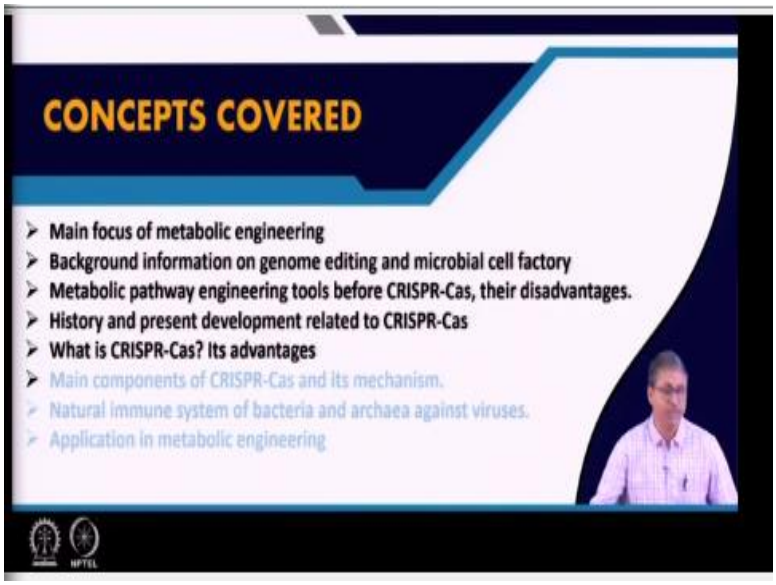
Metabolic Engineering
Prof. Pinaki Sar
Department of Biotechnology
Indian Institute of Technology-Kharagpur

Lecture-29
CRISPR-Cas System and its Application in Metabolic Engineering-Part I

In this lecture we are going to discuss about CRISPR-Cas system and it is application in metabolic engineering.

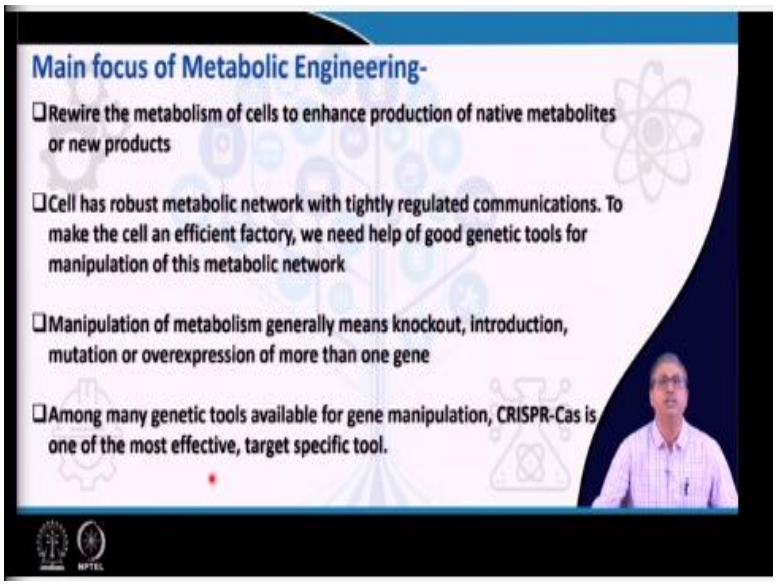
(Refer Slide Time: 00:39)



So, the major concept that we would like to cover during this deliberation and in subsequent follow up lectures would be recapitulating the major focus of metabolic engineering followed by background information on genome editing and microbial cell factory. And how metabolic engineering helped and progressed the overall improvement of cellular pathway engineering.

And the metabolic pathway engineering tools which were available before the CRISPR-Cas and their disadvantages which actually paved the progress of CRISPR-Cas as a genome engineering tool. History and the present developments related to CRISPR-Cas and what is CRISPR-Cas about the basic features of CRISPR-Cas and some of it is advantages will be discussed.

(Refer Slide Time: 01:38)



So, the main focus of metabolic engineering as we understand are to rewire the metabolism of cells to enhance the production of native metabolites or new products or even engineering the cells with respect to production of molecules or the properties of the cells, even the physiological properties are also included in metabolic engineering. Now generally it has been found at cell has robust metabolic network which is kind of intrinsic metabolic network with tightly coupled communications through different modes of cellular interaction.

Now to make the cell as an efficient factory as per our desire, we need to help of good genetic tools for manipulation of this metabolic network because it is highly complex and highly interconnected. Now manipulation of metabolism generally means a number of approaches which include the knockout of genes, knock-in or introduction of genes, introduction of different specific mutations or even over expression of more than one gene.

Now among many genetic tools which have been developed in the past 3, 4 decades and available for the metabolic engineering of any kind of cell or cell types. CRISPR-Cas is emerged as one of the most effective target specific tool for genome engineering which enabling the metabolic engineers to successfully engineer the cellular pathways in spite of the complex nature of the cellular metabolism.

(Refer Slide Time: 03:38)

Now metabolic engineering of cell factories has been driven by changes of genetic content of the cells. Fundamentally any metabolic engineer would like to change the genetic content by either introducing a small plasmid or a large plasmid or by changing the chromosomal DNA content. Now the spontaneous or induced random mutations which were conventionally used for a long period of time for the isolation and breeding of cells with improved properties or better production took a long time.

And effort and were actually non specific, what we have discussed earlier also. Transgenic technologies which have been developed in the past decades have made possible to introduce foreign gene that is the heterologous expression of genes and modified genes to implement a new metabolic pathway. So, in addition to individual gene alteration, alteration of the entire metabolic pathway was possible through different advense or different methods or different approaches of transgenic technology.

That the host cell generally they do not have or we would like to knock down if they have the genes, sometimes knock out the genes or knock down the genes or controlling their expression patterns in order to avoid the unwanted functions or unwanted flux distribution of the desired compounds. Now transformation efficiency of those cells which generally relied on the conventional transgenic technologies.

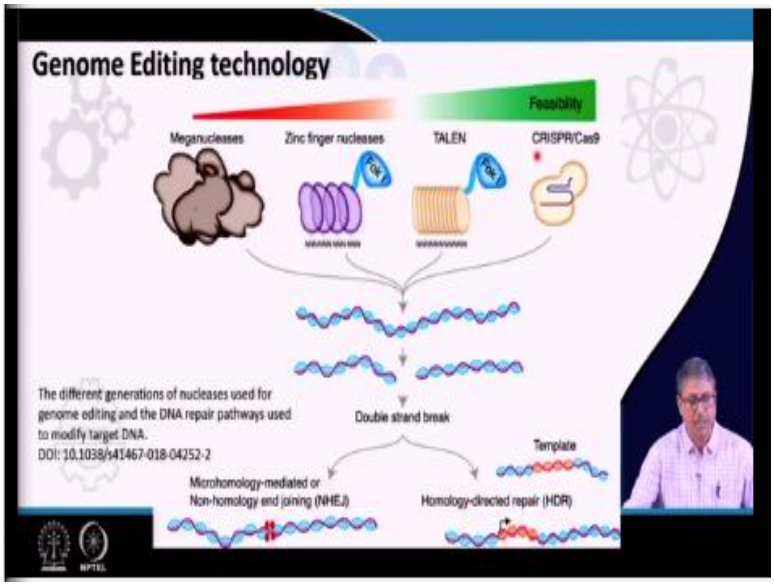
They are considerably varying depending upon the cell types because all cell types are all host cell types are not equally amenable for a high transformation efficiencies.

(Refer Slide Time: 05:30)

Now intensive studies have been done on the understanding of cell function and engineering for host production. Because microbial selection of microbial host and identifying the suitable host for genetic engineering became a very important aspect of microbial cell factory and intensive studies have primarily found that in selected organisms, such as E.coli and *Saccharomyces cerevisiae.*

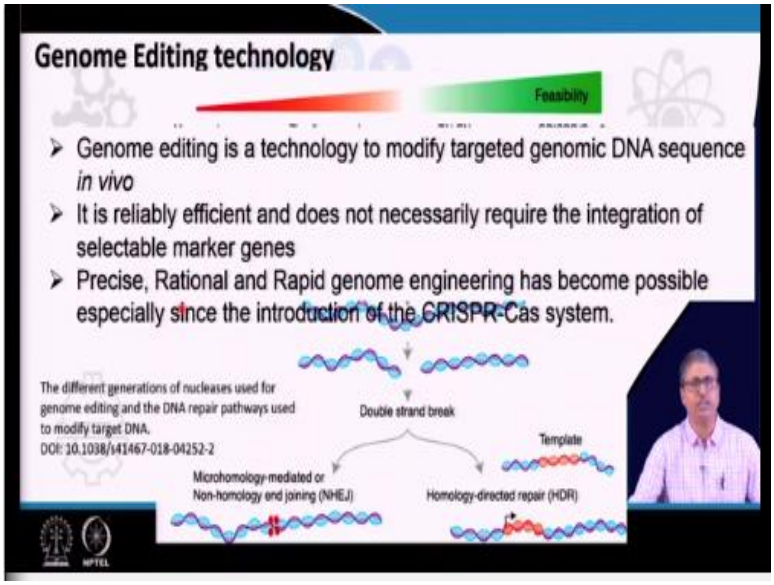
Now in addition, transgenic technology requires the selection of transformants using selective marker like antibiotic selection marker, limiting serial manipulation and multiplexing. So, these are some of the important limitations or bottlenecks which were identified during the course of application of transgenic technologies.

(Refer Slide Time: 06:17)



Now the genome editing technology which actually utilizes the different generation of nucleases to edit the genome and the DNA repair pathways, like the meganucleases or zinc finger nucleases or TALEN or CRISPR-Cas in the recent time.

(Refer Slide Time: 06:38)

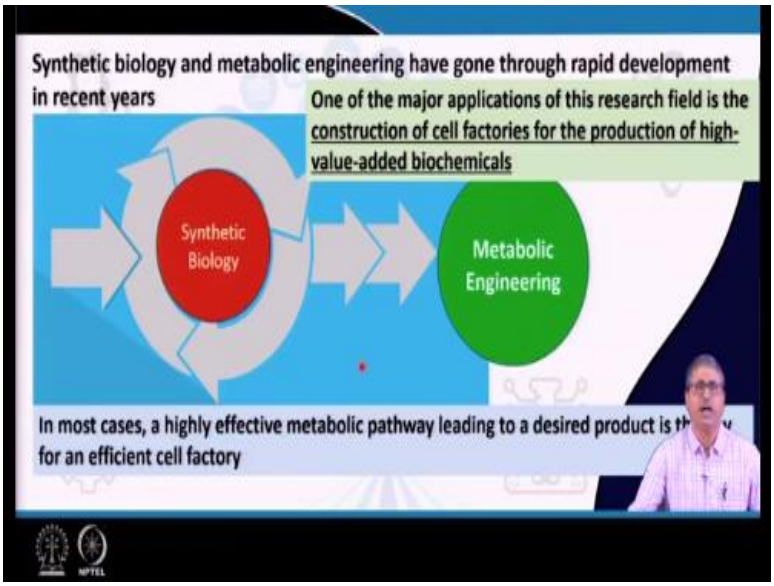


Enabled the modification of targeted genomic DNA sequences in vivo. So, compared to the use of different selection marker and all other time consuming laboratory procedure, this genome editing technology particularly that utilizes the CRISPR-Cas process that we are going to talk very soon. Enable scientist to modify the genome in a straightforward way and in vivo.

Now it is reliably efficient and does not necessarily require the integration of selectable marker genes. So, the issues encountered with the use of different type of marker genes were not there, it is also a precise. Like the use of CRISPR-Cas mechanism or genome editing technology is precise. It is rational because you can highly plan the entire process that which gene, where, when the gene has to be deleted or gene has to be knocked down or gene has to be knocked in.

All these processes can be highly controlled and rapid. The genome engineering tool has become possible especially since the introduction of the CRISPR-Cas system. So, CRISPR-Cas system enabled the scientist to perform a number of genome editing processes which were really very precise, very specific, highly rational and rapid.

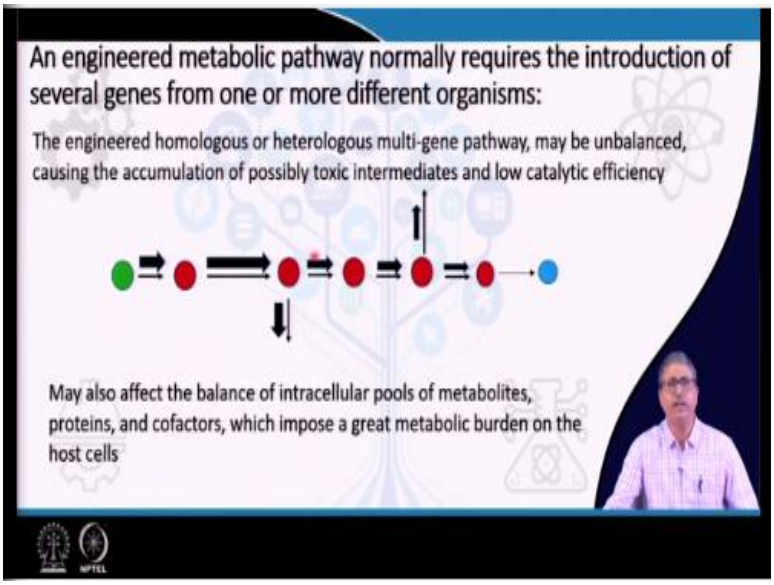
(Refer Slide Time: 08:03)



Now in the past decade particularly, we have observed a significant advancement with respect to synthetic biology and metabolic engineering. So, synthetic biology has helped the metabolic engineering and a prolonged growth, a rapid development has been observed in the past years, particularly in past decade or so. And what is most important over here is that in one of the major applications of this research field that is the amalgamation of synthetic biology and metabolic engineering for strain improvement is that construction of cell factories for the production of high value added biochemicals.

Particularly for a high product value or high value biochemicals, added biochemicals, microbial cell factories become very popular. And for that purpose synthetic biology become very evident and very crucial in metabolic engineering of such systems. Now in most cases a highly effective metabolic pathway leading to the desired product formation is and if is required for an efficient cell factory. So, we need to identify the most effective pathway and then manipulate that pathway in order to achieve the desired production of the high value biochemical compounds.

(Refer Slide Time: 09:31)



Now an engineered metabolic pathway normally requires the introduction of several genes as it is discussed and it is pretty well known. That metabolic pathways are quite complicated and there are requirements that more than one genes are often manipulated. And even genes from different organisms may be put together or brought into a single host in order to achieve the desired development in the metabolic pathway improvement.

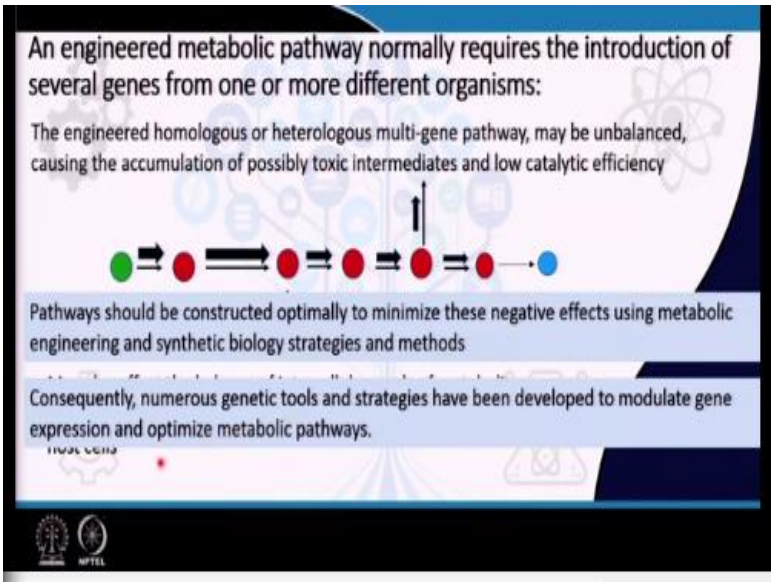
Now for example if we consider that here is the substrate which is the green ball, which is taken up by the cell and then subsequently with using the host native metabolic reactions, it is going to be converted to this product which is blue dot. Now different levels of metabolic pathway engineering through conventional methods have been introduced. And this could actually redirect the flux significantly in order to achieve the higher production of this particular desired product.

But while doing this kind of genetic engineering or genetic manipulation, we also encountered two important aspects or 2 important limiting factors of this particular process, is that the engineered homologous or heterologous multi gene pathways. So, it may be possible that all genes that we are engineering here in this reaction may not be from the same organism, just if you remember the process of production of artemisinin in *E.coli.*

So, you can actually bring the gene from eukaryotic host to the prokaryotic system, so ideally we can have multiple genes coming from different host. So, maybe it may be expected or it may be happening that an unbalanced expression of genes is achieved or occurring in the system, causing the accumulation of possibly toxic intermediates and low catalytic efficiency.

So, in spite of the genes being expressed there might be an irregularity in the natural balance or the homeostasis of the entire systems of the pathway like the distribution of enzymes, they are effective concentration etcetera, leading to accumulation of some of the metabolites which are actually toxic.

(Refer Slide Time: 11:56)



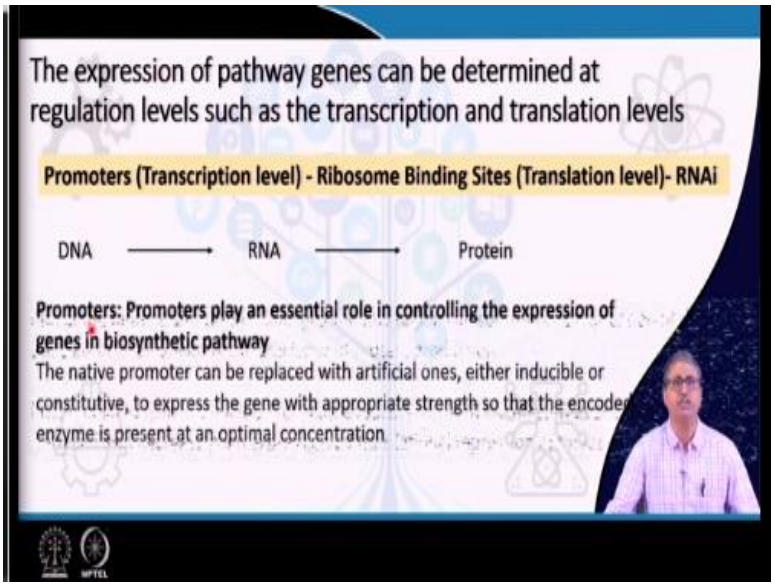
It may so also happen that the disturbance due to the balance of the intermediate products or intracellular pool of metabolites, some of the metabolites might be toxic and they will create a kind of metabolic burden to the host cell. Now when we do this kind of metabolic engineering considering source of genes from different host including the heterologous organisms.

So, pathways should be constructed optimally to minimize these negative effects during the metabolic engineering and synthetic biology strategy and method. So, whenever we develop the metabolic engineering strategies including the synthetic biology approach. We should be careful that this kind of imbalance or accumulation of toxic intermediates, toxic metabolites should be reduced or should be as minimum as possible.

Because they lead to imbalance and metabolic burden, which is considered to be a very important factor for the productivity and the well-being of the cells in it is final level of production system, when we are expecting the cells to be used as cell factories. Consequently numerous genetic tools and strategies have been developed, owing to this understanding or increasing knowledge that the conventional transgenic technologies might be resulting into some kind of metabolic burden.

So, there have been some caution and developments of genetic tools and strategies which are tried to modulate the gene expression and optimize the metabolic pathways accordingly.

(Refer Slide Time: 13:39)



Now in the subsequent slides we are going to talk about some of this very precise and very elegant gene alteration systems briefly. The expression of pathway genes, so it is multiple genes are there, can be determined at regulation level such as the transcription and translation levels.

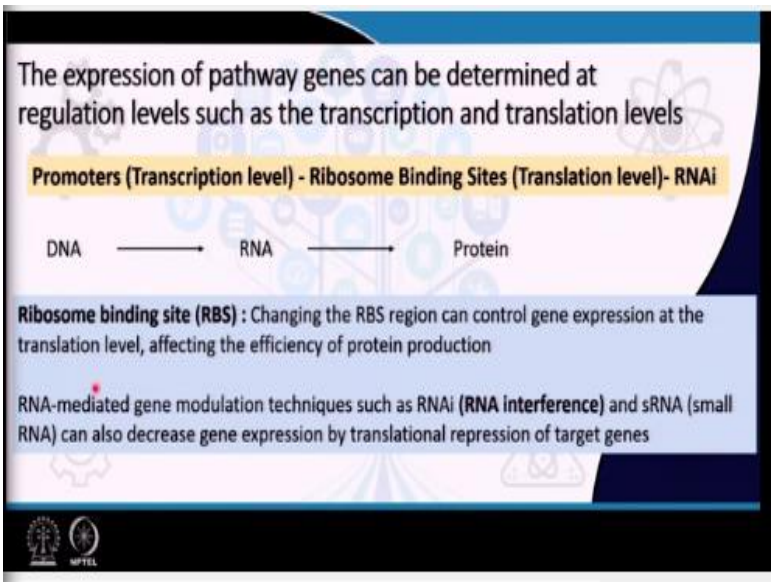
So, as we know this from the central dogma that the DNA double helix is allowing the transcription to follow first and forming the RNA.

And these RNA or RNAs are going to engage into the translation, forming the protein molecules. And these protein molecules will be processed to produce the effective biocatalyst that is the enzyme complexes and other things. Now 3 important aspects of engineering are envisaged in this aspect which is promoters. So, engineering the promoters at which will facilitate the expression of alteration of expression of genes or pathways involved in that at transcription level.

Then for the translational level modifications the ribosome binding sites and finally the RNA interference are targeted. Now the promoters are one of the most important and most successful sites for engineering the pathways, because, they play essential role in controlling the gene expression in any pathway, including the biosynthetic pathway. So, the promoters of different genes have been targeted for pathway engineering.

Now the native promoter can be replaced with artificial ones either inducible or constitutive or express the gene with appropriate strength. So, that the encoded enzymes present at an optimal concentration. So, there have been a significant research and development on this aspect that how to engineer the promoter. So, replacement of promoter the native promoter with advanced more sensible sensitive promoters which can be controlled in a much more strict way.

(Refer Slide Time: 15:50)

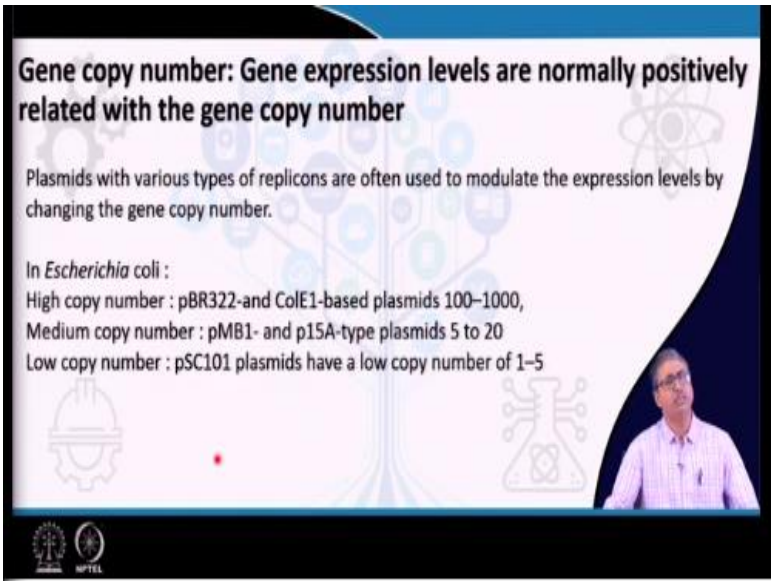


The next is the ribosome binding site manipulation which is basically enabling the metabolic engineer to control the translation of the specific pathway related enzymes. So, changing the ribosome binding site region may control the gene expression at the translational level affecting the efficiency of the protein production. Now also for the translational level modification of the protein production RNA mediated gene modulation technology.

Such as the RNA, RNAi interference and small RNA mediated interference, all small RNA mediated control of translation has also been attempted. And it has been found that they can also decrease gene expression particularly the RNAi or the sRNA mediated techniques, can decrease the gene expression by translational repression of the target gene. So, either we try to manipulate the promoter, replace it, modify it.

Or we control the ribosome binding sites to achieve some level of control on the translation of the particular pathway related enzyme proteins or we employ the RNA interference or small RNA mediated technologies, so to control often negatively or is a kind of a repression of the target genes.

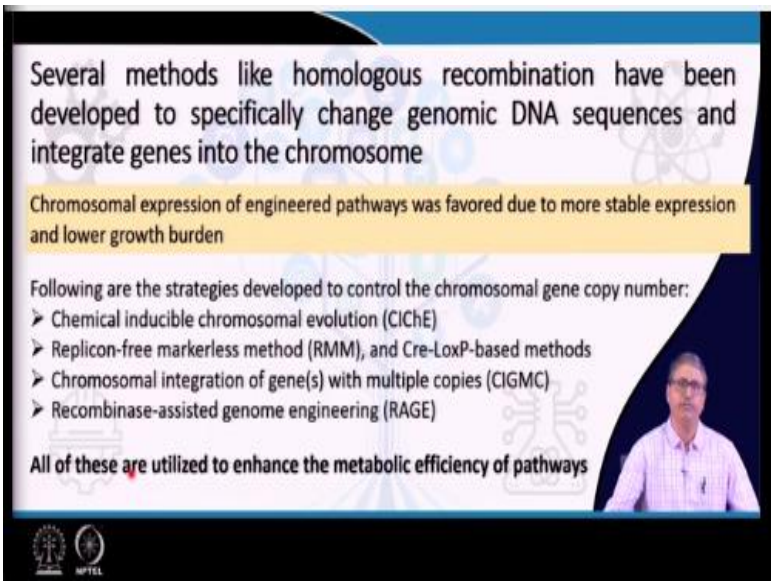
(Refer Slide Time: 17:15)



Now the other mode of controlling the gene expression or pathway engineering is the gene copy number control. Now gene expression levels are normally positively related with the gene copy number. So, higher the copy number of the gene it has been found the expression levels are normally found to be higher. Now high copy number of the particular target gene is generally achieved up to sometime using the plasmid.

So, plasmid with various types of replications are often used to modulate the expression level by changing the gene copy number. So, the larger the number of plasmid copies, the larger the higher will be the number of the desired or targeted gene. So, for example in E.coli, high copy numbers like pBR322 or medium copy number pMB1 or low copy number plasmids like pSC101 kind of plasmids have been used, tested and implemented successfully in a number of metabolic engineering attempts.

(Refer Slide Time: 18:16)



Now several methods like homologous recombination have been developed also, to specifically change the genomic DNA sequence and integrate genes into the chromosome. So, instead of working with the plasmid, because working with the plasmid showed some kind of issues and some kind of problems. So, homologous recombination based modification of the genomic sequences was attempted and we found some success in that.

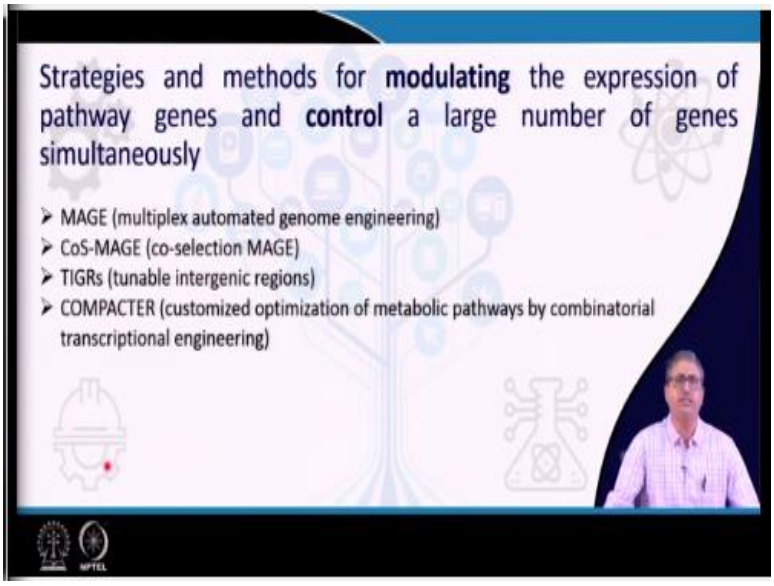
So, chromosomal expression of engineered pathway was favored due to more stable expression and lower growth burden. Because maintaining the plasmid and maintaining the copy number of the plasmid to the desired level was a challenge for the metabolic engineers. So, instead of doing that the scientist opted for often changing the chromosomal DNA configuration by relying on homologous recombination.

Now following are the strategies developed to control the chromosomal gene copy number. As we can see there are a number of chromosomal gene, copy number, alteration strategies were developed in the past decades or so which include, the chemical inducible chromosomal evolution or replication free markerless method or chromosomal integration of genes with multiple copies or even recombinase-assisted genome engineering.

So, a large number of reports are available which actually demonstrate the usefulness and also highlight some of the disadvantages or limitation of each of these methods. Now all of these are

utilized to enhance the metabolic efficiency of pathway. So, ultimate target was very simple. Engineer the pathway the entire pathway of the suitable host organism in order to improve the production of the desired compound.

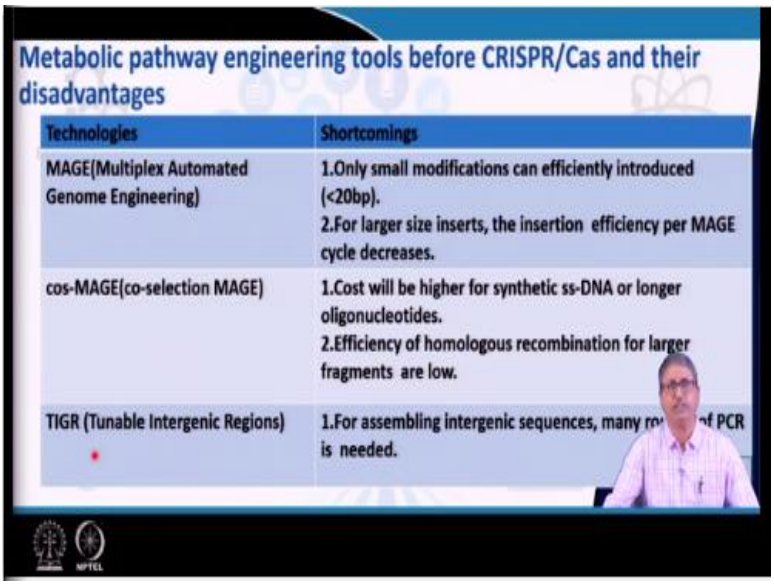
(Refer Slide Time: 20:11)



Now the strategies and methods for modulating the expression and pathway genes and control, a large number of genes simultaneously was also attempted. So, if we look at these processes the developments in the methods from random mutations to use of different specific transgene technologies including the alteration of the promoters or the ribosome binding site or application of RNA interference.

Then to change the plasmid copy numbers or homologous relying on homologous recombination. And then now it has been found at expression controlling the expression of multiple genes can be possible simultaneously. By using or adopting a very specific set of techniques which are the multiplex automated genome engineering tool or co-selection MAGE, another advanced method of MAGE and then the tunable intergenics regions and then the COMPACTER.

(Refer Slide Time: 21:16)



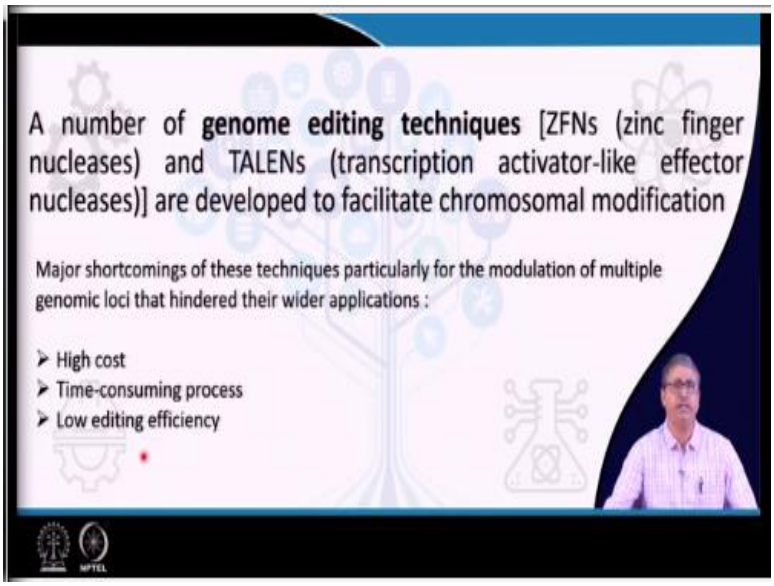
So, all these methods enabled a significant amount of developments. Now metabolic engineering tools before CRISPR-Cas and their disadvantage. So, out of so many genetic engineering tool including some of the genome engineering tool also. Because those who were more suitable for altering multiple genes simultaneously. So, addressing some of the aspects of the genome alteration or genome engineering tools.

So, the advantages and disadvantages are pretty well highlighted in the current literature. So, in this slide I am going to emphasize on few disadvantages or shortcomings of this method, this may be considered as shortcomings instead of disadvantages, because they have contributed a lot in progressing the metabolic engineering of desired pathways.

For example the multiplex automated genome engineering tool is only suitable for small modifications can effectively introduce less than 20bp base pair. For larger size inserts, the insertion efficiency decreases. For the advanced MAGE process also the cost is going to be higher as it is calculated. And efficiency of homologous recombination for larger fragments are low.

Similarly for the tunable intergenic region technology also, the assembling intergenic sequences many rounds of PCR are needed. So, if these are either cost intensive or time intensive or with respect to the higher larger size of gene fragments they were found to be quite less effective.

(Refer Slide Time: 22:59)

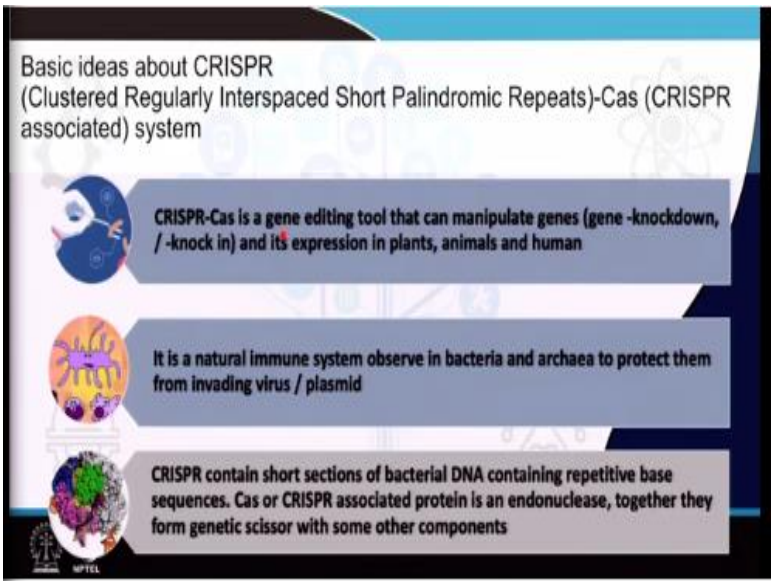


So, a number of genome editing technologies, very specific genome editing technologies subsequently were identified, developed in addition to MAGE and all other things, which are the zinc finger nucleases and transcription activator like effector nucleus. So, again all are these are nucleases, just before the application of CRISPR-Cas. These all methods were in place and tested very successfully and developed significantly to facilitate the chromosomal modification.

Now again with these zinc finger nuclease or TALENs like nuclease which were used for several reputed investigations. The major shortcomings of these techniques are found to be particularly for the modulation of multiple genomic loci at a time. Because that was one of the major target for the genome editing technologies, compared to other techniques like in transgenic technology or so, we were more bothered about individual genes.

So, when we try to edit multiple genes simultaneously within the genome without using the plasmid directly, we faced some challenges which are very important. And that actually enabled us to understand the efficacy or application of CRISPR-Cas that is the high cost. Most of the cases this TALEN or ZFN are found to be expensive, they are time consuming processes and low editing efficiency, their editing efficiency remains very poor.

(Refer Slide Time: 24:32)



So, in the meanwhile the developments on the CRISPR-Cas system as we will discuss more about CRISPR-Cas in our other lectures as well. So, CRISPR-Cas was investigated as a kind of bacterial immune or archaeal immune system. So, lot of discoveries, lot of mechanistic understanding were going on. Now some of the basic ideas are highlighted here, now what is CRISPR?

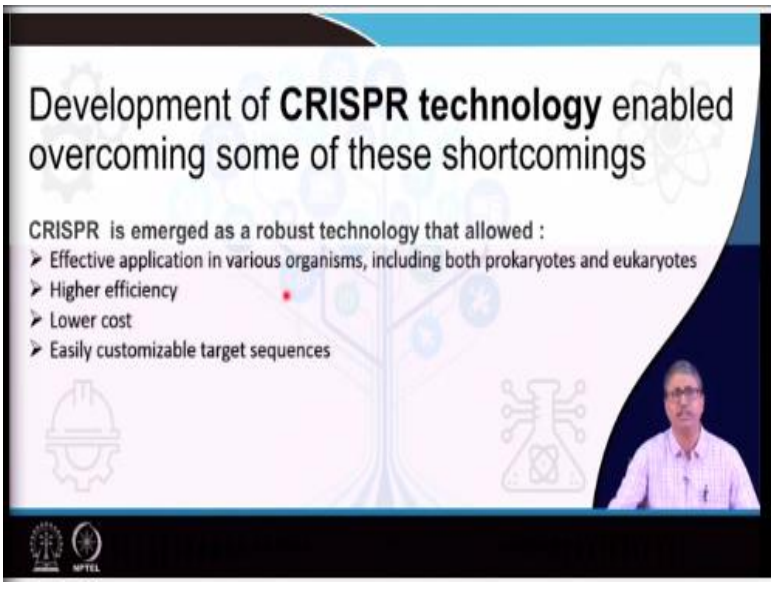
CRISPR is basically stand for the Clustered Regularly Interspaced Short Palindromic Repeats. And Cas which is associated system which is CRISPR associated system, so CRISPR is a very specific set of repeats and some other small DNA molecules which is there in the genomic DNA. And Cas is a separate entity which is closely associated with CRISPR. So, these Cas and CRISPR are often found to work together in order to facilitate it is natural function. That is providing the immunity towards invading virus or invading plasmids.

So, CRISPR-Cas is a gene editing technology, now it is naturally present in bacteria and Archaea. But we have found and tried to implement this CRISPR-Cas as a gene editing technology. And what we found that in recent time in last of 5 to 7 years time. This technology can help us to manipulate genes very successfully including the gene knock down, gene knock in and controlling the expression of genes even which are coming from plants to bacteria or within the plant system or within the animal system or even within the human system.

It is fundamentally a natural immune system, so CRISPR-Cas represents the kind of a microbial immune system which is observed in bacteria and archaea to protect them from invading virus or plasmid. CRISPR contains a short section of bacterial DNA containing repetitive base sequence. As it is the name stands for short palindromic repeat, so a number of repeats are there. And within the repeats we have some interesting pieces of DNA; those DNAs are actually obtained from previously invaded virus or plasmid. And as a kind of a memory the bacterial genome or archaeal genome maintain those small pieces very selectively and very regularly, they maintain these pieces.

And that actually provides them the memory kind of thing. Now Cas or CRISPR associated protein is an endonuclease, so there is a Cas encoding DNA also corresponding DNA. And this DNA it expresses and it produces the Cas associated protein which is an endonuclease. And together these CRIPSR that is the repeated repeats containing DNA portion, and the Cas they form a genetic scissor with some other component. So, it is basically they form a natural gene cutting method like which the genes can be make into small pieces; naturally it is the invading viral DNA or the plasmid DNA.

(Refer Slide Time: 28:14)

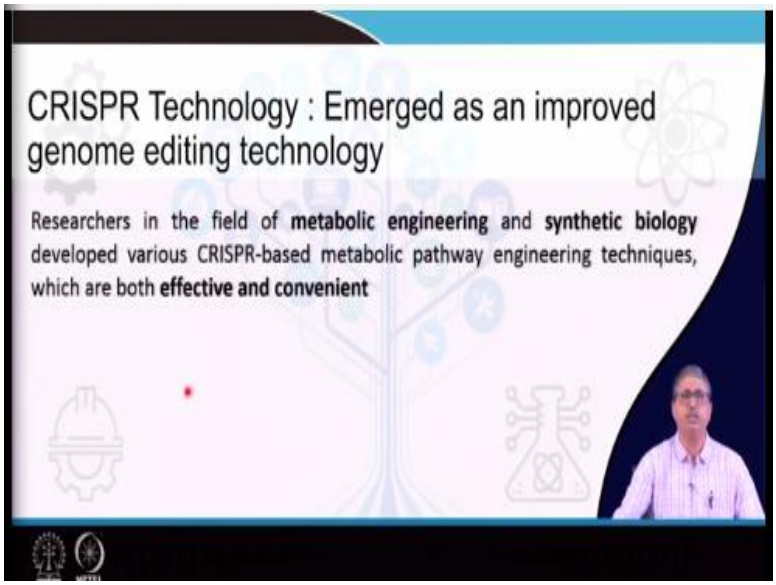


Now the development of CRIPSR technology enabled overcoming some of the shortcomings. So, we will be talking about the CRISPR-Cas technology, the fundamental processes in detail in some of the other lectures. But here I would like to highlight some of the major things that

enabled us to overcome these shortcomings which were presented by the transgenic technology and other available genome editing technologies.

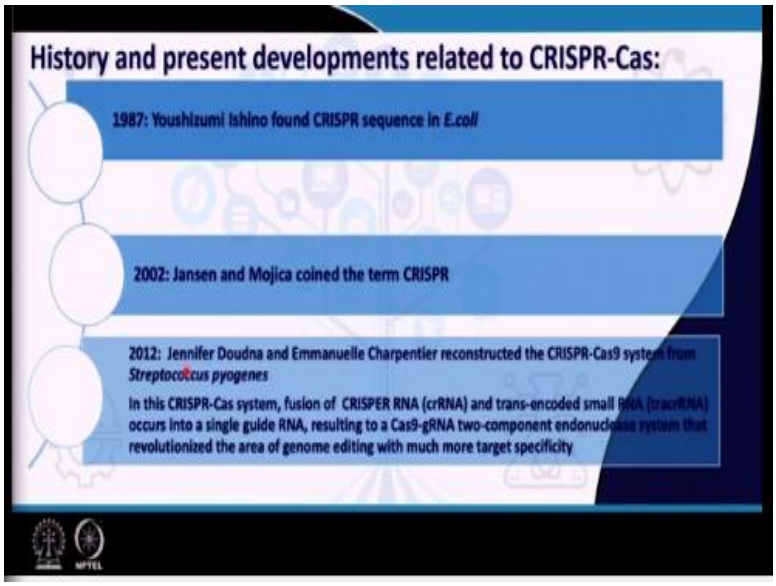
Now the CRIPSR is emerged as a robust technology, that allowed effective application in various organism including both prokaryotic and eukaryotic systems or cells. It has a very high efficiency, low cost and easily customizable target sequences.

(Refer Slide Time: 28:57)



The CRIPSR technology emerged as a result as an improved genome editing technology. And researcher in the field of metabolic engineering and synthetic biology both are developing various CRIPSR-Cas based metabolic pathway engineering techniques, which are both effective and convenient.

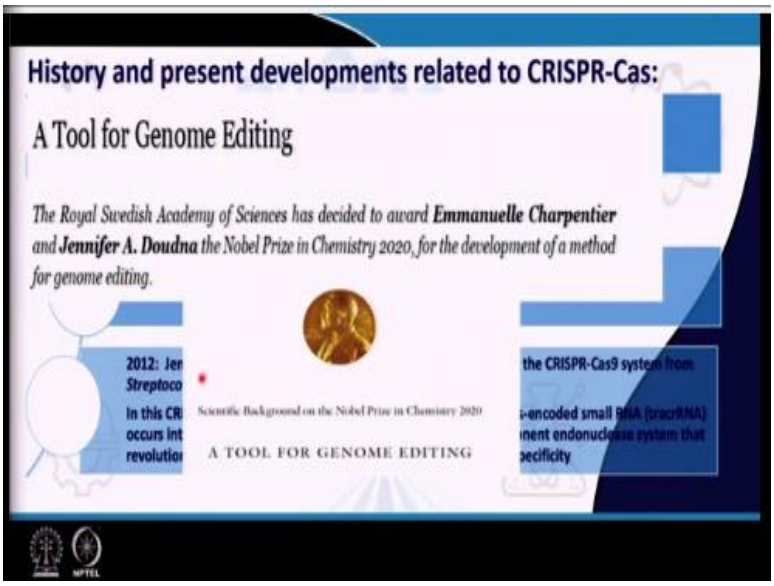
(Refer Slide Time: 29:12)



Now here is a brief history and the developments related to CRIPSR-Cas. That in 1987 Youshizumi Ishino found CRIPSR sequence first in *E.coli* although the name CRIPSR was not coined during that time. It is in the 2002 that Jansen and Mojica coined the term CRISPR. In 2012 Jennifer Doudna and Emmanuelle Charpentier reconstructed the CRISPR-Cas9 system from *Streptococcus pyogenes*.

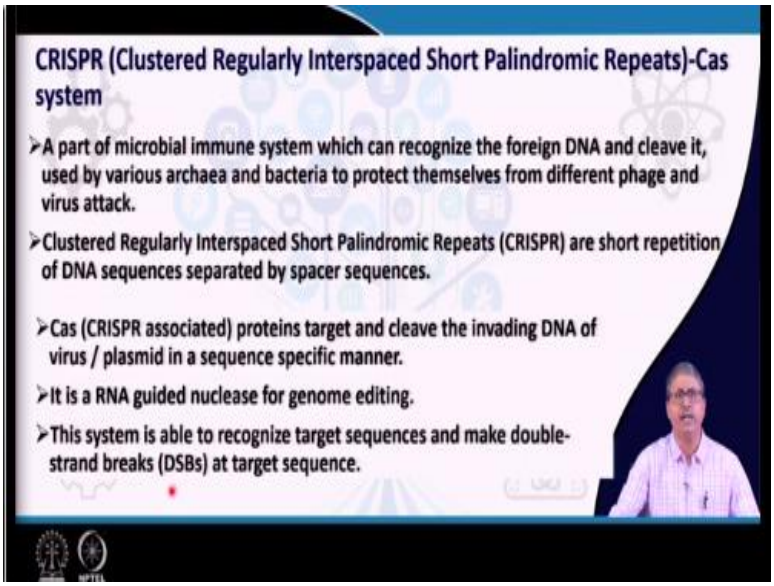
And in this CRISPR-Cas system fusion of CRIPSR RNA which is called crRNA and trans encoded small RNA occurs into a single guide RNA resulting into a Cas9-gRNA two component endonuclease system. That revolutionized the area of genome editing with much more target specificity.

(Refer Slide Time: 30:06)



And in 2020 Emmanuelle Charpentier and Jennifer Doudna they were awarded with the Nobel prize in chemistry for the development of the method for genome editing.

(Refer Slide Time: 30:19)

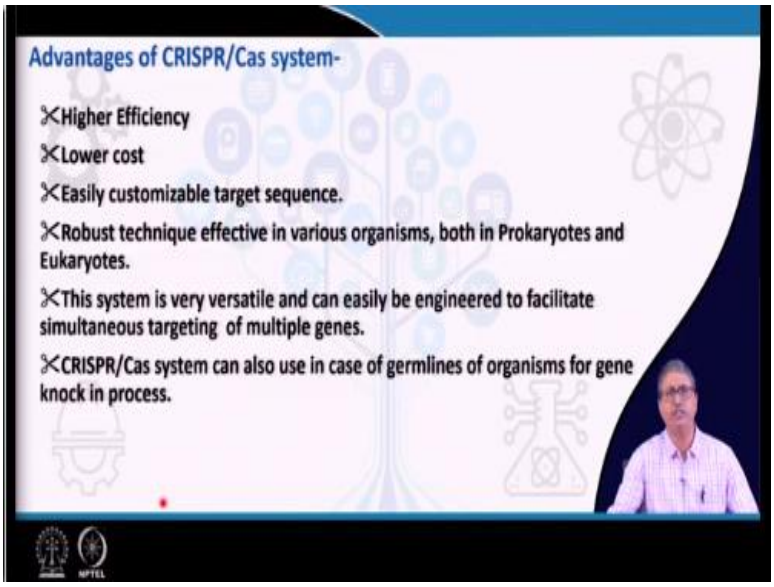


Now this particular system which is briefly presented here as a part of microbial immune system, which can recognize the foreign DNA and cleave it. And it is used by various archaea and bacteria to protect the host organism or the bacteria or archaea themselves from different phage and other virus attack. Now the Clustered Regularly Interspaced Short Palindromic Repeats that is the CRIPSR.

A short repetition of the DNA sequence separated by spacer sequence**.** And the Cas system which is the CRISPER associated protein which is already mentioned about an endonuclease. Target and cleave the invading DNA of virus or plasmid in a sequence specific manner, and it is a RNA guided nuclease for genome editing. So, Cas is the endonuclease or nuclease which is used for the genome editing.

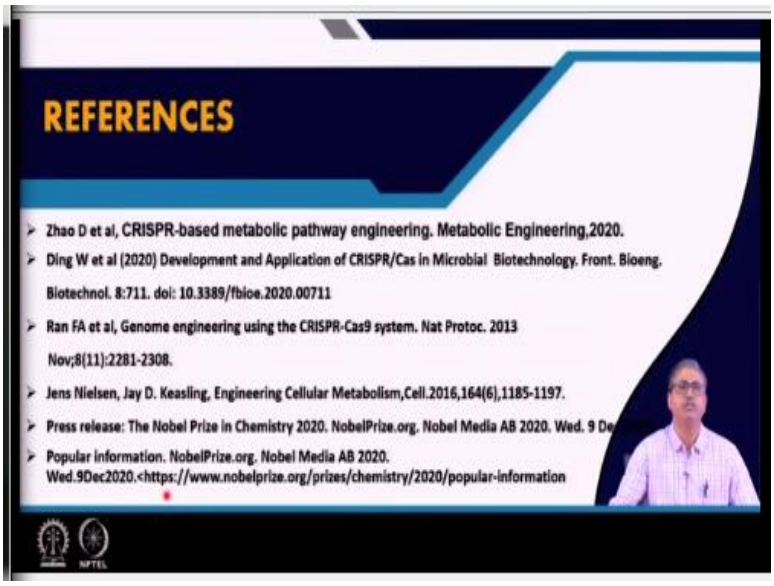
And naturally also it facilitates or it catalyzes the cutting of the DNA and cleavage of the DNA invading DNA in a very precise manner. And the entire process is naturally guided by a guide RNA molecule. Now this system is able to recognize the target sequences and make double stranded breaks that is called DSBs at target sequence.

(Refer Slide Time: 31:39)



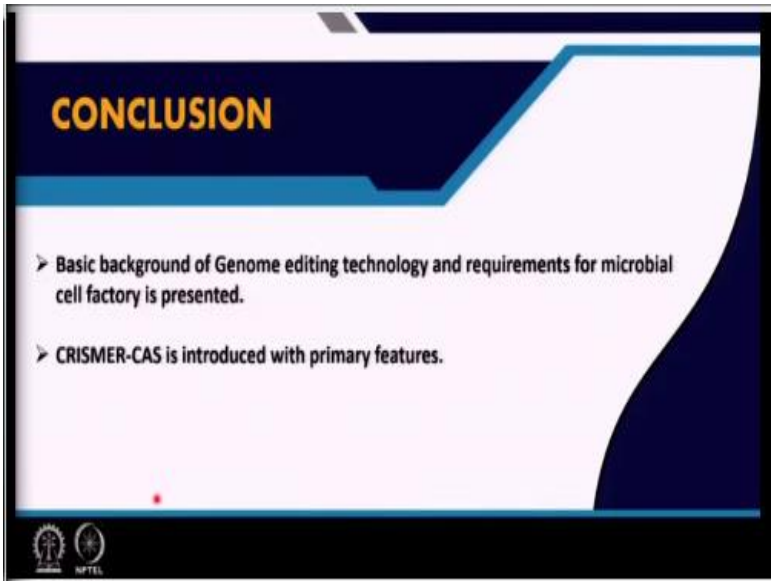
Now in the last few years we have found that a number of advantages which are offered by the CRISPR-Cas, including the high efficiency, low cost, easily customizable target sequences, robust technique effective in various organisms eukaryotic, prokaryotic. And is a very versatile and can easily be engineered to facilitate simultaneously targeting multiple genes. And CRIPSR-Cas system can also use in case of germlines of organism for gene knockout/in process.

(Refer Slide Time: 32:09)



So, this part of the lecture on CRISPR-Cas and it is role application in metabolic engineering is relied on a couple of papers, and here are the references.

(Refer Slide Time: 32:23)



And to conclude the basic background of genome editing technology including the major processes of transgenic technology and other genome editing technology which were very relevant for synthetic biology. And of course the metabolic engineering for cell microbial cell factory are presented. And CRISPR-Cas is introduced with some primary features, thank you.