

Analysis of Reactive Modules in the Metabolic Pathways

Masaaki Kotera

kot@kuicr.kyoto-u.ac.jp

Susumu Goto

goto@kuicr.kyoto-u.ac.jp

Masahiro Hattori

hattori@kuicr.kyoto-u.ac.jp

Minoru Kanehisa

kanehisa@kuicr.kyoto-u.ac.jp

Bioinformatics Center, Institute for Chemical Research, Kyoto University, Uji, Kyoto
611-0011, Japan

Keywords: enzymatic reaction, chemical compound, metabolic pathway

1 Introduction

Huge amount of putative genes have been unveiled by sequencing of genomes for various kinds of organisms, and the biological interest has shifted to the whole picture of the network and its dynamics of molecular interactions. The metabolic pathway consists of two major components: the enzymes that catalyze biological reactions and the chemical compounds that are converted by enzymes. Hitherto, the EC (Enzyme Commission) number has frequently been used as an efficient way to describe relationships between enzymes and chemical compounds, to give systematic names to enzymes, and to study various biological reactions and metabolic pathways. However, the EC numbers also contain several kinds of problems, which cause difficulties to analyze common reaction features in the pathways. This is because the EC number is supposed to be derived from reaction mechanisms but actually contains features about enzyme molecules and genes. To avoid these difficulties we have been developing de novo classification scheme of enzymatic reactions [4] that focuses only on the reaction mechanisms. The classification is based on the physicochemical environment of atoms [2] during reactions.

In this study, we analyze the metabolic pathways from the viewpoint of finding clusters of enzymatic reactions along the pathways, as one of the applications of our classification scheme. Here, we report the overall picture of this analysis and show some interesting properties, which we call reactive modules in the metabolic pathways.

2 Method and Results

The classification scheme we took in this study had already been described [4], where we prepared the initial dataset of chemical compounds and biological reactions from the KEGG/LIGAND database [1]. In this classification, each ID number represents the reaction characteristics and consists of three numbers that are hierarchically assigned as the EC numbers. The first numeral of numbering stands for the classification of reactions based on the organic chemistry; 1: nucleophilic substitution, 2: nucleophilic addition-elimination, 3: electrophilic aromatic substitution, 4: nucleophilic addition, 5: electrophilic addition, 6: electrophilic addition to aromatic compounds, 7: coordination to metals, and 8: stereometric or geometric isomeration. The second numerals correspond to chemical bonds that have changed over the reactions, and the third numerals represent unchanged atoms that are adjacent to the changed atoms. We analyzed all metabolic pathways in the KEGG PATHWAY database [3] to perform analysis of reactive modules, but here we show only the result for the glycolysis pathway. We demonstrate that reactive modules can be found by converting the EC numbers to our classification codes on each pathway map, as illustrated in Fig. 1. In this figure, any triplet of numbers represents the RC (Reaction Classification) number in our classification scheme.

Interestingly, we can point out that there are many enzymes with different EC numbers but with the same RC numbers; for example, EC5.4.2.2, 2.7.1.69 and 3.1.3.10 are classified into the same class of reactions in our scheme and all of those enzymes represent phosphorylation, denoted by RC 1.29.53

or 1.29.98. It is not easy to grasp the same functionality during reactions only from the EC numbers, which are based on subjective assignments by humans. In contrast, it is much easier to understand such a feature by our sophisticated classification that is based on computational comparisons of chemical structures. We have found that the groups of the same or similar RC numbers tend to be localized on the pathway maps; namely, the pathway maps may be divided into several sub-pathways on the basis of commonly used reactive modules representing continuous and closely correlated reaction steps along the pathways. For example, glycolysis consists of several modules of reaction mechanisms and there are two major modules: the phosphorylation module (1.29.53 and 1.29.98, surrounded by dotted lines in Fig. 1) and the addition/elimination module (5.71.150 and 5.57.131). Thus, it was interesting to observe that similar reactive mechanisms are localized and correlated along the pathway.

In conclusion, our classification scheme enables us to analyze similar reaction steps in the metabolic pathways and to identify reactive modules in different pathways. Now we are performing further analyses to better understand what types of reactions tend to form reactive modules and which pathways contain such modules.

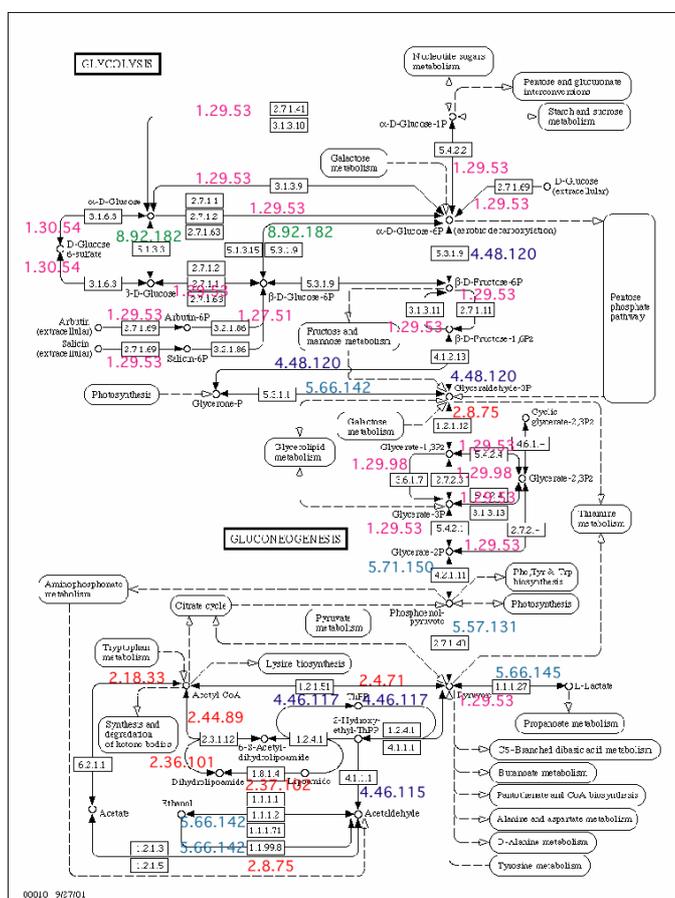


Figure 1: Example of the reaction characteristics on the metabolic pathway.

References

- [1] Goto, S., Okuno, Y., Hattori, M., Nishioka, T., and Kanehisa, M., LIGAND: database of chemical compounds and reactions in biological pathways, *Nucleic Acids Res.*, 30:402–404, 2002.
- [2] Hattori, M., Okuno, Y., Goto, S., and Kanehisa, M., Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways, *J. Am. Chem. Soc.*, 125:11853–11865, 2003.
- [3] Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M., The KEGG resource for deciphering the genome, *Nucleic Acids Res.*, in press.
- [4] Kotera, M., Okuno, Y., Hattori, M., Goto, S., and Kanehisa, M., Classification and analysis of KEGG/REACTION database, *Genome Informatics*, 13:490–491, 2002.