detection and identification of high-risk chemicals with LC/HMRS and machine learning

anneli kruve

Swedish drinking water

an offence more than every fifth day

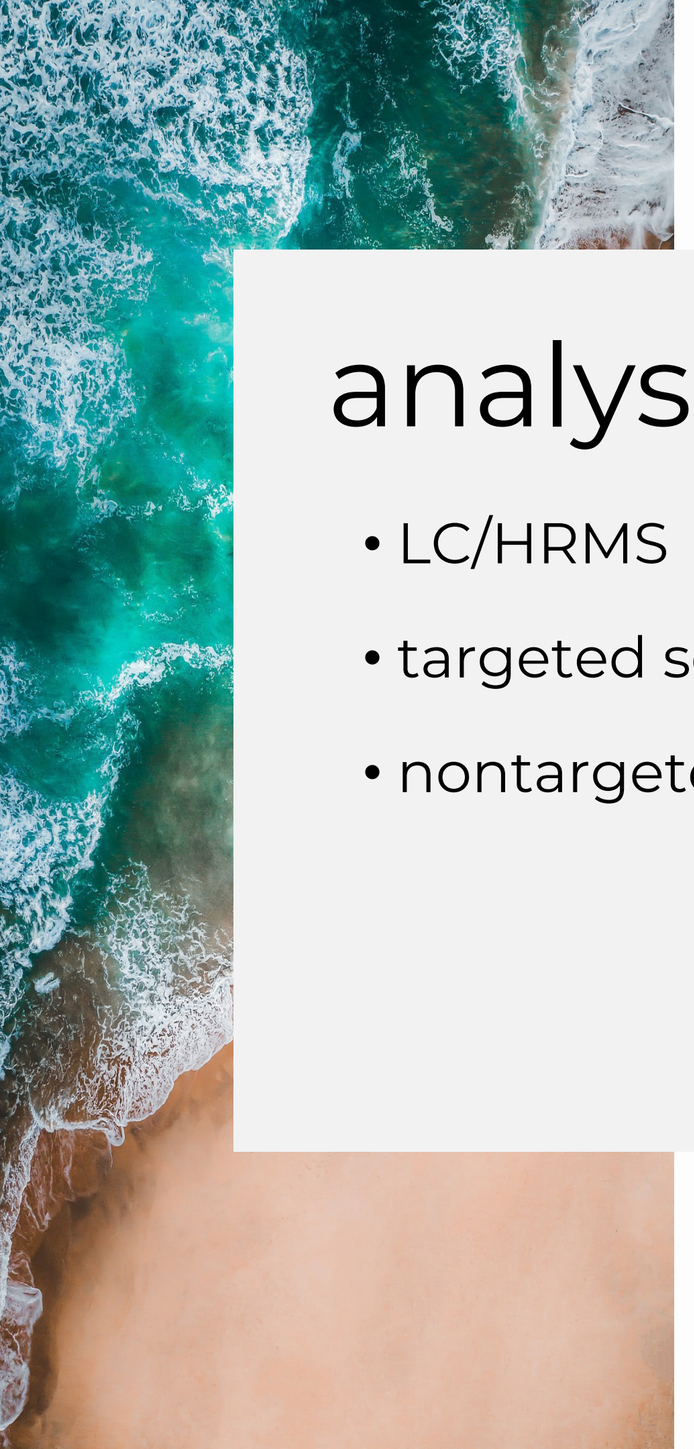
is chemical risk introduced?

DAGENS NYHETER.

SVERIGE

Ökat hot mot svenskt dricksvatten – brott mer än var femte dag

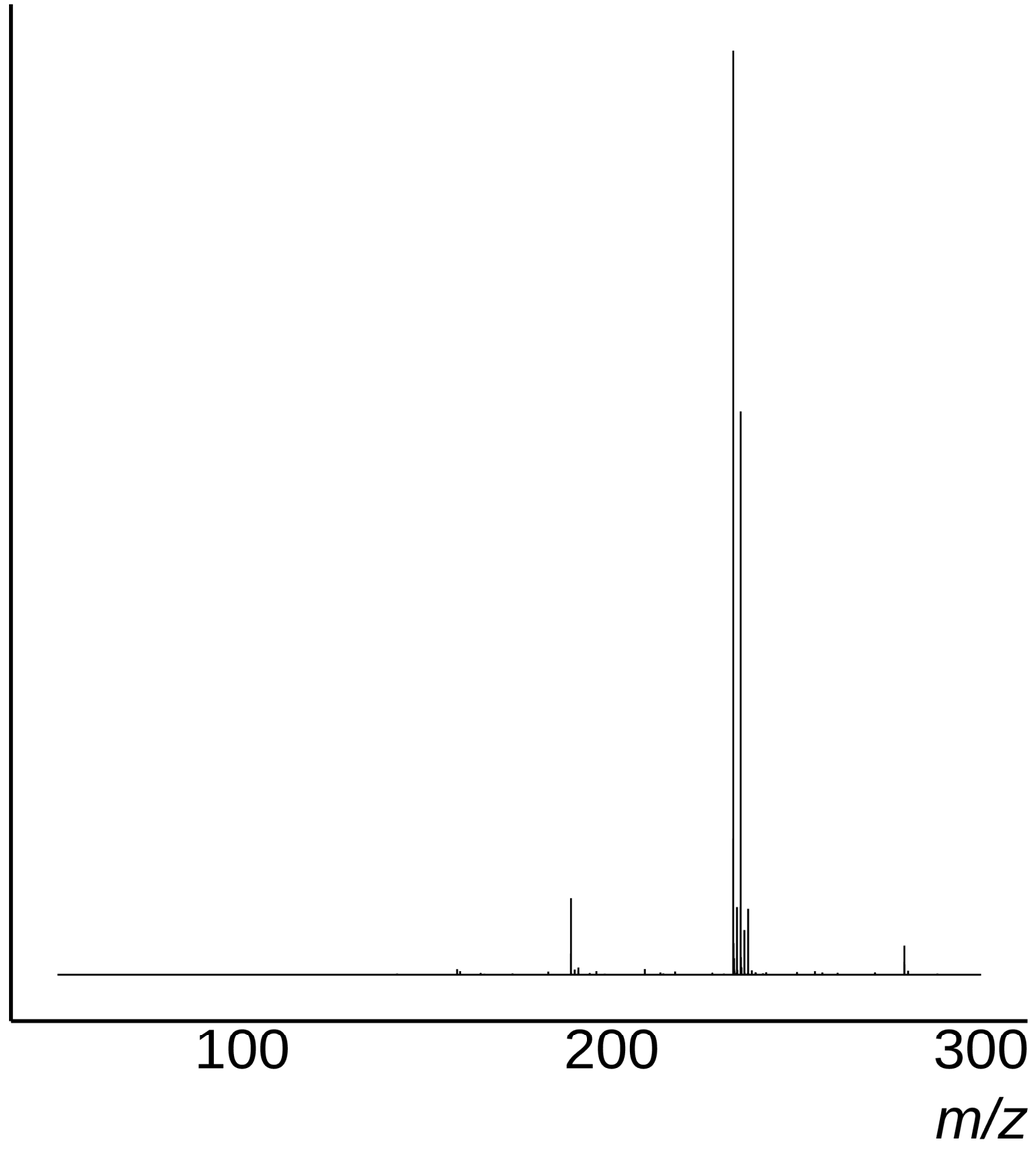
Uppdaterad 2024-12-02 Publicerad 2024-12-01



analysis

- LC/HRMS
- targeted screening
- nontargeted screening

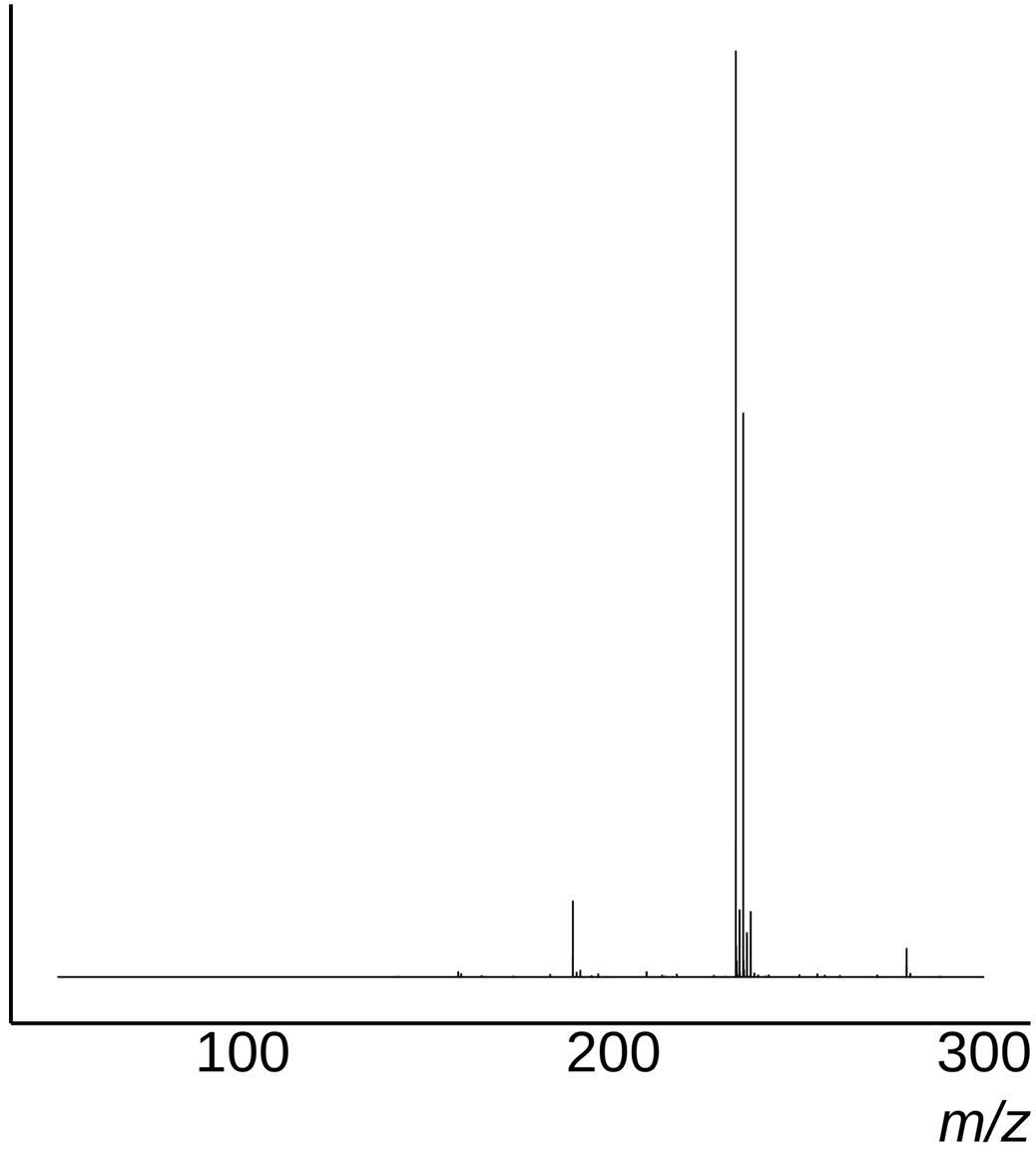
mass spectra



Hupatz, Rahu, et al. Anal. Bioanal. Chem. 417 (2025) 473

mass spectra

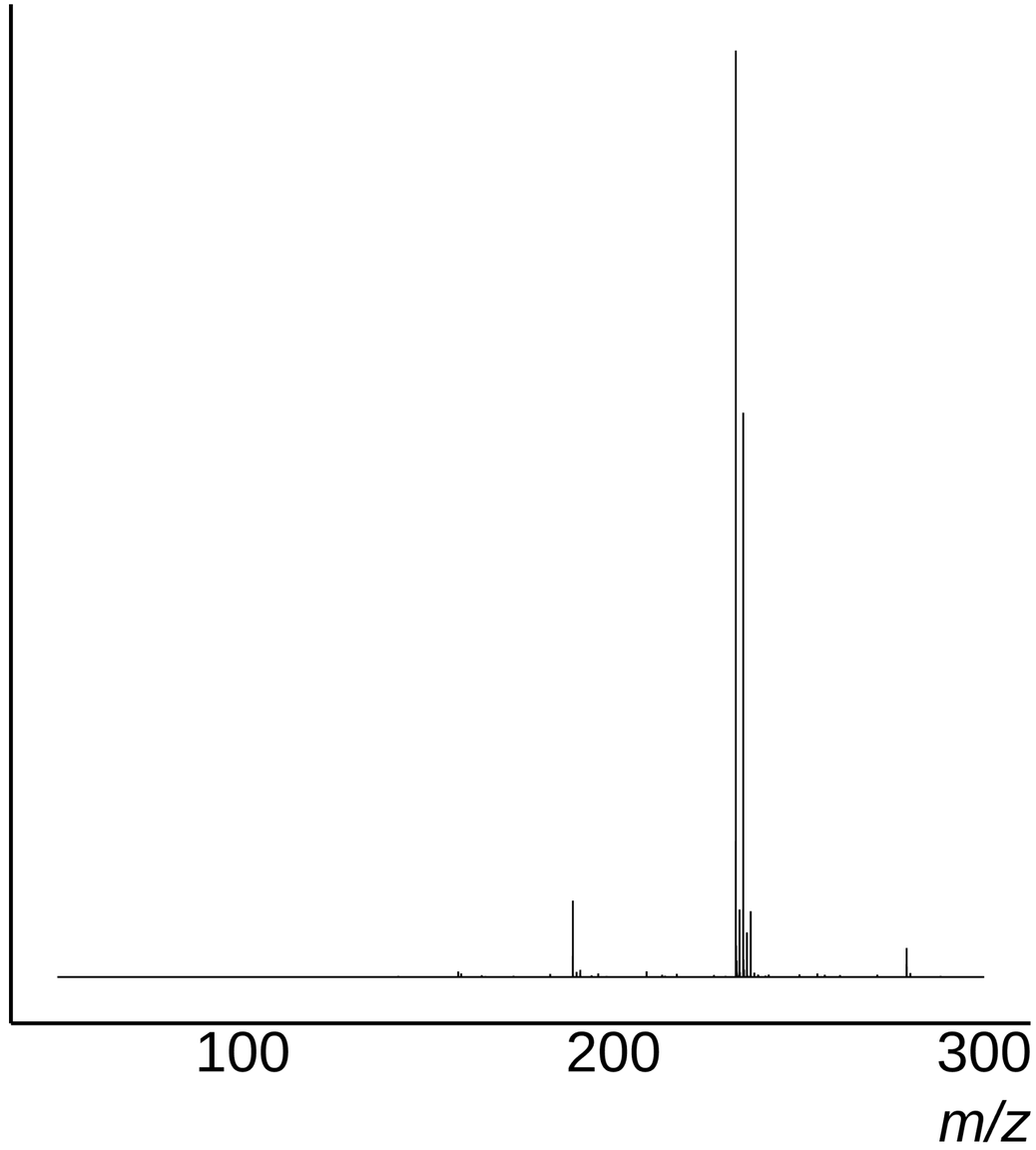
- library matching
  MassBank, MoNA, GNPS, NIST, ...



Hupatz, Rahu, et al. Anal. Bioanal. Chem. 417 (2025) 473

mass spectra

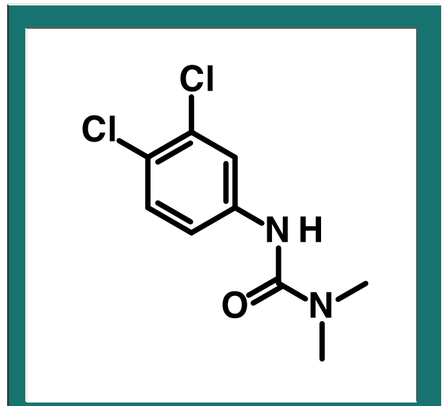
- library matching
  MassBank, MoNA, GNPS, NIST, …
- *in silico* interpretation
  SIRIUS, MetFrag, CFM-ID, …



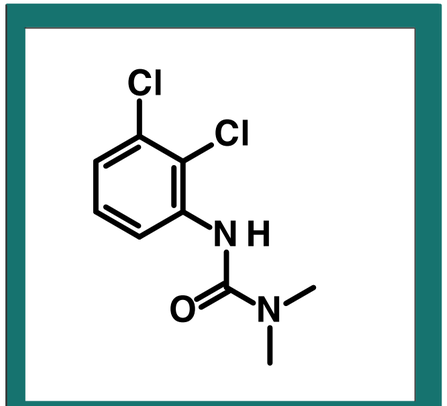
Hupatz, Rahu, et al. Anal. Bioanal. Chem. 417 (2025) 473

candidate structures

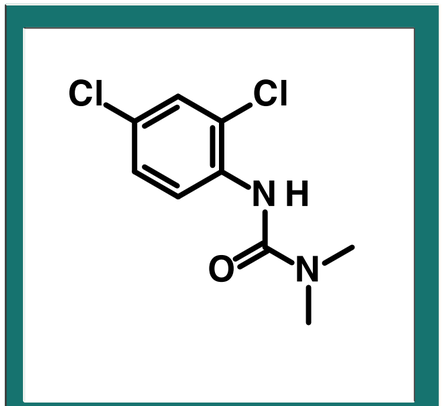
structural/positional isomers



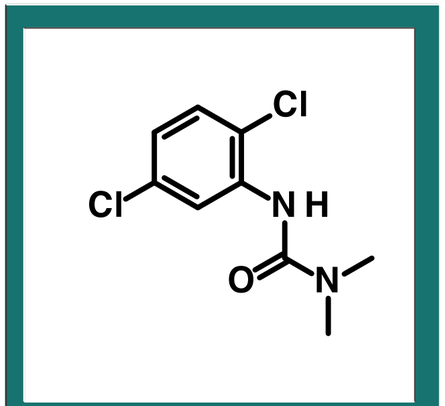
#1



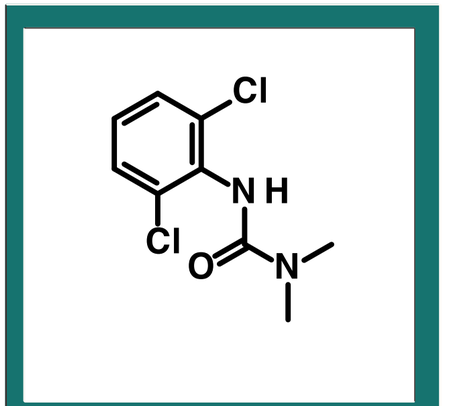
#2



#3



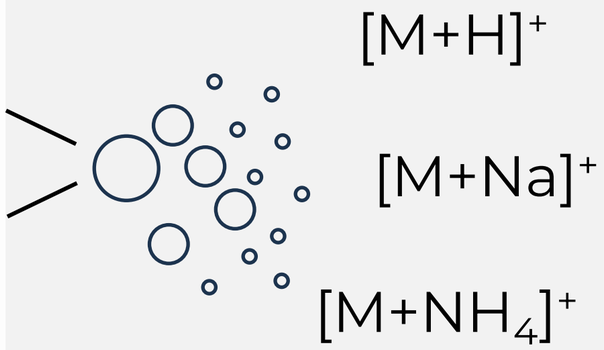
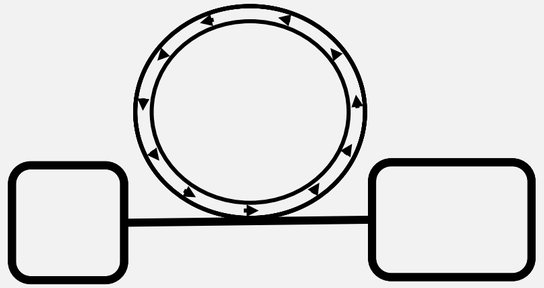
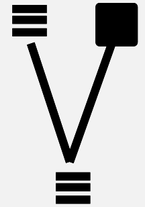
#4



#5

Akhlaqi et al. Anal. Bioanal. Chem. 415 (2023) 5247

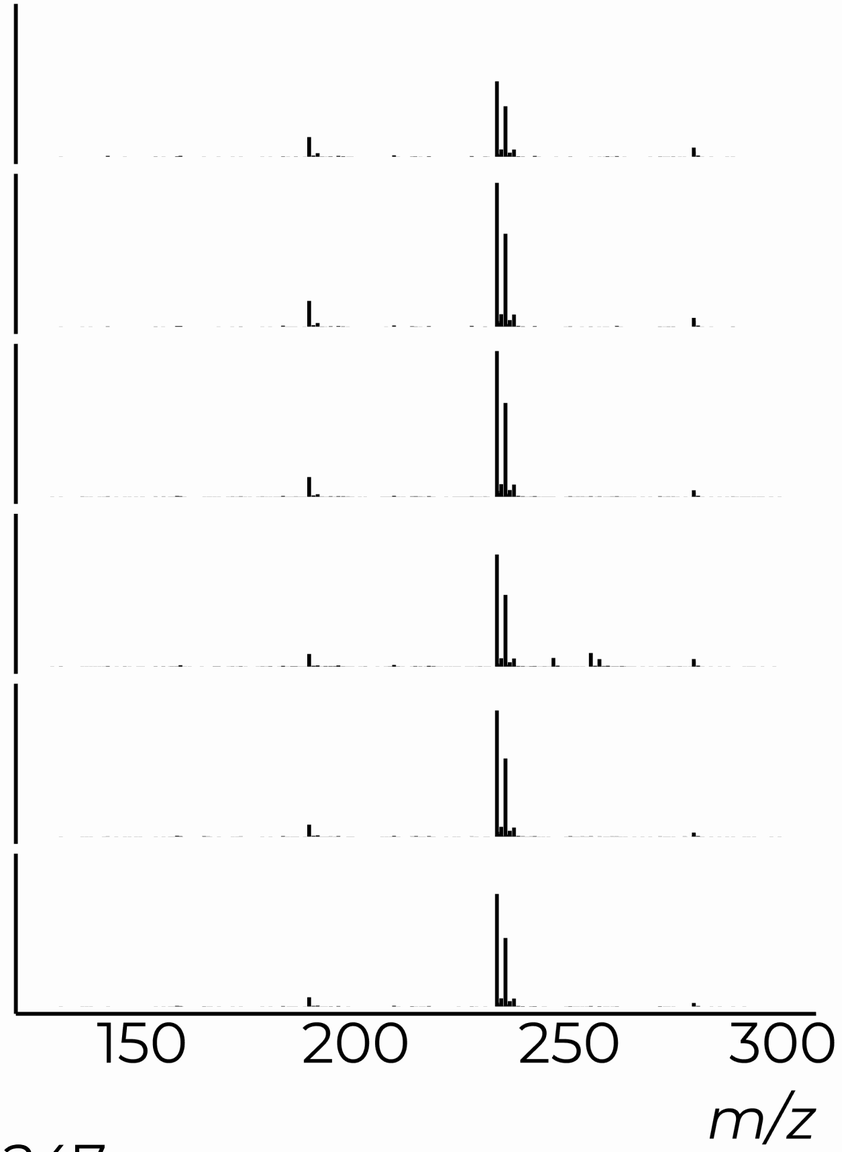
analytical information

chromatography	ionization	ion mobility	fragmentation
			

Akhlaqi et al. Anal. Bioanal. Chem. 415 (2023) 5247
Hupatz, Rahu, et al. Anal. Bioanal. Chem. 417 (2025) 473

confirmation

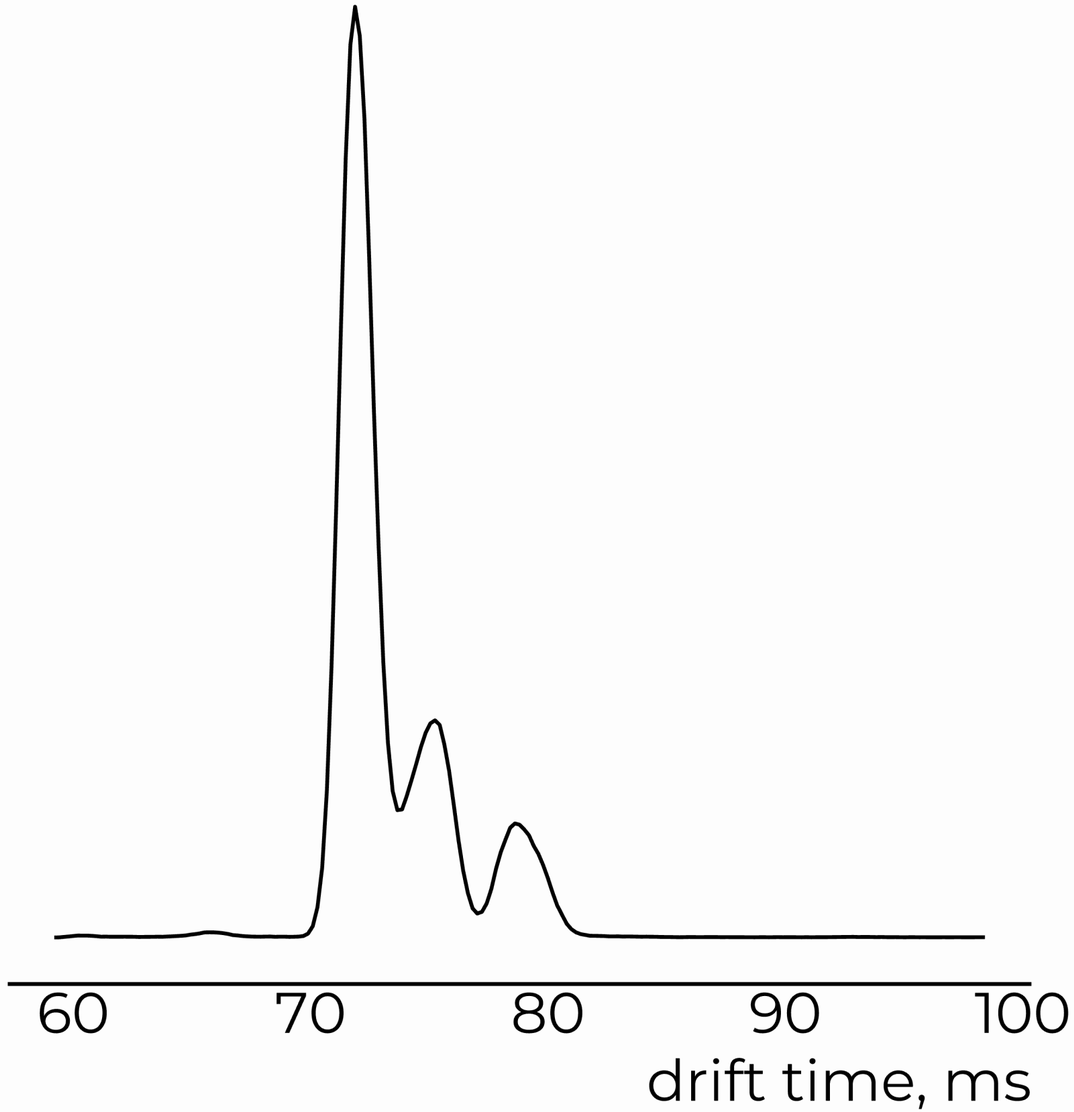
- MS^2 indistinguishable



Akhlaqi et al. Anal. Bioanal. Chem. 415 (2023) 5247

confirmation

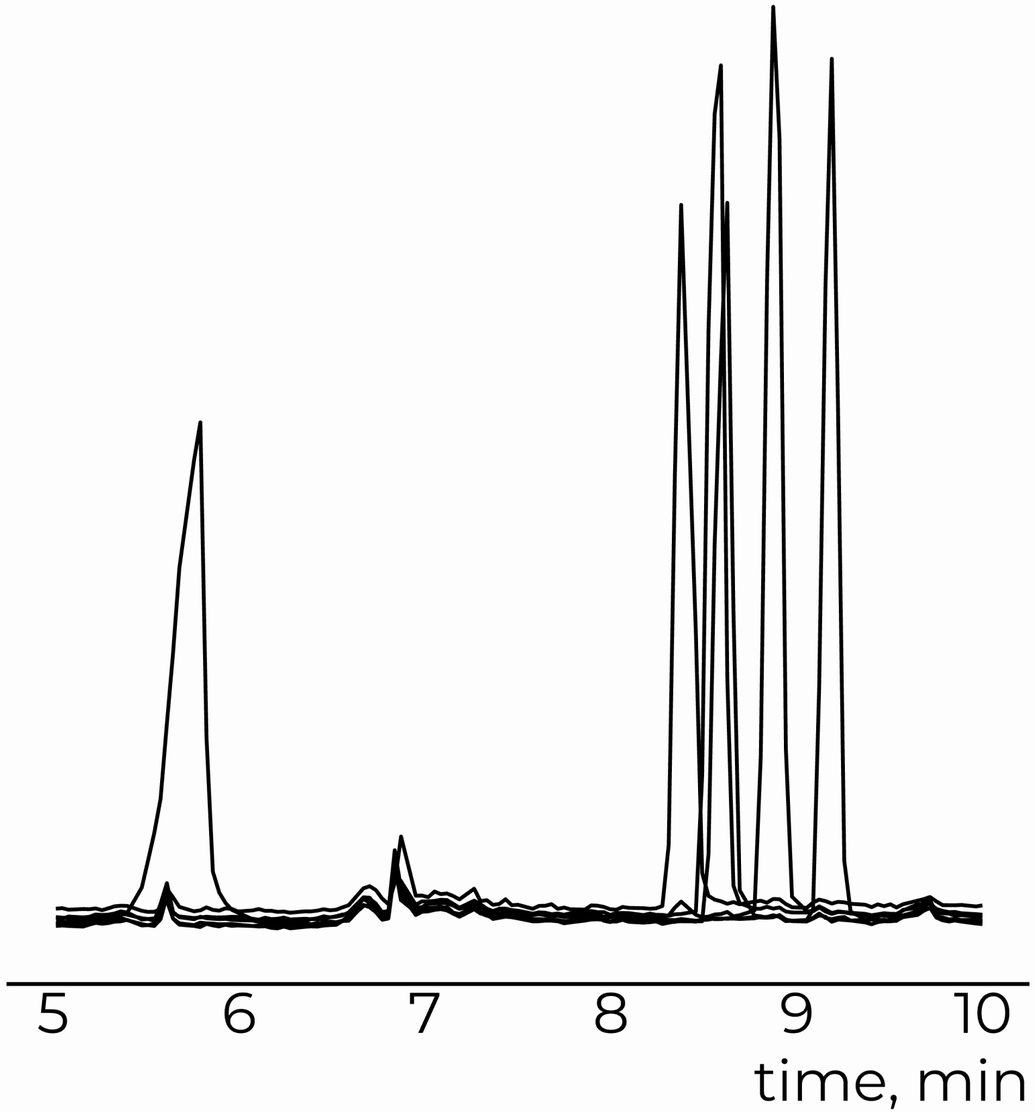
- MS^2 indistinguishable
- IMS indistinguishable



Akhlaqi et al. Anal. Bioanal. Chem. 415 (2023) 5247

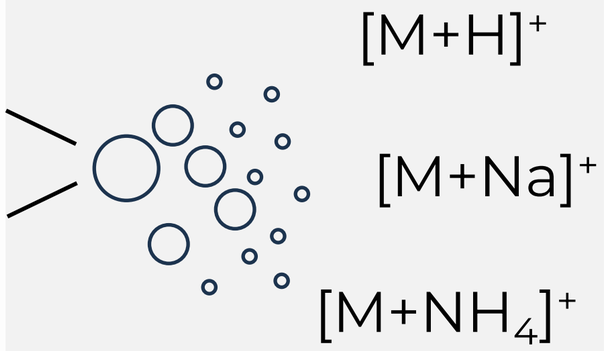
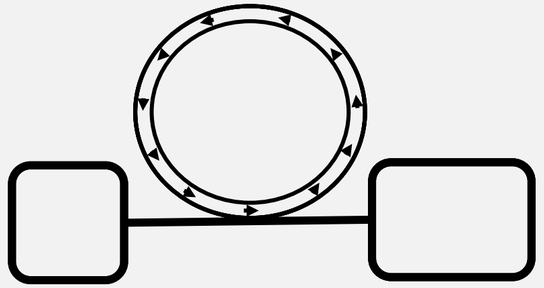
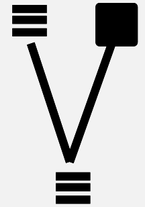
confirmation

- MS^2
  indistinguishable
- IMS
  indistinguishable
- LC
  50% distinguishable



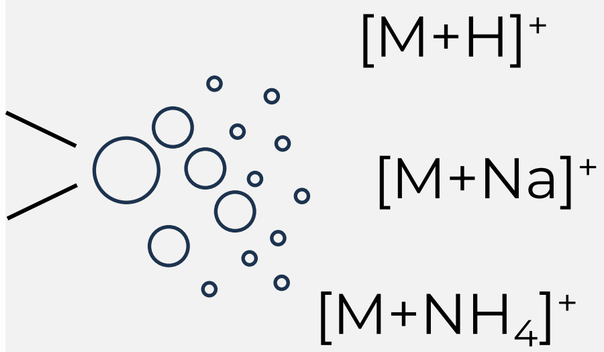
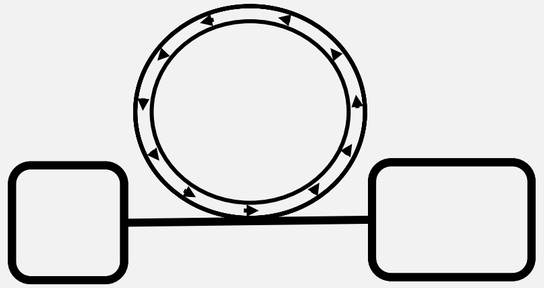
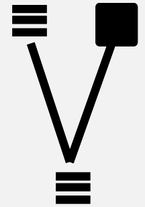
Akhlaqi et al. Anal. Bioanal. Chem. 415 (2023) 5247

analytical information

chromatography	ionization	ion mobility	fragmentation
			

Akhlaqi et al. Anal. Bioanal. Chem. 415 (2023) 5247

analytical information

chromatography	ionization	ion mobility	fragmentation
			
~3 of 5 cases	0 of 5 cases	~1 of 5 cases	~2 of 5 cases

Akhlaqi et al. Anal. Bioanal. Chem. 415 (2023) 5247

#1a

take-home message

wrong candidate structure
≠
false positive

experimentally inseparable isomeric candidate structures are common

#1b

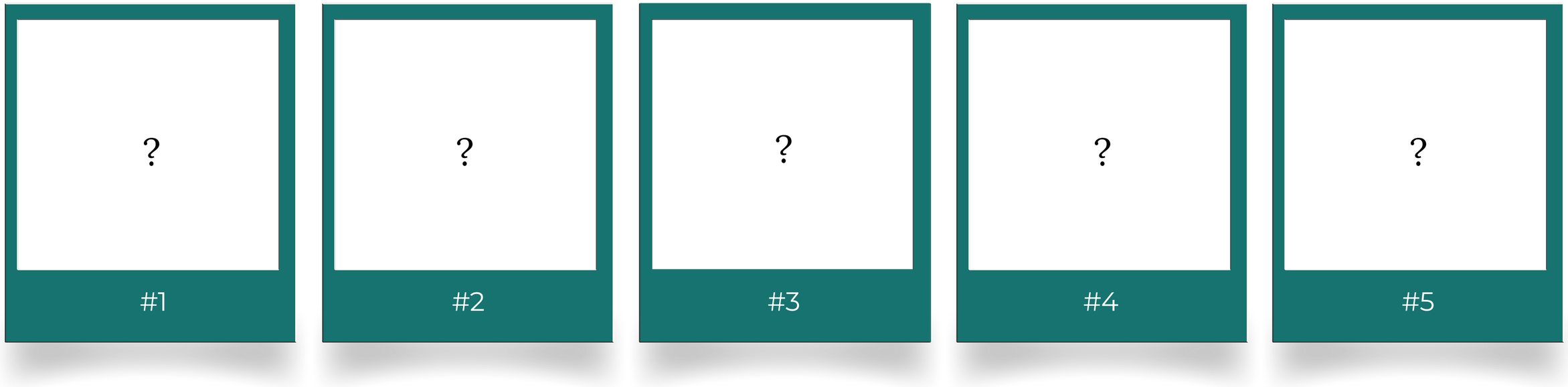
take-home message

selectivity of analytics matters

in silico tools cannot distinguish chemicals that are not separable analytically

candidate structures

no (likely) candidate structures



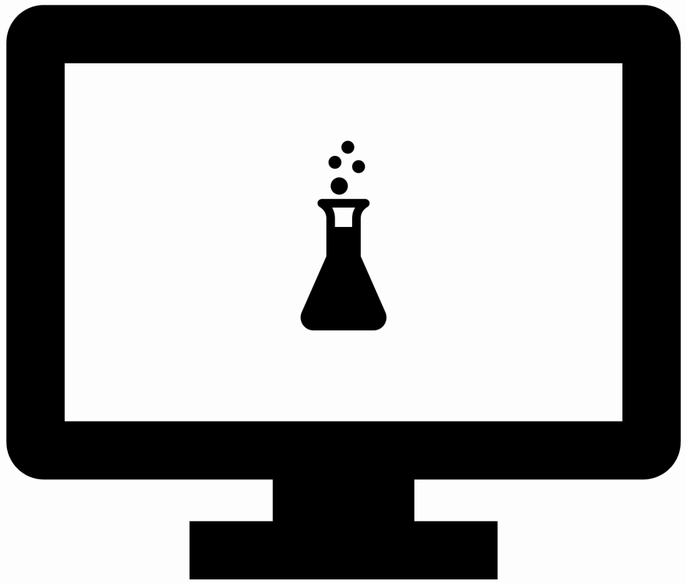
generate structures

- expert knowledge
  time consuming

generate structures

- expert knowledge
  time consuming

- in silico
  MSNovelist

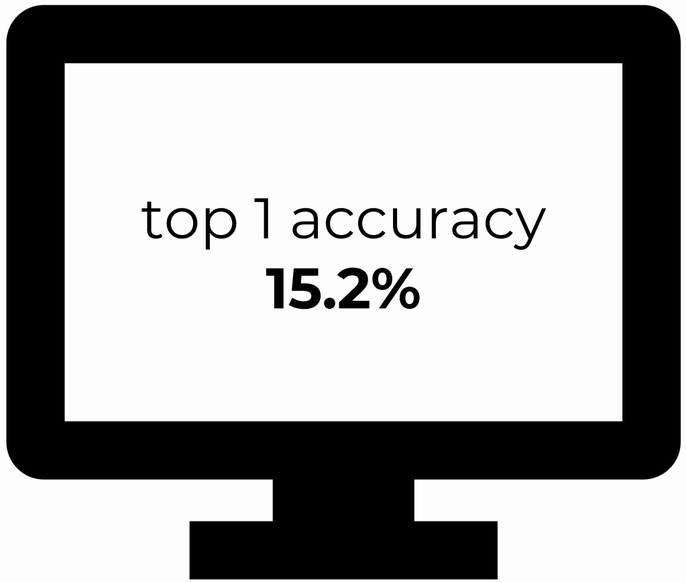


Stravs et al. Nature Mehtods 19 (2022) 865

generate structures

- expert knowledge
  time consuming

- in silico
  MSNovelist



Hättestrand et al. in preparation

#2

take-home message

evaluating unknowns is hard

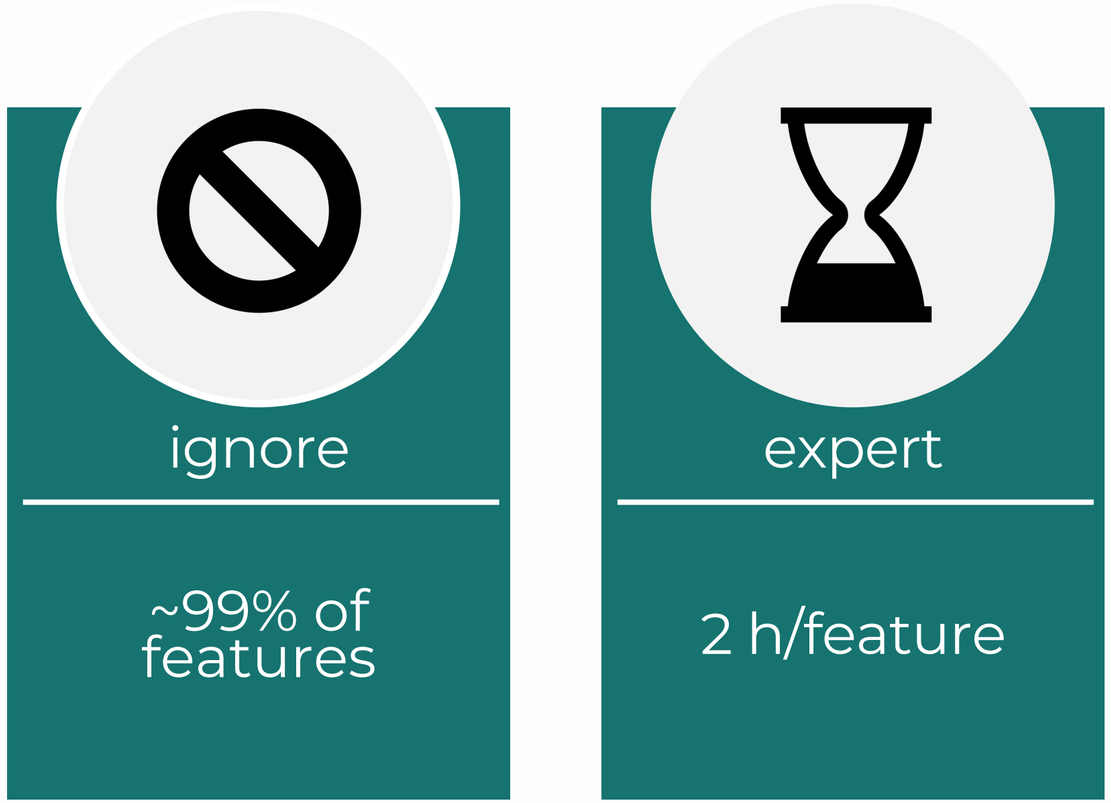
what is the ground truth?

unidentified LC/HRMS features

how to handle them?

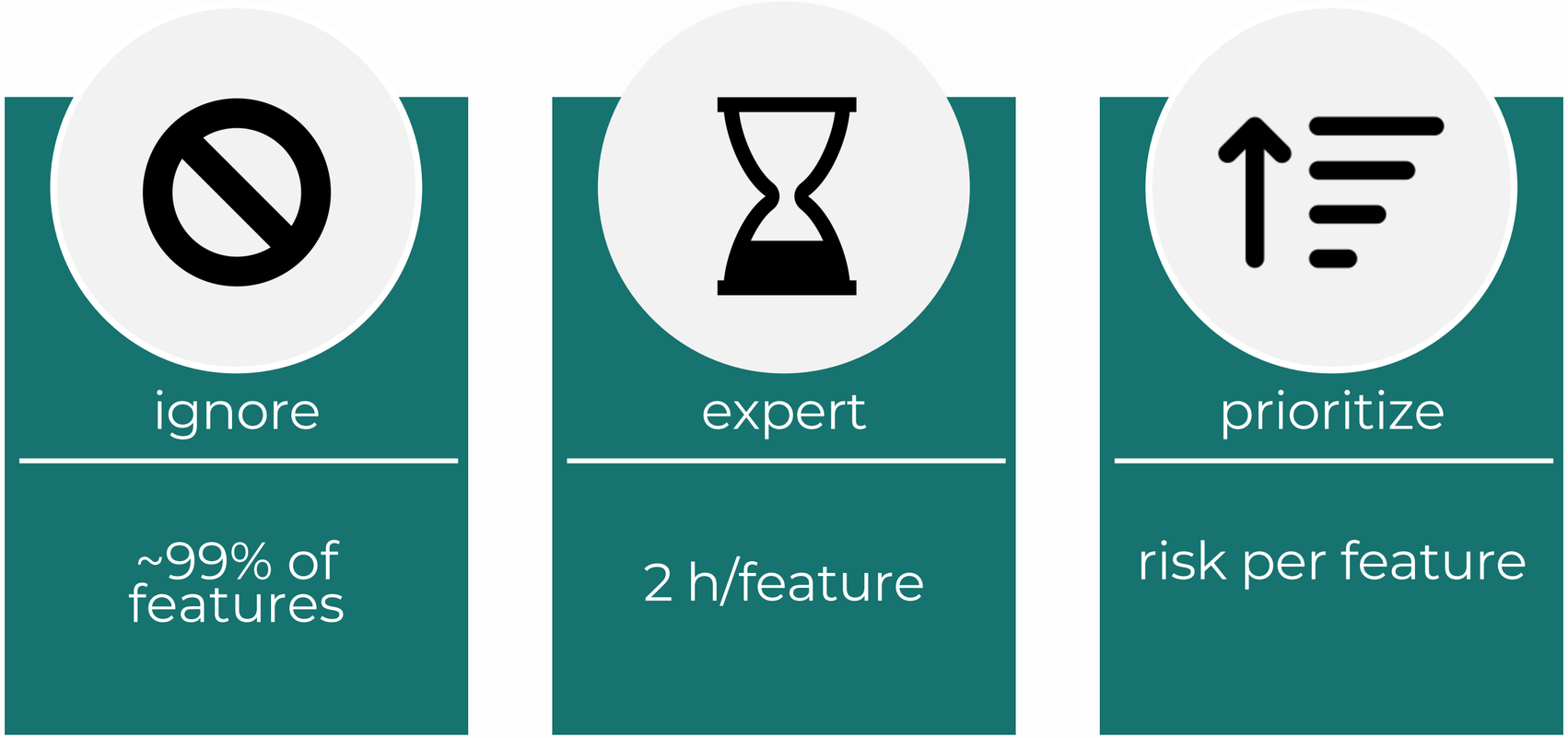
unidentified LC/HRMS features

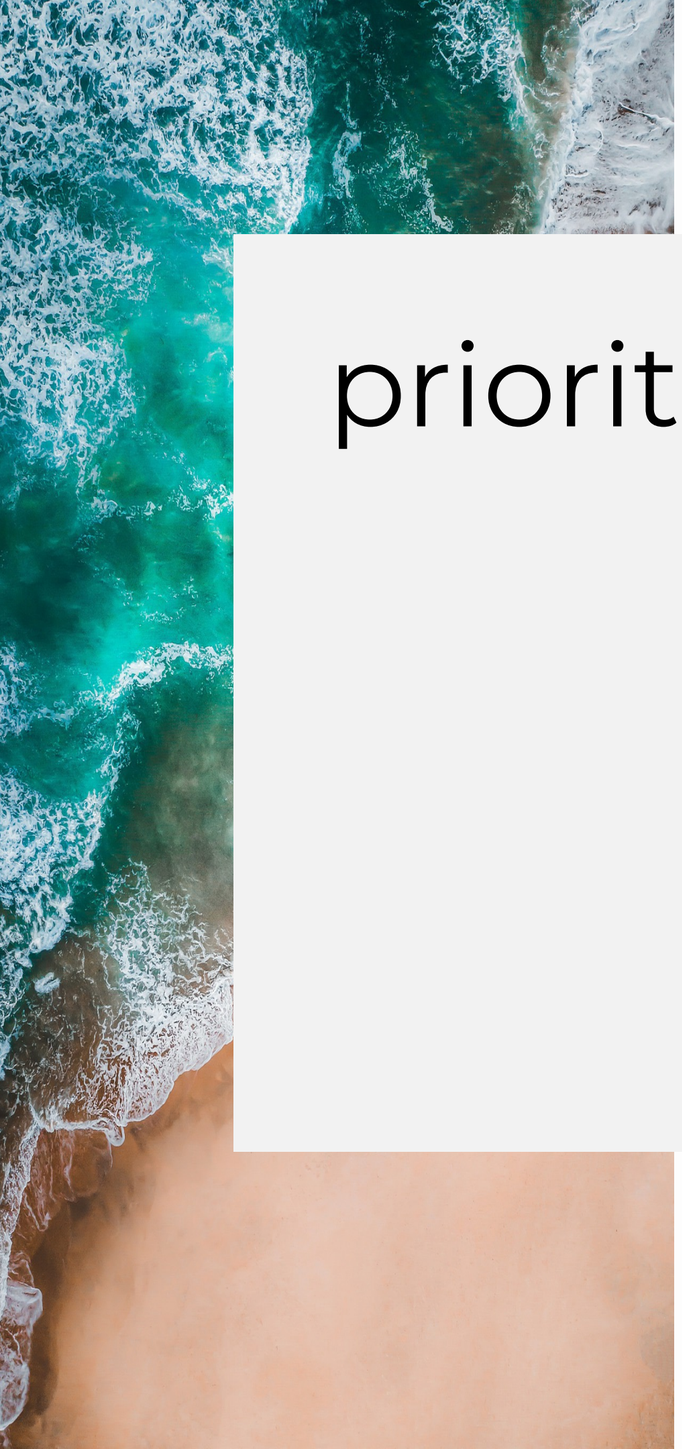
how to handle them?



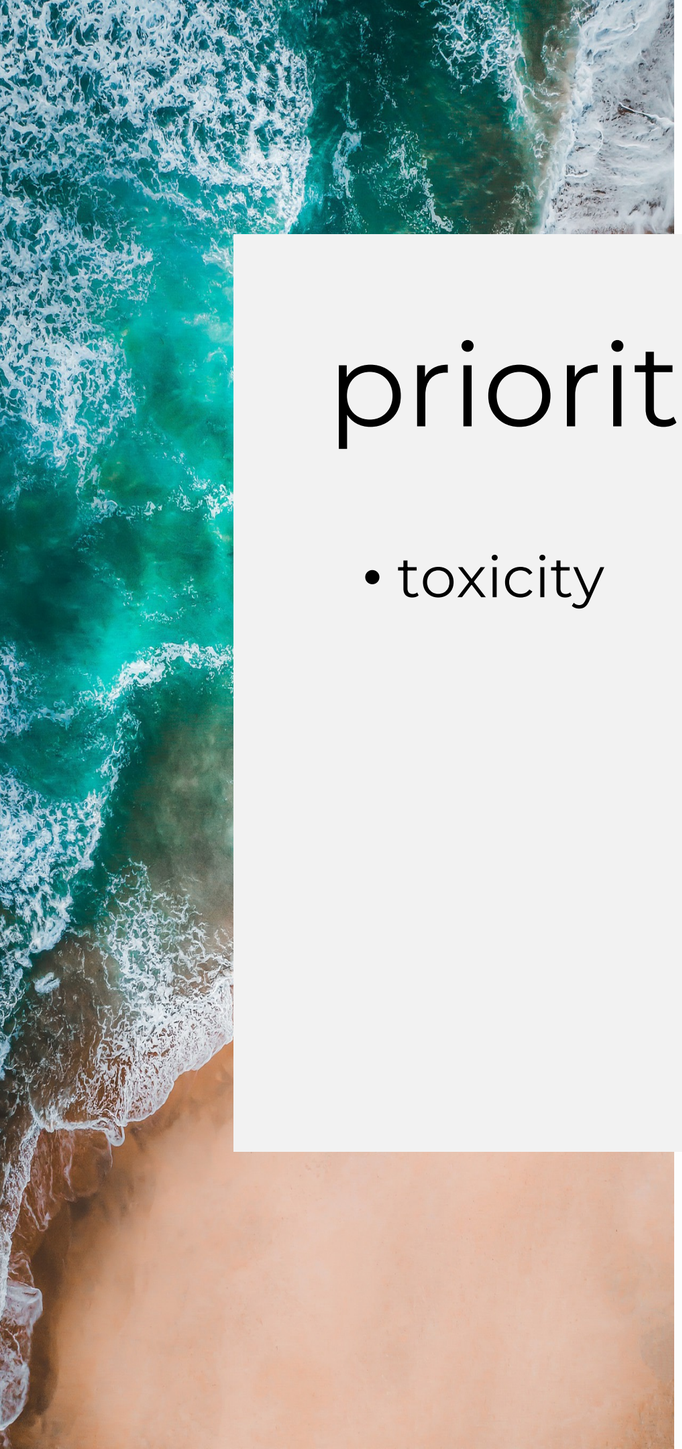
unidentified LC/HRMS features

how to handle them?



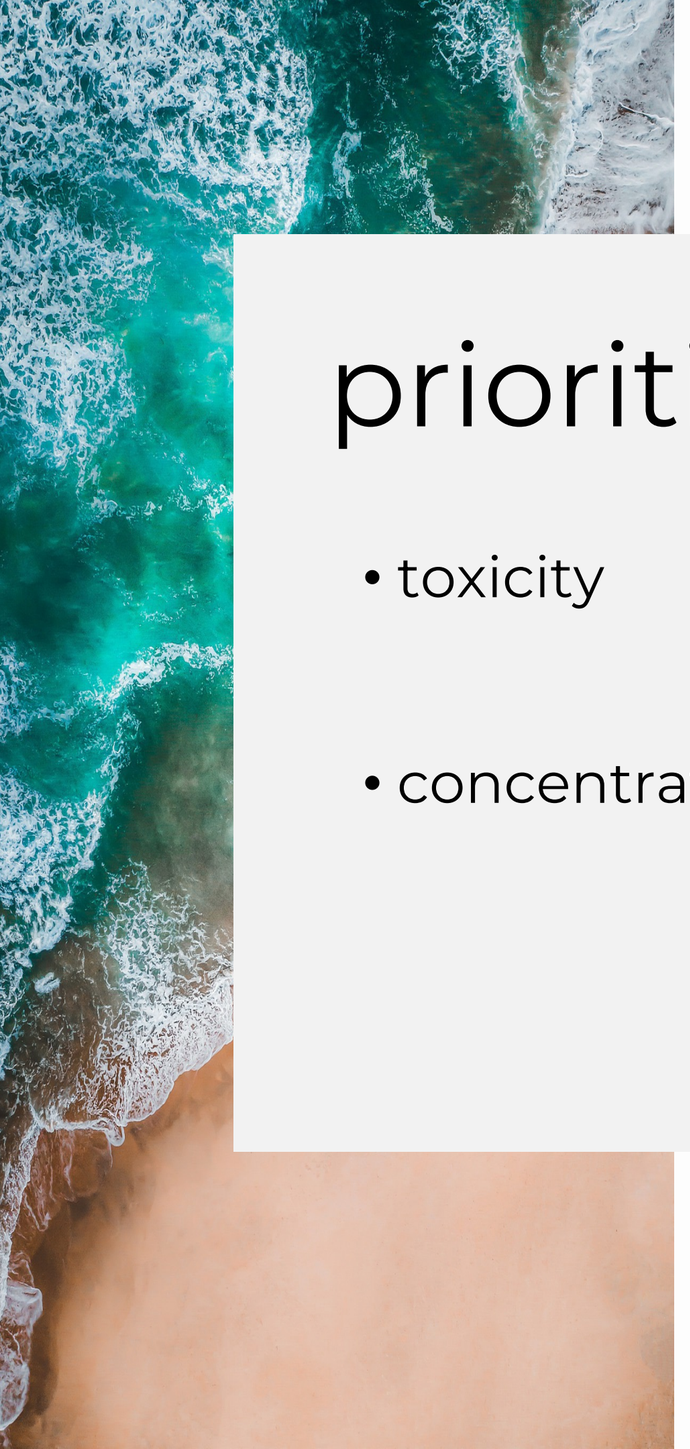


prioritization



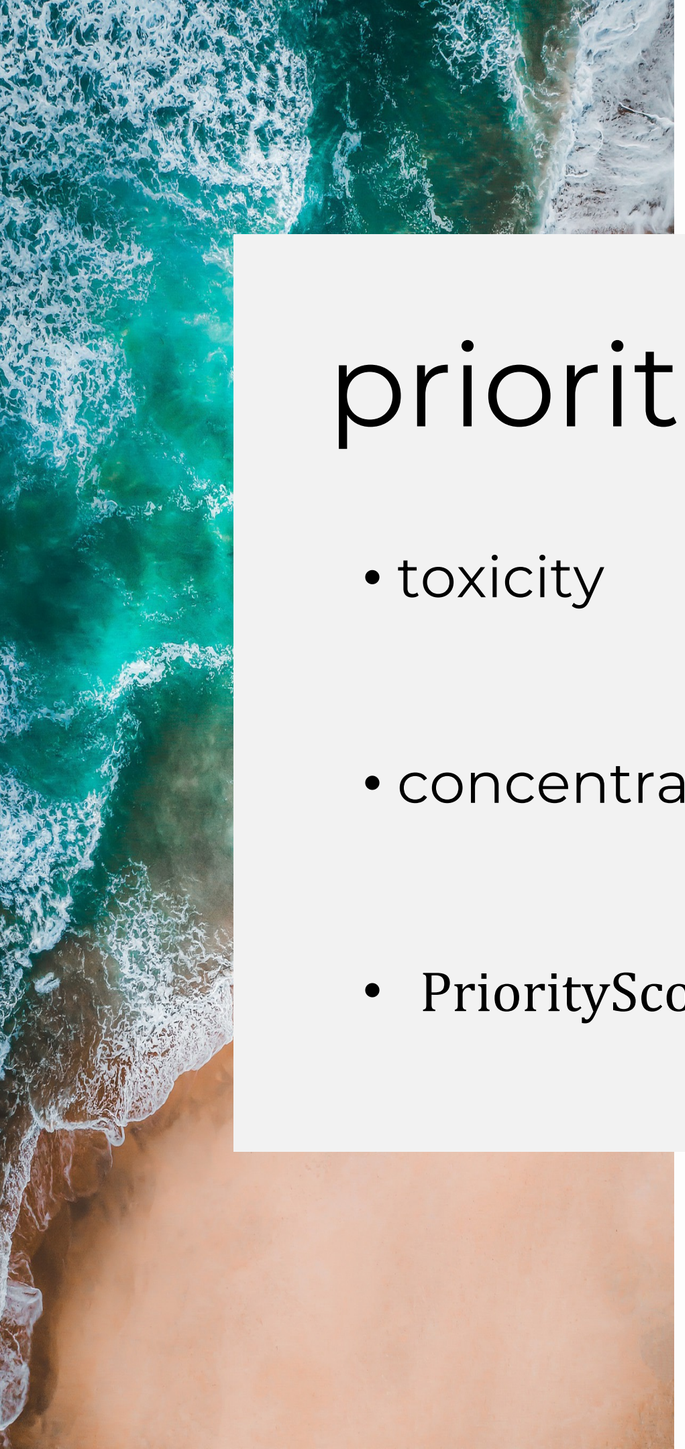
prioritization

- toxicity



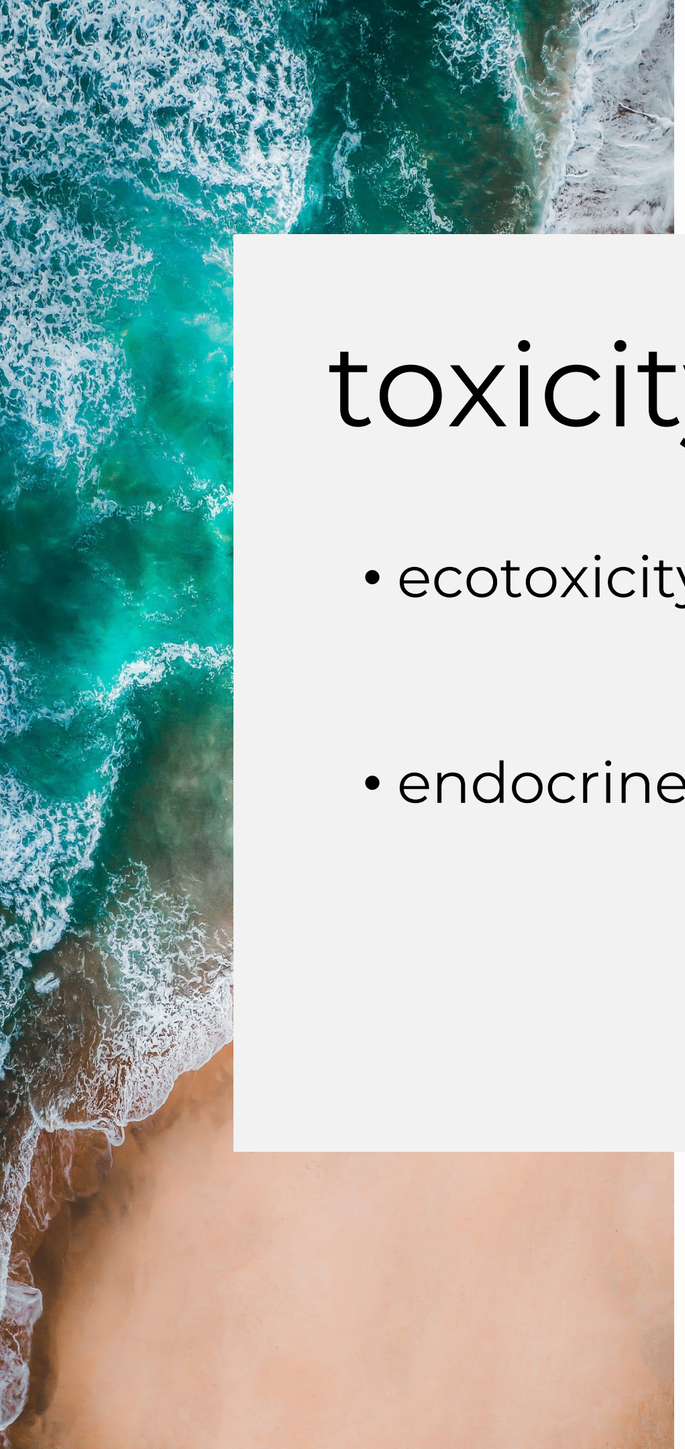
prioritization

- toxicity
- concentration



prioritization

- toxicity
- concentration
- $$\text{PriorityScore} = \frac{\text{C}_{\text{predicted}}}{\text{AC}_{50}^{\text{5th percentile}}}$$



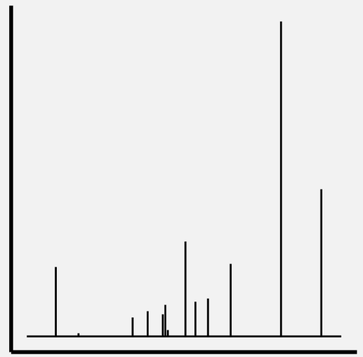
toxicity

- ecotoxicity

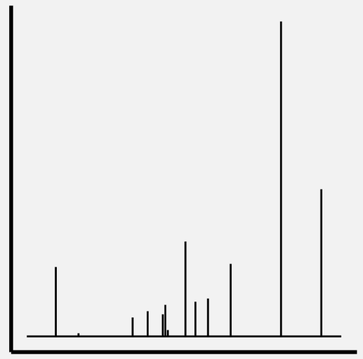
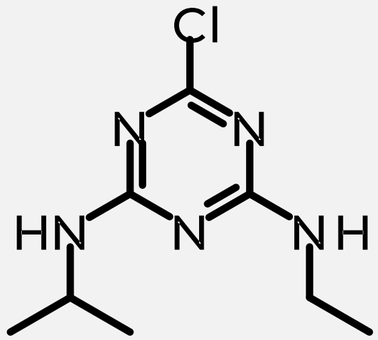
- endocrine disruptors

toxicity

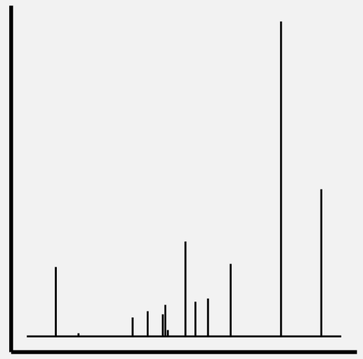
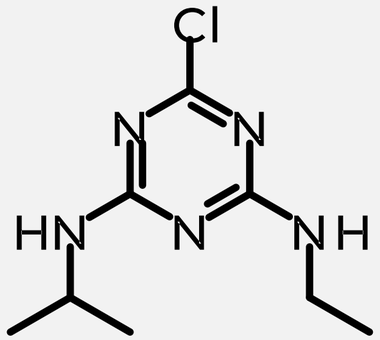
MS^2 spectra



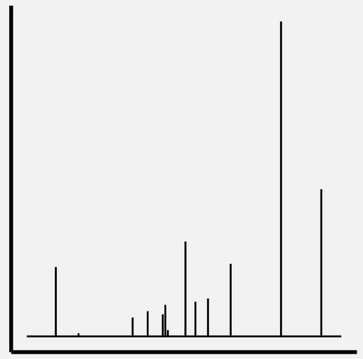
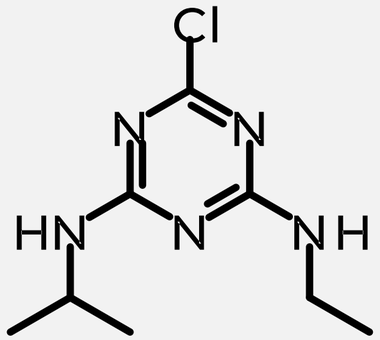
toxicity

MS^2 spectra	structure
	

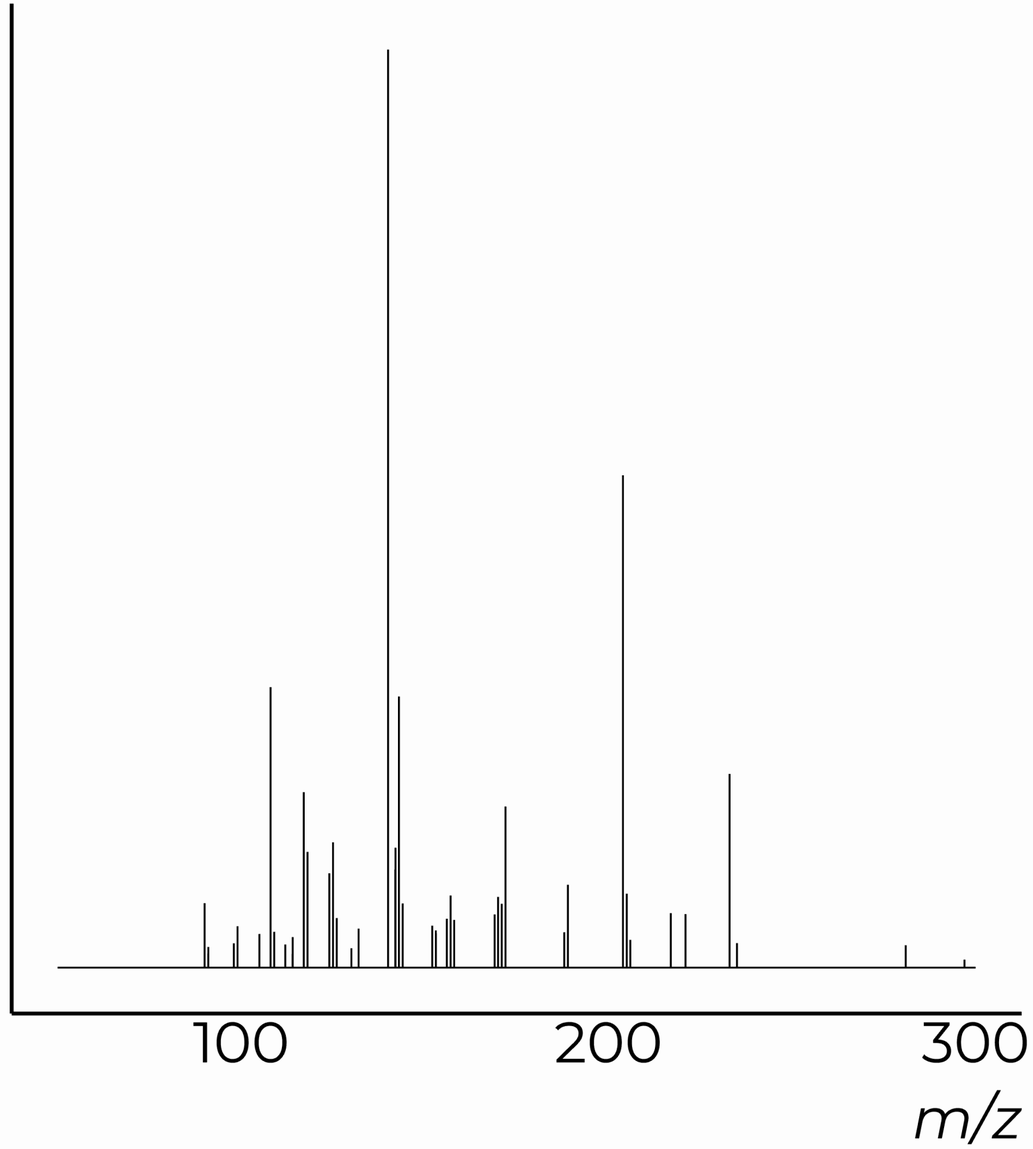
toxicity

MS^2 spectra	structure	descriptors
		

toxicity

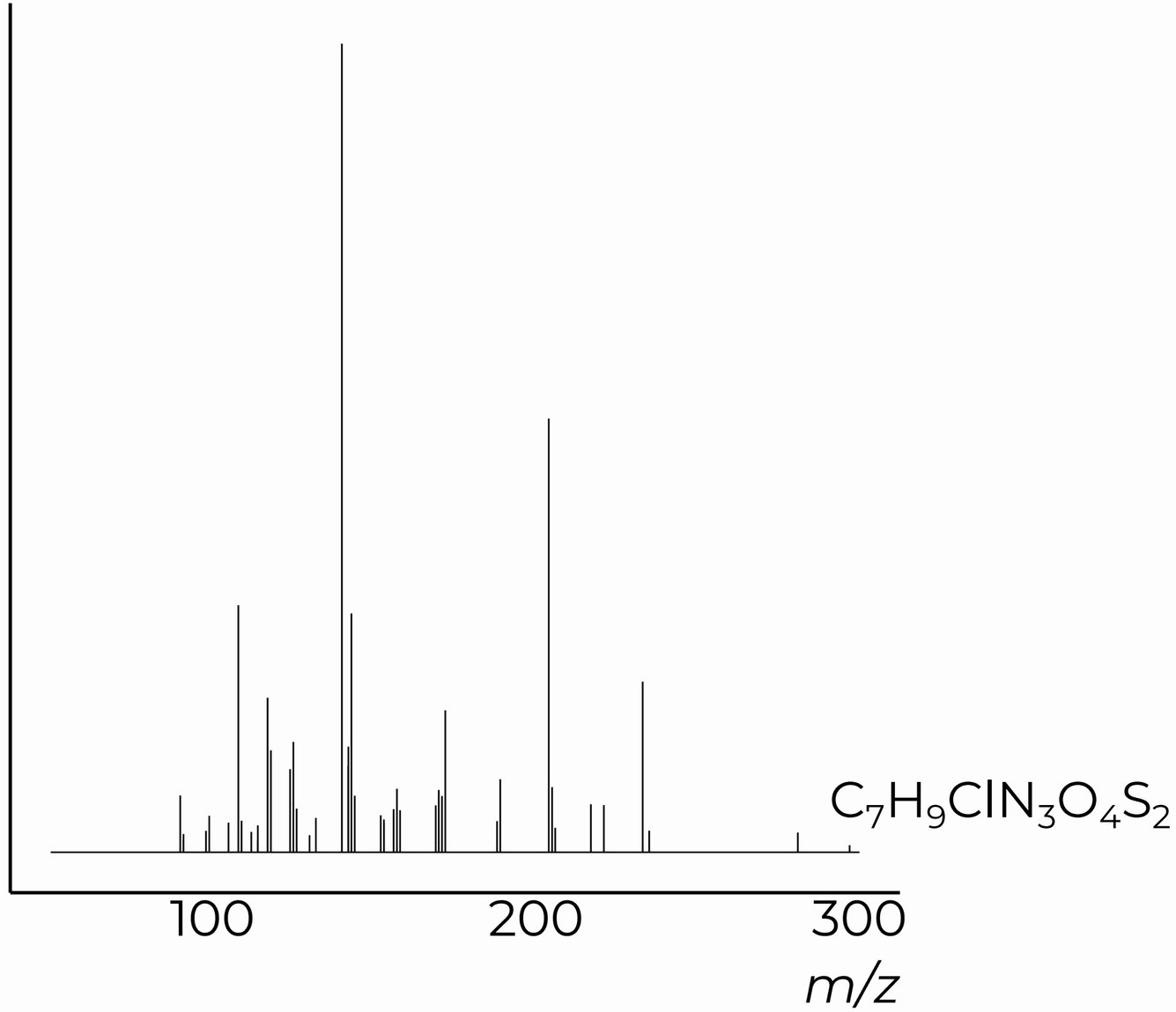
MS^2 spectra	structure	descriptors	predictions
			

information in MS^2



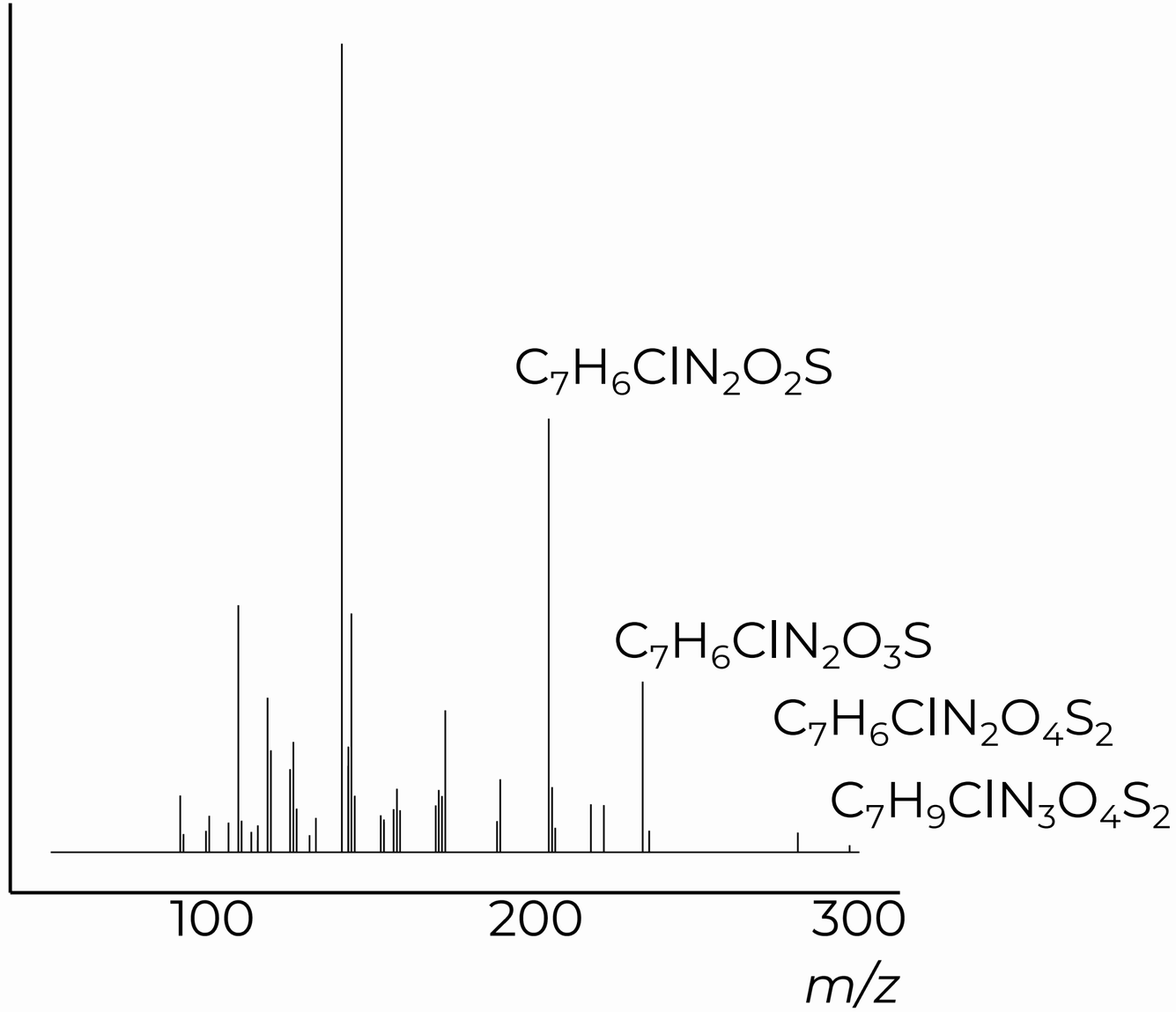
Dührkop et al. Nature Methods 16 (2019) 299.

information in MS^2



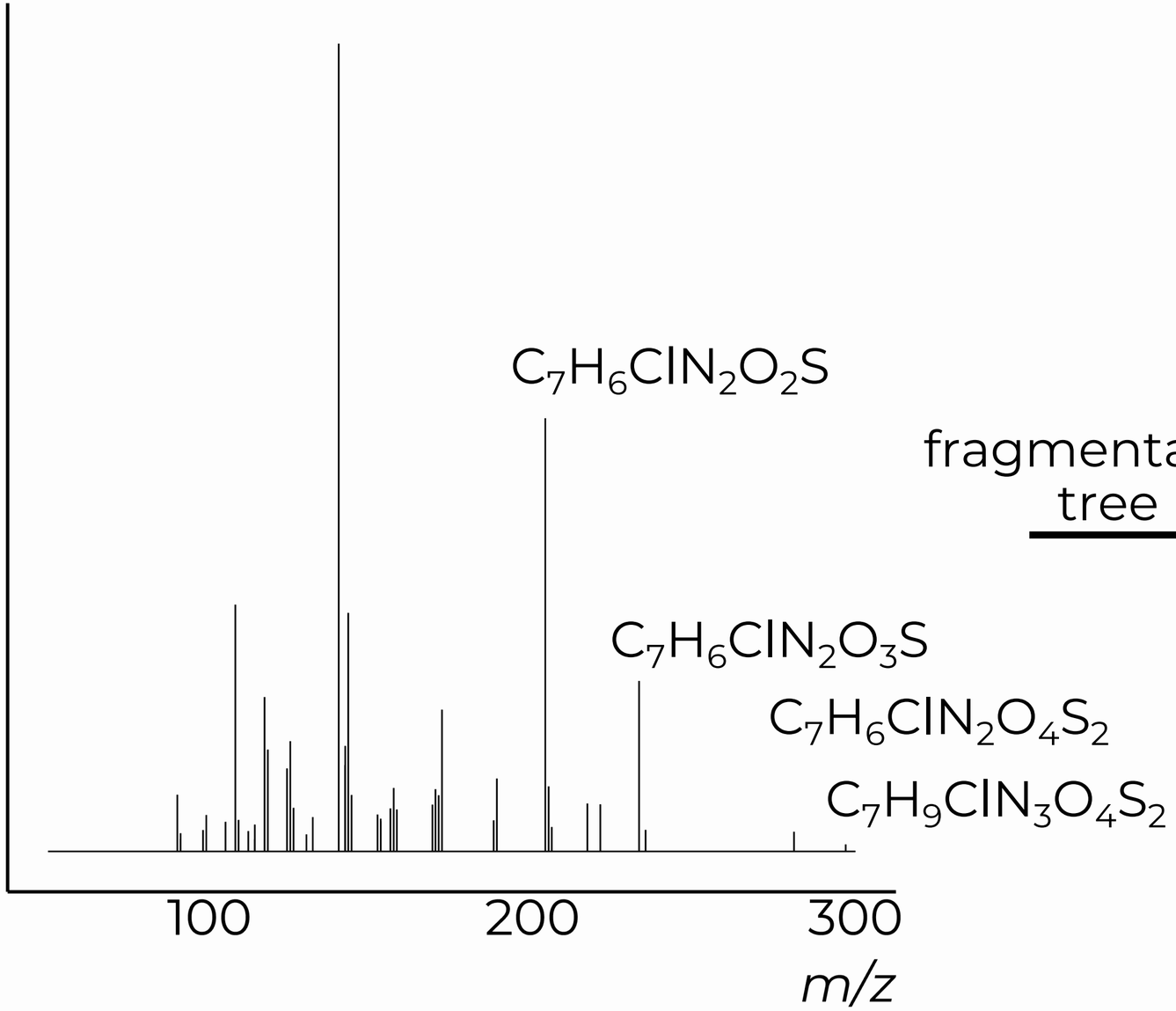
Dührkop et al. Nature Methods 16 (2019) 299.

information in MS^2

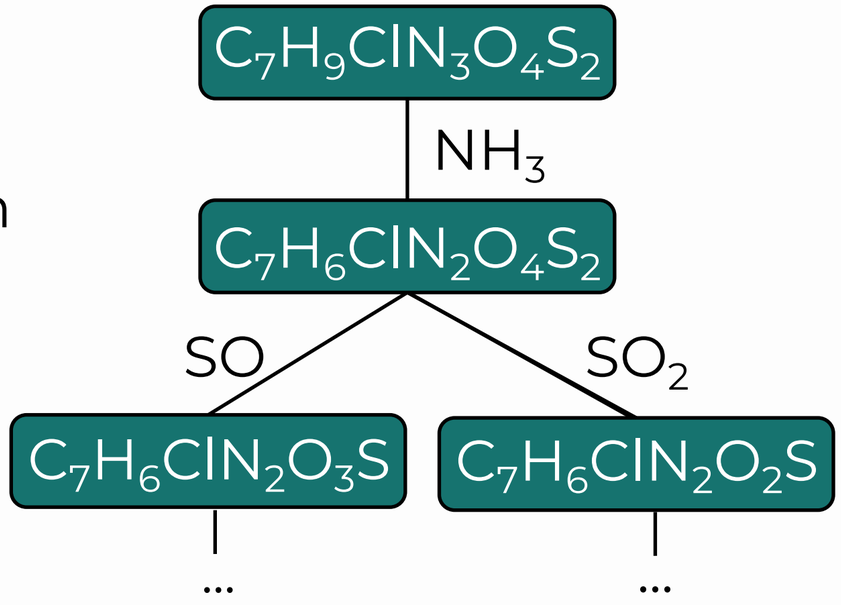


Dührkop et al. Nature Methods 16 (2019) 299.

information in MS^2

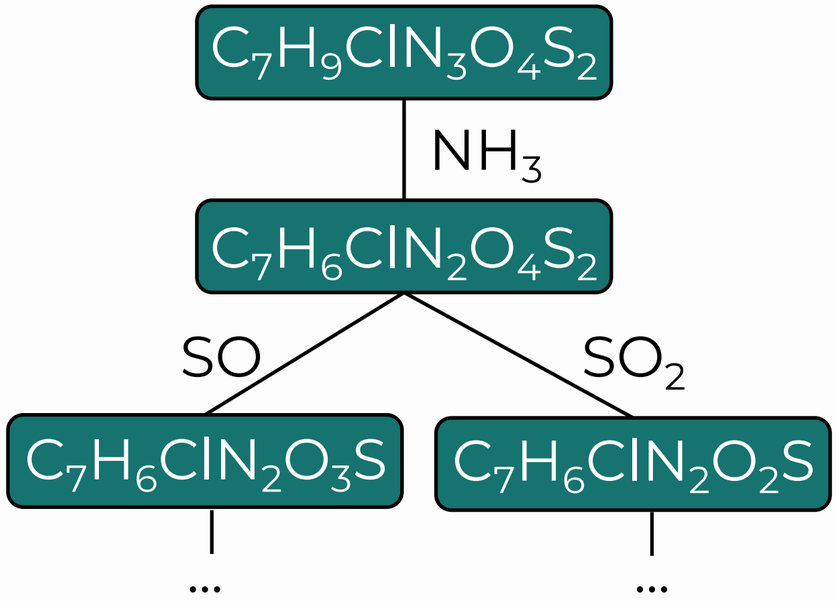


fragmentation tree



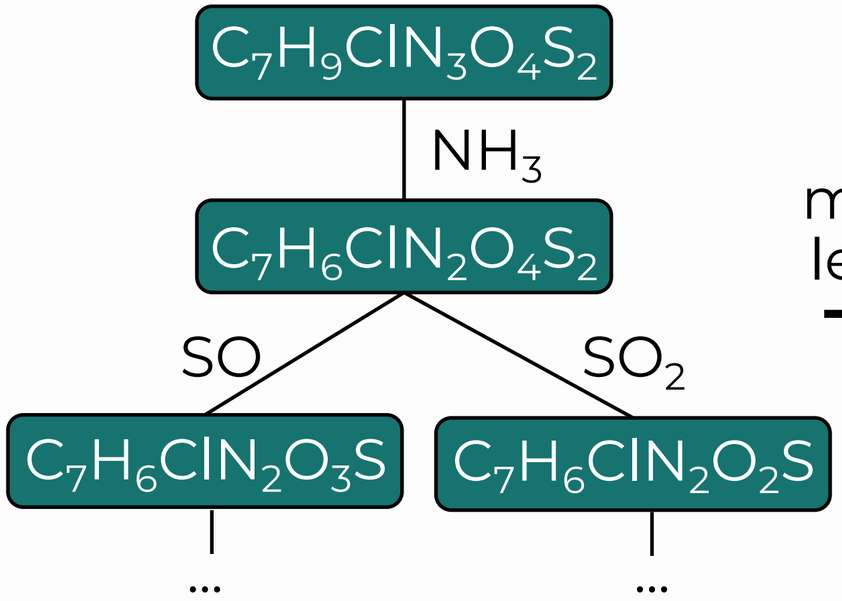
Dührkop et al. Nature Methods 16 (2019) 299.

information in MS^2



Dührkop et al. Nature Methods 16 (2019) 299.

information in MS^2

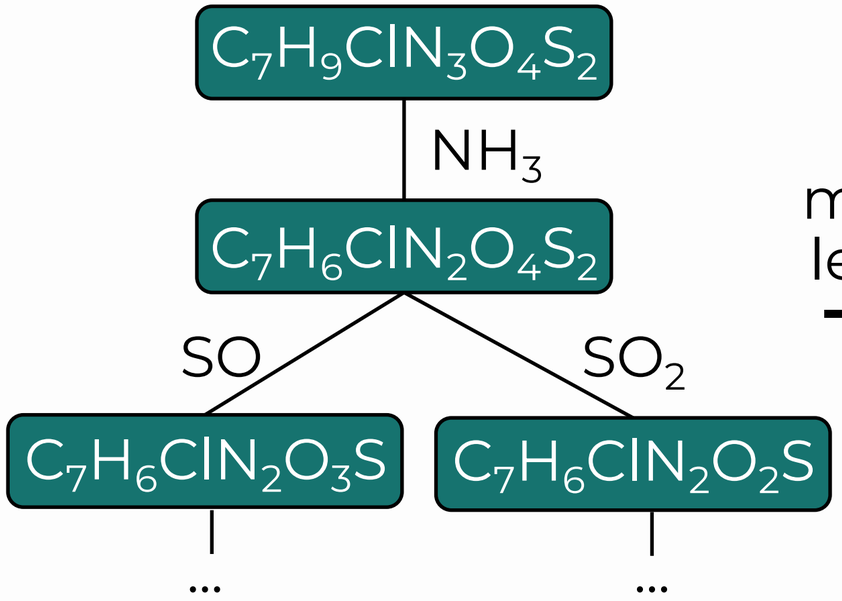


machine learning →

0.08	O-P
0.87	S=O
0.15	C-F
0.93	NH_2
0.99	O

Dührkop et al. Nature Methods 16 (2019) 299.

information in MS^2

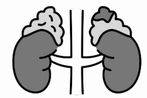


machine learning →

0.08	O-P
0.87	S=O
0.15	C-F
0.93	NH_2
0.99	O

machine learning →

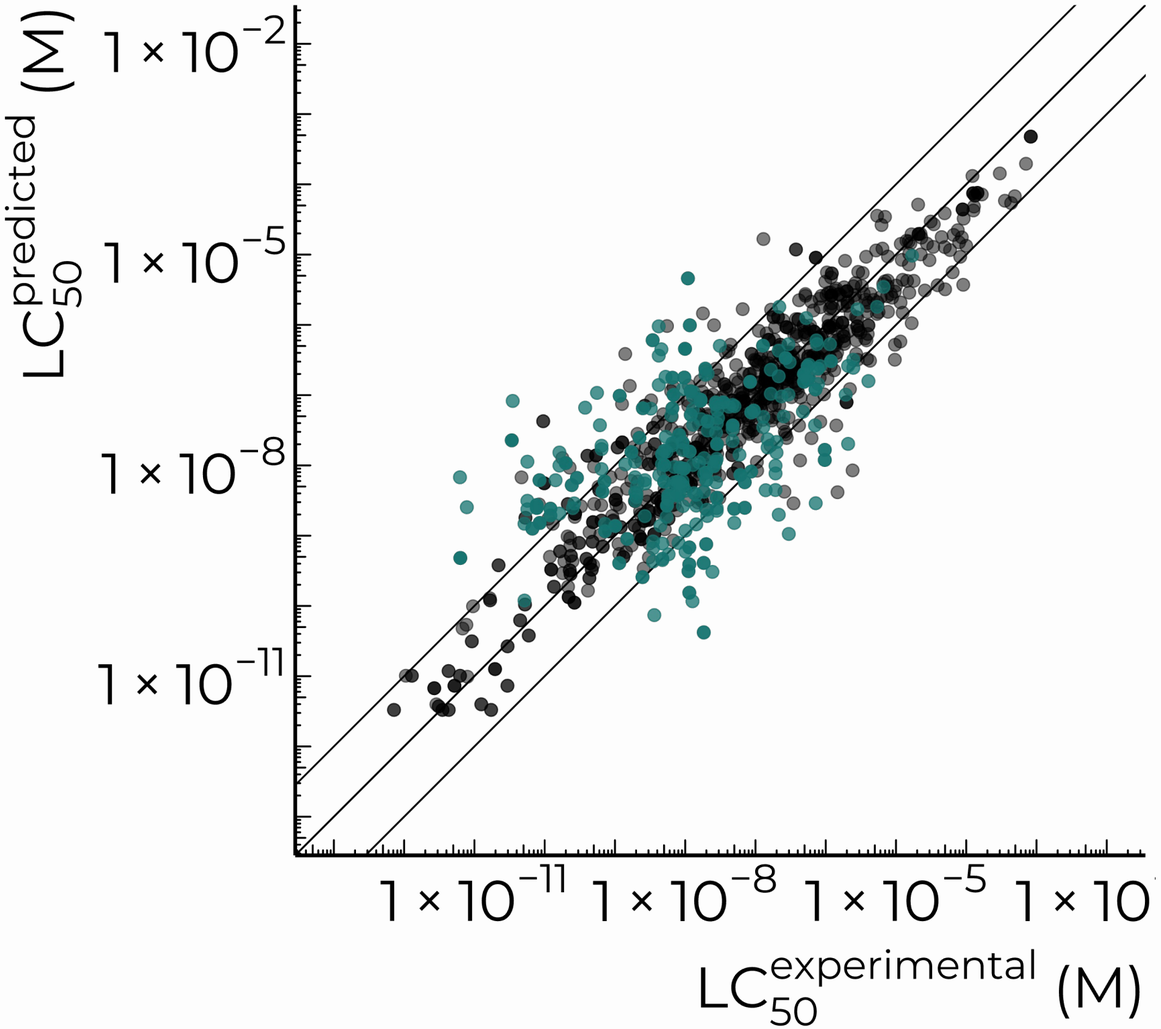
LC_{50} fish



endocrine disruption potential

LC_{50} predictions

- from structure
  RMSE 0.78 log(M)
- from MS^2 data
  RMSE 0.88 log(M)
- experimental
  SD 0.44 log(M)



Peets et al. ES&T 2022

endocrine disruption potential

- Tox21 Data Challenge
- 12 endpoints
- active vs inactive

Rahu et al. JCIM 2024

endocrine disruption potential

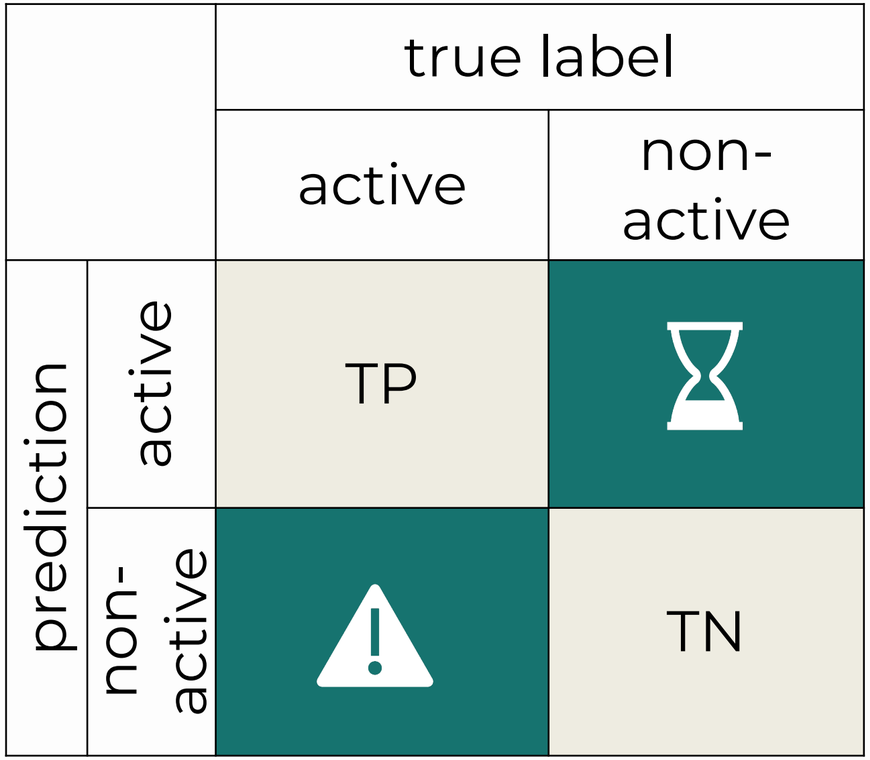
- Tox21 Data Challenge
- 12 endpoints
- active vs inactive

		true label	
		active	non-active
prediction	active	TP	FP
	non-active	FN	TN

Rahu et al. JCIM 2024

endocrine disruption potential

- Tox21 Data Challenge
- 12 endpoints
- active vs inactive



Rahu et al. JCIM 2024

endocrine disruption potential

- Tox21 Data Challenge
- 12 endpoints
- active vs inactive

		true label	
		active	non-active
prediction	active	TP	FP
	non-active	FN	TN

FPR @ TPR = 0.9

Rahu et al. JCIM 2024

endocrine disruption potential

- Tox21 Data Challenge
- 12 endpoints
- active vs inactive

bioassay	FPR
sr.mmp	25.1%
sr.p53	25.4%
nr.ahr	41.8%
...	...
nr.ar	82.4%
nr.er	85.0%

Rahu et al. JCIM 2024

#3

take-home message

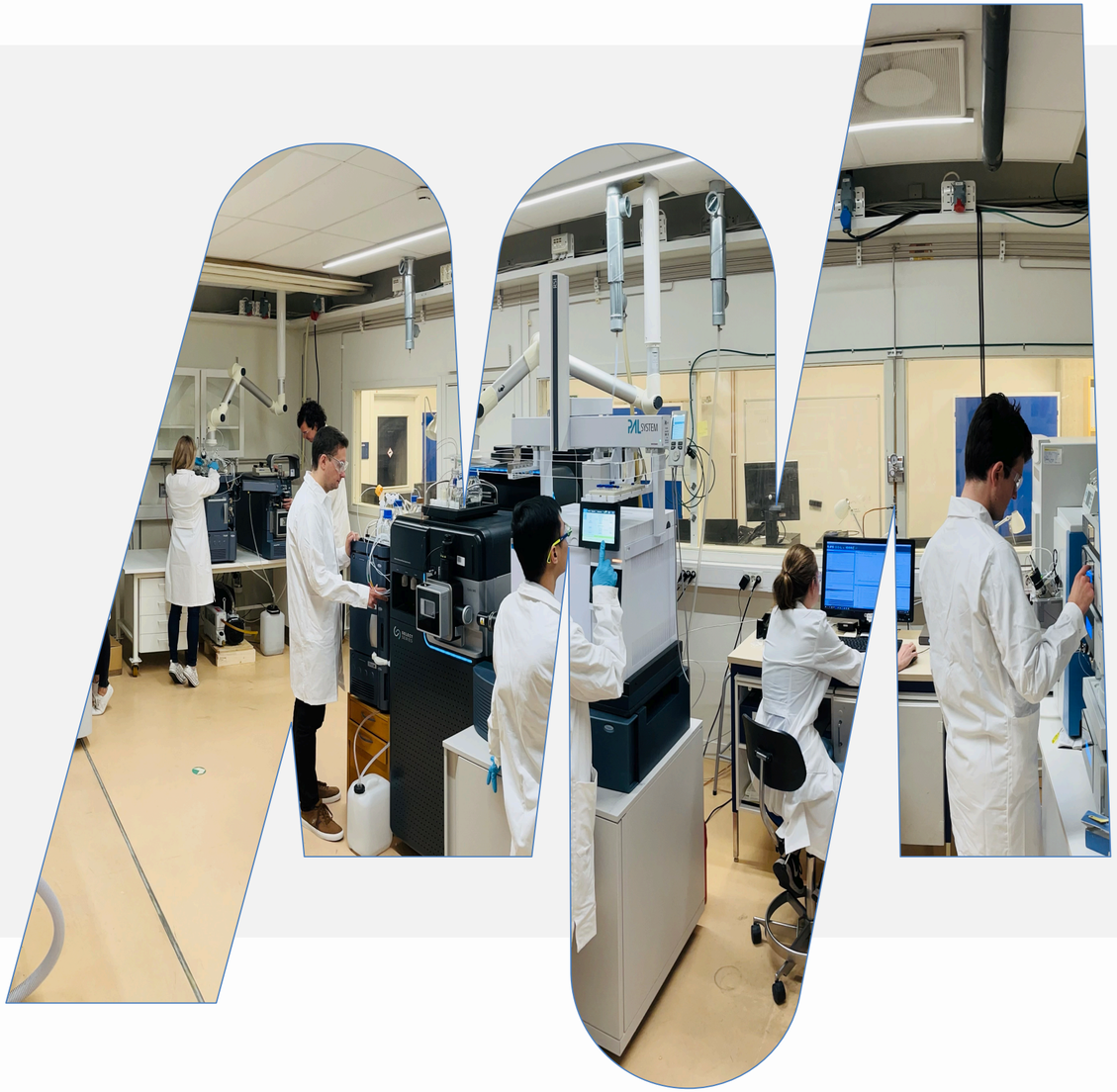
metrics matter

performance needs to be measured with metric relevant for the task

interlab

- organized by
  SLV, FOI, SVA

- aim
  detect and identify
  potentially hazardous
  chemicals in drinking water
  under time stress



interlab

- full sample activity BioCell Analytica

Machine Translated by Google

BioCell ANALYTICA

Table 1. The following BEQ values were measured in the current samples:

	Nrf2 activity	AntiÿAR activity	AR activity	ER activity	AhR activity
	µg/L (tBHQ equivalents)	ng/L (OHF equivalents)	ng/L (DHT equivalents)	pg/L (E2 equivalents)	ng/L (TCDD equivalents)
Reference to sample 1	<LOD	<LOD	<LOD	<LOD	<LOD
Sample 1	<LOD	<LOD	79300	784	0.0814
Reference to sample 2	21.1	73.6	<LOD	21.7	<LOD
Sample 2	992	2670	<LOD*	<LOD*	<LOD*
Detection limit	8.34	43.8	0.122	12.5	0.0196
detection limit*			6.93	50.0	0.156

Table 2. Genotoxicity.

	Genotoxic?
Reference to sample 1	No
Sample 1	No
Reference to sample 2	No
Sample 2	Could not be determined*

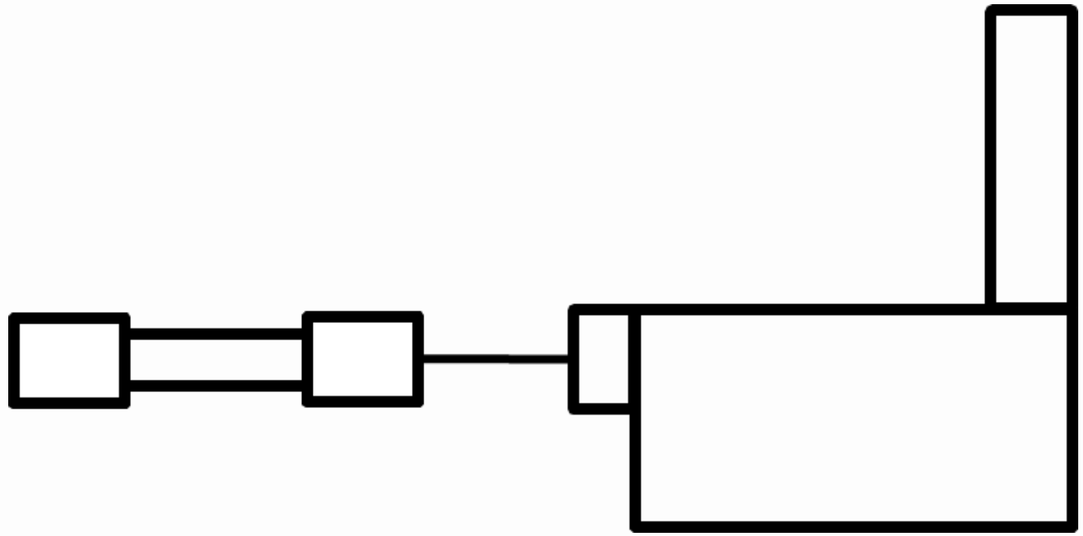
*Due to extensive cytotoxicity, despite repeated analyses, it could not be determined whether sample 2 was genotoxic or not. The sample was tested down to the concentration REF 12.5, but even then was too cytotoxic to be able to determine if it was genotoxic.

BioCell Analytica Uppsala AB
Ulls väg 29C, 756 51 Uppsala
biocellanalytica.se
kontakt@biocellanalytica.se

Sandberg, Rahu, et al. in preparation

interlab

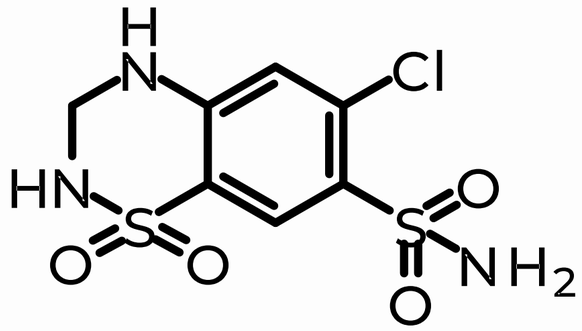
- full sample activity
  BioCell Analytica
- LC/HRMS
  4700 features



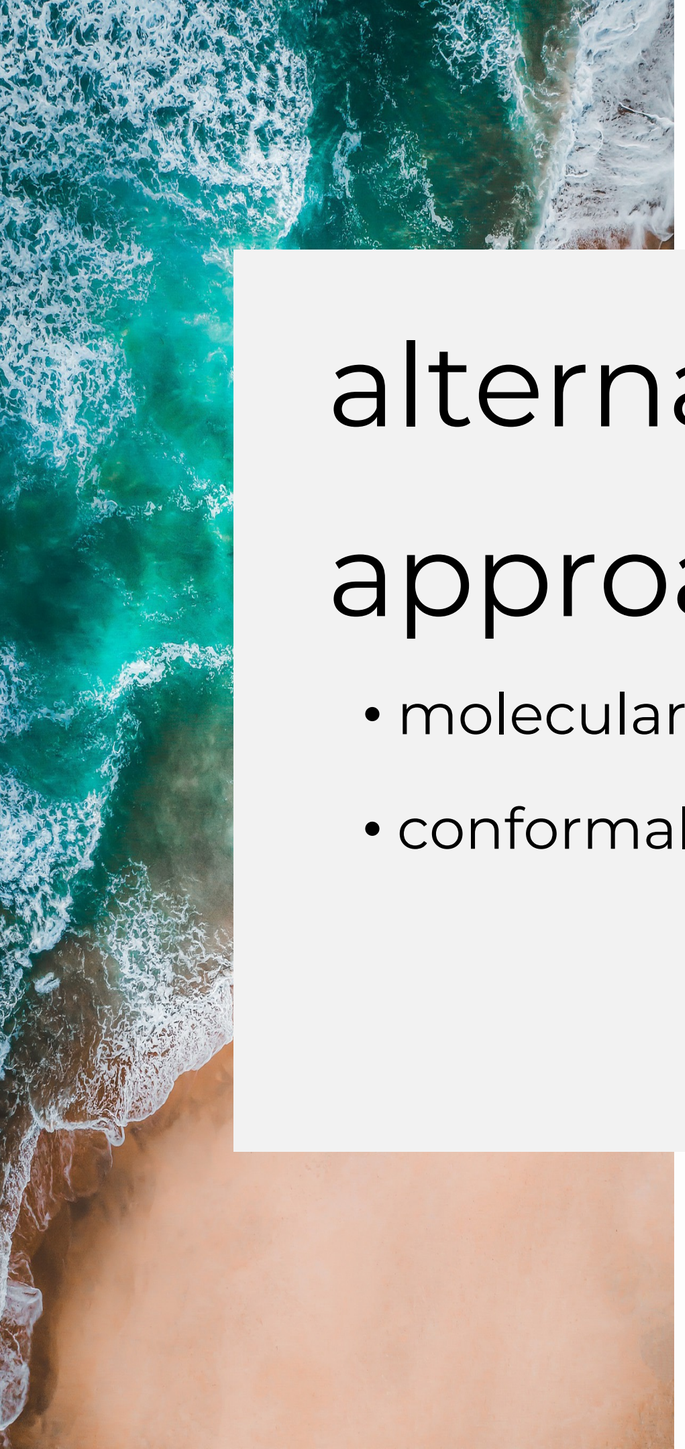
Sandberg, Rahu, et al. in preparation

interlab

- full sample activity
  BioCell Analytica
- LC/HRMS
  4700 features
- predicted AhR activity
  55 features



Sandberg, Rahu, et al. in preparation

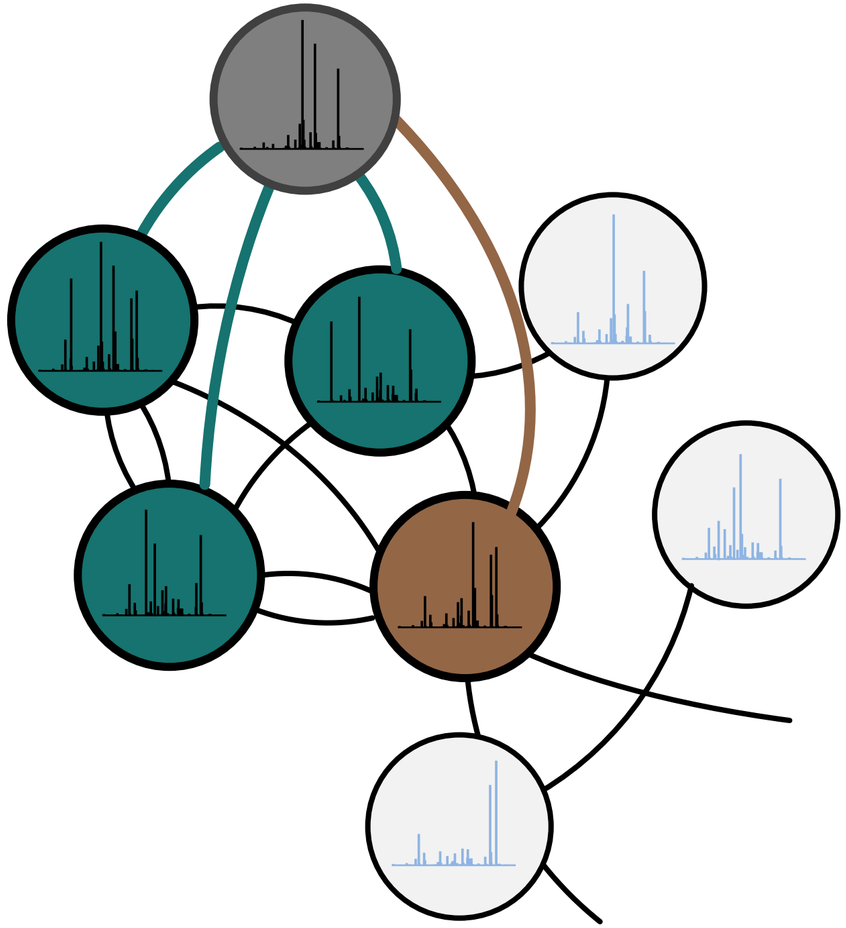


alternative approaches

- molecular networking
- conformal predictions

molecular networking

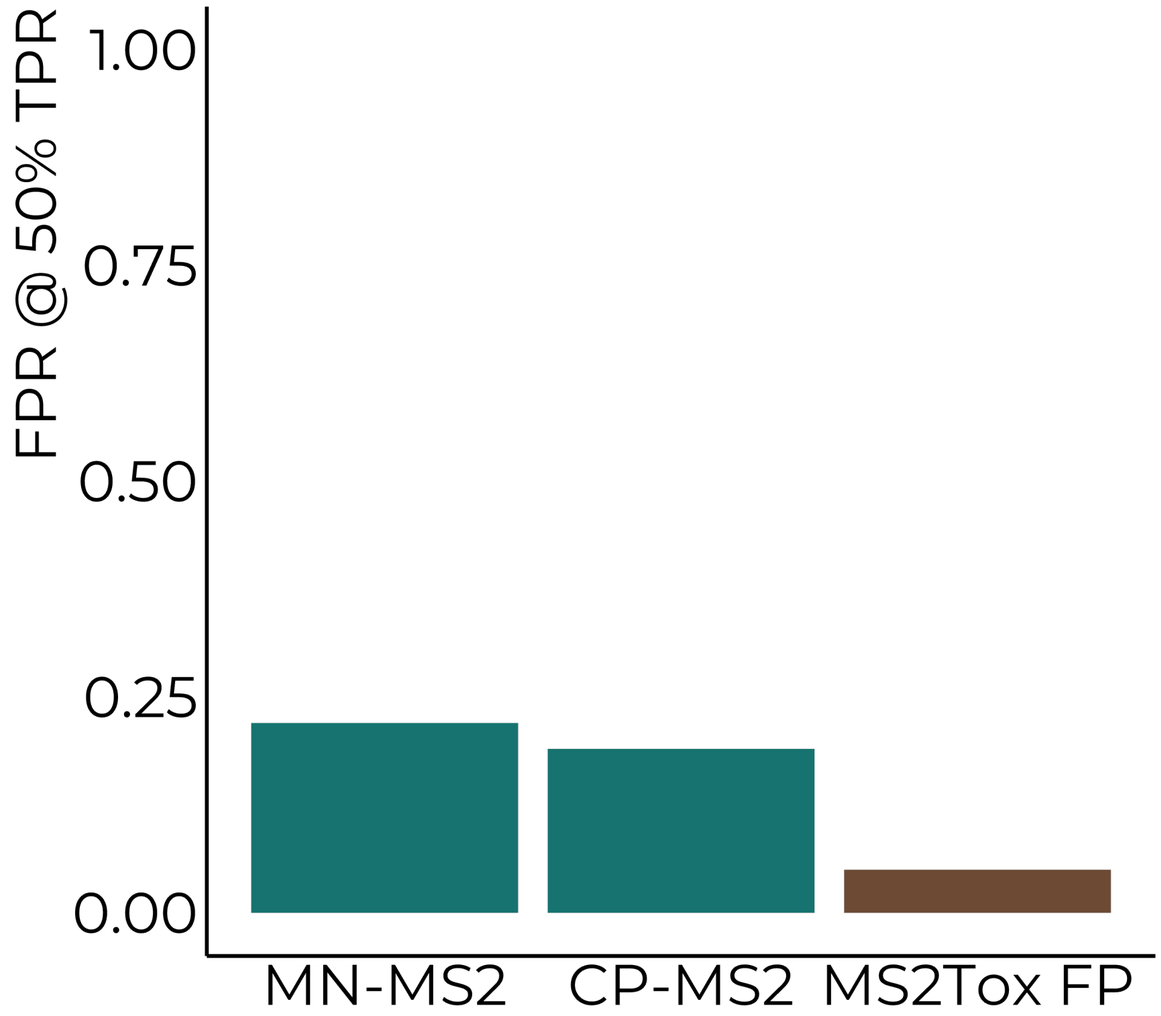
- similarity of MS^2 spectra
  MS2DeepScore
- consensus vote
  immediate neighbors



Kreutzer et al. in preparation

comparison of approaches

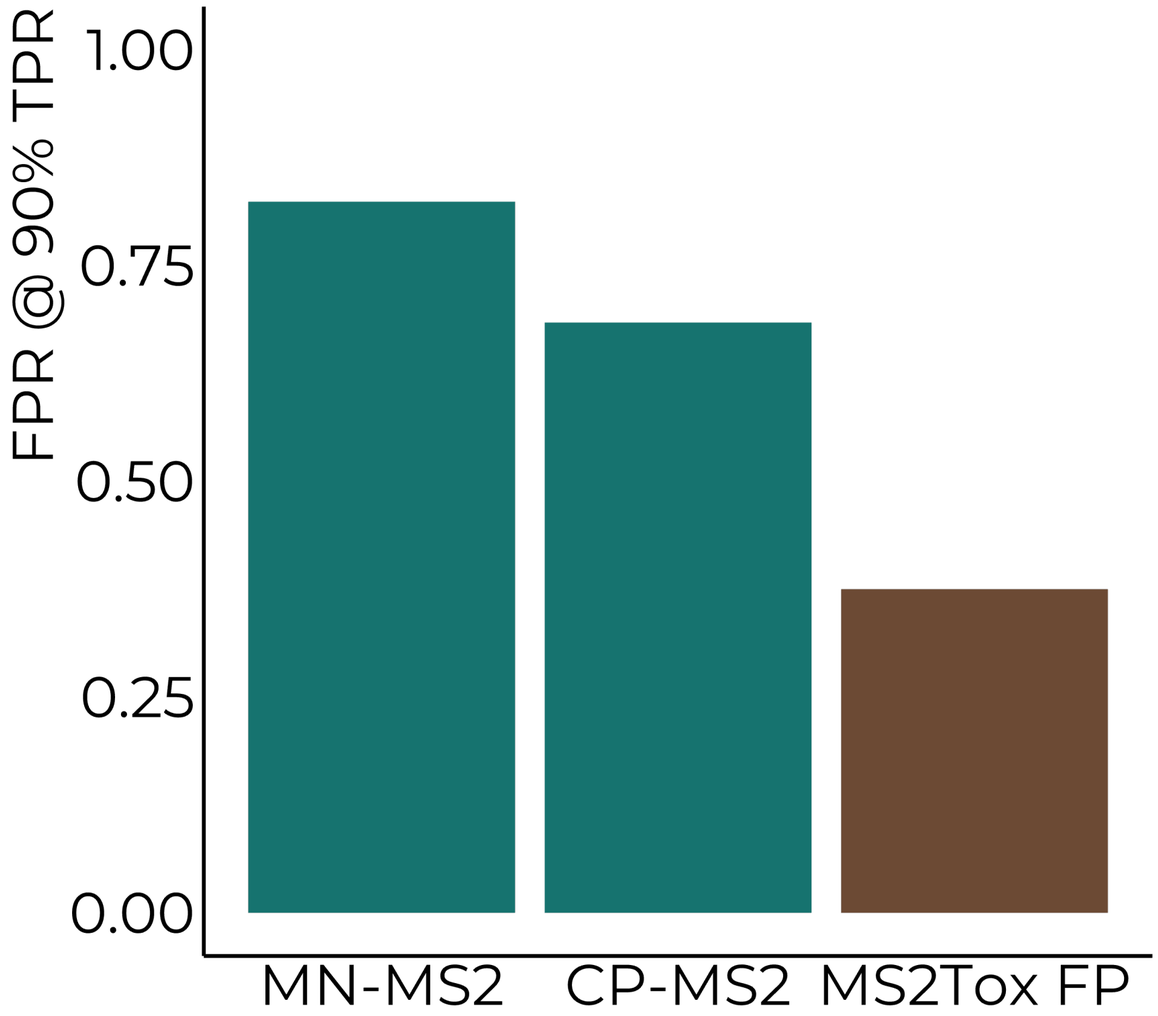
- unsupervised approach yields high FPR
- MS^2 similarity is insufficient



Kreutzer et al. in preparation

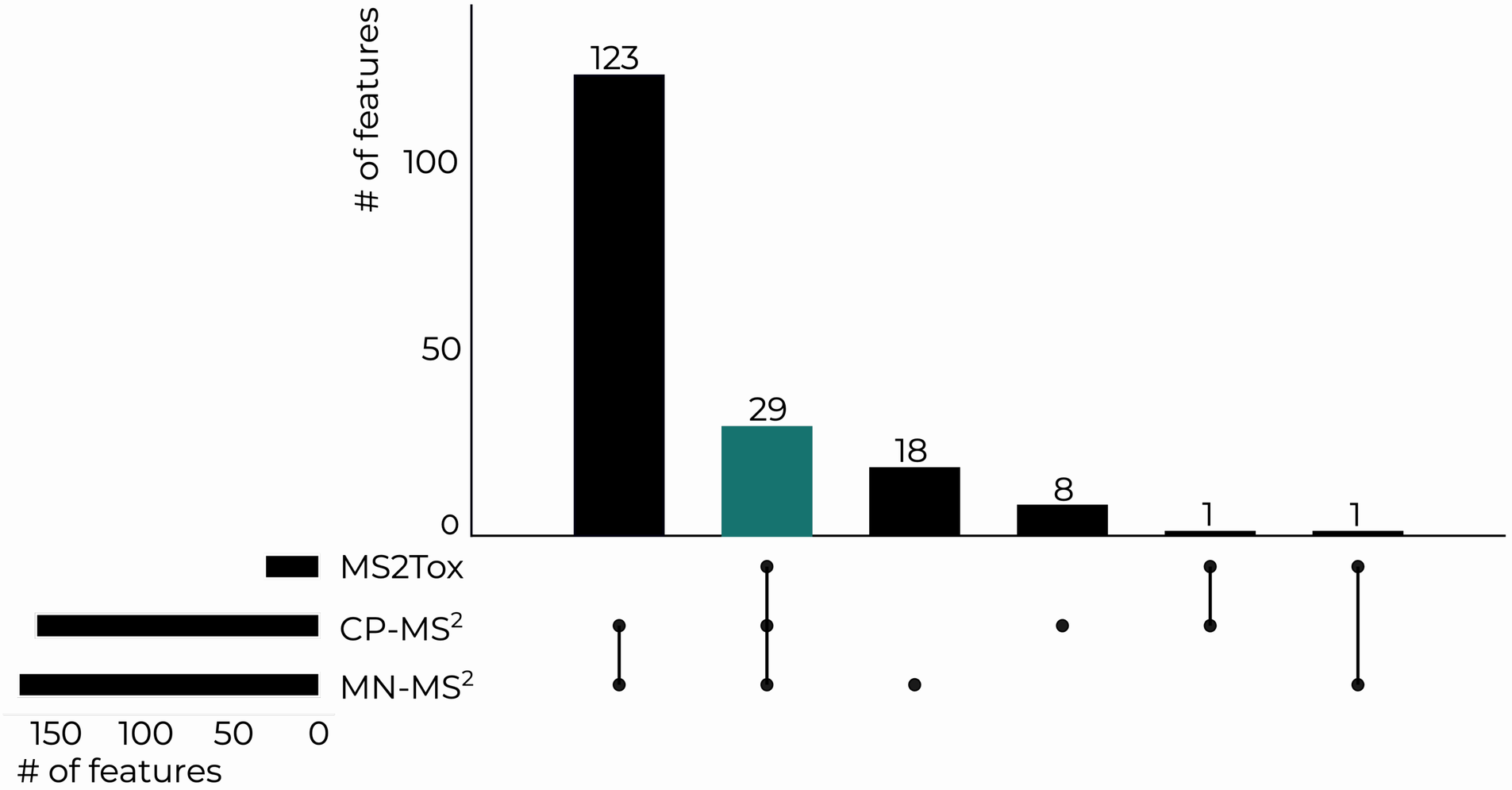
comparison of approaches

- unsupervised approach yields high FPR
- MS^2 similarity is insufficient



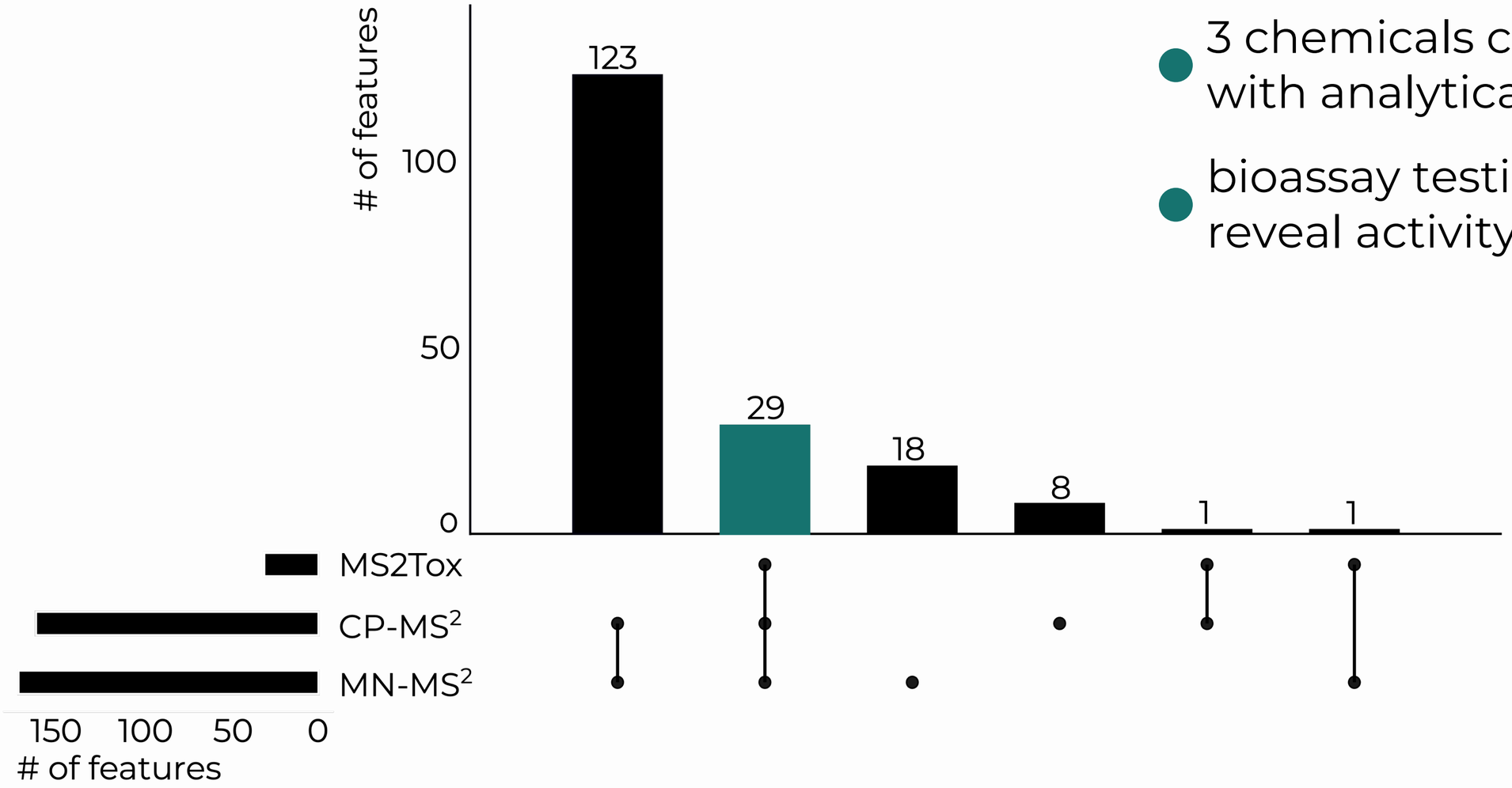
Kreutzer et al. in preparation

combining approaches



Kreutzer et al. in preparation

combining approaches



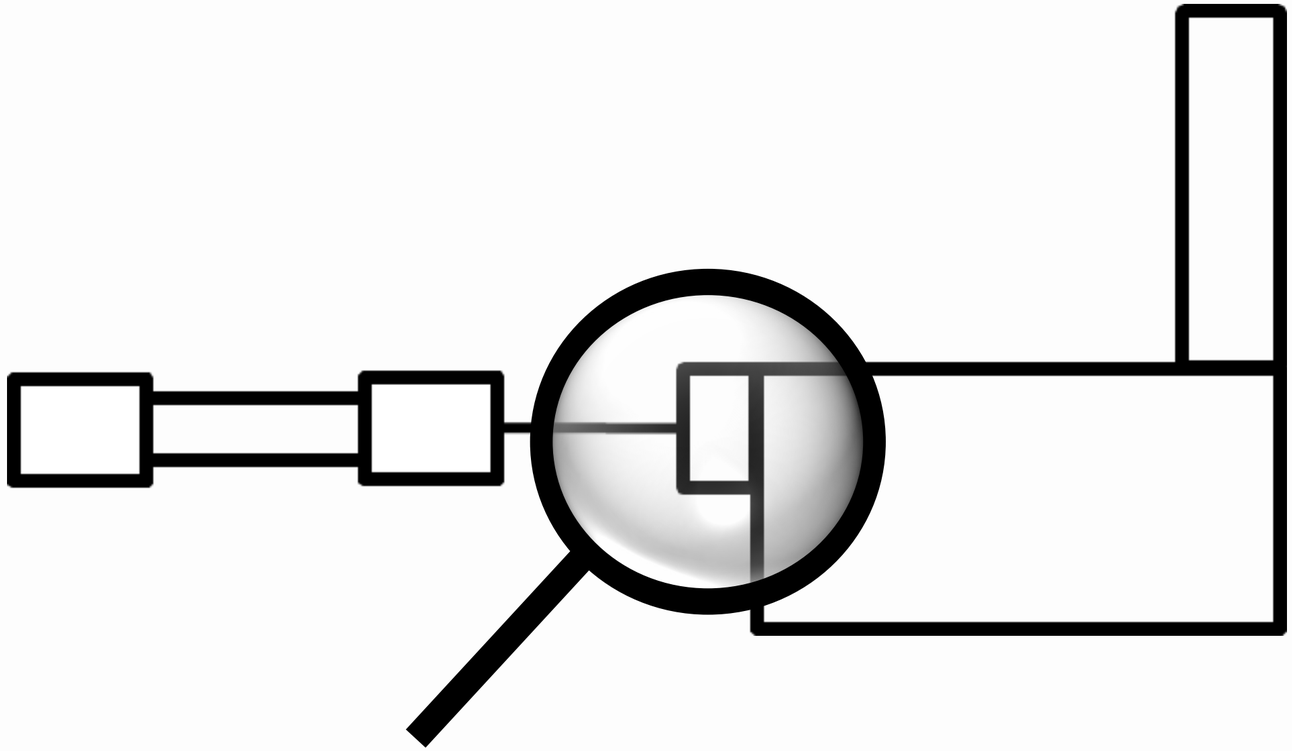
- 3 chemicals confirmed with analytical standards
- bioassay testing did not reveal activity

Kreutzer et al. in preparation

quantification

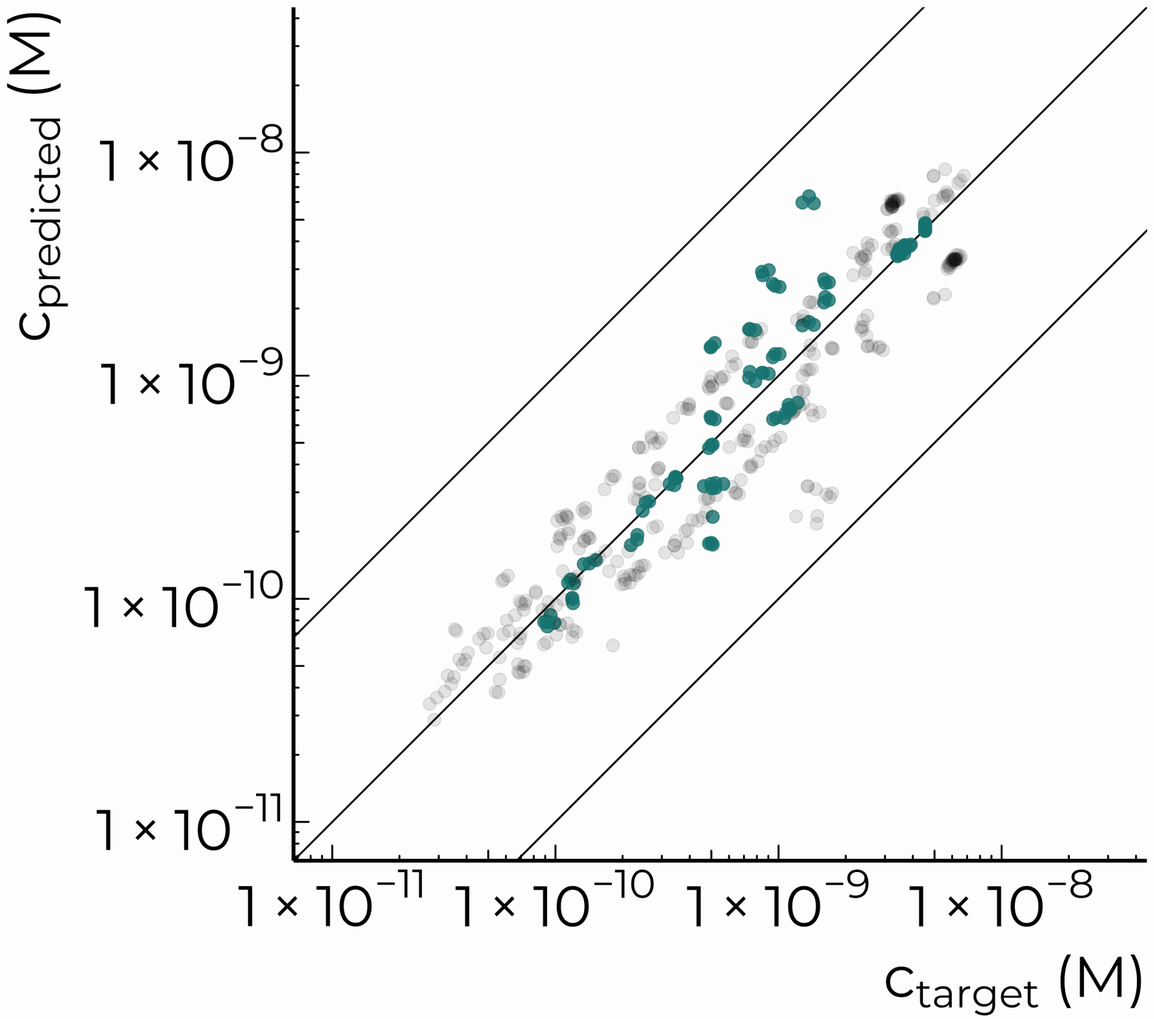
- lack of analytical standards
- different ionization efficiencies

electrospray ionization



quantification

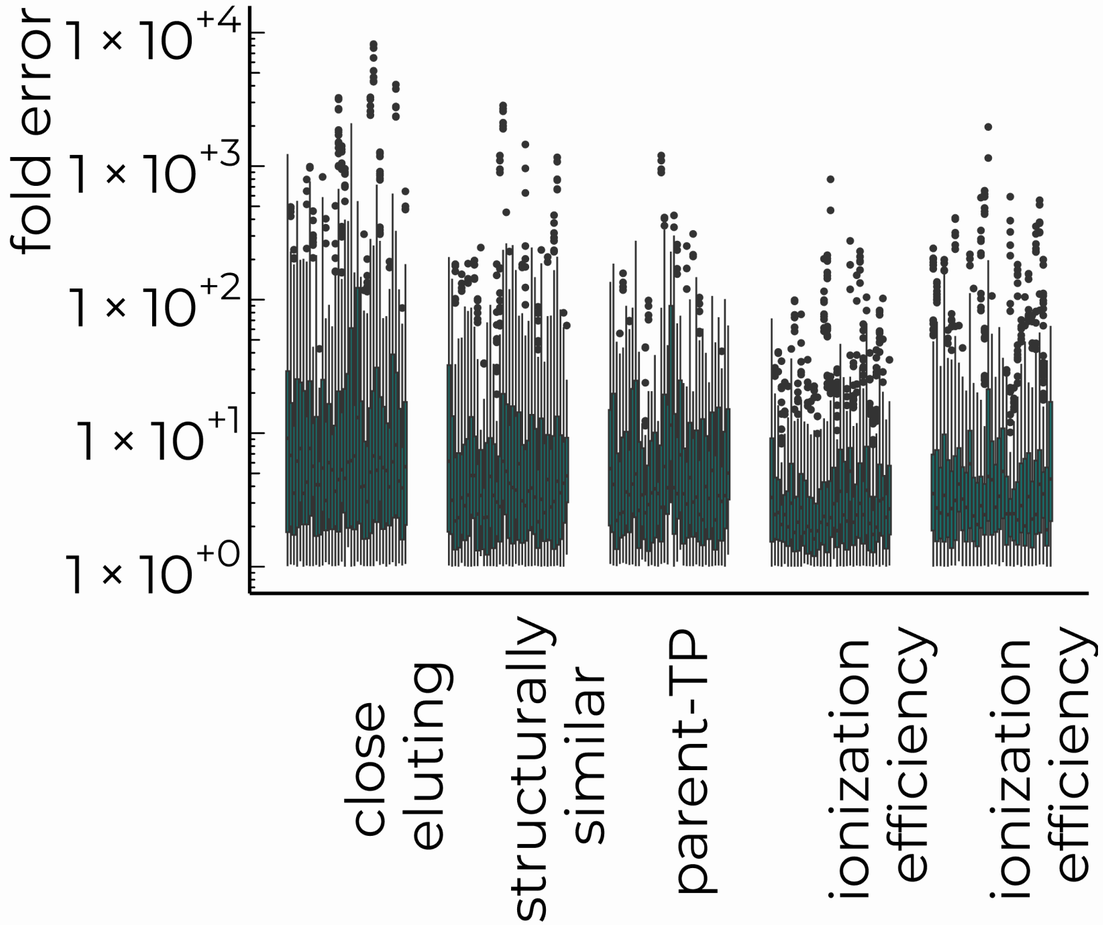
- from structure
  mean error 1.7x



Been et al. Water Research 204 (2021) 117612

quantification

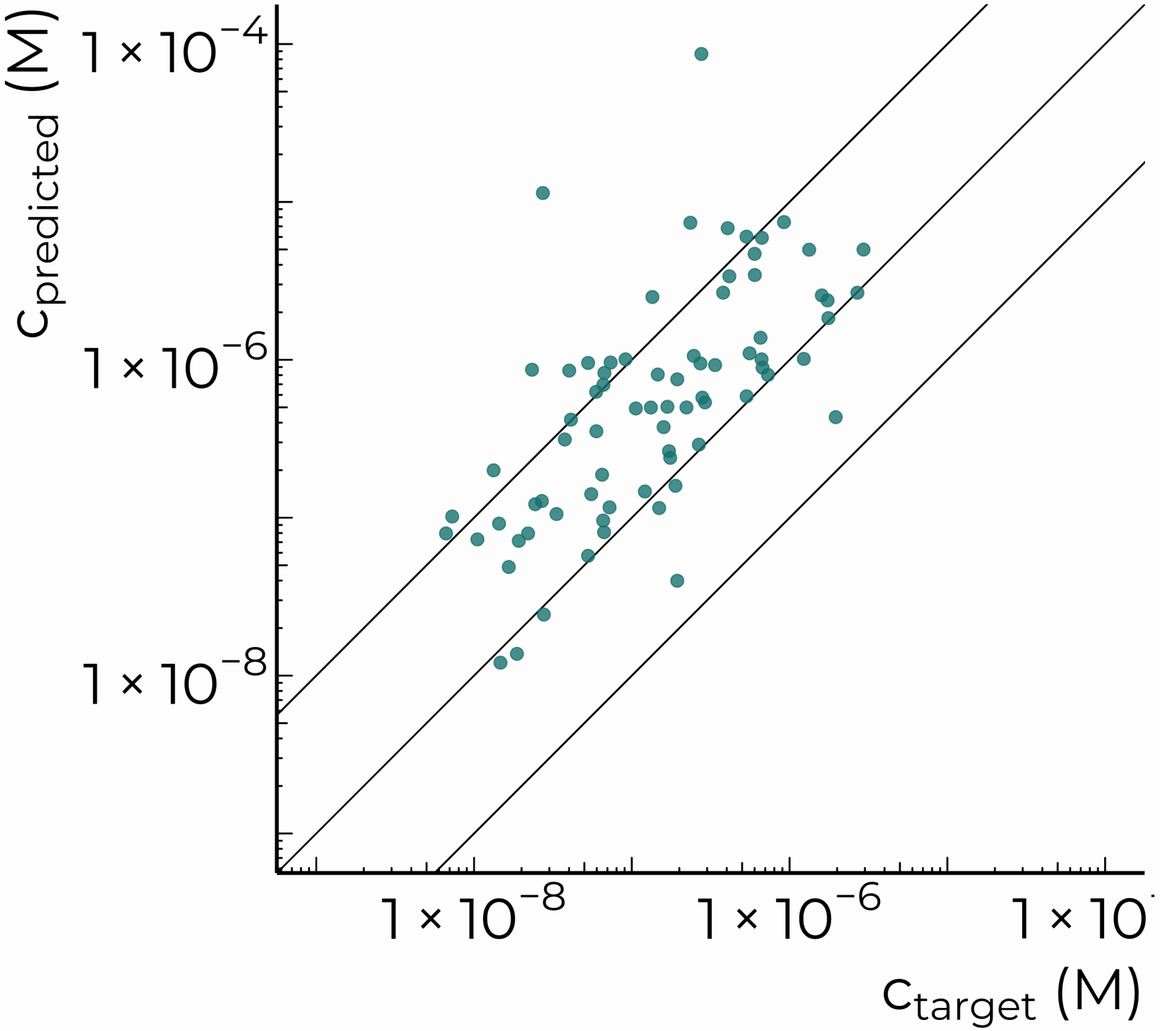
- from structure
  mean error 1.7x
- tested in interlab
  RMSE 0.88 log(M)



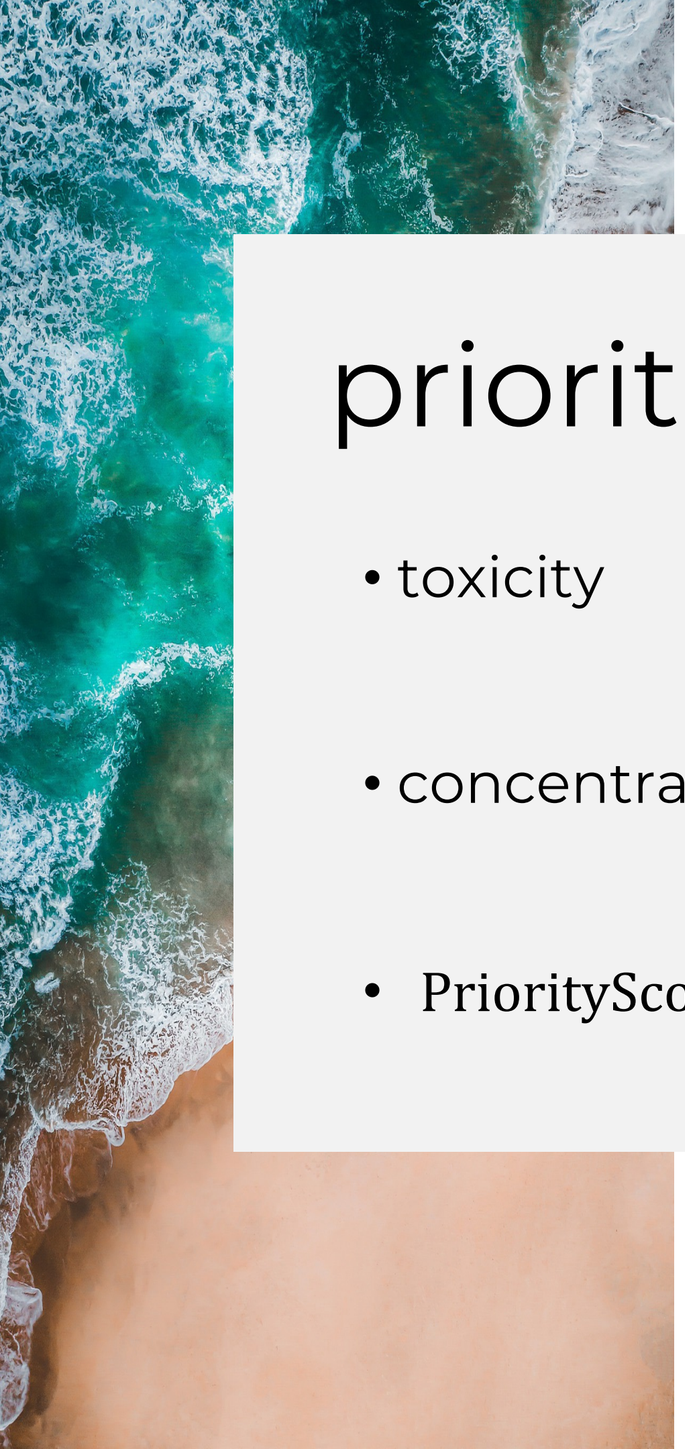
Malm et al. Anal. Chem. 96 (2024) 16215

quantification

- from structure
  mean error 1.7x
- tested in interlab
  RMSE 0.88 log(M)
- from MS^2 data
  mean error 9.5x



Sepman et al. Anal. Chem. 95 (2023) 12329

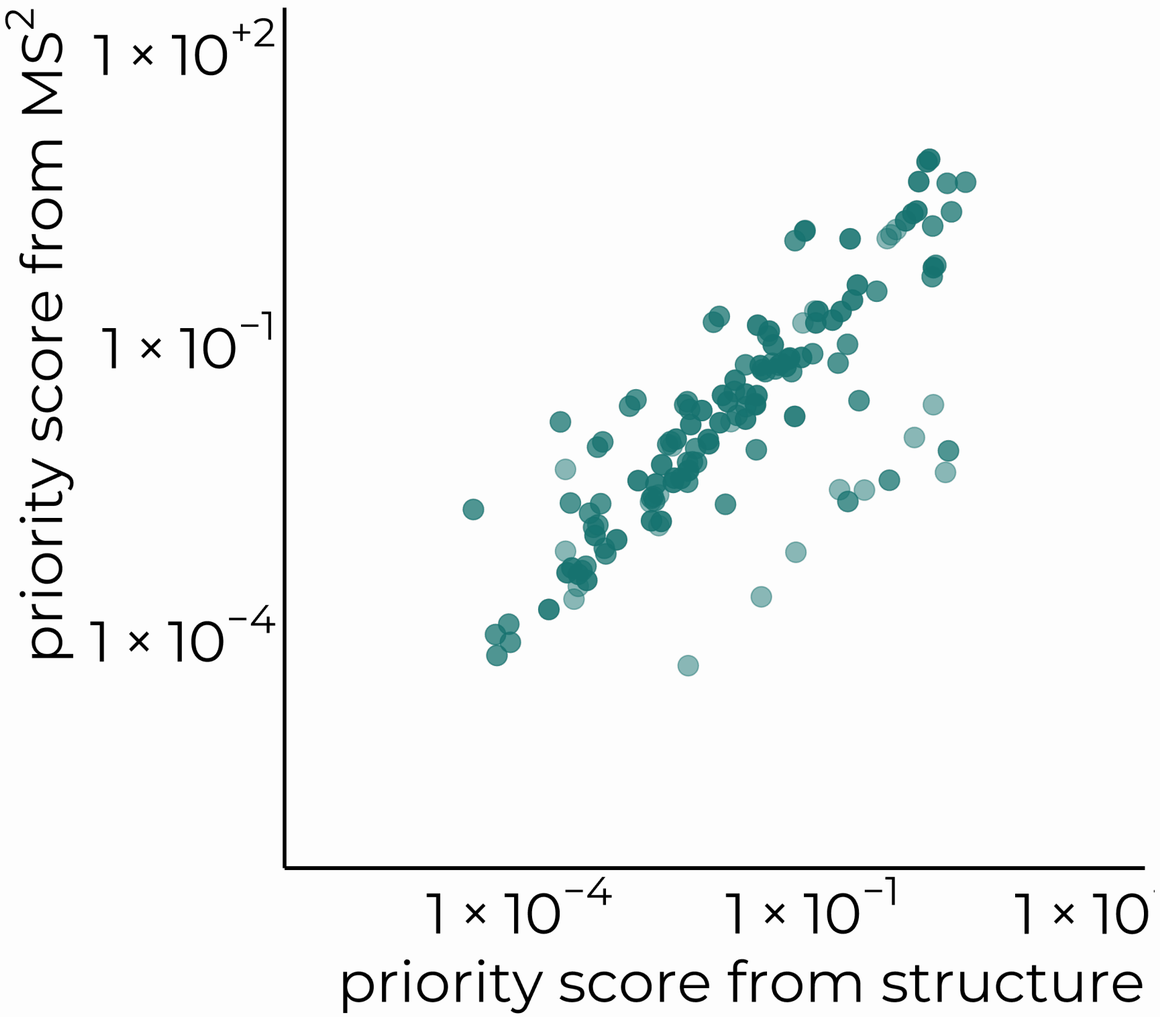


prioritization

- toxicity
- concentration
- $$\text{PriorityScore} = \frac{C_{\text{predicted}}}{AC_{50}^{\text{5th percentile}}}$$

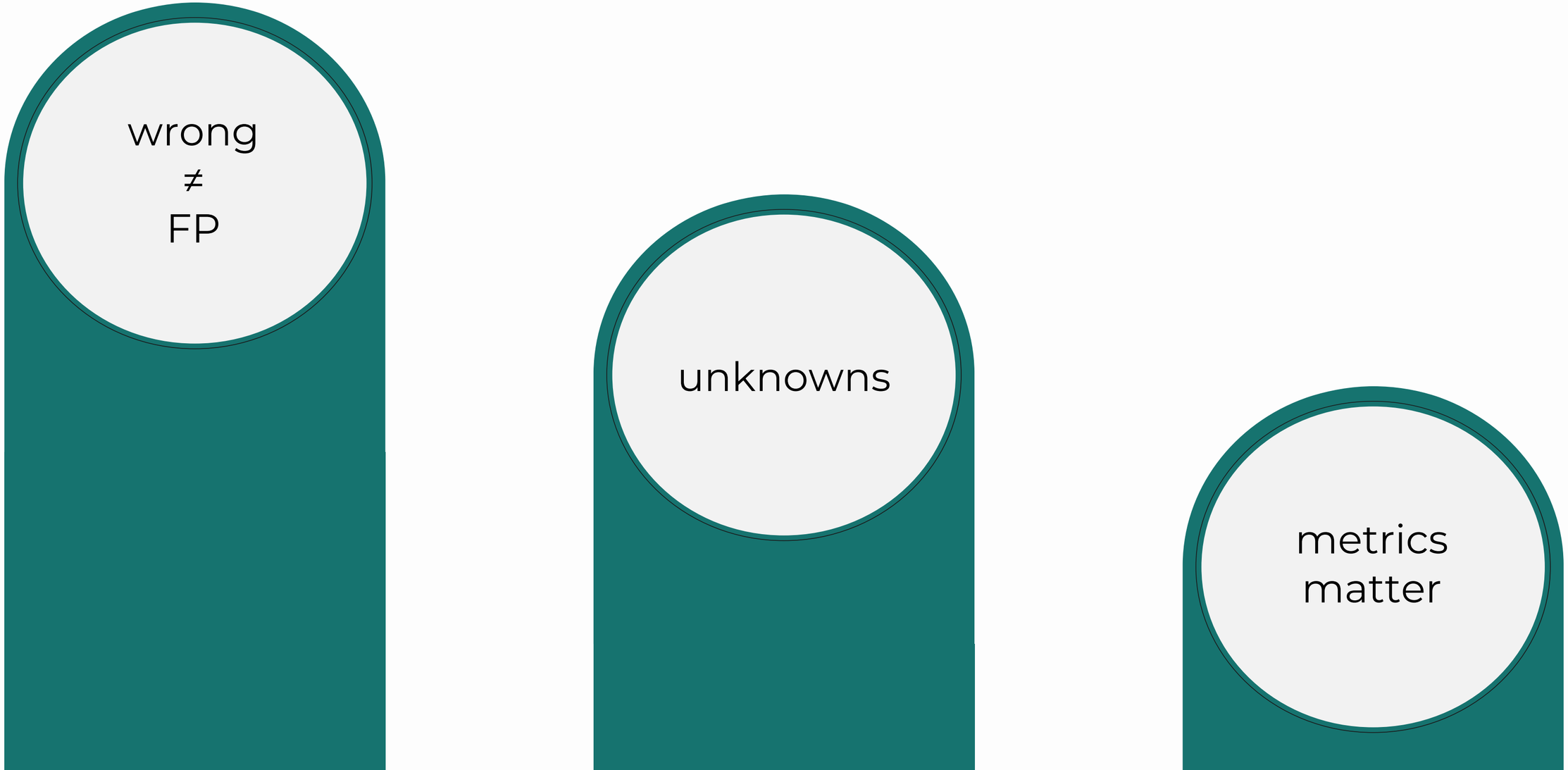
priority score

- fish LC_{50} & concentration
- MS^2 and structure based predictions agree
- acquisition mode impacts



Sepman et al. Anal. Chem. 95 (2023) 12329

take-home messages



anneli kruve

anneli.kruve@su.se

erc

Swedish Research Council

FORMAS

CARL TRYGGERS STIFTELSE FÖR VETENSKAPLIG FORSKNING

Stockholm University

SUCCeSS