

New machine learning method to analyze complex scientific data of proteins

September 21 2021, by Laura Arenschield



Credit: CC0 Public Domain

Scientists have developed a method using machine learning to better analyze data from a powerful scientific tool: Nuclear magnetic resonance (NMR). One way NMR data can be used is to understand proteins and chemical reactions in the human body. NMR is closely related to magnetic resonance imaging (MRI) for medical diagnosis.

NMR [spectrometers](#) allow scientists to characterize the structure of

molecules, such as proteins, but it can take highly skilled human experts a significant amount of time to analyze that data. This new machine learning method can analyze the data much more quickly and just as accurately.

In a study recently published in *Nature Communications*, the scientists described their process, which essentially teaches computers to untangle complex data about atomic-scale properties of proteins, parsing them into individual, readable images.

"To be able to use these data, we need to separate them into features from different parts of the molecule and quantify their specific properties," said Rafael Brüschweiler, senior author of the study, Ohio Research Scholar and a professor of chemistry and biochemistry at The Ohio State University. "And before this, it was very difficult to use computers to identify these individual features when they overlapped."

The process, developed by Dawei Li, lead author of the study and a research scientist at Ohio State's Campus Chemical Instrument Center, teaches computers to scan images from NMR spectrometers. Those images, known as spectra, appear as hundreds and thousands of peaks and valleys, which, for example, can show changes to proteins or complex metabolite mixtures in a biological sample, such as blood or urine, at the atomic level. The NMR data give important information about a protein's function and important clues about what is happening in a person's body.

But deconstructing the spectra into readable peaks can be difficult because often, the peaks overlap. The effect is almost like a mountain range, where closer, larger peaks obscure smaller ones that may also carry important information.

Think of the QR code readers on your phone: NMR spectra are like a

QR code of a molecule—every protein has its own specific 'QR code,'" Brüscheiler said. "However, the individual pixels of these 'QR codes' can overlap with each other to a significant degree. Your phone would not be able to decipher them. And that is the problem we have had with NMR spectroscopy and that we were able to solve by teaching a computer to accurately read these spectra."

The process involves creating an artificial deep neural network, a multi-layered network of nodes that the computer uses to separate and analyze data.

The researchers created that network, then taught it to analyze NMR spectra by feeding spectra that had already been analyzed by a person into the computer and telling the computer the previously known correct result. The process of teaching a computer to analyze spectra is almost like teaching a child to read—the researchers started with very simple spectra. Once the computer understood that, the researchers moved on to more complex sets. Eventually, they fed highly complex spectra of different proteins and from a mouse urine sample into the computer.

The computer, using the deep neural network that had been taught to analyze spectra, was able to parse out the [peaks](#) in the highly complex sample with the same accuracy as a human expert, the researchers found. And more, the computer did it faster and highly reproducibly.

Using machine learning as a tool to analyze NMR [spectra](#) is just one key step in the lengthy scientific process of NMR data interpretation, Brüscheiler said. But this research enhances the capabilities of NMR spectroscopists, including the users of Ohio State's new National Gateway Ultrahigh Field NMR Center, a \$17.5 million center funded by the National Science Foundation. The center is expected be commissioned in 2022 and will have the first 1.2 gigahertz NMR spectrometer in North America.

Other research scientists involved in this study include Alexandar Hansen, Chunhua Yuan and Lei Bruschweiler-Li, all of Ohio State's Campus Chemical Instrument Center.

More information: Da-Wei Li et al, DEEP picker is a deep neural network for accurate deconvolution of complex two-dimensional NMR spectra, *Nature Communications* (2021). [DOI: 10.1038/s41467-021-25496-5](https://doi.org/10.1038/s41467-021-25496-5)

Provided by The Ohio State University

Citation: New machine learning method to analyze complex scientific data of proteins (2021, September 21) retrieved 13 March 2024 from <https://phys.org/news/2021-09-machine-method-complex-scientific-proteins.html>

<p>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.</p>
--