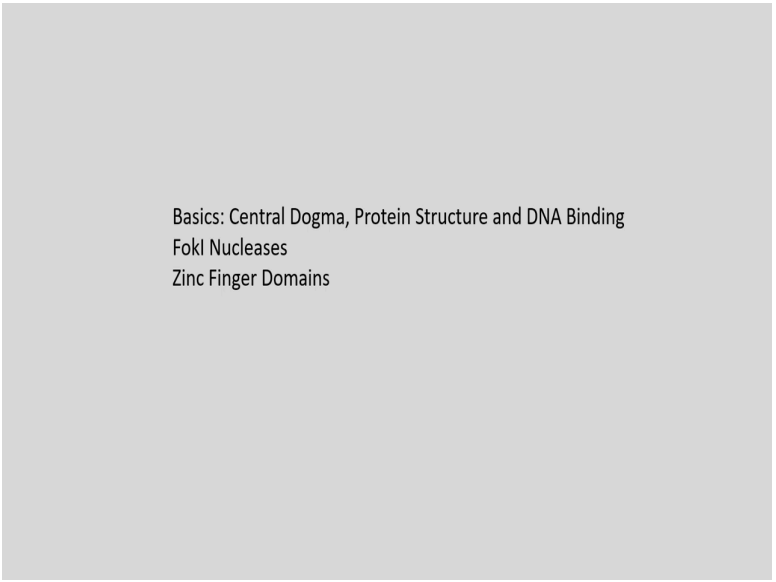
Genome Editing and Engineering
Prof. Utpal Bora
Department of Bioscience and Bioengineering
Indian Institute of Technology, Guwahati

Module - 05
Zinc Finger Nuclease (ZFN) Technology
Lecture - 10
Basics of Zinc Finger Nucleases

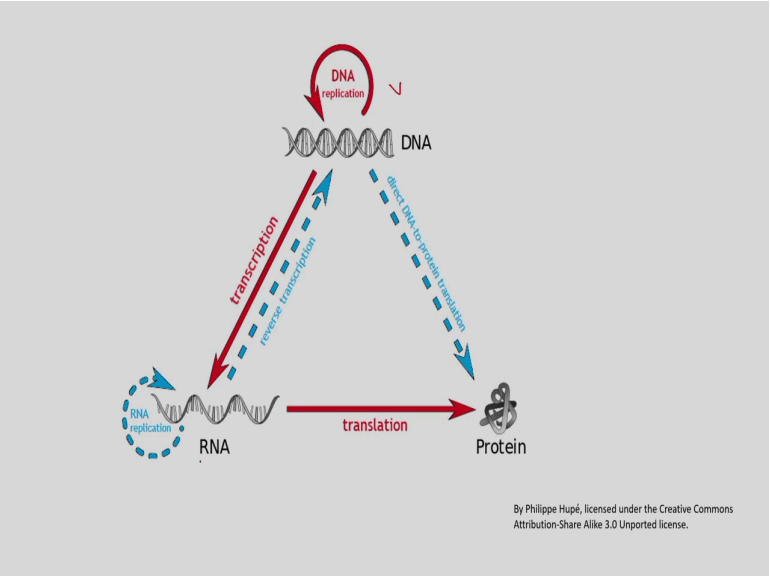
Welcome to my course on Genome Editing and Engineering, we are going to discuss about basics of Zinc Finger Nucleases today.

(Refer Slide Time: 00:40)



So, we will be first learning about some of the basic things which are important for this lecture. We will learn about the central dogma, the protein structures and how proteins bind to DNA. Then we will be discussing about one nucleus called fok1 nucleus and certain special motifs in proteins called zinc finger domains.

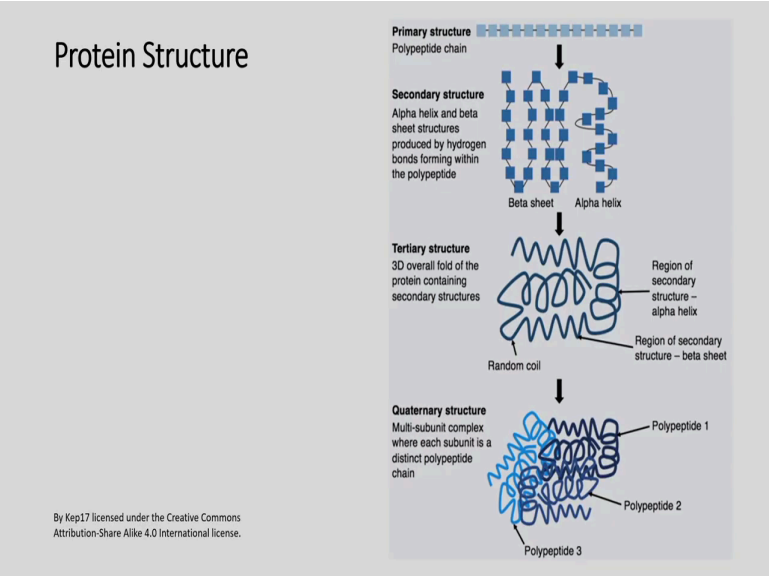
(Refer Slide Time: 01:05)



So, this figure may be familiar to many of you which depicts the central dogma of molecular biology or central dogma of life, where DNA is the information molecule which is copied by RNA, you can see in these arm by the method of transcription. And this RNA is translated into proteins, also the DNA replicates on its own or make copies of itself by the process of DNA replication.

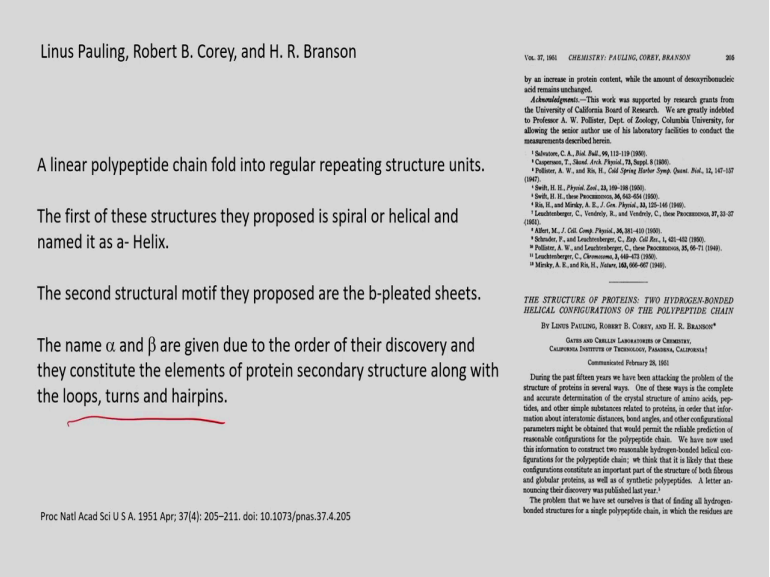
And there are also methods by which RNA may be copied into RNA, but we are not going to discuss much about those molecular mechanisms. Today we are going to use some of the concepts of protein structure important for this lecture. You know about the primary structure of proteins which are basically linear polymers of amino acids. And the secondary structure elements are there, known as beta sheets and alpha helix.

(Refer Slide Time: 02:15)



And then the tertiary structure which is basically the overall 3D folding of proteins or the arrangement of the secondary structures in space and that makes a protein functional or active. And then you have the quaternary structure which is basically the arrangement of the tertiary structures with respect to one another in three-dimensional space.

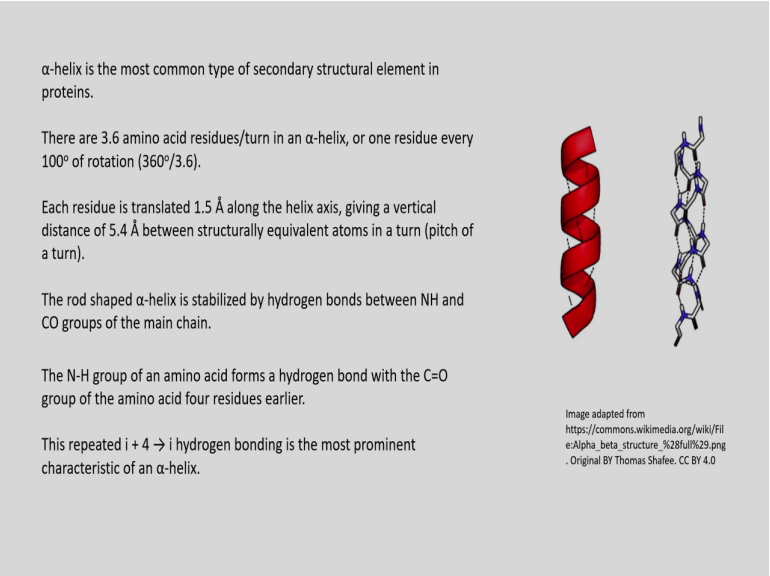
(Refer Slide Time: 02:44)



This is one of the landmark papers in science by Linus Pauling, Robert Corey and Branson where they described some of the important elements of protein structures. They came to a conclusion that linear polypeptide chains in the primary structure fall into regular repeating

structure units. And they named these repeating structure units as alpha helix which are basically a spiral arrangement of the amino acids in 3D space. And the second structure as beta pleated sheets, the name alpha and beta is due to the order of their discovery and they constitute the elements of protein secondary structure along with other elements like the loops, turns and the hairpins.

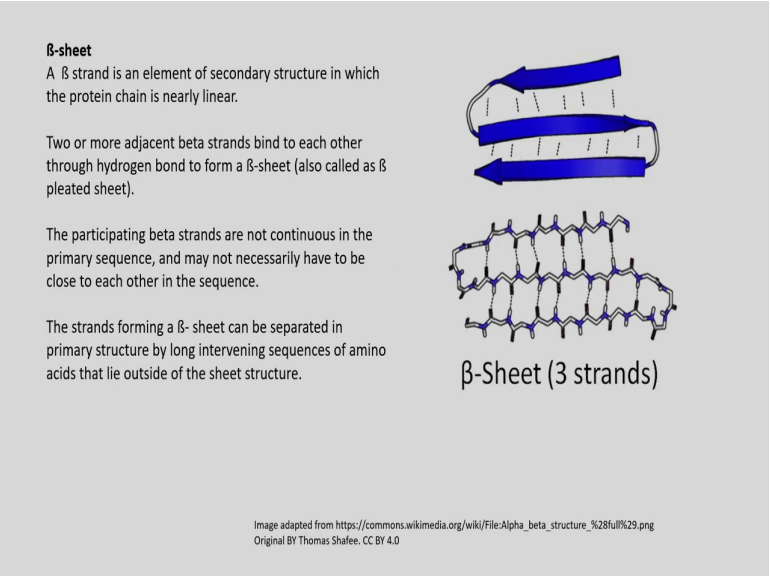
(Refer Slide Time: 03:39)



Let us focus a little bit about the alpha helix and this is very important from the point of view of studying zinc finger domains. Alpha helix is the most common type of secondary structural element in proteins. For every turn in these helix, there are around 3.6 amino acid residues or one residue per 100 degree rotation. These residues translates 1.5 angstrom along the helix axis and giving a vertical distance of around 5.4 angstrom between structurally equivalent atoms in a turn or the pitch is something around 0.5 nano meters.

As you can see this is a rod-shaped structure and this rod-shaped alpha helix structure is stabilized by hydrogen bonds between NH and CO groups of the main chain. The NH group of one amino acid forms a hydrogen bond with the CO group of the amino acid 4 residues earlier. So, there is a repeating pattern of i plus 4 to i hydrogen bonding and it is the prominent characteristics of an alpha helix.

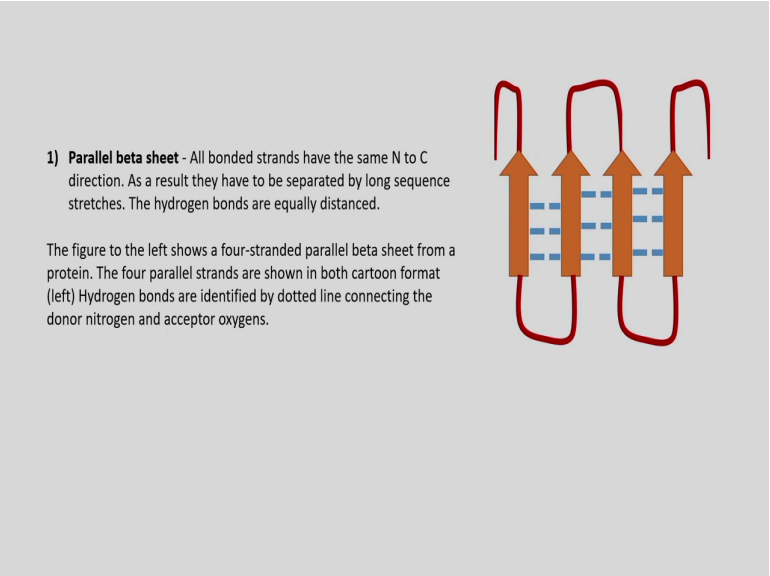
(Refer Slide Time: 04:54)



Next are the beta sheets or beta strands, a beta strand is an element of secondary structure in which the protein chain is nearly linear. Here, there is no any rod-shaped structure like the alpha helix. So, two or more adjacent beta strands will bind to each other through hydrogen bonding to form a beta sheet, these are also called as beta pleated sheets.

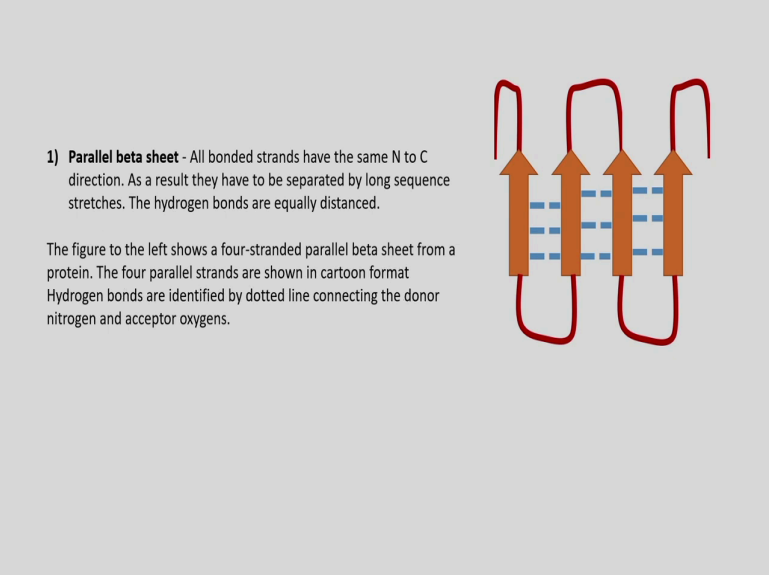
The participating beta strands are not continuous in the primary sequence and may not necessarily have to be close to one another in the sequence. The strands forming a beta sheet can be separated in primary structure by long intervening sequences of amino acids that lie outside of the sheet structure.

(Refer Slide Time: 05:42)



Now, these beta sheets may be arranged in different orientations, one of is the parallel beta sheet orientation, here all bonded strengths have the same N to C direction. As a result, they have to be separated by long sequence stretches, the hydrogen bonds are equally distanced. The figure to the left, you can see the four stranded parallel beta sheet from a protein.

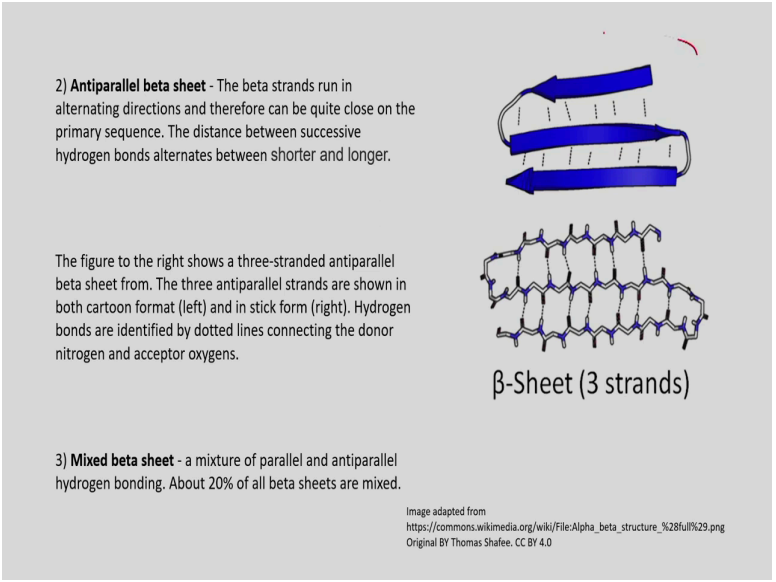
(Refer Slide Time: 06:10)



So, let us discuss some of the various orientations of the beta pleated sheets, the first one being the parallel beta sheets. All bounded strands have the same N to C direction as a result they have to be separated by long sequence stretches, the hydrogen bonds are also equally

distanced. In the figure in the left, you can see a four stranded beta parallel sheet from a protein. The four parallel strands are shown in cartoon format. The hydrogen bonds are identified by the dotted lines containing the donor nitrogen and the acceptor oxygens.

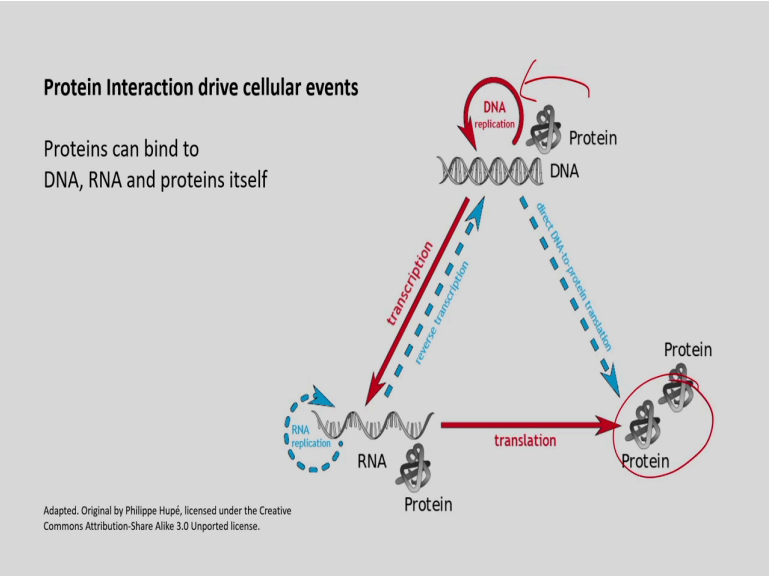
(Refer Slide Time: 06:45)



Then you have the antiparallel beta sheets and you can see these the direction of the beta strands are opposite to one another, but they are parallel. And here, the beta strands run in alternating directions and therefore, can be quite close on the primary sequence. The distance between successive hydrogen bond alternates between shorter and longer in this case.

And on the right, you can see the figure of a three stranded antiparallel beta sheet. The three antiparallel strands are shown in cartoon format and in the stick form. Hydrogen bonds are identified by the dotted lines connecting the donor nitrogen and acceptor oxygens. Then there is a third type of arrangement which we call as the mixed beta sheet, this is basically a mixture of parallel and antiparallel hydrogen bonding. About 20 percent of all beta sheets are found to be mixed beta sheets.

(Refer Slide Time: 07:45)

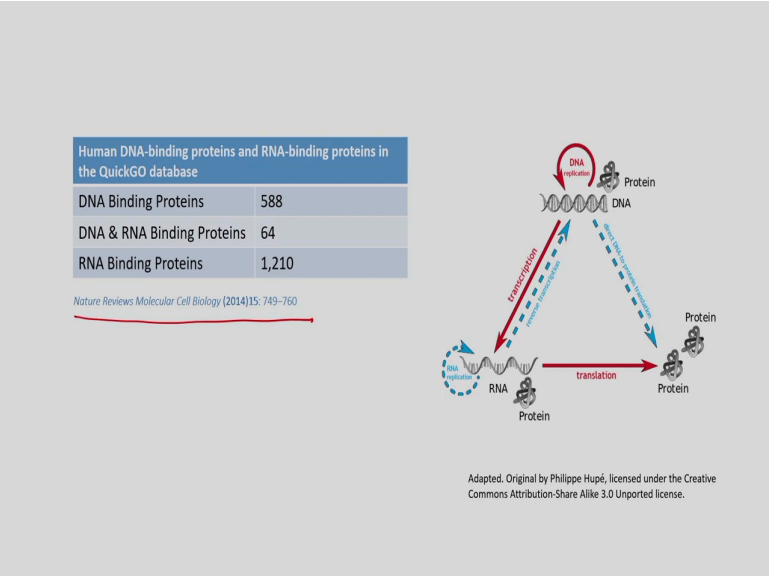


Now, we know about the central dogma of molecular biology. You also know about the various secondary structures of protein, as well as the tertiary structure and the quarterly structure in brief. Why do we need to know this? We have to know that a DNA is the informational molecule except storing and providing information, it cannot acts on its own. The work is being carried mostly by proteins and RNA; and RNA copies these information from DNA and translates them into the executors or the proteins.

And in certain cases, the RNA molecule is also a functional molecule. RNA may have structural role, RNA may have some kind of catalytic role. But the role which is being played by protein is highly diverse when we compare it with the roles being carried out by RNA. So, for doing all these work protein needs to talk to all other molecules inside the cell. Protein has to talk to DNA, it has talk to RNA.

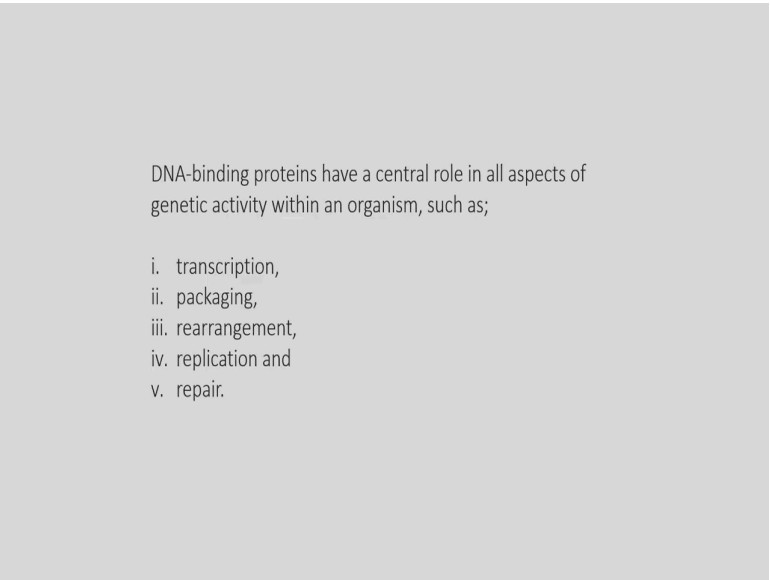
So, protein has to bind to DNA for whether replication to be carried out or transcription to be carried out or any other activity related to DNA. Similarly, proteins also bind to RNA, and then carries out translation and besides that, proteins also binds to proteins themselves or there are protein interactions. So, proteins are known as the players which drives cellular events by binding to DNA, RNA and proteins themselves.

(Refer Slide Time: 09:41)



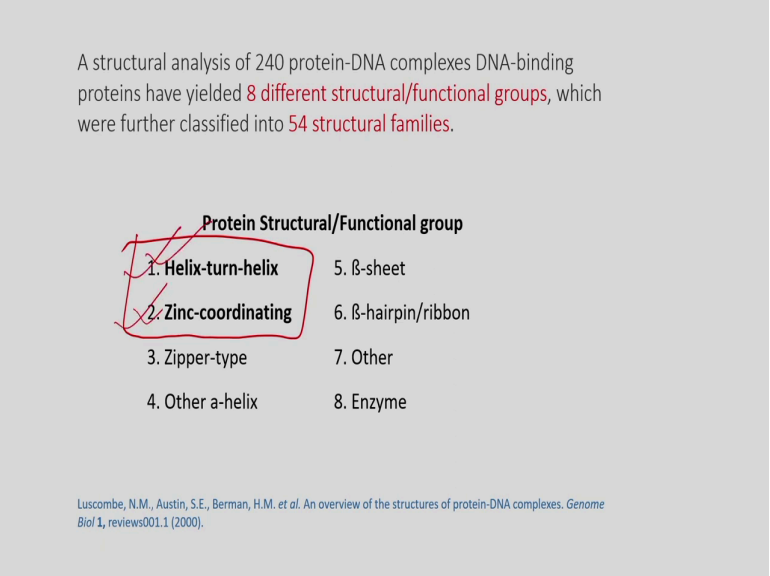
So, we now know that the proteins bind to DNA and RNA. So, there are proteins which we can call as DNA binding proteins and this study has reported that we have roughly around 588 DNA binding proteins. And we have a large number of RNA binding proteins which is more than double of the DNA binding proteins, being 1210 roughly. And there are some proteins a small set numbering something around 64 which binds to both DNA and RNA at the same time. So, these study is from the QuickGo database and this is the reference which you can consult for knowing more.

(Refer Slide Time: 10:35)



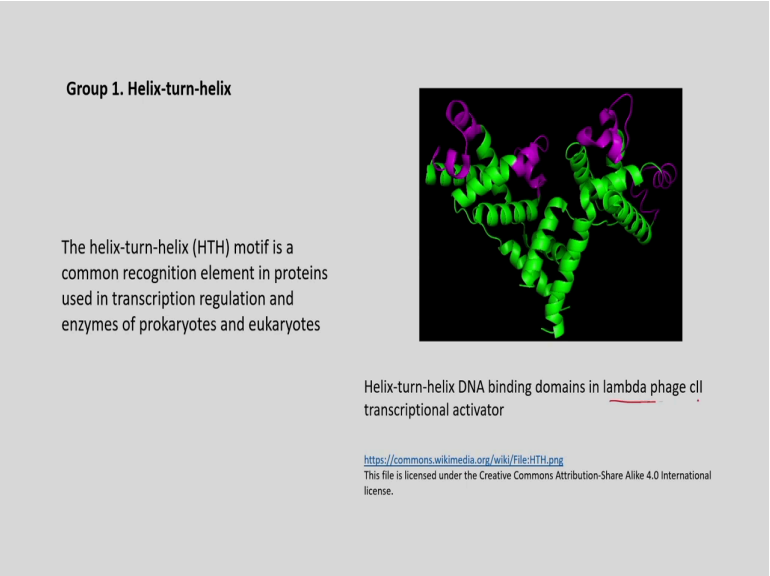
So, already little has been told to you about the importance of proteins in driving the cellular events. Now, let us focus on the DNA binding proteins which play a central role in all its aspects of genetic activity within an organism such as in transcription, packaging of DNA that we have discussed at length in the introductory classes. Then rearrangement of DNA that also has been discussed while studying homologous recombination and etcetera and replication and repair.

(Refer Slide Time: 11:14)



So, a study was carried out to analyze around 240 protein-DNA complexes of DNA binding proteins and all these yielded around 8 different structural functional groups, which were classified for into 54 structural families. So, basically these 8 large groups of protein structural functional group are the helix-turn-helix, the Zinc coordinating proteins, the zipper-types, then other alpha helix, and beta sheets, then beta-hairpin/ ribbon, then a miscellaneous group having diverse functions with some commonalities and there is eighth group, the enzymes. Some of the enzymes also fall into the other six classes. In this particular lecture, we will focus mostly on two protein structural or functional groups they are helix-turn-helix group and the Zinc coordinating group.

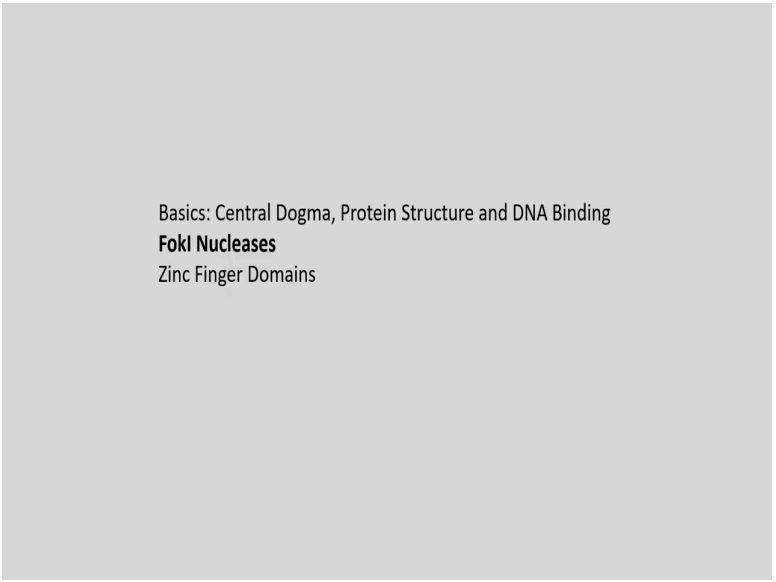
(Refer Slide Time: 12:22)



Let us first start with a discussion on the group one which is the helix-turn-helix. So, as you can see there are many helix over here; these are basically the alpha helix, helices which we discussed in the beginning of this lecture.

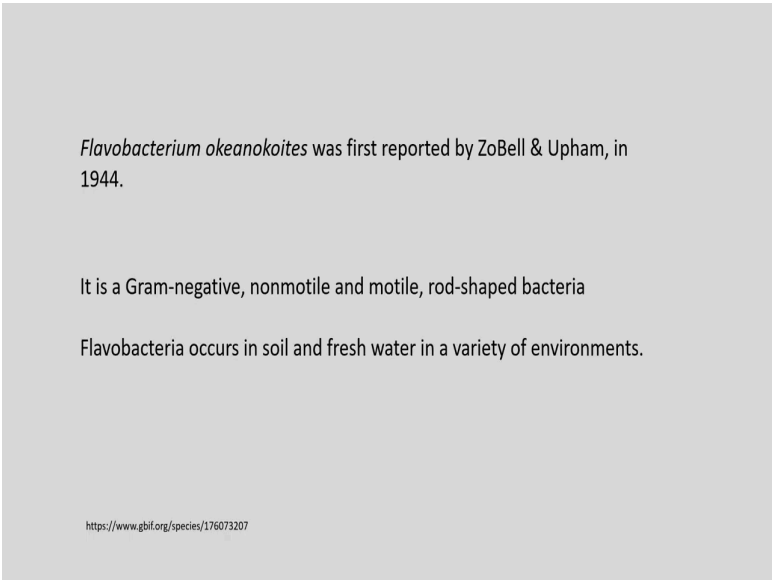
So, this HTH motif or the helix-turn-helix motif is a common recognition element in proteins used in transcription regulation, and enzymes of prokaryotes and eukaryotes. This is one important point we need to carry forward along with us for the remaining part of the lecture. Here you can see the helix-turn-helix DNA binding domains in lambda phase cII transcriptional activator.

(Refer Slide Time: 13:12)



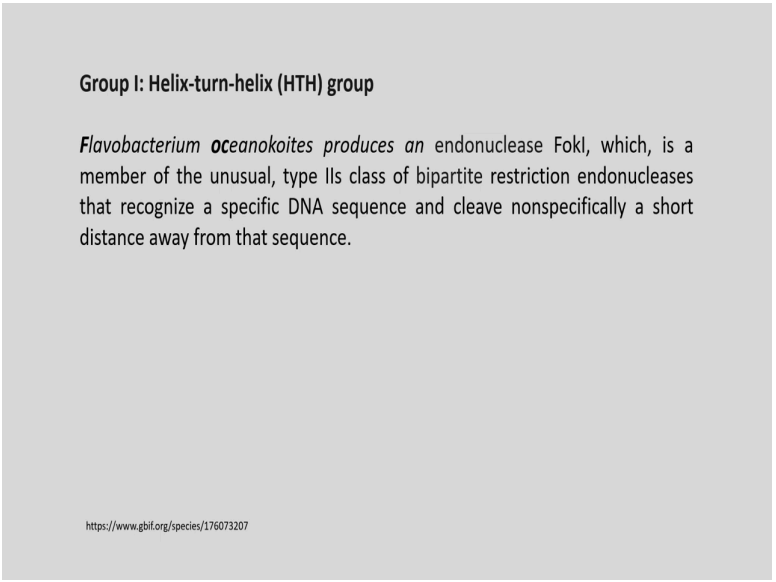
Let us now move on to the next part of this lecture the Fokl nucleuses. Let us now discuss about the Fokl nucleases which are important players in the ZFN technology platform.

(Refer Slide Time: 13:30)



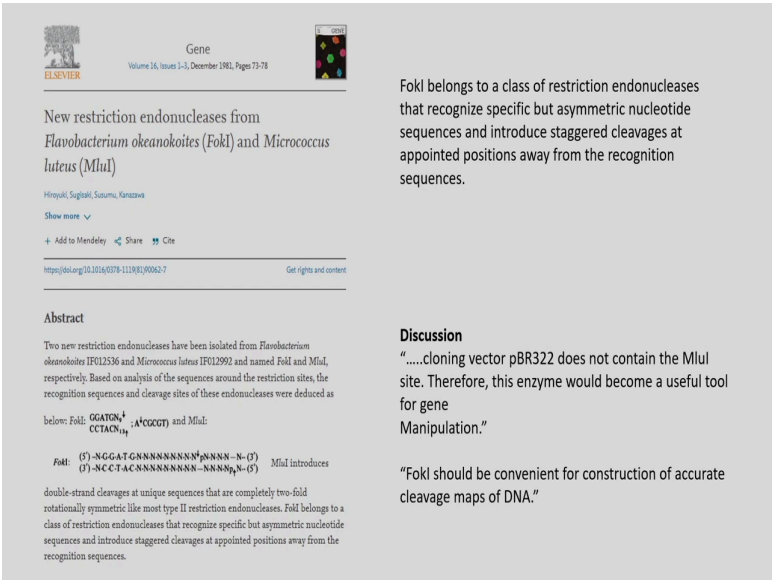
Flavobacterium okeanokoites, was first reported by ZoBell and Upham in 1944, which is a Gram-negative nonmotile and motile, rod shaped bacteria. It occurs in soil and fresh water in a variety of environments.

(Refer Slide Time: 13:48)



We just discussed about the helix-turn-helix group, structural functional group. This particular organism Flavobacterium Oceanokoites produces an endonuclease called FokI which is a member of the unusual, type two class of bipartite restriction endonucleases that recognizes a specific DNA sequence and cleave non-specifically a short distance away from that sequence.

(Refer Slide Time: 14:23)



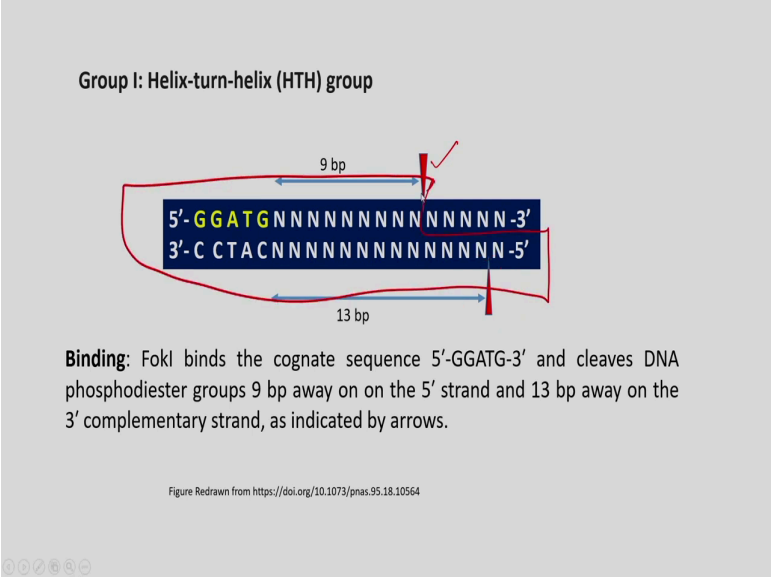
This is one of the very important discoveries in this particular field where Hiroyuki and his associate discovered a new restriction endonuclease from flavobacterium oceanokoites FokI

and another one from Micrococcus luteus. So in this paper, you can see from the abstract the importance of these two clones which produces these two type of endonucleases, restriction endonucleases. And here the Fok1 belongs to a class of restriction endonucleases that recognizes specific, but asymmetric nucleotide sequences and introduces staggered cleavages at the pointed positions away from the recognition sequences.

So, these are not cutting and recognizing at the same place and they produce blunt ends or staggered cleavages. In the discussion of this paper, the authors write about one observation where cloning vector pBR322 does not contain the Mlul one site; therefore, this enzyme could become a useful tool for gene manipulation. ok.

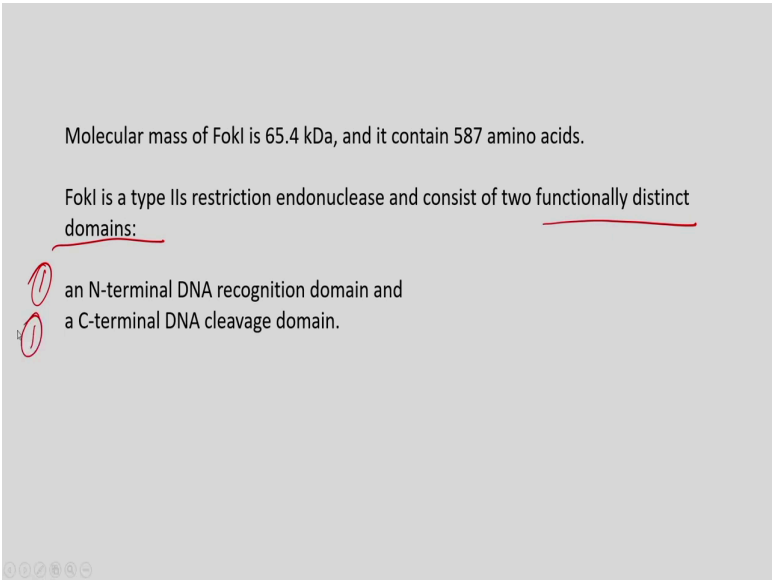
And they could not anticipate the potential of Fokl much and they concluded that Fokl should be convenient for construction of accurate cleavage maps of DNA, because Fokl cut non-specifically. So, with this; however, the scientific world came to know about the existence of this particular enzyme fokl. And soon it was forgotten, the discovery was something around sometime around 1981.

(Refer Slide Time: 16:33)



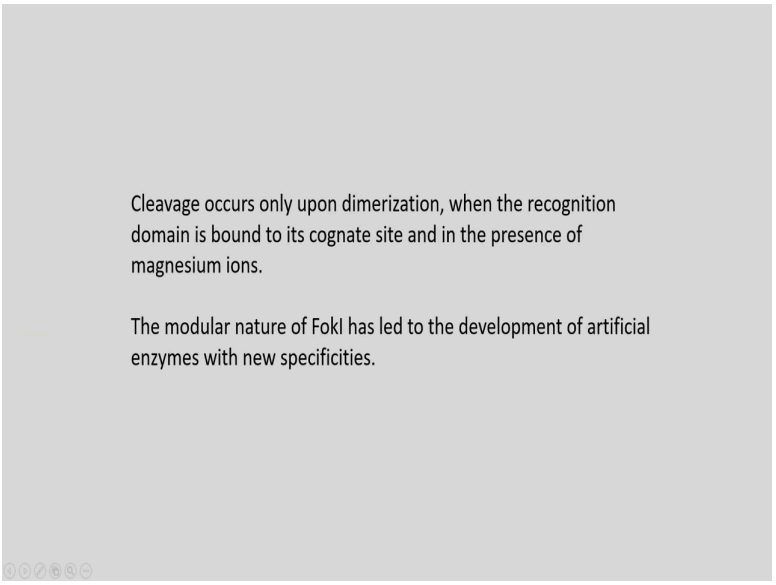
So, this is the binding sequence of the Fokl which binds cognate sequence 5 prime GGATG here and cleaves DNA phosphodiester groups 9 base pairs away here. So, it will bind here and it will cleave here on the 5 prime strand and 13 base pairs away on the 3 prime complementary strand and it will give a staggered cut as a result of this reaction.

(Refer Slide Time: 17:14)



The molecular mass of these enzyme is around 65.4 kilo Delton and it contains around 587 amino acids. And as already told you, this is a type two restriction in endonuclease and it consists of two functionally distinct domains, number one domain is the N terminal DNA recognition domain and a C terminal DNA cleavage domain. So, the domain which is in the N terminal will bind to this cognate sequence and domain at the C terminal will cleave at the 9 base pairs to 13 base pairs away.

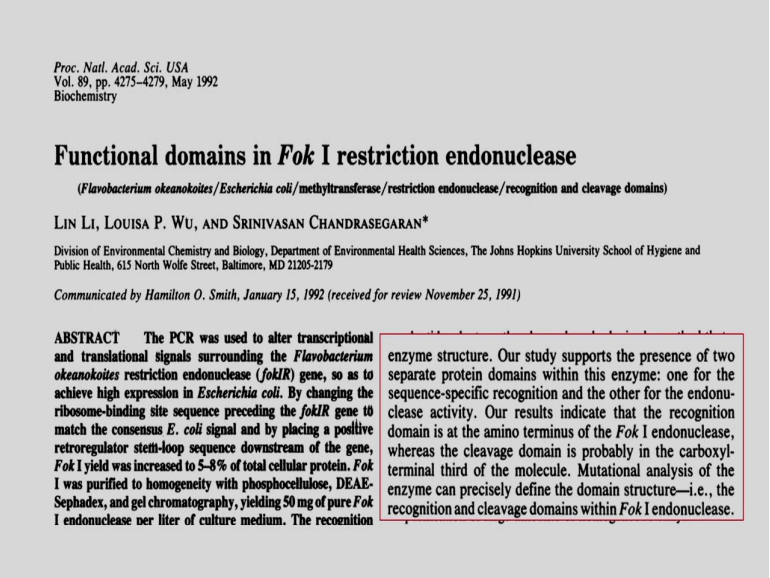
(Refer Slide Time: 18:02)



Now, cleavage occurs only upon dimerization, if only one Fokl enzyme binds to the DNA, it will not cleave. It needs a partner on the complementary strand as well. So, when the recognition domain is bound to its cognate side and in the presence of magnesium ions; so, these are some of the requirements for this enzyme to function.

So, as you can see that it has its two modules one is the N terminal domain and another is the C terminal domain. The modular nature of Fokl has led to the development of artificial enzymes with new specificities which we are going to study at length in the next lecture and little bit of that in this lecture too.

(Refer Slide Time: 18:50)



Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 4275–4279, May 1992
Biochemistry

Functional domains in *Fok* I restriction endonuclease

(*Flavobacterium okeanokoites*/*Escherichia coli*/methyltransferase/restriction endonuclease/recognition and cleavage domains)

LIN LI, LOUISA P. WU, AND SRINIVASAN CHANDRASEGARAN*

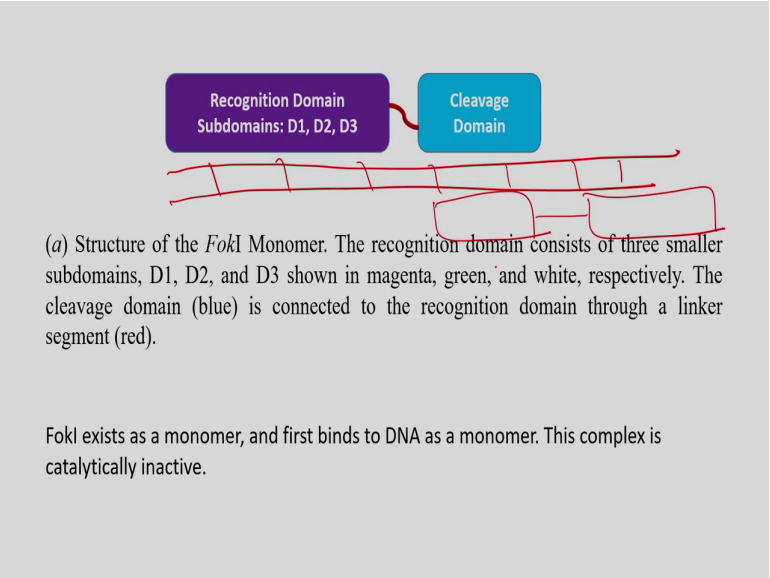
Division of Environmental Chemistry and Biology, Department of Environmental Health Sciences, The Johns Hopkins University School of Hygiene and Public Health, 615 North Wolfe Street, Baltimore, MD 21205-2179

Communicated by Hamilton O. Smith, January 15, 1992 (received for review November 25, 1991)

ABSTRACT **The PCR was used to alter transcriptional and translational signals surrounding the *Flavobacterium okeanokoites* restriction endonuclease (*fokIR*) gene, so as to achieve high expression in *Escherichia coli*. By changing the ribosome-binding site sequence preceding the *fokIR* gene to match the consensus *E. coli* signal and by placing a positive retroregulator stem-loop sequence downstream of the gene, *Fok* I yield was increased to 5–8% of total cellular protein. *Fok* I was purified to homogeneity with phosphocellulose, DEAE-Sephadex, and gel chromatography, yielding 50 mg of pure *Fok* I endonuclease per liter of culture medium. The recognition** enzyme structure. Our study supports the presence of two separate protein domains within this enzyme: one for the sequence-specific recognition and the other for the endonuclease activity. Our results indicate that the recognition domain is at the amino terminus of the *Fok* I endonuclease, whereas the cleavage domain is probably in the carboxyl-terminal third of the molecule. Mutational analysis of the enzyme can precisely define the domain structure—i.e., the recognition and cleavage domains within *Fok* I endonuclease.

Now, this is one of the important work carried out by Chandrasekaran and his colleagues and they published their results in 1982 where they identified the functional domains of fok1 restriction endonucleases. So, their study proved the presence of two separate protein domains within this enzyme which we have already discussed, one is for the site specific recognition and another one for the endonuclease activity.

(Refer Slide Time: 19:35)

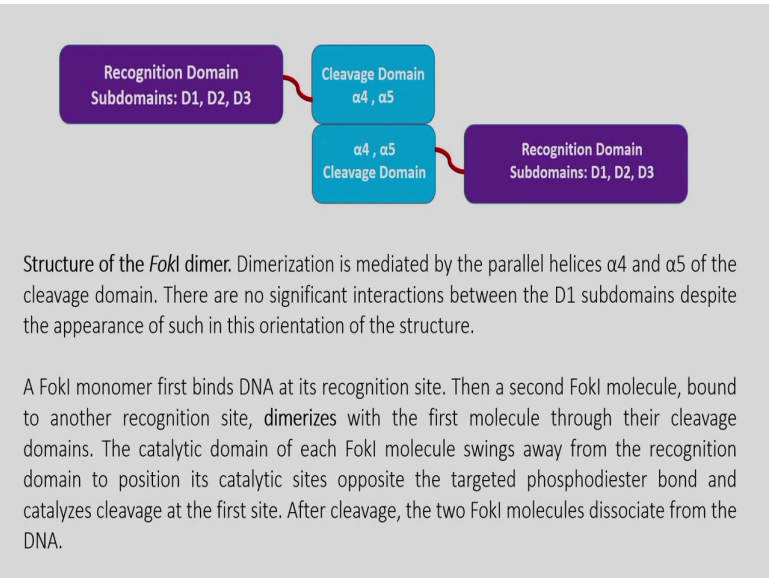


(*a*) Structure of the *Fok*I Monomer. The recognition domain consists of three smaller subdomains, D1, D2, and D3 shown in magenta, green, and white, respectively. The cleavage domain (blue) is connected to the recognition domain through a linker segment (red).

FokI exists as a monomer, and first binds to DNA as a monomer. This complex is catalytically inactive.

So, this recognition domain as you can see may have several subdomains D1, D2, and D3 and there is a cleavage domain in the N terminus. And this is connected through a linker molecule as shown in this particular picture cartoon. The FokI exist as a monomer, and first binds to DNA as a monomer, this complex is catalytically inactive. So, first time these particular monomer will bind to a DNA strand; ok, as a monomer, and then another FokI will bind on the other strand later.

(Refer Slide Time: 20:25)



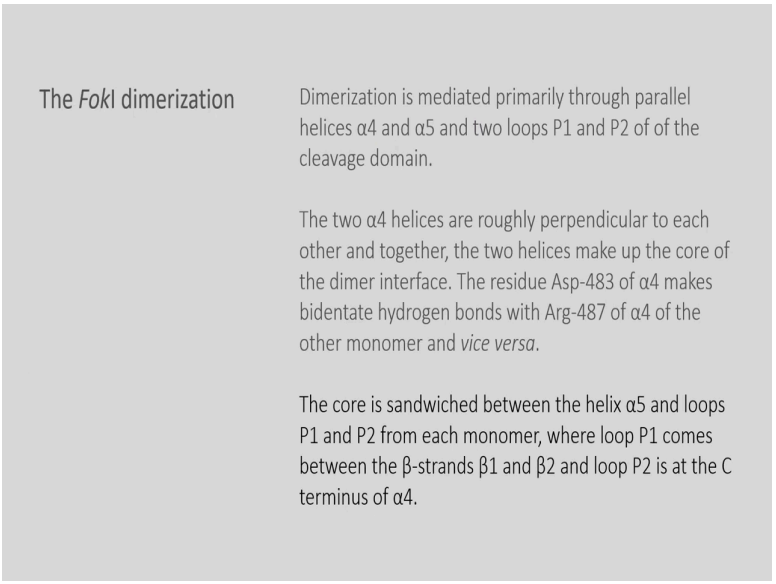
Structure of the *Fok*I dimer. Dimerization is mediated by the parallel helices α4 and α5 of the cleavage domain. There are no significant interactions between the D1 subdomains despite the appearance of such in this orientation of the structure.

A FokI monomer first binds DNA at its recognition site. Then a second FokI molecule, bound to another recognition site, dimerizes with the first molecule through their cleavage domains. The catalytic domain of each FokI molecule swings away from the recognition domain to position its catalytic sites opposite the targeted phosphodiester bond and catalyzes cleavage at the first site. After cleavage, the two FokI molecules dissociate from the DNA.

Structure of the Fokl dimer. Dimerization is mediated by the parallel helices alpha 4 and alpha 5 of the cleavage domain. There are no significant interactions between the D1 sub domains despite the appearance of such in this orientation of these structures. So, they remain further away from one another and they do not have any role in this dimerization. But the domains the alpha helices 4 and 5 in this cleavage domain play a major role.

A Fokl monomer first binds DNA at its recognition site; then a second Fokl molecule, bound to another recognition site would dimerize with the first molecule through their cleavage domains. The catalytic domain of each Fokl molecule swings away from the recognition domain to position its catalytic sites opposite to the targeted phosphodiester bond and catalyzes cleavage at the first site. After cleavage the two Fokl molecules dissociate from the DNA.

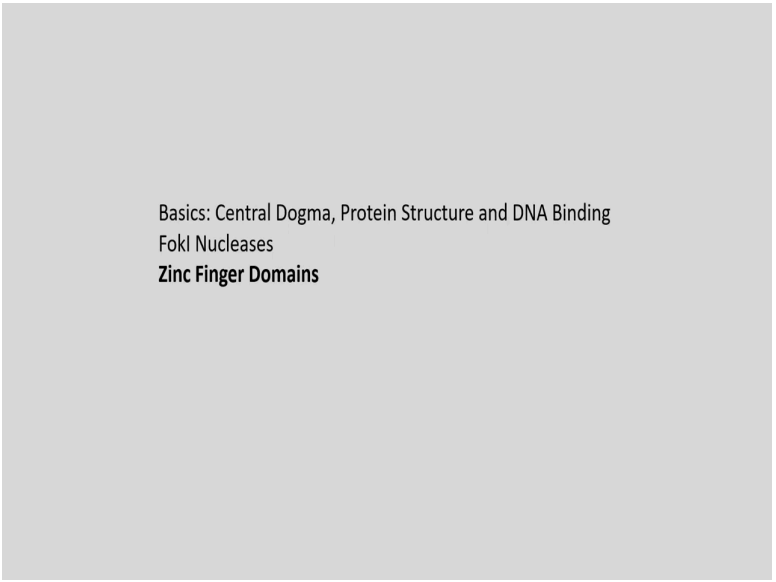
(Refer Slide Time: 21:32)



As already told to you the dimerization is primarily mediated through parallel helices alpha 4 and alpha 5 and two loops P1 and P2 of the cleavage domain. The two alpha 4 helices are roughly perpendicular to each other and together the two helices make up the core of the dimer interface.

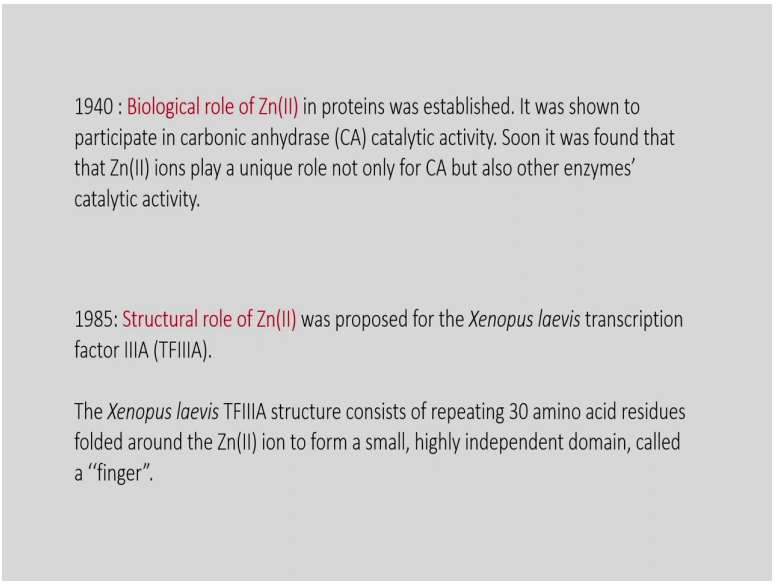
The residue Asp-483 of alpha 4 makes bidentate hydrogen bonds with Arg-487 of alpha 4 of the other monomer and vice versa. The core is sandwiched between the helix alpha 5 and loops P1 and P2 from each monomer, where loop P1 comes between the beta strands beta 1 and beta 2 and loop P2 is at the C terminus of alpha 4.

(Refer Slide Time: 22:21)



So, let us now move on to another important concept required for the ZFN technology, the Zinc Finger Domains. Let us now, discuss about the zinc finger domains which are very important for understanding the ZFN technology.

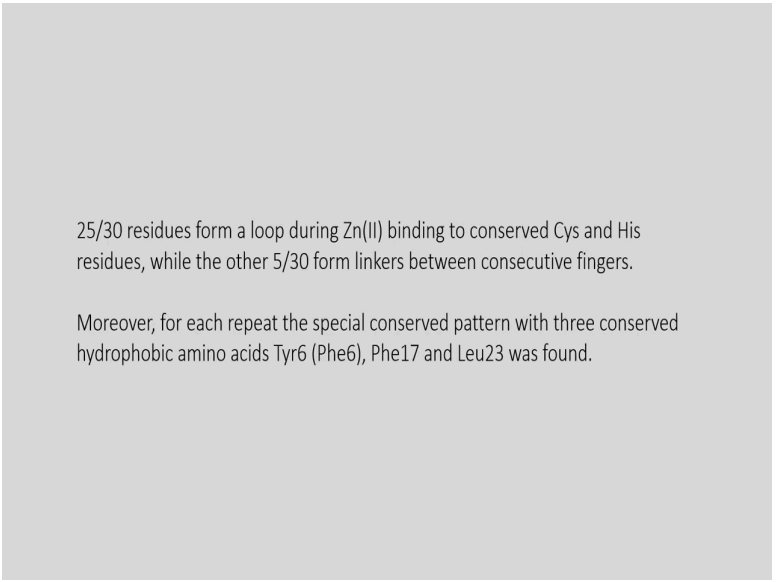
(Refer Slide Time: 22:47)



The biological role of zinc in proteins was established in 1940 and it was shown to participate in carbonic anhydrase catalytic activity. Soon it was found that the zinc ions play unique role not only for CA but also other enzymes catalytic activity. In 1984, the structural role of zinc was proposed for the Xenopus laevis transcription factor IIIA or TFIIIA. The Xenopus laevis

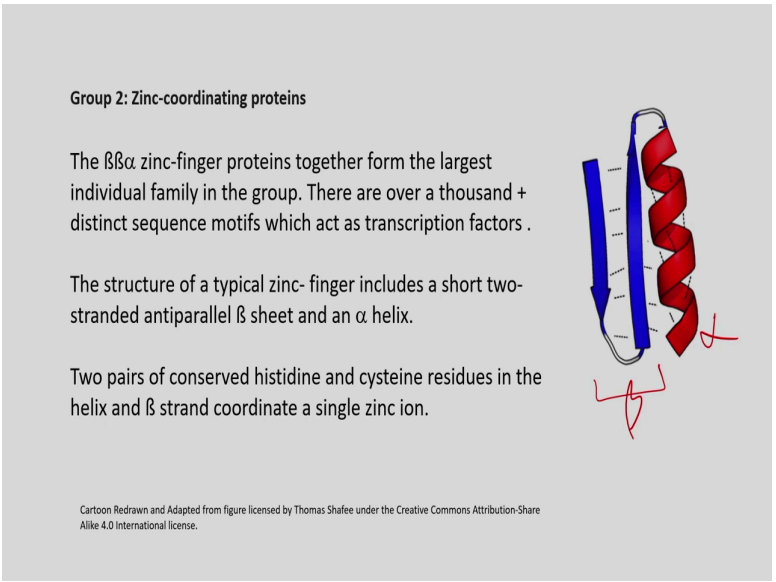
TFIIIA structure consists of repeating 30 amino acid residues folded around the zinc ion to form a small highly independent domain called a finger.

(Refer Slide Time: 23:34)



Around 20 to 30 residues form a loop during zinc binding to conserved cysteine and histidine residues, while the other 5 residues form linkers between conjugative fingers. Moreover, each repeat the special conserved pattern with three conserved hydrophobic amino acids tyrosine 6, phenylalanine 17 and leucine 23 was also reported.

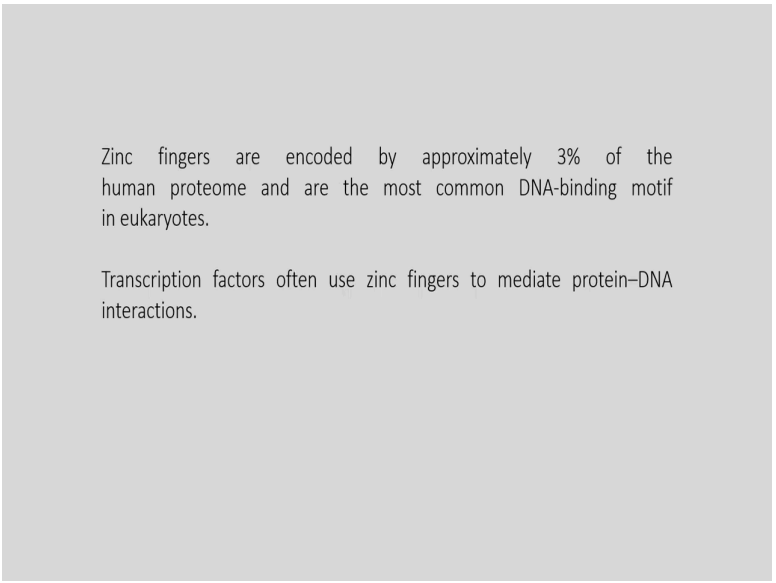
(Refer Slide Time: 24:01)



Let us focus on some of the features of zinc coordinating proteins, the beta beta alpha zinc finger proteins together from the largest individual family in the group. There are over a thousand plus distinct sequence motifs which acts as transcription factors. The structure of a typical zinc finger includes a short two stranded antiparallel beta sheet and an alpha helix.

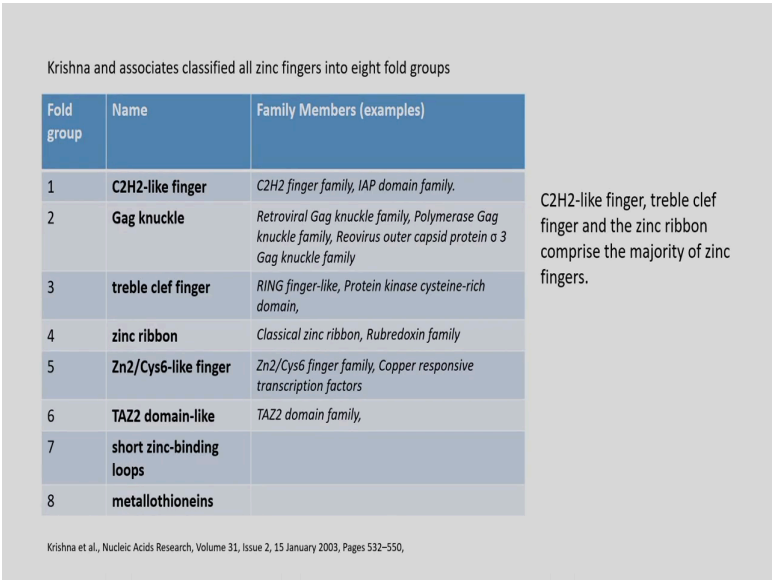
So, you can see here the beta sheets are antiparallel and it has a alpha helix over here. Two pairs of conserved histidine and cysteine residues in the helix and beta strand coordinate a single zinc ion.

(Refer Slide Time: 24:52)



Zinc fingers are encoded by approximately 3 percent of the human proteome and are the most common DNA binding motif in eukaryotes. Transcription factors often use zinc fingers to mediate protein DNA interactions.

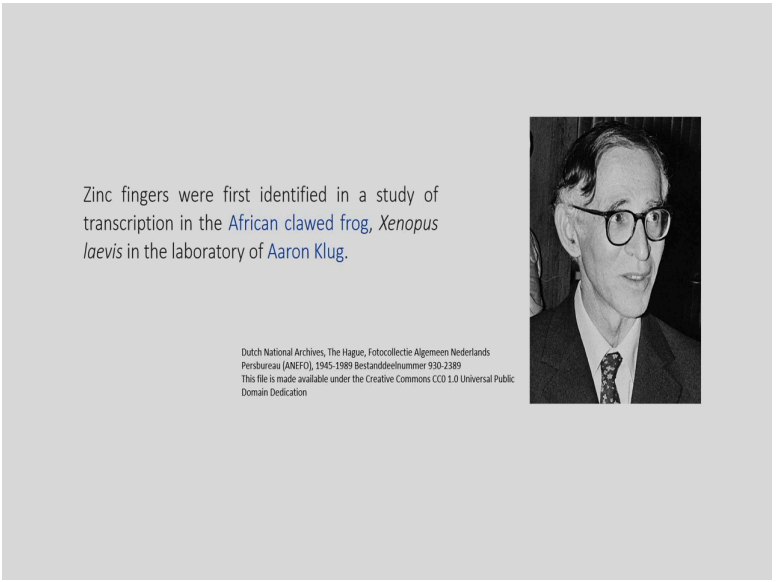
(Refer Slide Time: 25:11)



Krishna and associates classified all zinc fingers into eight-fold groups and you can see here the eight different fold groups over here. And having given different names the C2H2 like finger, Gag knuckle, treble clef, a zinc ribbon and so on. Among these you can have many family members for each whole group; for example, two of the family members that we have mentioned here are the C2H2 finger family IAP domain family and in a Gig knuckle you can see couple of families being listed over here.

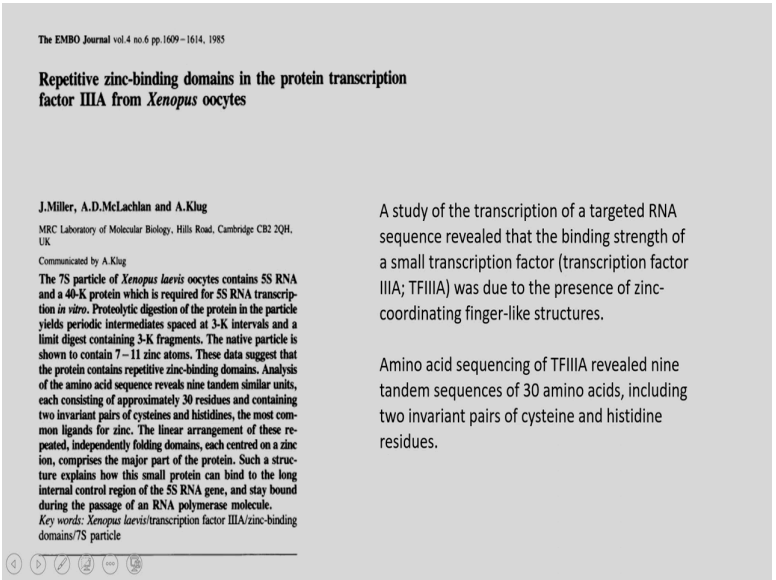
And accordingly, also in the treble clef you have the ring finger like protein kinase cysteine type domains and so on. Among all these 8 folds the C2H2 like finger, the treble clef finger, and the zinc ribbon finger, constitute the majority of the zinc fingers available in the biological world.

(Refer Slide Time: 26:21)



These zinc fingers were first identified in a study of transcription in the African clawed frog, Xenopus laevis as already told to you earlier in the laboratory of Aaron Klug.

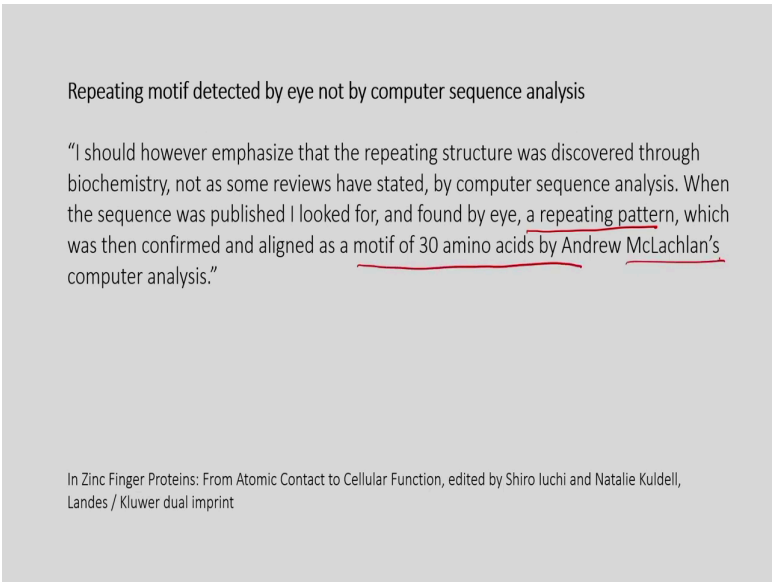
(Refer Slide Time: 26:36)



And this is the landmark paper Aaron Klug and his co-workers, Miller, McLachlan published in the EMBO journal in 1985. You can refer to this paper for full details of this particular discovery. Briefly, in this paper a study of the transcription of a targeted RNA sequence revealed that the binding strength of a small transcription factor TFIII3A was due to the presence of zinc coordinating finger like structures.

To know about these gene coordinating finger like structures, amino acid sequencing was done. And the amino acid sequencing of these particular TFIII3A factor revealed nine tandem sequences of 30 amino acids including two invariant pairs of cysteine and histidine residue.

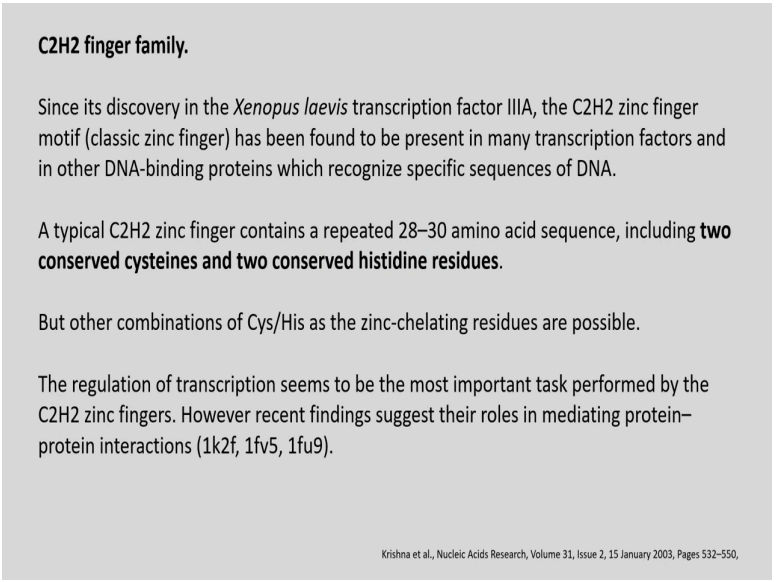
(Refer Slide Time: 27:46)



Interestingly in this work, they did not use much of computational power and these repeating motives were detected by the eye, scanning by the eye and not by computer analysis. So, it is very important to know some of the discovery is being done in very simple ways and these are the words of Aaron Klug.

I should however emphasize that the repeating structure was discovered through biochemistry and not as some reviews have stated by computer sequence analysis. When the sequence was published, I looked for and found by eye, a repeating pattern which was then confirmed and aligned as motif of 30 amino acids by Andrew McLachlan's computer analysis.

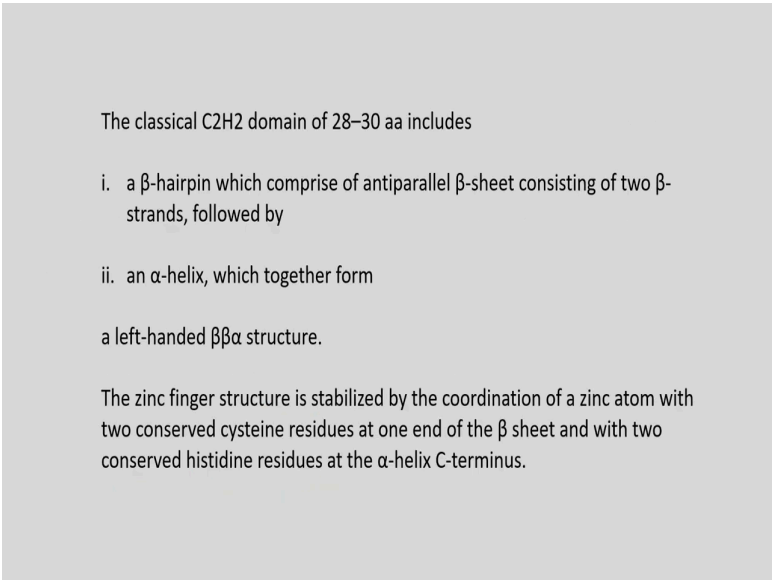
(Refer Slide Time: 28:46)



C2H2 finger family. So, you have two cysteine and two cysteine histidine as signature amino acids in these particular families. And since its discovery in the Xenopus laevis transcription factor IIIA, the C2H2 zinc finger motif also referred to as the classic zinc finger has been found to be present in many transcription factors and in other DNA binding proteins which recognize specific sequences of DNA.

A typical C2H2 zinc finger contains a repeated 20 to 30 amino acid sequences, including two conserved cysteine and two conserved histidine residues. But other combinations of cysteine and histidine is possible; for example, in zinc chelating, as zinc chelating residues. The regulation of the transcription seems to be the most important task performed by the C2H2 zinc fingers. However, recent studies have suggested their roles in mediating protein-protein interactions as well.

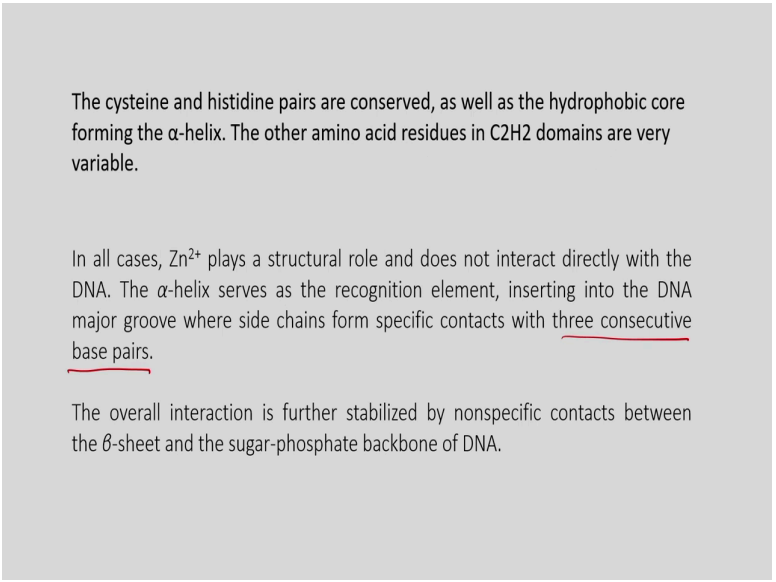
(Refer Slide Time: 30:10)



So, we now know that in Xenopus laevis, they had found around 30 amino acid residues, but further studies reveal that the C2H2 domain has a range of around 28 to 30 amino acids. And these C2H2 domains include certain secondary structure elements. A beta hairpin which comprise of antiparallel beta sheets consisting of two beta strands. So, two beta strands in anti-parallel orientation give rise to a beta hairpin and which is a important component of these classical C2H2 domain.

Then it has in addition to this, an alpha helix and together they form a left-handed beta beta alpha structure. Two beta coming from here and one alpha coming from here and the beta hair pin and the alpha helix. The zinc finger structure is stabilized by coordination of a zinc atom with two conserved cysteines residues at one end of the beta sheet and with two conserved histidine residues at the alpha helix C terminus.

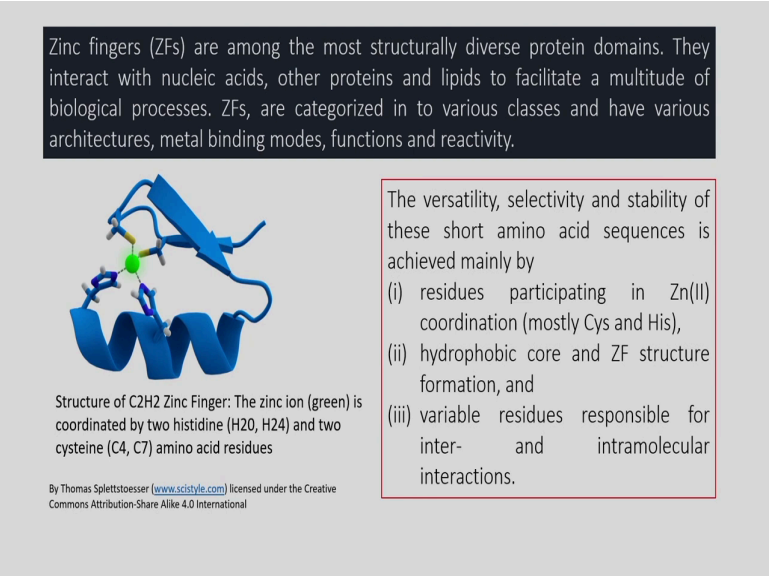
(Refer Slide Time: 31:37)



Not only the cysteine and histidine pairs are conserved, there is a highly conserved hydrophobic core forming the alpha helix. The other amino acid residues in C2H2 domains are largely variable. In all cases, zinc ion plays a structural role and does not interact directly with the DNA.

The alpha helix serves as the recognition element inserting into the DNA major groove; whereas, side chains form specific contacts with three consecutive base pairs. The overall interaction is further stabilized by non-specific contacts between the beta sheet and the sugar phosphate backbone of DNA.

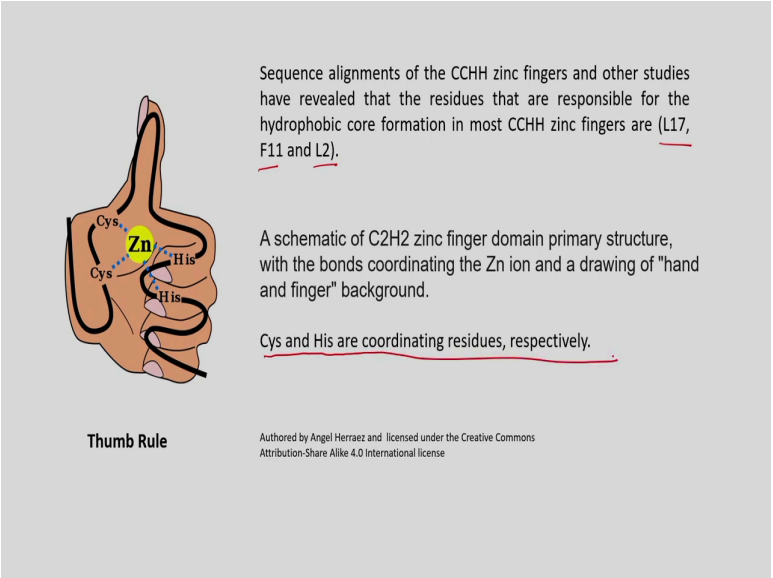
(Refer Slide Time: 32:30)



The zinc finger domains are the among the most structurally diverse protein domains. They interact with nucleic acids, other proteins and lipids to facilitate a multitude of biological processes. Zinc fingers are categorized into various classes and have various architectures, metal binding modes, functions and reactivity. So, here you can see the structure of C2H2 zinc finger and the zinc ion which is shown in green is coordinated by two histidine and two cysteine amino acid residues about which we have discussed in the previous slide.

The versatility, selectivity, and stability of these short amino acid sequences is achieved mainly by three things. Residue participating in zinc coordination mostly by cysteine and histidine. Then hydrophobic core and ZF structure formation and variable residues responsible for inter and intramolecular interactions.

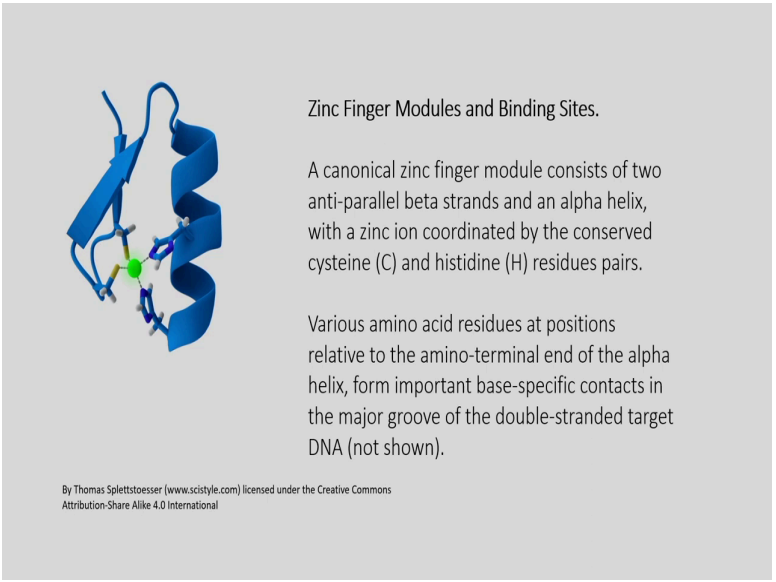
(Refer Slide Time: 33:42)



So, this is the thumb rule which you may remember to understand the zinc coordination. If you put your wrist in this orientation with the thumbs up, then in this side you have the two cysteine residues and in on these two fingers you have the zinc residues. And in your palm below the thumb, you have the zinc molecule and this is how they bind and stabilize the ZF structure.

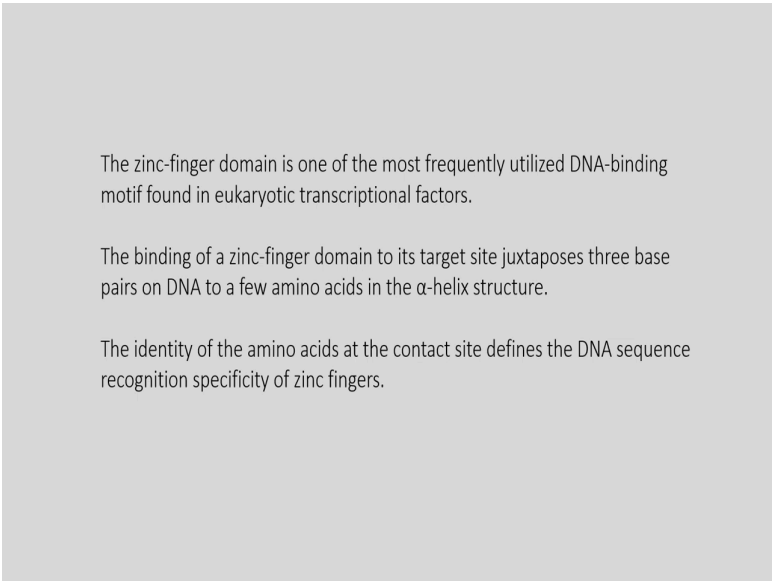
Sequence alignments of the C2H2 zinc fingers and other studies have revealed that the residues that are responsible for the hydrophobic core formation in most C2H2 zinc fingers are L7, F11, and L2. And this is a semantic of C2H2 zinc finger domain primary structure, with the bonds coordinating the zinc ions and a drawing of hand in finger background, the cysteine and histidine are coordinating residues respectively.

(Refer Slide Time: 34:53)



A canonical zinc finger module consists of two antiparallel beta strands and an alpha helix with a zinc ion coordinated by the conserved cysteine and histidine residue pairs. A various amino acid residues at positions relative to the amino terminal end of the alpha helix form important base specific contacts in the major group of the double stranded target DNA.

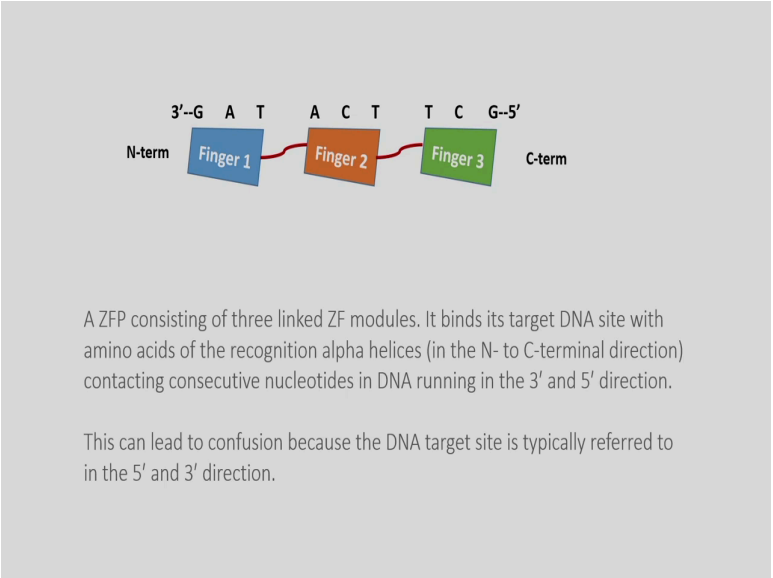
(Refer Slide Time: 35:19)



The zinc finger domain is one of the most frequently utilized DNA binding motif found in eukaryotic transcription factors as well. The finding of a the binding of a zinc finger domain to its target site juxtaposes three base pairs on DNA to a few amino acids in the alpha helix

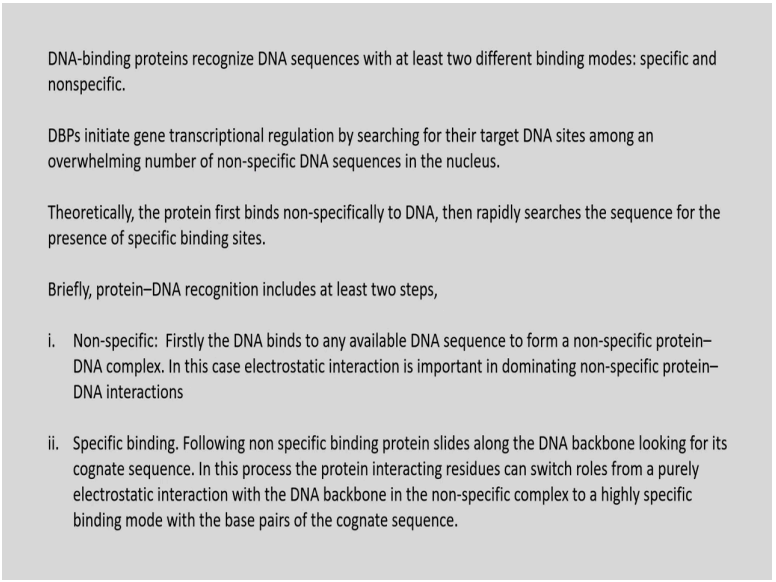
structure. The identity of the amino acids at the contact site defines the DNA sequence recognition specificity of the zinc fingers.

(Refer Slide Time: 35:47)



So, you have here three fingers, finger 1 2 3 and you have some cognate sequences which are in the DNA. So, basically a ZFP consisting of three linked ZF modules are being shown in this cartoon and it binds its target DNA site with amino acids of the recognition alpha helices in the N to C terminal direction contacting consecutive nucleotides in DNA running in the 3 prime to 5 prime directions. This can sometimes lead to confusion because the DNA target site is typically referred in the other direction which is the 5 to 3 prime direction.

(Refer Slide Time: 36:37)



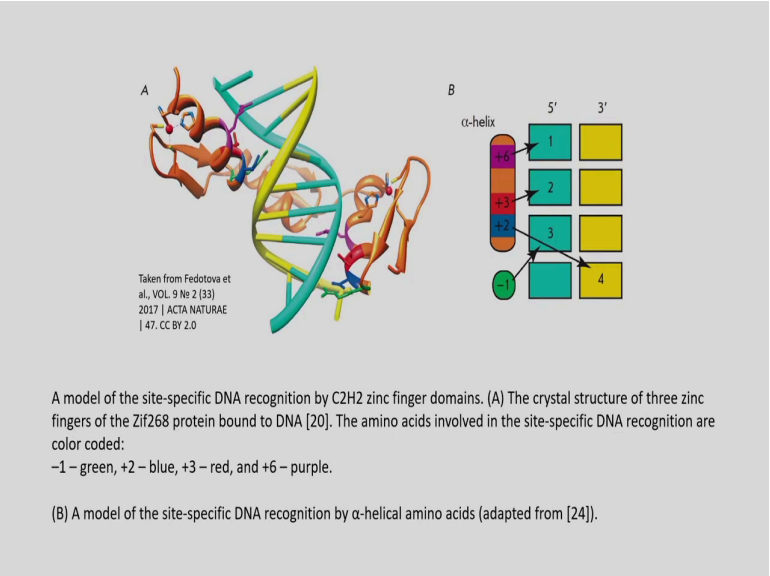
Let us now discuss how the proteins recognize DNA sequences and bind to them. The DNA binding proteins recognize DNA sequence with at least two different binding modes- specific and non-specific modes. DNA binding proteins would initiate gene transcriptional regulation by searching for their target DNA sites among an overwhelming number of non-specific DNA sequences in the nucleus.

Theoretically, the protein first binds non-specifically to any given sequence of DNA. And then rapidly searches the sequence for the presence of specific winding sites. In doing so, it sometimes would move along the DNA strand to find for its cognate sequences. So, briefly the protein DNA recognition involves at least two steps; the first step is the nonspecific binding, the protein would bind to any given sequence of DNA without bothering for the target or the cognate sequence.

So, firstly, the DNA binds to any available DNA sequence to form a non-specific protein DNA complex. And in these nonspecific reaction, the electrostatic interactions are the important dominating factors; later on its gradually would shift towards a specific binding. So, following nonspecific binding a protein will slide along the DNA back bone looking for its cognate sequence.

In this process, the protein interacting residues can switch roles from a purely electrostatic interaction with the DNA backbone in the non-specific complex to a highly specific binding mode with the base pairs of the cognate sequence.

(Refer Slide Time: 38:47)

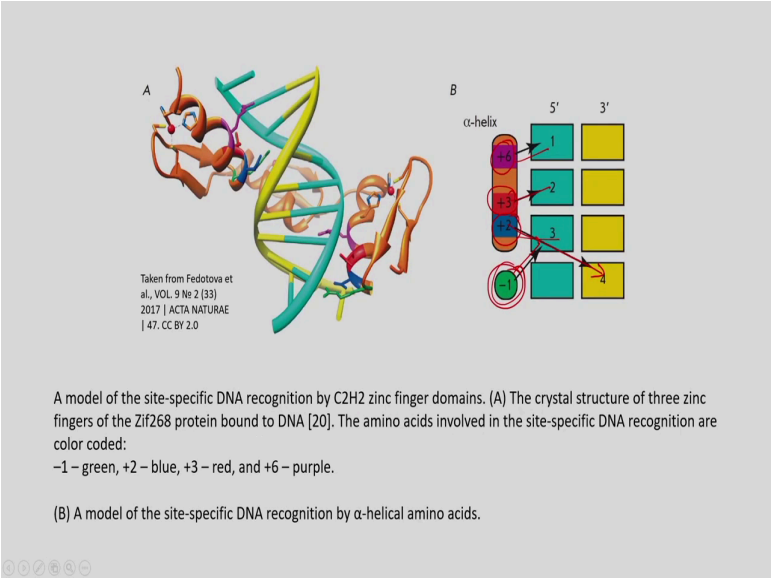


A model of the site-specific DNA recognition by C2H2 zinc finger domains. (A) The crystal structure of three zinc fingers of the Zif268 protein bound to DNA [20]. The amino acids involved in the site-specific DNA recognition are color coded:
−1 – green, +2 – blue, +3 – red, and +6 – purple.

(B) A model of the site-specific DNA recognition by α-helical amino acids (adapted from [24]).

So, here is a model of the site-specific DNA recognition by C2H2 zinc finger domains. So, in figure A, you can see a crystal structure of three zinc fingers; 1, 2, and 3 and this is the DNA molecule in the center. The amino acids involved in the site specific DNA recognition are color coded; minus 1 is green and plus 2 is blue, plus 3 is red and plus 6 is purple. On the corresponding panel figure B is a model of the site specific DNA recognition by alpha helical amino acids.

(Refer Slide Time: 39:41)



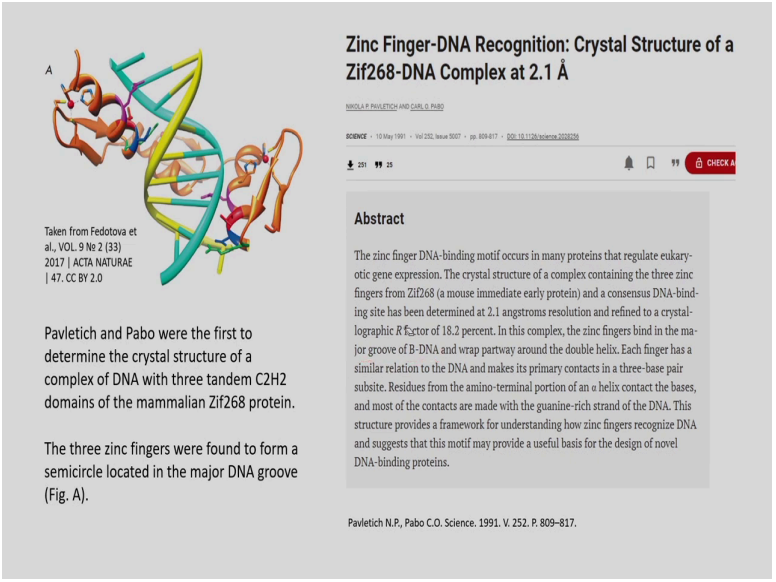
A model of the site-specific DNA recognition by C2H2 zinc finger domains. (A) The crystal structure of three zinc fingers of the Zif268 protein bound to DNA [20]. The amino acids involved in the site-specific DNA recognition are color coded:
−1 – green, +2 – blue, +3 – red, and +6 – purple.

(B) A model of the site-specific DNA recognition by α-helical amino acids.

A model of the site-specific DNA recognition by C2H2 finger domains is being shown in this figure; figure A, you can see the crystal structure of three zinc fingers of the Zif268 protein bound to DNA. So, the amino acids involved in the site-specific DNA recognition are color coded; minus 1 green, plus 2 blue, plus 3 red and plus 6 purple and these are the three zinc fingers which are binding to the DNA molecule in the center.

In panel B, you can see a model of site-specific DNA recognition by alpha helical amino acids. So, here you can see plus 6 corresponding to 1, plus 2 corresponding to 2, plus 2 corresponding to 4 and minus 1 corresponding to 3; so, we will discuss about these relationship in a later slide.

(Refer Slide Time: 40:54)

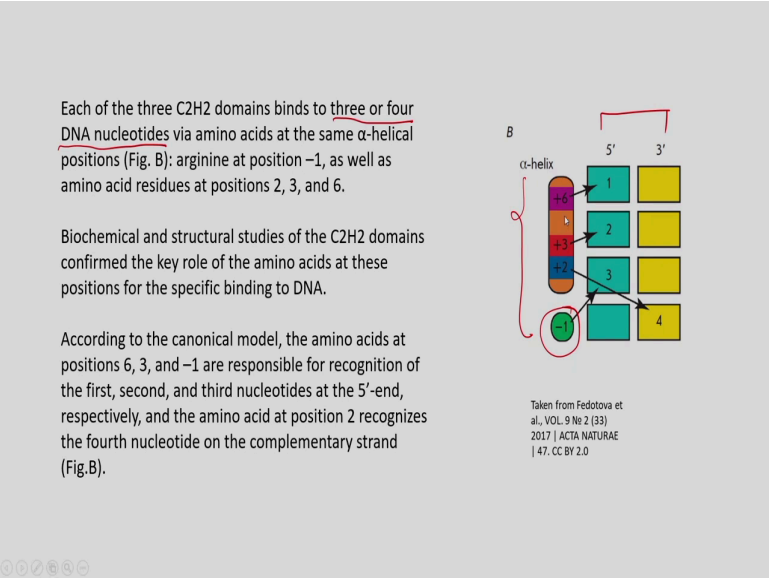


So, this is one of the important paper in finding out how zinc fingers recognize DNA molecule. And these work by Pavletich and Pabo were the first to determine the crystal structure of a complex of DNA with three tandem C2H2 domains of the mammalian ZF 268 proteins. The three zinc fingers were found to form a semicircle located in the major DNA groove.

So, in this complex as you can see in their abstract, the zinc fingers bind in the major group of beta DNA and wrap partway around the double helix. Each finger has a similar relation to the DNA and makes its primary contact in a three base pair subsite. Residues from the amino terminal portion of an alpha helix contact the basis and most of the contacts are made with the guanine rich strands of the DNA. And their work provided a framework for understanding

how zinc fingers recognize DNA and suggest that this motif may be useful basis for the design of novel DNA binding proteins.

(Refer Slide Time: 42:18)

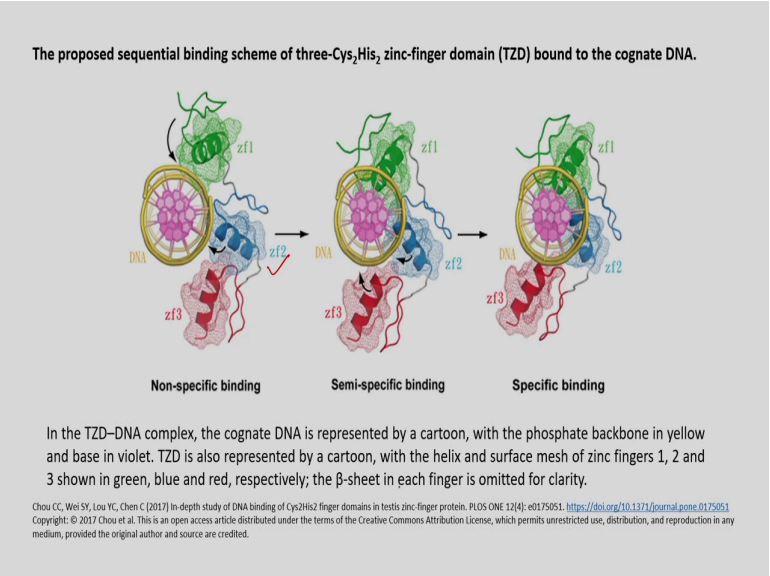


So, each of the three C2H2 domains bind to three or four DNA nucleotides via amino acids at the same alpha helical position. So, these amino acids positions are in the alpha helix; so, they are in the positions minus 1, plus 2, plus 3 and plus 6. And these are the positions of the DNA nucleotides in the forward strand and in the reverse strand as you can see 5 prime and this is the 3 prime.

So, arginine at position minus 1 as well as the amino acid residues at position 2, 3 and 6 respectively take part in this interaction. The biochemical and structural studies of these C2H2 domains confirm the key role of the amino acid at these positions for the specific binding to DNA. And there is a canonical model which suggest that the amino acids at the positions 6, 3 and 1 are responsible for recognition of the first, second and third nucleotides at the 5 prime end or the forward strand of the DNA respectively.

And the sorry….. these 3, 6 and minus 1 are responsible for binding to 1, 2 and 3 nucleotides in the forward strand or the 5 prime strand. And the amino acid at the second position binds to the fourth nucleotides on the complementary strand. So, this is very important to remember how the zinc fingers C2H2 domains bind to a double stranded DNA molecule. It is not just the 3 basis on the forward strand, a fourth base also sometimes play an important role and this fourth base is interacted by the second amino acid residue.

(Refer Slide Time: 44:52)



Now, we studied about the binding modes how it starts with a non-specific binding in the beginning and then it the complex slides along the DNA molecule. And then upon finding its cognate partner it transforms into a more specific binding interaction. So, this is a proposed sequential binding scheme of three Cys2His2 zinc-finger domain or TZD bound to the cognate DNA and this work was carried out in the testis zinc finger protein domain.

And you can refer to these paper to understand in details how this work was carried out. Now, let us look into these figures; so, these are the zinc finger 1, zinc finger 2, and zinc finger 3 and as discussed, all earlier, they are wrapping the DNA molecule in a semi-circle as you can see over here. So, in this TZD-DNA complex the cognate DNA represented by cartoon with the phosphate backbone in yellow here, and the basis inside in violet, ok, and they are inside these molecule.

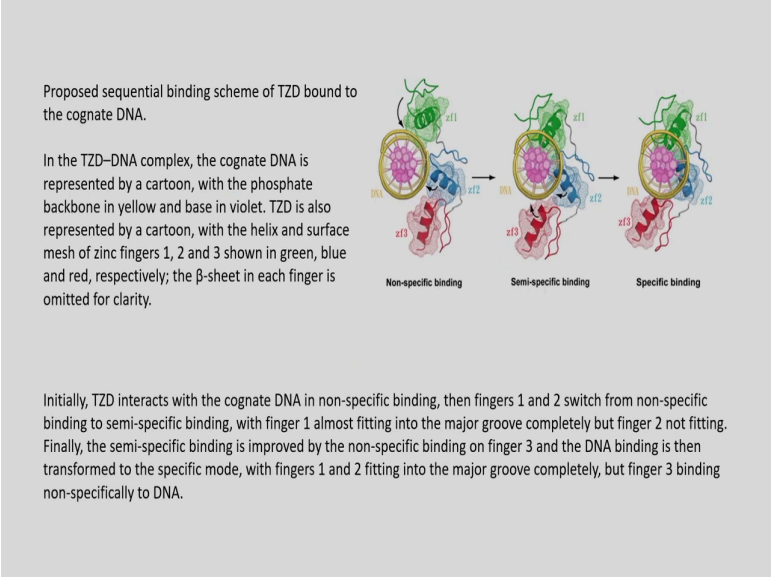
TZD is also represented by a cartoon with the helix and surface mesh of zinc fingers 1, 2 and 3 shown in green, blue and red respectively. The beta sheet in each finger is omitted in this case for clarity. So, initially the C2H2 zinc fingers bind to the DNA in a nonspecific manner and then you can see that zinc fingers 1 and 2 gradually trying to interact with the bases in the center to find out its cognate sequence. And this type of condition is known as the semi-specific binding state.

Now, this will keep on occurring until and unless the protein molecule find its cognate sequence. So, it always attempts to find out by reading or interacting with the bases in the

core its cognate sequence and if it fails to find it; so, it will slide to a next location and try to read it.

And this goes on and on until it will find its cognate sequence and whenever that cognate sequence is being found, it changes to the specific binding, already discussed to you. So, you can see that this is not a very simple reaction, it is quite dynamic and it is a heat and trial effort or method after which a zinc finger finds its cognate DNA sequence.

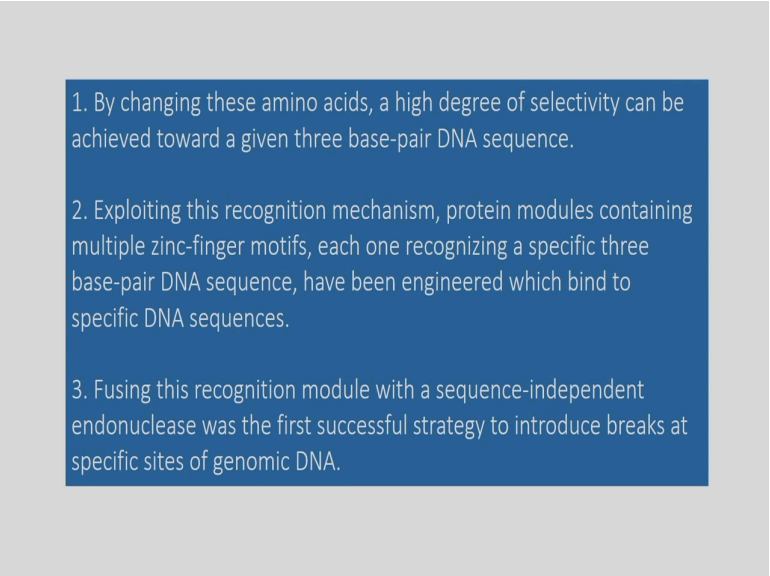
(Refer Slide Time: 48:38)



So, in the TZD-DNA complex, the cognate DNA is represented by a cartoon as already told to you. Initially, the TZD interacts with the cognitive DNA in a nonspecific binding, the fingers want to switch from non-specific binding to semi specific binding with finger 1 almost fitting into the major group completely, but finger 2 not fitting. Finally, the semi specific binding is improved by the non-specific binding on finger 3.

And the DNA binding is then transformed to the specific mode with fingers 1 and 2 fitting into the major group completely, but finger 3 binding non-specifically to the DNA. So, we now come to end of the discussion on the basics of zinc finger proteins as well as other proteins like FokI and together they are the important concepts for the zinc finger nucleus technology.

(Refer Slide Time: 49:55)



So, now we know that the zinc fingers binds to DNA cognate sequences. By changing these amino acids, a high degree of selectivity can be achieved toward a given three base pair DNA sequence. And exploiting this recognition mechanism, protein modules containing multiple zinc finger motifs each one recognizing a specific three base pair DNA sequence have been engineered which bind to specific DNA sequences.

And fusing these recognition module with a sequence independent endonucleus or DNA cutter was the first successful strategy to introduce breaks at specific sites of genomic DNA. And, we have studied in detail about the double strand DNA breaks and how those are helpful in homologous recombination and other mutational events. So, in the next lecture, we are going to discuss at length about utilizing these basic concepts in the development of zinc finger nucleases.

Thank you.