

RPAIR: a reactant-pair database representing chemical changes in enzymatic reactions

Masaaki Kotera
kot@kuicr.kyoto-u.ac.jp

Masahiro Hattori
hattori@kuicr.kyoto-u.ac.jp

Min-A Oh
mina@kuicr.kyoto-u.ac.jp

Rumiko Yamamoto
rumiko@scl.kyoto-u.ac.jp

Tomoko Komeno
ktomoko@scl.kyoto-u.ac.jp

Junko Yabuzaki
yzjuno@scl.kyoto-u.ac.jp

Koichiro Tonomura
tonomura@scl.kyoto-u.ac.jp

Susumu Goto
goto@kuicr.kyoto-u.ac.jp

Minoru Kanehisa
kanehisa@kuicr.kyoto-u.ac.jp

Bioinformatics Center, Institute for Chemical Research, Kyoto University, Uji, Kyoto

Keywords: secondary metabolism, enzymatic reaction, chemical compound, RC (Reaction Classification)

1 Introduction

Basic metabolism involves all metabolic pathways necessary for the survival of cells. Almost all of the chemical compounds in these pathways are already identified, and thus our interest is to investigate how they act to maintain dynamic systems such as living organisms. On the other hand, secondary metabolism is seen in differentiated cells in special species. Antibiotics in fungi is one example which is important for medical and many other purposes. They are believed to have some advantages for individuals as a whole. However, they are not necessary for each cell to survive, and it is often not understood why they are synthesized. Only small portions of the pathways involving secondary metabolism are known despite many efforts.

Currently, large numbers of compounds are registered in such databases as KEGG/LIGAND [1, 2], in the hopes of uncovering the whole picture and the basic rules governing secondary metabolism. In the present, however, the proportion of compounds involved in known enzymatic reactions or metabolic pathways are less than half of those registered in these databases. Most of these compounds are supposed to be catalyzed with the same types of reactions as those in the basic metabolism pathways. Thus, collecting the known chemical changes of basic metabolites can help us to predict the types of enzymatic reactions occurring in secondary metabolism as well as the possible intermediates in synthetic and degradation pathways.

We have already constructed the RC (Reaction Classification) method [3] based on reactant-pair relationships, using a computerized method of chemical structure comparison (molecular alignment) [4]. A reactant pair is defined as a pair of a substrate and a product that preserve chemical substructures through enzymatic reactions. We extracted "reaction center" atoms, which are the atoms directly involved in the reactions. This is a powerful scheme to systematically and comprehensively analyze enzymatic reactions on a genomic scale. However, there were some cases where biologically suitable molecular alignments could not be obtained without any knowledge of enzyme preferences. Therefore, we have developed a new database containing manually compiled reactant pairs and molecular alignments incorporating biochemical knowledge.

2 Method and Results

6,261 reactions in the LIGAND database are decomposed into 9,125 reactant pairs (without redundancy). Detailed information describing the enzymatic reactions as combinations of reactant pairs is written in the original article [3]. Reactant pairs are classified into five categories; (1) main-pairs, describing main changes on substrates, (2) cofac-pairs, describing changes on cofactors for oxidoreductases, (3) trans-pairs, focused on transferred groups for transferases, (4) ligase-pairs, describing the consumption of nucleoside triphosphates for ligases, and (5) leave-pairs, describing the separation or addition of inorganic compounds for such enzymes as lyases and hydrolases. The numbers of these pairs are 5699, 78, 1518, 13, 1817, respectively.

All atoms consisting of each reactant pair except hydrogen atoms are represented in KCF (KEGG Chemical Function) format [4]. Every atom in the former compound is manually assigned, if possible, to a corresponding atom in the latter compound, which on the whole represents the transformation of compounds.

Fig. 1 shows an example of the phosphorylation of a sugar. In the Structure section, the chemical structures of the two compounds are represented. The preserved atoms are given serial identification numbers and are highlighted in green. Corresponding alignment information in text is also described in the KCF section.

3 Discussion

Manually compiled molecular alignments reliably improve the definition of reaction center atom and the analysis of substrate specificity for each enzymatic reaction with our RC method. They also improve the prediction of unknown metabolic pathways as mentioned in the Introduction.

The main advantage of manual alignments over computational alignments is clearly observed especially in pairs involving compounds with high symmetry as well as in cases where the number of preserved atoms is much lower than all the atoms consisting of a reactant pair. For example, the manual alignments successfully retrieved the correct trans-pair of certain acyl-transferase reactions (EC: 2.3.1), whereas the computational alignment was unable to do so.

There are some cases where different alignment results may actually be allowed. Once this issue is addressed, the RPAIR database can also be applied to flux analyses in certain cellular processes.

4 Accessibility and Availability

The RPAIR database is available via the DBGET system in KEGG/LIGAND [1].

Acknowledgements

This work was supported by grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan, the Japan Society for the Promotion of Science, and the Japan Science and Technology Corporation. The computational resource was provided by the Bioinformatics Center, Institute for Chemical Research, Kyoto University.

References

- [1] Goto, S., Nishioka, T., and Kanehisa, M., LIGAND: chemical database for enzyme reactions, *Bioinformatics*, 14: 591-599, 1998, <http://www.genome.jp/kegg/ligand.html>
- [2] Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M., The KEGG resource for deciphering the genome, *Nucleic Acid Res*, 32: D277-D280, 2004.
- [3] Kotera, M., Okuno, Y., Hattori, M., Goto, S., and Kanehisa, M., Computational Assignment of the EC numbers for Genomic-scale Analysis of Enzymatic Reactions, *J. Am. Chem. Soc.*, 2004, in press.
- [4] 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.

