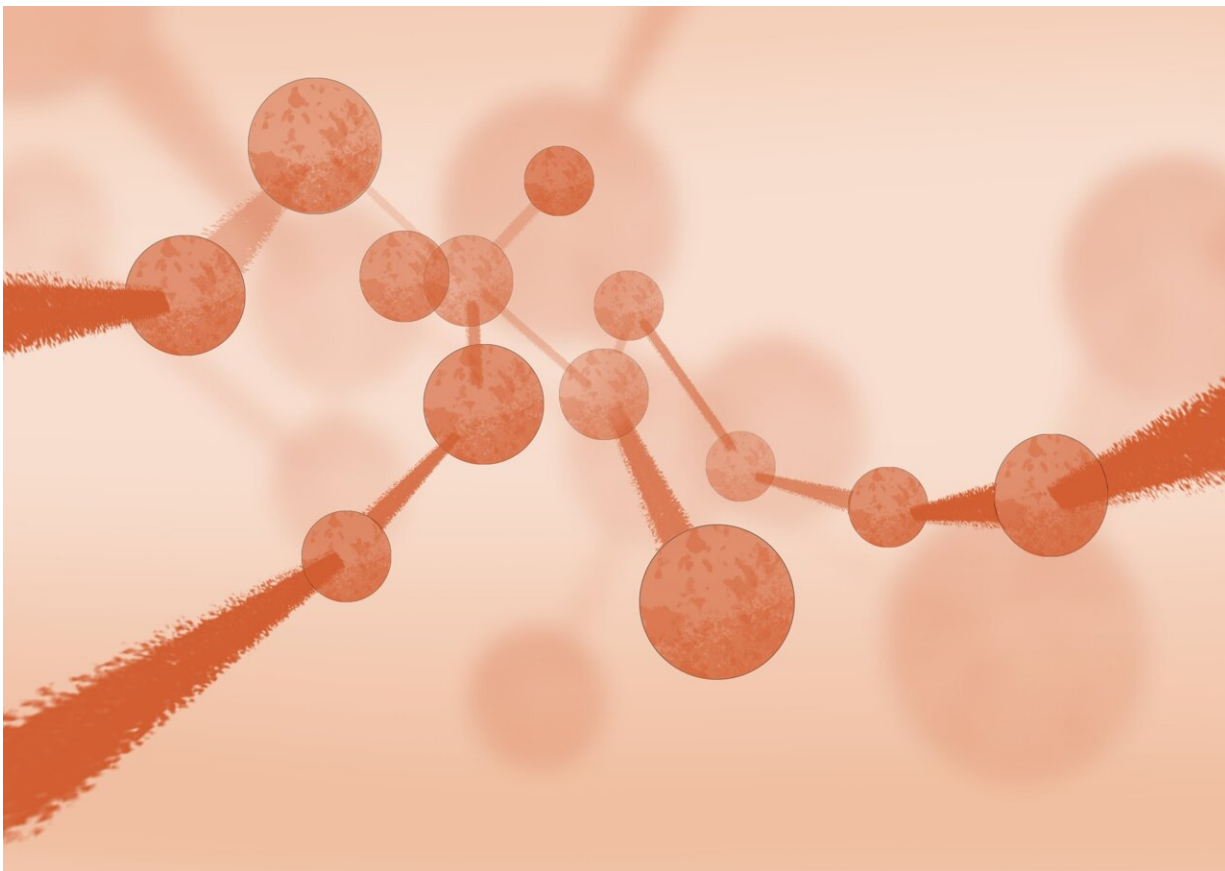


Scientists use machine learning to get an unprecedented view of small molecules

December 19 2022



Metabolites are extremely small – the diameter of a human hair is 100,000 nanometers, while that of a glucose molecule is approximately one nanometer. Credit: Matti Ahlgren/Aalto University

A new machine learning model will help scientists identify small

molecules, with applications in medicine, drug discovery and environmental chemistry. Developed by researchers at Aalto University and the University of Luxembourg, the model was trained with data from dozens of laboratories to become one of the most accurate tools for identifying small molecules.

Thousands of different small molecules, known as [metabolites](#), transport energy and transmit cellular information throughout the human body. Because they are so small, metabolites are difficult to distinguish from each other in a blood sample analysis—but identifying these molecules is important to understand how exercise, nutrition, [alcohol use](#) and metabolic disorders affect well-being.

Metabolites are normally identified by analyzing their mass and retention time with a separation technique called liquid chromatography followed by mass spectrometry. This technique first separates metabolites by running the sample through a column, which results in different flow rates—or retention times—through the measurement device.

Mass spectrometry is then used to fine-tune the identification process by sorting metabolites according to their mass. Researchers can also break metabolites into [smaller pieces](#) to analyze their composition using a technique called tandem [mass spectrometry](#).

"Even the best methods can't identify more than 40% of the molecules in samples without making some additional assumptions about the candidate molecules," says Professor Juho Rousu of Aalto University.

Now, Rousu's group has developed a novel machine learning model to identify small molecules. It was recently published in *Nature Machine Intelligence*.

"This new open-source model offers the whole research community an

enriched view of [small molecules](#). It will help research into methods to identify [metabolic disorders](#), such as diabetes, or even cancer," says Rousu.

The new approach elegantly sidesteps one of the challenges facing conventional methods. Because the retention times of molecules vary from lab to lab, data cannot be compared between labs. Eric Bach, a doctoral student at Aalto, came up with an alternative during his Ph.D. research that solved the problem.

"Our research shows that while absolute retention times may vary, the retention order is stable across measurements by different labs," Bach explains. "This allowed us to merge all publicly available data on metabolites for the first time ever and feed it into our machine learning model."

With the incorporation of data from dozens of laboratories around the globe, the machine learning model is accurate enough to distinguish between mirror image molecules, known as stereochemical variants. So far, identification tools have not been able to tell stereochemical variants apart, and the new capability is expected to open up new avenues in drug design and other fields.

"The fact that using stereochemistry improved the identification performance is a revelation for all developers of [metabolite](#) identification methods' says Emma Schymanski, associate professor at the Luxembourg Centre for Systems Biomedicine (LCSB) of the University of Luxembourg. "This method could also be used to help identify and trace micropollutants in the environment or characterize new metabolites in plant cells."

More information: Bach, E. et al, Joint structural annotation of small molecules using liquid chromatography retention order and tandem mass

spectrometry data, *Nature Machine Intelligence* (2022). DOI: [10.1038/s42256-022-00577-2](https://doi.org/10.1038/s42256-022-00577-2).
www.nature.com/articles/s42256-022-00577-2

Provided by Aalto University

Citation: Scientists use machine learning to get an unprecedented view of small molecules (2022, December 19) retrieved 21 June 2024 from <https://phys.org/news/2022-12-scientists-machine-unprecedented-view-small.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.