

APRIL 05, 2001

Researchers Evolve New Proteins in the Test Tube

To explore how functional proteins evolved, researchers have subjected proteins derived from a massive library of 400 trillion random DNA sequences to a kind of natural selection in the test tube. Their selection technique has evolved proteins unlike any found in nature. Yet the proteins do possess a key feature of many enzymes—the ability to specifically recognize and bind small molecules such as the common biological substrate, ATP.

According to the researchers, the experiment helps clarify how evolution might originally have selected proteins that function as enzymes and perform critical duties in the cell.

"The first proteins, we assume, had to come from random sequences. But it isn't really clear how many random sequences could actually do anything, and thus how easy or how hard it would be for new proteins to arise."

— Jack W. Szostak

In an article published in the April 5, 2001, issue of the journal *Nature*, Howard Hughes Medical Institute investigator Jack W. Szostak and colleague Anthony Keefe at Massachusetts General Hospital reported that they created a random DNA library from which they derived six trillion proteins. Each of the proteins derived from the DNA library was subjected to *in vitro* selection and tested for the ability to bind ATP. The purpose of the experiment, said Szostak, was to understand how frequently natural selection would produce proteins that could fold themselves into a shape that could be functional.

"The first proteins, we assume, had to come from random sequences," said Szostak. "But it isn't really clear how many random sequences could actually do anything, and thus how easy or how hard it would be for new proteins to arise. With this experiment, we wanted to try to put some kind of number on that process—to see whether evolving new proteins is something that could happen easily and frequently, or whether it was an incredibly rare event."

Keefe and Szostak chose ATP binding as the selection criterion both because it represents an easily measured activity and because they had performed similar experiments previously with RNA libraries and they wanted to be able to compare the results.

Starting with a pool of 400 trillion random DNA sequences, the researchers first removed those DNA sequences that contained stop codons that would otherwise halt protein synthesis. They next manipulated the library to yield DNA sequences that would yield proteins that were 80 amino acids long—a length that is sufficient to produce a protein that can fold into a stable three-dimensional shape.

Keefe and Szostak then produced from the DNA sequences the corresponding messenger RNA (mRNA)—the molecule that carries genetic information from DNA to the protein-making machinery. They engineered these mRNA segments in a way that would ensure that the proteins produced would remain attached to their mRNAs. This attachment allowed the scientists to use genetic methods to amplify the genetic information for the rare proteins that emerged from their *in vitro* selection process.

In the *in vitro* selection process, the researchers washed the libraries of fused mRNA-proteins through columns containing ATP attached to plastic beads. The scientists then collected those proteins that bound ATP and converted their attached mRNA to DNA. The DNA was then amplified to produce more mRNA, so that the protein synthesis and selection processes could be repeated. Through successive rounds of *in vitro* selection, which also included mutating the proteins to alter them slightly, the scientists artificially evolved new proteins that could tightly bind ATP.

"From the initial nine rounds of selection, we knew that we had proteins that were binding to ATP and that they had descended from four original molecules," said Szostak. "But they weren't very good at ATP binding, so we began introducing the mutations. After another nine rounds of selection, everything that was left had descended from one of the starting molecules.

"And when we looked at the four original sequences, and more closely at the optimized surviving sequence, as far as we can tell they don't really look much like anything we've seen in nature." Szostak cautioned that when he and his colleagues determine the three-dimensional structures of their artificially evolved proteins, they might, indeed, see some structures reminiscent of natural enzymes.

"We did this experiment to see whether we would come up with structures that were found in nature or were completely different, because we don't really know *a priori* if nature uses every possible structure that could form, or just a small subset of them," said Szostak. "The approach we used was the only way of getting an unbiased view of what proteins can actually do." At least at this early point in their studies, said Szostak, it appears as if nature had some choice in ATP-binding protein structures.

"As we do more experiments, if it turns out that there are basically lots and lots of possible solutions for such a biochemical problem, then that's intriguing," he said. "It suggests that biology just uses a small subset of those possibilities, maybe the first ones to evolve by chance.

"Or, alternatively, there may be many choices and the ones that survive might do so because they can be optimized for other functions in addition to simply ATP binding." For example, said Szostak, the living cell might select for protein structures with certain qualities of stability or ability to interact with other proteins.

The research was supported by grants from the National Institutes of Health and the NASA Astrobiology Institute.