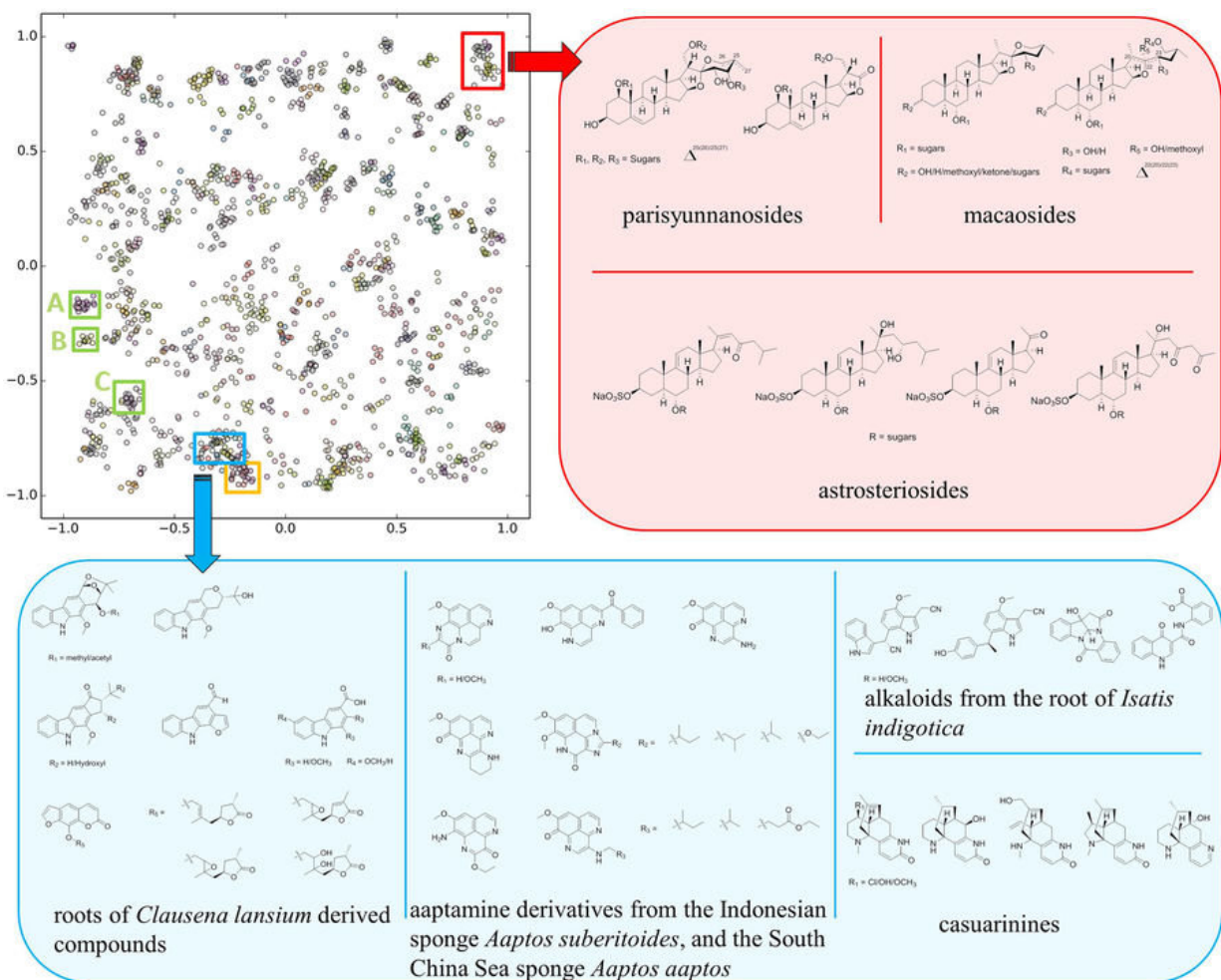


SMART: Facial recognition for molecular structures

October 31 2017, by Katherine Connor



The SMART cluster map based on training result of 2,054 HSQC spectra over 83,000 iterations, with inset boxes representing different compound classes discussed in the text. Credit: University of California - San Diego

An interdisciplinary team of researchers at the University of California San Diego has developed a method to identify the molecular structures of natural products that is significantly faster and more accurate than existing methods. The method works like facial recognition for molecular structures: It uses a piece of spectral data unique to each molecule and then runs it through a deep learning neural network to place the unknown molecule in a cluster of molecules with similar structures.

The new system is called "SMART," which stands for Small Molecule Accurate Recognition Technology, and has the potential to accelerate the molecular [structure](#) identification process ten-fold. This development could represent a paradigm shift in the chemical analysis, pharmaceutical and drug discovery fields since 70 percent of all Food and Drug Administration (FDA)-approved drugs are based on [natural products](#) such as soil microorganisms, terrestrial plants and, increasingly, marine life forms such as algae.

"The structure of a molecule is the enabling information," said Bill Gerwick, professor of oceanography and pharmaceutical sciences at UC San Diego's Scripps Institution of Oceanography. "You have to have the structure for any FDA approval. If you want to have intellectual property, you have to patent that structure. If you want to make analogs of that molecule, you need to know what the starting molecule is. It's a critical piece of information."

Chen Zhang, a nanoengineering Ph.D. student at UC San Diego collaborating with Gerwick and the first author of the paper published in *Nature Scientific Reports*, said that determining a molecule's structure can be a bottleneck in the natural product research process, taking experts months and even years to accurately determine the correct and complete structure. While each molecule and its identification timeline is different, the SMART approach gives researchers an early clue into what

family a new molecule falls under, drastically reducing the time it takes to characterize a new natural product.

"The way we were able to accelerate the process is by essentially using [facial recognition](#) software to look at the key piece of information we obtain on the molecules," Gerwick said. The key piece of information the team uses is called a heteronuclear singular quantum coherence nuclear magnetic resonance, or HSQC NMR, spectrum. It produces a topological map of spots that reveal which protons in the molecule are attached directly to which carbon atoms, an arrangement unique to every molecule.

Zhang and Gerwick teamed up with Gary Cottrell, a computer science and engineering professor at the UC San Diego Jacobs School of Engineering, to develop a [deep learning](#) system trained with thousands of HSQC spectra pulled from previous research. This convolutional neural network takes a 2-D image of the HSQC NMR spectrum of an unknown molecule and maps it into a 10-dimensional space clustered near similar [molecules](#), making it easier for researchers to elucidate an unknown molecule's structure.

"Chen took this approach to getting NMR spectra of over 4,000 compounds from the literature by literally cutting out the images from the PDFs of the papers," Cottrell said. "It was an awesome effort! Even so, this is normally not enough data to train a deep network, but we used a technology called a Siamese network, in which you train on pairs of images. This amplifies your training set by roughly the square of the number of compounds in a family, and is what made this project feasible."

This collaboration is the first time Gerwick has mentored an engineering student, and the exchange of ideas proved fruitful.

"It's been a wonderful interaction. UC San Diego has something really quite magical about it, and that is the depth of collaboration that occurs between departments—it's phenomenal," Gerwick said. "When you try and thoughtfully take from another discipline something that is maybe even commonplace in that discipline and apply it in a new and unique way in our discipline, it's an opportunity to really have this kind of paradigm-shifting thing. And I think this technology, with some advancement, could be a real [paradigm shift](#) in the way we do all kinds of chemistry and chemical analysis."

More information: Chen Zhang et al. Small Molecule Accurate Recognition Technology (SMART) to Enhance Natural Products Research, *Scientific Reports* (2017). [DOI: 10.1038/s41598-017-13923-x](https://doi.org/10.1038/s41598-017-13923-x)

Provided by University of California - San Diego

Citation: SMART: Facial recognition for molecular structures (2017, October 31) retrieved 18 April 2024 from <https://phys.org/news/2017-10-smart-facial-recognition-molecular.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.