iGRASP: A Modular Platform for Improved, Graph-Based Spectra Prediction

Nick Wierich [1] **Philipp M. Pflüger** [1] **Frank Glorius** [1]

[1] *University of Münster, Institute of Organic Chemistry, Corrensstraße 40, 48149 Münster, Germany*. Correspondence to: Frank Glorius glorius@uni-muenster.de.

1. Introduction

Structural elucidation of unknown compounds is of central importance for a diverse range of processes, such as analyzing the (by)products of chemical reactions, characterizing drug metabolites or identifying environmental contaminants. Out of the analytical methods available for this purpose, the combination of gas chromatography (GC) and electron ionization mass spectrometry (EI-MS) is particularly versatile, providing comparably deep structural information while allowing the analysis of crude mixtures using nanogram amount. [1]

While the EI mass spectrum of a given compound is the result of stochastic, yet reproducible and characteristic fragmentation processes, deducing an unknown compound's structure based on its spectrum remains challenging to date. In contrast to methods such as nuclear magnetic resonance (NMR), the extent to which humanly interpretable rules for the outcome of the probability-based fragmentation processes in EI-MS can be formulated is significantly lower. [2]

At the same time, the reproducible outcome allows for matching of an unknown compound's spectrum against a library of experimental spectra. Inherently, this library matching approach can only identify compounds that have been measured before, significantly restricting its practical applicability. [3] Since in many applications candidate structures are available, however, the ability to reliably predict their EI-MS spectra would allow for quick verification of these hypotheses by matching such predictions against the experimental spectrum.

Motivated by this need, several ML models have been developed for the purpose of predicting EI-MS spectra during recent years. [4–7] Yet, while they do allow for significantly faster predictions than conventional, ab-initio computational methods – which can practically only be used for very small molecules [8] – the quality of their predicted spectra still leave room for improvement. At the same time, the currently most accurate model is significantly restricted regarding the size and type of processable molecules, severely limiting its practical applicability. [4]

2. Spectra Prediction Framework

We herein present a versatile, modular platform for the development of graph-based EI-MS spectra prediction models, which we call *improved graph-based spectra prediction* (iGRASP) (Figure 1). In two separate modes of operation, both training and forward spectra prediction are supported. Altogether, iGRASP covers all steps within the ML development

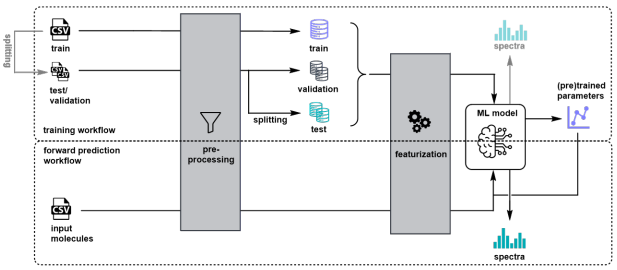

Fig. 1: iGRASP platform supporting model training, evaluation and forward spectra prediction.

lifecycle: From selecting suitable model architectures and parameters, via filtering the raw input dataset based on defined criteria, to subsequent training and validation to comparing different models.

In contrast to previous implementations, iGRASP is written in pure Python and can easily be modified, adapting to any desired use case. The entire training and testing process is built in a modular fashion by combining interchangeable classes, representing fundamental building blocks.

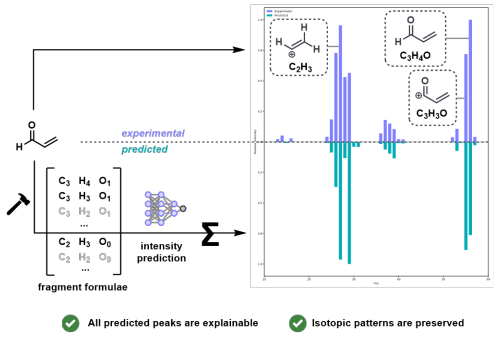

Fig. 2: Spectra prediction *via* fragment formulae, eliminating chemically impossible peak patterns.

Using this platform, we developed and trained a model achieving state-of-the-art accuracy of 0.88 mean spectral dot product across the NIST2017 [9] dataset (Figure 2). At the same time, 91% of molecules within that database can be processed, and any remaining restrictions can easily be tuned or removed by changing the configuration (Table 1). iGRASP thus represents a versatile basis for accelerating further research in the field of EI-MS spectra prediction.

Table 1: Comparison between our best candidate model and the current state of the art. [4]

	RASSP (current SOTA)	**iGRASP (this work)**
NIST2017 coverage	48%	**95%**
molecules/s	8.7	**84**
mean dot product	0.86	**0.88**

Acknowledgments

The European Research Council (ERC Advanced Grant Agreement No. 101098156, HighEnT F.G.) is gratefully acknowledged for generous financial support of this work.

References

[1] Nadine Strehmel, Joachim Kopka, Dierk Scheel, and Christoph Böttcher. Annotating unknown components from GC/EI-MS-based metabolite profiling experiments using GC/APCI(+)-QTOFMS. *Metabolomics*, 10(2):324–336, 2014.

[2] Emma L. Schymanski, Christine M. J. Gallampois, Martin Krauss, Markus Meringer, Steffen Neumann, Tobias Schulze, Sebastian Wolf, and Werner Brack. Consensus Structure Elucidation Combining GC/EI-MS, Structure Generation, and Calculated Properties. *Analytical Chemistry*, 84(7):3287–3295, 2012.

[3] Sebastian Böcker. *Algorithmic Mass Spectrometry*. Lehrstuhl Bioinformatik, Friedrich-Schiller-Universität Jena, 2019.

[4] Richard Licheng Zhu and Eric Jonas. Rapid Approximate Subset-Based Spectra Prediction for Electron Ionization–Mass Spectrometry. *Analytical Chemistry*, 95(5):2653–2663, 2023.

[5] Jennifer N. Wei, David Belanger, Ryan P. Adams, and D. Sculley. Rapid Prediction of Electron–Ionization Mass Spectrometry Using Neural Networks. *ACS Central Science*, 5(4):700–708, 2019.

[6] Felicity Allen, Allison Pon, Russ Greiner, and David Wishart. Computational Prediction of Electron Ionization Mass Spectra to Assist in GC/MS Compound Identification. *Analytical Chemistry*, 88(15):7689–7697, 2016.

[7] Christoph Ruttkies, Emma L. Schymanski, Sebastian Wolf, Juliane Hollender, and Steffen Neumann. MetFrag relaunched: incorporating strategies beyond in silico fragmentation. *Journal of Cheminformatics*, 8(1):3, 2016.

[8] Christoph Alexander Bauer and Stefan Grimme. How to Compute Electron Ionization Mass Spectra from First Principles. *The Journal of Physical Chemistry A*, 120(21):3755–3766, 2016.

[9] National Institute of Standards and Technology. Mass Spectral Library. `https://chemdata.nist.gov`, 2017. Accessed: 26.01.2026.