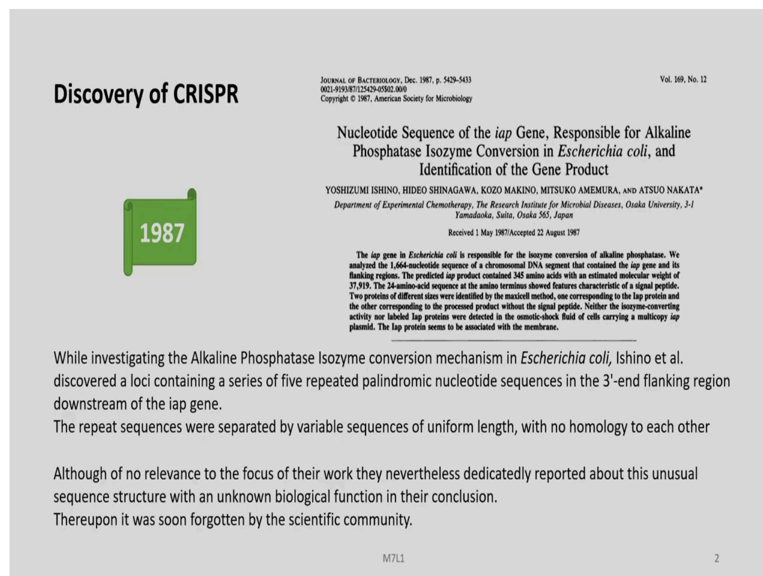
Genome Editing and Engineering
Prof. Utpal Bora
Department of Bioscience and Bioengineering
Indian Institute of Technology, Guwahati

Module - 07
Clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 technology
Lecture - 01
CRISPR system in bacteria - Part A

Welcome to my course on Genomic Editing and Engineering. In this module, we are going to discuss about the CRISPR Cas9 technology. As many of you may be knowing CRISPR stands for Clustered Regularly Interspace Short Palindromic Repeats. In this lecture 1, we are going to discuss about the basic biology of CRISPR system in bacteria.

(Refer Slide Time: 00:52)



Let us start with the work of Ishino et al, who in 1987 was working on an interesting enzyme called alkaline phosphatase which has many isozymes forms. They are looking into the mechanism of conversion of the various isozyme forms in E. coli. And while doing so, they discovered a series of five repeated palindromic nucleotide sequences in the 3 prime end, flanking region downstream of the iap gene.

The repeat sequences are separated by variable sequences of uniform length, with no homology to each other. Although of no relevance to their work they nevertheless dedicatedly reported about these unusual sequence structure with an unknown biological function in their

concluding part of their manuscript. Thereupon it was soon forgotten by the scientific community.

(Refer Slide Time: 01:53)

Universitat d'Alacant
Universidad de Alicante

Ramon Margalef
Multidisciplinary Institute for Environmental Studies

FRANCISCO JUAN MARTÍNEZ MOJICA

In 1992 Francisco Mojica at the University of Alicante in Spain while reviewing genome-sequence data from the extremely halophilic microbe *Haloferax mediterranei,* noticed 14 unusual DNA sequences, each 30 bases long. Each sequence read roughly the same backwards and forwards, and they repeated every 35 bases or so.

Mojica however had to struggle for funds to pursue research on his new discovery and interest and was even ridiculed by peers and decision makers for his interest in these "strange little repeats".

Source University website
https://imem.ua.es/en/about-us/francisco-juan-martinez-mojica.html

"There are many repeats in many organisms — we've known about them for years and still don't know how many of them work" they would say.

And would discourage him by telling "Don't care about repeats so much".

At that time none including Mojica knew that these repeats were going to change the course of the biotechnology revolution in the coming decade.

M7L1 3

In 1992 another researcher called Francisco Mojica working at the University of Alicante in Spain, was reviewing the genome sequence data from an extremely halophilic microbe called Haloferax mediterranei and he notice that in this particular organism there were 14 unusual DNA sequences, each of which was 30 bases long. And, each of these sequences read roughly the same backwards and forwards and we repeat it every 35 bases or so.

Mojica developed a lot of interest in these sequences, however he had to struggle to find funds to pursue research on this new discovery and he was even ridiculed by his peers and decision makers for his interest in these strange little repeats. There are many repeats in many organisms so, we have known them for years and still do not know how many of them work, many of his peers and criticizers would say. And, they would discourage him by telling do not care about repeats so much. At the time no one including Mojica knew that these repeats were going to change the course of biotechnology revolution in the coming decades.

(Refer Slide Time: 03:15)

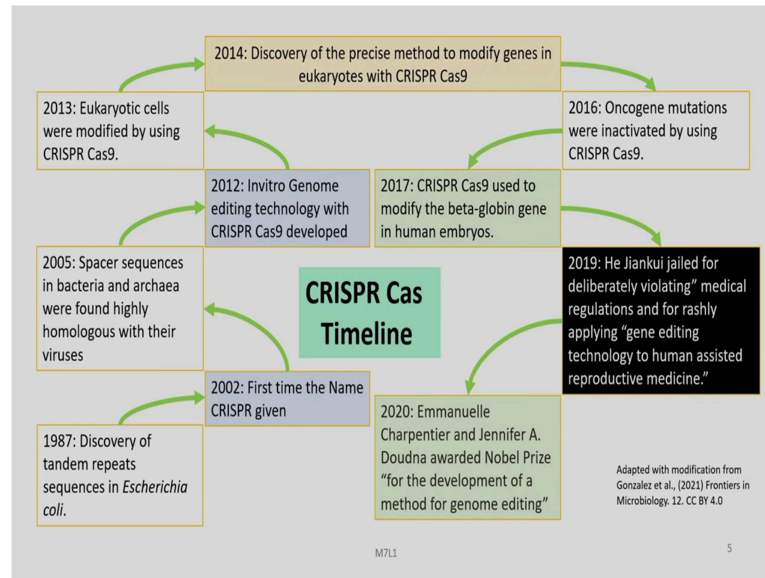
- The term **‘Clustered regularly interspaced palindromic repeats (CRISPR)’** was coined by Mojica in correspondence with a team of Dutch scientists led by ***Rudd Jansen*** at Utrecht University, in 2002.
- Jansen in the same year identified another set of sequence present with CRISPR which is called **CRISPR-associated genes (Cas genes).**
- By 2005, *Alexander Bolotin* found that **'spacer'** sequences between the CRISPR sequences shared similarities with the DNA of viruses.
- They also discovered 2-6 bp long **protospacer adjacent motif (PAM)**, a conserved sequence present at the end of sequence similar to spacers in target .
- It was hypothesised that CRISPR has some role in immunity of host bacteria.

M7L1 4

The term ‘Clustered regularly interspaced palindromic repeats CRISPR was coined by Mojica in correspondence with a team of Dutch scientists led by Rudd Jansen at the Utrecht University in the year 2002. Jansen in the same year identified another set of sequence present with CRISPR which he named as CRISPR associated genes or Cas genes. By 2005, Alexander Bolotin found that space sequence between the CRISPR sequences shared similarities with the DNA of viruses.

They also discovered 2 to 6 base pair long protospacer adjacent motif or PAM, a conserved sequence present at the end of sequence similar to spacers in the target. It was hypothesized that CRISPR has some role in immunity of host bacteria, although it was not understood exactly how.

(Refer Slide Time: 04:14)



Let us look into some of the important discoveries in the field of CRISPR. As we have already discussed about the work of Ishino et al, in 1987 they discovered some tandem repeat sequences downstream of the iap gene in E. coli, while looking for the mechanism of isozymes conversion of alkaline phosphatase. Then, we know about the work of Mojica and it was in 2002 which we have already discussed, that Mojica along with Jansen gave the name CRISPR.

In 2005, we know that the special sequences in bacteria and archaea were found to be highly homologous with the viruses which attack them. By 2012, this basic knowledge was soon converted into a technology. It was possible to do invitro genome editing with CRISPR Cas9 and this was developed in 2012.

The very next year this technology was used for modifying eukaryotic cells by using these CRISPR Cas9 technology. And, in 2014 there was the discovery of the precise method to modify genes in eukaryotes with these CRISPR Cas9. In 2016, oncogene mutations were inactivated by using CRISPR Cas9 technology and 2017 CRISPR Cas9 was used to modify the beta globin gene in human embryos.

In between these discoveries there are many many important discoveries which we are not discussing in detail. Kindly, refer to available literature if you are interested. 2019 is considered a black year in a way, He Jiankui in China was jailed for deliberately violating medical regulations and for rashly applying gene editing technology to human assisted

reproductive medicine. In 2020, amidst the Corona pandemic, Emmanuelle Charpentier and Jennifer A. Doudna was awarded Nobel Prize for the development of a method for genome editing which was basically the CRISPR Cas9 technology.

So, from 1987 up to 2020 you can see the trajectory through which this technology and discovery moved through, right up to the year of 2020 when it was honoured with the highest award. The development has not stopped, many new developments are coming in of late.

(Refer Slide Time: 07:05)

2007

It was Barrangou et al. who found out that these sequences once ignored for their unknown biological function and being ridiculed by peers as mere repeats, now called as CRISPR together with the associated Cas genes, form an adaptive immunity, which provides resistance against bacteriophage infection.

Science

CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes

Abstract

Clustered regularly interspaced short palindromic repeats (CRISPR) are a distinctive feature of the genomes of most Bacteria and Archaea and are thought to be involved in resistance to bacteriophages. We found that, after viral challenge, bacteria integrated new spacers derived from phage genomic sequences. Removal or addition of particular spacers modified the phage-resistance phenotype of the cell. Thus, CRISPR, together with associated *cas* genes, provided resistance against phages, and resistance specificity is determined by spacer-phage sequence similarity.

M7L1 6

We will discuss some of those in this particular course and these two few lectures. Let us go back to the year 2007, where when Barrangou et al. found out that these sequences which were once ignored for their unknown biological functions and was being ridiculed by peers as mere repeats and now named as CRISPR, together with associated Cas genes form an adaptive immunity which provides resistance against bacteriophage infection.

Basically, the immune system provides these kind of defence in higher organisms which are multicellular. But, findings such a corresponding system in a very small single cellular bacteria was quite astonishing. This was published in science as you can see this paper, CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes and is considered a landmark paper.

(Refer Slide Time: 08:04)

Markov chains: Introduced by Andrey Markov in 1906 is used to model a discrete-time, discrete space Stochastic Process of various domains like Finance (stock price movement),, Engineering, Physics (Brownian motion), Genetics (mutations) etc..

The Markov Chain can be described as a triple (S, X, P), a set of states S, with X random variables and a transition probability matrix P.

A Markov Chain is a sequence of random variables, within a finite state space with values in S, for which the transitional probability P, of the state at the time t, is given by the transitional from the state and the time t−1, with probability p (Markov assumption).

The random variables transition from one to state to the other, based on an important mathematical property called Markov Property.

Discrete Time **Markov Property** states that the calculated probability of a random process transitioning to the next possible state is only dependent on the current state and time and it is independent of the series of states that preceded it.

M7L1 7

Let us now go to a concept called Markov chains, many of you may be already knowing. This was introduced by Andrey Markov in 1906 and is used to model a discrete time, discrete space stochastic process of various domains like finance, engineering physics and even genetics. The Markov chain can be described as a triple of S, X and P, a set of states is denoted by S, with X random variables and a transition probability matrix of P.

In simple terms, a Markov chain is a sequence of random variables within a finite state space with values in S for which the transitional probability P, of the state at the time t is given by the transitional form the state and the time t minus 1 with probability p which is called as the Markov assumption. This time t minus 1 is important for the next event, we will discuss about that how.

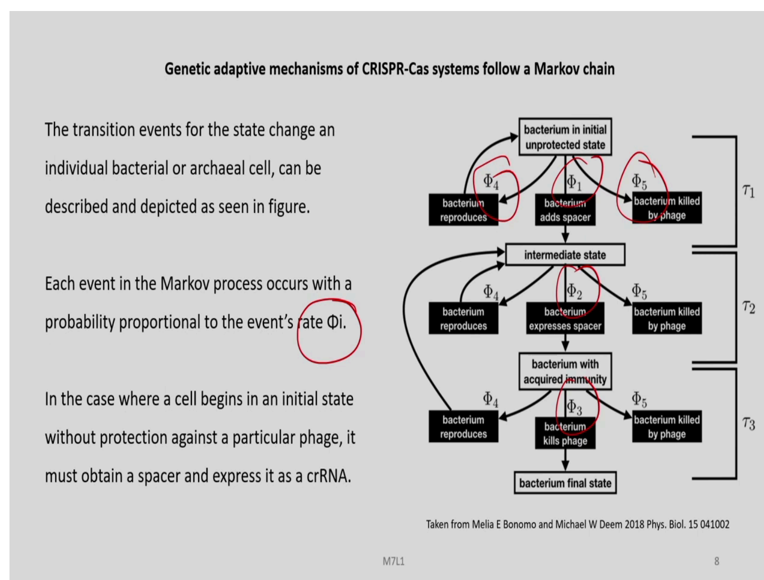
The random variables transition from one state to the other based on an important mathematical property called Markov property. Discrete time Markov property states that the calculated probability of a random process transitioning to the next possible state is only dependent on the current state and the time and it is independent of the series of states that preceded it.

To make it simple, right from your kindergarten days to middle school to high school, you have various levels of performances or you acquired knowledge step by step, right from learning the alphabets. So, finally, when you graduate out into college, what important what is important is your final score in the 10th or 10 plus 2, your score at 10th leads you to the 10

plus 2, where you decide whether to go to the art streams or the commerce streams or the science stream.

And, supposingly you have taken up the science stream, again your performance at the 10 plus 2 decide whether you are going to management, science or engineering or some other discipline. And, this decision is not actually having any impact from your earlier results prior to 10 plus 2 ok. So, now, we are going to use these example to understand some of the important concepts in the CRISPR biology.

(Refer Slide Time: 10:57)



[FL] generic adaptive mechanism of CRISPR Cas follows a Markov chain model. So, here you can see there is an initial unprotected state, where you have a bacteriu, which is not immune to certain viruses. But, something happens in the due course of time, the bacteria acquires some capability and it lens up or goes to a intermediate phase and due to which it acquires immunity.

So, from an unprotected state here, the bacteria gains a protected state and this leads to the final bacterium state. Now, if you look into its state, there are two probabilities. This is a bacteria which is unprotected. This bacteria will undergo reproduction and survive because it is not attacked by any virus. But, its other partners or other daughter cells were attacked by the bacteria and were killed, sorry attacked by the virus and were killed.

In the meantime, the bacteria acquires some weapons and it is in intermediate state. This also again would reproduce and this is not being attacked by any virus so, it survives. But, some of its other daughter cells are attacked by viruses and these are killed. Since, these acquisition of the weapon becomes permanent and the bacteria again reproduces and this is not yet attacked by the virus. So, it survives while a population of it is killed ok.

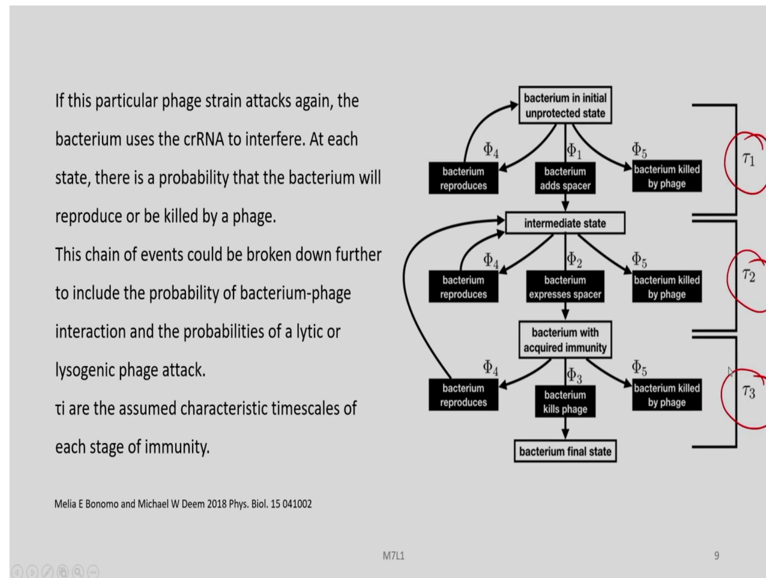
So, which means the immunity although it has been in the stage of acquisition is not yet getting expressed properly and the organism is still being killed. But, after the stage it enters a final stage where the immunity against the invading virus becomes permanent and in this case the bacterium kills the phage. So, you can see from this stage to this stage, the scenario is totally altered.

Now, this bacterium final stage is dependent on the immediate state which is the t minus 1 and it is although dependent on this stage and this stages, but it is the events in this stage which decides the final outcome. And, this is what we call as a Markov chain model. So, basically the transition events for the state change on individual bacterial or archaeal cell can be described and depicted as shown in this figure which we have discussed in detail.

Each event in this Markov process occurs with a probability proportional to the events rate called phi I (Φi). So, this is phi 1, phi 2, phi 3 and the other states phi 4 and phi 5, each state depicting one kind of event. Phi 4 is the event where the bacteria reproduces and survives, phi 5 is the event in which the bacterium reproduces, but it is killed by the phage. And, phi 3 is the stage where the bacterium kills the phage which is altogether the opposite.

In the case where a cell begins in an initial state without protection against a particular phage, it must obtain a spacer and express it as a crRNA which we actually call as a CRISPR RNA.

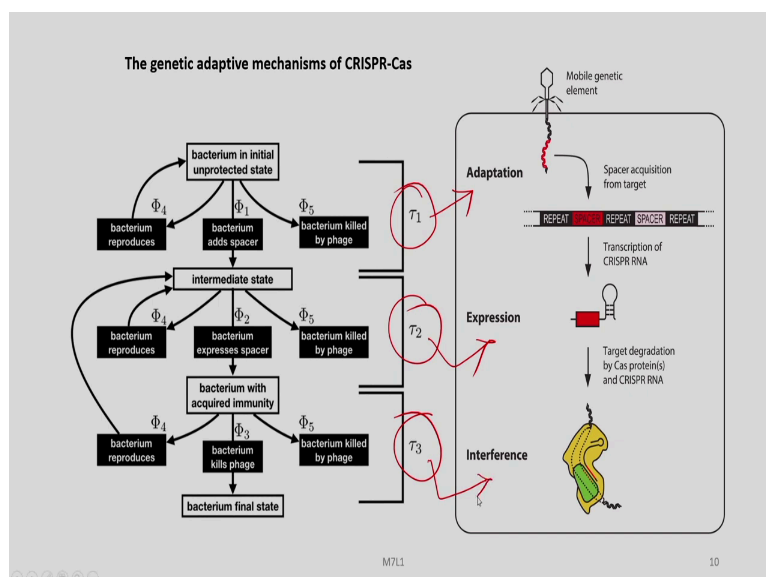
(Refer Slide Time: 15:33)



So, these are the things that we are going to discuss, how this acquisition is being done. If this particular phage strain attacks again, the bacterium uses crRNA to interfere. At each state, there is a probability that the bacterium will reproduce or be killed by a phage.

The chain of events could be broken down further to include the probability of bacterium phage interaction and the probabilities of a lytic or lysosenic phage attack. Tau i (τi) are the assumed characteristics timescales of each stage of immunity. So, we have here tau 1, tau 2 and tau 3 and you can remember this which we are going to use in the future discussions.

(Refer Slide Time: 16:13)



So, the genetic adaptive mechanisms of CRISPR Cas9, CRISPR Cas follows the Markov chain, what we have discussed. And, we also discussed about this tau 1, tau 2 and tau 3. Now, if you look into the figure on the right side, this corresponds to the stage called adaptation, this corresponds to the stage called expression and these correspondence to the stage called interference.

So, we are now trying to describe this Markov chain model in the physical terms, what goes on in the bacteria's adaptability against a viral attack. So, you can see here a virus attacking the bacterial cell and it is pushing inside the genetic material. Something called spacer acquisition from this target is going on. So, this DNA is being cut and incorporated into somewhere, that place is actually basically the CRISPR Cas loci.

And, then these may have been actually attacked by another virus in the past and that genetic information was also included here. And now, this is part of CRISPR Cas loci. So, the bacteria keeps on accumulating all the information DNA fragments from viruses which attacked it in the past. Now, we know that bacteria are very short lived organisms and they divide E. coli divides as fast as in 22, 20 to 22 minutes.

So, these sequences which are acquired in one single generation would actually be passed on to the next generation. And, that next generation would add up a new information resulting out of the viral attack. So, we know from this small discussion that this is also a heritable character. So, we are going to discuss about that again in the future. Now, there is a state is called expression where these loci is transcribed and to produce CRISPR RNA. And, then these CRISPR and this is called the expression stage and order intermediate stage. And then finally, these will lead to the interference or the where the bacterium will kill the virus.

(Refer Slide Time: 18:44)

The CRISPR locus, is present in about 84% of archaea and 45% of bacteria according to CRISPRdb.

CRISPR-Cas locus function as adaptive and heritable immune system of prokaryotes against invading virus.

Upon infection by virus, the viral DNA is acquired and integrated into CRISPR sequence

Future infections leads to targeted cleavage of the invading viral DNA having similarity to integrated spacer sequence in CRISPR

Grissa et al., 2007. Nucleic Acids Res., 35 (Web Server issue), pp. W52-W57

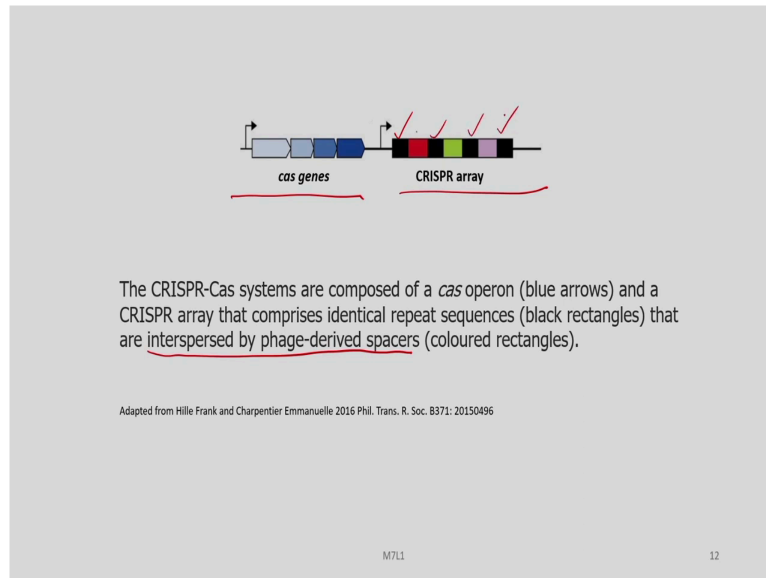
M7L1 11

So, the CRISPR locus is now little bit familiar to you, this is present in about 84 percent of archaea and 45 percent of bacteria according to CRISPRdb. CRISPR Cas locus functions as an adaptive and heritable immune system of prokaryotes against invading viruses. So, in our case whenever we take a vaccine that last our lifetime and then if an individual is vaccinated in childhood, that immunity can go up to adolescence and with booster doses, it may be going a little bit farther.

But, in the case of CRISPR Cas immunity in bacteria, it actually passes on from generation to generation because, here the immunity is stored in the form of information in its chromosome. While the immunity that we acquire by say vaccination does not go to the next generation. However, of course, in certain cases there is a mother to child transmission of immunity, but it would not happen in the male population.

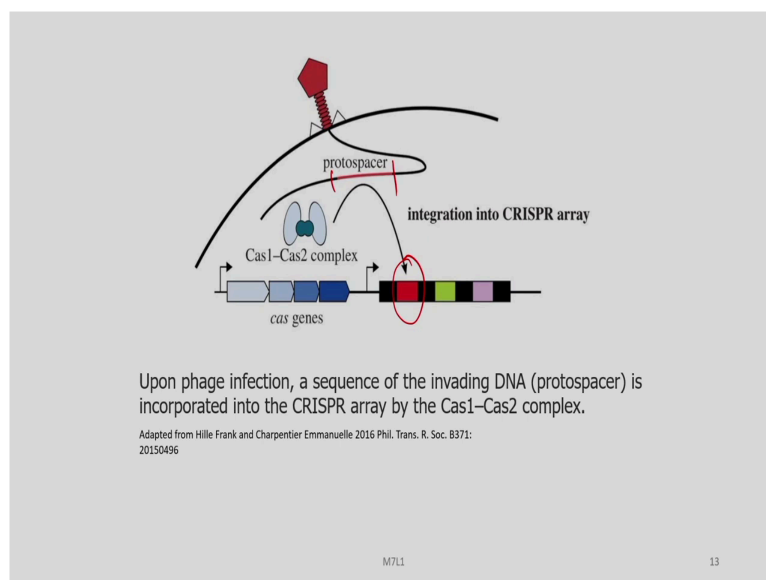
So, upon infection of the virus, the viral DNA is acquired and integrated in the CRISPR sequence as we have shown in the figure earlier. Future infections would lead to targeted cleavage of the invading viral DNA having similarity to integrated space sequence of CRISPR.

(Refer Slide Time: 20:22)



Now, you can see this in simple terms the Cas CRISPR-Cas loci where you have the Cas genes and you have the CRISPR array. Now, these are all interspersed by phage derived spacers, the coloured angles and these are the dark ones are the repeats.

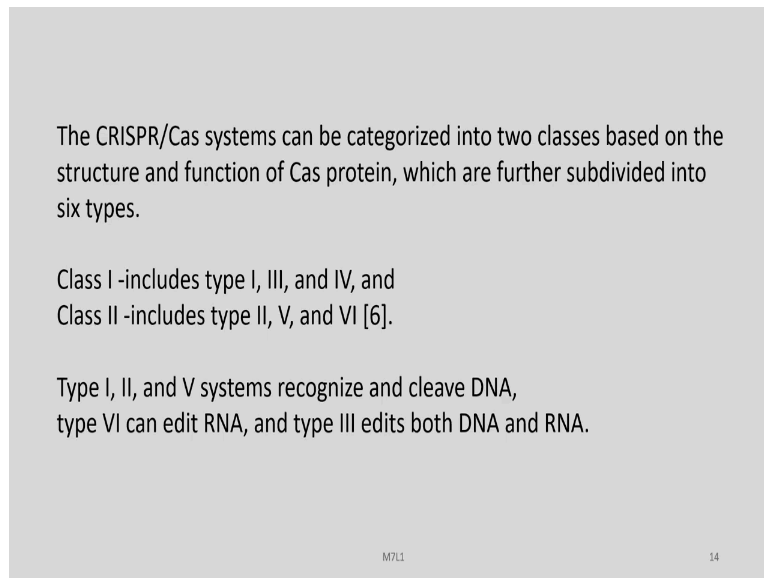
(Refer Slide Time: 20:46)



Let us see a little bit closer how these system works. So, once a virus attacks a bacterial cell and it injects the DNA inside the bacterial cell, there is a system inside the bacteria; the CRISPR Cas loci as you already know. They produce a complex called Cas 1-Cas 2 complex.

This Cas 1-Cas 2 complex will go and cleave a selected portion of these viral DNA and insert it into the existing CRISPR array. So, with each infection, the length of this array keeps on increasing. So, these CRISPR array is also kind of a scoreboard or scorecard where we can know how many times this particular bacterial strain has been infected by virus in the past.

(Refer Slide Time: 21:52)



The CRISPR Cas system can be categorized into two classes based on the structure and function of Cas protein, which are further subdivided into six types. So, class I includes type I, III and IV; class II includes type II, V and VI. Type I, II and V systems recognize and cleave DNA, type VI can edit RNA and type III edits both RNA and DNA.

(Refer Slide Time: 22:20)

All known CRISPR-Cas systems contain Cas1 and Cas2.

The three different types (Type I, II, and III) are each characterized by unique Cas proteins involved in maturation of the crRNAs, targeting of foreign nucleic acid, and nucleic acid cleavage (Makarova et al., 2011).

The Type II CRISPR-Cas system is defined by the presence of a large (~1000–1600 amino acids) endonuclease, Cas9.

M7L1 15

As we have already shown that Cas1-Cas2 system selects the fragment of DNA to be incorporated into the CRISPR Cas loci. And, today we know that all known CRISPR Cas systems contain these two proteins Cas1 and Cas2.

The three different types I, II and III are each characterized by unique Cas proteins involved in maturation of the CRISPR RNAs, targeting of foreign nucleic acid and nucleic acid cleavage. The type II CRISPR-Cas system is defined by the presence of a large 1000 to 1600 amino acid endonuclease called as Cas9.

(Refer Slide Time: 23:03)

Role of Cas1 and Cas 2 in psDNA acquisition by the host CRISPR array for the CRISPR-Cas adaptive immunity

Acquisition of enemy DNA happens in 4 steps:

i. protospacer binding and selection,
ii. 3′ overhang cleavage,
iii. integration, and
iv. DNA synthesis and repair.

In integration step, PAMc sequence C′-OH (cleavage at site1 of Cas1-Cas2) is integrated into the spacer side of Repeat1 in the CRISPR array (likely via the second nucleophilic attack), whereas non-PAMc (cleavage at site2) is integrated into the leader side of Repeat1 (likely via the first nucleophilic attack) .

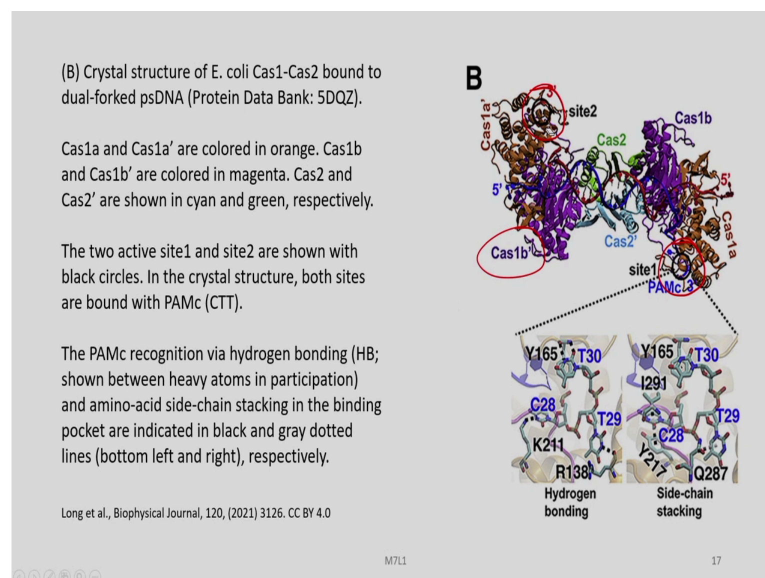
Long et al., Biophysical Journal, 120, (2021) 3126. CC BY 4.0

M7L1 16

Let us look into the role of Cas1 and Cas2 in protospacer DNA acquisition by the host CRISPR array for the CRISPR-Cas adaptive immunity. So, acquisition of the enemy DNA which is basically the viral DNA happens in 4 steps. The first step is the protospacer binding and selection. Second step is a 3 prime overhang cleavage as you can see here in this figure. The third is the integration and the fourth is the DNA synthesis and repair.

In the integration step, PAMc sequences in C prime which the cleavage at site1 of Cas1-Cas2 is integrated into the spacer side of Repeat1 in the CRISPR array, likely via the second nucleophilic attack. Whereas, the non-PAMc cleavage at site2 is integrated into the leader side of Repeat1, likely via the first nucleophilic attack.

(Refer Slide Time: 24:17)



Let us look into the crystal structure of these Cas1-Cas2 bound to dual forked proto spacer DNA from E. coli. So, you can see here Cas1a, Cas1a prime, this is Cas1a and this is Cas1a prime. Similarly, you have Cas2 and Cas2 prime, then you have Cas1b and Cas1b prime. And, they are shown in their respective colours to make them distinct from each other. The two active sites 1 and 2 are shown with black circles, you can see here a black circle maybe not visible clearly.

But, within this circle if you look into, you can see a black circle over here which has become red, now due to my colouring. So, these are the active sites 1 and 2. In a crystal structure, both sides are bound with PAMc CTT. The PAMc recognition via hydrogen bonding and

amino acid side chains stacking in the binding pocket are indicated in black and grey dotted lines respectively, here you can see the hydrogen bonding and the side chain sticking.

(Refer Slide Time: 25:45)

Type II CRISPR-Cas systems are further characterized by the requirement for a unique, accessory RNA, the trans-activating CRISPR RNA (tracrRNA), as well as RNase III, for maturation of crRNAs (Deltcheva et al., 2011; Jinek et al., 2012; Chylinski et al., 2013).

Cas9 is involved as a scaffold for maturation of crRNAs, and is required for cleavage of the double stranded DNA target.

tracrRNA Cas9 cas genes crRNA crRNA crRNA

M7L1 18

Type II CRISPR-Cas system are further characterized by the requirement for a unique, accessory RNA, the trans-activating CRISPR RNA or the tracrRNA as well as RNase III, for maturation of CRISPR RNAs. Cas9 is involved in this scaffold for maturation of CRISPR RNAs and is required for cleavage of the double stranded DNA target. So, you can see here the various genes in these particular loci.

(Refer Slide Time: 26:26)

Class 1 and class 2 CRISPR-Cas systems: key features, modular organization.
(*a*) The general architectures of class 1 (multiprotein effector complexes) and class 2 (single-protein effector complexes) CRISPR-Cas systems. Genes are shown as arrows; homologous genes are shown by the same colour. Gene names follow the current nomenclature and classification.

(*a*) class 1 and class 2 systems and 13 core gene families

class 1: cas3, cas8 or cas10 (large subunit), cas11 (small subunit), cas7, cas5, cas6, cas1, cas2, cas4, spacers, CRISPRs

class 2: cas9 or cas12 or cas13 (single effector protein), cas1, cas2, cas4, spacers, CRISPRs

Figure adapted from Koonin and Makarova (2019) Phil. Trans. R. Soc. B 374: 20180087. CC BY 4.0

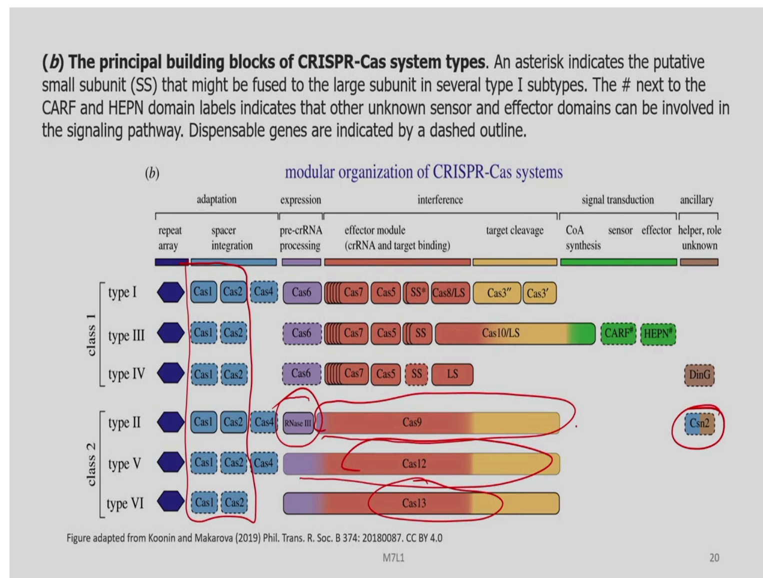
M7L1 19

Let us look into class 1 and class 2 CRISPR system, a little bit closely. What are the key features and the modularity of their organization? So, you can see here in figure a, the general architectures of class 1 which is having multiprotein effector complexes and class 2 which is having single protein effective complexes. Here, the genes are shown as arrows, homologous genes are shown by the same colour. Genes names follow the current nomenclature and classification.

And, here are the CRISPR repeats and here are the spacers. So, spacers and repeats you already know in detail and these are the associated genes, CRISPR associated genes. And, you can see here in class 1, Cas1 and Cas2 are present in both in fact, in all CRISPR Cas system as we have already discussed.

But, you will see that besides these two conserved highly conserved genes in all the systems, there is variation in the other Cas genes. As already told to you class 2 has only a single effector protein which maybe Cas9, Cas12 or Cas13 whereas, in type II you have so, many Cas 3, Cas 8, 8 or 10, 11, 7, 5, 6 and so on.

(Refer Slide Time: 28:09)



So, this is very important for our future understanding of the CRISPR Cas system in bacteria. So, what are the principal building blocks of CRISPR-Cas system which constitute the diverse types? So, here you see the class 1 which we described earlier and class 2 and the types under these classes, respective classes. And, you can see here the various modules used for adaptation and expression and interference. And, also some modules associated with the

signal transaction and others are having certain ancillary sequences which may having some helping role or some kind of unknown roles.

So, we already spoke about this. For example, type II is having Cas9 and then V is having Cas Cas12 and Cas13. And, in all of these you can see Cas1 and Cas2 are invariably present. In type II CRISPR Cas9 system you also have a protein called RNase III and then you have Csn2. So, we have to remember these proteins in future while discussing the CRISPR Cas9 system.

(Refer Slide Time: 29:39)

Table 1 Differences between type I, II, III, and V CRISPR/Cas systems

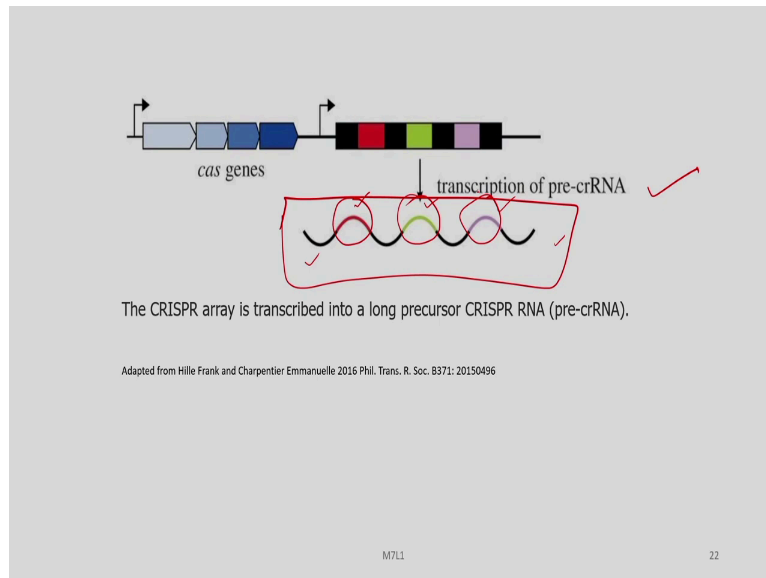
Classification	Type I	Type II	Type III	Type V
Signature protein	Cas3 (or Cas3′)	Cas9 (1368 amino acids)	Csm (III-A) or Cmr (III-B)	Cas12a (1200–1300 amino acids)
Effector	Cascade	crRNA and tracrRNA (sgRNA)	Cascade	crRNA
PAM sequence	3-nt	G-rich sequence, 5′-NGG-3′	Without PAM	5′-YTN-3′(FnCas12a), 5′-TTTN-3′(AsCas12a, LbCas12a)
Cleavage product	SSBs	DSB (flat end)	SSBs at every 6-nt	DSB (Sticky end with 5 nucleotides protruding)

Liu, Z., Dong, H., Cui, Y. *et al.* Application of different types of CRISPR/Cas-based systems in bacteria. *Microb Cell Fact* **19**, 172 (2020). CC BY 4.0

M7L1 21

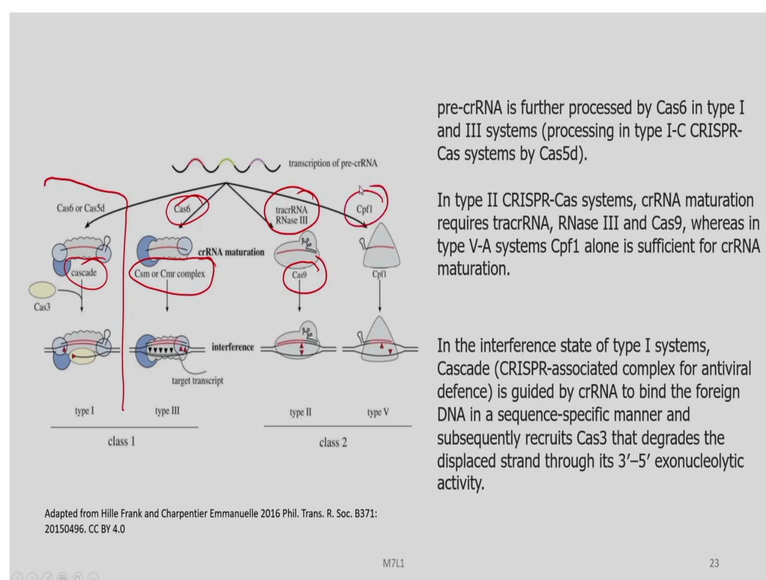
So, this is a tabulated form about the differences between the various types of CRISPR-Cas system. And, you can see here the type and you can see here the signature protein against each type. You have type II Cas9, very famous protein and then in type III you have Csm Cmr. In type I, you have Cascade, we will have a little bit of discussion about what is Cascade. Then, you have the various cleavage products, here you have single strand SSBs and you have DSBs and so on.

(Refer Slide Time: 30:29)



Now, let us look into the expression of these CRISPR-Cas loci. Now, let us look into the expression of the CRISPR-Cas loci. So, we know about this CRISPR loci, CRISPR-Cas loci now. So, these loci or genes get transcribed into a pre CRISPR RNA. So, here you have all the repeats and all the spacers expressed as a single transcript. And, you have to remember that these are basically the sequences of the viruses which attack these bacterial strain in the historical past. So, they are basically antecedent DNA.

(Refer Slide Time: 31:28)



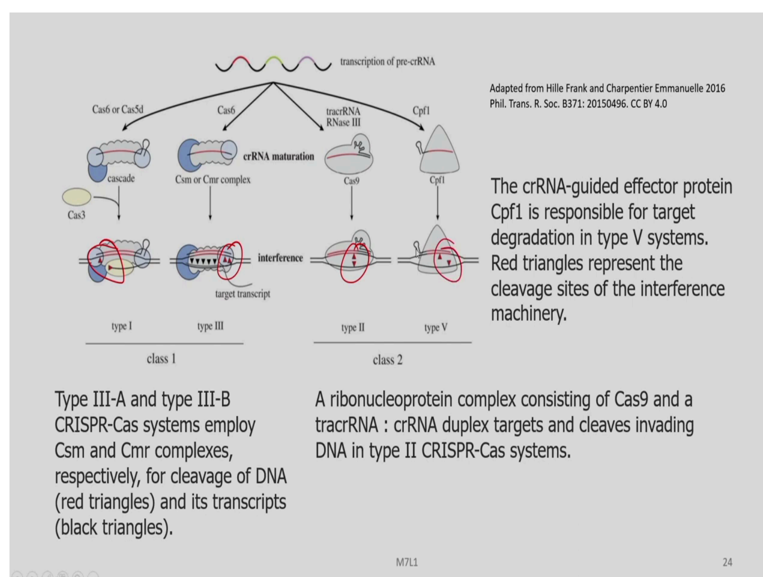
So, once these transcription happens and the formation of the pre CRISPR RNA happens; now depending on the class of the CRISPR-Cas system which we just discussed prior to this slide and the type, the processing would be different ok. So, this is the diversity and variation in this processing. So, you can see here in under class 1, we have type I and type III, under class 2, type II and type V.

So, these pre CRISPR RNA is handled in different ways, but different proteins in this different class and different types. So, for example, in this class 1 type you have the role of Cas6 or Cas5d or cascade here. Here, Cas6 plays a role and it involves the Csm or Cmr complex. And, in type II you have class 2, type II you have tracrRNA, RNase III and Cas9. And, here you have the Cpf1 and then the interference mechanism will also be different.

We will be discussing some of these in detail in the next slides. So, as already told you here in this description of the figure, these pre-CRISPR RNA is further processed by Cas6 in type I and III systems. In type II CRISPR-Cas system, CRISPR RNA maturation requires this tracrRNA, RNase III and Cas9, whereas in type V-A system cpf is alone enough for CRISPR RNA maturation.

In the interference state of type I systems, cascade which is basically CRISPR associated complex for antiviral defence is guided by CRISPR RNA to bind the foreign DNA in a sequence-specific manner and subsequently recruits Cas3 that degrades the displaced strand through its 3 prime to 5 prime exonucleolytic activity.

(Refer Slide Time: 33:54)



So, here the type III-A and type III-B CRISPR Cas system employs Csm and Cmr complexes respectively for cleavage of DNA and its transcripts here shown by the black and red triangles. And, class 2 a ribonucleoprotein complex consisting of Cas9 and a tracrRNA, CRISPR RNA duplex targets the cleavage invading DNA in type II CRISPR-Cas system.

And, while in V, the CRISPR RNA guided effector protein Cpf1 is responsible for target degradation. And, the red triangles here represent the cleavage sites of the interference machinery. So, you can see these sites are all in different locations depending on the class and type.

(Refer Slide Time: 34:48)

Type I and type II systems use short 3–7 bp protospacer-adjacent motifs (PAMs) for the identification of suitable protospacer for acquisition.

During naive adaptation by type II systems, PAM is recognized by Cas9 while in type I-E systems, Cas1 and Cas2 are sufficient for recognition of PAM.

Following the protospacer selection and processing, the acquisition machinery performs site-specific integration of the new spacer into the CRISPR array at the leader end, concurrent with duplication of the first repeat.

It has been found that both the leader sequence and the first repeat are essential for this process. Studies of the *Escherichia coli* type I-E and the *S. thermophilus* type II-A systems revealed that the leader-repeat boundary serves as an anchor for spacer integration.

(Detailed discussion later)

Sternberg et al.,(2016) Molecular Cell 61: 797

M7L1 25

Type I and type II systems use short 3 to 7 base pair protospacer adjacent motifs or PAMs for identification of suitable protospacer for acquisition. During naive adaptation by type II systems, PAM is recognized by Cas9 while in type I-E systems, Cas1 and Cas2 are sufficient for recognition of PAM.

Following the protospacer selection, the acquisition machinery performs site specific integration of the new spacer into the CRISPR array at the leader end, concurrent with duplication of the first repeat. It has been shown that both the leader sequence and the first repeat are essential for this process. Studies of the E. coli type I-E and S thermophilus type II-A systems revealed that the leader repeat boundary serves as an anchor for spacer integration. We will be discussing this soon.

(Refer Slide Time: 35:49)

New Spacer Acquisition in Type II-A CRISPR-Cas Immunity
Viral DNA
Cas1, Cas2, Cas9, Csn2
CRISPR array
New Spacer Acquisition
Cas1-Cas2-Csn2-DNA Spacer Capture Complex
4x Cas1 Dimers
2x Cas2 Dimers
2x Csn2 Tetramers
~30 bp of dsDNA protected within the complex

Figure from Wilkinson et al., 2019, Molecular Cell 75, 90. CC BY 4.0

In addition to Cas1, Cas2 and Cas9, type II-A CRISPR–Cas systems also include Csn2, which is considered a subtype-specific signature Cas protein. Csn2 has been implicated in the adaptation stage of CRISPR-mediated immunity since it was shown to be required for the acquisition of new spacers during in vivo adaptation experiments.#

When S. pyogenes type II-A Cas operon was expressed in Escherichia coli, four Cas proteins were co-purified*, suggesting the formation of a multi-protein Cas complex.

#Nucleic Acids Research. 2018; 46:9805–9815,
*Nature. 2015; 519:199–202

M7L1 26

Now, let us look into the arrangement of the various components in the CRISPR-Cas immunity system, in type II-A CRISPR-Cas immunity system in particular. So, you have a CRISPR array over here and then this is the existing CRISPR array ok. So, this is basically the tau 1 state or yes and the initial state ok. Now, if the virus is attacked by sorry, if the bacteria is attacked by a new virus in the presence of Cas1, Cas2, Cas9, Csn2, a new spacer is acquired ok.

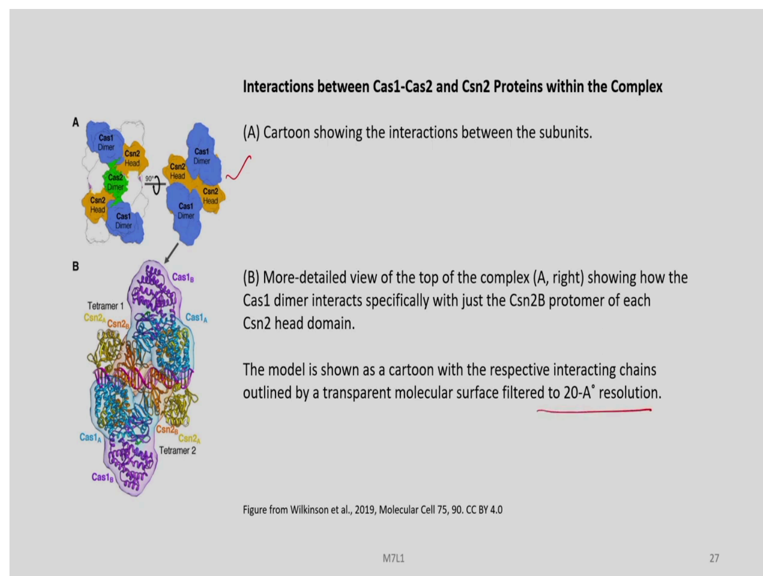
So, these viral DNA having this particular colour code is added here; so, this is the tau 3 state. So, here you can see this up to the yellow, up to this point it is same. So, new one is added over here. So, this is just schematics, it does not happen that straight away. We will discuss how it exactly happens. This is just to make the simple conceptual presentation that the viral DNA is added to the existing CRISPR array with the help of the Cas1, Cas2, Cas9 and Crn, Csn2 proteins.

And, these Cas1, Cas2, Csn2 DNA spacer capture complex. So, these complex will, leaving aside the Cas9, these complex will bind to these viral DNA. So, you can see here, this is the viral DNA having the similar colour code. And, these are the 4; 1, 2, 3, 4 Cas1 dimers and then you have 2 Cas2 dimers and you have 2 Csn2 tetramers. So, this is a 90 degree rotation and a different view of these same structure, where these Cas1, Cas2, Csn2 forms a complex with the viral DNA to cleave it and to cut it from the viral DNA and paste it into the CRISPR array, existing CRISPR array.

So, you can see here roughly around 30 base pair of dsDNA. This is protected within this complex and this will be pasted into the existing CRISPR array. And, if you remember the work of Ishino and Mojica and they spoke about something around 35 base pair length spacers or repeats ok. So, in addition to the Cas1, Cas2, Cas9 type II-A CRISPR Cas system also include this Csn2 which is considered a subtype specific signature Cas protein.

Csn2 has been implicated in the adaptation stage of CRISPR mediated immunity since was shown to be required for the acquisition of new spacers during in vivo adaptation experiments. When S. pyogenes type II-A cast operation was expressed in E. coli, four Cas proteins were co-purified, suggesting the formation of a multi-protein Cas complexes. So, these are the experimental evidences which led to the you know suggestion of these kind of structures occurring.

(Refer Slide Time: 39:46)

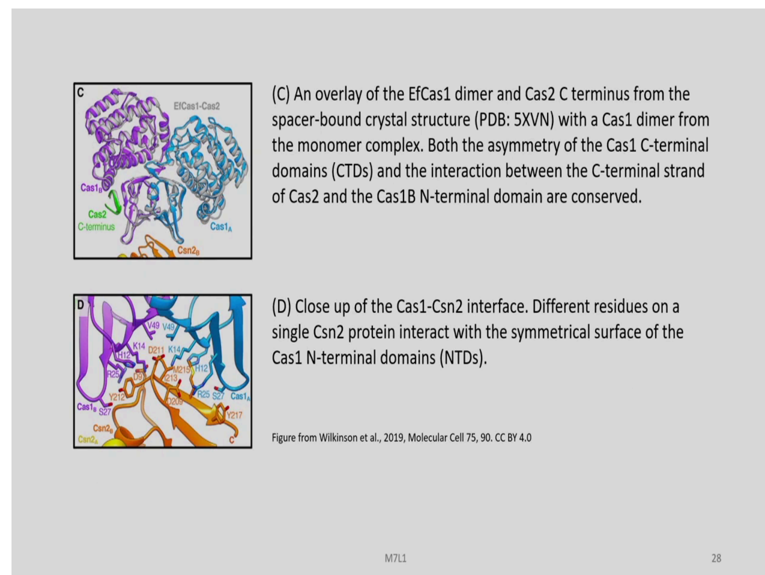


How does the interaction between Cas1, Cas2 and Csn2 proteins occur exactly within these complex. So, we spoke about these 4 Cas1, 2×Cas2, 2×Csn2 and you can see here this is a Cas1 dimer here, this is a Cas again here Csn2 head and this is a Cas2 dimer over here. And, this is as the 90 degree rotation to show the underside of these particular complex. And these are in more details, the structure of these particular complex.

So, this is a more detailed view of the top of the complex here right. So, in the custom dimer interacts specifically with just a Csn2B promoter of each Csn2 head domain. This model is

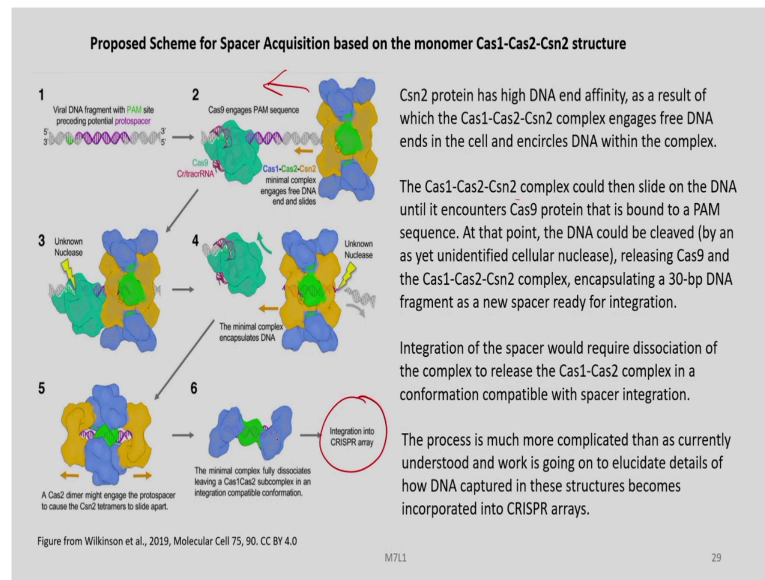
shown as a cartoon in the repetitive interacting chains outlined by a transparent molecular surface filtered to 20 Armstrong resolution.

(Refer Slide Time: 40:51)



Now, this is an overlay of the EfCas1 dimer and Cas2 terminus from the spacer bound crystal structure with a Cas1 dimer from the monomer complex. Both the asymmetry of the Cas1 C-terminal domains and the interacting interaction between the C-terminal strand of Cas2 and the Cas1 Cas1B and terminal domains are conserved over here. This is a closeup of the Cas1-Csn2 interface. Different residues on a single Csn2 protein interact with the symmetrical surfaces of the Cas1 N-terminal domains.

(Refer Slide Time: 41:33)



Now, this is a purpose scheme for spacer acquisitions. Now, we know that these Cas1-Cas2-Csn complex binds to the viral DNA. But, how do these acquisition really takes place? So, this is a proposed scheme as presented over here. So, you have this viral DNA fragment which has a PAM site ok, which precedes the potential protospacer.

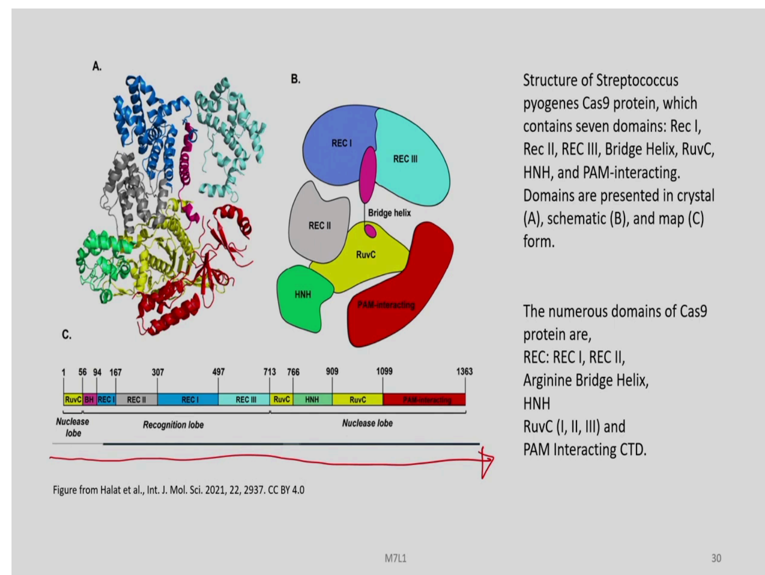
Now, these Cas1-Cas2-Csn complex engages on the free DNA and sites and while the Cas9 tracrRNA would bind to the PAM site. So, Csn2 protein has high DNA end affinity, it binds to DNA ends. As a result of which this Cas1-Cas2-Csn complex would engage free DNA ends in the cells and encircle DNA within the complex.

These Cas-Csn complex would then slide on the DNA, upon binding it goes towards the Cas9 complex. And, it will keep on moving until it encounters a Cas9 protein that is bound to a PAM sequence. At that point, DNA could be cleaved by as it identified cellular nucleus releasing Cas9 and Cas1-Cas2-Csn2 complex, encapsulating a 30 base pair DNA fragment as you can see in the figure and as a new spacer ready for integration. So, this cleaving is or picking up of the viral DNA is complete at this stage.

Integration of the spacer would require dissociation of the complex to release the Cas1-Cas2 complex in the conformation compatible with spacer integration. The unknown, the process is much more complicated than as currently understood and work is going on to elucidate the details of how DNA is captured in these structures and how it is incorporated into the this

CRISPR arrays. A lot of details already known, but certain informations are missing. So, this it is still remains many of these things still remains unknown in certain aspects.

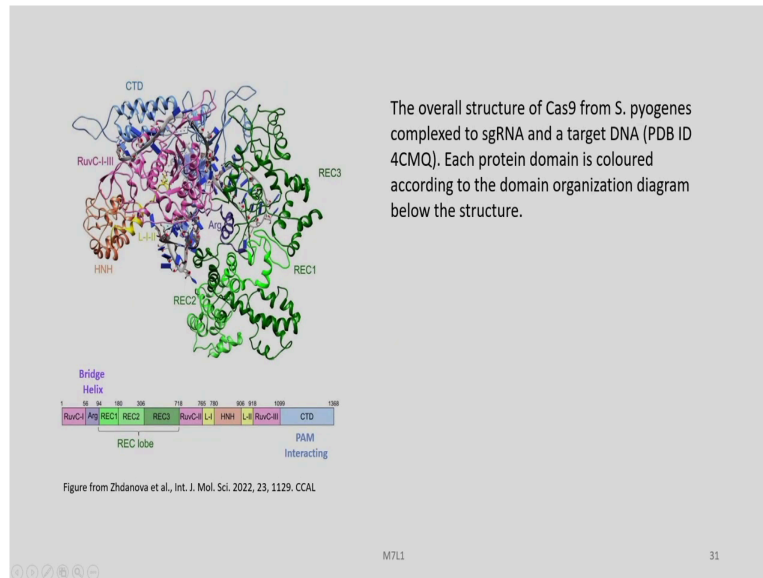
(Refer Slide Time: 44:11)



So, let us look into the structure of the Cas9 protein. For example, we are looking into the structure of the Streptococcus pyrogenes and these contains about seven domains with some basically are subdomains. So, you have these RecI, RecII and RecIII, then you have a bridge over here is a bridge helix, then you have RuvC and RuvC may actually be distributed into 3 Ruv 1, 2 and 3 ok.

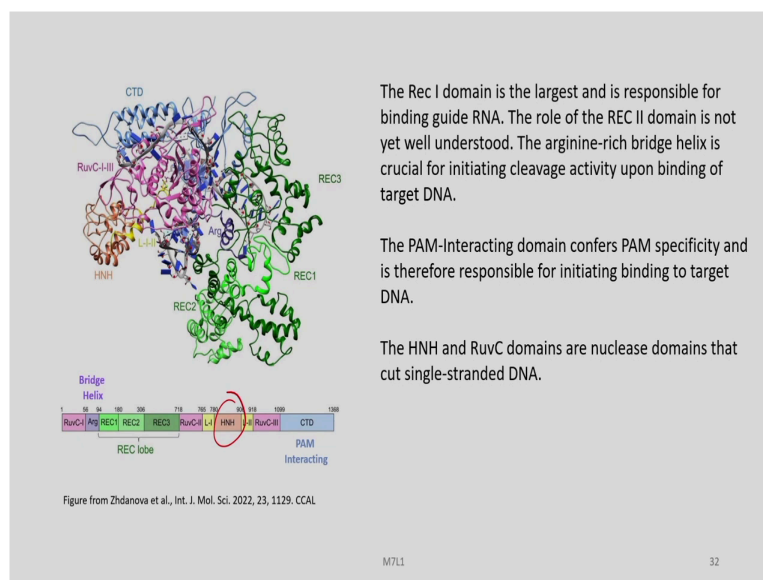
Then, there is a PAM interacting domain and we know about this PAM sequence now and these are overall the domains of Cas9 protein which we have already discussed. And, in three-dimensional space they are arranged in these spatial distribution. And, this is the linear genetic gene sequence arrangement. But, when it is expressed into a protein, it will fall in three-dimensional space. And, then you find the various domains in three-dimensional space in some structure simplified like this ok.

(Refer Slide Time: 45:34)



Now, let us look into the structure of a Cas9 which we have seen here, when it is complex to a single guide RNA ok and a target DNA. Each protein domain is coloured here according to the domain organization diagram below these structures.

(Refer Slide Time: 46:02)



The Rec I domain is the largest and it is responsible for binding the guide RNA. The role of the REC II domain is not understood currently. The arginine-rich bridge helix is crucial for initiating cleavage activity upon binding of target DNA. The PAM interacting domain confers a PAM specificity and is therefore, responsible for initiating binding to target DNA. The

HNH and RuvC domains are nuclease domains that cut the single-stranded DNA. So, with these, we complete the understanding of the structure of these very important proteins.

Thank you, for your patient hearing. In the next lectures, we look into the various stages by which the CRISPR-Cas immunity system matures.

Thank you.