The University of Minnesota Biocatalysis/Biodegradation Database: Developing a Formal Representation for the Prediction of Metabolic Pathways

S. Kramer[1], L. Ellis[2], C. Helma[3], S. Parsons[4], E. Gottmann[3]
[1]Austrian Research Institute for Artificial Intelligence, Vienna, Austria
stefan@ai.univie.ac.at
[2]Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, USA
[3]Institute for Tumor Biology–Cancer Research, University of Vienna, Vienna, Austria
[4]Intelligent Systems Group, Department of Electronic Engineering, Queen Mary & Westfield College, University of London, UK

The University of Minnesota Biocatalysis/Biodegradation Database (UM-BBD, http://www.labmed.umn.edu/umbbd/) is an unique source of information about experimentally determined microbial biocatalytic reactions and biodegradation pathways primarily for xenobiotic, chemical compounds (Ellis et al., 1999). Our aim is to apply various Artificial Intelligence methods to extract knowledge about biodegradation mechanisms from this data.

We are therefore developing a new, formal representation of the UM-BBD that lends itself to processing by Artificial Intelligence methods. This representation will have a deductive database format which is suitable for reasoning about pathways in the so-called logic of argumentation and for the induction of degradation pathway models in a logical framework.

The first step will be to use the knowledge in the UM-BBD as part of a Knowledge-Based System (KBS) that reasons about pathways for unknown compounds. The argumentation will be similar to Knowledge-Based Systems predicting toxicological properties (e.g., StAR (Langowski et al., 1997)) – constructing arguments for or against the possibility of certain pathways based on different grounds, such as structural similarity, chemical properties, and presence or absence of certain substructures.

The prediction of degradation pathways and products from examples and background knowledge will also be treated as a learning problem concerned with the prediction of complex transformations of relational structures. One goal of this work is to find a suitable formulation of this learning problem in Inductive Logic Programming (ILP) terms and to find a preliminary solution to this problem. Due to the complexity of the task and the limited amount of data, we aim at a tight integration of Knowledge-Based and Machine Learning techniques.

Ellis L.B.M., Hershberger C.D. and Wackett L.P.: The University of Minnesota Biocatalysis/Biodegradation Database: Specialized Metabolism for Functional Genomics, *Nucleic Acid Research*, 27:373-376, 1999.

Langowski J.J., Judson P.N., Patel M. and Tonnelier C.A.G.: StAR: A Knowledge-Based Computer System for Carcinogenic Risk Assessment, *Animal Alternatives, Welfare and Ethics*, 747-752, 1997.