Supplementary Information: Identification of persistent substructures in transformation products with zebrafish embryos using cheminformatics and a suspect screening approach

Parviel Chirsir*[1], Maria Lorena Cordero-Maldonado[1], Maxim P. Carlier[2], Timo Hamers[2], Emma L. Schymanski*[1]

[1] Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 6 avenue du Swing, 4367, Belvaux, Luxembourg

[2]Amsterdam Institute for Life and Environment, Section Environment and Health, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands

* Corresponding authors: P.C. parviel.chirsir@uni.lu, and E.L.S. emma.schymanski@uni.lu

ORCIDs: PC: 0000-0002-9932-8609, MLCM: 0000-0001-5631-0022, MPC: 0000-0001-5516-1846, TH: 0000-0002-5733-3290, ELS: 0000-0001-6868-8145

Table of Contents

List of Figures

List of Abbreviations

Abbreviations	Full Meaning
ADONA	3H-Perfluoro-4,8-dioxanonanoic acid
AI	Artificial Intelligence
AMMT	2-Amino-4-methoxy-6-methyl-1,3,5-triazine
CID	Chemical Identifier
DMT-DES	Desmethylthio-desethyl-simetryn or 2-N-ethyl-1,3,5-triazine-2,4-diamine
ECHA	European Chemical Agency
ESI	Electronic Supplementary Information
FA	Formic Acid
FAIR	Findable, Accessible, Interoperable, and Reusable
HMMM	Hexa-methoxy-methyl-melamine
hpf	Hours post fertilisation
K_{OC}	Organic-carbon-water partition coefficient
LC_{50}	Lethal concentration 50
LC-HRMS	Liquid chromatography coupled with high-resolution mass spectrometry
LCSB	Luxembourg Centre for Systems Biomedicine
LogK_{OC}	Logarithm of the organic-carbon-water partition coefficient
MeOH	Methanol
Milli-Q H_2O	Ultrapure water (Type 1)
ML	Machine Learning
NORMAN-SLE	NORMAN Suspect List Exchange
P	Persistence
PCID	PubChem Chemical Identifier
PFAS	Per- and poly-fluoroalkyl substances
PFBS	Perfluorobutane sulfonic acid
PFOA	Perfluorooctanoic acid
PMT	Persistent, Mobile and Toxic
PMT/vPvM	Persistent, Mobile and Toxic or very Persistent and very Mobile
QSAR	Quantitative Structure-Activity Relationship
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals
RT	retention time
SVHC	Substances of Very High Concern
T	Toxicity
TFA	Trifluoroacetic acid
TPs	Transformation products
vM	very Mobile
vP	very Persistent
vPvM	very Persistent and very Mobile
VUA	Vrije Universiteit Amsterdam
ZFPMTTPs	Zebrafish Persistent Mobile and Toxic Transformation Products
ZPMS90TPs	ZeroPM S90 Transformation Products

S1: Plate design for exposure experiments

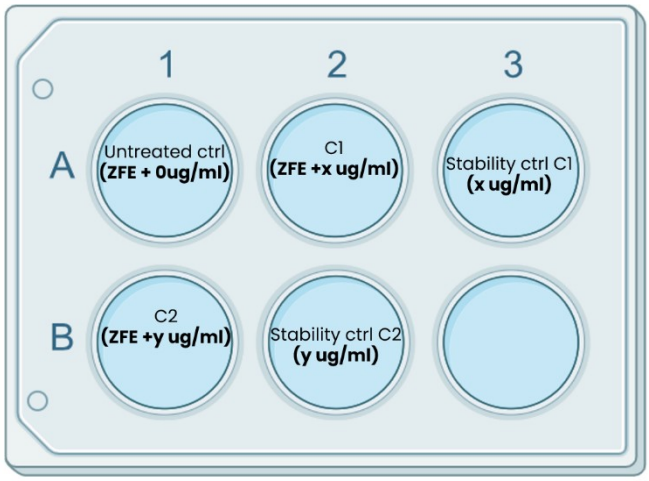

Figure S1: Plate design for the exposure experiment with zebrafish embryo(ZFE). Ctrl = control, C1 = treatment condition 1, x = lower concentration of compound, y = higher concentration of compound, C2 = treatment condition 2. (Partly created in BioRender, https://BioRender.com)

S2: Details of the sample extraction

The sample preparation and extraction protocols were adapted from the LCSB Metabolomics Platform (RRID: SCR_024769) protocols previously reported by Heins-Marroquin et al [46] and Talavera Andújar et al [47]. The same extraction and measurement protocol was used for both of the above-listed experiments. Before extraction, samples were briefly thawed on ice. For extraction, 600 mg of ceramic beads (size 1.4 mm; bead for Precellys homogenisers, QIAGEN, 13113-325) and 1000 µL of extraction fluid containing methanol (MeOH)/Milli-Q H_2O/internal standard (15:3:0.5 ratio, v/v/v) were added to the tubes on ice (4°C). The mixture was homogenised at 6000 rpm for 30 seconds, then 120 µL of Milli-Q H_2O at 4°C and 400 µL of chloroform were added, and the mixture was incubated for 10 minutes at maximum speed. After centrifugation at 15,000 rpm for 5 minutes, 760 µL of the supernatant was transferred to a 1.5 mL Eppendorf tube, along with 200 µL of Milli-Q H_2O and 200 µL of chloroform. This mixture was vortexed and centrifuged again for 5 minutes. The lower phase (300 µL) was collected, evaporated to dryness using a SpeedVac, and then reconstituted with 100 µL of 0.1% formic acid (FA) in Milli-Q H_2O and MeOH solution in a 90/10 (v/v) ratio. Finally, samples were filtered through PHENEX PTFE 4mm syringe filters (0.2 µm) into LC vials for measurement. Similarly, 100 µL of stability controls (when used) was extracted. Additional details can be found in the SI Section 2. Although this extraction technique has not been optimised for PFAS recovery, it was used in this study for consistency with other compounds. For quality control, two extraction blanks were made by extracting 100 µL of Milli-Q H_2O, and two pooled quality samples were made by aliquoting 5 µL of each sample into LC vials. These samples were injected first, then after every five samples and again at the end of the measurement, as recommended by Broadhurst et al. [48]. In addition, Milli-Q H_2O blanks were prepared using 100 µL of Milli-Q H_2O to test system stability and measured alongside the samples.

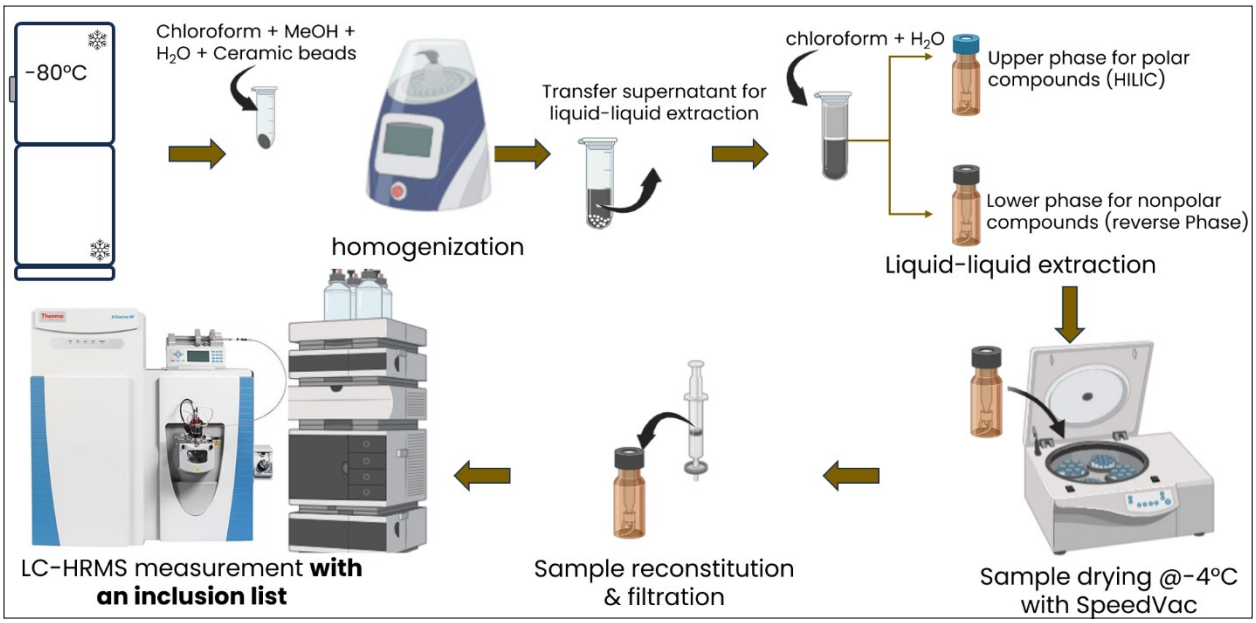

Figure S2: Sample extraction process. LLE with chloroform and water protocol from the LCSB metabolomic platform, previously reported by Heins-Marroquin et al and from Talavera et al. [1, 2]

S3: Generation of analogous TPs from literature

The analogous TPs were generated by first comparing compounds structurally similar to the compound of interest (i.e., compounds structurally similar to Simetryn, Chlorsulfuron, Metconazole, Bisphenol AF, and ADONA). These structurally similar compounds can be seen in SI Table 2. The SMILES of the compounds of interest (e.g., Simetryn) is used to search for structurally similar compounds from the PubChem database. Filtering of the results is done by filtering for compounds with Pharmacology and Biochemistry content in PubChem (this is the contents section in which transformation products are found), Figure S3. The transformation products were then retrieved from the PubChem transformations section, Figure S4A. Then ACD/ChemSketch (Freeware) 2021.2.1[3] is used to draw analogous TPs to those of structurally similar compounds, Figure S4B. For example, to generate the TPs of Simetryn, terbutryn, a structurally similar compound to simetryn, with TPs such as Hydroxy-dethiomethyl-terbutryn, Deethylterbutryne, and t-Butylhydroxy-terbutryn in the PubChem database is used. The transformation of terbutryn to Hydroxy-dethiomethyl-terbutryn, Deethylterbutryne, and t-Butylhydroxy-terbutryn a thioether substitution, N-dealkylation and oxidation, respectively. A similar transformation is then applied to the Simetryn Hydroxyethyl-simetryn, Desethyl-simetryn, and Hydroxy-simetryn. See Figure S3Figure S4 for more details.

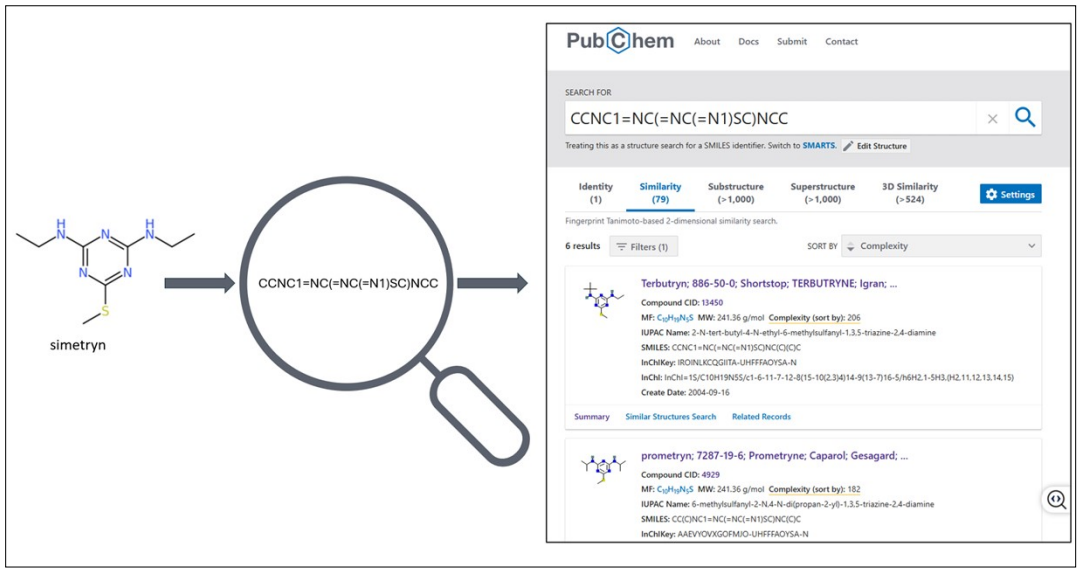

Figure S3: Expert curation of analogous TP: Step 1, searching for structurally similar compounds with documented TPs in PubChem database.

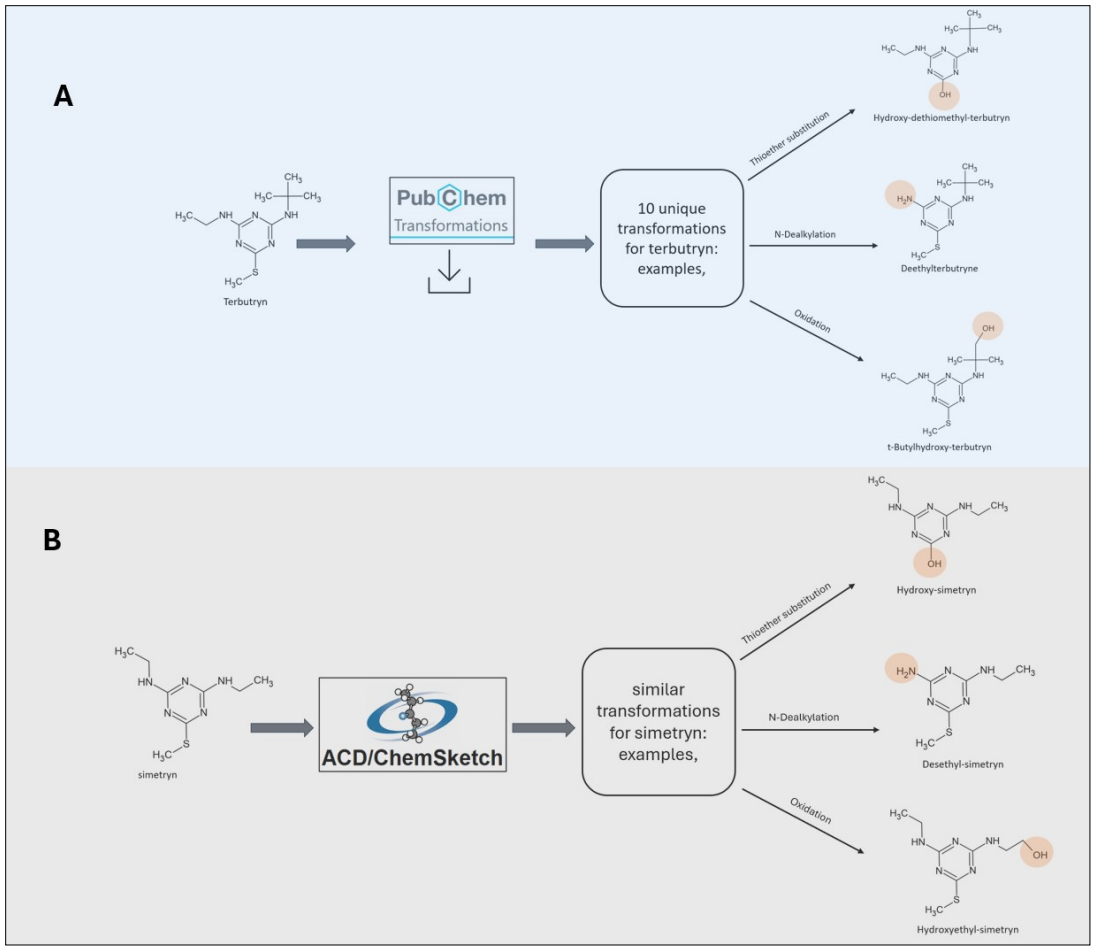

Figure S4: Expert curation of analogous TP: Step 2 (A), obtaining TPs of structurally similar compounds from PubChem Transformation, Step 2 (B), drawing the analogous TP ChemSketch.

S4: Validation of the suspect screening results

For extra validation of the suspect screening results, each relevant feature from the *featureGroupScreening.csv* file from patRoon is double-checked in the .Raw file is explored using the Qual Browser app in the Xcalibur software. The *mz* value of the feature is used to check for peak shape, peak intensity, RT and spectral data acquisition Figure S5. A good peak is a symmetrical or Gaussian peak, while the peak intensity threshold is set at 1.0×10^5 for a good peak intensity. The RT should match the RT where the MSMS or MS^2 spectral data is retrieved. The spectral data is then performed *in silico* fragmentation with the MetFrag web application [4], using the custom databases (ZFPMTTPs and ZPMS90TPs) as a dataset or annotation Figure S6. Results from MetFrag, including the MoNA scores and number of *in silico* peak matched, are used to assign identification confidence levels based on criteria similar to those used in the patRoon identification system (Figure S7) [5, 6]. However, there are specific distinctions for levels 3b and 3c. For level 3b, a good in-silico MSMS match occurs when more than three experimental fragments correspond with MetFrag fragments. In level 3c, the suspect mass, formula, and tentative structure are identified, but no MSMS fragments are available.

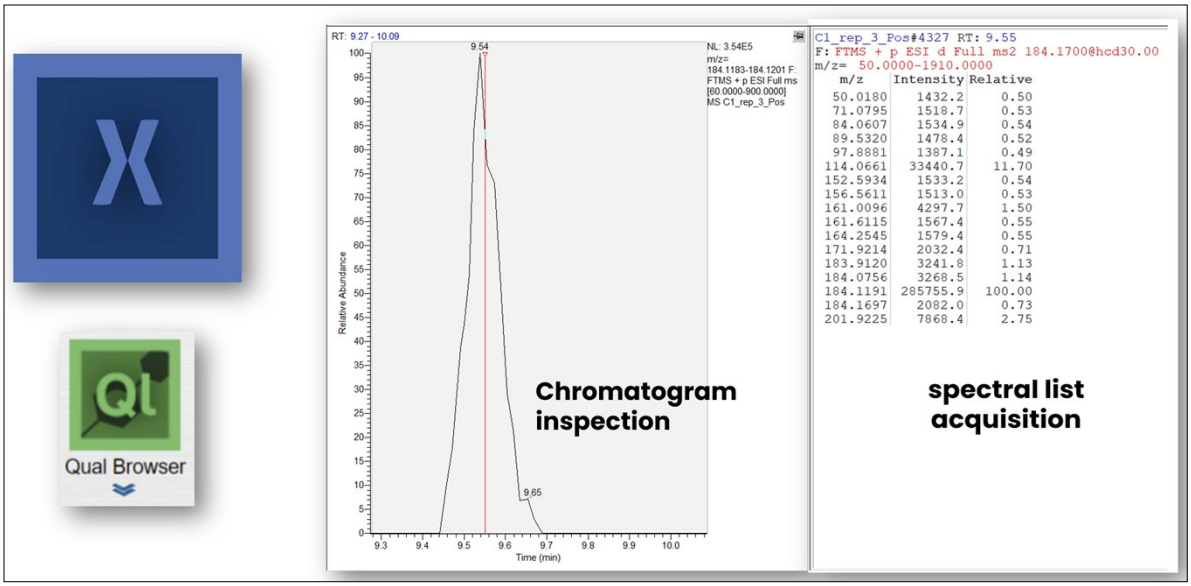

Figure S5: Step1, in data validation, chromatographic peak inspection and spectral list acquisition

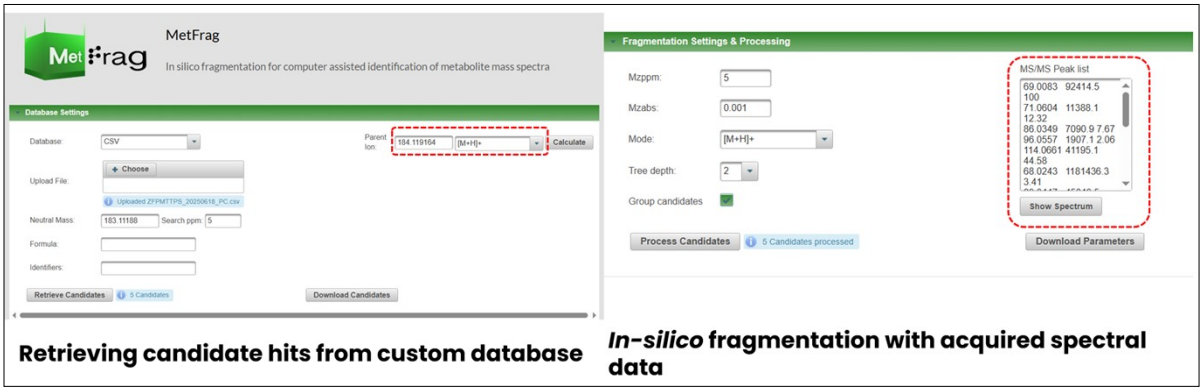

Figure S6: Step 2, MetFrag in silico fragmentation and annotation with the custom database (ZFPMTTPs or ZPMS90TPs)

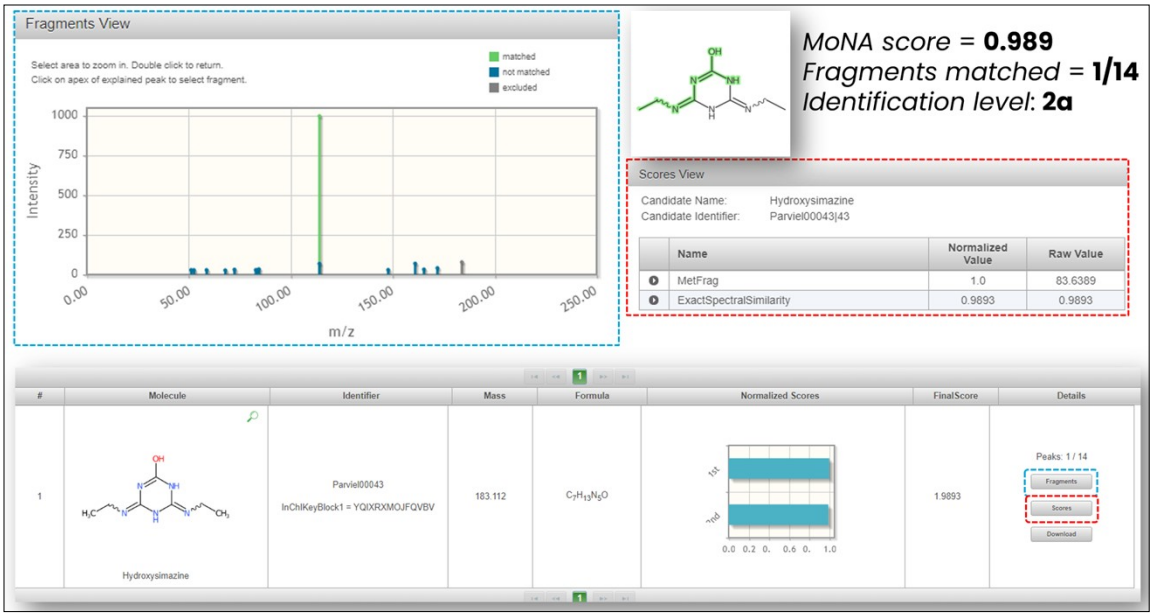

Figure S7 Step 3, results from MetFrag showing the peaks, number of in silico peak matched and MoNA score.

S5: Persistent substructure characterisation for 36 PM compounds

S5.1. Triazines

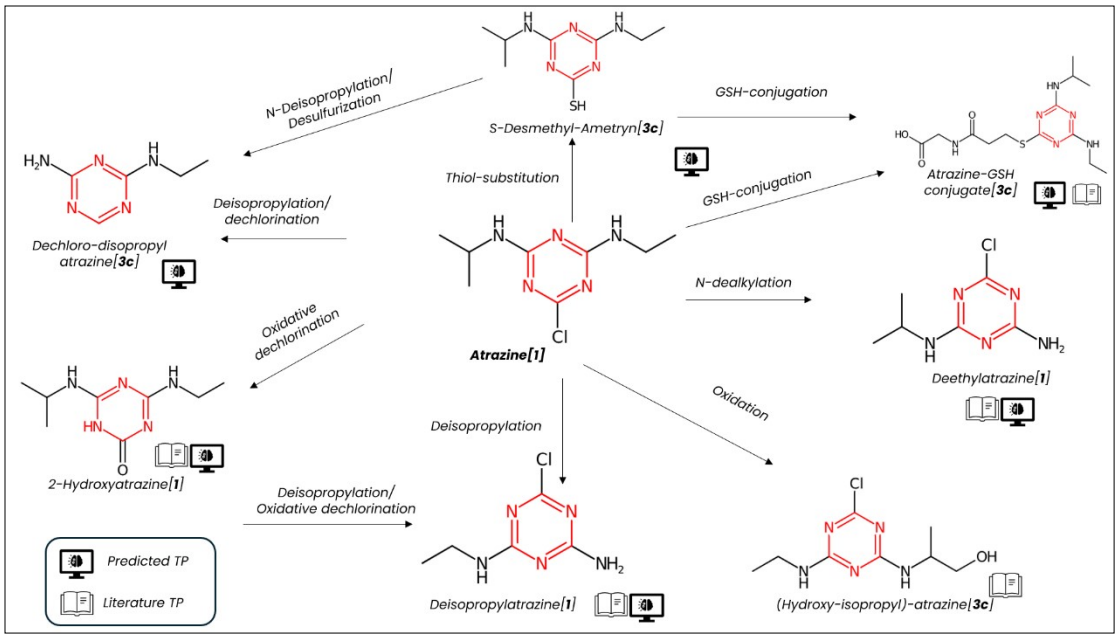

Figure S8: Proposed transformation pathway for atrazine, highlighting the persistent substructure (1,3,5-triazine ring) in red.

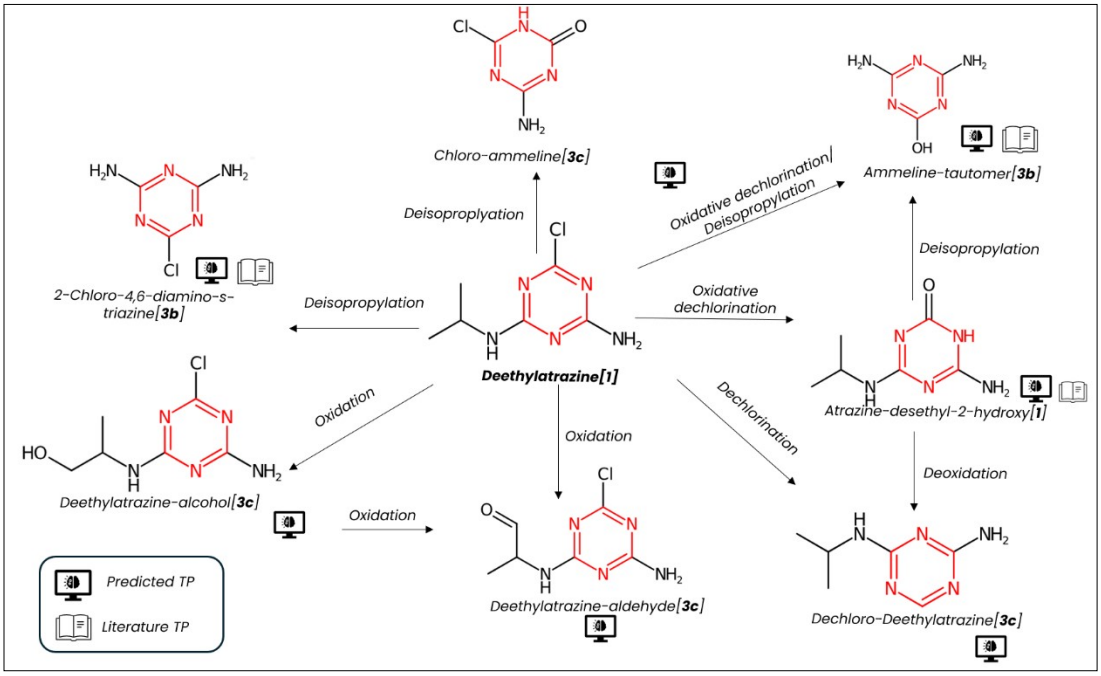

Figure S9: Proposed transformation pathway for deethylatrazine, highlighting the persistent substructure (1,3,5-triazine ring) in red.

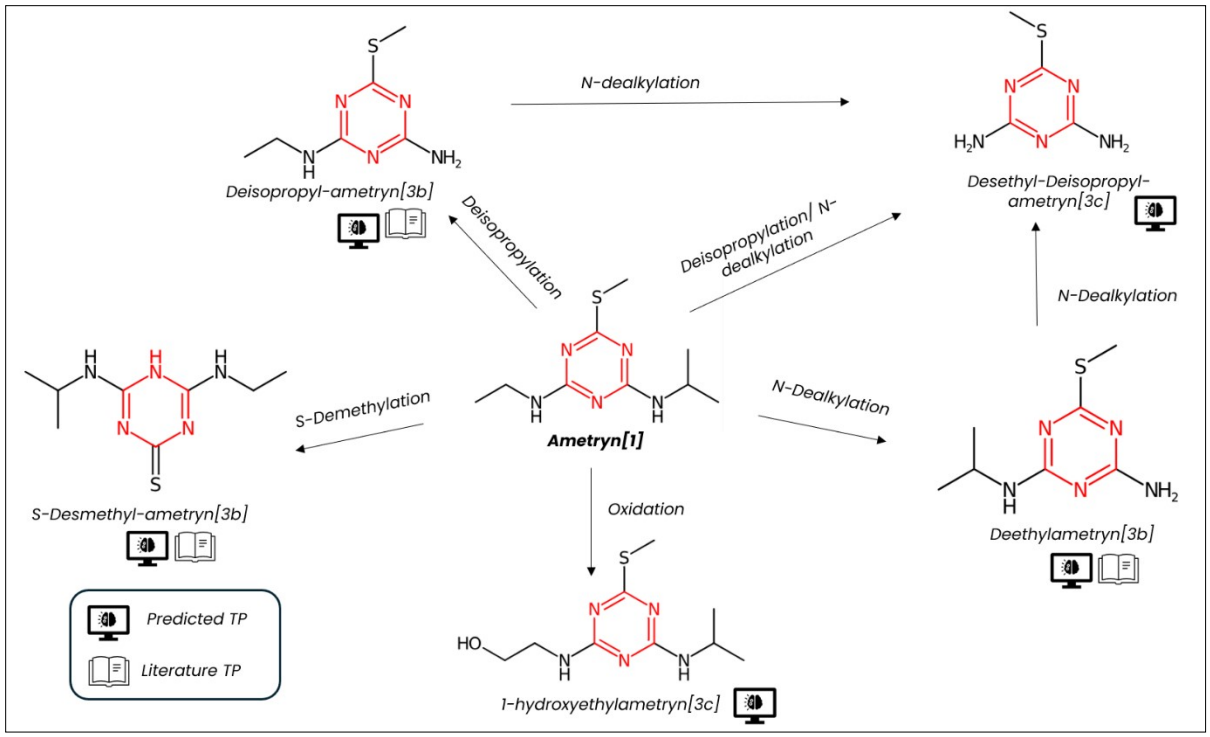

Figure S10: Proposed transformation pathway for Ametryn, highlighting the persistent substructure (1,3,5-triazine ring) in red.

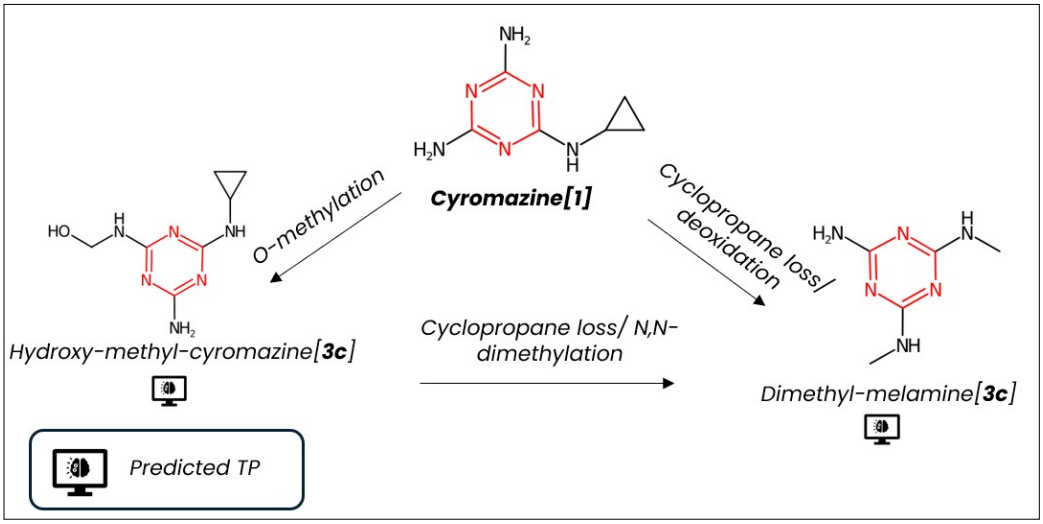

Figure S11: Proposed transformation pathway for cyromazine, highlighting the persistent substructure (1,3,5-triazine ring) in red.

S5.2. Triazoles

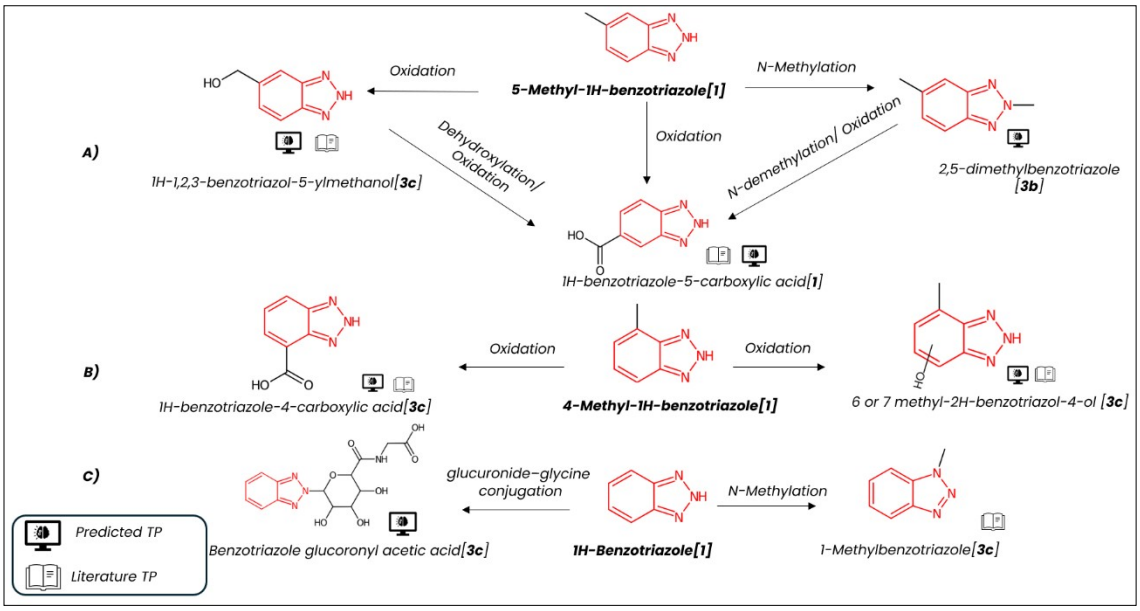

Figure S12: Proposed transformation pathway for A)5-methyl-1H-benzotriazole, B) 4-methyl-1H-Benzotriazole & C) 1H-benzotriazole, highlighting the persistent substructure (benzotriazole molecule) in red

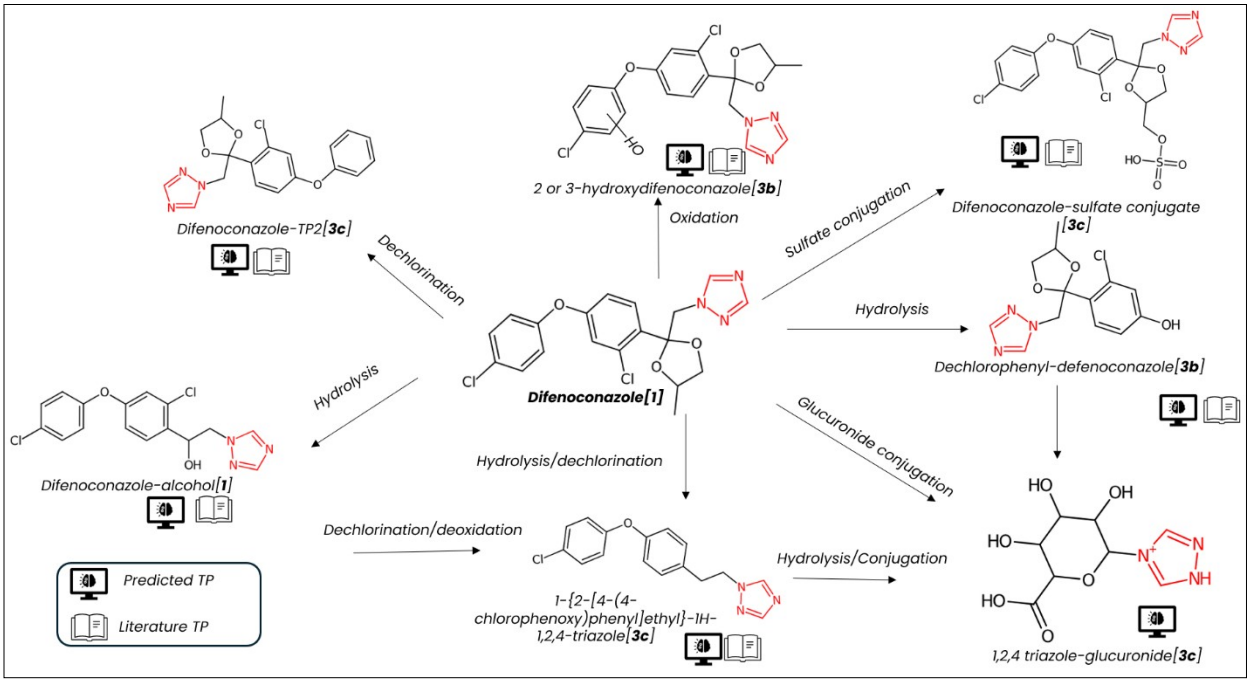

Figure S13: Proposed transformation pathway for difenoconazole, highlighting the persistent substructure (1,2,4-triazole) in red

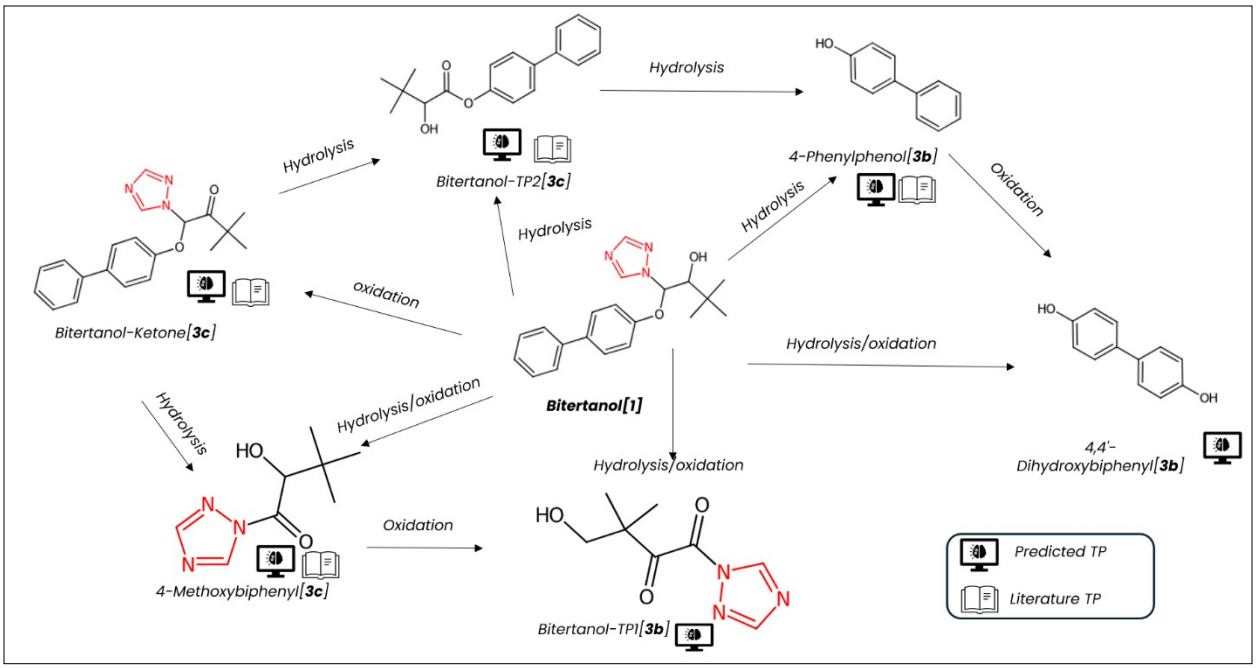

Figure S14: Proposed transformation pathway for bitertanol, highlighting the persistent substructure (1,2,4-triazole) in red.

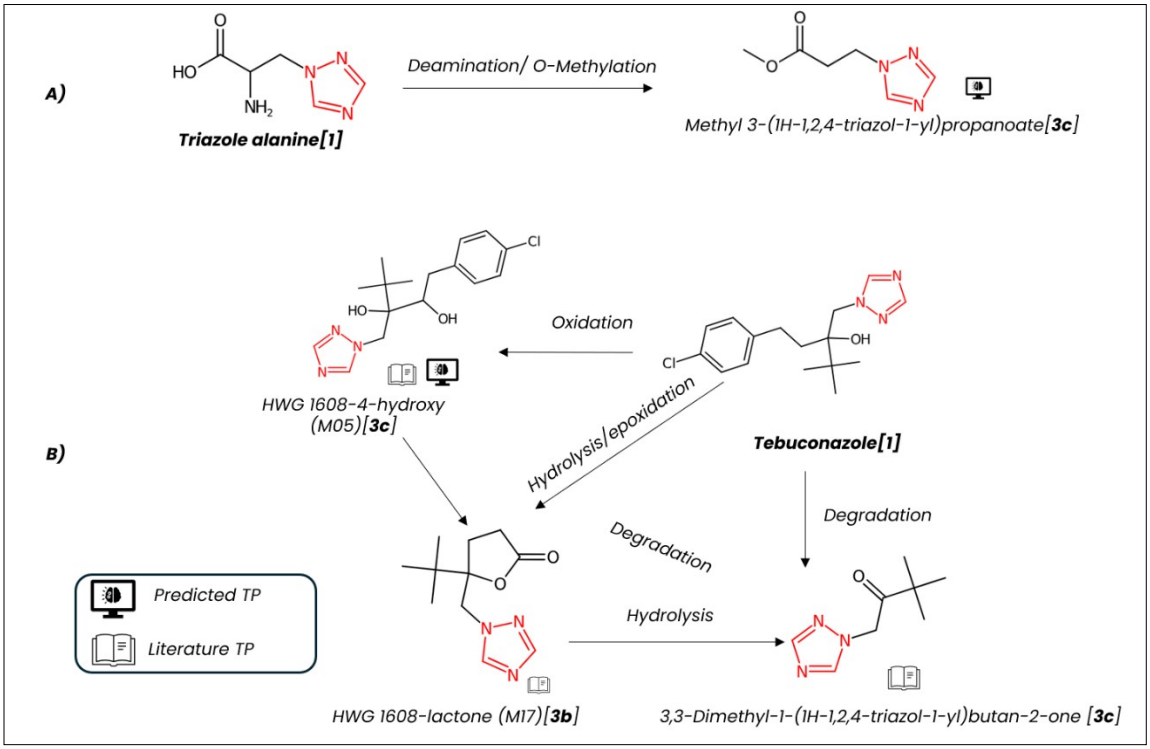

Figure S15: Proposed transformation pathway for A) triazole alanine & B) tebuconazole, highlighting the persistent substructure (1,2,4-triazole) in red.

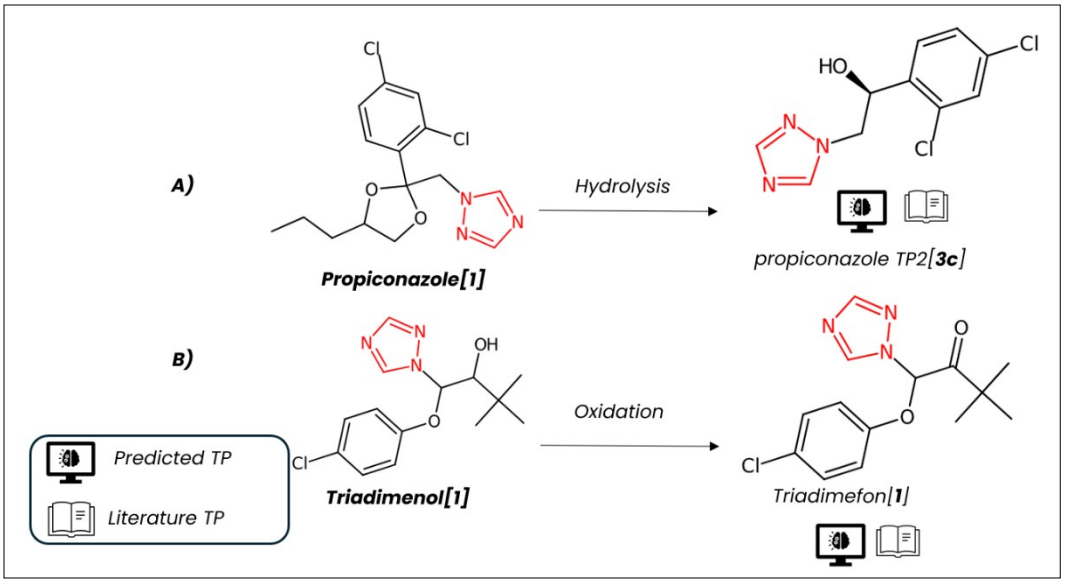

Figure S16: Proposed transformation pathway for A) propiconazole & B) triadimenol, highlighting the persistent substructure (1,2,4-triazole) in red.

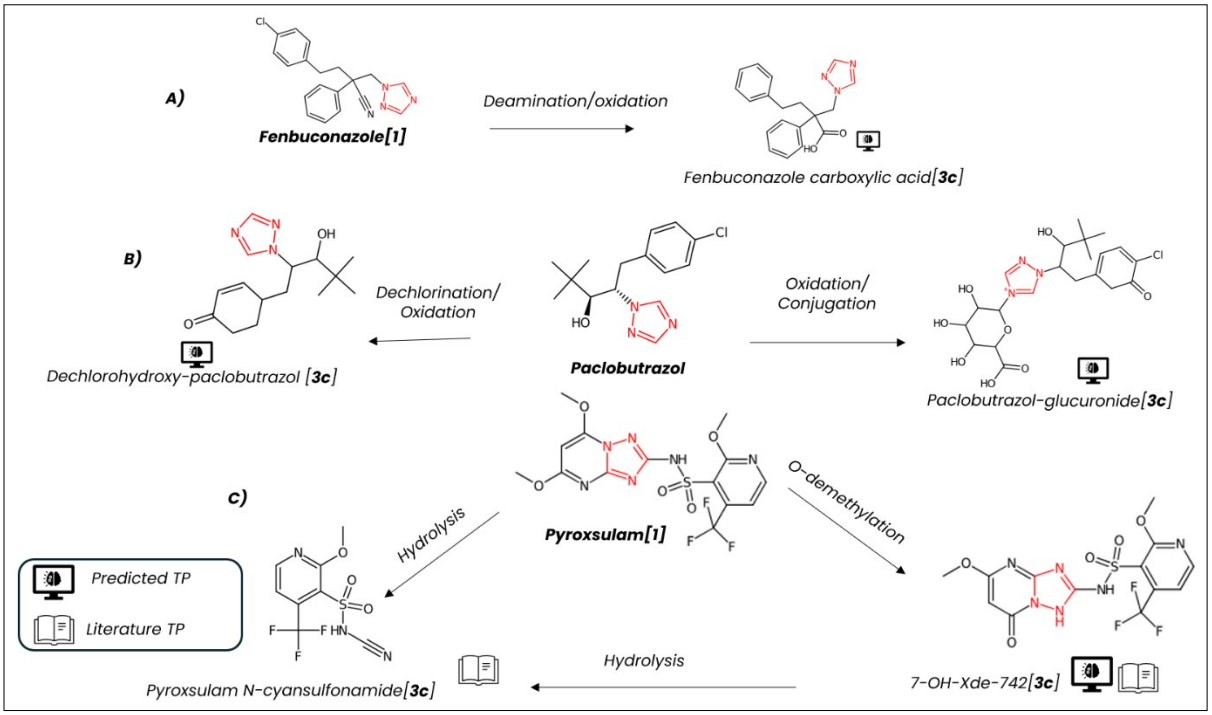

Figure S17: proposed transformation pathway for A) fenbuconazole, B) paclobutrazol and C) Pyroxsulam, highlighting the persistent substructure (1,2,4-triazole) in red.

S5.3. PFAS

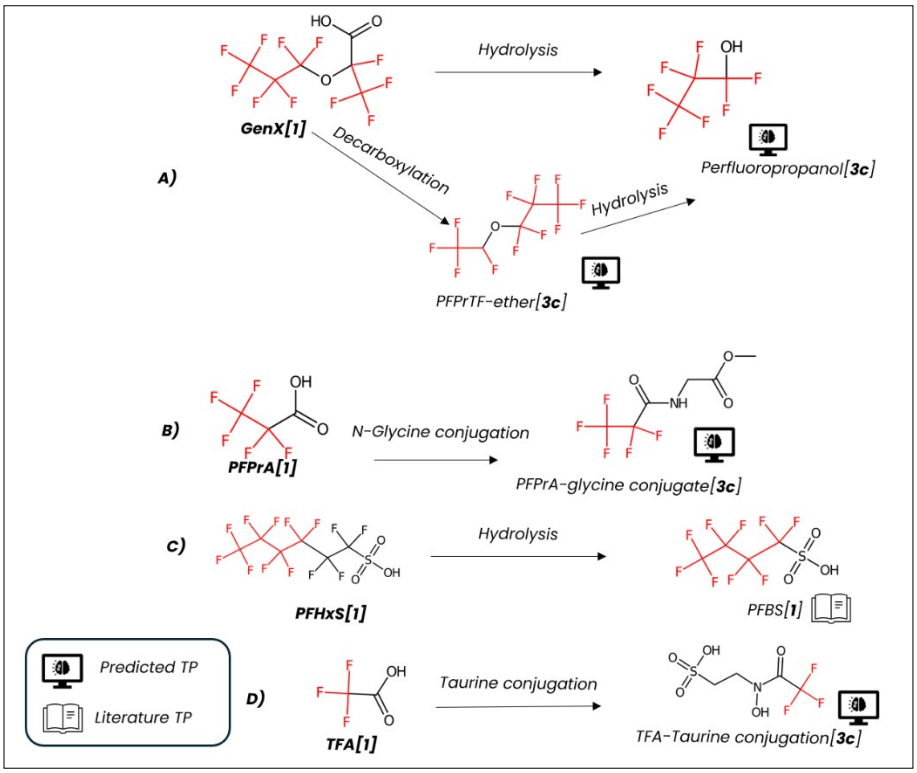

Figure S18: Proposed transformation pathway for A) GenX, B) PFPrA, C) PFHxS & D) TFA.

References

1. Talavera Andújar B, Aurich D, Aho VTE, et al (2022) Studying the Parkinson's disease metabolome and exposome in biological samples through different analytical and cheminformatics approaches: a pilot study. Anal Bioanal Chem. https://doi.org/10.1007/s00216-022-04207-z

2. Heins-Marroquin U, Singh RR, Perathoner S, et al (2024) CLN3 deficiency leads to neurological and metabolic perturbations during early development. Life Sci Alliance 7:e202302057. https://doi.org/10.26508/lsa.202302057

3. ACD/Labs (2025) ChemSketch Freeware. ACD/Labs. https://www.acdlabs.com/resources/free-chemistry-software-apps/chemsketch-freeware/

4. Ruttkies C, Schymanski EL, Wolf S, et al (2016) MetFrag relaunched: incorporating strategies beyond in silico fragmentation. J Cheminform 8:3. https://doi.org/10.1186/s13321-016-0115-9

5. Helmus R, van de Velde B, Brunner AM, et al (2022) patRoon 2.0: Improved non-target analysis workflowsincluding automated transformation product screening. JOSS 7:4029. https://doi.org/10.21105/joss.04029

6. Schymanski EL, Jeon J, Gulde R, et al (2014) Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. Environ Sci Technol 48:2097–2098. https://doi.org/10.1021/es5002105