

October 1, 2007

*Science Daily* — The present can tell you a lot about the past, but you need to know where to look. A new study appearing this month in *Genome Research* reveals that protein architectures – the three-dimensional structures of specific regions within proteins – provide an extraordinary window on the history of life.

In the study, researchers at the University of Illinois describe contemporary protein architectures as “molecular fossils” or “historical imprints” that mark important milestones in evolutionary history. The research team compiled a global census of protein architectures, and used these relics to plot the emergence, diversification and refinement of each of the three superkingdoms of life: Archaea, Bacteria and Eukarya.

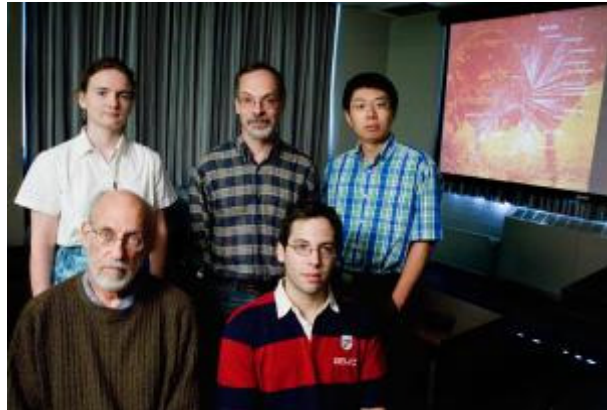
All proteins are composed of architectural elements, called domains, which can be identified by their structural and functional similarities to one another. Protein domains are the gears, belts, springs and motors that allow the larger protein machinery to function as it should. Every protein contains one or more of them, and proteins that perform very different tasks can contain identical domains.

Protein domains are grouped into what are called fold families and fold superfamilies. Members of a fold superfamily may differ in their underlying amino acid sequences, but retain structural and functional similarities and are evolutionarily related. Fold superfamilies are grouped together into broad categories, called folds. The new study tracks the evolution of folds and fold superfamilies from the ancient world to the present.

Protein folds turn out to be reliable markers of evolutionary events because they are quite stable over time, said Gustavo Caetano-Anollés, a professor of crop sciences and a principal investigator on the study. Even mutations in the genes that code for them rarely change their three-dimensional structures. “Structures are highly conserved because they were important discoveries in the history of the world,” Caetano-Anollés said. “It’s very difficult to come up with a new design to do something in a way that an existing structure cannot already do.”

The idea that protein folds are highly refined and profoundly flexible machines is supported by the fact that there are so few of them. Scientists have identified only about 1,000 folds and 1,500 fold superfamilies across all the organisms for which full genomes have been sequenced. Many of these protein folds are found in every organism. Other folds appear only in certain subsets of organismal life.

The Illinois team’s findings add a new dimension to a long and contentious debate about the earliest



*Crop sciences professor Gustavo Caetano-Anollés, back row center, describes contemporary protein architectures as “molecular fossils” or “historical imprints” that mark important milestones in evolutionary history. His research team compiled a global census of protein architectures, and used these relics to plot the emergence, diversification and refinement of each of the three superkingdoms of life: Archaea, Bacteria and Eukarya. Members of the team: front row, left to right, Jay Mittenthal, cell and developmental biology, and undergraduate Derek Caetano-Anollés; back row, Liudmila Yafremava and Minglei Wang, postdoctoral students in crop sciences.*

stages of evolutionary divergence. By looking at protein architectures across all organisms for which genomic information is available, the team found evidence that the archaeal microbes, the one-celled organisms that inhabit some of the most forbidding environments on the planet, were the first to emerge as an evolutionarily distinguishable group. Their evidence: The repertoire of architectures that would one day belong to the superkingdom known as the Archaea was the first to lose a fold. That fold, a huge class of protein fold superfamilies, simply disappeared from the archaeal lineage altogether.

Eventually, more and more folds joined the list of architectures abandoned by the Archaea, in what the authors describe as a process of "reductive evolution." The folds belonging to organisms that eventually evolved into what we now call bacteria and the multicellular eukaryotes also began to lose folds, but they started downsizing their repertoires much later than the Archaea.

Prior to this, the authors write, the world of protein folds was large and diverse, containing many of the fold architectures still in use today. This was the time of the "communal ancestor," before the emergence of superkingdoms and the myriad organisms that would eventually populate each group. This overview of protein architectures adds to the picture of how the superkingdoms emerged and diverged. The Archaea jettisoned many of the folds that had been part of their original heritage.

As a group, the bacteria lost fewer folds, although those that were parasites or obligate parasites retained only a minimalist repertoire of folds. Their strategy was to take advantage of the protein machinery available in their hosts. The multicellular Eukarya, a group that includes humans, retained the largest repertoire of fold architectures. "We are the keepers of everything. We have the largest repertoire that there is," Caetano-Anollés said. The eukaryotes' evolution into large, multicellular bodies that could live in diverse environments relied on an extensive library of protein architectures, he said.

The new study divides the evolution of protein architectures into three phases. First, there was a common world, with a large collection of protein folds available to all. The researchers call this a period of architectural diversification. Next came a period, called superkingdom specification, during which the three superkingdoms emerged. The third phase, organismal diversification, saw an explosion of inventiveness in protein architectures, particularly among the Eukarya.

Caetano-Anollés stressed that any attempt to build an evolutionary tree of life is limited by the type of data used to populate the tree. He compared the task to that of writing a history of building architecture by analyzing the changes over time that occurred in a single building component, such as the window.

"The window is a good element for studying the history of buildings," he said. "But it can be misleading because windows may have their own pace of historical change."

Caetano-Anollés said he believes his team's tree of protein architectures is robust because it is rooted in two axiomatic statements that seem to yield consistent and reliable results: First, the structure of protein folds is more stable than their genetic – and protein – sequences. Second, the folds that are the most common in life are also the most ancient.

The researchers suggest that the evolution of protein architectures is a key mechanism by which the three superkingdoms of life emerged from a communal ancestor.

Adopting some protein architectures while abandoning others, perhaps in response to environmental pressures, may have been the first steps on the path of evolutionary divergence.

The research team included postdoctoral researchers Minglei Wang and Liudmila Yafremava, undergraduate student Derek Caetano-Anollés, and professor emeritus in cell and developmental biology Jay Mittenthal. Gustavo Caetano-Anollés and Jay Mittenthal are affiliated with the Institute for Genomic Biology.