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. Nyheter Sverige Världen Ekonomi Kultur Sport Klimatet Ledare DN Debatt Me

SVERIGE

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

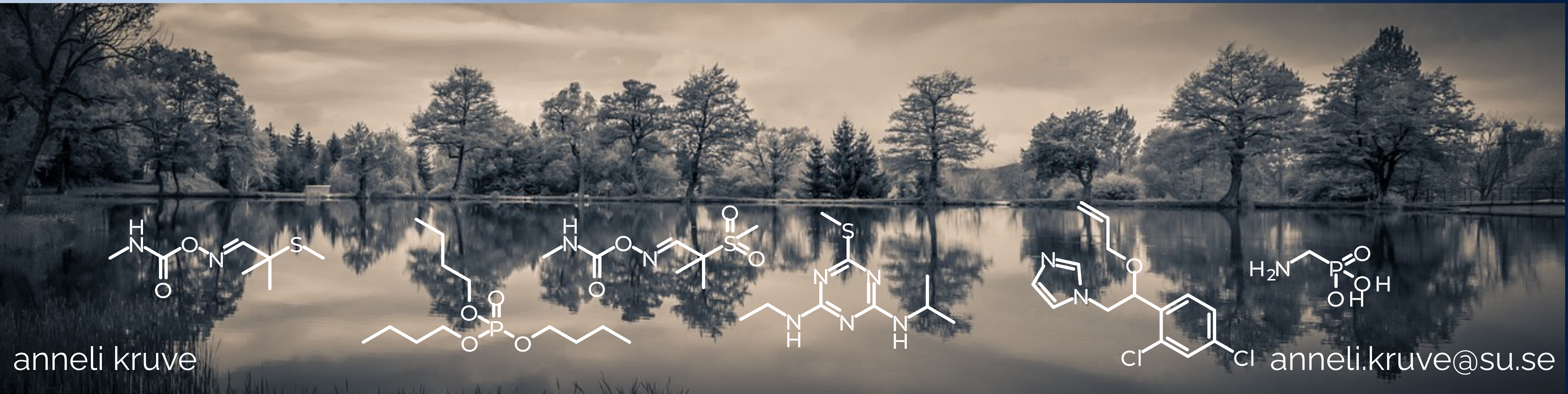
Uppdaterad 2024-12-02 Publicerad 2024-12-01

water analysis

anneli kruve

anneli.kruve@su.se

water analysis

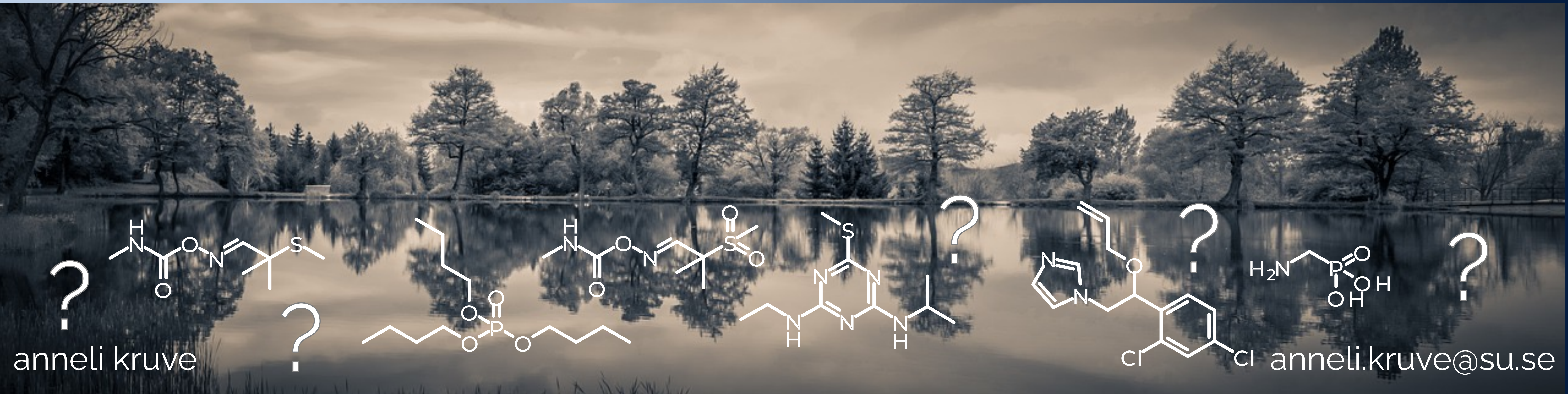


anneli kruve

anneli.kruve@su.se

water analysis

thousands of chemicals
detected
in the environment

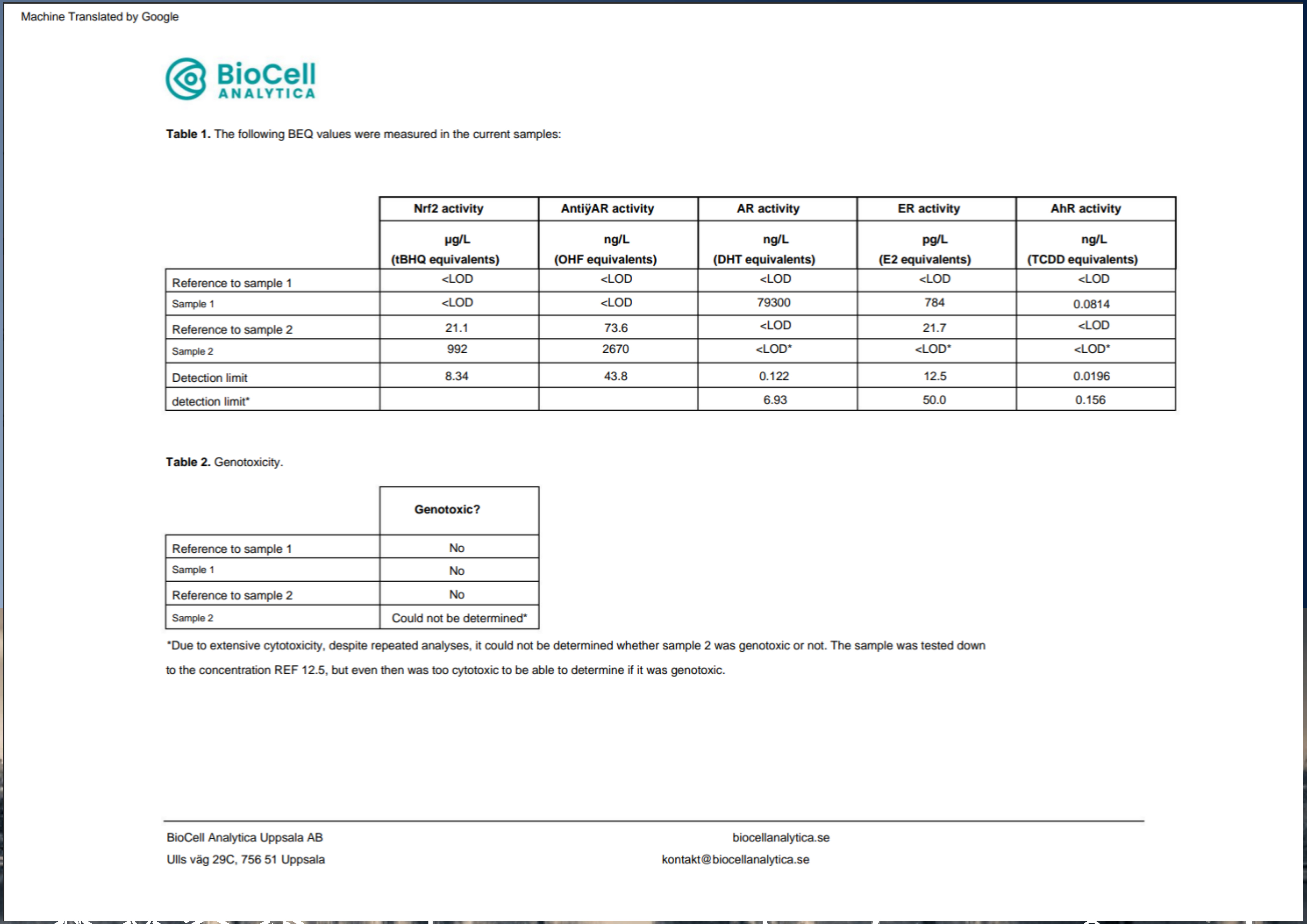


anneli kruve

anneli.kruve@su.se

water analysis

thousands of chemicals
detected
in the environment



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

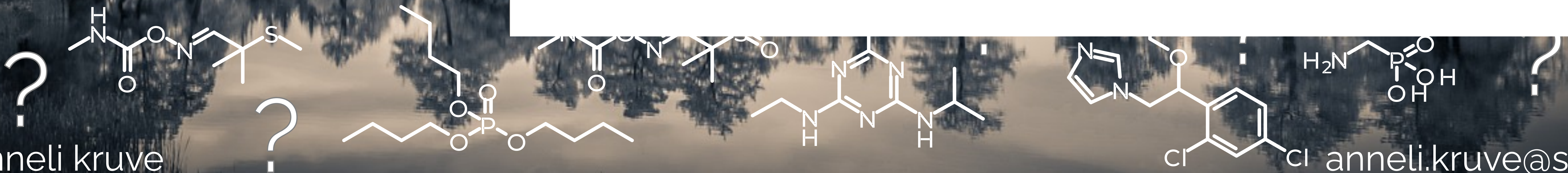
Table 2. Genotoxicity.

	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

Szabo et al. Anal Chem 2024

sample → online prioritization → HRMS → offline prioritization → chemical risk

online prioritization:
- top N
- inclusion list
- isotopic abundance
- adduct formation
- MS^2 analysis

offline prioritization:
- recursive inc/exc
- relative abundance
- differential analysis
- metabolic logic
- modelling

anneli kruve

anneli.kruve@su.se

prioritization of chemicals

TOXICITY

ecotoxicity and endocrine disruptors

CONCENTRATION

exposure to potentially toxic chemicals

RISK

$$\mathrm{PriorityScore} = \frac{c_{\mathrm{predicted}}}{\mathrm{AC}_{50}^{\mathrm{5th\ percentile}}}$$

anneli kruve

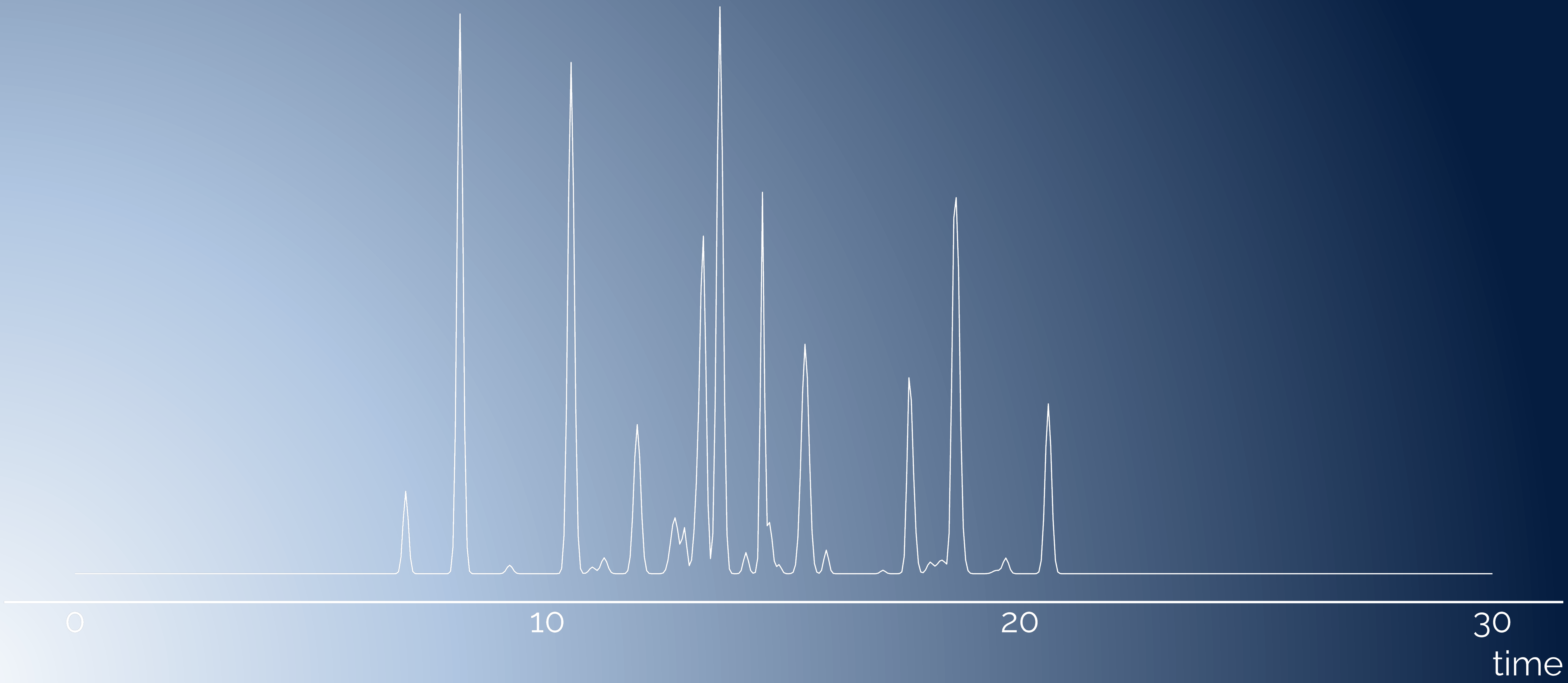
anneli.kruve@su.se

toxicity

of detected chemicals

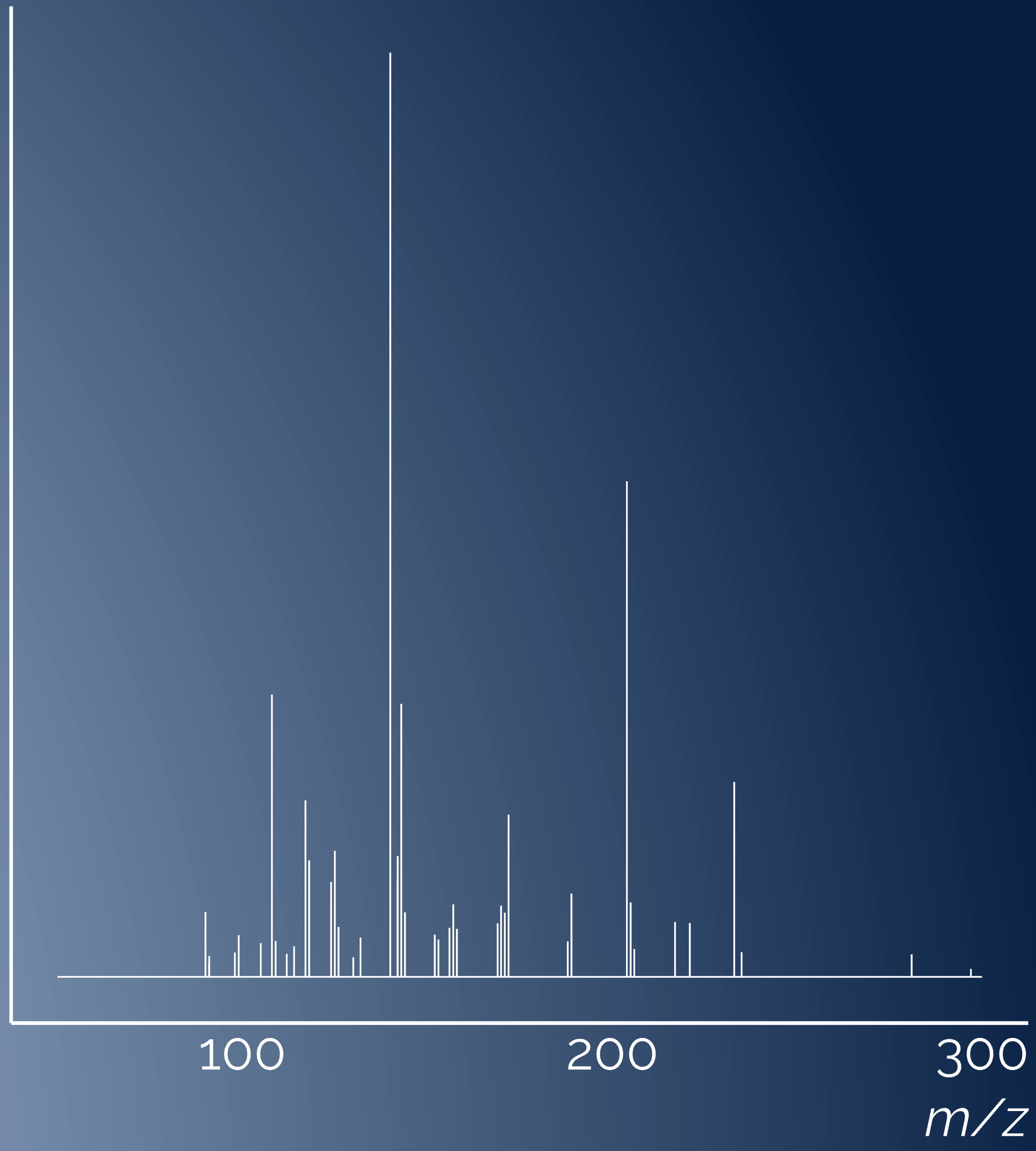
nontarget screening

with LC/HRMS



toxicity assessment

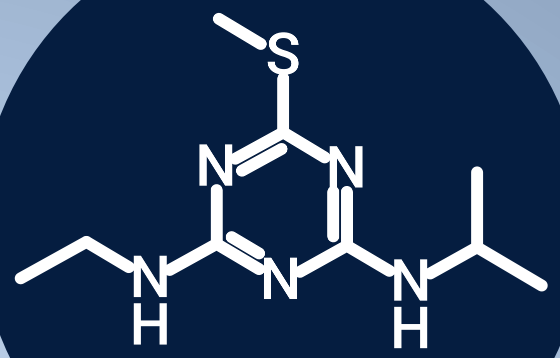
from spectra
to structure
to toxicity



anneli kruve

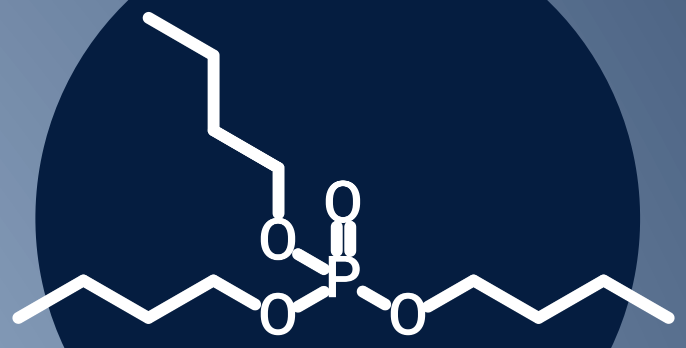
anneli.kruve@su.se

toxicity assessment



LC_{50} = 9.3 mg/L

known structure
known toxicity



LC_{50} = ? mg/L

known structure
unknown toxicity

?

LC_{50} = ? mg/L

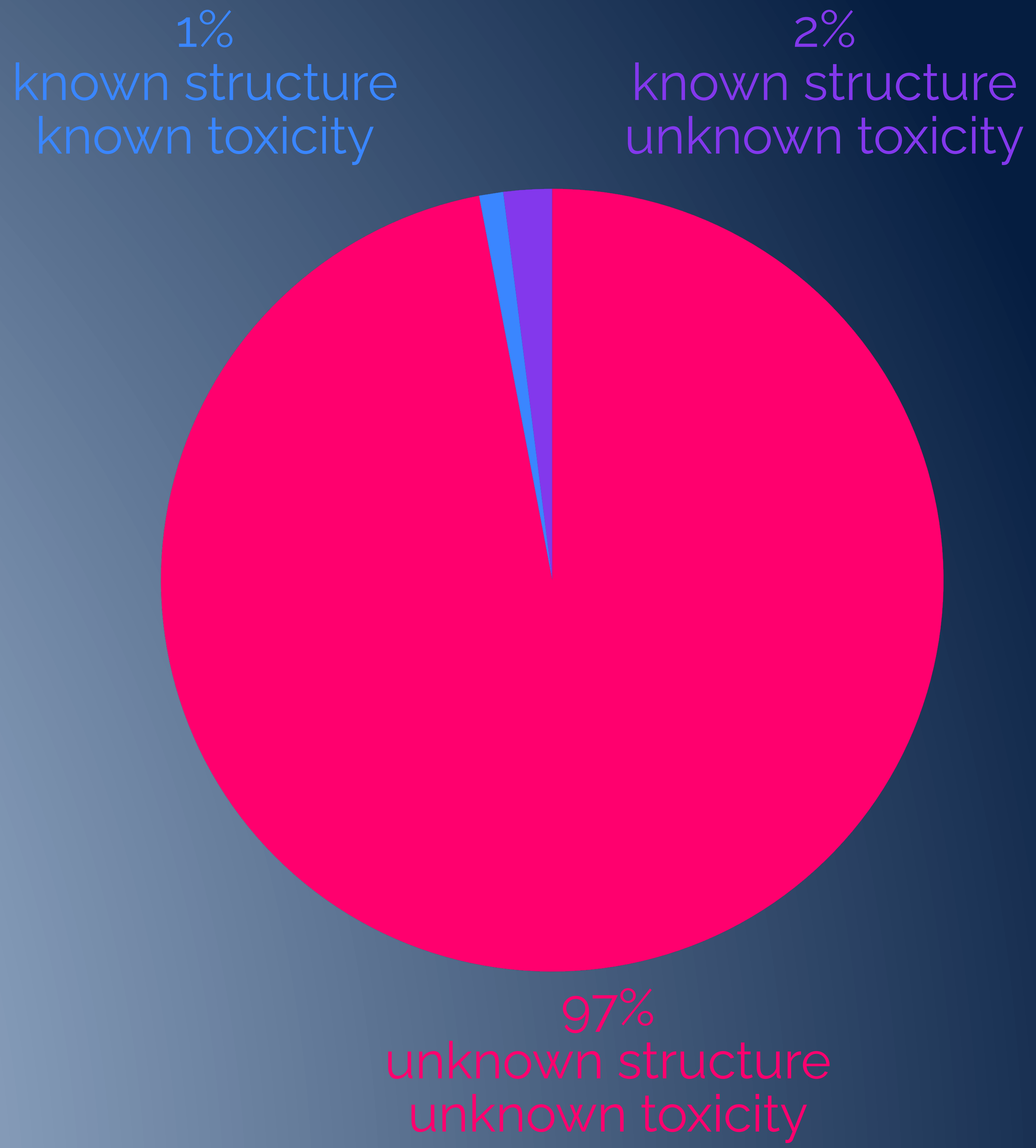
unknown structure
unknown toxicity

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

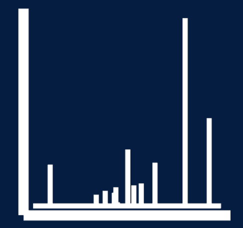
ENDOCRINE DISRUPTION

AhR, AR, ER, MMP, P53, …

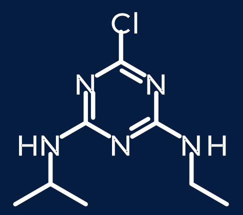
anneli kruve

anneli.kruve@su.se

workflow



MS^2 spectra



structure as SMILES

molecular descriptors

toxicity prediction

anneli kruve

anneli.kruve@su.se

workflow

MS² spectra

molecular descriptors

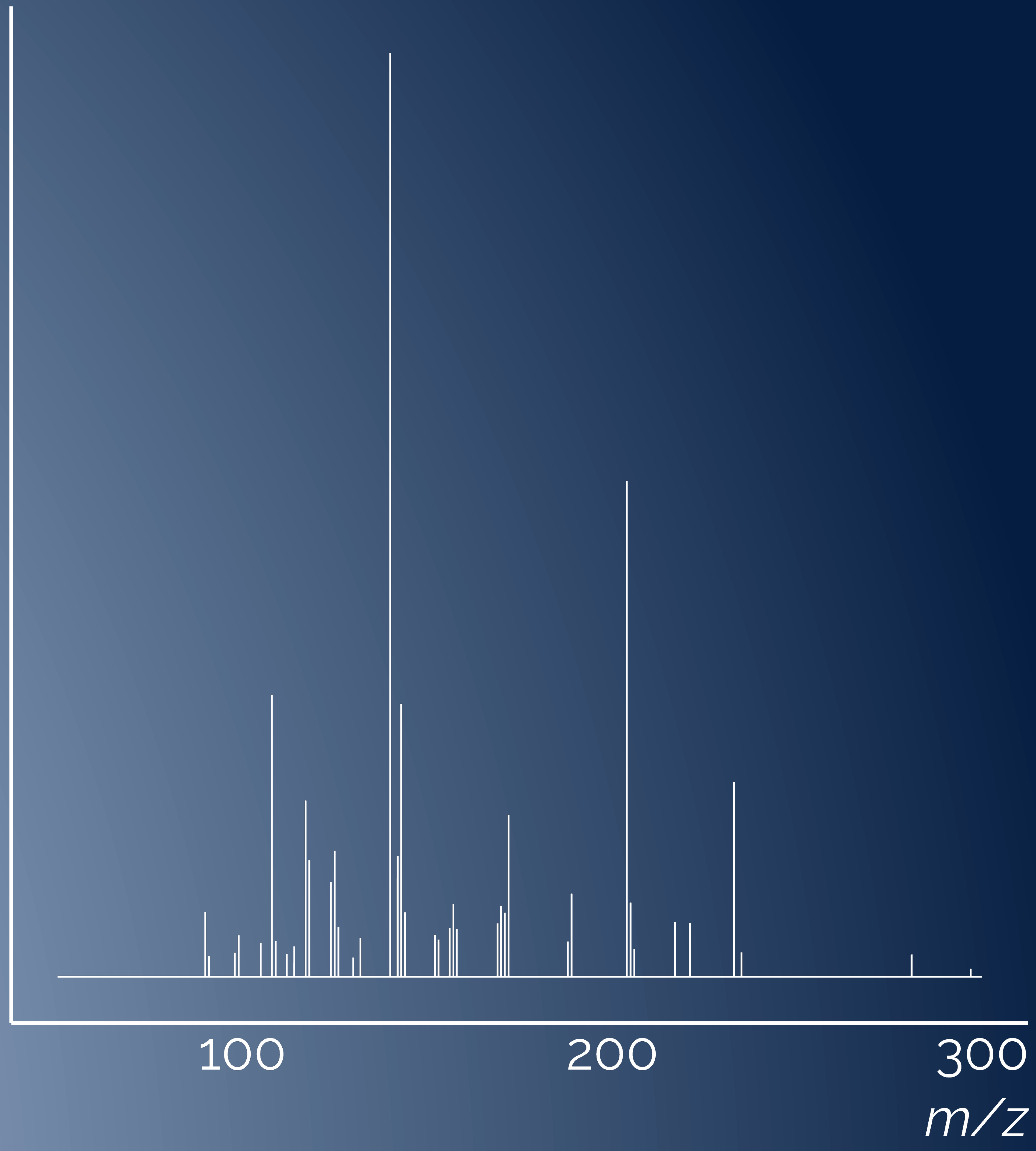
toxicity prediction

anneli kruve

anneli.kruve@su.se

information available

in mass spectra

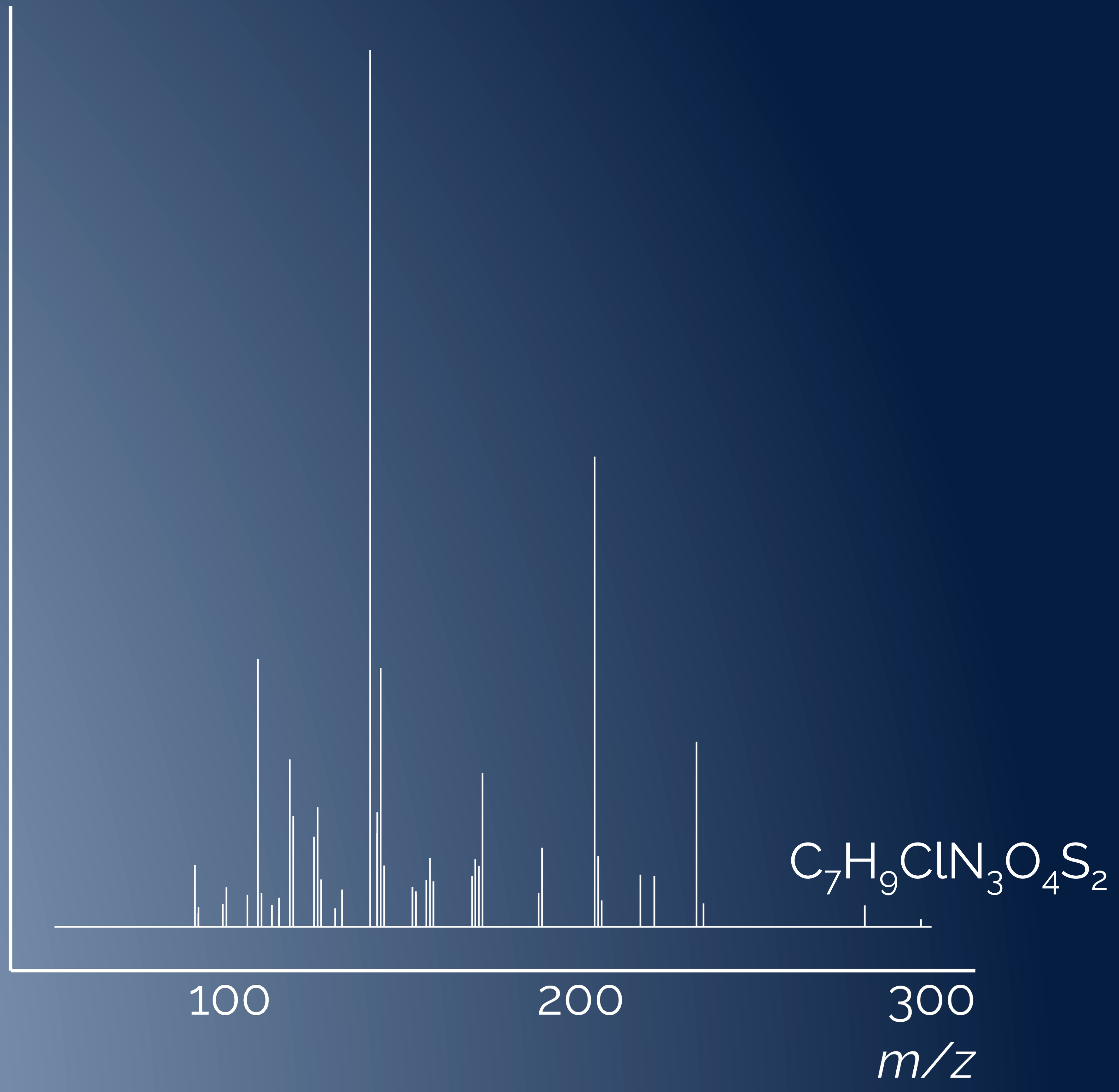


anneli kruve

anneli.kruve@su.se

information available

in mass spectra

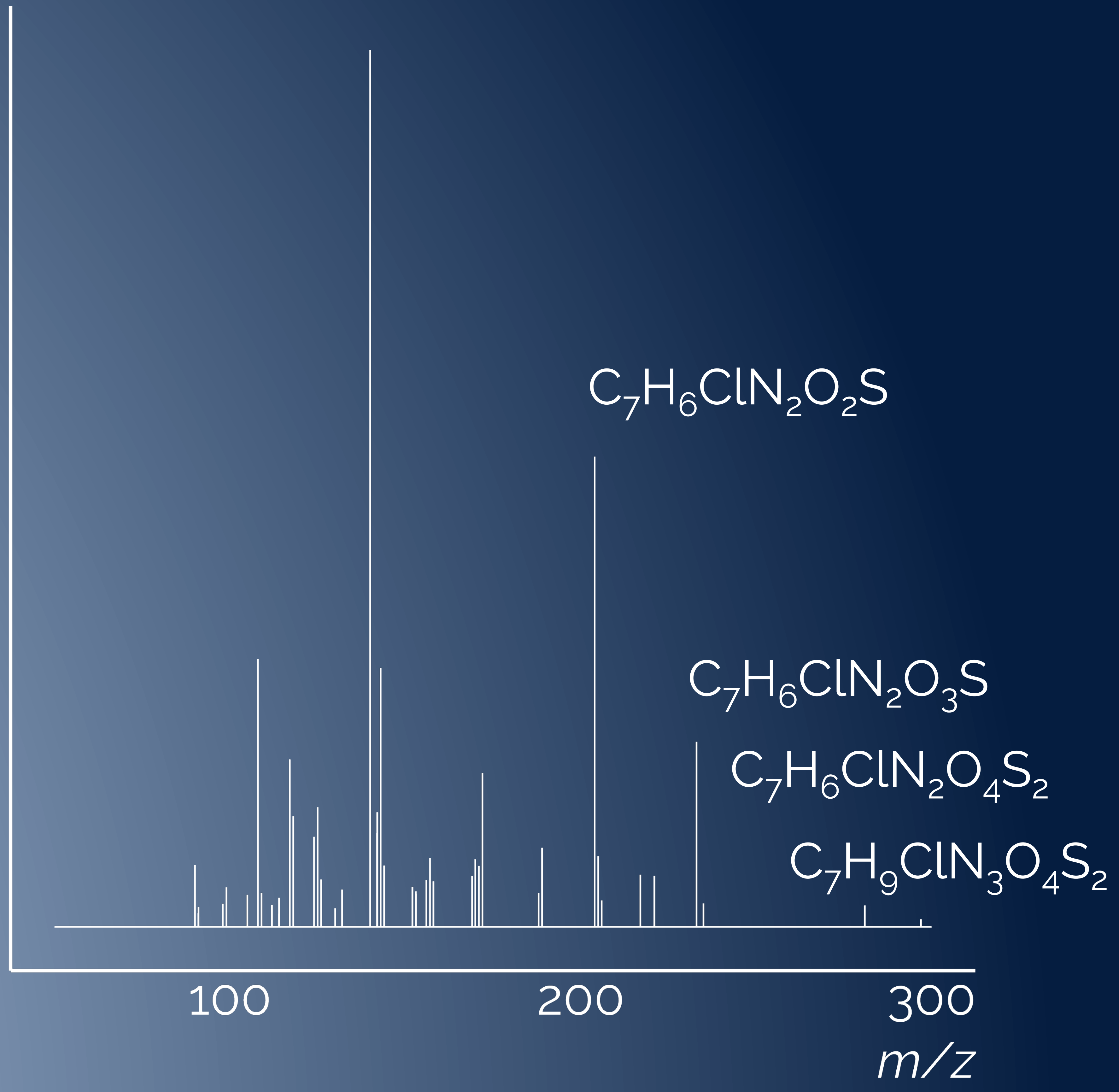


anneli kruve

anneli.kruve@su.se

information available

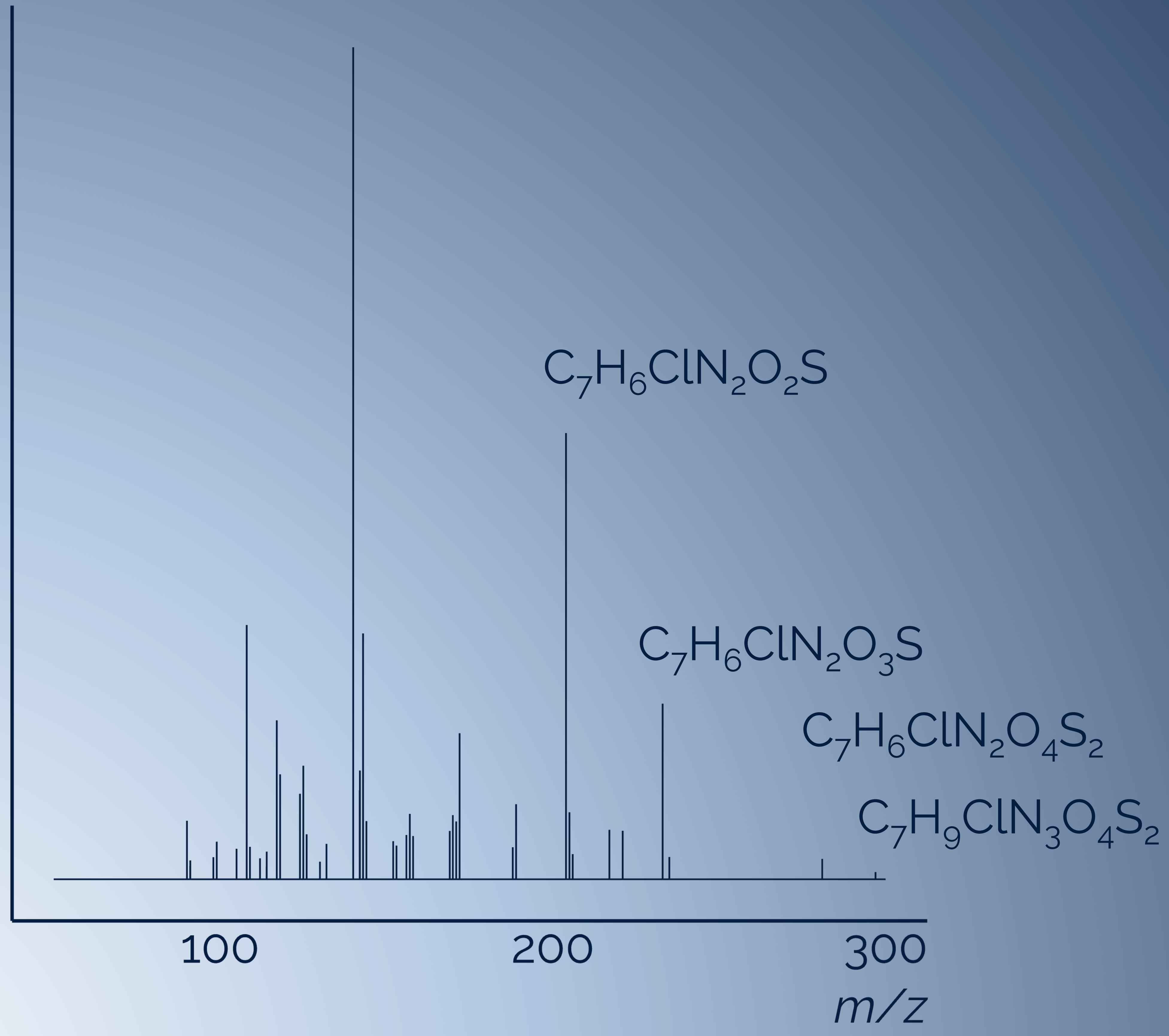
in mass spectra



anneli kruve
anneli.kruve@su.se

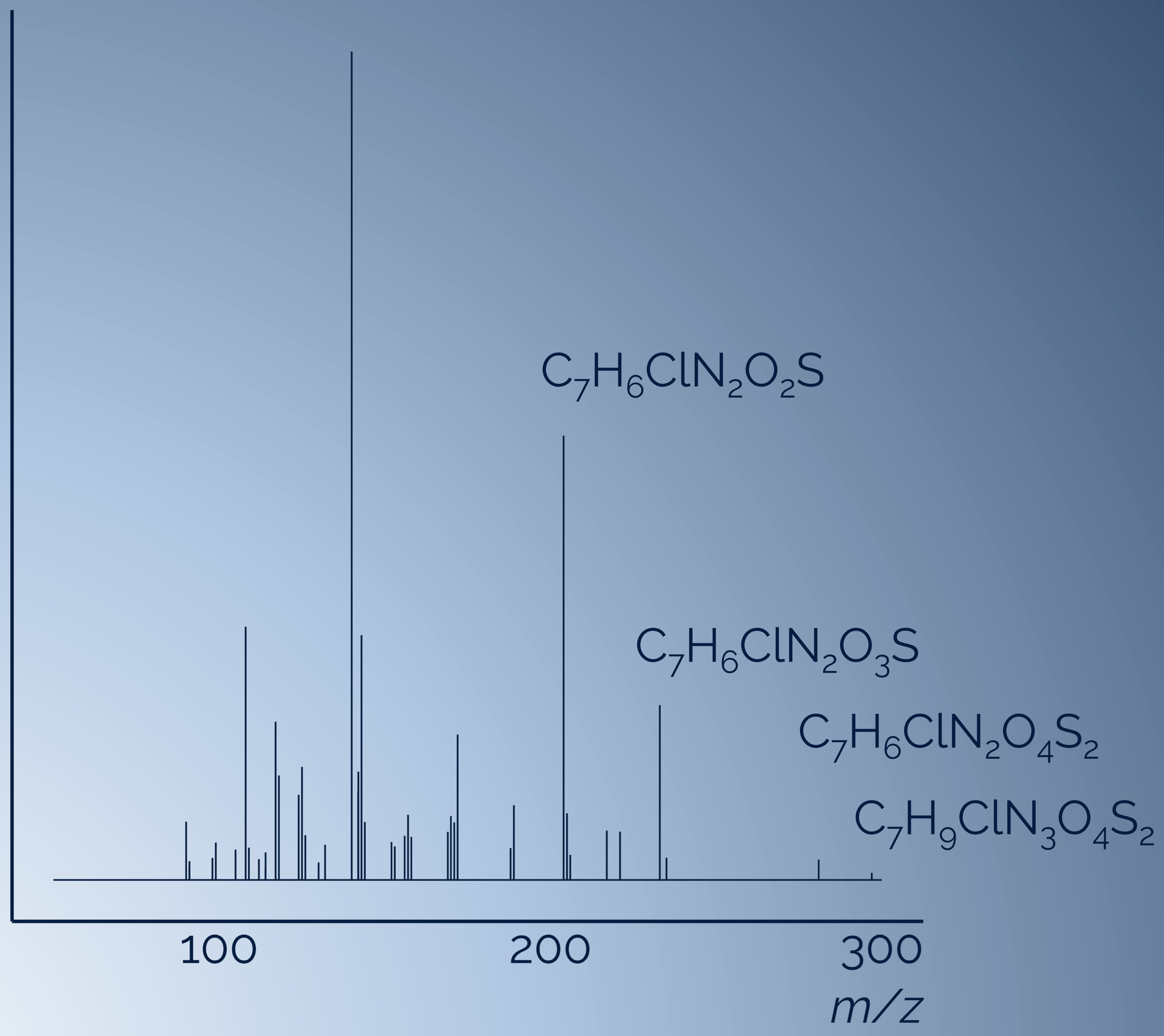
information available

in mass spectra



anneli kruve

anneli.kruve@su.se

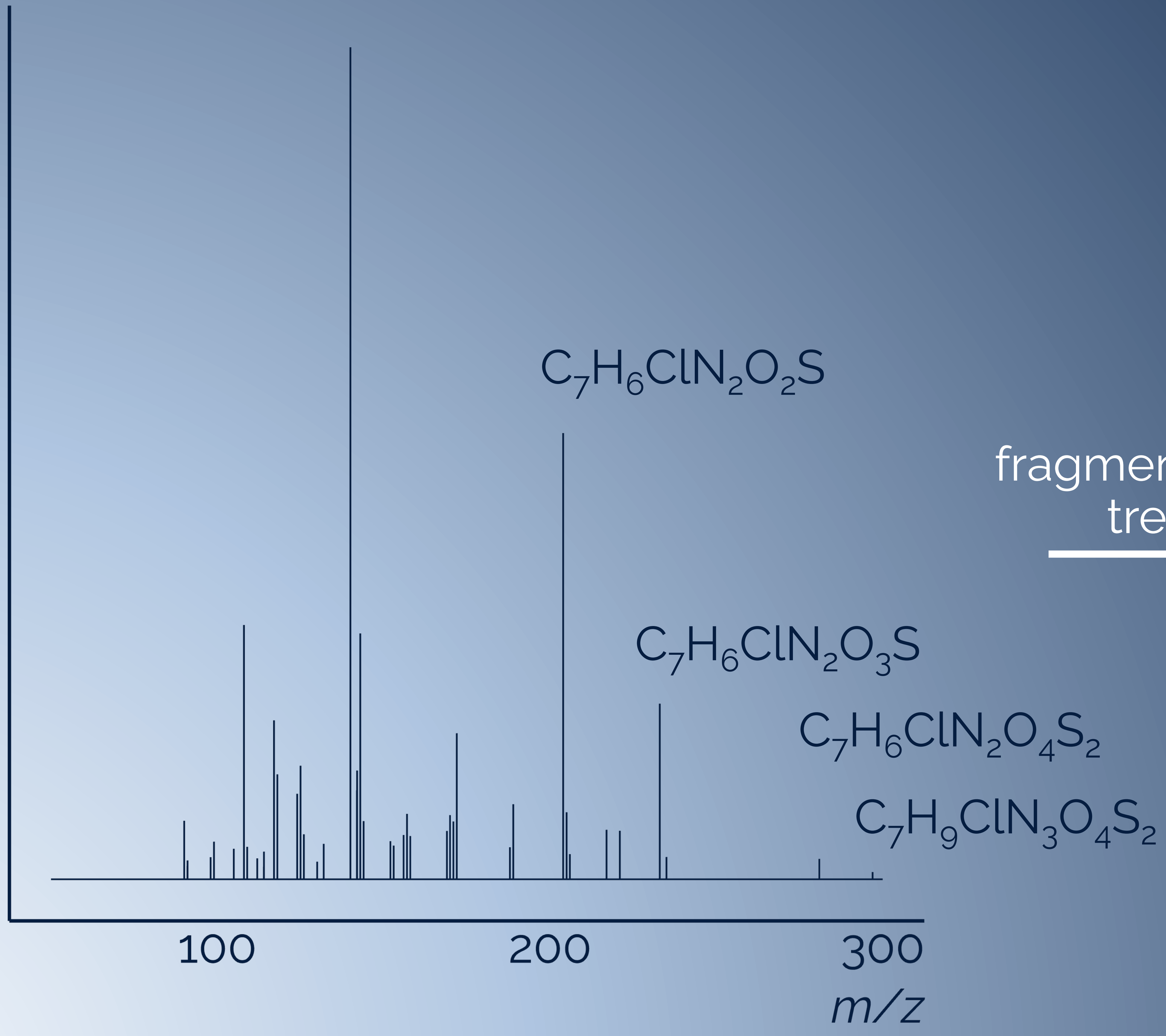


predict toxicity

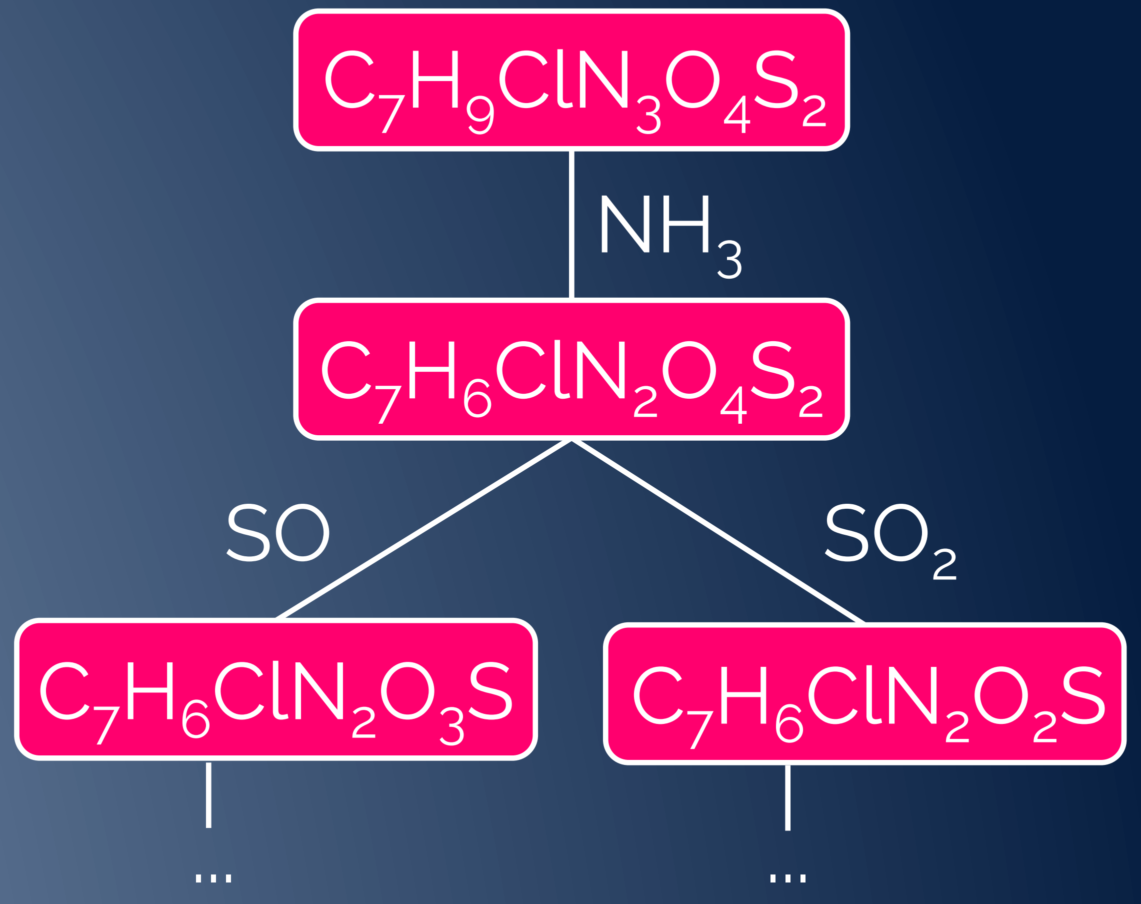
for unknown chemicals

anneli kruve

anneli.kruve@su.se

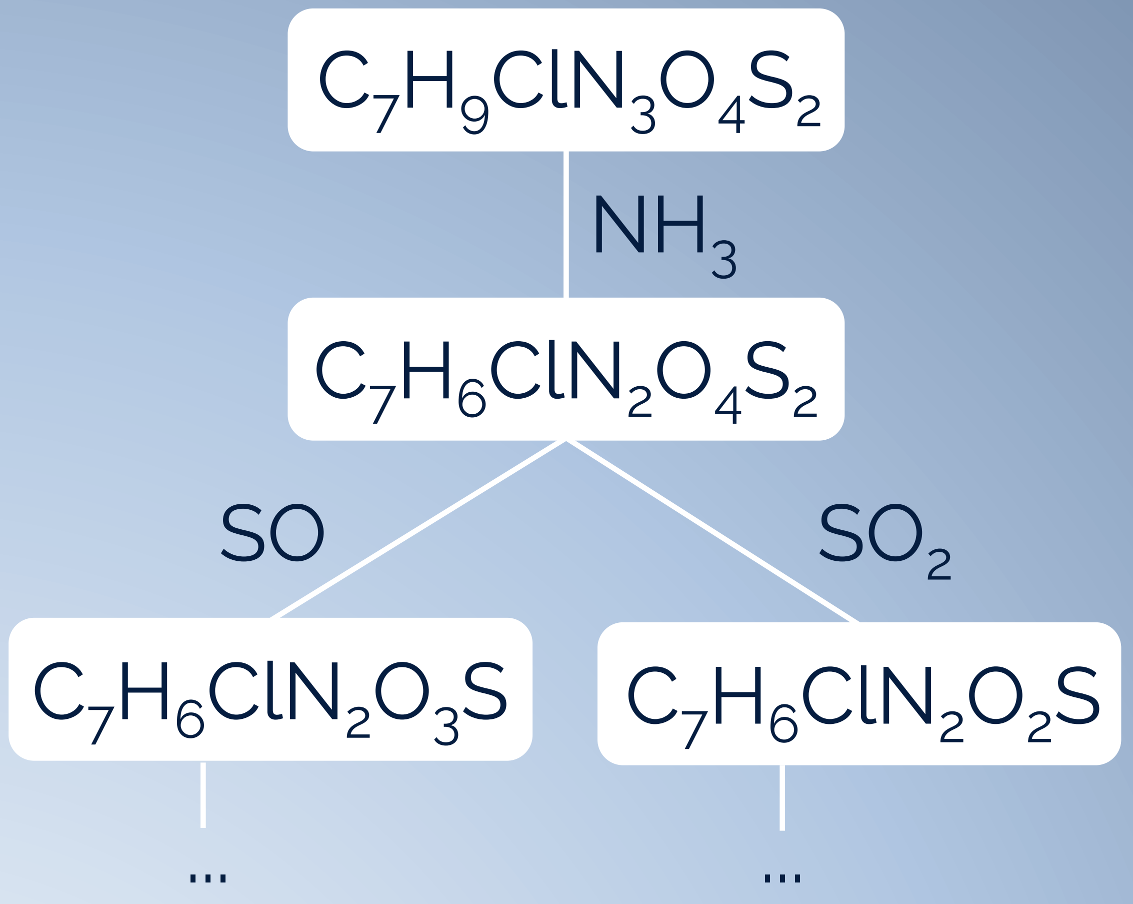


fragmentation tree

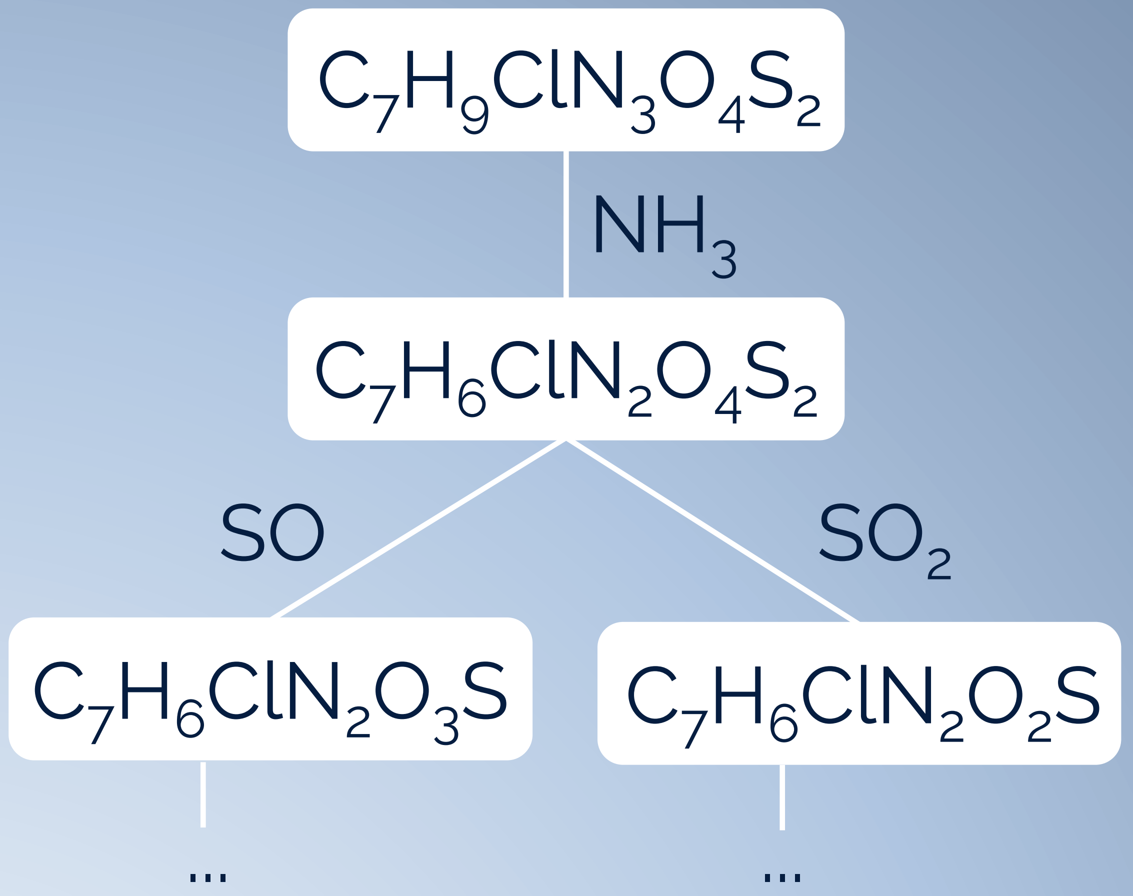


anneli kruve

anneli.kruve@su.se



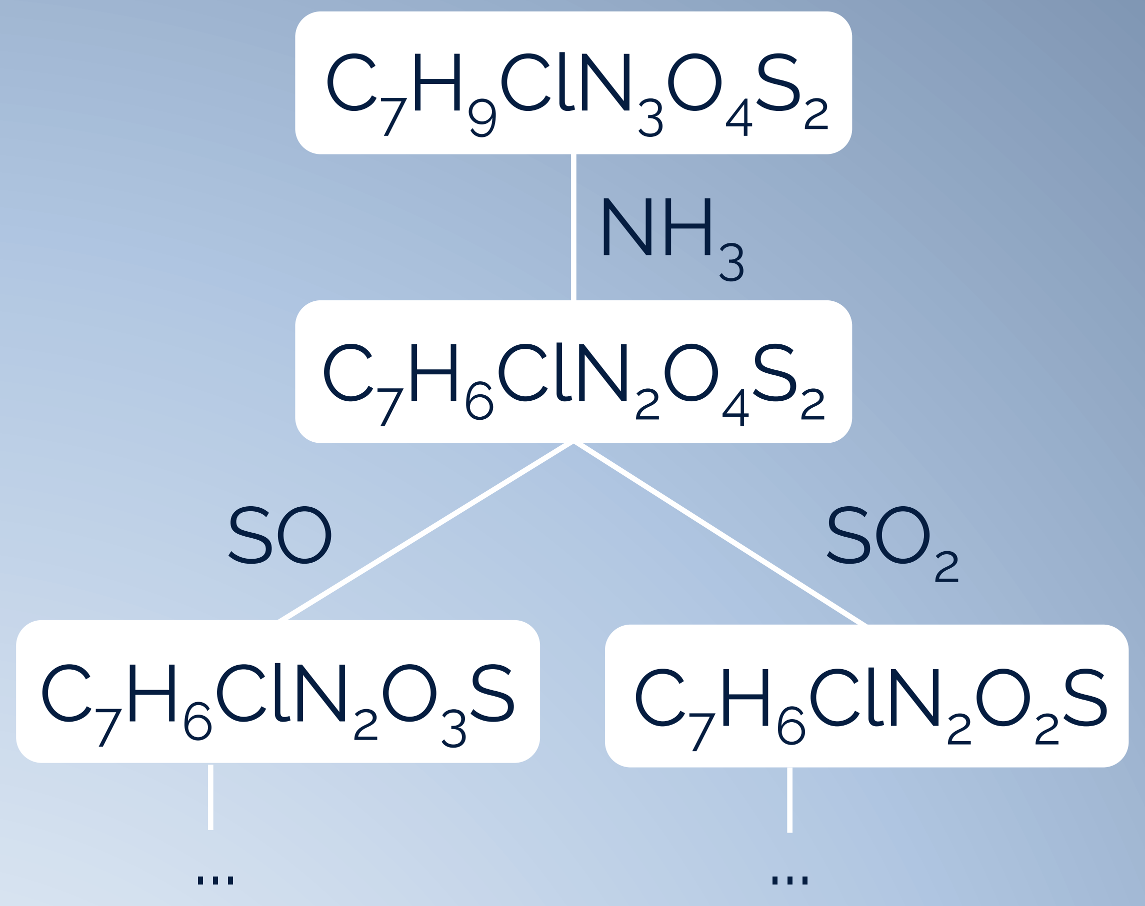
anneli kruve
anneli.kruve@su.se



machine learning →

0	O-P
1	S=O
0	C-F
1	NH_2
1	O

anneli kruve
anneli.kruve@su.se



machine learning →

0	O-P
1	S=O
0	C-F
1	NH_2
1	O

machine learning →

p(AhR active)= 0.83

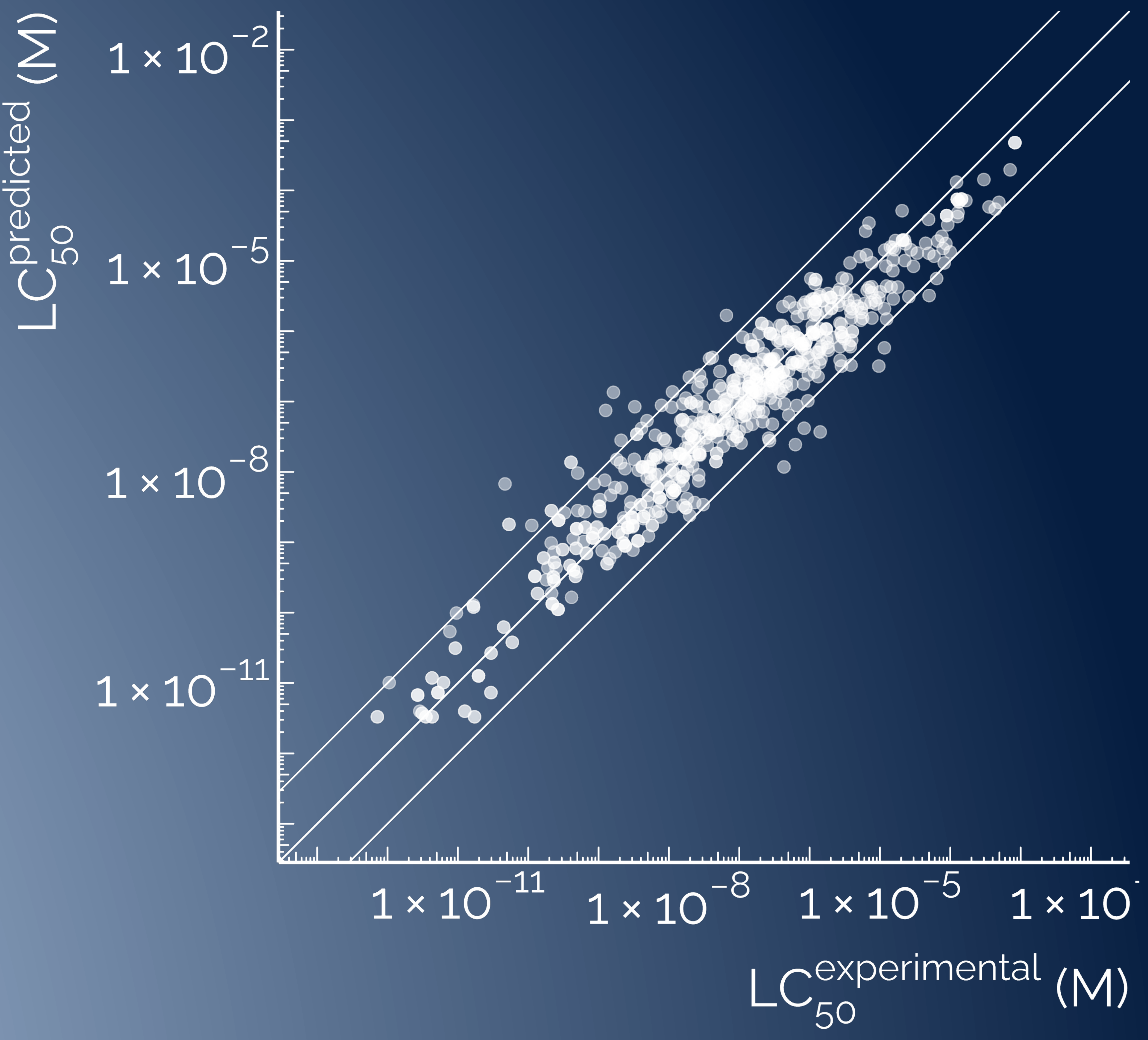
anneli kruve

anneli.kruve@su.se

LC_{50} predictions

Peets et al. ES&T 2022

fish LC_{50}



test set

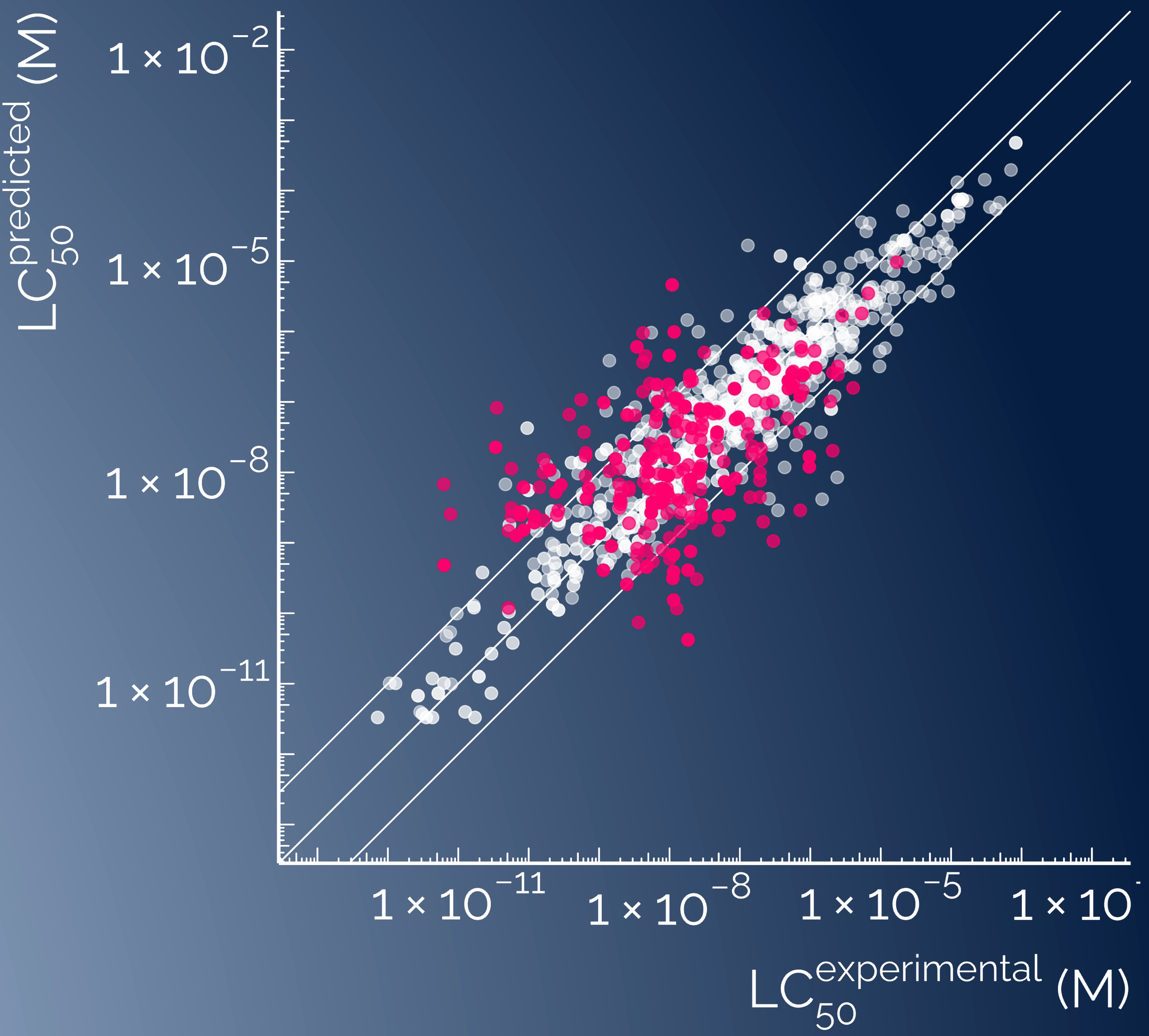
RMSE 0.78 log(M)

anneli kruve

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

Rahu et al. JCIM 2024
Tox21 endpoints

true label	
active	non-active

anneli.kruve@su.se

endocrine disruption

Rahu et al. JCIM 2024
Tox21 endpoints

		true label	
		active	non-active
prediction	active		
	non-active		

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

		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 compounds

anneli kruve
anneli.kruve@su.se

alternative approaches

Kreutzer et al. in preparation
Tox21 endpoints

 anneli.kruve@su.se

molecular networks

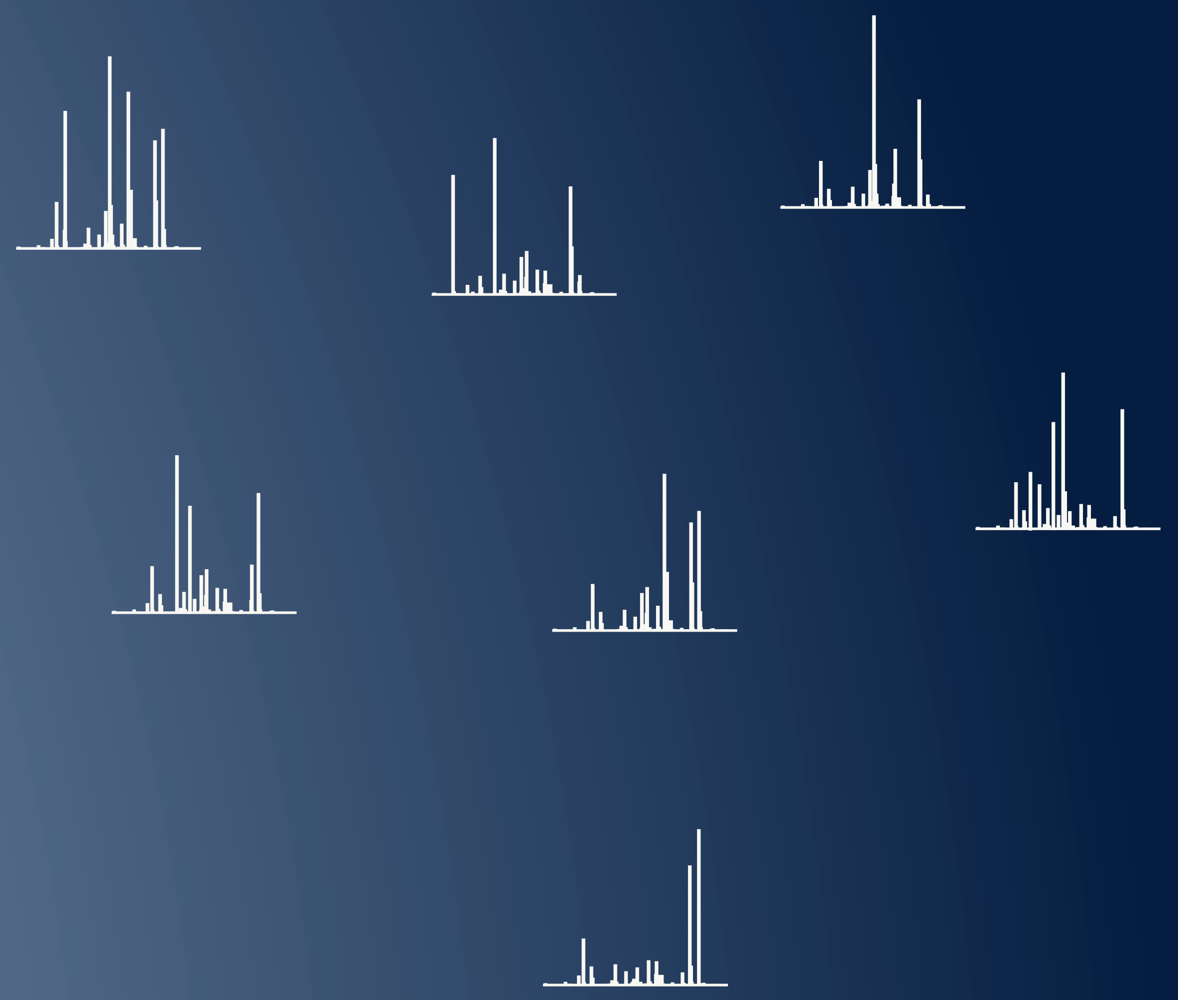
Kreutzer et al. in preparation
Tox21 endpoints

anneli kruve

anneli.kruve@su.se

molecular networks

Kreutzer et al. in preparation
Tox21 endpoints

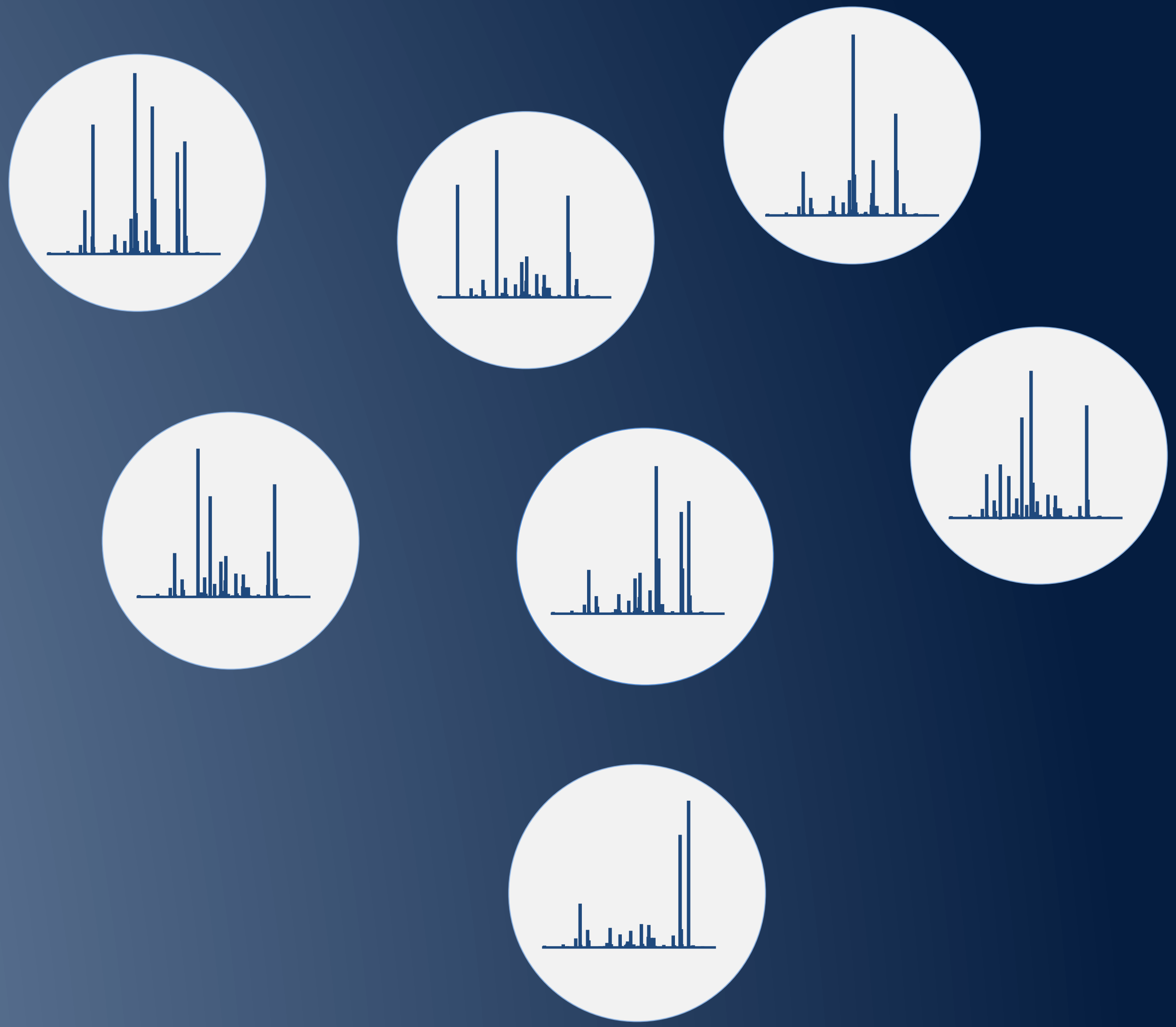


anneli kruve

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

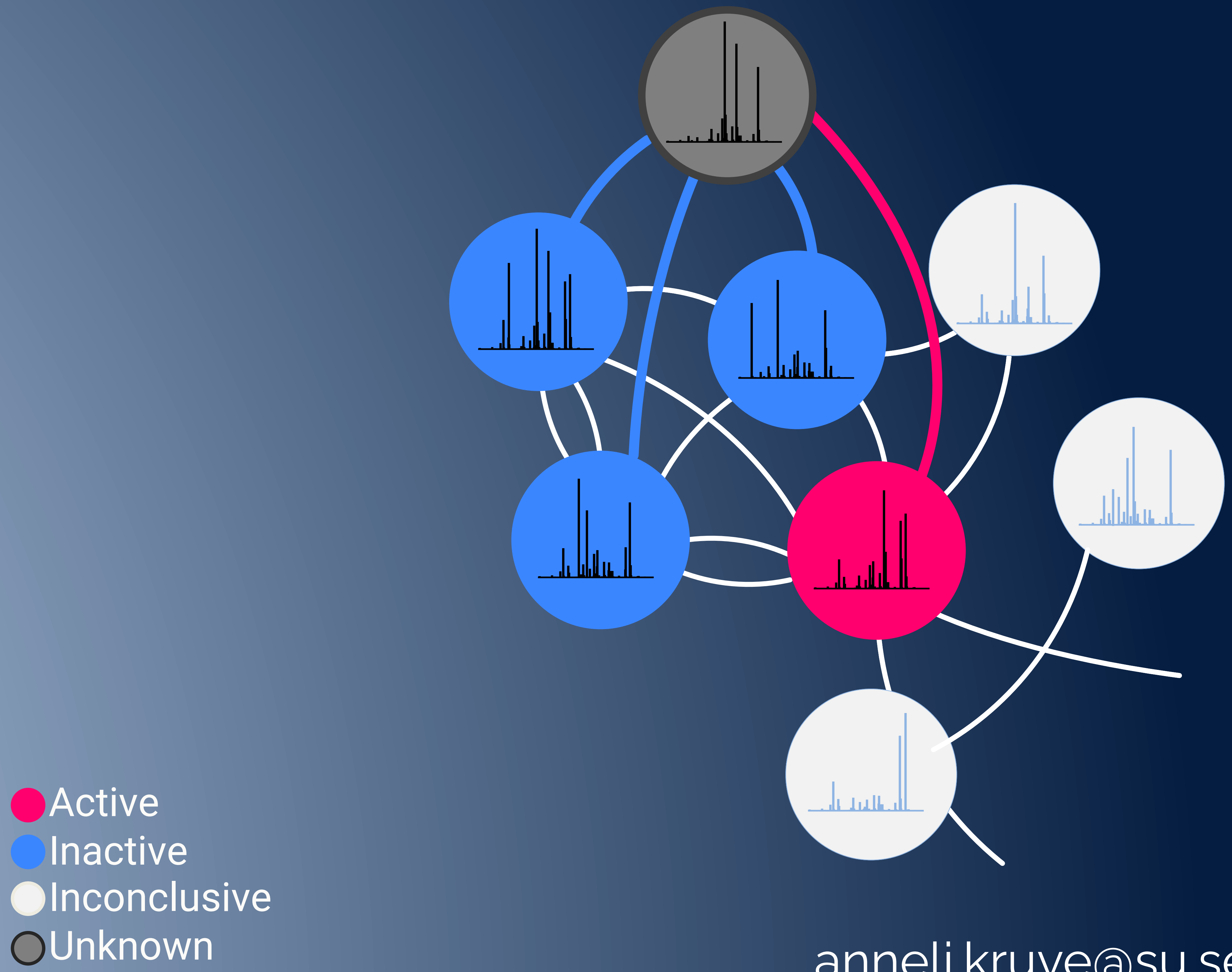
Active
Inactive
Inconclusive

anneli kruve

anneli.kruve@su.se

molecular networks

Kreutzer et al. in preparation
Tox21 endpoints

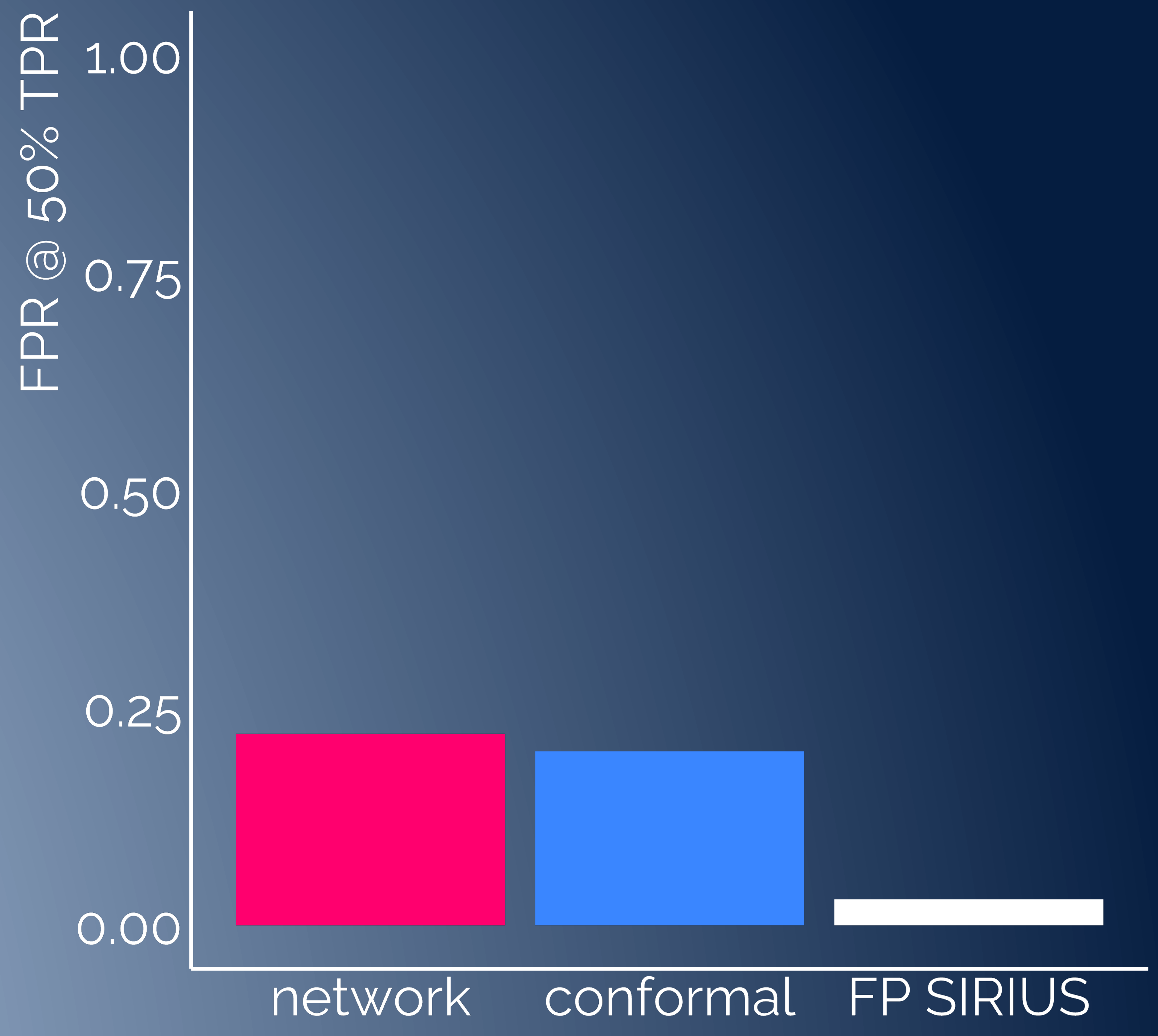


anneli kruve

anneli.kruve@su.se

AhR activity predictions

Kreutzer et al. in preparation
Tox21 endpoints



anneli kruve
anneli.kruve@su.se

case study:
interlaboratory comparison

water analysis

thousands of chemicals detected

in the environment

Machine Translated by Google

BioCell ANALYTICA

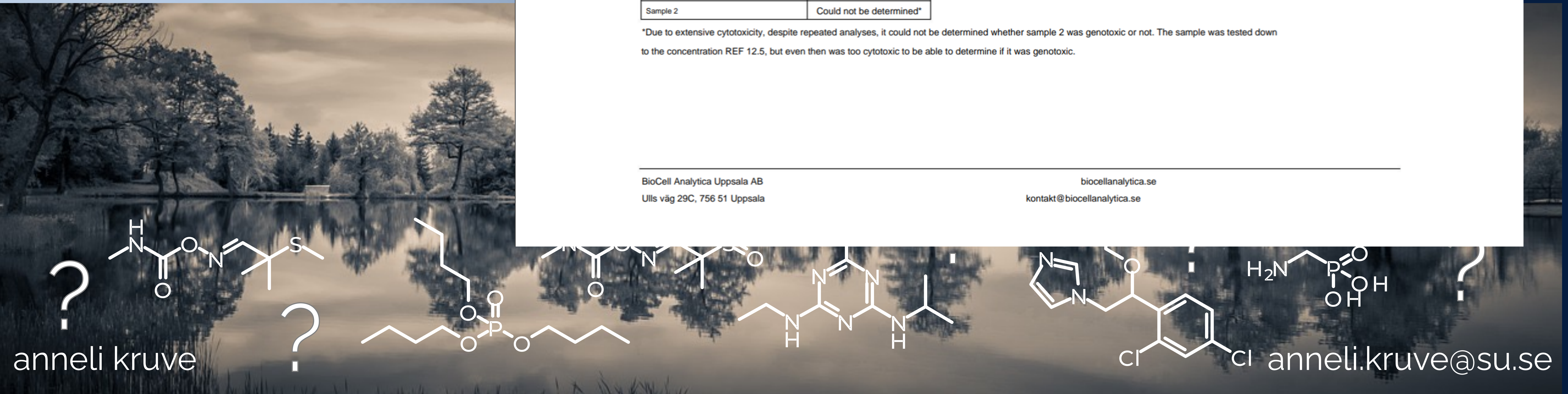
Table 1. The following BEQ values were measured in the current samples:

	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.



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@su.se

results

4700

LC/HRMS features detected

238

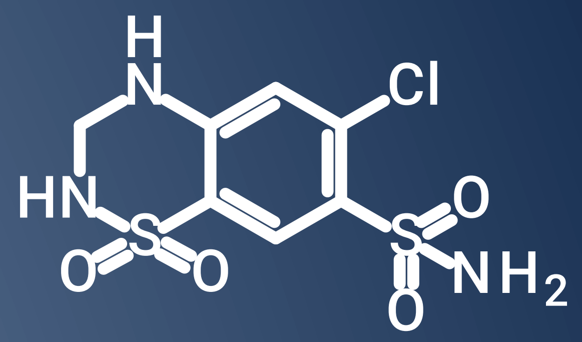
features with MS^2 spectra

55

features predicted active

anneli.kruve@su.se

AhR active



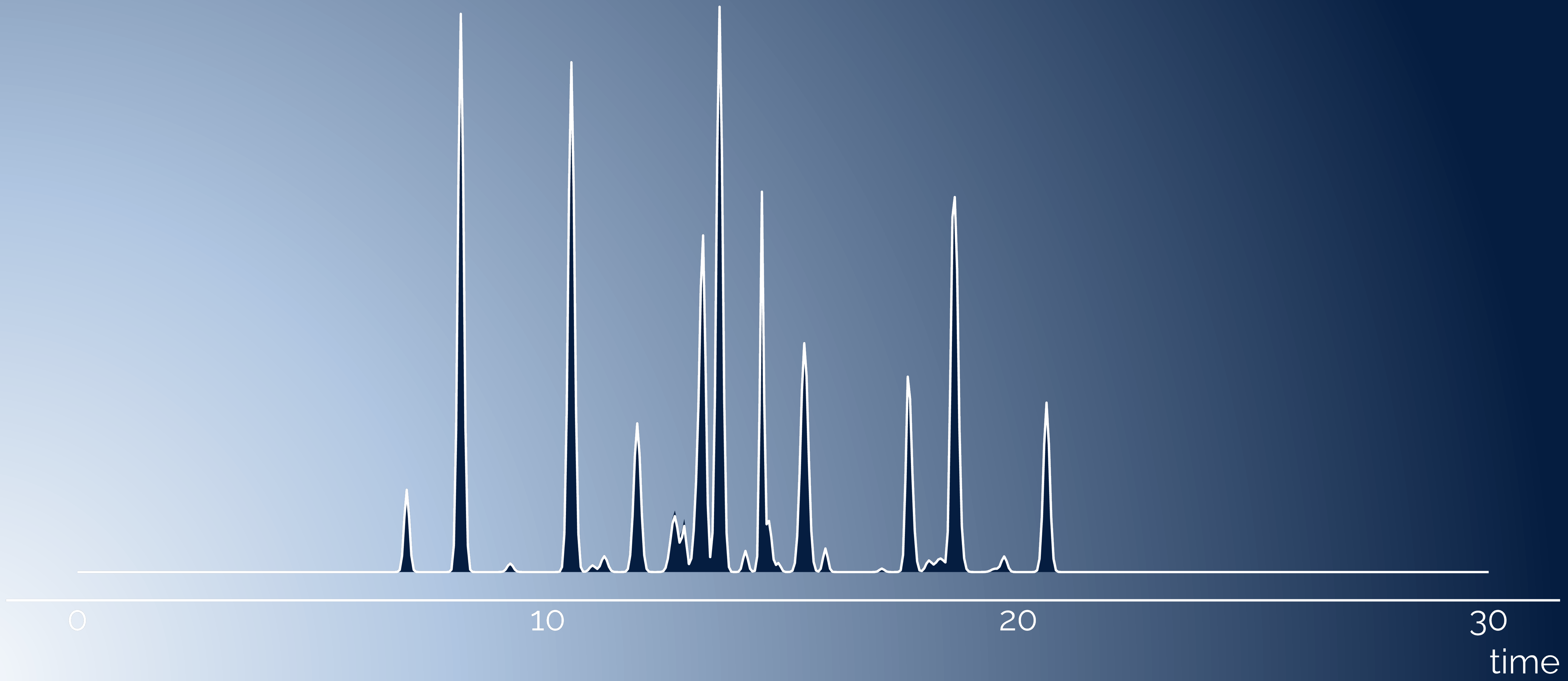
hydrochlorothiazide

quantification

of detected chemicals

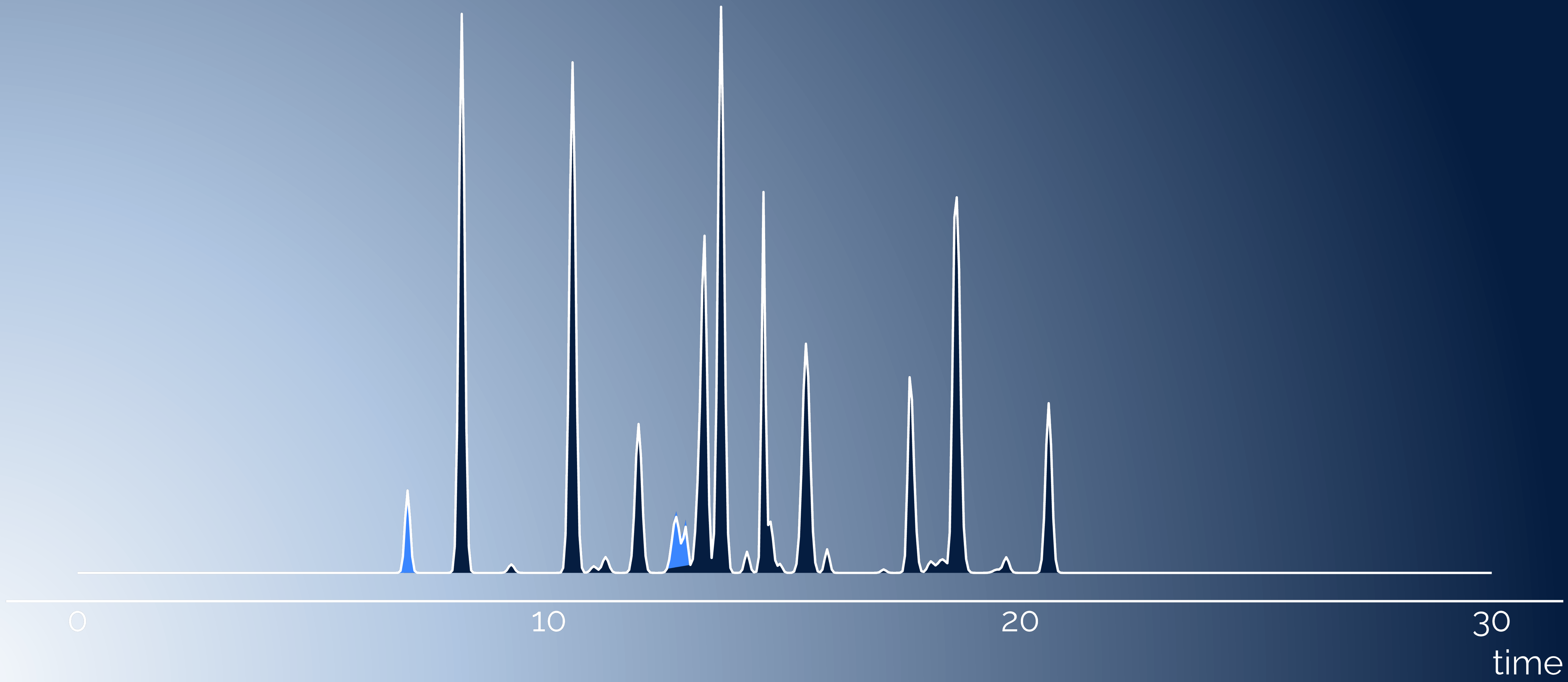
quantification in ESI/HRMS

Malm et al. Molecules 2021



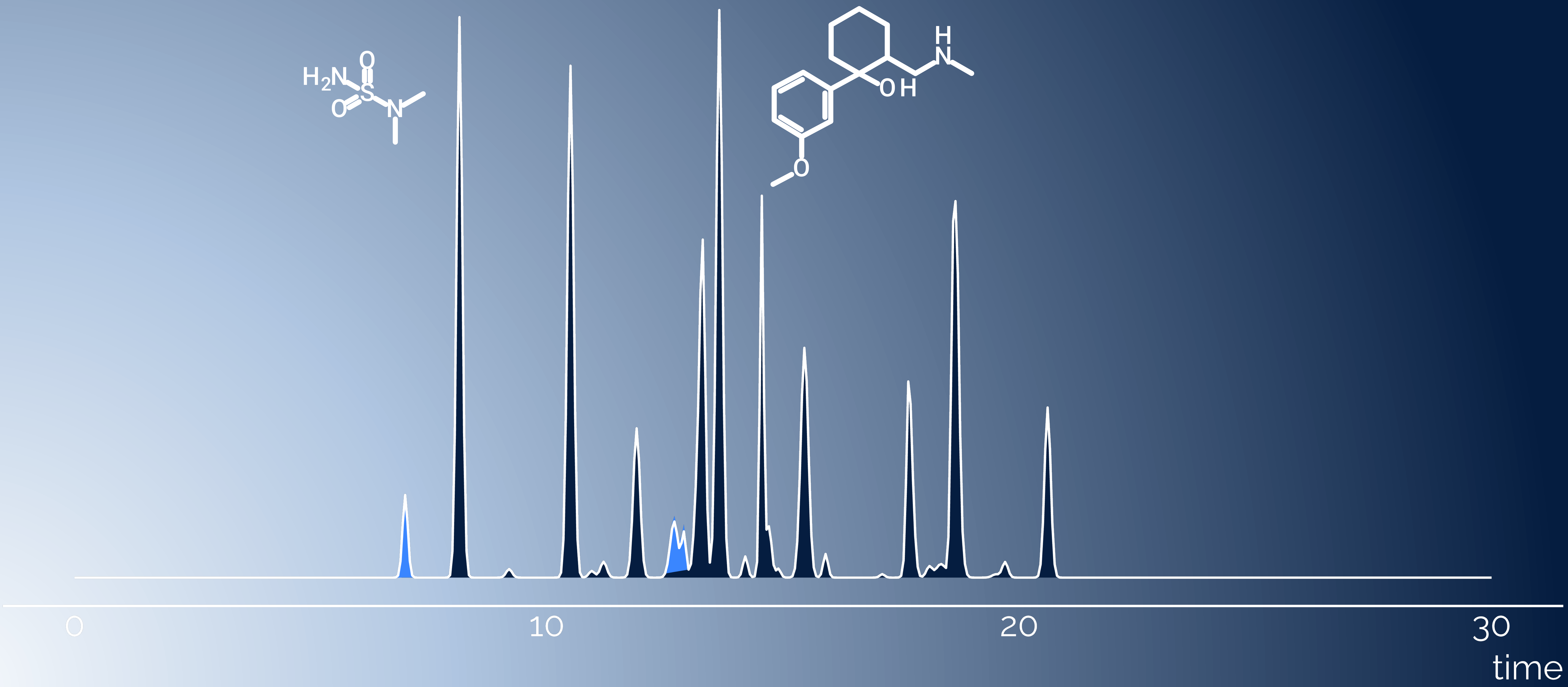
quantification in ESI/HRMS

Malm et al. Molecules 2021



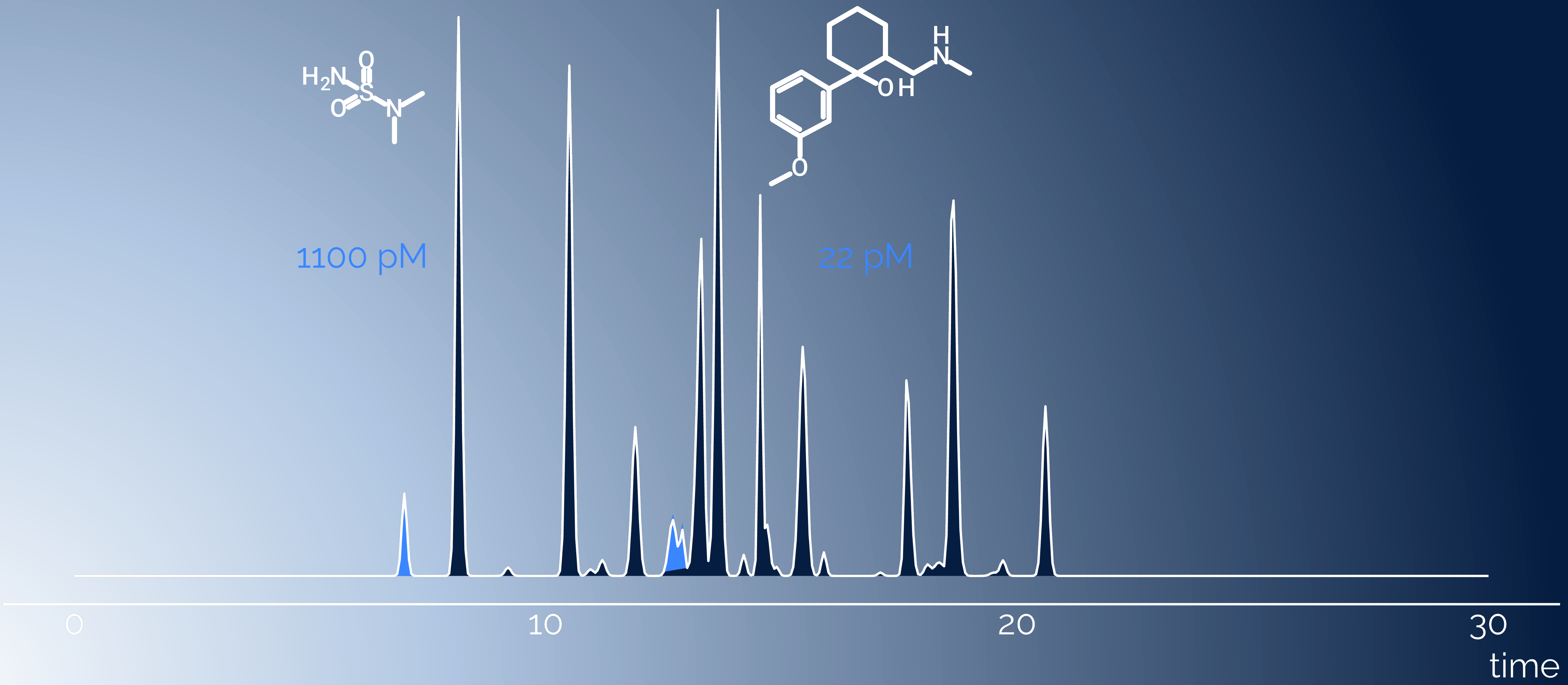
quantification in ESI/HRMS

Malm et al. Molecules 2021



quantification in ESI/HRMS

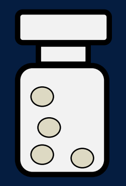
Malm et al. Molecules 2021



electrospray

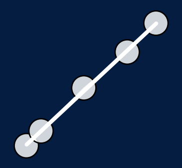
 anneli.kruve@su.se

ionization efficiency



ANALYSIS

flow injections or LC



DATA

calibration graphs

$$\frac{slope_1}{slope_2} \rightarrow IE$$

IONIZATION EFFICIENCY

relative measurements

anneli kruve

anneli.kruve@su.se

ionization efficiency

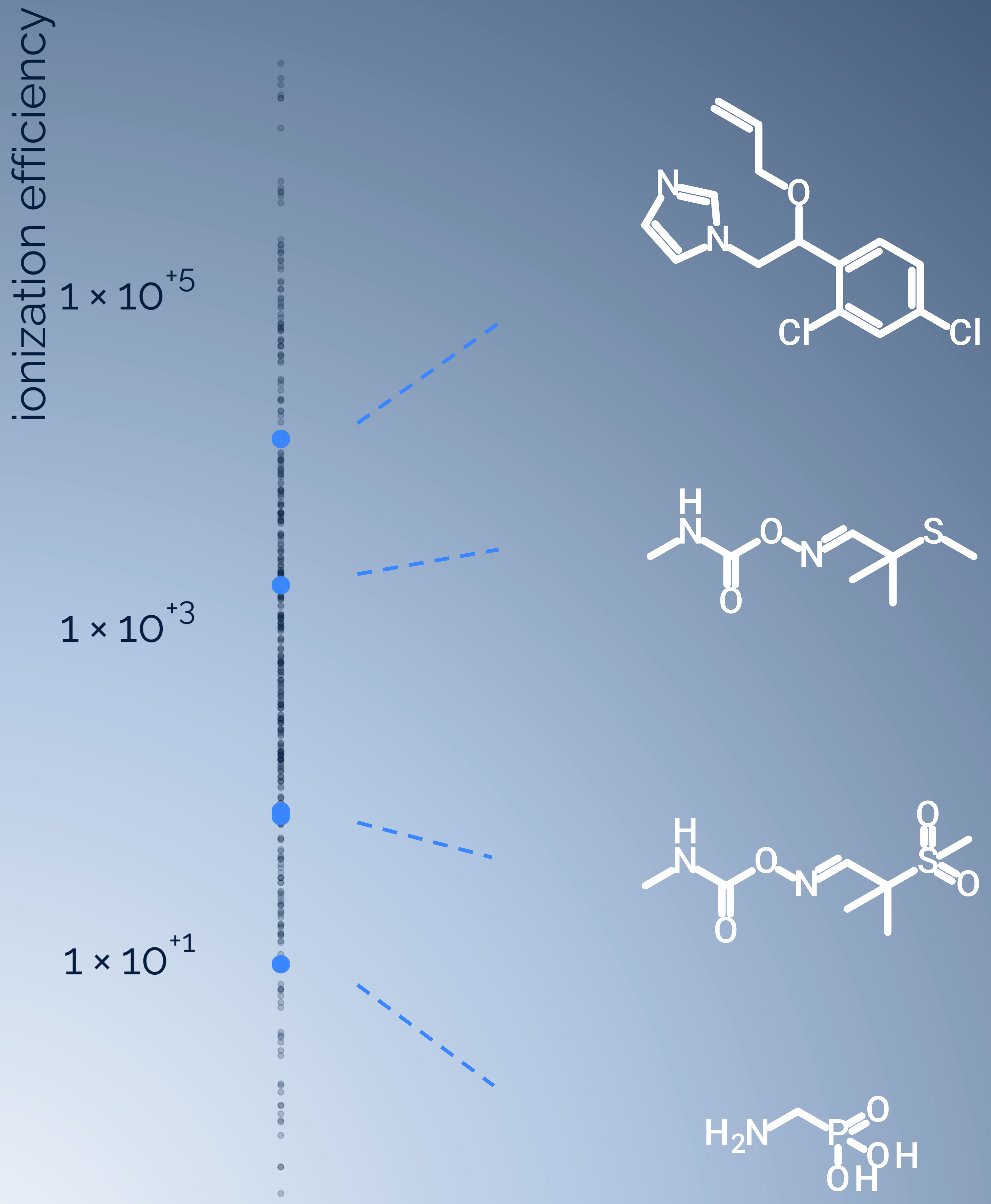
one solvent, purely analyte properties

377 chemicals

10,000,000x difference in *IE*

anneli kruve

anneli.kruve@su.se

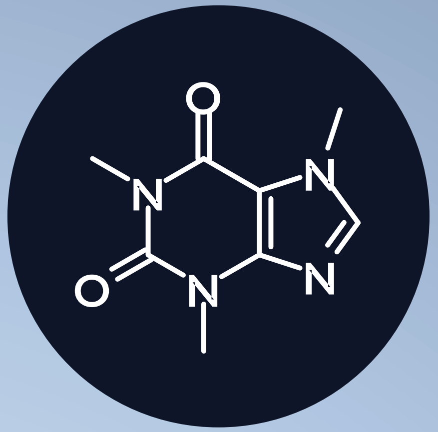


ionization efficiency

one solvent, purely analyte properties
377 chemicals
10,000,000x difference in *IE*

anneli kruve
anneli.kruve@su.se

ionization efficiency



structure

mobile phase

instrument

matrix

anneli kruve

anneli.kruve@su.se

quantification

with machine learning

quantification approaches

CLOSE ELUTING

calibration graph

STRUCTURALLY SIMILAR

calibration graphs

IONIZATION EFFICIENCY

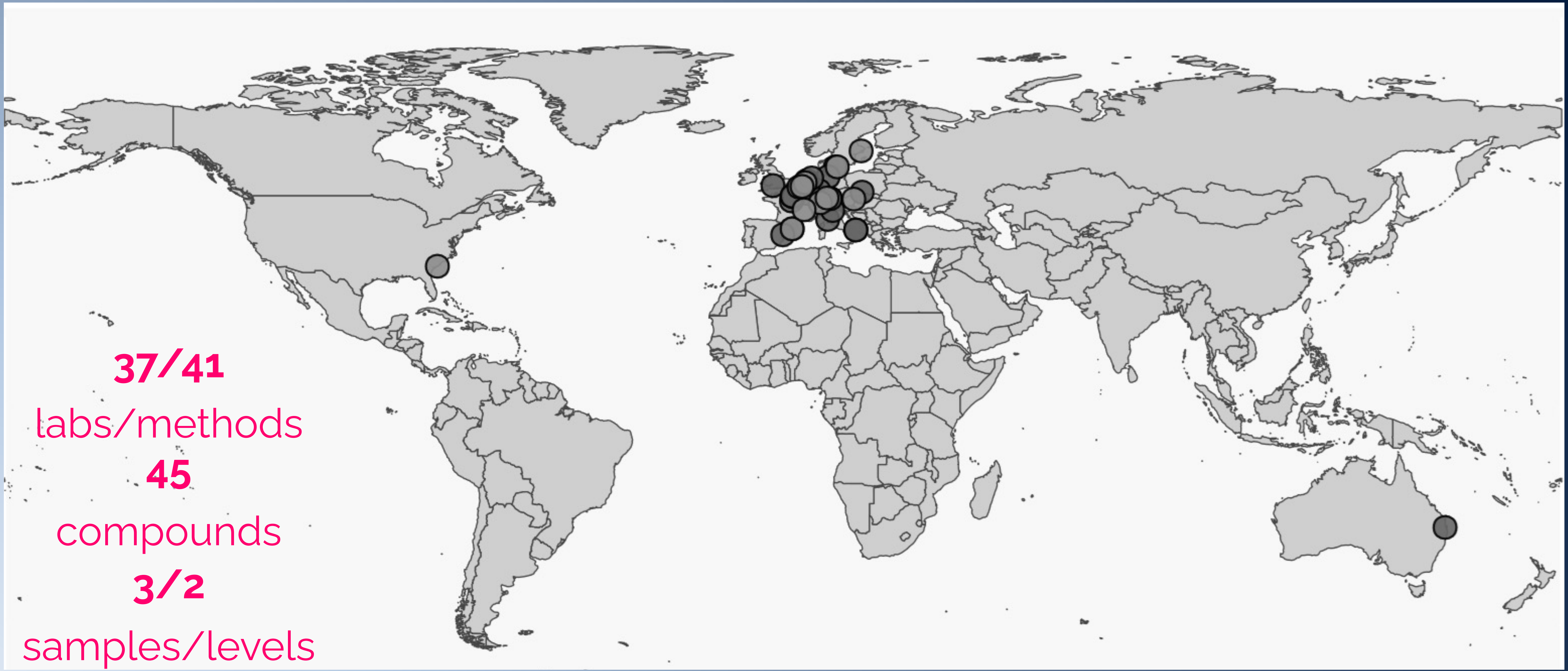
prediction model

anneli kruve

anneli.kruve@su.se

interlaboratory comparison

Malm et al. Anal Chem 2024

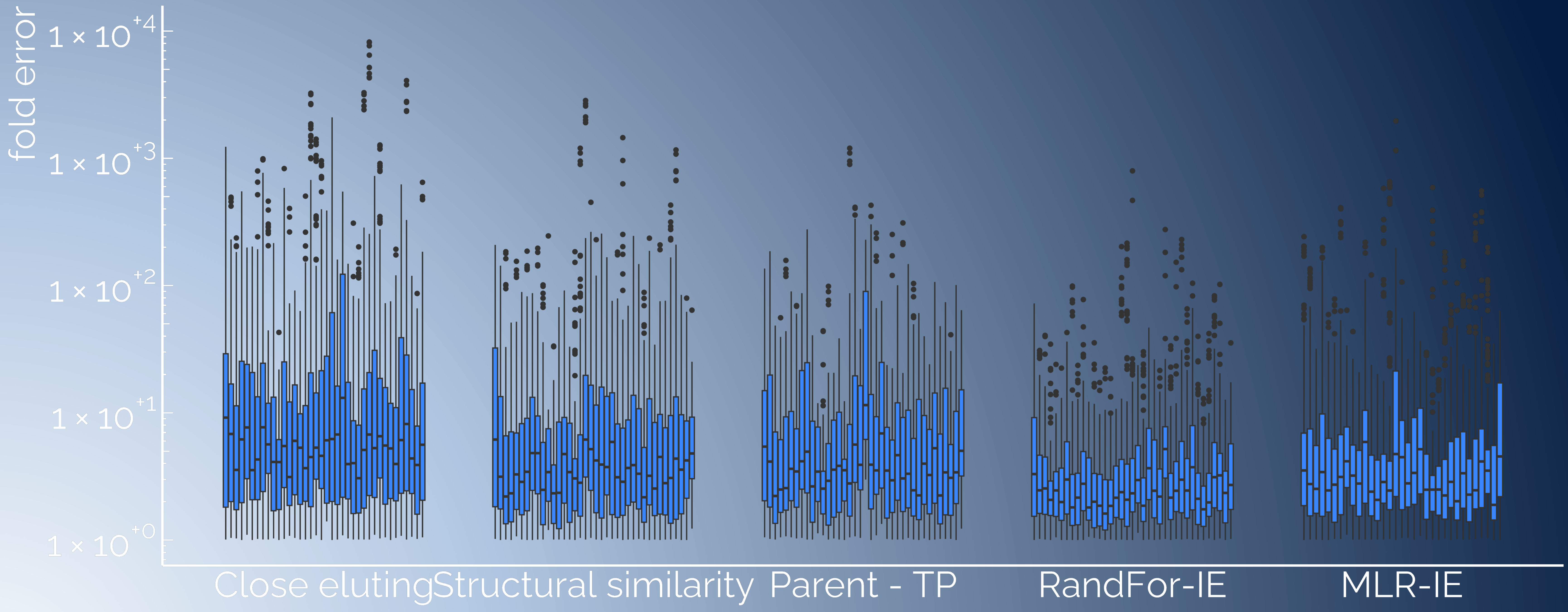


anneli kruve

anneli.kruve@su.se

interlaboratory comparison

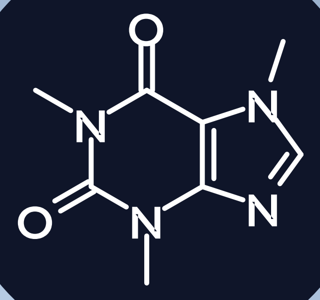
Malm et al. Anal Chem 2024



anneli kruve

anneli.kruve@su.se

workflow



SMILES or MS²

+ mobile phase composition

features

PaDEL, Mordred, SIRIUS fingerprints

machine learning

RandomForest, xgbTree

model selection

Root Mean Square Error (RMSE)

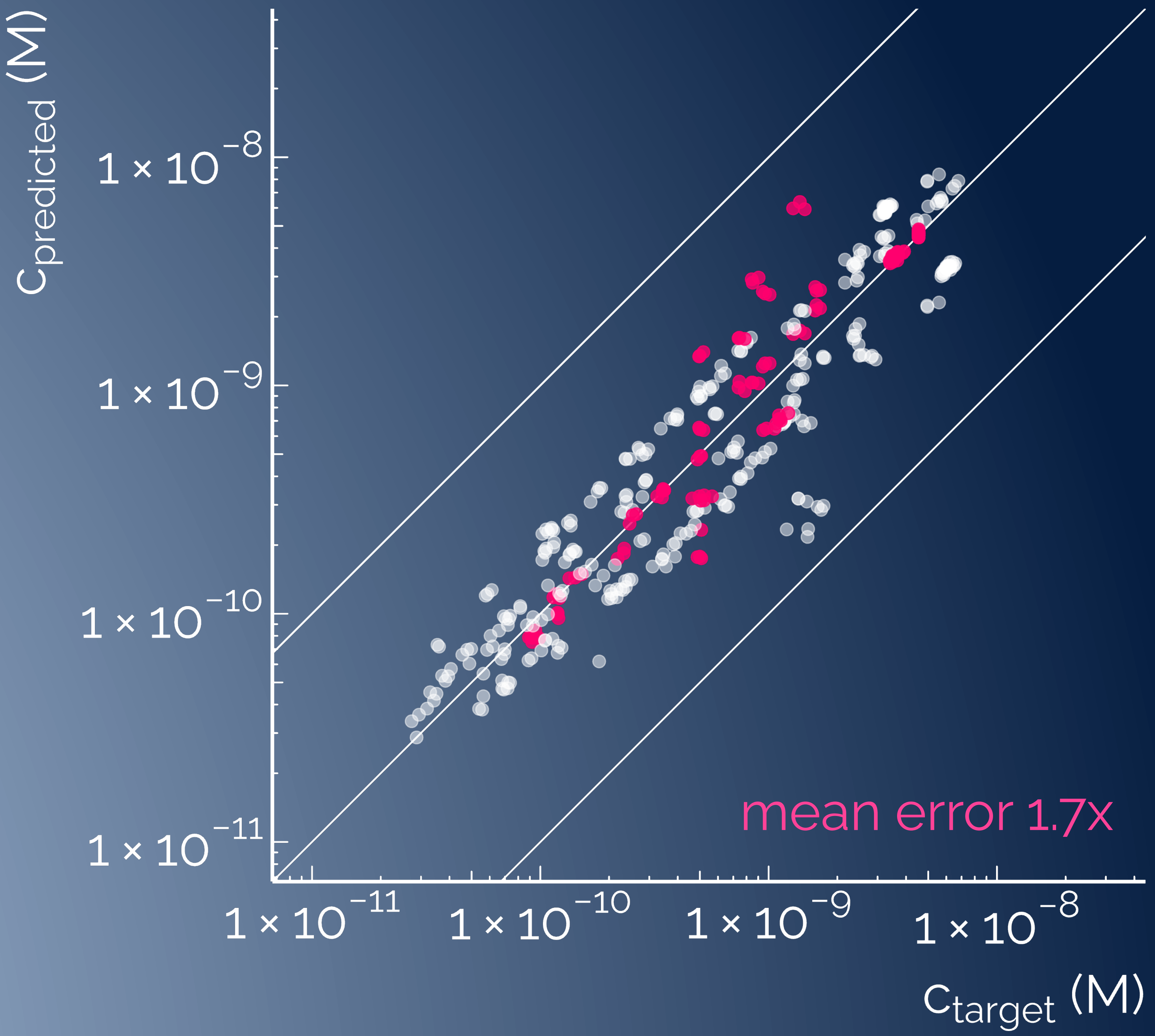
anneli kruve

anneli.kruve@su.se

quantification

from structure

Been et al. Water Research 2021

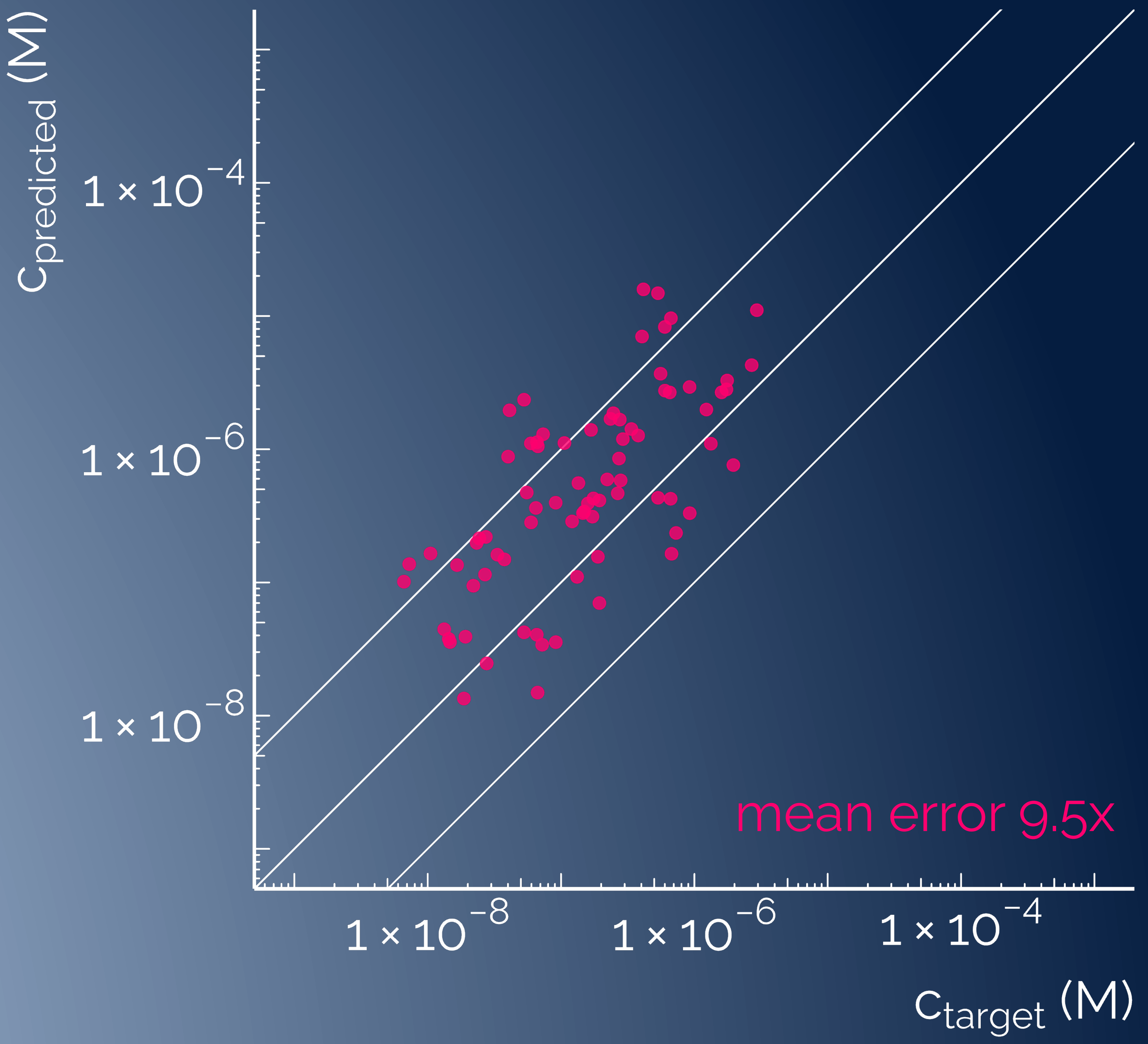


anneli.kruve@su.se

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

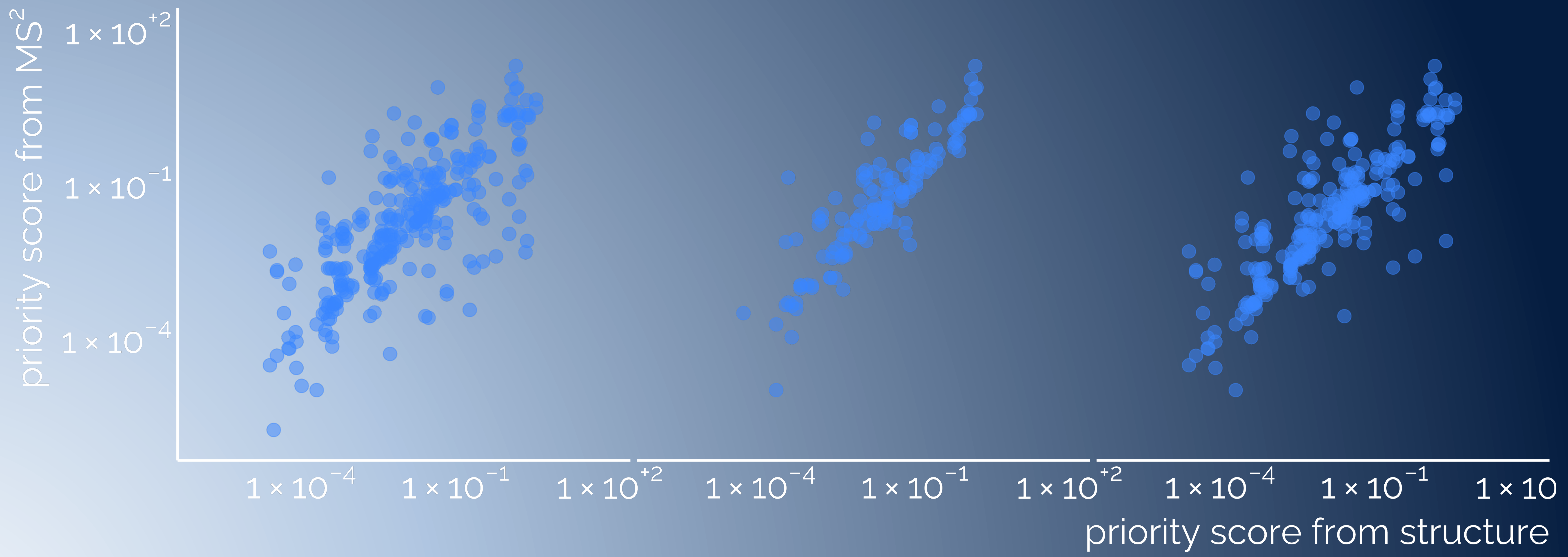
$$\text{PriorityScore} = \frac{c_{\text{predicted}}}{\text{AC}_{50}^{\text{5th percentile}}}$$

anneli kruve

anneli.kruve@su.se

interlaboratory comparison

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

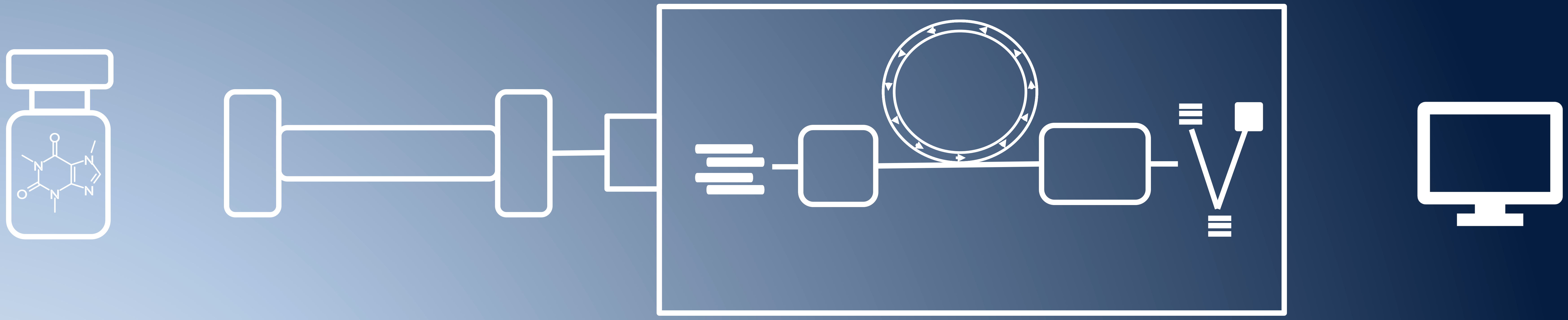
structure

anneli kruve

anneli.kruve@su.se

orthogonal separation

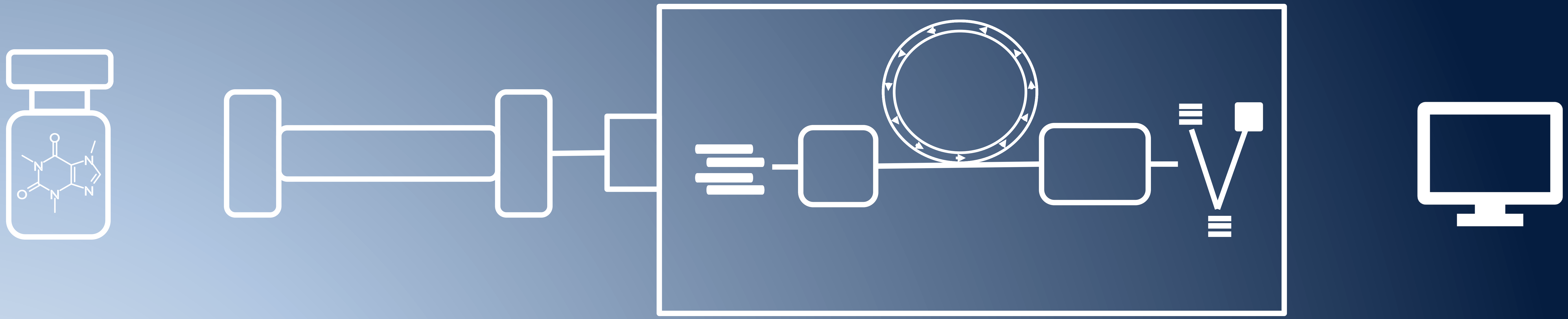
Akhlaqi et al. Anal Bioanal Chem 2023



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

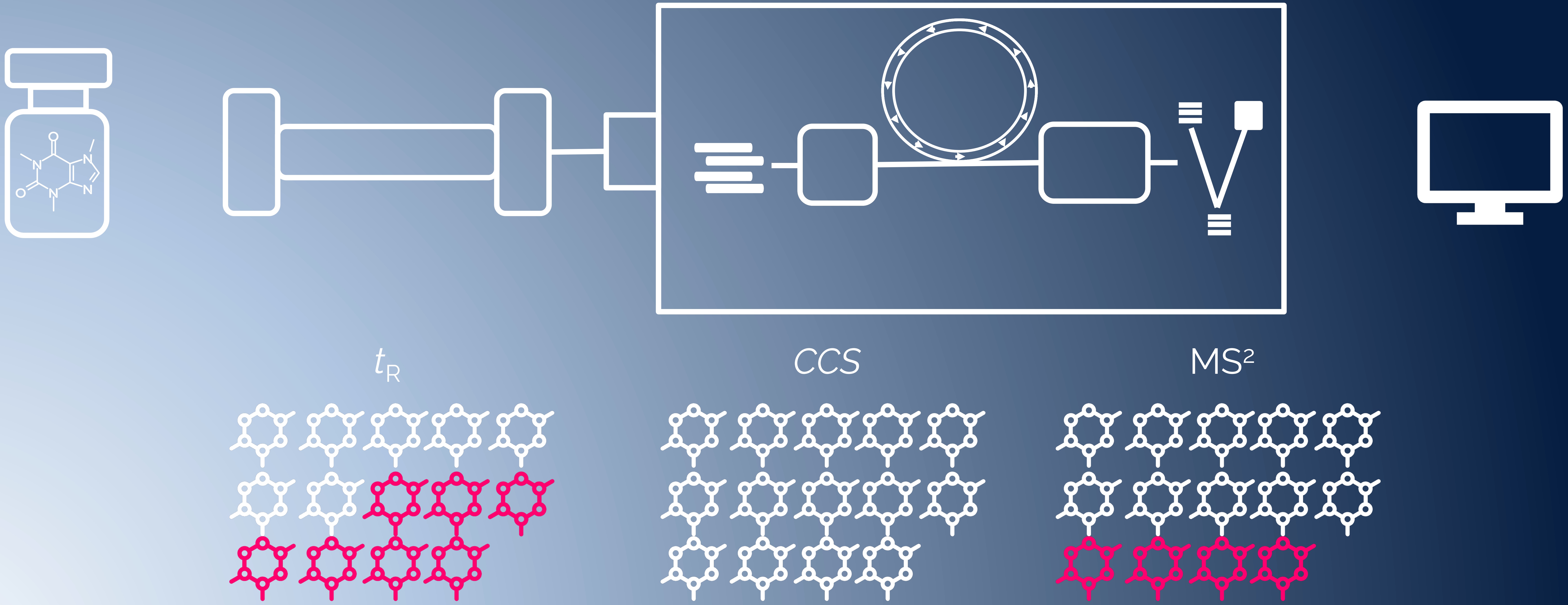
SIRIUS+ CSI:FingerID & CFM-ID

anneli kruve

anneli.kruve@su.se

orthogonal separation

Akhlaqi et al. Anal Bioanal Chem 2023



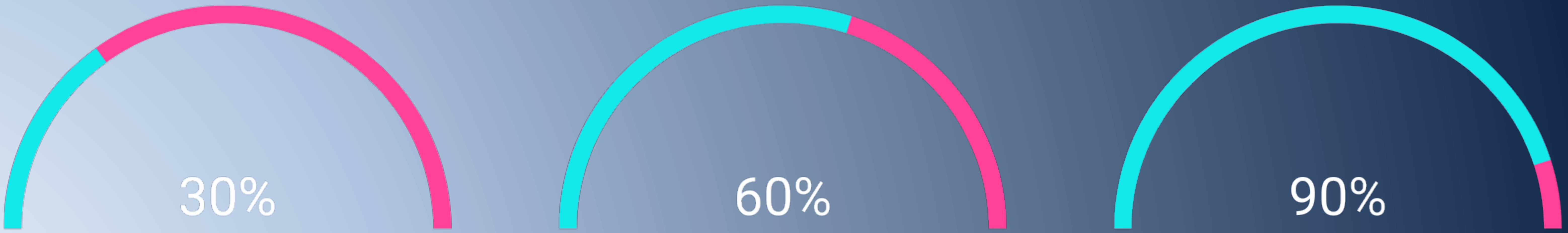
anneli kruve

anneli.kruve@su.se

machine learning

for prioritization and identification in nontarget screening

TOXICITY QUANTIFICATION RETENTION TIME



anneli kruve

anneli.kruve@su.se

erc

Swedish Research Council

FORMAS

WISE Wallenberg Initiative Materials Science for Sustainability

Wenner-Gren Foundations
Wenner-Gren Stiftelserna

CARL TRYGGERS STIFTELSE
FÖR VETENSKAPLIG FORSKNING

anneli kruve
anneli.kruve@su.se

Kruvelab.com

Exploring the research space...

anneli kruve
anneli.kruve@su.se