DISCUSSION

In this study the PAL gene has been isolated from the genome of the small crucifer, *Arabidopsis thaliana*. *Arabidopsis* has many useful features, to be exploited as an experimental system for the study of plant molecular biology (Meyerowitz and Pruitt, 1985). It has a short life-cycle (only 6 weeks), only five chromosomes and a genome size of $7 \times 10^7$ bp which is less than half that of *Drosophila* and is the smallest higher plant genome characterized to date. Leutwiler et al (1984) have performed a kinetic analysis on total *Arabidopsis* DNA and found it to be composed of 10-14% rapidly annealing sequences, 23-27% middle repetitive sequences and 50-55% single copy sequences. They have also determined that the majority of the middle repetitive DNA is derived from the chloroplast genome. The small genome size and the near absence of middle repetitive DNA in the nuclear genome allows rapid screening of *Arabidopsis* genome. Chromosomal walking is particularly suited for cloning of *Arabidopsis* genes closey linked to a RFLP marker. Genetic linkage map of the *Arabidopsis* nuclear genome consisting of 90 molecular markers that are distinguished on the basis of restriction fragment
length polymorphisms, has been constructed that enhances the utility of Arabidopsis for experiments in plant molecular genetics (Chang et al., 1988).

Numerous mutations affecting a range of developmental processes, various enzyme activities and hormone synthesis and response have been isolated in Arabidopsis, and approximately 80 of them have been mapped to the five linkage groups (Estelle and Somerville, 1987). Thus, it is possible to induce new mutations, screen or select for desired phenotypes and determine the locations of the genes revealed by mutant phenotypes. Recently attention has been focussed on the study of molecular basis of pattern formation in developing Arabidopsis flowers by cloning the homeotic agamous gene, the mutations of which result in the transformation of floral sex organs. The gene product of the agamous, has a high degree of sequence similarity to the DNA binding region of transcription factors from yeast and humans and to the product of a homeotic gene from Antirrhinum majus, suggesting that it probably encodes a transcription factor that regulates genes determining the floral development (Yanofsky et al., 1990). Other useful features of Arabidopsis include the small size of the plant and rapid seed setting which
allow large scale mutagenesis and rapid screening of the mutants. All these suggest that major contributions to our understanding of the molecular mechanisms involved in the regulation of gene expression in higher plants will emanate from studies using *A. thaliana* as an experimental model system.

The PAL gene of *Arabidopsis* has been cloned using a heterologous cDNA probe from *Phaseolus vulgaris*. A preliminary Southern blot analysis of *Arabidopsis* genomic DNA using the *Phaseolus* cDNA probe suggested that *Phaseolus* cDNA can be used as the probe to screen the genomic library of *Arabidopsis thaliana* to isolate PAL gene sequence of *Arabidopsis*.

**GENOMIC LIBRARY SCREENING**

In this study two genomic libraries of *Arabidopsis* were screened for PAL gene sequences. The genomic library of *Arabidopsis* (Landsberg erecta) constructed in the \( \lambda \)-EMBL4 vector, which was screened initially contained a high proportion (30%) of background non-recombinant plaques derived from the reconstituted vector molecules (Meyerowitz, personal communication). This library was plated on *E. coli* K803 host strain for screening. Since K803 is not a P2 lysogen and it does
not discriminate between recombinant and nonrecombinant phages both grow equally well on it. This effectively reduces the representation of recombinant phages; necessitating the much higher number of plaques to be screened. The genome size of *A. thaliana* requires only 1600 plaques to be screened for a single copy gene and usually while screening the genomic library, the library equivalent to several times the genome size is plated to maximize the probability of identifying single gene sequences. Considering all these, about 50,000 pfu of the library was plated for screening. But only one positive clone was obtained. Since the library obtained was already amplified, it may be possible that the PAL clones were under represented in the library. Amplification of a genomic library may lead to under representation of some of the recombinant clones. It is likely that certain recombinant phages propagate poorly possibly due to the detrimental effect of some of the foreign DNA sequences on the growth and life cycle of the phage. This could be one of the reason why only one positive clone was identified. Since the clone obtained from the first library turned out to be a truncated version of the PAL gene, to clone the complete gene, the coding sequences along with its flanking regions, screening a second library was inevitable.
The second genomic library of Arabidopsis (Columbia) constructed in the EMBL3 vector, was plated on E. coli NM539 host strain NM539 is a P2 lysogen and so does not allow propagation of the nonrecombinant phages (Frischauff et al., 1984). When 2,0000 pfu of the library were plated about 40 positive signals were identified. Because of practical inconvenience only 9 of these which gave strong plaque signals were picked-up for further screening and all of them were found to be containing PAL complementary sequences.

Since the first library was constructed in λ-EMBL4, the insert DNA is flanked by the EcoRI sites and can be excised by digestion of the DNA with EcoRI. On the other hand the second library was constructed into EMBL3 and the insert DNA could be excised by digestion with SalI (Fig.4). Even though both EcoRI and SalI are six bp cutter salI sites occur much less frequently than the EcoRI. Thus upon SalI digestion of the DNA very few fragments of the insert are produced. Therefore for the construction of genomic libraries of plant DNA (also for other eukaryotes) λ-EMBL3 should normally be preferred.

It has been shown by genomic Southern analysis that PAL in Arabidopsis is encoded by a small multigene
family. It is possible that the initial 40 clones represent all the members and since 9 strongly hybridizing plaques were picked, it is not surprising that the clones used for further screening contained the same members as it was shown by restriction and Southern blot analysis of the clones (Fig.11).

CHARACTERIZATION OF gATPAL-1 FROM A. THALIANA (LANDSBERG) LIBRARY

The cloning capacity of vector \( \lambda \)-EMBL 3/4 is upto 24 kb. Normally for the library construction, fragments of 15-20 kb are size selected. In order to get the representative genomic library, these 15-20 kb fragments should have been produced randomly. For random generation of fragments, shearing of the high molecular weight DNA is best. However, these fragments may contain uneven termini, which ought to be repaired by S1 digestion. Then linkers are to be added to make the fragments double. This difficulty can be overcome by using a restriction enzyme which cuts frequently, has its sites more or less randomly distributed and generate termini which are complementary to the 6 bp restriction site in the vector. For this purpose, genomic DNA is partially digested with Sau3A or MboI (GATC) for cloning
into the BamHI (GGATCC) of the vector DNA. One always hopes that GATC sequences are more or less randomly distributed throughout the genome.

Since the clone gATPAL-1 is derived from such a library, it was expected that its insert size would be of the order of 15-20 kb. The estimated insert size of the clone gATPAL1, obtained by adding the sizes of all the EcoRI fragments of the insert generated, is 15 kb, well within the expected range. The first step in characterizing the gene would be to localize the PAL complementary sequences within the 15 kb insert. Of course generation of a number of fragments of the insert would pose certain difficulties in arranging these subfragments into linear order.

First, localization as has been done involves the Southern blot hybridization of the digests of gATPAL with pPAL5 inserts as a probe. A single 2.8 kb fragment hybridizes strongly with the probe, suggesting that the PAL complementary sequences are present on this fragment. Since the fragment is too small to contain the entire gene. There were two possibilities.

(i) The clone gATPAL-1 contains a truncated PAL gene,
(ii) since the cDNA was used as a probe, the intron sequence might still be in the adjoining fragment. Therefore, it was imperative to exactly pinpoint the location of 2.8 kb EcoRI fragment within the 15 kb insert. Whether it is located in the middle or to one side of the insert. If it is in the middle, one would expect other fragments to contain adjoining sequences and some of them might be a part of the intron, so would not show up in the Southern. The digestion of gATPAL DNA with various restriction enzymes and subsequent Southern hybridization revealed that 2.8 kb EcoRI fragment is at one of the ends of the clone, adjacent to the 21 kb arm.

When the Southern hybridization experiments were performed with restriction digests of gATPAL-1 with heterologous cDNA as a probe, the vectors arms were also giving signals. This problem has to be resolved before the authenticity of the PAL clone could be established. One possibility was that since cDNA construction involves dC and dG tailing, these poly C or poly G tracts in the forms of tails of cDNA might be hybridizing to such G and C rich sequences in genome. By selectively removing these sequences by digestion with enzymes which remove both the termini of the cDNA, the hybridization to arms could be eliminated.
The authenticity of the PAL clone was established only by sequencing and comparing the sequences with that of the bean cDNA. For this purpose, 2.8 kb EcoRI fragment, which gave a strong signal in Southern experiments, was subcloned.

By determining the nucleotide sequence of the 0.5 kb and 0.3 kb HindIII fragments and comparing it with the sequence of PAL5, the Phaseolus PAL cDNA, the identity of the clone was confirmed. Such a comparison revealed almost 90% homology between the deduced amino acid sequences of Phaseolus PAL cDNA and that of 0.5 kb and 0.3 kb HindIII fragments of the Arabidopsis clone gATPAL-1.

Nucleotide sequence of the most of the 2.8 kb EcoRI fragment was determined and compared with that of one of the PAL genes (gPAL2) of Phaseolus vulgaris. This revealed that there is an intron of 453 bp in Arabidopsis PAL gene in the same location as it is in Phaseolus and that the Exon-I was truncated in the clone gATPAL-1. There was a very good degree (above 80%) of homology between the deduced amino acid sequences of PAL from both the systems.
CHARACTERIZATION OF CLONES FROM THE GENOMIC LIBRARY OF
A. THALIANA (COLUMBIA)

In order to get the complete gene along with its 5' upstream sequences a second library was screened with the 2.8 kb EcoRI fragment as probe. Three positive clones from the second library were analysed by Southern blot hybridization using a 5'-end specific probe (the 0.45 kb PstI-XbaI fragment) and a 3'-end specific probe (the 0.5 kb HindIII fragment which is in the middle of Exon-II). The results showed that the PAL complementary sequences were present in the middle of the inserts of gATPAL-2 and 4 whereas the 4.8 kb fragment was truncated in the clone gATPAL-2.

SANDWICH HYBRIDIZATION

One of the technical strategy to be mentioned at this point is the indirect hybridization of the insert, in the form of ssDNA (M13 construct containing the insert) of interest as the probe in the first stage of hybridization and nick translated RF of M13 DNA as the probe in the second stage (principle explained in methods). The 3'-end specific probe in the above Southern blot hybridization was the ss-M13 DNA containing the 0.5 kb HindIII insert (the template that was used for sequencing).
Since the DNA probe is in the single stranded form there is no reannealing of probe molecules and hybridization can be carried out for short durations and the reaction is driven toward the formation of hybrids between the ssDNA probe and the DNA on the filter. Hybridized with the nick translation RF-M13 DNA can be carried out for very short durations (as short as 1 hr) thus reducing exposure to $^{32}$P. The main advantage of this strategy is when different probes have to be used.

It requires the synthesis of only one labeled probe (M13 RF DNA) for detecting all the target sequences. The filters after hybridization with ss DNA containing different inserts should be treated separately when the second stage of hybridizations are being carried out. Filters may be washed together during high stringency washes.

A potential advantage of this strategy is that the signals produced from short and long inserts should be the same size, it is the labeled vector sequences which produce the signal. This type of approach should prove to be beneficial for gene quantitation experiments since signal strength variations due to differences in probe sizes are eliminated.
Since the hybridization signal was found to be enhanced many times, this strategy may prove useful for genomic blot hybridizations but whether such a strategy can be used when hybridization has to be done at low stringency is yet to be assessed. (This is questioned because washing off of $^{32}$P-labelled probe molecules need to be done at high stringency and low stringent washes may result in high background).

**M13 CLONING AND SEQUENCING STRATEGY**

The sequencing strategy is as shown in Fig.17. The different fragments were sequenced after cloning in forced orientations in M13 vectors and sequencing from both ends. Deletion constructs were made in order to sequence the fragments which did not have convenient restriction sites.

The 2.8 kb PstI EcoRI fragment which contained most of the coding sequences and the intron had many sites for TaqI. The 450 bp sequence from PstI site of the 2 kb PstI-EcoRI fragment also contained two sites for TaqI. So it was thought TaqI would be an ideal frequent cutter to use for the construction of nested deletions by partial digestion with 4 bp restriction enzyme.
Lack of overlapping deletion constructs obtained may be because only limited number of recombinant clones were screened for the desirable size of the ss DNA template. It may also be due to the lack of many sites for the frequent cutter used in the insert. It will be useful to use a combination of frequent cutters in separate batches of partial digestions. Ligation, and post ligation digestion to linearize the undigested starting DNA and transfection, all can be done in a single tube after pooling the desired partial digestion products of all the enzymes. Considering the simplicity, less amount of labour, and time involved in the construction of deletions by this strategy, this method will prove very efficient and useful if enzyme of proper choice or a combination of frequent cutting enzymes are used to generate the partial digestion products. Screening of ss DNA from multitude of clones can be done with minimum effort by the 'DIGE' electrophoresis, which does not involve purification of ss DNA from each and every clone. After screening them, desired ss DNA templates can be isolated for sequencing. Priming test, which involves a short extension of the primer using Klenow and a very low activity of $^{32}$P, need to be done when the enzymes used for the partial digestions have site(s) in between the polylinker.
cloning region and the gene II since the sequence present in this region is not necessary for the growth and propagation of the phages, and clones which do not contain sequences in this region will also survive and give rise to plaques but ss DNA templates from such clones cannot be primed due to the lack of the primer annealing site.

When DNA fragments were cloned into a single site the orientation of the insert can be in both ways. In order to sequence both the strands clones containing inserts in the opposite orientations were identified by 'one lane analysis'. In this strategy, ss DNA from several clones was isolated and sequencing reactions were performed for any one of the four bases but the same base sequence was analysed for all the clones. Upon running them on a sequencing gel, two categories of the base patterns should be seen, if clones containing opposite strands are being analysed. Thus the clones containing opposite strands could be identified and used for performing the sequencing reactions to get the nucleotide sequence information of both the strands. Such a strategy was followed for sequencing 0.5 kb and 0.3 kb HindIII fragments.
The dideoxy chain terminating method was carried out using either Klenow or the modified form of T7 DNA polymerase (sequenase), in which its exonuclease activity has been inactivated. The enzyme is composed of two subunits: T7 DNA polymerase—a product of T7 gene 5 and E. coli encoded TrxA gene product, thioredoxin. The thioredoxin component confers high degree of processivity to the reaction. High processivity of the enzyme makes it suitable for reading longer sequences (upto 450-500 bases could be read) and therefore, it is preferred over Klenow fragment.

ANALYSIS OF NUCLEOTIDE SEQUENCE OF ARABIDOPSIS PAL:
The nucleotide sequences of the 2.8 kb EcoRI fragment of the clone gATPAL-1 from Arabidopsis thaliana (Landsberg) library and that of the 2.8 kb PstI EcoRI fragment derived from the clone gATPAL-2 from Arabidopsis thaliana (Columbia) were 100% identical, showing that the same gene was isolated from both the libraries. When the nucleotide sequence was compared with the Phaseolus cDNA and one of the PAL genes (gPAL2) of Phaseolus (bean), presence of an intron of 453 bp was revealed. It also showed that the nucleotide sequence of the 3'-end of the gene was not determined. The 2 kb PstI-EcoRI fragment contained a short stretch of the
**Fig. 25:** A comparison of PAL genes of different plants.
coding region at the 5'- and the 5'-flanking region. The translation start site was assigned to the first initiation codon in frame with the rest of the coding region.

The 5'-end of Arabidopsis PAL has 12 amino acids in excess when compared to the deduced amino acid sequence of the Phaseolus PAL. The first exon of Arabidopsis PAL gene is 426 bp and 1539 bp of the nucleotide sequence of the second exon has been determined. The size of the exon II in Phaseolus and other plant PAL is 1.7 kb.

PAL GENE STRUCTURE: From the analysis of the nucleotide sequence by comparing with the PAL gene of Phaseolus the structure of Arabidopsis PAL gene could be derived as shown in Fig. 22. It has an intron of 453 bp. The exon I is 426 bp long and the expected size of the exon II is 1.7 kb.

A COMPARISON OF PAL GENES FROM DIFFERENT PLANT SYSTEMS: The Arabidopsis PAL gene is compared with the PAL genes of Phaseolus (Cramer et al., 1989), Parsley (Louis et al., 1989) and rice (Minami et al., 1989). In all the systems as shown in Fig. 23, the intron in the PAL gene
TABLE 1

Features of 5' upstream sequence of Arabidopsis PAL gene. Numbering with reference to translation start codon.

<table>
<thead>
<tr>
<th></th>
<th>From</th>
<th>To</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TATA boxes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-130</td>
<td>-135</td>
<td>TAATTT</td>
</tr>
<tr>
<td></td>
<td>-228</td>
<td>-223</td>
<td>TATAAT</td>
</tr>
<tr>
<td></td>
<td>-315</td>
<td>-320</td>
<td>TATAAT</td>
</tr>
<tr>
<td><strong>CCAAT boxes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-175</td>
<td>-179</td>
<td>CCAAC</td>
</tr>
<tr>
<td></td>
<td>-261</td>
<td>-265</td>
<td>CCAAC</td>
</tr>
<tr>
<td></td>
<td>-355</td>
<td>-358</td>
<td>CAAT</td>
</tr>
<tr>
<td><strong>Candidates for cis elements</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>-177</td>
<td>-186</td>
<td>CTTACCTACC</td>
</tr>
<tr>
<td>b</td>
<td>-258</td>
<td>-268</td>
<td>CAACCAACTCC</td>
</tr>
<tr>
<td>c</td>
<td>beyond 600 bp</td>
<td>ATATATATAGTTATA</td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>120 bp upstream of 'c' element</td>
<td>ATATATATATTTTTTACTTT</td>
<td></td>
</tr>
<tr>
<td>e</td>
<td>beyond 1 kb</td>
<td>AAATTAATATA</td>
<td></td>
</tr>
<tr>
<td>f</td>
<td>40 bp upstream of 'e' element</td>
<td>TATAAATCAA</td>
<td></td>
</tr>
<tr>
<td>g</td>
<td>-211 -221 and beyond 1 kb</td>
<td>ATGGAAGAAAA</td>
<td></td>
</tr>
<tr>
<td>g</td>
<td>-153 -161 and 52 bp upstream of 'c' element</td>
<td>TTATTTAAA</td>
<td></td>
</tr>
</tbody>
</table>

~(V) 
| c | I- |
is located at the same position i.e. on the codon of the same amino acid (arginine). The sizes of both the exons are almost the same in all these plant species. Inspite of the conservation of the intron site there is a high degree of divergence in the intron size as well as its nucleotide sequences in the various plant species. The two PAL genes (gPAL2 and gPAL3) of Phaseolus, have introns of different size with a high degree of divergence at the nucleotide sequence level (Cramer et al., 1989).

FEATURES OF THE 5' UPSTREAM SEQUENCE OF ARABIDOPSIS PAL GENE:

There are three sets of potential candidates for the TATA and CAAT boxes between -129 and -360 (with respect to translation start codon). Since the transcription start site has not yet been determined the exact TATA and CAAT boxes could not be assigned. It may be that PAL promoter in Arabidopsis has more than one TATA and CAAT boxes. Presence of more than one boxes may be implicated in the differential use of promoter sequences for the gene expression under different conditions (developmental and environmental factors). A list of promoter sequences and other possible putative
Table 2: List of UV and elicitor Inducible cis-elements in various promoters

<table>
<thead>
<tr>
<th>Distal (with respect to transcription startsite or translation startsite)</th>
<th>Proximal</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAACCAACTC // CTTACCTACC</td>
<td>Arabidopsis PAL (this study)</td>
</tr>
<tr>
<td>CTCCAACAAACCCC // TCTCACACCCCA</td>
<td>Parsley PAL (Louis et al 1989)</td>
</tr>
<tr>
<td>TGCCAACTGACCCG // TCTCAGCTACCA</td>
<td>Antirrhinum majus (CHS (Sommer and Saedler, 1986))</td>
</tr>
<tr>
<td>CCACCAAACCTCC // ACTCACCTACCC</td>
<td>Phascolus vulgaris CHS (Dron et al 1988)</td>
</tr>
<tr>
<td>TTACCAACCCGG // CTCACCAACCC</td>
<td>Parsley 4 CL (Douglas et al 1987)</td>
</tr>
<tr>
<td>// TCTAACCTACCA</td>
<td>Arabidopsis CHS (Feinbaum and Ausubel 1988)</td>
</tr>
<tr>
<td>// TCCAACTAACT</td>
<td>Parsley CHS (Herrman et al 1988)</td>
</tr>
</tbody>
</table>

Consensus:--

(a) TCTC CCTA C CC
(b) CCA A c A AAC C CC
cis-elements in the 5'-upstream sequence of *Arabidopsis* PAL gene is given in table 1.

A comparison of the 5' upstream sequences of the *Arabidopsis* PAL gene with a few stress light inducible genes reveals several interesting features. The sequence motif CTTACCTACC present from -178 to -188 (A element) and the sequence motif CAACCAACTCC present in between -255 to -271, (B element) are similar to the conserved motifs present in the promoters of several other genes of phenylpropapanoid pathway. These conserved motifs in these genes, having the consensus TACTACCTAC and CCACATCAC respectively display inducible footprints involved in the responses to both UV-irradiation and elicitor application (Louis et al., 1989). A list of these sequence motifs in the promoters of several genes is shown in table 2. In all these promoters, the sequence motif 'a' is present immediately next to the transcription start site whereas the sequence motif 'b' is present away from the transcription start site. The same arrangement is observed in *Arabidopsis* PAL promoter also. The element 'a' is proximal to the translation start site (transcription start site not determined) and element 'b' is distal to the translation start site. The
significance of these sequence motifs is emphasized by the studies on the promoters of other eliciter and UV-inducible genes. In the bean CHS gene, the sequence motif a is present in the region between -52 and -74 of the promoter and sequence motif 'b' is present between -155 and -166. Mutation analysis in the bean CHS promoter defined an elicitor regulation activator located between positions -30 and -173 (Dron et al., 1988), a stretch which includes both the conserved elements a and b. A parallel situation occurs in the case of the Antirrhinum CHS gene, where an orientation-independent UV-light responsive element has been defined using a heterologous parsley protoplast system between -39 and -197. (Lipphardt et al., 1988). Moreover, the fact that the cis-acting elements involved are conserved between these species, may suggest a similarity of trans-acting factors interacting with these element. Taken together these observations point towards an involvement of the conserved elements in elicitor or light mediated gene activation. This notion is strengthened by the fact that in parsley similar motifs display footprints inducible by such stimuli in both the PAL1 and 4-coumarate-CoA ligase gene which a coordinately regulated with PAL (Douglas et al., 1987). The same is true for one of the conserved motifs in the
parsley CHS promoter which not only displays a UV-light inducible in vivo footprint but also includes an element shown to be necessary for activation of CHS by UV-light (Schulze-Lefert et al., 1989). 

Further evidence that these elements play a role in the regulation of transcription comes from an analysis of snap dragon mutants (Schulze-Lefert et al., 1988). Where short deletions (3-15 bp) centered at position -55 (defining sequence motif 'a') in the CHS promoter leads to 65-75% reduction in the levels of CHS mRNA. All these observations suggest that these conserved sequence motifs a and b may be the site of binding of regulatory factors that promoter or facilitate gene activation as response to UV-light or an elicitor.

Another observation in the 5'upstream sequence of Arabidopsis PAL gene is the occurrence of an AT repetitive region (element 'c') the location of which is assumed to be somewhere between -600 and -700. (This part of 5'upstream sequence was derived from a deletion construct of the 2 kb PstI RI fragment. The ss DNA template of this construct was of the size of 1.5 kb. From this it was assumed that the break between position -362 and the II part of the 5'upstream sequence must
include a missing sequence of about 100-200 base only. Similar AT rich sequence stretches are present in the 5'upstream sequence of Arabidopsis CHS gene (Feinbaum and Ausubel, 1988) in two locations between -273' and -283 and between -255 and -265. Also there is an AT rich sequence of about 35 bases present in the gPAL2 gene of Phaseolus vulgaris located upstream of -500 from the transcription start site (Cramer, et al., 1989). In Arabidopsis PAL also this element is located far away from the translation start site.

Such an element is also present in the 5'upstream sequence of a HSP gene 17 (encoding the low molecular weight heat shock protein of 17 kD) of soybean. The AT repetitive element is reminiscent of certain upstream enhancer elements in yeast (Struhl, 1984) and nuclear scaffold attachments in Drosophila (Gasser and Laemmli, 1987). The functional significance of these elements in plants is yet to be assessed.

Another upstream sequence element, ('d'), ATAATATATTTTTTTACTTT located somewhere beyond position -700 from the translation codon may be implicated in having a role in the UV-induced gene expression. This notion comes from the fact that a similar conserved sequence exists in the upstream region of Antirrhinum
CHS gene (Sommer and Saedler, 1986) and is implicated in the UV activation of CHS gene. This element ATAATATATTTTTTTTTTTAT in Antirrhinum is located between -283 and -301 from the transcription start site. The presence of three regulatory sequence regions in the upstream region of the Antirrhinum CHS gene has been identified by transient expression studies in parsley protoplasts, upon UV-irradiation (Lipphardt et al., 1988). Out of these three regulatory regions the first region is immediately upstream of the TATA box with coordinates -39 to -197 and functions as an orientation independent UV-light responsive element. The next upstream -197 to -357 of the Antirrhinum CHS gene contains sequences that do not by themselves cause UV-induced expression but specifically potentiate the level of UV-induced expression when combined with the UV-responsive element proximal to the TATA box. A third element (-661 to -564) acts as a general enhancer for the level of both induced and uninduced expression (Lipphardt et al., 1988). The Antirrhinum sequence motif that is similar to the Arabidopsis PAL sequence motif 'd' is present in the -197 to -357 region of the Antirrhinum CHS gene which is involved in potentiating the effect of the proximal cis element of the Antirrhinum CHS gene. Such an observation suggests a
similar role for sequence motif 'd' of Arabidopsis PAL in the UV-induced transcription of the gene.

Apart from the above mentioned putative sequence elements, the presence of other sequence motifs, e, f and g in the upstream sequence of Arabidopsis PAL gene could be visualized on the basis of the similarity of such sequences in other stress-inducible genes. A comparison of upstream sequences of the Arabidopsis PAL with those from the Arabidopsis CHS gene (Feinbaum and Ausubel, 1988) reveals a similarity of certain conserved segments. The sequence AAATTAATAT (element 'e') present between -224 to -235 of Arabidopsis CHS gene is also present in a region of almost 1000 bp upstream of translation codon of the Arabidopsis PAL gene. Another sequence TATAAAATCAA, (element 'f') between -192 to -205 of Arabidopsis CHS gene is found at a position further upstream of the above sequence in the Arabidopsis PAL gene. The conservation of such sequences in the two genes which are possibly activated by similar set of signals may argue for a certain functional role for these elements.

A comparison of the 5'upstream sequence of the Arabidopsis PAL gene with that of the rice PAL gene reveals a conserved region in both the upstream
sequences. The sequence motif ATGGAACAAAAAAG is present between -170 to -180 of the rice PAL gene (Minami et al., 1989) and the equivalent sequence motif 'g' ATGGAAGAAAAAAG is located between -211 to -221, and beyond 41 kb in the III part of the 5'-upstream region of Arabidopsis PAL gene. The sequence motif may have a certain role in the expression of PAL gene in both. The sequence motif TTATTTAAA, ('h') is located at -154 and is repeated again in a distal region approximately beyond the position -600 in the II part of 5'-upstream region of Arabidopsis PAL gene. It may have a significance in the molecular of PAL gene expression. These plant species.

The role of these putative upstream DNA elements in expression and/or modulation of the PAL gene could only be assessed from the functional analysis of these cis-elements in response to the various factors that regulate the gene expression of PAL.

Organization of PAL gene in Arabidopsis

The complexity of the gene whether it is present as unique or is present as members of a gene family can be determined from the genomic Southern blot analysis with appropriate gene specific probes. At stringent
conditions of hybridization with the gene specific probe, the fragments corresponding to the gene probe used, can be identified. At a relatively low stringent conditions, fragments corresponding to the divergent forms of the gene also will hybridize. Such an experiment is exploited in the analysis of organisation of PAL genes sequences in the genome of Arabidopsis thaliana.

The results of the Southern blot analysis of Arabidopsis genomic DNA probed with the 2.8 kb PstI-EcoRI fragment comprising of most of the gene sequences, at high, moderate and low stringent conditions indicate the presence of a small multigene family of PAL in Arabidopsis thaliana. The genomic blot hybridization experiments done at high stringency identify several genomic fragments generated by various restriction enzymes corresponding to the isolated PAL gene from Arabidopsis. When the stringency is lowered many more fragments appear in all the digests. The occurrence of more than the expected number of bands and then low intensity suggest the possibility of other polymorphic forms of PAL gene in Arabidopsis with certain degree of sequence divergence. The Arabidopsis PAL gene that was cloned in this work is within the 4.8 kb EcoRI fragment,
and this fragment shows up very strongly. Apart from this 4.8 kb EcoRI band, three other bands representing approximately 6, 9 and 13.5 kb EcoRI fragments and possibly a 2.8 kb EcoRI fragment also light up. The first three fragments could be identified even at high stringency as faint signals. The presence of four distinct bands in the EcoRI digests suggest that PAL is encoded by at least four genes in Arabidopsis. The relative intensity of signals under high stringent condition may reflect the degree of relatedness or divergence of sequences from the PAL gene sequences cloned in this work. Multiplicity of bands under low stringent condition suggest the presence of (i) a number of copies of the PAL genes with greater degree of restriction fragment length polymorphism (ii) other gene sequences with some degree of relatedness to the PAL gene.

The actual number of the different members of gene family can be obtained by identifying the DNA sequences complementary to the transcripts of the various members from a cDNA library. The presence of many polymorphic forms of PAL gene, in Arabidopsis may suggest the possibility of involvement of complex regulatory network of the PAL gene expression in Arabidopsis either with
respect to various environmental stimuli, developmental cues or tissue specific expression of various members of the gene family.

The enzyme PAL has been shown to be encoded by a small multigene family in other plant systems also (Cramer et al., 1989, Minami et al., 1989, Louis et al., 1989). A common feature of multigene families is that their members are expressed in different ways in response to environmental and developmental signals. The multigene family of PAL is a good example of gene family that is differentially expressed in plants in response to environmental and developmental cues. As it has been already discussed under Literature Review different members of the PAL gene family in Phaseolus which encode the distinct subunit isoforms exhibit, markedly different patterns of regulation with respect to both organ specificity during normal development and within a single organ in response to diverse environmental stimuli (Liang et al., 1989). These observations suggest that the selective expression of PAL genes encoding functional variants is governed by a complex set of regulatory net-works for developmental and environmental control of phenylpropanoid biosynthesis.
Modifications of the pattern of expression of specific PAL genes in transgenic plants will help delineate the functional significance of the variant enzyme isoforms and functional assays of cis-acting elements involved in selective expression of PAL genes will allow molecular dissection of underlying regulatory networks. Hence further analysis of PAL gene expression should give an insight into the molecular mechanisms underlying the flexible integration of defense responses within the overall program of plant development.