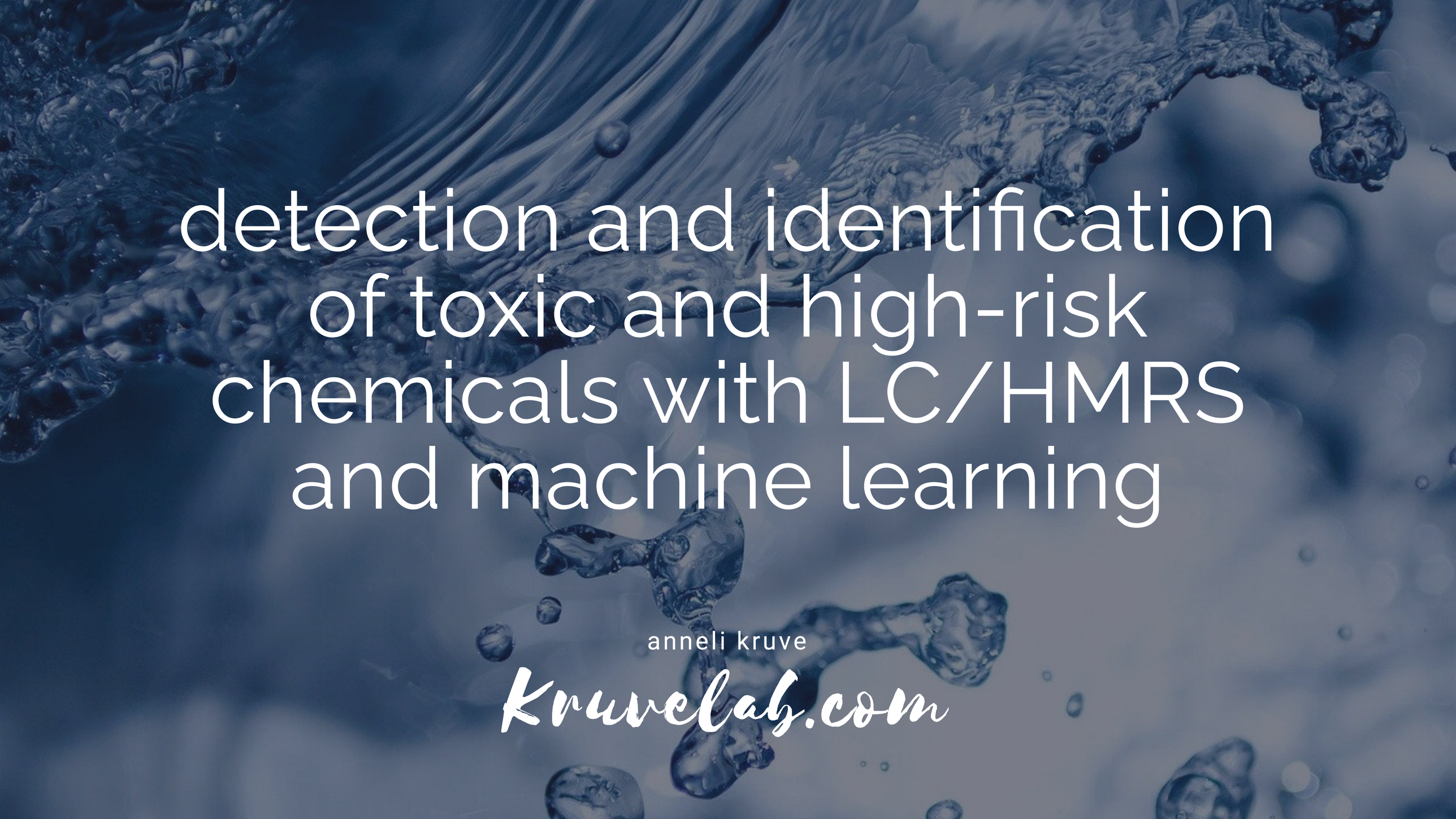


detection and identification of toxic and high-risk chemicals with LC/HMRS and machine learning

anneli kruve

Kruvelab.com

increased threat to Swedish drinking water

an offence more than every fifth day

anneli kruve

DAGENS NYHETER.

SVERIGE

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

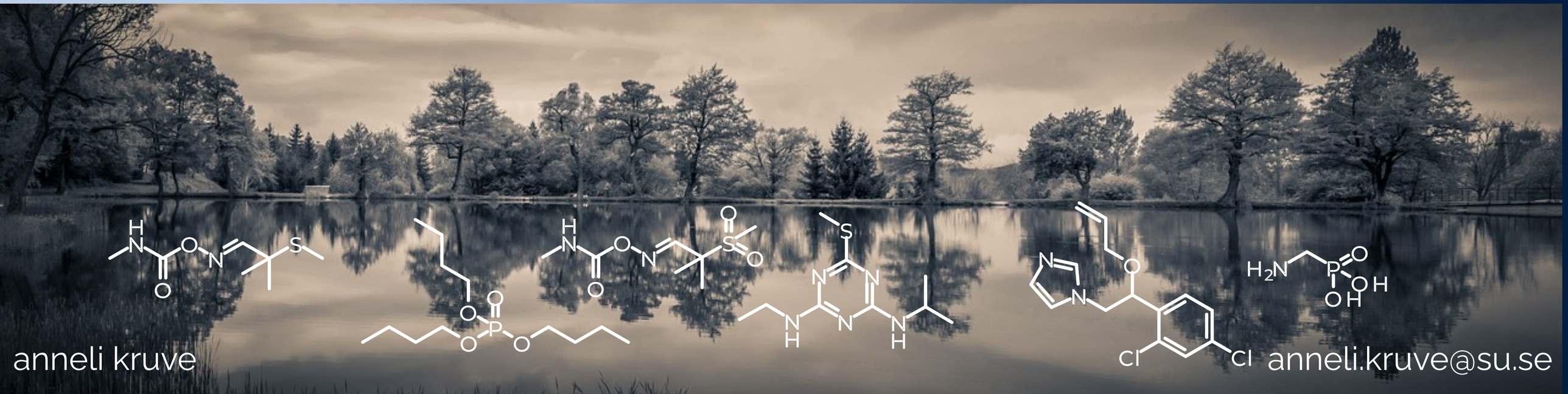
Uppdaterad 2024-12-02 Publicerad 2024-12-01

water analysis

anneli kruve

anneli.kruve@su.se

water analysis

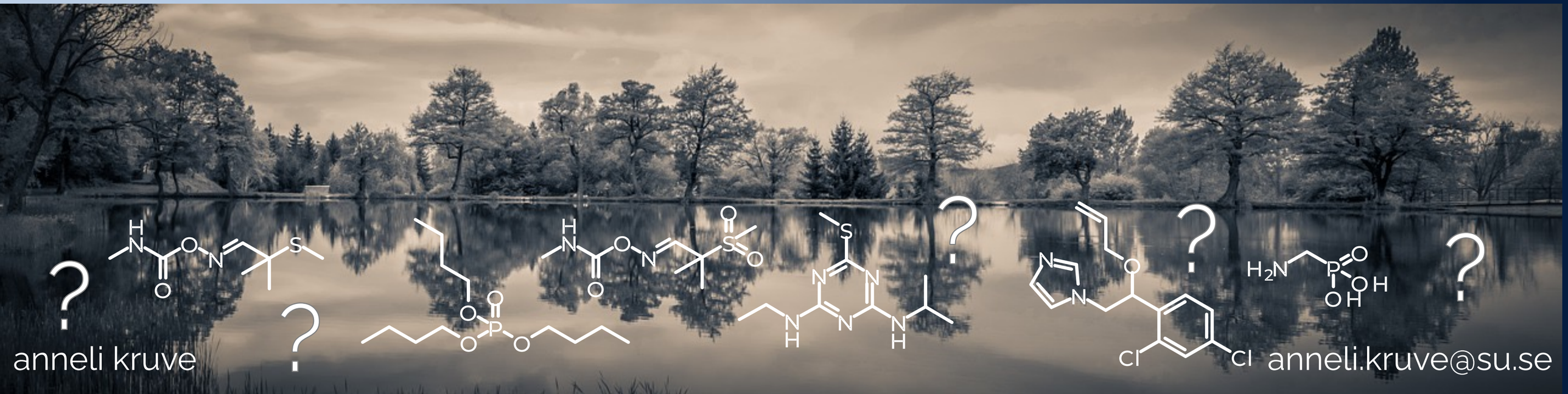


anneli kruve

anneli.kruve@su.se

water analysis

thousands of chemicals detected



anneli kruve

anneli.kruve@su.se

water analysis

thousands of chemicals detected

Machine Translated by Google

BioCell ANALYTICA

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

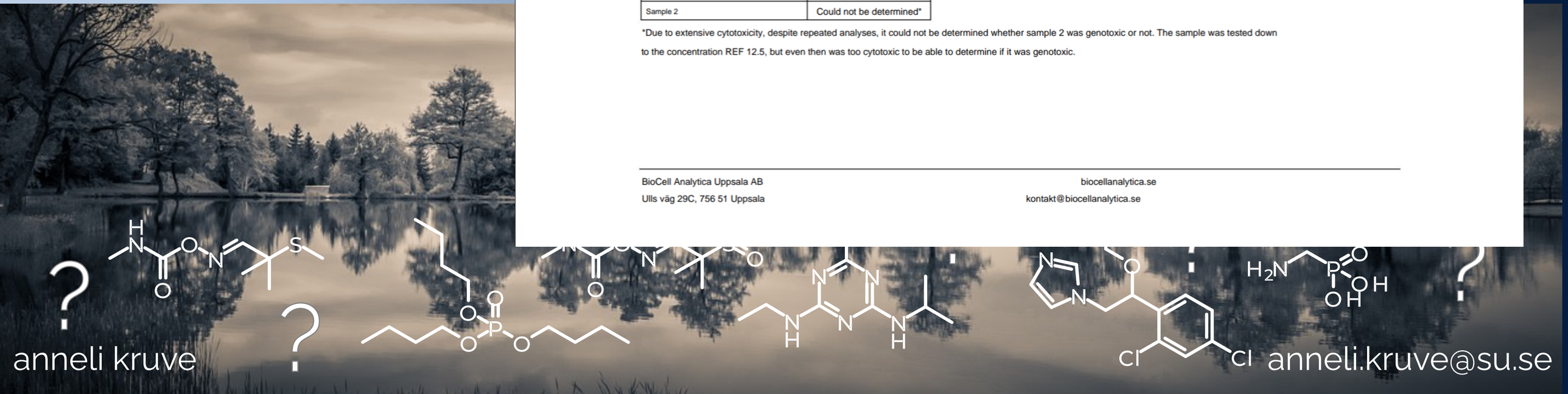
	Nrf2 activity	AntiÿAR activity	AR activity	ER activity	AhR activity
	µg/L (tBHQ equivalents)	ng/L (OHF equivalents)	ng/L (DHT equivalents)	pg/L (E2 equivalents)	ng/L (TCDD equivalents)
Reference to sample 1	<LOD	<LOD	<LOD	<LOD	<LOD
Sample 1	<LOD	<LOD	79300	784	0.0814
Reference to sample 2	21.1	73.6	<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



anneli kruve
anneli.kruve@su.se

how to ...

PRIORITIZE

risk

IDENTIFY

structure

anneli kruve

anneli.kruve@su.se

how to ...

PRIORITIZE

risk

IDENTIFY

structure

anneli kruve

anneli.kruve@su.se

prioritization of chemicals

TOXICITY

ecotoxicity and endocrine disruptors

CONCENTRATION

exposure to potentially toxic chemicals

RISK

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

anneli kruve

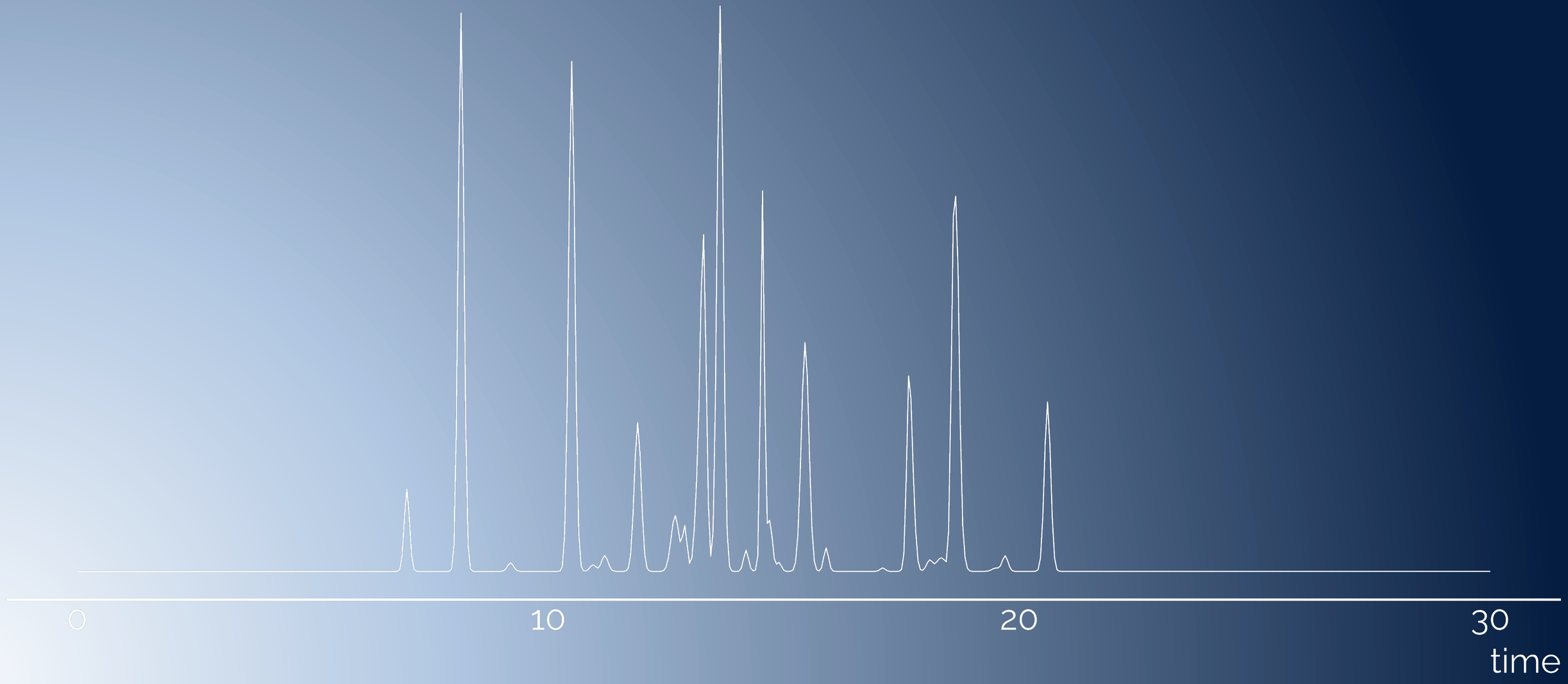
anneli.kruve@su.se

toxicity

of detected chemicals

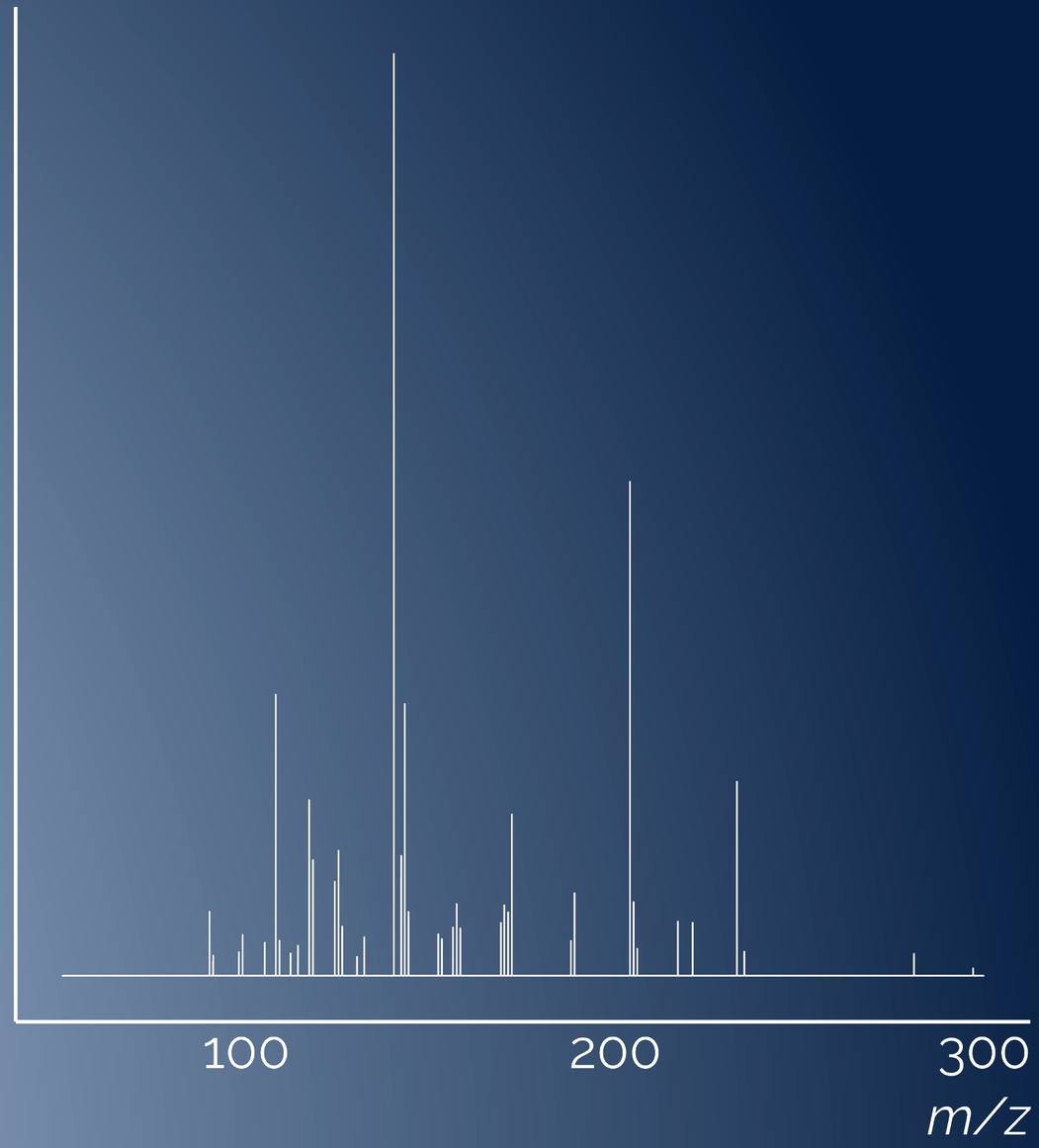
nontarget screening

with LC/HRMS



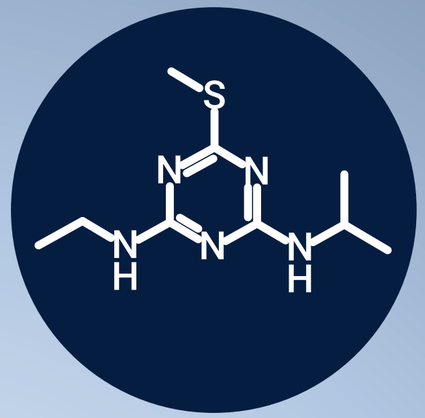
toxicity assessment

from spectra
to structure
to toxicity



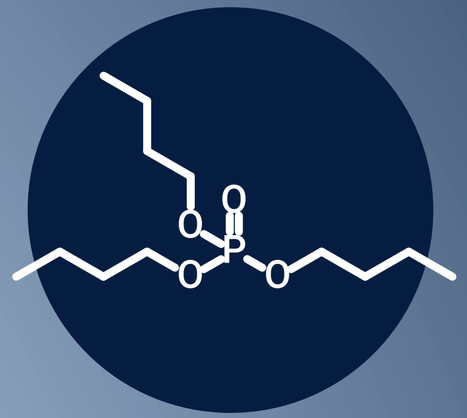
anneli kruve
anneli.kruve@su.se

toxicity assessment



LC_{50} = 9.3 mg/L

known structure
known toxicity



LC_{50} = ? mg/L

known structure
unknown toxicity

LC_{50} = ? mg/L

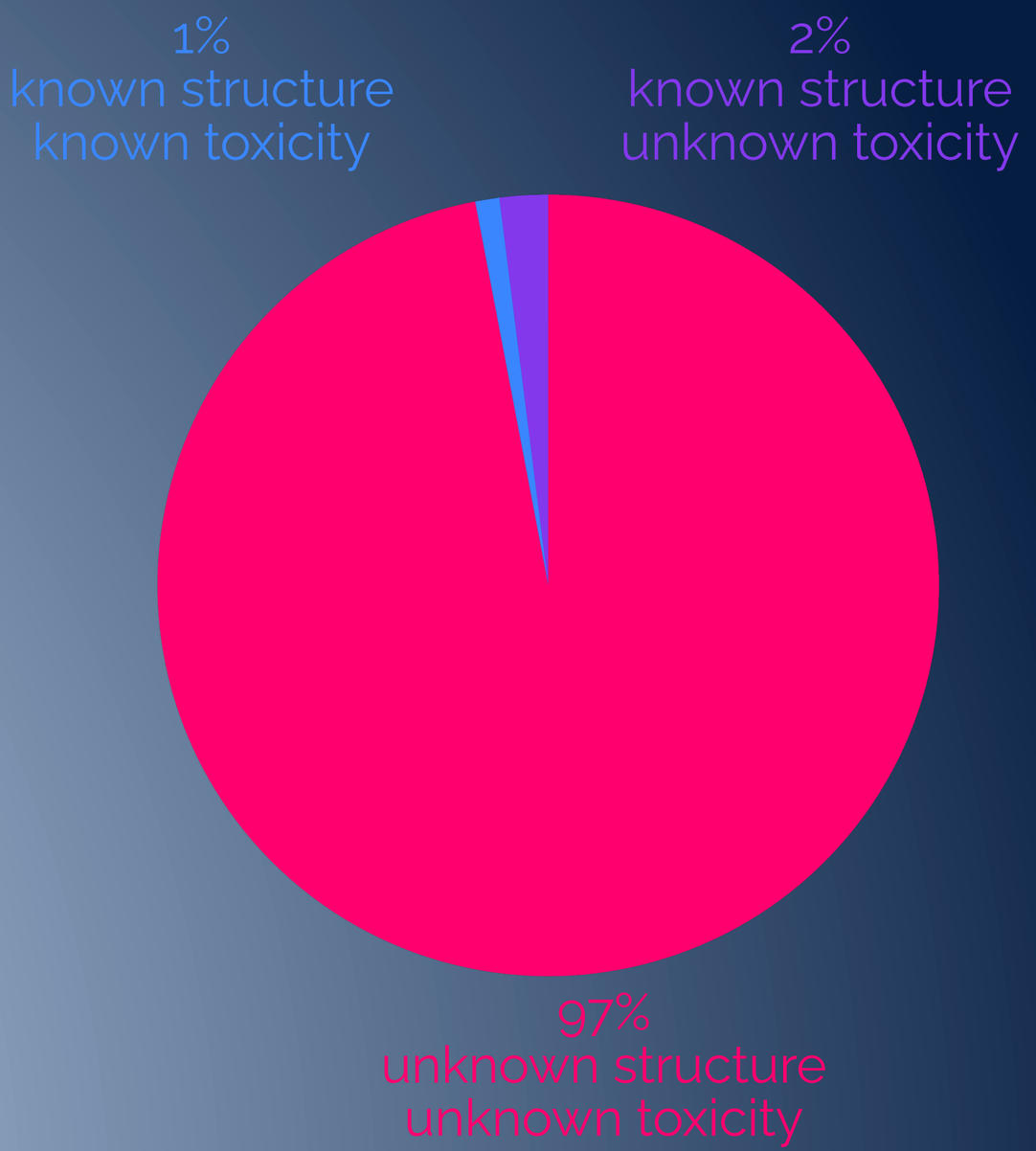
unknown structure
unknown toxicity

anneli kruve

anneli.kruve@su.se

toxicity assessment

vast majority of detected chemicals remain unknown



anneli kruve

anneli.kruve@su.se

predicting toxicity

of detected chemicals

endpoints

ECOTOXICITY

fathead minnow, bluegill,
and rainbow trout

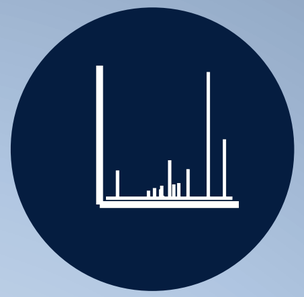
ENDOCRINE DISRUPTION

AhR, AR, ER, MMP, P53, ...

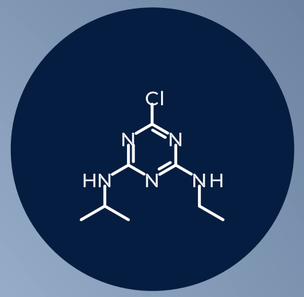
anneli kruve

anneli.kruve@su.se

workflow



MS² spectra



structure as SMILES

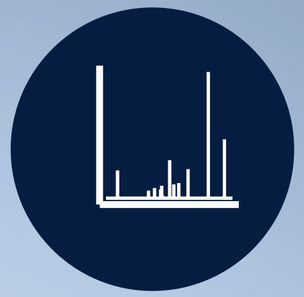
molecular descriptors

toxicity prediction

anneli kruve

anneli.kruve@su.se

workflow



MS^2 spectra

molecular descriptors

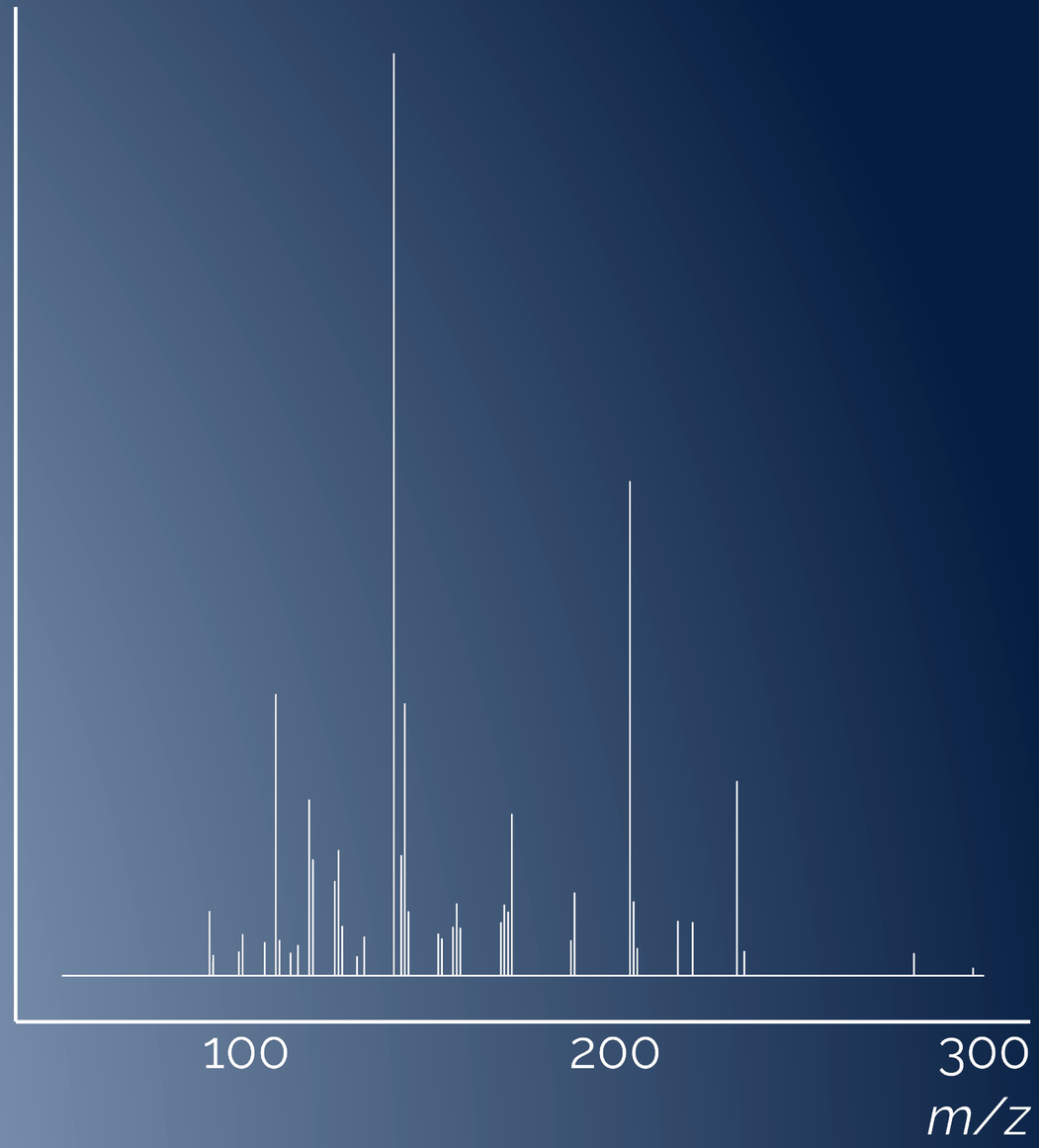
toxicity prediction

anneli kruve

anneli.kruve@su.se

information available

in mass spectra

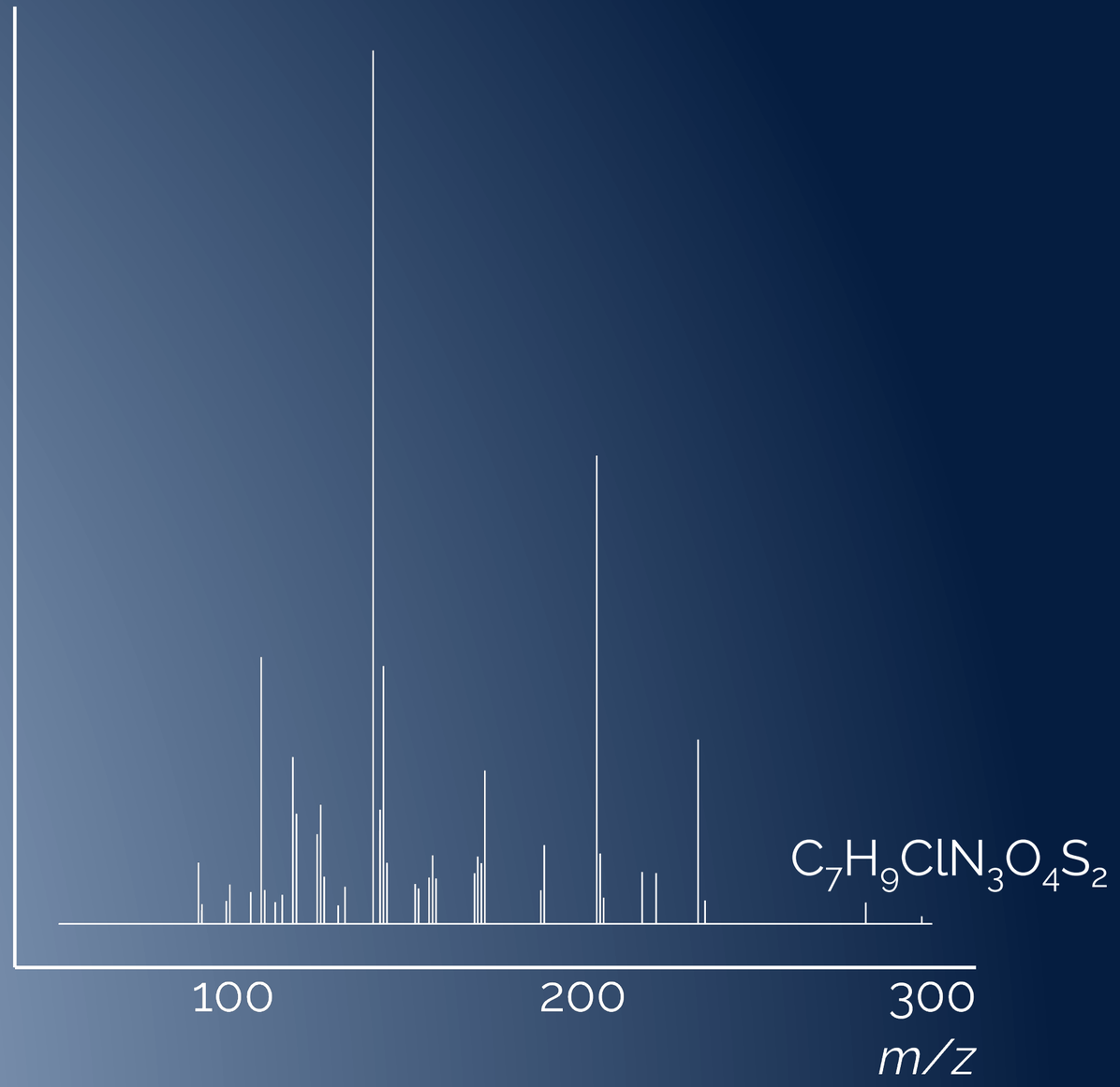


anneli kruve

anneli.kruve@su.se

information available

in mass spectra

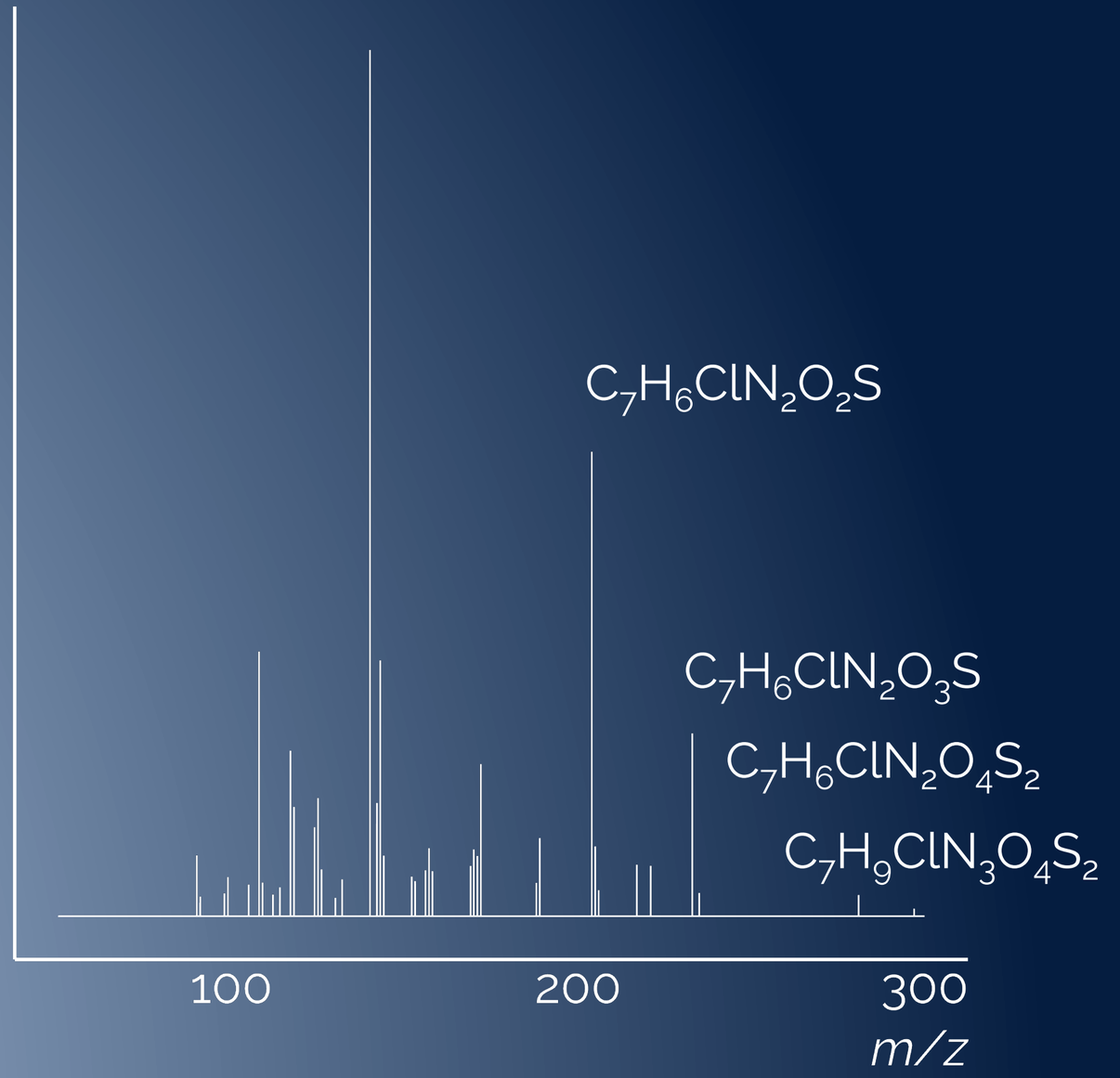


anneli kruve

anneli.kruve@su.se

information available

in mass spectra

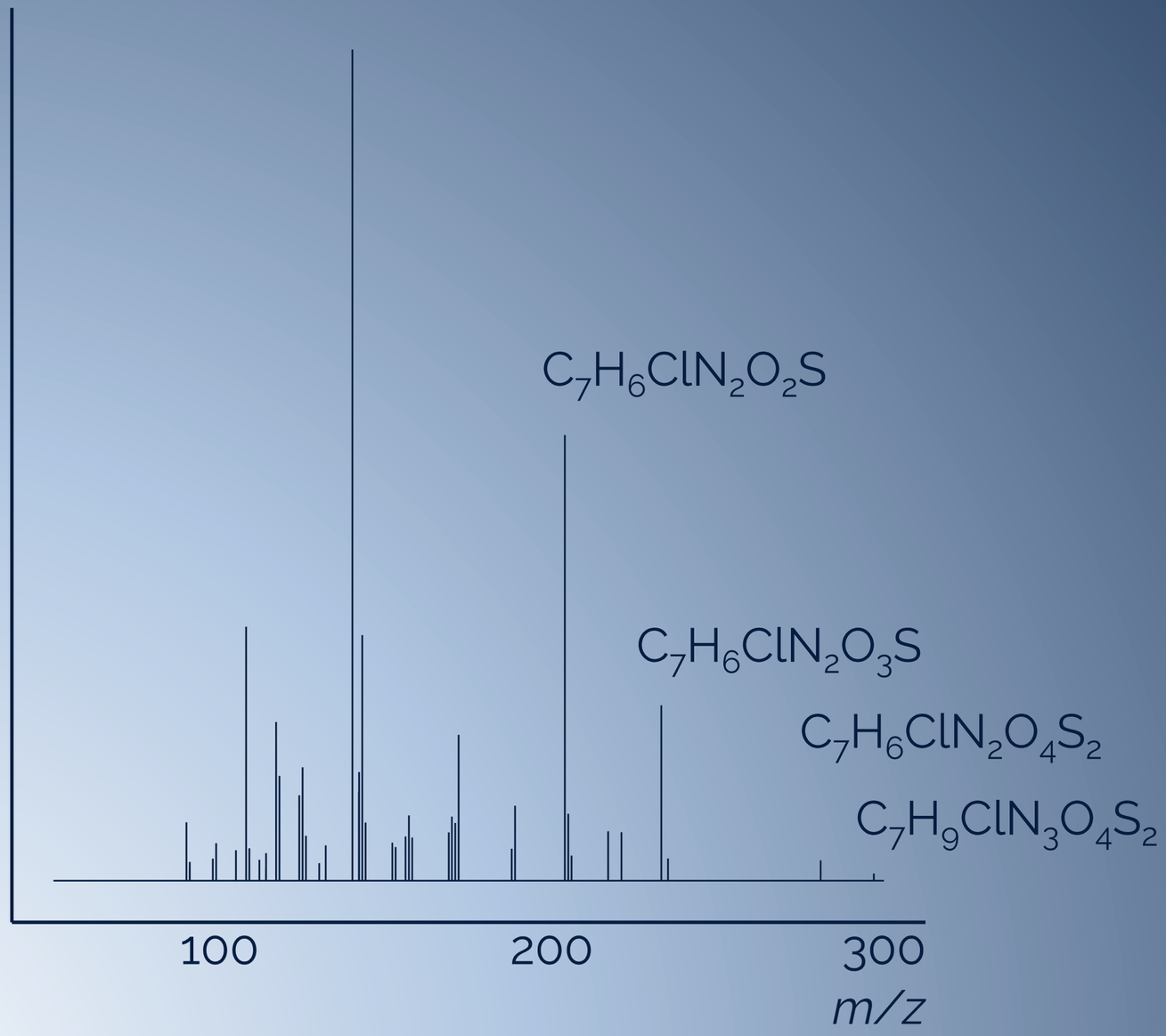


anneli kruve

anneli.kruve@su.se

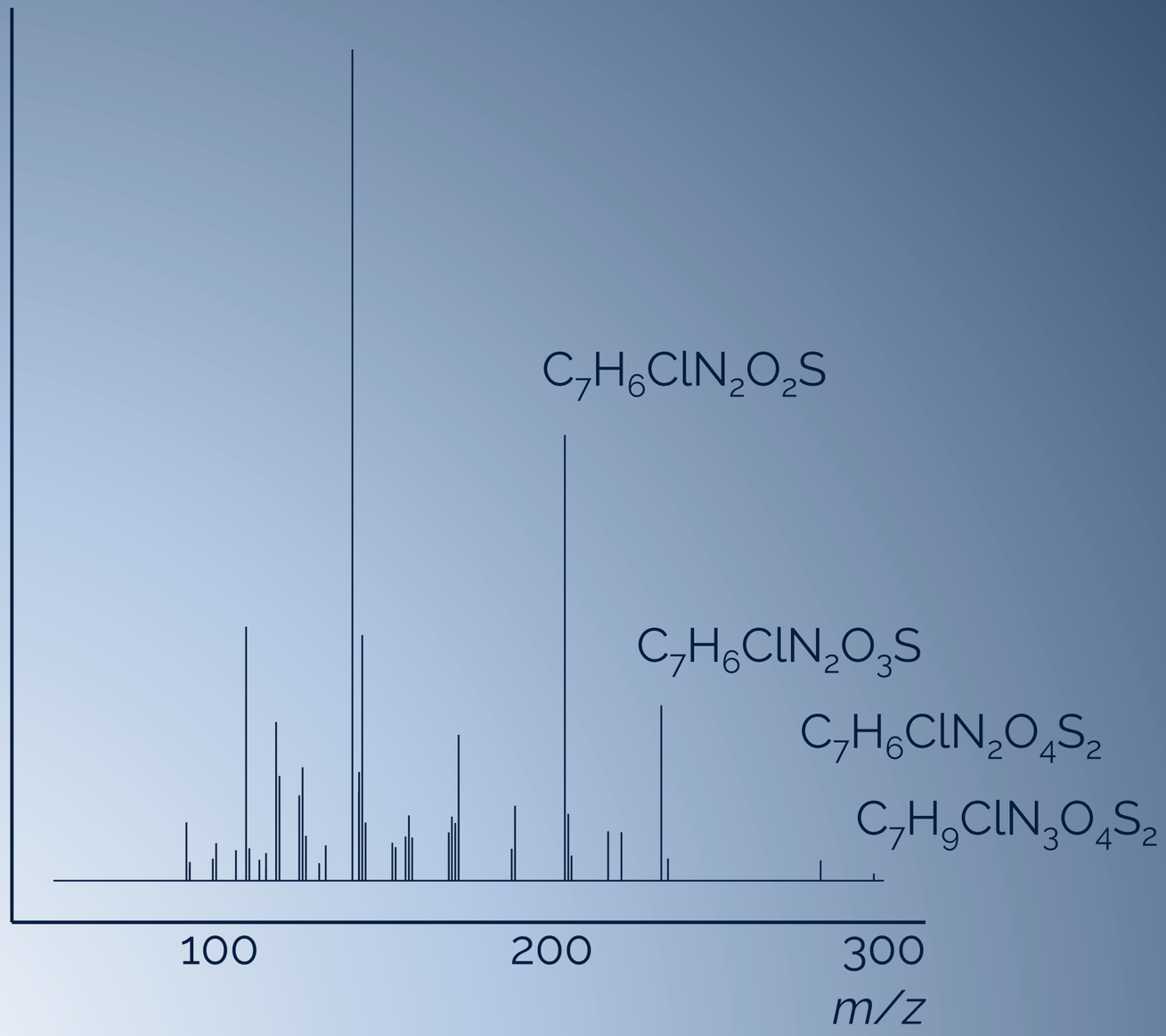
information available

in mass spectra



anneli kruve

anneli.kruve@su.se



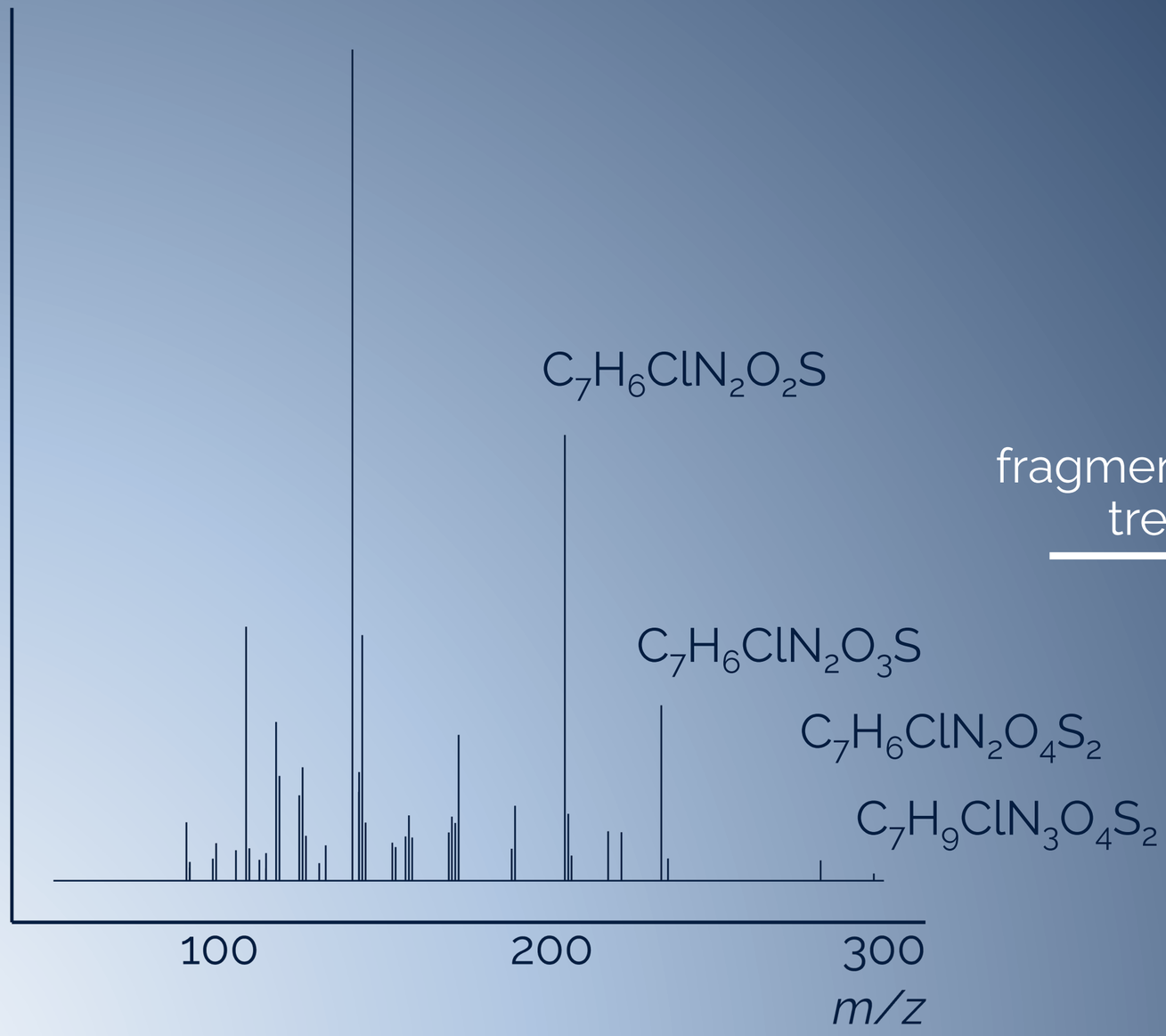
predict toxicity

for unknown chemicals

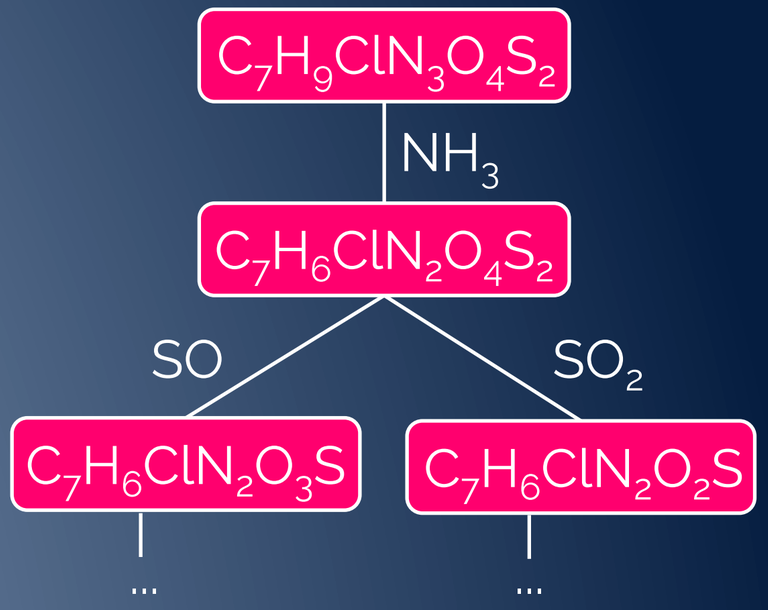
anneli kruve

anneli.kruve@su.se

Dührkop et al. SIRIUS + CSI:FingerID



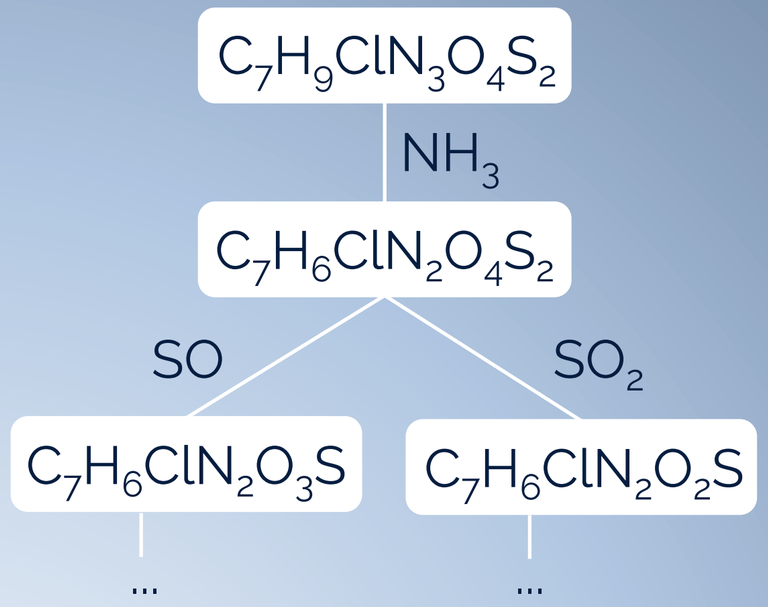
fragmentation tree



anneli kruve

anneli.kruve@su.se

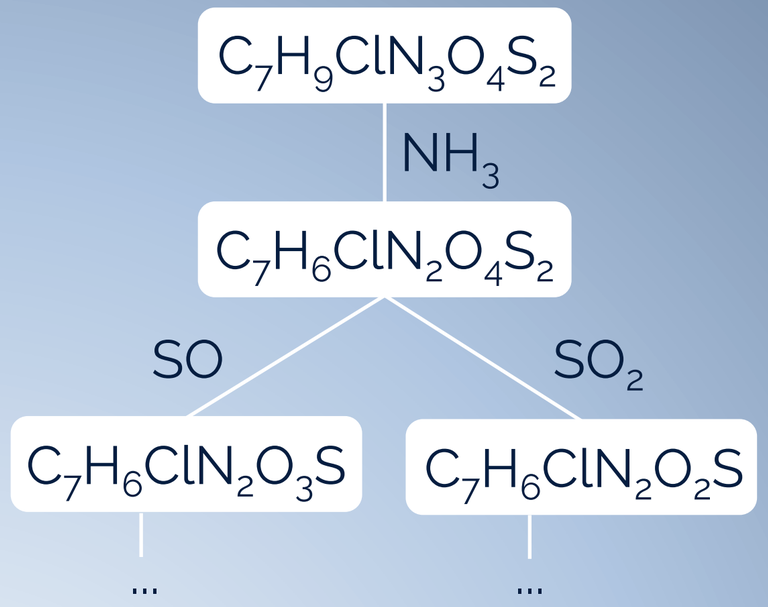
Dührkop et al. SIRIUS + CSI:FingerID



anneli kruve

anneli.kruve@su.se

Dührkop et al. SIRIUS + CSI:FingerID

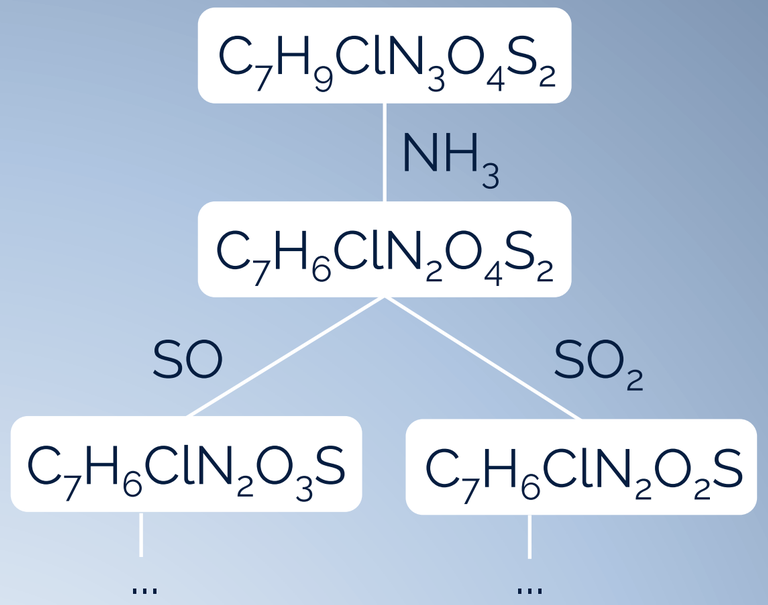


machine learning →

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

anneli kruve

anneli.kruve@su.se



machine learning →

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

machine learning →

p(AhR active)= 0.83

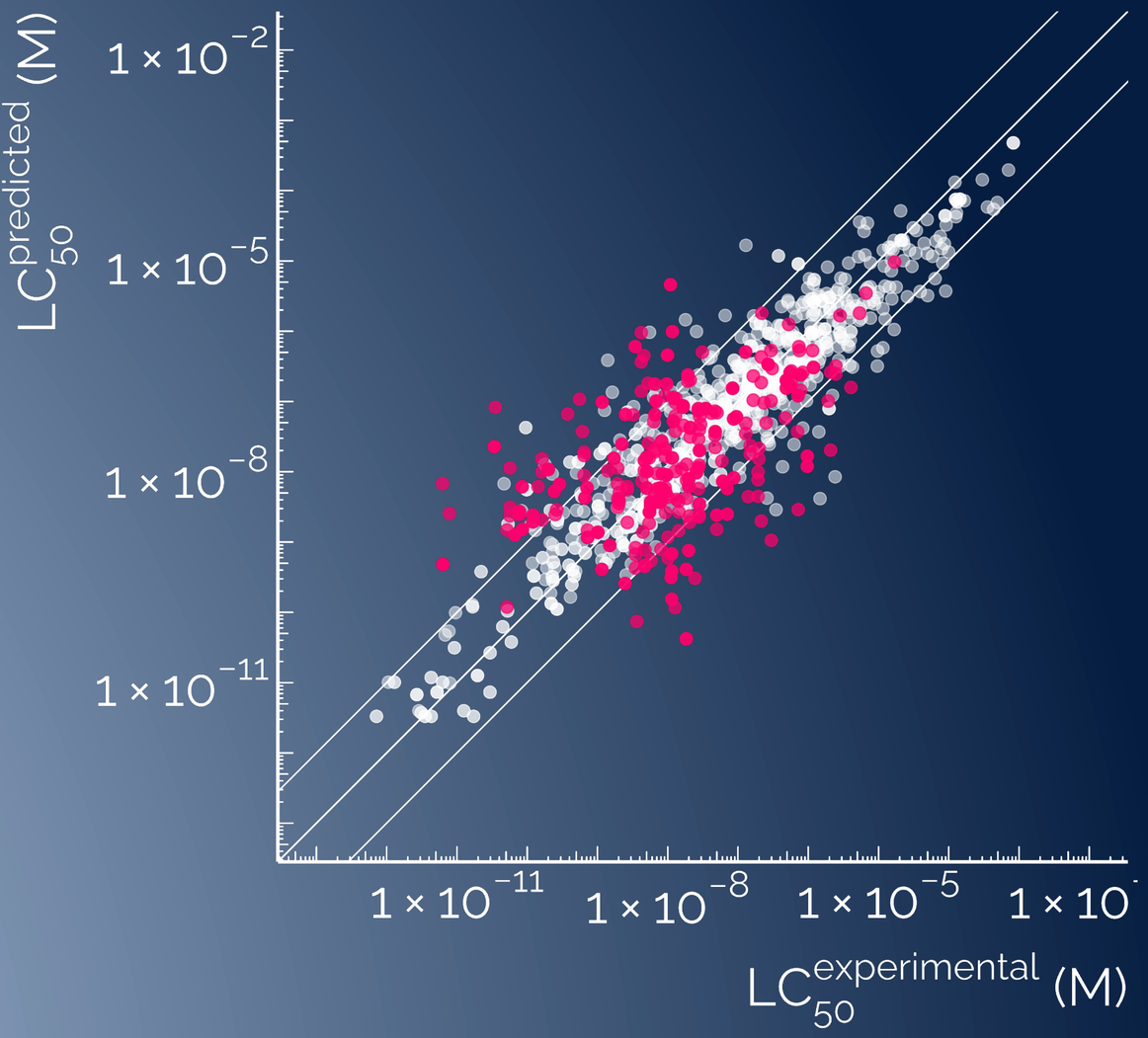
anneli kruve

anneli.kruve@su.se

LC_{50} predictions

Peets et al. ES&T 2022

fish LC_{50}



validation on MassBank

$RMSE_{model}$ 0.88 log(M)

$SD_{experimental}$ 0.44 log(M)

anneli kruve

endocrine disruption

Rahu et al. JCIM 2024
Tox21 endpoints

anneli kruve

anneli.kruve@su.se

endocrine disruption

true label	
active	non-active

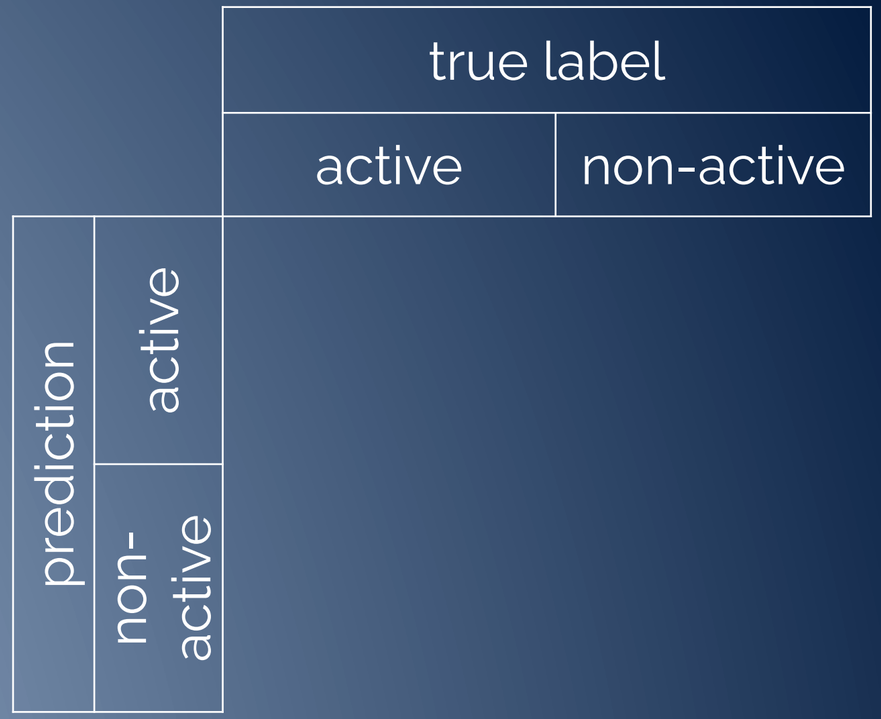
Rahu et al. JCIM 2024
Tox21 endpoints

anneli kruve

anneli.kruve@su.se

endocrine disruption

Rahu et al. JCIM 2024
Tox21 endpoints



anneli kruve

anneli.kruve@su.se

endocrine disruption

Rahu et al. JCIM 2024
Tox21 endpoints

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

anneli kruve

anneli.kruve@su.se

endocrine disruption

Rahu et al. JCIM 2024
Tox21 endpoints

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

anneli kruve

anneli.kruve@su.se

endocrine disruption

Rahu et al. JCIM 2024
Tox21 endpoints

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

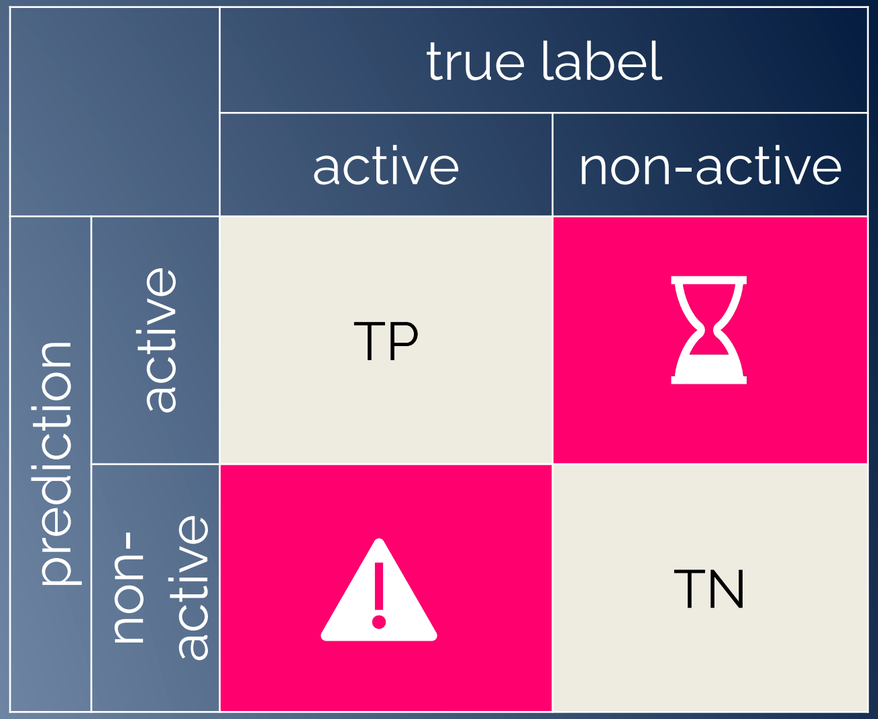
which is more dramatic:
FP or FN?

anneli kruve

anneli.kruve@su.se

endocrine disruption

Rahu et al. JCIM 2024
Tox21 endpoints



which is more dramatic:
FP or FN?

anneli kruve
anneli.kruve@su.se

endocrine disruption

Rahu et al. JCIM 2024
Tox21 endpoints

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

FPR @ TPR = 0.9

anneli kruve

anneli.kruve@su.se

endocrine disruption

Rahu et al. JCIM 2024
Tox21 endpoints

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

MassBank & MoNA
748 chemicals

anneli kruve

anneli.kruve@su.se

case study: interlaboratory comparison

water analysis

thousands of chemicals detected in the environment

BioCell ANALYTICA

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

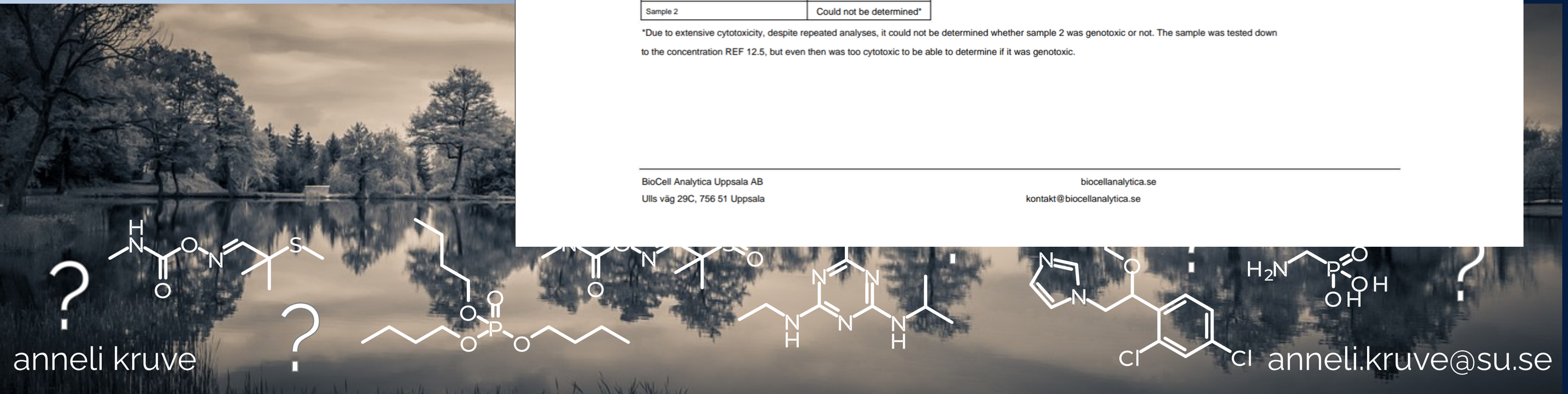
	Nrf2 activity	AntiÿAR activity	AR activity	ER activity	AhR activity
	µg/L (tBHQ equivalents)	ng/L (OHF equivalents)	ng/L (DHT equivalents)	pg/L (E2 equivalents)	ng/L (TCDD equivalents)
Reference to sample 1	<LOD	<LOD	<LOD	<LOD	<LOD
Sample 1	<LOD	<LOD	79300	784	0.0814
Reference to sample 2	21.1	73.6	<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



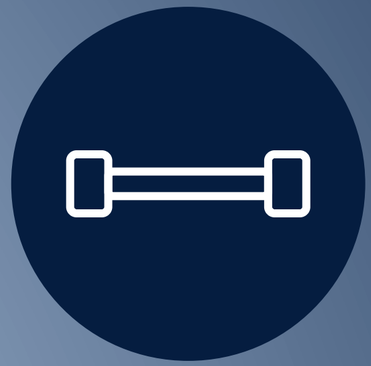
anneli kruve
anneli.kruve@su.se

interlaboratory comparison

Sandberg, Rahu, in preparation

SAMPLES

spiked water samples



ANALYSIS

HRMS, etc. characterization

DATA PROCESSING

AhR activity

anneli kruve

anneli.kruve@su.se

results

4700

LC/HRMS features detected

238

features with MS^2 spectra

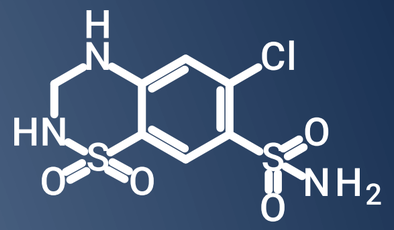
55

features predicted active

anneli kruve

anneli.kruve@su.se

AhR active



hydrochlorothiazide

anneli kruve

anneli.kruve@su.se

alternative approaches

Kreutzer et al. in preparation
Tox21 endpoints

anneli kruve

anneli.kruve@su.se

molecular networks

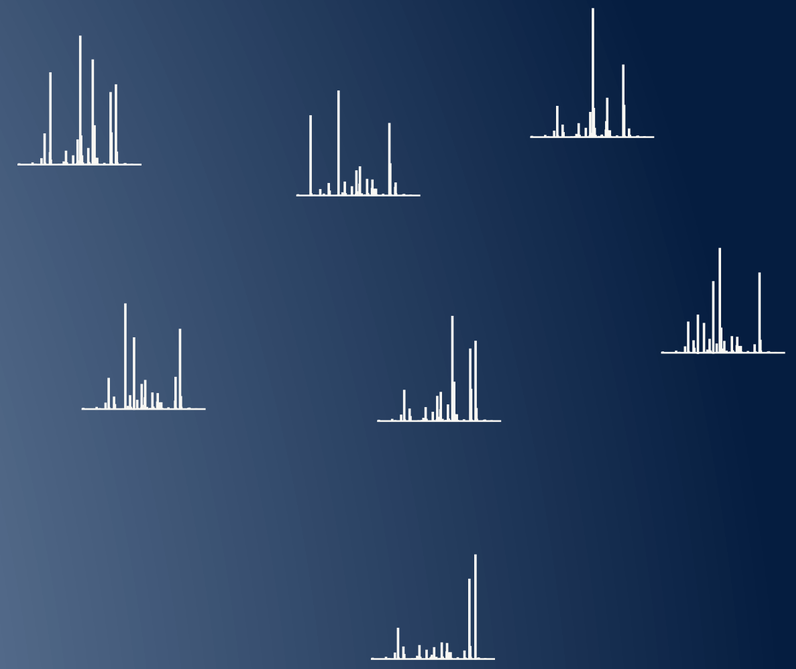
Kreutzer et al. in preparation
Tox21 endpoints

anneli kruve

anneli.kruve@su.se

molecular networks

Kreutzer et al. in preparation
Tox21 endpoints

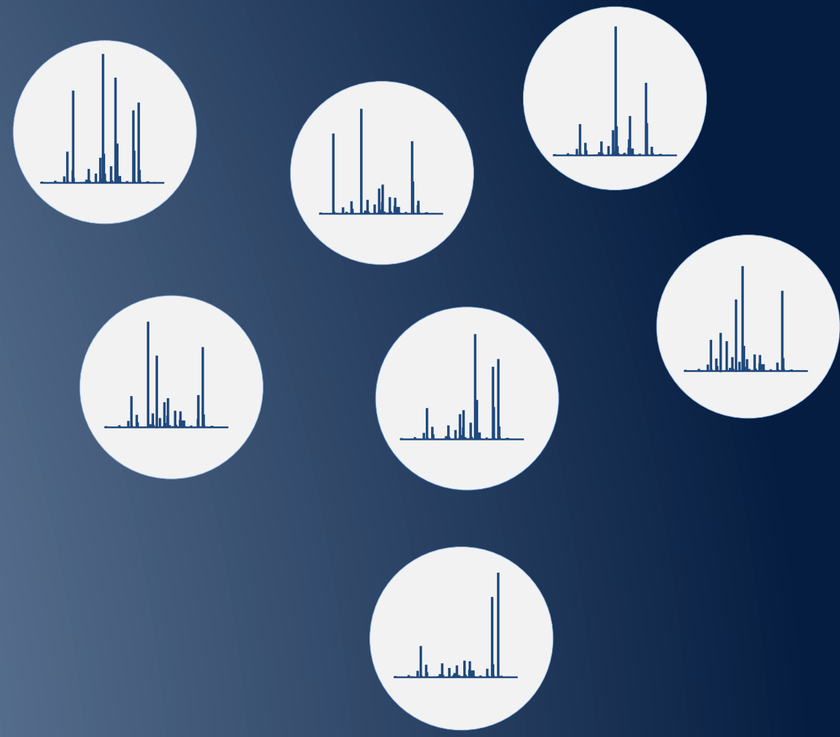


anneli kruve

anneli.kruve@su.se

molecular networks

Kreutzer et al. in preparation
Tox21 endpoints



anneli kruve

anneli.kruve@su.se

molecular networks

Kreutzer et al. in preparation
Tox21 endpoints

anneli kruve

anneli.kruve@su.se

molecular networks

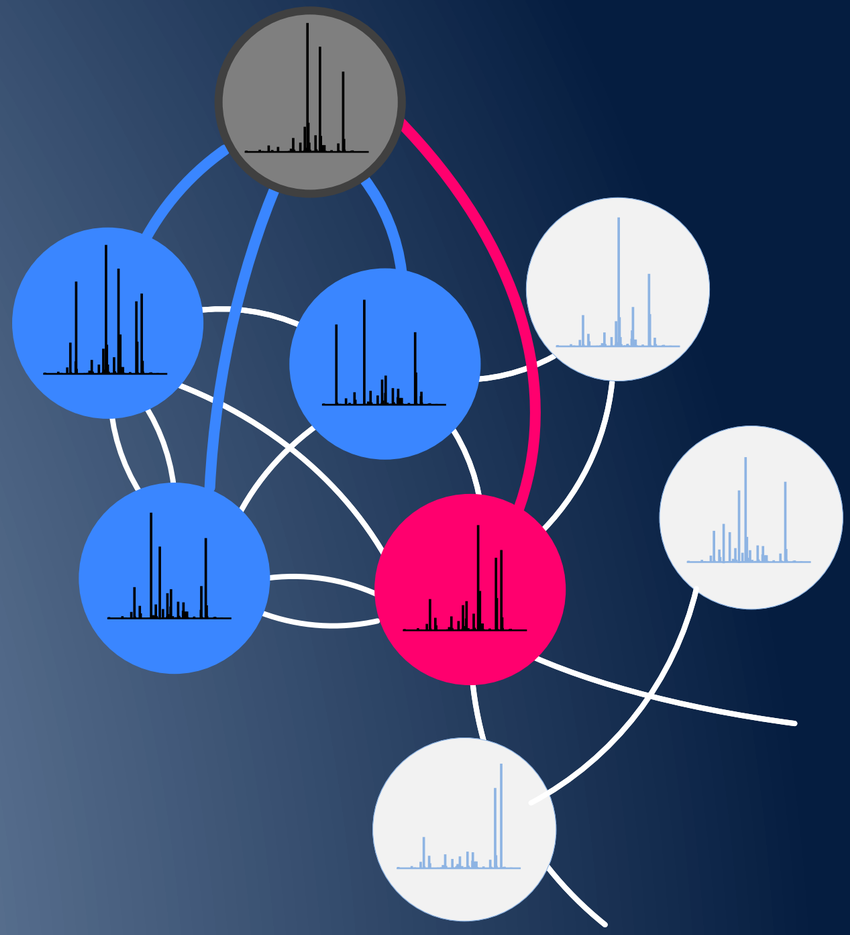
Kreutzer et al. in preparation
Tox21 endpoints

anneli kruve

anneli.kruve@su.se

molecular networks

Kreutzer et al. in preparation
Tox21 endpoints

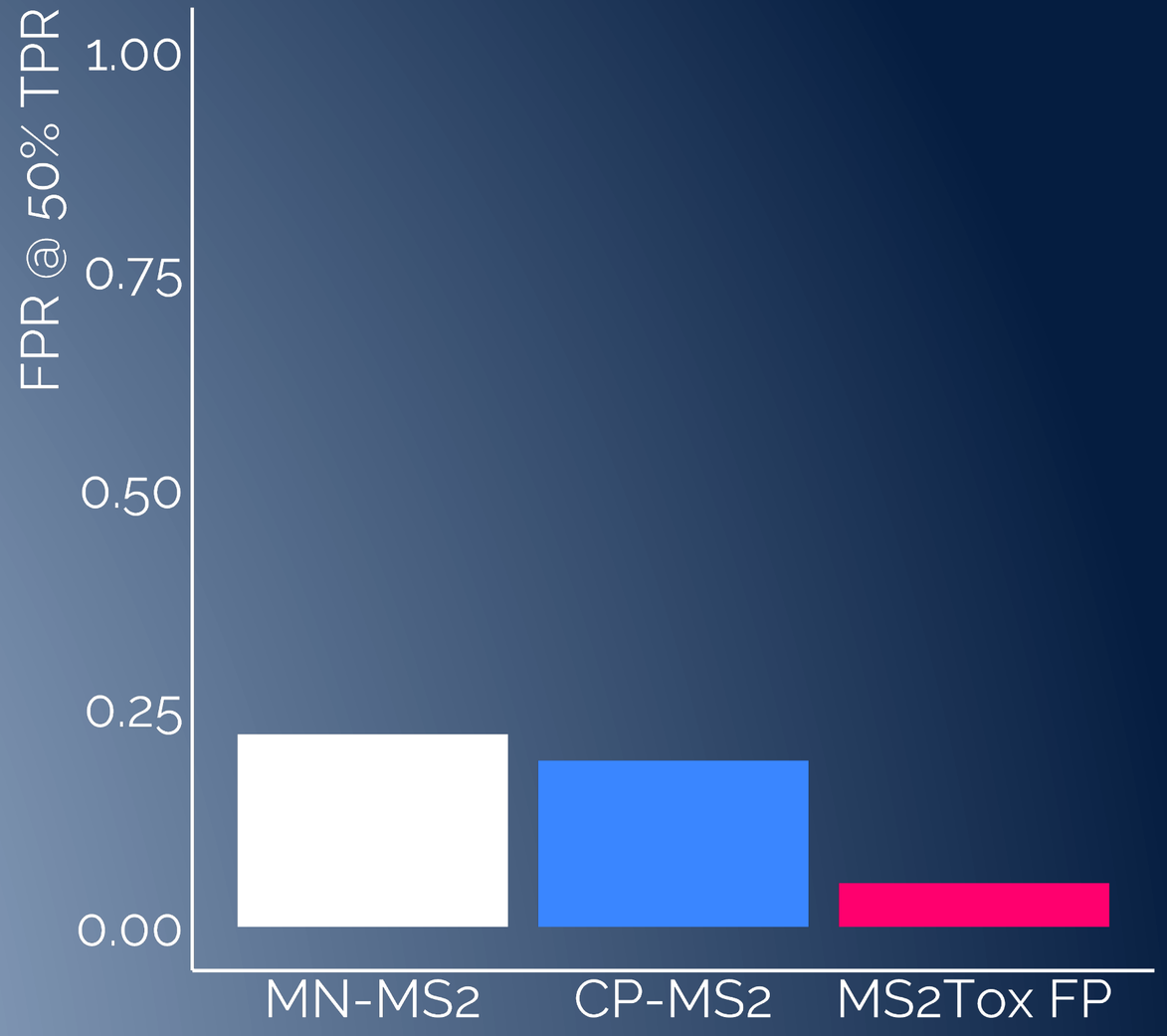


anneli kruve

anneli.kruve@su.se

AhR activity predictions

Kreutzer et al. in preparation
Tox21 endpoints

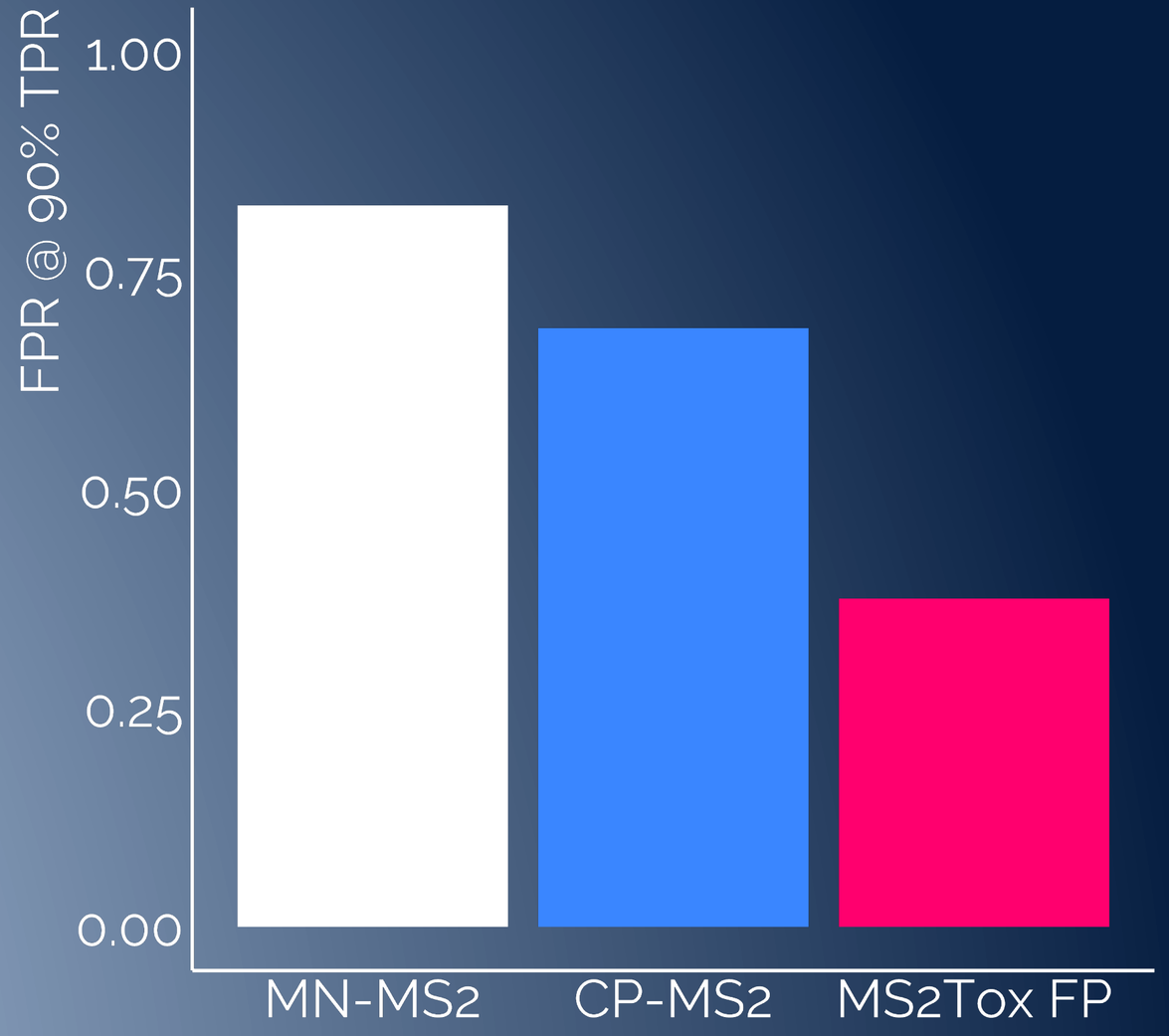


anneli kruve

anneli.kruve@su.se

AhR activity predictions

Kreutzer et al. in preparation
Tox21 endpoints



anneli kruve
anneli.kruve@su.se

application:
wastewater analyses

LC/HRMS features

986 features

LC/HRMS features

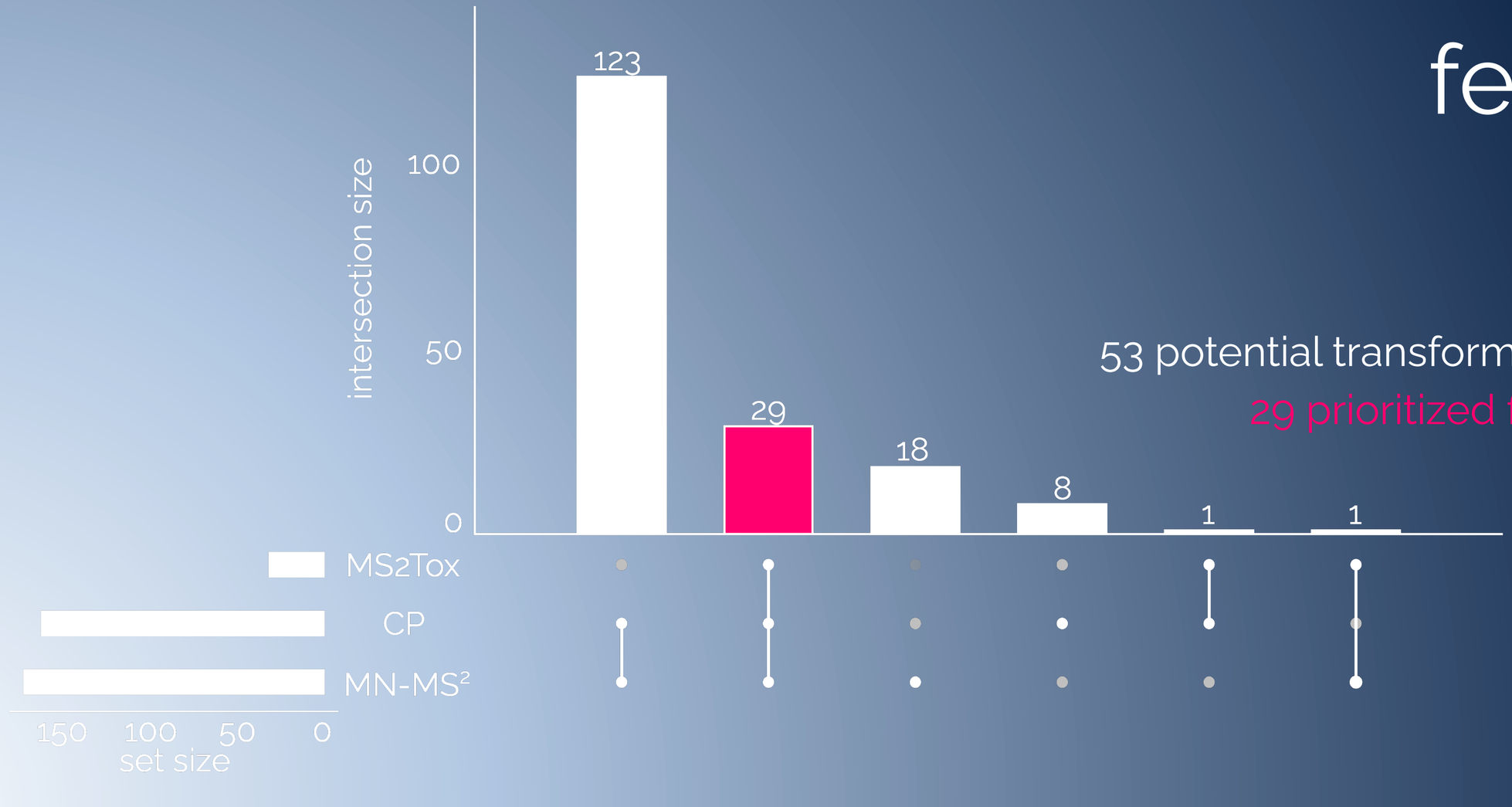
986 features
136 persistent

LC/HRMS features

986 features
136 persistent
53 potential transformation products

LC/HRMS features

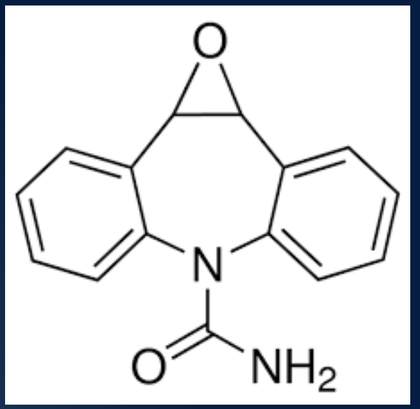
986 features
136 persistent
53 potential transformation products
29 prioritized for AhR activity

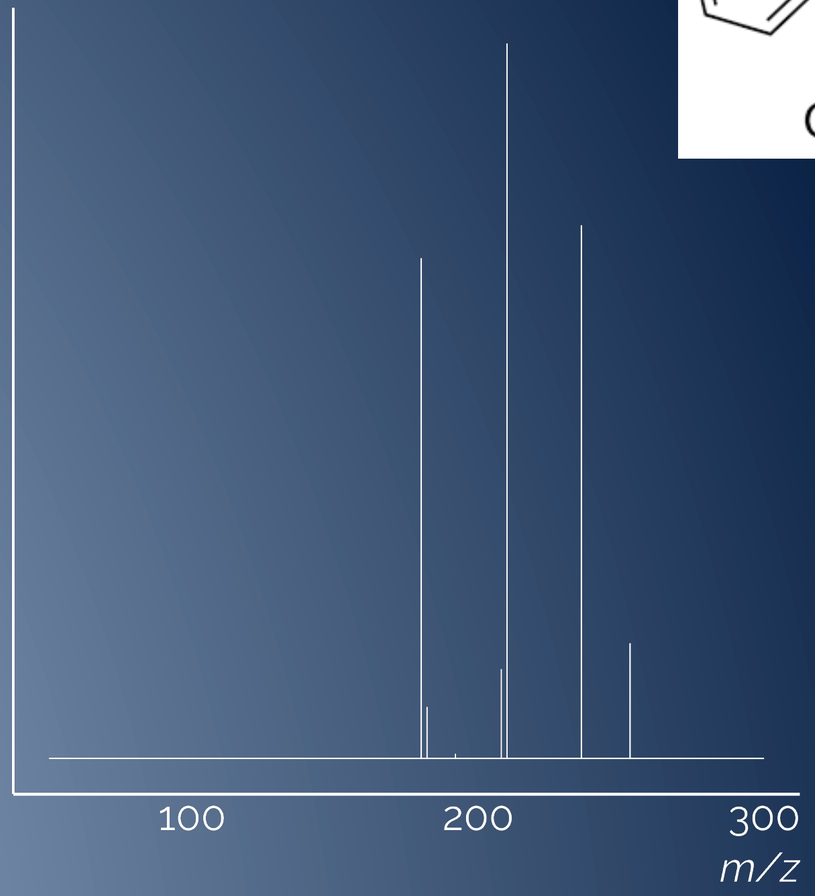


carbamazepine-10,11-epoxide

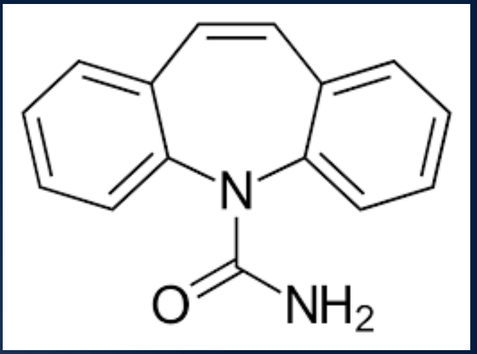
RT 8.01 min *m/z* = 252.090

confirmed with analytical standard





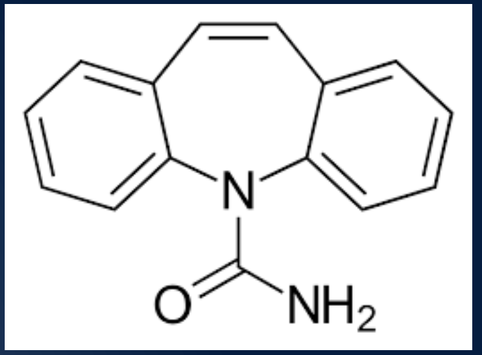
carbamazepine

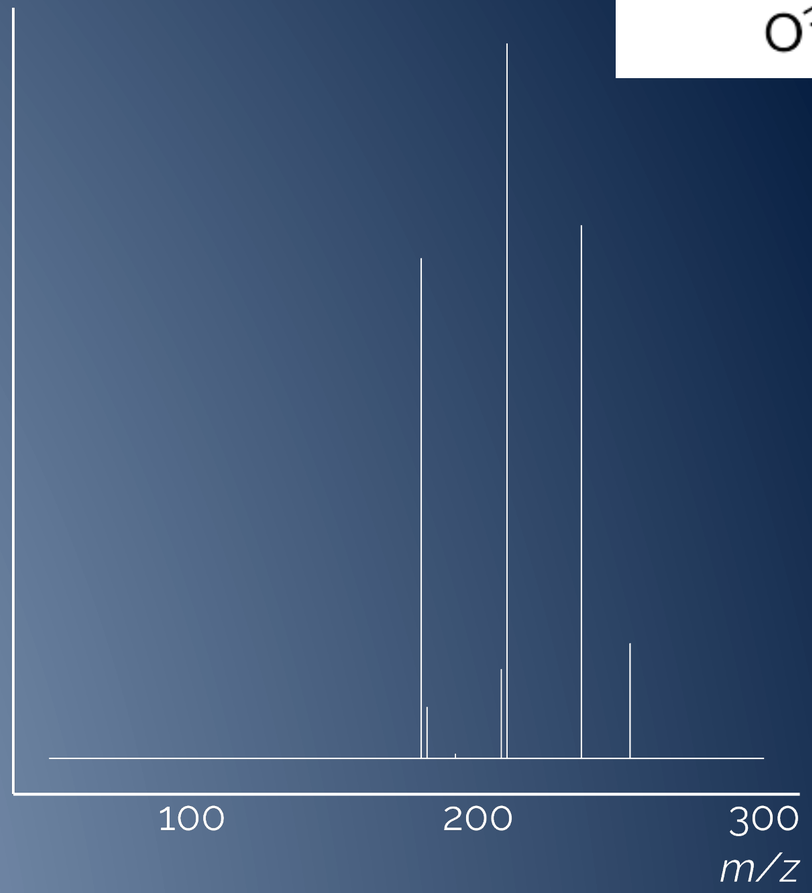


carbamazepine

RT 9.42 min *m/z* = 236.095

confirmed with analytical standard



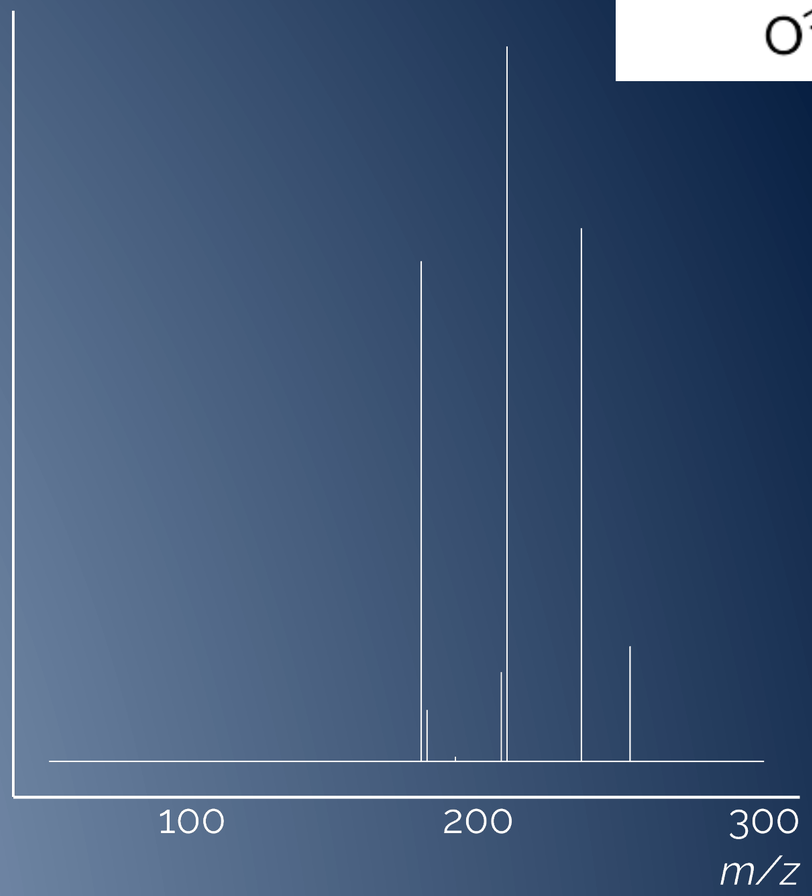


carbamazepine

RT 9.42 min *m/z* = 236.095

confirmed with analytical standard

prioritized by MS2Tox & MN-MS2

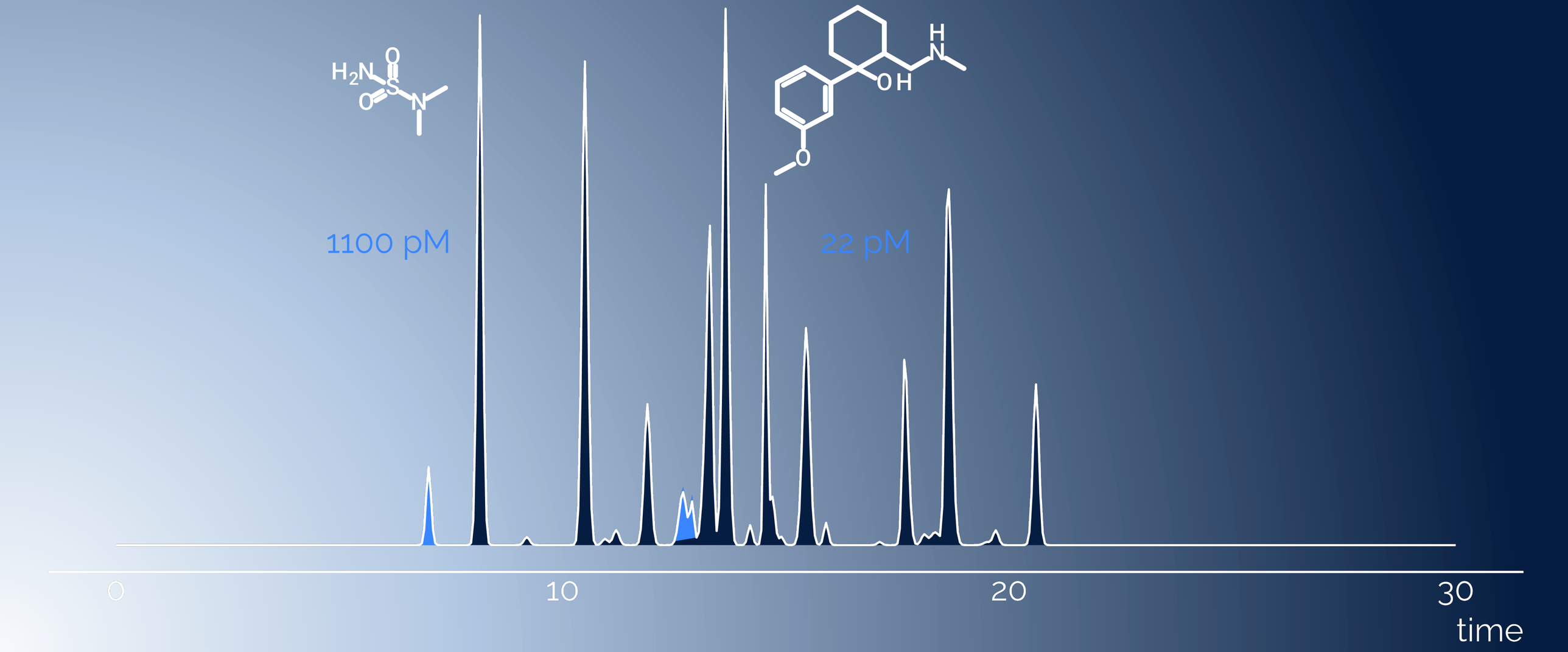


quantification

of detected chemicals

quantification in ESI/HRMS

Malm et al. Molecules 2021



electrospray

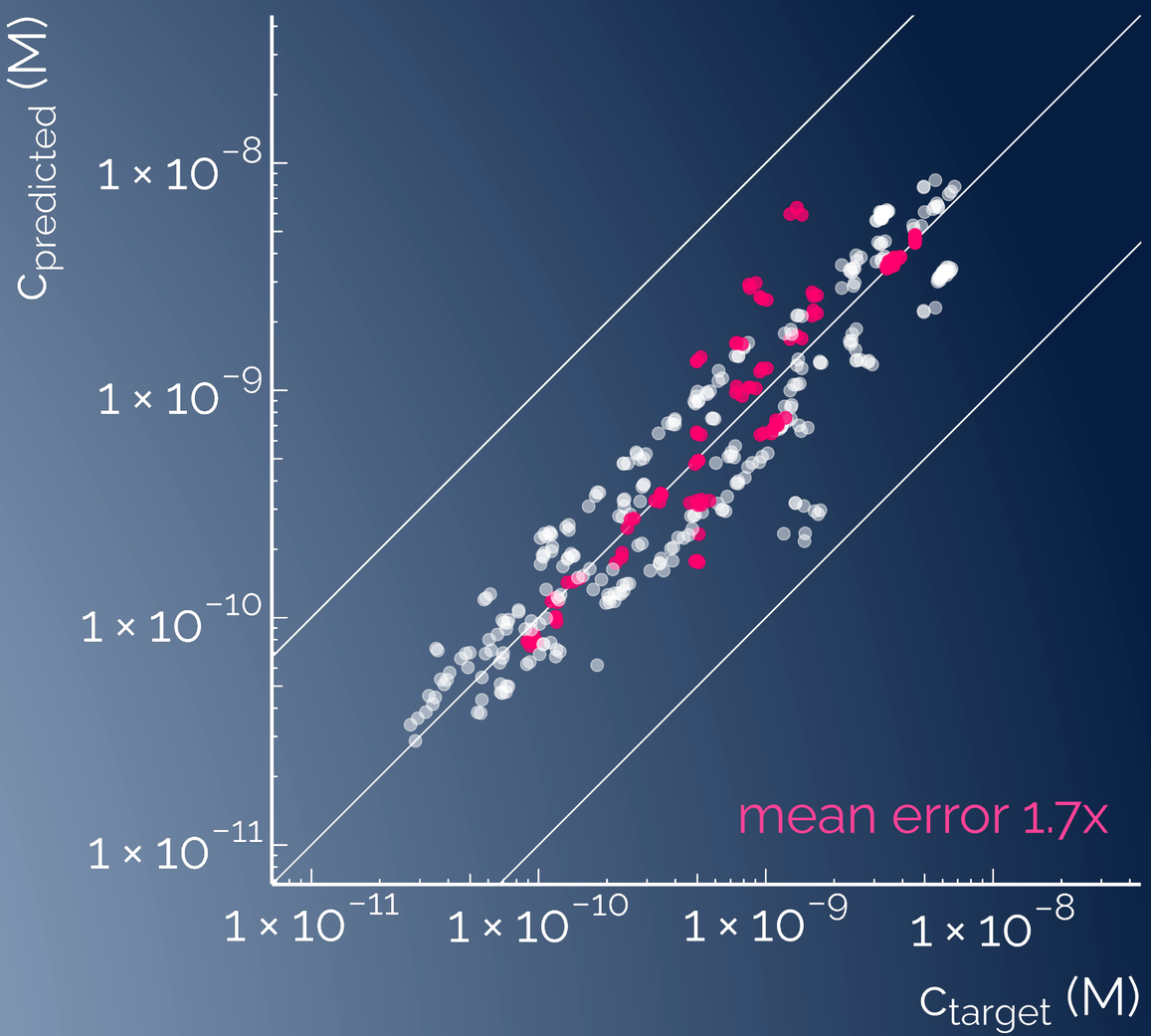
anneli kruve

anneli.kruve@su.se

quantification

from structure

Been et al. Water Research 2021



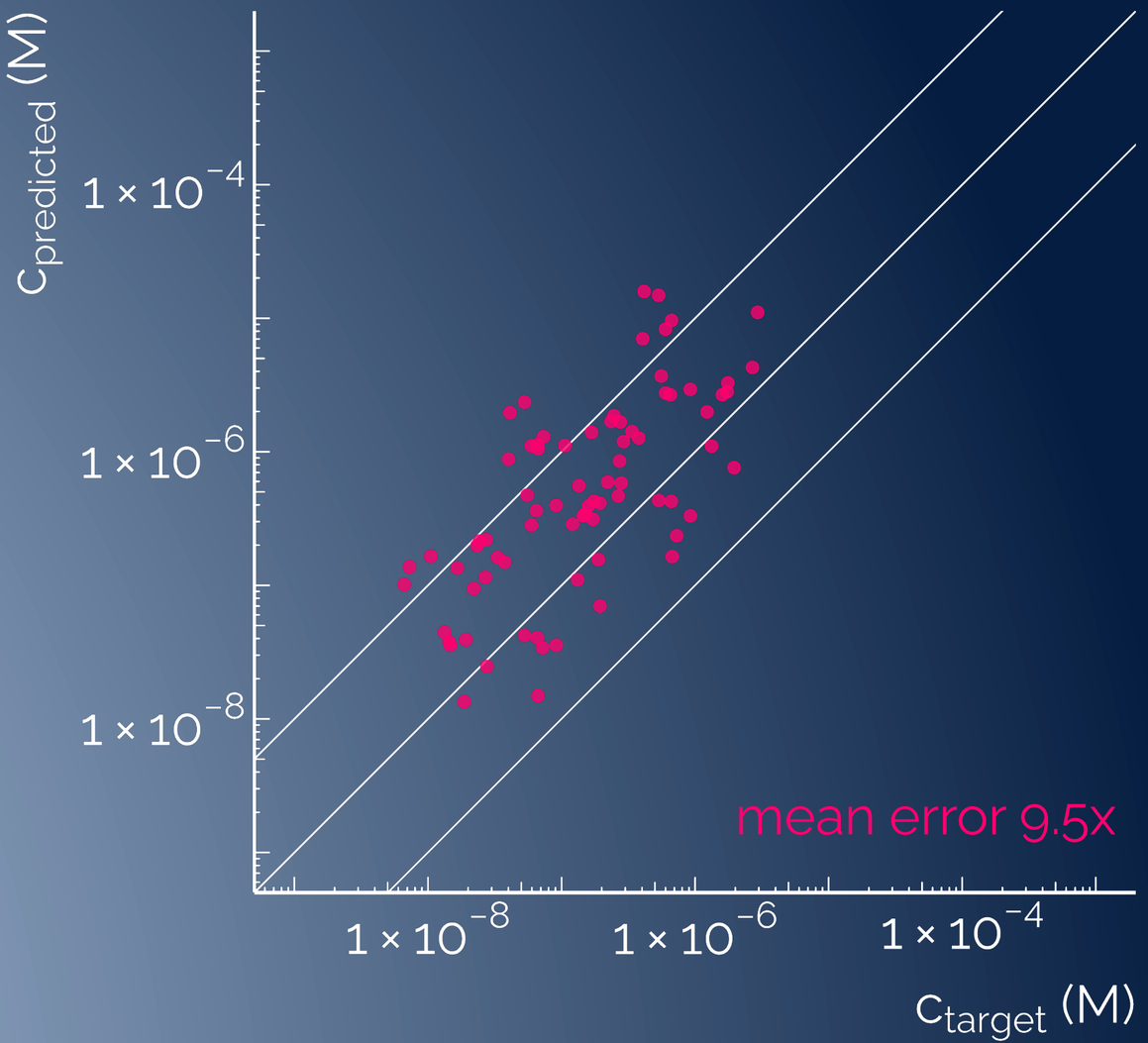
anneli kruve

anneli.kruve@su.se

quantification

from MS² spectra

Sepman et al. Anal Chem 2023



anneli kruve

anneli.kruve@su.se

risk

based prioritization

prioritization of chemicals

TOXICITY

ecotoxicity and endocrine disruptors

CONCENTRATION

exposure to potentially toxic chemicals

RISK

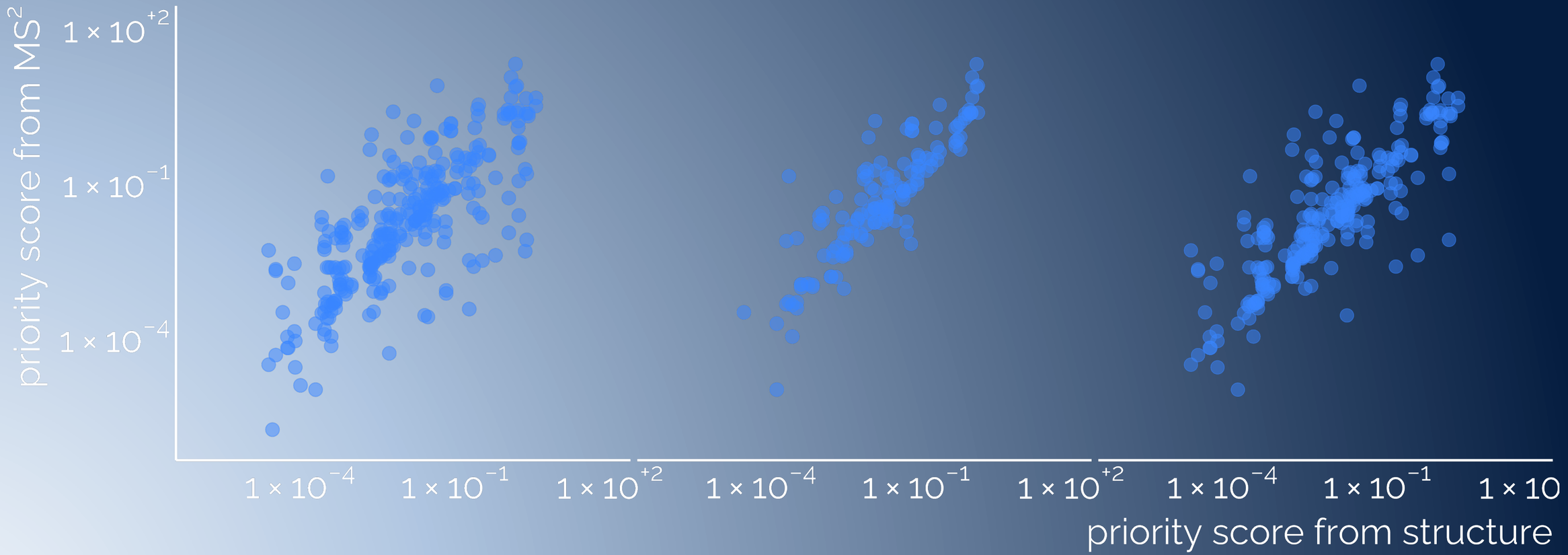
$$\text{PriorityScore} = \frac{C_{\text{predicted}}}{AC_{50}^{\text{5th percentile}}}$$

anneli kruve

anneli.kruve@su.se

wastewater analysis

Sepman, in preparation



anneli kruve

anneli.kruve@su.se

how to ...

PRIORITIZE

risk

IDENTIFY

structure

anneli kruve

anneli.kruve@su.se

how to ...

PRIORITIZE

risk

IDENTIFY

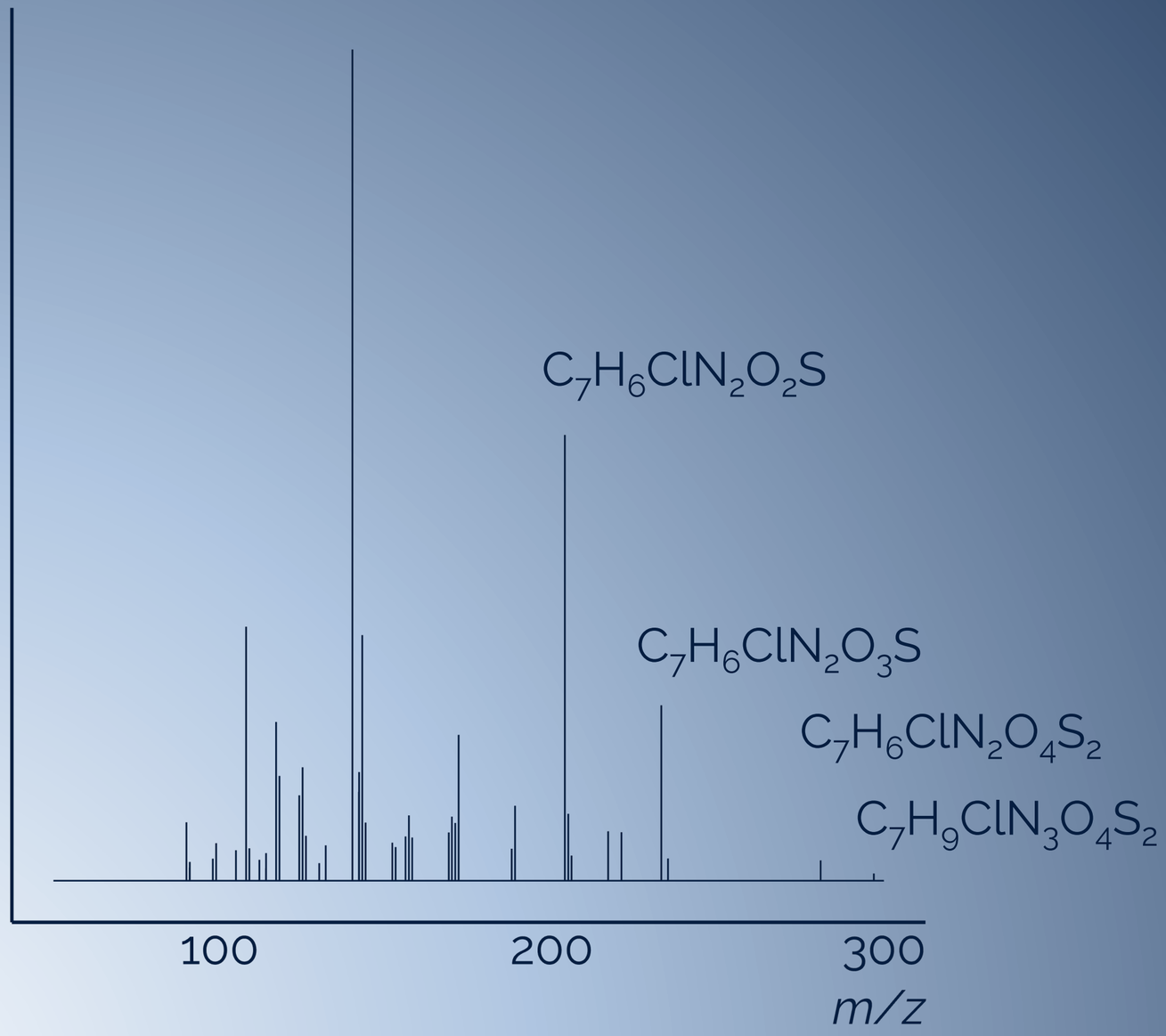
structure

anneli kruve

anneli.kruve@su.se

structural identification

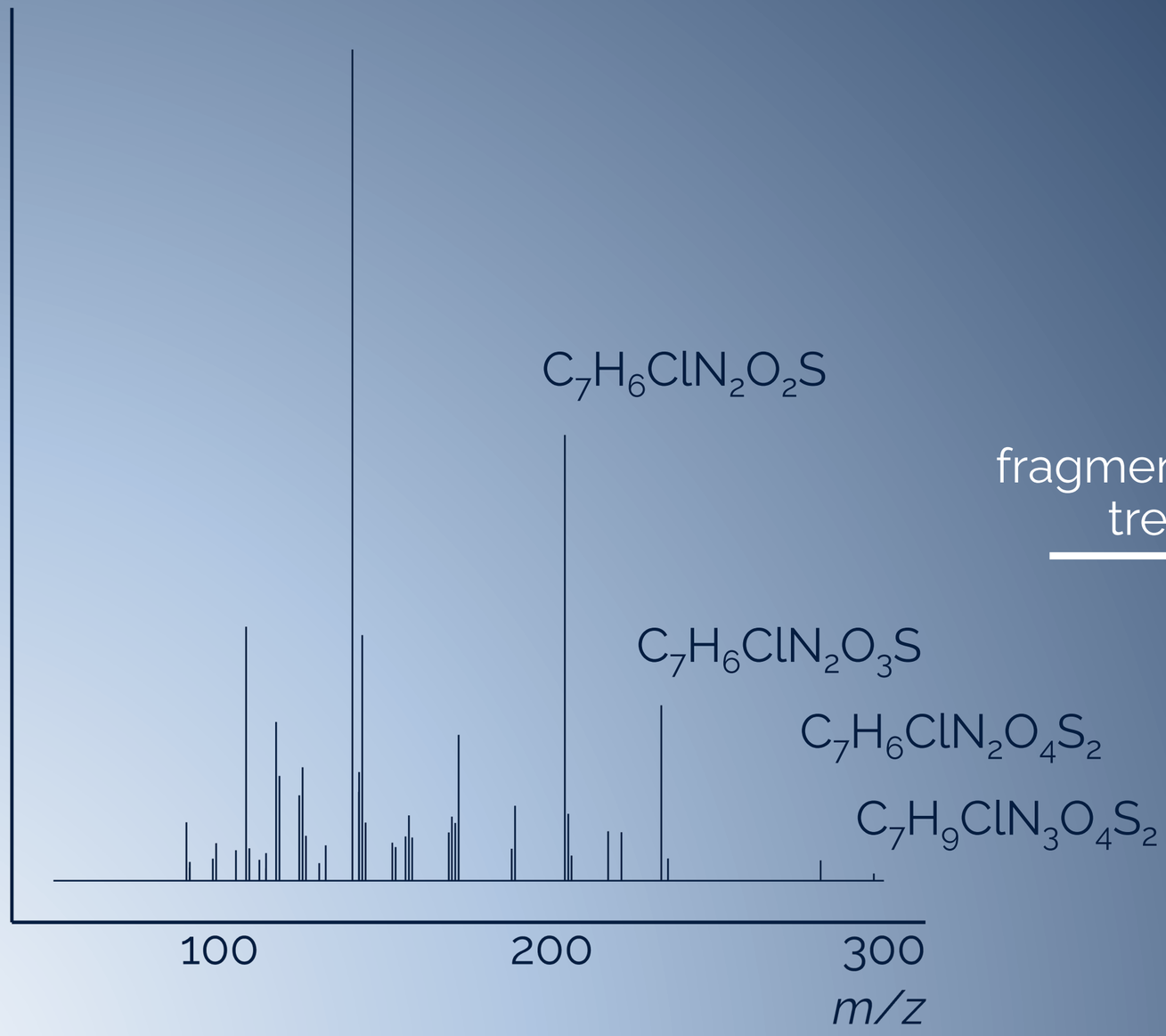
from MS^2 spectra



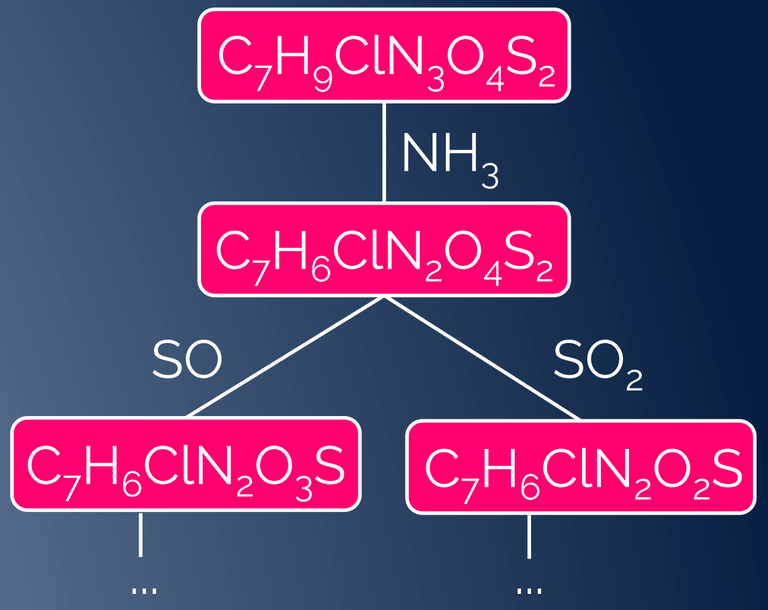
anneli kruve

anneli.kruve@su.se

Dührkop et al. SIRIUS + CSI:FingerID



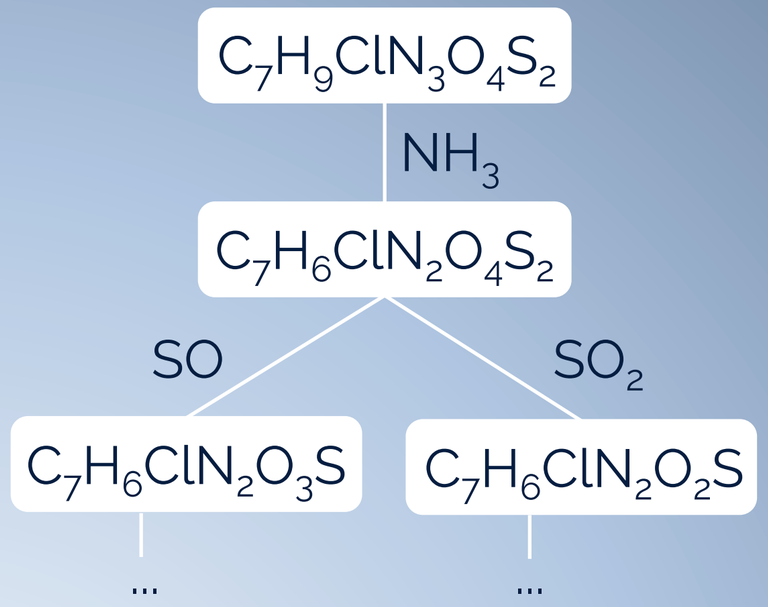
fragmentation tree



anneli kruve

anneli.kruve@su.se

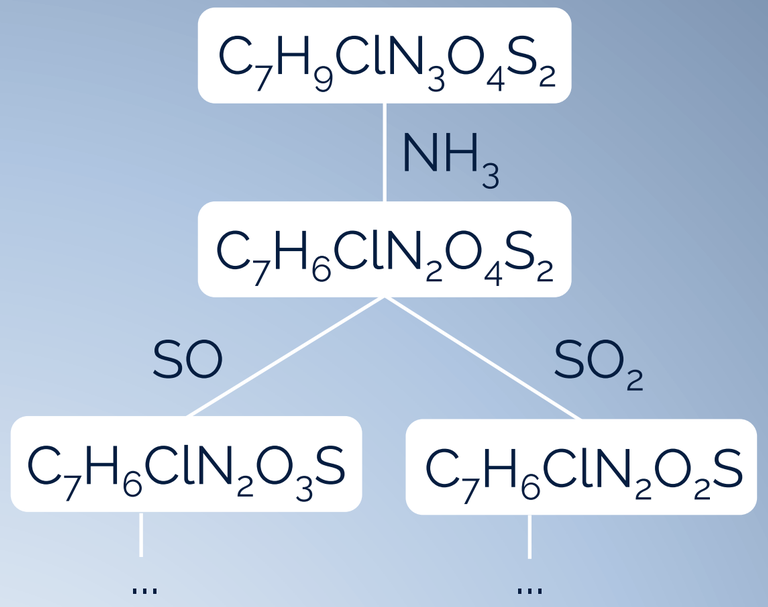
Dührkop et al. SIRIUS + CSI:FingerID



anneli kruve

anneli.kruve@su.se

Dührkop et al. SIRIUS + CSI:FingerID



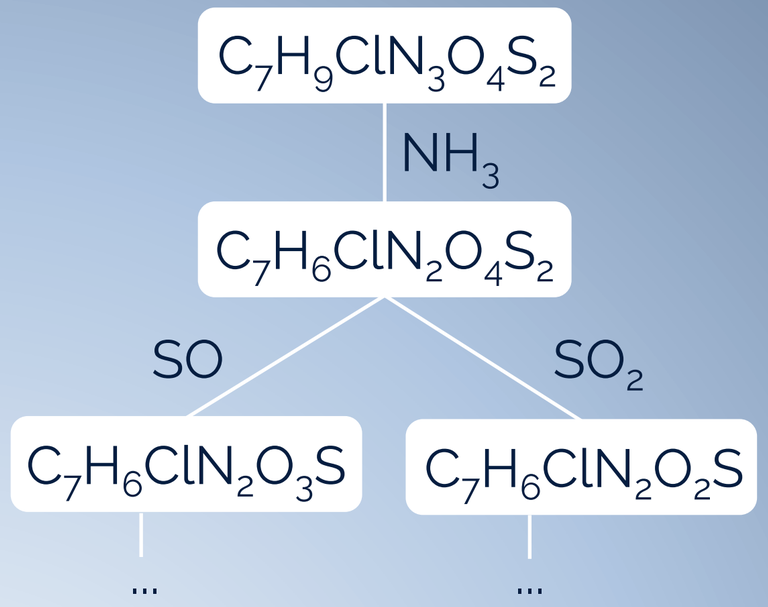
machine learning →

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

anneli kruve

anneli.kruve@su.se

Dührkop et al. SIRIUS + CSI:FingerID



machine learning →

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

structure →

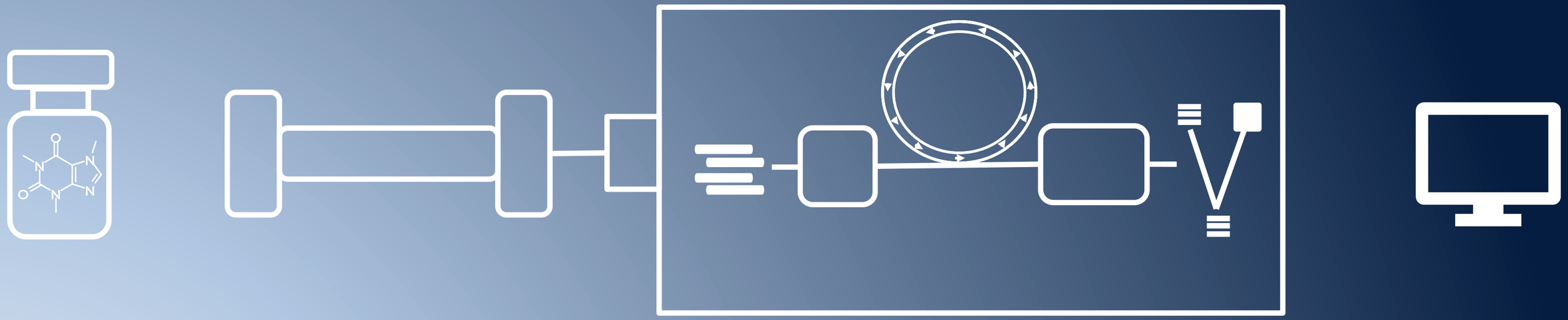
database matching

anneli kruve

anneli.kruve@su.se

orthogonal separation

Akhlaqi et al. Anal Bioanal Chem 2023



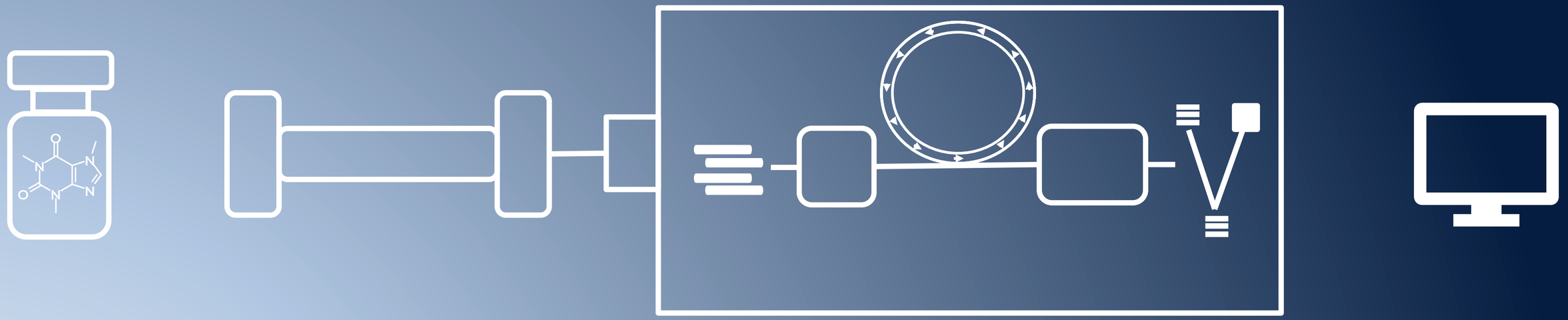
chromatography ion mobility MS^2

anneli kruve

anneli.kruve@su.se

orthogonal separation

Akhlaqi et al. Anal Bioanal Chem 2023



14
isomeric
chemicals

C_{18} RP

Cyclic IMS
&
MS^2 with ToF

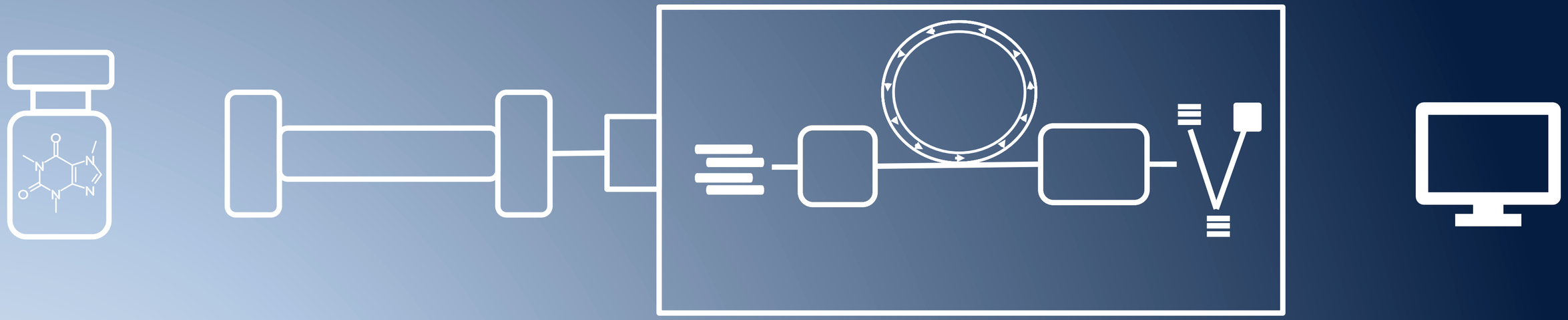
SIRIUS+
CSI:FingerID
&
CFM-ID

anneli kruve

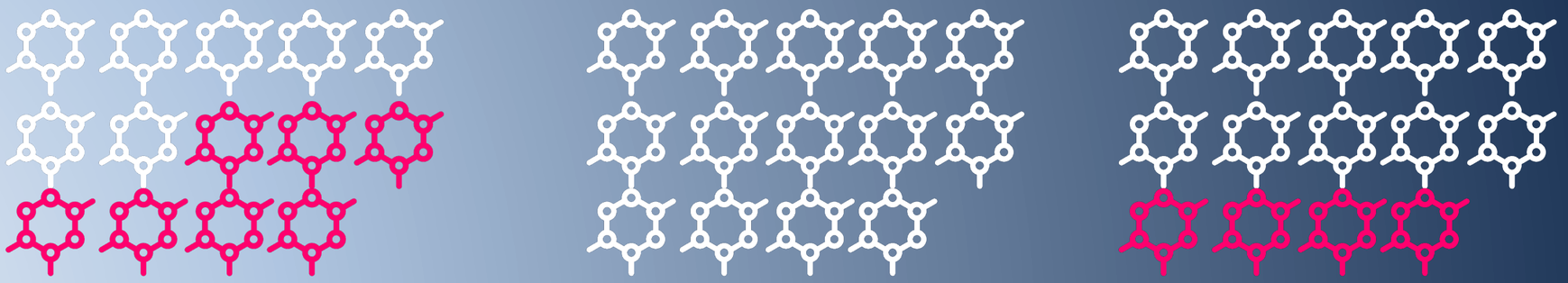
anneli.kruve@su.se

orthogonal separation

Akhlaqi et al. Anal Bioanal Chem 2023



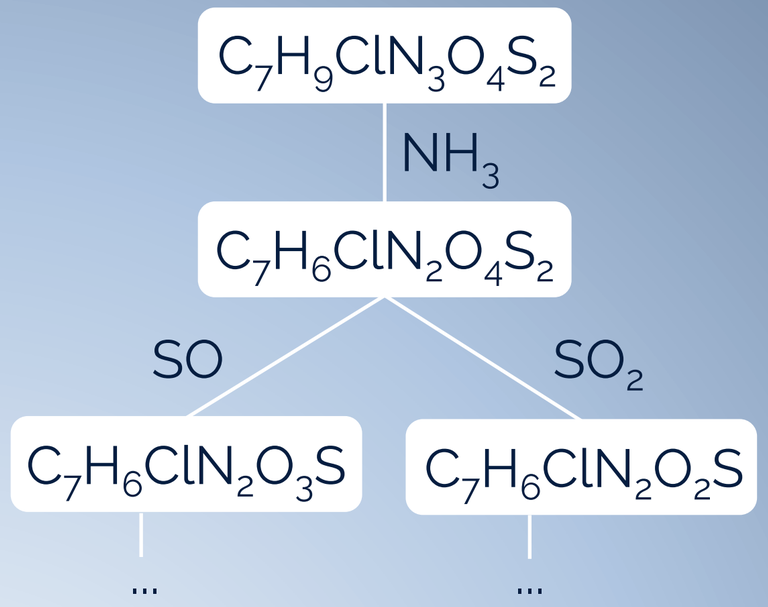
t_R *CCS* MS^2



anneli kruve

anneli.kruve@su.se

Dührkop et al. SIRIUS + CSI:FingerID
Stravs et al. MS Novelist



machine learning →

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

structure →

database matching

De Novo

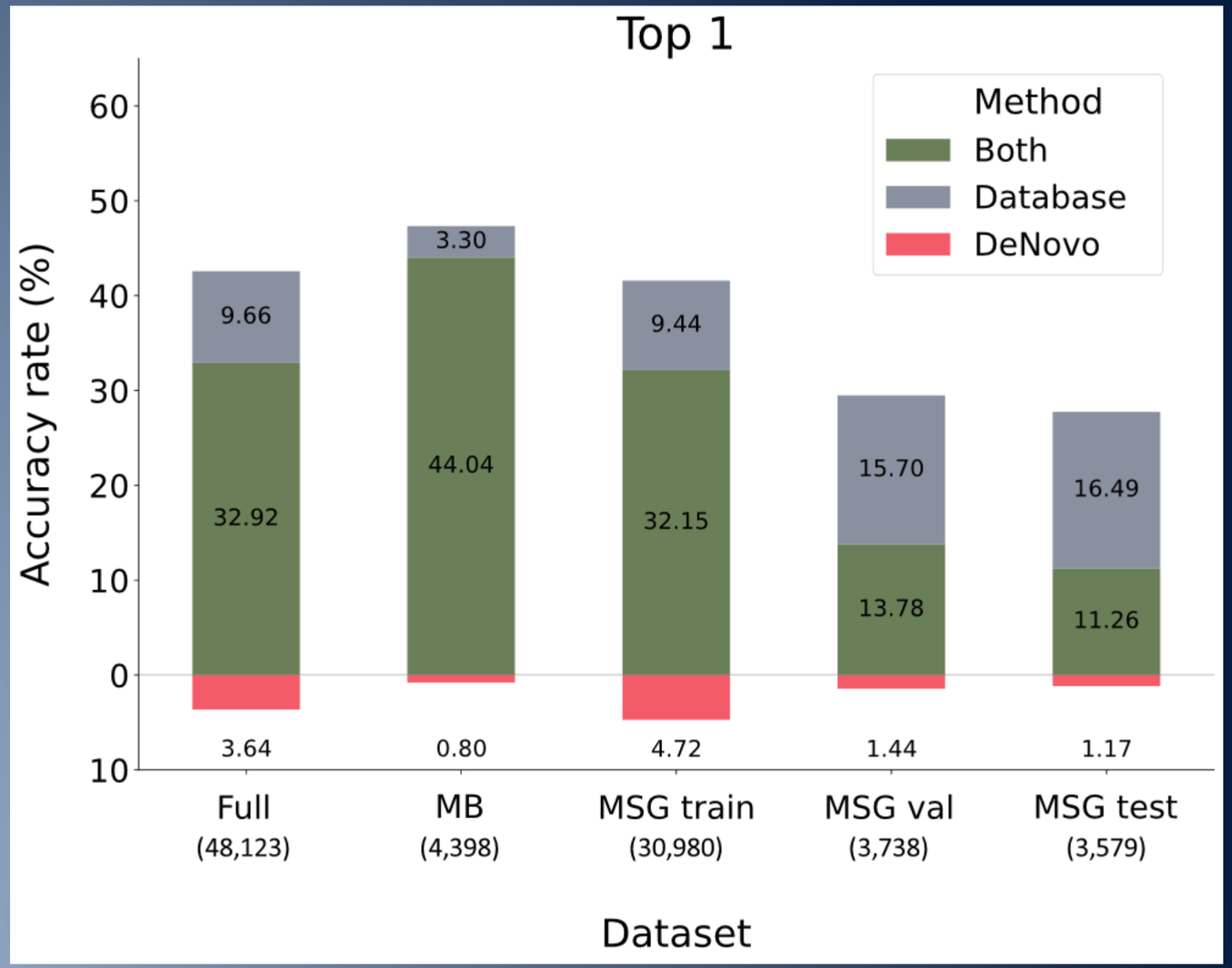
anneli kruve

anneli.kruve@su.se

structural identification

from MS^2 spectra

Hättestrand et al., in preparation



in training set 36.0%

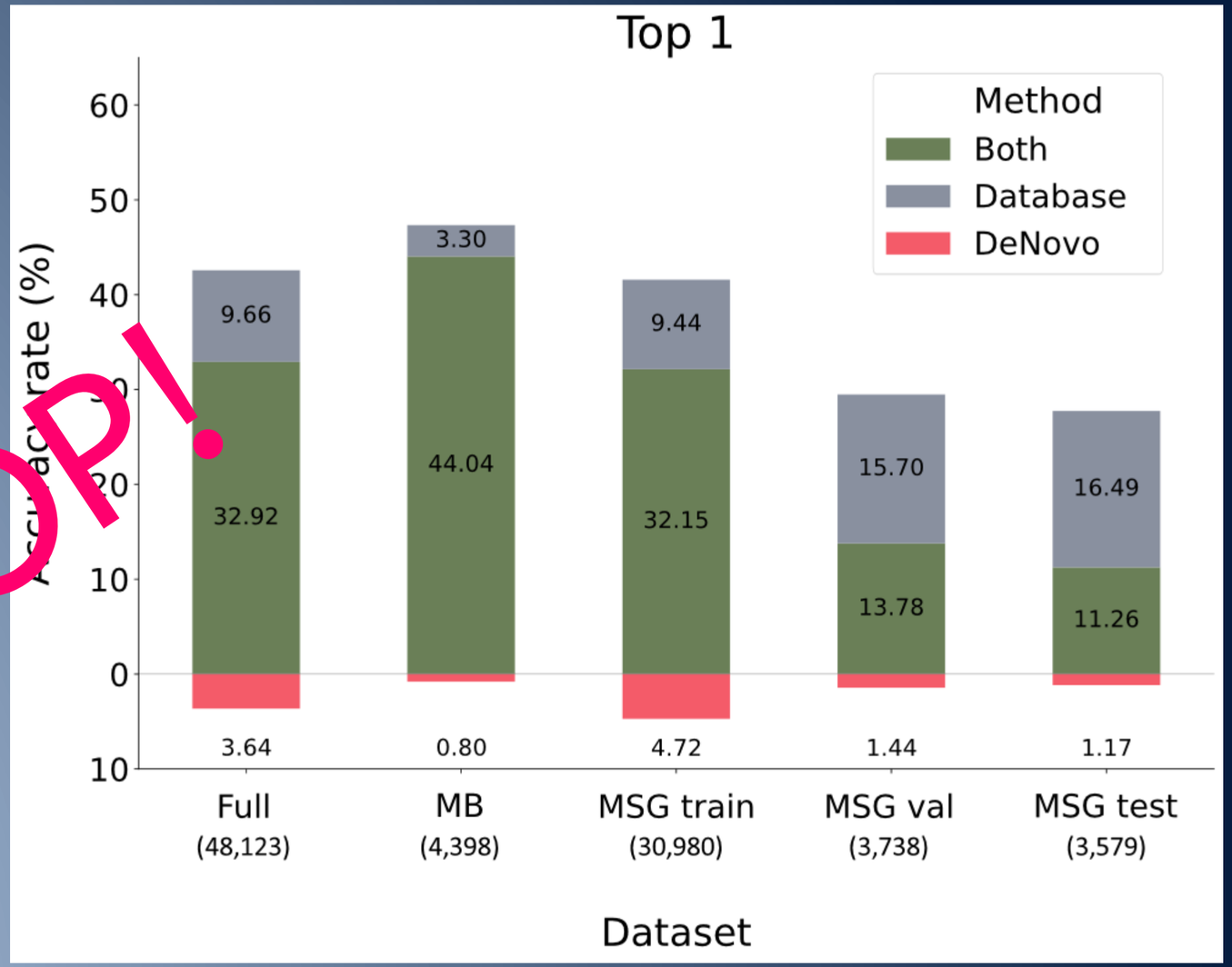
not in training 8.3%

anneli kruve

structural identification

from MS^2 spectra

STOP!



in training set 36.0%

not in training 8.3%

anneli kruve

take-home message

PRIORITIZE

(1) measuring performance is very important!
(2) chemistry still matters!

IDENTIFY

(1) all wrong structures are NOT false positives!
(2) multi-technique approach is important!

anneli kruve

anneli.kruve@su.se

anneli kruve

anneli.kruve@su.se