frontiers | Frontiers in Environmental Science

TYPE Mini Review
PUBLISHED 26 August 2022
DOI 10.3389/fenvs.2022.980284

OPEN ACCESS

EDITED BY
Joginder Singh,
Lovely Professional University, India

REVIEWED BY
Vineet Kumar,
National Environmental Engineering Research Institute (CSIR), India
Shweta Jaiswal,
Maharshi Dayanand University, India
Hanuman Singh Jatav,
Sri Karan Narendra Agriculture University, India

*CORRESPONDENCE
Pankaj Kumar Arora,
arora484@gmail.com

SPECIALTY SECTION
This article was submitted to Toxicology, Pollution and the Environment,
a section of the journal
Frontiers in Environmental Science

RECEIVED 29 June 2022
ACCEPTED 05 August 2022
PUBLISHED 26 August 2022

CITATION
Arora PK, Kumar A, Srivastava A, Garg SK and Singh VP (2022), Current bioinformatics tools for biodegradation of xenobiotic compounds.
Front. Environ. Sci. 10:980284.
doi: 10.3389/fenvs.2022.980284

Current bioinformatics tools for biodegradation of xenobiotic compounds

Pankaj Kumar Arora[1]*, Ashish Kumar[1], Alok Srivastava[2], Sanjay Kumar Garg[2] and Vijay Pal Singh[2]

[1]Department of Environmental Microbiology, Babasaheb Bhimrao Ambedkar University, Lucknow, India, [2]Department of Plant Science, Faculty of Applied Sciences, MJP Rohilkhand University, Bareilly, India

The review summarizes different bioinformatics tools used in studying the degradation of various xenobiotic compounds. Databases and Pathway Prediction Systems are the key bioinformatics tools involved in biodegradation. Several bio-degradative databases including EAWAG-BBD (Biocatalysis/Biodegradation Database), Plastics Microbial Biodegradation Database, ONDB (Organonitrogen Degradation Database), Food risk component database (Food risk component database, Aromadeg (Aromatic hydrocarbon Degrading Database), OxDBase (A database of Biodegradative oxygenases), and RHObase database (ring-hydroxylating oxygenase database) have been developed for biodegradation and bioremediation studies. Users can use pathway prediction systems to predict degradation of xenobiotic whose degradation has never been reported in the literature. This review will help to design a strategy for biodegradation of chemicals therefore, also help in improved bioremediation process of chemicals.

KEYWORDS

biodegradation, bioinformatics, database, pathway prediction system, EAWAG-BBD, scopus

1 Introduction

Biodegradation and Bioinformatics are two fundamental areas of environmental sciences and biotechnology. Biodegradation is a process of microbes-mediated breakdown of various toxic chemicals and other xenobiotic compounds (Arora et al., 2012). Microbes degrade toxic chemicals through the process of mineralization or co-metabolism. During the mineralization process, microbes use toxic molecules as their food materials to gain carbon, and energy and completely utilized them via enzymatic reactions. However, co-metabolism is a process by which biochemical reactions convert toxic compounds into less toxic ones. Figure 1 illustrates the bacterial mineralization and cometabolism of 4-nitrophenol. Mineralization of 4-nitrophenol results in carbon dioxide and water, which are utilized by bacteria to gain energy. In contrast, 4-nitrophenol cometabolized into 4-aminophenol, which was not utilized by bacteria. A toxic chemical can only be

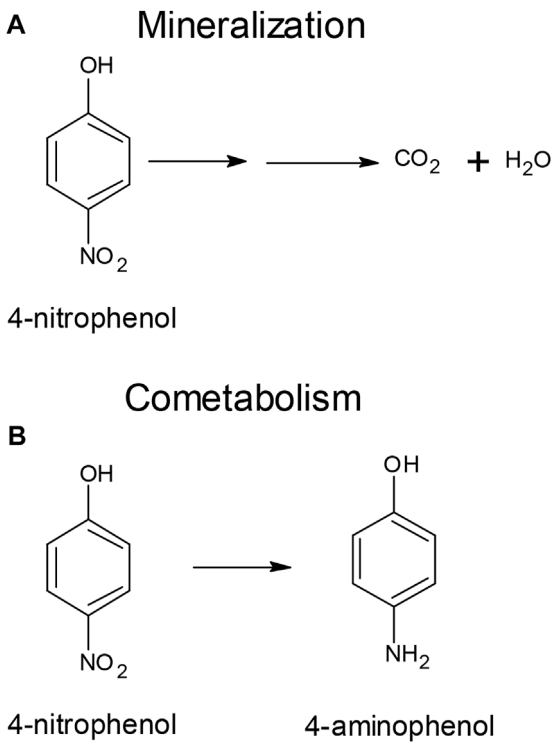

FIGURE 1
Bacterial mineralization **(A)** and cometabolism **(B)** of 4-nitrophenol.

degraded by a specific microbe that is responsive to the chemical's structure and the enzymes present in that microbe to tackle that specific chemical.

Microbial remediation has emerged as a technology aimed at removing hazardous materials from our soil and water. Many microbes have been identified that have ability for the removal of xenobiotic substances from the environment. Several microbial enzymes including hydrolases, dehalogenases, dioxygenases and monooxygenase are directly involved in biodegradation process. Bioremediation can now be enhanced by cloning the genes encoding these enzymes into bacteria.

Bioremediation can be improved by having knowledge about various xenobiotic compounds and their biodegradation (Arora and Bae, 2014). Due to the integration of bioinformatics and biodegradation, many bioinformatic tools have been developed to determine the fate of xenobiotic compounds in environment as well as to understand their degradation mechanism (Arora et al., 2009; Ellis et al., 2006: Gan and Zhang, 2019) Databases and Pathway Prediction tools are more prominent examples of these tools. An overview of various databases and pathways prediction tools can be found in Figure 2. This article summarizes the various databases and pathway prediction tools used in biodegradation studies.

2 Databases

Literature databases, Chemical databases, Biodegradative Databases are examples of these tools. Literature databases contain a collection of research publications and review articles related to various fields of sciences. SCOPUS and PubMed are the two literature databases. Users can search papers related to biodegradation and bioremediation using these databases.

Chemical databases store information about structure, toxicity, physical and environmental properties of chemicals. PubChem is the largest database of the world that offers free access 8to chemical research information. Chemicals can be searched by name, structure, molecular formula, and other criteria. Information on chemical and physical properties, biological activities, safety and toxicity, literature citations and patents can be found here. Other environmentally important chemical databases are summarizing in Table 1

2.1 Biodegradative databases

Biodegradative databases are those databases, which store the information primarily on biodegradation of various xenobiotic compounds. Examples are EAWAG-BBD (Biocatalysis/ Biodegradation Database), PMBD (Plastics Microbial Biodegradation Database), ONDB (Organonitrogen Degradation Database), FRCD (Food risk component database), Aromadeg (Aromatic hydrocarbon Degrading Database), OxDBase (A database of Biodegradative oxygenases), RHObase database (ring-hydroxylating oxygenase database). Table 2 provides a list of bio-degradative databases.

EAWAG-BBD was the first well-maintained database dedicated to biodegradation and bioremediation, initially maintained by the University of Minnesota and currently is running by the EAWAG. Using this database, users can get information of metabolic degradation pathways of a wide range of xenobiotic compounds (Ellis et al., 2006). Various fields of interests including microbes (543), bio-transforming rules (250), genes, proteins 993) metabolic pathways 219) and chemical reactions (1503) that contribute to the decomposition of xenobiotic compounds are included in this database.

Another database, OxDBase, was developed by Chandigarh, India's CSIR-Institute of Microbial Technology and freely available at http://www.imtech.res.in/raghava/oxdbase. OxDBase is an information-storage database that provides details of monooxygenases and dioxygenases involved in the aerobic degradation of various aromatic compounds (Arora et al., 2009).

PMBD (http://pmbd.genome-mining.cn/home) is dedicated to microbial degradation of plastics and contained information about 949 microorganisms-plastic degradation relationships and

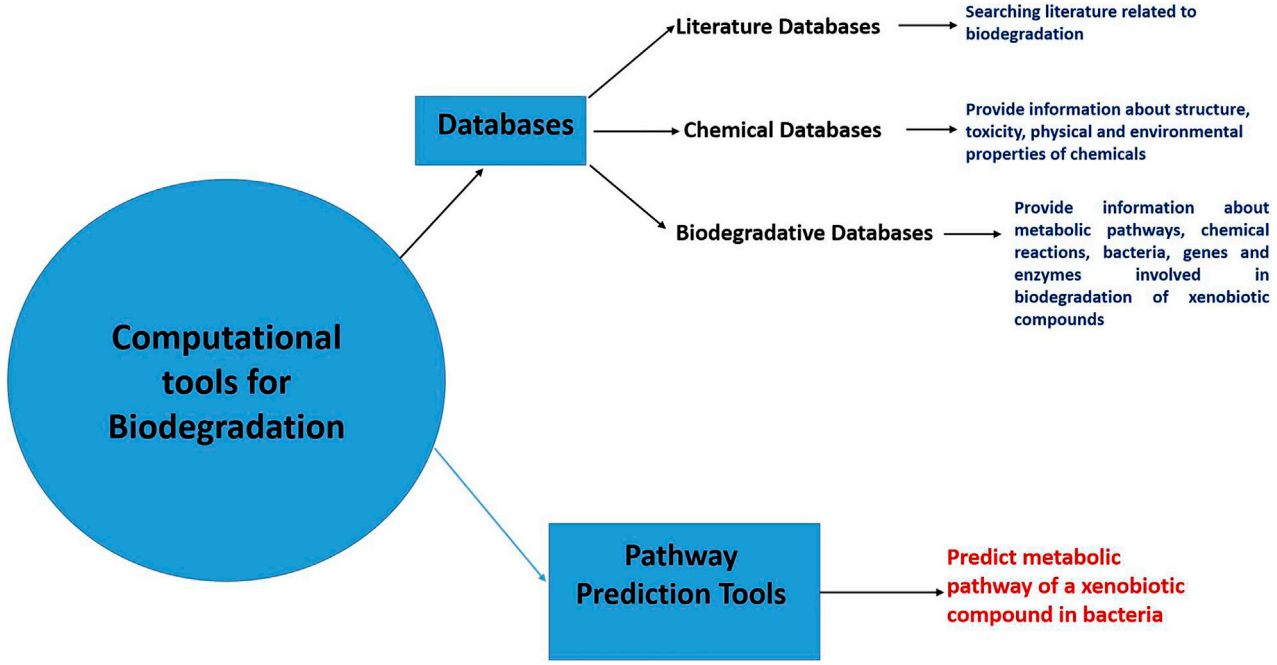

FIGURE 2
Role of various databases and pathways prediction systems in biodegradation

TABLE 1 List of environmentally important chemical databases.

Database	Description	References
ZINC 15	Database of compounds that are commercially available	Sterling et al. (2015)
SDBS	Spectral database for organic compounds	Saito and Kinugasa, (2011)
SureChEMBL	Freely available chemical data retrieved from patent literature	Papadatos et al. (2016)
NIST Chemistry WebBook	Contain thermochemical, ion-energetic, thermophysical and spectral data of various chemicals	Linstrom and Mallard, (2001)
EU Pesticide database	Database of chemicals used in plant protection products	Commission (2018)
Chemical and Products Database (CPDat)	This database maps more than 49,000 chemicals to a set of terms used to categorize or describe their use or function in 16,000 consumer products (e.g., shampoos, soaps)	Dionisio et al. (2018)
CompTox Chemistry Dashboard	A community data resource for environmental chemistry	Williams et al. (2017)
Aggregated Computational Toxicology Online Resource (ACToR)	The EPA's online aggregator of chemical toxicity data from all publicly available sources	Judson et al. (2008)
Distributed Structure-Searchable Toxicity (DSSTox)	The DSSTox offers a comprehensive public chemistry resource for improving predictive toxicology. One of its most unique features is the accurate mapping of bioassay results and physicochemical properties to corresponding chemicals	Richard and Williams, (2002)
ContaminantDB	Combines detailed contaminant data from different online references and databases on contaminants. The database currently houses 54,249 compound	Wishart, (2017)
T3DB: the toxic exposome database	Unique bioinformatics resource that combines detailed toxin data with comprehensive toxin target information	Wishart et al. (2015)

79 genes involved in microbial degradation of plastics (Gan and Zhang, 2019). In addition, information of 8000 predicted sequences of the enzymes involved in plastic biodegradation are also provided in this database.

ONDB includes information on chemical properties and biodegradability of commonly used organonitrogen compounds (Robinson et al., 2021). It provides insight into the pathways, reactions, microbes, and enzymatic reactions

TABLE 2 A list of Biodegradative Databases.

Database	Description	References
EAWAG-BBD	Biocatalysis/Biodegradation database that provides information about biodegradation of various xenobiotic compounds. This information include metabolic pathways, chemical reactions, bacteria, genes and enzymes involved in biodegradation	Ellis et al. (2006)
PMBD	Plastics Microbial Biodegradation Database that focuses on microbial degradation of plastics	Gan and Zhang, (2019)
ONDB	Organonitrogen Degradation Database that provides information about chemical properties and biodegradation nature of organonitrogen compounds	Robinson et al. (2021)
FRCD	Food risk component database that provides information about food risk components	Zhang et al. (2020)
OxDBase	A database of Biodegradative Oxygenases that includes both monooxygenases and dioxygenases involved in biodegradation of various xenobiotic compounds	Arora et al. (2009)
Bionemo	Describes how the biodegradation metabolism is mediated by proteins and genes	Carbajosa et al. (2009)
MetaCyc	The largest collection of metabolic pathways of all life forms	Caspi et al. (2020)
BioCyc	It includes 20,005 Pathway/Genome Databases (PGDBs) for model eukaryotes and thousands of microbes	Karp et al. (2019)
AromaDeg	Dedicated to aerobic degradation of aromatic compounds	Duarte et al. (2014)
RHObase	Aromatic ring-hydroxylating oxygenases database that provides information on bacterial Rieske-type ring-hydroxylating oxygenases	Chakraborty et al. (2014)

associated with degradation of organonitrogen compounds including urea, methylenediurea**,** Guanyl thiourea (1-amidino-2-thiourea, GTU), cyanuric acid, Cyanoguanidine, biuret (carbamoylurea), Allophanate.

FRCD ((http://www.rxnfinder.org/frcd/) is dedicated to food risk components which may be defined as chemicals which are present in foods and create problems to human health if they consume (Zhang et al., 2020). Examples of these compounds are toxic metals, pesticides, microbial metabolites and food additives. This database contains information of 12,018 toxic molecules from more than 150,,000 literature reports. Apart of the toxicity and biodegradability of these compounds, FRCD also provides data on molecular scaffold and chemical diversity of these compounds.

AromaDeg (http://aromadeg.siona.helmholtz-hzi.de) is a freely accesses web-based database that focuses on aerobic degradation of aromatic compounds. This database develops based on phylogenomic approach (Duarte et al., 2014). Through AromaDeg, it is possible to attempt to identify the amino acid sequences of key aerobic aromatic degradation enzyme families by querying novel genomic, metagenomic, or metatranscriptomic data sets. In initial step, each query sequence that belongs to one of the protein families considered in AromaDeg is aligned with other members of the respective protein family and thereby grouped in a specific cluster on the phylogenetic tree. In next step, functional annotation with substrate specificity can be determined based on experimentally validated functions of neighbouring cluster members. In this way, AromaDeg provides not only a comprehensive characterization of each protein superfamily but also high-throughput functional classifications of protein. This approach overcomes the limitations of homology-based function prediction and results in more accurate annotations of new biological functions associated with aerobic degradation pathways of aromatic compounds.

Bionemo (http://bionemo.bioinfo.cnio.es) is a manually curated database that was created by structural computational biology researchers at the Spanish National Cancer Research Center (Carbajosa et al., 2009). It contains information about the proteins and genes responsible for the metabolic process of biodegradation. There are 145 biochemical pathways, 945 reactions - 342 which have complexes associated with them, 537 enzymatic complexes, 1107 proteins, 234 microbial species, 90 transcription factors, 90 effectors, 128 TF DNA binding sites, and 100 promoters in this database.

MetaCyc contains information more than 2937 experimentally verified metabolic pathways from more than 3295 different organisms, which have been derived from experimental literature (Caspi et al., 2020). MetaCyc is the largest database that provide information about primary and secondary metabolic pathways as well as associated compounds, enzymes and genes. The database can be accessed for free at http://metacyc.org. MetaCyc is applicable to a variety of scientific fields. Additionally, it can serve as a database of reference data for computations of metabolic pathways by using sequenced genomes, (i) support metabolic engineering, (iii) facilitate comparisons of biochemical networks, and (iv) serve as a repository of metabolic knowledge from genomes. BioCyc group at SRI International developed and curated this database.

3 Pathway prediction systems

The ability to predict metabolic and degradation pathways of xenobiotics has great interest in the field of environmental sciences. Many researchers have developed bioinformatics-based pathway prediction systems that use knowledge-based

models, machine learning algorithms, or a hybrid approach. Both knowledge based and machine learning methods have disadvantages due to their dependency on existing data or transformation rules based on the available literature. Combining knowledge- and machine learning-based approaches, hybrid methods use machine learning-based relative reasoning models in order to predict the chances of individual transformation reactions. Example are BNICE (Biochemical Network Integrated Computational Explorer), and EAWAG-Pathway Prediction System (PPS).

BNICE is a computational prediction tool that produces all reactions of compounds know so far (Finley et al., 2009). This program uses enzyme reaction rules derived from the enzyme commission (EC) classification system. To predict a metabolic pathway, BNIC first identified functional group present in parent compounds then produces related compounds based on reaction rules. Each product undergoes this process until pathway is complete.

In EAWAG-PPS, pathway prediction is performed by identifying the functional chemical groups in a starting compound, and then determining the transformed product based on biotransformation rules (Gao et al., 2011)

4 Conclusion and future prospective

The biodegradation of chemicals is described in several databases. Users can use various databases to gain information about chemicals and their degradation. The biodegradative databases can be used to retrieve information about bacterial degradation of xenobiotic compounds including associated genes and enzymes. Similarly, using chemical databases, users can retrieve information about various properties of chemicals including their risk assessment and environmental properties. There are also several pathway prediction systems that can be used to predict degradation pathways for chemicals whose degradation pathways are unknown. Users can predict metabolic pathway of the compounds using various pathway prediction systems. These information will help to design a strategy for biodegradation of chemicals therefore, also help in improved bioremediation process of chemicals.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Acknowledgments

The author acknowledges the Department of Biotechnology, India to provide him Ramalingaswami Re-entry Fellowship.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Arora, P. K., and Bae, H. (2014). Integration of bioinformatics to biodegradation. *Biol. Proced. Online* 16 (1), 8–10. doi:10.1186/1480-9222-16-8

Arora, P. K., Kumar, M., Chauhan, A., Raghava, G. P., and Jain, R. K. (2009). OxDBase: A database of oxygenases involved in biodegradation. *BMC Res. Notes* 2 (1), 67–68. doi:10.1186/1756-0500-2-67

Arora, P. K., Sasikala, C., and Ramana, C. (2012). Degradation of chlorinated nitroaromatic compounds. *Appl. Microbiol. Biotechnol.* 93 (6), 2265–2277. doi:10.1007/s00253-012-3927-1

Carbajosa, G., Trigo, A., Valencia, A., and Cases, I. (2009). Bionemo: Molecular information on biodegradation metabolism. *Nucleic Acids Res.* 37 (1), D598–D602. doi:10.1093/nar/gkn864

Caspi, R., Billington, R., Keseler, I. M., Kothari, A., Krummenacker, M., Midford, P. E., et al. (2020). The MetaCyc database of metabolic pathways and enzymes - a 2019 update. *Nucleic Acids Res.* 48 (D1), D445–D453. doi:10.1093/nar/gkz862

Chakraborty, J., Jana, T., Saha, S., and Dutta, T. K. (2014). Ring-hydroxylating oxygenase database: A database of bacterial aromatic ring-hydroxylating oxygenases in the management of bioremediation and biocatalysis of aromaticcompounds. *Environ. Microbiol. Rep.* 6 (5), 519–523. doi:10.1111/1758-2229.12182

Commission (2018). *Regulation (EC) No 396/2005 : EU pesticide database*. Available online: http://ec.europa.eu/food/plant/pesticides/eu-pesticides-database/public/?event=pesticide.residue.CurrentMRLandlanguage=EN.

Dionisio, K. L., Phillips, K., Price, P. S., Grulke, C. M., Williams, A., Biryol, D., et al. (2018). The Chemical and Products Database, a resource for exposure-relevant data on chemicals in consumer products. *Sci. Data* 5 (1), 180125–180129. doi:10.1038/sdata.2018.125

Duarte, M., Jauregui, R., Vilchez-Vargas, R., Junca, H., and Pieper, D. H. (2014). AromaDeg, a novel database for phylogenomics of aerobic bacterial degradation of aromatics. *Database* 2014, bau118. doi:10.1093/database/bau118

Ellis, L. B., Roe, D., and Wackett, L. P. (2006). The university of Minnesota biocatalysis/biodegradation database: The first decade. *Nucleic Acids Res.* 34 (1), D517–D521. doi:10.1093/nar/gkj076

Finley, S. D., Broadbelt, L. J., and Hatzimanikatis, V. (2009). Computational framework for predictive biodegradation. *Biotechnol. Bioeng.* 104 (6), 1086–1097. doi:10.1002/bit.22489

Gan, Z., and Zhang, H. (2019). Pmbd: A comprehensive plastics microbial biodegradation database. *Database* 2019, baz199. doi:10.1093/database/baz119

Gao, J., Ellis, L. B., and Wackett, L. P. (2011). The university of Minnesota pathway prediction system: Multi-level prediction and visualization. *Nucleic Acids Res.* 39 (2), W406–W411. doi:10.1093/nar/gkr200

Judson, R., Richard, A., Dix, D., Houck, K., Elloumi, F., Martin, M., et al. (2008). ACToR—Aggregated computational toxicology resource. *Toxicol. Appl. Pharmacol.* 233 (1), 7–13. doi:10.1016/j.taap.2007.12.037

Karp, P. D., Billington, R., Caspi, R., Fulcher, C. A., Latendresse, M., Kothari, A., et al. (2019). The BioCyc collection of microbial genomes and metabolic pathways. *Brief. Bioinform.* 20 (4), 1085–1093. doi:10.1093/bib/bbx085

Linstrom, P. J., and Mallard, W. G. (2001). The nist chemistry WebBook: A chemical data resource on the internet. *J. Chem. Eng. Data* 46 (5), 1059–1063. doi:10.1021/je000236i

Papadatos, G., Davies, M., Dedman, N., Chambers, J., Gaulton, A., Siddle, J., et al. (2016). SureChEMBL: A large-scale, chemically annotated patent document database. *Nucleic Acids Res.* 44 (1), D1220–D1228. doi:10.1093/nar/gkv1253

Richard, A. M., and Williams, C. R. (2002). Distributed structure-searchable toxicity (DSSTox) public database network: A proposal. *Mutat. Research/Fundamental Mol. Mech. Mutagen.* 499 (1), 27–52. doi:10.1016/s0027-5107(01)00289-5

Robinson, S. L., Biernath, T., Rosenthal, C., Young, D., Wackett, L. P., and Martinez-Vaz, B. M. (2021). Development of the organonitrogen biodegradation database: Teaching bioinformatics and collaborative skills to undergraduates during a pandemic. *J. Microbiol. Biol. Educ.* 22 (1), 22.1.49–2351. doi:10.1128/jmbe.v22i1.2351

Saito, T., and Kinugasa, S. (2011). Development and release of a spectral database for organic compounds. *Synth. Engl. Ed.* 4 (1), 35–44. doi:10.5571/syntheng.4.35

Sterling, T., and Irwin, J. J. (2015). ZINC 15–ligand discovery for everyone. *J. Chem. Inf. Model.* 55 (11), 2324–2337. doi:10.1021/acs.jcim.5b00559

Williams, A. J., Grulke, C. M., Edwards, J., McEachran, A. D., Mansouri, K., Baker, N. C., et al. (2017). The CompTox chemistry dashboard: A community data resource for environmental chemistry. *J. Cheminform.* 9 (1), 61–27. doi:10.1186/s13321-017-0247-6

Wishart, D., Arndt, D., Pon, A., Sajed, T., Guo, A. C., Djoumbou, Y., et al. (2015). T3DB: The toxic exposome database. *Nucleic Acids Res.* 43 (D1), D928–D934. doi:10.1093/nar/gku1004

Wishart, D. S. (2017). ContaminantDB. http://contaminantdb.ca.

Zhang, D., Gong, L., Ding, S., Tian, Y., Jia, C., Liu, D., et al. (2020). Frcd: A comprehensive food risk component database with molecular scaffold, chemical diversity, toxicity, and biodegradability analysis. *Food Chem. x.* 318, 126470. doi:10.1016/j.foodchem.2020.126470