

Good morning. So, I see we have a lot of parents here.

How many parents have we got here?

Welcome to the parents. How many of the parents have done the reading for today? [LAUGHTER] Good, because we'll call the parents too, right? We'll see what happens.

All right, so, where are we? We've talked about this diagram that I keep coming back to. If you want to study biological function the two traditional ways to do that were to look at genetics or to look at biochemistry: genetics, the study of an organism with one broken component, those components being genes; biochemistry: the study of the purification of individual components from an organism away from the organism, particularly the most important such components being proteins.

What do they have to do to each other?

The unification in molecular biology that occurred in the middle of the century from the 1950s into the '60s and really up to 1970 or so, we came to a conceptual understand that genes encode proteins, and therefore these two different ways of looking at the organism: organism minus a component, components minus an organism were complementary points of view, and in theory, you could go from a gene sequence to a protein Sequence, a protein sequence back to a gene sequence, to go for a gene sequence to its function, its function to a protein, except for one to a tiny detail. This was all just conceptual. Conceptually we understood by about 1970 that the DNA made the RNA made the protein. The protein carried out the function but as of then, you couldn't individually work with or purify the DNA corresponding to any particular gene. All of the inferences had been indirect inferences: indirect inferences from bacterial genetics, bacterial regulation or Meselson-Stahl experiments, and all sorts of interesting indirect ways working out the genetic code, but it didn't let you read anything.

This was a problem. Some people in the late 1960s said, great, molecular biology is over.

We understand in principle how life works. Now let's go understand how the brain works. And there was an exodus of some people from molecular biology into neurobiology to now go nail the brain, figured that would be worth another ten years or so. But in fact, remarkably, people began to focus on how you could get to work with individual specific genes. Now, what's so hard about that? I mean, it's not very hard to crack open a red blood cell and purify different proteins. You can purify hemoglobin.

You can purify different enzymes. Biochemistry allows you to purify different components from each other. I want to purify an enzyme: let's crack open a yeast cell, separate the proteins over some column that separates them based on their size or their charge, and I'll get purer and purer fractions. I'll assay each fraction to see which one

has the enzymatic activity. But basically I use the physical chemical properties of the proteins to separate them into different buckets. Why not do that with, say, the human DNA and purify out the gene for beta globin, that encodes the beta component of hemoglobin?

What would be the problem of just using physical chemical purification to purify one human gene from another? Well, I mean, it's one very big molecule. Well, I could shear it up.

Maybe I'll just break it up. Now, let's purify the beta globin containing part. It all looks the same. It's just DNA. It's one chemical polymer with pretty boring properties, and they're not very different.

Any particular DNA sequence in any other DNA sequence basically about the same molecular weight, same charges, there's nothing to separate them by. How are you going to purify beta globin? That was the problem. That's where recombinant DNA came in was recombinant DNA was a remarkable and totally different way of purifying individual components. And the basis of it was this notion of cloning.

If I want to purify out from the human genome, how big is the human genome? The human genome is about three billion bases long. If I want to purify a particular gene, let's say beta globin or some other gene, typical gene, might be on the order of 30,000 letters long.

This is one part in  $10^5$  purification I've got to achieve.

Any given gene is only one part in  $10^5$  of the human genome. And then, what about a typical mutation? Maybe the mutation that causes sickle cell anemia by changing a single nucleotide and beta globe, well, that's one base pair. So that means I'm trying to identify something that's on the order of one part in  $10^9$  actually, a little less than one part in  $10^9$  of the whole genome.

Carrying out purifications like: really kind of hard to imagine. But the way it was done was by the invention of cloning.

Let me briefly overview of the idea of cloning, and then we'll dive into the details. The idea of cloning was, the way to purify individual molecules would just be to take the molecules and just dilute them so that there was only one of each model. That's very pure, isn't it? The problem is it's not very much, so you need a way to take a single copy of a molecule, and then make many copies of it. So purification's not hard. You just dilute it down so you work with single molecules but then you need to copy it back again and again and again, and no biochemical technique involves, say, fractionating a cell and replicating some enzyme, you know, copying some enzyme.

You can't copy enzymes, but you can copy DNA, and that was the basis of it. So here's the way it goes. The basic overview we'll look at is take your DNA and cut your DNA of interest, maybe the human genome, into pieces at

defined sites Then, paste your DNA, which is more technically ligate, the word we use.

Paste your DNA to some other DNA called a vector. So, cut your DNA and paste your DNA. Each piece of your, say, human DNA gets stuck to some piece of vector.

Insert this DNA into vectors that can replicate in bacteria.

So, I'm going to actually take my piece of human DNA and not just ligate it to any piece of DNA.

I'm going to take my human DNA, and I'm going to ligate it to a vector that has all of the machinery, all of the ability to be copied in a bacteria. Then what I'm going to do is I'm going to transform my DNA into a host cell, a host bacterial cell. Transform means introduce. When we talk about transforming DNA, we're not talking about changing it. It's the word that's used for taking my DNA, stuck into a vector, and introducing it into bacterial cells. Ideally, each bacterial cell would carry one such DNA molecule, and then what I want to do is I want to plate my cells, and select those that carry human DNA, my DNA; DNA I've put on it. So, I'm going to put them on a Petri plate and I want only the bacteria that happen to have picked an individual piece of human DNA to grow. So, that's the trick. It's a very simple trick. Take total human DNA, cut it up into pieces, glue it to a vector that's able to be copied so that it's able to be replicated in bacteria, put the vectors into bacterial cells; every bacterial cell picks up no more than one vector.

You plate it out, and you simply arrange so that the only cells that grow are those that picked up the piece of human DNA.

And then, every one of these colonies is the descendent of a single bacterial cell that picked up a single human molecule, but is obligingly copying that molecule for you again and again and again and again.

And thus, you have what we refer to; this whole collection here is called a library of clones.

This is called a recombinant library because every piece of the human genome is somewhere in here.

You know, this one here probably is active, and maybe this one here maybe is collagen-11 and that one there might, ah, there's beta globin. OK, actually when you look at the plate there's no way to tell but in principle they're all there. So, there will be this question of, how do we look at a library and pull out what the right one is? But somewhere in there should be a bacterial colony that has pure beta globin gene, the DNA for beta globin.

The next lecture will be about how you actually find it.

But today let's just build this library. So our goal is to be able to build a library like this.

So, we have to figure out how to cut DNA, paste, DNA, vectors, etc., etc. So that's what our subject will be today. Let's dive in. First, cutting DNA, how do you cut DNA? Restriction enzymes, etc. It turns out that the way you could cut DNA at particular places is as follows. Let me take a piece of DNA.

Here's a double-stranded piece of DNA. We'll go A, G, C, T, A, G, A, A, T, T, C, T, T, A, C, C, hydroxyl there, three prime. Let's go back on the other strand. What do we have? G, G, T, A, A, G, A, A, T, T, C, T, A, G, C, T, hydroxyl there, three prime. There's my double stranded piece of DNA. It turns out that there exists an enzyme that recognizes that exact sequence: G, A, A, T, T, C. The enzyme goes by the name Eco R1.

This protein, this enzyme, scans along the the DNA, and it finds this sequence: G, A, A, T, T, C.

Actually it's on this strand. What about on the other strand does it say? Same thing. But it's a reverse palindrome.

It's symmetric. That's very good.

And it turns out most restriction enzymes do that. OK, so what it does when it finds that, with the benefit of colored chalk that has just shown up here is it cleaves the DNA fragment like that.

And what it gives you then is a broken double strand with an overhang, T, T, A, A, five prime, three prime, three prime, five prime. This has a hydroxyl here this. This has a phosphate there. And then this other fragment here is A, A, T, T, C, T, T, A, C, C, G, G, T, A. So, what happens is, and this has a hydroxyl five prime, three prime, three prime, five prime I get into two fragments of DNA that have been broken there and have it over. The overhang is complementary.

Those two sequences match each other. There's what's called a five prime overhang and they're complementary. So, we have complementary, that is matching, five prime overhang. This is called Eco R1 because it's purified, this particular enzyme, from E coli strain R and it's the number one such enzyme that was purified from it. So, it is very simple nomenclature here. Now, here's a question. Why do bacteria have an enzyme like this? There are some people who feel that the reason is that this enzyme is here precisely to allow molecular biologists to cut and paste DNA, and this represents impressions likely, me among them.

How did anybody find this stuff? Well, shaggy dog story, I have to tell you the following shaggy dog story.

So, this is a fun shaggy dog story, and it's an MIT shaggy dog story because it comes from the work of Salvador Luria, who is a very famous biologist who worked here at MIT. So, Salvador Luria was studying bacteriophage. Remember, bacteriophage are the viruses that infect bacteria. So, he was studying bacteriophage, and he took his bacteriophage and used it to infect a strain of bacteria, strain A, and he also used it to infect a strain of bacteria, strain B. So when he did that, what you do is you plate a lawn of bacterial cells.

You kind of have a slush of bacterial cells that you plate here with virus mixed in, and wherever there's a virus, the virus grows, replicates, and either kills or slows down the growth of the cells so that bacterial cells grow everywhere else, but where a viral particle landed there's an absence of bacterial cells and that hole in the lawn, this whole thing is called a lawn of bacteria, and the holes in the lawn are called plaques.

So, when he did this, he found that when he did it on strain A he got a bunch of plaques and when he did it on strain B, he didn't, no plaques. So what what's the simplest explanation for this?

Strain B is different somehow. It's resistant to the virus. I don't know, the virus has to come in and do various things, and strain B isn't compatible with the virus or something like that. No big deal. So it's a resistant strain. But, occasionally you'd get a plaque.

Very occasionally, you'd have an occasional plaque.

So now, how would this be? I said the strain was resistant. How could there be an occasional plaque?

Mutation in, could it be imitation in the bacteria?

Sorry. Well, if it was a mutation in the bacteria there would be one bacteria that had the mutation. It was now susceptible, and it would die. But, the lawn would kind of grow because the cells around it wouldn't have a mutation.

So it's probably not a mutation in the bacteria but what could be? Maybe a mutation of the virus: what if it was a mutation in the virus that was able to overcome the resistance? Ah, so that's OK.

So, what this must be is the existence of a resistant virus that is a virus that can overcome the resistance of the bacteria. So far: perfectly normal, no problem. Now, let's do the following experiment. Let's take this resistant virus, and grow it, again, on strain A and grow it on strain B. What do you think is going to happen when I grow it on strain A? It'll grow lots of plaques. It still grows on strain A, and now what's going to happen when I grow it on strain B? If this was really a mutation that made it able to grow on strain B then it gets lots of plaques because it's now gained the ability to grow on strain B, and sure enough, that's what happens. So, there's nothing funky yet. But now, suppose I take one of these resistant viruses that I isolated here on strain B, I grow it again here on strain A. It grows. I grow it on strain B.

It grows. If I take it again from strain B and I repeat this, it'll still grow on strain A and still grow on strain B. Let's take one, though, from strain A. It's the resistant one which we have just now happened to have grown on strain A. And now, let's grow it again on strain A versus on strain B. And sure enough, it continues to grow on strain A, no problem.

And we grow it now on strain B. And, what shall we get? Well, it should grow on strain B, right, because it was a mutant virus, and it gained the ability to grow on either.

We passage it through B, it grows. We passage it through A.

But the answer was nothing, no growth. How can that be?

We had a virus. We agreed that was a mutant virus that had picked up the ability to grow on strain B, and we demonstrated it has now on either A or B.

We then reached in, and grabbed a copy of it here from strain A, having grown on strain A, and we try it again and it now won't grow on strain B. If this was a mutation, I mean, maybe the mutation reverted, right?

It was a reversion of the mutation. It mutated back. Is that plausible?

No, come on. The chance that all of the copies there would mutate back, come on. I mean, you could repeat this several times and this is always what happens.

What does that tell you about this mutation in the virus?

It can't be a mutation of the virus because if it was a mutation, it would be transmitted through.

But, passing through strain A makes it lose its ability to grow on strain B. But as long as you keep passing it through strain B, it can grow on strain B. This is not your typical genetics. So, Salvador Luria loved this.

And, he really worked out what was going on. And somehow, well, so anyway, they referred to this as strain B having the ability to restrict the growth of the virus. Strain B can restrict the growth of the virus.

That's where this word restriction enzyme comes from. What's really, truly going on here underneath the shaggy dog story? It took a long time before the shaggy dog story that Salvador Luria was the one to really demonstrate is fully worked out. But, what turns out to be the case is that strain B has a restriction enzyme.

That's how it restricts the growth. It has one of these enzymes that can cut DNA at a specific place.

When the virus comes into strain B, it injects its DNA, and the enzyme comes along and cuts the virus's DNA, protecting the bacteria. It's got its own little defense mechanism: pretty cool, pretty cool. So, any DNA that's introduced, if it has the sequence here, it'll take G, A, A, T, T, C, the bacteria cuts it. Wait a second, the bacteria has its own DNA. Why doesn't it chop up its own chromosome? Well, I mean, so one simple possibility would be that if this thing is looking for the sequence, G, A, A, T, T, C in the genome, maybe it's the case that the bacteria

has arranged that its own DNA never has a G, A, A, T, T, C. That would be a simple solution, right? But is it a plausible solution? Why not? But just statistically, how often do I expect to encounter a G, A, A, T, T, C? What's the frequency of any given six letter word in a four letter alphabet? It's about one in 46. So, about one in 46 positions will be a G, A, A, T, T, C, and that's about 4,000 letters. So, every 4, 00 letters, I expect to encounter a G, A, A, T, T, C. How big is the E coli genome?

4 million letters. So, how many G, A, A, T, T, Cs will there be? About 1,000 of them. It's just not plausible to imagine that it doesn't have the sites. So, your idea is that if it has these sites, it's got to arrange to protect its own sites. So, how is it going to protect its own sites?

Covers it or something. You could imagine something covers it or something, but you want to alter your own, so it turns out you're exactly right. What happens is there is an enzyme that comes along, and at this position, attaches a methyl group. It modifies the DNA by attaching a methyl group. It turns out that that methyl group is enough to prevent the restriction enzyme from binding.

So, this blocks the restriction enzyme. So, that way the bacteria is able to distinguish between its own DNA, which is methylated, and the viral DNA. So, wait a second, how does that explain my virus that manage to grow? How did my virus manage to grow? It would need to have gotten itself modified also to be protected.

Could that happen by chance? What if the methylation enzyme, the methylase, which is floating around in the cell, "accidentally" methylated the virus's DNA? What would happen then? The virus would become immune.

So, suppose the bacteria was pretty clever, and had a lot more restriction enzyme, and only a little bit of methylase?

Well, you'd imagine that most of the time the restriction enzyme would cut up the viral DNA first.

But every once in a while, the methylase would get there first and protect the virus's DNA.

That becomes an immune virus because it can't be cut by the enzyme anymore. And, if I take that, and I grow it again on strain B, it'll now produce lots of plaques because it was methylated. And, if I grow it again on strain B, it remains methylated because once it's methylated and comes into the cell, it's not cut. And so, its descendants will get methylated.

But, what happens if I ever grow that methylated virus on strain A? Strain A doesn't have the restriction enzyme, and it doesn't have the methylase. So, the progeny phage that grew up on strain A aren't methylated.

They're no longer protected. The protection that the virus has is the protection that comes from this methylation

enzyme.

It's not the sequence of the DNA. It's the attachment to these methyl groups. And so, it turns out that if you ever pass this virus through strain A, passage through strain A, the resulting DNA loses its methylation. And now, it can be cut.

And it can be cut. Well, this explained the weird results of Luria, that somehow bacteria had a complex defense mechanism of a restriction enzyme and a cognate methylase. The restriction enzyme would cut the sequence. The chromosome would be protected by methylating that site, and usually it would work fine.

Occasionally the bacterial virus would get methylated.

It would be protected as long as it continues to go through strains that have this restricted methylation system. That was it. Now, this shaggy dog story took a couple of decades to work out, and eventually led to Nobel prizes for the discovery of restriction enzymes. They're extremely important because although bacteria do this to protect themselves, they have also given us the perfect tool to now cut DNA where we want to cut DNA. Now, what if you wanted to cut at a G, A, A, T, T, C? You've got Eco R1.

But what if you wanted to cut it in another sequence?

Well, it turns out that if you want to cut it at G, G, A, T, C, C there's an enzyme called Bam H1. If you want to cut it at A, A, G, C, T, T or A, A, G, C, T, T, there's an enzyme called Hmd 3. If you want to cut it at just G, A, T, C like this, C, T, A, G, an enzyme called Mbo 1. And, there are enzymes that cut it this way, enzymes that cut it this way, enzymes that cut it this way, enzymes that recognize four bases, six bases. There are even enzymes that recognize eight bases.

It turns out that bacteria have elaborated zillions of different restriction enzymes that recognize different sequences. This is perfect for molecular biologists. Bacteria, of course, are much smarter than we are, having been out this much longer, have developed all of these tools for engineering.

All we have to do is borrow them. So how do you get Eco R1?

We grow out that strain of E coli; you purify Eco R1.

And how do you get Hmd 3? You grow up strain of haemophilus influenza. You purify the enzyme. At least, that's how primitive molecular biologists did it. If you wanted to work with a restriction enzyme, you'd grow up the bacteria.

You'd purify the enzyme yourself, and you would just use it in your laboratory. Of course today what does a

modern molecular biologist do if he or she should want Hmd 3?

It's in the catalog. So the catalog has 200 restriction enzymes. Yup, PSI-1 is new, on sale, 500 units for \$400.

Let's see what Eco R1 is going for.

Eco R1: look at this, 50,000 units \$200. That's a good price for Eco R1 because it's a very famous enzyme here. So all you have to do is you give them your credit card number and you have it tomorrow by FedEx. So that's how restriction enzymes are obtained today. So, next up, we can cut DNA any place we want to. We now need to glue DNA together.

Suppose I cut DNA, human DNA, and I'm going to cut it. I'll just take human DNA, your DNA, which I've purified, and I'm going to cut it at all its Eco R1 sites. I can take any other DNA I want.

I don't know, I could take zebra DNA. I could take anything and I could also cut it at Eco R1 sites. I could mix them together, and after mixing them together the fragments will float around and remember this down here has T, T, A, A. This fragment over here from some other piece T, T, A, A, this could be human DNA. This could be zebra DNA if you want to. It doesn't matter. It could be bacterial DNA.

These fragments overlap. They'll hydrogen bond a little bit, but that of course won't introduce a covalent bond here.

I'd really like to make a covalent bond. I would like to attach the piece of DNA from one source to the piece of DNA from the other source by doing the opposite of the restriction enzyme. The restriction enzyme cut at these locations. I would now like to catalyze the rejoining of the sugar phosphate backbone here.

So I would like to rejoin the sugar phosphate backbone. I have a hydroxyl here. I have a phosphate here, and I would like to ligate them together. So how I manage to ligate? What kind of fancy chemistry do I do to ligate these pieces of DNA together? I don't do any fancy chemistry. I again sit at the feet of bacteria who have solved all these problems before. And I ask bacteria, how do you do this? And they say, well, we have an enzyme called ligase. So, you purify ligase from bacteria, you add that, and ligase ligates the fragments together. Why do bacteria have an enzyme ligase?

For a pair of their own DNA. Things go wrong this is part of the DNA maintenance scheme of bacteria. They have an enzyme ligase to appear their own breaks in DNA and, obligingly, you can purify DNA ligase. So you add ligase, today, of course, if you need a ligase, how do you get it? It's in the catalog, absolutely. So, you can glue together any of those things you want. All right, next up, what DNA do I want to stick together? I mean, here I made a silly example.

I'm going to stick some human DNA to some zebra DNA. Why do that? I mean, just to show you that I can do it, right? I'm just demonstrating that I could stick any DNA to any DNA. Remember, once I've got a piece of DNA it doesn't know whether it came from a human or a zebra. It's just the molecule. You can stick the molecules together, right? But what do I really want to attach my human DNA to?

I want to attach it to some other DNA that has the ability to grow on its own within bacteria. Vectors: I need to make, here's what I would really like. I would like to have a piece of DNA that has some sequences that contain the recognition sites for replication. I'd like to have some replication initiation sites here. So, a piece of DNA that, remember, because the bacterial chromosome itself, here's my bacteria, bacteria's chromosome replicates itself, and it has the ability to start DNA replication at multiple sites called origins of replication. But, what I would really like is to be able to construct in the laboratory a synthetic piece of DNA that also would function as an origin of replication because then what I could do is in vitro take my piece of DNA, attach it to this vector, and it would now have the ability to grow the bacteria. How am I going to make a piece of DNA?

What kind of engineering tricks can we do to create a small piece of DNA that has all the machinery needed to be able to be copied and replicated just like bacterial chromosomes? That's a pretty fancy feat of engineering.

How are you going to do that? Sorry? OK, so who are you going to ask? If you wanted to do this, you're going to ask the experts. Who are the experts? Viruses or bacteria, or basically, if you want to do anything, the place to ask is the folks who have the most experience. And, the folks who have the most experience are almost always prokaryotic organisms because they are by far the most evolved things on this planet. Anything that can replicate itself and grow every 20 minutes or something like that has had a lot more generations of evolution than you have. And therefore, they are much more optimized than we are. And so you go ask and say, has any bacteria worked out how to do this? Turns out bacteria have worked out how to do this just fine. In fact, most bacteria, at least many bacteria, contain within them, in addition to their own chromosome, small circles of DNA. These are called episomes.

This is the chromosome. Epi means on top of or in addition to. So in addition to the chromosome, there's an episome. The episome is in fact an autonomously replicating piece of DNA that has an origin. And it replicates. Why do bacteria have episomes? It turns out episomes often contain genes. One of the genes they contain, or some of the types of genes they contain, are resistance genes. There might be, for example, a penicillin resistance gene contained on an episome, or a streptomycin resistance gene.

It turns out the bacteria have these episomes containing resistance genes, and they're not in the chromosome.

They're separate. Now, why would they do that?

It turns out when a bacterium dies and a cell cracks open, the DNA spills out. The next door neighbored bacteria has mechanisms to suck up DNA from the environment. You never know. It might find something interesting out there.

So, it turns out that bacteria are rather promiscuously exchanging pieces of DNA all the time. And so, a bacteria that has an episome that has a penicillin resistance gene can spread it to other bacteria, and it's very nice. It's compact.

It's on its own little episome, autonomously replicating piece of DNA. This is great for bacteria wanting to spread drug resistance. It's not good for human populations, for example, because this is how drug resistance spread through populations. This is why we have spreads of penicillin resistance. Now, of course, wait a second, this whole mechanism of spreading drug resistance, we've only had antibiotics since the 1940s. How did bacteria devise this so quickly? Sorry? Many generations since 1945?

That would be very impressive.

Yeah, but, I mean, why do they have this episome mechanism, the ability to spread DNA and all that? That's an awful lot to evolve in 50 years? Yeah? Something natural like penicillin. It turns out, we didn't think of penicillin. Who thought of penicillin?

Fungi. Right, again, we learn from the lower organisms. Penicillin comes from fungi. Bacteria have been fighting off penicillin for millions and tens of millions of years. So, we may be very proud of our penicillin and all that.

But, they've been at this for a very long time.

This is about war between bacteria and fungi.

That's what this is, OK? So, that's why these things are here. They're here so that bacteria can have these resistance genes against fungi and things like that that make antibiotics. Antibiotics are natural. We've made a few new ones, but most of the antibiotics have been made by nature.

And so, if I wanted to replicate DNA, if I wanted to attach my human DNA to a piece of DNA that's capable of autonomous replication, autonomously replicating circles of DNA, these autonomously replicating circles of DNA are also called plasmids. And that's the word we'll mostly use for them, plasmids. All I need to do is purify a plasmid from a bacteria. So, I find a bacteria that has plasmids.

I purify the plasmid, and then I can cut open the plasmid at the Eco R1 site, OK? So, this plasmid will have an ORI, an origin of replication. I'll cut it open at the Eco R1 site. I'll take human DNA fragments that I've cut with Eco R1.

I'll mix them with plasmid DNA that has been opened up, has an origin. Ligase will come along, join this up, and now I have a circle of DNA that has all the machinery to autonomously replicate, plus my human DNA. Now, if I wanted to get a vector, or an honest to goodness plasmid, I can go to a bacteria, grow it up, purify the plasmid, and cut it. Or alternatively, if I needed the plasmid, say, tomorrow, it's in the catalog. The next section of the catalog has a long list of plasmids here. There's a plasmid there, right? It's a nice plasmid.

Oh yes, let's see, pucK is a very good plasmid.

PBR 322 is a good plasmid. The whole section, all this purple stuff are the plasmids. So, you can get the plasmids too.

You place one order, you get the restriction enzymes, you get the ligases, you get the plasmids, no problem. So, I can then take total human DNA, cut up, cut up, cut up, cut up, add in plasmid, and I'm going to ligate together. And then, having ligated my human DNA to my plasmids, I'm going to mix with bacteria. I take some bacterial cells. I add my mixture of these plasmids containing human DNA. And now all I have to do is persuade the bacteria to suck up my plasmids containing human DNA.

How do I teach bacteria to suck up DNA? They do that for a living. That's what they do. They're always spreading material.

They have that ability. All we're doing is we're using their ability. So you get the sense that the kind of engineering that really works in biology is engineering that exploits what nature has been doing for a very long time.

Rather than butting your head up against the problem, usually somebody has solved it, and it's almost always bacteria. So, you've transformed the bacteria.

Now, there are a few tricks you can use to make them a little more transformable.

You can add calcium phosphate, and blah, blah, blah, but you can sort of persuade them to take up the DNA. And then all you have to do is plate them out on a plate.

Plate them out fairly dilutely so there are a lot of single bacterial cells that land on the plate, and wait for them to grow up. Each one of these had a single plasmid, a different plasmid than the next guy over. Wait a second, each one?

How do I guarantee that every bacteria in my test tube took up a plasmid? Is that plausible? I mean, I can't guarantee that every bacteria is going to take up a plasmid. Maybe I'll add so much plasmid that every bacteria will take one up. Oh, but that's a bad idea because why? Because then a lot of them will take up more than one.

You don't want to do that. You really only want to have at most one. So, if you were going to arrange so that at random you only have about one, you've got to have a lot that are zero. So, this is a problem. I mean, it's a real waste. My library is going to have large numbers of bacteria that don't have any plasmid. In fact, this transformation process is not so efficient. It's not so efficient. So, we have a little bit of a problem here is that some of these guys will have human DNA.

But, most of them won't. So, what can I do to arrange that any bacteria that did not pick up a plasmid was incapable of growing? Add a resistance gene to the plasmid. Suppose I were so clever as to add to that plasmid, penicillin resistance. So, not just an origin of replication, but suppose I also had a resistance gene here, say, for penicillin resistance or streptomycin resistance, or ampicillin tends to be a very big favorite, ampicillin resistance. Then, my plasmid would have ampicillin resistance gene encoded on it, an enzyme that can, say, break down ampicillin, so on and one way to to my perch you plate I just said ampa cell and now i'll even though most of the bacteria have not picked up a plasmid, only those bacteria that have picked up a plasmid have the ampicillin resistance gene and can grow on an ampicillin containing plate.

Now, how do I get a plasmid with an ampicillin resistance gene? It's in the catalog. It's all there, right? In fact, these occur naturally.

You can, with restriction enzymes, move the ampicillin resistance gene to your favorite plasmid. If you don't like that, you can put in kanamycin resistance, etc., etc., etc. So, that's how you do it. So, we've got the big picture here.

We have now gotten a library, the Library of Human Fragments contained in E coli. The library is a big Petri plate or many Petri plates, each one of which is a colony.

Each colony has a single vector with an origin, a resistance marker, and a distinct piece of human DNA. In this library lives somewhere the gene for Huntington's disease. Over here is a gene for cystic fibrosis, over here a gene for Duchenne muscular dystrophy, over here a gene for diastrophic dysplasia, over here a gene for etc. etc. The only detail, now, you've got a library. You've managed to purify each piece of human DNA away from every other piece of human DNA.

The only question now is how do you use the library?

How do you go to the library and withdraw the correct volume from the shelf? How do you find the one you're looking for? So, we have converted the problem of purification, which in every other form of biochemistry starts by saying, I'm going to purify something based on its distinctive properties, to I'm going to randomly purify everything.

Everything would be purified in its own bacteria, and I've now converted to the problem of finding the one that I

want in my library. Next time, we'll talk about how you go to the library and find what you want.

See you then.