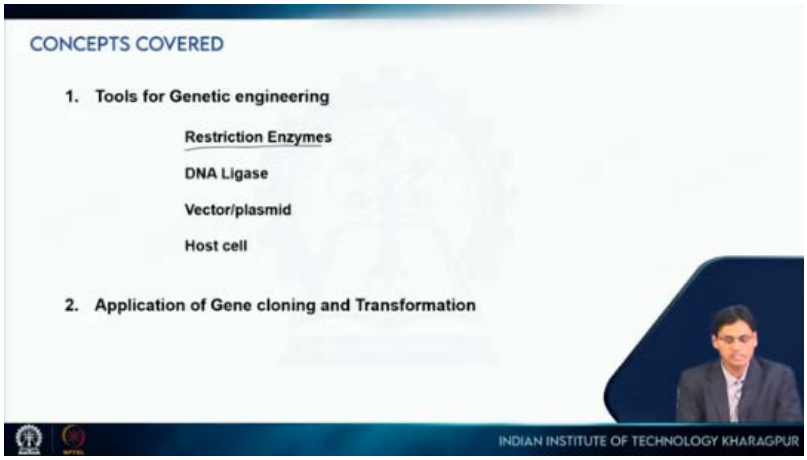
Introduction to Complex Biological Systems
Professor Dibyendu Samanta and Professor Soumya De
Department of Bioscience and Biotechnology
Indian Institute of Technology, Kharagpur

Lecture 54
Genetic engineering: Methods and applications

Hello everyone, we are discussing genetic engineering now. So, today in particular I will focus on fundamental methods of genetic engineering. Here I will be mostly focusing on different tools involved in genetic engineering such as restriction enzymes, DNA ligase, plasmid and a variety of host cells. If you see genetic engineering started, I would say after the 1970s. Part of the reason is genetic engineering, so you have to somehow cut and paste our genes for the engineering purpose.

Now the problem is the dimension of DNA in between 2 nucleotides; we have only a 3.4 angstrom gap. So, as a result of that conventional engineering tool that we cannot use to cut DNA precisely at a specific sequence, it is almost impossible. So as a result, only after the discovery of a few basic ingredients, for example restriction enzymes that can only precisely cut DNA in a specific sequence, and then DNA ligase: its utility and how it works. When we came to know all those things together, then the field of genetic engineering evolved. Then I will also be discussing the application of gene cloning and transformation.



Particularly restriction enzymes so those are also commonly called restriction endonucleases. So, they are naturally produced in bacteria. So particularly I would like to

mention here that these enzymes are present in some bacteria not present in all bacteria, but they are helping bacteria for their defense mechanism. So, if you see here, restriction enzymes are part of what you can say is their kind of immune system.

So, as you know, the phage virus means bacteriophage. They attack bacteria and multiply inside the bacterial cell. Now, these bacteria have some weapons, for example, this restriction endonuclease, and this enzyme digests foreign DNA. What happens when the phage virus attacks bacteria? They inject their DNA, phage viral DNA, inside the E. coli cell, for example, or any other bacteria. Now, this enzyme will degrade this foreign DNA. So, as a result, they cannot multiply inside that bacteria or kill the bacteria. So, this is some kind of their own strategy to save them, but we are utilizing this enzyme for our own good. So, now, if you see, as I already mentioned, they destroy bacteriophage DNA in bacterial cells. Also, it is very important to note here that they cannot digest host DNA with methylated C. That means the question now is: the same enzyme is present in bacterial cells, and bacteria also have DNA.

But then why is their DNA not getting digested? So, in bacteria, there are different types of modification systems. They modify their DNA in such a way that this endonuclease can understand. So, they will not cleave or digest their own DNA. They will only mount a response against foreign DNA, for example, in this case, the phage virus DNA. Now, if I mention a little bit here, the restriction endonuclease is somehow restricted by DNA sequence. This enzyme will not cut everywhere. It will specifically recognize a certain sequence. There are different types of restriction enzymes available in bacteria, and nowadays, those enzymes are also available commercially. For our own molecular cloning purposes in the laboratory, we just purchase those from some company. That is all.

But the thing is, if you see from the name that is endonuclease. So, nucleus means something which will cleave nucleic acid that is nucleus. Endonuclease means, for example, if you see this is a DNA double strand. So, I am just putting the direction here.

So, if this is DNA, then exonuclease will cut from one side; they will cut DNA from one side, but endonuclease cuts from the inside. That is why their name is endonuclease or

restriction endonuclease. Now, what is their substrate? I already mentioned their substrate is DNA, and they recognize specific nucleotide sequences in DNA and cleave the bond between two nucleotides. So, the bond between two nucleotides means not the hydrogen bond; hydrogen bond is just for the base pairing in DNA. So, the bond between two nucleotides means the covalent bond, the phosphodiester bond present in between the two nucleotides. So, that can be broken by this enzyme, restriction endonuclease. So, now, based on their action, there are many types of restriction enzymes. So, if we classify them broadly, we would say blunt-end cutter and sticky-end cutter. So, for example, if I say this is one DNA here.

So, now if you see here, the sequence CCCGGG, and the complementary sequence should be GGG CCC, and then here it is going like this, 3′. I am not writing the whole sequence here, 5′ and here 5′ to 3′. So, this is a recognition sequence of a particular enzyme, and this enzyme is known as SmaI. So, you do not need to remember this enzyme's name, like where they were isolated from some bacteria and their initials or other details, but it is not important to remember these names. But what is important here is that this enzyme cleaves this DNA. So, these nucleotides have phosphodiester bonds between them.

So, now, it will cleave here. As a result, this enzyme will work on this sequence and cleave this DNA, which means it will break two phosphodiester bonds, one here and the second one here. So, two phosphodiester bonds will be broken. So, as a result, what will be the end product?

You will get something like this: 5′ CCC, 3′ GGG 5′. So, these two fragments are produced now. That is why this enzyme, SmaI, is called a blunt-end cutter because, as you can see, the ends are blunt, not sticky. When I explain sticky ends, you will understand even better. So, let me give you one example of a sticky end.

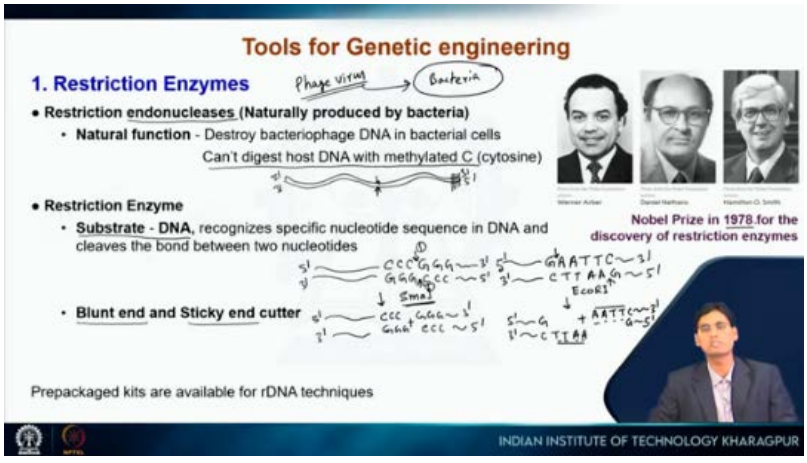
So, now if I say this is another sequence I am writing here 5′ GAATTC and then 3′. Similarly, 3′ you have some sequence here I am not writing just CTTAAG 5′. So, this is another portion of DNA I just wrote the sequence here. Now, if you see another enzyme, EcoR1.

This is another very common restriction enzyme we use in our lab, in every lab I would say. So, this enzyme will cleave in this way. So, here it will break one phosphodiester bond and here it will break another phosphodiester bond. So, as a result of that what will be the product? Now if you see the product will be something like this.

So, 5′ G and then 3′ CTTAA plus here AATTC 3′ and then you have G 5′. So here you can see that their end here this part is why we say sticky end because again this end can be joined together because they have some complementary sequence overhanging. So, as a result of that you can see this AATT. They can again form base pairing with this TTAA. So, they are complementary to each other. So, that is why they are called sticky ends.

Particularly, this enzyme, like sticky end cutter, we use very often in our laboratory for genetic engineering purposes. Blunt end cutters can also be used for some special purposes, but sticky end cutters are very common. As I already mentioned, these enzymes are available as prepackaged kits, and you can buy them commercially. Now, I should mention that the discovery of restriction enzymes and the understanding of restriction enzymes are very important for genetic engineering. As you can see, the scientists Werner Arber, Nathans, and Hamilton Smith received the Nobel Prize in 1978 in Physiology and Medicine for their work on restriction enzymes.

Without restriction enzymes, we cannot do genetic engineering. So, this is very important and now the next tool. So, we just discussed restriction enzymes; the next one is ligase. So, I mentioned.

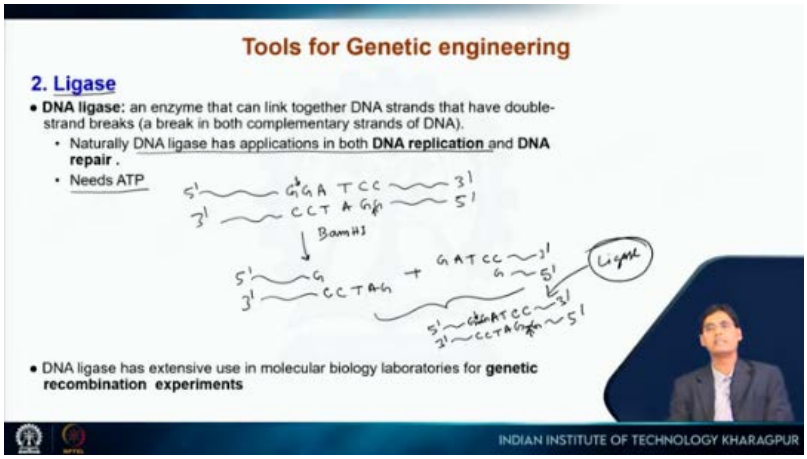


In genetic engineering, we have to cut and paste DNA. So that you can modify some DNA, you can add a gene to somewhere in a different cassette of DNA, so as a result, not only do you have to cut it, but you also have to paste it. Restriction enzymes can be used to specifically cut a DNA segment, but now you need another tool to paste them, and that is ligase. So, DNA ligase is an enzyme that can link together DNA strands that have a double-stranded break, meaning a break in both complementary strands of DNA. For example, when I explained in the previous slide how restriction enzymes break the phosphodiester bond in DNA; you noticed that restriction enzymes break two phosphodiester bonds in both the strands so as a result of that one phosphodiester bond in each strand. So, a total two phosphodiester bonds are breaking. So this double stranded break can be ligated again by this enzyme. That is why this enzyme is called ligase, and naturally this enzyme has a role in DNA replication, as you can remember in the first week we discussed replication. That time I told how ligase finally seals the gap between short segments of DNA. This is the function of ligase naturally, and for this work it needs ATP or adenosine triphosphate for energy.

Here, DNA ligase also has extensive use in molecular biology laboratories as I already mentioned for genetic engineering experiments. So, as a result of that, for example, if I am giving you another sequence here like 5′ so some sequences are present here, then I would say GGATCC and some more sequence here CCTAGG. So, again this is our DNA and we are showing only a portion of sequence which can be recognized by one enzyme that is called BamH1. This is also a very common enzyme. Very often we use this enzyme and this enzyme will cut here. So, as a result of that after digestion if you treat this DNA with this enzyme BamH1 then you will get something like this GCCTAG plus GATCCG 5′. So, this is done, but now as I mentioned they have a sticky end. So, as a result of that

if you miss this fragment in higher concentration, what will happen? So, they can again come to each other. They can again fall back based on their sequence. So, as a result of that, you will be seeing something like this: 5′G 3′CCTAG, then you have here GATCC 3′ G 5′, something like this. But here, the problem is the phosphorylated bond is not yet formed because that has already been cleaved by restriction endonuclease. So, as a result

of that, here you have to add the ligase enzyme. Then it will form this phosphodiester bond and it will happen in the presence of ATP or adenosine triphosphate. Now, this is a very important slide: the restriction enzymes and ligase are the key players in genetic engineering.



So, this cut-paste mechanism you can do if you use these two enzymes properly, the restriction enzyme as well as ligase. I am trying to explain to you with some examples; let us see. So, if this is a DNA and I am putting this sequence here: GAATTC. Similarly, here: CTTAAG so 3′ and 5′. So, now, this DNA, if I treat that with one enzyme called EcoR1, I already mentioned.

So, if I digest this DNA with EcoR1 and I assume that this EcoR1 site is not present in the rest of the DNA, only it is present in one portion here, as I mentioned. So, that you have to check during the cloning and genetic engineering experiment. So, as a result of that what will happen? So, after digestion, you will get 5′, and this fragment is then cut here so as a result of that G and then C then TTAA. So, this is the big fragment here.

So, I am just mentioning here, this is the large fragment and this side is the small fragment. Now, I am not writing the sequence again and again to save some time. So, now, this is the small fragment. Now you can actually separate these two fragments based on some other technique that you have already learned, for example, gel electrophoresis.

After this digestion, if you run the product like the digested product in agarose gel electrophoresis, then what will happen? So, this is the agarose gel. So, this is the negative

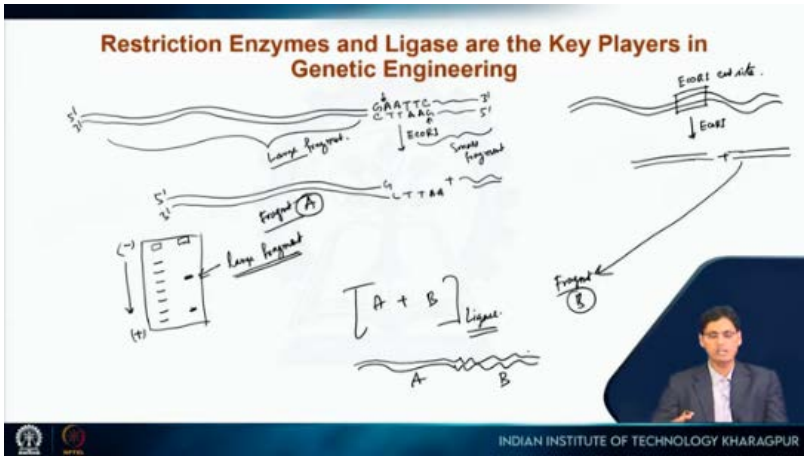
side, this is positive. So, you loaded your sample somewhere here. So, after the gel run you will be seeing that one band is here, one small band is here.

So, this is the large fragment. Now, the thing is you can load here some ladder that the molecular weight you know. So, on the basis of that you can actually detect your fragment of interest. Now the question is if this large fragment, if I say this is Fragment A. Now you have another DNA molecule; I am not writing the sequence here. So, this DNA you have also cut site the same cut site for example, EcoR1 cut site.

So, as a result of that you can digest this DNA with the EcoR1 enzyme, then again you will get two fragments, and now just run a gel and again you purify these fragments, and as a result of that, finally what you can do, you can take this fragment after purification, for example fragment B and here you have fragment A. So, you just mix this fragment A and fragment B in presence of enzyme ligands. So, as a result of that you will be seeing that A, this is A part and this is B part just for representation I am drawing this way.

So, this way now A and B should be attached together. So, this A and B they are coming from different DNA but you just use the same enzyme. So, they have the same sequence as their sticky ends, and then again, those sticky ends can attach to each other based on their best complementarity, and now if you add ligase as I mentioned here, then A and B should be attached together. This is the basic technique of genetic engineering, then only a little bit more complexity and modification, that is all. But this is the basic idea of how you can cut and paste DNA molecules together.

So, you have to select the proper restriction site and then design your experiment that way. Because you have to be careful that the same sequence is not present in the rest of the DNA; otherwise, your enzyme will cleave here and there, and you will get many fragments, which will complicate your experiment. So, you have to be careful about that. Now, this is all about restriction enzymes and DNA ligase. Another very important thing is the vector.



So, here, particularly, I am mentioning plasmids. So, plasmids are examples of vectors. So, what are plasmids? So, plasmids are small pieces of circular DNA used for cloning. So, as I already mentioned before, like vectors.

So, you have to put your gene of interest inside a vector because that vector must be transferred into a host cell so that you can produce your product, the protein of interest, for example. I will explain all those things soon. The thing is the plasmid. So, if this is E. coli. So, this is the bacteria E. coli, and now, E. coli have their one circular DNA that is the chromosome. This is the bacterial chromosome. Similarly, all other bacteria have this kind of chromosome or DNA. But plasmids are small pieces of circular DNA that are extrachromosomal, meaning they exist apart from the chromosome. Besides the chromosome, they have some additional small circular DNA, and this is called a plasmid. So, not all bacteria have plasmids, but some naturally possess them. These plasmids actually help bacteria acquire some additional properties for example, resistance against some antibiotics or sometimes resistance against heavy metals. So, there are many other functions of plasmids, as I already mentioned. Naturally, some bacteria have plasmids, and we use plasmids for our own interests in genetic engineering approaches.

So, why are we using plasmids? Why are they beneficial? I am going to explain a few points here. The first thing is self replication. That means it can replicate in the host that is independent of the origin of replication. So, it seems a little bit complicated.

The thing is like a bacteria cell, as I already mentioned; they have one chromosome, for example, E. coli. Now, whenever that chromosome replicates into two, the bacteria will

also divide into two. So, as a result, every bacterium always has only one chromosome. Now, plasmid replication is independent of chromosome replication. So, as a result, the number of plasmids can be higher in one bacterium.

So in one bacterium, it might contain only one chromosome, but it can contain multiple plasmids inside the bacterial cell. Now, the cloning site is the region containing multiple restriction sites so a suitable vector. I am just mentioning: say, for example, this is a vector and this is a plasmid. So, the first thing we mentioned is that it should have its own origin of replication, which is called the Ori site, the origin of replication.

So, it can replicate, and that is independent of chromosome replication, as I already mentioned. Now I am saying the cloning site. So, I would say if this is the cloning site. The cloning site means here you have multiple restriction sites sequences. Why? What is the purpose?

So this is the cloning site. Sometimes we say MCS, that means multiple cloning site, that means here you have some sequences which can be recognized by different types of restriction enzymes, and what is the purpose? So we will use a specific restriction endonuclease, and we cut this plasmid, and then we will put our gene of interest inside this plasmid. So, that is our goal. I will explain in more detail soon. So, that is our goal.

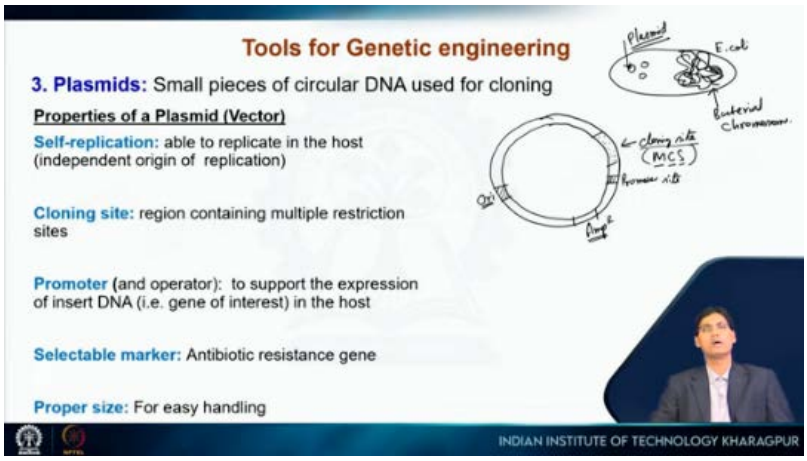
So, multiple cloning sites. So, natural plasmids, in their sequence, you might not get many restriction sites there. But, the thing is, nowadays, whatever plasmid we use in the lab are kind of engineered plasmids. So, this region like a cloning site is engineered in such a way that you have an option of many different types of restriction enzymes you can use and you can cut that particular sequence. So, this is a little bit engineered version of plasmid nowadays we use.

Now, the promoter, so promoter and operator, all those things are important for gene expression, as you already know, to support the expression of insert DNA. For example, you are interested in making some protein and the gene of interest you are working on. That gene of interest can be put into this plasmid cassette and then if you are putting it properly under some promoter so that in the right condition this gene will be expressed

into mRNA or messenger RNA and this messenger RNA will be converted into protein also so translated into protein.

So, as a result of that, you should have promoter operator all those sites for proper function and then selectable markers so this is also very important, selectable markers for example, some antibiotic resistant gene for example, the ampicillin resistance genes. So, ampicillin is an antibiotic and this should be present in the plasmid itself, and then this ampicillin resistant gene will provide resistance to that bacteria, and then proper size for easy handling.

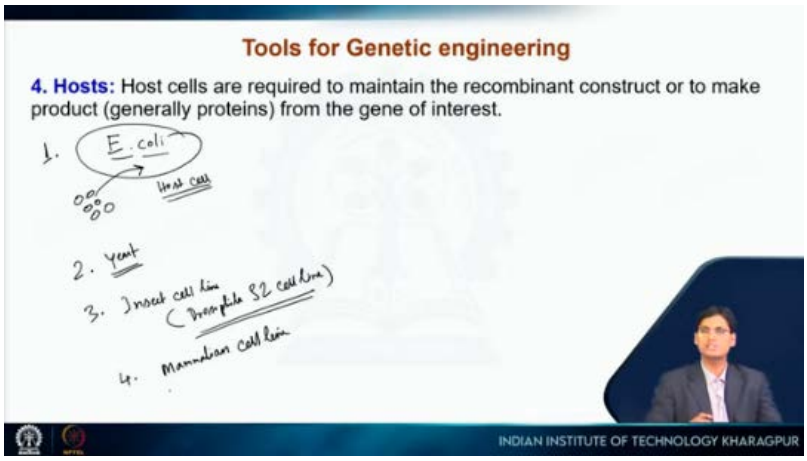
So, if it is too big then it is a little bit difficult to handle. So, that is why another important thing is that the size should be somewhere it will be more useful, not very big or very small. So, this is your promoter site which I mentioned before. So, these are the basic features of plasmid which are very beneficial for our genetic engineering approach and then host cell. So our goal is, for example, you are interested in some protein. So the corresponding gene you clone inside the plasmid, as I already just mentioned about the plasmid. You just put your gene of interest in the plasmid. Now the question is: you have to express the protein. So as a result of that, you have to transfer this plasmid into some host cell.



Very frequently we use the laboratory strain of E. coli or Escherichia coli; this is the laboratory strain, a non-pathogenic strain. So, you can use this E. coli strain as a host cell to express your protein of interest. So, as a result of that whatever engineered plasmid that means plasmid With your gene of interest, you have to transfer that into the E. coli

cell, and then if you grow the E. coli cell and if you put proper conditions, the E. coli cell will make protein for you because that protein-coding gene is already present in the plasmid; that is the overall idea of this experiment. Apart from E. coli cells, we can use different types of cells; for example, this is a prokaryotic expression system where we use E. coli, and many proteins we work with are eukaryotic membrane proteins that might not express in E. coli cells or might not fold properly.

So, as a result of that many times we need to use a eukaryotic expression system. Many times we use yeast expression systems. You know that yeast is a eukaryotic cell like fungus and then we sometimes use insect cell lines. For example, Drosophila S2 cell line. This is very commonly used for eukaryotic protein expression systems and on top of that we have defined types of mammalian cell line as well. So, it all depends on your application what kind of protein you are interested in and whether you can express it in bacteria, then you have to rely on some other system. So, bacterial expression is much cheaper that is why sometimes we start with a bacterial system, but the thing is if your protein requires a lot of post translational modification which is not present in the bacterial system then you have to rely on a eukaryotic expression system. So, this is all about different types of host cells that can be used for different applications.



Now, cloning and transformation of a gene. So, here we are trying to explain to you whatever I mentioned in a brief way. So, as you can see this is the plasmid and it contains an ampicillin resistance gene and this is the DNA fragment as you can see. Now, you can

cleave this plasmid with a particular restriction enzyme and the similar set of enzymes you have to use to get this DNA fragment as well.

So, as a result of that finally, you will be getting this recombinant plasmid. Why is it called a recombinant plasmid? Because this plasmid is not the natural plasmid anymore, you put your gene of interest here, the recombinant plasmid. Now, you have to use this plasmid for transformation; transformation means E. coli cell. So, you have this plasmid and you have your gene of interest here.

Now, if you transfer this into an E. coli cell, this E. coli cell has its own chromosome, but this E. coli cell is not antibiotic resistant. So, as a result of that, this E. coli cell, if you put it in some media, if you grow this E. coli cell in some liquid media, for example then If that media contains antibiotic ampicillin, for example, then this bacterium will not survive, but now if you have this plasmid inside this bacterium. So, this plasmid has an ampicillin resistance gene.

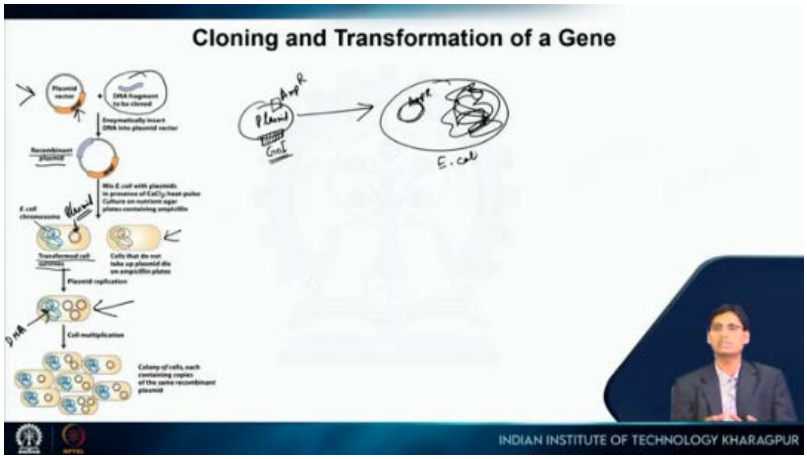
So, as a result of that If the plasmid is present inside this bacterium then this bacterium will survive in presence of ampicillin. So, that is why you can actually select which E.coli cell you have your plasmid and intentionally you put your gene of interest in the plasmid itself. So, as a result of that, whenever the plasmid is going inside the cell, you are also transferring your gene of interest. So, that is what you can see here.

After this transformation, you can see that this is a transformed E. coli cell. It has the plasmid, as you can see here; this is the plasmid. So, as a result of that it will survive in the presence of ampicillin, but this one, in this cell you cannot see any plasmid present here. So, as a result of that, this cell will die in the presence of an antibiotic.

So here we are using antibiotics for selection purposes. So, we want bacteria which contain plasmids, and you have already engineered the plasmid in such a way that every plasmid contains your gene of interest. So, as you can see here in this cell, this is the DNA, the bacterial DNA, and you have a plasmid in this bacterium so now if you grow this bacterium. In bacteria, the division cycle is very fast. Very quickly, the bacteria will multiply. If you keep your antibiotic concentration at such an extent that the bacteria will survive only if they have a lot of antibiotic-resistant genes, and since antibiotic-resistant

genes are present in their plasmids. So, they will replicate their plasmids. They will try to keep the number high; the plasmid number high so that they can defend against this antibiotic ampicillin. If the number of plasmids is high, that means the gene of interest is also present in every plasmid. So, that is the overall idea.

So, in every cell, you have multiple plasmids, and all these plasmids contain your gene of interest. Now, you can express your protein of interest. So for example, when we discuss the Lac operon, I mentioned that some inducers can be added. So, similarly, you can clone in such a way that you can control the expression of your protein of interest. Now this is a little bit of an application purpose: the production of a protein of interest through a genetic engineering approach.



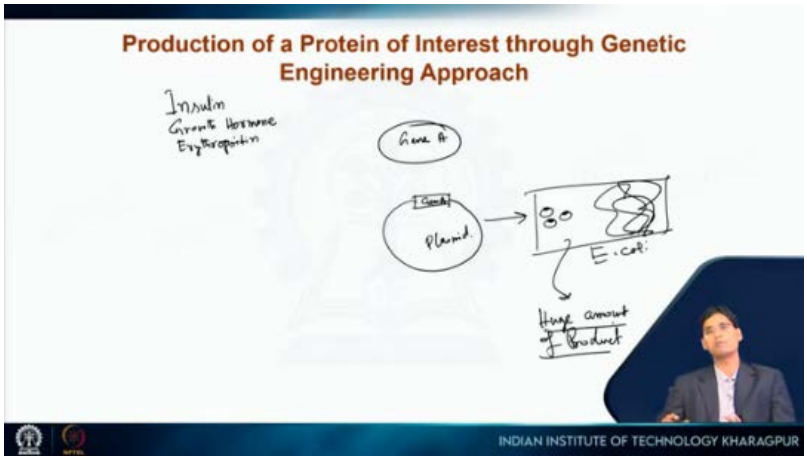
Whatever I mentioned based on that discussion, I am going to say that many proteins, even human proteins, are produced in bacterial systems. For example, I would say insulin. All of you know that insulin is produced in bacterial systems, human insulin, I am saying, and then, many growth hormones, and another example I would say is erythropoietin. So, all these proteins can be produced using recombinant DNA technology or genetic engineering. Then the advantage here is, for example, insulin or any other human protein in one cell in our body. In one cell, this information to make this product is written only twice, so we are actually diploid.

So, we have 23 pairs of chromosomes. So mostly I would say, if I say gene A, I am just giving one example; gene A is present in both pairs of chromosomes. So, as a result of that, I would say it is present just two times in every cell. Now, the thing is when you are

cloning this gene A inside this plasmid. So, now you clone this gene A in this plasmid. Now there are different types of plasmids. This plasmid can be a high-copy-number plasmid so that plasmid can exist in very high copy numbers in one bacterial cell. So when you are putting this plasmid in some E. coli cell. So, this is, for example, some E. coli cells.

So, this is the plasmid here and this plasmid contains your gene of interest here, gene A. Now when this E. coli is growing in some liquid media, if you keep proper antibiotic concentration because your plasmid is providing resistance against antibiotics, as a result, the number of plasmids will be very high. Sometimes the number of plasmid can be more than 1000 per E. coli cell.

So, now you can understand that even a small cell like E. coli, but it contains 1000 copies of your gene of interest that gene A. So, as a result of that you will get a huge amount of product, the product here means protein. So, the beauty of this system is that the recombinant DNA technology allows you to clone a gene of interest in a plasmid and then send this plasmid into an E. coli cell, a laboratory strain of bacteria, for example, and now you can give instructions to make the protein. So, that can be done and if you properly keep the antibiotic concentration and if you select the correct plasmid in every cell the number of plasmid can be as high as 1000 or more. So, as a result, you can understand that a huge amount of protein you can make is not possible from a eukaryotic cell without cloning if you try to purify this protein from some tissue sample or somewhere else. So, this is the power of genetic engineering or recombinant DNA technology.



You can go through any of these textbooks for that additional understanding and that is all.