

A new approach to reconstructing protein evolution

October 18 2019

There are an estimated 20,000 to 30,000 proteins at work in cells, where they carry out numerable functions, says computational molecular biologist Roman Sloutsky at the University of Massachusetts Amherst. "One of the central questions in all of biochemistry and molecular biology," he adds, is how their precisely-tuned functions are determined.

Most proteins belong to families related to each other in the same way species are, by having descended from a common ancestor, Sloutsky says. One angle scientists have taken to explore how their functions arise is to trace [protein](#) family evolution and relatedness, he notes, but reconstructing the twists and turns of past genetic divergence is very difficult.

In a paper just released in *eLife*, Sloutsky and his former advisor, associate professor Kristen Naegle, now at the University of Virginia, propose an unusual, new and more accurate way to trace how proteins diverged over time. "It can yield powerful insights into the relationship between [protein sequence](#), structure and function for that family," he says. The paper represents part of his doctoral work with Naegle at Washington University in St. Louis.

As Sloutsky explains, "Many protein scientists, including us, rely on reconstructions of the evolution of their proteins of interest in designing their research strategies. When those reconstructions are incorrect, we risk incorrectly interpreting the results of our experiments. Although, as we show, perfectly accurate reconstructions are often impossible,

understanding the limits of reconstruction accuracy allows us to design studies that are robust to that uncertainty and help us to avoid misinterpreting the results."

Now in Margaret Stratton's lab at UMass Amherst, she and Sloutsky are particularly interested in a family of cell-signaling proteins called Calcium/calmodulin-dependent kinase II (CaMKII), key players in neuropsychiatric disorders, some cancers, cardiac arrhythmias and fertility disorders caused by dysregulated cell signaling, he says. "If we understand the rules governing functional specificity well enough, we can apply that to design highly specific treatments to maximize efficacy and minimize side effects."

"A fundamental problem in evolution in general, though, is that the past truth is unknowable," Sloutsky adds. "So when you're trying to develop a method for reconstructing [evolutionary history](#), you have to simulate the evolutionary process."

To do this, the computational researchers build a known-sequence ancestor protein, then simulate a series of amino acid substitutions to arrive at a collection of realistic protein sequences for which they know the whole evolutionary history. "That gives us a data set in which we can test the reconstruction accuracy, because we know what happened at every stage," he notes. "It tells us how well the method is working, so when we go to work on proteins in a real family, we have an idea of how accurate the reconstruction is in general."

"We are introducing for the first time a method for assessing how accurate a reconstruction will be, based on observables—the relatedness among a group of proteins. Doing this over and over gives us a range within which the true history might be expected to occur."

Naegle uses a travel analogy to show how their new approach puts

bounds on a set of solutions: If one asks 1,000 people to predict what route a driver took on a past multi-stage trip, pieces shared most across all 1,000 answers are most likely to be true. One can never know the real path—in this case, evolution—but if no predictions agree over 1,000 answers, one has little confidence in their accuracy. But if all 1,000 agree on some stages, confidence is higher that a new model based on it will be true. It still may not be wholly accurate, but it's probably closer than any single route suggestion, she says.

Sloutsky adds that instead of reconstructing all possible protein divergence, their method focuses on only part of the history to generate many different versions. This yields an ensemble of possible solutions. "We reconstruct less than other methods but the part we do reconstruct we do more accurately," he notes.

He acknowledges, "It's not a traditional method. Our approach takes advantage of what we already knew about how protein families evolve, but in a new way. For many years this fact was known but no one took advantage of it. We focus on it. By reconstructing multiple plausible models, depending on how accurate the prediction is expected to be, our method can deliver one or a collection of candidates with the promise that they are all as accurate as can be. We show in the paper that they are all more accurate than any existing method."

More information: ASPEN, a methodology for reconstructing protein evolution with improved accuracy using ensemble models, eLife (2019). [DOI: 10.7554/eLife.47676.001](https://doi.org/10.7554/eLife.47676.001)

Provided by University of Massachusetts Amherst

Citation: A new approach to reconstructing protein evolution (2019, October 18) retrieved 26

April 2024 from <https://phys.org/news/2019-10-approach-reconstructing-protein-evolution.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.