

Y. Kreutzer, H. Sepman, A. Kruve  
yvonne.kreutzer@mmk.su.se

# Autoencoders for extracting structure and toxicity- relevant information from MS<sup>2</sup> spectra



Stockholm  
University

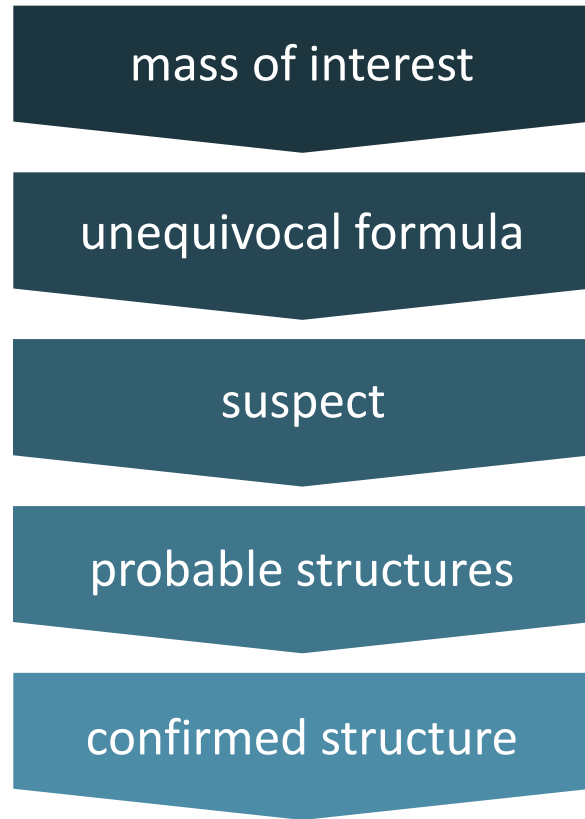


VR 3R  
2022-01353

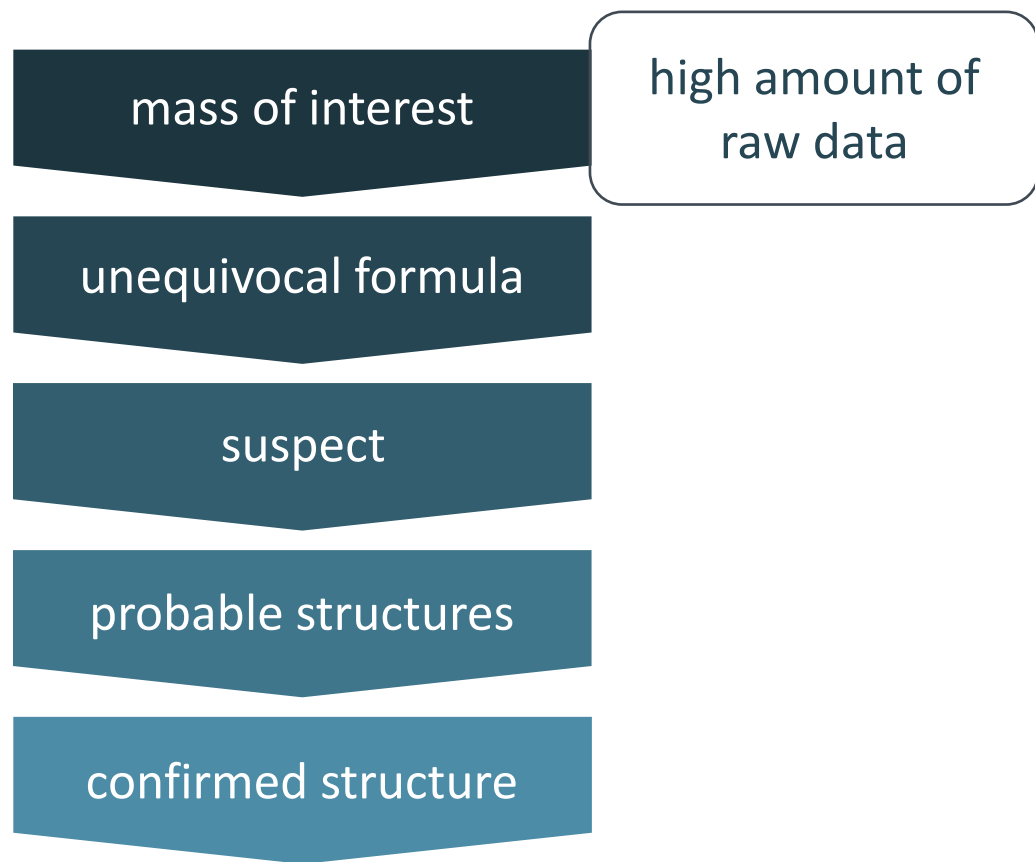
*Kruve lab*



# Introduction - risk prioritization



# Introduction - risk prioritization

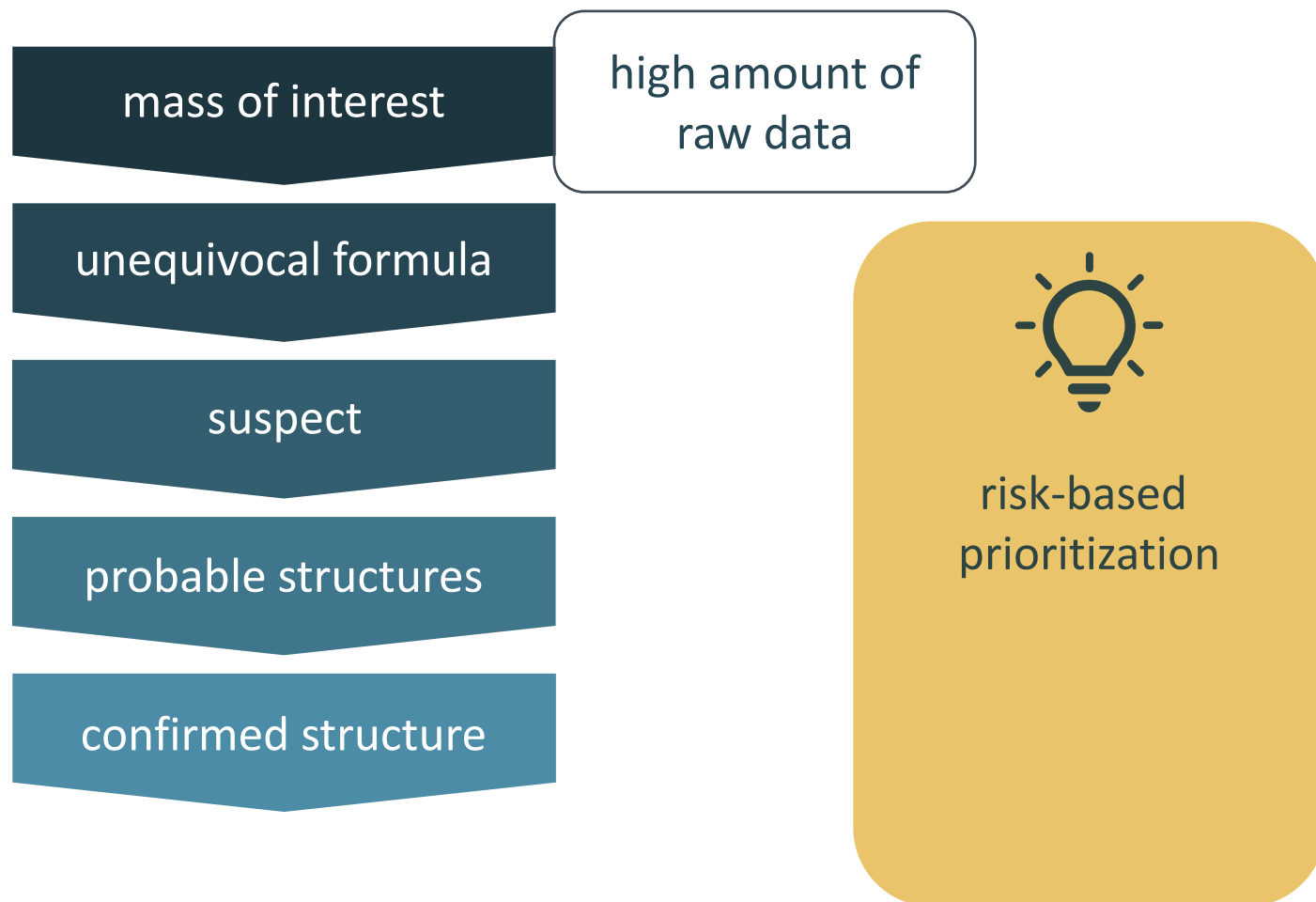


Alves VM, Muratov EN, Capuzzi SJ, et al. *Green Chem.* 2016

Schymanski EL, Singer HP, Slobodnik J, et al, *Anal Bioanal Chem.* 2015

Papazian S, D'Agostino LA, Sadiktsis I, et al. *Commun Earth Environ.* 2022

# Introduction - risk prioritization

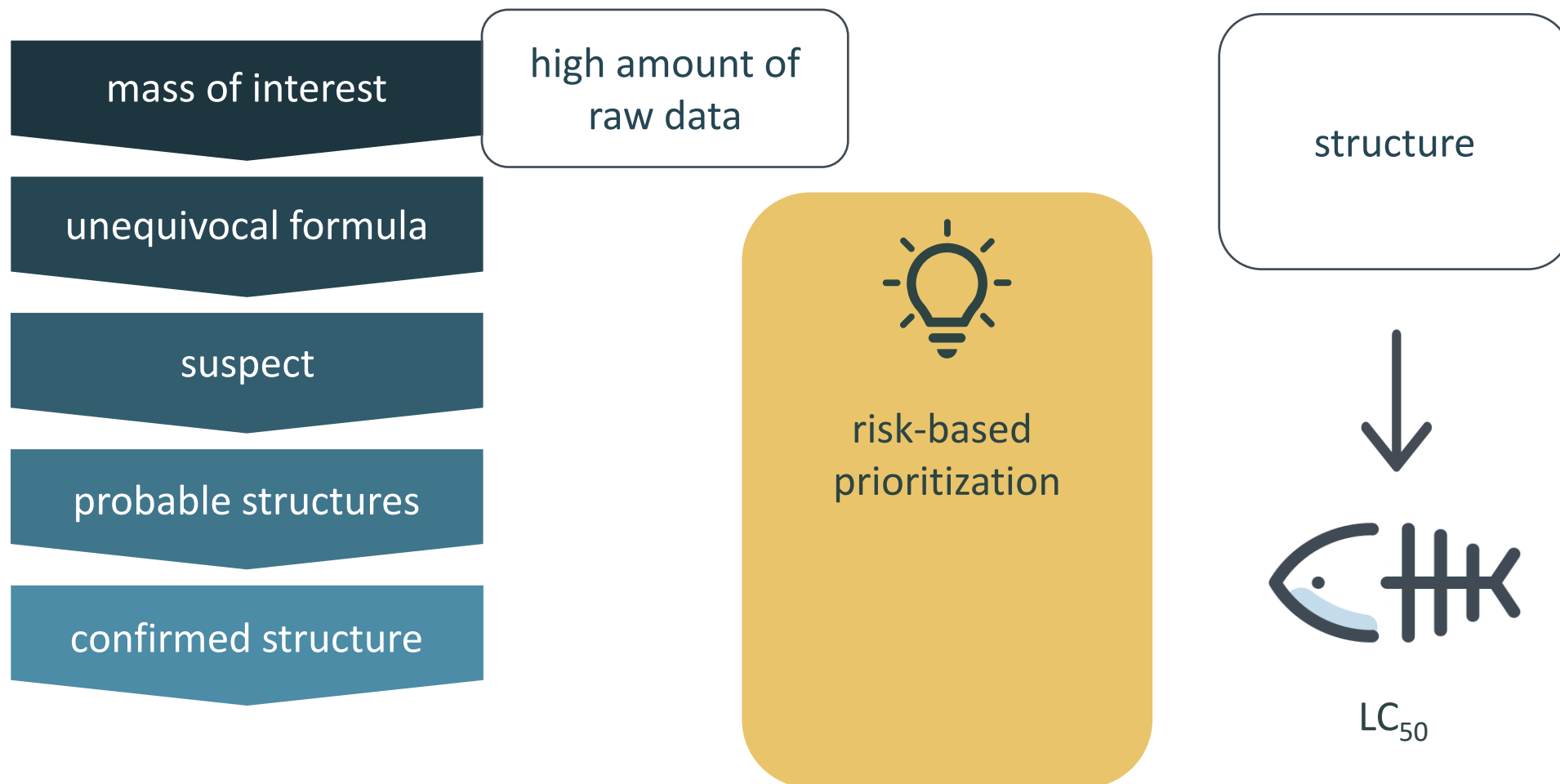


Alves VM, Muratov EN, Capuzzi SJ, et al. *Green Chem.* 2016

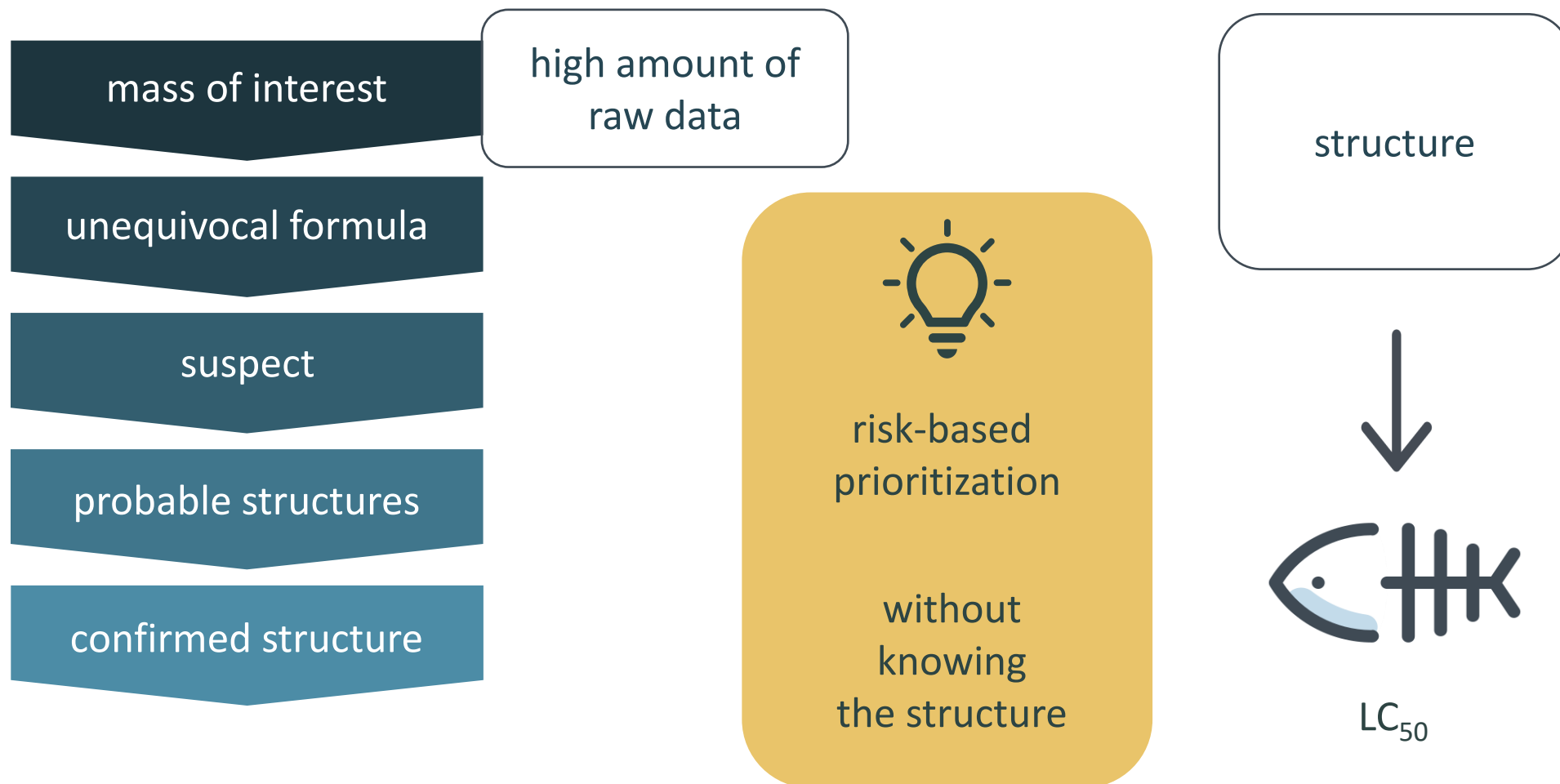
Schymanski EL, Singer HP, Slobodnik J, et al. *Anal Bioanal Chem.* 2015

Papazian S, D'Agostino LA, Sadiktsis I, et al. *Commun Earth Environ.* 2022

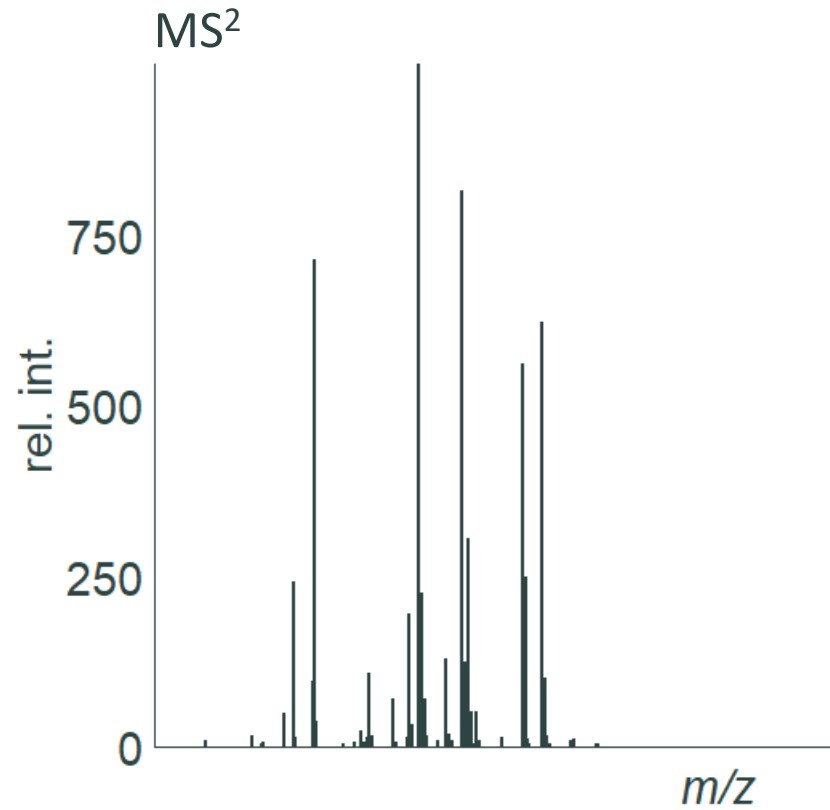
# Introduction - risk prioritization



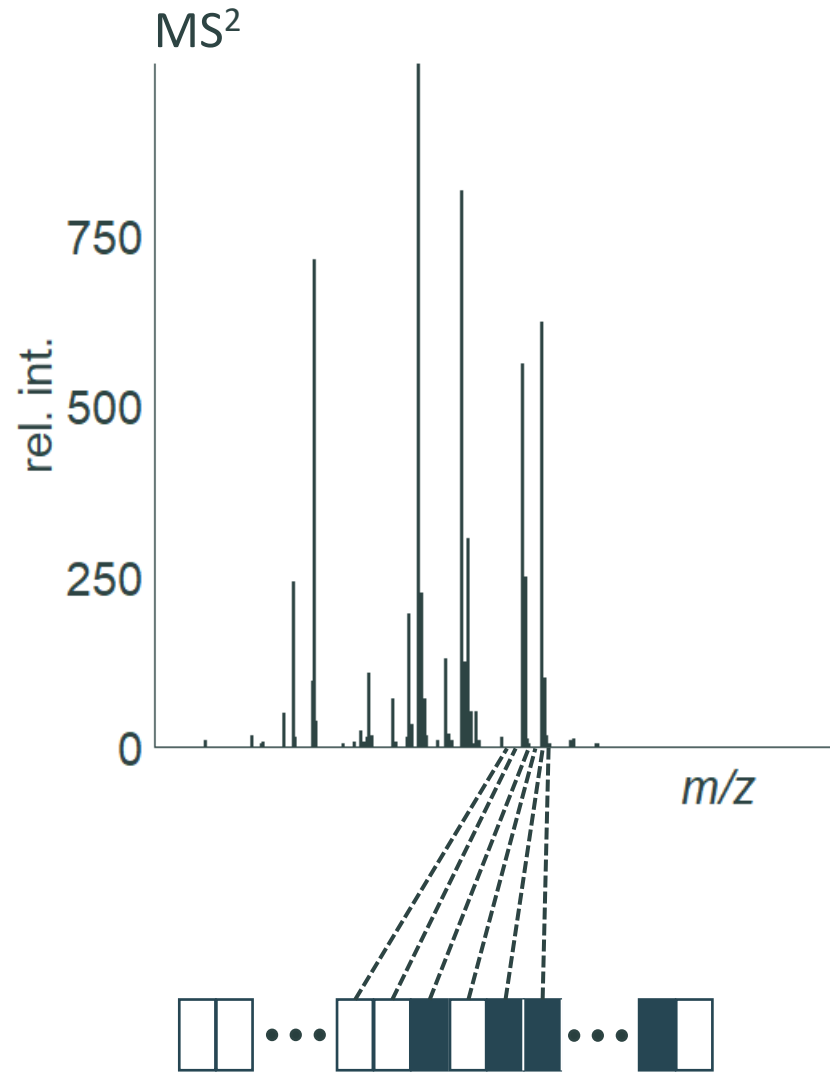
# Introduction - risk prioritization



# Introduction – dimension reduction

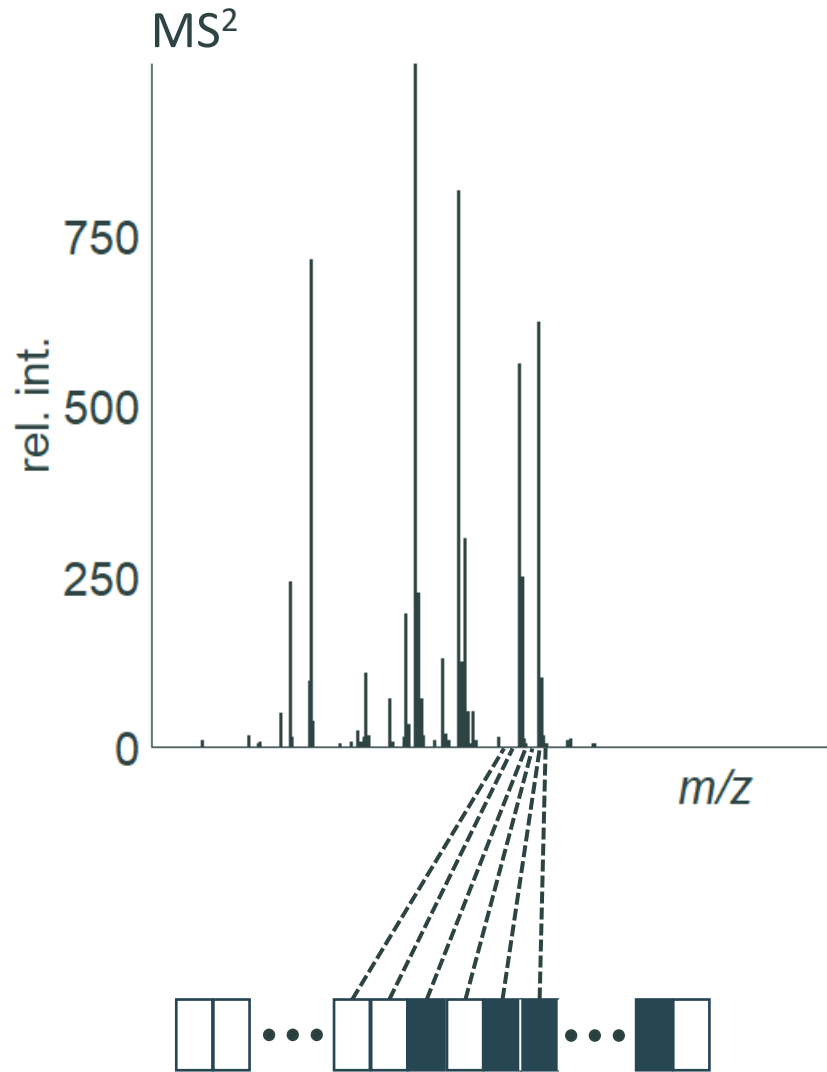


# Introduction – dimension reduction





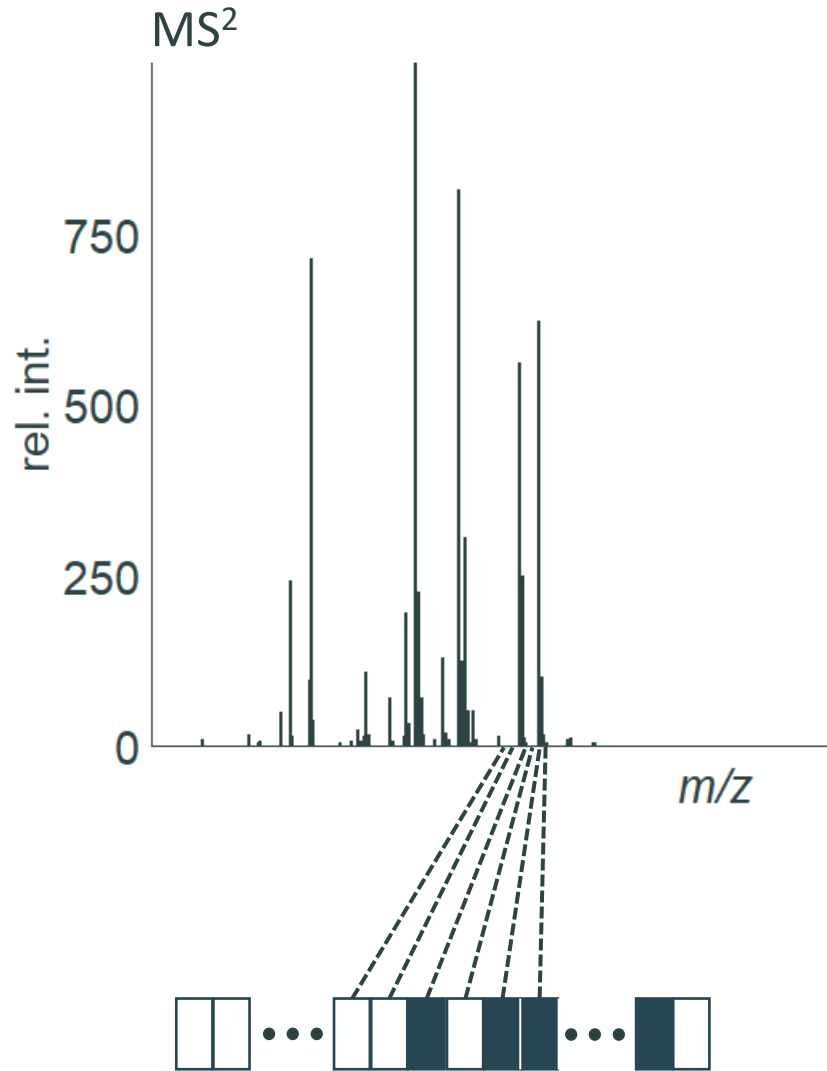
# Introduction – dimension reduction



## principal component analysis

- ✓ easy
- ✓ fast
- ✓ less computational power
- × performance on heterogenous data sets

# Introduction – dimension reduction



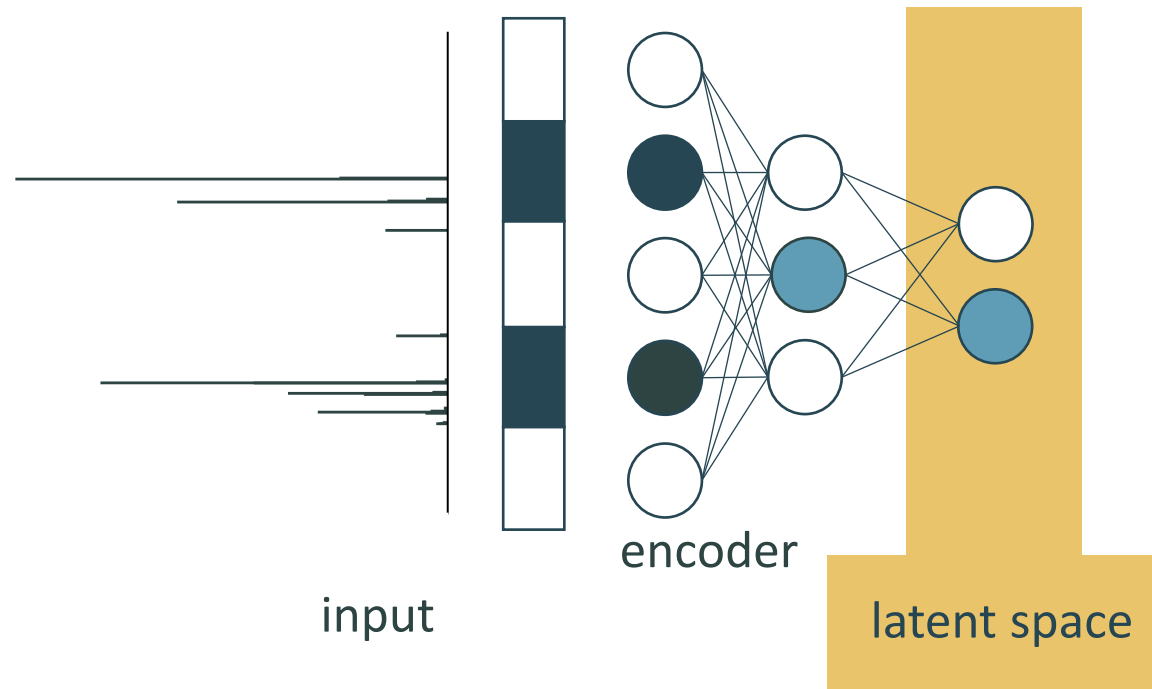
## principal component analysis

- ✓ easy
- ✓ fast
- ✓ less computational power
- × performance on heterogenous data sets

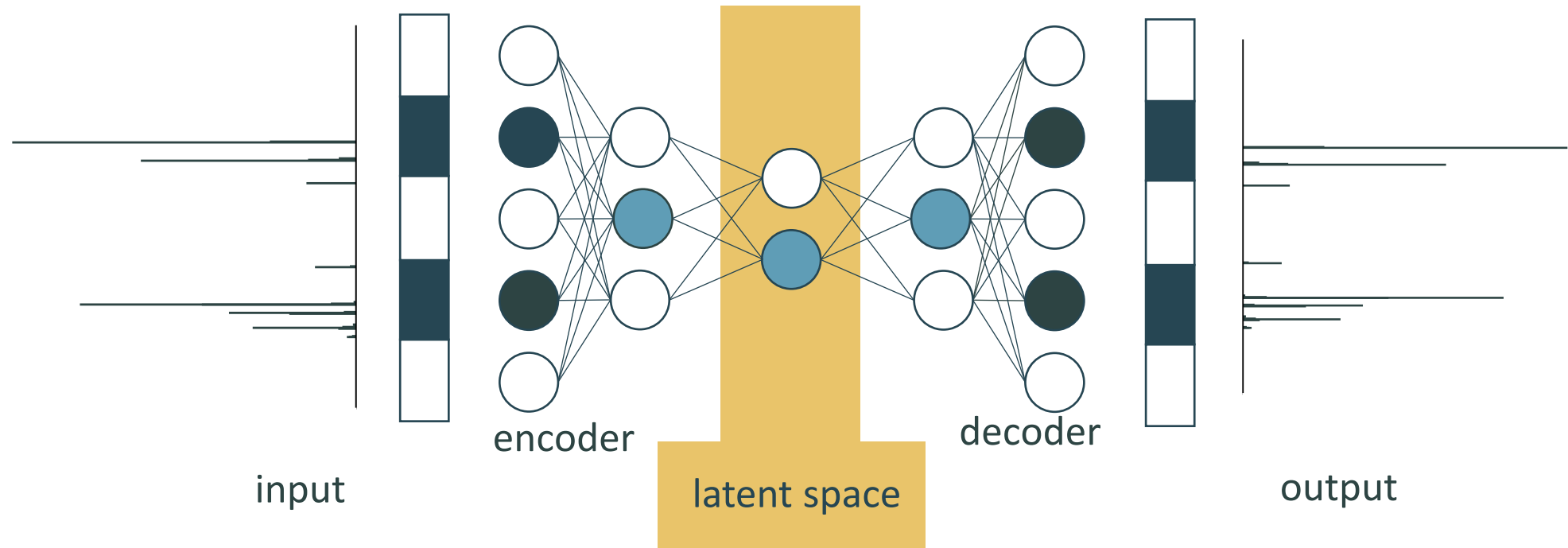
## autoencoder

- ✓ effective
- ✓ non-linear dimension reduction

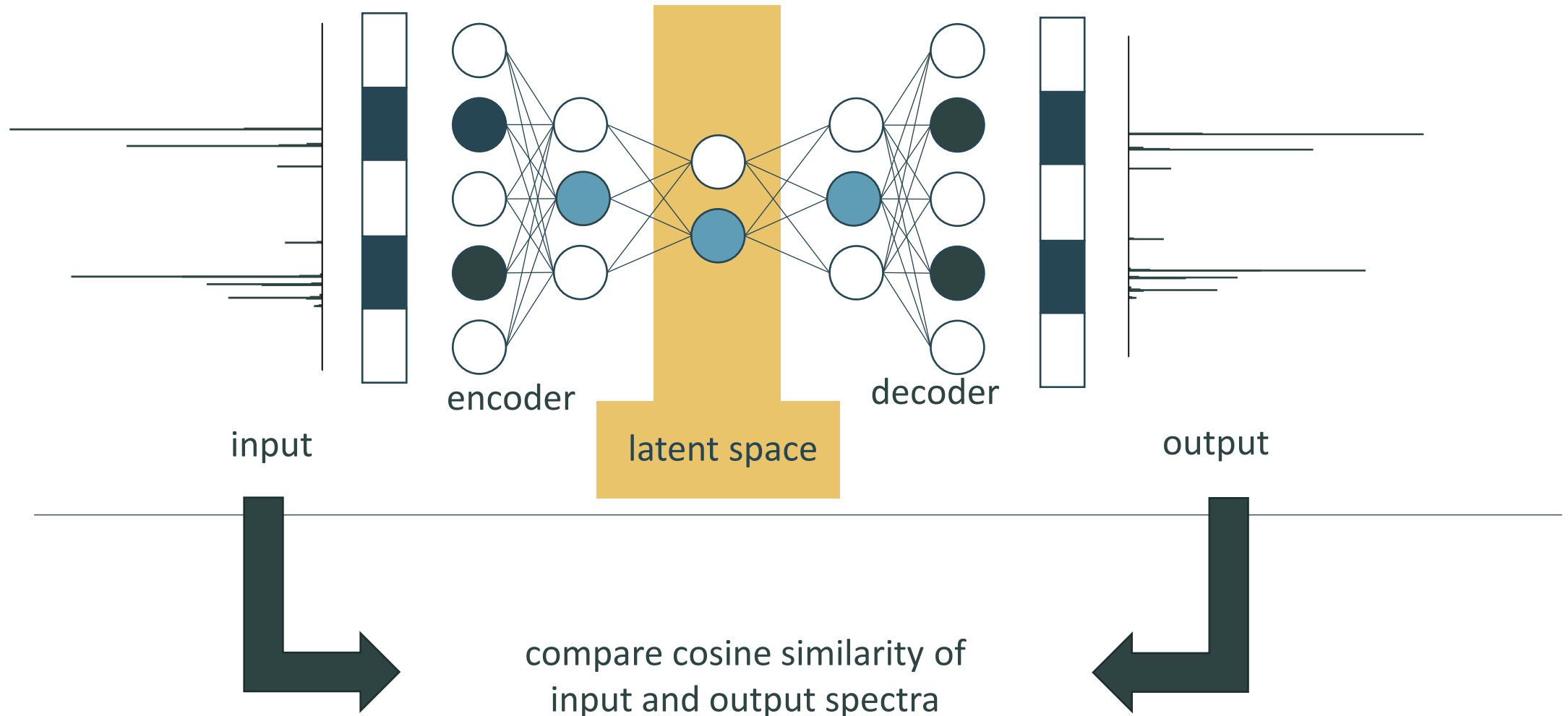
# Introduction - Autoencoders



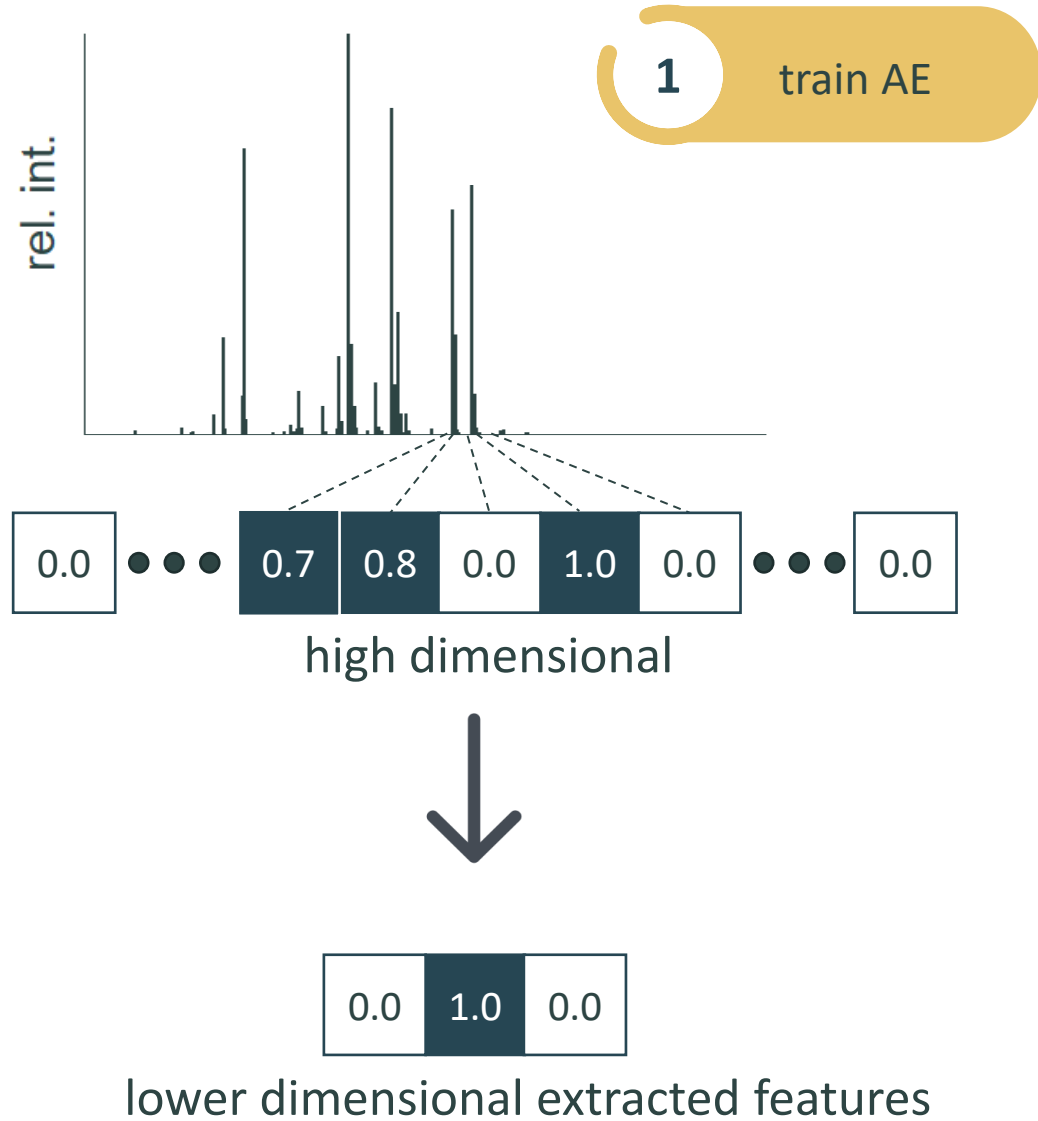
# Introduction - Autoencoders



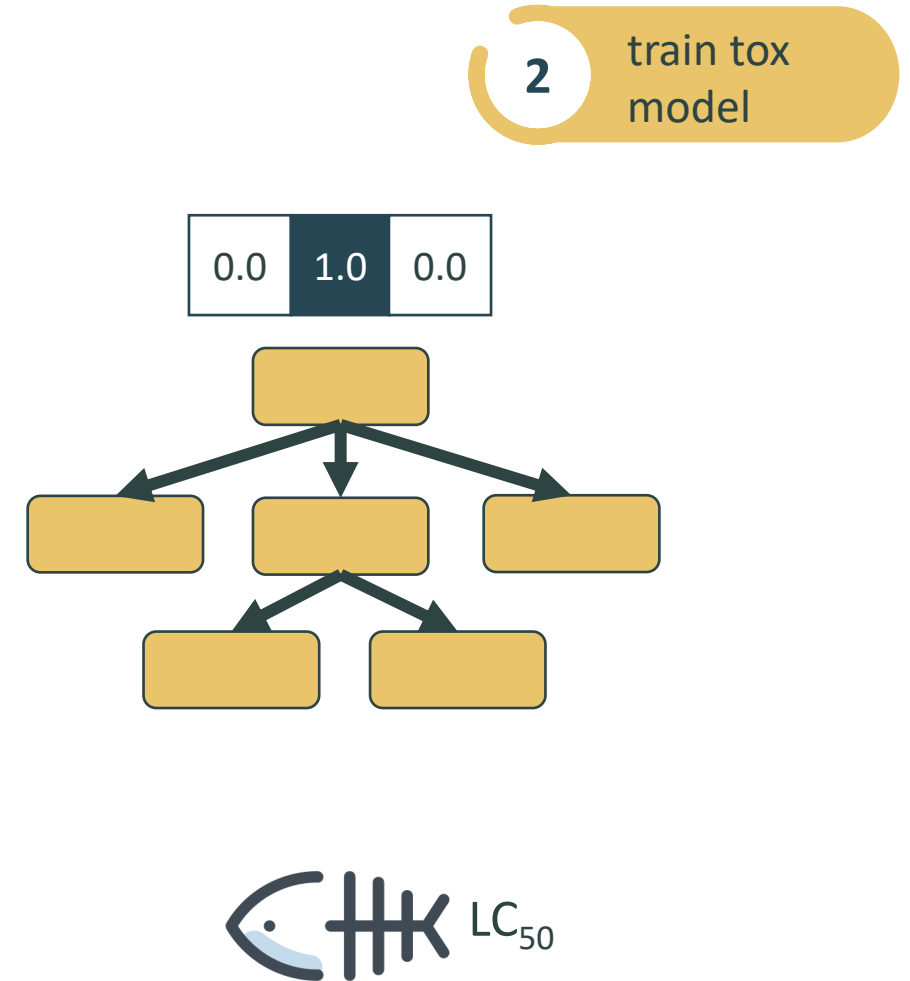
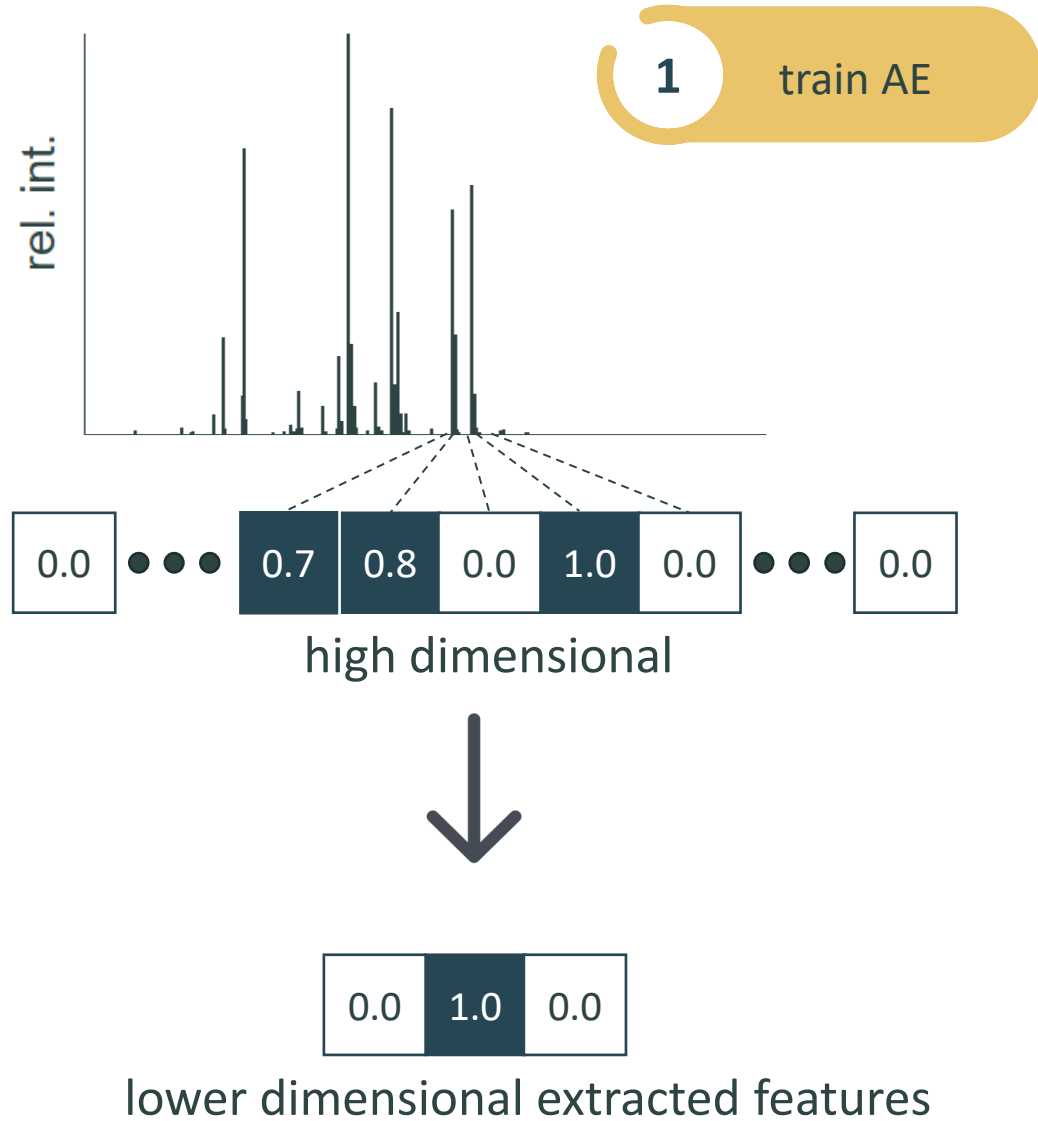
# Introduction - Autoencoders



# Aim

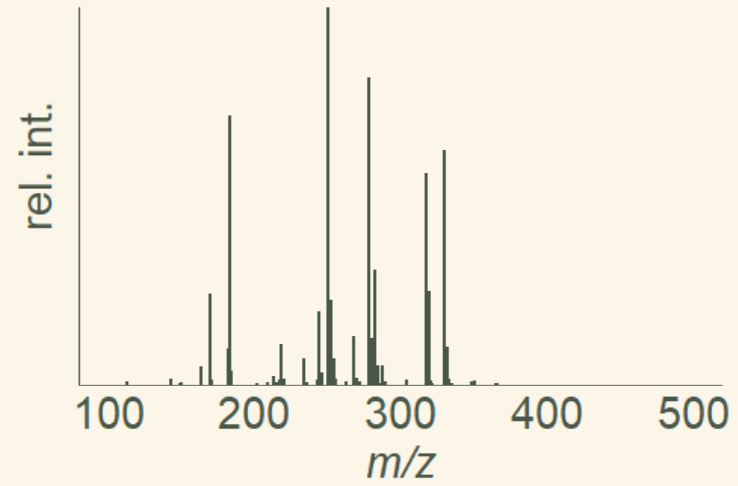


# Aim



# Data

## MassBank

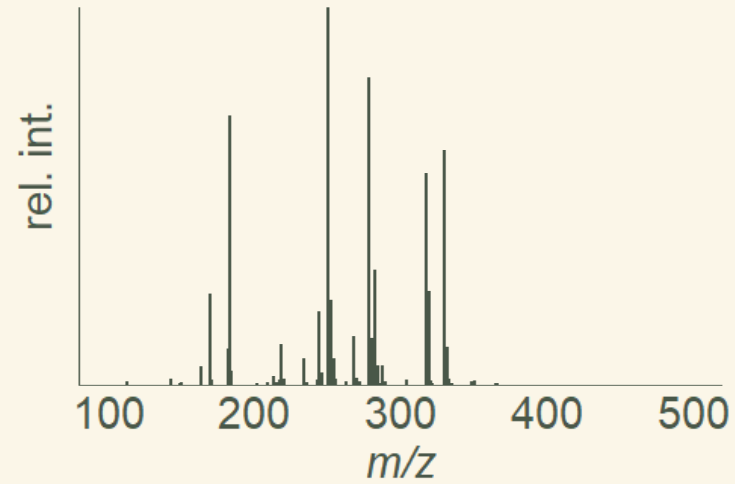


MassBank  
ESI +  
1689  
compounds



# Data

## MassBank



## CompTox



MassBank  
ESI +  
1689  
compounds

187  
compounds

CompTox  
LC<sub>50</sub>

# Results – information extraction

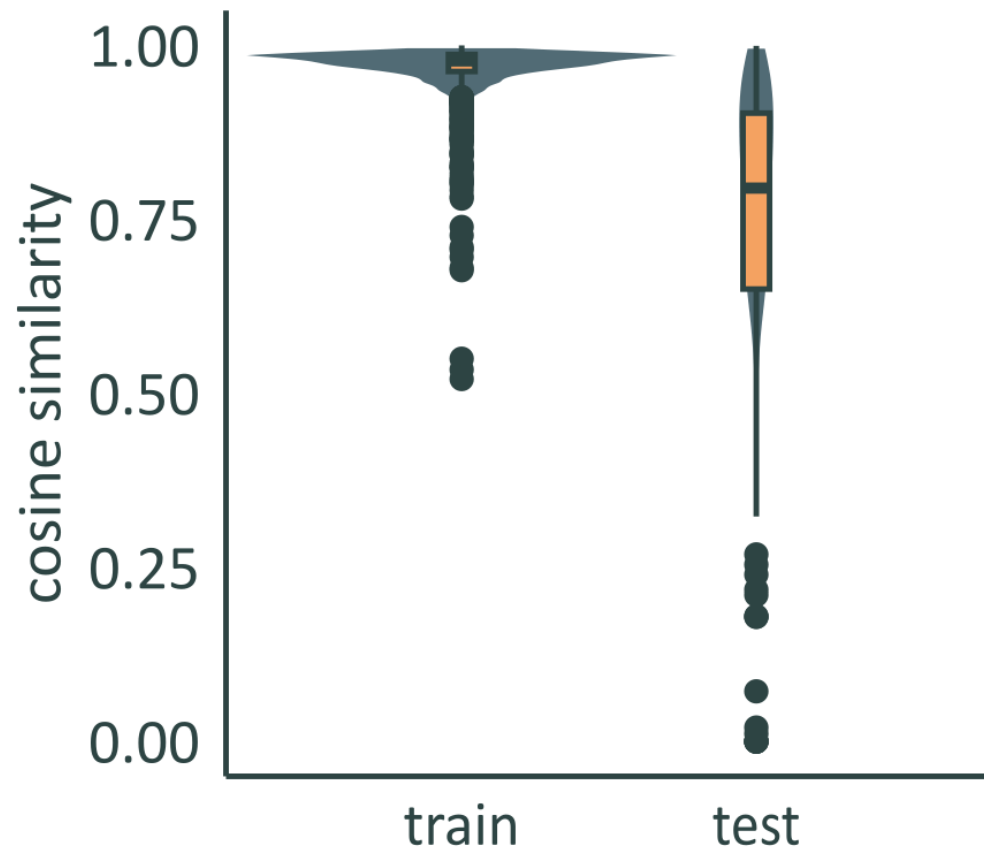
train = 1080 compounds

test = 270 compounds

# Results – information extraction

train = 1080 compounds

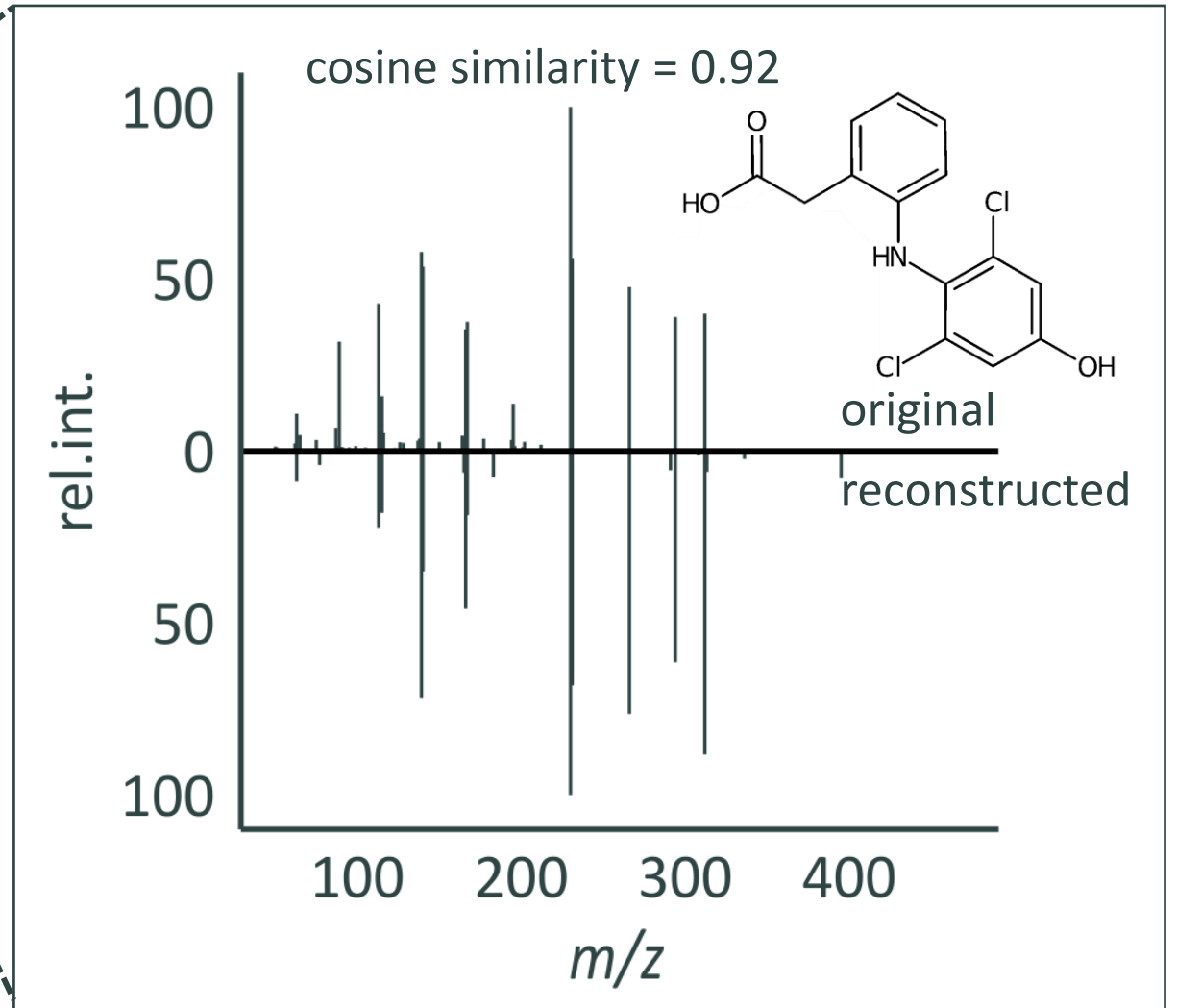
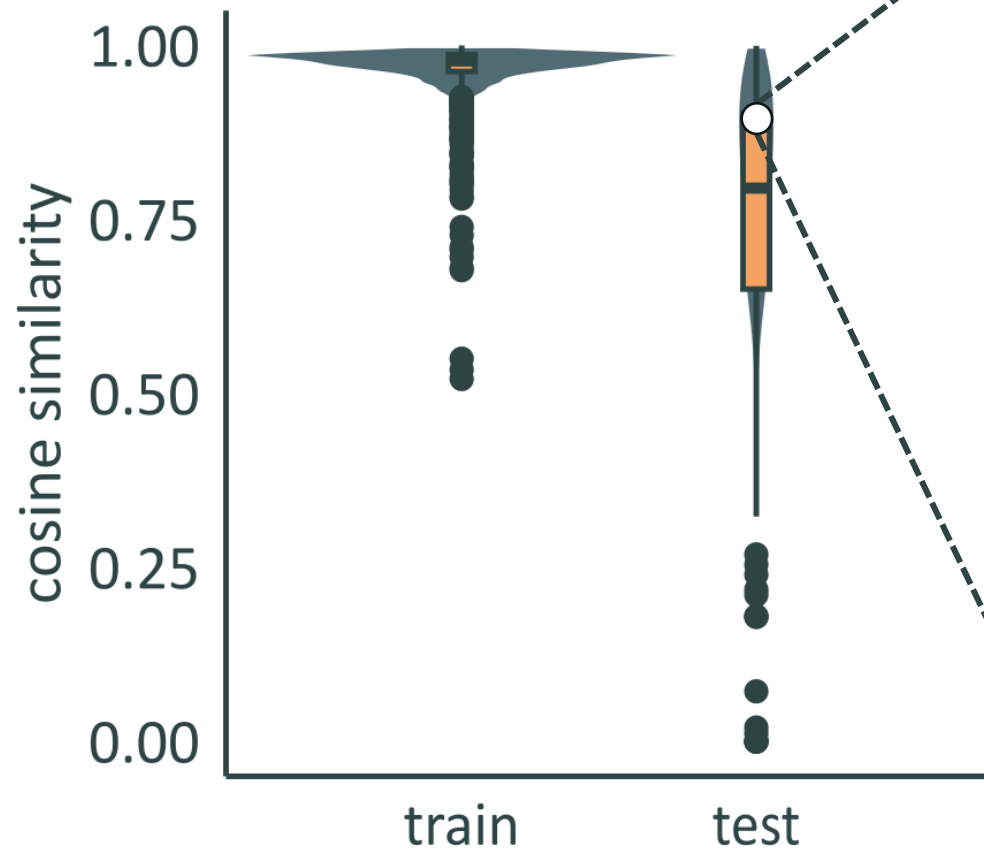
test = 270 compounds



# Results – information extraction

train = 1080 compounds

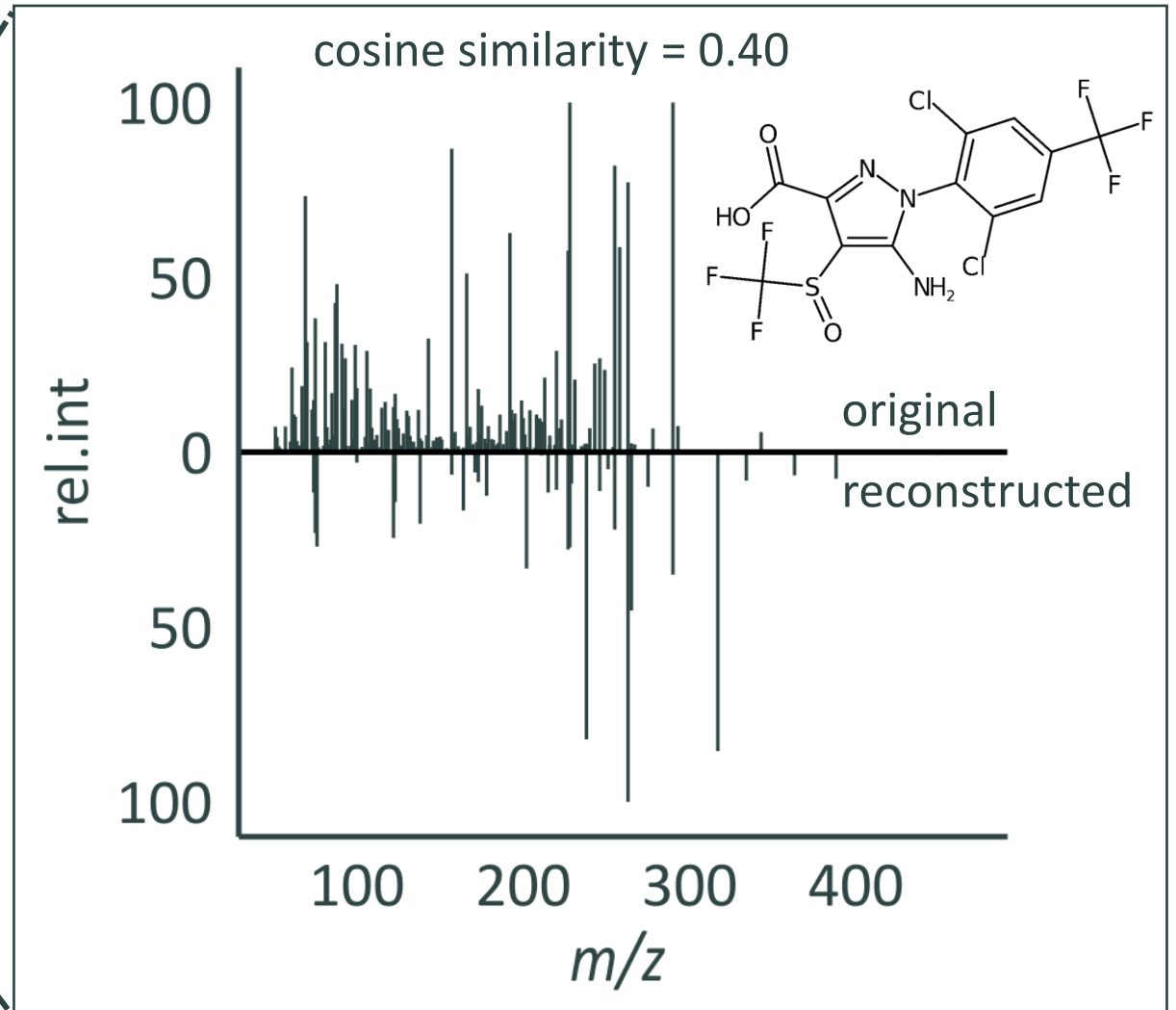
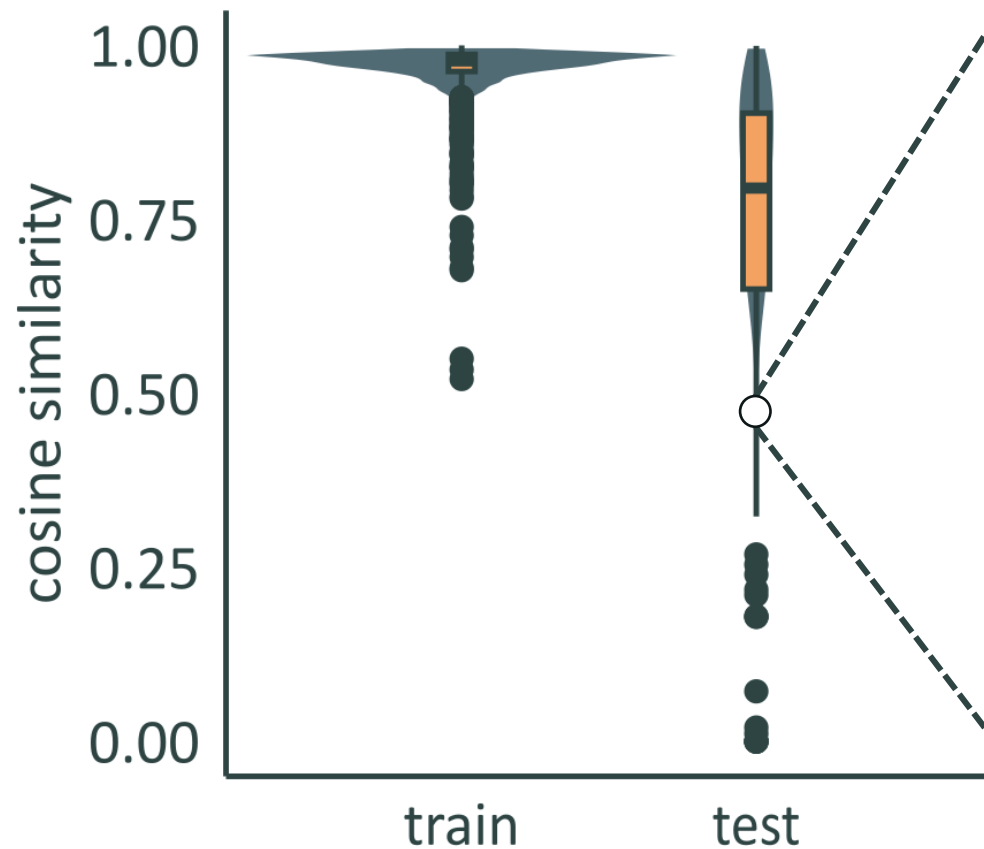
test = 270 compounds



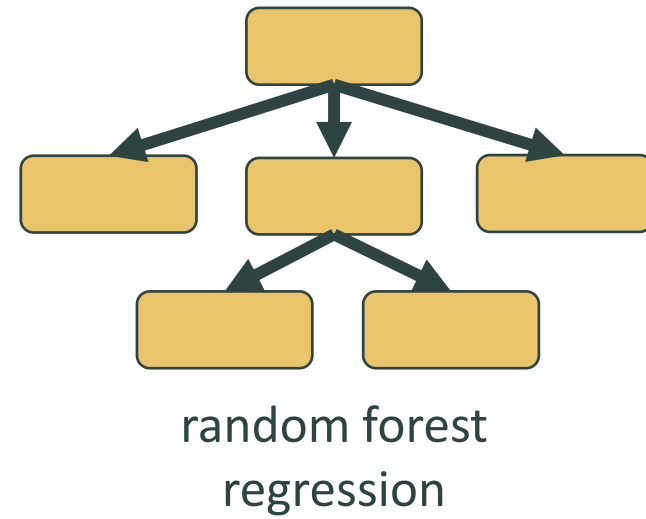
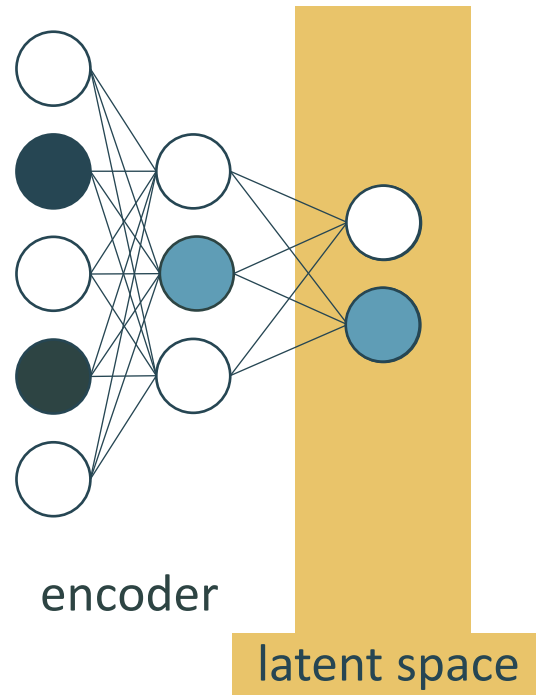
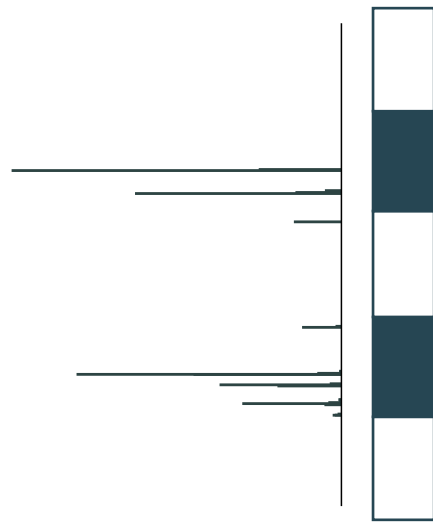
# Results – information extraction

train = 1080 compounds

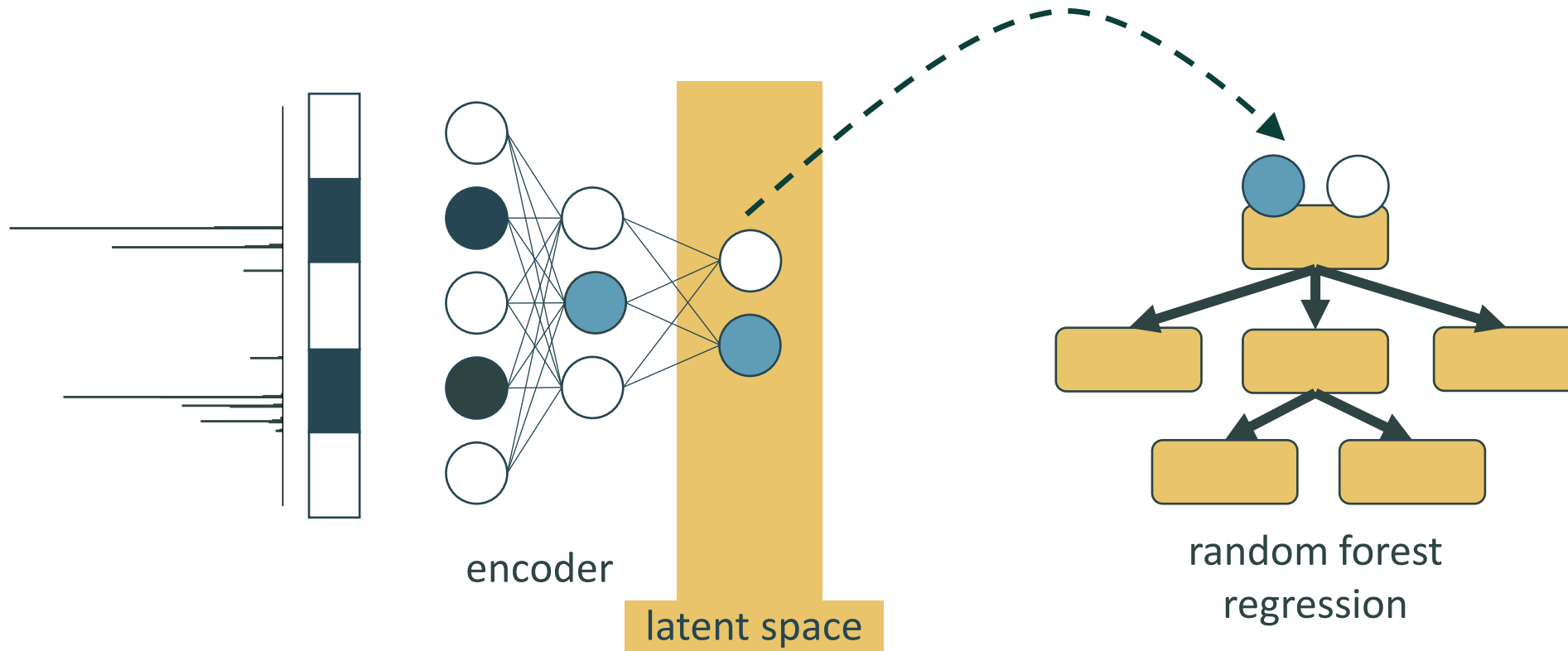
test = 270 compounds



# Results – information extraction

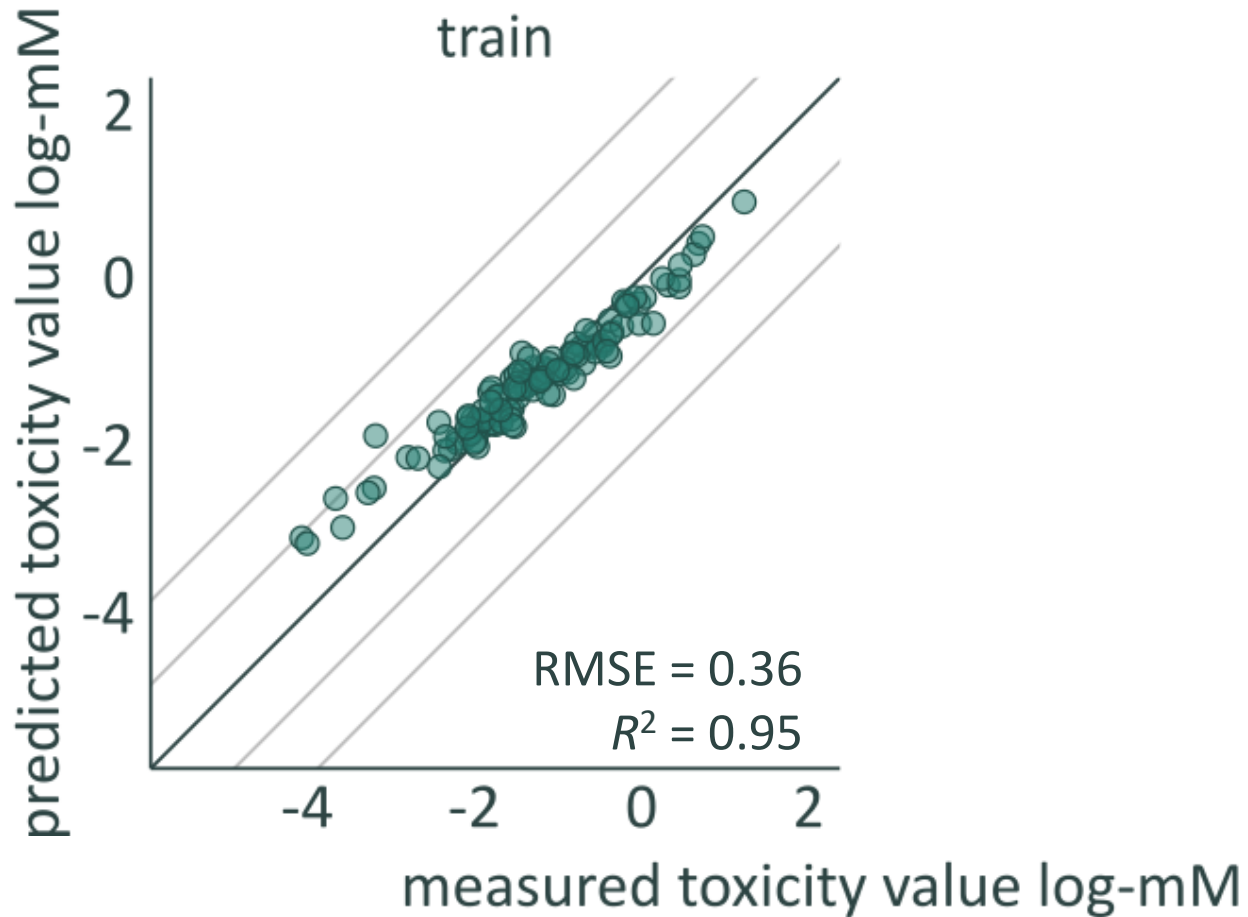


# Results – information extraction



# Results – LC<sub>50</sub> modelling

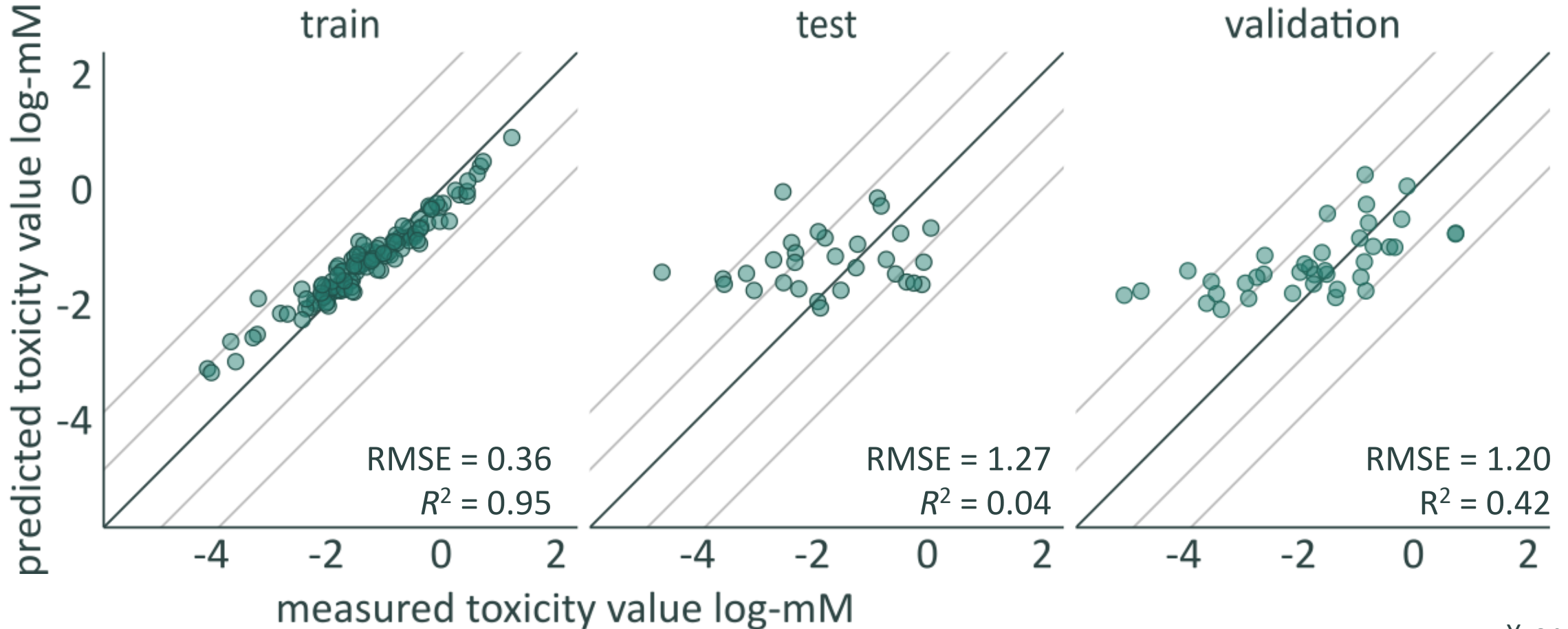
train = 119 compounds  
test = 30 compounds  
validation = 38 compounds





# Results – LC<sub>50</sub> modelling

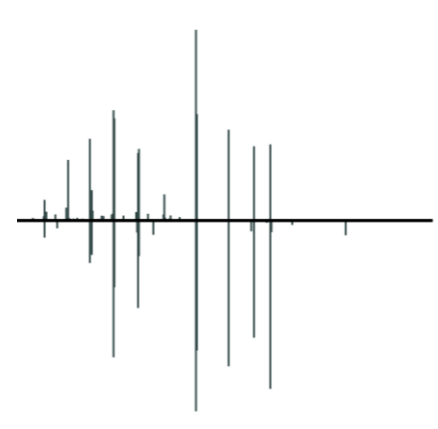
train = 119 compounds  
test = 30 compounds  
validation = 38 compounds



# Conclusions and future perspective

## Autoencoder

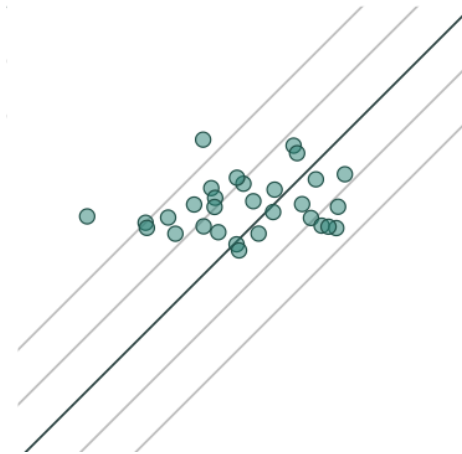
able to extract spectra-specific information



## LC<sub>50</sub> modelling

high prediction error

→ latent space do not yet hold toxicity relevant information





Stockholm  
University



VR 3R

2022-01353

*Kruve lab*



# Autoencoders for extracting structure and toxicity-relevant information from MS<sup>2</sup> spectra

Y. Kreutzer, H. Sepman, A. Kruve  
yvonne.kreutzer@mmk.su.se

