

MIT OpenCourseWare
<http://ocw.mit.edu>

7.012 Introduction to Biology, Fall 2004

Please use the following citation format:

Eric Lander, Robert Weinberg, and Claudette Gardel, *7.012 Introduction to Biology, Fall 2004*. (Massachusetts Institute of Technology: MIT OpenCourseWare). <http://ocw.mit.edu> (accessed MM DD, YYYY). License: Creative Commons Attribution-Noncommercial-Share Alike.

Note: Please use the actual date you accessed this material in your citation.

For more information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>

MIT OpenCourseWare
<http://ocw.mit.edu>

7.012 Introduction to Biology, Fall 2004
Transcript – Lecture 11

OK, so what I'd like to do today is pick up where we left off last time, with respect to how this genetic material actually functions.

We discussed last time the experiments that identified DNA as the fundamental genetic material, the transforming principle. We identified the eventual work

by Crick and Watson's work out of the structure of DNA as a double helix. We mentioned why that was so tremendously important, because it contained within it in principle the secret of replication, namely two strands, each of which contained the full information, and therefore each of which could, in principal, serve as a template for making the other strand. And that is, after all, the big issue about life is how do you, in fact, copy life? And then, I mentioned briefly these experiments

by these post-docs, Matt Meselson and Frank Stahl about 50 years ago to demonstrate that the semi-conservative model of DNA replication was right by virtue of actually labeling DNA during the course of its replication in one generation, and demonstrating that DNA actually changed in its density when you added in an isotope of nitrogen. And, it changed in its density in such a way as to be intermediate between what you'd expect from heavy, heavy, light, light.

You have the intermediate. So, that was all good experimental confirmation that this model was probably right. But now, how does it really work? After all the excitement calms down for a moment you say, OK, that's great. We now know in principal it's there, but what actually goes on? How is DNA really replicated? How is it really read out into information? How does it really, as Archibald Garrett noted, and as Beadle and Tatum noted, how does it really make protein as well?

How does it encode the instructions for that? Well, that was what was on people's minds in the late '50s. And, it was Francis Crick who was the real intellectual thinker about this. And, the eventual synthesis

that you guys all know, because, again, all this stuff gets taught in elementary school these days, was encapsulated in the central dogma of molecular biology,

which I will summarize here diagrammatically. The DNA is replicated to make copies of DNA. It's read out into the intermediate RNA, and then it is translated into protein. This process: translation. This process is called transcription.

And this process: replication. And what I'd like to do is go into some detail today about how each of these processes work. Now, at the beginning, when people were trying to patch this together, it wasn't as obvious as it is to you today, that DNA goes to RNA, goes to protein. And, in fact, it was a real struggle to figure out what this RNA stuff was doing in the middle, how it could possibly give rise to protein. I want to talk about some of that. Let me briefly mention, though, Francis Crick's term, the central dogma, because it sometimes

gets criticized, the word dogma there as being like religious belief and molecular biologists treated in this way. I've read a couple of social scientists who sort of say, "oh, dogma". In fact, Francis Crick deliberately named this the central dogma because he said there was no proof for it at the time it was put forward. He put it forward with that word precisely to emphasize that this was a working guess. But, it was merely a matter of belief that this is sort of how they were putting together the pieces.

And it was really a question of demonstrating how all these pieces work. We still call it the central dogma, but it's now, of course, extraordinarily well established. Let's look at this first piece. DNA is replicated. All right, so Meselson and Stahl tell us that, yeah, the DNA weight look like the new strand, the old strand, all that. How would you really demonstrate DNA replication? If you wanted to show me that DNA replication really

happens, this DNA goes to DNA, that somehow we had to take a double strand of DNA, and it gives rise to, it's one thing to show this in a bacterium by adding the nitrogen and all that. The way to really prove this was to be in a test tube. In vitro, reconstitute for me DNA replication. Show me that in a cell free system,

you can take DNA, and you can copy it as you would expect according to the Crick-Watson model here. Well, that is what Arthur Kornberg set out to do. Arthur Kornberg was a biochemist, and so his interest

was crack open the cell, and purify an enzyme that was able to copy DNA. Now, how do you do that? What cells should you pick?

Sorry? Why E coli? What a bacteria? It's simple, exactly. Good answer. You can grow up a lot of it, and presumably, if this DNA replication thing is right, it will apply to any organism. So, we'll go with E coli. So, what do you do? You just crack open a cell and purify components, and throw them in a test tube, and look for DNA synthesis? Well, you've got to put something in the test tube. What should we put in the test tube? Sorry?

Nucleotides, because we think that this is going to be made out of nucleotides. So, we'd better add some nucleotides to our test tube. So, actually, deoxynucleotides, we'll add some dATP, dCTP, dGTP, and dTTP, the deoxynucleotide triphosphates, altogether known as the dNTPs.

OK, that's good. So, we're going to take different fractions of the cell. We'll add it here. We'll add some nucleotides, and what else should we add? Well, if we were going to copy DNA, maybe we ought to put in a DNA strand. Let's put in a DNA template. So, let's put in a template strand of DNA

that we'll copy, here we go, and we've got our nucleotides floating around here. And, here's our template strand, a single strand of DNA, and now we add enzymes, and we hope that it's going to somehow copy the DNA. Now, it turns out that that's a little bit optimistic because in order to copy the DNA, and I think Kornberg had this insight,

it's helpful to give it a start. So, instead of just adding a single template strand, he also added a short complementary primer strand with the hope that he would be able to purify an enzyme, which even if it couldn't manage to start the synthesis of DNA, would be able to extend the synthesis of DNA.

That's a reasonable thing. Let's not ask for it all at once. Maybe it won't be a single fraction. Maybe multiple enzymes would be needed to get going. So, he needed a primer strand, a template strand, and some nucleotides. And then he added fractions, and he looked to see whether he could get incorporation of DNA. So now, let's look at this a little more closely. The primer strand goes like this. Five prime, ah,

This direction is going to matter a lot, I told you. Phosphate T, phosphate A, phosphate C, phosphate G, phosphate T, phosphate A, stop there. Template strand, the complement to that, will start in the opposite direction. These are anti-parallel. What matches the T: A. Keep going: T, G, C,

A, T, and phosphate, phosphate, phosphate, phosphate, phosphate; I'll stop writing the phosphates in a while. Let's say T, A, G, G, C, etc. This is the five prime end. That is the three prime end, OK? And, this one will go on further, let's say. All right, so what is the enzyme

that Kornberg hopes to find going to do? What's it going to add to the strand? It's going to add an A. All right, it wants to put in an A here. So, it's going to take a triphosphate, and it's going to catalyze the addition of a triphosphate to the growing end of this DNA chain, and which is its growing end? The three prime end

of the chain there, right? It's adding it to the three prime carbon there. And, when it does that, where is it going to get the energy for catalysis here for this chemical reaction here? It's going to get it from the dehydration synthesis and the breaking of this triphosphate bond, which is a high-energy bond. You'll take off your inorganic pyrophosphate and you'll add in an A. That's it.

Then, it will go off and it'll look for, what, a T, a triphosphate with T, DTTP, and then DCTP, etc. And it adds them in. This enzyme, this hypothetical enzyme, that can polymerize DNA like that is called polymerase. It's all very simple stuff. This is DNA polymerase. OK, and the nomenclatures here make tremendous sense. This is called DNA polymerase. Anyway, Kornberg isolated by a lot of work

DNA polymerase, and was able to demonstrate that it could in fact catalyze this reaction. This was incredibly exciting. He got a Nobel Prize for this amongst other things, but he really demonstrated that there were proteins that could copy DNA according to this double helical model for replication. I call your attention to the fact that the replication goes

five prime to three prime always, ever, all the time. This is universal. No one has ever found a DNA polymerization system in nature where it goes the other way. And, why would that be? This is just a digression. But tell me why that would be? Let's take our strand here, T, G, C, A,

T, T, A, G, C, G, T, why not go this way? Why not go, let's say, A, G, C, G. Let's see, what base should we put in? We'll take our triphosphate T, right? We'll put that in.

Let's see, where are we going to get the triphosphate bond; where are we going to get the energy? The triphosphate's on the wrong end. Oh, that's not a problem because when we put this G in, it must be that its triphosphate was still there, right? So, now we'll take the next one, a triphosphate T, and now why don't we just

carry out the polymerization using the triphosphate bond, the energy from the triphosphate bond, on a growing chain going in that direction? That would work, right? Just stick this guy here. It'll supply a new triphosphate at the end, and that triphosphate can be used to catalyze the next monomer. So, what's the problem? You could put the triphosphates on the growing chain. If we went this way, the triphosphate bond would be on the growing chain, rather than in this way the triphosphate is on the monomer.

But who cares? Who might care? If you were designing it, which way would you prefer to do it? The one with the energy first, well, why do you care whether the triphosphate is on this big, long chain that you've made, or whether it's on this monomer because either way you've got a triphosphate bond that could be on the monomers floating around, or it could be in that last position with the growing chain. Yeah?

Could be, could be. What kind of mistake might I make? Yep. And, you know, what other kind of mistakes can happen? What about these high-energy triphosphate bonds: unstable? What if they should just spontaneously hydrolyze?

Oops: big trouble, right? You've lost your triphosphate bond, and but what if this one spontaneously hydrolyzes? Aren't you in trouble? No, get another monomer, right? Clearly, it's no big deal if one of the monomers spontaneously hydrolyzes from a triphosphate to a monophosphate, but it's a big deal if you've invested all of this energy going in the other direction, and it should spontaneously hydrolyze.

So, it makes a great deal more sense to leave that high-energy bond on the monomer for the growing polymer rather than on the polymer itself. And, in fact, of course, nature hasn't told me why it chose to do this. This is my reason why I think nature chose to do this, but I think it's very reasonable, and I think it's right. So, this is not the way it's

done. This is the way it's done, and it's always done that way. No one has ever found a case where it's not. OK, so now let's look a little more closely at DNA replication.

Suppose I take not just this teeny little piece that Kornberg gives, but suppose I now look at what's going on in an organism. An organism might have a big, long chromosome. DNA replication is occurring along this chromosome. We've got to go five prime to three prime, five prime to three prime.

Let's suppose there's a primer here. Wait a second, where's the primer going to come from? If Kornberg's not there to add the primer, what does the organism do? To kind of make one itself, and I'm going to need some enzyme to make it. So, what enzyme's going to make it? Or, primase: it turns out to be remarkably, coincidentally

it's primase that makes the primer. It's funny how that works out. And so, primase makes the primer, and then what happens? Then, DNA polymerase comes along and catalyzes the addition, and works beautifully. What about on the other strand? So, it's got a what? Why does it have to play catch-up? Let's see, what kind of primer here?

It's got to go the other way. OK, so let's get a primer here. So, but wait a second, now it breathes and opens up a little more. We've got to get a primer here. And then, when it's going to open up even more we've got to get a primer there. See, this guy's going the wrong way. So, in fact, this is what happens. When the DNA opens like this, one primer here is sufficient to keep going, but here as you begin to open this up, the other strand needs

the continual addition of new primers, and then what happens when this DNA sequence here, growing, meets this DNA sequence there? They've got to be ligated together. They've got to be joined together. So, this is actually getting kind of complicated. We have little DNA fragments that have to be ligated together on this strand. Now, how are you going to ligate them together? Chemically, you've got to catalyze a

covalent bond between this little growing DNA chain and the previous growing DNA chain that was there. How are you going to ligate them? Ligase: yes! Coincidentally, it turns out that ligase does that. It's just wonderful the way this worked out, that ligase should do the ligation, and primase should do the primer, and all that. All right, so this goes

on and on. Now, this model, which is what would be compelled by what we're thinking about is experimentally proven.

There was a scientist who demonstrated that on this strand, this one goes slower, right, because it's got to, just as you said, catch up. Playing catch up, this is what's called the lagging strand. This guy is called the leading strand. The lagging strand plays catch-up to the leading strand. And, these little fragments can actually be really, truly identified biochemically. They were identified, in fact, by somebody called Okazaki. And, do you know what they're called, those fragments?

Okazaki fragments, exactly. That's what they're called. So, that's how it goes, and it goes with this continuous replication, and then this discontinuous replication there. Now, here's another problem. This upset people a lot. Try to take a long chromosome. In fact, let's even imagine that it's a circular chromosome like bacteria have, a big DNA circle.

Imagine trying to replicate this. All right, we're going to pull this apart some. We'll start replicating as we'll continue to pull this apart, etc., etc., but the problem is that we're going to end up with this DNA helix and this DNA helix wrapped around each other

so that we're going to have double helices, or we're going to have interlaced double helices. It's really very messy. Topologically, if I take a double helix and I copy the two strands, and the double helices went around each other 800 times before they got to the end and joined up, I've now got two circles of DNA that are

inextricably linked together with what's mathematically called the linking number of 800. That's not very good when I try to now divide my cell and say, in one chromosome to one cell and one chromosome to the other cell because I've got these two long, continuous ropes that are just so totally knotted with each other. This bothered people tremendously. You can prove, mathematically, some of you take the topology courses

that there is no way without cutting to pull apart two strings that are so intertwined with each other. So, how in the world is life going to do that? It's mathematically impossible to do that without actually cutting. So, it cuts it because it's got no choice, right? There's a theorem that says you have to cut it. So, it cuts it. You would actually need, it turns out, that if you're

going to separate out these two different double helices that are all wound up around each other, you're going to need to somehow cut the DNA, separate it, and pass it through the other side. And, you're going to need to do that to un-knot this thing. Now, does it change it chemically when you cut it and bring it around to the other side of the string?

It's still the same molecule, right? It's the same DNA, but topologically it's different. The two circles are now not linked to 800 times their links, 799 times, and if I keep doing that, so they are, you could call them topoisomers because they differ only in their topology, their topoisomers. So, you would need an enzyme that actually cuts the DNA, and is clever enough to pass it to the other side and then seal it back up, and cut the DNA, and pass it through the side and seal it back up.

What enzyme does that? Topoisomerase does that, that's right. And, there are topoisomerase enzymes that cut and paste the DNA to resolve this terrible linking number problem. So, life has worked all this stuff out, and there's just fascinating work that goes on to understand, woops, all of the steps there of DNA replication.

Now, I mentioned that these are actually pretty important things because processes like this are very important to rapidly growing cells. It turns out that some very good anti-cancer drugs are inhibitors of topoisomerase because rapidly growing cancer cells are highly sensitive to the need to continue to topologically untangle your DNA. And so, topoisomerase inhibitors turn out to be pretty good, well, they're not great, but they turned out to be acceptable cancer drugs.

Here's another issue: fidelity. The fidelity of DNA replication. If I'm copying the DNA, I'm going to put in my next base. It's a T. I want to put in an A, a G, I want to put in a C; how do I get it right? I have my DNA polymerase enzyme here.

How do I manage to get this right? Why don't I put in a G next to the T instead of an A? Well, it's energetically less favored, right? Energetically, there's some cost. There's a ΔG , an energetic difference between the right base and the wrong base.

Now, if I know ΔG , I from biochemistry know the equilibrium constant. I should be able to calculate, based on the energetic difference between putting in the right base and the wrong base how

often DNA polymerase makes a mistake, and it turns out you can do that. It turns out that the equilibrium constant is about 10^3 .

That means that DNA polymerase, remarkably, gets it right 99.9% of the time, it puts it in the right base. Isn't that impressive? No, it's terrible. Why is that terrible? Yeah, 99.9%

this is no Six Sigma performance or anything. This is pretty unimpressive stuff. I mean, a typical gene is more than 1,000 letters. That means we're going to actually make a mistake on average in every gene. This won't do. So, what happens? Sorry? Well, clearly the energetics say that the delta G is only enough to get us a factor of 10^3 . We're going to need an additional mechanism, and the additional mechanism's a proofreading. It's absolutely right.

We need to proofread this because we know that initially we're going to get it wrong at an unacceptably high rate. And so, it turns out that there are two kinds of DNA proofreading that go on. First off, DNA polymerase itself has a proofreading activity. Whenever DNA polymerase adds a base, it kind of also has an activity that will remove a base. So, it doesn't just add bases going forward.

It also has what's called an exonuclease activity that removes bases going backwards. Now, that may seem silly, right, because it's adding and subtracting, and adding and subtracting, but it adds more than it subtracts. And, the trick is that if there's a mismatched base, it's much more likely to subtract than to add, or much more likely to subtract than if there's not a mismatched base. So, the presence of a mismatch induces the enzyme to do its removal more than if there was a match.

In that fashion, DNA polymerase is able to substantially increase its proofreading ability to about one error in 10^5 or 10^6 , much better in one in 10^3 . Then, it turns out that there are mismatched detection and repair enzymes. They come along after

DNA polymerase has done its job, and they feel along the DNA for any mismatches. Mismatches are going to create funny structures. They're going to bulge in some way. And, mismatch repair enzymes are able to detect that something's funny, and they chop out some sequence, and they get copied back in. Now, with the proofreading that comes from these mismatched repair enzymes, you can get down to the neighborhood of one mistake in about 10^8 bases.

In the course of the human, yes? Oh, what a great question! Because, when it has a mistake, how does it know who to correct? In bacteria, I can tell you the answer. Wouldn't it be cool if you could leave a mark on the old strand? If the old strand could be temporarily

marked in some way so that the enzyme, when it sees a mismatch, would also know which strand to cut out and re-synthesize? It turns out that bacteria do that. Methylation enzymes actually mark the old strand. And, it takes a while before those methylation enzymes come along to mark the new strand, and it leaves a temporary mark as to who's the old strand. I wasn't going to mention that today, but it's a great question. So, it leaves breadcrumbs for a while that tells it who's the old strand.

So, all of this gets worked out. Yes? So, the exonucleases go backwards. They go three prime to five prime because, that's right, they only work in that direction. There are other exos that go in the other direction, but this exo on the polymerase go backwards, three prime to five. Now, this is not just theoretical stuff. It turns out that about one person in 400, that is, probably at least one person in this class,

is heterozygous for a mutation in one of the mismatch repair enzyme genes. One of the genes like MSH-2 or MLH-1 that encode the mismatch repair enzymes. What do you think happens if you are missing one of your two copies of these mismatch repair enzymes? Nothing much. The other copy's enough. But, what do you think would happen if by chance a single cell in your body were to lose the one remaining working copy of that enzyme, the gene-encoded remaining working copy?

Then it would have no copies. What do you think the response of the cell would be? High mutation rates, and cancer. It turns out that familial, hereditary, nonpolyposis coli, a familial form of colon cancer, is caused by, in many cases, mutations in the gene or genes, actually, encoding the mismatch repair enzymes.

So, our theoretical understanding of the central dogma here is an incredibly practical disease because getting DNA replication right is important. And, that provides a very good proof that the difference between 10^5 or 10^6 here and 10^8 accuracy matters a great deal, that without that mismatch repair enzyme present in the cells, one is in fact going to create new mutations at an unacceptably high rate and

lead to cancer. A few other random nice facts about DNA polymerases.

They're very fast speed. The speed of a DNA polymerase is about 2,000 nucleotides per second: very impressive. And then, one last point I can't help but mention, Arthur Kornberg discovers this enzyme, shows in a test tube, it works, people work out, how it works in detail, leading strands, lagging strands, topoisomerases, workout

fidelity, all these kinds of things, great. But Kornberg's enzyme, the enzyme he purifies that copies DNA, is it actually the right enzyme? Is it the enzyme that the bacterial cells he used actually use to copy their DNA? Well, a biochemist would say, I cracked open the cell. I purified a component. It's able to carry out this function.

There you go. But, what would the geneticist say? Sorry? Take out the component, and demonstrate now what? That the cell can't replicate. It's DNA. Until you've shown that, you haven't got the other half of the proof. So, of course, some geneticists decided to put this to the test. They took many mutant bacteria. One at a time, they grew them up,

and they did Kornberg's purification to purify DNA polymerase. This is unbelievably tedious stuff, guys. You've got to take each one. You've got to purify it; get DNA polymerase. OK, it's there. Next one, next one, next one, next one. But, suppose you found a mutant which couldn't make Kornberg's DNA polymerase but still grew and replicated its DNA. That would prove that Kornberg's enzyme was not essential. They did. It turns out that Kornberg's enzyme,

DNA polymerase 1, although it can replicate DNA in the test tube is not the enzyme that cells actually use for their major DNA replication. It turns out to be a relatively more minor repair enzyme used to fill in gaps. The actual enzyme is DNA polymerase 3, not that it matters to you a great deal, but this duality between the biochemistry and the genetics is very important because just the biochemical side of the story, without showing that it was

essential to the function in the organism misses a very important point there. So, the combination of genetics and biochemistry, biochemistry pointed us to a class of enzymes. The genetics, then, identifies which ones are used for which purposes in vivo, which is not that easy to do in the test tube. Anyway, I mentioned that, and obviously being a

geneticist, I like tweaking the biochemists about things like that. All right, onward. So, in our picture of DNA replication,

in our picture of the central dogma, we've got DNA goes to DNA, and what about the step of transcription, DNA goes to RNA? Well, we've got to copy out our DNA into an intermediate molecule called RNA, which is going to then be used as a template for protein synthesis.

Where do we start? Somewhere in here, there's some information. We want to make a copy of that information. How do we know where to start? Well, there's something. There's some information that says start here, right?

There's a little sign that says, start here. Such a thing is called a promoter. And, the promoter, which we'll come and talk about more in a while, probably in a lecture or two, the promoter says here's the place to start copying the DNA into RNA, and

it gets copied into the RNA by an enzyme that starts here, let's say, I don't know, T, A, T, G, G, T, A, T. On the other strand I guess it's going to be A, T, A, C, C, A, T, A. It's going to start copying here, and it's going to put in an A.

Then opposite the A, it's going to put in a U, because RNA has U, A, C, C, A, U, A, etc., except this time it's doing it not out of DNA but out of RNA. How does RNA differ from DNA? So, first off, instead of deoxyribose,

this is deoxyribose. In fact, it's two prime deoxyribose. This is just plain old ribose. Remember down there on the two prime carbon, DNA had just a hydrogen, whereas RNA has a hydroxyl.

All right, that's one difference, and it turns out that that hydroxyl is important because it would interfere in making long double helices of RNA. RNA doesn't make good, long double helices. Let's entirely do that, oxygen. And, the other major difference between DNA and RNA? The only other difference between DNA and RNA is that this has U where this has T, and what's the difference between T and U?

A single methyl group. That's the only difference between T and U. In this six-member ring over here, there is a methyl group. And here in the six membered ring, there's no methyl group. That's it. Why does RNA use U, and DNA use T? Anybody know?

It's not a big difference. That would be interesting, although I don't think it's true. I actually have no idea. I think this is fascinating. I've never had a good accounting of why it uses U and T. You need to know this, and it's true, but I don't actually have a,

whereas I have a good explanation for this I don't have a good explanation for that, although maybe some of my Origin of Life colleagues have an explanation. But I've always been a little puzzled. Why does it use U instead of T? Anyway, I do know why it doesn't have the hydroxyl. Well, it has the hydroxyl there. That really does affect the base stacking, and all sorts of things like that. All right, so you, don't go away, come back. So, the DNA is used as a template to copy here a strand of RNA.

Some important names: the strand that is being copied that is being transcribed is called the transcribed strand. This is called the non-transcribed strand that makes good sense. This is also called the coding strand. And, you will find it in your books as the coding strand. Why is it called the non-coding strand?

This is called the coding strand. Why is the top strand called the coding strand? Because the RNA that I copy out will have the same sequence as the coding strand, except for T's and U's. So, the RNA copy that is made from

the transcribed strand matches the sequence of the non-transcribes strand, or the coding strand. So, you will find this confusing, but you will probably find it on tests and some things like that to know which strand you're looking at. The coding strand is this strand which has the code that ends up, but in fact it's the template for the coding strand, the complement to the coding strand, the non-coding strand, the transcribed strand that is copied. Anyway, I've said that now, and you can,

So, how does it know where to stop? Sorry? Stop codons. Stop codons are actually about translation into protein, right, because we're going to come to stop codons in a second. There was some start signal there called a promoter, which is a start of transcription. It turns out there was also a stop signal that says stop of transcription. And, you guys haven't probably met that before.

But, there's a start signal, a stop signal, and all over the genome there are these things. So, here's some genome. Here's some gene that's got to be read out. And, it's read out this way, let's say. This is the

coding strand. This is what, I'll make two strands here. Now, in the next gene over here,

does it go in the same direction? It might. Or, it might not. It turns out that the orientation of genes along the chromosome, which way you read, is not a fixed thing across the entire length of the chromosome. So, when I refer to the transcribed strand or the non-transcribed strand, that's just a local definition

that says, with respect to that gene, this strand is coding, and this strand is non-coded. But with respect to the next gene over, it could be the other way. Now, this is not a very orderly way to do things, right? If a good engineer did this, they'd probably get all the pieces going in line and all that. But life did this, and it turns out that evolvable systems, you know, couldn't possibly maintain that order. Things are happening all the time, and genes can come in any order. In addition, how does RNA polymerase know when to turn on the gene?

Oh, sorry, what's the enzyme that polymerizes RNA? RNA polymerase, yes. How does it know when to turn on the gene? How does it turn on the right genes in the right tissues? We'll come to that. That's gene regulation. That's a big non-trivial thing. We'll save that one. All right, so we have all of this transcription. Let's now look at the last important part of our picture here, which is translation. So, RNA goes to protein. So, if RNA goes to protein, we take our messenger, our RNA over there. This is an RNA. What's the direction it's been copied? Five prime to three prime.

It's a single strand of RNA that we've copied here, a single strand and molecule, and let's give it a sequence, A, U, A, C, G, A, U, G, A, A, G, C, C, C, etc. Eventually we'll get to U, A, G. How is this RNA interpreted?

Well, in an abstract sense, the way this RNA is interpreted is by a triplet code. The cell could come along and start reading three letter codons. But, does it just start anywhere? No, it always starts at the same codon, and that codon is A, U, G. This is an initiator codon.

And it encodes a methionine. Then, the next codon down encodes lysine, arginine, etc. The interesting challenge is how in the world you get from a sequence of nucleotides to a sequence of amino acids.

So, we have to now get this funny translation step between nucleotides and amino acids. This concerned people greatly because transcription was pretty easy. Transcription was going to be the RNA, actually first replication,

each nucleotide would match a nucleotide on the DNA sequence. Then, RNA polymerization, each nucleotide of RNA would match. But how are we going to get amino acids to match specific RNA sequences? How are we going to get amino acids? Now, this bothered people a great deal. And, you know what some of the ideas were? Well, proteinase, right.

Some enzyme...well, actually the first ideas were very physical ideas. It was that the RNA message there would fold up into some kind of a funny shape that would just happen to match a lysine, and then the next little bit would fold up to match, I don't know, histidine, a methianine, and a serine, and a this, because people were thinking the complementarity of DNA bases all just physical matching that it would work that the amino acids would be

directly read off the RNA message. But, it was kind of crazy to imagine that because the amino acids all have such wildly different physical properties: positive charges, negative charges, hydrophilic, hydrophobic, different sizes. It just didn't make sense, but it bothered people a great deal. But, I would say that a lot of biochemists thought that that was sort of how it was going to have to work. The guy who really figured out what was going on did it with no experimental data whatsoever.

He did it by just sitting down and saying, that doesn't make any sense. There's got to be another solution. And, that was Francis Crick. Francis Crick just had an incredible mind. He, Mendel, and a few other people had this incredible insight into things. He said, look, this just makes no sense that the physical properties are going to do it. He said, what's got to be going on is that what I want to put in a certain amino acid

into a growing protein chain, I'm going to take my amino acid here. I'm going to take my codon here, and I'm going to build me some kind of an adapter. And, this adapter molecule will, in fact, solve the problem. So, he said, because Francis Crick, in addition to being brilliant, really didn't do any experiments.

He didn't do any experiments both because he wasn't that fond of doing experiments, and because he was legendarily not very good at the bench. But, what Francis did was he exhorted all of his colleagues to go find the adapter. He had what he called the adapter hypothesis. And sure enough, Crick was dead on, just right. The adapter hypothesis turned out to be that there was an adapter molecule

who was made itself out of RNA called transfer RNA. And, transfer RNA matched up by base pairing to each codon you see, and had amino acids attached to it and so the problem of how you mediate between a three-letter code of DNA or RNA, of nucleotides, and amino acids was solved by a clever intermediate.

It turned out that they looked, they found the molecule. So, it's just one of these great examples of somebody having thought up an idea, sent people off to look for it, and it was there. And then, of course, you've got to ask, how did the amino acids get stuck onto the right transfer RNAs? And the answer is there's a bunch of specific enzymes that do precisely that job, that look at the transfer RNA, attach the amino acid, and handle that whole problem. I will next time briefly end with the ribosome, and how those transfer RNAs work to

catalyze together the protein chain, and then what I want to do is turn to how this common picture of DNA, RNA, and protein varies amongst organisms. Until next time.