PROFESSOR: Good morning. Come on down. All right. So last time, we talked about the first couple of steps of the central dogma. The central dogma is this name given to the statement DNA goes to DNA by replication. DNA goes to RNA by the process of transcription. We call it transcription because it's such a direct copying of letter to letter. RNA goes to protein by a process of translation, because translation is the kind of word we would use between two different languages. And there are two different languages-- the languages of nucleotides and the language of amino acids.

For what it's worth, it's name, the central dogma, goes back to Francis Crick, who called this the central dogma. He did it in a kind of a light-hearted way, although others since then have criticized it, saying, oh, this is dogmatic. Actually, Crick said not that DNA goes to RNA goes to the protein. He says that all information flows from nucleic acids to proteins and not back again. Because even then, Crick knew that in theory, there was no reason you couldn't go back from RNA to DNA, and as we'll see today, that happens.

So sometimes people say, oh, well, the central dogma was proved wrong because RNA can go back to DNA. Well, actually, that was even anticipated right then at the very beginning, when they realized RNA and DNA were essentially equivalent information. The key step of converting nucleic acid information into protein information is translation. And that's the last bit we have to fill in and talk about today.

So let me start by just reminding you-- because most of you know it-- how translation works. So, in translation, you have a particular RNA that has been made by the cell. Here's my RNA. It goes five prime to three prime. That's always the way we write these things. And it has some particular sequence. I'll make up a sequence here. A, U, A, C, G, A, U, G, A, A, G, A, G, G, C, C, C, dot dot dot dot dot, U, A, G, dot dot dot dot, three prime.

All right. Somehow, that's going to be translated into a protein. It's translated

according to a fantastic look-up up table. This look-up table is called the genetic code. And the rule, the algorithm-- and it's a fairly simple algorithm, well, nothing's ever really simple in biology, but it's close to simple-- is you run along the sequence from the beginning. And you guys could write this. Anybody who's taken basic computer programming. Run along the sequence and find the first occurrence of A, U, G. Why A, U, G? Because A, U, G is the place you start. That's how life worked it out, and that's what it does.

After that, you parse the sequence in triplets. These triplets get the name codons. And then you keep going until you hit one of three possible triplets. U, A, G. U, A, A. U, G, A. And you stop there. Any MIT student should be able to write an algorithm that takes a string, finds the first occurrence of U, A, G, breaks it up into triplets passed there, keeps going until you encounter one of these three triplets.

What do you do for a given triplet? You look it up against the table. How many triplets are there? How many three-letter words are there with the four nucleotides?

AUDIENCE: 64.

PROFESSOR: It's 4 to the 3rd-- 64 possible words. I've used up three of them to be stop, so there are 4 to the 3rd possible codons. That's 64. Three stops. The other 61 possible codons specify an amino acid. That's it. There's a look-up table. The genetic code. How many amino acids are there? 20. And there are 61 possible codons, so that implies there is some redundancy. Some codons-- some amino acids are coded for by the same codon. Okay. That's fine.

And in your book is the genetic code that is the look-up table here. The genetic code translates these codons into amino acids. Or to stop, in the case of the three stop signals.

So for example, this A, U, G at the front is always translated into methionine. Met. And if I've got it right, this should be a lysine. Here we go. Arginine, proline, et cetera-- you just look it up. That's it. That's the order in which you make the proteins. So you send off an order written in RNA, you send it off to the factory, the

factory sends you back a protein that is methionine, lysine, arginine, proline, blah, blah, blah, blah, blah. That's it.

This genetic code is essentially universal amongst all of life. That's pretty stunning. What does that tell you? The fact that all of life uses virtually the identical genetic code? There's actually a tiny difference between prokaryotes and eukaryotes affecting a codon and there's a tiny difference somewhere else. But essentially, it's the exact same genetic code that all of life uses.

It's very unlikely that this genetic code is the only possible way you could make a genetic code, right? So the fact that all of life uses, essentially, exactly the same code is pretty strong evidence that all currently existing life descends from a common ancestor. Because if these were evolved independently, it's extremely unlikely that you would have gotten exactly the same genetic code. So that's an interesting point that you can see from just the fact that everybody uses essentially the same genetic code. It's a universal genetic code.

All right. So I've expressed this to you in a completely computer sciencey kind of way. But, of course, the cell doesn't do this by computer science because cells are unable to write C code. The reason that cells are unable to write C code is C wasn't really developed until the last several decades, and cells, pretty sure, precede the development of C code by Kernighan and Ritchie. So it's got to be the case that it's done some other way. How is it done? Well, it's done like this. And I'll just be very schematic and you'll get it in your book.

There's a big machine. The big machine here is called the ribosome. It consists, itself, of proteins and RNAs, and it's a huge structure. The huge structure needs to read codons. And this is a case where Francis Crick drove everybody nuts. Francis Crick, back in the 1950s, just sat at his desk and thought. He was terrible at doing experiments-- nobody really wanted to let him do experiments. Francis was a great thinker. He thought. He said, golly, how can this sequence be translated into amino acids? Well, people at the time had all sorts of nutty ideas. Some of the nutty ideas was that the sequence of the RNA folded up into pockets that just fit a proline. And

another pocket that just an arginine. And if you just think about the constraints to get that to work, it's nuts. That the sequence itself would form perfect binding pockets for the necessary amino acids.

Crick said impossible. He said the really sensible way to do this, if I were running life, what I would do is I would have an adapter molecule. The adapter molecule would be some kind of a nucleic acid, and the nucleic acid would kind of match the codon on one end and have the amino acid on the other end. And then there'd be another adapter, and the adapter would have the next amino acid. And then you would catalyze a bond between them. Therefore, I predict, says Francis, there will be small adapter molecules, probably made out of RNAs themselves. And Francis called this the Adapter Hypothesis. It drove people crazy because, of course, he was right. People found the adapters and they would use transfer things that transfer information-- get called transfer RNAs.

What happens is the ribosome has pockets in which these transfer RNAs basically come in, match their sequence, there's a codon each of these transfer RNAs has a matching anticodon that matches the triplet, and it has already attached to it an amino acid-- the right amino acid for that anticodon. How does that right amino acid get attached to the right tRNA?

There's an enzyme. The job of that enzyme is to attach this amino acid, proline, to this transfer RNA. It's a Prolyl-tRNA synthetase. And its job is to put proline on the right tRNAs. There's another one that puts arginine on the right tRNAs. There's a whole business that's set up to get the right tRNAs, have the right amino acids attached to them through a bunch of enzymes floating around. So then these tRNAs with the amino acid attached drop in, they drop into the next position, and a bond-- the peptide bond-- is catalyzed.

Interesting factoid. The catalysis, this enzymatic catalysis to join together those amino acids is actually carried out not by the proteins in a ribosome, but actually by RNA in the ribosome. The RNA is the enzyme. You know why that's kind of cool? If this bothers you, just forget it, but one of the mysteries about how you ever go from

DNA to RNA to protein and all of that is how the whole thing ever got started. How could you possibly have gotten protein synthesis started if the things that were needed to make protein synthesis were proteins?

So this is actually an echo of an ancient world 3 billion years ago, where this was all probably carried out by RNAs. RNA was probably the early catalysts for most things, and we still see evidence of the fact that even today your peptide bonds are catalyzed actually by RNA and they're doing the enzymatic work. Anyway, it's kind of cool. If you didn't get that, don't worry about it, but it's kind of cool.

So then what happens after you attach the first two amino acids? Well, the ribosome chugs down here and grabs the next code-- these two shift over. You can either think about the ribosome moving this way or the RNA moving that way. The next tRNA drops in, the next bond gets made, chugs over, the next one drops in, the next one gets made, and onward like that until it hits a stop codon at which, it releases it. The ribosome knows to release it. There's actually a little factor that drops in and tells it to release it there. And that's how you make proteins-- kind of cool. It works very well. It chugs along in that fashion.

For you computer scientists, what you basically have is a two-tape turing machine-- you're reading one tape and writing to the other tape. There's a nucleic acid tape and there's an amino acid tape. If you're not a computer scientist, forget I said that.

OK. So it's basically a two-tape tape turing machine where the RNA is coming through here and the protein is coming out that way, but the amino acids attach to each other until it comes off. In fact, actually, very recently, a Nobel Prize was awarded last year for beautiful, beautiful work on how this actually takes place-- the molecular details of the ribosome. Really gorgeous. Any questions about that? Yeah.

AUDIENCE: Is that not susceptible to the same error as replication is?

PROFESSOR: Oh, is that not susceptible to the same error? Does the ribosome ever make mistakes? It does. What happens if you make the wrong protein?

AUDIENCE: [INAUDIBLE].

PROFESSOR: Oh well. The answer turns out to be "oh well" because for any given DNA, you make lots of RNAs. For any RNA, you use it again and again to make lots of proteins. And if the occasional protein is not so good, if the quality control is not perfect, it's much less serious than if your master instructions in the DNA were not perfect. So the cell actually devotes a lot less attention to quality control.

Now. That's not to say there isn't important quality control. There are quality control mechanisms, but it doesn't have to be as accurate as one error in a billion, like you want to be in copying your DNA. And that's really an important point, is the archival copy has to be really good, but little yellow sticky notes-- what you make from it, which are basically RNA or the little yellow sticky notes you copied down-- they don't have to be perfect. And each copy, the protein doesn't have to be-- and if the protein is not perfect, there are mechanisms that take unfolded proteins and degrade them. So it turns out there are ways to achieve quality control in that sense. It's a great question.