Nucleic Acids Research, 1999, Vol. 27, No. 1 373–376

The University of Minnesota Biocatalysis/ Biodegradation Database: specialized metabolism for functional genomics

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]Center for Biodegradation Research and Informatics, Department of Biochemistry, University of Minnesota, St Paul, MN 55108, USA

Received August 31, 1998; Revised September 30, 1998; Accepted October 14, 1998

ABSTRACT

The University of Minnesota Biocatalysis/Biodegradation Database (UM-BBD, http://www.labmed.umn.edu/umbbd/index.html) first became available on the web in 1995 to provide information on microbial biocatalytic reactions of, and biodegradation pathways for, organic chemical compounds, especially those produced by man. Its goal is to become a representative database of biodegradation, spanning the diversity of known microbial metabolic routes, organic functional groups, and environmental conditions under which biodegradation occurs. The database can be used to enhance understanding of basic biochemistry, biocatalysis leading to speciality chemical manufacture, and biodegradation of environmental pollutants. It is also a resource for functional genomics, since it contains information on enzymes and genes involved in specialized metabolism not found in intermediary metabolism databases, and thus can assist in assigning functions to genes homologous to such less common genes. With information on >400 reactions and compounds, it is poised to become a resource for prediction of microbial biodegradation pathways for compounds it does not contain, a process complementary to predicting the functions of new classes of microbial genes.

INTRODUCTION

Current genome projects are generating a staggering number of gene sequences. GenBank (1; http://www.ncbi.nlm.nih.gov/Web/Genbank/index.html) contained 2.4×10^6 sequences in June 1998 and is growing exponentially, with a doubling time of 1.5 years (National Center for Biotechnology Information News, February, 1998, http://www.ncbi.nlm.nih.gov/Web/Newsltr/feb98.html). While presently workers struggle just to store this massive amount of information, the next major challenge will be functional genomics—determining each gene's function. Functional genomics is relatively straightforward when dealing with genes which code for enzymes involved in intermediary metabolism, the metabolism common to most organisms, such as glycolysis and the citric acid cycle. There are several intermediary metabolism databases such as KEGG (2; http://www.genome.ad.jp/kegg/kegg.html), many homologues for each coding sequence, and the metabolic function of each is well-studied. Functional genomics is more of a challenge when dealing with specialized metabolism. Forty-four percent of the coding sequences in the *Escherichia coli* genome have functions which are presently unknown (3); a large fraction of these may be involved in specialized metabolism.

The microbial world which includes *E.coli* flourishes in part due to its specialized catabolic metabolism. This permits microorganisms to gain energy from compounds which others cannot eat; for example, some *Pseudomonas* species can catabolize >1000 compounds. This catabolic metabolism is termed biodegradation. These esoteric feedstocks can include environmental pollutants, and microbes can thus be used to clean up pollution, a process called bioremediation. Specialized microbial catabolism can also be used synthetically, a process called biocatalysis. The University of Minnesota Biocatalysis/Biodegradation Database (UM-BBD, http://www.labmed.umn.edu/umbbd/index.html) started on the web in 1995 to provide information on the specialized microbial catabolic pathways and metabolic reactions which are important in biotechnology. Besides improving bioremediation efforts and biosynthetic processes, the UM-BBD is also a resource for functional genomics. UM-BBD users should cite this paper or its successors as its primary reference.

DATABASE CONTENT AND METHODS

The UM-BBD now contains information on >400 reactions and compounds and 300 enzymes, organized in 68 annotated metabolic pathways, for compounds from acrylonitrile to xylene. Figure 1 is an overview of UM-BBD contents. The circled compounds are routes to intermediary metabolism. The other compounds shown in Figure 1 funnel into these circles via the specialized metabolism that is the province of the UM-BBD. The database contains detailed information for each of these metabolic pathways, in both textual and graphical form. As an example of this, the information on the biodegradation of the compound acrylonitrile, boxed on the left in Figure 1, is shown

*To whom correspondence should be addressed. Tel: +1 612 625 9122; Fax: +1 612 625 1121; Email: lynda@email.labmed.umn.edu

in more detail in Figure 2. A graphical pathway for acrylonitrile biodegradation is shown in Figure 2A. Each boxed enzyme or compound name in Figure 2A links to a UM-BBD reaction (Fig. 2B) or compound (Fig. 2D) page. Each UM-BBD reaction page (Fig. 2B) includes a text depiction of a reaction, balanced for mass and charge; internal links to graphics of the reaction (Fig. 2E) and optionally to graphics of the reaction mechanism (Fig. 2C); and external links to static documents from, and dynamic searches of, relevant enzyme, gene and bibliographic databases. The UM-BBD is the only metabolic database to include such labor-intensive dynamic links. Since UM-BBD compounds are often interesting toxicologically or chemically, each UM-BBD compound page (Fig. 2D) has links to toxicology and/or other chemical databases.

Database format

The UM-BBD was prototyped in HTML form. When usage demonstrated that its format was used and useful, its information was transferred to the Compounds, Organisms, Reactions and Enzyme (CORE) database management system, written in Java (4). CORE contains the information needed to generate reaction and compound pages and can dynamically generate pathway maps starting from any UM-BBD reaction (Fig. 2B).

Database update

The UM-BBD is updated at least monthly. Updates include new pathways, new reactions added to existing pathways, new search tools, and other features. The pathways to be added are usually based on the scientific literature and are occasionally suggested by our users. When a contribution is based on unpublished work, our International Scientific Advisory Board [Ellis,L.B.M. (1998) UM-BBD Contributors Page. http://www.labmed.umn.edu/umbbd/contrib.html] reviews the submission for scientific accuracy. Much of the data entry and graphics are done by database staff but some pathways are entered by distant learners enrolled in a Biocatalysis and Biodegradation course taught completely over the Internet [Ellis,L.B.M. and Wackett,L.P. (1998) BioC/MicE 5309, Biocatalysis and Biodegradation. http://www.cee.umn.edu/biodeg/].

Data access

The home page (http://www.labmed.umn.edu/umbbd/index.html) contains a scrollable list of UM-BBD pathway maps, each of which links to compound and reaction pages. It also contains links to searchable lists of UM-BBD pathways, reactions, enzymes, compounds and microorganisms, and lists of pathway map and enzyme mechanism graphics. Other more specialized lists include UM-BBD reactions which have not been studied in enough detail to be assigned enzymes. Such a list can be used to define areas in specialized microbial metabolism where further work is needed.

Compounds can be searched for by full or partial name, CAS Registry Number and chemical formula. Compound searches return links to UM-BBD compound pages and any UM-BBD reaction pages in which the compound is produced or consumed. Enzymes can be searched for by full or partial name and full or partial EC code. Enzyme searches return a list of links to UM-BBD reaction pages catalyzed by that enzyme or, if there is only one such reaction, link directly to that reaction page. Microorganism entries can be searched by full or partial name and return links to UM-BBD pathways that involve that microorganism.

As mentioned earlier, the UM-BBD links to related databases. Some of these links are reciprocal and UM-BBD data can be accessed from those databases. Ligand Chemical Database at Kyoto University (5; http://www.genome.ad.jp/dbget/ligand.html) links to UM-BBD enzyme pages; Entrez PubMed, Nucleotide and Protein Databases (6; http://www.ncbi.nlm.nih.gov/Entrez/) link to UM-BBD reaction pages; and ChemFinder from CambridgeSoft, Inc. (7; http://chemfinder.camsoft.com/) links to UM-BBD compound pages.

Applications

The UM-BBD can be used for understanding of basic biochemistry, biocatalysis leading to speciality chemical manufacture, and biodegradation of environmental pollutants. Based on the 1998 [Ellis,L.B.M. (1998) UM-BBD 1998 User Survey. URL, http://www.labmed.umn.edu/umbbd/stats/results3.html] and earlier user surveys, UM-BBD information primarily supports pure and applied research. Industries increasingly need to know the environmental fate of their commercial chemicals, and this is largely governed by the metabolism of these chemicals by soil and water microorganisms. Commercial users have used our information as part of EPA reports; environmental law firms have used it to prepare their cases. Biotechnology companies increasingly turn to biocatalysis for new advances in speciality chemical manufacture. One example that predates the UM-BBD is the application of naphthalene dioxygenase, used naturally to catabolize polycyclic aromatic hydrocarbons, to produce the blue jean dye indigo in fermentation vessels (8). This is an example of the emerging field of green chemistry which will draw more heavily on specialized bacterial enzymes than on the more commonly studied enzymes of intermediary metabolism.

The UM-BBD can also be used to predict biodegradation reactions, both practically and theoretically. As an example of the former, functional genomic analysis based on the UM-BBD is currently being used to predict the metabolism of *Deinococcus radiodurans*, an organism which is of interest because of its extreme resistance to ionizing radiation but whose metabolic pathways are poorly studied. Since numerous hazardous waste sites contain high level radioactivity and toxic organic waste, a radiation-resistant bacterium is needed for bioremediation purposes. Genome analysis of *D.radiodurans* revealed a paucity of biodegradative enzymes of the type found in soil *Pseudomonas* species and thus recombinant biodegradative genes were cloned and expressed in the organism (9). The UM-BBD was used both in the functional genomic analysis and in deciding which biodegradative genes to experimentally express in *D.radiodurans*.

The theoretical prediction of biodegradation metabolism is also very important to industry and requires information of the type provided in the UM-BBD. The United States Environmental Protection Agency must decide annually on the environmental acceptability of perhaps 500 new compounds. In most cases, information on their pathways of biodegradation is lacking and too expensive and time-consuming to obtain experimentally. Thus, it is increasingly important to predict the biodegradative metabolism of new organic compounds based on known biodegradation reactions. This requires information databases such as the UM-BBD and expert system approaches whereby that knowledge can be used to predict new metabolism. This is in some sense the converse of the functional genomics problem in that the goal is determining the likelihood that microorganisms

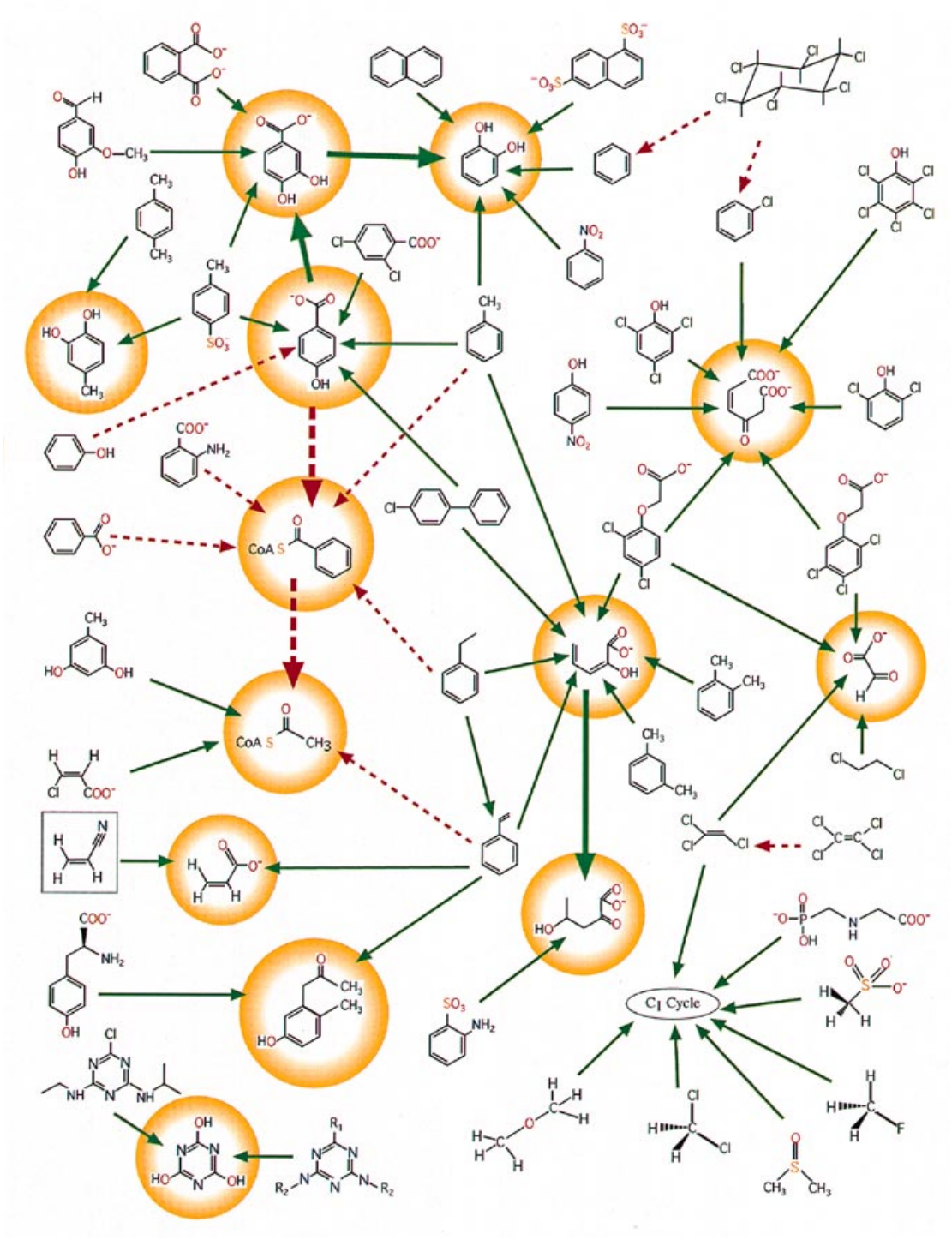

Figure 1. Graphical overview of UM-BBD content. The circled compounds lead to intermediary metabolism. The UM-BBD contains information on the metabolic reactions which funnel the other compounds (and others not shown) to the circles. Solid arrows indicate reactions carried out under aerobic conditions; dashed arrows, under anaerobic conditions. UM-BBD information for the biodegradation of acrylonitrile, the boxed compound on the left, is shown in more detail in Figure 2. URL, http://www.labmed.umn.edu/umbbd/meta/meta_map.html

collectively will contain genes and enzymes to metabolize a given compound. We are beginning to predict biodegradation pathways in a project called Predict-BT [Wackett,L.P., Ellis,L.B.M., Speedie,S. and Hershberger,C.D. (1998) PredictBT: The University of Minnesota Predictive Biodegradation Project. URL, http://www.labmed.umn.edu/umbbd/predictbt/], building on the information contained in the UM-BBD, intermediary metabolism, and gene sequence databases.

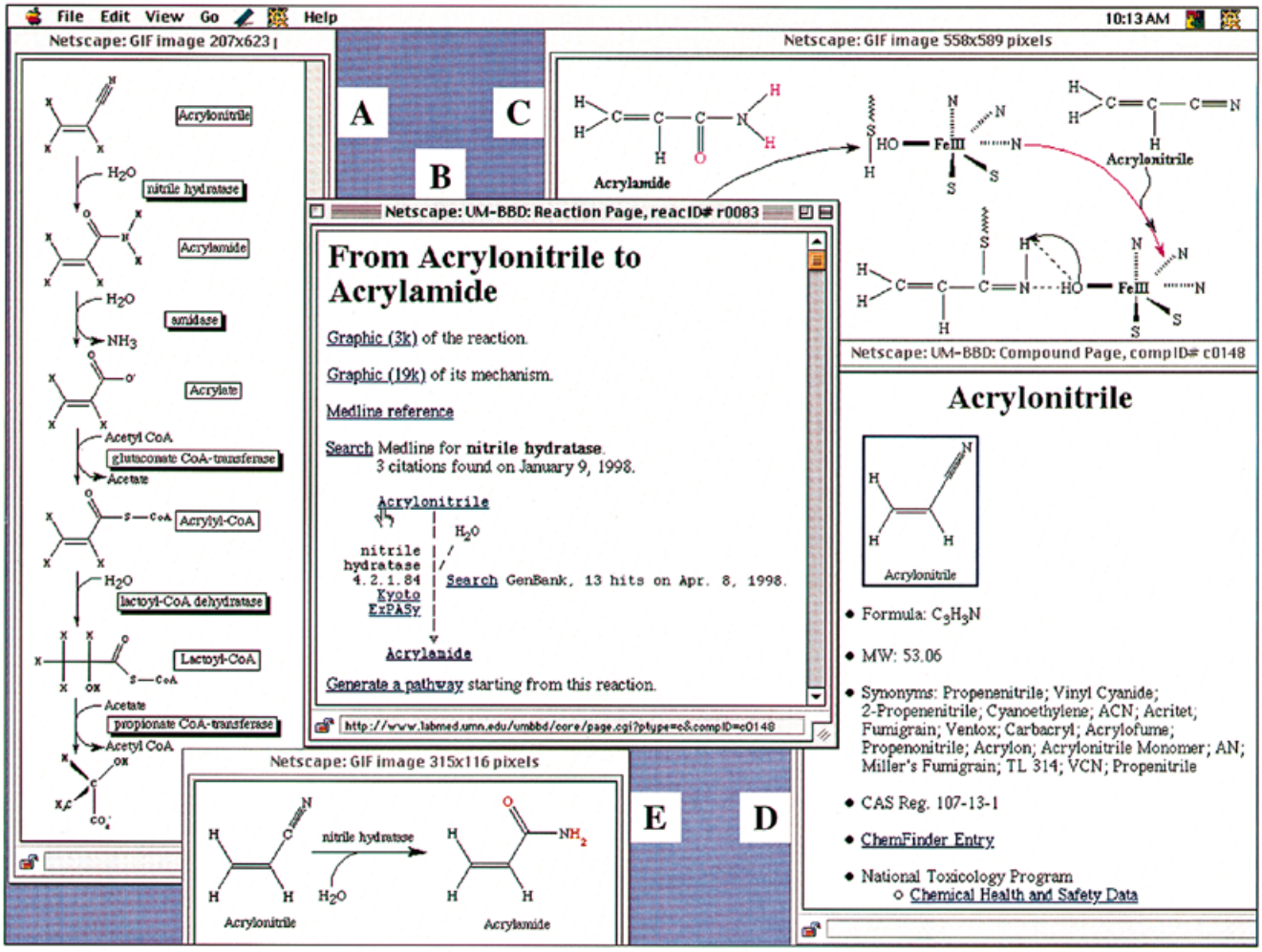

Figure 2. Example UM-BBD pathway information. (**A**) Graphical pathway map for the acrylonitrile pathway; (**B**) reaction page for the nitrile hydrolase reaction; (**C**) reaction mechanism graphic for nitrile hydrolase; (**D**) compound page for acrylonitrile; (**E**) reaction graphic for nitrile hydrolase. URL, http://www.labmed.umn.edu/umbbd/acr/acr_map.html

Future directions

The UM-BBD now has ~400 reactions and compounds, a very small fraction of the 10 million organic compounds currently known. It will never include all compounds and that is not its goal. Instead it is to become a representative database of biodegradation, spanning known metabolic routes, organic functional groups metabolized, and environmental conditions under which biodegradation occurs. The next goal will be to use this information to predict the metabolism of compounds the UM-BBD does not contain.

CONCLUSIONS

Functional genomics has the potential to decode the physiological meaning of an organism's genetic information. Its success will be limited without a greater knowledge, in experiment and representation, of the enzymes which mediate the more esoteric metabolism found in the bacterial world. The UM-BBD can assist in this and in fostering biotechnology and proper environmental stewardship.

ACKNOWLEDGEMENTS

We thank Jingfeng Feng and Stuart Speedie for helpful discussions. This work was supported in part by National Science Foundation Award 9630427, National Institutes of Health Award R01-GM56529 and an NIH Training Grant LM07041 traineeship to C.D.H.

REFERENCES

1 Benson,D.A., Boguski,M.S., Lipman,D.J., Ostell,J. and Ouellette,B.F.F. (1998) *Nucleic Acids Res.*, **26**, 1–7.
2 Kanehisa,M. (1998) In Letovsky,S. (ed.) Molecular Biology Databases. Kluwer Academic Press, Dordrecht, The Netherlands. In press.
3 Blattner,F.R., Plunkett,G., Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.K., Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J., Mau,B. and Shao,Y. (1997) *Science*, **277**, 1453–1462.
4 Ellis,L.B.M., Speedie,S. and McLeish,R. (1998) *Bioinformatics*, in press.
5 Goto,S., Nishioka,T. and Kanehisa,M. (1998) *Bioinformatics*, **14**, 591–599.
6 McIntyre,J. (1998) *Trends Genet.*, **14**, 39–40.
7 Brecher,J. (1998) *Chimia*, in press.
8 Bialy,H. (1997) *Nature Biotechnol.*, **15**, 110.
9 Lange,C.C., Wackett,L.P., Minton,K. and Daly,M. (1998) *Nature Biotechnol.*, **16**, 929–933.