Chemical Knowledge for Predicting Biosynthetic Pathways of Secondary Metabolites

Takaaki Nishioka [1] nishioka@scl.kyoto-u.ac.jp

Kimito Funatsu [2] 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

Metabolism is a chemical aspect of life that produces metabolites and energy by chemical reactions catalyzed by enzymes. Metabolic reactions that synthesize basic metabolites such as nucleic acids, carbohydrates, and amino acids are shared by various organisms. Those that synthesize secondary metabolites such as antibiotics, toxins, dyes and hormones are specific to individual organisms. Most of the known secondary metabolites have biological activities that are valuable for medical and agricultural uses. While basic metabolic pathways are well studied and collected in the LIGAND, PATHWAY, REACTION databases in Kyoto Encyclopedia of Genes and Genomes [2, 3], pathways for the biosynthesis of secondary metabolites are only partly or scarcely studied. Genome sequence analysis provides a source of the information necessary for predicting the biosynthesis pathways for secondary metabolites because the sequence analysis could reveal all the enzymes specific to each organism from their genes coded on the genome.

Metabolic pathway is reconstructed by predicting intermediate reactions stepwise in the direction reverse to biosynthesis; from a given secondary metabolite to a starting basic metabolite. At each prediction step, more than one reactions are usually proposed as the candidates for the next intermediate reaction because each intermediate metabolite usually have several possible reaction sites in its chemical structure. To select the most probable reaction, knowledge is required to evaluate the chemical reactivity of the candidates. We have analyzed metabolic reactions to obtain chemical knowledge on the reactions in the cells. We are accumulating two different types of chemical knowledge for the prediction of metabolic pathways. The one is related to the substrate and reaction specificity of the enzymes. The other type of chemical knowledge is related to the preference in the chemical reactions in the cells. These two are supplemental to each other.

2 Methods

To accumulate the first type of chemical knowledge, the enzyme catalyzed reactions in the KEGG databases were analyzed by the Kohonen self-organizing network simulator TUT-SOM [1, 4] that is originally devised for data classification of organic reactions and has been successfully applied for learning their reaction rules.

To extract the second type of chemical knowledge, the PATHWAY database was manually analyzed to find the reactions favored in the biological system; that is, to find a set of the reactions frequently used in the cells.

3 Results and Discussions

We defined a set of 66 structural features, for example OH, NH2 and COOH, as the chemical descriptor of metabolites. We tried to express the substrate specificity of each enzyme as a combination of the structural features characteristic to its substrates and products. TUT-SOM simulator was applied to extract such features from 284 metabolites of 66 enzyme reactions in the PATHWAY database. Learning by the simulator are not completed because the number of compounds allowed for each enzyme as a substrate is much less than that necessary for the learning.

The PATHWAY database was manually inspected to find the second type of chemical rules. Several rules are extracted. For example, four consecutive reactions catalyzed by different enzymes, (1) addition of acetyl group at the carbonyl carbon, (2) elimination of water, (3) addition of water, and (4) oxidation of OH group, are included in, at least, three different metabolic pathways; in the metabolisms of oxaloacetate to 2-oxoglutarate in the TCA cycle, 2-oxoglutarate to 2-oxoadipate in the lysine biosynthesis, and 2-oxoisovalerate to 2-oxo-4-methylvalerate in the leucine biosynthesis. This common four-step transformation could be a series of chemical reactions preferred in the cells or an evolutionally conserved unit of metabolism.

Chemical transformations commonly found in more than one metabolic pathways can be automatically detectable as the conserved sequence of enzyme EC numbers. In the above example, the consecutive reactions are expressed as a sequence of the EC numbers of the enzymes that catalyze the reactions; 4.1.3.X, 4.2.1.X, and 1.1.1.X, in which X means any number. When metabolic pathways are described as a sequence of EC numbers rather than a sequence of metabolites, any sequence of EC numbers conserved among different pathways could be detected by applying a method for searching sequence motifs.

The second type of chemical knowledge is favorable in the pathway prediction because it can predict a sequence of several reaction steps as a set.

Acknowledgments

This work is supported by a Grant-in-Aid (11149101) 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*, 27:377–379, 1999.

[4] Kohonen, T., Self-organized formation of topologically correct feature maps, *Biol. Cybern.*, 43:59–69, 1982.