CHIMIA **2023**, *77*, 424–431, https://doi.org/10.2533/chimia.2023.424

Microbial Biocatalysis within Us: The Underexplored Xenobiotic Biotransformation Potential of the Urinary Tract Microbiota

Thierry D. Marti, Milo R. Schärer and Serina L. Robinson*

Supplementary Information

Supplementary Glossary

Term	Definition
Bacterial species	A controversial term to define. A popular and controversial definition: a group of bacteria with 97% sequence similarity of the 16S ribosomal RNA gene
Biotransformation	Biochemical modification/transformation of compounds mediated by enzymes.
BLAST	The basic local alignment search tool “finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance”
CFU	Colony forming unit. The number of bacteria that grow on culture media
Co-metabolism	Co-metabolism is a process in which a compound is transformed by an organism's metabolic activity in the presence of another compound, which may or may not be directly utilized as a source of energy or nutrients.
Enzyme commission number	A numeric method for classifying enzymes based on their catalytic activity and substrates
Gene	DNA sequence coding for a protein
Metabolism	The ensemble of chemical processes occurring in a living organism
Metabolomics	Describing a method or context which looks at all the metabolites of all organisms present in a sample
Metagenomics	Describing a method or context which looks at all the DNA of all organisms present in a sample
Microbiome	The ensemble of genes encoded by the microbiota
Microbiota	The ensemble of microorganisms (e.g., bacteria, archaea, protozoa, viruses, fungi) in a given environment
Phylogenetic	Describing the evolutionary relationship of species or genes

Term	Definition
Shotgun sequencing	A DNA sequencing technique that involves randomly breaking up large DNA strands into smaller fragments, which are re-assembled computationally post-sequencing
Xenobiotic	Foreign compound not expected to be naturally found in an organism

Supplementary Text

The biotransformation of xenobiotics by the human microbiota can have physiological consequences for the host. To date, the activity of gut microbial enzymes on pharmaceuticals is best understood.[1] Microbial enzymes can lead to desired activation of biologically inactive prodrugs. For example, sulfasalazine is an anti-inflammatory drug which is cleaved by widespread gut microbial azoreductases to its bioactive product 5-aminosalicylic acid.[2] However, microbial enzymes also perform unwanted modifications to metabolized or bioactive drugs resulting in either their reabsorption (enterohepatic recycling), inactivation or toxification.[1]

As example of enterohepatic recycling, gut microbial β-glucuronidases can release the glucuronide moiety of non-steroidal anti-inflammatory drugs, such as diclofenac, reaching the gut by biliary excretion. The removal of the glucuronide moiety increases the lipophilicity of the drug allowing its reabsorption by enterocytes, where it is further metabolized into reactive metabolites promoting inflammation and compromising mucosal integrity.[3]

Microbial enzymes can also inactivate or reduce the bioactivity of a drug. Digoxin, a heart medication, is transformed to dihydrodigoxin by reduction through gut microbial enzymes leading to subtherapeutic concentrations of the bioactive form of the drug.[4] This contributes to the so-called firstpass phenomenon, which describes a loss of bioactive drug at the drug's target site due incomplete absorption, and degradation by host gut and liver enzymes, and microbial gut enzymes. Another unwanted microbial-xenobiotic interaction is toxification. Cyclamate is a notorious example of a food additive transformed by several members of the human microbiota to its more toxic cyclohexylamine product.[5,6] Also azo dyes, which are food colorants containing an azo bond, can be reduced by azoreductases, releasing potentially carcinogenic aromatic amines.[7] Recently, gut bacteria have also been found to interact with xenobiotics by storing them. Duloxetine, an antidepressant, is bioaccumulated by *Clostridium saccharolyticum.*[8] Finally, the microbiota can also influence host xenobiotic metabolism through the production of microbial metabolites or by modifying host metabolites.[1]

Supplementary text references:

[1] P. Spanogiannopoulos, E. N. Bess, R. N. Carmody, P. J. Turnbaugh, *Nat. Rev. Microbiol.* **2016**, *14*, 273, DOI: 10.1038/nrmicro.2016.17.

[2] M. A. Peppercorn, P. Goldman, *J. Pharmacol. Exp. Ther.* **1972**, *181*, 555.

[3] K. S. Saitta, C. Zhang, K. K. Lee, K. Fujimoto, M. R. Redinbo, U. A. Boelsterli, Xenobiotica 2014, 44, 28, DOI: 10.3109/00498254.2013.811314.

[4] J. Lindenbaum, D. G. Rund, V. P. Butler Jr, D. Tse-Eng, J. R. Saha, *N. Engl. J. Med.* **1981**, *305*, 789, DOI: 10.1056/NEJM198110013051403.

[5] B. S. Drasar, A. G. Renwick, R. T. Williams, Biochem. J. **1972**, *129*, 881, DOI: 10.1042/bj1290881.

[6] J. M. Price, C. G. Biava, B. L. Oser, E. E. Vogin, J. Steinfeld, H. L. Ley, *Science* **1970**, *167*, 1131, DOI: 10.1126/science.167.3921.1131.
[7] J. Feng, C. E. Cerniglia, H. Chen, *Front. Biosci.* **2012**, *4*, 568, DOI: 10.2741/e400.
[8] M. Klünemann, S. Andrejev, S. Blasche, A. Mateus, P. Phapale, S. Devendran, J. Vappiani, B. Simon, T. A. Scott, E. Kafkia, D. Konstantinidis, K. Zirngibl, E. Mastrorilli, M. Banzhaf, M.-T. Mackmull, F. Hövelmann, L. Nesme, A. R. Brochado, L. Maier, T. Bock, V. Periwal, M. Kumar, Y. Kim, M. Tramontano, C. Schultz, M. Beck, J. Hennig, M. Zimmermann, D. C. Sévin, F. Cabreiro, M. M. Savitski, P. Bork, A. Typas, K. R. Patil, *Nature* **2021**, *597*, 533, DOI: 10.1038/s41586-021-03891-8.

Supplementary Methods

Taxonomic analysis of microbiota

A data set previously described by Biehl et al.[1] was used to analyze the taxonomic composition of the gut, urinary and vaginal microbiota. Briefly, microbial samples were collected at these body sites from 15 premenopausal women and characterized by 16S rRNA sequencing with taxonomic assignment at the ASV level.[1] This data was kindly provided by Dr. Lena Biehl, Dr. Fejda Farowski and Prof. Maria Vehreschild in the .biom format. All data processing and visualization was done using R (version 4.2.2) within the RStudio IDE (version 2022.12.0.353). A list of packages used can be found below. Samples collected from midstream urine were removed from the dataset prior to analysis. ASV counts were normalized using the TMM method and agglomerated at the genus level.[2,3] The sum of genus counts per body site (gut, urine and vagina) was calculated. Genera with an overall relative abundance of < 1% were binned as other.

Software	**Version**	**Use**
BLAST+	2.9.0	Sequence alignment
R	4.2.3	Data processing and visualization
RStudio	2023.3.0.386	IDE for R
Meka (Java package)	1.0	Ensembles of Classifier Chains

Biotransformation prediction

A representative set of xenobiotics frequently detected in urine with > 80% relative excretion in urine relative to stool were selected from across different chemical classes (Supplementary table 1). For each compound, we used a published machine learning approach to predict probable microbial chemical biotransformations[4] implemented in the enviPath tool.[5] Briefly, enviPath uses Ensembles of Classifier Chains[6] to predict the probabilities of different biotransformation rules. Predicted biotransformation rules are depicted in Supplementary table 2. Each biotransformation rule was paired with candidate Enzyme Commission (EC) numbers predicted using enviLink[7], resulting in 338 distinct EC numbers corresponding to 49 numbers at the third level EC number and 13 numbers at the second level EC number. The ECs spanned three first-level numbers: EC 1 oxidoreductases, EC 2 transferases, and EC 3 hydrolases (Fig. 4). Although lyases (EC 4) are also relevant in many xenobiotic biotransformations in the gut,[24] biotransformation rules using these enzymes were not triggered. To cross-reference these predictions, we compared our EC number hits with results from the DrugBug prediction tool built specifically for gut xenobiotic biotransformations[8] but overall results were in good accordance. For each EC number, 1-3 representative enzymes at the second EC number level were selected as query sequences (Supplementary table 4). Query sequences were limited to characterized enzymes reported in the literature with either documented biotransformation

activity for the study set xenobiotics or with experimental evidence in the curated SwissProt database.[9] Proteins with high sequence similarity can catalyze different reactions, thus these computational analyses generate hypotheses which must be experimentally verified. Moreover, this method does not account for horizontal gene transfer e.g., of plasmid-encoded xenobiotic-degrading enzymes.[25] Nonetheless this analysis provides an initial picture of xenobiotic biotransformation potential to highlight future targets for experimental investigation.

Package	Version	Use
janitor [10]	2.2.0	Clean text within imported data
ggalluvial [11]	0.12.4	Create alluvial plots
ggnewscale [12]	0.4.8	Multiple color scales in one plot
ggpattern [13]	1.01	Add stripes to plot
ggpubr [14]	0.6.0	Modify plot aesthetics
pacman [15]	0.5.1	Package manager
pBrackets [16]	1.0.1	Add brackets to plot
phyloseq [17]	1.42.0	Process taxonomic data
RColorBrewer [18]	1.1-3	Color palettes for plots
readxl [19]	1.4.2	Import excel tables into R
tidyverse [20]	2.0.0	Data transformation and visualization

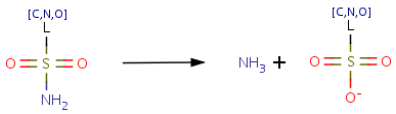
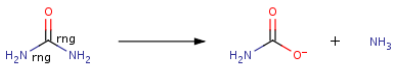
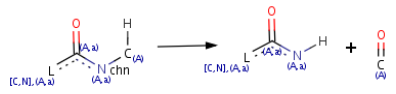
Xenobiotic degradation enzymes in urinary microbial isolates

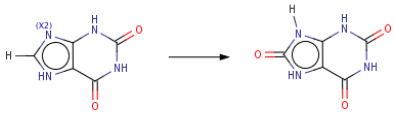
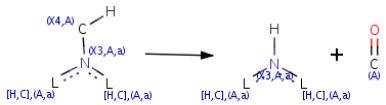
The genomes of urinary microbial isolates previously described by Thomas-White et al. were used in this analysis.[21] Genomes from females asymptomatic for UTIs were selected for this study, which matched NCBI assembly accession numbers and the strain classification. Of the 67 genomes of bacterial isolates from asymptomatic patients, 56 assemblies with matching strain classifications could be retrieved from the NCBI assembly database and were used in the study (Supplementary table 3). Sequence alignment of query protein sequences to reference genomes was done using a reverse protein alignment search with BLAST+. Significance cutoffs of evalue ≤ 0.1, query coverage ≥ 20 % and bitscore ≥ 50 were used.[22,23] All data processing and visualization was done using R (version 4.2.2) within the RStudio IDE (version 2022.12.0.353). A list of packages used can be found above.

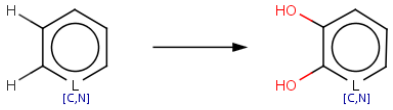
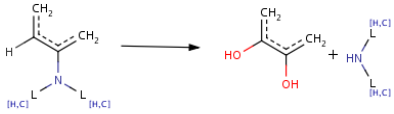
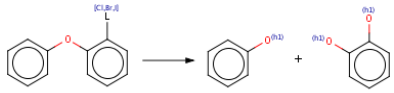
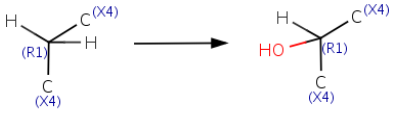
Supplementary table 1. Xenobiotics included in this study

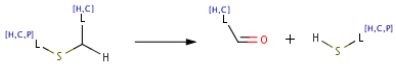
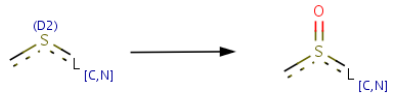
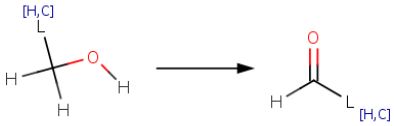
Name	% Excretion*	Xenobiotic class	Reference
Acesulfame	>99	Dietary component	Magnuson et al. 2016[26]
Caffeine	85	Dietary component	Wilson et al.[27]
Diclofenac	100	Analgesic	Köpping et al. 2020[28]
Emtricitabine	84	Virostatic	Köpping et al. 2020[28]
Gabapentin	100	Antiepileptic	Lienert et al. 2007[29]
Paracetamol	85-95	Analgesic	Forrest et al. 1982[30]
Saccharin	85-95	Dietary component	Magnuson et al. 2016[26]
Sulfamethoxazole	100	Antibiotic	Köpping et al.[28]
Tramadol	>90	Analgesic	Lienert et al. [29]
Trimethoprim	100	Antibiotic	Köpping et al.[28]

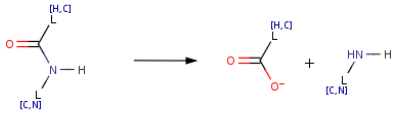
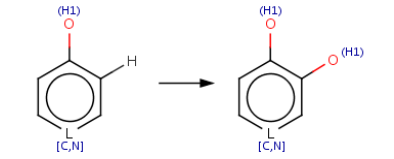
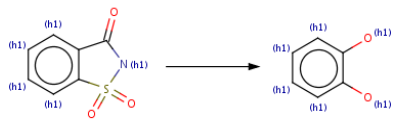
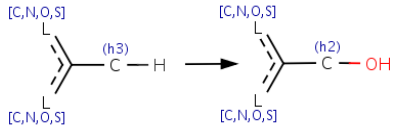
Supplementary table 2. Table of enviPath biotransformation rules triggered by included compounds including rule ID, description of activity and schematic of reactions [5].

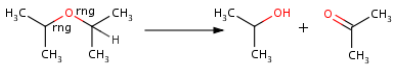
Rule	Description	Formula
bt0144	Sulfamate -> Amine Sulfonamide -> Amine + Sulfonate	
bt0291	Alkene -> Alkane	
bt0402	cyclic Urea derivative -> 1-Aminocarbamate derivative	
bt0243	N-substituted Amide -> Amide + Aldehyde or Ketone N, N-disubstituted Amide -> N-substituted Amide + Aldehyde or Ketone N-substituted Urea derivative	

	-> Urea derivative + Aldehyde or Ketone N,N-disubstituted Urea derivative -> N-substituted Urea derivative + Aldehyde or Ketone	
bt0316	Xanthine derivative -> Uric acid derivative	
bt0063	primary Amine -> Aldehyde or Ketone secondary Amine -> Amine + Aldehyde or Ketone tertiary Amine -> secondary Amine + Aldehyde or Ketone Methylammonium derivative -> Trimethylamine + Aldehyde or Ketone	

Rule	**Description**	**Formula**
bt0005	*vic*-unsubstituted Aromatic -> *vic*-Dihydroxyaromatic	
bt0065	1-Amino-2-unsubstituted aromatic -> *vic*-Dihydroxyaromatic + Amine 1-Amide-2-unsubstituted aromatic -> *vic*-Dihydroxyaromatic + Amide	
bt0374	polynuclear Aromatic system -> 1,2-dioxygenation and cleavage at connecting atom	
bt0242	secondary Aliphatic -> secondary Alcohol	

bt0029	organoHalide -> RH	
bt0402	cyclic Urea derivative -> 1-Aminocarbamate derivative	
bt0259	disubstituted Sulfide -> Aldehyde + Sulfide Monoalkylthiol -> Aldehyde + H_2S Thiol alkene -> Alcohol Monoalkylthiophosphate -> Aldehyde + Thiophosphate	
bt0162	disubstituted Sulfide -> disubstituted Sulfoxide	
bt0001	primary Alcohol -> Aldehyde	

Rule	Description	Formula
bt0067	secondary Amide -> Carboxylate + primary Amine Lactam -> Aminecarboxylate	
bt0014	1-Hydroxy-2-unsubstituted aromatic -> 1,2-Dihydroxyaromatic 4-Hydroxypyridine derivative -> 3,4-Dihydroxypyridine derivative	
bt0425	Saccharin -> Catechol	
bt0036	aromatic Methyl -> primary Alcohol	

bt0241	tertiary Aliphatic -> tertiary Alcohol	
bt0023	dialiphatic Ether -> Alcohol + Aldehyde aromatic-aliphatic Ether -> Phenol derivative + Aldehyde	

Supplementary table 3. NCBI Assembly database accession numbers of urinary microbial isolate genomes used in this study from Thomas-White et al.[21]

NCBI accession	Species	Strain
GCA_002860635.1	*Actinomyces naeslundii*	UMB0731
GCA_002861505.1	*Aerococcus christensenii*	UMB0844
GCA_003286755.1	*Aerococcus urinae*	UMB0722
GCA_002861485.1	*Alloscardovia omnicolens*	UMB0006
GCA_002858875.1	*Bacillus sp. UMB0728*	UMB0728
GCA_002858905.1	*Bacillus sp. UMB0893*	UMB0893
GCA_002861455.1	*Bifidobacterium breve*	UMB0915
GCA_002861445.1	*Bifidobacterium longum*	UMB0788
GCA_002847785.1	*Corynebacterium amycolatum*	UMB0042
GCA_002861385.1	*Corynebacterium aurimucosum*	UMB0043
GCA_002884935.1	*Corynebacterium tuscaniense*	UMB0792
GCA_002861285.1	*Enterococcus faecalis*	UMB0048
GCA_003892485.1	*Escherichia coli*	UMB0731
GCA_002861265.1	*Escherichia coli*	UMB0789
GCA_002861205.1	*Escherichia coli*	UMB0727
GCA_002884795.1	*Gardnerella vaginalis*	UMB1642

NCBI accession	Species	Strain
GCA_002862015.1	*Gardnerella vaginalis*	UMB0032A
GCA_002862005.1	*Gardnerella vaginalis*	UMB0032B
GCA_002861905.1	*Gardnerella vaginalis*	UMB0830
GCA_002861885.1	*Gardnerella vaginalis*	UMB0833
GCA_002861145.1	*Gardnerella vaginalis*	UMB0913
GCA_002861125.1	*Gardnerella vaginalis*	UMB0912
GCA_002847865.1	*Gordonia terrae*	UMB0777
GCA_002848015.1	*Lacticaseibacillus rhamnosus*	UMB0004
GCA_002863245.1	*Lactobacillus crispatus*	UMB0040
GCA_002863505.1	*Lactobacillus crispatus*	UMB1398
GCA_002863485.1	*Lactobacillus crispatus*	UMB0054
GCA_002861765.1	*Lactobacillus crispatus*	UMB0803
GCA_002861775.1	*Lactobacillus crispatus*	UMB0044
GCA_002847905.1	*Lactobacillus delbrueckii*	UMB0003
GCA_002863425.1	*Lactobacillus gasseri*	UMB0045
GCA_002884735.1	*Lactobacillus gasseri*	UMB0056
GCA_002884695.1	*Lactobacillus iners*	UMB1051

NCBI accession	Species	Strain
GCA_002863405.1	*Lactobacillus jensenii*	UMB0007
GCA_026184175.1	*Lactobacillus mulieris*	UMB0021
GCA_002863365.1	*Micrococcus luteus*	UMB0031
GCA_002863375.1	*Micrococcus luteus*	UMB0867
GCA_002863345.1	*Micrococcus luteus*	UMB0038
GCA_002863305.1	*Neisseria perflava*	UMB0023
GCA_012030315.1	*Proteus mirabilis*	UMB0038
GCA_002847605.1	*Pseudoglutamicibacter albus*	UMB0722
GCA_002861015.1	*Rothia mucilaginosa*	UMB0024
GCA_002847525.1	*Schaalia odontolytica*	UMB0018
GCA_002884615.1	*Staphylococcus pettenkoferi*	UMB0834
GCA_002871535.1	*Streptococcus agalactiae*	UMB0776
GCA_002861005.1	*Streptococcus agalactiae*	UMB0767
GCA_002884585.1	*Streptococcus agalactiae*	UMB0049
GCA_002848125.1	*Streptococcus anginosus*	UMB0050
GCA_002860945.1	*Streptococcus anginosus*	UMB0839
GCA_012030555.1	*Streptococcus anginosus*	UMB0839

NCBI accession	Species	Strain
GCA_002860805.1	*Streptococcus macedonicus*	UMB0733
GCA_002860865.1	*Streptococcus mitis*	UMB1341
GCA_002860905.1	*Streptococcus oralis subsp. dentisani*	UMB0832
GCA_002860885.1	*Streptococcus oralis subsp. dentisani*	UMB0008

Supplementary table 4. Enzyme query sequences used in reverse blast search annotated by name and EC class.

Name	EC class	Sequence
caffeine dehydrogenase	1.1	MFADINKGDAFGTWVGKSVPRREDADILAGRAEYIADIKLPGML EAAFLRSPFAHARIVSIDVSQALALPGVYDVMVGADIPDYVKPL PLMITYQNHRETPTSPLARDIVRYAGEPVAVVAAINRYVAEDAL ELIVVKYEELPVVASIDASLAVDGPRLYEGWPDNVVAKVSSEIG DVDAAMASADLVFEERFEIQRCHPAPLETRGFIAQWDFKGENL NVWNGTQIINQCRDFMSEVLDIPASKIRIRSPRLGGGFGAKFHF YVEEPAIVLLAKRVKAPVRWIEDRLEAFSATVHAREQVIDVKLC AMNDGRITGIVADIKGDLGASHHTMSMGPVWLTSVMMTGVYLI PNARSVAKAIVTNKPPSGSYRGWGQPQANFAVERMVDLLAHK LQLDPAAVRRINYVPEARMPYTGLAHTFDSGRYEVLHDRALKT FGYEAWLERQAAAQAQGRRIGIGMSFYAEVSAHGPSRFLNYV GGRQGGYDIARIRMDTTGDVYVYTGLCDMGQGVTNSLAQIAA DALGLNPDDVTVMTGDTALNPYTGWGTGASRSITIGGPAVMR AATRLREKILSIARHWLQADPDTLVLANRGVMVRDDPGRYVSF ASIGRAAYCQIIELPEDVEPGLEAVGVFDTVQLAWPYGMNLVAV EVDEDTGAVSFLDCMLVHDMGTIVNPMIVDGQLHGGIAQGIAQ ALYEELRYDENGQLGTGSFADFLMPTASEIPNMRFDHMVTESP LIPGGMKGVGEGGTIGTPAAVVNAIENALRPITNSKLNRTPVTP DRILTAISAGACA
sulfonamide monooxygenase SadA	1.1	MVDSSLPPDISALLERIDEIRPLIEKNAAQGEAERRVSQESIDAL EAIGAFRVTQPAKYGGYEGDSRAQVDVGAAVGKGDGGTAWV VALTNIANWLTALYPEKAQDDVWGEDRNAKVSVVLATNGKTQ RVDGGYLVSGEWSYNSASWHTQWAILGAELVDENGDFVDTA QLLIPRSDLGFKDIWHVAGMRSSGSNALSATDVFVPDHRVMR GEPALRGTYPGTTEDTPAVYRAGWIPVLNIILVGPQLGMGRAVL ERVISKADSKAIAYTSFERQSDSIAFQLDIAKAALLLEAAEGFAH RATDEIDIPAAQGVYPDYLTRARNRAYVGWIVEHTARAIEMLLT AHGSGAFAEVNPLQRLWRDQAVASRHAFVLPALGYELYGKAL LGREDGDSVTPLV

Name	EC class	Sequence
caffeine demethylase subunit alpha	1.1	MATRSWGAANKITSSTKATYMEQAIINNDREYLRHFWHPVCTV TELEKAHPSSLGPMAVTLLSEQLVVAKLGGEYVAMHDRCAHR SAKLSLGTIANDRLQCPYHGWQYDSEGACKLVPACPNSPIPNR AKVQRFDCEERYGLIWVRLDSSYACTEIPYFSAASDPRLRIVIQ EPYWWNATAERRWENFTDFSHFAFIHPGTLFDPNNAEPPIVPM DRFNGQFRFIYDTPEDMAVPNQAPIGSFSYTCSMPFAINLEVAK YSSKSLHVLFNVSCPIDDTTTKNFLLFAREQADDSDYLHIAFNDL VFAEDKPVIESQWPKDAPADEVSVVADKVSIQYRKWLRELKEA HKEGAQAFRSALLDPVIESDRSYT
reductive dehalogenase	1.2	MGEINRRNFLKVSILGAAAAAVASASAVKGMVSPLVADAADIVA PITETSEFPYKVDAKYQRYNSLKNFFEKTFDPEANKTPIKFHYD DVSKITGKKDTGKDLPTLNAERLGIKGRPATHTETSILFHTQHLG AMLTQRHNETGWTGLDEALNAGAWAVEFDYSGFNATGGGPG SVIPLYPINPMTNEIANEPVMVPGLYNWDNIDVESVRQQGQQW KFESKEEASKIVKKATRLLGADLVGIAPYDERWTYSTWGRKIYK PCKMPNGRTKYLPWDLPKMLSGGGVEVFGHAKFEPDWEKYA GFKPKSVIVFVLEEDYEAIRTSPSVISSATVGKSYSNMAEVAYKI AVFLRKLGYYAAPCGNDTGISVPMAVQAGLGEAGRNGLLITQK FGPRHRIAKVYTDLELAPDKPRKFGVREFCRLCKKCADACPAQ AISHEKDPKVLQPEDCEVAENPYTEKWHLDSNRCGSFWAYNG SPCSNCVAVCSWNKVETWNHDVARVATQIPLLQDAARKFDEW FGYNGPVNPDERLESGYVQNMVKDFWNNPESIKQ
dihydropyrimidine dehydrogenase subunit PreA	1.3	MIKKDLSVDFLGVHFENPFCLSSSPVGNCYEMCKNAYDAGWG GIVYKTLSPTHFKIDEVSPRFDELAKEDMHFVAFKNMEQLSEHP LEQDLADMRRLKEEYPNKVLIASIMGETLEDWTNLAKLVTETGA DMIELNFSCPQMTSHTMGSDVGTNPELCKKNCEAVKRGTSLP VLAKMTPNITTMVPVVKACLEGGADGFSAINTVKSIVDVDLKKK VGLPNIDGKSSVSGLSGKAVKPIALRFLQQLRSAAGLEQLPISGI GGIETWEDAAEFILLGATTLQVTTAIMEYGYRIIDDLTNGLMHYM EEQHVDHLQDLVGLANKNIIPTNQLDRNYKVYPKIDWDKCIGCG RCFISCQDGAHQALTWDDEKRQPVFDKSKCVGCQLCALVCPV GAIKLGLVEIKPGHKGNPAEIDVGSQRLHRYHPVKANQEN

Name	EC class	Sequence
NADP-dependent oxidoreductase PreT	1.3	MTSKYETESKGYTMTTVMQEAARCLLCHDAPCSQACPAHTNP AKFIRSVLFRNVKGAAETIRENNALGSICARVCPTERYCEKACT RAKIDGPIDIGGIQRYVTDMERKMNMKILKAGKPNGMSIAIIGSG PSGLQAATTLREKGYAVDIYEKAAKAGGYLTYGIPEYRLPEEIV DYEVQRIVDLGAHIIYNTTVGKDISMDDLKARYNAVIVAIGTSEA KMLPMFEHNICTESAISFLARAKESKGNLEDLPQNVLVIGGGDV AMDVVTTLKKLNVPYVTDVIYEQFDEFKASKKELAGAQEAGVTI VDGYVPKEVHQNRATFTHRKIKSELTITADKIILAVGQKANAEGL DIDLQHNEIPFREPRFRTKDPKVFATGDIVAGDKTVVVAVQKGK EVAEEIDRLLGGQEND
amine oxidase	1.4	MGSPSLYSARKTTLALAVALSFAWQAPVFAHGGEAHMVPMDK TLKEFGADVQWDDYAQLFTLIKDGAYVKVKPGAQTAIVNGQPL ALQVPVVMKDNKAWVSDTFINDVFQSGLDQTFQVEKRPHPLN ALTADEIKQAVEIVKASADFKPNTRFTEISLLPPDKEAVWAFALE NKPVDQPRKADVIMLDGKHIIEAVVDLQNNKLLSWQPIKDAHG MVLLDDFASVQNIINNSEEFAAAVKKRGITDAKKVITTPLTVGYF DGKDGLKQDARLLKVISYLDVGDGNYWAHPIENLVAVVDLEQK KIVKIEEGPVVPVPMTARPFDGRDRVAPAVKPMQIIEPEGKNYTI TGDMIHWRNWDFHLSMNSRVGPMISTVTYNDNGTKRKVMYE GSLGGMIVPYGDPDIGWYFKAYLDSGDYGMGTLTSPIARGKDA PSNAVLLNETIADYTGVPMEIPRAIAVFERYAGPEYKHQEMGQP NVSTERRELVVRWISTVGNYDYIFDWIFHENGTIGIDAGATGIEA VKGVKAKTMHDETAKDDTRYGTLIDHNIVGTTHQHIYNFRLDLD VDGENNSLVAMDPVVKPNTAGGPRTSTMQVNQYNIGNEQDAA QKFDPGTIRLLSNPNKENRMGNPVSYQIIPYAGGTHPVAKGAQ FAPDEWIYHRLSFMDKQLWVTRYHPGERFPEGKYPNRSTHDT GLGQYSKDNESLDNTDAVVWMTTGTTHVARAEEWPIMPTEW VHTLLKPWNFFDETPTLGALKKDK
pseudooxynicotine oxidase	1.4	MANDKGDISKDGVSRRKFLGGAVIGAAAAAGVGSQILSLSATA QGADKERVGPLQSNVDYDAVVIGGGFAGVTAARELSRSGLKT LVLEGRSRLGGRTFTSKLDGEKVELGGTWVHWTQPNVWTEV MHYGLEIEETVGLASPETVIWVTDNQVKRAPAAEAFEIFGAACT EYYKEAHNIYPRPFDPFFAKKALQEMDGLSASEYLNKLSLTRE QKDMMDSWLSGNGHNYPETIAYSEIMRWFALSNFNMPTMFDS IARYKIKSGTVSLLEAMVAESDMEVQLSTPVLKVKQDSHRVLITT EEGTIAASAVVMAVPLNTMGDVEYSPRLSDAKSEIASQGHAGK GVKGYIRIKQDVGNVMTYAPARNDVTPFTSVFTDHVGENGTLLI AFSADPKLVDINDSKAVEKALHPLLPGVEVTSSYGYDWNLDPF SKGTWCTYRPGQTTRYLTELQKREGRLFFAGSDMANGWRGFI DGAIESGREVGYQVASYLKGKNSNA

Name	EC class	Sequence
flavin reductase SadC	1.5	MTSESPTPKHAQESSSGGGALVLDTLSADDFRAIFRRHPAGVT VVTADAGTGPVALTATSVASMSADPPLLIFSVSSSSSSARILKD ANTVVVHFIGPDSIEIAKLGATSGIDRFADSTIWSRLPTGEVVFD AVRTWIRARVVNRLEASGSTIVIAQAIESNYTAAELDQQPEREG LVFMNRSWHQVGERSAII
urate oxidase	1.7	MMRLKQLNEMSASEFIHLLGGVFENSSWVAERAEPNRPYSSF QSLYNKMVEIVETASENEQLKLIQMHPHLGTNVKITDF SQEEQKHAGLNELTEDEHNHLMLLNKEYMDKFGFPFVMAVRG KTKQDIYRTIKERLKNNYRTEFEQALEEIKKIAMFRLQ EIINGGEMISMTNYKERVMYYGKGDVFAYRTYLKPLTGVRTIPE SPFSGRDHILFGVNVKISVGGTKLLTSFTKGDNSLV VATDSMKNFIQKHLASYTGTTIEGFLEYVATSFLKKYSHIEKISLI GEEIPFETTFAVKNGNRAASELVFKKSRNEYATA YLNMVRNEDNTLNITEQQSGLADLQLIKVSGNSFVGFIRDEYTT LPEDTNRPLFVYLNIKWKYKNIEDSFGDNPEYYVAA EQIRDIATSVFHETETLSIQHLIYLIGCRILERFPQLQEVNFESQN HTWDKIVEEIPGSQGKVYTEPRPPYGFQCFTVTQ EDLQHKNIPMLSAEIQ
vanillate-*O*-demethylase	2.1	MSAPTNLEQVLAAGGNTVEMLRNSQIGAYVYPVVAPEFSNWR TEQWAWRNSAVLFDQTHHMVDLYIRGKDALKLLSDTMINSPKG WEPNKAKQYVPVTPYGHVIGDGIIFYLAEEEFVYVGRAPAANW LMYHAQTGGYNVDIVHDDRSPSRPMGKPVQRISWRFQIQGPK AWDVIEKLHGGTLEKLKFFNMAEMNIAGMKIRTLRHGMAGAPG LEIWGPYETQEKARNAILEAGKEFGLIPVGSRAYPSNTLESGWI PSPLPAIYTGDKLKAYREWLPANSYEASGAIGGSFVSSNIEDYY VNPYEIGYGPFVKFDHDFIGRDALEAIDPATQRKKVTLAWNGD DMAKIYASLFDTEADAHYKFFDLPLANYANTNADAVLDAAGNV VGMSMFTGYSYNEKRALSLATIDHEIPVGTELTVLWGEENGGT RKTTVEPHKQMAVRAVVSPVPYSVTARETYEGGWRKAAVTA

Name	EC class	Sequence
pyrimidine phosphorylase	2.4	MRMVDIIHKKRSGNVLSDQEIQFFVDGVVSGEIPDYQISALLMAI YFQGMTDSEQATLTMKMMTSGDHLDLSSIPGIKVD KHSTGGVGDKVSIPLAAVIAAMGIPIPMISGRGLGHTGGTLDKL EAIPGYQVEMSEAKFIEQIKRDKCAIIGATGNIAPA DKKIYALRDVTDTVDSIPLIASSIMSKKIASGTDALIIDVKTGAGAF MKTLDDSRALAKALVSIGKGVGMQCMALITDMN QPLGRAIGNALEIQESIDVLKGNGPTDLEKLITAIGGYMAVMGG KAKTIGEGQKLAETVIHNGQGLKSFKRMIQDQGGDS NVVDEPTDILPQAAYQIDLPAKRTGIISKMVADEIGVASMLLGGG RQKANDKLDYSVGIYLNKKIGDPISEGESILTIHS NRKDVEDIKKILYDNIEISATAKKPKLIYETVGWSVED
pyridoxamine pyruvate transaminase	2.6	MMRYPEHADPVITLTAGPVNAYPEVLRGLGRTVLYDYDPAFQL LYEKVVDKAQKAMRLSNKPVILHGEPVLGLEAAAASLISPDDVV LNLASGVYGKGFGYWAKRYSPHLLEIEVPYNEAIDPQAVADML KAHPEITVVSVCHHDTPSGTINPIDAIGALVSAHGAYLIVDAVSS FGGMKTHPEDCKADIYVTGPNKCLGAPPGLTMMGVSERAWAK MKANPLAPRASMLSIVDWENAWSRDKPFPFTPSVSEINGLDVA LDLYLNEGPEAVWARHALTAKAMRAGVTAMGLSVWAASDSIA SPTTTAVRTPDGVDEKALRQAARARYGVVFSSGRGETLGKLT RIGHMGPTAQPIYAIAALTALGGAMNAAGRKLAIGKGIEAALAVI DADA
N-acetylgalactosamine sulfatase	3.1	MAVQPNFLFIFMDDMGWRDLACTGSTFYETPNIDRLCRQGMV FANSYASCPVCSPSRASYLTGQYPARLGVTDWIDMEGTSHPL RGKLIDAPYIKHLPEGEYTIAQALKDAGYETWHVGKWHLGGRE YYPDHFGFDVNIGGCSWGHPHEGYFSPYGIETLPEGPEGEYLT DRITDEAVRLLKERKAGGSRKPFYMNLCHYAVHTPIQVKDEDR ERFEKKAREQGLDQETALVEGEFHHTEDKKGRRVVRRVIQSD PSYAGMIWNLDQNIGRLLEALSECGEEENTVVVFTSDNGGLAT SEGSPTCNLPASEGKGWVYEGGTRVPLIVKYPGHVAPGSRCD VPVTTPDFYPTFLELAGVPQKSGIPIDGRSIVPLLAGNHMPERP VFWHYPHYGNQGGTPAASVVLGDYKYIEFFEDGRGELYDLKA DFSETNNICENMPEMAARLRMLLHGWQREVCARFPEVNEAYG EV

Name	EC class	Sequence
sulfoglucosamine sulfohydrolase	3.1	MTFRHRFLLLLVTAAALVSGWAIEKAPERRDAGSAPTVPPNILW ISCEDMSPRLGCYGDTTIPTPNIDRLAREGIRFTNAFCTAGVCA PSRNAIITGMYQTSTGGHNMRTQYDTYPAKTGLPKEYSVVMAP EVKAFPEFLRAAGYYATNNVKTDYQFEAPPTVWDEVSNKAYW KNRPDNRPFFAVFNNTVTHESQVWQRKDLPLRADPARIKVPPY YPDTKTVRQDMARFYSNIRDMDDWVGDILKQLETDGLLDKTIIF FWSDHGDGLPFVKREIYDRGLRVPLIVRFADGRFAGTTRDELIS MIDLAPTVLTLAGLTPPTYMQGRAFLDPQTGRRPATGQPRRYV FGARDRLDSEYDRVRTVHDGRYQYVRNFFPDRPLYMDIAFRK QQPMMAELLQLRDAGKLNPTQMLWFRPNKPAEELYDLTTDPY ELTNLAEKPAYAGHLKRLRKEMDKWLTETNDLGKIPEKELVQQ MWQGADKPPVTAAPQATRSGDKVALSCSTPGASIAYRIGDSK SWQVYTKPIDIPRQQYMTAVAMRIGYTRSPEVPVTP
acetoacetamide-*N* -sulfonic acid (ANSA) hydrolase	3.1	MAKALMLGSGAALPAKGRFNTSLAILEGARTLLIDCAQPASELL YHHGVDIVSVDTVVITHMHADHVTGVGQLAHLKHILFDNKPPRV FLDRNDGFIKDNLRHPHRNELNNVNPWLNIYVPAGVEETMTQY LSALYMRPEVFAKYSVSVLPYGEGEFHSDENFKLVAYPNAHIR EFYPELQNTDAVLSSYTIMIETLGRKCLYSSDLASFEEIDHLVDG ADTIFVEGAHFSPSELIRFAQEHKLEQVFVHHILATREEEFARLA RELAQANVKLTFDGFEVEL

Name	EC class	Sequence
beta-glucoronidase	3.2	MLRPVETPTREIKKLDGLWAFSLDRENCGIDQRWWESALQES RAIAVPGSFNDQFADADIRNYAGNVWYQREVFIPKGWAGQRIV LRFDAVTHYGKVWVNNQEVMEHQGGYTPFEADVTPYVIAGKS VRITVCVNNELNWQTIPPGMVITDENGKKKQSYFHDFFNYAGIH RSVMLYTTPNTWVDDITVVTHVAQDCNHASVDWQVVANGDVS VELRDADQQVVATGQGTSGTLQVVNPHLWQPGEGYLYELCVT AKSQTECDIYPLRVGIRSVAVKGEQFLINHKPFYFTGFGRHEDA DLRGKGFDNVLMVHDHALMDWIGANSYRTSHYPYAEEMLDW ADEHGIVVIDETAAVGFNLSLGIGFEAGNKPKELYSEEAVNGET QQAHLQAIKELIARDKNHPSVVMWSIANEPDTRPQGAREYFAP LAEATRKLDPTRPITCVNVMFCDAHTDTISDLFDVLCLNRYYGW YVQSGDLETAEKVLEKELLAWQEKLHQPIIITEYGVDTLAGLHS MYTDMWSEEYQCAWLDMYHRVFDRVSAVVGEQVWNFADFA TSQGILRVGGNKKGIFTRDRKPKSAAFLLQKRWTGMNFGEKP QQGGKQ
dinitroanisole-*O*-demethylase subunit beta	3.3	MTGRQRTTVVAPDRPVQDATISQLTTRVWTVAIDGYRTIVVEG ETGIVAINSFGTPSAQTKYRELITQTFGDKPVVAVVASIDHLDHT GRLGPFANGAEVIGHELGQAIAFGRGLPEQKLADTVVTGPVTEI ERAGVRLVLRYPAPTVGTGNLAVDLPDDDVVFMVGLQSGARY GIFPDFHFKHFLRATSEIAALGRRYFVPGRSEVMDAGQVRQAL EYVNDFQNACQRCLAGGEVPHWLLEPTTAYLHDELSSKWSHL EGYDPVAVGLGGLRVVCHYYMGGWWLDDTDHHELLYDHLTV RTYREYRERLATAGTGRA
cytosol non-specific dipeptidase	3.4	MSELSQLSPQPLWDIFAKICSIPHPSYHEEQLAEYIVGWAKEKG FHVERDQVGNILIRKPATAGMENRKPVVLQAHLDMVPQKNNDT VHDFTKDPIQPYIDGEWVKARGTTLGADNGIGMASALAVLADE NVVHGPLEVLLTMTEEAGMDGAFGLQGNWLQADILINTDSEEE GEIYMGCAGGIDFTSNLHLDREAVPAGFETFKLTLKGLKGGHS GGEIHVGLGNANKLLVRFLAGHAEELDLRLIDFNGGTLRNAIPR EAFATIAVAADKVDVLKSLVNTYQEILKNELAEKEKNLALLLDSV ANDKAALIAKSRDTFIRLLNATPNGVIRNSDVAKGVVETSLNVG VVTMTDNNVEIHCLIRSLIDSGKDYVVSMLDSLGKLAGAKTEAK GAYPGWQPDANSPVMHLVRETYQRLFNKTPNIQIIHAGLECGL FKKPYPEMDMVSIGPTITGPHSPDEQVHIESVGHYWTLLTELLK EIPAK

Name	EC class	Sequence
acesulfame hydrolase	3.5	MTSIDEITRKPAQVLAGLIRERELSCLEVTSAFLKRIDDINPKINA FCTVLHEAALAAAAQADQAFTSGSPIGPLHGLPVALKDLTPTKG VRTTRGSRLFENAVPAEDAELVRRLKKAGAIVIGKTNTPEFGHK GETDNLIFGPTRNPWRLDRTPGGSSGGSAAAVAAGLVPFAEG SDGAGSIRIPASMCGIFGFKPSYGRVPDVAGPFSSHTPFFHNG PLARSVGDATLLYQAMVGADSADPFSVPTDQDVLMSLDHGVA GLRVAFSVNLGYFEVSDEVKLACTRATEAFAALGCVVDEVEVD FDRELEAAFFTLWCAKLATVYSNITDSEFSLLEPVVQGLIEQGR RLSAVEFGRANLMREVVWSRLCSIFDKYDVLICPTTAVSAFPIE NGPPATINGASINRLLGWFLTYPFNFTGNPAASVPCGFSHDGL PIGMQIIGRRLDDGLVLRASRTFERLSPWPKLANPSF
paracetamol amidase TccA	3.5	MTATDELYYLPVTELSELIAERRLAPSELMSAVIARAEEVNPKLN ALVAQRFEAATREAAAADNEPSRGVLHGIPITLKDLAYETPDLP STYGSRAFAGYEPGFETVVGQRLRAAGTIAIGRTNSPEFGLTN TCESAQFGPTSNPWRPEHTPGGSSGGAGAAVAAGIAPLAAAN DGGGSCRVPASSCGVVGLKPSRGRVPWAPTSYEYWAGFATN GPIARTVEDVALLLDAMSGPVVGEPYGLPAPSESFLTASRRRP GPLRIAFSCTPPKPHDRLNAEVKQTFLAAVANFEALGHTVTEID HGLDGIFDSFIRVIAANTALSVTQTVPLGSLNLLEPNTLGLAQRG WGLSAMDYCEAINHLRTTAALSMARWTEDFDVLLTPTLTDLPP LTGQMPSYDGDLDACYLHMLGHNAFTYPFNVTGQPALSIPCG WSTSGLPIGLQIIGGMGQEARVLALAAAYEEAHPWAARKPPL
paracetamol dimethoate hydrolase	3.5	MARTGFYTDERTFWHATGMQALFLPVGDWVQPPNGTAGADT PDSKRRLLNLAHASGLIRKLTLPEAIPATVEDVCRVHPRDYIDRF KATSDAGGGDLGHLAPFSKGGYEIAMLSCGLAIAAVDDVLSGK VDNAYALCRPAGHHCLADTPMGFCLLANIPIAIEAAKARHGISR VAVVDWDVHHGNGTQSIFYDRADVLTISIHQDRCFPPGYSGAE DRGEGAGLGYNLNVPLPAGAGHDAYVQAFDDIVVPALDDFKP DLIIVASGLDANSVDPLARMLLHSESYRLLTQKMLDAAARLCGG KLVVVHEGGYAEAYVPFCGHALLEALSGERTAVVDPVLEMAEA WQPGPEAAAFHRQWIDRLVADLGK

Supplementary methods references:

[1] L. M. Biehl, F. Farowski, C. Hilpert, A. Nowag, A. Kretzschmar, N. Jazmati, A. Tsakmaklis, I. Wieters, Y. Khodamoradi, H. Wisplinghoff, M. J. G. T. Vehreschild, *PLoS One* 2022, 17, e0262095.

[2] M. D. Robinson, A. Oshlack, *Genome Biol*. 2010, 11, 1. DOI: 10.1186/gb-2010-11-3-r25.

[3] M. B. Pereira, M. Wallroth, V. Jonsson, E. Kristiansson, *BMC Genomics* 2018, 19, 1.

[4] J. Wicker, K. Fenner, S. Kramer, A Hybrid Machine Learning and Knowledge Based Approach to Limit Combinatorial Explosion in Biodegradation Prediction, *Computational Sustainability*, 2016, DOI: 10.1007/978-3-319-31858-5_5.

[5] J. Wicker, T. Lorsbach, M. Gütlein, E. Schmid, D. Latino, S. Kramer, K. Fenner, *Nucleic Acids Res.* 2016, 44, D502.

[6] Classifier Chains for Multi-Label Classification, J. Read, B. Pfahringer, G. Holmes, E. Frank, 2011, DOI: 10.1007/s10994-011-5256-5.

[7] E. Schmid, K. Fenner, *bioRxiv*, DOI: 10.1101/2021.05.20.442588.

[8] A. K. Sharma, S. K. Jaiswal, N. Chaudhary, V. K. Sharma, *Sci. Rep.* 2017, 7, 9751. DOI: 10.1038/s41598-017-10203-6.

[9] S. Poux, C. N. Arighi, M. Magrane, A. Bateman, C.-H. Wei, Z. Lu, E. Boutet, H. Bye-A-Jee, M. L. Famiglietti, B. Roechert, T. UniProt Consortium, *Bioinformatics* 2017, 33, 3454.

[10] S. Firke, janitor: Simple Tools for Examining and Cleaning Dirty Data, https://CRAN.R-project.org/package=janitor, accessed 2022.

[11] J. Brunson, J. Open Source Softw. 2020, 5, 2017.

[12] ggnewscale: Multiple Fill and Colour Scales in “ggplot2,” https://CRAN.R-project.org/package=ggnewscale, accessed 2022.

[13] ggpattern: 'ggplot2' Pattern Geoms, https://CRAN.R-project.org/package=ggpattern, accessed 2022.

[14] ggpubr: “ggplot2” Based Publication Ready Plots, https://CRAN.R-project.org/package=ggpubr, accessed 2022.

[15] pacman: Package Management for R, http://github.com/trinker/pacman, accessed 2022.

[16] pBrackets: Plot Brackets, https://CRAN.R-project.org/package=pBrackets, accessed March 24, 2023.

[17] P. J. McMurdie, S. Holmes, *PLoS One* 2013, 8, e61217.

[18] RColorBrewer: ColorBrewer Palettes, https://CRAN.R-project.org/package=RColorBrewer, accessed 2022.

[19] readxl: Read Excel Files, https://CRAN.R-project.org/package=readxl, accessed 2022.

[20] H. Wickham, M. Averick, J. Bryan, W. Chang, L. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. Pedersen, E. Miller, S. Bache, K. Müller, J. Ooms, D. Robinson, D. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, H. Yutani, J. *Open Source Softw.* 2019, 4, 1686.

[21] K. Thomas-White, S. C. Forster, N. Kumar, M. Van Kuiken, C. Putonti, M. D. Stares, E. E. Hilt, T. K. Price, A. J. Wolfe, T. D. Lawley, *Nat. Commun*. 2018, 9, 1557. DOI: 10.1038/s41467-018-03968-5.

[22] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J. Mol. Biol.* 1990, 215, 403.

[23] W. R. Pearson, *Curr. Protoc. Bioinformatics* 2013, Chapter 3, Unit 3.1.

[24] N. Koppel, V. Maini Rekdal, E. P. Balskus, *Science* 2017, 356, DOI: 10.1126/science.aag2770.

[25] M. L. Bonatelli, T. Rohwerder, D. Popp, Y. Liu, C. Akay, C. Schultz, K.-P. Liao, C. Ding, T. Reemtsma, L. Adrian, S. Kleinsteuber, *bioRxiv*, 2022, DOI: 10.1101/2022.08.17.504299.

[26] B. A. Magnuson, M. C. Carakostas, N. H. Moore, S. P. Poulos, A. G. Renwick, *Nutr. Rev.* 2016, 74, 670.

[27] C. Willson, *Toxicol Rep* 2018, 5, 1140.

[28] I. Köpping, C. S. McArdell, E. Borowska, M. A. Böhler, K. M. Udert, *Water Res X* 2020, 9, 100057.

[29] J. Lienert, T. Bürki, B. I. Escher, *Water Sci. Technol.* 2007, 56, 87.

[30] J. A. Forrest, J. A. Clements, L. F. Prescott, *Clin. Pharmacokinet.* 1982, 7, 93.