Chinese Science Bulletin

Expanded flux variability analysis on metabolic network of *Escherichia coli*

CHEN Tong[1], XIE ZhengWei[1] & OUYANG Qi[1,2†]

[1] Center for Theoretical Biology and School of Physics, Peking University, Beijing 100871, China;
[2] Department of Physics, Hong Kong Baptist University, Kowloon Tong, Hong Kong, China

Flux balance analysis, based on the mass conservation law in a cellular organism, has been extensively employed to study the interplay between structures and functions of cellular metabolic networks. Consequently, the phenotypes of the metabolism can be well elucidated. In this paper, we introduce the Expanded Flux Variability Analysis (EFVA) to characterize the intrinsic nature of metabolic reactions, such as flexibility, modularity and essentiality, by exploring the trend of the range, the maximum and the minimum flux of reactions. We took the metabolic network of *Escherichia coli* as an example and analyzed the variability of reaction fluxes under different growth rate constraints. The average variability of all reactions decreases dramatically when the growth rate increases. Consider the noise effect on the metabolic system, we thus argue that the microorganism may practically grow under a suboptimal state. Besides, under the EFVA framework, the reactions are easily to be grouped into catabolic and anabolic groups. And the anabolic groups can be further assigned to specific biomass constitute. We also discovered the growth rate dependent essentiality of reactions.

metabolic network, flux balance analysis, modularity, essentiality

The complete sequences of the genome of several model organisms combined with stoichiometric information of biochemical reactions allow biophysics researchers to reconstruct the metabolic network at the genome-scale level[1,2]. However, the lack of kinetic information on *in vivo* biochemical reactions and concentrations of metabolites hinders the development of dynamic models. To overcome these difficulties, the constraint-based and parameter-free models such as flux balance analysis (FBA)[3] have been well developed to interpret the properties of metabolic networks and predict cellular behaviors. The metabolic flux distributions, growth rate or by-product production provided by FBA are consistent with experiments[4]. The applications of these approaches are well documented in literature[5,6].

Under the framework of FBA, several novel methods have been developed to study the properties of the networks. Firstly, the flexibility and robustness of metabolic networks have been studied through elementary modes analysis[7] and multiple equivalent phenotypic state analysis[8]. The flexibility or redundancy is an important property because it results in robustness to the network that may face enzyme loses for some reasons. The flexibility can be represented by the number of elementary modes or that of vertexes under specific conditions[7,8]. Secondly, the modules of the network have been obtained from flux coupling analysis[9] or minimal cut sets[10] from the topological perspective. It enables us to reveal the global organization and functional units of the metabolic network. Thirdly, the essentiality of reactions can be elucidated by FBA[11]. If the maximum biomass flux becomes zero after deletion of one reaction, this reaction is essential under current conditions. High

Received December 28, 2008; accepted February 19, 2009; published online May 14, 2009
doi: 10.1007/s11434-009-0341-x
†Corresponding author (email: qi@pku.edu.cn)
Supported by the National Natural Science Foundation of China (Grant No. 10721403), National Basic Research Program of China (Grant Nos. 2006CB910706, 2007CB814800, 2009CB918500), Chun-Tsung endowment at Peking University and National Fund for Fostering Talents of Basic Science (Grant No. J0630311)

Citation: Chen T, Xie Z W, Ouyang Q. Expanded flux variability analysis on metabolic network of *Escherichia coli*. Chinese Sci Bull, 2009, 54: 2610−2619, doi: 10.1007/s11434-009-0341-x

consistence with the experiments has been obtained[11]. Though the above topics are different and independent, we find out they can all be well elucidated by our newly developed method, named expanded flux variability analysis (EFVA).

EFVA is an expansion of the flux variability analysis (FVA), which calculates the flux variability under predefined optimal states based on linear programming[12]. Compared with the mixed integer linear programming (MILP) algorithm, FVA is a computationally affordable method to discover multiple states of genome-scale models. In this work, we expand the flux variability method by tracing the trend of the variable range, the maximum and the minimum flux values of each biochemical reaction under the constraint of fixed growth rate, ranging from zero to optimal growth rate. We use the metabolic reaction network of *E. coli*. (iJE660a), reconstructed by Edwards and Palsson[11] as model organism. We find that the optimal growth rate is at an expense of decreasing flexibility and robustness, indicating that the microorganism may practically grow at a suboptimal rate. Our method can also be applied in the modular analysis of the network. All the reactions in the metabolic network can be separated into two groups according to their distinct behaviors in maximum possible value of the flux. One group supplies rudimentary disposal of nutrients and energy; the other is in charge of assembling the biomass components. The latter group can be further divided into subsets that synthesize each biomass component for cellular growth. We then interpret the underlying mechanism of biochemical reactions' essentiality based on the subsets and define the growth rate dependent essentiality.

1 Materials and methods

The subject of our study is *E. coli* iJE660a, reconstructed by Edwards and Palsson[11]. All isoenzymes catalyzing one reaction are merged to a single flux. This model finally includes 629 unique biochemical reactions and accounts for 660 genes. The biosynthetic demands, i.e. the production of amino acids, nucleotides, phospholipid together with energy and redox potential for growth are determined from the biomass composition and expressed as the growth flux, using the biosynthetic precursors (X_i) in experimental determined ratios (d_i):

$$\sum_i^m d_i \cdot X_i \xrightarrow{\text{growth flux}} \text{biomass}. \quad (1)$$

The primary constraints of FBA include the pseudo-steady-state assumption and the bounds for particular fluxes. The former constraint requires the balance of all inner metabolites, i.e. the sum of all influx equals to the sum of outflow; and it is based on the mass conservation law. The latter requires that the values of fluxes are restricted in an interval, because of the enzymatic capacity, the thermodynamical feasibility and the nutritional availability. These constraints of networks can be written as linear equations and inequalities. The feasible solution space is thus been defined as a high-dimensional polyhedron. The information on the function and structure of networks, i.e. viability, fitness, robustness, etc., can be obtained by analyzing this feasible solution space.

The assumption of the steady state of concentrations of the intracellular metabolites requires a mass balance of all the internal metabolites. Thus, the following equation must be satisfied:

$$\sum_{i=1}^{n} S_{ji} v_i = 0, \ \forall j \in \mathbb{N}, \quad (2)$$

where S_{ji} is an element of the $m \times n$ stoichiometric matrix S, m and n are the number of metabolites and the number of reactions respectively; v_i is an element in the net reactions rates vector v. Usually S is a singular matrix[13,14], so that eq. (1) has multiple solutions. These solutions are subjected to two constraints: the nutrition (i.e. carbon, oxygen, etc.) uptake constraint is specified by

$$v_i^{\text{uptake}} \leqslant v_i^{\text{uptake max}}, \ \forall i \in M_{\text{transport}}, \quad (3)$$

and the thermodynamic constraint is specified by

$$v_i \geqslant 0, \ \forall i \in M. \quad (4)$$

The traditional FBA is seeking the maximum growth rate in eq. (1) under those constraints.

In our expanded flux variability analysis, the growth rate is treated as a constraint to the network. This constraint is a fixed value v_{object} defined by

$$v_{\text{object}} = r \times v_{\text{optimal}}, \ r \in [0, 1], \quad (5)$$

where v_{optimal} is derived from the classical FBA; r is the relative growth rate ratio, indicating the difference between traditional and expanded flux variability analysis. If the value of r equals one, the expand approach will be identical to the traditional one. The expanded flux variability analysis can be mathematically interpreted as follows: For given i, maximize and minimize v_i subject to eqs. (2)—(5). The boundaries of all the reactions are

obtained. The range of a flux i (δ_i) is calculated in eq. (6) as the difference between the maximum and the minimum value:

$$\delta_i = v_{\max,i} - v_{\min,i}. \quad (6)$$

In order to identify and compare the effects of different levels of constraint on the variable ranges of reactions, we apply the reference range of the reaction. We derive the reference range by removing the constraint in eq. (5). That is, the $V_{\max,i}$ ($V_{\min,i}$) are deduced by maximizing (minimizing) v_i subject to the constraint in eqs. (2)—(4). Then the range can be normalized into the interval from 0 to 1 by

$$\Delta_i = \frac{\delta_i}{V_{\max,i} - V_{\min,i}}. \quad (7)$$

Under a given nutrient condition, some reactions always carry zero flux[9]. The variable ranges of this set of reactions are zero. Hence, we can first identify them using flux variability analysis to reduce the computational complexity.

The solutions of the liner programming (LP) problems (using GNU Linear Programming Kit) outlined above for each reaction i in the network determine the upper and lower bounds of every reaction flux. However, due to the coupling between reactions, this approach provides the bounds of all feasible flux distribution rather than the exact shape of allowable solution space. A schematic description of feasible set is presented in supplementary Figure 2. Note that in our expanded model, this growth constraint could be not only set as the biomass composition but also arranged as ATP composition or synthesization of single metabolite, which is usually used to guide metabolic engineering.

2 Results and discussion

2.1 *E. coli* practically grows at a suboptimal rate

The feasible solution space of a metabolic network, in which the fluxes of metabolites meet all the physiological constraints, includes all allowable metabolic flux distributions. It defines the metabolic capacity of the system[15] and determines the robustness of the network. Recent advancements have shown that the feasible solution space is biologically meaningful[16–18]: high variations in gene expression levels are detected between genetically identical cells taken from the same culture. Additionally, Fong et al.[12,19] have shown that multiple biologically meaningful flux states can be active in different conditions. All these evidences encourage us to study the variability of the reactions. Specifically, we study the variation of the feasible solution space as a function of the growth rate constraint.

The volume of the feasible solution space is a natural order parameter in this study. However, in practice it is very difficult to calculate the volume of the feasible solution space, which is a high-dimensional irregular polyhedron. Previously proposed Monte Carlo sampling method[17] is computationally intractable for genome scale networks. Herein, we try to find an order parameter to quantitatively measure the volume under the constraint of a fixed growth rate. Assume that the flux distribution of an organism evolves in the feasible solution space. If one ignores newly introduced reactions in the process and the influence of coupling between reactions, one evolutionary event is usually only related to one existing biochemical reaction, resulting in a random flux fluctuation of this reaction. Bilu et al.[16] have shown that the variable range of fluxes gives the evolutionary constraints on the enzymes activities and fluxes. We thus assume that the probability for an organism to stay in a particular growth rate is proportional to the variable range of the flux under this growth rate. In other words, the variable range of a particular reaction at a given growth rate measures the tolerance of the system encountering perturbations in this reaction. If treating all the reaction equally, one can thus use the average of variable ranges of all reactions (defined as evolutionary volume) to measure the volume of the feasible solution space at a specific growth rate. This is the order parameter we used in this study.

We calculated the evolutionary volume of the metabolic network of *E. coli*. (iJE660a) under the constraint of a fixed relative growth rate varying from 0 to 1, normalized by optimal growth rate in the minimal media. Figure 1 presents the main result of the study. It shows that the evolutionary volume decreases almost linearly with the increase of the growth rate constraint. Because a state of suboptimal growth rate corresponds to a larger solution space, the metabolism in this state is more flexible to mutational perturbations. We think that an organism should balance the advantages in growth rate and flexibility. Analytically, we can partition the system into different sub-groups according to its growth rate, and the growth rate of cell population in the ith partition can be described by the following equation:

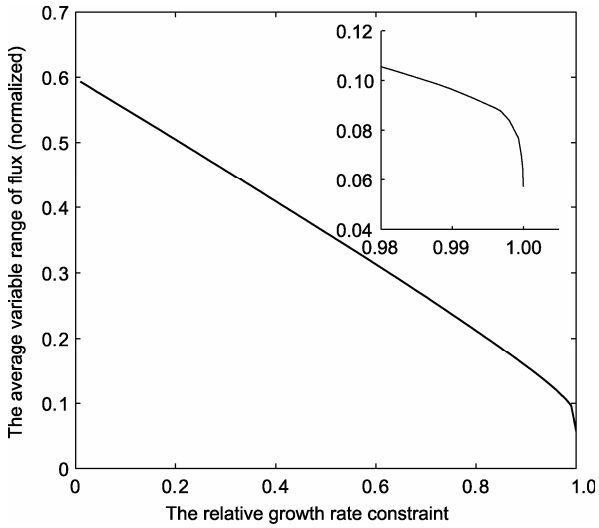

Figure 1 The variability of reactions under different growth rates in *Escherichia coli*. The average variability of reactions versus the fixed growth rate as constraint. The evolutionary volume shrinks as the growth rate increases.

$$\frac{\mathrm{d}p_i}{\mathrm{d}t} = x_i p_i + \sum_j b_{ij} p_j - \sum_j b_{ij} p_i,$$

where x_i is the growth rate of population i and b_{ij} is the transition rate of cell jump from jth sub-group to the i sub-group, which in our model is supposed to be proportional to the solution space of the ith sub-group (details in calculation of this transition matrix are analyzed and presented elsewhere). The above equation can be written as a vector form:

$$\frac{\mathrm{d}\boldsymbol{P}}{\mathrm{d}t} = M\boldsymbol{P}.$$

Due to the evolutionary competition, we find the largest eigenvalue of the equation gives the growth rate of the cell population, and the corresponding eigenvector tells us the composition of the cell population. Our calculation shows that the average growth rate of the *E. Coli* population should be about 2/3 of the maximum growth rate (details are presented else where), so that the cell may grow at a suboptimal growth rate.

This suboptimal growth phenomena have been validated by Fischer and Sauer[20] in the study of Bacillus subtilis. They found that a part of the null mutants of the species grew faster than the wild type, indicating that the metabolism of the species is regulated to a suboptimal growth rate in the wild type. We further evaluated the distribution of the growth rate obtained by Fischer et al., and found that the wild type was located at 0.76 normalized by the maximum growth rate of all mutations (refer to supplementary material for details). This observation supports our argument.

We also computed the variable range of each reaction under the constraint of a fixed relative growth rate. 400 out of 629 reactions carry non-zero flux without growth constraint. Under the optimal growth condition, the fluxes of only 37 reactions can be variable, whereas under any suboptimal growth condition the fluxes of more 60% (249) of the reactions are variable. Note that for the first approximate, we trait each reaction with equal weight, which is valid in FBA analysis but may not exactly reflect the real system: some reactions may be more important than others in other control systems.

Although the common perception is that the evolved networks operate at an optimal metabolic state, at least in their normal environment[4], this perception ignores the evolutionary approachability and stability of the system. Our results show that the network has less flexibility when operating at the optimal growth state. In fact, because of the inevitable fluctuations in enzymatic activities and genetic perturbations, there is always a trade-off between the growth rate and the robustness[20,21]. Therefore the growth rate is not the only dominate factor in the evolution process; robustness is another important aspect that should be considered.

Biological robustness is a ubiquitous property and fundamental feature of complex evolvable systems[7,22]. It is defined as the ability to maintain the performance in face of perturbations and uncertainties; and associates tightly with the flexibility[23]. This issue has been well addressed in *in vitro* evolution[24], where the thermodynamic stability and mutational robustness are critical to the functions and the evolutionary longevity of DNA molecules. In the case of the metabolic network, external perturbations and internal mutations may introduce instabilities and in-approachability of the optimal state, whereas the suboptimal ones are more stable against these fluctuations. This situation can be likened to a thermodynamic system, where the competition between enthalpy and entropy determines the fate of the system. Similarly, in the evolutionary process of *E. coli*, we suggest that there is also a trade-off between the growth rate (enthalpy) and evolutionary space (entropy). The evolutionary approachable and stable states should be located at a suboptimal growth.

2.2 The modular division of the metabolic network

We next report our investigation of the maximum value

of each flux as a function of the growth rate constraint. The result of the calculation is presented in Figure 2(a). We found out that the reactions in the metabolic network exhibited three different behaviors as the increase of the growth rate constraint. The first group of 20 reactions keep a constant maximum value because of the futile cycles in the network; the second (161 reactions) and the third (219 reactions) group of reactions have, respectively, an increasing and a decreasing maximum value as the growth rate constraint goes higher. The maximum fluxes in the second group are in higher order of magnitude than those in the third group. A detailed study revealed that the second group is the assembling group, which contributes to the synthesis of biomass components; the third group is in charge of supplying energy and the precursors of biomass components. It is composed of most reactions involved in glycolysis, TCA cycle, anaplerotic pathway and oxidative phosphorylation pathway. We define it as the preparing group.

If we use ATP consumption rate as the constraint instead of the growth rate constraint in eq. (5), the reactions in the assembling group switch to a decreasing maximum values with the increase of ATP consumption rate constraint, as shown in Figure 2(b). This switch is due to the following reasons: (i) there is a competition between the ATP yield and the growth for the carbon resource and oxygen; (ii) there is no coupling between the reactions in the assembling group and the ATP consumption reactions. Therefore a higher constraint of energy consumption rate will reduce the cellular growth rate, and thus decreases the maximum value of the coupled reactions, i.e. the reactions of the assembling biomass components. The experimental evidence for the competition between yield and growth rate has been discovered by Novak et al.[25]. If we set the biomass flux to zero and set the ATP consumption rate as the constraint, none of the reactions in the assembling group can carry flux, as shown in Figure 2(c). These results show that the assembling group is coupled with biomass composition.

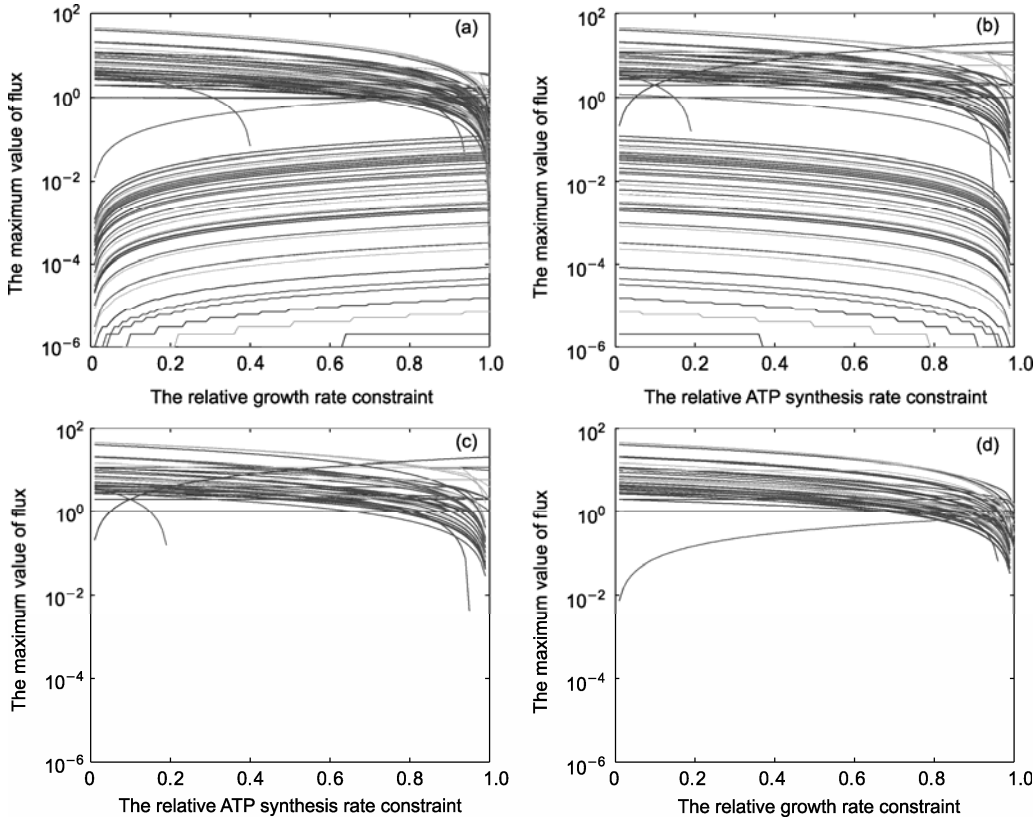

Figure 2 The maximum value of each reaction under different constraints in *Escherichia coli*. (a) The maximum feasible values of metabolic reactions as a function of the biomass composition threshold; (b) the assembling reactions group has a decreasing maximum value when setting ATP composition as the threshold and allowing cellular growth; (c) the assembling reactions group is suppressed to zero when setting ATP composition as the threshold and cellular growth to zero; (d) the maximum feasible values of the assembling reactions group all become null except reactions in isoleucine pathway when setting the composition of isoleucine as the threshold and not allowing growth. Each line represents one flux. Some reactions have identical viable range, thus lines may overlap.

With the implementation of our approach, we further tried to divide the assembling group into subsets associated with the synthesis of individual biomass component. It was implemented by tracing the behavior of the maximum value when we set the product rate of a particular biomass component as the constraint. With this approach, the biomass-coupled reaction group (i.e. the assembling group) can be broken up into subsets when the biomass reaction is replaced by independent drains of biomass components. Here, we take the procedure in isoleucine biosynthesis as an example. The behaviors of maximum value of fluxes are presented in Figure 2(d). The six overlaying increasing lines represent the maximum fluxes of six reactions, as shown in Figure 3. The combination of these reactions forms the assembling subset of isoleucine.

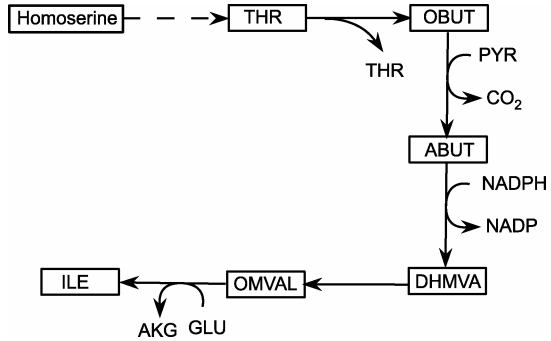

Figure 3 The isoleucine subset. The isoleucine biosynthesis pathway from threonine. For reaction details see supplementary information 4.

The same method is applied to calculating the subsets for all the biomass components (totally 47) and the assembling group is successfully divided. Figure 4 summarizes the results. This modular division agrees well with analytical results based on functional classification[11].

In the study, we observed that only less than one-tenth of the reactions in the preparing group are essential for synthesizing the biomass constitutes (see supplementary Table 3). Meanwhile, most of the reactions in the assembling group (the reactions in subsets) are essential for synthesizing the biomass constitutes. A few exceptions can be attributed to the alternative reactions in the long linear synthetic pathway. The union of all the subsets in Figure 4 includes 159 reactions, with the exception of the fake growth flux and DAMP +ATP ↔ ADP + DADP in the salvage pathway. Because the synthesis of some constitutes may share the same pathways or the same reactions, the subsets are not mutually exclusive. The composition of purine serves as a perfect example: Twelve reactions that are functionally related to the purine synthesis are simultaneously presented in the thirteen other subsets. The multiple involvement of reaction in subsets mainly results from the bow-tie structure of metabolic network[26]. Approximate 90 reactions only appear in one subset, while the uptake of ammonia flux is involved in 44 of the total 47 subsets.

Modularity of complex biological networks contributes to the robustness and flexibility of the organism[27–29]. Such modular descriptions of biochemical network function are shaping current researches in system biology[27]. Historical definition of modules as groups of reactions that have a related function is subject to intuitions of researchers, whereas our proposed methodology can be applied to dividing the metabolic network hierarchically by setting diversified constraint functions, i.e. cellular growth and synthesize of certain metabolites. Our method can include non-obvious groups of reactions that differ from the methods based on visual impression of topological property[30]. The key advantage of expanded flux variability analysis over previous approaches is that it does not need a priori assumption of dividing networks into subsystems for analysis, and its computational feasibility for the division of genome-scale metabolic network. Previous methods such as elementary mode and extreme pathway[31], which have been applied to identification of enzyme subset, require breaking the network into subsystems. They can only handle network with approximately one hundred reactions[32].

2.3 The explanation of the essentiality of reactions

We consider a reaction essential in the metabolic network if the growth of the organism will halt without this reaction[33,34]. In our framework, the range of a reaction flux for a given growth rate constraint can reveal the essentiality of the reaction. If the range contains zero, this reaction is not essential for the given constraint, because the flux can be equal to zero under this condition. On the other hand, if the range does not contain zero for a particular growth rate constraint, this reaction is essential under this condition, because the deletion of the reaction will render the network impossible to finish the given task. In this way, we can calculate the essentiality of the reactions. As shown in Table 1, our results show that the reactions in the assembling group are mostly

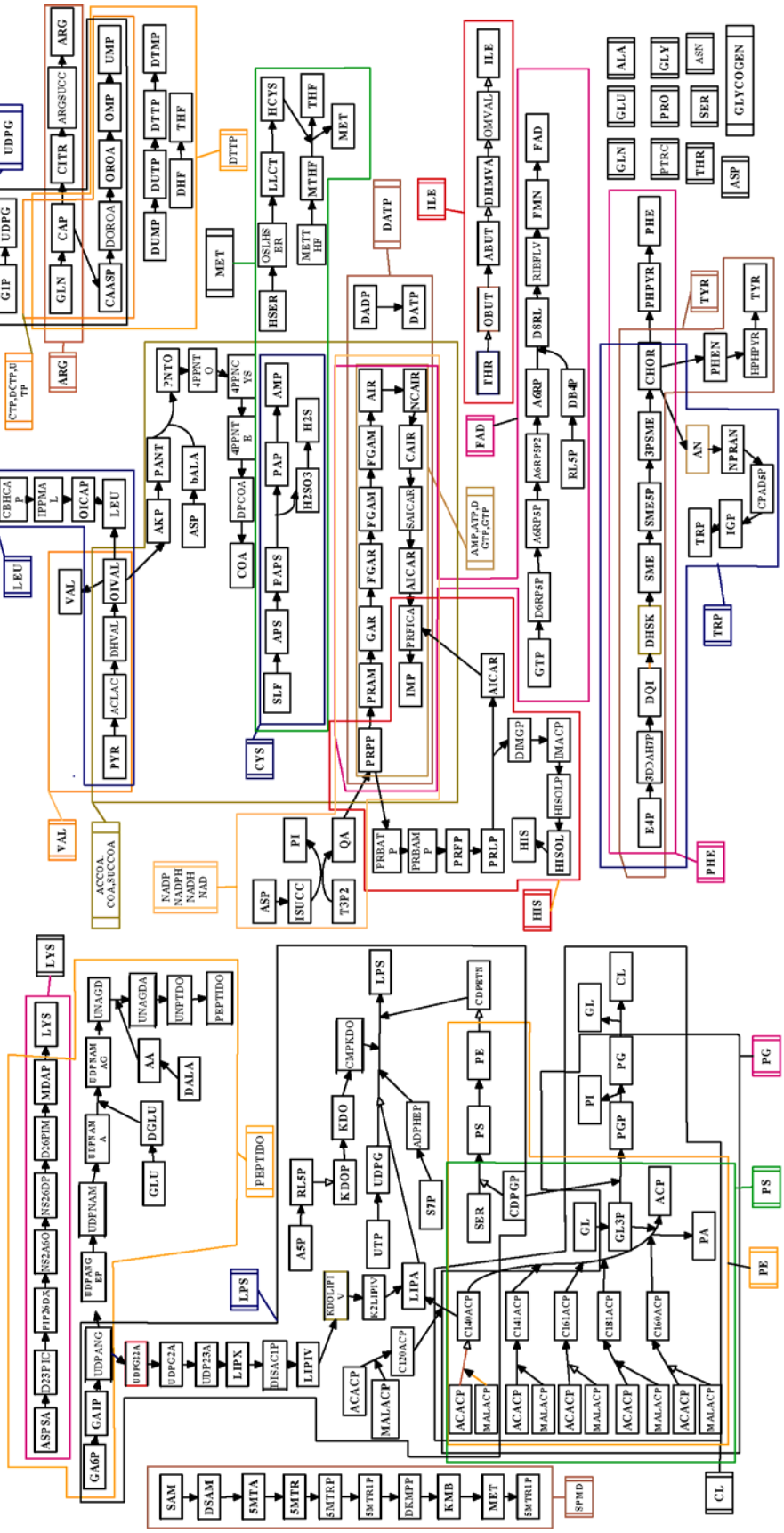

Figure 4 The detailed picture of subset division. Each frame involves the reactions in the subset for a particular biomass component. The flag to the frame denotes the name of subset. Several subsets may share the same reactions. For the sake of clarity, the ubiquitous elements in the subsets, i.e. the uptake reactions of ammonia and phosphate, are not included here.

Table 1 The explanation of essentiality: The number of reactions in each category. 70% of the optimal growth rate is used as the growth rate constraint in the calculation. The result is insensitive to the value of the growth rate constraint, as long as it is smaller than 70% of the optimal value. AS. denotes the assembling group, while PR. denotes the preparing group

Category	Essential	Nonessential
UP/UC and AS.	150	9
Not UP/UC and AS.	2	0
UP/UC and PR.	51	143
Not UP/UC and PR.	6	19

essential compared with those in the preparing group. 152 out of 161 (94%) deletions of reactions in the assembling group are lethal, compared with 57 out of 219 (23%) ones in the preparing group. The nine unessential reactions in the assembling group are unexpected because these reactions are directly coupled with biomass flux. The bypass in the synthetic pathways is attributed to the phenomenon.

The essentiality of a reaction can also change when the growth rate changes. Some genes which are responsible for the reactions in the metabolic network of *E. coli*. (iJE660a) are listed in Table 2. One observes that certain essential reactions may become nonessential when the growth rate constrain drops. This implies that we can define the growth rate as dependent essentiality. It is noteworthy that the essential threshold of oxygen uptake is 0.3, while that of carbon dioxide secretion is 0.59. This indicates that the glucose is not necessary to be oxidized to carbon dioxide under anaerobic condition; its secretion steps in only when the bacteria grow very fast.

Table 2 The listed genes are essential when the relative growth rate is greater than the essential threshold in the left column

Essential threshold	Genes
0.99	*araD*, *tktB*, *talB*, *sucAB*, *sucCD*, *ppc*, *fdhF*, *glpK*, *asnA*, *speF*, *proB*, *proA*, *purN*, *ndk*, *adk*, *deoD*, *cmk*, *gcvHTP*, *lpdA*, *pta*, *purT*, *deoB*, *pyrH*, *fabH*
0.98	*pgi*, *lpdA aceEF*, *zwf*, *pgl*, *gnd*
0.97	*cyoABCD*, *cycBC*, *cydAB*
0.96	*pfkB*, *fbaA*, *mdh*, *tpiA*
0.95	*gdhA*
0.94	*pykF*
0.92	*gpmB*, *eno*, *fumC*
0.86	*gapC1C2*, *pgk*
0.82	*nuoABEFGHIJKLMN*
0.59	carbon dioxide secretion
0.36	*atpABCDEFGHI*
0.3	oxygen uptake
0.04	*ribA*, *ribD*, *ribD*, *ribH*, *ribF*, *ribF*
0.02	*ribB*, *ribE*

There has been a long history to uncover the underlying mechanism of the essentiality of the metabolic reactions[35]. Previous studies tried to attribute the essentiality to the low degree of metabolites[11,36], especially the "UP/UC" nature of the reactions. Metabolites in the networks show a large variation in their degree of connectivity[35,36]. A metabolite was defined as "uniquely produced" or "UP" ("uniquely consumed" or "UC") if there is only one reaction in the network that produces (consumes) the metabolite[11]. The UP/UC reactions can largely explain the essentiality of the reaction because they are designated as "essential" for the balance of a particular metabolite. However, not all metabolites are indispensable for cellular growth. In Table 1, one finds that among 194 UP/UC reactions in the preparing group, 73.7% of the reactions are not essential. This is because the metabolites involved in the preparing group may not be required for cellular growth. For example, several reactions in pentose phosphate pathway are UP/UC reactions. However, the EMP pathway provides a bypass, so that these reactions are not essential for cellular growth. The central carbon metabolic network, i.e. glycolysis, TCA cycle, anaplerotic reactions, and oxidative phosphorylation reactions bears the most flexibility and redundancy of the system. The alternative pathway will buffer for the low degree reactions that seem fragile.

In contrast, in the assembling group 152 out of 161 reactions are essential. Among those, 150 reactions are UP/UC reactions (Table 1). Their essentiality can be explained by a low degree of reactions as explained by Samal et al.[33]. We can also explain the essentiality through our subset division. As shown in supplementary Table 2, almost all the reactions in the subsets are essential for the synthesis of certain biomass component. For instance, THR → OBUT + NH_3 is critical for the cellular growth, since it is involved in the synthesis of isoleucine. If a reaction is involved in several subsets, the deletion of the reaction will affect the synthesis of several biomass components. For instance, if we delete the reaction PRPP + GLN → PPI + GLU + PRAM, the biosynthesis of twelve biomass components cannot be conducted successfully.

From this analysis, we conclude that UP/UC analysis may not precisely explain the mechanism of the essentiality without a functional category analysis of the metabolites. Our approach can overcome the shortcomings of this topological study of network structure, and pro-

vide a method for the functional analysis of the metabolic system.

3 Conclusions

To summarize, we introduced the expanded flux variability analysis to study the intrinsic properties of reactions in the metabolic network under different environmental conditions. The approach was based on the solution of linear programming, thus allowing it to remain tractable for large metabolic network involving hundreds even thousands of reactions. We investigated not only the flux distributions at optimal growth condition but also the distributions at suboptimal growth rate. The variation trend of the range, the maximum and the minimum value of reactions as a function of the growth rate constraint were informative for us to obtain different characteristics of the network, such as the robustness, modular structure, and reaction essentiality. The analysis approach presented here has also been applied to analyzing other nutritional conditions in the metabolic network of *E. coli*, as well as other metabolic networks, such as *Saccharomyces cerevisiae*. Similar results have been obtained. This indicates that the expanded flux variability analysis can provide a useful framework for both modelers and experimentalists to extract biological information from metabolic reconstruction.

The authors thank M. Ni and S.Y. Wang for their valuable suggestions and critical reading of the manuscript.

1 Reed J L, Vo T D, Schilling C H, et al. An expanded genome-scale model of Escherichia coli K-12 (iJR904 (GSM/GPR). Genome Biol, 2003, 4: R54

2 Duarte N C, Herrgard M J, Palsson B O. Reconstruction and validation of saccharomyces cerevisiae iND750, a fully compart-mentalized genome-scale metabolic model. Genome Res, 2004, 14: 1298－1309

3 Edwards J S, Ibarra R U, Palsson B O. In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. Nature Biotech, 2001, 19: 125－130

4 Segre D, Vitkup D, Church G M. Analysis of optimality in natural and perturbed metabolic networks. Proc Natl Acad Sci USA, 2002, 99: 15112－15117

5 Edwards J S, Covert M, Palsson B O. Metabolic modelling of microbes: the flux-balance approach. Environ Microbiol, 2002, 4: 133－140

6 Price N D, Papin J A, Schilling C H, et al. Genome-scale microbial in silico models: the constraints-based approach. Trends Biotech, 2003, 21: 162－169

7 Stelling J, Sauer U, Szallasi Z, et al. Robustness of cellular Functions. Cell, 2004, 118: 675－685

8 Reed J L, Palsson B O. Genome-scale in silico models of *E. coli* have multiple equivalent phenotypic states: Assessment of coorelated reaction subsets that comprise network states. Genome Res, 2004, 14: 1797－1805

9 Burgard A P, Nikolaev E V, Schilling C H, et al. Flux coupling analysis of genome-scale metabolic network reconstructions. Genome Res, 2004, 14: 301－312

10 Klamt S, Gilles E D. Minimal cut sets in biochemical reaction networks. Bioinformatics, 2004, 20: 226－234

11 Edwards J S, Palsson B O. The Escherichia coli MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. Proc Natl Acad Sci USA, 2000, 297: 5528－5533

12 Mahadevan R, Schilling C H. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab Eng, 2003, 5: 264－276

13 Famili I, Foster J, Nielsen J, et al. Saccharomyces cerevisiae phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. Proc Natl Acad Sci USA, 2003, 100: 13134－13139

14 Schuster S, Fell D A, Dandekar T. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nature Biotech, 2000, 18: 326－332

15 Edwards J S, Pallson B O. Robustness analysis of the Escherichia coli metabolic network. Biotech Prog, 2000, 16: 927－939

16 Bilu Y, Shlomi T, Barkai N, et al. Conservation of expression and sequence of metabolic genes is reflected by activity across metabolic states. Plos Comput Biol, 2007, 2: 932－938

17 Wiback S, Famili I, Greenberg H, et al. Monte Carlo sampling can be used to determine the size and shape of the steady-state flux space. J Theoret Biol, 2004, 228: 437－447

18 Price N, Reed J, Papin J, et al. Network-based analysis of metabolic regulation in the human red blood cell. J Theoret Biol, 2003, 225: 185－194

19 Fong S, Joyce A, Palsson B. Parallel adaptive evolution cultures of Escherichia coli lead to convergent growth phenotypes with different gene expression states. Genome Res, 2005, 15: 1365－1372

20 Fischer E, Sauer U. Large-scale *in vivo* flux analysis shows rigidity and suboptimal performance of Bacillus subtilis metabolism. Nature Genet, 2005, 37: 636－640

21 Pfeiffer T, Schuster S, Bonhoeffer S. Cooperation and competition in the evolution of ATP-producing pathways. Science, 2002, 292: 504－507

22 Kitano H. Biological robustness. Nature Rev Genet, 2004, 5: 826－837

23 Stelling J, Klamt S, Bettenbrock K, et al. Metabolic network struc-

ture determines key aspects of fuctionality and regulation. Nature, 2002, 420: 190－193
24 Meyers L A, Lee J F, Perthwaite M, et al. The Robustness of naturally and artificially selected nucleic acid secondary structures. J Mol Evolut, 2004, 58: 681－691
25 Novak M, Pfeiffer T, Lenski R E, et al. Experimental tests for an evolutionary trade-off between growth rate and yield in *E. coli*. Am Nat, 2006, 168: 242－251
26 Caste M E, Doyle J. Bow ties, metabolism and disease. Trends Biotech, 2004, 22: 446－450
27 Papin P A, Reed J L, Palsson B O. Hierarchial thinking in network biology: the unbiased modularization of biochemical networks. Trends Biochem Sci, 2004, 29: 641－647
28 Guimera R, Amaral A N. Functional cartography of complex metabolic networks. Nature, 2005, 433: 895－900
29 Hartwell L H, Hopfield J J, Leibler S, et al. From molecular to modular cell biology. Nature, 1999, 402: C47－C52
30 Papin J A, Price N D, Wiback S J, et al. Metabolic pathways in the post-genome era. Trends Biochem Sci, 2003, 28: 250－258
31 Schuster S, Dandekar T, Fell D A. Detection of elementary flux mode in biochemical networks: A promising tool for pathway analysis and metabolic engineering. Trends Biochem Sci, 1999, 17: 53－60
32 Schilling C H, Cover M W, Famili I, et al. Genome-scale metabolic model of Helicobacter pylori. J Bacteriol, 2002, 184: 4582－4593
33 Samal A, Singh S, Giri V, et al. Low degree metabolites explain essential reactions and enhance modularity in biological networks. BMC Bioinformatics, 2006, 7: 118－127
34 Mahadevan R, Palsson B O. Properties of metabolic networks: Structure vs. function. Biophys J, 2005, 88: L7－L9
35 Jeong H, Tombor B, Albert R, et al. The large-scale organization of metabolic network. Nature, 2000, 407: 651－654
36 Jeong H, Mason S P, Barabasi A L, et al. Lethality and centrality in protein networks. Nature, 2001, 411: 41－42