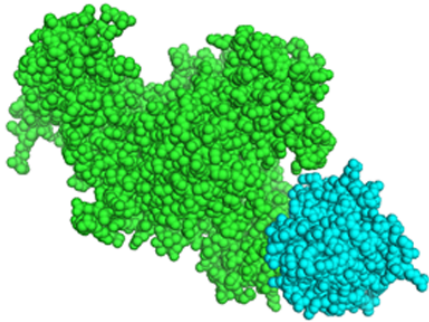


Algorithm for predicting protein pairings could help show how living systems work

21 September 2016



Interacting proteins. Credit: Lucy Colwell

An algorithm which models how proteins inside cells interact with each other will enhance the study of biology, and sheds light on how proteins work together to complete tasks such as turning food into energy.

Researchers have developed an [algorithm](#) that aids our understanding of how living systems work, by identifying which proteins within cells will interact with each other, based on their genetic sequences alone.

The ability to generate huge amounts of data from genetic sequencing has developed rapidly in the past decade, but the trouble for researchers is in being able to apply that sequence data to better understand living systems. The new research, published in the journal *Proceedings of the National Academy of Sciences*, is a significant step forward because biological processes, such as how our bodies turn food into energy, are driven by specific protein-protein interactions.

"We were really surprised that our algorithm was powerful enough to make accurate predictions in the absence of experimentally-derived data," said study co-author Dr Lucy Colwell, from the University of Cambridge's Department of

Chemistry, who led the study with Ned Wingreen of Princeton University. "Being able to predict these interactions will help us understand how proteins fit and work together to complete required tasks – and using an algorithm is much faster and much cheaper than relying on experiments."

When proteins interact with each other, they stick together to form protein complexes. In her previous research, Colwell found that if the two interacting proteins were known, [sequence data](#) could be used to figure out the structure of these complexes. Once the structure of the complexes is known, researchers can then investigate what is happening chemically. However, the question of which proteins interact with each other still required expensive, time-consuming experiments. Each cell often contains multiple versions of the same protein, and it wasn't possible to predict which version of each protein would interact specifically – instead, experiments involve trying all options to see which ones stick.

In the current paper, the researchers used a [mathematical algorithm](#) to sift through the possible interaction partners and identify pairs of proteins that interact with each other. The method correctly predicted 93% of protein-protein interactions present in a dataset of more than 40,000 protein sequences for which the pairing is known, without being first provided any examples of correct pairs.

When two proteins stick together, some amino acids on one chain stick to the amino acids on the other chain. The boundaries between interacting proteins tend to evolve together over time, causing their sequences to mirror each other.

The algorithm uses this effect to build a model of the interaction. It first randomly pairs protein versions within each organism – because interacting pairs tend to be more similar in sequence to one another than non-interacting pairs, the algorithm can quickly identify a small set of

largely correct pairings from the random starting point.

Using this small set, the algorithm measures whether the amino acid at a particular location in the first protein influences which amino acid occurs at a particular location in the second protein. These dependencies, learned from the data, are incorporated into a model and used to calculate the interaction strengths for each possible protein pair. Low-scoring pairings are eliminated, and the remaining set used to build an updated model.

The researchers thought that the algorithm would only work accurately if it first 'learned' what makes a good protein-protein pair by studying pairs that have been discovered in experiments. This meant that the researchers had to give the algorithm some known protein pairs, or 'gold standards,' against which to compare new sequences. The team used two well-studied families of proteins, histidine kinases and response regulators, which interact as part of a signaling system in bacteria.

But known examples are often scarce, and there are tens of millions of undiscovered protein-protein interactions in cells. So the team decided to see if they could reduce the amount of training they gave the algorithm. They gradually lowered the number of known histidine kinase-response regulator pairs that they fed into the algorithm, and were surprised to find that the algorithm continued to work. Finally, they ran the algorithm without giving it any such training pairs, and it still predicted new pairs with 93 percent accuracy.

"The fact that we didn't need a set of training data was really surprising," said Colwell.

The algorithm was developed using proteins from bacteria, and the researchers are now extending the technique to other organisms. "Reactions in living organisms are driven by specific [protein interactions](#)," said Colwell. "This approach allows us to identify and probe these interactions, an essential step towards building a picture of how living systems work."

More information: Anne-Florence Bitbol et al. 'Inferring interaction partners from protein

sequences.' *Proceedings of the National Academy of Sciences* (2016). [DOI: 10.1073/pnas.1606762113](https://doi.org/10.1073/pnas.1606762113)

Provided by University of Cambridge

APA citation: Algorithm for predicting protein pairings could help show how living systems work (2016, September 21) retrieved 17 May 2021 from <https://phys.org/news/2016-09-algorithm-protein-pairings.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.