

Das Dilemma unbekannter EI-Massenspektren in Wasserextrakten:

- Wege und Hilfen zur Identifizierung von Stoffklasse,
Teilstruktur oder komplettem Strukturvorschlag**

Das Dilemma unbekannter EI-Massenspektren in Wasserextrakten: Wege und Hilfen zur Identifizierung von Stoffklasse, Teilstruktur oder komplettem Strukturvorschlag



Michael Oehme
Institut für Angewandte Analytische Chemie
CH-9052 Niederteufen AR
und
Universität Basel

Das „normale“ Vorgehen:

Suche der Spektren in der NIST-Bibliothek im Standardmodus („identity“).
→ Viele Vorschläge, die nicht passen.

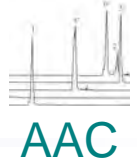
Was nun?

Gibt es Möglichkeiten, Informationen über Teilstrukturen, funktionelle Gruppe oder über das Kohlenstoffgerüst der unbekanntes Verbindung zu erhalten?

Wie kann man diese abrufen?

Wichtig: Man muss wissen, wie die Suchalgorithmen funktionieren und wie man Teilinformation erhalten kann!

Was geschieht eigentlich bei der Bibliothekssuche?



3 Suchalgorithmen:

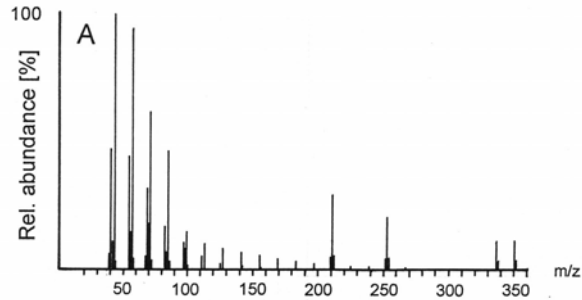
- **Komponentenvergleich („identity, forward search, fit“):**
Ist das unbekannte Spektrum in der Bibliothek enthalten? Harter Suchalgorithmus, ergibt recht sichere Resultate, versagt aber bei schlechten Spektren oder ähnlichen Verbindungen. Fokussiert auf Moleküle (falls vorhanden).
- **Bibliotheksvergleich („similarity, reversed Search, retrofit“):**
Sind Teile eines oder das gesamte Bibliotheksspektrum im unbekanntem Spektrum enthalten? Kann hilfreiche Hinweise ergeben, wenn ähnliche Verbindungen, aber nicht die Gesuchte in der Datenbank vorhanden sind. Suchresultate müssen wesentlich kritischer interpretiert werden.
- **Fragmentierungsmustervergleich:**
Sind typische Fragmentierungsmuster und Information über Teilstrukturen im Massenspektrum vorhanden? Kann sinnvolle Strukturinformationen geben, ohne dass das Spektrum in der Datenbank vorhanden ist.

- Da die Suche mit dem vollständigen Spektrum zu zeitaufwendig ist, wird dieses zunächst auf die signifikanteste Information reduziert.
- Die dadurch erhaltenen kondensierten Spektren von etwa 15-30 Massen werden dann mit einer auf die gleiche Weise reduzierten Bibliothek verglichen.

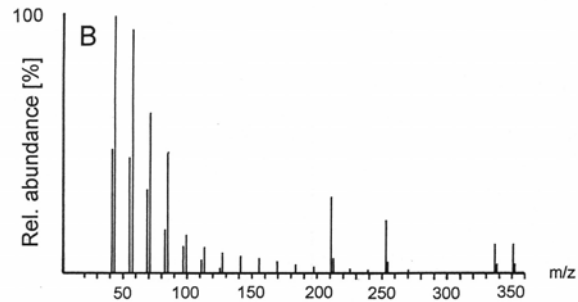
Eine tückische Falle: Was ist signifikant?

- Signifikant = Intensiv: Ganz falsch.
- Signifikant im Vergleich zu Nachbarfragmenten (Bereich 14 u): Wesentlich besser.
- Was sonst noch... ?

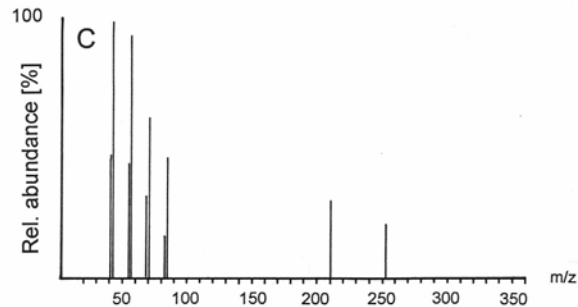
Reduktion eines MS auf signifikante Information



Originalspektrum

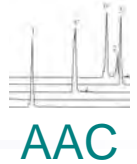


1-2 Massen pro Bereich von 14 u



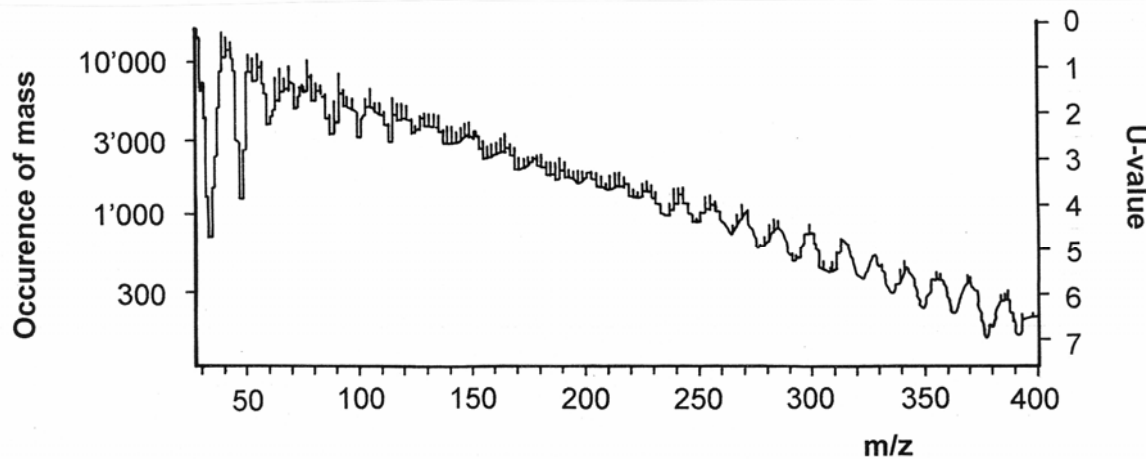
Die 10 intensivsten Massen

Reduktion eines MS auf signifikante Information II



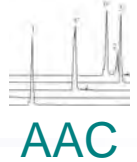
Viel besser:

Reduktion durch Auswahl von 10-30 Fragmentationen mittels eines Algorithmus, der sowohl Intensität als auch Signifikanz der Masse berücksichtigt: Gewichtung jeder Masse bezüglich ihre statistische Häufigkeit (U) und Intensität (A).



Häufigkeitsverteilung von Masse und Intensität bei 19'000 Massenspektren

Gefährlich: Suche mit reduzierten MS



Die dadurch erhaltenen kondensierten Spektren von etwa 15-30 Massen werden mit einer auf die gleiche Weise reduzierten Bibliothek verglichen („presearch on“).

Die dadurch vorselektionierten MS werden dann mit vollständigen MS verglichen.

Suchresultate werden meist innerhalb von 1-5 s erhalten und nach einem Ähnlichkeitsindex sortiert.

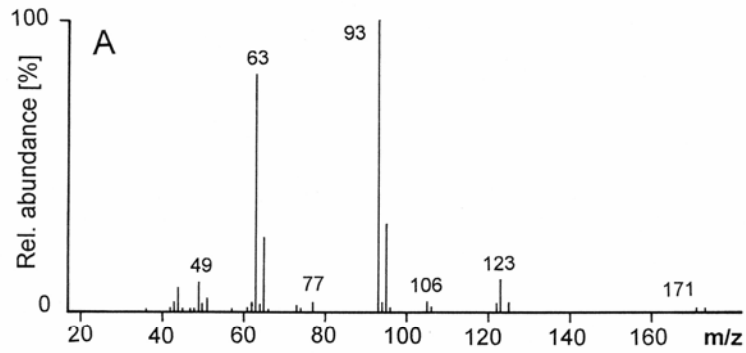
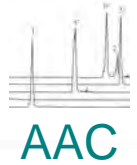
Kann schiefgehen, d.h. falsche Datenreduktion und komische Resultate (siehe Beispiel).

Daher: Versuche abschalten!

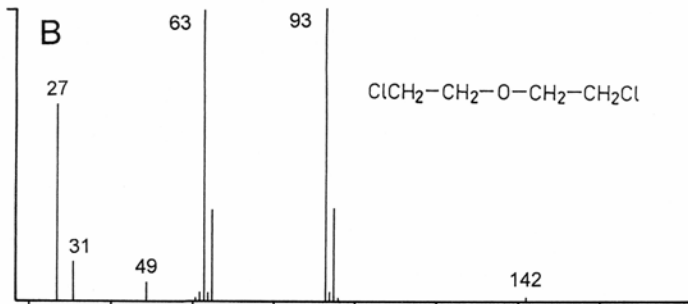
Resultat: Erste Suche ca. 1 min, alle weiteren Suchen, ca. 5 s. Warum?

Bei der ersten Suche werden parallele Teilsuchen angelegt. Danach wird die Bibliothek in Untergruppen verglichen.

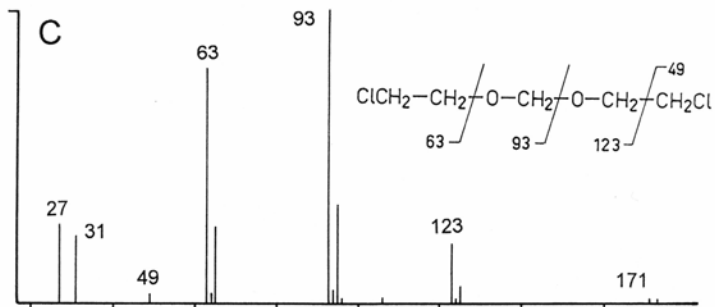
Suchresultat mit und ohne Versuche



A: Probenspektrum

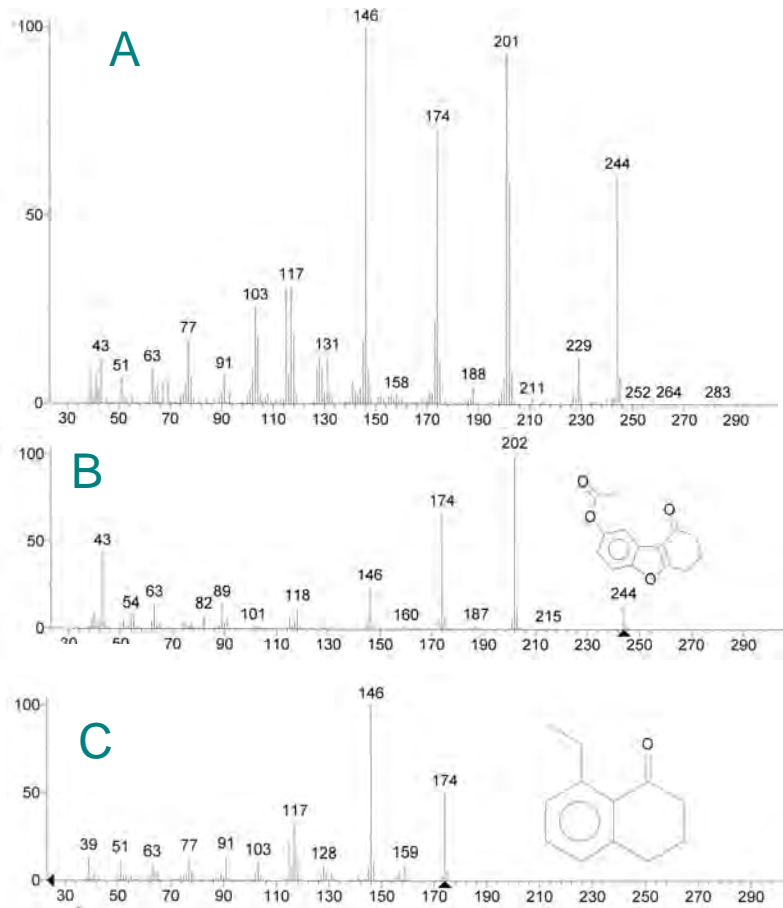


B. Suchresultat mit Versuche "an"



C: Suchresultat mit Versuche "aus"

Vergleich: Suche mit "identity" und "similarity"

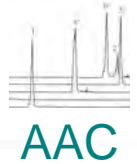


A: Probenspektrum, nicht in der Datenbank vorhanden.

B: Suchresultat auf der Basis "identity"

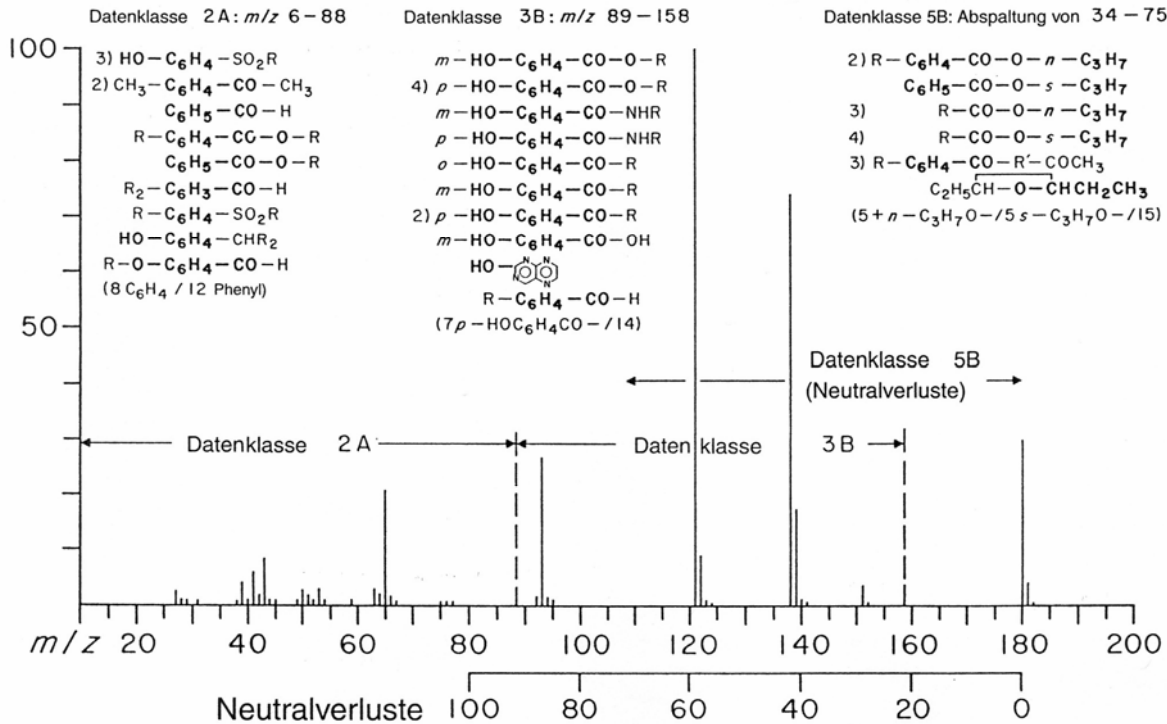
C: Suchresultat auf der Basis "similarity"

Fragmentierungsmustervergleich: Wie geht das?



Zwei Möglichkeiten:

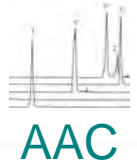
- STIRS ("Self training Interpretative and Retrieval System"), Teil in NIST-"browser".
- SISCOM ("Search for Identical and Similar Compounds"), Teil von MASSLIB.



Vergleich von Teilbereichen des MS mit Referenzbibliothek

Ergibt Informationen über Teilstrukturen und funktionelle Gruppen, die nach Wahrscheinlichkeit geordnet sind.

Wo finde ich das STIRS-Teilprogramm?



The screenshot displays the NIST MS Search 2.0 interface. At the top, the title bar reads "NIST MS Search 2.0 - [Simple, Presearch Default - 100 spectra]". The main window shows a mass spectrum plot for "1. ES-8-pH2d#868 RT: 12.25 AV: 1". The x-axis represents m/z from 30 to 220, and the y-axis represents relative intensity from 0 to 100. The base peak is at m/z 111. Other significant peaks are labeled at m/z 34, 38, 42, 46, 50, 54, 58, 63, 67, 75, 81, 85, 89, 93, 99, 105, 121, 127, 139, 144, 157, 170, 175, 190, and 207. A red circle highlights the search icon in the toolbar.

Below the mass spectrum, the "Substructure Information" window is open. It displays the following data:

Name of Unknown: ES-8-pH2d#868 RT: 12.25 AV: 1

Chlorine/Bromine information: Cl=1, Br=0 Probability=86%; Probability of presence of Cl=91%, of Br=0%

MW	Prob.
190	89
189	1
191	1

Substructural information:

Prob.	Present	Prob.	Absent
99	ArS	99	N
99	halogen	99	NoAr
99	ClBr	99	Ar-C
99	S	99	PhCsat
99	Cl	99	PhCO
99	Ar-S	99	Si(CH3)3
99	S-C	99	ArOR
98	Ph-all	99	hetcyc
98	O	99	PhOCH3

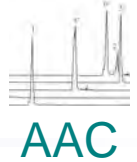
Set of Substructures in use:

#	Substruct.
1	OH
2	CO2H
3	ArOH
4	ROH
5	SH
6	7OH
7	SiH3
8	CH3
9	NH2

The window also shows a list of substructures and their corresponding chemical structures. The first structure is "1 (replb) Benzene, 1-c hloro-... 980 982R" with a chemical structure of 4-chlorobenzenesulfonamide. The second structure is "2 (mainlb) Benzene, 1-c hloro-... 924 954R" with a chemical structure of 4-chlorobenzenesulfonamide. The third structure is "3 (replb) Benzenesulfonami... 845 903R" with a chemical structure of benzenesulfonamide.

Gut versteckt und sehr schlechte graphische Oberfläche....

Zurück zum Beispiel: Anwendung von STIRS



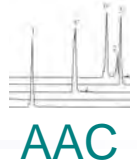
Structures present:

Number	Probability	Short Name And Description
1	98	RDB5_PLUS, rings + double bonds count >= 5
2	96	CH2/3, primary or secondary saturated carbon
3	94	O, contains oxygen
4	94	>C=O, carbonyl group
5	92	CH2, methylene group
6	91	.C=O, ring carbonyl
7	91	AR, aromatic ring
8	86	-CH2/3-, methylene or methyl group (chain)
9	86	CH3, methyl
10	82	C2H5/CH2CH2, primary or secondary ethyl linkage
11	71	hetcyc, ring containing at least one heteroatom
12	71	het_ring, ring containing non-carbon atom(s)
13	66	CH2C=O, methylene carbonyl bond
14	63	C3+-alkyl, methylene attached to saturated C-atoms (chain)
15	59	CH&O, contains only carbon, hydrogen and oxygen
16	54	CH2.C=O, methylene carbonyl bond (ring)
17	52	CO.CH2.CH2, methylene attached to methylene and carbonyl (ring)
18	52	>=8-ring, ring containing 8 or more atoms
19	51	-CH2/3-1, exactly one methylene or methyl group (chain)
20	49	ArC, aromatic carbon attached to carbon

Die gefundenen Strukturelemente stimmen gut mit der gefundenen Teilstruktur mit dem Modus "similarity" überein.

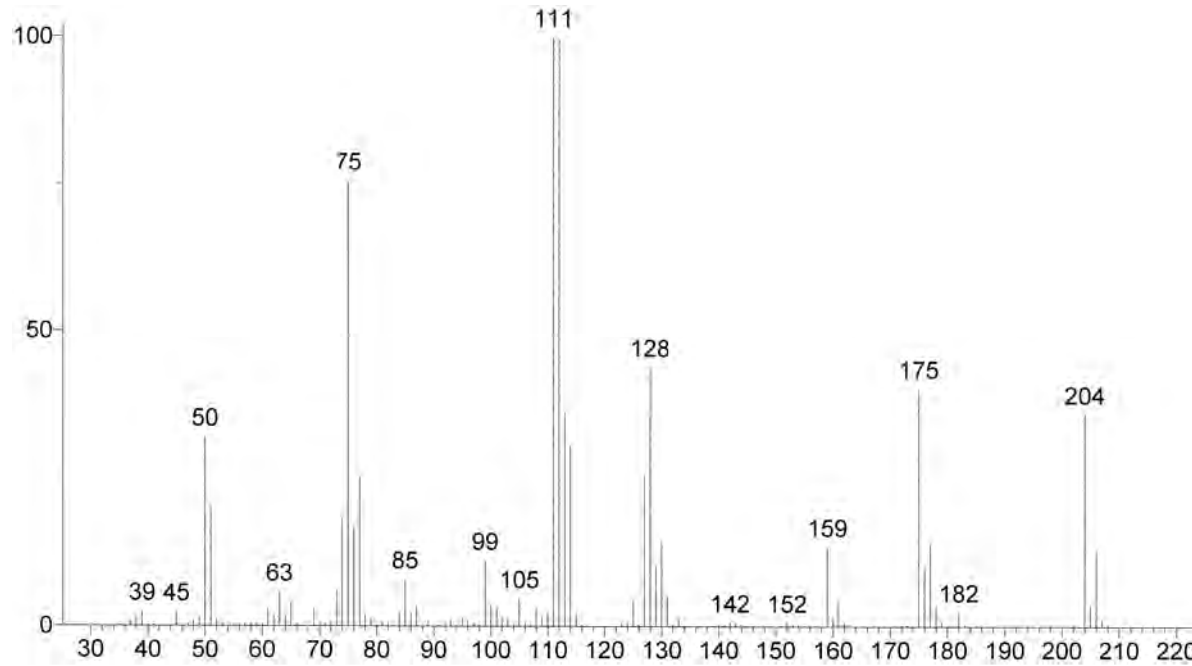
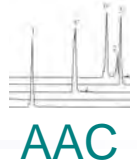
Zusätzlich werden Informationen über das wahrscheinliche Molekulargewicht etc. gegeben

Ein weiteres Problem: Was ist identisch?

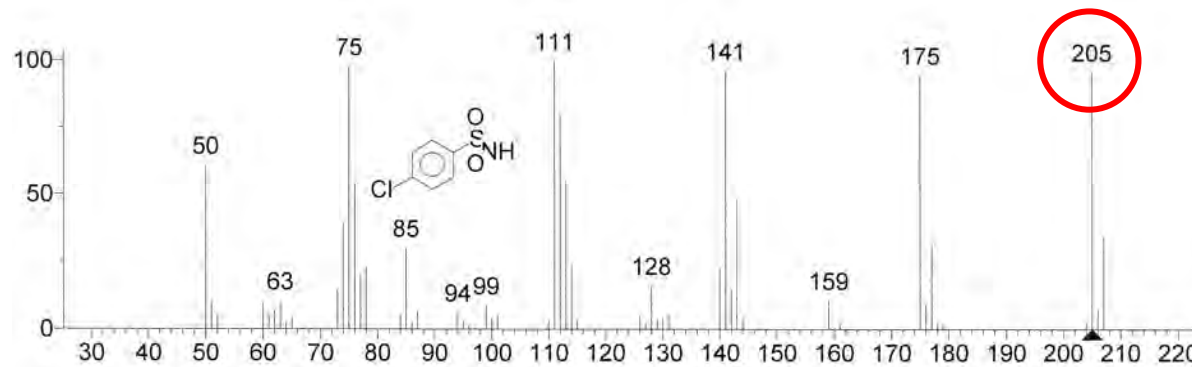


- Variable Qualität der Bibliotheksspektren (Instrumentbedingungen, Aufnahmequalität, Qualitätskontrolle bei der Aufnahme in die Datenbank). Siehe Beispiel.
- Eigene Aufnahmebedingungen wie z.B. das Intensitätsverhältnis zwischen niedrigem und hohem Massenbereich (abhängig von der Ionenquellen/-Massenfilteroptimierung).
- Konzentrationsveränderungen während der Registrierung (z.B. an GC-Signalfanke) ergeben Intensitätsverschiebungen (können teilweise durch Korrekturfaktor ("tilting") eliminiert werden).
- Überladung der Ionenquelle sowohl bei eigener Registrierung als auch im Bibliotheksspektrum (siehe Beispiel).
- Fehler bei der Hintergrundsubtraktion (Fehlen wichtiger Ionen geringer Intensität).

Beispiel: Überladenes Spektrum in der NIST-Bibliothek

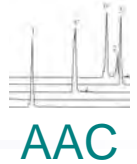


Probenspektrum



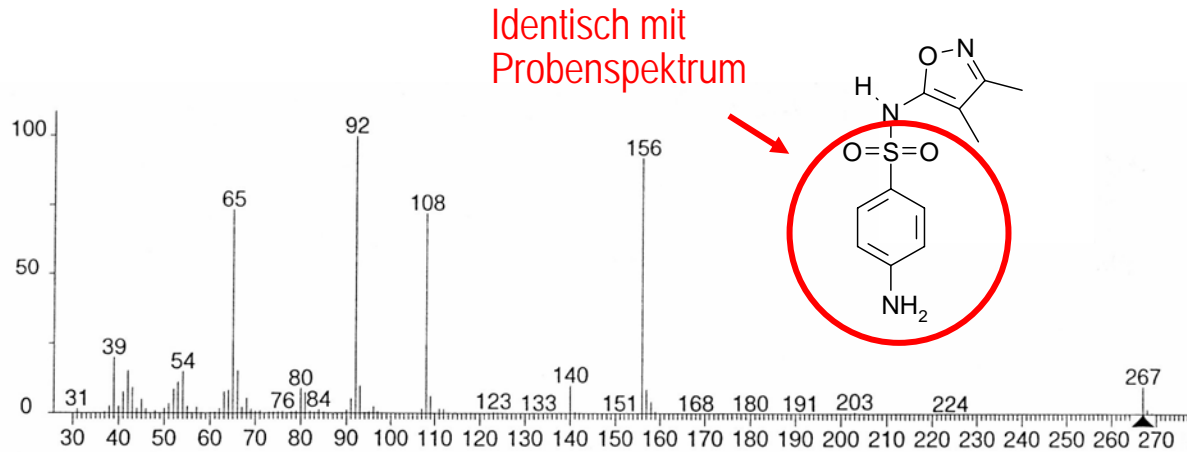
NIST-Spektrum

Bevor man Unbekannte interpretiert: Plausibilitätstest!



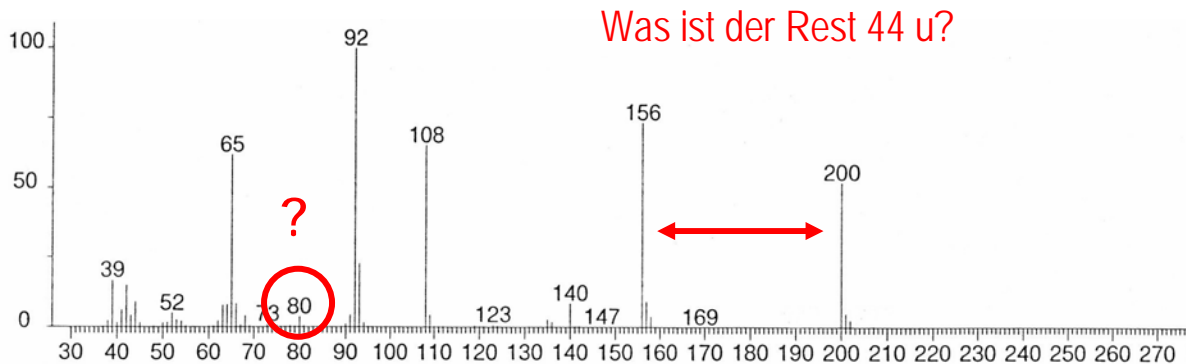
- Elution einer Verbindung mit M^+ m/z 278 bei 70 °C?
- Polare Verbindung in unpolarer HPLC-Fraktion?
- Ist das Vorhandensein realistisch? (Vulkanisierungsbeschleuniger in Butter?)
- Kann die Verbindung ein System(Geister)signal sein? (Kontaminierung während Probeaufarbeitung).

Auf dem Weg zur Struktur: Weiterinterpretation



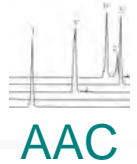
Strukturvorschlag mit "identity".

95% Übereinstimmung mit Sulfisoxazol

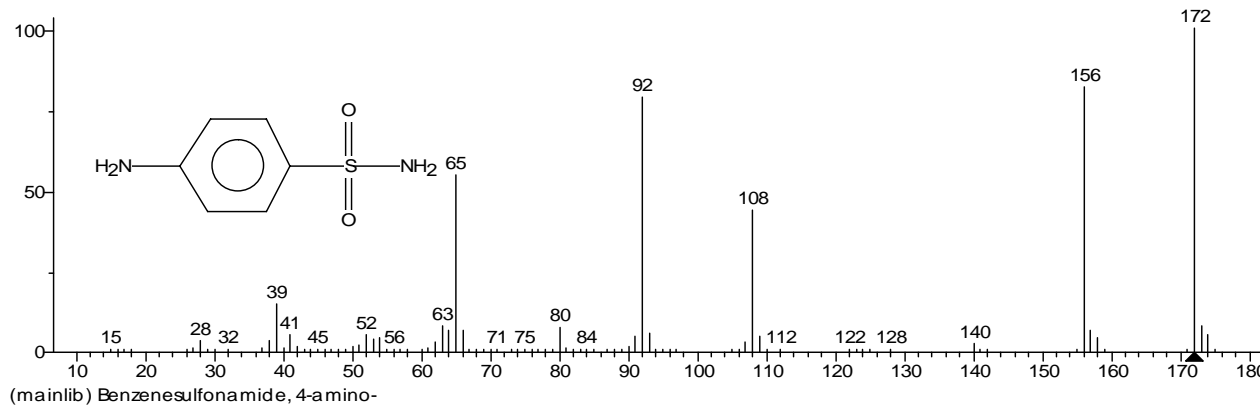


Probenspektrum

Rest 44 u und m/z 80?

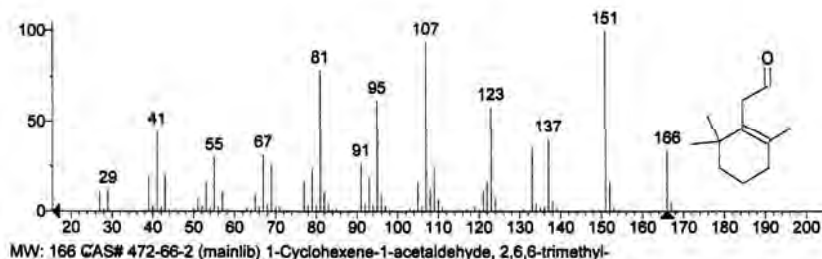
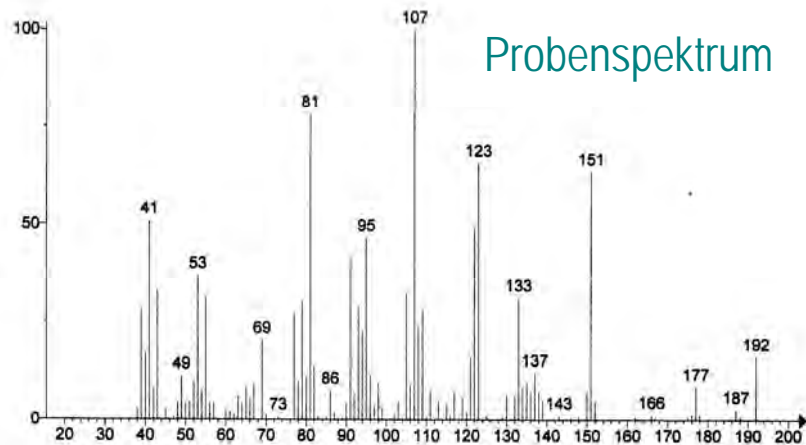


- Die unbekannte Verbindung enthält 2 Stickstoffatome (gerade Molekülmasse), eines davon im Rest 44 u (200-156 u). Differenz 30 u, CH₂O. Damit ist der Rest formell -NH-CHO.
- Da das Massenspektrum wie Sulfanilamid (unten) m/z 80 (-SO₂-NH₂) zeigt, ist die Struktur wahrscheinlich N4-Formyl-Sulfisoxazol (CHO-NH-Phenyl-SO₂-NH₂), ein Metabolit von Sulfisoxazol (Abbau durch Bodenorganismen?)



Manchmal das Wichtigste bei Grundwasserscreening:

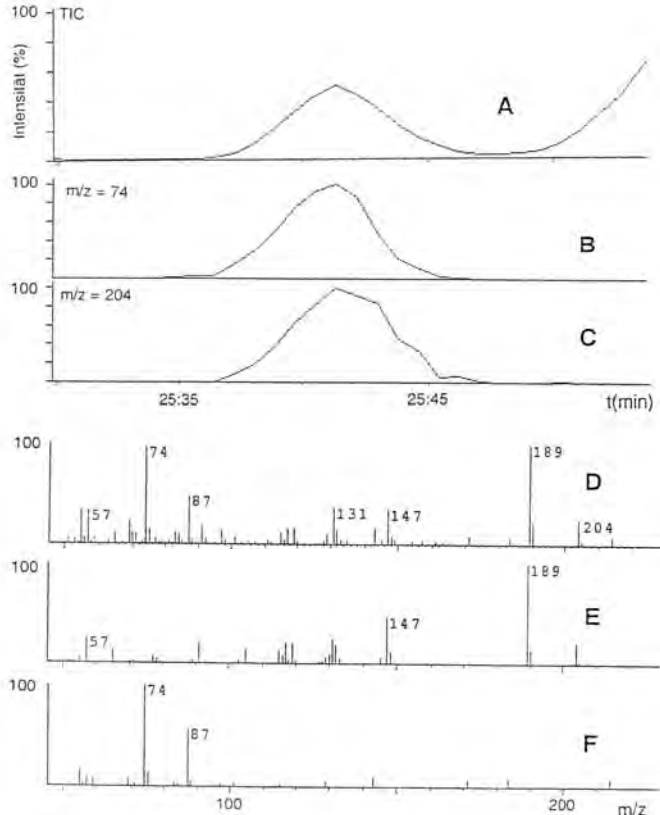
Biogen oder anthropogen?



Vorschlag mit "similarity"

- Das Massenspektrum einer unbekannteren Verbindung zeigt die typische Fragmentierung einer terpenoiden Struktur.
- Da isomere Terpenstrukturen fast identische Massenspektren besitzen, ist kein detaillierter Strukturvorschlag möglich. Es langt jedoch, die Stoffklasse festzustellen (vermutlich biogen ev. Parfümierung).
- Der Unterschied in der Molekülmasse (unbekannt 192 u, Vorschlag 166 u) wird durch einen zusätzlichen ungesättigten Kohlenwasserstoffrest verursacht. Eine weitere Interpretation ist in diesem Fall nicht nötig.

Ko-Elutionen: "similarity" findet beide Verbindungen



Massenspektren von einem Waschnittelextrakt.

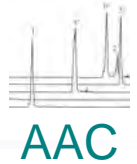
A: Totalionenstromchromatogramm, B und C: Massenchromatogramme

D: Massenspektrum vom Maximum des GC-Signals.

E: Vorschlag Lilial der Suche mittels „similarity“

F: Vorschlag Dodekansäuremethylester nach Abzug des Lilialspektrum und erneuter Suche.

Und schliesslich: Die eigene Erfahrung ist entscheidend



Beurteilung des Rohspektrums:

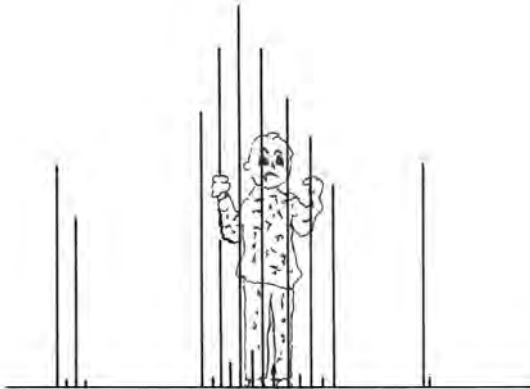
- Molekülion vorhanden?
- Vernünftige Fragmentierungen?
- Hat man eine eigene Meinung über die Stoffgruppe?
- Entspricht die Molekülmasse in etwa der zu erwartenden Retentionszeit?
- Wahl von signifikanten Ionen:
Koelutionen? Kann man die Verbindungsspektren voneinander subtrahieren?
- Was ist Hintergrund?

Beurteilung der Spektrensuche:

- Molekülion sicher erkannt?
- Stimmen nur Teile des Massenspektrums gut überein?
- Genügend Information vorhanden für eine sichere Identifizierung?
- Wird eventuell ein Gemisch erkannt und verschiedene Verbindungen vorgeschlagen?

- Die Bibliothekssuche kann wesentlich mehr und ergibt auch bei unbekanntem Verbindungen hilfreiche Informationen zur Strukturaufklärung.
- Die Identifizierung von Unbekanntem braucht Zeit und Erfahrung und kann nicht „so nebenbei“ gemacht werden.
- Die Kenntnisse, wie man (praxisnahe) EI-Spektren interpretiert, sind langsam am Aussterben.

Oder kürzer: Die Kunst der MS-Interpretation



Sich nicht durch die Information eingesperrt fühlen, sondern....



überflüssige Information eliminieren.



damit virtuos spielen, und....