

OK, so I'd like to go now to the next segment of the course.

Think you can probably appreciate little bit better this triangle I had on before about how what biochemists did was they tended to break cells open, look at the component parts, through other things in there, and then proteins.

But an awful lot of stuff having to do with function is proteins, and what geneticists, the discipline of genetics would do, which made mutants of living organisms, and that looked at how function was affected by mutating individual genes, how those were both very powerful approaches. Genetics told you what was really important, and biochemistry told you how it worked at a molecular level, but the real problem is knowing whether this thing you had doing something in the test tube was actually the one that did it in life.

And I think it'll do when Arthur Kornberg isolated the very first DNA polymerase, he was able to copy DNA. And he got a Nobel prize, and it was the first enzyme that could copy DNA, and then John Karyns made a mutant that was lacking the enzyme. And the organism was still alive.

So therefore, it couldn't have been the DNA polymerase that was copying the chromosome. It actually turned out to be a DNA repair enzyme. So, if you can actually unite genetics and biochemistry, if you can take a mutant that's broken in a function and you found your protein was missing, or vice versa, then you had a very, very powerful insight because you connect your knowledge of what was physiologically important to the biochemistry that you're doing. But it was really, really hard for years, and only in very rare occasions did some geneticists have a mutant that suggests so strongly that some biochemists would look, or some biochemists would have such a powerful result that they talked to a geneticist and seen if anybody had found the result.

And the power of recombinant DNA, although it started a biotech industry, and it made possible the sequencing of the genome, there's another level up one higher in conceptual understanding.

And what it did was, it let you go back and forth between here and here.

You wouldn't have any problem now if I gave you the sequencing of the gene, you can order it. You could stick it in a cell and make massive amounts of the protein. You purified it twenty-fold and it would be pure, whereas before you might have had to purified it fifteen thousand-fold out of 1,000 g of cells, and you would have had to been a very good biochemists with 15 steps in order to purify it. So, you can go from the sequence of the gene to the protein, or if we got a protein and we wanted to know which gene we'd just sequence a bit of the protein, use that genetic code, work ourselves backwards to some possible sequences, then go looking for the sequences, and then go find the gene. So, what recombinant DNA allows us to do is close that loop. You can go from genetics to biochemistry at back and forth. Now everybody does everything instead of it being isolated disciplines, which it was when I entered the field. So, all of the stuff depended on the development to clone particular pieces of DNA.

And I want to make clear right at the beginning, there's a couple of uses of cloning that are in popular usage right now. What we're talking about in this lecture is cloning a piece of DNA. What that means, is I'm going to take a particular segment of DNA, say, cut it here, and cut it there.

And I'm going to take that piece, and I'm going to do something to it that lets me amplify it and make many, many copies of that piece of DNA. And cloning of anything else you make a whole lot of copies.

So that's one use of the word cloning.

The other use, which you see in the popular press all the time, is cloning an individual, but not being there but you would take the nucleus from the cell of the individual, he put that nucleus into an egg that didn't have its own nucleus moved, and now you hope what you get out of that is an organism that has all the same genetic content as the starting individual.

And in fact, although it sounds very good in paper, is you're probably beginning to see it's not the panacea that people thought it was, or that we'd have to worry that in 10 years, all of my MIT students would be clones of the brightest person in the class or anything like that, because other stuff happens, because unless you go on to advanced biology courses, but there are modifications of DNA. There's all sorts of stuff that happens to it, so it's not identical.

And so many of these cloned organisms, like Dolly the sheep, that was famous, died early with, I forget, arthritis and things. So, there are a lot of problems on that score. But that's the other use of the word cloning and that's not what these next three lectures that we are talking about cloning a piece of DNA. And that was the big problem that faced the field, certainly when I was an undergrad and even when I was a grad student I was interested in synthesizing pieces of DNA, and it was one of those things that people said, why are you doing it? Well, because you could try to do it. What if you got a piece of DNA, like Gobind Khorana, who's my colleague, who got the Nobel prize for synthesizing the first gene.

He synthesized it. It was a tRNA gene that was 120 nucleotide base pairs long or something. He synthesized it.

He'd shown you could do it. But you couldn't do anything with it.

And there were sort of two big problems. One was the fact that this DNA, although it's not its a monotonous tetranucleotide.

It's pretty hard to tell. Each one of these things is a base pair, and human DNA has 3 billion of those. And a bit down here, doesn't look very different than the bit out there.

And it certainly wouldn't look very different than the bit of DNA that's 2 billion base pairs over on the other side of campus or something. So, there is no way to take DNA and cut it reproducibly, so you to get fragments.

What you could see from first principles was what you would need was magic scissors. And what would the scissors look like? Well, it would have to be scissors that could be sequenced because there's nothing else different. You know it's a regular backbone, and it's only four nucleotides. So, if you wanted to cut DNA in particular places, you had to have scissors that could see a sequence.

And furthermore, you can see they couldn't just, there are the hydrogen bonding parts of ANT or GNC because those are stuck together there in the middle of the DNA.

So, you'd need scissors that could somehow find a sequence and make a cut. And those were found. I'm going to tell you about those.

They're called restriction enzymes. So that was part of the thing. The other thing was, imagine I could cut out this fragment. And they gave it to you, and I said, great. Now I've got it.

Would you make me a lot of copies of this DNA?

Could you do that? Let's say, you now know how to transform the principle that we saw back with Avery.

We could take naked DNA and put this fragment into the cell.

Would it replicate? What do you think? Anybody remember?

No? OK, we talked about some other languages, right?

But one of the things that's in the DNA is the genetic code with all the genes. And we can find the reading frame. Remember when we talked about an origin of replication. I said that was sort of, at least for E. coli there's one origin. In eukaryotes the origins are spaced out along the DNA. And every time you have a round of replication, it starts with one of those origins and then goes.

So the chance of this piece of DNA by chance is going to have an origin is pretty small. So if I put it into an organism, it's going to sit there, if you're lucky, make it degraded because it's got [blends in?], or even if I made it into a circle it probably wouldn't replicate because it probably doesn't have the word in the DNA that says start a round of DNA replication.

So, the other overarching principle of DNA replication is you somehow have to take the fragment of DNA that you're looking at, and you have to attach it to an origin of replication.

Now, if you have an origin of replication and you have a fragment of DNA, and you put it in the cell, now you'll get a lot of copies of that piece of DNA. So that's what recombinant DNA is all about in a really, really simple form. I'm just going to take you now into increasing sort of levels of detail.

So, let me just sort of give you just sort of a really broad view of this cloning, and then we'll sort of start to dive in to some of the fancier techniques that have come out of this.

We'll talk about DNA sequencing, and PCR, and stuff in the next lecture. So the first principle here is to cut the DNA, and I know you may think this is sort of baby talk, but this is how I think. If you really think about this stuff, this is what it really is. With sequence specific molecular scissors, these have the rather odd name.

They're called restriction enzymes. I don't know if any of you know why they're called restriction enzymes, and although I'm sure that some of you have used them in your op to cut up pieces of DNA.

But what that does, these are enzymes, as I'll tell you, that recognize a particular sequence. And they always cut at that sequence. In the value of that is you can reproducibly cut DNA exactly the same spot, and the spots are specified by whatever sequence that particular pair of scissors knows how to read.

Then the next thing we have to do is we need to join the piece of DNA to an origin of replication. So the thing that carries the origin of replication is called a vector. And usually, not always, these are circles. We'll consider the ones that grow in *E. coli* are most of the time, or in bacteria, mostly circles.

They are the ones that are broadly used for most cloning.

So, we'll talk about those. What makes a vector? When it has to have is an origin of DNA replication. They usually have something else. We could call it a selectable marker, but something like a drug resistance.

If any of you have done cloning in a UROP usually it's something like a gene for making the cell ampicillin resistant, or tetracycline resistant, some antibiotic that would normally kill the cell.

So you can tell, does that cell have that vector or not because the cell starts at either ampicillin sensitive.

It acquires the vector that's replicating it, also acquires the gene that gives it the drug resistance.

But if we're going to cut that, if we're going to join a piece of DNA to that, we can't join into a circle without breaking it.

So, we need to cut the vector at a unique site. And we would use a restriction enzyme for that. And you can also see in designing a vector, you'd want something that only has one site.

So, what we would have achieved from this conceptually as we've now got this, this is the vector, its origin of replication here, and let's say ampicillin resistance, for example, as a selectable marker, the gene for that could be encoded here. And we have the fragment of, if you want to that down, just put it down on the floor, I think. We've made the point at this stage. Thanks.

What one has to do is to join this piece of DNA to that.

And we'll go to the molecular details of this.

But, we'll join the fragment to the vector, and actually this was something that was already in molecular biologists' toolkits, have been studying DNA replication. That's DNA ligase.

When we finish an Okazaki fragment, we had to seal [UNINTELLIGIBLE] and the enzyme that did that was an enzyme called DNA ligase.

So, molecular biologists basically had the scotch tape or the glue to join stuff back together. What they were missing for many, many years were the sequence specific molecular scissors.

So at this stage, if we were doing the recombinant DNA, we now have a vector. We now have a piece of DNA joined to it. In fact, we probably have a whole other mess of things that happens along the way.

But at the moment, they're in a test tube.

So, if we want to have this thing grow, what do we have to do next?

We are going to have to get the DNA from outside the cell inside the cell. That's the word; we need to transform the DNA into a cell.

Again, the word transform, that goes back to those transformation experiments with the *Streptococcus pneumoniae* going from smooth to rough, and you are taking stuff from the cell that transformed them from rough to smooth, whatever, that's where the word came from, but we now know it's getting naked DNA inside the cell where it can be replicated.

And then the next thing we need to know is, what cells have acquired this vector that at least as the vector.

We'll settle for that in the beginning, and to do that, you need to select for the marker on the vector. In the case of this one, we would start with a strain that killed my ampicillin, and then we just play it out and ask for guys that are ampicillin resistant. And, you can see that there is another class of problem because if we had uncut vector, and there would probably be, for sure, some of that in our mix, that would make the cells ampicillin. And if we had an

insert, it would also be ampicillin. So, if we really wanted to get into this, we'd have to do some more work to sort out what's on there.

But that's the basic stuff. I suspect most of you know this practically since kindergarten. But that's the overall framework into which, now, I'm going to start layering different pieces of detail. And the next part, again, some of you may know. I don't think it will be a totally foreign concept. You are probably familiar with this, that what are these restriction enzymes? The actual word is restriction endonuclease. They are often usually called restriction enzymes in a lab [parlons?]. Nuclease is something that cuts the nucleic acid, and endonuclease is one that doesn't need a free end. So, it can cut in the middle of the sequence instead of nibbling at the end. That would be an exonuclease. So, these things have names that tend to be something like ECO-R1, which has something to do with where they are derived from. And a typical one, one of the very first ones that is still in really wide use, is ECO-R1.

And this recognizes the sequence G A A T T C.

Now, you'll notice that if you read the sequence in this way it's the same sequence when you read it on the other strand.

It's called a palindrome but be careful because palindromes in English, those are words that you read from the front to the back; they're the same. In an English letter, it doesn't matter whether it's in A here or an A there.

But you guys know something about DNA structure.

There's a five to three prime polarity. So, reading this way doesn't look at all the same. It's totally different.

But reading in this strand, we say that's five to three.

The thing that's identical is the reciprocal sequence on the other side. So there is this, you see G A A and G A A but it's not like the English word palindrome, so get yourself mixed up about that.

Anyway, what this will then, what this enzyme then does, is it cuts to the side here. It cuts symmetrically.

And what it generates them is a G three prime hydroxyl.

Remember the ribose? If we have, say, an A there, this is the three prime position, and that's the five prime position in the sugar. So, it leaves a three prime hydroxyl, and is also then leaves a five prime phosphate.

So, we'd have A A T T C here, and then on the other side, we would get the reciprocal thing. So, we'd have G with a three prime hydroxyl, and then over A A T T C like that.

So now we've got a break here. We can pull those apart. But one of the nice things you can see right from this is that we're generating five prime single stranded ends, and this one is the sequence A A T T C. This is A A T T C here, and these guys, if they could get together and line up as they would here, they'd be able to form hydrogen bonds. So, if you take an enzyme like ECO-R1, and we took, let's say, a circle that had a single ECO-R1 site, G A A T T C, if we cut it with the restriction enzyme we would make [nicks?]. And if we kept a cold, all that we'd have is DNA nicks. And if we warm things up a little bit, there's only four hydrogen bonds that are holding that together.

So, the thing would linearize and just flop around in the breeze.

If we cooled it slowly, the thermodynamically most favorable state, the lowest energy state, would be with those ends coming back together. So, we could then add DNA ligase.

If we added these up and added DNA ligase, we could reverse the process and go back and forth, ECO-R1 to cut it, DNA to ligate it.

And then, the beauty of recombinant DNA is this rejoining part doesn't see what's out here or what's out there.

All it sees is the little ends that are generated by an ECO-R1 site.

So they take some of my DNA, and I'll cut it up.

I'll get a zillion ECO-R1 fragments, but they'll all have the same little overhanging bit that's complementary to the vector.

So if I take a vector cut with ECO-R1, and I take some of my own DNA and I mix them, I can get a little fragment, get in between the vector, and it does exactly that joining that I was diagramming right here. So again, it was the discovery of these restriction enzymes that made possible almost all the stuff that's happened in biology since 1975. The development of restriction enzymes was essentially, I was a postdoc at Berkeley at that point and the labs, Stan Cohen at Stanford, Herb Boyer at UCSF, and a two others around the country were working on this. They were almost all labs that had worked on bacterial plasmids. Plasmids are little circles of DNA, so the labs that started were ones who have been busy studying little circles of DNA that usually carry drug-resistances between cells.

And so that was happening while I was a postdoc.

And when I got to MIT and '76 the technology was just beginning.

I was one of the first labs trying to cut pieces of DNA and join them back together. So, it's a pretty recent

development.

At that point, DNA sequencing hadn't been invented.

The idea that you could pull out a piece of DNA and do something with it or produce a protein was just a thought. It didn't exist.

So it's hard to overemphasize how critical the discovery of these restriction enzymes were. Now, I just want to tell you where they came from, or how people found them.

And they'll try and do this quickly because I know some of you get impatient with history. But this is really important because it's very easy to make fun of basic research. You can ridicule anything pretty easily, and you might just ask because I'm telling you the story. Somebody proposed doing this.

I'm going to tell you the experiment that basically is the basis of the biotech industry, and would you have been smart enough to recognize that it was the discovery of a phenomenon called restriction, restriction in bacteriophage growth on bacteria?

And it was, here are actually a couple of EM's and these little plasmids. This is an electron micrograph one.

In these little circles it's been shadowed. And this is actually artificially colored, but that was the kind of plasmids that people were cutting up. So, as I said, trying to get through this DNA, and the stuff, what's made possible the sequencing at the Whitehead Genome Center and stuff that I'll tell you about is going on. I didn't really set this one up.

But that's Eric Lander who teaches 701 in the fall.

I told you a picture from that DNA 50th. Well, they had a banquet at the end of it, and I was there.

This is [Savandi Pabo? from Europe who is sequencing the chimp genome. And that's Francis Collins who is head of the entire Human Genome Project. This is Evelyn Witkin, who was a big discoverer of early DNA repair events.

And I put that one in because it was sort of interesting.

There was Eric, and Savandi, and Francis were talking about what would happen when they knew the sequence of the chimp genome, which wasn't done. And there was an advertisement, a poster advertising Jim Watson's latest book. And they ripped that in half, and were writing notes on the back all the way through dinner.

So if you want to see what scientists on the cutting edge, including someone who teaches 701 the fall looks like when they are not teaching 701, there is a picture.

So anyway, the discovery of restriction enzymes was Salvador Luria, who I've mentioned. He was a member of the biology department, and one of our Nobel Prize winners.

He started the cancer center. He also trained at Jim Watson, when I showed you that picture. This is Salvador standing over here.

Another thing that Salvador did, he was a Nobel Prize winner but he thought introductory biology. So I am basically following in the footsteps of Salvador. He wrote a book called, even though he was a Nobel Prize winner, a book called 36 lectures in introductory biology. And some universities, the intro to biology is taught by whoever is at the bottom of the food chain. The most junior professor gets stuck with intro to biology.

And here it's the other way. I mean you're getting Eric and Bob, for example, Weinberg to teach in the fall, tells you that.

And really where that comes from is the fact that Salvador Luria had such an interest in replication. So he trained at Jim Watson, started the cancer center here, and he also carried out this phenomenon of restriction. And to get this working with bacteriophage. And I know a [couple he wrote?

, he didn't like to see old guys on porches. So I got freaked out, and I took this next picture out for this morning. Oh, I've got to show it anyway for two reasons. This is Salvador sitting on a porch at Cold Spring Harbor with Max Delbrook who started, really, much of the work on bacteriophage that gave us the underpinnings of microbiology. And then put it partly on A because it shows the informality of the molecular biology culture which persists to this day, and also because Salvador had such an impish sense of humor. You would have really enjoyed had he been teaching this course. Anyway, Salvador was studying this bacteriophage. Remember we talked about it?

And they basically [with syringe? they injected their DNA into the cell. There's an electron micrograph. The DNA is up top there.

It goes in, and then the DNA takes over the cell, reprograms it, and makes babyphage. And I showed you how we make plaques.

So that was what Salvador was studying. And it's a little like what we are talking about with Mendel. He didn't have very many techniques available to him at the time. He couldn't sequence DNA.

He couldn't do a lot of things. But he could [plate phage?

and count, and things like that. And what Salvador was looking at, he had a bacteriophage, and he had two strains of bacteria.

I'll call them A and B. OK, here comes the experiment that founded the biotech industry. You ready? You going to fund me?

All right, so what I propose doing is I'm going to grow the phage on strain A, and now I'm going to plate on strain A and strain B.

I laid awake all last week thinking of this experiment.

So what did I get? I got a lot of plaques on strain A, probably something like 10^9 or 10^{10} per mil because that's what this phage lysate usually looks like once you've grown them up.

And if I plate them over on strain B, no phage, maybe an occasional phage. So, I bristly the bacteriophage can grow mine on strain A. It can't grow on strain B most of the time, but some variant has managed to figure out how to grow on strain B. Give our foray into genetics, I think many of you would think, probably as I suspect Salvador did at the time, the things mutated. It's learned. It's made some change in its genomes that's allowed it to grow on strain B. So, I wonder if it learned to grow on strain B, couldn't grow on strain A? So, basically what he took was the phage from that experiment, and then he plated them on strain A and strain B. And, well, as you might guess, since it was growing on strain B, lots of plaques, and there were lots of plaques over here. OK, so it didn't forget how to grow on strain A. So, better check over here, too, need a control experiment, so take this guy, plate it out, strain A, strain B, some plaques over here.

That's not a surprise going in strain A.

We are back to where we started from. It doesn't sound like a mutant, does it? And if it was a mutant, everybody should have been the same.

Instead, when you grew a phage on strain A, it didn't have the ability to grow on strain B. But if you gave it a chance to grow on strain B, most of them wouldn't make it.

But if it ever did, it had now acquired the ability to grow on strain B. So it could still grow on both.

But if you take somebody who had been growing, something had been growing at strain A, it lost it. And so, the idea there was then that it wasn't a mutation. Something was happening in the strain B that enabled it to grow on strain B.

And if it ever got away from that environment, [it?

lost it. So, the phenomenon was called restriction.

It wasn't a mutation. It was something else.

And it turned out, then, that restriction was due to an enzyme. An example of this kind of thing, then, would be this ECO-R1 activity that's able to cut at a very specific sequence. Now, if you are going to have a set of molecular scissors inside of you that could cut a G A T C sequences, you'd have a problem unless you did something else because every one of your G A T C sequences would be cut by the restriction enzymes.

So what the cells that have a restriction enzyme have, as a modification enzyme that recognizes the same sequence, and then modifies it in some way that makes it resistant to the restriction enzyme. And in the case of the ECO-R1, it puts on a methyl group on this A. You might have thought that that would interfere with base pairing, but it doesn't because adenine looks like this. These are the guys that do the base pair refining, if you look back, and you'll see that you could put a methyl group in there. It wouldn't interfere with the base pairing, but would allow this to go. So, it was the discovery of this phenomenon of restriction of bacteriophage grew on one strain, not on another. It could learn to grow on the other strain.

It could lose that acquisition. It was that phenomenon. People, didn't have any other reason other than it was an interesting problem in biology to understand it. Once they understood the basis of it, another whole world opened up because you could see from basic principles, now I could cut any piece of DNA. I could generate these little overhanging sticky ends. I could take a plasmid.

People knew about those. I could find one that only has one restriction site. I could stick things in.

Now I've attached an origin of replication to each of those pieces.

And I'm in business. I can now, for the first time ever, take a particular piece of DNA and make as many copies as I want.

And that was an absolute transformation to the way people were able to think about biology. So I'm going to just kind of give you an idea of how people would start. So the way people began and still began most things is they'd call it, the usual term is you call constructing a recombinant DNA library.

And there are a variety of different ways of doing this.

But this principle is the same. We'd take the DNA from whatever organism you're interested in, [and studying?].

And we cut with some restriction enzyme. And this restriction enzyme will cut wherever there happened to be sites. They might be close together. They might be far apart. But whatever, still generate some characteristic set of fragments. And we'll now have, in this case, fragment number one, fragment number two, fragment number

three, fragment number four, and so on. And if it's my DNA, there's a lot of fragments. And of course, they're all mixed up.

I can't tell where any of them are. They're just all mixed together in the test tube. Then we'll take that vector that we've opened. And now, we'll mix all of these fragments together with this vector. And then we join it just the way [that it's cut?]. And now what we'll get, it's a collection of plasmids that have different inserts.

So, one of the plasmids will have that fragment number one.

Another one of them will have fragment number two, number three, and so on. Then this whole thing is what's known as a library. You can see if it's DNA from me, there were 3 billion base pairs to start. Given the human DNA, G A T C sequences are pretty common. You can calculate the frequency yourself for how many sites on average there would be for a restriction enzyme within a piece of DNA and figure out roughly how the fragments there would be in the library of human DNA.

So, we are partway there. We can now make a library.

We can make it from bacterial DNA. We can make it from human DNA. But the next thing that people had to learn how to do was to figure out how to find a particular fragment that had the gene that you are interested in. And there's a whole variety of things. I mean, ultimately today since the human genome is sequenced, you go on a computer and a type and you find it because the sequence is all known. But the only reason we can do that is because of all the work that was done in between.

So, I'll give you several ways of doing this. But one of the ways I think you can see very easily, and it's actually going back to the term complementation. Remember complementation?

We had something that was mutant, and then we'd put in a wild type gene, and fixed it up again. So, for example, suppose I was studying histidine biosynthesis in E.

coli, and I wanted to find the gene that encoded the enzyme that I had just disabled in my histidine minus mutant. So, if I have a [hisoxotroph?], I'll call it a [his G?

gene, for example, that's one of the genes involved in making histidine. So, since it's a histidine auxotroph, if I have it on just minimal glucose plates, and I streak it out, it's not going to grow. But if I grow it on minimal glucose plus histidine, then it will be able to grow, right? So, I've got a variant of this organism.

It's got a single mutation in it that's affecting one gene.

And because I don't have that gene, I can't grow on minimal.

If I made a library of E. coli DNA, which is going to have a lot of fragments as well, and I took that library and I put it into this mutant, I'm going to get a big mess of things, all the different plasmids with all the different fragments that go into that mutant. How am I going to find the one that I want? Anybody see that? It's not that hard. I'm the mutant.

I can't grow on minimal because I can't make this enzyme.

Therefore, I can't make histidine. What do I need? How could you fix me up if you were a doctor? What's the gene we want out of here?

The one that makes that particular enzyme that makes histidine.

Yeah, take the whole library, stick it in this mutant. If the gene coming in encodes DNA polymerase, I'm not going to help this guy. It still won't grow in minimal, take a gene involved in making part of the cell wall would help.

But if I put in the, I get a fragment of DNA that includes the his G plus gene, and I put it in here, it's going to grow up. If it has the plasmid that has, or let's say the vector that has the his G plus gene.

So what you've done is a really, sort of, you've used that principle of complementation that some of you were sort of wondering about when we were doing genetics. So, you'd break a copy of a gene.

In the things we talked about, we bring in a whole chromosome that included in it just a wild type copy of the gene. With recombinant DNA, we can really narrow it down in the extreme. We can bring in a piece of DNA that is only the gene that's broken.

And we can take the gene back to the wild type. One thing, just to close, you'll see, if you remember back when I talked about language that are not universal, although the genetic code is universal, promoters and things are not. So I couldn't ever do this with a human DNA, could I, because it wouldn't get expressed. So, we need some other ways of finding those. We'll talk about those on the next lecture, OK?