**Metabolic network properties help assign weights to elementary modes to understand physiological flux distributions**

Qingzhao Wang*†, Yudi Yang†, Hongwu Ma and Xueming Zhao

Metabolic Engineering Laboratory, Department of Biochemical Engineering, School of Chemical Engineering and Technology, Tianjin University, Tianjin, 300072, People’s Republic of China

Received on October 31, 2006; revised on February 21, 2007; accepted on February 24, 2007

Advance Access publication March 6, 2007

Associate Editor: Alfonso Valencia

---

**ABSTRACT**

**Motivation:** Elementary modes (EMs) analysis has been well established. The existing methodologies for assigning weights to EMs cannot be directly applied for large-scale metabolic networks, since the tremendous number of modes would make the computation a time-consuming or even an impossible mission. Therefore, developing more efficient methods to deal with large set of EMs is urgent.

**Result:** We develop a method to evaluate the performance of employing a subset of the elementary modes to reconstruct a real flux distribution by using the relative error between the real flux vector and the reconstructed one as an indicator. We have found a power function relationship between the decrease of relative error and the increase of the number of selecting EMs, and a logarithmic relationship between the decreases of the number of non-zero weighted EMs and that of the number of selecting EMs. Our discoveries show that it is possible to reconstruct given flux distribution by a selected subset of EMs from a large metabolic network and furthermore, they help us identify the ‘governing modes’ to represent the cellular metabolism for such a condition.

**Contact:** diana_kingson@yahoo.com.cn
(or) Wangqingzhao@eyou.com

**Supplementary information:** Supplementary data are available at Bioinformatics online.

---

**1 INTRODUCTION**

Metabolic network analysis is one of the research focuses of systems biology. In recent years, a growing number of genome-scale metabolic networks of different species have been reconstructed with the aid of genome sequencing and high-throughput technologies (Covert et al., 2004), offering us a great opportunity to study them in an unprecedented manner, and acquire new knowledge about life science (Edward et al., 2000). Two aspects of metabolic networks-network topology and stoichiometry are what current researchers are most interested in, and both studies had revealed significant information. The study of network topology by the means of graph theory indicates that the metabolic network, organizing in a modular, hierarchical manner (Ma et al., 2004), resembles a small-world network.

*To whom correspondence should be addressed.
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

The research of the stoichiometric matrix of metabolic network has generated a series of powerful methodologies such as metabolic flux analysis (MFA), metabolic control analysis (MCA), flux balance analysis (FBA), etc. Considering both topological and stoichiometric characteristics of metabolic networks, metabolic pathway analysis (Schilling et al., 1999) may provide a more insightful and comprehensive means to study cellular metabolism than the above methodologies.

Recently, two related approaches for metabolic pathway analysis, elementary modes (EMs) (Schuster et al., 1994) and extreme pathways (ExPas) (Schilling et al., 2000), have demonstrated their power in studying the properties of metabolic networks (Klamt and Stelling, 2003; Papin, 2004). The robustness of the metabolic network, the potential of the species to convert a desired product, and even the gene regulation can be predicted by EMs analysis (Stelling et al., 2002). Regulatory structures for metabolic network of human red blood cell had been studied by ExPas analysis (Barrett et al., 2006). In addition to these applications, metabolic pathway analysis has been utilized to find the clue for strain optimization (Carlson et al., 2002). Since every steady-state flux distribution can be expressed as a non-negative linear combination of EMs (ExPas), understanding how probably and to what extent every mode devote to the real flux distribution would shed light on the complex cellular metabolism. For such a purpose, using EMs (ExPas) to reconstruct actual physiological flux distribution is the first and indispensable step.

To date, several means had been published to use EMs (ExPas) to reconstruct flux distributions of real metabolic networks. One way is to seek the minimal norm of weight vector by solving a quadratic programming problem (Schwartz and Kanehsia, 2005). The biological sense of this approach is apparent in that this algorithm finds the modes that are identical to the real flux distribution pattern. An earlier research used the Moore-Penrose generalized inverse of E (E is an \( n \times m \) matrix, where \( m \) denotes the number of elementary modes, and \( n \) denotes the number of reactions of the metabolic network) to assign weights to each mode of the metabolic network (Poolman et al., 2004). A common defect of the two methods is that their application is confined to analyzing small metabolic networks whose limited number of EMs (ExPas) would not pose a threat to a normal PC’s computation ability. However, in most situations, larger-scale metabolic networks considering not only...
the central metabolism but also the synthesis of precursors, the excretion of byproducts and (or) even the balance for co-enzymes are adopted to acquire a more systematic and precise information of cellular metabolism. The resulting huge numbers (from ten thousands to millions) of EMs (ExPas) of these systems (Gagneur and Klamt, 2004) make directly using these computation procedures impossible, since handling such a large number of variables simultaneously is currently beyond the ability of both software and hardware. Besides, to our knowledge, the ‘optimal’ solution of the weights acquired by the previous methods still needs more tests and improvements for interpreting the cells’ complex metabolic behaviors.

Eigenpathway, a definition from the singular value decomposition (SVD) of extreme pathway matrix, also had been used to reconstruct flux distribution (Price et al., 2003). For different patterns of flux distributions of the same metabolic network, the only difference of the reconstruction results is the coefficients of the eigenpathway vectors and the numbers of eigenpathway used for the calculation. Compared with the reconstruction by ExPas the effects of gene manipulation and (or) transcriptional regulation (Covert and Palsson, 2003) cannot be easily evaluated in terms of eigenpathways, since the change of extreme pathway matrix may eventually alter eigenpathways. This methodology is therefore, short of biological sense compared with its mathematical convenience.

Instead of seeking for an optimized weights vector, we use an original approach to resolve the problem of assigning weights for a large set of EMs to reconstruct real flux distributions. First, we calculate the set of EMs of a revised Escherichia coli metabolic network (Stelling et al., 2002). Then, we devise a quadratic program to explore the possibility and performance of using a subset of the EMs to reconstruct flux distributions. We discover that it is possible to use a part of EMs set to well represent physiological flux distributions, and the number of non-zero weighted EMs calculated to reconstruct flux distribution increases logarithmically with the increasing scale of EMs subset. Further analysis shows that there do exist a special subset of EMs that receive non-zero weights with a more frequent rate than any other ones do, and the number of EMs belonging to this special subset is much smaller than the total number of EMs. The EMs belonging to this special subset can help us understand physiological flux distributions.

## 2 METHODS

### 2.1 Models and fluxes data

A revised (the acetate uptake reaction was replaced by the succinate excretion reaction) E.coli central metabolic network and a previously constructed central metabolic network of purple non-sulfur bacteria were used as our models (Stelling et al., 2002). Flux Analyzer 6.0 (Klamt et al., 2003) was used to calculate all the EMs of the two networks. The metabolic network of E.coli had 89 metabolites, 110 reactions and a total of 711 984 EMs. The metabolic network of purple non-sulfur bacteria had 76 metabolites, 87 reactions and a total of 149 835 EMs. The flux distribution from anaerobic growth of E.coli (Schmidt et al., 1999) was used for the study of \( RE(k) \) (average relative error) and \( N(k) \) (non-zero weight), and for the subsequent analysis to seek the dominating modes. Different patterns of flux distributions within the constraints of the two metabolic networks were used to evaluate the universality of the changing styles of \( RE(k) \) and \( N(k) \) in terms of \( k \) (\( k \) denotes the number of selected EMs).

### 2.2 Calculation for \( RE(k) \) and \( N(k) \)

The performance of using subset of EMs to reconstruct flux distribution was first evaluated. A total of \( k \) EMs were randomly selected and used for reconstruction. Since all the EMs have equal chance to be selected, for each \( k \), the reconstruction procedure was repeated \( n \) times to guarantee that all the EMs can be selected and concerned for flux reconstruction. \( RE_i(k) \) denotes the \( i \)th relative error between the reconstructed flux vector \( v_i(k) \) (where \( k \) denotes the number of EMs being used, \( i \) means \( i \)th reconstruction result) and the target vector \( v \) from the \( i \)th reconstruction result. The \( i \)th reconstructed flux vector \( v_i(k) \) was calculated by solving the following non-negative constraints quadratic program(specifically, the function is \( lqpnonneg \), which is provided by MATLAB):

\[
\text{Minimize} \quad ||P_i(k)^T x_i(k) - v ||_2
\]

s.t. \( x_i(k) \geq 0 \)

Where \( x_i(k) \) was the calculated weights for \( k \) modes, and \( P_i(k) \) was the matrix whose columns were composed of \( k \) modes. The \( i \)th relative error \( RE_i(k) \) was given by:

\[
RE_i(k) = \frac{||v_i(k) - v||_2}{||v||_2}
\]

\( N_i(k) \) denotes the number of non-zero weights for the \( i \)th result when \( k \) EMs were selected to reconstruct the flux distribution. According to different number of \( k \), \( n \) ranged from 1000 to 10 000 and the average relative error \( RE(k) \) and number of non-zero weights \( N(k) \) were calculated.

### 2.3 Identification for the governing modes

Again, the steady-state flux distribution of E.coli for anaerobic growth was used to hunt for the governing modes. The same quadratic program was adopted, and \( k \) and \( n \) were chosen to be 5000 and 2000, respectively. All ten millions of the calculated weights and the corresponding indices of the selected modes were recorded. The elements of the 2000 weights vectors helped divide all the EMs into two subsets \( Z \) and \( NZ \). The corresponding modes that acquire a zero weight at least one time belong to \( Z \), and these that always acquire non-zero weights belong to \( NZ \). The numbers of elements for both subsets, \( N_Z \) and \( N_{NZ} \), were counted by utilizing the corresponding relationship between the indices and the weights. The number of non-zero elements of the weights vectors was calculated, and \( W_Z \) and \( W_{NZ} \) were the numbers of the non-zero weights belonging to \( Z \) and \( NZ \), respectively. Therefore, when used for reconstructing the physiological flux distribution, the average frequencies to receive a non-zero weight for the modes from each subset were given by:

\[
F_Z = \frac{W_Z}{N_Z}
\]

\[
F_{NZ} = \frac{W_{NZ}}{N_{NZ}}
\]

## 3 RESULTS

### 3.1 Randomly selecting EMs to reconstruct real flux distribution

We randomly selected a series of fixed numbers of EMs from a total number of 711 984 modes of the E.coli central metabolic network, and used a quadratic program to assign weights to the
selected EMs to reconstruct the actual physiological flux distribution (see Methods section for details). We found a negative power function relationship between the decrease of average relative error $\overline{RE}(k)$ (where $k$ denotes the number of selected modes) and the increase of the numbers of selected modes $k$ (Fig. 1). In order to exclude the possibilities of our results coming from the particular steady state being used, we further test more flux distributions and once again the same rules have been observed.

The relationship between $\overline{RE}(k)$ and $k$ implies us that for randomly selecting EMs to reconstruct flux distributions, there exists a boundary number of EMs by which the accuracy would not increase effectively when that number is exceeded. Another worth noting finding is the logarithmic relationship between the increase of the numbers of average non-zero weights $\overline{N}(k)$ and the increase of $k$ (Fig. 2). According to Figure 2, if the entire EMs of E.coli central metabolic network could be used to reconstruct the same flux distribution, the number of non-zero weights would be no more than 16. We performed the same calculation procedures for different patterns of flux distributions of E.coli, and for those of purple non-sulfur bacteria to evaluate the universality of these rules. Despite differences between parameters obtained from linear fitting, all the results showed the same tendencies as above (data not shown).

It is a rational idea that using a larger subset of the total EMs would receive a better result for reconstructing flux distributions, since there is a greater possibility that the ‘more proper’ modes could be found within the larger subset. However, the power function relationship between $\overline{RE}(k)$ and $k$, and the logarithmic relationship between $\overline{N}(k)$ and $k$ cannot be simply explained by probability. As mentioned before, metabolic networks have a series of properties. They have a few so-called ‘hub metabolites’ (Ma and Zeng, 2003), a power law distribution of connection degree among the metabolite nodes, a log normal distribution of the flux distributions (Sariyar et al., 2006), and so forth, we think that our discoveries are also determined by the topological and stoichiometric properties of metabolic networks. In order to prove that we have used some counter conditions such as uniform distributions and normal distributions which are not within the constraints of metabolic networks to test this hypothesis (Supplementary Material I). These distributions do not observe such relationships.

### 3.2 Identifying the ‘governing’ modes

Former algorithms for assigning weights to EMs to reconstruct flux vector try to seek a special solution. However, a single weight vector of EMs cannot completely reveal the characteristics of a given phenotype, since at most conditions there are infinite combinations of modes to reconstruct the flux distribution. Due to the complexity and redundancy of cellular metabolism, the biological sense behind these algorithms is therefore not the proof of the accuracy and efficiency of these algorithms. Instead of searching an optimal solution of weight vector, we identified a subset of EMs that would be more easier to receive non-zero weights from the quadratic program than the other modes. Previous study about the relationship between $\overline{RE}(k)$ and $k$ helped us select a proper $k$ for subsequent calculations, since although increasing $k$ would improve the accuracy; it would lengthen the computation time as well (Table 1).

The frequency for the modes from the subset NZ to reconstruct the given flux distribution is ~400 times higher than that of the modes from the subset Z. Another important finding is that NZ has an extremely small size compared with that of the set of E.coli central metabolic network EMs, which makes it more easier to handle than the original one. We used the modes from the subset NZ to represent current flux distribution to evaluate its effectiveness, and the result showed a perfect fitting (the relative error is under $10^{-4}$).

**Fig. 1.** The distribution of $\overline{RE}$ for $k$. The circle denotes the distribution of $\overline{RE}$ for $k$, and the asterisk represents a log-log plot for the distribution. The solid line is a linear fitting, and the relative coefficient is $-0.9993$. A non-negative constraints quadratic program was used to fit the anaerobic growth flux distribution of E.coli. $\overline{RE}$ was calculated from 2000 times repeated simulations.

**Fig. 2.** The distribution of $\overline{N}$ for $k$. The circle denotes the distribution of $\overline{N}$ for $k$, and the asterisk represents a lin-log plot for the distribution. The solid line is a linear fitting, and the relative coefficient is 0.9962.
metabolism than the previous ideas. we think is more flexible and systemic to study cellular algorithms when dealing with large set of EMs (ExPas).

Second, it helps seek a set of proper modes rather than a

an algorithm to identify a subset of EMs whose elements
distinguish our method. First of all, it bypasses the demanding
modes appear much more frequently in reconstructing flux distribution. The relationship between

of assigning weights to a large set of EMs to reconstruct real

5 CONCLUSION

For model organisms whose relationship between mRNA abundance and DNA binding has been well studied such as E.coli, Saccharomyces cerevisiae and so on (Herrgard et al., 2004), transcriptional regulatory rules can be used to discard infeasible EMs for a given phenotype (Covert et al., 2003). However, for most organisms, such constraints are not easy to acquire compared with the topology and stoichiometry of their metabolic networks. Although, for some cases, special growth conditions can be used to dispose those EMs with improper substrate uptake or by-products excretion reactions, the original huge set of EMs can seldom be reduced to a handy scale for further analysis. In this work, we used only the information about network topology and stoichiometry on purpose to show how our method can lead to the identification of the ‘governing modes’ from the enormous set of EMs.

Our methods share some similarity with \( \alpha \)-spectrum, which defines a weight range for each extreme pathway to reconstruct a given flux distribution (Wiback et al., 2003). The method of \( \alpha \)-spectrum considers the weight ranges of all the ExPas when reconstructing physiological flux distribution. This method has been used to understand changing metabolisms brought by environment and (or) regulation. However, based on our study, we think that the number of non-zero weights, which derive from flux reconstruction calculation, is extremely limited, and it is therefore not necessary to consider all the modes. Especially, for a large set of ExPas with an identical scale of our model, the computation for \( \alpha \)-spectrum may become time consuming or even impossible. Our method, on the contrary, can be easily performed without such a constraint. Therefore, it is especially a useful tool for dealing with large-scale metabolic networks.

4 DISCUSSION

\[ F_Z = \frac{N_{FZ}}{F} \]

\( F_Z \) is the average frequency of the modes belonging to the set \( Z \) to receive a non-zero weight for all the 2000 simulations, and \( F_{NZ} \) is that of the modes belonging to the set \( NZ \).

\[ \frac{F_{NZ}}{F} = 403.20 \]

Table 1. Statistical result for the repeated simulations

\[ N_{Z} = 21769 \quad N_{NZ} = 1048 \]

\[ W_{Z} = 711899 \quad W_{NZ} = 85 \]

\[ F_{Z} = 0.031 \quad F_{NZ} = 12.329 \]

\( F_Z \) is the average frequency of the modes belonging to the set \( Z \) to receive a non-zero weight for all the 2000 simulations, and \( F_{NZ} \) is that of the modes belonging to the set \( NZ \).

\[ \text{Conflict of Interest: none declared.} \]

REFERENCES


ACKNOWLEDGEMENTS

This work is financially supported by the National Natural Science Foundation of China (NSFC-20536040), the State Key Development Program for Basic Research of China (No. 2003CB716003 and 2007CB707802) and the Program of Introducing Talents of Discipline to Universities (No. B06006).