Components of life: the origin of the elements

Our course is a bit bold to talk about life in the universe, given that we only know about life on Earth. If we're not careful, it will be easy to slip into the way of thinking that what is critical for life *here* is critical for life *anywhere*. For example, must life originate on a planet the size of the Earth orbiting at our distance from a star very like the Sun, or can there be large differences from our situation?

In this lecture, we'll try to go back to the basics. We will assert that for life to not only originate but also develop requires:

1. A long time. On Earth it took about three billion years to go from single-celled organisms to multicellular life. We don't know if that is slow, fast, or average. However, we can say that based on our observations of life around us, big changes in things large enough to interest us likely take at least millions of years.
2. Some kind of complex chemistry. Atoms found in all life on Earth include hydrogen, carbon, nitrogen, oxygen, phosphorus, and sulfur as the dominant components, with trace amounts of many other elements. Is this critical? We certainly don't know that every one of those is necessary. However, if we are limited to just hydrogen, or hydrogen and helium, our prospects don't look good (see below). We therefore assume that at a minimum we need elements heavier than helium.
3. A universe that is not completely uniform. A uniform universe contains nothing that would distinguish itself, so structures have to form somehow.

As we move through the course we will address some of these speculations in more depth. For example, on Earth, all life needs liquid water to go through its full biological cycle; that is, some microbes and spores can survive indefinitely when completely dry, but need liquid water to reproduce and grow. Is this a universal requirement? How about carbon; is its chemistry unique enough that it is essential to the formation of complex molecules?

Here, however, we give the cosmological background. In this lecture we will discuss what we know about the age of the universe, as well as the formation of elements and planets.

The content of the universe

If we agree that uniformly spread stuff isn't alive, we need some way to cause it to clump. Our understanding of how the clumping happens depends on the content of the universe, so let's survey that briefly.

As you probably know, in the past several years it has become clear that the stuff that we are made of is only a tiny fraction of the mass and energy in the universe. Even more disconcertingly, we don't have a good idea about what the rest is!

More specifically, it turns out that currently only about 5% of the total mass-energy content of the universe is in protons, neutrons, and electrons. 27% is in particles of an unknown type that barely interacts with matter. This is called "dark matter", and believe it or not we can actually rule out every single type of particle ever detected as a dark matter candidate! Many people hope that dark matter can eventually be detected in the laboratory (indeed, there is one group that claims they've already seen it), and in particular that it is something beyond the standard model of particle physics; there are theories suggesting that other types of particles are necessary.

How, though, can we infer that such stuff exists if we can't detect it directly? The way this has been done is through gravity. All mass-energy gravitates, so if there are dim but massive things around they can make their presence known. After all, this is the way that we detected the supermassive black hole in the center of the Galaxy.

One approach is to look at stellar orbits at various distances from the centers of galaxies. Of course, we don't wait around the requisite several hundred million years to see the full orbit! Instead, observers take spectra of galaxies including our own, determine the line-of-sight speed from Doppler shifts, and then use Kepler's laws to estimate the mass contained inside the orbit of the star.

Starting in the late 1950s, Vera Rubin and others began to find that the rotation curves (i.e., orbital speed versus orbital radius) did *not* behave as expected. A spiral galaxy such as the Milky Way has widely distributed mass, but at the Sun's orbital radius and beyond the number density of stars is much less than it is further in. One therefore does not predict a rotation curve $v \propto r^{-1/2}$ as exists in the Solar System, where there is a single dominant mass. In fact, one expects that the rotation speed should *increase* with increasing radius in the inner ~ 1 kpc. However, the diminishing density at greater radii would be expected to lead to a gradually decreasing orbital speed.

This is not what is seen. Instead, the orbital speed of stars more or less flattens out, to a speed that is therefore called $v_{\rm flat}$. Most of the community interprets this as the result of matter we can't see (i.e., dark matter), but some think it might be because gravity itself is modified at low enough accelerations. I personally lean towards dark matter being real, but there is room for argument.

Even weirder, though, is that the rest of the universe, a full 69% currently, is something that is pushing the universe apart faster and faster! This has been labeled "dark energy" and may well be equivalent to Einstein's "cosmological constant" that he introduced in 1917 in a failed attempt to allow for a static universe. We *really* don't know what causes this.

Fortunately for our discussion of structure formation, the importance of dark energy has only come up in the relatively recent history of the universe, as opposed to when the first stars and galaxies were forming. At those earlier epochs, dark energy could be ignored, so we'll drop it (whew!).

How structure forms

First of all, let's sharpen the structure formation question. Right now, there are some parts of the universe that are a lot denser than others. For example, you and I are a modest factor of 10^{30} times denser than an average part of the universe! It is that kind of structure formation that (as far as we know) is necessary for life.

However, in the early days the universe was really, really, uniform. For example, 380,000 years after the big bang (at the time that the cosmic microwave background started streaming freely), the deviations from uniformity were only at the few parts in 10^5 level. How did we get from there to here, and how long did it take?

Remarkably, gravity by itself does the initial job, followed by a little bit of electromagnetism. Since early on dark matter was the major component of the universe (dark energy was not yet important, and electrons/protons/neutrons were just about 1/6 of the total density of the dark matter), we have a relatively easy picture. Dark matter basically doesn't interact in any way except gravitationally, meaning that it is relatively easy to treat. What simulations and theory show is that early on only small concentrations of dark matter can come together and stop expanding with the universe, but that these collections grow larger and larger as time goes on (until dark energy causes the expansion to accelerate, at which point no larger collections of matter come together).

From the standpoint of life, though, this is still unsatisfactory because dark matter just floats hither and yon, simply orbiting rather than forming complex structures. We therefore have to concentrate on what the ordinary matter is doing at this time.

First, recall that ordinary matter has only about 1/6 as much stuff as dark matter. Therefore, the ordinary matter follows where the dark matter goes, at least early on. The ordinary matter has some temperature, so for it to collect and stay in one spot requires that the gravitational potential in a dark matter concentration be enough to hold the matter in. This is thought to have happened first when the universe was maybe 1/200 of its current age, i.e., about 70 million years old.

When ordinary matter is comfortably settled in a collection of dark matter, then the ordinary matter can start to cool. When it does so, it falls closer to the center of the matter distribution; one of the results of this is that visible galaxies, such as our Milky Way, are concentrated in a much smaller volume than is occupied by the dark matter (say, 100,000 light years across versus 2 million light years across). The falling ends up heating up the matter, and eventually, at high enough temperature and density, nuclear fusion sets in and a star is born. It is these stars that have taken the original hydrogen and helium and produced heavier elements: as Carl Sagan said, we are literally made of star stuff. See our second supplement for a discussion of how the elements were produced in our universe.

For our purposes, though, we need more than elements. We need molecules to have the required complexity of life. Amazingly, more than 100 different molecules have been detected from space, so let's talk about them now.

Molecules in space

Before going over some of the molecules that have been detected, let's back up a bit and marvel at our ability to say anything about this at all! Indeed, the power of spectroscopy to inform us about compositions was so unexpected that in 1840 Augustus Comte used the composition of stars as an example of something that would never be known. However, precise measurements of absorption and emission lines has given us remarkably detailed information about stars and galaxies.

To detect molecules requires spectra that depend on their molecular form rather than the individual atoms that make up the molecules. That is, if we are simply looking at a normal transition of an electron, such a transition will be basically the same for a given atom whether or not it is in a molecule (not quite, but close). Therefore, the transitions that are studied are vibrational or rotational transitions. These are much lower energy than electronic transitions, so we observe them using radio waves. Large molecules themselves tend to be more fragile than small molecules, so typically one should look in cold, dense regions such as molecular clouds.

This has been an extremely productive enterprise. There are, of course, plenty of inorganic molecules. H_2 is the most common molecule, CO is seen easily, and SiO, OH, NaCl and many others have been observed. However, many organic molecules have been observed as well; indeed, nearly 200 at last counting. For example, simple sugars have been found, as have different types of alcohol (followed by much drinking, I'm sure!). More impressively from the standpoint of life is that the simplest amino acid, glycine, has also been discovered in interstellar clouds. The fundamental role that amino acids play in life on Earth makes this interesting because it suggests that organic molecules, the building blocks of life, could well be common in the universe.

What we learn from all this (including the supplementary material) is that the universe had to age a fair amount, between a hundred million and a billion years depending on the location, before heavy elements were present. We also find that when enough such elements are present, and cold, dense environments are available, many molecules form. Interstellar clouds, though, despite being dense by astronomical standards, are nearly perfect vacuums by terrestrial standards: maybe 10^{-13} of atmospheric pressure even for the cores of such clouds! We presume (and we take a bit of a leap here) that much greater condensation is needed for life. Now, it could be that somewhere around a distant star one can find life on a gas giant such as Jupiter. We really don't know. However, since we live on a rocky planet, we will now discuss the formation of planets and in particular the terrestrials.

Formation of terrestrial planets

With our penchant for seeing everything in terms of ourselves, it is natural that our first instinct is to think that life is most likely to arise on a planet similar to Earth. We do have a lot of advantages: liquid water on the surface, thick enough atmosphere to protect us from high energy radiation such as X-rays, a magnetic field that deflects high-energy cosmic rays, and so on. Of course, it could be that somewhere else in the universe a species is congratulating itself on being on a perfect gas planet. We don't know. Still, it does make sense that planets would be good places for life. Stars are too hot for complicated molecules to exist, and cold places in space (such as interstellar clouds) are so low-density that it's difficult to imagine much activity taking place on time scales short enough to be of interest. We therefore need to know a little more about how stellar systems form, and how in particular Earth-like (terrestrial) planets form.

The formation of stars

The details of this process are surprisingly complicated, and there is a great deal that we still do not know. The basics, however, are reasonable. We start out with an interstellar gas cloud, which is a giant thing: many parsecs across, high-density by astronomical standards (hundreds of atoms per cubic centimeter; of course, our atmosphere is more like 10^{20} atoms per cubic centimeter, so this is a really good vacuum!), and cool to the tune of $T \sim 10$ K. Such a cloud might contain a million solar masses or more, and it thus has significant gravity within itself. Little pockets of the cloud start coming together, cooling and radiating, and getting denser and denser. You can think of this as equivalent to stuff falling in a gravitational field, which means that the kinetic energy (and thus temperature) of the stuff in the middle gets larger and larger. Eventually, the center of such a pocket gets dense enough and hot enough that it undergoes nuclear fusion and becomes a star.

Sounds straightforward, right? The difficulty is that hidden in this description is a problem. The problem is that because the pocket starts out so large, any little bit of rotation means that the gas has a huge amount of angular momentum. In fact, the lowest reasonable amount of angular momentum is many orders of magnitude greater than what any star has. But you can't just cause angular momentum to vanish, because it is conserved. Think of an ice skater who brings her arms in while spinning around. As the gas pocket contracts, it spins faster and faster. This causes it to flatten out into a disk, but to become something as tiny as a star the gas that forms the star has to transfer almost all its angular momentum to something else.

This is a problem that isn't 100% solved. Most of the angular momentum probably leaves via the outermost portions of the gas disk escaping from the system. However, a lot of the rest can end up in planets. That is, if you think about the current-day Solar System, Jupiter's orbit has about 100 times as much angular momentum as the Sun's spin does. What we do know is that the angular momentum barrier means that matter can't just plunge in and form a star. Instead, a gas disk is formed, with matter near the inner edge flowing slowly in to form the star while the outer matter moves away. This disk is thought to last a few million years, based on observations of stars at this stage. It is from the disk that planets must form, so let's examine that process.

Formation of planets: considerations of composition

How can planets form out of the disk? One thing that might occur to you is that this could be a smaller version of the original formation of stars from interstellar gas clouds. That is, perhaps some extra-dense pocket of the gas in the disk could become self-gravitating, pulling itself together to form a planet. This is something that may indeed happen for some large extrasolar planets. However, it has difficulties explaining the composition of many planets. Let's explore this a bit.

The Sun contains some 99.8+% of the mass of the entire Solar System, so we would expect that the mass fractions of different elements in the Sun would be representative of the whole nebula from which it formed. Additionally, we can check out the mass fractions of elements in current nebulae. In both cases, we find that by mass, hydrogen is about 73% of matter and helium is about 25%, with the remaining elements making up about 2%.

How about for the Earth? Here the fractions are dramatically different. Iron is the most common element by mass, constituting some 32% of the Earth. Oxygen is next, at about 30%, followed by silicon, magnesium, sulfur, and others. Hydrogen and helium aren't even remotely competitive. The same is true for the other inner planets (Mercury, Venus, and Mars, as well as our Moon). Jupiter and Saturn's overall compositions are close to that of the Sun, but for Uranus and Neptune we again run into discrepancies. There must be some other mechanism.

Planet formation: the standard model

Rather than straight collapse from a gas cloud, how else could a planet come to be? The key is to realize that atoms and molecules can stick together in grains (for rocky stuff like silicon and iron) or small ice particles. These grains then have enough electric charge on them that they can come together in larger and larger groups. These rocks/iceballs/whatever then collide, grow larger, and eventually produce planets. The mass of the planets, however, depends on how far out they are. Close enough to the host star, things are hot enough that water ice and other ices cannot be formed. This means that hydrogen is a gas rather than being in a solid form with oxygen, so very little mass can be brought together. This restricts the inner planets to starting out with iron, silicon, and their oxides. Since this is a tiny fraction of the original nebula, the inner planets are small.

However, far enough out, water ice can form. This involves hydrogen, by far the most abundant element. It is then thought that as the ice balls get big enough, the gravity can attract and retain the gaseous helium and hydrogen that is floating around (this in itself is made easier because far away from the star the temperature is low enough that the hydrogen and helium move slowly). At the right location, then, giant gaseous planets can form. Jupiter, being the closest giant planet, was able to scoop up more mass than the others.

Caveat: much of this picture was developed before other solar systems were formed, hence we might be arguing more specifically than we should. Still, we do get some bits of useful insight:

- Planets that *form* close to their host star are likely to be pretty small, because ices can't form. There may be ways to form giant planets far out and then *bring* them close, but that is a different story.
- Innumerable grains/rocks/planetesimals would have formed early on, and they would grow by collisions with their fellows. This means that the early solar system was a violent and dangerous place. As an example, it is thought based on many independent lines of evidence that our Moon was formed when a Mars-sized object hit the proto-Earth (note that it wasn't Mars itself, just a similar-sized object). Any life that somehow cropped up at this stage (from 4.6 billion years ago to about 4 billion years ago) was almost certainly wiped out by these collisions. Indeed, as we'll discuss when we hit mass extinctions, the baby versions of such collisions still happen occasionally.

  Let's sum this up and then think about what it means. Our current understanding is that Earth and the other terrestrial planets are mainly made out of stuff that could make grains even in the hot solar system. That's why we have lots of iron, silicon, and their oxides (iron oxide is rust; silicon dioxide is sand). It is *not* the case that "heavy stuff sank to near the Sun", which is something you commonly hear. Initially, all the nebula everywhere had the primordial composition; it's just that light gases such as hydrogen and helium didn't take part in grain formation.

Reflections on terrestrial planets

If life really does require terrestrial planets, what does the formation scenario mean for how common life is? To answer that question partially, we can imagine a first-generation star, i.e., one that forms in the universe early enough that nuclear fusion plus winds and supernovae haven't had enough time to populate the interstellar medium with heavy elements. In that case, a star will form with only hydrogen and helium. No ices can form, so the only path to a planet would be direct collapse. Even this wouldn't make a terrestrial planet.

We therefore realize that formation of Earth-like planets requires plenty of heavy elements. Indeed, even a second-generation star would likely not have nearly enough iron, silicon, and oxygen to make a planet like Earth. Therefore, the formation of terrestrial planets, as well as the development of complex chemistry, seems to require that significant time go by so that the interstellar medium be somewhat enriched in heavier elements. There is some evidence that even more massive planets form more easily around stars that have extra amounts of heavy elements, but this is disputed. It is even possible that our Solar System is one of the earliest systems that had enough heavy elements, but (1) we don't know what the threshold is, and (2) a billion years here or there would make no significant difference compared to variations in the local environment.

Properties of the Earth

When we compare the different inner planets in our Solar System in a few lectures, we will realize that apparently small differences in mass or orbit can make a huge difference. Let's start with the Earth's size.

As you may have gathered from the formation scenario we discussed, when the Earth and the other planets were formed they were rather hot (indeed, molten) due to the violence of their many collisions. However, the inner planets cooled off at different rates. You can think of it like this: imagine that you have taken a freshly-cooked pie out of the oven. If you want to cool it off as quickly as possible, are you better off letting it stand as is, or cutting it into many smaller pieces and letting those cool off? The latter, of course. This also works for planets. Large ones, such as the Earth, cool off much more slowly than small ones, such as Mercury or Mars (or the Moon, if you want to consider that an honorary planet).

This has a surprising number of important consequences. One is that the Earth has a relatively thin crust on top of a molten interior. Convection in that molten interior breaks up and moves the crust around in plates, leading to plate tectonics (and causing earthquakes, volcanoes, the beautiful fit of Africa and South America, and lots of other good stuff). Historically, this has caused ecosystems to shift around as the shore-to-interior ratio of continents has altered. This may have played an important role in evolution. The liquid nature of the Earth's interior has also allowed us to sustain a magnetic field that is strong enough to deflect high-energy charged particles from the Sun. It's difficult to forecast what would have happened to life here without the field, but current life would be severely damaged by those particles. In addition, of course, the size of the Earth allows us to retain a significant atmosphere.

In contrast, the Moon and Mercury are small enough to have cooled off almost completely. There do appear to be liquid components in both, but not enough for plate tectonics. The same is true for Mars. Venus, which is just slightly smaller than the Earth, is a special case. We see no active tectonics at the moment, but at the same time it appears that there are no craters older than about 500 million years on the surface. Why might this be? One intriguing possibility is that the crust is thicker than on Earth, but that energy continues to be released in the center because of radioactive decay. In this picture, when the energy buildup is large enough, the granddaddies of all eruptions explode in many places, causing lava to resurface Venus. This might have happened many times in the past, but in this picture the last one was half a billion years ago.

By the way, let me forestall one possible misconception. We think of the Sun as providing our heat, so it would be natural to imagine that the inner planets Mercury and Venus would be kept hot mainly by the Sun, with their interior heat playing a minor role. However, note that here we are concerned with whether *rock* is solid or molten, and the Sun's heat is nowhere close to enough to melt rock even at Mercury. Therefore, yes, the surface temperature does depend significantly on the Sun's illumination, but the thickness of the crust depends on how long the interior takes to cool.

How special is Earth? A first look

As we move through the course we'll occasionally return to the issue of how special our situation is. We can now revisit this by noting that some people think that Earth is in a uniquely privileged environment, so that very few planets should be expected to support life. Examples of some of the proposed specialness include:

- Earth had to have formed right about now in cosmic history. Earlier, and not enough heavy elements would be present. Much later, and maybe so much mass would be present that the surface would be constantly wracked with quakes or even molten.
- The low eccentricities of the Solar System are unusual. Stuff smashing into other stuff tends to decrease eccentricity because the random motions average out, but near the end of major planet formation there are a few large bodies moving around and these deflect each other gravitationally rather than colliding. Simulations indicate that this leads to high eccentricities, and indeed many extrasolar systems are like that as well.

What do you think?

In any case, even if we have a planet of the right type in the right orbit, life still has to develop. The origin of life is difficult to constrain strongly (although we'll discuss it in a later class). However, once life has taken hold, further development occurs via a powerful and elegant mechanism called evolution. We will discuss many aspects of evolution over the next two lectures.