

Bibliographie RBA pour les plantes

S. Fischer - Biosys - MAIAGE

21 octobre 2016

Table des matières

1	Les concepts de bases	3
1.1	FBA	3
1.2	RBA	6
1.3	Description du métabolisme	7

1 Les concepts de bases

1.1 FBA

Formulation Je reprends la façon dont c'est cité dans l'intro de Goelzer et al. (2011). Le formalisme des FBA est très simple : on se donne un certain nombre de métabolites et de réactions, une matrice de stoechiométrie S qui représente les réactions et des flux ν pour chaque réaction. Le problème d'optimisation est juste donné par

$$\begin{cases} \text{Maximiser } c^T \nu \\ \text{sous les contraintes } S\nu = b, \alpha_i \leq \nu_i \leq \beta_i \end{cases}$$

b est nul s'il n'y a pas de flux entrant.

Remarque 1.1. Dans le domaine, on identifie volontiers une réaction par l'enzyme qui la catalyse. Pour l'instant je ne sais pas trop ce qui se passe si une enzyme catalyse deux réactions différentes. En tout cas il ne faut pas s'étonner si les dimensions de S sont données en fonction du nombre d'enzymes et pas du nombre de réactions.

Remarque 1.2. Implicitement, il semble que le système soit réécrit en tenant compte des sens de réactions et en forçant $\nu \geq 0$ (Savinell and Palsson, 1992a)!!! Ça rajoute pas mal de contraintes.

Remarque 1.3. La matrice de stoechiométrie peut être simplifiée en agrégeant des réactions (Varma and Palsson, 1993a).

Remarque 1.4. La formulation duale du problème est utilisée pour juger l'importance de chaque métabolite à réaliser un objectif précis. Ça a été formalisé avec les idées de *shadow price* et *reduced costs* (Savinell and Palsson, 1992a) et *scaled shadow price* (Varma and Palsson, 1993b).

Problèmes (1) on est obligé de minorer/majorer certaines valeurs de ν pour borner l'espace des solutions, (2) le vecteur c est fixé *a priori*. Il donne les flux nécessaires à soutenir un certain taux de croissance (notamment en contrôlant les flux des métabolites finaux), mais la composition dépend en réalité du taux de croissance.

Interprétation mathématique La première contrainte est un simple système linéaire. Si $b = 0$, ν vit dans le noyau de S (typiquement un espace vectoriel non borné de type \mathbb{R}^n). Si b est non nul, ν vit dans un espace affine $\nu_0 + \text{Ker} S$ (sensiblement la même chose). Une base naturelle du noyau est donné par les cycles minimaux du système. Imposer b ne diminue pas la dimension de l'espace des solutions...

$c^T \nu$ est un produit scalaire. Sur une sphère unité du noyau, il est maximal dans la direction de la projection de c sur l'espace des solutions. Mais vu que cet espace n'est pas borné, il est évident qu'il n'y a pas de solution (sauf si c est orthogonal à l'espace des solutions...). On peut même ajouter que dès qu'une des dimensions de l'espace (non orthogonale à c) n'est pas bornée, il n'y a pas de solution. Il faut donc autant de contraintes (indépendantes) sur ν que de dimensions au noyau de S .

En imposant des bornes sur chaque composante de ν on découpe le s.e.v de façon à le borner en un espace convexe. Le produit scalaire est forcément maximal sur un bord de l'espace découpé. Vu que le bord est linéaire par morceau et que $c^T \nu$ est proportionnel à la projection de ν sur c , on peut se convaincre que c'est maximal sur un sommet (sauf si une arête/face est orthogonale à c , blablabla - rien de nouveau pour de l'optimisation linéaire en espace convexe, c'est juste pour donner une image). Vu que le sommet est donné par le découpage des α et β , il y a un risque que ce découpage détermine la solution finale. Normalement, si on est obligés de définir des β_i , c'est qu'ils devraient intervenir dans la solution. Par contre difficile d'évaluer en grande dimension à quel point l'orientation de c peut faire basculer la solution d'un sommet à l'autre ou à quel point ce basculement peut avoir des effets catastrophiques ?

Certains auteurs n'imposent pas de bornes sur ν (Varma and Palsson, 1993a,b) et se donnent un certain b . Ici intervient la contrainte cachée $\nu \geq 0$ et la décomposition des réactions bidirectionnelles en réactions unidirectionnelles. Cela supprime de l'espace des solutions tous les cycles infaisables d'un point de vue thermodynamique et limite la dégénérescence des solutions aux cycles faisables. La solution est unique ssi le noyau de A ne s'intersecte pas avec le sous-espace $\nu \geq 0$.

Remarque 1.5. La dimension la plus grand de S est visiblement le nombre de réactions, pas le nombre de métabolites. Il faudrait réfléchir au rang (plein ?) de S (et du coup à la dimension du noyau).

Variantes ? Le problème est clairement formulé comme Goelzer et al. (2011) dans Fell and Small (1986) (parfois considéré comme l'article fondateur). Pourtant dans Varma and Palsson (1993b, 1994), la formulation de la fonction objectif est différente, l'objectif est exprimé en fonction des métabolites. Dans tous les cas, il semble qu'on identifie un métabolisme à son flux de sortie, mais comment on fait pour construire c quand plusieurs réactions produisent un élément (comme l'ATP ou le NADH) ? Le seul article qui explique clairement sa procédure est Savinell and Palsson (1992a), où les coûts sont répartis uniformément. Ceci étant dit, dans cet article on ne cherche pas à maximiser la croissance de biomasse, mais à minimiser le coût de la croissance !

Le choix de la contrainte qui permet de fixer la solution n'est pas toujours la même. Dans Savinell and Palsson (1992a); Varma and Palsson (1993a,b) A semble bien fichue donc $\nu \geq 0$ impose assez de contraintes. Par contre les choix de b et c varient. Dans Savinell and Palsson (1992a), b inclut les flux de précurseurs biosynthétiques et c comprend les coûts de production/importation des nutriments de base. Pareil pour Savinell and Palsson (1992b), sauf qu'on retire des réactions pour que le problème LP ait une solution ! Dans Varma and Palsson (1993b), on ne sait pas trop trop ce qu'il y a dans b et c'est c qui comprend les flux de précurseurs biosynthétiques nécessaires à la croissance. Dans tous ces papiers, l'existence d'une solution est empirique, il n'y a pas de justification théorique *a priori*.

Applications Savinell and Palsson (1992a) pose le problème mathématique et l'applique à la croissance de cellules tumorales (souris). Comme les exigences biosynthétiques

et énergiques sont imposées dans b , les auteurs imposent des optimisations *ad hoc* sur c , comme la minimisation de production d'ATP ou d'import d'éléments. Ils comparent l'efficacité des différents a.a. et du glucose comme source de croissance. L'article rentre ensuite dans des détails du fonctionnement du métabolisme que je n'ai pas lus. L'analyse du fonctionnement des cellules tumorales continue dans Savinell and Palsson (1992b), notamment en fonction des conditions environnementales (densité d'oxygène, sources de carbone et d'azote) et des éléments rejetés (notamment alanine et glutamine). Lire l'article pour plus de détails, difficile de faire la synthèse vu mes connaissances sur le métabolisme.

Varma and Palsson (1993a) commence par s'intéresser au catabolisme d'*E. coli*. Le problème est simplifié dans la mesure où le glucose est envisagé comme seule source de carbone et on ne cherche à synthétiser qu'une cible. La fonction objectif est donc réduite à une forme extrêmement simple. Pour avoir une solution, le flux entrant en glucose est normalisé à un. Les cibles envisagées sont celles qui servent de base à l'anabolisme : les 3 cofacteurs (ATP pour l'énergie, NADH, NADPH pour la puissance réductrice) et les 12 précurseurs biosynthétiques. Les auteurs analysent un certain nombre d'aspects, notamment le nombre d'atomes de carbones qui sont effectivement transférés du glucose vers les précurseurs, et le coût en ATP pour réaliser le pathway optimal proposé (appelé *shadow price*).

Dans la suite (Varma and Palsson, 1993b), les auteurs cherchent à intégrer ces résultats en définissant c à partir des coûts connus en chaque cofacteur et précurseur. Le formalisme utilisé semble peu rigoureux mais c'est l'idée. On reste sur une contrainte de flux d'entrée pour obtenir une solution : 1g de glucose. Ils développent une nouvelle métrique le *scaled shadow price* qui permet de montrer la pression qu'il y a sur chaque cofacteur ou précurseur pour la création de biomasse. Si la valeur est proche de 1, ça veut dire qu'on est au niveau de production maximal pour ce précurseur/cofacteur. Dans un premier temps, la solution obtenue n'est pas adéquate. Le taux de croissance est trop élevé et les pathways utilisés ne correspondent pas à la réalité.

Les auteurs étudient ensuite la sensibilité du système. Forcer certains pathways n'a pas forcément un effet dramatique. Changer marginalement les valeurs de c non plus. Changer le rapport P/O qui détermine la stoechiométrie des transporteurs d'électrons non plus. Reste alors les coûts de maintenance en énergie pour tous les processus hors croissance (motilité, etc.) qui sont traduits arbitrairement en coût ATP. La première étude a montré qu'un glucose pouvait générer au maximum 18.7 ATP, qui est donc le coût maximal. On tombe sur des valeurs proches des expériences pour un coût de 4 ATP, qui sélectionne automatiquement les pathways connus expérimentalement. Pour finir, les auteurs analysent la stabilité des solutions en fonction des coûts de maintenance. Il y a des sauts assez brutaux dans les différentes métriques. Cela correspond à des changements dans les pathways utilisés et la capacité d'échanger des cofacteurs redox contre de l'énergie. Ils citent notamment un cas où l'augmentation de l'utilisation du cycle de Krebs permet de libérer à la fois plus d'ATP et de NADH, ce qui relâche la pression sur les besoins en NADH dont le surplus est converti en ATP également.

Varma and Palsson (1994) fait de l'application pure à différentes conditions de croissance. Le formalisme mathématique est très bizarre parce qu'il semble que la population

de cellule soit assimilée à une grosse bactérie (on considère la biomasse totale). Pour une fois, une partie de l'article est consacré à la manière de contraindre le système. Les auteurs estiment expérimentalement le flux max de glucose, d'oxygène, les coûts de maintenance en ATP et la participation des précurseurs à la biomasse. Malgré ça, la façon de contraindre les FBA en pratique est floue et ça semble toujours extrêmement sous-contraint... La deuxième partie de l'article fait des comparaisons avec des expériences de croissance en chemostat et en milieu limité. L'article fait un focus sur un switch de sécrétion et d'utilisation d'acétate. Celui-ci servirait de puit rédox à haut taux de croissance à la place de l'oxygène et de source de carbone à haute densité (à voir plus en détail peut-être).

1.2 RBA

Principe Le principe est expliqué dans Goelzer et al. (2011) et est brièvement reproduit sur la Figure 1. L'article contient les définitions et preuves du modèle.

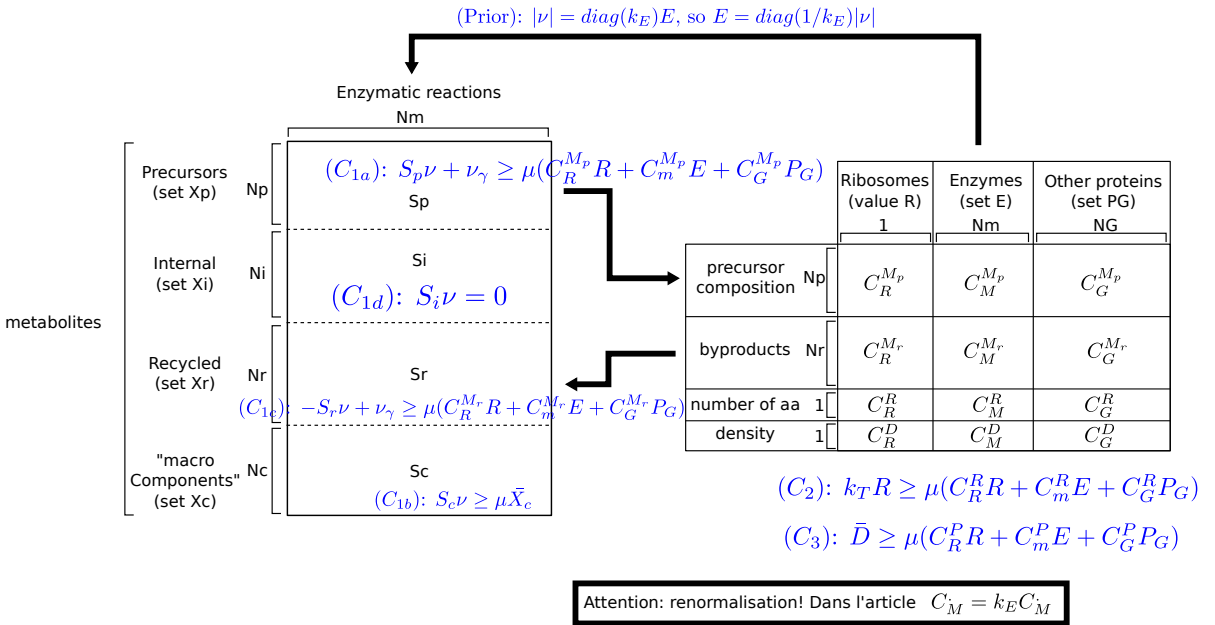


FIG. 1: Schéma de fonctionnement des RBA.

Formulation mathématique

Interprétations mathématiques

Applications Goelzer et al. (2011) applique les RBA à la croissance chez *B. subtilis*. Les cellules croissent dans un milieu qui varie de riche à pauvre. Plus le milieu est riche, plus les bactéries croissent rapidement. Le premier résultat positif est l'augmentation de

la part de ribosome et la diminution de la part de protéines métaboliques dans le pool total des protéines. Les auteurs montrent mathématiquement que les RBA permettent de désactiver complètement un pathway à cause de la quantité de protéines nécessaires à maintenir des pathways alternatifs. Quand un a.a. est ajouté au milieu, on observe que sa synthèse *de novo* cesse complètement (ou presque).

La solution des RBA fournit un jeu de régulations qui permettrait d'obtenir un taux de croissance optimal. On peut s'attendre à ce que ces régulations soient effectivement implémentées. D'après Goelzer et al. (2011), on a un assez bon accord entre prédictions et données. Un exemple avec un switch sur des transporteurs à basse et haute affinités est donné pour fournir une illustration simple du phénomène.

Une analyse un peu différente est présentée dans Goelzer et al. (2015). Le but est d'estimer les répartitions globales des ressources chez *B. subtilis*. Le modèle prend en compte le folding via l'introduction de chaperones (mathématiquement de la même façon que les ribosomes). Grâce à de nombreuses données, les efficacités enzymatiques sont estimées (elles étaient fixées de façon *ad hoc* dans Goelzer et al. (2011)). Cela permet de donner un coût en nombre de protéines aux différents sous-processus de la cellule. Une analyse de sensibilité montre que les paramètres liés à la répartition des protéines (densité, fraction de protéines cytosoliques, etc.). Dans le supplementary material, les auteurs montrent que les résultats se recoupent partiellement avec une analyse FBA, mais les FBA échouent à allouer correctement les protéines. Globalement le modèle proposé est beaucoup plus riche que dans l'article précédent. J'ai raté beaucoup de subtilités il faudra que je revienne un peu dessus au fur et à mesure.

1.3 Description du métabolisme

Varma and Palsson (1993a,b) décrivent assez précisément le catabolisme de base (avec beaucoup de sources à l'appui). On y trouve la description du pathway glycolitique (récupération du glucose), la voie des pentoses phosphates (production de R5P et E4P), le cycle de Krebs (production d'ATP/GTP, NADH) et ses coréactions (shunt glyoxalique, réactions anaplérotiques). On trouve également la description des chaînes de transport d'électrons et déshydrogénases (production d'ATP, échanges NADH/NADPH/ATP).

Références

- D A Fell and J R Small. Fat synthesis in adipose tissue. An examination of stoichiometric constraints. *Biochem J*, 238(3) :781–786, September 1986. ISSN 0264-6021.
- Anne Goelzer, Vincent Fromion, and Gérard Scorletti. Cell design in bacteria as a convex optimization problem. *Automatica*, 47(6) :1210–1218, June 2011. ISSN 00051098. doi : 10.1016/j.automatica.2011.02.038.
- Anne Goelzer, Jan Muntel, Victor Chubukov, Matthieu Jules, Eric Prestel, Rolf Nölker, Mahendra Mariadassou, Stéphane Aymerich, Michael Hecker, Philippe Noirot, Dörte

- Becher, and Vincent Fromion. Quantitative prediction of genome-wide resource allocation in bacteria. *Metabolic Engineering*, 32 :232–243, November 2015. ISSN 10967176. doi : 10.1016/j.ymben.2015.10.003.
- Joanne M. Savinell and Bernhard O. Palsson. Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism. *Journal of Theoretical Biology*, 154(4) :421–454, 1992a. ISSN 0022-5193. doi : 10.1016/S0022-5193(05)80161-4.
- Joanne M. Savinell and Bernhard O. Palsson. Network analysis of intermediary metabolism using linear optimization. II. Interpretation of Hybridoma Cell Metabolism. *Journal of Theoretical Biology*, 154(4) :455–473, 1992b. ISSN 0022-5193. doi : 10.1016/S0022-5193(05)80162-6.
- A. Varma and B. O. Palsson. Metabolic capabilities of Escherichia coli : I. synthesis of biosynthetic precursors and cofactors. *J. Theor. Biol.*, 165(4) :477–502, December 1993a. ISSN 0022-5193.
- A Varma and B O Palsson. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. *Appl Environ Microbiol*, 60(10) :3724–3731, October 1994. ISSN 0099-2240.
- Amit Varma and Bernhard O. Palsson. Metabolic capabilities of Escherichia coli : II. Optimal Growth Patterns. *J. Theor. Biol.*, 165(4) :503–522, 1993b. ISSN 0022-5193. doi : 10.1006/jtbi.1993.1203.