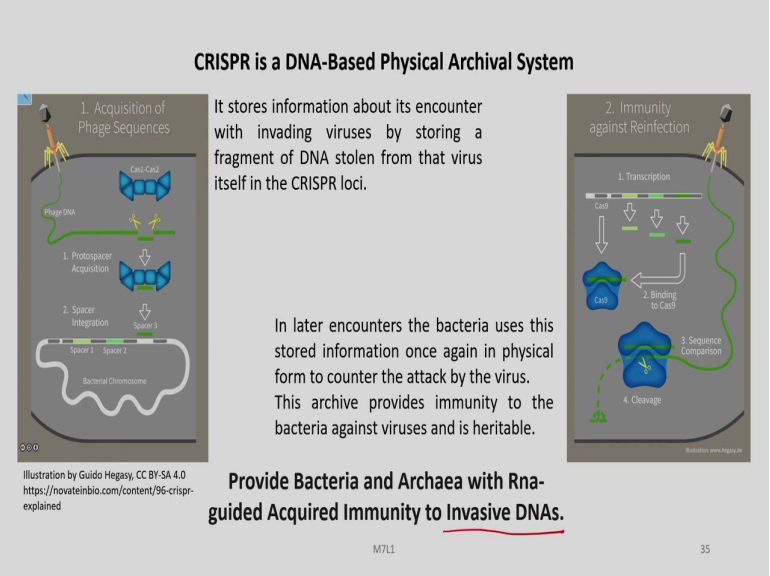
Genome Editing and Engineering
Prof. Utpal Bora
Department of Bioscience and Bioengineering
Indian Institute of Technology, Guwahati

Module - 07
Clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 technology
Lecture - 15
CRISPR system in bacteria - Part B

Welcome back to my course on Genome Editing and Engineering. We are discussing here the CRISPR Cas9 technology. And, in this part we are continuing to discuss on the CRISPR system in bacteria. So, you already know about the CRISPR Cas system in detail by this time.

(Refer Slide Time: 00:46)

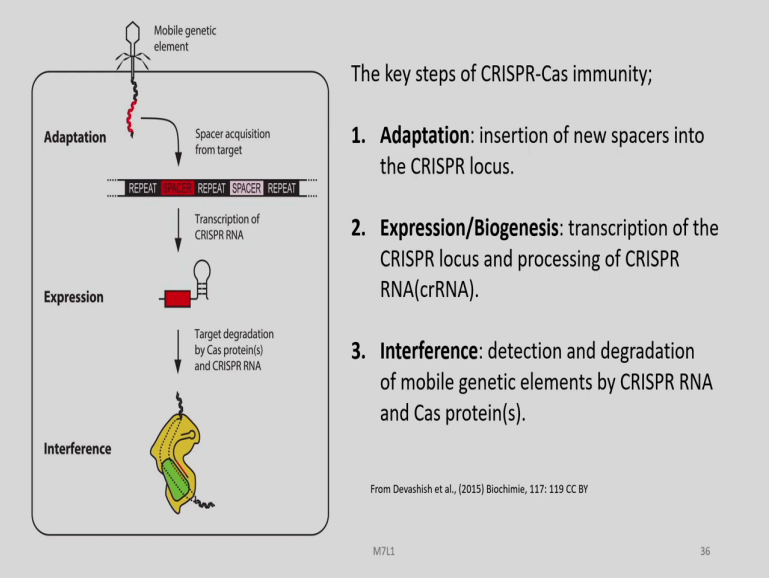


Let us remember that CRISPR is a DNA based physical archival system. And it provides bacteria and archaea with RNA guided acquired immunity against invasive DNAs. And these invasive DNAs may come from various sources like viruses, etc. Now, there are a few important things about this acquired immunity. So, there has to be some kind of acquisition of the DNA, which we already discussed in the earlier part.

So, in this acquisition of the fast sequences or NEME sequences there is storage of information about its encounter with the invading virus by storing a fragment of the DNA stolen from that virus itself into the CRISPR loci. And in the immunity against this reinfection phase, whenever in the future some virus attack is going to happen, in this later

encounters the bacteria would use the stored information once again in the physical form to counter the attack by the virus. This archive provides immunity to the bacteria against virus and is heritable and this is already discussed and known to you.

(Refer Slide Time: 02:13)



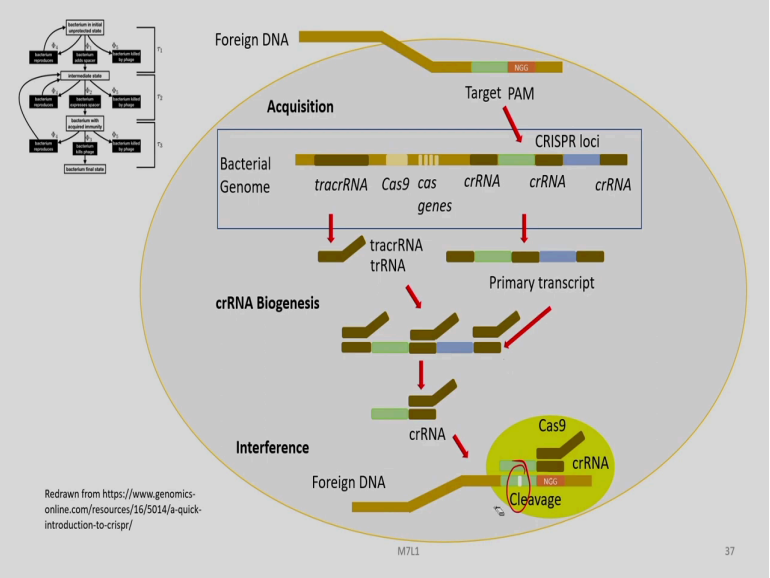
Now, what are the key steps of CRISPR Cas immunity? We were discussing about the Markov chain model in this regard. Now, let us look into exactly what happens inside the bacteria when it encounters a virus or a mobile genetic element of pathogenic nature. So, the key steps are the adaptation stage, second step is the expression in the biogenesis stage and the third is the interference stage.

In the adaptation stage there is the insertion of the new spacers into the CRISPR loci and we have discussed how the Cas1, Cas2 and CSN 2 in and along with Cas9 in the CRISPR Cas9 system does that. And in the expression and biogenesis the transcription of the CRISPR loci and processing of CRISPR RNA takes place. And this is again dependent on the class N type which we have discussed. In certain types cascade will be involved, in certain other types Cas9 will be involved and so on.

In interference also there will be variability we will discuss these in future, but exactly what happens in interference is that the bacteria detects and degrades the mobile genetic elements by CRISPR RNA and Cas9 in later encounters. The adaptation is a outcome of the earlier encounters and at that stage there is no any immunity and so there is no any interference.

But as these are being stored information is being stored and the final state, immune state of the bacteria is obtained after that any infection by that particular virus will be countered by these defense system. So, inference happens at much later point.

(Refer Slide Time: 04:14)



Now, this is the simplified Markov chain model. We have discussed in length earlier. Now, if you look into these in a very simple schematic way you have these foreign DNA here which has invaded the bacterial cell. And we know that PAM sequences are very important in the CRISPR Cas9 system. This does not represent all the class N types, we are discussing the CRISPR Cas9 system here specifically.

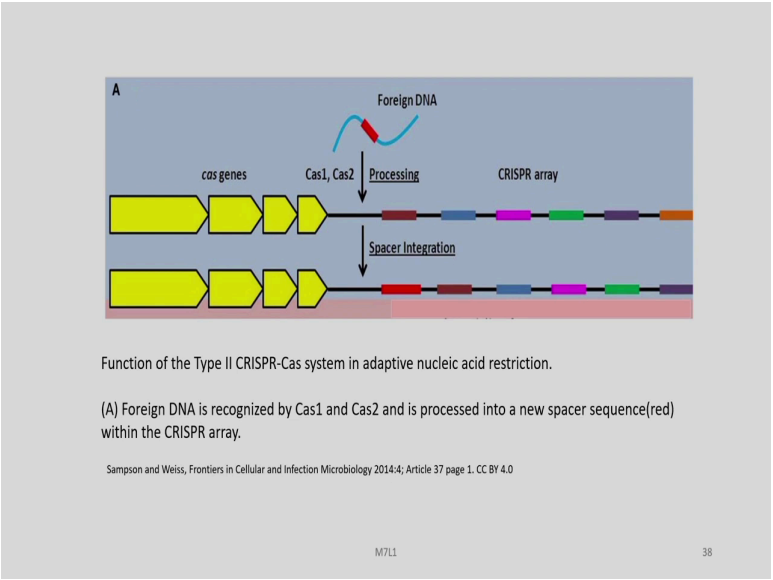
So, this helps in the recognition as well as the adaptation. So, here we know about the bacterial genome which has this loci and of CRISPR Cas9 genes. And then we also know that a tracrRNA is very important. And due to the adaptation or acquisition process, this one small fragment of the invading viral DNA will be incorporated into the existing CRISPR loci.

Now, let us focus a little bit on this tracrRNA . So, this tracrRNA is being transcribed here as you can see. And also the primary transcript or pre CRISPR RNA is transcribed. Now, this tracrRNA has some complementarity with these repeat sequences. And so, these tracrRNA will form complexes or you can see here many of these small tracrRNAs are forming complexes with the primary transcript.

Now, there is some kind of a processing at this stage which cleaves these primary transcripts which are complexed with the tracrRNA into single partners. And forms the matured CRISPR RNA. And this CRISPR RNA goes and in the next generation or next round when the same virus attacks the bacteria the spacer molecule or the DNA which was originally stolen from these virus and stored into the CRISPR RNA forms complementarity, due to the complementarity we will form these assembly here. And in the presence here the Cas9 also plays a role and this will lead to the cleavage of this viral DNA. And thereby in the final stage the bacterium would be able to kill the virus or annihilate the virus.

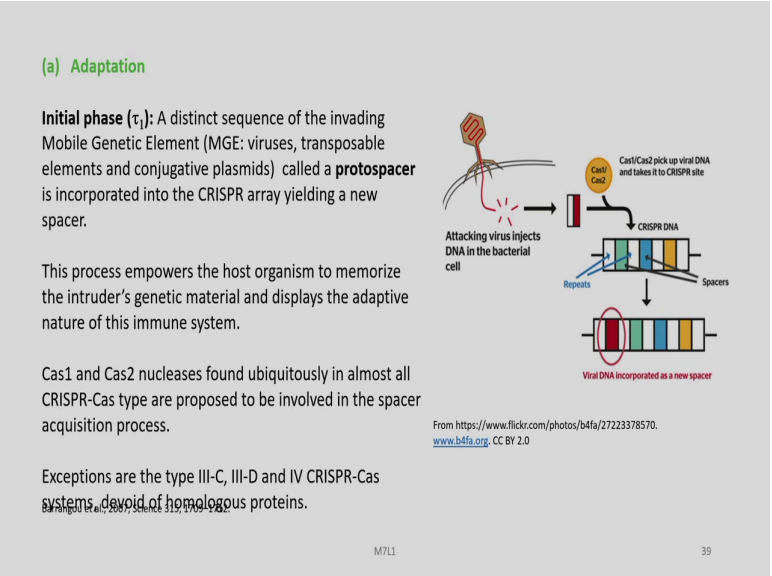
So, this is exactly what happens at the molecular level, but there are very different proteins and various factors involved in each of these stages. We will be trying to discuss what are those proteins in details.

(Refer Slide Time: 07:53)



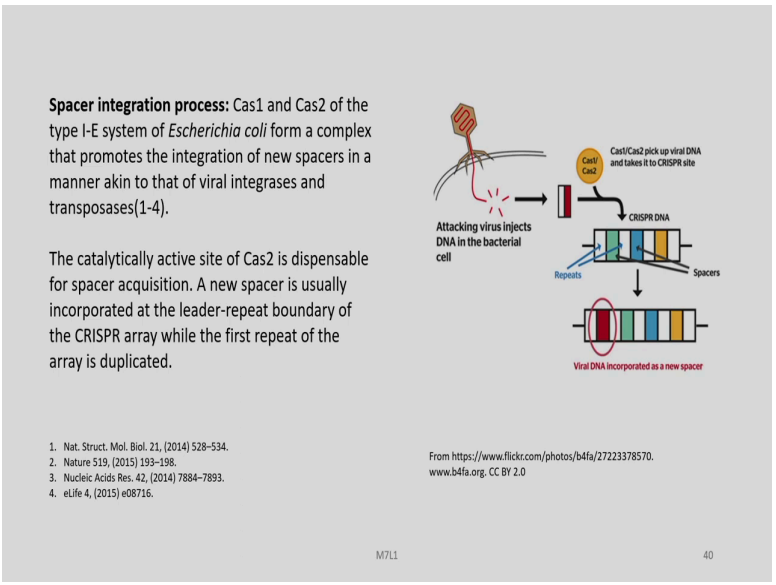
So, now we know the function of the type II CRISPR-Cas system. It is an adaptive nucleic acid restriction or virus annihilation by cleaving its DNA. And then we know about this role of Cas1, Cas2 which is conserved in all types of class and types of the CRISPR Cas system and foreign DNA is recognized by these two.

(Refer Slide Time: 08:23)



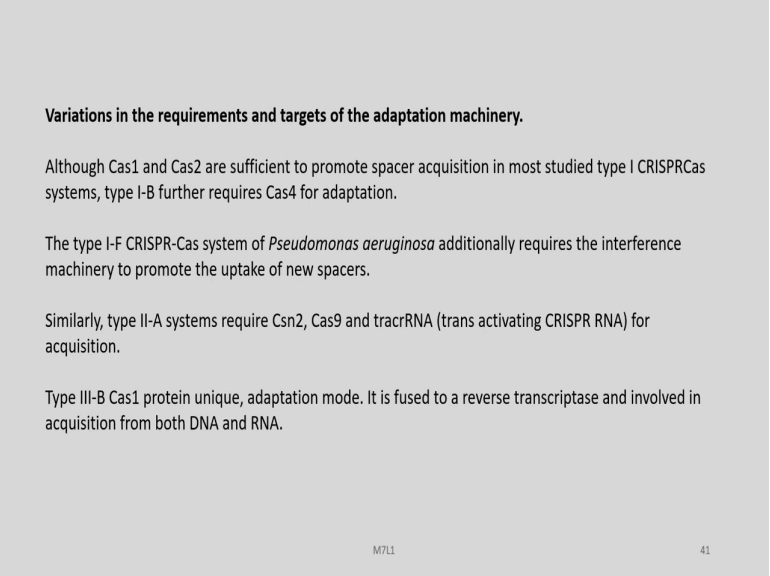
Let us discuss in detail about the adaptation stage. So, this is the initial phase or tau 1. A distinct sequence of the invading mobile genetic elements which may be viruses transposable elements and conjugative plasmids. And these distinct DNA sequence, we call as a protospacer, is incorporated into the CRISPR array yielding a new spacer. This process empowers the host organism to memorize the intruder genetic material and displays the adaptive nature of this immune system Cas1, Cas2 nucleases are found ubiquitously we know about them.

(Refer Slide Time: 09:04)



And in the spacer integration process the Cas1 and Cas2 forms a complex that promotes the integration of the new spacers in a manner which is similar to that of viral integrases and transposases. And we have discussed a lot about these integrases in the preliminary classes. The catalytically active site of Cas2 is dispensable for spacer acquisition. A new spacer is usually incorporated at the leader repeat boundary of the CRISPR array while the first repeat of the array is duplicated.

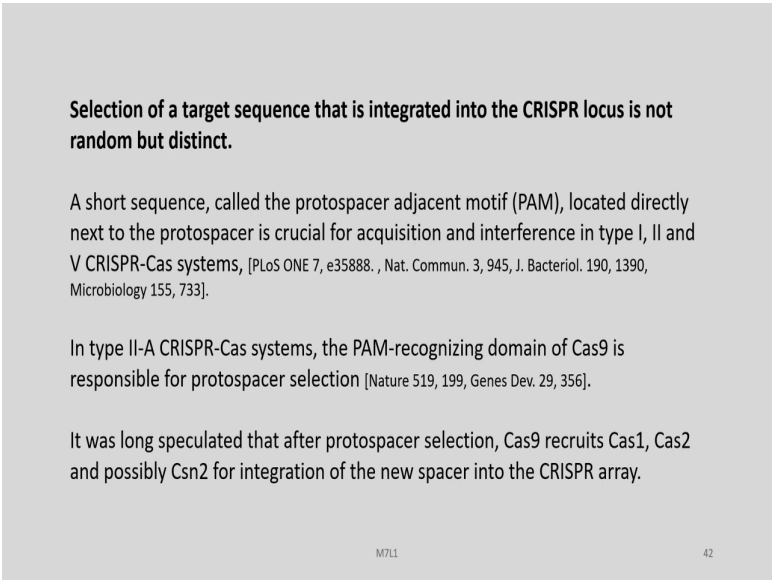
(Refer Slide Time: 09:39)



There are variations in the requirements and targets of the adaptation machinery. Cas1 and Cas2 although is sufficient to promote spacer acquisition in most studied types of type 1 CRISPR systems type I-B requires Cas 4 for adaptation. Type I-F CRISPR Cas system of *Pseudomonas aeruginosa* additionally requires the interference machinery to promote the uptake of new spacers.

Similarly, type I-A systems require CSN2, Cas9 and tracrRNA which we have discussed for acquisition. The type III-B Cas1 proteins has a unique adaptation mode. It is fused to a reverse transcriptase and involved in acquisition from both DNA and RNA.

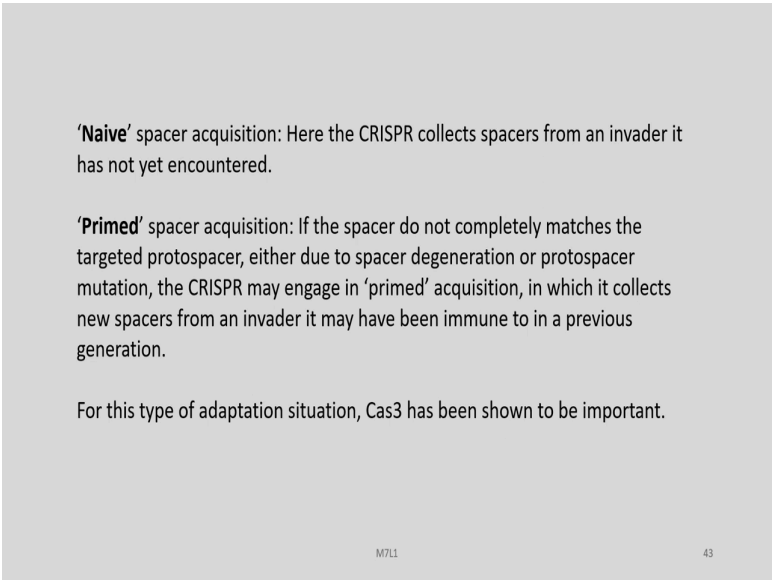
(Refer Slide Time: 10:28)



The selection of a target sequence that is integrated into the CRISPR loci is not random, but it is distinct. We know that a short sequence called the Protospacer Adjacent Motif, PAM which is located directly next to the protospacer is crucial for acquisition in reference in type I, II and V CRISPR Cas systems.

In type II-A CRISPR Cas systems, the PAM recognition domain of Cas9 is responsible for this protospacer selection. And it was speculated for long that after protospacer selection Cas9 recruits is Cas1, 2 and possibly Csn2 for integration for the new spacer into the CRISPR array.

(Refer Slide Time: 11:10)

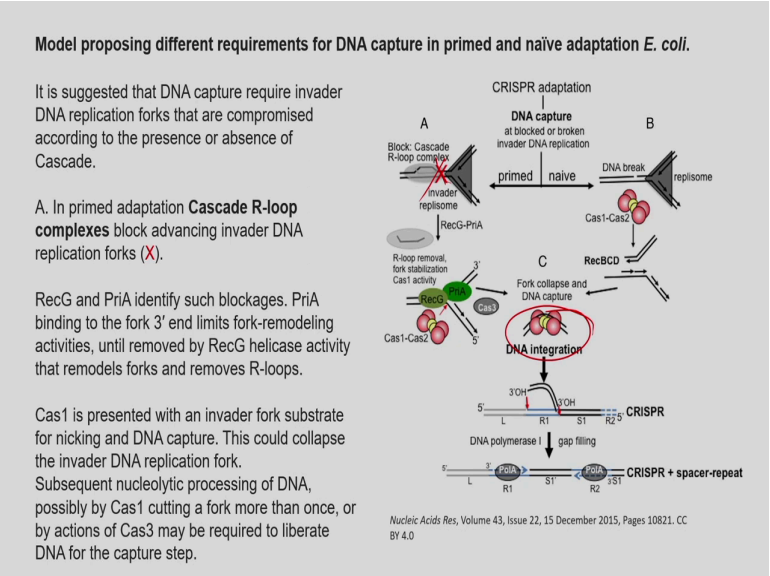


Now, there are two kind of spacer acquisition processes. One is the naive spacer acquisition, the other one is the primed spacer acquisition. In naive spacer acquisition the CRISPR collects the spacers from an invader it has not yet encountered. This is some kind of a debut encounter or the first time a virus attacks a bacteria or a bacteria is exposed to a first time attack of a particular strain of a virus.

While primed spacer acquisition happens when a bacteria encounters a virus which has already attacked it in the past, but due to some reasons the memory may have been lost in certain cases. So, if the spacer do not completely matche the targeted proto spacer either due to spacer degeneration, the spacer had got degenerated or protospacer mutation, the CRISPR may engage in prime acquisition.

So, although the antecedent DNA would ease there due to the earlier attacks, some change in these antecedent DNA has taken place. It may have got totally lost. So, this is not naive in a way, this is called primed spacer acquisitions. So, here it collects new spacers from the invader. It may have been immune to in the previous generation for this type of adaptation situation Cas3 has been shown to be very very important.

(Refer Slide Time: 12:50)



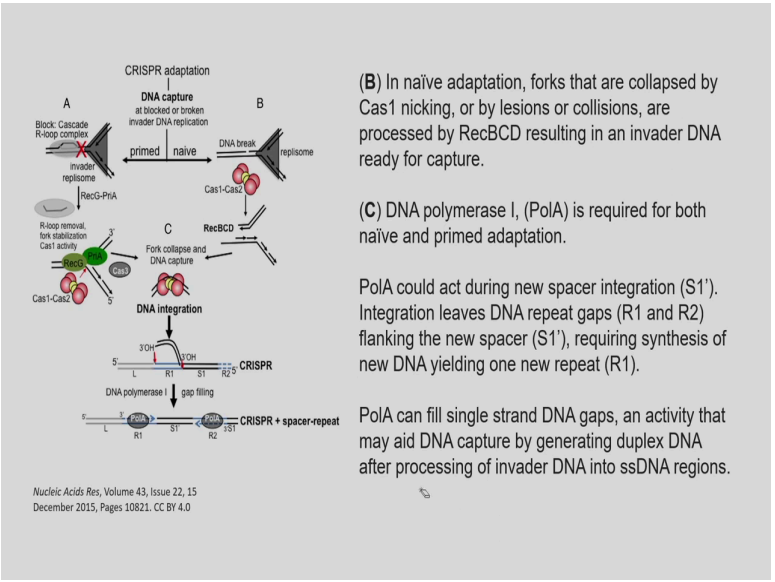
So, you can see here the CRISPR adaptation DNA capture a blocked or broken invader DNA replication. So, you have the primed spacer acquisition on the one side and the naive spacer acquisition on the other side. We will be discussing these at length just for a quick look into these, you have cascade R-loop complex here. And you have a invader replisome here and certain players like RecG, PriA involved here.

And then here you have again Cas1, Cas2 system and you have RecBCD involved here. And both these pathways would lead DNA integration over here and then you have involvement of DNA polymerases and so on. Now, let us discuss in detail exactly what is naive and primed spacer acquisition. It is suggested that the DNA capture requires invader DNA replication forks that are compromised according to the presence or absence of cascade.

In prime adaptation these cascade R-loop complexes block advancing invader DNA replication fork. So, it will be blocked over here. RecG and PriA identify such blockages. This blockage will be identified by these RecG PriA. And PriA binding to the fork 3 prime end limits fork remodeling activities until removed by RecG helicase activity that remodels fork and removes the R-loops.

Cas1 is presented with an invader fork substrate for nicking and DNA capture. This could collapse the invader DNA replication fork. Subsequent nucleolytic processing of DNA possibly by Cas1 cutting a fork more than once or by actions of Cas3 may be required to liberate DNA for the capture step.

(Refer Slide Time: 14:46)

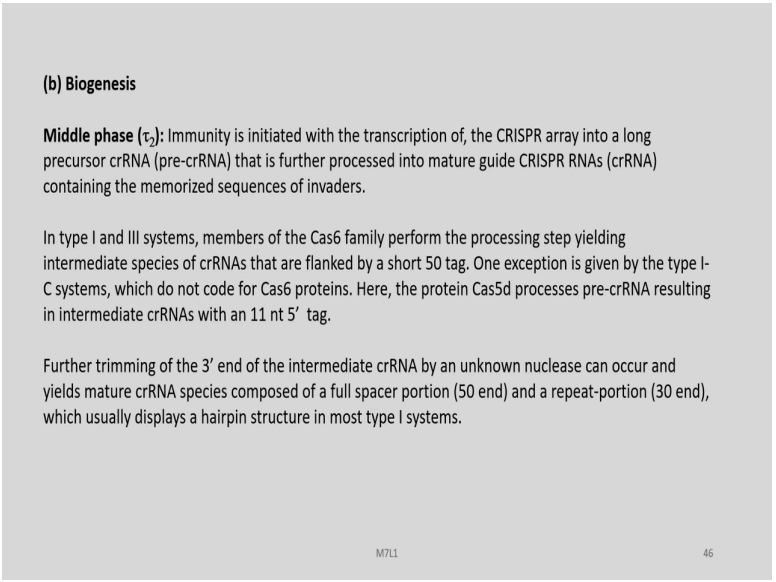


In naive adaptation fork data are collapsed by Cas1 nicking or by lesions or collisions are processed by RecBCD resulting in an invader DNA which is ready for capture. Here in c you can see the DNA polymerase 1 represented by simply PolA. This is required for both naive and primed adaptation

So, the pathway is actually primed and naive merges from these point onwards. So, PolA is required for both the processes. PolA could act during new spacer integration here you can see in S 1 prime integration leaves DNA repeat gaps R 1 and R 2 flanking the new spacer.

So, here S 1 prime requiring synthesis of new DNA yielding one new repeat. PolA can fill the single stranded DNA gaps which is an activity that may aid DNA capture by generating duplex DNA after processing or invader DNA into a single stranded DNA regions.

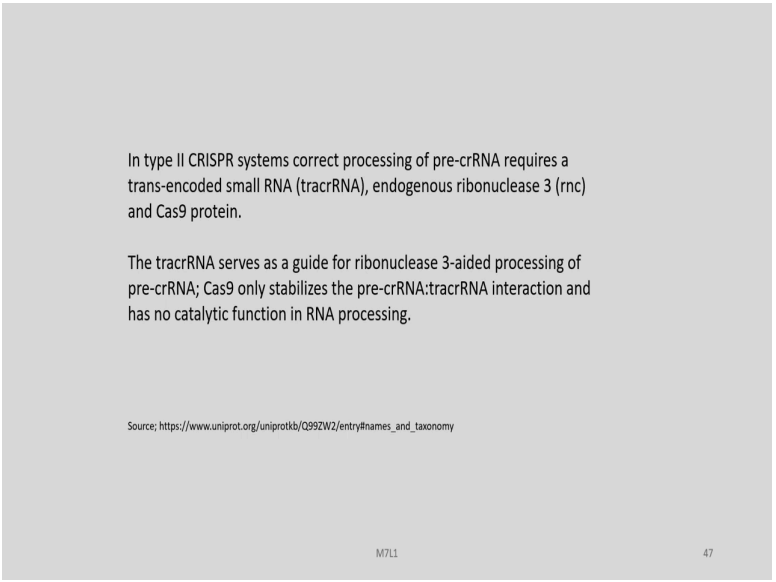
(Refer Slide Time: 16:10)



Let us now go into the middle phase or the tau 2 phase where biogenesis or expression occurs. In this phase immunity is initiated with the transcription of the CRISPR array into a long precursor CRISPR RNA, pre-crRNA that is further processed into mature guide CRISPR RNAs containing the memorized sequences of invades. In type I and III systems members of the Cas6 family perform the processing step yielding immediate species of crRNAs that are flanked by a short 50 tag.

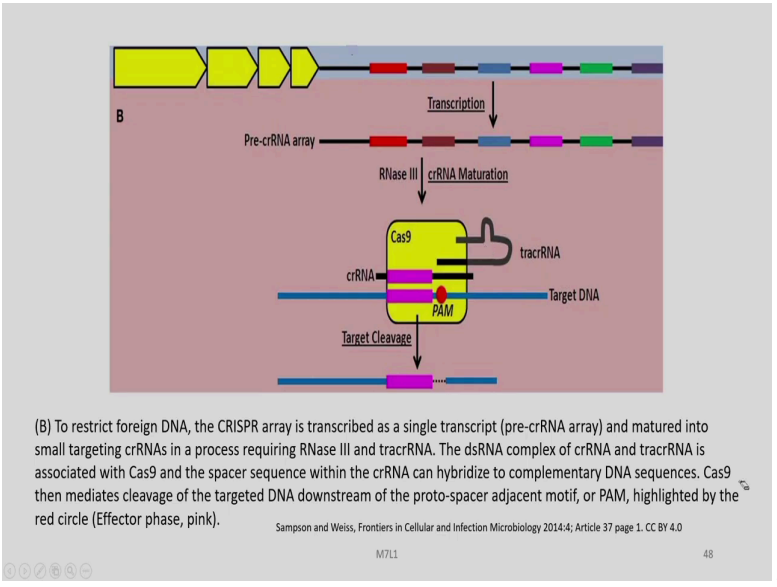
One exception is that by the type one c system which do not code for Cas6 proteins. Here the protein Cas5d processes pre CRISPR RNA resulting in intermediate CRISPR RNA is with an 11 nucleotide 5' tag. Further, trimming of the 3' end of the intermediate CRISPR RNA by an unknown nucleus can occur and yields mature crRNA species composed of a full spacer portion and a repeat portion which usually displays a hairpin structure in most type I systems.

(Refer Slide Time: 17:34)



In type II CRISPR system correct processing of pre CRISPR RNA requires a trans encoded small RNA endogenous ribonuclease 3 and Cas9 protein. The tracrRNA serves as a guide for ribonuclease 3 aided processing of pre CRISPR RNA Cas9 only stabilizes the pre CRISPR RNA tracrRNA interaction and has no catalytic function in RNA processing.

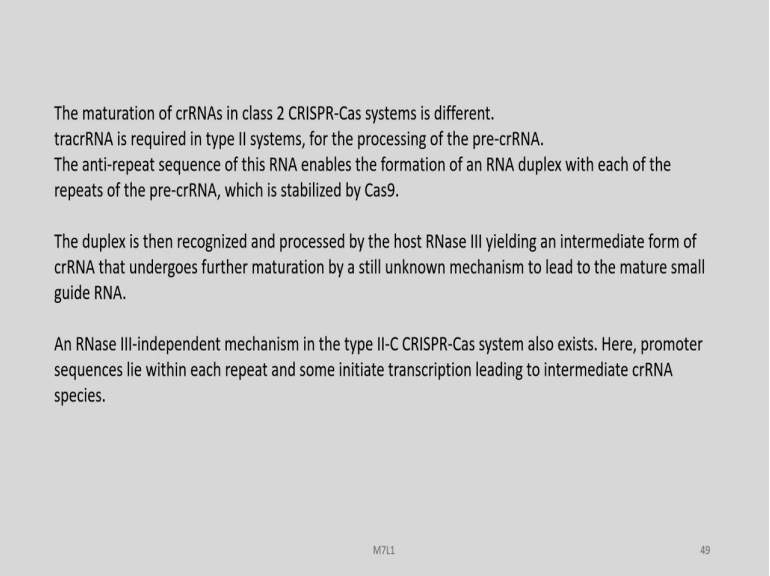
(Refer Slide Time: 18:02)



So, to restrict the Foreign DNA, the CRISPR array is transcribed as a single transcript which we have already discussed. And matured into small targeting crRNAs in a process which requires RNase III as well as tracrRNA. The double stranded RNA complex of crRNA and

tracrRNA is associated with the Cas9. And the spacer sequence within the CRISPR RNA can hybridize to complementary DNA sequences. Cas9 then mediates cleavage of the targeted DNA downstream of the protospacer adjacent motif or PAM which is highlighted by the red circle.

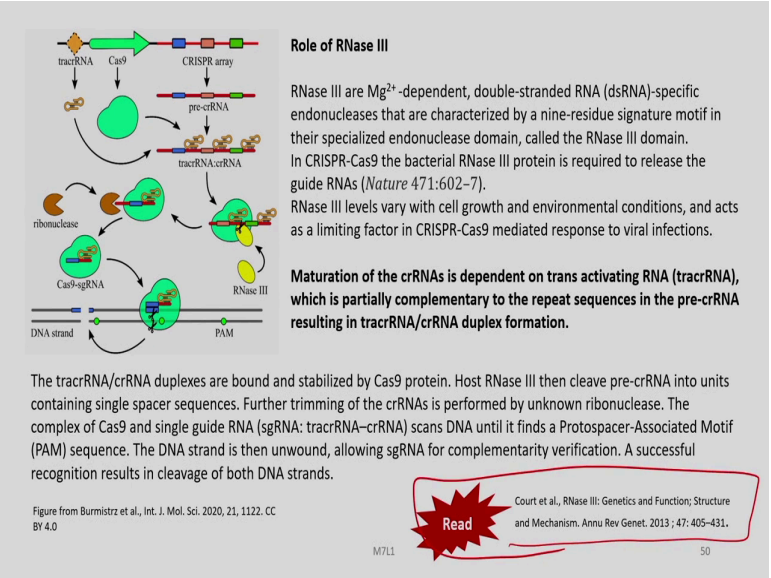
(Refer Slide Time: 18:58)



The maturation of CRISPR RNA Class 2 CRISPR-Cas system is different; tracrRNA is required in type II systems for the processing of the pre-CRISPR RNA. The anti repeat sequence of these RNA enables the formation of an RNA duplex with each of the repeats of the pre-CRISPR RNA which is stabilized by Cas9.

The duplex is then recognized and processed by the host RNase III endonuclease yielding an intermediate form of CRISPR RNA, that undergoes further maturation by a still unknown mechanism to lead to the mature small guide RNA. And RNase III independent mechanism in the type II CRISPR system also exists. Here the promoter sequence lie within each repeat and some initiate transcription leading to the intermediate CRISPR RNA species.

(Refer Slide Time: 19:58)



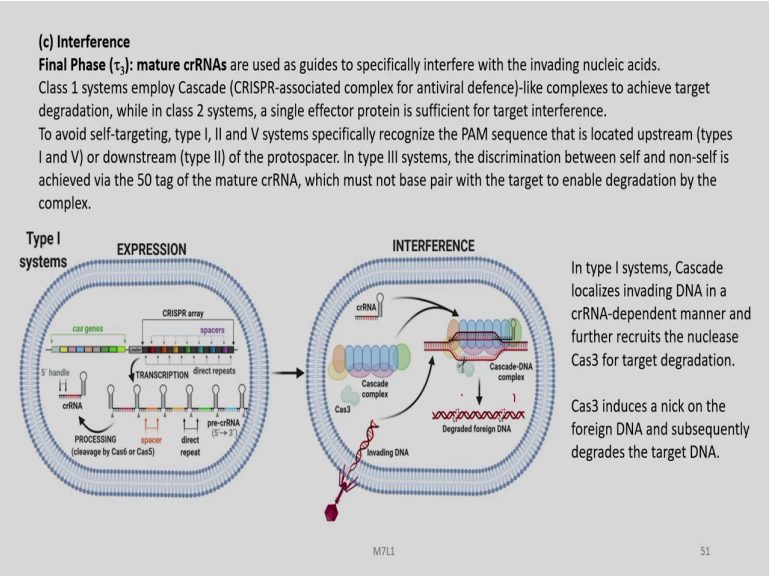
Let us discuss the role of RNase III. The RNase III are magnesium ion dependent double stranded RNA specific endonucleases that are characterized by a 9 residue signature motif in their specialized endonuclease domain called the RNase III domain. In CRISPR Cas9 the bacterial RNase III protein is required to release the guide RNase, RNase III levels vary with cell growth and environmental conditions. And acts as a limiting factor in CRISPR Cas9 mediated response to viral infections.

Maturation of the CRISPR RNase is dependent on trans activating RNA which is partially complementary to the repeat sequences in the pre CRISPR RNA resulting in tracrRNA crRNA, duplex formation as shown in the figure. The tracrRNA crRNA duplexes are bound and stabilized by the Cas9 protein.

Host RNase III then cleave pre CRISPR RNA in the units containing single spacer sequences. Further trimming of the CRISPR RNA is performed by unknown endonuclease as we have shown here. And also discussed in the earlier slide. The complex of Cas9 and single guide RNA will scan DNA until it finds a PAM sequence.

The DNA strand is then unwound allowing sgRNA for complementary verification. A successful recognition will result in the cleavage of both DNA strands. So, for further details in these RNase III genetic structure and function you may refer to this paper in annual review of genetics.

(Refer Slide Time: 21:40)

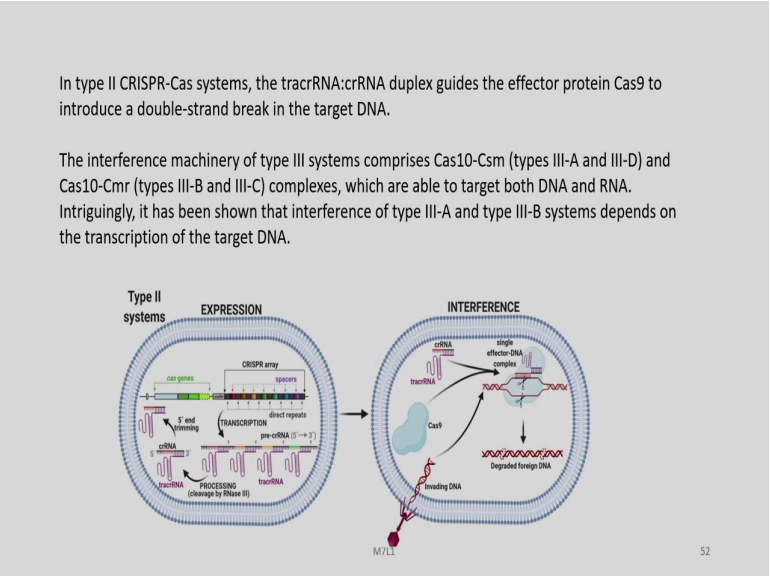


Let us now look into the events that occur at the third phase or the final phase or the tau 3 stage interference. Here the mature CRISPR RNAs are used as guides to specifically interfere with the invading nucleic acids. Class 1 system employs cascade like complexes to achieve target degradation while in class 2 systems, a single effector protein is sufficient for target interference.

To avoid self targeting type I, II and V systems specifically recognize the PAM sequence that is located upstream in type I and V or downstream in type II of the protospacer. In type III systems the discrimination between self and non self is achieved via the 50 tag of the mature crRNA CRISPR RNA which must not base pair with the target to enable degradation by the complex.

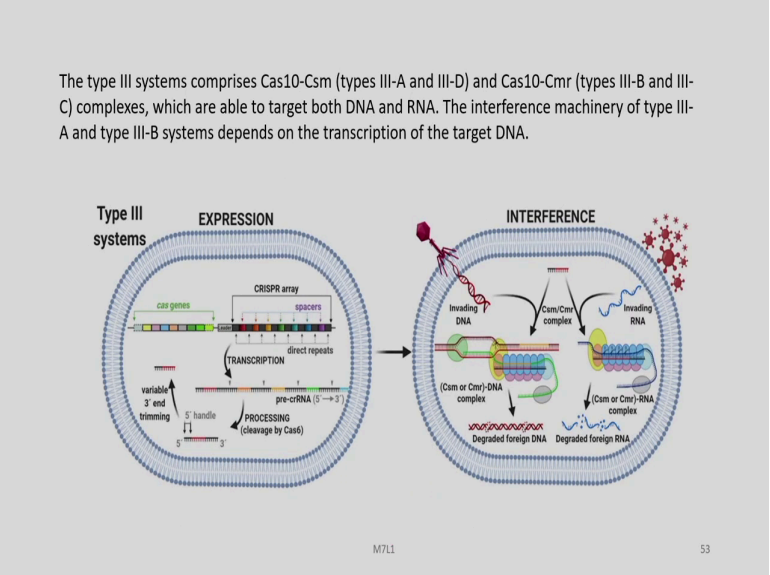
In type I systems cascade localizes invading DNA in a CRISPR RNA dependent manner, and further recruits the nuclease Cas3 for target degradation. Cas3 induces a nick on the foreign DNA and subsequently degrades the target DNA.

(Refer Slide Time: 23:06)



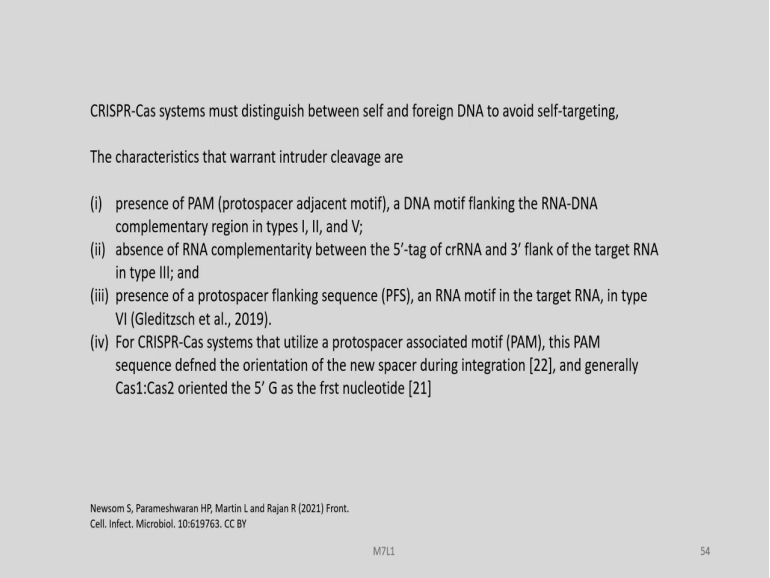
In type II, Cas system as you can see in this picture the tracrRNA crRNA duplex guides the effect of protein Cas9 to introduce a double stranded break in the target DNA. The interference machinery of type III systems comprise Cas10 Csm (type III-A and III-D) and Cas10 Cmr complexes which are able to target both DNA and RNA intriguingly. It has been shown that the interference of type III, and in type III-B systems depends on the transcription of the target DNA.

(Refer Slide Time: 23:39)



This is a type III system which comprises Cas10 Csm and Cas10 Cmr complexes which are able to target both DNA and RNA. The interference machinery of type III-A and type III-B system depends on the transcription of the target DNA.

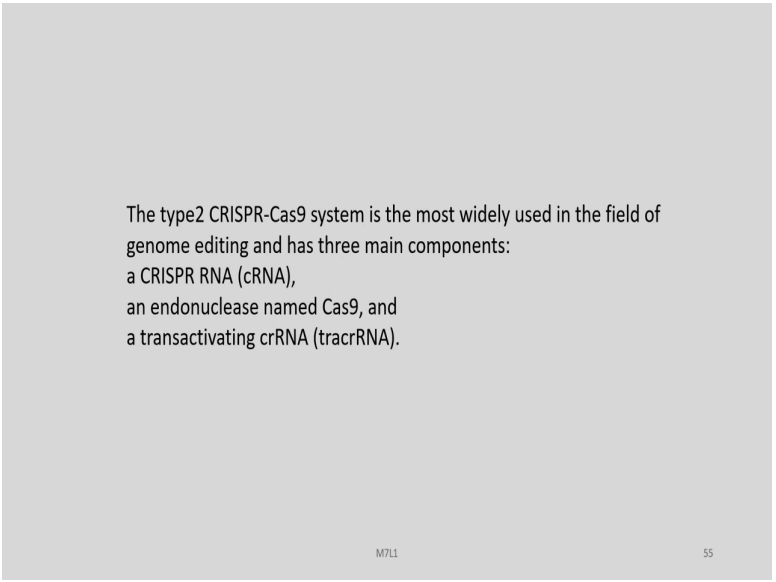
(Refer Slide Time: 24:04)



The CRISPR-Cas systems must distinguish between self and foreign DNA to avoid self targeting. The characteristic that warrant intruder cleavages are presence of PAM, a DNA motif flanking the RNA-DNA complementary region in types I, II and V. Absence of RNA complementarity between the 5 prime tag of CRISPR RNA and 3 prime flank of the target RNA in type III. And presence of a protospacer flanking sequence, PFS and RNA motif in the target RNA in type VI.

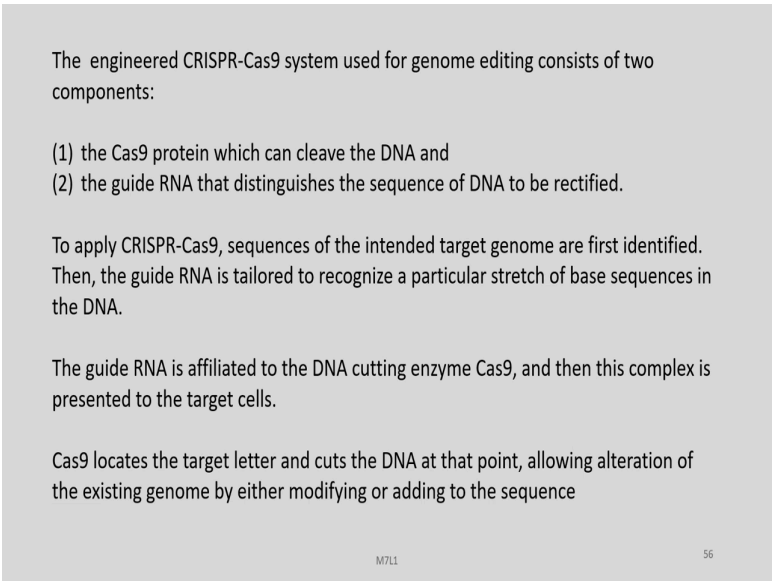
For CRISPR-Cas systems that utilizes a protospacer associated motif PAM. This PAM sequence define the orientation of the new spacer during integration. And generally Cas1, Cas2 oriented the 5’ G as the first nucleotide.

(Refer Slide Time: 25:01)



The type II CRISPR Cas9 system is the most widely used in the field of genome editing. And has three main components as we now know a CRISPR RNA, an endonuclease named Cas9 and a transactivating crRNA.

(Refer Slide Time: 25:18)

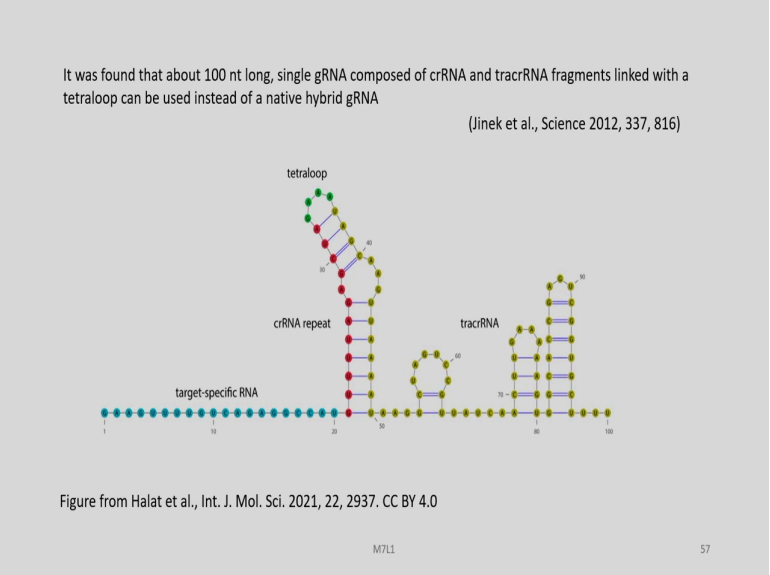


The engineered CRISPR Cas system used for genome editing consists of two components. The Cas9 protein which can cleave the DNA, and the guide RNA that distinguishes the sequence of DNA to be rectified. To apply CRISPR Cas9, sequences of the intended target

genome are first targeted identified. Then the guide RNA is tailored to recognize a particular stretch of the base sequence in the DNA.

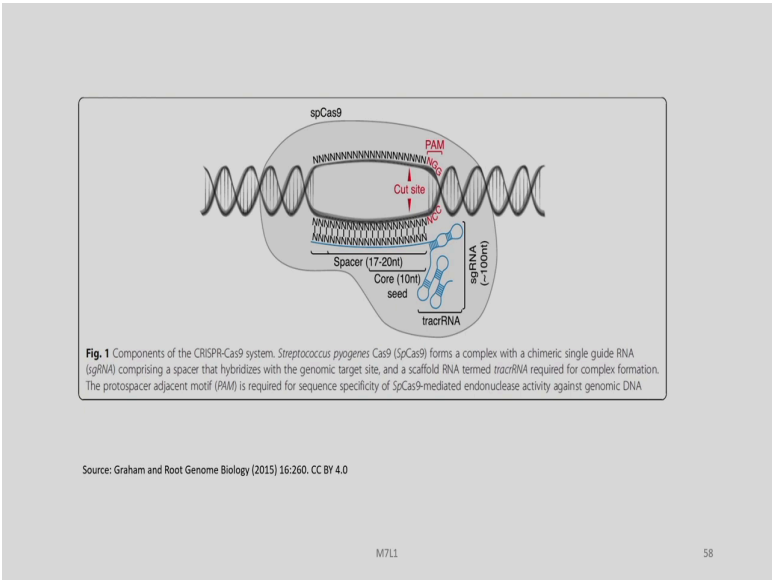
The guide RNA is affiliated to the DNA cutting enzyme Cas9 and then these complex is presented to the target cells. The Cas9 locates the target letter and cuts the DNA at that point allowing alteration of the existing genome by either modifying or adding to the sequence.

(Refer Slide Time: 26:03)



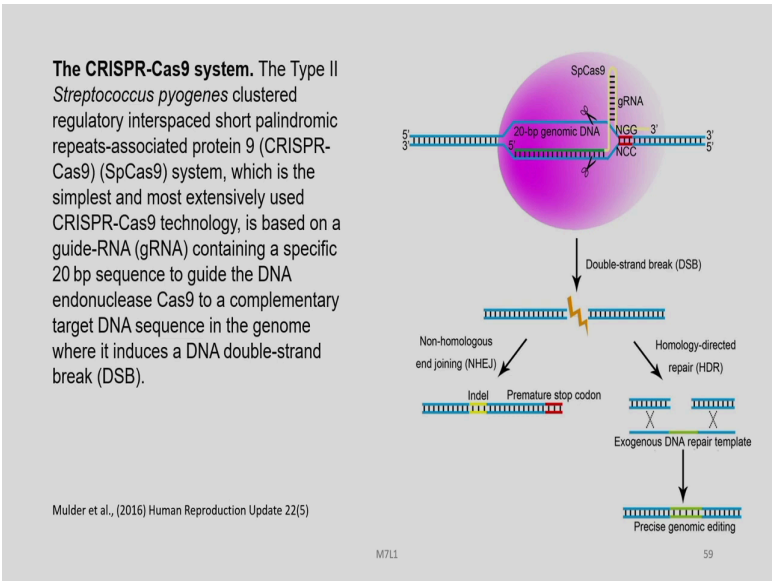
Now, look into this figure little bit closely, it was found at about 100 nucleotide long single guide RNA composed of CRISPR RNA and tracrRNA fragments linked with a tetra loop can be used instead of a native hybrid gRNA. So, this is a synthetic construct. So, we can use this for the CRISPR Cas9 based genome editing.

(Refer Slide Time: 26:45)



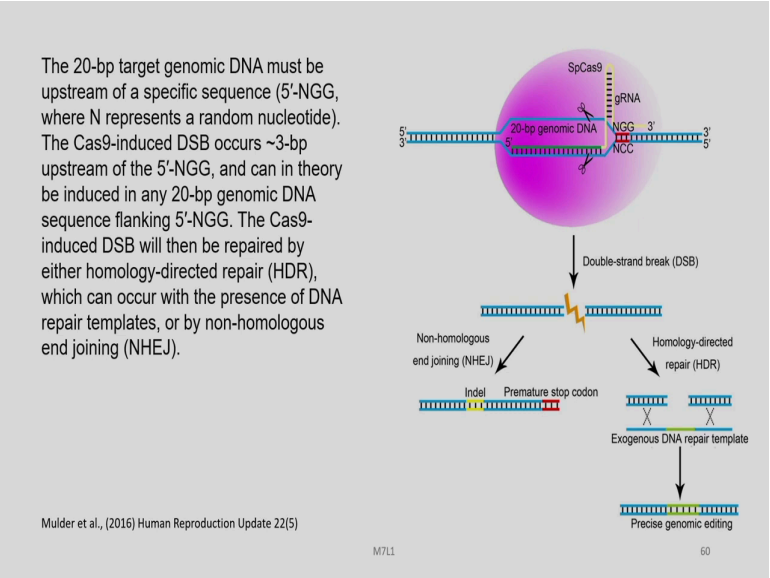
So, this is the components of the CRISPR-Cas9 system. You can see here *Streptococcus pyogenes* Cas9. These big bubble, its form being a complex with chimeric single guide RNA which we just discussed earlier. Comprising a spacer that hybridizes with the genomic target site here, this is the spacer and the scaffold RNA termed tracrRNA required for complex formation. The protospacer adjacent motif is required for sequence specificity of the *Streptococcus* Cas9 mediated endonuclease activity against genomic DNA.

(Refer Slide Time: 27:39)



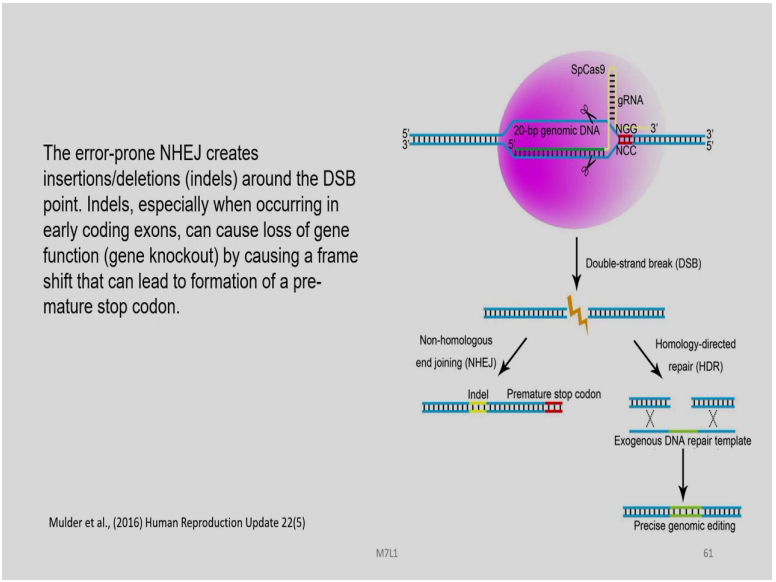
The CRISPR Cas9 system of these *Streptococcus pyogenes* is the simplest and most extensively used CRISPR Cas9 technology. And this is based on the guide RNA containing a specific 20 base pair sequence to guide the DNA endonuclease Cas9 to a complementary target DNA sequence in the genome where it induces a double strand break as you can see here. And then once double strand break happens the part may go towards the NHEJ or towards the homology directed repair. And we have discussed about these two pathways in detail in the earlier lectures.

(Refer Slide Time: 28:26)



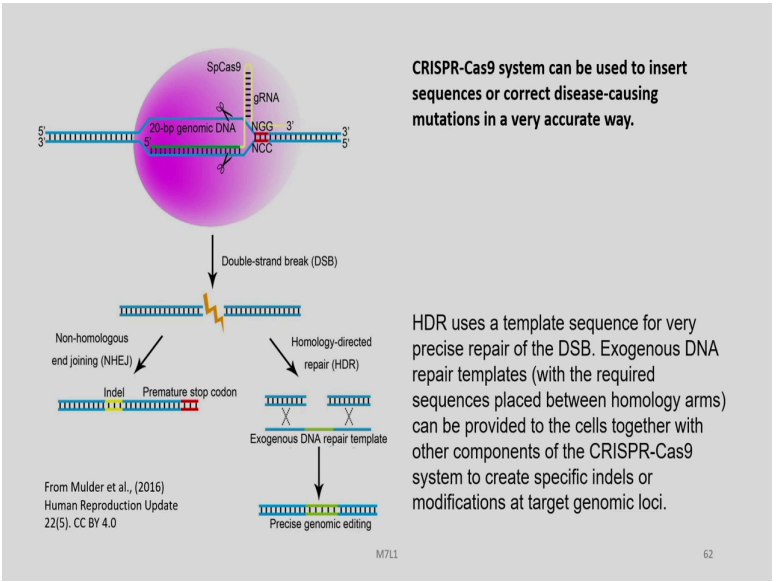
The 20 base pair target genomic DNA must be upstream of a specific sequence 5’ NGG where N represents a random nucleotide, the Cas9 induced DSB occurs about 3 base pair upstream of the 5’ NGG. And can in theory be induced in any 20 base pair genomic DNA flanking sequences. The Cas9 induced DSB will then be repaired by either the homology directed repair which can occur with the presence of a DNA repair template or by the null homologous end joining method.

(Refer Slide Time: 29:03)



The error-prone null homologous end joining creates insertions and deletions indels around the DSB points here especially when occurring in early coding exons and can cause loss of gene function or gene knockout by causing a framework shift that can lead to formation of a premature stop codon.

(Refer Slide Time: 29:26)



CRISPR-Cas9 system can be used to insert sequences or correct disease causing mutations in a very accurate way. HDR uses a template sequence for very precise repair of the DSB. Exogenous DNA repair templates with the required sequences placed between homology

arms can be provided to the cells together with other components of the CRISPR-Cas9 system to create specific indels or modifications at the target genomic loci. We will be discussing about the applications of CRISPR Cas9 technology in the in the next lecture in detail. So, with this we come to an end of today's lecture.

Thank you for your patient hearing.