

Data-driven selection of stationary phases for MS²-similar environmental isomers

Presenter: Ruchun Chen

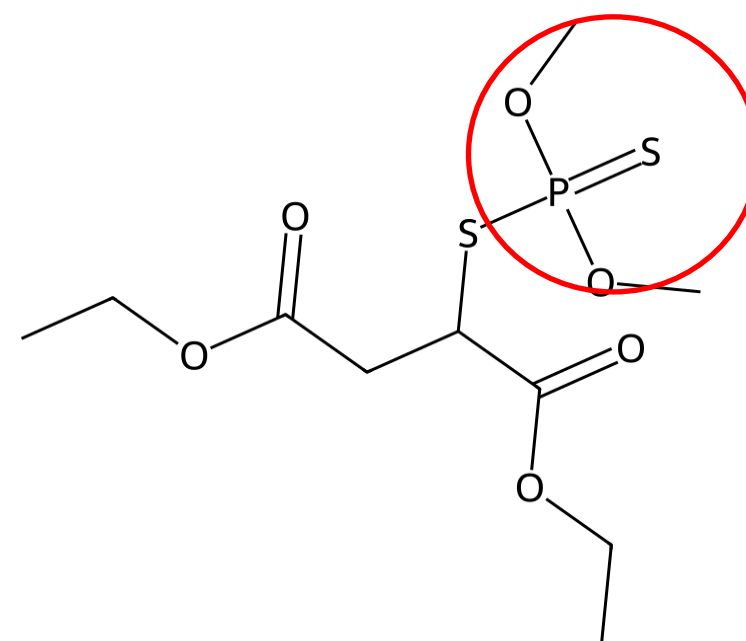
Supervisor: Prof. Anneli Kruve, Dr. Eliise Tammekivi

ruchun.chen@su.se

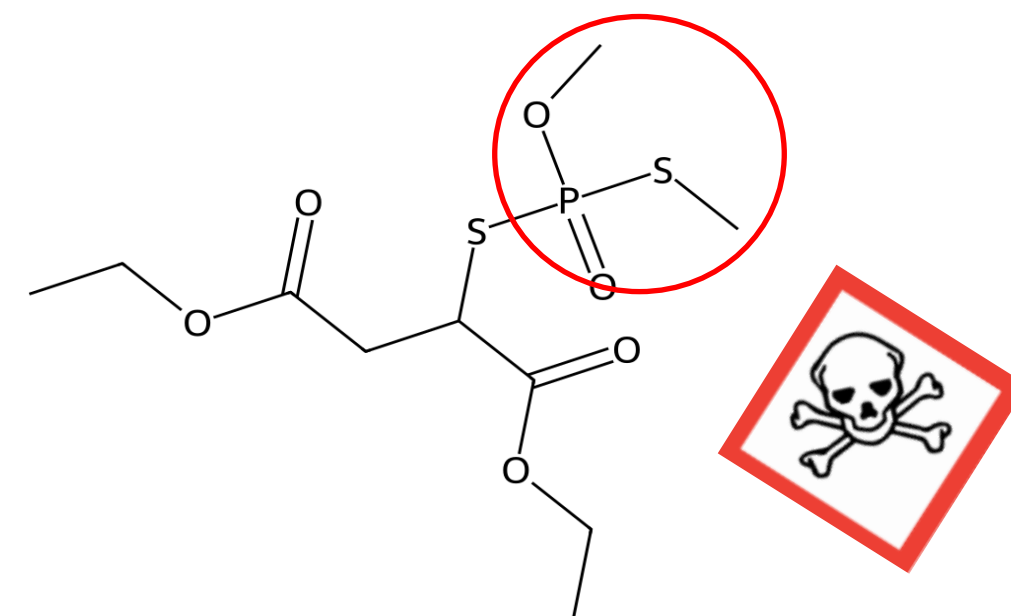
Introduction

Why isomer identification matters

Pakistan 1976 – Malaria Epidemic



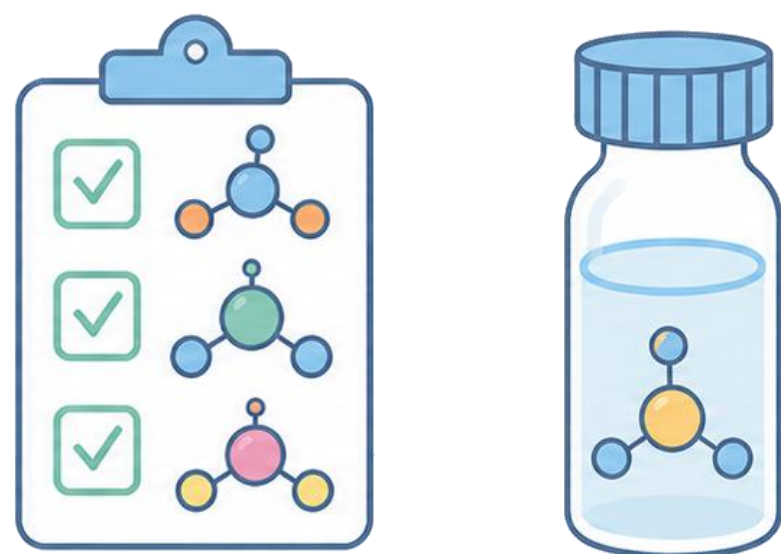
Malathion



Isomalation

**How do we know
what's there?**

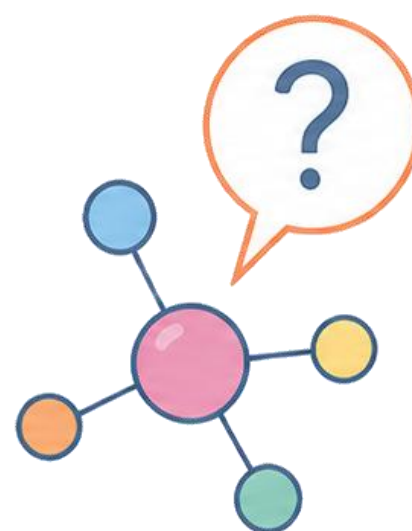
Non-Target Screening (NTS)



Targeted Analysis

Predifined chemical

Referenrece standards



What may remain unseen

Unexpected pollutants

Emerging contaminants

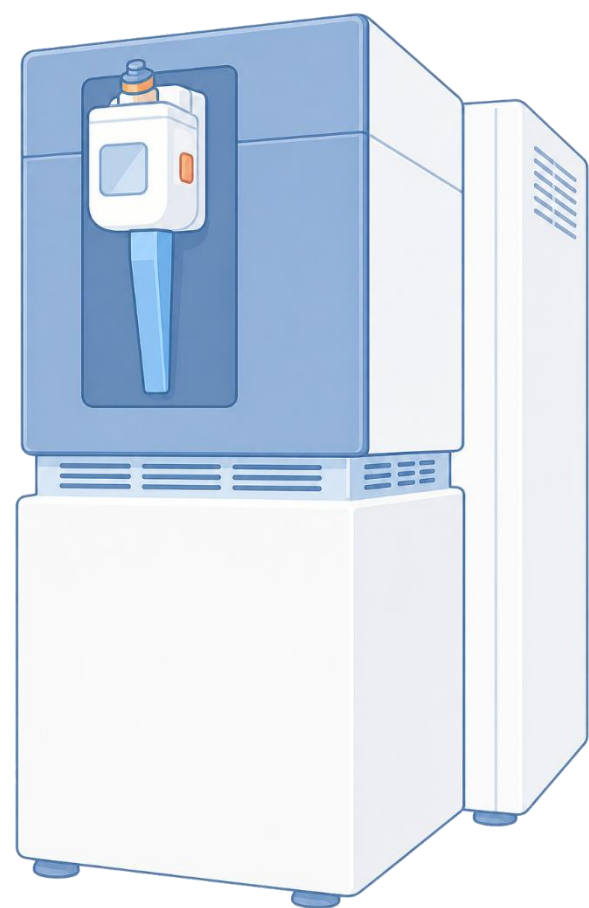


Non-target Screening

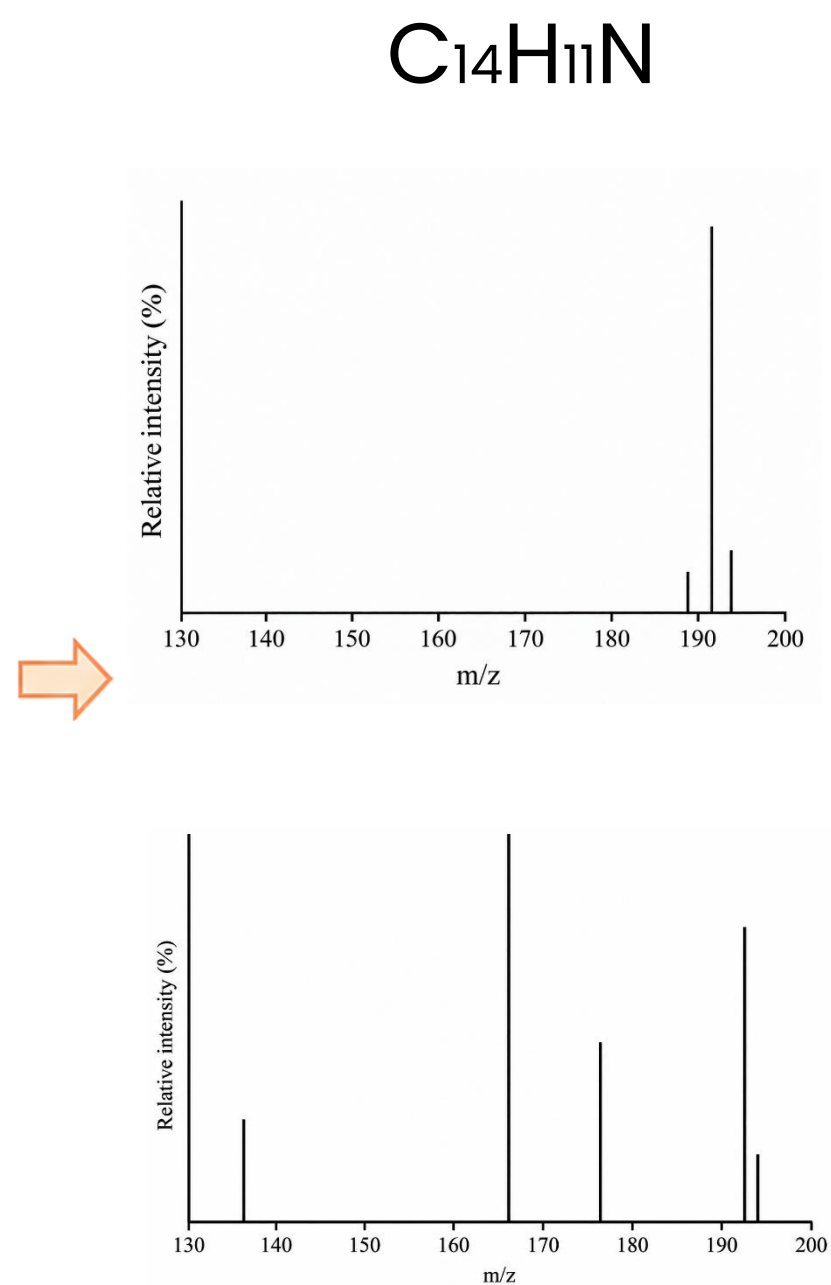
Collect broad range of signals

NTS with LC-HRMS

Liquid Chromatography-High Resolution Mass Spectrometry

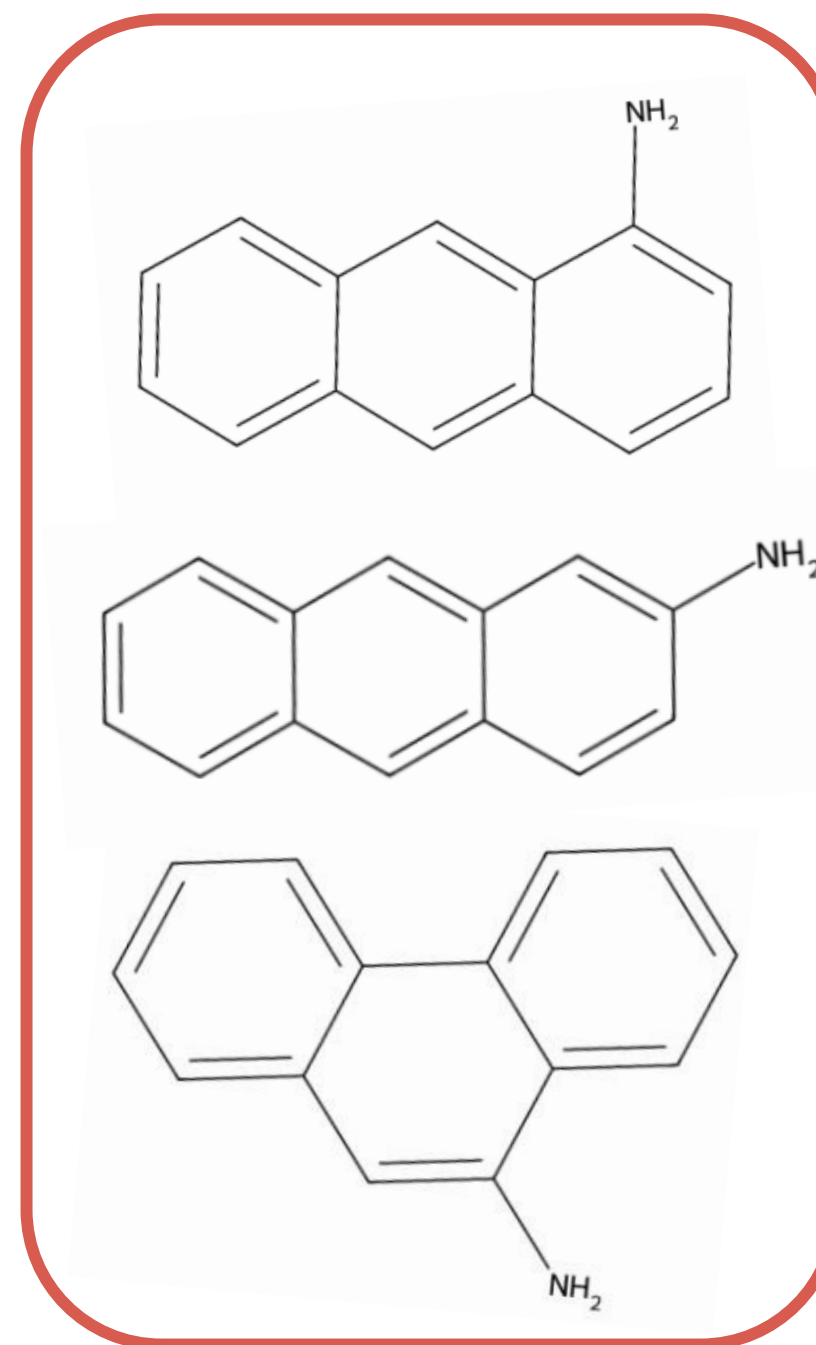


LC-HRMS

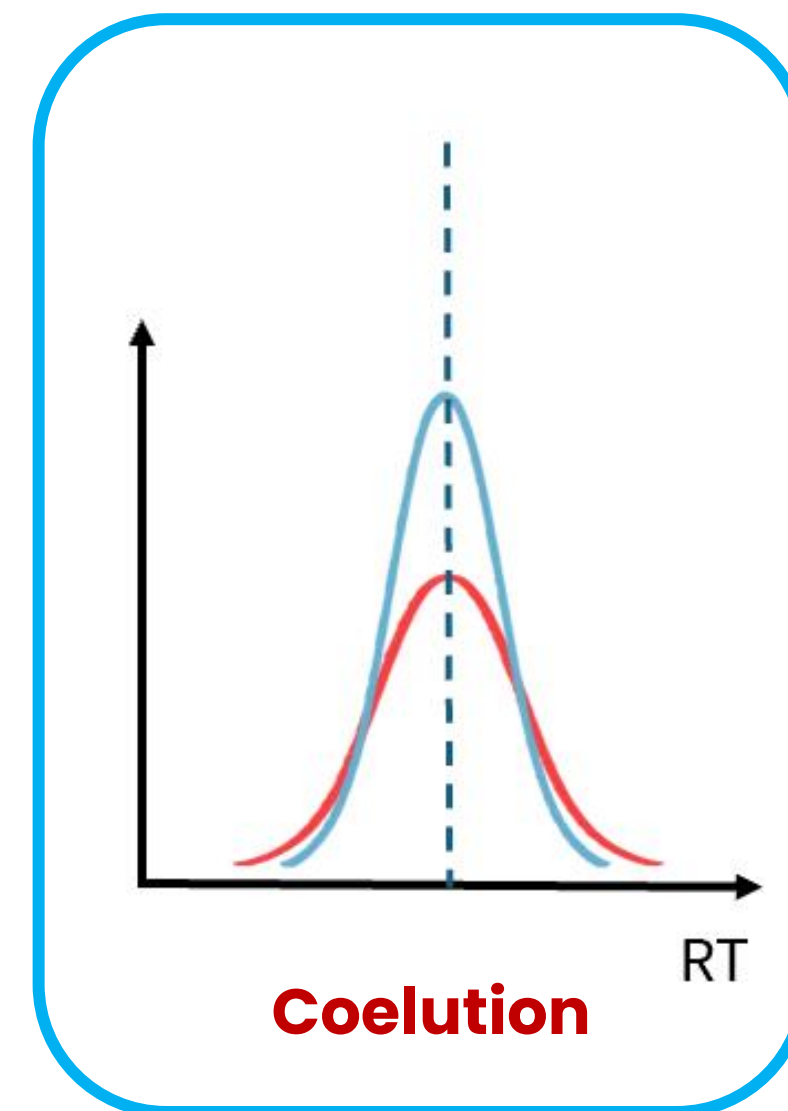


MS^1

MS^2

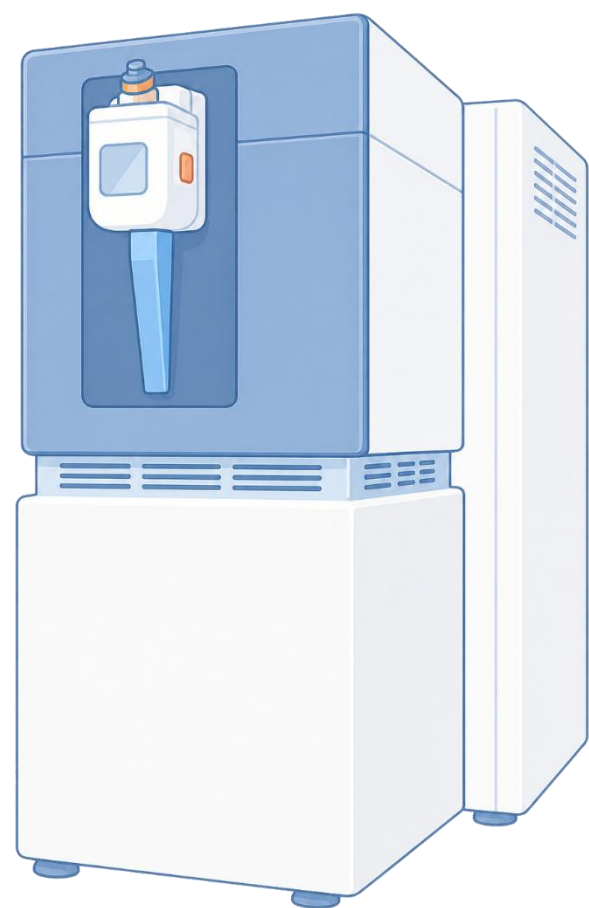


Candidate Structure

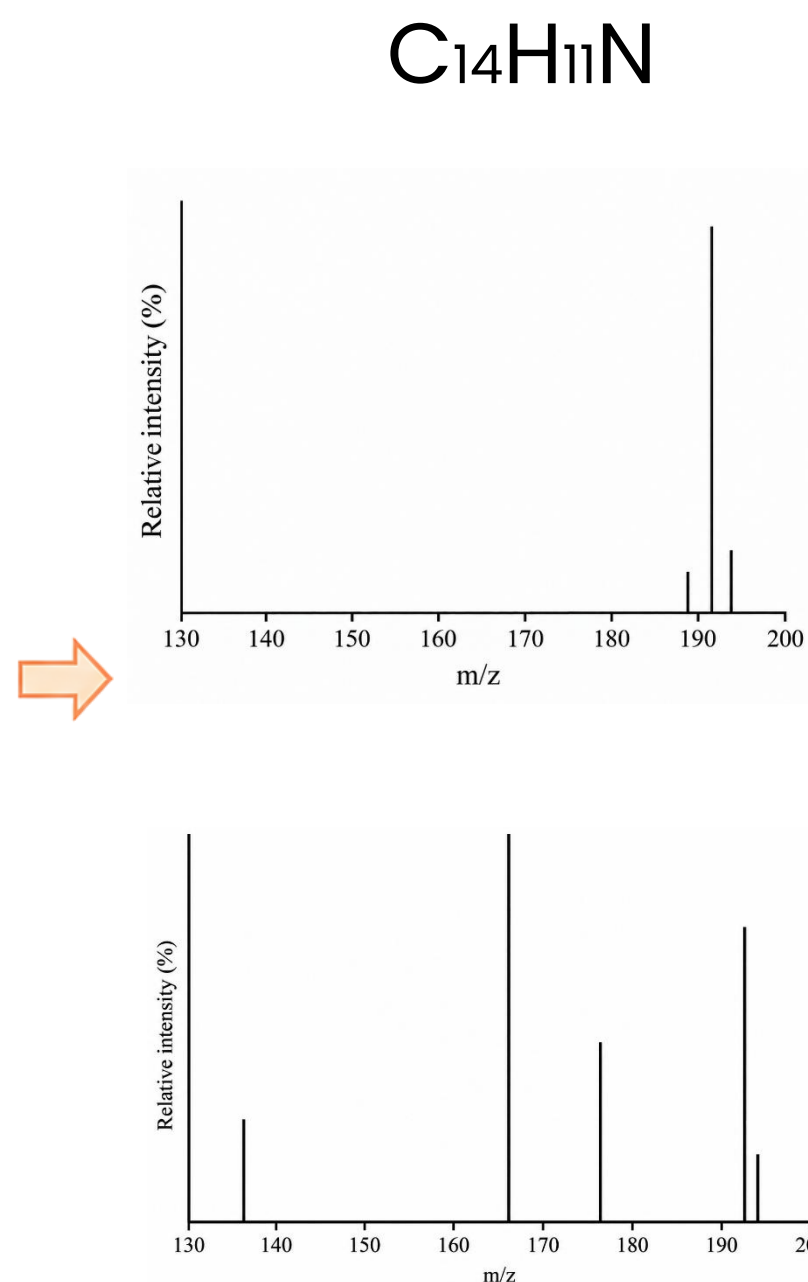


NTS with LC-HRMS

Liquid Chromatography-High Resolution Mass Spectrometry

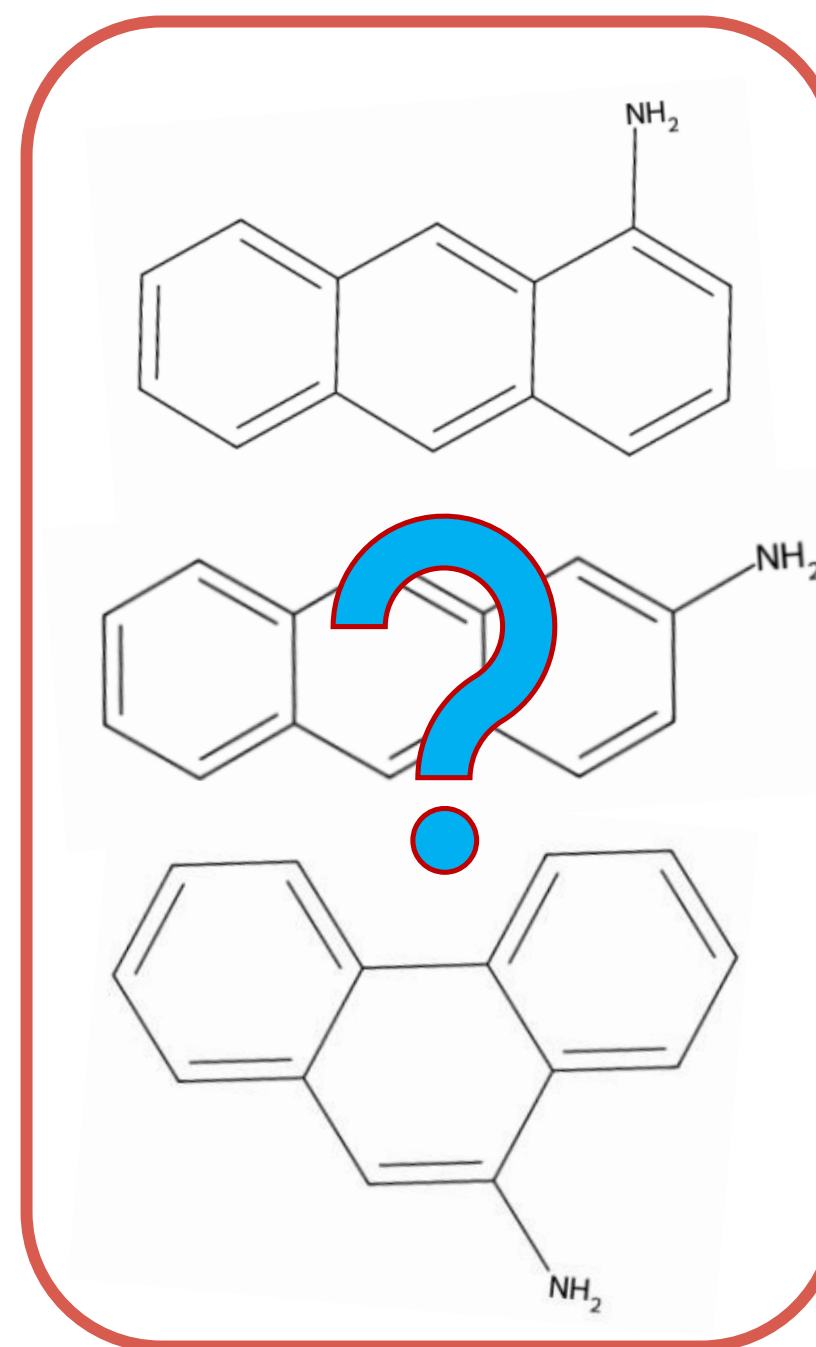


LC-HRMS

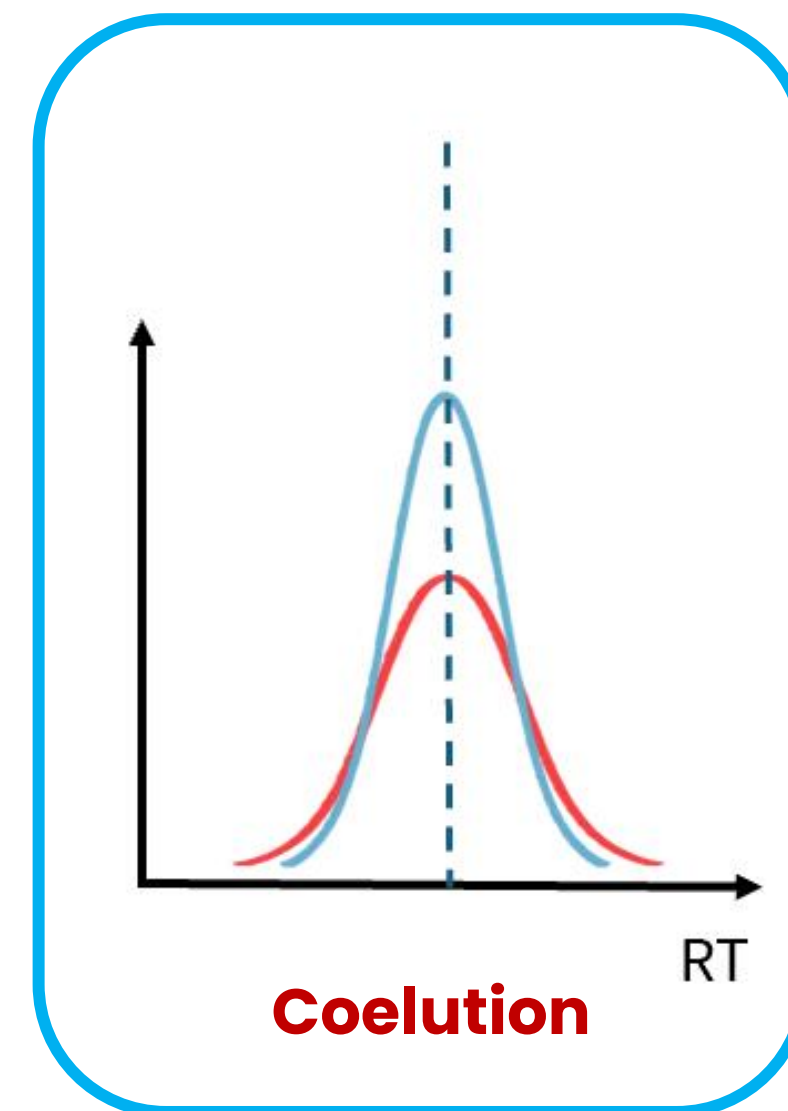


MS^1

MS^2



Candidate Structure

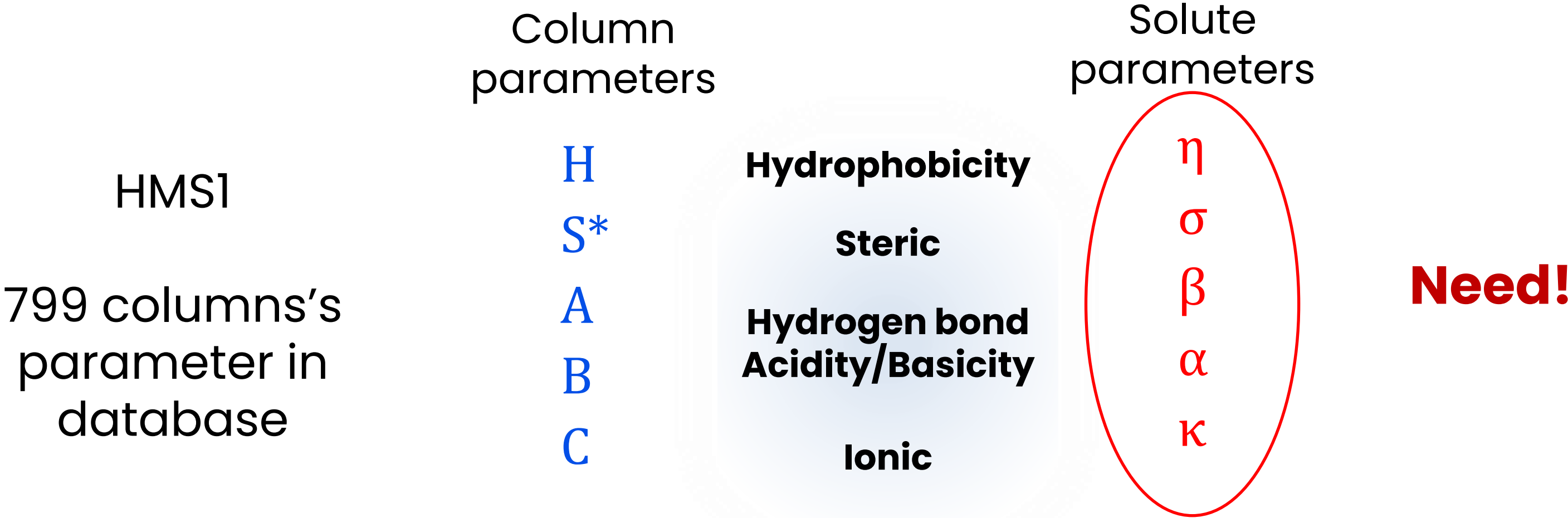


How do we select the stationary phase to help with separating isomers without extensive experimental screening?

HSM & Machine Learning

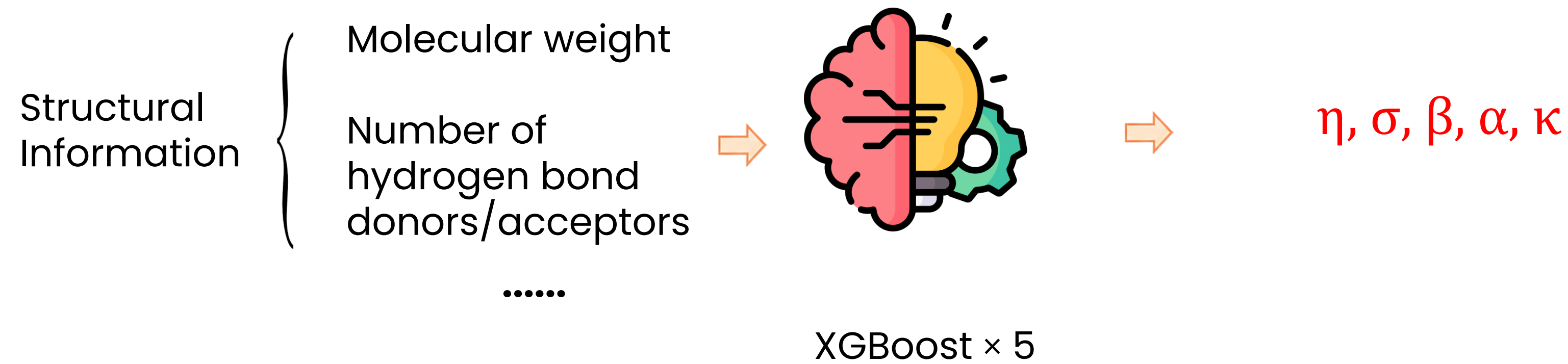
HSM- Hydrophobic Subtraction Model of Selectivity

$$\log\left(\frac{k_x}{k_{ref}}\right) = \log\alpha = \eta H - \sigma S^* + \beta A + \alpha B + \kappa C$$



HPLCColumns.org, Column Selectivity Database (accessed Sep 2026), Snyder et al. (2004).

Build Models for Solute Parameter Prediction



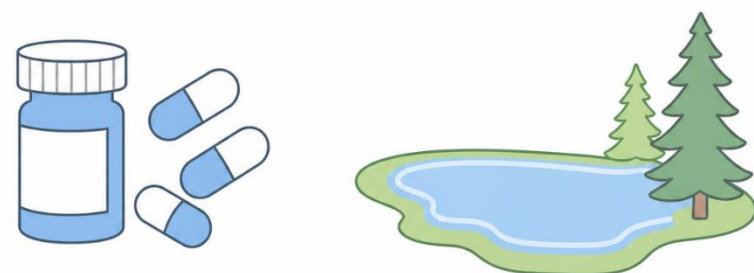
HSM1 solute-parameter data: Wen et al., *J. Chromatogr. A* 1541 (2018), Wen et al., *Electrophoresis* 40 (2019)

Selecting Representative Isomer Pairs



48 Pairs, 83 individual chemicals

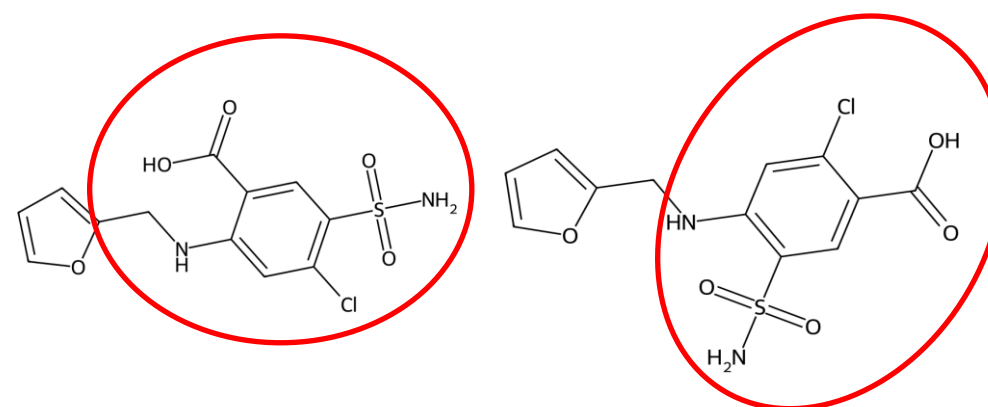
PubChemLite
EXPOSOMICS



Tanimoto
similarity >0.85

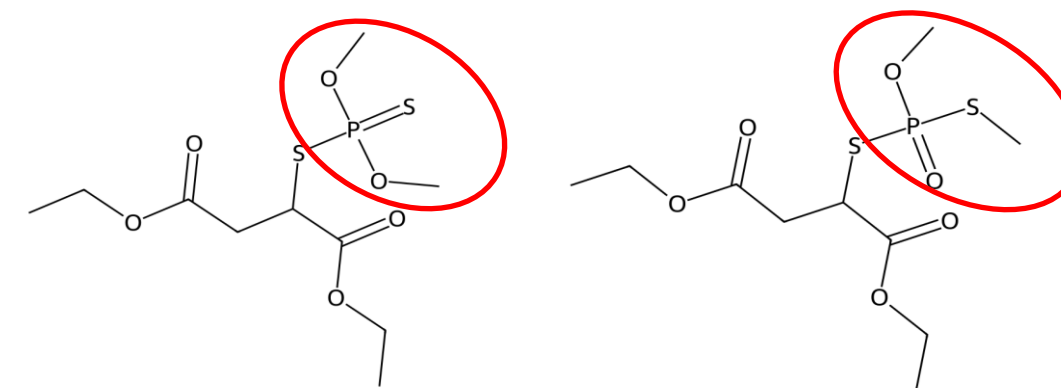
Pharmaceuticals

Furosemide & Isofurosemide



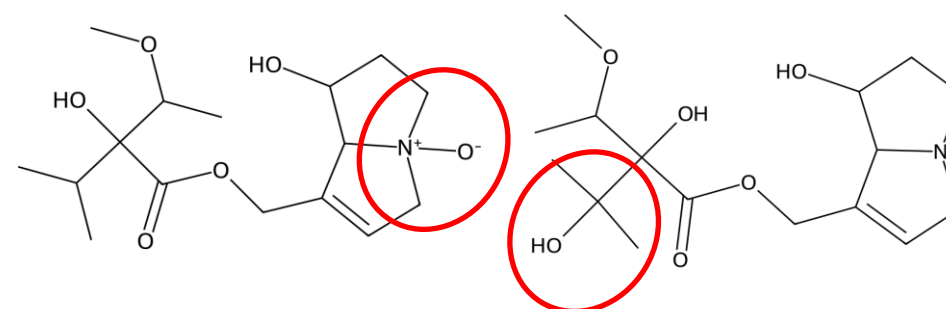
Pesticides

Malathion & Isomalathion

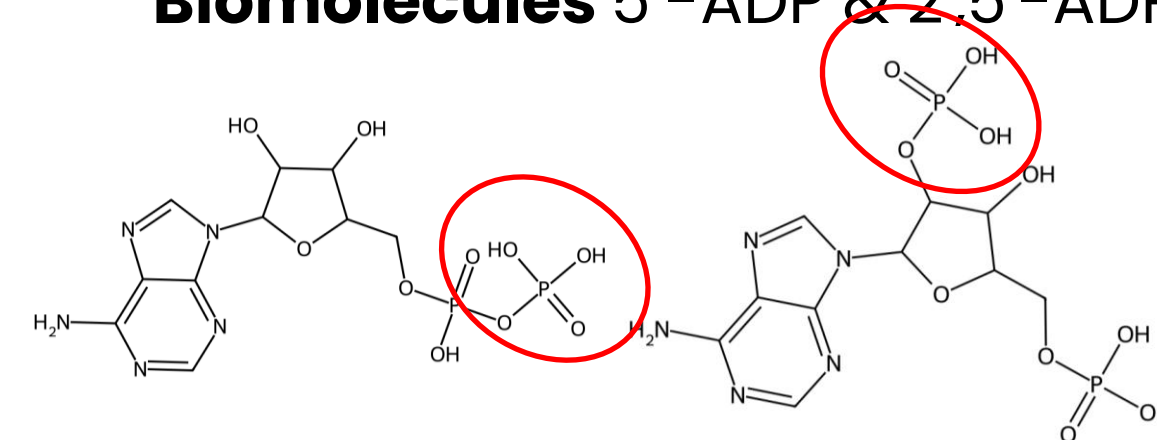


Alkaloids

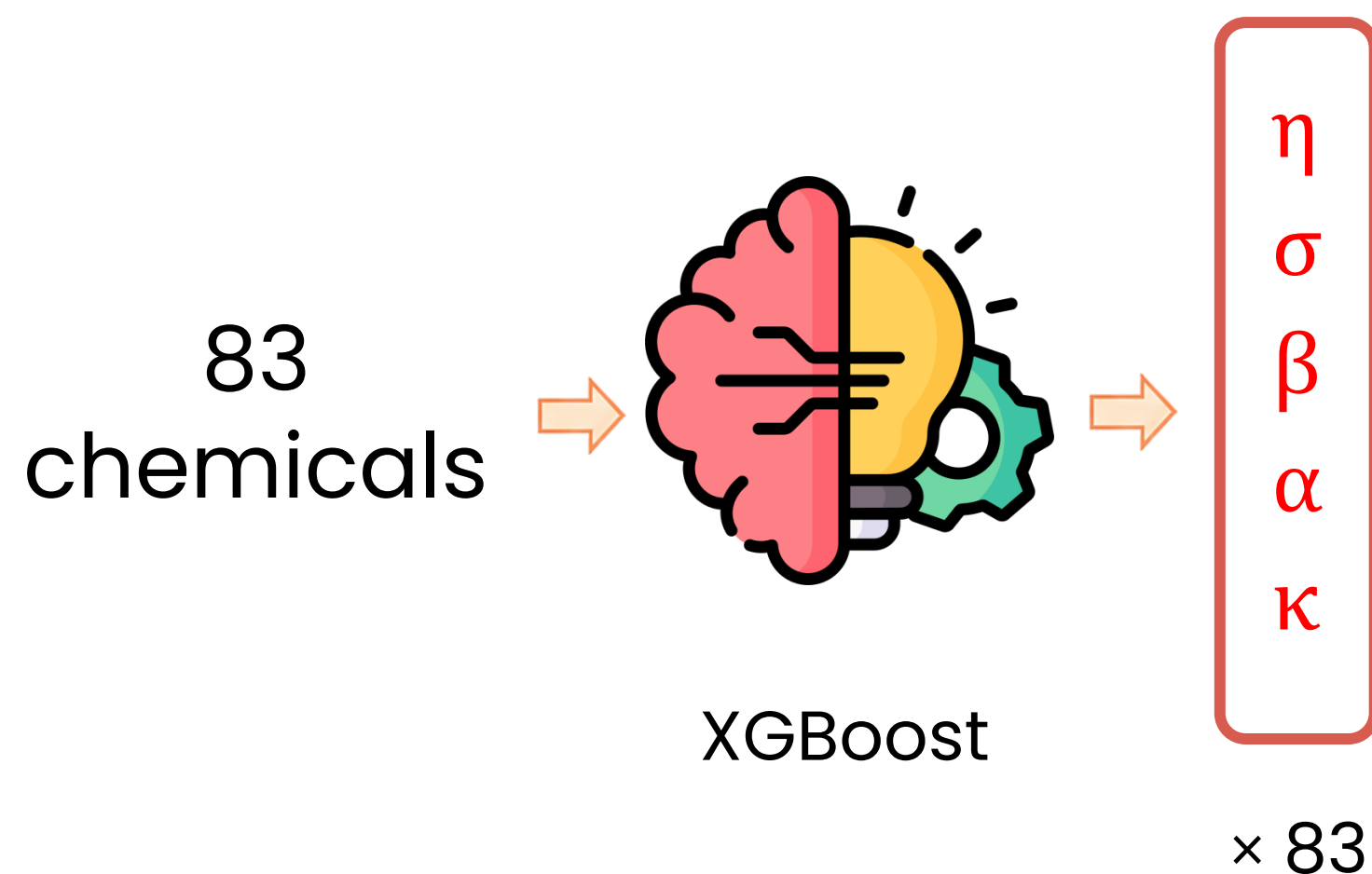
Heliotrine N-oxide & Europine



Biomolecules 5'-ADP & 2',5'-ADP



Predicting isomer selectivity across 799 columns



Isomer 1 Isomer 2

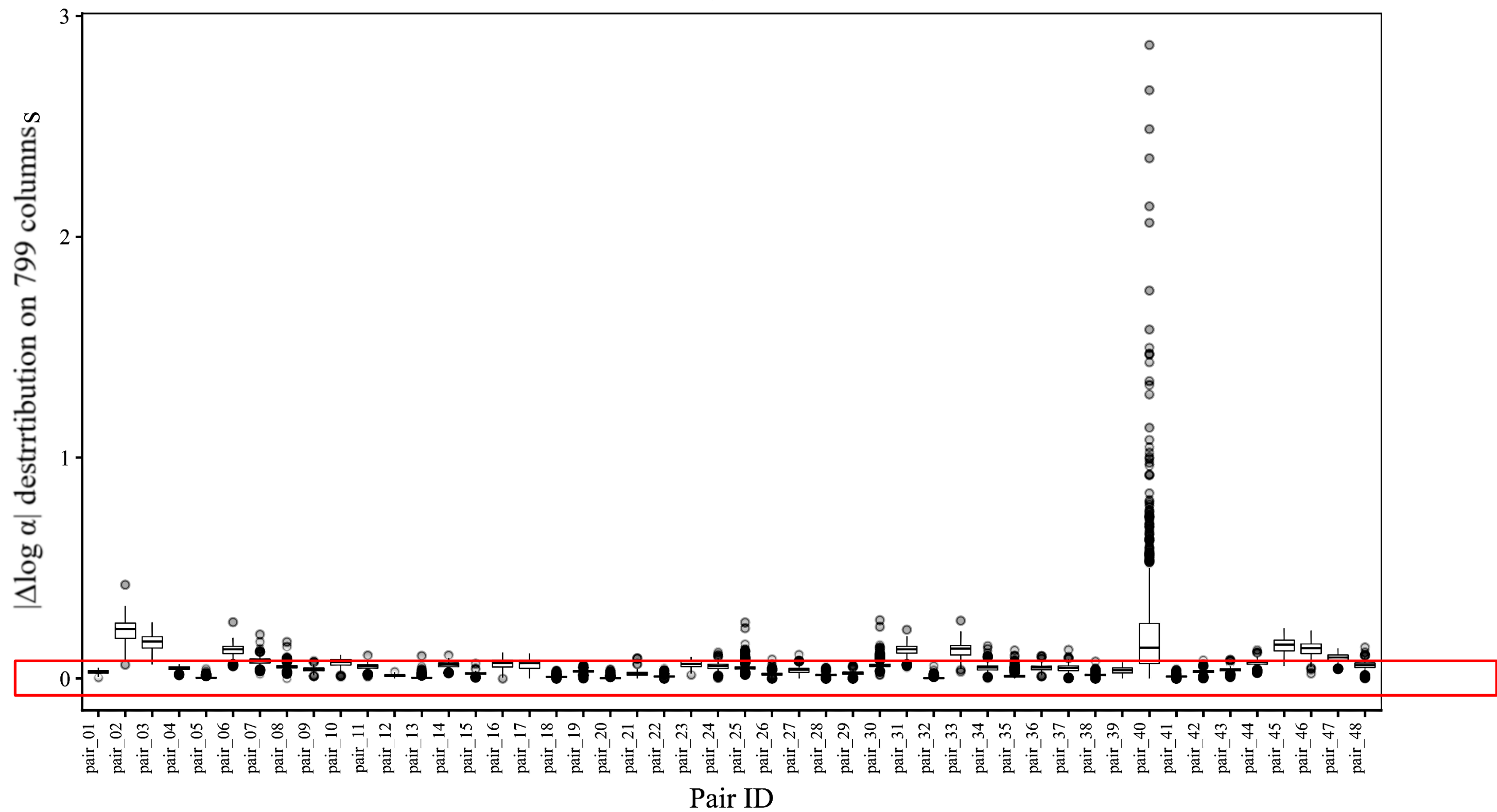
$$\left\{ \begin{array}{l} \log \left(\frac{k_1}{k_{ref}} \right) = \eta_1 H - \sigma_1 S^* + \beta_1 A + \alpha_1 B + \kappa_1 C \\ \log \left(\frac{k_2}{k_{ref}} \right) = \eta_2 H - \sigma_2 S^* + \beta_2 A + \alpha_2 B + \kappa_2 C \end{array} \right.$$

$$\Delta \log \alpha = \log \left(\frac{k_1}{k_{ref}} \right) - \log \left(\frac{k_2}{k_{ref}} \right) = \log \left(\frac{k_1}{k_2} \right)$$

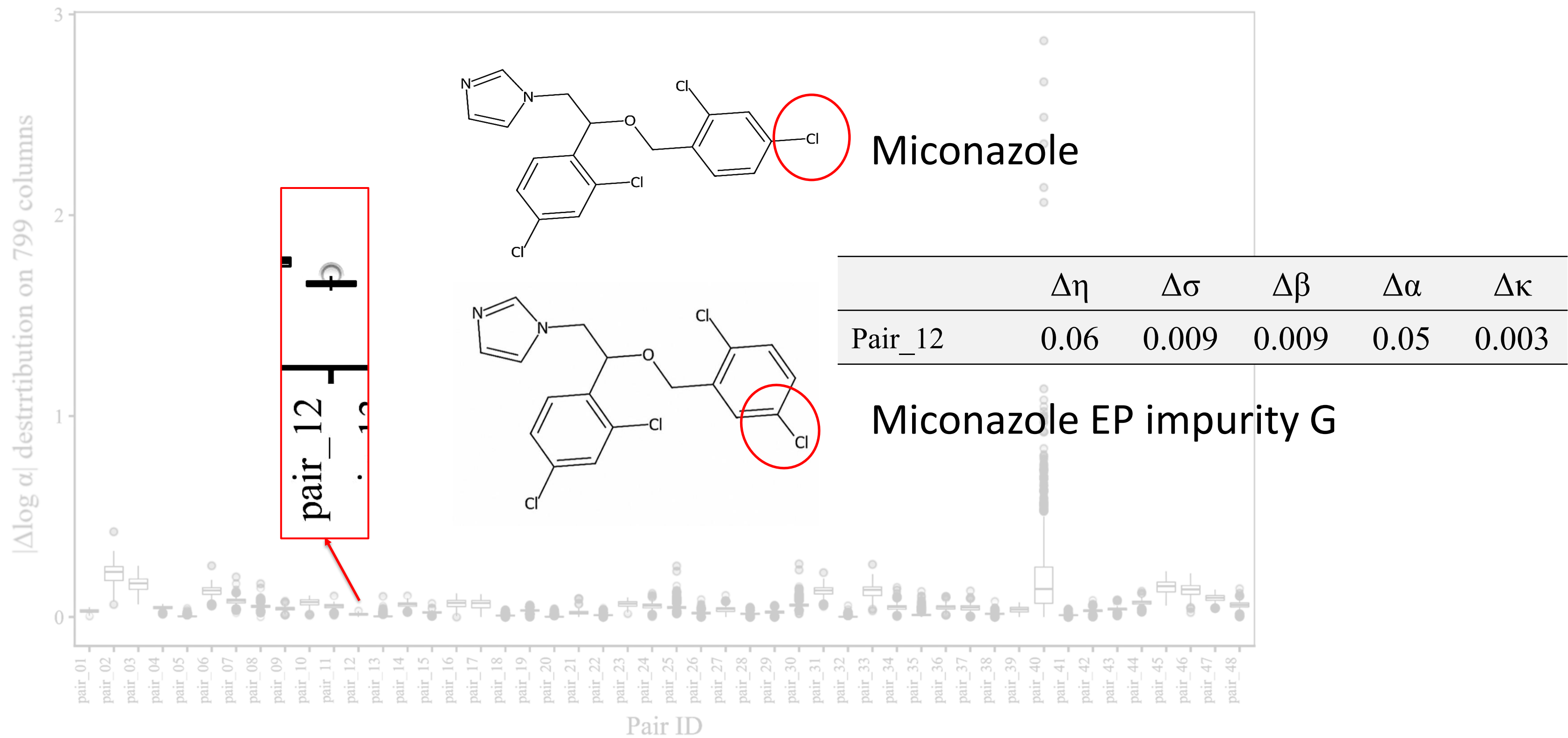
48 × 799

Overview of Prediction Result

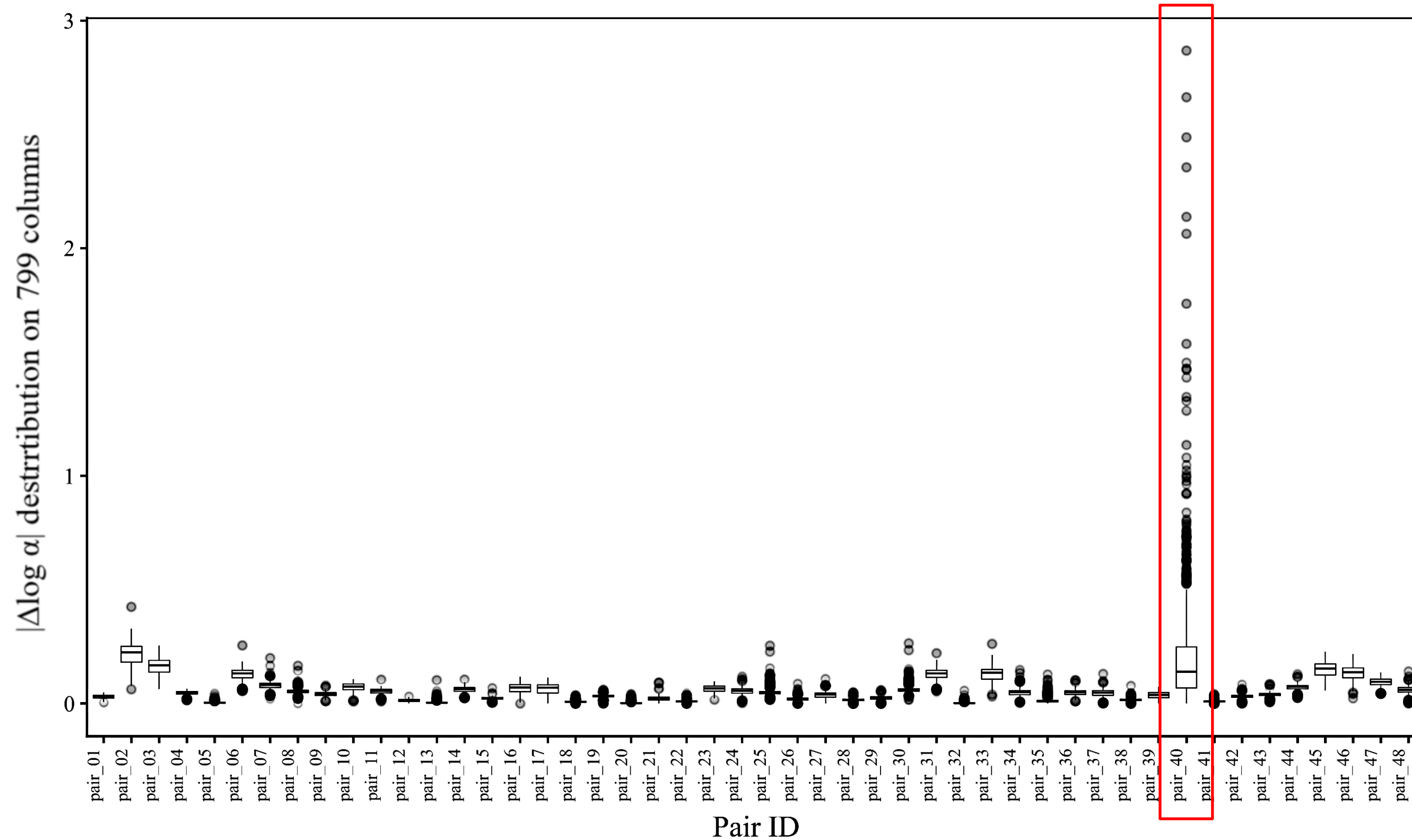
Pair wise distribution across all columns



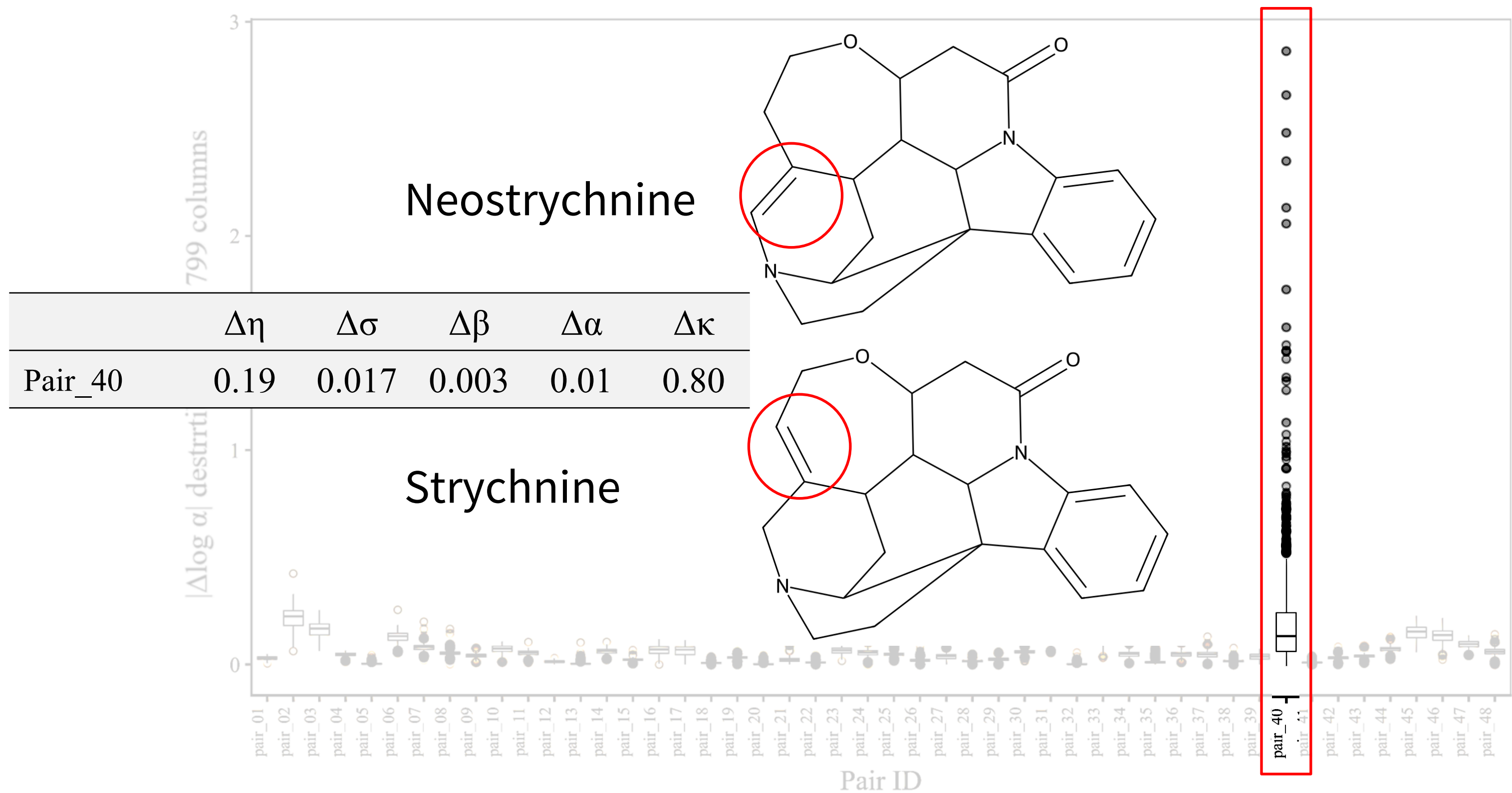
Pair wise distribution across all columns



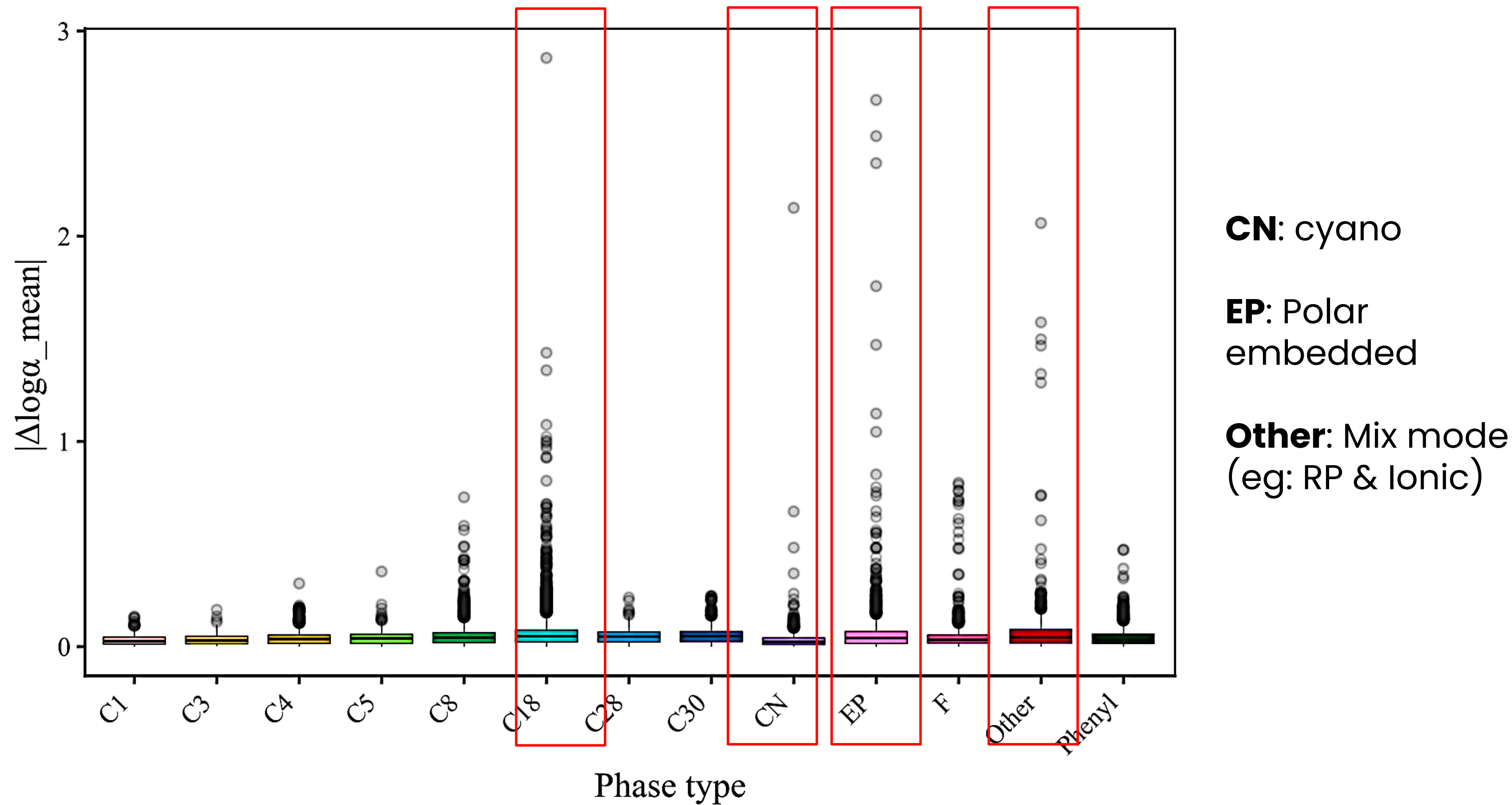
Pair wise distribution across all columns



Pair wise distribution across all columns



Phase wise distribution



Experimental Evaluation

Gradient vs. Isocratic

Isocratic : HSM1 experimental condition



Gradient : Real LC-MS NTS routine

83
chemicals

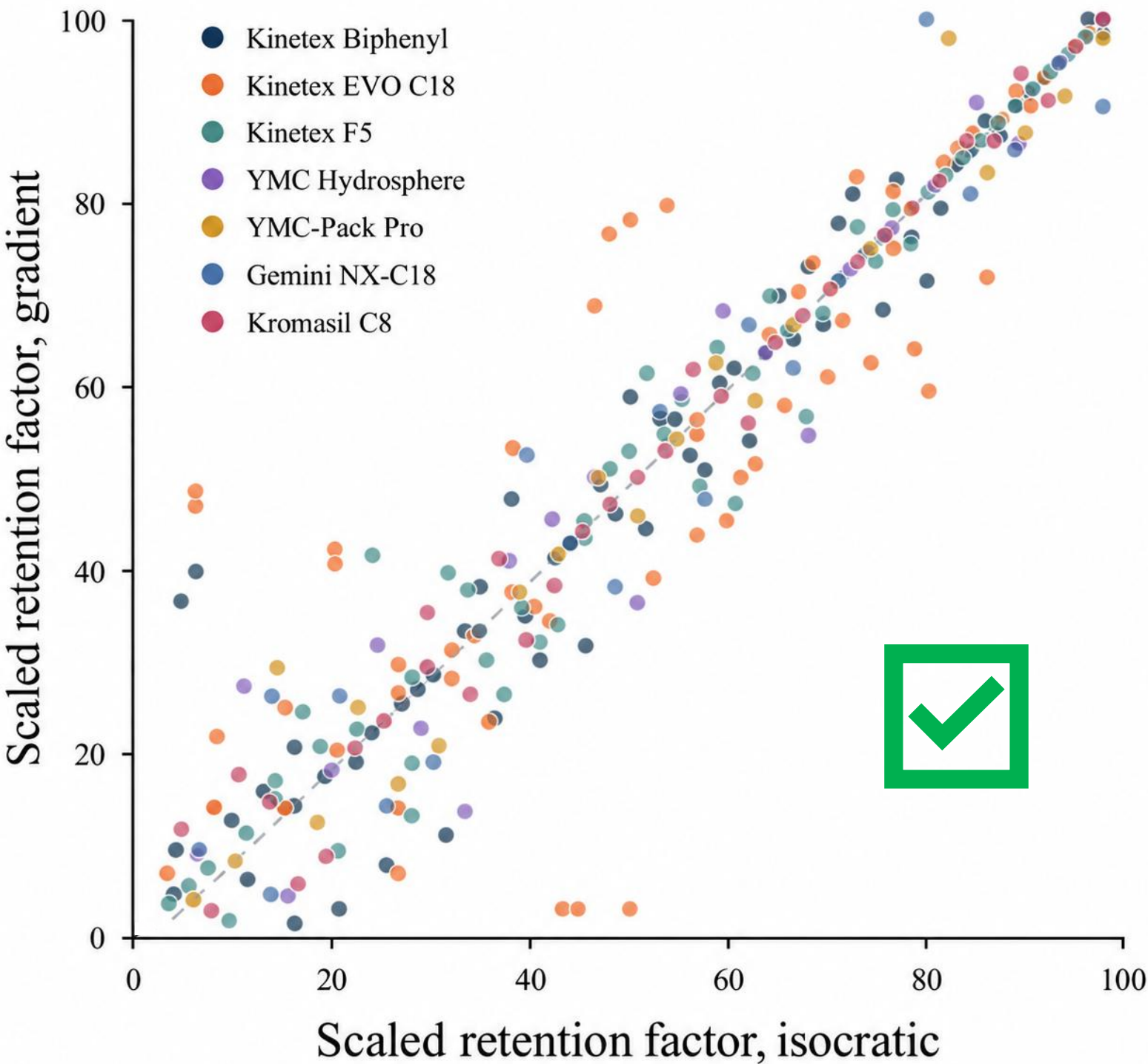


7 columns:
Biphenyl, C8, C18, F5



Gradient
&
Isocratic

Spearman correlation = 0.94
Across pooled 281 valid
measurements

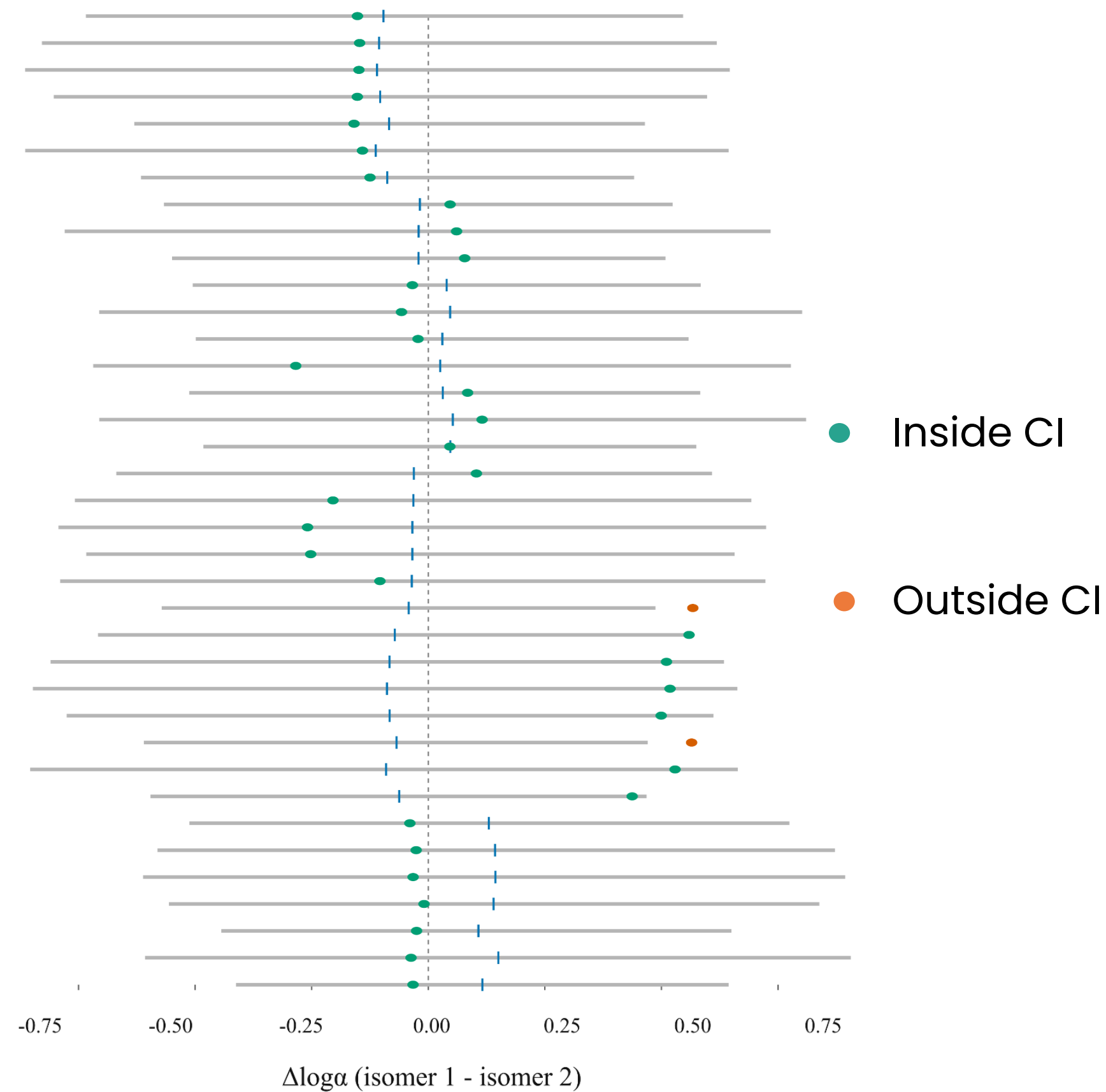


Experimental vs Prediction

35/37

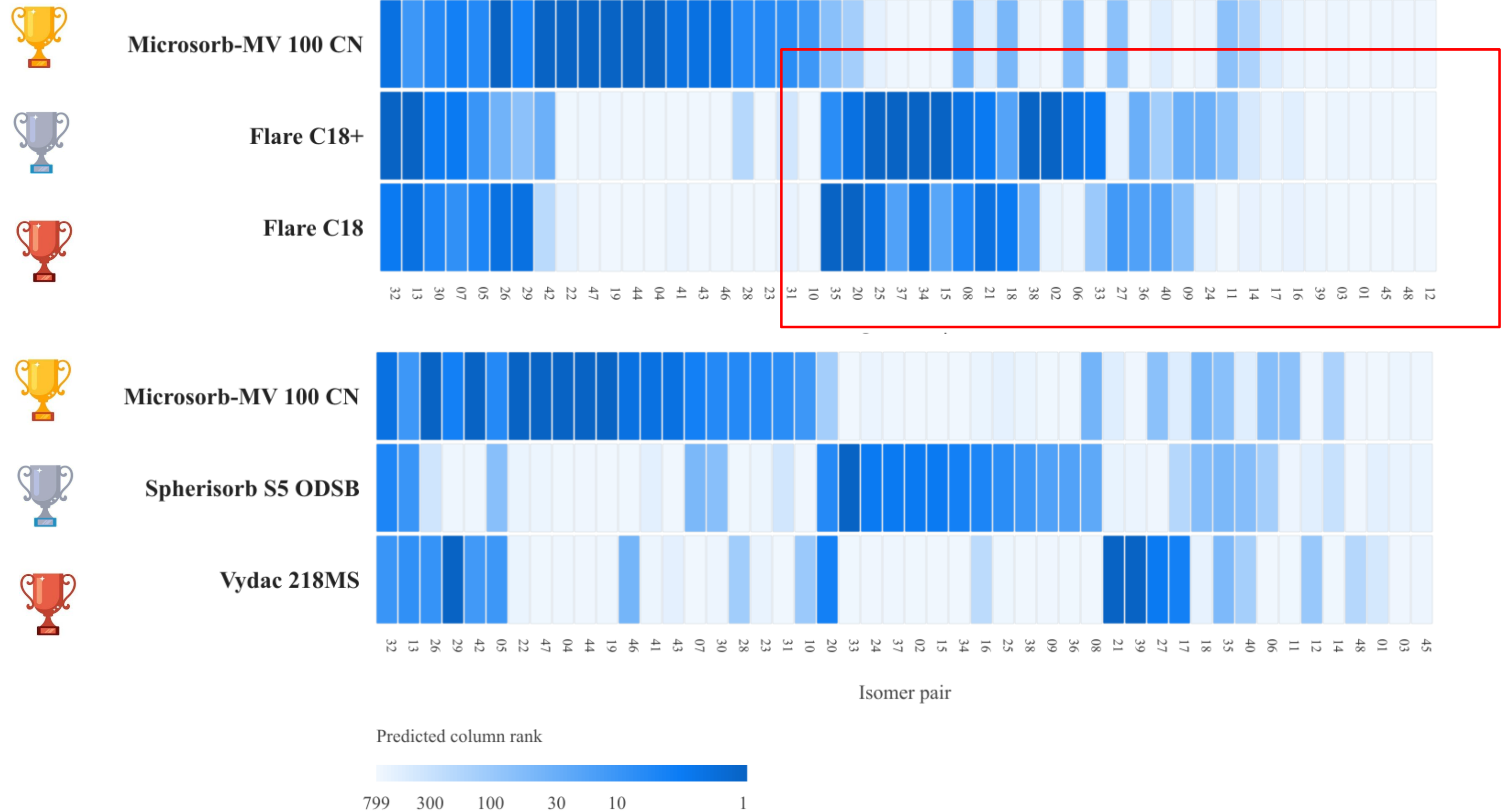
From 8 pairs on 7 columns
94.6% coverage

Experimental values fell inside
the predicted 95%
Confidence Interval



Column Suggestion

Preliminary Column Suggestion



Summary

Conclusion

- XGBoost–HSM predictions: 48 isomer pairs × 799 columns
- Experimental evaluation and preliminary column selection
- Potential to reduce experimental column screening

Limitation

- Limited in-house experimental data & column diversity

Future

- Further test suggested columns
- Apply to NTS of real environmental samples.

Thank You!

Data-driven selection of
stationary phases for MS²-
similar environmental isomers

Presenter: Ruchun Chen

Supervisor: Prof. Anneli Kruve, Dr. Eliise Tammekivi

ruchun.chen@su.se

Kruvelab.com



European Research Council
Established by the European Commission

