Strukturaufklärung aus LC-HRMS-Daten: Gängige Softwaretools für den Umweltbereich

(Ausgabe 1.0 2025)

Wasserchemische Gesellschaft

Fachgruppe in der Gesellschaft Deutscher Chemiker

GDCh

Gesellschaft Deutscher Chemiker

Impressum

Ausgabe 1.0 2025

Erschienen in Koblenz, April 2025

Verantwortlich für den Inhalt: Susanne Brüggen, Marco Scheurer

Dieser Leitfaden wurde von einer Teilgruppe folgender Mitglieder des Fachausschusses ‘Non-Target-Screening‘ der Wasserchemischen Gesellschaft erarbeitet.

Mitglieder:	
Armbruster, Dominic	TZW: DVGW-Technologiezentrum Wasser
Bader, Tobias	Zweckverband Landeswasserversorgung
Brüggen, Susanne*	Landesamt für Natur, Umwelt und Verbraucherschutz NRW
Letzel, Thomas	Analytisches Forschungsinstitut für Non-Target Screening GmbH
Liebmann, Diana	Labor der Berliner Wasserbetriebe
Macherius, André	Bayerisches Landesamt für Umwelt
Ruppe, Steffen	Amt für Umwelt und Energie Basel
Scheurer, Marco*	Landesanstalt für Umwelt Baden-Württemberg
Schlüsener, Michael	Bundesanstalt für Gewässerkunde
Singer, Heinz	Eawag-Swiss Federal Institute of Aquatic Science and Technology
Zahn, Daniel	Helmholtz-Zentrum für Umweltforschung
Zwiener, Christian	Universität Tübingen

* koordinierende Bearbeitung

Herausgeber
Wasserchemischen Gesellschaft
Fachausschuss ‘Non-Target-Screening‘ des Hauptausschusses Forschung
Bundesanstalt für Gewässerkunde
Am Mainzer Tor 1
56068 Koblenz
Telefon: +49 69 7917-368
E-Mail: Wasserchemische-Gesellschaft@bafg.de
Web: http://www.wasserchemische-gesellschaft.de

©Wasserchemische Gesellschaft
Das Urheberrecht für dieses Dokument liegt bei den mitwirkenden Autoren. Alle Anfragen zur Wiedergabe und Reproduktion in jedem Medium, einschließlich Übersetzungen, sollten direkt an das Sekretariat der Wasserchemischen Gesellschaft gerichtet werden. Der Text darf zwecks Weiterverkaufs nicht kopiert werden.

Inhaltsverzeichnis

Abbildungsverzeichnis 4

1 Zielsetzung 5

2 Einleitung 5

3 Software Lösungen für die Strukturaufklärung 7

3.1 MassBank 7

3.2 MetFrag Web beta 11

3.3 SIRIUS 13

3.4 mzCloud™ 15

3.5 CFM-ID 21

3.6 PF∆Screen 24

3.7 FOR-IDENT 26

4 Zusammenfassung 28

Danksagung 28

Abbildungsverzeichnis

Abbildung 1: Möglicher Ablauf der Strukturaufklärung eines unbekannten Features beginnend mit einer *mass of interest* 7

Abbildung 2: WebUI MassBank 9

Abbildung 3: Beispielhafte Ergebnisanzeige einer Suchanfrage in MassBank 10

Abbildung 4: Eingabemaske von MetFrag und der beispielhaften Suche in der PubChem-Datenbank inklusive Referenzen 12

Abbildung 5: Bestandteile des SIRIUS-Softwarepakets mit Funktionalitäten und Lizensierungsanforderungen 13

Abbildung 6: Spektrenbibliothek LC/MS Reference 16

Abbildung 7: Chemische Struktur und Angaben zu Spektrenbäumen für eine Substanz. Über das „+“ vor der Spalte *ID* sowie über den Button *Expand all* oben rechts können einzelne bzw. alle Einträge aufgeklappt werden 17

Abbildung 8: Detaillierte Ergebnis einer Substanzsuche mzCloud 17

Abbildung 9: Ergebnisansicht der Substanzsuche in mzCloud 18

Abbildung 10: Ergebnis der Spektrensuche in mzCloud 19

Abbildung 11: Detailansicht eines Ergebnisses der Spektrensuche in mzCloud 20

Abbildung 12: Von CFM-ID generierte Spektrenvorhersage für Methyltryptamin für drei Kollisionsenergien. Die Molekülstrukturen der Spektrenpeaks lassen sich mit einem *Mouseover* anzeigen und sind tabellarisch zusammengefasst.22

Abbildung 13: Graphische Benutzeroberfläche von PFΔ*Screen*. Die drei Schritte der Datenauswertung sind im Folgenden beschrieben 25

1 Zielsetzung

Die Intention dieses Dokuments besteht darin, die in der Wasseranalytik meist genutzten Datenbanken und Software-Tools zur Strukturaufklärung ausgehend von priorisierten Features aus Non-Target Analysen (NTA) vorzustellen, um den Einstieg in deren Nutzung zu erleichtern. Ziel ist es, ergänzend zum NTS-Leitfaden der Wasserchemischen Gesellschafft, detaillierte Informationen zur Strukturaufklärung zusammenzustellen.

Das Dokument ist von Anwendern für Anwender geschrieben und enthält als Arbeitshilfe somit auch Tipps und Tricks zur Nutzung der portraitierten Datenbanken und Programme. Es weist darüber hinaus auf vorhandene Hilfen, Lern- und Erklärvideos sowie korrespondierende Literatur hin, lässt aber auch Limitierungen nicht unerwähnt.

Diese Zusammenstellung enthält eine Liste ausgewählter Werkzeuge, die aufgrund ihrer besonderen Nützlichkeit, Effizienz und Popularität ausgewählt wurden und beschränkt sich im Wesentlichen auf Datenbankabfragen mit Referenzspektren und/oder In-Silico-Daten. Dabei handelt es sich jedoch nicht um eine komplette Übersicht aller verfügbaren Tools. Vielmehr spiegelt diese Auswahl die subjektiven Einschätzungen und Erfahrungen der Autoren wider.

2 Einleitung

Der Einsatz des Non-Target Screenings mittels Flüssigkeitschromatographie gekoppelt mit hochauflösender Massenspektrometrie (LC-HRMS) hat in der Wasseranalytik in jüngster Vergangenheit zu einem vermehrten Nachweis von unbekannten und bisher nicht berücksichtigten Substanzen geführt. Der qualitative Nachweis von organischen Spurenstoffen kann dabei unter Zuhilfenahme von statistischen Methoden bereits zur Prozessbeschreibung oder zur Abgrenzung von Proben genutzt werden. Hinsichtlich priorisierter Signale (Features) ist häufig eine Strukturaufklärung (Identifizierung) und die anschließende Quantifizierung von Interesse, was meist mit einem großen analytischen und zeitlichen Aufwand verbunden ist.

Die Identifizierung von unbekannten Substanzen ist als Flaschenhals der Non-Target-Analytik zu sehen.

In den letzten Jahren sind zahlreiche Datenbanken und Programme entwickelt worden, die die Strukturaufklärung der priorisierten Features durch Datenbankabgleiche, in silico-Fragmentierungen oder durch künstliche Intelligenz (KI) unterstützte Strategien erleichtern sollen. Die Nutzung von Datenbanken zur Unterstützung der Identifizierung von unbekannten Features ist eine schnelle und effektive Vorgehensweise, für die im Internet zum einen kostenlose MS^2-

Spektren-Datenbanken als auch Stoff-Datenbanken zur Verfügung stehen. Hierbei können potentielle Kandidaten gefunden werden bzw. lässt sich deren Anzahl durch den unmittelbaren Abgleich der MS^2-Spektren einschränken. Durch den direkten Abgleich von detektierten Fragmenten mit Fragment-Datenbanken können zudem Hinweise auf strukturelle Elemente erhalten werden. Neben allgemeinen chemischen Datenbanken mit mehreren Millionen Einträgen sind auch solche mit primär gewässerrelevanten Substanzen vorhanden, die die Anzahl potentieller Kandidaten bereits sinnvoll einschränken. Darüber hinaus versuchen KI-gestützte Ansätze, Strukturvorschläge durch Abgleich mit Strukturdatenbanken zu generieren. Dabei wird auf Grundlage des Fragmentierungsmusters ein molekularer Fingerabdruck erstellt und mit solchen verglichen, die aus Strukturcodes, wie SMILES oder InChI, abgeleitet werden können.

Dieses Dokument beschreibt die letzten Schritte im Non-Target Workflow, bei denen priorisierte Features identifiziert werden sollen. Ausgangspunkt ist folglich eine mass of interest für die u. a. Ladungszustand, Ionen-Spezies, Addukte und eine eventuelle In-Source-Fragmentierung geprüft werden müssen. Das Vorhandensein von Isotopen der mass of interest wird von einigen HRMS-Nutzen als Voraussetzung für weitere Anstrengungen zur Strukturaufklärung angesehen. Im Idealfall liegen zu diesem Zeitpunkt bereits MS/MS-Daten vor (Abbildung 1). Sind diese nicht vorhanden, sollten MS/MS-Spektren in nachfolgenden Messungen gezielt generiert werden, vorzugsweise bei mehreren Kollisionsenergien, um die hier vorgestellten Werkzeuge sinnvoll nutzen zu können. Diese können zu unterschiedlichen Identifizierungslevels führen (Schymanski et al., 2014). Ziel sollte es sein das Spektrum der mass of interest zumindest mit einem Datenbanktreffer zu verifizieren (Level 2), um eine chemische Struktur vorschlagen zu können. Eine finale Bestätigung erfolgt, falls möglich, mit einem Referenzstandard (Level 1)

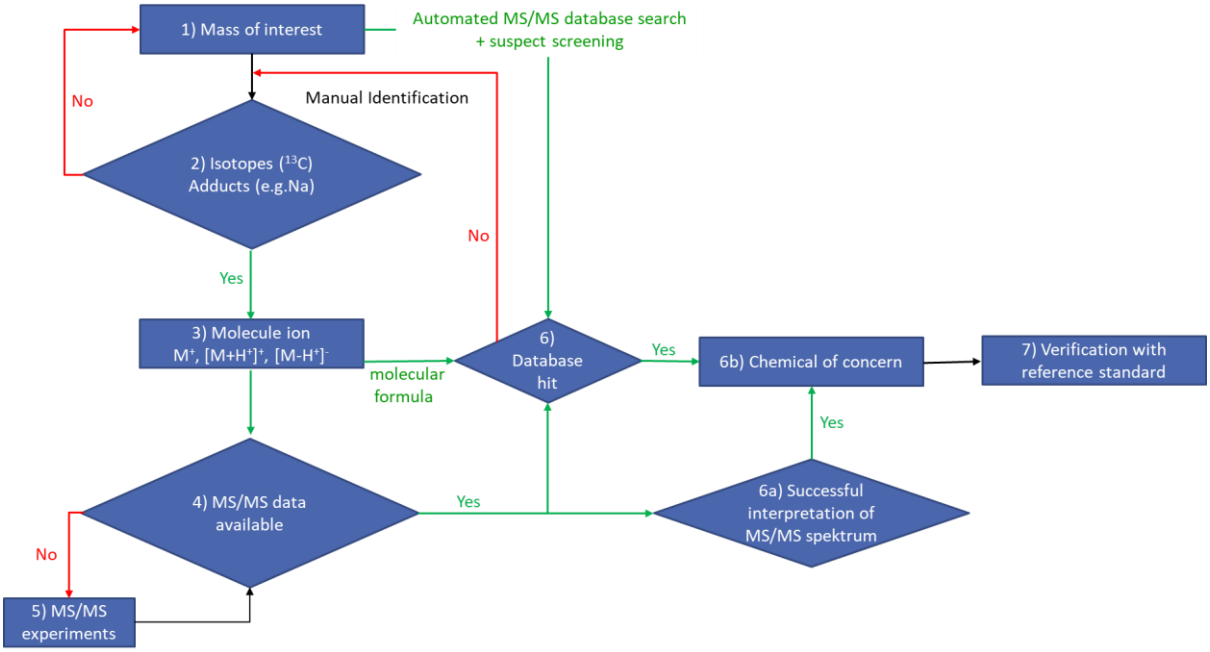

Abbildung 1: Möglicher Ablauf der Strukturaufklärung eines unbekannten Features beginnend mit einer *mass of interest*

Weiterführende Informationen

Schymanski; E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H.P.; Hollender, J. Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence, Environ Sci Tech, 48(4).

3 Software Lösungen für die Strukturaufklärung

3.1 MassBank

MassBank ist eine Open-Source-Datenbank, die Massenspektren organischer Stoffe enthält. Sie wird von einem großen Konsortium gepflegt und weiterentwickelt. Im Fokus stehen kleine Moleküle, weshalb auch der Umweltbereich stark vertreten ist. Die meisten Spektren wurden mit hochauflösenden Massenspektrometern bei verschiedenen Kollisionsenergien erzeugt.

Suchoptionen

Die Datenbank kann über das WebUI genutzt werden (https://massbank.eu/MassBank/; Abbildung 2), wobei verschiedene Suchoptionen zur Verfügung stehen:

- ‘Basic search‘ ermöglicht die Spektrensuche ausgehend vom Stoffnamen, berechneter Masse oder Summenformel. Vor der Suche lassen sich über Filter (z. B. Geräte, Ionisationstechniken, Polarität) die Ergebnisse bereits eingrenzen.
- Stoffname: Es können auch nur Teile des Strings eingetragen werden. In der Datenbank werden auch alle hinterlegten Namen (Synonyme, teilweise IUPAC-Bezeichnungen) durchsucht. Häufig existieren mehrere Bezeichnungen für eine Substanz. Viele Beitragende geben jedoch nur einen Namen an, weshalb die Suche der Substanz ausschließlich über den Namen nicht zu empfehlen ist.
- Summenformel: Es muss darauf geachtet werden, dass keine Halbstrukturformeln genutzt werden (Bsp. Angabe von $C_2HF_3O_2$ anstatt CF_3COOH). Außerdem muss die richtige Reihenfolge (C, H, alle weiteren Elemente alphabetisch) eingehalten werden, da sonst kein Datenbanktreffer erzielt wird.
- Die ‘Peak list‘-Suche erlaubt den klassischen Abgleich von gemessen MS/MS-Spektren mit den in der Datenbank hinterlegten Spektren. Bei dieser „Similarity“-Suche wird die Ähnlichkeit der Spektren in Betracht gezogen. Hierbei muss, im Gegensatz zur „Peaks“-Suche (s. u.), nicht jedes Fragment aus dem Eingabefenster im Spektrum enthalten sein.

  Über das Eingabefenster lässt sich eine Liste von m/z-Werten und relativen Intensitäten der Fragmentionen eintragen. Bei einigen Geräteherstellern lassen sich MS/MS-Spektren direkt in der benötigten Form exportieren. Weniger intensive Signale, welche häufig Rauschen darstellen, lassen sich durch einen relativen Intensitäts-Schwellenwert entfernen. Der Standardwert liegt bei 5%, bezogen auf das größte Signal im Spektrum. Über das Eingabefenster lässt sich keine Massentoleranz angeben, d.h. die Ergebnisse umfassen auch Spektren, die von anderen Vorläuferionen stammen. Bei unbekannten Stoffen lassen sich durch MS/MS-Ähnlichkeiten ggf. bereits Strukturmerkmale zuordnen.
- Mit der ‘Peaks‘-Suche lassen sich Massenspektren gezielt nach bestimmten m/z durchsuchen. Dabei lassen sich ein oder mehrere m/z-Werte in der Suchmaske definieren, welche in allen Datenbankspektren enthalten sein müssen. Intensitäten werden hier nicht berücksichtigt. Beim Eintragen mehrerer m/z-Werte sind diese in den Default-Einstellungen durch UND-verknüpft, d.h. jede der Fragmentmassen muss im Datenbankspektrum gefunden werden. Optional lässt sich auch eine ODER-Verknüpfung auswählen. Bei der ‘Peaks‘-Suche lassen sich Massentoleranzen für die Signale berücksichtigen.

- Mit „Peak Differences“ kann gezielt nach Massendifferenzen in MS/MS-Spektren gesucht werden. In der Eingabemaske lassen sich m/z-Differenzen angeben oder alternativ die Summenformel des Neutralverlustes (z. B. H_2O, CO_2), woraus automatisch die m/z-Differenzen berechnet werden.
- Eine Suche kann außerdem über den ‘InChiKey‘ (eindeutiger Strukturidentifier) oder ein ‘Splash‘ (eindeutiger Spektrenidentifier) durchgeführt werden.

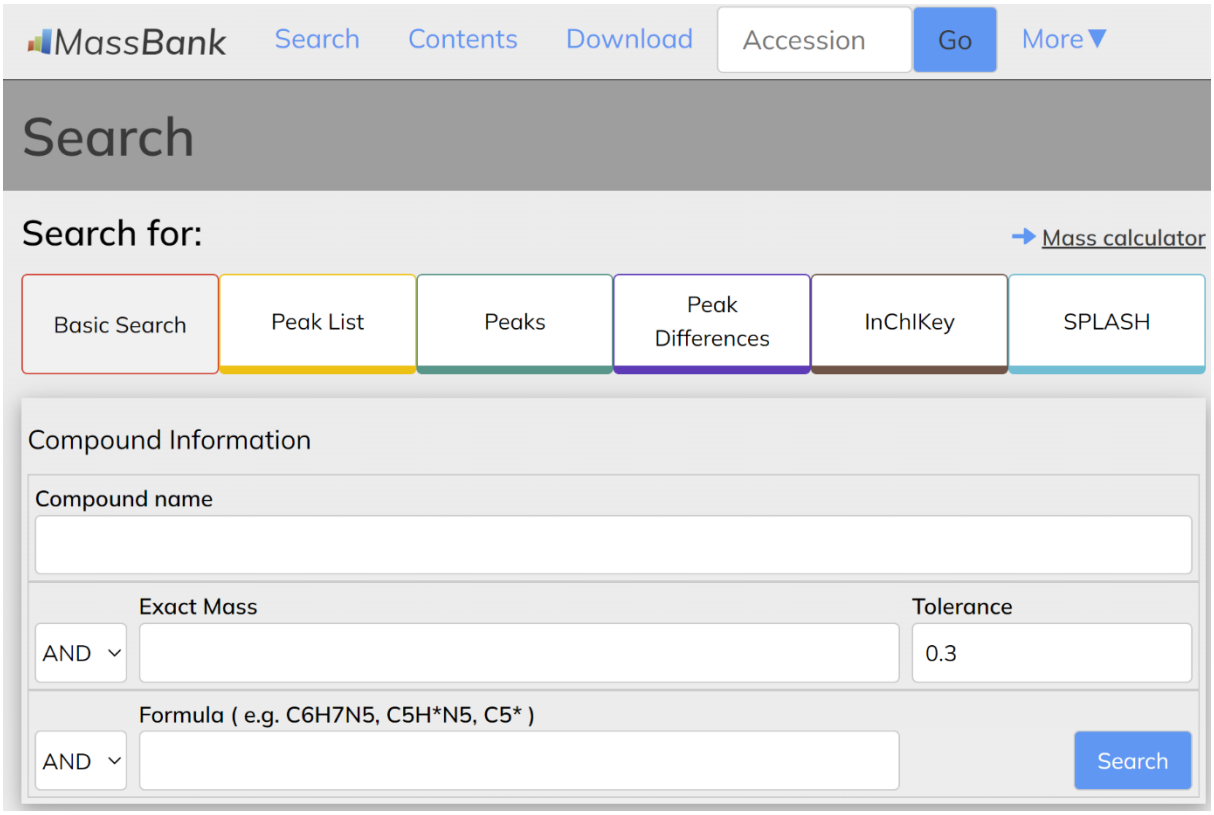

Abbildung 2: WebUI MassBank

Datenbankeinträge

Die Ergebnisse der oben beschriebenen Suchanfragen werden in einer Tabelle dargestellt, welche den Namen, die Summenformel und die Strukturformel enthält (Abbildung 3). Weitere Spalten sind abhängig von der Art der Suche. Werden Spektren über die ‘Peak list‘ abgefragt, werden zusätzlich noch die Anzahl der Hits (wie viele Fragmentmassen stimmen im Upload mit dem jeweiligen Datenbankeintrag überein) sowie ein Score zwischen 0 und 1 (wie gut die Muster zusammen passen) ausgegeben.

Neben dem Namen befindet sich ein +-Zeichen, unter welchem für diesen Treffer alle Spektren (verschiedener Geräte und Kollisionsenergien, sofern vorhanden) verlinkt sind. Ferner lässt sich über die Spalte ID die Herkunft des Spektrums erkennen.

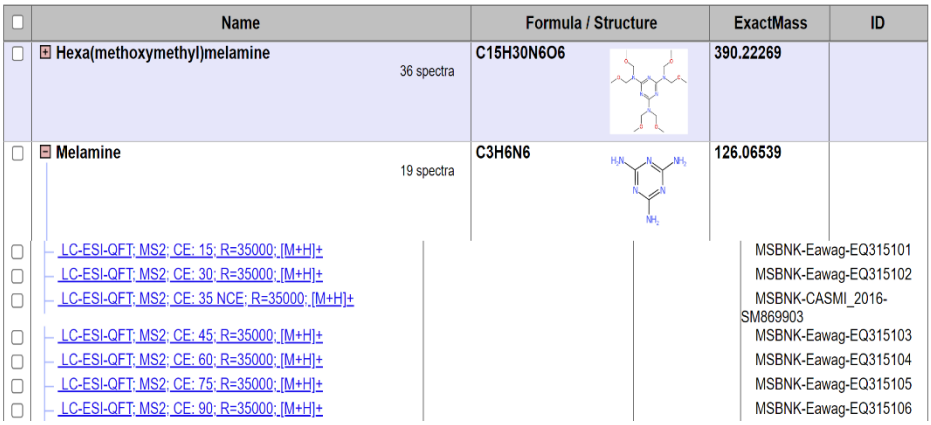

Abbildung 3: Beispielhafte Ergebnisanzeige einer Suchanfrage in MassBank

Über das Klicken der Links wird der eigentliche Datenbankeintrag geöffnet. Dieser stellt zu Beginn das Spektrum dar. Eine automatische Annotierung der Fragmentmassen ist nicht umgesetzt, allerdings lässt sich durch ein „Mouseover“ die Masse und Intensität anzeigen. Außerdem sind zum Ende der Seite alle Fragmentmassen mit weiteren Informationen aufgelistet. Der Datenbankeintrag enthält auch weitere Metadaten zur Herkunft der Daten (Autoren, Datum etc.), weitere Substanz-Identifier (Namen, SMILES, InchIKey, CAS etc.) sowie weitere methodische Metadaten (Gerät, Polarität, Fragmentierung, Säule, Gradient etc.).

Limitierungen

Aktuell gibt es kaum eine Dokumentation, in welcher die Funktionsweise beschrieben wird. Die Spektrenabfrage über das WebUI dauert teilweise lange und ist nicht selten von Time-out-Fehlern begleitet. Ferner lässt sich beim klassischen Spektrenabgleich (‘Peak list‘) aktuell weder eine Masse für das Vorläuferion definieren, noch kann die Massengenauigkeit für die Fragmentmassen definiert werden.

Eine lokale Nutzung von MassBank kann diese Limitierungen teilweise ausräumen. Die MassBank-Daten lassen sich in verschiedenen Formaten herunterladen und in existierende oder weitergehende Workflows/Tools einbinden. Das .msp-Format wird von zahlreichen kommerziellen und Open-Source-Programmen als Datenbank- oder Importformat unterstützt. Eine einfache Möglichkeit ist die Nutzung über das Programm ‘NIST MS Search‘, des US National Instituts of Standards and Technology. Dieses ermöglicht schnelle Abfragen und

stellt Ergebnisse übersichtlich in Mirror-Plots dar, die übereinstimmende Spektren gespiegelt abbilden, z. B. um Identifizierungen aus Spektrenbibliotheken zu visualisieren.

Weiterführende Informationen

Aufgenommene Spektren für „neue Substanzen“ können der Community über RMassBank zur Verfügung gestellt werden:

https://github.com/MassBank/RMassBank

Download der Spektren:

https://github.com/MassBank/MassBank-data/releases/tag/2024.06

NIST MS Search:

https://chemdata.nist.gov/dokuwiki/doku.php?id=chemdata:nistlibs

3.2 MetFrag Web beta

MetFrag (MetFrag; ipb-halle.de)) ist eine frei verfügbare Software für die Annotierung von hochaufgelösten Produktionenspektren organischer Substanzen. Dies ist der erste Schritt zur Identifizierung der Struktur eines Moleküls. Kandidatenmoleküle aus verschiedenen Datenbanken werden in silico fragmentiert und mit den Fragmenten abgeglichen. Ein anhand der Fragmentpeak-Übereinstimmungen berechneter Score gibt Hinweise auf die Qualität der Zuordnung des Kandidatenspektrums.

Die Verwendung von MetFrag erfolgt über ein Webtool. Für die Verarbeitung von Hochdurchsatzdaten ist die Kommandozeilenversion am besten geeignet (Batch-Upload), ebenso wie das R-Paket für Benutzer, die mit der Programmiersprache R vertraut sind.

Eingangsdaten:

- mass of interest oder Summenformel (erforderlich)
- MS/MS-Spektren (erforderlich)
- Retentionszeit mit Hilfe von RT data model (optional)

Ein Abgleich kann mit verschiedenen Datenbanken (z. B. PubChem, KEGG, ChemSpider, NORMANSusDat, PubChemLite, …) durchgeführt werden. Es gibt viele verschiedene Filteroptionen, um die Anzahl an Kandidaten sinnvoll zu reduzieren. Bei der Suche in PubChem ist beispielsweise anzuraten, sowohl bei den Basiseinstellungen als auch den Filteroptionen die Referenzen der Datenbank zu inkludieren (Abbildung 4). Diese Einstellungen führten bei der Testung von MetFrag mit Beispieldatensätzen zu deutlich besseren Ergebnissen.

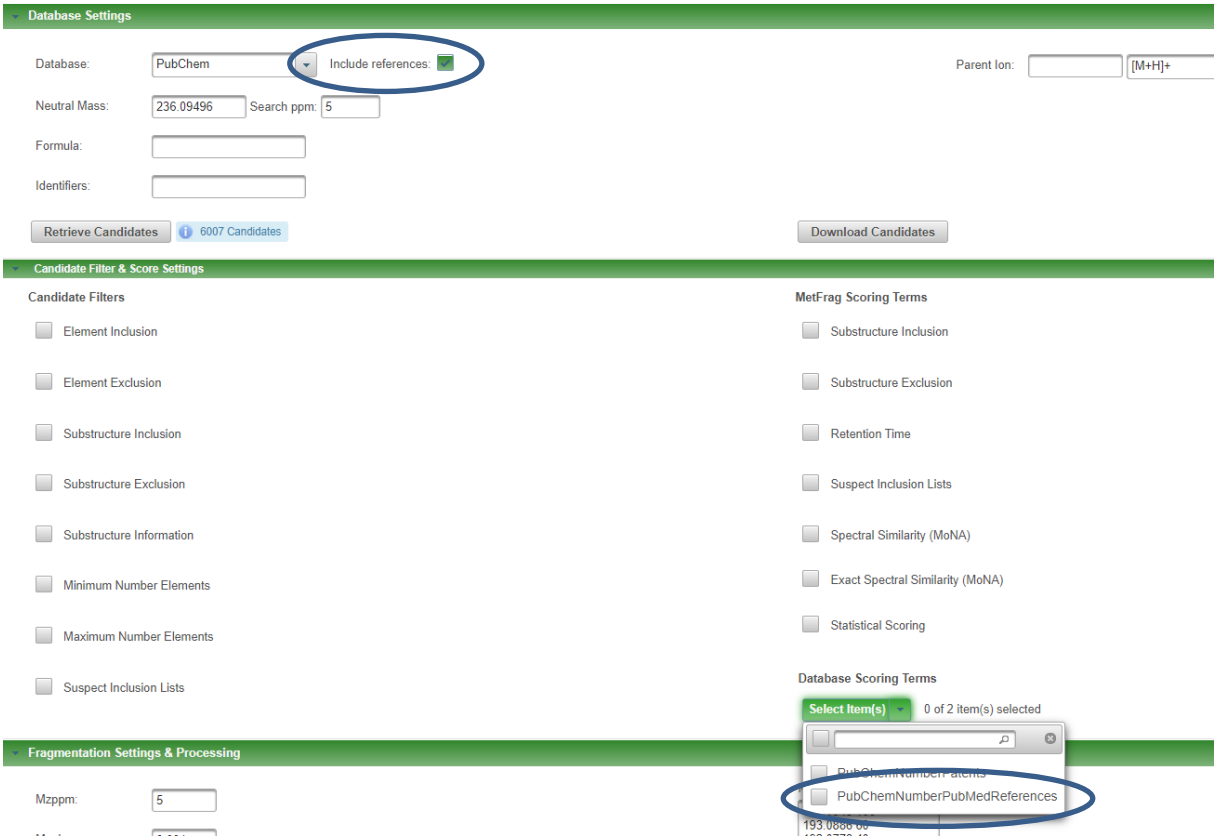

Abbildung 4: Eingabemaske von MetFrag und der beispielhaften Suche in der PubChem-Datenbank inklusive Referenzen

Die Ergebnisse werden entsprechend gerankt angezeigt. Eine Gesamtliste der Ergebnisse kann zudem exportiert werden. Unter „Fragments“ werden alle eingegeben MS/MS-Fragmente angezeigt; dabei werden Fragmente farblich markiert, die einem Strukturelement zugeordnet werden können. Diese Zuordnung kann ebenfalls heruntergeladen werden.

Weiterführende Informationen

MetFrag bietet eine bebilderte Eingabehilfe mit Beispieldaten an. Auch ein Trainings-Video ist vorhanden:

- https://ipb-halle.github.io/MetFragTraining/ (Pfeil rechts unten)
  MetFrag - project (ipb-halle.github.io); Achtung: Für optimale Funktionalität ist bei der Massenabweichung 4.7 statt 5 ppm einzugeben.

Weitere Infos:

- https://github.com/ipb-halle/MetFragTraining
- https://www.youtube.com/watch?v=V5M-dcE3qz8 (ab 3.45 min)

3.3 SIRIUS

Das SIRIUS-Softwarepaket wurde vorrangig für die Prozessierung von LC-HRMS/MS-Daten im Feld der Metabolomics, d. h. für Identifizierung niedermolekularer Biomoleküle, entwickelt. In der aktuellen Version 5 umfasst es die Tools SIRIUS, ZODIAC, CSI:FingerID und CANOPUS (Abbildung 5) und ermöglicht die Summenformelannotation, das Ableiten von Strukturvorschlägen und die Verbindungsklassenannotation.

Um das Softwarepaket in vollem Umfang nutzen zu können, ist die Installation, das Anlegen eines Nutzerkontos und eine Nutzerlizenz erforderlich. Für den akademischen Bereich werden die Lizenz und die entsprechenden Webservices kostenfrei zur Verfügung gestellt. Für andere Nutzer werden Lizenz und Webservices durch die Bright Giant GmbH bereitgestellt und sind nicht oder nur im begrenzten Umfang kostenfrei. Mit SIRIUS und ZODIAC werden die Daten lokal prozessiert, während für CSI:FingerID und CANOPUS der Zugriff auf die Webservices/Server erforderlich ist. Die Nutzung der Tools SIRIUS und ZODIAC ist daher auch ohne gültige Lizenz möglich.

Nicht-kommerzielle Institutionen wie Ämter oder gemeinnützige Organisationen können 500 Verbindungen pro Jahr kostenfrei prozessieren; darüber hinaus muss eine Lizenz erworben werden (persönliche Mitteilung durch Bright Giant GmbH).

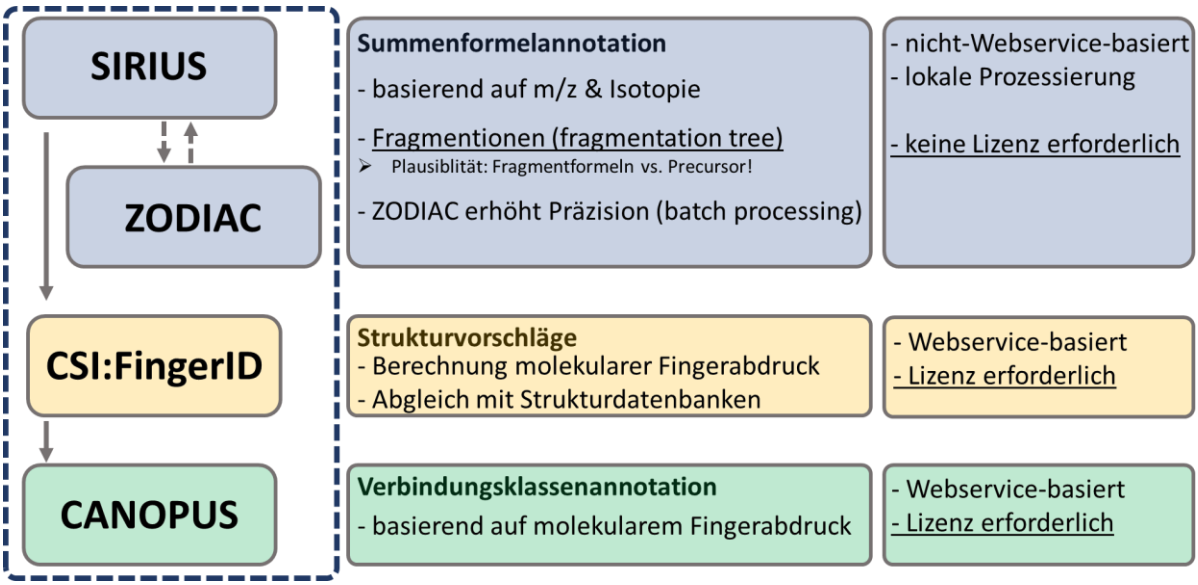

Abbildung 5: Bestandteile des SIRIUS-Softwarepakets mit Funktionalitäten und Lizensierungsanforderungen

Für die Anwendung des SIRIUS-Softwarepakets sind zwingend LC-HRMS/MS-Daten erforderlich, die im DDA-Modus (Data-Dependent-Acquisition) aufgezeichnet wurden. Zudem müs-

sen die Rohdaten in das mzML-Format konvertiert und centroidisiert werden. SIRIUS 5 ermöglicht die wesentlichen Schritte des Preprocessings, wie Peak picking, Komponentisierung sowie Feature alignment (optional) inkl. gap filling bei der Auswertung größerer Datensätze. Dabei werden nur Features importiert für die MS/MS-Daten vorliegen. Zudem ist es möglich, Daten aus mzMine und OpenMS zu importieren und so weitere Kriterien für Feature-Extraktion und Priorisierung zu nutzen.

Exkurs: Datenkonvertierung von Rohdaten in das mzML- bzw. mzXML-Format

Einige der hier vorgestellten Software-Tools erfordern die Konvertierung bzw. Centroidisierung der Rohdaten. Ein von der Autorengruppe hierzu viel genutztes Werkzeug ist MSConvert, das kostenfrei über die Website von ProteoWizard (https://proteowizard.sourceforge.io/) verfügbar ist. Beschreibungen zu dessen Nutzung sind als Videotutorials (z. B. https://www.youtube.com/watch?v=hOJ6ibCA5Pk) oder in Textform (z. B. https://www.envibee.ch/eng/enviMass/topics/PW.php) vielfach im Internet zu finden.

Die extrahierten/priorisierten Features können durch das SIRIUS-Softwarepaket einzeln und auch im Batch prozessiert werden. Die Berechnung von Summenformelvorschlägen erfolgt durch das SIRIUS-Tool basierend auf MS1-Daten, d. h. Masse-zu-Ladungsverhältnis und Isotopie, sowie Informationen aus den Fragmentierungsmustern (DDA-MS/MS-Spektren). Bei der Prozessierung einer Vielzahl von Verbindungen des LC-HRMS-Laufes im Batch-Modus verbessert ZODIAC durch den Vergleich ähnlicher Fragmentierungsmuster das Ranking der Summenformelkandidaten. Das basiert auf der Annahme, dass mehrere ähnliche Stoffe in derselben Probe vorhanden sind. Vor allem für Fragestellungen im Metabolomics-Bereich soll ZODIAC zur Präzision der Annotation beitragen. Für die Prozessierung einzelner Features ist ZODIAC nicht sinnvoll und steht nicht zur Verfügung.

Strukturvorschläge werden mit CSI:FingerID durch den Abgleich mit Strukturdatenbanken (im Gegensatz zu MS/MS-Spektrendatenbanken) erarbeitet. Dabei leitet das Tool auf Grundlage des Fragmentierungsmusters einen molekularen Fingerabdruck ab und vergleicht diesen mit Fingerabdrücken, die aus Strukturcodes wie SMILES oder InChI abgeleitet werden, ab. Als Strukturdatenbanken stehen u. a. PubChem, PubMed, NORMAN sowie viele weitere zur Verfügung. Eigene Datenbanken können mit eingebunden werden. Als Ergebnis werden Strukturvorschläge mit einer Zuordnung der Fragmentionen zu Struktureinheiten des Kandidaten generiert.

CANOPUS nutzt die mit CSI:FingerID erstellten Fingerabdrücke und leitet daraus mögliche Zuordnungen zu Verbindungsklassen ab. Dies erlaubt die weitere Charakterisierung von Kandidaten, die mittels Strukturdatenbankabgleich nicht oder nicht eindeutig identifizierbar sind.

Die Ergebnisse können im .csv- bzw. im .tsv-Format gespeichert werden, was den einfachen Import und die weiterführende Prozessierung außerhalb des SIRIUS-Softwarepakets ermöglicht.

Weiterführende Informationen

Für das SIRIUS-Softwarepaket liegt bereits eine umfangreiche Sammlung von Dokumentationen und Tutorials vor:

Homepage des Lehrstuhl Bioinformatik der FSU Jena, SIRIUS Download und Quick Start
https://bio.informatik.uni-jena.de/software/sirius/

Dokumentation des Böker Lab
https://boecker-lab.github.io/docs.sirius.github.io/

Functional Metabolomics Lab - NTMSS2023
https://www.youtube.com/watch?v=Gj_h7fvzL_g
https://www.youtube.com/watch?v=9NXU3KdqYT4

Functional Metabolomics Lab - NTMSS2022
https://www.youtube.com/watch?v=Wo9GIga7cbI&t=205s

3.4 mzCloud™

mzCloud™ (mzCloud – Advanced Mass Spectral Database) ist eine kostenlose, webbasierte Spektrenbibliothek der Firma Thermo Fisher Scientific™ und dient der Identifikation kleiner Moleküle mittels Tandem-Massenspektrometrie.

Für die *New mzCloud* ist eine Registrierung notwendig. Aktuell kann die Registrierung über ein Google- oder Facebook-Konto bzw. eine beliebige Mailadresse erfolgen. Als Browser werden Google Chrome und Mozilla Firefox empfohlen (siehe auch mzCloud – Browser and OS Compatibility). Der Prozess der Registrierung/Anmeldung wird außerdem in kurzen Videoanleitungen auf der Webseite unter *Training* erläutert.

Der Zugriff auf mzCloud™ ist auch über die Software*Mass Frontier*™ und *Compound Discoverer*™ (beide Thermo Fisher Scientific™) möglich, in diesem Fall ist keine gesonderte Registrierung für den Zugriff auf die Datenbank nötig.

In Zusammenarbeit mit Forschungs- und Industriepartnern sind zwei Datenbanken (LC/MS Reference, LC/MS Autoprocessed) entstanden, die Stand Februar 2024 insgesamt 8,7 Millionen hoch- und niedrigaufgelöste MS^n-Spektren von ca. 26.000 Substanzen umfassen und ständig erweitert werden. Für die Datenbanken wurden die MS/MS- und MS^n-Spektren unter einer Reihe von Versuchsbedingungen (CID, HCD, mehrere Kollisionsenergien) aufgenommen und in sogenannten Spektrenbäumen (spectral trees) zusammengeführt. Während die Spektren in der Datenbank *LC/MS Reference* manuell kuratiert wurden, wurden die Spektren in der Datenbank *LC/MS Autoprocessed* einer Autoprozessierung unterzogen. Die *LC/MS Reference*-Datenbank enthält außerdem eine Annotierung von Fragmenten, die mit Hilfe der Software Mass Frontier™ vorhergesagt und von erfahren Massenspektrometrikern bewertet und manuell zugewiesen wurden. Die Einträge können über die *Spalten ID, Legacy ID, Compound Name, Formula* und *Monoisotopic Mass* sortiert und gefiltert werden (Abbildung 6).

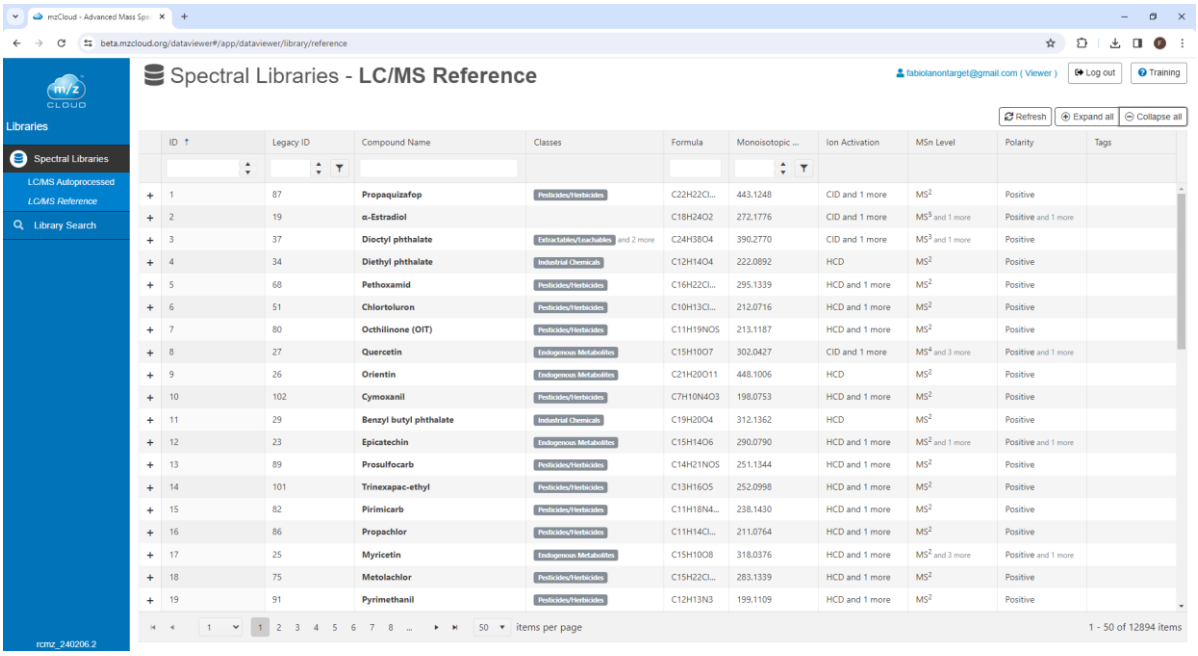

Abbildung 6: Spektrenbibliothek LC/MS Reference

Neben den experimentellen Daten wie Spektrenbäumen (Abbildung 7) und Zerfallskurven (Fragmentintensität gegen Kollisionsenergie) enthält jeder Datenbankeintrag den Namen der Verbindung mit Synonymen, die chemische Struktur, Summenformel, InChI, InChIKey und andere Identifikatoren sowie detaillierte experimentelle Informationen und Verlinkungen zu anderen chemischen Datenbanken wie z. B. ChemSpider und PubChem.

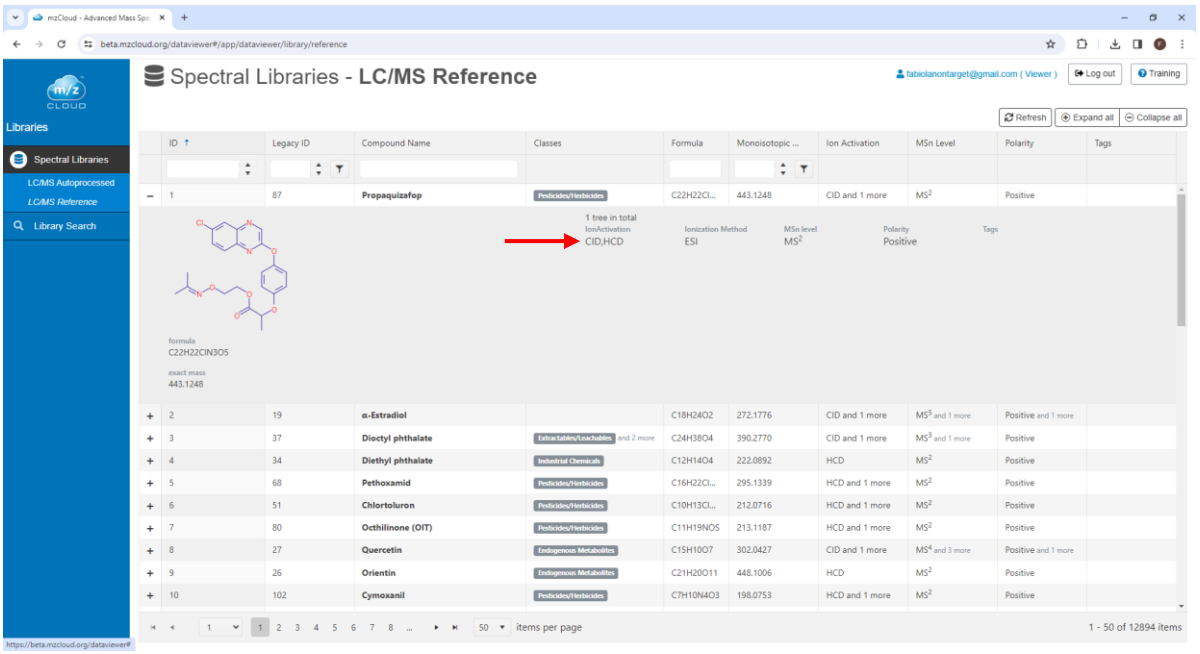

Abbildung 7: Chemische Struktur und Angaben zu Spektrenbäumen für eine Substanz. Über das „+“ vor der Spalte *ID* sowie über den Button *Expand all* oben rechts können einzelne bzw. alle Einträge aufgeklappt werden

Durch einen Klick auf die Zeile CID, HCD öffnet sich die detaillierte Übersicht zu der jeweiligen Substanz. Diese Übersicht zeigt den Spektrenbaum, Metadaten inkl. Verlinkungen zu anderen Datenquellen, die experimentellen Bedingungen, die Fragmentspektren und Zerfallskurven (Abbildung 8).

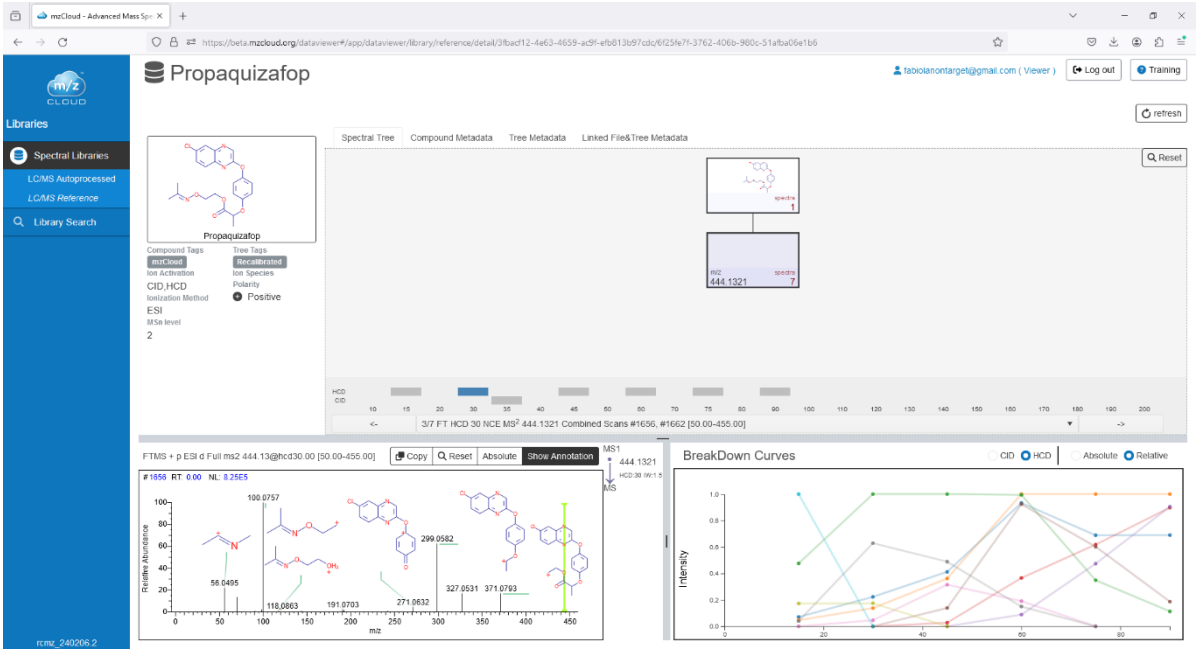

Abbildung 8: Detaillierte Ergebnis einer Substanzsuche mzCloud

Substanzsuche in Datenbanken

Es stehen mehrere Suchfunktionen zur Verfügung, die in kurzen Videoanleitungen ebenfalls auf der Webseite unter *Training* erläutert werden. So kann nach einer Substanz u. a. über den Substanznamen und andere Identifikatoren (einzeln und im Batch), sowie über die chemische Struktur (mol-Datei) gesucht werden. Als Suchkriterien stehen Compound Name, Legacy ID, Canonical SMILES, CAS Nr., InChI und InChIKey bzw. eine beliebige Kombination der Suchkriterien zur Verfügung. Die Suche kann zusätzlich über die *Compound Class* oder *Monoisotopic Mass* eingeschränkt werden.

Die Substanzsuche erzeugt eine Liste mit Treffern, in der man durch Klicken auf Show In Library unter dem jeweiligen Compound Name (Abbildung 9) zum entsprechenden Eintrag in einer der beiden Spektrenbibliotheken gelangt. Man hat die Möglichkeit über den Button *Columns* (oben rechts) beliebige Spalten mit Informationen z. B. zu InChIKey, PubChem und viele mehr zu der Ansicht hinzuzufügen oder herauszunehmen. Es kann über viele der Spalten gefiltert und/oder sortiert werden.

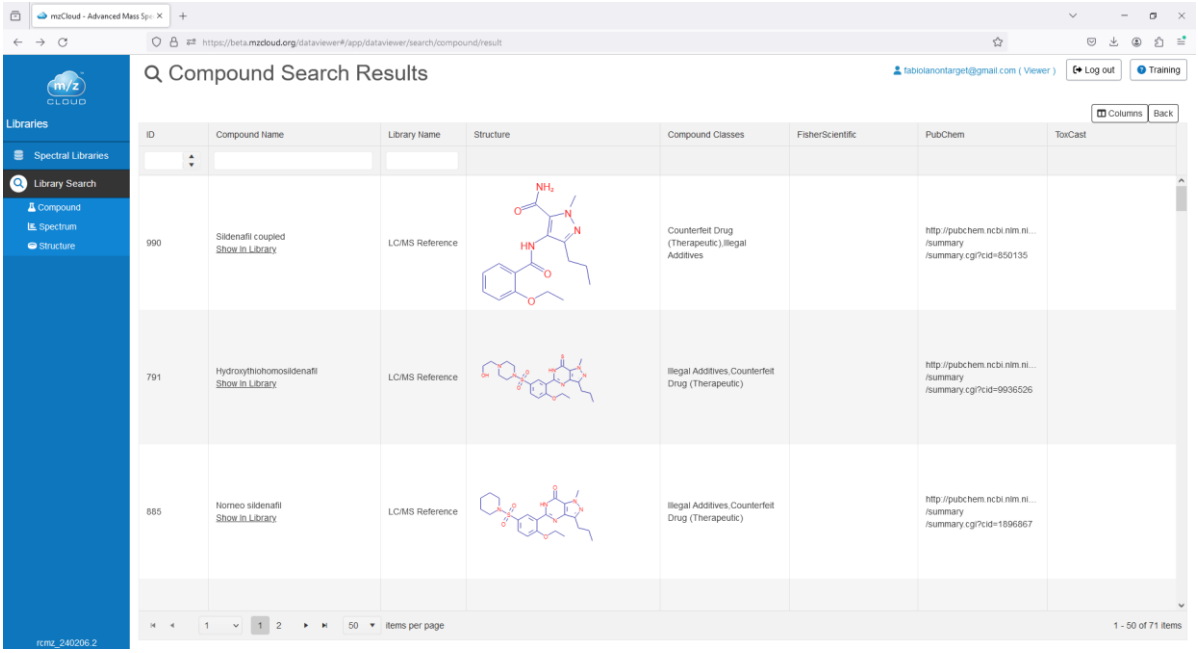

Abbildung 9: Ergebnisansicht der Substanzsuche in mzCloud

Auch wenn eine Substanz nicht in der Bibliothek vertreten ist, bietet die Technik des *Precursor Ion Fingerprinting* (PIF) eine Ähnlichkeitssuche (*Similarity Search*), die bei der Identifizierung unterstützen soll. Dazu werden substrukturelle Informationen durch den Vergleich von Produktionenspektren von strukturell verwandten Verbindungen identifiziert. Beiträge mit detaillierteren Informationen zu PIF sind auf der Webseite verlinkt.

Substanzsuche mit experimentell erzeugten Spektren

Das Hauptaugenmerk von mzCloud™ liegt jedoch auf der Identifizierung einer Substanz über experimentell erzeugte MS^n-Spektren (nur Einzelsuche). Diese können als msp- oder json-Datei sowie in Textform mit den Fragmentmassen und den dazugehörigen (relativen) Intensitäten in das Suchfeld eingeladen werden. Dabei ist zu beachten, dass die Einstellmöglichkeiten rechts unter dem Spektrum nur genutzt werden können, wenn das Spektrum als msp-Datei und NICHT in Textform eingeladen wurde. Bei Auswahl von *Identity* als *Search Type* ist die Angabe einer Precursor-Masse nötig.

Es gibt keine Informationen zu den unterschiedlichen *Search Types* und *Scoring Methods*. Die Suchergebnisse werden automatisch absteigend nach dem *Match Score* der ausgewählten *Scoring Method* sortiert. Dabei wird der *Match Score* immer für alle drei verfügbaren *Scoring Methods* angegeben. Man hat die Möglichkeit die Ergebnisse u.a. nach *Compound Name*, einem der *Match Scores, Number of matched Peaks (No. Matched Peaks)* und anderen Kriterien zu sortieren. Die Ergebnisse können außerdem gefiltert werden nach: *Ion Activation, Polarity, Compound Class, Tree Depth, No. Matched Peaks, Intensity Overlap, ScanRange Miss, Precursor Accuracy* und *Library* (Abbildung 10).

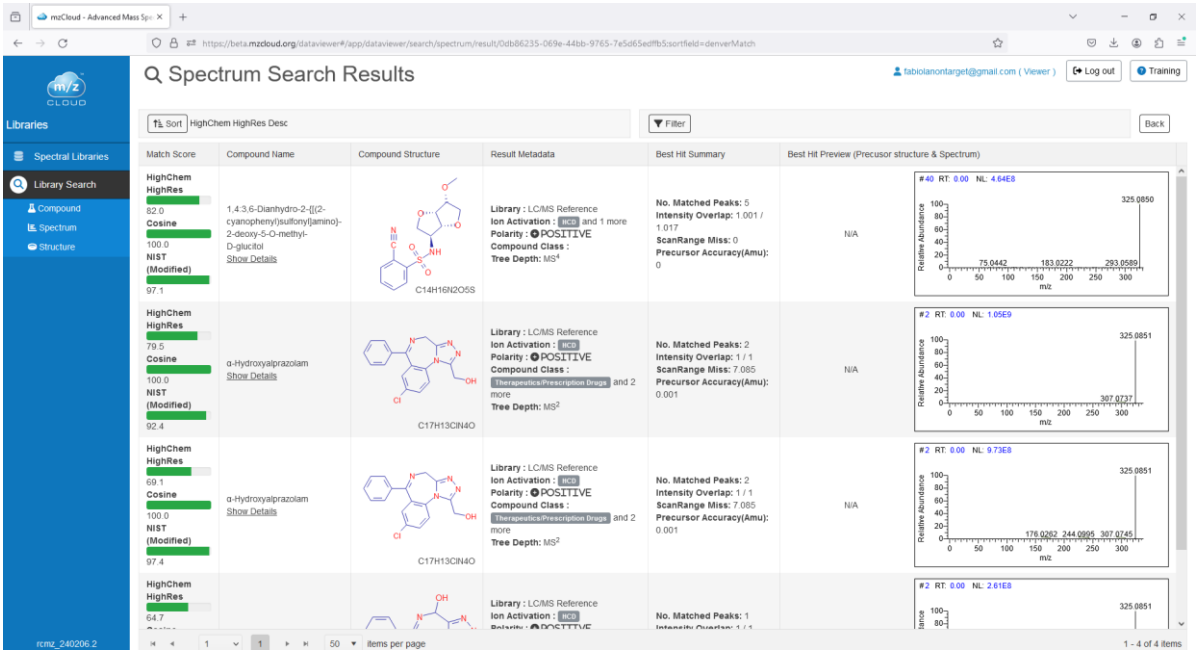

Abbildung 10: Ergebnis der Spektrensuche in mzCloud

Durch klicken auf *Show Details* unter dem jeweiligen *Compound Name* gelangt man anschließend zur Detail-Ansicht des jeweiligen Treffers inkl. Metadaten und experimentellen Details (Abbildung 11).

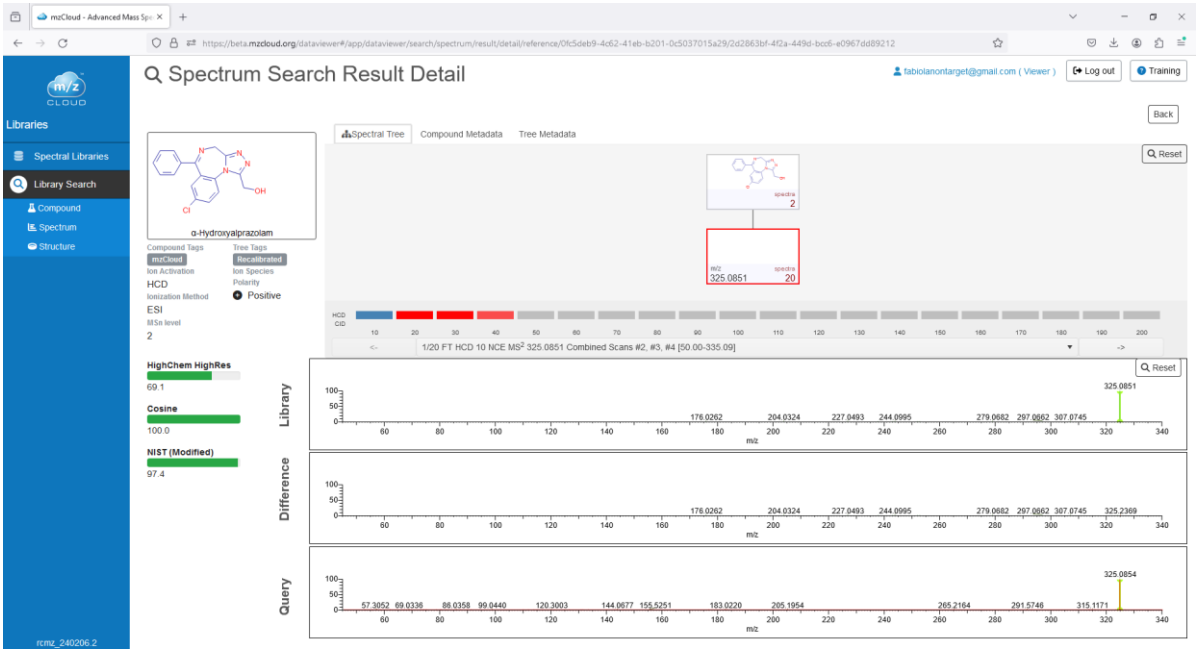

Abbildung 11: Detailansicht eines Ergebnisses der Spektrensuche in mzCloud

Weiterführende Informationen

Bei Tests mit selbst erzeugten Orbitrap-Spektren sowie TOF-Spektren aus MassBank wurden folgende Erfahrungen gemacht:

- Score-Einstellung von > 50 führen häufig zu richtigen Ergebnissen
- Verwendung von *Symmetric* als *Match Algorithm* ist zu empfehlen, da weniger falsch positive Ergebnisse erzeugt werden
- Beispiel: Desisopropylatrazin-Spektren → viele ähnliche Pestizide (Triazine) → aussagekräftige MS/MS-Spektren sind notwendig (CE-Werte!)
- Bei wenigen Fragmenten (z. B. Moclobemid) wird es schwieriger → Score-Threshold 60 besser

Für die gezielte Struktursuche muss eine mol-Datei hochgeladen werden. Es steht kein integrierter Strukturediteor zur Verfügung, wie auf der Webseite angegeben wird. Die Ergebnisübersicht entspricht der Übersicht der Substanzsuche und kann ebenso nach den eigenen Bedürfnissen angepasst, gefiltert und sortiert werden.

- Auf der Webseite werden Übersichten zu den Verbindungen je Verbindungsklasse, die in den Datenbanken enthalten sind, zur Verfügung gestellt. Diese lassen sich als csv-Dateien herunterladen und können zusätzlich als Massenlisten im Suspect- und Non-Target Screening verwendet werden.
- Kurze Reaktionszeit des Supports
- Die App ist nicht mehr verfügbar, da diese durch die *new mzCloud* nicht länger unterstützt wird.

3.5 CFM-ID

CFM-ID bietet eine kostenlose Browseranwendung, die auf Competitive Fragmentation Modeling (CFM) basiert und unter CFM-ID (wishartlab.com) aufgerufen werden kann. Es handelt sich um ein probabilistisches, generatives Modell für MS/MS-Fragmentierungsprozesse, das maschinelles Lernen zur Datengenerierung nutzt. CFM-ID wurde ursprünglich für die Metabolomics-Forschung entwickelt und besteht aus den nachfolgend beschriebenen drei Modulen.

Spektrenvorhersage

Diese Funktion dient der Vorhersage von QToF-MS/MS-Spektren für mehrere Kollisionsenergien. Als Eingabeinformation wird der InChI- oder SMILES-Code benötigt. Eine Auswahl von Ionisierungmodus und Addukttyp ist ebenfalls möglich.

Die Spektren werden für niedrige (10 eV), mittlere (20 eV) und hohe (40 eV) Kollisionsenergien berechnet und durch eine Liste von "Masse-Intensitätspaaren" in zwei verschiedenen Formaten dargestellt: (1) eine herunterladbare Textdatei, die das vorhergesagte MS/MS-Spektrum enthält und (2) eine interaktive Grafik des vorhergesagten MS/MS-Spektrums in der jedem Spektrenpeak ein m/z-Wert, ein relativer Intensitätswert und eine Strukturformel des wahrscheinlichsten Fragmentions zugeordnet ist.

Es stehen über neun Millionen Spektren für über eine Million Substanzen zur Verfügung, von denen knapp 145.000 experimentell erzeugt wurden.

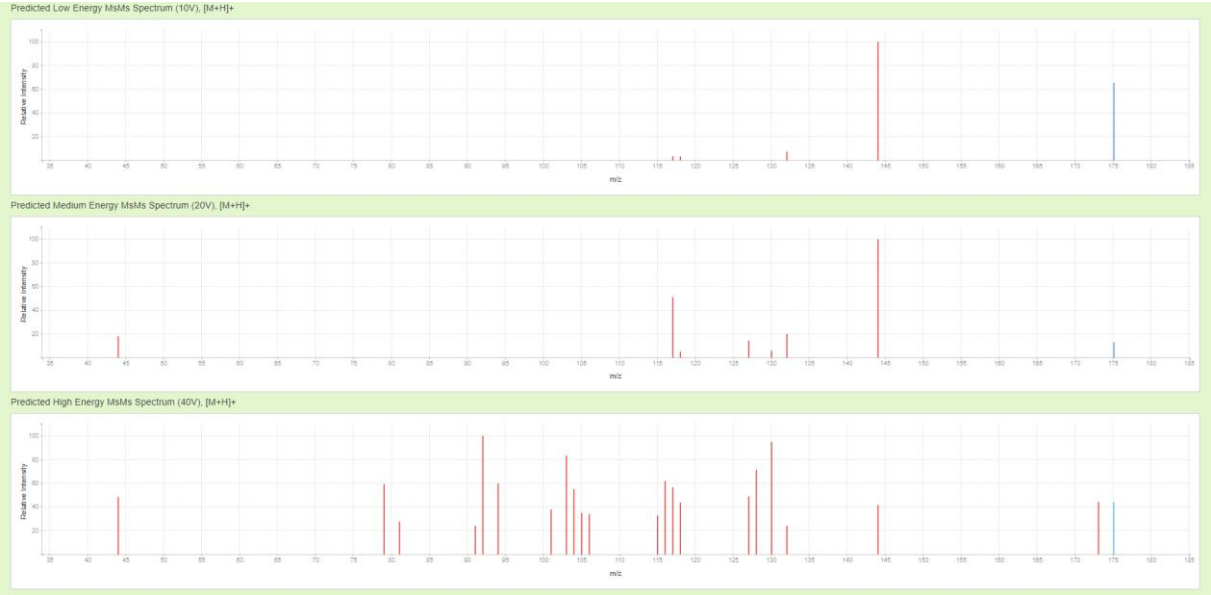

Abbildung 12: Von CFM-ID generierte Spektrenvorhersage für Methyltryptamin für drei Kollisionsenergien. Die Molekülstrukturen der Spektrenpeaks lassen sich mit einem *Mouseover* anzeigen und sind tabellarisch zusammengefasst.

Spektren-Peak-Annotierung

Mit dieser Funktion werden die Fragmentionen-Peaks in einem gegebenen MS/MS-Spektrum für ein bestimmtes Molekül annotiert. Der Benutzer muss die chemische Struktur (SMILES oder InChI) des fragmentierten Moleküls und eine Spektral-Peakliste aus einem MS/MS-Experiment bereitstellen. Nach der Übermittlung der Daten wird von CFM-ID eine vollständige Liste möglicher Fragmente berechnet. Anschließend werden die wahrscheinlichsten Fragmente für jeden Spektralpeak mithilfe eines vorab trainierten SE-CFM-Modells bestimmt. Die Ergebnisse liefern eine kommentierte Form der übermittelten experimentellen Spektren, wobei jedem Peak eine Liste von IDs angehängt ist, die allen Fragmenten mit den korrekten m/z entsprechen und in der Reihenfolge eines Wahrscheinlichkeits-Rankings aufgeführt sind. Es wird auch ein hypothetisches Fragmentierungsschema generiert, das alle kommentierten Fragmente und Fragmentierungswege enthält. Dieses wird in Form einer Liste von Fragmenten und einer Liste von Übergängen zwischen diesen Fragmenten bereitgestellt.

Identifizierung von Verbindungen

Diese Funktion ermittelt die Verbindungen, die am ehesten mit einem gegebenen MS/MS-Spektrum übereinstimmen. Die eingegebenen MS/MS-Spektren (mit einer oder mehreren Kol-

lisionsenergien) werden mit in silico vorhergesagten MS/MS-Spektren und/oder experimentellen MS/MS-Spektren aus 14 frei auswählbaren Datenbanken verglichen. Hypothetische Kandidaten werden entsprechend ihres Gesamtscores (Übereinstimmung) gereiht und zusammen mit vielen weiteren Informationen in Tabellenform angezeigt. Die Kandidatenspektren können darüber hinaus jeweils zusammen mit dem Input-Spektrum in einem MS-Spiegeldiagramm anzeigt werden.

Limitierungen

Zum Zeitpunkt der Testung im Januar 2025 war die Anwendung nicht voll funktionsfähig. Bei der Eingabe von MS/MS-Spektren und dem angestrebten Vergleich mit Spektren aus Datenbanken konnten die Anfragen nicht wie beschrieben ausgeführt werden. Bei der Nutzung der Beispieldatensätze trat derselbe Fehler auf. Nach mehreren Anfragen per E-Mail an die Betreiber der Seite wurde die Auskunft erteilt, dass Serverprobleme der Grund für die beobachteten Unzulänglichkeiten seien.

Weiterführende Informationen

CFM-ID bietet für jedes Modul eine bebilderte Eingabehilfe in der beispielhaft auch Ergebnisse dargestellt sind. Es besteht zudem die Möglichkeit Beispieldatensätze zu laden, um mit diesen die Anwendung zu testen. Darüber hinaus sind einige wissenschaftliche Veröffentlichungen zu CFM-ID erschienen von denen die meisten als Open-Access-Publikationen zur Verfügung stehen:

- Wang, F.; Allen, D.; Tian, S.; Oler, E.; Gautam, V.; Greiner, R.; Metz, T.O.; Wishart, D.S. (2022) CFM-ID 4.0–a web server for accurate MS-based metabolite identification. Nucleic Acids Research 50 (W1), W165-W174.
- Wang, F.; Liigand, J.; Tian, S.; Arndt, D.; Greiner, R.; Wishart, D. (2021) CFM-ID 4.0: More Accurate ESI MS/MS Spectral Prediction and Compound Identification. Anal Chem. 93(34):11692-11700.
- Djoumbou-Feunang, Y.; Pon, A.;, Karu, N.; Zheng, J.; Li, C.; Arndt, D.; Gautam, M.; Allen, F.; Wishart, D.S. (2019) Significantly Improved ESI-MS/MS Prediction and Compound Identification. Metabolites. 9(4):72.
- Allen, F.; Greiner, R. Wishart D.S. (2016) Computational prediction of electron ionization mass spectra to assist in GC-MS compound identification. Anal Chem. 88(15):7689-97.
- Allen, F.; Greiner, R.; Wishart, D. (2015) Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification. Metabolomics. 11:98–110.

- Allen, F.; Pon, A.; Wilson, M.; Greiner, R.; Wishart, D.S. (2014). CFM-ID: a web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra. Nucleic Acids Res. 42(Web Server issue):W94-9.

3.6 PFΔScreen

PFΔ*Screen* ist ein Python-basiertes Tool, das die Identifizierung von PFAS in HRMS-Rohdaten unterstützen soll und muss lokal installiert werden. Es kann unter folgendem Link als ZIP-Datei über den grünen „Code"-button von GitHub heruntergeladen werden: https://github.com/JonZwe/PFAScreen.

Die Installation kann auf Windows Betriebssystemen einfach per Doppelklick auf die Batch-Datei „Installation.bat" durchgeführt werden. Im Folgenden werden dann Python und die benötigten Packages automatisch installiert. Auf anderen Betriebssystemen können nach der Installation von Python die genutzten Packages für PFΔ*Screen* manuell installiert werden.

Die LC-HRMS Rohdaten können im mzML-Format eingelesen werden. Eine vorherige Konvertierung kann beispielsweise über die MSConvert-Software durchgeführt werden (siehe Exkurs in Kapitel 3.3). Als Ausgangspunkt werden momentan centroidierte mzML-Dateien, welche mittels dem DDA Verfahren aufgezeichnet wurden, akzeptiert.

PFΔ*Screen* kann in der momentanen Version (v0.1.1) für die Datenauswertung einzelner Messfiles verwendet werden (eine Probe und eine Blankkontrolle, keine automatische Auswertung ganzer Batches).

Für die Ausführung von PFΔ*Screen* muss die Batch-Datei „Run_PFAScreen.bat" geöffnet werden, eine einfache graphische Oberfläche öffnet sich, die zur Dateneingabe, Anpassung der Parameter und interaktiver Datenvisualisierung genutzt werden kann:

Nun können drei Schritte zur Priorisierung und Identifizierung von PFAS durchgeführt werden:

I. Featuredetektion, MS/MS Alignment und Blankkorrektur:

Es können die mzML-Dateien einer Probe und eines Blanks geladen werden, um chromatographische Features mittels OpenMS [3, 4] zu detektieren (Peaks mit zugehörigen Isotopen) gefolgt von einer MS/MS Zuordnung und einer optionalen Blankkorrektur (Run FeatureFinding). Optional kann auch eine CSV-Datei mit m/z, RT, Intensität und Intensität von M+1 Isotopen von einer Featuredetektion einer anderen Software zusammen mit den mzML-Dateien geladen werden, wenn ein anderer Algorithmus als OpenMS (FeatureFinderMetabo) erwünscht ist.

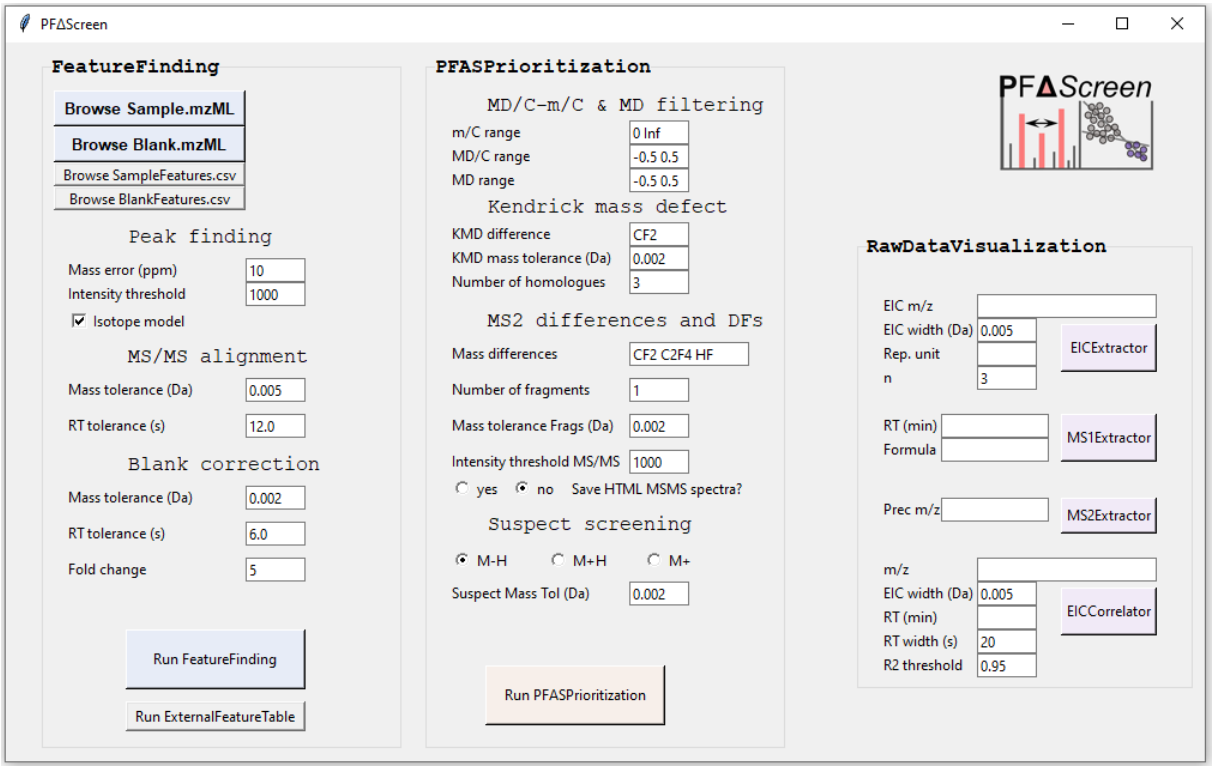

Abbildung 13: Graphische Benutzeroberfläche von PFΔ*Screen*. Die drei Schritte der Datenauswertung sind im Folgenden beschrieben

II. Priorisierung und tentative Identifizierung von PFAS:

Im zweiten Schritt können die detektierten Signale hinsichtlich potenzieller PFAS priorisiert werden. Dafür können folgende Techniken genutzt werden: Massendefektfilterung, Filterung mittels MD/C-m/C Ansatz [5, 6], Detektion von homologen Reihen (z. B. CF_2-Wiederholeinheit) über die Kendrickmassendefektanalyse, sowie im MS/MS eine Detektion von PFAS-typischen diagnostischen Fragmenten (zugrundeliegende Datenbank mit etwa 800 Fragmenten) und spezifischen Fragmentmassendifferenzen (z. B., ΔCF_2, ΔHF). Abschließend wird mit einer Suspectliste (leicht anpassbar, aktuell PFAS-NIST) über einen m/z-Abgleich und Isotopenabgleich nach potenziellen PFAS gesucht. Die Ergebnisse werden in einer Excel- und CSV-Tabelle gespeichert. Außerdem werden wichtige interaktive Plots (z. B. KMD, MD/C-m/C-Plot) gespeichert.

III. Datenvisualisierung:

Mittels der graphischen Benutzeroberfläche von PFΔ*Screen* können verschiedene Plots für eine Verifizierung generiert werden. Dazu zählen Extracted Ion Chromatograms (EICs), MS^1 und MS/MS-Spektren (mit detektierten Fragmenten und Massendifferenzen), Überlagerung von gemessenen mit theoretischen Isotopenmustern, sowie eine Korrelation von EICs, um Addukte und In-Source-Fragmente manuell zu detektieren.

Limitierungen

In der aktuellen Version (v0.1.1) können nur einzelne Messfiles (und ein Blank) analysiert werden. Die LC-HRMS Rohdaten können nicht als Eingangsdaten verwendet werden, sondern müssen zuvor in das mzML-Format konvertiert werden. Die Priorisierung mittels des MD/C-m/C baut auf eine akkurate Integration der M und M+1 Isotope auf (Abschätzung der Kohlenstoffanzahl). Es sollte also überprüft werden, ob der genutzte Algorithmus dafür genaue Ergebnisse liefert (z. B. mit Standards).

Weiterführende Informationen

Ein detailliertes Onlinetutorial über Installation und Funktionsweise (36 Minuten) ist unter folgendem Link zu finden: PFAScreen Tutorial — PFAS Feature Prioritization in Non-Target HRMS Data - YouTube.

Die Open-Access-Publikation von Zweigle et al. (2023) mit hilfreichen Erklärungen (auch in der SI) ist hier zu finden: https://link.springer.com/article/10.1007/s00216-023-05070-2.

Literatur:

- Zweigle, J., et al., PFΔScreen - an open-source tool for automated PFAS feature prioritization in non-target HRMS data. Anal Bioanal Chem, 2023.
- Chambers, M.C., et al., A cross-platform toolkit for mass spectrometry and proteomics. Nat Biotechnol, 2012. 30(10): p. 918-20.
- Röst, H.L., et al., OpenMS: a flexible open-source software platform for mass spectrometry data analysis. Nat Methods, 2016. 13(9): p. 741-8.
- Sachsenberg, T., et al., OpenMS 3 expands the frontiers of open-source computational mass spectrometry. Preprint, 2023.
- Kaufmann, A., et al., Simplifying Nontargeted Analysis of PFAS in Complex Food Matrixes. J AOAC Int, 2022. 105(5): p. 1280-1287.
- Zweigle, J., B. Bugsel, and C. Zwiener, Efficient PFAS prioritization in non-target HRMS data: systematic evaluation of the novel MD/C-m/C approach. Anal Bioanal Chem, 2023. 415(10): p. 1791-1801.
- Zweigle, J., B. Bugsel, and C. Zwiener, FindPFΔS: Non-Target Screening for PFAS - Comprehensive Data Mining for MS2 Fragment Mass Differences. Anal Chem, 2022. 94(30): p. 10788-10796.

3.7 FOR-IDENT

FOR-IDENT ist eine Online-Datenauswertungsplattform, die dabei hilft relevante Verbindungen in Proben zu identifizieren bzw. grundsätzlich zu charakterisieren. Die Plattform entstand

in vom BMBF geförderten Forschungsprojekten und wird vom zukünftigen Verein FOR-IDENT übernommen und weiter gepflegt werden.
Die Plattform kann über https://water.for-ident.org/ aufgerufen werden. Die Suchfunktionen können ohne Registrierung genutzt werden. Für die molekulare Screening-Funktion ist ein Login erforderlich; die Registrierung und Nutzung sind kostenlos.
Die Plattform kann mit verschiedenen Dateiformaten umgehen. Eine Übersicht über die unterstützten Formate findet sich auf der Hauptseite des molekularen Screening-Bereichs unter https://water.for-ident.org/#!processing. Die meisten der unterstützten Importformate basieren auf Exportformaten von NTS-Softwarepaketen. Damit soll eine möglichst einfache Integration von FOR-IDENT in NTS-Workflows ermöglicht werden.

Der Kern der Plattform besteht aus eine Reihe von anwendungsspezifischen Substanzdatenbanken, die mit analytischen Daten durchsucht werden können, um potenzielle Übereinstimmungen zu finden. Die Datenbanken umfassen wasserrelevante anthropogene Verbindungen (so die Datenbank STOFF-IDENT; (Letzel et al., 2015; Minkus et al., 2021)), natürlich in Gewässern vorkommende Verbindungen (PLANT-IDENT; Whaman et al., 2021 und POLLEN-IDENT) sowie per- und polyfluorierte Verbindungen (PFAS-IDENT; Ulrich et al., 2024). Weitere Datenbanken können in Zukunft hinzugefügt werden. Die Datenbanken sind mittels Suchfunktion nach bestimmten Verbindungen und Bedingungen durchsuchbar. Auch können nach einem Daten-Upload die detektierten Massen mittels einer molekularen Screening-Funktion mit Verbindungen in der Datenbank abgeglichen werden. Wenn im hochgeladenen Datensatz Informationen zur Retentionszeit und RTI-Kalibrierungsdaten verfügbar sind, ist es auch möglich diese zu verwenden, um Kandidatenverbindungen auf der Grundlage der Hydrophobizität in Bezug auf die Retentionszeit zu priorisieren (Minkus et al., 2021). Weiter können hochgeladene MS/MS-Spektren von Kandidatenverbindungen mit in silico-generierten MS/MS-Spektren (aus dem MetFrag-Tool) verglichen werden. Außerdem sind Verlinkungen zu Spektralbibliotheken (z. B. MassBank) und anderen Datenbanken (z. B. EPA Dashboard) enthalten.

Weiterführende Informationen

Ein Benutzerhandbuch kann unter https://water.for-ident.org/download/manual.pdf aufgerufen werden. Darüber hinaus sind einige wissenschaftliche Veröffentlichung zur Online-Datenauswertungsplattform FOR-IDENT erschienen:

- Letzel, T.; Bayer, A.; Schulz, W.; Heermann, A.; Lucke, T.; Greco, G.; Grosse, S.; Schüssler, W.; Sengl, M. and Letzel, M.: LC-MS Screening Techniques for Waste Water Analysis and Analytical Data Handling Strategies: Sartans and Their Transformation Products as an Example. Chemosphere 2015, 137, 198-206..

- Minkus, S.; Bieber, S. and Letzel, T.: (Very) polar organic compounds in the Danube river basin: a non-target screening workflow and prioritization strategy for extracting highly confident features. Analytical Methods 2021, 13(17), 2044 – 2054.
- Wahman, R.; Moser, S.; Bieber, S.; Cruzeiro, C.; Schröder, P.; Gilg, A.; Lesske, F.; and Letzel, T.: Untargeted analysis of Lemna minor metabolites: Workflow and prioritization strategy comparing highly confident features in between different mass spectrometers. Metabolites 2021, 11(12), 832.
- Ulrich, H.; Macherius, A.; Kunkel, U.; Sengl, M.; Letzel, T.: Novel PFAS-specific monitoring approach for highly impacted surface waters. Chemosphere 2024, 349 (February), 140893.

4 Zusammenfassung

Diese kurze Sammlung zeigt, dass verschiedene und frei verfügbare Werkzeuge durch Datenbankabgleiche, In-Silico-Fragmentierungen und/oder künstliche Intelligenz (KI) die Strukturaufklärung priorisierter Massen aus LC-HRMS-Messungen unterstützen können. Diese Ausführungen decken jedoch nur einen Teil des insgesamt zu Verfügung stehenden Werkzeugsortiments ab. Anwender sollten sich deshalb durch das Dokument nicht nur ermutigt fühlen, die vorgestellten Werkzeuge für ihre Zwecke zu testen, sondern auch weitere zu erkunden – auch in der Hoffnung, dass dieses Dokument durch neue Beiträge in zukünftigen Versionen weiterwachsen wird. Umfangreiche Zusammenstellungen bzw. Linksammlungen hierzu bieten BP4NTA (Additional Resources – BP4NTA) oder auch die Seite https://www.ms-utils.org/ (Software List — ms-utils.org).

Der nächste logische Schritt, um die Arbeit von HRMS- bzw. NTS-Anwendern weiter zu erleichtern, wäre die Entwicklung von NTS-Pipelines zur Unterstützung von kompletten Workflows, mit denen Daten auch batchweise abgearbeitet werden können. Für „echte“ Unbekannte, deren Strukturen in Stoffdatenbanken bisher nicht beschrieben sind, wären zudem niederschwellige Hilfen für die Interpretation von Massenspektren wünschenswert.

Danksagung

Wir danken Michael Stravs (Eawag) für die Zuarbeit zum MassBank-Kapitel, Jonathan Zweigle (Universität Kopenhagen) für die Ausführung zu PF∆Screen, Stefan Bieber (AFIN) für seinen Beitrag zu FOR-IDENT sowie Kai Dührkop (Universität Jena) für seine Hilfe bei der Erstellung des Textes zu SIRIUS