

PROFESSOR: So, first step, we need to cut our DNA. Step one, cut, which is going to be DNA restriction enzymes.

It turns out, quite remarkably, that if I have a sequence of DNA, five prime A G C T A G A A T T C T T A C C three prime, and we'll come backwards filling in the sequence. It turns out that molecular biologists are so cool that they invented a protein that is able to recognize that six letter sequence, G-A-A-T-T-C.

And it is able-- actually it's G-A-A-T-T-C on this strand, but what is it coming back? It's G-A-A-T-T-C. It's the same thing. So actually this is a palindrome. That's kind of nice. It's a palindrome-- it's a reverse palindrome. It's the same word spelled backwards on the other strand.

So what it does is it cuts it like that. And what it produces is a DNA molecule like this and a DNA molecule like that, that's mostly double stranded, but has a four base pair overhang. The overhang reads T-T-A-A here. It reads A-A-T-T there. And remember this is five prime to three prime, five prime to three prime. And there you go.

This guy has its little phosphate at the end there. This guy has his little hydroxyl over there. And it cuts it. Now that is an incredible piece of engineering. To come up with a protein, to devise a protein, that is able to recognize those six bases and cut at those six bases. And cut in just this way making a really clean overhang here. It's this cool five prime overhang.

Who do you think invented this cool protein? What engineer came up with this cool protein?

AUDIENCE: MIT engineers.

PROFESSOR: MIT engineers, yeah. Not a chance. Not a chance. This is a really tough feat. This is something that can only be done by the smartest engineers on the planet. And MIT engineers are unfortunately only the smartest human engineers on the planet.

Who came up with this is E. coli.

AUDIENCE: So you found it somewhere in nature?

PROFESSOR: Sorry?

AUDIENCE: You found it somewhere in nature?

PROFESSOR: Of course you find it somewhere in nature. Almost everything important that we say molecular biologists have come up with, it means molecular sat at the feet of the true masters, bacteria, and learned from the true masters. This protein is found in nature. And it's found in E. coli.

In fact, it's found in E. coli strain R. And it was the first such protein found in E. coli strain R, so it gets the name EcoR1. And it cuts the DNA like this. Pretty cool. Pretty cool.

Now, it turns out that E. coli has this EcoR1. How often does E. coli-- so whenever I take EcoR1, this protein, purified from E. coli, and I add it to DNA it always cuts at this site, which we call an EcoR1 site. How frequently do we expect, what's the frequency of EcoR1 sites? G-A-A-T-T-C, how often will that occur at random? One in--

AUDIENCE: Two to the sixth?

PROFESSOR: One in two to the sixth? How many letters do I have?

AUDIENCE: Four, oh, four to the sixth.

PROFESSOR: One in four to the sixth. My frequency should be about one in four to the sixth, which is about what? What's four to the sixth? It's two to the 12th. It's about 4,000. It's about one in 4,000 letters. One in 4,000 bases.

So it's very convenient. One in every 4,000 bases it'll roughly cut. It'll cut at roughly one 4,000 bases.

Why doesn't E. coli cut its own DNA? If it's got this protein floating around in its cell,

why isn't it chopping up its own DNA? Doesn't have G-A-A-T-T-C? Yeah, the problem is it's so frequent. That'd be really hard to make sure-- I mean, E. coli has 4 million letters in its genome. This should cut every 4,000 bases. You expect about 1,000 such sequences. It might be hard to arrange not to cut-- not to have any such sequences. It's a good idea, is not to have any, but an alternative--

AUDIENCE: [INAUDIBLE].

PROFESSOR: It protects them. It turns out E. coli, instead of avoiding the sequence altogether, has another trick up its sleeve. E. coli protects this sequence whenever it occurs. So it turns out that whenever you have a stretch of the E. coli genome that has this G-A-A-T-T-C in it, what E. coli does is it puts--

I'm just writing M-E here for a methyl group. Right, C-H three up there. It puts a methyl group-- I'll write C-H three. There we go. It puts a methyl group on the A, that middle A.

Well, that is a cute trick that E. coli uses, putting a methyl group there. Because what happens is, when there's a methyl group, right at that position, the enzyme no longer recognizes and no longer cuts there. So that's kind of clever.

E. coli makes this protein that can recognize G-A-A-T-T-C, but it has a second protein that puts methyl groups there. And this protein happens by accident to be called a methylase. It has a methylase. And the methylase protects that sequence.

So now, this is really cool engineering, but kind of dumb. What's it doing there? It has something that cuts the sequence and it protects the sequence, why bother having this? Yeah?

AUDIENCE: You can use it to cut it at places to unwrap the true strands.

PROFESSOR: That's an interesting idea. We could use it to cut our DNA and open it up to unravel our true strands. It's a thought. Yes?

AUDIENCE: To protect the bacteria from viruses?

PROFESSOR: Protect the bacteria from viruses. How do you protect yourself from viruses? Well, you have an immune system with immune cells and antibodies and all that. Does E. coli have an immune system? Why doesn't it have immune cells? Because it's like one cell. How's it going to have an immune system, right?

So suppose E. coli gets a cold. Suppose it gets infected by a virus. How's it going to protect itself? Cut at a frequently occurring DNA sequence. Now the virus, of course, isn't methylated there, bingo.

That's how it tells its own-- you can tell cell from an invader. E. coli can tell cell from an invader because it's methylated its own G-A-A-T-T-C sites, but the virus isn't methylated there. Way cool. This is an immune system for E. coli.

Now, it turns out-- so this is protection. These restriction enzymes protect E. coli from viruses. It turns out that E. coli is not alone in this clever trick. It turns out that other bacteria have also thought of this trick.

So it turns out that there is another restriction enzyme that cuts at G-G-A-T-C-C. And on the other strand it goes G-G-A-T-C-C. It, again, cuts in that distinctive pattern. And it's called BamH1.

And there's another guy. And he cuts at A-A-G-C-T-T, A-A-G-C-T-T. And it also cuts like that. And it's called HindIII. And there's some that cut at G-A-T-C, just the four letter word. And they cut like that. And there's some that cut at C-A-G-C-T-G, C-A-G-C-T-G, and this, cuts smack in the middle.

In other words, there's a wide number of different tricks. Some cut at six bases. Some cut at four bases. Some cut at eight bases. Some cut leaving an overhang. Some cut smack in the middle. Some cut leaving the overhang in the other direction. Some allow a degenerate base in the middle. It doesn't care which base is in the middle.

There's a zillion different solutions that bacteria have come up with for their immune system. And so, if I want to cut up some human DNA all I need is say, this protein EcoR1 or BamH1 or HindIII or MVL 1 or PVU 2 or et cetera. And I can do that by

growing up *E. coli* and purifying the protein.

And if I wanted HindIII, I would grow up *Haemophilus influenzae* and purify the protein. So in a molecular biology lab, today, if you want to cut up human DNA, you could grow up some *E. coli* and purify EcoRI or *Haemophilus influenzae*.

And that is indeed what ancient molecular biologists did in prehistoric days in the 1970s and 1980s. They would purify their own restriction enzymes. They're still alive today. You can talk to them. There are many of them on the faculty. And they'll tell you how it put hair on their chest to be able to purify their own restriction enzymes.

What do you do today? Order it online from the catalog, right? You know, there's the catalog, the New England Bio Labs catalog. Let's see what we got here.

Restriction enzymes, modifying enzymes, polymerases, all right, EcoRI, sale on EcoRI right now. \$100 buys you 10,000 units of EcoRI. It's in the catalog. You can go online. You can order it. You can have it tomorrow by FedEx. So, but that's how it works. It's in the catalog. So you can get any restriction enzyme you want to cut DNA anywhere you want to.