# Classification and Analysis of KEGG/REACTION Database

**Masaaki Kotera** kot@kuicr.kyoto-u.ac.jp
**Yasushi Okuno** okuno@kuicr.kyoto-u.ac.jp
**Masahiro Hattori** hattori@kuicr.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 611-0011, Japan

**Keywords:** EC (Enzyme Commission), chemical compound, metabolic pathway

## 1 Introduction

For recent decades, the EC (Enzyme Commission) number has been used to research various biological reactions and metabolic pathways. However, the EC classification is unsuitable for bioinformatics analyses because of the, at least, following two problems. One is inconsistent criteria for classification, utilizing reaction patterns, substrates, transfered groups, and acceptor groups in different ways in different hierarchies. Another is that when one EC number is given to an enzyme which catalyzes an overall reaction consisting of multiple elementary reactions steps, the same number is often assigned to all elementary reactions. For example, there are several reactions which oxidize substrate and then dissociate inorganic compounds (CO2, NH3, etc.), but once these reactions are classified into an overall oxidoreductase reaction (EC.1), it is not easy to reveal inherent property of dissociation only from the EC number.

In order to solve these problems, we are constructing a new classification scheme for enzymatic reactions as accumulated in the LIGAND/REACTION database [3]. Here, we report the decomposition of reactions into the combinations of reactant-product pairs and its application to the classification of reactions.

## 2 Method and Results

A reactant-product pair is a chemical compound pair consisting of a reactant and a product that share atoms or atom groups excluding hydrogen atoms. We suppose the features of each reaction can be represeted by the combination of those pairs. Thus, we analyzed reactions in LIGAND and constructed a database of reactant-product pairs. After compiling the database, we computed matched and other regions of pairs using the graph theoretical method [1]. Finally, the usefulness of those regions to classify reactions is evaluated.

### 2.1 Collection of Reactant-Product Pairs

We used 5,237 reactions in LIGAND/REACTION database [2, 3]. From those reactions, we picked up 7,990 reactant-product pairs without considering whether each compound of a pair is organic or inorganic. Each pair was manually examined from the viewpoint of biochemical knowledge. Of those data, the number of pairs consisting of only organic compounds is 3,955, which is used for the following procedures.

### 2.2 Post-Processing of the Database of Pairs

For the sake of reaction classification, we have computed three types of characteristic regions of each pair: matched region, difference region, and reactive region, by the chemical structure comparison method [1]. (1) Matched region: the region that does not change through the reaction and is expected to be involved in substrate specificity of the enzymes; atoms and atomic bonds are aligned in this region in the chemical structure comparison. (2) Difference region: the region that changes through the reaction; atoms and atomic bonds are not aligned in the chemical structure comparison. (3) Reactive region: the reaction center responsible for the reaction mechanismsof an enzyme; atoms and atomic bonds are part of the difference region and adjacent to the matched region.

According to the identity of these three regions, reactant-product pairs are classified into groups, and the grouping information is stored in the database.

### 2.3 Evaluation of Regions of Pairs

To evaluate the usefulness of our database, we examined the correspondence between our grouping and the EC numbers. Consequently, many groups in our database can be assigned to only one class of the first level of the EC nomenclature, and some groups possess two or more classes. However, there are also many cases where a single group is assigned to multiple numbers, especially those assigned to both EC.2 (Transferase) and EC.3 (Hydrolase), and additionally to EC.6 (Ligase). We explain this observation as follows: a hydrolysis reaction can be described as "transferring reaction using a water molecule", and a ligase reaction can be described as a combination of hydrolysis and its counterreaction. Therefore, the hydrolysis and the ligase reaction partially have the same aspects of transferases.

## 3 Discussion

These results indicate that enzymatic reactions can be more consistently classified by using the three characteristic regions of reactant-product pairs, and provide our basis for constructing a new classification method of enzymatic reactions. Our classification is expected to be more effective and precise than the traditional EC nomenclature, because it is based on elementary reaction steps of every enzymatic reaction.

## Acknowledgments

We thank Rumiko Yamamoto, Tomoko Komeno and Koichiro Tonomura for improvement and maintenance of the LIGAND database. This work was supported by grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan and from the Japan Society for the Promotion of Science.

## References

[1] Okuno, Y., Hattori, M., Kotera, M., Igarashi, Y., Goto, S., and Kanehisa, M., Vzyme: A template-based method to predict reactions between two chemical compounds, *Genome Informatics*, 13:355–356, 2002.

[2] Yamamoto, R., Goto, S., Okuno. Y., Hattori, M., Nishioka, T., and Kanehisa, M., Improvement of the LIGAND chemical database, *Genome Informatics*, 13:492–493, 2002.

[3] LIGAND database, `http://www.genome.ad.jp/ligand/`