Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Theoretical Biology 241 (2006) 823–829
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

Exploring local structural organization of metabolic networks using subgraph patterns

Young-Ho Eom[a], Soojin Lee[b], Hawoong Jeong[a,*]

[a]Department of Physics, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea
[b]Department of Microbiology, Chungnam National University, Daejeon 305-764, Korea

Received 31 August 2005; received in revised form 17 January 2006; accepted 17 January 2006
Available online 28 February 2006

Abstract

Metabolic networks of many cellular organisms share global statistical features. Their connectivity distributions follow the long-tailed power law and show the small-world property. In addition, their modular structures are organized in a hierarchical manner. Although the global topological organization of metabolic networks is well understood, their local structural organization is still not clear. Investigating local properties of metabolic networks is necessary to understand the nature of metabolism in living organisms. To identify the local structural organization of metabolic networks, we analysed the subgraphs of metabolic networks of 43 organisms from three domains of life. We first identified the network motifs of metabolic networks and identified the statistically significant subgraph patterns. We then compared metabolic networks from different domains and found that they have similar local structures and that the local structure of each metabolic network has its own taxonomical meaning. Organisms closer in taxonomy showed similar local structures. In addition, the common substrates of 43 metabolic networks were not randomly distributed, but were more likely to be constituents of cohesive subgraph patterns.

Keywords: Metabolic network; Network motif

1. Introduction

Many chemical reactions occur within cells to produce energy or to build various molecules. The sum of all these chemical reactions within a living organism is referred to as metabolism. Cellular metabolism involves a complex network of metabolic substrates produced through enzyme-catalysed biochemical reactions. It is important to understand the topological structure of metabolic networks (Barabási and Oltvai, 2004). A series of recent works has provided a thorough understanding of the global and large-scale organization of metabolic networks (Jeong et al., 2000; Ravasz et al., 2002). These are scale-free networks in which the degree of distribution follows a power-law distribution $P(k) \sim k^{-\gamma}$ where k is the number of reactions. The network diameter, the average path length of a network, of each organism is small and the same for all organisms (Jeong et al., 2000), indicating that the average number of reactions in which a certain substrate takes part increases with the number of substrates found within a given organism. The networks also have a hierarchical organization of modularity, that is, they have modular structures and the organization of modules is hierarchical (Ravasz et al., 2002). However, the local structure of metabolic networks is still not fully understood.

Subgraph patterns and network motifs have been applied recently to understand the local structure of complex networks (Milo et al., 2002, 2004; Vazquez et al., 2004). Subgraph patterns consist of more than three nodes and the links connecting only these nodes, which represent the minimum subnetworks of complex networks. Examples of triad subgraph patterns are shown in Fig. 1A. Network motifs are the subgraph patterns that occur in a complex network at numbers that are significantly higher than those in a random network (Milo et al., 2002). These are believed to represent the simplest building blocks of complex networks and the topologically characteristic

*Corresponding author. Tel.: +82 42 869 2543; fax: +82 42 869 2510.
E-mail address: hjeong@kaist.ac.kr (H. Jeong).

0022-5193/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2006.01.018

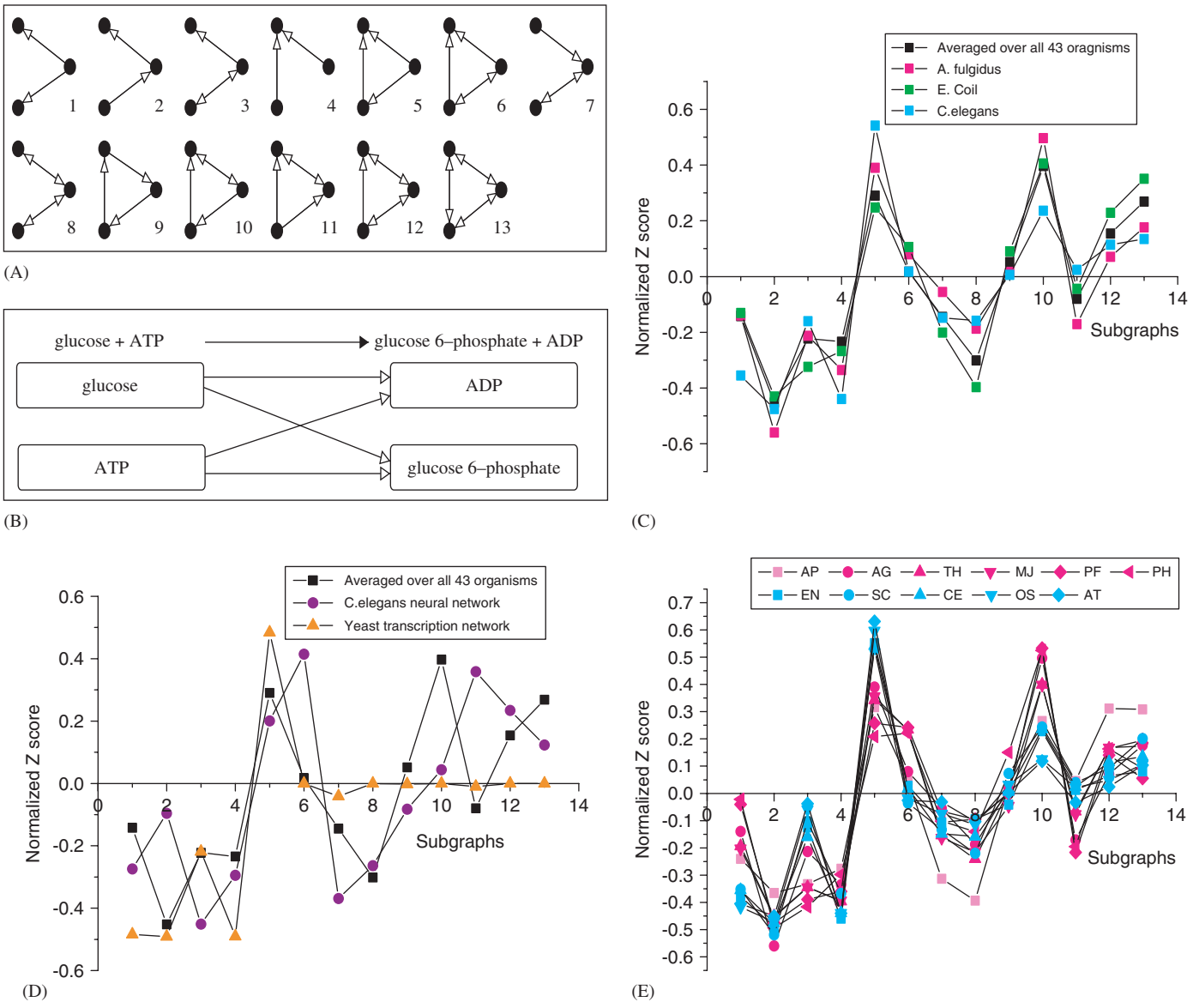

Fig. 1. (A) All possibilities of 13 types of three-node connected subgraphs. A number was assigned to each pattern. (B) Theoretical graphical representation of a chemical reaction in the metabolic networks space. (C) The triad significance profiles (TSPs) of metabolic networks. (D) TSPs of three different biological networks: metabolic networks of 43 organisms, *C. elegans* neural network, and yeast transcription network. (E) TSPs of six Archaea (pink). AP, *A. pernix*; AG, *A. fulgidus*; TH, *M. thermoautotrophicum*; MJ, *M. jannashii*; PF, *P. furiosus*; PH, *P. horikoshii*. TSPs of five Eukaryotes (blue). EN, *E. nidulans*; SC, *S. cerevisiae*; CE, *C. elegans*; OS, *O. savita*; AT, *A. thaliana*.

interaction patterns within complex networks. Recently, it was also shown that certain motifs have been enhanced through the evolution of a network, which supports the functional importance of the motifs (Vazquez et al., 2004). For example, in transcription networks, a biochemical network responsible for regulating the expression of genes in cells, the network motifs are thought to be circuit elements that perform key information-processing functions (Milo et al., 2002; Shen-Orr et al., 2002; Mangan and Alon, 2003). The feed-forward loop, one network motif of transcription networks, can act as a circuit that reduces noise and responds only to a persistent signal.

We analysed the local structure of metabolic networks of 43 organisms from three different domains of life. We examined the triad subgraph patterns of metabolic networks of 43 organisms and identified their network motifs involved in the metabolic networks of 43 different organisms based on data deposited in the WIT database (Overbeek et al., 2000), which is also available at http://www.nd.edu/~networks/resources. These organisms cover three domains of life including six Archaea, 32 Bacteria, and five Eukaryotes. In the metabolic network, the nodes of the network represent metabolic substrates and the links represent the chemical reactions in which the substrates participate (Fig. 1A). The direction of each link implies the

direction from an input substrate (educt) to an output substrate (product) (Fig. 1B).

2. Network motifs and significance profile

We used the following algorithm to obtain the network motifs (Milo et al., 2002). We scanned for all possible three-node subgraphs in the network and recorded the number of occurrences of each subgraph. To identify a statistically significant subgraph pattern, we compared the network to an ensemble of suitably randomized networks. In our implementation, we used a Markov-chain algorithm. Starting with the original network, we randomly switched chosen pairs of links ($S1 \to P1$, $S2 \to P2$ is replaced by $S1 \to P2$, $S2 \to P1$) repeatedly until the network is well randomized. Switching is prohibited if the either of the links $S1 \to P2$ or $S2 \to P1$ already exist. Each node in the randomized networks contained the same number of incoming and outgoing links as the corresponding node in the original network. In addition, the randomized networks that were used to estimate the significance of n-node subgraphs were generated to preserve the same number of appearances of all $(n-1)$-node subgraphs as in the original network. For three nodes case, to preserve the number of bidirectional links, we only switched a bidirectional link for a different bidirectional link ($S1 \leftrightarrow P1$, $S2 \leftrightarrow P2$ to $S1 \leftrightarrow P2$, $S2 \leftrightarrow$ P1), only if both ($S1$ and $P2$) and ($S2$ and $P1$) are not connected by a link in any direction. In case of the unidirectional link, we only switched a unidirectional link for a different unidirectional link ($S1 \to P1$, $S2 \to P2$ to $S1 \to P2$, $S2 \to P1$), only if they do not generate new bidirectional links and the either of the links $S1 \to P2$ or $S2 \to P1$ do not already exist. For each subgraph i, the statistical significance of the subgraph is described by the Z score

$$Z_i = (N_i^{real} - \langle N_i^{rand} \rangle)/std(N_i^{rand}). \quad (1)$$

N_i^{real} is the number of appearances of the subgraph i in the network, and $\langle N_i^{rand} \rangle$ and $std(N_i^{rand})$ are the average and standard deviation of its appearances in the ensemble of randomized networks, respectively. The subgraph pattern exhibiting a high Z score is the statistically significant pattern. Here, the network motifs are those subgraph patterns having a Z score greater than 2. We applied this algorithm to detect network motifs from the metabolic networks of 43 organisms and found that all metabolic networks have their own network motifs. To provide a more quantitative analysis, we investigated the local structure of metabolic networks of each organism in detail and identified the significance profile (SP) of each metabolic network (Milo et al., 2004). The SP is the vector of Z scores normalized to a unit length, of which the ith component is given by

$$\mathbf{SP}_i = Z_i \Big/ \left(\sum_j Z_j^2 \right)^{1/2}. \quad (2)$$

The SP of a given network represents the relative significance of the subgraphs in that network. It is important to compare networks of different sizes because network motifs in large networks tend to have higher Z scores than network motifs in small networks (Milo et al., 2004).

The triad significance profile (TSP) for each metabolic network is presented in Fig. 1C. The TSPs of these networks are found to be almost insensitive to a removal of 20% of edges or to an addition of 20% new edges randomly, representing that our results are robust to possible missing or false-positive data errors. All metabolic networks showed similar TSPs and three network motifs of triads 5, 10, and 13 were found frequently. These motifs, especially 5 and 10, are well-known feed-forward loop and its variation of function is a prevalence of short detours in metabolic network (Gleiss et al., 2001; Heinrich and Schuster, 1996). In contrast, triads 2, 4, and 8 were antimotifs that were significantly underrepresented. The correlation coefficient between the TSPs of metabolic networks in 43 organisms was about 0.78. The correlation coefficient between a pair of vectors $\mathbf{SP}(u)$ and $\mathbf{SP}(v)$ was defined as $\sum_{i=1}^{N} (SP(u)_i - m_u)(SP(v)_i - m_v)/(|\mathbf{SP}(u)||\mathbf{SP}(v)|)$, where $m_u = \sum_{i=1}^{N} SP(u)_i/N$ is the average SP of vector $\mathbf{SP(u)}$ over all N (here 13) types of subgraphs and $|\mathbf{SP(u)}| = \sqrt{\sum_{i=1}^{N} (SP(u) - m_i)^2}$ is the norm of the centered $\mathbf{SP}(u)$. A high correlation coefficient implies high similarity. Thus, the correlation of 0.78 shows that metabolic networks have the same topological structure in both large-scale organization (inhomogeneous power-law degree distribution) and in local organization (sharing common topological substructures).

To clarify whether these local structural properties apply to metabolic networks only, we identified TSPs of other biological networks, including the neural network of *C. elegans* (Achacoso and Yamamoto, 1992) and the transcription network of yeast (Fig. 1D), which is available at http://www.weizmann.ac.il/mcb/UriAlon/Papers/networkMotifs/. Interestingly, the TSPs of the tested network differed from those of the metabolic network. In the metabolic network, triads 5, 10, and 13 had high normalized Z scores and triads 2, 4, and 8 had low normalized Z scores. In contrast, in the yeast transcription network, only triad 5 had a high normalized Z score and triads 1, 2, and 4 had low normalized Z scores. In the *C. elegans* neural network, triads 5, 6, and 11 had high normalized Z scores and triads 3, 4, and 7 had low normalized Z scores. Although triads 5 (high Z score) and 4 (low Z score) share a similar behavior, the overall profiles indicate that the design principle of metabolic networks differs from that of other biological networks.

3. Taxonomy based on local structure

Another interesting feature of the local organization of metabolic networks is that the TSP of each metabolic network has taxonomical meaning. Although the overall

TSPs of metabolic networks are similar, the TSPs of the organisms belonging to same taxonomic group are closer. For example, the correlation coefficient of TSPs from only the metabolic network of the six Archaea species tested is 0.91 and that of the five Eukaryotes is 0.99, indicating that metabolic networks in the same taxonomy group show similar local characteristics (Fig. 1C). However, the correlation coefficient from the TSPs of 32 Bacteria is 0.8, which is only slightly higher than the average value of all 43 organisms (0.78). Bacteria can be divided into several subgroups based on their evolutionary details (e.g. parasitic or non-parasitic bacteria) (Andersson and Andersson, 1999) (see Fig. 2). When these bacterial subgroups are considered, the correlation coefficient between TSPs of metabolic networks that belong to the same bacteria subgroups becomes much higher. For example, the correlation coefficient between TSPs of metabolic networks for six γ-proteobacteria is 0.92. Using TSP analysis, we also showed that the taxonomical classification in Archae corresponds highly with the metabolic network system. Archaea can be classified into two phylum, Crenarchaeota and Euryarchaeota. The TSP of *Aeropyrum pernix*, which belongs to Crenarchaeota, differs somewhat from the TSPs of the other five Archaea, which belong to the different phylum, Euryarchaeota (Fig. 1E).

Fig. 2 shows the correlation coefficient matrix of TSPs for all metabolic networks. These results clearly show that the species-specific feature is reflected in the local organization of metabolic networks. Recently, Milo et al. suggested a method of dividing various complex networks into "superfamilies" using the network TSPs (Milo et al., 2004) and found several superfamilies of previously unrelated networks with similar SP networks. For example, one superfamily includes the protein signaling network, developmental genetic network, and neural network. This superfamily is a group of biological information-processing networks. Although the details of the networks differ, the design principles of networks are reflected in the local structure of the networks. Our results represent biologically meaningful 'families' of metabolic networks of 43 organisms, or in this case, taxonomical groups.

From an evolutionary point of view, it is also interesting that the system-level organization of archaeal and eukaryotic metabolic networks is closely related (Podani et al., 2001). In local-level organization, Bacteria have four subgroups, as illustrated in Fig. 2. Bacteria belonging to subgroups 1, 2, and 4 are closely related to Crenarchaeota (one phylum of Archaea), Euryarchaeota (another phylum of Archaea), and Eukaryote, respectively. Bacteria belonging to subgroup 3 are not closely related to Archaea or Eukaryote. There are several ways to produce hierarchical classification from the correlation matrix. One of the most frequently used algorithms is the average-link algorithm, a type of hierarchical algorithm. In the

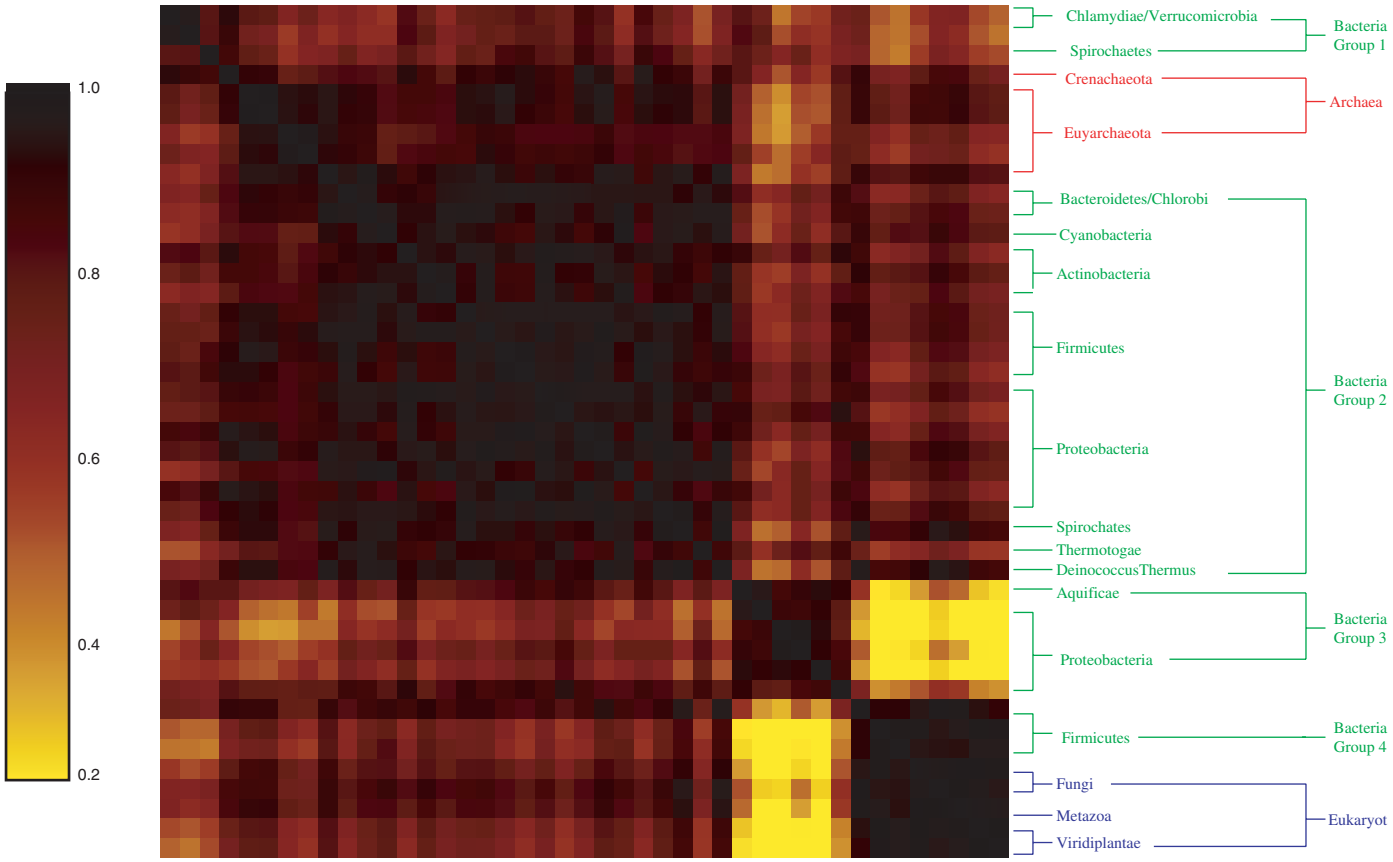

Fig. 2. The correlation coefficient matrix of triad significance profiles for the metabolic networks of 43 organisms.

hierarchical algorithm, objects are connected by a tree structure called the dendrogram (Fig. 3) in which objects in the same subtree are closer to each other than objects in different subtrees. In the average-link algorithm, the cluster similarity of two clusters is the average pairwise similarity between organisms in the two clusters (the correlation between SPs in this case). Using the average-link algorithm, we analysed the metabolic networks of 43 organisms (Fig. 3) and identified the subgroups of bacteria and the separation of Crenarchaeote, *A. pernix* from five Euryarchaeotes. This separation was previously observed in the system-level organization (Podani et al., 2001). We also applied our method to the alternative bipartite network where intermediate reactions are nodes, and found that still taxonomical classification is still valid although details of TSP are not the same.

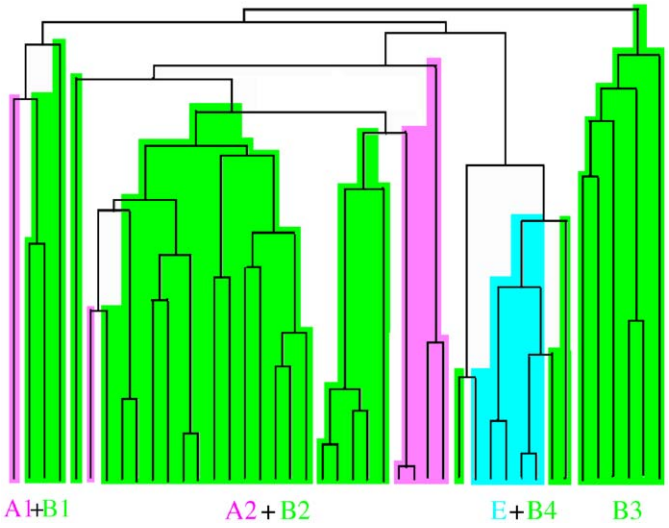

Fig. 3. Hierarchical dendrogram based on the significance profiles of the metabolic networks. Here we use the average-link algorithm. A, Archaea; B, Bacteria; E, Eukaryotes.

4. Correlation between cohesive subgraph patterns and common substrates

In the protein-interaction network of *S. cerevisiae*, proteins organized in cohesive subgraph patterns are more conserved than those not associated in such cohesive patterns (Wuchty et al., 2003). Orthologs are not randomly distributed through the protein-interaction network in yeast, but are the constituents of cohesive subgraph patterns. In other words, proteins that have evolutionary importance make more cohesive subgraph patterns in the protein-interaction network. And using conserved network motif, protein-interaction prediction algorithm is suggested (Albert and Albert, 2004). From a similar motivation, what subgraph patterns are generated by common substrates in the metabolic networks is an interesting question. We investigated the correlation between the cohesiveness of subgraph patterns and the probability that a given subgraph pattern is composed of substrates common to all 43 organisms. Cohesive patterns contain more internal links than non-cohesive patterns. In directed patterns, cohesive patterns also have more bidirectional links than non-directed patterns. First, we selected 62 common substrates of all 43 organisms to analyse, such as ATP, ADP, water, and glycine, which are basic materials required for survival. We then analysed the subgraph patterns of these 62 common substrates only as they appeared in a given metabolic network. For comparison, we used 62 randomly selected substrates as controls and identified the subgraph patterns for these randomly selected substrates only.

The results for the three-node subgraph in metabolic networks of *S. cerevisiae* are shown in Table 1. We obtained similar results in the other 42 organisms. We found remarkably different conservation rates for substrates in the different subgraphs. The conservation rate for a given subgraph pattern is defined as the number of

Table 1
Relationship between cohesiveness of subgraph patterns and the probability that a given subgraph is composed of common substrates only

Subgraph	Number of subgraphs	Natural conservation rate (%)	Random conservation rate (%)	Conservation ratio
1	2512	3.60	0.16	23.23
2	1375	2.84	0.14	20.05
3	9143	4.98	0.15	33.77
4	1452	2.75	0.13	21.12
5	130	6.15	0.13	45.71
6	35	28.57	0.12	238.1
7	7230	4.83	0.13	36.44
8	18 313	5.97	0.14	42.51
9	5	0.00	0.20	0
10	72	16.67	0.12	137.93
11	57	19.30	0.13	144.74
12	274	36.13	0.15	242.65
13	493	28.40	0.15	195.26

The natural conservation rate is the fraction of the original *S. cerevisiae* subgraph that consists of the common substrates only. The random conservation rate represents the fraction of subgraph patterns consisting of the randomly selected common substrates. The conservation ratio is the natural conservation rate divided by random conservation rate.

subgraphs of which all constituents come from only the common substrates divided by the number of all subgraphs. As shown in Table 1, all three constituents of 3.6% of triad 1 are the common substrates, whereas all three constituents of 36.1% of triad 12 are made only with the common substrates. We found a tendency for the subgraph patterns with a loop structure and a double link (i.e. more cohesive patterns) to have high conservation rates (e.g. triads 6, 10, 11, 12, and 13).

These results indicate that the common substrates are not distributed randomly in the metabolic network but are the constituents of cohesive patterns. In other words, substrates which are essential for survival are more likely to generate cohesive subgraph patterns. To test the validity of this finding, we made random sets of 62 common substrates and calculated the conservation ratio, defined as the natural conservation rate divided by the random conservation rate. All the subgraph patterns except triad 9 had a conservation ratio higher than 20, indicating that the network topology influences the natural placement of common substrates in the metabolic network. In addition, the cohesive subgraph patterns have a high conservation ratio. We found similar results in the four-node case, verifying that the common substrates are more likely to be the constituents of cohesive patterns.

5. Conclusion

Through an intensive analysis of the local structure of metabolic networks of 43 organisms, we identified network motifs of the metabolic networks from three domains of life: Archaea, Bacteria, and Eukaryotes (Table 2). The triad

Table 2
Statistics of each network

Network	AP	AG	TH	MJ	PF	PH
Number of nodes	201	493	427	421	312	320
Number of links	911	2247	2022	1939	1384	1401
Network motif	5, 12, 13	5, 10, 13	5, 6, 10, 12, 13	5, 6, 10, 12, 13	5, 6, 10	5, 6, 10, 12
Network	AA	CQ	CT	CY	PG	MB
Number of nodes	414	187	211	539	415	421
Number of links	1911	646	766	2576	1835	1894
Network motif	10, 12, 13	12, 13	12, 13	5, 10, 13	5, 10, 13	5, 10, 12, 13
Network	ML	MT	BS	EF	CA	MG
Number of nodes	417	580	772	375	486	199
Number of links	1904	2738	3809	1862	2377	783
Network motif	5, 10	5, 10, 13	5, 10, 12, 13	5, 10, 13, 13	5, 10, 13	5, 10, 12
Network	MP	PN	ST	CL	RC	RP
Number of nodes	171	404	390	386	663	206
Number of links	721	1962	1915	1773	3139	824
Network motif	5, 10	5, 10, 13	10, 13	10, 13	5, 10, 12, 13	10, 12, 13
Network	NG	NM	CJ	HP	EC	TY
Number of nodes	399	374	373	369	765	806
Number of links	1907	1798	1715	1764	3904	4049
Network motif	10, 13	10, 13	10, 13	5, 10, 12, 13	5, 10, 12, 13	5, 10, 12, 13
Network	YP	AB	HI	PA	TP	BB
Number of nodes	555	385	508	725	202	179
Number of links	2556	1711	2421	3511	864	696
Network motif	5, 10, 13	5, 10, 13	5, 6, 10, 12, 13	5, 10, 13	5, 6, 10	13
Network	TM	DR	EN	SC	CE	OS
Number of nodes	333	803	377	551	452	288
Number of links	1543	3889	1704	2789	2159	1218
Network motif	10	5, 10, 13	5, 10	5, 10, 13	5, 10	5, 10, 13
Network	AT	CE neural	Yeast transcription			
Number of nodes	299	297	688			
Number of links	1276	2345	1079			
Network motif	5, 10	5, 6, 10, 11, 12, 13	5			

AP, *A. pernix*; AG, *A. fulgidus*; TH, *M. thermoautotrophicum*; MJ, *M. jannaschii*; PF, *P. furiosus*; PH, *P. horikoshii*; AA, *A. aeolicus*; CQ, *C. pneumoniae*; CT, *C. trachomatis*; CY, *Synechocystissp.*; PG, *P. gingivalis*; MB, *M. bovis*; ML, *M. leprae*; MT, *M. tuberculosis*; BS, *B. subtilis*; EF, *E. faecalis*; CA, *C. acetobutylicum*; MG, *M. genitalium*; MP, *M. pneumoniae*; PN, *S. pneumoniae*; ST, *S. pyogenes*; CL, *C. tepidum*; RC, *R. capsulatus*; RP, *R. prowazekii*; NG, *N. gonorrhoeae*; NM, *N. meningitidis*; CJ, *C. jejuni*; HP, *H. pylori*, EC, *E. coli*; TY, *S. typhi*; YP, *Y. pestis*; AB, *A. actinomycetemcomitans*; HI, *H. influenzae*; PA, *P. aeruginosa*; TP, *T. pallidum*; BB, *B. burgdorferi*; TM, *T. maritima*; DR, *D. radiodurans*; EN, *E. nidulans*; SC, *S. cerevisiae*; CE, *C. elegans*; OS, *O. sativa*; AT, *A. thaliana*.

significance profiles of metabolic networks were used to compare the local structure of metabolic networks. The local organization of metabolic networks is similar to each other, but is also species specific. These results indicate that, despite significant variation in their individual substrates and pathways, these metabolic networks have the same topological properties in both the large scale and local sense. Moreover, the local organization of metabolic networks clearly shows characteristic features of each individual metabolic network.

By investigating the correlations between common substrates of 43 organisms and the cohesiveness of subgraph pattern, we found that the common substrates are not randomly distributed in the metabolic network but are the constituents of cohesive patterns.

Acknowledgments

This work was supported by the Ministry of Science and Technology through a Korean Systems Biology Research Grant M10309020000-03B5002-00000 (Y. E.), by ABRL through Grant no. R14-2002-059-01002-0 (H. J.) and by Chungnam National University through Grant no. 2005-6223 (S. L.).

References

Achacoso, T.B., Yamamoto, W.S., 1992. Neuroanatomy of *C. elegans* for Computation. CRC Press, Boca Raton, FL.

Albert, I., Albert, R., 2004. Conserved network motifs allow protein–protein interaction prediction. Bioinformatics 20, 3346–3352.

Andersson, J.O., Andersson, S.G., 1999. Insights into the evolutionary process of genome degradation. Curr. Opin. Genet. Dev. 9, 664–671.

Barabási, A.-L., Oltvai, Z.N., 2004. Network biology: understanding the cell's functional organization. Nature Reviews. Genetics 5, 101–113.

Gleiss, P.M., Stadler, P.F., Wagner, A., Fell, D.A., 2001. Relevant cycles in chemical reaction networks. Adv. Complex Systems 4, 207–226.

Heinrich, R., Schuster, S., 1996. The Regulation of Cellular Systems. Chapman & Hall, New York.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., Barabási, A.-L., 2000. The large-scale organization of metabolic networks. Nature (London) 407, 651–654.

Mangan, S., Alon, U., 2003. Structure and function of the feed-forward loop network motif. Proc. Natl Acad. Sci. USA 100, 11980–11985.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U., 2002. Network motifs: simple building blocks of complex networks. Science 298, 824–827.

Milo, R., Itzkovitz, S., Kashtan, N., Levitt, R., Shen-Orr, S., Ayzenshtat, I., Sheffer, M., Alon, U., 2004. Superfamilies of evolved and designed networks. Science 303, 1538–1542.

Overbeek, R., et al., 2000. WIT: integrated system for high-throughput genome sequence analysis and metabolic reconstruction. Nucleic Acids Res. 28, 123–125.

Podani, J., Oltvai, Z.N., Jeong, H., Tombor, B., Barabási, A.-L., Szathmary, E., 2001. Comparable system-level organization of Archaea and Eukaryotes. Nat. Genet. 29, 54–56.

Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., Barabási, A.-L., 2002. Hierarchical organization of modularity in metabolic networks. Science 297, 1551–1555.

Shen-Orr, S., Milo, R., Mangan, S., Alon, U., 2002. Network motifs in the transcriptional regulation network of *Escherichia coli*. Nat. Genet. 31, 64–68.

Vazquez, A., Dobrin, R., Sergi, D., Eckmann, J.-P., Oltvai, Z.N., Barabási, A.-L., 2004. The topological relationship between the large-scale attributes and local interaction patterns of complex networks. Proc. Natl Acad. Sci. USA 101, 17 940.

Wuchty, S., Oltvai, Z.N., Barabási, A.-L., 2003. Evolutionary conservation of motif constituents in the yeast protein interaction network. Nat. Genet. 35, 176–179.