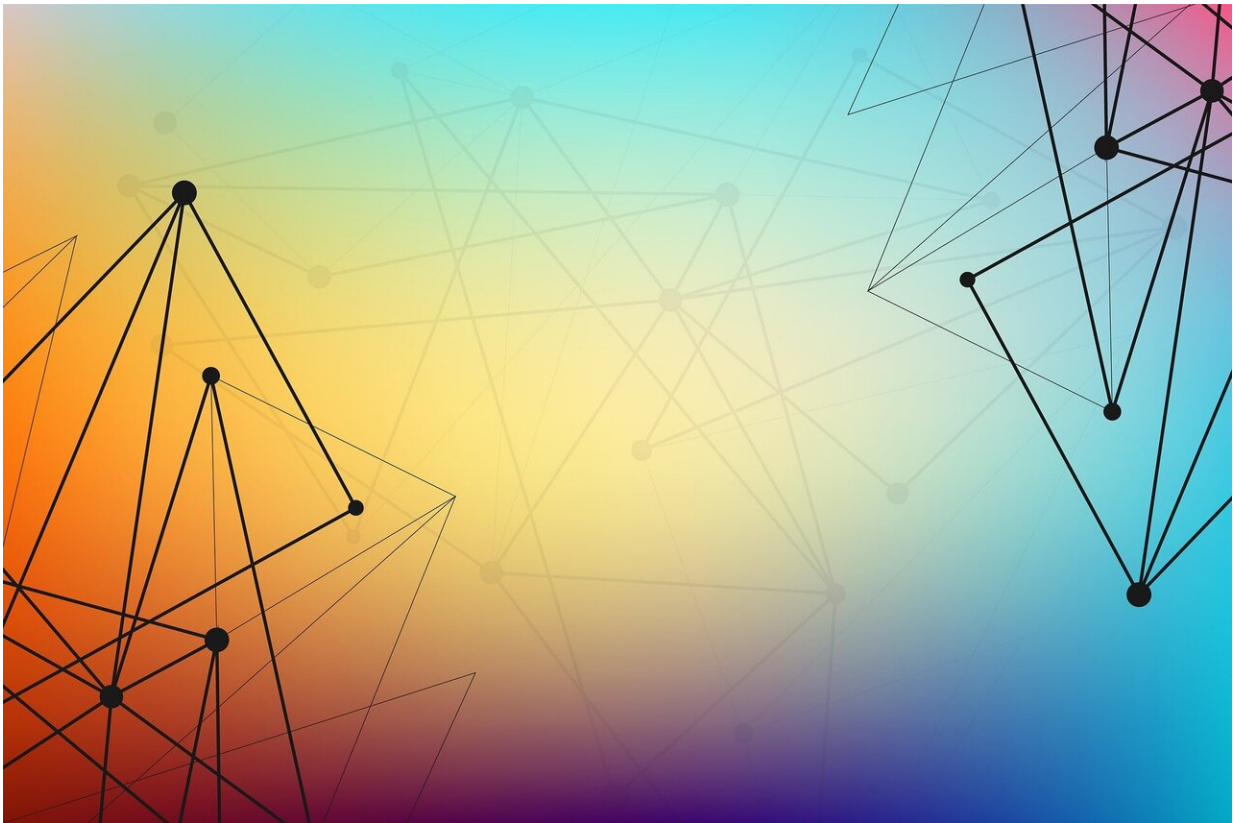


# Algorithm uses mass spectrometry data to predict identity of molecules

June 17 2021

---



Credit: Pixabay/CC0 Public Domain

An algorithm designed by researchers from Carnegie Mellon University's Computational Biology Department and St. Petersburg State University in Russia could help scientists identify unknown molecules.

The algorithm, called MolDiscovery, uses mass spectrometry data from molecules to predict the identity of unknown substances, telling scientists early in their research whether they have stumbled on something new or merely rediscovered something already known.

This development could save time and money in the search for new naturally occurring products that could be used in medicine.

"Scientists waste a lot of time isolating [molecules](#) that are already known, essentially rediscovering penicillin," said Hosein Mohimani, an assistant professor and part of the research team. "Detecting whether a molecule is known or not early on can save time and millions of dollars, and will hopefully enable [pharmaceutical companies](#) and researchers to better search for novel natural products that could result in the development of new drugs."

The team's work, "MolDiscovery: Learning Mass Spectrometry Fragmentation of Small Molecules," was recently published in *Nature Communications*. The research team included Mohimani; CMU Ph.D. students Liu Cao and Mustafa Guler; Yi-Yuan Lee, a research assistant at CMU; and Azat Tagirdzhanov and Alexey Gurevich, both researchers at the Center for Algorithmic Biotechnology at St. Petersburg State University.

Mohimani, whose research in the Metabolomics and Metagenomics Lab focuses on the search for new, naturally occurring drugs, said after a scientist detects a molecule that holds promise as a potential drug in a marine or soil sample, for example, it can take a year or longer to identify the molecule with no guarantee that the substance is new. MolDiscovery uses [mass](#) spectrometry measurements and a predictive machine learning model to identify molecules quickly and accurately.

Mass spectrometry measurements are the fingerprints of molecules, but

unlike fingerprints there's no enormous database to match them against. Even though hundreds of thousands of naturally occurring molecules have been discovered, scientists do not have access to their [mass spectrometry data](#). MolDiscovery predicts the identity of a molecule from the mass [spectrometry](#) data without relying on a mass spectra database to match it against.

The team hopes MolDiscovery will be a useful tool for labs in the discovery of novel natural products. MolDiscovery could work in tandem with NRPminer, a machine learning platform developed by Mohimani's lab, that helps scientists isolate natural products. Research related to NRPminer was also recently published in *Nature Communications*.

**More information:** Liu Cao et al, MolDiscovery: learning mass spectrometry fragmentation of small molecules, *Nature Communications* (2021). [DOI: 10.1038/s41467-021-23986-0](https://doi.org/10.1038/s41467-021-23986-0)

Provided by Carnegie Mellon University

Citation: Algorithm uses mass spectrometry data to predict identity of molecules (2021, June 17) retrieved 2 May 2024 from <https://phys.org/news/2021-06-algorithm-mass-spectrometry-identity-molecules.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.