# Microbial Pathway Prediction: A Functional Group Approach

Bo Kyeng Hou,[†] Lawrence P. Wackett,[†] and Lynda B. M. Ellis*[,‡]

Biological Technology Institute, University of Minnesota, St. Paul, Minnesota 55108, and Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, Minnesota 55455

Received January 29, 2003

We have developed a system to predict microbial catabolism, using the University of Minnesota Biocatalysis/Biodegradation Database (UM-BBD, http://umbbd.ahc.umn.edu/) as a knowledge base. The present system, available on the Web (http://umbbd.ahc.umn.edu/predict/), can predict biodegradation of most of the major aliphatic and aromatic organic functional groups containing C, H, N, O, and halogens. It can duplicate at least one known biodegradation pathway for 60% of the compounds in a 84-member validation set; most pathways that did not completely duplicate known metabolism could plausibly occur in nature. Users are encouraged, and have begun, to submit additional biotransformation rules and comment on existing rules; the system will further develop under the direction of the scientific community.



## INTRODUCTION

Over 18 million chemical compounds are known, with more than 65 000 currently used in commerce. Though the biodegradability of chemicals in the environment is largely predicated on their ability to serve as substrates for microbial enzymes, this ability has been determined for only a small percentage of these chemicals. To fill this gap and since these compounds may have more toxic metabolites, there is increasing interest in using computational methods to predict biodegradation pathways.[1]

Two such systems are META and METEOR. META[2−4] and METEOR[5−9] were both initially designed to predict pathways for the mammalian detoxification of drugs and environmental pollutants. META later added prediction of microbial biodegradation.[10,11] The accuracy of META, METEOR, and similar systems in predicting biodegradation reactions will be as good as the databases and rules that these systems draw upon. While knowledge of biodegradation reactions is incomplete, any system should draw on the entire breadth of known information.

The University of Minnesota Biocatalysis/Biodegradation Database[12] (UM-BBD, http://umbbd.ahc.umn.edu/) provides curated information on over 800 microbial catabolic reactions, emphasizing the breadth of biochemical reaction types. The UM-BBD has compiled and organized information on biochemical reactions involving over 50 chemical functional groups. Some functional groups can undergo multiple types of biodegradation reactions; the database attempts to represent them all. This information has the potential to represent the metabolism of millions of chemical compounds comprised of mixtures of these functional groups.

The present study was undertaken to mine the curated UM-BBD information on functional group biotransformation to develop a tool for predicting microbial catabolism. To avoid combinatorial explosion, that is, generating too many metabolic pathways, the current system predicts all plausible biotranformations for a given compound and permits the user to decide which route(s) to follow. The prediction software is available to all on the Web (http://umbbd.ahc.umn.edu/predict/), allowing the international scientific community to use and comment on it.

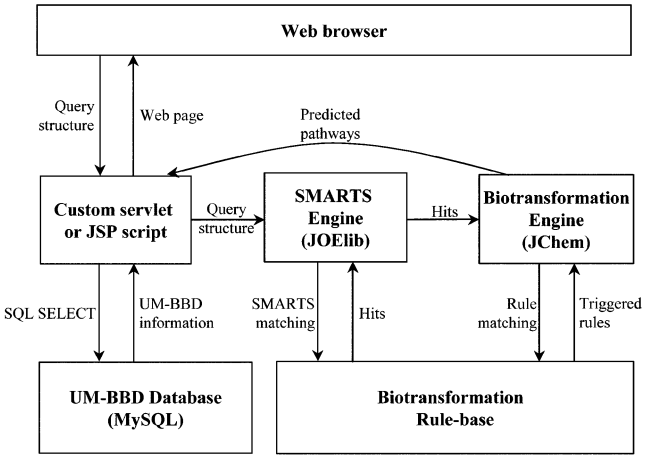


**Figure 1.** System architecture.

## MATERIALS AND METHODS

The Pathway Prediction System uses hardware, software, chemical representations, and databases to perform three main tasks: SMARTS matching, rule matching, and biotransformation. System architecture is shown in Figure 1. It was evaluated by examining its ability to predict known UM-BBD pathways.

**Hardware, Software, Chemical Representations, and Databases.** A SUN Ultra 4 computer running the SunOS 5.7 operating system was used. Code was written in Java, SQL, and JSP (JavaServer Pages).[13] Java code was developed using the Java compiler, Java Virtual Machine, and classes supplied with the Java Development Kit (JDK) 1.3.1 from Sun.[14] Two Marvin Java applets[15] from ChemAxon, Inc.[16] are part of the system. The MarvinSketch applet (shown in

* Corresponding author phone: (612) 625-9122; fax: (612) 624-6404; e-mail: Lynda@tc.umn.edu. Corresponding author address: Mayo Mail Code 609, 420 SE Delaware Street, Minneapolis MN 55455.
[†] Biological Technology Institute.
[‡] Department of Laboratory Medicine and Pathology.

10.1021/ci034018f CCC: $25.00 
Published on Web 04/04/2003

**Table 1.** UM-BBD Organic Functional Groups Containing C, H, O, N, X[a]

| | | |
|---|---|---|
| alcohol, primary | alcohol, secondary | alcohol, tertiary |
| aldehyde | alkane, primary | alkane, secondary |
| alkane, tertiary | alkene | alkyne |
| amide | amine, primary | amine, secondary |
| amine, tertiary | bicycloaliphatic ring | biphenyl-type benzenoid ring |
| carboxylic acid | carboxylic acid ester | cycloaliphatic ring |
| cyanamide | diazo | epoxide |
| ether | N-heterocyclic ring, saturated | N-heterocyclic ring, unsaturated |
| O-heterocyclic ring | ketone | monocylic aromatic hydrocarbon |
| nitrate ester | nitrile | nitro |
| organohalide | oxime | peroxide |
| polycyclic aromatic hydrocarbon | tricycloaliphatic ring | |

[a] See text.

**Table 2.** 10 Biotranformation Rules for Guided Pathway Prediction[a]

| ruleID | description |
|---|---|
| bt0001 | primary alcohol → aldehyde |
| bt0002 | secondary alcohol → ketone |
| bt0003 | aldehyde → carboxylic acid |
| bt0004 | mono- or unsubstituted benzenoid → 3,4-*cis*-dihydroxydihydrobenzenoid |
| bt0005 | mono- or unsubstituted benzenoid → 1,2-*cis*-dihydroxydihydrobenzenoid |
| bt0006 | 1,2-*cis*-dihydoxydihydrobenzenoid → 1,2-dihydroxybenzenoid |
| bt0007 | 3,4-*cis*-dihydoxydihydrobenzenoid → 3,4-dihydroxybenzenoid |
| bt0008 | 1,2-dihydroxybenzenoid → proximal extradiol ring cleavage |
| bt0009 | 1,2-dihydroxybenzenoid → intradiol ring cleavage |
| bt0010 | 1,2-dihydroxybenzenoid → distal extradiol ring cleavage |

[a] A complete list of rules is found at http://umbbd.ahc.umn.edu/servlets/pageservlet?ptype=allrules.

Figure 8) can be used to draw the compound whose biodegradation is to be predicted. The MarvinView applet (shown in Figures 9 and 10) is used to display predicted pathways. Chemical structures are represented by SMILES (Simplified Molecular Input Line Entry System) strings.[17−19] The SMARTS (SMiles ARbitrary Target Specification) language is an extension of SMILES for creating patterns of organic functional groups.[20] The SMARTS engine to carry out chemical substructure matching was developed using the JOElib library.[21] The biotransformation engine was developed using the JChem package[15] from ChemAxon, Inc.[16]

Two databases, the UM-BBD and a database of biotransformation rules, support the system. At the end of 2002, the UM-BBD contained information on over 130 metabolic pathways, over 800 reactions, almost 800 compounds, and over 500 enzymes. The information on UM-BBD compounds, reactions, and enzymes is stored in a MySQL[22] database and its Web pages are dynamically generated through JSP scripts. The 80 rules in the biotransformation rule database (rule-base) specify the parts of a query compound that are transformed and the nature of these biotransformations. JSP scripts create the list of UM-BBD reactions related to each biotransformation rule. The rules are described in more detail below.

**Biotransformation Rules.** Rules were developed for biotransformations if at least one example of such metabolism is found in the UM-BBD or is otherwise known to occur in the environment. Thirty-five organic functional groups containing only C, H, O, N, F, Cl, Br, and I found in the UM-BBD (Table 1) are the focus of the present Pathway Prediction System (PPS). Each functional group may be transformed in several ways. For example, a cyano group might be hydrated with one water molecule to form an amide, hydrolyzed by two water molecules to form carboxylic acid and ammonia, or reduced by six electrons (for example, by nitrogenase) to yield a methyl group and ammonia. These three types of reactions are represented by rules bt0028, bt0030, and bt0031, respectively. The Nomenclature Commission of the International Union of Biochemistry and Molecular Biology (IUBMB) has developed a four-level hierarchical classification scheme based on enzyme reaction types.[23] UM-BBD organic functional groups and IUBMB classifications of UM-BBD enzymes are used to create biotransformation rules.

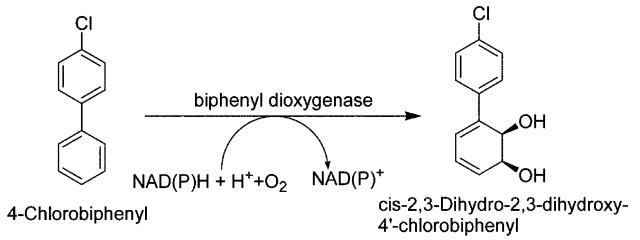


**Figure 2.** UM-BBD reaction example of the first step in ring cleavage, bt0005.

Each rule identifies the functional group it transforms and the biotransformation; the first 10 are shown in Table 2. An example of a rule is aldehyde → carboxylic acid (bt0003). The system can identify an aldehyde and convert it to a carboxylic acid, wherever it appears in an organic compound. A more complicated rule (bt0008) is shown in Figure 2: A mono- or unsubstituted benzenoid with two consecutive unsubstituted ring positions can be converted to a 1,2-*cis*-dihydroxydihydrobenzenoid (the first step leading to ring cleavage). Each of the biotransformation rules has a Web page (Figure 3), including a link to one or more UM-BBD reactions, each with bibliographic references, or the word "spontaneous" for abiotic reactions. An e-mail link allows users to make comments on each rule. With input from the scientific community, present rules can be improved and additional rules can be created.

# Rule bt0003

[Pathway Prediction Engine] [All Rules List] [BBD Main Menu]

**Description:**
bt0003: Aldehyde -> Carboxylic acid

**UM-BBD Reaction(s):**
Benzaldehyde -----> Benzoate (reacID# r0269)
...
2-Methylbenzaldehyde -----> *o*-Methylbenzoate (reacID# r0222)
3-Methylbenzaldehyde -----> *m*-Methylbenzoate (reacID# r0214)
1-Naphthaldehyde -----> 1-Naphthoic acid (reacID# r0787)
2-Naphthaldehyde -----> 2-Naphthoic acid (reacID# r0772)
1-Octanal -----> Octanoate (reacID# r0023)
6-Oxohexanoate -----> Adipate (reacID# r0175)
Salicylaldehyde -----> Salicylate (reacID# r0339)
*p*-Tolualdehyde -----> *p*-Toluate (reacID# r0177)
Vanillin -----> Vanillate (reacID# r0145)

If you have any comments on rule bt0003, email BBDMaster@mail.ahc.umn.edu

[Pathway Prediction Engine] [All Rules List] [BBD Main Menu]

**Figure 3.** A portion of a biotransformation rule Web page. Ten of the 22 UM-BBD reactions demonstrating this rule are shown. Reactions link to UM-BBD reaction pages. The complete rule can be seen at http://umbbd.ahc.umn.edu:8015/umbbd/rule.jsp?rule=bt0003.

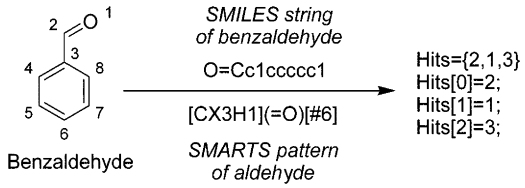


**Figure 4.** Example of SMARTS pattern matching.

**SMARTS Matching.** To find and list the organic functional groups in a query compound, and use UM-BBD data to predict how each functional group is transformed, the SMARTS pattern of each functional group is first defined. The presence and location of functional groups in a query compound guide the rule matching by the zero-based indices of the matching atoms in each SMARTS pattern (IF-part) of all biotransformation rules. For example, the oxygen atom from an aldehyde defined with the SMARTS pattern "[CX3H1](=O)[#6]" is referred to as "Hits[1]". If the query compound is benzaldehyde ("O=Cc1ccccc1" in SMILES format), the mappings result of the query compound and the SMARTS pattern is "Hits={2,1,3}", and the position of the oxygen atom in the aldehyde is "Hits[1]=1" as shown in Figure 4.

A more complex example is "bt0021: alkene → alcohol." Markovnikov's Rule[24] is followed in most, but not all, UM-BBD alkene hydration reactions. Rule bt0021 implements Markovnikov's rule and thus considers all four R groups around the double bond in an aliphatic C=C substructure and classifies them as electron donating (most C groups) or electron withdrawing (C=O, etc.). If all R groups are electron donating, the side with fewer R groups (more H) receives the "H" and the other side receives the "OH". If some R groups are electron withdrawing, the side with the largest number of electron withdrawing groups receives the "H" and the other side receives the "OH". If the sides have equal numbers, then both alternatives are shown. Users are informed of this and similar information for other complex rules in an optional "Comments" field on rule Web pages.

Multiple recursive SMARTS definitions are needed to match some atoms around the double bond for bt0021. However, such complex constraints cannot be encoded in the single SMARTS pattern used in a rule. Instead, the simplified SMARTS pattern, "[C;!R1]=[C;!R1]", which means "a carbon−carbon double bond not in a ring", is used in the SMARTS matching step, and the more complex constraints are checked in the rule matching step, using a Constraint Check function described below.

**Rule Matching.** As mentioned under System Architecture, a biotransformation rule consists of two parts: a SMARTS pattern (IF-part) and an assembly execution code (THEN-part). The SMARTS pattern defines the particular organic functional group targeted by each rule. The assembly execution code provides explicit instructions on how to transform the query molecule containing this functional group into a product. This code is translated into a sequence of parametrized transformation functions in the rule matching step.

The system uses the transformation functions in the JChem package[15,16] including addition of predefined functional groups, addition of atoms, removal of bonds, changing of the bond order, and changing of double bond stereochemistry. The system also uses a Constraint Check function written using the JOElib library.[17] This function finds an exact reaction point in case of multiple conditional constraints. For example, for "bt0021: alkene → alcohol", the code "$a[] = CC(Hits, bt0021);" returns reaction points after classifying all four R groups around the double bond into electron donating or electron withdrawing groups. The Constraint Check function is used to manage constraints too complicated for the SMARTS engine to check.

The sequence of transformation functions is executed with the appropriate parameters, which include the atom list for the query compound matched by the SMARTS pattern of each rule. The rule scripts for "bt0003: aldehyde → carboxylic acid", "bt0008: 1,2-dihydroxybenzenoid → proximal extradiol ring cleavage", and "bt0021: alkene → alcohol" are shown in Figure 5.

**Biotransformation.** When the above rule matching procedure is finished, a sequence of parametrized biotransformation functions in the THEN-part of each matched rule executes to transform a query compound into its product or products. For example, if rule bt0003 is triggered by the

```
a) ;;bt0003: aldehyde -> carboxylic acid.

IF  {"[CX3H1](=O)[#6]"}
THEN {
    AA(r_point=Hits[0], atom="O", bond_order=1, down_up=0, charge=-1);
}

b) ;;bt0008: 1,2-dihydroxybenzenoid->proximal extradiol ring cleavage.
IF  {"[cR1;H0,H1]:[cR1]:[cR1]:[cR1;H0]([O;H1]):[cR1;H0][O;H1]"}
THEN {
    BC(atom1=Hits[0], atom2= Hits[1], bond_order=1, down_up=0);
    BC(atom1= Hits[1], atom2= Hits[2], bond_order=2, down_up=0);
    BC(atom1= Hits[2], atom2= Hits[3], bond_order=1, down_up=0);
    BC(atom1= Hits[3], atom2= Hits[4], bond_order=2, down_up=0);
    BC(atom1= Hits[4], atom2= Hits[6], bond_order=1, down_up=0);
    BC(atom1= Hits[6], atom2= Hits[0], bond_order=2, down_up=0);
    AC(atom= Hits[7], charge=-1);
    AA(r_point=Hits[0], atom="O", bond_order=2, down_up=0, charge=0);
    AA(r_point=Hits[6], atom="O", bond_order=2, down_up=0, charge=0);
    BR(atom1=Hits[0], atom2=Hits[6]);
    SC(atom1= Hits[0], atom4= Hits[3], stereo_flag=CIS);
    SC(atom1= Hits[2], atom4= Hits[6], stereo_flag=CIS);
}

c) ;;bt0021: alkene -> alcohol.
IF  {"[C;!R1]=[C;!R1]"}
THEN {
    $a[] = CC(Hits, bt0021);
    AA(r_point=Hits[$a], atom="O", bond_order=1, down_up=0, charge=0);
}
```

**Figure 5.** Example rule scripts. (a) bt0003: aldehyde → carboxylic acid; (b) bt0008: 1,2-dihydroxybenzenoid → proximal extradiol ring cleavage; (c) bt0021: alkene → alcohol.

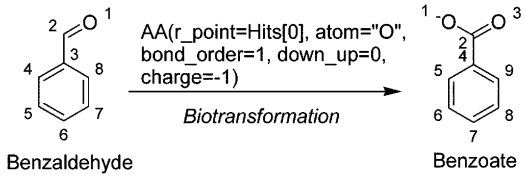


**Figure 6.** Example biotransformation.

query compound benzaldehyde, then the biotransformation engine adds an oxygen with a single bond into an carbon atom of the query compound, "Hit[0]=2", as shown in Figure 6.

**Evaluation.** On December 9, 2002, 119 organic compounds initiated one or more UM-BBD pathway branches; that is, were not the product of a UM-BBD reaction. Of these "highest" compounds, 84 compounds contain only C, H, O, N, F, Cl, and/or Br and initiated pathways containing two or more reactions. These form the validation set for the project, since biotransformations of other atoms are not presently included in the Pathway Prediction System (PPS). Names of the 84 compounds are in Supporting Information.

The PPS Evaluation system automatically submitted each compound in the validation set to the PPS and followed the branch or branches that mirrored UM-BBD metabolism as known in early December, 2002. A successful pathway prediction was one in which at least one such branch continued until a compound is reached that, in the UM-BBD, was either a "dead-end" compound whose metabolism has not yet been determined or one that linked to a commonly metabolized compound in the KEGG database.[26] An unsuccessful prediction reached a point in all branches when it no longer mirrored UM-BBD metabolism. In addition, when a successful prediction was made, the number of compounds "downstream" of the initiating compound that also have correctly predicted degradation was determined.

## RESULTS

**Pathway Prediction System.** The user can access the microbial pathway prediction Web page through most Web browsers with Java and JavaScript support. The Pathway Prediction System (PPS) main Web page is shown in Figure 7. The structure of the chemical compound whose degradation is to be predicted can be entered as a SMILES string or drawn. If the latter input method is used, a SMILES string is generated from the drawn structure. The input SMILES string is checked, and a warning message is displayed if it is not valid.

The SMARTS engine performs SMARTS matching, the process of finding particular patterns of functional groups in a query compound. Biotransformation rules, stored in a biotransformation rule-base, contain SMARTS patterns in their IF-parts. The SMARTS engine attempts to map each

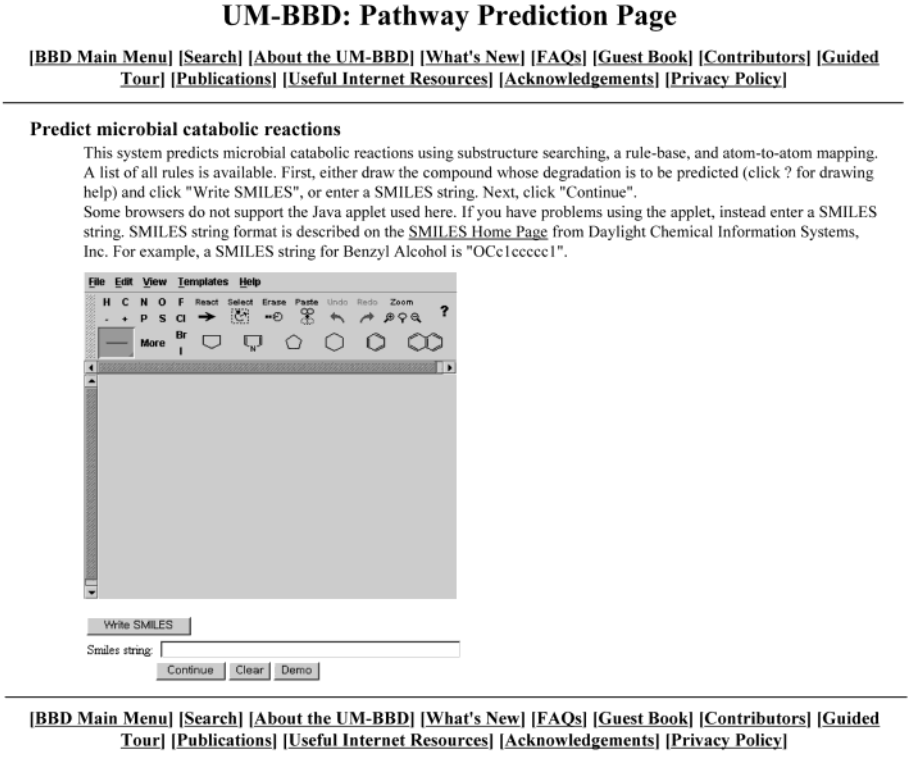


**Figure 7.** The pathway prediction home page, http://umbbd.ahc.umn.edu/predict/.

**Table 3.** 27 Pathway Termination Compounds[a]

| | | | |
|---|---|---|---|
| acetaldehyde | acetate | acetoacetate | acetylene |
| butyrate | ethane | ethanol | ethylene |
| formaldehyde | formate | glycol | glycolate |
| glyoxylate | 2-hydroxybutyrate | 3-hydroxybutyrate | 2-ketobutyrate |
| L-lactate | malate | malonate | methane |
| methanol | oxalate | oxaloacetate | propane |
| propionate | pyruvate | succinate | |

[a] See text.

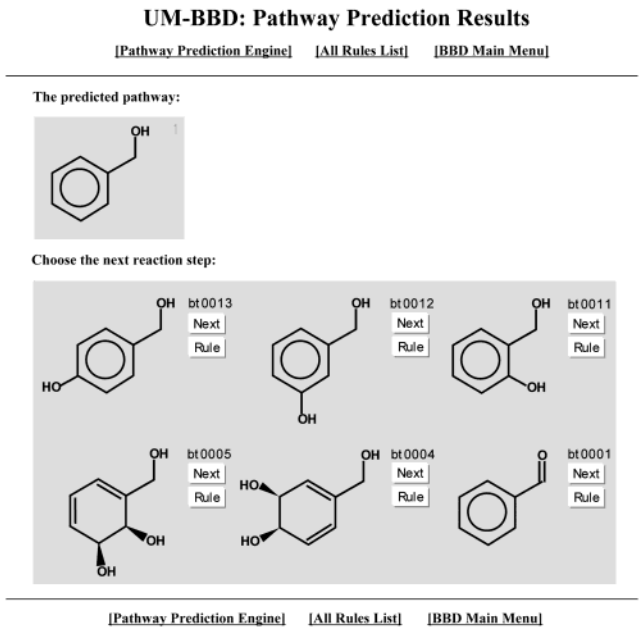


**Figure 8.** First step, showing six possible biotransformations.

rule's SMARTS pattern onto the SMILES string of the query compound. Hits are sent to the biotransformation engine, which collects and parametrizes them and transforms a query compound into plausible products by reference to a sequence of compiled transformation code (THEN-part) of each triggered rule. The prediction terminates when the pathway reaches a compound which either cannot be degraded using existing rules or is on a list of 27 termination compounds (Table 3). These latter are either easily metabolized by most microorganisms and/or are gases easily lost from the local environment, such as ethane.

The UM-BBD Guided Pathway Prediction starting from the compound benzyl alcohol generates six possible initial biotransformations (Figure 8). The user proceeds by selecting "next" button to indicate the next compound to be transformed or "rule" button to see the biotransformation (bt) rule for that transformation, including, where possible, a link to one or more UM-BBD reactions, each with a bibliographic reference. Figure 9 shows a further step in one pathway branch prediction. The compounds in the pathway are numbered 1−6, in pathway order. This pathway continues by adding water to a double bond (bt0021) or decarboxylation (bt0051).

**Evaluation.** The Pathway Prediction System has been tested in order to evaluate its performance. The validation set contained 84 UM-BBD "highest" compounds with known biodegradation. One or more known UM-BBD biodegradation pathways can be duplicated for 50 of them (60%). The names of the 84 compounds, and whether at least one known biodegradation pathway could be, or could not be, completely

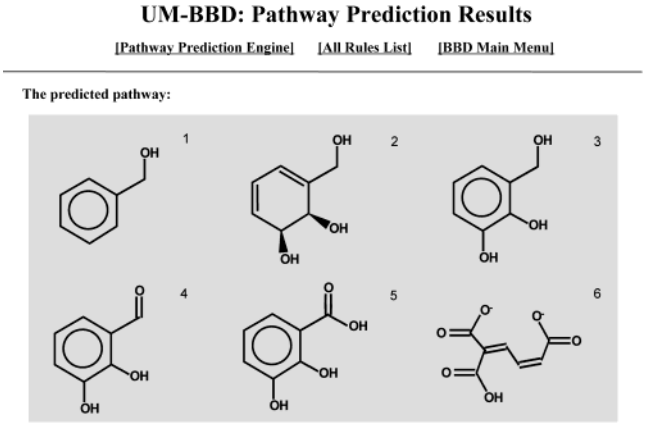


**Figure 9.** First six steps in one predicted pathway branch initiated in Figure 8.

predicted by the present system, are found in the Supporting Information.

## DISCUSSION

Biodegradation content experts have compiled biotransformation rules from an in-depth analysis of the UM-BBD information, and chemical informaticians have implemented these rules in the present biodegradation prediction engine. These are important steps toward the goal of developing an expert system to automatically predict plausible biodegradation pathways with high accuracy. The present guided system requires a knowledgeable user's input to limit the combinatorial explosion possible when generating a metabolic tree.

Evaluation of the present system is conservative. The correct prediction of the biodegradation of 60% of the 84 highest UM-BBD compounds includes correct predictions of 196 compounds below these in the pathways. Though the known biodegradation was not completely predicted for 34 highest compounds, pathways that might plausibly occur in nature were predicted for a majority of them. For example, the UM-BBD depicts biodegradation of the compound 2,4,6-trichlorophenol through 2,6-dichlorophenol, 2-chloromaleyl acetate, maleyl acetate, and 3-oxoadipate.[26] One branch of the predicted pathway for this compound includes oxidative dehalogenation (using bt0022), followed by reductive dehalogenation to hydroquinone (bt0029), oxidative hydroxylation to 1,2,4-benzentriol (bt0016), and ring cleavage and further transformation to maleylaceate (bt0018 and bt0044). Both schemes are shown in Figure 10. While this predictive logic is foreign to some experimentalists, who may carry out reactions exclusively under either oxidative or reductive conditions, compounds in many environments readily shuttle between the two. This predicted pathway is plausible.

Future work will focus on refining present rules and generating new rules to predict the following: the known metabolism of compounds in the validation set whose biodegradation was not completely predicted (Table 2, Supporting Information); metabolism of functional groups containing S, P, and other elements not handled in the present system; and anerobic microbial catabolism. As an example of present rules to be refined, the present system correctly handles ring oxidation for mono- and unsubstituted benzenoids. While di-, tri-, tetra-, or pentasubstituted benzenoids may also be subject to microbial ring oxidation, predicting plausible attack sites requires information on physicochem-

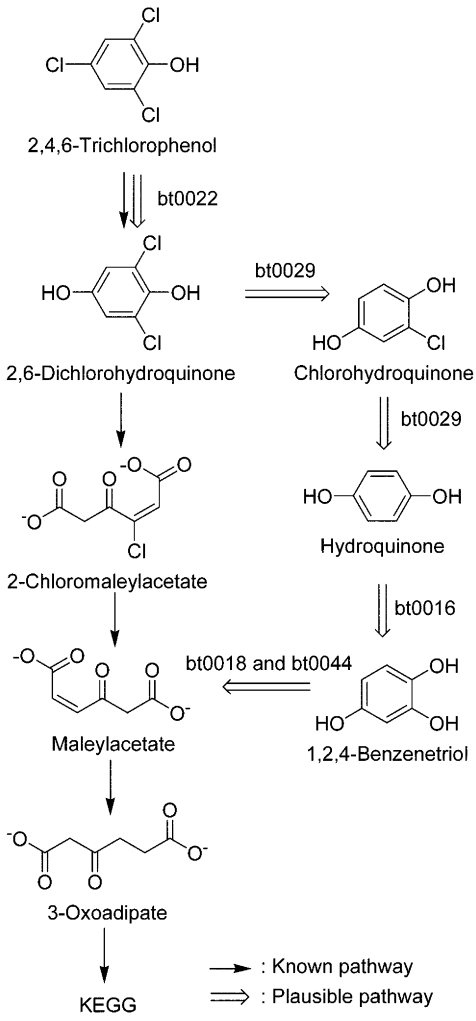


**Figure 10.** Known and plausible predicted biodegradation for 2,4,6-trichlorophenol. The biotransformation rules triggered for each predicted reaction are shown (see text).

ical properties of substituents, not presently available to the system. Regarding anaerobic microbial metabolism, the present system does contain a rule for anaerobic reductive dehalogenation (bt0029). However, some anaerobic biodegradation pathways include an anabolic step; for example, the addition of coenzyme A (CoA) in the anaerobic biodegradation of benzoate.[27] The present system will be expanded to handle this type of metabolism.

The knowledge-based prediction systems META and METEOR generate a metabolic tree of possible detoxification or catabolic reactions via a more complex process, using rules that assign priority for some reactions over others,[28,29] and basing metabolic rules on physicochemical parameters such as logP, the octanol/water partition coefficient.[6,10] We plan to incorporate priority rules and chemical information including chemical similarity and physicochemical property estimation, into our system; to further develop machine-learning programs to extract biodegradation knowledge from the UM-BBD; and define more elaborate heuristics on how to apply this knowledge.

While it may be problematic to determine an appropriate generalized rule for specific UM-BBD biotransformations, the task of implementing that rule is equally challenging, involving as it may constructing a complicated SMARTS expression and numerous constraint checks. Care must be taken to ensure that the rule implemented is the rule proposed. The METEOR system includes a sophisticated, graphical, rule-generation engine. We plan to implement a similar system to allow content experts to propose rules or rule modifications in graphical form, as part of an integrated rule-implementation system.

One advantage the PPS has over systems such as META and METEOR is that the latter are proprietary. The PPS and its rule-base are part of the public Web. A rule contribution has already been received from Nagaraja Tejoprakash of the Centre of Relevance & Excellence in Agro & Industrial Biotechnology, Thapar Institute of Engineering & Technology, Patiala, India. The PPS is being developed with input from the worldwide scientific community.

## CONCLUSIONS

The microbial pathway prediction system can predict known or plausible biodegradation pathways for compounds the UM-BBD does not contain. Challenges include the implementation of transformation rules and incorporating functional group interactions. To minimize combinatorial explosion of pathways, heuristics from biodegradation experts and information on compound physicochemical properties are needed. Users are encouraged to submit additional biotransformation rules and comment on the system and its existing rules. In this way, the prediction system grows under the direction of the scientific community.

## ACKNOWLEDGMENT

This research was supported by the Office of Science (BER), U.S. Department of Energy, Grant No. DE-FG02-01ER63268. We thank Wenjun Kang, Sean Anderson, Philip Judson, John Carlis, Nagaraja Tejoprakash, Masanori Arita, and C. Doug Hershberger for helpful discussions.

**Supporting Information Available:** Fifty UM-BBD compounds, known biodegradation predicted (Table 1) and 34 UM-BBD compounds, known biodegradation not completely predicted (Table 2). This material is available free of charge via the Internet at http://pubs.acs.org.

## REFERENCES AND NOTES

(1) Wackett, L. P.; Hershberger, C. D. *Biocatalysis and Biodegradation*; ASM Press: Washington, DC, 2001; pp 157−170.
(2) Klopman, G.; Dimayuga, M.; Talafous, J. META I. A Program for the Evaluation of Metabolic Transformation of Chemicals. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1320−1325.
(3) Talafous, J.; Sayre, L. M.; Mieyal, J. J.; Klopman, G. META 2. A Dictionary Model of Mammalian Xenobiotic Metabolism. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1326−1333.
(4) Klopman, G.; Tu, M.; Fan, B. T. META 4. Prediction of the Metabolism of Polycyclic Aromatic Hydrocarbons. *Theor. Chem. Acc.* **1999**, *102*, 33−38.
(5) Langowski, J. J.; Long, A. Computer Systems for the Prediction of Xenobiotic Metabolism. *Adv. Drug Delivery Rev.* **2002**, *54*, 407−415.
(6) Greene, N. Knowledge-based Expert Systems for Toxicity and Metabolism Prediction. In *Drug metabolism*; Erhardt, P. W., Ed.; Blackwell Science Ltd.: London, 1999; pp 289−296.
(7) Greene, N.; Judson, P. N.; Langowski, J. J.; Marchant, C. A. Knowledge-based Expert Systems for Toxicity and Metabolism Prediction: DEREK, StAR, and METEOR. *SAR QSAR Environ. Res.* **1999**, *10*, 299−313.
(8) Judson, P. N.; Fox, J.; Krause, P. J. Using New Reasoning Technology in Chemical Information Systems. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 621−624.
(9) Judson, P. N. Rule Induction for Systems Predicting Biological Activity. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 148−153.

(10) Klopman, G.; Tu, M. Structure-biodegradability Study and Computer-automated Prediction of Aerobic Biodegradation of Chemicals. *Environ. Toxicol. Chem*. **1997**, *16*, 1829−1835.
(11) Klopman, G.; Zhang, Z.; Balthasar, D. M.; Rosencranz, H. S. Computer-automated Predictions of Aerobic Biodegradation Transforms in the Environment. *Environ. Toxicol. Chem*. **1995**, *14*, 395−403.
(12) Ellis, L. B. M.; Hou, B. K.; Wenjun, K.; Wackett, L. P. The University of Minnesota Biocatalysis/Biodegradation Database: Post-Genomic Data Mining. *Nucleic Acids Res*. **2003**, *31*, 262−265. http://umbbd.ahc.umn.edu/
(13) Sun Microsystems, Inc., 4150 Network Circle, Santa Clara, CA 95054. Java Server Pages (JSP), version 2.0, 2002. http://java.sun.com/products/jsp.
(14) Sun Microsystems, Inc., 4150 Network Circle, Santa Clara, CA 95054. Java Development Kit 1.3.1. http://java.sun.com/.
(15) Csizmadia, F. J Chem: Java Applets and Modules Supporting Chemical Database Handling from Web Browsers. *J. Chem. Inf. Comput. Sci*. **2000**, *40*, 323−324.
(16) ChemAxon Ltd., Máramaros köz 3/a, Budapest, 1037 Hungary. http://www.chemaxon.com.
(17) Weininger, D. SMILES: a Chemical Language for Information Systems. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci*. **1988**, *28*, 31−36.
(18) Weininger, D. SMILES 2: Algorithm for Generation of Unique SMILES Notation. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 97−101.
(19) James, C. A.; Weininger, D.; Delany, J. *Daylight Theory Manual*, *SMILES Theory;* 2000. http://www.daylight.com/dayhtml/doc/theory/theory.smiles.html.
(20) James, C. A.; Weininger, D.; Delany, J. *Daylight Theory Manual*, *SMARTS Theory;* 2000. http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html.
(21) Wegner, J. K. JOELib. Version Nov 16, 2002. http://joelib.sourceforge.net.
(22) MySQL version 3.23, MySQL Inc., 2510 Fairview Avenue East, Seattle WA 98102. http://www.mysql.com/.
(23) Webb, E. C. Enzyme Nomenclature: Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes; Academic Press: San Diego, CA, 1992. http://www.chem-.qmw.ac.uk/iubmb/enzyme/.
(24) Johnson, W. A. *Invitation to Organic Chemistry*; Jones and Bartlett: Boston, MA, 1999; p 194.
(25) Kanehisa, M.; Goto, S.; Kawashima, S.; Nakaya, A. The KEGG Databases at GenomeNet. *Nucleic Acids Res.* **2002**, *30*, 42−46. http://www.genome.ad.jp/kegg/.
(26) Zeng, Y. UM-BBD Pentachlorophenol Family Pathway. http://umbbd.ahc.umn.edu/pcp/pcp_map.html.
(27) Spormann, A.; Oh, D. J. UM-BBD Anaerobic Benzoate Pathway. http://umbbd.ahc.umn.edu/benz/benz_map.html.
(28) Klopman, G.; Tu, M.; Talafous, J. META 3. A Genetic Algorithm for Metabolic Transform Priorities Optimization. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 329−334.
(29) Long A. Rule-based Prioritisation of Metabolites. Some Recent Developments in METEOR. *Drug Metabolism Rev.* **2002**, *34* Supplement 1, 71.

CI034018F