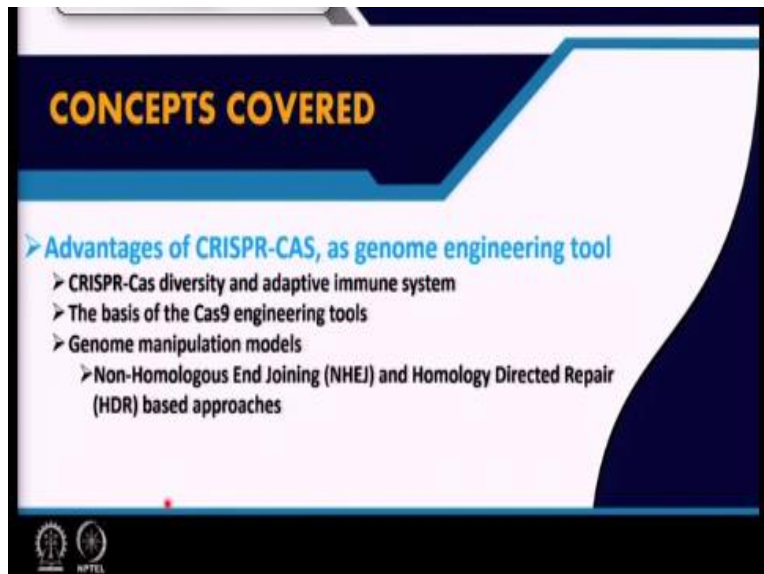
Metabolic Engineering
Prof. Pinaki Sar
Department of Biotechnology
Indian Institute of Technology-Kharagpur

Lecture-31
CRISPR-Cas System and its Application in Metabolic Engineering-Part III

Application of CRISPR-Cas system in metabolic engineering is going to be discussed in today's lecture. This is considered to be the part 3 of the lecture on CRISPR-Cas system and its application in metabolic engineering.

(Refer Slide Time: 00:46)



In this talk, the following concepts are going to be covered. We will start with advantages of CRISPR-Cas system as genome engineering tool. And then briefly we will talk about the diversity of CRISPR-Cas and it is adaptive immune system. The basis of CRISPR-Cas9 engineering tools are going to be highlighted. And then the genome manipulation models will be emphasized. Non-Homologous End Joining and Homology Directed Repair based approaches are going to be discussed.

(Refer Slide Time: 01:26)

CRISPR-Cas systems have been applied in single cell microbes to:

Maintain genomic integrity by mitigating the effects of foreign or mobile genomic elements

Modify and manipulate genomic DNA through the introduction of double-strand breaks (DSBs), which give rise to sequence alterations when repaired by endogenous repair pathways

Control gene expression via construction of artificial transcription factor

Now CRISPR-Cas systems have been applied in single cell microbes to maintain genome integrity by mitigating the effects of foreign or mobile genomic elements. Now to modify and manipulate genomic DNA through the introduction of double strand breaks DSBs which give rise to sequence alteration when repaired by endogenous repair pathways were utilized during this application of CRISPR-Cas system in metabolic engineering. Now control gene expression via constructive or construction of artificial transcription factors are also studied.

(Refer Slide Time: 02:14)

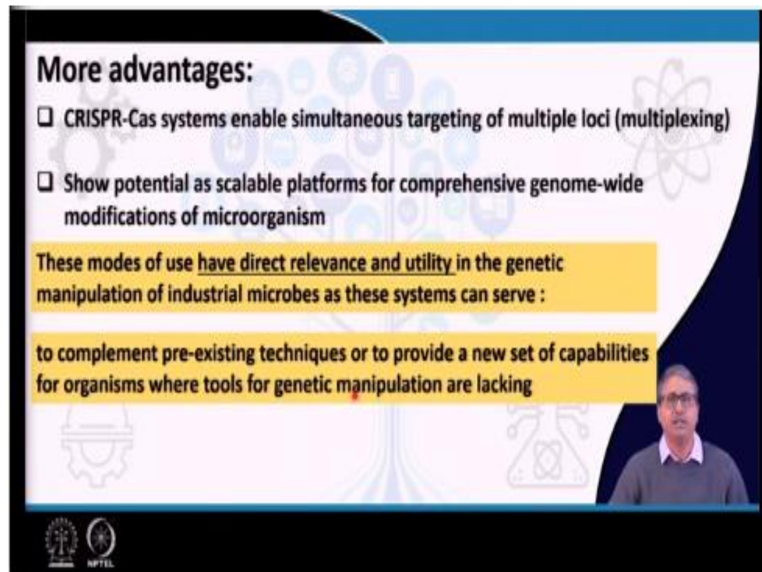
The repurposing of the RNA-guided DNA endonuclease from the type IIa CRISPR-Cas system of *Streptococcus pyogenes* (and of other Cas9 orthologues) **as genome editing tools brought an unprecedented revolution to the life sciences field**

Mougiakos et al, Current Opinion in Biotechnology, 50, 146-157, 2018

Now the repurposing of RNA guided DNA endonuclease from the type II CRISPR-Cas system of Streptococcus pyogenes and of other Cas9 orthologues emerged as a genome editing tools.

And that brought an unprecedented revolution to the metabolic engineering in particular and in general the life science field including industrial biotechnology.

(Refer Slide Time: 02:46)



Now the specific advantages of this application of CRISPR-Cas system with respect to industrial biotechnology or metabolic engineering in particular are as follows. The CRISPR-Cas system enables simultaneous targeting of multiple loci that is called the multiplexing. So, a number of ah genetic loci can be targeted or altered. And it shows potential as scalable platforms for comprehensive genome-wide modifications of microorganisms.

Now these modes of use have direct relevance and utility in the genetic manipulation of industrial microbes as this system can serve couple of very important aspects. That is to complement the pre-existing techniques or to provide a new set of capabilities for organisms where tools for genetic manipulations are lacking.

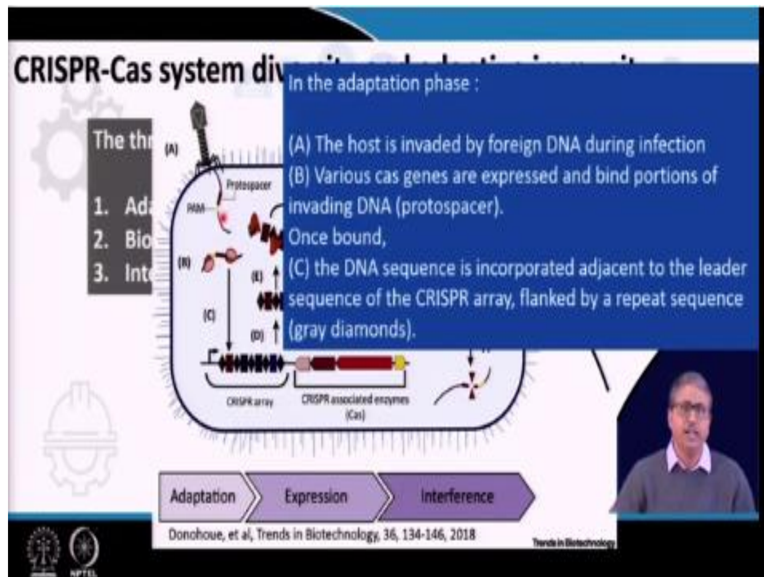
(Refer Slide Time: 03:48)

Now CRISPR-Cas system diversity and adaptive immunity: The three stages of CRISPR-Cas adaptive immunity are adaptation, second is the biogenesis or expression, and third is the interference.

(Refer Slide Time: 04:03)

As we have discussed in other lectures, so this is going to be a kind of a summarizing the entire the process where we can see the phase of adaptation.

(Refer Slide Time: 04:18)

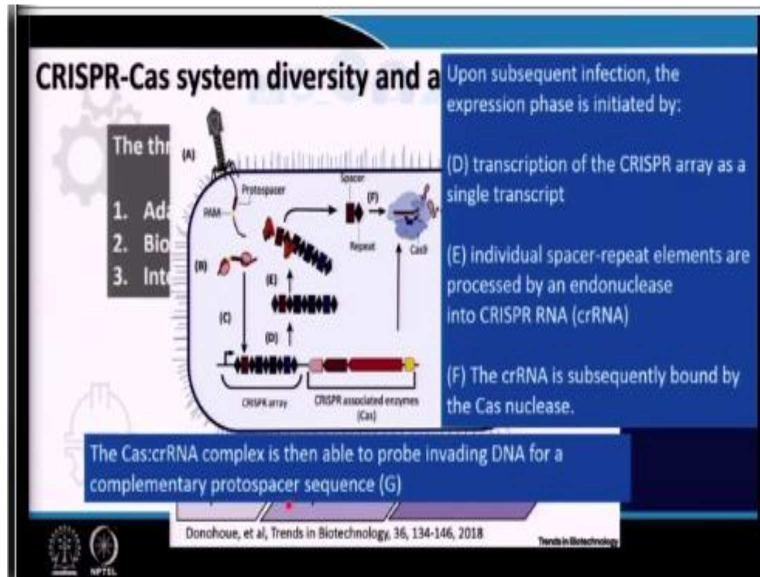


That is basically as the host is invaded by foreign DNA including the infection of the various including during the infection process by phage or other mobile genetic element etcetera. Various Cas genes are expressed and the bind portion of the invading DNA. So, here is the invading DNA which is now targeted and bound by different Cas genes which are expressed in the Cas protein.

They are able to identify because of the recognition of the PAM and the protospacer, they are able to locate the specific region of binding and they bind it. Now once bound the DNA sequence is incorporated adjacent to the leader sequence of the CRIPSR array flanked by the repeat sequence that is the picture is depicting the entire process. So, once the Cas protein is able to recognize the protospacer.

And then it cleaves the protospacer producing the small piece which is processed. And then form the spacer and then it is included within the CRIPR array just after the leader sequence.

(Refer Slide Time: 05:38)



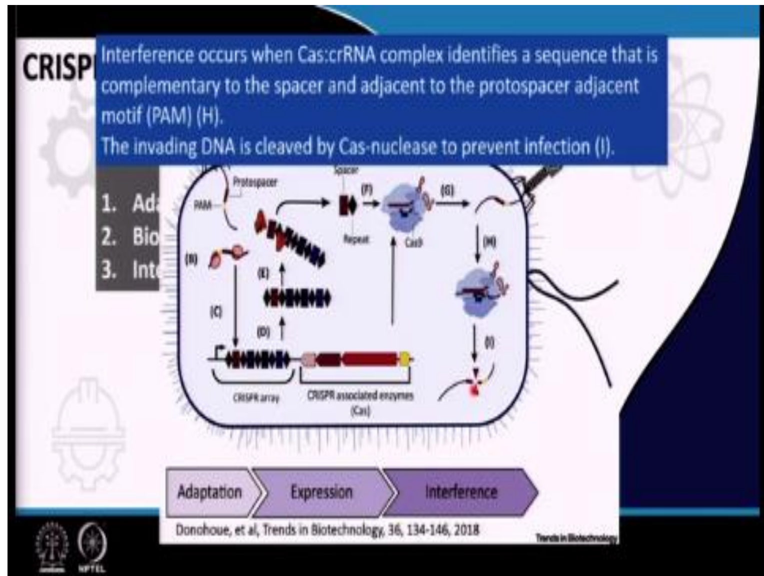
Now upon subsequent infection, so this is the CRIPSR which is now having one another spacer included in that, and that spacer is actually a kind of a token of the memory of the recent or past infection or past invasion of the foreign DNA. Now upon subsequent infection the expression phase or the biogenesis phase is initiated, that is the transcription of the CRIPSR array as a single transcript occur.

Now the entire CRIPSR array which is now expressing will produce the transcript which is called the CRIPSR RNA will be produced. And individual spacer repeat elements will be processed by an endonuclease that is again a part of the Cas itself into CRIPSR RNA or crRNA. So, we will have the small crRNA fragments, so these crRNAs will be produced and then the crRNA is subsequently bound by the Cas nuclease itself.

So, the Cas which is the a kind of a complex protein unit, it will be able to accommodate this small RNA fragment which will have this part of the repeat and the spacer, particular special region, so it will be bound to this Cas region. Now the Cas CRISPR RNA complex, so that is once the crisper RNA is bound to the CAS, we call this Cas CRIPSR RNA complex.

It is then able to probe the invading DNA for a complementary protospacer sequence. So, now this complex which is now the Cas plus the CRIPSR RNA in combination it will be able to interrogate or scan the invading DNA in order to find out complementary protospacer region.

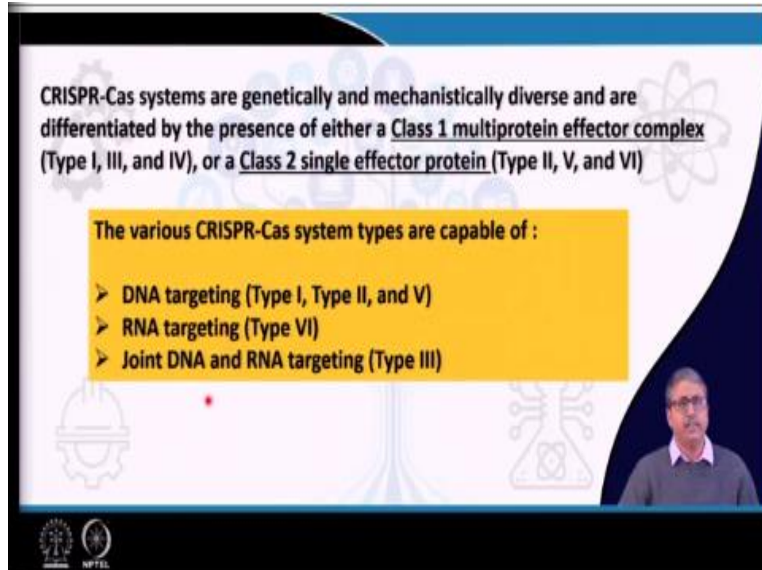
(Refer Slide Time: 07:35)



Now as soon as the protospacer region is identified, the interference will occur when the CRISPR-Cas system and the CRIPSR RNA complex identify the sequence that is complementary to the spacer and adjacent to the protospacer we have the PAM region. So, once this PAM and the protospacer, the PAM is over here in the yellow colour and the protospacer region is identified.

The invading DNA is cleaved by the Cas nuclease to prevent the further infection. So, the specifically close to the PAM region, this invading DNA will have a double stranded break and the DNA will be cleaved, so that is the interference.

(Refer Slide Time: 08:16)



Now this particular event of catalyzing the double stranded break by the entire process of Cas and it is the RNA molecule which has the spacer along with the repeat sequence is exploited in the developing metabolic engineering process or developing the genome engineering tools. Now the CRISPR-Cas systems are genetically and mechanistically diverse.

And are differentiated by the presence of either a class 1 multiprotein effector complex which is having basically three, types type I, type III and type IV, or a class 2 which has very characteristically single effector protein. As we know that Cas9 which is most widely used in metabolic engineering belongs to class 2 CRISPR-Cas system because it is having a single effector protein.

And it is having different types like type II, type V, and type VI. And the various CRISPR-Cas system types are capable of either DNA targeting, RNA targeting or joint DNA and RNA targeting.

(Refer Slide Time: 09:29)

Now the basis of the; CRISPR-Cas9 engineering tools are going to be discussed. Now the simple way in which the Cas9 nuclease can be guided to the desired DNA target denoted as protospacer. So, the protospacer is kind of an distinct or very characteristic region within the DNA. And we can actually target the Cas nuclease to reach to that target which is basically the protospacer region.

And if there is the adjacent PAM residue then the CRISPR-Cas system if it is able to activate the Cas through the CRISPR RNA and then trans activating RNA complex. Then it would actually help the Cas9 to deliver the catalytic mechanisms and carry out the double stranded brake.

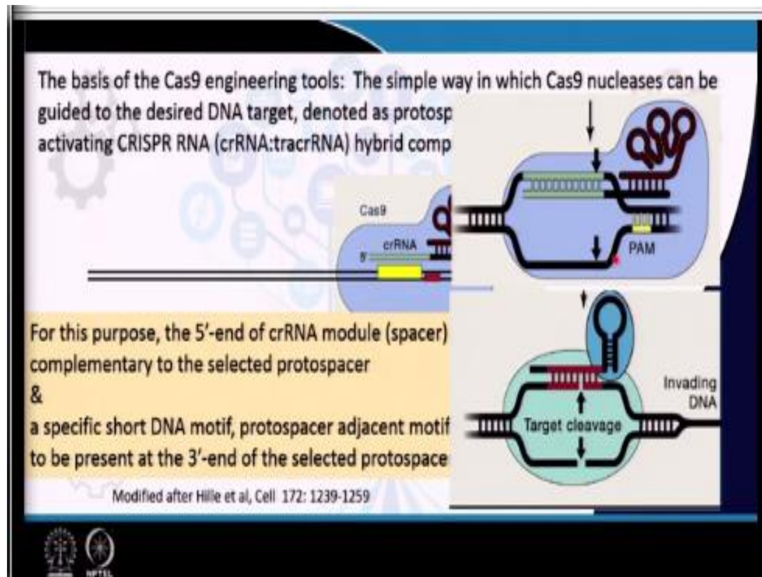
(Refer Slide Time: 10:24)

So, let us see what is going to happen, so once it is going to the scan or go through the DNA, it will try to locate that if that the protospacer region is there. And the protospacer region can be very clearly identified because the spacer which is already there in the CRISPR RNA and a trans activating RNA which is complementarily bound to this CRIPSR spacer region will enable the entire Cas enzyme or Cas complex to identify and then establish the complementary region and also with the help of the PAM region over here.

It will be very specifically capable of identifying that this is the region of protospacer where the binding between the complementary spacer and the protospacer region can be achieved. Now for this purpose the 5 prime end of the CRISPR RNA module which is basically the spacer has to be complementary to the selected protospacer. So, if we know that we are going to target a particular gene, then that particular gene is going to be our protospacer.

And we can actually engineer our cell to produce the spacer accordingly the CRIPSR RNA. Because we know that our target gene within the genome is this, so we can actually design the CRIPSR RNA accordingly. So, the CRIPSR RNA will be complementary to any target gene, and then if it is able to identify it, it will bind. Now a specific short DNA motif which is already known is known as the protospacer adjacent motif PAM has to be present at the 3 prime end of the selected protospacer. So, that the entire CRISPR RNA and trans activating RNA is able to identify and then enable the Cas system to carry out the rest of the catalytic activities.

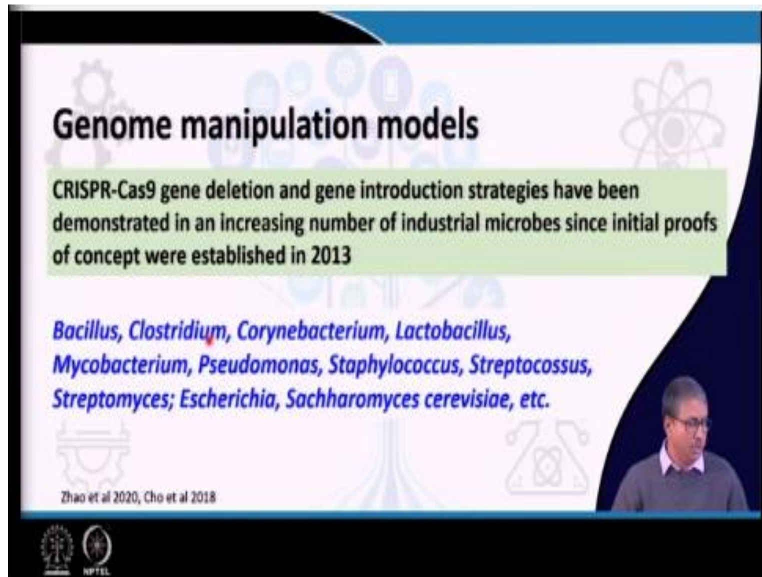
(Refer Slide Time: 12:35)



So, here is the picture with a better resolution that we can see that this is the PAM region which is adjacent to the protospacer. And if the PAM is identified then this CRISPR RNA and trans activity, so this is basically trans activating RNA and this is the CRIPSR RNA. So, the entire combination is going to start the process rest of the process, so that the CRIPSR the Cas system can start the double-stranded break process.

So, if everything goes well then the target DNA that is the invading DNA in case of a natural system or in case of a genetically engineered system, the DNA of our target region will be subjected to a cleavage. And it is going to be a double stranded break both the DNA stands will be cleaved at a very specific location.

(Refer Slide Time: 13:33)

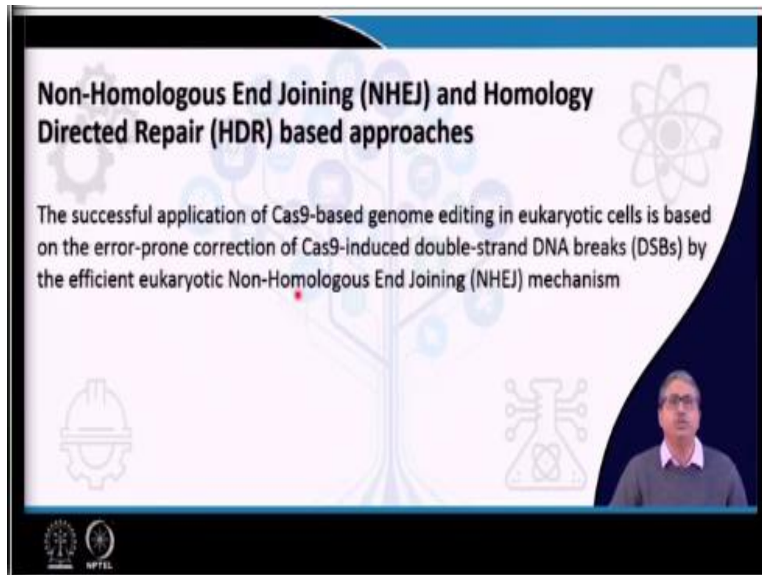


Now based on this concept the genome manipulation models are developed. So, the CRISPR-Cas gene deletion and gene introduction strategies have been demonstrated in an increasing number of industrial microbes since the initial proof of concept established in 2013. And a large number of bacterial and fungal and other the plant cells are subjected to metabolic engineering.

To be of particular interest of the *Bacillus, Clostridium, Corynebacterium, Lactobacillus, Pseudomonas, Staphylococcus, Mycobacterium, Streptomyces, E.coli Saccharomyces cerevisiae* and other yeast as well. As you can understand that most of these organisms are well studied, well used already in different applications with respect to industrial biotechnology or they are being subject to metabolic engineering.

So, the advent of CRISPR-Cas base metabolic engineering approach helped scientist to achieve their goals, their targets more comprehensively. Because as we have discussed earlier the minimum disturbance to the genome or the burden to the genome is achieved through CRIPSR-Cas system.

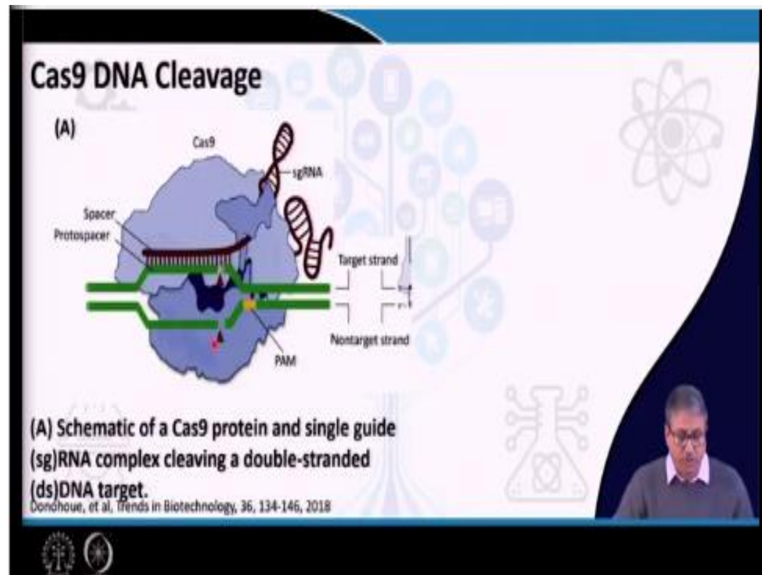
(Refer Slide Time: 14:54)



Now we are going to talk about the details of the methods which are applied while achieving the metabolic engineering of the important or relevant strains. So, two methods are very relevant to begin with are the Non-Homologous End Joining and Homology Directed Repair based approaches. So, the successful application of cas9 based genome editing in eukaryotic cell is based on the error prone correction of cas9 induced double stranded break that is the DSBs by efficient eukaryotic Non-Homologous End Joining mechanism.

So, that means in eukaryotic system the Non-Homologous End Joining mechanisms are already in place. So, if we are able to create a double stranded break in a particular gene through the CRISPR-Cas system, and CRISPR-Cas the entire system that is the CRISPR RNA and trans activating RNA. Then this Non-Homologous End Joining mechanism which is already there in eukaryotic cells will try to repair that. And that repairing process will help us to achieve some of the goals of metabolic engineering, so let us see that.

(Refer Slide Time: 16:13)

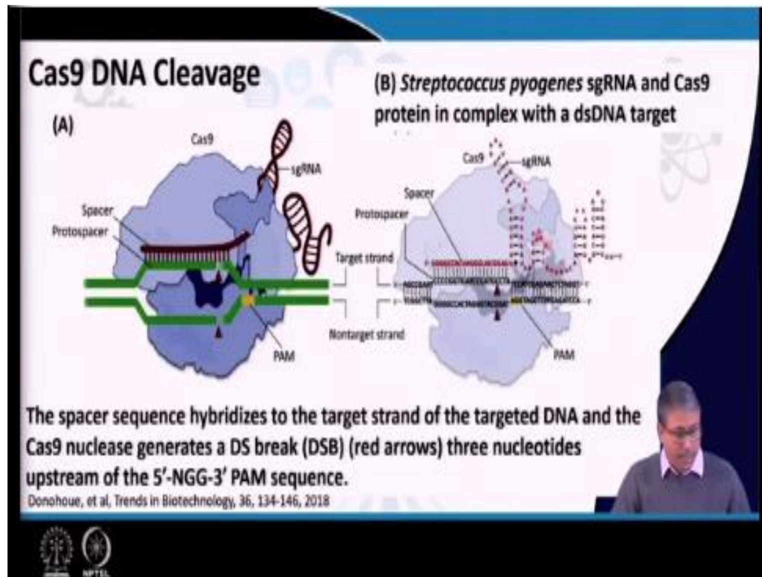


Now Cas9 catalyzes the DNA cleavage that is what is clearly discussed earlier. Now this is again a schematic of Cas9 protein and the single guide RNA complex. So, this single guide RNA complex which is a kind of a hybrid complex between the 2 components, one is the trans activating RNA and another part is the spacer. So, this spacer is the complementary to the target gene region.

So, if we know that our target gene within a genome, we can design the spacer accordingly and then put this spacer sequence along with the trans activating RNA sequence through a plasmid. And then let the entire single guide RNA complex to be expressed, the RNA will be produced. And that RNA if it binds to the Cas9 will be capable of facilitating the rest of the reaction.

Now this reaction which is achieved by the complementary binding of the spacer to the protospacer helped by this the trans activating RNA will finally lead to the double stranded break. So, here is the double stranded break, again it is shown here.

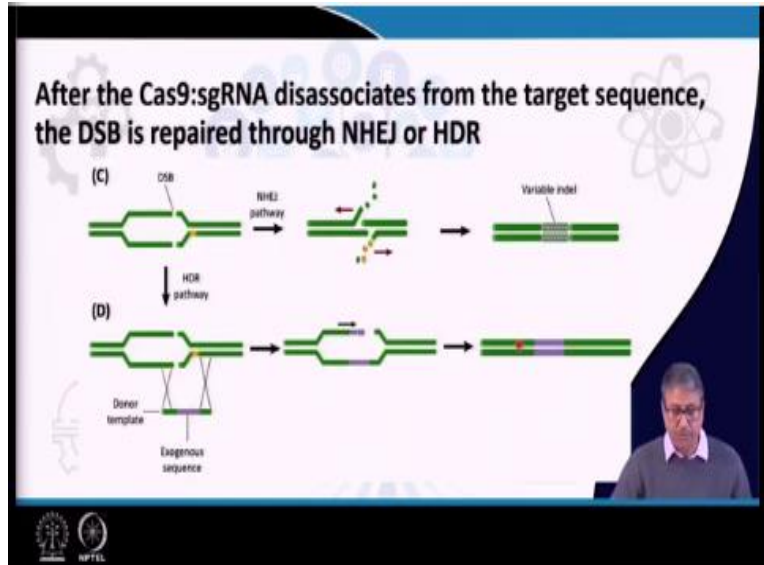
(Refer Slide Time: 17:32)



Now in *Streptococcus pyogenes* the single guide RNA and Cas9 protein in the complex with a double stranded DNA target is well elucidated. And as you can see this is the spacer region and this entire part is the partly the repeat portion, and then you have the trans activating RNA portion, and here is the PAM region. So, if we know that the particular gene we are going to target as I mentioned we have to select the appropriate spacer and try to express it within this.

Now this spacer sequence hybridizes to the target strand of the targeted DNA and Cas9 nuclease generate the double stranded break that DSB which is indicated here as red arrows through 3, nucleotide upstream of the PAM sequence. So, it is exactly 3 sequences upstream to the PAM sequence. So, it is highly specific and it is in the control of only that we are able to express this single guide RNA and the PAM is recognized, and the Cas is able to catalyze the entire process.

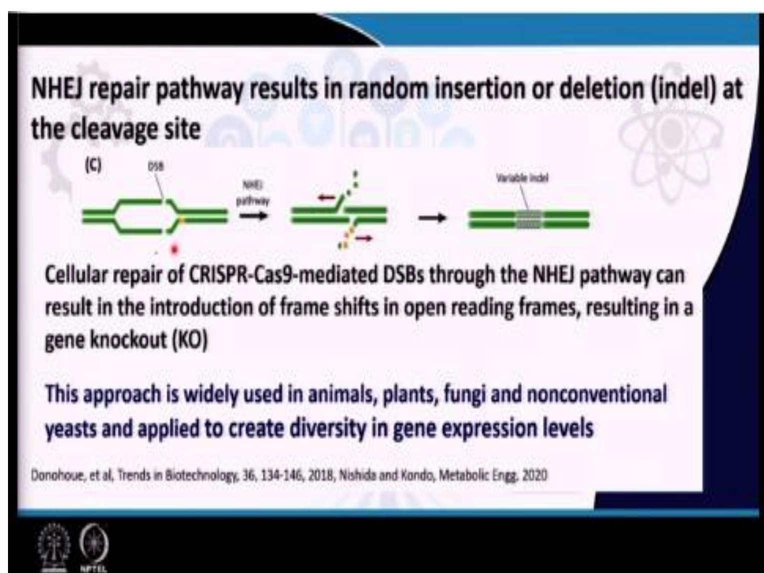
(Refer Slide Time: 18:43)



Now after the Cas9 single guide RNA dissociates from the target sequence, the double stranded break is repaired through the non homology based end joining method or the homology based repair mechanisms. Now here we are portraying both the mechanism, so you have the double stranded break over here. So, either the NHEJ pathway will lead to some kind of mutation into that.

Or this HDR or homology directed recombination process will enable us to introduce a new sequence or new gene or a modified gene, if it is provided through another plasmid or through other sources within the target region.

(Refer Slide Time: 19:34)

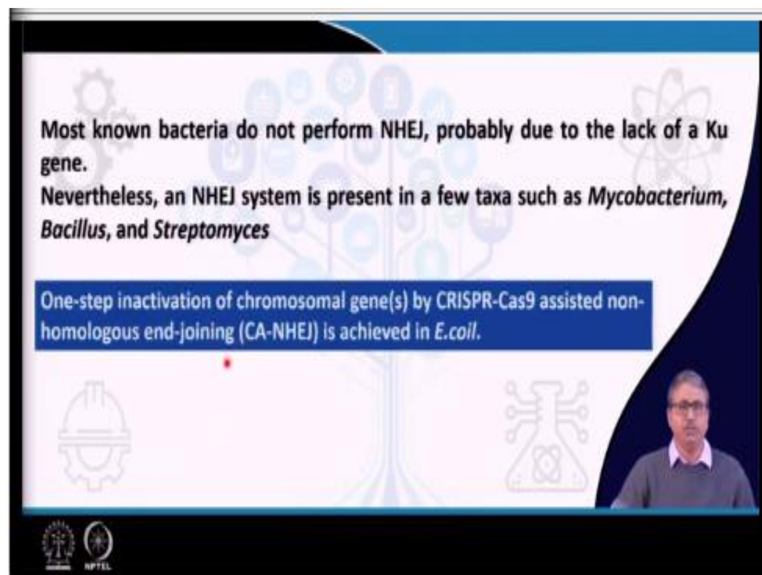


So, the NHEJ repair pathway results in random insertion or deletion indel at the cleavage site. So, we can actually target our Cas nuclease through a specific location within the genome and then facilitate the double stranded break. So, once the double stranded break is there because this is again a very site specific double stranded break, and then the variable indel will be obtained.

Now the cellular repair system of CRISPR-Cas mediated double stranded break through the NHEJ pathway can result in the introduction of frame shift in the open reading frame, resulting into engine knockout or KO we abbreviate often as. So, it is very commonly observed in eukaryotic systems, that if you are able to create a double stranded break by Cas9 system then ultimately it will lead to a gene knockout within a particular gene where the double stranded break is achieved.

Now this approach is widely used in animals, plants, fungi and in also nonconventional yeasts and applied to create diversity in gene expression levels, because the expression level depends on what are the genes are disrupted because of the knock down or knockout process. And by each of these knock out process, we can characterize these strains and then determine their flux and other things which are important for metabolic engineering studies.

(Refer Slide Time: 21:12)



Now most known bacteria, however do not perform this particular mode of NHEJ probably due to the lack of a couple of very important gene like Ku genes. However, the system is found to be

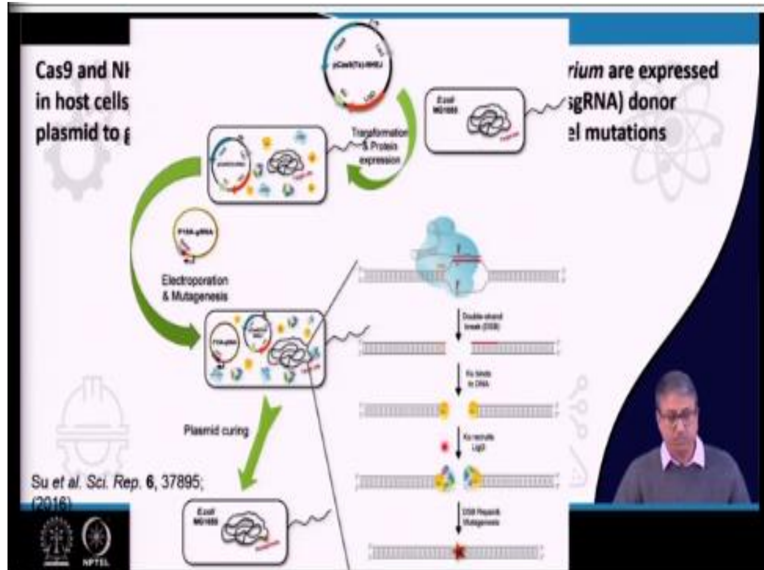
present in few taxa such as Mycobacterium, Bacillus or even Streptomyces. So, scientist tried to use these relevant genes which are involved in NHEJ that is the Ku and other genes from Mycobacterium and try to express it into *E.coli* system.

And then they could actually achieve some amount of knock down or alteration in the gene expression. Now one step inactivation of the chromosomal genes by CRIPSR-Cas9 assisted non homologous end-joining process is achieved in *E.coli*.

(Refer Slide Time: 22:07)

Now here is that case study. Now the Cas9 and NHEJ related proteins that is the Mycobacterium Ku and LigD genes from Mycobacterium are expressed in host cells. So, *E.coli* host cell are engineered first with respect to these two genes, so that these two genes are there in a plasmid, and they are transformed with a single guide RNA. So, another plasmid will have the guide RNA, donor plasmid to generate the double stranded break and trigger the indel mutation.

(Refer Slide Time: 22:40)



So, here you can see that one of the plasmids are going to have this Ku and LigD gene which are responsible for this known non homology end joining process. And they will produce the plenty of these protein molecules along with the expression of the Cas9, because Cas9 is also there in the first plasmid. And then the target gene guide RNA will be incorporated into the same system. Now this guide RNA will enable specific binding of the Cas nuclease and then the double stranded break will be there.

And followed by the; double stranded break this Ku and LigD kind of proteins, now they are expressed because in the plasmid these genes are there. So, that will facilitate the end joining process resulting into the mutagenesis. So, this was successful and in subsequent time, we found that a number of such elegant steps were used to produce a large number of very good results.

(Refer Slide Time: 23:47)

DNA cleavage may cause cell death or unexpected chromosomal rearrangement, more prominent in prokaryotes

The advantage of using CRISPR-Cas9 over conventional tools is that the introduction of DSBs greatly increases the rates of recombination when used in conjunction with an appropriate DNA donor molecule

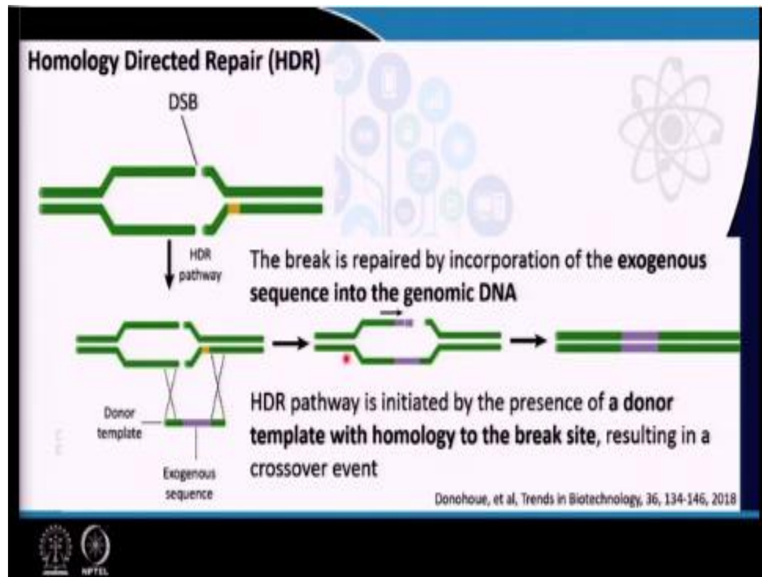
The advantage of using CRISPR-Cas9 over conventional tools is that the introduc DSBs greatly increases the rates of recombination when used in conjunction appropriate DNA donor molecule

Now the DNA cleavage may cause cell death or unexpected chromosomal rearrangement more prominent in prokaryotes. So, we already talked about this particular feature that in prokaryotic cells like bacteria, we do not have the scope for the end joining based DNA repair. So, the double stranded break based mechanisms are naturally not possible. But as I stated with this particular example that there are other methods by which we can circumvent this problem.

And we can still go ahead with different gene altering the expression of different genes. Now the advantage of using CRIPSR-Cas9 over conventional tools is that the introduction of double stranded break greatly increases the rate of recombination when used in conjunction with an appropriate DNA donor molecule. So, followed by this double stranded break if we have the required machinery for the homology based recombination and also a donor, DNA donor molecule.

Then perhaps we can take the advantage of the double stranded break and introduce a gene of our interest within the break site. Now, the advantage of using the CRISPR-Cas9 over conventional tools is that the introduction of double stranded break greatly increase the rates of recombination and when used in conjunction with appropriate DNA donor molecule. So, this is one of the very interesting parts.

(Refer Slide Time: 25:14)

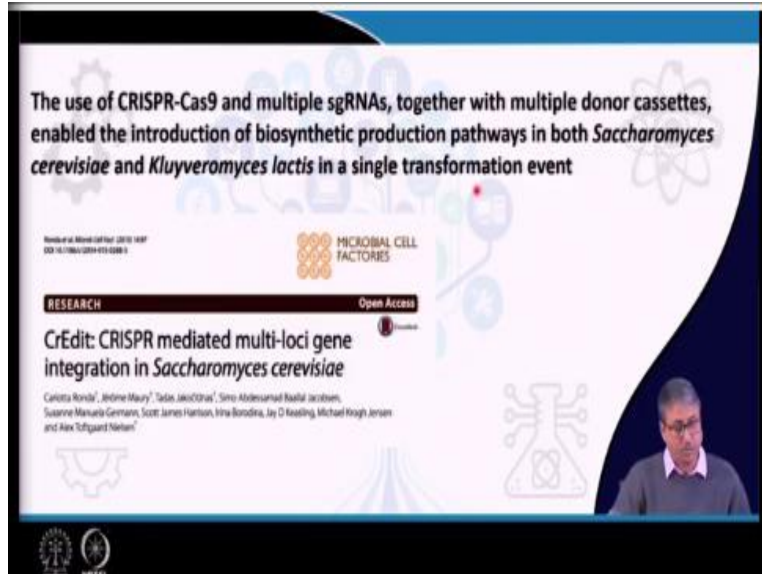


Now if we see that this is particularly known as homology directed repair or HDR. So, here is the break catalyzed by the CRISPR-Cas9 system. So, we can actually target this break very specifically as we discussed earlier. And then if this donor template is provided by exogenous sequence maybe through plasmid or by other mechanism. Then perhaps this could be inserted as you can see over here.

This is inserted perfectly with this double stranded break site. So, HDR pathway is initiated by the presence of a donor template with homology to the break site. So, we must you know that in order to achieve genetic engineering or genome editing through CRISPR-Cas HDR system. So, we must be providing a donor template with homology to the brake site and then only the crossover event will occur.

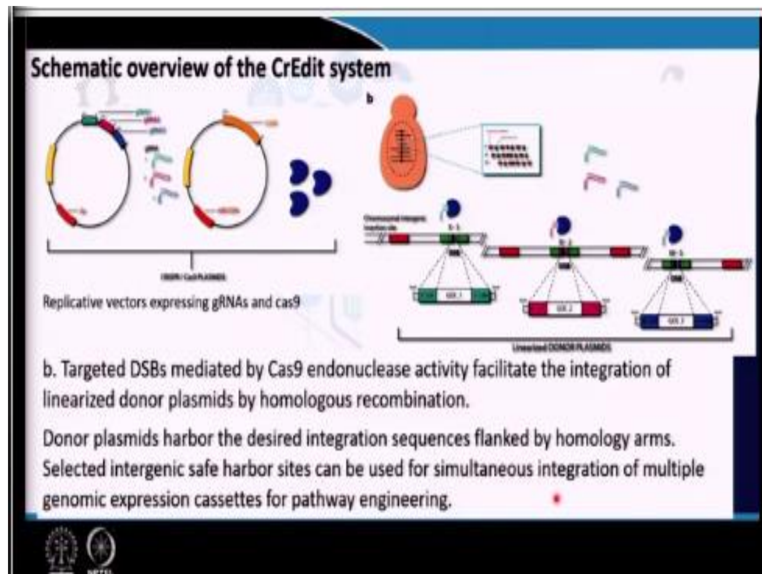
And the brake is repaired by the incorporation of the exogenous sequence into the genomic DNA. So, once we have this exogenous DNA whose sites are actually complementary and is suitable for recombination or homology directed repair, then this will be incorporated into the target site.

(Refer Slide Time: 26:35)



And again this method is very well established to achieve different extent of metabolically engineered cells. Like the use of CRIPSR-Cas9 and multiple guide RNAs together in with multiple donor cassettes, enabled the introduction of biosynthetic production pathway in both *Saccharomyces cerevisiae* and *Kluyveromyces lactis* in a single transformation event.

(Refer Slide Time: 27:04)

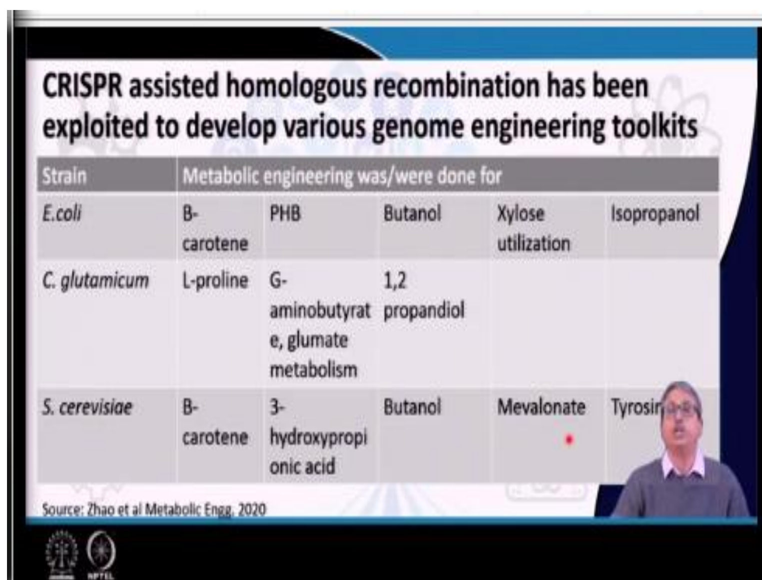


So, a large number of data are available with that. Like in this case you can see that different guide RNAs are produced through the plasmid. And then also the Cas9 gene is expressed. So, you have the Cas9 over here, we have the guide RNA, so each of these guide RNAs are specific for a particular region in the genomic DNA where we want the HDR event to occur.

Now if we follow the next of the steps, so the targeted double stranded breaks mediated by the Cas9 endonuclease activity, facilitate the integration of linearized donor plasmid of homologous recombination. Now all of these like GOI 1, GOI 2 and GOI 3, these genes are actually gene of interest that we want to insert over here. So, with a set of plasmids we can actually incorporate these genes.

So, along with the gRNA these target genes are to be provided through different plasmids, and when they are available through recombination they will be inserted here. Now donor plasmids harbor the desired integration sequence flanked by the homology arms. Selected intergenic safe harbor sites can be used for simultaneous integration of multiple genomic expression cassettes for pathway engineering.

(Refer Slide Time: 28:22)

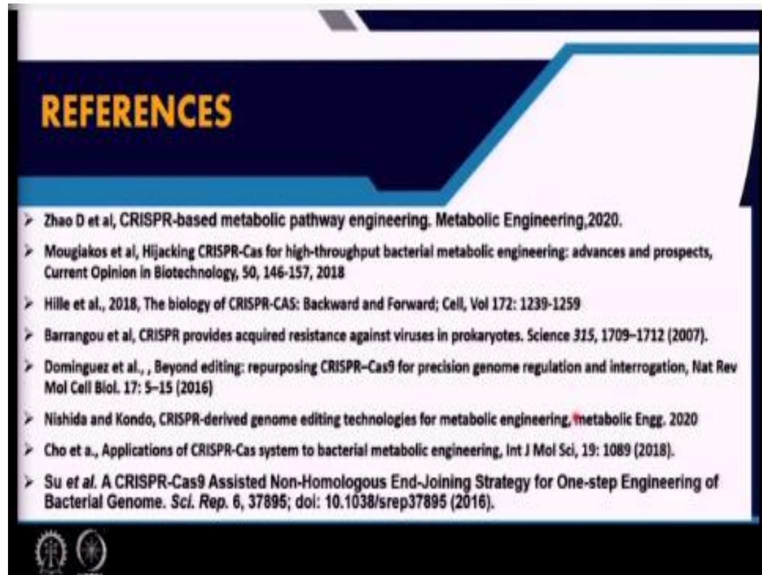


Strain	Metabolic engineering was/were done for				
E.coli	B-carotene	PHB	Butanol	Xylose utilization	Isopropanol
C. glutamicum	L-proline	G-aminobutyrate, glumate metabolism	1,2 propandiol		
S. cerevisiae	B-carotene	3-hydroxypropionic acid	Butanol	Mevalonate	Tyrosin

And several such studies have been reported in the literature in last 5 or 10 years. So, some of the important achievements are summarized in this table. As you can see that the CRIPSR assisted homologous recombination data is available. And part of this data like with respect to *E.coli,* different studies have achieved the metabolic engineering of E.coli with respect to beta carotene, polyhydroxy butyric acid, butanol, xylose utilization, isopropanol production, *Corynebacterium glutamicum* is engineered through the same technology for enhanced production of proline, gamma aminobutyrate or glutamate metabolism, 1, 2-propandiol.

Or *Saccharomyces cerevisiae* is used for beta carotene production, 3-hydroxypropionic acid production, butanol, Mevalonate production or even tyrosine production.

(Refer Slide Time: 29:25)



So, with this I complete this part and for this the following references are to be followed and thank you.