DISCUSSION

The primary objective of the present work is to gather a better understanding of the mechanism of expression of the reporter gene, GUS fused to At2S3 promoter and to determine how specific is the At2S3 promoter using *Agrobacterium*, yeast and plant systems. It has been found that GUS gene product under the same promoter is highly expressed in the seed as has been determined fluorometrically as well as by Northern blot analysis (Table 1.1, Fig. 1.7). There is also no associated problems of instability and degradation in case of GUS gene product.

The analysis of seed protein genes in transgenic plant has already identified elements (CIS elements) in the 5' flanking regions of seed genes which are postulated to play an important role in the activation of transcription at the appropriate time in the specific tissue during seed development. The activation is primarily due to the binding at these regions (CIS elements) of transacting nuclear proteins. Recently, the promoter analysis as a measure of gene expression has been questioned (Sieburth and Meyerowitz, 1997). Because these authors demonstrated that sequences controlling the expression of *Arabidopsis* floral meristem and organ identity gene (AG) lie within the transcribed region of the gene and not located exclusively in the AG 5' untranscribed region. In the present case the involvement of such elements in the expression of 2S3 gene fused with At2S3 promoter has been ruled out from the earlier observation on the 2S3 gene expressed in yeast (Pal and Biswas, 1995).

In general, the 5' flanking sequences of genes appear to be divided into two regions, 1) controlling the tissue specific temporal and spatial regulation of a given gene, and 2) an upstream region to increase the transcriptional activity of the entire gene. The identification of the genes which code for the nuclear binding protein(s) that interacts with the 5' control sequences, and the determination of the structure and regulation is expected to give a greater understanding of gene regulation in developing seeds.

In order to assess the minimal sequence elements to have significant and sufficient expression of GUS gene, 0.94 Kb of the total 1.5 Kb At2S3 DNA starting from ATG has been taken and a vector containing this 0.94Kb At2S3 promoter fragment has been found to be expressed almost to the same extent (90%) in tobacco plant as compared to that of full length 1.5 Kb At2S3 promoter as evidenced from Northern blot analysis (Fig. 1.11).

At this stage, the involvement of transacting factor(s) which might be the key regulator(s) in this system was contemplated. In order to make a preliminary idea about the nature of the polypeptide(s) if at all binding with DNA fragment of interest, i.e. At2S3p, Southwestern analysis was performed (Miskimins et al., 1985). What appeared in Southwestern experiment, looking at the lane containing nuclear extract of flowering stages of the plant (Fig. 1.12) is that a number of factors that bind with the At2S3p are present. But attention has been given only to the lowest mol. wt. polypeptide viz. ~16kDa factor as this appeared to be the common factor, detected in the nuclear extract of yeast as well as in the cell extract of *Agrobacterium* (Fig. 2.9). The

appearence of ~16kDa factor in all the three systems also lends credence to our earlier observation that indeed, this At2S3 promoter is recognized by *S. cerevisiae* (Pal and Biswas, 1995). Thus the recognition of a seed specific plant promoter by both yeast and *Agrobacterium* through a common transacting factor(s) might be very interesting and unique. At the same time our earlier observation also prompted us performing the experiments to delve deep into the problem and to answer pertinent questions pertaining to this recognition event. It has been shown that At2S3 promoter is active only in the developing phase of seeds with very little or no expression of GUS gene under this At2S3 promoter in leaves. But interestingly, abscisic acid (ABA) treatment of the leaves made the At2S3 promoter active in the leaves. It is known that dehydration of plants or ABA treatment can induce new proteins (Galan et al., 1986) and during seed formation dehydration and ABA accummulation are significant. Therefore, it was reasoned that if 16kDa protein be induced in pMP17 transformed tobacco plant by the treatment of leaves with ABA, is it possible to show the expression of transgene in the leaf? That is what has been demonstrated and is evident from the expression of the GUS gene in the leaves of transformed tobacco plants treated with ABA (Table 1.2). When Southwestern blot was performed taking crude nuclear extract of leaves of tobacco plant pretreated with ABA, similar factor of molecular wt. in the region of 16kDa appeared (Fig. 1.14). Polyethylene Glycol(5%) also produced the same effect (Fig. 1.14). The appearance of this 16kDa factor in the leaves pretreated with ABA is

concomitant with the corresponding mRNA transcript of the GUS gene in the leaves in Northern blot analysis using GUS gene as a probe (Fig. 1.15). The synchronised appearence of GUS gene transcript in the leaves pretreated with ABA is a pointer to the fact that most probably the ~16KDa factor has made At2S3 promoter active in the leaves. Again, the appearance of this ~16KDa protein in the leaves is due to it's synthesis is clarified by the fact that prior treatment of the leaves with cycloheximide before treatment with ABA failed to produce this ~16KDa protein in the transformed tobacco plant (Fig. 1.14). Under this condition, no GUS activity was detected in the leaves also. This observation rather raises the question about tissue specificity of At2S3 promoter. The tissue nonspecificity of At2S3 promoter has also been demonstrated by the observation that the GUS gene under this promoter is expressed in *Agrobacterium* and yeast in a constitutive manner (Table 2.1). The expression of GUS gene in the *Agrobacterium* has been found to be comparable to that in tobacco. In the yeast, however, it was 30% of what has been recorded in the case of tobacco (Table 2.1) suggesting the participation of other factor(s) necessary for effective expression in the yeast. In order to assess which part of the promoter is essential for the expression of GUS gene, the 0.94 Kb At2S3 promoter starting from ATG of GUS gene has been found to retrieve almost 90% and 80% of the activity of the full length 1.5 Kb At2S3 in case of *Agrobacterium* and yeast respectively (Fig. 2.7 and 2.6, Table 2.1), indicating that 0.94Kb pAt2S3 contains the characteristic features and sequences of this promoter to remain functional

almost to the same extent as that obtained with 1.5 Kb pAt2S3. It has been documented that a 16KDa protein is detected in *Agrobacterium* and yeast as well as in the flowering stages of tobacco plant in Southwestern experiment using 0.94 Kb pAt2S3(Fig. 1.12 and Fig. 2.9). Same experiment with cauliflower mosaic virus 35S promoter with tobacco developing seed nuclear extract leads to different band pattern confirming the specificity as well as the selectivity of different nuclear protein factors to bind to different promoters (Fig.1.13). In order to clarify how a seed specific promoter is being recognised by yeast and *Agrobacterium*, an analysis of the At2S3 promoter has been attempted. In fact, *Saccharomyces cerevisiae*, being an evolutionary conserved unicellular organism, are widely used as a model system to unravel the complex eukaryotic regulation of gene expression (Buckholz and Gleeson, 1991). Several plant seed storage protein genes expressed in *S. cerevisiae* have been reported. But only two reports are there where plant protein genes were found to be expressed under the control of it's native unmodified promoter. One is phaseolin gene (Cramer et al., 1985). But it was a single subunit protein. Pal and Biswas, (1995) was able to express a multi subunit protein like 2S albumin gene 3 under the control of it's native promoter in yeast. Indeed, it has been shown by this group that in *Saccharomyces cerevisiae*, the complex post translational processing of the 2S protein was also similar to that in plant. Moreover, the extent of production of those two seed protein genes expressed under their natural promoters is also comparable; the concentration being 0.03% of the yeast intracellular total

protein. From primer extension analysis, it has been shown that At2S3 promoter is recognized by *S. cerevisiae* RNA polymerase II. It is also interesting to note that GUS gene under a heterologous plant promoter (At2S3) starts it's mRNA transcript in yeast at a point found to be same in plant (Fig. 2.4). It was earlier reported that the different tsp of phaseolin mRNA from its' own promoter (though some of the minor tsp remained the same) in *S. cerevisiae* than that in bean (Cramer et al., 1985).

The sequence analysis of the 0.94 Kb At2S3 promoter has been found to contain significant homology with other isomers of *Arabidopsis thaliana* 2S gene family (Krebbers et al.,1988). The similarities is significant mainly upto approximately 200bp 5' of the ATG. This upstream sequence of At2S3 gene contain the characteristic features of a seed specific promoter. The putative TATA box, CAAT box and the conserved G-box (CACGTG) are present (Fig. 3.2). An A/T rich motif AATTTTATG found to be involved in seed nuclear protein binding in other storage protein promoters (Jensen et. al., 1988; Jordano et. al., 1989; Bustos et. al., 1989) is also present in this At2S3 promoter. This promoter also contains enhancer like sequence, GCCA.

Further, the analysis of the nucleotide sequence of this promoter has revealed features, which is related to the binding of transacting protein(s). From the gel retardation assay the positions for regulatory protein(s) of this promoter were determined. Mobility shift assay was performed taking G-box sequence as well as sequences upstream and downstream of the G-box sequence, which have been considered potentially very important for

binding with nuclear factors in plant, yeast and *Agrobacterium*. This G-box motif (CACGTG) found mainly in the upstream sequences of the genes stimulated by the environmental cues, is also shown to be involved in nuclear protein binding in case of a seed specific promoter like 2S promoter from *B. juncea*. Binding studies with nuclear extracts have shown that proteins interact with this sequence in different plant species (Guiliano et. al., 1988; Staiger et. al., 1989; Delisle and Ferl, 1990). Different cDNAs encoding proteins that bind to oligonucleotides containing the CACGTG sequence or closely related sequences have been cloned. The G-box binding factors (GBFs) are members of a family of bZIP proteins that contain a basic DNA binding domain and a Leucine-Zipper motif (Ellenberger et. al., 1992; Vinson et. al. 1989). Outside of the bZIP domain and an N-terminal proline rich region shared by some of these proteins, they show little similarity to each other. cDNAs for more than twenty clones of G-box binding protein (GBF) have been isolated from different plants species (Brunelle & Chua, 1993; Izawa et al., 1994). G-box element is present in the promoters of many plant genes and is responsive to different environmental cues (Menkens & Cashmore, 1994). Assignment of particular transcription factors to specific roles in promoter elements can be problematic, specially in systems such as the G-box, where multiple factors of overlapping specificity exist. Four different GBFs have been identified in *Arabidopsis* alone. The fourth member of the *Arabidopsis* G-box binding factor (GBF), GBF4 is distinct from GBF1, GBF2 and GBF3. GBF4 is unable to recognize the palindromic G-box as a homodimer, although it

contains the basic region capable of specifically recognizing the G-box (Menkens and Cashmore, 1994). De Pater et al.(1994) reported a unique bZIP protein that binds to a palindromic sequence without an ACGT core located in a seed specific element of the pea lectin promoter. There are three GBF like proteins in parsley that bind the G-box element in the chalcone synthase promoter (Weisshaar et. al., 1991) and two GBF like proteins from wheat interact with the related hexamer sequence element in the promoter of the histone H3 gene (Tobata et al., 1991). Meier and Gruissem (1994) reported the isolation and characterisation of four cDNAs from young tomato fruit that encode bZIP G-box binding protein (GBF2, GBF4, GBF9 & GBF12) with the sizes of 35kDa, 28kDa, 28kDa and 25kDa respectively. Delisle and Ferl (1990) found G-box binding proteins in *Arabidopsis* alcohol dehydrogenase gene of molecular weight 31kDa and 18kDa respectively. These results suggest that families of related bZIP proteins exist in most plants which recognize the CACGTG motif, most likely including those plants for which at present only one GBF has been reported (Oeda et. al., 1991).

Two lines of evidence demonstrate a general transcriptional activation function for G-box binding factors. First, the tobacco protein TAF-1 enhances transcription from a chimeric CaMV '-90' 35S promoter fused to six copies of the TAF-1 binding site GGTACGTGGC (Oeda et. al., 1991). Second, fusion of the proline rich N-terminal fragment of the *Arabidopsis* factor GBF1 to the DNA binding domain of the yeast GAL4 protein can activate transcription from a promoter containing the GAL4 binding site in mammalian cells (Schindler

et. al., 1992). These results support a model for GBF function in plants first established by cis analysis, suggesting that GBFs have a general enhancing activity which is necessary for full activity of the respective promoters. Considering the multiplicity of GBF like proteins two questions emerge regarding their potential function in transcription activation. First how is the transmission of the different regulatory signals linked to GBFs? And second, are all GBFs within one plant species functionally equivalent or do specific GBFs link the transcriptional machinery to different signal transduction pathways? G-box (CCACGTGG) and I-box (GATAAG) sequences of the *A. thaliana* Rubiscobisphosphate carboxylase small subunit promoter are required for expression mediated by the *Arabidopsis* rbcs-1A promoter in transgenic tobacco plants and are bound in vitro by factors from plant nuclear extracts termed GBF and GA-1 respectively. It has been shown that the promoter fragment (-390 to -60 rbcs-1A) containing the G-box and two I-boxes activates transcription from a truncated iso-1-cytochrome C (CYC 1) gene promoter in *S. cerevisiae*. Mutagenesis of either of the box sequences eliminated the expression mediated by this fragment. Single copy of G-box sequences from *Arabidopsis* rbcs-1A, Adh and tomato rbcs-3A promoters were more potent activators and were used in mobility shift assays to identify a DNA binding activity in yeast functionally similar to GBF (Donald et al., 1990). The question arises as to whether the putative 16 kDa GBF present in tobacco, yeast and *Agrobacterium* is similar to the GBF thus far reported from different sources? The GBFs of different molecular weight ranging from 35 kDa to 18 kDa are reported (Delisle

and Ferl, 1990; Meier and Gruissem, 1994). On the basis of these findings the present GBF might be analogous to that reported to bind the Adh gene of *Arabidopsis* and yeast . However, it is premature to categorize these until the detailed sequences of the present 16 kDa from three systems are known.

Now, transcriptional regulation depends on the sequence specific interaction of transacting protein factors and cis-acting DNA elements. Our understanding of the structure and function of transacting factors, based on analysis of their cloned genes, has been largely restricted to factors from animals and yeast; comparatively little is known about those of plants (Levine and Hoey 1988; Hope and Struhl 1986; Johnson and McKnight 1989).

Several plant genes showing structural similarities with yeast and mammalian transcription factors have been recently identified (Katagiri & Chua 1992, for a review). The maize O2 locus is of considerable interest in this context. It has been shown on the basis of genetical and molecular studies that O2 is a transcriptional activator of the 22kDa zein and b-32 genes (Lohmer et al. 1991) and encodes a protein that belongs to the bZIP class of transcription factors (Hartings et. al. 1989). A computer search for relative protein sequences revealed further similarity between O2 and GCN4, a yeast transcription factor (Hinnebusch 1984). Mauri et. al. (1993) reported that O2 could activate H1S3 gene, gene normally under control of GCN4 in *S. cerevisiae*. Again, the yeast transcription activator GAL4 has been shown to function in cells derived from various organisms, including

plants (Fischer et al., 1988; Ma et al., 1988). Tobacco bZIP binding protein TGA1a has been found to activate transcription in both human and plant in vitro transcription system (Katagiri et al. 1990). Hilson et al. (1990) found that the fos and jun oncogene products transactive chimaeric or native promoters containing AP1/GCN4 binding sites in plant cells. This is in agreement with findings indicating that unrelated eukaryotes may share common basic mechanism for transcriptional activation (Ptashne 1988) and that in yeast, plant and animal cells such sets of cis-elements and transacting factors may commonly be utilized for regulation of gene transcription.

A mutant construct of G box sequence (CG to AT) in the 2S3 promoter (Fig. 3.4 B) when used for DNA mobility shift assay, no DNA- protein complex formation could be detected (Fig. 3.8).

In agreement with the findings of Williams et. al (1992), it was also found from DNA protein binding experiments that sequences outside the G-box core, nucleotides-4, -3, 3 and 4 relative to the axis of symmetry, affect the binding specificities of GBF, strongest binding occurs when a G is at position -4 and weaker when T is at -4 (Fig.3.4). Williams et. al., (1992) found that the size of the nucleotide at position -3, is not an important determinant, it can be either a pyrimidine (C) or a purine (A). They propose that the charged groups within the major grove of the DNA are critical. Within this major grove, both C:G and A:T base pairs present an amino group, hydrogen donor to the left and a carbonyl group hydrogen acceptor to the right, looking 5' to 3' (Schleif, 1988). This postulate is supported by the finding that the cloned G-box

binding proteins are bZip proteins; it is thought that the basic region of these proteins forms all alpha helix which lies within the major groove where it can easily recognize the orientation of this carbonyl: amino pair (Vinson et. al., 1989).All of the G-box binding protein cloned from plants thus far are members of the bZip family (Oeda et. al., 1991; Weisshaar et. al., 1991). These proteins are expected to bind to their target sequences as dimers, with the DNA binding domain of each monomer contacting one half of the symmetrical DNA binding sites (Vinson et al., 1989). Although the fourth *Arabidopsis thaliana* G-box biding factor (GBF4) cannot bind to DNA as homodimer, but as heterodimer (Menkens and Cashmore ; 1994).

In the present findings it has been shown that the ~16KDa is the one of the factors responsible for activity of the At2S3 promoter since it has made At2S3 promoter active even in older leaves pretreated with ABA. Moreover, the binding of 16KDa factor with the G-box motif has been documented in all the three systems viz. tobacco, yeast and *Agrobacterium*. The assignment of particular transcription factors to specific roles in promoter elements can be difficult especially in systems having the G-box, where multiple factors of overlapping specificity exist. In the *Arabidopsis* alcohol dehydrogenase (Adh) promoter the G-box regulates expression in the response to cold and dehydration presumably through the action of ABA. To assess the interaction of protein(s) with the Adh promoter, G-box and nearby half-G-box elements of the promoter were used. Typical in vitro assays demonstrated specific interaction of GBF3 with the G-box and half G-box element. In vitro

binding signature of GBF3 matches the foot print signatures detected in vivo at the G-box. Because RNA gel blot data indicated that GBF3 is itself induced by ABA, it is concluded that GBF3 is indeed the GBF responsible in cell cultures for binding to the Adh G-box and therefore responsible for ABA regulated expression of Adh. Potential limitations of this conclusion are considered by the fact that other GBFs bind the G-box with the same signature as GBF3, and subtle differences between in vivo and in vitro footprint signatures indicate that factors other than or in addition to GBF3 interact with the half G-box element (Lu et al., 1996). In the present context, the question arises as to what extent the putative 16 KDa protein detected in plant, yeast and *Agrobacterium* is similar. The preliminary data from DNA foot printing analysis is not conclusive because of two reasons. First, the 16 KDa protein could not be purified from the three systems worked with and second the same protein could not be obtained in sufficient quantity. Until the gene for this 16 KDa protein is cloned and expressed, no definitive DNA foot printing data could be evaluated. The common denominators by which 16 KDa protein in the three systems so far tried appeared to be similar are 1) the molecular weight of the protein is in the same range, 2) the protein binds with the At2S3 promoter even after denaturation and renaturation during the process of Southwestern analysis, 3) the putative 16 kDa protein from the three systems forms complex with the G box element of the 2S3 promoter and 4) mutation in G box sequence (ACGT) prevents its binding. A possible explanation to the question as to the seed specific plant promoter being recognised by

yeast might be to compare an endogenous, strong yeast promoter viz pyruvate kinase promoter with this At2S3 promoter (Table 2.1). All highly expressed yeast genes have been observed to contain the sequence CAAG or the consensus PyAAPu very close to CT rich domain and it is believed that the Pyrimidine block in concert with the TATA and PyAAPu elements may be correlated in determining the messenger RNA start sites (Dobson et al., 1982). Like yeast promoters, the pAt2S3 contains the consensus PyAAAPu within -24 and -36 nt upstream of translational initiator ATG (+1) and by primer extension the tsp was mapped within this sequence. As the sequence motifs of different promoters are aligned, it is seen that pAt2S3 satisfies most of the requirements of a *S. cerevisiae* gene though the nt positions of the consensus sequences vary with respect to the translational initiator ATG. The 5' end of the pyruvate kinase mRNA has been mapped and starts within the DNA sequence CAAG at -37 to -27 nt upstream from the first ATG whereas tsp of GUS gene in At2S3 promoter in yeast has been found to be 36 nt upstream from A of ATG. The 36 base domain immediately adjacent to the initiating codon ATG is very A rich in both PyK and pAt2S3. There is only a single G in this region in case of PyK promoter and in case of pAt2S3, two G residues are found to be there. An A is found at position -3 in case of At2S3 promoter which has been found to be conserved in all sequenced yeast genes (Kozak, 1981; Dobson et al., 1982). All these common features together with the documented evidence that plant G box sequence could activate transcription in *S. cerevisiae* (Donald et al., 1990) provided insights in

elucidating the salient point of recognition of a seed specific plant promoter by yeast in a constitutive manner.

Specifically, the question is asked as to how *Agrobacterium* could recognise the plant promoter and what has been revealed from the present observation is that a host of factors are present in *Agrobacterium* which can bind At2S3p in vitro. If 16KDa protein is taken as one of the transactivating factors in this case as verified by the expression of GUS gene transcript under At2S3p in transformed tobacco leaf pretreated with ABA, there is concomitant appearance of a 16KDa protein, without which little expression of GUS gene was detectable, the question arises as to why this protein is present at all in *Agrobacterium* unless this protein performs other specific function(s). Until cloning and sequencing of the gene(s) for 16KDa proteins are available from the three systems, no other function can be speculated. However, it is not surprising how At2S3 promoter is being recognised by *Agrobacterium*. Since typical *Agrobacterium* nopaline synthase (nos) promoter reveals that its controlling signals share most of the characteristics of other plant genes in particular and of other eukaryotic genes in general (Breathnach and Chambon, 1981). At the 5' end, the nos gene contains sequences TATA or Hogness box, 35 bp upstream of the start of transcription, and a sequence similar to AGGA box consensus sequence for plant genes 60 to 80 bp upstream of the 5' end of the transcript. Moreover, Boliver (1978) used nos promoter and termination signals to express heterologous gene, (CAT) in plants. The close association of plant and

Agrobacterium also helped acquisitioning some typical eukaryotic traits to *Agrobacterium*. So, recognition of At2S3 promoter by *Agrobacterium* once again suggests that At2S3 promoter signal is enough to produce a transcript of the GUS gene that is recognized by the translational machinery of the *Agrobacterium* cells to produce an active protein. Transcription of GUS gene as well as its subsequent translation into the protein product constitutively in *Agrobacterium* under At2S3 promoter is very significant. Because there are reports e.g. yeast protein, a transcription factor, GCN4 mRNA is constitutively expressed at relatively high levels, but its' mRNA is efficiently translated exclusively when amino acids are limiting (Hinnebusch, 1988). Actually, considering the primary sequences of At2S3p with nopaline synthase promoter, no striking homology was detected. But there is now an emerging concept to consider DNA as a three dimensional molecule not as a mere sequence of bases. So, global properties of a sequence operating within a large domain (Hutchinson, 1996) dictate the three dimensional structure of protein and DNA. An attempt has been made to see whether there is any homology at the level of three dimensional conformation of At2S3p and NosP using a software package for nucleic acid generation (Bansal et al., 1995) (NUCGEN). NUCGEN predicted on the basis of crystallographic data of oligonucleotides that the two promoters have significant intrinsic curvature as calculated on the basis of end to end distance/actual path length. For At2S3p this is about 0.90 and for nopaline synthase promoter, it is about 0.877. Implicitly, this lends credence to the possibility of seed specific promoter

being recognised by *Agrobacterium*. Because crucial point emerges is that not only the primary sequences of nucleotides in the promoter but its overall three dimensional structure might be important in making the promoter effective in a system. How this 16 KDa factor regulates the transcription and overall expression of the gene in *Agrobacterium*, yeast and plant as well as its homology with other GBFs are yet to be elucidated.