



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

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

Mention Informatique
École doctorale Matisse

présentée par
Oumarou ABDOU ARBI

*préparée à l'unité de recherche : UMR 6074 IRISA
 Institut de Recherche en Informatique et Système Aléatoires
 Composante Universitaire : IFSIC*

Intitulé de la thèse :

Etude de la variabilité

des contributions de

nutriments à un réseau

métabolique :

modélisation, optimisation

et application en nutrition

**Thèse soutenue à Rennes
 le 30/09/2013**

devant le jury composé de :

Mireille REGNIER

Directrice de recherche, INRIA, Saclay / *Rapporteuse*

Daniel KAHN

Directeur de recherche, INRA, Université Lyon 1, LBBE /
Rapporteur

Rumen ANDONOV

Professeur, Univ. Rennes 1, IRISA / *Examineur*

Marie BEURTON-AIMAR

Maître de conférence, Univ. Bordeaux, LABRI /
Examinatrice

Jérémie BOURDON

Maître de conférence, Univ. Nantes, LINA / *Examineur*

Sophie LEMOSQUET

Chargé de recherche, INRA, Centre Inra Pégase /
Examinatrice

Anne SIEGEL

Directrice de recherche, CNRS, IRISA /
Directrice de thèse

Théodore TAPSOBA

Professeur, Université polytechnique de Bobo-Dioulasso
 (Burkina faso) / *Co-directeur de thèse*

Remerciements

Je tiens ici à remercier très sincèrement Anne Siegel, DR CNRS, qui a accepté de diriger cette thèse. Je remercie également Théodore Tapsoba, Professeur à l'université polytechnique de bobo-dioulasso, co-directeur de cette thèse.

Je remercie Mireille Regnier, DR INRIA, Inria Saclay, et Daniel Kahn, DR Inra, Université Lyon 1, LBBE, d'avoir bien voulu accepter la charge de rapporteur.

Je remercie Rumen Andonov, Professeur à l'université de Rennes 1, IRISA, Marie Beurton-Aimar, Maître de conférence à l'université de Bordeaux, LABRI, d'avoir bien voulu juger ce travail.

Je remercie Jérémie Bourdon, Maître de conférence à l'université de Nantes, LINA, et Sophie Lemosquet, CR Inra, Centre Inra Pégase qui ont beaucoup participé à la réussite de ce travail.

Je remercie Marie-Francoise Roy, Professeur à l'Université de Rennes1, qui a également participé à la réalisation de cette thèse qu'elle trouve ici toute ma gratitude et mon respect.

Mes remerciements également aux équipes Symbiose, Dyliss, Genscale, Genouest à Rennes et NT-AGATA au Burkina Faso chez qui j'ai préparé cette thèse.

Enfin, je remercie ma famille qui m'a énormément soutenu à la réussite de ce projet. Qu'elle reçoive ici toute ma reconnaissance.

Table des matières

Table des matières	0
1 Introduction	5
Introduction	5
1.1 Réseaux métaboliques	5
1.2 Modélisation du métabolisme	10
1.2.1 Modélisation structurelle	10
1.2.2 Modélisation dynamique	10
1.2.3 Modélisation à base de contraintes	13
1.2.3.1 Modes élémentaires et voies extrêmes	13
1.2.3.2 <i>Flux Balance Analysis</i> (FBA)	14
2 Analyse de la distribution de flux dans un réseau métabolique en nutrition	21
2.1 Les réseaux métaboliques entrées/sorties	21
2.2 Un exemple de réseau entrées/sorties : la glande mammaire	22
2.3 Etude du métabolisme de la glande mammaire par les méthodes issues de la littérature	28
2.3.1 La caleulette métabolique	28
2.3.2 Modélisation dynamique	29
2.3.3 Maximisation d'ATP	29
2.4 Explorer les distributions extrémales connaissant le bilan d'ATP	31
2.4.1 Etude de la flexibilité à ATP fixé	31
2.4.2 Identification des distributions extrémales	32
2.4.3 Validation biologique	33
3 Contribution des entrées (Inputs) dans les sorties (Outputs) : AIO	37
3.1 Problème posé	37
3.2 Formalisation et modélisation des AIOs	39
3.3 Calcul des AIOs par la méthode du parcours des graphes	45
3.4 Calcul des AIOs dans un cas général	46
3.5 Propriétés	50

3.6	Algorithme général de calcul des AIOs	51
3.7	Application : calcul des AIOs de sommets extrémaux	52
4	Algorithme de recherche locale pour explorer la flexibilité dans le cas des organismes complexes	57
4.1	Problème posé	58
4.2	Résolution par une méthode de points intérieurs	59
4.2.1	Points intérieurs	60
4.2.2	Implémentation et application	62
4.2.3	Limites du processus calcul formel et méthodes de résolution existantes	62
4.2.3.1	Complexité des méthodes basées sur le calcul formel de l'inverse d'une matrice	63
4.2.3.2	Distance des solutions optimales par rapport au bord de l'espace	65
4.3	Algorithme de recherche locale pour calculer les optimums des AIOs . . .	66
4.3.1	Modélisation	66
4.3.1.1	Centre de Chebyshev d'un polyèdre	66
4.3.1.2	Discrétisation du bord du polyèdre	68
4.3.1.3	Calcul des coordonnées des points du bord du polyèdre	68
4.3.2	Principe de l'algorithme	69
4.3.3	L'algorithme	69
4.3.4	Intérêt fondamental de l'algorithme de recherche locale	71
4.4	Application biologique	72
4.4.1	Tables des résultats	72
4.4.2	Distances des solutions optimales par rapport au bord de l'espace de recherche	77
5	NutritionAnalyzer : un logiciel bio-informatique d'analyse de réseau métabolique	81
5.1	Contexte et motivations	81
5.2	Analyse d'un réseau métabolique par l'approche du calcul des AIOs . . .	82
5.2.1	Les différentes approches	82
5.2.2	Calcul de AIO d'une distribution extrémale	84
5.2.3	Architecture des différents fichiers pour l'analyse d'un réseau métabolique entrées/sorties par les AIOs	87
6	Conclusion et perspectives	91
6.1	Conclusion	91
6.2	Perspectives	92
	Table des figures	95

A Générer l'ensemble des sommets extrémaux d'un polyèdre	97
A.1 Algorithme du simplexe	97
A.1.1 Définitions et Notations	98
A.1.2 Présentation de l'algorithme	98
A.1.2.1 L'algorithme du simplexe proprement dit : la phase 2 . .	99
A.1.2.2 Initialisation : la phase 1	99
A.2 Algorithme de calcul sommets d'un polyèdre	101
B Données de la glande mammaire	103
B.1 Les différentes réactions stœchiométriques faisant intervenir les cofacteurs et les pivots	103
B.1.1 Bilans d'utilisation des pivots	105
C Publications	107
Bibliographie	157

Chapitre 1

Introduction

Nous présentons dans ce chapitre, les réseaux métaboliques et les différentes méthodes de modélisation utilisées pour les analyser. Nous verrons en quoi ces méthodes permettent de comprendre le métabolisme à l'échelle d'un organe comme celui de la glande mammaire.

1.1 Réseaux métaboliques

La cellule est l'unité structurale et fonctionnelle de tout être vivant (excepté les virus). Dans le cas des organismes multi-cellulaires, chaque cellule est associée à une fonction bien déterminée. Les cellules de même type sont réunies en tissus, eux mêmes réunis en organes. Les organes à leur tour forment l'organisme vivant.

La cellule est le siège de nombreuses réactions métaboliques. Ces réactions permettent à un organisme de transformer des substrats en produits, lui permettant ainsi de croître, de se reproduire, de maintenir sa structure et d'interagir avec son environnement. Ces transformations sont rendues possibles par des enzymes. Chaque cellule contient plusieurs centaines d'enzymes de types différents et chaque enzyme catalyse une réaction spécifique à un substrat spécifique. La Figure 1.1 montre le processus de transformation de substrats en produits. Le substrat se fixe à un site particulier de l'enzyme appelé *site actif* pour former un complexe enzyme-substrat. Le substrat se mue en un ou plusieurs produits qui se dégageront du site actif. Une suite de réactions enzymatiques liée à un produit ou substrat est appelée *voie métabolique*. Un exemple de voie métabolique, la plus répandue au sein du monde du vivant, est la *glycolyse*. Elle transforme le glucose en Acétyl CoA pour produire de l'énergie via le *cycle de krebs* (Figure 1.2).

Le métabolisme cellulaire correspond donc à l'ensemble des réactions métaboliques qui peuvent avoir lieu dans une cellule. Il se compose d'un grand nombre de réactions et de métabolites. Par exemple, pour *Escherichia coli* qui est un des organismes les plus étudiés, plus de 1400 réactions différentes ont été identifiées [KCVSZ⁺11]. La Figure 1.3 ci-dessous donne le nombre de réactions et de métabolites qui interviennent dans des organismes étudiés depuis 1999.

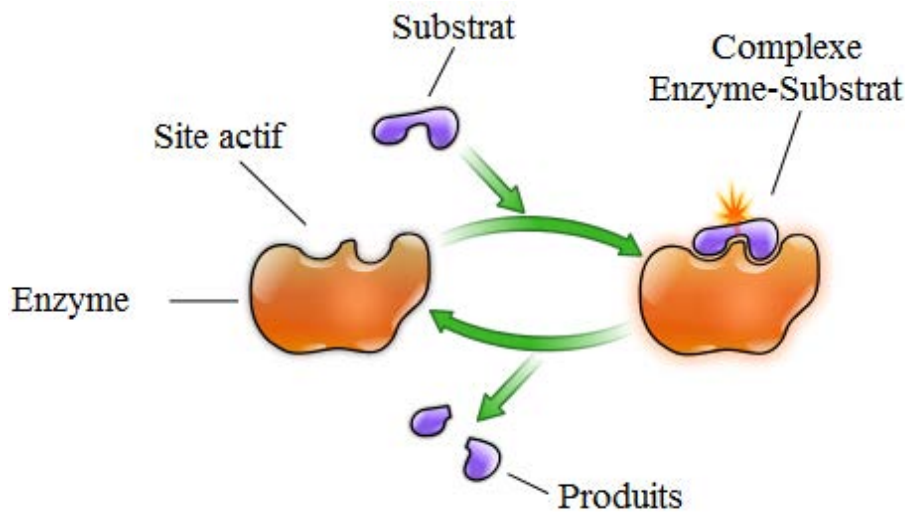


FIGURE 1.1 – Action d’une enzyme sur un substrat pour former deux produits.

Les réactions qui composent un organisme forment un réseau complexe d’enzymes et de métabolites. Un *réseau métabolique* est donc une représentation du métabolisme sous forme de graphe dont les arêtes représentent les réactions (enzymes) et les nœuds, les métabolites.

Le métabolisme intéresse beaucoup la communauté des biologistes car il définit des processus biologiques essentiels à la vie de la cellule. L’analyse des réseaux métaboliques est couramment utilisée en biotechnologies pour améliorer les performances des productions industrielles. Parmi de nombreux exemples elle a été utilisée pour maximiser la production d’acides aminés aromatiques à partir d’hydrate de carbone (glucides) en utilisant une cellule souche de *E. coli* [PPW⁺03].

Les réseaux métaboliques ont été également utilisés dans le cadre de l’essor des technologies d’observation dites ”omiques” pour comprendre le fonctionnement d’une cellule. Par exemple, pour analyser, interpréter les données d’expression issues des puces à ADN [GDS⁺05, GW07, LC04, WPM⁺06, YEW⁺05] afin de comprendre les processus mis en place par un organisme en réponse à différents environnements/changements (stress, maladies, pollution, etc.).

Matrice stœchiométrique, Métabolites d’entrées, intermédiaires et de sorties

Un réseau métabolique est caractérisé par un certain nombre de réactions entre les métabolites. Les métabolites sont en nombre fini $n_{\mathcal{M}}$ que nous désignerons par $m_1, m_2, \dots, m_{n_{\mathcal{M}}}$. Les réactions sont elles aussi en nombre fini $n_{\mathcal{R}}$. Elles relient les différents métabolites du système, nous les désignerons par $r_1, r_2, \dots, r_{n_{\mathcal{R}}}$.

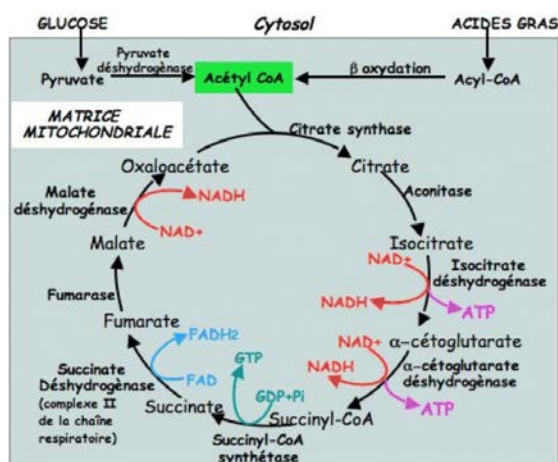
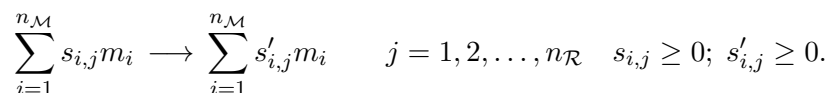


FIGURE 1.2 – Glycolyse : un exemple de voie métabolique qui transforme le glucose en Acetyl-CoA pour produire de l'énergie sous forme d'ATP.

Chacune des réactions r_j est de la forme suivante :



Les coefficients $s_{i,j}$ et $s'_{i,j}$ sont des nombres réels positifs appelés *coefficients stœchiométriques*. Ils expriment la quantité du métabolite m_i qui est consommée ou produite dans la $j^{\text{ième}}$ réaction.

Nous introduisons les notations matricielles suivantes :

$$S = [s_{i,j}] : \text{ la matrice de dimension } n_{\mathcal{M}} \times n_{\mathcal{R}},$$

$$S' = [s'_{i,j}] : \text{ la matrice de dimension } n_{\mathcal{M}} \times n_{\mathcal{R}}.$$

La *matrice stœchiométrique* est définie comme suit : $M = S' - S$, de dimension : $n_{\mathcal{M}} \times n_{\mathcal{R}}$. Les lignes de M correspondent à l'ensemble des métabolites du système et les colonnes à l'ensemble des réactions reliant les différents métabolites.

Pour présenter un réseau métabolique, le plus simple est de commencer par un exemple.

Exemple 1.1.1. Soit le réseau représenté par la Figure 1.4 qui comprend trois réactions $\{r'_1, r'_2, r'_3\}$ mettant en œuvre six métabolites $\{m_1, m_2, m_3, m_4, m_5, m_6\}$.

Dans l'exemple du réseau de la Figure 1.4, r'_1 est appelé *réaction réversible* et r'_2, r'_3 sont dites *réactions irréversibles*.

Par la suite, nous serons amenés à considérer des sous-ensembles particuliers des métabolites intervenant dans un réseau métabolique.

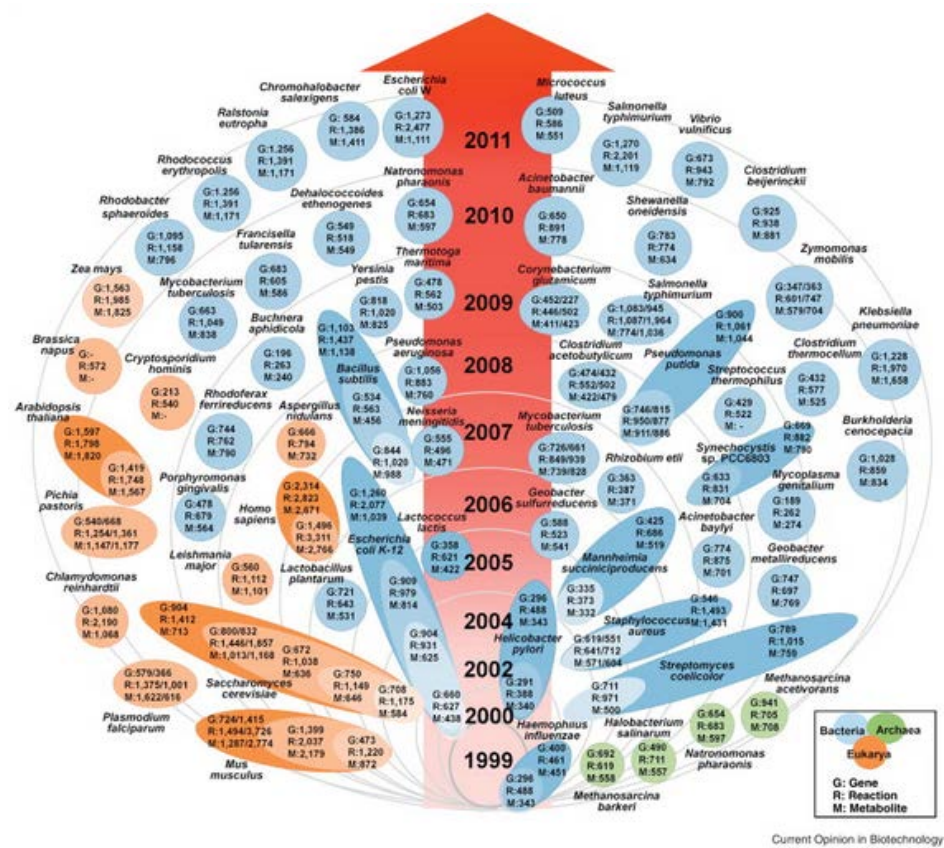


FIGURE 1.3 – Organismes dont le métabolisme a été reconstruit depuis 1999 [KSK⁺11] Les ellipsoïdes orange, bleue et verte représentent respectivement les eucaryotes, les procaryotes et les archées. Les lettres G,R et M représentent le nombre de gènes, de réactions et de métabolites identifiés dans ces modèles.

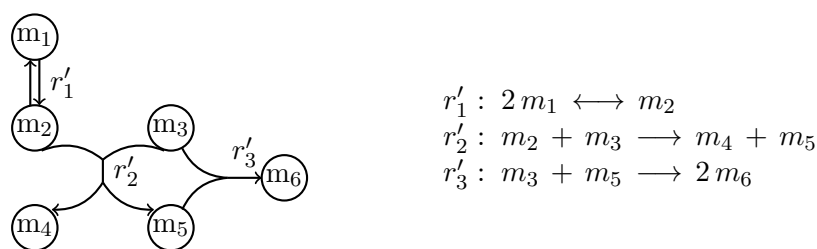


FIGURE 1.4 – Exemple de réseau métabolique.

Un métabolite d'entrée ou simplement une entrée, peut être désigné comme tout métabolite dont la production n'est pas intégralement associée par les réactions du système.

Il apparaît donc uniquement du côté gauche des réactions avec un coefficient $s_{i,j} > 0$. L'ensemble des métabolites d'entrées est en nombre fini p et est noté : $\mathcal{I} = \{I_1, I_2, \dots, I_p\}$ où $I_i = m_i$ avec $1 \leq i \leq p$.

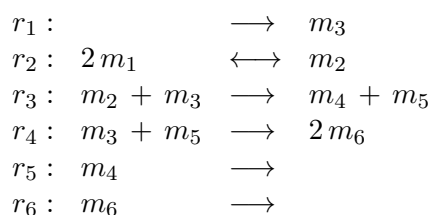
Un métabolite intermédiaire ou pivot, est un métabolite qui apparaît du côté gauche des réactions avec un coefficient $s_{i,j} > 0$ et du côté droit des réactions avec un coefficient $s'_{i,j} > 0$. L'ensemble des métabolites intermédiaires est en nombre fini n et est noté : $\mathcal{P} = \{P_1, P_2, \dots, P_n\}$ où $P_i = m_i$ avec $p + 1 \leq i \leq p + n$.

Un métabolite de sortie ou simplement une sortie, peut être désigné comme tout métabolite dont la consommation n'est pas intégralement associée par les réactions du système. Il apparaît donc uniquement du côté droit des réactions avec un coefficient $s'_{i,j} > 0$. L'ensemble des métabolites de sorties est en nombre fini q et est noté : $\mathcal{O} = \{O_1, O_2, \dots, O_q\}$ où $O_i = m_i$ avec $p + n + 1 \leq i \leq p + n + q$. Ainsi l'ensemble des métabolites est $\mathcal{M} = \mathcal{I} \cup \mathcal{P} \cup \mathcal{O}$.

Dans l'exemple du réseau de la Figure 1.4 nous avons :

$$\begin{aligned} \text{Métabolites intermédiaires :} & \quad \mathcal{P} = \{m_1, m_2, m_3, m_5\} \\ \text{Métabolites d'entrées (Entrées) :} & \quad \mathcal{I} = \{m_3\} \\ \text{Métabolites de sorties (Sorties) :} & \quad \mathcal{O} = \{m_4, m_6\} \end{aligned}$$

Pour mieux identifier les *entrées* et les *sorties* dans le réseau, l'exemple peut être réécrit comme suit :



“ $\longrightarrow m_3$ ” est appelée réaction d'entrée. “ $m_4 \longrightarrow$ ” et “ $m_6 \longrightarrow$ ” sont appelées réactions de sorties.

A chaque réaction r , on peut associer un flux v_r . Le terme de flux est équivalent à la vitesse *nette* de la réaction (en anglais “net rate”), c'est-à-dire la différence entre la vitesse de la réaction directe et celle de la réaction inverse. On note :

- $v_{\mathcal{I}}$ le vecteur correspondant aux flux des réactions d'entrées ;
- $v_{\mathcal{O}}$ le vecteur correspondant aux flux des réactions de sorties ;
- v le vecteur correspondant aux flux des réactions intermédiaires.

On appelle *vecteur de flux* (ou distribution de flux), un $n_{\mathcal{R}}$ -vecteur V de l'espace des réactions dont chaque élément décrit le flux à travers une réaction, c'est-à-dire $V = (v_{\mathcal{I}}, v, v_{\mathcal{O}})$.

Etat stationnaire d'un réseau métabolique. L'état stationnaire, d'un réseau métabolique défini par une matrice stœchiométrique, correspond à l'ensemble des distributions V de flux, solutions du système d'équations linéaires $MV = 0$. Ce système est généralement sous contraint et l'ensemble des états stationnaires a une forme d'un convexe.

Dans cette thèse, nous allons voir en quoi l'étude fine des réseaux métaboliques permet de mieux comprendre la réponse de l'ensemble d'un métabolisme à l'échelle d'un organe. L'application visée ici sera le métabolisme de la glande mammaire. En particulier, dans ce domaine, les principales études de modélisation se font à l'aide de méthodes *dynamiques* [WB84, Han94].

Or, dans le monde des biotechnologies cellulaires, l'analyse *statique* des réseaux métaboliques des micro-organismes a pris un essor considérable ces 20 dernières années, avec des modèles à base de contraintes, notamment les *modes élémentaires/voies extrêmes* [SDF99, PSP⁺04] et le *flux balance analysis* [EP00, KPE03, OTP10].

Un des objectifs de cette thèse sera de comprendre comment tirer profit de cet essor pour comprendre le fonctionnement de la glande mammaire. Avant cela, dans cette introduction, nous allons décrire les principales caractéristiques des différents types de modélisation rencontrés sur les réseaux métaboliques.

1.2 Modélisation du métabolisme

Il existe trois principaux types de modélisations du métabolisme selon le niveau d'analyse des réseaux considérés, à savoir la modélisation structurelle, la modélisation dynamique et la modélisation à base de contraintes [Ste04] (Figure 5.1).

1.2.1 Modélisation structurelle

Les modélisations structurelles nécessitent généralement peu (ou pas) de données quantitatives et sont dédiées à une analyse qualitative des réseaux. Elles s'adressent principalement à l'analyse et la vérification structurelle des réseaux reconstruits. Elles se basent sur les interactions entre métabolites et effecteurs (enzymes, transporteurs,...) pour créer un graphe d'interaction. Ces réseaux d'interaction sont très utilisés pour étudier l'organisation structurelle du métabolisme, et à ce titre, la théorie des graphes fournit beaucoup de méthodes pour analyser la connectivité des métabolites ou des effecteurs [BIB04], les chemins du graphe utilisés pour analyser leurs longueurs [MZ03] ou leurs modes élémentaires [SDF99], ou encore pour identifier des motifs particuliers tels que les *choke points* (enzymes consommant ou produisant un unique métabolite) utilisés comme cible de médicaments [YHT⁺04, RS06].

Cette modélisation n'est pas adaptée à une exploitation quantitative de données biologiques comme c'est le cas par exemple de la glande mammaire où les analyses peuvent porter sur la répartition quantitative des nutriments prélevés par la mamelle sur les constituants du lait.

1.2.2 Modélisation dynamique

Lorsqu'on souhaite étudier la réponse dynamique du réseau, on utilise généralement des équations différentielles.

Dans ce type de modélisation, on considère généralement que la structure du réseau est connue et on s'intéresse uniquement à sa dynamique. Il existe en fait différents types

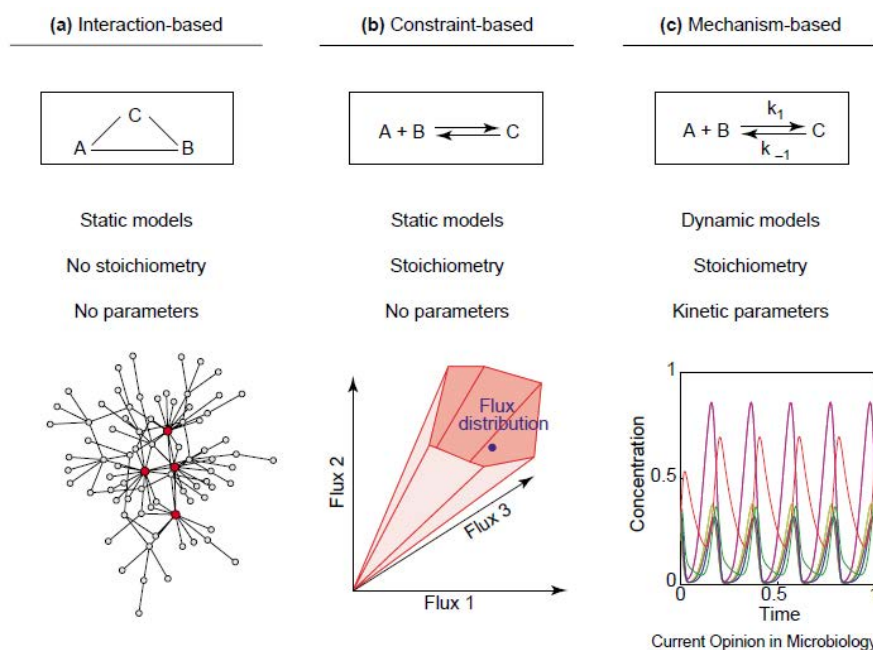


FIGURE 1.5 – Différentes approches de modélisation des réseaux métaboliques [Ste04] (a) Modélisation basée uniquement sur les interactions, (b) Modélisation à base de contraintes, (c) Modélisation cinétique tenant compte de l'évolution au cours du temps des quantités d'enzyme et de métabolites.

de modèles à base d'équations différentielles. L'idée commune est de décrire par des équations les variations de concentrations de chacun des composants du système. La variante la plus simple utilise des équations différentielles ordinaires. Différentes lois permettent de modéliser les vitesses des réactions dont les lois *d'action de masse* et de *Michaelis Menten*.

Lorsqu'une réaction r_j obéit à la loi *d'action de masse*, la vitesse, V_j de la réaction est donnée par l'expression suivante :

$$V_j = k_j \prod_{i \in I^{r_j}} m_i^{s_{i,j}},$$

où k_j est un coefficient de proportionnalité appelé *la constante de vitesse* de la réaction r_j , I^{r_j} désigne l'index de l'ensemble des réactifs mis en œuvre dans la réaction d'indice j , m_i est la concentration du i^{eme} métabolite (par abus de langage on a confondu le métabolite m_i à sa concentration) et $s_{i,j}$ correspondent aux coefficients stœchiométriques des réactifs.

La loi d'action de masse consiste donc à exprimer chaque vitesse de réaction comme étant proportionnelle au produit des concentrations des réactifs de la réaction, chaque concentration étant élevée à la puissance positive $s_{i,j}$.

Dans les réactions enzymatiques, on est amené à généraliser les vitesses des réactions de la façon suivante :

$$V_j = k_j \prod_{i \in I^{r_j}} \rho_{ij}(m_i)$$

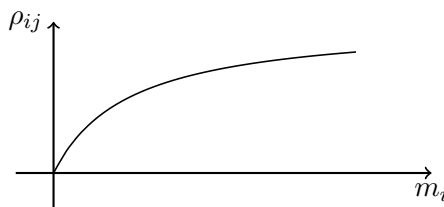


FIGURE 1.6 – Fonction cinétique monotone croissante

Souvent, les fonctions ρ_{ij} sont monotones croissantes, comme sur la Figure 1.6. Un des exemples les plus courants est connu sous le nom de cinétique de *Michaelis-Menten* représentée par la fonction suivante :

$$\rho_{ij}(m_i) = \frac{m_i}{K_{ij} + m_i}.$$

La variation de la concentration de chaque métabolite du système est égale à la somme des vitesses des réactions produisant le métabolite moins la somme des vitesses des réactions consommant le métabolite. Ainsi, la variation de concentration d'un métabolite m_i dans un système de réactions est donnée par la formule suivante :

$$\frac{dm_i(t)}{dt} = \sum_{j \in I} s_{i,j} V_j - \sum_{j \in I} s'_{i,j} V_j. \quad (1.1)$$

Où I désigne l'ensemble des indices des réactions du système.

Un frein à l'utilisation des équations différentielles pour étudier le métabolisme est la modélisation des vitesses des réactions. Aussi, le nombre de paramètres à estimer peut être très important. En général, plus le modèle est réaliste, plus il contient de paramètres.

Par exemple des fonctions complexes ont été utilisées pour modéliser les vitesses des réactions dans [Han94] comme l'indique la Figure 1.7. Mais, il a fallu ajuster plusieurs paramètres pour avoir des simulations numériques quantitativement correctes.

Pour contourner tous ces problèmes, l'étude des réseaux métaboliques est ramenée à l'état stationnaire du système.

Equation Forms for Metabolic Transactions

<i>Form number</i>	<i>General equation form^a</i>
1	$U_{A,B} = K_{A,B} * (S_1/C_1) * (S_n/C_n)$
2	$U_{A,B} = K_{A,B} * (S_1/C_1) * (S_n/C_n) * (C_f/I)$
3	$U_{A,B} = V_{A,B}/(1 + (k_1/S_1) + (k_n/S_n))$
4	$U_{A,B} = V_{A,B}/(1 + (k_1/S_1)^e + (k_n/S_n))$
5	$U_{A,B} = V_{A,B}/(1 + (k_1/S_1) + (k_n/S_n) + (I/j_i))$
6	$U_{A,B} = V_{A,B}/(1 + (k_1/S_1) + (k_n/S_n) + (I/j_i)^e)$
7	$U_{A,B} = V_{A,B} * (I/j_i)^e / (1 + (k_1/S_1) + (k_2/S_2)^e)$

FIGURE 1.7 – Modélisation des vitesses des réactions pour un réseau métabolique de la glande mammaire [Han94].

1.2.3 Modélisation à base de contraintes

Dans les modèles à base de contraintes, deux principales hypothèses sont considérées. La première est l'hypothèse d'état stationnaire (chaque métabolite interne, m_i , qui est produit doit être consommé). C'est-à-dire dans l'équation 1.1, on a : $\frac{dm_i}{dt} = 0$ (avec, $\nu_{i,j} = s_{i,j}$ et $\nu'_{i,j} = s'_{i,j}$). La deuxième concerne les contraintes dues à l'irréversibilité de certaines réactions.

Pour le $n_{\mathcal{R}}$ -vecteur V de l'espace des réactions, notons par V_j le flux à travers la réaction r_j . Alors, en considérant l'ensemble du réseau, les deux contraintes mentionnées précédemment s'expriment de la manière suivante :

- 1) $MV = 0$;
- 2) $V_j > 0, \quad \forall r_j$ irréversible.

D'autres contraintes peuvent être ajoutées comme par exemple des bornes supérieures sur les vitesses des réactions. C'est-à-dire pour chaque réaction r , on a : $v_r \leq v_r^{max}$.

Ces contraintes définissent un ensemble de distribution de flux qui a la forme d'un cône convexe. Ce cône est dit *pointé* dans le cas où toutes les réactions sont irréversibles et *non pointé* dans le cas contraire. Deux approches sont principalement utilisées pour étudier cet espace de distribution à savoir *le flux balance analysis* et *les modes élémentaires/voies extrêmes*.

1.2.3.1 Modes élémentaires et voies extrêmes

Un mode élémentaire ou une voie extrême est défini comme un vecteur de flux admissible (c'est-à-dire satisfaisant les contraintes (1) et (2), de support minimal. Formellement,

si on note $R(V) = \{j \text{ tel que } V_j \neq 0\}$ l'ensemble des réactions ayant un flux non nul dans le vecteur V , alors la condition de support minimal s'écrit :

- 3) Il n'existe pas de vecteur de flux admissible F tel que $R(F) \subset R(V)$ avec $F \neq V$.

Les concepts de modes élémentaires et de voies extrêmes ont été introduits pour traiter le cas plus général d'un réseau contenant des réactions réversibles (cône non pointé). Pour calculer les modes élémentaires ou les voies extrêmes, il faut commencer par identifier les métabolites externes qui sont ceux se trouvant aux extrémités du réseau. Pour chaque paire de métabolites externes on calcule l'ensemble des chemins, réactions pour aller du premier métabolite au second. Les deux définitions sont différentes précisément dans leur façon de traiter les réactions réversibles. Les algorithmes permettant de calculer les voies extrêmes [SLP00] procèdent de la même manière que les algorithmes permettant de calculer les modes élémentaires [SFD00], sauf que pour les voies extrêmes les réactions réversibles sont découpées en deux et considérées comme irréversibles.

La Figure 1.8, issue de [PPW⁺03] illustre la différence entre les concepts sur un exemple.

L'ensemble des voies extrêmes constitue une base convexe du cône non pointé alors que l'ensemble des modes élémentaires ne constitue pas une base convexe mais représente cependant une famille génératrice.

L'interprétation biologique d'un mode élémentaire ou d'une voie extrême est qu'il s'agit d'un ensemble minimum de réactions qui peut fonctionner à l'état stationnaire dans la cellule [SFD00, SLP00, PPW⁺03]. Les modes élémentaires ont été utilisés par exemple pour analyser l'impact des enzymes déficientes, en étudiant l'importance et la pertinence des réactions ou la flexibilité et la robustesse des voies métaboliques [GK04], pour comprendre et décrire les propriétés des réseaux de signalisation et de régulation de la transcription [GPP⁺06, KSRL⁺06].

1.2.3.2 Flux Balance Analysis (FBA)

Les analyses de type FBA reposent sur l'hypothèse que les organismes ont évolué de manière à ce que leur réseau métabolique se comporte de manière optimale à son état stationnaire [PPSP03, KPE03].

Ceci se traduit par l'optimisation d'une fonction objectif linéaire (combinaison linéaire de flux) sur le cône mentionné ci-dessus pour faire des prédictions en accord avec des données expérimentales.

La fonction objectif ainsi employée permet de formuler différents types d'hypothèses et contraintes correspondant par exemple à la croissance cellulaire, la synthèse d'un produit, la minimisation de la consommation d'ATP, la consommation de nutriment [SVC02]... Pour cela, une réaction artificielle est généralement ajoutée au réseau. Par exemple dans le cas d'un objectif de production de biomasse (constituants nécessaires à la croissance de l'organisme), cette réaction (appelée réaction de biomasse) consomme des métabolites précurseurs à une stœchiométrie qui simule la production réelle de ses constituants

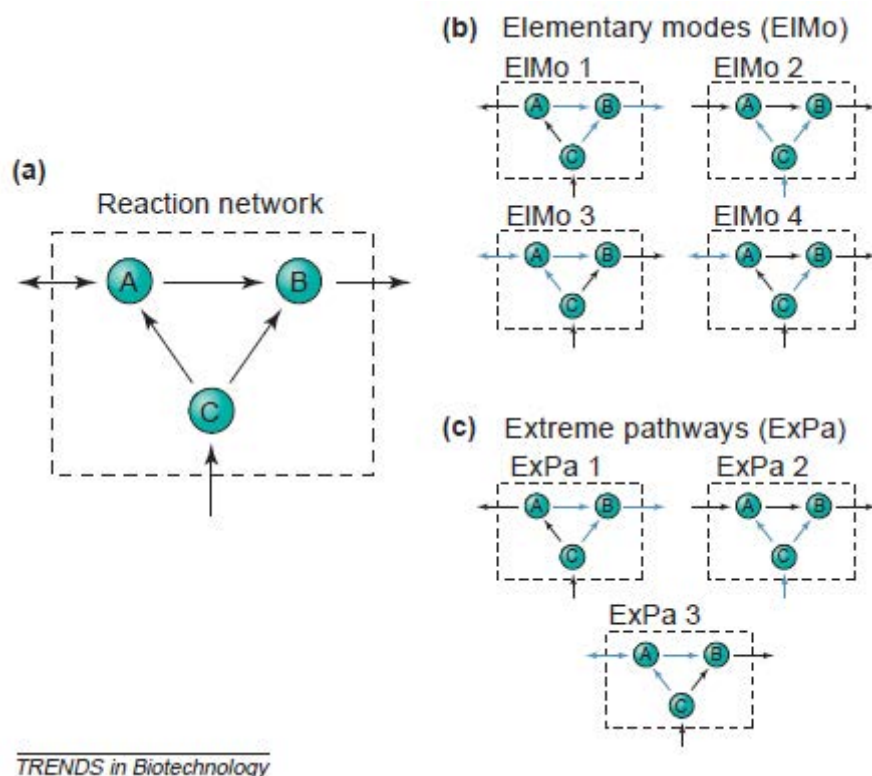


FIGURE 1.8 – Exemple de modes élémentaires et voies extrêmes sur un réseau métabolique [PPW⁺03]. *Modes élémentaires et voies extrêmes sur un exemple simple de réseau métabolique. Il est constitué de trois métabolites, trois réactions internes et trois externes (a). Le réseau admet quatre modes élémentaires (c) et trois voies extrêmes (b). La différence entre les deux ensembles est liée à l'utilisation de la réaction réversible au niveau du métabolite A. Cet exemple démontre la principale différence entre les deux approches.*

[OTP10]. Cette stratégie a ainsi permis de modéliser efficacement le comportement de grands réseaux métaboliques tels que celui d'*Escherichia coli* [EP00,CP02] et *Helicobacter pylori* [SCF⁺02], pour lesquels les prédictions de maximisation de la biomasse sont apparues consistantes avec les mesures expérimentales [KPE03].

Le *flux balance analysis* a été également appliqué avec succès pour prédire la létalité de délétions de gènes sur certains organismes en maximisant la combinaison linéaire des flux des réactions qui ne sont pas catalysées par ces gènes [EP00]. Le flux balance analysis a été utilisé pour prédire des relations épistatiques entre 890 gènes de *Saccharomyces cerevisial* [SDCK05].

Le succès du FBA a beaucoup attiré l'attention des bio-informaticiens et les bio-mathématiciens. Ainsi plusieurs autres méthodes ont été développées dans l'objectif de

diversifier et d'intégrer plus de connaissances biologiques dans les analyses effectuées par le FBA. Ces méthodes peuvent être regroupées dans deux catégories selon qu'elles modifient le principe de fonctionnement du FBA ou qu'elles intègrent le FBA dans une analyse plus large.

Dans la première catégorie (modification du principe de fonctionnement du FBA), on peut citer les approches permettant d'incorporer, lorsqu'elles sont disponibles, des données thermodynamiques [BLQ02], de cinétique [CP02, CKR⁺04] ou des valeurs de flux issues de l'expérimentation pour certaines réactions [YBL⁺10].

Dans la deuxième catégorie (intégration du FBA dans l'analyse des réseaux), des algorithmes du type *noninferior set estimation* (NISE) permettent d'optimiser successivement des fonctions objectifs sur un réseau métabolique et de faire ensuite une analyse multi-objectif de ces fonctions sur le réseau [OLLP09]. On peut également citer, *Flux Variability Analysis* (FVA) qui permet d'étudier la flexibilité dans un réseau métabolique en calculant pour chaque réaction la valeur minimale et maximale du flux sous certaines conditions sur la fonction objectif [GT10].

Plusieurs autres méthodes pour les deux catégories existent dans la littérature et différents logiciels tels que *OptFlux* [RME⁺10] ou Fasimu [HHG⁺11] tentent à présent de regrouper toutes ces méthodes d'analyse basées sur le FBA.

L'analyse basée sur une fonction objectif montre ses limites quand la fonction objectif n'est pas unique et que la *vraie* fonction objectif peut être difficile à déterminer dans différentes conditions ou après différentes modifications génétiques. Ainsi, dans [SZZ⁺12], les auteurs ont analysé 5 fonctions objectifs, à savoir la maximisation de la biomasse, de l'ATP, de l'Acétate, du CO₂ et la minimisation en valeur absolue de la somme des flux sur des bactéries pour prédire des flux expérimentaux. En combinant deux à deux et trois à trois les fonctions objectifs, ils ont montré que la combinaison de la maximisation de la biomasse, de l'ATP et la minimisation en valeur absolue de la somme des flux donne des résultats cohérents par rapport à la littérature. Récemment, un outil permettant d'intégrer diverses connaissances biologiques (en termes de qualité) dont peut disposer un réseau, a été développé en se basant sur de l'optimisation multi-objectif [GBS⁺].

Contrairement à l'optimisation mono-objectif, il n'existe pas de solution unique pour un problème multi-objectif (*PMO*), mais un ensemble de solutions, connu comme l'ensemble des solutions Pareto Optimales (*PO*).

Dans la littérature deux groupes de méthodes sont principalement utilisés pour la résolution des PMOs

- Les méthodes exactes ;
- les métaheuristiques.

Les méthodes exactes telles que : *branch and bound* [SRD88], *l'algorithme A** [SW91] et la *programmation dynamique* [Whi82] sont efficaces pour des problèmes bi-objectifs et de petite taille. Pour les problèmes à plus de deux critères ou de grande taille, il n'existe pas de méthodes exactes efficaces de résolution. Ainsi, on assiste depuis les années 1980 à un accroissement d'intérêt sur l'utilisation de métaheuristiques pour la résolution de PMO.

Elles comprennent notamment la méthode du recuit simulé, les algorithmes génétiques, la méthode de recherche tabou, les algorithmes de colonies de fourmis, etc.

En effet, ces heuristiques ont la particularité de s'adapter à toutes sortes de PMO. Elles ne garantissent pas de trouver de manière exacte l'ensemble PO , mais plutôt une approximation de cet ensemble. Ces méthodes ont une ambition commune : résoudre au mieux les problèmes d'optimisation difficiles.

D'une façon générale, un problème d'optimisation multi-objectifs (PMO) peut être défini de la manière suivante :

$$(PMO) \begin{cases} \min F(x) = (f_1(x), f_2(x), \dots, f_n(x)) \\ \text{s.c. } x \in E \end{cases}$$

où $n \geq 2$ est le nombre de fonctions objectifs, $x = (x_1, x_2, \dots, x_n)$ est le vecteur représentant les variables de décision, E représente l'ensemble des solutions réalisables associé à des contraintes d'égalités et d'inégalités et $F(x) = (f_1(x), f_2(x), \dots, f_n(x))$ est le vecteur des critères à optimiser.

Un point $x \in E$ est dit Pareto-optimal s'il n'est dominé par aucun autre point de E . La reformulation mathématique donne :

$$x \in E \text{ domine } x' \in E \quad \text{si } \forall i, f_i(x) \leq f_i(x') \quad \text{avec } \exists i \text{ tel que } f_i(x) < f_i(x')$$

Ces points sont également appelés solutions non inférieures ou non dominées. Dans l'exemple ci-dessous, les points a , c et e ne sont dominés par aucun autre. Alors que le point d est dominé par c , et que le point b est dominé par les points a et c .

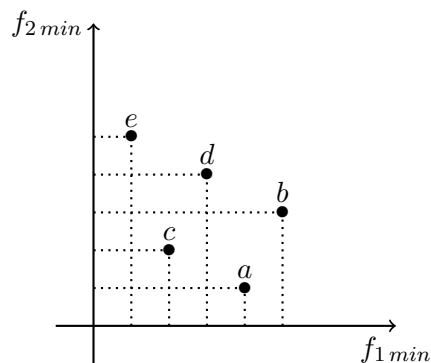


FIGURE 1.9 – Exemple de dominance.

L'ensemble des points non dominés constitue la frontière de Pareto et tous les points Pareto-optimaux. Dans la Figure 1.9, la frontière de Pareto est composée des points a , c et e .

Les méthodes de résolution par les métaheuristiques consistent à générer uniformément quelques milliers de points dans l'espace de recherche pour les faire converger vers la frontière de Pareto. Par exemple, dans le cas d'un espace de recherche qui a la forme d'un polyèdre borné, on peut déterminer le centre de Chebyshev du polyèdre, centre de la plus grande boule contenue dans le polyèdre, puis générer uniformément des directions à partir de ce centre pour prendre aléatoirement des points dans l'espace [KS98].

Les points ainsi générés permettent de déterminer la frontière de Pareto. Par exemple, les algorithmes génétiques consistent à prendre aléatoirement des points un à un ou deux à deux et les transformer en d'autres points. Les transformations consistent à appliquer des *opérateurs* appelés *opérateurs de mutation* ou de *croisement* aux points aléatoirement choisis [Gol89].

Description par chapitre

Chapitre 2. Dans ce chapitre, il s'agit de comprendre si le métabolisme de la glande mammaire peut être modélisé par une optimisation d'une combinaison linéaire des flux des réactions. Nous appliquons ainsi le *Flux Balance Analysis* à l'étude du métabolisme de la glande mammaire sur des jeux de données réels.

Nous présentons les résultats d'analyse des différentes distributions extrémales (sommets extrémaux) susceptibles d'optimiser une combinaison linéaire des flux des réactions en les confrontant aux données de la littérature.

Chapitre 3. Nous avons formalisé, dans ce chapitre, une technique inspirée par le domaine des plantes et de la nutrition. Il s'agit de quantifier les précurseurs qui interviennent dans la composition des sorties du système. Cette composition est appelée *contribution des entrées (inputs) dans les sorties (outputs)* [AIO]. Nous avons ensuite utilisé cette approche d'analyse par la méthode des AIOs pour discuter des distributions extrémales qui soulèvent un doute de pertinence biologique dans le chapitre 2.

Chapitre 4. Nous présentons une extension du *Flux Variability Analysis* (FVA) qui consiste à optimiser des flux de réactions pour analyser la flexibilité dans un réseau métabolique. Cette approche que nous avons proposé est une méthode consistant à calculer les valeurs minimales et maximales des AIOs plutôt que des flux. Pour cela, nous avons développé des approches originales pour optimiser des fonctions non linéaires des flux. Nous nous sommes en particulier inspirés des heuristiques existant en multi-objectif pour pouvoir discrétiser l'espace des solutions. Cette approche permet de discriminer des régimes alimentaires sans formuler d'hypothèses sur le comportement interne du système. Nous présentons des résultats d'application sur des régimes alimentaires de la glande mammaire.

Chapitre 5. Dans ce chapitre, nous présentons un logiciel bio-informatique, *NutritionAnalyzer*, que nous avons développé pour l'analyse de flexibilité dans les réseaux métaboliques en se basant sur le calcul des AIOs.

Chapitre 6. En conclusion, nous faisons un bilan des méthodes développées pour l'analyse des réseaux métaboliques et les résultats de leurs applications au cas de la glande mammaire. Nous présentons des améliorations à apporter au niveau du modèle de la glande mammaire et des difficultés qu'on peut rencontrer. Nous avons également proposé des solutions pour améliorer les méthodes développées. Enfin, nous avons présenté des directions de recherche qui peuvent faire suite à cette thèse.

Annexes. En annexes, nous présentons : le modèle stœchiométrique utilisé pour l'analyse du métabolisme de la glande mammaire, un algorithme du simplexe qui a permis de déterminer les différentes distributions extrémales et les publications scientifiques issues de cette thèse.

Chapitre 2

Analyse de la distribution de flux dans un réseau métabolique en nutrition

Dans ce chapitre, nous appliquons le *flux balance analysis* à l'analyse des données sur un modèle du métabolisme de la glande mammaire. L'objectif est de chercher à savoir si la transformation des nutriments par la glande mammaire pour produire du lait peut être modélisée par l'optimisation d'une combinaison linéaire des flux des réactions du réseau.

2.1 Les réseaux métaboliques entrées/sorties

L'analyse des réseaux métaboliques entrées/sorties est très utilisée dans le domaine de la nutrition. En particulier dans la production laitière où les expérimentations consistent à étudier l'impact des différents régimes alimentaires sur les constituants du lait.

Comme indiqué en introduction, un réseau métabolique est caractérisé par un ensemble de métabolites et de réactions reliant ces métabolites. L'ensemble des métabolites est séparé en trois sous-ensembles :

- les métabolites d'entrées $\mathcal{I} = \{I_1, I_2, \dots, I_p\}$ où $I_i = m_i$ avec $1 \leq i \leq p$;
- les métabolites intermédiaires (pivots) $\mathcal{P} = \{P_1, P_2, \dots, P_n\}$ où $P_i = m_i$ avec $p + 1 \leq i \leq p + n$.
- les métabolites de sorties $\mathcal{O} = \{O_1, O_2, \dots, O_q\}$ où $O_i = m_i$ avec $p + n + 1 \leq i \leq p + n + q$;

A chaque métabolite d'entrée, on associe une réaction de la forme $\longrightarrow I_i$ (réaction d'entrée) et à chaque métabolite de sortie, on associe une réaction de la forme $O_i \longrightarrow$ (réaction de sortie). Les flux des réactions d'entrées et de sorties peuvent être fournis par des mesures expérimentales assez difficiles à obtenir puisque réalisées sur des mammifères qui doivent être en lactation.

A l'état quasi-stationnaire les flux des réactions doivent satisfaire l'équation :

$$M \begin{pmatrix} v_{\mathcal{I}} \\ v \\ v_{\mathcal{O}} \end{pmatrix} = 0. \quad (2.1)$$

Ici, M désigne la matrice stœchiométrique du réseau, $v_{\mathcal{I}}$ le vecteur correspondant aux flux des réactions d'entrées, $v_{\mathcal{O}}$ le vecteur correspondant aux flux des réactions de sorties, v le vecteur correspondant aux flux des réactions intermédiaires.

Une des particularités en nutrition dans les modèles dynamiques est de réduire l'espace de recherche en rajoutant des contraintes additionnelles globales sur les répartitions des sorties. On les modélise par un système d'équations supplémentaires

$$Bv = 0. \quad (2.2)$$

Ici, B est la matrice représentant les coefficients des contraintes additionnelles.

2.2 Un exemple de réseau entrées/sorties : la glande mammaire

L'exemple d'application de réseau métabolique entrées/sorties à l'échelle de l'organe que nous allons considérer est celui de la glande mammaire. Ce modèle a été étudié et modélisé avec différentes approches, numériques [WB84, Han94] ou d'inspiration statique [VM02].

Dans le cadre d'une collaboration avec l'INRA Saint Gilles, nous avons amélioré ce modèle statique et pris en compte les spécificités des ruminants tirées des modèles numériques.

Dans le cas du métabolisme de la mamelle des vaches laitières, les expérimentations consistent souvent à proposer différents régimes alimentaires aux vaches et étudier l'impact sur la synthèse des constituants du lait (lactose, protéines, acides gras etc.). Ainsi donc, pour chaque régime alimentaire nous avons des données expérimentales entrées/sorties mesurées en mmol/h.

Les entrées. Ce sont les flux des métabolites ou gaz prélevés par la glande mammaire. Il s'agit de :

- les flux en acides aminés (AA) à utiliser dans des voies métaboliques autres que la synthèse protéique ;
- les AA indispensables sont histidine, isoleucine, leucine, lysine, méthionine (methionine), phénylalanine (phenylalanine), thréonine (threonine), tryptophane, valine
- les AA non indispensables sont arginine, alanine, asparagine, aspartate, cystéine (cysteine), glutamate, glutamine, glycine, proline, sérine (serine), tyrosine. L'ornithine, la citrulline et la taurine ne sont pas encore renseignées dans le modèle.

- l'acétate (acetate) ;
- le β -hydroxybutyrate (BHBA) ;
- le glucose (glucose ou GLC) ;
- le glycérol (glycerol) ;
- le lactate (lactate) ;
- l'acide propionique ou propionate.

Les sorties. Ce sont les flux des métabolites sécrétés par la glande mammaire dans le lait ou rejetés au niveau de la veine mammaire. Il s'agit de :

- le nombre de liens peptidiques des protéines du lait (peptide output) ;
- les flux en acides aminés non indispensables à synthétiser dans la glande mammaire ;
- les acides gras du lait synthétisés dans la glande mammaire comptant 4 à 16 atomes de carbone ;
- la quantité totale de triglycérides et de glycérol-3P contenu dans ces triglycérides ;
- le lactose.

Dans le modèle, la synthèse protéique et la dégradation des protéines ne sont prises en compte qu'au travers du nombre de lien peptide (lien entre deux acides aminés). La différence entre les flux de synthèse protéique (« peptide synthesis ») et dégradation protéique (peptide hydrolysis) est supposée représenter le flux de protéines exportées dans le lait (« peptide output ») qui est renseigné dans les sorties.

Certains métabolites sont considérés comme entrée ou sortie en fonction du régime alimentaire. C'est le cas, par exemple des acides aminés.

Un acide aminé est supposé être une entrée si la quantité prélevée (par unité de temps) est supérieure à sa quantité exportée dans les protéines du lait. Dans ce cas le flux renseigné correspond à cette différence (exemple : l'alanine dans le traitement (Ctrl) dont le flux est représenté par v_{128} dans la Table 2.1). Cet acide aminé est utilisé dans différentes voies métaboliques que la synthèse protéique (peptide synthesis) pour produire d'autres métabolites constituants du lait (output : acides aminés non indispensables, lactose etc.) ou peut contribuer à la production de CO_2 . Par abus de langage l'équation d'utilisation de cette entrée en acide aminé est notée par Van Milgen(2002) dans notre modèle par « Amino Acid Catabolism ».

Un acide aminé non indispensable est supposé être une sortie si la quantité prélevée par la glande mammaire est inférieure à la quantité produite dans les protéines du lait. Dans ce cas, il doit être synthétisé dans la glande. La donnée renseignée dans les sorties est la valeur absolue de cette différence pour l'acide aminé considéré (exemple : l'alanine dans le traitement (CN) dont le flux est représenté par v_{121} dans la Table 2.1).

Nous présentons et utilisons dans cette thèse des **données entrées/sorties** pour trois régimes alimentaires :

- Un régime alimentaire appelé contrôle (Ctrl) [LRL⁺09, RLL⁺06] correspondant à une alimentation proche des recommandations de l'INRA pour les vaches laitières ;

- Un régime alimentaire appelé caséine (CN) [LRL⁺09, RLL⁺06] correspondant une alimentation où la quantité de protéines digérées dans l'intestin était augmentée de 30% ;
- Un régime alimentaire appelé (HB) [Han94]. Il s'agit d'un jeu de données qui a été constitué à partir de résultats de différentes expériences.

Les réactions biochimiques transformant les différents nutriments prélevés pour produire le lait sont écrites en utilisant un nombre restreint de 11 **métabolites intermédiaires** appelés **pivots biochimiques**.

Les métabolites intermédiaires. Ces métabolites sont :

- le glucose (GLC) ;
- le glucose-6-phosphate (G6P) ;
- le glycéraldéhyde-3-phosphate (G3P) ;
- le glycérol 3-phosphate (glycerol-3P) ;
- le pyruvate (PYR) ;
- l'acétyl-coenzyme A cytosolique (ACoAc) ;
- le mitochondrien (ACoAm) ;
- le α -cétoglutarate (aKG) ;
- l'oxaloacétate (OAA) ;
- l'amorce des acides gras à 2 ou 4 carbones (FA primer) ;
- la sérine (SER).

Ces pivots correspondent à des carrefours importants entre différentes voies métaboliques.

Le **réseau métabolique** utilisé pour les trois jeux de données est donné à la Figure 2.1. Il est constitué de 83 réactions et 49 métabolites (entrées, sorties, métabolites intermédiaires et cofacteurs) correspondant aux réactions des principales voies biochimiques à savoir : la synthèse des matières grasses, la synthèse des matières protéiques, la synthèse du lactose et les voies d'oxydation du NADH, NADPH et FADH₂. L'ensemble des réactions du réseau faisant intervenir les métabolites intermédiaires et les cofacteurs est donné en annexe B.

La Figure 2.2 donne un réseau simplifié de la glande mammaire présentant uniquement les métabolites d'entrées, de sorties et les intermédiaires.

Les contraintes additionnelles. Les contraintes additionnelles ajoutées au modèle sont des contraintes issues des modèles dynamiques de la littérature [Han94] et qui expriment la relation existant entre certaines réactions biochimiques. Il s'agit :

- du taux de synthèse d'acides gras acylés ($C(n : m) - acylCoA$). Il est fixé à trois fois au taux de synthèse des triglycérides en sortie, en supposant que 100% des lipides sont des triglycérides ;
- du taux des différents acides gras. Ils sont supposés être synthétisés à 50% à partir de l'acétate (c'est-à-dire des $primerC(2 : 0)CoA$) et à 50% à partir du BHBA

flux entrée/sortie	(Ctrl) mmol/h/demi- mamelle	(CN) mmol/h/demi- mamelle	(HB) mol/j/mamelle	mmol/h/demi- mamelle	
v_2 entrée glucose	237	232	12.21	254	
v_{95} entrée glycerol	5.84	5.74	0.033	0.69	
v_{96} entrée acetate	510	462	18.42	384	
v_{97} entrée BHBA	84	167	7.25	151	
v_{98} entrée lactate	0	0	0.023	0.48	
v_{62} 3C(n :m)- acycoA+glycerol- 3P→triglyceride(sortie)	32.96	39.11	1.52	31.67	
acides gras synthétisés					
v_{100} sortie C(4 :0)	10.08	11.59	0.46	9.48	
v_{101} sortie C(6 :0)	4.51	5.58	0.18	3.79	
v_{102} sortie C(8 :0)	2.23	2.87	0.10	2.06	
v_{103} sortie C(10 :0)	4.66	6.46	0.19	3.96	
v_{104} sortie C(12 :0)	4.23	6.12	0.17	3.56	
v_{105} sortie C(14 :0)	13.90	17.89	0.45	9.31	
v_{106} sortie C(16 :0)	18.82	21.44	0.64	13.40	
v_{99} sortie Lactose	73.80	83.52	3.81	79.28	
acides aminés (entrée ou sortie)					
v_{128} entrée Alanine	3.11	0	0.105	2.19	alanine ca- tabolism
v_{121} sortie Alanine	0	3.26	0	0	alanine syn- thesis
v_{119} entrée Arginine	4.40	4.48	0.526	10.96	arginine ca- tabolism
v_{134} sortie Asparagine	0	0	0.023	0.48	asparagine synthesis
v_{125} sortie Aspartate	3.43	4.13	0.247	5.15	aspartate synthesis
v_{122} sortie Glutamate	0.54	6.33	0.230	4.79	glutamate synthesis
v_{131} entrée Glutamine	1.22	1.79	0.072	1.50	glutamine catabolism
v_{120} sortie Glycine	4.98	3.44	0.248	5.17	glycine syn- thesis
v_{124} sortie Proline	10.65	10.99	0.670	13.96	proline syn- thesis
v_{136} sortie Serine ⁽⁷⁾	7.21	7.50	0.090	1.88	Serine syn- thesis - Se- rine used in other path- ways
v_{118} entrée Histidine	0.23	0	0	0	histidine catabolism
v_{113} entrée Isoleucine	2.19	3.57	1.518	31.63	isoleucine catabolism
v_{114} entrée Leucine	2.02	3.76	0	0	leucine ca- tabolism
v_{108} entrée Lysine	2.68	3.58	0.191	3.98	lysine cata- bolism
v_{111} entrée Threonine	0.35	0	0	0	threonine catabolism
v_{115} entrée Valine	2.54	3.86	0.438	9.13	valine cata- bolism
v_{107} sortie Peptide	124.5	150.0	7.2	149.17	

TABLE 2.1 – Trois jeux de données entrées/sorties.

Données renormalisées en mmol/h/demi-mamelle pour les régimes alimentaires (*Ctrl*) [RLL⁺06, LRL⁺09] (*CN*) [RLL⁺06, LRL⁺09] et (*HB*) [Han94]. Ces données déterminent les valeurs des vecteurs $v_{\mathcal{I}}$ et $v_{\mathcal{O}}$

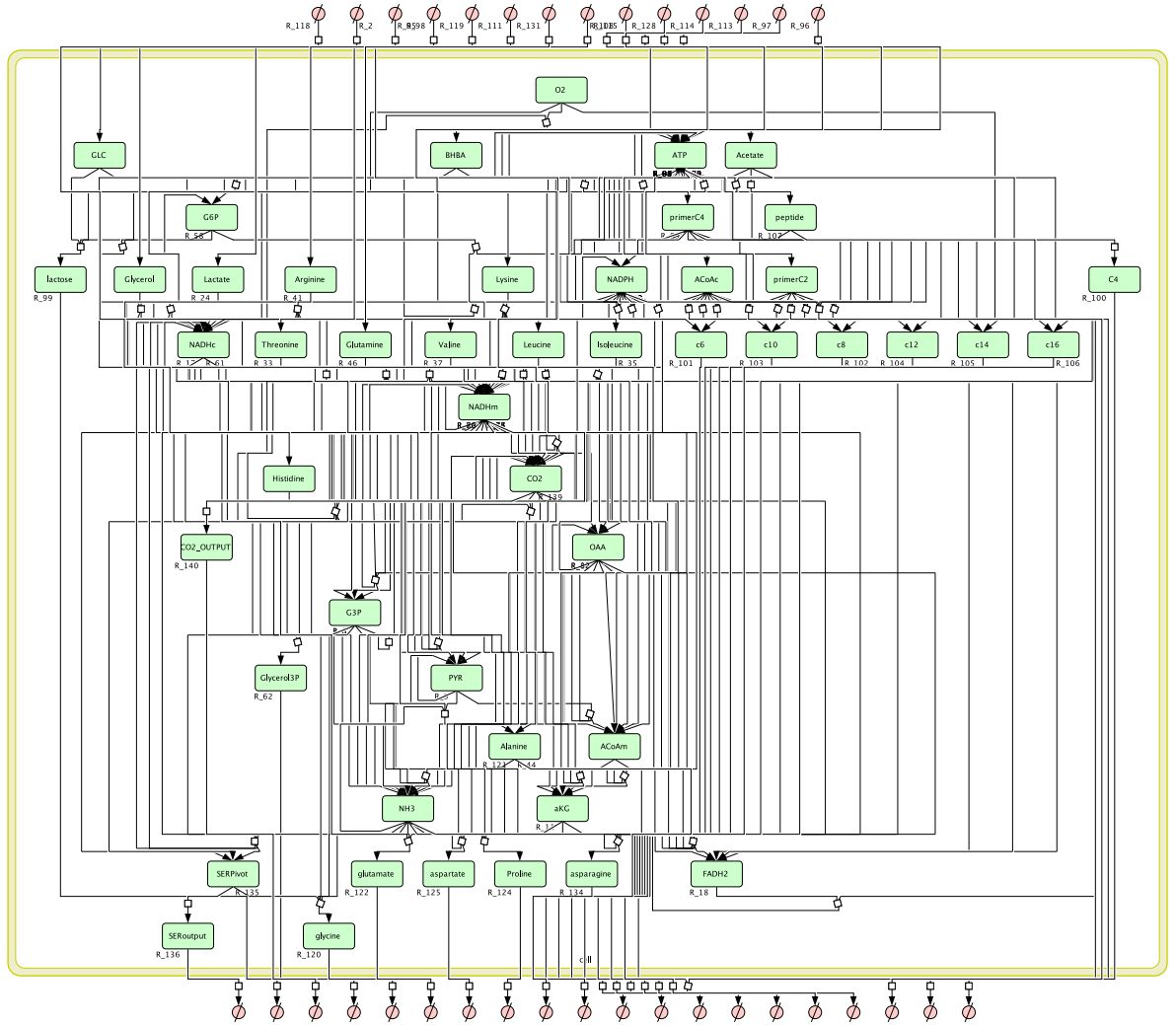


FIGURE 2.1 – Réseau métabolique complet de la glande mammaire correspondant aux trois jeux de données (Ctrl), (CN) et (HB).

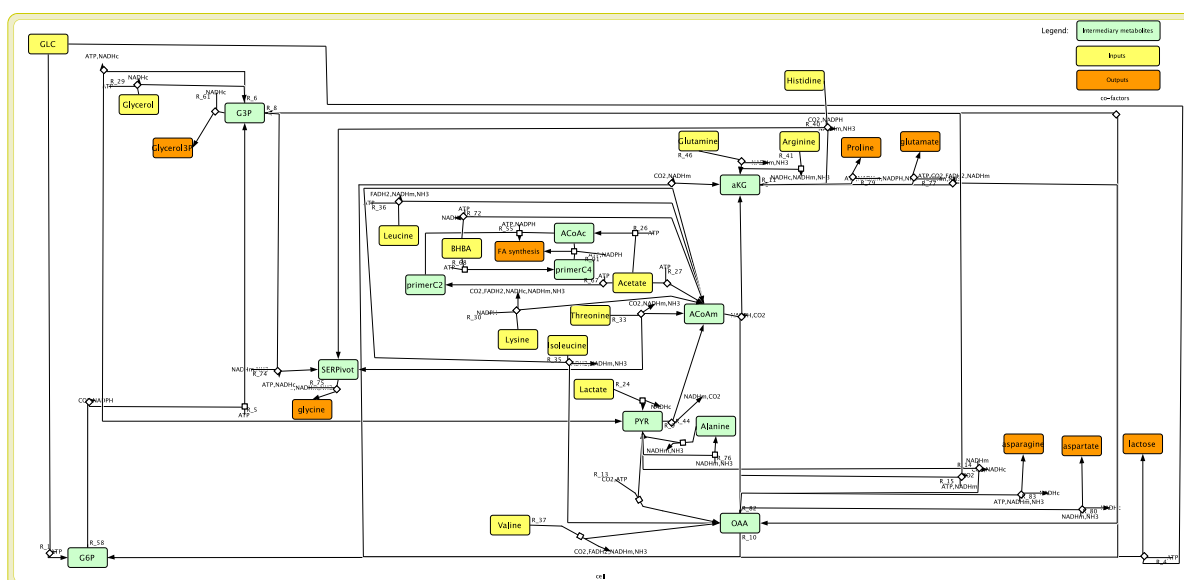


FIGURE 2.2 – Réseau métabolique simplifié de la glande mammaire correspondant aux trois jeux de données (Ctrl), (CN) et (HB).

(c'est-à-dire des $primerC(4:0)CoA$) sauf l'acide gras $C(4:0)$ qui est supposé être synthétisé à 100% du BHBA ;

- du taux de NADPH. Il est supposé que 30% du NADPH est synthétisé par la voie des isocitrates déshydrogénases et de 70% par la voie des pentoses phosphates.

Ces contraintes additionnelles représentant le système $Bv = 0$ de la section 2.1 sont données par les équations : $v_{56} - 3v_{62} = 0$; $v_{51} - v_{52} = 0$; $v_{54} - v_{90} = 0$; $v_{55} - v_{91} = 0$; $v_{86} - v_{92} = 0$; $v_{87} - v_{93} = 0$; $v_{88} - v_{94} = 0$; $0.7v_{82} - 0.3v_{58} = 0$.

Les jeux de données (Ctrl), (CN) et (HB) présentent l'avantage que les données renseignées dans les modèles (appelées entrées v_I et sorties v_O) sont cohérentes, c'est-à-dire qu'il n'y a aucun flux d'entrées trop petits (nutriments prélevés par la glande) ne permettant pas de produire les flux de sorties (nutriments produits par la glande dans le lait et CO_2) ou inversement qu'il n'y a pas de flux de nutriments appelés des sorties qui soit trop grand relativement aux flux de nutriments considérés comme des entrées dans notre système (la glande mammaire de vache). De plus, l'ensemble de ces flux renseignés en entrée et en sortie permettait aussi d'équilibrer l'ensemble des réactions de métabolisme intermédiaire. C'est-à-dire le système 2.1 et 2.2 admet au moins une solution pour $v \geq 0$.

Nous présentons dans la section qui suit une étude du réseau métabolique entrées/sorties de la glande mammaire basée sur le calcul de bilan par des méthodes issues de la littérature.

2.3 Etude du métabolisme de la glande mammaire par les méthodes issues de la littérature

Nous présentons trois approches basées sur des bilans en ATP d'utilisation des nutriments prélevés par la mamelle pour synthétiser les constituants du lait. La première approche est une estimation du bilan d'ATP basée sur des *bilans mammaires* [LRL⁺09] à partir d'une méthode appelée *calcullette métabolique* [VM02]. La deuxième concerne les modèles différentiels et enfin la troisième est une maximisation du bilan d'ATP basée sur le FBA.

Comme nous le verrons, il est apparu que la démarche implicite de la calcullette métabolique est la maximisation du bilan d'ATP sur un polyèdre convexe.

2.3.1 La calcullette métabolique

Pour analyser des données entrées/sorties telles que celles présentées dans la Table 2.1 afin de comparer différents régimes alimentaires, des chercheurs de l'Inra Saint Gilles utilisent une méthode désignée par le terme de *calcullette métabolique* proposée par un autre chercheur rennais [VM02] qui consiste à exploiter la stœchiométrie du système et les contraintes additionnelles tirées de la littérature pour déterminer "à la main" les réactions métaboliques empruntées par les nutriments apportés par le régime alimentaire.

La calcullette métabolique a été à la base élaborée pour l'étude du métabolisme énergétique des monogastriques, mais elle peut être utilisée afin d'étudier le métabolisme de tout type de cellules, de tissus ou d'organes de mammifères. Il suffit d'ajouter les équations des voies métaboliques que l'on souhaite étudier.

Cependant, elle a été adaptée au modèle mammaire décrit dans la section 2.2 en rajoutant des réactions du métabolisme mammaire chez la vache laitière publiées dans d'autres modèles [Han94, LAAS⁺10, VMGR03, WB84] afin de calculer des bilans d'utilisation en ATP, CO₂, O₂, et NH₃ des nutriments prélevés par la mamelle pour produire les différents constituants du lait.

Concrètement, la calcullette est une feuille *excel* comme le montre la Figure 2.3 ci-dessous. Les colonnes contiennent les cofacteurs (colonnes E à K) et les métabolites intermédiaires à équilibrer (colonnes M à T). Les lignes correspondent aux réactions à utiliser, elles intègrent également la stœchiométrie des réactions. Par exemple, supposons qu'un utilisateur connaît expérimentalement l'entrée du glucose dans la glande mammaire et la sortie du lactose. Il choisit ensuite les équations appropriées (colonne B) pour synthétiser le lactose et oxyder le glucose. Le calcul de bilan revient à rechercher un facteur de multiplication acceptable pour chaque réaction, de manière à équilibrer les bilans des pivots, et à rendre le bilan d'ATP positif. Cette recherche de coefficient se fait manuellement. Les bilans sont alors estimés à partir des coefficients calculés à partir de la calcullette.

Il faut bien noter qu'il existe plusieurs manières d'équilibrer un tel bilan. En pratique, l'utilisateur va rechercher à minimiser les cycles futiles du système.

A	B	C	D	E	F	G	H	J	K	L	M	N	O	P	Q	R	S	T
1	eq.	Flux value.	reaction	ATP	cNADH	mNADH	FADH2	CO2	O2	NH3	OAA	α KG	PYR	mAcCoA	cAcCoA	GLC	G6P	G3P
100	4	$F_1=1,0$	lactose synthesis	-1	0	0	0	0	0	0	0	0	0	0	0	-1	-1	0
101	2	$F_2=3,0$	glucose entry	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0
102	1	$F_3=3,0$	GLC \rightarrow G6P	-2	0	0	0	0	0	0	0	0	0	0	0	-2	2	0
103	5	$F_4=1,0$	G6P \rightarrow G3P	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	-1
104	6	$F_5=2,0$	G3P \rightarrow PYR	4	2	0	0	0	0	0	0	0	2	0	0	0	0	-2
105	8	$F_6=0,0$	G3P \rightarrow G6P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
106	9	$F_7=2,0$	PYR \rightarrow mAcCoA	0	0	2	0	2	0	0	0	0	-2	2	0	0	0	0
107	10	$F_8=2,0$	OAA + mAcCoA \rightarrow α KG	0	0	2	0	2	0	0	0	-2	2	0	-2	0	0	0
108	11	$F_9=2,0$	α KG \rightarrow OAA	2	0	4	2	2	0	0	2	-2	0	0	0	0	0	0
109	13	$F_{10}=0,0$	PYR \rightarrow OAA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
110	15	$F_{11}=0,0$	OAA \rightarrow G3P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
111																		
112	17	$F_{12}=2,0$	cNADH \leftrightarrow mNADH	0	-2	2	0	0	0	0	0	0	0	0	0	0	0	0
113	16	$F_{13}=10,0$	mNADH oxidation	30	0	-10	0	0	-5	0	0	0	0	0	0	0	0	0
114	18	$F_{14}=2,0$	FADH2 oxidation	4	0	0	-2	0	-1	0	0	0	0	0	0	0	0	0
			balance	36	0	0	0	6	-6	0	0	0	0	0	0	0	0	0
				B_{ATP}	B_{cNADH}	B_{mNADH}	B_{FADH2}	B_{CO2}	B_{O2}	B_{NH3}	B_{OAA}	$B_{\alpha KG}$	B_{PYR}	B_{mAcCoA}	B_{cAcCoA}	B_{GLC}	B_{G6P}	B_{G3P}

FIGURE 2.3 – Exemple d'utilisation de la calculette métabolique.

La calculette métabolique Excel : trois moles de glucose ont été utilisées pour produire une mole de lactose. Les taux (coefficients) d'utilisation des différentes réactions sont déterminés manuellement pour équilibrer les métabolites intermédiaires et déterminer les bilans d'utilisation en ATP, CO₂, O₂ et NH₃.

Les bilans d'ATP calculés par la calculette métabolique en rejetant les cycles futiles sont respectivement de 3081, 2045 et 6628 mmol/h/demi-mamelle pour les jeux (Ctrl), (CN) et (HB) comme présentés dans la Table 2.2.

2.3.2 Modélisation dynamique

Dans le modèle numérique qui tient lieu de référence par le métabolisme mammaire [Han94], l'ATP généré dans la glande mammaire chez la vache laitière est estimé à 6500 mmol/h/demi-mamelle (c'est-à-dire 312 mol/j/mamelle) alors que 2125 mmol/h/demi-mamelle (c'est-à-dire 102 mol/j/mamelle) sont estimés être utilisés pour la synthèse du lait [HFM⁺09]. Donc le bilan d'ATP après la synthèse des constituants du lait est de 6500 - 2125 = 4375 mmol/h/demi-mamelle.

2.3.3 Maximisation d'ATP

Une des caractéristiques de l'application du FBA en nutrition est que les données entrées/sorties contraignent fortement la variabilité de l'espace de distribution des flux. Ainsi, l'espace correspondant aux équations 2.1 et 2.2 qui a plusieurs dizaines de flux extrémaux sans les données entrées/sorties admet seulement cinq flux libres et indépendants avec les données entrées/sorties. Ces flux libres correspondent aux flux des réactions : G3P \rightarrow G6P (R_8), OAA \rightarrow PYR (R_{14}), OAA \rightarrow G3P (R_{15}), NADPH oxidation (R_{19}), peptide hydrolysis (R_{64}) de flux respectifs $v_8, v_{14}, v_{15}, v_{19}, v_{64}$. En rajoutant une contrainte d'irréversibilité sur chaque réaction r du réseau, c'est-à-dire $v_r \geq 0$, l'espace des solutions a la forme d'un cône infini pointé comme l'illustre la Figure 2.4 ci-dessous.

Le bilan d'utilisation de l'ATP, s'exprimant comme une combinaison linéaire des flux des différentes réactions du réseau, peut également s'écrire comme une combinaison

linéaire des variables v_8 , v_{14} , v_{15} , v_{19} , v_{64} . Ainsi, la maximisation de l'ATP se ramène à un problème classique dans le FBA [KPE03].

Différents algorithmes permettent de résoudre ce problème d'optimisation linéaire, dont l'algorithme du simplexe [Dan51], présenté en annexe A. Il permet de trouver la solution optimale exacte ou de montrer que la fonction objectif n'est pas bornée.

Il a été montré que si la solution optimale existe et est unique alors elle se trouve sur un sommet de l'espace des solutions [Ker89]. Par la suite, il est apparu que l'unique sommet de notre cône est celui qui maximise le bilan d'ATP pour les trois jeux de données (Ctrl), (CN) et (HB).

Les valeurs maximales d'ATP générées pour les trois jeux de données sont : 3081 pour (Ctrl), 2045 pour (CN) et 6628 pour (HB) comme présentées dans la Table 2.2. Elles correspondent exactement aux bilans d'ATP calculés en utilisant la calculette métabolique. Ainsi, la calculette métabolique peut être vue comme un cas particulier d'application du FBA avec un objectif de maximisation de l'ATP, ou plus généralement de minimisation des cycles futiles. Ce résultat, nouveau dans le domaine de la nutrition, a été présenté dans les actes de la conférence *modnut* 2011 et dans diverses réunions organisées par l'Inra [LAAS⁺10].

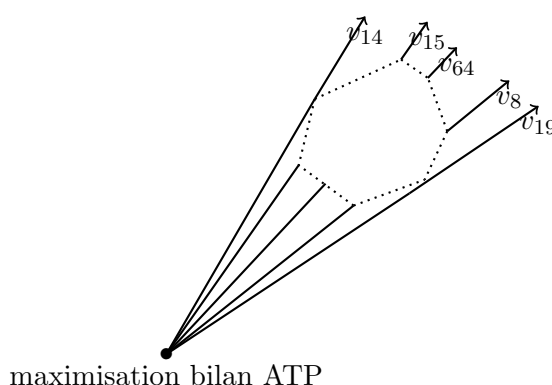


FIGURE 2.4 – Espace de distribution de flux à ATP non fixé.

Cependant, on peut constater que les modèles numériques et statiques prédisent des bilans ATP très variables. Il est de 3081 mmol/h pour (Ctrl), 2045 mmol/h pour (CN) et 6628 mmol/h pour (HB) pour la maximisation alors qu'il est de 4375 pour le modèle différentiel (cf Table 2.2). Une question naturelle est de comprendre quel est le modèle le plus pertinent. De manière intéressante, différents arguments ont tendance à n'en choisir aucun. En effet, la maximisation du bilan d'ATP implique que la dégradation des protéines s'annule. Ceci est contraire aux observations expérimentales où la synthèse protéique mammaire (peptide synthesis) est toujours supérieure à la production de

protéines dans le lait (peptide output). Sachant que la quantité de protéines dans la glande mammaire est constante, cela signifie nécessairement qu'il y a une dégradation non négligeable des protéines (peptide hydrolysis) dans la glande mammaire. Par exemple pour les jeux de données (Ctrl), (CN) et (HB) la dégradation des protéines (R_{64}) correspond respectivement à 0.25 ; 0.23 ; 0.67 fois la synthèse des protéines (R_{63}) [HFM⁺09, RLL⁺06].

Le cycle de synthèse et de dégradation des protéines de la glande mammaire qui représente le nombre de liens peptidiques synthétisé et hydrolysé est appelé *turnover protéique*.

La première conclusion que nous faisons ici est que le fonctionnement de la glande mammaire ne peut pas être décrit par la maximisation du bilan d'ATP. Pour étudier la variabilité dans l'espace de distributions des flux et identifier une stratégie d'optimisation pertinente dans le métabolisme de la glande mammaire, une contrainte additionnelle prenant en compte le bilan d'ATP doit être ajoutée au système.

Modèle Biologique (mmol/h/demi-mamelle) <i>Critère de sélection des solutions</i>	jeu de données dégradation des protéines	Bilan ATP	
Modèle de la glande mammaire de la Figure 2.1 <i>Résultats de la calculatrice métabolique</i> [VM02]	(HB)	6628	0
	(Ctrl)	3081	0
	(CN)	2045	0
Modèle de la glande mammaire de la Figure 2.1 <i>résultats de la ATP optimization</i>	(HB)	6628	0
	(Ctrl)	3081	0
	(CN)	2045	0
Modèle d'Hannigan-Baldwin [Han94] <i>étude d'états d'équilibre de modèle différentiel</i> [HFM ⁺ 09]	(HB)	4375=6500-2125	0
	(Ctrl)	<i>indisponible</i>	
	(CN)	<i>indisponible</i>	

TABLE 2.2 – Trois méthodes différentes de calcul de bilans d'ATP pour les trois jeux de données (Ctrl), (CN) et (HB). *Les méthodes de la maximisation du bilan d'ATP et de la calculatrice métabolique donne les mêmes résultats pour les 3 jeux de données. Le bilan d'ATP est estimé à 4375 mmol/h pour le jeu de données (HB) dans le modèle différentiel Hanigan [Han94]. Les trois méthodes donnent des valeurs nulles sur la dégradation des protéines (contraire à la littérature [HFM⁺09, RLL⁺06]).*

2.4 Explorer les distributions extrémales connaissant le bilan d'ATP

2.4.1 Étude de la flexibilité à ATP fixé

La section précédente montre que les deux modèles (calculatrice métabolique et maximisation du bilan d'ATP) ne sont pas biologiquement valides à ATP libre. Ainsi, dans cette étude nous fixons le bilan d'ATP égal à 1250 mmol/h/demi-mamelle (c'est-à-dire

60/mol/j/mamelle comme estimé dans [HFM⁺09]). Cette valeur correspond à la quantité d'ATP supposée être utilisée pour des fonctions autres que la synthèse du lait (synthèse des acides nucléiques et de fonctions de maintenance de potentiel de membrane), elle est indépendante du régime alimentaire considéré. Les résultats présentés ci-dessous restent également valables pour une variation de l'ordre de 10% de cette valeur pour les bilans d'ATP.

Il faut noter que dans les 4375 mmol/h/demi-mamelle d'ATP correspondant au bilan d'ATP (après synthèse des constituants du lait) indiqué ci-dessus pour les modèles différentiels, les 1250 mmol/h/demi-mamelle correspondent à la quantité d'ATP pour les fonctions indiquées dans le paragraphe précédent. Le reste, c'est-à-dire 4375 - 1250 serait d'après Hanigan entièrement consommé par le cycle de turnover protéique qu'il définit comme le fait qu'une grande partie des protéines synthétisées dans la glande mammaire serait re-dégradée.

L'espace de distribution des flux qui était précédemment non borné devient borné avec 4 flux extrémaux lorsque la contrainte d'ATP est fixée à 1250. Les flux extrémaux correspondent aux flux des réactions : peptide hydrolysis (R_{64}), NADPH oxidation (R_{19}), OAA \rightarrow PYR (R_{14}), OAA \rightarrow G3P (R_{15}). Plus précisément le flux de la réaction G3P \rightarrow G6P (R_8) qui précédemment était libre devient contraint par les flux des 4 réactions (R_{14} , R_{15} , R_{19} , R_{64}).

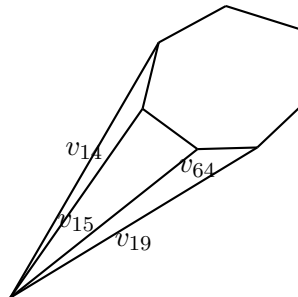


FIGURE 2.5 – Espace de distribution de flux à ATP fixé à 1250 mmol/h/demi-mamelle.

2.4.2 Identification des distributions extrémales

Les sommets extrémaux de l'espace de distribution des flux sont calculés pour les trois jeux de données (Ctrl), (CN) et (HB) en utilisant l'algorithme de tirage aléatoire présenté en annexe A. Il consiste à se donner aléatoirement, un nombre de tirages de fonctions objectifs à effectuer. Pour chaque tirage de fonction objectif, on exécute l'algorithme du simplexe pour obtenir un sommet extrémal. Les différents sommets extrémaux sont déterminés au cours des tirages successifs. L'ensemble des sommets extrémaux est déterminé après avoir opéré plusieurs tirages.

Il faut noter que pour le calcul des sommets extrémaux n'importe quelle autre méthode de FBA existante peut être utilisée. Dans notre cas, les différentes données ont contraint le modèle et ont rendu l'approche probabiliste utilisable et efficace. Le choix de cette méthode plutôt qu'une approche conventionnelle a été motivé par un souci d'uniformité avec les approches développées dans les chapitres suivants, en particulier le chapitre 4.

Ainsi, pour les jeux de données (Ctrl) et (CN), 8 sommets extrémaux ont été trouvés et 13 sommets pour le jeu de données (HB). Les sommets extrémaux sont classés dans la Table 2.3 ci-dessous en fonction de l'activation ou non des réactions R_8 , R_{14} , R_{15} , R_{19} , R_{64} .

Ensuite pour chaque sommet extrémal, nous avons identifié une fonction objectif qu'il permet d'optimiser. Chacune des fonctions objectifs s'écrit sous la forme : $av_{14} + bv_{15} + cv_{19} + dv_{64}$ où v_{14} , v_{15} , v_{19} et v_{64} sont les flux des réactions R_{14} , R_{15} , R_{19} , R_{64} et a , b , c , d des nombres réels.

Pour (Ctrl) et (CN) les distributions extrémales (sommets extrémaux) sont classifiées de manière semblables. D'une part le flux de la réaction NADPH oxydation (R_{19}) est fixé à zéro ou à sa valeur maximale. Ensuite, un des flux des réactions $OAA \rightarrow PYR$ (R_{14}), $OAA \rightarrow G3P$ (R_{15}), $G3P \rightarrow G6P$ (R_8), *peptide hydrolysis* (R_{64}) est fortement activé alors que les autres sont bloqués.

Pour (HB), on peut remarquer que R_{14} , R_{15} et R_{19} ne peuvent pas être activées (flux non nuls) ou inactivées (flux nuls) en même temps. Plus précisément si R_{15} est activée, alors R_{14} ou R_{19} est inactivée. Si R_{15} est faiblement activée alors R_8 ou R_{64} est fortement activée. C'est ce qui explique le nombre plus important de distributions extrémales associé à ce jeu de données.

D'autres part, les valeurs minimales et maximales d'utilisation des flux correspondant aux 4 réactions R_{14} ($OAA \rightarrow PYR$), R_{15} ($OAA \rightarrow G3P$), R_8 ($G3P \rightarrow G6P$) et R_{64} (*peptide hydrolysis*) sont déterminées par la méthode de *flux variability analysis* (FVA) [GT10]. Les valeurs maximales des flux de ces réactions sont respectivement : 1831, 1831, 669, 305 mmol/h pour (Ctrl) ; 795, 795, 22, 133 mmol/h pour (CN) et 5407, 5407, 1196, 896 mmol/h pour (HB).

La valeur minimale de chaque flux est égale à 0 pour chacun des 3 jeux de données.

En conclusion, le jeu de données (HB) présente plus de flexibilité, puis le jeu de données (Ctrl) et enfin le jeu (CN).

2.4.3 Validation biologique

L'analyse faite dans la Table 2.3 pour discuter de la pertinence biologique des distributions extrémales est basée essentiellement sur les données de la littérature sur le flux de la dégradation des protéines (R_{64}) et la synthèse des protéines (R_{63}) [HFM⁺09, LRL⁺09].

En effet, la dégradation des protéine représentée par le flux v_{64} de la réaction R_{64} doit être toujours non nulle et la synthèse des protéines doit être au maximum égale à trois fois le flux des protéines exportées dans le lait ($v_{63} \leq 3(v_{63} - v_{64})$) [HFM⁺09, LRL⁺09]. Cette deuxième condition, $v_{63} \leq 3(v_{63} - v_{64})$, nous donne un rapport $\frac{v_{64}}{v_{63}}$ au maximum égal à 0.67. Avec ce critère, toutes les distributions extrémales de A à F pour (Ctrl) et (CN) sont biologiquement très peu réalistes. La totalité des distributions extrémales pour (HB) est sujette à caution car le rapport $\frac{v_{64}}{v_{63}}$ est supérieur ou égal à 0.85.

Pour les distributions extrémales A à F, on peut également remarquer que les valeurs des flux des réactions R_8 ou R_{13} sont plus grandes que les valeurs maximales (de V_{\max} et d'activité) données dans la littérature [SBC76]. Toutes les valeurs des flux des distributions extrémales sont données dans la Table 2.3 ci-dessus.

Par contre, avec la stœchiométrie du modèle de la Figure 2.1, les distributions extrémales G et H, maximisant respectivement les fonctions objectifs $v_{64} + v_{19}$ et $v_{64} + v_{19}$, sont cohérentes avec les données de la littérature sur les flux des réactions R_8 , R_{13} , R_{63} et R_{64} pour les jeux de données (Ctrl) et (CN). En effet, pour ces deux distributions extrémales, les flux des réactions R_8 et R_{13} n'excèdent pas les valeurs maximales données dans la littérature [SBC76]. Le rapport $\frac{v_{64}}{v_{63}}$ est égal à 0.7 pour (Ctrl) et 0.47 pour (CN). Les rapports sont donc approximativement inférieur ou égal à 0.67 donné dans littérature [HFM⁺09].

Cependant, on s'attendait à ce que la synthèse des protéines (peptide synthesis) soit plus élevée dans (CN) que dans (Ctrl) car la synthèse des protéines devrait augmenter avec l'apport protéique comme c'est le cas de (CN) [RLL⁺06]. Curieusement, c'est le contraire qui se produit pour les distributions extrémales G et H. En effet, la synthèse des protéines décroît pour (CN) lorsqu'on le compare à (Ctrl). Le flux correspondant à la synthèse des protéines est égal à 430 ou 410 mmol/h/demi-mamelle pour (Ctrl) et 283 ou 282 mmol/h/demi-mamelle pour (CN).

En conclusion la totalité des distributions extrémales associée au jeu (HB) et six distributions extrémales sur les huit, pour chacun des jeux de données (Ctrl) et (CN), sont en contradiction avec les données de la littérature et ne peuvent pas être biologiquement valides. Il reste pour chacun des deux jeux de données (Ctrl) et (CN), deux distributions extrémales qui vérifient toutes les conditions disponibles mais elles sont sujettes à caution vis à vis des différents critères qui ne peuvent pas être intégrés dans le processus d'analyse formellement. Cette étude constitue la première partie acceptée pour publication *BMC Systems Biology* [AALVM⁺on].

Dans le chapitre suivant, nous formalisons une approche qui est très utilisée dans le domaine de la nutrition la contribution des entrées (Inputs) dans les sorties (Outputs), AIO, pour davantage discuter de la pertinence biologique de chacune des deux distributions extrémales de (Ctrl) et (CN).

Jeux de données	Exemple de fonction à maximiser	Valeurs des flux des distributions extrémales					Validation		Flux incohérents avec la littérature					
		R_{19} NADPH oxydation	R_{15} OAA → G3P	R_{14} OAA → Pyr	R_8 G3P → G6P	R_{64} peptide hydrolysis	R_{63} peptide synthesis	R_{13} Pyr → OAA						
Distributions extrémales de flux dans l'espace de distribution des flux	(Ctrl)	B	$v_{15}-v_{19}$	0	1831	0	0	0	125	1835	R_{13}, R_{64}			
	(CN)			795	795			150	803					
	(HB)			5407	5407			150	5378					
	(Ctrl)	F	$v_{14}-v_{19}$	0	0	1831	0	0	125	1835	R_{13}, R_{64}			
	(CN)				795			150	803					
	(HB)				5407			150	5378					
	(Ctrl)	D	v_8-v_{19}	0	0	0	3662	0	125	4	R_8, R_{64}			
	(CN)						1590		150	8				
	(HB)								150	0				
		D1	$v_8-v_{19}-v_{14}$	0	29	0	29	10757	0	150	0			
													D2	$v_8-v_{19}-v_{15}$
	(Ctrl)													
	(CN)													
	(HB)													
	(Ctrl)	H	$v_{64}-v_{19}$	0	0	0	0	305	430	4				
	(CN)							133	283	8				
	(HB)													
		H1	$v_{64}-v_{19}-v_{14}$	0	29	0	29	0	896	1046	0	R_{63}		
													H2	$v_{64}-v_{19}-v_{15}$
	(Ctrl)													
	(CN)													
	(HB)													
	(Ctrl)	A	$v_{15}+v_{19}$	669	1714	0	0	0	125	1718	R_{13}, R_{64}			
	(CN)			22	791			150	799					
(HB)	1196			5198			150	5170						
(Ctrl)	E	$v_{14}+v_{19}$	669	0	1714	0	0	125	1718	R_{13}, R_{64}				
(CN)			22	791			150	799						
(HB)			1196	29	5169		150	5169						
	E1	$v_{14}+v_{19}$	1196	29	5169	0	0	150	5213					
											E2	$v_{14}+v_{19}-50v_{15}$		
(Ctrl)														
(CN)														
(HB)														
(Ctrl)	C	v_8+v_{19}	669	0	0	3428	0	125	4	R_8, R_{64}				
(CN)			22	1583			150	8						
(HB)			1196	29	10338		150	0						
	C1	v_8+v_{19}	1196	29	0	10338	0	150	0					
											C2	$v_8+v_{19}-50v_{15}$		
(Ctrl)														
(CN)														
(HB)														
(Ctrl)	G	$v_{64}+v_{19}$	669	0	0	0	286	410	4					
(CN)			22	132			282	8						
(HB)			1196	29	0	861	1011	0						
	G1	$v_{64}+v_{19}$	1196	29	0	0	861	1011	1019					
											G2	$v_{64}+v_{19}-50v_{15}$		
(Ctrl)														
(CN)														
(HB)														
bornes supérieures de flux tirées de la littérature						≤ 591 mmol/h/demi- mamelle [SBC76]	Non nul [HFM ⁺ 09, LRL ⁺ 09]	$\leq 3(v_{63}-v_{64})$ [HFM ⁺ 09]	≤ 266 mmol/h/demi- mamelle [SBC76]					

TABLE 2.3 – Propriétés des différentes distributions extrémales pour un bilan d'ATP constant pour les jeux de données (Ctrl), (CN) et (HB). 8 distributions extrémales (A-F) pour chacun des jeux de données (Ctrl) et (CN), 13 distributions extrémales pour (HB) (A, B, C1, C2, D1, D2, E1, E2, F). A chaque distribution on peut associer une fonction objectif qu'elle optimise. Par exemple, la distribution extrême A maximise la fonction objectif qui est la somme des flux des réactions R_{15} et R_{19} ($v_{15} + v_{19}$). On peut classer les distributions extrémales en fonction des voies qu'elles activent ou non et des fonctions objectifs qu'elles peuvent optimiser. Par exemple, la distribution B consiste à ne pas emprunter R_{19} , R_{14} , R_8 et R_{64} et à faire circuler un maximum de flux dans R_{15} . La valeur de v_{15} dépend alors des jeux de données (Ctrl), (CN) et (HB). Ainsi, 3 distributions (A, B, F) présentent les mêmes caractéristiques pour les trois jeux de données. 5 distributions extrémales (C, D, E, G, H) sont identiques pour (Ctrl) et (CN), chacun produit 2 distributions (C1, C2); (D1, D2); (E1, E2); (G1, G2) et (H1, H2) pour (HB).

Chapitre 3

Contribution des entrées (Inputs) dans les sorties (Outputs) : AIO

Dans le chapitre 2 nous avons introduit un modèle de la glande mammaire et montré que, en conditions expérimentales, la plupart des distributions extrémales susceptibles d'être l'unique extrémum d'un objectif linéaire, c'est-à-dire les sommets de l'espace de recherche, sont sujettes à caution, car l'ordre de grandeur des valeurs de certains flux est largement supérieur aux valeurs maximales proposées dans la littérature. L'analyse de deux seules distributions extrémales dont les ordres de grandeurs sont raisonnables soulève un doute sur leur pertinence biologique. Mais il existe également, dans la littérature, des données sur les proportions de flux de carbones entrant de certains nutriments prélevés pour produire des constituants du lait. Par exemple, dans [BAL74], il est montré que la totalité du flux de carbone du Lactose ne vient pas du Glucose. Pour pouvoir tirer partie de cette information, il est nécessaire de déterminer la contribution des nutriments prélevés dans les produits formés (constituants du lait).

La *contribution des entrées dans les sorties* est une notion très liée au domaine biologique de la production animale mais aucune formalisation précise de cette notion n'est disponible dans la littérature. Ainsi, dans ce chapitre nous introduisons une formalisation générique et une définition de cette contribution que nous désignerons par AIO. Nous présentons un calcul des AIOs dans des cas particuliers en utilisant des parcours de graphe. Nous allons montrer les limites de cette approche et présenter une nouvelle méthode qui consiste à inverser une matrice pour calculer ces contributions dans un cas général, en particulier en présence de cycles.

3.1 Problème posé

Lorsqu'on étudie un métabolisme à l'échelle d'un organe comme dans le domaine de la nutrition, on procède souvent à des comparaisons *d'efficacité* ou *rendement* de deux situations métaboliques différentes. Cependant, le terme d'efficacité dépend du champ d'application considéré. Il peut s'agir d'une efficacité en terme d'énergie, de biomasse ou

de transformation de protéines [BABM⁺11, VMGR03].

Des exemples d'estimation d'efficacité sont la contribution des carbones du glucogène pour produire les carbones du Glucose dans le foie [RT91] ou la contribution des carbones du Glucose pour produire les carbones du Lactose dans la glande mammaire [LRL⁺09].

Pour expliciter cette notion, considérons un réseau métabolique utilisant deux nutriments d'entrée I_1 et I_2 pour produire deux produits O_1 et O_2 qui sont des produits finaux et qui sortent du système (par exemple : Lactose et CO_2 produits à partir du Glucose et Glycérol3P). Les nutriments I_1 et I_2 sont transformés en différents métabolites P_1 , P_2 , et P_3 selon le réseau décrit à la Figure 3.1.

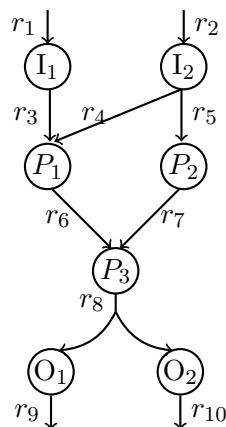


FIGURE 3.1 – Réseau métabolique transformant deux nutriments d'entrées I_1 et I_2 pour produire deux sorties O_1 et O_2 .

La question à laquelle on cherche à répondre est la suivante : quelles proportions des entrées I_1 et I_2 se répartissent exactement dans les sorties O_1 et O_2 ?

Cette question est généralement abordée en étudiant les répartitions de carbones. Autrement dit, la question devient : comment les carbones des entrées I_1 et I_2 sont-ils répartis dans les carbones des sorties O_1 et O_2 ? En production animale la réponse à cette question permet de comparer des situations alimentaires différentes en terme d'efficacité à transformer les nutriments prélevés pour produire du lait ou de la viande.

Le calcul d'efficacité consiste donc à déterminer la contribution des carbones des entrées dans les carbones des sorties (en anglais "allocation of input over output" ou AIO). Cependant aucune formalisation précise des AIOs n'est disponible dans la littérature. Nous allons introduire ici une formalisation et une définition générique des AIOs. Notre formalisation permet de répondre à la question ci-dessus posée et peut donc être vue comme un calcul d'efficacité car elle permettra, par exemple, de déterminer la contribution des carbones du glucogénique aux carbones du Glucose ou des carbones du Glucose aux carbones du Lactose. Elle permet également de comparer des distributions de flux dans un réseau métabolique comme nous allons le voir dans ce chapitre.

Le calcul des AIOs, que nous allons introduire, dans un cas général de réseau métabolique avec cycles s'appuie sur un résultat mathématique d'inversion de matrice. Cette méthode a permis de contourner les limites des approches de parcours de graphe liées à la présence de cycles dans les réseaux.

3.2 Formalisation et modélisation des AIOs

Nous avons défini dans l'introduction :

- $\mathcal{I} = \{I_1, I_2, \dots, I_p\}$, désigne l'ensemble des métabolites d'entrées ;
- $\mathcal{P} = \{P_1, P_2, \dots, P_n\}$, l'ensemble des métabolites intermédiaires ;
- $\mathcal{O} = \{O_1, O_2, \dots, O_q\}$, l'ensemble des métabolites de sorties ;
- $m_i = I_i$ pour $1 \leq i \leq p$;
- $m_{p+i} = P_i$ pour $1 \leq i \leq n$;
- $m_{p+n+i} = O_i$ pour $1 \leq i \leq q$.

Nous choisissons un élément à travers lequel faire le calcul de répartition. Nous allons noter cet élément φ , qui peut être du carbone, de l'azote etc.

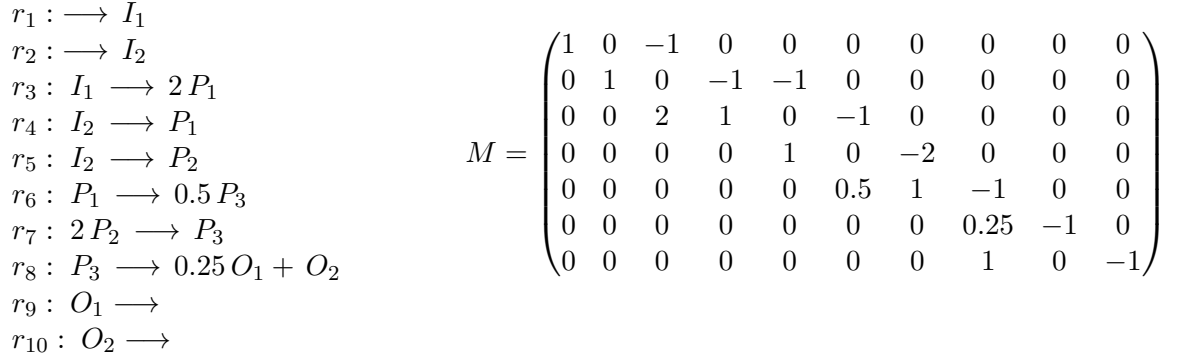
On suppose connu le vecteur $c \in \mathbb{N}^{p+n+q}$ correspondant au nombre d'atomes de l'élément de répartition contenu dans chaque métabolite (nombre d'atomes de carbone par exemple). Ce vecteur est de taille $n + p + q$ (n : nombre de métabolites d'entrées, p : nombre de métabolites intermédiaires, q : nombre de métabolites de sorties).

Comme indiqué en introduction, on note $M = (a_{i,j})$, de dimension $(n + p + q) \times n_{\mathcal{R}}$, la matrice stœchiométrique du réseau métabolique, avec $n_{\mathcal{R}}$ le nombre de réactions du système. Une hypothèse fondamentale pour appliquer notre méthode est la suivante : on suppose que la stœchiométrie de chaque réaction du système satisfait la propriété de la conservation de la matière par rapport à φ . Ce qui signifie que :

$$\sum_{m_i \in \mathcal{M}} a_{i,j} c(m_i) = 0 \text{ pour toute réaction } r_j \in \mathcal{R}, \quad (3.1)$$

avec \mathcal{M} l'ensemble des métabolites du système.

La Figure 3.1 est une représentation sous forme de graphe d'un réseau métabolique. Supposons que $c = (4, 2, 2, 2, 4, 4, 3)$ est le vecteur correspondant au nombre d'atomes de carbones des métabolites $(I_1, I_2, P_1, P_2, P_3, O_1, O_2)$. En supposant que pour chaque réaction la propriété (3.1) est satisfaite pour les carbones, le réseau peut être décrit par les réactions, r_1 à r_{10} , ci-dessous. Il est de matrice stœchiométrique M (les lignes correspondent aux métabolites et les colonnes aux réactions).



Pour toute réaction r_j et pour chaque substrat de la réaction, on définit le pourcentage du substrat par rapport à l'élément φ qui contribue à la formation d'un produit m_k de cette réaction. Soit $In = (In^{r_j}(m_k))$, la matrice de dimension $(n + p + q) \times n_{\mathcal{R}}$, correspondant à ces pourcentages. Elle est définie comme suit :

$$In^{r_j}(m_k) = \begin{cases} \frac{a_{k,j}c(m_k)}{\sum_{m_i \in \mathcal{M}, a_{i,j} < 0} |a_{i,j}|c(m_i)} & \text{si } m_i \text{ est substrat dans } r_j \text{ (i.e. } a_{i,j} < 0) \\ 0 & \text{sinon} \end{cases} \quad (3.2)$$

Avec cette définition de In , on a fait l'hypothèse que les carbones jouent un rôle équivalent dans leurs transformations chimiques. C'est-à-dire que chaque produit de la réaction r_j reçoit un nombre de carbones proportionnel à la quantité de substrat de la réaction. Ceci pourrait être amélioré par exemple en spécifiant les types de carbones sur chaque produit lorsqu'on dispose de données expérimentales de transfert de carbones entre les substrats et les produits dans les différentes réactions.

En conséquence de la loi de conservation de la matière nous avons :

$$\sum_{m \in \mathcal{M}} In^r(m) = 1 \text{ pour toute réaction } r \in \mathcal{R}, \quad (3.3)$$

où \mathcal{R} désigne l'ensemble des réactions du système.

Avec les données ci-dessus pour la Figure 3.1 (matrice M et le vecteur c), la matrice In correspondante aux pourcentages de carbones reçus par les métabolites à travers chaque réaction, est la suivante (les lignes correspondent aux métabolites et les colonnes aux réactions)

$$In = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.25 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.75 & 0 & 0 & 0
\end{pmatrix} \quad (3.4)$$

Par exemple, sur la huitième colonne $In^{r_8}(O_1) = 0.25$ signifie que 25 % des carbones de P_3 sont utilisés pour produire O_1 . $In^{r_8}(O_2) = 0.75$ signifie que 75 % des carbones de P_3 sont utilisés pour produire O_2 .

Soit v un vecteur de distribution de flux tel qu'il n'y ait pas d'accumulation de métabolites intermédiaires. C'est-à-dire un vecteur v qui vérifie $M \begin{pmatrix} v_{\mathcal{I}} \\ v \\ v_{\mathcal{O}} \end{pmatrix} = 0$, où $v_{\mathcal{I}}$ est le vecteur des flux des réactions des métabolites d'entrées et $v_{\mathcal{O}}$, le vecteur des flux des réactions des métabolites de sorties.

Connaissant les flux des réactions d'entrées et de sorties, l'état stationnaire permet de déterminer les flux de toutes les autres réactions pour les métabolites intermédiaires soient équilibrés. C'est-à-dire tout ce qui est créé est également consommé.

Par exemple, en posant $v_{\mathcal{I}} = (5, 10)$, et $v_{\mathcal{O}} = (2.5, 10)$, comme l'indique la Figure 3.2, l'état stationnaire permet de déterminer le vecteur $v = (v_3, v_4, v_5, v_6, v_7, v_8)$ des flux des réactions $r_3, r_4, r_5, r_6, r_7, r_8$ pour que les métabolites intermédiaires P_1, P_2 et P_3 soient équilibrés. Ce qui revient à résoudre le système (3.5) suivant :

$$\begin{cases} v_3 = 5 \\ v_4 + v_5 = 10 \\ 2v_3 + v_4 = v_6 \\ v_5 = 2v_7 \\ 0.5v_6 + v_7 = v_8 \\ 0.25v_8 = 2.5 \\ v_8 = 10 \end{cases} \quad (3.5)$$

Le système 3.5 est équivalent au système 3.6 suivant :

$$\begin{cases} v_3 = 5 \\ v_8 = 10 \\ v_5 = 10 - v_4 \\ v_6 = 10 + v_4 \\ v_7 = 5 - 0.5v_4 \end{cases} \quad (3.6)$$

Le système 3.6 est un système d'équations à une indéterminée v_4 . Il admet une infinité de solutions en fixant des valeurs à v_4 . Par exemple pour $v_4 = 5$, $v = (5, 5, 5, 15, 2.5, 10)$ est un vecteur de flux qui permet d'équilibrer les métabolites intermédiaires P_1, P_2 et P_3 .

On introduit maintenant une matrice W pour représenter les pourcentages d'utilisation du flux à travers le pool de chaque métabolite dans les différentes réactions. Pour l'état stationnaire correspondant à $v = (5, 5, 5, 15, 2.5, 10)$ indiquée sur la Figure 3.3 dans les différentes réactions pour chaque métabolite est la suivante (les lignes correspondent aux réactions et les colonnes aux métabolites)

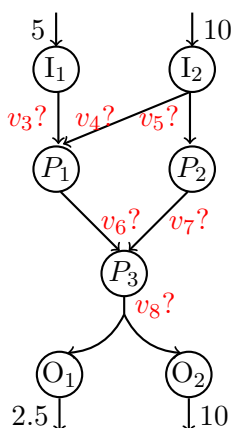


FIGURE 3.2 – Connaissant $v_{\mathcal{I}} = (5, 10)$ le vecteur des flux des réactions d’entrées et $v_{\mathcal{O}} = (2.5, 10)$ le vecteur des flux des réactions de sorties, l’état stationnaire permet de déterminer les vecteurs $v = (v_3, v_4, v_5, v_6, v_7, v_8)$ des flux des autres réactions pour que tout ce qui est créé pour les métabolites intermédiaires P_1 , P_2 et P_3 soit également consommé.

$$W = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad (3.7)$$

Chaque colonne de matrice W , correspond à la proportion de flux du métabolite correspondant dans les différentes réactions. En considérant la deuxième colonne, elle correspond au métabolite I_2 du réseau. Ce métabolite n’est utilisé que dans les réactions r_4 et r_5 . Or, le flux total en I_2 est égal à 10. Le flux de chacune des réactions r_4 et r_5 est égal à 5. Donc la proportion du flux du métabolite I_2 dans chacune des réactions r_4 et r_5 est égal à $5/10 = 0.5$.

Le calcul précédent se fait manuellement sur un exemple précis. Pour généraliser le concept de pourcentage d’utilisation, on introduit la matrice $W = (w_{j,k})$, de dimension $n_{\mathcal{R}} \times (n + p + q)$, définie par :

$$w_{j,k} = \begin{cases} \frac{|a_{k,j}|v_j}{F(m_k)} & \text{si } m_k \text{ est substrat dans } r_j \text{ (i.e. } a_{k,j} < 0) \\ 0 & \text{sinon} \end{cases} \quad (3.8)$$

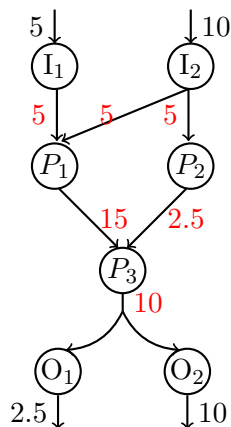


FIGURE 3.3 – Exemple de vecteur de flux v qui équilibre P_1 , P_2 et P_3 . Avec $v_I = (5, 10)$ le vecteur des flux des réactions d'entrées et $v_O = (2.5, 10)$ le vecteur des flux des réactions de sorties, $v = (5, 5, 5, 15, 2.5, 10)$ permet d'équilibrer les métabolites intermédiaires P_1 , P_2 et P_3 en tenant compte de la stœchiométrie du réseau.

avec

$$F(m_k) = \sum_{r_\eta \in \mathcal{R}, a_{k,\eta} > 0} a_{k,\eta} v_\eta = \sum_{r_\eta \in \mathcal{R}, a_{k,\eta} < 0} |a_{k,\eta}| v_\eta,$$

le flux à travers le pool du métabolite intermédiaire, m_k , produit avant d'être dégradé par d'autres réactions.

Enfin, on peut remarquer que pour tout métabolite d'entrée m , le flux d'un métabolite d'entrée par rapport à l'élément de répartition φ est égal à $x^m = c(m)v_m$, où v_m désigne le flux de l'entrée m . A l'aide des différentes matrices introduites précédemment, il devient possible d'établir un ensemble de relations entre les proportions de carbones reçues par chaque métabolite (intermédiaire ou de sortie).

Proposition 3.2.1. Notons $x[v, m](m_k)$ la proportion du flux de φ du métabolite d'entrée m dans le métabolite m_k pour $p + 1 \leq k \leq p + n + q$.

Ainsi, $x[v, m](m_k)$, pour $p + 1 \leq k \leq p + n + q$, vérifie la relation suivante :

$$x[v, m](m_k) = \underbrace{\sum_{r_j \in \mathcal{R}, a_{k,j} > 0}}_{z_1} \underbrace{\sum_{m_i \in \mathcal{M}, a_{i,j} < 0}}_{z_2} \underbrace{In^{r_j}(m_k)}_{z_3} \times \underbrace{w_{j,i}}_{z_4} \times \underbrace{x[v, m](m_i)}_{z_5} \quad (3.9)$$

$\underbrace{\hspace{15em}}_{z_6}$

Où, pour $1 \leq i \leq p$

$$x[v, m](m_i) = \begin{cases} c(m)v_m & \text{si } m = m_i \\ 0 & \text{sinon.} \end{cases} \quad (3.10)$$

- Démonstration.* z_1 représente les réactions, r_j , produisant le métabolite m_k et z_2 les substrats dans r_j ;
 z_3 le coefficient correspondant à la proportion de substrat m_i par rapport à l'élément de répartition φ transformé en produit m_k à travers r_j ;
 z_4 le rapport est égal à la proportion du flux total de m_i transformé en m_k dans r_j ;
 z_5 est la proportion du flux de l'entrée m par rapport à l'élément de répartition arrivant en m_i ;
 z_6 le flux total de l'entrée m par rapport à l'élément de répartition arrivant en à travers tous les substrats de la réaction r_j .

□

Sur l'exemple de la Figure 3.3, pour $v = (5, 5, 5, 15, 2.5, 10)$, les coefficients $x[v, m]$ de flux de carbone I_1 et I_2 dans les métabolites de sortie O_1 , et O_2 sont contraints par les relations suivantes :

$$\begin{cases} x[v, I_1](P_1) = x[v, I_1](I_1) \\ x[v, I_1](P_2) = 0 \\ x[v, I_1](P_3) = x[v, I_1](P_1) \\ x[v, I_1](O_1) = 0.25x[v, I_1](P_3) \\ x[v, I_1](O_2) = 0.75x[v, I_1](P_3) \end{cases}$$

et

$$\begin{cases} x[v, I_2](P_1) = 0.5x[v, I_2](I_2) \\ x[v, I_2](P_2) = 0.5x[v, I_2](I_2) \\ x[v, I_2](P_3) = x[v, I_2](P_1) + x[v, I_2](P_2) \\ x[v, I_2](O_1) = 0.25x[v, I_2](P_3) \\ x[v, I_2](O_2) = 0.75x[v, I_2](P_3) \end{cases}$$

avec $x[v, I_1](I_1) = x^{I_1} = 5 * 4 = 20$ et $x[v, I_2](I_2) = x^{I_2} = 10 * 2 = 20$.

Ainsi, finalement on note par :

$x_{\mathcal{O}}[v, m] = (x[v, m](m_{p+n+1}), x[v, m](m_{p+n+2}), \dots, x[v, m](m_{p+n+q}))^{\top} \in \mathbb{R}^q$ le vecteur correspondant aux proportions du flux de φ du métabolite d'entrée m dans les différentes sorties. $(.)^{\top}$ désigne la transposée et \mathcal{O} l'ensemble des métabolites de sorties. Par exemple, la proportion du flux de *carbone* du glucose qui entre dans la composition du lactose et du CO_2 libéré dans le sang.

Définition 3.2.1. *Les coefficients du vecteur $x_{\mathcal{O}}[v, m]$ sont appelés contribution du métabolite d'entrée m dans l'ensemble des métabolites de sorties \mathcal{O} .*

Dans notre exemple, on constate que ces valeurs sont uniquement déterminées par :

$$\begin{pmatrix} x[(5, 5, 5, 15, 2.5, 10), I_1](O_1) & x[(5, 5, 5, 15, 2.5, 10), I_2](O_1) \\ x[(5, 5, 5, 15, 2.5, 10), I_1](O_2) & x[(5, 5, 5, 15, 2.5, 10), I_2](O_2) \end{pmatrix} = \begin{pmatrix} 5 & 5 \\ 15 & 15 \end{pmatrix}$$

Cela veut dire que, dans les 20 mol/h de carbones entrant dans le système par I_1 , 5 mol/h de carbones se retrouvent dans le flux de O_1 et les 15 autres dans le flux de O_2 . Il en est de même pour I_2 .

3.3 Calcul des AIOs par la méthode du parcours des graphes

On peut appliquer l'algorithme 1, de parcours de graphe en profondeur, ci-dessous pour calculer les AIOs, $x_{\mathcal{O}}[v, m]$, pour toute distribution de flux v dans le cas d'un réseau métabolique sans cycle.

Algorithme 1 Algorithme de calcul des AIOs pour les réseaux métaboliques sans cycles

Entrées : $\mathcal{I}, \mathcal{P}, \mathcal{O}, v$ (une distribution de flux), $x_{\mathcal{I}} = \{x^{I_i}, i \in \{1, 2, \dots, p\}\}$, matrice In , matrice W

Sorties : $x_{\mathcal{O}}[v, m] = \{x_{O_i}[v, m], i \in \{1, 2, \dots, q\} m \in \mathcal{I}\}$

Pour i allant de 1 à p **Faire**

$I_i \leftarrow m; 1 \leftarrow \alpha; x^m \leftarrow Q; x_s[v, I_i] = 0, s \in \mathcal{O}$ (Initialisation)

AIO (m, α, Q)

{

Si NonMarquer(m) **Alors**

Marquer(m)

Pour toute réaction r_j consommant m **Faire**

Pour tout produit m_k de la réaction r_j **Faire**

$m_k \leftarrow m$

$In^{r_j}(m_k) * w(j, i) \leftarrow \alpha$

$\alpha * Q \leftarrow Q$

$k \leftarrow i$

Pour l allant de 1 à q **Faire**

Si ($m_k == O_l$) **Alors**

$x_{O_l}[v, I_i] + \alpha * Q \leftarrow x_{O_l}[v, I_i]$

Fin Si

Fin Pour

AIO (m, α, Q)

Fin Pour

Fin Pour

Fin Si

}

Fin Pour

Sur l'exemple de la Figure 3.3, toutes les données d'entrées pour l'algorithme 1 sont réunies pour le calcul des AIOs de flux de carbone. En effet, on connaît :

- les matrices W et In , les ensembles $\mathcal{I} = \{I_1, I_2\}$, $\mathcal{P} = \{P_1, P_2, P_3\}$, et $\mathcal{O} = \{O_1, O_2\}$, une distribution de flux $v = (5, 5, 5, 15, 2.5, 10)$;
- les flux de carbone des métabolites d'entrées, c'est-à-dire : $x_{\mathcal{I}} = (x^{I_1}, x^{I_2})$ avec $x^{I_1} = 5 * 4 = 20$ et $x^{I_2} = 10 * 2 = 20$.

La sortie de l'algorithme sur ces données donne les résultats décrits précédemment.

Dans l'exemple de la Figure 3.3, le calcul des contributions a pu être réalisé puisque le système d'équations reliant ces contributions était triangulaire. Plus généralement, un phénomène similaire a lieu lorsque le réseau métabolique ne possède aucun cycle.

Proposition 3.3.1. *L'algorithme 1 permet de déterminer la répartition des entrées dans les sorties pour tout réseau métabolique sans cycle.*

Or les réseaux métaboliques sont rarement sans cycle. En particulier dans la glande mammaire où par exemple le Glucose prélevé est métabolisé dans le cycle de Krebs. Si nous introduisons une réaction supplémentaire $P_3 \rightarrow 2P_2$ dans le réseau précédent alors nous aurons un cycle comme le montre la Figure 3.4. Deux problèmes majeurs se présentent dans ce cas :

- si chaque métabolite n'est parcouru qu'une seule fois dans l'algo 1 alors le calcul des AIOs ne sera pas correct car le flux de carbone de I arrivant en P_3 à partir de P_2 ne prendra pas en compte le recyclage dans le cycle ;
- si un métabolite peut être parcouru plusieurs fois alors l'algorithme ne termine pas.

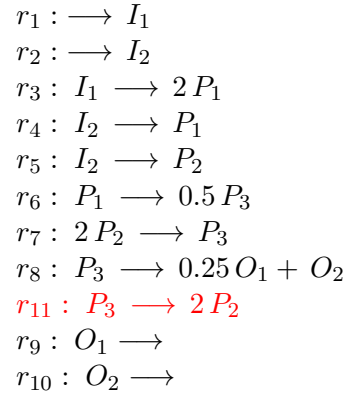
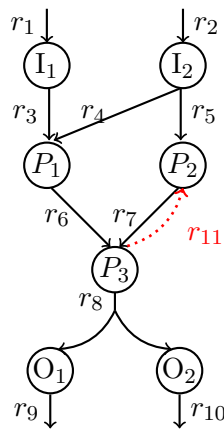


FIGURE 3.4 – Réseau métabolique avec cycle transformant deux nutriments d'entrées I_1 et I_2 pour produire deux sorties O_1 et O_2 .

Pour contourner ce problème, nous avons proposé de modéliser le calcul des AIOs en inversant une matrice appelée M-matrice (toujours inversible).

3.4 Calcul des AIOs dans un cas général

Pour déterminer les coefficients du vecteur $x_{\mathcal{O}}[v, m] \in \mathbb{R}^q$, nous introduisons le vecteur $x_{\mathcal{P}}[v, m] \in \mathbb{R}^n$ correspondant aux proportions du flux de φ du métabolite d'entrée m qui sont nécessaires pour produire les métabolites intermédiaires avant qu'ils ne soient dégradés en d'autres métabolites. En notant $x_{\mathcal{I}}^m$ le vecteur dont toutes les composantes sont nulles sauf celle correspondant à l'entrée m i.e. $x_{\mathcal{I}}^m = (0, \dots, 0, c(m)v_m, 0, \dots, 0)^\top$, $x_{\mathcal{O}}[v, m] \in \mathbb{R}^q$ est donné par le Théorème 3.4.1 ci-dessous.

Théorème 3.4.1. *le vecteur $x_{\mathcal{O}}[v, m]$ correspondant à la répartition du flux du métabolite d'entrée m dans les différentes sorties est solution du système suivant :*

$$D_1[v] \begin{pmatrix} x_{\mathcal{P}}[v, m] \\ x_{\mathcal{O}}[v, m] \end{pmatrix} + D_2[v] x_{\mathcal{I}}^m = 0 \quad (3.11)$$

où

$$D_1[v] = (d_{k,i})_{p < k, i \leq p+n+q}$$

et

$$D_2[v] = (d_{k,i})_{p < k \leq p+n+q, 1 \leq i \leq p}$$

avec

$$d_{k,i} = \begin{cases} 1 & \text{si } k = i \\ - \sum_{r_j \in \mathcal{R}} w_{j,i} * In^{r_j}(m_k) & \text{sinon.} \end{cases} \quad (3.12)$$

$D_1[v]$ est une matrice carrée de dimension $n + q$ dont les éléments diagonaux sont tous égaux à 1 et les éléments non diagonaux sont négatifs.

Démonstration. Comme indiqué dans la section 3.2 la proportion du flux de φ du métabolite d'entrée m dans le métabolite m_k , $x[v, m](m_k)$, pour $p + 1 \leq k \leq p + n + q$ vérifie la relation suivante :

$$x[v, m](m_k) = \sum_{r_j \in \mathcal{R}, a_{k,j} > 0} \sum_{m_i \in \mathcal{M}, a_{i,j} < 0} In^{r_j}(m_k) \times w_{j,i} \times x[v, m](m_i) \quad (3.13)$$

Où, pour $1 \leq i \leq p$

$$x[v, m](m_i) = \begin{cases} c(m)v_m & \text{si } m = m_i \\ 0 & \text{sinon.} \end{cases} \quad (3.14)$$

Notons

$$\begin{aligned} x_{\mathcal{I}}^m &= (0, \dots, 0, c(m)v_m, 0, \dots, 0)^\top \\ x_{\mathcal{P}}[v, m] &= (x[v, m](m_{p+1}), x[v, m](m_{p+2}), \dots, x[v, m](m_{p+n}))^\top \\ x_{\mathcal{O}}[v, m] &= (x[v, m](m_{p+n+1}), x[v, m](m_{p+n+2}), \dots, x[v, m](m_{p+n+q}))^\top \end{aligned}$$

Où $(.)^\top$ désigne la transposée.

On déduit, les équations 3.11 et 3.12. □

La matrice D_1 peut être réduite pour ne prendre en compte que les variables libres du système comme suit :

Propriétés 3.4.1. *Soit v_{libre} un ensemble de variables libres et génératrices de l'ensemble de solution du système :*

$$M \begin{pmatrix} v_{\mathcal{I}} \\ v \\ v_{\mathcal{O}} \end{pmatrix} = 0, \quad Bv = 0.$$

Alors, il existe une fonction $v = f(v_{\text{libre}}) = (f_1(v_{\text{libre}}), f_2(v_{\text{libre}}), \dots, f_k(v_{\text{libre}}))$, telle que $f(\mathbb{R}^*)$ décrive l'ensemble des vecteurs v solutions du système ci-dessus.

Alors

$$\bar{D}_1[v_{\text{libre}}] \begin{pmatrix} \bar{x}_{\mathcal{P}}[v_{\text{libre}}, m] \\ \bar{x}_{\mathcal{O}}[v_{\text{libre}}, m] \end{pmatrix} + \bar{D}_2[v_{\text{libre}}] x_{\mathcal{I}}^m = 0 \quad (3.15)$$

Soit :

$$\bar{D}_1[v_{\text{libre}}] = D_1[f(v_{\text{libre}})]; \quad \bar{D}_2[v_{\text{libre}}] = D_2[f(v_{\text{libre}})]$$

$$\bar{x}_{\mathcal{P}}[v_{\text{libre}}, m] = x_{\mathcal{P}}[f(v_{\text{libre}}), m]; \quad \bar{x}_{\mathcal{O}}[v_{\text{libre}}, m] = x_{\mathcal{O}}[f(v_{\text{libre}}), m]$$

Exemple 3.4.1. Dans l'exemple du réseau avec cycle de la Figure 3.4, le système 3.5 avec $v_{\mathcal{I}} = (5, 10)$, $v_{\mathcal{O}} = (2.5, 10)$ et $v = (v_3, v_4, v_5, v_6, v_7, v_8, v_{11})$ devient le système suivant

$$\begin{cases} v_3 = 5 \\ v_4 + v_5 = 10 \\ 2v_3 + v_4 = v_6 \\ v_5 + 2v_{11} = 2v_7 \\ 0.5v_6 + v_7 = v_8 + v_{11} \\ 0.25v_8 = 2.5 \\ v_8 = 10 \end{cases} \quad (3.16)$$

et les matrices $D_1[v]$ et $D_2[v]$, avec $v = (v_3, v_4, v_5, v_6, v_7, v_8, v_{11})$, sont les suivantes :

$$D_1[v] = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & \frac{-v_{11}}{0.5v_6+v_7} & 0 & 0 \\ \frac{-v_6}{2v_3+v_4} & 0 & 1 & 0 & 0 \\ 0 & \frac{-2v_7}{v_5+2v_{11}} & \frac{-v_8}{0.5v_6+v_7} & 0.25 & 1 & 0 \\ 0 & 0 & \frac{-v_8}{0.5v_6+v_7} & 0.75 & 0 & 1 \end{pmatrix} \quad (3.17)$$

$$D_2[v] = \begin{pmatrix} \frac{-v_3}{5} & \frac{-v_4}{10} \\ 0 & \frac{-v_5}{10} \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \quad (3.18)$$

Le système 3.16 donne :

$$\begin{cases} v_3 = 5 \\ v_5 = 10 - v_4 \\ v_6 = 10 + v_4 \\ v_{11} = -5 + 0.5v_4 + v_7 \\ v_8 = 10 \end{cases} \quad (3.19)$$

En utilisant la propriété 3.4.1 ci dessus, on a :

$$\begin{cases} v_3 = f_1(v_4, v_7) = 5 \\ v_4 = f_2(v_4, v_7) = v_4 \\ v_5 = f_3(v_4, v_7) = 10 - v_4 \\ v_6 = f_4(v_4, v_7) = 10 + v_4 \\ v_7 = f_5(v_4, v_7) = v_7 \\ v_8 = f_6(v_4, v_7) = 10 \\ v_{11} = f_7(v_4, v_7) = -5 + 0.5v_4 + v_7 \end{cases}$$

$$\bar{D}_1[v_{libre}] = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & \frac{5-0.5v_4-v_7}{5+0.5v_4+v_7} & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ 0 & -1 & \frac{-2.5}{5+0.5v_4+v_7} & 1 & 0 \\ 0 & 0 & \frac{-7.5}{5+0.5v_4+v_7} & 0 & 1 \end{pmatrix} \quad (3.20)$$

$$\bar{D}_2[v_{libre}] = \begin{pmatrix} -1 & \frac{-v_4}{10} \\ 0 & \frac{-10+v_4}{10} \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \quad (3.21)$$

Analyse de la matrice des AIOs

La matrice $D_1[v]$ est une matrice carrée d'ordre $n + q$ dont :

- tous les éléments diagonaux sont égaux à 1 ;
- chaque élément diagonal correspond à un métabolite intermédiaire (pivot) ou un métabolite de sortie (sortie) ;
- pour chaque élément diagonal correspondant à un pivot les éléments non diagonaux correspondent aux pourcentages d'utilisation du pivot pour produire d'autres métabolites. Avec l'hypothèse de non accumulation des pivots la somme des pourcentages d'utilisation est égale à 1 ;
- pour chaque élément diagonal correspondant à une sortie, les éléments non diagonaux sont tous égaux à 0 car par définition une sortie est un métabolite qui ne produit aucun autre métabolite.

Ainsi, la matrice $D_1[v]$ se redéfinit comme suit :

$$D_1 = \begin{pmatrix} D_{\mathcal{I}}[v] & 0 \\ D_{\mathcal{O}}[v] & I_{\mathcal{O}} \end{pmatrix} \quad (3.22)$$

avec :

$$\sum_{k=1, k \neq i}^{n+q} d_{ki} = 1, \quad 1 \leq i \leq n.$$

$D_{\mathcal{I}}[v]$ est une matrice carrée d'ordre n , 0 est la matrice nulle d'ordre $n \times q$, $I_{\mathcal{O}}$ est la matrice identité d'ordre q et $D_{\mathcal{O}}[v]$ est une matrice d'ordre $q \times n$.

3.5 Propriétés

Les coefficients du vecteur $x_{\mathcal{O}}[v, m]$, que nous voulons calculer, forment une solution d'un système linéaire lorsque le flux de distribution est supposé fixé préalablement.

Dans cette section, nous allons montrer que sous certaines hypothèses ce système admet une solution unique.

Définition 3.5.1. Une matrice $A \in \mathcal{M}_n(\mathbb{R})$ est appelée une *Z-matrice* si tous ses coefficients non diagonaux sont négatifs. Une *Z-matrice* dont toutes les valeurs propres sont de partie réelle strictement positive est appelée une *M-matrice*.

Définition 3.5.2. Soit A une matrice carrée d'ordre n à coefficients complexes, le rayon spectral de A est le réel positif

$$\rho(A) = \max_{\lambda \in Sp(A)} |\lambda|,$$

où $Sp(A)$ désigne l'ensemble des valeurs propres complexes de A .

Lemme 3.5.1. [Ada09] Une matrice M est une *M-matrice* si et seulement si elle peut s'écrire sous la forme $A = sI_n - B$ où B est une matrice positive et $s > \rho(B)$, $\rho(B)$ désigne le rayon spectral de la matrice B .

La famille des *Z-matrices* a les propriétés particulières suivantes.

Théorème 3.5.1. [Ada09] Soit $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{R})$ une *Z-matrice*. Alors les propriétés suivantes sont équivalentes.

1. A^{-1} est positive, c'est-à-dire que tous ses coefficients sont positifs.
2. A est une *M-matrice*.
3. Les mineurs principaux dominants de A sont strictement positifs.

Le rayon spectral dans le cas des matrices réelles possède les propriétés suivantes.

Propriétés 3.5.1. Soit $\mathcal{M}_n(\mathbb{R})$ l'ensemble des matrices carrées d'ordre n . Pour toute matrice carrée positive $A \in \mathcal{M}_n(\mathbb{R})$ on a :

$$\min_{1 \leq i \leq n} \sum_{j=1}^n a_{ij} \leq \rho(A) \leq \max_{1 \leq i \leq n} \sum_{j=1}^n a_{ij}$$

$$\min_{1 \leq j \leq n} \sum_{i=1}^n a_{ij} \leq \rho(A) \leq \max_{1 \leq j \leq n} \sum_{i=1}^n a_{ij}.$$

Or, la matrice $D_{\mathcal{I}}[v]$ de dimension n de l'équation 3.22, vérifie la propriété suivante :

Proposition 3.5.1. La matrice $D_{\mathcal{I}}[v]$ définie dans l'équation 3.22 ci-dessus s'écrit sous la forme $D_{\mathcal{I}}[v] = I_n - B[v]$, avec B une matrice positive.

Démonstration. Découle de l'équation 3.12 du théorème 3.4.1. □

Théorème 3.5.2. *Si pour $1 \leq i \leq n$*

$$\sum_{k=1, k \neq i}^n |d_{ki}| \neq 1, \quad (3.23)$$

alors $D_1[v]$ de dimension $n + q$ est inversible.

Démonstration. Découle du fait que le déterminant de $D_1[v]$ est égal au déterminant de $D_{\mathcal{I}}[v]$ en valeur absolue d'après l'équation 3.22 . Donc l'inversibilité de $D_1[v]$ dépend de l'inversibilité de $D_{\mathcal{I}}[v]$. D'après la proposition 3.5.1, on a : $D_{\mathcal{I}}[v] = I_n - B[v]$. Si on a la condition 3.23 du théorème 3.5.2, d'après la propriété 3.5.1, $\rho(B[v]) < 1$. En utilisant le lemme 3.5.1 on déduit que $D_{\mathcal{I}}[v]$ est inversible et par la suite $D_1[v]$ est aussi inversible. \square

Corollaire 3.5.1. *Si chaque pivot d'un réseau métabolique produit au moins une sortie alors $D_1[v]$ est inversible et le système d'équations admet une unique solution.*

Démonstration. Chaque colonne de la matrice $D_{\mathcal{O}}$ correspond aux pourcentages d'utilisation des pivots pour produire des sorties. Le calcul des AIOs est basé sur le modèle stoechiométrique [LAAS⁺10] dans lequel un pivot produit au moins une sortie c'est-à-dire que pour chaque colonne de $D_{\mathcal{O}}[v]$ au moins un élément est non nul. Donc

$$\sum_{k=1, k \neq i}^{n+q} |d_{ki}| < 1, \quad 1 \leq i \leq n$$

On déduit de ce qui précède que la matrice des AIOs est inversible. \square

Ainsi, dans le cas de cycles, nous venons de montrer que les coefficients des AIOs sont déterminées de manière uniques par l'inversion d'une matrice.

3.6 Algorithme général de calcul des AIOs

La modélisation des AIOs sous la forme matricielle est un moyen d'étendre le calcul des AIOs à des grands réseaux métaboliques avec cycles. Ainsi, si pour une distribution de flux v donnée, la matrice $D_1[v]$ est une matrice inversible alors les AIOs s'obtiennent en inversant la matrice $D_1[v]$. L'algorithme 2 ci-dessous prenant en entrées : une distribution de flux v , le vecteur des flux des entrées par rapport à l'élément de répartition φ , $x_{\mathcal{I}}$, les matrices $D_1[v]$ et $D_2[v]$ permet d'automatiser très facilement le calcul des sorties $x_{\mathcal{O}}[v, m]$, $m \in \mathcal{I}$.

Le calcul complet des AIOs se fait donc en $O((n + q)^3 + npq)$ opérations pour une distribution de flux donnée. En effet, l'inversion de la matrice D_1^{-1} nécessite $O((n + q)^3)$ opérations et $O(npq)$ opérations sont également nécessaires pour calculer les contributions des entrées dans les sorties.

Algorithme 2 Algorithme de calcul des AIOs dans le cas général

Entrées : v (une distribution de flux), $D_1[v]$, $D_2[v]$, $x_{\mathcal{I}} = \{x^{I_1}, x^{I_2}, \dots, x^{I_p}\}$
Sorties : $x_{\mathcal{O}}[v, m] = \{x_{O_i}[v, m], i \in \{1, 2, \dots, q\} m \in \mathcal{I}\}$

 On calcule l'inverse D_1^{-1} pour v donné

Pour i allant de 1 à p **Faire**
 $x_s[v, I_i] = 0, s \in \mathcal{O}$ (Initialisation)

Pour k allant de $n + 1$ à $n + q$ **Faire**
Pour j allant de 1 à $n + q$ **Faire**
 $x_{O_{k-n}}[v, I_i] + D_1^{-1}(k, j) * (-D_2(j, i)) * x^{I_i} \leftarrow x_{O_{k-n}}[v, I_i]$
Fin Pour
Fin Pour
Fin Pour

3.7 Application : calcul des AIOs de sommets extrémaux

Nous avons vu, dans le chapitre 2, que la plupart des distributions extrémales susceptibles d'optimiser un objectif linéaire sont sujettes à caution car les valeurs de certains flux sont en dehors des ordres de grandeurs des valeurs issues de la littérature. Cependant sur certaines distributions extrémales les données de la littérature sur les flux des réactions ne permettent pas de finaliser la discussion sur leur pertinence biologique.

Ainsi nous avons calculé les AIOs des flux de carbones des nutriments prélevés (Entrées) dans les constituants du lait (Sorties) pour deux régimes alimentaires pour discuter de la pertinence de quatre distributions extrémales qui vérifient toutes les conditions disponibles par rapport aux flux des réactions dans le chapitre 2. Il s'agit de :

- deux distributions extrémales nommées G et H pour le régime alimentaire (Ctrl);
- deux distributions extrémales nommées également G et H pour le régime alimentaire (CN).

Pour chacun des régimes alimentaires (Ctrl) et (CN), G et H sont les distributions qui maximisent respectivement les fonctions objectifs $(v_{64} + v_{19})$ et $(v_{64} - v_{19})$.

Les Tables 3.1, 3.2, 3.3, 3.4 donnent des AIOs calculés pour les sommets G et H, présentés dans le chapitre 2, pour les régimes alimentaires (CN) et (Ctrl).

Les entrées (Glucose, Glycérol, Acétate, BHBA, Lactate) correspondent aux nutriments prélevés par la mamelle. Les acides aminés (Isoleucine, Leucine, Lysine, Thréonine, Valine, Histidine, Arginine, Alanine et Glutamine) considérés comme entrées sont des acides aminés satisfaisant la propriété : la quantité prélevée par la mamelle moins la quantité utilisée pour produire les protéines du lait est positive.

Les sorties (Glycérol3P, Lactose, C4, C6, C8, C10, C12, C14, C16) sont des données mesurées dans le lait. Les acides aminés (Glycine, Glutamate, Proline, Aspartate, Asparagine et SerineOutput) sont des acides aminés tels que la quantité prélevée par la mamelle moins la quantité utilisée pour produire les protéines du lait est négative.

Les différentes tables montrent que, pour les AIOs des distributions extrémales des régimes (CN) et Ctrl, le Glucose est unique précurseur du Lactose. Ceci est en

Entrées	Glucose	Glycerol	Acetate	BHBA	Lysine	Threonine	Isoleucine	Leucine	Valine	Histidine	Arginine	Alanine	Glutamine
Sorties	Pourcentage des flux de carbone des sorties venant de chaque entrée												
Glycerol3P	87.4	12.6	0	0	0	0	0	0	0	0	0	0	0
Lactose	100	0	0	0	0	0	0	0	0	0	0	0	0
C4	0	0	0	100	0	0	0	0	0	0	0	0	0
C6	0	0	66.7	33.3	0	0	0	0	0	0	0	0	0
C8	0	0	75	25	0	0	0	0	0	0	0	0	0
C10	0	0	80	20	0	0	0	0	0	0	0	0	0
C12	0	0	83.3	16.7	0	0	0	0	0	0	0	0	0
C14	0	0	85.7	14.3	0	0	0	0	0	0	0	0	0
C16	0	0	87.5	12.5	0	0	0	0	0	0	0	0	0
Glycine	73	9.2	9.7	4	0.3	1.4	0.3	0.2	0.3	0.8	0.4	0.2	0.1
Glutamate	1.3	0.2	61.5	25.5	1.4	0.1	1.7	1.6	1.3	0.2	3.4	1	0.9
Proline	1.3	0.2	61.5	25.5	1.4	0.1	1.7	1.6	1.3	0.2	3.4	1	0.9
Aspartate	1.8	0.2	60.1	25	1.3	0.1	2.2	1.5	2	0.2	3.3	1.4	0.9
Peptide	0	0	0	0	0	0	0	0	0	0	0	0	0
SerineOutput	84.8	12.2	0	0	0	1.8	0	0	0	1.1	0	0	0
CO2Output	37.6	0.1	38.7	16.1	1.3	0.1	1.1	1	1	0.1	1.7	0.8	0.5

TABLE 3.1 – Calcul des AIOs pour le sommet extrémal G de (Ctrl). G sommet qui maximise la fonction objectif ($v_{64} + v_{19}$). Pour chaque colonne (entrée), les valeurs, correspondent aux pourcentages des sorties provenant de cette entrée. La première colonne montre que 100% du lactose provient du glucose (ce qui est contraire à la littérature [BAL74]).

Entrées	Glucose	Glycerol	Acetate	BHBA	Lysine	Threonine	Isoleucine	Leucine	Valine	Histidine	Arginine	Alanine	Glutamine
Sorties	Pourcentage des flux de carbone des sorties venant de chaque entrée												
Glycerol3P	95.3	4.7	0	0	0	0	0	0	0	0	0	0	0
Lactose	100	0	0	0	0	0	0	0	0	0	0	0	0
C4	0	0	0	100	0	0	0	0	0	0	0	0	0
C6	0	0	66.7	33.3	0	0	0	0	0	0	0	0	0
C8	0	0	75	25	0	0	0	0	0	0	0	0	0
C10	0	0	80	20	0	0	0	0	0	0	0	0	0
C12	0	0	83.3	16.7	0	0	0	0	0	0	0	0	0
C14	0	0	85.7	14.3	0	0	0	0	0	0	0	0	0
C16	0	0	87.5	12.5	0	0	0	0	0	0	0	0	0
Glycine	78.3	3.7	9.8	4.1	0.3	1.4	0.3	0.2	0.3	0.8	0.4	0.2	0.1
Glutamate	17.2	0.8	51.2	21.3	1.1	0.1	1.4	1.3	1.1	0.1	2.8	0.7	0.8
Proline	17.2	0.8	51.2	21.3	1.1	0.1	1.4	1.3	1.1	0.1	2.8	0.7	0.8
Aspartate	17.6	0.9	50.3	20.9	1.1	0.1	1.9	1.3	1.6	0.1	2.8	0.7	0.8
Peptide	0	0	0	0	0	0	0	0	0	0	0	0	0
SerineOutput	92.5	4.6	0	0	0	1.8	0	0	0	1.1	0	0	0
CO2Output	35.7	1	39.3	16.3	1.4	0.1	1.1	1	1.1	0.1	1.8	0.8	0.5

TABLE 3.2 – Calcul des AIOs pour le sommet extrémal H de (Ctrl). H sommet qui maximise la fonction objectif ($v_{64} - v_{19}$). Pour chaque colonne (entrée), les valeurs, correspondent aux pourcentages des sorties provenant de cette entrée. La première colonne montre que 100% du lactose provient du glucose (ce qui est contraire à la littérature [BAL74]).

contradiction avec des études déjà réalisées sur le métabolisme de la vache laitière qui suggèrent que le Glycérol et peut être même les acides aminés contribuent à la synthèse du Lactose [BSEKO06]. Plus précisément environ 15% des carbones du Lactose pourrait ne pas provenir du Glucose [BAL74].

En conclusion, cette étude des AIOs suggère que les distributions extrémales G et H ne sont pas pertinentes biologiquement. D'autres part dans le chapitre 2, on a également vu que, toutes les autres distributions extrémales sont non pertinentes car des valeurs de certains flux sont plus grandes ou plus petites par rapport aux valeurs de la littérature.

Par la suite nous pouvons dire qu'avec les informations dont nous dis-

Entrées	Glucose	Glycerol	Acetate	BHBA	Lysine	Isoleucine	Leucine	Valine	Arginine	Glutamine
Sorties	Pourcentage des flux de carbone des sorties venant de chaque entrée									
Glycerol3P	90.3	9.7	0	0	0	0	0	0	0	0
Lactose	100	0	0	0	0	0	0	0	0	0
C4	0	0	0	100	0	0	0	0	0	0
C6	0	0	66.7	33.3	0	0	0	0	0	0
C8	0	0	75	25	0	0	0	0	0	0
C10	0	0	80	20	0	0	0	0	0	0
C12	0	0	83.3	16.7	0	0	0	0	0	0
C14	0	0	85.7	14.3	0	0	0	0	0	0
C16	0	0	87.5	12.5	0	0	0	0	0	0
Glycine	73.8	7.3	5.6	10.8	0.5	0.5	0.5	0.4	0.5	0.2
Alanine	90.3	9.7	0	0	0	0	0	0	0	0
Glutamate	2.5	0.2	28.7	56.1	1.6	2.4	2.5	1.7	3	1.2
Proline	2.5	0.2	28.7	56.1	1.6	2.4	2.5	1.7	3	1.2
Aspartate	3.8	0.4	27.8	54.2	1.6	3.2	2.4	2.6	2.9	1.2
Peptide	0	0	0	0	0	0	0	0	0	0
SerineOutput	90.3	9.7	0	0	0	0	0	0	0	0
CO2Output	24.2	0.2	22.2	43.4	1.9	1.8	2	1.7	1.9	0.8

TABLE 3.3 – Calcul des AIOs pour le sommet extrémal G de (CN). G sommet qui maximise la fonction objectif ($v_{64} + v_{19}$). Pour chaque colonne (entrée), les valeurs, correspondent aux pourcentages des sorties provenant de cette entrée. La première colonne montre que 100% du lactose provient du glucose (ce qui est contraire à la littérature [BAL74]).

Entrées	Glucose	Glycerol	Acetate	BHBA	Lysine	Isoleucine	Leucine	Valine	Arginine	Glutamine
Sorties	Pourcentage des flux de carbone des sorties venant de chaque entrée									
Glycerol3P	90.7	9.3	0	0	0	0	0	0	0	0
Lactose	100	0	0	0	0	0	0	0	0	0
C4	0	0	0	100	0	0	0	0	0	0
C6	0	0	66.7	33.3	0	0	0	0	0	0
C8	0	0	75	25	0	0	0	0	0	0
C10	0	0	80	20	0	0	0	0	0	0
C12	0	0	83.3	16.7	0	0	0	0	0	0
C14	0	0	85.7	14.3	0	0	0	0	0	0
C16	0	0	87.5	12.5	0	0	0	0	0	0
Glycine	74.1	7	5.6	10.8	0.5	0.5	0.5	0.4	0.5	0.2
Alanine	90.7	9.3	0	0	0	0	0	0	0	0
Glutamate	3	0.3	28.6	55.8	1.6	2.4	2.5	1.7	3	1.2
Proline	3	0.3	28.6	55.8	1.6	2.4	2.5	1.7	3	1.2
Aspartate	4.3	0.4	27.6	53.9	1.5	3.2	2.4	2.6	2.9	1.2
Peptide	0	0	0	0	0	0	0	0	0	0
SerineOutput	90.7	9.3	0	0	0	0	0	0	0	0
CO2Output	24.1	0.2	22.2	43.4	1.9	1.8	2	1.7	1.9	0.8

TABLE 3.4 – Calcul des AIOs pour le sommet extrémal H de (CN). H sommet qui maximise la fonction objectif ($v_{64} - v_{19}$). Pour chaque colonne (entrée), les valeurs, correspondent aux pourcentages des sorties provenant de cette entrée. La première colonne montre que 100% du lactose provient du glucose (ce qui est contraire à la littérature [BAL74]).

posons sur le fonctionnement de la mamelle, on ne peut pas identifier une fonction linéaire est optimale sur un sommet de l'espace de distribution des flux et qui est biologiquement compatible avec les informations dont nous disposons.

Pour réaliser cette analyse, il a fallu modéliser en détail les contributions des entrées dans les sorties et montrer qu'elles peuvent se déterminer de manière unique sous certaines conditions peu restrictives sur le réseau. La modélisation des AIOs avec une approche générique constitue le contenu principal de la

partie méthode de l'article accepté pour publication à *BMC System Biology* [AALVM⁺ on].

Cependant, il existe plusieurs distributions de flux qui satisfont toutes les hypothèses de la littérature. Par exemple pour le traitement (Ctrl) il existe une distribution tels que : $v_8 < 591$ et $v_{13} < 266$ avec $v_{64}/v_{63} = 0.62 < 0.67$ et que le glucose est précurseur du lactose à 90.79%. Pour comprendre le fonctionnement global de la glande mammaire et comparer des régimes alimentaires différents, nous avons introduit dans le chapitre suivant une méthode permettant d'étudier la variabilité des AIOs dans l'espace de distribution des flux. Dans l'esprit, il s'agit, comme pour les techniques de *flux variability analysis*, de faire abstraction d'une hypothèse d'optimisation pour le fonctionnement du système. L'étude de cette variabilité nécessite de résoudre des problèmes d'optimisation non linéaires. Ainsi, nous avons proposé un algorithme efficace permettant de calculer les valeurs minimales et maximales des AIOs et avoir une vision complète des comportements possibles de la glande mammaire pour un régime alimentaire donné.

Chapitre 4

Algorithme de recherche locale pour explorer la flexibilité dans le cas des organismes complexes

Les résultats auxquels nous sommes arrivés dans les chapitres 2 et 3 ont montré qu'aucune fonction objectif linéaire ne peut être identifiée de manière unique pour modéliser la réponse de la glande mammaire avec notre modèle métabolique. En d'autres termes les sommets de l'espace de recherche correspondant aux distributions extrémales, c'est-à-dire susceptibles d'optimiser un objectif linéaire, ne déterminent pas les distributions adaptées par la mamelle pour transformer les nutriments. Ainsi, on se retrouve exactement dans le contexte où les approches de FVA (*Flux Variability Analysis*) ont été développées. Cependant, comme vu dans le chapitre précédent, les résultats du FVA ne sont pas assez précis pour prendre en compte les répartitions des nutriments. Dans ce chapitre, nous avons développé un algorithme de recherche locale pour étudier la variabilité des coefficients des AIO en extension du FVA. Cette approche permet de calculer les contributions maximales et minimales des nutriments prélevés par la mamelle dans les différents constituants du lait. Cette approche ne favorise aucune fonction objectif et permet d'avoir une vision globale de l'effet d'un régime alimentaire sur les constituants du lait.

Etudier la variabilité des AIOs se ramène à optimiser différentes fonctions obtenues à partir de la matrice $\overline{D}_1^{-1}[v_{libre}, m]$, introduite dans le chapitre 3. Ici *v_{libre}* devient la variable et les problèmes d'optimisation deviennent non linéaires. La question principale sera alors d'obtenir une forme analytique exploitable pour la fonction à optimiser. Ainsi lorsque la formule analytique est disponible, nous utilisons un algorithme de programmation non linéaire appelé *méthode de points intérieurs* (une bibliothèque de la fonction *fmincon* de matlab) pour calculer les optima des fonctions à optimiser. Une particularité des méthodes de points intérieurs est qu'elles n'explorent pas exactement le bord de l'espace de recherche pour la recherche de solutions optimales. Nous avons proposé un algorithme de recherche locale permettant d'explorer de manière plus fine le bord de l'espace de recherche.

D'autre part la forme analytique des AIOs, nécessaire à l'utilisation de *fmincon*, n'est pas toujours disponible. L'algorithme de recherche locale que nous proposons est une approche qui permet de contourner cette difficulté. En effet, il permet de calculer les optima des AIOs sans avoir besoin d'une formule analytique pour les fonctions à optimiser.

4.1 Problème posé

La détermination des contributions maximales et minimales des nutriments prélevés par la mamelle dans les différents constituants du lait revient à déterminer les minima et les maxima d'une matrice de fonctions non linéaires et non convexes sur un espace convexe. Le problème est de la forme : pour $n + 1 \leq i \leq n + q$ et $1 \leq k \leq p$,

$$(\text{Max/Min})\text{imize } \left\{ AIO[v][i, k]; v \in \mathbb{R}^n, M \begin{pmatrix} v_I \\ v \\ v_O \end{pmatrix} = 0, Bv = 0, v \geq 0 \right\}. \quad (4.1)$$

avec $AIO[v] = x_{\mathcal{O}}[v, m]$.

$$\begin{pmatrix} x_{\mathcal{P}}[v, m] \\ x_{\mathcal{O}}[v, m] \end{pmatrix} = -D_1[v]^{-1} D_2[v] \begin{pmatrix} c(m_1)v_I(m_1) & & 0 \\ & \ddots & \\ 0 & & c(m_p)v_I(m_p) \end{pmatrix}.$$

En utilisant la Propriété 3.4.1 du chapitre 3, c'est-à-dire en exprimant le vecteur v en fonction des variables libres du système sous la forme $v = f(v_{\text{libre}}) = (f_1(v_{\text{libre}}), f_2(v_{\text{libre}}), \dots, f_k(v_{\text{libre}}))$, le problème devient :

$$(\text{Max/Min})\text{imize } \{ \overline{AIO}[v_{\text{libre}}][i, k]; v_{\text{libre}} \in \mathbb{R}^s \}. \quad (4.2)$$

$\overline{AIO}[v_{\text{libre}}] = \overline{x}_{\mathcal{O}}[v_{\text{libre}}, m]$.

$$\begin{pmatrix} \overline{x}_{\mathcal{P}}[v_{\text{libre}}, m] \\ \overline{x}_{\mathcal{O}}[v_{\text{libre}}, m] \end{pmatrix} = -\overline{D}_1[v_{\text{libre}}]^{-1} \overline{D}_2[v_{\text{libre}}] \begin{pmatrix} c(m_1)v_I(m_1) & & 0 \\ & \ddots & \\ 0 & & c(m_p)v_I(m_p) \end{pmatrix}.$$

Ici le vecteur v_{libre} n'est plus supposé connu, c'est-à-dire que les coefficients de v_{libre} sont des variables de l'espace des solutions. Ainsi les $\overline{AIO}[v_{\text{libre}}]$ sont des fonctions non linéaires des coefficients de v_{libre} . Cependant un problème supplémentaire est d'obtenir de manière formelle la forme analytique de $\overline{D}_1[v_{\text{libre}}]^{-1}$. Comme nous allons le voir dans la suite de ce chapitre une des limites des méthodes existantes pour la résolution du problème 4.2 est qu'elles reposent sur la disponibilité de cette forme analytique. En effet, le calcul de cette forme analytique avec des approches de calcul formel a une complexité en espace mémoire exponentiel. Pour contourner cette difficulté nous avons proposé une nouvelle approche à la section 4.3 permettant de calculer les min-max des AIOs indépendamment de cette disponibilité.

Exemple 4.1.1. Dans l'exemple du réseau avec cycle présenté au chapitre 3, nous avons utilisé le logiciel de calcul formel *sagemath* [SO12] pour calculer $\bar{D}_1[v_{libre}]^{-1}$. Ainsi, $\overline{ATO}[v_{libre}] = \bar{x}_{\mathcal{O}}[v_{libre}, m]$ est donné par l'expression suivante :

$$\begin{pmatrix} \bar{x}_{\mathcal{P}}[v_{libre}, m] \\ \bar{x}_{\mathcal{O}}[v_{libre}, m] \end{pmatrix} = -\bar{D}_1[v_{libre}]^{-1} \begin{pmatrix} -1 & \frac{-v_4}{10} \\ 0 & \frac{-10+v_4}{10} \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 20 & 0 \\ 0 & 20 \end{pmatrix}$$

On obtient $\bar{D}_1[v_{libre}]^{-1}$ exprimé ci-dessous en utilisant le logiciel de calcul formel *sagemath*.

$$\bar{D}_1[v_{libre}]^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ a/b & 1 & a/b & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ a/b + 2.5/b & 1 & a/b + 2.5/b & 1 & 0 \\ 7.5/b & 0 & 7.5/b & 0 & 1 \end{pmatrix}$$

$$a = 0.5v_4 + v_7 - 5 \quad b = 0.5v_4 + v_7 + 5$$

$$\bar{D}_1[v_{libre}] = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & \frac{5-0.5v_4-v_7}{5+0.5v_4+v_7} & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ 0 & -1 & \frac{-2.5}{5+0.5v_4+v_7} & 1 & 0 \\ 0 & 0 & \frac{-7.5}{5+0.5v_4+v_7} & 0 & 1 \end{pmatrix}$$

Les matrices $\bar{D}_1[v_{libre}]$ et $\bar{D}_2[v_{libre}]$ permettant de déterminer les contributions maximales et minimales des nutriments prélevés par la mamelle dans les différents constituants du lait pour les régimes alimentaires (Ctrl), (CN) et (HB) sont exprimées en fonctions de 4 variables libres. C'est-à-dire à l'état stationnaire, en renseignant les valeurs entrées/sorties ($v_{\mathcal{I}}$) et ($v_{\mathcal{O}}$) pour chacun des régimes alimentaires l'espace de distribution des flux correspondant au système :

$$M \begin{pmatrix} v_{\mathcal{I}} \\ v \\ v_{\mathcal{O}} \end{pmatrix} = 0, Bv = 0, v \geq 0$$

a 4 flux libres et indépendants (4 degrés de liberté).

4.2 Résolution par une méthode de points intérieurs

Dans cette section, nous présentons les fondements des *méthodes de points intérieurs*. Nous présentons une comparaison de résultats de calcul de la méthode de points intérieurs

de la bibliothèque *fmincon* de matlab et d'autres méthodes itératives (active-set et sqp) pour le problème 4.1. Enfin, nous discutons des limites de ces méthodes de résolution pour ce même problème 4.1.

4.2.1 Points intérieurs

L'optimisation de chaque coefficient du problème 4.1 revient à résoudre un problème de la forme :

$$(P_{EI}) \begin{cases} \text{Minimiser } f(x) \\ c_i(x) = 0, \quad i \in E = \{1, 2, \dots, m_E\} \\ c_i(x) \leq 0, \quad i \in I = \{m_E + 1, m_E + 2, \dots, m_E + m_I\} \\ x \in \mathbb{R}^n \end{cases} \quad (4.3)$$

On note

$$c_E = (c_1, c_2, \dots, c_{m_E})^\top, \quad c_I = (c_{m_E+1}, c_{m_E+2}, \dots, c_{m_E+m_I})^\top, \quad c = (c_E, c_I)^\top.$$

$$I^o(x) = \{i \in I : c_i(x) = 0\}.$$

$I^o(x)$ est appelé ensemble des indices des contraintes d'inégalité *actives* en x . De ce faite une contrainte d'égalité est par définition *active*.

$C = \{x \in \mathbb{R}^n : c_E(x) = 0, c_I(x) = 0\}$ est appelé *ensemble des solutions réalisables*. Où, les fonctions f , et c_i , $i \in E \cup I$, sont des fonctions de \mathbb{R}^n vers \mathbb{R} .

On suppose dans toute la suite que ces fonctions sont indéfiniment différentiables sur \mathbb{R}^n .

Qualification des contraintes

- Les contraintes sont dites *qualifiées en* $x^o \in c$ si les gradients des contraintes actives en x^o , $\{\nabla c_i(x^o), i \in E \cup I \text{ et } c_i(x) = 0\}$ sont linéairement indépendantes.
- Si c est un polyèdre convexe (i.e. : toutes les contraintes sont des fonctions affines) alors les contraintes sont qualifiées en tout point de l'ensemble admissible.

D'autres hypothèses de qualification des contraintes peuvent être consultées dans [BGLS01].

Soit la fonction

$$\begin{aligned} \ell : \mathbb{R}^n \times \mathbb{R}^{m_E} \times \mathbb{R}^{m_I} &\longrightarrow \mathbb{R} \\ (x, \lambda_E, \lambda_I) &\longrightarrow \ell(x, \lambda_E, \lambda_I) = f(x) + \lambda_E^\top c_E(x) + \lambda_I^\top c_I(x) \end{aligned}$$

ℓ est appelé Lagrangien associé au problème (P_{EI}) . $\lambda_E^\top = (\lambda_1, \lambda_2, \dots, \lambda_{m_E})$ et $\lambda_I^\top = (\lambda_{m_E+1}, \lambda_{m_E+2}, \dots, \lambda_{m_E+m_I})$ sont appelés les multiplicateurs de Lagrange associé à x .

Théorème 4.2.1. [BGLS01] (*conditions nécessaires de Karush-Kuhn-Tucher (KKT)*)

Soit x^* un minimum local de (P_{EI}) . On suppose que les contraintes sont qualifiées en x^* . Alors il existe $(\lambda_E^*, \lambda_I^*) \in \mathbb{R}^{m_E} \times \mathbb{R}^{m_I}$ tels que

$$(KKT) \begin{cases} \nabla_x \ell(x^*, \lambda_E^*, \lambda_I^*) = \nabla f(x^*) + A_E(x^*)^\top \lambda_E^* + A_I(x^*)^\top \lambda_I^* = 0 \\ c_E(x^*) = 0 \\ c_I(x^*) \leq 0 \\ (\lambda_I^*)^\top c_I(x^*) = 0 \\ \lambda_I^* \geq 0. \end{cases} \quad (4.4)$$

Où, $\nabla f(x^*)$ désigne le gradient (vecteur des dérivées partielles) de la fonction f au point x^* et $A_E(x^*)$ désigne la Jacobienne (vecteur des dérivées partielles) de c_E en x^* .

Le triplet $(x^*, \lambda_E^*, \lambda_I^*)$ satisfaisant les conditions de (KKT) est appelé solution *primal-dual* de (P_{EI}) et x^* est appelé point *stationnaire*.

Remarque 4.2.1. Si f , c_E et c_I sont des fonctions convexes, les conditions de (KKT) sont à la fois nécessaires et suffisantes pour que x^* soit un minimum global [KT51].

Le théorème 4.2.1 a par la suite été utilisé pour la recherche de solutions numériques des fonctions non linéaires convexes.

Cependant, l'une des difficultés de l'utilisation du théorème pour la résolution numérique repose sur la satisfaction de la condition, $(\lambda_I^*)^\top c_I(x^*) = 0$, du théorème. Cette condition est difficile à résoudre à cause de son caractère combinatoire. En effet, il y a 2^{m_I} façons de les réaliser en choisissant pour tout $i \in I$ si c'est λ_i^* ou $c_i(x^*)$ qui est nul. Afin d'éviter ce problème, les méthodes de points intérieurs cherchent à perturber chaque condition $\lambda_i^* c_i(x^*) = 0$ en $\lambda_i^* c_i(x^*) = \mu$ (μ est un paramètre) et en imposant la stricte négativité à $c_I(x^*)$ [Fri54, FM68]. On obtient ainsi les conditions d'optimalité perturbées suivantes :

$$\begin{cases} \nabla_x \ell(x^*, \lambda_E^*, \lambda_I^*) = \nabla f(x^*) + A_E(x^*)^\top \lambda_E^* + A_I(x^*)^\top \lambda_I^* = 0 \\ c_E(x^*) = 0 \\ c_I(x^*) < 0 \\ \lambda_I^* \cdot c_I(x^*) - \mu e = 0 \\ \lambda_I^* > 0. \end{cases} \quad (4.5)$$

Où, e un vecteur de dimension $m_I \times 1$ dont toutes les composantes sont égales à 1.

En substituant la valeur des multiplateurs λ_I^* dans la première condition à l'aide de la quatrième condition dans le système (4.5), on obtient le système suivant :

$$\begin{cases} \nabla f(x^*) + A_E(x^*)^\top \lambda_E^* + \mu A_I(x^*)^\top c_I^{-1}(x^*) = 0 \\ c_E(x^*) = 0 \\ c_I(x^*) < 0. \end{cases} \quad (4.6)$$

Ces conditions d'optimalité perturbées peuvent se voir comme les conditions d'optimalité du problème suivant :

$$(P_{EI}^\mu) \begin{cases} \text{Minimiser}(\Phi_\mu(x) = f(x) - \mu \sum_{i=m_E+1}^{m_E+m_I} \ln(-c_i(x))) \\ c_E(x) = 0 \\ x \in \mathbb{R}^n, \mu < 0. \end{cases} \quad (4.7)$$

La fonction $\Phi_\mu(x)$ est appelée *fonction barrière logarithmique*.

Les méthodes de points intérieurs ont d'abord été appliquées au cas linéaire avant d'être généralisées de manière itérative aux problèmes quadratiques [BW00, Kad01, VS99] puis aux cas non linéaires.

4.2.2 Implémentation et application

Une quantité impressionnante de méthodes de points intérieurs ont été proposées depuis la fin du 20^e siècle mais juste une partie a été réellement implémentée parmi lesquelles *LOQO* de Vanderbrei et Shanno [VS99], *KNITRO* de Nocedal et Waltz [BGN00, BHN99], *MINOS* de Murtagh et Saunders [MS95], *SNOPT* de Gill, Muray et Saunders [GMS02], *filterSQP* de Fletcher et Leyffer [FL97], et *Interior-Point* de la fonction *fmincon* de Matlab [R2011].

Le solveur *fmincon* contient également plusieurs autres méthodes itératives de résolution. Seulement trois de ces méthodes peuvent être utilisées pour la résolution du problème posé à la section 4.1 à savoir : les méthodes de *points intérieurs*, *sqp* (*sequentiel quadratic programming*) et *active-set*.

Nous avons testé ces trois méthodes pour déterminer les quantités minimales de flux de carbones des nutriments prélevés qui sont répartis dans les constituants du lait pour le jeu de données (Ctrl) présenté au chapitre 2. Comme mentionné dans la section 4.1, l'expression formelle des AIOs est une matrice de fonctions non linéaires avec 4 flux libres et indépendants (4 degrés de liberté).

Les résultats sont donnés dans les Tables 4.1, 4.2, 4.3.

Les trois méthodes donnent des résultats assez similaires, mais la méthode de *points intérieurs* est celle qui donne des résultats légèrement meilleurs.

4.2.3 Limites du processus calcul formel et méthodes de résolution existantes

Dans cette sous section, nécessitant une expression analytique obtenue par des méthodes de calcul formel, nous allons montrer que les méthodes d'optimisation basées sur du calcul formel de l'inverse d'une matrice ont une complexité exponentielle. Ensuite, nous allons montrer que les méthodes de points intérieurs n'explorent pas exactement le bord de l'espace de recherche. Ces deux faits tracent les limites des méthodes de résolution précédentes.

Entrées	Glucose 1422	Glycerol 17,5	Acetate 1020	BHBA 336	Lysine 16,1	Threonine 1,40	Isoleucine 13,1	Leucine 12,1	Valine 12,7	Histidine 1,38	Arginine 26,4	Alanine 9,33	Glutamine 6,10	
Sorties	Résultats de valeurs Minimales d'utilisation des entrées dans les sorties par la méthode de <i>points intérieurs</i> sur (Ctrl)													
Glycerol3P	98.9	32.9425	1.2428	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
Lactose	886	750.8260	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
C4	40.3	0	0	0	40.3200	0	0	0	0	0	0	0	0	
C6	27.1	0	0	18.0400	9.0200	0	0	0	0	0	0	0	0	
C8	17.8	0	0	13.3800	4.4600	0	0	0	0	0	0	0	0	
C10	46.6	0	0	37.2800	9.3200	0	0	0	0	0	0	0	0	
C12	50.8	0	0	42.3000	8.4600	0	0	0	0	0	0	0	0	
C14	195	0	0	166.8000	27.8000	0	0	0	0	0	0	0	0	
C16	301	0	0	263.4800	37.6400	0	0	0	0	0	0	0	0	
Glycine	10	3.4777	0.1150	0.9661	0.4016	0.0335	0.1384	0.0263	0.0245	0.0259	0.0836	0.0433	0.0188	0.0121
Glutamate	2.7	0.0352	0.0007	1.0139	0.4214	0.0256	0.0016	0.0211	0.0257	0.0148	0.0026	0.0530	0.0121	0.0148
Proline	53.3	0.6939	0.0146	19.9964	8.3115	0.5043	0.0313	0.4168	0.5063	0.2927	0.0507	1.0451	0.2387	0.2909
Aspartate	13.7	0.2452	0.0052	3.8069	1.5823	0.1071	0.0064	0.1205	0.0964	0.1119	0.0091	0.1882	0.0894	0.0524
SerineOutput	21.6	6.9991	0.2639	0.0076	0.0032	0.0003	0.3954	0.0002	0.0002	0.0002	0.2359	0.0003	0.0002	0.0001
CO ₂ Output	1126	396.1066	1.3422	342.7613	142.4687	12.1432	0.6533	9.0907	8.6779	9.0506	0.7338	15.1323	6.3353	4.2126

TABLE 4.1 – Calcul de valeurs minimales des AIOs pour (Ctrl) par la méthode de points intérieurs. Les résultats sont obtenus en environ 3 minutes de simulation. Les chiffres en gras sont les différences observées avec *active-set* (Table 4.3) et *sqp* (Table 4.2).

Entrées	Glucose 1422	Glycerol 17,5	Acetate 1020	BHBA 336	Lysine 16,1	Threonine 1,40	Isoleucine 13,1	Leucine 12,1	Valine 12,7	Histidine 1,38	Arginine 26,4	Alanine 9,33	Glutamine 6,10	
Sorties	Résultats de valeurs Minimales d'utilisation des entrées dans les sorties par la méthode des <i>sqp</i> sur (Ctrl)													
Glycerol3P	98.9	33.0292	1.2760	0	0	0	0	0	0	0	0.0000	0	0	
Lactose	886	750.8260	0	0	0	0	0	0	0	0	0	0.0000	0	
C4	40.3	0	0	0	40.3200	0	0	0	0	0	0	0	0	
C6	27.1	0	0	18.0400	9.0200	0	0	0	0	0	0	0	0	
C8	17.8	0	0	13.3800	4.4600	0	0	0	0	0	0	0	0	
C10	46.6	0	0	37.2800	9.3200	0	0	0	0	0	0	0	0	
C12	50.8	0	0	42.3000	8.4600	0	0	0	0	0	0	0	0	
C14	195	0	0	166.8000	27.8000	0	0	0	0	0	0	0	0	
C16	301	0	0	263.4800	37.6400	0	0	0	0	0	0	0	0	
Glycine	10	3.4777	0.1168	0.9748	0.4052	0.0336	0.1384	0.0267	0.0247	0.0262	0.0836	0.0439	0.0191	0.0122
Glutamate	2.7	0.0352	0.0121	1.0717	0.4455	0.0317	0.0022	0.0335	0.0342	0.0254	0.0039	0.0560	0.0227	0.0203
Proline	53.3	0.6939	0.0173	19.9964	8.4558	0.5437	0.0399	0.4192	0.5351	0.2927	0.0512	1.1043	0.2768	0.2937
Aspartate	13.7	0.2452	0.0061	3.8069	1.5823	0.1202	0.0103	0.1216	0.0969	0.1121	0.0153	0.1896	0.1138	0.0528
SerineOutput	21.6	6.9995	0.2710	0.0077	0.0032	0.0003	0.3954	0.0002	0.0002	0.0002	0.2359	0.0003	0.0002	0.0001
CO ₂ Output	1126	396.1066	1.3444	342.7613	142.4687	12.2048	0.6533	9.0907	8.6779	9.0506	0.7338	15.1323	6.3353	4.2126

TABLE 4.2 – Calcul de valeurs minimales des AIOs pour (Ctrl) par la méthode *sqp*. Les résultats sont obtenus en environ 2 minutes de simulation. Les chiffres en gras sont les différences observées avec la méthode de *points intérieurs* présenté dans la Table 4.1. La méthode de *points intérieurs* est celle qui donne les meilleurs résultats.

4.2.3.1 Complexité des méthodes basées sur le calcul formel de l'inverse d'une matrice

Soit $\mathcal{M}_s(\mathbb{K})$, l'ensemble des matrices carrées à coefficients dans \mathbb{K} ($\mathbb{K} = \mathbb{R}$ ou \mathbb{C}). Pour toute matrice carrée $A = (a_{ij})_{1 \leq i \leq s, 1 \leq j \leq s} \in \mathcal{M}_s(\mathbb{K})$, le déterminant de A est noté par $\det(A)$ ou encore $|A|$ et est donné par la définition suivante

Définition 4.2.1. *Le déterminant d'une matrice carrée $A : s \times s$ est la somme de tous les produits élémentaires signés de A :*

$$\det(A) = \sum_{\sigma \in \mathcal{P}_s} \text{sign}(\sigma) a_{i_1 1} a_{i_2 2} \dots a_{i_s s}$$

où \mathcal{P}_s est l'ensemble des permutations de s et $\text{sign}(\sigma) = +1$ ou -1 .

Entrées	Glucose	Glycerol	Acetate	BHBA	Lysine	Threonine	Isoleucine	Leucine	Valine	Histidine	Arginine	Alanine	Glutamine	
	1422	17,5	1020	336	16,1	1,40	13,1	12,1	12,7	1,38	26,4	9,33	6,10	
Sorties	Résultats de valeurs Minimales d'utilisation des entrées dans les sorties par la méthode des <i>active-set</i> sur (Ctrl)													
Glycerol3P	98.9	32.9425	1.2428	0	0	0	0.0000	0	0	0	0	0.0000	0	
Lactose	886	750.8260	0	0	0	0	0	0	0	0	0.0000	0.0000	0.0000	
C4	40.3	0	0	0	40.3200	0	0	0	0	0	0	0	0	
C6	27.1	0	0	18.0400	9.0200	0	0	0	0	0	0	0	0	
C8	17.8	0	0	13.3800	4.4600	0	0	0	0	0	0	0	0	
C10	46.6	0	0	37.2800	9.3200	0	0	0	0	0	0	0	0	
C12	50.8	0	0	42.3000	8.4600	0	0	0	0	0	0	0	0	
C14	195	0	0	166.8000	27.8000	0	0	0	0	0	0	0	0	
C16	301	0	0	263.4800	37.6400	0	0	0	0	0	0	0	0	
Glycine	10	3.4777	0.1175	0.9661	0.4035	0.0336	0.1384	0.0267	0.0247	0.0262	0.0836	0.0439	0.0191	0.0122
Glutamate	2.7	0.0352	0.0009	1.0139	0.4214	0.0275	0.0022	0.0222	0.0271	0.0160	0.0039	0.0560	0.0138	0.0156
Proline	53.3	0.6939	0.0173	19.9964	8.3115	0.5043	0.0333	0.4192	0.5063	0.2927	0.0536	1.0451	0.2387	0.2909
Aspartate	13.7	0.2452	0.0061	3.8069	1.5823	0.1075	0.0070	0.1216	0.0969	0.1128	0.0099	0.1896	0.0949	0.0528
SerineOutput	21.6	6.9991	0.2640	0.0077	0.0032	0.0003	0.3954	0.0002	0.0002	0.0002	0.2359	0.0003	0.0002	0.0001
CO ₂ Output	1126	396.1066	1.3422	342.7613	142.4687	12.1432	0.6966	9.0907	8.6779	9.0506	0.7338	15.1323	6.3353	4.2126

TABLE 4.3 – Calcul de valeurs minimales des AIOs pour (Ctrl) par la méthode *active-set*. Les résultats sont obtenus en environ 45 minutes de simulation. Les chiffres en gras sont les différences observées avec la méthode de *points intérieurs* présenté dans la Table 4.1. La méthode de *points intérieurs* est celle qui donne les meilleurs résultats.

Définition 4.2.2. a) On appelle mineur de la matrice A , le déterminant d'ordre $(s - 1)$ de A noté Δ_{ij} et obtenu en supprimant dans A la $i^{\text{ème}}$ ligne et la $j^{\text{ème}}$ colonne.

b) On appelle cofacteur de A le nombre noté C_{ij} et défini par : $C_{ij} = (-1)^{i+j} \Delta_{ij}$.

c) On appelle comatrice de A , notée \tilde{A} , la matrice des cofacteurs, c'est-à-dire $\tilde{A} = (C_{ij})_{1 \leq i \leq s, 1 \leq j \leq s}$.

Proposition 4.2.1. Soit A une matrice carrée inversible, alors si A^{-1} désigne son inverse on a :

$$A^{-1} = \frac{1}{\det(A)} \tilde{A}^{\top},$$

où $(.)^{\top}$: désigne la transposée.

Plusieurs algorithmes permettent de calculer l'inverse d'une matrice dont l'algorithme du pivot de Gauss [PT86] et plus efficacement l'algorithme de Bareiss [Lio96].

Quelque soit l'algorithme utilisé, le calcul de l'inverse A^{-1} d'une matrice carrée A nécessite de stocker en mémoire au moins un mineur de A .

Théorème 4.2.2. Tout mineur d'une matrice carrée dont les éléments sont des variables libres et indépendantes a une complexité en mémoire exponentielle par rapport aux variables de la matrice.

Démonstration. Ceci découle du fait que si les éléments de la matrice sont des variables libres et indépendantes alors chaque mineur est une somme de monômes de degré $s - 1$ en $(s - 1)^2$ variables. \square

Corollaire 4.2.1. *Toute méthode d'optimisation basée sur du calcul formel de l'inverse d'une matrice a une complexité en mémoire exponentielle par rapport aux variables de la matrice.*

Démonstration. Ceci découle immédiatement du Théorème 4.2.2. □

4.2.3.2 Distance des solutions optimales par rapport au bord de l'espace

Il s'agit ici de présenter des résultats sur les méthodes de points intérieurs par rapport au bord de l'espace de recherche.

Proposition 4.2.2. *Les méthodes de points intérieurs n'explorent pas exactement le bord de l'espace.*

Démonstration. Découle de la présence du logarithme dans les conditions d'optimalité présentées dans le système 4.7 ci-dessus. En effet, le logarithme n'est défini que pour des arguments strictement positifs. Par la suite la fonction barrière logarithmique n'est définie que pour les points se trouvant à l'intérieur du domaine réalisable (c'est-à-dire vérifiant $-c_I(x) > 0$). □

Distances des solutions minimales des AIOs pour (Ctrl) par rapport au bord de l'espace des solutions. Comme exemple d'application, nous présentons dans la Table 4.4 les distances des solutions minimales des AIOs pour (Ctrl) par rapport au bord de l'espace de recherche. Ici la distance d'une solution optimale par rapport au bord de l'espace de recherche est définie comme étant la distance minimale entre la solution optimale et les différents hyperplans qui définissent l'espace de recherche.

Entrées	Glucose	Glycerol	Acetate	BHBA	Lysine	Threonine	Isoleucine	Leucine	Valine	Histidine	Arginine	Alanine	Glutamine
Sorties	Distances des solutions minimales par rapport au bord de l'espace des solutions pour (Ctrl)												
Glycerol3P	1e-7	3e-6	2e-8	4e-8	8e-9	1e-7	5e-9	7e-9	6e-9	6e-8	3e-7	8e-9	1e-8
Lactose	3e-6	9e-8	3e-010	7e-8	1e-8	2e-7	8e-9	1e-8	9e-9	1e-7	5e-9	1e-8	2e-8
C4	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7
C6	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7
C8	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7
C10	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7
C12	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7
C14	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7
C16	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7	5.7
Glycine	1e-8	2e-6	3e-9	1e-8	3e-5	1e-6	1e-7	4e-6	2e-5	2e-6	1e-5	2e-7	8e-6
Glutamate	4e-8	2e-6	1e-7	1e-6	2e-5	3e-5	5e-6	9e-7	7e-6	8e-6	8e-6	2e-5	6e-6
Proline	2e-9	2e-8	2e-8	1e-8	4e-8	1e-5	3e-5	9e-7	8e-8	8e-6	8e-8	8e-6	7e-8
Aspartate	6e-8	1e-8	8e-6	1e-6	1e-5	1e-5	9e-6	2e-5	1e-5	3e-5	4e-7	8e-6	3e-5
SerineOutput	5e-9	4e-6	2e-8	2e-7	3e-8	5e-7	2e-8	3e-8	2e-8	3e-7	1e-8	3e-8	5e-8
CO ₂ Output	3e-7	1e-9	3e-8	2e-6	4e-7	3e-5	4e-7	3e-7	2e-6	1e-5	8e-7	2e-7	1e-5

TABLE 4.4 – Distances des solutions minimales des AIOs pour (Ctrl) par rapport au bord de l'espace des solutions. *Les valeurs des distances sont de l'ordre de 1e-6, donc très proche de l'espace de recherche* (les valeurs 5.7 sont des distances pour lesquelles les fonctions correspondantes sont des fonctions constantes par rapport aux variables).

Les résultats des distances des solutions minimales (Table 4.4) permettent intuitivement de dire que les solutions optimales sont sur le bord de l'espace de recherche. Pour

être plus précis nous avons proposé un algorithme de recherche locale pour explorer le bord de l'espace de recherche.

4.3 Algorithme de recherche locale pour calculer les optimaux des AIOs

Dans la section 4.2.2, nous avons montré que les solutions optimales calculées sont très proches du bord de l'espace de recherche en utilisant une méthode de *points intérieurs*.

Nous présentons et utilisons ici un algorithme de recherche locale qui permet d'explorer de manière plus fine le bord de l'espace de recherche contrairement aux méthodes de *points intérieurs*.

Cet algorithme de recherche locale est également une approche alternative pour calculer les valeurs minimales et maximales d'utilisation des entrées dans les sorties lorsque la forme analytique des AIOs n'est pas disponible.

4.3.1 Modélisation

La méthode est inspirée des heuristiques existant sur le multi-objectif qui consistent à discrétiser l'espace de recherche des solutions pour générer des points afin de les faire converger vers les solutions optimales.

Nous proposons ici une discrétisation du bord de l'espace des solutions et une recherche locale de solutions optimales des AIOs. Nous faisons ici l'hypothèse que l'espace de recherche est un polyèdre borné, ensuite nous déterminons le *centre de Chebyshev*, x_c , de cet espace. En considérant une boule ayant pour centre x_c , tout point de la boule peut être paramétrisé. Ceci, implique une paramétrisation de tout point du polyèdre, en particulier les points du bord. Ainsi, on peut définir une discrétisation autour de chaque point du polyèdre pour appliquer une recherche locale d'optimum.

Cette recherche locale sur le bord a permis de retrouver les solutions optimales calculées par la méthode de points intérieurs présentée ci-dessus.

4.3.1.1 Centre de Chebyshev d'un polyèdre

Considérons le polyèdre \mathcal{H} défini par :

$$\mathcal{H} = \{x \in \mathbb{R}^s : (A^i)^\top x \leq b_i; i \in \{1, 2, \dots, m\} \text{ et } x \geq 0\}, \quad (4.8)$$

où s est le nombre de variables libres du système, A^i est un vecteur réel de dimension n et b_i un réel.

On appelle *centre de Chebyshev* du polyèdre \mathcal{H} , le centre de la plus grande boule, $\mathcal{B}(x_c, r)$, contenue dans \mathcal{H} où x_c représente le centre et r le rayon de la boule.

Proposition 4.3.1. *Le calcul du centre et du rayon de Chebyshev revient à résoudre le problème linéaire suivant*

$$\begin{cases} \text{Maximiser } k^T v \\ Mf(v_{libre}) \leq B \end{cases} \quad (4.9)$$

où, $v = (x_c, r)$ est un vecteur de dimension $s + 1$. k , un vecteur de dimension $n + 1$ dont seule la $(s + 1)^{ième}$ composante est non nulle et est égale à 1.

$M = (A^T, (A^{s+1})^T)$ une matrice de dimension $m \times (s + 1)$.

Avec, $A = (A^1, A^2, \dots, A^m)$ et $A^{s+1} = (\|A^1\|_2, \|A^2\|_2, \dots, \|A^m\|_2)$, pour $A^i, i = 1, 2, \dots, m$ définie dans l'équation 4.8. $B = (b_1, b_2, \dots, b_m)$, b_i définie dans l'équation 4.8 pour $i = 1, 2, \dots, m$.

Démonstration. Par définition et comme le montre la Figure 4.1, on cherche à déterminer la valeur maximale du rayon r tel que la boule, $\mathcal{B}(x_c, r)$, de centre x_c et de rayon r soit contenu dans \mathcal{H} . Ce qui revient donc à résoudre le problème

$$\begin{cases} \text{Maximiser } r \\ \text{s.c. } \mathcal{B}(x_c, r) \subset \mathcal{H} \end{cases}$$

$$\begin{aligned} \mathcal{B}(x_c, r) \subset \mathcal{H} &\Leftrightarrow x_c + u \in \mathcal{H}, \forall u \in \mathcal{B}(0, r) \\ &\Leftrightarrow (A^i)^T(x_c + u) \leq b_i, \forall u : \|u\|_2 \leq r, i = 1, 2, \dots, m \\ &\Leftrightarrow \sup_{u: \|u\|_2 \leq r} (A^i)^T u \leq b_i - (A^i)^T x_c, i = 1, 2, \dots, m \end{aligned}$$

Or, d'après l'inégalité de Cauchy-Schwartz :

$$(A^i)^T u \leq \|A^i\|_2 \|u\|_2 \leq r \|A^i\|_2 = \sup_{u: \|u\|_2 \leq r} (A^i)^T u \quad \forall i = 1, 2, \dots, m$$

Donc,

$$\mathcal{B}(x_c, r) \subset \mathcal{H} \Leftrightarrow (A^i)^T x_c + r \|A^i\|_2 \leq b_i \quad \forall i = 1, 2, \dots, m$$

□

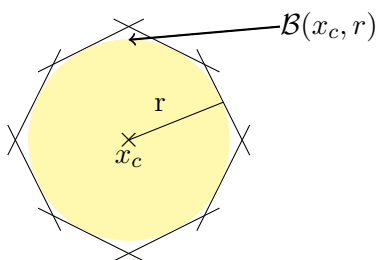


FIGURE 4.1 – Centre de Chebyshev.

4.3.1.2 Discrétisation du bord du polyèdre

La discrétisation du bord du polyèdre permet d'appliquer une recherche locale d'optimum autour de chaque point du bord.

Soit un point x de la frontière de la boule $\mathcal{B}(x_c, r)$. Ce point est identifié de manière unique par un vecteur $\theta = (\theta_1, \dots, \theta_s)$ avec $\theta_l \in [0, 2\pi[$ pour $1 \leq l < s$ et $\theta_s \in [0, \pi[$. Comme \mathcal{H} est un convexe, alors il existe un unique point d'intersection entre la demi-droite $[x_c, x)$ et le polyèdre \mathcal{H} . On note $P(x)$, ce point. Inversement, pour tout point y du bord du polyèdre, il existe un unique point x appartenant à la frontière de la boule $B(x_c, r)$ tel que $y = P(x)$. Par extension nous dirons que, $(\theta_1, \dots, \theta_s)$ est également une paramétrisation du bord du polyèdre. Ainsi, on peut utiliser cette paramétrisation pour déterminer une discrétisation du bord.

Soit p un entier strictement positif, chaque θ_l est supposé être sous la forme $2\pi j_l/p$, pour $j_l \in \{0, \dots, p\}$, $i \in \{0, \dots, s-1\}$, et θ_s sous la forme $\pi j_s/p$.

Soit $(2\pi j_1/p, \dots, 2\pi j_i/p, \dots, \pi j_s/p)$, une paramétrisation d'un point y du bord.

On appelle *points voisins de y* de niveau 1 par rapport à la discrétisation p et on note $\mathcal{N}_1^p(y)$ les points obtenus en ajoutant $+1$ ou -1 à chaque coordonnée de paramétrisation. Par exemple $(\theta_1, \dots, 2\pi(j_i + 1)/p, \dots, \theta_s)$ et $(\theta_1, \dots, 2\pi(j_i - 1)/p, \dots, \theta_s)$ sont définis comme des voisins de $(\theta_1, \dots, 2\pi j_i/p, \dots, \theta_s)$ en modifiant une coordonnée de paramétrisation. En procédant ainsi, l'ensemble $\mathcal{N}_1^p(y)$ contiendra $2s$ voisins pour y .

On définit également les *voisins de niveau 2* en ajoutant $+2$ ou -2 et on note l'ensemble $\mathcal{N}_2^p(y)$. De façon générale l'ensemble des *voisins de niveau N* , obtenu en ajoutant $+N$ ou $-N$ sera, noté $\mathcal{N}_N^p(y)$.

4.3.1.3 Calcul des coordonnées des points du bord du polyèdre

Le calcul des coordonnées des points du bord permet pour chaque $P(x)$ et de ses voisins, définis ci-dessus, de déterminer les coordonnées qui les identifient de manière unique dans l'espace de recherche. Ces coordonnées permettent d'évaluer la valeur de la fonction à optimiser en chacun de ces points et de déterminer la valeur optimale. Plus précisément, dans l'expression de $AIO[v][i, k]$ de la section 4.1, on détermine la valeur de v pour chaque $P(x)$ et de ses voisins.

Considérons les ensembles : $\mathcal{H}_i = \{x \in \mathbb{R}^s : (A^i)^\top x - b_i = 0\}$, $i \in \{1, 2, \dots, m\}$, appelés hyperplans de \mathbb{R}^n . Soit $x \in B(x_c, r)$, on note $q = x - x_c$ la direction de la demi-droite $[x_c, x)$. On a la proposition suivante :

Proposition 4.3.2. *Si $[x_c, x)$ possède une intersection avec un hyperplan \mathcal{H}_i alors elle le coupe en un point $x_{\mathcal{H}_i}$ défini par*

$$x_{\mathcal{H}_i} = x - q * \frac{(A^i)^\top x - b_i}{(A^i)^\top q} \quad (4.10)$$

Le point $P(x)$ intersection de la demi-droite $[x_c, x)$ et du bord du polyèdre est défini

par $P(x) = x_{\mathcal{H}_{i_0}}$ tel que

$$\|x_{\mathcal{H}_{i_0}}\|_2 = \min_i (\|x_{\mathcal{H}_i}\|_2, i \in \{1, 2, \dots, m\}) \quad (4.11)$$

4.3.2 Principe de l'algorithme

Le principe de l'algorithme consiste à tirer des points et à évaluer la fonction à optimiser en ces points. Ainsi, pour la recherche locale sur le bord de l'espace de recherche, on tire aléatoirement avec une probabilité de :

1. $a\%$ d'avoir un point y au bord du polyèdre (en générant une direction aléatoire d suivant la discrétisation de la boule de Chebyshev à partir de x_c) pour lequel on évalue la fonction à optimiser en y puis aux points en $\mathcal{N}_1^p(y), \mathcal{N}_2^p(y), \dots$ tant qu'on peut améliorer la fonction.
2. $(1 - a)\%$ d'améliorer la recherche autour du meilleur y obtenu au 1) en raffinant le niveau de discrétisation, par exemple en prenant $p' = 10 * p$, ensuite en évaluant la fonction à optimiser en $\mathcal{N}_1^{p'}(y), \mathcal{N}_2^{p'}(y), \dots$ tant qu'on peut améliorer la fonction.

Ici a est un paramètre qu'on peut utiliser pour affiner aléatoirement la recherche de solution. La valeur de $a = 80$ donne globalement des meilleurs résultats pour $p = 1000$ et $p' = 10000$. Ainsi toutes les simulations ont été faites en considérant ces paramètres.

Notations

- On note ω le bord du polyèdre ;
- d_+ le sens positif de la direction de d et $y_+ = d_+ \cap \omega$;
- d_- le sens négatif de la direction de d et $y_- = d_- \cap \omega$.

4.3.3 L'algorithme

L'algorithme ci-dessous permet de faire une recherche locale d'optimum pour une fonction non linéaire donnée sur un polyèdre borné. On a la proposition suivante :

Proposition 4.3.3. *Pour toute fonction non linéaire continue (f) et pour un nombre de tirage donné ($nbreDetirages$), l'algorithme ci-dessous effectue un nombre fini d'itérations sur un polyèdre borné pour la recherche de solution optimale. Autrement dit, il termine.*

Démonstration. On sait que toute fonction continue sur un ensemble fermé borné est bornée. C'est-à-dire qu'elle admet une plus petite et plus grande valeur. Comme le polyèdre est discrétisé de manière fini alors chaque boucle *tant que* termine après un nombre fini d'itérations. \square

Algorithme 3 Recherche locale.

Entrées : $nbreDetirages \in \mathbb{R}$, f (fonction à optimiser), $D_1 \in \mathbb{R}$ (paramètre de discrétisation), $D_2 \in \mathbb{R}$ (paramètre de discrétisation avec $D_2 > D_1$),

Sorties : y_{opt} (point optimal pour la fonction f), f_{opt} (valeur optimal de la fonction f)

On pose : $p_D = 1$, $val = \infty$, $valmin = \infty$

Pour i allant de 1 à $nbreDetirages$ **Faire**

$(2\pi r_1, 2\pi r_2, \dots, 2\pi r_{s-1}, \pi r_s) \leftarrow r$ avec $(r_i \in \{1, \dots, D_1 - 1\}, i \in \{1, \dots, s\})$
 $k_1 = 1, k_2 = 1, k = 1$

Tant que $(p_D > 0.2)$ **Faire**

$r/D_1 \leftarrow d$, On calcule : y_+ et y_- en utilisant les équations (4.10 et 4.11)

$f(t_0) = \min_{t \in \mathcal{N}_1^{D_1}(y_+)} (f(t))$

Tant que $(f(y_+) > f(t_0))$ **Faire**

$k_1 + 1 \leftarrow k_1$, On pose : $f(\tau) = \min_{t \in \mathcal{N}_{k_1}^{D_1}(y_+)} (f(t))$

Si $(f(\tau) < f(t_0))$ **Alors**

$f(\tau) \leftarrow f(t_0), \tau \leftarrow t_0$

Fin Si

Fin Tant que

$f(t'_0) = \min_{t \in \mathcal{N}_1^{D_1}(y_-)} (f(t))$

Tant que $(f(y_-) > f(t'_0))$ **Faire**

$k_2 + 1 \leftarrow k_2$, On pose : $f(\tau') = \min_{t \in \mathcal{N}_{k_2}^{D_1}(y_-)} (f(t))$

Si $(f(\tau') < f(t'_0))$ **Alors**

$f(\tau') \leftarrow f(t'_0), \tau' \leftarrow t'_0$

Fin Si

Fin Tant que

Si $(val < f(t_0))$ **Alors**

$f(t_0) \leftarrow val, t_0 \leftarrow \zeta$

Sinon

$f(t'_0) \leftarrow val, t'_0 \leftarrow \zeta$

Fin Si

Si $(val < valmin)$ **Alors**

$val \leftarrow valmin$

Fin Si

On génère une nouvelle valeur $\alpha \in [0, 1]$ et $\alpha \leftarrow p_D$

Fin Tant que

$valmin \leftarrow f_{opt}, \zeta \leftarrow y_{opt}$

$r/D_2 \leftarrow d$, On pose : $f(\zeta') = \min_{t \in \mathcal{N}_1^{D_2}(\zeta)} (f(t))$

Tant que $(f(\zeta') > f(\zeta))$ **Faire**

$k + 1 \leftarrow k$, On pose : $f(\zeta'') = \min_{t \in \mathcal{N}_k^{D_2}(\zeta)} (f(t))$

Si $(f(\zeta'') < f(\zeta'))$ **Alors**

$f(\zeta'') \leftarrow f_{opt}, \zeta'' \leftarrow y_{opt}$

Fin Si

Fin Tant que

Fin Pour

Calcul des solutions minimales des AIOs pour (Ctrl) en utilisant l'algorithme de recherche locale. Comme exemple d'application, nous présentons dans la Table 4.5 les solutions minimales des AIOs pour (Ctrl). Comme mentionné dans la section 4.1, l'algorithme de recherche locale est appliqué ici à un espace de recherche à 4 dimensions. Sur plusieurs résultats l'algorithme trouve exactement les mêmes résultats que la méthode de points intérieurs de *fmincon* mais aucun résultat n'est strictement inférieur à celui de *fmincon*. On peut donc conclure que les solutions optimales sont sur le bord de l'espace de recherche.

Entrées	Glucose 1422	Glycerol 17,5	Acetate 1020	BHBA 336	Lysine 16,1	Threonine 1,40	Isoleucine 13,1	Leucine 12,1	Valine 12,7	Histidine 1,38	Arginine 26,4	Alanine 9,33	Glutamine 6,10
Sorties	Résultats de valeurs Minimales d'utilisation des entrées dans les sorties en utilisant l'algorithme de recherche locale sur (Ctrl)												
Glycerol3P	98.9	32.9897	1.2466	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Lactose	886	751.0898	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
C4	40.3	0	0	0	40.3200	0	0	0	0	0	0	0	0
C6	27.1	0	0	18.0400	9.0200	0	0	0	0	0	0	0	0
C8	17.8	0	0	13.3800	4.4600	0	0	0	0	0	0	0	0
C10	46.6	0	0	37.2800	9.3200	0	0	0	0	0	0	0	0
C12	50.8	0	0	42.3000	8.4600	0	0	0	0	0	0	0	0
C14	195	0	0	166.8000	27.8000	0	0	0	0	0	0	0	0
C16	0	0	0	263.4800	37.6400	0	0	0	0	0	0	0	0
Glycine	10	3.4819	0.1177	0.9679	0.4016	0.0335	0.1384	0.0264	0.0245	0.0260	0.0836	0.0434	0.0188
Glutamate	2.7	0.0662	0.0033	1.0329	0.4314	0.0256	0.0016	0.0211	0.0257	0.0150	0.0026	0.0543	0.0121
Proline	53.3	0.7159	0.0237	20.5745	8.4642	0.5084	0.0323	0.4281	0.5214	0.3034	0.0515	1.0511	0.2500
Aspartate	13.7	0.3343	0.0197	3.8069	1.6104	0.1071	0.0064	0.1263	0.1003	0.1145	0.0091	0.1931	0.0908
SerineOutput	21.6	7.0460	0.2715	0.0076	0.0032	0.0003	0.3954	0.0002	0.0002	0.0002	0.2359	0.0003	0.0002
CO ₂ Output	1126	396.3780	1.4097	342.8778	143.3149	12.2325	0.6548	9.1491	8.8390	9.0989	0.7338	15.1780	6.3648

TABLE 4.5 – Calcul de valeurs minimales des AIOs pour (Ctrl) en utilisant l'algorithme de recherche locale. La plupart des valeurs correspondent aux valeurs trouvées par la méthode des points intérieurs données dans la Table 4.1. Les résultats sont obtenus en environ 300 minutes de simulation.

4.3.4 Intérêt fondamental de l'algorithme de recherche locale

L'algorithme de recherche locale est une approche alternative pour calculer les valeurs minimales et maximales d'utilisation des entrées dans les sorties lorsque la forme analytique des AIOs n'est pas disponible.

Proposition 4.3.4. *L'algorithme de recherche locale est un algorithme polynomial en mémoire et qui permet de calculer les contributions des entrées dans les sorties indépendamment de la disponibilité des fonctions analytiques des AIOs avec un nombre d'opérations de $O(K[(n + q)^3 + npq])$. Avec : p , le nombre de métabolites d'entrées ; q , le nombre de métabolites de sorties ; n , le nombre de métabolites intermédiaires et K , le nombre de points de l'espace de recherche utilisé.*

Démonstration. Ceci, découle du fait que la recherche locale peut être utilisée sans que la forme analytique de $\bar{D}_1^{-1}[v_{libre}]$ de la section 4.1 ne soit disponible. C'est-à-dire on a juste besoin de stocker en mémoire $\bar{D}_1[v_{libre}]$ et non $\bar{D}_1^{-1}[v_{libre}]$, ce qui donne une complexité polynomiale en mémoire. Ensuite, pour chaque point I de l'espace de recherche, on détermine la distribution de flux v_I correspondant au point I , on calcule $D_1^{-1}[v_I]$ sans passer par le calcul formel de $D_1^{-1}[v]$. Le calcul de $D_1^{-1}[v_I]$ permet d'avoir les AIO

correspondant au point *I*. Enfin, le calcul des AIO sur plusieurs tirages permet de déterminer des valeurs approchées des Min/Max.

□

4.4 Application biologique

Les contributions maximales et minimales des flux de carbones des nutriments prélevés (Entrées) dans les constituants du lait (Sorties) sont calculées pour trois régimes alimentaires : un régime alimentaire appelé contrôle (Ctrl) [LRL⁺09,RLL⁺06] correspondant à une alimentation proche des recommandations INRA pour les vaches laitières ; un régime alimentaire appelé caséine (CN) [LRL⁺09,RLL⁺06] correspondant une alimentation où la quantité de protéines digérées dans l'intestin était augmentée de 30% et un régime alimentaire appelé (HB) [Han94], qui est jeu de données qui a été constitué à partir de résultats de différentes expériences.

Les données correspondant à (Ctrl) et (CN) et (HB) sont présentées dans la Table 2.1 du chapitre 2.

4.4.1 Tables des résultats

Les Tables 4.6, 4.7 et 4.8 donnent les utilisations minimales des entrées dans les sorties pour les jeux de données (Ctrl), (CN) et (HB).

Entrées	Glucose	Glycerol	Acetate	BHBA	Lysine	Threonine	Isoleucine	Leucine	Valine	Histidine	Arginine	Alanine	Glutamine
	1422	17,5	1020	336	16,1	1,40	13,1	12,1	12,7	1,38	26,4	9,33	6,10
Sorties	Utilisations Minimales des entrées dans les sorties (en mmol/h/démi-mamelle de Carbone) pour (Ctrl)												
Glycerol3P	98.9	32.9	1.2	0	0	0	0	0	0	0	0	0	0
Lactose	886	751	0	0	0	0	0	0	0	0	0	0	0
C4	40.3	0	0	0	40.3	0	0	0	0	0	0	0	0
C6	27.1	0	0	18	9	0	0	0	0	0	0	0	0
C8	17.8	0	0	13.4	4.5	0	0	0	0	0	0	0	0
C10	46.6	0	0	37.3	9.3	0	0	0	0	0	0	0	0
C12	50.8	0	0	42.3	8.5	0	0	0	0	0	0	0	0
C14	195	0	0	166.8	27.8	0	0	0	0	0	0	0	0
C16	301	0	0	263.5	37.6	0	0	0	0	0	0	0	0
Glycine	10	3.5	0.1	1	0.4	0	0.1	0	0	0.1	0	0	0
Glutamate	2.7	0	0	1	0.4	0	0	0	0	0	0.1	0	0
Proline	53.3	0.7	0	20	8.3	0.5	0	0.4	0.5	0.3	0.1	0.2	0.3
Aspartate	13.7	0.2	0	3.8	1.6	0.1	0	0.1	0.1	0	0.2	0.1	0.1
SerineOutput	21.6	7	0.3	0	0	0	0.4	0	0	0.2	0	0	0
CO ₂ Output	1126	396.1	1.3	342.8	142.5	12.1	0.7	9.1	8.7	9.1	0.7	15.1	4.2

TABLE 4.6 – Utilisations minimales des AIOs pour (Ctrl). La table présente les valeurs des quantités des nutriments prélevés par la mamelle, des nutriments dans le lait et les quantités minimales des entrées qui sont réparties dans les sorties. Par exemple, 1422 mmol/h de Glucose sont prélevés et 886 mmol/h sont mesurés dans le lait. La quantité minimale du Glucose qui contribue dans le Lactose est de 751 pour le régime alimentaire (Ctrl). Le pourcentage de la contribution minimale du Glucose dans le Lactose est de $\frac{751}{886} = 85\%$.

Les Tables 4.9, 4.10 et 4.11 donnent les utilisations maximales des entrées dans les sorties pour les jeux de données Ctrl, CN et HB.

Entrées	Glucose	Glycerol	Acetate	BHBA	Lysine	Isoleucine	Leucine	Valine	Arginine	Glutamine
	1392	17.2	924	668	21.5	21.4	22.6	19.3	26.9	8.95
Sorties	Utilisations Minimales des entrées dans les sorties (en mmol/h/démi-mamelle de Carbone) pour (CN)									
Glycerol3P	117	36.1	1.8	0	0	0	0	0	0	0
Lactose	1002	866	0	0	0	0	0	0	0	0
C4	46.4	0	0	0	46.4	0	0	0	0	0
C6	33.5	0	0	22.3	11.2	0	0	0	0	0
C8	23	0	0	17.2	5.7	0	0	0	0	0
C10	64.6	0	0	51.7	12.9	0	0	0	0	0
C12	73.4	0	0	61.2	12.2	0	0	0	0	0
C14	250	0	0	214.7	35.8	0	0	0	0	0
C16	343	0	0	300.2	42.9	0	0	0	0	0
Glycine	6.9	2.1	0.1	0.4	0.7	0	0	0	0	0
Alanine	9.8	1.4	0	0	0	0	0	0	0	0
Glutamate	31.7	0.8	0	6.7	13	0.4	0.4	0.6	0.3	0.7
Proline	55	1.4	0	11.6	22.6	0.7	0.8	1	0.5	1.2
Aspartate	16.5	0.6	0	2.7	5.3	0.2	0.3	0.2	0.2	0.3
SerineOutput	22.5	6.9	0.3	0	0	0	0	0	0	0
CO ₂ Output	1021	243.6	2	180	351.1	16.3	14.2	15.8	13.4	14.8
										5.9

TABLE 4.7 – Utilisations minimales des AIOs pour (CN). La table présente les valeurs des quantités des nutriments prélevés par la mamelle, des nutriments dans le lait et les quantités minimales des entrées qui sont réparties dans les sorties. Par exemple, 1392 mmol/h de Glucose sont prélevées et 1002 mmol/h de Lactose sont mesurées dans le lait. La quantité minimale du Glucose qui contribue dans le Lactose est de 866 mmol/h pour le régime alimentaire (CN). Le pourcentage de la contribution minimale du Glucose dans le Lactose est de $:\frac{866}{1002} = 85\%$.

Entrées	Glucose	Glycerol	Acetate	BHBA	Lactate	Lysine	Isoleucine	Valine	Arginine	Alanine	Glutamine
	1526	2.06	768	604	1.44	23.9	190	45.6	65.8	6.56	7.50
Sorties	Utilisations Minimales des entrées dans les sorties (en mmol/h/démi-mamelle de Carbone) pour (HB)										
Glycerol3P	95	9.1	0.1	0	0	0	0	0	0	0	0
Lactose	953	741	0	0	0	0	0	0	0	0	0
C4	38	0	0	0	38.3	0	0	0	0	0	0
C6	23	0	0	15	7.5	0	0	0	0	0	0
C8	17	0	0	12.5	4.2	0	0	0	0	0	0
C10	40	0	0	31.7	7.9	0	0	0	0	0	0
C12	43	0	0	35.4	7.1	0	0	0	0	0	0
C14	131	0	0	112.5	18.7	0	0	0	0	0	0
C16	213	0	0	186.7	26.7	0	0	0	0	0	0
Glycine	10	1.7	0	0.6	0.7	0	0.3	0.1	0.1	0	0
Glutamate	24	0	0	5.1	6.7	0	0.2	1.9	0.3	0.8	0.1
Proline	70	0	0	14.8	19.6	0	0.7	5.6	1	2.4	0.1
Aspartate	21	0	0	3.3	4.3	0	0.2	1.9	0.4	0.5	0.1
Asparagine	2	0	0	0.3	0.4	0	0	0.2	0	0	0
SerineOutput	6	0.5	0	0	0	0	0	0	0	0	0
CO ₂ Output	1546	439.3	0	267.5	353.4	1	17.7	134.7	33.1	38.1	4.6
											5.2

TABLE 4.8 – Utilisations minimales des AIOs pour (HB). La table présente les valeurs des quantités des nutriments prélevés par la mamelle, des nutriments dans le lait et les quantités minimales des entrées qui sont réparties dans les sorties. Par exemple, 1526 mmol/h de Glucose sont prélevées et 953 mmol/h de Lactose sont mesurées dans le lait. La quantité minimale du Glucose qui contribue dans le Lactose est de 741 mmol/h pour le régime alimentaire (HB). Le pourcentage de la contribution minimale du Glucose dans le Lactose est de $:\frac{741}{953} = 78\%$.

Pour chaque colonne des tables, les valeurs, correspondent aux utilisations minimales de flux de carbone des entrées dans les sorties. Par exemple, pour la Table 4.6 la première colonne de la table montre que sur les 1422 flux de carbone du glucose en entrée la quantité minimale qui va dans le lactose est de 751.

Entrées	Glucose	Glycerol	Acetate	BHBA	Lysine	Threonine	Isoleucine	Leucine	Valine	Histidine	Arginine	Alanine	Glutamine	
	1422	17,5	1020	336	16,1	1,40	13,1	12,1	12,7	1,38	26,4	9,33	6,10	
Sorties	Utilisations Maximales des entrées dans les sorties (en mmol/h/démi-mamelle de Carbone) pour (Ctrl)													
Glycerol3P	98.9	97	12.5	38.4	16	1.1	0.1	1.3	1	1.2	0.1	2	1	0.6
Lactose	886	886	7.7	79.8	33.2	2.3	0.1	2.6	2	2.4	0.2	4	2	1.1
C4	40.3	0	0	40.3	0	0	0	0	0	0	0	0	0	0
C6	27.1	0	0	18	9	0	0	0	0	0	0	0	0	0
C8	17.8	0	0	13.4	4.5	0	0	0	0	0	0	0	0	0
C10	46.6	0	0	37.3	9.3	0	0	0	0	0	0	0	0	0
C12	50.8	0	0	42.3	8.5	0	0	0	0	0	0	0	0	0
C14	195	0	0	166.8	27.8	0	0	0	0	0	0	0	0	0
C16	301	0	0	263.5	37.6	0	0	0	0	0	0	0	0	0
Glycine	10	8	0.9	3.7	1.5	0.1	0.1	0.1	0.1	0.1	0.1	0.2	0.1	0.1
Glutamate	2.7	1.1	0	1.6	0.7	0	0	0	0	0	0	0.1	0	0
Proline	53.3	20.8	0.7	32.4	13.5	0.7	0	0.9	0.8	0.7	0.1	1.8	0.6	0.5
Aspartate	13.7	7.4	0.3	8.2	3.4	0.2	0	0.3	0.2	0.3	0	0.4	0.2	0.1
SerineOutput	21.6	20.6	2.6	8.2	3.4	0.2	0.4	0.3	0.2	0.3	0.3	0.4	0.2	0.1
CO ₂ Output	1126	564.6	14.5	15.3	0.8	12.4	11.3	12.1	1	20.3	8.8	5.6		

TABLE 4.9 – Utilisations maximales des AIOs pour (Ctrl). La table présente les valeurs des quantités des nutriments prélevés par la mamelle, des nutriments dans le lait et les quantités minimales des entrées qui sont réparties dans les sorties. Par exemple, 1422 mmol/h de Glucose sont prélevées et 886 mmol/h de Lactose sont mesurées dans le lait. La quantité minimale du Glucose qui contribue dans le Lactose est de 886 pour le régime alimentaire (Ctrl). Le pourcentage de la contribution minimale du Glucose dans le Lactose est de $\frac{886}{886} = 100\%$.

Entrées	Glucose	Glycerol	Acetate	BHBA	Lysine	Isoleucine	Leucine	Valine	Arginine	Glutamine	
	1392	17.2	924	668	21.5	21.4	22.6	19.3	26.9	8.95	
Sorties	Utilisations Maximales des entrées dans les sorties (en mmol/h/démi-mamelle de Carbone) pour (CN)										
Glycerol3P	117	114.8	11.3	22.8	44.5	1.6	2.3	2	2	2.2	0
Lactose	1002	1002	9.2	37.9	74	2.6	3.9	3.3	3.3	3.7	0
C4	46.4	0	0	46.4	0	0	0	0	0	0	0
C6	33.5	0	0	22.3	11.2	0	0	0	0	0	0
C8	23	0	0	17.2	5.7	0	0	0	0	0	0
C10	64.6	0	0	51.7	12.9	0	0	0	0	0	0
C12	73.4	0	0	61.2	12.2	0	0	0	0	0	0
C14	250	0	0	214.7	35.8	0	0	0	0	0	0
C16	343	0	0	300.2	42.9	0	0	0	0	0	0
Glycine	6.9	5.5	0.5	1.3	2.6	0.1	0.1	0.1	0.1	0.1	0
Alanine	9.8	9.6	0.9	2.4	4.7	0.2	0.3	0.2	0.2	0.2	0
Glutamate	31.7	8.8	0.4	9	17.6	0.5	0.7	0.8	0.5	0.9	0
Proline	55	15.3	0.6	15.6	30.5	0.9	1.3	1.4	0.9	1.6	0
Aspartate	16.5	7	0.3	4.5	8.9	0.3	0.5	0.4	0.4	0.5	0
SerineOutput	22.5	22	2.2	4.4	8.5	0.3	0.4	0.4	0.4	0.4	0
CO ₂ Output	1021	402.1	12.5	227.4	443.6	19.8	19.1	20	17.4	19.2	0

TABLE 4.10 – Utilisations maximales des AIOs pour (CN). La table présente les valeurs des quantités des nutriments prélevés par la mamelle, des nutriments dans le lait et les quantités minimales des entrées qui sont réparties dans les sorties. Par exemple, 1392 mmol/h de Glucose sont prélevées et 1002 mmol/h de Lactose sont mesurées dans le lait. La quantité minimale du Glucose qui contribue dans le Lactose est de 1002 pour le régime alimentaire (CN). Le pourcentage de la contribution minimale du Glucose dans le Lactose est de $\frac{1002}{1002} = 100\%$.

Le calcul des AIO pour le jeu de données (HB) avait pour simple objectif de montrer que les méthodes développées s'appliquent à d'autres jeux de données que ceux issus des expérimentations de Lemosquet et al. [LRL⁺09, RLL⁺06]. De plus, l'emploi du jeu de

Entrées		Glucose	Glycerol	Acetate	BHBA	Lactate	Lysine	Isoleucine	Valine	Arginine	Alanine	Glutamine
		1526	2.06	768	604	1.44	23.9	190	45.6	65.8	6.56	7.50
Sorties		Utilisations Maximales des entrées dans les sorties (en mmol/h/démi-mamelle de Carbone) pour (HB)										
Glycerol3P	95	94.8	1.8	24.7	32.6	0.1	1.3	16.9	3.7	4.3	0.6	0.6
Lactose	953	952.5	0.9	64.4	85	0.3	3.8	36.3	8.5	9.8	1.3	1.3
C4	38	0	0	0	38.3	0	0	0	0	0	0	0
C6	23	0	0	15	7.5	0	0	0	0	0	0	0
C8	17	0	0	12.5	4.2	0	0	0	0	0	0	0
C10	40	0	0	31.7	7.9	0	0	0	0	0	0	0
C12	43	0	0	35.4	7.1	0	0	0	0	0	0	0
C14	131	0	0	112.5	18.7	0	0	0	0	0	0	0
C16	213	0	0	186.7	26.7	0	0	0	0	0	0	0
Glycine	10	8.5	0.1	2.5	3.3	0	0.1	1.6	0.4	0.4	0.1	0.1
Glutamate	24	8.4	0	7.5	9.9	0	0.3	3.7	0.7	1.3	0.1	0.2
Proline	70	24.3	0.1	21.9	29	0.1	1	10.8	2.1	3.9	0.3	0.5
Aspartate	21	9.7	0	5.9	7.9	0	0.3	4.2	0.9	1	0.2	0.1
Asparagine	2	0.9	0	0.6	0.7	0	0	0.4	0.1	0.1	0	0
SerineOutput	6	5.6	0.1	1.5	1.9	0	0.1	1	0.2	0.3	0	0
CO ₂ Output	1546	689.2	1.8	344.6	455.2	1.3	22.6	177.1	43	49.9	6.2	6.9

TABLE 4.11 – Utilisations maximales des AIOs pour (HB). La table présente les valeurs des quantités des nutriments prélevés par la mamelle, des nutriments dans le lait et les quantités minimales des entrées qui sont réparties dans les sorties. Par exemple, 1526 mmol/h de Glucose sont prélevées et 953 mmol/h de Lactose sont mesurées dans le lait. La quantité minimale du Glucose qui contribue dans le Lactose est de 953 mmol/h pour le régime alimentaire (HB). Le pourcentage de la contribution minimale du Glucose dans le Lactose est de $\frac{953}{953} = 100\%$.

données (HB) a été l'occasion de créer des modèles où les prélèvements mammaires d'acides gras longs (à 16 ou 18 carbones) à partir des triglycérides sanguins étaient renseignés. Ces acides gras longs étaient utilisés pour synthétiser les triglycérides du lait (sans être métabolisés) et une partie (10%) était oxydée (catabolisée en acetyl-CoA) [AALVM⁺on].

Par contre, le calcul des AIO pour (CN) a permis de suggérer une hypothèse pour la question de recherche qui suit traitée dans l'article [LRL⁺09].

Question de recherche pour le régime (CN) traitée dans [LRL⁺09]. Comment et pourquoi lorsqu'on augmente l'apport de protéines à des vaches laitières, elles produisent plus de lactose dans le lait ; sachant que le lactose est synthétisé par la glande mammaire et que le principal précurseur du lactose est le glucose ?

Les mesures effectuées dans Lemosquet et al. [LRL⁺09] (de prélèvements nets mammaires nets mammaires et de composition du lait) ne permettent pas de répondre à cette question. En effet, la technique de mesure des bilans d'organes (prélèvements nets de nutriments par un organe, et productions de nutriments) ne renseigne pas sur le métabolisme interne à l'organe.

Cependant les deux hypothèses suivantes ont été émises dans [LRL⁺09].

Hypothèses émises dans [LRL⁺09] pour la question de recherche ci-dessus du régime (CN) :

- Première hypothèse : l'augmentation du prélèvement mammaire de certains acides aminés indispensables (Lysine, Leucine, Isoleucine, Valine, semi indispensables : Arginine) entraîne une augmentation de leur catabolisme (CO₂) fournissant alors

plus d'énergie sous forme d'ATP à la glande mammaire. Ceci permettrait alors une réorientation de la répartition du glucose qui serait moins oxydé (CO_2 , ATP) et pourrait être utilisé en proportion plus importante pour synthétiser le lactose.

- Deuxième hypothèse : les acides aminés aminés prélevés en excès (Isoleucine, Valine) pourraient contribuer à la synthèse du lactose via la synthèse de galactose (représenté dans nos modèles par la contribution au glucose-6-P).

Cependant, le résultat des AIO permet de suggérer la première hypothèse comme hypothèse à valider dans le futur par une expérimentation. En effet si on compare les Tables 4.7 et 4.10 donnant respectivement les valeurs minimales et maximales pour le régime (CN) aux Tables 4.6 et 4.9 donnant respectivement les valeurs minimales et maximales des AIO pour le régime (Ctrl), on constate que pour des prélèvements mammaires de glucose non statistiquement différents (1396 vs. 1422 mmol/h/demi-mamelle, dans (CN) et (Ctrl)), que **le régime (Ctrl) est caractérisé par une plus forte oxydation du glucose en CO_2 par rapport à (CN)**. Car, $\frac{396}{1422} = 27.9\%$ à $\frac{565}{1422} = 39.7\%$ du Glucose est oxydé en CO_2 pour (Ctrl) alors qu'il est de $\frac{244}{1392} = 17.5\%$ à $\frac{402}{1392} = 28.9\%$ pour (CN).

On constate aussi que pour des quantités significativement plus importantes de lysine (16.1 vs. 21.5 mmol/h/demi-mamelle), leucine (12.1 vs. 22.6 mmol/h/demi-mamelle), isoleucine (12.1 vs. 22.6 mmol/h/demi-mamelle) et valine (12.7 vs. 19.3 mmol/h/demi-mamelle) à cataboliser (données renseignées comme entrées), une même proportion est oxydée en CO_2 pour une production de CO_2 non différente statistiquement entre (Ctrl) et (CN) (1126 vs. 1021 mmol/h/demi-mamelle) conduisant à des quantités de CO_2 produits à partir de ses acides aminés supérieures. Par exemple, la contribution minimale de la lysine au CO_2 dans (CN) est de 16.3 mmol/h/demi-mamelle (Table 4.7), la contribution maximale dans (Ctrl) est de 12.4 (Table 4.9). On remarque également une quantité un peu plus importante de carbone du lactose qui proviendrait de certains acides aminés dans le régime (CN). Par exemple, l'isoleucine contribuerait au maximum à 3.9 mmol/h/demi-mamelle de carbone de lactose (Table 4.10) dans le régime (CN) et au maximum à 2.6 mmol/h/demi-mamelle dans le traitement (Ctrl) (Table 4.9). Cependant ces petites différences ne permettent pas d'expliquer la différence significative de production de lactose (886 vs. 1002 mmol/h/ demi-mamelle entre (Ctrl) et (CN)).

On constate également que 85% ($\frac{751}{886}$ (Table 4.6) ou $\frac{866}{1002}$ Table 4.7) à 100% (Tables 4.9, 4.10) pour les régimes (Ctrl) et (CN). Cependant on constate qu'il existe des distributions de flux tels que le Glucose n'est pas unique précurseur du Lactose [BSEKO06] et ceci est conforme aux estimations de la proportion des carbones du Lactose venant Glucose publié dans Bickerstaffe et al. 1974 [BAL74].

On voit bien que la stratégie du calcul des contributions minimales et maximales des nutriments prélevés dans les constituants du lait permet d'explorer la flexibilité de glande mammaire et de discriminer des situations nutritionnelles différentes. Il s'agit de la conclusion de l'article accepté pour publication à *BMC System Biology* [AALVM⁺on].

4.4.2 Distances des solutions optimales par rapport au bord de l'espace de recherche

Les différentes distances des solutions optimales par rapport au bord de l'espace de recherche, en utilisant l'algorithme de recherche locale, pour les trois jeux de données sont données dans les Tables 4.12, 4.13, 4.14, 4.15, 4.16, 4.17. Ici également, la distance d'une solution optimale par rapport au bord de l'espace de recherche est définie comme étant la distance minimale entre la solution optimale et les différents hyperplans qui définissent l'espace de recherche. Les différentes valeurs trouvées montrent que les solutions optimales appartiennent au bord de l'espace de recherche. Ce que nous avons intuité en utilisant *fmincon* mais pas totalement prouvé.

Finalement, les coefficients des AIOs sont tous optimisés sur des points du bord de l'espace de recherche, mais pas aux sommets. Cela veut dire qu'il y a des points sur le bord de l'espace de recherche autres que les sommets qui ont une interprétation formelle (optimisation des rendements). Or, on a vu sur l'exemple de la glande mammaire comme sur différents exemples dans la littérature [SZZ⁺12, GBS⁺] que les sommets du simplexe peuvent ne pas être suffisants pour décrire la réponse d'un système métabolique. Il pourra être intéressant avec de nouvelles analyses de voir quelles sont les distributions qui optimisent les coefficients des AIOs, sur un modèle plus classique comme *E.coli* et voir s'ils ont un rapport avec les distributions identifiées par des approches multi-objectifs.

Entrées	Glucose	Glycerol	Acetate	BHBA	Lysine	Threonine	Isoleucine	Leucine	Valine	Histidine	Arginine	Alanine	Glutamine
Sorties	Distances des solutions maximales par rapport au bord de l'espace des solutions pour (Ctrl)												
Glycerol3P	0	5e-14	0	6e-14	6e-14	0	0	0	0	0	0	0	0
Lactose	0	0	0	6e-14	0	0	0	0	0	0	0	0	0
C4	0	6e-14	6e-14	1e-14	0	1e-14	1e-14	6e-16	6e-14	6e-14	0	0	0
C6	0	6e-14	6e-14	1e-14	0	1e-14	1e-14	6e-16	6e-14	6e-14	0	0	0
C8	0	6e-14	6e-14	1e-14	0	1e-14	1e-14	6e-16	6e-14	6e-14	0	0	0
C10	0	6e-14	6e-14	1e-14	0	1e-14	1e-14	6e-16	6e-14	6e-14	0	0	0
C12	0	6e-14	6e-14	1e-14	0	1e-14	1e-14	6e-16	6e-14	6e-14	0	0	0
C14	0	6e-14	6e-14	1e-14	0	1e-14	1e-14	6e-16	6e-14	6e-14	0	0	0
C16	0	6e-14	6e-14	1e-14	0	1e-14	1e-14	6e-16	6e-14	6e-14	0	0	0
Glycine	0	0	0	4e-14	0	0	0	1e-14	0	0	0	0	0
Glutamate	0	0	0	0	0	0	6e-14	0	0	0	0	0	0
Proline	0	0	6e-14	0	0	5e-14	6e-14	0	1e-14	6e-14	0	6e-14	5e-14
Aspartate	3e-14	0	0	0	0	0	5e-14	0	6e-14	0	0	0	5e-14
SerineOutput	0	0	4e-14	0	4e-14	0	0	0	0	0	6e-16	0	0
CO ₂ Output	0	0	5e-14	5e-14	0	5e-14	0	5e-14	0	5e-14	5e-14	0	5e-14

TABLE 4.12 – Distances des solutions maximales des AIOs pour (Ctrl) par rapport au bord de l'espace des solutions en utilisant l'algorithme de recherche locale. *Les valeurs des distances sont de l'ordre de 1e-14 avec algorithme de recherche locale alors qu'elles sont de l'ordre de 1e-6 avec la méthode des points intérieurs. L'algorithme de recherche locale a permis de conclure que les solutions maximales des AIOs pour (Ctrl) sont sur le bord de l'espace de recherche.*

Entrées	Glucose	Glycerol	Acetate	BHBA	Lysine	Threonine	Isoleucine	Leucine	Valine	Histidine	Arginine	Alanine	Glutamine
Sorties	Distances des solutions minimales par rapport au bord de l'espace des solutions pour (Ctrl)												
Glycerol3P	6e-14	0	5e-14	5e-14	5e-14	5e-14	5e-14	5e-14	5e-14	5e-14	0	5e-14	1e-13
Lactose	6e-14	4e-13	5e-14	0	5e-14	5e-14	5e-14	5e-14	5e-14	5e-14	5e-14	5e-14	5e-14
C4	0	0	0	0	5e-14	4e-13	0	5e-14	5e-14	0	4e-13	5e-14	5e-14
C6	0	0	0	0	5e-14	4e-13	0	5e-14	5e-14	0	4e-13	5e-14	5e-14
C8	0	0	0	0	5e-14	4e-13	0	5e-14	5e-14	0	4e-13	5e-14	5e-14
C10	0	0	0	0	5e-14	4e-13	0	5e-14	5e-14	0	4e-13	5e-14	5e-14
C12	0	0	0	0	5e-14	4e-13	0	5e-14	5e-14	0	4e-13	5e-14	5e-14
C14	0	0	0	0	5e-14	4e-13	0	5e-14	5e-14	0	4e-13	5e-14	5e-14
C16	0	0	0	0	5e-14	4e-13	0	5e-14	5e-14	0	4e-13	5e-14	5e-14
Glycine	0	0	0	0	5e-14	0	0	0	5e-14	0	5e-14	5e-14	0
Glutamate	0	6e-14	0	0	8.9809e-14	0	0	0	0	0	1e-13	0	0
Proline	0	0	5e-14	0	0	0	0	0	6e-14	0	0	0	0
Aspartate	0	0	0	0	9.4206e-15	0	5e-14	0	0	0	0	0	0
SerineOutput	0	0	5e-14	5e-14	0	5e-14	0	5e-14	5e-14	5e-14	0	5e-14	5e-14
CO ₂ Output	5e-14	0	0	0	0	0	0	0	0	0	6e-14	0	0

TABLE 4.13 – Distances des solutions minimales des AIOs pour (Ctrl) par rapport au bord de l'espace des solutions en utilisant l'algorithme de recherche locale. *Les valeurs des distances sont de l'ordre de 1e-14 avec algorithme de recherche locale alors qu'elles sont de l'ordre de 1e-6 avec la méthode des points intérieurs. L'algorithme de recherche locale a permis de conclure que les solutions minimales des AIOs pour (Ctrl) sont sur le bord de l'espace de recherche.*

Entrées	Glucose	Glycerol	Acetate	BHBA	Lysine	Isoleucine	Leucine	Valine	Arginine	Glutamine
Sorties	Distances des solutions maximales par rapport au bord de l'espace des solutions pour (CN)									
Glycerol3P	5e-14	0	0	2e-14	0	0	2e-14	0	0	0
Lactose	2e-14	3e-14	2e-14	0	0	2e-14	2e-14	0	0	2e-14
C4	2e-14	0	0	0	0	0	0	0	0	0
C6	2e-14	0	0	0	0	0	0	0	0	0
C8	2e-14	0	0	0	0	0	0	0	0	0
C10	2e-14	0	0	0	0	0	0	0	0	0
C12	2e-14	0	0	0	0	0	0	0	0	0
C14	2e-14	0	0	0	0	0	0	0	0	0
C16	2e-14	0	0	0	0	0	0	0	0	0
Glycine	5e-14	2e-14	0	0	2e-14	0	0	0	0	0
Alanine	2e-15	0	7e-15	1e-14	3e-14	1e-14	1e-14	7e-15	0	3e-14
Glutamate	2e-14	0	7e-15	7e-15	7e-15	1e-14	3e-14	7e-15	1e-14	7e-15
Proline	2e-14	0	0	3e-14	2e-14	2e-14	7e-15	7e-15	0	0
Aspartate	0	2e-14	0	3e-14	7e-15	7e-15	4e-14	0	0	0
SerineOutput	5e-14	2e-14	0	2e-14	0	0	0	2e-14	0	2e-14
CO ₂ Output	5e-14	0	0	0	0	0	0	2e-14	0	6e-14

TABLE 4.14 – Distances des solutions maximales des AIOs pour (CN) par rapport au bord de l'espace des solutions en utilisant l'algorithme de recherche locale. *Les valeurs des distances sont de l'ordre de 1e-14 avec algorithme de recherche locale alors qu'elles sont de l'ordre de 1e-6 avec la méthode des points intérieurs. L'algorithme de recherche locale a permis de conclure que les solutions maximales des AIOs pour (CN) sont sur le bord de l'espace de recherche.*

Entrées	Glucose	Glycerol	Acetate	BHBA	Lysine	Isoleucine	Leucine	Valine	Arginine	Glutamine
Sorties	Distances des solutions minimales par rapport au bord de l'espace des solutions pour (CN)									
Glycerol3P	0	0	7e-15	1e-14	3e-14	0	3e-14	3e-14	3e-14	2e-14
Lactose	0	4e-14	0	2e-14	2e-14	2e-14	4e-14	4e-14	2e-14	4e-14
C4	0	0	2e-14	2e-14	0	0	0	0	2e-14	0
C6	0	0	2e-14	2e-14	0	0	0	0	2e-14	0
C8	0	0	2e-14	2e-14	0	0	0	0	2e-14	0
C10	0	0	2e-14	2e-14	0	0	0	0	2e-14	0
C12	0	0	2e-14	2e-14	0	0	0	0	2e-14	0
C14	0	0	2e-14	2e-14	0	0	0	0	2e-14	0
C16	0	0	2e-14	2e-14	0	0	0	0	2e-14	0
Glycine	0	0	0	0	0	0	3e-14	7e-15	2e-14	7e-15
Alanine	1e-14	7e-15	0	2e-15	0	0	2e-15	2e-14	0	2e-14
Glutamate	1e-14	7e-15	2e-14	0	2e-14	0	2e-14	2e-14	2e-14	0
Proline	7e-15	7e-15	2e-14	2e-14	2e-14	0	2e-14	2e-14	0	0
Aspartate	1e-14	0	2e-14	2e-14	0	0	2e-14	2e-14	0	2e-14
SerineOutput	0	0	2e-14	3e-14	3e-14	7e-15	0	2e-14	1e-14	0
CO ₂ Output	1e-14	0	2e-14	2e-14	0	0	3e-14	2e-14	0	2e-14

TABLE 4.15 – Distances des solutions minimales des AIOs pour (CN) par rapport au bord de l'espace des solutions en utilisant l'algorithme de recherche locale. *Les valeurs des distances sont de l'ordre de 1e-14 avec algorithme de recherche locale alors qu'elles sont de l'ordre de 1e-6 avec la méthode des points intérieurs. L'algorithme de recherche locale a permis de conclure que les solutions minimales des AIOs pour (CN) sont sur le bord de l'espace de recherche.*

Entrées	Glucose	Glycerol	Acetate	BHBA	Lactate	Lysine	Isoleucine	Valine	Arginine	Alanine	Glutamine
Sorties	Distances des solutions maximales par rapport au bord de l'espace des solutions pour (HB)										
Glycerol3P	6e-15	3e-15	4e-15	3e-15	4e-15	2e-15	7e-15	9e-15	2e-15	4e-15	3e-15
Lactose	8e-15	0	0	3e-14	3e-14	3e-14	1e-14	0	0	4e-16	0
C4	4e-15	5e-15	0	2e-15	0	0	0	0	0	0	2e-15
C6	4e-15	5e-15	0	2e-15	0	0	0	0	0	0	2e-15
C8	4e-15	5e-15	0	2e-15	0	0	0	0	0	0	2e-15
C10	4e-15	5e-15	0	2e-15	0	0	0	0	0	0	2e-15
C12	4e-15	5e-15	0	2e-15	0	0	0	0	0	0	2e-15
C14	4e-15	5e-15	0	2e-15	0	0	0	0	0	0	2e-15
C16	4e-15	5e-15	0	2e-15	0	0	0	0	0	0	2e-15
Glycine	6e-15	4e-15	3e-15	1e-15	4e-15	4e-15	9e-15	2e-15	3e-15	5e-17	5e-16
Glutamate	3e-15	4e-16	0	7e-15	0	0	4e-15	1e-15	1e-14	0	7e-15
Proline	3e-15	3e-15	0	0	0	4e-15	0	1e-14	4e-16	0	1e-14
Aspartate	4e-16	7e-16	9e-15	0	2e-15	7e-15	0	0	0	9e-16	5e-15
Asparagine	3e-15	5e-15	0	6e-15	0	3e-15	3e-15	0	1e-14	0	3e-15
SerineOutput	3e-15	0	4e-16	6e-15	2e-16	1e-15	4e-15	9e-16	3e-15	5e-15	5e-15
CO ₂ Output	6e-15	3e-14	3e-15	1e-15	2e-15	4e-15	4e-15	2e-15	3e-15	5e-17	5e-16

TABLE 4.16 – Distances des solutions maximales des AIOs pour (HB) par rapport au bord de l'espace des solutions en utilisant l'algorithme de recherche locale. *Les valeurs des distances sont de l'ordre de 1e-14 avec algorithme de recherche locale alors qu'elles sont de l'ordre de 1e-6 avec la méthode des points intérieurs. L'algorithme de recherche locale a permis de conclure que les solutions maximales des AIOs pour (HB) sont sur le bord de l'espace de recherche.*

Entrées	Glucose	Glycerol	Acetate	BHBA	Lactate	Lysine	Isoleucine	Valine	Arginine	Alanine	Glutamine
Sorties	Distances des solutions minimales par rapport au bord de l'espace des solutions pour (HB)										
Glycerol3P	0	3e-14	9e-15	8e-15	3e-15	9e-15	3e-15	7e-15	8e-15	9e-15	7.4837e-15
Lactose	9e-15	9e-15	1e-14	8e-15	8.6036e-15	8.4387e-15	7.9989e-15	9.4282e-15	9e-15	1.0198e-14	1.0830e-14
C4	0	4e-16	3e-15	3e-15	3e-15	9e-16	3e-14	3e-15	0	0	4e-16
C6	0	4e-16	3e-15	3e-15	3e-15	9e-16	3e-14	3e-15	0	0	4e-16
C8	0	4e-16	3e-15	3e-15	3e-15	9e-16	3e-14	3e-15	0	0	4e-16
C10	0	4e-16	3e-15	3e-15	3e-15	9e-16	3e-14	3e-15	0	0	4e-16
C12	0	4e-16	3e-15	3e-15	3e-15	9e-16	3e-14	3e-15	0	0	4e-16
C14	0	4e-16	3e-15	3e-15	3e-15	9e-16	3e-14	3e-15	0	0	4e-16
C16	0	4e-16	3e-15	3e-15	3e-15	9e-16	3e-14	3e-15	0	0	4e-16
Glycine	2e-15	3e-14	26e-15	1e-15	9e-16	5e-15	3e-15	1e-15	3e-15	2e-15	6e-15
Glutamate	5e-15	0	3e-15	1.0441e-14	0	3e-15	0	0	3e-15	3e-14	3e-15
Proline	0	0	3e-15	4e-16	3e-14	2e-15	9e-15	0	3e-15	0	3e-15
Aspartate	2e-16	0	4e-16	3e-15	3e-15	3e-15	3e-15	3e-14	3e-14	0	3e-15
Asparagine	7e-15	0	3e-15	1e-14	0	3e-15	0	3e-15	0	9e-15	0
SerineOutput	2e-15	0	5e-15	6e-15	9e-15	7e-15	3e-15	6e-15	8e-15	6e-15	5e-15
CO ₂ Output	5e-15	0	3e-15	1e-14	0	3e-15	0	0	3e-15	3e-14	3e-15

TABLE 4.17 – Distances des solutions minimales des AIOs pour (HB) par rapport au bord de l'espace des solutions en utilisant l'algorithme de recherche locale. *Les valeurs des distances sont de l'ordre de $1e-14$ avec l'algorithme de recherche locale alors qu'elles sont de l'ordre de $1e-6$ avec la méthode des points intérieurs. L'algorithme de recherche locale a permis de conclure que les solutions minimales des AIOs pour (HB) sont sur le bord de l'espace de recherche.*

Chapitre 5

NutritionAnalyzer : un logiciel bio-informatique d'analyse de réseau métabolique

5.1 Contexte et motivations

Comme nous venons de le voir dans les chapitres précédents, les mathématiques et l'informatique interviennent de plus en plus dans le processus d'analyse des données d'observation et de mesure au sein des cellules.

Un exemple précurseur est connu depuis une dizaine d'année sous le nom de *Flux Balance Analysis*. Cette notion revient à étudier les distributions de flux des nutriments assimilés par l'organisme à l'état stationnaire qui se caractérise comme la solution d'un système d'équations linéaires. Le système est généralement sous contraint et donc admet une infinité de solutions qui a la forme d'un cône convexe. Les sommets du cône appelés *distributions extrémales* correspondent aux distributions qui optimisent des objectifs linéaires [Ker89]. Plusieurs algorithmes permettent de déterminer ces distributions extrémales par la résolution de problèmes d'optimisation linéaires dont l'algorithme du simplexe [Dan51]. Dans, [EP00, SCF⁺02], les auteurs ont montré que la bactérie E.coli a tendance à sélectionner les voies métaboliques qui optimisent la biomasse.

Une question biologique a priori indépendante qui intéresse fortement l'INRA (Institut National de Recherche en Agriculture) est l'influence de différents régimes alimentaires et traitements sur la qualité et la composition de différents organes animaux. En particulier, Sophie Lemosquet travaille à l'INRA UMR1080 production du lait de Saint Gilles (Bretagne) sur le métabolisme de la mamelle des vaches laitières. L'enjeu finalisé est de prédire la composition du lait. Il s'agit plus particulièrement d'étudier l'impact de l'équilibre des nutriments sur les régulations biochimiques de la synthèse des constituants du lait (lactose, protéines, acides gras) chez la vache laitière. La technique utilisée est celle des bilans mammaires [LRL⁺09] qui quantifie d'une part les prélèvements de nutriments précurseurs du lait par la mamelle et d'autre part la composition du lait. L'utilisation

du modèle de [VM02] permet d'étudier des scénarios de répartition d'utilisation des nutriments dans les différentes voies métaboliques de la glande mammaire dans le but de comprendre les règles de répartition des nutriments qui expliquent les modifications de composition du lait. D'autres équipes de l'Inra s'intéressent à d'autres constituants du lait (acides gras...), et/ou à d'autres espèces animales (porcs, caprins...), ou d'autres niveaux d'observation (cellule, tissu, organe ou organisme). Les expérimentations dans ces situations consistent souvent à proposer différents régimes aux animaux et à mesurer la composition finale en constituants d'intérêts pour identifier les stratégies d'alimentation répondant aux critères attendus de qualité. Par exemple, les expérimentations à l'Inra Saint Gilles consistent à proposer des régimes différents sur la quantité et la nature de l'énergie apportée et sur la quantité et la nature de protéines apportées chez les vaches laitières et à analyser les modifications de production et de qualité du lait produit. Cette qualité est estimée, par le calcul de rendements énergétiques, à l'aide de différents taux d'utilisation des nutriments dans les différentes voies biochimiques en rejetant les cycles futiles.

La modélisation faite dans le chapitre 2 a permis de comprendre que la stratégie d'utilisation des voies métaboliques par la technique des bilans mammaires revient alors à rechercher dans l'ensemble des solutions celle qui maximise le bilan d'ATP (maximisation d'une combinaison linéaire des inconnues), similaire à l'optimisation de biomasse mise en évidence chez les bactéries avec le *flux balance analysis* (FBA).

Plusieurs méthodes et outils ont été développées dans l'objectif de diversifier et d'intégrer plus de connaissances biologiques dans les analyses effectuées par le FBA. Le lecteur intéressé peut consulter les articles tels que OptFlux [RME⁺10] ou Fasimu [HHG⁺11].

Cependant aucune de ces méthodes n'offre, à notre connaissance, une approche permettant d'étudier la pertinence biologique d'une distribution extrême en calculant la contribution des nutriments prélevés (Inputs) dans les produits formés (Outputs) AIO. Cette approche est très utilisée dans le domaine de la nutrition. Ainsi, dans le chapitre 3, nous avons formalisé et modélisé l'approche sous forme de calcul de l'inverse d'une matrice. Dans ce chapitre nous présentons l'outil, *nutritionanalyzer*, qui permet non seulement de simuler le calcul de ces AIOs pour discuter de la pertinence biologique d'une distribution extrême mais aussi d'explorer de façon plus générale un réseau métabolique entrées/sorties et de comparer des situations alimentaires différentes.

5.2 Analyse d'un réseau métabolique par l'approche du calcul des AIOs

5.2.1 Les différentes approches

NutritionAnalyzer offre trois approches d'analyse du cône de l'espace de distribution des flux d'un réseau métabolique entrées/sorties présenté au chapitre 1. La première approche est une amélioration du *flux balance analysis*. On part d'un modèle stœchiométrique avec des données entrées/sorties, on choisit une fonction objectif d'optimisation (ATP,

biomasse etc.), on calcule une distribution extrême (un sommet du cône) correspondant à cette fonction objectif et enfin, on détermine les AIOs correspondant à cette distribution extrême. La deuxième approche consiste à déterminer toutes les distributions extrêmes (tous les sommets du cône) correspondant au modèle stœchiométrique et aux données entrées/sorties, à déterminer une fonction objectif associée à chaque distribution et à calculer les AIOs correspondantes à chaque distribution extrême. Comme présenté dans les chapitres précédents, l'objectif de ces deux approches est de chercher à comprendre si on peut prédire le fonctionnement d'un organisme par le FBA. Pour ces deux premières approches, nous utilisons un algorithme du simplexe ainsi que des approches probabilistes présentées en annexe A. En pratique, ceci est adapté en présence de jeux de données entrées/sorties qui restreignent fortement le nombre de comportements extrémaux possibles. La troisième et dernière approche est une extension de *flux variability analysis* qui permet d'explorer la flexibilité d'un réseau métabolique en résolvant des problèmes d'optimisation linéaires par le calcul des min/max des flux des réactions, introduits au chapitre 4. Cette approche quant à elle permet d'explorer la flexibilité d'un réseau métabolique entrées/sorties en résolvant des problèmes d'optimisation non linéaires par le calcul des min/max des AIOs correspondant à ce dernier.

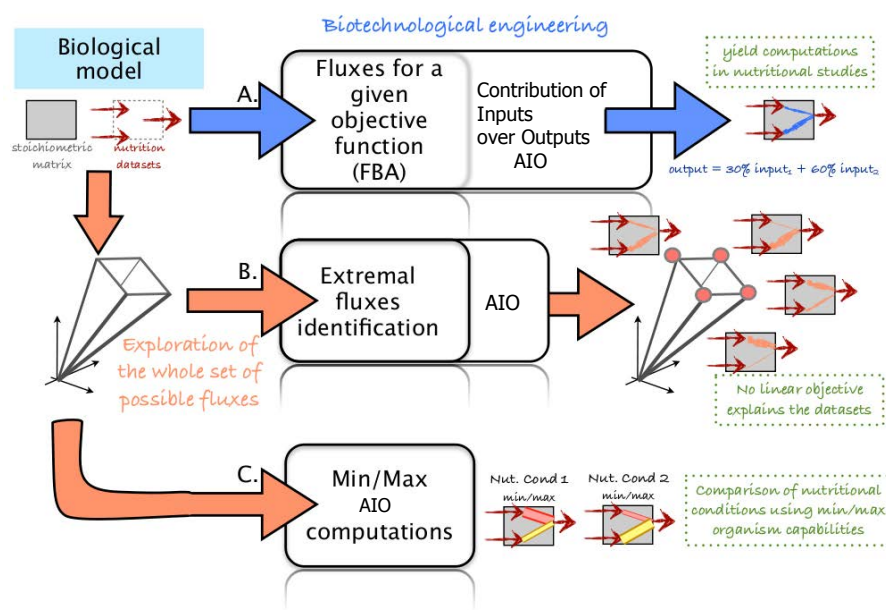


FIGURE 5.1 – Différentes approches d'analyse d'un réseau métabolique avec Nutritio-nAnalyzer A. Calcul des AIO d'une distribution extrême, B. Calcul des AIO de toutes les distributions extrêmes de l'espace de distribution, C. Comparaison de régimes alimentaires par le calcul de min/max des AIOs.

5.2.2 Calcul de AIO d'une distribution extrême

Etape1 : chargement du modèle stœchiométrique

Dans cette étape (Figure 5.3) l'utilisateur peut choisir un modèle prédéfini (simple input mode) qui est dans un format .mdl. Il peut également utiliser son propre modèle stœchiométrique sous un format .csv contenant l'ensemble des réactions métaboliques (les champs étant séparés par des points virgule). A la fin de chaque simulation, toutes les informations renseignées sont réécrites dans un fichier .mdl. Ce fichier peut être téléchargé pour une utilisation ultérieure. Un éditeur de texte permet également d'ajouter ou de modifier des réactions métaboliques comme le montre la Figure 5.2.



FIGURE 5.2 – Editeur de réactions métaboliques.

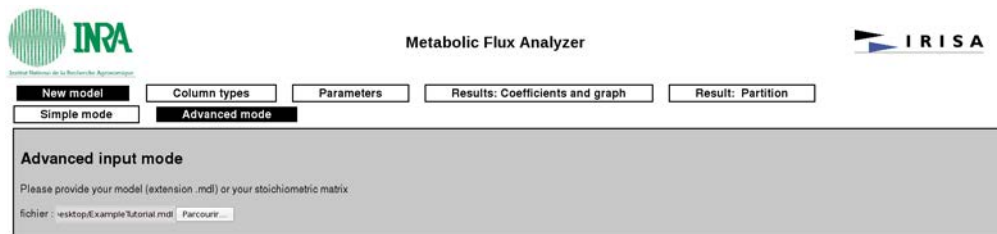


FIGURE 5.3 – Chargement du modèle stœchiométrique.

Etape2 : choix du type de chaque métabolite

Cette page consiste à renseigner le type de chaque métabolite (Figure 5.4). Un métabolite peut être :

- une entrée/sortie. On peut choisir de renseigner la valeur mesurée en sélectionnant (Measured as Input or Output) ou prédire l'entrée/sortie en sélectionnant (Predicted Input or Output)

- un cofacteur. On peut choisir que le bilan d'utilisation de ce cofacteur soit nul (c'est-à-dire que tout ce qui est produit est également consommé) ou bien la fixer à une valeur.
- un métabolite intermédiaire. Dans ce cas le bilan d'utilisation est fixé à zéro (c'est-à-dire tout ce qui est produit est également consommé)
- choisi comme un objectif à optimiser. Dans ce cas une distribution extrême permettant d'optimiser la combinaison linéaire des flux décrivant la consommation et la production de ce métabolite sera calculée à la fin de l'étape 4

Dans cette étape on doit également choisir l'élément chimique servant de base de calcul des AIOs (ça peut être du carbone, de l'azote etc.)

The screenshot shows the 'Metabolic Flux Analyzer' software interface. At the top, there are logos for INRA and IRISA. Below the logos are navigation buttons: 'New model', 'Column types', 'Parameters', 'Results: Coefficients and graph', and 'Result: Partition'. The main window is titled 'Analyze of information from stoichiometric matrix' and contains a table with the following columns: Equation, Reaction, and Intermediary metabolite. Each cell in the table has a dropdown menu. The 'Equation' column lists various reactions like ATP, NADHc, NADHm, FADH2, NADPH, CO2, O2, and NH3. The 'Reaction' column lists reactions like GLC, G6P, G3P, Glycerol3P, primerC2, lactose, C16, Leucine, peptide, and SERPivot. The 'Intermediary metabolite' column lists various metabolites like Glycerol3P, primerC2, lactose, C16, Leucine, peptide, and SERPivot. Each cell in the table has a dropdown menu with options like 'Measured as input or Output' and '0'.

FIGURE 5.4 – Choix du type de chaque métabolite.

Etape3 : choix des différentes réactions/renseignement des contraintes additionnelles

Etape3a : choix des différentes réactions

Dans ce premier onglet (Figure 5.5) de cette étape l'utilisateur peut cocher les différentes réactions qu'il veut utiliser dans sa simulation. Il peut également renseigner les valeurs des flux mesurés pour toutes les entrées/sorties

The screenshot shows the 'Metabolic Flux Analyzer' software interface. At the top, there are logos for INRA and IRISA. Below the logos are navigation buttons: 'New model', 'Column types', 'Parameters', 'Results: Coefficients and graph', and 'Result: Partition'. The main window is titled 'Choice of reactions' and contains a table with the following columns: Intermediates reactions, Choice inputs, Choice values, Output, and Choice values. The 'Intermediates reactions' column lists various reactions like 1 - GLC → G6P, 3 - G6P → GLC, 4 - lactose synthesis, 5 - G6P → G3P, 6 - G3P → PYR, 8 - G3P → G6P, 9 - PYR → ACoAm, and 10 - OAA + ACoAm → αKG. The 'Choice inputs' column lists various inputs like 2 - glucose input, 20 - C(16:0) oxydation, 21 - C(18:0) oxydation, 22 - C(16:1) oxydation, 23 - C(18:1) oxydation, 25 - fructose → G3P, 32 - cysteine catabolism, and 34 - typtophan catabolism. The 'Choice values' column lists various values like 237, 50, 62, 64, 71, 99, 100, and 101. The 'Output' column lists various outputs like 7 - G6P → GLC, 50 - urea synthesis, 62 - 3C(n)m-acylCoA + glycerol-3P → triglyceride, 64 - peptide hydrolysis, 71 - ACoAm → acetoacetate, 99 - lactose output, 100 - C(4:0) output, and 101 - C(6:0) output. The 'Choice values' column lists various values like 32.96, 73.8, 10.08, and 4.51.

FIGURE 5.5 – Choix des différentes réactions.

Étape3b : renseignement des contraintes additionnelles

Dans ce deuxième onglet (Figure 5.6) de cette étape l'utilisateur peut renseigner des contraintes linéaires pouvant lier les flux des différentes réactions intervenant dans la simulation. Il peut également choisir comme objectif d'optimisation une combinaison linéaire quelconque des flux des réactions intervenant dans la simulation.

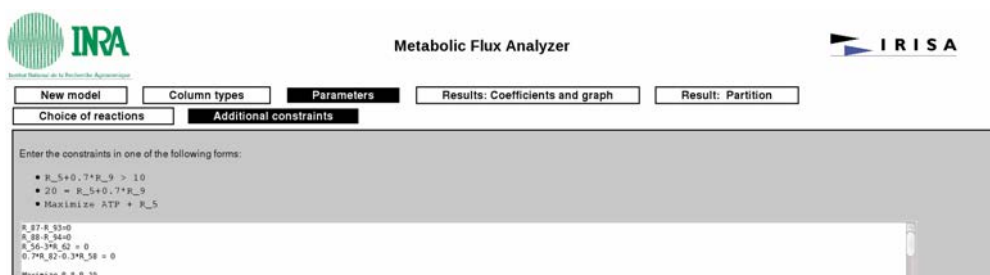


FIGURE 5.6 – Renseignement des contraintes additionnelles.

Étape4 : résultat du calcul d'une distribution extrême

Cette étape donne la distribution extrême correspondant aux différentes contraintes renseignées et à l'objectif linéaire d'optimisation choisi.

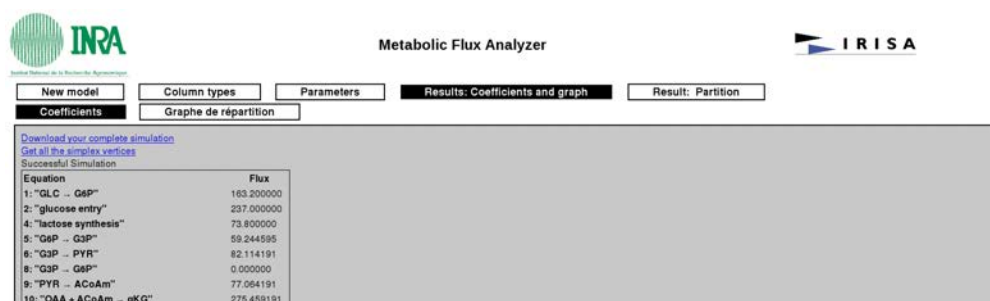


FIGURE 5.7 – Valeurs des flux de la distribution extrême.

Étape5 : calcul des AIOs de la distribution extrême calculée

Cette étape (Figure 5.8) donne le calcul des AIOs pour la distribution extrême calculée. En colonne nous avons les entrées et en ligne les sorties. Pour chaque colonne, les valeurs, correspondent à la contribution de l'entrée correspondante dans les différentes sorties

Distribution computation for ""carbones""

Explicit relations

```

X_O2 =
X_Glycerol3P = +0.066 X_GLC +0.265 X_Glycerol
X_Lactose = +0.029 X_GLC
X_C4 = +0.120 X_BHBA
X_C6 = +0.033 X_Acetate +0.027 X_BHBA
X_C8 = +0.033 X_Acetate +0.033 X_BHBA
X_C10 = +0.027 X_Acetate +0.028 X_BHBA
X_C12 = +0.041 X_Acetate +0.022 X_BHBA
X_C14 = +0.164 X_Acetate +0.093 X_BHBA
X_C16 = +0.254 X_Acetate +0.112 X_BHBA
X_PEP154 =
X_Glycine = +0.006 X_GLC +0.024 X_Glycerol +0.129 X_Threonine +0.086 X_Histidine
X_Glutamate = +0.000 X_GLC +0.001 X_Glycerol +0.001 X_Acetate +0.001 X_BHBA +0.001 X_Lys +0.001 X_Threonine +0.002 X_Isoleucine +0.002 X_Leucine +0.002 X_Valine +0.002 X_Histidine +0.002 X_Arginine +0.002 X_Alanine +0.003 X_Glutamine
X_Proline = +0.007 X_GLC +0.027 X_Glycerol +0.050 X_Acetate +0.033 X_BHBA +0.027 X_Lys +0.026 X_Threonine +0.025 X_Isoleucine +0.025 X_Leucine +0.043 X_Valine +0.025 X_Histidine +0.025 X_Arginine +0.040 X_Alanine +0.066 X_Glutamine
X_Aspartate = +0.002 X_GLC +0.007 X_Glycerol +0.000 X_Acetate +0.000 X_BHBA +0.000 X_Lys +0.007 X_Threonine +0.020 X_Isoleucine +0.014 X_Leucine +0.017 X_Valine +0.014 X_Histidine +0.014 X_Arginine +0.011 X_Alanine +0.017 X_Glutamine
X_Pararagine = +0.002 X_GLC +0.010 X_Glycerol +0.009 X_Acetate +0.011 X_BHBA +0.015 X_Lys +0.010 X_Threonine +0.027 X_Isoleucine +0.020 X_Leucine +0.024 X_Valine +0.020 X_Histidine +0.020 X_Arginine +0.020 X_Alanine +0.024 X_Glutamine
X_SERoutput = +0.034 X_GLC +0.005 X_Glycerol +0.270 X_Threonine +0.100 X_Histidine

```

Raw distribution

OUTIN	GLC	peptide	Glycerol	Acetate	BHBA	Lys	Threonine	Isoleucine	Leucine	Valine	Histidine	Arginine	Alanine	Glutamine
O2														
Glycerol3P	94.24		4.64											
lactose	885.60													
C4					40.32									
c6				18.04	9.02									
c8				13.38	4.46									
c10				37.28	9.32									
c12				42.50	8.46									
c14				166.80	27.80									
c16				263.48	37.64									
peptide														
glycine	9.11		0.45				0.21				0.13			
glutamate	0.36		0.02	1.01	0.42	0.02	0.00	0.03	0.03	0.02	0.00	0.06	0.01	0.02
Proline	0.47		0.47	28.50	11.01	0.61	0.05	0.74	0.71	0.57	0.08	1.47	0.38	0.44
aspartate	2.60		0.13	8.82	2.75	0.15	0.01	0.25	0.18	0.22	0.02	0.37	0.11	0.11
pararagine	3.71		0.18	9.46	3.92	0.22	0.02	0.36	0.25	0.32	0.03	0.53	0.15	0.16
SERoutput	19.63		0.97				0.46				0.28			

FIGURE 5.8 – Table des AIOs de la distribution extrême calculée.

5.2.3 Architecture des différents fichiers pour l'analyse d'un réseau métabolique entrées/sorties par les AIOs

A partir d'une simulation donnée (modèle stœchiométrique + données entrées/sorties) deux types de fichiers sont générés : un fichier *Modele.txt* et un fichier *Matrice.sage* (comme le montre la Figure 5.9).

Le fichier *Modele.txt* contient les différentes contraintes décrivant le cône mentionné ci-dessus. Quant au fichier *Matrice.sage*, il contient les matrices des AIOs ($D_1[v]$ et $D_2[v]$) écrites en fonction des variables v de l'espace de distributions.

Utilisation du fichier *Modele.txt* : le fichier *Modele.txt* est utilisé pour générer l'ensemble des distributions extrêmes. Comme expliqué dans le chapitre 2, les sommets extrêmes de l'espace de distribution des flux pour les trois jeux de données (Ctrl), (CN) et (HB) sont calculés en utilisant l'algorithme de tirage aléatoire présenté en annexe A. Il consiste à se donner un nombre de tirages aléatoires de fonctions objectifs à effectuer. Pour chaque tirage de fonction objectif, on exécute l'algorithme du simplexe pour obtenir un sommet extrême. Les différents sommets extrêmes sont déterminés au cours des tirages successifs. L'ensemble des sommets extrêmes est déterminé après avoir testé plusieurs nombres de tirage aléatoire.

On peut par la suite calculer les AIOs correspondant à chaque distribution extrême en utilisant le logiciel *nutritionanalyzer.org* après avoir associé à chaque distribution extrême une fonction objectif qui permet de l'identifier comme présenté dans la Table 2.3 du chapitre 2.

Il faut noter que pour générer l'ensemble des distributions extrêmes n'importe quelle

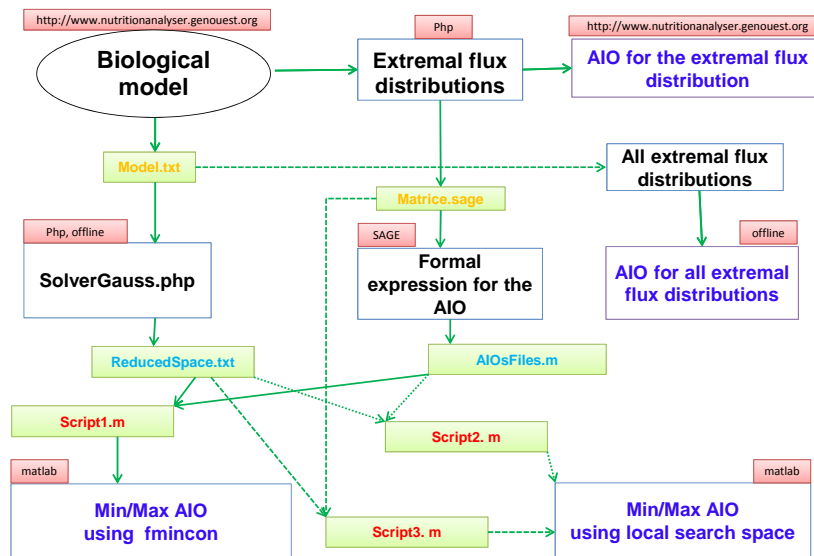


FIGURE 5.9 – Architecture des différents fichiers de calcul des AIOs

autre méthode de FBA existante peut être utilisée.

Ce fichier (*Modele.txt*) est également utilisé par un solveur de système d'équations linéaires (pivot de Gauss) pour décrire le cône par un ensemble minimal d'équations. Ce solveur permet de générer un fichier *ReducedSpace.txt*.

Utilisation du fichier *Matrice.sage* : le fichier *Matrice.sage* est utilisé par un logiciel de calcul formel (sagemath [SO12]) pour générer des fichiers matlab, *AIOsFiles.m*, contenant des fonctions non linéaires des variables v de l'espace de distribution correspondant aux expressions analytiques des AIOs (comme le montre la Figure 5.9).

Chaque fichier *AIOsFiles.m* peut être exécuté dans l'espace de recherche *ReducedSpace.txt* par un programme matlab (*Script1.m*) pour calculer les AIOs avec des solveurs de résolution de problèmes d'optimisation non linéaires existant dans la fonction *fmincon* de matlab (*interior-point*, *sqp*, *active-set*, etc.)

Chaque fichier *AIOsFiles.m* peut être exécuté dans l'espace de recherche *ReducedSpace.txt* par un programme matlab (*Script2.m*) pour calculer les AIOs par l'algorithme de recherche locale présenté au chapitre 4.

Enfin, lorsque les expressions analytiques des AIOs ne sont pas disponibles le fichier, *Matrice.sage*), peut être directement exécuté dans l'espace de recherche *ReducedSpace.txt* par un programme matlab (*Script3.m*) pour calculer les AIOs par l'algorithme de recherche locale présenté chapitre 4.

Chapitre 6

Conclusion et perspectives

6.1 Conclusion

Dans cette thèse, nous avons formalisé et modélisé une approche pour analyser le métabolisme dans les réseaux métaboliques entrées/sorties (modèle stœchiométrique et des données entrées/sorties) : contribution des entrées dans les sorties (AIO). Cette approche se définit comme une amélioration du *flux balance analysis*. L'application de cette approche à l'analyse du métabolisme de la glande mammaire a permis de montrer que son métabolisme ne peut pas être modélisé par l'optimisation unique d'une combinaison linéaire des flux des réactions.

Pour comprendre le fonctionnement global de la glande mammaire et comparer des régimes alimentaires différents, nous avons introduit une méthode permettant d'étudier la variabilité des AIOs dans l'espace de distribution des flux. Dans l'esprit, il s'agit, comme pour les techniques de *flux variability analysis*, de faire abstraction d'une hypothèse d'optimisation pour le fonctionnement du système. L'étude de cette variabilité nécessite de résoudre des problèmes d'optimisation non linéaires. Ainsi, nous avons proposé un algorithme efficace permettant de calculer les valeurs minimales et maximales des AIOs et avoir une vision complète des comportements possibles de la glande mammaire pour un régime alimentaire donné.

L'étude des contributions minimales et maximales des nutriments prélevés dans les constituants du lait a permis de comparer le métabolisme de la glande mammaire dans deux situations nutritionnelles différentes (Ctrl et CN). Cette comparaison a permis de suggérer une hypothèse à une question biologique posée dans [LRL⁺09] à savoir : comment et pourquoi lorsqu'on augmente l'apport de protéines à des vaches laitières, elles produisent plus de lactose dans le lait ; sachant que le lactose est synthétisé par la glande mammaire et que le principal précurseur du lactose est le glucose ?

Les mesures effectuées dans Lemosquet et al. [LRL⁺09] (de prélèvements nets mammaires nets mammaires et de composition du lait) ne permettent pas de répondre à cette question. En effet, la technique de mesure des bilans d'organes (prélèvements nets de nutriments par un organe, et productions de nutriments) ne renseigne pas sur le métabolisme interne à l'organe.

Cependant, deux hypothèses étaient émises dans [LRL⁺09] : première hypothèse, l'augmentation du prélèvement mammaire de certains acides aminés indispensables (Lysine, Leucine, Isoleucine, Valine, semi indispensables : Arginine) entraîne une augmentation de leur catabolisme (CO_2) fournissant alors plus d'énergie sous forme d'ATP à la glande mammaire. Ceci permettrait alors une réorientation de la répartition du glucose qui serait moins oxydé (CO_2 , ATP) et pourrait être utilisé en proportion plus importante pour synthétiser le lactose. deuxième hypothèse, les acides aminés aminés prélevés en excès (Isoleucine, Valine) pourraient contribuer à la synthèse du lactose via la synthèse de galactose (représenté dans nos modèles par la contribution au glucose-6-P).

Le résultat des AIO a permis de suggérer la première hypothèse comme hypothèse à valider dans le futur par une expérimentation. En effet on a constaté que **le régime (Ctrl) est caractérisé par une plus forte oxydation du glucose en CO_2 par rapport à (CN).**

De façon générale, le logiciel *NutritionAnalyzer* qui a été développé dans cette thèse permet bien d'analyser des hypothèses concernant le métabolisme interne au système considéré (ici un organe la glande mammaire) sans mesures propres de ce métabolisme. Il est un moyen d'exploiter de façon plus approfondies des données difficiles à collecter. C'est un moyen de mieux cibler les expérimentations futures à effectuer en cernant mieux les hypothèses à privilégier. Les outils (interface) et les méthodes (AIO) développées s'appliqueraient très facilement sur d'autres mammifères et à l'étude d'autre organe comme par exemple le foie.

6.2 Perspectives

Comme nous l'avons vu l'algorithme de recherche locale que nous avons proposé est une approche permettant de calculer les valeurs minimales et maximales des AIOs sans avoir besoin d'une formule analytique pour les fonctions à optimiser contrairement aux méthodes existantes pour lesquelles une formule analytique est nécessaire. Cependant, l'algorithme prend assez de temps de calcul mais ce temps peut être amélioré en parallélisant la recherche de solution optimale par découpage de l'espace de recherche. Ceci permettra également de calculer les valeurs minimales et maximales des AIOs sans utiliser le logiciel payant *matlab*. C'est à dire les calculs seront exécutés en utilisant exclusivement le langage de programmation libre *php*.

Nous avons également vu que le calcul des valeurs minimales et maximales des AIOs a permis de montrer que le régime alimentaire (Ctrl) est caractérisé par une plus forte oxydation du glucose en CO_2 par rapport à (CN) indépendamment du choix d'une fonction objectif linéaire à optimiser. C'est-à-dire les min/max des AIOs permettent d'avoir une vision complète et globale du fonctionnement de la glande mammaire. Ainsi, une valorisation de ce travail de recherche serait son utilisation à privilégier des régimes alimentaires polluant moins l'environnement.

D'autres part, les résultats auxquels nous sommes arrivés avec l'algorithme de recherche locale ont montré que les coefficients des AIOs sont tous optimisés sur des points du bord de l'espace de recherche, mais pas aux sommets. Cela veut dire qu'il y a des points sur le bord de l'espace de recherche autre que les sommets qui ont une interprétation formelle (optimisation des rendements). Or, des études sont entrain de montrer que les sommets de l'espace de recherche peuvent ne pas être suffisants pour décrire la réponse d'un système métabolique. Et, de plus en plus des approches multi-objectifs sont entrain d'être utilisées pour analyser ces systèmes métaboliques bactériens [SZZ⁺12, GBS⁺]. Ainsi, il pourra être intéressant de voir si les résultats obtenus à l'échelle bactérienne se vérifient à l'échelle d'un organe. Pour cela, on pourra utiliser le calcul des AIOs pour explorer ces modèles bactériens et voir si les distributions qui optimisent les coefficients des AIOs ont un rapport avec les distributions identifiées par les approches multi-objectifs.

Table des figures

1.1	Action d'une enzyme sur un substrat pour former deux produits.	6
1.2	Glycolyse : un exemple de voie métabolique qui transforme le glucose en Acetyl-CoA pour produire de l'énergie sous forme d'ATP.	7
1.3	Organismes dont le métabolisme a été reconstruit depuis 1999	8
1.4	Exemple de réseau métabolique.	8
1.5	Différentes approches de modélisation des réseaux métaboliques	11
1.6	Fonction cinétique monotone croissante	12
1.7	Modélisation des vitesses des réactions pour un réseau métabolique de la glande mammaire [Han94].	13
1.8	Exemple de modes élémentaires et voies extrêmes sur un réseau métabolique [PPW ⁺ 03]. <i>Modes élémentaires et voies extrêmes sur un exemple simple de réseau métabolique. Il est constitué de trois métabolites, trois réactions internes et trois externes (a). Le réseau admet quatre modes élémentaires (c) et trois voies extrêmes (b). La différence entre les deux ensembles est liée à l'utilisation de la réaction réversible au niveau du métabolite A. Cet exemple démontre la principale différence entre les deux approches.</i>	15
1.9	Exemple de dominance.	17
2.1	Réseau métabolique complet de la glande mammaire correspondant aux trois jeux de données (Ctrl), (CN) et (HB).	26
2.2	Réseau métabolique simplifié de la glande mammaire correspondant aux trois jeux de données (Ctrl), (CN) et (HB).	27
2.3	Exemple d'utilisation de la caleulette métabolique.	29
2.4	Espace de distribution de flux à ATP non fixé.	30
2.5	Espace de distribution de flux à ATP fixé à 1250 mmol/h/demi-mamelle.	32
3.1	Réseau métabolique transformant deux nutriments d'entrées I_1 et I_2 pour produire deux sorties O_1 et O_2	38
3.2	Connaissant $v_{\mathcal{I}} = (5, 10)$ le vecteur des flux des réactions d'entrées et $v_{\mathcal{O}} = (2.5, 10)$ le vecteur des flux des réactions de sorties, l'état stationnaire permet de déterminer les vecteurs $v = (v_3, v_4, v_5, v_6, v_7, v_8)$ des flux des autres réactions pour que tout ce qui est créé pour les métabolites intermédiaires P_1, P_2 et P_3 soit également consommé.	42

3.3	Exemple de vecteur de flux v qui équilibre P_1 , P_2 et P_3 . Avec $v_I = (5, 10)$ le vecteur des flux des réactions d'entrées et $v_O = (2.5, 10)$ le vecteur des flux des réactions de sorties, $v = (5, 5, 5, 15, 2.5, 10)$ permet d'équilibrer les métabolites intermédiaires P_1 , P_2 et P_3 en tenant compte de la stœchiométrie du réseau.	43
3.4	Réseau métabolique avec cycle transformant deux nutriments d'entrées I_1 et I_2 pour produire deux sorties O_1 et O_2	46
4.1	Centre de Chebychev.	67
5.1	Différentes approches d'analyse d'un réseau métabolique avec NutritionAnalyzer A. Calcul des AIO d'une distribution extrême, B. Calcul des AIO de toutes les distributions extrêmes de l'espace de distribution, C. Comparaison de régimes alimentaires par le calcul de min/max des AIOs.	83
5.2	Editeur de réactions métaboliques.	84
5.3	Chargement du modèle stœchiométrique.	84
5.4	Choix du type de chaque métabolite.	85
5.5	Choix des différentes réactions.	85
5.6	Renseignement des contraintes additionnelles.	86
5.7	Valeurs des flux de la distribution extrême.	86
5.8	Table des AIOs de la distribution extrême calculée.	87
5.9	Architecture des différents fichiers de calcul des AIOs	88

Annexe A

Générer l'ensemble des sommets extrémaux d'un polyèdre

Générer l'ensemble des sommets extrémaux d'un polyèdre est un problème mathématique difficile à résoudre. Dans le pire des cas un polyèdre défini par n inconnues et m contraintes linéaires avec $n > m$ admet $C_n^m = \frac{n!}{m!(n-m)!}$ sommets, même si dans la pratique on en a beaucoup moins.

Plusieurs travaux ont été réalisés pour générer des sommets extrémaux dans le cas borné [Bal61, Mat73, DP77]

Pour traiter le cas non borné, nous présentons ici une approche probabiliste qui génère aléatoirement des fonctions objectives à optimiser en utilisant l'algorithme du simplexe.

A.1 Algorithme du simplexe

L'algorithme du simplexe est un algorithme de résolution de programmes linéaires. Il a été développé par Georges dantzig [Dan51] dans les années 50 à l'université de Princeton à New Jersey et a marqué le début de l'ère de l'optimisation moderne.

Il a été montré que lorsqu'un programme linéaire admet une solution, cette solution se trouve sur un sommet de l'espace des solutions (qui est un polyèdre convexe) [Ker89]. L'algorithme du simplexe part d'un sommet à un autre en améliorant l'objectif, d'où la nécessité de disposer d'un sommet initial réalisable pour appliquer l'algorithme. Plusieurs méthodes existent dans la littérature pour l'initialisation du simplexe. La méthode des deux phases que nous présenterons ici est la plus connue pour l'initialisation.

La force du simplexe est qu'on arrive à trouver le point optimal exact en un nombre d'itérations fini et si le point optimal n'existe pas ou que la fonction objectif n'est pas bornée alors la méthode arrive à déceler ces deux cas aussi.

A.1.1 Définitions et Notations

Considérons le programme linéaire suivant :

$$(PL) \begin{cases} \text{Minimiser } c^t x \\ \text{s.c.} \\ Ax = b \\ x \geq 0 \end{cases} \quad (\text{A.1})$$

où c et x sont deux vecteurs de \mathbb{R}^n , $b \in \mathbb{R}^m$ et A une matrice de dimensions $m \times n$. Le système (A.1) est appelé forme *standard* d'un programme linéaire. On suppose que $m < n$.

- On appelle ensemble de solutions réalisables l'ensemble $K = \{x \in \mathbb{R}^n : Ax = b \text{ et } x \geq 0\}$.
- Une base de A est une sous matrice B formée de m colonnes linéairement indépendantes de A .
- Soit B une base de A , le point $x = (x_B, x_N) \in \mathbb{R}^n$, solution de l'équation $Ax = b$ et tel que : $x_B = B^{-1}b$, $x_N = 0$ est appelée solution de base, où N est une matrice de dimension $m \times n - m$. On note B_x la base associée à x .
- Une solution de base $x = (x_B, x_N)$ telle que $x_B \geq 0$ est dite solution de base réalisable.
On pose : $\beta_x = \{i \in \{1, 2, \dots, m\} : A^i \in B_x\}$ et $I_x = \{i \in \{1, 2, \dots, n\} : x_i > 0\}$ cet ensemble sera le plus souvent confondu à la base B_x elle même.
- Un point $x \in K$ est un sommet si et seulement s'il est une solution de base réalisable.
- Un sommet est dit non dégénéré si $x_B > 0$. Il est dit dégénéré dans le cas contraire (au moins une composante de x_B est nulle)

A.1.2 Présentation de l'algorithme

L'algorithme du simplexe évolue sur la frontière du domaine réalisable de sommet en sommet adjacent, en réduisant l'objectif jusqu'à l'optimum. Il tient compte systématiquement des résultats établis précédemment. Ainsi pour $j \in \{1, 2, \dots, n\}$

- chaque colonne de la matrice A^j , se transforme successivement en $A^{j,k} = (\alpha_1^{j,k}, \alpha_2^{j,k}, \dots, \alpha_m^{j,k})^t$ d'une base B_{x_k} à une autre ($k \in \{1, 2, \dots, \}$) avec $A^{j,0} = A^j$ et $(\alpha_1^{j,0}, \alpha_2^{j,0}, \dots, \alpha_m^{j,0}) = (\alpha_1^j, \alpha_2^j, \dots, \alpha_m^j)$. $\alpha_i^{j,k}$ sont les coordonnées de $A^{j,k}$ dans la base β_{x_k} (c'est-à-dire $A^{j,k} = \sum_{i \in \beta_{x_k}} \alpha_i^{j,k} A^{i,k}$).
- Chaque coût marginal, $Z_j = c_j$ de la fonction objectif, se transforme successivement en $Z_{j,k} = Z_{j,k} - \sum_{i \in \beta_{x_k}} \alpha_i^{j,k} c_{i,k}$ avec $Z_{j,0} = Z_j$.

Un critère d'optimalité simple permet de reconnaître le sommet optimal. Le nombre de sommet étant fini, l'algorithme ainsi défini converge en un nombre fini d'itérations n'excédant pas le nombre $C_n^m = \frac{n!}{m!(n-m)!}$, sous l'hypothèse que tous les sommets visités sont non dégénérés. Dans le cas dégénéré, l'algorithme risque de cycler, cependant il existe des techniques convenables pour éviter ce phénomène ([DOW55], [Bla77]). Cependant ceci est rare en pratique. La méthode des deux phases du simplexe permettant de résoudre (A.1) se présente comme suit

- **Phase 1 - Initialisation** : Trouver une solution de base réalisable (ou bien détecter l'impossibilité)
- **Phase 2 - Progression** : On passe d'un sommet à un sommet adjacent pour diminuer la fonction objectif (ou bien détecter que la fonction objectif n'est pas minorée)

A.1.2.1 L'algorithme du simplexe proprement dit : la phase 2

L'algorithme 1 ci-dessous donne les différentes étapes du simplexe

A.1.2.2 Initialisation : la phase 1

Si les contraintes peuvent se mettre sous la forme $Ax \leq b, x \geq 0$ (appelée forme *canonique*), on peut facilement déterminer une solution de base réalisable dans le cas où $b \geq 0$. En effet, sous forme standard les contraintes deviennent $Ax + e = b$, avec $x, e \geq 0$ où e désignent les variables d'écart. Une solution de base réalisable évidente est alors $x = 0, e = b \geq 0$. Cependant, pour un (PL) sous forme standard, il n'y a pas toujours de solution de base réalisable évidente. On va voir comment contruire des solutions de base réalisable : c'est la phase d'initialisation du simplexe. On considère le (PL) sous forme standard, pour obtenir une solution de base réalisable ou bien pour détecter l'impossibilité, on introduit un problème de programmation linéaire auxiliaire pour des variables supplémentaires appelées *variables artificielles*.

Soit $a = (a_1, a_2, \dots, a_m)$ les variables artificielles et on considère le programme linéaire suivant :

$$(PLA) \begin{cases} \text{Minimiser } \sum_{i=1}^m a_i \\ \text{s.c.} \\ Ax + a = b \\ x \geq 0 \\ a \geq 0 \end{cases} \quad (A.2)$$

On a la propriété évidente suivante.

Proposition A.1.1. *Un (PL) admet une solution réalisable si et seulement si le problème auxiliaire (PLA) admet une solution de base optimale avec $a = 0$.*

Ce résultat fournit un procédé pratique pour obtenir une solution de base réalisable pour le problème (PL). On applique l'algorithme du simplexe au problème auxiliaire (PLA). A la fin du simplexe, le coût minimal est nul sinon on a détecté l'impossibilité pour (PL). A la fin du simplexe sur (PLA), si tout s'est déroulé normalement (coût nul), on cherche à éliminer de la base toutes les variables artificielles.

Algorithme 4 Phase 2 du simplexe

Entrées : x_0 (sommet initial), β_{x_0} (base initiale), c^t (fonction objectif)

Sorties : x_{opt} (sommet optimal)

$x_0 \leftarrow x$

$\beta_{x_0} \leftarrow \beta_x$

Si ($Z_j \geq 0 \forall j \notin \beta_x$) **Alors**

$x \leftarrow x_{opt}$

Fin Si

Si ($\exists j \notin \beta_x$ tel que $Z_j < 0$) **Alors**

On pose $J = \{j \notin \beta_x : Z_j < 0\} \neq \emptyset$

Si ($\exists j \in J$ tel que $\alpha_i^j \leq 0 \forall i \in \beta_x$) **Alors**

(A.1) admet une solution infinie

Fin Si

Si ($\forall j \in J \exists i \in \beta_x$ tel que $\alpha_i^j > 0$) **Alors**

Si ($\exists j^+ \in J$ tel que $\{i \in \beta_x : \alpha_i^{j^+} > 0\} \subset I_x$) **Alors**

On pose $\theta^{j^+} = \min\{\frac{x_i}{\alpha_i^{j^+}} : i \in \beta_x, \alpha_i^{j^+} > 0\}$, $\theta^{j^+} > 0$

Soit $j^- \in \beta_x$ tel que $\alpha_{j^-}^{j^+} > 0$ et $\theta^{j^+} = \frac{x_{j^-}}{\alpha_{j^-}^{j^+}}$

Soit

$$\bar{x} = \begin{cases} \bar{x}_i = x_i - \theta^{j^+} \alpha_i^{j^+} & \text{si } i \in \beta_x \\ \bar{x}_{j^+} = \theta^{j^+} \\ \bar{x}_i = 0 & \text{si } i \notin \beta_x \cup j^+ \end{cases}$$

\bar{x} est un sommet de K différent de x tel que $c^t \bar{x} < c^t x$ et

$\beta_{\bar{x}} = (\beta_x \setminus \{j^-\}) \cup \{j^+\}$ est une base réalisable associée à \bar{x}

$\bar{x} \leftarrow x$

$\beta_{\bar{x}} \leftarrow \beta_x$

Sinon

i.e. $\forall j \in J \exists i \in \beta_x$ tel que $\alpha_i^j = 0$ et $x_i = 0$

Soit j^+ un indice quelconque dans J et soit $j^- \in \beta_x$ tel que $\alpha_{j^-}^{j^+} > 0$ et $x_{j^-} = 0$

$\beta'_x = (\beta_x \setminus \{j^-\}) \cup \{j^+\}$ est une base réalisable, différente β_x , associée au même sommet x

$\beta'_x \leftarrow \beta_x$

Fin Si

Fin Si

Fin Si

A.2 Algorithme de calcul sommets d'un polyèdre

L'algorithme 2 ci-dessous permet de déterminer les différents sommets extrémaux en générant aléatoirement un objectif linéaire d'optimisation

Algorithme 5 Calcul sommets d'un polyèdre

Entrées : x_{opt}^0 (sommets initial), β_{x_0} (base initiale), nombreDeTirage

Sorties : X (ensemble de sommets différents)

Initialisation : $X = \{x_{opt}^0\}$, $n = 1$

Pour i allant de 1 à nombreDeTirage **Faire**

 générer c_i^t (une direction aléatoire)

 calculer x_{opt}^i correspondant c_i^t avec l'algo précédent

$|X| = 0$ (compteur du nombre d'éléments de X visité pour chaque itération)

Pour j allant de 0 à n ($j < n$) **Faire**

Si ($x_{opt}^j \neq x_{opt}^i$) **Alors**

$|X| + 1 \leftarrow |X|$

Fin Si

Fin Pour

Si $|X| == n$ **Alors**

$x_{opt}^i \leftarrow x_{opt}^n$

$n + 1 \leftarrow n$

$X \cup \{x_{opt}^n\} \leftarrow X$

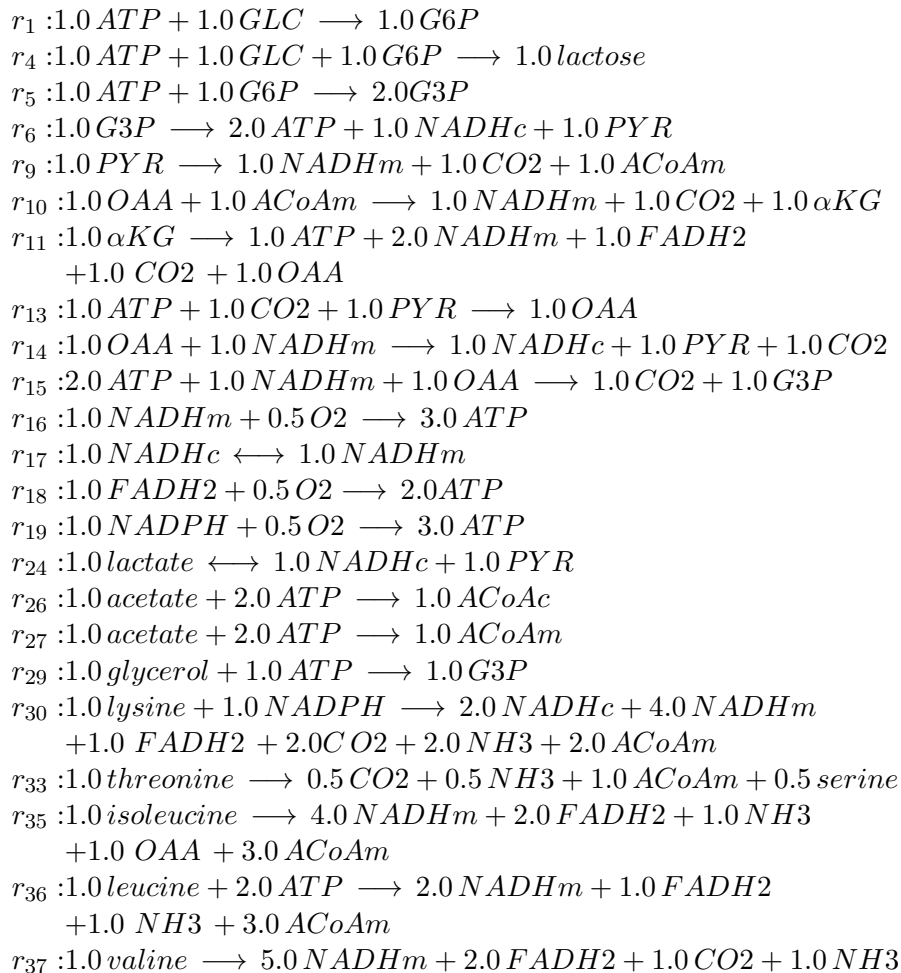
Fin Si

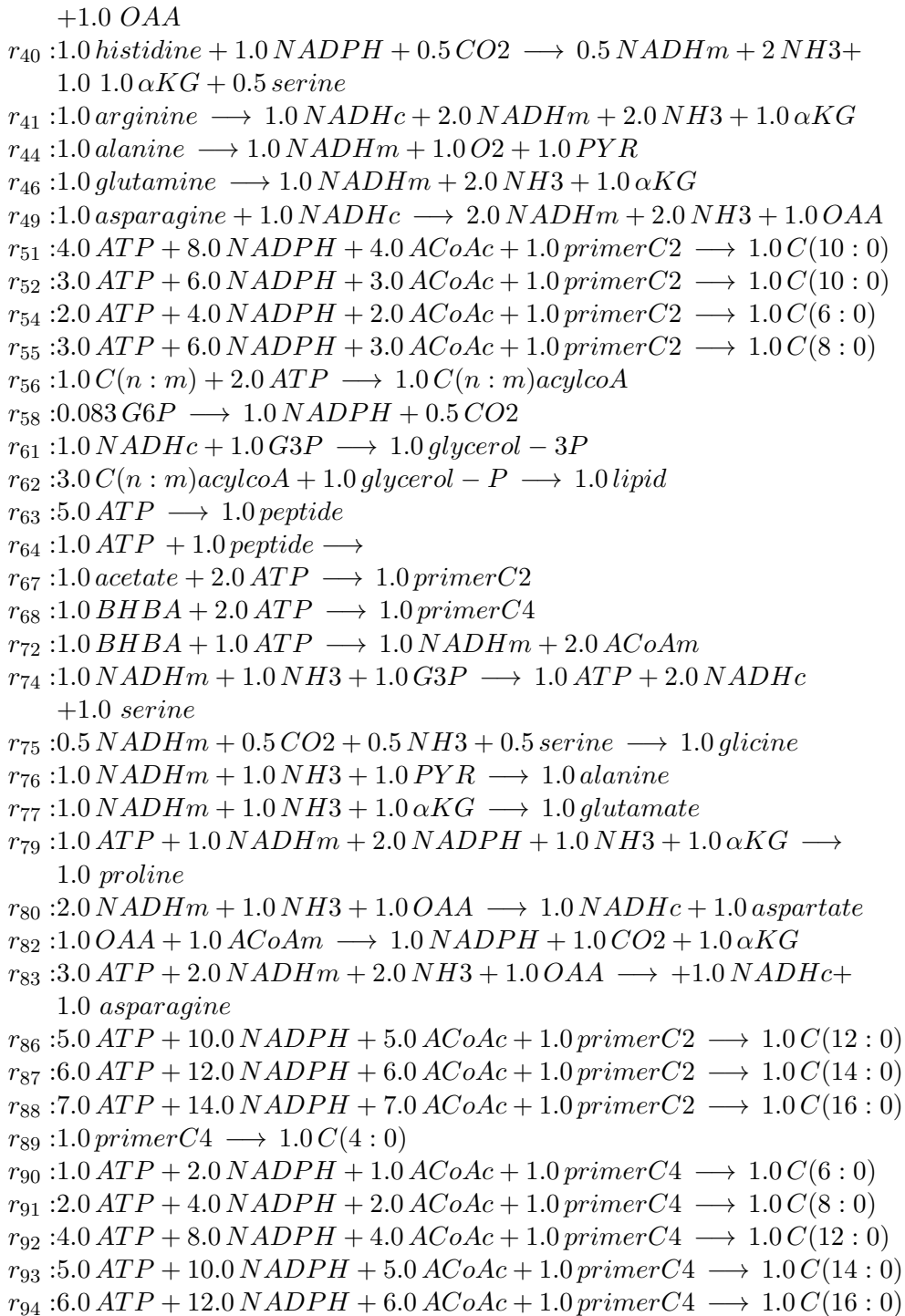
Fin Pour

Annexe B

Données de la glande mammaire

B.1 Les différentes réactions stœchiométriques faisant intervenir les cofacteurs et les pivots





B.1.1 Bilans d'utilisation des pivots

En désignant par α_i le taux d'utilisation d'une réaction r_i et en considérant la stoechiométrie du système, les variations des concentrations des pivots (métabolites intermédiaires) et cofacteurs sont donnés par les équations suivantes :

$$\begin{aligned}
 \frac{d(SER)}{dt} &= \alpha_{74} - 0.5\alpha_{75} \\
 \frac{d(Fapimer)}{dt} &= -\alpha_{51} - \alpha_{52} + \alpha_{67} + \alpha_{68} \\
 \frac{d(Glycerol-3P)}{dt} &= \alpha_{61} - \alpha_{62} \\
 \frac{d(G3P)}{dt} &= 2\alpha_5 - \alpha_6 + \alpha_{15} + \alpha_{29} - \alpha_{61} - \alpha_{74} \\
 \frac{d(G6P)}{dt} &= \alpha_1 - \alpha_4 - \alpha_5 - 0.08\alpha_{58} \\
 \frac{d(GLC)}{dt} &= -\alpha_1 + \alpha_2 - \alpha_4 \\
 \frac{d(ACoAc)}{dt} &= \alpha_{26} - 4\alpha_{51} - 3\alpha_{52} \\
 \frac{d(ACoAm)}{dt} &= \alpha_9 - \alpha_{10} + \alpha_{27} + 2\alpha_{30} + \alpha_{35} + 3\alpha_{36} + \alpha_{72} - \alpha_{82} \\
 \frac{d(PYR)}{dt} &= \alpha_6 - \alpha_9 - \alpha_{13} + \alpha_{24} \\
 \frac{d(\alpha KG)}{dt} &= \alpha_{10} - \alpha_{11} + \alpha_{41} - \alpha_{77} - \alpha_{79} + \alpha_{82} \\
 \frac{d(OAA)}{dt} &= -\alpha_{10} + \alpha_{11} + \alpha_{13} - \alpha_{15} + \alpha_{35} + \alpha_{37} + \alpha_{49} - \alpha_{80} - \alpha_{82} \\
 \frac{d(NADPH)}{dt} &= -\alpha_{19} - \alpha_{30} - 8\alpha_{51} - 6\alpha_{52} + \alpha_{58} - 2\alpha_{79} + \alpha_{82} \\
 \frac{d(FADH2)}{dt} &= \alpha_{11} - \alpha_{18} + \alpha_{30} + 2\alpha_{35} + \alpha_{36} + 2\alpha_{37} \\
 \frac{d(NADHm)}{dt} &= \alpha_9 + \alpha_{10} + 2\alpha_{11} - \alpha_{15} - \alpha_{16} + \alpha_{17} + 4\alpha_{30} + 4\alpha_{35} + 2\alpha_{36} \\
 &\quad + 5\alpha_{37} + 2\alpha_{41} + 2\alpha_{49} - \alpha_{74} - 0.5\alpha_{75} - \alpha_{77} - \alpha_{79} - 2\alpha_{80} + \alpha_{72} \\
 \frac{d(NADHc)}{dt} &= \alpha_6 + \alpha_{14} - \alpha_{17} + \alpha_{24} + \alpha_{29} + 2\alpha_{30} + \alpha_{41} - \alpha_{49} - \alpha_{61} + 2\alpha_{74} + \alpha_{80} \\
 \frac{d(ATP)}{dt} &= -\alpha_1 - \alpha_4 - \alpha_5 + 2\alpha_6 + \alpha_{11} - \alpha_{13} - 2\alpha_{15} + 3\alpha_{16} + 2\alpha_{18} + 3\alpha_{19} \\
 &\quad - 2\alpha_{29} - 2\alpha_{36} - 4\alpha_{51} - 3\alpha_{52} - 2\alpha_{56} - 5\alpha_{63} - 2\alpha_{67} - 2\alpha_{68} + \alpha_{74} - \alpha_{79} \\
 \frac{d(CO2)}{dt} &= \alpha_9 + \alpha_{10} + \alpha_{11} - \alpha_{13} + \alpha_{15} + 2\alpha_{30} + \alpha_{37} + 0.5\alpha_{58} - 0.5\alpha_{75} + \alpha_{82} \\
 \frac{d(O2)}{dt} &= -0.5\alpha_{16} - 0.5\alpha_{18} - 0.5\alpha_{19} \\
 \frac{d(NH3)}{dt} &= 2\alpha_{30} + \alpha_{35} + \alpha_{36} + \alpha_{37} + 2\alpha_{41} + 2\alpha_{49} - \alpha_{74} - 0.5\alpha_{75} - 0.5\alpha_{77} - \alpha_{79} \\
 &\quad - \alpha_{80}
 \end{aligned}$$

Annexe C

Publications

Cet Annexe présente les différentes publications issues de cette thèse

A generic stoichiometric model to analyse the metabolic flexibility of the mammary gland in lactating dairy cows

S. Lemosquet^{1,4}, O. Abdou Arbi², A. Siegel^{2,3}, J. Guinard-Flament^{4,1}, J. Van Milgen^{5,6} and J. Bourdon^{7,2}

¹INRA UMR1080 Dairy Production, 35590 Saint-Gilles, France; sophie.lemosquet@rennes.inra.fr

²INRIA Rennes Bretagne Atlantique, Symbiose project, 35042, Rennes, France

³CNRS Université de Rennes 1, UMR 6074 IRISA, 35042 Rennes, France

⁴Agrocampus Ouest, UMR1080 Dairy Production, 35000 Rennes, France

⁵INRA UMR1079 SENAH, 35590 Saint-Gilles, France

⁶Agrocampus Ouest, UMR1079 SENAH, 35000 Rennes, France

⁷LINA UMR 6241, Université de Nantes, 44322 Nantes, France

Abstract

The stoichiometry model tool of Van Milgen (2002) was adapted to study the metabolism of the ruminant mammary gland. It was used to study the uptake of nutrients in several metabolic reactions using two studies of mammary gland net balance (net nutrient uptake and milk yield and composition) in dairy cows receiving casein infusions in the duodenum. The rules applied to find the partitioning were based on the hypotheses that there is no accumulation of intermediary metabolites in the mammary gland (rule 1) and that there is no deficiency of ATP or cofactors (rule 2). The system was solved manually using a spreadsheet, while respecting the 2 rules. A new automatic tool using Flux Balance Analysis theory was developed. The stoichiometry of the model appeared to be appropriate since the predicted CO₂ production was close to the measured CO₂ production. The nutrient partitioning obtained by manual solving corresponded to one solution among possibly infinite number of solutions. The manual solving of the equations typically corresponded to the maximisation of ATP production while respecting nutrient output for milk synthesis. This maximization of ATP production is a third, explicit rule required to obtain the same partitioning with the automatic tool based on Flux Balance Analysis theory.

Keywords: nutritional model, biochemistry, flux balance analysis, ATP control

Introduction

The main biochemical pathways in the different metabolisms were recently integrated in a generic stoichiometric model (Van Milgen, 2002; Van Milgen *et al.*, 2003) called the ‘metabolic calculator’. The structure of this calculator is based on a restricted number of intermediary metabolites called carbon-chain ‘pivots’, which correspond to important cross points between metabolic pathways (i.e. glucose, pyruvate, acetyl-CoA, oxaloacetate, α -ketoglutarate and serine). Each biochemical reaction is then described by the variations of these intermediary metabolites and also integrates the variation in ATP and other cofactors. The user then has to manually combine these reactions while respecting the rule that there will be no accumulation or deficit of pivots and cofactors. This approach allows the computation of the global cost of product synthesis by a biological entity (e.g. cell, tissue, whole organism). For example, the calculator allows comparison of the efficiency of ATP production from direct glucose oxidation or from glucose oxidation via (temporary) storage as glycogen (Van Milgen, 2002). This tool was also used to check the consistency of the overall nutrient balance in the mammary gland of the sow (Reneau *et al.*, 2003; Van Milgen *et al.*, 2003). A major characteristic of the calculator is its large flexibility, since the choice of reactions to be equilibrated is left entirely to the user.

From a mathematical viewpoint, the metabolic calculator can be seen as a computational environment for studying the solution set of a system of linear equations. Each reaction is associated to a variable, estimating how often the reaction is used in the biological process. Equilibrating the concentration of given nutrients (the pivots) gives rise to linear equations. Thus, performing a computation with the metabolic calculator involves solving a system of linear equations. Unfortunately, the linear system is generally over-dimensioned, since the number of nutrients to be equilibrated (i.e. the number of equations) is much smaller than the number of reactions (i.e. variables). Therefore, from a mathematical viewpoint, the system has an infinite number of solutions, even though the user of the metabolic calculator usually finds what is often claimed to be a unique solution after computation. It appears that the user usually follows an *a priori* ‘deterministic’ choice when performing the computation.

Therefore, one can question which unconscious hypotheses are assumed by the user to navigate in the infinite space of solutions, and how these hypotheses can be formalised as equations and constraints over this space of solutions to design a completely automatic tool. This question will be examined through an example of data of mammary metabolism in dairy cows (Guinard *et al.*, 1994; Guinard and Rulquin, 1994; Lemosquet *et al.*, 2009; Raggio *et al.*, 2006).

Materials and methods

The metabolic calculator: reaction as a function of carbon chain pivots and of cofactors

In the generic stoichiometric model of Van Milgen (2002), reactions were mainly described by a restricted number of intermediary metabolites, called carbon-chain ‘pivots’, at cross points between metabolic pathways presented in Figure 1. These pivots were chosen to describe the main possible conversions between metabolites. As mentioned above, reactions also integrate the variations of ATP, other cofactors and metabolites (i.e. NADH, NADPH, FADH₂, O₂, CO₂ and NH₃).

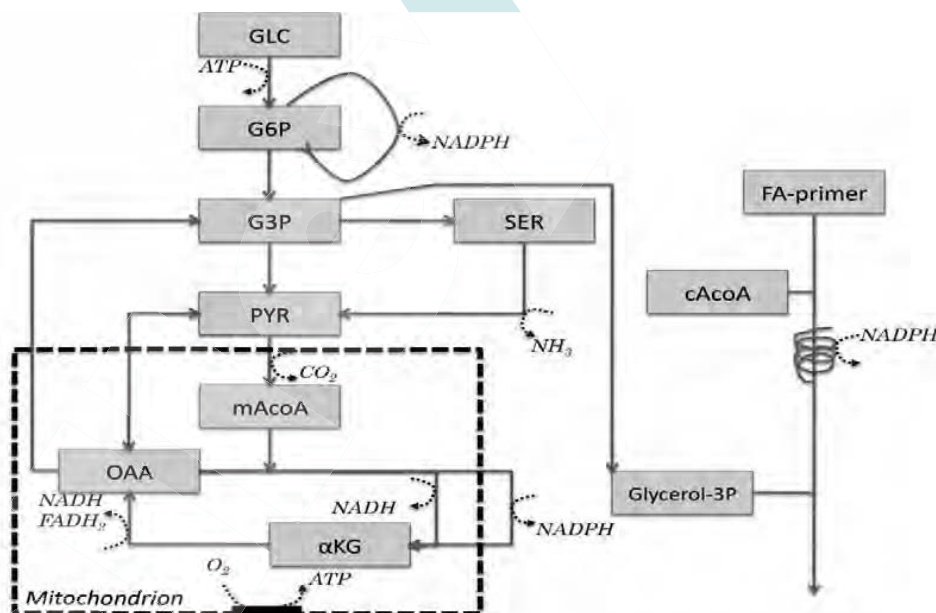


Figure 1. Possible conversions between intermediary pivots (11) in the example of the ruminant mammary gland (GLC: glucose; G6P: glucose 6-phosphate; G3P: triose phosphate; SER: serine; PYR: pyruvate; ACoA: mitochondrial (m) and cytosolic (c) acetyl CoA; αKG: α-ketoglutarate; OAA: oxaloacetate; FA primer: Fatty acid primer: the first two or four carbon chain of a fatty acid; Glycerol 3-P: glycerol 3-Phosphate).

The calculator is provided as a spreadsheet which computes the nutrient balance resulting from the combination of metabolic reactions. As shown in Figure 2, columns contain both cofactors (columns E to K) and products to be equilibrated (pivots – columns M to T). Rows correspond to available reactions and include the stoichiometry of the reactions. Each reaction will be used with a given rate (column C), which will be used as a multiplicative factor for the stoichiometry reactions. The final balance (last row) is obtained by summation of the balances of the selected reactions.

For example, the user may have measured experimentally the uptake of glucose by the mammary gland and lactose output and may have observed that the glucose uptake was greater than the glucose use for lactose synthesis. The user then has to select the appropriate equation (column B) for lactose synthesis and glucose oxidation and manually compute the reaction rates (column C), so that the balance of each pivot equals zero. This computation has two main interests: first, it allows the calculation of the theoretical balance of non-measured nutrients, such as ATP. Second, the value of the reaction rates permit the estimation of the way each nutrient is consumed by the different pathways. This is used to find, analyse and compare scenarios of nutrients partitioning within the biological entity from balance studies, and as mammary net balance studies.

Mathematical modelling and flux balance analysis

Studying metabolic networks is crucial for detecting products that play key roles in biochemical processes. There exists a large body of literature describing tools that address this problem from different viewpoints, included in the large domain of Flux Balance Analysis (Schuster *et al.*, 2000). One of the most important hypothesis of Flux Balance Analysis theory relies in the mass conservation hypothesis (i.e. all the pivots are balanced). It allows one to revisit metabolic networks features as problems of linear algebra in which the stoichiometric matrix plays a central role, tackled with optimization algorithms.

Let us revisit the stoichiometric model tool in this framework. Determining reactions rates in column C of Figure 2 is formalised by setting a variable for each coefficient rate. The vector that gathers all these rate variables is denoted by *F*. We also introduce a vector *B* which describes the balance between outputs and inputs of each nutrient. Experimentally measured data allow the computation of the coefficients of *B* which correspond to unbalanced measured nutrients (such as cofactors). Assuming that there is no accumulation of pivots in the cell (commonly referred to as the balance hypothesis) means that the coefficients of *B* corresponding to pivots equal zero. Independently,

A	B	C	D	E	F	G	H	J	K	L	M	N	O	P	Q	R	S	T
1	eq.	Flux value.	reaction	ATP	cNADH	mNADH	FADH2	CO2	O2	NH3	OAA	αKG	PYR	mACoA	cACoA	GLC	G6P	G3P
100	4	$F_4 = 1,0$	lactose synthesis	-1	0	0	0	0	0	0	0	0	0	0	0	-1	-1	0
101	2	$F_2 = 3,0$	glucose entry	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0
102	1	$F_1 = 3,0$	GLC → G6P	-2	0	0	0	0	0	0	0	0	0	0	0	-2	2	0
103	5	$F_5 = 1,0$	G6P → G3P	-1	0	0	0	0	0	0	0	0	0	0	0	0	-1	2
104	6	$F_6 = 2,0$	G3P → PYR	4	2	0	0	0	0	0	0	0	2	0	0	0	0	-2
105	8	$F_8 = 0,0$	G3P → G6P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
106	9	$F_9 = 2,0$	PYR → mACoA	0	0	2	0	2	0	0	0	0	-2	2	0	0	0	0
107	10	$F_{10} = 2,0$	OAA + mACoA → αKG	0	0	2	0	2	0	0	-2	2	0	-2	0	0	0	0
108	11	$F_{11} = 2,0$	αKG → OAA	2	0	4	2	2	0	0	2	-2	0	0	0	0	0	0
109	13	$F_{13} = 0,0$	PYR → OAA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
110	15	$F_{15} = 0,0$	OAA → G3P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
111																		
112	17	$F_{17} = 2,0$	cNADH ↔ mNADH	0	-2	2	0	0	0	0	0	0	0	0	0	0	0	0
113	16	$F_{16} = 10,0$	mNADH oxidation	30	0	-10	0	0	-5	0	0	0	0	0	0	0	0	0
114	18	$F_{18} = 2,0$	FADH2 oxidation	4	0	0	-2	0	-1	0	0	0	0	0	0	0	0	0
			balance	36	0	0	0	6	-6	0	0	0	0	0	0	0	0	0
				B_{ATP}	B_{cNADH}	B_{mNADH}	B_{FADH2}	B_{CO2}	B_{O2}	B_{NH3}	B_{OAA}	$B_{αKG}$	B_{PYR}	B_{mACoA}	B_{cACoA}	B_{GLC}	B_{G6P}	B_{G3P}

Figure 2. The metabolic calculator spreadsheet: three moles of glucose (GLC) were used and 1 mol of lactose was produced (eq: equations; Coef: coefficient; OAA: oxaloacetate; αKG: α-ketoglutarate; PYR: pyruvate; mACoA and cACoA: mitochondrian (m) and cytosolique (c) acetyl CoA; G6P: glucose 6-phosphate; G3P: triose phosphate).

the stoichiometry of reactions is described by a matrix S , called the stoichiometric matrix ($S_{i,j}$ is the stoichiometric coefficient of the i -th metabolite in the j -th reaction). Computing the reaction rates (i.e. the coefficients of the vector F) needed to explain how the balance vector B may be equilibrated by using the stoichiometry S is then performed by solving the matrix equation $S \cdot F = B$, that is studying the solutions of a system of linear equations. Figure 3a presents the mathematical modelling of the metabolic calculator for the simple example depicted in Figure 2.

Unfortunately, this system commonly has an infinite number of solutions. Flux Balance Analysis is a set of mathematical methods that rely on additional constraints over the space of solutions to explore it and enhance solutions with a biological relevance. Reactions rates (called flux coefficients in this theory) are constrained by both thermodynamic constraints (that impose some reactions to be irreversible) and by Michaelis-Menten dynamics (yielding bounds for the productions of metabolites in the cell). All together, such bounds and constraints imply that the space of solutions has a very particular shape: it is a simplex (i.e. a convex polyhedra) as depicted in Figure 3b.

The very specific shape of the space of solution allows a better understanding of the biological strategy used by the entity to adapt to uptakes. By taking benefit of the linear structure and optimisation algorithms, it becomes possible to compute the solution satisfying a global objective function. For instance, as stated by Edwards and Palsson (2000) and Edwards *et al.* (2001), it was shown that bacteria select reactions that maximise biomass growth. Similarly, other species have been shown to evolve to use the smallest quantity of an exogenous nutrient such as glucose (Fell and Small, 1986).

From these examples, it turns out that exploring the infinite space of solution of linear equations yielded by stoichiometry can be handled in a formalised framework, by solving a linear optimisation problem based on three types of information: (1) A stoichiometric model that defines the equations (this is exactly what appears in the stoichiometric model tool of Van Milgen); (2) Constraints deduced from thermodynamics or obtained from the literature, allowing to restrict the combinations between reactions; (3) An objective function, which is the main phenotype to globally optimise the system to select a unique solution within the space of solutions.

Notice that the objective function defines an optimisation direction. The optimal solution is thus placed at a vertex node of the simplex. The simplex method is an efficient method to solve such a problem (Chvatal, 1983). It consists in a walk on the vertices of the simplex passing from one

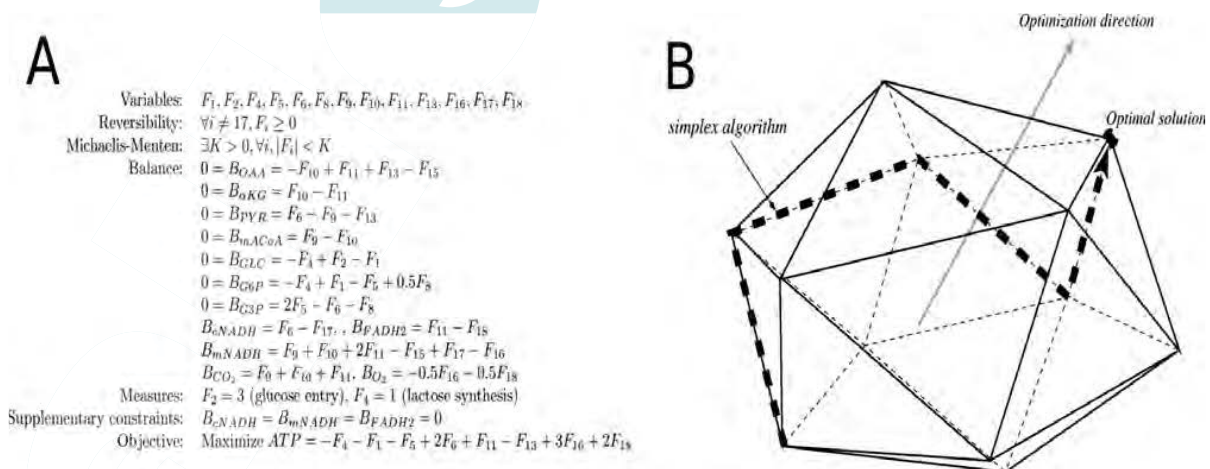


Figure 3. (A) a mathematical viewpoint for the metabolic calculator (same example as Figure 2) and (B) the space of solutions of the stoichiometric model has a polyhedra shape.

vertex to its best neighbour according to the objective function. The algorithm ends when it is not possible to improve the objective function.

With this Flux Balance Analysis theory at hand, a natural question is now to understand whether the computations performed with the metabolic calculator can be fully included in this framework, by exhibiting some constraints and objective functions which can model the implicit search of a solution. We studied this question on a specific example related to metabolism of the mammary gland.

Adaption of the stoichiometry to ruminant mammary gland

To adapt the original model to mammary gland metabolism of ruminants, five additional intermediary metabolites (i.e. pivots) were introduced (Figure 1). Mitochondrial (m) and cytosolic (c) acetyl-CoA (AcCoA) were used to distinguish between utilisation of cAcCoA for fatty acid (FA) synthesis and utilisation of mAcCoA in the tricarboxylic acid cycle (TCA). In addition, glucose-6-phosphate (G6P) and triose-phosphate (G3P) were introduced to account more specifically for the partitioning of glucose (in this case, the use of G6P) between the pentose phosphate pathway, lactose synthesis and glycerol utilisation. For lipid metabolism, an intermediary metabolite called fatty acid primer (FA primer) was introduced. It corresponds to the primers of two or four carbon-units necessary to initiate fatty acid synthesis before elongation. In ruminants, this primer can be acetate or β -hydroxybutyrate (BHBA). Additional reactions were introduced to describe milk synthesis. For example, NADPH synthesis through the isocitrate deshydrogenase pathway was included. The final spreadsheet included a total of 11 pivots and 92 reactions. However, some reactions were considered as non-existing in the ruminant mammary gland, in particular the NADPH synthesis through malic enzyme. Overall, the reactions taken into account were close to those retained in other models of mammary metabolism (Waghorn and Baldwin, 1984; Hanigan and Baldwin, 1994).

Dataset and input data

Two experiments studying mammary metabolism in response to casein infusions in the duodenum in supplement to diet were analysed. In the first experiment (Exp. A: Guinard *et al.*, 1994; Guinard and Rulquin, 1994), treatments corresponding to 0 (Ctrl), 177 g/d, 762 g/d of casein (CN) infused were compared to control (Ctrl: 0 g/d of CN infused) and CN (743 g/d) in the second experiment (Exp. B: Lemosquet *et al.*, 2009; Raggio *et al.*, 2006). The input data corresponded to half-udder uptakes of major nutrients (GLC, all individual amino acids (AA), acetate, BHBA, glycerol, lactate) and of half-udder milk output (lactose, AA in protein synthesised in the mammary gland and all FA synthesised within the mammary gland, glycerol-3-P required for milk triglycerides), and mammary blood CO₂ production. Precise calculations of milk component output were described in Lemosquet *et al.* (2009). We used similar hypotheses to those adopted by Hanigan and Baldwin (1994) for AA composition of milk protein (Raggio *et al.*, 2006) and to those adopted by Waghorn and Baldwin (1984) for milk fatty acid synthesised in the mammary gland. It was assumed that all FA from C4 to C12, 85% of C14 and C15, and 60% of C16 and C17 were synthesized in the mammary gland as in Waghorn and Baldwin (1984).

In the present computation, protein and long chain FA directly taken up by the mammary gland and not synthesised were not taken into account. In addition, if the net balance (uptake – output in milk protein) of a non-essential AA (NEAA) was significantly different from zero and negative then this NEAA was considered to be synthesized within the mammary gland. Similarly, an AA was considered to be catabolised (including NEAA synthesis) if its net balance (uptake-output in milk protein) was significantly different from zero and positive.

Results*Performing computations with the metabolic calculator*

Several hypotheses were used to select the reactions and finally compute all reactions rates describing the nutrient partitioning between metabolic pathways. Two types of rules could be distinguished: 'strong' rules absolutely necessary to find a solution and 'weak' rules. The first two strong rules were based on the hypotheses that there is no accumulation or deficiency of intermediary metabolites in the mammary gland (rule 1) and that there is no deficiency of ATP or cofactors (rule 2). Most of the weak rules were deduced from the literature from ruminant mammary metabolism. These rules are open for debate and may be modified in future computations. Among these weak rules, glucose was assumed to be the sole precursor of lactose, uptake of glycerol by the mammary gland was not used for glycerol-3P synthesis but only oxidised. For NADPH production, it was assumed that two-third of NADPH was produced in the pentose phosphate pathway (Hanigan and Baldwin, 1994) and the rest by the isocitrate dehydrogenase pathway.

One important point, in terms of the mathematical problem, is that these weak rules allowed selection of a set of reactions in the spreadsheet from which to compute coefficient rates such that intermediary metabolites and cofactors (in columns) and reactions (rows) respect the two strong rules.

Biological validation of dataset and results of 'metabolic calculator' computations

This model allowed us to test easily the validity of the datasets for several aspects.

First, it analysed carbon and nitrogen balances through CO_2 and NH_3 , respectively (Table 1). With respect to the stoichiometry of the reactions, CO_2 calculated by the model corresponded to carbon uptake minus carbon output in milk and NH_3 production to nitrogen balance. In Exp. B, carbon (Lemosquet *et al.*, 2009) and nitrogen balances (Raggio *et al.*, 2006) were quite well equilibrated and close to the model value. In both experiments, CO_2 production was measured by a blood gas apparatus method, and the model gave very similar results, thereby indirectly validating the measurement. The differences in nitrogen balance were linked to differences in AA uptake retained. In Exp. A, N balances were not equilibrated. For example, in the CN (743 g/j) treatment of Exp. A, NH_3 was positive. This positive nitrogen balance was also highlighted by Guinard and Rulquin (1994) and they hypothesised that peptide production may occur, which was not accounted for in the present model.

Second, it allowed the determination of whether enough nutrients were taken up to explain milk synthesis based on other criteria than carbon or nitrogen balances. In the Ctrl treatment of Exp. A, in addition to the deficit in NH_3 group (-29.7 mmol/h) to synthesise NEAA, it was not possible to respect the rule of no deficit in pivots since intermediary metabolites such as G3P were negative. These observations may indicate an underestimation of the mammary blood flow in Ctrl treatment of Exp. A.

Third, in this example by applying the same rules of computation, the partitioning of nutrient utilisation can be analysed and compared between experiments. As an illustration, ATP balance was very similar in high CN treatments in both experiments (in Table 1: 1,941 vs. 1,835 in Exp. A and B, respectively). In Exp. A, ATP balance and ATP gain from AA catabolism greatly increased in response to increasing CN infusions. However, in Exp. B, ATP balance decreased but ATP gain by BHBA oxidation greatly increased in response to CN infusions.

Table 1. Example of the results obtained after manual computation of net mammary balances of CO₂, NH₃ and ATP in two experiments (Exp. A and Exp. B) in dairy cows.

	Exp. A ¹			Exp. B ²	
	Ctrl ³	CN ⁴ : 177 g/j	CN ⁴ : 743 g/j	Ctrl ³	CN ⁴ : 762 g/j
(mmol/h/half udder)					
CO ₂ measured	628	760	864	930	1,030
CO ₂ by the model	526	720	841	1,047	988
N balance calculated					
NH ₃ by the model	-29.7	10.1	61.4	-3.7	-2.2
Balance of ATP ⁵	232	1,438	1,941	2,579	1,835
ATP from BHBA ⁶	827	940	911	505	1,887
ATP from AA ⁷	60	298	990	280	391

¹ Guinard *et al.* (1994) and Guinard and Rulquin (1994).

² Lemosquet *et al.*, 2009 and Raggio *et al.* (2006).

³ Control treatment.

⁴ Casein treatments (duodenal infusions).

⁵ Obtained by the model.

⁶ ATP produced during BHBA oxidation through isocitrate dehydrogenase pathways.

⁷ ATP produced during amino acids (AA) catabolism.

Automatically recovering identical results with Flux Balance Analysis together with an ATP optimisation objective

The computations with the metabolic calculator presented above might suggest a unique solution of partition, which is not the case. Indeed, the stoichiometric model adapted to the ruminant mammary gland generates a linear system $S.F = B$ composed by 23 equations and 40 variables (only a subset of all the 92 possible reactions described by Van Milgen model are used). This leads to a rectangular stoichiometric matrix S of size 23×40, a flux vector F with 40 coefficients rates to be estimated and a balance vector B with 23 nutrients to be balanced (11 of these corresponds to pivots). For all irreversible reactions i , the constraint F_i strictly positive is added. The resulting linear problem is clearly under-parameterised and has an infinite number of solutions. Several additional constraints are simply added to formalise the weak rules used to perform the computation with the stoichiometric tool (see above). For instance, one of these constraints concerns lactose synthesis. In our study, we impose that the unique origin of lactose is glucose (for GLC and G6P parts of lactose). Thus equation 8 ($G3P \rightarrow G6P$) is not active leading to $F_8 = 0$. Unfortunately, even with these new constraints, the system of equations is still under-parameterised: it admits an infinite number of solutions.

As shown in the Flux Balance Analysis theory, focusing on a particular point of the space of solutions requires the definition of a general objective function that formalises the general phenotype that is optimised during milk synthesis. To find a unique solution, it would turn the mathematical problem into a linear optimisation problem. Notice that linear means that both constraints defining the space of solutions and the objective function are linear combinations of the variables F_p , such as balances of products that we are currently considering.

As mentioned before, several objective functions have been proposed in the literature depending on biological applications. However, they are often related to the optimisation of biomass production. In our case, the consumption of every nutrient is known and appears in the balance. Furthermore, the quantity of nutrients is exactly the same for all solutions. Therefore, the biomass growth depends only on the consumption of ATP. Thus, maximising the biomass growth is equivalent here to a minimisation of ATP consumption.

Following this idea, we used the simplex algorithm (see the Materials and Methods section) to estimate reaction rates that maximise the production of ATP while satisfying the stoichiometric constraints given by the model and the literature constraints discussed above. Performing this computation automatically resulted in the same reaction rates as those obtained with the manual computation of the 'metabolic calculator'. This suggests that computations with the metabolic calculator spreadsheet tool actually follow an optimisation strategy just as those used in Flux Balance Analysis. In the context of ruminant mammary gland, the optimisation is performed such that ATP production globally reaches its maximum.

Discussion

The formalisation as equations and constraints by the completely automatic tool based on Flux Balance Analysis revealed that an objective function was required to find a unique solution of nutrient partition in each treatment. In the present case, manually using the metabolic calculator appears to correspond to the hypothesis of a maximisation of ATP production (i.e. the objective function). This means that the usual choice of pathways for milk synthesis corresponded to a minimisation of ATP consumption. This part was not a conscious choice. However, this also means that when nutrient uptake is higher than the amount required for milk synthesis, the remainder of the nutrients is catabolised, which is necessary to satisfy the rule (constraint) of non-accumulation of pivots and cofactors. For example, it is evident that mammary glucose uptake should always exceed lactose synthesis. The question remains whether it is relevant to take the maximisation of the ATP as the objective function?

At a cell scale this objective function seemed to be relevant. In fact, several objective functions, mainly based on maximisation of biomass growth, have been proposed in the literature, especially for bacterial metabolism (Fell and Small, 1986; Edwards and Palsson, 2000). In our context, as discussed in the result section, maximising the biomass growth is equivalent here to a minimisation of ATP consumption.

At the scale of the mammary gland metabolism, the question concerning the appropriate objective function remains open. To our knowledge, the ruminant mammary metabolism model of Hanigan and Baldwin (1994) is based on other constraints than the maximisation of ATP. It is a model of simulation of the literature knowledge on mammary metabolism. It was validated by using *in vitro* measurements of partitioning through tracer studies. However, Hanigan *et al.* (2009) seemed to find similar quantities of ATP used in milk component (3.3 mol/kg of milk) and of ATP generated by mammary gland in their model compared with the present computation. With the Ctrl treatment of Exp. B, the quantity of ATP used for milk synthesis was also 3.3 mol/kg of milk and, the quantity of ATP generated for a whole udder (7.8 mol/kg of milk) exceed by 2-fold the quantity of ATP used for milk as reported by Hanigan *et al.* (2009). The similarity of the results between both models suggests that the biological rule of maximisation of ATP may be acceptable to explore mammary metabolism. This concept related to ATP will be also interesting to explore since mammary nutrient uptake is regulated by ATP availability in the model proposed by Cant and McBride (1995).

At a level of the whole organism, the rule of maximisation of ATP seems to be less obvious. In Molly cow model, the efficiency of ATP in several metabolic pathways was assumed to decrease to avoid a too high increase of the cow body weight in high fat diet treatments (Johnson *et al.*, 2009). Another example which occurred at whole body metabolism is the variation of efficiencies of ATP production from GLC (Van Milgen, 2002). It can vary from 38 ATP/mol of GLC (direct oxidation) to 31 ATP/mol of GLC, if GLC is converted in glutamate and if glutamate is incorporated. However, the maximisation rule may still hold if additional constraints will be considered. For example, glutamate is extensively catabolised in the gut, and the oxidation of glucose to ATP via glutamate can be seen as a spatial organisation of metabolism (Van Milgen, 2002). Similarly, glucose oxidation via lipid can be interpreted as a temporal organisation of metabolism and the constraint (glucose to ATP via lipid) may be included as a constraint. The rule of maximisation of ATP, under this constraint, may still hold.

In conclusion, the mathematical formalisation using the flux balance analysis theory of the stoichiometric tool of Van Milgen (2002, 2003) required an objective function to find a unique solution of nutrient partitioning. The validity of an objective function based on the maximization of ATP as a rule of nutrient partition needs further confirmation.

References

- Cant, J.P. and McBride, B.W., 1995. Mathematical analysis of the relationship between blood flow and uptake of nutrients in the mammary glands of a lactating cow. *Journal of Dairy Research* 62:405-422.
- Chvatal, V., 1983. *Linear Programming*. W.H. Freeman and Co. New York, NY, U.S., 478 pp.
- Edwards, J.S., Ibarra, R.U. and Palsson, B.O., 2001. *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nature Biotechnology* 19:125-130.
- Edwards, J.S. and Palsson, B.O., 2000. The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proceedings of the National Academy Sciences* 97:5528-5533.
- Fell, D.A. and Small, J.R., 1986. Fat synthesis in adipose tissue. An examination of stoichiometric constraints. *Biochemical Journal* 238:781-786.
- Johnson, H.A., Calvert, C.C. and Fadel, J.G., 2009. Impact of adjusting energy values on predictions of energy metabolism using molly. In: Sauvante, D., Agabriel, J., Faverdin, P., Lescoat, P. and Van Milgen, J. (eds.) 7^e International workshop. *Modelling Nutrient Digestion and Utilization in Farm Animals*, September 10-12, 2009. Paris, France, p 42.
- Guinard, J., Rulquin, H. and Vérité, R., 1994. Effect of graded levels of duodenal infusions of casein on mammary uptake in lactating cows. 1. Major nutrients. *Journal of Dairy Science* 77:2221-2231.
- Guinard, J. and Rulquin, H., 1994. Effect of graded levels of duodenal infusions of casein on mammary uptake in lactating cows. 2. Individual amino acids. *Journal of Dairy Science* 77:3304-3315.
- Hanigan, M.D. and Baldwin, R.L., 1994. A mechanistic model of mammary gland metabolism in the lactating cow. *Agricultural Systems* 45:369-419.
- Hanigan, M.D., France, J., Mabweesh, S.J., McNabb, W.C. and Bequette, B.J., 2009. High rates of mammary tissue protein turnover in lactating goats are energetically costly. *Journal of Nutrition* 139:1118-1127.
- Lemosquet, S., Raggio, G., Lobley, G.E., Rulquin, H., Guinard-Flament, J. and Lapierre, H., 2009. Whole-body glucose metabolism and mammary energetic nutrient metabolism in lactating dairy cows receiving digestive infusions of casein and propionic acid. *Journal of Dairy Science* 92:6068-6082.
- Raggio, G., Lemosquet, S., Lobley, G.E., Rulquin, H. and Lapierre, H., 2006. Effect of casein and propionate supply on mammary protein metabolism in lactating dairy cows. *Journal of Dairy Science* 89:4340-4351.
- Renaudeau, D., Noblet, J. and Dourmad, J.Y., 2003. Effect of ambient temperature on mammary gland metabolism in lactating sows. *Journal of Animal Science* 81:217-231.
- Schuster, S., Fell, D.A. and Dandekar, T., 2000. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature Biotechnology* 18:326-332.

Part 4

- Van Milgen, J., 2002. Modeling biochemical aspects of energy metabolism in mammals. *Journal of Nutrition* 132:3195-3202.
- Van Milgen, J., Gondret, F. and Renaudeau, D., 2003. The use of nutritional models as a tool in basis research. In: Souffrant, W. B. and Metges, C. C. (eds.) *Progress in research on energy and protein metabolism*. September 13-18, 2003. Rostock, Germany, pp.259-263.
- Waghorn, G.C. and Baldwin, R.L., 1984. Model of metabolite flux within mammary gland of the lactating cow, *Journal of Dairy Science* 67:531-544.

Exploring metabolism flexibility in complex organisms through quantitative study of precursor sets for system outputs

Oumarou Abdou Arbi^{1,2,3}, Sophie Lemosquet^{4,5}, Jaap Van Milgen^{4,5}, Anne Siegel^{2,1,3}, Jérémie Bourdon^{*6,3}

¹Université Rennes 1 IRISA UMR 6074. Campus de Beaulieu. 35042 Rennes Cedex. France

² CNRS. IRISA UMR 6074. Campus de Beaulieu. 35042 Rennes Cedex. France

³ INRIA. Campus de Beaulieu. 35042 Rennes Cedex. France

⁴ INRA. UMR1348 Pegase. F-35590 Saint-Gilles. France

⁵ Agrocampus Ouest. UMR1348 Pegase. F-35000 Rennes. France

⁶LINA. UMR 6241, Université de Nantes, France.

Email: oumarou.abdou-arbi@irisa.fr; Sophie.Lemosquet@rennes.inra.fr; Jaap.VanMilgen@rennes.inra.fr; anne.siegel@irisa.fr; Jeremie.Bourdon@univ-nantes.fr;

*Corresponding author

Abstract

Background

When studying metabolism at the organ level, a major challenge is to understand the matter exchanges between the input and output components of the system. For example, in nutrition, biochemical models have been developed to study the metabolism of the mammary gland in relation to the synthesis of milk components. These models were designed to account for the quantitative constraints observed on inputs and outputs of the system. In these models, a compatible flux distribution is first selected. Alternatively, an infinite family of compatible set of flux rates may have to be studied when the constraints raised by observations are insufficient to identify a single flux distribution. The precursors of output nutrients are traced back with analyses similar to the computation of yield rates. However, the computation of the quantitative contributions of precursors may lack precision, mainly because some precursors are involved in the composition of several nutrients and because some metabolites are cycled in loops.

Results

We formally modeled the quantitative allocation of input nutrients among the branches of the metabolic network (AIO). It corresponds to yield information which, if standardized across all the outputs of the system, allows a precise quantitative understanding of their precursors. By solving nonlinear optimization problems, we introduced a method to study the variability of AIO coefficients when parsing the space of flux distributions that are compatible with both model stoichiometry and experimental data. Applied to a model of the metabolism of the mammary gland, our method made it possible to distinguish the effects of different nutritional treatments, although it cannot be proved that the mammary gland optimizes a specific linear combination of flux variables, including those based on energy. Altogether, our study indicated that the mammary gland possesses considerable metabolic flexibility.

Conclusion

Our method enables to study the variability of a metabolic network with respect to efficiency (i.e. yield rates). It allows a quantitative comparison of the respective contributions of precursors to the production of a set of nutrients by a metabolic network, regardless of the choice of the flux distribution within the different branches of the network.

Background

When studying metabolism, it is important to elucidate how fluxes are distributed among the different pathways of the metabolic network with respect to the available quantitative information about the system behavior. Several methods can be used to address this issue. The first approach consists of building a mechanistic description of transformations and identifying the regulations involved in the system. Continuous dynamical models are often used for this purpose, especially when time-series responses to different treatments are available to infer the dynamics of the network. Static approaches such as Petri net can also identify qualitative distributions of fluxes in a metabolic network [51], and their stochastic extensions can even take into account stoichiometric and kinetic information [26]. With a complementary approach, one can study a system at the functional level, based on the study of fluxes at steady states. This is the purpose of the Flux Balance Analysis (FBA) framework, which has evolved considerably over the past decades [29, 36]. With an FBA, the identification of external regulations is not necessary because it is assumed that the global behavior of the system can be modeled by optimizing linear combinations of selected fluxes (i.e. the *objective function*). Roughly, the methods developed in this field aim to explore a convex space of plausible

flux distributions and to study the *extreme* flux distributions obtained when optimizing a linear objective function. It allows checking whether an extreme flux distribution is consistent with experimental data and to predict new experimental observations [18, 28, 40, 41]. Nonetheless, the consistency of the solutions obtained by FBA depends on the quality of the constraints integrated in the model. To overcome this limitation, several extensions have been proposed in the literature. These extensions can be broken down into two parts. The first incorporates additional biological knowledge, such as reaction thermodynamics [5] or multioptimization [35]. The second is based on the use of FBA to globally analyze large-scale metabolic networks. For example, in flux variability analysis [34], the minimum and the maximum flux for each reaction in the network are computed under some (sub-)optimal conditions.

In this paper, our main purpose is to extend this framework to study the variability of a metabolic network at the level of *efficiencies* instead of *fluxes*. Indeed, when studying a metabolism at the organ level, a major challenge consists in comparing the efficiency, or yield rates, of two metabolic situations, i.e. the response to various input patterns. A typical example of such studies are those concerning animal nutrition, which aim to predict the quality and quantity of animal production (meat, fat, milk, etc.) in response to breeding factors. In this field, the energy and protein conversion efficiencies are derived from the study of the flux distribution of input nutrients between the different branches of the metabolic network [20]. More generally, although its definition depends on the field of application, the concept of efficiency is often linked to energy, mass growth and protein conversion [8, 25, 49]. However, the computing of efficiencies is prone to difficulties and errors, since it requires computing the quantities precursors sets which are required to explain the measured composition of outputs, although some of the internal products may be recycled within cycles [1]. With this goal, we defined the *allocation of an input towards an output* (with the abbreviation AIO) to be the proportion of a matter component (such as carbon) in a given input flux that is recovered in the selected output flux. Our definition is based on the choice of a material component, such as carbon or nitrogen, which allows a comparison of the contributions of input metabolites to the composition of output products. It can be seen as a yield rate, which is uniform among all the outputs of the system, allowing a precise understanding of the precursors used. As a first methodological contribution, we prove that AIO can be uniquely described and computed, even in the case when there are metabolic system cycles, by a matrix whose coefficients are nonlinear functions of the flux variables. Introducing nonlinearity in the definition of AIO cannot be avoided because of the presence of these cycles.

Studying the variability of AIO within a complete space of plausible flux distributions requires the solving of nonlinear optimization problems which are underdetermined in tangible applications. As a second

methodological contribution, we have proposed efficient algorithms to compute lower and upper bounds for AIOs over the family of flux distributions which are compatible with both the system’s stoichiometry and the experimental datasets, regardless of the choice of a flux distribution for the internal branches of the network. An important aspect is that when the metabolic network is provided with input-output data, the complete space of plausible distributions appears to have a relatively small size, and can therefore be studied with our method.

Our main example of application is related to milk production. In this context, several models have been introduced, in relation with the aforementioned classification of models. One class of small-size dynamical mechanistic models predicts the blood flow and input nutrients of the metabolic system (i.e. the mammary gland) [13,50] or, alternatively, the nutrients produced by the metabolic system (in terms of milk composition) [23,24]. Another class of models predicts the distribution (i.e. partitioning) of fluxes over the pathways from the input and output nutrients of the system [22,52]. In the latter family, both dynamic and static approaches exist. Indeed, numerical models based on mass-action equations were initially proposed to describe the fluxes to and from individual metabolites, with different levels of description [22,52]. To that goal, optimizers were used to determine a reasonable set of parameter estimates for the dynamical model of the system. Although such a set of parameter estimates is not unique. With a totally different approach, in another study, the main biochemical pathways in the different metabolisms were integrated in a generic stoichiometric model called the *metabolic spreadsheet* [48,49]. Its structure was based on a restricted number of intermediary metabolites called carbon-chain *pivots*, which correspond to important cross-over points between metabolic pathways. This allowed the computation of the flux rates for all reactions, constrained by a general rule stating that there is no accumulation of intermediary metabolite or cofactors. A study of manual calculations performed with the metabolic spreadsheet showed that it works, in practice, by maximizing an objective function (ATP production) in a convex solution space [31]. Therefore, this model can be considered as an application of the Flux Balance Analysis framework [37].

Nonetheless, in this field of study, there has been little discussion on the impact of the choice of a single model among several (possibly infinitely many) reasonable models [25]. To investigate this issue in a more automatic way, we compared the aforementioned conventional models to the convex space of plausible flux distributions associated with steady states of a model of mammary gland metabolism, as computed in FBA. We checked the consistency of extreme flux distributions of nutrients with experimental data of the mammary gland and milk production, including the contribution of nutrient input-output and isotope balance studies [6,9]. Based on our AIO computation framework, our analysis highlighted that the metabolic

behavior of the mammary gland cannot be modeled by maximizing ATP production or by optimizing a linear combination of flux variables of the model of the mammary gland. In other words, although an infinite number of flux distributions are compatible with the data, none are extreme within the space of feasible models. Selecting any of these nonoptimal flux distributions to predict the system behavior appears difficult without additional experimentation.

To gain better understanding of the system response regardless of the choice of a flux distribution for the internal branches of the network, we applied our method to estimate the variability of AIO coefficients in our model and compared the effects of two different diets on mammary gland metabolism. Our results suggest that the bounds of AIO are sufficient to distinguish the effects of different nutritional treatments without selecting a flux distribution for the internal reactions of the metabolic network by any method - optimization of a linear combination of fluxes or a residual score. Overall, the complete study suggests considerable flexibility in mammary gland metabolism. It provides a view of the functioning of the system although its internal processes still cannot be clarified because of limitations on experimentation on large animals such as ruminants.

Results

We first investigated the set of flux distributions that are compatible with the stoichiometry of our mammary gland model (depicted in Fig. 1), without taking datasets into account. The model exhibited a large variability, since thousands of extreme pathways could be identified.

We then successively computed the set of flux distributions compatible with the model and the real datasets of lactation metabolism in dairy cows. The datasets are given in Table 1. They include a control diet (Ctrl), a diet related to an increased protein supply through casein infusion into the duodenum (CN), and a complementary dataset (HB) previously used in a mechanistic model of the mammary metabolism [22]. For the three datasets, the set of plausible flux distributions – solutions to a linear system detailed in Eq. (1) – was an unbounded convex cone of dimension 5. In all cases, the solution spaces shared the same set of five independent variables. This suggests that there are five independent levels of variability within the system: peptide hydrolysis (R_{64}), NADPH oxidation (R_{19}), OAA→PYR (R_{14}), OAA→G3P (R_{15}) and G3P→G6P (R_8). Therefore, by including this dataset the model became a fairly small and constrained network. Nevertheless, it was not uniquely determined since there were still several degrees of freedom.

Investigating the relevance of the optimization strategies for mammary metabolism

The balance between the ATP generated by the system and the ATP used by the system (including the ATP cost of milk component synthesis) was computed for the three datasets (Ctrl), (CN) and (HB). The results are detailed in Table 2. In this table, the FBA approach based on optimization of the ATP balance is compared to the natural functioning of the system.

First, we considered the manual computation of fluxes with a tool named “metabolic spreadsheet” [48,49]. We applied the rules of no accumulation on any intermediary metabolites and cofactors with a utilization rejecting all the cycles. As expected, since some cycles are ATP-consuming, both flux distributions were equivalent in both approaches (Table 2). The ATP balances were respectively 3081, 2045 and 6628 mmol/h/half udder (i.e. 318 mol/d/udder) in the (Ctrl), (CN) and (HB) datasets.

According to this model, the ATP generated is estimated to be 6500 mmol/h/half udder (i.e. 312 mol/d/udder) while 2125 mmol/h/half udder (i.e. 102 mol/d/udder) were estimated to be used for milk component synthesis [21]. Therefore, the ATP-balance obtained for this model was slightly smaller than the optimal one obtained for our mammary gland model. Independently of the approach considered and the model or dataset at hand, the remaining ATP (ATP balance), after use for milk synthesis, appeared to be rather high and nonconstant. Indeed, as shown in [21], an ATP balance of 1250 mmol/h/half udder (60 mol/d/udder) can be expected to be used for other functions not accounted for in the models, such as maintaining membrane potential and synthesizing nucleic acids. The numerical values obtained here were far from this estimated balance, suggesting that each of these three models have a nonrelevant ATP balance.

In addition, both in the ODE model and the ATP-optimization approach, peptide hydrolysis was obtained at zero, implying an absence of any protein turnover. This contradicts all the observations about this pathway: considerable use of this pathway has been evidenced in several publications, although the peptide hydrolysis rates differed significantly depending on the technique used for the measurements: peptide hydrolysis (R_{64} i.e. mammary protein degradation) spans from 0.25, 0.23 to 0.67 of peptide synthesis (R_{63} i.e. total mammary protein synthesis) in (Ctrl), (CN) and (HB), respectively [21,39].

Overall, we concluded with this analysis that the energy-based optimization function may not allow an appropriate simulation of mammary gland metabolism. This was expected, considering that the system is studied at the complete organ level, involving competing processes which can rarely be modeled with a single linear objective function.

Exploring all extreme flux distributions in a refined simplex

In order to study the variability within the space of plausible flux distributions and to identify alternative relevant optimization strategies, an additional constraint was placed on the ATP balance of the system. We considered that an ATP-balance of 1250 mmol/h/half udder (60 mol/d/udder) was a relevant measure for this study [21], although we checked that all the results discussed below were still valid when introducing a 10% tolerance on this estimation of ATP-balance. Providing a bound for ATP yielded a new constraint on several variables of the system. For the two datasets (Ctrl) and (CN), the flux G3P→G6P (R_8) was no longer an independent variable in the system. The simplex, which was previously unbounded, appeared to be bounded with four independent variables: OAA→PYR (R_{14}), OAA→G3P (R_{15}), NADPH oxidation (R_{19}), peptide hydrolysis (R_{64}).

Applying flux variability tools [38], it appeared that the minimum of each of the fluxes $v_{14}, v_{15}, v_{19}, v_{64}$ was equal to zero for all treatments. The maxima of the fluxes ($v_{14}, v_{15}, v_{19}, v_{64}$) were (1831, 1831, 669, 305) for the (Ctrl) treatment and (795, 795, 22, 133) for the (CN) treatment. This suggests that the (Ctrl) treatment generates a more flexible space of plausible flux distributions than the (CN) treatment.

We computed all extreme vertices of the simplex of plausible flux distributions for the two treatments (Ctrl), (CN). The simplex structure of the polyhedron implies that the optimum of any linear combination of metabolic fluxes involved in the model is either uniquely attained for one of these extreme points or attained by all flux distributions positioned on a face of the simplex. To gain insight on the pathways involved in the variability of our model, we also computed the linear combination of fluxes optimized for one of the extreme flux distributions. Eight extreme vertices were found for the (Ctrl) and (CN) datasets. They were obtained when optimizing the same set of linear functions ¹. In Table 3, the optimal flux distributions for (Ctrl), (CN) are classified according to the activation or inactivation of several pathways in the model. It is worth noting that, according to this table, the two sets of plausible flux distributions associated with (CN) and (Ctrl) have the same topological structure. Indeed, the eight extremal behaviors for (CN) and (Ctrl) have clear combinatorics: first, NADPH oxidation is set either at zero or at its maximal value. Then, a single flux within OAA→PYR (R_{14}), OAA→G3P (R_{15}), G3P→G6P (R_8), peptide hydrolysis (R_{64}) is strongly activated whereas the three other remaining fluxes are blocked.

As discussed in a previous paragraph, flux distributions with no peptide hydrolysis cannot be considered as relevant [21,32]. This suggests the extreme flux distributions A-F in (CN) and (Ctrl) are not biologically relevant. On the contrary, two extreme flux distributions (distributions G and H) are consistent with the

¹These functions are: $v_8 + v_{19}, v_8 - v_{19}, v_{14} + v_{19}, v_{14} - v_{19}, v_{15} + v_{19}, v_{15} - v_{19}, v_{64} + v_{19}, v_{64} - v_{19}$

stoichiometry of the system for the (CN) and (Ctrl) treatments. These two flux distributions consist in optimizing peptide synthesis and hydrolysis after NADPH oxidation. They corresponded to a ratio between peptide hydrolysis and synthesis $\frac{v_{64}}{v_{63}}$ of 0.70 in the (Ctrl) treatment and of 0.47 in the (CN) treatment. These ratios are equal or lower to the maximum ratio of $0.67v_{63}$ [21]. However, in (CN), treatment peptide synthesis was expected to be higher than in Ctrl treatments since in mammals protein synthesis is reported to increase with increasing protein intake (as CN in treatment) [39]. Curiously, on the contrary, in both flux distributions G and H, the total mammary protein synthesis decreased for the (CN) treatment when compared to the (Ctrl) treatment: v_{63} equals 430 or 410 mmol/h/half udder in (Ctrl) and 283 or 282 mmol/h/half udder in (CN).

Study of quantitative contributions of precursors (AIO) for plausible extreme flux distributions

As a further investigation to check the relevance of distributions G and H in the (CN) and (Ctrl) treatments, we studied the quantitative contributions of precursors of output nutrients. This study was inspired by the usual techniques in the field of nutrition - or any domain concerned with organ studies. These techniques consist in computing yield rates to elucidate how an input nutrient may contribute to the composition of an output product, for instance to clarify what proportion of glucose, acetate or alanine taken up by the mammary gland can be recovered in the milk components (lactose, fatty acids, protein) or oxidized and recovered in CO₂ released in blood. To formalize this issue, we first selected carbon as the component according to which the contributions of precursors were to be computed. Then, in order to determine how much carbon introduced into the system through a given input flux can be recovered in the rate of production of an output metabolite, we introduced a precise model for the *allocation of nutrients in measured outputs (AIO)* (see Eq. (2)). As shown in Table 4, the nutrient allocation corresponding to both flux distributions G and H in the (CN) and (Ctrl) treatments provides evidence that glucose is the unique precursor of lactose synthesis. This contradicts studies of dairy cows suggesting that glycerol and perhaps amino acids contribute to lactose synthesis [6], and that fifteen percent of lactose carbon could not derive from glucose [9].

This precise analysis of the origin of carbon in lactose synthesis with AIO suggests that extreme flux distributions G and H have to be rejected, although both flux distributions are consistent with flux variability criteria. We concluded that none of the extreme vertices of the set of plausible flux distributions could be considered biologically relevant with respect to the model and data at hand. Notice that the optimization of any linear combination of metabolic fluxes is either reached by these extreme distributions or reached by the infinite number of flux distributions lying in a face of the simplex. This suggests that the functioning of

the mammary gland cannot be uniquely modeled by the optimization of any linear combination of metabolic fluxes involved in the current model.

Discriminate treatments despite mammary gland flexibility: maxima of AIO on the complete polyhedron of flux distributions

Previous studies suggest that the response of the mammary gland cannot be modeled uniquely by the optimization of a linear objective function of fluxes. However, there are several nonoptimal flux distributions that satisfy the literature-based information that we have used so far ². More generally, many flux distributions compatible with the additional constraints, including the condition on carbon precursors for lactose can be shown for both (CN) and (Ctrl) treatments. Topological arguments prove that an infinite set of flux distributions exist. Nonetheless, without an idea about the exact shape of the space of feasible fluxes (because of the nonlinear nature of the condition of carbon precursors), we cannot select a plausible point within this space. In other words, the available knowledge appears insufficient to determine uniquely a flux distribution of nutrients among the different branches of the proposed model.

To understand the functioning of the mammary gland and despite this difficulty, we introduced a method to estimate the variability of nutrient allocation among pathways on a carbon basis (see Eq. (3)) by computing the range (min-max) of AIO coefficients. As these coefficients are nonlinear functions of flux variables, computing these min and max over the complete space of plausible flux distribution requires solving nonlinear optimization problems. Our dedicated algorithm detailed in the method section scaled properly to the real case that we are studying ³. Interestingly, with this approach, we did not favor any internal functioning of the system since we parsed all objective functions (linear combinations of flux variables), in order to have a complete description of the space of plausible flux distributions.

The min-max tables for allocations of nutrients in the different pathways provided a clearer view of nutrient utilization within the mammary gland. Unlike the functioning of the extremal distributions (G) and (H) shown in Table 4, the contribution of carbon from glucose to lactose carbon was quite variable within the full set of plausible distribution. Indeed, as shown in Table 5, glucose was the precursor of 85% ($\frac{1751}{886}$ or $\frac{866}{1002}$) to 100% of the carbon lactose in the (Ctrl) and (CN) treatments. Therefore, there exist flux distributions such that glucose is not the only precursor of lactose synthesis [6], and these distributions are consistent with quantitative estimations of the ratio of carbon glucose recovered in lactose [9].

²For instance, for the (Ctrl) treatment, there is a distribution such that $v_8 < 591$ and $v_{13} < 266$ are lower than the bounds introduced in [6, 9], while $v_{64}/v_{63} = 0.62 < 0.67$ and glucose is the precursor for 90.79% of the lactose carbon

³All the computations were performed by using *Matlab* release 2011 on a Xeon E5645 multicore computer.

More generally, the intervals of distributions of nutrients in different pathways were used to compare the effects of the (Ctrl) and (CN) treatments. Biologically, in comparison to the (Ctrl) treatment, the (CN) treatment was characterized by a lower proportion of glucose (on a carbon basis) which is oxidized in CO₂, and a larger ratio used for lactose synthesis. This hypothesis can be sustained since the intervals of distribution of carbon from glucose in CO₂ and lactose almost did not overlap in the (Ctrl) and (CN) treatments. More precisely, in the (Ctrl) treatment, from $\frac{396}{1422} = 27.9\%$ to $\frac{565}{1422} = 39.7\%$ of glucose carbon was oxidized to produce CO₂, whereas this ratio ranged from 17.5% to 28.9% for the (CN) treatment. In addition, in the (Ctrl) treatment, 52.8% to 62.3% of the glucose carbon was required to produce lactose carbon, whereas 62.2% to 72% of glucose was required during the (CN) treatment.

Altogether, this analysis allowed us to discriminate the effects of the different treatments whatever the internal functioning of the system may be: the (CN) treatment (increase in protein supply to cows) was characterized by a lower proportion of glucose oxidized in CO₂ than in (Ctrl). It appears to be a suitable strategy to analyze the metabolism flexibility without selecting a precise flux distribution or making any assumption on the internal metabolic fluxes.

As a final study, we used an interior point exploration method to estimate the AIO variability over the boundary of the simplex, rather than the complete space that was explored previously. As both tables appeared to be equal, we concluded that optimized AIO are reached on the boundary of the simplex. However, not all extreme vertices of the simplex are relevant biologically and they do not optimize AIO. This suggests that the flux distributions that optimize an AIO coefficient are placed at the interior of the simplex faces. In the future, it may be interesting to study the biological significance with multi-objective approaches [42] and check whether these can be considered as a "characteristic" point of these faces.

Impact of long-chain fatty acids oxidation over the model predictions

As a last study, we studied the robustness of our conclusion with respect to changes in the modeling of input (net uptake) of long-chain fatty acids (LCFA), C16 and C18, in the tricarboxylic acid (TCA) cycle. In the study of the (Ctrl) and (CN) treatments, such an input of LCFA was not considered in the system since they are not synthesized by the mammary gland and we hypothesized that they are not oxidized within the mammary gland, based on isotope measurements in studies in fed lactating goats and in nonruminants [3,44,45]. This hypothesis was sustained by the measured carbon balance in the (Ctrl) and (CN) datasets [32], which did not require increasing the prediction of CO₂ by introducing any LCFA in the TCA cycle (see also Table 6(a)). Nonetheless, in other contexts and models, LCFA oxidation was either measured in

starvation condition [4] or introduced in order balance carbon consumption and production with respect to CO₂ prediction [22, 24, 50, 52]. This LCFA oxidation hypothesis was for instance retained in models studying the (HB) dataset [22, 24].

From this literature study, it appears that LCFA oxidation may depend on both the environmental and experimental contexts, and no single model can be favored yet. To study the impact of LCFA oxidation, we successively introduced a ratio of LCFA in the TCA cycle (10%, 20%, 25%), and assumed that 50% of C16 output are synthesized within the mammary gland [22] (see Figure 2). Then we performed a complete study of the (Ctrl), (CN) and (HB) datasets for these hypotheses (see Table 6).

For the (Ctrl) treatment, by comparing the predicted and measured CO₂ quantities, we concluded that the hypotheses of 20% and 25% of FA oxidized in the TCA cycle were not in agreement with our experimental data: the increase of CO₂ prediction was too large both when compared to measured CO₂ (see Table 6(a)) and when confronted to the measured increase of 9% of CO₂ deriving from long LCFA in lactating goats in an extreme starvation condition [4]. Similarly, the hypothesis of 0% of LCFA oxidation was not consistent with the (HB) treatment [52].

For all remaining compatible pairs of model and datasets, we first studied the ATP maximization hypothesis. In all cases, our results, shown in Table 6(a), suggest that ATP maximization is not biologically relevant: ATP balance was even larger than the ATP balance of the models where no LCFA was introduced in the TCA cycle (0%).

Then we enumerated extreme flux distributions and studied their biological relevance. As shown in Table 6(b), for each ratio of LCFA in the TCA cycle, the structure of the simplex of plausible flux distributions for the (Ctrl) and (CN) datasets was similar to that shown in Table 3 (0%). Using similar arguments, no extreme distribution could be considered as biologically relevant. When studying the space of solution associated with the (HB) dataset, a more complex structure appeared, based on 13 extreme flux distributions instead of 8. Explicit (although not unique) linear functions that can be optimized to obtain these extreme distributions are detailed in Table 6(b)⁴. More precisely, for the (HB) treatment, we still obtain a division according to the activation of the four fluxes OAA→PYR (R_{14}), OAA→G3P (R_{15}), G3P→G6P (R_8), peptide hydrolysis (R_{64}). Nonetheless, an intricate phenomenon appears. Indeed, measurements imply that OAA→G3P (R_{15}) is very constrained. It cannot vanish at the same time as OAA→PYR (R_{14}) or when NADPH oxidation (R_{19}) is at its maximal value. Therefore, for several optimal conditions related to

⁴These functions are: $v_8 + v_{19}$, $v_8 + v_{19} - 50 v_{15}$, $v_8 - v_{19} - v_{14}$, $v_8 - v_{19} - v_{15}$, $v_{14} + v_{19}$, $v_{14} + v_{19} - 50 v_{15}$, $v_{14} - v_{19}$, $v_{15} + v_{19}$, $v_{15} - v_{19}$, $v_{64} + v_{19}$, $v_{64} + v_{19} - 50 v_{15}$, $v_{64} - v_{19} - v_{14}$, $v_{64} - v_{19} - v_{15}$.

NADPH oxidation (R_{19}), OAA→PYR (R_{14}), G3P→G6P (R_8), peptide hydrolysis (R_{64}), we need to decide whether OAA→G3P (R_{15}) is slightly non-zero, or whether OAA→G3P is assumed to vanish although the optimal is not exactly reached for the initial flux considered. Despite this difference of structure between the (HB) dataset and the (Ctrl) and (CN) datasets, our analysis suggested that extreme distributions were not biologically relevant, independently from the ratio of LCFA oxidation or the dataset under study.

Finally, comparing the variability of the AIO coefficients for the (Ctrl) and (CN) treatments when oxidizing 10% of LCFA in the TCA cycle still suggested that the (CN) treatment is characterized by a lower proportion of glucose (on a carbon basis) which is oxidized in CO₂, and a larger ratio used for lactose synthesis.

Altogether, this study suggests that the main characteristics of the (Ctrl) and the (CN) treatments are robust and could be elucidated despite lacking information on the precise internal behavior of LCFA oxidation.

Discussions

Analyzing the distribution of nutrients in a metabolic network to study the flexibility of a metabolism at the organ level

An important challenge in applicational fields of metabolism studies at the organ level is to understand how the components of inputs are transformed into some expected outputs, under some assumptions about the functioning of a system. To that end, great use is made of comparisons between yield rates describing the allocation of input nutrients within the set of outputs. Nonetheless, to allow a precise comparison of nutrients, these studies require insights on the distribution of matter components across the output of *each* reaction involved in the system. Such information may be provided by several experimental marking techniques (C^{13} MFA) that make use of carbon isotopes [53]. Nonetheless, their application to mammals is challenging [16].

To elucidate how input nutrients are allocated among the output nutrients of a metabolic system despite experimental limitations, we have introduced novel methods which refine the flux balance analysis of a metabolic system related to an organ of a large animal. Our method can be seen as an extension of Flux Variability Analysis [34], where the emphasize is put on *efficiency* or *yield rate* variability rather than on flux variability.

As an example of application, we have studied the mammary metabolism in ruminants (dairy cows)

[31,48,49]. Compared to the conventional models available, our stoichiometric model describes very precisely the variations of energy consumptions as ATP, allowing us to investigate some optimization hypotheses related to energy variations.

As a methodological innovation, we introduced a method to estimate the *Allocation of Inputs in Outputs* (AIO), that is, the ratio of transformation of each input nutrient into outputs, provided that we are given a flux distribution balancing the production and consumption of intermediary metabolites. Two AIO computations are possible: one may fix a linear combination of flux variable to optimize, leading to an extreme flux distribution (i.e. extreme pathways) whose allocations (AIO) are precisely computed with our method. Alternatively, we can reason with regard to the possibly infinite complete convex space of plausible flux distributions, in the spirit of FVA analysis, without solving the set of equations. In this case, our method estimates the minima and maxima of each allocation (AIO) coefficient within the complete space.

The latter point is where our main algorithmic innovation lies. Indeed, AIO coefficients are not linear with respect to the flux variables v . Therefore, they have no reason to be extremal for flux distributions corresponding to vertices, edges or faces of the simplex, unless the gradient of every component of the AIO is a nonzero function. Therefore, we have introduced two algorithms allowing us to compute the extremal values of AIO coefficients when the flux space is bounded. The fastest algorithm requires the exhibition of an analytic expression of the AIO matrix. If this is not possible, a local-search algorithm can be used. These algorithms can be used to check whether the extremal values are reached for flux distributions lying on the boundary of the simplex of plausible flux distributions.

This approach applied to a stoichiometry model taking ATP variations into account permitted a better understanding of ruminant mammary gland metabolism in comparison to previous studies based on similar models and datasets [31,52]. For instance, we were able to characterize the differences in the effects of two treatments, such as the quantitative range of the proportion of glucose which is oxidized in CO₂, or used for lactose synthesis. This information was derived from the carbon composition of each metabolite, expert knowledge which is easily accessible and is therefore compatible with the experimental limitations regarding mammals.

Optimization strategies within tissue or organs

Our study allowed us to revisit optimization-based hypotheses on the functioning of the mammary metabolism. We have provided evidence that flux distributions corresponding to an optimal production of energy (ATP) cannot describe appropriately the metabolism of the mammary gland, as would be the case

for bacterial metabolisms, which tend to optimize biomass-related functions [18,19]. The main reason is that the ATP balance is too large and variable among the responses to different treatments, confirming previous observations [21]. As a consequence, the underlying hypotheses used to drive previous studies of mammary metabolism have to be carefully reformulated [22, 48, 49].

As an alternative, we have hypothesized that ATP balance remains nonoptimal and almost constant in response to several treatments, and we have introduced in the model a recent estimation for this quantity [21]. Our goal was to check whether the observed responses to the system could be explained by the optimization of a linear combination of fluxes, that is, an extremal vertex of the simplex of plausible flux distributions. This hypothesis was rejected, first because it led to non-realistic orders of magnitude for some fluxes such as peptide hydrolysis, and second because the precursors of some components such as glucose were not biologically relevant [6,9]. Notably, it was necessary to trace back the quantitative origin of the nutrients to complete this analysis and reject all extreme flux distributions, illustrating the advantage of the AIO approach.

Therefore, no optimization of a linear combination of flux variables could be found to uniquely describe the metabolism of the mammary gland, a multicellular organ, as an extreme flux distribution of our model. To overcome this limitation of FBA-inspired analyses at the mammalian tissue or organ level, we studied the variability of AIO coefficients by introducing the computation of min-max ranges of AIO. This led to a general overview of the effects of treatments without precluding any steady state internal behavior.

Benefits of studying the variability of the allocation of input in output (AIO) and future model refinements

The study of the variability of AIO, that is, intervals of allocation of nutrients, made it possible to distinguish the metabolism of the mammary gland in two different nutritional conditions corresponding to an increase in protein supply (Ctrl and CN). Notably, the intervals of allocation of glucose in CO₂ and lactose were different in the two nutritional conditions, reflecting the mammary gland's metabolic flexibility. Nonetheless, several improvements can be made to the model.

First, the simplex of plausible distributions could be reduced by introducing knowledge currently available on the kinetic bounds for enzymatic activities, introduced in previous numerical models [22]. This may for instance prevent some non-constrained behaviors shown in Table 5, such as the possible nonzero contribution of acetate and β -hydroxybutyrate (BHBA) carbon to lactose. The main issue in this area will be to aggregate the enzymatic knowledge to fit with the format of the reactions in the model, such as the reaction R_6 which

transforms G3P to PYR under the regulation of five different enzymes, together with NAD⁺ and ATP availability. Therefore, proposing relevant maximal values requires the use of advanced methods for model reductions.

A second improvement of the model is dependent on the production of additional observations to clarify the set of possible behaviors of the system. Particularly, the contribution of acetate and β -hydroxybutyrate (BHBA) to lactose shown in Table 5 remains questionable in ruminants since it suggests that the carbon of acetyl-CoA could leave the TCA cycle of the mitochondria through the OAA \rightarrow G3P pathway. This could be explained by the fact that the carbon allocation in the model did not take account of the carbon positions. However, one isotope model and one study of gluconeogenesis in man suggested that labeled carbon within glucose came from [2-¹⁴C]-acetate (through ¹⁴C-acetylCoA in TCA) [14, 27]. It remains challenging to obtain such precise information for large ruminants such as dairy cows. Nevertheless, the model could be improved by modeling with additional constraints the few items of knowledge we have on the flux distribution of positioned carbons.

The last improvement of the model consists of including constraints that do not correspond to rule-based metabolic equations. More precisely, in several numerical models, the effects of external fluxes were introduced to predict the response of the mammary gland in a consistent way [13, 25, 50]. Among external fluxes, we noticed the regulatory effect of long-chain fatty acids on the activity of enzymes involved in fatty acid synthesis, although the mechanisms are not clearly understood [10]. Although some extensions of the FBA formalism to dynamic regulation exist [15], the data availability was insufficient to apply the extensions in the model. Therefore, existing regulations of enzyme activities cannot be included directly in the stoichiometric model. To overcome this limitation, we plan to investigate the effects of external fluxes by introducing additional constraints in the model, to check whether the nutrient output in the milk can be predicted from the nutrient input.

Conclusion

We have introduced a method in the framework of flux-balance analysis of a metabolic network. As a main novelty, our approach allows studying the variability of efficiencies (or yield rates) of a metabolic model provided with input-output measurements. More precisely, our approach allows a quantitative estimation of the minimum and maximum proportions of the carbon quantity of each input nutrient which is recovered in each output component of the system. The main innovation is to propose a method which does not require determining the quantitative distribution of nutrients between the branches of the system. To that end,

we have performed a parsing of the space of flux distributions which are compatible with both the model stoichiometry and input-output measurements.

This method was applied to study the response of the mammary gland to several treatments. It allowed us to distinguish two different metabolic responses of the system, corresponding to two nutritional situations and accurately reflecting metabolic flexibility. Overall, our method appears to be configured to study the variability of the yield rates of a metabolic system at the multicellular or organ level without making any hypothesis on the internal behavior of the system.

Method

Flux modeling based on Flux Balance Analysis

Metabolic models are described according to the generic framework of Flux Balance Analysis [28, 30, 37]. Metabolites split into three non-intersecting sets. First, *input metabolites* are gathered in set \mathcal{I} . The rate vector of all input fluxes (with in the form " $\rightarrow I$ ") is denoted by $v_I \in \mathbb{R}^p$. Second, *output metabolites* are denoted by \mathcal{O} . We denote by $v_O \in \mathbb{R}^q$ the rate vector of the corresponding output fluxes (in the form " $O \rightarrow$ "). Finally, *intermediary metabolites* (or *pivots*) are denoted by \mathcal{P} , and its cardinality is denoted by n . The complete set of metabolites is $\mathcal{M} = \mathcal{I} \cup \mathcal{P} \cup \mathcal{O} = \{m_1 \dots m_{p+n+q}\}$.

Let $\mathcal{R} = \{r_1 \dots r_t\}$ be the set of t reactions which produce some metabolites while degrading others. The rate vector of these reactions is denoted by $v \in \mathbb{R}^t$. A reaction r_j has the form $\sum_{m \in \mathcal{I} \cup \mathcal{P}} s_{m,j} m \rightarrow \sum_{m' \in \mathcal{P} \cup \mathcal{O}} s'_{m',j} m'$ where $s_{m,j}$ and $s'_{m',j}$ are input and output stoichiometric coefficients. The metabolic system is described by the so-called *stoichiometric matrix* $M = (a_{i,j})_{i \leq p+n+q, j \leq t}$, where $a_{i,j} = s'_{m_i,j} - s_{m_i,j}$.

As a usual assumption, intermediary metabolites cannot accumulate in the cell: the consumption and production flux rates of intermediary metabolites are balanced. A linear constraint is derived on the rate vector $v \in \mathbb{R}^t$. Additional biological knowledge about the distribution of fluxes within the system is modeled by a matrix B . Overall, plausible flux distributions v satisfy the following equations:

$$M \begin{pmatrix} -v_I \\ v \\ v_O \end{pmatrix} = 0 \text{ [Stoichiometry constraints]; } \quad Bv = 0 \text{ [Biological knowledge constraints].} \quad (1)$$

Assuming that reactions are irreversible and fluxes have physical upper-bounds, the set of solutions to Eq.(1) is a convex polyhedron, called *a simplex, of plausible flux distributions*. The simplex can be described by means of its vertices, edges or faces [28,30,37], which are related to the optimization of objective functions and have shown their biological significance in several contexts [17]. In the example described below, vertices of the simplex (i.e. *extreme pathways*) were computed with standard linear-based methods in the bound case,

and with probabilistic methods in the unbound case. As a refinement of these methods, we used a random sampling of the set of possible linear objective functions to obtain a complete description of the set of extreme pathways. This approach was efficient since the dimension of the space of feasible distribution is quite small, because it is strongly constrained by the input-output vectors v_I and v_O .

Modeling the quantitative contribution of input metabolites to output nutrients: AIO

In studies at the organ level such as in nutrition, the choice of a plausible flux distribution v aims to elucidate how an input nutrient may contribute to the composition of an output product. To formalize this issue, we introduce a *component* according to which the allocation of input nutrients will be computed. For dietary applications, carbon is a relevant element since it appears in the composition of all metabolites in the system. Nitrogen is less generic but more relevant when specifically studying nitrogen metabolism. Our issue therefore reads as follows: how much carbon introduced into the system through a given input flux can be recovered in the rate of production of an output metabolite?

We introduce the following formalism. Let $c \in \mathbb{R}^{p+n+q}$ be a vector describing the *component composition* of metabolites (its carbon composition for instance). Assume that the stoichiometry of each reaction in the system satisfies a matter-invariant property with respect to the component, that is: $\sum_{m_i \in \mathcal{M}} a_{i,j} c(m_i) = 0$ for every reaction $r_j \in \mathcal{R}$. Of the total component mass provided as a substrate to a reaction r , only a part contributes to the composition of a given product m . Let $In^r(m)$ denote this ratio. Assuming that the system follows a *proportional matter distribution*, a consequence of the mass-invariant property is that $\sum_{m \in \mathcal{M}} In^r(m) = 1$ for all $r \in \mathcal{R}$.

Let v be a plausible flux distribution and a, solution to Eq.(1). Let m be an input metabolite, for example glucose. The rate of the flux of component mass brought into the system by m equals $C(m)v_I(m)$. We denote by $x_O[v, m] \in \mathbb{R}^q$ the vector of proportions of input component fluxes recovered in each output flux (for instance, the proportion of carbon from the input glucose appearing in the composition of the CO2 in blood). The coefficients of $x_O[v, m]$ are called the *allocation of the metabolite m in output nutrients*.

To determine these ratios, we introduce $x_P[v, m] \in \mathbb{R}^n$ the *vector of ratios of input component fluxes* which are required to produce each intermediary metabolite m before its breakdown into other metabolites. The law of matter conservation and the assumption that intermediary metabolites do not accumulate put several constraints on $x_P[v, m]$ and $x_O[v, m]$, which are detailed in Table 6. We deduce that the vector of allocations $x_O[v, m]$ is a solution to the equation $D_2[v]x_I^{(m)} = -D_1[v] \begin{pmatrix} x_P[v, m] \\ x_O[v, m] \end{pmatrix}$ for the input parameter vector $x_I^{(m)} = (0, \dots, 0, C(m)v_I(m), 0, \dots, 0)$, where D_1 and D_2 are defined in Table 7. To elucidate whether

this system of matrices determines uniquely $x_O[v, m]$, we noticed that D_1 is a square matrix such that all diagonal coefficients are equal to 1 and the others are nonpositive. This family of matrices is named the M -matrix in [2], and it has been shown that such a matrix is invertible. We deduce that D_1 is invertible so that the vectors of allocations $x_O[v, m]$ and $x_P[v, m]$ are determined uniquely. With these formulas, when the flux distribution v has been chosen, any algorithm or method to inverse matrix D_1 can be used.

Overall, the *complete table showing the distribution of nutrient inputs in nutrient outputs* is defined to be the $q \times p$ matrix $AIO[v]$, the columns of which are $x_O[v, m_1], \dots, x_O[v, m_p]$, as follows:

$$AIO[v] = x_O[v, m], \quad \begin{pmatrix} x_P[v, m] \\ x_O[v, m] \end{pmatrix} = -D_1[v]^{-1} D_2[v] \begin{pmatrix} C(m_1)v_I(m_1) & & 0 \\ & \ddots & \\ 0 & & C(m_p)v_I(m_p) \end{pmatrix}. \quad (2)$$

With these formulas at hand, as soon as the flux distribution v has been chosen, any algorithm or method to inverse the matrix D_1 can be used. Altogether, the computation of a complete AIO table takes $O((n + p + q)^3 + np)$ operations for a given flux distribution v .

Computing extrema of the AIO coefficients among the complete polyhedron of plausible flux distributions: solving nonlinear optimization problems

Computing the extrema of the AIO coefficients among the convex polyhedron of plausible flux distributions v , leads to an optimization problem in a nonlinear context. The problem rewrites in an optimization scheme, leading to $2q \times p$ optimization problems of the form

$$\forall i \leq q, \forall k \leq p, (\text{Max/Min})\text{imize} \left\{ AIO[v][i, k]; v \in \mathbb{R}^n, M \begin{pmatrix} -v_I \\ v \\ v_O \end{pmatrix} = 0, Bv = 0, v \geq 0 \right\}. \quad (3)$$

These problems are nonlinear programming (NLP) problems [7] since they aim to optimize a nonlinear objective function over a possibly nonbounded simplex search space.

When both the simplex search space was bounded and an analytical expression for $AIO[V]$ as a function of all the free variables of the eq.(1) solution space was producible with formal algebra software [46], the Matlab *fmincon* function was used to solve the NLP problem. Among all available optimization routines, the best performance was achieved by the interior point algorithm, which has a polynomial complexity for these problems [12]. Other optimization routines were used to confirm our results, but the computation time was much longer.

When an analytical expression was not available, we implemented an alternative strategy making use of some local search routines together with the fact that $AIO(v)$ can be computed for any v in a cubic time. First, we compute the Chebyshev ball $B(x_0, r)$ of the simplex [11]. Any point x of the n -dimensional

ball is uniquely identified by a vector $(\theta_1, \dots, \theta_n)$ with $\theta_i \in [0, 2\pi[$ for $1 \leq i < n$ and $\theta_n \in [0, \pi[$. The intersection $P(x)$ of the half line $[x_0, x)$ with the simplex of plausible flux distributions is unique if it exists. Conversely, for any point y on the boundary of the simplex, there exists a unique $x \in B(x_0, r)$ such that $y = P(x)$. By extension, we say that $(\theta_1, \dots, \theta_n)$ is a parameterization of the boundary of the simplex. This parameterization is used to determine a discretization of the simplex boundary. Given an integer $p > 0$, every θ_i is assumed to be of the form $2\pi j/p$, for $j \in \{0, \dots, p-1\}$, whereas θ_n rewrites as $\pi j/p$. The finite neighborhood of any point y on the boundary, denoted by $\mathcal{N}(y)$ is obtained by modifying each single coordinate of the parameterization $(\theta_1, \dots, 2\pi j/p, \dots, \theta_n)$ of y : for instance, among the maximal $2n$ neighbors, we find the point with parameterization $(\theta_1, \dots, 2\pi(j-1)/p, \dots, \theta_n)$ and $(\theta_1, \dots, 2\pi(j+1)/p, \dots, \theta_n)$. We can finally apply a local search algorithm based on a dichotomy principle to improve the discretization level and reach the solution to the optimization problem on the simplex boundary.

The same algorithm is extended to the full space by introducing a supplementary coordinate standing for the distance between $P(x)$ and the center x_0 of the Chebyshev ball. This makes it possible to check whether the optimum identified previously lies on the boundary of the simplex.

Analysis workflow

A web application dedicated to the computation of AIO for metabolic networks is freely available⁵. It enables us to build a stoichiometric model, define its inputs and outputs, solve the associated FBA problem and finally compute the AIO of the selected flux distribution. Offline tools enabling the computation of the extrema of the AIO are also provided in a companion web page⁶.

The complete workflow of analysis requires the online webpage together with the use of several environments. First, *php* is required for the computation of extremal vertices of the convex polyhedron of plausible flux distributions, including the case when this space is not bounded. Second, *SAGE* is used to performed formal algebra tasks involved in the computation of AIO matrices. Finally, *matlab* is needed to compute the extrema of the AIO coefficients. The main functionalities of the workflow are depicted in Fig. 2.

Mammary gland stoichiometry model

As the basis of the ruminant mammary gland metabolism that is studied in this paper, we used a generic model of mammary metabolism [48, 49]. This generic model contained 54 reactions involving six interme-

⁵<http://nutritionanalyzer.genouest.org>

⁶<http://nutritionanalyzer.genouest.org/SupplementaryMaterial>. The access to the pre-formatted model and datasets supporting the results is planned to be on request. A valid access is provided by: login : SuppMat; password : tampus

diary metabolites, called carbon-chain *pivots*, at cross-over points between metabolic pathways, chosen to describe the main possible conversions between metabolites: oxaloacetate (*OAA*), α -ketoglutarate (α -*KG*), pyruvate (*PYR*), acetyl coenzyme A (*AcoA*), glucose (*GLC*) and serine (*SER pivot*). The reactions also take into account the variations in ATP, four cofactors (*NADHc*, *NADHm*, *NADPH* and *FADH2*) and other metabolites (*O2*, *CO2* and *NH3*).

The model was first extended and detailed in order to include the reactions included in other models of mammary metabolism [22,31,49,52]. Reactions for lactose synthesis, milk protein, fatty acids and glycerol-3P of triglycerides were included in the model, as well as four new pivots. Mitochondrial (*mAcoA*) and cytosolic (*cAcoA*) acetyl-CoA were used to distinguish between utilization of *cAcoA* for fatty acid (*FA*) synthesis and utilization of *mAcoA* in the tricarboxylic acid (*TCA*) cycle. In addition, glucose-6-phosphate (*G6P*) and triose-phosphate (*G3P*) were introduced to account more specifically for the transformation of glucose through the pentose phosphate pathway, lactose synthesis and glycerol-3P utilization. For lipid metabolism, three new intermediary metabolites were introduced (*Glycerol-3P*) and two fatty acid primers (*Primer (C2:0)CoA* and *Primer (C4:0)CoA*). They correspond to the primers of two (from acetate) or four (from β -hydroxybutyrate; *BHBA*) carbon-units necessary to initiate fatty acid synthesis before elongation, since in ruminant mammary glands fatty acids are synthesized from acetate or *BHBA* and not glucose. Other additional reactions were included to describe milk synthesis, such as NADPH synthesis through the isocitrate dehydrogenase pathway. The input (net uptake) of long-chain fatty acids was not considered in the system since these are not synthesized by the mammary gland and we hypothesized that they are not oxidized within the mammary gland of lactating animals, as measured in well-fed goat and nonruminant through isotopes [3, 45]. This hypothesis was sustained by carbon balance measurements based on experimental measures of CO₂ [32]. This latter hypothesis was relaxed in a second step (see Figure 2) to study how our results are impacted by the introduction of oxidation of long-chain fatty acid within TCA cycle, as modeled in alternative contexts [22, 24, 50, 52]. In addition, the action of long-chain fatty acids to regulate enzyme activities involved in C(4:0) ... C(16:0) production was taken into account by the fact that C(4:0), ... C(16:0) fluxes of synthesis are explicitly given in the dataset.

Protein synthesis was summarized by the number of peptide links synthesized (peptide synthesis) and the ATP required for each peptide synthesis was fixed at 5 in the present model [48]. Similarly, protein degradation was summarized by the number of hydrolyzed peptide links (peptide hydrolysis). The ATP consumption for each peptide hydrolysis was set at a single ATP [33].

Finally, four output reactions were included in the model in order to calculate the overall allocations in

carbon (R_{137}), nitrogen (R_{138}), CO_2 (R_{139}) and Serine (R_{135}).

As a last step, this model was restricted to the case of ruminants. Some reactions, such as fatty acid oxidations (R_{20} to R_{23}) and pyruvate synthesis through malic enzyme (R_{66}) and some input and output fluxes present in the generic stoichiometric model [48] were not considered to occur in ruminant mammary glands (fructose (R_{25}); glucose production from $G6P$ neoglucogenesis (R_{10}), acetone synthesis (R_{69}), $BHBA$ synthesis (R_{70}), acetoacetate synthesis (R_{71}), and ureogenesis (R_{50}).

Altogether, this mammary gland model contained 140 reactions, involving eleven intermediary metabolites, four cofactors and ATP, CO_2 , O_2 and NH_3 . The stoichiometry of the system implies that the balance (production minus consumption) of these four last outputs can be uniquely deduced from the choice of a plausible flux distribution. The full model is shown in Fig. 1 and the corresponding stoichiometric model is provided in the supplementary webpage (SBML format).

Additional information: datasets, inputs, outputs, additional knowledge, component

Two datasets [32,39] were used, in relation to the response of dairy cow mammary gland metabolism to some increases in the protein supply: a control diet (Ctrl) and increased protein supply through casein infusion into the duodenum (CN). A complementary dataset (HB), previously used in a mechanistic model of the mammary metabolism, was used as a reference response [22]. Datasets are shown in Table 1. All data were converted into mmol/h/half udder.

Fluxes of nutrients taken up on a net basis by the mammary gland were considered as inputs whereas fluxes of nutrients produced in milk, such as lactose, milk protein, milk fatty acid arranged in triglycerides or CO_2 produced and released in blood, were considered as outputs. Special attention was paid to amino acid inputs and outputs, milk protein output, fatty acids synthesized within the mammary gland and total triglyceride output according to established calculation rules [22,47]. Milk proteins were modeled by their number of peptide links (peptide output). The amino acid inputs or outputs corresponded to the balance between their real net uptake minus their contribution to peptide outputs (utilization in milk protein). When the balance of an amino acid was positive, it corresponded to an input flux. When it was negative, it corresponded to an output flux (see Table 1). Two input reactions were also assumed to be inactive in all data sets since the corresponding net uptake were not measured in the datasets (Table 1): *propionate* input (R_{28}) and triglyceride hydrolysis (R_{65}) in *long-chain fatty acids* input.

Eight additional linear constraints were introduced to the fluxes to ensure that the model was relevant

biologically [22]. These are detailed in Table 4 and allow building matrix B appearing in Eq.(1)⁷. The rate of fatty-acid acylCoA ($C(n:m)$ -acylCoA) was fixed to be three times higher than the triglyceride output, assuming that 100% of milk lipids are triglycerides: it included all fatty acid acylation, that is, fatty acid synthesized in the mammary gland and long fatty acid taken up. 50% of fatty acids were assumed to be synthesized from acetate, i.e. from *primer* $C(2:0)CoA$ and the other 50% from BHBA (*primer* $C(4:0)CoA$), with the exception of $C(4:0)$ fatty acid, which was supposed to be synthesized only from BHBA. It was assumed that 30% of NADPH was generated by the isocitrate deshydrogenase pathway and 70% of NADPH was generated in the pentose phosphate pathway [22].

Carbon was the component chosen to compute and analyze the allocation of nutrient inputs in nutrient outputs. The stoichiometry of the metabolic network satisfies the mass-invariant property according to this component. This required a very precise description of CO₂ production and consumption in the model reactions and was validated with external measurements of the CO₂ balance provided with the datasets.

Availability of supporting data

The datasets supporting the results of this article are included within the article (Table 1). The SBML file for the mammary gland model and tools enabling the computation of all results are provided in a companion web page: <http://nutritionanalyzer.genouest.org/SupplementaryMaterial>.

Author's contributions

O.A.-A., J.B. and A.S. formalized the computation AIO as an optimization issue. O.A.-A. and J.B. conceived the algorithms to compute AIO. O.A.-A. performed experimentations. O.A.-A., S.L. and J.V.-M. built the stoichiometric model. S.L. prepared experimental datasets and performed the biological interpretation of the results. A.S. and S.L. studied the structure on the simplex of plausible flux distributions. All authors contributed to the redaction of the paper.

Acknowledgements

Part of this program was supported by the French National Agency for Research (ANR-10-BLANC- 0218). The authors wish to thank Damien Eveillard for fruitful discussions.

⁷Mathematically, these constraints can be derived as: $0.7v_{82} - 0.3v_{58} = 0$; $v_{51} - v_{52} = 0$; $v_{54} - v_{90} = 0$; $v_{55} - v_{91} = 0$; $v_{86} - v_{92} = 0$; $v_{87} - v_{93} = 0$; $v_{88} - v_{94} = 0$; $v_{56} - 3v_{62} = 0$.

References

1. V. Acuña, A. Marchetti-Spaccamela, M-F. Sagot, and L. Stougie. A note on the complexity of finding and enumerating elementary modes. *Biosystems*, 99(3):210–214, 2010.
2. O Adamou. *Study of some epidemiological models by the methods of computer algebra*. PhD thesis, University of Rennes-1 (France) and UAM (Niger), 2009.
3. E. F. Annison, J. L. Linzell, S. Fazakerley, and B. W. Nichols. The oxidation and utilization of palmitate, stearate, oleate and acetate by the mammary gland of the fed goat in relation to their overall metabolism, and the role of plasma phospholipids and neutral lipids in milk-fat synthesis. *Biochem. J.*, 102(3):637–647, Mar 1967.
4. E. F. Annison, J. L. Linzell, and C. E. West. Mammary and whole animal metabolism of glucose and fatty acids in fasting lactating goats. *J. Physiol. (Lond.)*, 197(2):445–459, Jul 1968.
5. D. A Beard, S-D. Liang, and H. Qian. Energy balance for analysis of complex metabolic networks. *Biophysical journal*, 83(1):79–86, July 2002.
6. B J Bequette, N E Sunny, S W El-Kadi, and S L Owens. Application of stable isotopes and mass isotopomer distribution analysis to the study of intermediary metabolism of nutrients. *J Anim Sci*, 84 Suppl:E50–9, 2006.
7. Dimitri P Bertsekas. *Nonlinear Programming*. Cambridge, MA.: Athena Scientific, 1999.
8. M Beurton-Aimar, B Beauvoit, A Monier, F Vallee, M Dieuaide-Noubhani, and S Colombie. Comparison between elementary flux modes analysis and ¹³C-metabolic fluxes measured in bacterial and plant cells. *BMC Syst Biol*, 5:95, 2011.
9. R Bickerstaffe, E F Annison, and J L Linzell. The metabolism of glucose, acetate, lipids and amino acids in lactating dairy cows. *The Journal of Agricultural Science*, 82(01):71–85, 1974.
10. P. Blavy, F. Gondret, H. Guillou, S. Lagarrigue, P.G.P. Martin, J. van Milgen, O. Radulescu, and A. Siegel. A minimal model for hepatic fatty acid balance during fasting: Application to {PPAR} alpha-deficient mice. *Journal of Theoretical Biology*, 261(2):266 – 278, 2009.
11. N D Botkin and V L Turova-Botkina. An algorithm for finding the chebyshev center of a convex polyhedron. *Applied Mathematics and Optimization*, 29(2):211–222, 1994.
12. R H Byrd, J C Gilbert, and J Nocedal. A trust region method based on interior point techniques for nonlinear programming. *Mathematical Programming*, 89(1):149–185, 2000.
13. J P Cant and B W McBride. Mathematical analysis of the relationship between blood flow and uptake of nutrients in the mammary glands of a lactating cow. *The Journal of dairy research*, 62(3):405–22, August 1995.
14. A. Consoli, F. Kennedy, J. Miles, and J. Gerich. Determination of Krebs cycle metabolic carbon exchange in vivo and its use to estimate the individual contributions of gluconeogenesis and glycogenolysis to overall glucose output in man. *J. Clin. Invest.*, 80(5):1303–1310, Nov 1987.
15. M W. Covert, C. H. Schilling, and B. Palsson. Regulation of gene expression in flux balance models of metabolism. *Journal of Theoretical Biology*, 213(1):73 – 88, 2001.
16. S B Crown, W S Ahn, and M R Antoniewicz. Rational design of ¹³C-labeling experiments for metabolic flux analysis in mammalian cells. *BMC Syst Biol*, 6:43, 2012.
17. J S Edwards, R U Ibarra, and B O Palsson. In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. *Nat. Biotechnol.*, 19(2):125–130, 2001.
18. J S Edwards and B O Palsson. The Escherichia coli MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences of the United States of America*, 97(10):5528–33, May 2000.
19. D A Fell and J R Small. Fat synthesis in adipose tissue. An examination of stoichiometric constraints. *Biochemical Journal*, 238:781–786, 1986.
20. A Ghozlane, F Bringaud, H Soueidan, I Dutour, F Jourdan, and P Thebault. Flux Analysis of the Trypanosoma brucei Glycolysis Based on a Multiobjective-Criteria Bioinformatic Approach. *Adv Bioinformatics*, 2012:159423, 2012.
21. M Hanigan, J France, S Mabjeesh, W C McNabb, and B Bequette. High rates of mammary tissue protein turnover in lactating goats are energetically costly. *Journal of Nutrition*, 139:1118–1127, 2009.

22. M D Hanigan. A mechanistic model of mammary gland metabolism in the lactating cow. *Agricultural Systems*, 45(4):369–419, 1994.
23. M D Hanigan, L A Crompton, B J Bequette, J A N Mills, and J France. Modelling mammary metabolism in the dairy cow to predict milk constituent yield, with emphasis on amino acid metabolism and milk protein production: model evaluation. *Journal of theoretical biology*, 217(3):311–30, August 2002.
24. M D Hanigan, L A Crompton, J A Metcalf, and J France. Modelling mammary metabolism in the dairy cow to predict milk constituent yield, with emphasis on amino acid metabolism and milk protein production: model construction. *Journal of theoretical biology*, 213(2):223–39, November 2001.
25. M.D. Hanigan, L.A. Crompton, C.K. Reynolds, D. Wray-Cahen, M.A. Lomax, and J. France. An integrative model of amino acid metabolism in the liver of the lactating dairy cow. *Journal of Theoretical Biology*, 228(2):271 – 289, 2004.
26. S. Hardy and P. N Robillard. Modeling and simulation of molecular biology systems using petri nets: modeling goals of various approaches. *Journal of bioinformatics and computational biology*, 2(4):595–613, December 2004.
27. J. Katz. Determination of gluconeogenesis in vivo with ^{14}C -labeled substrates. *Am. J. Physiol.*, 248(4 Pt 2):R391–399, Apr 1985.
28. K. J Kauffman, P. Prakash, and J. S Edwards. Advances in flux balance analysis. *Current opinion in biotechnology*, 14(5):491–6, October 2003.
29. M. Lakshmanan, G. Koh, B. K S Chung, and D-Y. Lee. Software applications for flux balance analysis. *Briefings in bioinformatics*, November 2012.
30. A Larhlimi and A Bockmayr. Minimal Metabolic Behaviors and the Reversible Metabolic Space. Technical report, Matheon, Berlin, Germany, 2009.
31. S Lemosquet, O Abdou-Arbi, A Siegel, J Guinard-Flament, J Van Milgen, and J Bourdon. A generic stoichiometric model to analyse the metabolic flexibility of the mammary gland in lactating dairy cows. In P Faverdin D. Sauvant J. Van Milgen and N Friggens, editors, *Modelling nutrient digestion and utilization in farm animals*. Wageningen Academic Publishers, 2010.
32. S Lemosquet, G Raggio, G Lobley, H Rulquin, J Guinard-Flament, and H Lapierre. Whole-body glucose metabolism and mammary energetic nutrient metabolism in lactating dairy cows receiving digestive infusions of casein and propionic acid. *Journal of Dairy Science*, 92(12):6068–6082, 2009.
33. G E Lobley. Energy metabolism reactions in ruminant muscle: responses to age, nutrition and hormonal status. *Reprod Nutr Dev*, 30(1):13–34, 1990.
34. R Mahadevan and C H Schilling. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic engineering*, 5(4):264–76, October 2003.
35. Y-G. Oh, D-Y. Lee, S. Y. Lee, and S. Park. Multiobjective flux balancing using the NISE method for metabolic network analysis. *Biotechnology progress*, 25(4):999–1008, 2009.
36. J. D Orth, I. Thiele, and B. ØPalsson. What is flux balance analysis? *Nature biotechnology*, 28(3):245–8, March 2010.
37. J A Papin, J Stelling, N D Price, S Klamt, S Schuster, and B O Palsson. Comparison of network-based pathway analysis methods. *Trends Biotechnol.*, 22:400–405, 2004.
38. T Pfeiffer, I Sanchez-Valdenebro, J C Nuevo, F Montero, and S Schuster. Metatool: for studying metabolic networks. *Bioinformatics*, 15(3):251–257, 1999.
39. G Raggio, S Lemosquet, G Lobley, H Rulquin, and H Lapierre. Effect of casein and propionate supply on mammary protein metabolism in lactating dairy cows. *Journal of Dairy Science*, 89:4340–4351, 2006.
40. C H Schilling, D Letscher, and B O Palsson. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *Journal of theoretical biology*, 203(3):229–48, April 2000.
41. C H Schilling, S Schuster, B O Palsson, and R Heinrich. Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnology progress*, 15(3):296–303, 1999.
42. R. Schuetz, N. Zamboni, M. Zampieri, M. Heinemann, and U. Sauer. Multidimensional optimality of microbial metabolism. *Science*, 336(6081):601–604, May 2012.

43. R A Scott, D E Beuman, and J H Clark. Cellular gluconeogenesis by lactating bovine mammary tissue. *J. Dairy Sci.*, 59(1):50–56, January 1976.
44. P. R. Shorten, T. B. Pleasants, and G. C. Upreti. A mathematical model for mammary fatty acid synthesis and triglyceride assembly: the role of stearoyl CoA desaturase (SCD). *J. Dairy Res.*, 71(4):385–397, Nov 2004.
45. G. H. Smith, B. Crabtree, and R. Smith. *Energy metabolism in the mammary gland*, pages 121–141. Elsevier Amsterdam; New York, 1983.
46. W.A. Stein and Others. *Sage Mathematics Software (Version 5.5)*. The Sage Development Team, 2012.
47. H E Swaisgood. *Handbook of Milk Composition*, volume 1, chapter Protein an, pages 464–471. Academic Press, San Diego, 1995.
48. J Van-Milgen. Modeling biochemical aspects of energy metabolism in mammals. *The Journal of Nutrition*, 132(10):3195–3202, 2002.
49. J Van-Milgen, F Gondret, and D Renaudeau. The use of nutritional models as a tool in basis research. In W B Souffrant and C C Metges, editors, *Progress in research on energy and protein metabolism*, pages 259–263, 2003.
50. V. Volpe, J. P. Cant, R. C. Boston, P. Susmel, and P. Moate. Development of a dynamic mathematical model for investigating mammary gland metabolism in lactating cows. *The Journal of Agricultural Science*, 148(01):31, September 2010.
51. K. Voss, M. Heiner, and I. Koch. Steady state analysis of metabolic pathways using Petri nets. *In silico biology*, 3(3):367–87, January 2003.
52. G C Waghorn and R L Baldwin. Model of metabolite flux within mammary gland of the lactating cow. *Journal of Dairy Science*, 67:531–544, 1984.
53. W. Wiechert. 13C Metabolic Flux Analysis. *Metabolic Engineering*, 3(3):195–206, 2001.

Figures

Figure 1 - Simplified view for the stoichiometric model of ruminant mammary metabolism with no long-chain fatty acid oxidation

The complete model is detailed in a SBML supplementary file. Eleven nodes are considered as pivots (green nodes), that is, intermediary metabolites which are not accumulated in the cell. The model is built by adding 26 specific reactions to the ruminant mammary to the generic model of [48]. A treatment is characterized by a set of input nodes (yellow nodes), quantitatively described in mmol/h/half udder. Nutrients contained in the milk are considered as output nodes. The node "Fatty acid synthesis" is an abstraction for 14 reactions corresponding to C(4:0), C(6:0), ... C(16:0) from primerC2 and primerC4. Note that, depending on the treatment, the role of nonessential amino acids may change from input to output according to the balance of the amino-acid considered. The node "peptide" summarizes the ATP cost of protein synthesis and protein degradation. Notably, the stoichiometry of reactions was adjusted in order to balance carbon exchanges, including CO₂. This is a key point in order to compute exact allocation tables in the following stages. Additional literature-based information allows us to generate additional linear constraints on some reactions fluxes.

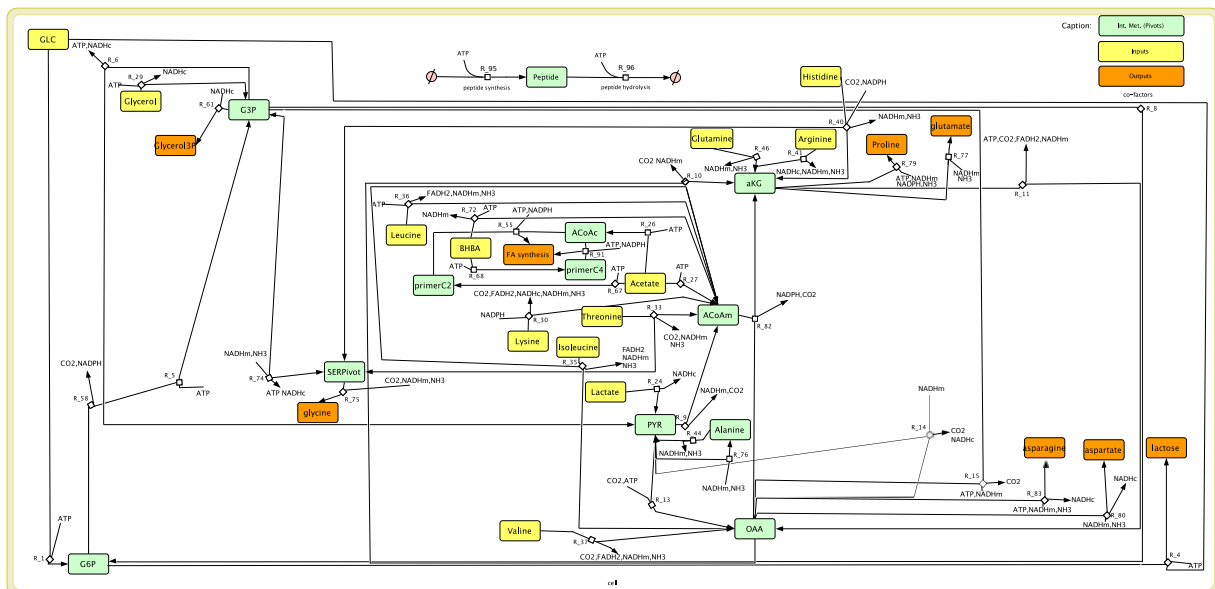


Figure 2 - Modeling long-chain fatty acids oxidation (LCFA) in tricarboxylic acid (TCA) cycle

Introducing LCFA (C(16:0), C(18:0)) oxidation in the TCA cycle may be required to consistently model the response to several treatments, such as the (HB) dataset (Table 1). In this case, the model shown in Figure 1 is extended by introducing inputs of C(16:0) (R_{141}) and C(18:0) (R_{142}), and output of C(18:0) (R_{143}). C(16:0) oxidation (R_{20}) and C(18:0) oxidation (R_{21}) are modified accordingly.

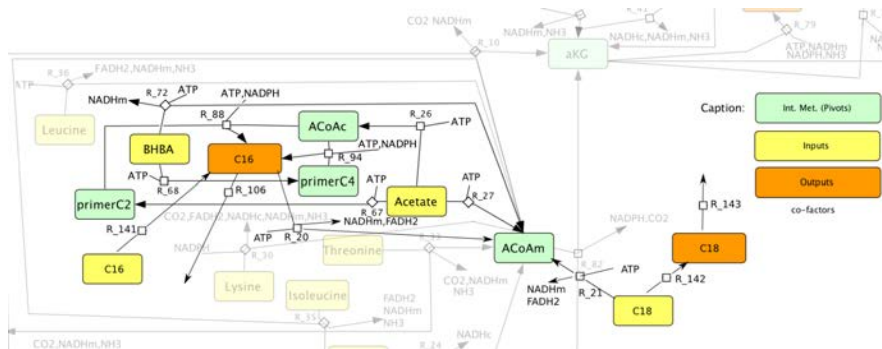
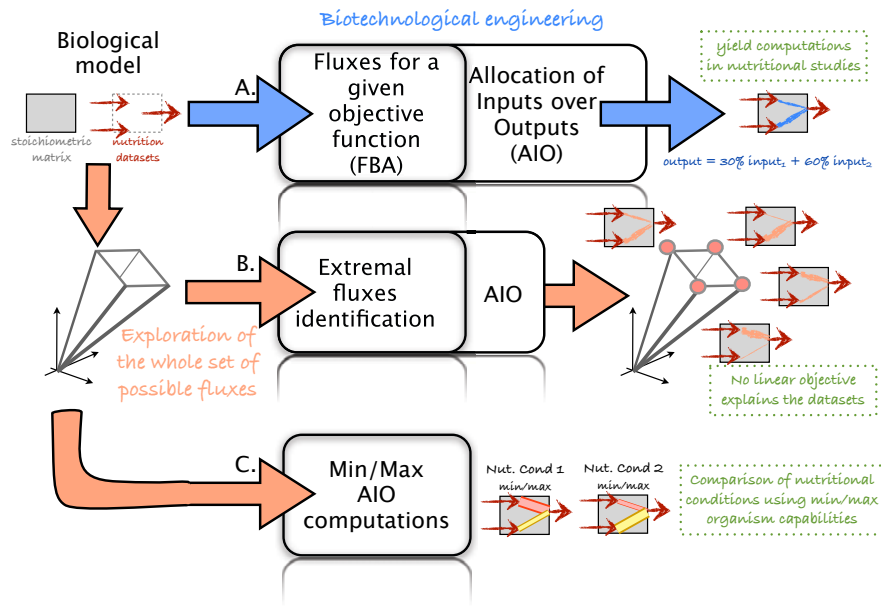


Figure 3 - Main functionalities of the analysis workflow

First, extremal vertices of the simplex polyhedron of plausible flux distributions have to be computed, including the case when this space is not bounded. Then, formal algebra is required to obtain a symbolic representation of AIO matrices, expressed as a formal function, where the variables are the coefficients of a plausible flux distribution. Finally, extrema of the AIO coefficients are computed among the complete simplex of plausible flux distributions, either with the existing optimization routine or with dedicated local-search algorithms.



Tables

Table 1 - Net uptake and milk component output of the mammary gland in three treatments

Data are renormalized in mmol/h/half udder. (a) Control diet (*Ctrl*) [32, 39] (b) Higher protein supply by casein infusion in the duodenum (*CN*) [32, 39] (c) Generic dataset (*HB*) [22]. This table is used to parameterize input and output vectors v_I and v_O together with additional biological linear constraints on some reaction fluxes.

Input or output flux		(Ctrl) [32, 39] mmol/h/half udder	(CN) [32, 39] mmol/h/half udder	(HB) [22] mol/d/udder mmol/h/half udder	
v ₂	glucose input ⁽¹⁾	237	232	12.21	254
v ₉₅	glycerol input	5.84	5.74	0.033	0.69
v ₉₆	acetate input	510	462	18.42	384
v ₉₇	BHBA input ⁽²⁾	84	167	7.25	151
v ₉₈	lactate input	0	0	0.023	0.48
v ₆₂	3C(n:m)-acyCoA+glycerol- 3P→triglyceride ⁽³⁾ output ⁽⁴⁾	32.96	39.11	1.52	31.67
Fatty acid output (synthesized) ⁽⁵⁾					
v ₁₀₀	C(4:0)	10.08	11.59	0.46	9.48
v ₁₀₁	C(6:0)	4.51	5.58	0.18	3.79
v ₁₀₂	C(8:0)	2.23	2.87	0.10	2.06
v ₁₀₃	C(10:0)	4.66	6.46	0.19	3.96
v ₁₀₄	C(12:0)	4.23	6.12	0.17	3.56
v ₁₀₅	C(14:0)	13.90	17.89	0.45	9.31
v ₁₀₆	C(16:0)	18.82	21.44	0.64	13.40
v ₉₉	Lactose output	73.80	83.52	3.81	79.28
Amino Acids balance ⁽⁶⁾ i.e. entry or output					
v ₁₂₈	Alanine input	3.11	0	0.105	2.19
v ₁₂₁	Alanine output	0	3.26	0	0
v ₁₁₉	Arginine input	4.40	4.48	0.526	10.96
v ₁₃₄	Asparagine output	0	0	0.023	0.48
v ₁₂₅	Aspartate output	3.43	4.13	0.247	5.15
v ₁₂₂	Glutamate output	0.54	6.33	0.230	4.79
v ₁₃₁	Glutamine input	1.22	1.79	0.072	1.50
v ₁₂₀	Glycine output	4.98	3.44	0.248	5.17
v ₁₂₄	Proline output	10.65	10.99	0.670	13.96
v ₁₃₆	Serine output ⁽⁷⁾	7.21	7.50	0.090	1.88
v ₁₁₈	Histidine input	0.23	0	0	0
v ₁₁₃	Isoleucine input	2.19	3.57	1.518	31.63
v ₁₁₄	Leucine input	2.02	3.76	0	0
v ₁₀₈	Lysine input	2.68	3.58	0.191	3.98
v ₁₁₁	Threonine input	0.35	0	0	0
v ₁₁₅	Valine input	2.54	3.86	0.438	9.13
v ₁₀₇	Peptide output ⁽⁸⁾	124.5	150.0	7.2	149.17
Additional constraints					
v ₈₂	NADPH through ICDH pathways ⁽⁹⁾	30%	30%	30%	30%
v ₅₈ = $\frac{0.7}{0.3} v_{82}$	NADPH through Pentose Phosphate ⁽⁹⁾	70%	70%	70%	70%
v ₅₆ = 3v ₆₂	C(n:m)→C(n:m)-acylCoA	98.87	117.32	4.56	95.00
FA primer from Acetate ⁽¹⁰⁾					
v ₅₃	C(4:0)	0	0	0	0
v ₅₄ = v ₉₀	C(6:0)	2.256	2.790	0.091	1.896
v ₅₅ = v ₉₁	C(8:0)	1.113	1.437	0.050	1.031
v ₅₁ = v ₅₂	C(10:0)	2.331	3.230	0.095	1.979
v ₈₆ = v ₉₂	C(12:0)	2.116	3.061	0.086	1.781
v ₈₇ = v ₉₃	C(14:0)	6.951	8.946	0.223	4.655
v ₈₈ = v ₉₄	C(16:0)	9.410	10.719	0.322	6.698
Other constraints ⁽¹¹⁾					
v ₂₄	Lactate →Pyruvate	0	0		
v ₄₄	Alanine catabolism	0	0		
v ₇₆	Alanine synthesis	0	0	0	0
v ₈₃	Asparagine synthesis	0	0		
v ₄₀	Histidine catabolism	0	0	0	0
v ₃₆	Leucine catabolism	0	0	0	0
v ₃₃	Threonine catabolism	0	0	0	0

¹ Input i.e. taken up by the stoichiometric system considered (i.e.net uptake in our example for the mammary gland)

² β-Hydroxybutyrate

³ Total triglycerides secreted in milk considering that milk fat was composed of 100% triglycerides and that all the triglycerides were secreted in milk fat [22]

⁴ Output i.e. leaving the system (secreted in milk)

⁵ All fatty acids synthesized within the mammary gland i.e. all C4 to C14 and 50% of C16 [22]

⁶ The balance between amino acid net uptake and amino acid net output in milk protein is calculated with established rules [22,47]. If the balance is positive it corresponds to an amino acid input that will be catabolized. If this balance is negative, the values corresponded to an amino acid output (that will be synthesized to meet its requirement for milk protein)

⁷ Serine output corresponded to Serine synthesized minus Serine utilized in other pathways i.e. Serine required in addition to Ser uptake to synthesize milk protein

⁸ Peptide output: number of peptide links required to synthesize the proteins exported out of the system (i.e. in milk protein)

⁹ [22]

¹⁰ Fatty acid primers were synthesized for 50% from acetate and for 50% from BHBA except C4 FA primer which was supposed to be synthesized only from BHBA [22]

¹¹ Set at zero because their inputs or outputs are set at zero (to avoid futile cycle)

Table 2 - Three different computations of ATP balance for the mammary gland in different treatments

Three natural assumptions are considered to model the mammary gland behavior: removal of all cycles, optimization of ATP production and study of the equilibria of a dedicated ODE-based model. All models exhibit considerable variability in their ATP balance (in mmol/h/half udder), which contradicts the assumption about the behavior of this organ. Moreover, quantitatively, the computed ATP balances are much higher than recent measurements. This suggests that ATP maximization cannot be considered as a natural objective function to model cow mammary behavior.

Biological model <i>Criteria of selection of a solution</i>	Dataset	ATP balance	proteic turnover
Mammary-gland model (Fig. 2) <i>Manual study based on cycle removal</i> [48]	(HB) (Ctrl) (CN)	6628 3081 2045	0 0 0
Mammary-gland model (Fig. 2) <i>ATP optimization</i>	(HB) (Ctrl) (CN)	6628 3081 2045	0 0 0
Model of Hannigan-Baldwin [22] <i>Study of numerical equilibria of an ODE model</i> [21]	(HB) (Ctrl) (CN)	4375=6500-2125 <i>non available</i> <i>non available</i>	0

Table 3 - Main properties of the simplex vertices under the assumption of constant ATP-production

The qualitative properties of all vertices are shared in (Ctrl) and (CN) treatments. Both correspond to a simplex with height vertices. So are six of each of the (Ctrl) and (CN) simplex vertices. H and G vertices, in the (Ctrl) and (CN) treatments, are plausible with respect to the literature.

	Dataset	Model name	Example of maximized function	Combinatorics of pathways					Validation		Pathways with nonrelevant flux values
				R_{19}	R_{15}	R_{14}	R_8	R_{64}	R_{63}	R_{13}	
				NADPH oxidation	OAA → G3P	OAA → PYR	G3P → G6P	peptide hydrolysis	peptide synthesis	Pyr → OAA	
Extreme flux distributions within the set of plausible solutions	(Ctrl)	B	$v_{15}-v_{19}$	0	1831	0	0	0	125	1835	R_{13}, R_{64}
	(CN)			795	0	0	0	150	803		
	(Ctrl)	F	$v_{14}-v_{19}$	0	0	1831	0	0	125	1835	R_{13}, R_{64}
	(CN)			795	0	0	0	150	803		
	(Ctrl)	D	v_8-v_{19}	0	0	0	3662	0	125	4	R_8, R_{64}
	(CN)			0	0	0	1590	0	150	8	
	(Ctrl)	H	$v_{64}-v_{19}$	0	0	0	0	305	430	4	
	(CN)			0	0	0	0	133	283	8	
	(Ctrl)	A	$v_{15}+v_{19}$	694	1714	0	0	0	125	1718	R_{13}, R_{64}
	(CN)			22	791	0	0	0	150	799	
	(Ctrl)	E	$v_{14}+v_{19}$	694	0	1714	0	0	125	1718	R_{13}, R_{64}
	(CN)			22	0	791	0	0	150	799	
(Ctrl)	C	v_8+v_{19}	694	0	0	3428	0	125	4	R_8, R_{64}	
(CN)			22	0	0	1583	0	150	8		
(Ctrl)	G	$v_{64}+v_{19}$	669	0	0	0	286	410	4		
(CN)			22	0	0	0	132	282	8		
Literature-based upperbounds for fluxes							≤ 591 mmol/h/half udder [43]	Non-zero [21, 32]	lower than whole body protein synthesis [39]	≤ 266 mmol/h/half udder [43]	

Table 4 - Origin of carbon mass within outputs for the two optimal flux distributions shown in Table 3

Both models have empty flux through the reactions OAA → PYR (R_{14}), OAA → G3P (R_{15}) and G3P → G6P (R_8). Model (G) shows strong NADPH oxidation whereas model (H) has zero NADPH oxidation.

Input	GLC	Glycerol	Acetate	BHBA	Lys	Threonine	Isoleucine	Leucine	Valine	Histidine	Arginine	Alanine	Glutamine	
Output	Origin of the carbon mass of each output within input (in percentage of total carbon mass of each output)													
model	(G)	(H)	(G)	(H)	(G)	(H)	(G)	(H)	(G)	(H)	(G)	(H)	(G)	(H)
Glycerol3P	87.4	95.3	12.6	4.7	0	0	0	0	0	0	0	0	0	0
lactose	100.0	0	0	0	0	0	0	0	0	0	0	0	0	0
C4	0	0	0	100.0	0	0	0	0	0	0	0	0	0	0
c6	0	0	66.7	33.3	0	0	0	0	0	0	0	0	0	0
c8	0	0	75.0	25.0	0	0	0	0	0	0	0	0	0	0
c12	0	0	80.0	20.0	0	0	0	0	0	0	0	0	0	0
c14	0	0	83.3	16.7	0	0	0	0	0	0	0	0	0	0
c16	0	0	85.7	14.3	0	0	0	0	0	0	0	0	0	0
glycine	73.0	78.3	9.2	3.7	9.7	9.8	4.0	4.1	0.3	1.4	0.3	0.2	0.3	0.8
glutamate	1.3	17.2	0.2	0.8	61.5	51.2	25.5	21.3	1.4	1.1	0.1	1.7	1.4	1.6
Proline	1.3	17.2	0.2	0.8	61.5	51.2	25.5	21.3	1.4	1.1	0.1	1.7	1.4	1.6
aspartate	1.8	17.6	0.2	0.9	60.1	50.3	25.0	20.9	1.3	1.1	0.1	2.2	1.9	1.5
peptide	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SERoutput	84.8	92.5	12.2	4.6	0.0	0.0	0.0	0.0	1.8	0.0	0.0	0.0	1.1	0.0
CO2Output	37.6	35.7	0.1	1.0	38.7	39.3	16.1	16.3	1.3	1.4	0.1	1.1	1.0	1.0

(a) Origin of carbon mass in outputs for (Ctrl) treatment.

Input	GLC	Glycerol	Acetate	BHBA	Lys	Isoleucine	Leucine	Valine	Arginine	Glutamine
Origin of the carbon mass of each output within input (in percentage of total carbon mass of each output)										
Output	(G)	(H)	(G)	(H)	(G)	(H)	(G)	(H)	(G)	(H)
model	(G)	(H)	(G)	(H)	(G)	(H)	(G)	(H)	(G)	(H)
Glycerol3P	90.3	90.7	9.7	9.3	0	0	0	0	0	0
lactose	100.0	0	0	0	0	0	0	0	0	0
C4	0	0	0	0	100.0	0	0	0	0	0
c6	0	0	0	0	66.7	33.3	0	0	0	0
c8	0	0	0	0	75.0	25.0	0	0	0	0
c10	0	0	0	0	80.0	20.0	0	0	0	0
c12	0	0	0	0	83.3	16.7	0	0	0	0
c14	0	0	0	0	85.7	14.3	0	0	0	0
c16	0	0	0	0	87.5	12.5	0	0	0	0
glycine	73.8	74.1	7.3	7.0	5.6	10.8	0.5	0.5	0.5	0.4
Alanine	90.3	90.7	9.7	9.3	0	0	0	0	0	0
glutamate	2.5	3.0	0.2	0.3	28.7	28.6	56.1	55.8	1.6	2.4
Proline	2.5	3.0	0.2	0.3	28.7	28.6	56.1	55.8	1.6	2.4
aspartate	3.8	4.3	0.4	0.4	27.8	27.6	54.2	53.9	1.6	1.5
peptide	0	0	0	0	0	0	0	0	0	0
SEROutput	90.3	90.7	9.7	9.3	0	0	0	0	0	0
CO2Output	24.2	24.1	0.2	0.2	22.2	43.4	1.9	1.8	2.0	1.7

(b) Origin of carbon mass in outputs for (CN) treatment.

Table 5 - Minimum and maximum utilization of input in each output

A local-search algorithm allowed us to compute the minima and maxima of each AIO coefficient for the two treatments (Ctrl), (CN) (in mmol/h/half udder of Carbon). These tables allow discriminating the response of the mammary gland to the two treatments without requiring selection of a flux distribution for reactions in the metabolic network. (CN) treatment (protein intake by food) is characterized by a lower proportion of glucose which is oxidized in CO₂ than in (Ctrl).

Input	Glucose	Glycerol	Acetate	BHBA	Lysine	Threonine ¹	Isoleucine ¹	Leucine ¹	Valine ¹	Histidine ¹	Arginine ¹	Alanine ¹	Glutamine ¹	
	1422	17.5	1020	336	16.1	1.40	13.1	12.1	12.7	1.38	26.4	9.33	6.10	
Output	Minimum and maximum utilisation of input in each output (in mmol/h/half udder of Carbon)													
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
Glycerol3P	98.9	32.9	97.0	1.24	12.5	0	38.4	0	16.0	0	1.1	0	0.1	0
Lactose	886	751	886	0	7.7	0	79.8	0	33.2	0	2.3	0	0.1	0
C4	40.3	0	0	0	0	40.3	0	0	0	0	0	0	0	0
c6	27.1	0	0	0	18.0	9.0	0	0	0	0	0	0	0	0
c8	17.8	0	0	0	13.4	4.5	0	0	0	0	0	0	0	0
c10	46.6	0	0	0	37.3	9.3	0	0	0	0	0	0	0	0
c12	50.8	0	0	0	42.3	8.5	0	0	0	0	0	0	0	0
c14	195	0	0	0	167	27.8	0	0	0	0	0	0	0	0
c16	301	0	0	0	263	37.6	0	0	0	0	0	0	0	0
Glycine ²	10.0	3.5	8.0	0.1	0.9	1.0	3.7	0.4	1.5	0.0	0.1	0.1	0.0	0.1
Glutamate ²	2.7	0.0	1.1	0.0	0.1	1.0	1.6	0.4	0.7	0.0	0.1	0.0	0.1	0.0
Proline ²	53.3	0.7	20.8	0.0	0.7	20.0	32.4	8.3	13.5	0.5	0.7	0.0	0.1	1.0
Aspartate ²	13.7	0.2	7.4	0.0	0.3	3.8	8.2	1.6	3.4	0.1	0.2	0.0	0.1	0.1
Serine ²	21.6	7.0	20.6	0.3	2.6	0.0	8.2	0.0	3.4	0.0	0.2	0.4	0.0	0.2
CO ₂	1126	396	565	1.3	14.5	343	446	142	185	12.1	15.3	0.7	0.8	9.1

(a). (Ctrl) treatment.

Input	Glucose	Glycerol	Acetate	BHBA	Lysine ¹	Isoleucine ¹	Leucine ¹	Valine ¹	Arginine ¹	Glutamine ¹
	1392	17.2	924	668	21.5	21.4	22.6	19.3	26.9	8.95
Output	Minimum and maximum utilisation of input in each output (in mmol/h/half udder of Carbon)									
	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max
Glycerol3P	117	36.1	115	1.8	11.3	0	22.8	0	44.5	0
Lactose	1002	866	1002	0	9.2	0	37.9	0	74.0	0
C4	46.4	0	0	0	0	46.4	0	0	0	0
C6	33.5	0	0	0	22.3	11.2	0	0	0	0
C8	23.0	0	0	0	17.2	5.7	0	0	0	0
C10	64.6	0	0	0	51.7	12.9	0	0	0	0
C12	73.4	0	0	0	61.2	12.2	0	0	0	0
C14	250	0	0	0	215	35.8	0	0	0	0
C16	343	0	0	0	300	42.9	0	0	0	0
glycine ²	6.9	2.1	5.5	0.0	0.5	0.4	1.3	0.7	2.6	0.0
Alanine ²	9.8	1.4	9.6	0.0	0.9	0	2.4	0	4.7	0
glutamate ²	31.7	0.8	8.8	0.0	0.4	6.7	9.0	13.0	17.6	0.4
proline ²	55.0	1.4	15.3	0.0	0.6	11.6	15.6	22.6	30.5	0.7
aspartate ²	16.5	0.6	7.0	0.0	0.3	2.7	4.5	5.3	8.9	0.2
serine ²	22.5	6.9	22.0	0.3	2.2	0	4.4	0	8.5	0
CO ₂	1021	244	402	2.0	12.5	180	227	351	444	16.3

(b). (CN) treatment.

⁽¹⁾ Amino acid input corresponded to positive balance between amino acid net uptake and amino acid and utilization in milk protein (i.e. peptide output).

⁽²⁾ Amino acid Output corresponded to negative balance between, amino acid net uptake and utilization in milk protein (i.e. peptide output)

Table 6 - Effect of long-chain fatty acids (LCFA) oxidation in the tricarboxylic acid (TCA) cycle over model analysis

To study the impact of the variability of FA oxidation, a ratio of long-chain FA (10%, 20%, 25%) was introduced in the TCA cycle. For datasets (CN), (Ctrl) and (HB) which were compatible with a given ratio of LCFA oxidation, extreme flux distributions and AIO coefficients variability were studied. Our main conclusions are robust to the introduction of LCFA oxidation (Table 6(a)). Interestingly, as shown in Table 6(b), the structure of the simplex generated by the (HB) diet is more complicated than the (Ctrl) and (CN) treatments, with 13 vertices for all hypotheses. This is due to the fact that R_{15} is highly constrained by measurements, so that this flux cannot vanish when R_{14} is optimized or when R_{19} is maximized. All the vertices for the (HB)-simplex contradict knowledge-based literature.

Model	Dataset	CO ₂ prediction		ATP optimum	ATP=1250 mmol/h/half udder			Variability of AIO coefficients	
		predicted CO ₂	ratio of predicted CO ₂ and measured CO ₂		Extreme flux distribution (see Table 6(b))			Glucose carbon required to produce lactose	
Ratio of long-chain FA oxidated in TCA					Total number	Nonplausible flux values	Nonplausible AIO	Glucose carbon required to produce lactose	Glucose carbon oxidized
0%	(HB)	1546	non available	6628	nonrelevant hypothesis [22, 24, 52]				
	(CN)	1021	99%	2045	8	6	2	[62.2 ; 72.0]	[17.5 ; 28.9]
	(Ctrl)	1126	121%	3081	8	6	2	[52.8 ; 62.3]	[27.8 ; 39.7]
10%	(HB)	1640	non available	7395	13	13	0	[47.7 ; 62.4]	[28.8 ; 45.9]
	(CN)	1107	107%	2739	8	6	2	[59.4 ; 72.0]	[17.5 ; 32.0]
	(Ctrl)	1202	129%	3701	8	6	2	[51.6 ; 62.3]	[27.9 ; 41.2]
20%	(HB)	1756	non available	8336	13	13	0	[46.9 ; 62.4]	[29.0 ; 47.0]
	(CN)	1214	118%	3607	8	6	2	[57.3 ; 72.0]	[17.5 ; 34.6]
	(Ctrl)	1298	140%	4476	nonrelevant hypothesis: predicted CO ₂ is not compatible with measured CO ₂ [32]				
25%	(HB)	1826	non available	8396	13	13	0	[46.5 ; 62.4]	[29.0 ; 47.5]
	(CN)	1279	124%	4128	8	6	2	[56.3 ; 72.0]	[17.6 ; 35.8]
	(Ctrl)	1355	146%	4938	nonrelevant hypothesis: predicted CO ₂ is not compatible with measured CO ₂ [32]				
				ATP balance is too high	Extreme distributions are not biologically relevant			For plausible ratios of long-chain FA in TCA, (CN) treatment is characterized by a lower proportion of glucose (on a carbon basis) which is oxidized in CO ₂ , and a larger ratio used for lactose synthesis.	

(a). Main characteristics of models with different ratios of LCFA oxidized in TCA.

Dataset	% of FA oxidized in TCA	Model name	Example of maximized function	Combinatorics of pathways					Validation			
				R_{19} NADPH oxidation	R_{15} OAA→G3P	R_{14} OAA→PYR	R_8 G3P→G6P	R_{64} peptide hydrolysis	R_{63} peptide synthesis	R_{13} Pyr → OAA		
(Ctrl)	0%	-			1831					125	1835	Non relevant flux values for R_{13}, R_{64}
	10%	-			2451					2455		
	0%	-			795					803		
(CN)	10%	B	v_{15-v19}	0	1489	0	0	0		150	1497	
	20%	-			2357					2365		
	25%	-			2878					2886		
(HB)	10%	-			6173					150	6145	
	20%	-			7115					7086		
	25%	-			7675					7646		
(Ctrl)	0%	-				1831				125	1835	Non relevant flux values for R_{13}, R_{64}
	10%	-				2451				2455		
	0%	-				795				803		
(CN)	10%	F	v_{14-v19}	0	0	1489	0	0		150	1497	
	20%	-				2357				2365		
	25%	-				2878				2886		
(HB)	10%	-				6173				150	6145	
	20%	-				7115				7086		
	25%	-				7675				7646		
(Ctrl)	0%	-						3662		125	4	Non relevant flux values for R_8, R_{64}
	10%	-						4902				
	0%	-						1590				
(CN)	10%	D	v_8-v19	0	0	0	0	2978		150	8	
	20%	-						4714				
	25%	-						5756				
	10%	-						12289				
(HB)	20%	D1	$v_8-v19-v14$	0	29	0		14172			0	
	25%	-						15292		150		
	10%	-						12289				
	20%	D2	$v_8-v19-v15$	0	0	29		14172				
	25%	-						15292				
(Ctrl)	0%	-						305		430	4	Glucose is the unique precursor of lactose synthesis (AIO)
	10%	-						409		533		
	0%	-						133		283		
(CN)	10%	H	v_{64-v19}	0	0	0	0	248		398	8	
	20%	-						393		543		
	25%	-						480		630		
	10%	-						1024		1174		
(HB)	20%	H1	$v_{64-v19-v14}$	0	29	0		1181		1331	0	
	25%	-						1274		1424		
	10%	-						1024		1174		
	20%	H2	$v_{64-v19-v15}$	0	0	29		1181		1331		
	25%	-						1274		1424		
(Ctrl)	0%	-			669	1714				125	1718	Non relevant flux values for R_{13}, R_{64}
	10%	-			694	2330				150	2334	
	0%	-				791					799	
(CN)	10%	A	v_{15+v19}	22	1485	0	0	0		150	1493	
	20%	-			2353					2361		
	25%	-			2874					2882		
(HB)	10%	-			5961					150	5932	
	20%	-			1216	6902				6873		
	25%	-				7462				7433		
(Ctrl)	0%	-			694	1714				125	1718	Non relevant flux values for R_{13}, R_{64}
	10%	-				2330					2334	
	0%	-				791					799	
(CN)	10%	E	v_{14+v19}	22	0	1485	0	0		150	1493	
	20%	-				2353				2361		
	25%	-				2874				2882		
	10%	-			1216	32	5929				5932	
(HB)	20%	E1	v_{14+v19}	1216	32	6902				150	6873	
	25%	-			1216	32	7462				6920	
	10%	-					6008	0			5979	
	20%	E2	$v_{14+v19-50v15}$	946	0	6949					6920	
	25%	-				7509					7481	
(Ctrl)	0%	-			694			3428		125	4	Non relevant flux values for R_8, R_{64}
	10%	-						4659				
	0%	-						1583				
(CN)	10%	C	v_8+v19	22	0	0	0	2971		150	8	
	15%	-						4706				
	20%	-						5749				
	10%	-						11858				
(HB)	20%	C1	v_8+v19	1216	32	0		13840			3	
	25%	-						14961				
	10%	-						11958		150		
	20%	C2	$v_8+v19-50v15$	946	0	29		13840			0	
	25%	-						14961				
(Ctrl)	0%	-			694			286		410	4	Glucose is the unique precursor of lactose synthesis (AIO)
	10%	-						388		513		
	0%	-						132		282		
(CN)	10%	G	v_{64+v19}	22	0	0	0	248		398	8	
	20%	-						392		542		
	25%	-						479		629		
	10%	-						989		1139		
(HB)	20%	G1	v_{64+v19}	1216	32	0		1145		1295	3	
	25%	-						1238		1388		
	10%	-						996		1146		
	20%	G2	$v_{64+v19-50v15}$	946	0	29		1146		1296	0	
	25%	-						1247		1397		
Litterature-based upperbounds for fluxes								≤ 591 mmol/h/half udder [43]	Non-zero [21, 32]	lower than whole body protein synthesis [39]	≤ 266 mmol/h/half udder [43]	

(b). Main properties of the simplex vertices, assuming constant ATP-production, with different ratios of LCFA oxidized in TCA.

Table 7 - Modeling the quantitative allocation of input nutrients in output products.

We are given a stoichiometric matrix A and a vector $c \in \mathbb{R}^{p+n+q}$ which describes the component composition of all metabolites. If v is a fixed flux distribution which is compatible with the stoichiometry of the system, $AIO[v]$ is a matrix whose (i, j) input describes the proportion of component quantity contained in m_i which is recovered in the flux of the output m .

$In^{r_j}(m_k) = \begin{cases} \frac{a_{k,j}C(m_k)}{\sum_{m_i \in \mathcal{M}, a_{i,j} < 0} a_{i,j} C(m_i)} & \text{if } a_{k,j} > 0 \\ 0 & \text{otherwise.} \end{cases}$	Ratio of the flux of component c provided as substrate to a reaction r_j recovered in the composition of the product m_k . It is the sum of individual substrate contributions.
$F(m_i) = \sum_{r_\eta \in \mathcal{R}, a_{i,\eta} > 0} a_{i,\eta} v_\eta = \sum_{r_\eta \in \mathcal{R}, a_{i,\eta} < 0} a_{i,\eta} v_\eta$	Total metabolite rate involved in the production of an intermediary metabolite m_i , before its degradation by other reactions
$d_{k,i} = \begin{cases} 1 & \text{if } k = i \\ - \sum_{r_j \in \mathcal{R}, a_{i,j} < 0} \frac{ a_{i,j} v_j}{F(m_i)} * In^{r_j}(m_k) & \text{otherwise.} \end{cases}$	Ratio of a product flux ($m_k \in \mathcal{P} \cup \mathcal{O}$) on the production of $m_i \in \mathcal{I} \cup \mathcal{P} \cup \mathcal{O}$
$D_1[v] = (d_{k,i})_{p < k, i \leq p+n+q}$	Linear transformation of matter components contained in intermediary or output metabolites
$D_2[v] = (d_{k,i})_{p < k \leq p+n+q, 1 \leq i \leq p}$	Linear transformation of matter component contained in input metabolites
$x[v, m] = \begin{pmatrix} x_I^{(m)} \\ x_P[x, m] \\ x_O[x, m] \end{pmatrix}$	rates of fluxes of component brought by the m -input flux appearing in the composition of each metabolites. $x_I^{(m)} = (0, \dots, 0, C(m)v_I(m), 0, \dots, 0)$.
$D_2[v]x_I^{(m)} = -D_1[v] \begin{pmatrix} x_P[x, m] \\ x_O[x, m] \end{pmatrix}$	Constraints on component fluxes deduced from the matter-invariance law, derived from Eq.(4) below.
$x[v, m](m_k) = \underbrace{\sum_{r_j \in \mathcal{R}, a_{k,j} > 0}}_{\text{reactions producing } m_k} \underbrace{In^{r_j}(m_k)}_{\text{proportion of component flux brought to the composition of } m_k \text{ through } r_j} \times \underbrace{\sum_{m_i \in \mathcal{M}, a_{i,j} < 0}}_{\text{substrate of } r_j} \underbrace{\frac{ a_{i,j} v_j}{F(m_i)}}_{\text{Proportion of } m_i \text{ - component consumed by } r_j} \times \underbrace{x[v, m](m_i)}_{\text{rate of the flux of component appearing in the composition of } m_i} \quad (4)$	
<hr/> total flux of component provided as substrate to r_j	

Additional Files

Additional Files

Additional file 1 - SBML version of the metabolic model

The complete metabolic model of mammary gland metabolism is provided in the free and open standard SBML representation format (Systems Biology Markup Language).

Bibliographie

- [AALVM⁺on] O. Abdou-Arbi, S. Lemosquet, J. Van-Milgen, A. Siegel, and J. Bourdon. Exploring metabolism flexibility in complex organisms through quantitative study of precursor sets for system outputs. *Bmc Systems Biology*, 2013. accepted for publication.
- [Ada09] O Adamou. *Study of some epidemiological models by the methods of computer algebra*. PhD thesis, University of Rennes-1 (France) and UAM (Niger), 2009.
- [Ari06] Rantanen Ari. *Algorithms for 13C metabolic flux analysis*. PhD thesis, University of Helsinki (Finland), 2006.
- [Arr51] K. J. Arrow. *A gradient Method for Approximating Saddle Points and Constrained Maxima*. RAND Corporation, Santer Monica, California, Paper, 1951.
- [BABM⁺11] Marie Beurton-Aimar, Bertrand Beauvoit, Antoine Monier, François Vallée, Martine Dieuaide-Noubhani, and Sophie Colombié. Comparison between elementary flux modes analysis and 13c-metabolic fluxes measured in bacterial and plant cells. *BMC systems biology*, 5 :95, January 2011.
- [Bal61] M L Balinski. An algorithm for finding all vertices of convex polyhedral sets. *SIAM Journal on Applied Mathematics*, 9 :72–81, 1961.
- [BAL74] R Bickerstaffe, E F Annison, and J L Linzell. The metabolism of glucose, acetate, lipids and amino acids in lactating dairy cows. *The Journal of Agricultural Science*, 82(01) :71–85, 1974.
- [BGLS01] J.F. Bonnans, J. C. Gilbert, C. Lemaréchal, and C. Sagastizabal. *Numerical Optimization-Theoretical and Practical Aspects*. Springer Verlag, Berlin, 2001.
- [BGN00] R. H. Byrd, J. C. Gilbert, and J. Nocedal. A trust region method based on interior point techniques for nonlinear programming. *Mathematical Programming*,, 89(1) :149–185, 2000.
- [BHN99] R. H. Byrd, M. E. Hribar, and J. Nocedal. An interior point algorithm for large-scale nonlinear programming. *SIAM Journal on Optimization*, 9(4) :877–900, 1999.
- [BIB04] S. Bergmann, J. Ihmels, and N. Barkai. Similarities and differences in genome-wide expression data of six organisms. *PLoS Biol*, 2(1) :E9, 2004.

- [Bla77] R. G. Bland. New finite pivoting rules for the simplex method. *Mathematics of Operations Research*, 2 :103–107, 1977.
- [BLQ02] Daniel A Beard, Shou-dan Liang, and Hong Qian. Energy balance for analysis of complex metabolic networks. *Biophysical journal*, 83(1) :79–86, July 2002.
- [BSEKO06] B J Bequette, N E Sunny, S W El-Kadi, and S L Owens. Application of stable isotopes and mass isotopomer distribution analysis to the study of intermediary metabolism of nutrients. *J Anim Sci*, 84 Suppl :E50–9, 2006.
- [BW00] L. T. Biegler and A. Wächter. Failure of a class of interior point methods for nonlinear programming. *Math. Program.*, 88 :565–587, 2000.
- [CBJS05] A. Cornish-Bowden, M. Jamin, and V. Saks. *Cinétique enzymatique*. EDP Sciences, 2005.
- [CKR⁺04] M. W. Covert, E. M. Knight, J. L. Reed, M. J. Herrgard, and B. O. Palsson. Integrating high-throughput and computational data elucidates bacterial networks. *Nature*, 429(6987) :92–96, 2004.
- [CP02] M. W. Covert and B. O. Palsson. Transcriptional regulation in constraints-based metabolic models of escherichia coli. *J. Biol. Chem*, 277(31) :28058–28064, 2002.
- [Dan51] G. B. Dantzig. Maximization of a linear function of variable subject to linear inequalities. In T. C. Koopmans, editor, *Activity Analysis of Production and Allocation*. John Wiley and Sons, New York, 339-347, 1951.
- [DOW55] G. B. Dantzig, A. Orden, and P. Wolfe. The generalized simplex method for minimizing a linear form under inequality restraints. *Pacific Journal of Mathematics*, 5(2) :183–195, 1955.
- [DP77] M E Dyer and L G Proll. An algorithms for determining all extreme points of a convex polytope. *Mathematical Programming*, 12 :81–96, 1977.
- [EBTTZ96] A. S. El-Bakry, R. A. Tapia, T. Tsuchiya, and Y. Zang. On the formulation and theory of the newton interior point method for nonlinear programming. *JOTA*, 89(3) :507–541, 1996.
- [EIP01] J S Edwards, R U Ibarra, and B O Palsson. In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. *Nature Biotechnology*, 19(2) :125–130, 2001.
- [EP00] J S Edwards and B O Palsson. The Escherichia coli MG1655 in silico metabolic genotype : its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences of the United States of America*, 97(10) :5528–33, May 2000.
- [FL97] R. Fletcher and S. Leyffer. *Nonlinear Programming without a penalty function*. Numerical Analysis Report NA/171, Departement of Mathematics, University of Dundee, Dundee, Scotland, 1997.

- [FLT02] R. Fletcher, S. Leyffer, and P. L. Toint. On the global convergence of a filter-sqp algorithm. *SIAM Journal on Optimization*, 13(4) :44–59, 2002.
- [FM68] A. V. Fiacco and G. P. McCormick. *Nonlinear programming : Sequential Unconstrained Minimization Techniques*. John Wiley and Sons, Inc., New York-London-Sydney, 1968.
- [Fri54] K. R. Frisch. *Principles of Linear Programming. With particular Reference to the Double Gradient Form of the Logarithmic Potential Method, Memorandum of October 18, 1954*. University of Oslo, 1954.
- [GBS⁺] A Ghozlane, F Bringaud, H Soueidan, I Dutour, F Jourdan, and P Thebault. Flux Analysis of the Trypanosoma brucei Glycolysis Based on a Multiobjective-Criteria Bioinformatic Approach. *Adv Bioinformatics*, 2012 :159423.
- [GDS⁺05] A. Ghazalpour, S. Doss, S. S. Sheth, L. A. Ingram-Drake, E. E. Schadt, A. J. Lusis, and T. A. Drake. Genomic analysis of metabolic pathway gene expression in mice. *Genome Biol*, 6(R59), 2005.
- [GK04] J Gagneur and S Klamt. Computation of elementary modes : a unifying framework and the new binary approach. *BMC Bioinformatics*, 5 :175, 2004.
- [GMS02] P. E. Gill, W. Murray, and M. A. Saunders. Snopt : An sqp algorithm for large-scale constrained optimization. *SIAM Journal on Optimization*, 12(4) :979–1006, 2002.
- [Gol89] D.E. Goldberg. *Genetic algorithms in search, optimization, and machine learning*. Addison-Wesley, 1989.
- [GPP⁺06] E. P. Gianchandani, J. A. Papin, N. D. Price, A. R. Joyce, and B. O. Palsson. Matrix formalism to describe functional States of transcriptional regulatory systems. *PLoS Comput Biol*, 2 :e101, 2006.
- [GT10] S. Gudmundsson and I. Thiele. Computationally efficient flux variability analysis. *BMC Bioinformatics*, 11 :489, 2010.
- [GW07] N. Goffard and G. Weiller. Pathexpress : a web-based tool to identify relevant pathways in gene expression data. *Nucleic Acids Res*, 35 :W176–81, 2007.
- [Han94] M D Hanigan. A mechanistic model of mammary gland metabolism in the lactating cow. *Agricultural Systems*, 45(4) :369–419, 1994.
- [HFM⁺09] M Hanigan, J France, S Mabweesh, W C McNabb, and B Bequette. High rates of mammary tissue protein turnover in lactating goats are energetically costly. *Journal of Nutrition*, 139 :1118–1127, 2009.
- [HHG⁺11] Andreas Hoppe, Sabrina Hoffmann, Andreas Gerasch, Christoph Gille, and Hermann-Georg Holzhütter. FASIMU : flexible software for flux-balance computation series in large metabolic networks. *BMC bioinformatics*, 12 :28, January 2011.

- [Kad01] A. Kadiri. *Analyse numérique des méthodes de points intérieurs pour les problèmes de complémentarité linéaire et la programmation quadratique convexe*. PhD thesis, INSA de Rouen (France), 2001.
- [KBB⁺08] L Klachiyani, E Boros, K Borys, K Elbassioni, and V Gurvich. Generating all vertices of a polyhedron is hard. Technical report, Rutgers University, New Jersey, United States, 2008.
- [KCVSZ⁺11] I. M. Keseler, J. Collado-Vides, A. Santos-Zavaleta, M. Peralta-Gil, S. Gama-Castro, L. Muñoz-Rascado, C. Bonavides-Martinez, S. Paley, M. Krummenacker, T. Altman, P. Kaipa, A. Spaulding, J. Pacheco, M. Latendresse, C. Fulcher, M. Sarker, A. G. Shearer, A. Mackie, I. Paulsen, R. P. Gunsalus, and P. D. Karp. Ecocyc : a comprehensive database of escherichia coli biology. *Nucleic Acids Res*, 39 (Database issue :D583–D590, 2011).
- [Ker89] A. Keraghel. *Etude adaptative et comparative des principales variantes dans l'algorithme de Karmarkar*. PhD thesis, Université Joseph Fourier Grenoble (France), 1989.
- [KPE03] K Kauffman, P Prakash, and J Edwards. Advances in flux balance analysis. *Current Opinion in Biotechnology*, 14 :491–496, 2003.
- [KS98] D. E. Kaufman and R. L. Smith. Direction choice for accelerated convergence in hit-and-run samplings. *Op. Res.*, 46 :84–95, 1998.
- [KSK⁺11] T. Y. Kim, S. B. Sohn, Y. B. Kim, W. J. Kim, and S. Y. Lee. Recent advances in reconstruction and applications of genome-scale metabolic models. *Curr Opin Biotechnol.*, 2011.
- [KSRL⁺06] S. Klamt, J. Saez-Rodriguez, J. A. Lindquist, L. Simeoni, and E. D. Gilles. A methodology for the structural and functional analysis of signaling and regulatory networks. *BMC Bioinformatics*, 7 :56, 2006.
- [KT51] H. W. Khun and A. W. Tucker. Nonlinear programming. In Berkeley University of California Press and Los Angeles, editors, *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 481–492, 1951.
- [LAAS⁺10] S Lemosquet, O Abdou-Arbi, A Siegel, J Guinard-Flament, J Van Milgen, and J Bourdon. A generic stoichiometric model to analyse the metabolic flexibility of the mammary gland in lactating dairy cows. In P Faverdin D. Sauvant J. Van Milgen and N Friggens, editors, *Modelling nutrient digestion and utilization in farm animals*. Wageningen Academic Publishers, 2010.
- [LB09] A Larhlimi and A Bockmayr. Minimal Metabolic Behaviors and the Reversible Metabolic Space. Technical report, Matheon, Berlin, Germany, 2009.
- [LC04] Z. Li and C. Chan. Integrating gene expression and metabolic profiles. *J Biol Chem*, 279 :27124–37, 2004.

- [Lio96] D Lionel. Algorithme de Bareiss, algorithme des sous-résultants. *Informatique théorique et applications*, pages 319–347, 1996.
- [Lob90] G E Lobley. Energy metabolism reactions in ruminant muscle : responses to age, nutrition and hormonal status. *Reprod Nutr Dev*, 30(1) :13–34, 1990.
- [LRL⁺09] S Lemosquet, G Raggio, G Lobley, H Rulquin, J Guinard-Flament, and H Lapierre. Whole-body glucose metabolism and mammary energetic nutrient metabolism in lactating dairy cows receiving digestive infusions of casein and propionic acid. *Journal of Dairy Science*, 92(12) :6068–6082, 2009.
- [Mat73] T H Mattheiss. An algorithm for determining irrelevant constraints and all vertices in systems of linear inequalities. *Operations Research*, 21 :247–260, 1973.
- [MS95] B. A. Murtagh and M. A. Saunders. *MINOS 5.4 User's Guide*. Systems Optimization Laboratory. Stanford University, 1983, Revised 1995.
- [MZ03] H. Ma and A. P. Zeng. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. *Bioinformatics*, 19(2) :270–277, 2003.
- [OLLP09] Young-Gyun Oh, Dong-Yup Lee, Sang Yup Lee, and Sunwon Park. Multiobjective flux balancing using the NISE method for metabolic network analysis. *Biotechnology progress*, 25(4) :999–1008, 2009.
- [OTP10] Jeffrey D Orth, Ines Thiele, and Bernhard ØPalsson. What is flux balance analysis? *Nature biotechnology*, 28(3) :245–8, March 2010.
- [PPSP03] N. D. Price, J. A. Papin, C. H. Schilling, and B. O. Palsson. Genome-scale microbial in silico models : the constraints-based approach. *Trends Biotechnol*, 21(4) :162–169, 2003.
- [PPW⁺03] J. A. Papin, N. D. Price, S. J. Wiback, D. A. Fell, and B. O. Palsson. Metabolic pathways in the post-genome era. *TRENDS in Biochemical Sciences*, 28(5) :250–258, 2003.
- [PSP⁺04] J A Papin, J Stelling, N D Price, S Klamt, S Schuster, and B O Palsson. Comparison of network-based pathway analysis methods. *Trends Biotechnol.*, 22 :400–405, 2004.
- [PT86] L. Patrick and Raymond T. *Analyse numérique matricielle appliquée à l'art de l'ingénieur, tome 1 : Méthodes directes*. Masson, Paris, 1986.
- [R2011] Optimization Toolbox User's Guide R2011b. *Matlab manual, The Math-work*. 2011.
- [RLL⁺06] G Raggio, S Lemosquet, G Lobley, H Rulquin, and H Lapierre. Effect of casein and propionate supply on mammary protein metabolism in lactating dairy cows. *Journal of Dairy Science*, 89 :4340–4351, 2006.

- [RME⁺10] Isabel Rocha, Paulo Maia, Pedro Evangelista, Paulo Vilaça, Simão Soares, José P Pinto, Jens Nielsen, Kiran R Patil, Eugénio C Ferreira, and Miguel Rocha. OptFlux : an open-source software platform for in silico metabolic engineering. *BMC systems biology*, 4 :45, January 2010.
- [RS06] S. A. Rahman and D. Schomburg. Observing local and global properties of metabolic pathways : 'load points' and 'choke points' in the metabolic networks. *Bioinformatics*, 22(14) :1767–1774, 2006.
- [RT91] C. K. Reynolds and H. F. Tyrrell. Effects of Mesenteric Vein L-alanine infusion on liver metabolism in beef heifers fed on diets differing in forage : concentrate ratio. *British Journal of Nutrition*, 66 :437–450, 1991.
- [SBC76] R A Scott, D E Beuman, and J H Clark. Cellular gluconeogenesis by lactating bovine mammary tissue. *J. Dairy Sci.*, 59(1) :50–56, January 1976.
- [SCF⁺02] C. H. Schilling, M. W. Covert, I. Famili, G. M. Church, J. S. Edwards, and B. O. Palsson. Genome-scale metabolic model of helicobacter pylori 26695. *J. Bacteriol*, 184(16) :4582–4593, 2002.
- [SDCK05] D. Segr, A. DeLuna, G. Church, and R. Kishony. Modular epistasis in yeast metabolism. *Nature Genetics*, 37(1) :77–83, 2005.
- [SDF99] S. Schuster, T. Dandekar, and D. A. Fell. Detection of elementary flux modes in biochemical networks : a promising tool for pathway analysis and metabolic engineering. *Trends Biotechnology*, 17(2) :53–60, 1999.
- [SFD00] S Schuster, D A Fell, and T Dandekar. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature Biotechnology*, 18 :326–332, 2000.
- [SLP00] C H Schilling, D Letscher, and B Palsson. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J. Theor. Biol.*, 203 :229–248, 2000.
- [SO12] W.A. Stein and Others. *Sage Mathematics Software (Version 5.5)*. The Sage Development Team, 2012.
- [SRD88] T. Sen, M. Raiszadeh, and P. Dileepan. A branch and bound approach to the bicriterion scheduling problem involving total flowtime and range of lateness. *Management Science*, 34(2) :254–260, 1988.
- [Ste04] J Stelling. Mathematical models in microbial systems biology. *Current Opinion in Microbiology*, 7 :513–518, 2004.
- [SVC02] D. Segrè, D. Vitkup, and G. M. Church. Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci U S A*, 99(23) :15112–15117, 2002.
- [SW91] B. Stewart and C. White. Multiobjective a*. *Journal of the ACM*, 38(4) :775–814, 1991.

- [SZZ⁺12] R. Schuetz, N. Zamboni, M. Zampieri, M. Heinemann, and U. Sauer. Multidimensional optimality of microbial metabolism. *Science*, 336(6081) :601–604, May 2012.
- [VM02] J Van-Milgen. Modeling biochemical aspects of energy metabolism in mammals. *The Journal of Nutrition*, 132(10) :3195–3202, 2002.
- [VMGR03] J Van-Milgen, F Gondret, and D Renaudeau. The use of nutritional models as a tool in basis research. In W B Souffrant and C C Metges, editors, *Progress in research on energy and protein metabolism*, pages 259–263, 2003.
- [VS99] R. J. Vanderbei and D. Shanno. An interior point algorithm for nonconvex nonlinear programming. *Computational Optimization and Applications*, 13 :231–252, 1999.
- [WB84] G C Waghorn and R L Baldwin. Model of metabolite flux within mammary gland of the lactating cow. *Journal of Dairy Science*, 67 :531–544, 1984.
- [Whi82] D. White. The set of efficient solutions for multiple-objectives shortest path problems. *Computers and Operations Research*, 9 :101–107, 1982.
- [WPM⁺06] H. Wei, S. Persson, T. Mehta, V. Srinivasasainagendra, L. Chen, G. P. Page, C. Somerville, and A. Loraine. Transcriptional coordination of the metabolic network in arabidopsis. *Plant Physiol*, 142 :762–74, 2006.
- [YBL⁺10] K. Yizhak, T. Benyamini, W. Liebermeister, E. Ruppin, and T. Shlomi. Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. *Bioinformatics*, 26(12) :i255–i260, 2010.
- [YEW⁺05] Y. Yang, L. Engin, E. S. Wurtele, C. Cruz-Neira, and J. A. Dickerson. Integration of metabolic networks and gene expression in virtual reality. *Bioinformatics*, 21 :3645–50, 2005.
- [YHT⁺04] I. Yeh, T. Hanekamp, S. Tsoka, P. D. Karp, and R. B. Altman. Computational analysis of plasmodium falciparum metabolism : organizing genomic information to facilitate drug discovery. *Genome Res*, 14(5) :917–924, 2004.

Résumé

Nous développons une approche générique pour comprendre comment différents régimes alimentaires peuvent influencer la qualité et la composition du lait. Cette question s'intègre dans le cadre du *Flux Balance Analysis* (FBA), qui consiste à analyser un réseau métabolique en optimisant un système de contraintes linéaires.

Nous avons proposé une extension du FBA pour analyser la transformation des nutriments en intégrant des hypothèses biologiques utilisées par différents modèles numériques dans un modèle générique de la glande mammaire. Notre méthode permet de quantifier les précurseurs qui interviennent dans la composition des sorties du système, en calculant des *contributions des entrées dans les sorties [AIO]*. A l'aide de cette approche, nous avons montré que la transformation des nutriments du lait ne peut pas être modélisée par l'optimisation d'une combinaison linéaire des flux des réactions sur un modèle du métabolisme mammaire.

Pour étudier plus précisément la flexibilité d'un réseau métabolique, nous avons proposé un algorithme efficace de recherche locale pour calculer les valeurs extrémales des coefficients des AIOs. Cette approche permet de discriminer les traitements sans formuler d'hypothèses sur le comportement interne du système.

Abstract

This thesis proposes a generic approach to understanding how different diets affect the quality and composition of milk. This question is addressed in the framework of Flux Balance Analysis (FBA), which considers metabolic network analysis as an optimization issue on a system of linear constraints.

In this work, we extended FBA to take into account nutrients transformation by incorporating general assumptions made by various numerical methods in a generic stoichiometric model of the mammary gland. Our method tries to quantify the precursor composition of each system output and to discuss the biological relevance of a set of flux in a given metabolic network. The composition is called contribution of inputs over outputs [AIO]. Using this method on the mammary metabolism, we could show that nutrients transformation cannot be properly modelled by optimizing a linear combination of reactions fluxes in the mammary gland model.

In order to further investigate metabolic network flexibility, we have proposed an efficient local search algorithm computing the extremal values of AIO coefficients. This approach enables to discriminate diets without making any assumption on the internal behaviour of the system.