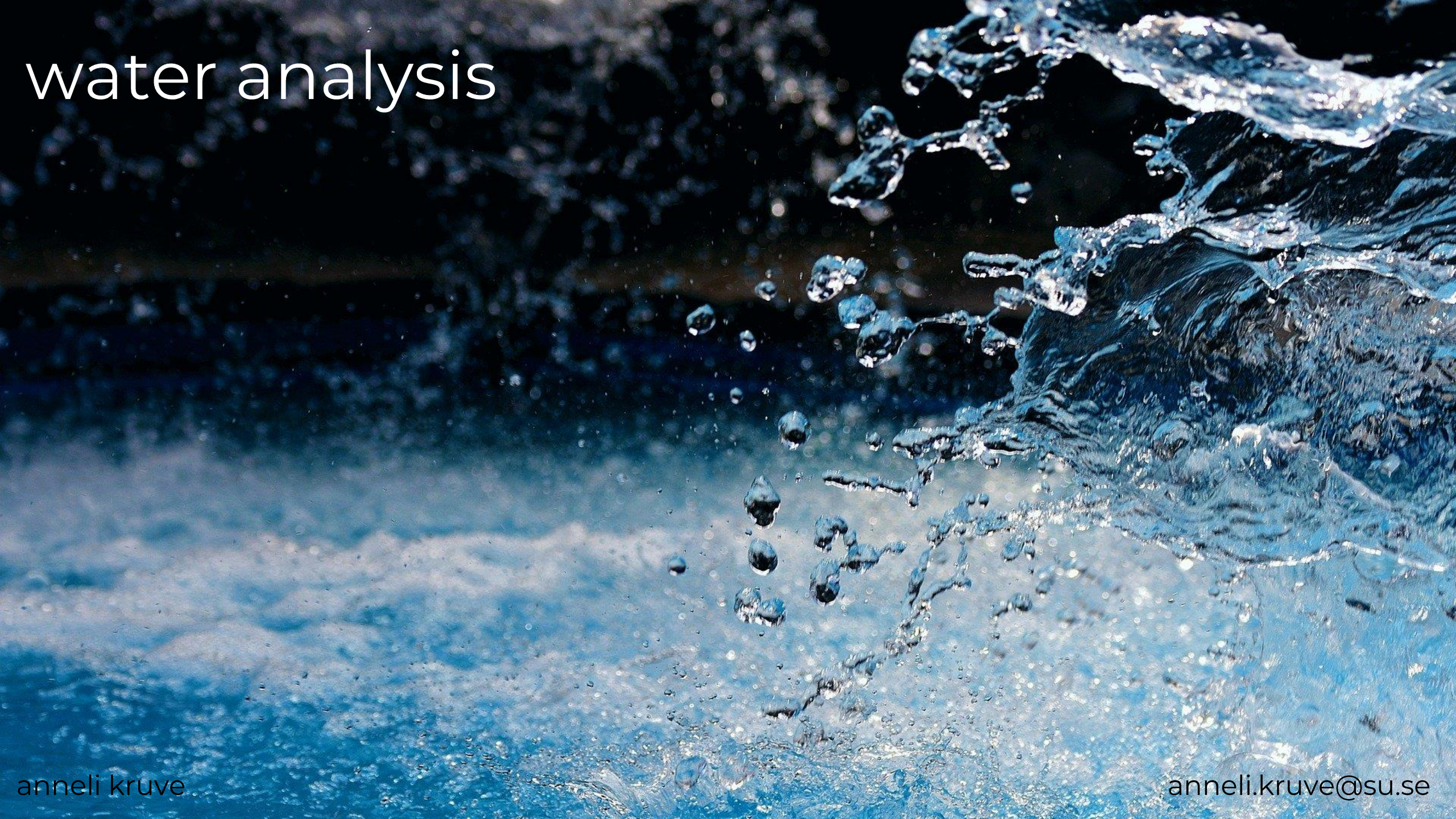
toxicity predictions of unidentified chemicals in water by HRMS & ML: how and why?

anneli kruve
anneli.kruve@su.se
kruvelab.com

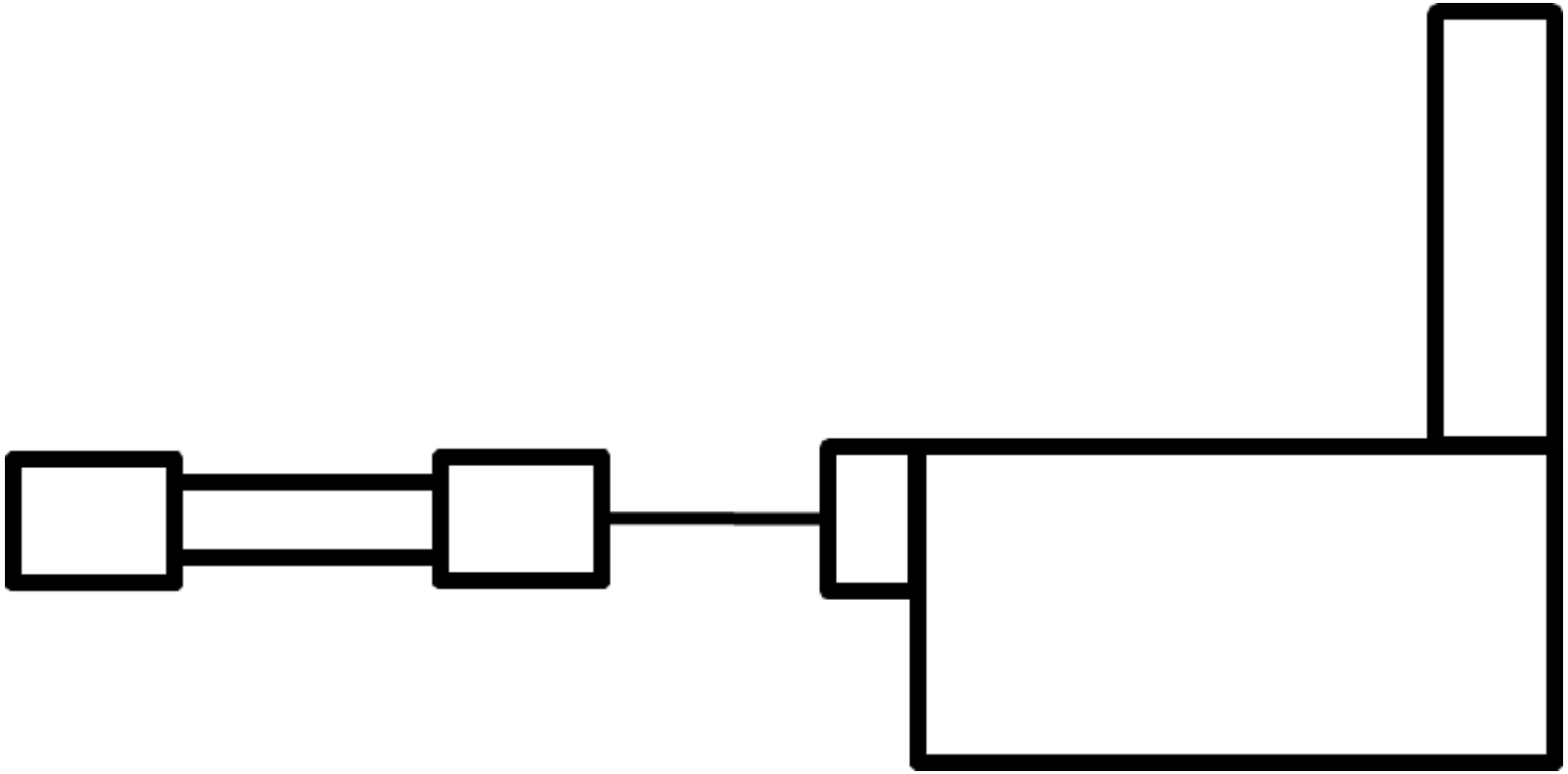
water analysis



anneli kruve

anneli.kruve@su.se

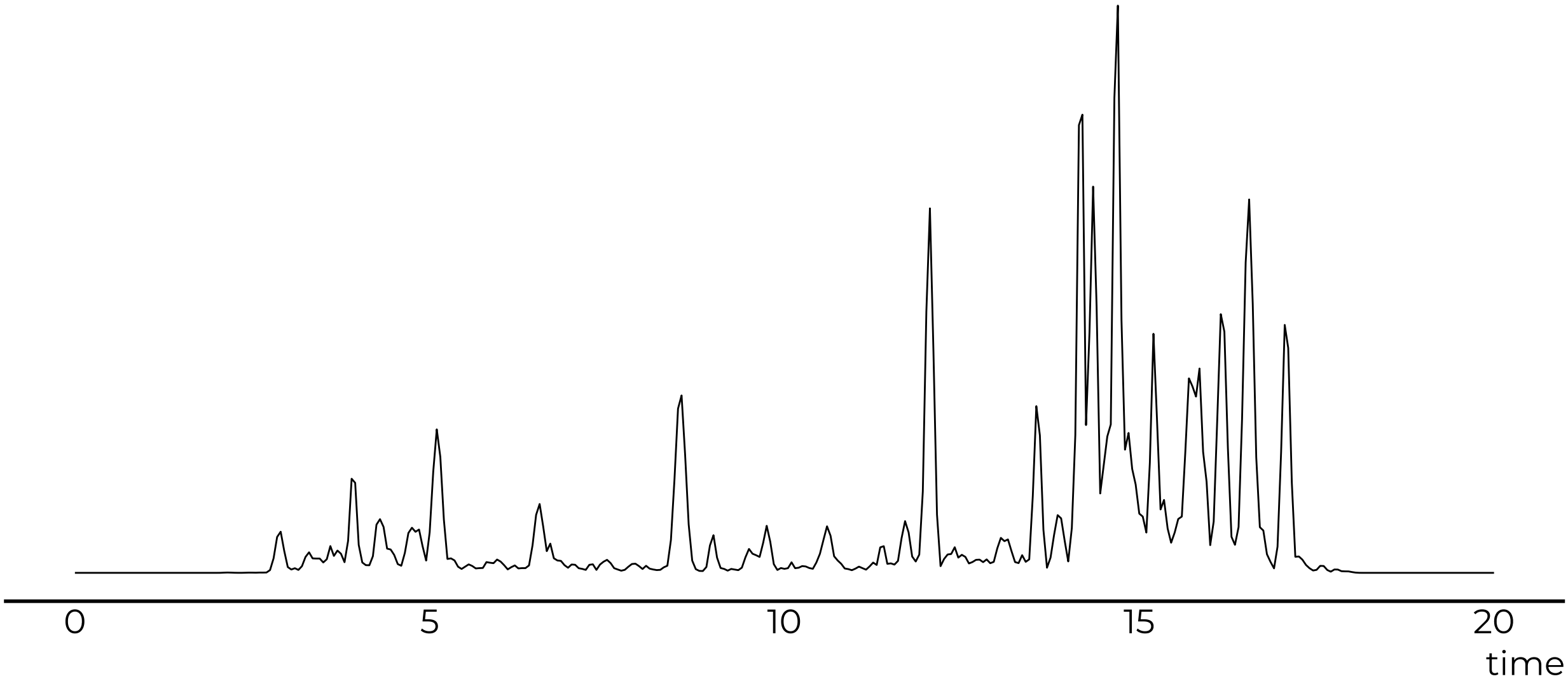
nontarget screening with LC/HRMS



anneli kruve

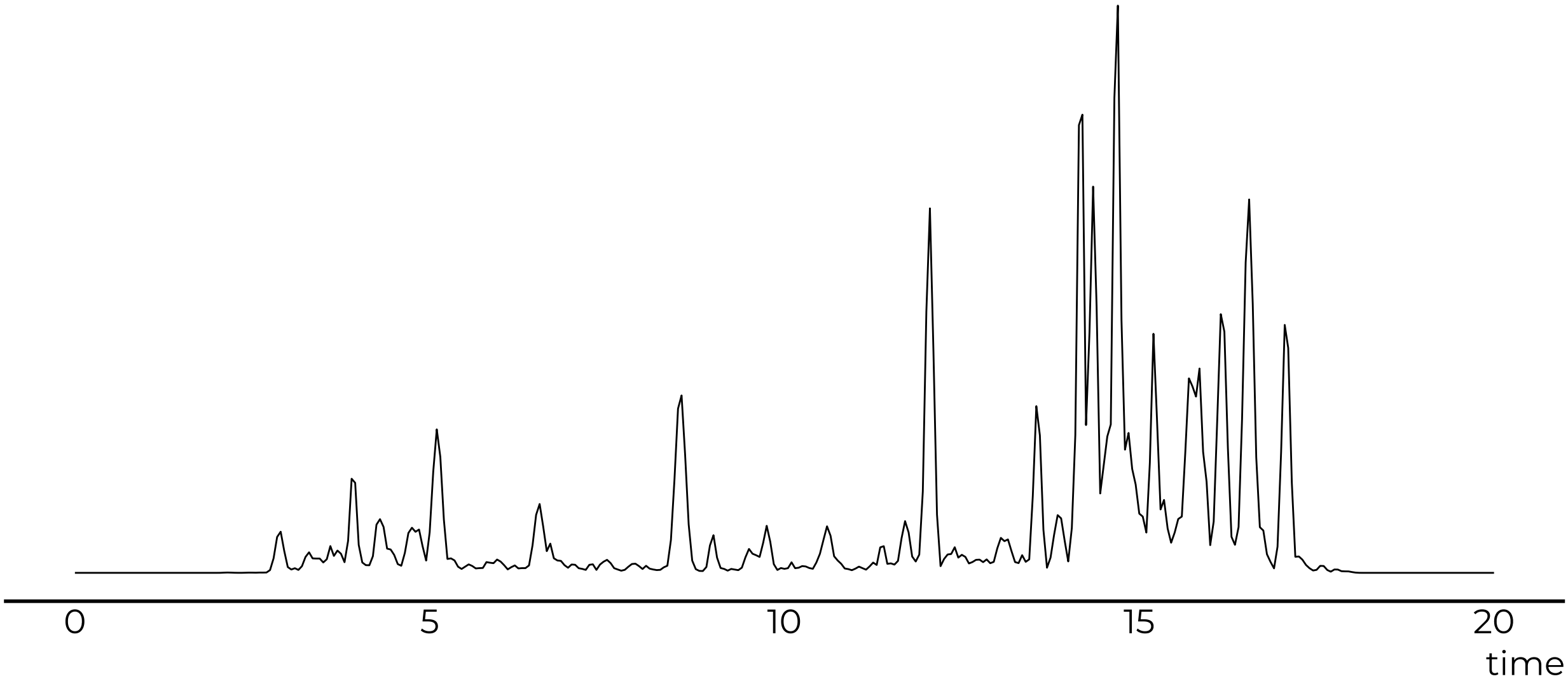
anneli.kruve@su.se

nontarget screening with LC/HRMS



anneli kruve
anneli.kruve@su.se

what next?



anneli kruve
anneli.kruve@su.se

prioritization

toxicity

anneli.kruve@su.se

prioritization

toxicity

concentration

anneli kruve

anneli.kruve@su.se

prioritization

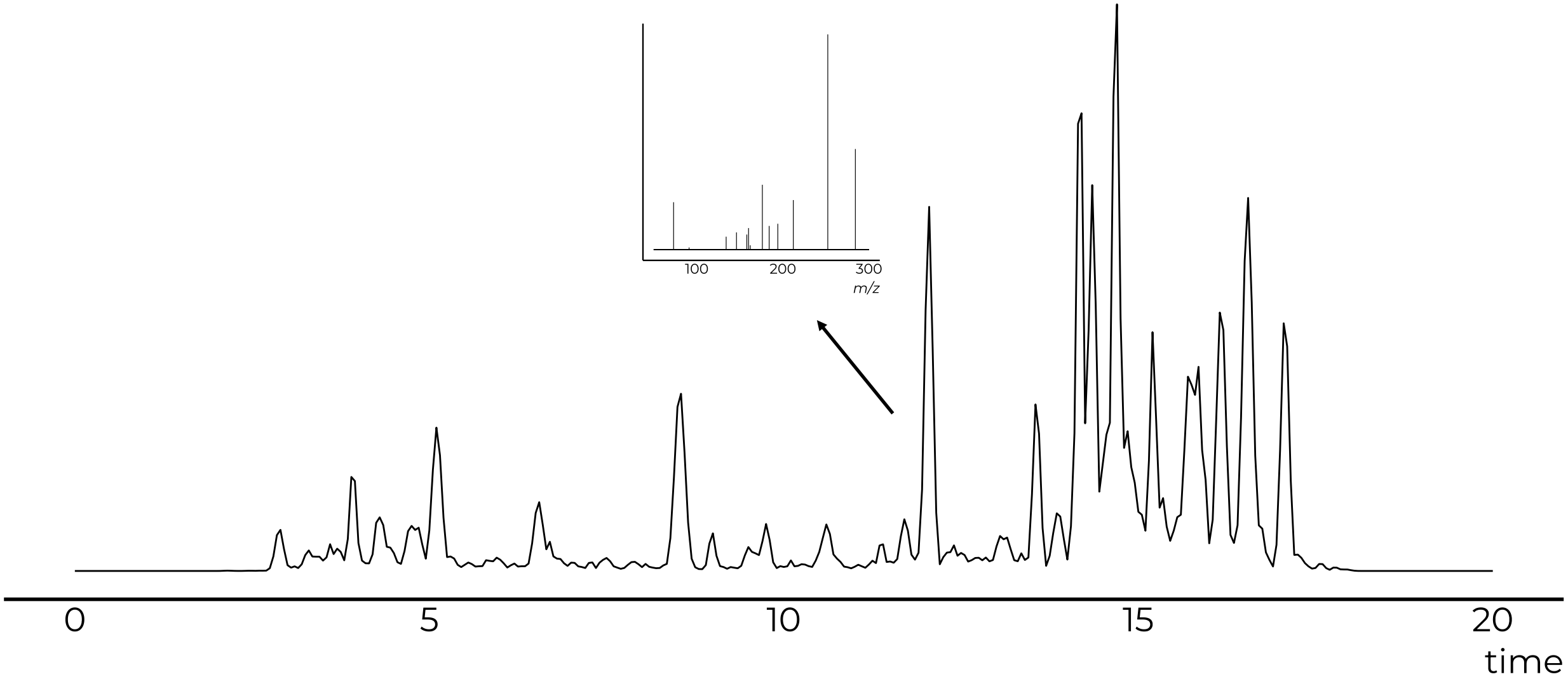
toxicity

concentration

risk

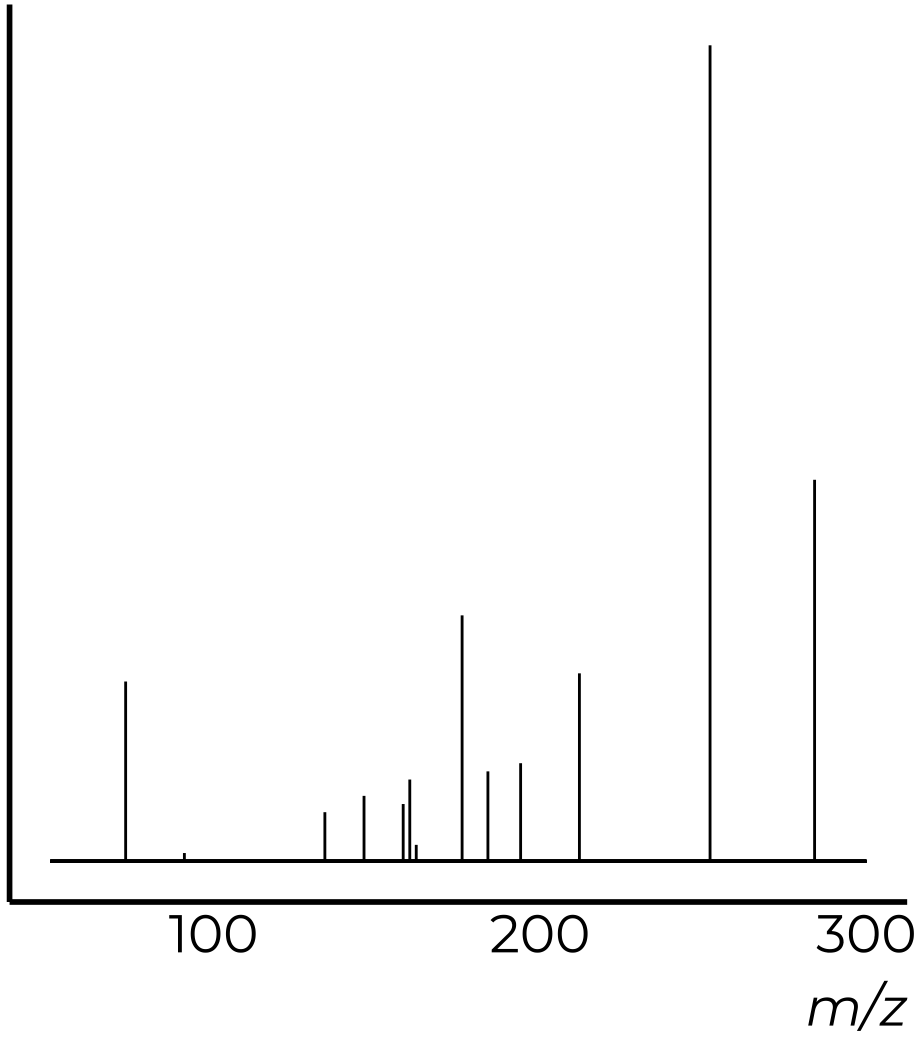
anneli.kruve@su.se

nontarget screening with LC/HRMS



anneli kruve
anneli.kruve@su.se

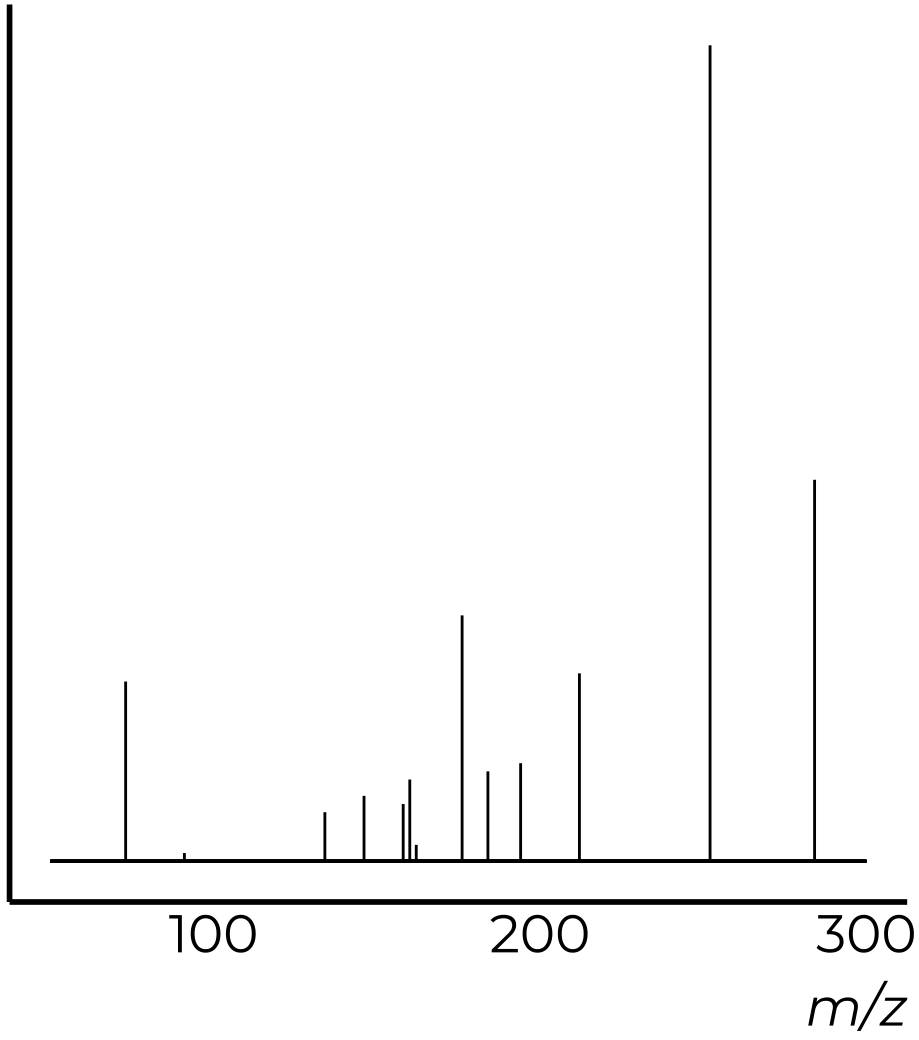
toxicity assessment

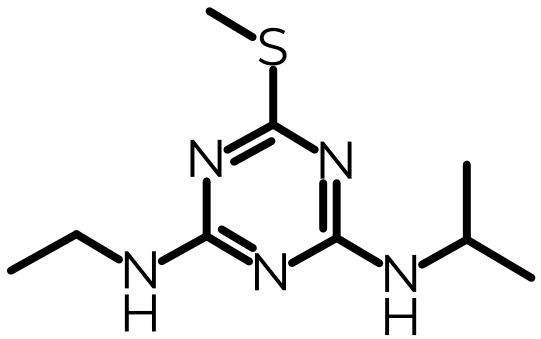


anneli kruve

anneli.kruve@su.se

toxicity assessment

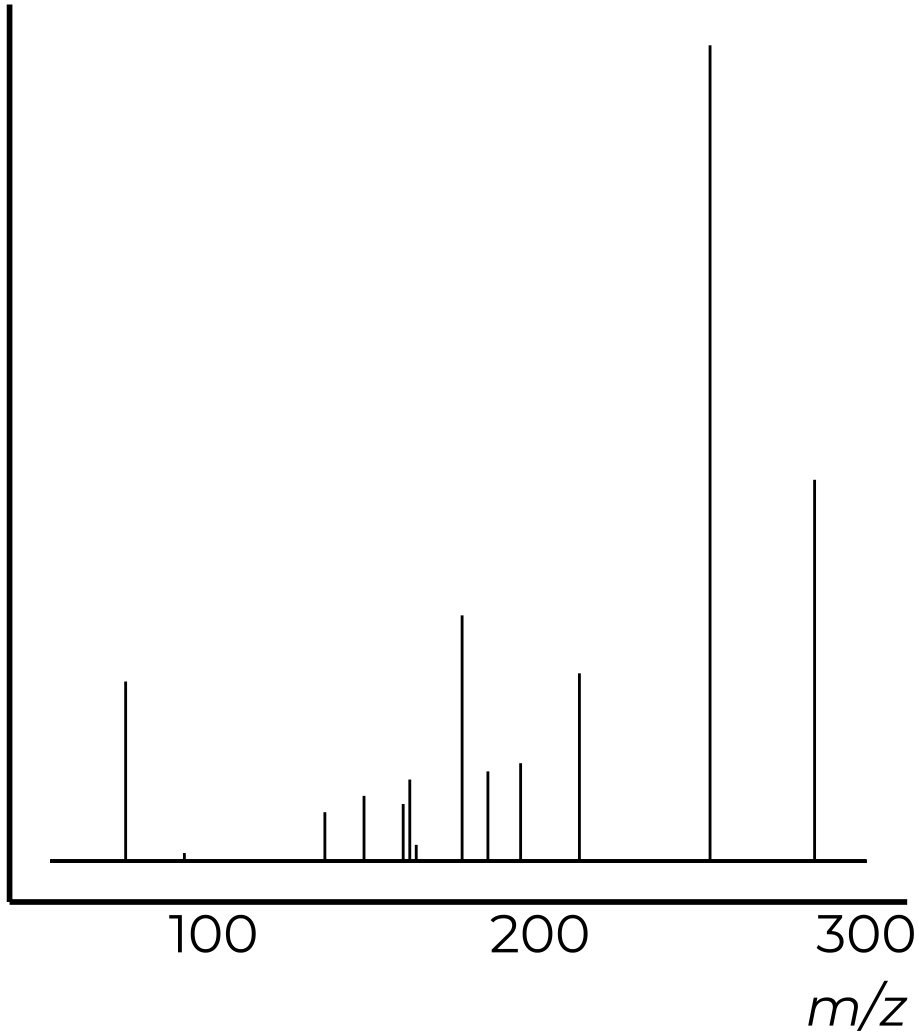


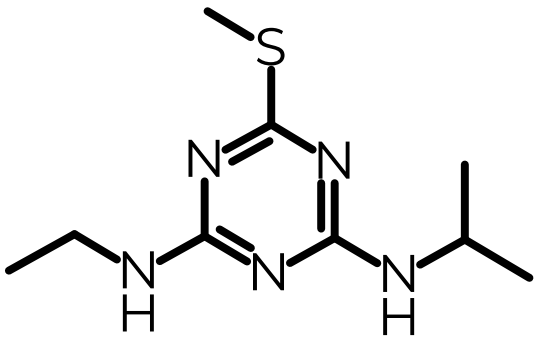


anneli kruve

anneli.kruve@su.se

toxicity assessment



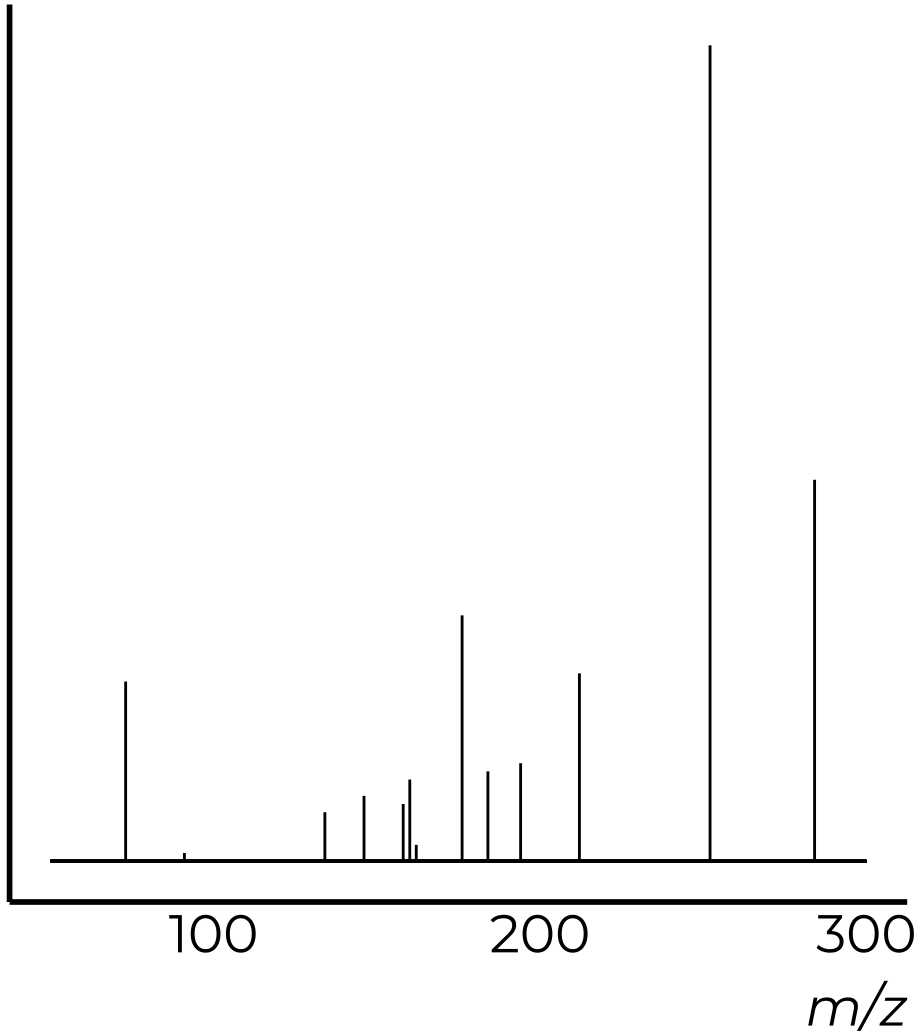


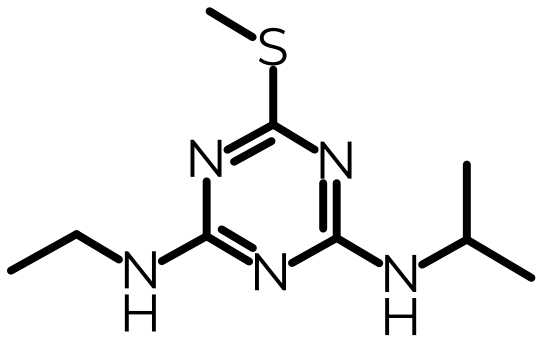
LC_{50} = 9.3 mg/L

anneli kruve

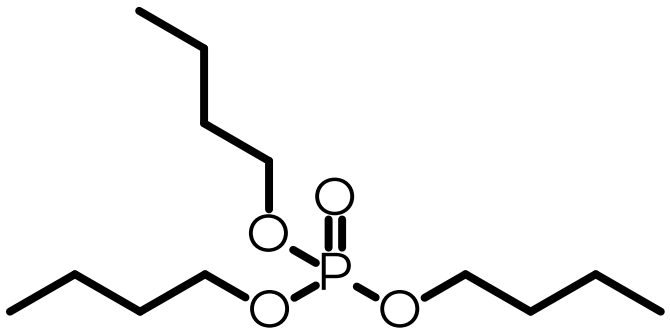
anneli.kruve@su.se

toxicity assessment





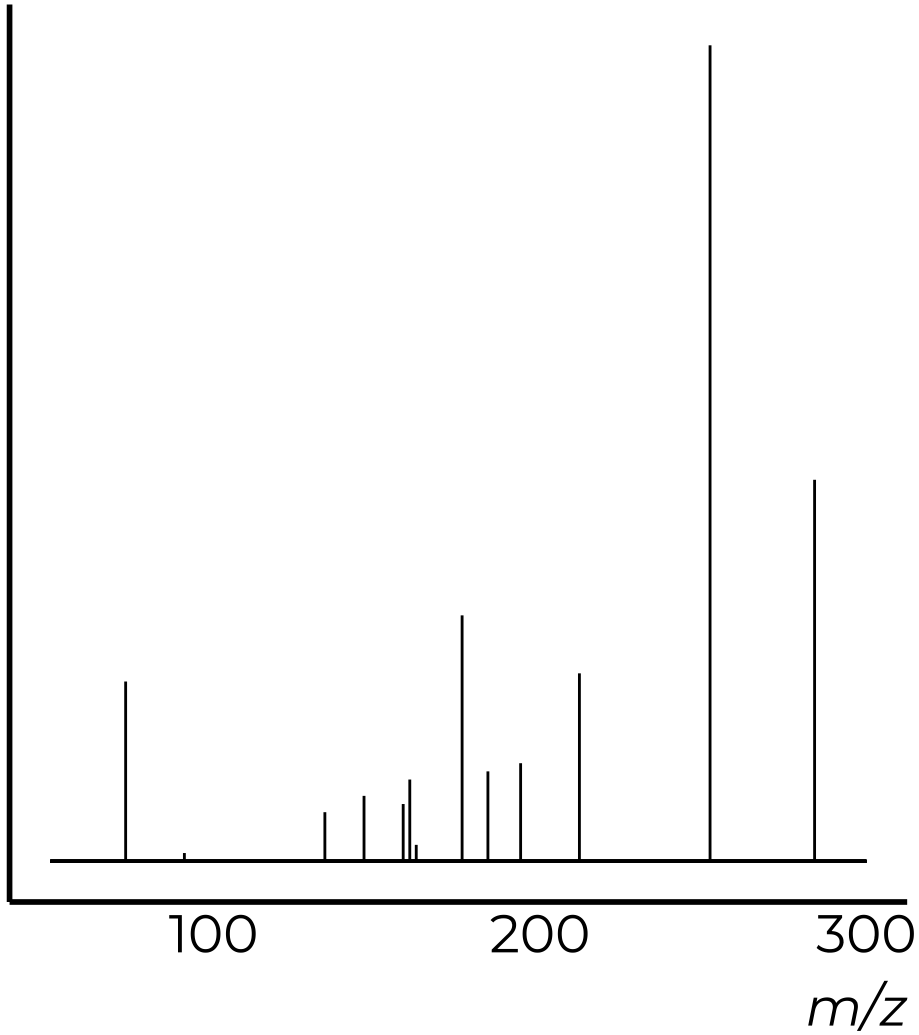
LC_{50} = 9.3 mg/L

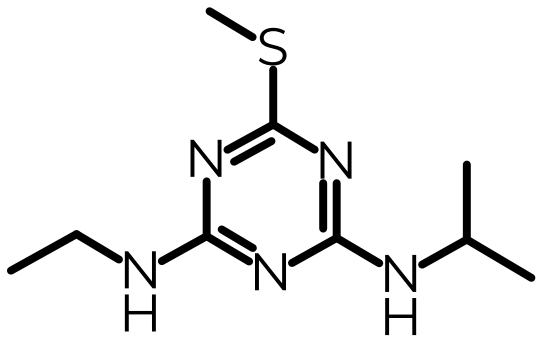


anneli kruve

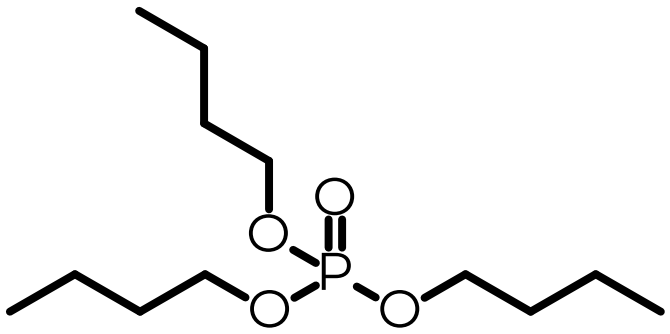
anneli.kruve@su.se

toxicity assessment





LC_{50} = 9.3 mg/L

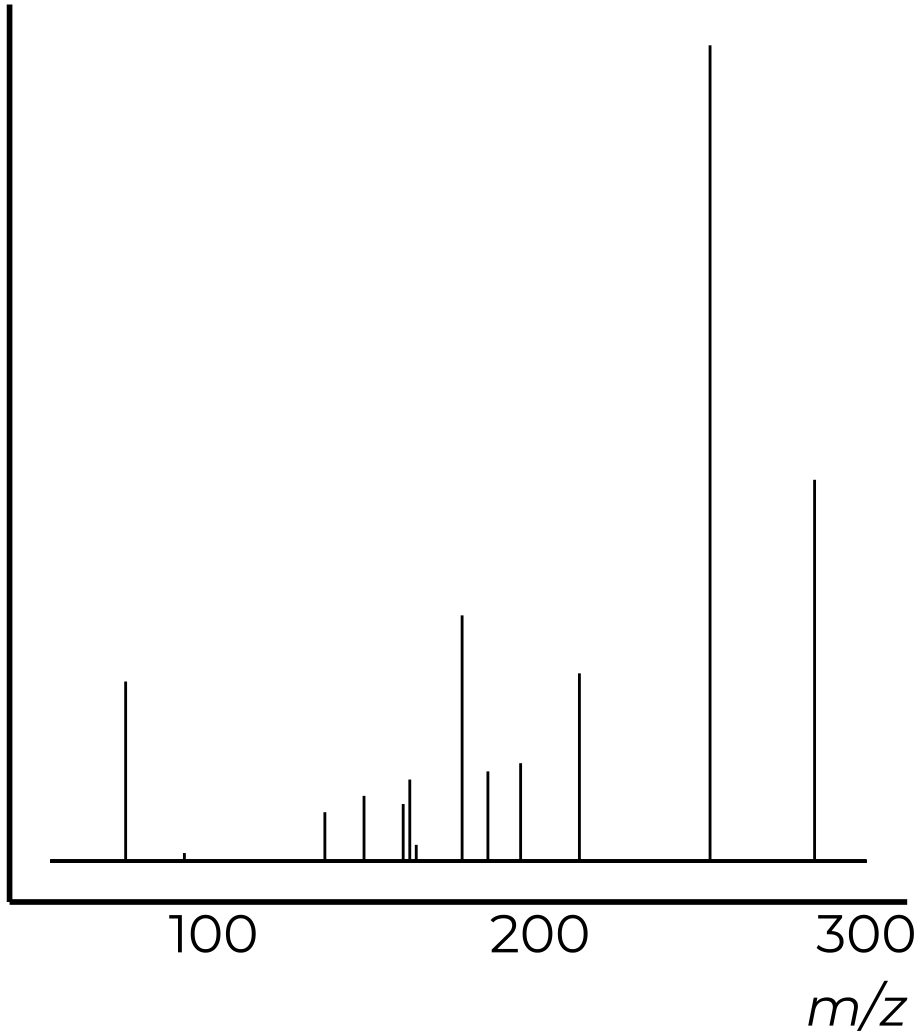


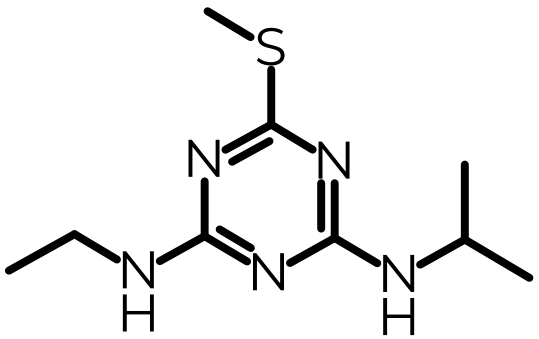
LC_{50} = ? mg/L

anneli kruve

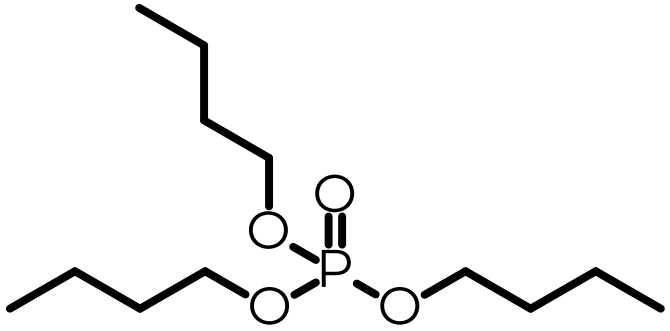
anneli.kruve@su.se

toxicity assessment





LC_{50} = 9.3 mg/L



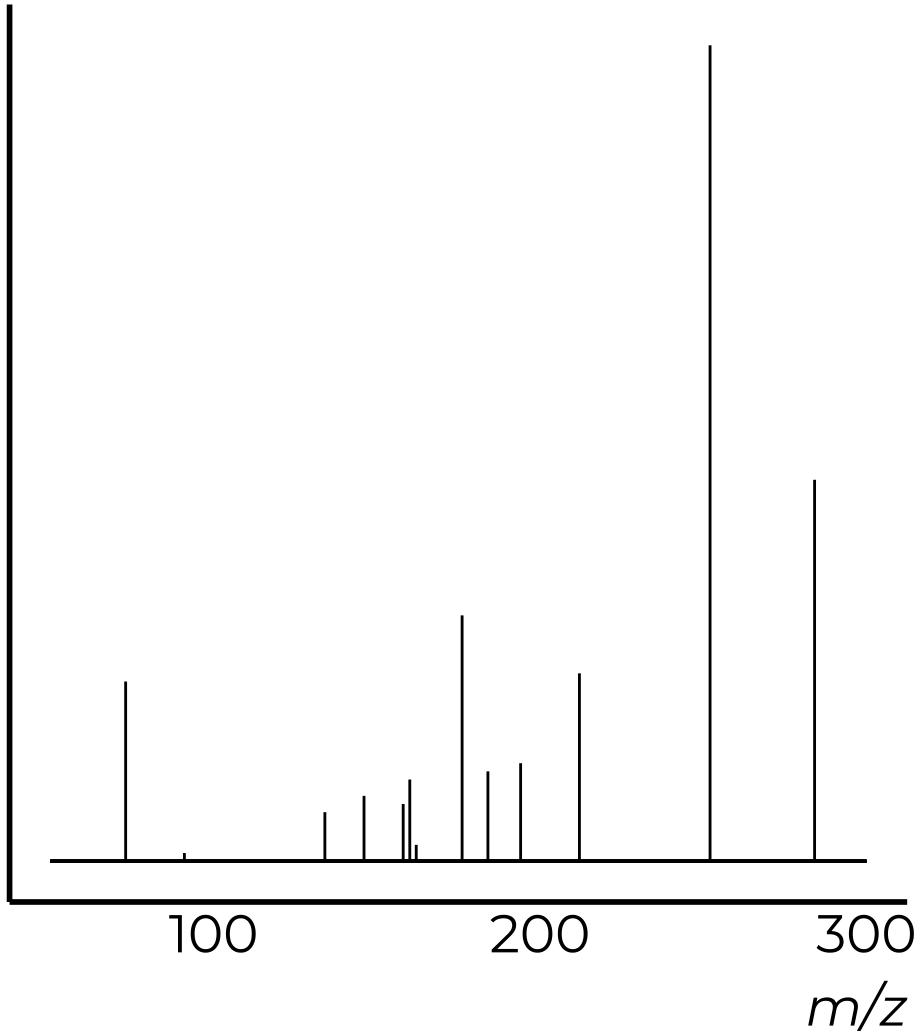
LC_{50} = ? mg/L

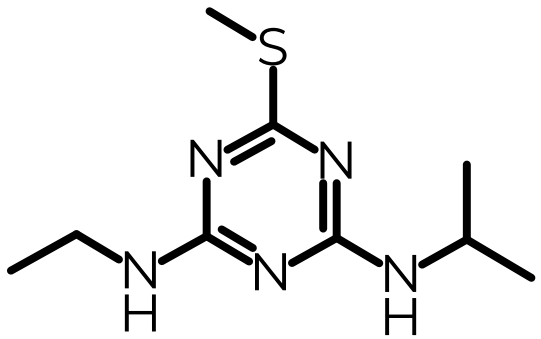
?

anneli kruve

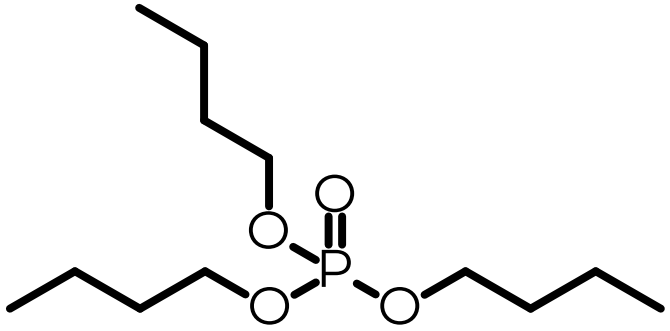
anneli.kruve@su.se

toxicity assessment





LC_{50} = 9.3 mg/L



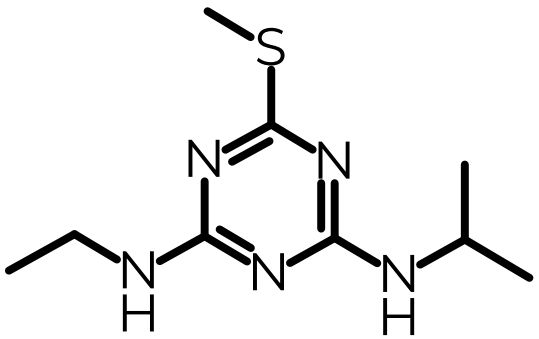
LC_{50} = ? mg/L

LC_{50} = ? mg/L

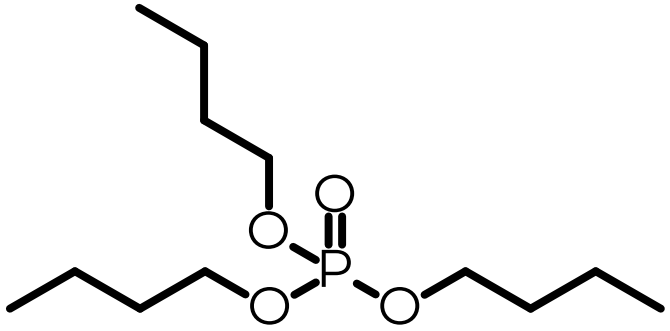
anneli kruve

anneli.kruve@su.se

toxicity assessment



LC_{50} = 9.3 mg/L



LC_{50} = ? mg/L

?

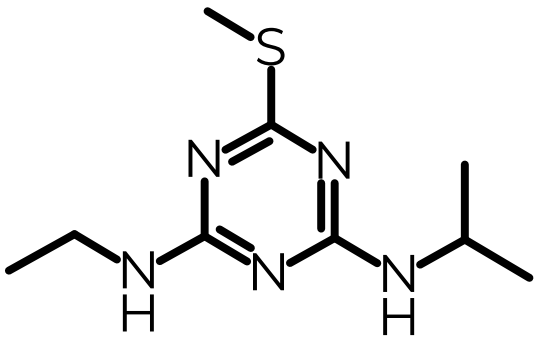
LC_{50} = ? mg/L

anneli kruve

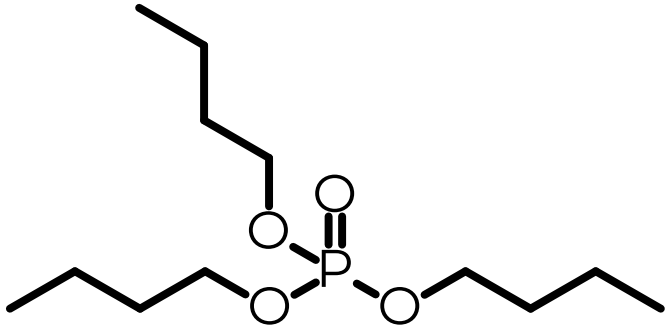
anneli.kruve@su.se

toxicity assessment

<1%



LC_{50} = 9.3 mg/L



LC_{50} = ? mg/L

?

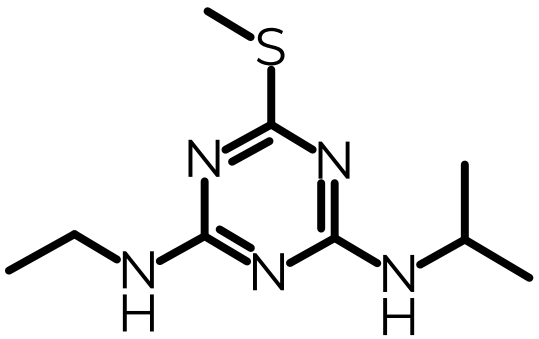
LC_{50} = ? mg/L

anneli kruve

anneli.kruve@su.se

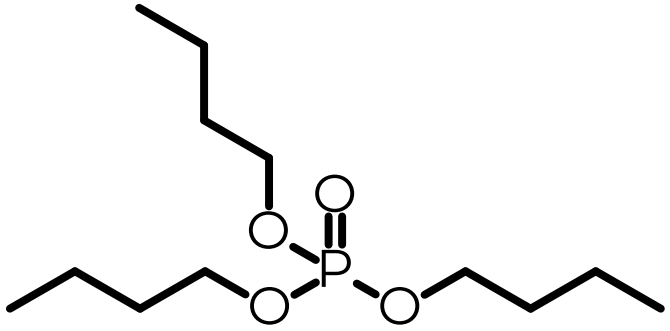
toxicity assessment

<1%



LC_{50} = 9.3 mg/L

<2%



LC_{50} = ? mg/L

?

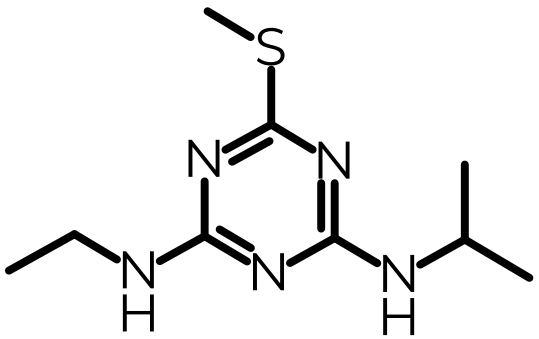
LC_{50} = ? mg/L

anneli kruve

anneli.kruve@su.se

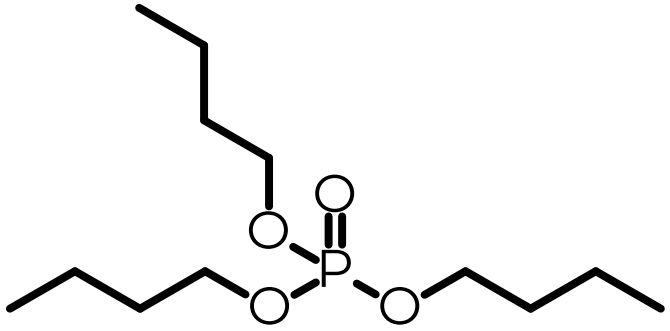
toxicity assessment

<1%



LC_{50} = 9.3 mg/L

<2%



LC_{50} = ? mg/L

~98%

?

LC_{50} = ? mg/L

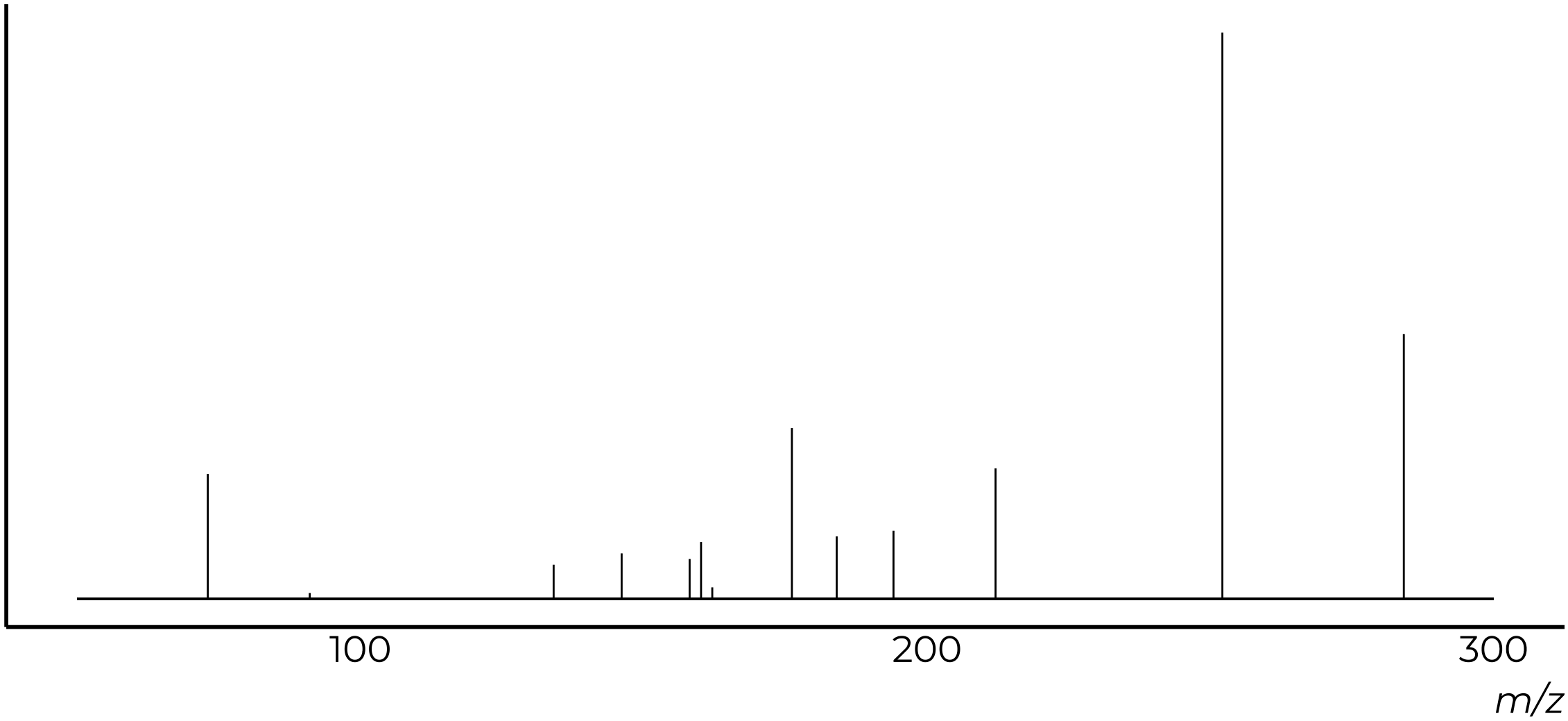
anneli kruve

anneli.kruve@su.se

information available

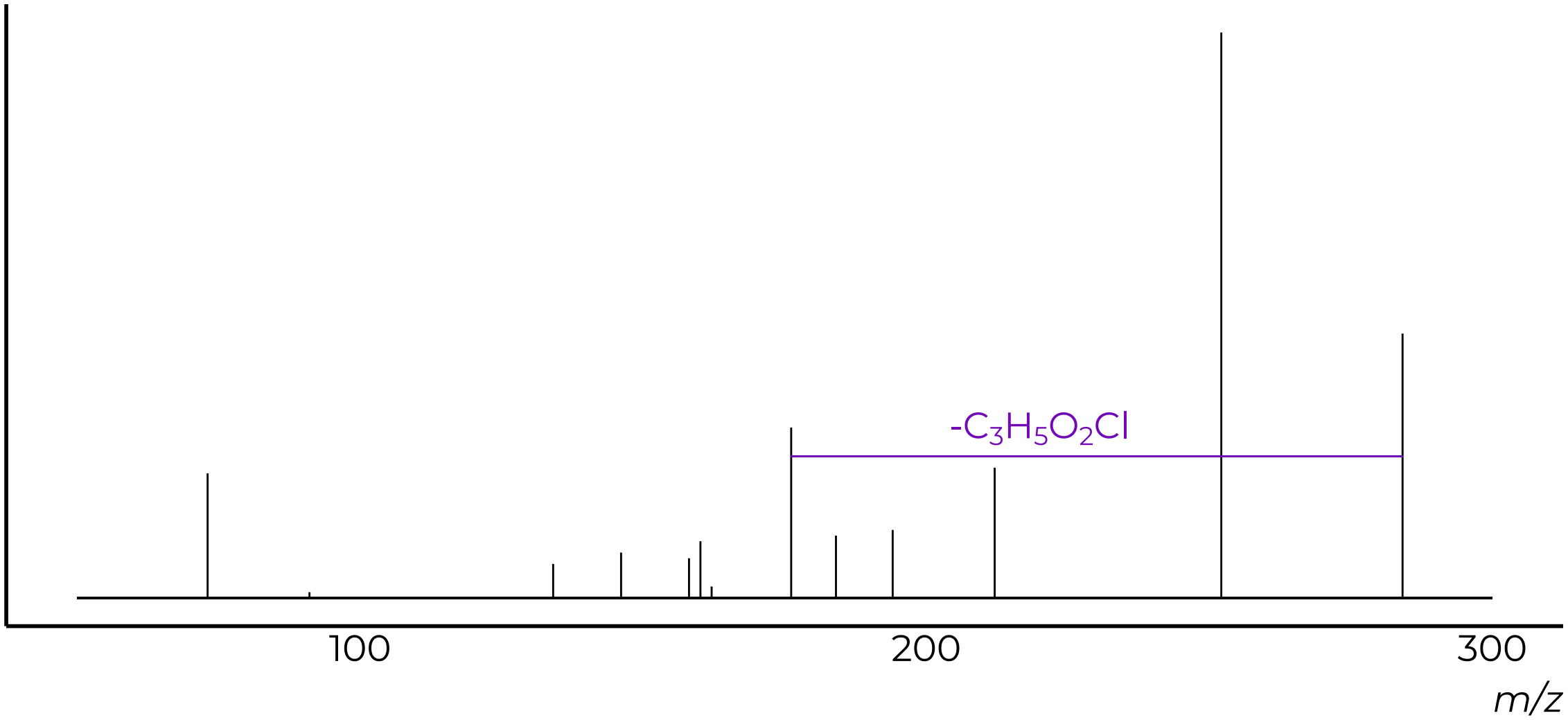
in MS^2 spectra

MS^2 spectra



anneli kruve
anneli.kruve@su.se

MS^2 spectra



anneli kruve
anneli.kruve@su.se

modelling toxicity

selected endpoint

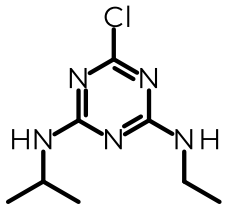
anneli.kruve@su.se

selected endpoint

fathead minnow, bluegill and rainbow trout

anneli.kruve@su.se

workflow

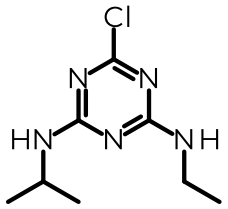


structure as SMILES

anneli kruve

anneli.kruve@su.se

workflow



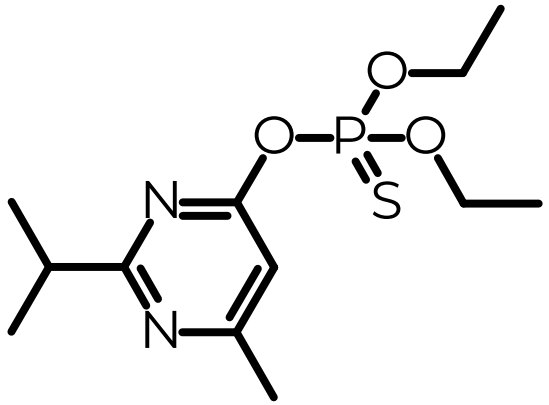
structure as SMILES

molecular descriptors

anneli kruve

anneli.kruve@su.se

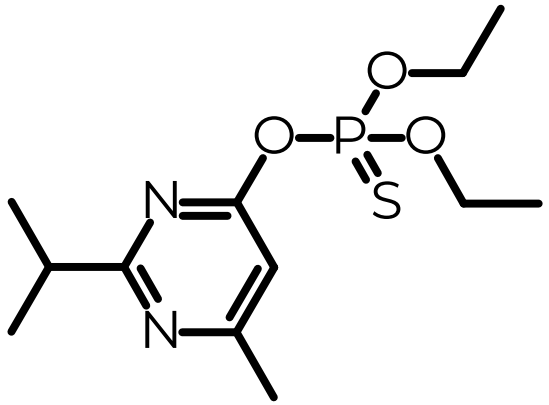
structural fingerprints



anneli kruve

anneli.kruve@su.se

structural fingerprints

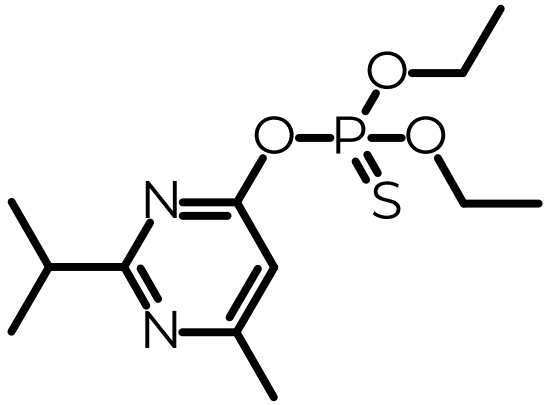


R: rcdk

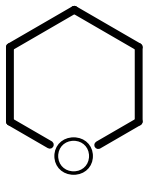
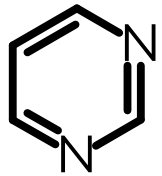
anneli kruve

anneli.kruve@su.se

structural fingerprints



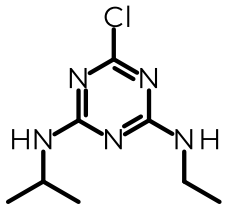
R: rcdk

0	
1	
1	
0	
1	

anneli kruve

anneli.kruve@su.se

workflow



structure as SMILES

molecular descriptors

machine learning for predicting LC_{50}

anneli kruve

anneli.kruve@su.se

model training

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

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

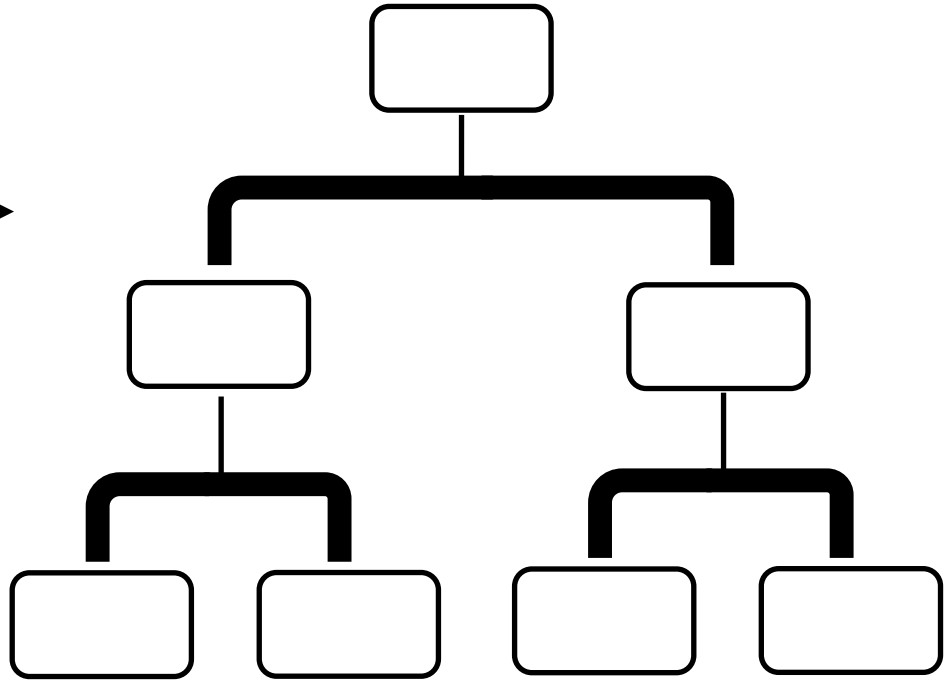
model training

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

gradient
boosting



anneli.kruve@su.se

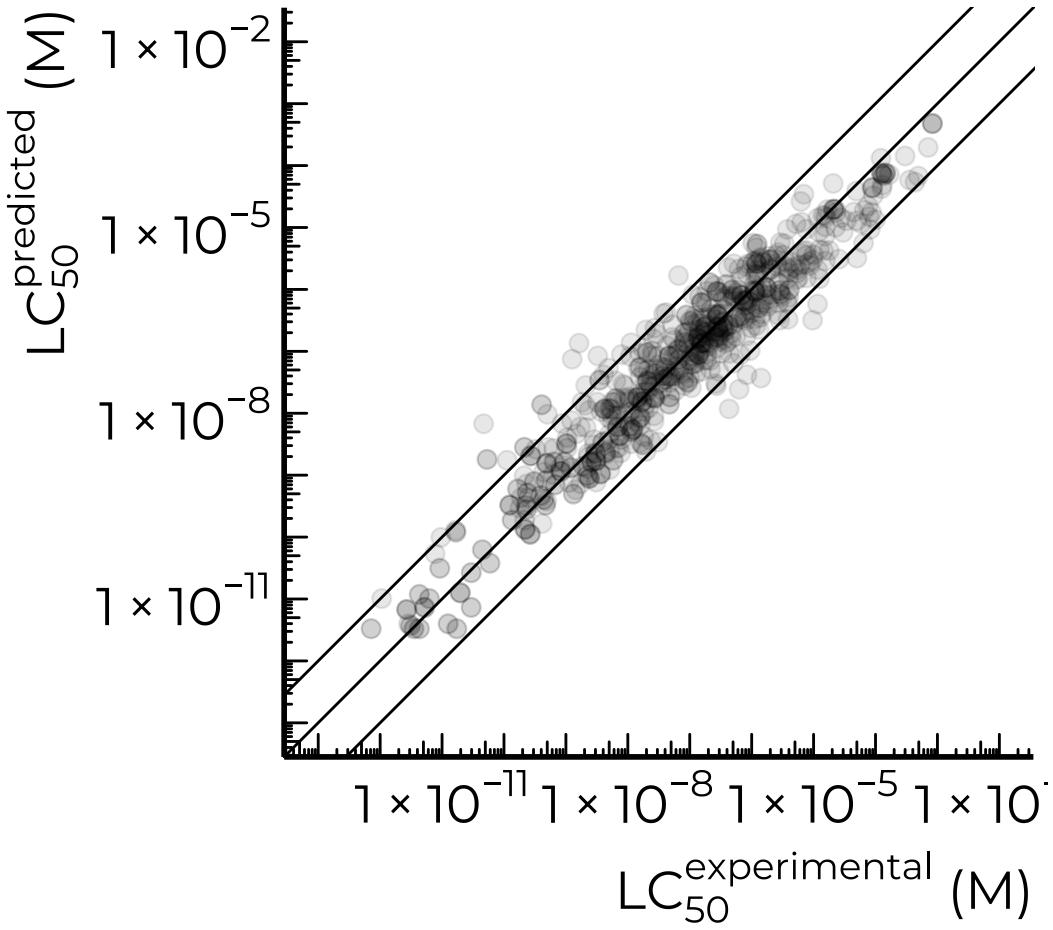
performance

of LC_{50} predictions with molecular fingerprints

LC_{50} predictions

Peets et al. ES&T 2022

fish LC_{50}



training set

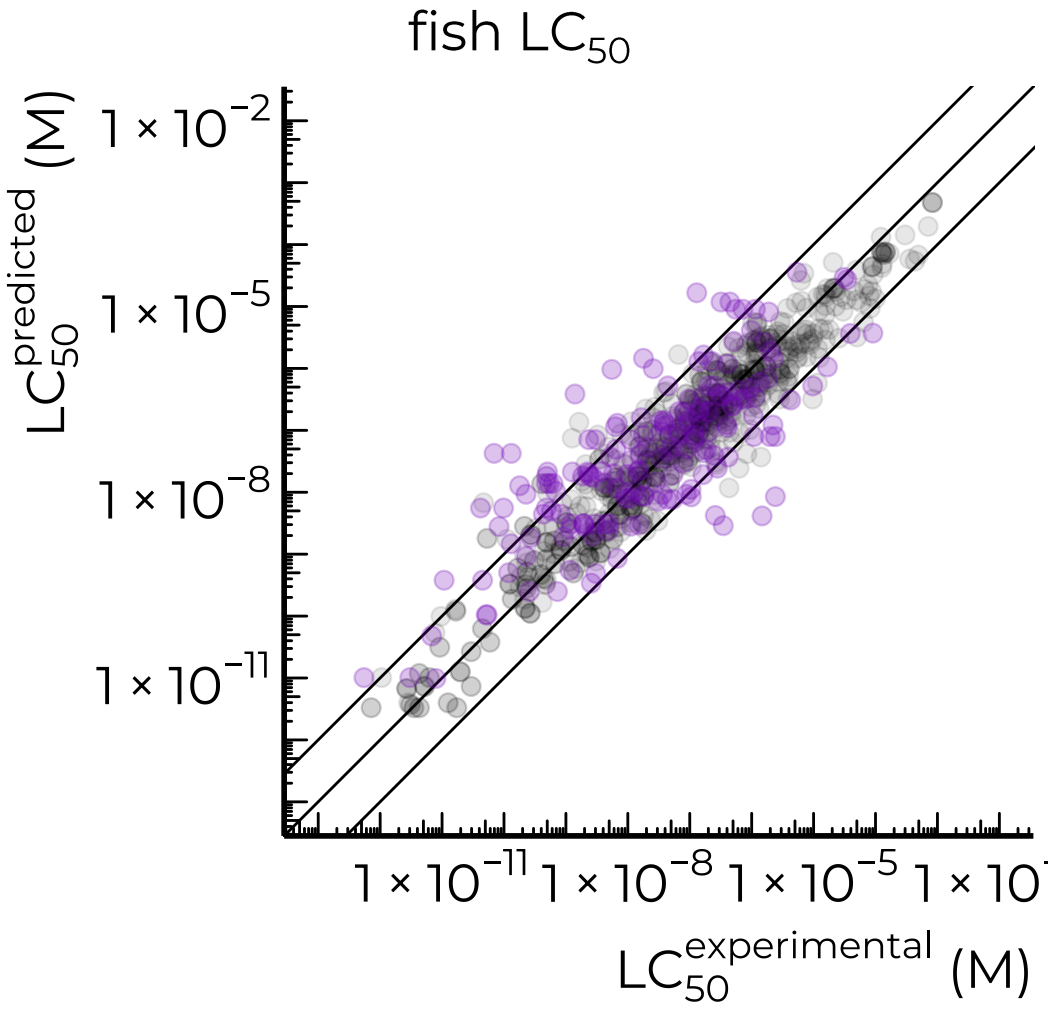
RMSE 0.52 log(M)

anneli kruve

anneli.kruve@su.se

LC_{50} predictions

Peets et al. ES&T 2022



training set

RMSE 0.52 log(M)

test set

RMSE 0.78 log(M)

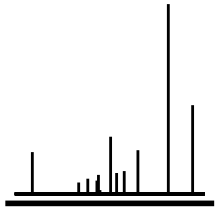
anneli kruve

anneli.kruve@su.se

unidentified chemicals

from MS^2 spectra

workflow



MS^2 spectra

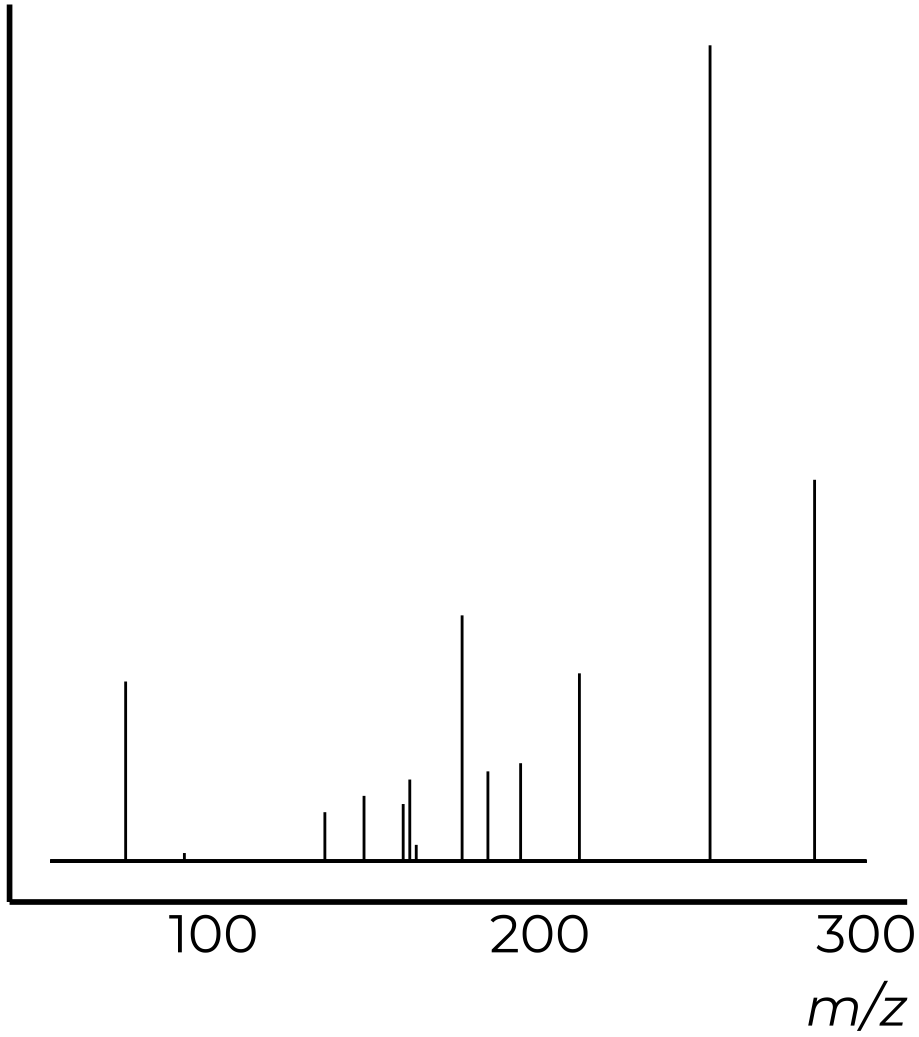
molecular fingerprints with SIRIUS+CSI:FingerID

predict LC_{50}

anneli kruve

anneli.kruve@su.se

predict for unknown chemicals

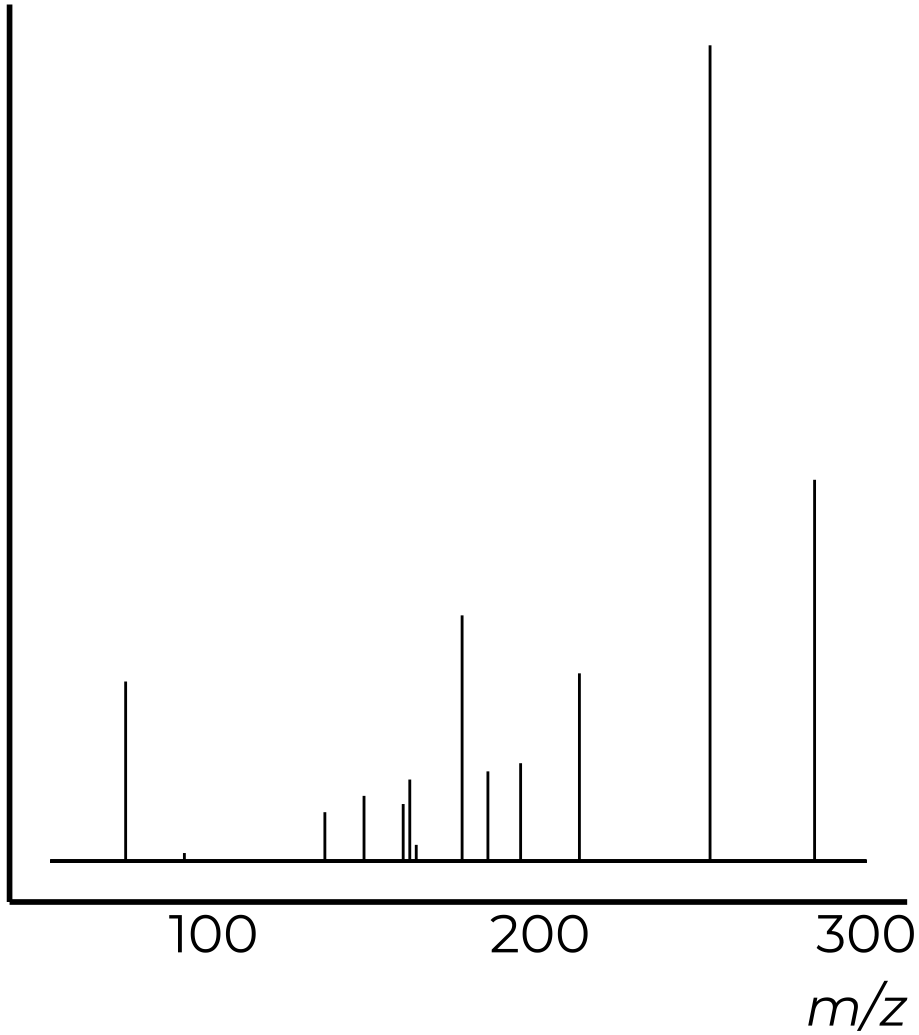


?

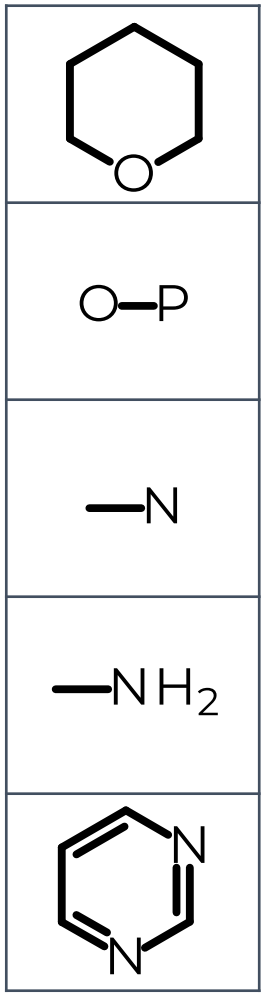
anneli kruve

anneli.kruve@su.se

predict for unknown chemicals



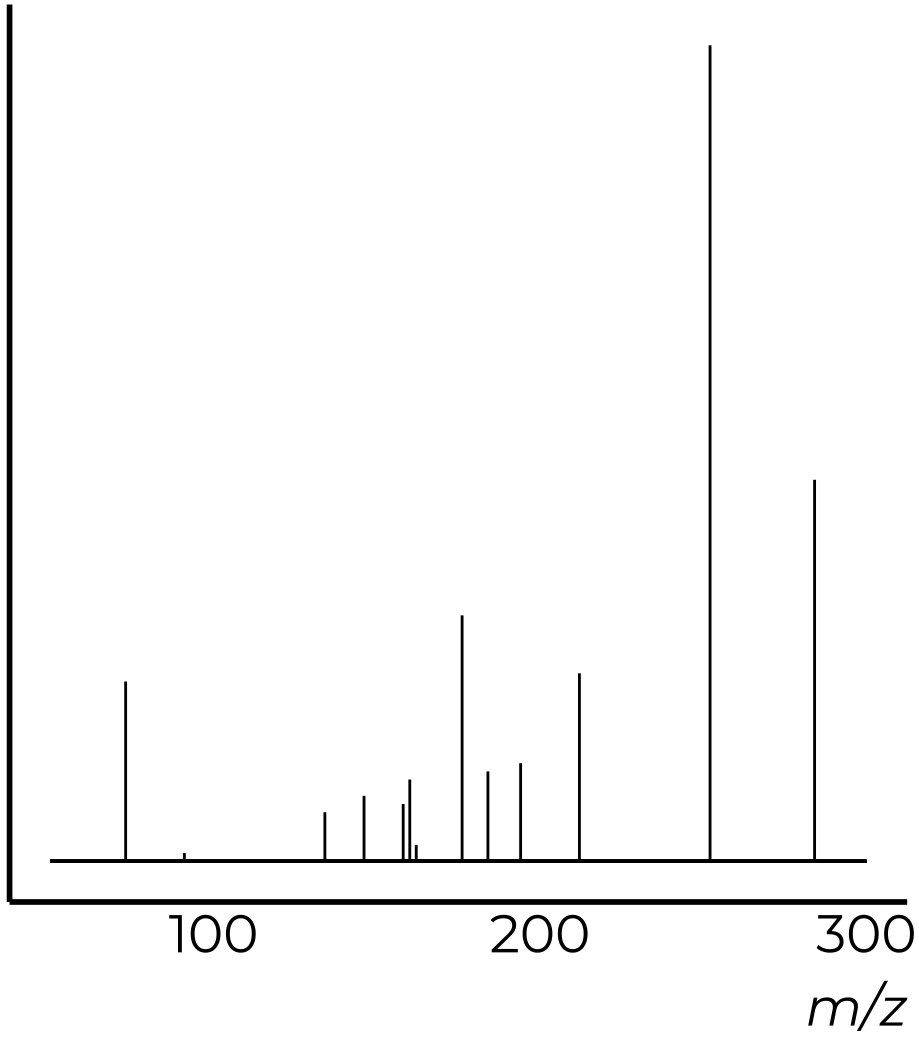
SIRIUS+
CSI:FingerID



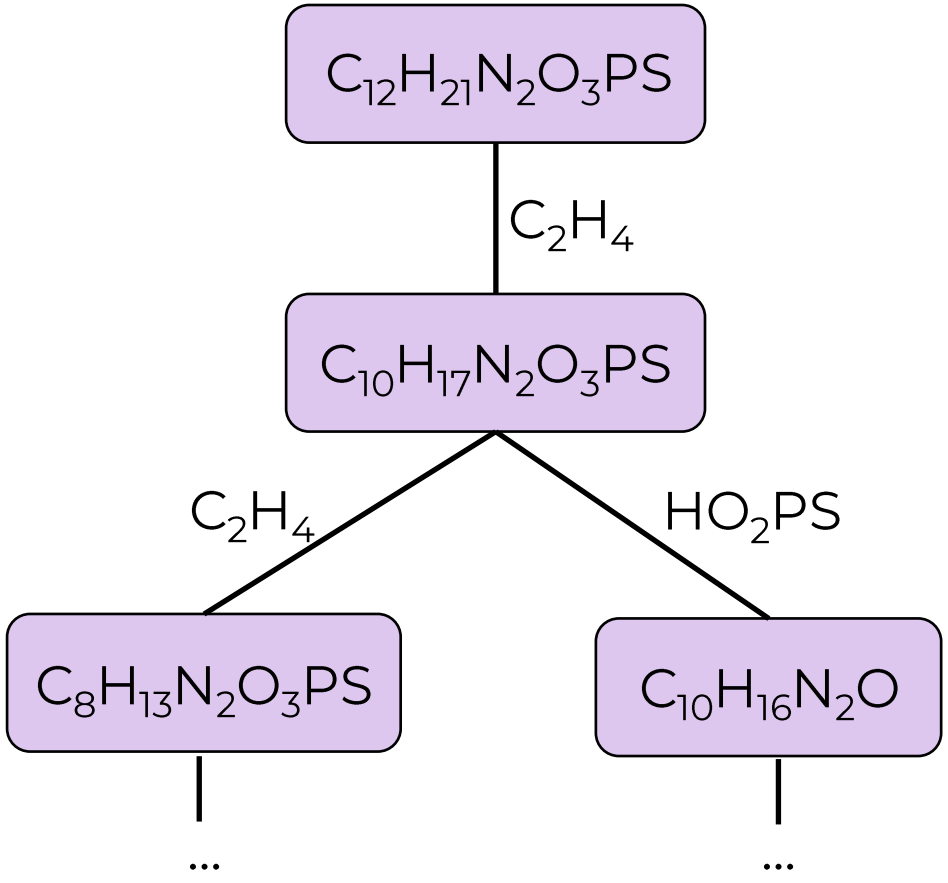
anneli kruve

anneli.kruve@su.se

predict for unknown chemicals



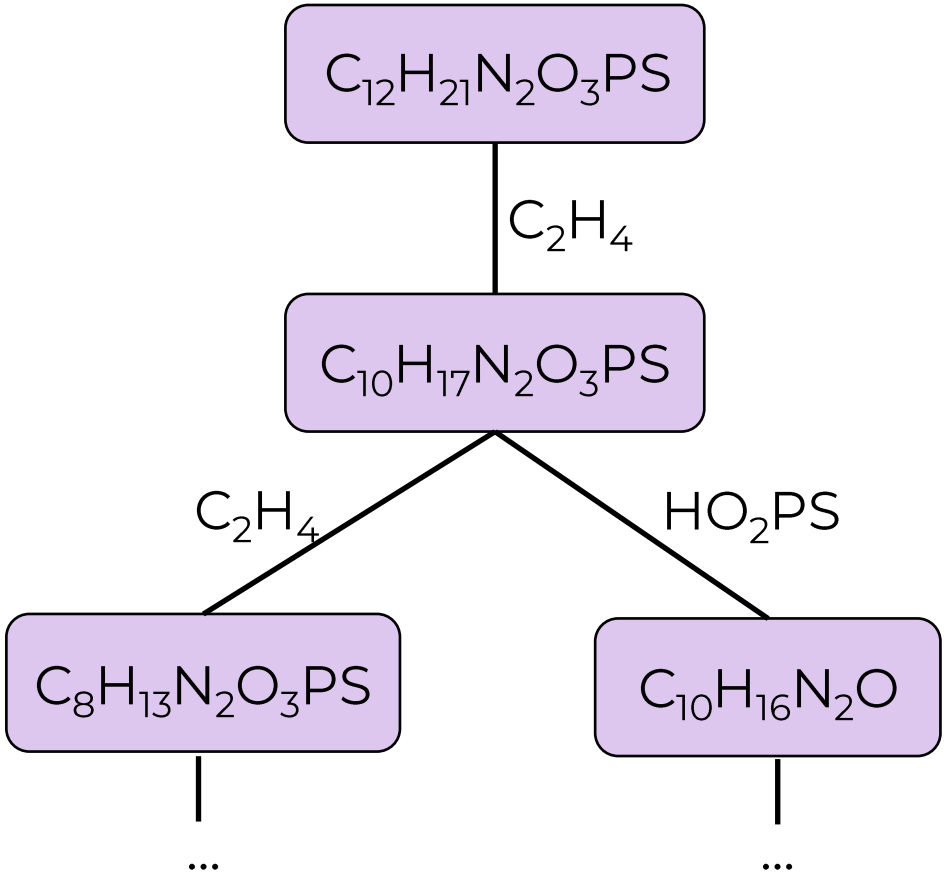
SIRIUS+
CSI:FingerID



anneli kruve

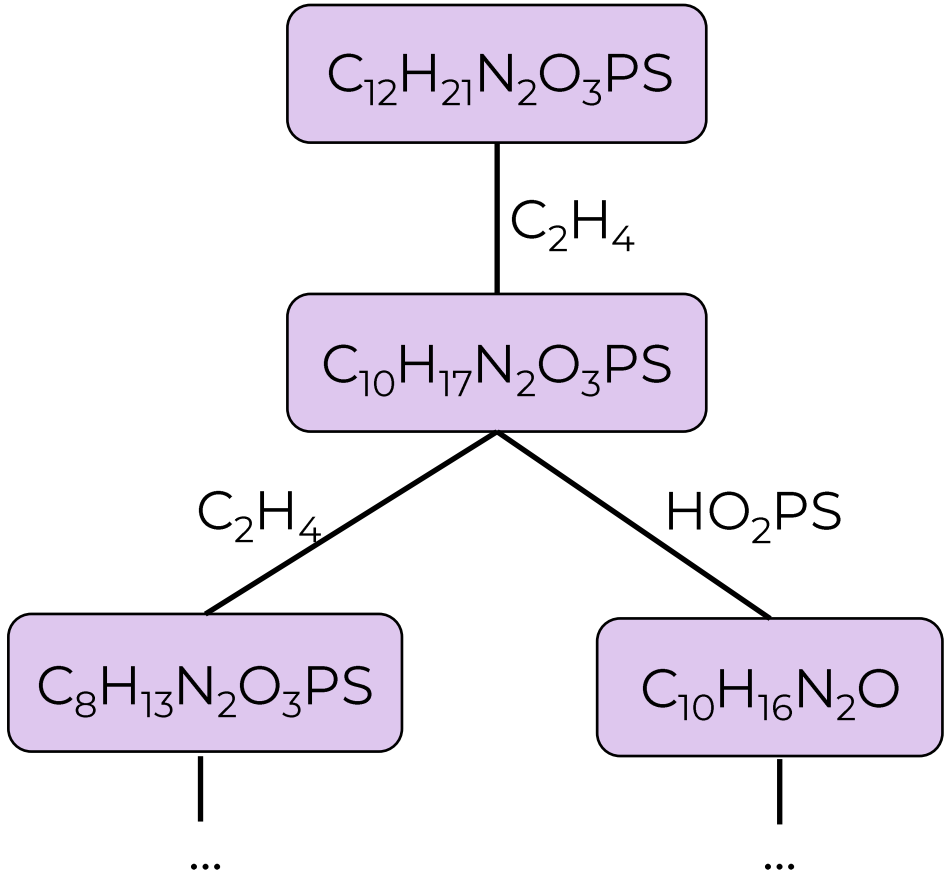
anneli.kruve@su.se

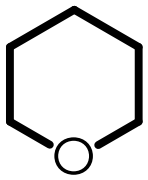
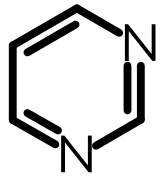
predict for unknown chemicals



anneli.kruve@su.se

predict for unknown chemicals

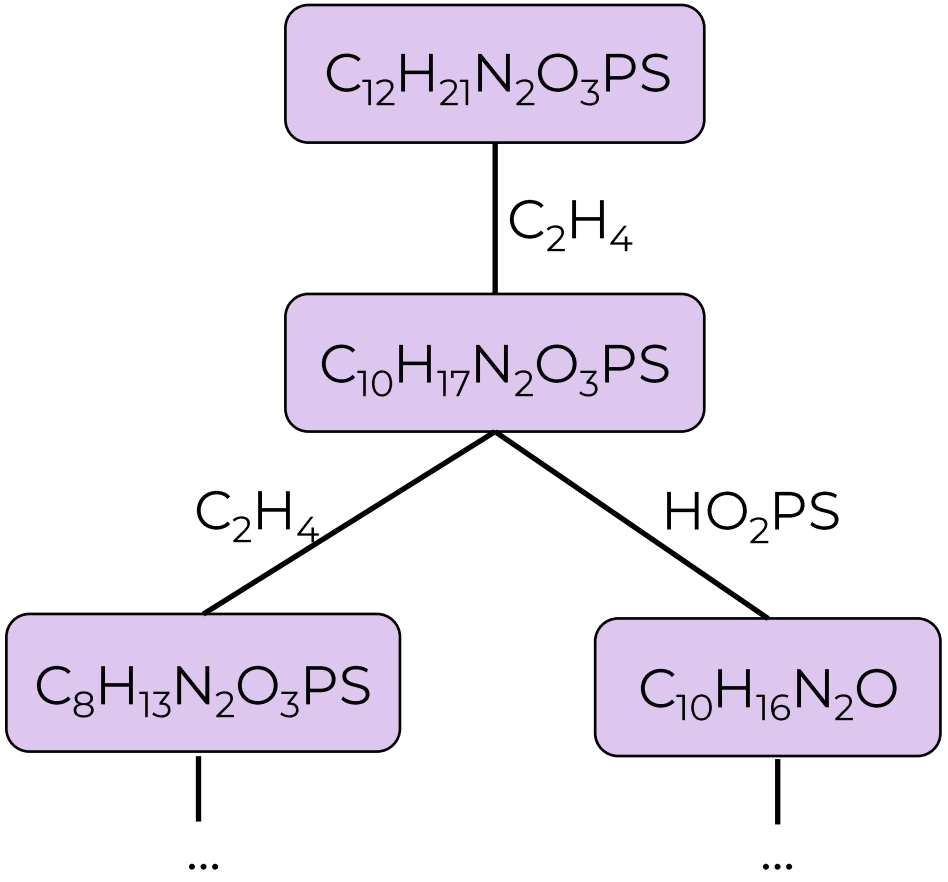


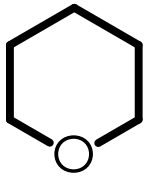
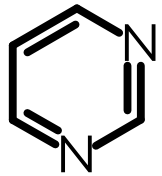
0.001	
0.999	
0.999	
0.198	
0.988	

anneli kruve

anneli.kruve@su.se

predict for unknown chemicals

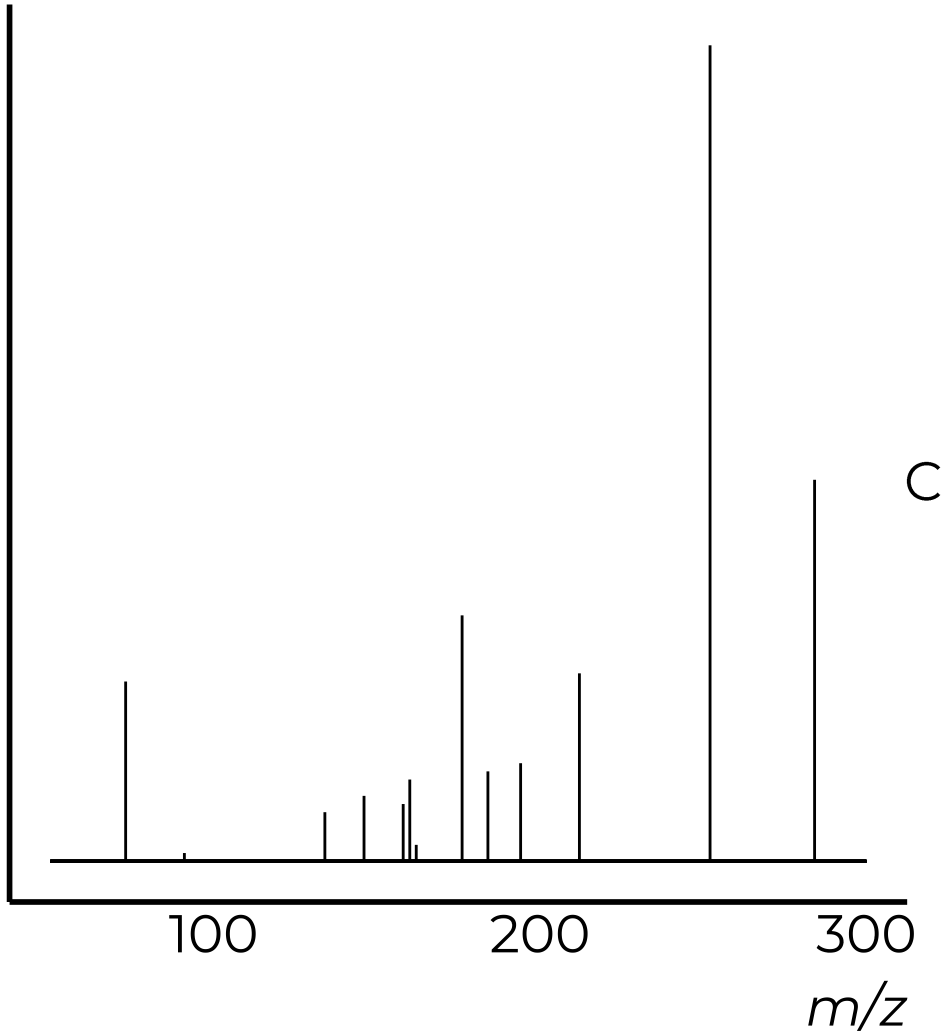


0	
1	
1	
0	
1	

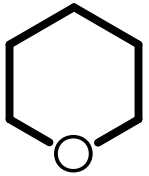
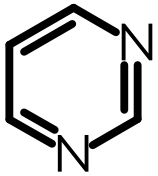
anneli kruve

anneli.kruve@su.se

predict for unknown chemicals



SIRIUS+
CSI:FingerID

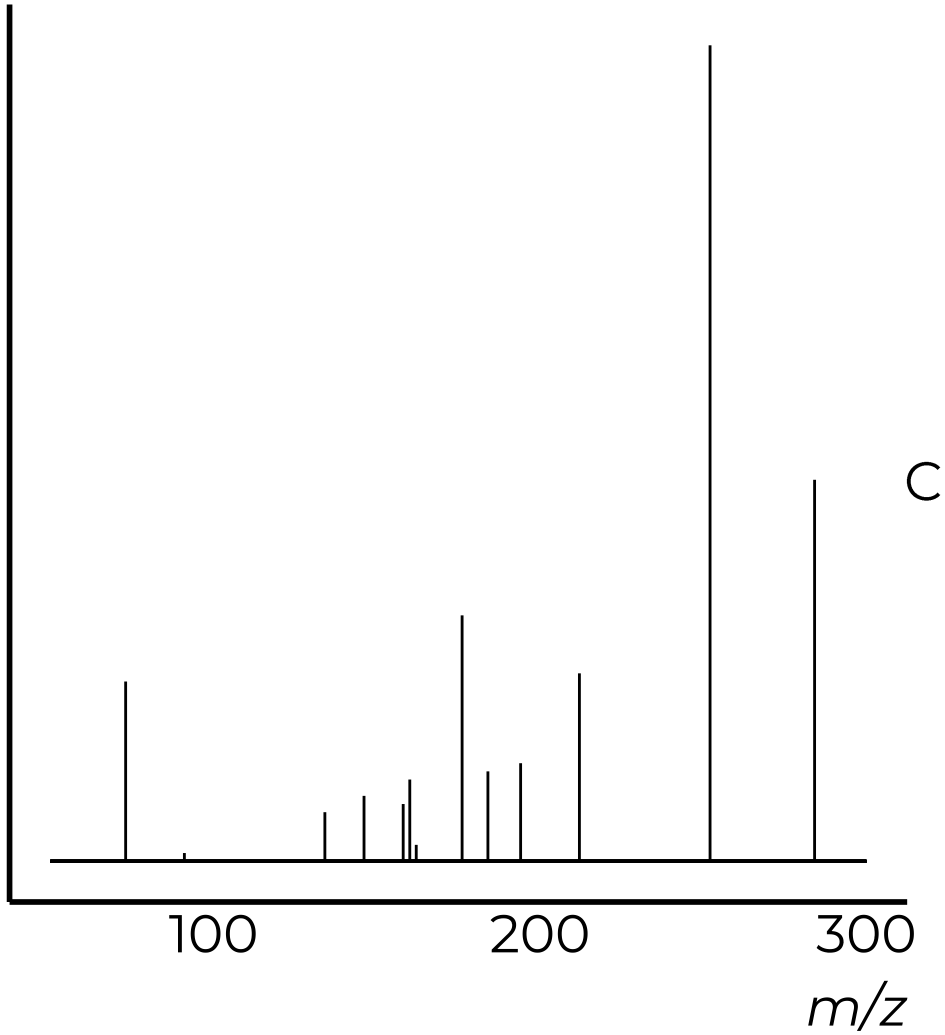
0	
1	
1	
0	
1	

gradient
boosting

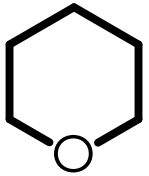
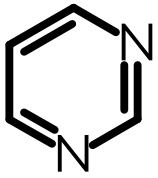
anneli kruve

anneli.kruve@su.se

predict for unknown chemicals



SIRIUS+
CSI:FingerID

0	
1	
1	
0	
1	

gradient
boosting

LC_{50} =
-2.2 log(mM)

anneli kruve

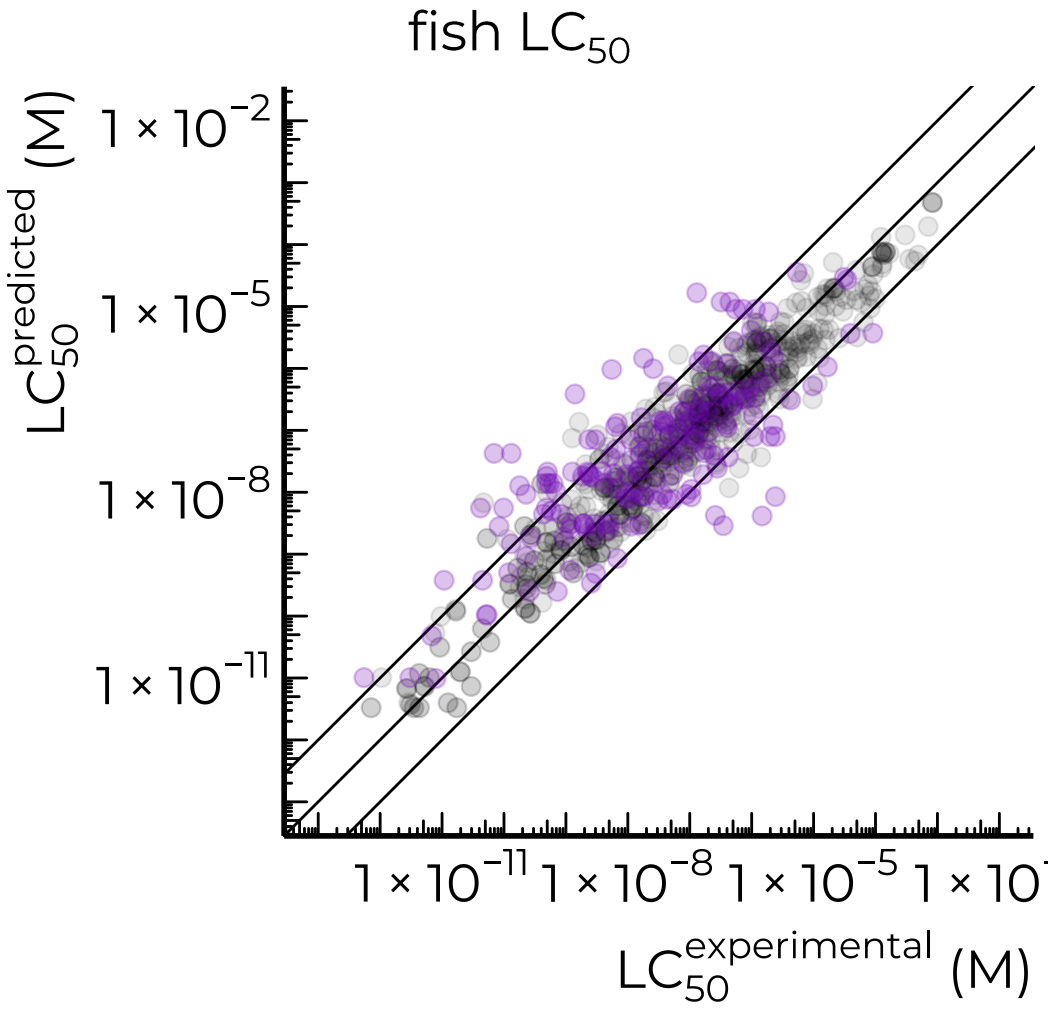
anneli.kruve@su.se

LC_{50} predictions

anneli.kruve@su.se

LC_{50} predictions

Peets et al. ES&T 2022



test set on structures

RMSE 0.78 log(M)

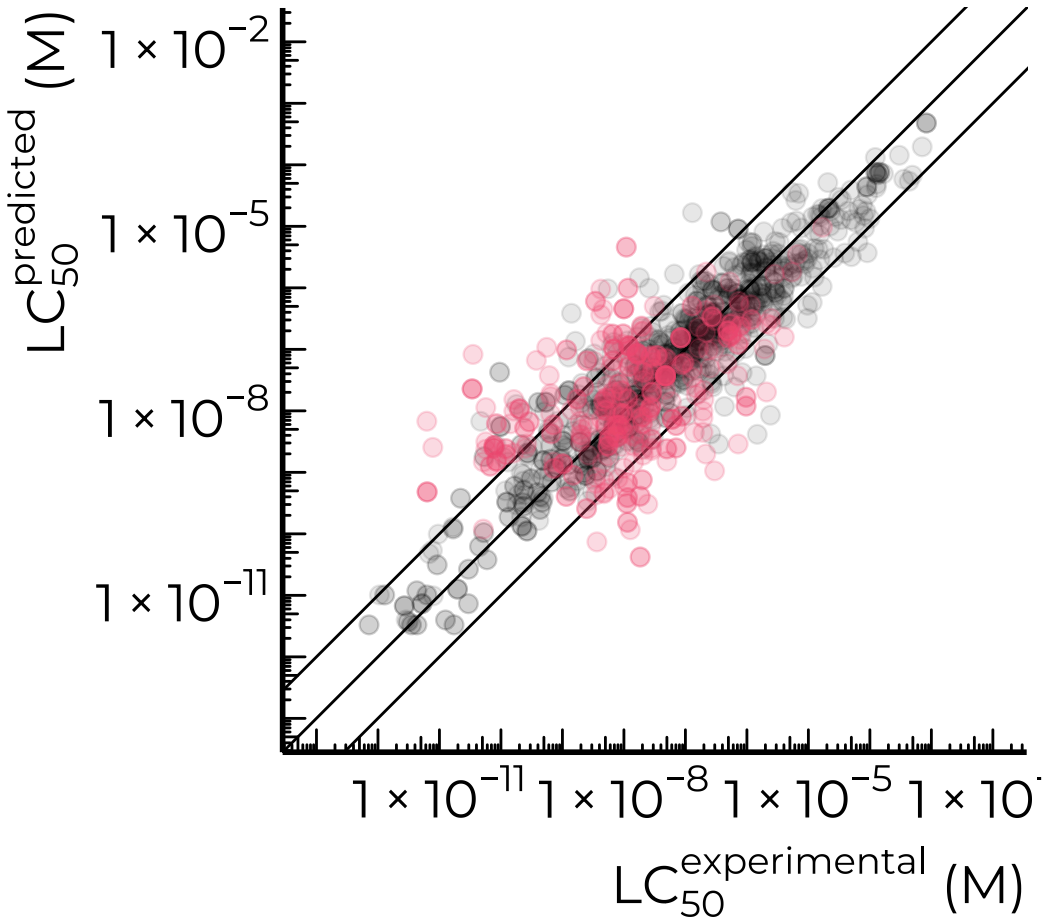
anneli kruve

anneli.kruve@su.se

LC_{50} predictions

Peets et al. ES&T 2022

fish LC_{50}



test set on structures

RMSE 0.78 log(M)

validation on MassBank

$RMSE_{model}$ 0.88 log(M)

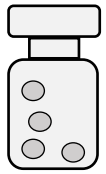
$SD_{experimental}$ 0.44 log(mM)

anneli kruve

anneli.kruve@su.se

pinpointing toxic chemicals

case study on wastewater

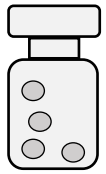


wastewater samples

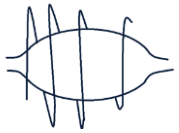
anneli kruve

anneli.kruve@su.se

case study on wastewater



wastewater samples

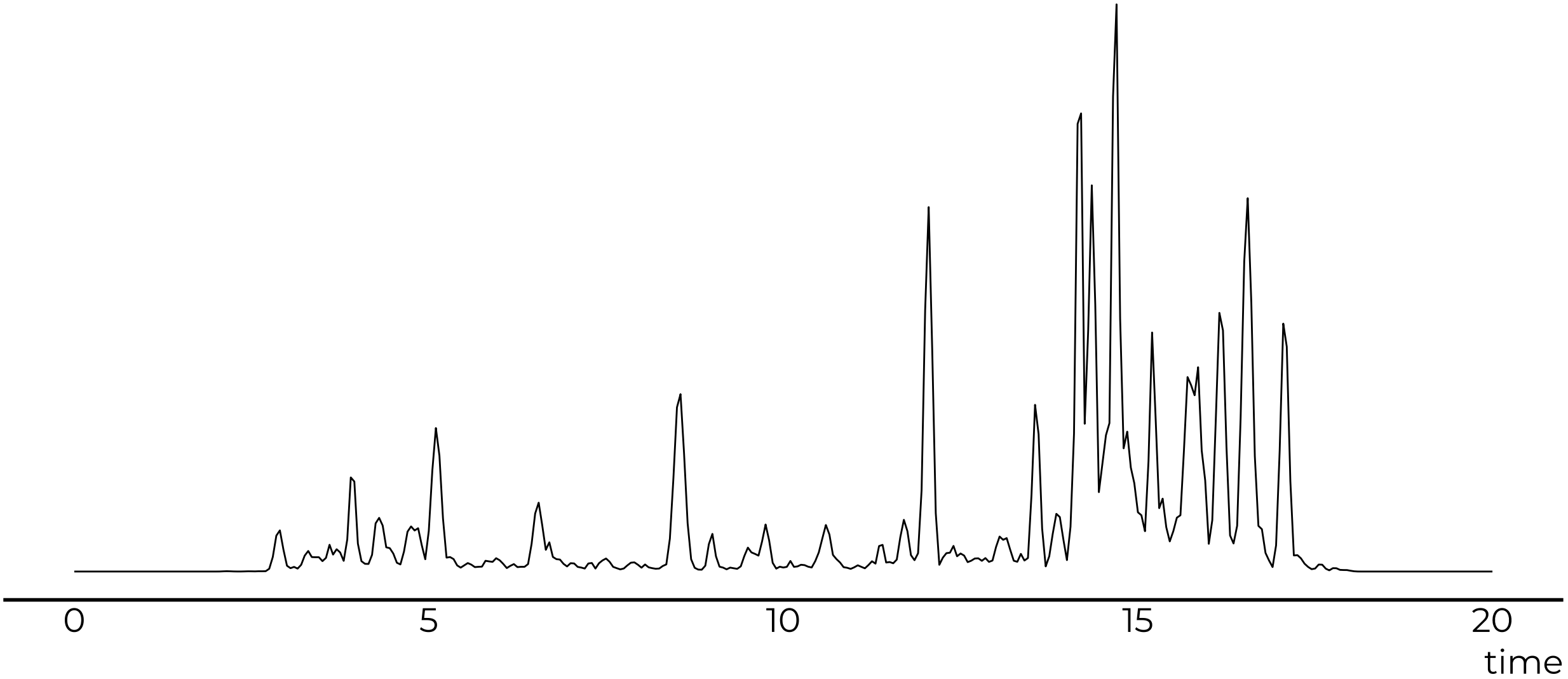


LC/HRMS analysis

anneli kruve

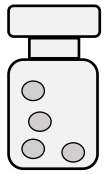
anneli.kruve@su.se

case study on wastewater

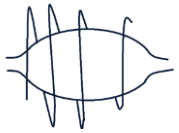


anneli kruve
anneli.kruve@su.se

case study on wastewater



wastewater samples



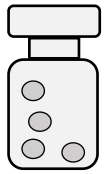
LC/HRMS analysis

MS-DIAL peak picking

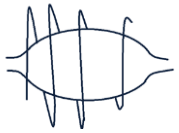
anneli kruve

anneli.kruve@su.se

case study on wastewater



wastewater samples



LC/HRMS analysis

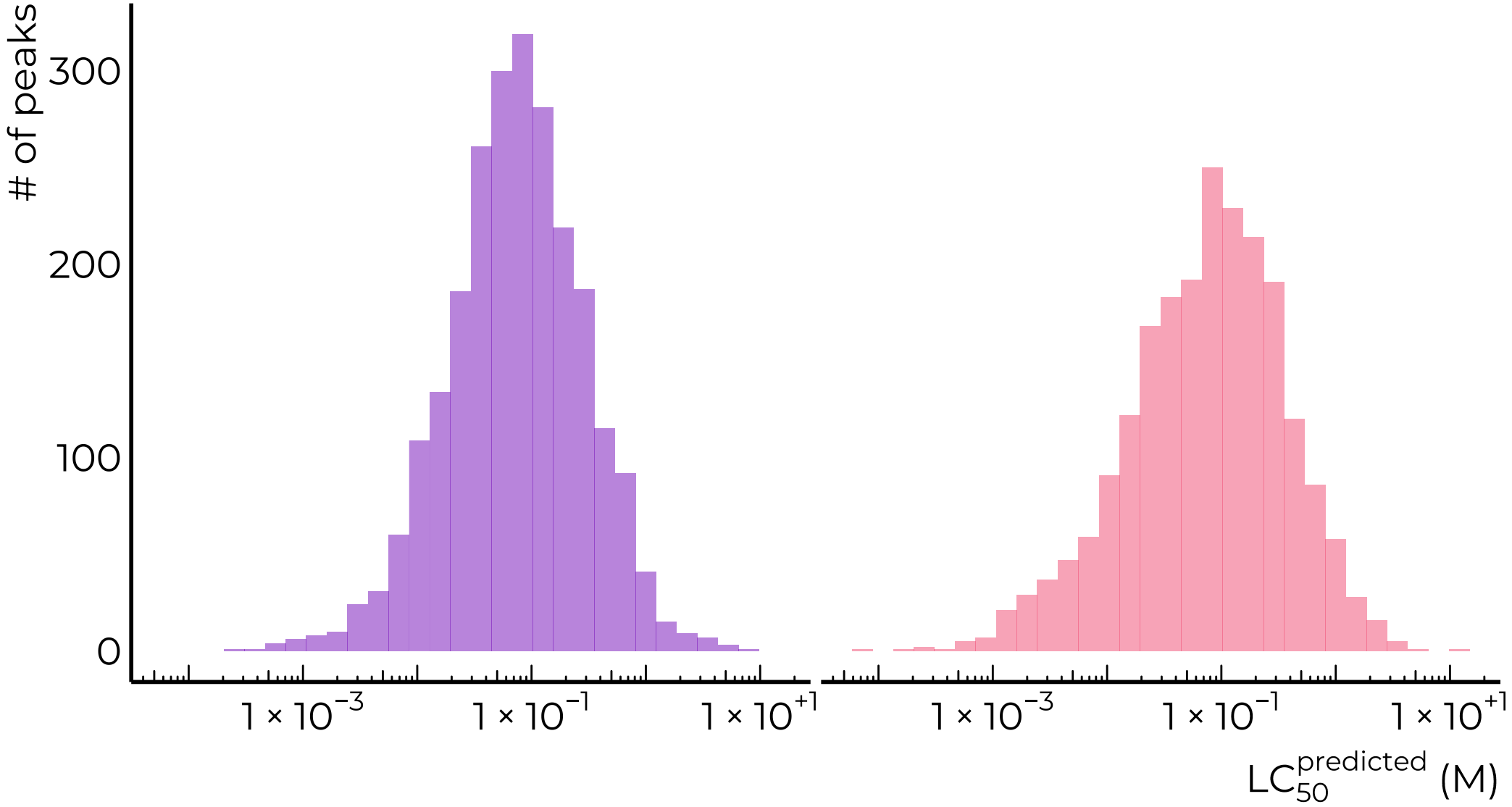
MS-DIAL peak picking

SIRIUS+CSI:FingerID

anneli kruve

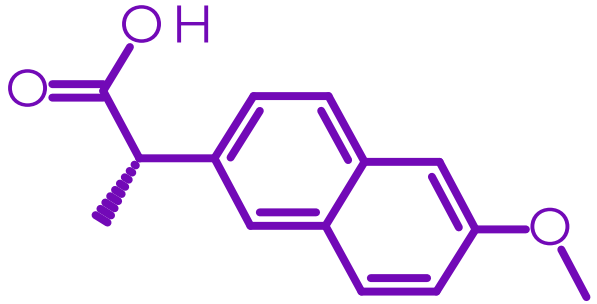
anneli.kruve@su.se

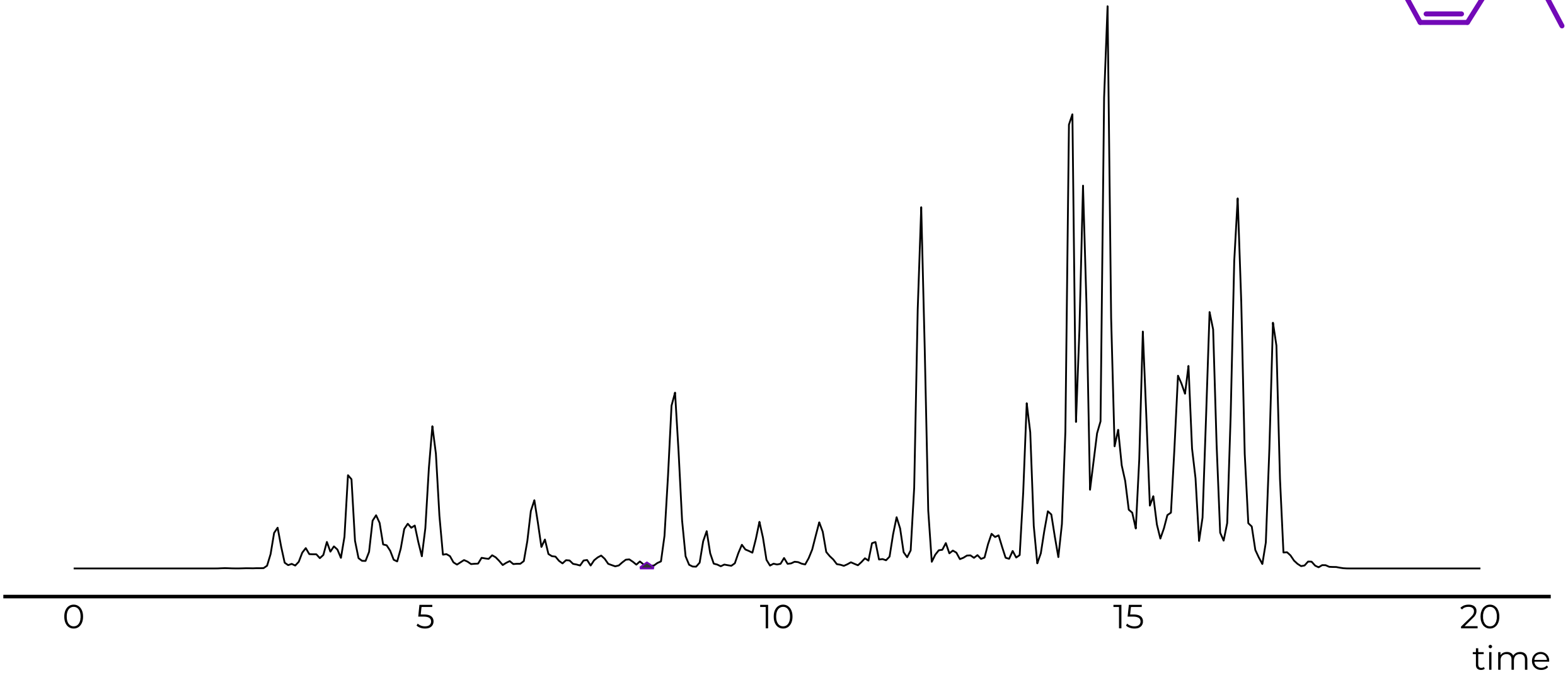
LC_{50} distribution



anneli kruve
anneli.kruve@su.se

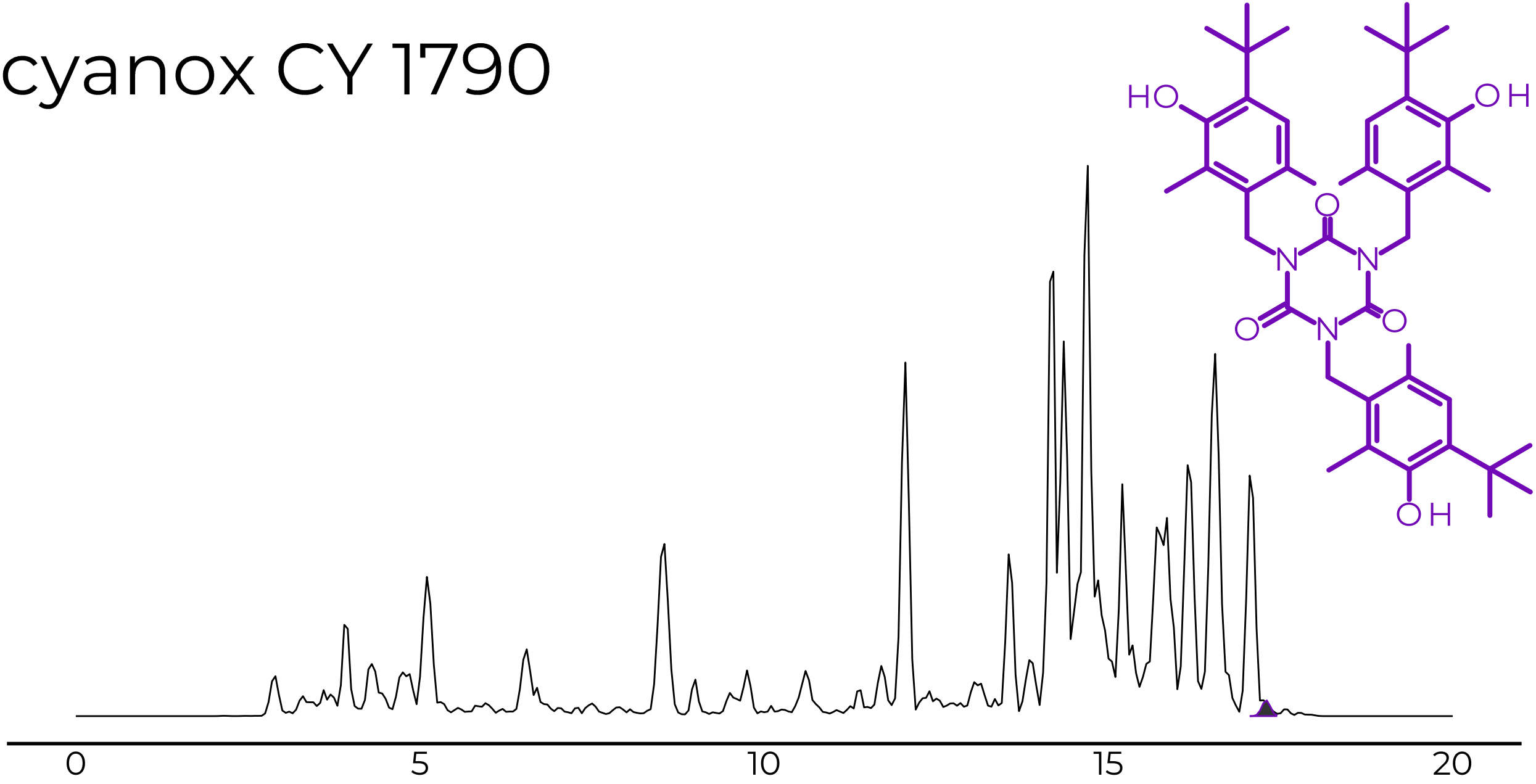
naproxen





anneli kruve
anneli.kruve@su.se

cyanox CY 1790

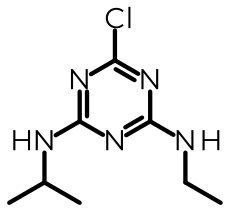


anneli kruve
anneli.kruve@su.se

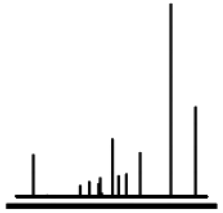
summary

prioritization in NTS

toxicity



structure



MS^2 spectrum

anneli kruve

anneli.kruve@su.se

kruvelab.com

anneli.kruve@su.se