Automatic Assignment of Full EC Numbers Based on Structural Changes of Chemical Compounds

Nobuya Tanaka1
tanaka@kuicr.kyoto-u.ac.jp
Tadashi Kadowaki1
kadowaki@kuicr.kyoto-u.ac.jp
Susumu Goto1
goto@kuicr.kyoto-u.ac.jp
Minoru Kanehisa1
kanehisa@kuicr.kyoto-u.ac.jp

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

Keywords: enzyme, reaction, EC number

1 Introduction

We have developed an algorithm for automatically assigning full EC numbers given chemical structures of substrates and products. The EC system is a hierarchical classification of enzymatic reactions divided into four levels, each represented by a unique number. The first three levels define the reaction type, and the fourth, assigned serially, defines the substrate specificity. Although EC numbers have been widely used as standard nomenclature, the recent explosion of putative enzymes have become problematic in the current procedures used to assign EC numbers, which require full experimental characterization and publication of enzymes. Therefore, for a genomic scale analysis of enzymes and metabolic pathways, it becomes necessary to develop a new classification scheme of enzymatic reactions using systematic techniques.

2 e-zyme system

The e-zyme that predicts the first three levels of EC numbers has been previously published[3, 4]. The e-zyme system encodes reaction patterns by computationally detecting changes of the compound structures around the reaction center before and after the reaction[1]. This coding scheme mechanically organizes reaction types, which are compatible up to the third level of EC numbers. However, it has not been able to predict the fourth level.

3 Algorithm

The fourth level corresponds to the substrate specificity of enzymes. We assume that the substrate specificity depends on the substructures around the reaction center. Based on this assumption, we have developed an algorithm for assigning full (including the fourth level) EC numbers by extending the e-zyme system. In this algorithm, the variations of substrate substructures are enumerated incrementally extending from the reaction center. By recursively joining the same substructures, we obtain a classification tree. We then labeled the nodes of this tree with known EC numbers, so that it could be used to classify the substrate specificity of new reactions.
4 Results

Figure 1 is the classification tree for EC 1.1.1. The compounds corresponding to EC 1.1.1.1 were scattered into four clusters, while those corresponding to EC 1.1.1.21 mapped to a single cluster. This illustrates the range of substrate specificity of different EC numbers. We are currently developing a prediction method based on this tree, which will be incorporated into the e-enzyme system.

5 Acknowledgements

This work was supported by grants from the Ministry of Education, Sports, Culture, Science and Technology Agency and the Japan Science and Technology Agency. The computational resources were provided by the Supercomputer Laboratory, Institute for Chemical Research, Kyoto University.

References


