

The origin of life

I hope I have convinced you in the last four lectures that once life arises, evolution will allow it to diversify and become better adapted to its environment. This doesn't guarantee anything in particular about progress towards intelligence. Indeed, development of intelligence or even complex life might be rare. However, once life gets going evolution will allow it to fill the available niches.

This leaves open a big question: how did life get started? Is it inevitable given the right conditions, or did it require a lot of luck? This is a question to which no one knows the full answer. It is useful to break it up into subparts. We will first consider the simplest lifeforms on Earth to see whether we could imagine them forming from non-life. We will discover that the answer is no: even the simplest life on Earth must have formed from still more primitive organisms. We will then start from the other end: how complicated a set of chemicals can form from simpler chemicals? We will conclude by speculating about what might bridge the two regimes: the so-called "RNA world" hypothesis and its alternatives.

The simplest life on Earth

Consider one of your cells. You are a eukaryotic organism, meaning that your cells are pretty complicated, as are those of any eukaryote (including the single-celled variety). Your cells have membranes that separate the inside from the outside. These membranes are not completely impermeable, but instead have vesicles to allow transport of substances into and out of your cell. Inside your cell there are many "organelles", which are self-contained entities that perform various functions. For example, your cell nucleus contains your DNA; mitochondria generate power (and have their own separate DNA); rough endoplasmic reticulum synthesizes proteins; smooth endoplasmic reticulum synthesizes lipids; and so on. They all function together beautifully, at least most of the time. This is vastly too complicated to have come together from non-life in this form. Consistent with this idea, the simplest eukaryotes don't appear in the fossil record until about two billion years after the first life.

We can therefore take a step back into prokaryotes, i.e., bacteria and archaea. These are far simpler: the cells still have a membrane, but no organelles. Their DNA, for example, is not in a nucleus but is instead free inside the cell. The DNA tends to have far fewer base pairs than that of eukaryotes, and whereas much of eukaryotic DNA is "junk" (in the sense of not coding for proteins; it may have other functions), prokaryote DNA gets as much mileage out of itself as possible. For example, there are often overlapping regions that code for different genes. The cells are smaller than eukaryotic cells as well; the smallest eukaryotic cell is about 0.8 microns across, whereas the smallest prokaryotic cell can be less than 0.2 microns in diameter.

That's pretty small, right? Still, you have to realize that atoms are much smaller still. A water molecule, for example, is about 0.2 nanometers across, about 1/1000 of the diameter of the smallest prokaryotic cell. Therefore there are something like 1000^3 or about a billion atoms in a very small cell. This is way too much to have arisen in one jump, especially because there are some intricate arrangements of proteins in, e.g., the secretory system.

We can then take a step into the even more primitive and ask about viruses. These are not considered alive by some researchers, because they need to hijack a host's DNA to replicate. Still, they do manage to replicate, usually to the detriment of the host. These can be some 20 nanometers across, containing about a million atoms.

Even simpler are the viroids or prions, which can contain just a few thousand atoms. These do not contain the protein coats that characterize viruses, but they are the culprits for some nasty diseases (such as mad cow disease). However, like viruses, prions and viroids require a more sophisticated organism around to reproduce, so we are left with a chicken and egg problem.

Before starting from the other side and asking how complicated molecules build up from simple ones, are we surprised that much more primitive self-replicating organisms aren't found now? I think that we should *not* be surprised. Consider that, by definition, the first self-replicating things were much less adapted to their environment than today's creatures are to theirs. In addition, the environment itself has changed dramatically. Therefore, we expect only relatively sophisticated and well-adjusted organisms currently.

This puts us in a difficult position, but as scientists we won't just give up! We know that molecular clouds have some amino acids in them as well as other organic molecules, but in such small concentrations that it seems unlikely that much more complicated things would form. Do we have any evidence that complex chemicals can form spontaneously in the right environment?

The Miller-Urey Experiment

To answer that question, Stanley Miller and Harold Urey performed a classic experiment in 1953. They put a mixture of water, methane, ammonia, carbon monoxide, and hydrogen gas in a chamber and pumped electricity through it. This was intended to simulate the composition of the early atmosphere and discharges of lightning. After a week they discovered that more than 10% of the carbon had been incorporated into organic molecules, including amino acids, sugars, lipids, and some building blocks for nucleic acids (such as DNA or RNA, although these themselves were not detected). Indeed, complex organic molecules had formed from simple non-organic molecules.

Note, though, that self-replicating molecules had not formed. In addition, later under-

standing of the early Earth suggested that the atmosphere was not as reducing as the above mixture would suggest (i.e., it wasn't as rich in hydrogen compounds, and may have been richer in oxidizing compounds such as carbon dioxide, nitrogen, and hydrogen sulfide). Oxidizing atmospheres have been shown to be harsher for the formation of organic molecules. On the other hand, this has in turn been challenged, and the early atmosphere may have been as much as 40% hydrogen, which is hospitable to the formation of organic molecules. So we don't know for sure, but it is at least possible that the Miller-Urey experiment was not far wrong.

The Miller-Urey experiment and its descendants have not produced life. Indeed, this might be a bit much to ask, given that it probably took on the order of a hundred million years after the end of heavy bombardment from planetesimals to generate life on Earth that we can recognize in the fossil record. However, it did help close one gap: from simple molecules to more complicated molecules. This makes it at least plausible that with enough time and a large enough playing field even more complicated molecules, such as nucleic acids, could emerge.

Note that liquid water is the ideal medium for such drama. It is the "universal solvent", meaning that many different chemicals can float around in water without being destroyed. It is also dense enough that the reaction rate between chemicals can be high. Additionally, since water is a polar molecule (it is positively charged on one side, and negatively on the other) it attracts some molecules and repels others. In particular, there is a class of simple organic molecules called phospholipids; the heads of these molecules are attracted to water, and the tails are repelled from water. Therefore, if a number of these molecules are in the same vicinity within a watery medium, they tend to form a double layer, with the water-hating sides towards each other on the inside. From this stage, it is not a large stretch towards cell membranes.

At this point it is useful to step back a bit. The details of biology, e.g., the specific nature of DNA including its nucleotide bases and the particular amino acids for which it codes, are almost certainly largely the product of chance and specific conditions on Earth. We would not therefore expect these specifics to be reproduced on other worlds. However, simple chemistry works the same everywhere. Put the same combination of chemicals in water at the same temperature and you'll get the same reactions. This means that, at least if you have liquid water and other raw materials, you can expect to get as far as the Miller-Urey experiment did in the formation of complex organic molecules.

In our current knowledge, however, there is still a significant gap. For all but the simplest organisms, DNA contains the genetic code, and DNA is far too complicated to have been the first such genetic molecule. It has, however, been discovered that some organisms use RNA instead. This is a simpler molecule. In fact, it has been suggested that RNA may have

served multiple purposes, from genetic carriers to the transcribers of the code to (possibly) even enzymes. If so, one could imagine that very early life used RNA for all of these purposes and only later did DNA take over because of its greater stability. This is known as the “RNA World” hypothesis, which we will now explore.

The RNA World hypothesis

To set the stage for this proposal it is useful to contemplate the intricate dance done by the molecules in current-day organisms. DNA contains the genetic code, but it gives instructions to messenger RNA (mRNA), which then has those instructions transcribed by ribosomes. Ultimately, proteins are formed. Proteins themselves serve a wide variety of purposes, including as hormones, cell signalers and receptors, and as structural components for many parts of the body. Their best-known function, however is as *enzymes*. Enzymes are molecules that speed up chemical and biological reactions that would otherwise be extremely slow. A simple early explanation for how they work that has held up well is the “lock and key” mechanism, by which an enzyme of a particular shape is able to insert into a target molecule, thereby allowing another molecule to bind. When this is done, the enzyme decouples, unchanged, and goes on to the next reaction.

The point of all this is that our current biochemical system is amazingly complicated. DNA signals mRNA, which produces proteins via ribosomal intermediaries, and enzymes are required to unzip the double-stranded DNA in the first place. Indeed, there is something of a chicken and egg problem: genes require enzymes, but enzymes require genes. For this system to have emerged by small steps it is necessary that there was a simpler system early on that evolved. Do we have candidates?

Yes! In the 1980s it became clear that various forms of RNA can play a greater role than had been suspected (see the as-usual informative Wikipedia article on “RNA World” for more information). In particular, it appears that RNA can (1) duplicate itself or other RNA molecules, (2) catalyze reactions that produce subunits of RNA, hence increasing the number of RNA building blocks, (3) catalyze chains of amino acids, which ultimately might produce proteins. It can also store genetic information, just like DNA. Thus storage, transmission, and duplication of genetic information could all be done by just RNA, without requiring DNA or even, at the start, proteins.

This is not, of course, the current state of life. That is a hint that our current setup of DNA, RNA, and proteins is fitter in some ways. One issue is that DNA is more stable than RNA, meaning that duplications are vastly more reliable and that far more information can be stored in DNA than in RNA. Another is that proteins are far more effective and flexible molecules for enzymatic activity than are “ribozymes” (RNA acting as enzymes). The suggestion, therefore, is that from RNA doing all the work in a sort of pre-life environment,

various mutations eventually produced more stable molecules that took over.

There is very active research going on in this area. For example, Gerald Joyce and his group at the Scripps Institute have recently set up Richard Lenski-like experiments with RNA. That is, they put RNA in a container along with some basic building blocks and let it duplicate and mutate. This occurs on a much faster timescale than Lenski's bacteria can reproduce, so they have set up automated equipment (called "microfluidics") that fill a container with the RNA and building blocks, wait for a while, then take a small part of that solution and add it to a fresh container with just building blocks and no RNA. They have seen overwhelming increases in the "fitness" of the RNA (defined as time to duplication), and because the molecule is so simple they have been able to analyze the molecular structure as a function of time. These are promising experiments but are in their first stages, so keep your eyes open for more developments!

Other ideas for the development of life

A number of people have proposed substitutes for, or addendums to, the RNA world hypothesis. Some think that other molecules are better bets, e.g., different types of nucleic acid such as peptide nucleic acid (PNA), threose nucleic acid (TNA), and glycol nucleic acid (GNA). The idea is that some of these, e.g., PNA, are more stable than RNA and more easily synthesized. It is still far too early to tell, though.

We should end, though, with an idea that is completely unproven and yet interesting enough to merit mention in a course like this. Suppose that although once life starts it can go far via evolution, the steps necessary to achieve first life are extremely difficult. Suppose therefore that the overwhelming majority of even those planets in the right orbits around the right stars don't develop life on their own. Is there another option?

An intriguing suggestion is then the concept of "panspermia", meaning that little viruses or whatever are flying around between the stars on dust grains or whatever. The idea is that life managed to buck the odds and originate *somewhere*, then subsequent asteroidal collisions or whatever kicked off parts of the planet and carried microscopic life. These then drifted around and occasionally encountered a planet that had an environment that could allow evolution and development. Can we make any estimates bearing on the realism of this idea?

One sanity check is that there has to be enough time for dust grains or whatever to travel the needed distances across the galaxy. The escape speed from the Solar System at the orbit of the Earth is about 40 km/s, so let's say that this is roughly the speed at which dust escapes (it could be more or less than this, but that's probably a decent estimate). Then how long would it take to travel $30,000 = 3 \times 10^4$ light years, which is the distance to our galactic center and therefore encompasses much of the galaxy? One light year is about

10^{13} km, so the time is

$$T = 3 \times 10^4 \times 10^{13} \text{ km} / 40 \text{ km s}^{-1} = 7.5 \times 10^{15} \text{ s} . \quad (1)$$

One year is about 3×10^7 seconds, so the time needed to drift across the galaxy would be about 2.5×10^8 yr, or 250 million years. The galaxy is more than 10 billion years old, so in principle there is plenty of time. By the way, keep this calculation in mind later in the semester when we talk about the Fermi paradox!

There are, of course, many further questions. For example, how much matter needs to be blasted out of life-containing planets for a realistic chance to see ours? How likely is it that even viruses would survive a many km/s kick out of the atmosphere of their host planet? There have also been rather extravagant claims about panspermia, notably Fred Hoyle's idea that mass epidemics such as the flu are caused by microbes from outer space. How would you test that? Hint: such microbes would have no reason to be similar to Earth life. Still, seeding of life is interesting and might help if conditions are rare for its start.