The University of Minnesota Biocatalysis/Biodegradation Database: a pathway/genome dataset and system

Lynda B.M. Ellis[1], I. Saira Mian[2], Lawrence P. Wackett[3]

[1]Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN 55455 USA, lynda@tc.umn.edu, [2]Department of Radiation Biology and DNA Repair, Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720 USA, SMian@lbl.gov, [3]Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, St Paul, MN 55108 USA, wackett@biosci.cbs.umn.edu

The used and useful University of Minnesota Biocatalysis/Biodegradation Database (UM-BBD, Ellis *et al.*, 1999) (http://www.labmed.umn.edu/umbbd/index.html) is a curated collection of microbial metabolic pathways. In contrast to other pathway databases such as KEGG and WIT, the primary focus of the UM-BDD is bacterial and archaeal specialized pathways for the biodegradation of xenobiotic chemical compounds. As of May 1999, the UM-BBD contained 90 carefully vetted pathways, 625 reactions, 567 compounds, 410 enzymes and 232 microorganism entries. Our long-term goal is prediction of metabolic pathways and development of an ontology of biological function that can address complex queries spanning pathways and genomes. To meet this goal, a pathway/genome dataset and system for large-scale, high throughput, semi-automated navigation and analysis is required.

Here, we describe the start of such a dataset and system. The protein building blocks of UM-BDD pathways are being more completely characterized through annotation using a library of generative (hidden Markov model) and discriminative (support vector machine) models of known structural and other domains (Mian and Dubchak, 1999; Chow *et al.*, 1999). Phylogenetic analysis and, ultimately, atomic level models of each domain provide insights into areas such as substrate specificity, intra- and inter-domain relationships, strategies for engineering new functionality, and evolution of specialized microbial (and possibly eucaryotic) metabolism. This research provides the framework for integrating expression and other biological data into the UM-BDD. The system being developed will allow exploration of the interplay between local molecular form and larger integrated biological processes.

_____

Ellis LBM, Hershberger CD and Wackett LP (1999) "The University of Minnesota Biocatalysis/Biodegradation Database: Specialized Metabolism for Functional Genomics" Nucleic Acid Research 27: 373-376.

Mian IS and Dubchak I (1999) "Representing and reasoning about protein families: combining generative (hidden Markov model) and discriminative (neural network) based-methods derived from different projections of a family." In preparation.

Chow ML, Austin AS and Mian IS (1999). "A hidden Markov model/Support Vector Machine Protein Family Library". In preparation.