

Strategy for Exploring Metabolic Pathways: Generation of Hypothetical Metabolic Network

Jun Ohta*

Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, Japan

Abstract

Development of metabolomics has made completion of metabolic map an important issue. Rational suggestion of specific metabolic pathways as candidates for real pathway will promote study to complete metabolic map. Exploring hypothetical metabolic networks containing candidate pathways enable this. Three different hypothetical metabolic networks abbreviated as BRP-dependent network, metabolite-dependent network and CFB network are considered. BRP-dependent and CFB networks are generated by network expansion from seed compounds via balanced reaction pattern (BRP) and via simple cleavage/formation of chemical bonds, respectively. Concept of network expansion, a term introduced by Heinrich's group, is generalized and a previous approach by Hatzimanikatis and his colleagues for generation of BRP-dependent network equivalent is re-defined in the context of network expansion. Metabolite-dependent network is generated from a given set of metabolites based on stoichiometry. Hypothetical metabolites and BRPs are considered to appear in BRP-dependent and metabolite-dependent networks, respectively. Concept of atom network is introduced.

Keywords: Metabolomics; Metabolic network; Hypothetical metabolites

Introduction

Development of metabolomics has been resulting in accumulation of detailed information of the entire metabolism. This increases importance of metabolic map as the basis of analysis of metabolome data and opportunity to validate the present knowledge of metabolic pathways. Study on metabolism has a long history. But, this does not guarantee completeness of the present metabolic map. There are genes which have not been assigned their function, whilst genome information is available in many organisms. Thus, novel enzyme reactions may be found as functions of those genes which do not have assigned function. In this circumstance, informatics technique to explore metabolic pathways including new reactions will be useful.

Viewing the entire metabolism as a network has been popular since the report by Barabasi's group [1]. Studies have been conducted about the global topological nature of metabolic networks. In most of those studies, metabolic network is defined as network of metabolites connected via enzyme reactions [1-3], where each reaction is decomposed and reduced to set of binary relationships between a pair of substrate and product, for example, substrate-product relationships. Binary relationships between substrate and product can be used for calculating metabolic pathway as metabolite sequence to connect 2 compounds. We call such an approach for exploring pathways binary relationship approach. Arita constructed a database for atomic tracing based on substrate-product relationship extracted from known enzyme reactions [4,5], where possible pathways can be found as combinations of established substrate-product relationships. RPAIR in KEGG describes pattern of chemical transformation between a pair of substrate and product in enzyme reactions [6], which may be used for prediction of pathways including hypothetical metabolites.

Whilst a given reaction is decomposed to a set of binary relationships between substrate and product, reconstruction of the original stoichiometric reaction only from the set of binary relationships between substrate and product is not an easy problem. This means that the original reaction may be more informative than the set of binary relationships between substrate and product. Thus,

alternative approaches which use characteristics of the entire of each individual reaction can be considered for exploring metabolic pathways, where the entire of hypothetical reactions are generated, then, hypothetical metabolic network is formed from the generated hypothetical reactions and established reactions. Metabolic pathways are explored on the resulting hypothetical network. Such approaches are called entire reaction approach in this paper. An approach proposed by Hatzimanikatis et al. [7] is an example of entire reaction approaches. In their approach, reaction patterns are extracted from enzyme reactions and applied to hypothetical substrates, resulting in generation of hypothetical reaction. In this paper, we show the concept of a new type of entire reaction approach where hypothetical reaction is generated from given or known metabolites only based on chemical stoichiometry. Then, we theoretically characterize difference between our new approach and the approach by Hatzimanikatis and his colleagues in resulting hypothetical reactions and metabolic networks and show complementary relationship between these 2 entire reaction approaches. Further, we introduce concept of chemical reaction and metabolic network as atom network to describe all the atom-level information concerning reaction/network.

Hypothetical Reactions

General consideration on chemical reaction

Apparently, in chemical reaction, cleavage of plural chemical bonds in the given reactant(s) occurs and, then, each of the resulting edges of the cleaved bonds gets connection with a new partner edge

*Corresponding author: Dr. Jun Ohta, Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, 2-5-1 Shikatacho, Okayama 700-8558, Japan, Tel: +81-86-235-7125; Fax: +81-86-235-7126; E-mail: jo25@md.okayama-u.ac.jp

Received December 02, 2014; Accepted December 20, 2014; Published January 21, 2015

Citation: Ohta J (2015) Strategy for Exploring Metabolic Pathways: Generation of Hypothetical Metabolic Network. J Comput Sci Syst Biol 8: 055-058. doi:10.4172/jcsb.1000171

Copyright: © 2015 Ohta J. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

different from the previous one to form a new chemical bond, which is called cleavage and formation of chemical bonds and abbreviated as CFB. Products are formed as a result of a CFB of the given reactants. Chemical structure of the product(s) can be deduced from the structure of the reactant(s) if the CFB in the reaction is specified. In a chemical equation, a one-to-one correspondence exists between the set of the atoms included in the reactant(s) and the set of atoms included in the product(s) as the basis of stoichiometry, because each atom in the reactant(s) is transformed into a specific atom in the product(s). Thus, complete description of a chemical reaction gives us the following 4 kinds of information:

- i. Chemical structure of the reactant(s)
- ii. Chemical structure of the product(s)
- iii. Chemical bonds cleaved or formed by CFB
- iv. Atom-to-atom correspondence between the reactant(s) and product(s)

In this article, connection between 2 atoms through multiple bonds is counted as assembly of separate single bonds and aromatic structure is considered as resonance or rapid inter-conversion between/among different structures. Stereochemistry of compounds is not considered at the present stage of study.

Atom network structure of chemical reaction and metabolic network

We introduce concept of chemical reaction as atom network by explaining atom network structure of chemical reaction. In the atom network structure of chemical reaction, nodes are atoms included in the reactant(s) and product(s), and edges are 2 different types of atom-to-atom connection, that is chemical bonds which define the chemical structure of the reactant(s) and product(s) and atom-to-atom correspondences between the reactant(s) and product(s). Atom-to-atom correspondences as edges are sub-classified and distinguished depending on which of reactant-product binary relationships the correspondence belongs to. Information of atom network structure includes chemical structure of the reactant(s) and product(s), which is defined by atom-to-atom connection via chemical bonds, as well as atom-to-atom correspondence between the reactant(s) and product(s). Therefore, when atom network structure of a reaction is given, chemical bonds cleaved by CFB are identified as those which exist between a pair of atoms in the reactant(s), but, not between the corresponding atoms in the product(s). Similarly, chemical bonds formed by CFB are identified as those which exist between a pair of atoms in the product(s), but, not between the corresponding atoms in the reactant(s). Thus, atom network structure of a chemical reaction includes information of chemical bonds cleaved or formed by CFB, giving us all of the above 4 kinds of information about the reaction. Ordinary description of chemical reaction does not indicate whether atom-to-atom correspondences between the reactant(s) and product(s) are considered or not. We can avoid such ambiguousness by using the term atom network structure.

Concept of chemical reaction as atom network leads us to concept of metabolic network as atom network. In the atom network structure of metabolic network, nodes are atoms included in metabolites, and edges are chemical bonds which define the structure of metabolites and atom-to-atom correspondences via reaction between metabolites. Atom-to-atom correspondences as edges are further classified and distinguished depending on which reaction each edge belongs to.

Atom network works a mathematical object which carries information both of chemical structure and atomic tracing. Atom network view is useful for study to integrate chemical structure and atomic tracing.

Balanced reaction pattern (BRP)

The chemical bonds cleaved or formed by CFB in an enzyme reaction depend on reaction specificity of the responsible enzyme. We introduce balanced reaction pattern, BRP, as a carrier of information about the surroundings of the chemical bonds affected by CFB. BRP is defined as the portion of the reactant(s) including all the atoms connected via bonds affected by CFB and its corresponding part of the product(s). BRP is stoichiometric and has atom network structure. The size and boundary of BRP is arbitrary and depends on the purpose of BRP usage. The minimum BRP is the portion of reactant(s) composed just of all the atoms connected via bonds affected by CFB and its corresponding part of product(s). The maximum BRP is the entire of the reaction. The boundary of BRP can be selected to specify the corresponding CFB uniquely in a given reaction.

Generation of hypothetical reactions

It is desirable that hypothetical reactions as candidates for new biochemical reactions keep certain biological character. Two different approaches for generation of such biological hypothetical reactions, BRP-dependent reaction generation and metabolite-dependent reaction generation, are described. CFB reaction generation is also introduced as procedure for generation of general hypothetical reactions.

BRP-dependent reaction generation was proposed by Hatzimanikatis and his colleagues. We call their approach BRP-dependent reaction generation. Enzyme reaction has substrate specificity and reaction specificity. Substrate specificity is sometimes not so strict. Thus, reaction generation based on known reaction specificity will result in biological hypothetical reaction. Appropriate BRP is expected to be able to express reaction specificity of enzyme. Further, enzyme categories at different depths in the EC number classification are expected to be expressed by using the different sizes of BRP. In the approach by Hatzimanikatis and his colleagues [7], biological BRP equivalents are extracted from reaction formulae in the EC number classification and used to generate hypothetical reactions by applying the BRP equivalents to hypothetical substrates which is not the original substrates but have the substructure near the CFB site included in BRP.

Metabolite-dependent reaction generation is a new approach we propose in this paper. In this approach, hypothetical reactions are generated as a stoichiometrically possible arrangement of given compounds, usually metabolites or biological compounds, in reaction formulae. Outline of a possible procedure for metabolite-dependent reaction generation is shown below together with a simple example to help understanding of the concept of this approach.

Step 1. Give a set of metabolites to be used as substrates or products.

Glucose- $C_6H_{12}O_6$

Glucose 6-P (G6P)- $C_6H_{13}O_9P$

Phosphoric acid (P_i)- H_3PO_4

Water (H_2O)- H_2O

Step 2. List up reaction candidates to be considered as hypothetical reaction



Step 3. Select stoichiometrically possible reaction candidates.



At the end of Step 3, the reaction candidates selected do not have atom network structure. The atom network structure is obtained only after atom-to-atom correspondences between the reactant(s) or substrate(s) and product(s) etc. are assigned. BRP can be deduced from the obtained atom network structure. Atom network structure, especially its BRP, can be criteria for further selection of real candidate for biochemical reaction. This criterion is arbitrary, but should be determined depending on the purpose of candidate selection.

In BRP-dependent reaction generation, generated hypothetical reactions keep biological character in the fact that their BRP are derived from biochemical reactions. Hypothetical substrates or products may appear, but any hypothetical BRPs not. In metabolite-dependent reaction generation, generated hypothetical reactions keep biological character in the fact that all of their substrates and products are biological compounds. Hypothetical BRP may appear, but any hypothetical substrates or products not. BRP-dependent reaction generation and metabolite-dependent reaction generation are complementary in the character of generated hypothetical reactions.

CFB reaction generation means hypothetical reaction generation by applying a CFB to given reactant(s). The total number of chemical bonds seen in the given reactant(s) is countable, indicating all the possible CFBs can be listed up. Thus, all the possible hypothetical reactions using given reactant(s) can be calculated theoretically. While the number of chemical bonds affected by CFB is small, computation of this calculation would not require so much CPU power. Such computation would be useful for giving atom network structure to hypothetical reaction by metabolite-dependent reaction generation as well as established enzyme reactions. Atom network structure of enzyme reaction can be used for extraction of its BRP.

Hypothetical metabolic networks

Hypothetical metabolic networks and generalization of network expansion

Two approaches for biological hypothetical reaction generation and a method to generate general hypothetical reactions are shown above. Each of them can be used for generation of hypothetical metabolic network corresponding to the approach or method. New metabolic pathways exist on such hypothetical metabolic networks.

Hatzimanikatis and his colleagues introduced generation of hypothetical metabolic network using the approach we call BRP-dependent reaction generation [7]. On the other hand, Heinrich's group proposed concept of network expansion for analysis of architecture of metabolic networks [8,9]. The former intended to synthesize networks, whereas the latter intended to analyze networks. However, correspondences are observed between the above two. Therefore, we propose to generalize network expansion and understand hypothetical network generation by Hatzimanikatis and his colleagues from the context of network expansion. For hypothetical network generation, the expansion process is defined by the following algorithm.

- Selection of one or more biochemical compounds acting as a seed of the expanding network. The seed represents the first generation in the series of expanding networks.
- 2' Generation of those hypothetical reactions by a specific rule A which use as substrates only compounds already present in the current network.
- Incorporation of the generated reactions and their products into the network. This results in the next generation of expanding networks.
- 4' Repetition of steps b and c

2' and 4' are modification of the original definition [9]. Under this definition of network expansion, hypothetical network generation by Hatzimanikatis and his colleagues is understood as network expansion where rule A is BRP-dependent reaction generation using biological BRPs. We use BRP expansion to indicate hypothetical network generation by Hatzimanikatis and his colleagues. Sets of seed compounds to fully recover metabolic network composed of established reactions by the original network expansion can be identified. Assume BRP expansion from such a set of elementary seed compounds is performed using the set of biological BRPs covering all the known enzyme reaction. This BRP expansion results in hypothetical metabolic network composed all the established reactions and hypothetical reactions. We name this hypothetical network BRP-dependent network.

Generation of hypothetical metabolic networks subsequent to CFB reaction generation can also be understood in the context of network expansion, where rule A is CFB reaction generation, and we call this CFB expansion. Theoretically, CFB expansion from elementary seed compounds can result in hypothetical metabolic network including all the possibilities of chemical structure and reaction pattern. We name this hypothetical network CFB network.

Generation of hypothetical metabolic networks subsequent to metabolite-dependent reaction generation is called stoichiometric metabolic network generation. Assume that metabolic network composed of established reactions. Stoichiometric metabolic network generation from all the compounds included in metabolic network composed of established reactions results in hypothetical metabolic network including all the established reactions and hypothetical reactions. We call this hypothetical network metabolite-dependent network.

BRP-dependent network and metabolite-dependent network overlaps each other. Reactions are classified into the following 3 categories:

Type 1. Reaction belongs to both BRP-dependent and metabolite-dependent networks.

Type 2. Reaction belongs to BRP-dependent network, but not to metabolite-dependent network.

Type 3. Reaction belongs to metabolite-dependent network, but not to BRP-dependent network.

All the established reactions are classified in type 1 reaction. Type 1 reactions have known BRP and are composed of known metabolites as substrate(s)/product(s). Therefore, hypothetical type 1 reactions, if any, are good candidate for real biochemical reaction. Type 2 reactions use hypothetical metabolites as substrates or products. Type 3 reactions have hypothetical BRP. Existence of types 2 and 3 reactions reflects complementary relationship between BRP-dependent network and

metabolite-dependent network. It is of interest whether types 2 and 3 reactions are included in real metabolic network or not. See the supplement for an example of classification of reactions.

Discussion

In this paper we have introduced metabolite-dependent reaction generation and stoichiometric network generation as a new strategy for generation of hypothetical reaction and metabolic network. The approach by Hatzimanikatis and his colleagues have been re-defined by generalizing the concept of network expansion. This may give a chance of utilization of results from network analysis by network expansion for the approach by Hatzimanikatis and his colleagues, BRP-dependent reaction generation and BRP expansion. Comparison between our new approach and the approach introduced by Hatzimanikatis and his colleagues clearly indicates complementary relationship of these 2 approaches both at the reaction level and at the network level.

Both BRP in BRP-dependent reaction generation and criteria for reaction selection in metabolite-dependent reaction generation are arbitrary. Modification of these can be tried to increase similarity of the resulting hypothetical networks to network composed of established reactions. Cycle of modification and check of similarity in such effort would teach us about principle for organization of metabolic network as well as increases the quality of hypothetical networks where metabolic pathways are explored. Existence of known substrate-product relationship and the number of chemical bonds cleaved in the responsible CFB as well as similarity of BRP to known biological BRP can be criteria for reaction selection in metabolite-dependent reaction generation,

Concept of metabolic network as atom network was introduced. Methodology to describe information about atom-to-atom correspondence between a pair of substrate and product as a matrix was introduced [10,11]. Arita has made a software for genome scale atomic tracing equipped with visualization system of chemical structure [4,5]. Thus, importance of atom-level consideration has been recognized. However, metabolic networks have generally been analyzed as network of metabolites [1-3]. It is difficult to describe all the atom-level information both of chemical structure and atomic tracing in the architecture of network of metabolites. In contrast, it is easy to describe the atom-level information in the architecture of atom network or network of atoms. In atom network as a mathematical object, 2 types of edges, chemical bonds and atom-to-atom correspondences, are of the same importance. This is good for study of relationship between these 2 types of edges. We expect that mathematical study of atom network structure provides new insights in metabolic network study.

Reference network is a useful tool for studies on general networks. What is a reference network suitable for studies on metabolic networks is not an easy question, because there is constraint in metabolic network structure, for example, stoichiometric constraint in metabolite network structure. CFB network and metabolite-dependent network are introduced in this paper. CFB network can be understood as

most dense atom network and metabolite-dependent network can be understood as most dense metabolite network composed of the same set of metabolites. These networks can be the basis to develop rational reference metabolic network.

References

1. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi AL (2000) The large-scale organization of metabolic networks. *Nature* 407: 651-654.
2. Ma H, Zeng AP (2003) Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. *Bioinformatics* 19: 270-277.
3. Wagner A, Fell DA (2001) The small world inside large metabolic networks. *Proc Biol Sci* 268: 1803-1810.
4. Arita M (2003) In silico atomic tracing by substrate-product relationship in *Escherichia coli* intermediary metabolism. *Genome Res* 13: 2455-2466.
5. Arita M (2004) The metabolic world of *Escherichia coli* is not small. *Proc Natl Acad Sci USA* 101: 1543-1547.
6. Kotera M, Okuno Y, Hattori M, Goto S, Kanehisa M (2004) Computational Assignment of the EC Numbers for Genomic-Scale Analysis of Enzymatic Reactions. *J Am Chem Soc* 126: 16487-16498.
7. Hatzimanikatis V, Li C, Ionita JA, Henry CS, Jankowski MD, et al. (2005) Exploring the diversity of complex metabolic networks. *Bioinformatics* 21: 1603-1609.
8. Ebenhoh O, Handorf T, Heinrich R (2004) Structural analysis of expanding metabolic networks. *Genome Inform* 15: 35-45.
9. Handorf T, Ebenhoh O, Heinrich R (2004) Expanding Metabolic Networks: Scopes of Compounds, Robustness, and Evolution. *Genome Inform* 15: 35-45.
10. Ohta J (2006) Connectivity matrix method for analyses of biological networks and its application to atom-level analysis of a model network of carbohydrate metabolism. *IEE Proc Syst Biol* 153: 372-374.
11. Zupke C, Stephanopoulos G (1994) Modeling of isotope distributions and intracellular fluxes in metabolic networks using atom mapping matrices. *Biotechnol Prog* 10: 489-498.