Analysis of large set of elementary flux modes : application to energetic mitochondrial metabolism

Sabine Pérès[1], Marie Beurton-Aimar[1,2] & Jean-Pierre Mazat[1]

[1]Laborartoire de Physiopathologie mitochondriale, INSERM U688 & Université Bordeaux 2
[2]LABRI UMR 6800 Université Bordeaux 1
email: sabine.peres@etud.u-bordeaux2.fr, aimar@labri.fr, jpm@u-bordeaux2.fr

Abstract

The concept of elementary flux mode is a promising approach for pathway-oriented perspective of metabolic networks. This concept defines a unique set of pathways, which represents a set of generating vectors of the solution space of feasible steady states. This set can be determined from the stoichiometric matrix of the network only. However, in large networks the combinatorial explosion of their number prevent to derive simple conclusions from their analysis. We applied a clustering method to describe the decomposition in elementary flux modes of the bioenergetic mitochondrial metabolism (Krebs cycle, β-oxidation of fatty-acids, oxidative phosphorylation, etc.). This network involves 41 enzymatic reactions, 31 metabolites. 7,250 elementary flux modes are derived from the stoichiometry matrix of the network. We clustered them by similarity and described physiological properties of most of the groups, which could in some instances be attached to specific types of mitochondria.

Introduction

With the increasing number of reconstructed reaction networks [11], there is a growing need to assess the emerging properties of these networks. One current approach comes from the concept of *elementary flux mode* (*EFM*). An *EFM* is a minimal set of enzymes that can operate at steady state with all irreversible reactions proceeding in the appropriate direction. A metabolic network with m metabolites and r reactions can be represented by a stoichiometric matrix N of m rows and r columns [7] such that :

$$N_{ij} = \begin{cases} a & \text{if } a \text{ moles of metabolite } i \text{ comes from the reaction } j \\ -a & \text{if } a \text{ moles of metabolite } i \text{ are consumed in reaction } j \\ 0 & \text{otherwise} \end{cases}$$

A vector $e \in \mathbb{R}^r$ is an *EFM* if it fulfills the following conditions [9, 10] :

1. $Ne = 0$. (*Steady state*)
2. For all indice i of an irreversible reaction $e_i \geq 0$. (*feasibility*)
3. Let $supp(v) = \{i \in \mathbb{N} : v_i \neq 0\}$. For all e' in the set of all elementary flux modes of N, $supp(e') \subseteq supp(e) \Rightarrow \exists \alpha \in \mathbb{R}$ such that $e' = \alpha e$. (*minimality*)

The concept of elementary fux modes provides a mathematical tool to define all metabolic pathways that are feasible in a given metabolic network.

Most applications of *EFMs* avoid to analyze complexe networks due to the combinatorial explosion in the number of *EFMs* [5, 2]. In these conditions, it is difficult to give an interpretation of their biological meaning. For this reason we developed a classification of a great number of *EFMs* in few clusters with biological relevance. We applied this method to the decomposition of mitochondrial metabolism in 7,250 *EFMs* .

1 Classification of *EFMs*

Our objective is to make an unsupervised classification (as clustering) of the *EFMs* set because the label of the clusters and their number are *a priori* unknown. Our major concern becomes to "reveal" the organization of *EFMs* into clusters with biological meanings. In order to organize these *EFMs* into clusters, we need to define a clustering criterion based on features selection. This is an important step, because the process of assigning *EFM* to clusters may lead to very different results, depending on the specific criterion used. This lead us to discuss an *EFM* representation and a clustering criterion based on a proximity measure to quantifies how "similar" or "dissimilar" two feature vectors are.

1.1 Feature selection : representation of the *EFM*

Our feature selection is the presence or the absence of a reaction in an *EFM* , without taking into account the stoichiometry. Thus an *EFM* can be represented as a binary vector in the order of the reactions $v \in \mathbb{R}^r$ where r is the number of reactions such that

$$\forall 1 \leq j \leq r,\ v_j = \begin{cases} 1 & \text{if the reaction } j \text{ is present in the } \textit{EFM} \\ 0 & \text{otherwise} \end{cases}$$

1.2 Proximity notion

The similarity between *EFMs* can be seen as the number of reactions they have in common (i.e the number of places where two vectors have 1 in common). Let us consider the well-known Euclidean distance, d :

$$d(x, y) = \sqrt{\sum_{i=1}^{r} (x_i - y_i)^2}$$

where $x,\ y \in \mathbb{R}^r$ and $x_i,\ y_i$ are the i^{th} coordinates of x and y respectively. This is a dissimilarity measure. For binary vector, the square of this measure give the number of places where two vectors differ. The zero componants between vectors are taken as similarities as well as the one componants, but the clusters which contain a very small distance due the presence of zeros are not absorbed in the clusters which contain a lot of ones because they are very dissimilar. Moreover, they contain fewer reactions motifs. This kind of group are not significative and do not prevent the emerging pattern of the other group.

1.3 Clustering algorithm

The number of clusters where the set of *EFMs* can be merge are *a priori* unknown. A hierarchical clustering [3] based on agglomerative method is used. The clusters are built using the function *hclust* in the *R* software [1]. This function performs a hierarchical cluster analysis using a set of dissimilarities for the n objets being clustered. Initially, each object is assigned to its own cluster and then the algorithm proceeds iteratively, at each stage joining the two most similar clusters, continuing until there is just a single cluster. At each stage, distances between clusters are recomputed. *hclust* returns a dendrogram (figure 1) and the number of clusters can be chosen by cutting the dendrogram at an appropriate distance.

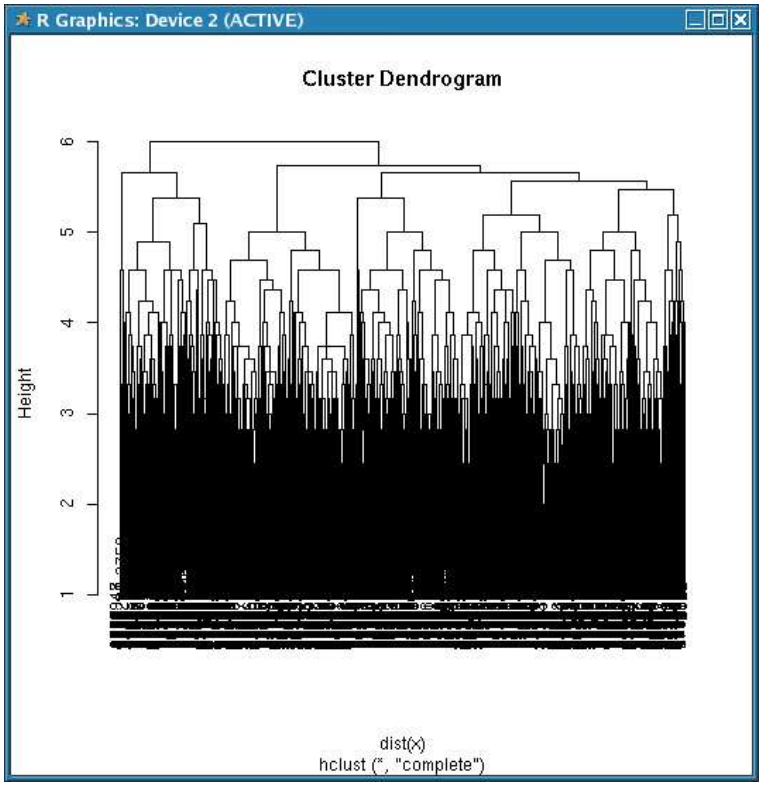

Figure 1: Dendrogram of the mitochondria clusters

2 Application to the energetic metabolism of the mitochondria

2.1 Mitochondria model

We describe the bioenergetic mitochondrial metabolism (Krebs cycle, β-oxidation of fatty-acids, oxidative phosphorylation, etc.) with 41 enzymatic reactions and carriers through the inner mitochondrial membrane and 31 internal metabolites. Using Metatool [6, 8] or FluxAnalyzer [4], 7,250 elementary flux modes are derived from the stoichiometry matrix of the network.

2.2 Results

By applying *hclust* to our set of 7,250 *EFMs* we obtain the dendrogram represented in figure 1. We decided to cut this dendrogram at a distance giving 12 groups. In order to characterize each cluster, we looked for specific reactions motifs, which could be largely represented in a given cluster. In most of the clusters, we found one motif which is present in all the *EFMs* of the cluster. The results are summarized in table 1, indicating the size of a cluster i (i.e. the number of *EFMs* it contains) and the size of the common motif present in all *EFMs* of a cluster. The last column indicates the occurrence of a given motif, common to a cluster, in the whole set of *EFMs* . As a matter of fact, if a reactions motif appears in an entire cluster, it is very probable that it will also appear frequently in other *EFM* outside the cluster used to define the motif. For instance, the second pattern m_2 appears in 39% of all the set, m_{11} appears in 28.7% and m_9 appears in 25%.

The analysis of the common motifs allows a characterization of each class :

clusters	size of the cluster	size of the commmon pattern	General occurence (%)
1	1272	0	0
2	734	3	39
3	171	0	0
4	861	0	0
5	130	3	9
6	789	7	11
7	1039	1	69
8	143	5	6
9	1713	7	25
10	299	8	4.2
11	57	4	28.7
12	42	3	2.5

Table 1: Pattern and size of the clusters

- Class 2 corresponds to the synthesis (or the consumption) of ketone bodies.
- Class 5 corresponds to the mitochondrial metabolism linked to the urea cycle in hepatocytes.
- Class 6 corresponds to the mitochondrial metabolism linked to the synthesis (or the consumption) of ketone bodies and the respiration on fatty acids compulsory linked to FADH2 entry in respiratory chain.
- Class 7 corresponds to all other elementary modes involving FADH2 utilization by the respiratory chain.
- Class 8 corresponds to the mitochondrial metabolism linked to the urea cycle and part of the Krebs cycle.
- Class 9 correspond to the respiration on fatty acids with part of Krebs cycle.
- Class 10 corresponds to an entry with pyruvate, part of the Krebs cycle and synthesis of ketone bodies.
- Class 11 links the synthesis of FADH2 in the Krebs cycle to its utilization in respiratory chain.
- Class 12 links the respiration on glutamate to ATP synthesis (or consumption ?).

Conclusion

We have presented a method, based on classical clustering techniques, to classify, with biological meaning, the great number of *EFMs* obtained in the study of metabolic networks. An important step in our analysis is the distance at which the dendrogram is cut. Cutting at too high value will give a too low number of class with no longer biological signification. Cutting at a too low distance value will introduce a too great number of non-significant differences and the biological meaning will again be lost. In our analysis, a slightly greater number of classes could evidence common motifs in the sub-classes of the classes devoid of common motif at step 12. We are analyzing other cutting distances leading to a number of

cluster between 12 and 18, the motifs involved in each cases, the way they appear when the clusters are split.

References

[1] *http://www.r-project.org/.*

[2] Julien Gagneur, David B. Jackson, and Georg Casari. Hierarchical analysis of dependency in metabolic networks. *Bioinformatics*, 19(8):1027–1034, 2003.

[3] S.C. Johnson. Hierarchical clustering schemes. *Psychometrika*, 2:241–254, 1967.

[4] J. Klamt, S.and Stelling, M. Ginkel, and E.D. Gilles. Fluxanalyzer : exploring structure, pathways and flux distributions in metabolic networks on interactive flux maps. *Bioinformatics*, 19(2):261–269, 2003.

[5] S. Klamt and J. Stelling. Combinatorial complexity of pathway anaysis in metabolic networks. *Mol Bio Rep*, 29:233–236, 2002.

[6] T. Pfeiffer, I. Sanchez-Valdenebro, J.C. Nuno, F. Montero, and S. Schuster. Metatool: for studying metabolic networks. *Bioinformatics*, 1999.

[7] C. Reder. Metabolic control theory : a structural approach. *J. theor. Biol.*, 1988.

[8] S. Schuster, T. Dandekar, and D.A. Fell. Description of the algorithm for computing elementary flux modes. http://www.bioinf.mdc-berlin.de/projects/metabolic/metatool/.

[9] S. Schuster and C. Hilgetag. On elementary flux modes in biochemical reaction systems at steady state. *Journal of Biological Systems*, 1994.

[10] S. Schuster, J.H. Hilgetag, C.and Woods, and D.A. Fell. Reaction routes in biochemical reaction systems : Algebraic properties, validated calculation procedure and example from nucleotide metabolism. *J. Math. Biol*, 2002.

[11] S. Schuster, T. Pfeiffer, F. Moldenhauer, I. Koch, and T. Dandekar. Exploring the pathway structure of metabolism : decomposition into subnetworks ans application to mycoplasma pneumoniae. *Bioinformatics*, 18(2):351–361, 2002.