

Chemical Genomic Study of Endocrine Disrupting Chemicals in Metabolic Pathways

Tadashi Kadowaki¹
kadowaki@kuicr.kyoto-u.ac.jp

Tetsuya Adachi²
adachi@tom.life.h.kyoto-u.ac.jp

Shinobu Okamoto¹
okamoto@kuicr.kyoto-u.ac.jp

Nobuya Tanaka¹
tanaka@kuicr.kyoto-u.ac.jp

Koichiro Tonomura¹
tonomura@scl.kyoto-u.ac.jp

Gozoh Tsujimoto²
gtsuji@pharm.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

² Department of Genomic Drug Discovery Science, Graduate School of Pharmaceutical Sciences, Kyoto University, Yoshida-Shimoadachi-cho, Sakyo-ku, Kyoto 606-8501, Japan

Keywords: endocrine disrupting chemical, metabolic pathway, structural learning, chemical genomics

1 Introduction

Endocrine disrupting activities in industrial chemicals have been a serious problem, because they are distributed broadly, they are active at very low concentrations, and some of them are difficult to detoxify. Moreover, only a small number of active chemicals are known, while most chemicals have not been examined yet. To reduce the number of examinations required for unknown chemicals, some computational prediction methods have been proposed.

In this study, we propose a chemical genomic approach to understand the relationship between activities and chemical structures. Our approach revealed the existence of synthetic and degradation chains in metabolic pathway maps. At the end of the degradation chains, we found certain oxidization enzymes that appeared frequently. We also confirmed our prediction by performing a bioassay experiment.

The results would help us to find potential detoxifying microorganisms. In addition, our method can be applied to other types of toxic chemicals, such as cancer-causing agents.

2 Method and Results

We used two data sets of chemicals. One was a training data set, which had chemical structures known to have endocrine disrupting activity. The other was a test data set of chemical structures without activity information. The training data set was obtained from the Endocrine Disruptor Knowledge Base (EDKB) [1], while the test data set of chemicals was provided by the KEGG COMPOUND database [2].

A structural learning method based on the Adaboost [3], which consists of weighted weak classifiers, was applied. As chemical structures are expressed as graphs, subgraph-based decision stumps can define the weak classifiers [4]. The decision stump has two parameters, a graph and a class label. It returns either the label or its opposite sign depending on whether it contains the subgraph or not.

The performance of the learning method was validated by 10-fold cross validation. The results showed that the best performance was obtained when the number of weak classifiers was around 30. Therefore, we used 30 weak classifiers for our analysis.

The prediction model constructed by this method does not have enough accuracy for the actual test data set, KEGG COMPOUND; the model predicted about one third of the input chemicals as active. This was not acceptable, because the previous large-scale analysis of known industrial chemicals showed that only less than 5% were active chemicals. In general, the bias of the training data set caused this bias in prediction. To avoid this bias, we randomly sampled 60 chemicals from KEGG COMPOUND to add inactive chemicals to the training data set. The final classification was decided by a major vote of 100 individual strong classifiers,

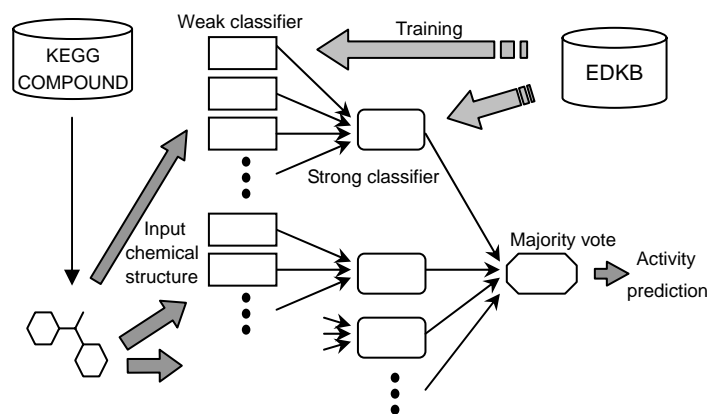


Figure 1. Overview of the method.

which were trained with distinct randomly-sampled chemicals. Figure 1 shows the overview of our method.

Our method predicted 614 active chemicals from the test data set. This is about 5% of the data set, is consistent with the previous study. These chemicals were mapped onto the KEGG PATHWAY and 21 metabolic pathways were found. The results showed that the chemicals with disrupting activity appear consecutively on metabolic reaction pathways. Some of these belong to synthetic pathways and others belong to degradation pathways. At the end of the degradation chains, enzymes with EC numbers 1.13.-.- and 1.14.-.- appeared most frequently.

Our prediction method was confirmed by a MCF-7 cell proliferation assay. In particular, the endocrine disrupting chemical bisphenol A (BPA) was active, but three degraded chemicals in the pathway, 4-HBAC, 4-HBAL and 4-HAP were inactive. This was consistent with our prediction.

3 Discussion

A chemical genomic approach for prediction of endocrine disrupting chemicals was performed. The relationship between chemicals and enzymes can be extracted from pathway maps. The key enzymes in detoxification were found in the degradation pathway maps. Some predicted active and inactive chemicals were confirmed by a bioassay.

The active and inactive chemicals and enzymes on the degradation pathways will help us to find microorganisms that have the potential to degrade environmental pollutants. The approach proposed in this study can also be applied to other types of toxicity, such as cancer-causing agents and, hepatotoxic and nephrotoxic chemicals.

References

- [1] <http://edkb.fda.gov/>
- [2] Kanehisa M, Goto S, Kawashima S, Okuno Y, and Hattori M., The KEGG resource for deciphering the genome, *Nucleic Acids Res.* 32:D277-280, 2004.
- [3] Freund T. and Schapire R. E., Experiments with a new boosting algorithm, *Proc. 13th International Conference on Machine Learning*, 148-146, 1996.
- [4] Kudo, T., Maeda, E., and Matsumoto, Y., An Application of Boosting to Graph Classification, *Proc. NIPS 2004*, 729-736, 2004.