

Laboratory Studies and the Origin of Life

Tutorial by Jack W. Szostak, February 21, 2003¹

Clouds of inanimate molecules swirling in space once formed nebulae, stars, and planets—and, eventually, the biomolecules from which all living things are made. How did life as we know it emerge from such far-flung beginnings?

Today's sophisticated life forms offer few clues. The eukaryotic cells, one of the three major branches of the tree of life, are surprisingly complex. They have highly differentiated internal structures, including cell membranes, nuclei, genetic materials, and more, each with its own function—plus the ability to metabolize and self-replicate. Might we retrace the chain of events that led simple chemicals to combine into these exquisitely ordered living cells?

Molecular biologist Jack W. Szostak, Ph.D. reports that scientists are using advanced laboratory technologies to break down the chemical beginnings of life into small, observable steps. When they complete the job, we will know more about how life may have developed from nonliving molecules. We might also learn more about how easily and frequently life can develop in the universe, and how living things respond to environmental change.

Szostak, a professor in the Department of Genetics at Harvard Medical School, runs a laboratory in the Department of Molecular Biology at the Massachusetts General Hospital. On February 21, 2003 he spoke at a NASA and Templeton Foundation funded workshop, on the Origin of Life, hosted by the Dialogue on Science, Ethics, and Religion, a program of the American Association for the Advancement of Science.

Thinking about how life begins, Szostak said, requires an act of envisioning. “Presumably, the first cells to come into existence had a much simpler level of organization. How can we reason back from the structure of existing living things to imagine much simpler earlier cells?”

The discovery some 20 years ago of RNA molecules that can catalyze certain cell functions helped scientists make this imaginative leap backwards. Perhaps the simpler predecessors of today's cells relied on RNA to perform a wide range of internal biochemical functions. As scientists generate high-resolution images of ever more subtle cell structures, this idea, known as the RNA World hypothesis, has gained support. Szostak showed an image made in the laboratory of Thomas A. Steitz at Yale depicting the inside of a subunit of a ribosome, where amino acids are assembled into proteins. RNA surrounds that region of the cell, clearly governing the chemistry of protein assembly (The Structural Basis of Ribosome Activity in Peptide Bond Synthesis, Paul Nissen, Jeffrey Hansen, Nenad Ban, Peter B. Moore, and Thomas A. Steitz, *Science* 2000 August 11; 289: 920-930).

¹ Refer to CD for accompanying presentation slides

“This is consistent with the idea that early cells used RNA to do all kinds of biochemical functions,” said Szostak, “and gradually RNA molecules learned how to assemble proteins—and that led to the more complex kind of cellular organization that we have today.”

So early cells must have had RNA or an RNA-like genetic polymer (a chainlike biomolecule composed of up to thousands of look-alike subunits) that could self-replicate and perform other biochemical functions. But, Szostak noted, “To get from several chemically activated nucleotides to some kind of very sophisticated macromolecule that could carry out the functions of self-replication—it almost seems like a magical transformation.”

To show how genetic polymers could have developed spontaneously in the early natural world, researchers have demonstrated two ways of assembling activated nucleotides into polymers. One method uses clay minerals to catalyze the growth of RNA chains. The second method, discovered accidentally in another laboratory, freezes activated nucleotides to stimulate them to link up into long polymers.

How might relatively simple, short single-stranded oligonucleotides (short polymers) become the long complex, nucleic acids, like the double-stranded DNA and generally single-stranded RNA polymers in human cells? Scientists know that the replication of a single strand requires two steps. First, the original strand, or template, must generate a complementary strand. Second, those two joined strands, original and complement, must then split apart, and the second strand become a template for another complement. This second complement will be a copy of the first template strand.

In human cells an intricate enzymatic process induces complementary strands to form spontaneously. Over the last 30 years, Leslie Orgel of the Salk Institute and others have partially succeeded using nonenzymatic methods to try to stimulate the self-replication of single-stranded oligonucleotides. But the nonenzymatic processes used in those experiments were relatively slow and built nonuniform chemical linkages. They could not force the second complement to generate a copy of itself, which would complete the self-replication of the oligonucleotide, and so could not show how oligonucleotides might self-replicate in a prebiotic (before cells existed) environment.

Other lines of research have looked at later steps in this reconstruction of the past. David Bartel’s lab at the Whitehead Institute for Biomedical Research has built an RNA molecule that can catalyze the type of polymerization reaction needed for RNA replication. Although it is far from replicating completely and perfectly, Szostak noted, “It’s an important molecule, in that it proves the principle that RNA could copy itself.”

Szostak’s research team constructs scenarios that test the evolutionary process. They use chemically synthesized, information-rich RNA and DNA molecules instead of living organisms to simplify the experiments, imposing selection criteria on the molecular populations and then forcing the molecules to replicate. For example, researchers can require molecules to bind to a target molecule, or to catalyze a specific chemical reaction,

in order to survive the next round of replication. The researchers run a replication cycle several times, making the molecules conform to these selections again and again. In response to these pressures, the populations of molecules change. “That is evolution,” said Szostak.

A remarkable sequence of experiments offers insights about the biochemistry of evolution. In the first experiment, researchers in Szostak’s lab performed a selection forcing RNA molecules to fold into a three-dimensional shape that would bind to the shape of ATP. The RNA structure that emerged from this selection turned up again a few years later, to everyone’s surprise, in a different experiment in a different laboratory. In this second instance, a selection was performed to bind RNA molecules to a compound called NAD (nicotinamide adenine dinucleotide), which, like ATP, contains adenosine as a component. Was there a connection between the presence of adenosine and that particular RNA structure?

It seems so. An experiment in a third lab, designed to bind RNA to yet another adenosine compound, SAM (S-adenosylmethionine), produced the very same RNA structure. Three labs doing three different experiments thus converged in the RNA structure they generated. (This RNA structure is not found in nature.)

What does this convergence mean? “Sometimes,” said Szostak, “one solution is much simpler than others—and that is the one you tend to see.” RNA may have thousands of ways to bind to adenosine, but the structure that emerged independently in these three experiments must be the simplest, and therefore the most common and likely solution.

In some instances no one molecular structure is simpler than the alternatives. Szostak cited one selection for RNA molecules that bind to GTP (guanosine triphosphate) instead of ATP (adenosine triphosphate). “This time,” he said, “we found all kinds of solutions in one experiment. It just happened that there are a number of solutions of a similar informational complexity.” In this scenario, the way an organism evolves would be completely dominated by chance and by history. Whatever sequences the organism population happens to start with would probably determine the population’s evolutionary path.

“As we begin to understand what all of these processes were, we will be able to recreate at least parts of that in the lab,” said Szostak, “and we will be able to build new life forms. In fact we are explicitly trying to do that as a way of testing our ideas and theories about the origin of life.” Scientists will accomplish this by combining nonliving components and organic chemicals to develop something that behaves like a living organism.

Szostak raised a corollary question about evolution, displaying a picture of a catalytic RNA molecule, widely found in nature, which can cut itself in two. “Does the fact that this molecule was found in diverse biological locations mean that all of the molecules have descended from a common ancestor? Or has the molecule just evolved independently many times, so we have an example of convergent evolution?”

To address those questions, his research team built a population of random sequence RNA molecules and selected for those that could cut themselves in two. When the self-cutting activity approached the efficiency seen in living populations, the selection produced the same kind of RNA molecule that was found in nature. Apparently, chemical and physical laws that are constant in the environment influence evolutionary trajectories so that the outcomes will sometimes converge.

The concept of randomness is deeply embedded in modern science (in contrast with earlier Newtonian scientific notions that were more deterministic). For example, chaos theory suggests that the state of a physical system depends so closely on its initial conditions that we cannot predict with any degree of accuracy how the system will change over time. That is, a system can evolve in a myriad of different directions, determined by whatever small changes randomly occur in the initial condition.

In biological evolution randomness is best conveyed through a concept called “sequence space.” A sequence is the order in which subunits appear in DNA, RNA, or protein molecules. The sequence determines the function of the molecule.

The number of possible sequences in genetic polymers is incomprehensibly vast. Even a relatively small protein molecule with only 100 amino acids contains about 20^{100} possible sequences. The entire observable universe could be filled with a lattice of points spaced at the radius of a proton holding about 10^{124} points. Although that amount could fill the known universe, it falls short, by a factor of a million, of the number needed to map all possible sequences in the small protein molecule.

Thus sequence space, which contains all possibilities, is immense. Every evolutionary trajectory starts at a random point somewhere in that space of possibilities. The outcome of that trajectory is indeterminate.

Darwinian evolution, which depends on competition among divergent versions within the same species, can only emerge where a union of two kinds of physical systems occurs. An array of varied heritable information must be present—and arranged in an organizational structure that allows variations to competitively propagate.

For example, a population of polymerase molecules (those that can copy other molecules) floating freely in solution will not be able to engage in Darwinian evolution. If there develops a mutant polymerase that copies more accurately or faster than the originals, it will bump into other molecules in the population and copy them more effectively—but will be copied less effectively by its neighbors. In an unconfined environment, its superior characteristics cannot feed back into the system, and so, because that kind of variation does not bring about a selective advantage, Darwinian evolution cannot take place. A second condition must be present: compartmentalization.

If the same population of replicating genetic molecules with a superior mutant polymerase is confined within a compartment, and that compartment itself is dividing, a

different dynamic develops among the molecules. They randomly copy each other inside the compartment. The compartment grows and divides and the molecules get distributed at random into new compartments. After a few generations some compartments contain only the mutant polymerase molecules, some contain mixtures of mutants and originals, and some have only the original molecules.

In compartments with only polymerase molecule mutants, the replicators copy only each other, and do it faster and more accurately—and are more likely to continue to propagate. Within compartments containing only the original molecules, the chance of propagating is lower because they are not replicating each other as well.

Only where there exists a combination of a replicating compartment system along with replicating genetic information can the potential exist for Darwinian evolution. Living cells have achieved that by forming membranes to confine their contents.

By Nancy Ellen Roth, Science Writer