

expérimental. Ceci revient à propager dans le graphe d'interactions l'information fournie par les expérimentations. On prédit ainsi le comportement d'un certain nombre de constituants non observés.

En pratique, sur le réseau d'interactions d'*E. coli* incluant les facteurs sigma, les données concernant le stress nutritionnel portent sur 40 produits et prédisent la variation de 381 molécules supplémentaires, qui sont validées à 70 % par des données « transcriptome » (figure 2) (10). Les 30 % de prédictions non validées indiquent des défauts du modèle qui doit être précisé.

Mesure alternative

Partant de connaissances et de données incomplètes sur un système, cette démarche permet d'évaluer la validité d'un modèle, puis de guider les expé-

rimentations qui permettront de le préciser, en quantifiant l'importance des produits pour la validation du modèle. Plus généralement, ces méthodes suggèrent une mesure alternative de l'importance fonctionnelle d'un groupe de composants d'un réseau, basée sur le pouvoir prédictif de leur observation et prenant en compte le pourcentage du réseau qui est contraint par l'observation d'un jeu de variables.

On pourra en particulier comparer cette approche à la théorie statistique des réseaux biologiques (11), où l'importance d'un sommet est en rapport avec le nombre total de connexions. Les réseaux transcriptionnels des procaryotes suggèrent ainsi que les nœuds de grande valence sont les plus conservés au cours de l'évolution. Il sera intéressant de voir si le pouvoir prédictif défini par les contraintes qualitatives confirme ces suggestions. ●

(11) Barabasi AL, Albert R (1999) *Science* 286, 509-12

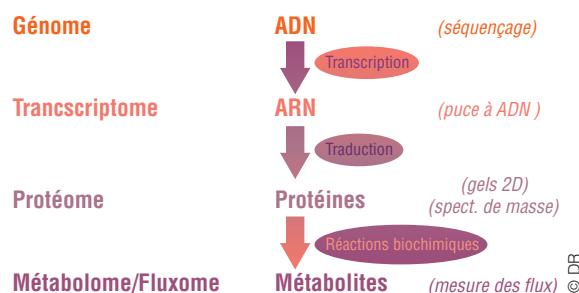
Structure et modélisation des réseaux métaboliques

La connaissance du génome d'un organisme, associée au travail antérieur des biochimistes, permet une bonne connaissance de son métabolisme. La complexité des réseaux métaboliques oblige cependant à se concentrer, dans un premier temps, sur l'étude de leur structure, indépendamment des propriétés cinétiques des enzymes. Ces analyses ainsi limitées produisent cependant des résultats extrêmement informatifs et pertinents.

Jean-Pierre Mazat^{*,**}, Christine Nazaret^{***}, Sabine Pérès^{*}

La connaissance du génome d'un organisme, associée au travail antérieur des biochimistes, permet une bonne connaissance de son métabolisme. La complexité des réseaux métaboliques oblige cependant à se concentrer, dans un premier temps, sur l'étude de leur structure, indépendamment des propriétés cinétiques des enzymes. Ces analyses ainsi limitées produisent cependant des résultats extrêmement informatifs et pertinents.

En principe, la connaissance du génome entraîne la connaissance du métabolisme de l'organisme (figure 1). Les biologistes ont pendant longtemps suivi une approche « réductionniste » dans laquelle les composants individuels du système vivant ont été étudiés séparément. Avec l'arrivée des données de masse (séquençage, puce à ADN, etc.), la procédure s'in-



verse et permet désormais d'étudier comment ces composants interagissent pour former un système complexe en utilisant une approche « intégrée ». L'objectif de la biologie intégrative est ainsi de comprendre comment les différents éléments interagissent

* Inserm U688 et Programme d'épigénomique, Genopole®, Évry
 ** CNRS 5466, université Bordeaux 2, 146 rue Léo Saignat, 33076 Bordeaux cedex
 *** jpm@u-bordeaux2.fr

Figure 1
 Du génome au métabolome

dans un système donné afin d'établir des relations qui conduisent d'un génotype à son expression phénotypique. C'est pourquoi l'étude du métabolome est une étape nécessaire.

Une des premières questions est de comprendre comment est structuré le métabolisme et plus particulièrement comment cette structuration peut être mise en relation avec des phénotypes particuliers. De plus, nous montrerons que l'étude structurale d'un modèle peut permettre des prédictions et/ou des pistes expérimentales.

Qu'est ce qu'un réseau métabolique ?

Un réseau métabolique est constitué de réactions élémentaires catalysées par des enzymes. Ces dernières sont des catalyseurs qui permettent d'accélérer, voire de rendre simplement possibles, les réactions du métabolisme. Il n'est donc pas étonnant que pendant de nombreuses années les biochimistes se soient penchés sur les propriétés cinétiques des enzymes isolées, leur associant une équation de vitesse traduisant le mécanisme plus ou moins intime de la transformation du substrat en produit. L'archétype d'équation de cinétique enzymatique est l'équation de Michaelis-Henri (1) :

$$v = \frac{V_M \cdot S}{K_M + S}$$

où $v = -dS/dt$ et K_M est la constante de Michaelis et V_M est la vitesse maximale représentant la quantité d'enzyme et sa capacité catalytique. Malheureusement, ces paramètres ont beaucoup été étudiés à des concentrations de substrats et de produits particulières, souvent éloignées de leurs concentrations cellulaires. Cependant, dans la cellule, le réseau n'est pas l'addition de ces réactions isolées. Toute perturbation d'une réaction sera transmise via des variations des métabolites intermédiaires aux autres réactions du réseau ; on peut imaginer qu'ils agissent comme des ressorts dont la tension s'établit et varie en fonction des vitesses des différentes réactions.

En fait, dans un réseau, les concentrations des métabolites sont déterminées par la structure du réseau (c'est-à-dire les réactions qui utilisent ou produisent ce métabolite) et les cinétiques des réactions impliquées (qui dépendent des concentrations des métabolites intermédiaires, des substrats, des produits et éventuellement de régulateurs). La complexité d'un réseau métabolique réside dans cette interconnexion des réactions.

Première hypothèse : le milieu cellulaire¹ est homogène

Ce principe n'est évidemment en général pas vérifié. Il existe de nombreux exemples de compartimentation et de canalisation (*channeling* en anglais) (2). Mais il s'avère que la plupart des études quantitatives du métabolisme considèrent, souvent sans le dire explicitement, que ce principe est satisfait. En effet, l'étude d'un réseau métabolique dans un réseau inhomogène est beaucoup plus difficile alors que celle du même réseau dans un

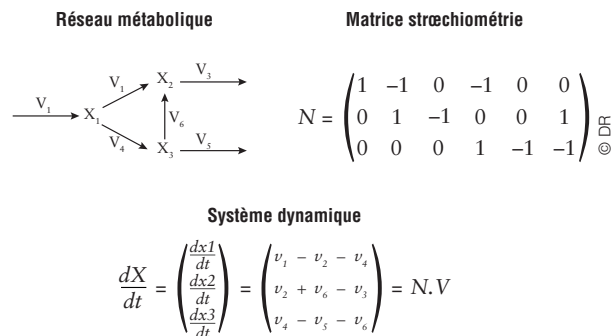


Figure 2 Exemple de matrice de stœchiométrie

X_1 , X_2 et X_3 représentent les métabolites intermédiaires du réseau métabolique et dx_1/dt , dx_2/dt et dx_3/dt , la dérivée de leur concentration, c'est-à-dire leur variation instantanée. v_1, \dots, v_6 représentent les vitesses de chaque étape.

milieu homogène peut fournir des indications semi-quantitatives sur le comportement du réseau métabolique réel, que le milieu soit homogène ou non ; cela revient à considérer un milieu homogène comme un cas limite, idéal, d'un milieu hétérogène. Dans toute la suite de ce chapitre, nous considérerons donc ce principe vérifié.

Représentation d'un réseau métabolique : la matrice de stœchiométrie

Considérons un réseau métabolique comprenant r réactions repérées par leur enzyme e_1, e_2, \dots, e_r ou leur vitesse v_1, v_2, \dots, v_r et m métabolites x_1, x_2, \dots, x_m avec des stœchiométries n_{ij} , c'est-à-dire que n_{ij} molécules du métabolite x_i réagissent dans la réaction v_j . Il est utile pour les calculs et l'analyse théorique de la structure des réseaux métaboliques de représenter un réseau métabolique par un objet mathématique manipulable théoriquement : la matrice de stœchiométrie (3), définie par $N = (n_{ij})_{1 \leq i \leq m, 1 \leq j \leq r}$. Un exemple en est donné dans la figure 2.

On définit aussi le vecteur des vitesses $V = \begin{pmatrix} v_1 \\ M \\ v_r \end{pmatrix}$

Notons que les stœchiométries sont comptées positivement pour l'apparition du métabolite et négativement pour sa consommation. Cela suppose, en particulier pour les réactions réversibles, que l'on définit un sens de réaction.

Trois remarques sur le réseau et la matrice :

- dans notre réseau métabolique, la simple flèche représentant une réaction n'implique pas que cette dernière soit irréversible dans le sens de la flèche. Elle indique simplement une orientation arbitraire de la réaction qui permet de donner un signe à la fonction de vitesse v_j ;
- il existe une correspondance bi-univoque entre un tel réseau métabolique et sa matrice de stœchiométrie (en ne tenant pas compte de la réversibilité ou non des réactions) : la construction de cette matrice est complètement déterminée par le réseau métabolique et on peut reconstruire un réseau métabolique et un seul à partir d'une matrice de stœchiométrie donnée ;

(1) Henri V (1902) *Compte Rend Acad Sci* 135, 916-19
 (2) Agius L, Sherratt HSA (ed) (1997) *Channelling in intermediary metabolism*, Portland Press, London
 (3) Reder C (1988) *J Theor Biol* 135, 175-201

- notons enfin que les coefficients ne sont pas nécessairement des 1, des -1 ou des 0 ; des coefficients entiers différents (2, -3, etc.) peuvent aussi être rencontrés en fonction des réactions constituant le réseau métabolique.

Système dynamique

Dans un réseau métabolique, la variation du métabolite x_i est naturellement le résultat de sa production et de sa consommation ce qui peut s'écrire à chaque instant, à l'aide d'une équation différentielle :

$$\frac{dx_i}{dt} = \sum_{j=1}^r n_{ij} v_j$$

Ce système dynamique peut s'écrire sous forme matricielle :

$$d[X]/dt = N.V (X,p)$$

où N est la matrice de stœchiométrie du réseau et V le vecteur des vitesses (figure 2), qui dépend des métabolites x_i symbolisés par le vecteur X et d'un certain nombre de paramètres représentés par p .

Notons que, sans l'hypothèse d'homogénéité, ce système serait remplacé par un système d'équations aux dérivées partielles plus complexe faisant intervenir les dérivées des concentrations par rapport à des variables d'espace.

La séparation de la structure et de la cinétique

Dans l'équation $d[X]/dt = N.V$, la structure du réseau métabolique représenté par la matrice de stœchiométrie N (qui traduit l'existence de boucles, de branches, ou de parties linéaires etc.) est séparée des équations de vitesse qui sont rassemblées dans le vecteur V .

C'est l'étude de la matrice de stœchiométrie N qui va permettre de déduire des propriétés structurelles des solutions du système différentiel.

L'état stationnaire : noyau de la matrice de stœchiométrie

Pour calculer les flux dans un réseau métabolique, on a l'habitude de considérer que les concentrations des métabolites sont constantes, c'est-à-dire que leur production s'équilibre avec leur consommation. Cette hypothèse est assez bien confirmée expérimentalement dans de nombreux cas sur des temps de l'ordre de quelques minutes et parfois plus. Si ce n'était pas le cas, ces mesures seraient d'ailleurs beaucoup plus délicates.

Cependant, il est bien clair que cette hypothèse d'état stationnaire ne tient pas lorsque l'on change de conditions physiologiques : passage pour le muscle de l'état de repos à l'exercice, variation de la température ou de l'ensoleillement pour des plantes au cours de la journée.

Enfin, même dans des états physiologiques donnés, certains systèmes peuvent être oscillants ; c'est le cas bien connu de la glycolyse ; c'est aussi celui de la sécré-

tion d'insuline qui s'accompagne de tout un métabolisme oscillant, en particulier calcique dans les cellules β du pancréas.

Pour la suite, on acceptera l'hypothèse d'état stationnaire, tout en sachant qu'il s'agit d'une hypothèse usuelle qui n'est pas toujours satisfaite.

Remarquons pour terminer que l'étude expérimentale d'états non stationnaires peut être plus riche en résultats que celle d'états stationnaires, qui peut être d'ailleurs difficile à réaliser expérimentalement. L'étude dynamique d'un réseau métabolique peut en particulier permettre d'isoler des comportements avec différentes échelles de temps (4).

La représentation matricielle permet aussi d'écrire simplement les conditions d'état stationnaire sous la forme :

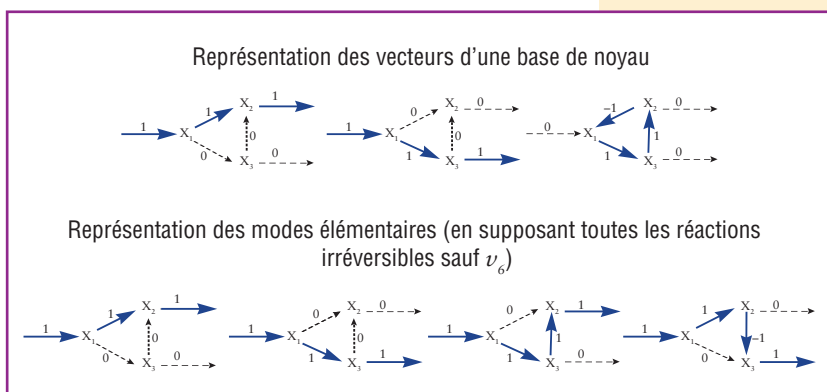
$$d[X]/dt = N.V = 0$$

et d'en tirer d'importantes conséquences en termes d'algèbre matricielle : cela signifie que le vecteur vitesse V appartient à ce que l'on appelle le « noyau » de la matrice. La notion de noyau de la matrice N traduit de manière mathématique et condensée la propriété biochimique simple selon laquelle, à l'état stationnaire, pour tous les métabolites, la production est égale à la consommation. Son avantage est que l'on peut profiter pour l'analyse des réseaux métaboliques des propriétés générales (que nous ne détaillerons pas ici) des noyaux d'applications linéaires représentées par des matrices. En particulier, nous utiliserons le fait que le noyau est un espace vectoriel.

En principe, la description du noyau de N permet de décrire l'ensemble des vitesses (i.e. le flux) possibles à l'état stationnaire. Des algorithmes permettent de trouver une base du noyau et de faire cette description (5). On a en apparence résolu le problème de la description des flux dans un réseau métabolique. Mais les biochimistes aimeraient bien avoir tous les chemins réellement possibles dans un réseau métabolique ; certains sont en effet impossibles car un certain nombre de réactions sont irréversibles. Ce problème a conduit à la définition des modes élémentaires de flux.

Figure 3 Base du noyau et modes élémentaires de l'exemple

Les flèches pleines indiquent les étapes (réactions) empruntées et les flèches en pointillés les étapes qui ne fonctionnent pas dans le schéma considéré. Les chiffres 0, 1, etc. indiquent le flux relatif à travers l'étape.



(4) Raïs B *et al.* (2001) *Biochem J* 356, 425-32
(5) Letellier T *et al.* (1991) *Comp Appl Biosci* 7, 383-90

Modes élémentaires de flux

Un mode élémentaire de flux (*elementary flux mode* ou *efm* en anglais) est un ensemble minimal de réactions qui opèrent à l'état stationnaire et respectent le sens des réactions irréversibles (6). La minimalité signifie que la suppression d'une quelconque des réactions de l'ensemble empêchera tout état stationnaire.

Notons qu'une description pratiquement identique, appelée *extreme pathways*, a été proposée par Palsson (7).

L'ensemble des modes élémentaires est un sous-ensemble du noyau de la matrice de stœchiométrie et correspond à tous les chemins (minimaux) possibles, conduisant d'un ou de plusieurs métabolites d'entrée à un ou plusieurs métabolites de sortie ou à des cycles (figure 3).

Il apparaît que pour certains couples de métabolites d'entrée et de sortie plusieurs chemins sont possibles. Cela participe sans aucun doute à la robustesse bien connue des réseaux métaboliques et au fait assez souvent observé que l'inactivation d'un gène (enzyme, réaction) a peu d'effet sur la survie d'une cellule.

Intérêts d'une décomposition de réseau métabolique en modes élémentaires

L'utilisation des modes élémentaires permet de déterminer des propriétés biologiques liées à la structure du réseau :

- **identification de voies (métaboliques)** : l'ensemble des efms permet de déterminer toutes les voies admissibles du réseau. Le fait que ces voies soient réellement empruntées dépendra bien entendu des cinétiques des différentes enzymes catalysant les réactions du réseau ;
- **identification de voies optimales** : les efms permettent de déterminer la production optimale d'un métabolite à partir d'un ensemble de métabolites d'entrée (8) ;
- **corrélation des réactions** : des ensembles corrélés, appelés *enzyme subsets* (9), sont des groupes de réactions qui opèrent toujours ensemble avec une proportion de flux fixe dans tous les états stationnaires. Ils peuvent jouer un rôle important dans la compréhension des structures de régulation. Dans certains cas, chez les procaryotes, ces ensembles peuvent être dans le même opéron et chez les eucaryotes dans le même régulon ;
- **robustesse du réseau** : c'est la capacité du réseau à répondre à un changement de structure, comme l'inhibition d'une de ses étapes (10). Par exemple, cela permet d'étudier si la production d'un métabolite est toujours possible après suppression d'une ou de plusieurs réactions. Si une réaction apparaît dans tous les efms d'un réseau, alors elle est essentielle. L'absence de cette réaction peut être considérée comme létale, car elle a comme conséquence la suppression de tous les efms ;
- **délétions létales minimales, ou *minimal cut set*** : c'est un ensemble minimal de réactions d'un réseau tel que son inactivation empêche définitivement certaines fonctions du réseau, comme par exemple la production d'un métabolite (11).

Limitations

Le problème est que le nombre de modes élémentaires augmente très rapidement avec le nombre de réactions (en fait avec le degré de connexion des métabolites

intermédiaires). Par exemple, pour une représentation du métabolisme énergétique mitochondrial comprenant une quarantaine de réactions (selon les tissus ou les organismes) et une trentaine de métabolites, le nombre de modes élémentaires de flux varie autour de 6 000 (12). Même si l'on imagine qu'un certain nombre d'entre eux, pour des raisons cinétiques, sont à éliminer, on peut penser qu'il va en rester un grand nombre.

Pour des réseaux plus importants (au-delà de 500 réactions), il est actuellement impossible matériellement de calculer l'ensemble des efms.

La théorie du contrôle du métabolisme

Cette théorie (3, 13, 14) décrit comment un réseau métabolique répond à de petites perturbations au voisinage d'un état stationnaire. Deux types de coefficients sont définis : les coefficients d'élasticité, qui quantifient les variations des vitesses des étapes isolées, et les coefficients de contrôle, qui expriment la réponse globale du réseau aux perturbations d'une étape donnée (figure 4).

Une des conséquences importantes de la théorie est le fait que, pour tout réseau métabolique, la somme des coefficients de contrôle d'un flux est égale à 1 (2).

Cela a deux conséquences importantes :

- le contrôle d'un flux peut être distribué : ce résultat va contre l'idée de l'existence d'une étape limitante unique (qui donnerait sa vitesse au flux). Plusieurs étapes peuvent concourir à l'établissement d'un état stationnaire ;
- la distribution des coefficients de contrôle d'un flux peut changer lorsque l'on change d'état stationnaire : une étape contrôlante dans certaines conditions peut ne plus l'être dans d'autres. C'est un point qu'il faut garder en mémoire pour juger de l'action de médicaments (supposés agir sur des étapes qui peuvent n'être contrôlantes que dans certaines conditions).

La mesure des coefficients de contrôle montre qu'en général leur valeur est faible (la plupart du temps entre 0 et 0,3) ; cela veut dire que les réseaux métaboliques sont en général peu sensibles aux variations des activités enzymatiques des étapes qui le composent.

C'est une bonne nouvelle pour nous car cela signifie que notre métabolisme peut tolérer de fortes variations dans l'activité de nos enzymes jusqu'à un certain seuil (chez les hétérozygotes, les flux métaboliques sont en général identiques à ceux des homozygotes, malgré une activité environ deux fois plus faible (15)) ; et, de fait, dans la population normale, les activités de certaines enzymes peuvent varier d'un facteur 2,5.

Mais c'est une mauvaise nouvelle pour les biotechnologistes, car il y a peu d'espoir d'augmenter un flux par l'amplification d'un seul gène dans un réseau métabolique ; c'est un ensemble de gènes qu'il faudra amplifier.

Quelle réponse d'un réseau métabolique à l'élimination d'une ou plusieurs de ses étapes ?

Il s'agit de mesurer l'importance d'une réaction dans un réseau métabolique. Plusieurs méthodes peuvent être employées pour une telle étude :

- (6) Schuster S, Hilgetag C (1994) *J Biol Systems* 2, 165-82
 (7) Schilling CH *et al.* (2000) *J Theor Biol* 203, 229-48
 (8) Fell DA (1998) *Biotechnol Bioeng* 58, 121-4
 (9) Pfeiffer T *et al.* (1999) *Bioinformatics* 15, 251-7
 (10) Stelling J *et al.* (2002) *Nature* 420, 190-3
 (11) Klamt S, Gilles ED (2004) *Bioinformatics* 20, 226-34
 (12) Pérès S *et al.* (2007) TSI, sous presse
 (13) Kacser H, Burns JA (1973) *Symp Soc Exp Biol* 27, 65-104
 (14) Heinrich R, Rapoport RA (1974) *Eur J Biochem* 42, 89-95
 (15) Kacser H, Burns JA (1981) *Genetics* 97, 639-66

- Lemke *et al.* (16) utilisent une approche de type théorie des graphes et montrent que l'inactivation d'une large fraction d'enzymes (91 %) cause peu d'effet ;

- Lmielinski *et al.* (17) utilisent la méthode FBA (*Flux Balance Analysis*) développée par Palsson (18) pour décrire l'ensemble de toutes les espèces métaboliques qui sont éliminées par l'inactivation d'un gène ; cela leur permet d'analyser le caractère essentiel d'un métabolite pour un phénotype donné. Leurs résultats font apparaître le rôle déterminant d'un certain nombre de métabolites membranaires ou de la paroi et de quinones.

On peut aussi rechercher les modes élémentaires de flux qui sont éliminés lorsque l'on invalide un gène et donc une réaction.

Ces approches théoriques permettent de dégager un petit nombre (une trentaine) de gènes essentiels dans le métabolisme (et qui dépendent de l'organisme ou tissu étudié) et sont en assez bon accord avec les résultats expérimentaux d'inactivation de gènes. Les points de désaccords entre théorie et expérience sont évidemment intéressants à approfondir.

Quelles sont les voies métaboliques empruntées ?

L'analyse de la structure des réseaux métaboliques est une étape essentielle pour la compréhension de leur fonctionnement ; cela permet d'aborder de manière rigoureuse leur fonctionnement et dans un premier temps la sensibilité des réseaux métaboliques au voisinage des états stationnaires comme nous l'avons vu dans le paragraphe consacré à la théorie du contrôle du métabolisme, en séparant ce qui est dû au réseau (symbolisé par la matrice N) et ce qui est dû aux cinétiques. Cette étude montre que la réponse d'un réseau dans sa globalité (c'est-à-dire la réponse des flux à l'état stationnaire) peut être très différente de l'action sur les activités enzymatiques individuelles. Il y a un phénomène d'amortissement par le réseau des variations individuelles.

L'analyse de la structure d'un réseau métabolique en termes de modes élémentaires a le grand intérêt de représenter toutes les voies métaboliques qui peuvent être empruntées dans ce réseau. Le problème est que leur nombre croît extrêmement rapidement avec le nombre de réactions et de métabolites impliqués. À titre d'exemple, leur détermination pour le métabolisme d'*E. Coli*, avec plus de 700 réactions et 500 métabolites, dépasse les capacités de calculs actuels et même si ce n'était pas le cas, leur nombre considérable rendrait leur analyse difficile.

La question demeure néanmoins : tous les modes élémentaires d'un réseau métabolique sont-ils réellement empruntés ou bien simplement quelques-uns, et dans ce cas lesquels ?

La réponse à cette question n'est pas facile. Cela remet en cause le fait que les seules voies métaboliques empruntées sont celles des manuels universitaires. De temps en temps, une nouvelle voie possible ou un nouveau mutant est décrit, pour lequel de nouveaux chemins métaboliques sont proposés ; mais cela n'a rien de commun avec le grand nombre de voies métaboliques qu'il semble nécessaire d'envi-

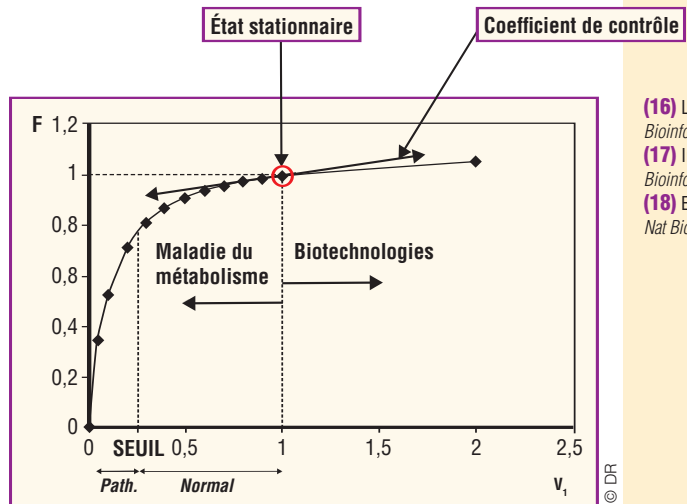


Figure 3 Coefficient de contrôle et effet de seuil

Le cercle entourant le point de coordonnées (1 ; 1) représente l'état stationnaire de départ. Dans les biotechnologies (flèche vers la droite), on cherche à augmenter le flux d'un produit d'intérêt en augmentant la vitesse d'une étape (par amplification de gène par exemple). En revanche, dans les maladies métaboliques (flèche vers la gauche), on a affaire au déficit d'une étape. Dans les deux cas, un coefficient de contrôle faible (ce qui est la plupart du temps le cas) entraîne une variation faible du flux, même pour des variations importantes de l'étape. Ce n'est que quand l'activité de l'étape passe en-dessous d'un certain seuil que le flux commence à diminuer significativement.

sager. Dans notre analyse du métabolisme énergétique mitochondrial, on peut imaginer que la majorité des 6 000 efms n'est pas empruntée pour des raisons cinétiques et qu'il n'en reste qu'un petit nombre, emprunté avec une fréquence raisonnable. Mais, qu'entendre par « un petit nombre » : 2 ou 3, une dizaine, voire une centaine ? Même une dizaine apparaîtrait comme un chiffre élevé et inattendu pour la plupart des biochimistes.

Pour répondre à cette question, plusieurs approches sont envisageables : tout d'abord, une approche théorique peut permettre d'éliminer une partie des efms. Certains d'entre eux font intervenir des stœchiométries de réactions importantes (plusieurs dizaines) (11). On peut penser (mais cela reste à démontrer) que ces combinaisons de stœchiométries risquent d'être difficiles à réaliser au niveau des molécules individuelles. Des considérations cinétiques sur certaines enzymes peuvent éliminer de nombreux efms et des enzymes peuvent avoir une influence prépondérante en ce sens.

La deuxième approche, sûrement difficile, est expérimentale ; il faudrait pouvoir suivre le devenir d'une molécule qui entre dans le réseau et savoir quel chemin elle emprunte au cours de ses transformations avant de ressortir. Si cela est possible d'un point de vue théorique, cela semble plus difficile d'un point de vue expérimental. ●

REMERCIEMENTS

Ce travail a reçu le soutien du programme STIC Santé-INSERM et de l'ACI IMPBio 2004.

- (16) Lemke N *et al.* (2004) *Bioinformatics* 20, 115-9
- (17) Imielinski M (2005) *Bioinformatics* 21, 2008-16
- (18) Edwards JS *et al.* (2001) *Nat Biotechnol* 19, 125-30