



HAL
open science

Contributions des entrées sur les sorties pour les réseaux métaboliques sur génomes entiers: performances et utilisation pour des études en nutrition humaine

Oumarou Abdou Arbi, Anne Siegel, Jérémie Bourdon

► **To cite this version:**

Oumarou Abdou Arbi, Anne Siegel, Jérémie Bourdon. Contributions des entrées sur les sorties pour les réseaux métaboliques sur génomes entiers: performances et utilisation pour des études en nutrition humaine. CARI 2022 - Colloque Africain sur la Recherche en Informatique et en Mathématiques Appliquées, Oct 2022, Yaoundé, Cameroun. pp.1-12. <hal-03689107v2>

HAL Id: hal-03689107

<https://inria.hal.science/hal-03689107v2>

Submitted on 19 Jun 2022

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

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



HAL Authorization

Contributions des entrées sur les sorties pour les réseaux métaboliques sur génomes entiers : performances et utilisation pour des études en nutrition humaine

Oumarou ABDOU ARBI ^{*1}, Anne SIEGEL², Jérémie BOURDON³

¹Département de mathématiques, Université Dan Dicko Dankoulodo de Maradi, Niger

²Univ Rennes, Inria, CNRS, IRISA UMR 6074. Rennes, France

³Nantes Université, Ecole Centrale Nantes, CNRS, LS2N, UMR 6004, F-44000 Nantes, France

*E-mail : oabdouarbi@gmail.com

Résumé

Metabolic networks are composed by all the biochemical reactions that can take place within an organism, as well as the exchanges of metabolites between the organism and its environment. The metabolic network of many organisms has been reconstructed, either manually or automatically, including at the scale of the entire genome. These models are stored in biological reference databases. Diets can be applied to these networks to model the behavior of organisms, when subjected to a specific diet. In this work, we automate an approach that allows, for each input nutrient in the network, to determine the percentages that are distributed in the different outputs when the organism is forced to evolve in a given diet. We name this approach nAIO (normalized Allocation of inputs on outputs). The nAIO is computed thanks to the inversion of a large-scale matrix and is combined with linear optimization problems. We apply this calculation to all known bacterial networks from studies of the gut microbiota and stored in the Virtual Metabolic Human database. The calculation of nAIOs shows that computation times do not depend on the size of the network but rather on the selected diet. The nAIO calculation also shows that for some bacteria the nAIOs are independent of diet. For these bacteria the nAIOs can be used to make predictions that result in a linear relationship between the inputs of the system and its outputs.

Keywords : Metabolic Network, Flux Balance Analysis, flux distribution, Diets

I INTRODUCTION

Les nouvelles techniques d'observation dont le séquençage d'ADN ont induit une évolution des méthodes de recherche très importante au sein de la communauté des chercheurs en biologie. L'identification des gènes codant pour des enzymes, par ces nouvelles techniques, a permis de reconstruire *les réseaux métaboliques* de plusieurs organismes, c'est-à-dire l'ensemble des réactions biochimiques qui se déroulent au niveau de ces organismes. Les mathématiques et l'informatique

interviennent à travers des approches de modélisation pour comprendre les transformations intracellulaires.

Un exemple précurseur de réussite de modélisation à travers les réseaux métaboliques est connu sous le nom de *Flux Balance Analysis (FBA)*. Cette approche repose 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 [4, 5].

Ceci se traduit par l'optimisation d'une fonction objectif linéaire (combinaison linéaire de flux) sur un convexe pour faire des prédictions en accord avec des données expérimentales.

Cette stratégie a ainsi permis de modéliser efficacement le comportement de grands réseaux métaboliques tels que celui d'*Escherichia coli* [2] et *Helicobacter pylori* [3], pour lesquels les prédictions de maximisation de la biomasse sont apparues consistantes avec les mesures expérimentales [4].

Le succès du FBA a beaucoup attiré l'attention des bio-informaticiens et les biomathématiciens. 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.

Différents logiciels tels que OptFlux [8] ou Fasimu [10] tentent de regrouper toutes ces méthodes. Récemment, nous avons proposé une extension du FBA pour analyser le métabolisme dans les réseaux métaboliques entrées/sorties : contribution des entrées dans les sorties (*Allocation of Input in Output : AIO*) [11]. La modélisation par les AIO consiste à identifier les nutriments en entrées de la cellule et les nutriments en sorties de la cellule par dégradation successive des nutriments en entrées puis pour chaque entrée déterminer la quantité de flux de carbones qui se distribue dans les différentes sorties par inversion d'une matrice. Cette approche se définit comme une amélioration du FBA car elle permet de discuter de la pertinence biologique d'une solution optimale obtenue par le FBA. Le calcul des AIO a été réalisé par du code *php* à travers une interface web que nous avons développée (<http://nutritionanalyzer.genouest.org/>) qui prend en entrée un réseau métabolique pour donner en sorties les AIO d'une solution optimale en exécutant l'algorithme du simplexe.

L'objectif de ce travail est d'automatiser le calcul des AIO, par du code python, pour pouvoir l'appliquer à des modèles de réseaux métaboliques des bactéries du microbiote intestinal (flore intestinale) humain présentés au format standard sbml et issus de la plateforme *Virtual Metabolic Human* (<https://www.vmh.life/>) puis de faire une analyse de performance de ces calculs. Ceci permet de combiner les calculs faits avec des résultats obtenus via les bibliothèques classiques de manipulation des réseaux métaboliques comme *cobrapy* et de faire un vrai passage à l'échelle et un traitement de réseaux métaboliques sur génomes entiers.

II NORMALISATION DE LA CONTRIBUTION DES ENTRÉES DANS LES SORTIES (nAIO)

La contribution des entrées dans les sorties (contribution des entrées dans les sorties : AIO) est une notion très liée au domaine biologique de la production animale. Elle consiste à comparer l'efficacité ou le rendement de deux situations métaboliques différentes. Il peut s'agir d'une efficacité en terme d'énergie, de biomasse ou de transformation de protéines [6, 9].

Des exemples d'estimation d'efficacité sont la contribution des carbones du glucogène pour produire les carbones du Glucose dans le foie [1] ou la contribution des carbones du Glucose pour produire les carbones du Lactose dans la glande mammaire [7]. La formalisation des AIO consiste

à se donner un réseau métabolique [11]. C'est à dire un ensemble de métabolites et des réactions qui lient ces métabolites [4, 5]. L'ensemble M des métabolites est séparé en trois sous-ensembles : les métabolites d'entrées $I = \{i_1, i_2, \dots, i_p\}$ où $i_i = m_i$ avec $1 \leq i \leq p$; les métabolites intermédiaires (pivots) $P = \{\pi_1, \pi_2, \dots, \pi_n\}$ où $\pi_{i-p} = m_i$ avec $p+1 \leq i \leq p+n$; les métabolites de sorties $O = \{o_1, o_2, \dots, o_q\}$ où $o_{i-p-n} = m_i$ avec $p+n+1 \leq i \leq p+n+q$. Ainsi $M = I \cup P \cup O$. A chaque métabolite d'entrée, on associe une réaction de la forme : $\longrightarrow i_k$ (réaction d'entrée) et à chaque métabolite de sortie, on associe une réaction de la forme : $o_k \longrightarrow$ (réaction de sortie). On note R l'ensemble des réactions et n_R le nombre de réactions du réseau.

Les réseaux métaboliques stockés dans les bases de données biologiques sont des fichiers au format sbml. Dans ces fichiers, les métabolites d'entrées et de sorties sont représentés par des réactions d'échanges, c'est à dire les organismes peuvent importer des nutriments par ces réactions et produire des nutriments en sorties par ces réactions. Une réaction d'échange dont le flux de la réaction est négatif correspond à un métabolite d'entrée et une réaction d'échange dont le flux de la réaction est positif correspond à un métabolite de sortie. Cultiver un organisme dans un régime alimentaire, consiste à définir une fonction $Diet : I \longrightarrow \mathbb{R}^+$ qui à chaque métabolite d'entrée associe une valeur maximale de nutriment importée dans le régime. A l'état stationnaire, pour chaque métabo-

lite, tout ce qui est produit est également consommé. Un vecteur de flux d'état stationnaire vérifie alors les relations : (i) $Sv = 0$ avec (ii) $LB \leq v \leq UB$, où $S = (s_{i,j})_{1 \leq i \leq p+n+q, 1 \leq j \leq n_R}$ désigne la matrice stœchiométrique du réseau métabolique, LB et UB les bornes inférieure et supérieure du vecteur des flux des réactions. Contraindre un organisme à évoluer dans un régime revient à ajouter la contrainte additionnelle suivante : (iii) $\forall i \in I, |v_{input(i)}| \leq Diet(i)$. Soit $v_I \in \mathbb{R}^p$ le vecteur correspondant aux flux des réactions d'entrées, $v_O \in \mathbb{R}^q$ le vecteur correspondant aux flux des réactions de sorties, $v_P \in \mathbb{R}^n$ le vecteur correspondant aux flux des réactions intermédiaires.

On note $v = \begin{pmatrix} v_I \\ v_P \\ v_O \end{pmatrix}$ et $x_O[v, m] \in \mathbb{R}^q$ le vecteur dont chaque élément correspond à la proportion de flux d'un composé (carbone, hydrogène, oxygène etc.) d'un métabolite d'entrée m qui se retrouve dans chaque sortie. Pour déterminer ces quantités, on détermine le vecteur $x_P[v, m]$ qui correspond à la proportion de flux d'un composé d'un métabolite d'entrée m qui arrive dans chaque métabolite intermédiaire avant de se retrouver dans les sorties. La contribution des entrées

dans les sorties (AIO) est définie par la matrice $AIO[v]$ avec

$$AIO[v] = x_O[v, m]. \quad \begin{pmatrix} x_P[v, m] \\ x_O[v, m] \end{pmatrix} = -D_1[v]^{-1} D_2[v] \begin{pmatrix} C(m_1)v_I(m_1) & & 0 \\ & \ddots & \\ 0 & & C(m_p)v_I(m_p) \end{pmatrix}. \quad (1)$$

$C(m_i)$ désigne la quantité d'un composé (carbone, hydrogène, oxygène, ...) contenu dans le métabolite m_i et $v_I(m_i)$ le flux de la réaction du métabolite m_i avec $i \in \{1, 2, \dots, p\}$. La définition des coefficients des matrices $D_1[v]$ et $D_2[v]$ sont donnés dans la table ci-dessous pour un vecteur v fixé par exemple obtenu par optimisation d'une fonction objectif.

$D_1[v]$	$= (d_{k,i})_{p < k, i \leq p+n+q}$	Transformation linéaire du flux d'un composé contenu dans un métabolite intermédiaire en d'autres métabolites intermédiaires ou sorties
$D_2[v]$	$= (d_{k,i})_{p < k \leq p+n+q, 1 \leq i \leq p}$	Transformation linéaire du flux d'un composé contenu dans un métabolite d'entrée
$d_{k,i}$	$= \begin{cases} 1 & \text{if } k = i \\ - \sum_{r_j \in R, s_{i,j} < 0} \frac{ s_{i,j} v_j}{F(m_i)} * In^{r_j}(m_k) & \text{otherwise.} \end{cases}$	Proportion de flux d'un composé du métabolite $m_i \in I \cup P \cup O$ arrivant en ($m_k \in P \cup O$)
$In^r(m_k)$	$= \begin{cases} \frac{s_{k,j} C(m_k)}{\sum_{m_i \in M, s_{i,j} < 0} s_{i,j} C(m_i)} & \text{if } s_{k,j} > 0 \\ 0 & \text{otherwise.} \end{cases}$	Proportion d'un composé de chaque substrat de la réaction r_j arrivant à un produit m_k de la réaction. C'est une contribution individuelle des substrats au produit m_k .
$F(m_i)$	$= \sum_{r_\eta \in R, s_{i,\eta} > 0} s_{i,\eta} v_\eta = \sum_{r_\eta \in R, s_{i,\eta} < 0} s_{i,\eta} v_\eta$	Quantité totale de flux arrivant à un métabolite intermédiaire avant d'être dégradé par d'autres réactions

Soit la fonction $AIO(S, UB, LB, Diet, v, atom) : I \times O \rightarrow \mathbb{R}^+$ qui dépend du réseau métabolique (S, UB, LB), d'un régime alimentaire $Diet$ donné, d'un composé de répartition $atom$ donné et d'un flux v qui satisfait les relations (i), (ii) and (iii).

La fonction *contribution des entrées dans les sorties normalisée* (nAIO) est définie comme suit :

$$nAIO(S, UB, LB, Diet, atom, v) : (i, o) \in I \times O \rightarrow \frac{AIO(i, o)}{\sum_{o1 \in O} AIO(i, o1)} \in [0, 1].$$

III DONNÉES D'APPLICATION DES nAIO

Les données que nous utilisons pour le calcul des nAIO sont issues de la base de données Virtual Metabolic Human (<https://www.vmh.life/>). Sur la plateforme vmh, les réseaux métaboliques de 818 bactéries ont été reconstruits. Pour chaque bactérie, son réseau est stocké dans un fichier sbml. Il contient l'ensemble des métabolites et les différentes réactions liant ces différents métabolites. A chaque flux de réaction, des bornes inférieures et supérieures sont associées. Les métabolites d'entrées et de sorties sont représentés par des réactions d'échanges.

La plateforme contient 11 régimes alimentaires qui ont été établis en se basant sur des données de la littérature et des exemples réels. Les compositions moléculaires des régimes alimentaires peuvent être téléchargées en *millimoles/jour/personne*. Dans ce cas, les valeurs peuvent être utilisées comme bornes supérieures des flux des métabolites d'entrées des réseaux métabolites permettant ainsi de lier l'analyse des flux à un régime alimentaire comme nous l'avons fait dans le cas des AIO. Les 11 régimes alimentaires sont : "vegetarian diet" (un régime végétarien), "vegan diet" (un régime végétalien), "EU diet" (un régime alimentaire défini comme régime de l'UE), "Mediterranean diet" (un régime méditerranéen), "DACH diet" (c'est un régime basé sur la quantité recommandée de nutriments établie par la société de nutrition en Suisse, en Allemagne et en Autriche), "High protein diet" (régime riche en protéines), "gluten free diet" (régime sans gluten), "high fiber diet" (régime riche en fibres), "diet for type 2 diabetes" (régime alimentaire de patient diabétique de type 2), "high fat diet" (régime riche en graisses), "unhealthy diet" (un régime alimentaire qualifié de malsain). Pour chaque régime alimentaire, on dispose de la quantité en *mmol/jour/personne* pour 91 éléments nutritifs (cuivre , zinc, magnésium, eau, lactose , fructose etc.)

IV COMPARAISON DE FLUX ET DE MATRICES nAIO

Pour chaque souche de bactérie et chaque régime alimentaire définie à la section III, une valeur de flux de maximisation de la biomasse est calculée pour établir la matrice $nAIO(S, UB, LB, Diet, atom, v)$ telle que définie à la section II avec $atom="carbone"$. Ainsi, pour chaque bactérie, 11 valeurs de flux de maximisation de la biomasse et 11 matrices nAIO sont calculées. Pour les 818 bactéries disponibles sur la plateforme vmh $818*11=8998$ flux de maximisation de biomasse et 8998 matrice nAIO sont calculées.

Pour chaque souche, pour comparer deux flux de maximisation de biomasse issus de deux régimes alimentaires $Diet1$ et $Diet2$ nous calculons la distance euclidienne $distFlux(souche, Diet1, Diet2)$ entre ces deux flux définie par :

$$distFlux(souche, Diet1, Diet2) = \sum_{i=1}^{n_R} (v_{(souche, Diet1)}(i) - v_{(souche, Diet2)}(i))^2,$$

où n_R désigne le nombre de réactions du réseau métabolique de la souche. En outre, $v_{(souche, Diet1)}$ et $v_{(souche, Diet2)}$ sont des vecteurs qui maximisent la biomasse pour la bactérie $souche$ sous les contraintes (i), (ii) et (iii) de la section II.

De même, pour chaque souche, pour deux matrices nAIO de deux régimes alimentaires $Diet1$ et $Diet2$ nous calculons la distance euclidienne $distnAIO(souche, Diet1, Diet2)$ entre ces deux matrices définie par :

$$distnAIO(souche, Diet1, Diet2) = \sum_{o \in \bar{O}} \sum_{i \in \bar{I}} (nAIO_1(i, o) - nAIO_2(i, o))^2,$$

où $nAIO_1$ et $nAIO_2$ sont respectivement les fonctions de contribution obtenues respectivement pour les régimes $Diet1$ et $Diet2$. Autrement dit,

$$nAIO_1 = nAIO(S, UB, LB, Diet1, atom, v_{(souche, Diet1)}), et$$

$$nAIO_2 = nAIO(S, UB, LB, Diet2, atom, v_{(souche, Diet2)}).$$

En outre, dans la mesure où les métabolites d'entrée et de sortie dépendent de chacun des régimes, la distance ne peut être calculée que sur les entrées et sorties communes aux deux régimes. L'ensemble des entrées (resp. sorties) communes à $Diet1$ et $Diet2$ est noté \bar{I} (resp. \bar{O}).

Pour chaque souche et pour deux régimes quelconques des 11 régimes, les distances peuvent être calculées par $distFlux(souche, Diet1, Diet2)$ et $distnAIO(souche, Diet1, Diet2)$.

Les valeurs $distnAIO(souche, Diet1, Diet2)$ sont utilisées pour comparer pour une souche donnée, les distances de nAIO entre les 11 régimes pris deux à deux. Des valeurs $distFlux(souche, Diet1, Diet2)$ grandes avec des valeurs $distnAIO(souche, Diet1, Diet2)$ proches de zéro, signifient que pour la souche donnée est indépendante des régimes.

V RÉSULTATS DE PERFORMANCE DES CALCULS DES nAIO

Le calcul des nAIO consiste à calculer des inverses de matrices carrées de grandes tailles. Par exemple la bactérie *Achromobacter_xylooxidans_A8* est une bactérie dont le réseau métabolique est constitué de 1668 réactions et de 1382 métabolites. Pour cette bactérie, le calcul des nAIO pour la répartition des flux de carbones des entrées dans les sorties, a consisté à calculer l'inverse d'une matrice carrée de dimension 436 et la maximisation de la biomasse pour cette bactérie, a consisté à maximiser une fonction linéaire sur des contraintes linéaires dont la matrice est de dimension 1382×1668 .

Pour évaluer les performances des calculs des nAIO, pour chacun des 11 régimes alimentaires, nous avons donné les temps des calculs des nAIO sur un ordinateur de processeur $2.5GHz$ (1 seul coeur utilisé) et de 8 Go Ram. Pour chaque régime ces temps sont calculés en secondes en considérant l'ensemble des 818 bactéries de la plateforme vmh.

La Figure 1 fournit les temps de calcul des nAIO de l'ensemble des 818 bactéries pour chaque régime en fonction du nombre de métabolites pour chaque réseau métabolique. Les temps de calcul dépendent des régimes et non du nombre de métabolites. A titre d'exemple, la bactérie *Rudanella_lutea_DSM_19387* compte 1341 métabolites et son temps de calcul des nAIO 48.432 secondes pour le régime "High Fiber" alors que la bactérie *Faecalibacterium_pra-usnitzii_M21_2* compte 1361 mais son temps de calcul des nAIO est 0.232 secondes pour le régime "High Fiber".

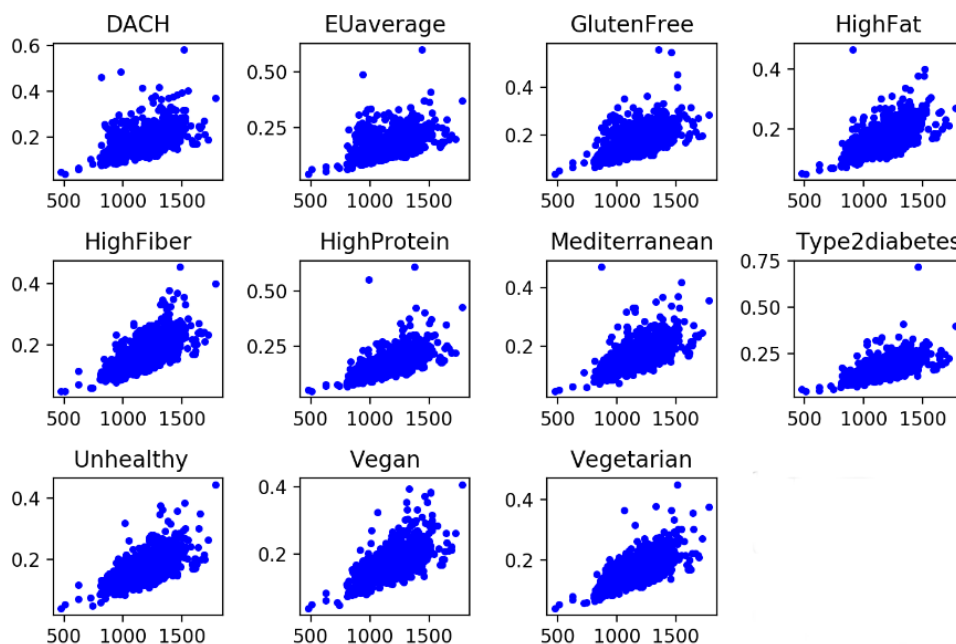


FIGURE 1 – **Temps nAIO (en secondes) des 818 réseaux métaboliques en fonction du nombre de métabolites.** En abscisse le nombre de métabolites pour chaque souche et en ordonnée le temps en secondes.

VI DISCUSSION ET CONCLUSION

Le calcul des nAIO consiste à déterminer pour chaque entrée, les pourcentages qui se distribuent dans les différentes sorties. Le calcul des AIO dépend également du régime alimentaire considéré. Pour une souche de bactérie, on peut déterminer, les valeurs des distances des nAIO et des flux pour deux régimes alimentaires considérés. Nous avons calculé les distances euclidiennes entre les nAIO et entre les flux pour la bactérie *Escherichia_albertii_KF1* pour les 11 régimes alimentaires pris deux à deux (voir Annexe 3). Parmi tous les résultats obtenus, on observe par exemple que la distance de nAIO entre le régime "High Fat" (ligne "AIO High Fat") et le régime "High Fiber" (colonne "High Fiber") est nulle, ce qui traduit une indépendance du nAIO sur le régime. Ceci n'est pas vrai au niveau des flux eux mêmes car il s'agit ici de deux flux différents qui donnent la même matrice nAIO pour la bactérie. Les différentes comparaisons des distances nAIO entre les régimes pris deux à deux donnent des valeurs presque nulles et des valeurs de distances de flux non nulles. Ainsi, pour ces souches présentant, la matrice nAIO peut servir pour prédire des valeurs d'entrées en fonction des valeurs sorties.

Le calcul de la contribution des entrées dans les sorties (AIO) dans les réseaux métaboliques entrées-sorties consiste à inverser des matrices pour quantifier la distribution des entrées dans les sorties. Dans ce travail, nous avons appliqué ce calcul à l'ensemble des réseaux métaboliques bactériens stocké dans la base de données *virtual metabolic human* lorsque les bactéries sont contraintes à évoluer dans des régimes alimentaires donnés. Il ressort que le temps de calcul ne dépend pas de la taille du réseau mais plutôt du régime alimentaire appliqué au réseau. Ce travail a également permis de se rendre compte que le calcul des nAIO peut servir pour prédire des entrées connaissant

des sorties. Ces calculs développés et appliqués en nutrition humaine n'ont été possible que grâce à une nouvelle implémentation d'un code php en code python qui est plus intégré aux autres outils d'analyse des réseaux métaboliques et plus performant. Les études de performance ont révélé des disparités d'un régime à l'autre, il va maintenant s'agir d'étudier précisément de manière plus systématique les linéarisations obtenues.

Références

- [1] C. K. REYNOLDS et H. F. TYRRELL. « Effects of Mesenteric Vein L-alanine infusion on liver metabolism in beef heifers fed on diets differing in forage : concentrate ratio ». In : *British Journal of Nutrition* 66 (1991), pages 437-450.
- [2] M. W. COVERT et B. O. PALSSON. « Transcriptional regulation in constraints-based metabolic models of escherichia coli ». In : *J. Biol. Chem* 277.31 (2002), pages 28058-28064.
- [3] C. H. SCHILLING, M. W. COVERT, I. FAMILI, G. M. CHURCH, J. S. EDWARDS et B. O. PALSSON. « Genome-scale metabolic model of helicobacter pylori 26695 ». In : *J. Bacteriol* 184.16 (2002), pages 4582-4593.
- [4] K. KAUFFMAN, P. PRAKASH et J. EDWARDS. « Advances in flux balance analysis ». In : *Current Opinion in Biotechnology* 14 (2003), pages 491-496.
- [5] N. D. PRICE, J. A. PAPIN, C. H. SCHILLING et B. O. PALSSON. « Genome-scale microbial in silico models : the constraints-based approach ». In : *Trends Biotechnol* 21.4 (2003), pages 162-169.
- [6] J. VAN-MILGEN, F. GONDRET et D. RENAUDEAU. « The use of nutritional models as a tool in basis research ». In : *Progress in research on energy and protein metabolism*. Sous la direction de W. B. SOUFFRANT et C. C. METGES. 2003, pages 259-263.
- [7] S. LEMOSQUET, G. RAGGIO, G. LOBLEY, H. RULQUIN, J. GUINARD-FLAMENT et H. LAPIERRE. « Whole-body glucose metabolism and mammary energetic nutrient metabolism in lactating dairy cows receiving digestive infusions of casein and propionic acid ». In : *Journal of Dairy Science* 92.12 (2009), pages 6068-6082.
- [8] I. ROCHA, P. MAIA, P. EVANGELISTA, P. VILAÇA, S. SOARES, J. P. PINTO, J. NIELSEN, K. R. PATIL, E. C. FERREIRA et M. ROCHA. « **OptFlux: an open-source software platform for in silico metabolic engineering.** » In : *BMC systems biology* 4 (jan. 2010), page 45. ISSN : 1752-0509.
- [9] M. BEURTON-AIMAR, B. BEAUVOIT, A. MONIER, F. VALLÉE, M. DIEUAIDE-NOUBHANI et S. COLOMBIÉ. « **Comparison between elementary flux modes analysis and ¹³C-metabolic fluxes measured in bacterial and plant cells.** » In : *BMC systems biology* 5 (jan. 2011), page 95. ISSN : 1752-0509.
- [10] A. HOPPE, S. HOFFMANN, A. GERASCH, C. GILLE et H.-G. HOLZHÜTTER. « **FASIMU: flexible software for flux-balance computation series in large metabolic networks.** » In : *BMC bioinformatics* 12 (jan. 2011), page 28. ISSN : 1471-2105.
- [11] O. ABDOU-ARBI, S. LEMOSQUET, J. VAN MILGEN, A. SIEGEL et J. BOURDON. « **Exploring metabolism flexibility in complex organisms through quantitative study of precursor sets for system outputs.** » In : *BMC Systems Biology* 8 (2014), page 8.

Annexes

A EXEMPLE DÉTAILLÉ DE CALCUL DE nAIO

Pour expliciter le calcul de nAIO, considérons le réseau métabolique de la Figure 2 composé de cinq métabolites. C'est un réseau utilisant deux métabolites d'entrées A et B pour produire deux métabolites de sorties D et E qui sont des produits finaux et qui sortent du système. Les métabolites A et B sont transformés en C avant de produire D et E comme suit :

$R1 : 2A + B \rightarrow C$; $R2 : B + C \rightarrow D + 2E$; $R3 : 4A \rightarrow D$; $R4 : \rightarrow A$; $R5 : \rightarrow B$; $R6 : D \rightarrow$; $R7 : E \rightarrow$. Comme l'indique la Figure 2. Le réseau est équilibré en carbones, c'est à dire pour chaque réaction, la somme des carbones des substrats de la réaction est égale à la somme des carbones des produits de la réaction.

La matrice stœchiométrique du réseau est :

$$S = \begin{pmatrix} & R_1 & R_2 & R_3 & R_4 & R_5 & R_6 & R_7 \\ A & -2 & 0 & -4 & 1 & 0 & 0 & 0 \\ B & -1 & -1 & 0 & 0 & 1 & 0 & 0 \\ C & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ D & 0 & 1 & 1 & 0 & 0 & -1 & 0 \\ E & 0 & 2 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}$$

Le vecteur de flux $v = (2, 2, 1.5, 10, 4, 3.5, 4)$ qui maximise la fonction objective $v_6 + v_7$ la contrainte d'état stationnaire : $Sv = 0$ mais aussi les contraintes aux bornes et les contraintes liées au régime telles qu'elles sont données au niveau de la Figure 2. La matrice des AIO est :

$$AIO = \begin{pmatrix} & A & B \\ D & 8 & 6 \\ E & 2 & 6 \end{pmatrix} \quad nAIO = \begin{pmatrix} & A & B \\ D & 0.8 & 0.5 \\ E & 0.2 & 0.5 \end{pmatrix}$$

Le flux de carbone qui rentre par $A = \text{flux de la réaction } R_4 * \text{nombre de carbone de } A = 10 * 1$. Ce flux de carbone se divise en deux à travers $R1$ et $R3$. $2 * v_1 / v_4 = 2/5$ passe dans $R1$ et $4 * v_3 / v_4 = 3/5$ passe dans $R3$. Le flux de carbone de A qui rentre dans C est :

$(2/5) * \text{carb}(C) / (2 * \text{carb}(A) + \text{carb}(B)) * \text{flux carbone de } A = (2/5) * (5/5) * 10 = 4$. Le flux de carbone de A qui rentre dans D est : $(3/5) * \text{carb}(D) / 4 * \text{carb}(A) * \text{flux carbone de } A = (3/5) * (4/4) * 10 = 6$.

Les 4 flux de carbones de A qui vont C se répartissent dans D et E à travers $R2$:

$\text{carb}(D) / (\text{carb}(B) + \text{carb}(C)) = 1/2$ dans D et $2 * \text{carb}(E) / (\text{carb}(B) + \text{carb}(C)) = 1/2$ dans E . Le flux de carbone de C venant de A qui rentre dans D est : $(2/5) * (1/2) * \text{flux de carbone de } A = (2/5) * (1/2) * 10 = 2$. Le flux de carbone de C venant de A qui rentre dans E est : $(2/5) * (1/2) * \text{flux de carbone de } A = (2/5) * (1/2) * 10 = 2$. Donc pour les 10 flux de carbone de A , $6 + 2 = 8$ se retrouvent dans D et 2 se

retrouvent dans E . Flux de carbone qui rentre par $B = v_5 * \text{carb}(B) = 4 * 3 = 12$. Ce flux de carbone se divise en deux à travers $R1$ et $R2$. $v_1 / v_5 = 1/2$ passe dans $R1$ et $v_2 / v_5 = 1/2$ passe dans $R2$. Le

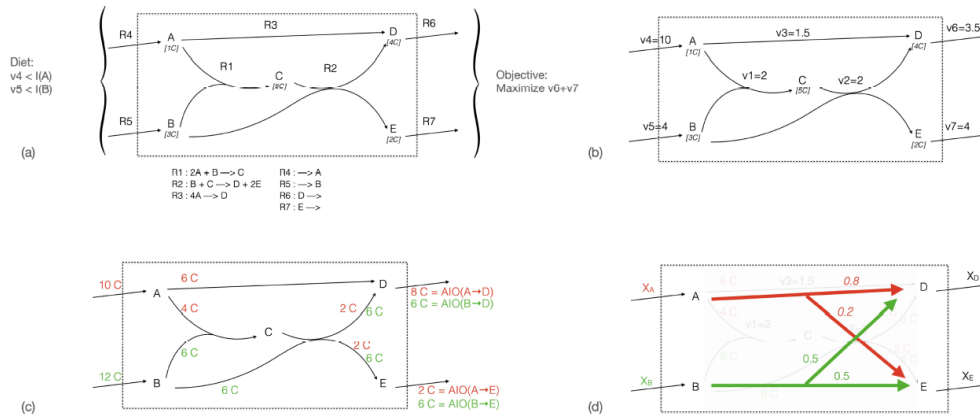


FIGURE 2 – Exemple de réseau métabolique pour le calcul de nAIO (a) Représentation du réseau avec l'ensemble des réactions, l'ensemble des métabolites et le nombre de carbones dans chaque métabolite (ici $\text{carbone}(A) = 1$, $\text{carbone}(B) = 3$, $\text{carbone}(C) = 5$, $\text{carbone}(D) = 4$, $\text{carbone}(E) = 2$). La fonction objective à maximiser est : ($obj = v_6 + v_7$). Les bornes supérieures produits par le régime alimentaire sur les entrées sont : $I(A) = 10 \text{mmol.l}^{-1}.\text{h}^{-1}$ et $I(B) = 4 \text{mmol.l}^{-1}.\text{h}^{-1}$. (b) Les différentes valeurs de flux après maximisation de la fonction objective avec comme contraintes sur les flux ($LB = 0$ and $UB \leq 10 \text{mmol.l}^{-1}.\text{h}^{-1}$). (c) Calcul des valeurs des flux de carbone A arrivant en D et E (en rouge) et des flux de carbone B arrivant en D et E (en vert). Ce calcul tient compte des valeurs des flux des réactions et des coefficients. Le résultat du calcul des flux de carbones des entrées dans les sorties est appelé AIO (d) Calcul normalisé des AIO appelé *normalized AIO*.

flux de carbone de B qui rentre dans C à travers $R1$ est : $(1/2) * \text{carb}(C) / (2 * \text{carb}(A) + \text{carb}(B)) * \text{flux}$
 carbone de B = $(1/2) * (5/5) * 12 = 6$. Le flux de carbone de B qui passe à travers $R2$ est : $(1/2) * \text{flux}$
 carbone de B = 6. Ces 6 flux de carbones se répartissent $1/2$ dans D et $1/2$ dans E à travers $R2$ c'est à dire 3 dans D et 3 dans E.

Le flux de carbone de B qui rentre dans C à travers $R1$ se répartit également à $1/2$ dans D et $1/2$ dans E c'est à dire 3 dans D et 3 dans E. Donc pour les 12 flux de carbone de B, $3+3=6$ se retrouvent dans D et $3+3 = 6$ se retrouvent dans E. En résumé, le flux de carbone dans D venant de A est 8 (6 à travers $R3$ et 2 à travers $R1 - R2$) et le flux de carbone dans D venant de B est 6. Ainsi, $AIO(A \rightarrow D) = 8$ et $AIO(B \rightarrow D) = 6$. De la même manière, le flux de carbone dans E venant de A est égal 2 (à travers $R1 - R2$) et le flux de carbone dans E venant de B est égal 6. Ainsi, $(AIO(A \rightarrow E) = 2$ et $AIO(B \rightarrow E) = 6$. Au total sur les 10 flux de carbones de A : 80% sont transformés en D et 20% sont transformés en E. Sur les 12 flux de carbones de B : 50% sont transformés en D et 50% sont transformés en E. Ce qui correspond à la matrice des AIO normalisée et qui donne la relation linéaire entre les entrées et sorties du système comme suit :

$$X_D = 0.8X_A + 0.5X_B$$

$$X_E = 0.2X_A + 0.5X_B$$

où $X_A = carb(A) * v_4$ correspond au flux de carbones venant de A et $X_D = carb(D) * v_6$ correspond au flux de carbones sortant en D. De même $X_B = carb(B) * v_5$ correspond au flux de carbones venant de B et $X_E = carb(E) * v_7$ correspond au flux de carbones sortant en E.

B TEMPS DE CALCUL nAIO EN FONCTION DU NOMBRE DE RÉACTIONS

En complément de la Figure 1 qui fournit le temps de calcul des nAIO des 818 réseaux métaboliques en fonction du nombre de métabolites, nous donnons ici (Figure 3) l'ensemble des temps de calcul des nAIO des 818 réseaux métaboliques en fonction du nombre de réactions du réseau. Les résultats obtenus sont tout à fait similaires à ceux obtenus pour le nombre de métabolites.

A titre d'exemple, la bactérie *Faecalibacterium_prausnitzii_M21_2* compte 1767 réactions et son temps de calcul des nAIO est 0.254 secondes pour le régime "EU average" alors que la bactérie *Rudanella_lutea_DSM_19387* compte 1341 mais son temps de calcul des nAIO est 49.042 secondes pour le régime "EU average".

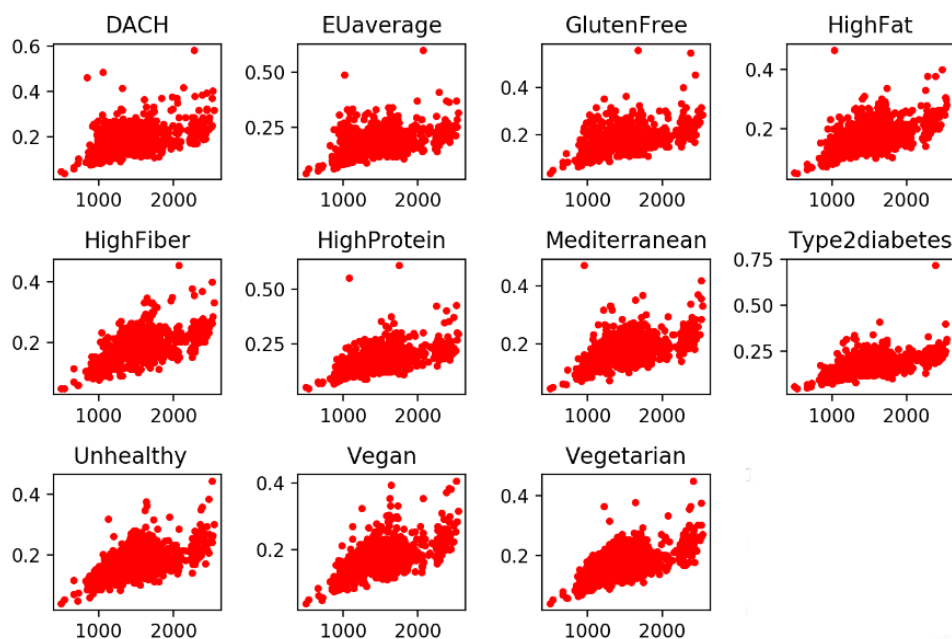


FIGURE 3 – Temps nAIO (en secondes) des 818 réseaux métaboliques en fonction du nombre de réactions pour chaque réseau métabolique. En abscisse le nombre de métabolites pour chaque souche et en ordonnée le temps en secondes

C TABLE DE COMPARAISON DISTANCES FLUX ET nAIO

La table suivante synthétise les calculs de distance des distributions de flux et des nAIO pour la souche *Escherichia_albertii_KF1* et les 11 régimes de l'étude.

AIO/FBA	DACH	EU average	Gluten Free	High Fat	High Fiber	High Protein	Mediterranean	Type2 diabetes	Unhealthy	Vegan	Vegetarian
AIO DACH	0	0	0	0	0	0	0	0	0	0	0.007
FBA DACH	0	2.575	15.91	63.31	18.372	11.752	24.849	11.654	11.823	10.21	17.204
AIO EUaverage	0	0	0	0	0	0	0	0	0	0	0.007
FBA EUaverage	2.575	0	14.883	59.258	21.604	11.593	24.965	11.323	11.212	13.63	18.678
AIO GlutenFree	0	0	0	0	0	0	0	0	0	0	0.007
FBA GlutenFree	15.91	14.883	0	81.687	12.804	9.809	17.247	9.315	13.382	15.99	20.893
AIO HighFat	0	0	0	0	0	0	0	0	0	0	0.007
FBA HighFat	63.31	59.258	81.687	0	111.154	76.125	117.743	70.295	57.417	98.161	82.727
AIO HighFiber	0	0	0	0	0	0	0	0	0	0	0.008
FBA HighFiber	18.372	21.604	12.804	111.154	0	11.83	7.631	13.408	29.701	9.73	24.582
AIO HighProtein	0	0	0	0	0	0	0	0	0	0	0.007
FBA HighProtein	11.752	11.593	9.809	76.125	11.83	0	9.88	1.48	13.13	8.858	21.753
AIO Mediterranean	0	0	0	0	0	0	0	0	0	0	0.008
FBA Mediterranean	24.849	24.965	17.247	117.743	7.631	9.88	0	11.141	31.256	11.239	32.282
AIO Type2diabetes	0	0	0	0	0	0	0	0	0	0	0.007
FBA Type2diabetes	11.654	11.323	9.315	70.295	13.408	1.48	11.141	0	12.191	10.299	22.106
AIO Unhealthy	0	0	0	0	0	0	0	0	0	0	0.007
FBA Unhealthy	11.823	11.212	13.382	57.417	29.701	13.13	31.256	12.191	0	17.307	24.138
AIO Vegan	0	0	0	0	0	0	0	0	0	0	0.007
FBA Vegan	10.21	13.63	15.99	98.161	9.73	8.858	11.239	10.299	17.307	0	19.796
AIO Vegetarian	0.007	0.007	0.007	0.007	0.008	0.007	0.008	0.007	0.007	0.007	0
FBA Vegetarian	17.204	18.678	20.893	82.727	24.582	21.753	32.282	22.106	24.138	19.796	0