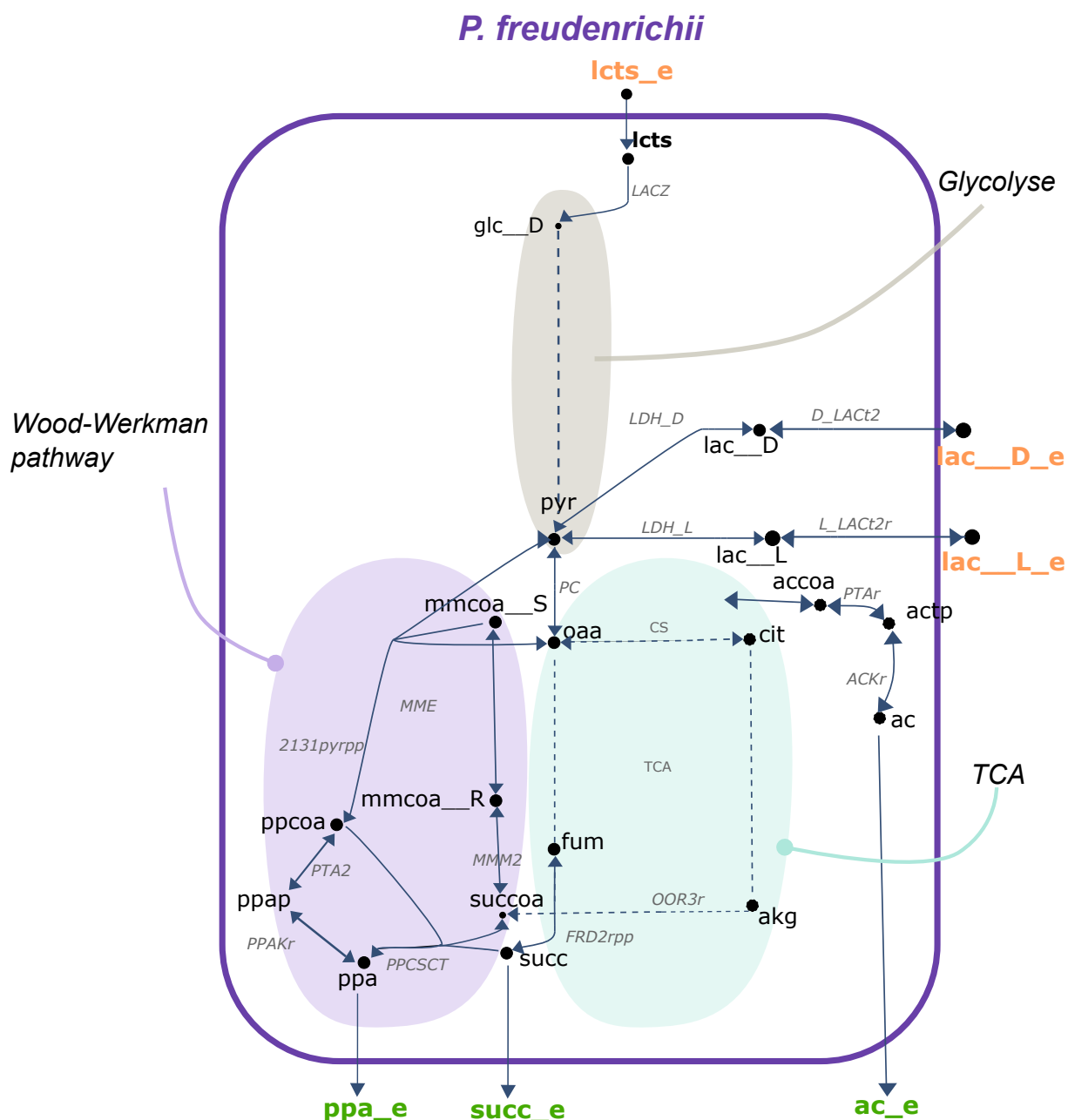


FIGURE 1.4 – Cartes des voies métaboliques du métabolisme carboné central d'intérêt de *P. freudenreichii*. La voie métabolique de la glycolyse est représentée en gris, le cycle du TCA en bleu clair, et la voie de Wood-Werkman en violet. Les métabolites en orange et vert correspondent aux entrées et sorties des modèles métaboliques. La liste des métabolites et des réactions est répertoriée à la fin du chapitre à la section 1.4

Nous avons par la suite vérifié la qualité de chaque réseau métabolique : une faible fraction de réactions universellement bloquées ($< 3\%$), et une faible fraction de réactions sans association de gène-proteine-réactions (GPR) ($< 30\%$) (?). Puis, nous les avons analysés par la méthode d'équilibre des flux (FBA) (??) assurant ainsi la production des molécules de la biomasse à partir de l'environnement nutritionnel du lait. Nous avons obtenu des taux de croissances 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 (?), 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 dans le compartiment cytosolique. Les caractéristiques générales des trois modèles sont décrites dans la Table 1.6. 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 (?).

	<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 $\text{mmol.gDW}^{-1}.\text{hr}^{-1}$	0.187	0.645	2.172

TABLE 1.6 – Caractéristiques générales des 3 modèles métaboliques

Le raffinement des trois modèles FBA a permis la prédiction de production et de consommation de plusieurs composés d'arômes tels que les acides lactique, acétique et propionique, le diacétyl et le butanediol (??) (Fig. 1.4,1.2,1.3). De plus, 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

Afin d'obtenir des modèles individuels suffisamment bien curés, nous avons utilisé les données cinétiques à disposition pour affiner chaque modèle (voir 1.2.1.2). Cette étape de calibration consiste à minimiser l'écart entre les données expérimentales et la simulation de façon dynamique en estimant des paramètres (voir 1.7). 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 que les concentrations des métabolites pour *P. freudenreichii*.

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
	GALUi	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
	PPAKr	borne inférieure borne supérieure	0 0	Inhibe la surproduction de l'acétate
	LACZ	borne inférieure borne supérieure	0 5	Autorise la consommation de lactose
	PTAr	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 1.7 – Raffinement manuel des GEMs avec les justifications correspondantes

Pour estimer les paramètres, nous avons résolu le problème de minimisation décrit dans les équations 1.8 pour les bactéries lactiques et 1.9 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 se 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. 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 lactique, les bornes supérieures des réactions d'échange du lactate EX_lac__L_e et EX_lac__D_e, sont modulées comme suit :

$$\min\left(\frac{m_{lcts_e}}{\Delta t * \sum_{i \in \mathcal{M}(lcts_e)} b_i}, \mu_{max,lcts} * 10^{(-k_{lac} * \phi_{undiss})} + \mu_{min,lcts}\right) * 4 \quad (1.10)$$

Toutes les constantes sont décrites dans la sous section 1.2.3. Pour rappel, m_{lcts_e} est la concentration du lactose, $\mathcal{M}(lcts_e)$ le sous ensemble des bactéries métabolisant le lactose, $\mu_{max,lcts}$ et $\mu_{min,lcts}$ valeurs maximales et minimales 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. (1.11)). 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 $v_{i,ac_e}^{exp} = 0.5$ 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 (?) dans lesquelles l'acidité libérée par les bactéries lactiques est représentée par la 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 :

$$\max\left(-\frac{m_{lcts_e}}{\Delta t * \sum_{i \in \mathcal{M}(lcts_e)} b_i}, -\mu_{max,lcts} * 10^{(-k_{lac} * \phi_{undiss})} - \mu_{min,lcts}\right) \quad (1.11)$$

Concernant la régulation par le pH, la quantité d'acide lactique non dissocié est calculé à partir de la concentration de lactate au moyen de la fonction ϕ_{undiss} (voir équation 1.5 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 :

$$\mu_{i,j} = \alpha \frac{m_j(T) - m_j(0)}{\int_0^T b_i(t) dt} \quad (1.12)$$

où $\mu_{i,j}$ correspond à la valeur de flux, m_j la concentration du métabolite, T valeur finale, 0 valeur initiale, $b_i(t)$ la densité bactérienne, 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 limitant, ainsi nous avons modéliser la phase plateau par une croissance logistique.

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

$$\max\left(-\frac{m_{lac_{L,e}}}{\Delta t * \sum_{i \in \mathcal{M}(lac_L)} b_i}, F_{lac} * K_{freud} * \frac{b_{freud}}{B_{p,freud}}\right) \quad (1.13)$$

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 consommatrice 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 1.12) 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. Soient 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 1.12). Le même terme de modulation que dans l'équation 1.13 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 modélisé par $v_{i,j}^{int} = -8$ pour $j = L. lactis$ 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 1.5, 1.6 et 1.7. 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 1.5a), 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. 1.5a). 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 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. 1.3). 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 transkétole, 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* (?). 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. 1.7a). Cependant, nous retrouvons une activation simultanée des cycles du TCA et de Wood-Werkman (?), 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 1.5, 1.6, 1.7 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 1.5, 1.6 (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, notre modèle métabolique de *P. freudenreichii* retrouve les concentrations finales prédites par les expériences biologiques (Figure 1.7(c)). Afin de vérifier statistiquement que l'apprentissage de nos modèles est correct, nous avons représenté sur la figure Fig. 1.8 (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 obtenir des modèles métaboliques complets 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 en créant un modèle de communauté

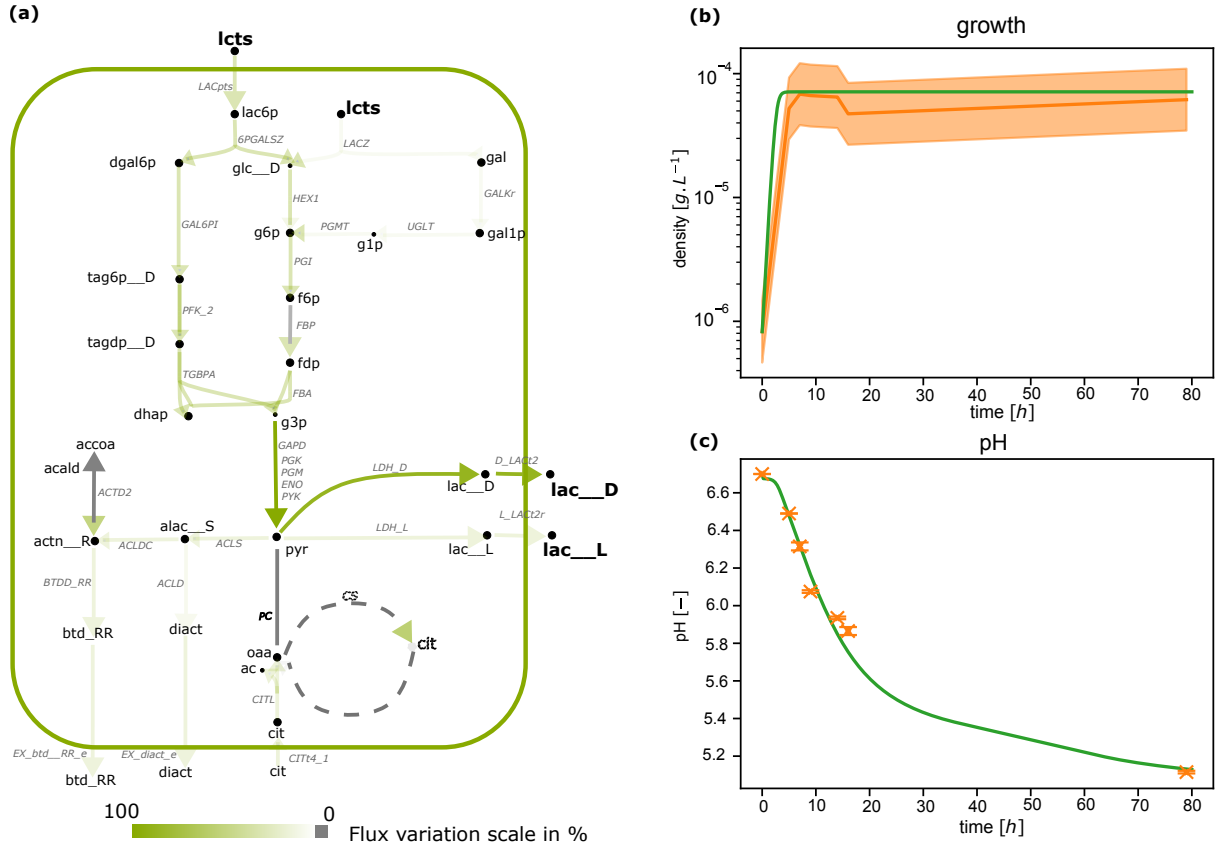


FIGURE 1.5 – Réseau métabolique de *L. lactis* après calibration sur les données de culture pure. (a) Représentation d'un FBA appliqué sur le métabolisme carboné central lors de l'inoculation de *L. lactis*. L'échelle de couleur représente les valeurs de flux prédites par le FBA et normalisées par la plus haute valeur de flux prédites 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 (courbe orange, with $\pm 1/4$ log, which is the admitted range of precision for plating numbering). (c) Courbes de pH dans des conditions de culture pure simulé par le modèle (courbe verte) et issues des données expérimentales (croix orange, with standard deviation). La liste des métabolites et des réactions est répertoriée à la fin du chapitre à la section 1.4

afin de prédire les croissances de chaque souche en communauté, 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.

1.3.2 Simulations à l'échelle de la communauté

En premier lieu, nous avons créé un modèle de communauté où 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 chaque modèle individuel en intégrant la valeur de "quorum sensing" des données expérimentales en co-culture (voir la Table 1.2.1.2 et l'équation 1.3 dans le but de reproduire le comportement métabolique de la communauté *in silico* (voir eq. 1.6-1.7). Ce modèle de communauté

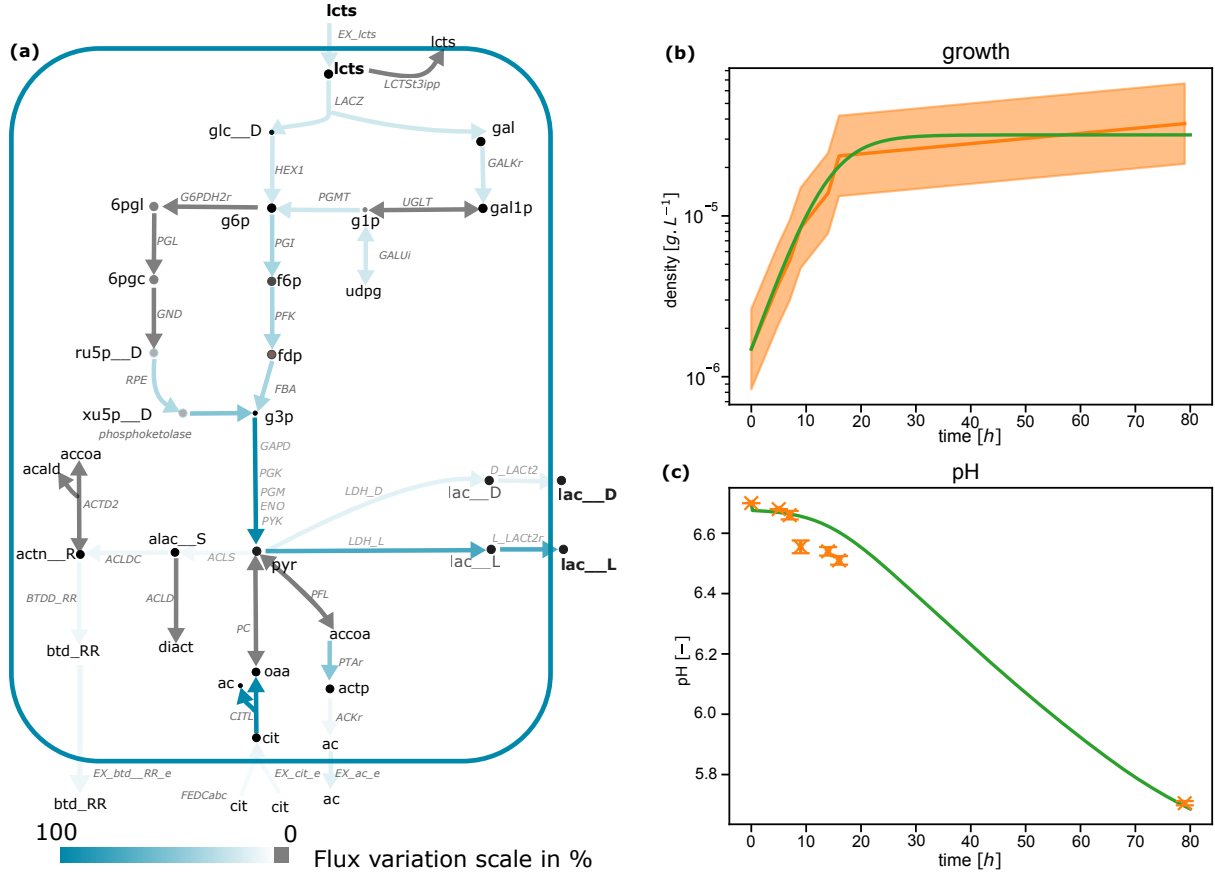


FIGURE 1.6 – Réseau métabolique de *L. plantarum* après calibration sur les données de culture pure. (a) Représentation d'un FBA appliqué sur le métabolisme carboné central lors de l'inoculation de *L. plantarum*. L'échelle de couleur représente les valeurs de flux prédites par le FBA et normalisées par la plus haute valeur de flux prédites 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 (courbe orange, with $\pm 1/4$ log, which is the admitted range of precision for plating numbering). (c) Courbes de pH dans des conditions de culture pure simulé par le modèle (courbe verte) et issues des données expérimentales (croix orange, with standard deviation). La liste des métabolites et des réactions est répertoriée à la fin du chapitre à la section 1.4

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 utilisé afin d'affiner les prédictions. Ces données vont en revanche être consacrées à la validation des prédictions du modèle.

Les simulations de la croissance, du pH et de la métabolomique du modèle de communauté ont correctement expliqué les valeurs expérimentales (Fig. 1.9(a)-(c), Table supplémentaire 1.3). 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 d'une sur-acidification du fromage durant les étapes de moulage et de démou-

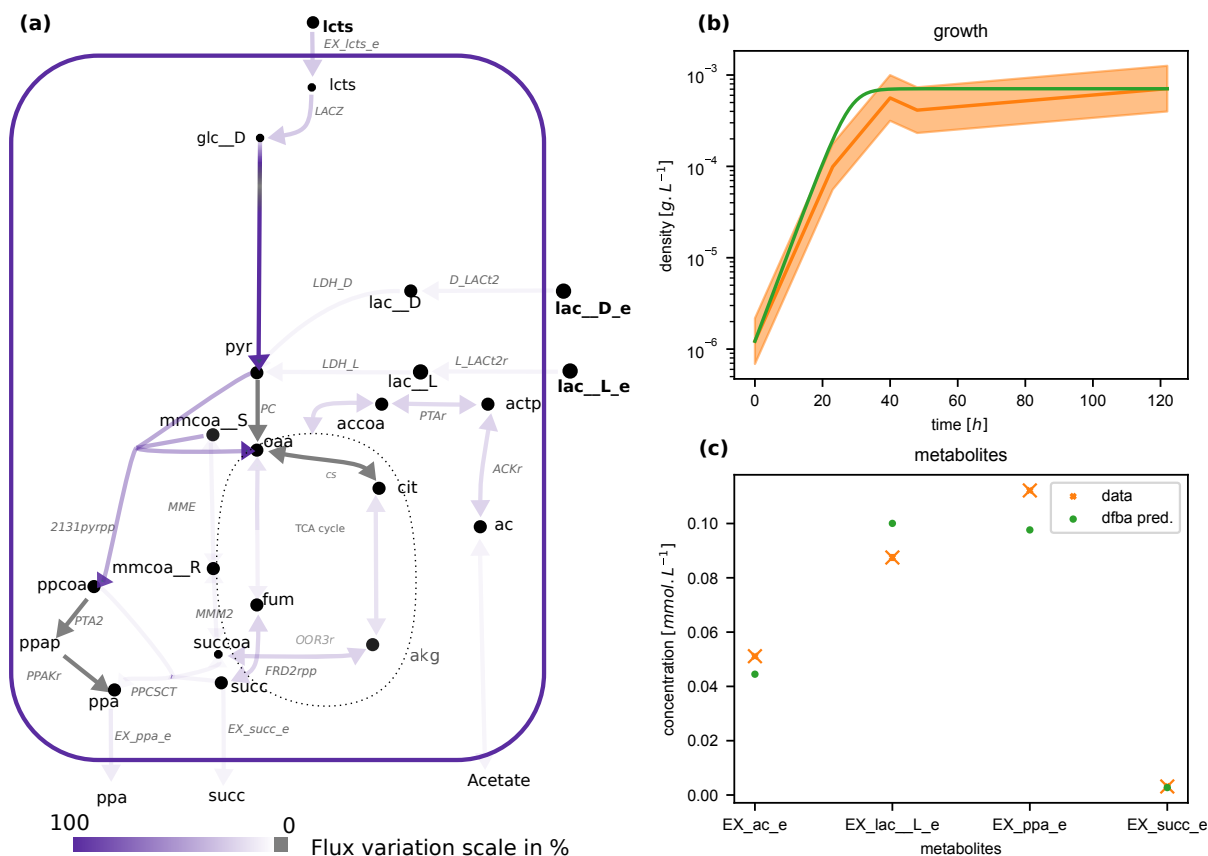


FIGURE 1.7 – Réseau métabolique de *P. freudenreichii* après calibration sur les données de culture pure. (a) Représentation d'un FBA appliqué sur le métabolisme carboné central lors de l'inoculation de *P. freudenreichii*. L'échelle de couleur représente les valeurs de flux prédites par le FBA et normalisées par la plus haute valeur de flux prédites 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 (courbe orange, with $\pm 1/4$ log, which is the admitted range of precision for plating numbering). (c) Concentrations de lactate, acetate, propionate et succinate simulées (points verts) observées (croix en orange avec un écart-type) au temps final. La liste des métabolites et des réactions est répertoriée à la fin du chapitre à la section 1.4

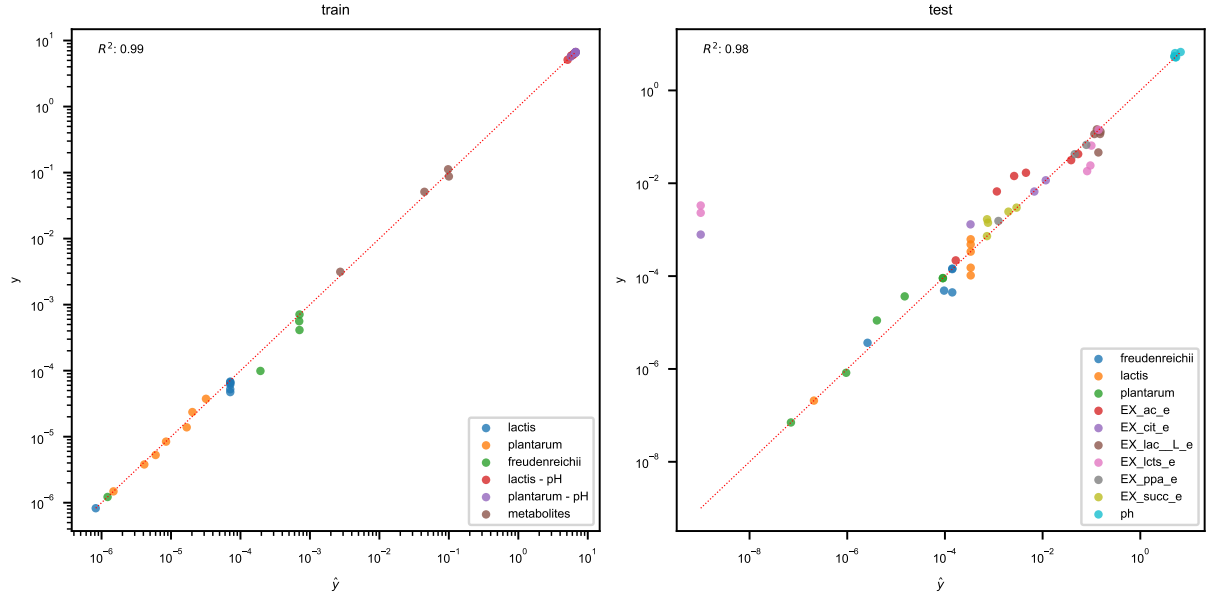


FIGURE 1.8 – **Validation des prédictions après apprentissage.** Les valeurs simulées issues des modèles (\hat{y}) et celles des données expérimentales respectives (y) lors de la phase d'apprentissage sur les données de cultures pures (a) et lors de la phase de test en communauté (b). La couleur représente le type de la donnée et la bissectrice est affichée par une ligne en pointillée 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.

lage. 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 1.10). 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 1.8, b).

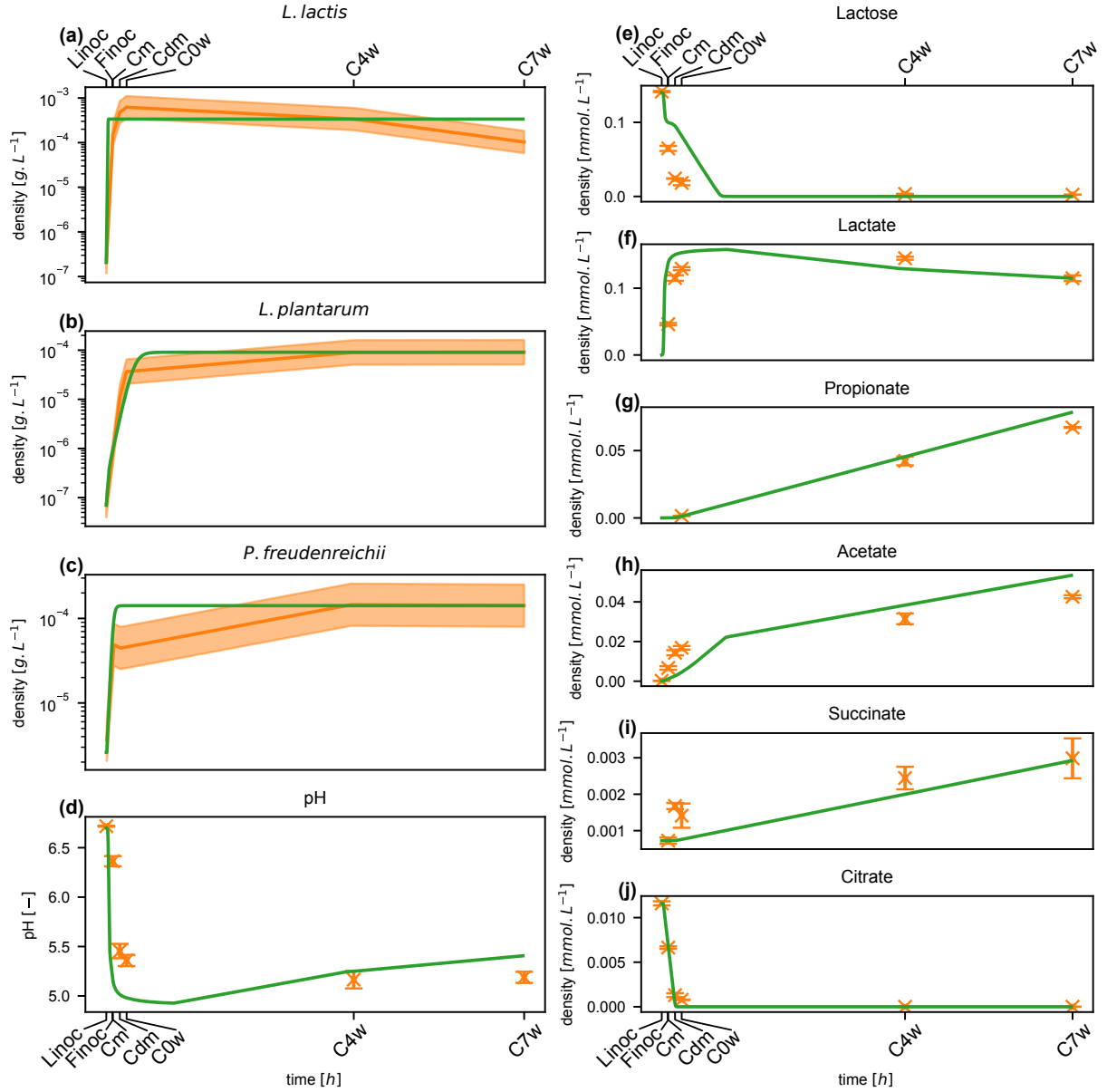


FIGURE 1.9 – (a-c) Courbes de croissance de respectivement *L. lactis*, *P. freudenreichii* and *L. plantarum* calculées par le modèle (ligne verte) et celles observées expérimentalement (en orange). (d-j) Profil métabolique et du pH en condition de co-culture calculé par le modèle et expérimentalement 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).

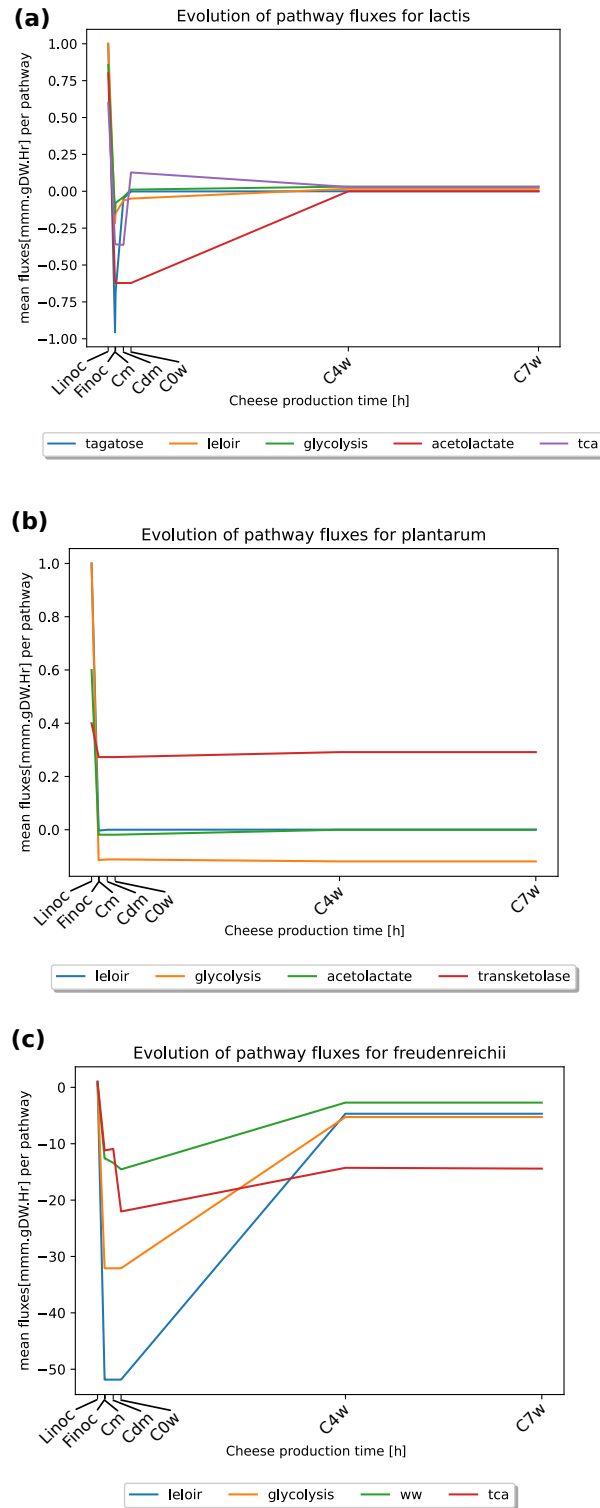


FIGURE 1.10 – Illustration du changement de voies métaboliques durant le processus de production du fromage. Pour chaque bactérie nous avons lancé une simulation dynamique (dFBA) et les flux réactionnelles pour chaque voies métaboliques d'intérêt des figures 1.2,1.3,1.4 on été retrouvé à l'incoculation de bactéries lactiques (Linoc, t=0 heure), de *P. freudenreichii* (Finoc, t=18 heures), au moulage (t=19.5 heures), demoulage (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 normalisés les flux par la valeur de flux lors de l'inoculum.

En somme, nous avons montré qu'utiliser uniquement des modèles métabolique curés et calibrés individuellement permet de retrouver les observations en communauté. Malgré les deux étapes itératives coûteuses en temps, l'interactivité entre modèle numérique et 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 du point de vue des interactions bactériennes possibles au sein de cette communauté. Dans les paragraphes suivant, 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.

1.3.3 Interactions bactériennes et exploration du modèle de communauté

Notre modèle numérique de communauté peut révéler des interactions microbiennes du type syntrophiques, dans laquelle une espèce se nourrit des produits de l'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. 1.11, (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. Concernant le lactose et le citrate, nous observons que *L. lactis* les consomment très rapidement 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 ce dernier, 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 sont mis 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.

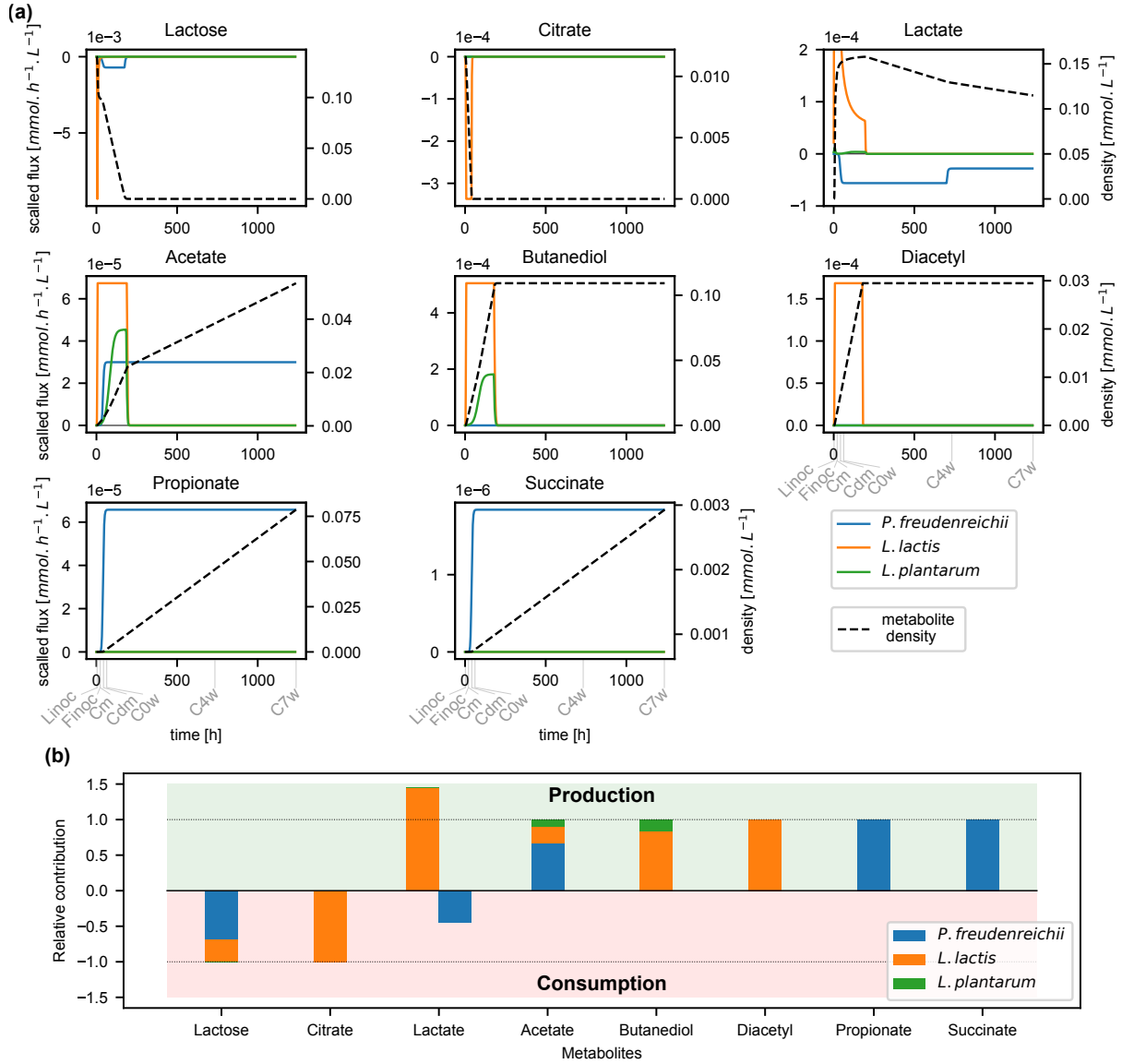


FIGURE 1.11 – (a) **Flux dynamiques de consommation et de production de métabolites d'intérêts.** Nous représentons pour chaque bactérie les flux dynamiques de consommation et de production de chaque métabolites que l'on en simulation (lignes colorées). La dynamique de la concentration de ces mêmes métabolites dans le fromage est aussi représenté par les traits en pointillés. (b) **Contribution relative des bactéries dans le devenir des métabolites.** Nous avons calculé pour chaque souche sa contribution totale dans la consommation et la production de ces métabolites et intégré en temps par $\int_0^t \mu_{i,j} b_i$ pour le métabolite j et la bactérie i , voir eq.1.7). Par la suite, nous avons normalisé le résultat par la somme des bactéries. 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 ($t = 18$ hours) ; 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, 4 semaines d'affinage ($t = 1236$ hours).

La production de lactate étant corrélée avec la consommation de lactose, nous observons une très forte production par *L. lactis* pendant 250 heures, correspondant à l'affinage

dans le processus de fabrication du fromage. Malgré la plus faible croissance de *L. plantarum*, ce dernier contribue peu à la production de lactate. Durant la phase de l'affinage, le lactate est consommé par *P. freudenreichii* contribuant ainsi à l'augmentation du pH dans la communauté comme observé dans la Figure 1.9 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, temps à partir duquel la phase plateau commence. Ce schéma est retrouvé chez les bactéries lactiques intervenant également dans la production du butanediol *via* la voie de l'acéto-lactate. 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. 1.11, (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. A cette fin, nous avons utilisé deux méthodes basées sur les flux, SMETANA (?) et MiCOM (?), et une approche qualitative basée sur le raisonnement, Metage2Metabo (?) 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 échangés (Table 1.8 et 1.9), alors que l'approche par raisonnement, Metage2Metabo, identifie 11 métabolites qui ne peuvent pas être produits par une espèce sans interactions dans la communauté (Table 1.10).

Une première observation est qu'un nombre assez important (11) de métabolites sont communs dans les prédictions des composés échangeables de SMETANA et MiCOM : lactate, qui est aussi prédit dans le modèle dynamique, phénylalanine, serine, malate, succinate, xanthine, H_2S , 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 H_2S 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 methodes et Figure 1.12). Les résultats suggèrent que les interactions pour H₂S, ribose et le glycerol sont plausibles à plusieurs étapes de la fabrication du fromage. A 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	acetaldéhyde	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 1.8 – Métabolites échangeable 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	acetalaldéhyde	Aldehydes-Or-Ketones, Aldehydes	Pf 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 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 1.9 – Métabolites échangeable sur la communauté du fromage identifiés par SMETANA. Abréviations : Ll, *L. lactis* ; Lp, *L. plantarum*, Pf, *P. freudenreichii*

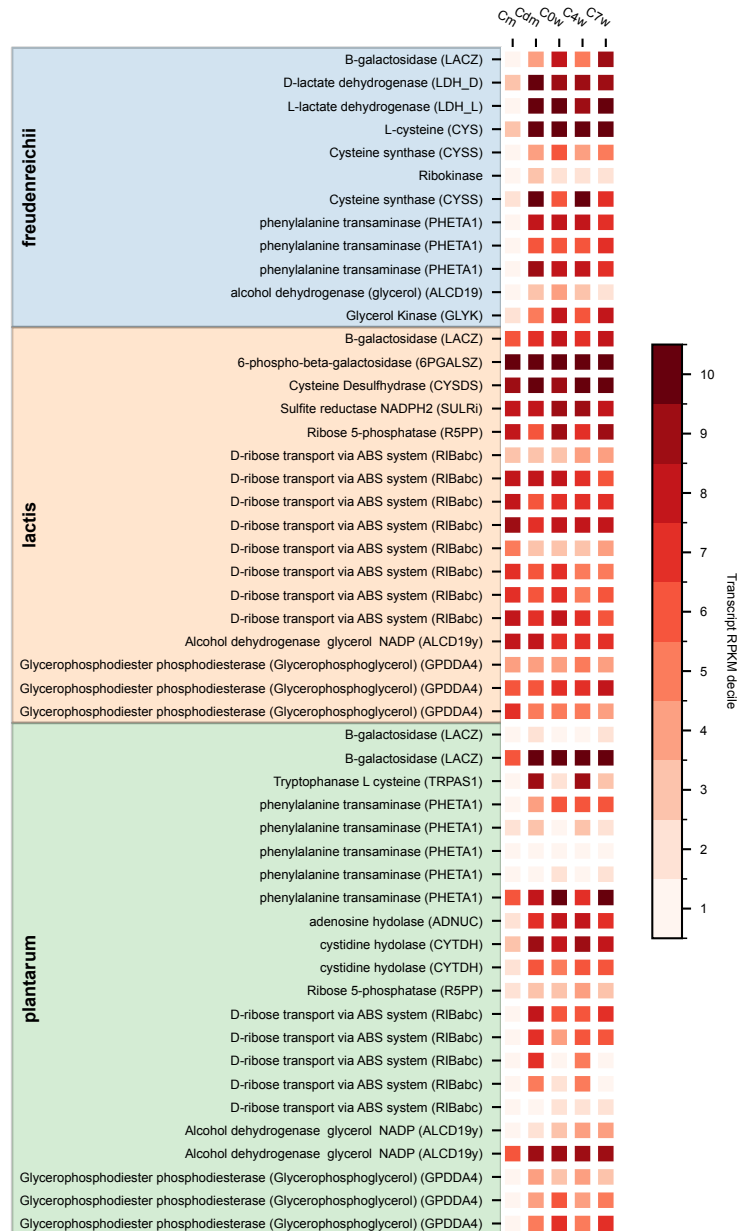


FIGURE 1.12 – **Expression des gènes des données métatranscriptomiques** Cette carte thermique affiche le décile du nombre de RPKM de transcrits de gènes à un moment donné (colonnes) et dans un micro-organisme donné (les zones colorées indiquent à quelle bactérie le gène appartient). Les déciles ont ensuite été codés par couleur, 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).

Enfin, 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 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 ø	glyceraldehyde ø	ø
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 1.10 – Liste de métabolites prédits par Metage2Metabo comme étant productibles uniquement par coopération métabolique

1.4 Discussion et Conclusion

Nous avons fourni 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 inoculées 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 peu de paramètres pour chacun, limitant ainsi le sur apprentissage du modèle durant les simulations communautaires. 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, de la littérature scientifique et des données multi-omiques étaient nécessaires pour accomplir 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 (???). Pour la première fois, l'utilisation du citrate et de la fermentation hétérolactique ont été observé dans le modèle curé de *L. plantarum* en lait. La vérification de modèle métabolique individuel et de l'activation des différentes voies métaboliques décrites plus haut représentait ainsi une étape importante pour valider les modèles, permettant la reproduction d'activités enzymatiques observées dans (??). La bactérie propionique *P. freudenreichii* convertit le lactose, et préférentiellement le lactate en propionate d'après (???). Nous avons ajouté l'enzyme propionyl-CoA :succinate CoA transférase (2.8.3.-) pour compléter le cycle de Wood-Werkman, et ainsi assurer la production de propionate.

Dans (?), les paramètres individuels de calibration pour chaque nutriment du milieu de culture sont imposés dans le but d'expliquer la métabolomique à l'échelle de l'espèce, et de prédire les composés de la métabolomique à l'échelle de la communauté. De plus, 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 (?), qui a utilisé des paramètres de calibration spécifiques aux souche 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. Il est montré que la production de propionate peut être attribuée uniquement à *P. freudenreichii*, comme indiqué dans (?), 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 régulés 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 de milieu, réprimant probablement le métabolisme de *L. plantarum* (voir equation sur la régulation du pH sur les bactéries lactiques). *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, la famille des acides aminés était bien représentée.

Dans le modèle de co-culture, *P. freudenreichii* garde son métabolisme activé dans la dernière phase de production d'un fromage : l'affinage. Ce métabolisme est reflété par la consommation de lactate durant tout l'affinage, et le relargage de composés organoleptiques à flux constant (Figure 1.11). 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 stoppant 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 1.12)

Ce chapitre a révélé comment le processus itératif, la combinaison d’expertise biologique, de données hétérogènes et de la modélisation métabolique permet 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 potentiel 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 du métabolisme. Un dernier chapitre se consacrera 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