

# Untangling life's origins

March 11 2013

---

Researchers in the Evolutionary Bioinformatics Laboratory at the University of Illinois in collaboration with German scientists have been using bioinformatics techniques to probe the world of proteins for answers to questions about the origins of life.

Proteins are formed from chains of [amino acids](#) and fold into [three-dimensional structures](#) that determine their function. According to [crop sciences](#) professor Gustavo Caetano-Anollés, very little is known about the evolutionary drivers for this folding.

In collaboration with scientists at the Heidelberg Institute for [Theoretical Studies](#), he has been working at the interface of molecular evolution and [molecular dynamics](#), looking back to when proteins first appeared approximately 3.8 billion years ago to determine changes in folding speed over time.

To do this, they looked at all known protein structures as defined in the Structural Classification of Proteins (SCOP) database and mined their presence in 989 fully sequenced genomes. In a previous study, researchers in Caetano-Anollés's group used SCOP and genomic information to reconstruct phylogenomic trees that describe the history of the protein world. The current research is based on these types of trees.

"They are not the standard trees that people see in phylogenetic analysis," he said. "In phylogenetic analysis, usually the tips of the trees, the leaves, are organisms or microbes. In these, they are entire [biological](#)

[systems.](#)"

In contrast, the leaves of these new trees are [protein domains](#), which are compact evolutionary units of structure and function. Proteins are usually complex combinations of several domains.

"We have a world of about 90,000 of these structures, but they seem to be always producing the same designs," he said. Over the last 10 years, he has been part of the effort to map these designs, or folds, because they are determined by the way the protein chains fold on themselves. To date, approximately 1,300 folds have been characterized.

For the current study, the researchers identified protein sequences in the genomes that had the same folding structure as known proteins. They then used bioinformatics techniques to compare them to each other on a time scale to determine when proteins became part of a particular organism. This allowed them to map protein structures and organisms onto a timeline.

Directly calculating the folding speed for all of these proteins would be impossible with today's technology, so the researchers took advantage of the fact that a protein always folds at the same points and used a measure called Size Modified Contact Order (SMCO).

Contact order is the ability of a protein to establish links between segments of the polypeptide chain. When points that are close together on the chain come together, they generally form helical structures; when distant points come together, they form beta strands that interact with each other and form sheets. Contact order measures how many of the connections are local and how many are distant. Experimental studies have shown that it is correlated with folding speed. The measure is normalized (size modified) to take protein length, which affects folding speed, into account.

They saw a peculiar pattern in the results.

"What we see is an hourglass," said Caetano-Anollés. "At the beginning, proteins seem not to be folding so fast. And then, as time progresses, there's a tendency to fold faster and faster. And then it reaches a critical point, and at this point we have a tendency that reverses, that seems to go back again to slow folding." However, the tendency toward higher speed dominates.

This point coincides with what he calls the "Big Bang" in protein evolution. Approximately 1.5 billion years ago, more complex domain structures and multi-domain proteins emerged with the appearance of multicellular organisms. Amino acid chains, which make up proteins, also became shorter at this point in time.

Why does folding speed matter?

"If the protein does not fold, in the vast majority of cases it will not have a function. So folding implies functionality. And speed of folding implies speed of achieving that functionality," he explained. "For a cell, that's very important, because if proteins are very slow folders, there is a time lag to when that function will be accessible to the cell."

Fast folders are also less susceptible to aggregation, or clumping together, so they work faster. Moreover, proteins that fold rapidly are more likely to fold correctly. [Protein](#) misfolding has been linked with diseases such as Alzheimer's.

Caetano-Anollés said, however, that this research makes an important contribution to understanding how molecules work. "The complexities of the biological functions of molecules are still poorly understood," he said.

"If we mix the world of molecular dynamics with the world of [molecular evolution](#), we can then determine what aspects of sequences are important for molecular dynamics, and therefore, we can apply them to genetic engineering, synthetic biology, and so on."

Provided by University of Illinois at Urbana-Champaign

Citation: Untangling life's origins (2013, March 11) retrieved 23 June 2024 from <https://phys.org/news/2013-03-untangling-life.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.