Identification of unknown organic compounds based on comparison of electron ionization mass spectra

Andrey S. Samokhin
Igor A. Revelsky

Chemistry Department
Lomonosov Moscow State University

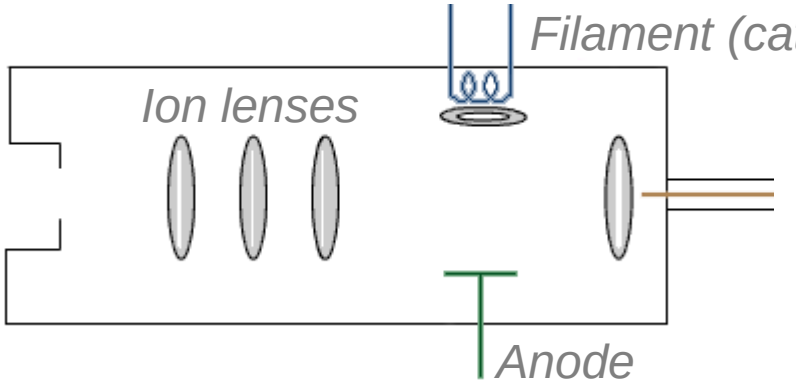
Identification of unknown organic compounds

- Nuclear magnetic resonance (**NMR**) spectroscopy
- Infrared (**IR**) spectroscopy
- Ultraviolet (**UV**) spectroscopy
- Mass spectrometry (**MS**)
  - High-performance liquid chromatography/mass spectrometry (**HPLC/MS**)
    - Electrospray ionization
    - Atmospheric pressure chemical ionization
    - Atmospheric pressure photoionization
  - Gas chromatography/mass spectrometry (**GC/MS**)
    - Electron ionization (**EI**)
    - Chemical ionization

Features of electron ionization mass spectrometry (70 eV)

$$M + e^- \rightarrow M^{+\bullet} + 2e^-$$

Electron ionization ion source

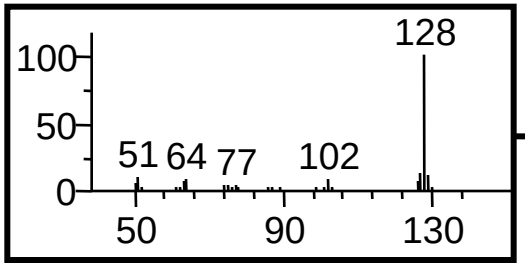


- Formation of a large number of fragment ions
- High (interlaboratory) reproducibility of mass spectra (in comparison with other ionization techniques)

Electron ionization mass spectral databases

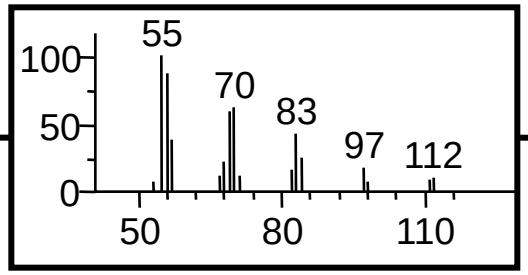
- **NIST'11** (213000 compounds)
- **Wiley 9th** (592000 compounds)

Identification based on search against mass spectral database (Library search)



Experimental mass spectrum of unknown compound

Mass spectral database



Similarity index

...

Result of library search

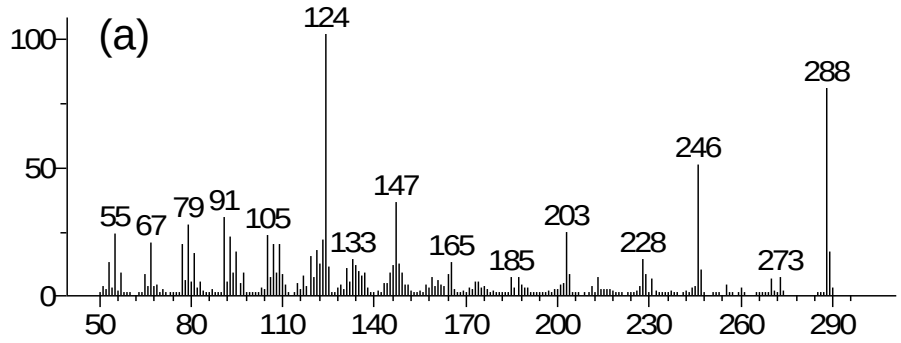
- Result of library search is the list of possible candidates (“Hit table”). Possible candidates are ranked by Similarity index
- “Hit table” obtained using MS Search 2.0 software:

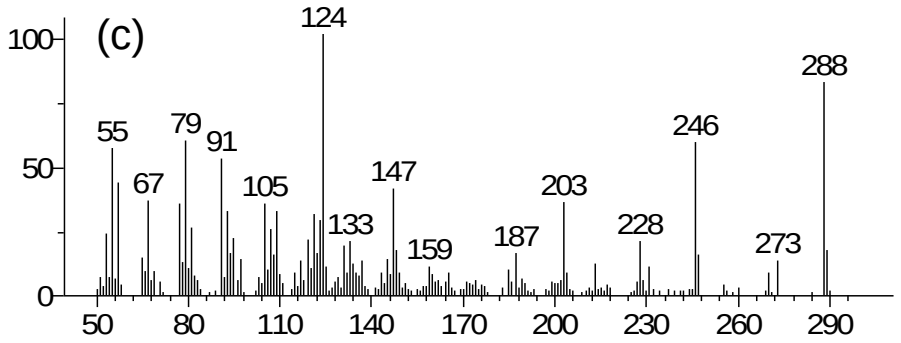
#	Lib.	Match	R.Match	Prob.	Name
1	M	855	855	27.4	Nonanoic acid, methylester
2	M	849	849	21.6	Undecanoic acid, methylester
3	M	845	845	18.2	Decanoic acid, methylester
4	M	838	838	14.0	Tridecanoic acid, methylester
5	M	819	819	6.77	Dodecanoic acid, methylester

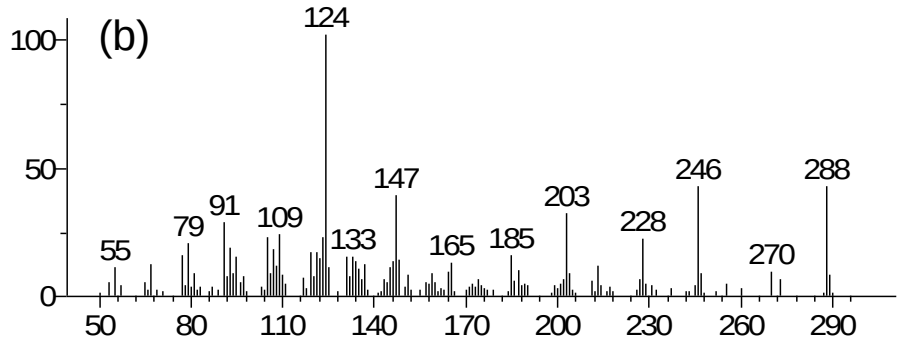
Correct compound does not occupy the first position in the “hit table” in approximately 25% of cases[1]

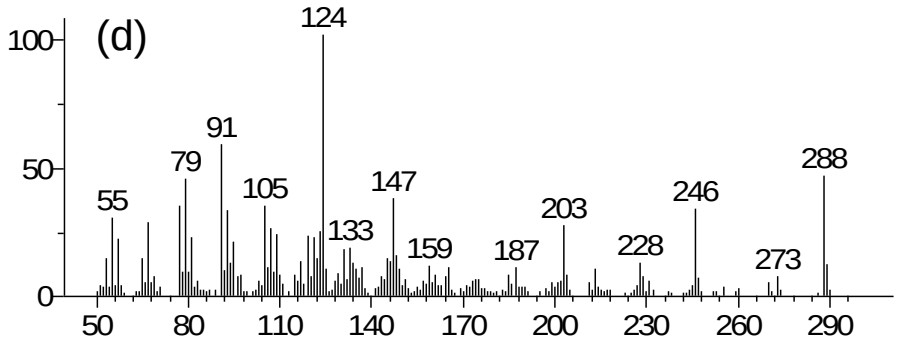
[1] Stein S.E., Scott D.R., *J. Am. Soc. Mass Spectrom.*, 1994, vol. 5, p. 859–866.

Differences between experimental and library mass spectra of testosterone









(a) – NIST mass spectral database (http://webbook.nist.gov/)
(b) – mass spectra registered by us using quadrupole instrument “DSQ” (Thermo)
(c) – Thevis M. Mass Spectrometry in Sports Drug Testing, Wiley, 2010. 360 p.
(d) – Sigma-Aldrich, Product Information, Testosterone (T 5411).

Dissimilarity of registration conditions

- Technique used for introducing sample into a mass spectrometer
  - direct inlet
  - combination with gas chromatograph
- Temperature of ion source
- Type of mass analyzer
  - quadrupole
  - time-of-fligth
  - ion trap
  - magnetic sector
- Method of instrument calibration

Prediction of electron ionization mass spectra

Electron ionization mass spectra cannot be fully predicted by experts and simulated using any software:

- Theory of molecular ion fragmentation gives only general information about set of ions
- Software (e.g. Mass Frontier, ACD Labs, MOLGEN) can be used to predict only the set of *m/z* values corresponding to possible fragment ions (intensities of mass spectral peaks cannot be predicted)

Therefore identification of unknown compounds (by means of electron ionization mass spectrometry) should be based on comparison with experimental mass spectra

Reliable identification by means of electron ionization mass spectrometry

For the most reliable identification by means of electron ionization mass spectrometry the full mass spectra of unknown compound and possible candidate should be registered under identical experimental conditions and compared

The main problem is how to compare full mass spectra, which may contain hundreds of peaks

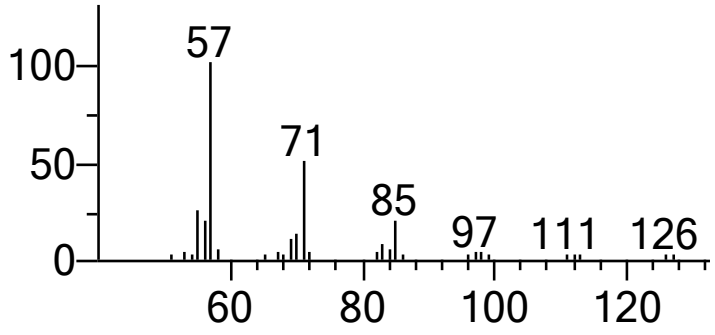
Possible approaches for mass spectra comparison

- Visual comparison
- Application of mathematical algorithms used in library search programs
- Comparison of intensities of individual mass spectral peaks

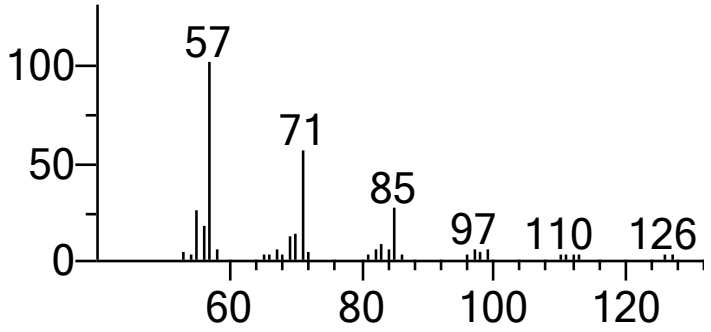
Visual comparison

- Visual comparison is very subjective
- Mass spectra similarity cannot be estimated quantitatively

Do these mass spectra correspond to the same compound?

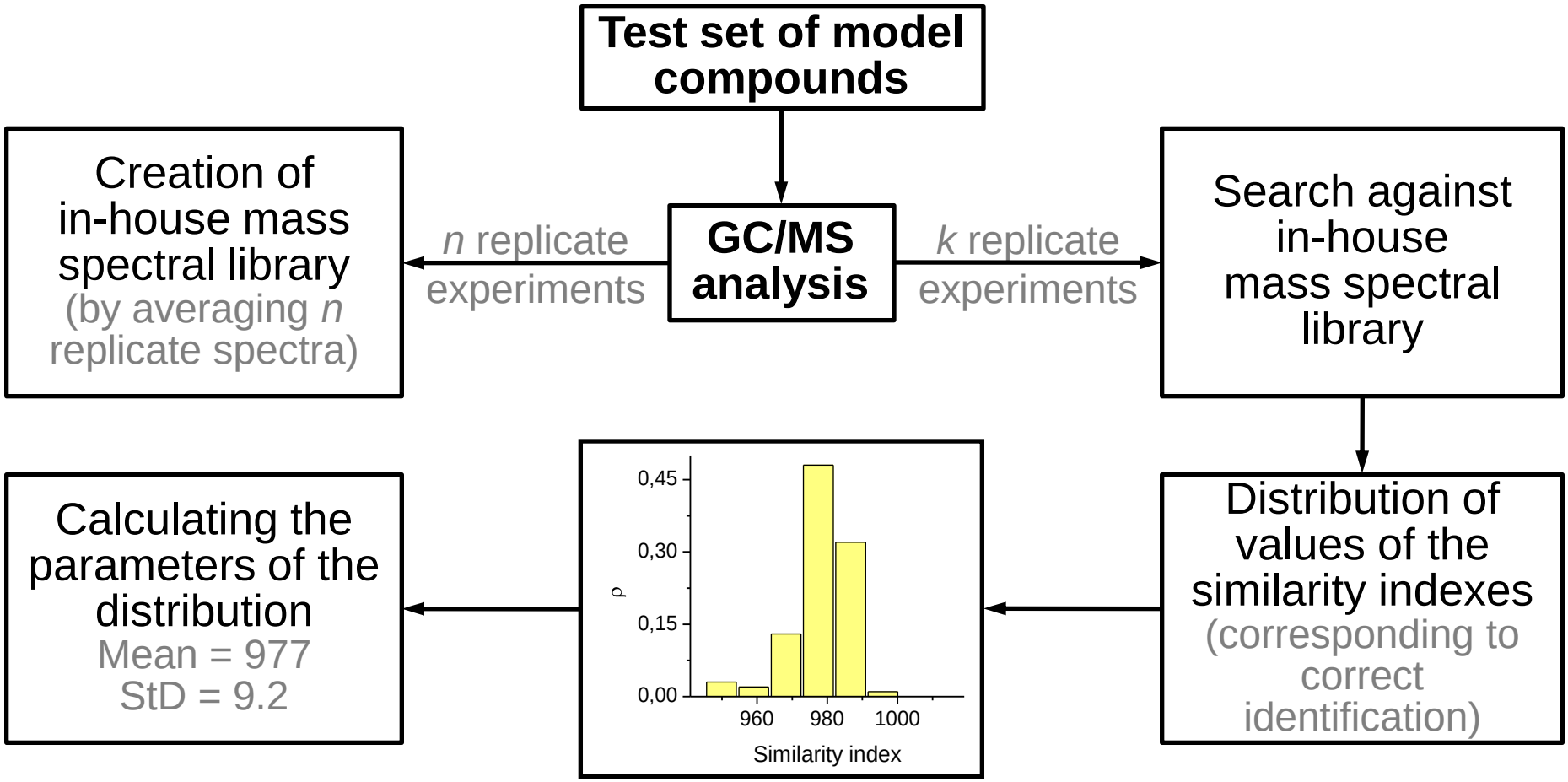


***n*-Dodecane**
(MW = 170)



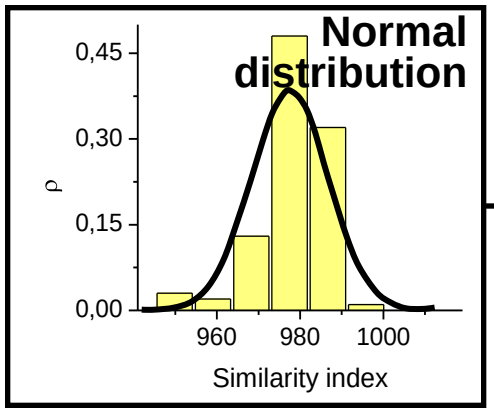
***n*-Tridecane**
(MW = 184)

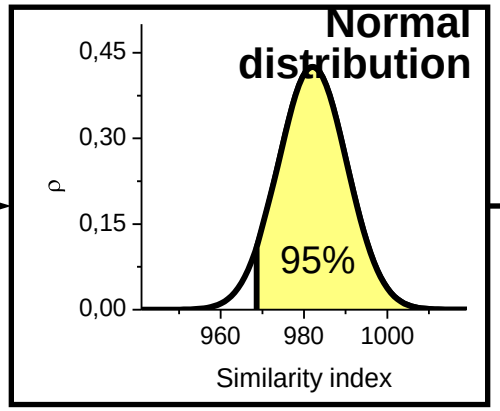
Application of mathematical algorithms used in library search programs



Samokhin A.S., Revelsky I.A., Revelsky A.I. Book of Abstracts. Fourth All-Russian Conference "Fundamental issues of mass spectrometry and its analytical application". October 10–14, 2010, Zvenigorod, Russia. P. 58–59.

Application of mathematical algorithms used in library search programs





Critical value of **Similarity index = quantile$_{95\%}$** (corresponding to reliable identification)

Mass spectra correspond to the same compound, if similarity index between these spectra is more than critical value (equaled to quantile$_{95\%}$)

Samokhin A.S., Revelsky I.A., Revelsky A.I. Book of Abstracts. Fourth All-Russian Conference "Fundamental issues of mass spectrometry and its analytical application". October 10–14, 2010, Zvenigorod, Russia. P. 58–59.

Application of mathematical algorithms used in library search programs

The main limitation is that the magnitude of critical value depends on registration conditions

- Magnitudes of critical values calculated for different instruments are not equal

- The experiments (including registration of hundreds of spectra) should be repeated every time, when conditions are changed
    - ion source contamination
    - method of instrument calibration
    - etc.

Samokhin A.S., Revelsky I.A., Revelsky A.I. Book of Abstracts. Fourth All-Russian Conference "Fundamental issues of mass spectrometry and its analytical application“. October 10–14, 2010, Zvenigorod, Russia. P. 58–59.

Comparison of intensities of individual mass spectral peaks

Tolerance windows for relative intensities of mass spectral peaks[1–3]

- only several (usually 3) mass spectral peaks are considered
- width of tolerance window is quite large

Intensity[3] (normalized to the base peak)	**Width of window[3]**
>50%	$\pm 10\%_{relative}$
20–50%	$\pm 15\%_{relative}$
10–20%	$\pm 20\%_{relative}$
≤10%	$\pm 50\%_{relative}$

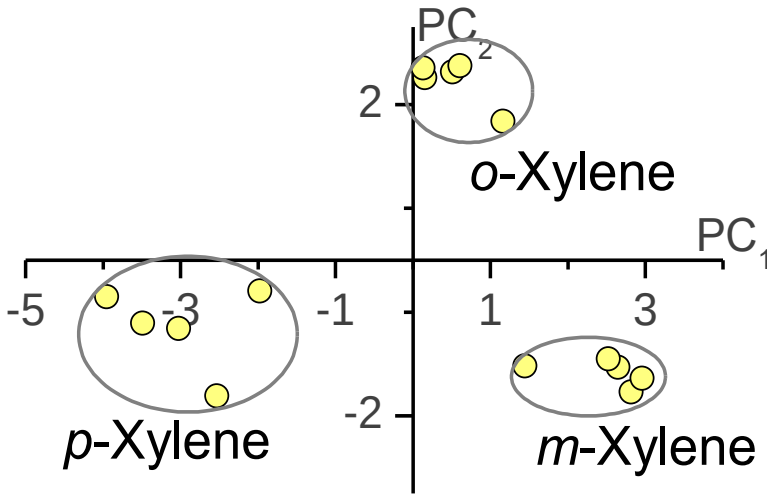
[1] Commission Decision 2002/657/EC, *Off. J. Eur. Commun.*, 12 August. 2002

[2] Guidance for industry: mass spectrometry for confirmation of the identity of animal drug residues: final guidance. FDA, Washington, DC; 2003.

[3] WADA Technical Document—TD2010IDCR, 2011.

Distinguishing electron ionization mass spectra of isomers by PCA

- Hejazi et. al have used PCA to distinguish geometrical isomers of α-linolenic acid methyl ester[1]
- We have shown possibility of distinguishing between *o*-xylene, *m*-xylene and *p*-xylene[2]



[1] Hejazi L., Ebrahimi D., Guilhaus M., Hibbert D.B., *J. Am. Soc. Mass Spectrom.*, 2009, vol. 20, p. 1272–1280.

[2] Samokhin A., Revelsky I., *Eur. J. Mass Spectrom.* 2011, vol. 17, p. 477–480.

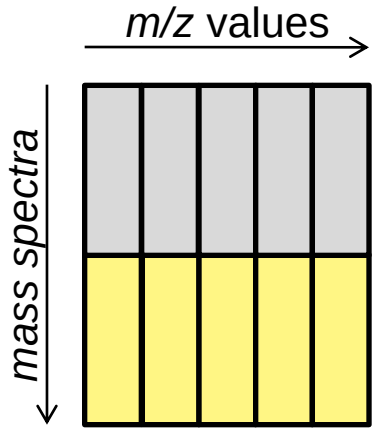
Goal

The goal of this work was to develop simple approach (based on using principal component analysis) for reliable comparison of mass spectra registered under identical experimental conditions

Scheme of analysis

n_1 replicate mass spectra of unknown compound

n_2 replicate mass spectra of possible candidate



Principal component analysis

Intensities of mass spectral peaks were normalized to the total ion current:

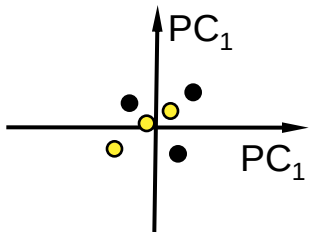
$$I^{\text{relative}}_{m/z=j} = \frac{I^{\text{absolut}}_{m/z=j}}{\sum_i I^{\text{absolut}}_{m/z=i}} \cdot 100\%$$

What can we expect?

Principal component analysis

Unknown compound **is** possible candidate

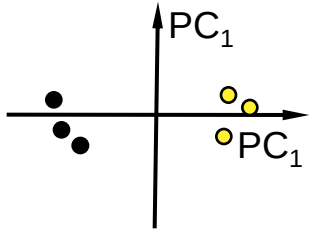
Score plot



1. There is only one group of objects on the score plot
2. All PCs contain only noise

Unknown compound **is not** possible candidate

Score plot

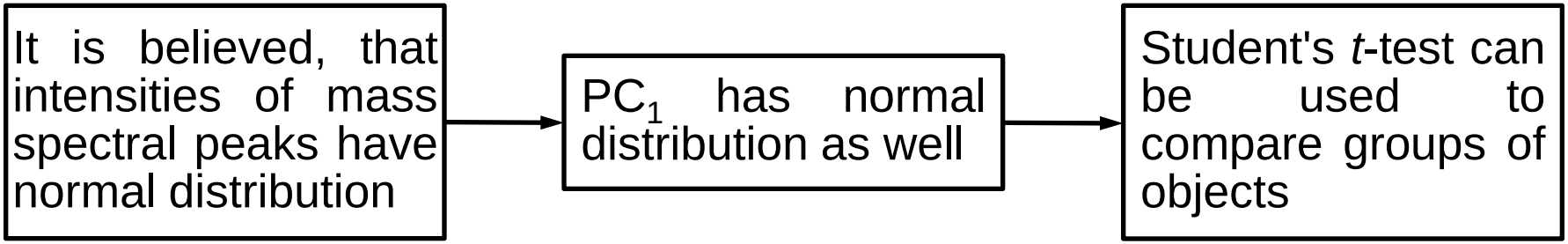


1. There are two groups of objects on the score plot
2. PC_1 should separate objects corresponding to different groups

Therefore we can use only PC_1 to compare mass spectra

The first principal component

$$PC_1 = p_1 \cdot I_1 + p_2 \cdot I_2 + \ldots + p_n \cdot I_n$$

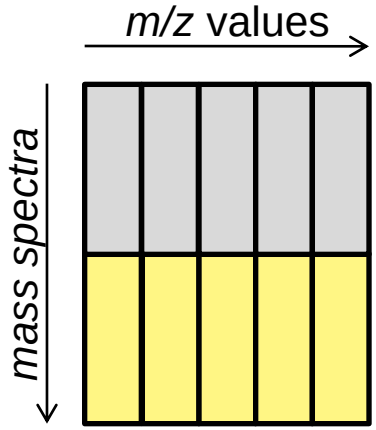


Therefore we transform multidimensional analytical signal to the one-dimensional one

Scheme of analysis

n_1 replicate mass spectra of unknown compound

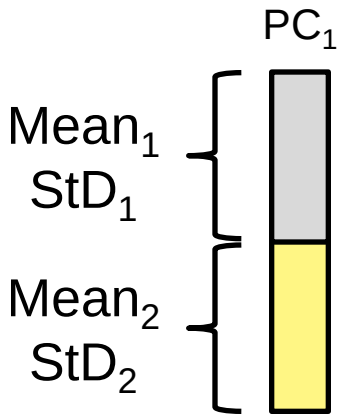
n_2 replicate mass spectra of possible candidate



Principal component analysis

Student's *t*-test:

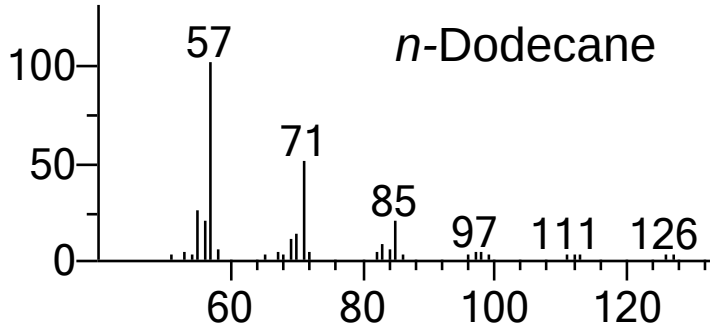
$$\xi = \frac{|\text{Mean}_1 - \text{Mean}_2|}{\text{StD}} \cdot \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}$$

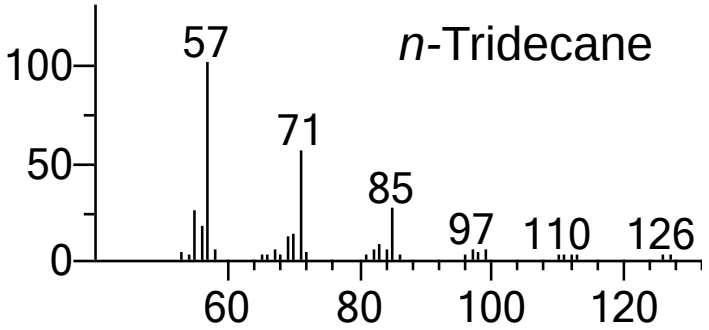


If $\xi \leq t\,(P, f = n_1 + n_2 - 2)$,
unknown compound **is** possible candidate

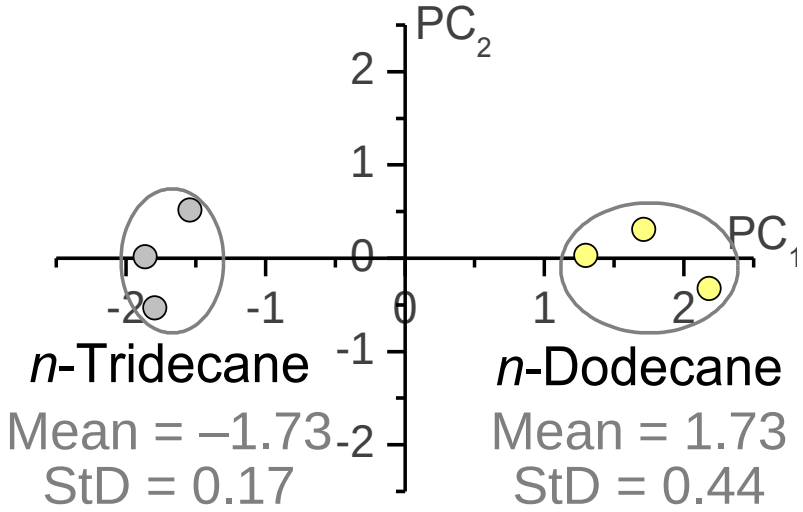
If $\xi > t\,(P, f = n_1 + n_2 - 2)$,
unknown compound **is not** possible candidate

Example #1 (discrimination of hydrocarbons)





Mass spectra corresponded to 1 ng of respective analyte
23 mass spectral peaks were taken into account

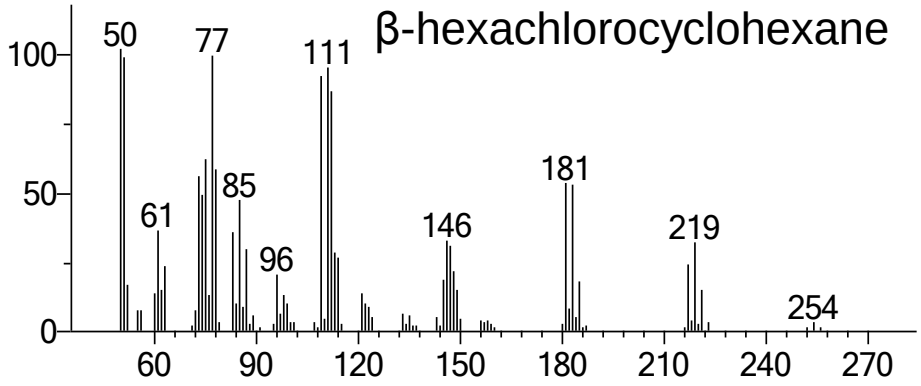


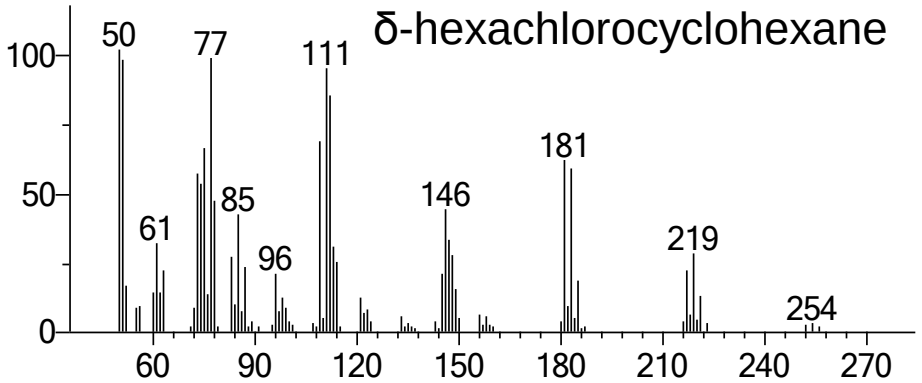
$\xi = 12.7$

P	*t* (*P*, 4)
0.95	2.8
0.99	4.6
0.999	8.6

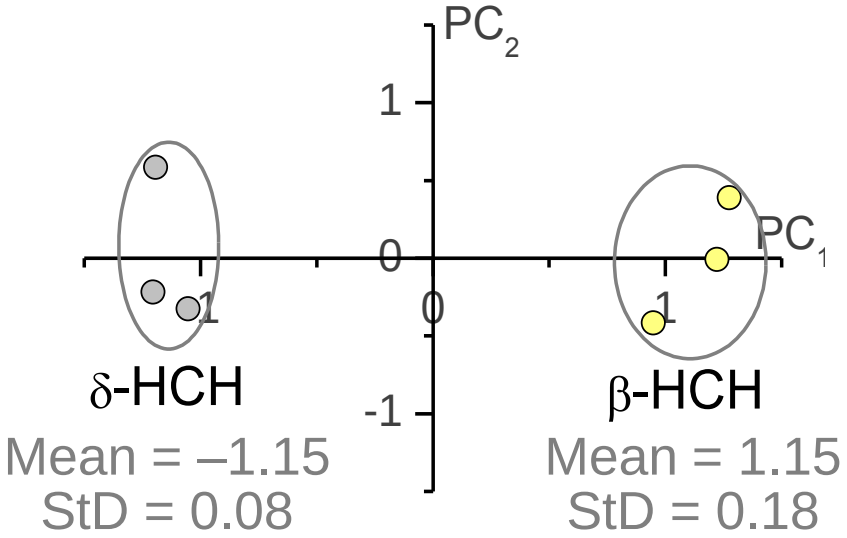
Mass spectra correspond to the different compounds

Example #2 (discrimination of β-HCH and δ-HCH)





Mass spectra corresponded to 0.5 ng of respective analyte
45 mass spectral peaks were taken into account



ξ = 20.5

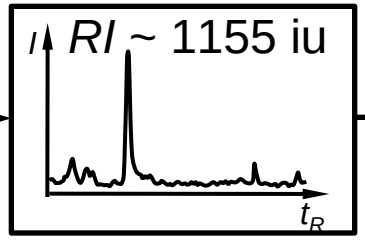
P	*t* (*P*, 4)
0.95	2.8
0.99	4.6
0.999	8.6

Mass spectra correspond to the different compounds

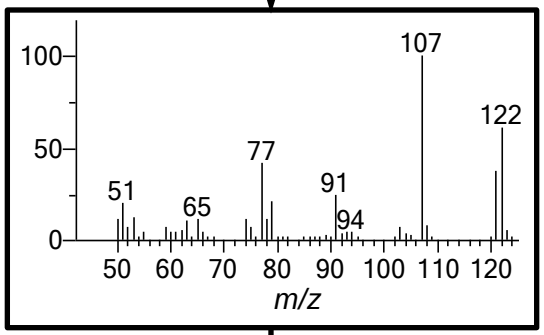
Example #3 (identification of "unknown" compound)

"Unknown" sample

GC/MS analysis



Extraction of pure mass spectrum



Name	*RI* [1], iu
2,6-xylenol	1115
2,4-xylenol	1150
2,5-xylenol	1152
3,5-xylenol	1170
2,3-xylenol	1180
3,4-xylenol	1196

Library search against mass spectral database

#	Name	Similarity index
1	2,3-xylenol	896
2	2,5-xylenol	888
3	2,6-xylenol	875
4	2,4-xylenol	870
5	3,4-xylenol	859

[1] Mjos S., Meier S., Boitsov S, *J. Chromatogr. A* 2003, vol. 1123, p. 98–105.

Example #3 (identification of "unknown" compound)

"Unknown" compound

Possible candidate (2,5-xylenol; C = 0.1 ng uL^{-1})

GC/MS analysis

GC/MS analysis

Retention times

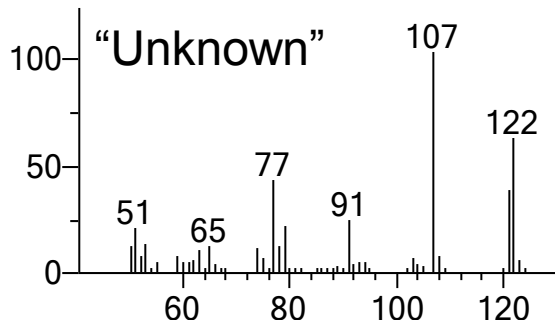
$t_R^{(\text{"unknown"})}$ = 261.70 sec

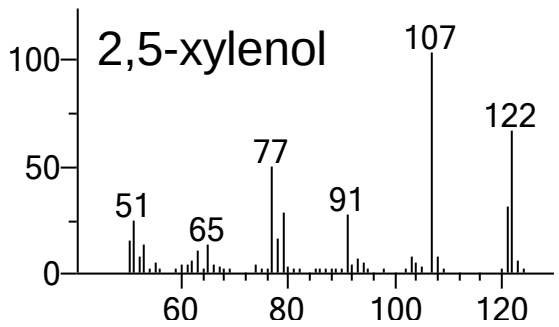
$t_R^{(2,5\text{-xylenol})}$ = 261.85 sec

Δt_R = 0.15 sec

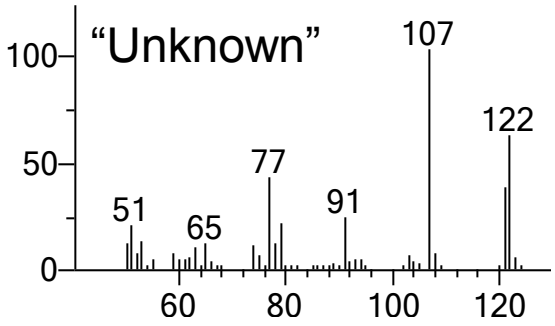
Δt_R = 0.06%

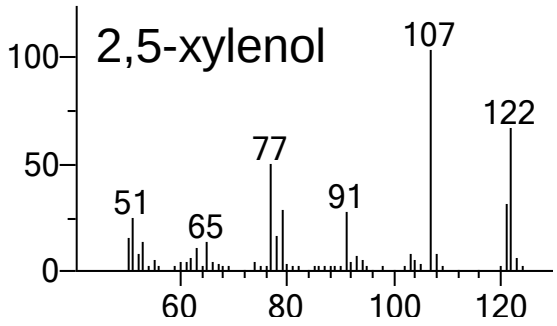
Mass spectra





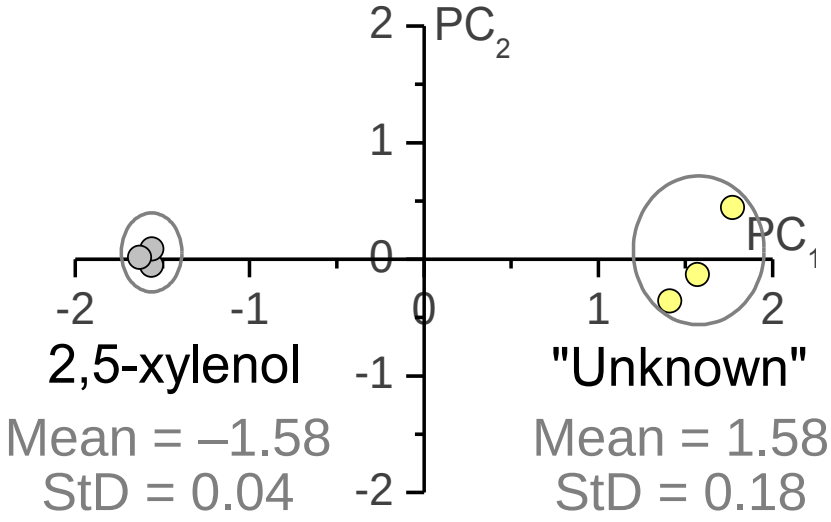
Example #3 (identification of "unknown" compound)





Intensities of chromatographic peaks (of "unknown" compound and 2,5-xylenol) were approximately equal

26 mass spectral peaks were taken into account



$\xi = 29.8$

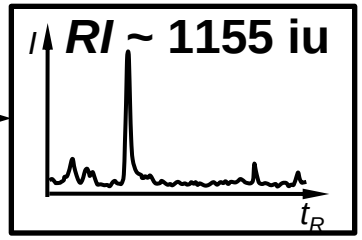
P	*t* (*P*, 4)
0.95	2.8
0.99	4.6
0.999	8.6

Mass spectra correspond to the different compounds

Example #3 (identification of “unknown” compound)

“Unknown” sample

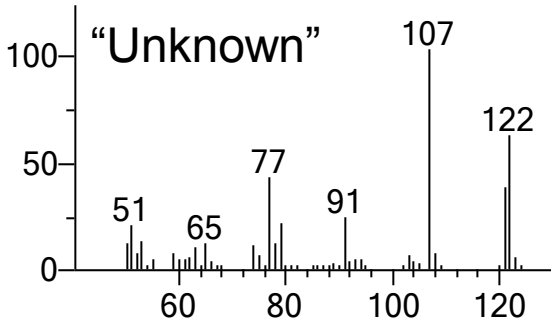
GC/MS analysis

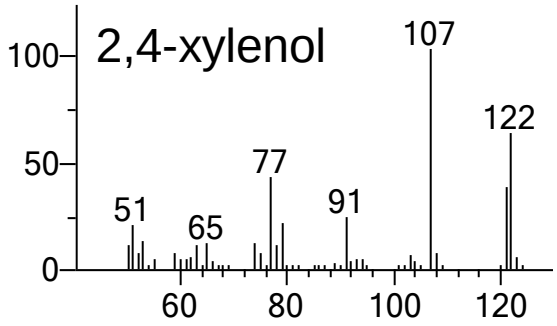


Name	*RI* [1], iu
2,6-xylenol	1115
2,4-xylenol	1150
2,5-xylenol	1152
3,5-xylenol	1170
2,3-xylenol	1180
3,4-xylenol	1196

[1] Mjos S., Meier S., Boitsov S, *J. Chromatogr. A* 2003, vol. 1123, p. 98–105.

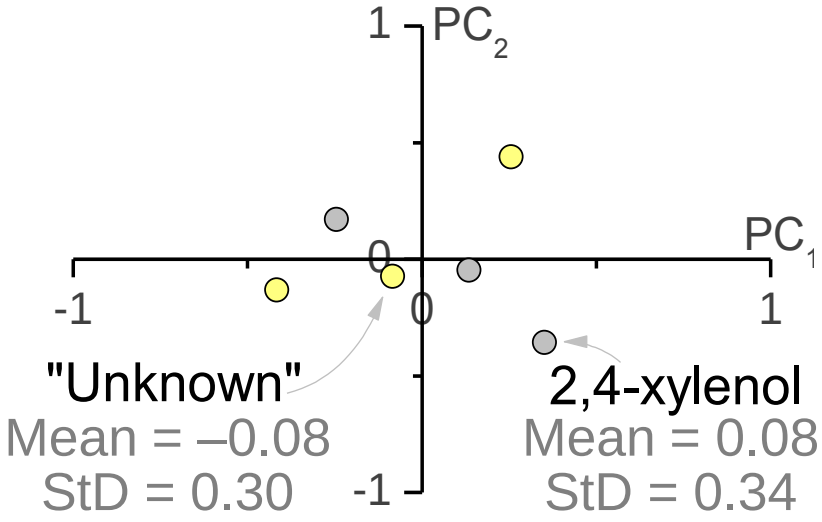
Example #3 (identification of "unknown" compound)

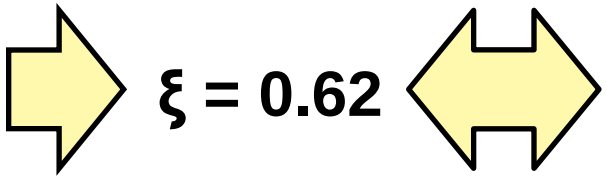




Intensities of chromatographic peaks (of "unknown" compound and 2,4-xylenol) were approximately equal

26 mass spectral peaks were taken into account



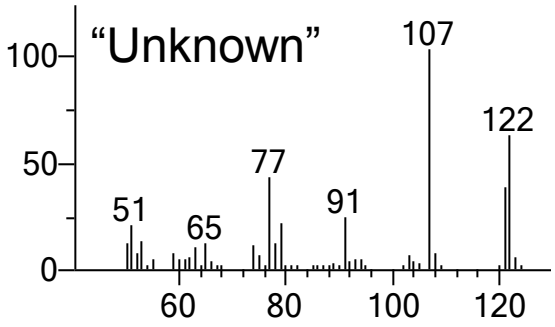


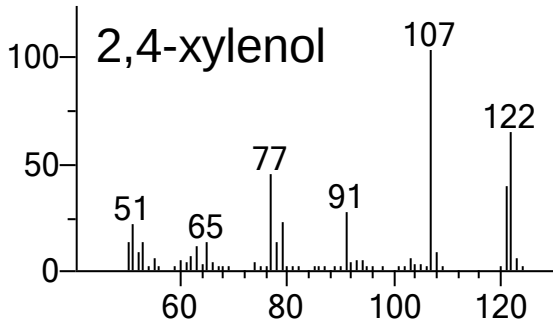
ξ = 0.62

P	*t* (*P*, 4)
0.95	2.8
0.99	4.6
0.999	8.6

Mass spectra correspond to the same compound

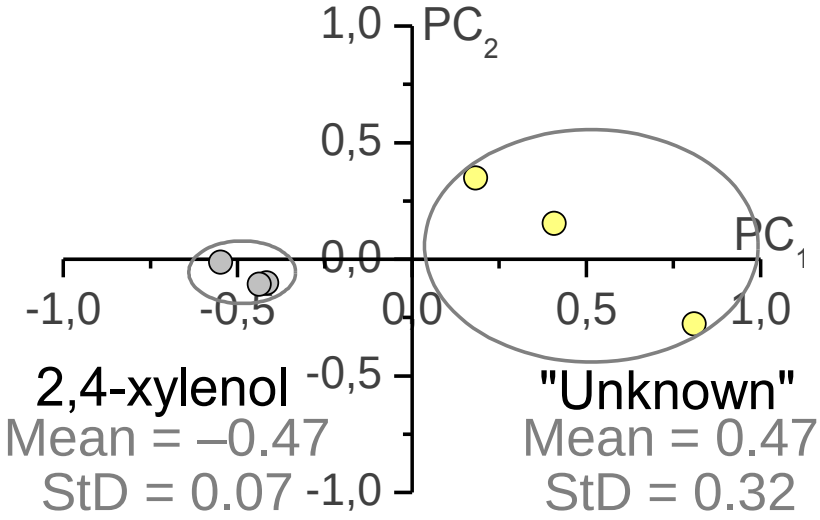
Example #3 (identification of “unknown” compound)

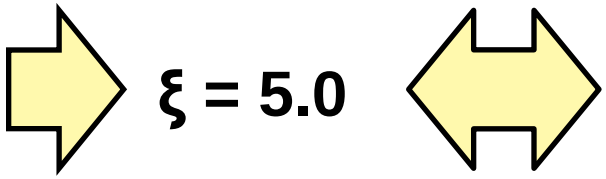




Intensity of chromatographic peak corresponding to 2,4-xylenol was at 5 times more than intensity of peak corresponding to “unknown” compound

26 mass spectral peaks were taken into account





P	*t* (*P*, 4)
0.95	2.8
0.99	4.6
0.999	8.6

Conclusions

- We have developed simple approach (based on using principal component analysis) for reliable comparison of full electron ionization mass spectra
  - Mass spectra should be registered under identical experimental conditions
  - Mass spectra should correspond approximately the same amounts of analytes
  - Both unknown compound and possible candidate should be analyzed at least three times
- Applicability of developed approach was shown in a number of examples
- Dependence of electron ionization mass spectra on amount of substance was shown

Acknowledgement

- We are grateful to Leco company for providing us with a time-of-flight mass spectrometer coupled with a gas chromatograph for our research
- We are grateful to Dr. A.V. Garmash for discussion of presented results