

# New technique is quantum leap forward in understanding proteins

December 23 2008

---

In this ongoing quest, a group of Scripps Research Institute scientists, along with colleagues from the University of California, San Diego, (UCSD) have borrowed from physics to deliver one of those research rarities -- an unmitigated success. The group has devised a computational method that, with remarkable accuracy, predicts how bacterial proteins fold and interact.

In the short term, this new capability, described in this week's Early Edition of the *Proceedings of the National Academy of Sciences*, should allow development of new antibiotics that target and block newly identified protein interactions vital to the survival of pathogenic bacteria. In the longer term, as the collective body of genomics data for humans and other animals grows, some version of the new technique may allow similar protein predictive capabilities for higher organisms, spawning a wealth of new and highly effective drug discovery options.

"I think it's a quantum leap," says team leader James Hoch, a professor at The Scripps Research Institute, of the work, "This is one thing I really am proud of."

Ever since genomic data has been available, researchers have been looking for ways to understand protein interactions, but no method has proven even close to sufficient. "It's really the last frontier in proteins," says Hoch, "figuring out who they interact with and the structures they make."

One way to study proteins is to actually image their interactions using x-ray crystallography. This has provided invaluable, but very limited, information, because the method is fraught with drawbacks including extreme labor intensity and great difficulty in actually capturing the intended protein interactions. Assistant Professor Hendrik Szurmant, another leader of the project from Scripps Research, says the process is so difficult with x-ray crystallography that it only rarely works for transient interactions.

Another available means for studying protein interactions is a statistical method known as covariance, and the Scripps Research-UCSD team's new method relies on this as a foundation. Covariance analysis of proteins involves studying the amino acids found at specific locations on various protein sequences culled from genomics data. Covariance analysis between two proteins identifies residue positions that vary together from residue positions that vary at random.

Covariance has proven quite effective at identifying critical residues that bind directly with other proteins or other spots on the same protein, which is the goal. But, unfortunately, the method also identifies a high percentage of residues that turn out to not be involved in these direct interactions. Research groups have developed various techniques to winnow out such indirect interactions, but with only limited success—until now.

## **Outside Help**

Years ago, frustrated by the inadequacy of available techniques, Hoch and his colleagues set out to find some means beyond the normal bounds of biology to solve the problem of identifying the directly interacting protein residues without crystallography. The search eventually brought them to Professor Terry Hwa at UCSD and Martin Weigt, an expert in a computational technique known as message passing, in Turin, Italy. This

method, used mainly in an area known as spin glass physics, is a computer-intensive means of finding patterns in certain types of data.

For the first test of message passing with proteins, the group focused on the proteins involved in the well-studied two-component signaling system, which is responsible for a range of critical functions in bacteria. The first step of the work was to analyze the countless proteins involved in this system applying standard covariance techniques to available genomics data. The full analysis included about 2,500 different protein pairings and considered the potential interactions between about 100 residues on each protein in a pair.

To visualize this computational challenge, think of a grid that is 100 residues tall, for the first protein, and 100 residues wide, from the second protein. The resulting 10,000 boxes in this grid represent all of the potential residue interactions, and the overall analysis forms a cube 2,500 blocks deep because there is a similar grid for each of the 2,500 protein pairings. Covariance can rank each of these 25 million blocks to identify the target residues that interact directly, along with those numerous indirect pairings that need to be winnowed away.

The innovative next step was for the UCSD group to feed this covariance data into a message-passing program. Over the course of about a week of computing, the program analyzed this seemingly unfathomable mass of information and in time identified patterns in the highest-ranking cubes. Continued analysis ultimately yielded predictions about which pairings were in fact direct interactions.

Because the two-component signaling system has been the focus of intense research efforts at Scripps Research and elsewhere, including extensive x-ray crystallography, many of the direct residue interactions had already been identified. That meant all-or-nothing results for the very first message-passing experiment. Either the technique would

accurately identify the direct pairings or not.

The results came back overwhelmingly positive, and it was the culmination of a very long quest for Hoch. "It felt absolutely great," he says, "I thought, 'We finally got it! We got it and it works!'"

With a given protein binding site, on average, the message passing identified ten direct interactions accurately before giving a single false positive. Given that researchers can identify the active binding site for proteins by knowing as few as three directly interaction residues, this success rate is more than enough, for instance, to identify a new drug target. In the case of proteins that interact with themselves, there were 23 correct pairings identified before a first false positive.

"Based on test models so far, it appears that the method is absolutely, astonishingly accurate," says Szurmant.

The next step, already under way, is to use the message passing in the drug discovery process. The two-component signaling system is responsible for countless essential functions in bacterial cells including adjustment to growth conditions, and can control virulence. That means interruption of strategic direct interactions can kill pathogenic bacteria. Though many direct interactions in the system had been identified, the message-passing work has also identified new ones.

The message-passing technique is dependant on the availability of extensive genomic data, and some 800 or so bacterial genomes have been fully sequenced. But applying message passing to animals will have to wait until a similar volume of genomic data is available for them. Ultimately, some form of the technique could identify important protein interactions in humans, which would open a wealth of new drug targeting possibilities.

Paper: "Identification of direct residue contacts in protein-protein interaction by message passing", PNAS.

Source: Scripps Research Institute

Citation: New technique is quantum leap forward in understanding proteins (2008, December 23) retrieved 25 April 2024 from <https://phys.org/news/2008-12-technique-quantum-proteins.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.