

## A Workshop for Predicting Chemical Fate in the Environment

Final Report to the  
Consortium on Law and Values in Health, Environment & the Life Sciences

Submitted by Lynda Ellis, PhD

### Introduction

The University of Minnesota Biocatalysis/Biodegradation Database (UM-BBD, <http://umbbd.msi.umn.edu/>, Ellis et al., 2006), freely available on the web for over a decade, contains information on the bacterial metabolism of chemicals, knowledge that is used to both degrade pollutants and biosynthesize industrial chemicals in non-polluting processes. Eight years ago, a project was begun to use the knowledge in the UM-BBD to predict the fate of chemicals in the environment, which has led to development of the University of Minnesota Pathway Prediction System (PPS, <http://umbbd.msi.umn.edu/predict/>, Ellis et al., 2008). As part of this project, we have held three PredictBT workshops convening biodegradation and computer experts to help guide PPS development. Consortium support was added to existing Graduate School support for the 4<sup>th</sup> PredictBT workshop, focused on integrating genomic and thermodynamic information into the PPS. This required expertise different than what we have obtained previously; thus, a new workshop with different expert scientists was essential. The workshop was held at the University of Minnesota, May 4-5, 2007. Its final report is attached; its website is: <http://umbbd.msi.umn.edu/predictbt/workshop4/>.

### Background

The University of Minnesota has hosted research programs on microbial biocatalysis and biodegradation (collectively, biotransformations) for over 40 years. Currently, biocatalysis and biodegradation are priority areas under the President's University-wide initiatives. The Institute for Renewable Energy and the Environment, the Center for Microbial and Plant Genomics, and the BioTechnology Institute are part of the constellation of resources working under this initiative.

Another of the University's jewels in this research area is the University of Minnesota Biocatalysis/Biodegradation Database (UM-BBD, <http://umbbd.msi.umn.edu/>, Ellis et al., 2006). The UM-BBD, freely available on the web, has been used worldwide for over 10 years. The UM-BBD contains information on the bacterial metabolism of chemicals, knowledge that is used to both degrade pollutants and biosynthesize industrial chemicals in non-polluting processes.

### Workshop 4

Eight years ago, a project was begun to use the knowledge in the UM-BBD to predict the fate of chemicals in the environment. As part of this project, we have held four PredictBT workshops convening biodegradation and computer experts (<http://umbbd.msi.umn.edu/predictbt/>). The first, held at the University of Minnesota in 1997 (Wackett et al., 1999), led to several publications (Wackett & Ellis, 1999; Hou et al., 2003; Hou et al., 2004) and the implementation of the Pathway Prediction System (PPS, <http://umbbd.msi.umn.edu/predict/>), a software tool for biodegradation prediction. In 2005, two additional workshops were held, funded by the European Union (EU); the system is heavily used by industry and environmental regulatory agencies in Europe. The first EU-funded workshop was held at the University of Minnesota in May 2005. The second was held in Brussels, the EU capital, in December 2005. These workshops improved the PPS, brought greater recognition to the University and the system, and have contributed to attracting research funding from Lhasa Limited, a company in the United Kingdom. Following the third workshop, one participant, Assistant Professor Kathrin Fenner from the

Swiss Federal Institute of Technology, received a grant to the Swiss National Science Foundation to come to Minnesota for one year to work on the PPS; she started in November 2006.

We are continuing to develop the PPS with funding from NSF, and believe it is essential to hold periodic workshops to guide and ground development of the system. The knowledge required to predict environmental fate of chemicals is more extensive than that possessed by any single person, requiring expertise in physical chemistry, organic chemistry, biochemistry, microbiology, molecular biology, and computer science. Our past workshops have brought this expertise together and further teach us what users want in such a system. The fourth workshop was focused on integrating genomic and thermodynamic knowledge into chemical fate prediction.

The workshop was held at the University of Minnesota, on May 4-5, 2007. There were 14 participants, 7 external to UM: John Bumpus, Department of Chemistry and Biochemistry, University of Northern Iowa; Kathrin Fenner, Environmental Chemistry Group, ETH, Switzerland; Susumu Goto, Bioinformatics Center, Kyoto University, Japan; Stefan Kramer, Faculty for Informatics, Technical University Munich, Germany; Fangping Mu, Theoretical Biology and Biophysics, Los Alamos National Laboratory; Mukesh Patel, Lhasa Limited, Leeds, UK; Jeanne vanBriesen, Department of Civil and Environmental Engineering, Carnegie-Mellon University; and 7 UM faculty, postdocs, and students: Lynda Ellis, Laboratory Medicine and Pathology; Larry Wackett, Biochemistry, Molecular Biology & Biophysics; Yevgeniy (Eugene) Podolyan, Computer Science and Engineering; Chunhui Li, Biochemistry, Molecular Biology & Biophysics; Junfeng (Jeff) Gao, Health Informatics; Rachael Long, BioTechnology Institute; Michael Turnbull, Laboratory Medicine and Pathology.

**References** (\* = contains fourth workshop content or ideas generated by collaborations formed there)

Ellis, L.B.M., Roe, D., Wackett, L.P. (2006) The University of Minnesota Biocatalysis/ Biodegradation Database: the first decade. *Nucleic Acids Research*, **34**: D517 – D521.

\*Ellis, L.B.M., Gao, J., Fenner, K., Wackett, L.P. (2008) The University of Minnesota Pathway Prediction System: predicting metabolic logic. *Nucleic Acids Research*, **36**: W427 – W432.

\*Fenner, K., Gao, J., Kramer, S., Ellis L.B.M., Wackett, L.P. (2008) Data Mining to Limit Combinatorial Explosion in Biodegradation Pathway Prediction. *Bioinformatics*, in press.

Hou, B.K., L.P. Wackett, and L.B.M. Ellis (2003) Microbial pathway prediction: A functional group approach. *J. Chem. Inf. Comp. Sci.* **43**:1051-1057.

Hou, B.K., L.B.M. Ellis, and L.P. Wackett (2004) Encoding microbial metabolic logic: Predicting biodegradation. *J. Ind. Microbiol. Biotechnol.* **31**:261-272.

Wackett, L.P., L.B.M. Ellis, S.M. Speedie, C.D. Hershberger, H-J. Knackmuss, A.M. Spormann, C.T. Walsh, L.J. Forney, W.F. Punch, T. Kazic, M. Kanehisa, and D.J. Berndt (1999) Predicting microbial biodegradation pathways. *ASM News* **65**:87-93.

Wackett, L.P. and L.B.M. Ellis (1999) Predicting biodegradation. *Environ. Microbiol.* **1**: 119-124.

**Summary, PredictBT Workshop, May 4-5, 2007**  
**<http://umbbd.msi.umn.edu/predictbt/>**

### **Introduction/Background**

The Fourth PredictBT workshop was held May 4-5, 2007 at the University of Minnesota, St Paul campus, in the Cargill Building for Microbial and Plant Genomics. External participants were: John Bumpus, University of Northern Iowa; Kathrin Fenner, EAWAG/ETH, Switzerland; Susumu Goto, Kyoto University, Japan; Stefan Kramer, Technical University Munich; Fangping Mu, Los Alamos National Laboratory; Mukesh Patel, Lhasa Limited, UK; Jeanne VanBriesen, Carnegie-Mellon University.

Participants at the past two PredictBT workshops, held in May 2005 at the University of Minnesota and in December 2005 in Brussels, Belgium, made many suggestions to improve the University of Minnesota Biocatalysis/Biodegradation Database Pathway Prediction System (PPS, <http://umbbd.msi.umn.edu/predict/>). Most of these have been implemented: Results are now sorted by aerobic likelihood from green to yellow to red. A Cpd button now appears when a result compound is in the BBD. CoA is now compressed in rule graphics. It remains expanded in results, but one click in a result compound image will open a separate, rescalable compound window. Loops and futile cycles have been eliminated. All atoms in result compounds are black. CoA thioester formation was divided into aromatic (unlikely) and aliphatic (likely). Fatty acids were excluded from decarboxylation (bt0051). Oxidative cleavage of amines (bt0063) is now likely. Dephosphorylation of aryl phosphates (bt0104) was generalized to handle allyl phosphates.

Other improvements, in addition to those suggested at PredictBT workshops, were made: We have moved to a new server, with more room for growth, faster speed, and better system support. All web pages are now served from the standard port 80. All PPS rules have been moved to the ChemAxon Reactor software, which provides a GUI interface for rule creation and update. The results page now includes a navigation bar, for easy navigation among previous prediction steps, the PPS Home, or the UM-BBD Home. Results no longer contain acetate, CO<sub>2</sub> and similar small molecules, whose degradation should not be predicted. The results page now uses .png format graphics, rather than Java Applets, to show results. Thus, PPS results appear more rapidly, and their display is more stable. Compounds in predicted pathways now contain Rule and, where possible, Cpd buttons, and are now separated by arrows colored with the aerobic likelihood of the relevant btrule. A web page on "Compounds whose biodegradation should not be predicted" has been developed, and the PPS tutorial has been improved.

Some workshop recommendations are still in progress: A draft statement of PPS purpose and scope (see below) has been developed. Simplification of facile transformation sequences: we now can skip single "immediate" reaction steps; work on simplifying longer cascades, such as beta-oxidation, is in progress. The number of rules for reactions of sulfur-containing compounds has been increased; this is continuing. Some rules should handle substrate rings of varying sizes; this is under study [slides Update050407.pdf].

### Challenges to Expert Rule Ranking

Lynda Ellis discussed the fact that only 21/230 rules are based on 10+ reactions, and it is harder to rank rules when there are fewer examples. There have been errors in ranking. Over half the rules are ranked as Neutral. The major problem: combinatorial explosion. [slides Predict050407.pdf]

### Suggested Ways to Limit Combinatorial Explosion

- Ask experts to rerank neutral rules. Would have to rerank all rules, no funding for this.
- Use relative reasoning – **Mukesh Patel** - Lhasa Limited uses this in METEOR to further subdivide the rules. For example, methyl groups are oxidized to primary alcohols (bt0036) frequently when the PPS predicts pathways for atrazine. Oxidative cleavage of amines (bt0063) should be preferred. Add relative reasoning that bt0063 > bt0036. Then, when bt0063 is triggered, biotransformations from bt0036 will not be shown. A simple version of relative reasoning (not as complex as used in METEOR) could be implemented in PPS.
- More rationally restrict rules using biochemical and chemical knowledge.
- Create a list of Readily Degradable Intermediates (RDIs) as goals for degradation.
- Create a separate, batch system that would generate the entire tree, then prune it.
- Prune by thermodynamics – **Kathrin Fenner** – Using ATP counting, known pathways are in the upper 10<sup>th</sup> percentile of possible PPS pathways (abstract available). [slides FennerPPS-Workshop-050507.pdf ]
- Prune by shortest path to RDI. Even seeing two levels of the tree would help, because there will be more chance to find RDIs.
- Use similarity. **Susumu Goto** – Kyoto group developed a biodegradation prediction system that uses chemical similarity to avoid combinatorial explosion. **Larry Wackett** and **Jeanne VanBriesen** - Use biological, not chemical, similarity. 2,4-dimethylbenzamine is >0.9 similar (Tanimoto) to m-Xylene, but they have different RDIs. 2,4-dimethylbenzamine is biodegraded to 4-methylcatechol, m-Xylene to 3-methylcatechol. **Stefan Kramer** - chemical similarity could be used to prescreen queries. If a query is not chemically similar to UM-BBD compounds, PPS predictions are less accurate.
- Use machine learning. **Stefan Kramer** - developed software for determining substructures that may determine whether a transformation occurs. He needs positive and negative examples. 8/21 rules with 10+ reactions are problematic. These will be studied first.
- Use molecular properties. **Fangping Mu** – Los Alamos group used molecular properties to predict oxidoreductase enzymes using KEGG data (paper available). However there are still problems predicting  $-CH \rightarrow -COH$  and problems with the negative examples for  $-CHO \rightarrow -COOH$  [slides MU-UMN.pdf]

### Other comments

Ordering results by aerobic likelihood ranking (more to less likely), then by rule number, may imply that smaller numbered neutral rules are more likely than larger numbered ones. This is not necessarily incorrect, since more common rules were created first (have lower numbers).

**Draft Purpose and Scope**

The PPS predicts plausible pathways for the biodegradation of chemical compounds. Its predictions are most accurate for compounds similar to compounds in the UM-BBD. Its predictions are most accurate for normal environments: exposed to air, in moist soil or water, at moderate temperatures and pH, with no competing chemicals or toxins. Its predictions are most accurate for compounds that are the sole source of energy, carbon, nitrogen or other essential elements for the microbes in these environments, rather than cometabolized, or present in trace amounts.

**Sponsors**

The Graduate School

**Contact Us**

We welcome comments on the Draft Purpose and Scope and the rest of this summary, and in general on the PPS and the UM-BBD on which it is based.

Send your comments to Larry Wackett <[wacke003@umn.edu](mailto:wacke003@umn.edu)> and Lynda Ellis <[lynda@umn.edu](mailto:lynda@umn.edu)>.