ERIC LANDER: And so the issue became how does DNA replication work. And so I'm about to go into it. Now, I'm going to note we're going to be starting this DNA goes to RNA, goes to protein, and DNA goes to itself. DNA is replicated. It makes RNA. The RNA is used to make protein. This will be what we'll be talking about today and tomorrow. So the first step of that is, how does DNA give rise to more DNA?

Well, how do you find an enzyme? How do you do biochemistry? What do you do?

AUDIENCE: Assays.

ERIC LANDER: Assay. So you've got to grind up the cell. I got to choose a cell in which I'm likely to find an enzyme, grind it up, break it up into different fractions, and test each fraction. That's all biochemists do, right? So what cell might have the enzyme we're looking for? What cells might be able to copy DNA? How about all cells?

So let's use a simple cell. What's a simple cell? Let's use bacteria. So we'll take some bacteria, we'll grow it up, we'll grind it up. We'll fractionate it into different fractions, and we'll see if one of those fractions has the ability to copy DNA. If we're going to run an assay, we have to give it a substrate. What substrate would you like to give it? What do you think it needs?

AUDIENCE: [INAUDIBLE].

ERIC LANDER: It better have some free nucleotides otherwise, how are we going to make DNA. What else? Are you going to ask it to make DNA all by itself? We want something that can copy one of the strands of a double helix. So what should we give it?

AUDIENCE: [INAUDIBLE].

ERIC LANDER: Sorry?

AUDIENCE: Half a helix.

ERIC LANDER: Half a helix. A strand of DNA, the strand to be used as a template. So let's give it a template strand. So we'll take a template strand of DNA. There's my template of

DNA. Let's actually give it a little sequence actually, here. Let's say A phosphate, T phosphate, G phosphate, C phosphate, A phosphate, T phosphate, T phosphate, A phosphate, G phosphate, G phosphate. I'm going to not write the phosphates too much longer, guys, but anyway C phosphate, C phosphate, T phosphate, like that.

Pretty soon, in fact, almost immediately, I'm going to start dropping the phosphates in here. But that's the way it goes. All right. That's a template. We need floating around in the solution some trinucleotides. We have some nucleotides floating around. And now will this enzyme work? We would try different fractions and see if it's able to just install the right letters in the right place.

Now, it turned out it needed one more thing, and the person who discovered this, Arthur Kornberg, thought of it. It needed a head start. It needed a primer. So the primer goes let's say, phosphate T, phosphate A, phosphate C, phosphate G, phosphate T, phosphate A, let's say like that. So this is the five prime end of DNA. Remember the phosphate is hanging off the five prime carbon, right?

What's look at the other end. The other end ends in the hydroxyl on the three prime end of the ribose. Since this is anti-parallel, this strand is going five prime phosphate to three prime hydroxyl. You're going to need to know five prime and three prime. So I'm doing this so you get used to five prime and three prime.

There you go. If you're handed a primer to get a head start, and you're handed a template, and you hand it some nucleotides, you then assay different fractions exactly as you suggested and we see is one of them capable of extending this strand by putting in an A, putting in a T, putting in a C, putting in a C, putting in a C, putting in a G. That's the assay.

And Arthur Kornberg discovered an enzyme that could do this. And the biochemists went nuts. They thought, wow. This is so cool. Kornberg is able to discover an enzyme that can accomplish this. The enzyme polymerizes DNA. Coincidentally, what is the enzyme called?

AUDIENCE: DNA polymerase.

ERIC LANDER: DNA polymerase. Accidentally, has a nice name. Good. DNA polymerase. Excellent. Now, notice what it does. It takes this triphosphate, puts it in here, and it joins it into a sugar phosphate chain. Where does it get the energy for that synthesis? Hydrolysis of the triphosphates, right? It's the hydrolysis of the triphosphate. That's the energy.

What direction is the synthesis proceeding? Starts here at the five prime end, and it moves adding on to the three prime end. So it's five prime to three prime direction. That's the direction it moves. It adds to the three prime end. It adds to the free nucleotides to the three prime end. Why not do it the other way?

AUDIENCE: [INAUDIBLE].

ERIC LANDER: Sorry?

AUDIENCE: Phosphates.

ERIC LANDER: Can't hear you. Shout loud.

AUDIENCE: Phosphates.

ERIC LANDER: Phosphates, yes. You see, suppose we were going the other way. Suppose the primer was this way. Where would as we added each base, the triphosphate would be on the strands, right? And we'd be adding to the three prime end here. That means the energy supplied by the triphosphate would be on the growing strands rather than in the free nucleotides.

Why would it be a terrible idea to put your energy source on the growing strand?

MIKE: [INAUDIBLE].

ERIC LANDER: Well Mike, you know, those triphosphate bonds are pretty unstable. They hydrolyzed by themselves at some frequency. If you're a free nucleotide and the triphosphate hydrolyzes, big deal. That free nucleotide floating around loses its triphosphate. But what if I'm the growing strand, and I lose my triphosphate?

AUDIENCE: [LAUGHS]

ERIC LANDER: Exactly.

AUDIENCE: There goes your chain.

ERIC LANDER: There goes my chain. So you know, life's not stupid. It doesn't do it that way. It does it this way. No one has ever found a polymerase that goes this way. They find them all going that way for just that reason. Exactly. Bingo. That was why life evolved it that way, because you want your triphosphates, those hydrolyzable triphosphates to be floating around freely rather than investing.

Now just think about that. It's a kind of cool thing. It doesn't matter. Your book doesn't talk about it. But to me, it helps me remember which way it's going and how it is, and it's kind of interesting. Any way.

All right. So Kornberg wins the Nobel Prize for this. Good stuff. It's very deserved, but you know, there's some questions. Where does the primer come from in life? See, Kornberg gave this test tube a primer. But suppose I'm replicating some DNA. So let's suppose I have a double strand of DNA, and I'm just going to open it up here, five prime to three prime, five prime to three prime.

I need to get like a primer here. Then the primer can be extended by polymerase. Well, where's the primer come from? It turns out there is an enzyme specially devoted to making those primers. Kornberg didn't know it, but there's an enzyme. And by coincidence, it is called primase. Exactly. Primase makes the primer. So you need a primer here, and the primer is made by primase.

Once primase makes a primer, polymerase can chug along and do it just fine. Let's check out the other strand. Primer here, polymerase chugs along. But now as this double helix opens up, what happens over here? The synthesis going this way. So what do I have to do here?

AUDIENCE: [INAUDIBLE].

ERIC LANDER: Another primer. Need another primer. Then as it opens up more, what do I need?

AUDIENCE: Another primer.

ERIC LANDER: Another primer. So the two strands are experiencing very different kind of replication. In one place, one primer in the five prime to three prime direction is enough to keep going. In the other strand, as it keeps opening up, you gotta keep making primers. You have all these little fragments there.

Now, those little fragments were discovered by Okazaki, and they are called Okazaki fragments. Again, I just mention these things. They are known to molecular biologists. But these little guys are Okazaki fragments, and they tell you that you're on the right track here. This is indeed how it's working. You can see those little fragments there. But now, what's the problem with the Okazaki fragments?

They're not connected, right? The primase makes a primer. The polymerase copies the DNA, it bumps into the next primer, but you've got to connect them. So that's a problem. That's a real problem. I'll redraw that here. Here was my primer. I got a new primer over here. I got a new primer over here. Right there. Right there. They're not contiguous connected.

The word we use for connecting two pieces of DNA, which is a standard English word not used that often is to ligate two things together. Ligature, for example, in music. You ligate things together. How do you think the cell deals with ligating these things together? An enzyme called--

AUDIENCE: Ligase.

ERIC LANDER: Exactly. So ligase does the ligation. Ligase ligates. It is so lucky that these words turn out to have accidentally made sense. It's really cool. So ligase ligates.

Now, I'll tell you a factoid, but don't worry about it too much. Primase actually doesn't make DNA. We haven't gotten there yet, but it turns out primase makes RNA. Turns out to be easier to start an RNA than a DNA from scratch. Cell doesn't like to start DNA from scratch. It likes to start RNA from scratch as we'll get to a moment with transcription.

So as a factoid, I'll mention to you that those little primers are actually RNA primers, and what happens is they get extended into DNA, and they bump into and kind of displace the previous RNA, so it's slightly more complicated than I told you. You're welcome to forget that. If you would like to believe that primase is actually making little segments of DNA, it'll be just fine.

But in fact, it doesn't actually. It's making little segments of RNA so there's a whole other machinery that has to deal with that. But the basic concept five prime to three prime, little primers, getting extended, getting ligated, that's how you make your DNA. And you can check it out, and it works. All right.

Well, it turns out to even be a little more complicated. That was how we got the synthesis going, but we also have a little bit of a topological problem. This again, says a lot about how people do science. You gotta just like not worry about certain things.

If Kornberg had said, oh my goodness. I can't give my test tube a primer, because I don't know how the cell would make a primer, he wouldn't have made any progress. So he throws in the primer and says, the cell will figure it out. I'm just giving it a primer, and I'll see what happens.

Now, there's another problem, this topological problem that also can make your head hurt. Let me try to explain what the topological problem is. Suppose I have DNA like that. Make that a little prettier. So I have some DNA like that. And maybe it goes around for a very long distance like a circle or something like that. I now want to copy that DNA.

So I have one strand, and I'm copying it. I have this other strand, and I'm copying it. And remember, these two strands are wrapped around, and around, and around, and around each other. One is going like this. One is going like that, and there's some wrapped around. And as I tug them apart to make a new strand, to synthesize a new strand, those two new double helices are so totally intertwined with each other. Every turn that there was in the double helix is now a twist and turn

connecting the two, sort of entangling the two helices.

So I have the two new double helices entangled with each other. Why is that going to be a problem? I'm going to send these to two daughter cells. These are the two genomes for the two daughter cells. In fact in particular, if this thing was a circle, the two new circles will be totally wrapped around each other with a gazillion wraps. No way they're going to two daughter cells.

Now, here is where mathematicians are very useful, because it is a theorem that if I take two circles wrapped around each other like that, there is no topological deformation possible that can separate them. It's like these puzzles, you get some strings wrapped around each other separate them. It's a theorem that two circles wrapped around each other like that cannot be separated unless, of course, you cheat.

What's cheating?

AUDIENCE: You cut it.

ERIC LANDER: You cut it, obviously. If you cut it, then you can separate it. But otherwise, it's mathematically impossible to separate them. So this could concern people. How could a cell do this? So what does the cell do?

AUDIENCE: It cuts it.

ERIC LANDER: It cuts it. It's got no choice, right? It's a theorem, right? Even cells can't violate theorems. So it cuts it. The only way to get these things apart is to cut it. Now, what it does, is it takes those double helices. I'll represent the double helix as a thicker kind of thing now. That was my double helix, this other double helix wrapped around it. It's got to cut it.

Now, when I take two DNAs that are wrapped around each other or two DNAs that are separate, have I done any chemistry on them? I'm sorry. Are they chemically different? They're chemically the same molecules. But they're topologically different. Topologically means wrapped around. In one case, they were topologically

entangled. In the other case, they're topologically separated from each other.

So they're still the same chemical bonds, the same molecules, but when I separate these two double helices now, the difference between these is that they are what are called topoisomers. They are isomers because they're exactly the same chemical formula. But they're topoisomers because they have different topology. They're not wrapped around each other anymore.

So it turns out there is an enzyme that just gets in there and makes a double stranded cut in one of the double helices, grabs the two ends, passes it around the other side, and ligates them back together, and keeps doing that until they're disentangled. Pretty clever. Cut, paste, cut, paste till it can separate those two double helices from each other. Remarkably, this enzyme is called topoisomerase.

This job is done by topoisomerase, actually, by topoisomerase II. There's a couple of different topoisomerases, and it's topoisomerase II that does this particular job, cuts and seals up that double-stranded break.

All right. It is amazing how this works. Let's take another problem in how we do DNA replication. So let's deal with fidelity. The fidelity, accuracy of replication. I have my strand. Which direction do we go? We go, for this template, five prime to three prime. This way goes five prime to three prime, the opposite direction there. I now add on. If this is a T, what do I add in?

AUDIENCE: [INAUDIBLE].

ERIC LANDER: If it's a GCGTAAT, et cetera. Why does the right base go in? Why does the right base go in? Yeah?

AUDIENCE: Hydrogen bonding.

ERIC LANDER: Hydrogen bonding. It's got that these hydrogen bonds. AT makes two hydrogen bonds. GC makes three hydrogen bonds. The wrong base could never go in. Sorry. In biochemistry, do you ever say never? No, we say K equilibrium.

We say how much more unfavored is it for the wrong base to go in? It's not

impossible, it's just disfavored, because it's energetically less good. How much energetically less good is it? What is the delta G for putting in the wrong base? It's not infinity.

It turns out that there is an equilibrium constant for putting in the wrong base, and that is K equilibrium is about 10 to the third for the right base, 10 to the minus third for the wrong base. Thank goodness. So only one time in 1,000 does it put in the wrong base. That's what that has to mean, right? If it's 1,000 times less favored energetically, it means you only make a mistake one letter in 1,000.

How do you feel about that for your own genomes? Is that a level of quality control you are satisfied with?

AUDIENCE: No.

ERIC LANDER: No. How big is a typical gene? Typical gene is, in terms of its protein coding information, you guys already know about DNA goes to RNA goes to protein. It's about 2,000 bases of protein coding information. That guarantees two mistakes per cell division. Not good. Two mistakes per cell division. That's not OK. That's two mistakes per cell division. That would be two errors per cell division, and you have a lot of cell divisions, you're in a lot of trouble.

So it turns out something more is needed. Quality control is needed. So later, it was discovered that the enzyme DNA polymerase, which has a five prime to three prime polymerization activity also does a second thing. That same enzyme, DNA polymerase, is also a three prime to five prime exonuclease.

What do you think an exonuclease is?

AUDIENCE: [INAUDIBLE].

ERIC LANDER: Take stuff out. So it adds bases in the forward direction, but it also goes backwards and takes bases out. Isn't that dumb? I thought we were trying to synthesize, but we're also unsynthesizing. With some probability, it goes backwards and takes out bases.

Turns out that the probability of taking out a base backwards is higher if it's the wrong base. It's proofreading as it goes as I hope you are. It's proofreading.

It goes backwards and takes bases out more often. Sometimes it takes out the right bases, but it is proofreading its work. And more often when it's the wrong base, it goes backwards, and so you get the benefit of a K equilibrium from the original base. And then there's a separate K equilibrium for the proofreading, and that helps you. And when you combine the proofreading with the original accuracy, now, we're down to something like 10 to the minus five or 10 to the minus six errors per base, per cell division.

It's only making on the order of one error per million. Now are we satisfied? No. You guys pretty hard nosed. Not good enough, because you have 50 cell divisions to make more and some cells go through many, many, many more cell divisions. Not acceptable. But it's a start.

So proofreading helps. So we have the fidelity of replication. Replication makes an error at a rate of 10 to the minus third. Proofreading brings you down to 10 to the minus six, and there's another process.

There are a set of enzymes that go around and feel the DNA double helix after it's finished, and if you put in the wrong base, the width of the helix is not right. The shape is wrong. It feels for mismatches.

So there is a mismatch repair system. Mismatch repair comes along, and if there was an error right here, the helix bulges out too much let's say. Mismatch repair cuts, removes some DNA, and gives the cell another chance to do it again. Mismatch repair gets you down to something in the neighborhood of 10 to the minus eighth, 10 to the minus ninth.

Let's say for the sake of argument, 10 to the minus ninth. You're genome is about three times 10 to the ninth. Now making that's one or two errors per genome, that's not so bad. Why do we care? Why am I bothering you with this? Who cares between 10 to minus sixth, 10 to the minus ninth? Big deal.

Well, a few percent of you in this class are heterozygous for a mutation in the mismatch repair enzymes. Don't worry. Your cells have the other copy that's good. But suppose one of your cells were to lose, by mutation, the good copy of the mismatch repair enzyme? And now that cell in your body had no copies of mismatch repair enzyme. What do you think is going to happen to your DNA replication?

Instead of being one in a billion, it would be one in a million accuracy. Turns out you have an extremely high risk of colon cancer. There are hereditary colon cancer syndromes that are due to inherited defects in the mismatch repair system. It is not at all trivial. Hereditary polyposis coli is due to a defect in this enzyme.

It matters. You've got to get it down to that level because otherwise, you're getting mutations that cause cancer, that is, when you lose both copies, if you lost both copies. Most of your cells would be fine, but if you'd lose the other good copy, by chance, that cell can go on to cause cancer. So this stuff actually matters.

Finally, finally, speed. Kind of fun to talk about speed. How fast does polymerase work? It turns out that polymerase is able to polymerize 2,000 nucleotides per second. That's very impressive to me.

It zips along at 2,000 nucleotides per second, installing the right base, getting it right only 99.9% of the time, proofreading as it goes, and gets the whole thing done 2,000 letters in a second. That is impressive engineering. That is really impressive engineering.

So that's kind of how DNA replication works well, except for one thing. Kornberg was a biochemist. Biochemists purify things in test tubes. He discovered an enzyme, Kornberg's polymerase. How do we know it's the enzyme the cell actually uses to copy its DNA?

See, I'm a geneticist. I look at Kornberg and I say, nice job. You showed me an enzyme that in a test tube is capable of polymerizing DNA. How do I know that's the enzyme that's actually doing it from the cell copies its whole genome? What does a geneticist want to see?

AUDIENCE: A mutant.

ERIC LANDER: A mutant. Show me a mutant then I'll believe. So someone went along and took E. colis one at a time because what else could they do. And for every single E. coli they grew up from a plate, they purified Kornberg's enzyme. And you know what they found? They found a mutant E. coli that lacked Kornberg's enzyme, and it could replicate its DNA just fine. What does that tell us?

Kornberg actually had the wrong enzyme. He still deserves a Nobel Prize for it because he got an enzyme that could copy DNA. It's actually not the main enzyme that does the job. Because we can make a mutant that lacks that enzyme and it can still copy the DNA, it can't be the main enzyme.

Turns out what Kornberg found was a minor polymerase that was used in those mismatch repair situations that would come along and do the tidying and clean up. The main enzyme turned out to be another enzyme, a more complicated enzyme. So my point about biochemistry and genetics both having to talk to each other, you only really know something when you have it from a biochemical point of view and the genetic point of view. The two have to go together.

Kornberg's enzyme is a great enzyme, it's a fantastic enzyme. It just happens not to be the main enzyme, and you can only know that by genetics. Of course, you can only purify it by biochemistry. All right. So that's DNA replication. Any questions about DNA replication before I go on? Yes?

AUDIENCE: [INAUDIBLE].

ERIC LANDER: Polymerase III or polymerase II, depending on the organism. They're all called polymerases. They're all DNA polymerases. They just get different names and numbers. Turns out most cells have multiple polymerases and Kornberg found the kind of simpler polymerase. The main replication polymerase also called polymerase but with a different number, is a different more complicated enzyme. Yes?

AUDIENCE: How does the enzyme know which one is the right..?

ERIC LANDER: how does it know which one is right?

AUDIENCE: [INAUDIBLE].

ERIC LANDER: Because 50% of the time you get it wrong. Do you know what bacteria do? What a great question. How would it know which one to get right? Know what bacteria do? They're very tricky. They mark their DNA, don't worry about this. They mark their DNA with methyl groups. There is an enzyme that comes along and put methyl groups at certain positions, but that enzyme is kind of slow.

So I have a methyl-marked DNA double helix. When I replicate it, the new strand is made, and what does the new strand lack?

AUDIENCE: Little methyl groups.

ERIC LANDER: Little methyl groups. It'll get them eventually because that slow enzyme will come along and put them on, but mismatch repair is fast. So what is mismatch repair looking for? The little methyl groups that are kind of breadcrumbs that say, this was the old strand, and this guy is the new stand.

It's thought of everything. It's really smart. Very, very smart.