Docket #: S24-328

A method to predict molecular identity from routine H^{1} and C^{13} NMR

Stanford scientists have developed a machine learning framework that predicts the complete molecular structure of an unknown compound directly from routine 1D NMR spectra with no additional information required, automating one of the most time-consuming and expertise-dependent steps in chemical analysis.

Nuclear magnetic resonance (NMR) spectroscopy is the workhorse technique for characterizing small molecules across the pharmaceutical, chemical, and materials industries, yet interpreting the resulting spectra remains a laborious manual exercise that depends heavily on trained chemists. Structure elucidation from one-dimensional ^{1}H and ^{13}C spectra, the most routinely acquired NMR data, is especially difficult because the number of possible structures grows combinatorially with the number of heavy atoms, quickly reaching into the range of 10^{35} possible structures for commonly encountered molecular sizes. Researchers at Stanford addressed this challenge by building an end-to-end multitask machine learning framework that predicts both molecular formula and atomic connectivity directly from 1D NMR spectra, with no prior chemical knowledge required. The system pairs a convolutional neural network for spectral feature extraction with a transformer-based generative model that assembles molecular fragments into complete candidate structures. On molecules containing up to 40 heavy atoms (C, N, O, P, S, Si, B, F, Cl, Br, I), in which $> 10^{35}$ of structures are theoretically possible, the current prototype of the model identifies the exact correct structure within its top 15 predictions 60.4% of the time, reducing the chemist's search space by up over 35 orders of magnitude. Further improvements in accuracy are expected with ongoing work to access larger training datasets. By automating an analysis step that has historically gated everything from drug discovery to quality control, this technology offers chemical and pharmaceutical organizations a powerful tool to accelerate compound identification and unlock new applications in automated chemical synthesis within a global NMR market currently estimated at $760 million and

growing.

Stage of Development
Prototype

Applications

- Automated interpretation of NMR spectra for chemical structure identification
- High-throughput compound identification for chemical manufacturers and suppliers
- Research tool for academic and industrial chemistry laboratories

Advantages

- Predicts full molecular structure from routine 1D ^{13}H and ^{13}C spectra alone
- Requires no prior chemical knowledge, including molecular formula
- Reduces the candidate search space by over 35 orders of magnitude
- Substantially lowers the expertise required for NMR-based structure elucidation

Publications

- Hu, Frank, et al. Pushing the Limits of One-Dimensional NMR Spectroscopy for Automated Structure Elucidation Using Artificial Intelligence. *Journal of Chemical Information and Modeling/i> June 2026.*
- Hu, Frank, et al. Accurate and Efficient Structure Elucidation from Routine One-Dimensional NMR Spectra Using Multitask Machine Learning. *ACS Central Science/i> November 2024.*

Innovators

- Grant Rotskoff
- Matthew Kanan
- Frank Hu
- Thomas Markland

Licensing Contact

Michael Spaid

Technology Licensing Associate 2

Email