International Journal of the Physical Sciences Vol. 5(11), pp. 1744-1751, 18 September, 2010
Available online at http://www.academicjournals.org/IJPS
ISSN 1992 - 1950 ©2010 Academic Journals

Full Length Research Paper

The structure and function of complex *Halobacterium salinarum* metabolic network

X. D. Liu[1]*, C. W. Wu[1] and Ding D. W.[2]*

[1]State Key Laboratory of Structural Analysis for Industrial Equipment, Faculty of Vehicle Engineering and Mechanics, Dalian University of Technology, Dalian 116023, China.
[2]Department of Mathematics and Computer Science, Chizhou College, Chizhou 247000, China.

Accepted 24 August, 2010

Genome-scale metabolic networks are widely used in industrial and medical biotechnology and complex networks methods are increasingly becoming important in the investigation of these models. The present paper performed a study of *Halobacterium salinarum* metabolic network by structural and functional analysis. First, we extracted high-quality *H. salinarum* metabolic network model from a recent reconstruction and based on its "bow tie" structure, we then extracted and studied the giant strong component (GSC) with its functional significance. Global structural properties such as average path length, degree distribution and self-similar exponent were computed and it indicated that the GSC is also a small-world, scale-free and self-similar network. Furthermore, the top 10 hub metabolites and functional modules in giant strong component were studied with their biological significance.

Key words: Average path length, complex networks, degree distribution, giant strong component, metabolic network, modularity, scale-free, small-world, systems biology.

INTRODUCTION

Remarkable advances in systems biology technologies (for example, genomics, proteomics, microarrays, etc.) enabled us to reconstruct more and more genome-scale biological networks (for example, metabolic, protein interaction, signaling networks, etc.) (Feist et al., 2009). Due to its completeness and reliability, the metabolic network plays a central role in biological networks research. Current research mainly include its reconstruction, structural and functional analysis, aid to industrial production, drug target discovery, etc (Feist et al., 2009; Barabasi and Oltvai, 2004; Albert, 2005; Aittokallio and Schwikowski, 2006; Ma and Goryanin, 2008; Song, 2009).

How to understand these large metabolic (or other biological) networks is a daunting challenge in the post-genomic era. As detailed kinetic parameters are hardly available, lots of recent studies have focused on structural and functional analysis of these networks (Barabasi and Oltvai, 2004; Albert, 2005; Aittokallio and Schwikowski, 2006; Ma and Goryanin, 2008). Structural-oriented methods such as complex networks analysis (Barabasi and Oltvai, 2004; Albert, 2005; Aittokallio and Schwikowski, 2006), Petri nets analysis (Chaouiya, 2007; Ding and Li, 2009) and stoichiometric network analysis (Palsson, 2006) have been well established gradually. Generally, we represented the metabolic networks by the so-called metabolite graph in complex networks analysis, that is, the nodes (metabolites) are linked by arcs or edges (metabolic reactions accordingly). Then, the fundamental organizational principles that underlie networks could be discovered based on global topological structural properties such as "small-world", "scale-free" and "self-similar". Furthermore, to discover the functional units involved in metabolic networks, it is suggested that metabolic networks should have modularity which is similar to other complex networks, such as social networks, internet, worldwide web, etc (Guimera and Amaral, 2005).

Halobacterium salinarum, an extremely halophilic marine gram-negative archaeon, has been reconstructed and discussed recently (for example, investigation of aerobic essential amino acid degradation) (Gonzalez et al., 2008). In the present article, firstly, we use the high

*Corresponding author. E-mail: dw.ding@hotmail.com, xdliu76@gmail.com.

quality metabolic network model of *H. salinarum* to generate a metabolite graph with 489 nodes and 803 links. Then, the structure of *H. salinarum* metabolic networks is explained and discussed based on "bow tie" structure, with emphasis on the giant strong component (GSC) part. At last, the functional significance, global structural properties and modularity of giant strong component in *H. salinarum* metabolic networks are studied.

MATERIALS AND METHODS

Metabolic networks and its bow tie structure

To investigate the topological properties of *H. salinarum* metabolism, we first obtain all metabolic reactions involved in the metabolic network of *H. salinarum* from a recent reconstruction (Gonzalez et al., 2008) and then, use the number of each metabolite corresponding to compounds in the KEGG LIGAND database. For instance, metabolite ID 22 corresponds to compound C00022 (pyruvate, PYR) in the KEGG database. Subsequently, to reflect biologically, meaningful transformations, all the reactions were revised by Ma and Zeng's database (Ma and Zeng, 2003). The advantages of their database are: (1) corrected obvious inconsistencies, (2) confirmation of the reversibility of every reaction and (3) exclusion of the current metabolites and small molecules (for example, ATP, ADP, NADH, etc). At last, the metabolic network that is reconstructed is represented by the so-called metabolite graph in which the nodes are metabolites and the links are reactions. For example, the irreversible reaction, 64 + 26 → 25 is represented by two directed arcs 64 → 25 and 26 → 25.

Since the "bow tie" structure of metabolic networks is proposed, it is increasingly recognized as being a conserved property of complex networks, as highlighted by recent studies and the results suggest that this structure property is functionally meaningful for metabolism, disease and the design principle of biological robustness (Zhao et al., 2007). Here, the so-called "bow tie" structure means that the network could be decomposed into four parts: (1) giant strong component (GSC), which is the biggest strongly connected component of the metabolic network. Note: the strongly connected component of a network is defined as the largest cluster of some nodes where any pair of nodes is mutually reachable; (2) substrate subset (S), which consists of the nodes that can go to the GSC, but cannot come from it; (3) product subset (P), which consists of the nodes that can come from the GSC, but can not go back to it and (4) isolated subset (IS), which contains some isolated nodes that can not come from the GSC and also can not go back to it (Zhao et al., 2007; Bondy and Murty, 1976; Ma and Zeng, 2003).

Global structural properties

It is suggested that average path length of metabolic networks is very small, showing itself as the property of "small-world". Another structural parameter is network diameter, which is defined as the path length of the longest pathway among all the shortest pathways (Barabasi and Oltvai, 2004). Furthermore, the direct reflection of difference among numerous metabolites in metabolic networks is the connection degree k, which is the link 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, that is, $P(k) \sim k^{-r}$ ($2<r<3$) (Barabasi and Oltvai, 2004), which means that most of the nodes in the network have a low degree, while a few nodes have a very high degree. In other words, metabolic network is a sort of typical "scale-free" network (Barabasi and Oltvai, 2004; Albert, 2005). The third structural property, "self-similarity", studied here, which relates to that of any part of the network looks like the entire property. Generally, the self-similarity of a network could be quantitatively characterized by its self-similar exponent (or fractal dimension) d_B, and d_B that could be calculated by a well-known box covering algorithm according to $N_B(\ell_B)/N \approx \ell_B^{-d_B}$, where: N_B is the minimum number of boxes covering the network fully, ℓ_B is the box size and N is the number of nodes in the network (Song et al., 2005; Ding and He, 2010).

Modularity and modules identification

Currently, there are too many methods used to module identification (da Silva, 2006; Fortunato, 2010), but the most important property 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 follows:

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

where r is the number of modules, l_s is the number of links between nodes in modules, d_s 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 would be widely used for identification of modules (Guimera and Amaral, 2005).

As simulated annealing is approved as a superexcellent method for modules identification in complex networks (especially in metabolic networks), thus we mainly engaged the method in this study (Guimera and Amaral, 2005). Here, simulated annealing 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 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:

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

where C_2 and C_1 are the cost after and before the update respectively, while T is computational temperature. Specifically, at each temperature T, there would be $n_i = fS^2$ nodes individual movements from one module to another and $n_c = fS$ nodes collective movements, where S is the number of nodes in the network and f is the recommended range of 0.1 to 1. Certainly, at each temperature T, the system would be cooled down to $T' = cT$.

RESULTS AND DISCUSSION

Metabolic network and its bow tie structure

The metabolite graph for the metabolic network of

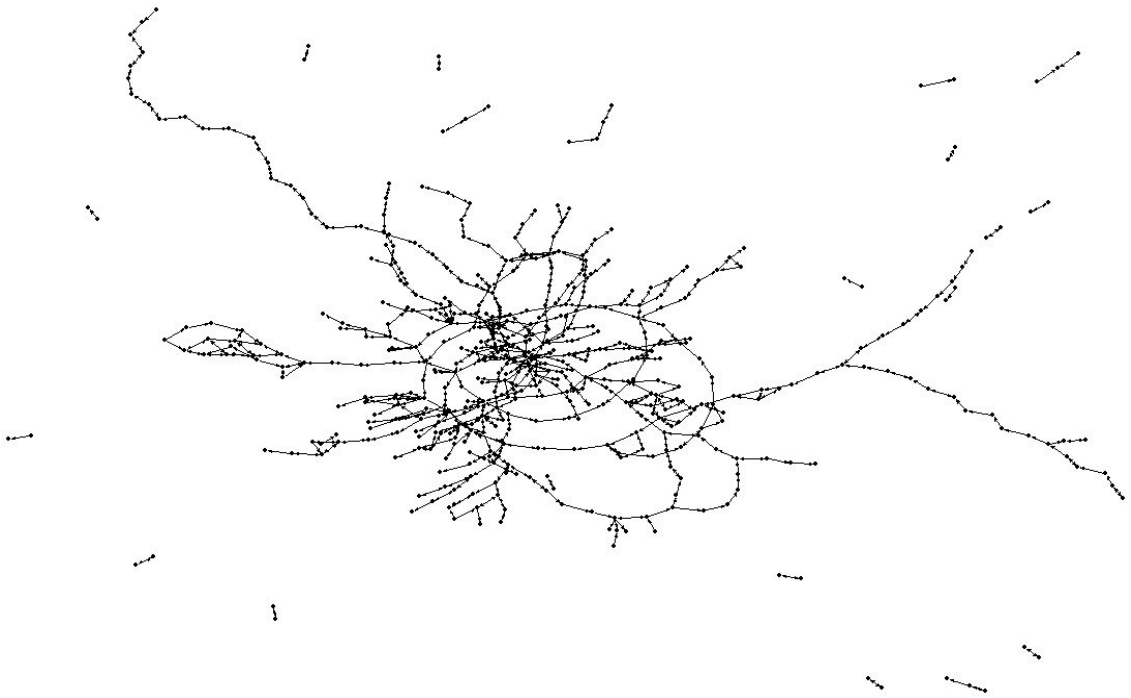

Figure 1. Metabolic network topology structure of *H. salinarum*. The nodes correspond to metabolites and the lines correspond to reactions. The picture was drawn using the Pajek program with Kamada-Kawai layout.

Table 1. The bow tie structure of *H. salinarum* metabolic network. Metabolites and reactions in giant strong component (GSC), substrate subset (S), product subset (P) and isolated subset (IS).

Subsets	GSC	S	P	IS	Total
No. of metabolites	126	47	191	125	489
Percentage of metabolites	25.8	9.6	39.1	25.5	100
No. of reactions	273	66	271	193	803
Percentage of reactions	34	8.2	33.7	24.1	100

H. salinarum is obtained based on the methods introduced in metabolic network and its bow tie structure (under materials and methods). The network contains 489 nodes and 803 links and the global topology structure is shown in Figure 1. It is clear that the whole network is far from strong component and included many isolated reactions. As a result, the whole metabolic network of *H. salinarum* is decomposed into four parts based on 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 involved herein, while the metabolites and reactions involved in the giant strong component part (the global topology structure is shown in Figure 2) are 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 (Schilling et al., 2000) and elementary modes (Schuster et al., 2000). Furthermore, this may be due to the fact that the giant strong component: (1) is the biggest strongly connected components of a metabolic network, (2) determines the structure of the entire network at a high degree, (3) plays an important role in metabolism, disease and biological robustness and (4) have been widely investigated for some other organisms (for example, human, *B. thuringiensis*, etc) (Ma and Goryanin, 2008; Zhao et al., 2007; Bondy and Murty, 1976; Ma and Zeng, 2003; Ma et al., 2007; Ding et al., 2009), thus a more detailed analysis of *H. salinarum* is given below.

All of the 273 metabolic reactions in the giant strong component are compared to KEGG pathways and it is shown that they are mainly concentrated on carbohydrate (50.2%) and amino acid (41.0%) metabolism (Table 2). The reactions of carbohydrate metabolism accurately correspond to glycolysis and TCA cycle, and partly correspond to pyruvate, propanoate and butanoate

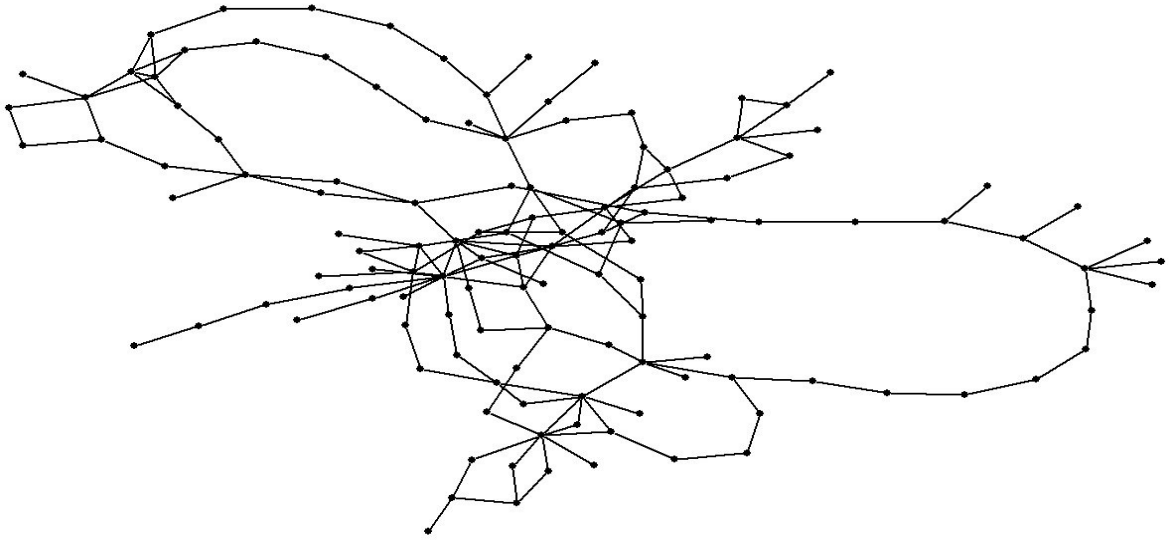

Figure 2. Giant strong component topology structure of *H. salinarum*. The nodes correspond to metabolites and the lines correspond to reactions. The picture was drawn using the Pajek program with Kamada- Kawai layout.

Table 2. Reactions in giant strong component (GSC) of *H. salinarum* metabolic network.

Reactions in GSC	No. of reactions	Percentage of reactions
Carbohydrate metabolism	137	50.2
Amino acid metabolism	112	41.0
Lipid metabolism	5	1.8
Others	19	7.0
Total	273	100

Table 3. Average path length (AL) and diameter (D) of multi-organism.

Organisms	Abbreviation	AL	D
Escherichia coli	eco	8.16	23
Haemophilus influenzae	hin	8.35	27
Saccharomyces cerevisiae	sce	9.71	31
Rattus norvegicus	rno	10.99	38
Homo sapiens	hsa	11.33	46
Caenorhabditis elegans	cel	10.87	49

metabolism. From the point of view of topological network, the results show that metabolites in carbohydrate metabolism (in particular glycolysis and TCA cycle, important part of the central metabolism) have the highest probability of 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 glycine, serine and threonine metabolism; valine, leucine and isoleucine degradation; phenylalanine, tyrosine and tryptophan biosynthesis, these might reveal the nutrient requirement in *H. salinarum.*

Global structural properties

The average path length is 10.71 and the network diameter is 35 for the giant strong component of *H. salinarum* metabolic network, which is similar to other

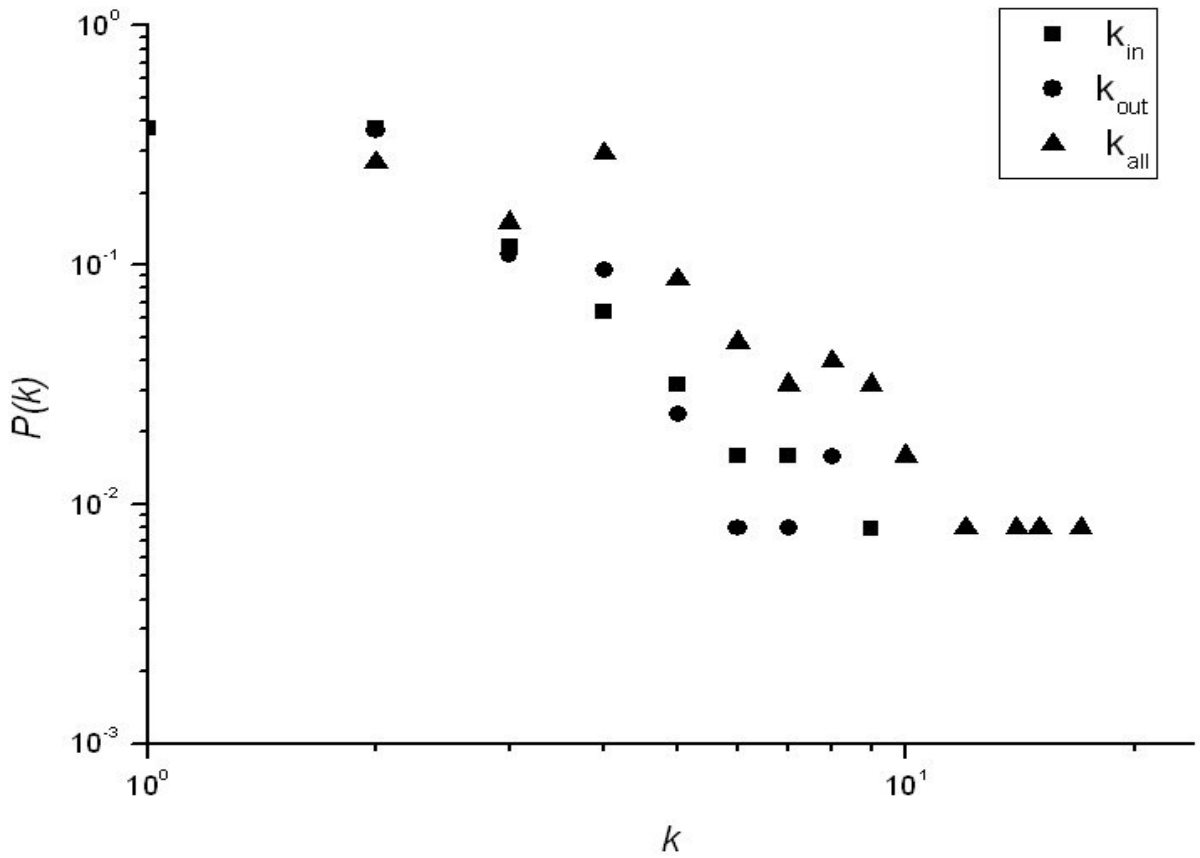

Figure 3. Log-log plot of the degree distributions for the giant strong component of *H. salinarum* metabolic network.

Table 4. The top 10 hub metabolites of the giant strong component of *H. salinarum* metabolic network.

Degree	ID	Metabolite name	Abbreviation
17	22	Pyruvate	PYR
15	111	Glycerone phosphate	GlyP
14	118	(2R)-2-Hydroxy-3-(phosphonooxy)-propanal	2HPP
12	311	Isocitrate	ICIT
10	441	L-Aspartate	ASP
10	579	Dihydrolipoamide	DIHY
9	25	L-Glutamate	GLU
9	251	Chorismate	CHOR
9	158	Citrate	CIT
9	877	Crotonoyl-CoA	CrCoA

multi-organism via Ma and Zeng (2003) (Table 3). As such, we then checked the scale-free property of the giant strong component of *H. salinarum* metabolic network (Figure 3). As it is known, the nodes with high degree of scale-free network would dominate the network structure and make the network robust against random errors such as mutation and environmental changes. We identified 10 primary metabolites with the highest degree for *H. salinarum* metabolic network (Table 4). Among these top 10 central metabolites, PYR, DIHY and ICIT are important intermediates in the glycolysis pathway, while GlyP plays a key role in glycolysis pathway, fructose and mannose metabolism, glycerophospholipid metabolism, carbon fixation, nicotinate and nicotinamide metabolism. 2HPP is the metabolite linking glycolysis pathway, pentose phosphate pathway and carbon fixation, whereas

Table 5. Decomposed results of the giant strong component of *H. salinarum* metabolic network based on simulated annealing algorithm.

Module	Nodes	Total links	Within links	Between links
1	13	17	12	5
2	14	23	17	6
3	18	33	23	10
4	18	27	22	5
5	16	25	19	6
6	16	19	15	4
7	17	27	23	4
8	14	17	13	4
Modularity	0.736192			

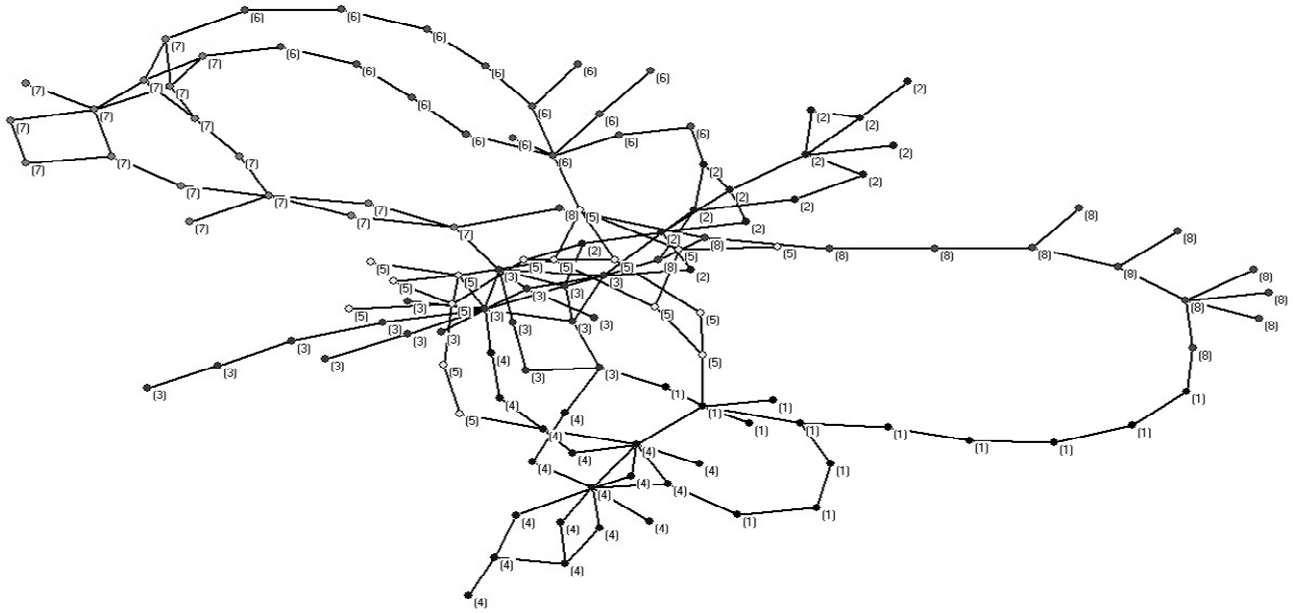

Figure 4. Modules in the giant strong component of *H. salinarum* metabolic network. The picture was drawn using the Pajek program with Kamada- Kawai layout (Notes: each module is signed by its module No.; the No. is also used in Tables 5 and 6).

ASP and GLU are two important amino acids that are directly produced in TCA cycle and could be converted to many other useful amino acids. In a like manner, CHOR links folate biosynthesis, biosynthesis of siderophore group, nonribosomal peptides and phenylalanine, tyrosine and tryptophan biosynthesis, whereas CIT is an important intermediate in TCA cycle, while CrCoA is an important intermediate in the butanoate metabolism and it links fatty acid metabolism and benzoate degradation via CoA ligation. As links among different functional metabolic pathways, these hub metabolites with their corresponding reactions play a key role in metabolic regulation and may be helpful in revealing the biological significance of *H. salinarum* metabolism. At last, according to the box covering algorithm, we get the relation between the box size ℓ_B and the corresponding number of boxes N_B for the GSC of *H. salinarum* metabolism. Following the above expression, the self-similar exponent for it is 1.64. The result shows that the GSC of *H. salinarum* metabolism is also self-similar, but the self-similar exponent is remarkably lower than the entire metabolism (average of 3.5 for 43 metabolic networks in Song et al. (2005) and Ding and He (2010).

Modularity and modules identification

Several decomposed results of the giant strong component of *H. salinarum* metabolic network based on simulated annealing algorithm are obtained due to different iteration factor (*f*) and cooling factor (c) as mentioned in 'modularity and modules identification under materials and methods'. At last, we chose the best decomposed result (Table 5 and Figure 4) after a number

Table 6. The decomposed results of the giant strong component of *H. salinarum* metabolic network reaffirmed by comparison to KEGG metabolic pathways.

Module	Pathways in KEGG
1	Phenylalanine, tyrosine and tryptophan biosynthesis
2	——
3	Pyruvate metabolism and citrate cycle (TCA cycle)
4	Glycolysis/gluconeogenesis and glycerolipid metabolism
5	Glycine, serine and threonine metabolism
6	Valine, leucine and isoleucine degradation and propanoate metabolism
7	Glycolysis/gluconeogenesis and butanoate metabolism
8	Tyrosine metabolism and phenylalanine; tyrosine and tryptophan biosynthesis

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

of computing. The result shows clearly, a partition with the number of metabolites, total links, within-module links and between-module links in each module and the modularity in the partition of the network is 0.736192. Then the decomposed result is also reaffirmed by a comparison of KEGG metabolic pathways, that is, most modules mainly correspond to one or two KEGG pathways (Table 6). For instance, module 1 corresponds to phenylalanine, tyrosine and tryptophan biosynthesis which demonstrated the anterior one, while module 3 corresponds to pyruvate metabolism and citrate cycle (TCA cycle) which demonstrated the latter one.

Conclusions

As the knowledge of interactions between biological molecules has been accumulated rapidly, more and more genome-scale metabolic networks are being reconstructed (Feist et al., 2009). Due to the absence of detailed kinetic parameters, a number of topological structural based approaches have already been developed to discover functional information involved in metabolic networks, and as such, the study suggested that these computational modeling and analysis could contribute a lot to the understanding of the structure and function of these networks (Barabasi and Oltvai, 2004; Albert, 2005; Aittokallio and Schwikowski, 2006; Ma and Goryanin, 2008; Song, 2009; Chaouiya, 2007; Ding and Li, 2009; Palsson, 2006; Guimera and Amaral, 2005; Ma and Zeng, 2003; Zhao et al., 2007; Ma and Zeng, 2003; Song et al., 2005; Ding and He, 2010; da Silva, 2006; Ma et al., 2007; Ding et al., 2009).

Taken together, this study provides an attempt at exploring the fundamental organizational principles that underlie *H. salinarum* metabolic network. We have initiated the study by integrating data from a recent reconstruction of *H. salinarum* and then represented the model by a metabolite graph. Considering the many isolated reactions included in the whole metabolic network, we extracted the most important part, which is the giant strong component and analyzed its global structural properties and biological implication. We validated the “small-world”, “scale-free” and “self-similar” characters and analyzed the first 10 hub metabolites of the giant strong component accordingly. Finally, the functional modules in the giant strong component were studied with their biological significance.

ACKNOWLEDGEMENTS

The authors would like to thank the anonymous reviewers for their valuable comments on the manuscript. Support for this work is provided by fund from Natural Science Fund of Anhui Education Department (KJ2010B133).

REFERENCES

Aittokallio T, Schwikowski B (2006). Graph-based methods for analysing networks in cell biology. Brief Bioinform., 7: 243-255.
Albert R (2005). Scale-free networks in cell biology. J. Cell Sci., 118: 4947-4957.
Barabasi AL, Oltvai ZN (2004). Network biology: understanding the cell’s functional organization. Nature Reviews Genetics, 5: 101-113.
Bondy JA, Murty USR (1976). Graph theory with applications. London: Macmillan.
Chaouiya C (2007). Petri net modelling of biological networks. Brief Bioinform. 8: 210-219.
da Silva MR (2006). Bioinformatics tools for the visualization and structural analysis of metabolic networks. Ph.D. thesis, Technische Universitat Carolo-Wilhelmina zu Braunschweig, Germany.
Ding DW, Ding YR, Li LN, Chen SW, Cai YJ, Xu WB (2009). Structural and functional analysis of giant strong component of Bacillus thuringiensis metabolic network. Brazilian J. Microbiol. 40: 411-416.
Ding DW, He XR (2010). The self-similarity of trunk metabolism. Rivista di Biologia / Biology Forum, 103:18-21.
Ding DW, Li LN (2009). Modeling and analyzing the metabolism of riboflavin production using Petri nets. J. Biological Syst., 17: 479-490.
Feist AM, Herrgard MJ, Thiele I, Reed JL, Palsson BO (2009). Reconstruction of biochemical networks in microorganisms. Nature Rev. Microbiol., 7: 129-143.
Fortunato S (2010). Community detection in graphs. arXiv:0906.0612 v2. http://arxiv.org/abs/0906.0612.
Gonzalez O, Gronau S, Falb M, Pfeiffer F, Mendoza E, Zimmer R,

Oesterhelt D (2008). Reconstruction, modeling and analysis of Halobacterium salinarum R-1 metabolism. Mol Biosyst., 4: 148-159.
Guimera R, Amaral LAN (2005). Functional cartography of complex metabolic networks. Nature, 7: 895-900.
Ma HW, Goryanin I (2008). Human metabolic network reconstruction and its impact on drug discovery and development. Drug Discov. Today, 13: 402-408.
Ma HW, Sorokin A, Mazein A, Selkov A, Selkov E, Demin O, Goryanin I (2007). The Edinburgh human metabolic network reconstruction and its functional analysis. Mol. Syst. Biol., 3: 135.
Ma HW, Zeng AP (2003). Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. Bioinformatics, 19: 270-277.
Ma HW, Zeng AP (2003). The connectivity structure, giant strong component and centrality of metabolic networks. Bioinformatics, 19: 1423-1430.
Palsson BO (2006). Systems Biology: Properties of Reconstructed Networks. Cambridge University Press, London.
Schilling CH, Letscher D, Palsson BO (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.
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.
Song B, Sridhar P, Kahveci T, Ranka S (2009). Double iterative optimisation for metabolic network-based drug target identification. Int. J. Data Mining Bioinformatics, 3: 124-144.
Song C, Havlin S, Makse HA (2005). Self-similarity of complex networks. Nature, 433: 392-395.
Zhao J, Tao L, Yu H, Luo JH, Cao ZW, Li YX (2007). Bow-tie topological features of metabolic networks and the functional significance. Chinese Sci. Bull., 52: 47-54.