

## From Spectra to Structure: Automated structure elucidation for organic chemistry

Marvin Alberts<sup>1,2</sup>, Teodoro Laino<sup>1</sup>, Nina Hartrampf<sup>2</sup>

<sup>1</sup> IBM Research Zürich, 8134 Rüschlikon, Switzerland

<sup>2</sup> Department of Chemistry, University of Zürich, 8006 Zürich, Switzerland

[marvin.alberts@ibm.com](mailto:marvin.alberts@ibm.com)

The initial phases of drug discovery are inherently driven by the speed at which new molecules can be synthesised, characterised and evaluated.<sup>1</sup> The introduction of robotics and automation into the chemical laboratory reactions promises to significantly speed up the rate at which molecules can be synthesised.<sup>2</sup> Despite this progress, the characterisation of the synthesized compounds is not straightforward and often carried out manually. Consequently, structure elucidation often becomes a time-consuming and tedious undertaking. To this end we explore Transformer models to automatically carry out structure elucidation from routine analytical spectra. We introduce the first model capable of predicting the chemical structure, scaffold and functional groups directly from Infrared spectra.<sup>3</sup> In addition we evaluate a model trained to predict the structure from either <sup>1</sup>H- or <sup>13</sup>C-NMR spectra.<sup>4</sup> We train both models on simulated spectra and evaluate them experimentally promising to significantly accelerate the synthesis and characterization of molecules.

1. Hughes, J., Rees, S., Kalindjian, S. & Philpott, K. Principles of early drug discovery. *Br J Pharmacol* **162**, 1239–1249 (2011).
2. Christensen, M. *et al.* Automation isn't automatic. *Chem. Sci.* **12**, 15473–15490 (2021).
3. Alberts, M., Laino, T. & Vaucher, A. C. Leveraging Infrared Spectroscopy for Automated Structure Elucidation. Preprint at <https://doi.org/10.26434/chemrxiv-2023-5v27f> (2023).
4. Alberts, M., Zipoli, F. & Vaucher, A. C. Learning the Language of NMR: Structure Elucidation from NMR Spectra Using Transformer Models. Preprint at <https://doi.org/10.26434/chemrxiv-2023-8wxcz> (2023).