STRUCTURAL AND FUNCTIONAL ANALYSIS OF GIANT STRONG COMPONENT OF *BACILLUS THURINGIENSIS* METABOLIC NETWORK

Ding, D.W.[1,2]; Ding, Y.R.[1]*; Li, L.N.[3]; Cai, Y.J.[4]; Xu, W.B.[1]*

[1]School of Information Technology, Jiangnan University, Wuxi 214036, China; [2]Department of Mathematics and Computer Science, Chizhou College, Chizhou 247000, China; [3]Department of Environmental Science, East China Normal University, Shanghai 200062, China; [4]School of biotechnology Key Laboratory of Industrial Biotechnology, Jiangnan University, Wuxi 214036, China

ABSTRACT

The purpose of this work was to study the giant strong component (GSC) of *B. thuringiensis* metabolic network by structural and functional analysis. Based on so-called "bow tie" structure, we extracted and studied GSC with its functional significance. Global structural properties such as degree distribution and average path length were computed and indicated that the GSC is also a small-world and scale-free network. Furthermore, the GSC was decomposed and functional significant for metabolism of these divisions were investigated by comparing to KEGG metabolic pathways.

Key words: *Bacillus thuringiensis*, Giant Strong Component, metabolic network

INTRODUCTION

Advancements in the emerging systems biology in recent years fuelled the expectation that we could understand cellular behaviors by discovering how function arises in the interactions of cellular components (19). Thanks to the high-throughput (HT) technologies, which allow us to list all of these cellular components for an organism on the genome scale, and thus more and more biochemical networks are reconstructed, such as metabolic networks (8-10), transcriptional regulatory networks (15) and signaling networks (28).

However, due to combinatorial explosion of pathways, it is difficult or even impossible to apply traditional pathway analysis methods (33-34) to these reconstructed networks. Help forward the way is provided by the rapidly developing complex networks, in which graph representation is widely used (1,3,5,16,23,26). For instance, the metabolic network could be represented by so-called metabolite graph in which the nodes are metabolites and the links are reactions. Then, the fundamental organizational principles that underlie networks could be discovered based global topological structural properties such as so-called "small-world" (36), "scale-free" (2) etc. Furthermore, to discover

functional units involved in metabolic networks, it is suggested that metabolic networks should have modularity (29,31,32) which is similar to other complex networks, such as social networks, Internet, Worldwide Web etc.

In this article, the use of metabolic reaction data to generate a metabolic network with 830 nodes and 1132 links of an important insecticidal bacterium *B. thuringiensis* (11,22) is achieved firstly. Subsequently, structural analysis of *B. thuringiensis* metabolic networks is explained and discussed based "bow tie" structure which is proposed by Ma and Zeng (24), emphasis is placed on the giant strong component (GSC) part. At last, the functional significance, global structural properties and modularity of GSC of *B. thuringiensis* metabolic networks are studied.

MATERIALS AND METHODS

Data Acquisition and Representation

To investigate the topological properties of metabolism of *B. thuringiensis*, we first obtained all metabolic reactions involved in metabolic network of *B. thuringiensis* from KEGG database (17), and use number of each metabolite correspond

*Corresponding Author. Mailing address: School of Information Technology, Jiangnan University, Wuxi 214036, China. Tel.: 86-0510-85912136 Fax: 86-0510-85912136. E-mail: sytu_xu@yahoo.com.cn

to compounds in the KEGG LIGAND database. For instance, metabolite 246 corresponds to compound C00246 (butanoate) in the KEGG database. Subsequently, all of the reactions are revised based a KEGG-based database developed by Ma and Zeng (25): 1) corrected obvious inconsistencies; 2) confirmed the reversibility of every reaction; 3) excluded the current metabolites and small molecules such as ATP, ADP, NADH and H_2O etc, with the purpose of reflecting biologically meaningful transformations. At last, the metabolic network reconstructed is represented by so-called metabolite graph in which the nodes are metabolites and the links are reactions. For example, the irreversible reaction, $64 + 26 \rightarrow 25$ is represented by two directed arcs $64 \rightarrow 25$ and $26 \rightarrow 25$.

Bow Tie Structure

Since Ma and Zeng proposed (24) the "bow tie" structure of metabolic networks, it is increasingly recognized as being a conserved property of complex networks, as highlighted by recent studies (6,20,21,37), and the results suggest that this structure property is functional meaningful for metabolism, disease and the design principle of biological robustness.

Generally speaking, a network with the "bow tie" structure could be decomposed into four parts: 1) giant strong component (GSC), 2) substrate subset (S), 3) product subset (P), and 4) isolated subset (IS) (24). The GSC is the biggest strongly connected components of a metabolic network.

Degree Distribution and Average Path Length

The direct reflection of difference among numerous metabolites in metabolic networks is the connection degree k, which is the links that the node has to others, and the degree distribution $P(k)$ gives the probability of a node with degree k. One of the most important properties of metabolic networks is the power law degree distribution, i.e. $P(k) \sim k^{\tau}$ ($r \approx 2.2$), which means that most of the nodes in the network have a low degree, while a few nodes have a very high degree (16, 35). In other words, metabolic network is a sort of typical scale-free network (2). It is suggested that average path length of metabolic networks is very small (16,25), shown itself the property of "small-world". Another structure parameter is network diameter, which is defined as the path length of the longest pathway among all of the shortest pathways (4).

Modularity and Simulated Annealing Algorithm

An important properties related to detection of modules is modularity. For a presumptive partition of the nodes of a network into modules, the modularity M of this partition is defined as following (14,27):

$$M \equiv \sum_{s=1}^{r} \left[\frac{l_s}{L} - \left(\frac{d_s}{2L} \right)^2 \right] \qquad (1)$$

where r is the number of modules, ls is the number of links between nodes in modules, ds is the sum of the degrees of the nodes in module s, and L is the total number of links in the network. It is suggested that maximization of the modularity function would yield the most accurate results for random networks and widely used for identification of modules (12,13).

Simulated annealing (18) is a stochastic optimization technique that could find 'low cost' configuration without getting trapped in 'high cost' local minima. As mentioned above, the method based on simulated annealing tries to find the optimal partitions of modules by maximizing the network modularity (12,13), and thus the cost is C= − M herein, where M is the modularity defined in equation (1). At each temperature T, some random updates are performed and accepted with probability:

$$p = \begin{cases} 1 & if \quad C_2 \leq C_1 \\ \exp(-\dfrac{C_2 - C_1}{T}) & if \quad C_2 \geq C_1 \end{cases} \qquad (2)$$

where $C2$ and $C1$ are respectively the cost after the update and before the update, while T is computational temperature. Specifically, at each temperature T there would be $ni = f S2$ nodes individual movements from one module to another, and $nc = f S$ nodes collective movements, where S is the number of nodes in the network, and f with the recommended range of 0.1 to 1. At each certain temperature T, the system would be cooled down to $T' = cT$.

RESULTS AND DISCUSSION

Bow Tie Structure and Extraction of GSC

The metabolic network of *B. thuringiensis* is reconstructed based the methods which is introduced in section 2.1. The network contains 830 nodes and 1132 links, and the global topology structure is shown in Fig. 1. It is clearly that the whole network is far from strong component and included many isolated reactions. Then the whole metabolic network of *B. thuringiensis* is decomposed into four parts based the "bow tie" structure (Table 1). It should be noted that most nodes in S, P and IS part are connected by some single link which are not interested herein, while the metabolites and reactions involved in the GSC part is clearly much less than the whole network, and would be used to reduce the complexity of applying other pathway analysis methods such as extreme pathways and elementary modes (33,34). Furthermore, the GSC is the biggest strongly connected components of a metabolic network and determined structure of the entire network at a certain extent (24,37), thus it would be more detailed analysis herein.

All of the 268 metabolic reactions are compared to KEGG pathways, and show that they are mainly concentrated on
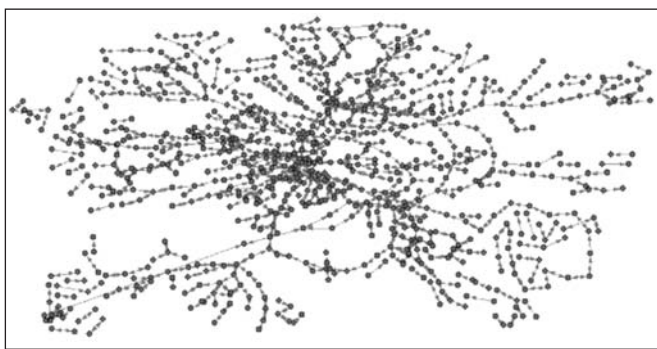
Figure 1. Metabolic network topology structure of *B. thuringiensis*, the nodes correspond to metabolites and the lines correspond to reactions. The picture was drawn using the Pajek program.

Table 1. The bow tie structure of *B. thuringiensis* metabolic network. GSC (giant strong component), S (substrate subset), P (product subset) and IS (isolated subset).

Subsets	GSC	S	P	IS	Total
No. of metabolites	118	73	190	449	830
Percentage of metabolites	14.2%	8.8%	22.9%	54.1%	100%
No. of reactions	268	82	252	530	1132
Percentage of reactions	23.7%	7.2%	22.3%	46.8%	100%

carbohydrate metabolism and amino acid metabolism (Table 2). The reactions of carbohydrate metabolism accurately correspond to glycolysis, TCA cycle, pentose phosphate pathway, and partly correspond to pyruvate metabolism and butanoate metabolism. From the point of view of network topological, the results show that metabolites in carbohydrate metabolism (in particular glycolysis, TCA cycle and pentose phosphate pathway, i.e. the central metabolism) have the higher probability of much more links and stronger robustness in network, and thus might have higher attack tolerance despite external cues, genetic variation and stochastic noise. While reactions of amino acid metabolism are mainly concentrated on urea cycle and metabolism of amino groups, arginine and proline metabolism, and glycerophospholipid metabolism, these might reveal the nutrient requirement in *B. thuringiensis*.

Degree Distribution and Average Path Length

We first checked the scale-free property of the GSC of *B. thuringiensis* metabolic network (Fig. 2). As it known, the nodes with high degree of scale-free network would dominate the

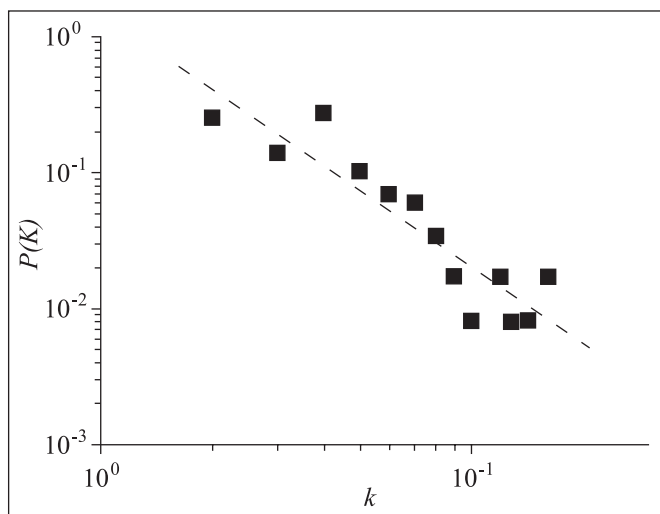**Table 2.** Reactions in GSC of *B. thuringiensis* metabolic network.

Reactions in GSC	No. of reactions	Percentage of reactions
Carbohydrate Metabolism	140	52.2%
Amino Acid Metabolism	84	31.3%
Energy Metabolism	24	9.0%
Lipid Metabolism	8	3.0%
Others	12	4.5%
Total	268	100%

network structure, and make the network robust against random errors such as mutation and environmental changes. Ma and Zeng have identified 20 primary metabolites with the highest degree for 80 fully sequenced organisms and suggested these metabolites are almost the same (25). The result is partly reaffirmed in this study (Table 3), 6 of 10 hub metabolites of the GSC of *B. thuringiensis* metabolic network are present in their list (PYR, GLU, AcCoA, ICIT, ASP and SUC), while the remained 4 metabolites are not. Among these 4 metabolites, BuCoA is the key metabolite linking butanoate metabolism and fatty acid metabolism, and it suggested that it is a key role related to novel pathway about PHB metabolism (7). 2HPP is the metabolite linking glycolysis pathway, pentose phosphate pathway and carbon fixation, E4P is the metabolite linking pentose phosphate pathway, aminosugars metabolism and carbon fixation, and GlyP play a key role among glycolysis pathway, fructose and mannose metabolism, glycerophospholipid metabolism, carbon fixation, nicotinate and nicotinamide metabolism. As links among different functional metabolic payhways, these hub metabolites (especially those 4 which are differ from Ma's universal hub

Table 3. The first 10 hub metabolites of the GSC of *B. thuringiensis* metabolic network.

Degree	Metabolite name	Abbreviation
16	Pyruvate	PYR
16	(2R)-2-Hydroxy-3-(phosphonooxy)-propanal	2HPP
14	Glycerone phosphate	GlyP
13	L-Glutamate	GLU
12	Acetyl-CoA	AcCoA
12	Isocitrate	ICIT
10	D-Erythrose 4-phosphate	E4P
9	L-Aspartate	ASP
9	Butanoyl-CoA	BuCoA
8	Succinate	SUC

metabolites) with their corresponding reactions play a key role in metabolic regulation and may helpful to reveal the biological significance about *B. thuringiensis* metabolism. The average path length is 8.63 and network diameter is 24 for the GSC of *B.*
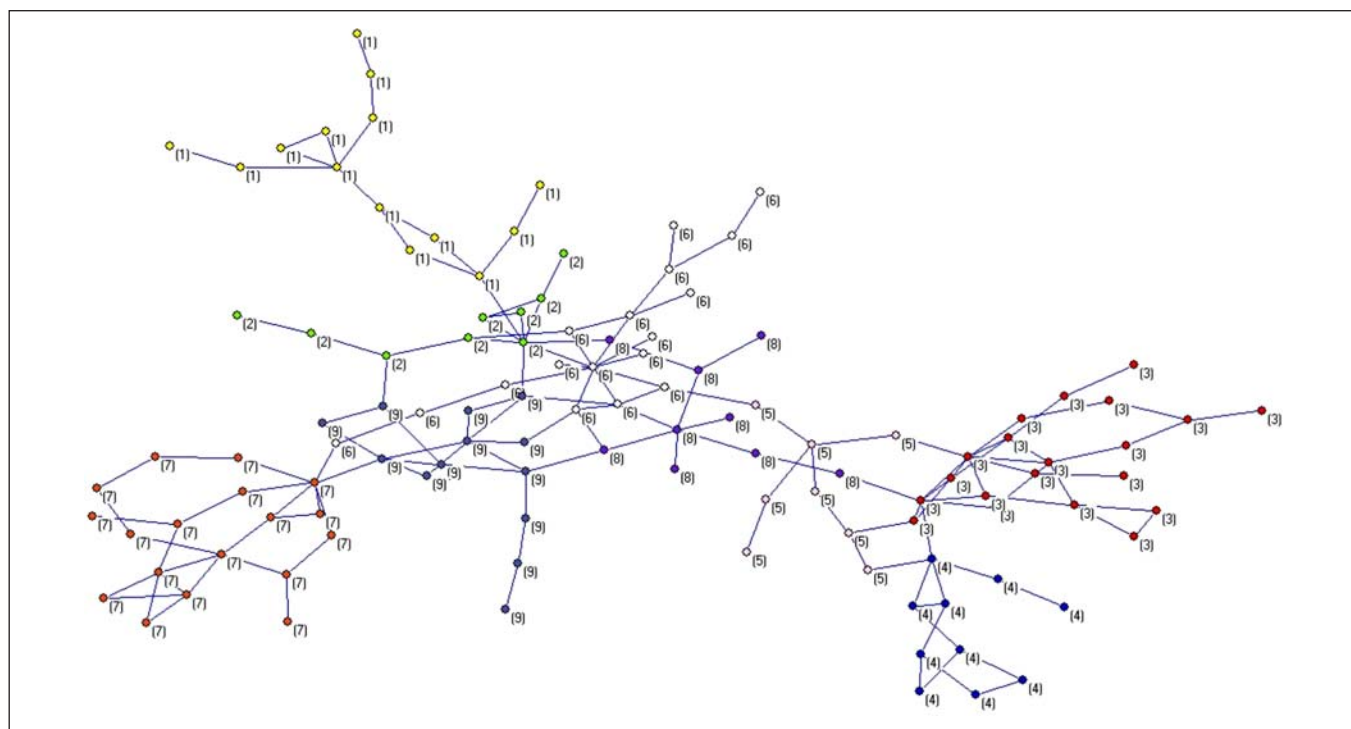
thuringiensis metabolic network, which is similar to other multi-bacteria via Pathway Hunter Tool (PHT) (30) and Ma and Zeng (25) (Table 4).

Modules of GSC

Various of decomposed results of the giant strong component of *B. thuringiensis* metabolic network based on simulated annealing algorithm are obtained due to different iteration factor (*f*) and cooling factor (c) as mentioned in section 2, at last we chosen the best decomposed result (Table 5, Fig. 3) after a number of computing. The result gives clearly partition with the number of metabolites, total links, within-module links and between-module links in each module and the modularity

Figure 2. Log-log plot of the degree distributions for the GSC of *B. thuringiensis* metabolic network.

Table 4. Average path length (AL) and diameter (D) of multi-bacteria.

Organisms	Abbreviation	AL	D
Bacillus subtilis	bsu	8.48	23
Escherichia coli	eco	8.16	23
Haemophilus influenzae	hin	8.35	27
Helicobacter pylori	hpj	7.91	24
Salmonella typhimurium	stm	8.22	24

Figure 3. Modules in the GSC of *B. thuringiensis* metabolic network, the picture was drawn using the Pajek program.

in the partition of the network is 0.752183. Then the decomposed result is also reaffirmed by compared to KEGG metabolic pathways, i.e. most modules are mainly corresponding to one or two KEGG pathways (Table 6). For instance, 11 of 12 within links in module 4 are corresponding to Glycerophospholipid metabolism and 11 of 15 within links in module 1 are corresponding to butanoate metabolism would demonstrate the anterior one. As for the latter one, 10 of 24 and 8 of 24 within links in module 7 are corresponding to arginine and proline metabolism, urea cycle and metabolism of amino groups, respectively.

Table 5. Decomposed results of the GSC of *B. thuringiensis* metabolic network.

Module	Nodes	Total links	Within links	Between links
1	14	16	15	1
2	9	16	10	6
3	20	32	28	4
4	10	14	12	2
5	8	11	7	4
6	17	27	19	8
7	18	26	24	2
8	9	13	8	5
9	13	22	16	6
Modularity	0.752183			

Table 6. The decomposed results of the GSC of *B. thuringiensis* metabolic network is reaffirmed by compared to KEGG metabolic pathways.

Module	Pathways in KEGG
1	butanoate metabolism
2	pyruvate metabolism
3	glycolysis, pentose phosphate pathway, carbon fixation
4	glycerophospholipid metabolism
5	——
6	pyruvate metabolism, several amino acid biosynthesis
7	arginine and proline metabolism, urea cycle and metabolism of amino groups
8	——
9	TCA cycle

—— represents that the corresponding module includes several pathways and it is difficult to assign it one or two simple pathways.

CONCLUSION

With the explosion of knowledge in 'X-mics' and systems biology, more and more genome-scale metabolic networks being reconstructed (8-10). To discover functional information involved in metabolic networks, a number of topological structural based approaches have already been developed, and it suggested that these computational modeling and analysis could contribute a lot to the understanding of the structure and function of these systems (1,3,5,16,23,26).

Taken together, this study provides an attempt at exploring the fundamental organizational principles that underlie *B. thuringiensis* metabolic network. We have initiated the study by integrating data from KEGG and correlative database, then the metabolic network reconstructed is represented by metabolite graph. Considering many isolated reactions are included in the whole metabolic networks, we extracted the most important part giant strong component (GSC) and analyzed its global structural properties and biological implication. We validated the "small-world" and "scale-free" characters and analyzed the first 10 hub metabolites of the GSC accordingly. Finally, the functional modules in GSC were studied with their biological significance.

RESUMO

Análise estrutural de funcional do GSC (Giant Strong Component) da rede metabólica de *Bacillus thurigiensis*

O objetivo deste trabalho foi realizar uma análise estrutural e funcional do GSC (Giant Strong Component) da rede metabólica de *Bacillus thurigiensis*. Baseando-se na estrutura *bow-tie*, o GSC foi extraído e analisado quanto ao sue significado funcional. Propriedades estruturais globais tais como grau de distribuição e tamanho médio da via metabólica foram mensuradas, concluindo-se que o GSC é também uma rede *small world* e *scale–free*. Além disso, a rede GSC foi decomposta e as divisões com significância funcional no metabolismo foram comparadas às vias metabólicas KEGG.

Palavras-chave: *Bacillus thurigiensis,* Giant Strong Component, rede metabólica

REFERENCES

1. Alon, U. (2003). Biological Networks: the Tinkerer as an Engineer. *Science* 301, 1866-1867.

2. Barabasi, A.L.; Albert, R. (1999). Emergence of scaling in random networks. *Science* 286, 509-512.

3. Barabasi, A.L.; Oltvai, Z.N. (2004). Network Biology: Understanding the Cell's Functional Organization. *Nat. Rev. Genet.* 5, 101-113.

4. Batagelj, V.; Mrvar, A. (1998). Pajek - program for large network analysis. *Connections* 21, 47-57.

5. Boccaletti, S.; Latora, V.; Moreno, Y.; Chavez, M.; Hwang, D.U. (2006). Complex networks: Structure and dynamics. *Physics Reports* 424, 175-308.

6. Csete, M.; Doyle, J. (2004). Bow ties, Metabolism and disease. *Trends Biotechnol* 22, 446-450.

7. Ding, D.W.; Ding, Y.R.; Cai, Y.J.; Chen, S.W.; Xu, W.B. (2008). Exploring poly-beta-hydroxy-butyrate metabolism through network-based extreme pathway analysis. *Riv. Biol. Biol. Forum* 101, 5-18.

8. Duarte, N.C.; Becker, S.A.; Jamshidi, N.; Thiele, I.; Mo, M.L.; Vo, T.D.; Srivas, R.; Palsson, B.O. (2007). Global reconstruction of the human metabolic network based on genomic and bibliomic data. *PNAS* 104: 1777-1782.

9. Duarte, N.C.; Herrgard, M.J.; Palsson, B.O. (2004). Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genomescale metabolic model. *Genome Res.* 14, 1298-1309.

10. Famili, I.; Forster, J.; Nielsen, J.; Palsson, B.O. (2003). *Saccharomyces cerevisiae* phenotypes can be predicted using constraint-based analysis of a genome-scale reconstructed metabolic network. *PNAS* 100, 13134-13139.

11. Gitahy, P.D.; de Souza, M.T.; Monnerat, R.G.; Arrigoni, P.D.; Baldani, J.I. (2007). A Brazilian Bacillus thuringiensis strain highly active to sugarcane borer Diatraea saccharalis (Lepidoptera : Crambidae). *Braz. J. Microbiol.* 38, 531-537.

12. Guimera, R.; Amaral, L.A.N. (2005). Functional cartography of complex metabolic networks. *Nature* 433, 895-900.

13. Guimera, R.; Amaral, L.A.N. (2005). Cartography of complex networks: modules and universal roles. *J. Stat. Mech. Theory Exp.*, P02001.

14. Guimera, R.; Sales-Pardo, M.; Amaral, L.A.N. (2004). Modularity from fluctuations in random graphs and complex networks. *Phys. Rev. E.* 70, 025101.

15. Herrgard, M.J.; Covert, M.W.; Palsson, B.O. (2004). Reconstruction of Microbial Transcriptional Regulatory Networks. *Curr. Opin. Biotechnol.* 15, 70-77.

16. Jeong, H.; Tombor, B.; Albert, R.; Oltvai, Z.N.; Barabasi, A.L. (2000). The Large-scale Organization of Metabolic Networks. *Nature* 407, 651-654.

17. Kanehisa, M.; Goto, S. (1998). KEGG: Kyoto encyclopedia of genes and genomes. *Nucl Acids Res* 28, 27-30.

18. Kirkpatrick, S.; Gelatt, C.D.; Vecchi, M.P. (1983). Optimization by simulated annealing. *Science* 220, 671-680.

19. Kitano, H. (2002). Computational systems biology. *Nature* 420, 206-210.

20. Kitano, H. (2004). Biological robustness. *Nat Rev Genet* 5, 826-837.

21. Kitano, H. (2007). A robustness-based approach to systems-oriented drug design. *Nature Rev. Drug Discov.* 6, 202-210.

22. Knaak, N.; Rohr, A.A.; Fiuza, L.M. (2007). In vitro effect of Bacillus thuringiensis strains and cry proteins in phytopathogenic fungi of paddy rice-field. *Braz. J. Microbiol.* 38, 526-530.

23. Lemke, N.; Heredia, F.; Barcellos, C.K.; dos Reis, A.N.; Mombach, J.C.M. (2004). Essentiality and damage in metabolic networks. *Bioinformatics* 20, 115-119.

24. Ma, H.W.; Zeng, A.P. (2003). The connectivity structure, giant strong component and centrality of metabolic networks. *Bioinformatics* 19, 1423-1430.

25. Ma, H.W.; Zeng, A.P. (2003). Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. *Bioinformatics* 19, 270-277.

26. Mombach, J.C.M.; Lemke, N.; da Silva, N.M.; Ferreira, R.A.; Isaia, E.; Barcellos, C.K. (2006). Bioinformatics analysis of mycoplasma metabolism: Important enzymes, metabolic similarities, and redundancy. Comput. *Biol. Med.* 36, 542-552.

27. Newman, M.E.J.; Girvan, M. (2004). Finding and evaluating community structure in networks. *Phys. Rev. E* 69, 026113.

28. Papin, J.A.; Hunter, T.; Palsson, B.O.; Subramaniam, S. (2005). Reconstruction of cellular signaling networks and analysis of their properties. *Nat. Rev. Mol. Cell Biol.* 6, 99-111.

29. Papin, J.A.; Reed, J.L.; Palsson, B.O. (2004). Hierarchical thinking in network biology: the unbiased modularization of biochemical networks. *Trends Biochem. Sci.* 29, 641-647.

30. Rahman, S.A.; Advani, P.; Schunk, R.; Schrader, R.; Schomburg, D. (2005). Metabolic pathway analysis web service (Pathway Hunter Tool at CUBIC). *Bioinformatics* 21, 1189-1193.

31. Ravasz, E.; Somera, A.L.; Mongru, D.A.; Oltvai, Z.N.; Barabasi, A.L. (2002). Hierarchical organization of modularity in metabolic networks. *Science* 297, 1551-1555.

32. Rives, A.W.; Galitski, T. (2003). Modular organization of cellular networks. *PNAS* 100, 1128-1133.

33. Schilling, C.H.; Letscher, D.; Palsson, B.O. (2000). Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J. Theor. Biol.* 203, 229-248.

34. Schuster, S.; Fell, D.A.; Dandekar, T. (2000). A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat. Biotechnol.* 18, 326-332.

35. Wagner, A.; Fell, D.A. (2001). The small world inside large metabolic networks. *Proc. R. Soc. Lond. B* 268, 1803-1810.

36. Watts, D.J.; Strogatz, S.H. (1998). Collective dynamics of 'small-world' networks. *Nature* 393, 440-442.

37. Zhao, J.; Tao, L.; Yu, H.; Luo, J.H.; Li, Y.X. (2007). Bow-tie topological features of metabolic networks and the functional significance. *Chin. Sci. Bull.* 52, 47-54.