ARTICLE

BIOTECHNOLOGY and BIOENGINEERING

Compensatory Ability to Null Mutation in Metabolic Networks

Da Jiang,[1,2] Shuigeng Zhou,[1,2] Yi-Ping Phoebe Chen[3,4]

[1]Shanghai Key Laboratory of Intelligent Information Processing, Fudan University, Shanghai, China; telephone: +86-21-55664298; fax: +86-21-55664298; e-mail: sgzhou@fudan.edu.cn

[2]School of Computer Science, Fudan University, Shanghai, China

[3]Faculty of Science and Technology, Deakin University, Melbourne, Victoria, Australia; telephone: +61-3-92517684; fax: +61-3-92517604; e-mail: phoebe@deakin.edu.au

[4]Australia Research Council Centre in Bioinformatics, Melbourne, Victoria, Australia

Received 12 September 2008; revision received 27 November 2008; accepted 15 December 2008

Published online 30 December 2008 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/bit.22237

ABSTRACT: Robustness is an inherent property of biological system. It is still a limited understanding of how it is accomplished at the cellular or molecular level. To this end, this article analyzes the impact degree of each reaction to others, which is defined as the number of cascading failures of following and/or forward reactions when an initial reaction is deleted. By analyzing more than 800 organism's metabolic networks, it suggests that the reactions with larger impact degrees are likely essential and the universal reactions should also be essential. Alternative metabolic pathways compensate null mutations, which represents that average impact degrees for all organisms are small. Interestingly, average impact degrees of archaea organisms are smaller than other two categories of organisms, eukayote and bacteria, indicating that archaea organisms have strong robustness to resist the various perturbations during the evolution process. The results show that scale-free feature and reaction reversibility contribute to the robustness in metabolic networks. The optimal growth temperature of organism also relates the robust structure of metabolic network.

Biotechnol. Bioeng. 2009;103: 361–369.

KEYWORDS: metabolic networks; scale-free; redundancy; robustness

Introduction

In general, robustness is the persistence of a system's characteristic behavior under perturbations or conditions of uncertainty (Stelling et al., 2004). The word robustness, when used with regard to metabolic network, refers to the organism that it is against single-gene or even multiple-gene mutations by using redundant or alternative pathways. Many knockout mutants of organisms are still able to grow, some with almost the same growth rate as the wild-type (Motter et al., 2008; Segre et al., 2002; Shlomi et al., 2005). Two primary mechanisms that compensate the null mutations in metabolic networks include: redundancy of components (genetic buffering) and modules with overlapping functions (functional complementation) (Stelling et al., 2004). The redundant metabolic pathways are not straightforward in single metabolic network. To understand robustness in complex metabolic networks, more theoretical tools are needed.

The flux balance analysis (FBA) method is well presented for analyzing the robustness of metabolic networks. Details of FBA have been described elsewhere (Kauffman et al., 2003; Lee et al., 2006). The FBA method needs objective function, such as maximizing the lactic acid (Fong et al., 2005), ethanol (Pharkya and Maranas, 2006), succinic acid (Lee et al., 2005; Wang et al., 2006), L-valine (Park et al., 2007) or L-threonine (Lee et al., 2007). A lot of researches are based on the assumptions that maximize biomass yield (Feist and Palsson, 2008). Under a given environmental condition, reaction/gene deletions were simulated by constraining the flux of the corresponding reaction(s) to zero and calculating the corresponding knockout flux distribution. A reaction/gene was classified as having no essentiality if the biomass production rate of the knockout strain was not less than a given cutoff compared to the original strain; if the rate of a simulated reaction/gene deletion was below the cutoff, the deletion was assumed to be essential (Famili et al., 2003). This method can find many essential reactions/genes. It also finds many reactions/genes are not essential for growth, which are redundant for organism. Currently, only a few model organisms' biomass reactions are given in detail, such as *Bacillus subtilis* (Oh et al., 2007), *Escherichia coli* (*E. coli*) (Feist et al., 2007), *Helicobacter pylori* (Thiele et al., 2005), *Saccharomyces*

Correspondence to: S. Zhou and Y.-P.P. Chen

cerevisiae (Herrgard et al., 2006), and *Staphylococcus aureus* (Becker and Palsson, 2005). For many other organisms, these biomass reactions are not given. Therefore, analyzing the robustness of those organisms does not make use of the FBA method.

A central concept in metabolic pathway analysis is that of elementary flux modes (EFMs) (Schuster et al., 2000). An EFM is a minimal set of enzymes that can operate at steady state. EFM analysis appears to be well suited to characterize network redundancy because each EFM is non-redundant. By examining which of these modes form the same substrates to the same products, one can detect parallel routes. This is achieved based on the relative number of elementary modes remaining after the knockout of enzymes. Wilhelm et al. (2004) demonstrated quantitatively that the metabolism of *E. coli* is more robust than the one of human erythrocyte. Extending this study of single knockout, Behre et al. (2008) studied the structural robustness of metabolic networks when considering the general case of double and multiple knockouts. Because the number of pathways in EFMs grows exponentially with the increasing network size (Klamt and Stelling, 2002), Behre et al. (2008) only studied the amino acid synthesis in *E. coli* (119 metabolites and 164 reactions) and the central metabolism of human erythrocytes (36 metabolites and 41 reactions). In recent years, many other studies used the concept of EFMs to investigate the robustness of biochemical systems (Çakır et al., 2004; Gabaldón et al., 2007; Krömer et al., 2006; Schuster and Kenanov, 2005; Schwender et al., 2004; Stelling et al., 2002). All of these analyses focus on the middle-scale metabolic networks. The link between robustness and pathway function is not established for large-scale metabolic network. Lots of analyses of network structures elucidate design principles of metabolic networks, providing valuable insights into the functional organization of organisms.

The compound graph is that the nodes represent metabolites and the edges between nodes represent enzymatic reactions (Ravasz et al., 2002). The reaction graph is a dual form of compound graph that nodes are reactions and edges are compounds (Wagner and Fell, 2001). Many graph theories are used to investigate the features of large-scale metabolic networks (Barabási and Oltvai, 2004), such as, degree distributions (Arita, 2004; Jeong et al., 2000; Wagner and Fell, 2001; Zhu and Qin, 2005), average pathway length (Arita, 2004; Zhu and Qin, 2005), and average clustering coefficient (Takemoto et al., 2007; Zhu and Qin, 2005). These global properties reflect the complex interaction machineries. However, using the graph theories to investigate the metabolic networks misses the important compound character that any compound in metabolic network cannot be substituted directly by other single compound.

Here we try to investigate the local property and how it impacts other reactions by deleting an initial reaction in the metabolic network. The impact degree is defined as the number of cascading failures of following and/or forward reactions when an initial reaction is deleted. We have determined this to be the global property when calculating the average value of impact degrees among all reactions in the metabolic network. This method establishes the link between robustness and pathway function for large-scale metabolic network. This method does not need the biomass reactions as usually used in FBA method. We try to investigate what determines the robust mechanism of the metabolic networks from three aspects, the percentage of reversible reactions, the average connected degree, and the optimal growth temperature.

Materials and Methods

The metabolic network of each organism represents its inner relation of various metabolites. We downloaded metabolic reactions from the KEGG database (Kanehisa et al., 2004) at August 2008. The reaction and pathway data locate at the FTP server (ftp://ftp.genome.ad.jp/pub/kegg/release/current/). There are 854 organisms in current KEGG database. The number and content of reactions are variable in these organisms.

To investigate the compensatory mechanism of reaction deletion, we develop a method to calculate the impact degree, which is defined as the number of cascading failure of following or/and forward reactions when an initial reaction is deleted. There are two kinds of reactions, reversible reaction and irreversible reaction. Hence, we divide each reversible reaction into two irreversible reactions with different directions. Generally, substrate and product are defined as the consumed metabolites and the produced metabolites of reaction, respectively. A reaction's deletion may impact other reactions, which takes on two aspects. One is that the forward reaction may be terminated when the products cannot be consumed, the other is that the subsequent reactions without the metabolite (substrate) produced by the initial reaction will be terminated. These two kinds of impacted reactions will cascade to impact others.

Most reactions have multiple substrates and multiple products. For clearly presentation, we just use a simple example to illustrate the process. There is a schematic diagram to illustrate how to calculate the impact degree for each reaction, as shown in Figure 1. There are seven metabolites and six reactions in this figure. The arrow represents the reaction from the substrate to the product. If

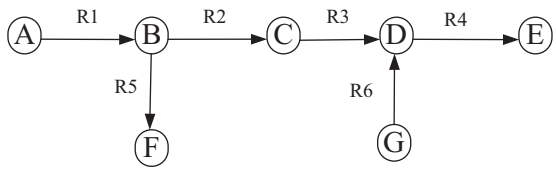

Figure 1. A sample of metabolic reaction network.

the reaction R1 is deleted, the direct following reactions R2 and R5 are influenced because no substrate produced by the R1 can supply these two reactions. The reaction R3 is also directly influenced by the reaction R2. But the reaction R4 and R6 are not influenced by the reaction R3. So, the impact degree of the R1 is 4, which contains R1, R2, R5, and R3. The deletion of reaction R1 only influences subsequent reactions. If the reaction R4 is deleted, it will influence the direct forward reaction R3 and R4. The metabolite D cannot be consumed so that these two reactions must be not active in the living organism. The reaction R2 also cannot be active because the metabolite C cannot be consumed. The reaction R1 and R2 are not influenced by the reaction R2. Hence, the impact degree of the reaction R4 is 4, which contains R4, R3, R6, and R2. Certainly, if the reaction R1 and R4 are deleted simultaneously, the impact degree of these two reactions is 6.

The metabolic networks have few cycles in which metabolites are transitive dependency. Because such cycles usually have many entrance and exit metabolites, not all the reactions in such cycles are influenced by the initial deleted reaction when reactions are cascading failure to the cycles. Therefore, we do not take account the reactions in the cycles. Certainly, the existence of these cycles may result in larger impact degrees of some reactions than that calculated by our method. However, the number of such reactions is small compared to the hundreds even thousands of reactions in an organism.

Results

Deletion of an intracellular reaction in biology primarily implies that the enzyme that catalyzes the particular reaction is either not produced or produced but rendered inactive. The inaction of an enzyme is also the result of knockout of the corresponding gene(s). However, the relations between these two joints, gene–enzyme and enzyme–reaction, are not straightforward. Some enzymes are independent proteins, which carry out identical reactions and some others need enzyme complexes (multiple enzymes) to be expressed for these reactions to occur. Generally, a reaction is catalyzed by a single-enzyme. Some other reactions can be catalyzed by many different enzymes, each of which can catalyze the reaction individually. Here, we mainly investigate the structure of the metabolic reaction network, and we also investigate the impact of the enzyme deletion on the metabolic reaction network. Due to the complex gene–enzyme relation and the absence of the gene–enzyme relationships for all the organisms, we do not take account of the knockout of genes.

The reaction 'R06432', which the equation is 'dTDP-L-olivose + *S*-Adenosyl-L-methionine ↔ dTDP-L-oleandrose + *S*-Adenosyl-L-homocysteine + H+', has the largest impact degree with the value of 27. This reaction only exists in the pathway of polyketide sugar unit biosynthesis in the bacteria of *Streptomyces avermitilis*. Only a few reactions have large impact degrees (Fig. 2). There are 170 reactions that have average impact degrees are greater than 5. Compared with the total number of reactions (3,377), the reactions with great impact degrees only occupy a small part (∼5%). Most reactions have small impact degrees. Especially, there are 955 reactions with the value of 1, which means that they do not influence any other reactions. In total, the average impact degree among all reactions in all organisms is ∼1.98. Such small value shows that the organisms have perfect compensatory mechanism.

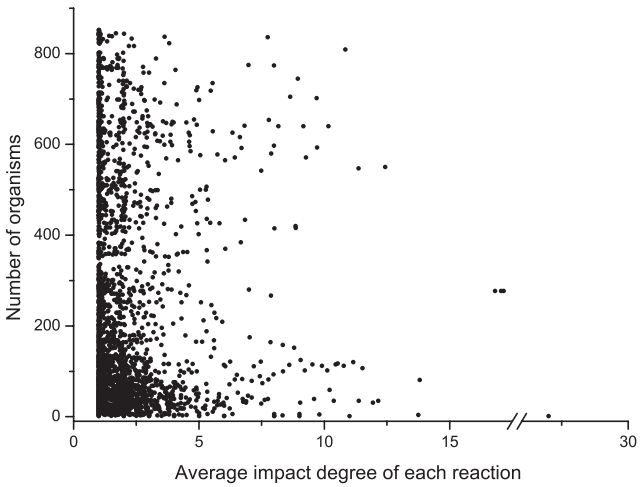

Figure 2. The relation between the average impact degree of the reaction and the number of organisms containing this reaction. Note that node represents reaction.

A reaction should be important if it exists in many organisms. Hence, our null hypothesis is that the reaction existing in many organisms should have a small impact degree because the organism should have strong compensatory mechanism to the absence of this reaction. However, there are cases where few reactions exist in many organisms and have great impact degrees as shown in Figure 2.

The gene knockout process is of higher biological relevance than the reaction deletion. The reaction essentiality is determined by the associated gene. Because there is no gene database to validate the reaction essentiality for all of the organisms, we can but refer to the *E. coli* gene essentiality published by Gerdes et al. (2003). Their work gave the enzyme–gene relation information of *E. coli* in detail. We also extract the reaction–enzyme information from the KEGG database and relate the reaction to the gene by the enzyme. Hence, the essentiality of a reaction can be inferred from the gene in the *E. coli*. However, some reactions are determined by multi enzymes or genes. Here, we consider that if a gene is essential, then the corresponding reaction is "essential," and if an enzyme does not exist or exist but the essentiality of this enzyme is not clear in the *E. coli*, then the corresponding reaction is defined as "unclear," and other reactions are defined as "non-essential." For presenting the essentiality of all reactions

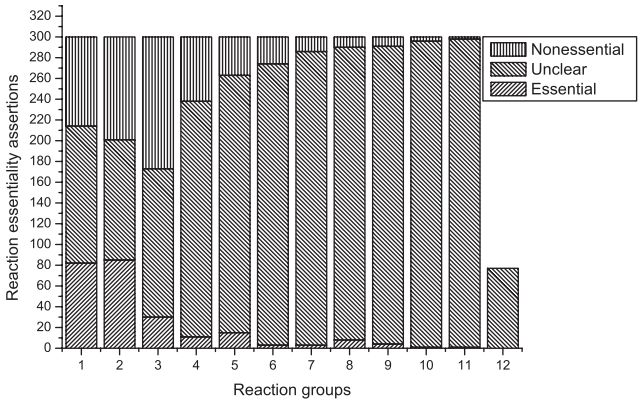

Figure 3. Reaction essentiality assertions in different groups. Note that the reactions are separated into 12 groups sorted by the total impact degree from large to small.

clearly, we sort the total impact degree of all reactions from large to small and separate these reactions into 12 groups, where each group has 300 reactions except for the last group (Fig. 3). The essential reactions occupy a big proportion in the first two groups (top 600 reactions sorted by the descendent total impact degree). Especially, the reactions with top largest total impact degrees almost are the essential ones as shown in Table I. Though many reactions with large impact degrees are not clear for the essentiality, they are most likely to be the essential ones in other organism. For example, the reaction R04109 (L-Glutamyl-tRNA(Glu) + NADPH + H+ ↔ (S)-4-Amino-5-oxopentanoate + tRNA(Glu) + NADP+) exists in 547 organisms and has a total of 6,213 impact degree among all the organisms. This reaction has an enzymes (EC 1.2.1.70) functioned as glutamyl-tRNA reductase. This enzyme is essential in many organisms, such as *Mycobacterium tuberculosis* H37Rv (Sassetti et al., 2003) and *Francisella novicida* U112 (Gallagher et al., 2007). Another example is that the reaction R00014 (Pyruvate + Thiamin diphosphate ↔ 2-(alpha-Hydroxyethyl)thiamine diphosphate + CO2) exists in 809 organisms and has a total of 8,762 impact degree among all the organisms. This reaction has enzyme (EC 1.2.4.1) functioning as pyruvate dehydrogenase. This enzyme is essential in many organisms, such as *B. subtilis* 168 (Kobayashi et al., 2003), *S. aureus* N315 (Ji et al., 2001), *Mycoplasma genitalium* G37 (Glass et al., 2006), and *M. pul monis* UAB CTIP (French et al., 2008).

When the total impact degree is decreasing, the proportion of the essential reaction is also decreasing. This result means that it is most likely to be an essential reaction if it has great impact degree. Contrarily, if a reaction has a small impact degree, then this reaction should not be essential.

We also separate all the reactions into 12 groups by the weight of the reaction where the weight is defined as the number of the different organisms containing this reaction. The reaction in the frontal groups exists in a lot of organisms

Table I. List of top 42 reactions with great impact degree.

Reaction	Impact degree	Essentiality
R00014	8,762	U
R00621	6,833	E
R00036	6,804	E
R01626	6,660	E
R00428	6,505	E
R05046	6,505	E
R05578	6,467	E
R04109	6,213	U
R01799	6,183	E
R00084	6,089	U
R05048	5,865	E
R02272	5,756	E
R07618	5,402	E
R03165	5,291	E
R04639	5,225	E
R00660	5,090	N
R05636	4,769	U
R05688	4,569	E
R03193	4,367	E
R02735	4,083	E
R02199	4,072	N
R05633	4,056	E
R05634	3,967	E
R00586	3,965	E
R00615	3,923	U
R03504	3,671	N
R01818	3,586	N
R07460	3,516	U
R03018	3,490	E
R03197	3,487	E
R05637	3,411	E
R02473	3,312	N
R02783	3,155	E
R01150	3,144	E
R02016	3,135	U
R00734	3,099	E
R03650	3,034	E
R04958	2,913	E
R05883	2,908	U
R00401	2,873	N
R00511	2,869	U
R04457	2,825	U

E, essential; N, non-essential; U, unclear.

(Fig. 4). It is distinct that the reaction with great weight is most likely to be an essential one (Table II). It is very possible that the reaction is essential when it exists in many organisms. These universal reactions are important for the organisms. However, the relation between the average impact degree of a reaction and the number of organisms holding this reaction does not exist with a significant correlation (Fig. 2). We speculate that the reaction with both great impact degree and a great weight should be most likely the essential one. Therefore, we testify the reactions exist both in the first group of the two kinds of sort orders (Figs. 3 and 4). We find the essential reactions occupy 44.54% of the total 119 reactions and the reactions with unclear essentiality occupy 29.41%. We also testify the reactions exist both in the first two groups of the two kinds of sort orders. There are 417 reactions having both great impact degree and great

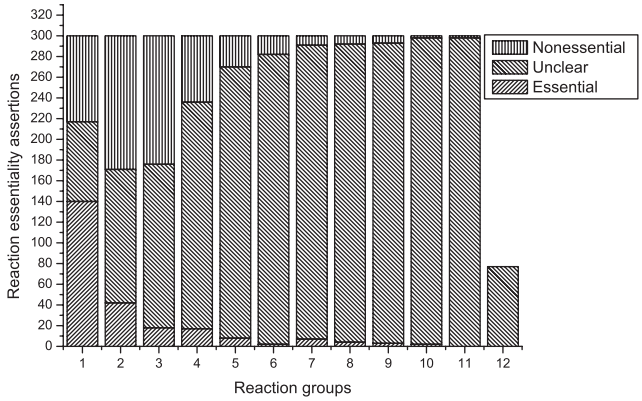

Figure 4. Reaction essentiality assertions in different groups. Note that the reactions are separated into 12 groups sorted by the weight, which is the number of organisms containing the corresponding reaction from large to small.

weight. The essential reactions occupy 37.41% and the reactions with unclear essentiality occupy 28.30%. The reaction in the unclear group is most likely to be the essential one though it is unclear in the *E. coli*. For example, the reaction R01518 (2-Phospho-D-glycerate ↔ 3-Phospho-D-glycerate) exists in 833 organisms and the total impact degree is also 833. This reaction has enzyme (EC 5.4.2.1) functioning as phosphoglycerate mutase. This enzyme is essential in the *B. subtilis* 168 (Kobayashi et al., 2003), *M. genitalium* G37 (Glass et al., 2006), and *Francisella novicida* U112 (Gallagher et al., 2007).

We take the metabolic network as reaction graphs where the node is the reaction and an edge exists between two reactions if there is a compound, which is both a product of one reaction and a substrate of the other one (Nacher et al., 2005; Ramezanpour et al., 2003; Wagner and Fell, 2001). We calculate the connected degree for each reaction in all the organisms. We then calculate the average connected degrees for each reaction among all the organisms. For example, the reactions R00006 exists in 688 organisms and all the total connected degrees for this reaction in 688 organisms are 20,147 hence the average connected degree of this reaction is 39.28. In order to calculate the average connected degree distribution, we round value of the average connected degree to the integer for all reactions. Hence, the average connected degree of the reaction R00006 is 39. The average connected degree distribution is shown in Figure 5. This distribution is in good agreement with power law, specifically, $P(k) \sim k^{-\gamma}$, where $P(k)$ is the probability of finding a vertex with degree k, and γ is the connected degree exponent, with its value being 1.08. This reaction graph is scale-free. The scale-free networks have been thought to be robust against accidental failures. This is because random failure affects mainly the node with small connected degree. Such failed nodes do not disrupt the networks' integrity (Albert et al., 2000).

We group all the reactions by the average connected degree from large to small (Fig. 6). The essential reactions

Table II. List of top 42 reactions with great weight.

Reaction	Weight	Essentiality
R00377	852	E
R00378	852	E
R00376	852	E
R00375	852	E
R03660	850	U
R03662	848	E
R03658	845	E
R00127	845	E
R01547	845	E
R03656	844	E
R03038	844	E
R03654	843	E
R03664	843	E
R02918	842	E
R03657	841	E
R03665	840	E
R05577	840	E
R03663	839	E
R03661	839	E
R03655	838	E
R00158	838	E
R03650	837	E
R05578	836	E
R03659	833	E
R01518	833	U
R03646	832	E
R01512	831	E
R00658	831	U
R01015	829	U
R00945	827	E
R01056	826	U
R00571	826	E
R00573	826	E
R04773	824	E
R02016	823	U
R01049	822	E
R04771	817	E
R00177	817	E
R00014	809	U
R02018	802	E
R02024	802	E
R02094	802	E

E, essential; N, non-essential; U, unclear.

almost distribute equally in each group. We cannot speculate the essential reaction by the distribution of the average connected degree of the reaction node. It is implied that the reaction node with high average connected degree need not be important for the real organisms. It is observed that the essentiality of reactions in a node is not correlated with node connectivity as structural analysis can suggest. This character partly determines the robustness of the organisms because the lethal probability is reduced even if the hub reaction of the metabolic network was attacked.

For each organism, it has a global property, which is the average connected degree among all reactions in this organism. Such global property can represent the complexity of organism and the denseness of pathways in organism. The more redundant pathways are, the higher of average connected degree is. We also investigate the ability of the

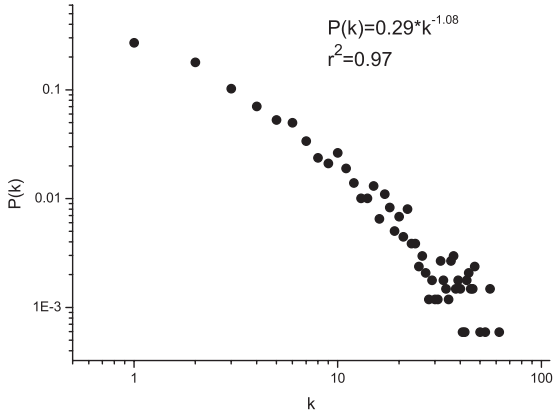

Figure 5. Distribution of degree in reaction graph with both axes plotted on log scales.

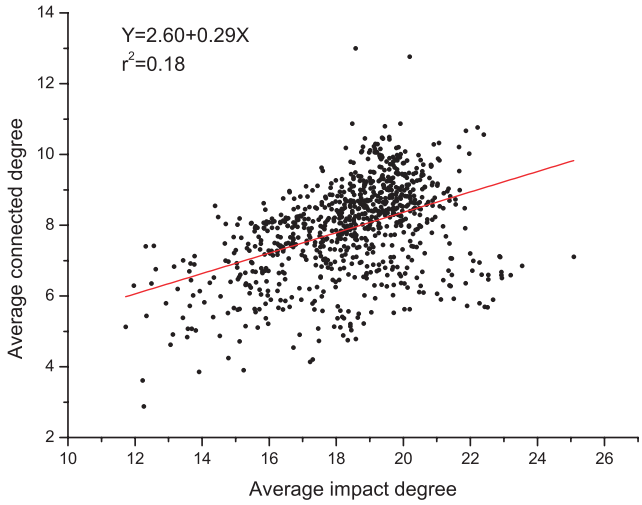

Figure 7. The relation between the average connected degree and the average impact degree among all organisms. Note the average impact degree is adopted from 1,000 repetitions of randomly deleting 10 reactions synchronously. Each node is a reaction. [Color figure can be seen in the online version of this article, available at www.interscience.wiley.com.]

organism to compensate for the deletion of multi-reactions simultaneously. Here, we take account of randomly deleting 10 reactions at a time for all the organisms and repeating this process to 1,000 times. Hence, every organism has a value representing the ability to compensate the deletion of ten reactions. Figure 7 represents the correlation between the average connected degree and the compensatory ability for each organism with the value of 0.18. The relativity between the average degree and the compensatory ability represents that redundant pathways contribute robustness of the organism.

For those 854 organisms, there are 52 archaea organisms, 137 eukayote organisms, and 665 bacteria organisms. We also separate these organisms into nine groups according to their average impact degrees from large to small (Fig. 8). We find the archaea organisms have a small average impact degree compared to two other kinds of organisms. Archaea organisms may be the oldest form of life on Earth. Most live in extreme habitats such as extremely hot thermal vents or hypersaline water. The small average impact degrees of archaea organisms indicate that these organisms have strong robustness to resist the various perturbations during the evolution process.

We also calculate the percentage of the reversible reactions for each organism. Figure 9 represents the relation between the average impact degree of random ten reactions and the percentage of the reversible reactions for each organism. Though the relativity of these two characters

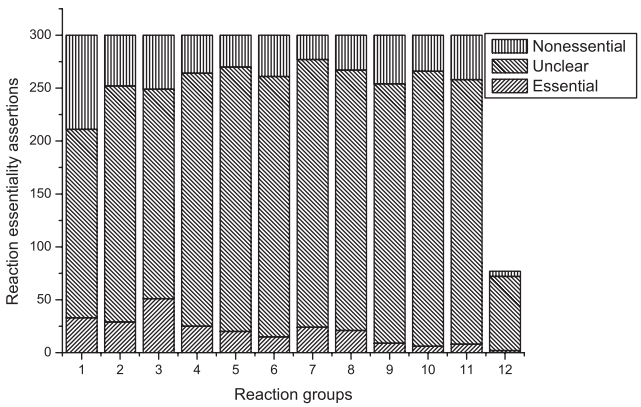

Figure 6. Reaction essentiality assertions in different groups. Note the reactions are separated into 12 groups sorted by the average reaction connected degree from large to small.

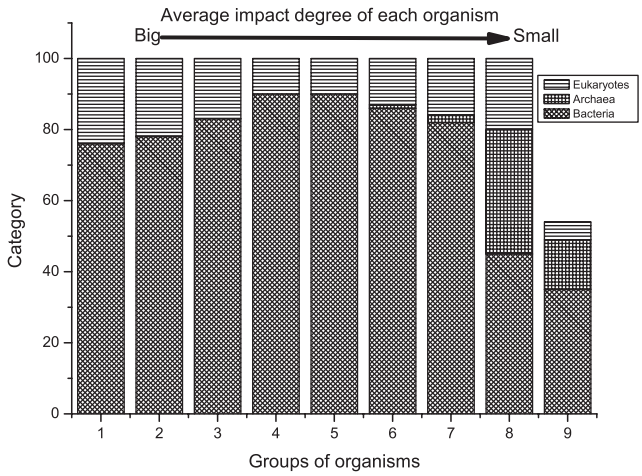

Figure 8. Category assertions in different groups. Note the average impact degree is adopted from deleting 10 reactions synchronously among 1,000 repetitions.

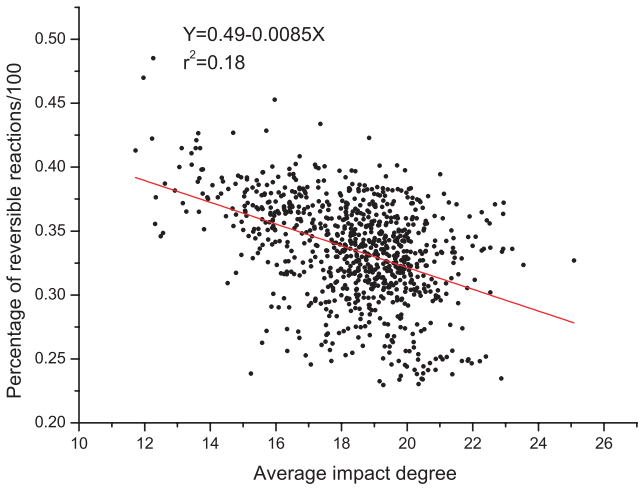

Figure 9. The relation between the percentage of reversible reactions and the average impact degree among all organisms. Note the average impact degree is adopted from 1,000 repetitions of randomly deleting 10 reactions synchronously. Each node is an organism. [Color figure can be seen in the online version of this article, available at www.interscience.wiley.com.]

is not strong with the value of 0.18, it implies the compensating null mutations of the organism are partly determined by the reversible reactions.

We further investigate whether or not the optimal growth temperature is correlated with the average impact degree. The optimal growth temperature of 113 prokaryotic organisms are obtained from the work of Takemoto et al. (2007), which contains 18 archaea organisms, and 95 bacteria organisms. Figure 10 shows significant negative corrections between optimal growth temperature and average impact degree. The percentage of the reversible

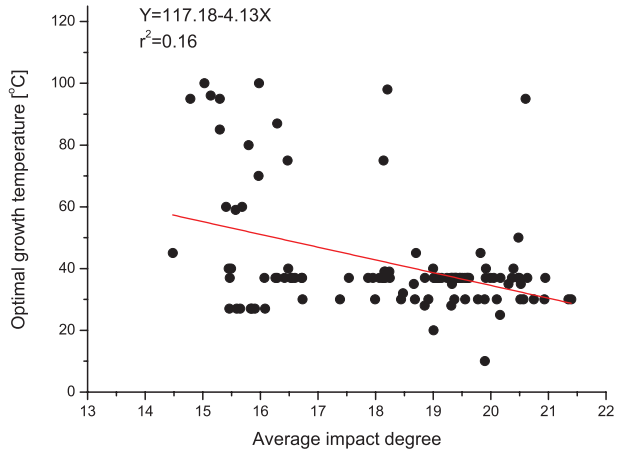

Figure 10. The relation between the optimal growth temperature and the average impact degree among 113 prokaryotic organisms, including 18 archaea organisms, and 95 bacteria organisms. Note the average impact degree is adopted from 1,000 repetitions of randomly deleting 10 reactions synchronously. Each node is an organism. [Color figure can be seen in the online version of this article, available at www.interscience.wiley.com.]

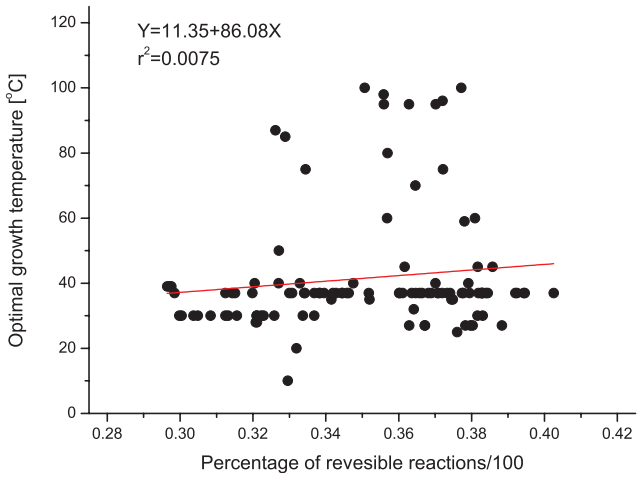

Figure 11. The relation between the optimal growth temperature and the percentage of reversible reactions among 113 prokaryotic organisms. [Color figure can be seen in the online version of this article, available at www.interscience.wiley.com.]

reactions is positively correlated with optimal growth temperature (Fig. 11).

Discussion

Among many researches to predict the essential genes by analyzing the metabolic networks, the FBA method based on optimization is the most important one. Currently, many extensions of FBA method have developed (Burgard et al., 2003; Dashika et al., 2006). FBA method typically invokes the optimization of a particular cellular objective. Most objectives are set as the maximizing biomass yield (Feist and Palsson, 2008). Some researches approve it (Gianchandani et al., 2008), but others do not confirm it (Nielsen, 2007; Schuster et al., 2008). Besides, the biomass reaction, which represents the growth of the organism is hard to give for many organisms. It still has many difficulties to predict essential reaction correctly by FBA method (Becker and Palsson, 2008). The theory of minimal cut set based on EFMs is provided firstly by Klamt and Gilles (2004), which is minimal set of reactions in the network whose knockout will definitely lead to a failure in certain network functions. The reactions in such set are essential for the target. The analysis of potential failure modes in metabolic networks will help to identify crucial parts in the network structure. Though Haus et al. (2008) provided an improved algorithm to calculate the minimal cut set and EFMs for solving even large-scale metabolic network, the time required was long. Another disadvantage of calculating the EFMs to sustain the growth is that the target reaction (biomass synthesize) is also not given for a lot of organisms' metabolic network. Our method does not need the target reaction and hence provides a compensatory mechanism to predict essential reaction.

Zhu and Qin (2005) found that archaea metabolic networks differentiate significantly from those of bacteria and eukayote organisms by analyzing the network indices, degree distribution, and motif profile. Maybe the archaea metabolic networks suffer small impaction from inner disturbance as illustrated in our work, which results in the structure difference from the other two categories. It is an interesting phenomenon that the structural properties are correlated with the optimal growth temperature (Takemoto et al., 2007). Here, we also find that optimal growth temperature is negatively correlated with the impact degree among 113 prokaryotic organisms. As we all known that it does not hold the absolute irreversible reactions. Usually, some irreversible reactions turn into reversible reactions when the temperature arises. Therefore, the high temperature results in high percentage of reversible reactions in the metabolic network. Because the average impact degree is negatively correlated with the percentage of reversible reactions and the small average impact degree implies the more robustness of metabolic network, the robustness of organism is also determined by the optimal growth temperature. The existence of reversible reaction enlarges the feasible space of flux distribution, which also results in more robustness. The scopes of compounds are investigated by Handorf et al. (2005). Such scopes comprise all compounds, which can be synthesized from the seed substrates. It demonstrated that most of the deletions of single reaction have only a small or even no effect on the scope size. The network expansion is in general very robust against elimination of single or few reactions. The method of scopes for compounds analyzed the robustness of all reactions in the KEGG database, not differentiating the organisms. They described the robustness of biosphere in some extent when all reactions are collected together. Their method is an expanded process from one compound or a few compounds, whereas our method is reduced processes that delete reactions from the complete metabolic network. Both methods provide new insight into robust analysis of metabolic network.

Conclusion

This article starts with the analysis of the local property of the metabolic network, which is the impact degree of each reaction to the metabolic network. It suggests that the reaction with larger impact degree to others is likely to be essential. Furthermore, the universal reactions are also likely essential. If the reaction has both a large impact degree and exists in a lot of organisms, the reaction should be classed as essential. Taking the metabolic reaction network as the graph where the nodes are the reaction and the edges are the relation of two reactions connected by the common metabolite, the metabolic networks have a scale-free property. In graph theory, the scale-free feature can resist the most uncertain perturbation where the chance to attack the hub of the network is small. Though graph theory suggests that hubs are the most important feature of the whole system, the cellular networks do not follow this rule that the essential reactions exist not only in the hub nodes but also in other non-hub nodes.

The average impact degree of an organism is the global property of this organism. Taking the metabolic network as a graph, the organism also has another global property, which is the average connected degree. For all the organisms, these two kinds of degrees (average impact degree and average connected degree) have distinct relativity, which represents redundant pathways contributing to the robustness of the organisms. Another global property of the organism is the percentage of reversible reactions. If an organism has many reversible reactions, the one would have more strong robustness. The robustness also correlates to the optimal growth temperature.

Furthermore, comparing the three categories of organisms, the archaea organisms have small average impact degrees. This discovery indicates the archaea organisms with strong robustness can resist perturbation and live in extreme environments through evolutionary history.

Da Jiang and Shuigeng Zhou are supported by National Natural Science Foundation of China under grant No. 60496327 and Shanghai Leading Academic Discipline Project (Project number B114). Yi-Ping Phoebe Chen is partially supported by Australia Research Council DP 0559251 and CE0561992.

References

Albert R, Jeong H, Barabási A-L. 2000. Error and attack tolerance of complex networks. Nature 406:378–382.

Arita M. 2004. The metabolic world of Escherichia coli is not small. Proc Natl Acad Sci USA 101:1543–1547.

Barabási A-L, Oltvai ZN. 2004. Network biology: Understanding the cell's functional organization. Nat Rev Genet 5:101–113.

Becker SA, Palsson BO. 2005. Genome-scale reconstruction of the metabolic network in Staphylococcus aureus N315: An initial draft to the two-dimensional annotation. BMC Microbiol 5:8.

Becker SA, Palsson BO. 2008. Three factors underlying incorrect in silico predictions of essential metabolic genes. BMC Syst Biol 2:14.

Behre J, Wilhelm T, von Kamp A, Ruppin E, Schuster S. 2008. Structural robustness of metabolic networks with respect to multiple knockouts. J Theor Biol 252:433–441.

Burgard AP, Pharkya P, Maranas D. 2003. OptKnock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol Bioeng 84:647–657.

Çakır T, Tacer CS, Ülgen KÖ. 2004. Metabolic pathway analysis of enzyme-deficient human red blood cells. Biosystems 78:49–67.

Dashika MS, Burgard A, Maranas D. 2006. A computational framework for the topological analysis and targeted disruption of signal transduction networks. Biophys J 91:382–398.

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

Feist AM, Palsson BO. 2008. The growing scope of applications of genome-scale metabolic reconstructions using Escherichia coli. Nat Biotechnol 26:659–667.

Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BO. 2007. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that

accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol 3:121.
Fong SS, Nanchen A, Palsson BO. 2005. In silico design and adaptive evolution of Escherichia coli for production of lactic acid. Biotechnol Bioeng 91:643–648.
French CT, Lao P, Loraine AE, Matthews BT, Yu H, Dybvig K. 2008. Large-scale transposon mutagenesis of Mycoplasma pulmonis. Mol Microbiol 69:67–76.
Gabaldón T, Peretó J, Montero F, Gil R, Latorre A, Moya A. 2007. Structural analyses of a hypothetical minimal metabolism. Philos Trans R Soc B 362:1751–1762.
Gallagher LA, Ramage E, Jacobs MA, Kaul R, Brittnacher M, Manoil C. 2007. A comprehensive transposon mutant library of Francisella novicida a bioweapon surrogate. Proc Natl Acad Sci USA 104:1009–1014.
Gerdes SY, et al. 2003 Experimental determination and system level analysis of essential genes in Escherichia coli MG1655. J Bacteriol 185:5673–5684.
Gianchandani EP, Oberhardt MA, Burgard AP, Maranas CD, Papin JA. 2008. Predicting biological system objectives de novo from internal state measurements. BMC Bioinformat 9:43.
Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M, Hutchison CA III, Smith HO, Venter JC. 2006. Essential genes of a minimal bacterium. Proc Natl Acad Sci USA 103:425–430.
Handorf T, Ebenhöh O, Heinrich R. 2005. Expanding metabolic networks: Scopes of compounds robustness and evolution. J Mol Evol 61:498–512.
Haus UU, Klamt S, Stephen T. 2008. Computing knockout strategies in metabolic networks. J Comput Biol 15:259–268.
Herrgard MJ, Lee B-S, Portnoy V, Palsson BO. 2006. Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in Saccharomyces cerevisiae. Genome Res 16:627–635.
Jeong H, Tombor B, Albert R, Oltvai ZN, Barabási A-L. 2000. The large-scale organization of metabolic networks. Nature 406:651–654.
Ji Y, Zhang B, Van Horn SF, Warren P, Woodnutt G, Burnham MK, Rosenberg M. 2001. Identification of critical staphylococcal genes using conditional phenotypes generated by antisense RNA. Science 293:2266–2269.
Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. 2004. The KEGG resource for deciphering the genome. Nucleic Acids Res 32: 277–280.
Kauffman KJ, Prakash P, Edwards JS. 2003. Advances in flux balance analysis. Curr Opin Biotech 14:491–496.
Klamt S, Gilles ED. 2004. Minimal cut sets in biochemical reaction networks. Bioinformatics 20:226–234.
Klamt S, Stelling J. 2002. Combinatorial complexity of pathway analysis in metabolic networks. Mol Biol Rep 29:233–236.
Kobayashi K, Ehrlich SD, Albertini A, et al. 2003. Essential Bacillus subtilis genes. Proc Natl Acad Sci USA 100:4678–4683.
Krömer JO, Wittmann C, Schröder H, Heinzle E. 2006. Metabolic pathway analysis for rational design of L-methionine production by Escherichia coli and Corynebacterium glutamicum. Metab Eng 8:353–369.
Lee SJ, Lee DY, Kim TY, Kim BH, Lee J, Lee SY. 2005. Metabolic engineering of Escherichia coli for enhanced production of succinic acid based on genome comparison and in silico gene knockout simulation. Appl Environ Microbiol 71:7880–7887.
Lee JM, Gianchandani EP, Papin JA. 2006. Flux balance analysis in the era of metabolomics. Brief Bioinform 7:140–150.
Lee KH, Park JH, Kim TY, Kim HU, Lee SY. 2007. Systems metabolic engineering of Escherichia coli for L-threonine production. Mol Syst Biol 3:149.
Motter AE, Gulbahce N, Almaas E, Barabási A-L. 2008. Predicting synthetic rescues in metabolic networks. Mol Syst Biol 4:168–178.
Nacher JC, Yamada T, Goto S, Kanehisa M, Akutsu T. 2005. Two complementary representations of a scale-free network. Physica A 349:349–363.
Nielsen J. 2007. Principles of optimal metabolic network operation. Mol Syst Biol 3:126–127.
Oh Y-K, Palsson BO, Park SM, Schilling CH, Mahadevan R. 2007. Genome-scale reconstruction of metabolic network in Bacillus subtilis based on high-throughput phenotyping and gene essentiality data. J Biol Chem 282:28791–28799.
Park JH, Lee KH, Kim TY, Lee SY. 2007. Metabolic engineering of Escherichia coli for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation. Proc Natl Acad Sci USA 104:7797–7802.
Pharkya P, Maranas CD. 2006. An optimization framework for identifying reaction activation/ inhibition or elimination candidates for overproduction in microbial systems. Metab Eng 8:1–13.
Ramezanpour A, Karimipour V, Mashaghi A. 2003. Generating correlated networks from uncorrelated ones. Phys Rev E 67:046107.
Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási A-L. 2002. Hierarquical organization of modularity in metabolic networks. Science 297:1551–1555.
Sassetti CM, Boyd DH, Rubin EJ. 2003. Genes required for mycobacterial growth defined by high-density mutagenesis. Mol Microbiol 48:77–84.
Schuster S, Kenanov D. 2005. Adenine and adenosine salvage pathways in erythrocytes and the role of S-adenosylhomocysteine hydrolase a theoretical study using elementary flux modes. FEBS J 272:5278–5290.
Schuster S, Fell DA, Dandekar T. 2000. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nat Biotechnol 18:326–332.
Schuster S, Pfeiffer T, Fell DA. 2008. Is maximization of molar yield in metabolic networks favoured by evolution? J Theor Biol 252:497–504.
Schwender J, Goffman F, Ohlrogge JB, Shachar-Hill Y. 2004. Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds. Nature 432:779–782.
Segre D, Vitkup D, Church GM. 2002. Analysis of optimality in natural and perturbed metabolic networks. Proc Natl Acad Sci USA 99:15112–15117.
Shlomi T, Berkman O, Ruppin E. 2005. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. Proc Natl Acad Sci USA 102:7695–7700.
Stelling J, Klamt S, Bettenbrock K, Schuster S, Gilles ED. 2002. Metabolic network structure determines key aspects of functionality and regulation. Nature 420:190–193.
Stelling J, Sauer U, Szallasi Z, Doyle FJ III, Doyle J. 2004. Robustness of cellular functions. Cell 118:675–685.
Takemoto K, Nacher JC, Akutsu T. 2007. Correlation between structure and temperature in prokaryotic metabolic networks. BMC Bioinformat 8:303.
Thiele I, Vo TD, Price ND, Palsson BO. 2005. An Expanded Metabolic Reconstruction of Helicobacter pylori (iIT341 GSM/GPR): An in silico genome-scale characterization of single and double deletion mutants. J Bacteriol 187:5818–5830.
Wagner A, Fell D. 2001. The small world inside large metabolic network. Proc R Soc Lond B 268:1803–1810.
Wang Q, Chen X, Yang Y, Zhao X. 2006. Genome-scale in silico aided metabolic analysis and flux comparisons of Escherichia coli to improve succinate production. Appl Microbiol Biotechnol 73:887–894.
Wilhelm T, Behre J, Schuster S. 2004. Analysis of structural robustness of metabolic networks. IEE Proc Syst Biol 1:114–120.
Zhu D, Qin ZS. 2005. Structural comparison of metabolic networks in selected single cell organisms. BMC Bioinformat 6:8.