J Ind Microbiol Biotechnol (2004) 31: 261–272
DOI 10.1007/s10295-004-0144-7

ORIGINAL PAPER

Bo Kyeng Hou · Lynda B. M. Ellis
Lawrence P. Wackett

Encoding microbial metabolic logic: predicting biodegradation

Received: 1 January 2004 / Accepted: 12 April 2004 / Published online: 10 July 2004

Abstract Prediction of microbial metabolism is important for annotating genome sequences and for understanding the fate of chemicals in the environment. A metabolic pathway prediction system (PPS) has been developed that is freely available on the world wide web (http://umbbd.ahc.umn.edu/predict/), recognizes the organic functional groups found in a compound, and predicts transformations based on metabolic rules. These rules are designed largely by examining reactions catalogued in the University of Minnesota Biocatalysis/Biodegradation Database (UM-BBD) and are generalized based on metabolic logic. The predictive accuracy of the PPS was tested: (1) using a 113-member set of compounds found in the database, (2) against a set of compounds whose metabolism was predicted by human experts, and (3) for consistency with experimental microbial growth studies. First, the system correctly predicted known metabolism for 111 of the 113 compounds containing C and H, O, N, S, P and/or halides that initiate existing pathways in the database, and also correctly predicted 410 of the 569 known pathway branches for these compounds. Second, computer predictions were compared to predictions by human experts for biodegradation of six compounds whose metabolism was not described in the literature. Third, the system predicted reactions liberating ammonia from three organonitrogen compounds, consistent with laboratory experiments showing that each compound served as the sole nitrogen source supporting microbial growth. The rule-based nature of the PPS makes it transparent, expandable, and adaptable.

Keywords Bacteria · Metabolism · Biodegradation · Prediction · Metabolic logic

L. P. Wackett (✉)
Department of Biochemistry, Molecular Biology and Biophysics and BioTechnology Institute, University of Minnesota,
1479 Gortner Avenue, St Paul, MN 55108, USA
E-mail: wackett@biosci.cbs.umn.edu
Tel.: +1-612-6253785
Fax: +1-612-6251700

L. B. M. Ellis
Department of Laboratory Medicine and Pathology, University of Minnesota,
Minneapolis, MN 55455, USA

B. K. Hou
Department of Laboratory Medicine and Pathology, University of Minnesota, St Paul, MN 55108, USA

Introduction

More than a century of studies of metabolism from many microbial species has revealed that extraordinary enzymatic diversity resides in microbes. Since microbes are exposed to thousands of natural products and a great number of synthetic chemicals, the number of potential substrates for microbial enzymes is enormous [25]. More than ten million chemical substances are known and hundreds more are synthesized each week. More than 65,000 chemical substances are in commerce and many are subject to microbial metabolism under environmental conditions. Microbial metabolic pathways for hundreds of commercial compounds are depicted on the University of Minnesota Biocatalysis/Biodegradation Database (UM-BBD, http://umbbd.ahc.umn.edu/) [8]. The UM-BBD developers, cognizant of the impracticality of covering all known microbial metabolism, prefer novel reaction types when further populating the database.

Novel microbial transformation reactions of the type found in the UM-BBD are largely responsible for the degradation of newly synthesized pesticides in those cases in which the metabolism is well-known. Unfortunately, the full range of potential substrates for microbial metabolism far outstrips the ability of scientists to study experimentally how each could be transformed. This poses an important problem for society. New pesticides and other products cannot be tested quickly enough to fully inform environmental regulatory agencies prior to use of each commercial compound.

Increasingly, this dilemma will require the use of computer-encoded metabolic logic for predicting the pathways by which compounds are transformed by microbial enzymes.

Pathway prediction systems (PPS) for xenobiotic metabolism have been recently reviewed [15]. The three most similar to the PPS (http://umbbd.ahc.umn.edu/predict/) described here, METEOR, MetabolExpert, and META, were all initially developed to predict mammalian detoxification reactions. At the present time, METEOR [10, 17] remains limited to mammalian metabolism; MetabolExpert [6] includes rules for photochemical transformations; and META [13, 15] includes rules for aerobic environmental biodegradation.

Given that the UM-BBD currently provides the most comprehensive set of microbial transformation reactions, a prediction system formulated from its data has the potential to provide a useful picture of a compound's possible environmental metabolic fate. Such a system could be interfaced with toxicological data to define the search for potential carcinogenic or toxic intermediates.

The overall logic of the PPS is illustrated in Fig. 1; the computational tools which allow that metabolic logic to be applied in the current work are described elsewhere [11]. The current paper details the logic behind the metabolic reaction rules that provide the underpinnings of the PPS, and assess the system's breadth and accuracy via computational and bench experiments.

Fig. 1 Pathway prediction system flow. The system depicts all the rule-derived metabolites for the user. The user may rerun the cycle with different predicted cycle metabolites (2, 3,..., *N*), extending a pathway with each cycle

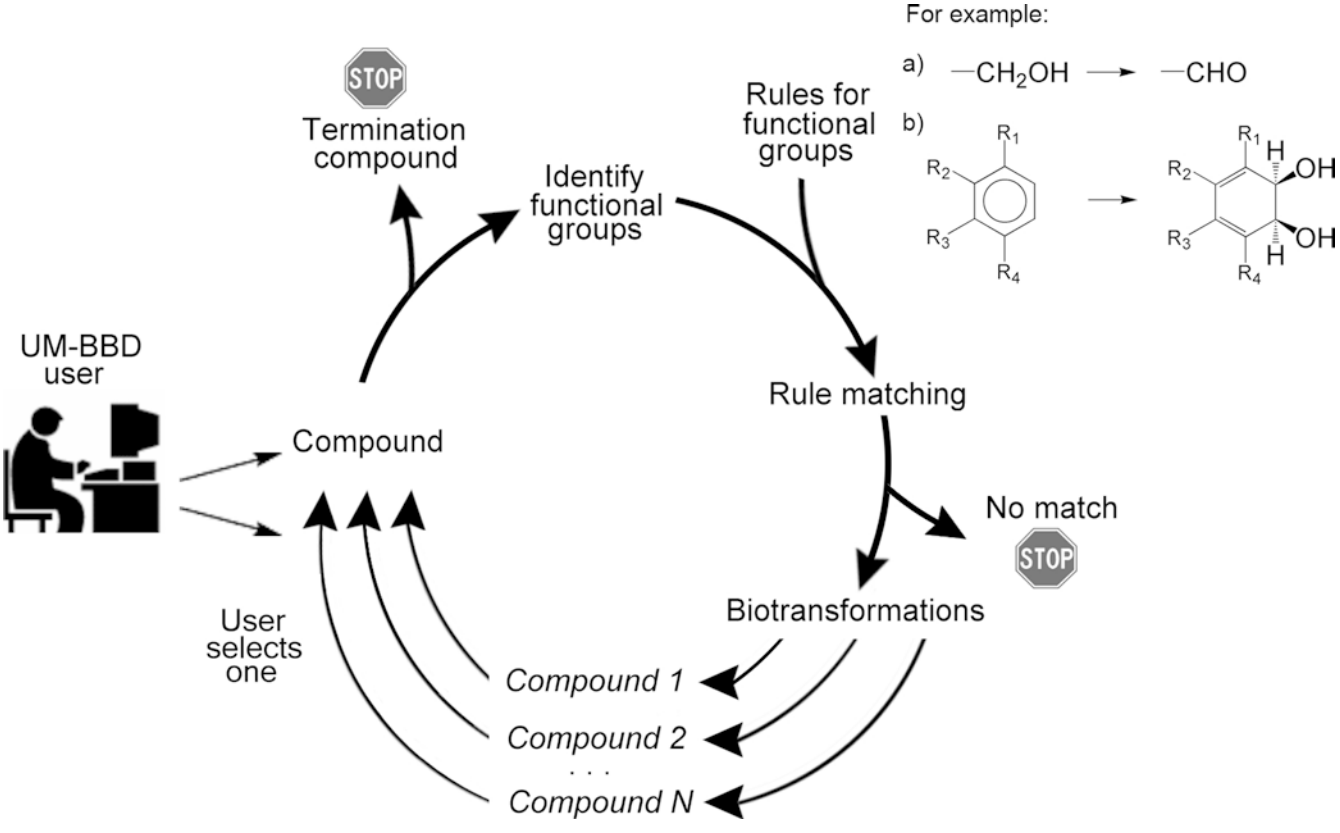


Fig. 2 The functional groups currently found on the UM-BBD that form the basis for rule-generation. The most current version of this list is on the web, http://umbbd.ahc.umn.edu/search/FG_image_map.html with links to UM-BBD compounds that exemplify each functional group

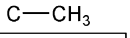
Methods

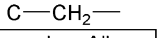
PPS development

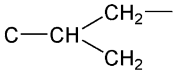
The PPS is based on rules describing the reactions of the chemical functional groups that define metabolism. The 50 organic functional groups presently found in the UM-BBD by manual examination (Fig. 2) contain, in addition to carbon, hydrogen, nitrogen, oxygen, phosphorus, sulfur, and halogen atoms, the metal or metalloid elements mercury, tin, arsenic and silicon. On average, each functional group undergoes several different metabolic transformations. Thus, 200–300 transformations should define all functional group metabolism currently found in the UM-BBD. The development of over 250 rules, the first major methodological accomplishment of the present effort, is described in Results.

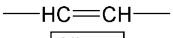
An additional method is rule generalization. Generalization is accomplished by determining, rule by rule, whether a rule describing a specific metabolic

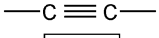

Methane


Primary Alkane


Secondary Alkane

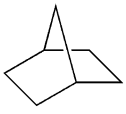

Tertiary Alkane

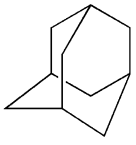

Alkene

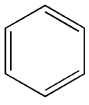

Alkyne

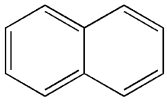

x = 1, 2, 3, 4

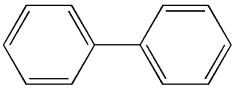
Cycloaliphatic ring

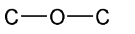

Bicycloaliphatic ring

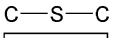

Tricycloaliphatic ring


Monocyclic aromatic hydocarbon


Polycyclic aromatic hydrocarbon

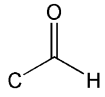

Biphenyl-type benzenoid ring

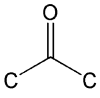

Oxygen ether


Thioether

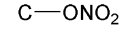

Alcohol

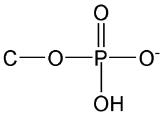

Thiol

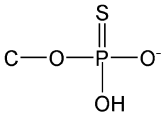

Aldehyde

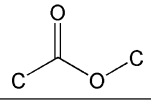

Ketone

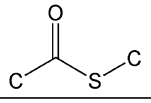

Thioketone

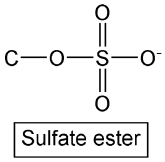

Nitrate ester

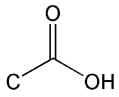

Organophosphate ester

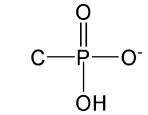

Thiophosphate ester

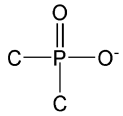

Carboxylic acid ester

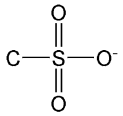

Carboxylic thioester

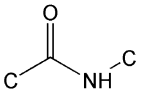

Sulfate ester

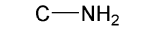

Carboxylic acid

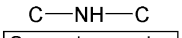

Phosphonic acid

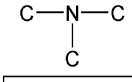

Phosphinic acid

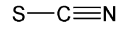

Sulfonic acid

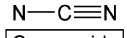

Amide

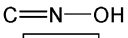

Primary amine

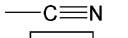

Secondary amine

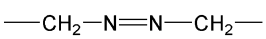

Tertiary amine

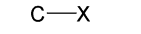

Thiocyanate

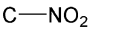

Cyanamide


Oxime


Nitrile


Diazo


Organohalide


Nitro

Organomercurial

Organoarsenical

Organotin

Organosilicon

S-heterocyclic ring

Saturated N-heterocyclic ring

Unsaturated N-heterocyclic ring

O-heterocyclic ring

Epoxide

Peroxide

transformation could be rewritten to cover a general class of compounds. The rule development process is illustrated with examples in the Results section. PPS biotransformations are carried out in two steps: a selection step that identifies a functional group to be transformed, followed by a transformation step that applies a virtual biotransformation to that functional group [11].

The rules are run by the PPS computer hardware, software and system architecture that has been previously described [11], with the following additions. We have added the ability to search the rulebase for the full or partial name of starting or ending organic functional group (e.g. aldehyde or carboxylate). Since the start of the project, rule web pages have included a list of hypertext links to all UM-BBD reaction web pages that exemplify the rule [11]; now reaction pages reciprocally list the rules with which they are in accord. These and other minor enhancements were implemented using Java 1.4.1 (Sun Microsystems, http://java.sun.com/).

Blocking unwanted biotransformations

It is necessary to block metabolic transformations that do not to occur in nature. For example, coenzyme A (CoA) moieties are blocked to prevent transformation of its functional groups. This blocking is carried out by first checking the Simplified Molecular Input Line Entry Specification (SMILES) string of the query compound to see whether it includes the SMILES string for CoA. If it does, the atom numbers assigned to the CoA in the query compound are recorded. In the rule matching process, if the atom numbers to be transformed by a matched rule form part of the CoA, the PPS ignores that transformation. The process can be easily extended if needed to block other chemical moieties.

Tests of PPS predictions

Since the rules were developed based on UM-BBD reactions, the PPS was tested to see whether it could reconstruct UM-BBD metabolism. As of 17 November 2003, the UM-BBD contained 113 compounds containing C and H, N, O, S, P and/or halides that initiated one or more pathways containing two or more well-characterized reactions (Table 1). The increasing ability of the PPS to predict the 569 known UM-BBD pathway branches for these compounds provided guidance for its development and assessed the breadth of PPS coverage.

To monitor PPS development, periodically it was automatically given each of these compounds. For each, the products of its first biotransformation based on PPS rules were compared to the products of that compound's reactions contained in the UM-BBD. Those that were found were subjected to a second round of PPS biotransformation. The cycle of biotransformation, comparison to known UM-BBD metabolism, and selection of known compounds to further transform, was continued until one of several endpoints was reached. The PPS could predict a compound's metabolism until no UM-BBD products were found; the UM-BBD pathway linked to a more general metabolism database (e.g. KEGG [12]), the UM-BBD compound was degraded to CO_2; or the next known reaction for that compound was a multistep reaction not characterized well enough for development of biotransformation rules. If any of those conditions were observed, that pathway branch met the test. If the PPS could not correctly transform a compound whose further metabolism was known to the UM-BBD in enough detail to model, then that pathway branch did not meet the test.

This was not the only test used, because PPS rules must predict the biodegradation of compounds not contained within the UM-BBD. In this context, PPS predictions were compared to human expert predictions for six compounds. Expert predictions of metabolic pathways for compounds whose biodegradation had not yet been studied were derived from the June 1998 PredictBT workshop [25], http://umbbd.ahc.umn.edu/workshop/. In that workshop, one of the authors (LW), predicted the biodegradation of two compounds, fentichlor and saccharine, to serve as examples for the other experts. Six experts, excluding the authors, then predicted the biodegradation of four additional compounds: barban, hydroxypivaldehyde, permethrin, and tetracyanoethene. The workshop summary includes the six predicted pathways and a rationale for each prediction (http://umbbd.ahc.umn.edu/workshop/summary.html).

Enrichment cultures

The PPS predicted the metabolism of three organonitrogen compounds: thionicotinamide, 4-hydrazinobenzoic acid, and thiohydantoin, and showed that ammonia would be released. To determine whether microbes could indeed liberate ammonia, enrichment cultures were conducted, using each organonitrogen compound as the sole source of nitrogen to support growth. Enrichment cultures were inoculated with soil obtained from outside Gortner Laboratory, University of Minnesota, St. Paul, Minn., USA, into minimal medium [24] containing each compound as the sole nitrogen source. Growth, measured as turbidity at 600 nm, continued to exceed 0.5 absorbance units (A_{600})after repeated transfers. Controls without any added nitrogen source showed turbidity less than 0.05 absorbance units.

Results

Rule development

Rules of metabolic transformation derive from chemical and biochemical logic. The functional group

Table 1 The 113-compound PPS known metabolism test set. The first 111 compounds had known metabolism correctly predicted by the PPS; the two on the last line (in italics) did not

Acetylene	Acrylonitrile	Adamantanone
2-Aminobenzenesulfonate	1-Aminocyclopropane-1-carboxylate	Atrazine
Benzonitrile	Benzyl sulfide	Biphenyl
Bromoxynil	(+)-Camphor	*epsilon*-Caprolactam
Carbazole	4-Carboxy-4′-sulfoazobenzene	Carbon disulfide
Carbon tetrachloride	2-Chlorobenzoate	4-Chlorobiphenyl
2-Chloro- *N*-isopropylacetanilide	Cyanamide	Cyclohexa-1,4-diene-1-carboxyl-CoA
Cyclohex-1-ene-1-carboxyl-CoA	Cyclohex-2,5-diene-1-carboxyl-CoA	Cyclohexanol
Cyclohexylsulfamate	Cyromazine	*p*-Cymene
Dibenzofuran	Dibenzo- *p*-dioxin	Dibenzothiophene
2,4-Dichloroaniline	2,4-Dichlorobenzoate	1,2-Dichloroethane
2,6-Dichlorophenol	2,4-Dichlorophenoxyacetic acid	1,3-Dichloro-2-propanol
Cis-1,3-Dichloropropene	*trans*-1,3-Dichloropropene	2,4-Dichlorotoluene
Dimethyl ether	Dimethyl sulfoxide	Dimethylphosphinic acid
Dodecyl sulfate	Ethylbenzene	Fluorene
Gallate	Glyphosate	*n*-Hexane
Hexahydro-1,3,5-trinitro-1,3,5-triazine	4-Hydroxyphthalate	Iprodione
Isoproturon	(+)-(*4R*)-Limonene	(*RS*)-Nicotine
2-Nitropropane	Methanesulfonic acid	L-Methionine
4-Methoxybenzoate	Methyl *tert*-butyl ether	Methyl fluoride
Trans-2-Methyl-5-isopropylhexa-2,5-dienal	1-Methylnaphthalene	2-Methylnaphthalene
3-Methylquinoline	Mordant yellow 3	Naphthalene
Naphthalene-1,6-disulfonate	Naphthalene-1-sulfonate	Naphthalene-2,6-disulfonate
Naphthalene-2-sulfonate	Nitrilotriacetate	Nitrobenzene
o-Nitrobenzoate	Nitroglycerin	*n*-Octane
Octyl hydroperoxide	Orcinol	Parathion
Pentachlorophenol	Pentaerythritol tetranitrate	Phenanthrene
Phenol	*Z*-Phenylacetaldoxime	3-Phenylpropionate
Alpha-Pinene	Propylene	Pyrene
Pyrrole-2-carboxylate	Resorcinol	Terephthalate
3,3′,5,5′-Tetrabromobisphenol A	1,2,3,4-Tetrachlorobenzene	Tetrachloroethene
Tetrahydrofuran	Thiocyanate	L-Threonine
Toluene	Toluene-4-sulfonate	1,2,3-Tribromopropane
1,1,1-Trichloro-2,2- *bis*-(4′-chlorophenyl)ethane (DDT)	1,1,1-Trichloroethane	2,4,6-Trichlorophenol
2,4,5-Trichlorophenoxyacetic acid	2,4,6-Trinitrotoluene	L-Tyrosine
Triethanolamine	Trifluoroacetate	Vanillyl alcohol
m-Xylene	*o*-Xylene	*p*-Xylene
beta-1,2,3,4,5,6-Hexachlorocyclohexane	*gamma-1,2,3,4,5,6-Hexachlorocyclohexane*	

transformations defined here are based on, in order of frequency: previously characterized enzyme-catalyzed reactions, reactions catalyzed by whole cells and assumed to be enzymatic, and chemical reactions that occur spontaneously and rapidly, usually going to completion in seconds or less. Over 250 biotransformation (bt) rules have been generated, encompassing functional groups containing the major UM-BBD elements: C, H, N, O, S, P, and the halogens.

Initially, rules were developed that covered the most common environmental biotransformations. As the system has moved to maturity, rules often cover more specific reactions and we seek improvement by making them more general; having them cover a broader range of compounds without generating reactions that would not likely occur in nature. Overly specific rules, strictly tailored to known UM-BBD metabolism, may score high in tests, but fail when challenged by the much greater variety of organic functional groups found in environmental compounds. For example, in the PPS version described in Hou et al. [11], rule bt0083 described the oxygenation of a methyl halide with subsequent spontaneous *gem*-elimination to yield formaldehyde. In the updated PPS version, rule bt0083 now describes the transformation of an alkyl halide to an aldehyde or ketone, encompassing the previous substrate and a large number of other alkyl halide substrates.

The earlier PPS, among its first seven rules, contained four that dealt with catechol (aromatic *vic*-dihydroxy) formation (Table 2). These rules restricted their application to a single aromatic ring, and related rules were written for naphthalenes, phenanthrenes, and other polyaromatic hydrocarbons known to be transformed by microbes in this fashion.

As an example of specificity for the initial dioxygenation, the specific rule for naphthalene and its derivatives, rule bt0038, creates *cis*-1,2-dihydroxydihydronaphthalene. The PPS predicts only this as a product of naphthalene dioxygenation; it is the only unsubstituted naphthalene *cis*-dihydrodiol observed in nature. While the PPS predicts many reactions that have not yet been observed in nature, there is, in this case, a logical chemical reason not to do so. The 2,3-dioxygenation of naphthalene would simultaneously remove aromaticity in both rings, while 1,2-dioxygenation only removes

Table 2 Evolution of PPS rules dealing with the upper catechol pathway (see text)

A. Original rules [11]
bt0004: 2,3-unsubstituted benzenoid → 2,3- *cis*-dihydroxydihydrobenzenoid
bt0005: 3,4-unsubstituted benzenoid → 3,4- *cis*-dihydroxydihydrobenzenoid
bt0006: 2,3- *cis*-dihydoxydihydrobenzenoid → 2,3-dihydroxybenzenoid
bt0007: 3,4- *cis*-dihydoxydihydrobenzenoid → 3,4-dihydroxybenzenoid
B. Current replacement rules
bt0004: 2,3-unsubstituted benzenoid → 2,3-cis-dihydroxydihydrobenzenoid
bt0005: 3,4-unsubstituted benzenoid → 3,4-cis-dihydroxydihydrobenzenoid
bt0255: dihydrodihydroxyaromatic → 1,2-dihydroxyaromatic (replaces bt0006 and bt0007)

aromaticity from one ring and is thus more favorable energetically.

Although the site(s), if any, for the initial dioxygenation are specific to a given ring system, once any dihydrodihydroxy derivative (single or polynuclear) forms, aerobic microbes plausibly dehydrogenate it to a *vic*-dihydroxy (catechol-like) aromatic compound. In this context, rules bt0006, bt0007, and others, were replaced by one new rule, bt0255 (Table 2), which virtually transforms dihydrodihydroxypolynuclear aromatics as shown in Fig. 3. We seek to generalize rules as much as possible, but they remain restrictive when metabolic or chemical logic dictates.

A further refinement of rules governing aromatic ring cleavage reactions is for polynuclear aromatic systems that undergo ring opening via reactions known as angular dioxygenation (Fig. 4). Angular dioxygenation reactions are described by rule: ''bt0196: polynuclear aromatic system → hemiacetal-like polynuclear aromatic system.'' Angular dioxygenation reactions occur with compounds such as carbazole, which contains a biphenyl ring connected by a nitrogen atom to form a central heterocyclic ring, or diphenylether that connects the rings only via an oxygen atom. Dioxygenation of the carbon atom bearing the heteroatom (O, N or S) and the adjacent carbon results in the formation of a *cis*-dihydrodiol that undergoes spontaneous ring opening to yield a biphenyl catechol. The PPS also includes systems with a connecting carbonyl group that, following angular dioxygenation, can undergo enzymatic ring opening [20].

Rule addition

New rules include those needed for anaerobic microbial catabolism and the catabolism of organophosphorous and organosulfur compounds. A description of anaerobic metabolism, for example in the metabolism of benzoate, requires rules for the addition and removal of CoA. This poses a challenge to biotransformation software, since the many functional groups in the CoA moiety are not to be metabolized; the PPS can now ignore them, as described in Methods. The addition of CoA is an anabolic (synthetic) reaction in the sense that it requires ATP and typically requires a ligase that catalyzes bond formation. Microbial anaerobic (and other) catabolism can require a small number of anabolic reactions; rules for anabolism are now incorporated into the PPS.

The PPS now includes both aerobic and anaerobic metabolic reactions and applies its rules indiscriminately to each compound it acts upon. The PPS can depict aerobic, anaerobic, or mixed pathways. As an example of mixed pathways, during the biodegradation of 4-chlorobenzoate by *Pseudomonas* sp. CBS3 and other aerobic bacteria, the carboxylic acid functional group is initially esterified with CoA, a reaction reminiscent of anaerobic benzoic acid metabolism [23]. However, in this case, after hydrolytic dechlorination and CoA removal, the ring is subjected to monooxygenase-catalyzed hydroxylation and oxygenative ring cleavage. Thus, reaction types typically found in either strictly anaerobic or strictly aerobic bacteria can be combined.

Pathway extension

Pathway prediction system users select the intermediate they wish to use for each step in pathway development (Fig. 1). This prevents the combinatorial explosion of pathways that could occur when many functional groups are present or if a large number of rules can be invoked for one or more functional groups. In those cases, the user would not benefit from a metabolic ''tree'' with a hundred or more branches, so user selectivity is desirable.

As an example of the permutations possible with even a relatively simple compound, consider biphenyl, which contains two phenyl rings connected by a carbon-carbon single bond. For the initial reaction, the PPS predicts 2,3- and 3,4-dioxygenation reactions, both of which are known in the literature, and monooxygenation at each of the three possible ring positions. The user would then pick one of those metabolites and proceed to further metabolism (Fig. 1). One of the compounds in a further selection is a catechol bonded by a single carbon-carbon bond to *cis*-2,3-dihydroxydihydrobenzene, a compound not commonly thought of as a biphenyl metabolite. As predicted though, *Alcaligenes eutrophus* strain H850 is able to transform 3,4-dihydroxybiphenyl to 3,4-dihydroxy- *cis*-2′, 3′-dihydrodihydroxybiphenyl (D.T. Gibson, University of Iowa , personal communication).

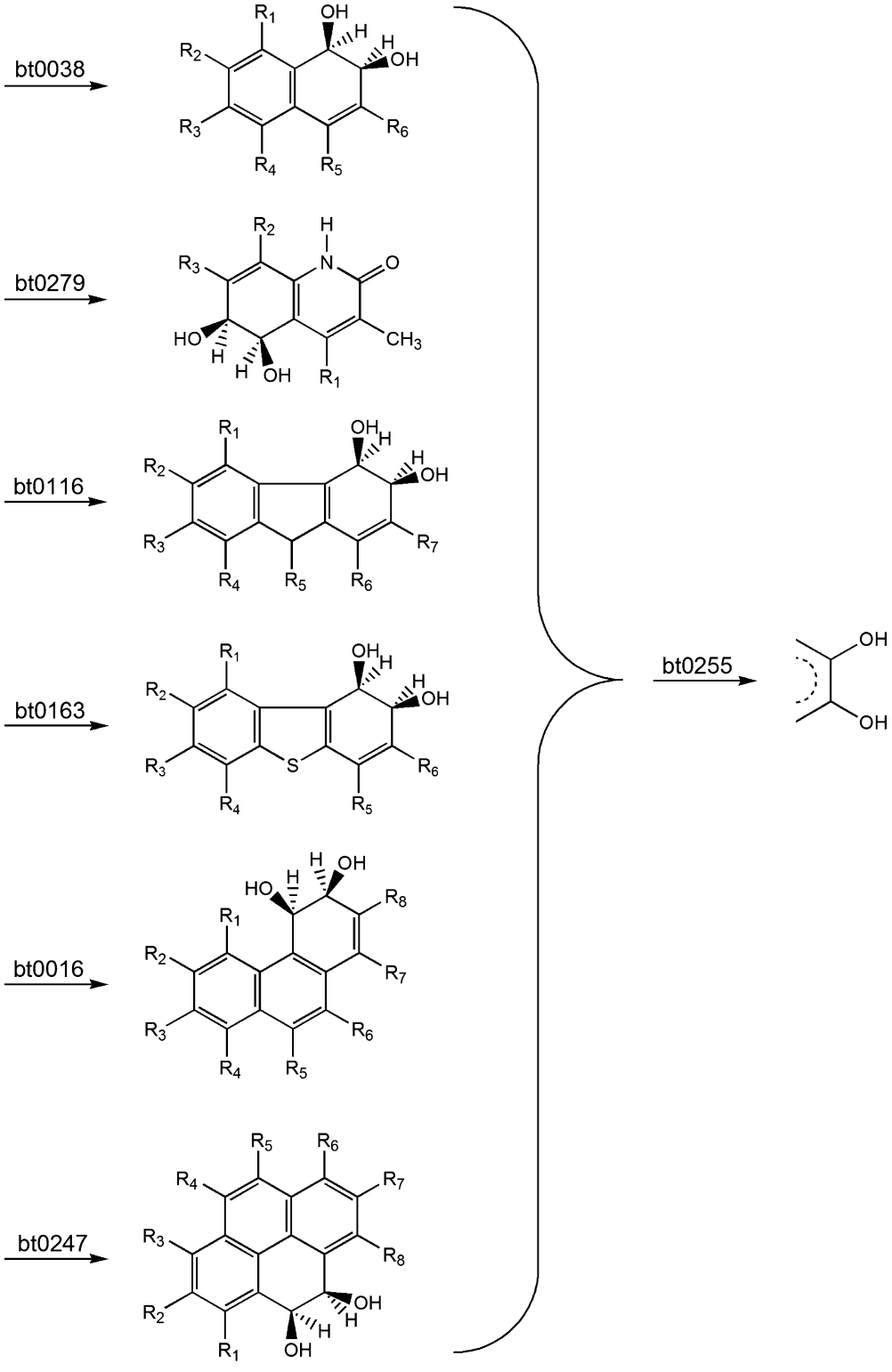


Fig. 3 Examples of six specific rules and one general rule. Six of the dihydrodihydroxy polynuclear aromatic ring systems created by specific rules. The rule that creates each one is to the left of the compound. All (and more) are virtually transformed to the corresponding aromatic dihydroxy compounds by the single, generalized, rule, bt0255. A list of all UM-BBD reactions in accord with this rule is available on the web: http://umbbd.ahc.umn.edu:8015/umbbd/rule.jsp?rule=bt0255

Comparison to known metabolism

As described in Methods, one PPS test set is the 113 UM-BBD compounds initiating pathways with two or more reactions and containing only the element C combined with H, O, N, S, P and/or halide atoms (Table 1). The UM-BBD contains information on 569 biodegradation pathway branches for these 113 compounds. The PPS duplicates at least one known biodegradation pathway for 111 of the 113 compounds, and can predict 410 (72%) of the 569 branches. The only two compounds for which it cannot yet predict at least one

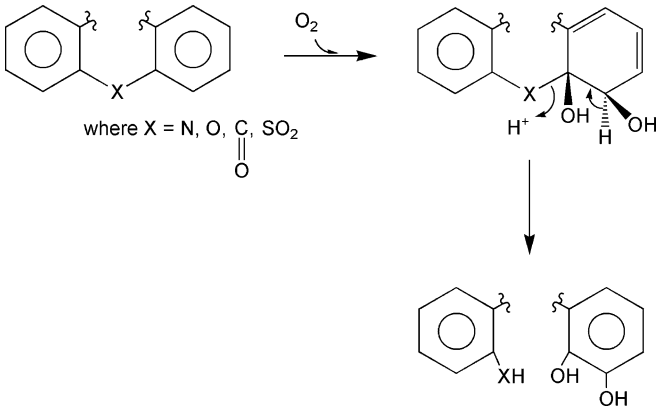


Fig. 4 The angular dioxygenation reactions forming the basis for transformation of a large number of heterocyclic ring compounds in nature. The break in the depicted structures represents either H, other functional groups, or atoms bridging the rings

known biodegradation pathway are β-1,2,3,4,5,6-hexachlorocyclohexane and γ-1,2,3,4,5,6-hexachlorocyclohexane.

Comparison to human expert predictions

As described in Methods, experts predicted biodegradation pathways for six compounds whose metabolism was not covered in the UM-BBB or other databases and for which biodegradation pathway information was not available in the open scientific literature [25]: fentichlor, saccharine, barban, hydroxypivaldehyde, permethrin, and tetracyanoethene. The current PPS can duplicate at least one expert-predicted pathway for all of these except fentichlor, a compound composed of two chlorophenolic rings linked via a thioether bridge. In one predicted pathway for fentichlor, an expert predicted the formation of two molecules of 4-chlorocatechol; the PPS instead predicted 4-chlorophenol and 1,2,3-trihydroxy-4-chlorobenzene, which is considered here to be more plausible.

The PPS also does not predict one of the reactions predicted by human experts for permethrin, an insecticide that contains two phenyl rings bridged by an oxygen ether. In agreement with two human experts, the PPS predicts that the ether group in permethrin will undergo angular dioxygenation leading to ether cleavage, via biochemical reactions shown in Fig. 4. However, two other experts instead suggested hydrolysis of the ether linkage, which was not predicted by the PPS. After evaluating the latter prediction, we now find that known ether hydrolases (EC 3.3.2) only act on epoxides (strained ethers) or enol ethers (products stabilized by enol/keto tautomerization) [29], http://www.chem.qmw.ac.uk/iubmb/enzyme/. Thus, by sticking to known biochemical rules, it is likely that the PPS is correct, and we are unaware of any biochemical precedent for this reaction.

However, this comparison to expert predictions also reveals some weaknesses of the present PPS. Some compounds are predicted by the PPS to undergo additional, implausible, reactions, such as the hydrolytic displacement of a vinylic cyano group (tetracyanoethylene); or contain such a wealth of different functional groups that the permutations are daunting (e.g. barban).

Prediction consistency with laboratory experiments

Another measure of the effectiveness of the PPS comes from a comparison of computational predictions with experimental observations. A computational prediction takes minutes; pathway elucidation by enzyme and metabolite isolation sometimes takes years or decades. However, data from readily conducted growth experiments can be compared with PPS predictions to determine whether they are consistent.

For this comparison, we used novel chemical compounds for which biodegradation information is currently not available. Bacteria were isolated on thionicotinamide, thiohydantoin, and 4-hydrazinobenzoic acid as described in the Methods section; each compound was used as the sole source of nitrogen and supported growth. The PPS predicts a single reaction for thionicotinamide, a hydrolytic cleavage of the thioamide to yield thionicotinic acid with the liberation of ammonium ion (Fig. 5a). This prediction is completely consistent with the growth of the bacteria on this compound as the sole source of nitrogen. Similarly, the PPS predicts that one molecule of thiohydantoin undergoes a hydrolytic ring opening reaction followed by other plausible reactions that overall liberate two ammonium ions to support growth (Fig. 5b).

The PPS prediction starting with 4-hydrazinobenzoic acid was more surprising, yet still plausible. We anticipated that 4-hydrazinobenzoic acid might undergo reductive cleavage of the hydrazine functional group to yield ammonium ion and *p*-aminobenzoic acid. Instead the PPS predicted decarboxylation followed by hydroxylation of a nitrogen atom to yield ammonium ion and hydroxylaminobenzene, an unexpected but completely plausible pathway for liberating ammonia to support growth (Fig. 5c). Five years ago, hydroxylaminobenzene was discovered to be an intermediate in nitrobenzene metabolism [22], lending credence to this PPS prediction.

Discussion

Prediction utility

Metabolic pathway prediction is an important approach for studying environmental chemical fate. Predicting metabolic pathways allows the the environmental and toxicological consequences of metabolic intermediates to be evaluated. In those cases in which an intermediate

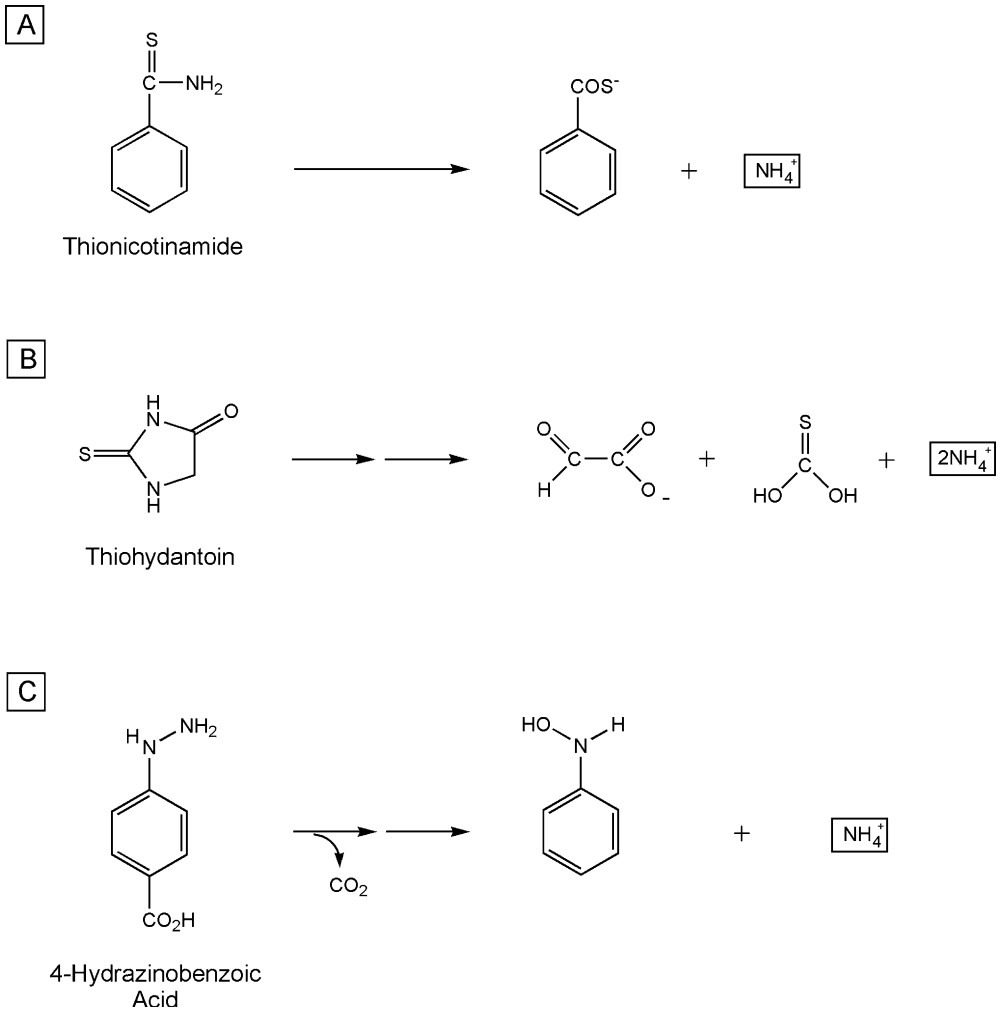


Fig. 5 PPS predictions for compounds used in microbial growth experiments. Each compound provided the only nitrogen source for microbial growth (see text). The PPS predicted reactions generating ammonia and plausible organic intermediates (shown) for: **a** thioamide, **b** N-heterocyclic ring compound, and **c** organohydrazine, respectively

could be more toxic to mammals than the starting compound, it would then become very important to monitor polluted environments for the accumulation of the toxic intermediate. Some examples in which accurate pathway prediction would have revealed environmental problems are the reductive dechlorination of perchloroethylene to generate carcinogenic vinyl chloride [18], and the reduction of nitro to mutagenic nitroso compounds [2].

The PPS is focused on the metabolism of the vast population of known microbes rather than metabolism by any individual microbe. This could be either a strength or a weakness depending on the interests of individual investigators. Classically, microbiologists have focused either on individual microbes or biodegradation of compounds by mixed natural populations [1]. There is increasing evidence that the catabolism of anthropogenic chemicals in soils is often consortial; that is, it is the sequential action of different microorganisms acting collectively [5, 7]. The catabolic genes and enzymes present in consortial metabolic assemblages can move into disparate bacterial genera via horizontal gene transfer. Horizontal gene transfer occurs via catabolic plasmids and transposable elements [31] that spread the genes and enzymes [27] that form the basis of the PPS. In this context, a metabolic predictive system that overlooks taxonomic lines may best serve scientists and regulators.

Other predictive methods

A prediction system generating metabolic pathways necessarily encodes extensive knowledge of many biodegradation reactions, rather than relying on some small number of parameters that might lead to oversimplification and hence greater inaccuracy. For example, one concept guiding biodegradability prediction has been the hydrophobicity of a compound [14, 19]. Highly hydrophobic compounds are presumed to be less biodegradable. This has less to do with the potential for metabolism than it does with bioavailability of the compound in a complex environmental matrix in which hydrophobic materials could sequester the compound away from the cells, preventing metabolism. For example, toluene is much more hydrophobic than phytic acid (inositol hexaphosphate) but is significantly more biodegradable. The UM-BBD shows at least six distinct

metabolic pathways and many bacterial species active in the catabolism of toluene, consistent with the idea that its metabolism is widespread. The idea that biodegradability is best predicted by anticipating which known enzymes might work on starting substrates and metabolites was also advanced by Rorije et al. [21].

The output of some previous computational methods has been a "biodegradability index," based on the concept of "xenophores," or metabolically unfavorable functional groups that purportedly inhibit biodegradation, such as halogen substituents or nitro groups [1]. However, biodegradability prediction based on xenophores is flawed because it ignores known microbial metabolic pathways. For example, in the xenophore model, a methyl group is considered favorable to biodegradation whereas a chlorine substituent is unfavorable. Yet there are many literature reports of hexachlorobenzene biodegradation [4, 9, 30] and none, to our knowledge, of that for hexamethylbenzene or pentamethylbenezene.

Pathway prediction system metabolic prediction is consistent with this assessment. The PPS predicts that hexamethylbenzene would require many metabolic steps to generate even a single intermediate for which microbial metabolism is known. This is borne out in experiments with 1,2,4-trimethylbenezene [3]. PPS predictions impart importance to the number of reactions required to metabolize a given substance, given that carrying many more enzymes and genes is probably constrained against in microbial evolution. The xenophore model does not take these important ideas, which are emerging from genomics studies, into account.

Prediction tests

Though we report success with the known metabolism test, correctly predicting 410 pathway branches for 111 starting compounds (Table 1), this is a conservative number. In the test process, the PPS correctly predicts biodegradation of 1,474 compounds, the starting compounds plus 1,364 intermediate compounds in pathway branches.

Since the UM-BBD was used as the basis for rules establishment, testing it using UM-BBD data does not test its ability to predict plausible biodegradation of compounds not found in the UM-BBD. To address this issue, the PPS was also tested on six compounds whose biodegradation is not presently known, and its predictions were compared with those of human experts [26]. PPS predictions agreed with one or more human experts for almost all reactions in at least one pathway for biodegradation of six compounds. For two reactions where the PPS and experts disagreed, the PPS prediction was considered more plausible.

For these compounds, the PPS is today no better or worse than human experts. It did not agree with all experts. Conversely, the PPS predicted some reactions not considered by other experts, most of which were plausible. It sometimes presented a much larger number of plausible reactions than did human experts. Users may prefer such human filtering and prioritization, though it may omit plausible reactions. The PPS needs further refinement to prune implausible reactions and offer "expert-like" guidance when numerous plausible reactions are generated.

Additionally, PPS predicted biodegradation of organonitrogen compounds that served as the nitrogen source for microorganisms obtained by enrichment culture. In these examples, plausible metabolic pathways were predicted to liberate nitrogen as ammonium ion that could support growth. Similarly, microbial growth on other organo-nitrogen, -sulfur or -phosphorous compounds could be readily tested to see whether bench results are consistent with PPS-predicted metabolism.

Current development

When the number of biotransformation rules was small, it was easy to browse the list of rules and, from their names, such as "bt0003: aldehyde $\rightarrow$ carboxylate", know exactly what chemistry they covered. Now that the number of rules exceeds 250, browsing is of less use. We have added the ability to search on rule names, and thus a search for "aldehyde" will return bt0003 and most other rules that act on or produce aldehydes.

Even so, some rules on that list, such as "bt0063: secondary or tertiary Amine $\rightarrow$ primary or secondary Amine + Aldehyde or Ketone" are more difficult to understand. A comment was added on the bt0063 rule web page: "Oxidative removal of an R group from a secondary or tertiary amine. Secondary amine $\rightarrow$ Monoamine + Aldehyde or Ketone. Tertiary amine $\rightarrow$ Secondary amine + Aldehyde or Ketone. An aldehyde is produced if the leaving R group is attached through a primary carbon. A ketone is produced if the leaving R group is attached through a secondary carbon." However, it is not possible to search on text in the comment field. Such a rule would more accurately be given more than one name. It includes at least both "secondary Amine $\rightarrow$ Monoamine + Aldehyde or Ketone" and "tertiary Amine $\rightarrow$ secondary Amine + Aldehyde or Ketone." We are adding abilities to have multiple names (descriptors) for rules, to display all of them in a browsable list, and to search on them.

PPS expansion

The PPS might expand in the future in many different, non-exclusive ways. It cannot yet predict at least one known biodegradation pathway for two of the 113 compounds in the known metabolic test set. These two compounds, β-1,2,3,4,5,6-hexachlorocyclohexane and γ-1,2,3,4,5,6-hexachlorocyclohexane, are geometric isomers that cannot be distinguished by the present PPS. We are working to overcome this limitation. Also, a

focus of future work will be to better prune away less likely metabolism, leaving only the most plausible pathways, and offer guidance when numerous plausible transformations are predicted.

While rule generalization is necessary as described under Results, some rules may be overgeneralized. For example, the rule for benzenoid dioxygenation (Fig. 1) allows the ring to have up to four substituents. Doubtless, some substituent combinations will preclude enzymatic dioxygenation; however, guidelines for restricting such substituents and improving other generalizations are as yet unknown. We invite input from the scientific community to the authors to improve this and other aspects of the PPS.

Future developments might allow the merging of biodegradation-predictive and toxicity-predictive software to anticipate potential health consequences for new industrial chemicals. This would be particularly useful for screening pesticides prior to their environmental release, in order to predict problems associated with a compound's widespread use. The METEOR system discussed in the Introduction, although limited to predicting mammalian metabolism, interfaces with the DEREK toxicity prediction system [16]. The free, web-based PPS has its toxicity counterpart in the EPA's persistent, bioaccumulative, and toxic (PBT) Profiler [28], http://www.pbtprofiler.net/. The two systems use a similar format for compound entry. It is possible, with small modifications of both systems, that a user query entered in the PBT Profiler could simultaneously be sent to the PPS, or, as the user enters or selects each compound in a predicted PPS pathway, the compound could simultaneously be sent to the PBT Profiler.

The PPS might also provide the initial basis to construct a plausible system depicting biocatalytic pathways for the synthesis of biotechnologically important compounds. The user would enter into the PPS-derived system both a starting and a desired final compound, and have the system determine whether the final compound can be produced from the starting compound in a reasonable number of steps. If a reasonable pathway were found, the PPS would then display the intervening reactions; each reaction would link to a biotransformation rule; each rule would link to a web page for each known UM-BBD reaction that exemplifies that rule. From a UM-BBD reaction page, the user could determine the availability of enzymes and their genes that could be used to construct such a pathway. This would require additional development of the present PPS, in both the rule base and computer system.

Pathway prediction system rules are currently based on 50 functional groups. More will be added as additional metabolism is discovered. The rules can be accessed throughout the prediction process and links are given to UM-BBD reactions that exemplify the rules. The rule-based nature of the PPS makes it transparent, expandable, and adaptable.

Acknowledgements This research was supported by the Office of Science (BER), US Department of Energy, grant no. DE-FG02-01ER63268 and a grant from LHASA Ltd, Leeds, UK. We thank Sean Anderson for development of part of the PPS system and authoring several biotransformation rules, including those that generalized formation of catechol-like compounds. We thank Ana Negrete for carrying out microbial enrichment cultures. We thank Jack Richman, Dave Roe, Philip Judson, Anthony Long, and Jeff Osborne for helpful discussions.

References

1. Alexander M (1994) Effect of chemical structure on biodegradation. In: Biodegradation and bioremediation. Academic Press, San Diego, pp 159–176
2. Beller HR (2002) Anaerobic biotransformation of RDX (hexahydro-1,3,5-trinitro-1,3,5-triazine) by aquifer bacteria using hydrogen as the sole electron donor. Water Res 36:2533–2540
3. Bestetti G, Galli E (1987) Characterization of a novel TOL-like plasmid from *Pseudomonas putida* involved in 1,2,4-trimethylbenzene degradation. J Bacteriol 169:1780–1783
4. Chen X, Christopher A, Jones JP, Bell SG, Guo Q, Xu F, Rao Z, Wong LL (2002) Crystal structure of the F87 W/Y96F/V247L mutant of cytochrome P-450$_{CAM}$ with 1,3,5-trichlorobenzene bound and further protein engineering for the oxidation of pentachlorobenzene and hexachlorobenzene. J Biol Chem 277:37519–37526
5. Dangmann E, Stolz A, Kuhm AE, Hammer A, Feigel B, Noisommit-Rizzi N, Rizzi M, Reuss MM, Knackmuss HJ (1996) Degradation of 4-aminobenzene-sulfonate by a two-species bacterial coculture. Physiological interactions between *Hydrogenophaga palleronii* S1 and *Agrobacterium radiobacter* S2. Biodegradation 7:223–229
6. Darvas F (1988) Predicting metabolic pathways by logic programming. J Mol Graph 6:80–85
7. DeSouza ML, Newcombe D, Alvey S, Crowley DE, Hay A, Sadowsky MJ, Wackett LP (1998) Molecular basis of a bacterial consortium: interspecies catabolism of atrazine. Appl Environ Microbiol 64:178–184
8. Ellis LBM, Hou BK, Wenjun K, Wackett LP (2003) The University of Minnesota Biocatalysis/Biodegradation Database: post-genomic data mining. Nucleic Acids Res 31:262–265
9. Fathepure BZ, Tiedje JM, Boyd SA (1988) Reductive dechlorination of hexachlorobenzene to tri- and dichlorobenzenes in anaerobic sewage sludge. Appl Environ Microbiol 54:327–330
10. Greene N (1999) Knowledge based expert systems for toxicity and metabolism prediction. In: Erhardt PW (ed) Drug metabolism. Blackwell, London, pp 289–292
11. Hou BK, Wackett LP, Ellis LBM (2003) Microbial pathway prediction: A functional group approach. J Chem Inf Comput Sci 43:1051–1057
12. Kanehisa M, Goto S, Kawashima S, Nakaya A (2002) The KEGG databases at GenomeNet. Nucleic Acids Res 30:42–46
13. Klopman G, Tu M (1997) Structure-biodegradability study and computer-automated prediction of aerobic biodegradation of chemicals. Environ Toxicol Chem 16:1829–1835
14. Klopman G, Wang S, Balthasar DM (1992) Estimation of aqueous solubility of organic molecules by the group contribution approach. Application to the study of biodegradation. J Chem Inf Comp Sci 32:474–482
15. Klopman G, Zhang Z, Balthasar DM, Rosenkranz HS (1995) Computer automated predictions of aerobic biodegradation transforms in the environment. Environ Toxicol Chem 14:395–403
16. Langowski JJ, Long, A (2002) Computer systems for the prediction of xenobiotic metabolism. Adv Drug Deliv Rev 54:407–415

17. Long A (2002) Rule based prioritisation of metabolites. Some recent developments in METEOR. Drug Metab Rev 34(Suppl 1):71
18. Maymo-Gatell X, Anguish T, Zinder SH (1999) Reductive dechlorination of chlorinated ethenes and 1,2-dichloroethane by "*Dehalococcoides ethenogenes*" 195. Appl Environ Microbiol 65:3108–3113
19. Meylan WM, Howard PH (1995) Atom/fragment contribution method for estimating octanol-water partition coefficients. J Pharm Sci 84:83–92
20. Nojiri H, Habe H, Ornori T (2001) Bacterial degradation of aromatic compounds via angular dioxygenation. J Gen Appl Microbiol 47:279–305
21. Rorije E, Germa F, Philipp B, Schink B, Beimborn DB (2002) Prediction of biodegradability from structure: Imidazoles. SAR QSAR Environ Res 13:199–204
22. Schenzle A, Lenke H, Spain JC, Knackmuss HJ (1999) 3-Hydroxylaminophenol mutase from *Ralstonia eutropha* JMP134 catalyzes a Bamberger rearrangement. J Bacteriol 181:1444–1450
23. Scholten JD, Chang KH, Babbitt PC, Charest H, Sylvestre M, Dunaway-Mariano D (1991) Novel enzymic hydrolytic dehalogenation of a chlorinated aromatic. Science 253:182–185
24. Stanier RY, Palleroni NJ, Duodoroff M (1966) The aerobic pseudomonads: a taxonomic study. J Gen Microbiol 43:159–271
25. Wackett LP, Ellis LBM, Speedie SM, Hershberger CD, Knackmuss H-J, Spormann AM, Walsh CT, Forney LJ, Punch WF, Kazic T, Kanehisa M, Berndt DJ (1999) Predicting microbial biodegradation pathways. ASM News 65:87–93
26. Wackett LP, Hershberger CD (2001) Predicting microbial biocatalysis and biodegradation. In: Biocatalysis and biodegradation. ASM Press, Washington DC, pp 157–170
27. Wackett LP, Sadowsky MJ, Martinez B, Shapir N (2002) Biodegradation of atrazine and related triazine compounds: from enzymes to field studies. Appl Microbiol Biotechnol 58:39–45
28. Walker JD, Carlsen L (2002) QSARs for identifying and prioritizing substances with persistence and bioconcentration potential. SAR QSAR Environ Res 13:713–725
29. Webb EC (1992) Enzyme nomenclature: recommendations of the nomenclature committee of the international union of biochemistry and molecular biology on the nomenclature and classification of enzymes. Academic, San Diego
30. Wu Q, Milliken CE, Meier GP, Watts JE, Sowers KR, May HD (2002) Dechlorination of chlorobenzenes by a culture containing bacterium DF-1, a PCB dechlorinating microorganism. Environ Sci Technol 36:3290–3294
31. Wyndham RC, Cashore AE, Nakatsu CH, Peel MC (1994) Catabolic transposons. Biodegradation 5:323–342