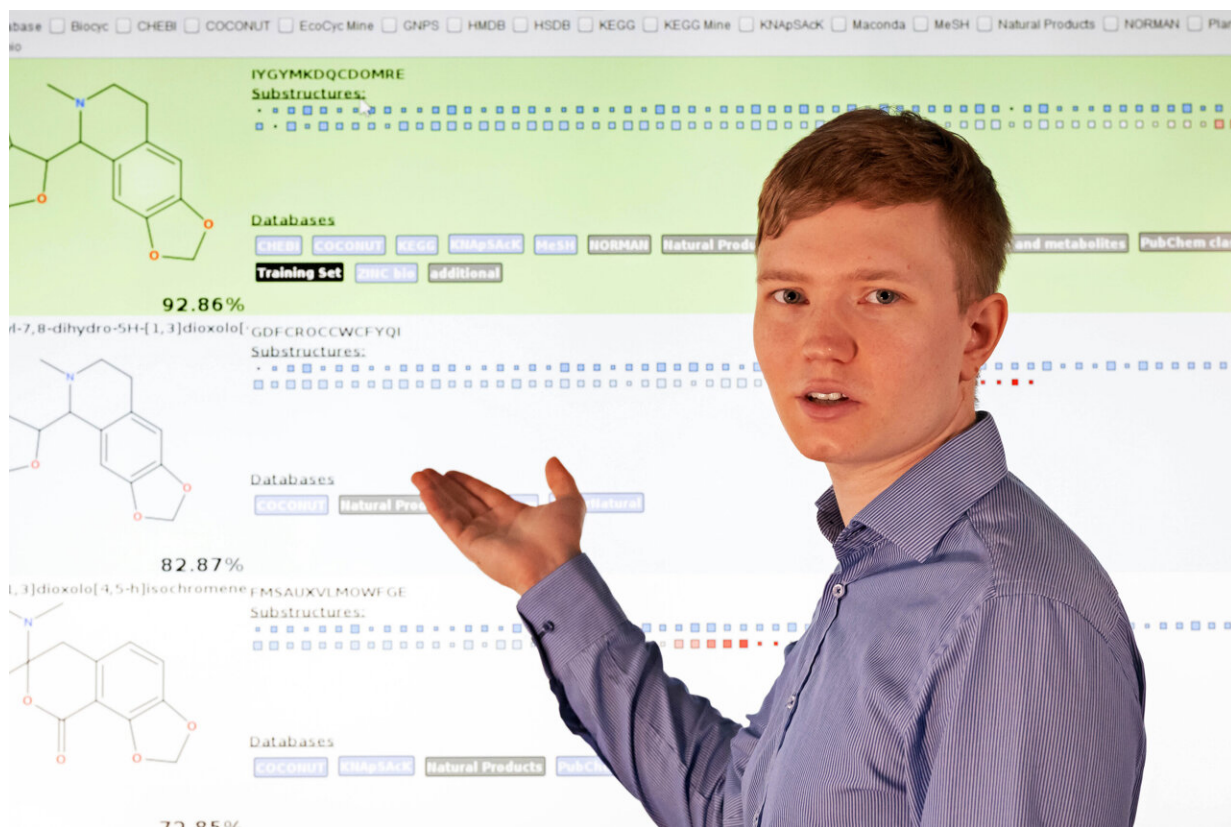


Artificial intelligence helps to find new natural substances

October 14 2021



Martin Hoffmann from the University of Jena presents COSMIC (Confidence Of Small Molecule IdentifiCations), a workflow that enables the identification of small molecules faster and easier than previously possible. Credit: Jens Meyer / University of Jena

More than a third of all medicines available today are based on active

substances from nature, and a research team from the University of Jena has developed a procedure to identify small active substance molecules much more quickly and easily. Secondary natural substances that occur in numerous plants, bacteria and fungi can be anti-inflammatory, can ward off pathogens or even prevent the growth of cancer cells. However, making use of the riches provided by nature's medicine cabinet and identifying new natural substances is time-consuming, costly and labor-intensive. A team of bioinformaticians at Friedrich Schiller University Jena has now developed a method that enables much faster and easier identification of small active substance molecules. The researchers present their method, called COSMIC (Confidence Of Small Molecule IdentifiCations), in the current issue of *Nature Biotechnology*.

Millions of structural data items not yet deciphered

To find out which substances are contained in a biological sample such as a plant extract, a researcher analyzes the sample using [mass spectrometry](#). In this process, the molecules are broken down into fragments and their mass is determined. "The CSI:FingerID molecule search engine we developed allows us to search specifically for [molecular structures](#) that match these fragments," says Prof. Sebastian Böcker of the University of Jena. "Whether this search is successful—i.e., whether the search result represents the correct [structure](#)—is not something we can distinguish in this way."

There are currently huge collections of data with billions of mass spectrometry data items from millions of analyzes of biological samples, the vast majority of which have not been identified as to their structure. This is where COSMIC is now coming into play, enabling structures to be deciphered automatically for a large proportion of these as yet unidentified molecules.

"To this end, we use machine-learning methods," explains Martin

Hoffmann, lead author of the new publication. "First, the mass spectrum of the sample under examination is compared with the available structural data." As a result, you get—as in a Google search—a more or less extensive list of possible hits. "Our method now indicates how confident one can be that the hit found in the first place is actually the structure we are looking for," Hoffmann adds. To do this, COSMIC determines a score that evaluates the quality of the suggested hit and deduces whether it is correct or incorrect.

New bile acids discovered

Böcker and his team have been able to demonstrate how well their method really works, in cooperation with colleagues from the University of California, San Diego. They studied mass spectrometry data from the digestive system of mice, searching for as yet unknown bile acids. For this purpose, more than 28,000 theoretically possible bile acid structures were constructed and compared with the measurement data from the microbiome of the mice. The subsequent analysis with COSMIC yielded a total of 11 new, previously completely unknown bile acid structures. Two of these have since been confirmed using specifically synthesized reference samples.

"This shows, firstly, that our method works reliably," emphasizes Sebastian Böcker. Secondly, COSMIC makes it possible to accelerate substantially the search for new and interesting substances, because the screening can be performed completely automatically, without any manual effort and in a very short time. Böcker expects that in the coming years, it will be possible to clarify thousands of new molecular structures in this way.

More information: Martin A. Hoffmann et al, High-confidence structural annotation of metabolites absent from spectral libraries, *Nature Biotechnology* (2021). [DOI: 10.1038/s41587-021-01045-9](https://doi.org/10.1038/s41587-021-01045-9)

Provided by Friedrich Schiller University of Jena

Citation: Artificial intelligence helps to find new natural substances (2021, October 14) retrieved 23 April 2024 from <https://phys.org/news/2021-10-artificial-intelligence-natural-substances.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.