Chapter 6

From Metabolic Reactions to Networks and Pathways

Masanori Arita

Abstract

Enzymatic reactions form a hypergraph structure and their translation into a graph structure accompanies an information loss. This chapter introduces well-known topological transformations from metabolic reactions to a graph, and discusses their advantages and disadvantages. Also discussed is the legitimacy of defining cofactors or currency metabolites, and suitable application area of each representation.

Key words: Metabolism, Network, Pathway, Modularity, Database

1. Introduction

After the publication of Barabási and Albert's (1) inspiring work, computing the degree exponent of a network became a standard process in analyzing biological networks. Irrespective of the network under investigation, researchers routinely discuss the evolutionary implication(s) of its hubs, motifs, or modules. They also associate characteristics of the network, such as the average path length or clustering coefficient with biological implications (Table 1). These properties, called network metrics, are actively applied to characterize and classify the network topology while ignoring dynamics (see Chapter 13) (2). Metabolism was the first such application (3) and ever since, many variations have been proposed.

The notable characteristic of network biology, i.e., metric-based understanding of biology, is its topological representation and associated discretization. In real metabolism, each reaction includes multiple substrates and products. Enzymes are regulated by multiple components and protein modules, and their reaction mechanism is characterized by various physicochemical quantitative parameters, such as the Gibbs energy and enzyme velocity. In the analysis of network biology, on the other hand, quantitative factors are reduced to a simple graph structure and the main focus is on the degree distribution, reachability between nodes, or modularity.

Jacques van Helden et al. (eds.), *Bacterial Molecular Networks: Methods and Protocols*, Methods in Molecular Biology, vol. 804, DOI 10.1007/978-1-61779-361-5_6,

Table 1
Network metrics: well-known parameters used to characterize the topology of networks

Degree, density, and assortativity	*Degree* is the number of edges for each node. High-degree nodes are called hubs *Density* is the proportion of edges for each node to the maximum degree. Hubs are high-density nodes *Clustering coefficient* (CC) is the average of the edge density of immediate neighbors for a given node. Natural networks are known to have high (ca.0.3 and up) CCs *Assortativity* is a correlation coefficient between the degree distributions of node neighbors. When many nodes are connected with nodes of different degrees, the network is called *dissortative*
Distance and path	*Distance* between nodes is the least number of edges to connect them, i.e., the length of their shortest path *Path* is the actual sequence of edges. *Average path length* (APL) is the average of distances between all pairs of nodes. Many natural networks are known to have small APLs
Centrality	*Centrality* is the criterion to determine the "center" of the network in an undirected graph. (In a directed graph, the equivalent notion is called *prestige*.) *Degree-centrality* chooses the most-connected hub as the center. *Closeness centrality* chooses the node whose average distance to all other nodes is the minimum *Betweenness centrality* chooses the node which are most used in shortest paths between all pairs of nodes
Motif	*Motif* is a small "building block" or frequently appearing pattern in the network. Typically, all patterns up to five nodes are exhaustively searched in the network. Well-known directed motifs include feed-forward and feed-back loops (see Chapters 12 and 14)
Modularity	*Modularity* is a degree of separation for sub-networks. Many natural networks are known to have high modularity. The study on modularity is immature and no clear definition exists

There are advantages and disadvantages in the discretized representation. One important benefit is its simplicity. Very large-scale networks are amenable to algorithmic analyses as a simple graph but not as a more complicated representation (e.g., differential equations). Also, facile computation and simplistic interpretation could attract many computational biologists into this research field. On the backstage, however, many biological aspects are overlooked; as will be explained, even the original network topology may be sacrificed. In this chapter, we introduce topological transformations from metabolic reactions to graph representation and discuss their advantages and disadvantages. We also discuss the legitimacy of defining cofactors or currency metabolites.

2. Strategies for Graph Transformation

The graph structure is a collection of directed/undirected binary relations defined as the paired set of nodes and edges. Each relation, or edge, has only one source and one target node. (The source and target are not distinguished in the case of undirected graphs.) Originally, standard metabolic reactions catalyze multiple substrates into products. Instead of a graph, they form a hypergraph where each edge may connect any number of nodes, as represented by a box in Fig. 1a. Information on metabolite

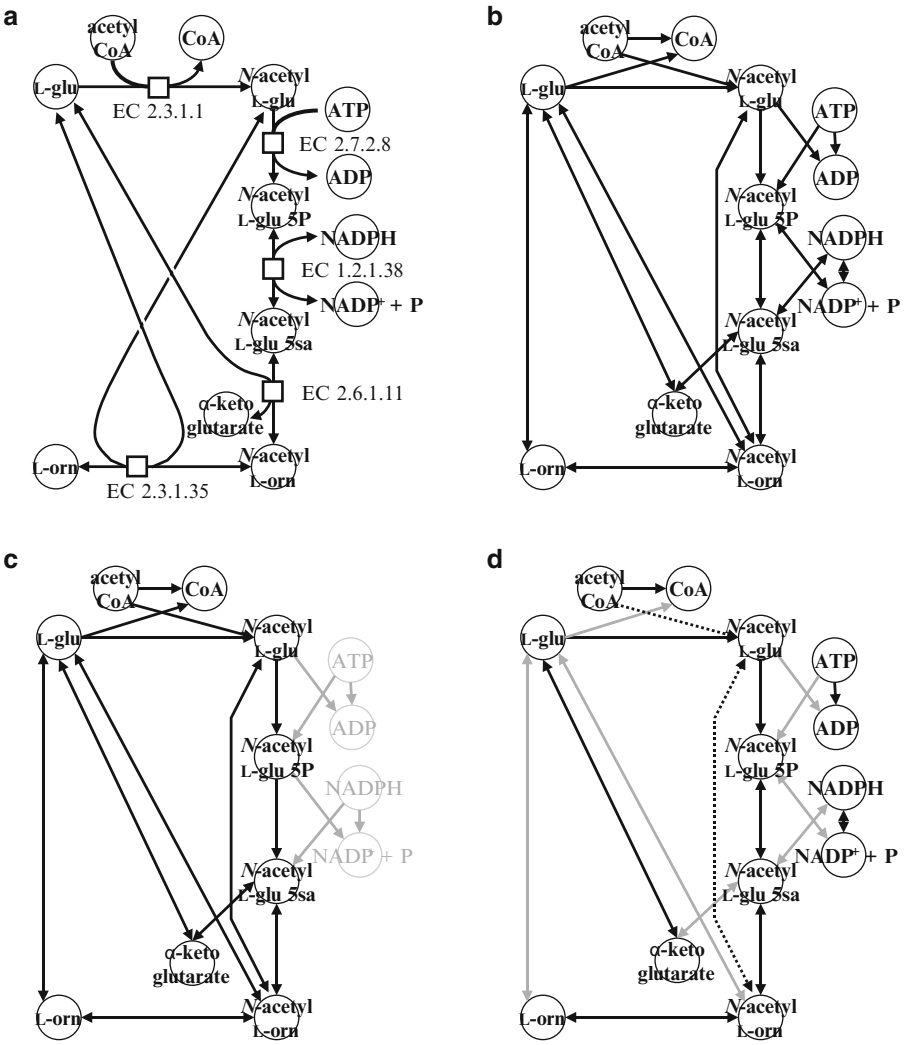

Fig. 1. Three network representations (**b–d**) for the L-ornithine biosynthetic pathway (**a**). (**a**) *Bipartite representation*; (**b**) *fully connected substrate graph*; (**c**) *substrate graph* without currency metabolites; and (**d**) substrate–product *graph*. *Dotted* and *solid lines* indicate the acetyl moiety and the carbon backbone of L-glutamate, respectively. *Dotted lines* are missed in the *substrate graph* with approximate structural relationships.

Table 2
Network hubs in different graph representations for *Escherichia coli* metabolism

Jeong et al. (3)	Fell and Wagner (4)	Ma and Zeng (9)	Arita (12)
H_2O	L-Glutamate	Glycerate 3P	CO_2
ADP	Pyruvate	D-Ribose 5P	Pyruvate
P	Coenzyme A	Acetyl CoA	Acetyl CoA
ATP	α-Ketoglutarate	Pyruvate	ATP
L-Glutamate	L-Glutamine	D-Xylulose 5P	D-Glucose
$NADP^+$	L-Aspartate	D-Fructose 6P	L-Glutamate
PP	Acetyl CoA	5P-D-ribose 1PP	D-Galactose
NAD^+	Phosphoribosyl PP	L-Glutamate	Coenzyme A
NADPH	Tetrahydrofolate	D-Glyceraldehyde 3P	*S*-adenosyl L-methionine
NADH	Succinate	L-Aspartate	5P-D-ribose 1PP

The list of top ten hub metabolites for each graph transformation. The list is a reproduction from Arita (12). *P* (ortho)phosphate, *PP* pyrophosphate

structures is implicitly represented in this hypergraph format, and the structural information can be easily lost in the reduction of metabolism into a graph. Depending on their purposes, several transformation strategies exist for mapping the metabolic hypergraph into a graph. Information loss is unavoidable to some extent, and no transformation strategy is perfect. To explain the respective benefit(s) of transformation, we introduce major strategies one by one, using L-ornithine biosynthesis, the pathway from L-glutamate to L-ornithine, as an example (Fig. 1).

2.1. Fully Connected Substrate Graph

As in the traditional metabolic chart, graph nodes represent metabolites in the substrate graph. Nodes are linked if they participate in the same reaction (Fig. 1b). In the fully connected representation, substrates in a reaction are fully linked with their products, i.e., each reaction is transformed to a complete bipartite graph. The network therefore portrays the relation by "reaction membership." Apparently, frequently occurring metabolites in reactions become the most connected metabolites or hubs.

In their pioneering study, Jeong et al. (3) constructed fully connected substrate graphs for 43 organisms representing three domains of life. They "unexpectedly" found that the average path length of the network (APL; see Table 1 for definition) was identical for all organisms (around 3–4 steps), and the ranking of the most connected metabolites was practically identical. In order of frequency, the hubs were H_2O, ADP, orthophosphate (P), ATP, L-glutamate, $NADP^+$, pyrophosphate (PP), NAD^+, NADPH, and NADH (Table 2). These metabolites are usually referred to as

inorganics or cofactors. Jeong et al. (3) argued that "the same highly connected substrates may provide the connections between modules responsible for distinct metabolic functions." Indeed, these metabolites function as a shortcut to achieve the small APL, as L-glutamate is directly linked with L-ornithine in Fig. 1b.

2.2. Substrate Graph Without Currency Metabolites

For biochemists, inorganics and cofactors are minor players. In the traditional metabolic chart, they are depicted in smaller fonts and their linkage is often unattended. Some cofactors have been called currency metabolites because of their ubiquity. To exclude such currency metabolites and inorganics from the substrate graph, Fell and Wagner (4) manually specified hub metabolites to be excluded in their network reconstruction in *Escherichia coli* (Fig. 1c). Their removal list included ATP, ADP, NAD^+, $NADP^+$, NADH, NADPH, CO_2, ammonia, sulfate, thioredoxin, P, and PP. That is, except for L-glutamate, all top ten hubs in the work of Jeong et al. were removed and the remaining hub metabolites were brought forward to form the new ten hubs: L-glutamate, pyruvate, Coenzyme A (CoA), α-ketoglutarate, L-glutamine, L-aspartate, acetyl CoA, phosphoribosyl PP, tetrahydrofolate, and succinate.

The removal of currency metabolites is the most popular procedure in metabolic analysis; importantly, it does not change the topological statistics that are found in the fully connected, bipartite graph representation (4). In other words, currency metabolites are carefully chosen so that the resulting network retains the network distance, modularity, and other important properties of the metabolism (5, 6). In this sense, the decision is inevitably artificial and context-dependent. For example, Wagner and Fell (7) argued that L-glutamate is an important hub based on its ranking, but its high degree originated from many amino-transfer reactions, where L-glutamate and α-ketoglutarate function as amino-group donor and acceptor, respectively. They not only play important roles in amino acid metabolism, but they also function as cofactors in the amino-group transfer. Likewise, currency metabolites should not be removed from their own biosynthesis/degradation pathways. Clearly, the dichotomy of metabolites between currency and commodity (noncurrency) is unacceptable. Recent trends focus on network modularity rather than hubs (6, 8), but the same observation applies. It is a fundamental problem in the modern network analysis.

2.3. Substrate Graph with Structural Relationship

One way to ensure the proper treatment of currency metabolites is to verify each transformation manually. Ma and Zeng (9) manually specified edges between substrates and products for over 3,000 reactions so that edges connect structurally related nodes only. Their rule of thumb was to link moieties that share more than two carbons, thus overlooking the methyl and acetyl transfers and

noncarbon atoms (Fig. 1d). Later, Rahman et al. (10) proposed the use of structural similarity to automatically obtain the same effect in their Pathway Hunter Tool. The good news is that this method does not raise the dichotomy between currency and commodity metabolites. Its disadvantage however, lies in the bias originating from the number of reactions used. Because of this bias, the top ten hubs Ma and Zeng identified in the *E. coli* network included D-xylulose 5P and D-ribose 5P (Table 2), which play relatively minor roles from the metabolic perspective.

The direct reflection of the number of reactions as node degrees represents another fundamental problem regardless of the assignment of currency metabolites. Metabolic reactions for genes are often incomplete and ambiguous. Many enzymes exhibit broad substrate specificity and they are sometimes found to perform alternative functions. Such ambiguity is exemplified in the over 500 reassignments of the EC numbers (denoted as transferred entries in the EC hierarchy) and the frequent use of generic terms, such as "amino acid," "alcohol," "acyl," or "long-chain," in the systematic names of enzymes. Since such generic terms correspond to multiple substrates, the number of reactions corresponding to a given set of EC numbers may vary depending on the interpretation. Therefore, direct counting of the number of reactions is not a robust way for estimating metabolic hubs. For example, EC 2.3.1.85 (acyl-CoA:malonyl-CoA *C*-acyltransferase) is associated with as many as 30 reactions in the KEGG RPAIR database (in fact, many of them are not acyltransferase reactions and are automatically associated with this EC number according to an attached comment by the IUBMB) (11). If all reactions are automatically transformed to a graph representation, the common cofactors for this EC reaction, CoA and NADPs, receive too many graph edges. Still, the molecular structure of the main substrate, acyl CoA, remains ambiguous. The resulting graph structure depends on how these reactions are curated for each target species (including the removal of inappropriate assignments). In general, such curation problem has been ignored, especially by computational biologists, possibly because it is time consuming and its discussion is of little computational interest.

2.4. Substrate–Product Graph

To reduce the bias originating from the interpretation of multiple reactions, a different abstraction has been proposed; we call it the substrate–product graph (12). Instead of counting the number of reactions, this method counts the number of structural transformations by decomposing each reaction into a set of atomic mappings, and is independently employed to construct the KEGG RPAIR database (11). Atomic mapping comprises a set of atomic position pairs for a substrate–product relationship corresponding

to a migrating moiety in the reaction. For example, atomic mapping between ATP and ADP represents a loss of orthophosphate from ATP (kinase) and mapping between L-glutamate and α-ketoglutarate represents an exchange of an amino and oxo group (aminotransferase). In the substrate–product graph, nodes correspond to metabolites as in the substrate graph, but edges correspond to atomic mappings. No matter how often each atomic mapping appears in the reaction set (e.g., the atomic mapping between ATP and ADP, or NAD and NADH, appears quite frequently), only one edge is drawn for each structural correspondence. Since a cofactor is always involved in the same reaction pattern, its function corresponds to only one atomic mapping and the node degree is kept small. Most importantly, this representation can correct network statistics. Specifically, the small APL of the metabolic network was lost in this representation (13). The top ten hubs became CO_2, pyruvate, acetyl CoA, ATP, D-glucose, L-glutamate, D-galactose, CoA, *S*-adenosyl L-methionine, and phosphoribosyl PP. CO_2 was the topmost hub because of the various decarboxylation reactions. Likewise, the high rank of acetyl CoA and CoA resulted from many acyl transfers. In this representation, we reconfirmed the importance of L-glutamate pointed out by Wagner and Fell (7), and the statistics also implies the ancientry of the purine-related pathways from the plasticity of the adenosine moiety. Based on protein folds, the purine pathway was reported to be one of the most ancient subnetworks (14).

By focusing on the number of structural changes, the substrate–product graph concisely represents the structural plasticity in the metabolic network (Fig. 1d). The reduction to atomic mappings has both merits and demerits. First, it is robust to the ambiguity in genome annotations, such as the interpretation of general reactions. On the other hand, it is computationally expensive and loses information derived from redundancies in the reaction network. Nonetheless, the method is slowly gaining popularity. Blum and Kohlbacher (15) proposed a degree-weighted version of the substrate–product graph. Pitkanen et al. (16) presented a new, practical method that can detect branched pathways at atomic resolution. Faust et al. (17) used the RPAIR database to reconstruct the substrate–product graph and optimized weighting schemes for metabolic pathway finding (also see Chapter 7). All approaches reported significant improvement in metabolic analyses through the consideration of structural relationships.

2.5. Reaction Graph

The reaction graph, or enzyme-centric graph, is a yet another representation of a metabolic network (7). In this graph, each reaction corresponds to a graph node; an edge is drawn if two reactions share a substrate or product. The representation is a dual

image of the substrate graph: nodes and edges in the reaction graph correspond to edges and nodes in the substrate graph. Since there is a one-to-one correspondence between reaction- and substrate graphs, both representations contain the same amount of information. Therefore, the benefits and disadvantages of the substrate graph also apply to the reaction graph and it is not worthwhile to make comparisons as long as the graphs are dual images.

2.6. Bipartite Graph

Bipartite graph is a graph with two types of nodes, where either type is linked to the other type only. A metabolic network can be naturally mapped to a bipartite representation with metabolite- and reaction nodes. It represents association of metabolites with their reactions, and is often used for network visualization (Fig. 1a). Strictly speaking, the bipartite representation does not constrain that all substrates in a reaction should be together consumed to produce all its products. Also, the representation does not distinguish incoming edges or outgoing edges (in this sense, Fig. 1a is not an accurate bipartite view because it uses two kinds of edges, straight and curved, to denote co-enzymes). In summary, the representation does not interfere with the above mentioned network interpretations because their advantages and disadvantages are directly applicable to the bipartite case.

3. Available Data Resources and Their Usage

In the previous section, we reviewed six major graph representations for enzymatic reactions. Reaction data are available on many Web sites (Table 3). The original authoritative source is the IUBMB enzyme nomenclature, where EC numbers are assigned and updated (18). A reliable source of kinetic information is the BRENDA database, where detailed kinetic parameters and substrate specificity are available for each organism (19). Most metabolic information is based on these databases, and the network reconstruction usually undergoes organism-specific modification/curation. Discretized metabolic data should be applied with care because each representation has its original purpose. Here, we introduce a few important tips to be considered.

3.1. Different Databases Use Different Metabolite Names

Unfortunately, metabolite names are no more standardized than protein names. Standardization of metabolite names must precede network reconstruction. For example, α-ketoglutarate can be written as 2-oxoglutarate, oxoglutaric acid, 2-ketoglutarate, or α-ketoglutaric acid. The LIGAND section in the KEGG database

Table 3
Freely available data for reaction information

IUBMB Enzyme Nomenclature (http://www.chem.qmul.ac.uk/iubmb/enzyme/)	Enzyme names and reactions for each EC number with references
BRENDA (http://www.brenda-enzymes.org/)	Organism, Km value, isozyme, and substrate specificity information with references
MetaCyc (http://biocyc.org/metacyc/index.shtml)	Species-specific enzyme and pathway information for over 350 organisms with references
KEGG (http://www.genome.jp/kegg/reaction/)	Substrate–product information for EC reactions without resolving general names for over 1,100 organisms
BiGG (http://systemsbiology.ucsd.edu/In_Silico_Organisms/)	Genome-scale metabolic reconstruction for over 30 organisms

is a good source of such synonyms (20). It must be noted that many enzymatic reactions also use generic names, such as "alcohol" or "amino acid." Many reactions do not specify stereochemical information (e.g., serine instead of L-serine). Such discrepancies must be manually resolved by referencing genomic information in network reconstruction. Already curated data may be available for some model organisms (Table 3), but otherwise, manual curation is a necessary step. Not a few computational studies bulk-download reaction data from the KEGG- and other databases to reconstruct metabolic networks. It is not easy to verify the extent of curation from the brief description of their reconstruction policy, but without standardization, the results are not sufficiently accurate for an estimation of the global network properties.

3.2. Not All Connected Reactions form Pathways

A metabolic pathway consists of a sequence of reactions through which some structural moiety must be passed down. This prerequisite is not always satisfied by a graph path in the metabolic network. In Fig. 1b, for example, the link between L-glutamate and CoA, or between *N*-acetyl L-glutamate 5-semialdehyde and α-ketoglutarate is not a valid pathway because there is no carbon or nitrogen transfer between them. However, many studies computed the shortest paths in the substrate graph to identify the metrics of metabolic networks. It must be noted that such computation does not reflect true metabolic pathways although it may capture some general characteristics.

The directionality of reactions is another issue that causes headaches. There are many discrepancies between databases on directionality, and it is difficult to estimate directionality from enzyme structures or catalytic formulas. Kümmel et al. (21) proposed a heuristic algorithm to check reversibility based on the Gibbs energy, but Maskow and von Stockar (22) argued that the Gibbs energy does not facilitate determination of feasibility without considering intracellular ionic strength, the pH value, and the concentration of abundant metabolites. For now, it is safer to assume that most reactions are reversible; irreversible reactions are, for example, decarboxylation and cyclization reactions.

3.3. Kinetic Parameters Are Measured In Vitro, Not In Vivo

Kinetic parameters reported in the BRENDA and other databases were measured in vitro assuming the classic model of Michaelis and Menten (23). These parameters are good indicators for estimating the relative efficiency of enzymes; however, their true catalytic velocity within cells remains unknown. This is why not every kinetic parameter is directly applicable to simulation studies. Likewise, these values alone cannot indicate the reversibility of reactions. These observations justify the importance of qualitative studies because currently available parameters are not sufficient to fully depict intracellular conditions.

4. Network Applications: Constraint-Based or Graph-Theoretic Approaches

There are two mainstreams in utilizing reconstructed metabolic networks: the constraint-based- and the graph-theoretic approach. The former focuses on the metabolic capability of an organism in terms of metabolite fluxes and attempts to find the basis of the possible flux space. In contrast, the graph approach focuses on the metrics of metabolic networks and tries to understand flux in terms of pathways.

4.1. Constraint-Based Approach

This approach is best represented by flux balance analysis (FBA) and the elementary flux mode (EFM) (24, 25). In these strategies, the metabolic hypergraph is implemented as a matrix to solve its linear constraints, and the basis of the flux space is called extreme pathways or elementary modes, respectively (see Chapter 20) (26). The main constraints are the steady-state hypothesis for metabolites (no increase or decrease in their amounts) and the capacity constraints for reactions (directionality and the maximum value for flux). The elementary mode is more constrained in the formalization of a metabolic pathway and therefore harder to compute. The lack of scalability has been a bottleneck in the application of the EFM, but a new scalable approach was recently

presented by the EFM group (27). A detailed comparison between FBA and EFM is available elsewhere (28).

The resolution of generic names and the gap filling of major pathways are crucial steps for both strategies. Currently available reconstructions need substantial updates to be called complete. In the "complete" in silico strain (iJR904) of *E. coli*, for example, biosynthesis for 64 metabolites is reported to be flawed (29, 30). While gaps in many biosynthetic/degradation pathways remain, manual verification of completeness is difficult. The graph-theoretic representation of pathways is needed to fill these pathway gaps.

4.2. Graph-Theoretic Approach

To find metabolic pathways, the approach concentrates on the topological aspects of the network. A commonly used method is finding the shortest paths by tracking node connectivity (10, 12, 15–17) (also see Chapter 7). Since the direct application of the shortest path algorithm to the substrate graph produces a huge number of false-positive pathways, the substrate–product graph or its alternative with atomic information is the proper choice. This requirement is obvious especially when nitrogen or sulfur metabolism is considered. Another often overlooked problem is the consideration of molecular symmetry in pathway finding algorithms (11, 16). Since many molecules have symmetric atomic positions, simple connection of atomic mappings, such as the RPAIR information may miss important pathways. For example, simplistic connection cannot reproduce the carbon fate in the TCA cycle, where many symmetric molecules are involved. However, the consideration of symmetry information is closely related with the stereo-specificity of enzymes (e.g., pro-chirality) and is not easily solvable.

4.3. Identifying the Network Modularity

Some recent works focused on network modularity and its evolutionary implications in addition to pathways (14, 31). Strangely, the substrate–product graph has not been used to identify the modularity of metabolic networks; this may be attributable to the ambiguous definition of modularity.

A common pitfall in assessing network modularity is the use of KEGG metabolic maps as functional categories. These maps are drawn with aesthetic considerations; e.g., the same pathways are drawn multiple times for an easier understanding, and there is no well-defined criterion for omitting currency metabolites. Consequently, most maps contain multiple functional modules in metabolism. This observation is evident when the KEGG maps are compared with the MetaCyc pathway repository, the other comprehensive pathway resource for all organisms (see Chapter 11). The MetaCyc contains over 1,400 pathway modules in contrast to 160 KEGG maps (32). The lack of atom-resolved guidelines for the modularity of metabolites hampers a proper

comparison between KEGG maps (substrate graphs with duplication), MetaCyc pathways (bipartite graphs in LISP format), and the substrate–product representation for computing pathways. Therefore, the study of modularity remains at an immature stage and provides a number of interesting research topics.

5. Future Perspectives

Many researchers fail to recognize the proper representation of metabolic networks and their existing variations. It is crucial, especially for computational biologists, to understand the purpose of network representation and its appropriate use in a relevant context. To fully utilize the power of network analysis, more biological understanding is necessary. In a short term, we need to define biologically justifiable network modules. The use of KEGG pathway maps or similar classifications as the modularity standard is scientifically indefensible. Furthermore, a public database for atomic mappings that is easily understandable and applicable must be established to facilitate application of the substrate–product graph with atomic information by nonexperts. It should be noted that the substrate graph can be easily biased by the treatment of generic reactions and by the definition of currency metabolites (the KEGG RPAIR database does not recognize this problem). In view of the lack of standard names or identifiers even for basic compounds, however, it would be a challenging task to establish a common repository for metabolic pathways and associated information in a single institution. In this respect, a publicly shared portal may be a solution so that we can establish a community-based repository for metabolic information (33). In a longer term, the use of artificial chemical network to study the evolution of metabolism provides an interesting perspective because the fortuity of the real world renders evolutionary analysis possible only with such theoretical simulations (34). However, these trials should incorporate, at least, elemental information, such as carbon, nitrogen and sulfur, because their balance is crucial in understanding the metabolism.

Acknowledgments

The author thanks the editors and reviewers for valuable comments. This work is supported by JST-BIRD and Grant-in-Aid for Scientific Research on Priority Areas “Systems Genomics” from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

1. Barabási A-L, Albert R. (1999) Emergence of scaling in random networks. Science, 286(5439):509–512.
2. Liu B. (2007) *Web Data Mining – Exploring Hyperlinks, Contents and Usage Data*. Springer Series on Data-Centric Systems and Applications.
3. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabási A-I. (2000) The large-scale organization of metabolic networks. Nature, 407(6804):651–654.
4. Fell DA, Wagner A. (2000) The small world of metabolism. Nat Biotechnol, 18(11): 1121–1122.
5. Huss M, Holme P. (2006) Currency and commodity metabolites: their identification and relation to the modularity of metabolic networks. IET Syst Biol, 1(5):280–285.
6. Parter M, Kashtan N, Alon U. (2007) Environmental variability and modularity of bacterial metabolic networks. BMC Evol Biol, 7:169.
7. Wagner A, Fell D. (2001) The small world inside large metabolic networks. Proc R Soc Lond B, 268(1478):1803–1810.
8. Kreimer A, Borenstein E, Gophna U, Ruppin E. (2008) The evolution of modularity in bacterial metabolic networks. Proc Natl Acad Sci U S A, 105(19):6976–6981.
9. Ma H, Zeng A-P. (2003) Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. Bioinformatics, 19(2):270–277.
10. Rahman SA, Advani P, Schunk R, Schrader R, Schomburg D. (2005) Metabolic pathway analysis web service (Pathway Hunter Tool at CUBIC). Bioinformatics, 21(7):1189–1193.
11. Kotera M, Hattori M, Oh M-A, Yamamoto R, Komeno T, Yabuzaki J, Tonomura K, Goto S, Kanehisa M. (2004) RPAIR: a reactant-pair database representing chemical changes in enzymatic reactions. Genome Informatics, 15:P062 (poster abstract)
12. Arita M. (2003) In silico atomic tracing by substrate–product relationships in *Escherichia coli* intermediary metabolism. Genome Res, 13:2455–2466.
13. Arita M. (2004) The metabolic world of *Escherichia coli* is not small. Proc Natl Acad Sci U S A, 101(6):1543–1547.
14. Caetano-Anollés G, Kim HS, Mittenthal JE. (2007) The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture. Proc Natl Acad Sci U S A, 104(22):9358–9363.
15. Blum T, Kohlbacher O. (2008) Using atom mapping rules for an improved detection of relevant routes in weighted metabolic networks. J Comput Biol, 15(6):565–576.
16. Pitkanen E, Jouhten P, Rousu J. (2009) Inferring branching pathways in genome-scale metabolic networks. BMC Syst Biol, 3(1):103.
17. Faust K, Croes D, van Helden J. (2009) Metabolic pathfinding using RPAIR annotation. J Mol Biol, 388(2):390–414.
18. IUBMB Enzyme Nomenclature. [http://www.chem.qmul.ac.uk/iubmb/enzyme/]
19. Chang A, Scheer M, Grote A, Schomburg I, Schomburg D. (2009) BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. Nucleic Acids Res, 37(Database issue):D588–D592.
20. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y. (2008) KEGG for linking genomes to life and the environment. Nucleic Acids Res, 36(Database issue):D480–D484.
21. Kümmel A, Panke S, Heinemann M. (2006) Systematic assignment of thermodynamic constraints in metabolic network models. BMC Bioinformatics, 7:512.
22. Maskow T, von Stockar U. (2005) How reliable are thermodynamic feasibility statements of biochemical pathways? Biotechnol Bioeng, 92(2):223–230.
23. Symbolism and Terminology in Enzyme Kinetics (IUBMB Recommendation 1981). Enzyme Reactions & Inhibition. [http://www.chem.qmul.ac.uk/iubmb/kinetics/ek4t6.html].
24. Price ND, Reed JL, Palsson BØ. (2004) Genome-scale models of microbial cells: evaluating the consequences of constraints. Nat Rev Microbiol, 2(11):886–897.
25. 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(3):326–332.
26. Schilling CH, Schuster S, Palsson BØ, Heinrich R. (1999) Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. Biotechnol Prog, 15(3):296–303.
27. Kaleta C, de Figueiredo LF, Schuster S. (2009) Can the whole be less than the sum

of its parts? Pathway analysis in genome-scale metabolic networks using elementary flux patterns. Genome Res, 19(10):1872–1883.

28. Klamt S, Stelling J. (2003) Two approaches for metabolic pathway analysis? Trends Biotechnol, 21(2):64–69.
29. Reed JL, Vo TD, Schilling CH, Palsson BØ. (2003) An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). Genome Biol, 4:R54.
30. Satish Kumar V, Dasika MS, Maranas CD. (2007) Optimization based automated curation of metabolic reconstructions. BMC Bioinformatics, 8:212.
31. Guimerà R, Nunes Amaral LA. (2005) Functional cartography of complex metabolic networks. Nature, 433(7028):895–900.
32. Caspi R, Altman T, Dale JM, Dreher K, Fulcher CA, Gilham F, Kaipa P, Karthikeyan AS, Kothari A, Krummenacker M, Latendresse M, Mueller LA, Paley S, Popescu L, Pujar A, Shearer AG, Zhang P, Karp PD. (2010) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res, 38(Database issue):D473–D479.
33. Arita M. (2009) What can metabolomics learn from genomics and proteomics? Curr Opin Biotechnol, 20(6):610–615.
34. Hintze A, Adami C. Evolution of complex modular biological networks. PLoS Comput Biol, 4(2):e23.