Comprehensive Molecular Diagnostics and Advanced Gene Expression Analysis

Prof. Aritri Bir

Dr. B.C. Roy Multi-Speciality Medical Research Centre

Indian Institute of Technology Kharagpur

Lecture 18 : Methods to study DNA-protein interaction I

Namaskar. Welcome back to the NPTEL lecture series comprehensive molecular diagnostics and advanced gene expression analysis. So, we were at our module 4, where we are discussing different tools for molecular diagnostics and gene expression analysis. Today we are going to discuss different methods by which we can study DNA and protein interaction. So, in this class we are going to learn different in vitro technique to study DNA protein interaction and in the next class we are going to discuss different in vivo techniques to study DNA protein interactions .

So, why this DNA protein interaction study is required? Basically these proteins which are interacting with the DNA these interaction are the fundamental understanding or fundamental basis of different gene regulation. So, based on the interaction of the protein it is decided whether a gene will be expressed whether it is on or will not be expressed that is off. So, this on off of the gene is basically regulated by different proteins interaction. Then there are different cellular process and function which are extensively related with DNA replication transcription translation DNA repair.

So, you already know there are multiple proteins like transcription factor, then termination factors in case of DNA repair there are different proteins like mutH, mutL which are in mismatch repair. So, these all these are proteins which interact with the DNA and their interaction decides different cellular process and function. Further in different diseases it has been seen that there is DNA protein interaction which is aberrant. So, in cancer neurodegenerative disorder genetic diseases there are multiple proteins which are aberrantly interacting with the DNA or some new aberrant proteins which are interacting with the DNA. So, these are the interactions by which we can detect different diseases.

Similarly multiple drugs can be designed which is relevant in regulating the DNA protein interaction also those drugs can competitively replace the protein and interact with the DNA. So, for that also the DNA protein interaction study is required. And lastly all the methods like gene editing via CRISPR or gene expression manipulation they all

rely over this DNA protein interaction. So, for studying or rather conducting these methods in different biotechnological studies or genetic engineering thorough knowledge of DNA protein interaction is required. So, in this lecture we are going to discuss different in vitro techniques by which we can study DNA protein interaction.

So, what are the different in vitro techniques? DNA footprinting assay, EMSA or electrophoretic mobility shift assay, south western blotting, yeast one hybrid assay and finally, phage display for DNA binding protein. So, these are the methods we are going to study or learn today. Coming to the footprinting assay. So, basically the principle of DNA footprinting assay is if there is a protein which interacts with DNA that protein bound DNA form is basically protected from degradation by different enzymes the commonest one is DNAs. So, different chemical and enzymatic digestion the protein DNA the DNA the part of the DNA which is bound to the protein that part is protected from these digestions.

So, those specific region where the protein is bound can be identified or sequenced by footprinting assay. Now, in footprinting there is the digestion of both the naked and protein bound DNA. So, what is the naked DNA? The DNA over which no protein interactions are there. So, the same DNA so, the same DNA which has no protein bound yet is our naked DNA and another the same DNA strand where different proteins are bound this is our target of interest. So, this region basically is protected from the chemical or enzymatic digestion.

Now, if we digest both of them what will happen? So, here you can see this is the control reaction, control reaction which contains only DNA without the protein interaction. So, basically both the naked and target DNA target DNA which is actually protein bound DNA they are digested similarly with the enzyme DNase remember DNase is the commonest example. So, here you can see in the control reaction the naked DNA is run which gives rise to different bands because by digestion different fragments are created of different molecular weight. Now, in case of the protein bound DNA what happens? Here you can see these region does not have any band similarly these are the very mild band you can see here. So, it indicates that this part is basically bound with the protein.

So, the segment of the DNA bound by the protein appear as an empty stretch and this is the foot print which is created by this protein bound fragment by identifying these sequence we can identify in which region the protein binds. Also by varying the concentration of DNA binding protein the binding affinity of the protein can be estimated and the minimum concentration of the protein at which the foot print will be observed can be determined. So, this is the foot printing assay. Coming to electrophoretic mobility shift assay or EMSA now electrophoretic mobility shift assay

these terms are important. So, basically there is electrophoretic separation of the protein DNA interaction or the protein RNA interaction and that electrophoretic separation is done over polyacrylamide gel or agarose gel.

So, the principle is the region of protein bound DNA that complex is heavier than the unbound DNA or naked DNA and because its molecular weight is more it will move slowly when subjected to a non denaturing gel polyacrylamide or agarose gel. And what will happen in the blot if the naked DNA reaches the gel here the protein bound DNA being heavy will create a band which is located in higher. So, basically the band is shifted upwards. So, the band here you can see is shifted upwards because of the slower run. So, it is also this EMSA also also called gel shift assay or gel retardation assay because the run of the gel the run of the sample is retarded.

Now for identification of this sequence DNA probes can be designed by radio labeling or labeling with dyes dye specific to the stained DNA and protein for identification of the site. So, what is done you can see here is the oligonucleotide the DNA which is the naked DNA and here is the cell lysate. So, the oligonucleotide DNA and the cell lysate they are incubated together. So, what we will get there will be interaction of DNA and protein this complex DNA protein complex now is incubated with specific antibodies to the protein of interest and this antibody bound protein. So, there are 2 methods where DNA is only bound with protein which is higher than the naked DNA molecular weight.

If we run the DNA protein along with the antibody it is molecular weight will be heavier. So, what will happen when we specifically target the protein with specific antibody this complex has high more more molecular weight than this DNA protein interaction. So, the shift will be more. So, this part is known as super shift means more shifting of the band. So, there are 3 kinds of band we can see for the naked DNA in the gel it will be at the lower region the DNA region which is bound with protein they will generate higher molecular weight.

So, if we identify this region of DNA we can identify in which specific sequence the protein is binding even we can identify the target protein by applying a specific antibody to the protein. So, another sample where both DNA protein and antibody all these 3 are present it will create a band higher than this. So, this is known as super shift. So, this is the basis of electrophoretic mobility shift assay and the modifications are multiple like as I told super shift assay then capillary electrophoretic mobility shift assay or CEMSA. Now here then by name it is clear that SAMSA is combined with capillary electrophoretic base capillary electrophoresis based separation.

So, basically separation and quantification of DNA protein interaction is done over uncoated capillaries with no gel matrices. Now, this separated part is now the DNA

protein part is now identified by high sensitivity laser induced fluorescence which label the DNA with fluorescent dye or sometimes there is radio leveling, but these days radio leveling are not done. Then capillary electrophoresis the principle is basically separating the analyte on the basis of their mass to charge ratio and therefore, the separation of the complex will be lie first the protein free protein, protein which is not bound with the DNA will come out then protein DNA complex will come out and lastly the DNA will come out only. So, this is the basis of elution. So, in the capillary so, this is the capillary where free protein is bound DNA protein sequence is bound also the DNA can bind.

So, when we elute this will come first the protein after that the DNA protein interaction will come then finally, the free or naked DNA will come out. Similarly, EMSA can be combined with immunodepletion followed by super shift assay. Now, what happens in case of immunodepletion the basic principle is to immunoprecipitate one specific protein by its by identifying it with the specific antibody and precipitating with protein a cephalos beads. So, what happen this in this case a specific protein is basically removed from the sample. Now, this IgA EMSA is done where multiple transcription factors are present or multiple proteins are present which can interact with a specific sequence with a single specific sequence.

So, consider there are two proteins A and B I want to check whether this B protein interact with the DNA, but in general experiments we will found a higher band of a protein bound DNA, but we do not know whether that DNA is bind binding with A or B. So, what we will do we will remove this A from the sample. So, immunodepletion of this A will be done by binding this A protein with anti A antibody attached with protein a cephalos. So, there is immunodepletion of A after that if there is any DNA protein complex then the protein has to be B fine. So, this depleted extract then analyzed by presence of protein by super shift assay if we target one specific antibody towards this B protein.

So, what we will get B anti B and this DNA. So, by identifying this we can check whether the B protein is binding with the specific region of DNA or not. So, these are the modifications of electrophoretic mobility shift assay. Coming to the next experiment that is south western blotting. Now southern blotting we have already read that is for the identification of a specific sequence over DNA and I told you that western blotting is for identification of a specific protein.

Now the protein is basically identified by a specific targeted antibody. Similarly the DNA fragment is identified by the complementary probes. Now when this experiment is done where we do not have any specific antibodies for identification of the protein. So, in those cases we use south western blotting. Now what happens in the experimental procedure we label the oligonucleotide.

So, we have labeled oligonucleotide instead of the antibodies. So, these are our probes. Now the cellular extract containing the DNA protein interaction that extract is resolved using SDS page over which this in this SDS page sample the extract is run. After the separation based on the charge to mass ratio mass to charge ratio the proteins are then transferred over the membrane. Now this membrane bound proteins are incubated with oligonucleotides labeled oligonucleotides or the probes.

So, what will happen because we cannot identify the protein because we do not have antibody we are identifying we are we are identifying the region by using complementary probes. And finally, the membrane is photographed only by the band corresponding to bound oligos will appear in the final picture because others will not bound with the complementary probe. So, we can identify that specific protein DNA complex via this probes. Now what is the advantage we can use different type of oligonucleotides different sequence can be identified if we use different label oligos in the same blot. So, various type of sequence which are interacting with different proteins can be identified in a same gel in southwestern blotting.

Then another method yeast 1 hybrid assay. Now this is one modification of yeast 2 hybrid assay which is basically applied where we need to study protein-protein interaction. Now when we identify different protein-protein interaction it has the basic principle like most eukaryotic transcription factors they have 2 domains 2 different domains over the same protein transcription factors are protein you know. So, those 2 domains are one is activation domain represented by AD another is DNA binding domain represented by DB and these 2 domains are physically separable and if separated the proteins become functionally inactive. So, basically if these 2 regions are separated suppose a protein is having one AD region and one DB region.

So, either one of these domain if separated then the residual protein becomes inactive and it cannot activate the RNA polymerase to the promoter site which is the function of the transcription factor it recruits the RNA polymerase to the promoter site. So, that the transcription can be initiated. So, when either of these 2 domain is separated RNA polymerase cannot perform to start the transcription. Now in these 2 physically separated domain 2 different proteins are attached or fused. So, consider in AD or activation domain a protein X is bound this is known as prey prey segment and in the sorry this is DB I am sorry .

So, in this DB region we are attaching another protein and this is known as bet. Now both of these are expressed in yeast now if these 2 proteins comes closer these 2 proteins come closer and interacts what will happen these 2 region will also come close they will join rejoin and will form the functionally active transcription factor. So, what will be the

effect RNA polymerase now can act and can start the transcription. So, what we will get downstream expression of the gene that reports that the protein is expressed.

So, this is known as reporter gene. So, this is yeast 2 hybrid model where the 2 segments or 2 different proteins are expressed over the yeast 1 hybrid assay is. So, the yeast 1 hybrid assay is just the modification because it is 1 hybrid assay because 1 protein only expressed. So, instead of having DNA binding domain of the protein the DNA binding domain is basically located over the DNA itself. So, this is our DNA binding domain which we already have in the oligonucleotide. So, we need to express the protein x suppose with the activation domain.

Now if this activation domain comes closer or in proximity with the DNA binding domain located over the oligonucleotide and they can interact what will happen on fusion of these 2 successful activation of the RNA polymerase will happen and what will be there the successful expression of the reporter gene. So, this is the signal that yes these 2 molecule 1 is protein and another is DNA they are interacting for which the reporter gene is expressed. Now this study has 1 very important advantage that we are detecting this DNA protein interaction in their native configuration. So, the sensitivity is very high in this condition. Next we are moving towards another method that is Fudge display.

Fudge display for DNA binding proteins the Fudge display basically refers to the method of expressing a peptide or protein domain on a bacteriophage capsid by genetically fusing its amino acid sequence to that of the coat protein coat protein of the fudge encoded by the fudge. So, these 2 are basically fused. So, what happen when it is fused with the the I mean the peptide protein domain when it is fused with the fudge coat proteins when there is expression of the coat protein there will be expression of this peptide or protein. So, we can use multiple types of protein domain over different fudges and can form a Fudge display library and from that library we can purify the specific target protein sequence protein domain by affinity purification using appropriate ligand. Now the probe for the affinity purification is basically the dsDNA oligo.

So, the ligand here is the dsDNA oligo which has the DNA binding domain over it. So, consider there are multiple ists which are expressing different protein like this is protein A, this is protein B, this is protein C and we have a plate or a microtiter plate or solid matrix over which different dsDNA are attached. Now if amongst this protein this B protein interacts with the DNA what will happen when we incubate this mixture with the microtiter plate only the B protein will bind here others will be free and when we wash it the unbound proteins like A, C it will be removed. So, what we have in our hand is the B protein which is bound to the dsDNA oligo ligand or probes. So, basically fudge that display the protein which can interact with the dsDNA oligo will remain bound to the

matrix.

After that the remaining part will be eluted and then more fudge will be produced by bacterial infection with helper fudge and finally, the clones will be identified by sequencing the fudge genome. So, all these are the different methods by which we can study different in vitro technique to study DNA protein interaction. To sum up we read footprinting assay where based on the principle that the protein which is binding with the DNA the segment of protein bound part of the DNA will be protected from the different chemical and enzymatic digestion and based on that if we will get some footprint over the gel electrophoresis. Then EMSA or electrophoretic mobility shift assay where electrophoretic separation of protein DNA or protein RNA complex can be done over different types of gel electrophoresis where the protein bound DNA fragment will run or will migrate in slow speed and will produce a band in a higher location. Then south western blotting which is a complex or combination of southern and western blotting where instead of identifying the protein because the proteins specific antibody is not known or not in hand we are identifying we are identifying the DNA segment by oligonucleotide probe.

Then yeast one hybrid assay where we are studying the protein and DNA interaction by exploiting the activation of repotarginine yeast and finally, fudge display for DNA binding protein where fudge display involving expression of peptide or protein in a bacteriophage capsid and finally, purifying them by affinity purification using the probes of dsDNA oligos and subsequent identification through fudge genome sequencing. So, these are all of in vitro techniques in the next class we will learn of different in vivo technique these are my references. Thank you all and see you in the next class.