IRonBranch.AI: Branching IR data into molecular form

AI-Driven Molecular Structure Prediction

***N. Charron, A. Krishnadas, J. Kansal, M. Läuter, A. Ragyanszki**, Zuse Institute Berlin*

In Short

- Infrared (IR) spectroscopy is widely used to determine molecular structures, yet the complexity of IR spectra often prevents straightforward manual interpretation.
- Artificial intelligence (AI) methods trained on quantum chemical data can effectively capture complex spectral patterns, enabling accurate predictions of molecular structures from IR spectra.
- The integration of computational chemistry, advanced ML algorithms, and experimental IR spectroscopy provides deeper insights into molecular reaction pathways.
- The developed AI-based framework significantly accelerates molecular structure prediction

In molecular sciences, one of the most complex challenges is predicting the structure of molecules based on experimental data. This task plays an important role in many practical scientific and industrial applications, where accurate structure prediction can lead to the discovery of new drugs or advances in sustainable energy solutions. For example, identifying accurate molecular structures in drug design allows for safer and more effective therapies. In materials science, molecular structure prediction plays an important role in designing materials with better sustainability. For renewable energy, proposing new molecular structures may lead to improved battery design and other innovative technologies.

While scientists have traditionally used manual methods for identifying molecular structures, such approaches are difficult and time-consuming. Additionally, even within the known laws of chemistry and physics, the “chemical space” of possible molecular structures is almost infinite 5. Today, advances in artificial intelligence (AI) offer potential solutions to these problems, allowing for faster and more accurate structure predictions by learning patterns and chemical rules from existing data. The direct prediction or design of molecule structures from various physical measurements or desired properties marks a significant frontier in molecular science 1

Machine Learning (ML), a subset of AI, has made great strides in answering this inverse, molecular design problem from known, measured, or desired properties. Leveraging traditional data, such as measurements from techniques like infrared spectroscopy (IR), presents an avenue to unravel unknown yet significant molecular structures 2.

Furthermore, recent developments of High-Performance Computing (HPC) hardware for AI provide new opportunities to address this challenge. By utilizing cutting-edge computational power, it is possible to combine the principles of computationally highly demanding quantum mechanics with the predictive capabilities of AI/ML in order to learn patterns and make predictions from large datasets.

Despite these advancements in HPC and AI/ML, currently these models often require highly specialized data, meaning their performance is strongly dependent on the type of experimental technique used to provide the data. Additionally, preparing datasets, curating, preprocessing and training ML models, is a time-consuming endeavor.

The primary focus of our project is to develop a structure prediction method for molecules based on computed (e.g. IR) spectra. While experimental techniques like IR generate detailed data about molecules, analyzing this data to uncover molecular structures is a complex task. AI models trained on datasets of spectral measurements can learn to identify the underlying molecular structures with high accuracy and efficiency 4. The ultimate goal is to develop an ML-based model, trained on highly accurate computational quantum chemistry data, which is able to predict molecular structures directly from IR spectra. While we plan to investigate accurate structure prediction from IR spectra, we will develop a set of universal and open-source software tools that can be adapted to any type of spectra.

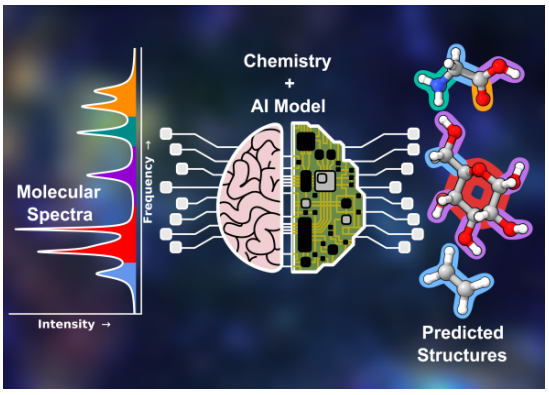

***Figure 1: Project overview:** This project will provide methods for inverse molecular structure prediction directly from spectroscopy measurements.*

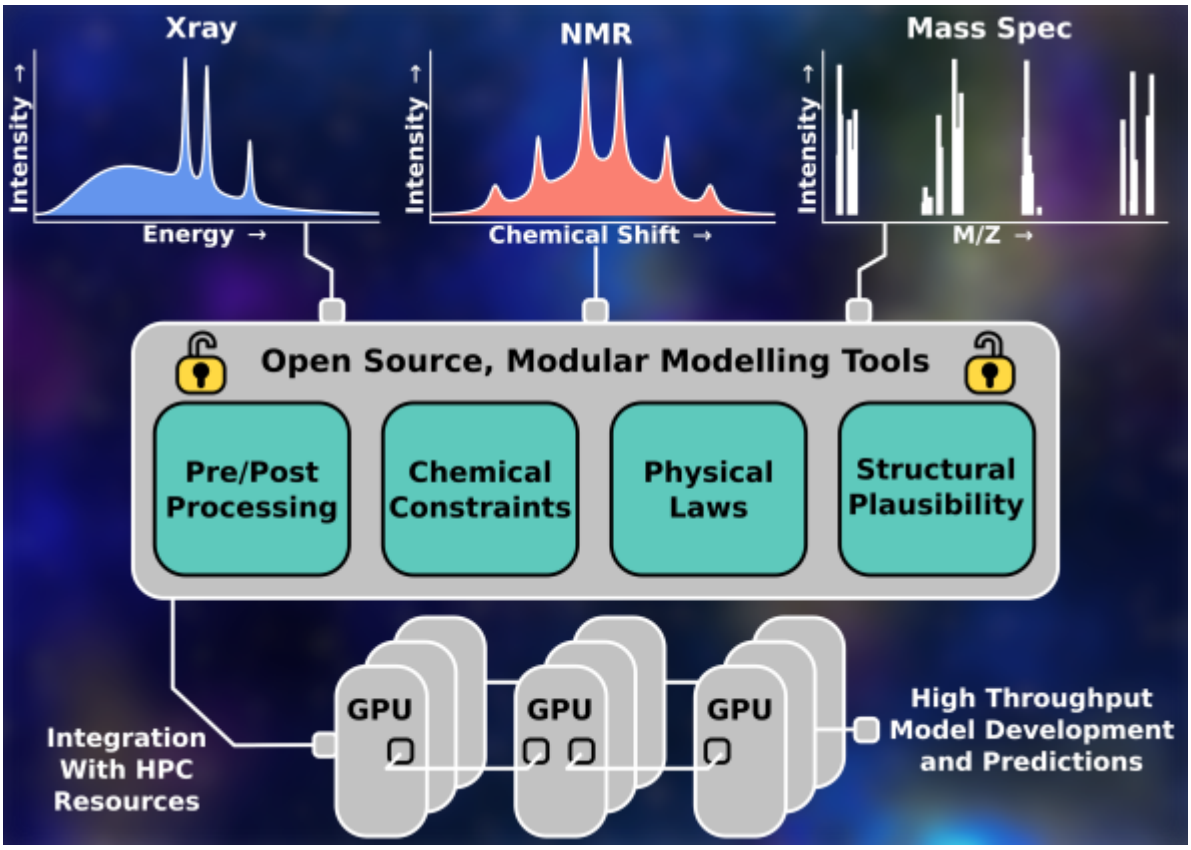

Figure 2: *An adaptable methodology leveraging multiple spectroscopic techniques to enable efficient AI model training and high-throughput molecular structure predictions.*

Ultimately, these tools have the potential for broad impact from many experimental measurements, and may lower the barrier for making chemical predictions using ML and HPC. 3 This interdisciplinary approach holds immense promise in advancing our understanding of molecular behavior and has far-reaching implications across various scientific domains.

WWW

https://www.nhr-verein.de/en/AIC

More Information

[1] S. A. Al, AR, Allouche, *Chem. Phys. Let.* **856**, 141603 (2024). doi: 10.1016/j.cplett.2024.141603.

[2] R. Beckmann, F. Brieuc, C. Schran, D. Marx, *J. Chem. Theory. Comput.* **9**, 5492 (2022). doi: 10.1021/acs.jctc.2c00511.

[3] G. Jung, SG. Jung, JaM Cole, *Chem. Sci.* **14**, 3600 (2023). doi:10.1039/D2SC05892H

[4] Litsa, Eleni E. and Chenthamarakshan, Vijil and Das, Payel and Kavraki, Lydia E., *Commun. Chem.* **6**, 1 (2023). doi:10.1038/s42004-023-00932-3

[5] Gómez-Bombarelli et al, *ACS Cent. Sci.* **4**, 268 (2018). doi:10.1021/acscentsci.7b00572

Funding

BMBF ZIB-ZP-1-24-26

DFG Subject Area

303-01