Detecting pharmaceuticals and their transformation products

Martin Hoffmann and Franziska Hufsky, Bright Giant GmbH, Jena, Germany

The presence of pharmaceuticals and their transformation products in water sources is an emerging environmental concern, posing potential risks to ecosystems and human health. Traditional targeted analysis focuses on known compounds but overlooks transformation products that may exhibit higher persistence or bioactivity than the original drug.

Non-targeted screening using high-resolution mass spectrometry has emerged as a powerful tool for detecting both known and unknown contaminants, unlike targeted approaches that only screen for a predefined list of suspects. Further expanding analytical strategies to include transformation products enables a deeper understanding of their environmental fate and behaviour, which is essential for developing effective mitigation strategies to safeguard public health. This approach is relevant not only for pharmaceuticals but also for pesticides and industrial chemicals, whose degradation products may have significant environmental impacts.

Here, we demonstrate how SIRIUS annotates pharmaceuticals in Luxembourgish rivers, from precursor drug screening in structure databases and spectral libraries to transformation product screening using a custom-generated database.

Experimental Setup

The analysed dataset was obtained from Singh et al. (2021). Surface water samples were collected from Luxembourgish rivers at 13 different locations. A total of 92 samples were gathered during routine monitoring events between 2019 and 2020.

A non-targeted analysis approach was employed. The water samples were concentrated using solid-phase extraction and subsequently analysed using liquid chromatography coupled to a high-resolution mass spectrometer (LC-HRMS). The LC-HRMS analysis was conducted on a Thermo QExactive HF mass spectrometer equipped with a Waters Acquity UPLC BEH C18 column (1.7 µm, 2.1 × 150 mm) using positive electrospray ionisation and data-dependent acquisition. Note that while only the positive ion mode was used in this analysis, SIRIUS can also process negative ion mode spectra. MS and MS/MS spectra were imported to SIRIUS.

For further details on data acquisition, please refer to Singh et al. (2021).

Data Resources

For our analysis, we utilised the following data resources.

Drug list: A compilation of 816[1] pharmaceutical compounds that have marketing authorisation in Luxembourg from the Ministry of Health and are therefore potentially in use domestically.

Spectral libraries: As spectral libraries we used MassBank[2] and MoNA[3]. Out of the drugs listed, 577 have reference spectra available in these spectral libraries.

Drug database (drug_suspects): A custom structure database for the pharmaceuticals from the drug list. This database contains 772 molecular structures. For stereoisomers, one representative is used; compounds with invalid SMILES notations were discarded.

Transformation product database (drug_TP): We generated transformation products using BioTransformer[4] (Djoumbou-Feunang et al. 2019; Wishart et al. 2022) resulting in over 1.06 million transformation products in total for 713 of the drugs from the drug list. These structures are not unique, as BioTransformer can produce the same transformation product for multiple drugs. The final drug_TP contains 483 203 unique structures, corresponding to 22 804 different molecular formulas.

Custom structure databases and spectral libraries can be imported to SIRIUS via the Databases dialog. All imported spectra will automatically be used for spectral library matching during molecular formula annotation. A spectral library is also a molecular structure database and thus can also be selected for structure database search in SIRIUS.

Methods

All 92 samples were imported to SIRIUS (Dührkop et al. 2019) for analysis. Preprocessing yielded a total of 29 646 features, of which 15 819 were valid for computation, i.e. meeting the criteria of having MS/MS data, not being multiply charged, and not being multimeric. Results were considered for features of all quality levels, ensuring inclusion of all computable data within SIRIUS. Be aware that all reported identifications are putative identifications, as annotation results were not experimentally verified.

Molecular formula annotation

Molecular formulas were assigned using database search. In that case, SIRIUS exclusively considers molecular formulas included in the chosen database(s). Here, we chose SIRIUS's biomolecule structure database[5], along with the imported databases (MassBank and MoNA, drug_suspects, drug_TP).

Structure annotation

The structure annotation workflow in SIRIUS is a non-targeted workflow. SIRIUS identifies the structure of a molecule by predicting its molecular fingerprint and searching for matches in a molecular structure database (Dührkop et al. 2015). Here, we demonstrate how SIRIUS detects precursor drugs as well as transformation products using custom databases (see *Figure 1*). We searched within the biomolecule structure database[5] (part of SIRIUS) as well as the imported drug_suspects and drug_TP databases. PubChem was used as fallback database: If the top hit in PubChem had a confidence score (measure of a hit being correct (Hoffmann et al. 2022)) at least twice as high as the top hit from the selected databases, the search was expanded to include PubChem results.

In addition, SIRIUS automatically performs spectral library searches against all available spectral libraries whenever the molecular formula annotation workflow is applied. We searched in MassBank and MoNA. Spectral matching is based on the cosine score, calculated with squared peak intensities while ignoring the precursor peak. A drug was considered putatively identified if the cosine similarity score of the best spectral match was at least 0.8 and a minimum of three peaks were matched.

Results

Precursor drug screening

A total of 80 precursor drugs from the target drug list were reported as top hits using SIRIUS. Among these, 30 could not be matched through spectral library search[6] (see *Figure 2*). Notably, 17 of these otherwise unidentified compounds were confidently annotated by SIRIUS (confidence score >0.64)[7] with SIRIUS. This highlights SIRIUS's ability to detect compounds that would have been missed by spectral library searches while still integrating the power of spectral library search for suspect identification.

One drug (**Aspirin**, acetylsalicylic acid) was putatively identified through spectral library search but not detected as top hit in SIRIUS. The best SIRIUS match for this feature was **monomethyl phthalate**, which had an similarly high spectral library search score (both cosine similarities ~0.97) but is not included in the drug list. Monomethyl phthalate is a breakdown product of dimethyl phthalate, used in

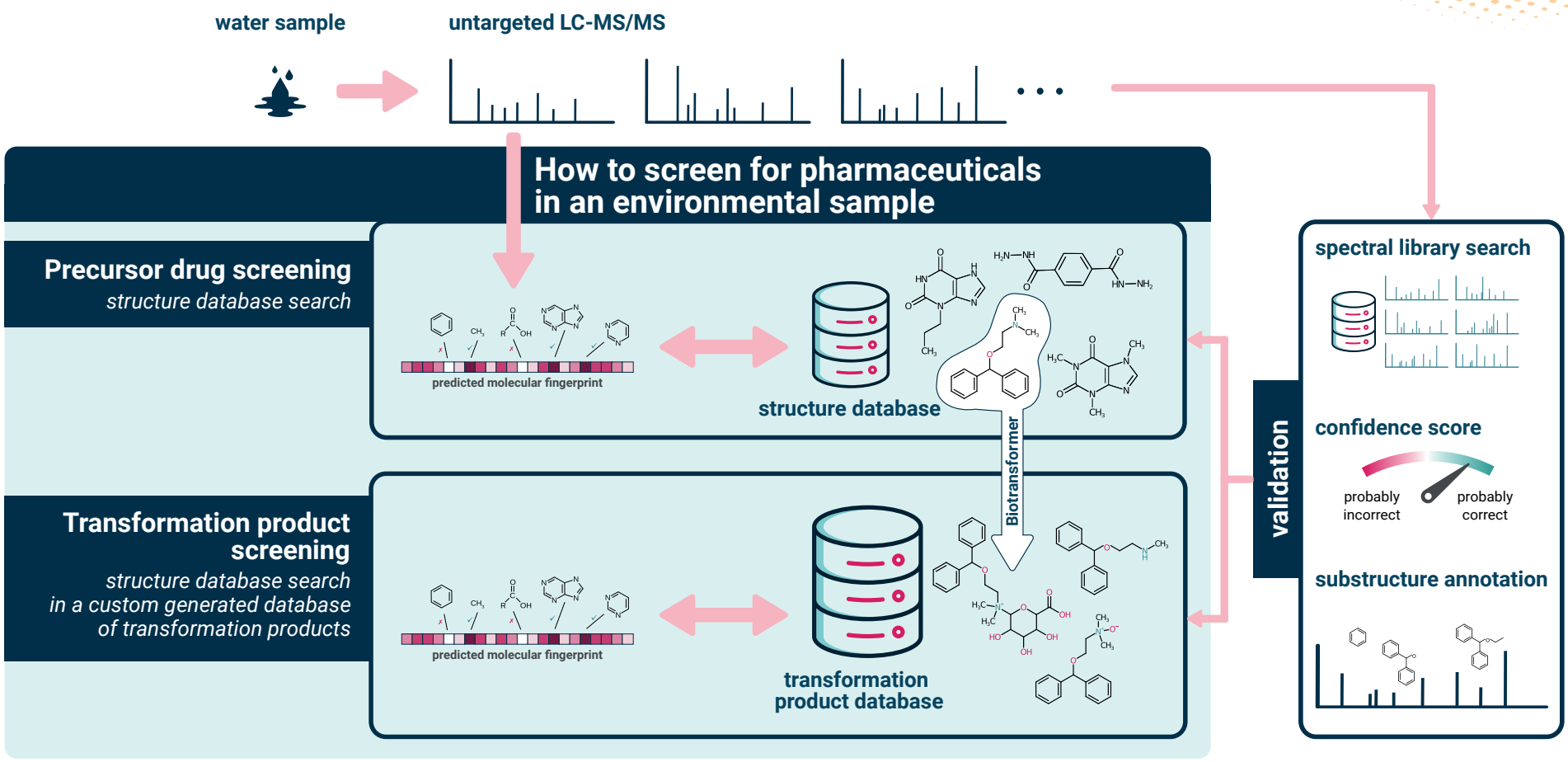

Figure 1: ***How to screen for pharmaceuticals in an environmental sample:*** *We screen for pharmaceuticals in environmental samples on two levels:* ***precursor drug screening*** *and* ***transformation product screening****. For precursor drug screening, we search biomolecule structure databases including a custom structure database of pharmaceuticals. For transformation product screening, we generate transformation products with BioTransformer (Djoumbou-Feunang et al. 2019; Wishart et al. 2022). Results are validated using spectral library search, confidence scores for structure database search hits, and substructure annotation. All methods are part of SIRIUS.*

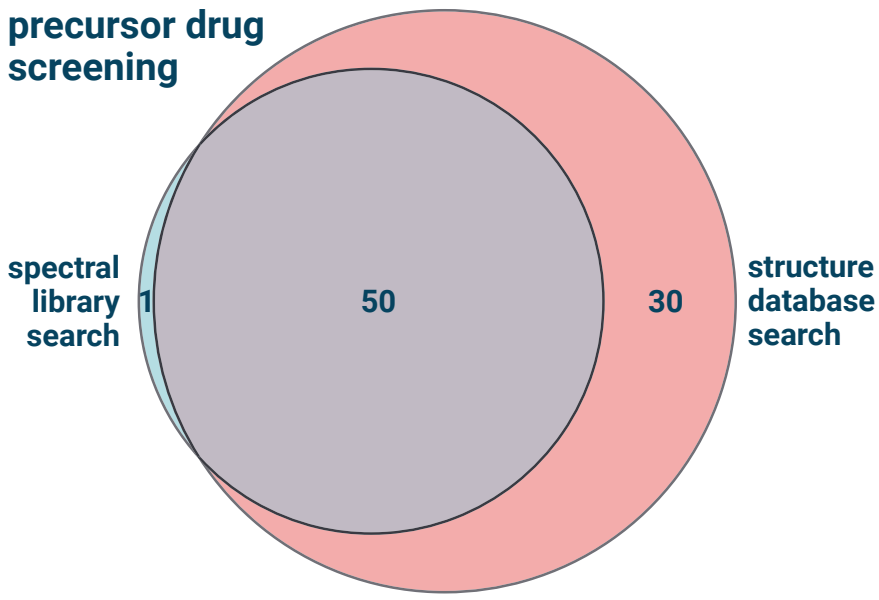

Figure 2: ***Precursor drug screening:*** *Number of pharmaceutical compounds from the drug list putatively identified by* ***spectral library search*** *(left, blue) and by* ***structure database search*** *using SIRIUS (right, pink). 50 compounds have been identified by both methods.*

insect repellent and plastic, justifying the possibility of this alternative putative identification[8]. Further investigation into both results could provide additional clarification (see *Figure 3*). As molecular structure annotation in SIRIUS is based on molecular fingerprinting, structural explanations can be explored within SIRIUS (see *Figure 3*). Moreover, substructure annotations visualise the direct connection to the input MS/MS spectrum (see *Figure 5*). The unambiguous identity of this feature can only be resolved by additional experimental data. The advantage of using SIRIUS lies in its ability to present both possibilities - Aspirin as the best spectral match and monomethyl phthalate as the best structure database match - preventing 'suspect-blindness' and showing that the annotation of Aspirin is not unambiguous.

Transformation product screening

For transformation product screening, we only considered transformation products that were not themselves on the drug list. A total of 292 unique transformation products were detected as top-ranked hits. Of these, only 33 transformation products were also detected through spectral library searches (see *Figure 4*).

In total 85 transformation products were classified as high-confidence identifications[7]. Notably, 56 of these high-confidence hits were not found through spectral library search, highlighting the complementary nature of SIRIUS in detecting transformation products that might otherwise be overlooked.

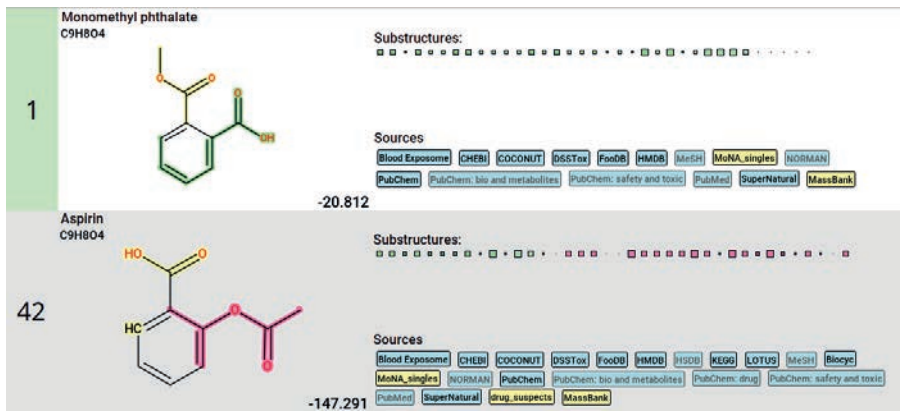

Figure 3: SIRIUS results for ***monomethyl phthalate*** *and* ***Aspirin*** *which have similarly highspectral library search scores for the same feature. The colour highlighting of the candidate structures reflects substructures that are supported by fingerprint evidence (green), substructures that contradict fingerprint evidence (pink), and substructures with mixed support (yellow).*

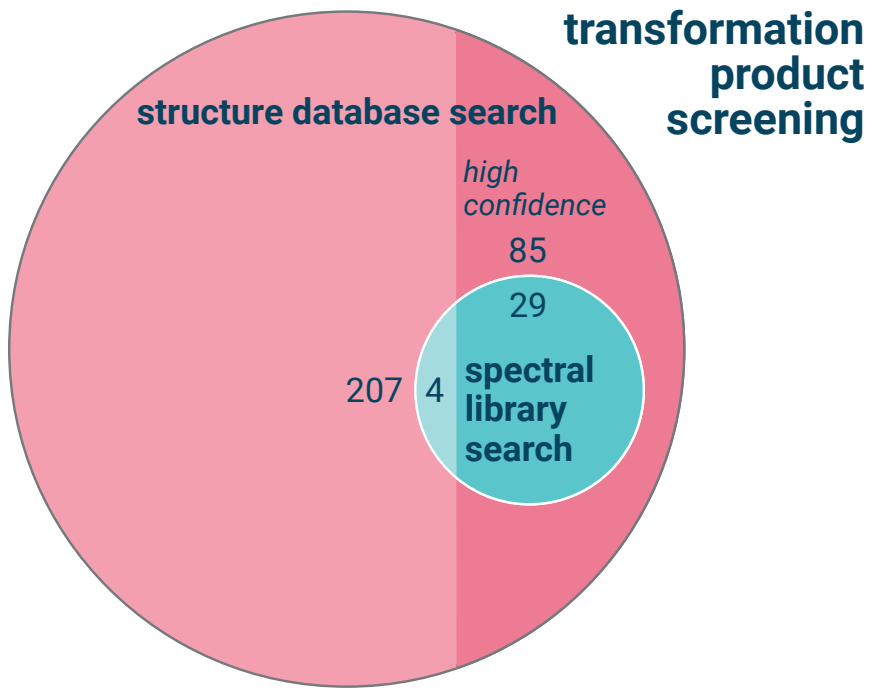

Figure 4: ***Transformation product screening:*** *Number of putatively identified transformation products using* ***spectral library search*** *(blue, 33 in total) and using SIRIUS* ***structure database search*** *(pink, 292 in total) in a custom database of transformation products generated with BioTransformer. The share of compounds displayed in dark pink represents those detected with high confidence.*

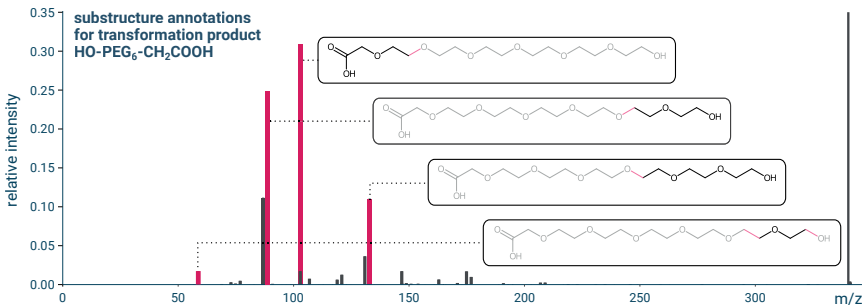

Figure 5: Substructure annotation for transformation product HO-PEG6-CH2COOH. Substructure annotation is directly visualised in SIRIUS.

Conclusion

SIRIUS offers a comprehensive approach to assessing pharmaceutical contamination by combining precursor drug screening and transformation product screening.

SIRIUS is not limited to public chemical databases, but can be used effectively to search custom databases of structure candidates generated, for example, by tools such as BioTransformer. Additionally, SIRIUS provides multiple validation levels, including spectral library searches, confidence scores for database hits, and substructure annotations (see *Figure 5*) for a more in-depth investigation of results.

The ability to screen transformation products without requiring reference spectra is particularly valuable not only for pharmaceuticals but also for pesticides and industrial chemicals, whose degradation products may have significant environmental impacts.

Additional Resources

Learn more about the method:
https://bright-giant.com/labmate-transformation-products

Download SIRIUS: https://bright-giant.com/labmate-sirius-release-latest

Notes

[1] Available online as LUXPHARMA (S76) on Zenodo (DOI: 10.5281/zenodo.4587356)

[2] MassBank-data release 2024.11 (DOI: 10.5281/zenodo.14221628)

[3] LC-MS/MS Spectra from https://mona.fiehnlab.ucdavis.edu/downloads

[4] For metabolism prediction we chose AllHuman, which predicts small molecule metabolism in the human superorganism. It covers biotransformations occurring both in human tissues as well as the gut microbiota. We predict 3 steps. https://biotransformer.ca/help

[5] The biomolecule structure database is an aggregation of several structure databases containing small molecules of biological interest, including metabolites and other compounds of biological relevance, natural products, synthetic products with potential bioactivity, and contaminants observed in experiments. https://v6.docs.sirius-ms.io/methods-background/#CSIFingerID

[6] Differences in identification numbers compared to the study by Singh et al.1 may be attributed to our approach of excluding the precursor peak during spectral matching.

[7] A confidence score threshold of 0.64 roughly corresponds to FDR 10 % in evaluations (Hoffmann et al. 2022)

[8] Dimethyl phthalate's production and use as a plasticizer for nitrocellulose and cellulose acetate, resins, rubber, and in solid rocket propellants; in lacquers; plastics; rubber; coating agents; safety glass and in molding powders (Lewis 2007) may result in its release to the environment through various waste streams(SRC). https://pubchem.ncbi.nlm.nih.gov/compound/8554

References

Djoumbou-Feunang, Yannick, Jarlei Fiamoncini, Alberto Gil-de-la-Fuente, Russell Greiner, Claudine Manach, and David S. Wishart. 2019. 'BioTransformer: A Comprehensive Computational Tool for Small Molecule Metabolism Prediction and Metabolite Identification'. Journal of Cheminformatics 11 (1): 2.

Dührkop, Kai, Markus Fleischauer, Marcus Ludwig, Alexander A. Aksenov, Alexey V. Melnik, Marvin Meusel, Pieter C. Dorrestein, Juho Rousu, and Sebastian Böcker. 2019. 'SIRIUS 4: A Rapid Tool for Turning Tandem Mass Spectra into Metabolite Structure Information'. Nature Methods 16 (4): 299–302.

Dührkop, Kai, Huibin Shen, Marvin Meusel, Juho Rousu, and Sebastian Böcker. 2015. 'Searching Molecular Structure Databases with Tandem Mass Spectra Using CSI:FingerID'. Proceedings of the National Academy of Sciences of the United States of America 112 (41): 12580–85.

Hoffmann, Martin A., Louis-Félix Nothias, Marcus Ludwig, Markus Fleischauer, Emily C. Gentry, Michael Witting, Pieter C. Dorrestein, Kai Dührkop, and Sebastian Böcker. 2022. 'High-Confidence Structural Annotation of Metabolites Absent from Spectral Libraries'. Nature Biotechnology 40 (3): 411–21.

Lewis, Richard J. Sr 2007. Hawley's Condensed Chemical Dictionary, Book and CD-ROM Set. Wiley-Interscience.

Singh, Randolph R., Adelene Lai, Jessy Krier, Todor Kondić, Philippe Diderich, and Emma L. Schymanski. 2021. 'Occurrence and Distribution of Pharmaceuticals and Their Transformation Products in Luxembourgish Surface Waters'. ACS Environmental Au 1 (1): 58–70.

Wishart, David S., Siyang Tian, Dana Allen, Eponine Oler, Harrison Peters, Vicki W. Lui, Vasuk Gautam, Yannick Djoumbou-Feunang, Russell Greiner, and Thomas O. Metz. 2022. 'BioTransformer 3.0-a Web Server for Accurately Predicting Metabolic Transformation Products'. Nucleic Acids Research 50 (W1): W115–23.