

MITOCW | 7.014-2005-L26

So I just want to back up a couple because I think there's still some confusion about what a restriction enzyme is and exactly what it does, although I indicated, a three prime hydroxyl and a five-prime phosphate.

Let me show you. This is the way a lot, at least a large class of restriction enzymes, went. We've seen a deoxyribose backbone before with a phosphate backbone going on up to the next nucleotide. This is the five-prime position.

This is the three prime position. There is a phosphate here. It goes down to the next one. It goes down like this.

So what I'm showing you, GAATTC, five prime to three prime, here's where I indicated the first cut comes from.

And I indicated that the cleavage generates a G with a three prime hydroxyl, and the G ends up with a five prime phosphate.

So what that means is that the hydrolysis happens right there, so that after cleavage, what you end up with is a five-prime phosphate and three prime hydroxyl. This is paired with the C on the opposite strand. It makes one cut here, and then it would make an identical cut on the opposite strand.

So if we were to pull those apart this strand here would have G with its three prime OH, and the other strand would be TTAA.

And if we were to pull the apart. We'd have A A T T C, a three prime OH here, five prime phosphate there, five prime phosphate there. Again, you have to remember, this strand is going five prime, three prime in that direction. This was going five prime, three prime in the end. And the beauty of these restriction enzymes, at least it's not true of all of them. But it's true of a lot of them as they generate these what are called sticky ends.

You can pull them apart. They come back together, though, and reform those base pairs. It's almost like having little bits of Velcro at the end. And when this end is looking for a complementary sequence to pair with. It doesn't know what's out here, and it doesn't know what's out there, all it sees as this.

So I can cut this and rejoin it or I can cut them and pull apart, take another piece of DNA that's been cut with the same enzyme and therefore has the same corresponding little sticky ends on each end, and it could insert right in the middle. And that's the principle of cloning. And it was the development of these, if you will, magic scissors that made it possible to take this DNA which looks so homogeneous, nothing but GA's, T's, and C's, and then cut it up in defined ways. When I was a postdoc, a friend of mine had just purified \$2 million worth of EcoRI because he purified some of this enzyme. At that point, the only way to get these things was to produce them yourself and purify them. Now there are literally hundreds of these and they recognize different sequences, and once people understood that they existed, then they just started to look in different, usually they're from bacteria, and they just

looked in different bacteria until they found another one. And then they purified it.

So if you go to, say, any of the companies that do stuff for recombinant DNA, you'll find lists like this.

Funny little abbreviations like EcoRI usually have something that tells you some abbreviation related to the organism, from which the restriction enzyme was isolated. And you can find things that will cut, I won't say every sequence, but very, very, many sequences. There are literally hundreds of these, and you just order them. And the next day, a FedEx package arrives with a little bit of the enzyme that will cut at that sequence. Another concept that seemed to be a problem, was what's a vector? So if you understand that there are sequence specific molecular scissors, that if we have a piece of DNA and there's an EcoRI site here, you would come to think of it like that because it's going to cut in a slightly skewed way.

Maybe there's another one right here, another one over here.

If we take this DNA and cut it with this particular enzyme, we get a break here. Then this will get this piece running from here to here. We'll get this little piece here. We'll get this piece here, and will get whatever goes off on those sides. So that's naked DNA in a test tube. So I could cut any piece of DNA at some sites generating a bunch of fragments and if I just took those fragments and transformed them into E. coli, I took naked DNA, took it from the outside, put it inside, is it going to replicate? No. Why not? Because there's a special signal called the origin of replication that says "start replicating DNA here". This came from a piece of human, let's say some of my DNA, it would not have a signal in it that said to the E. coli replication machinery, "start replicating DNA right here".

So the principle of, apart from being able to cut DNA fragments, is you have together to replicate so you can make lots of lots of copies. The trick is to attach the DNA, at least a widely used trick is to attach the DNA to something that has an origin of replication that will work in the organism in question.

And that was what we call a vector. So this is an E. coli cell.

Another thing that's very confusing is all the circles that show up in this course. This is huge. The vector was double-stranded DNA that maybe, let's say, had a unique EcoRI restriction site in it. That's the only EcoRI restriction site.

The other thing that we'd need to have is an origin of DNA replication.

So, that's why this plasmid is able to propagate itself.

This little circle of DNA is able to propagate itself, and then some kind of selectable marker. And, most of the time that's a drug-resistance, so although it doesn't have to be.

And if we cut that here, generate sticky ends, then we can take this fragment and stick it in here, to give this piece here joined to the vector.

And that's an insert. Let's say it was that piece there.

In fact, if you wanted to clone DNA in *E. coli*, and then have the plasmid work in yeast, if you just take that plasmid that works, the vector with its insert that works in *E.*

coli and put it in the yeast, it won't replicate either. And that's because these other languages other than the genetic code are not universal. So, you also have to put also in a sequence that said to the yeast replication machinery "start something here". People call that a shuttle vector, something that will replicate in *E. coli* or replicate in yeast, and the same principle applies to other organisms. Okay, now probably the trickiest thing, the thing where I sort of muddled it on Friday, and I apologize for that, was this discovery of restriction enzymes. And again, some of you were frustrated. You said, why do I waste time?

Why not just tell you stuff that's on the exam? Okay again, people were talking about molecular scissors when I was an undergrad and grad student, and chemists were trying to think if they could come up with some way to get some specificity in how to cutting DNA.

And the answer, the discovery of restriction enzymes didn't come from that kind of experiment. It came from somebody trying to understand what seemed to be a really obscure piece of biology.

Julie made up a lovely little slide. But I think it basically did was I left at one of the little layers that I usually show.

So, let me just talk. Here's, again, what Luria saw when he was doing these experiments. So, I'm going to tell you now what was in strain A and B, and maybe that will help.

But I wanted you first to see it without knowing what anything beyond what Luria knew. OK, strain A has no restriction enzyme, and no modification enzyme.

And although there are different types of modification enzymes, many of them are methylases. So, we'll call that. And this one has a restriction enzyme. And, it has corresponding methylase.

Just to review that again, if you can figure this pretty much out from first principles that if you were an organism, and you had a restriction enzyme that would cut this, there are two things you can possibly do to keep from cutting up your own DNA.

One is to never have that sequence of pure DNA. That would prevent you from cutting up your own DNA, even though it has it. It's pretty constraining now, because somewhere in a particular protein, you might need that little

bit of sequence to encode something that you need to make a critical protein. So instead what you find is organisms have a restriction enzyme, have those sequences, but they don't cut up their own DNA because they modify their own DNA, by putting, in the case of this one, they put a methyl here.

And I drew that out the other day. You can see, you can put a methyl group on the exocyclic amino group of adenine, and not interfere with base pairing. But what you can do, is interfere with the way that the restriction enzymes sees that sequence. And when the cell pulls the DNA apart, each of the old strands is methylated, and a new strand is initially not methylated.

But that's enough keep the restriction enzyme from doing its thing. And then the methylase will come along, find a sequence, and then the progeny strand, the daughter strand, will begin to get methylated.

So, once you get DNA methylated, you can propagate it as long as you have the methyl group. So this was what was really underlying what Luria did. But he didn't know that.

Let me quickly just go through this again. So he grew the strain, the phage on strain A, no restriction enzyme, just plain old DNA. So, he picks a plaque that's probably about a billion phage in a plaque, somewhere around 10^8 , 10^9 probably, no that's not true, a little less than that, somewhat less than that, but lots and lots of phage particles in the plaque. Resuspend them and then plate them out.

And of course it grew on strain A. That's what it was growing on.

That wouldn't surprise you at all. The surprise was, even though he knew he had lots and lots of phage, hardly any of them grew on strain B.

But, he found a rare plaque that had learned to grow.

And he tested then, and this thing grows on strain B.

That was not a surprise because it had to grow on strain B to be up there. And he tested, and it still grew on the strain A.

So up until now, everything we knew, everything we've talked about in that course, you think, ah-ha, this original phage couldn't grow on strain B, but it's mutated. Somehow. It's learned to grow on strain B.

That would be a perfectly reasonable explanation.

But if that were the case, what would've happened?

Well it was growing on those phage, you pick them again, it should still grow on A and B, and they still do.

The problem with that model was that they grew on strain A.

They had trouble growing on strain B. Now that we've learned to grow on strain B, but you hadn't forgotten how to grow on strain A.

But if you take the ones that were growing on strain A, they grow on strain A, not a surprise.

This is the problem right here. If this was a mutant, permanent change, then we should have been able to have it grow on strain B as well. So what was really happening in there? Well, at the beginning, the phage DNA was lacking any kind of modification. It grew fine on strain A, because there was no restriction enzyme. When it went into strain B, that now had an enzyme, a restriction enzyme that cut up any time it found that sequence. And so, most of the phage that injected their DNA, those DNAs were trashed by the restriction enzyme. But there is a methylase in there that's also able to methylate those sequences. And what happened somewhere along the way was that there was the methyl, on one phage DNA, and got enough methyls on there that the phage could be replicated before it got cut up by the restriction enzyme.

Once the phage molecule has methylations on the site, it's able to grow just fine on strain B, and it will be able to do that forever. However, if you take that DNA with the methyls on it, and we put it back strain A, it'll still grow, as this one doesn't have any kind of restriction enzyme, but while it's growing its busy losing all the methyls again.

We are right back to where we started from. An obscure experiment, one of the most obscure you could get, many people would have paid no attention. It does not seem worth it.

So, the phenomenon was called restriction. They had to give it a name. It wasn't mutation. They said this phage DNA was being restricted somehow when it grew on strain B. People called them, postulated there must be a restriction enzyme that was doing this restricting of phage growth. When they found out what was doing it, they had discovered magic scissors that would cut DNA at particular sequences. So, the point here again, to try and go back and do this, I hope some of you will get this anyway, many of the really important discoveries come out of basic research. They are easy to ridicule. Why would I spend money on cancer, human disease or something, for somebody studying some little weird phenomenon about phage. But if you want to trace back to the experiment, it sort of started the biotech industry. It was the discovery of restriction enzymes.

It took a little while to discover what it was, but the reason they were discovered, were people were trying to understand that phenomenology. Once we got restriction enzymes, we already had ligase, which is sort of the tape we'd need, and just to show you here, when we go back to this one, you can see that if we put these together again, we have a three prime hydroxyl, and a five-prime phosphate. That's what DNA ligase knows how to do,

because that's how you seal up the end of a Okazaki fragment. So, that particular part of the molecular biology toolkit was already known to molecular biologists who had been studying DNA replication.

OK, so, if we took some DNA from anything, and we cut up into pieces like this, and then we join them with a vector that had been cut, so let's just sort of open it up a little bit, this fragment would go in here into one vector molecule.

This fragment would insert in another vector molecule, and so on and so forth. Then we would have what I said was a library, and the problem at this point is, so you transform those into *E. coli*, and now we have a whole series of *E. coli*. They have their own chromosome, every one of them because they still have to be a bacterium. So, let's take three members of *E.*

coli from this library, and they will all have this vector, but they'll have, let's say, insert number 1, 2, 3. This insert is a little. This insert is bigger, and so on. If we did it right. We have every possible fragment of DNA from the original source sitting in its own vector.

And the whole collection of *E. coli* in this population in the certain library, and the next part of the trick was, how do you find the thing you want, especially if you take my DNA with 3 billion base pairs. That's an awful lot of restriction fragments no matter what you do. How do you go about doing it? So the experiment that I showed you at the end of the lecture, cloning by complementation, is fairly simple, and it was basically one of the first methods that was used to find genes in recombinant library. And that would be, for example, something that had thisGene mutation in the chromosomal DNA.

This is what the situation I'd described the other day.

So, if we put the library into every cell, such that the bacterium we transformed the library into was broken for the *hisG* gene, and that mutant couldn't grow on minimal medium unless we put in added histadine. But, if one member of that library had the wild type, *hisG* gene, let's say it was this one here, maybe it had several genes on it.

But let's say over here. We had *hisG*⁺, then the strain is back to being able to synthesize histidine because it's got all the enzymes. What I was pointing out was this really is complementation, just like we did in that phage cross. We've got one broken copy of the gene. We've got a good copy. And all you need is one good copy, and you're back in business. What I was saying at the end of the lecture was this is not a general solution, though. If I wanted to find the corresponding histidine gene from my DNA, and all of these biosynthetic pathways, pretty much they rose so early in evolution, the biochemistry is essentially identical in all cells.

Can I use this approach to find the same gene for my DNA?

What do you think? Why don't you turn to somebody beside you and see if you can talk for a minute, and then let's see if, I can think of at least a couple of problems. Let's see if you can come up with one or two of them. Find somebody near you and see if you can come up with anything.

Anybody want to volunteer? An idea of why it would not work?

Or some of you think it would? No ideas? God, it's Monday.

[LAUGHTER] I feel like most of you guys.

Somebody, come on. What do you think?

It's going to work? No idea? What has to happen for it to work? I'll give you a vector that has my gene corresponding to that enzyme. It's in *E. coli*. I need to make the protein.

Yeah? No, it's in the vector. We cloned it into an *E.*

coli vector. So, that's got it. Yeah?

Well it'll have a language that will say "start transcription", but whose language is it going to have? It's going to have my transcriptional stuff. Will that work in *E. coli*?

Even though the open reading frame is fine, that's good.

How about translation? I didn't even tell you about that.

There is actually some specific stuff needed. That's not universal, either. So when you get the RNA you still have to translate it. There is another thing that might mess us up. Do you remember anything else about, yeah? Introns and exons? What if my gene has introns in it, which it almost surely has? We have to get rid of those.

E. coli doesn't know what they are. It's not used to taking them out.

You see the issues? Although that's a cute thing and that helps you find a gene from *E. coli* by complementing *E. coli* mutant, or maybe you could do it with yeast if you had a factor that would replicate in yeast, it wasn't a general solution.

So people had to use a whole variety of different ways.

Here's another way. You know that you have that genetic code. That was worked out years ago. So let's say I was a biochemist, and I'd found a protein that I was interested in, and I purified it, and they got it out to single protein,

and then I could cut it up with things that proteases that will cut the protein into pieces, and there are ways of sequencing protein. I'm not going to tell you how it works in this course. We just don't have time.

But you can get the sequence of little pieces of protein.

And let's imagine that this was the sequence of part of the protein that I purified. It's one of my enzymes [SOUND OFF/THEN ON] and I'd like to find the gene. Well, how could I use that information to figure out where the gene is in this library?

So here's the strategy. We get out the genetic code, which Gobind Khorana and Marshall Nirenberg helped work out, and we say OK, alanine, and if you look it up, what you'll find is that it's GC, and then it can be A T C or G. It can be any of those.

If we look up what the codon for aspartate. We'll find that there's a G or an A, but it could be T, or it could be C. We look up lysine; it'll mostly be A, but it could be A or G. You'll notice the variation of those things is almost all in the third codon if that hadn't struck you. Same thing with threonine: A, C, and this is another one; therefore codons that encode this.

And this one asparagine is that. So knowing that piece of the protein doesn't define a unique sequence. Though, what we could do it is we could synthesize what's called a mixed probe.

And that would mean when we are going to synthesize this DNA and we'd start with a G building block. And then we'd add a C building block. So, now we made G and C. And at the next step, we'd add an equal mixture of A, T, G, and C. So, what we would get out of that would be we'd be getting G, and then the next biosynthetic step would give us G, C. And then the next biochemical step would give us GCA, GCT, GCC, or GCG.

At the next step, we'd add a G. So everyone of these would get a G. Everyone would get an A, and then the next step, they would branch. And if you follow that out, you'll see by the end you have a mixture of probes.

One of them is going to be the right one that you find in the DNA.

Now if you work out the number of possibilities, you'll discover that most of the time there is only going to be one probe that's unique. Once you get to about 20 nucleotides, any sequence, on average, is represented once in the human genome. So as long as you make the probe long enough, one of the things in your mixture will be a defined probe.

So, what we can take is we have all these different pieces of DNA that are the logical variants you can see here.

And then we would label the probe with P32. It's a radioactive isotope, and it's very easy. You can add it to a five-

prime phosphate. There's a special enzyme that will very easily take the terminal phosphate from ATP and put it over. It doesn't really [SOUND OFF/THEN ON] for this course how we get it there.

But we can do is radioactively label the probe. So now we've got this mixture. And, somewhere in this library is a piece of DNA that's going to have the gene that's encoding the protein that we're interested in. So, how do you go about trying to deal with that? So, what we'll do is we'll plate our E. coli library onto a bunch of Petri plates.

So, I won't put too many colonies on here, so we can sort of see a pattern. But, we'd have probably a lot of them, and we'd have a bunch of plates. You can work out statistically how many plates you have to have to have a chance of finding your gene of interest. Then what we do, is lay a membrane on the plate.

It's a particular type of membrane, and what that will do is it will make a copy of everything that's there, and we're going to save the plate. And then we're going to treat the membrane, OK? We've got a membrane that's got an identical pattern.

They've got some of the bacteria from the colonies stuck at the corresponding parts on the membrane. We're going to lyse the E. coli; that means break them open so all their insides spill out.

We will denature the DNA by treating with a condition.

You can, for example, vary the pH and make the strands come apart.

That gives single stranded DNA, "ss" I'm using as an abbreviation for single strand of the strands you pulled apart. And, this sticks to the membrane. So, now we've got, at every one of these little positions on the membrane something that looks like this. Here's the membrane, and there's some sort of single stranded DNA that's stuck to the membrane in that fashion. The DNA that's stuck here came from the bacterium here that had a particular insert.

Over here, we have all the E. coli DNA and the vector DNA, but we will have a different piece of DNA in the vector.

Everybody with me? OK, so if we were now to take our radioactive probe that we made up there and get the conditions just right, that single stranded probe will come in, and it will try and find its complement. It'll form hydrogen bonds because that's the lowest energy well. And, we think about it thermodynamically. And, if we can get it right, the temperature and the conditions right, nothing will stick unless it's an exact match to the sequence.

And, if we get the right probe that can form hydrogen bonds with everything on here, and it's got P32 at this point, what will happen is we'll have now, the probe will stick, say, to this particular colony, now with radioactivity right there.

So, put some photographic film over the membrane, and right here there's P32, and that'll expose the film and nowhere else. And then, when we develop, what we'll find is one, if this works well, anyway, one spot.

So, now we know that that piece, that colony had a piece of human DNA in it that was related to the sequence from the protein that I had purified. So we go back to this colony, I think our things have probably migrated around a little bit here. Let's move this up just a little bit, and make it a little better.

So this one is this one. So, now I can go back to this colony and pick it out. And, let's say it's this insert.

So, now I've found I can sequence the rest of that piece of DNA.

We'll talk about how we sequence DNA in the next lecture.

So that's an alternative way of identifying a clone of interest.

There was a particularly painful way of finding a gene in a library that we for the most part do not have to do any more.

It was called positional cloning. And for example the gene that when it's broken causes cystic fibrosis, it's a very difficult disease.

Humans who have cystic fibrosis have a very tough time.

So there's a great deal of interest in finding the gene that was broken in these patients. Human geneticists, I showed you something about pedigrees, they would have a chromosome. They might have banding patterns, and they would have figured out that somewhere along the chromosome that the gene for cystic fibrosis lay somewhere between two genetic markers that they had identified. Now, the amount of DNA between something that you knew that the gene was here, and something knew that the gene was there, could be huge.

It could be many, many, many times the size of the E.

coli chromosome. So, what people would do is they'd clone something from here, a little piece of DNA from there, and also clone something and they get a little piece of DNA from there.

And then they go into the library, and they try and find something that had this DNA and something that extended in this direction the little bit. By the sort of thing you'd have marker A somewhere in the middle of the cystic fibrosis gene but you didn't know exactly where.

You'd clone a little piece of DNA and use that to find another one that overlapped with it. And that you'd find, use

that to find another piece of DNA. You'd walk your way over this way, and you'd start the same process at the other end.

And, every one of these things, the same kind of operation that we've got here, so cycles, and cycles, and cycles of acquiring the next adjacent piece of DNA, and working your way along here. And you had to use more than one different restriction enzymes, otherwise you wouldn't be able to get these overlaps. And by doing that, eventually they were able to get all the DNA that was between these markers. They knew the cystic fibrosis gene was there from the maps they had made by studying human pedigrees. And then, once you knew the sequence, then you'd take candidate genes, and you'd take a bunch of cystic fibrosis patients, and you'd start to see if every person who had cystic fibrosis had a mutation in that gene. And, eventually they got it.

That process in the case of cystic fibrosis took five years to do that with a huge team of people. And, I guess it was from 1985-1990.

If you wanted to do that experiment today. We come now to one of the most widely used ways to finding a gene of interest, and that is you go to the computer. The whole human genome sequence is in there. If we were to do that experiment today, we'd say, well, I know what this gene is, so you look that gene up in the database. And you knew what this gene was and you look up the database. Then you just look at all the DNA that's in the middle.

And you'd see a whole series of open reading frames.

And you'd probably say, well, what do I know about the biology of cystic fibrosis? Could I make a guess? Is it a membrane protein? Is it not a membrane protein?

And, there are certain characteristics that would probably allow you to make a guess. And then, you could jump right in and start sequencing DNA. You could start the experiment practically that afternoon instead of five years later.

So, if you look back from the literature, you'll find some of the key genes; in fact, in human biology, we're isolated by this very painful process of positional cloning, and you hardly ever have to do that now. It may be the odd case where something's needed, but most of the stuff now, there are these amazing databases, and I'll give you the URL for it at the beginning of next lecture. And I'll show you when, this is an experiment in which someone took this gene for cystic fibrosis. It is a membrane protein, and it's one of those proteins that mediates the passage of chloride ions across the membrane.

And, if that gets broken, you end up with cystic fibrosis.

What someone has done here is they've taken that green fluorescent protein gene. And, they've fused it to the

end of the cystic fibrosis gene. So, you can tell where the cystic fibrosis gene is localized in a lung cell by looking to see where the fluorescence is. And, I think you can see that the fluorescence is out there along the membrane.

OK, so at the beginning of next lecture, I'll introduce you to how we take one of these recombinant plasmids, and make what's called a restriction map. It's using a very simple, little piece of apparatus like that, and we'll go in and tell you about DNA sequencing, and this PCR technique you've heard, preliminary chain reaction that you've heard so much about, OK?