Eukaryotic Gene Expression

Prof. P. N. Rangarajan

Department of Biochemistry

Indian Institute of Science, Bangalore

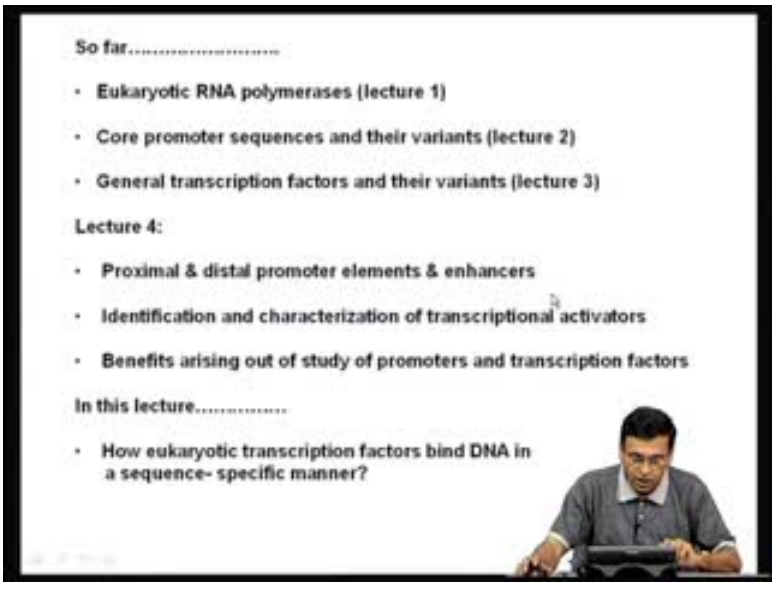
Module No. # 02

Lecture No. # 05

Eukaryotic Transcription Factors: The DNA Binding Domain

Today's lecture on eukaryotic gene expression, we are going to actually study the one important aspect of eukaryotic transcription factors, that is, we are going to understand how eukaryotic transcription factors go and bind to specific DNA sequences. We are going to focus on what is called as the DNA binding domain of eukaryotic transcription factors.

(Refer Slide Time: 00:40)



Now, just before I start, I just want to recapitulate what we have studied, so far. So far, what we have studied is, we have studied about eukaryotic RNA polymerases; how the eukaryotic RNA polymerases I, II, and III, actually go and bind to specific promoters and activate the transcription of messenger RNA, and make messenger RNA, transfer RNA, ribosomal RNA, and so and so forth.

Then, I told you, in order for eukaryotic RNA polymerase to bind to a protein coding gene promoter, you require what are called as general transcription factors. They are what are called as core promoter elements and the binding of these general transcription factors and to the core promoter sequences, actually, is responsible for initiation of transcription.

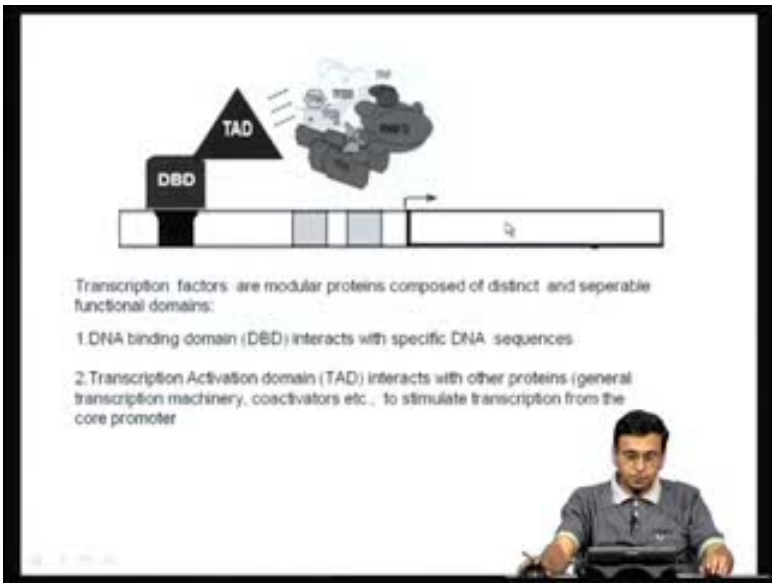
And I have also mentioned to you, how there are variations within the core promoter elements and there are also variants of general transcription factors, and this itself can contribute to differential gene regulation. Then, in the last class, we actually told, discussed, it is not just the core promoter elements; there are also what is called as proximal promoter elements, distal promoter elements, enhancer sequences, which are present much upstream of the core promoter region.

And there are what are called as sequence-specific DNA binding proteins or upstream activators or transcription factors, which, actually, can bind in a sequence-specific manner to these various promoter elements, and this binding, ultimately, enhances the rate of transcriptionization of eukaryotic protein coding genes, and I also briefly mentioned, this knowledge– that there are strong promoters, there are weak promoters; there are individual promoters, there are constitutive promoters.

This knowledge has actually helped in a…, what is called as a billion dollar biotechnology industry, because today, using this knowledge, we can now make any protein of our interest, in any expression system of our choice. You can make a protein in bacteria, yeast, insect cells, mammalian cells, plant cells, anywhere, by simply taking the gene and putting under a promoter of your choice, and putting them inside the particular organism or cells, you can actually make a recombinant protein production.

Now, in this lecture today, we are going to now discuss how exactly the eukaryotic transcription factors go and bind to specific DNA. Remember, there are two important things when you talk about transcription activation– one case, the eukaryotic transcription factor should go and bind to a specific DNA sequence, and then, once it recognizes specific DNA sequence and binds, it has to activate transcription.

(Refer Slide Time: 02:56)



So, the transcription factors should contain two important domains: one is called as a DNA binding domain and a transcription activation domain. Now, by studying a number of eukaryotic transcription factors, people have now found out that these two functions are actually modular in nature. That is, the transcription factors contain 2 modular domains: one is called as the DNA binding domain; another is a transcription activation domain.

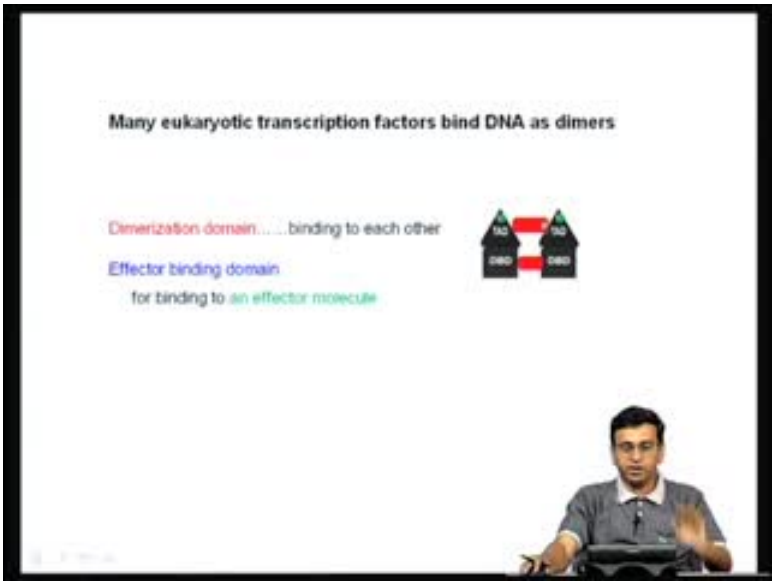
And these domains are actually separable. In fact, you can take the DNA binding domain of one transcription factor and replace it with the DNA binding domain of another transcription factor, and now, make the transcription factor recognize the DNA sequence of the other one.

So, you can take the DBD of, DBD of transcription factor a, and replace it with the DBD of b, and now, the transcription factor a will now start recognizing the… do the function of transcription factor b.

Similarly, you can swap the DNA binding domains or transcription activation domains between two different transcription factors. This knowledge has, actually, made a very important contribution for understanding function of a number of eukaryotic genes. We will discuss about this little bit later.

So, remember, for a transcription factor to function in eukaryotic cells, it should contain a DNA binding domain for interaction with specific DNA sequences, and should contain transcription activation domain, with which it can interact with proteins, which has the part of the general transcription machinery, or either directly, or with what are called as coactivators, corepressors, and so on, which we will discuss in the later stages.

(Refer Slide Time: 04:22)



The other important feature that you have to recognize when you talk about transcription factor function is that many transcription factors actually bind as dimers. Now, in the previous class, I gave you two examples: one is the glucocorticoid receptor, another is nuclear factor kappa B. Now, the glucocorticoid receptor actually binds DNA as a homodimer. That means, two monomers of glucocorticoid receptors come together and they dimerize, and then, they go and bind to the DNA, whereas because of N F kappa B, I give an example, where two different sub units, so it is heterodimer. Actually, the heterodimer now go and binds to DNA.

So, in order for the transcription factors to function, in many cases like glucocorticoid receptor N F kappa B, it is not on if they just have a DNA binding domain and transcription activation domain. They should also have a domain for dimerization, so they should have a dimerization domain for binding to each other.

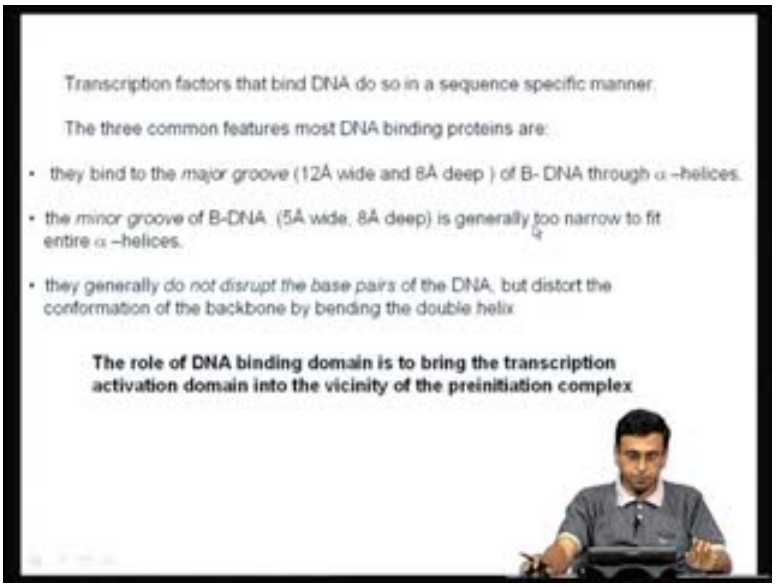
Now, this dimerization, as shown in this particular cartoon, can either facilitate dimerization of the 2 DNA binding domains, or it can be facilitate dimerization of the 2 transcription activation domains. That means, this dimerization domain can be present within the DNA binding domain or it can be present in the transcription activation domain, and the function of this dimerization domain is to bring the 2 monomers together.

It can be either 2 monomers of the same time, then it forms the homodimer; or it can be 2 different monomers; then it becomes a heterodimer. The other important thing, which again we discussed in the previous class, is many transcription factors like glucocorticoids; They…, their function is modulator by certain small molecules. These are called the effector molecules.

That means, I told you in the last class, the glucocorticoid hormone has to go and bind to the glucocorticoid receptor, and this binding of the ligand brings about a conformational change in the receptor so that it cannot interact with the Hsp90, and then it results in the exposure of the nuclear localization signal. Therefore, it can go inside the nucleus and activate transcription.

That means, the receptor also has a domain to which the small molecule can go and bind. So, the effector molecule, in this case, is glucocorticoid hormone, has to go and bind to a specific region within the protein, and therefore, this receptor, as I show in this particular cartoon, should also have a domain for binding to small effector molecules. So, in addition to the DNA binding domain and transcription activation domain, they must have a…, many of the transcribers have a domain for binding of these effector molecules, and small molecules can actually go and bind to the specific effector molecule binding domain, and bring about modulation of the transcription factor function.

(Refer Slide Time: 07:03)



So, in addition to a DNA binding domain, transcription activation domain, dimerization domain, and the effector binding domains are also very important parts of a eukaryotic transcription factor. Now, the important function, important criterion for a transcription factor function is transcription factor should bind DNA in a sequence-specific manner.

Now, there are 3 common features for most of the DNA binding proteins. You take any transcription factor– there are 3 important features you have to remember. One is, when they bind DNA, they bind the major groove of the B-DNA. The B-DNA is the normal form of DNA, which is present in the eukaryotic cells. You also have other kinds of DNA like Z-DNA, and so on and so forth.

But, we would remember, most of the transcription factors actually bind to B-form of DNA, and when they found the B-form of DNA, the B-form of DNA contains what is called as a major groove and minor groove, and most of the… these DNA-protein interactions involved in transcription function activator function, involves the major groove of the DNA. So, the alpha…, specific alpha helices of the transcription factors go and recognize specific bases in the major groove of the DNA, and as a result, the DNA-protein interaction takes place, culminating in transcription activation.

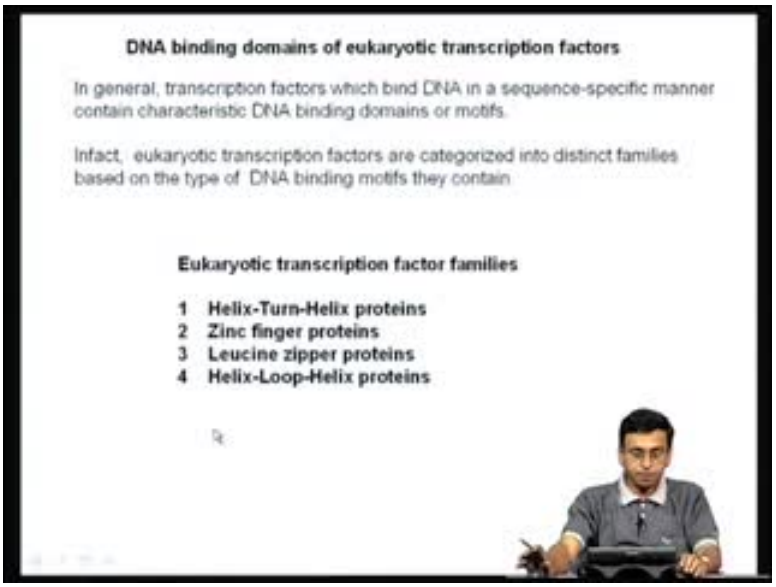
And usually, the minor groove of B-DNA is generally too narrow to fit the entire alpha helices, and therefore, most of the transcription modulator transcription functions involve major groove of the beta form, B-form of DNA.

Now, other important thing, when you talk about transcription factor binding to DNA, this does not actually involve disruption of base pairs of DNA; the DNA still remains double stranded. The base pairing of the two strands of DNA is not, in fact, is altered. So, the transcription factor binding DNA does not really alter or disrupt the base pairs of DNA, but it actually distorts the conformation of the backbone of the DNA by actually bringing about bending of the double helix.

So, this is how most of the transcription factors binding DNA and bring about transcription activation. So, the important part function of the DNA binding domain of a transcription factor is to bring the transcription activation domain to close proximity to the preinitiation complex, which contains the general transcription factors and RNA polymerase II.

So, unless a transcription factor is brought into the vicinity of the promoter, these transcription factors cannot facilitate enhancement in the rate of transcription initiation. So, the major function of the DNA binding domain of a transcription factor is to bring the transcription factor to a specific region in the promoter, in such a way that the transcription activation domain of this transcription factor can now interact with general transcription factor and bring about transcription activation.

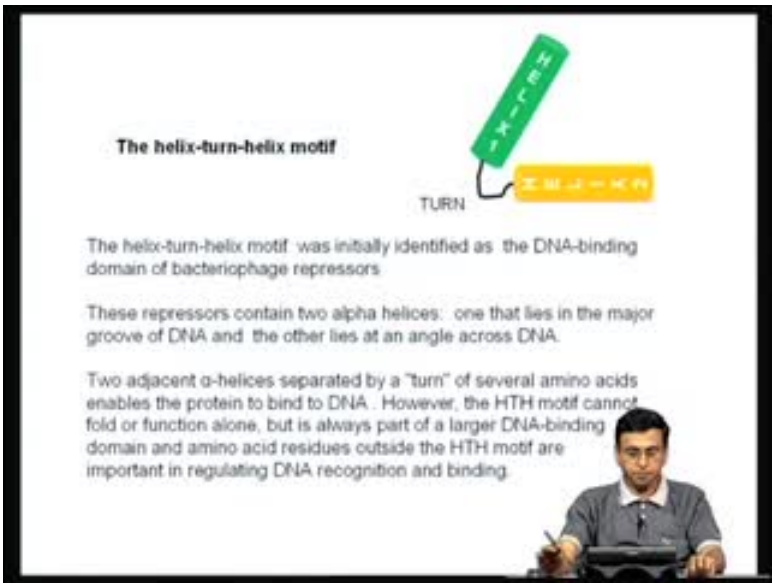
(Refer Slide Time: 09:31)



Now, we need to spend some time to understand this DNA binding property of the transcription activators, because it is based on the DNA binding domains that many transcription factors are being classified into different families. So, not all transcription factors bind DNA in the same manner; they contain very specific mechanisms, by which they can go and bind to DNA in a sequence-specific manner. So, transcription factors which bind DNA in sequence-specific manner contain characteristic DNA binding domains or DNA binding motifs.

And, in fact, based on the kind of DNA binding domain or motifs, eukaryotic transcription factors are classified into different families. So, we have, for example, what are called helix-turn-helix motifs, zinc finger proteins which contains zinc finger DNA binding motifs; we have what are called as leucine zipper motif proteins, which contain leucine zipper, which are involved in sequence-specific DNA recognition, and what are called as helix-loop-helix protein, in which the DNA recognition is brought about by what is called as the helix-loop-helix motif. There are number of other motifs, but let us confine to these 4 motifs for the present.

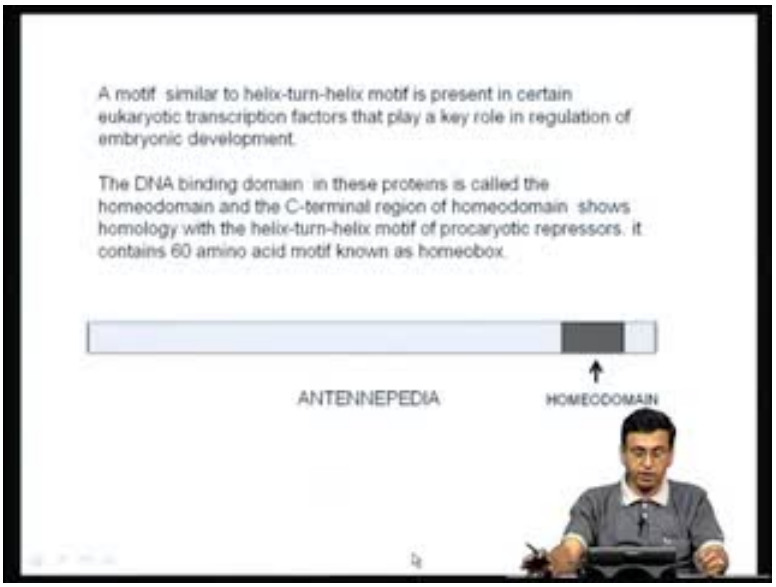
(Refer Slide Time: 10:33)



Now, the helix-turn-helix motif. Now, the helix-turn-helix motif was one of the most well studied motifs, because this motifs was initially identified in the DNA binding domain of bacteriophage repressors. So, when people started understanding how the phage actually goes and regulates the expression of the e coli cells or the bacterial cells, they found a phage actually have certain negative regulators, like the repressors. We have lambda repressor, core repressor, and so on and so forth, which actually bind to specific sequence of the host genome and shut off the expression of either genes involved in lytic phase or lysogenic phase, and these repressors actually contain a DNA recognition motif which is called as helix-turn-helix.

So, these repressor 2 alpha helices– one that lies in the major groove of the DNA, and the other that lies at a angle across the DNA. Two adjacent alpha helices separated by a turn of several amino acids enables the protein to bind DNA, so these are the helix-turn-helix motif, shown here. The helix-turn-helix motif cannot fold or function alone, but it is always part of a larger DNA binding domain, and the amino acids outside the helix-turn-helix motif are very important in regulating the DNA recognition and binding.

(Refer Slide Time: 11:52)



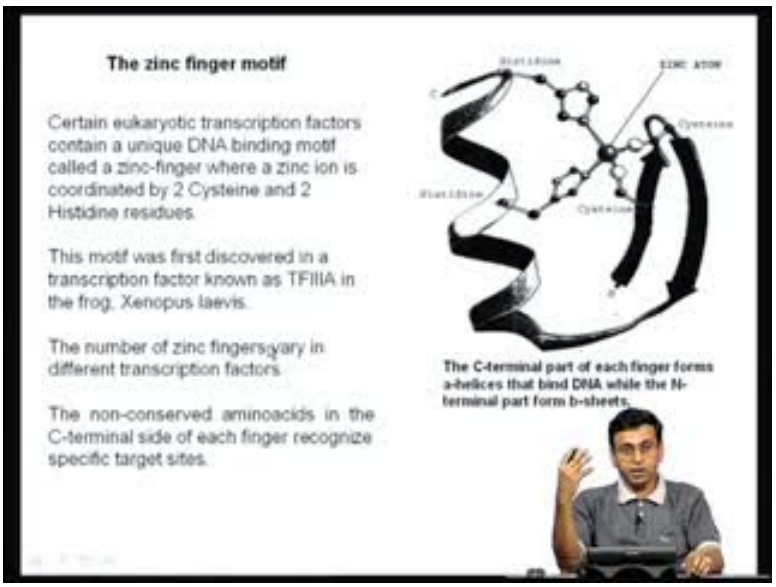
So, helix-turn-helix motif is one of the first identified DNA sequence-specific DNA recognition motifs in transcription factors, and originally found in the prokaryotic transcription regulators like the lambda repressor, core repressor, and so on and so forth. Now, a motif similar to helix-turn-helix motif is present in certain eukaryotic transcription factors, and especially those transcription factors which play a very important role in the regulation of development, and the DNA binding domain in these proteins is called as the homeodomain. So, proteins are called as hox proteins, and so on and so forth, which, again, we will discuss in detail when we talk about development regulation of gene expression.

Suffice to know now that many transcription factors in eukaryotes, which play a very important role in development regulation of gene expression, they contain a DNA recognition motif, which is very similar to the helix-turn-helix motif of prokaryotic transcription factors, and these are actually called as homeodomains.

One example what I have shown here, is that a transcription factor in Drosophila known as antennepedia, which plays a very important role in development, contains this particular helix-turn-helix turn of motif, which is actually known as the homeodomain.

So, the usually the C-terminal domain region of these transcription factors contain this homeodomain, and it contains about a 60 of these, homeodomain is comprised about 60 amino acid motif.

(Refer Slide Time: 13:07)



Now, the other important and most well studied DNA binding motif in eukaryotic transcription factor is what is called as a zinc finger motif. Now, a zinc finger motif is present in many eukaryotic transcription factors. It is called as a zinc finger because a zinc ion is coordinated between 2 cysteines and 2 histidine residues of a… the amino acid sequence of a protein.

So, for example, you have the histidine 1, histidine 2, here, and you have the cysteine 1 and cysteine 2 here. So, we have C2; we have C2H2. So, these 2 cysteine and 2 histidines actually coordinate a zinc atom, and as a result, this forms what is called as a zinc finger.

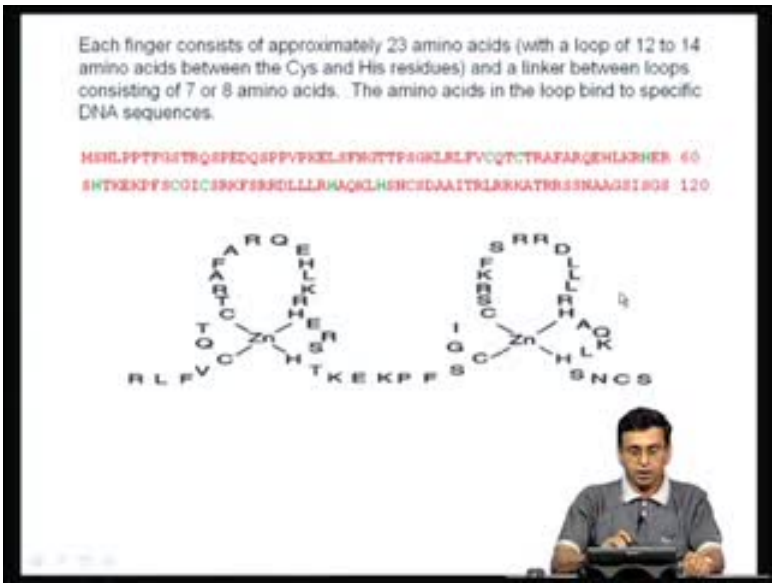
So, a zinc atom actually interacts with 2 cysteine residues and 2 histidine residues, and these are actually called as C2H2 zinc fingers. Now, the first report of such kind of a zinc finger motif being involved in a sequence-specific DNA recognition was identified in a transcription factor called as TF3A, in the frog Xenopus laevis. So, this was the first zinc finger protein which was actually discovered. In fact, the TF3A contains multiple such zinc fingers, and is involved in the recognition of the 5S RNA gene promoter.

And the number of zinc fingers in this transcription factors varies from 1 to, 1 to many; and this is exactly the cartoon actually showing, how exactly the 2 cysteines and 2 histidines actually coordinate with a zinc atom in these zinc finger transcription factors.

(Refer Slide Time: 14:34)

So a C2H2 zinc finger protein has a series of zinc fingers, and the consensus sequences of a single zinc finger is shown here; you will have a cysteine, about 2 to 4 amino acids, X means any amino acid. Again, you have a cysteine residue, and you have again 3 amino acids. Again, X means any amino acid; a phenylalanine, X3, leucine, X2, his, X 3, and his. So, I have C2H2 separated by a specific number amino acids, so the interspersed cysteine and histidine residues covalently bind zinc atoms, resulting in the folding the amino acids into loops known as zinc fingers.

(Refer Slide Time: 15:10)



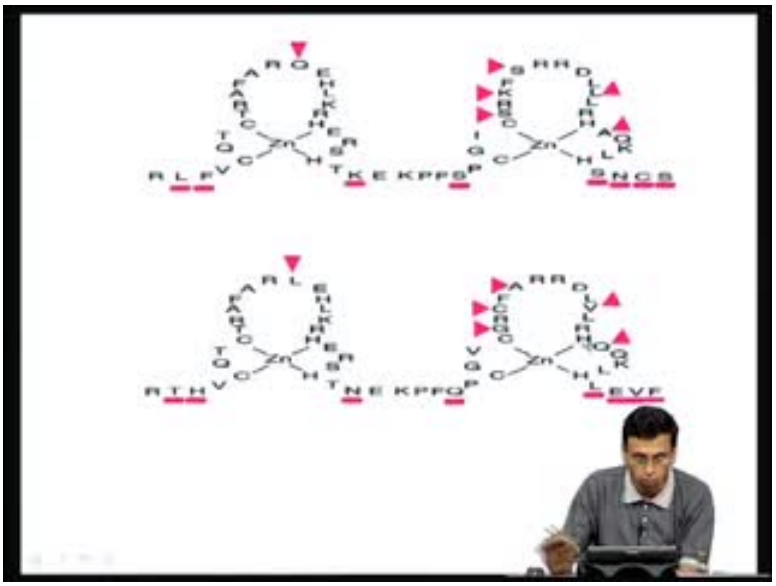
Now, let us exactly see how exactly this is formed. Here, I gave one example of a cartoon, how actually a zinc finger of a transcription factor is formed. So, as I told you, each zinc finger consists of approximately 23 amino acids, with a loop of about 12 to 14 amino acids. You can see, this is the 12 to 13, 14 amino acids which actually form a loop, and these are 2 cysteines and 2 histidines here, which actually coordinate with the zinc finger, and this is the rest of…, so the…, this is the amino terminus. This is towards the amino terminus; the protein comes here, and goes like this, and comes like this, and goes like this.

So, in this case, we have 2 C2H2 kind of a zinc fingers, so it is a C2H2 zinc finger protein. This is the actual amino acids here, so you have, actually, a linker between the loops consisting of 7 to 8 amino acids. When you have multiple zinc fingers, you have about 7 to 8 amino acids loop between the zinc fingers, and amino acids in the loop– these are the ones which are actually involved in sequence-specific DNA recognition.

So, here is an example. This is the linear amino acid sequence, and you can see, we have cysteine, 2 amino acids, cysteine, then you have a loop. This is the loop, which is formed here. Then, again, you have H2, and these are the 2 H2 residues here. So, C2, C2, H2, H2, and is the loop, and then we have a stretch of amino acids here, which actually acts as a linker. Then, you again… you have a C2, which is shown here, C2. Again, this is the

loop shown here, and you have H2, H2. So, this is an example of a C2H2 kind of a zinc finger motif.

(Refer Slide Time: 16:38)

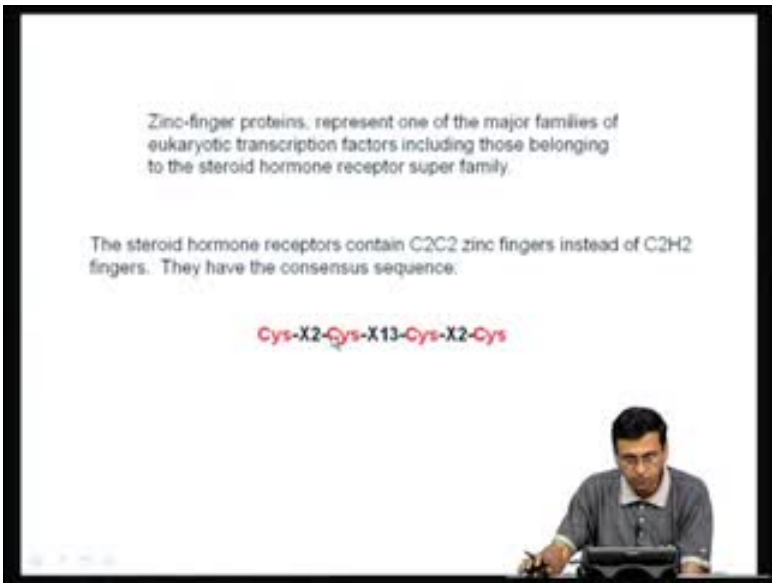


Now, today, zinc finger protein family is one of the largest transcription factors of families. So, a number of zinc finger kind of a transcription factors have been identified. So, I ask the question– how do all these things do they recognize the same sequence, or do they recognize different sequence?

Now, although they have a basic structure, in which the 2 cysteines and 2 histidines are conserved in all these zinc finger transcription factor proteins, but there are variation in the other amino acids. Here, I given, for example, 2 examples, where the underlined or the triangles actually show these are the amino acids, which are different between these two zinc finger proteins.

But, the 2 cysteines and 2 histidines are conserved, but the other amino acids in the either loop region or the linker region of these zinc finger proteins can be different, and it is these differences in these loops that actually impart differential recognition of DNA. So, these…, the…, for example, this zinc finger protein may recognize a DNA sequence, which is completely different from the sequence recognized by them, although both of them contain same kind of zinc fingers.

(Refer Slide Time: 17:49)



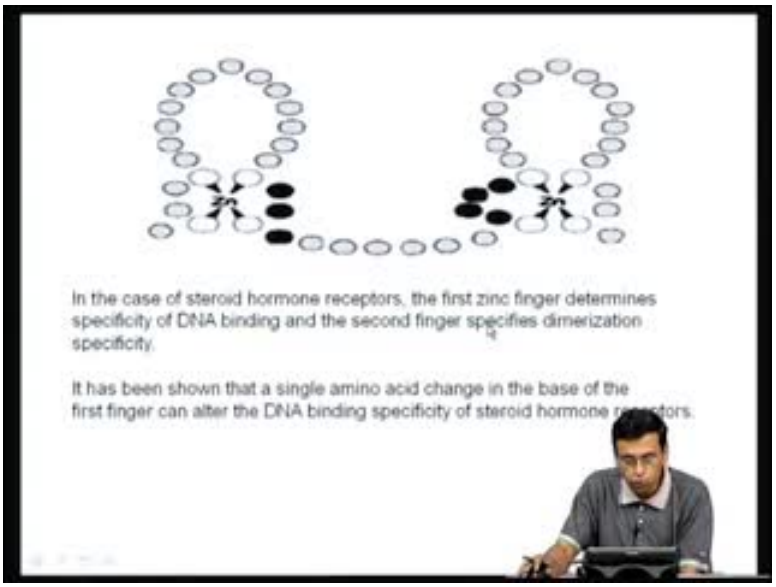
So, the amino acid differences in and around the C2H2 motif, actually, contribute to differential DNA recognition of this zinc finger proteins. So, zinc finger proteins represent one of the major families of eukaryotic transcription factors, including those belonging to the steroid hormone receptor family.

For example, in the last class, I told you, we have what is called as a glucocorticoid receptor, and glucocorticoid receptor is actually a member of the zinc finger transcription factor family. We will study in more detail how, actually, glucocorticoid receptor binds, and what is the organization of zinc fingers, little bit later.

So, one important thing you have to remember now, when we talk about steroid hormone receptors, is that in the previous slide, I told you the TF3A contains what is called as the C2H2 kind of a zinc finger, where your 2 cysteines and 2 histidines, which are involved in the coordination of the zinc finger family zinc atom. Whereas, in the case of proteins like steroid hormone receptors, which are also sequence-specific transcription factors, instead of a C2H2, we have what is called as C2C2 kind of a zinc finger.

So, the zinc atom is actually coordinated by 4 cysteines, so we have C2, loop, and again, 2 C2, and these 2 cysteines actually coordinate the zinc atom. So, this belong to the C2C2 type of zinc fingers, whereas the TFIIIAs are example of the C2H2 kind of a zinc fingers.

(Refer Slide Time: 19:13)



So, the consensus sequence for these kind of a C2C2 zinc fingers is– you have the first cysteine residues separate by 2 amino acids, the next cysteine residues, and then, you have a 13 amino acids, which actually form a loop; then you have another cysteine residue, 2 amino acids, and next cysteine residue. So, this is the consensus sequence for the C2C2 kind of a zinc finger.

Now, studies in the case of steroid hormone receptors, the zinc fingers in steroid hormone receptors have been extensively studied. Like I told you, steroid hormone means you have glucocorticoid hormone, progesterone, estrogen, and so on and so forth. Now, remember, each one of these hormones have a different physiological effect.

That means, if these receptors, but all the…, whether you take an estrogen receptor or progesterone receptor or glucocorticoid receptor, they are all zinc finger proteins. So, that means, although there are all zinc finger proteins, that genes which have to be activated by glucocorticoid receptor must be entirely different than the genes which have be have to be activated by progesterone or estrogen receptor, because each one of them varies with the different physiological response.

Now, how do these steroid hormone receptors, despite all of them containing a zinc finger, recognize different DNA sequences? Now, unless they recognize different DNA

sequences, they cannot go and bind to different promoters and activate different sets of chains.

So, how is differential gene regulation brought about by steroid hormone receptors, despite the fact that they are all zinc finger proteins? That is why I told you, as I told you in the previous slide, although they all contain the same zinc fingers, the amino acids in the base of the zinc finger are often different.

So, this amino acid difference, the base of the first zinc finger, actually, makes a glucocorticoid receptor recognize what is called as the glucocorticoid response element, whereas an estrogen receptor go and recognize the estrogen receptor response element, and there are differences between glucocorticoid response element and estrogen response element in terms of the sequence.

And therefore, because of these amino acid differences in the base of the first zinc finger, a glucocorticoid receptor will and only recognize a glucocorticoid response element, whereas an estrogen receptor will go and recognize only an estrogen response element. So, this is how differential regulation is brought about by these different members of the steroid hormone receptors. We will talk the exact mechanisms a little bit later, when we have a one entire class on regulation by nuclear hormone receptors. What has been very interestingly shown is that by simply making one amino acid change in the base of the zinc finger of glucocorticoid receptor.

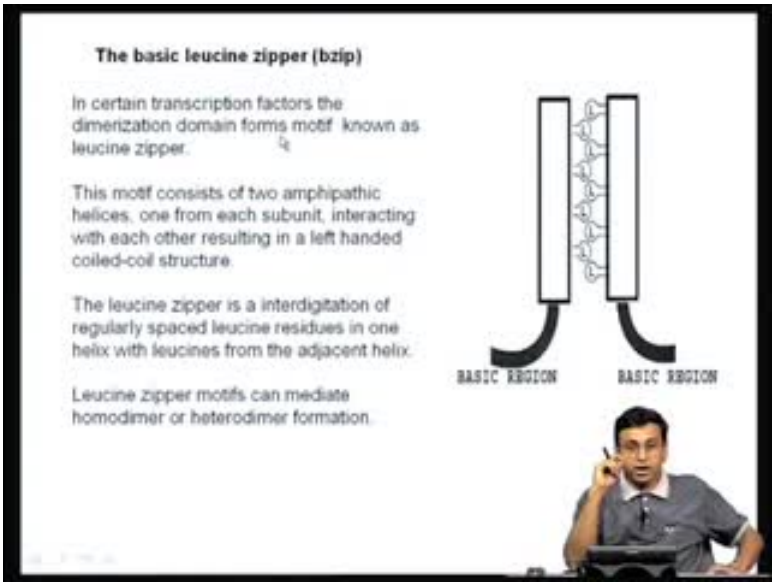
That means, if you for a…let us say, for example, this is a lysine residue, and let us say, in the estrogen receptor, instead of lysine, you have a leucine here. Now, if you now convert this lysine residue in glucocorticoid receptor into leucine residue, you can make the glucocorticoid receptor into an estrogen receptor. What I mean is, now a glucocorticoid receptor, instead of now recognizing glucocorticoid response element, will now start binding to an estrogen response element. So, the difference between whether glucocorticoid receptor has to recognize a glucocorticoid response element, or whether an estrogen receptor has to recognize an estrogen response element, is ultimately decided by a one amino acid, which is present in the base of the first zinc finger.

So, you can see, at the physiological level, you have 2 entirely different hormones. The physiological effects of a glucocorticoid hormone is to entirely different from that of

estrogen, which means, they have to activate totally different set of genes. But at the molecular level, you can see how finely the gene regulation is tuned. At the molecular level, the only difference between whether a glucocorticoid receptor has to bind a glucocorticoid response element, or whether an estrogen receptor has to bind a response element, is ultimately narrowed down to just 1 amino acid.

Say, if you change this 1 amino acid to that of estrogen receptor, now glucocorticoid receptor will now start binding to estrogen response elements, and it will be a big mess, right? We will be, in fact, that means, we will be activating genes estrogen response, which is will be activate in response of glucocorticoids, which will be a big mess in the system. So, you can see, at the molecular level, such fine difference in the amino acid sequences and nucleotide sequences make a big difference at the physiological level.

(Refer Slide Time: 22:52)



Now, the other important DNA binding motif that people have discovered in certain transcription factors, like in the case of yeast, there is a protein called Gcn4, and in the case of mammalian cell, they have what are called as c-Jun, Fos, CREB, and so on and so forth. Again, do not worry about these terminologies; we will discuss in detail in later classes.

Suffice to know that certain transcription factors in eukaryotes contain neither a helix-turn motif nor a zinc finger, but they contain a different kind of a DNA binding motif,

which is actually called as the basic leucine zipper or bzip. Now, this motif is slightly different from what we have discussed, so far, in the sense that this motif actually consists of two amphipathic helices– one from each subunit, interacting each with each with each other, resulting a left-handed coiled coil structure.

Now, these are the two amphipathic helices which I have shown here, and very interesting here is that, in which, remember, when you are talking about these things, its assume that all these transcription factors which contain these domains, they actually bind as dimers; it can be a homodimer or it can be a heterodimer, but they bind as dimer, and what, actually, this leucine zipper does is, it actually facilitates dimerization.

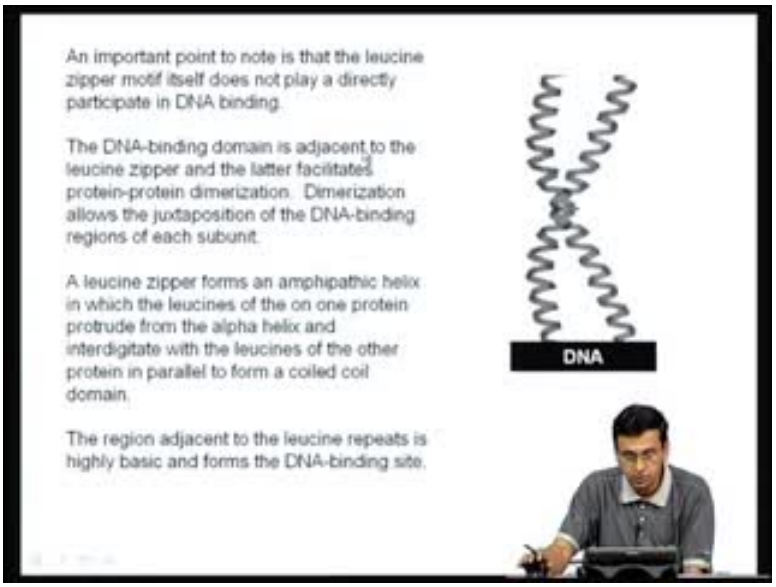
So, each 1 of these monomers have what are called as amphipathic helix, and the leucines in each one of these amphipathic helices are raised in such a way that they can actually form what is called as a zipper, that is, these leucine residues interdigitate with each other, and as a result, these 2 monomers are bought together, and the actual DNA binding is actually done by an adjacent region which is rich in basic amino acids.

So, the leucine zipper is not directly involved in DNA binding. They actually bring the 2 monomers together by the interdigitation of 2 the series of the leucine residues in these amphipathic helices, so that, the basic region can come together, and this basic region can now recognize the DNA and facilitate DNA binding.

And, very important thing is that, the leucine zippers can either facilitate homodimer formation or heterodimer formation. For example, if, in the case of some transcription factors belong to c-Jun family, there is a transcription factor called c-Jun which can either form a homodimer, or in some cases, it can also form a heterodimer with another this leucine zipper protein called c-Foss.

So, you can have a c-Jun c-Jun homodimer or you can have a c-Jun c-Foss heterodimer. So, leucine zippers can either form homodimer, or they can form heterodimer. So, this again brings in diversity, so the strength of a transcription signal response or diversity in transcription regulation by homodimer may by different from a heterodimer, and this itself can contribute to differential gene regulation.

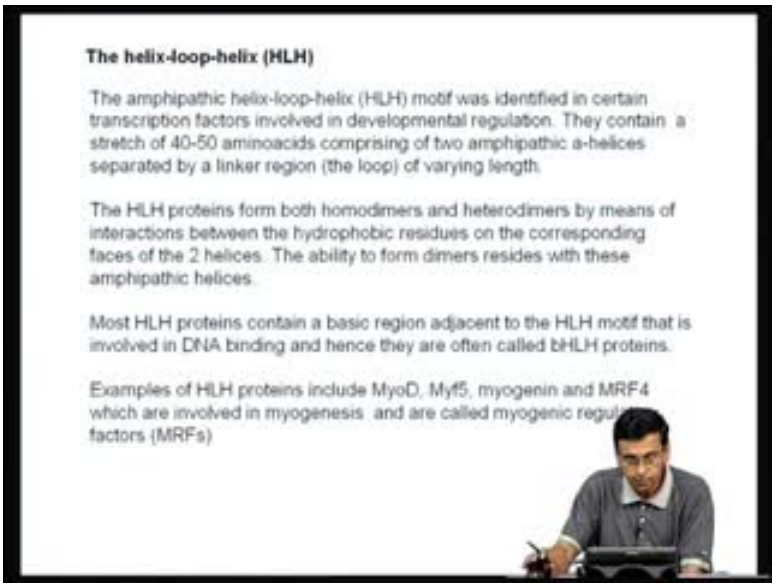
(Refer Slide Time: 25:33)



So by facilitating either homodimer formation or heterodimer formation, differential gene regulation can be brought about by the leucine zipper family of transcription factors. So, important point to note, when we talk about leucine zipper family, is that the leucine zipper motif itself direct… does not directly participate in DNA. Now, that is what I just told you, it actually facilitates in bringing the 2 monomers together, and the leucine zipper form promote dimer formation so that the adjacent basic region can go and bind DNA.

So, the DNA binding domain adjacent to the leucine zipper is the one that actually involved in the DNA binding. Now, the leucine zipper forms an amphipathic helix, in which the leucines of one protein protrude from an alpha helix, and interdigitate with the leucines of the other protein in parallel, to form a coiled coil domain, and the region adjacent to the leucine repeats, which is highly basic, is the one, actually, that forms the DNA binding site.

(Refer Slide Time: 26:25)



So, in fact, when you look at the leucine zipper protein, they actually look like scissors, as we shown here. The other important DNA binding motifs, which have been discovered in many transcription factors, is what is called as a helix-loop-helix motif. Remember, this is different from the helix-turn-helix motif, which we discussed earlier.

Now the amphipathic helix-loop-helix was actually identified in certain transcription factors involved in development regulation. Just like you found, the helix-turn-helix motif also is important in many of the homodomain-containing proteins. The helix-loop-helix also is a very important DNA binding motif, which is involved in many transform factors which play a very important role in developmental regulation of gene expression, and these transcription factors actually contain a stretch of about 40 to 50 amino acids, which compress about 2 alpha helices separated by a linker region or a loop of varying length.

Now, again, the helix-loop-helix proteins form both homodimers and heterodimers by means of interactions between hydrophobic residues on corresponding face of 2 helices, and the ability to form the heterodimers resides with these amphipathic helices.

Most helix-loop-helix proteins contain a basic region adjacent to the helix-helix motif that is involved in DNA binding, and hence, they are actually called as bHLH proteins. So, just that you have the bzip proteins, where a basic region involved in a binding is

actually brought together, or juxtaposed together, because facilitated by the dimerization of leucine zipper, here, the helix-loop-helix proteins also contain a DNA binding motif, which involve the basic region, and therefore, just like you have the bzip proteins, you have a bHLH class of proteins.

Some of the examples, these are of very important proteins of which belong to this particular family of transcription factors, are MyoD, MyF5, myogenin, MRF4, etcetera, and all these are very important transcription factor which all control, actually, muscle differentiation. So, difference during development, differentiation of cells into muscle cells, actually, involves turning on the expression of some of these important transcription factors.

And it is these transcription factors, which, actually, are important in converting these cells into cells of a particular lineage into skeletal muscle. So, many of the HLH family of proteins are involved in muscle differentiation.

(Refer Slide Time: 28:30)

Now, there is a very nice review on helix-turn-helix motif, zinc finger motifs, and leucine zipper motifs, by Kevin Struhl in Trends in Biochemical Sciences, in the April 89 issue. So, I have not given all the details of all these 3 motifs. So, I have just mentioned, given you a overview of various motifs. So, if you want to learn little bit more about how exactly all this motifs are organized, I strongly suggest you this very

nice and simple review, which was published in Trends in Biochemical Sciences or TIBS.

(Refer Slide Time: 29:03)

Now, these… Although these 4 domains which are discussed, so far, are the most important, or some of the examples of DNA recognition motifs. It do not… do not expect that all the transcription factors actually fall only into the 4 categories. As we learn more and more about transcription factors binding to DNA, more and more novel DNA binding motifs have been identified; and this just tells you there are at least about 33 different DNA binding motifs have been identified in various DNA binding proteins and transcription factors. So far, you have Homeo-box, POU domain, what is called as a bromodomain, and you have what is called as a match box proteins, which is called as a zinc ribbon proteins, RUN domain. I mean, you can just go on and on.

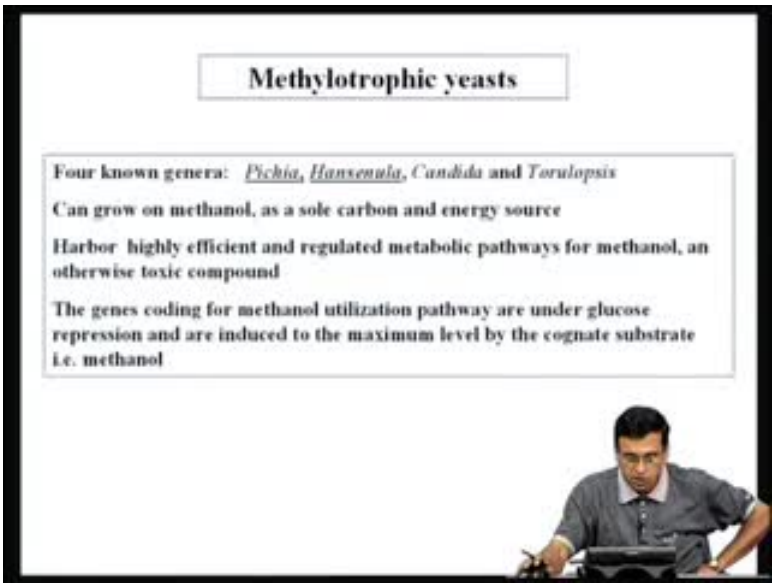
So, do not, do not think that all the transcription factors should contain only these 4 kinds of DNA recognition motifs. These are just examples what I gave you. In addition to these 4 DNA recognition motifs, there are number of other motif; they have been discovered in a number of transcription factors.

(Refer Slide Time: 30:03)

Now, what I will do in the next, maybe 10 or 15 minutes, is to give you an example– a very simple example. Now, we have studied, so far, about proximal promoter elements, distal promoter elements, core promoter elements, RNA polymerase II, general transcription factors, then upstream activators, and then, we discussed about how upstream activators go and bind to specific DNA sequences.

I thought, you will understand better, if I now discuss all these things I have told you, so far, with respect to one particular example– the simple example that I want to now do, is to discuss with you, gene regulation in yeast cells– how yeast cells respond to methanol.

(Refer Slide Time: 30:43)



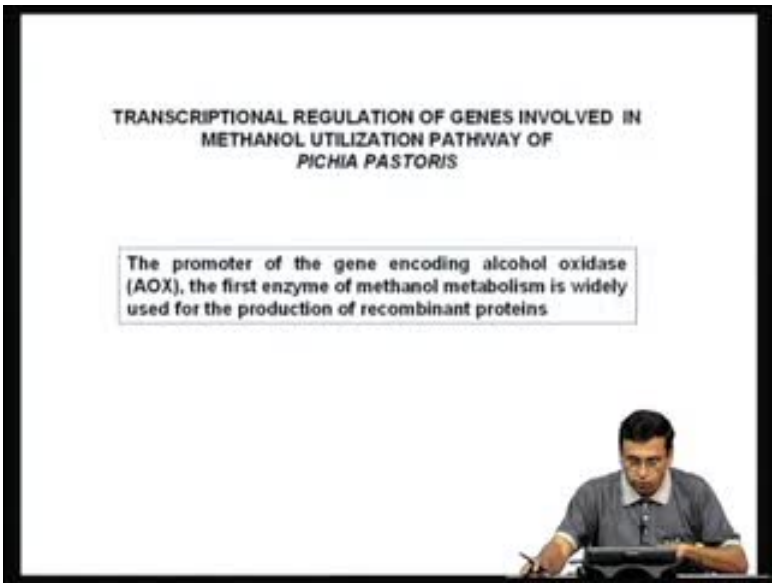
Now, normally, as yeast cells are grown using glucose as a carbon source. Yeast, like Saccharomyces cerevisiae– a baker yeast, and so on and so forth. Normally, you want to grow yeast cells, or even e coli cells, you actually put glucose, because glucose can be directly both taken to glycolysis, Kreb's cycle, electron transport, and you can generate lots of energy; and many cells systems like Saccharomyces cerevisiae, derive primarily by glycolysis.

And glucose is the most ubiquitous and most favored carbon source, not only by e coli, but also by yeast, and even higher organisms. Now, but there are certain yeasts, which are actually called as methylotrophic yeasts. Now, these methylotrophic yeasts, instead of glucose, they can also use methanol as a carbon source.

There are at least 4 such yeast belong the 4 genera– Pichia, Hansenula, Candida, and Torulopsis. So, they– Saccharomyces cerevisiae, which is the baker's yeast, it cannot use methanol as a carbon source. In fact, the cells will die.

Whereas these particular, use, in addition to glucose, can also use methanol as a carbon source. And it turns out, they can use methanol as a carbon source because they actually have a machinery for metabolizing methanol, can be convert to formaldehyde, can be convert to formic, and so on and so forth; and during this series of reactions, NAD is convert to NADH, and this NADH can then be used for generating ATP.

(Refer Slide Time: 32:29)



So, methanol can be used for efficiently deriving energy, and therefore, these yeast actually have genes which code for enzymes, which are involved in breakdown of methanol, what is called as a methanol utilization pathway, whereas Saccharomyces cerevisiae does not have it. But, what is very interesting about gene regulation is concerned, is that these genes which are involved in methanol metabolism, they are not normally expressed when you grow these yeast cells in glucose. The moment you change the carbon source from glucose to methanol, all these genes, which are involved in synthesis of enzymes involved in methanol utilization pathway, they are all turned on.

Say, very interesting to now understand how does these particular gene switch is brought about. How does methanol induce the activation of promoters of genes, which are required for methanol metabolism? Till recently, the transcription regulation, by which these genes of methanol utilization pathway of, pathway of… methylotrophic yeasts is turned on, is not very well understood.

Now, it turns out the first enzyme which is required in the methanol metabolism, is an enzyme called alcohol oxidase. The job of this alcohol oxidase is to break down methanol to formaldehyde. Now, this, the reason I am stressing this particular gene is that this alcohol oxidase promoter is widely used for making recombinant proteins. Now, in the previous class when you discussed about enhancer elements, I told you, one of the major benefits of understanding promoters and transcription factor binding is our ability

to produce recombinant proteins. You can produce recombinant protein in any system of your choice.

That is one of the major benefits out of understanding promoter-DNA interaction, promoter-protein interactions. Now, the alcohol oxidase promoter is one of the most widely used promoters for making recombinant proteins in yeast cells. Now, if I want to make large amounts of growth hormone or insulin, or hepatitis B surface antigen, all that I have to do is, I have to take the gene coding for these proteins and put in downstream of the alcohol oxidase promoter, and put this plasmid in Pichia pastoris cells, and you can grow this yeast in large amounts in glucose.

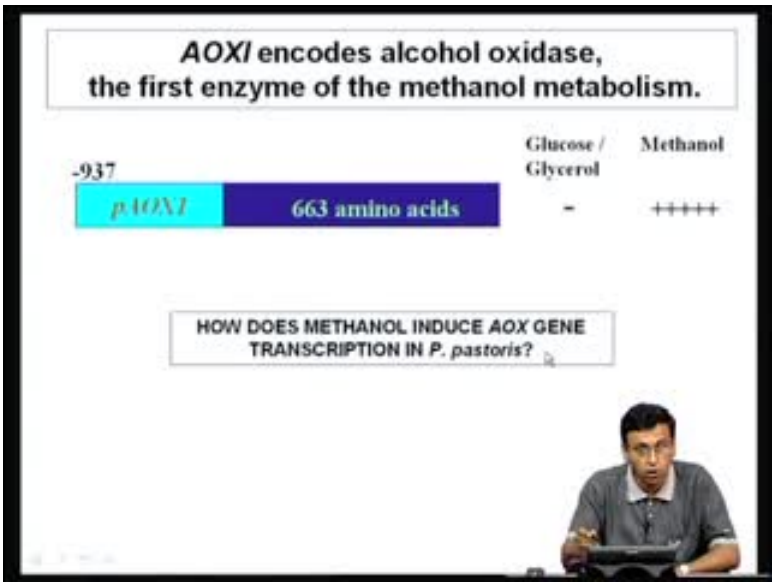
As long as you grow the cells in glucose, this promoter is not turned on. Therefore, your recombinant protein will not be made. But the moment you change the carbon source from glucose to methanol, this promoter is activated, and you get large amounts of your recombinant proteins is made. In fact, the alcohol oxidase promoter is one of the most potent eukaryotic promoters known.

Almost 30 percent of the total cellular proteins is contributed by this 1 protein, so large amounts of this alcohol oxidase is made when the carbon source is changed from glucose to methanol. But, although a large number of such recombinant proteins are being made in the… using this alcohol oxidase expression system, very little was known as to how exactly methanol is inducing the expression of these genes.

So, you can see, there are 2 situations– as long as you grow the yeast cells on glucose, the promoter is not turned on, but the moment you change the carbon source from methanol… glucose to methanol, all the genes are turned on. So, we are not again talking about one gene, that means, there are about, let us say, about 6 or 20 genes. All these genes, which are required for methanol metabolism, have to be simultaneously turned on when you change the carbon source from glucose to methanol. What is actually told you is that there must be a transcription factor just like, for example, in eukaryotic cells, when you add a glucocorticoid hormone or estrogen hormone, all the genes which have to be activated in response to the glucocorticoid hormone or estrogen hormone has to be turned on.

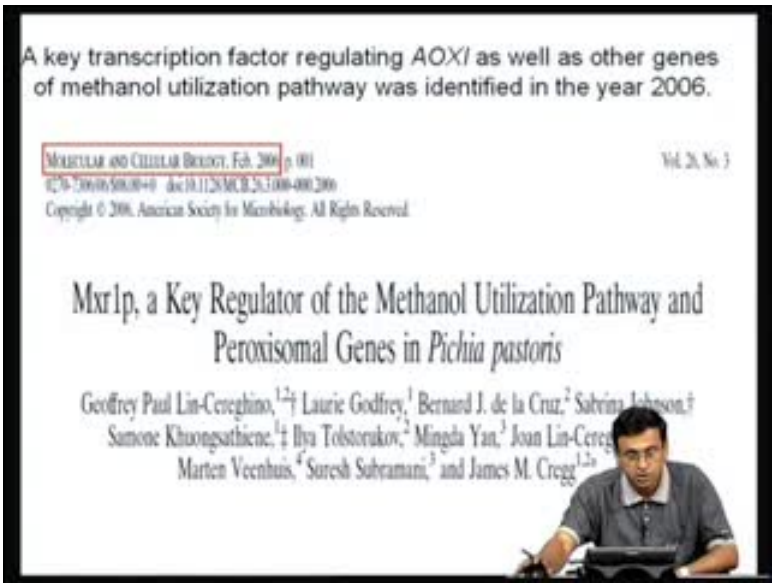
So, the estrogen receptor or glucocorticoid receptor, in response to the hormone, will now go and bind to the promoter regions of all these genes and activate the expression of all these target genes. In the same way, there must be a transcription factor, which must be going and binding to the promoters of all these genes, which are required for methanol metabolism. So, such proteins are called as master regulators; like you can actually call steroid hormone receptor as a master regulator, because it just does not regulate one gene, but it regulates the expression of number of genes involved in a specific pathway or a specific physiological response. So, in this particular case, there must be a master regulator, which must be activated when you change the carbon source from glucose to methanol. What is this master regulator was the question, which was asked.

(Refer Slide Time: 36:16)



So, as I told you, the alcohol oxidase promoter is widely used for expressing recombinant proteins, and the alcohol oxidase encodes the first enzyme in the methanol metabolism, and as long as you grow the yeast– Pichia pastoris yeast cells, in glucose or glycerol as the carbon source, this promoter is not activated; but the moment you change the carbon source to methanol, very high levels of this protein is made, and, in fact, one of these…, one of the most potent eukaryotic promoters known.

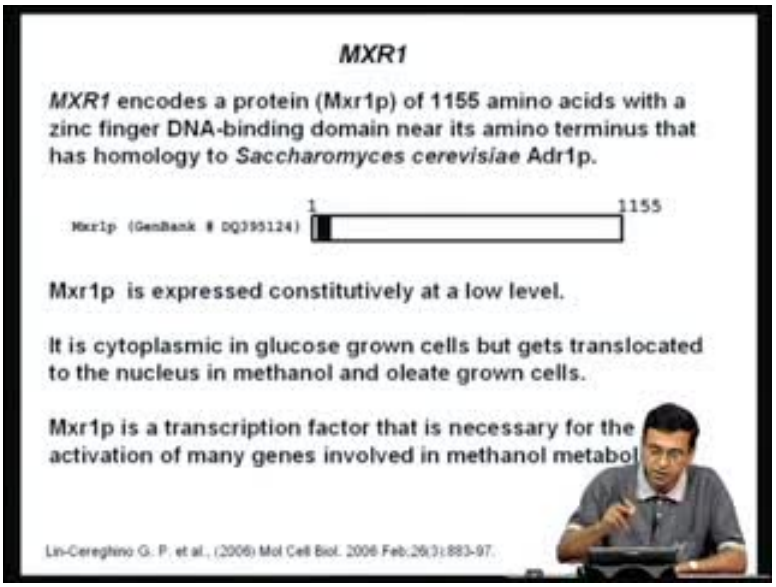
(Refer Slide Time: 36:47)



But a question that is to be asked is that, how does methanol induce the expression of alcohol oxidase gene and other genes which are involved in methanol metabolism? Now, it is only recently, in the year 2006, a group in United States actually identified a transcription factor, which they called as Mxr1p.

Now, how exactly they identified, I will not go into these details. I suggest, if you are interested, you can actually read this particular issue of Molecular and Cell Biology, where they exactly go over and describe how exactly they identified this Mxr1p, which stands for methanol expression regulator 1, how they have actually identified this as a master regulator of methanol metabolism in Pichia pastoris cells.

(Refer Slide Time: 37:19)



Now, what is interesting, what they show in this particular paper is that this particular protein is a zinc finger protein. Now, it contains a C2H2 kind of a zinc finger. There are 2 such zinc fingers, and it is a huge protein which contains about 1155 amino acids.

But the DNA binding domain, which consists of 2 zinc fingers, is present near the amino terminus of the zinc finger protein, and what they actually show is that this protein is expressed at very low levels. Whether you grow these Pichia pastoris cells either in glucose, or glycerol, or methanol, the protein is expressed.

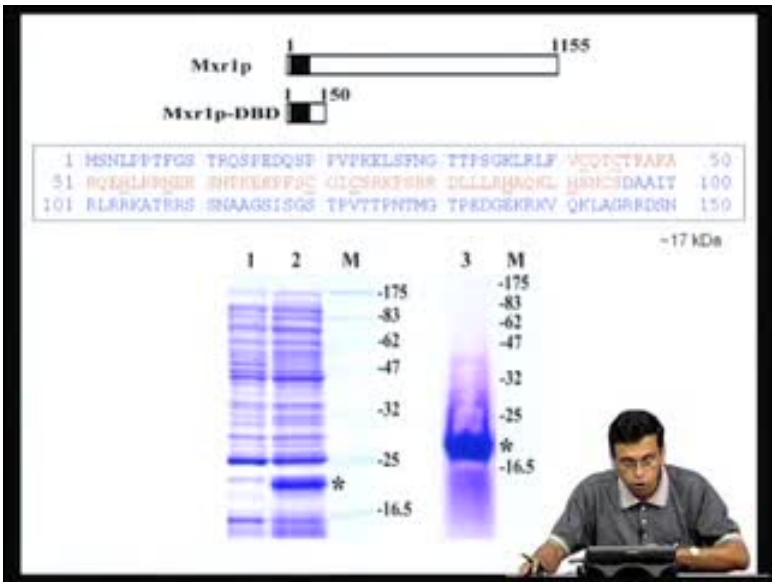
But remember, the promoters are turned on only when you change the carbon source to methanol. So, although this transcription factor is expressed all the time, it is a activating transcription of the target genes only when the carbon source changes from the glucose to methanol, and it turns out, how is this brought about? It turns out, as long as you grow this Pichia pastoris in either glucose or glycerol as a carbon source, this Mxr1p stays in the cytoplasm.

But the moment you change the carbon source from glucose to methanol, or glycerol to methanol, this Mxr1p now goes inside the nucleus, just like the steroid hormone induces translocation of the glucocorticoid receptor from cytoplasm to nucleus. Somehow, methanol is induced in the translocation of this Mxr1p from cytoplasm to nucleus, and once it goes inside the nucleus, it is now going and binding to the promoters of the all the

genes which are required for methanol metabolism, and turning on all the methanol utilization pathway.

So what they have shown in this paper is that, if you now delete the gene coding for Mxr1p, that is, if Pichia pastoris cells cannot make Mxr1p, then they cannot metabolize methanol. They cannot grow in methanol. All the genes that are required for methanol utilization pathway are not expressed by just deleting 1 transcription factor, clearly indicating that this transcription factor, actually, is the master regulator for activating all the genes of a methanol utilization pathway.

(Refer Slide Time: 39:13)



But what these particular people, in this particular paper, had not shown, is that how exactly this Mxr1p goes and binds. Now, if the Mxr1p has to function as a master regulator, it has to now go and bind to specific elements in the promoters of not only alcohol oxidase, which is the first enzyme in a methanol metabolism, it should also be binding to the promoters of various other genes which are also required for metabolizing methanol.

For example, the second enzyme– the is called as dihydroxyacetone synthase, then you have what is called as a formaldehyde dehydrogenase, formate dehydrogenase. All these enzymes are required for the metabolism of methanol.

So, you need to now identify how this Mxr1p goes and binds the promoters of not only alcohol oxidase, but also the promoter regions of all the genes. That means, we need to go back to some of the 2 or 3 classes behind; I told you that we have to now understand how actually Mxr1p binds to specific DNA sequences.

So, now let us recapitulate some of the techniques that we discussed. Now, how do you go about and study transcription, and how do you identify what kind of sequences that transcription factors goes and binds? I told you, the 2 popular experimental techniques that people use for identifying the transcription factor binding sites and promoters, is electrophoretic mobility shift assay, DNase1 footprinting, etcetera.

So, this is the techniques that we, in our own laboratory, actually, used to identify what kind of the sequences does this Mxr1p goes and binds to. Now, again I told you, Mxr1p, again, is a zinc finger protein, and just as in any transcription factor, it must be containing 2 domains– one is the DNA binding domain, which contains 2 zinc fingers, and somewhere, in the rest of the protein, there must be a transcription activation domain.

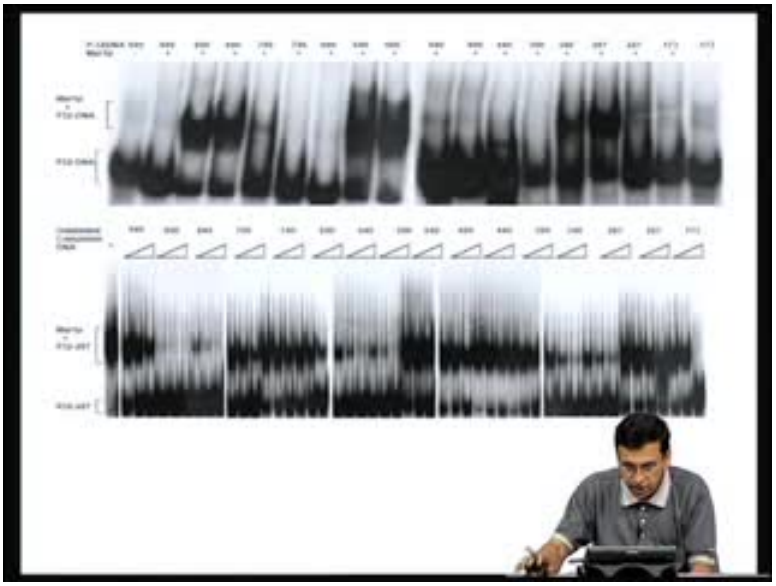
Now, as it is known for many transcription factors, these 2 domains must be separable. That means, the DNA binding domain must be separable from the transcription activation domain. So, assuming that what is shown in the other transcription factor has to be true for Mxr1p, what we did is, we took just…, because it is very difficult to express large proteins in e coli cells, because that is a huge protein of 1155 amino acids.

So, since we are, at this point, are interested only in understanding what kind of sequences Mxr1p actually recognizes, we actually took a 150 amino acid region in the Mxr1p protein, which consists of the zinc finger region, and over expressed them in e coli cells, and you can see, these are the e coli cells which are all the e coli proteins.

But, if you now induce the expression of the recombinant protein using what is called as the lactose operon, lac operon, using IPTG, we could see the Mxr1p is now made in large amounts, and the Mxr1p, actually, we have cloned it in such a way that it contains a histidine tag at the amino terminus, and therefore, we can actually purify these proteins in large amounts by passing on this entire extract through, what is called as a metal affinity column.

So, you can make a nickel agarose column, and if you pass all these proteins through nickel agarose column, only the histidine-tagged Mxr1p protein will bind to the column. Rest of the proteins will come in the flow-through, and you can actually elute this protein from column, and then you can, see, you can purify this protein.

(Refer Slide Time: 42:25)



So, now, I have a yeast transcription factor DNA binding domain in large amounts. Now, using large amounts of this Mxr1p transcription factor, now we ask the question– how does this transcription factor go and bind to specific regions of the promoter DNA? So, what I did, we did what is called as a electrophoretic mobility shift assay.

Now, the alcohol oxidase promoter is about 1000 base pairs. Now, these 1000 base pairs is what everybody is using for making a large number of recombinant proteins. So, what I did, I took, I took this alcohol oxidase promoter, because 1000 base pairs is too big a DNA fragment to do electrophoretic mobility shift assays. Usually, electrophoretic mobility shift assays are done with DNA fragments of about 50, maximum 50 to 60 base pairs; not more than that.

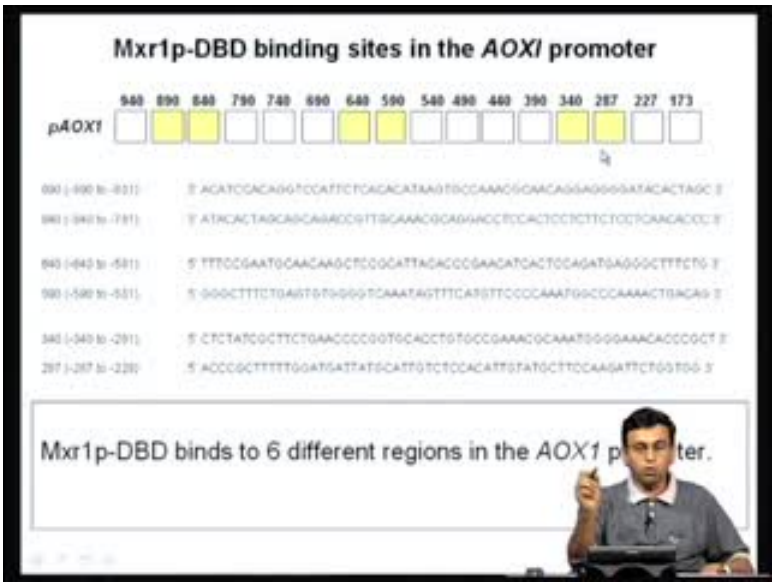
So, what we did, we took this 1000 base pair region and divided them into about 12 different 60 base pair regions, from minus 1000 to plus 1. This is what I shown here; 940 means, it has the first 940 base pair from… from the 1000 to 940, then you have 890, and so and so forth. The entire promoter region was divided into 12 different or 16 different

promoter regions, each spanning a different region of the alcohol oxidase, and we radio labeled each one of these promoter fragments, and then incubated with this Mxr1p DNA binding domain, which was made in e coli cells, and ask the question– what kind of these promoter regions within the alcohol oxxidae promoter is capable of binding Mxr1p?

And when you do this electrophoretic mobility shift assay, where you take the DNA and DNA-protein complexes and run them on a acrylamide gel, as I described my previous class, when the DNA binds to a protein, it becomes a bigger complex, and therefore, its mobility is retarded in the gel. Therefore, you get slow moving complex, and you see, that is what is happening here. Not all these promoter fragments bind Mxr1p. For example, 940, which actually contain 60 base pair region up to 940 base pairs of the promoter, does not form any protein complex. Similarly, the 890 region, which is the next adjacent region in the promoter, makes a nice complex with Mxr1p; so is the next 65 base pairs.

Whereas the next 3 different regions of AOX promoter do not interact with the Mxr1p, whereas 2 other regions interact, and so on and so forth. That means, in the entire alcohol oxidase promoter region, we have been able to identify about 2, 4, about 6 or 7 different regions where the alcohol oxidase, where the Mxr1p actually binds specifically.

(Refer Slide Time: 45:52)



That means, we have identified about 6 different regions in the alcohol oxidase promoter where Mxr1p can actually go and bind in a very specific manner. We have also now conclude, demonstrated convincingly that this binding is very specific, by reducing what is, we all say, competition. I told you in last class, you can also do a competition experiment, where we actually took, for example, you know that this 890 actually binds Mxr1p.

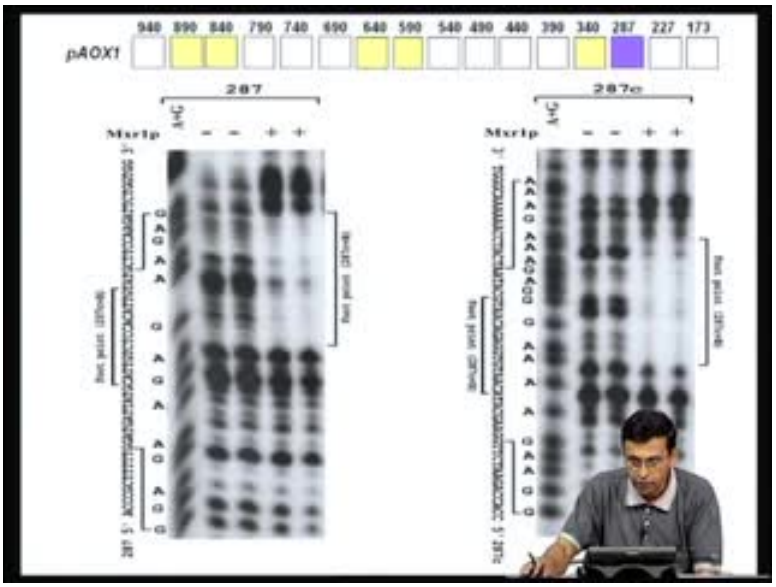
So, we took this 890, and then if you now chase the complex with protein complex with same homologous probe, it chases very well, but the heterologous probe do not chase. That is, those promoter sequences, which do not bind Mxr1p, they do not chase the protein complex.

But those DNA fragments which bind Mxr1p, when you put excess amounts of those DNA, they very effectively chase. So these kinds of experiments very clearly tell you, that what you are actually measuring is a sequence-specific DNA-protein interaction. Only when you add where you have a radio labeled fragment, when you have a homologous excess amount of the same homologous DNA, it effectively chases binding, but in a way, the excess amounts of a heterologous DNA fragment, that is a fragment, which does not bind, it does not chase.

So, by using these kinds of studies, we actually identified there are about 6 different regions in the alcohol oxidase promoter, where Mxr1p can actually bind in vitro. Now, we still do not know whether the same 6 regions are also bound by Mxr1p in vivo. This is, again we will discuss later. There are techniques called as chromatin immunoprecipitation and so on and so forth, by which you can actually identify whether the transcription factor, which is binding to a DNA fragment in vitro, can the same region be bound by the actual promoter in inside the cells in vivo. This, there are different techniques to study those things.

For the time being, remember, using a recombinant Mxr1p transcription factor DNA binding domain expressed in e coli cells, we have been able to demonstrate, that in the, in the alcohol oxidase 1 promoter, Mxr1p binds 6 different regions in the AOX1 promoter.

(Refer Slide Time: 46:37)

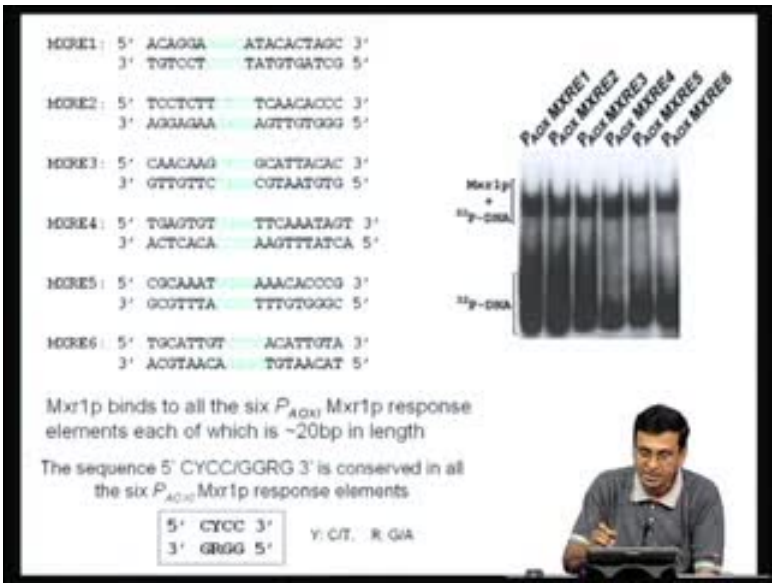


We have also then what is called as DNase1 footprinting studies, where, for example, in this case, we have taken this particular promoter region which is about 60 base pairs in length. Usually, transcription factors do not require 60 base pairs to bind. Usually, they bind about 15 to 20 base pairs; that is the usual sequence that is sufficient for a transcription factor to bind.

So, in this 16 base pair, what we have identified using EMSA, there must be a much smaller region where, actually, the protein is binding. So, to identify within the 60 base pair region where exactly the protein is binding, we carried out DNase1 footprinting studies.

For example, we have taken this particular fragment, labeled either the top strand or bottom strand, and you can see, and then carried out DNase1 footprinting studies, which I explained in my last class, and you can see in the absence of protein, DNase1 cleaves all over and generates all these band, but when incubated with protein, you can see, it has created a very nice foot print; I am clearly telling that. So, this is the region. In the entire 60 base pair region, in the 60 F 5 base pair region, Mxr1p is binding somewhere in this region, which is what is giving a foot print. This is from the top strand; this is in the bottom strand; and in both the strands, Mxr1p is able to make a contact. That means, it is recognizing, it is binding to both sides of the DNA.

(Refer Slide Time: 48:04)



So, using footprinting studies, we have actually identified, and then once we identified a foot print, we actually made 20 base pairs of this particular DNA oligo. All you go say all our earlier studies of EMSA was done with 65 base pair sequences, but now, you made only a 20 base pair oligonucleotide, double stranded oligonucleotide, and actually shown that, yes– this 20 base pair DNA is good enough for Mxr1p binding.

So, we identified about 65 base pair regions, 6 difference such regions by using EMSA studies, then by reading foot printing studies we narrowed down to about 20 base pair region. Now, we actually show, 20 base pair region in these 6 different regions can actually bind Mxr1p. We have done lots of other studies to actually show it is very specific binding, and so on and so forth.

But then, the interesting thing is, as I told you, Mxr1p, if it is activating transcription of specific genes, it must be a sequence-specific transcription factor; it cannot go and bind anywhere in the promoter. When I say it is binding to 6 different regions in the alcohol oxidase promoter, that mean, all the 6 regions must be having some common DNA elements. There must be some common recognition sequences in these 6 different regions, which the alcohol, which the Mxr1p is able to recognize.

So, once we identify 6 different regions the alcohol oxidase promoter, which the Mxr1p is binding, then we ask the question– what is common in these 6 different regions? And

interestingly, we found out, all these 6 different regions contain either a GGGG motif, or a GG, or GGGG motif, or a GG purine G motif, you can see here, they contain a CTCC or a CCCC motif, and a complementary sequence.

So, the common sequence which is present in all these sequences is either a CTCC in one of the strands, or a CCCC motif. So, what we said is that all the DNA for Mxr1p to bind to particular region the promoter, you require either a CYCC sequence, where Y is either a C or T. So, the top strand should contain CYCC, and the bottom strand should contain GGRG, where R is a purine.

So, we identified that Mxr1p requires this particular as the core sequence for it to bind. Now, I will not going into a more detail, because what we actually shown is that, in addition to the code sequence, it also requires an specific sequences flanking this particular code sequence for DNA recognition. But, so what we have shown is that Mxr1p is actually sequence-specific DNA binding proteins, and the 6 different promoter regions, which we have identified, actually, contain the same kind of sequence, which contains either a CTCC or CCCC motif must be present for Mxr1p to bind.

(Refer Slide Time: 50:23)

I will, when we are actually doing some of these experiments, there is another group, actually, which actually published a paper. They have actually deleted certain regions in the promoter region of AOX1, and then actually showed if you delete these regions, you

actually lose the ability of methanol to activate transcription from AOX1 promoter, indicating that these regions are actually important for transcription activation of the methanol-inducible activation of the AOX1 promoter.

(Refer Slide Time: 51:03)

So, now we have identified 6 different promoter regions. We now put these 2 stories together and ask the question– do the promoter regions which this group has identified in using their in vivo studies, do they actually contain binding sites for Mxr1p? In fact, it turns out, very interestingly, for example, they have actually showed, if you delete the region between minus 8 0 5 and minus 798, the promoter activity drops down only 60 percent of the normal.

And very interestingly, this region actually contains one of the Mxr1p binding sites we have identified, which you call as Mxra2. Similarly, in the nucleic acids paper they have shown, if you delete the region between minus 643 and minus 597, the promoter activity drops by 33 percent, and, in fact, what we have shown is that this region actually contains one of the Mxr1p binding sites we have identified, and so on and so forth.

So, every deletion that these people have identified seems to be containing a Mxr1p binding site. So, there is a very good correlation between the Mxr1p binding site that we have identified in vitro, and when you delete regions which contain this Mxr1p binding site, there seems to be a decrease in the promoter function, suggesting that these binding

sites actually serve as Mxra's or Mxr1p response elements in vivo. So, Mxr1p may be binding to the specific regions, and if you delete in this regions, you actually lose promoter activity.

(Refer Slide Time: 52:10)

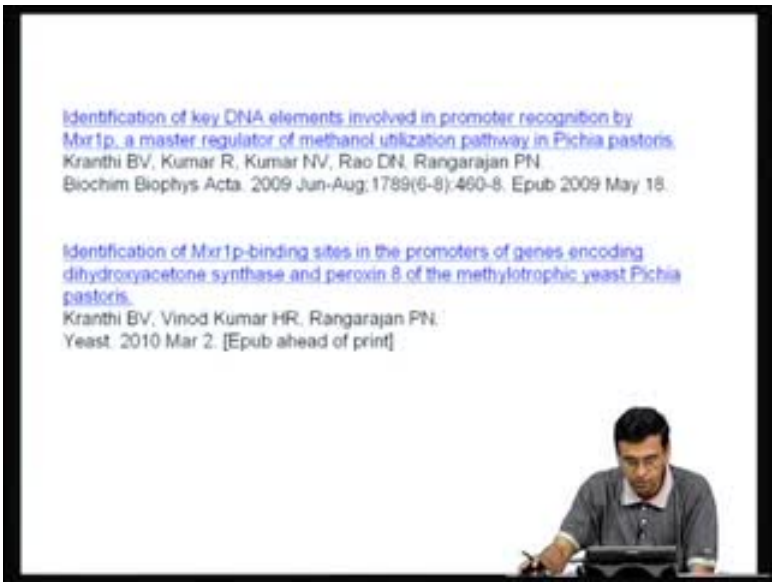
So, this is just the summary of what I told you. We have identified 6 different Mxr1p binding sites in the alcohol oxidase promoter regions, which are all shown in red, and I can see all of them containing either a GGGG or GAGG motif, or the complementary is there are CTCC or CCCC motif, which is the commons in all, the all of them. And what is very interesting is that deletion of all these promoter regions actually results in the drop in the promoter activity. For example, if you delete this particular region where the TCCC motif related, there is a decrease in the 40 percent promoter activity.

If you delete this particular region shown in the pink, there is only a 67 percent amount promoter activity, whereas if you delete this particular region, it… there is only 70 percent promoter activity is gone, and this deletion in this region requires only 60 percent decrease in promoter activity, deletion here requires decreases 58 percent promoter activity is gone.

So, clearly indicating that it is possible that the binding sites we have identified in vitro may actually be serving as binding site for Mxr1p in vivo as well. But, more conclusive experiment needs to be done to actually show that Mxr1p is actually binding to these

regions in a native state or in the actually in the nucleus of the actual cells. This how these studies are done; we will discuss little bit later.

(Refer Slide Time: 53:24)



We actually studied some of… published many of these findings. Not only we have shown that these CTCC or CCCC containing motifs are present in the alcohol oxidase promoter. These specific Mxr1p binding sites are also present in two other promoters of the methanol utilization pathway, namely, dihydroxyacetone synthase, as well as some other thing called peroxin 8. So, what we are proposing, is that, Mxr, when we change the carbon source from glucose to methanol, this Mxr1p zinc finger protein translocate for cytosol to nucleus, and then goes and binds the promoters, which must be containing these CTCC motifs.

So, at least we have shown 3 promoters which are actually activated by methanol, actually contain these kinds of Mxr1p binding sites in their promoters. So, the Mxr1p is binding as a global regulator, by actually binding these kind of a motifs in the promoter regions of the various methanol utilization pathway genes, and activating the expression of these target genes.

So, I think I will stop here, and what I have told you, so far, is that we have discussed in detail how these sequence-specific transcription factors go and bind to DNA. They contain a DNA binding domain and transcription activation domain. We will talk about

transcription activation domain in the next class, but today's class we focused primarily on the DNA binding domain. We discussed various DNA binding domains which are present in transcription factors, like the helix-turn-helix motif, zinc finger motifs, helix-loop-helix motif, leucine zipper, and so on and so forth.

And then I told you, there may be many, be many other transcription factor DNA binding motifs as well, and then I briefly gave an example, how our own research in our own laboratory, where we actually identified the binding site for transcription factor called Mxr1p in the promoter region of an alcohol oxidase 1 promoter, which is the target gene, and using techniques such as electrophoretic mobility shift assay, DNase1 foot printing, how we went about, and identified the binding sites for a specific transcription factor.

Now, one can study these kinds of studies using any eukaryotic promoter of your choice. I think I will stop here.