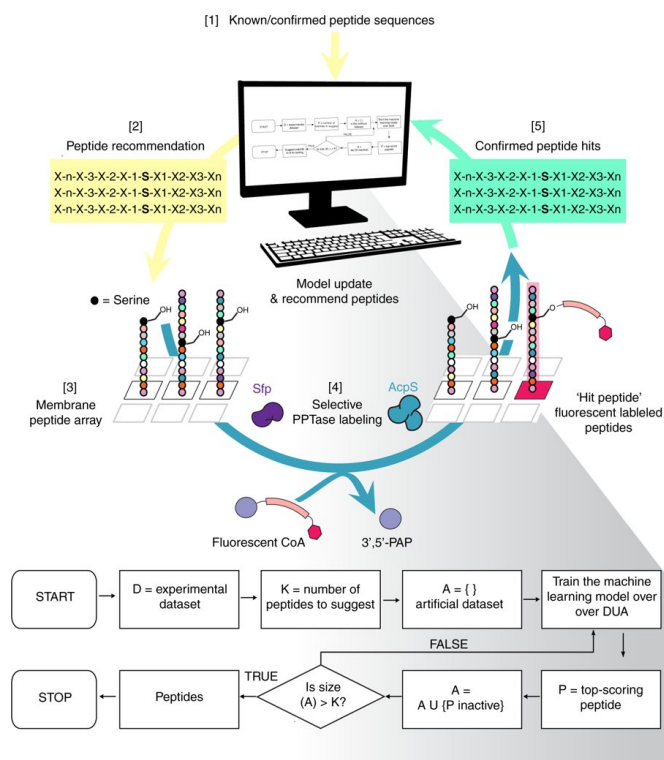


# Using machine learning to design peptides

10 December 2018



University of California, San Diego, have developed a new way of finding optimal peptide sequences: using a machine-learning algorithm as a collaborator.

The algorithm analyzes [experimental data](#) and offers suggestions on the next best sequence to try, creating a back-and-forth [selection process](#) that drastically reduces the time needed to find the optimal peptide.

The results, which could provide a new framework for experiments across materials science and chemistry, were published in *Nature Communications* on December 7.

"We view this as the next wave in how we design molecules and materials," said Northwestern professor Nathan Gianneschi, a corresponding author on the paper. "We can combine what we know from intuition with the power of an algorithm and find the solution with fewer experiments."

Overview of the iterative Peptide Optimization with Optimal Learning (POOL) method workflow. Credit: *Nature Communications* (2018). DOI: 10.1038/s41467-018-07717-6

Gianneschi is the Jacob and Rosaline Cohn Professor in the department of chemistry in Northwestern's Weinberg College of Arts and Sciences and in the departments of [materials science](#) and engineering and of biomedical engineering at Northwestern Engineering.

Scientists and engineers have long been interested in synthesizing peptides—chains of amino acids responsible for conducting many functions within cells—to both mimic nature and to perform new activities. A designed peptide, for example, could be a functional drug acting in certain areas in the body without degrading, a difficult task for many peptides.

But methods for discovering and synthesizing [peptides](#) are expensive and time-consuming, often involving months or years of guesswork and failure.

To create the method, Gianneschi, who is also the associate director of Northwestern's International Institute for Nanotechnology, teamed up with Peter Frazier, an associate professor at Cornell who works in operations research and machine learning, and Michael Burkart, a chemical biologist and expert in enzymology at UC San Diego, to find a better way to make peptides that could generate biomaterials—specifically nanostructures and microstructures that could modify proteins in certain ways. The first step was to find the right peptides that would act as enzymatic substrates for these structures.

Northwestern University researchers, teaming up with collaborators at Cornell University and the

Peptides are built from chains of amino acids that can be as many as 20 amino acids long, with 20

different possibilities for each acid. Since the sequence determines the peptide function, figuring out optimal sequences requires expensive experiments often conducted with guesswork.

The experimentalists, Gianneschi and Burkart, worked with Frazier over several years to develop a system that combined experimental data with a machine-learning algorithm to find the best strategies for creating new materials.

After Frazier designed the algorithm and the two worked together to train it, the experimentalists developed an array of 100 peptides, conducted experiments to figure out which ones worked as they were meant to, then fed that information into the algorithm. The algorithm then recommended what to change for the next round of peptide development, and also recommended strategies that it thought would fail.

"Now we were starting to get selectivity," Gianneschi said. By completing this process several times, they were able to home in on optimal peptides.

"Instead of guessing and looking at millions of peptides, we were able to look at hundreds of peptides and very quickly converge on sequences that behaved in completely new ways," he said. When compared against random mutations or guesswork, the [algorithm](#) method was statistically far more successful.

Though this work focused on substrates, this process could be used to discover peptides for any kind of purpose, like drug delivery, and perhaps even be used to discover DNA sequences, as well. Because any sort of optimal sequence could be discovered, researchers are also not limited to what amino acids sequences are found in the genetic code.

The next step will be automating the entire process. Gianneschi is also interested in using the method to find optimal surfaces for polymers, specifically polymers used in medical implants. Finding the right surfaces that will bind with tissue or muscle could help prevent scar tissue or implant rejection.

"You could essentially discover sequences that do specific things, which is really at the core of what peptides and nucleic acids do in nature," he said. "This could revolutionize how we make peptides."

**More information:** Lorillee Tallorin et al, Discovering de novo peptide substrates for enzymes using machine learning, *Nature Communications* (2018). [DOI: 10.1038/s41467-018-07717-6](https://doi.org/10.1038/s41467-018-07717-6)

Provided by Northwestern University

APA citation: Using machine learning to design peptides (2018, December 10) retrieved 21 September 2019 from <https://phys.org/news/2018-12-machine-peptides.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.*