PRAMANA
— journal of
physics

Vol. 80, No. 2
February 2013
pp. 337–348

Complex network perspective on structure and function of *Staphylococcus aureus* metabolic network

L YING[1] and D W DING[2,*]

[1]Education Department, National University of Defense Technology, Changsha 410073, China
[2]Department of Mathematics and Computer Science, Chizhou College, Chizhou 247000, China
*Corresponding author. E-mail: dw.ding@hotmail.com

MS received 4 May 2012; revised 22 June 2012; accepted 24 July 2012

Abstract. With remarkable advances in reconstruction of genome-scale metabolic networks, uncovering complex network structure and function from these networks is becoming one of the most important topics in system biology. This work aims at studying the structure and function of *Staphylococcus aureus* (*S. aureus*) metabolic network by complex network methods. We first generated a metabolite graph from the recently reconstructed high-quality *S. aureus* metabolic network model. Then, based on 'bow tie' structure character, we explain and discuss the global structure of *S. aureus* metabolic network. The functional significance, global structural properties, modularity and centrality analysis of giant strong component in *S. aureus* metabolic networks are studied.

Keywords. Centrality analysis; metabolic network; modularity analysis; systems biology.

PACS Nos 89.75.Hc; 02.50.−r; 05.90.+m

1. Introduction

With the advent of whole genome sequencing and high-throughput approaches (e.g., genomics, proteomics, etc.), the full components and their interactions in biological systems could be well-characterized. Reconstruction of comprehensive cellular networks has become a major area of systems biology research. Examples of such networks are gene regulatory networks, protein interaction networks, signalling networks and (in particular) metabolic networks [1,2].

To date, many reconstructed metabolic networks are available. Generally, there are many metabolites and metabolic reactions in reconstructed metabolic networks. Hence the number of traditional metabolic engineering methods to understand and interpret these large networks are very limited. However, progress in this area is aided by the rapidly developing complex networks. Following the work initiated by Jeong *et al* [3], the application of complex network to cellular networks has been greatly accelerated over the

last decade. Results suggest that these methods are invaluable in understanding cellular organizational principles, as well as for proposing new hypotheses [4–7].

We applied their methods to investigate the structure and function of *Staphylococcus aureus* (*S. aureus*, an important pathogen) metabolic network in the present paper. We first generated a metabolite graph with 855 nodes and 1353 links from a recent reconstructed high-quality *S. aureus* metabolic network model [8]. Then, based on the 'bow tie' structure character, we explained and discussed the global structure of *S. aureus* metabolic network. Finally, the functional significance, global structural properties, modularity and centrality analysis of giant strong components in *S. aureus* metabolic networks were studied.

2. Materials and methods

2.1 *Construction of metabolite graph*

To better understand the topological properties of *S. aureus* metabolic network, we first obtained a recently reconstructed high-quality *S. aureus* metabolic network model [8], and used a number of each metabolite instead of compounds in the KEGG LIGAND database. For instance, we used the metabolite 24 instead of the compound C00024 (acetyl-CoA) in the KEGG database. Subsequently, all the reactions were revised using Ma and Zeng's database [9], since their database: (1) confirmed the reversibility of every reaction and (2) excluded the current metabolites and small molecules such as ATP, ADP, NADH, H_2O, etc., with the purpose of reflecting biologically meaningful transformations. Finally, the metabolic network 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, $64 + 26 \rightarrow 25$ is represented by two directed links $64 \rightarrow 25$ and $26 \rightarrow 25$.

2.2 *Bow tie structure*

Since Ma and Zeng [10] proposed the 'bow tie' structure of metabolic networks, it has been 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 [11,12].

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). Here, the GSC is the biggest strongly connected component of a network and a strongly connected component is defined as a subgraph of a network in which any pair of nodes is mutually reachable [11].

2.3 *Average path length, small-worldness and degree distribution*

Average path length is the most basic and important network measure. Generally speaking, average path length is defined as the average of the shortest paths between all pairs of nodes, and the shortest path is defined as the path with the smallest number of links

between two nodes. Another structure parameter is the network diameter, which is defined as the path length of the longest pathway among all of the short pathways [4].

Humphries and Gurney [13] defined the small-worldness S^{Δ} as follows:

$$S^{\Delta} = \frac{C/L}{C_{\mathrm{ER}}/L_{\mathrm{ER}}}, \tag{1}$$

where C is the clustering coefficient, L is the average path length, C_{ER} and L_{ER} are the corresponding measures in an equal size Erdös–Rényis random network.

It is suggested that if the average path length is very small and the small-worldness S^{Δ} is larger than 1, the network will have the 'small-world' property [13].

On the other hand, the direct reflection of difference among numerous metabolites in metabolic networks is the connection degree k, which is the link that the node has with others, and the degree distribution $P(k)$ gives the probability of a node with degree k. One of the most important properties of the metabolic networks is the power law degree distribution, i.e. $P(k) \sim k^{-r}$ ($r \approx 2.2$), which means that most of the nodes in the network have a low degree, while a few nodes have a very high degree. In other words, metabolic network is a kind of typical scale-free network [4].

2.4 *Modularity analysis*

The basic principle for defining functional modules in biology networks is similar to that of community social networks, which are dense node–node links within modules, but have sparser links between them [14]. An important measure related to the 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:

$$M \equiv \sum_{s=1}^{r} \left[\frac{l_{\mathrm{s}}}{L} - \left(\frac{d_{\mathrm{s}}}{2L} \right)^2 \right], \tag{2}$$

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 M would yield the most accurate results for real-world complex networks, and thus is widely used for identifying modules in networks [13].

Guimera and Amaral [15] introduced a simulated annealing-based method to find the optimal partitions of modules by maximizing the network modularity. In their method, some random updates are performed and accepted with probability P:

$$p = \begin{cases} 1 & \text{if} \quad C_2 \le C_1 \\ \exp\left(-\dfrac{C_2 - C_1}{T} \right) & \text{if} \quad C_2 \ge C_1 \end{cases} \tag{3}$$

where C_2 and C_1 are respectively the cost after the update and before the update, while T is the computational temperature. Specifically, at each temperature T, there would be $n_{\mathrm{i}} (= f S^2$ nodes) individual movements from one module to another and $n_{\mathrm{c}} (= f S$ 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 a certain temperature T, the system would be cooled down to $T' = cT$.

Table 1. Definitions for the centrality measures used in this study. $d(v)$ denotes the degree of the vertex v, dist(v, w) denotes the length of the shortest path between the vertices v and w, $\sigma_{st}(v)$ denotes the number of the shortest path from s to t that use the vertex v, $\delta_{st}(v) = \sigma_{st}(v)/\sigma_{st}$, where σ_{st} denotes the number of the shortest paths from s to t, A denotes the adjacency matrix of the graph.

Name	Definition
Degree	$C_{deg}(v) = d(v)$
Eccentricity	$C_{ecc}(v) = 1/(\max_{w\in V} \text{dist}(v, w))$
Closeness	$C_{clo}(v) = 1/(\sum_{w\in V} \text{dist}(v, w))$
Radiality	$C_{rad}(v) = \sum_{w\in V}(\Delta_G + 1 - \text{dist}(v, w))/(n-1)$
Centroid value	$C_{cen}(v) = \min_{w\in V\setminus\{v\}}\{f(v, w)\}$
Shortest path betweenness	$C_{spb}(v) = \sum_{S\neq v\in V}\sum_{t\neq v\in V}\delta_{st}(v)$
Katz status	$C_{katz} = \sum a^k (A^T)^k \hat{1}$
Eigenvector	$\lambda C_{eig} = AC_{eig}$
Page rank	$C_{pr} = dFC_{pr} + (1-d)\hat{1}$
HITS-Hubs	$C_{hubs} = AC_{auths}$

2.5 *Centrality analysis*

Generally speaking, centrality is a function that assigns a numerical value $C(v)$ to each vertex of a network, and there are many different measures for computing such a centrality. The simplest measure degree centrality is used to show the number of connections of each vertex in the network, and thus is used to identify the hub metabolites. Betweenness centrality is corresponding to the number of shortest pathways going through the vertices, while closeness centrality is helpful to identify vertices in the core and periphery part of the network, and so on. Table 1 summarizes the centrality measures used in this study [16].

3. Results and discussion

3.1 *S. aureus metabolic network*

The metabolic network of *S. aureus* is reconstructed based on the methods which are introduced in §2.1. The network contains 855 nodes and 1353 links, and the global topology structure is shown in figure 1. It is clear that the whole network includes many isolated reactions.

3.2 *Bow tie structure*

The whole metabolic network of *S. aureus* is then decomposed into four parts based on the 'bow tie' structure (figure 2, table 2). It should be noted that most nodes in S, P and IS parts are connected by some single link which are not interested herein, while the metabolites and reactions involved in the giant strong component part are clearly much

Figure 1. Metabolic network topology structure of *S. aureus*. The nodes correspond to metabolites and the links correspond to reactions. The picture was drawn using the Pajek program with Kamada–Kawai lay-out.

less than the whole network, and would be used to reduce the complexity of applying other pathway analysis methods such as extreme pathways [17,18] and elementary modes [19]. Furthermore, as the giant strong component is the biggest strongly connected component of a metabolic network and determined structure of the entire network at a certain extent

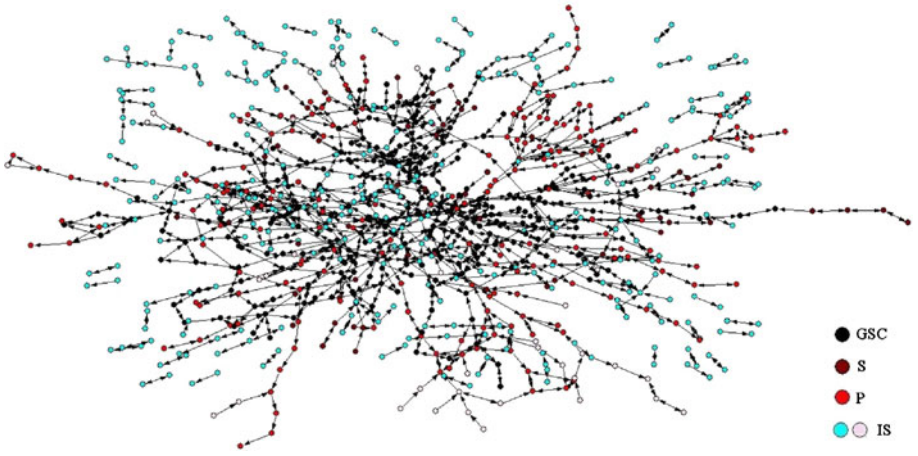

Figure 2. Bow tie structure of *S. aureus* network.

Table 2. The bow tie structure of *S. aureus* 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	250	61	188	356	855
Percentage of metabolites	29.2	7.1	22	41.6	100
No. of reactions	560	42	249	502	1353
Percentage of reactions	41.4	3.1	18.4	37.1	100

[11,12], a more detailed analysis is given below. The network contains 250 nodes and 560 links, and the global topology structure is shown in figure 3.

All of the 560 metabolic reactions are compared to KEGG pathways, and it is shown that they are mainly involved in carbohydrate metabolism (37.9%) and amino acid metabolism (33.6%) (table 3). The reactions of carbohydrate metabolism accurately correspond to glycolysis, TCA cycle, pentose phosphate pathway, and partly correspond to pyruvate metabolism, fructose and mannose metabolism, peptidoglycan biosynthesis, butanoate metabolism, glyoxylate and dicarboxylate metabolism, starch and sucrose metabolism. From the network topological point, 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 many more links and stronger robustness in network, and thus might possess higher attack tolerance despite external cues, genetic variation and stochastic noise. While reactions of amino acid metabolism are mainly involved in histidine metabolism, glycine, serine and threonine metabolism, phenylalanine, tyrosine and tryptophan biosynthesis, urea cycle

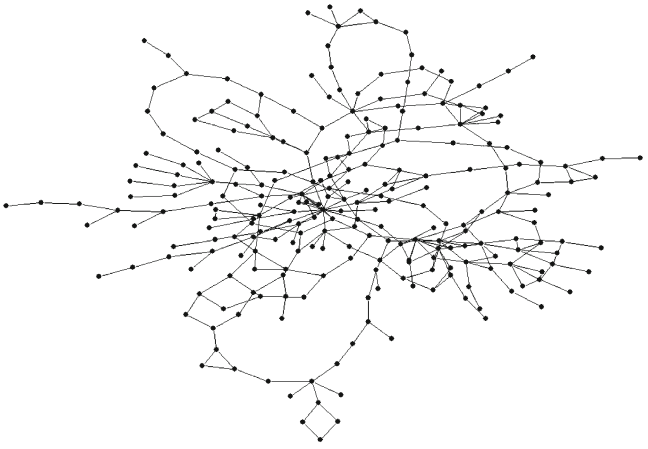

Figure 3. Giant strong component topology structure of *S. aureus*. The nodes correspond to metabolites and the links correspond to reactions. The picture was drawn using the Pajek program with Kamada–Kawai lay-out.

Table 3. Reactions in giant strong component (GSC) of *S. aureus* metabolic network.

Reactions in GSC	No. of reactions	Percentage of reactions
Carbohydrate metabolism	212	37.9
Energy metabolism	23	4.1
Lipid metabolism	36	6.4
Nucleotide metabolism	44	7.9
Amino acid metabolism	188	33.6
Others	57	10.2
Total	560	100

and metabolism of amino groups, lysine biosynthesis, arginine and proline metabolism, alanine and aspartate metabolism, cysteine metabolism, valine, leucine and isoleucine degradation, valine, leucine and isoleucine biosynthesis, these might reveal the nutrient requirement in *S. aureus.*

3.3 *Average path length, small-worldness and degree distribution*

Herein, we checked these two properties of the giant strong component in *S. aureus* metabolic networks. We firstly computed the average path length and network diameter. The average path length is 10.72 steps and network diameter is 35 steps for the giant strong component of *S. aureus* metabolic network, which is similar to other multibacteria (Ma and Zeng [9]) (table 4). The small-world S^{Δ} is 7.72. These results show itself the property of 'small-worldness'.

According to scientific literature and practical experience, if the network degree distributions follow power law, the network is 'scale-free'. We have then investigated the in-degree (the number of directed links that point to the node) distributions, out-degree (the number of directed links that start at the node) distributions and total-degree (the number of total links, i.e. the summation of in-degree and out-degree) distributions of the giant strong component in *S. aureus* metabolic network. The results show that all the 3-degree distributions approximately follow the power law (figure 4), i.e., the network possesses the 'scale-free' property.

Table 4. Average path length (AL) and diameter (D) of multiorganisms.

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

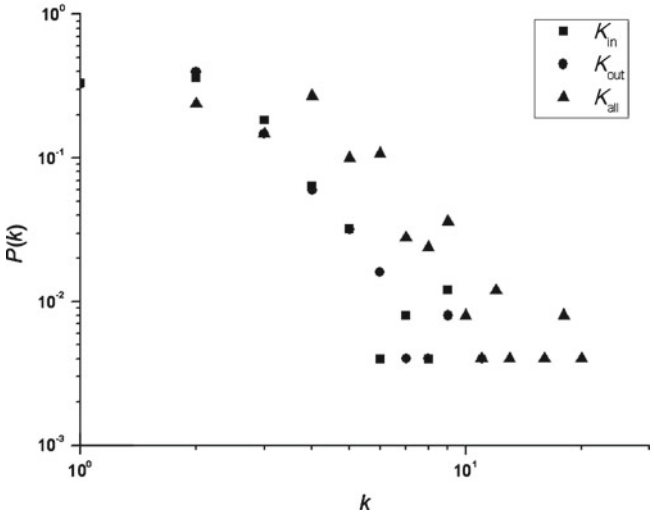

Figure 4. Log–log plot of degree distributions for the giant strong component of *S. aureus* metabolic network.

3.4 *Modularity analysis*

Various decomposed results of the giant strong component of *S. aureus* metabolic network based on simulated annealing algorithm are obtained due to different iteration factors (f) and cooling factors (c) as mentioned in §2. Finally, we selected the best decomposed result (table 5, figure 5) after a number of computings. The result gives a clear

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

Module	Nodes	Total links	Within links	Between links
1	20	27	19	8
2	17	26	20	6
3	10	10	9	1
4	19	23	20	3
5	24	45	33	12
6	24	40	33	7
7	19	24	21	3
8	21	46	30	16
9	23	36	28	8
10	29	42	36	6
11	14	17	14	3
12	17	21	17	4
13	13	13	12	1
Modularity	0.793489			

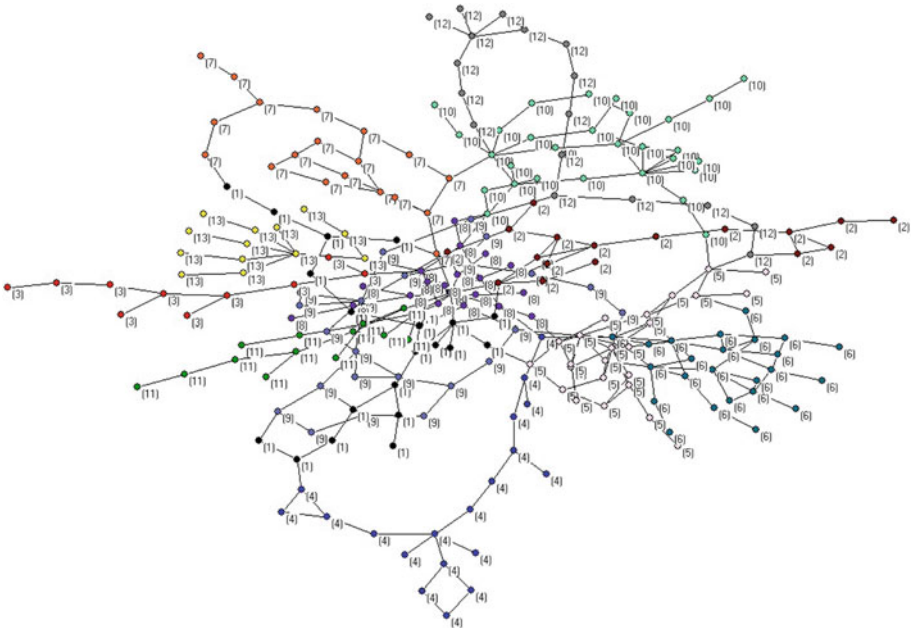

Figure 5. Modules in the giant strong component of *S. aureus* metabolic network, the picture was drawn using the Pajek program with Kamada–Kawai lay-out. (Note: Each module is assigned a module No., which is also used in tables 5 and 6.)

Table 6. Decomposed results of the giant strong component of *S. aureus* metabolic network is reaffirmed by comparing with KEGG metabolic pathways. – indicates that the corresponding module includes several pathways and it is difficult to assign it one or two simple pathways.

Module	Pathways in KEGG
1	KEGG MAP00260, KEGG MAP00630
2	KEGG MAP00271, KEGG MAP00230
3	KEGG MAP00062, KEGG MAP00650
4	KEGG MAP00400, KEGG MAP00790
5	KEGG MAP00561, KEGG MAP00010
6	KEGG MAP00010, KEGG MAP00030
7	KEGG MAP00550, KEGG MAP00300
8	KEGG MAP00620
9	KEGG MAP00230
10	KEGG MAP00220, KEGG MAP00330
11	–
12	KEGG MAP00340
13	KEGG MAP00120, KEGG MAP00280

partition with a 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 0.79. Then the decomposed result is also reaffirmed by comparing with KEGG metabolic pathways, i.e. most modules mainly correspond to one or two KEGG pathways (table 6). For instance, module 1 corresponds to glycine, serine and threonine metabolism (KEGG MAP00260) and glyoxylate and dicarboxylate metabolism (KEGG MAP00630), module 2 corresponds to methionine metabolism (KEGG MAP00271) and purine metabolism (KEGG MAP00230), ..., module 8 mainly corresponds to pyruvate metabolism (KEGG MAP00620), etc.

3.5 *Centrality analysis*

The multicentrality measures of the giant strong component in *S. aureus* metabolic networks are then computed. As different central metabolites correspond to different centrality measures (see metabolite indices in table 7), we have ranked top 10 central metabolites according to the number of centrality measures in which the metabolite is a central metabolite (table 8). Among these top 10 centre metabolites, GLU and ASP are the two important amino acids, 3PG, PYR and ICIT are the important intermediates in the

Table 7. The top 20 central metabolite indices corresponding to different centrality measures (Degree, deg; Eccentricity, ecc; Closeness, clo; Radiality, rad; Centroid value, cen; Shortest path betweenness, spb; Katz status, katz; Eigenvector, eig; PageRank, pr; HITS-Hubs, hubs).

Rank	V_{deg}	V_{ecc}	V_{clo}	V_{rad}	V_{cen}	V_{spb}	V_{katz}	V_{eig}	V_{pr}	V_{hubs}
1	22	36	22	22	22	22	118	118	24	118
2	111	22	36	36	36	84	111	111	251	111
3	118	49	41	41	84	118	5345	5345	62	5345
4	5345	149	149	149	673	673	22	5378	100	5378
5	25	898	65	65	118	24	24	354	111	279
6	24	41	74	74	24	36	279	279	118	85
7	62	65	49	49	49	25	84	231	5345	354
8	100	74	97	97	631	26	25	3785	5897	231
9	85	97	118	118	197	188	354	85	348	3785
10	279	122	6010	6010	149	133	231	447	22	447
11	900	186	900	900	258	37	197	1094	84	1094
12	26	546	631	631	62	311	5378	197	542	197
13	31	900	546	546	26	48	85	236	42	236
14	36	1352	168	168	168	158	900	673	944	5382
15	65	6010	186	186	1682	74	100	5382	311	673
16	84	42	133	133	1201	41	92	275	25	184
17	92	133	673	673	74	217	158	184	77	93
18	135	168	258	258	231	231	668	644	122	6893
19	197	231	898	898	301	160	447	93	4882	668
20	251	424	197	197	25	266	227	6893	158	275

Table 8. The top 10 central metabolites ranked by the number of centrality measures in which the metabolite is a centre metabolite.

Rank	Vertex	Metabolite name	Abbreviation	Number
1	118	(2R)-2-Hydroxy-3-(phosphonooxy)-propanal	2HPP	9
2	22	Pyruvate	PYR	8
3	197	3-Phospho-D-glycerate	3PG	7
4	36	Oxaloacetate	OAA	6
4	231	D-Xylulose 5-phosphate	Xu5P-D	6
4	673	2-Deoxy-D-ribose 5-phosphate	2Dr5P	6
5	24	Acetyl-CoA	AcCoA	5
5	25	L-Glutamate	GLU	5
5	74	Phosphoenolpyruvate	PEP	5
5	84	Acetaldehyde	Acald	5
5	111	Glycerone phosphate	GlyP	5
5	900	2-Acetolactate	Alac	5
5	5345	Beta-D-Fructose 6-phosphate	F6P	5

glycolysis pathway, OAA-linked pyruvate metabolism and TCA cycle, SUC-linked butanoate metabolism and TCA cycle, AcCoA-linked glycolysis pathway, citric acid cycle and fatty acid synthesis pathway, 2HPP is the metabolite linking glycolysis pathway, pentose phosphate pathway and carbon fixation, and GlyP plays a key role in glycolysis pathway, fructose and mannose metabolism, glycerophospholipid metabolism, carbon fixation, nicotinate and nicotinamide metabolism.

4. Conclusion

Owing to the fast development of genome-scale network reconstruction, analysing these networks is one of the most important challenges for post-genomic biology. However, there are too many metabolites and metabolic reactions in these networks, and thus there are limited number of traditional metabolic engineering methods to understand and interpret these large networks. Recent studies show that complex network methods are more promising for modelling and analysing genome-scale metabolic networks. The results suggest that the method is invaluable in understanding cellular organizational principles, as well as proposing new hypotheses [4–7].

We applied the methods to investigate the structure and function of *S. aureus* metabolic network in the present paper. We have initiated the study by extracting the model from a recently reconstructed high-quality *S. aureus* metabolic network. The obtained model is represented by a metabolite graph. Then, based on the 'bow tie' structure character, we explained and discussed the functional significance and global structure of *S. aureus* metabolic network, we validated the 'scale-free' and 'small-world' characters. At last, modularity and centrality analysis of giant strong component in *S. aureus* metabolic networks were studied with their biological significance.

Acknowledgements

The authors thank the anonymous reviewers for valuable comments on this study.

References

[1] J L Reed *et al*, *Nat. Rev. Genet.* **7**, 130 (2006)
[2] A M Feist *et al*, *Nat. Rev. Microbiol.* **7**, 129 (2009)
[3] H Jeong *et al*, *Nature* **407**, 651 (2000)
[4] A L Barabasi and Z N Oltvai, *Nat. Rev. Genet.* **5**, 101 (2004)
[5] T Aittokallio and B Schwikowski, *Brief Bioinform.* **7**, 243 (2006)
[6] D W Ding and L N Li, *J. Biol. Syst.* **17**, 479 (2009)
[7] D W Ding *et al*, *Braz. J. Microbiol.* **40**, 411 (2009)
[8] D S Lee *et al*, *J. Bacteriol.* **191**, 4015 (2009)
[9] H W Ma and A P Zeng, *Bioinformatics* **19**, 270 (2003)
[10] H W Ma and A P Zeng, *Bioinformatics* **19**, 1423 (2003)
[11] J Zhao *et al*, *Chin. Sci. Bull.* **52**, 47 (2007)
[12] D W Ding and L N Li, *Rivista di Biologia/Biology Forum* **101**, 12 (2009)
[13] M D Humphries and K Gurney, *PLoS ONE* **3**, e0002051 (2008)
[14] M E J Newman and M Girvan, *Phys. Rev.* **E69**, 026113 (2004)
[15] R Guimera and L A N Amaral, *Nature* **433**, 895 (2005)
[16] B H Junker, D Koschutzkiet and F Schreiber, *BMC Bioinform.* **7**, 219 (2006)
[17] J A Papin *et al*, *Trends Biochem. Sci.* **28**, 250 (2003)
[18] J A Papin *et al*, *Trends Biotechnol.* **22**, 400 (2004)
[19] S Schuster, D A Fell and T Dandekar, *Nat. Biotechnol.* **18**, 326 (2000)