Integrating Data Science in MS: Insights into MS/MS Fragmentation and Differential Mass Spectrometry Behavior WP 050

Stepan Stepanovic
Gérard Hopfgartner

University of Geneva
Life Sciences Mass Spectrometry
Geneva Switzerland

Overview

- The complexity of small molecule MS fragmentation and Differential Mobility Spectrometry (DMS) data presents a substantial challenge.
- Our study explores the intersection of data science and mass spectrometry we aim to decipher the complex MS data of various molecules, linking fragmentation patterns to molecular structures.

Introduction

In our research, we harness the power of cheminformatics, molecular modeling, and machine learning to decode MS data(m/z, intensity, mobility…). By systematically exploring small molecule and peptide behaviors across CID, EAD, UVPD, and DMS techniques, we unveil critical patterns that enhance our understanding of molecular structures and their influence on MS-based multiomics studies.

Methods

Data analysis pipeline

We developed a completely automated protocol to clean and featurize data from chromatography, ion mobility, and mass spectrometry. All available molecular data, analyte names and peptide sequence, were directly imported into pandas data frame, after which we converted molecular names to SMILES using pubchempy while peptide sequences were converted to .pdb files using AlphaFold2 (ColabFold). Both smiles and .pdb structures were converted to RDKit mol objects, which were a starting point for all fur-ther property calculations (MW, logP, TPSA), feature extraction (number of aromatic rings in a molecule, fingerprints) and proton affinity DFT calculations, labeled as Gb.

DFT calculations

All DFT calculations were carried out using the Amsterdam Density Functional (ADF) program within the Amsterdam Modelling Suite (AMS2021).[1] Initial structures were optimized with the PBE DFT method using TZ2P basis and Grimme G4 dispersion correction, and the nature of protonation during ESI in DMS analyses was evaluated using the Differential Proton Affinity approach.

Liquid chromatography - mass spectrometry analysis of metabolite, peptide and pesticide mix

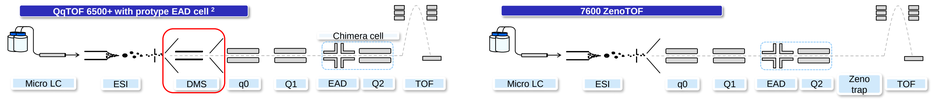

Figure 1. *QqTOF 6600+ with protype EAD cell; 7600 ZenoTOF (AB Sciex, Condord, ON, Canada)*

- MS/MS: Modifications to the ZenoTOF 7600 and Qtrap 6500+ (Sciex, ON) added UVPD, CID, and EAD capabilities, with the EAD cell enabling UVPD via 266 nm (Nd:YAG, 19 kHz, TeemsPhotonics) and 213 nm lasers (solid-state, 1 kHz, CryLas). SWATH was employed for initial detection, facilitating swift CID, EAD, and UVPD transitions.
- DMS: A TTOF 6600+ (Sciex, Concord, ON, Canada) with SelexION DMS analyzed a urine and plasma metabolite mix as well as 185 peptides from 92 proteins, using a DMS cell in a 5500 QTRAP.

Results

1. *CID, EAD and UVPD fragmentation of 180 pesticide molecules*

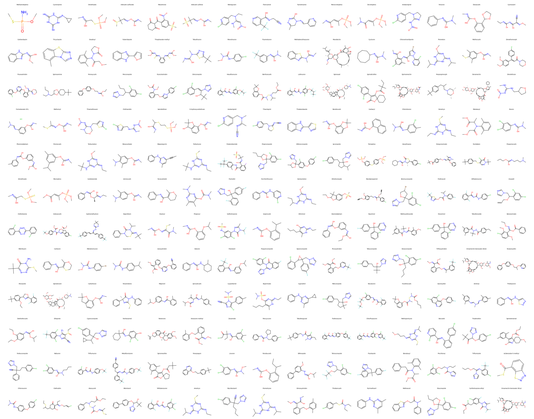

Figure 2. *180 small molecule pesticides subjected to CID, EAD and UVPD fragmentation*

- CID fragmentation at 10-100 eV, EAD at 0-25 eV, and UVPD at 266 nm.
- Data groups into two distinctive clusters based on total/matched peaks, best matching EAD/CID energies,
- and cos similarity scores.
- 90% accuracy in predicting cluster labels using 36 functional groups.
- Clusters indicate the extent of overlap between CID, EAD, and UVPD, showing how much new information
- each fragmentation technique provides.
- For molecules in green cluster, CID/EAD/UVPD all bring significant amount of complementary information

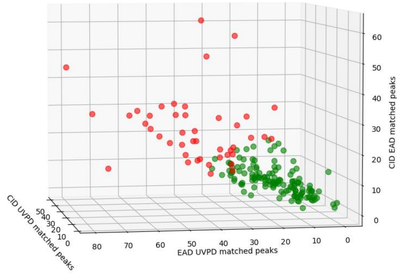

Figure 3. *K-means Clustering Using Selected Features*

2. *Liquid chromatography - differential mobility spectrometry - mass spectrometry on 50 metabolites present in plasma and urine*[3]

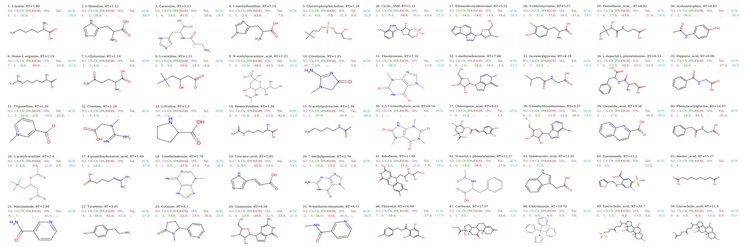

Figure 4. *Structure of mix 50 analytes with the respective retention time and Compensation Voltage (CoV) values (N 2 , 1.5% mole ratio cyclohexane (Ch), ethanol (EtOH), isopropanol (IPA), toluene (Tol), acetonitrile (ACN), and one binary modifier: 0.05% mole ratio IPA in Ch)*

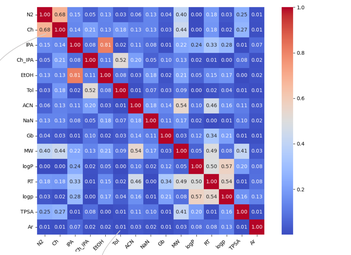

Figure 5. *R² matrix*

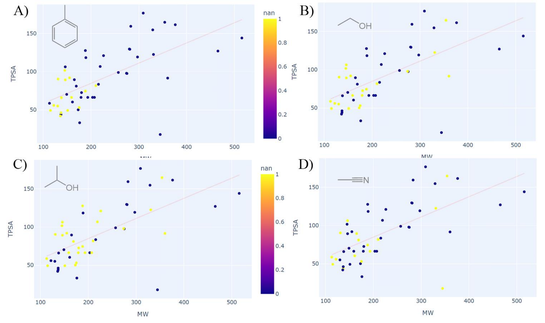

Figure 6. *Graphical representation of signal suppression with various modifiers. Total polar surface area (TPSA) versus molecular weight (MW); points are colored in accordance with the presence (blue) or absence (yellow) of CoV signal.*

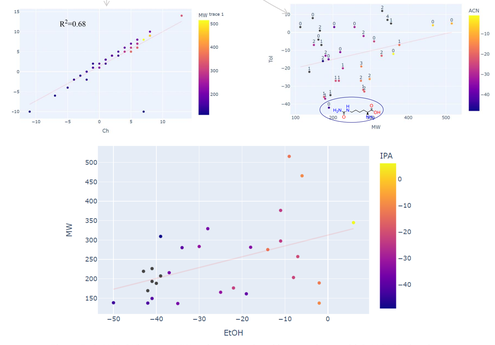

Figure 7. *CoV (EtOH) versus MW, points are colored in accordance withCoV (IPA) signal*

3. *Liquid chromatography - differential mobility spectrometry - mass spectrometry on 185 peptides*[3]

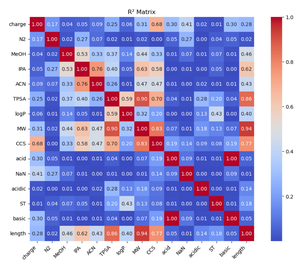

Figure 8. *R² matrix*

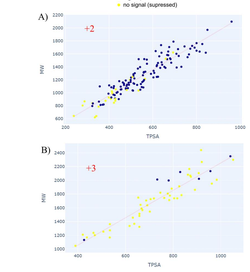

Figure 9. *: TPSA vs MW, points are colored in accordance with the signal suppression. A) peptide charge =2+, B) peptide charge =3+.*

- Objective: Predict peptide detection with polar modifier using ML classification.
- Features: Peptide charge, CoV(N2), TPSA, MW, number of basic AAs (base), Ser and Thr residues (ST), acidic sidechains (acid), peptide length.
- Methods: Scaled features, fivefold cross-validation, grid hyperparameter search.
- Models Tested: SVM, Decision Tree, Random Forest, XGBoost, KNN, Logistic Regression, Ridge, Logistic Regression CV.
- Best Performance: Random Forest and Logistic Regression CV.
- Selected Model: Logistic Regression CV - Accuracy: 0.89, Precision: 0.93, Recall: 0.81, F1 Score: 0.87, AUC-ROC: 0.90.
- Feature Importance: Peptide charge, MW, TPSA, peptide length identified via coefficient analysis, permutation feature importance, SHAP values, and RFE.

Conclusions

- Overlapping between CID, EAD, and UVPD fragmentation techniques for small molecule pesticides is clearly related to the FGs present in the molecules.
- Systematic exploration of MS challenges reveals significant correlations between the chemistry of small molecules/peptides and their molecular structures/properties

References

1. ADF 2021.1, SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands, http://www.scm.com.

2. Baba, T.; Campbell, J. L.; Le Blanc, J. C.; Hager, J. W.; Thomson, B. A. Electron capture dissociation in a branched radio-frequency ion trap. *Anal Chem* **2015**, *87* (1), 785-792.

3. Stepanovic S.; Ekmekciu L.; Alghanemac B.; Hopfgartner G. Insights into Modifiers Effects in Differential Mobility Spectrometry: A Data Science Approach for Metabolomics and Peptidomics *J. Mass Spectrom.* **2024**