



The Information Theory of Life

The polymath Christoph Adami is investigating life's origins by reimagining living things as self-perpetuating information strings.

By Kevin Hartnett



Chris

Adami on the Michigan State University campus.

There are few bigger — or harder — questions to tackle in science than the question of how life arose. We weren't around when it happened, of course, and apart from the fact that life exists, there's no evidence to suggest that life can come from anything besides prior life. Which presents a quandary.

[Christoph Adami](#) does not know how life got started, but he knows a lot of other things. His main expertise is in information theory, a branch of applied mathematics developed in the 1940s for understanding information transmissions over a wire. Since then, the field has found wide application, and few researchers have done more in that regard than Adami, who is a professor of physics and astronomy and also microbiology and molecular genetics at Michigan State University. He takes the analytical perspective provided by information theory and transplants it into a great range of disciplines, including microbiology, genetics, physics, astronomy and neuroscience. Lately, he's been using it to pry open a statistical window onto the circumstances that might have existed at the moment life first clicked into place.

To do this, he begins with a mental leap: Life, he argues, should not be thought of as a chemical event. Instead, it should be thought of as information. The shift in perspective provides a tidy way in which to begin tackling a messy question. In the following interview, Adami defines information as “the ability to make predictions with a likelihood better than chance,” and he says we should think of the human genome — or the genome of any organism — as a repository of information about the world gathered in small bits over time through the process of evolution. The repository includes information on everything we could possibly need to know, such as how to convert sugar into energy, how to evade a predator on the savannah, and, most critically for evolution, how to reproduce or self-replicate.

This reconceptualization doesn’t by itself resolve the issue of how life got started, but it does provide a framework in which we can start to calculate [the odds of life developing in the first place](#). Adami explains that a precondition for information is the existence of an alphabet, a set of pieces that, when assembled in the right order, expresses something meaningful. No one knows what that alphabet was at the time that inanimate molecules coupled up to produce the first bits of information. Using information theory, though, Adami tries to help chemists think about the distribution of molecules that would have had to be present at the beginning in order to make it even [statistically plausible for life to arise by chance](#).

Quanta Magazine spoke with Adami about what information theory has to say about the origins of life. An edited and condensed version of the interview follows.

QUANTA MAGAZINE: How does the concept of information help us understand how life works?

CHRISTOPH ADAMI: Information is the currency of life. One definition of information is the ability to make predictions with a likelihood better than chance. That’s what any living organism needs to be able to do, because if you can do that, you’re surviving at a higher rate. [Lower organisms] make predictions that there’s carbon, water and sugar. Higher organisms make predictions about, for example, whether an organism is after you and you want to escape. Our DNA is an encyclopedia about the world we live in and how to survive in it.

Think of evolution as a process where information is flowing from the environment into the genome. The genome learns more about the environment, and with this information, the genome can make predictions about the state of the environment.

If the genome is a reflection of the world, doesn’t that make the information context specific?

Information in a sequence needs to be interpreted in its environment. Your DNA means nothing on Mars or underwater because underwater is not where you live. A sequence is information in context. A virus’s sequence in its context — its host — has enough information to replicate because it can take advantage of its environment.

What happens when the environment changes?

The first thing that happens is that stuff that was information about the environment isn’t information anymore. Cataclysmic change means the amount of information you have about the environment may have dropped. And because information is the currency of life, suddenly you’re not so fit anymore. That’s what happened with dinosaurs.

Once you start thinking about life as information, how does it change the way you think about the conditions under which life might have arisen?

Life is information stored in a symbolic language. It's self-referential, which is necessary because any piece of information is rare, and the only way you make it stop being rare is by copying the sequence with instructions given within the sequence. The secret of all life is that through the copying process, we take something that is extraordinarily rare and make it extraordinarily abundant.

But where did that first bit of self-referential information come from?

We of course know that all life on Earth has enormous amounts of information that comes from evolution, which allows information to grow slowly. Before evolution, you couldn't have this process. As a consequence, the first piece of information has to have arisen by chance.

A lot of your work has been in figuring out just that probability, that life would have arisen by chance.

On the one hand, the problem is easy; on the other, it's difficult. We don't know what that symbolic language was at the origins of life. It could have been RNA or any other set of molecules. But it has to have been an alphabet. The easy part is asking simply what the likelihood of life is, given absolutely no knowledge of the distribution of the letters of the alphabet. In other words, each letter of the alphabet is at your disposal with equal frequency.

The equivalent of that is, let's say, that instead of looking for a self-replicating [form of life], we're looking for an English word. Take the word "origins." If I type letters randomly, how likely is it that I'm going to type "origins"? It is one in 10 billion.

Even simple words are very rare. Then you can do a calculation: How likely would it be to get 100 bits of information by chance? It quickly becomes so unlikely that in a finite universe, the probability is effectively zero.

But there's no reason to assume that each letter of the alphabet was present in equal proportion when life started. Could the deck have been stacked?

The letters of the alphabet, the monomers of hypothetical primordial chemistry, don't occur with equal frequency. The rate at which they occur depends tremendously on local conditions like temperature, pressure and acidity levels.

How does this affect the chance that life would arise?

What if the probability distribution of letters is biased, so some letters are more likely than others? We can do this for the English language. Let's imagine that the letter distribution is that of the English language, with *e* more common than *t*, which is more common than *i*. If you do this, it turns out the likelihood of the emergence of "origins" [increases by an order of magnitude](#). Just by having a frequency distribution that's closer to what you might want, it doesn't just buy you a little bit, it buys you an exponentially amplifying factor.

What does this mean for the origin of life? If you make a naive mathematical calculation of the likelihood of spontaneous emergence, the answer is that it cannot happen on Earth or on any planet anywhere in the universe. But it turns out you're disregarding a process that adjusts the likelihood.

There's an enormous diversity of environmental niches on Earth. We have all kinds of different places — maybe millions or billions of different places — with different probability distributions. We only need one that by chance approximates the correct composition. By having this huge variety of different environments, we might get information for free.

But we don't know the conditions at the time the first piece of information appeared by chance.

There are an extraordinary number of unknowns. The biggest one is that we don't know what the original set of chemicals was. I have heard tremendous amounts of interesting stuff about what happens in volcanic vents [under the ocean]. It seems that this kind of environment is set up to get information for free. It's always a question in the origins of life, what came first, metabolism or replication. In this case it seems you're getting metabolism for free. Replication needs energy; you can't do it without energy. Where does energy come from if you don't have metabolism? It turns out that at these vents, you get metabolism for free.

If you have achieved that, the only thing you need is a way of moving away from this source of metabolism to establish genes that make metabolism work.

Your take on the origins of life is very different from more familiar approaches, like thinking about the chemistry of amino acids. Are there ways in which your approach complements those?

If you just look at chemicals, you don't know how much information is in there. You have to have processes that give you information for free, and without those, the mathematics just isn't going to work out. Creating certain types of molecules makes you more likely to create others and biases the probability distribution in a way that makes life less rare. The amount of information you need for free is essentially zero.

Chemists say, "I still don't understand what you're saying," because they don't understand information theory, but they're listening. This is perhaps the first time the rigorous application of information theory is raining upon these chemists, but they're willing to learn. I've asked chemists, "Do you believe that the basis of life is information?" And most of them answered, "You've convinced me it's information."

Your models investigate how life could emerge by chance. Do you find that people are philosophically opposed to that possibility?

I've been under attack from creationists from the moment I created life when designing [[the artificial life simulator](#)] *Avida*. I was on their primary target list right away. I'm used to these kinds of fights. They've made kind of timid attacks because they weren't really understanding what I'm saying, which is normal because I don't think they've ever understood the concept of information.

You have footholds in lots of fields, like biology, physics, astronomy and neuroscience. In a [blog post last year](#) you approvingly quoted Erwin Schrödinger, who wrote, "Some of us should venture to embark on a synthesis of facts and theories, albeit with second-hand and incomplete knowledge of some of them." Do you see yourself and your work that way?

Yes. I'm trained as a theoretical physicist, but the more you learn about different fields, the more

you realize these fields aren't separated by the boundaries people have put upon them, but in fact share enormous commonalities. As a consequence, I have learned to discover a possible application in a remote field and start jumping in there and trying to make progress. It's a method that's not without its detractors. Every time I jump into a field, I have a new set of reviewers and they say, "Who the hell is he?" I do believe I'm able to see further than others because I have looked at so many different fields of science.

Schrödinger goes on to say that scientists undertake this kind of synthesis work "at the risk of making fools of ourselves." Do you worry about that?

I am acutely aware of that, which is why when I do jump into another field I'm trying to read as much as I can about it because I have a bias not to make a fool of myself. If I jump into a field, I need to have full control of the literature and must therefore be able to act as if I've been in the field for 20 years, which makes it difficult. So you have to work twice as hard. People say, "Why do you do it?" If I see a problem where I think I can make a contribution, I have a hard time saying I'm letting other people do it.