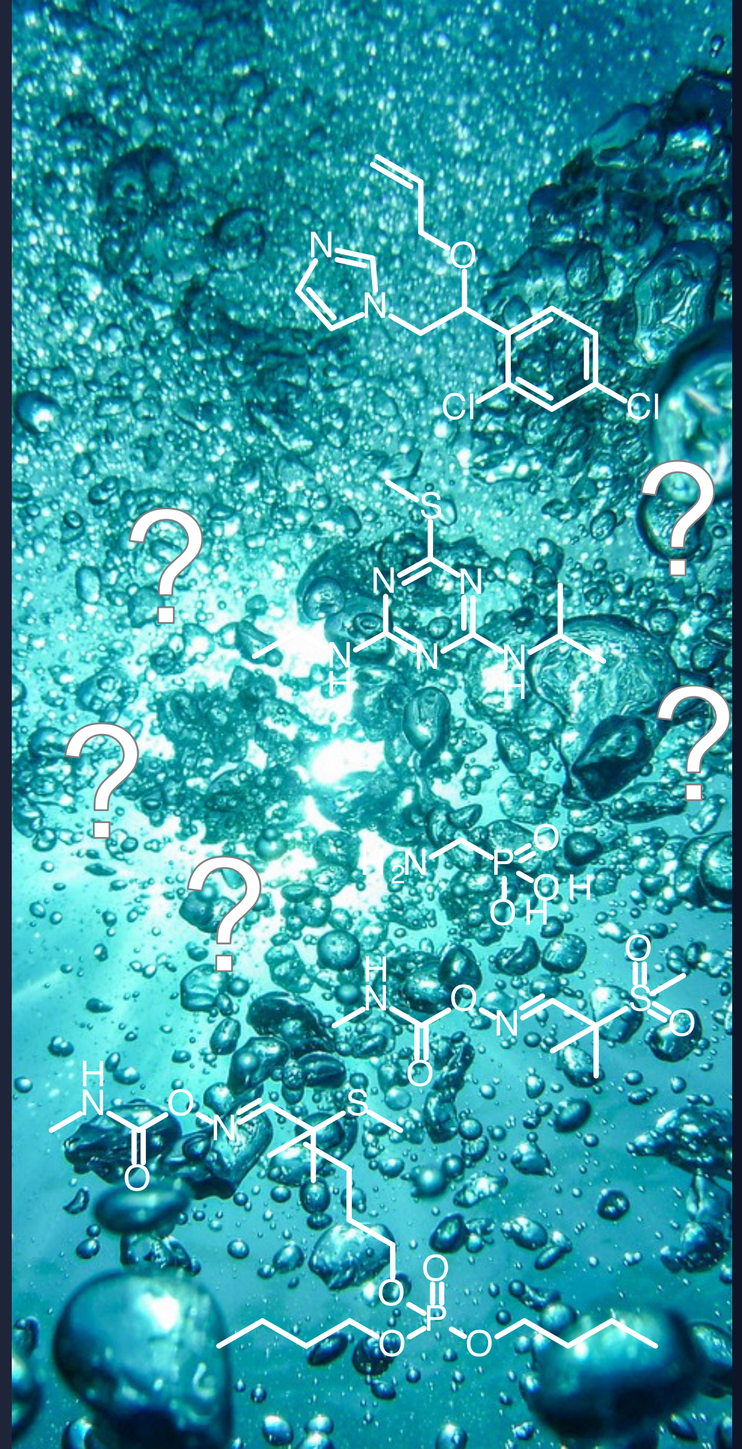
LC/HRMS: prioritizing & elucidating the structure of toxic chemicals

Anneli Kruve

Kruvelab.com

water analysis

thousands of chemicals
detected
in the environment



anneli kruve

anneli.kruve@su.se

how to ...

PRIORITIZE

risk

IDENTIFY

structure

anneli kruve

anneli.kruve@su.se

how to ...

PRIORITIZE

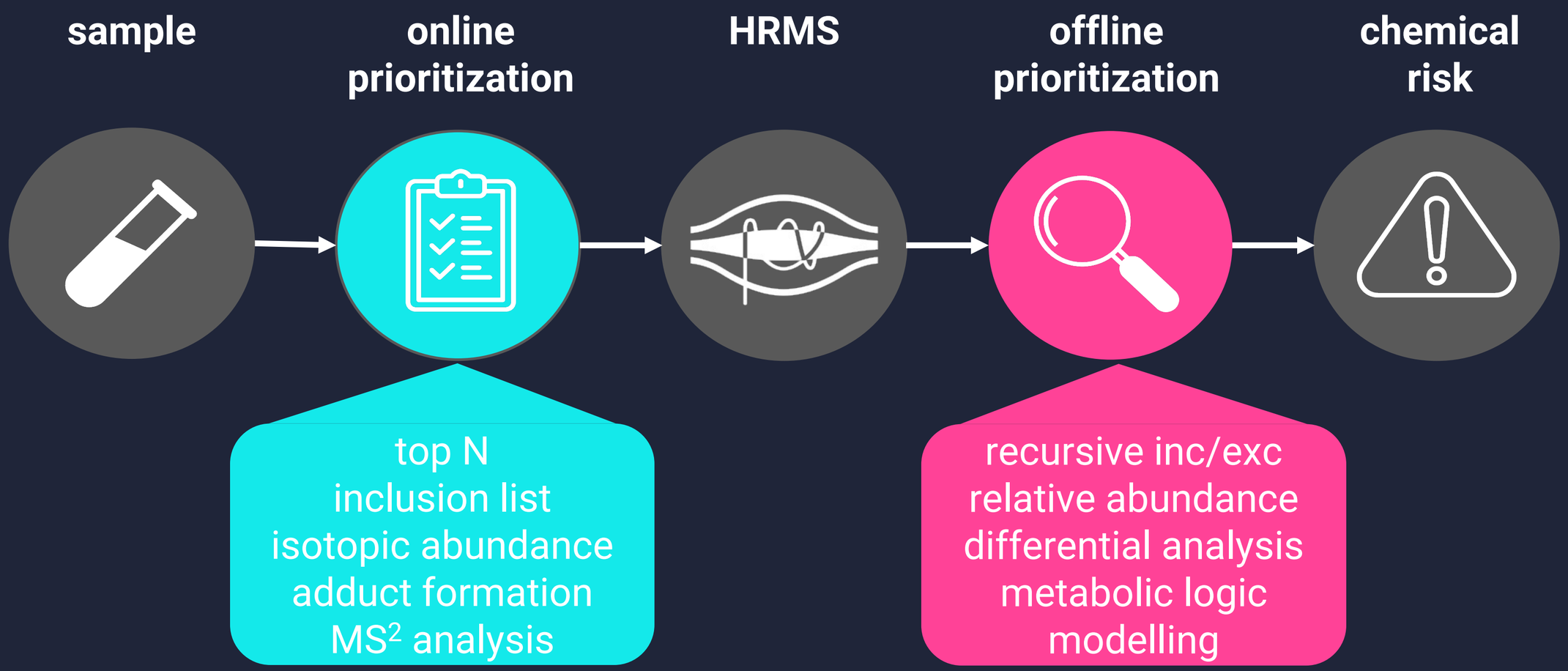
risk

anneli kruve

anneli.kruve@su.se

prioritization

Szabo et al. Anal Chem 2024



anneli kruve

anneli.kruve@su.se

prioritization of chemicals

TOXICITY

ecotoxicity and endocrine disruptors

CONCENTRATION

exposure to potentially toxic chemicals

RISK

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

anneli kruve

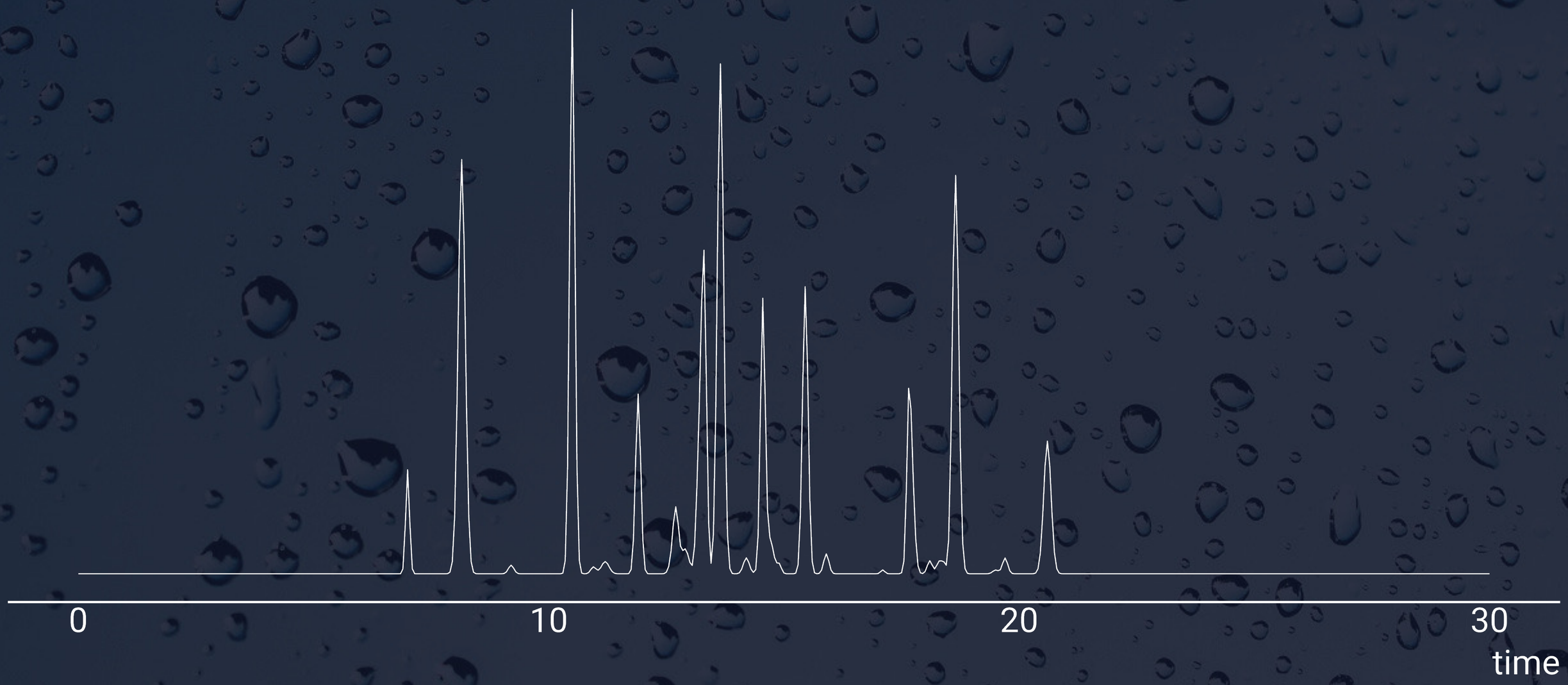
anneli.kruve@su.se

toxicity

of detected chemicals

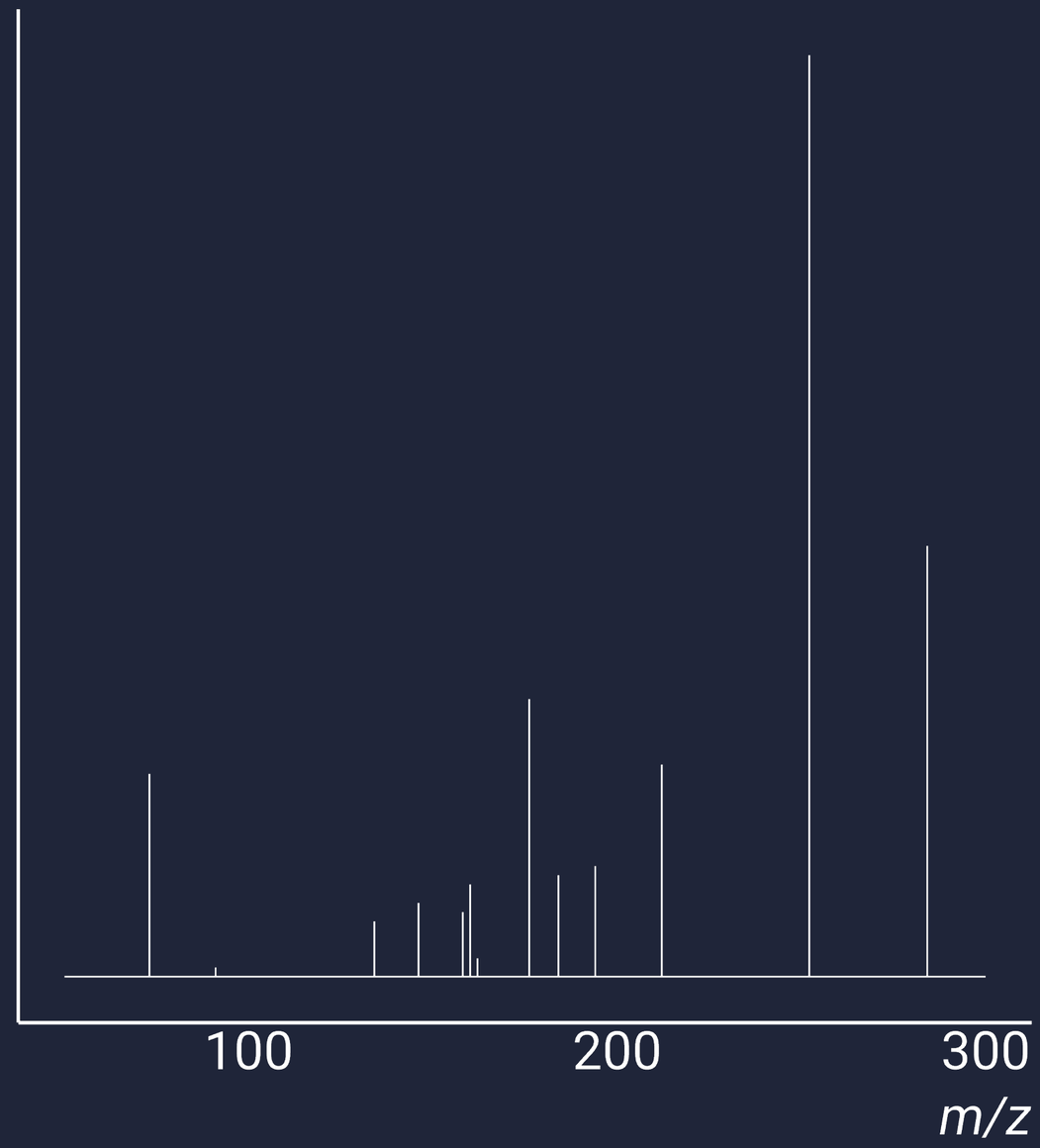
nontarget screening

with LC/HRMS



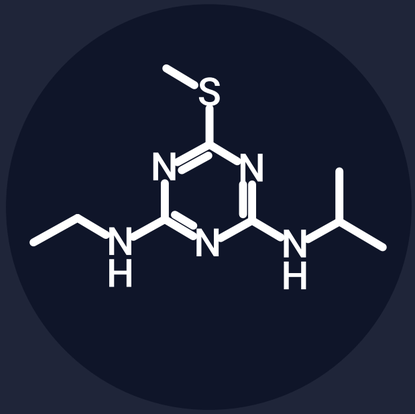
toxicity assessment

from spectra
to structure
to toxicity



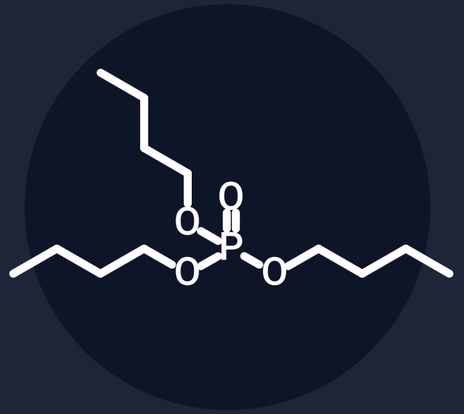
anneli kruve
anneli.kruve@su.se

toxicity assessment



LC_{50} = 9.3 mg/L

known structure
known toxicity



LC_{50} = ? mg/L

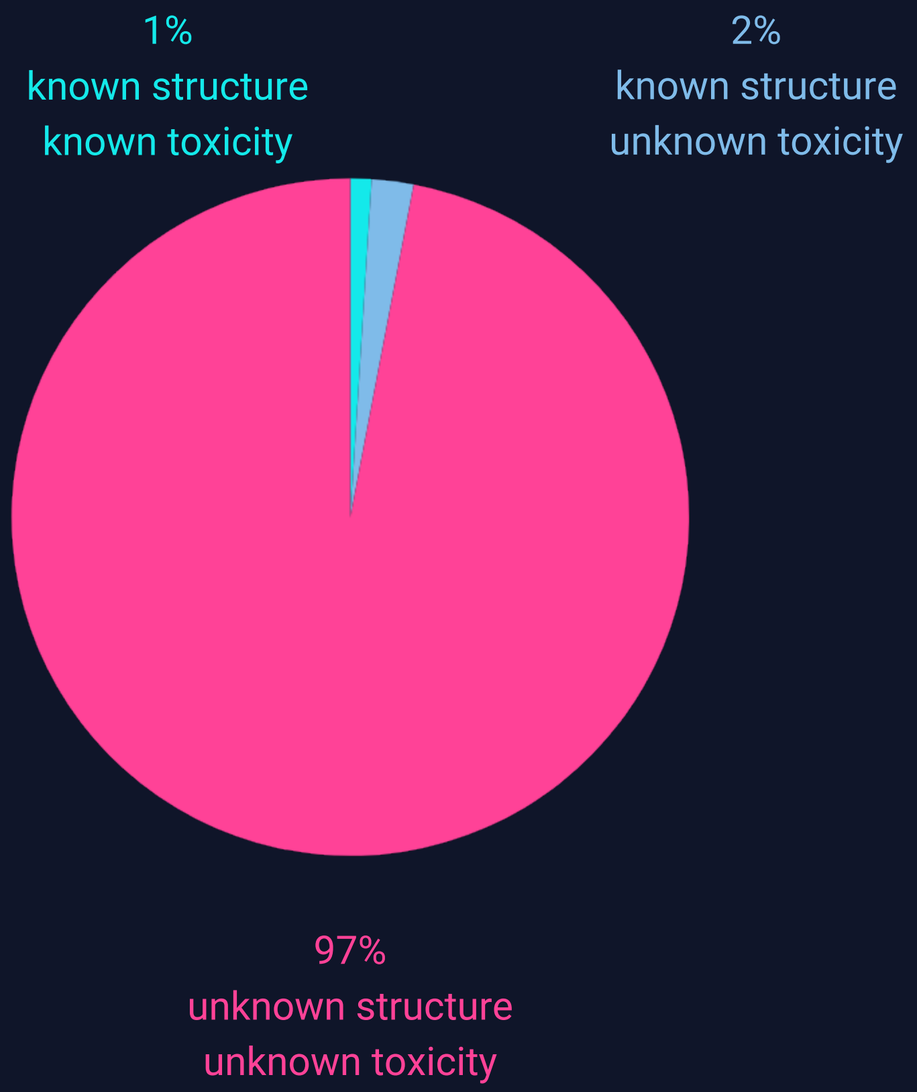
known structure
unknown toxicity

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

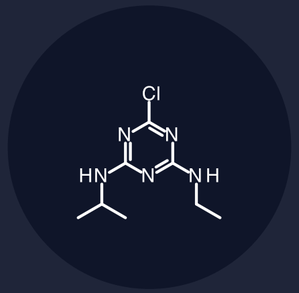
AhR, AR, ER, MMP, P53, ...

anneli kruve

anneli.kruve@su.se

workflow

MS^2 SPECTRA



STRUCTURE AS SMILES

MOLECULAR DESCRIPTORS

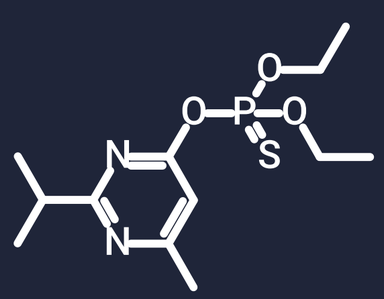
PREDICT TOXICITY

anneli kruve

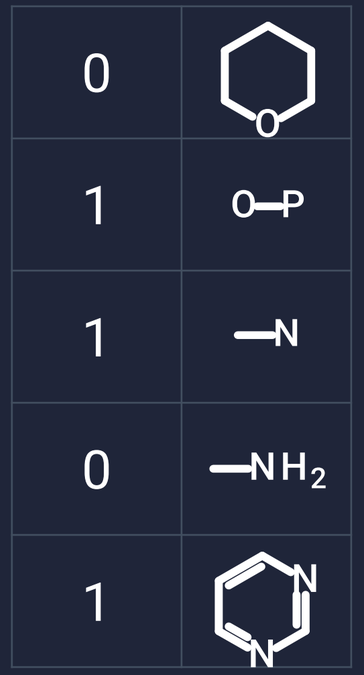
anneli.kruve@su.se

structural fingerprints

Peets et al. ES&T 2022



R:
rcdk



anneli kruve

anneli.kruve@su.se

model training

Peets et al. ES&T 2022

mass (Da)	fp1	...	fp243
317.32000	0	...	0
208.26100	1	...	0
240.21499	1	...	0
300.57998	0	...	0
201.22500	0	...	0

anneli.kruve@su.se

model training

Peets et al. ES&T 2022

mass (Da)	fp1	...	fp243
317.32000	0	...	0
208.26100	1	...	0
240.21499	1	...	0
300.57998	0	...	0
201.22500	0	...	0

training set
517
chemicals

test set
130
chemicals

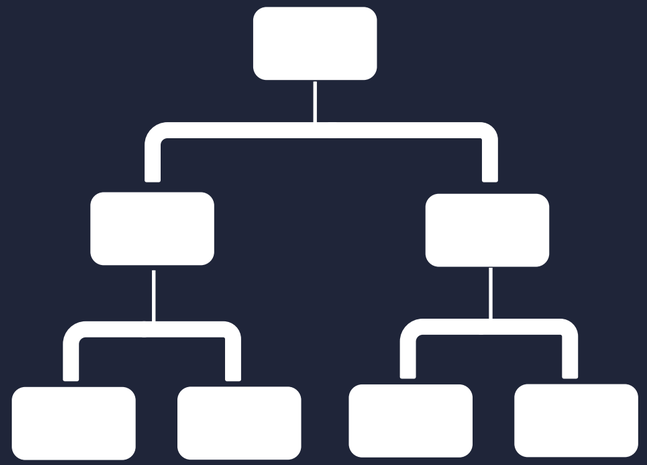
anneli.kruve@su.se

model training

Peets et al. ES&T 2022

mass (Da)	fp1	...	fp243
317.32000	0	...	0
208.26100	1	...	0
240.21499	1	...	0
300.57998	0	...	0
201.22500	0	...	0

gradient boosting

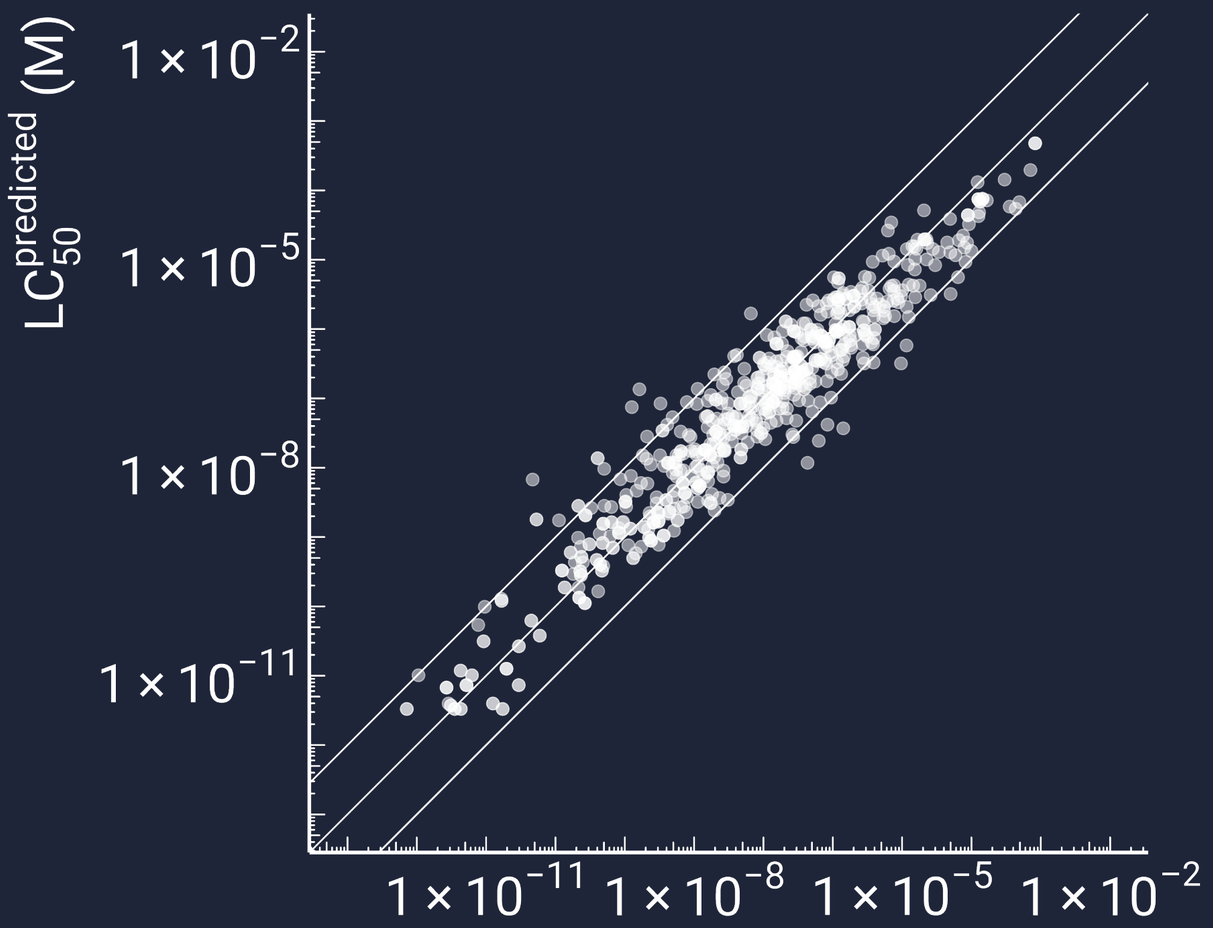


anneli kruve

anneli.kruve@su.se

LC_{50} predictions

Peets et al. ES&T 2022
fish LC_{50}



training set

RMSE 0.52 log(M)

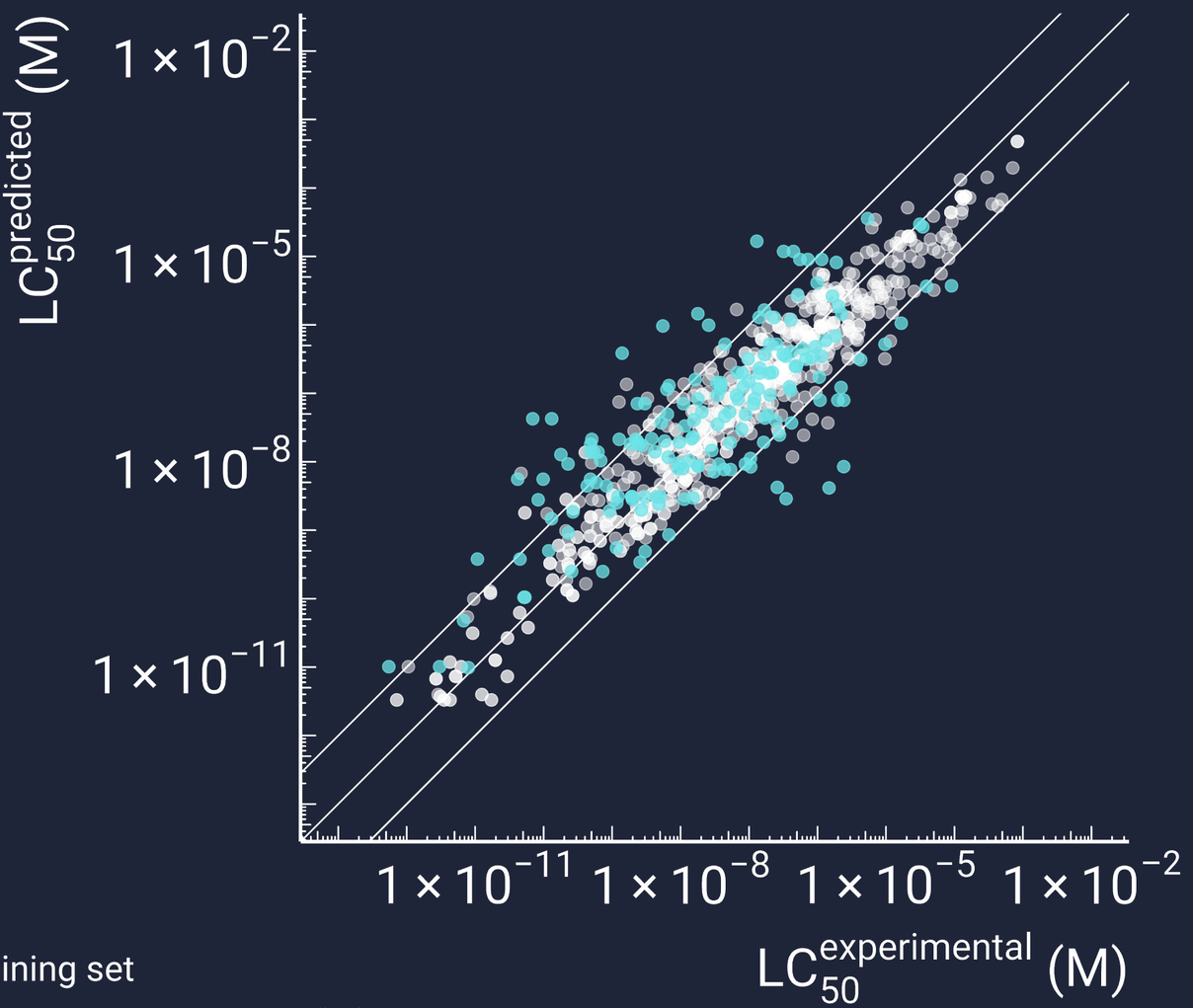
anneli kruve

anneli.kruve@su.se

LC_{50} predictions

Peets et al. ES&T 2022

fish LC_{50}



training set

RMSE 0.52 log(M)

test set

RMSE 0.78 log(M)

anneli kruve

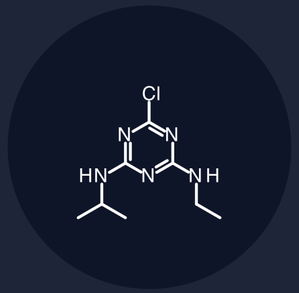
anneli.kruve@su.se

predicting toxicity

of unidentified chemicals

workflow

MS^2 SPECTRA



STRUCTURE AS SMILES

MOLECULAR DESCRIPTORS

PREDICT TOXICITY

anneli kruve

anneli.kruve@su.se

workflow

MS^2 SPECTRA

MOLECULAR DESCRIPTORS

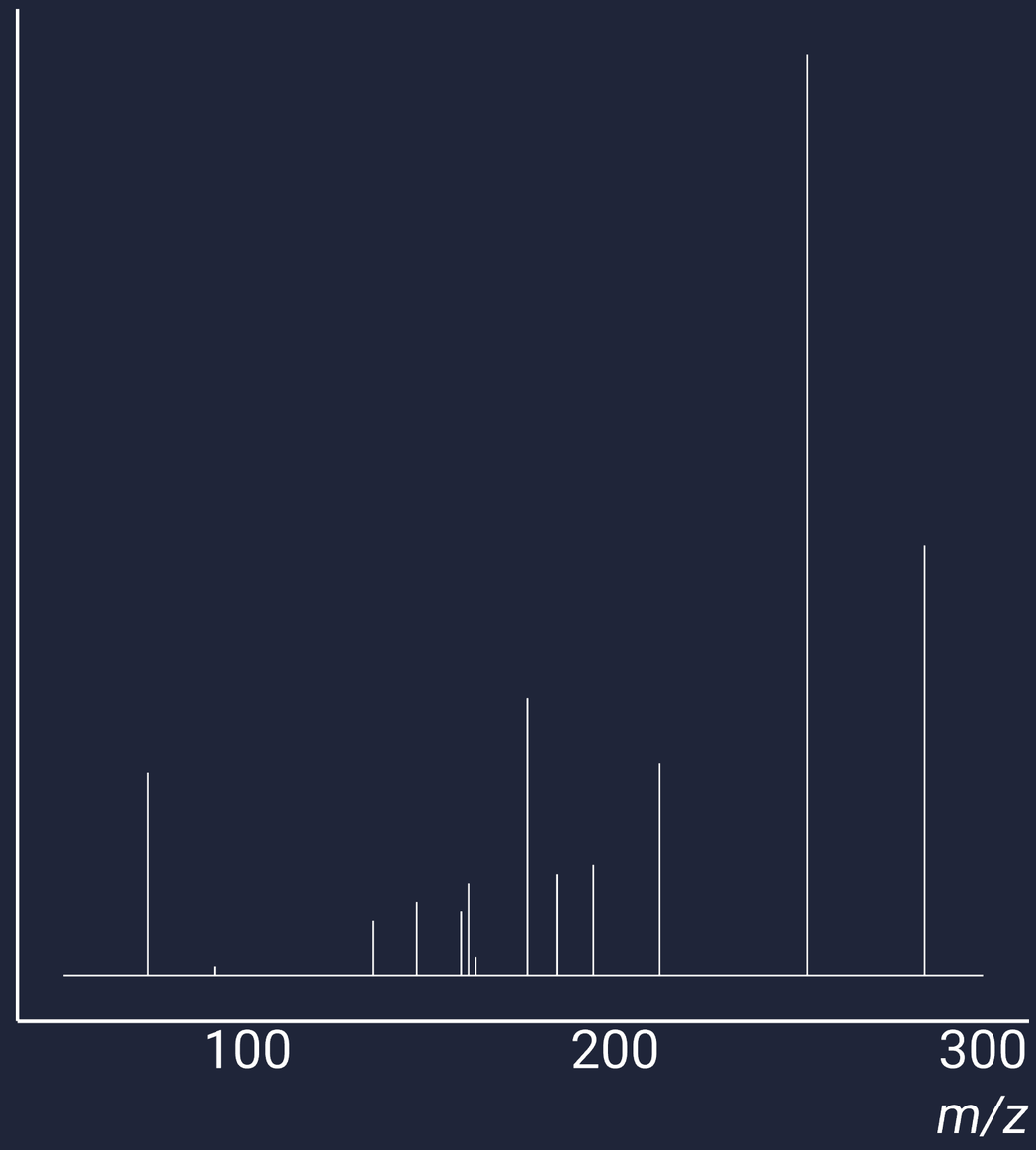
PREDICT TOXICITY

anneli kruve

anneli.kruve@su.se

information available

in MS^2 spectra

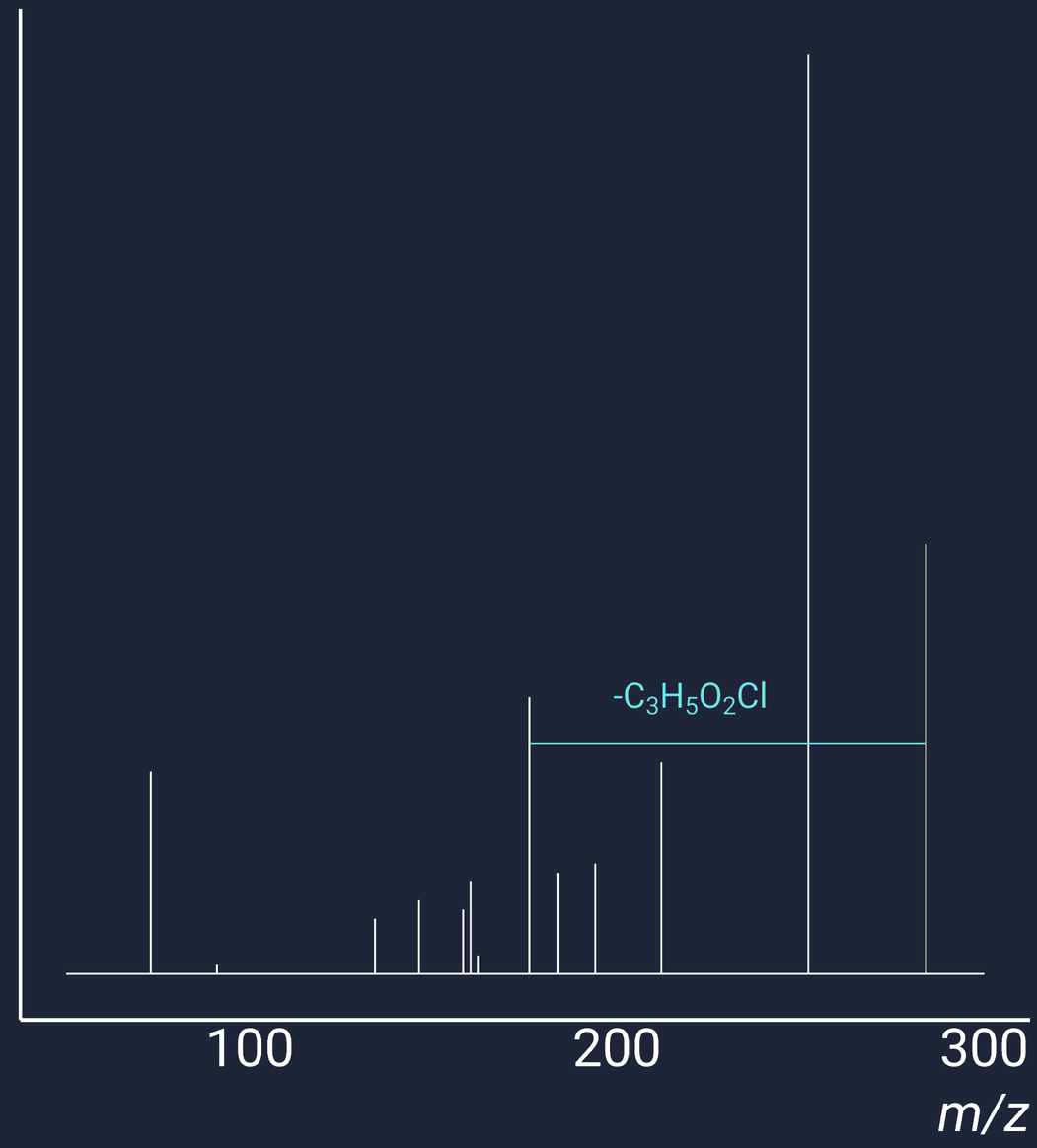


anneli kruve

anneli.kruve@su.se

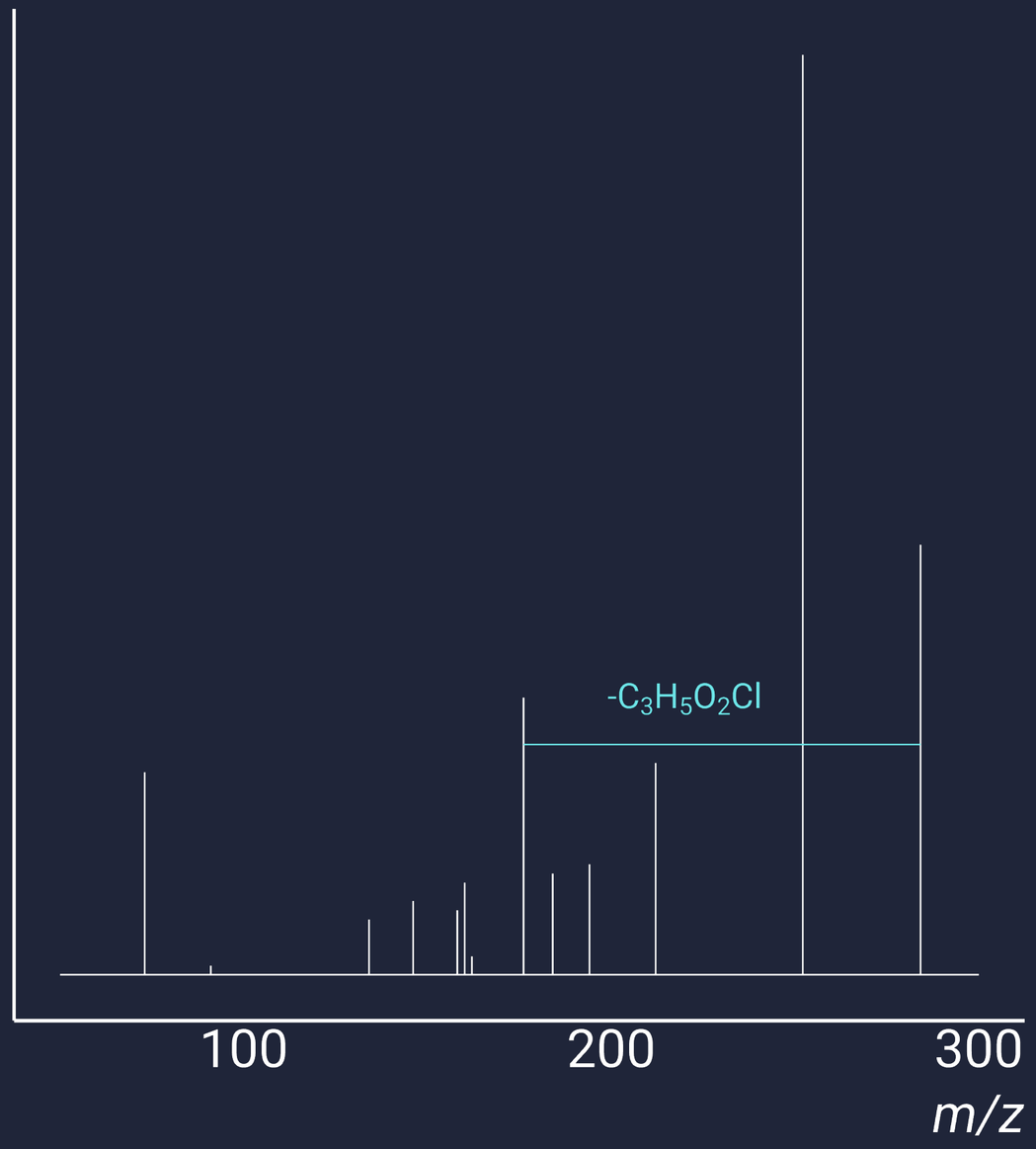
information available

in MS^2 spectra



anneli kruve

anneli.kruve@su.se

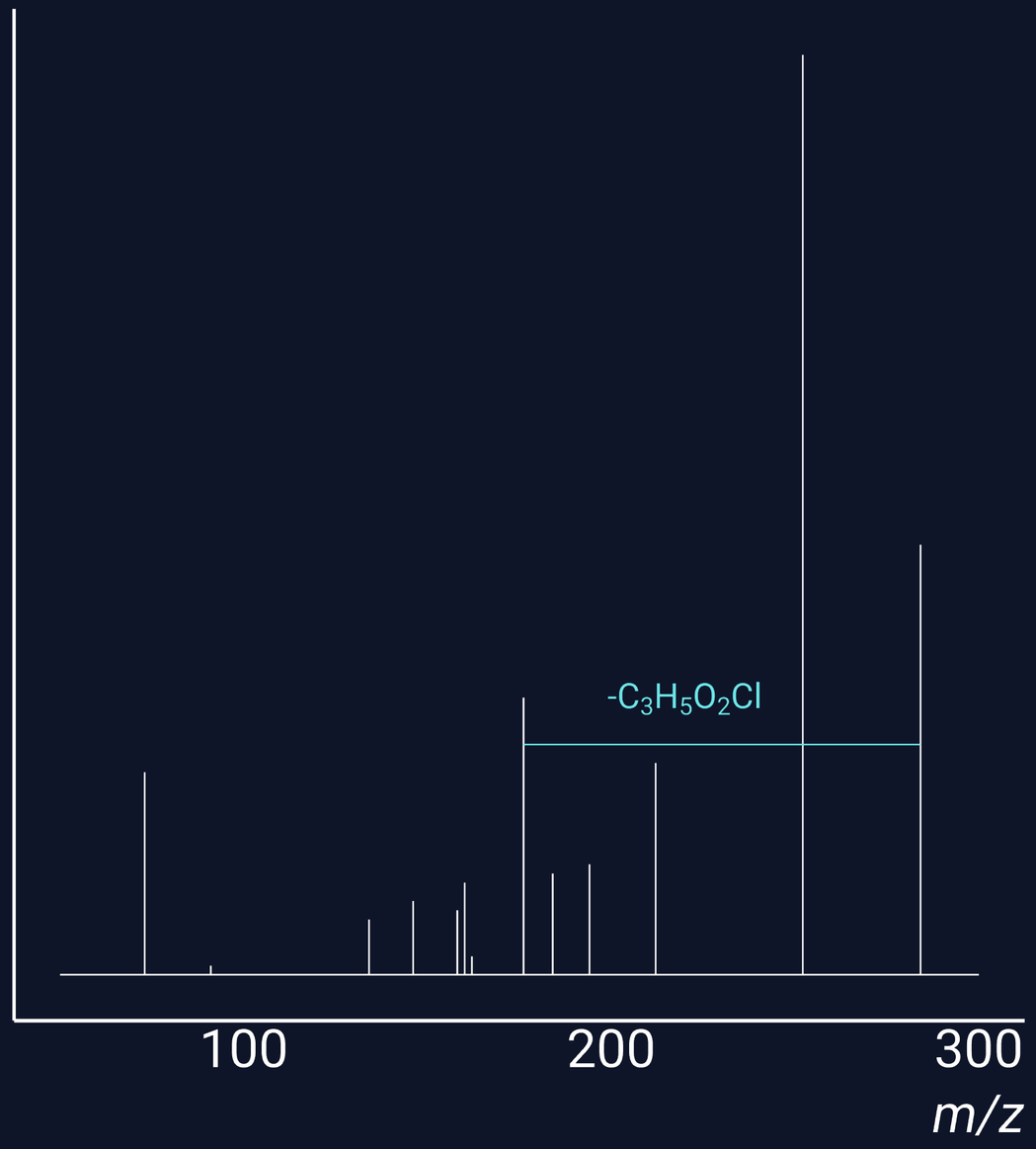


predict toxicity

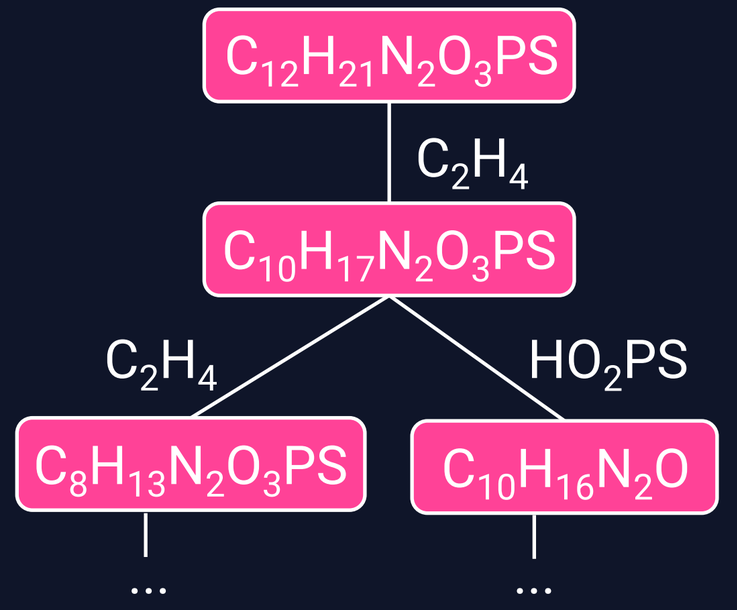
for unknown chemicals

anneli kruve

anneli.kruve@su.se

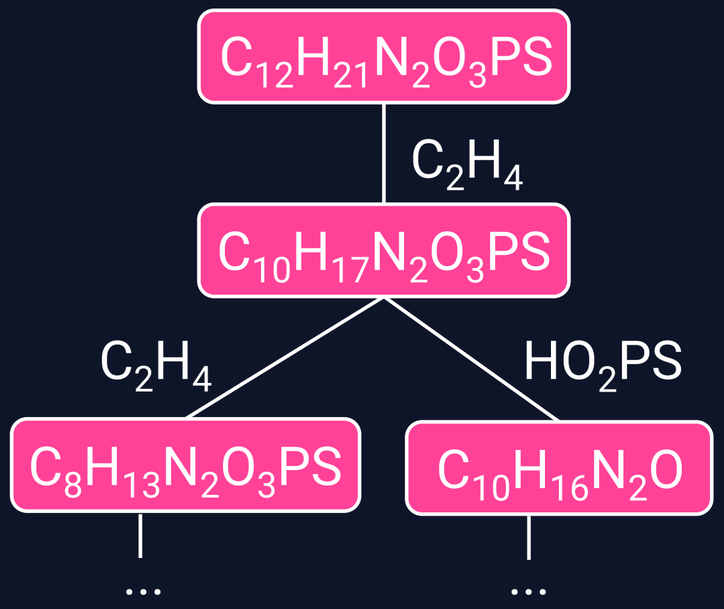


SIRIUS+
CSI:FingerID



anneli kruve

anneli.kruve@su.se



SIRIUS+
CSI:FingerID

0	
1	
1	
0	
1	

gradient
boosting

LC_{50} =
−2.2 log(mM)

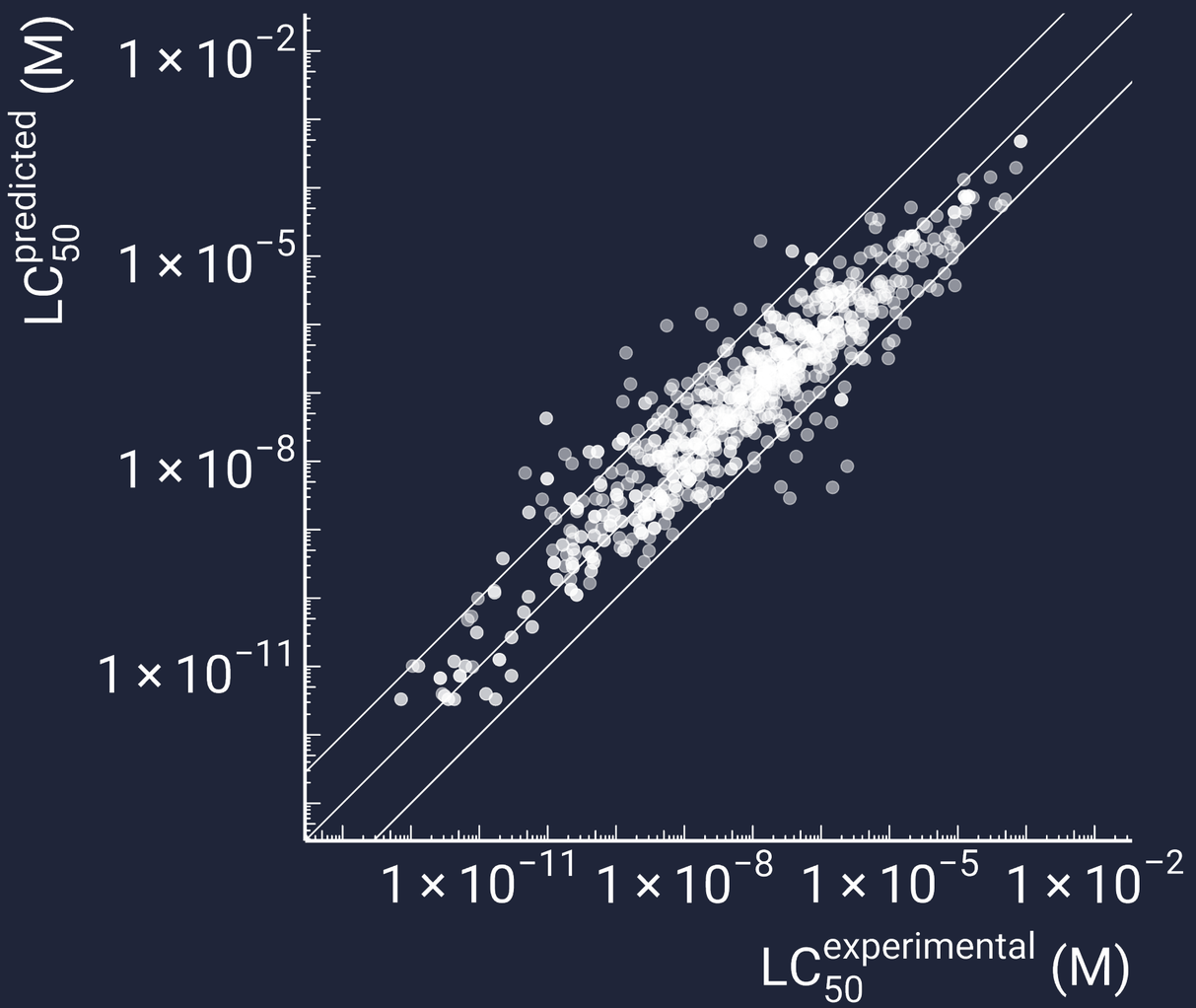
anneli kruve

anneli.kruve@su.se

LC_{50} predictions

Peets et al. ES&T 2022

fish LC_{50}



test set

RMSE 0.78 log(M)

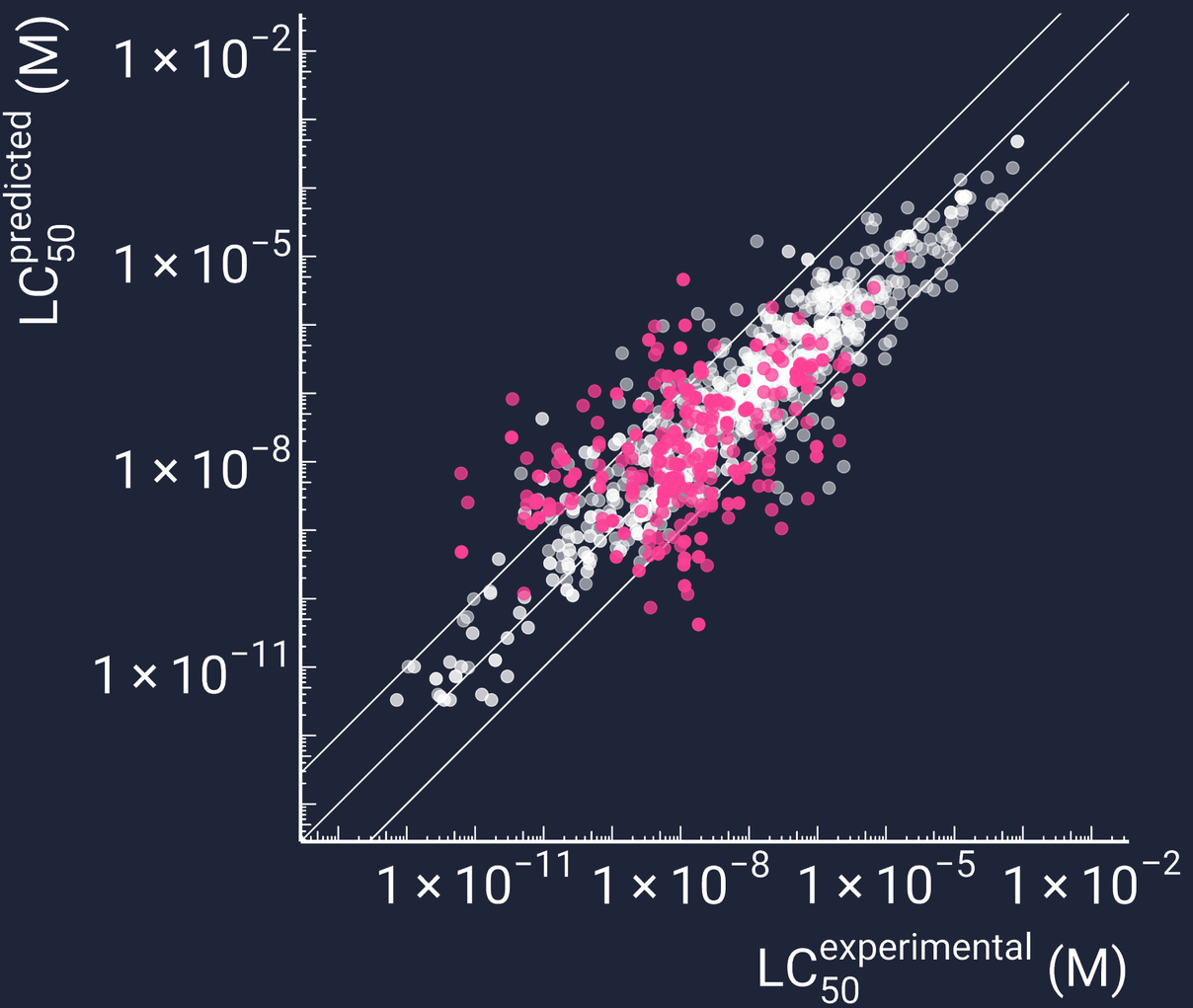
anneli kruve

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

anneli.kruve@su.se

endocrine disruption

Rahu et al. ES&T 2024
Tox21 endpoints

anneli kruve

anneli.kruve@su.se

endocrine disruption

Rahu et al. ES&T 2024
Tox21 endpoints

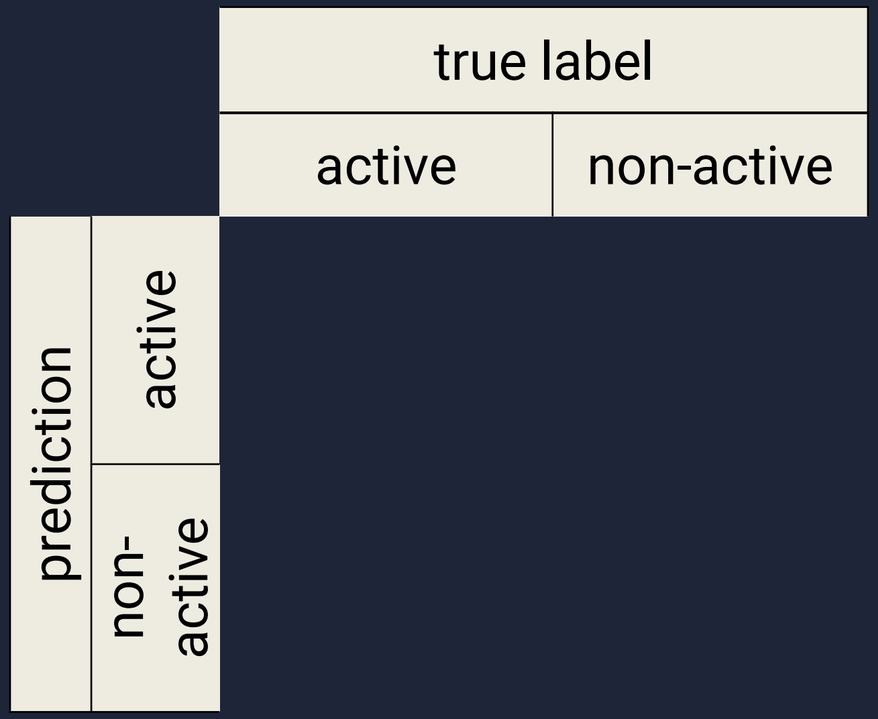
true label	
active	non-active

anneli kruve

anneli.kruve@su.se

endocrine disruption

Rahu et al. ES&T 2024
Tox21 endpoints



anneli kruve
anneli.kruve@su.se

endocrine disruption

Rahu et al. ES&T 2024
Tox21 endpoints

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

which is more dramatic:

type I error
type II error

anneli kruve

anneli.kruve@su.se

endocrine disruption

Rahu et al. ES&T 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. ES&T 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 with MS^2 & tox

anneli kruve

anneli.kruve@su.se

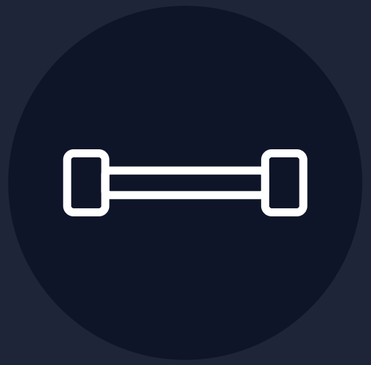
case study: interlaboratory comparison

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

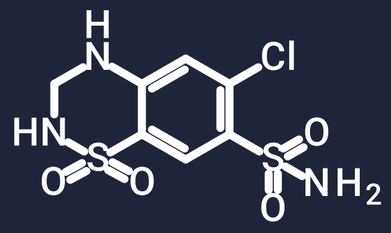
features predicted active

anneli kruve

anneli.kruve@su.se

AhR active

hydrochlorothiazide



anneli kruve

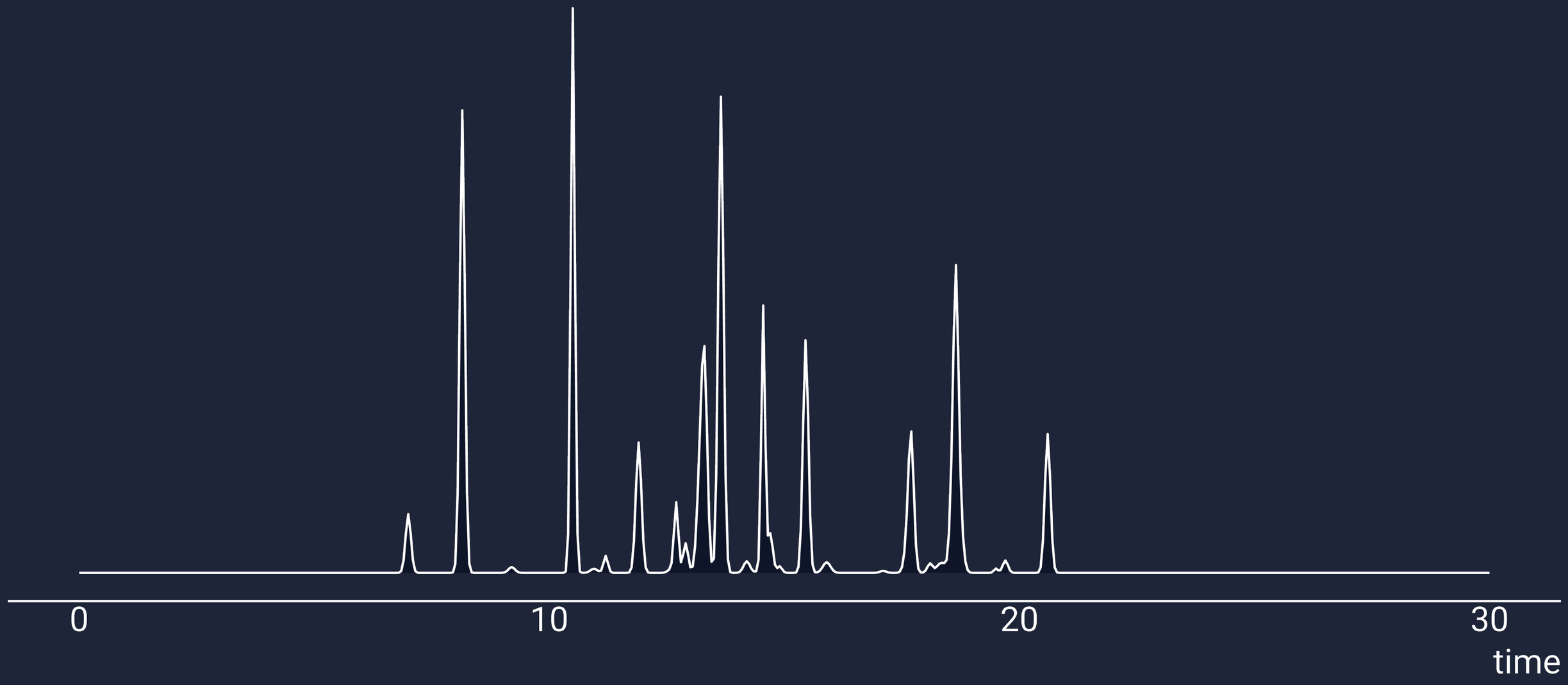
anneli.kruve@su.se

quantification

of detected chemicals

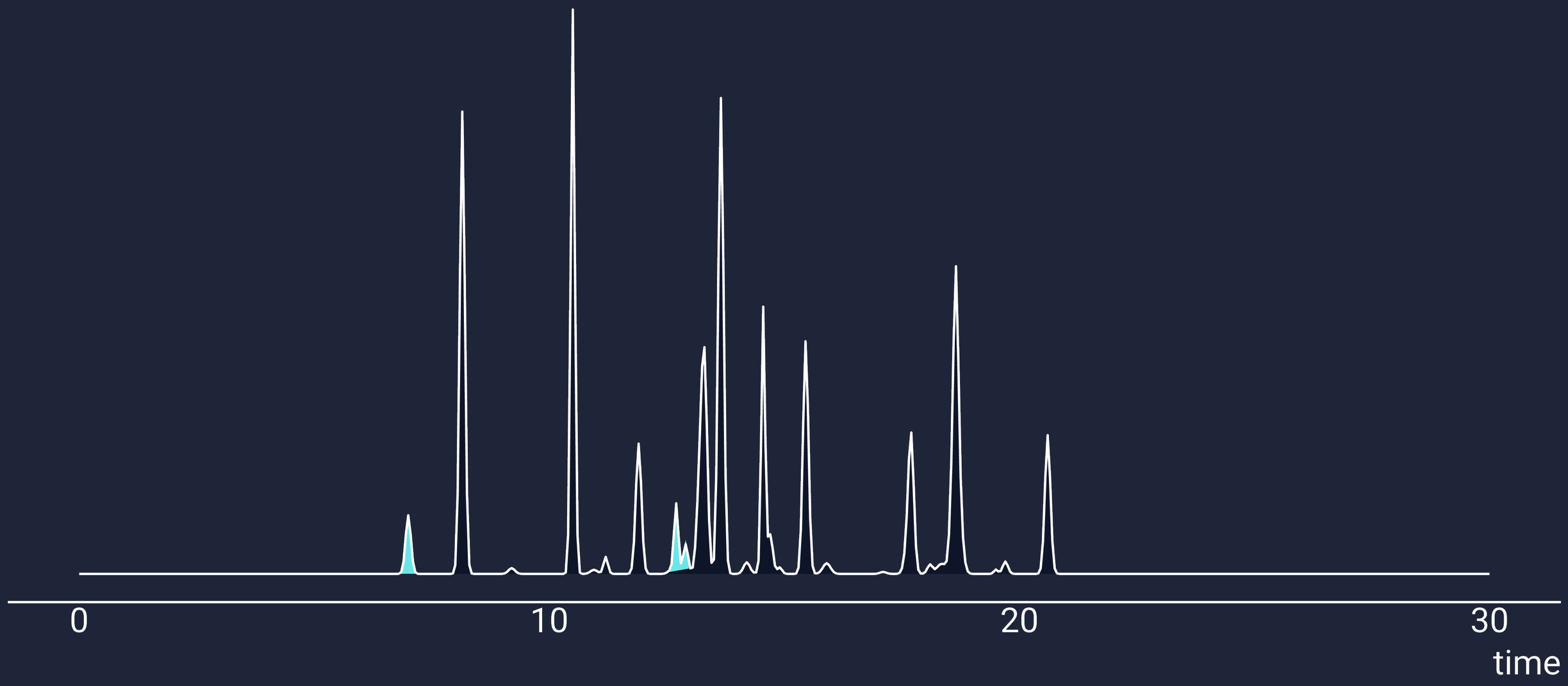
quantification in ESI/HRMS

Malm et al. Molecules 2021



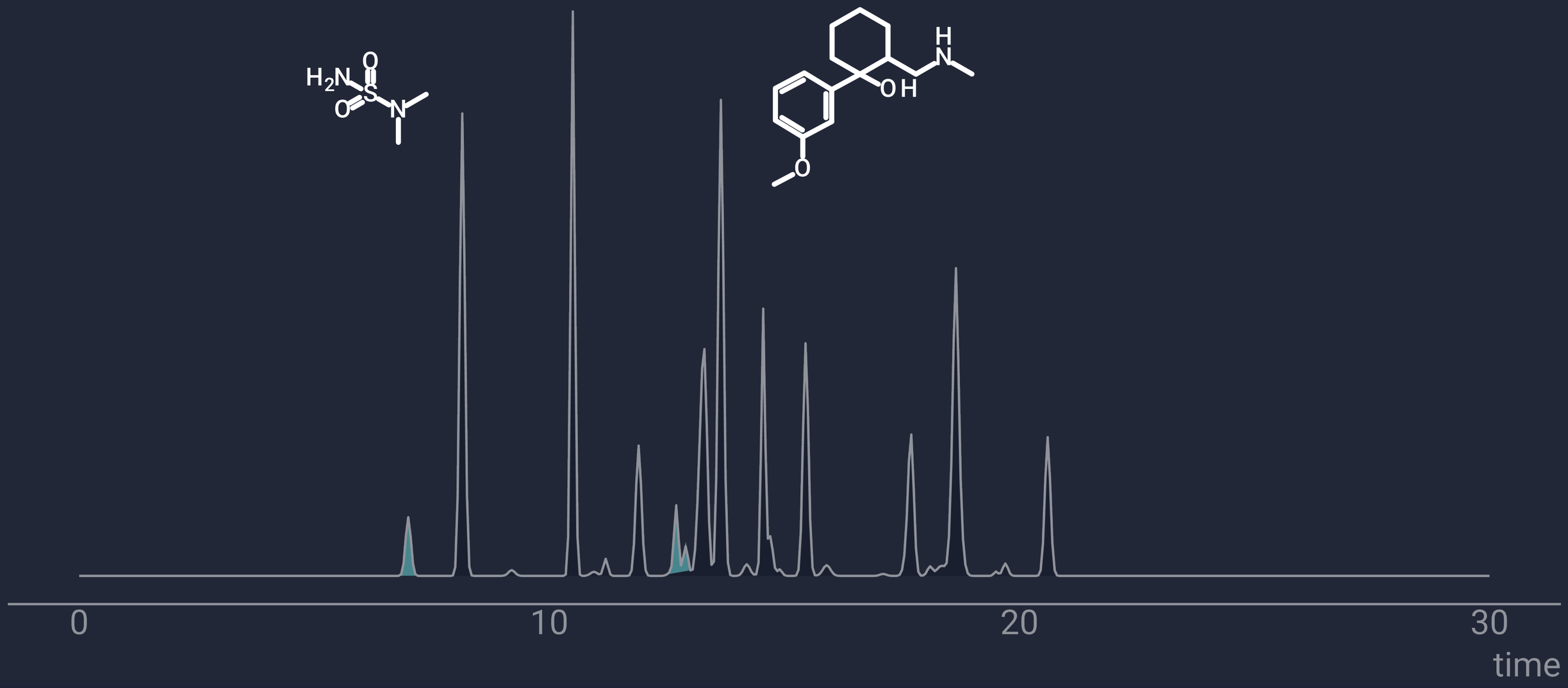
quantification in ESI/HRMS

Malm et al. Molecules 2021



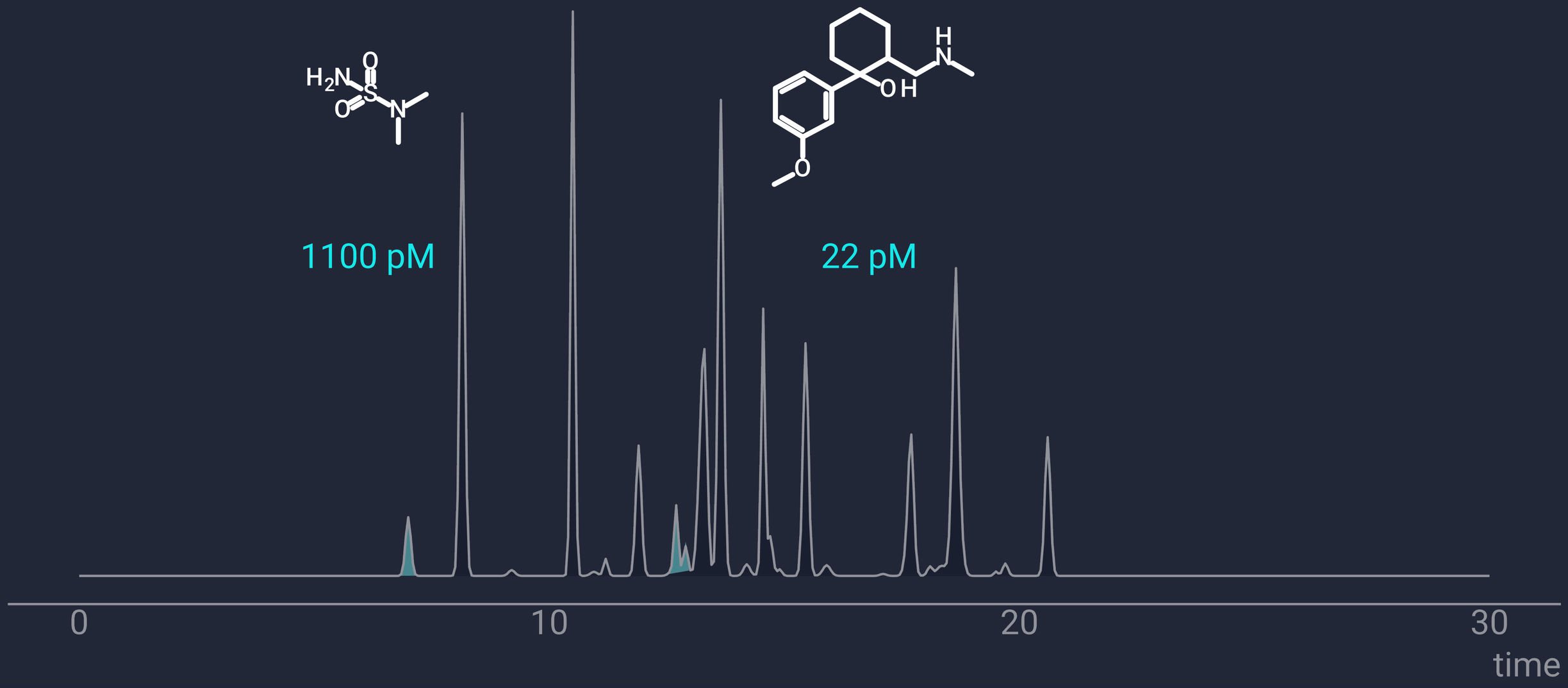
quantification in ESI/HRMS

Malm et al. Molecules 2021

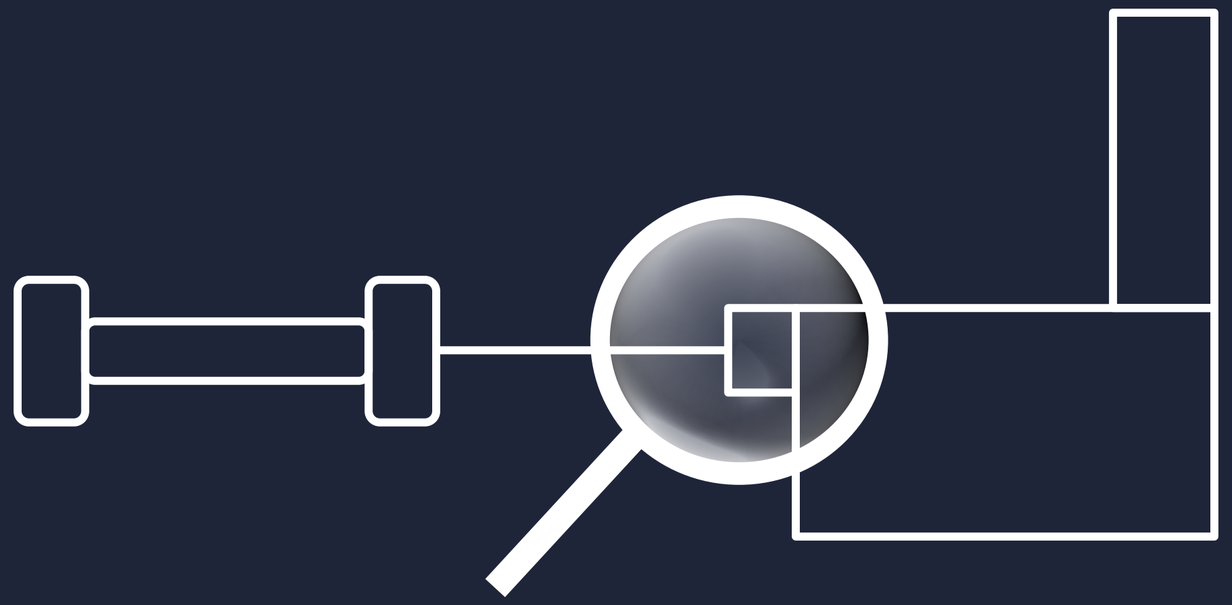


quantification in ESI/HRMS

Malm et al. Molecules 2021



electrospray



anneli kruve

anneli.kruve@su.se

ionization efficiency

ANALYSIS

flow injections

DATA

calibration graphs

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

IONIZATION EFFICIENCY

relative measurements

anneli kruve

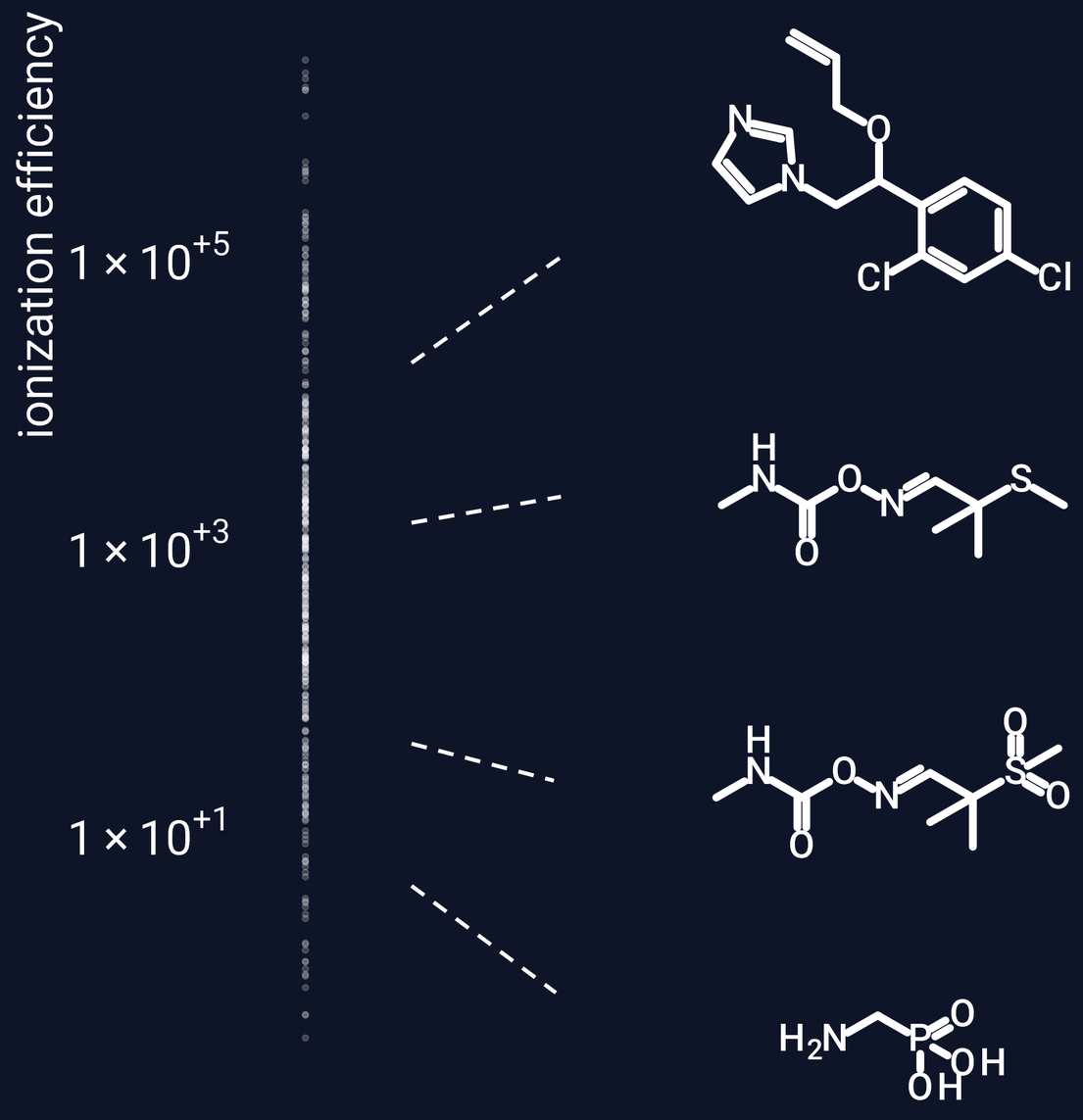
anneli.kruve@su.se

ionization efficiency

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

anneli kruve

anneli.kruve@su.se



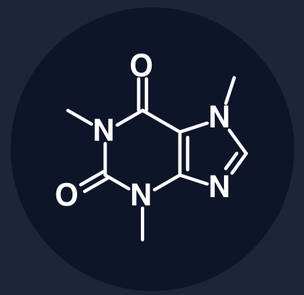
ionization efficiency

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

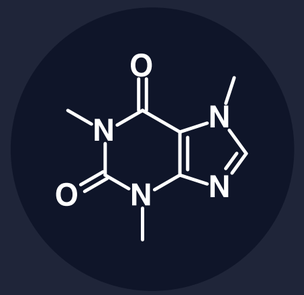
anneli kruve

anneli.kruve@su.se

quantification

with machine learning

workflow



SMILES or MS^2

+ mobile phase composition

MOLECULAR DESCRIPTORS

PaDEL, Mordred, SIRIUS fingerprints

MODEL TRAINING

RandomForest, xgbTree

MODEL SELECTION

Root Mean Square Error (RMSE)

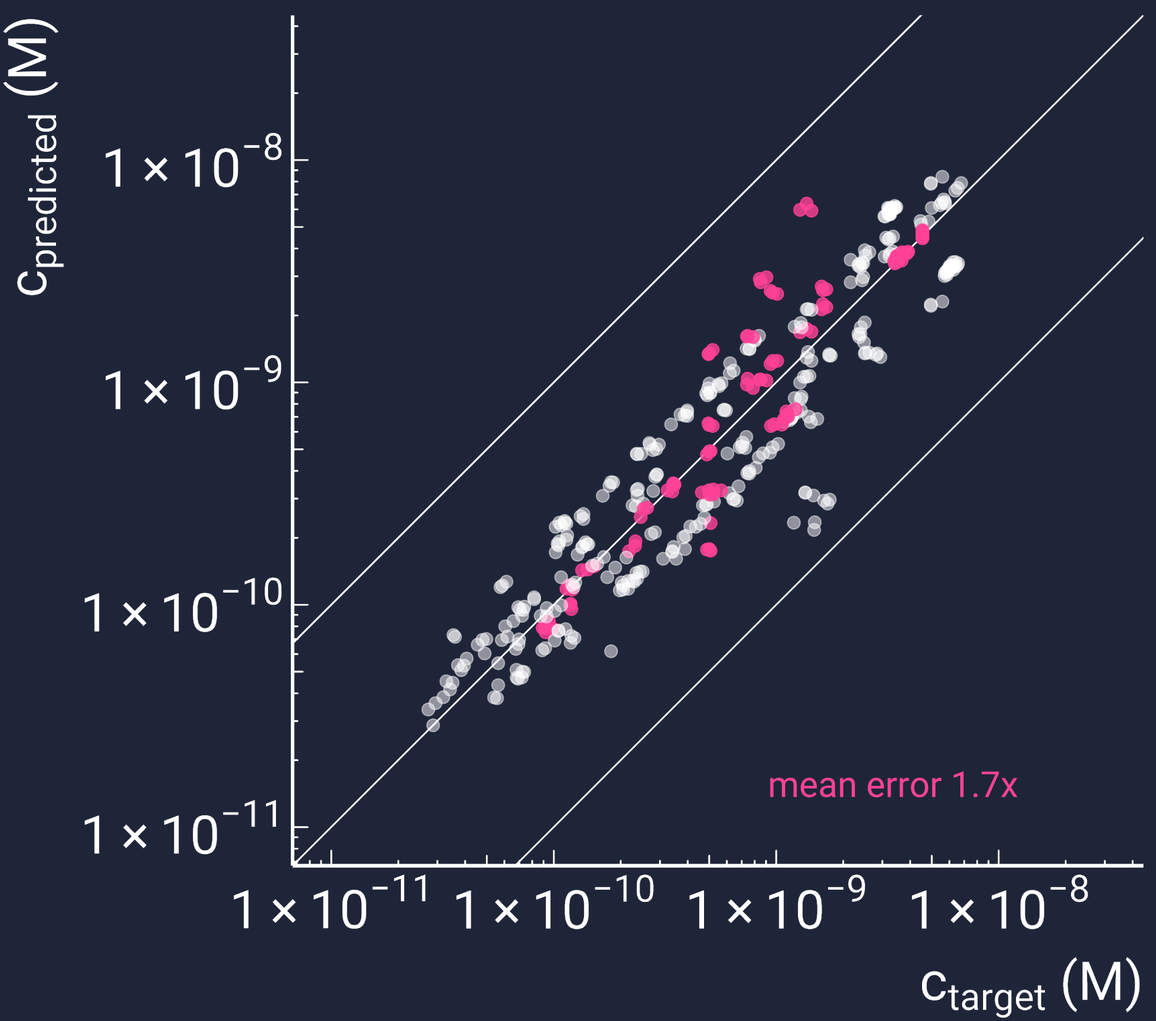
anneli kruve

anneli.kruve@su.se

quantification

from structure

Been et al. Water Research 2021



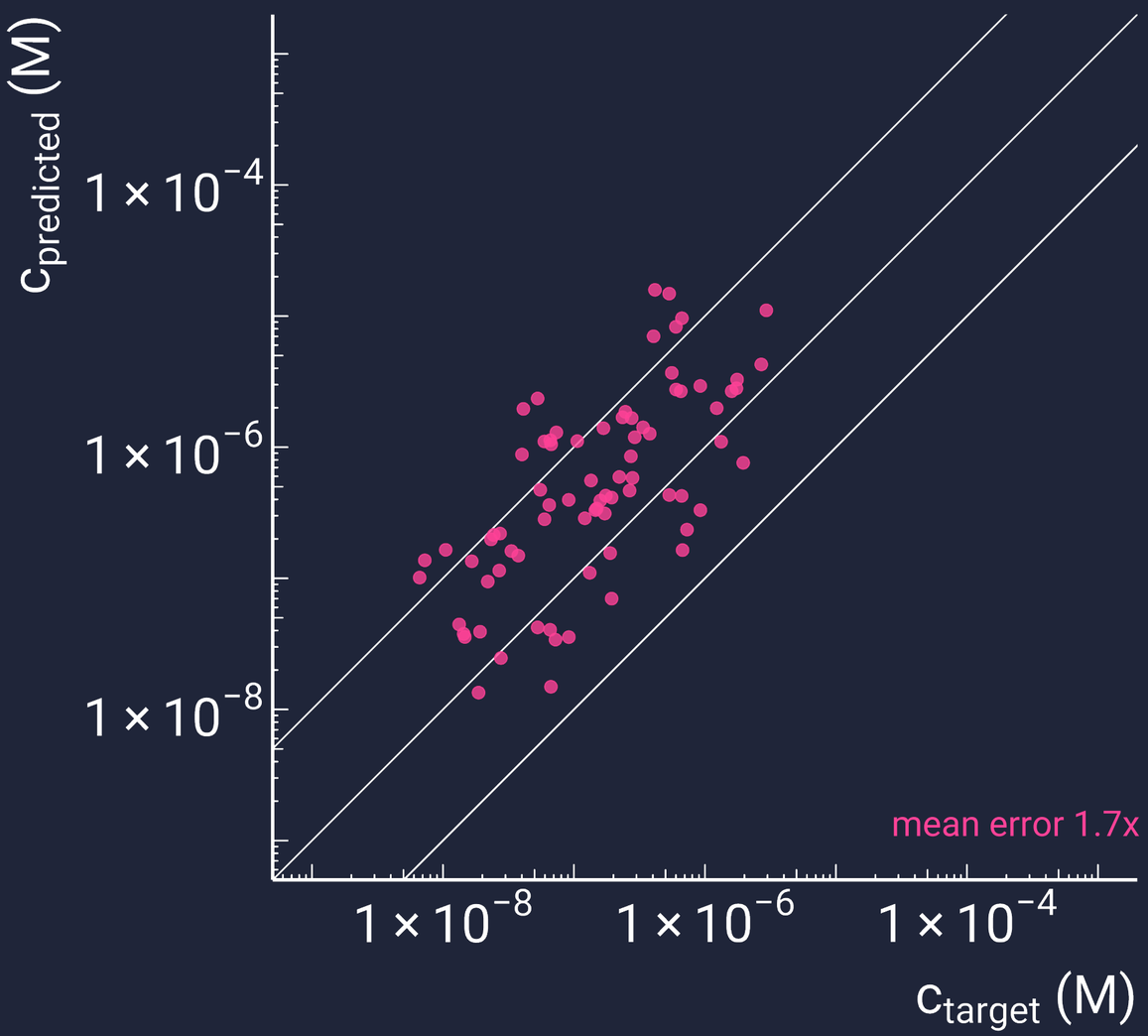
anneli kruve

anneli.kruve@su.se

quantification

from MS^2 spectra

Sepman et al. Anal Chem 2023



anneli kruve

anneli.kruve@su.se

how to ...

PRIORITIZE

risk

IDENTIFY

structure

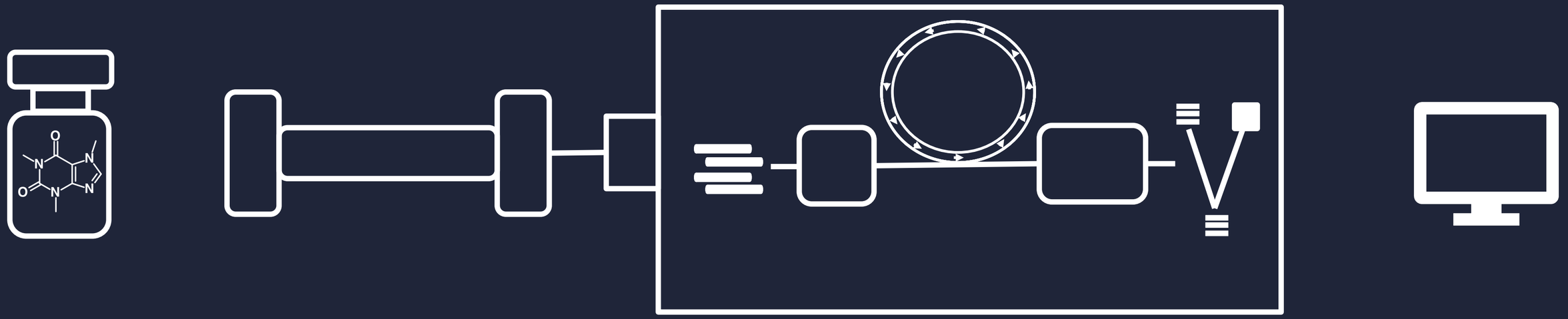
how to ...

IDENTIFY

structure

orthogonal separation

Akhlaqi et al. Anal Bioanal Chem 2023



14
isomeric
chemicals

C_{18} RP

Cyclic IMS
&
MS^2 with ToF

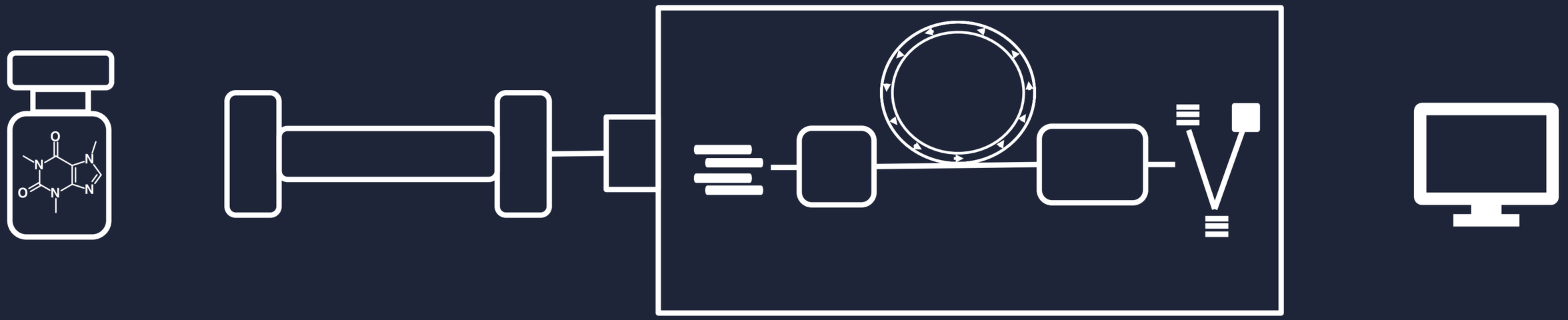
SIRIUS+
CSI:FingerID
&
CFM-ID

anneli kruve

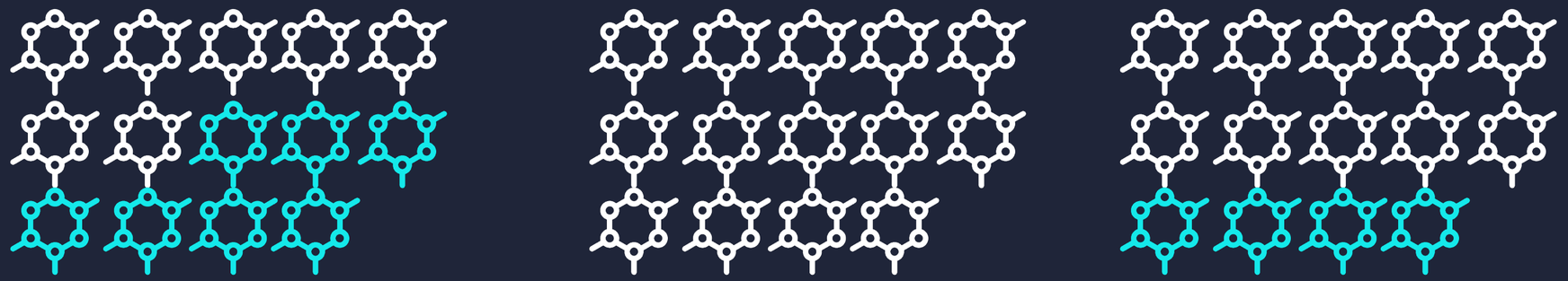
anneli.kruve@su.se

orthogonal separation

Akhlaqi et al. Anal Bioanal Chem 2023



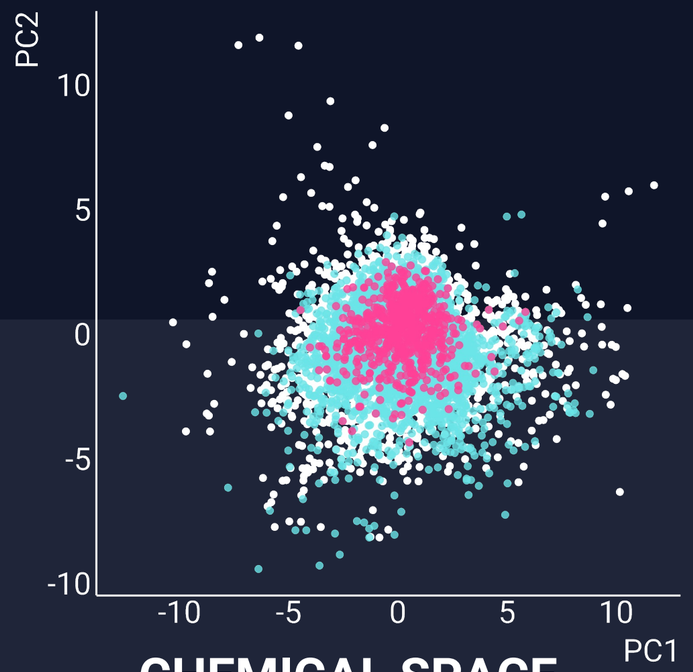
t_R *CCS* MS^2



anneli kruve

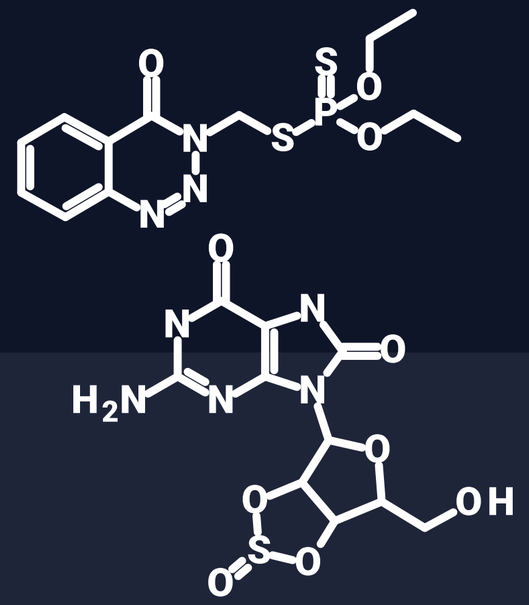
anneli.kruve@su.se

where is the problem



CHEMICAL SPACE

low chemical space coverage



CANDIDATE STRUCTURES

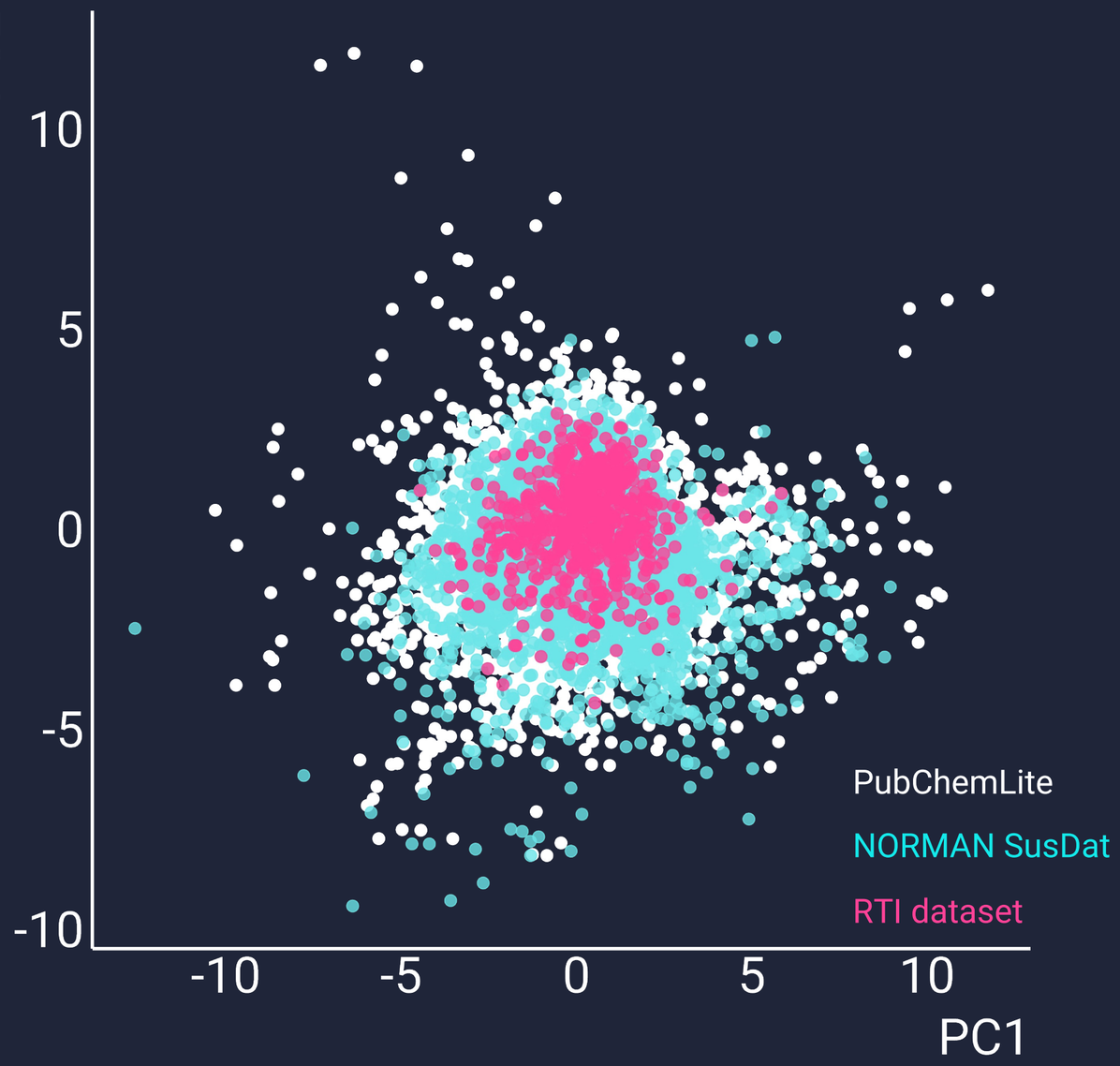
ability to distinguish isomers

anneli kruve

anneli.kruve@su.se

chemical space

covered by prediction algorithms

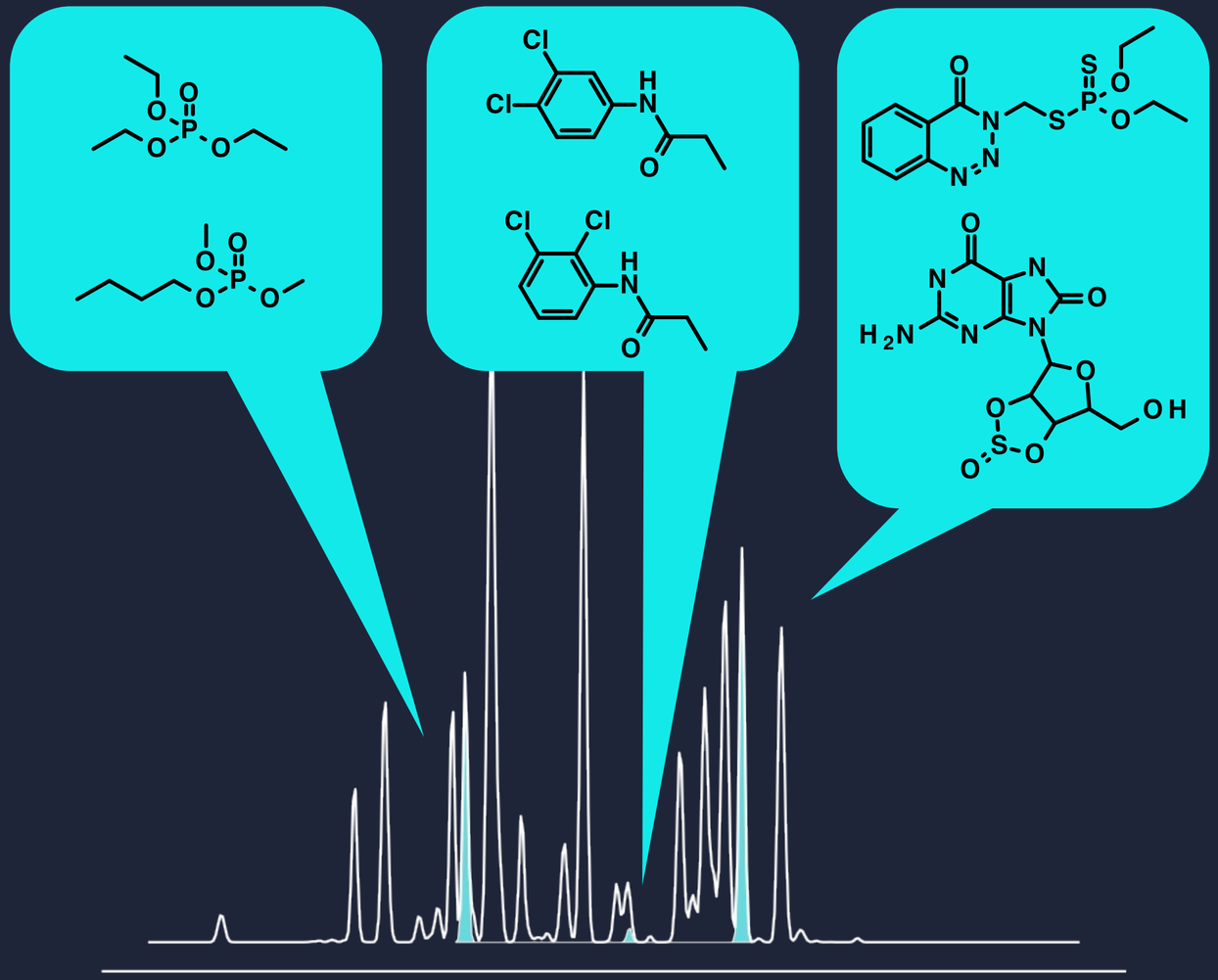


anneli kruve

anneli.kruve@su.se

semi-supervised learning for RT

Sandberg et al. in preparation

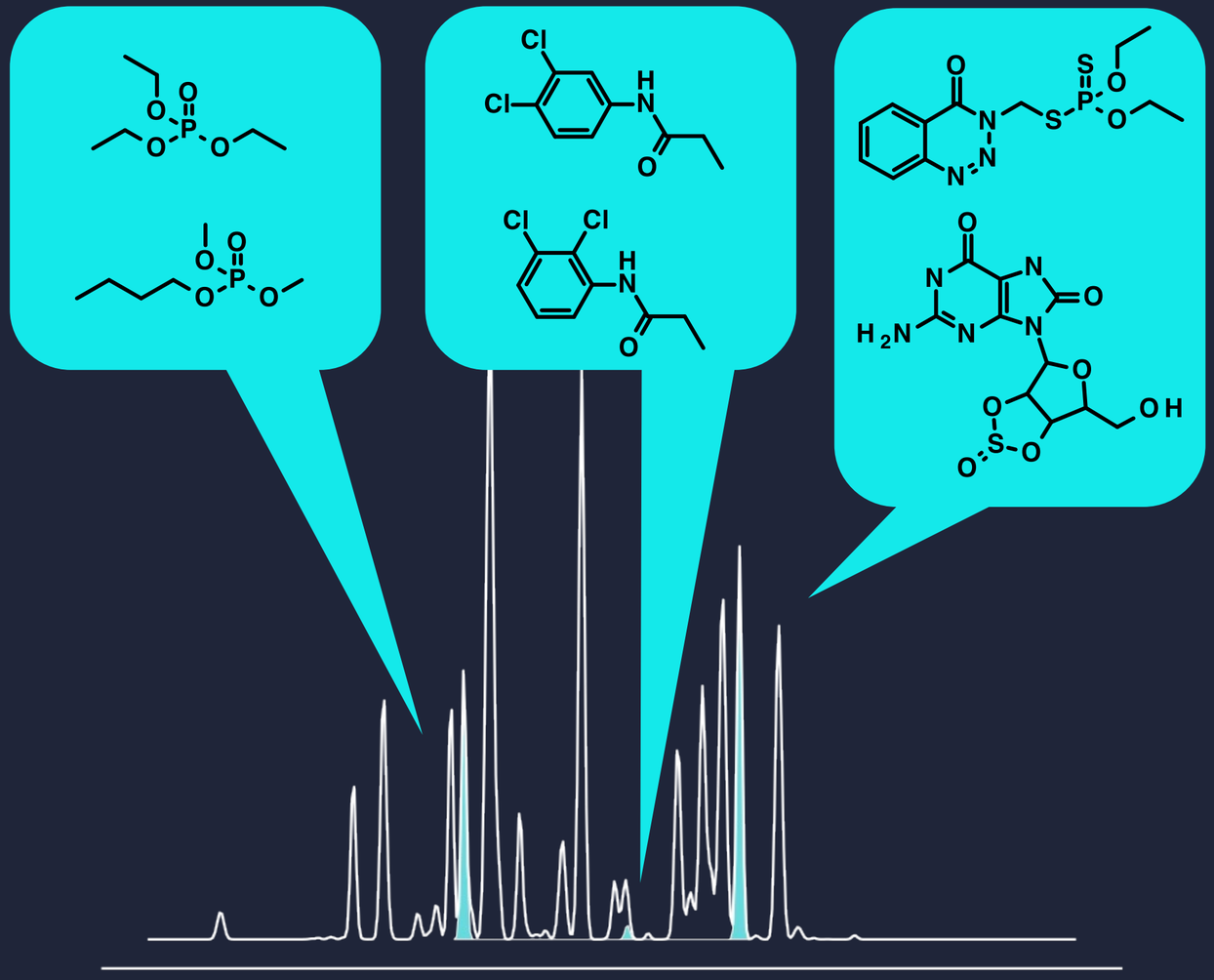


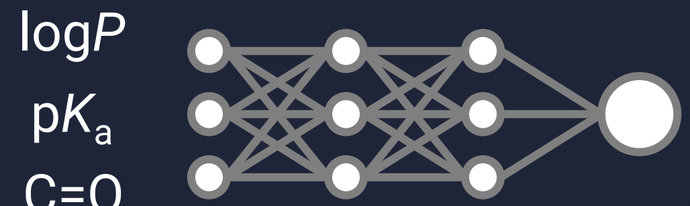
anneli kruve

anneli.kruve@su.se

semi-supervised learning for RT

Sandberg et al. in preparation



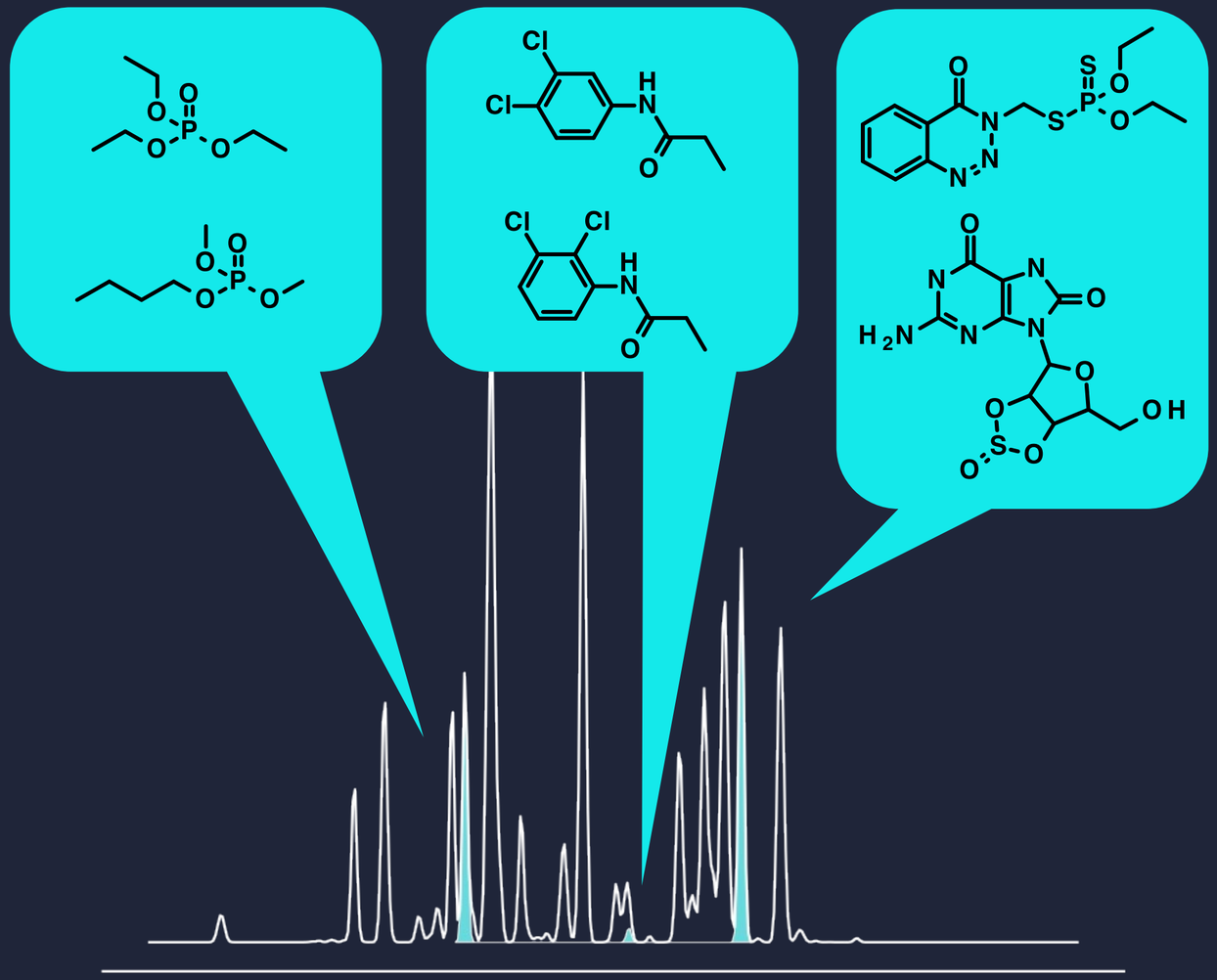


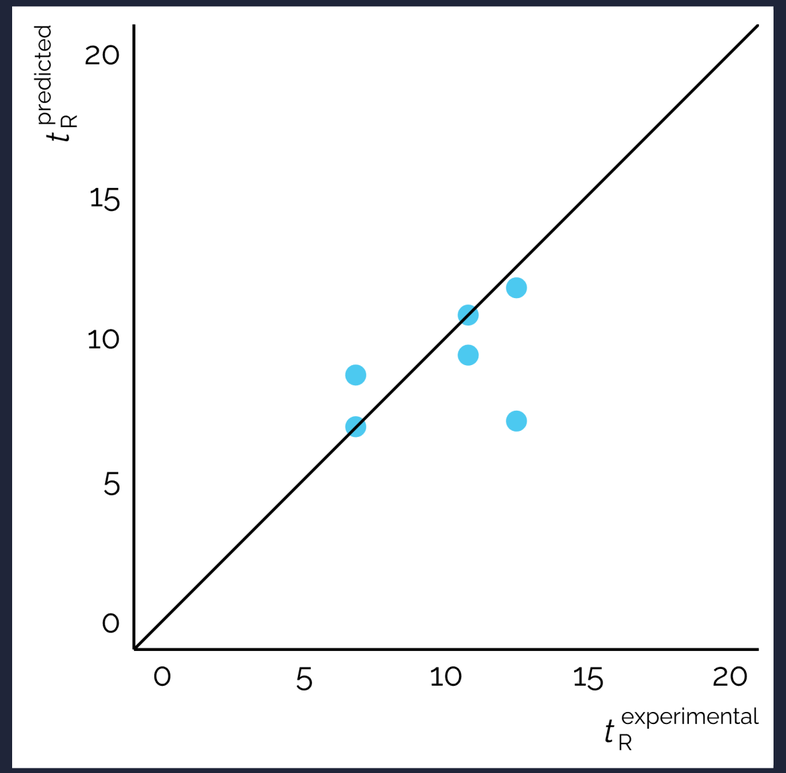
anneli kruve

anneli.kruve@su.se

semi-supervised learning for RT

Sandberg et al. in preparation

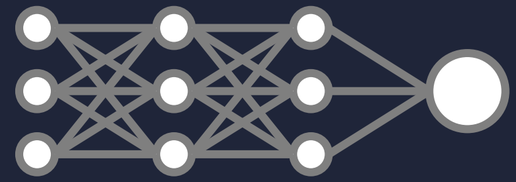




log*P*

pK_a

C=O



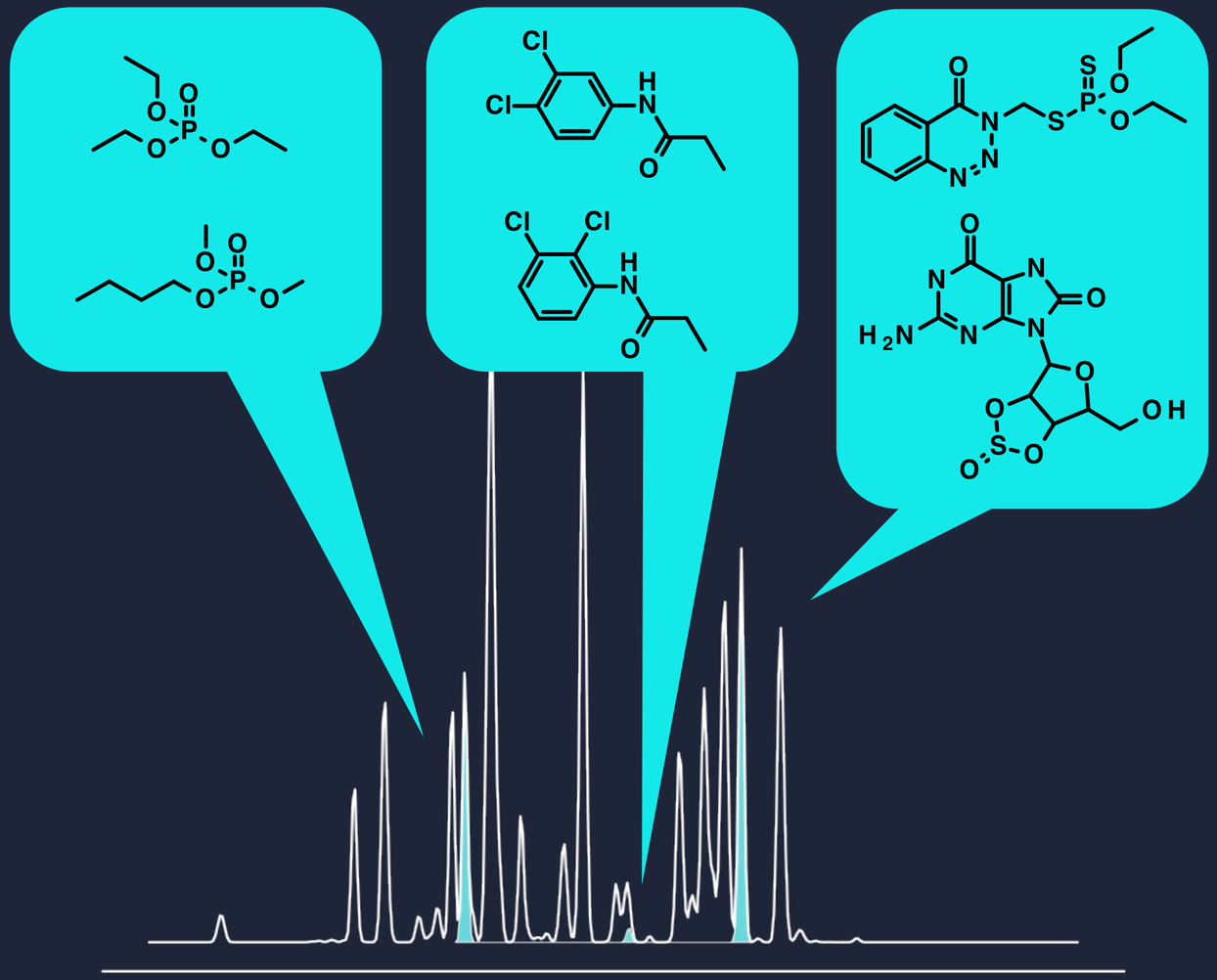
t_R

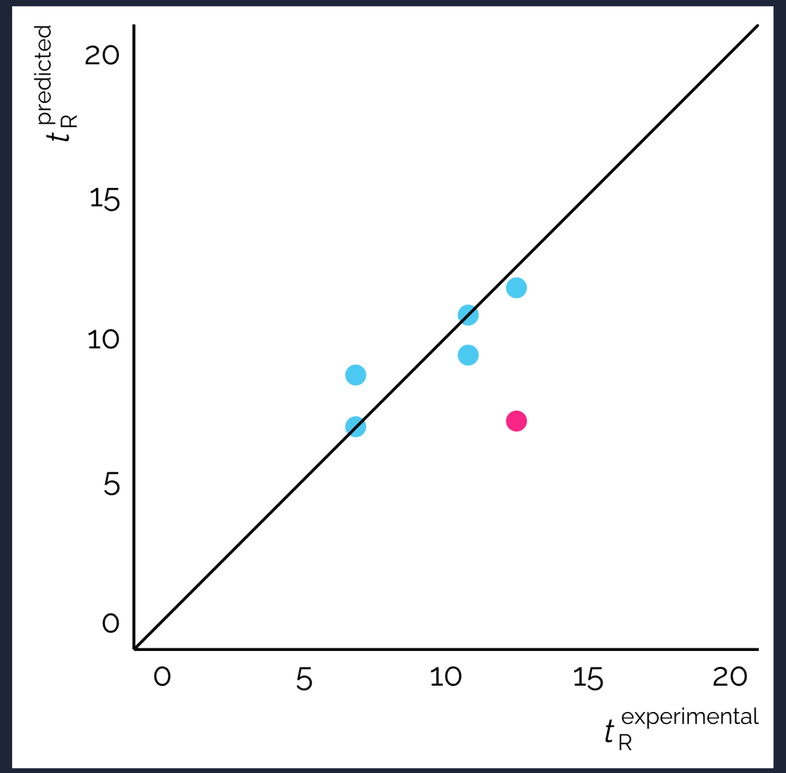
anneli kruve

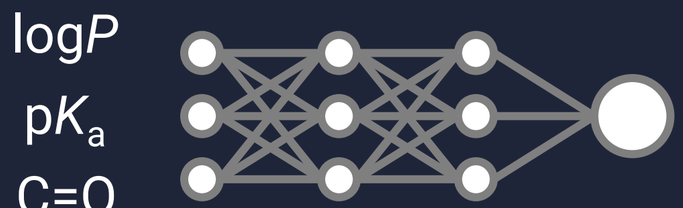
anneli.kruve@su.se

semi-supervised learning for RT

Sandberg et al. in preparation





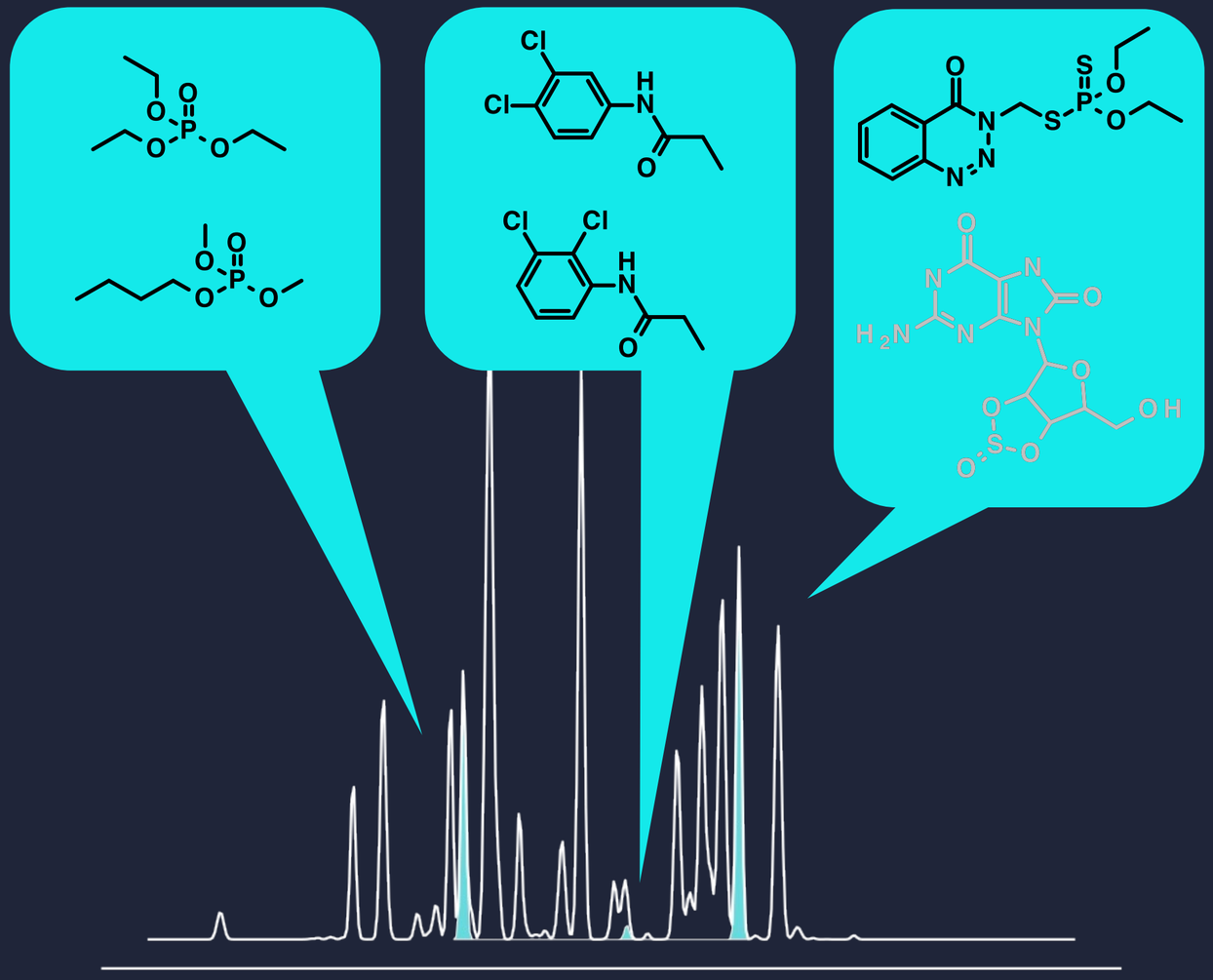


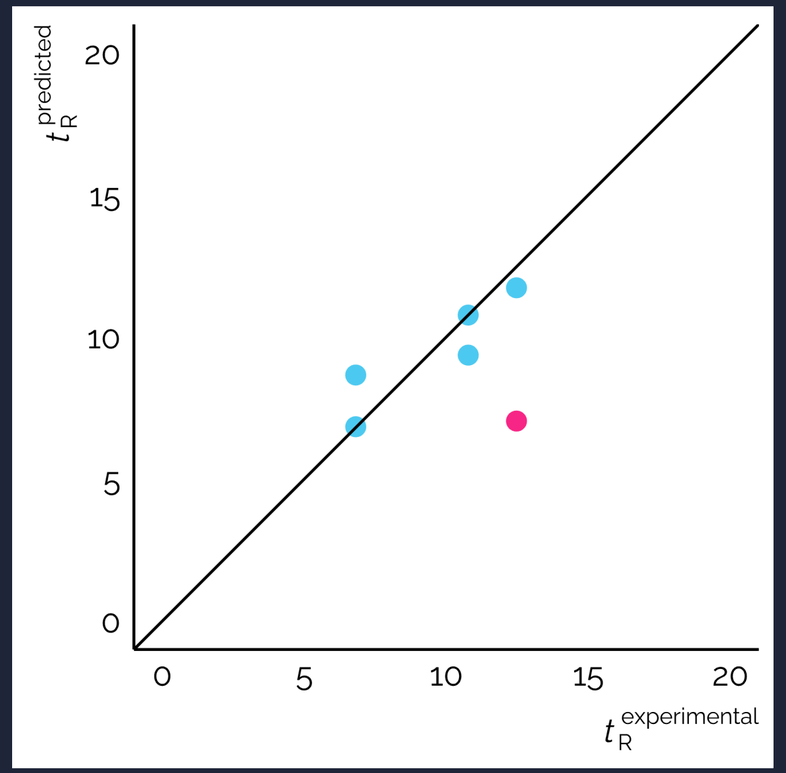
anneli kruve

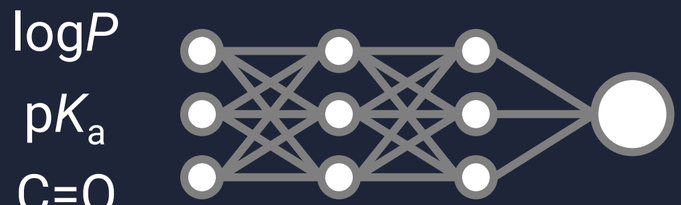
anneli.kruve@su.se

semi-supervised learning for RT

Sandberg et al. in preparation





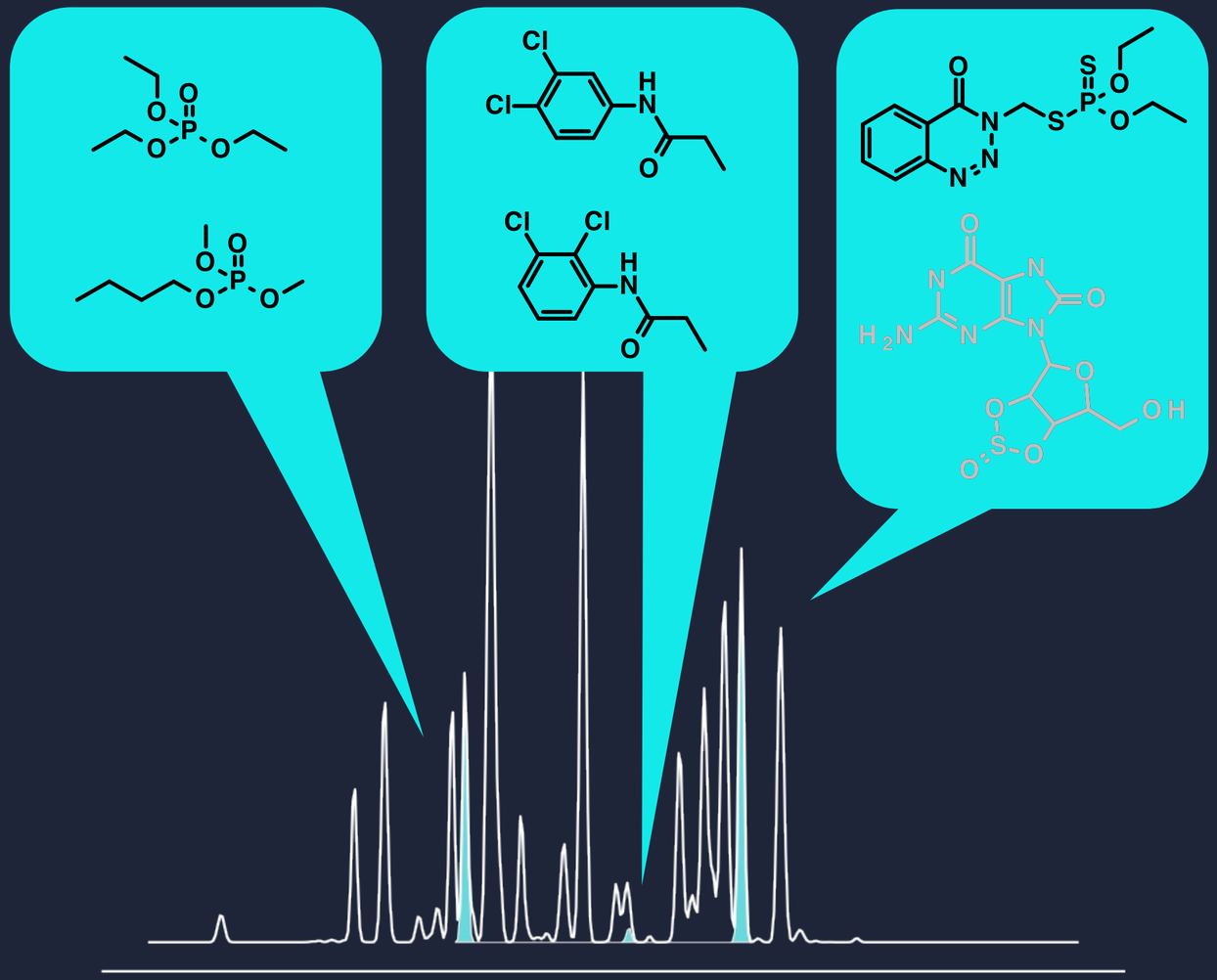


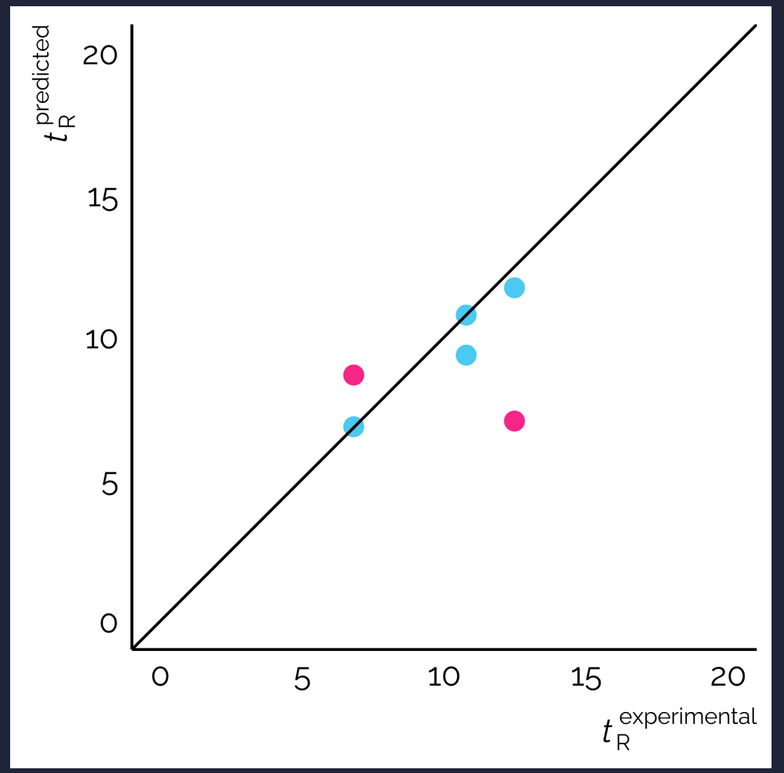
anneli kruve

anneli.kruve@su.se

semi-supervised learning for RT

Sandberg et al. in preparation

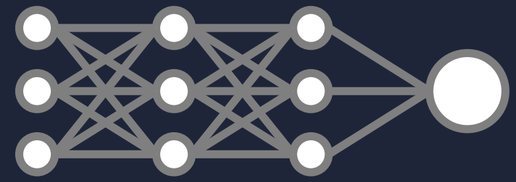




log*P*

pK_a

C=O



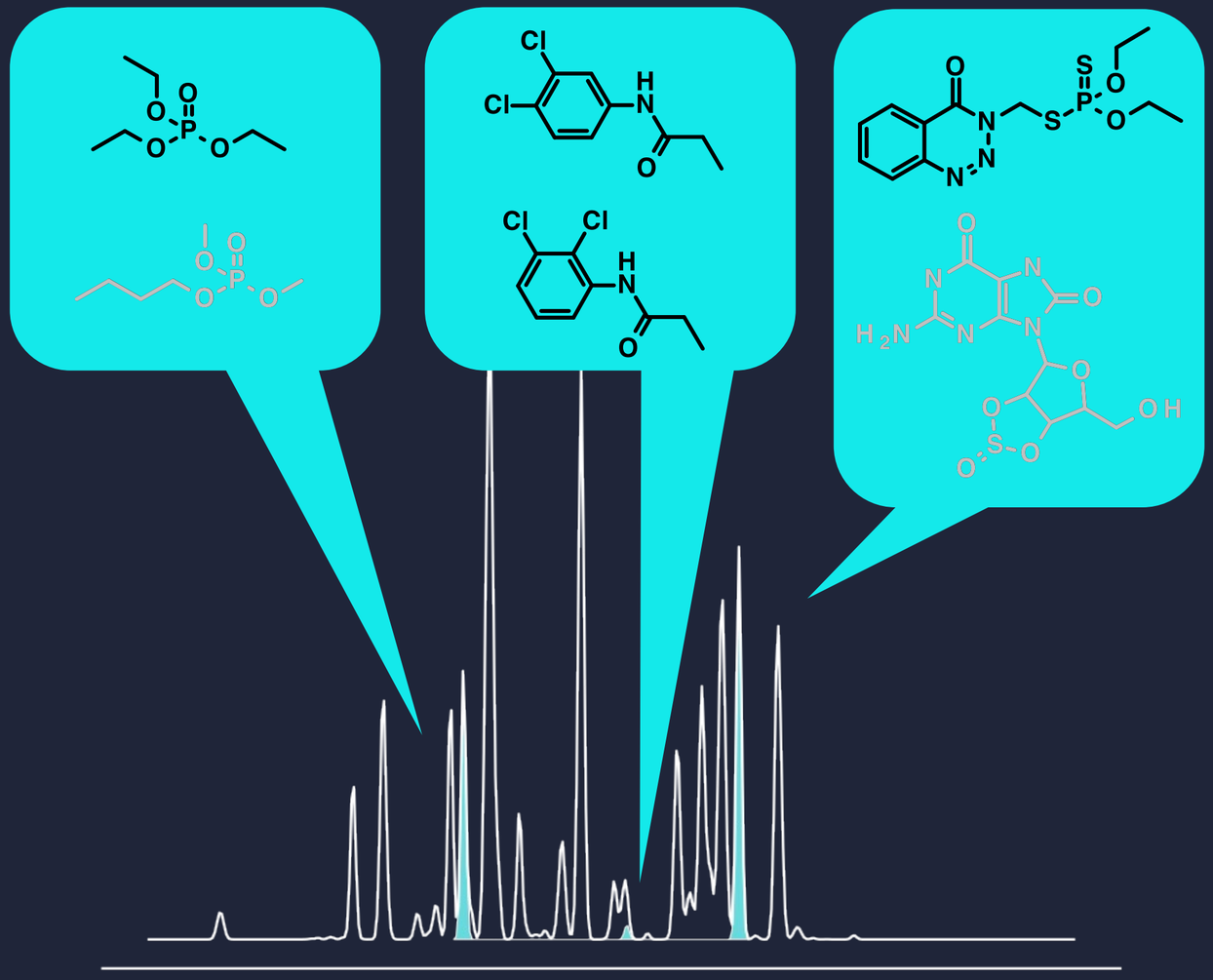
t_R

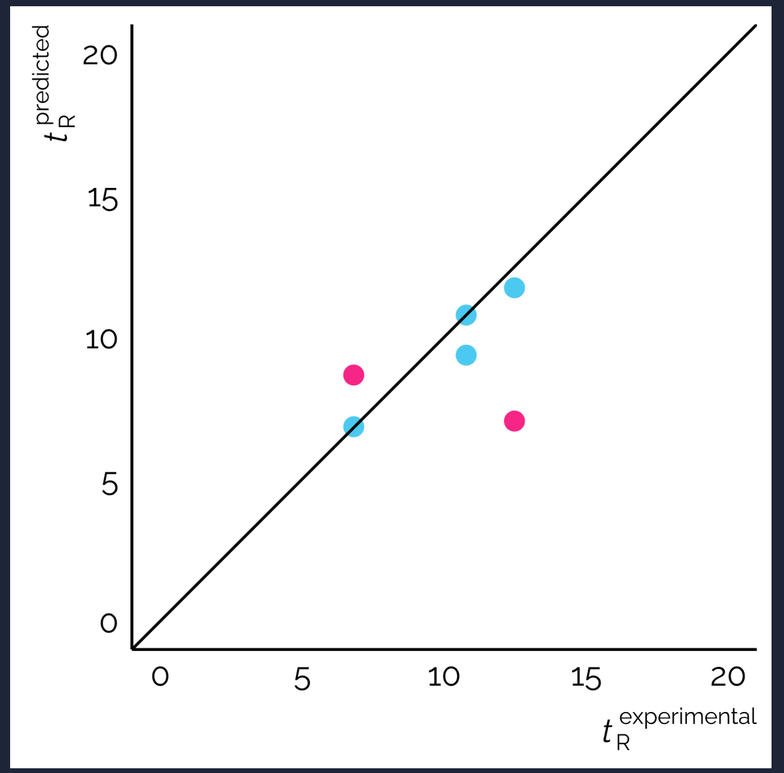
anneli kruve

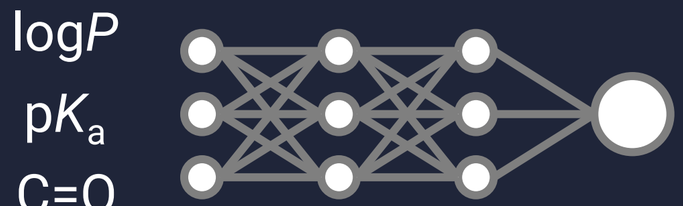
anneli.kruve@su.se

semi-supervised learning for RT

Sandberg et al. in preparation





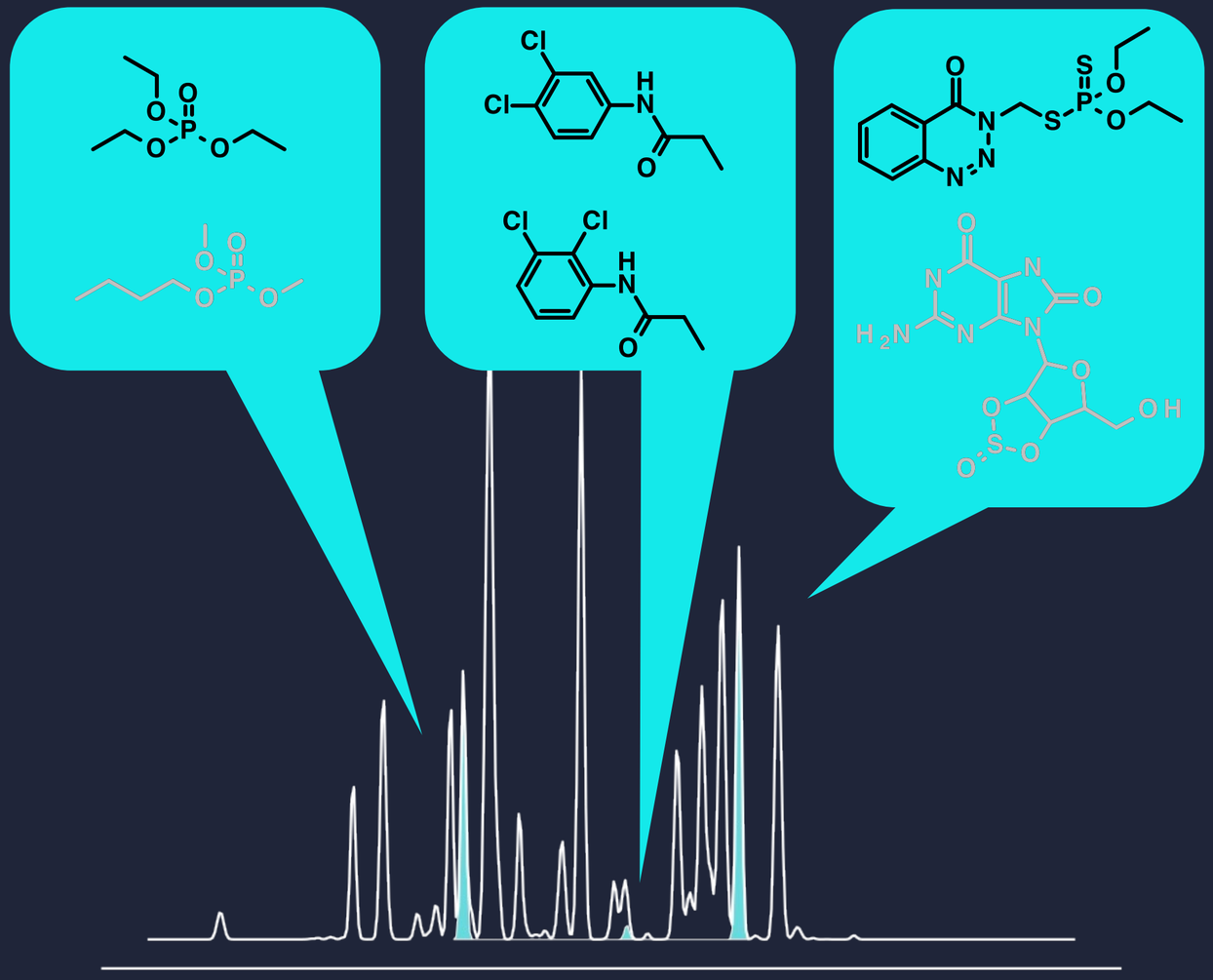


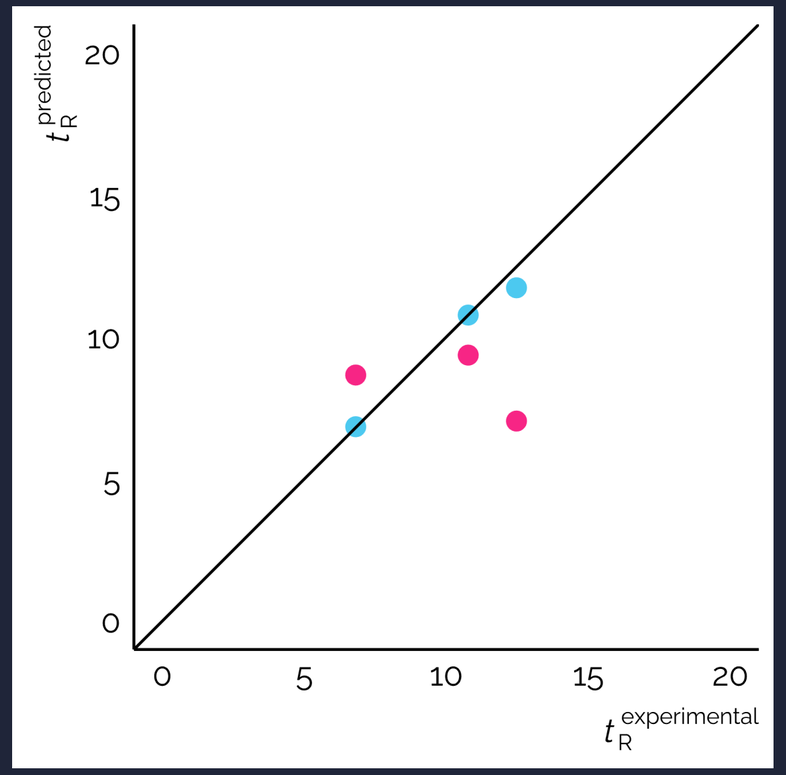
anneli kruve

anneli.kruve@su.se

semi-supervised learning for RT

Sandberg et al. in preparation

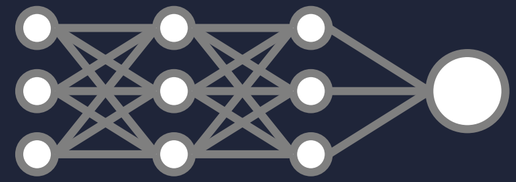




log*P*

pK_a

C=O



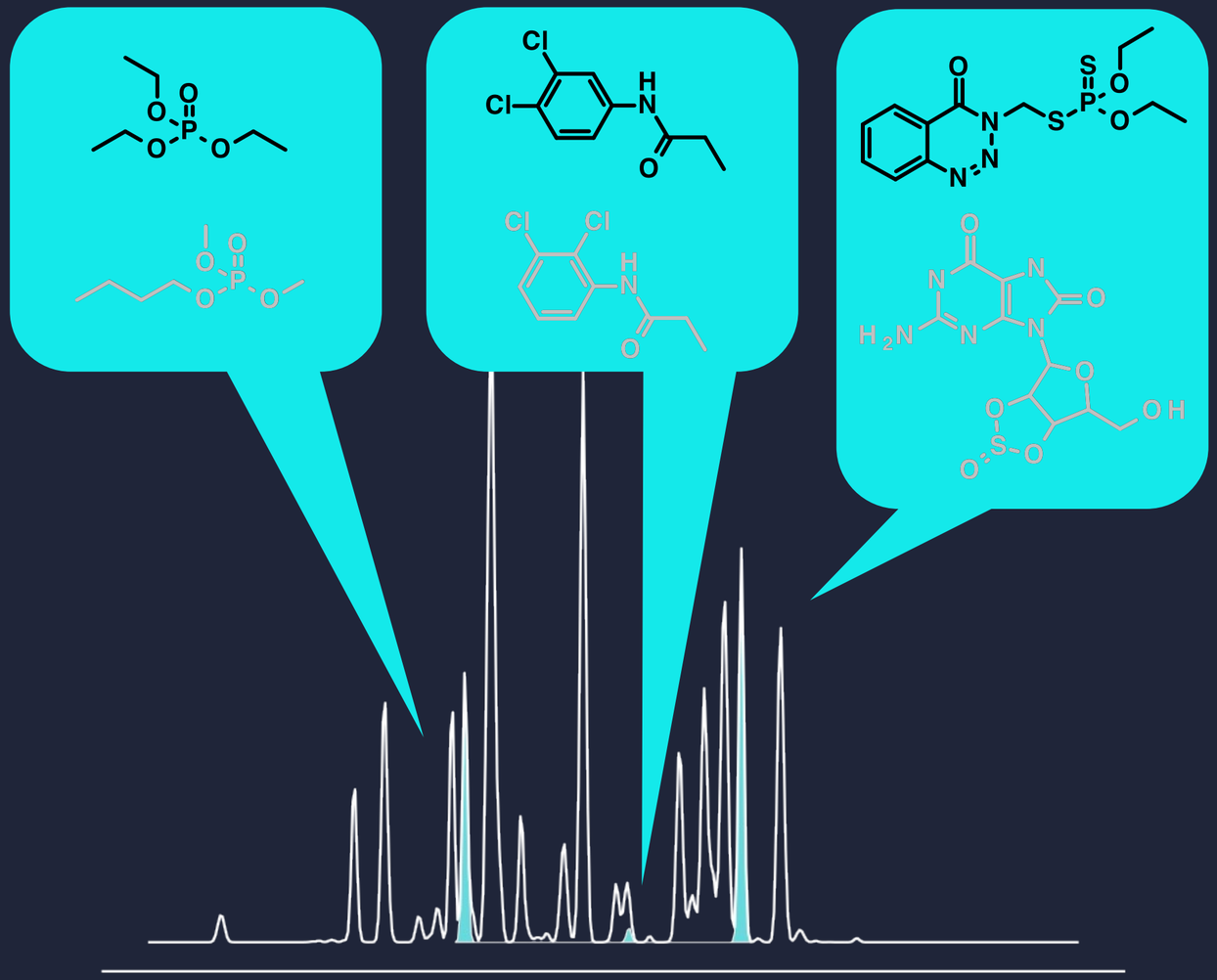
t_R

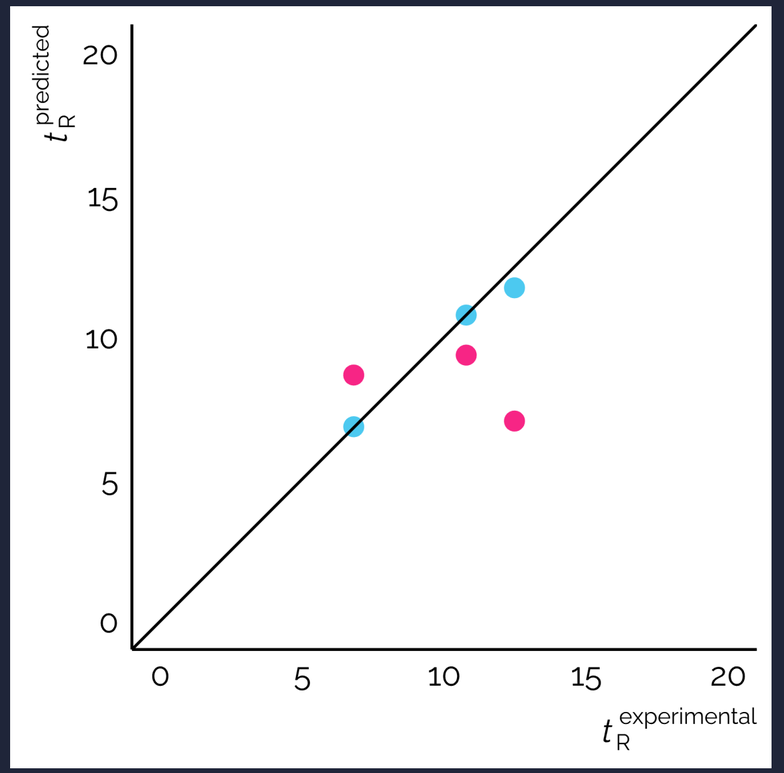
anneli kruve

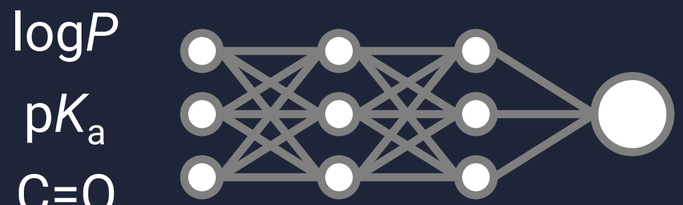
anneli.kruve@su.se

semi-supervised learning for RT

Sandberg et al. in preparation







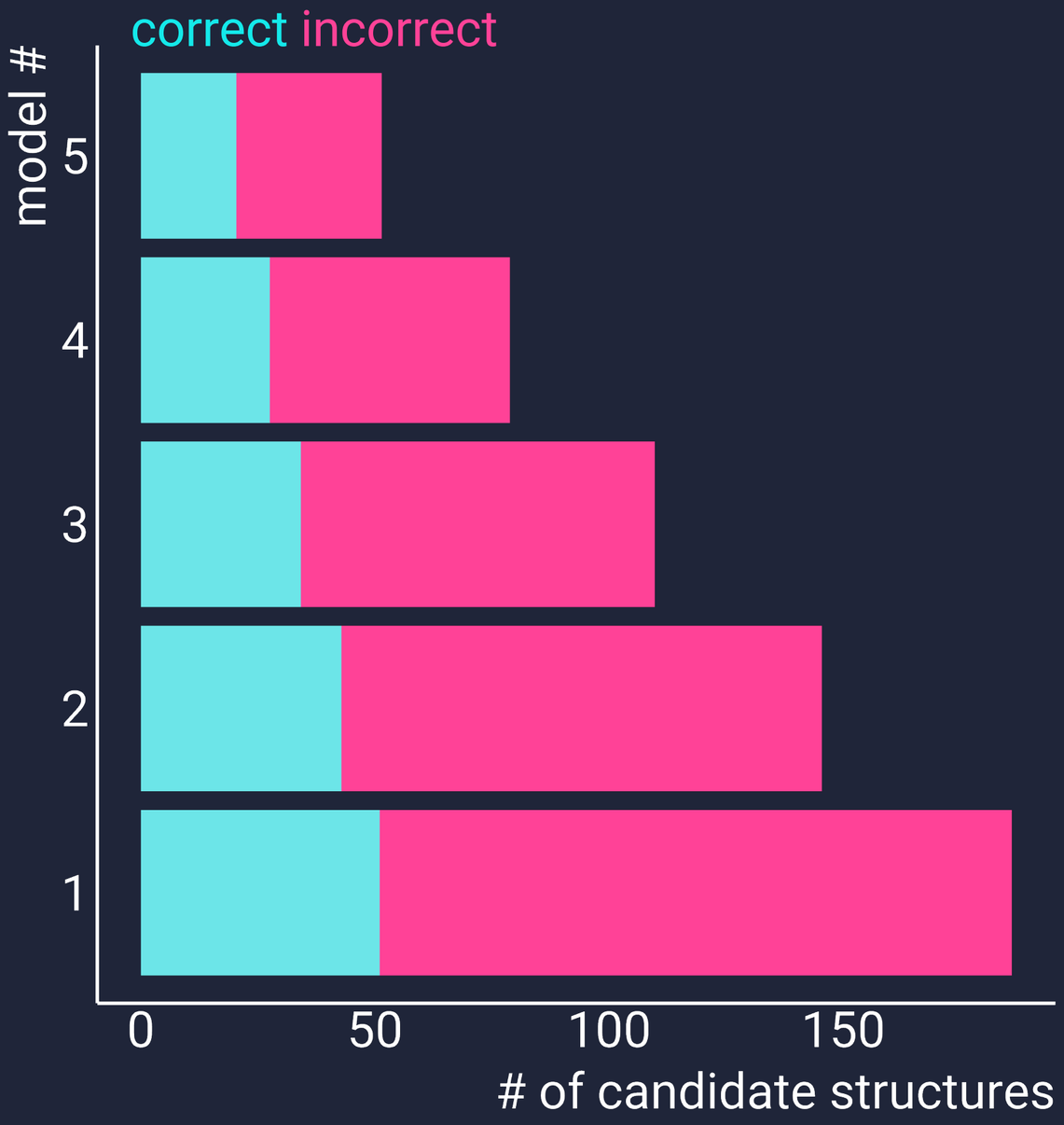
anneli kruve

anneli.kruve@su.se

candidate structure

evaluation

Sandberg et al. in preparation



anneli kruve

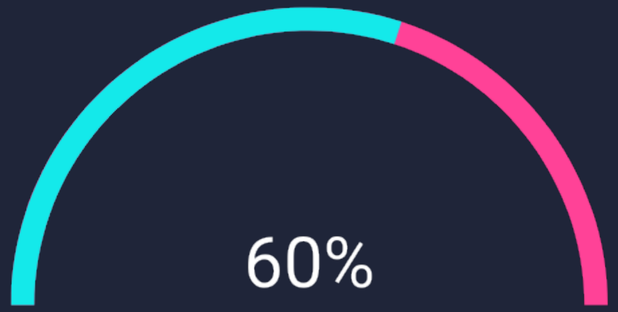
anneli.kruve@su.se

machine learning

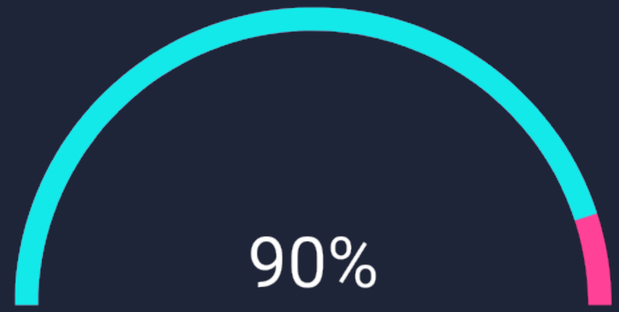
for prioritization and identification in nontarget screening

TOXICITY

QUANTIFICATION



RETENTION TIME



anneli kruve

anneli.kruve@su.se

Swedish Research Council

FORMAS

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

post-doc post-doc PhD PhD PhD

anneli kruve
anneli.kruve@su.se

Kruvelab.com