Search Pathways for Secondary Metabolites Based on Genome Analysis and Chemical Knowledge on Metabolism

Takaaki Nishioka [1] **Kimito Funatsu** [2]

nishioka@scl.kyoto-u.ac.jp funatsu@tutkie.tut.ac.jp

[1]Graduate School of Agricultural Sciences, Kyoto University, Oiwake-cho, Kitashirakawa, Kyoto 606-8502, Japan

[2]Department of Knowledge-Based Information Engineering, Toyohashi University of Technology, 1-1 Tempaku, Toyohashi, Aichi 441-8580, Japan

1 Introduction

To estimate the whole aspects of metabolic activity of a biological species, the first step is to reconstruct all the possible metabolic pathways from the enzymes coded by the genes on the genome. For the basic metabolites that are common products in various organisms, pathways for their mutual chemical conversion are well studied and have been collected as the PATHWAY and REACTION databases in Kyoto Encyclopedia of Genes and Genomes (KEGG) [2, 3]. On the other, those remain unknown for the secondary metabolites that are synthesized from basic metabolites via species specific pathways. Prediction of biosynthesis pathways for them might not only reveal the enzymes catalyzing the reactions and the metabolic intermediates, but also be applicable to design novel bioactive compounds through gene engineering of microorganisms.

Pathcomp developed and implemented on KEGG by Goto [2, 6] is the system that is now available for practical use of searching pathways. In their system, every pair of two metabolites that are mutually converted by one enzymatic reaction is extracted as a binary relation and registered in the REACTION database. When two compounds that are mutually converted via unknown pathway are given, it generates all possible pathways by connecting binary relations available. Their system, however, is not applicable to most secondary metabolites because starting basic metabolites are unknown. This is the reason why we develop another system that searches both the pathway and the starting basic metabolite for a given secondary metabolite.

2 Concept of System

Two procedures must be included in the system. The one is to list all the possible reaction sites in the chemical structure of a given compound. Reaction sites are usually on the functional groups, heteroatoms, and chemical bonds connecting them. The other is to rank the relative reactivity for the reaction sites listed. In this procedure, knowledge of metabolic reactions evaluates the reactivity at each site by comparing the structure of the given compound with the substrate specificity of enzymes. For this purpose, knowledge database is designed to analyze and to accumulate the reaction rules on the substrate specificity and reaction type for well-characterized enzyme reactions.

In the field of organic synthesis, a method of neural network developed by Kohonen [5] has been successfully applied for learning the reaction rules of various organic reactions [7]. This method has never been applied to enzyme catalyzed reactions.

In the present study, we apply Kohonen self-organizing network simulator TUT-SOM [1], which is originally devised for data classification of organic reactions, to enzyme catalyzed reactions. Learning in TUT-SOM repeats the following three steps; selection of a reference vector from a set of input vectors, matching of weight vectors to input vectors, and modification of weight vectors. TUT-SOM also supports counterpropagation method developed by Hecht-Nielsen [4].

3 Results and Discussions

In the application of Kohonen self-organizing network to the enzyme catalyzed reactions, the following problems are revealed.

(1) In organic synthetic reactions, two reactants locally interact with each other only at their reaction sites. Input parameters are quantitative physicochemical parameters such as electron density, ionization potential, and electron affinity of the reaction sites of reactants and products. In enzyme catalyzed reactions, two reactants first form a complex with enzyme molecule, then they interact with each other. Critical step is the complex formation, in which the enzyme strictly recognizes the whole structures of the reactants. Input parameters must include global structural features of the reactant molecules as well as local ones near the reaction sites. It is difficult to find proper parameters describing the global structural features.

(2) To extract reaction rules for one enzyme reaction, data should be collected for the reactions with the same enzyme and more substrates with different structures than the number of input parameters. In the databases, only limited number of data are collected for each enzyme; reactions for one or two different substrate analogues because of strict substrate specificity. This makes it difficult to extract reaction rules.

Acknowledgments

This work is supported in part by a Grant-in-Aid (10168101) for Scientific Research on Priority Areas from The Ministry of Education, Science, Sports and Culture in Japan.

References

[1] Funatsu, K., *Manual for TUT-SOM*, 1997.

[2] Goto, S., Bono, H., Ogata, H., Fujibuchi, W., Nishioka, T., Sato, K., and Kanehisa, M., Organizing and Computing Metabolic Pathway Data in terms of Binary Relations. *Pacific Symposium on Biocomputing '97*, (Ed. by Altman, R. B. and Dunker, A. K. and Hunter, L. and Klein, T. E.), World Scientific, 175–186, 1997.

[3] Goto, S., Nishioka, T. and Kanehisa, M., LIGAND Database for Enzymes, Compounds, and Reactions, *Nucleic Acids Research*, *in press.*

[4] Hecht-Nielsen, R., Counterpropagation Networks, *Appl. Optics.*, 26:4979–4984, 1987.

[5] Kohonen, T., Self-Organized Formation of Topologically Correct Feature Maps, *Biol. Cybern.*, 43:59–69, 1982.

[6] Ogata, H., Goto, S., Fujibuchi, W. and Kanehisa, M., Computation with the KEGG pathway database, *Biosystem*, 47:119–128, 1998.

[7] Satoh, H., Sacher, O., Nakata, T., Chen, L., Gasteiger, J. and Funatsu, K., Classification of Organic Reactions: Similarity of Reactions Based on Changes in the Electronic Features of Oxygen Atoms at the Reaction Sites, *J. Chem. Inf. Comput. Sci.*, 38:210–219, 1998.