 Nucleic Acids Research, 2000, Vol. 28, No. 1 377–379

The University of Minnesota Biocatalysis/Biodegradation Database: microorganisms, genomics and prediction

Lynda B. M. Ellis*, C. Douglas Hershberger and Lawrence P. Wackett[1]

Center for Biodegradation Research and Informatics, Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN 55455, USA and [1]Department of Biochemistry, University of Minnesota, St Paul, MN 55108, USA

Received September 13, 1999; Accepted September 15, 1999

ABSTRACT

The University of Minnesota Biocatalysis/Biodegradation Database (http://www.labmed.umn.edu/umbbd/) begins its fifth year having met its initial goals. It contains ~100 pathways for microbial catabolic metabolism of primarily xenobiotic organic compounds, including information on ~650 reactions, 600 compounds and 400 enzymes, and containing ~250 microorganism entries. It includes information on most known microbial catabolic reaction types and the organic functional groups they transform. Having reached its first goals, it is ready to move beyond them. It is poised to grow in many different ways, including mirror sites; fold prediction for its sequenced enzymes; closer ties to genome and microbial strain databases; and the prediction of biodegradation pathways for compounds it does not contain.

INTRODUCTION

On February 25, 2000, the University of Minnesota Biocatalysis/Biodegradation Database (UM-BBD, http://www.labmed.umn.edu/umbbd/index.html) will celebrate its fifth birthday. When it started on the web in early 1995, with only four metabolic pathways to its name, the goal was to represent the wide diversity of microbial catabolic reactions. Now, with ~100 pathways, 650 reactions, 600 compounds and 400 enzymes, it has reached this goal: it includes information on most known microbial catabolic reaction types and transformations of organic functional groups. Here we describe its present status and future directions.

PRESENT STATUS

Since the goal of the UM-BBD is to document the breadth of reaction types catalyzed by microbes, it has been necessary to define what constitutes 'difference' with respect to reactions. A 'functional group' is the smallest structural unit in an organic compound acted upon by a microbial enzyme. Examples are shown in Table 1. Each functional group can be transformed in one or more different enzymic reactions. For example, the nitrile functional group requires a specific enzyme, a nitrile hydratase or a nitrilase, to transform this more biologically rare functional group to an amide or carboxylic acid, respectively. The latter functional groups are more common and can be metabolized by many bacteria. The UM-BBD currently includes one or more compounds containing the 41 organic functional groups shown in Table 1; a list of the functional groups and representative compound(s) for each (http://www.labmed.umn.edu/umbbd/search/FuncGrps.html) and list of pathways organized by functional group (http://www.labmed.umn.edu/umbbd/search/sys_path.html) are available.

Table 1. UM-BBD functional groups[a]

Alkane, primary	Aldehyde
Alkane, secondary	Carboxylic acid
Alkane, tertiary	Carboxylic acid ester
Cycloaliphatic ring	Carboxylic thioester
Alkene	Amide
Alkyne	Nitrile
Monocyclic aromatic hydrocarbon	Thiocyanate
Polycyclic aromatic hydrocarbon	Nitro
Biphenyl-type benzenoid ring	Nitrate ester
Oxygen ether	Diazo
Thioether	Organohalide
S-heterocyclic ring	Organomercurial
N-heterocyclic ring	Organoarsenical
O-heterocyclic ring	Organosilicon
Ketone	Organotin
Thioketone	Organophosphate ester
Alcohol	Thiophosphate ester
Thiol	Phosphonic acid
Amine, primary	Sulfonic acid
Amine, secondary	Sulfate ester
Amine, tertiary	

[a]Taken from a list of UM-BBD functional groups and one or more UM-BBD compounds which contain them, available at URL: http://www.labmed.umn.edu/umbbd/search/FuncGrps.html

*To whom correspondence should be addressed at: Box 609 Mayo Building, 420 SE Delaware Street, Minneapolis, MN 55455, USA. Tel: +1 612 625 9122; Fax: +1 612 625 1121; Email: lynda@email.labmed.umn.edu

Table 2. Type (and number) of reactions catalyzed by naphthalene 1,2-dioxygenase, EC 1.14.12.12[a]

Dioxygenation reactions (28):	Acts on aromatic hydrocarbons (12)
	Acts on substituted aromatic compounds (11)
	Acts on heterocyclic aromatic compounds (5)
Monooxygenation reactions (24)	
Desaturation reactions (7)	
O- and N-Dealkylation reactions (5)	
Sulfoxidation reactions (9)	

[a]Taken from a complete list of all 73 reactions, available via URL: http://www.labmed.umn.edu/umbbd/naph/ndo.html

UM-BBD data content and methods, including data format, update and access, have been reported (1). With the bulk of data entry complete, a higher percentage of new pathways will be entered by distant learners enrolled in a Biocatalysis and Biodegradation course taught completely over the Internet (http://www.cee.umn.edu/biodeg/). This course is now offered once a year during a 15-week session each summer and can be taken for graduate and undergraduate credit.

Annotation of UM-BBD reaction pages has been carried out almost since the database began. The initial annotations were animations or pictures of the mechanisms for selected UM-BBD enzyme-catalyzed reactions. Twenty-four of these now exist and are listed in the UM-BBD Index to Graphics (http://www.labmed.umn.edu/umbbd/search/graph_BBD.html). This past year, we began to develop pages that document the biocatalytic versatility of these enzymes. For example, the UM-BBD contains seven reactions catalyzed by the enzyme naphthalene 1,2-dioxygenase (EC 1.14.12.12, http://www.labmed.umn.edu/servlets/pageservlet?ptype=e&enzymeID=e0002). But this enzyme is known to catalyze many more. We have documented 73 reactions (http://www.labmed.umn.edu/umbbd/naph/ndo.html); the major reactions groups and subgroups are shown in Table 2.

What does the future hold? Continued growth at a slower rate, of course, but as a representative, rather than all-inclusive, database, the UM-BBD will never include all information in the field. Also, as the database grows, an increasingly higher percentage of time is spent updating existing information. The past year has seen the UM-BBD reach a critical mass; it has reached its initial goals and it is now ripe for future development in several different, complementary, directions.

MIRROR SITES

As a database on the WWW, the UM-BBD attracts users from over 90 countries on six continents. However access may be more difficult the further a user is from the UM-BBD's midwestern US home. In 1999 UM-BBD adminstrators were approached by adminstrators of the Sequence Retrieval System (SRS) (2), maintained by the European Bioinformatics Institute (EBI), and agreed to permit EBI to mirror UM-BBD data. This mirror will offer most of the original, and some new, functionality. It is being tested now and will soon be available on the EBI's SRS server (http://srs.ebi.ac.uk/). This will provide UM-BBD users with several benefits: off-site data backup; improved access for European users; and reduced load on the US server, which should improve access for other users. Other mirror sites may be set up in the future.

GENOMICS

Last year, Ellis and colleagues stated that the UM-BBD could be a resource for functional genomics (1); developing its genome potential has since become a priority. I. Saira Mian and her colleagues at the Lawrence Berkeley National Laboratory are developing several different types of models for protein sequence annotation and the prediction of protein folds: Hidden Markov models, Support Vector Machine models and Markov Random Field models. These models will first be tested on UM-BBD enzymes, since these are a well-curated set with, as noted above, a wider than usual catalytic range. Since they are involved in the catabolism of diverse substrates, their folds are expected to span a greater diversity than an equal number of enzymes not so involved. The goal here is to develop a pathway/genome database and system.

BIODEGRADATIVE STRAIN DATABASE

Jim Tiedje and colleagues at Michigan State University are developing a Biodegradative Strain Database (BSD, http://www.cme.msu.edu/CME/BSD.html), an on-line database of information on described, biodegradative microbial strains. The BSD began with information in the UM-BBD's Microbe Index (http://www.labmed.umn.edu/umbbd/search/micro.html) and has now doubled in size. Links will be developed between the UM-BBD's metabolic and the BSD's strain information. In this way, users of either database will be able to determine how to grow the strain, its phylogeny, the reactions it can carry out, and the genes and enzymes it contains, all relevant to biodegradation (J.M.Tiedje and J.Urbance, personal communication).

PREDICTION

Since the UM-BBD will never contain all known biodegradation pathway information and much more biodegradation occurs than is known, it has long been a goal to use UM-BBD information to predict microbial catabolic pathways for compounds it does not contain. The topic can be approached in a number of ways (3,4).

For example, the vast majority of known environmental organic compounds are composed of combinations of one or more of the 41 functional groups shown in Table 1. If the individual reactions required to metabolize each functional group are known, then predicting biodegradation is largely anticipating the effects of neighboring groups and the order in which the different groups are metabolized in the sequence of a catabolic pathway. Though this process can be computationally prohibitive for multi-functional compounds, most organic chemicals of environmental interest have relatively few functional groups.

CONCLUSIONS

The UM-BBD, having met its initial goals, is poised to grow in many different, complementary ways, including: mirror sites; fold prediction for its sequenced enzymes; closer ties to genome and microbial strain databases; and the prediction of biodegradation pathways for compounds it does not contain.

ACKNOWLEDGEMENT

Supported in part by NSF grant DBI-9630427.

REFERENCES

1. Ellis,L.B.M., Hershberger,C.D. and Wackett,L.P. (1999) *Nucleic Acids Res.*, **27**, 373–376.
2. Etzold,T., Ulyanov,A. and Argos,P. (1996) *Methods Enzymol.*, **266**, 114.
3. Wackett,L.P., Ellis,L.B.M., Speedie,S., Hershberger,C.D., Knackmuss,H.-J., Spormann,A., Walsh,C.T., Forney,L.J., Punch,W.F., Kazic,T., Kanehisa,M. and Berndt,D.J. (1999) *Am. Soc. Microbiol. News*, **65**, 87–93.
4. Wackett,L.P. and Ellis,L.B.M. (1999) *Environ. Microbiol.*, **1**, 119–124.