JOANNE STUBBE: Because we haven't gotten that far in class to understand what this protein is that's the focus of the paper, I still think the paper is straightforward to understand. I'm just going to put it into context. So I was having trouble trying to decide what to do. And maybe I shouldn't have done this.

But the fact is that this technology we're going to be focusing on in a very sort of simple way, CRISPR-Cas, has taken the world by storm. And that's the take home message from this. So you can sort of get what it does. But really to look at the details, you have to go in and study it.

And every time you pick up another journal, you look at Google Journals or something like-- there's another 100, 200, 300 papers published on this. So this is a current technology that has taken off really since 2012. And so very rarely is technology successful in that short period of time.

And it happens to have been applied to one of the key enzymes that people are now focused on in terms of controlling cholesterol levels, which is what we're talking about. So I use it as an opportunity to just show you what this technology is. Have any of you ever done this technology? Nobody in the last class had done the technology either.

But my niece is a sophomore here. She spent a whole EUROP doing this technology. So this technology has moved into the lab.

My lab hasn't used it either. So I probably can't answer any of the details. But it's one of these things that it is extremely complicated. I think I can give you a cartoon overview of how it works. But if you're going to use it, just like every tool, you have to study it in more detail.

OK. So I am going to ask you questions. And this is going to be different from the one I did on Thursday, because I spent too much time talking about this article. And then I'll come back to-- how many of you read these articles? How many of you didn't read these articles?

OK. So your class is much worse than the other one. The other one had read all the articles.

OK. So we won't have a very good discussion about this. And I'll tell you why I think you should read that, but I'm not going to focus on that till we end.

OK. So the paper we are going to focus on is this one. And this is the gene product, the protein that has become a focus of attention of many people in terms of controlling cholesterol levels and as an alternative to statins or maybe better than statins, but we haven't gotten there yet. OK. So this is still on the drawing board in there.

Many people focused on clinical trials targeting this particular protein. And so one of the questions that this paper focused on and asked-- hopefully, you all have read the paper. It was only three pages, so it wasn't very hard to read. Is this protein important in terms of controlling cholesterol levels?

And they did experiments in tissue culture and in mice to try to address that issue using CRISPR-Cas as a way of destroying the gene, the gene which then destroys the protein. OK. So this other article sort of gives you an overview of the kinds of things we need to think about to make the technology better. And when technology is introduced-- just like if you look at unnatural amino acids that you guys looked at with the Schultz technology.

I mean, first 15 years Peter was doing that, he collaborated with my lab. We never published a single paper, OK? Because the technology was not good.

And so now, the technology is still not good. But it's getting there, and it's improved greatly. So you often see something.

It looks, oh my goodness, you know, this is going to be fantastic. But the devil is in the details, OK? And that's one of the take-home messages from this course anyhow.

OK. So what I want to do is just give you a very brief overview of what I had hoped to get to by the end of lecture today and didn't quite get there. And so we did get to the fact that we made LDL particles. And LDL is transferred in the blood and is a major carrier of cholesterol.

So it takes cholesterol from our diet. And it's going to deliver it into different kinds of cells, and it does so. There's a receptor on the surface of the cell. And these little things here, these little flags, are at the receptors and the receptors.

That's what we're going to talk about next lecture is low density lipoprotein receptor. OK. And this is the basic. Brown and Goldstein figured out that genetic mutations in this receptor and

other steps associated with getting the receptor to the surface of the plasma membrane are responsible for children for the disease familial hypercholesterolemia where kids die at age 7 heart attacks, because of inability to control cholesterol levels. OK.

So I need to just sort of briefly walk you through the model, because that model is related to the effect of this protein you were reading about in the paper that you were supposed to read for today. OK. So this lipoprotein can bind to the receptor. This is a plasma membrane.

You see there are three receptors. The receptors have to cluster to be successful at somehow, by mechanisms that are moderately well-understood, can engulf the LDL particle and form a little vesicle. And the little vesicle is coated with a protein called clathrin.

OK. And we'll see over the course of the rest of the semester this is used over and over again-- so is clathrin-- as a way of taking up nutrients into the cell. So this is a major mechanism of doing that. And then what happens is the clathrin is removed. Biochemically, it's removed enzymatically.

And what you're left with is a vesicle that then fuses with an endosome. And so that's a little organelle with lipid membranes that is acidic. And when the LDL protein gets into the interior of this little vesicle and the pH is lower than the normal pH, goes from 7 and 1/2 to 5, the LDL receptor dissociates from the LDL particle.

And so then what happens, by mechanisms that are really incompletely understood, the receptors can recycle to the surface. OK? And what happens is that, when some of them recycle to the surface, you're left with an LDL particle that fuses with another organelle called the lysosome. And then this LDL particle goes into the lysosome.

The lysosome is sort of like a proteasome. It's a bag of proteases and lipases. So it just degrades everything in there-- amino acids, fats, everything-- allowing you to produce amino acids and cholesterol, free cholesterol.

And then cholesterol in the liver is often stored. It gets esterified. And it's stored as triacylglycerol. Fatty acids esterify to cholesterols. OK?

So the process, of course, of getting the LDL receptor to the surface is done in the rough endoplasmic reticulum. Because it's a membrane protein, it's transferred by things called little coated vesicles. And then somehow these little coated vesicles deliver the receptor to the

protein. OK?

So this is a very complicated process. And, in fact, mutations that are responsible for heart attacks occur in every step in this process. It's not just the LDL receptor. We'll see that in class next time. OK.

So the key thing you need to know for today is that you have LDL receptors that interact with LDL. And that's key to taking the cholesterol into the cell. That's the take-home message. That's fairly easy to understand.

OK. So the protein we're focused on today is this guy, PCSK9-- horrible acronym which I've written down. I can't even remember it. But it stands for Proprotein Convertase Subtilisin/Kexin 9. OK.

So the important thing is subtilisin. Has anyone ever heard of subtilisin? So that's like [INAUDIBLE]. So it evolved convergently.

And so it is a protease that has a serine, a histidine, and aspartic acid, like you learned in protein media degradation in the first part of the module. OK. And this protein was discovered-- we'll see in a minute-- again, because of patients. OK. So the patients presented themselves in a funny way.

That's how the LDL receptor was discovered. If you've read Brown and Goldstein's article, which was one of the things I asked you to read, you've already gone through that. That was the thing that got Brown and Goldstein excited about this.

What's going on? Why do these kids have heart attacks at such an early age? Can we figure out what's wrong? And can we do something to fix it?

And so, here, what happens is this protein is made as a proprotein just like any kind of serine protease. Lots of times you are pre-proproteins. And they process, they usually self-process, into an active form. And why do they have that? Why does a protease have a pre-pro sequence on it?

AUDIENCE: So you have, like, spatial temporal control of the sectors?

JOANNE STUBBE: Yeah, over activities. So you're controlling the activity. Because if you produce a protease, nobody could ever overproduce proteases. Why? What happens inside the cell? Everything

gets degraded.

OK. Because proteins have specificity. But if you overproduce them, all your proteins have degraded. So it's not trivial to overproduce proteases.

And so they have a mechanism-- hopefully, you learned about that in introductory biochemistry course-- that makes it inactive till you're ready to use it. And then it cleaves itself. Something triggers it, it cleaves itself. And then it's ready to go. And that's true here, too.

So here you have this little purple worm that has to auto process to become active. And in some way, it's going to end up extracellularly. And so it's got to go through membranes.

So it goes through the Golgi stacks, just like I just showed you with a cholesterol, the LDL receptor. And it gets extruded extracellularly. And that's where it is. It's out there. OK. It's processed from the original version of it.

And so the working hypothesis is-- and this was based on a patient. They found a patient where the LDL levels were elevated, OK. And the child that had this had early coronary disease. That is heart attacks at an earlier age.

And they studied this in some detail. And they found out that what this protein does-- I'm not sure we really understand the details of what the protein does-- was that it could bind to the LDL receptor. OK, so this little orange thing is what you just saw on the previous slide.

And this little blue thing is LDL. So, now, what happens is instead of having just LDL, low density lipoprotein and the receptor, you've now got another protein stuck to this. OK. And so when this protein is bound, it also undergoes receptor mediated-- they don't show any steps here. I'm not sure if it's been studied in detail.

It also undergoes receptor mediated endocytosis. So it's taken into the cells. And normally, remember, with the LDL receptor, the receptor gets recycled. Here, what happens? Something changes because of this complex.

And so now this complex is in the endosome. But the LDL particle, which has a cholesterol, doesn't associate from the LDL receptor. The receptor doesn't recycle. But, instead, the whole gemisch, the protein, the receptor, and the LDL particle, fuse with the lysosome, which is a bag of proteases. And it's degraded.

So what are the consequences of that? The consequences of that are that you lower concentrations of the LDL receptor on the plasma membrane. OK? And if you lower the concentrations of the LDL receptor on the plasma membrane, what happens to the low density lipoprotein concentrations?

AUDIENCE: [INAUDIBLE]

JOANNE STUBBE: Yeah, it increases. And so then you're in trouble. OK? So that's the model. Again, I haven't read a lot of papers on this.

The discovery was made of this of patients that had that phenotype in 2003. But they also found patients that had a loss of function. And they found out that some of these patients-- they're different kinds of patients. They have different phenotypes. But they had a single mutation.

And these patients with single mutation had reduced LDL cholesterol. And they had the same amounts or elevated amounts of the LDL receptor. And because they had LDL receptor, they had lower cholesterol and more LDL receptor to take up the cholesterol, they had reduction in coronary disease. OK?

Everybody get that? Why do we care about that? OK. So can somebody tell me from the paper what was the take-home message from the paper?

Why do we care about that? What's unique about this particular protein protein, PCSK9, compared to using statins, for example? Did you guys read the paper?

OK. So the paper was pretty short. Even if you didn't understand all the details, I thought the paper was pretty easy to understand. So why do we care about? What was the take-home message? Why are we targeting this?

AUDIENCE: To change the expression of proteins that create new--

JOANNE STUBBE: PC.

AUDIENCE: Yeah.

JOANNE STUBBE: Yeah. So but why do we want to do that? We have statins. Statins, you know, everybody's gobbling statins a lot. I mean, you probably know 20 people that take statins.

I know many, many people that take statins. So it's a wonder drug in many ways. But when do you start giving statins?

When do people start taking-- I'm probably not allowed to ask that. So you don't have to answer if you don't want to. But are any of you taking statins?

No. OK. But there could be people that have, you know, high cholesterol. I mean, a lot of it is genetic. I eat McDonald's hamburgers all the time. And I eat huge amounts of ice cream.

And I have extremely low cholesterol levels. OK? And it's genetic. OK. Other people might not eat any of that stuff, and they might have extremely high cholesterol.

So when you see people, maybe your parents, basically, they're taking this. And it's after you have some issue, right? You have coronary heart problem. You have chest pains, whatever.

So they start looking for what could be causing that. And the first thing they look for is clogging of the arteries. And that's when they start some kind of therapy like statins.

The beauty of this is, if this model is correct that I just showed you, if you could figure out how to remove or greatly reduce that protein, then that would automatically, you know, prevent the normal function of this protein, which is to degrade the LDL receptor in the lysosome. And I'll get to that in the very end. So if you could figure out how to treat, you could diagnose the predisposition to having elevated cholesterol levels and start treating it much earlier. You have a much higher propensity for success compared if you take statins halfway through your life.

I mean, there's really good epidemiological data that support that. So people are extremely interested in figuring out-- I don't think we know the details of the function of what this protein is-- but lowering this protein. Because the consequences of that are lowering cholesterol in the plasma.

OK? So are we all on the same page? Everybody understand that? Because that's key to thinking about the paper. OK.

So the reason I picked this is people these people in this paper wanted to understand is this protein really important. And so what they did was they decided they were going to knock out the gene or do something to greatly reduce the gene, which then would reduce the amount of protein, which then would allow you to analyze the phenotypic consequences. OK? And what was the analysis they used in this paper?

They used two different kinds of analysis. Well, we're not in detail. Globally, what did they use? What were their model systems?

AUDIENCE: [INAUDIBLE]

JOANNE STUBBE: You need to talk louder because I'm deaf.

AUDIENCE: For [INAUDIBLE] they used a surveyor as--

JOANNE STUBBE: Yeah, so that used a surveyor on what, though? So that's too detailed. I want a bigger picture. So you're right.

They used surveyor cell assays. That's more detail than I want right now. So they looked at it two ways if you look at the figures. So what were the assays? In the surveyor assay, what were they assaying?

AUDIENCE: The blood samples.

JOANNE STUBBE: The blood samples of mice. So that's one of the things from the liver. OK. So they took liver cells from mice. So they were using animal models. OK.

So one of the questions that, if you read the paper carefully, you should be asking yourself-- and this is always a question when you're looking at therapeutics. Is this animal model any good? OK. And then the other way that they were looking at this was with tissue culture cells.

Because, in general, when you start studying something, you don't start on humans, or you don't start on whole animals. You need to start on something simpler. And we haven't gotten to this yet, but Brown and Goldstein, if you've read the reading, have used fibroblast cells.

And they showed fibroblast cells behave like liver. And it turns out it had great predictive power. It might not have, but it does.

So you need some kind of a model system. And so they used both of those systems to try to test the idea that, if you could get rid of this protein, you could alter in a way that these patients, these loss of function patients, behaved in terms of the levels of cholesterol. So that was it.

And so how did they decide to do this? And so I would say, in general, we don't talk about this kind of stuff very much in 508. But if you're ever going to be a biochemist, you can't do

biochemistry without being able to do gene knockouts inside the cell.

So 25 years ago, that was tough, OK? In the mid-1980s, you could first do that well in bacteria. We still do a lot of that in my lab. It takes four months, three, four months.

It's not easy. With the older technology, it works. But it's a rare event, and you've got to screen through a lot of things to find the ones that are interesting.

And this technology, CRISPR-Cas, allows you to do this in a couple of days. It's revolutionized what you can do. So you might be studying something really complicated in the test tube. But the question is is what you're studying relevant to what's happening in the cell.

And so if you're asking a chemical question, a mechanistic question, like how does isopentenyl pyrophosphate do its chemistry, you don't need to do that in a cell. You can do that in a test tube. If you're asking how things are regulated, which is what we're doing now, you must be in the cell.

And the issues within the cell are that people overproduce stuff. You know, they have to mess around, so they can see something. And whenever they do that, they change everything.

So the future, for anybody that's interested in biochemistry biology interface, is you've got to be able to do both. And so this technology, I guarantee, in some form you will be using if you pursue a career in doing biochemical and biological studies. OK. So the question really is we want to do manipulation of a gene. OK?

And so people have wanted to do this forever. So you might want to delete the gene and see what the phenotypic consequences are. You can do that. You know, there's been technology around.

They won the Nobel Prize for the technology in 1983. But, again, it takes months. And you have to screen through millions of cells to be able to figure out which one has your gene deleted or another gene inserted in place of the gene of interest where you've modified the gene of interest, which then gives you information about the function of the protein.

And so having technology that can turn around rapidly is important. And so I'm just going to show you what the state of the art has been up until two years ago. And, really, they also work by the same mechanisms.

It's just the CRISPR-Cas, even though it's really still early days, works much more efficiently. OK. So the idea is you have a piece of DNA that you care about, and you want to cleave it. And all of these cleavages are double-strand breaks.

So double-stranded breaks are lethal to the cell, so you have to repair them. OK. And you have to have a way to repair them. And I'll show you what those two ways are. You've all seen it in some form.

But you want to have cleavage at a specific site. And then when you have cleavage, the question, if you repair that, can you delete part of that gene which would make the entire gene inactive? Or, can you replace, in this cleavage site, a gene of interest with a mutation in it, et cetera? OK.

You can do many, many, many genetic engineering projects, which are sort of covered in review articles. The more creative people become, the more things you can actually do. OK. So how do you do that?

So what they do in the case of the zinc fingers, does anybody know what a zinc finger is? Has anybody seen a zinc finger before? So a zinc finger is a little small protein, I don't know, maybe 70, 80 amino acids that combine zinc and that its sequence specifically binds DNA.

OK. That's a major way of regulating transcription inside the cell. OK. And there's not just one zinc finger. There are many, many zinc fingers. OK.

So what people have done is taken these little motifs that combine zinc and designed these motifs, so they recognize a sequence. So these guys, these little zinc fingers, now are targeting the DNA that you want to cleave. So they're targeting it here, and the targeting it here.

So what that means is every time you want to do an experiment like this, you have to make a little zinc finger. String them together to get enough binding affinity, so you get specificity. That's really key. And you could do it.

We're pretty good at this, but it takes months. And so what they do is, once they have these binders, then they attach a nuclease. OK. So Fok1 is the nuclease.

So that just means you're cleaving a phosodiester bond over your nucleic acid. And you cleave on one strand. And on the other strand, you have a double-standard break.

And these enzymes work by giving you blunt-ended cleavage. There's no overhangs in the DNA cleavage. So people have used this for a long time.

In fact, Carl Pabo at MIT, who's an X-ray crystallographer that studied regulation by zinc finger transcription, was one of the people that founded the companies that got this technology off the ground. But it's hard. OK. So the second technology which I think is much more widely used-- but I think it'll be completely displaced. I might be wrong.

You can buy a kit Golden Gate, TALEN kit. That's right. You can buy it from some company. And it's the same idea.

So, I mean, I don't know anything about this in detail. But it turns out that these little proteins, which are 34 amino acids, you can actually look at a sequence of DNA and design 34 amino acid repeats in a way that it can bind to double-stranded DNA. OK. So this is like, so you have a double-stranded DNA helix.

You string a bunch of these little domains together. And you can actually design these little domains, the sequence of these little domains. And it forms a super helix around the double-stranded helix.

So the protein forms a helix around the nucleic acid helix. And what it does is it targets the nuclease for cleavage. So it's the same idea. It's just the mechanisms of targeting are different.

And so, I mean, they have structures of these things. It's sort of really an interesting problem in molecular recognition if any of you are interested. But I would say, if you want to use this to do something biochemical and biological, you probably want to go to CRISPR-Cas system now.

OK. So both of these are the same. They have a nuclease and something that targets it to a DNA sequence of interest. And if you've read the paper on the PCSK9, that's exactly what they're doing. They're targeting a sequence for double-strand cleavage.

OK. So this then brings us into the CRISPR-Cas system. And I've given you a hand out of this which is, again, a simplification. Now, I think there are six different moderately well-studied CRISPR-Cas systems. They're all different.

So they all have different numbers of proteins. Although, the idea of how they work is pretty

similar, I think this has turned out to be the best behaved in terms of biochemically putting it back together and having it work. OK. So what do we have here?

So, hopefully, you all know now that what you need for this to work is a Cas9. What's Cas9?

AUDIENCE: The CRISPR associated with [INAUDIBLE].

JOANNE STUBBE: So a CRISPR-- is that what the acronym is?

AUDIENCE: ...it's a nuclease.

JOANNE STUBBE: Yeah. It's a nuclease. OK. And what's special about this nuclease?

AUDIENCE: Sequence-specific. That's like--

JOANNE STUBBE: It has what?

AUDIENCE: A guide RNA that makes it--

JOANNE STUBBE: No. So just the nuclease, we're just talking about the protein now. We do have to worry about that, yeah. So if you look at the Cas9 sequence, what do you find out? That's not in the paper, but--

AUDIENCE: So it's got two different regions that can bind the two different strands--

JOANNE STUBBE: Right

AUDIENCE: --and, like, [INAUDIBLE] in a different [INAUDIBLE].

JOANNE STUBBE: Yeah. So you have two different nuclease domains. OK. I mean, this is not necessarily given. One is going to go to one strand. And one is going to go to the other strand. OK.

And we'll talk a little bit about that. And then, as you were saying, what's unique about this? In this picture, what's wrong with this picture? If you read the original discoveries in the bacterial system, what's unusual about this particular-- well, I guess-- OK, no. It's OK.

OK. So what do you have here? What is this part? This should be tracr. What's tracr?

AUDIENCE: It's the transactivator.

JOANNE STUBBE: Yeah, so it's transactivating. OK. And then what's the gRNA?

AUDIENCE: The guide.

JOANNE STUBBE: So that's the guide that is part of this bigger piece of DNA that we're going to look at in a second. OK. So what you need, although this isn't what people use now for the technology, is you need two pieces of RNA. And you need the target for double-strand cleavage.

And you only need a single nuclease. OK? And the key question is how do you make them assemble. OK. And how do you make it as simple as possible, so that you can use this in bacteria, but also use it in humans which is what Eric Lander focuses on.

So CRISPR, and we'll look at this in a [INAUDIBLE], has this horrible name, Clustered Regulatory Interspaced Short Palindrome Repeat. OK? So that's the name. And so this just summarizes-- and we're going to come back to this in a minute-- that all three of these methods, the zinc fingers, the towels, and the Cas9 system, all do the same thing.

They somehow recognize double-stranded DNA and cleave it, OK? And so they all give you a break in the DNA, which is lethal if you don't figure out how to deal with that break. OK. And there are two ways to deal with that break.

There are two ways of repairing the break that we're not going to talk about in detail, but you probably have heard about somewhere. So what's the way that they deal with this double-stranded break in the paper? Did anybody read the paper carefully enough?

And how do they know? So, somehow, you've got to put these things back together. Otherwise, your organism is completely dead, which is the goal of having this CRISPR locus for the bacteria. They want to kill the invading virus. OK. But in this particular paper, which one of these two methods did they show or did they propose from the data that they talked about was involved in repairing the double-stranded break?

If you look at this paper, they describe non-homologous end joining. Because in the end, if you looked at the paper carefully, when they were trying to tell whether they successfully got a double-stranded cleavage, they did a lot of polymerase chain reactions to figure out whether they got specific or non-specific cutting. And when they did the sequencing on this, they could tell, because of the different mechanisms between these two, that most of the damage was repaired by non-homologous end joining.

So what happens with this approach? What happens with this approach is that the repair is

putting the things back together. When you have blunt ends, you've lost the information from the sequence. And you have a disconnect and, if you got a couple of cleavage sites putting them back together, is really tough.

And so when you put them back together, you might have an insertion. You might have a deletion. You might have a frameshift. You get a mess.

But then when you look at the very ends of your gene using the polymerase chain reaction, what happens is you get a mixture of things. And you can sequence them, so you can tell something about how the repair happened at the double-stranded break. So if you have a double-stranded break, OK, so the question is here do you have a deletion, so it's a little bit shorter. Or, do you have an insertion? Or, do you have an rearrangement?

And what you do then is sequence these things using PCR. And then you can get information about the mechanism of repair. OK? So the alternative mechanism-- and this is really important if you want to replace one gene with another gene, a whole gene, rather than just removing the gene, which is what happens here.

Here, you've made a cut in the middle of this chain. You've removed a few amino acids. Or, you a removed amino acid, and it's rearranged a little bit, so the protein is never going to get formed.

Here, what you're doing with the homologous repair is you have a template. OK. So if you don't know anything about homologous DNA repair, you need to go back and look into it, your basic textbook, and at least read the definition of what's going on. But you have a template.

Once you have a template, you can copy that template and replace one gene with another gene. So this template becomes really key in replacing, site specifically, one gene with another gene and, as a consequence, one enzyme or protein of interest with another one. OK.

So this was taken from an article by Jay Keasling. And Jay Keasling is interested in synthetic biology. He's an Artemisinin in fame. We talked about that in class. That's the anti-malarial agent.

He's also been a major player in trying to figure out how to make bacteria use mevalonic acid pathway, which is what we're talking about in class, to make jet fuel. OK. So how do you make hydrocarbons that are really energy efficient compared to ethanol or butanol? And so his whole lab is focused on figuring out how to use CRISPR-Cas to engineer genes from many

different organisms back into the organism of choice.

And this technology, apparently, allows you to do five or six genes simultaneously once you figure out how to do it. And so you can do a lot of manipulation in a really fast time compared to the months it used to take before. And so what does this tell us?

I mean, I think this is the most amazing thing. If you read the Eric Lander historical perspective on the discovery of CRISPR-Cas, there was a guy in the late 1980s that lived in Spain and did all his research in a salt marsh. OK. And he got really interested in these archaebacteria, really weird bacteria. I don't think they're weird, but most people don't really think about they have really interesting chemistry.

And when he was sequencing part of this, for some reason, he found palindromic repeats, many palindromic repeats. And that's those purple spacers. And he says, well, what that heck is going on with that? What is this?

OK. So he had discovered this locus in the genome. OK. Now, bioinformatics over the years, if you read the history of this, played a huge role. So you go back, and you look at all of these sequences. You find even an E. coli, you have these little spacer repeats over and over again that are palindromic.

And so then the community got really interested in why you would have a locus that looks like this. And what they found is, if you start looking at the genes on either side of it, you found genes that were conserved, that coded for the Cas9 protein. In this case, S. pyogenes is the one that was used in this paper, which is the nuclease.

And they also found this transactivating RNA. And what's interesting about the transactivating RNA is it has a sequence that's homologous to one of the sequences in the spacer. OK. So they've got to be able to hybridize to each other. OK.

So that started. It became very interesting. And the question was focused on how is this editing going to happen when you get cleavage of your gene. Or does this act like, for example, in SI or an SH RNA in controlling levels of gene expression?

And so there were many people that contributed over the years to figure out how this locus is used. And that's what I'm going to briefly describe. So the idea is the following.

And I think that discovery, in my opinion, is really a seminal discovery by some guy who was

working in a marsh working on some bizarre archae, made this discovery, followed it through for the next 15 years, and discovered that bacteria have adaptive immunity systems. I mean, that's really sort of mind-boggling. I remember when this paper was published, a first paper was published where they knew this was happening.

Somebody in my lab gave a group meeting on it. And my mouth just dropped to the floor, because nobody predicted this at all. This is what I would call a revolutionary discovery.

And what they found was-- and, again, this is the bioinformatics data analysis now, which we can do better and better and better. What they discovered-- they were looking for what's in between these spacers, OK? So what's in between these repeats?

I'm calling it the wrong thing. These little purple things are the repeats. Again, they are the palindromic sequences over and over and over again. What is in between the repeats of the spacers?

OK. Where do the spacers come from? Well, they didn't know. OK. But when they started looking at sequences of many of these things, what they found was they came from phage, viruses.

OK. So, here, they have a bacteria, and they have phage DNA. Because people would sequence a lot of, at that time, phage DNA. And so what happens is the virus, or it could be a plasmid born piece of DNA where information is transferred from one bacteria to another, they get into the cell. And then they have proteins.

And these, again, are Cas genes. And people are still studying these that take the viral DNA or the plasmid DNA and cut it into little pieces and somehow insert it between these repeats. So all of these spacers are different sequences of DNA that come from the invading species, the virus, or a piece of plasmid that you got from another bacteria.

And so that became really exciting. OK? And so then the question is, how do you take all this information and convert it into something that can kill the idea-- if you have adaptive immunity, how do you use this information to kill the virus? OK. So what we now know happens is that the DNA can be transcribed into RNA. OK?

And so you have this piece of RNA with the repeat and the spacer. And then you can also transcribe the transactivating RNA. And they form stem loop structures. That's what those little

things are.

So they have palindromic sequences-- so, you know, the base pair, that's why they draw the picture like that-- and Cas9. OK. And so what we know now is that this strand of RNA, this pre-CRISPR RNA can interact with the transactivating RNA. OK? So they have a way of hybridizing to each other.

And that's what you see here. So you see this little purple repeat. And you see the hybridization there. And these two pieces of RNA can bind to Cas9. OK.

So Cas9 is the nuclease. And in this particular type of CRISPR, there's a ribonuclease, RNase III, which takes off all this stuff. So you only have a single spacer that's actually going to be recognized.

OK. So you typically could do this. You could do this again with different pieces of DNA and make many of these things, OK, and do many cuts. And so that's why people in engineering are excited about this.

You can do more than one cut at once. But we're just going to focus on a single set of cleavages, double-strand cleavage, with one spacer. And in this case, the spacer is brown. OK?

So we trim it. OK. And so this is our machine, two pieces of RNA and a protein. And then it goes searching for what happens.

The virus invades. The virus has this sequence somewhere in its genome. Somehow, the bacteria knows the virus has invaded. It makes this machinery.

It goes searching for this sequence. This sequence then is recognized, because it can hybridize to one of the two strands of the DNA. OK. And the nuclease then simply cuts it in two pieces.

So the idea is simple. I mean, obviously, this is an extremely complex process where it's going to be regulated at every step along the way. But, somehow, bacteria have figured out that, you know, if you have a virus that infects the bacteria, what often happens is the virus causes cells to lyse.

And the bacteria is dead. OK? So that happens. So to save yourself, you want to get rid of the

virus. OK. And so this is a way that bacteria have evolved to be able to kill this invading virus that would otherwise kill that, which is what adaptive immunity is all about.

OK. So this is the model. And so then what people have been focusing on and what was focused on in this paper is Cas9. OK. So the idea's easy. It cleaves double-stranded DNA and gives you blunt ends. Furthermore, it knows where that occurs relative to a P-A-M site, a PAM activating site.

It cuts in a certain region. So they've studied all of that. They know where it cuts. I'll show you that in a minute.

And then if you want to target any gene inside the cell, you now can put in the right spacer. OK? Then you put the whole thing together. Now, the key issue is getting all of this stuff, the protein and the two pieces of RNA.

They are going to go in as DNA, OK, getting them into the cell. And how did they get-- because if you can't get it into the cell, you can't do the double-stranded cleavage. So how did they get this into the cell then? Anybody notice that in the paper?

AUDIENCE: Adenovirus.

JOANNE STUBBE: Yeah, so adenovirus. So people are trying to do gene replacements using all kinds of methods. None of this is trivial.

In this case, they're working on a mouse liver. And adenovirus, I don't know very much about adenovirus. But, apparently, it likes to live in the liver. So that's one of the reasons they chose looking at the mouse liver, but it happens to also be where all the cholesterol metabolism or the predominant cholesterol metabolism occurs as well.

And so they wanted to try other things. In the end, you're probably never going to be able to use adenovirus. People have been trying to do that for years for gene replacement without success.

So a key issue is going to be how do you get this into the cell, I mean. And so that's what a lot of people are trying to focus on now. But to do this in tissue culture, you can do it without any problems. There are ways of getting it into the cell.

OK. So this is what we were talking about. Once you get the cleavage, you know, you can

repair the cleavage by this method. If you want to read about this, you can go read about it. But what this does is gives you a deletion of your gene. Or, if you have a template, then you can use this template to remake a protein with a mutation in it, for example, or with a tag on the end, so you can purify it by affinity column, chromatography.

And so then the question is in almost all cells-- and it depends on the organism-- you have both mechanisms of repair. And so one of the issues is how do you tweak the repair depending on what function you want to use the technology for to do this or to do that. And a lot of people are studying that. That's one of the focuses of many labs if you look at what are the issues, where are we going.

OK. So if you look here, we were just talking about Cas9. So the blue is one nuclease. And the green is another nuclease.

And in this case, this is the double-stranded DNA target they're after. And this little piece here, this TGG, is called the PAM site. And that's required for recognition by the Cas9 protein.

And do we understand the basis of that? The answer is yes. I don't know. Did any of you hear Jennifer Doudna talk?

Yeah. So, I mean, she's the one that discovered that. She just published two weeks ago the structure of this complex. And so, you know, I haven't had time to study it. And this is the kind of thing that, if you really want to use this, you got to roll up your sleeves and get in there and study it.

But what they did was instead of having the guide RNA and the transactivating RNA, they put them together. And so that makes the genetic engineering simpler. But the question is how do you put them together. OK-- not trivial. OK.

And the whole Eric Lander article about the history of this process, the Doudna, Charpentier group, figured out how to do this extremely efficiently for bacteria-- doesn't work in humans. So Lander's article is focusing on, you know, humans is much more important. And so Zhang, who is at MIT, had figured out another way.

You can't just use these two little pieces of RNA. You need something much bigger to have the Cas9 work successfully inside the cell. And what's the basis of that? I don't know. But they did a lot of experiments to figure out how you could get this to work most efficiently.

So these are the two partners. And the question is what's going on. And, you, know, frankly I haven't even had time to digest. My lab hasn't used this technology. I haven't had time to digest it.

But what you see-- let me just point out one other thing. There are two domains. So you have the nuclease domain, which are not contiguous.

And then you have a helical domain. And Doudna used Cryo-EM, which we've talked about at 25, 30 angstroms resolution-- not particularly good-- to show that when you started with Cas9, but you added the two pieces of RNA, you got a change in conformation. They could see that in the Cryo-EM, because it was huge.

OK. And then when you add the targeting DNA, what happens is the nuclease domains change tremendously. The conformation of the protein changes tremendously. Putting everything together, you have the double-stranded DNA. So here you have the double-stranded DNA.

Here, it's hybridizing to the guide RNA. And this is the tracr transactivating RNA. And the Cas9 simply surrounds this whole thing. That's what she talk about in the lecture this past week or past two weeks.

So this was at 30 angstroms. And they had that model. And this was the result of an atomic resolution structure that just came out a week ago.

And it just shows you, I've told you, you know, you have to separate the strands. I've given you a cartoon of that. And you need to stare at this a long time. But green is one nuclease. Blue is another nuclease.

One can see that the blue and the purple are the DNA. And you can see the two strands separating, because they need hybridize to two different parts of the guide and the tracr RNA, which is in orange. And so what they're doing is looking at a model for how this works. OK.

So the key issues, all of which were covered in this paper in some form, are shown here. And these are the key issues that everybody is facing. And I'm already over.

But the delivery method into the cell, OK, we talked about that. They use adenovirus in this paper. So I would suggest you go back and you look at what they did. OK.

Off target effects, did they look at that? Does anybody know? Did they look at that in this particular paper? Yeah, they did.

And so how did they figure out what's going to be off target? How did they choose what to look for? I mean, you know, you got a billion base pairs, right? So how did they tell what to look for?

One of the first things they did is what? How did they target the PCSK9? How did they figure out what to target? Can anybody tell me that?

All right, nobody can tell me that? I guarantee you're going to have this on the first exam. You're going to have something on this in the first exam. We'll see if you go back and you read this.

So the whole paper, really the whole first couple figures, is focused on how do you decide what to target in the PCSK9. And they looked of exon 1. And they looked at exon 2. And then they did experiments to try to look at that.

That's how they got this idea about what the mechanism of repair was. And the sequence they targeted, they then look for other sequences that were three or four base pairs different. And then they also did PCR reactions on all those genes to see if you got cleavage or not.

So if this is ever going to be used technologically in humans, which is the goal of this paper-- you know, we're very far removed from that. We have a lot of ethical questions. The bottom line is you need to remove all the off-target sites.

You need to control, as we've already talked about, the two methods of repair of the double-stranded breaks. And I think now that we have structure, we ought to be able to even better design these three pieces to make more efficient chemistry of cleavage. I mean, it's amazing how efficient this was.

So they did the whole thing in four or five days. I mean, that was really quite amazing. And so what I suggest you do now in the rest of the paper is, I think, straightforward.

It just tests this model by looking for what happens to low density lipoprotein cholesterol. What happens to the receptor? What happens to cholesterol levels? What happens to triacylglycerol levels? And does it conform to the model that people have for the function of this protein in controlling cholesterol levels?

So what I would suggest you do is you go back now. Hopefully, you're now interested in this a little more. And go back and read this. And if anybody has any questions, they can come back and talk to me.