

The University of Minnesota Biocatalysis/ Biodegradation Database: the first decade

Lynda B. M. Ellis*, Dave Roe and Lawrence P. Wackett¹

Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, Mayo Mail Code 609, 420 SE Delaware Street, MN 55455, USA and ¹BioTechnology Institute, University of Minnesota, St Paul, MN 55108, USA

Received September 12, 2005; Revised and Accepted October 8, 2005

ABSTRACT

As the University of Minnesota Biocatalysis/Biodegradation Database (UM-BBD, <http://umbbd.ahc.umn.edu/>) starts its second decade, it includes information on over 900 compounds, over 600 enzymes, nearly 1000 reactions and about 350 microorganism entries. Its Biochemical Periodic Tables have grown to include biological information for almost all stable, non-noble-gas elements (<http://umbbd.ahc.umn.edu/periodic/>). Its Pathway Prediction System (PPS) (<http://umbbd.ahc.umn.edu/predict/>) is now an internationally recognized, open system for predicting microbial catabolism of organic compounds. Graphical display of PPS rules, a stand-alone version of the PPS and guidance for PPS users are being developed. The next decade should see the PPS, and the UM-BBD on which it is based, find increasing use by national and international government agencies, commercial organizations and educational institutions.

INTRODUCTION

The University of Minnesota Biocatalysis/Biodegradation Database (UM-BBD, <http://umbbd.ahc.umn.edu/>) contains compound, enzyme, reaction and pathway information for microbial catabolism of primarily anthropogenic materials. It has been available on the web for over 10 years, and has grown from 4 to almost 150 pathways. As it starts its second decade, the UM-BBD contains information on over 900 compounds, over 600 enzymes, nearly 1000 reactions and about 350 microorganism entries. Its data content and methods, including data format, update and access, have been reported previously (1–4).

Along with pathway data, the UM-BBD now includes Biochemical Periodic Tables and a Biodegradation Pathway Prediction System (PPS). Its Biochemical Periodic Tables

have grown to include biological information for almost all stable, non-noble-gas elements, and an 'About the Biochemical Periodic Tables' web page has been created (<http://umbbd.ahc.umn.edu/periodic/>). The PPS (<http://umbbd.ahc.umn.edu/predict/>) is now an internationally recognized, open system for predicting microbial catabolism of organic compounds, described in more detail below.

DATABASE GROWTH AND UPDATES

Since its last report (4), the UM-BBD has grown overall ~20%. It has added about 200 reactions and compounds and 150 enzymes. Enzyme entries grow more slowly than reactions, not only because some new reactions are catalyzed by existing enzymes, but also because some are not well-characterized enough for their enzymes to be entered. This information was added to 17 new pathways and 34 existing pathways. Now that our pathways span a wide representation of microbial catabolism, we primarily limit additions to pathways that represent new carbon skeletons or novel metabolism, and spend more time and resources keeping existing information up-to-date. As further examples of the last, links on compound pages to the National Toxicology Program (<http://ntp.niehs.nih.gov/>), and on pathway pages and microorganism entries to the American Type Culture Collection (<http://www.atcc.org/>), were updated during this period.

PPS

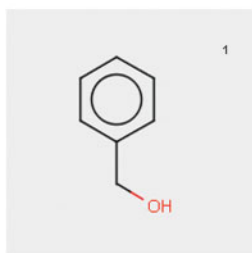
Earlier versions of the PPS have been described (4–6). Rules for microbial biotransformation of organic functional groups, derived from UM-BBD data, are developed. Users input a chemical structure, the PPS determines the functional groups it contains, and the compound is metabolically transformed *in silico* using PPS rules. The PPS displays the resulting compounds, the user selects one, the cycle repeats and a predicted biodegradation pathway grows (5,6). An 'About the PPS' web page has been developed (<http://umbbd.ahc.umn.edu/predict/aboutPPS.html>).

*To whom correspondence should be addressed. Tel: +1 612 625 9122; Fax: +1 612 624 6404; Email: lynda@tc.umn.edu

Over the past 3 years, the PPS has grown to include about 250 rules. Rules now can have multiple descriptions. For example, rule bt0024, Ester \rightarrow Alcohol + Carboxylate, is also known as Lactone \rightarrow Hydroxyacid. Rules can be now be searched on by name of substrate or product, or a complete list of all rules can be browsed. Rule web pages list all UM-BBD reactions that exemplify each rule; and UM-BBD reaction pages list the rules they exemplify. A UM-BBD reaction may exemplify up to three rules, in series.

As the number of rules increases and each rule increases in complexity, the average number of biotransformations shown to the user at each prediction cycle also increases. For example, the PPS initially predicted six biotransformations for benzyl alcohol [Figure 8 in Ref. (5)]. The present system predicts nine (Figure 1), a 50% increase. Users increasingly complain of 'too many choices'. To address this complaint, with support from the 6th EU Framework Assessing Large-scale environmental Risks with tested Methods (ALARM)

The predicted pathway:



Aerobic Likelihood:

■ Very likely
 ■ Likely
 ■ Neutral
 ■ Unlikely
 ■ Very unlikely
 ■ Unknown

Show BioTransformations: All Aerobic

Choose the next reaction step:

<p>b10012 Next Rule Cpd</p>	<p>b10011 Next Rule</p>	<p>b10013 Next Rule</p>
<p>b10001 Next Rule Cpd</p>	<p>b10005 Next Rule</p>	<p>b10005 Next Rule</p>
<p>b10005 Next Rule</p>	<p>b10005 Next Rule</p>	<p>b10270 Next Rule</p>

Figure 1. Excerpt from a PPS pathway prediction web page. The complete page appears if 'Demo' is chosen from the initial PPS page (<http://umbbd.ahc.umn.edu/predict/>). Predicted compounds 1 and 4 are in the UM-BBD; these are the only displays that include a Cpd button. That button links to their UM-BBD compound pages, and, from there, to reaction and pathway pages.

project (<http://www.alarm-project.ufz.de/>), on May 6–7, 2005, a PredictBT workshop was held to prioritize UM-BBD rules (<http://umbbd.ahc.umn.edu/predictbt/>).

The rules available in April 2005 were given to biodegradation content experts—who each scored subsets of them as follows: 1, Very Likely; 2, Likely; 3, Neutral; 4, Unlikely; 5, Very Unlikely—to occur under standard aerobic conditions [25°C, soil (moderate moisture) or water, neutral pH, no toxic or competing chemicals]. Two or more experts scored each rule. Rules with a wider range of scores were discussed at the workshop; participants were able to reach consensus on all of them. Arrangements have been made with selected workshop participants to similarly prioritize future rules.

Users may now see the color-coded score for each predicted biotransformation and optionally chose to see only the first three of the five groups, the ones most likely to occur under standard aerobic conditions (Figure 1). If a user starts with benzene and chooses the most probable aerobic reaction at each step, ring opening is predicted on the third step, in a pathway generated as if a personal biodegradation consultant guided each choice.

Participants also recommended that displays of predicted compounds in the UM-BBD indicate those predicted compounds in the UM-BBD, since this may guide user choices. A new 'Cpd' button is that indication (shown in Figure 1 for the first and fourth predicted compounds). That button links to the UM-BBD web page for that compound, and, from that, to information on its known reactions and catabolic pathways.

The original system was created, in part, using the JChem software (now JChemBase) (7) from ChemAxon, Inc. One limitation of the system was that rules were implemented as Java function calls (5), not readily understood by content experts. ChemAxon later introduced Reactor software, which

provides a graphic user interface for rule creation. Rules can now be created by non-programmers, and readily reviewed by content experts and general users. All new rules and modifications of existing rules are carried out in Reactor; existing rules are being converted to Reactor as time permits. Rules that have been moved to Reactor now display graphics on the rule page (Figure 2).

This conversion also allows us to generalize some rules that previously were too specific. For example, we first encoded the formation of aromatic dihydrodiols in several specific rules (6). Now there is one more general rule for most of them (Figure 2), since, as we state in the public Comment to this rule, '... aromatic hydrocarbon dioxygenases produce an activated dioxygen species that is thought to be sufficiently reactive to potentially functionalize most, if not all, aromatic ring carbon atoms ...'.

The UM-BBD has always been freely available on the web and that will continue for the foreseeable future. However, some users have expressed interest in a stand-alone version that can be used behind a firewall or in other environments where the Internet is not accessible. In 1999, UM-BBD users voted 'UM-BBD on CDROM' as what their institutions would be most willing to pay to use (<http://umbbd.ahc.umn.edu/stats3/results4.html>). Such a stand-alone version, including the PPS (but not the Biochemical Periodic Tables) is being developed with support from Lhasa Limited, Leeds, UK.

Lhasa Limited developed Meteor (8), stand-alone Windows software for predicting mammalian detoxification metabolism, using a functional group approach similar to that used in the PPS. Lhasa Limited is supporting transfer of UM-BBD microbial biodegradation rules into the Meteor framework, to create Meteor PPS (MEPPS). A prototype of MEPPS is available (Figure 3).

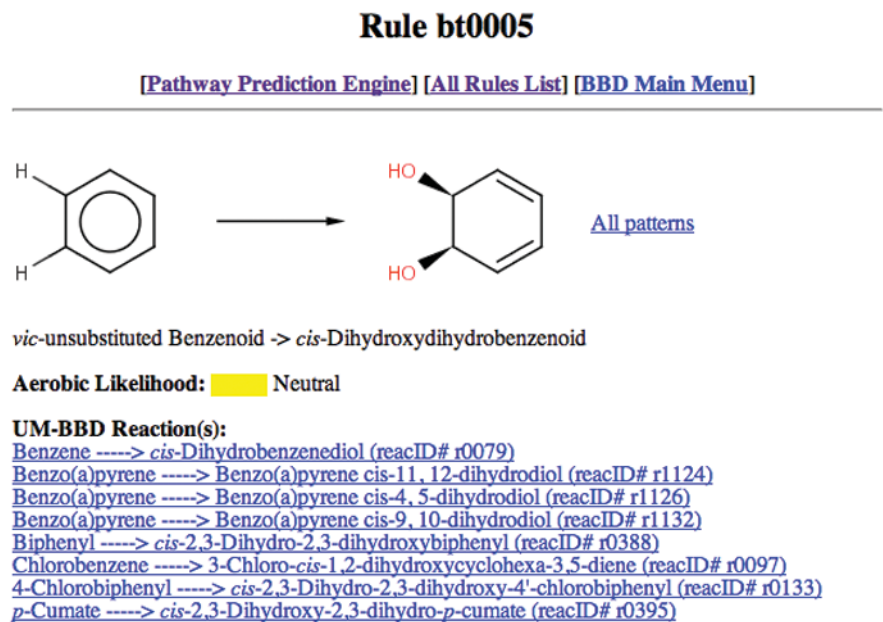


Figure 2. Excerpt from the web page for bt0005, a rule that has been moved to Reactor. One rule pattern graphic is displayed. Since more than one pattern is used for this rule, a link to a graphic showing all of them is to its right. Benzyl alcohol triggers both patterns for this rule at two different locations, producing predicted compounds 5 through 8 in Figure 1. A list of all rules is available at <http://umbbd.ahc.umn.edu/servlets/pageservlet?ptype=allrules>.

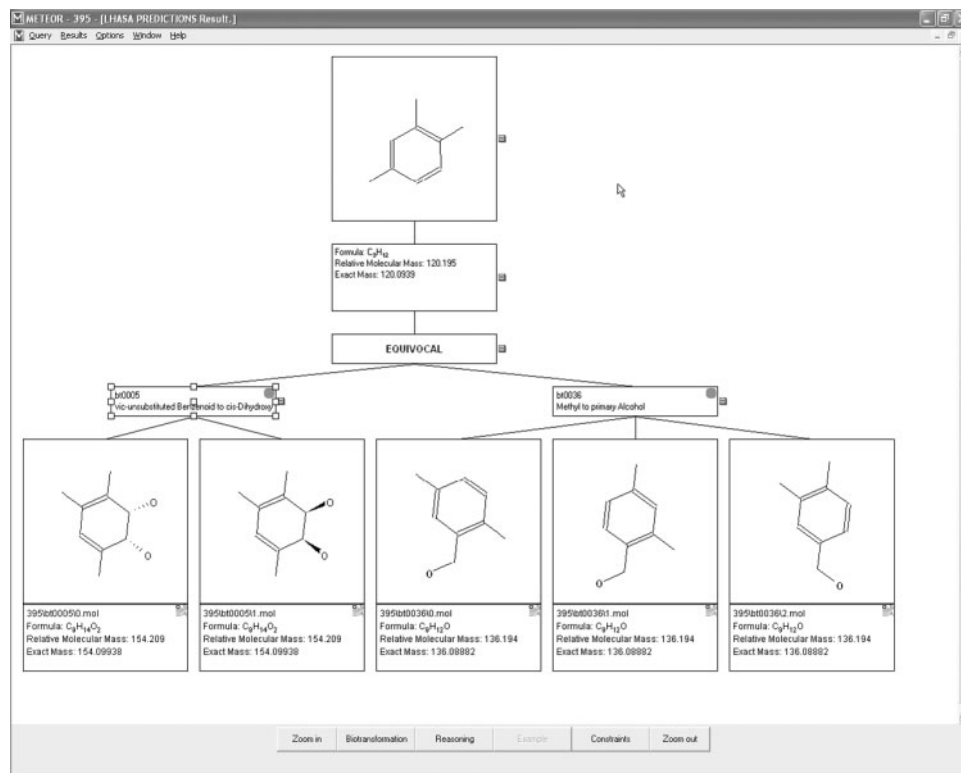


Figure 3. Excerpt from a prototype MEPPS page. The grayed out 'Examples' button at the bottom will link to the appropriate static UM-BBD rule web page (and from that to static reaction, enzyme, compound and pathway pages) once the prototype is complete.

MIRRORS AND DERIVATIVE WORKS

Since 2000, the European Bioinformatics Institute has mirrored the UM-BBD, as part of the EBI SRS server (9). SRS files are updated in synchrony with the UM-BBD; they omit the Biochemical Periodic Tables and the PPS.

The MEPPS system described above under Prediction, derived from the PPS, contains a static version of the UM-BBD. In principle, MEPPS UM-BBD files could be updated synchronously with the master UM-BBD. However, MEPPS is stand-alone software to be used at multiple sites, possibly behind a firewall. Timing of updates has not yet been determined, and may be less frequent.

In June 2000, Kyoto Encyclopedia of Genes and Genomes (KEGG) (10) began to add UM-BBD information to other metabolic pathways. This is a derivative work, not a mirror, since UM-BBD information is reformatted to KEGG standards, only a subset of UM-BBD information is used, and several UM-BBD pathways may be included in one KEGG graphic. Presently, about one-third (53/146) of UM-BBD pathways are included. KEGG documents and provides links to the UM-BBD pathways that are included in each KEGG graphic (see <http://www.genome.ad.jp/kegg/pathway.html>, section 1.11, Xenobiotics). KEGG does not update its UM-BBD-derived pathways in synchrony with the UM-BBD.

Other derivative works include Metarouter and CATABOL. Metarouter (11) is a static derivative, based on information taken from the UM-BBD in 2002. It is a proof-of-concept for alternative display and enhancement of UM-BBD information. CATABOL (12), a commercial biodegradability prediction

system, uses UM-BBD information for approximately half the pathways it contains (Ovanes Mekenyan, personal communication). It is unclear how often, or even whether, this information is updated.

CONCLUSIONS

With implementation of graphical display of PPS rules, development of a stand-alone version, and improvement guidance for PPS users, the next decade should see the PPS, and the UM-BBD on which it is based, find increasing use by national and international government agencies, commercial organizations and educational institutions. The UM-BBD is now approaching its teen years. As an adolescent, in the next stage of growth it will find new interests, explore new surroundings and develop new friends and colleagues. The next decade will be the deciding point: can it make the transition to maturity and independence?

ACKNOWLEDGEMENTS

We thank Sean Anderson, Jim Auer, John Carlis, Tony Dodge, Carla Essenberg, Mark Fischbach, Philip Judson, Yogesh Kale, Venkatesan Ramaswamy, Jack Richman, Ted Sands, John Schrom, Dan Smith and Jonathan Vessey for help in improving the UM-BBD in the past 3 years. We thank the May 2005 PredictBT workshop participants for their help in prioritizing UM-BBD rules. We thank the students in BioC/MicE 5309 at the University of Minnesota for creating many UM-BBD

pathways. This work was supported in part by DOE DE-FG02-01ER63268, Lhasa Limited, University of Minnesota 2005 Initiatives in Digital Technology and the 6th EU Framework ALARM project. Funding to pay the Open Access publication charges for this article was provided by Lhasa Limited.

Conflict of interest statement. None declared.

REFERENCES

1. Ellis,L.B., Hershberger,C.D. and Wackett,L.P. (1999) The University of Minnesota Biocatalysis/Biodegradation database: specialized metabolism for functional genomics. *Nucleic Acids Res.*, **27**, 373–376.
2. Ellis,L.B., Hershberger,C.D. and Wackett,L.P. (2000) The University of Minnesota Biocatalysis/Biodegradation database: microorganisms, genomics and prediction. *Nucleic Acids Res.*, **28**, 377–379.
3. Ellis,L.B., Hershberger,C.D., Bryan,E.M. and Wackett,L.P. (2001) The University of Minnesota Biocatalysis/Biodegradation Database: emphasizing enzymes. *Nucleic Acids Res.*, **29**, 340–343.
4. Ellis,L.B., Hou,B.K., Kang,W. and Wackett,L.P. (2003) The University of Minnesota Biocatalysis/Biodegradation Database: post-genomic data mining. *Nucleic Acids Res.*, **31**, 262–265.
5. Hou,B.K., Wackett,L.P. and Ellis,L.B. (2003) Microbial pathway predicting: a functional group approach. *J. Chem. Inf. Comput. Sci.*, **43**, 1051–1057.
6. Hou,B.K., Ellis,L.B. and Wackett,L.P. (2004) Encoding microbial metabolic logic: predicting biodegradation. *J. Ind. Microbiol. Biotechnol.*, **31**, 261–272.
7. Csizmadia,F. (2000) JChem: Java applets and modules supporting chemical database handling from web browsers. *J. Chem. Inf. Comput. Sci.*, **40**, 323–324.
8. Langowski,J.J. and Long,A. (2002) Computer systems for the prediction of xenobiotic metabolism. *Adv. Drug Deliv. Rev.*, **54**, 407–415.
9. Brooksbank,C., Cameron,G. and Thornton,J. (2005) The European Bioinformatics Institute's data resources: towards systems biology. *Nucleic Acids Res.*, **33**, D46–D53.
10. Kanehisa,M., Goto,S., Kawashima,S., Okuno,Y. and Hattori,M. (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res.*, **32**, 277–280.
11. Pazos,F., Guijas,D., Valencia,A. and De Lorenzo,V. (2005) MetaRouter: bioinformatics for bioremediation. *Nucleic Acids Res.*, **33**, D588–D592.
12. Dimitrov,S., Kamenska,V., Walker,J.D., Windle,W., Purdy,R., Lewis,M. and Mekenyan,O. (2004) Predicting the biodegradation products of perfluorinated chemicals using CATABOL. *SAR QSAR Environ. Res.*, **15**, 69–82.