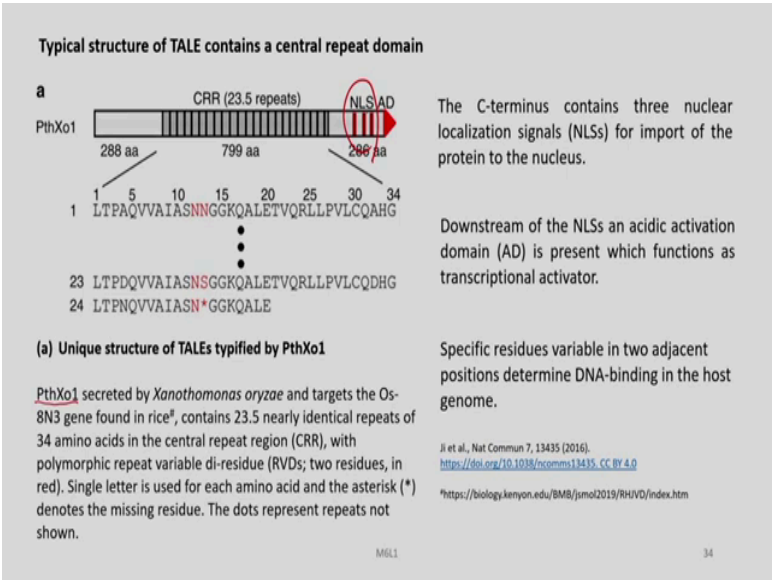
Genome Editing and Engineering
Prof. Utpal Bora
Department of Bioscience and Bioengineering
Indian Institute of Technology, Guwahati

Module - 06
Transcription activator-like effector nuclease (TALEN) Technology
Lecture - 01
Basics of TALEN-Part B

(Refer Slide Time: 00:35)

(Refer Slide Time: 00:40)



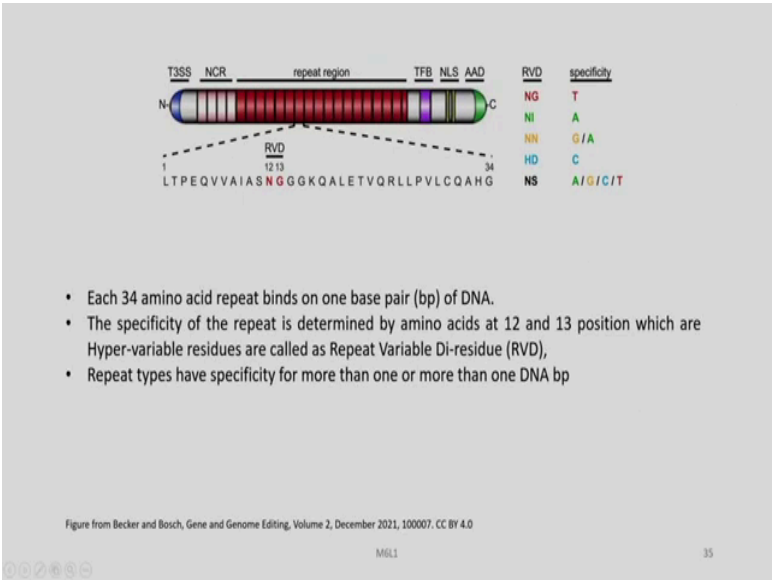
Welcome to part B of the lecture on the Basics of TALEN. So, here we will discuss about the structure of TALE effectors first and you can see here the typical structure of TALE effectors which contain a central repeat domain about which we have been speaking in the earlier part. So, this is the TALE of PthXo from *Xanothomonas* oryziae and this targets the Os8N3 gene which is found in rice and this contains around 23.5 nearly identical repeats of 34 amino acids in the central repeat region with polymorphic repeat variable di-residue.

So, here you can see the 24 members which is basically representing the 23.5 five repeats and in the centre, you can see the RVDs and except these everything else on the proteins are similar and the protein may be shorter than 34 here. But the sequence except the RVD will be similar in all cases. A single letter is used for each amino acid and the asterisks denotes the missing residue, the dots represent repeats which are not shown in other members.

A specific residue residual variable in the 2 adjacent positions determines the DNA binding in the host genome, so these will decide the DNA specificity. In these C-terminus there are three nuclear localization signals; these NLS and with these the protein is imported to the nucleus of the host cell. And downstream of this nuclear localization signal there is an acidic activation domain AD which function as transcriptional activator.

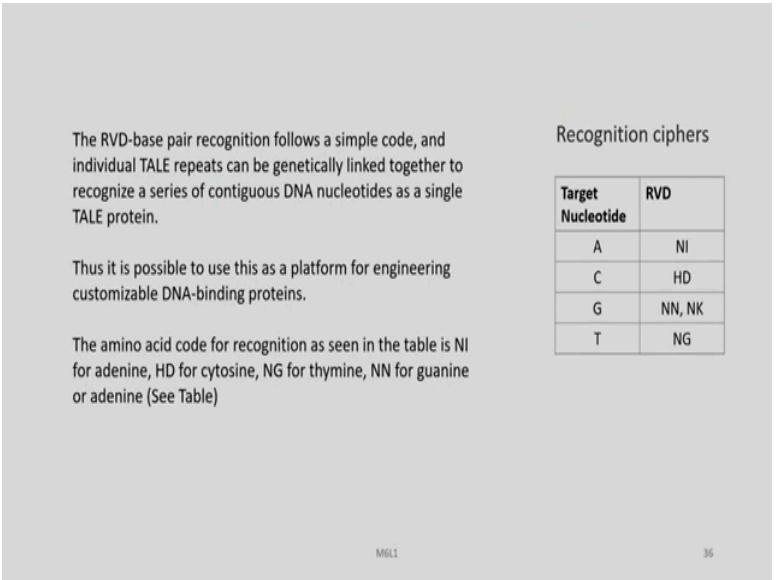
So, this is the typical structure we need to remember there is a central repeat region CRR, then there are these RVD in each of these repeats which are around 34 amino acids in land. And then there is a nuclear localization signal in the C terminal along with an acid activation domain.

(Refer Slide Time: 03:41)



So, each of these 34 amino acid repeats bind to one base pair of DNA, the specificity of these repeat is determined by the amino acids, its position number 12 and 13 which we refer to as the Repeat Variable Di-Residue RVD and the repeat types have specificity for one or more DNA basis. So, this is basically the kind of RVD against which specificity of the DNA base is listed, we will come to these again once more.

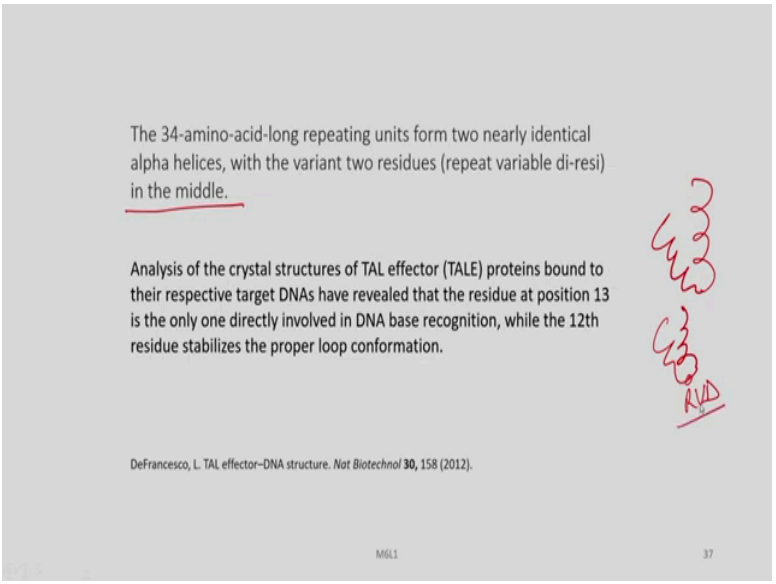
(Refer Slide Time: 04:23)



The RVD base pairs recognition follows a very simple code and the individual TALE repeats can be genetically linked together to recognize a series of contiguous DNA nucleotides as a

single TALE protein. Thus it is possible to use these as a platform for engineering customizable DNA binding proteins, the amino acid code for the recognition in the table which also we have discussed in the earlier slide is for A NI RVD for C HD, G NN NK and for, T NG.

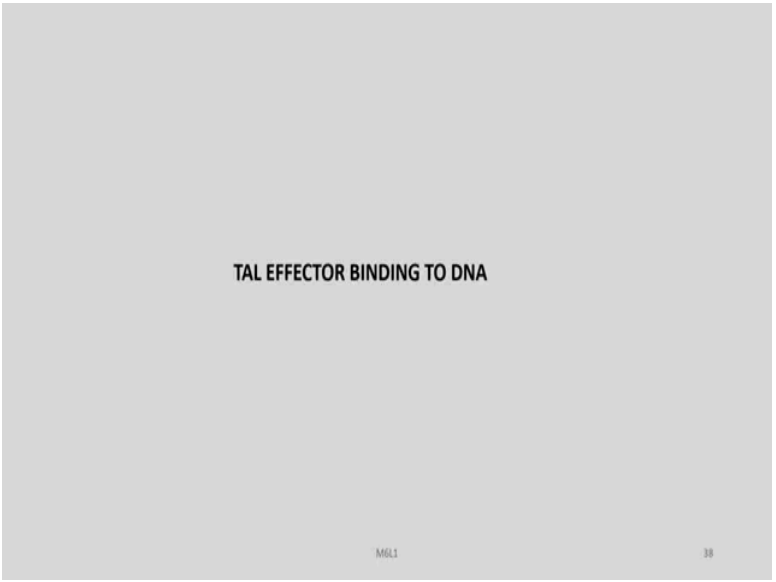
(Refer Slide Time: 05:09)



These 34 amino acid long repeating units form 2 nearly identical alpha helices we spoke about the alpha hairpin in one of the earlier slides. So, these alpha hairpins are coming from these 2 nearly identical alpha helices formed by the 34 amino acid long repeat units and the two variant residues, 12 and 13, lie in the middle. So, if you have one alpha helix like this; or so here it will be the RVD.

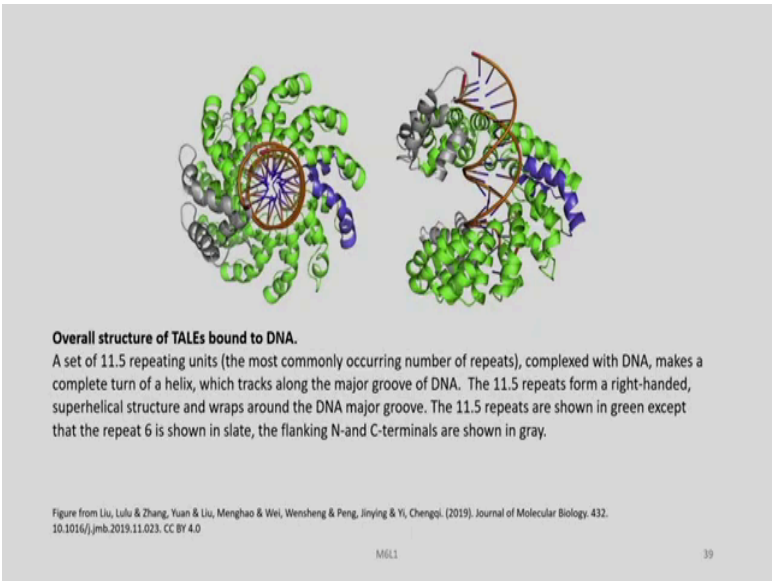
Analysis of the crystal structure of TAL effector proteins bound to their respective target DNA have revealed that the residual position 13 is the only one directly involved in DNA based recognition while, the 12th residue stabilizes the proper loop conformation. So, there are variables at both the positions, but 13 is thought to be directly involved in the DNA specificity and the 12th amino acid plays more of a structure stabilizing role.

(Refer Slide Time: 06:30)



So, how do these TAL effectors bind to DNA, we now know more or less that there are these 34 amino acids folded into 2 hairpin loops and they form a hairpin and in the centre the RVDs are located.

(Refer Slide Time: 06:56)

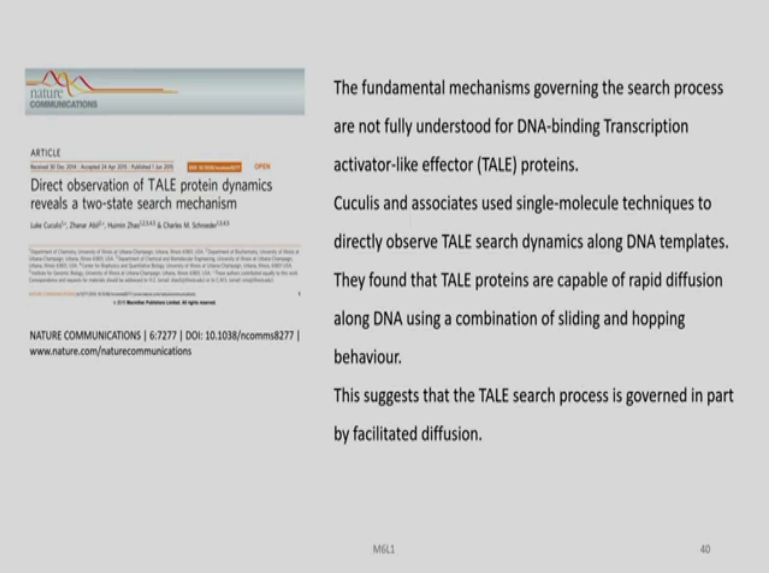


And you can see the cross sectional view of the overall structure of TALEs bound to DNA in the first picture and the cross sectional view in the right side. So, a set of around 11.5 repeating units, the most commonly occurring number of repeats complexed with DNA in this particular case, makes a complete turn of the helix which tracks along the major groove

of the DNA. The 11.5 repeats form a right handed super helical structure and wraps around the DNA major groove.

The 11.5 repeats are shown in green except that the repeat 6 is shown in slate the flanking N and C terminals are shown in gray. So, depending on the TALE these numbers may vary this is just to show the interaction of the TALEs with DNA. So, we know that they wrap around the DNA for the complexation and the track along the major groove of the DNA as can be seen in these pictures.

(Refer Slide Time: 08:11)



Now, we know how the TALEs bind to DNA. So, we have these RVDs concept where 13th amino acid is directly responsible, while the 12th provides some kind of structural ability. We also now know that the TALEs wrap around the DNA molecule. And from the cross sectional view we can see the interaction beautiful interaction and we know that it tracks around the major groove of the DNA.

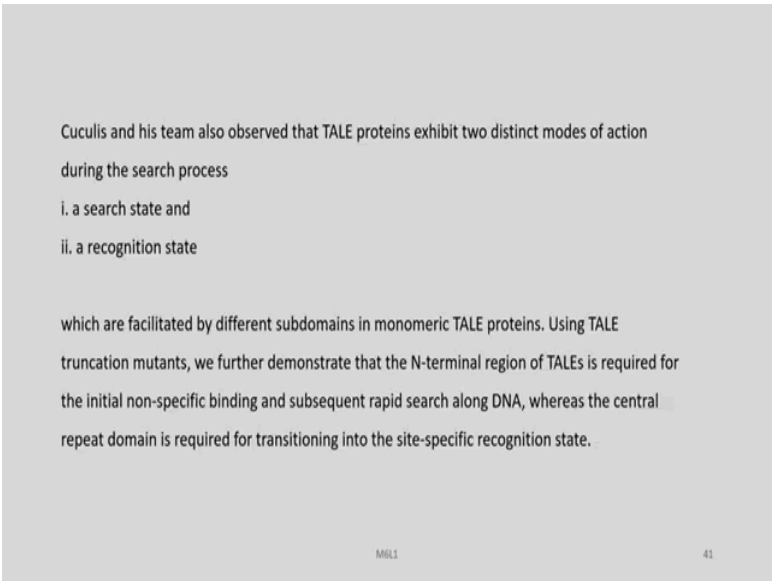
Let us try to learn a little bit of the mechanism by which the TALEs identify their cognate DNA sequences. So, each repeat unit has an RVD which has a base specificity and if a couple of such repeats are arranged they have different RVDs with different specificities and we know a combination of these different kinds of repeat units can give us specificity against very long DNA sequences.

But how does overall a TALE protein find out its cognitive partner or complementary DNA sequence? The fundamental mechanism governing the search process are not fully understood for DNA binding transcription activator like effectors and these paper by Cuculis proposes the TALE protein dynamics which reveals a 2 state search mechanism.

So, they used single molecule techniques to directly observe the TALE search dynamics if you are interested you can go to the experimental part of this paper through which they found out some of these interesting effects. And they found that TALE proteins are capable of rapid diffusion along DNA using a combination of sliding and hopping behaviour.

So, we have similar kind of phenomena in ZFN, but this is a little bit different here we call it as "sliding and hopping" behaviour. And this sliding and hopping behaviour is very much important for these search processes and the 2 state search mechanisms. What are these 2 state search mechanisms, we will understand soon in the next slide and later on. So, overall we need to remember that TALE search process is governed in part by facilitated diffusion.

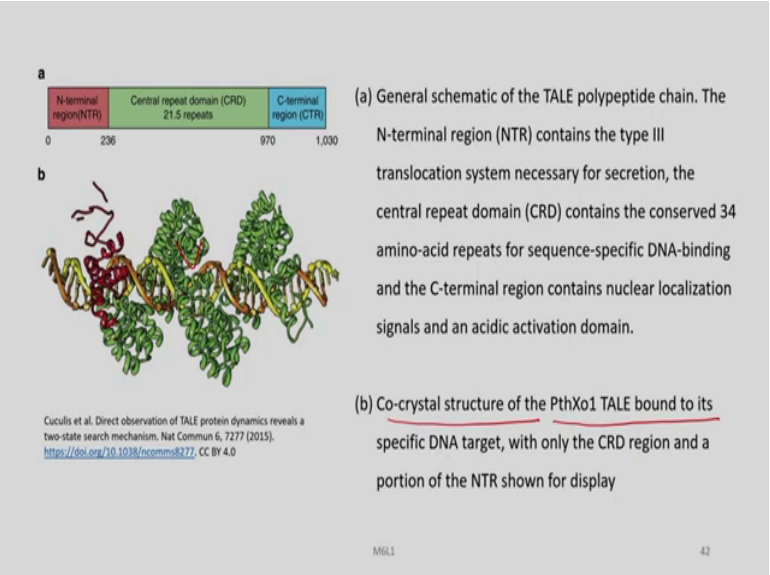
(Refer Slide Time: 11:17)



So, Cuculis and his team observed that TALE proteins exhibit 2 distinct modes of action during the search process: a search state and the second one is a recognition state. And these are facilitated by different subdomains in monomeric TALE proteins, using TALE truncation mutants they further demonstrated that the N terminal region of TALEs is required for the initial nonspecific binding and subsequent rapid search along DNA. Whereas, the central

repeat domain is required for transitioning into site specific recognition states. So, we have seen in details the typical structure of a TALE protein.

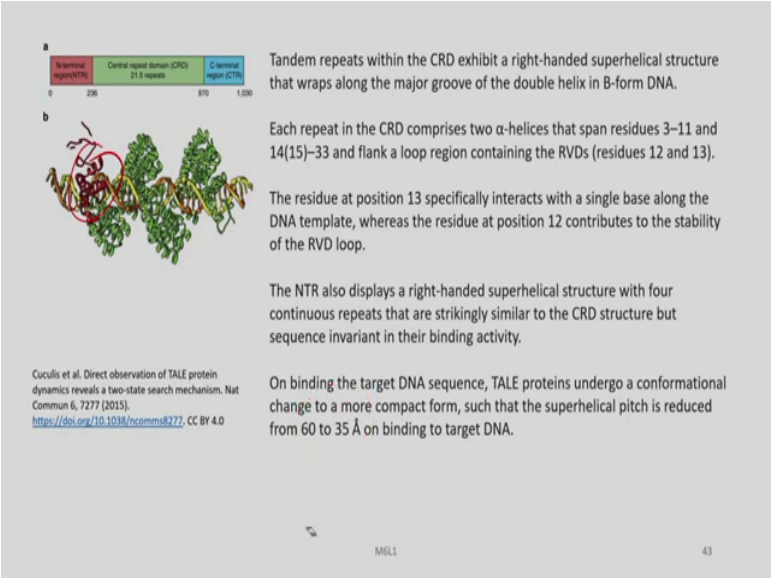
(Refer Slide Time: 12:11)



Now, let us see in a very simplified way, here is a TALE protein with N terminal region or NTR and then there is a central repeat domain CRD. In this case we are taking around 21.5 repeats or 22 repeats and there is a C terminal region it is a protein with N terminal and a C terminal and a central domain which is as simple as these. And you can see here the corresponding colors this is the N terminal and this is the C terminal and in this picture. Sorry, this is the central domain up to here and the C terminal is not shown here for the sake of simplicity.

So, this is the general schematic of the TALE polypeptide chain and the various domains as I have discussed. And these CRD contains the 34 amino acid repeats for sequence-specific DNA binding and the C terminal contains the nuclear localization factor and the acid activation domain, which we have studied earlier. So, they did some crystallization experiment from where they find out that the co-crystal structure of PthXo1 TALE bound to its specific DNA target, with only the CRD region and a portion of the NTR shown for display here.

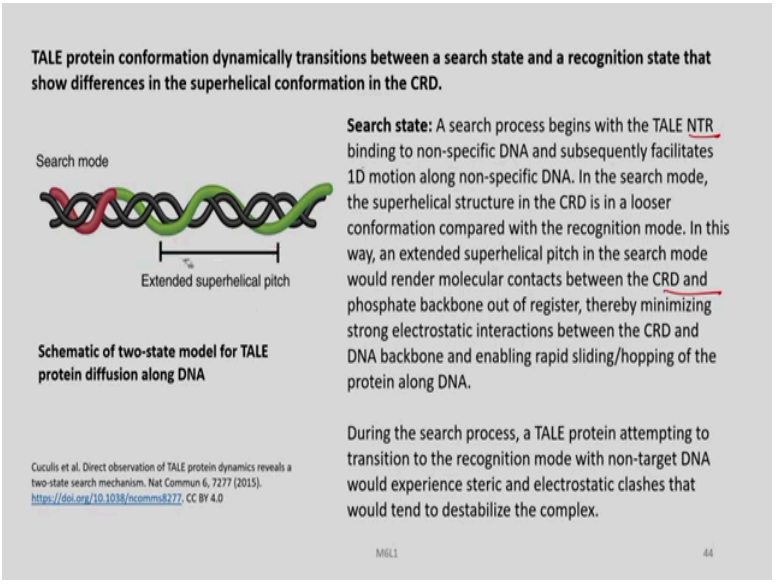
(Refer Slide Time: 13:45)



So, tandem repeats within the CRD exhibit a right-handed super helical structure that wraps along the major groove of the double helix in B-form DNA. Each repeat in the CRD comprises 2 alpha helices that span residues 3 to 11 and 14 (15) to 33, and flank a loop region containing the RVDs in the centre (positions 12 and 13).

The residue at position 13 specifically interacts with a single base along the DNA template, whereas the residue at position 12 contributes to the stability of the RVD loop which we have emphasized again and again. The NTR also displays a right-handed super helical structure with 4 continuous repeats that are strikingly similar to the CRD structure, but sequence invariant in their binding activity. This is very, very important these 4 continuous repeats similar to the CR structure and they bind to DNA, but they are sequence invariant; they are not specific and therefore they can initiate the binding easily. On binding the target DNA sequence the TALE proteins undergo a conformational change to a more compact form, such that the super helical pitch is reduced from 60 to 35 Armstrong's on binding to target DNA which is roughly 50 percent, OK, little bit more than that 50 say 5 percent, roughly around, ok.

(Refer Slide Time: 15:27)



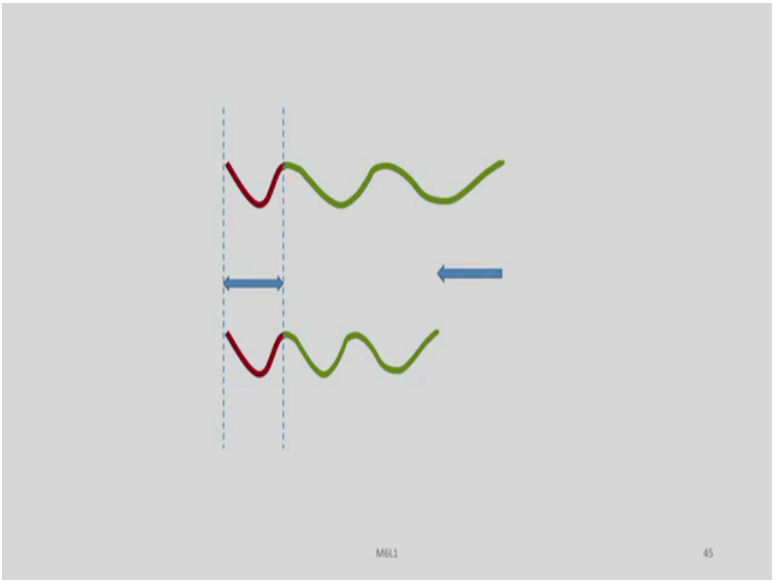
So, this is the TALE protein conformation you can see here: this is the N domain which has these 4 continuous repeats that are strikingly similar to the CRD structure, but sequence invariant in their binding. So, here in the search state or search mode, a search process begins with the TALE NTR binding to nonspecific DNA, because their sequence invariant and subsequently facilitates one dimensional motion along nonspecific DNA it will bind and just move around in the search mode.

The super helical structure in the CRD is in a looser conformation compared with the recognition mode. The N terminal NTR binds due to nonspecific interactions. And the CRD we will also bind, try to bind and because it requires specific basEs first it will bind loosely to the DNA. And this is kind of an engagement mode or search mode where the TALE proteins try to find its cognate partner.

In this way an extended super helical pitch in the source mode would render molecular contacts to get its partner or identify the partner, that the TALE protein has to come and interact with DNA. So, the first level of interaction is nonspecific: they just come and bind to the DNA and only after binding it tries to look for it is partner. So, in this way an extended super helical pitch in the search mode would render molecular contacts between the CRD and phosphate backbone out of register, thereby minimizing strong electrostatic interactions between the CRD and DNA backbone and enabling rapid sliding/hopping of the protein along the DNA. During the search process a TALE protein attempting to transition to the recognition mode with non-target DNA would experience steric and electrostatic clashes that

would tend to destabilize the complex. So, once it binds it will try to look for its partner by sliding and even hopping.

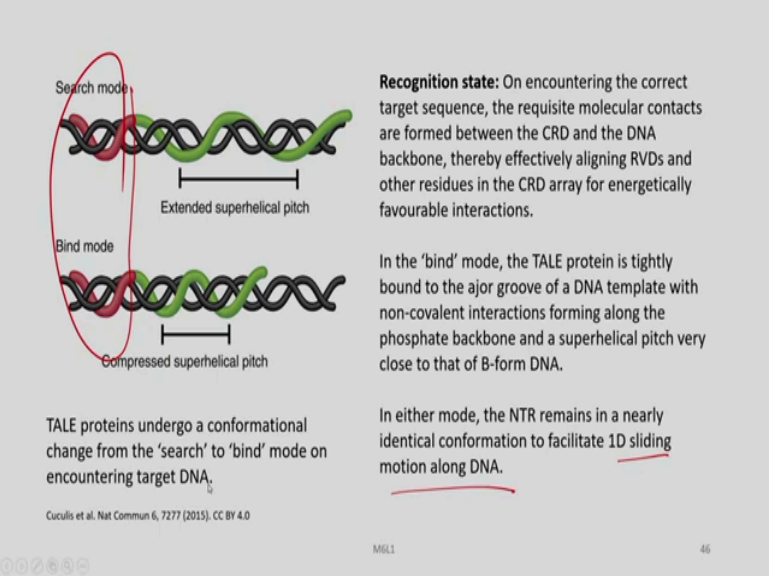
(Refer Slide Time: 18:00)



So, let us have a little bit of discussion on the shortening of these proteins which is having a NTR and a CRD. So, this is a protein as you can see representing the TALE protein and say this is the structure in which it binds to a DNA molecule in a nonspecific manner. And this NTR would bind to DNA in a nonspecific manner and it will force the CRD also to bind to DNA in a nonspecific manner.

Once it does, it goes for sliding and hopping and in that process this sequence may be able to find its cognate partner. And now let's say, if we take this as a spring and we try to put pressure on this green portion, we can shorten, compress this spring and this protein will be shortened. But you have to see that this particular red region remains unchanged, there is no compression on this particular zone, the compression is entirely on the green portion of the string. So, this concept is very important to understand the binding mode of the TALE proteins on the DNA.

(Refer Slide Time: 19:43)

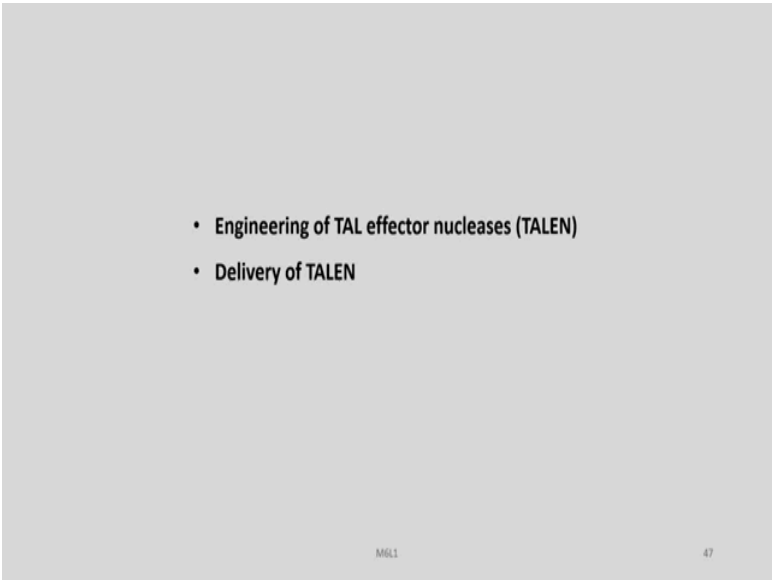


And we have already seen the search mode, where this extended super helical pitch is there because of the loose binding. But in a recognition state the shortening of this protein has taken place, if you can look into it and which is roughly around from 60 amino acids to 35, or roughly around 55 percent shortening.

So, this is the bind mode and in the recognition state, on encountering the correct target sequence, the requisite molecular contacts are formed between the CRD and the DNA backbone, thereby effectively aligning RVDs and other residues in the CRD array for energetically favourable interactions and this one remain unchanged; it is not having any specificity.

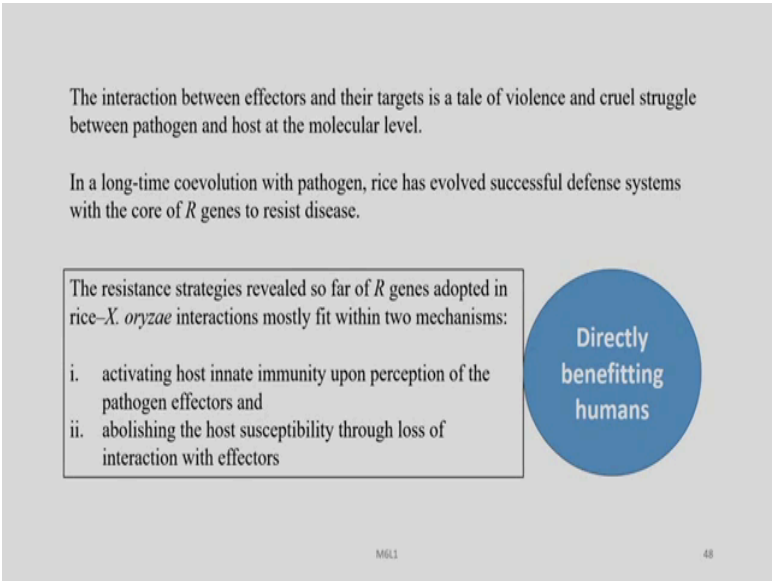
In the bind mode the TALE protein is tightly bound to the major groove of the DNA template with non-covalent interactions, forming along the phosphate backbone and a super helical pitch which is very close to that of the B-form DNA. In either mode the NTR remains in a nearly identical conformation to facilitate 1D sliding motion along the DNA and you can see the TALE proteins undergo a conformational change from the search to bind mode on encountering the target DNA.

(Refer Slide Time: 21:30)



So, this is how the TALEs finally find their complementary DNA sequence and bind to it in a very tight bonding. Let us now move to the next part of this lecture. We said about engineering of TAL effector nucleases or TALEN and once we engineer them or synthesize them how do we deliver them?

(Refer Slide Time: 22:07)

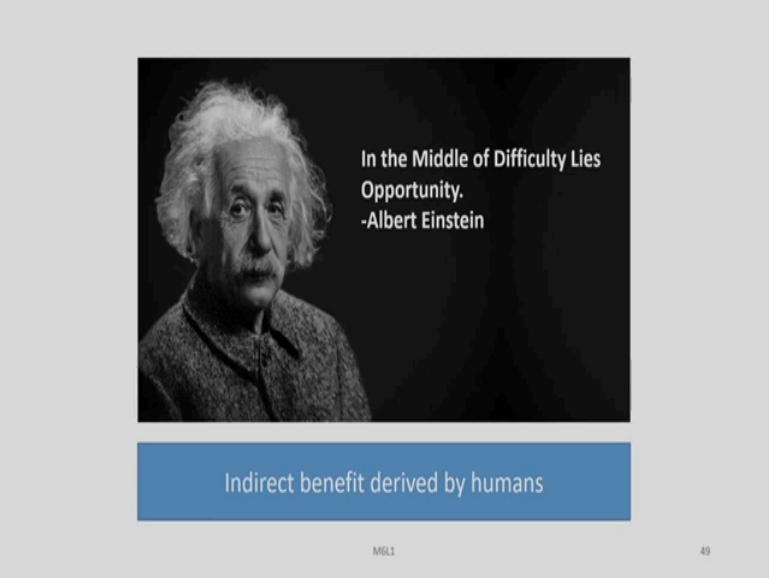


The interaction between effectors and their targets is a tale of violence and cruel struggle between pathogens and host at the molecular level. We know that the bacteria try to send in the TAL effectors through the T3 SS and which lends up finally in the nucleus, which

transcriptionally activates certain susceptible genes, thereby weakening or killing the host and the bacteria will finally draw the nutrients from the killed host or the weakened host.

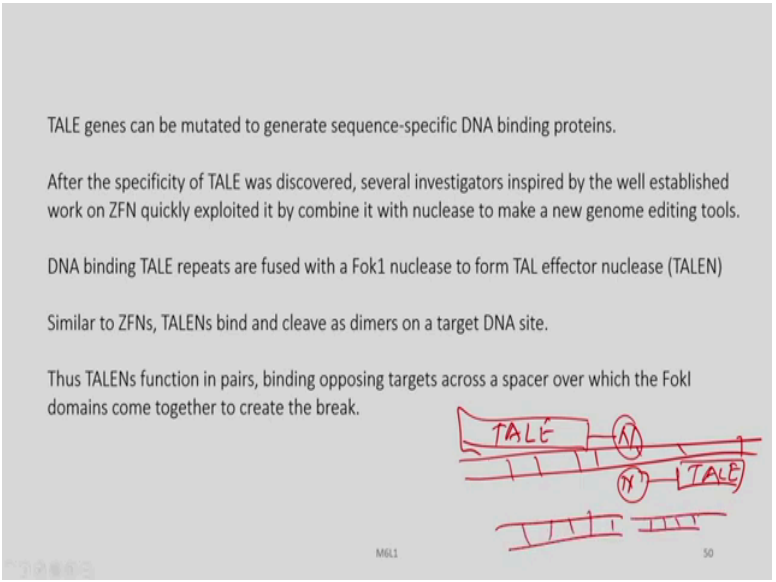
The plant cells also have evolved against these kind of infection in the evolutionary history. So, there has always been some kind of a war or battle which is ongoing and plant systems also have evolved a certain resistance mechanism, some of which we have studied briefly in the early part of this lecture. So, there is coevolution with the pathogen, and rice and other plants have successfully evolved various defence systems with the core of the R genes to resist. So, briefly these natural mechanisms directly benefit humans.

(Refer Slide Time: 23:43)



However, the pathogenicity remains in plant species which do not develop this kind of resistance mechanisms. However, in spite of all such difficulties and the negative states of pathogenic *Xanthomonas* are causing diseases in so many plants following the inspiring words of Einstein "in the middle of difficulty lies opportunity", we can use these basic biology concepts to build up genome editing tools as in the case of ZFNs which we have done studied earlier. And this is the indirect benefit humans can derive out of such a biological phenomenon.

(Refer Slide Time: 24:40)



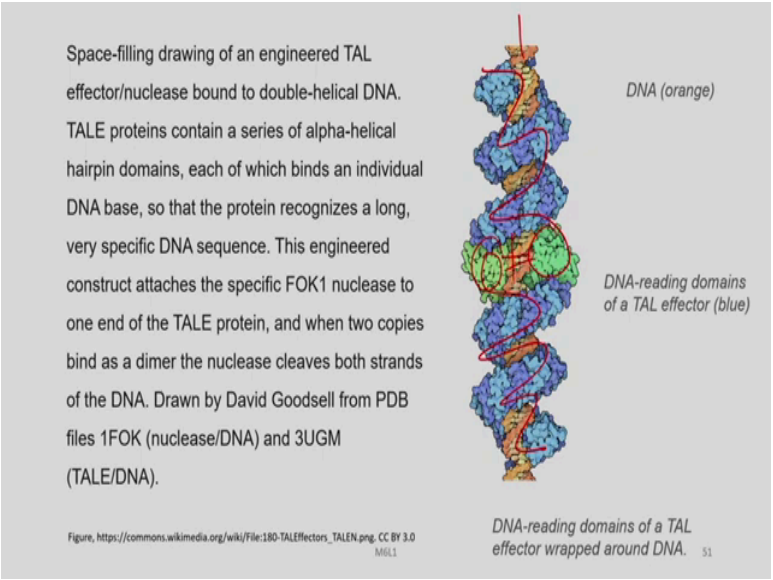
The TALE genes can be mutated to generate sequence specific DNA binding proteins, we just concluded a discussion on the DNA binding of TALEs. How they bind, how the RVDs are important, how the NTR binds and searches along with the CRD for the specific DNA sequence to which it has complementarity. So, now we can also design these TALE genes to bind to specific DNA sequences of our interest. So, given any gene on which we want to do some kind of engineering which may be knock in or knockout or a point mutation or where in the genomic location we want to insert certain genes, we can use the TALE genes to help us in the same way as the ZFNs have been used, which we discussed in the earlier module. So, after the specificity of TALE was discovered several investigators inspired by the well-established work on ZFN, quickly exploited it by combining it with a nuclease to make a new genome editing tool. So, here nothing is much different only the DNA binding part is different.

In the earlier case we are using zinc fingers now here we are using TAL effectors and these TALE effectors also will be engineered and designed synthetically. So, DNA binding TALE repeats are fused with Fok1 nuclease to form TAL effector nucleases. So, this is the TALE and this is the N nuclease which gives us the beautiful genome editing tool called as TALEN and similar to ZFNs they cleave as dimers on a target DNA site.

The dimerization is required not by the TALE domains, the dimerization is always required by the Fok domains. So, we know about the mechanism of dimerization from the last lecture,

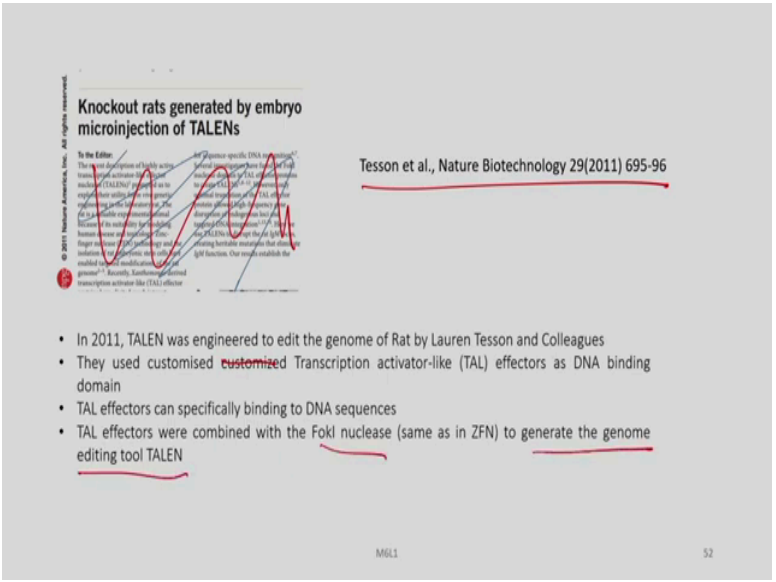
and we are not going to discuss here those topics again. And once they dimerize they will cleave the DNA and produce double-strand breaks. And once the double-strand breaks are produced the natural repair system of NHEJ and homologous recombination will take place and a similar type of genome engineering can be done as in the case of ZFNs.

(Refer Slide Time: 27:48)



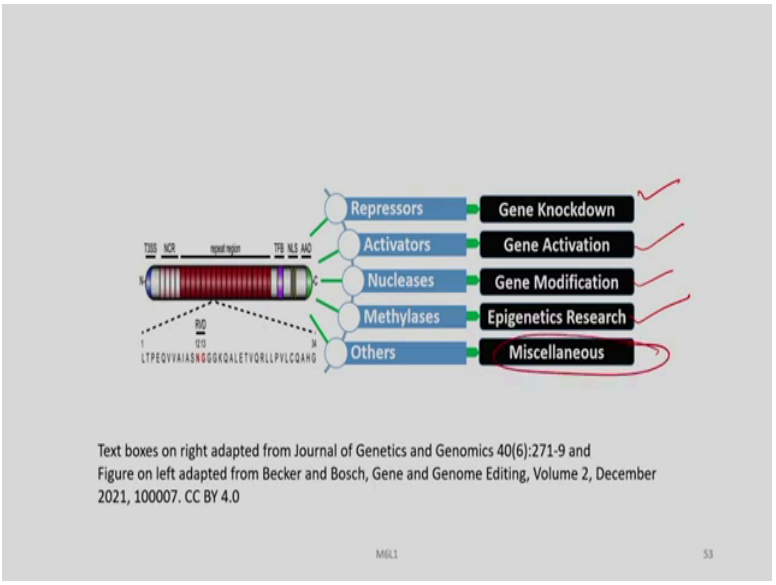
So, here is a space filling model of an engineered TAL effector nuclease. This is, these green ones are the nuclease portion; the DNA is the orange portion and the blue one are the TALEs and you have TALEs binding to both the strands. So, this is one TALE TALEN and this is the other TALEN and they are dimerizing and cleaving.

(Refer Slide Time: 28:27)



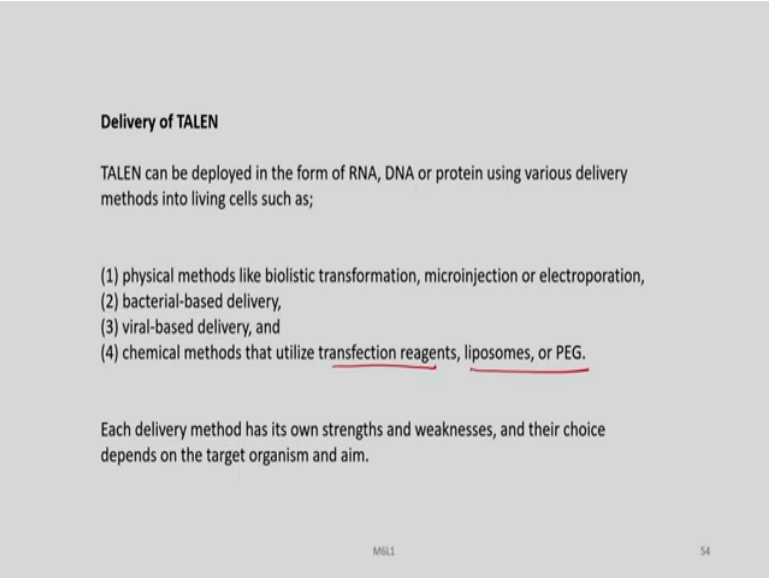
So, this is one important paper in 2011, so just a little time after TALEN was developed, the knockout rats were created by embryo injection of TALENs by Tesson et al and his colleagues. They used customized transcription activator-like effectors as DNA binding domain, the TAL effectors can specifically bind to the DNA sequences which is now known to us and they were combined with the Fok1 nuclease to generate the genome editing tool TALEN.

(Refer Slide Time: 29:13)



Now, not only cutting or cleaving DNA molecules and inducing double-strand breaks, we can do many other kind of modifications or manipulations using the TALE domain, fused to different kind of molecules, which may be repressors, or which may be activators and which may be nucleases, methylases and others; they are respectively used for gene knockdown, gene activation, then gene modification which we have already discussed. Then one of the most emerging fields today is the epigenetics research and there are many others which we have to get a club as being miscellaneous.

(Refer Slide Time: 30:06)



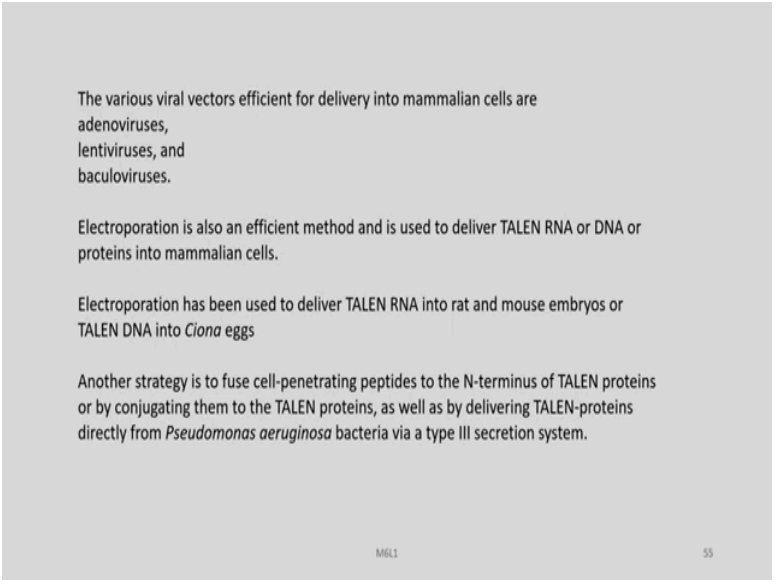
So, we can design TALEs and we can fuse them to many different kind of agents which carry out different genetic modifications. Now how do we deliver these TALENs inside a cell or inside a tissue or an organism? So, one good thing about TALEN is that we can deliver them in the form of mRNA. So, they will be transcribed inside the cell into TALENs, ok, or we can translate it into TALENs and then we can deliver them as DNA loaded into plasmids.

And they will go the full central dogma cycle: they will produce RNA and then proteins. Or we can also directly deliver the TALENs as proteins. We can synthesize them outside and directly deliver them inside and once they are inside the cell along with the nuclear localization signal they will be delivered to the nucleus and they will bind to the DNA molecules. And they will carry out the desired functions as per the type of functionality we have added to the TALE domain.

So, for delivery of these different type of TALE, TALEN constructs RNA, DNA or protein we have with us the typical delivery methods like physical methods using biolistic transformation, microinjection, or electroporation. And there are also bacteria bacterial-based delivery systems and viral-based delivery systems. As well as chemical methods that utilizes the transfection reagents liposomes or the PEG.

Each of these delivery methods have their own advantages and disadvantages, and the choice depends on the type of target organism and also the aim of the particular genome editing or genome manipulation that we intend to carry out.

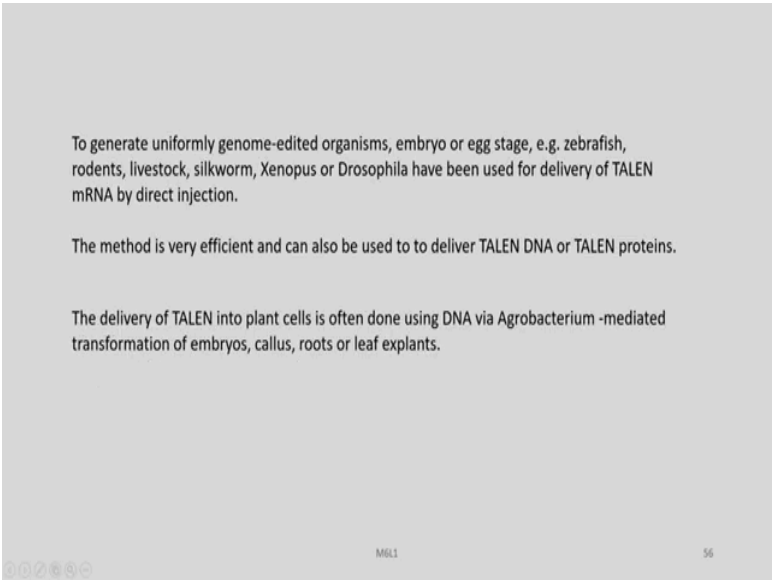
(Refer Slide Time: 32:34)



So, various viral vectors are available and they have various degrees of efficiency like adenoviruses, lentiviruses and baculoviruses. Electroporation is also a efficient method and this is used to deliver TALEN RNA or DNA and even proteins into the mammalian cells. And there have been certain applications of using electroporation to deliver TALEN RNA into rat and mouse embryos or TALEN DNA into *Ciona* eggs.

Another strategy is to fuse cell-penetrating peptides to the N terminus of the TALEN proteins or by conjugating them to the TALEN proteins as well as by delivering TALEN proteins directly from the *Pseudomonas aeruginosa* bacteria via a type 3 secretion system. And we know naturally *Xanthomonas* uses this type three secretion system to deliver the TALE effectors directly to the plants. So, if it is related to delivering into plant system we can use these natural pathway or doorway for delivering the TALEN constructs.

(Refer Slide Time: 33:54)



To generate uniformly genome edited organisms, embryo or egg stage is preferred. For example, in Zebra fish, rodents, livestock, silkworm, *Xenopus* or *Drosophila*, it has been used for delivery of TALEN by direct injection. And this is a very efficient method and can also be used to deliver TALEN DNA or the TALEN protein. The delivery of TALEN into plant cell is often done using the conventional *Agrobacterium*-mediated transformation of embryos, callus, roots or leaf explants, where the TALEN is sent as a DNA construct.

So, with this, we come to the end of this lecture, where we tried to discuss about the basics of the TALEs and how fusing these TALEs to nucleases and other proteins can be used to carry out various genetic modifications.

Thank you.