Course Name: Basics of Crop Breeding and Plant Biotechnology

Professor Name: Dr. Joydeep Banerjee

Department Name: Agricultural and Food Engineering

Institute Name: Indian Institute of Technology Kharagpur

Week: 09

Lecture-39: Promoters and Preparation of Overexpression Construct

Welcome back. So, we will continue again. Now, we will start our discussion on Promoters. So, what is promoter? Earlier, I have mentioned that promoters are the upstream regulatory region of a particular gene. So, if you think about bacterial promoter, in bacteria also different genes are there, in bacterial promoter, it has been found that suppose from this part the transcription is starting. So, from which particular base transcription is started that is known as TSS or Transcription Start Site, Transcription Start Site.

The first base wherefrom transcription is starting and before that the promoter region is available. The base prior to TSS is known as -1, there before that -2 in this way, if we go upstream different negative sequences are available, upstream sequences are mentioned by minus sign; - 1, - 2. In this way, in bacterial promoter, some regions are available close to -10 region or - 35 region, they play crucial role in bacterial gene expression. In some bacterial promoter along with this - 10 and - 35 region, some other upstream regions are also available, upstream regulatory region.

But if you think about the eukaryotic promoter, in eukaryotic promoter in case of eukaryotes, the gene regulation is much more complicated than the prokaryotic system. So, let us see how does it look like. So, suppose, this is the chromosomal DNA, here a particular gene transcription is started from here, that is the transcription start site. So, just before that the promoter region is supposed to be started. So, in this way - 1, - 2 different regions are there, it has been found that in most of the eukaryotic promoter

within the first 200 to 300 base that is in between - 200, - 1 … - 200 or - 300 within this region different TATA box binding elements are available.

And some sequences are available on the DNA, it is sequence is close to TAATT or TATATT. So, these sequences are part of TATA box, means, in those sequences some TATA box binding factors kind of protein binds. So, before that some other regions are also available wherever the CAAT box binding elements bind. So, this TATA box, CAAT box are mostly found in the eukaryotic promoter. And in addition to that, the RNA polymerase, the major thing which initiate the transcription, CAAT binds and the binding or efficacy of RNA polymerase will be determined due to the presence of this TATA box element, this CAAT box element etc.

So, although RNA polymerase is the major enzyme which cause the transcription in eukaryotes. So, different factors are also involved in the process like TATA box binding factors, CAAT box binding factors and other upstream regulatory regions. So, those are known as *cis* regulatory region or upstream regulatory region. So, what is available over there? In those upstream regulatory region or *cis* regulatory region, different small sequences are available. And on those specific sequences different transcription factor binds, different transcription factor binds.

So, different types of transcription factors are there like bZIP transcription factor, bHLH transcription factor, then CAMTA transcription factor I have mentioned earlier, then and WRKY transcription factor. So, these different transcription factors are available in eukaryotic system and they control various kind of stress regulation or hormonal regulation. Ok! So, for each and every transcription factors, some specific sequences are available on the upstream DNA region. If that sequence is available, then only that particular transcription factor can bind and once it will bind, the overall structure will be like this. Suppose, this is the RNA polymerase binding here, the TATA box binding element has bound, their CAAT box have bound and other transcription factors which are supposed to bind the DNA they will fold in this way and they interact with the RNA polymerase also and in this way, they enhance the transcription process.

Similarly, some sequences are also there in the promoter region that cause the retardation in the transcription process. So, whatever we got from here, promoter region is the upstream region of a particular gene, where some activator sequence might be there, some repressor sequence might be there. In activator sequence, some favorable transcription factors will bind that will enhance the transcription process, while in the repressor region, some other types of transcription factors bind that will try to repress the process. Because, suppose a plant is growing under control condition, while another plant is growing under drought condition. The plant which is growing under control condition, it does not need to express all the genes which are responsible for drought tolerance.

So, for drought tolerance, some genes are specifically expressed that is DREB genes or DREB transcription factors, then CBP kind of proteins, those genes are expressed under drought stress condition. In control, their expression is not at all needed. So, their expression is not up-regulated or retarded by some factors, while, if the plant encounter drought stress, the expression of those genes will be up-regulated due to availability of some transcription factors. Ok! So, in this way, promoter is a big part of the DNA which is available in the upstream part of a gene and several activators, repressor molecules might be there. And through deletion analysis or either from the 5′ end or 3′ end, suppose it is a transcription factor of a gene this is the 2kb region.

In this 2 kb region, this is known as the proximal part of the promoter. This is known as the distal part of the promoter and this is the median part of the promoter. In this way, we can characterize a promoter in the proximal part, which type of transcription factor binding sites are available we can analyze in the distal part, which types of transcription factors are available we can analyze how by making different deletion constants. Suppose this one was the promoter of 2kb. So, we can make different deletion constructs this is the 5′ to 3′, 5 ′ to 3′, 5′ to 3′.

Over here we have done 3′ deletion. Similarly, we can delete from this side also we can

cause this is the 5′ to 3′, 3′, 5′ to 3′. Over here, this part is common which that is the proximal part is common in all of them, but the distal part has been deleted. So, here we have done the 5′ deletion while in the earlier case, we had done the 3′ deletion, here the distal part is common. In this way, by making different deletion constructs or by making internal deletion constructs, we can amplify the distal part, we can amplify the proximal part and we can seal it by without taking the median region.

In this way also, we can make different deletion constructs of the promoter to characterize in which part important transcription factors are available. So, now we start our discussion on preparation of overexpression construct. So, for overexpression of a particular gene we need to have a promoter and a terminator in our shuttle vector. So, what is the objective? The objective is to make the overexpression construct so that we can over express a particular gene in heterologous system. Suppose, we have cloned a particular gene from rice, suppose *OsGLP1* has been cloned from rice.

Now, we need to express it in tobacco system to characterize that gene whether that gene is showing any type of stress tolerance or not. Because in tobacco, this gene is not available *OsGLP1* is not at all available if we over express it from here, new sets of protein will be produced in the heterologous system and we can see its features. If it is a disease resistance gene, once the transformed tobacco plants will be infected with that particular disease we can see some sort of resistance. So, in this way, for functional characterization of a gene we have to take two different approaches one is through overexpression strategy, another one is gene silencing strategy. So, let us start our discussion on overexpression strategy.

Suppose, we have already cloned our gene, the gene cloning has been discussed earlier. Suppose we have cloned our desirable gene or the coding DNA sequences, coding DNA sequence can be cloned if we have to start from the mRNA, from mRNA we have to make cDNA then we have to design the primer and we can clone the coding DNA sequences. And suppose it has been cloned in the restriction site *Eco*RI and *Sma*I and initial clone has been done in pUC18 vector. Now for preparation of overexpression

construct we need the promoter sequence also and we need the terminator sequence also. So, for expression in tobacco system, in most of the cases various scientific communities various research lab they try to express the target gene under CaMV 35S promoter.

So, we will take the example of CaMV 35S promoter, suppose in this way, we have cloned the CaMV 35S promoter also. The promoter has been cloned, let us assume in *Hin*dIII and *Eco*RI site. While, we have also cloned the terminator region, the nos terminator region suppose we have cloned and the nos terminator cloning has been done in *Sma*I and *Kpn*I site. So, then, how can we make the overexpression construct? For making any type of construct, first we will try to play with the smaller vector. We can try to make the initial construct in the pUC18 or blue script-based vector and once the construct is prepared, then it will be transferred into the binary vector for actual plant transformation.

So, if we think that we have a target vector that is pUC based vector at it is MCS, multiple cloning sites different restriction sites will be available. Suppose over here, the *Hind*III, then *Bam*HI, then *Eco*RI, then *Sma*I and *Kpn*I these sites are available this is our cloning vector we will try to make our clone initially in this vector. So, we have to digest the gene part, from here we have to digest this vector with *Eco*RI and *Sma*I and we will ligate both the things to put our gene over here. Once this part is ready, then we have to digest the promoter which has been already cloned in a vector, we can digest it from here using *Hin*dIII and *Eco*RI site and we can put the promoter over there. And then, we can put the terminator also by digesting it with *Sma*I and *Kpn*I and over here, we can clone the terminator.

So, finally, our vector will be like this, where we will be having the promoter is cloned in *Hin*dIII and *Eco*RI, then we will be having our gene, over here gene will be placed in *Eco*RI and *Sma*I and at last, we will be having the terminator region in *Sma*I and *Kpn*I. So, this will be the structure of our overexpression construct in the small cloning vector. So, for each and every cloning we have to digest the insert we have to digest the vector. Once a particular part is cloned, then we will use this backbone to clone the next part in

this way, eventually we can make this whole construct. And once this overexpression construct will be ready in small vector then we have to transfer it in the binary vector like pCAMBIA 1301.

It is a binary vector, I have mentioned earlier in this binary vector the left border and right border region is available it has origin of replication. Then within this left border and right border region, the plant selectable marker gene, hygromycin selectable marker gene is available that is also expressed under a promoter and a terminator is also there. It will be having MCS, region multiple cloning site region in that MCS the *Hin*dIII and *Kpn* site will be there. So, we can just cut this vector at *Hin*dIII and *Kpn* site and this part could be cloned over there. In the pCAMBIA 1301, a GUS gene is also available under CaMV 35S promoter.

So, once this binary construct will be ready, we have to do the *Agrobacterium* transformation, means within the *Agrobacterium* competent cells, different strains are used generally EHA 105 or LBA 4404. These strains are very much popular EHA 105 is mostly used for rice transformation, LBA 4404 is mostly used for tobacco transformation. So, once this binary vector is ready where we have put our overexpression construct, then first we need to transfer this vector into the competent *Agrobacterium* cell. Then those agro cells harboring the recombinant vector will be used for plant transformation. So, this is the process through which we can overexpress a particular gene, an overexpression construct can be made and the overall process will be performed. Thank you.