

Licence informatique — ALGO

Modélisation stœchiométrique

Les systèmes biologiques intracellulaires se modélisent couramment sous la forme de systèmes de réactions chimiques. Certaines propriétés de ces systèmes peuvent s'étudier par programmation linéaire.

Le texte qui suit constitue non seulement un projet mais aussi une préparation aux examens de l'UE ALGO pour l'année 2007–2008.

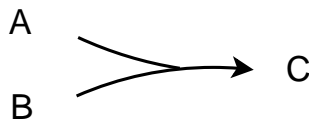
1 Réactions chimiques, stœchiométrie

Le but de cette section ? Expliquer la notion de coefficient *stœchiométrique* et la signification de certains diagrammes qui apparaissent dans les *réseaux de réactions* chimiques.

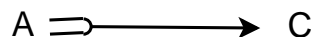
Quelques exemples. La plus simple des réactions chimiques décrit la transformation d'une *espèce* A (le *réactant*) en une espèce B (le *produit*) : à chaque instant, toute molécule de l'espèce A est susceptible de se transformer en une molécule de l'espèce B . Lorsqu'elle se produit, la réaction consomme une molécule de A : le coefficient stœchiométrique de A dans cette réaction vaut -1 . La réaction produit une molécule de B : le coefficient stœchiométrique de B dans cette réaction vaut 1 . Dans les réseaux de réactions, cette réaction est représentée par :



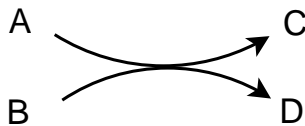
Une autre réaction chimique élémentaire classique décrit la formation d'un polymère C par association de deux réactants A et B : à tout instant, une molécule de l'espèce A et une molécule de l'espèce B qui se rencontrent peuvent s'associer pour former une seule molécule de l'espèce C . Les coefficients stœchiométriques de A , B et C valent respectivement -1 , -1 et 1 pour cette réaction. Dans les réseaux de réactions, cette réaction est représentée par :



Dans le cas particulier où les deux réactants appartiennent à la même espèce A , les coefficients stœchiométriques de A et C valent -2 et 1 . Dans les réseaux de réactions, cette réaction est représentée par :



La réaction suivante décrit la transformation de deux espèces A et B en deux autres espèces C et D . En voici une interprétation possible : à tout instant, une molécule de l'espèce A qui rencontre une molécule de l'espèce B peut s'associer à une partie de la molécule B (produisant ainsi une molécule de l'espèce C) et laisser libre l'autre partie de la molécule B (considérée comme appartenant à l'espèce D). Les coefficients stœchiométriques de A , B , C et D valent respectivement -1 , -1 , 1 et 1 . Dans les réseaux de réactions, cette réaction est représentée par :



Autres cas. On se restreint à des réactions comportant *au plus deux réactants*¹. Par contre, le nombre de produits peut être supérieur à deux. Il arrive parfois qu'une même espèce A apparaisse à la fois comme réactant et comme produit d'une réaction². Dans ce cas, le coefficient stœchiométrique de A est égal au nombre de molécules produites moins le nombre de molécules consommées.

Réversibilité. En toute rigueur, toute réaction chimique est *réversible* (toutes les flèches sont à double sens, en quelque sorte). En pratique, certaines réactions peuvent être considérées comme irréversibles parce que la vitesse de la réaction (nous n'en avons pas parlé) est beaucoup plus importante dans un sens que dans l'autre. Dans le cas d'une réaction réversible, le signe des coefficients stœchiométriques est *a priori* indéfini puisque chaque espèce est à la fois un réactant et un produit. Pour lever cette ambiguïté et fixer le signe des coefficients stœchiométriques, on fixe arbitrairement une *direction privilégiée* pour chaque réaction réversible et on calcule les coefficients stœchiométriques comme si la réaction était irréversible, dans la direction privilégiée. Dans les réseaux de réactions, les réactions réversibles apparaissent sous la forme de flèches à double sens. On fait ressortir la direction privilégiée en augmentant la taille de la pointe dans cette direction. Le dessin suivant correspond à une réaction réversible. La direction privilégiée est choisie de A vers B . Les coefficients stœchiométriques de A et de B valent respectivement -1 et 1 .

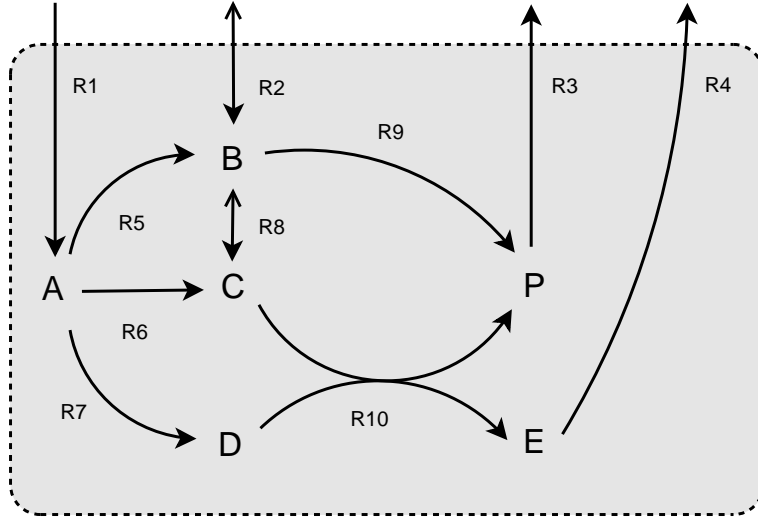


¹En décomposant les réactions et en introduisant des espèces intermédiaires, il est toujours possible de se ramener à ce cas-là.

²Du point de vue de la stœchiométrie, rien ne distingue les réactions $A \longrightarrow B$ et $A + A \longrightarrow A + B$. Toutefois, les interprétations sont différentes et, si l'on poussait l'étude jusqu'à la simulation numérique, on obtiendrait des évolutions différentes pour les deux cas.

2 Exemple de réseau de réactions

Le dessin suivant [3] représente un réseau de réactions. Il fait intervenir $m = 6$ espèces *internes* et $q = 10$ réactions dont deux sont réversibles : R_2 et R_8 . Les frontières (matérialisées en pointillé) et les connexions avec le « reste du monde » sont des caractéristiques importantes d'un réseau. Les espèces internes sont celles explicitement modélisées. Les espèces externes peuvent vivre à l'extérieur du système (c'est-à-dire de la cellule, du noyau etc.) mais aussi à l'intérieur. Ce peut par exemple être le cas d'espèces en très grande quantité, comme de l'eau.



À un réseau de réactions, on associe une *matrice de stœchiométrie*, notée N , de dimension $m \times q$. Il y a une ligne par espèce et une colonne par réaction. Ligne ℓ , colonne c , on trouve le coefficient stœchiométrique de l'espèce ℓ dans la réaction c . Sur l'exemple,

	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9	R_{10}
A	1	0	0	0	-1	-1	-1	0	0	0
B	0	1	0	0	1	0	0	-1	-1	0
C	0	0	0	0	0	1	0	1	0	-1
D	0	0	0	0	0	0	1	0	0	-1
E	0	0	0	-1	0	0	0	0	0	1
P	0	0	-1	0	0	0	0	0	1	1

3 Hypothèse d'équilibre et flux

On suppose que le système a atteint un état d'équilibre et on s'intéresse aux *flux de molécules* transformées par les réactions par unité de temps (un nombre de secondes ou de minutes). Supposer l'état d'équilibre, c'est supposer que les concentrations des différentes espèces (les quantités³ de molécules si l'on préfère) ne varient pas. C'est donc supposer que

³Une concentration est un nombre de molécules divisé par un volume. Les deux notions peuvent être utilisées indifféremment ici puisque les volumes ne sont pas modélisés. Toutefois, en chimie, les réactions

toute transformation d'une molécule d'une espèce A en une espèce B est compensée par la transformation d'une molécule d'espèce B en autre chose.

À chaque réaction R_i , on associe un nombre réel : le *flux* f_i emprunté par la réaction. Dans le cas d'un réseau comportant q réactions, un flux est un vecteur colonne f de \mathbb{R}^q . Dans le cas d'une réaction R_i dont les réactants et les produits sont distincts deux-à-deux, l'interprétation de f_i est simple : c'est le nombre de molécules de chaque espèce consommées ou produites par la réaction. Si R_i est irréversible alors $f_i \geq 0$. Si R_i est réversible alors $f_i \geq 0$ si la réaction s'effectue dans la direction privilégiée sinon on a $f_i \leq 0$. Dans le cas général, l'interprétation est plus compliquée. Le plus simple consiste à définir les flux en supposant le système à l'équilibre :

Définition 1 (*flux à l'équilibre*)

En notant m le nombre d'espèces internes et q le nombre de réactions, un flux est un vecteur colonne f de \mathbb{R}^q satisfaisant les conditions suivantes (où 0 désigne le vecteur colonne nul de dimension m) :

$$N \cdot f = 0 \quad \text{et} \quad f_i \geq 0 \text{ si } R_i \text{ est irréversible.}$$

Le vecteur suivant offre un exemple de flux. On peut l'interpréter ainsi : une molécule d'une espèce externe est transformée en une molécule de A par R_1 . Au même moment, une molécule de A est transformée en C par R_6 , une molécule de C est transformée en B par R_8 (comme la réaction réversible est prise dans la direction non privilégiée, le flux f_8 est négatif), une molécule de B est transformée en P par R_9 et une molécule de P est transformée en une espèce externe par R_3 .

$$f = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \\ f_6 \\ f_7 \\ f_8 \\ f_9 \\ f_{10} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ -1 \\ 1 \\ 0 \end{pmatrix}.$$

sont fonctions des concentrations et pas des quantités.

Un autre exemple est donné par le vecteur suivant.

$$f = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \\ f_6 \\ f_7 \\ f_8 \\ f_9 \\ f_{10} \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 1.5 \\ 1.5 \\ 0 \\ 0.5 \\ 1.5 \\ 1 \\ 0 \\ 1.5 \end{pmatrix}.$$

Question 1. Donner une méthode permettant de modéliser en AMPL un réseau de réactions chimiques. Le modèle ne devrait normalement pas comporter d'objectif économique. Illustrer votre méthode sur l'exemple précédent. Pour éviter la solution triviale $f = 0$, on peut supposer qu'on dispose de mesures pour certaines réactions (prendre par exemple $f_1 = 2$ et $f_2 = 1$). Quel est le flux calculé par AMPL ?

4 Le réseau de production de biomasse d'*E. Coli*

Bien que fortement inspiré de données compilées dans des articles sérieux [1, 4, 2], ce sujet est, sous certains aspects, un peu fantaisiste.

Escherichia Coli est une bactérie intestinale très commune chez les mammifères. On la rencontre chez l'être humain. Le réseau de réactions de la figure 1 décrit une partie du mécanisme de production de biomasse à partir de glucose [1]. Ce réseau est composé de trois sous-réseaux. Deux d'entre eux correspondent à des séquences de réactions qui convertissent du glucose en pyruvate (PYR) : la glycolyse et la voie métabolique des pentoses-phosphates. Le troisième représente le cycle de l'acide citrique, qui transforme le pyruvate ainsi produit.

L'ensemble du réseau est constitué de vingt et une réactions internes assez bien identifiées, notées v_1, \dots, v_{21} , d'une réaction introduisant du glucose depuis l'extérieur de la cellule, de onze réactions produisant de la biomasse et de quatre réactions produisant d'autres composés (acétate, pyruvate, oxoglutarate et fumarate). La biomasse et les quatre composés sont produits vers l'extérieur de la cellule.

Parmi les différentes espèces chimiques, le composé NADH, qui représente une source d'énergie, est consommé par la bactérie. Cette consommation du NADH n'est pas représentée dans le réseau de la figure 1. D'autres composés tels que NAD^+ et NADP^+ peuvent être considérés comme présents en très grande quantité dans la cellule.

La réaction chimique v_2 fait l'objet d'une remise en cause : certains travaux [2] suggèrent que les phénomènes représentés par cette réaction seraient mieux modélisés par l'ensemble des trois réactions v_{19} , v_{20} et v_{21} .

Ce sujet consiste à étudier cette question.

On s'intéresse pour commencer au type sauvage⁴ d'*E. Coli*. Pour certaines réactions du réseau, qui ont pu être mesurées, on dispose de fourchettes de valeurs possibles, récapitulées dans le tableau suivant.

<i>flux</i>	<i>Réactions productrices de biomasse</i>										
	G6P	P5P	F6P	T3P	E4P	PGA	PEP	PYR	ACA	OGA	OAA
<i>min.</i>	1	2	0	0	20	9	7	20	8	8	8
<i>max.</i>	3	4	1	2	30	11	9	25	20	12	12
	<i>Autres réactions</i>										
	v_4	v_5	v_6	v_{13}	v_{14}	v_{18}					
<i>min.</i>	2	30	10	70	70	83					
<i>max.</i>	10	130	50	130	130	∞					

Pour reconstituer les flux des réactions manquantes chez le type sauvage, on fait l'hypothèse que, sous l'effet de la sélection naturelle, le réseau maximise la production de biomasse à partir du glucose. Les onze réactions productrices de biomasse ne sont toutefois pas considérées comme équivalentes : la quantité de biomasse produite par chacune de ces réactions s'obtient en multipliant le flux par un coefficient déterminé empiriquement. Les onze coefficients sont synthétisés dans le tableau suivant.

<i>Coefficients empiriques</i>										
G6P	P5P	F6P	T3P	E4P	PGA	PEP	PYR	ACA	OGA	OAA
1.0	1.1	.9	.9	1.15	1.2	1.25	1.2	.9	1.3	1.25

Question 2. Ecrire un modèle AMPL pour le problème ci-dessus. Votre modélisation doit permettre à l'utilisateur de choisir facilement entre l'étude du modèle élémentaire (avec v_2) et celle du modèle plus sophistiqué (avec les trois réactions v_{19} , v_{20} et v_{21}). Déterminer un ensemble de flux qui maximise la production de biomasse dans les deux cas.

Les contraintes énoncées précédemment ne suffisent pas à assurer l'unicité de la solution optimale. Pour réduire davantage les possibilités, on fait l'hypothèse que, dans le cas des flux pour lesquels on dispose d'une fourchette avec des bornes *finies*, le flux est proche de la valeur du milieu de la fourchette.

Question 3. Modifier le programme précédent pour qu'il détermine, parmi les flux réalisant la quantité de biomasse optimale, une solution qui minimise le maximum des écarts (en valeur absolue) entre chaque flux et la valeur du milieu de la fourchette (quand elle existe).

Vous pouvez résoudre la question posée en deux temps.

Etudier les deux modèles (l'élémentaire et le sophistiqué).

⁴Le type d'*E. Coli* qu'on rencontre à l'état naturel.

En laboratoire, une souche mutante, viable, a été obtenue en réprimant le gène responsable de la réaction v_9 [1, *pyruvate kynase knockout*]. On estime que cette mutation a pour effet de fixer à 100 le flux transitant par v_9 .

Question 4. Résoudre le programme linéaire obtenu à la question précédente pour qu'il reproduise la mutation.

Que peut-on en déduire à propos des deux modèles ?

Il n'est pas forcément pertinent d'imaginer que, dans le cas du mutant, les flux maximisent la production de biomasse puisque la pression due à la sélection naturelle n'a pas eu le temps d'opérer. Une autre hypothèse (hypothèse de l'ajustement métabolique minimal [4, MOMA]) consiste à imaginer que chez le mutant, les flux sont aussi proches que possible des flux optimaux de l'espèce sauvage.

Question 5. Modifier le programme linéaire pour qu'il détermine, chez le mutant, une solution qui minimise le maximum des écarts (en valeur absolue) entre les flux chez le mutant et les flux optimaux chez l'espèce sauvage, obtenus à la question précédente.

Étudier la variante qui consiste à minimiser la somme des valeurs absolues des écarts.

Question 6. Selon certains auteurs [3], la non nécessaire unicité de la solution optimale des programmes linéaires limite la pertinence de la méthode précédente. Argumenter.

5 Consignes

Le projet est à rendre à votre enseignant de *Travaux Dirigés* le 3 mai 2008, au plus tard. Un projet peut être réalisé seul ou en binôme. Les n -nômes pour $n \geq 3$ sont interdits.

Le projet est optionnel. La note compte comme une note d'interrogation écrite. Une partie de sujet d'examen des deux sessions sera liée à ce projet. On vous conseille donc fortement de le réaliser.

Le projet doit être rendu par voie électronique, sous la forme d'une archive (fichier « .zip » ou « .tar.gz ») contenant les modèles AMPL correspondant aux différentes questions ainsi qu'un compte-rendu au format PDF. Le nom de l'archive doit être fabriqué à partir du ou des noms de famille de ses auteurs.

Le compte-rendu doit rappeler le nom du ou des auteurs. Il doit expliquer les choix de modélisation et permettre la compréhension des modèles AMPL. Les réponses aux questions doivent impérativement être rédigées (les tableaux de nombres produits par la commande *display* d'AMPL ne sont absolument pas suffisants : ils doivent être interprétés).

Les modèles AMPL doivent être lisibles et commentés. Ils doivent être fournis dans des fichiers séparés, faciles à exécuter. Leur identificateur doit comporter le numéro de la question à laquelle ils correspondent (ex : *modele-Q5.ampl*).

D'une façon générale, les projets considérés comme anormalement compliqués, mal expliqués ou difficiles à vérifier ne seront pas corrigés.

Références

- [1] Marcel Emmerling, Michael Dauner, Aaron Ponti, Jocelyne Fiaux, Michel Hochuli, Thomas Szyperski, Kurt Wüthrich, J. E. Bailey, and Uwe Sauer. Metabolic Flux Responses to Pyruvate Kinase Knockout in *Escherichia Coli*. *Journal of Bacteriology*, 184(1) :152–164, 2001.
- [2] Eliane Fischer, Nicola Zamboni, and Uwe Sauer. High-throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived ^{13}C constraints. *Analytical Biochemistry*, 325 :308–316, 2003.
- [3] Steffen Klamt and Jörg Stelling. Stoichiometric and Constraint-based Modeling. In Zoltan Szallasi, Jörg Stelling, and Vipul Periwal, editors, *System Modeling in Cell Biology : From Concepts to Nuts and Bolts*, pages 73–96. Cambridge, Massachussets : The MIT Press, 2006.
- [4] Daniel Segrè, Dennis Vitku, and George M. Church. Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of The National Academy of Science of the USA*, 99(23) :15112–15117, 2002.

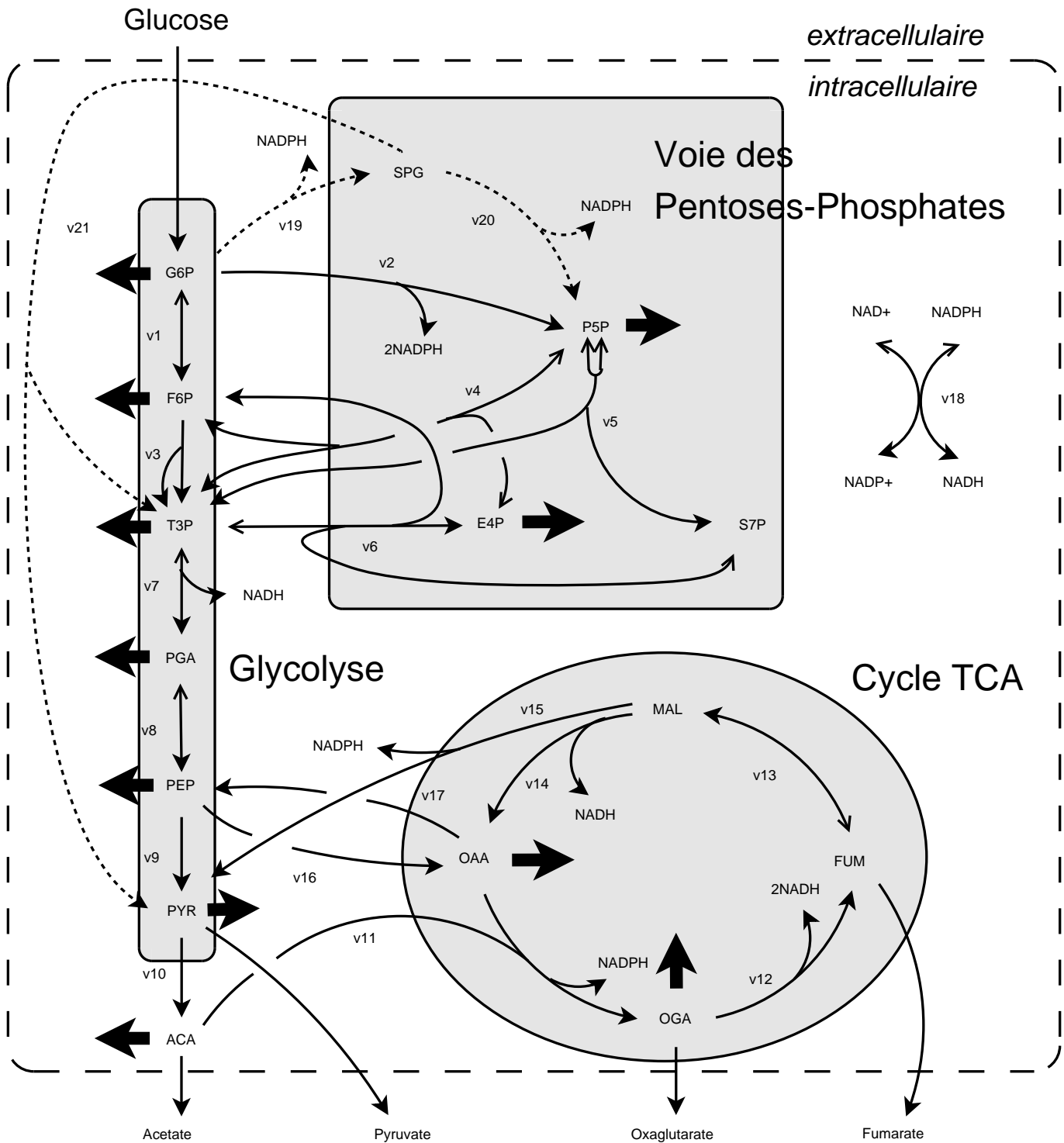


FIG. 1 – Les flèches en trait plein représentent des réactions connues. Les flèches en pointillé représentent des réactions hypothétiques. Les flux vers des unités de production de bio-masse sont représentés par les flèches noires épaisses. La direction privilégiée des réactions réversibles est matérialisée par une pointe plus épaisse.