CHAPTER V

Cloning and heterologous expression of herbivory induced transcripts in tea
5.1. Introduction

Herbivory induces significant changes in gene expression with the emergence of new and sometimes novel stress induced transcripts in plants. Some of them may have potential roles in stress mitigation either by way of repair mechanisms, alternative growth or by expressing defense related proteins as part of the direct defense machinery of the plant. Such direct plant defences against insect herbivores include limiting food supply, reducing nutrient value, reducing preference, disrupting physical structures, inhibiting chemical pathways and inviting potential predators of the attacking insect. Major known defense chemicals include plant secondary metabolites, protein inhibitors of insect digestive enzymes, proteases, lectins, amino acid deaminases and oxidases (Chen, 2008). Moreover, the role of signalling molecules in initiating and coordinating such direct or indirect defense responses cannot be overlooked.

In view of the above and the extent of damage caused by Helopeltis to tea productivity, the identification and characterization of defense related genes is a fundamental prerequisite towards any biotechnology-based pest control practice in tea. As such, this chapter discusses the full length decipheration of four differentially induced transcripts (cDNAs) in the Helopeltis tolerant cultivar 111/1 followed by their heterologous and purification.

5.1.1. Rapid amplification of Cdna ends (RACE)

Rapid Amplification of Cdna Ends (RACE) is a technique used in molecular biology to obtain the full length sequence of an RNA transcript found within a cell. RACE results in the production of a Cdna copy of the RNA sequence of interest, produced through reverse transcription, followed by PCR amplification of the Cdna copies. The amplified Cdna copies are then sequenced and, if long enough, should map to a unique Mrna present in the experimental Mrna pool, the full sequence of which is either novel or reported. RACE can provide the sequence of an RNA transcript from a small known sequence within the transcript to the 5’ end (5’ RACE-PCR) or 3’ end (3’ RACE-PCR) of the RNA. This technique is sometimes called one-sided PCR or anchored PCR.
RACE amplifies either terminals of a cDNA, even where they have unknown sequence (Frohman et al., 1988, Ohara et al., 1989). RACE only requires knowledge of a short sequence within the mRNA of interest. It is often used for cloning the remainder of incomplete cDNAs. In PCR, RACE uses a “specific primer” designed using the known mRNA sequence together with a “general primer” complementary either to the mRNA poly (A) tail (for 3’-RACE) or to a homopolymer added to the 3’-end of the cDNA (for 5’-RACE) (Fig.5.1). Because homopolymers do not make good PCR primers, and to facilitate cloning of RACE products, the general primers contain a sequence with a restriction endonuclease site at their 5’-end. The cDNA template for the PCR may be produced either using an oligo-Dt primer (for 3’- or 5’-RACE) or using a primer complementary to the known sequence within the mRNA (for 5’-RACE only). Where RACE gives a mixture of products, an aliquot of this mixture may be used as the template for another PCR using a second specific primer (nested within the first) and the original general primer (nested RACE).

Upon successful RACE analysis, the resultant DNA sequence is expected to represent a copy of a complete mature mRNA and therefore the encoded protein.

5.1.2. Heterologous expression of protein

Heterologous expression involves identification of genes and transfer of the corresponding DNA fragments to hosts other than the original source for
synthesis of the encoded proteins. Protein isolation, especially from plant sources, can be costly, cumbersome and lengthy, and heterologous expression provides a convenient alternative (Fig 5.2). This methodology allows large-scale production of plant proteins in microorganisms to study their biochemical and biophysical features. Foreign hosts may also provide a simpler system for studies on functions of proteins and for elucidation of their roles in complex mechanisms such as metabolic reactions and membrane transport. Recombinant plant proteins and peptides produced by heterologous expression are also used in industrial applications. It is worthwhile to mention that the selection of expression systems is done depending on whether the purpose of study is production of large quantities of protein or investigation of functional features of the cloned protein. The physicochemical properties of the investigated protein also play a role in this choice.

*Escherichia coli* (*E.coli*) is the first and most extensively used prokaryotic expression system for heterologous protein production (Frommer and Ninnemann, 1995). It remains generally the first choice due to its simplicity, rapid growth rate, and relatively low cost. Almost all commercially available inducible cloning vectors are compatible with *E. coli* and extensive biochemical and genetic information is available.

Heterologous expression is a powerful tool for functional and biochemical analyses of genes and gene families isolated from various organisms. It is particularly important for plants where the whole genome sequence is not available. Therefore, heterologous expression of differentially expressed genes from 111/1 will allow a better insight on the physiological, structural and functional aspects of their encoded proteins or some of their domains, an understanding which may significantly help us in addressing the criticality of these proteins towards herbivore deterrence.
5.1.3. Proteins under study

In this study, four herbivory induced differentially expressed transcripts in 111/1 were considered, encoding four different proteins namely – Ricin agglutinin (a component of direct defense with insecticidal properties), Ethylene responsive transcriptional co-activator (ERTC) like protein(part of stress induced signal transduction pathway), Formate dehydrogenase (a protein essential for alternate respiration during stress) and NBS-LRR like protein (a disease resistant and stress signal transducing protein).

The ricin-agglutinin (synonomous- lectins) family proteins are type II ribosome inactivating toxin proteins (RIP) (type I being the Shiga family) which inactivate 60S ribosomal subunits by an N-glycosidic cleavage which releases a specific adenine base from the sugar-phosphate backbone of 28S Rrna thereby inhibiting protein synthesis in eukaryotic cells (Igarashi et al., 1988; May et al., 1989; Funatsu et al., 1989). They depurinate the universally conserved α-sarcin loop of large rRNAs. This depurination inactivates the ribosome, thereby blocking its further participation in protein synthesis, indicating their cytotoxicity (Van Damme, 2008). Although many plant lectins are able to bind simple sugars such as Glc, Man, or Gal, they have a much higher affinity for oligosaccharides, which are not common or totally absent in plants (Peumans and Van Damme,1995), indicating their role in activities other than plants own.

Fig 5.2 Flowchart for heterologous expression
metabolism. The preferential association of lectins with those parts of the plant that are most susceptible to attack by foreign organisms is also an argument for a protective role (Peumans and Van Damme, 1995). A circumstantial argument in favour of a defense role of plant lectins is their marked stability under unfavorable conditions (Peumans and Van Damme, 1995). An investigation by Sadeghi et al., (2006) on the effect of lectins on oviposition revealed that lectin substantially reduces the egg laying capacity in seeds of chickpea. The most notorious example of a type-2 RIP is the extremely toxic ricin.

**Ethylene responsive transcriptional co-activator** is a kind of **multiprotein bridging factor 1** (MBF1) interact with transcription factors, such as c-Jun, GCN4, and ATF1, or with different hormone receptors and link them with the TATA-binding protein (Takemaru et al., 1997; Takemaru et al., 1998; Brendel et al., 2002; Liu et al., 2003; Busk et al., 2003). Transgenic *Arabidopsis thaliana* plants constitutively expressing the MBF1c gene enhances the tolerance of transgenic plants to bacterial infection, heat, and osmotic stress by partially activating, or perturbing, the ethylene-response signal transduction pathway (Suzuki et al., 2005), a central mediator of the emission of specific volatile organic compounds as indirect defense, accumulation of phenolic compounds and proteinase inhibitor activity (Dahl and Baldwin, 2007). They also found that such plants remain tolerant to osmotic stresses in the presence of an additional stress like heat. Given the established role of ethylene in coordinating herbivory induced defense responses in other plants, full length elucidation and heterologous expression of the same in tea was found necessary.

**Formate dehydrogenase** (FDH, EC 1.2.1.2.) is a soluble mitochondrial enzyme capable of oxidizing formate into CO2 in the presence of NAD+. It is abundant in non-green tissues and scarce in photosynthetic tissues. Norton (1963) suggested that in fresh potato tuber slices, the bulk of respiration is mediated by systems other than the TCA cycle. Cabassa and his group clearly demonstrate that formate is a good respiratory substrate for isolated potato tuber mitochondria and is consistent with the high level of FDH protein found in tuber (Cabassa et al., 1998). Under stress, FDH transcripts (and protein) accumulate in leaves, and leaf mitochondria acquire the ability to use formate as a respiratory substrate (Ambard-Bretteville et al., 2003). Cabassa et al., 1998 show that FDH Mrna expression is responsive to dark, hypoxia, wounding, cold, and drought wherein only, wounding led to a rapid
response of FDH mRNAs (20 min), accumulating 13-fold higher than the control for up to 1 h, remained high for 24 h, and then decreased after 24 h. in potato plants. The relatively slow responses to all the other stresses tested here (8 h on average) probably reflect regulation of FDH expression by changes in concentration of certain intracellular metabolites due to the activation of specific metabolic pathways; for example, dark, hypoxia, and chilling all induce glycolysis. The fact that the response of FDH transcripts to ABA is minor compared with their induction by most stresses suggests that ABA independent transduction pathways are probably involved in the FDH response to stress.

**Leucine-rich repeat (LRR)** class of proteins, typified by imperfect repeats of blocks of amino acids, usually with about 24 residues per repeat element. The LRR resistance gene proteins may also have nucleotide binding sites, leucine zipper domains, or kinase domains suggestive of signal transduction functions. In a few cases, disease resistance genes have been transferred to foreign plants by transformation and generally shown to be functional. A few LRR plant disease resistance genes have been shown to exhibit dual specificities—that is, the plant harbouring them either recognizes two different pests or two different elicitors. Especially exciting was the recent finding by Valerie Williamson and colleagues at the Univ. of California, Davis (Rossi et al., 1998) that the cloned *Mi*–resistance gene in tomato against the root knot nematode also recognizes a species of aphid. It is not known whether the nematode and aphid produce the same elicitor, as is likely, but the finding is of considerable importance and has practical implications that should stimulate the search for additional disease resistance genes that target insects. While several examples of insect-targeting resistance genes are recognized, they are relatively rare compared to resistance genes known against fungi, bacteria, nematodes, and viruses.

It is pertinent that DNA sequence for the entire encoded protein or at least the trait specific functional domain of these transcripts is identified and expressed as heterologous proteins. This will enable a better understanding of their physiological and structural attributes with significant implications to pest management strategies.

5.2 Materials and methods
5.2.1. Rapid Amplification of cDNA Ends (RACE)

Full length decipheration of the genes containing the ORF was accomplished by 5' RACE technique according to the instructions and components provided in the Invitrogen GeneRacer kit (cat. No. L1500-01) using the total RNA from 111/1 8hr. infested leaf samples.

The GeneRacer technique is based on RNA ligase-mediated (RLM-RACE) and oligo-capping rapid amplification of Cdna ends (RACE) methods, and results in the selective ligation of an RNA oligonucleotide to the 5′ ends of decapped Mrna using T4 RNA ligase (Maruyama and Sugano, 1994; Schaefer, 1995; Volloch et al., 1994). The prepared single strand Cdna acted as template for 5'RACE PCR reactions in order to obtain the entire protein coding sequence of the four genes mentioned above. The primers used in each of the RACE PCRs are given below (Table 5.1).

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Size(mer)</th>
<th>Tm(°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5' Gene racer primer</td>
<td>5'CGACTGGAGCACGAGGACACTGA3'</td>
<td>23</td>
<td>74</td>
</tr>
<tr>
<td>5' gene Racer nested primer</td>
<td>5'GGACACTGACATGGACTGAAGGAGTA3'</td>
<td>26</td>
<td>78</td>
</tr>
<tr>
<td>Right primer for Ricin agglutinin (PR_S3IO18_R)</td>
<td>5'GATTCCCATGGAACCGCCACAA3'</td>
<td>22</td>
<td>68.16</td>
</tr>
<tr>
<td>Right primer for ERTC (PR_S5P3A10_R)</td>
<td>5'CTGCTCGAGGCAGCTTTGTTTGAG3'</td>
<td>24</td>
<td>68.58</td>
</tr>
<tr>
<td>Right primer for FDH (PR_s3I_E12_R)</td>
<td>5'TCAGAGCCAATTCCAGCTGTGAGC3'</td>
<td>24</td>
<td>69.10</td>
</tr>
<tr>
<td>Right primer for LRR (PR_S5P2F1_R)</td>
<td>5'GTGCATCCGGGATTCTCCAGTAA3'</td>
<td>24</td>
<td>69.17</td>
</tr>
</tbody>
</table>

Table 5.1 Primers used in 5’RACE PCRs

5.2.1.1. 5’RACE PCR set up

All the reactions were set in a sterile, Dnase/Rnase free 0.2ML PCR tube in the following order per 10 µL of reaction volume (Table 5.2). All components (except the GSP) were part of the Platinum Taq. DNA polymerase High fidelity kit (Invitrogen corp., Cat. No.11304-011).

<table>
<thead>
<tr>
<th>Components</th>
<th>Volume</th>
</tr>
</thead>
</table>


The anti-Taq DNA polymerase antibody complexes with and inhibits polymerase activity at room temperature. Activity is restored after the initial denaturation step in PCR cycling at 94°C, providing an automatic “hot start” for increased specificity, sensitivity, and yield (Chou et al., 1992, Sharkey et al., 1994).

All the PCR reactions were performed in Eppendorf Mastercycler thermal cycler with different combinations of annealing temperatures (Table 5.3). The resultant PCR products were run in 1% TAE EtBr gel along with the appropriate ladder. Following cycling parameters were followed.

<table>
<thead>
<tr>
<th>Steps</th>
<th>Temperature</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial denaturation</td>
<td>94°C</td>
<td>1 min</td>
</tr>
<tr>
<td>Amplification(35 cycles)</td>
<td>Desired Ta</td>
<td>30 sec.</td>
</tr>
<tr>
<td></td>
<td>Extension(72°C)</td>
<td>2 min.</td>
</tr>
<tr>
<td>Final Extension</td>
<td>72°C</td>
<td>7 min</td>
</tr>
<tr>
<td>Store</td>
<td>4°C</td>
<td>∞</td>
</tr>
</tbody>
</table>

Table 5.3 Thermal cycling parameters for RACE PCR

5.2.1.2. Gel elution and cloning of RACE products

Upon successful amplification and gel image documentation, 214approx.214d bands were eluted and concentrated using the HiPurA Agarose Gel DNA Purification Spin Kit (Himedia Labs, Cat No. MB 511) and subsequently cloned into Pgem-T Easy vector system (Promega corp., Cat No. A1360) according to the instructions mentioned therein (Table 5.4).
Components | Volume
--- | ---
2X rapid Ligation Buffer | 2.5 µL
Template (RACE 215pprox.215d) (10ng/ µL) | 1.0 µL
Pgem-T Easy(50ng/ µL) | 0.5 µL
T4 DNA ligase(3U/ µL) | 0.5 µL
Water | 0.5 µL
Total volume | 5.0 µL

Incubate at 16°C overnight and store at -20°C until further use.

Table 5.4 Components of Ligation set up

5.2.2. Transformation of Competent cells by recombinant pGEM-T Easy

The ligation mix was mixed with 20 µL of ElectroMAX DH10β *E.coli* Cells (Invitrogen corp, Cat. No. 18290-015) and electroporated using the Eppendorf Electroporator (Eppendorf Inc., Germany, Cat No. 4307 000.658 ) at 2200 V in accordance with their respective recommended guidelines.

The transformants were plated in LB solid media with Ampicillin (100µg/Ml), X-Gal (40 µg/Ml) and 0.5Mm IPTG and incubated overnight at 37°C. A single white colony was marked and part of it was used as the template for colony PCR (Table 5.5) for confirmation and identity of insert. PCR profile same as outlined in 5.2.1.1.

Components | Volume
--- | ---
10X High fidelity PCR buffer | 1.0 µL
2.5 Mm Dntp | 0.8 µL
50Mm MgSO₄ | 0.4 µL
5’Gene Racer primer(10 µM) | 0.5 µL
Reverse GSP(10 µM) | 0.5 µL
White recombinant Colony* | 0.0µL
Platinum Taq. High fidelity (5U/ µL) | 0.06 µL
Nuclease free Water | 6.73 µL
Total | 10.00 µL

Table 5.5 Components of colony PCR*Part of the identified colony is picked by a sterile toothpick and inoculated in the PCR mix as template.

Upon electrophoresis in 1XTAE EtBr gel for correct identification of insert, the same colony was considered for plasmid extraction.
5.2.3. Plasmid Extraction

Plasmid extraction was carried out according to the alkaline lysis mini preparation protocol (Sambrook et al., 1989). The extracted plasmid was run in an 8% TAE EtBr gel and then quantified before sequencing (fig. 5.3).

![Fig.5.3. Purified plasmids from transformed bacterial cells.](image)

5.2.4. Sequencing

The recombinant plasmids were sequenced using Sangers dideoxy cycle sequencing method (Sanger et al., 1977) with the help of M13 forward primer (5'CGCCAGGGTTTCCCAGTCACGAC 3'). The process incorporates two steps-

a) Cycle sequencing PCR using the Big Dye terminator chemistry (Table 5.6).

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Big Dye Terminator V.3.1 ready reaction mix</td>
<td>1.0 µL</td>
</tr>
<tr>
<td>(1/8th. Reaction) (ABI, P/N. 4336197)</td>
<td></td>
</tr>
<tr>
<td>5X reaction buffer</td>
<td>1.5 µL</td>
</tr>
<tr>
<td>M13F primer(2pmoles/ µL)</td>
<td>1.6 µL</td>
</tr>
<tr>
<td>Plasmid template</td>
<td>1.5 µL</td>
</tr>
<tr>
<td>Nuclease free water</td>
<td>4.4 µL</td>
</tr>
<tr>
<td>Total volume</td>
<td>10.00 µL</td>
</tr>
</tbody>
</table>

Table 5.6. Components of cycle sequencing

The reaction volume is spun briefly for subsequent thermal cycling:

96°C- 1 min
96°C- 12 sec 
60°C- 4 min 
4°C- ∞ \{ 25 cycles \\

**Purification of sequencing reaction products:**

1. Transferred the reaction product in a 1.5 Ml tube.
2. Prepared the master mix I (10 µL of nuclease free water+ 2 µL of 125Mm EDTA Ph.8.0) per 10 µL PCR product volume.
3. Prepared master mix II (2 µL NaOAc Ph 4.6 + 50 µL 95% reaction) per 10 µL PCR product volume.
4. Add MM I and II to PCR product and mixed by inversion. Allowed to stand for 10 min at room temperature.
5. Spun @12,000g for 20 min. at room temperature.
6. Decanted supernatant and allowed to drain in tissue paper.
7. Washed with 70% etOH and spin @12000g for 10 min at room temperature.
8. Air dried and added 12 µL of Hi-Di formamide.

b) Electrophoresis of the PCR product for sequence analysis. (ABI 3130xl Genetic analyser): For sequencing the PCR products ABI 3130xl genetic analyser was used and the resultant generated sequence was visualized by ABI sequence scanner software V1.0 and trimmed for vector and M13F primer sequences manually. The trimmed RACE generated DNA sequence was then aligned with the original template sequence for overlapping regions to obtain complete Mrna sequence.

5.2.5. Heterologous Expression

5.2.5.1 Identification of ORF

The RACE derived complete sequence was fed into NCBI's online ORF finder tool (http://www.ncbi.nlm.nih.gov/projects/gorf/) to identify potential and inherent sense ORF(s).
5.2.5.2 Primer designing for ORF cloning into pET Vector.

The primers were designed according to the guidelines provided by Pet-Ek/LIC 43.1 (Novagen Inc., Cat.No.71072), with the objective of amplifying the respective ORF along-with adaptor sequences to facilitate efficient ligation of the same within the MCS of the vector (Fig.5.4, 5.5 and 5.6)

![Fig. 5.4 pET-43.1 Ek/LIC Vector map](Image)

![Fig. 5.5. Diagram of Ek/LIC strategy](Image)

Fig.5.5. **Diagram of Ek/LIC strategy.** After amplification with primers that include the indicated 5’ LIC extensions, the PCR inset is treated with LIC-qualified T4 DNA polymerase(-Datp), annealed to the Ek/LIC vector, and the resultant nicked, circular plasmid is transformed into competent *E.coli.*
Fig 5.6 **pET-43.1 Ek/LIC cloning/expression region showing the N and C terminal tag.** The N terminal tags comprise of Nus.Tag, His.Tag and S.tag accounting for 498 amino acids (approximately 54.78 kDa).

The primers were designed with the following common adaptor sequence:

**Sense primer:** 5’ GAC GAC GAC AAG ATX*–insert-specific sequence 3’

**Antisense primer:** 5’ GA GGA GAA GCC CGG TXX**–insert-specific sequence3’

*The first nucleotide of the insert-specific sequence on the sense primer must complete the codon ATX resulting in Met (X=G) or Ile (X=A, C or T).

**Since a C-terminal tag is not desired, an in-frame stop codon in the insert-specific sequence is included in the antisense primer.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PET_RA_SP</strong></td>
<td>5’GAC GAC GAC AAG ATG CAA AGG AAG AAG AAC GAG AAT GTG</td>
</tr>
<tr>
<td><strong>PET_RA_ASP</strong></td>
<td>5’ GA GGA GAA GCC CGG TCA AAG CAT TGG AAG CCA CTT CTG GTT</td>
</tr>
<tr>
<td><strong>PET_ERTC_SP</strong></td>
<td>5’GAC GAC GAC AAG ATG TCA GGA GTT GGT CGG ATC GCA CAA A</td>
</tr>
<tr>
<td><strong>PET_ERTC_ASP</strong></td>
<td>5’ GA GGA GAC CGG TCA GGT CGG TGG CAT TAT CAT GAG AA</td>
</tr>
<tr>
<td><strong>PET_FDH_SP</strong></td>
<td>5’GAC GAG GAC ATG AGG ATG AGG ATG CGT GTT GCC T</td>
</tr>
<tr>
<td><strong>PET_FDH_ASP</strong></td>
<td>5’ GA GGA GAA GCC CGG CTA TGA GCA TGA ATG GCT GCA ACA A</td>
</tr>
<tr>
<td><strong>PET_F1_SP</strong></td>
<td>5’ GAG GAC GAC AAG ATG AGG AAG AGG AAG TTG CAT TAT CAT CTG GTT</td>
</tr>
<tr>
<td><strong>PET_F1_ASP</strong></td>
<td>5’ GA GGA GAA GCC CGG TTA AAA CAA GAT GCT GTC AAA</td>
</tr>
</tbody>
</table>

**Table 5.7 pET primer sequences for peptide specific amplification of encoding ORFs.**
The above primers were used to amplify ORFs (Table 5.7) from the RACE template using the following PCR profile (Table 5.8)

<table>
<thead>
<tr>
<th>PCR step</th>
<th>Temperature</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial denaturation</td>
<td>94°C</td>
<td>1 min</td>
</tr>
<tr>
<td>Amplification(35 cycles)</td>
<td>Desired Ta</td>
<td>30 sec.</td>
</tr>
<tr>
<td></td>
<td>Extension(72°C)</td>
<td>2 min.</td>
</tr>
<tr>
<td>Final Extension</td>
<td>72°C</td>
<td>7 min</td>
</tr>
<tr>
<td>Store</td>
<td>4°C</td>
<td>∞</td>
</tr>
</tbody>
</table>

Table 5.8 Thermal cycling parameters for PCR based amplification of ORFs

5.2.5.3. Ligation Independent cloning (LIC)

Amplicons are ligated to the pET Ek/LIC-43.1 vector (Novagen, 71072-3, Germany) according to the manufacturer’s instructions.

5.2.5.4. Transformation of ligation mixture into primary cloning host

1. NovaBlue Giga Singles™ Competent Cells (Cat. No. 71127) were provided in Ek/LIC Vector Kits and were used as a convenient cloning host because the recA endA mutations facilitate high transformation efficiency and high yields of excellent plasmid DNA.

2. One aliquot of competent cell (in supplied snap cap tube) was thawed by holding within the fingers.

3. 1 µL of annealing reaction was mixed gently with the above and incubated in ice for 5 min.

4. Cells were transformed by placing the tube at 42°C for exactly 30 min. in a water bath. Thereafter, the tube was immediately quenched in ice.

5. 250 µL of the SOC medium was gently added while in ice and transferred into a 10 ml. glass test tube and shaken at 225 rpm at 37°C for 1 hour, before plating on the LB selection medium.
6. To select for transformants, a portion of the transformation mix was plated on medium containing antibiotic for the plasmid-encoded drug resistance (50 µg/ml ampicillin for Amp resistance marker). (Fig. 5.7)

7. Colony PCR for confirmation of insert was performed using respective Pet primers (see 5.2.2)

![Fig. 5.7. Transformed primary host colonies containing the pET construct.](image)

### 5.2.5.5 Plasmid purification from NovaBlue Giga Singles™ Competent Cells

(For detailed protocol see 5.2.3)

### 5.2.5.6. Transformation of BL21 (DE3) pLysS Competent Cells (Expression host)

Transformation was carried out according to the guidelines enumerated in section 5.2.5.4. Transformed cells were plated in a selection media containing Chloramphenicol 34 µg/Ml (bacterial selection marker) and Ampicillin 50µg/Ml and incubated at 37°C overnight. The recombinant colony was marked and part of it was used as template for colony PCR (as mentioned in 5.2.1.4). Upon confirmation, the remaining part of the colony was considered for expression.

### 5.2.5.7. Protein induction

The remainder of the marked colony was picked with a sterile toothpick and inoculated in LB broth with Cam₃₄Amp₅₀ and incubated in shaker incubator for 14 hours at 37°C. Two 50µL aliquots of the same each inoculated were in two separate conical flasks containing 50 Ml of LB broth with Cam₃₄ and Amp₅₀. The broth was incubated in shaker incubator at 37°C till O.D. reaches 0.6. Thereafter, protein induction was initiated by adding IPTG (0.1, 0.5 or 1.0Mm final conc.) while the other
with same amount of sterile water (under sterile conditions). The media was incubated at 30°C /220rpm for all the proteins studied. Thereafter, 1.0ML aliquots of the same were collected and pelleted from both induced and control samples at the end of 5th and 10th Hour intervals respectively. O.D. of each of the aliquots was also quantified. The labelled pellets were stored at -20°C till SDS-PAGE analysis and protein purification.

5.2.5.8. SDS-PAGE analysis of induced proteins.

The discontinuous and denaturing sodium dodecyl sulphate-Poly acrylamide gel electrophoresis (SDS-PAGE) technique was used for comparison of total protein profiles between induced and control bacterial pellets (Shapiro et al.,1967).Total bacterial proteins from both induced and control cell pellets was run in SDS-PAGE system (V-GES vertical gel electrophoresis system, WEALTEC) with 5% stacking and 10% separating gels in accordance with the instructions and guidelines recommended by the manufacturer. Sample volumes (bacterial total protein) with O.D=2 and 15 µL of purified protein were loaded alongwith protein ladders (Bangalore genei, cat no. RPMWM; New England Biolabs cat no.P7703S) and run at a constant 60V for 3 hours. Gels were then stained using Comassie brilliant blue solution (R-250) (Himedia labs, Cat. No.443282L) for one hour and then de-stained for one hour each by gentle rocking in destaining solution I and II subsequently (Table 5.9). The gel image is documented by digital white light photography.

<table>
<thead>
<tr>
<th>Staining solution(500ML)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Comassie brilliant blue(R-250)</td>
<td>1.25 g</td>
</tr>
<tr>
<td>Methanol</td>
<td>250ML</td>
</tr>
<tr>
<td>Acetic acid</td>
<td>35ML</td>
</tr>
<tr>
<td>Water</td>
<td>215ML</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Destaining solution I (500ML)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methanol</td>
</tr>
<tr>
<td>Acetic acid</td>
</tr>
<tr>
<td>Water</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Destaining solution II (500ML)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methanol</td>
</tr>
<tr>
<td>Acetic acid</td>
</tr>
<tr>
<td>Water</td>
</tr>
</tbody>
</table>

Table 5.9 Composition of staining and de-staining solution for PAGE gels.
5.2.5.9. Single protein isolation and purification.

Four induced bacterial pellets were processed and purified using the components and recommendations furnished by Dynabeads His-Tag isolation and pulldown kit (Invitrogen Inc., Cat no. 101.03D). The isolated His Tag containing fusion protein was again run on SDS-PAGE system along with its corresponding induced and control total protein counterparts and a suitable protein ladder for easy comparison and analysis.

5.3. Results and discussion

5.3.1. Ricin Agglutinin like protein

5' RACE analysis (Fig 5.8) revealed full length Mrna sequence of the above protein, which is exclusively induced and upregulated in 111/1 following infestation by *Helopeltis theivora*. The details are given below-
Fig. 5.8 5’RACE PCR amplification for S3I_O18

The sequence contains the following ORF (333 bases) encoding a 111 amino acid long Ricin agglutinin like protein.
Translation in frame 3 (5′-3′) (111 amino acids)

\[
\text{MQRNENNVGLGDCVSNSQTDQKWAIYPDGSIRPQQNDVCVTSN}
\]

\[
\text{QNIVSCSYGWSSQRWVTNDCAILNLYNGMVMDVKSNSPILO}
\]

\[
\text{QIILWPFGNPNQKWLPM}
\]

BLAST P of the amino acid sequence revealed the presence of putative conserved domains as sugar binding sites belonging to Ricin superfamily.

The expect value for the closest protein namely “Putative type 2 Ribosome inactivating protein” (gb\|ACV60361.1\|) was found to be \(4\times10^{-55}\) with a bit score of 216. The alignment details of the following is given under.

\begin{verbatim}
>gb|ACV60361.1| putative type 2 ribosome-inactivating protein [Camellia sinensis]
Length=570

Score = 216 bits (551), Expect = 4e-53, Method: Compositional matrix adjust.
Identities = 103/570 (19%), Positives = 21/570 (3%), Gaps = 0/570 (0%)
Query 1 MQRNENNVGLGDCVSNSQTDQKWAIYPDGSIRPQQNDVCVTSNHFQGTNNIIVS556
Query 61 SSQRWFNTNDCAILNLYNGMVMDVKSNSPILOQQQILPNFPNGNSPNQKWL

Sbjct 450 MQRNENNVGLGDCVSNSQTDQKWAIYPDGSIRPQQNDVCVTSNHFQGTNNIIVS
Sbjct 520 SSQRWFNTNDCAILNLYNGMVMDVKSNSPILOQQQILPNFPNGNSPNQKWL
\end{verbatim}

Heterologous expression of the protein (111 amino acids, 12.21 kDa) was accomplished as a fusion protein along with N terminal tag (549 amino acids or approximately 60 KDa) matching with the theoretically predicted molecular weight of 12.65 kDa. Maximum induction occurred when the cell culture was incubated for 10 hours at 30°C with a final concentration of 1Mm IPTG in the media. Upon electrophoresis, the purified (fusion) protein migrated as a single unit depicting a molecular weight of approximately 72 kDa (Fig 5.9), probably due to structural features and amino acid sequence-composition.
Fig. 5.9. Heterologous protein expression of *Camellia sinensis* Ricin agglutinin.

With the objective of identifying domains and functional sites (if any) within the putative protein, its amino acid sequence was fed into the ExPASy-Prosite available in the ExPASy proteomics server (http://expasy.org/prosite/). The protein was found to contain Ricin Lectin domain of ricin B chain (Prosite ID: PS50231). Primary structure analysis has shown the presence of a similar domain in many carbohydrate-recognition in plant and bacterial AB-toxins, glycosidases or proteases (Hirabayashi et al., 1998; Hazes and Read, 1995; Hazes, 1996). This domain, known as the ricin B lectin domain, can be present in one or more copies and has been shown in some instance to bind simple sugars, such as galactose or lactose. The ricin B lectin domain is composed of three homologous subdomains of 40 amino acids (α, β and γ) and a linker peptide of around 15 residues (lambda). It has been proposed that the ricin B lectin domain arose by gene triplication from a primitive 40 residue galactoside-binding peptide (Rutenber et al., 1987; Rutenber and Robertus, 1991). The most characteristic, though not completely conserved, sequence feature is the presence of a Q-W pattern. Consequently, the ricin B lectin domain has also been referred as the (QxW) 3 domain and the three homologous regions as the QxW repeats (Hazes and Read, 1995; Hazes, 1996). A 226 approximately 226 de bond is also conserved in some of the QxW repeats (Hazes and Read, 1995).

The 3D structure of the ricin B chain has shown that the three QxW repeats pack around a pseudo threefold axis that is stabilised by the lambda linker (Rutenber et al., 1987). The ricin B lectin domain has no major segments of a helix or β sheet but
each of the QxW repeats contains an omega loop (Rutenber and Robertus, 1991). An idealized omega-loop is a compact, contiguous segment of polypeptide that traces a 'loop-shaped' path in three-dimensional space; the main chain resembles a Greek omega.

In view of the above, it can be concluded that the protein under consideration is a kind of Ribosome inactivating type 2 protein with a Ricin B lectin domain. It needs mentioning that the protein is only 92% similar to its reported nearest homologue and therefore can be considered a new variant of same. Further structural analysis of the protein will enable better understanding of any functional ramifications due to such variation.

The 3D structure of the protein was predicted using Phyre² homology modelling server ([http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index](http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index)) (Kelley and Sternberg, 2009). The main results of the prediction suggest that the modelling of the protein is based on chain B of a PDB hydrolase molecule “Type 2 Ribosome inactivates protein cinnamomin” with a confidence and coverage of 100%.

**Secondary structure prediction**

![Secondary structure prediction of *Camellia sinensis* Ricin agglutinin chain B](image)

The N terminal region of the protein showed higher degree of unpredictability with respect to its reference model, since such regions may have important functional roles (Kelley and Sternberg, 2009). Based on the above prediction, the following 3D
model of *Camellia sinensis* Ricin agglutinin chain B has been proposed using the Phyre2 server (Fig. 5.11)

![Fig. 5.11 Predicted model of *Camellia sinensis* Ricin agglutinin chain B]

**5.3.2. Ethylene Responsive Transcriptional Co-activator like protein (ERTC)**

5′RACE PCR of S5P3A10 subtracted EST (445 bases) yielded full length cDNA counterpart of its mRNA containing 730 bases.
Fig. 5.12 Amplification from 5’RACE PCR for *Camellia sinensis* ERTC like protein

The CDS for the Mrna was found to posses 249bp ORF encoding 82 amino acids long protein.
BLAST P analysis revealed the presence of a putative conserved domain belonging to MBF1 (multiprotein bridging factor 1) superfamily with most nearest significant alignment with ethylene-responsive transcriptional coactivator \( [Elaeis \; guineensis] \) gb|ACF06517.1|. This is the first report of heterologous expression in tea. Multiprotein bridging factor 1 domain is found in the multiprotein bridging factor 1 (MBF1) which forms a heterodimer with MBF2. It has been shown to make direct contact with the TATA-box binding protein (TBP) and interacts with Ftz-F1, stabilising the Ftz-F1-DNA complex. It is also found in the endothelial differentiation-related factor (EDF-1). Human EDF-1 is involved in the repression of endothelial differentiation, interacts with CaM and is phosphorylated by PKC. The domain is found in a wide range of eukaryotic proteins including metazoans, fungi and plants. A helix-turn-helix motif (HTH)(pfam01381) is found to its C-terminus. The domain is an evolutionarily conserved transcriptional coactivator that connects a regulatory factor and TATA element-binding protein (Takemuru et al., 1991) indicating that S5P3A10 to be a yet another ortholog in tea.
Heterologous expression of the protein (82 amino acids, 9.02 kDa) was accomplished as a fusion protein along with commercially available Tag (Fig. 5.13). Maximum
induction occurred when the cell culture was incubated for 10 hours at 30°C with a final concentration of 1Mm IPTG in the media. The purified (fusion) protein migrated as a single unit when electrophoresed depicting a molecular weight of approximately 63800 daltons (or 9020 daltons upon enterokinase treatment and separation from the fused Tag protein).

Fig. 5.13. Heterologous expression of *Camellia sinensis* ERTC like protein

**In-silico Homolgy modelling**

The 3D structure of the protein was predicted using Phyre2 homology modelling server (http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index) (Kelley and Sternberg, 2009). The main results of the prediction suggest that the modelling of the protein is based on Methionine synthase activation domain-like fold belonging to Methionine synthase SAM-binding domain family with a confidence of only 12.4% and coverage of 33% (27 residues or 33% of protein sequence have been modelled with 12.4% confidence by the single highest scoring template). This suggests that the protein structure doesn’t have any homologue in public domain, since 63% of the sequence is predicted disordered, which are critical to the understanding of functional role of the same (Kelley and Sternberg, 2009). Phyre2 based secondary structure prediction of the amino acid sequence demonstrates the location and spectrum of disordered regions across the sequence (Fig.5.14). Further structural studies are therefore necessary for better understanding of structure-function relationship.
Based on the above prediction based available templates, the following partial 3D model of the protein has been proposed (Fig. 5.15) using the Phyre2 server, which is incomplete.

![Secondary structure and disorder prediction](image)

**Fig. 5.14** Predicted Secondary structure of *Camellia sinensis* ERTC like protein

![3-D model](image)

**Fig. 5.15** 3-D model of a *Camellia sinensis* ERTC like protein. Image coloured by rainbow N → C terminus

### 5.3.3. Formate Dehydrogenase like Hypothetical protein

5' RACE analysis revealed full length Mrna sequence of the above protein, which is exclusively induced and upregulated in 111/1 following infestation by *Helopeltis*
theivora. The PCR amplified a 558bp region upstream to the original subtraced sequence (Fig.5.16). The details are given below-

>GA3130XL_S3(R)_E12 (Subtracted sequence) (371 bases)

GAAGGAGCAAAATTGTAACGTGAAACATATTCTCTGATCTGTCATCTCATAACTACGCCATT
CCATCCGGCCTATGTGTTACAGGGGAAAGGATCAGAAGAGCTAAAATTACACATGTCAT
GCTGGAATTGGCTCTGACTATTTGATTCTGCAAGCTGAGCTGAGTGATTACAGTTGGAAG
AGGTCACTGGAAGCAATGTGAAGTGTCAACTGCAAGCTGAGTGAATATGTTCTCGT
TAGGAATTTTCTTGCCCTGGATAACCATCAAGTATCAATGCGAGTCAAGCTTATGCT
CATAAGGCGTTATTTTATTTTTATAACATTTGTTAACAAAATCTGTTTAAAAA

> GA3130XL_5*RACE amplicon (558 bases)

CGACTGGAGCACGAGGACACTGAAGCGGGGAAGTGAATACCGGATAATATCATACAACACAGTAA
TAGAGCTTTGGAAGGTCTGATCTCCTCATTCTCCTCTCTATCGATTCTCGGCCGAGGCTT
TACCAAATTCGATTCCGATTCGGATTCGGAGATGCGATGATGAGCGTGGTCTCGACCGCA
GTTCGTTCTTTTTGCTCTCTGGAGTTCAATTTTCTCTCAATCCAAAACCGAAGACCCTCGAT
TCACCTGGAAAGCAAAAAGGATAGTTGGAGTGGTCTACAAAGCACCACTGGAATGGAACAT
CGAATTTTGGGCTGTCGAGAGGATGATTGCGGATCATCGACTGTTGAGGATCAAGAGCCAA
CCAACTCACCATGGACGACAAAAGAGGACCAAAATTGTAACCTGAGAAAATCTATTTCTGAT
CTGCACTGTCCTCATACACTCGGACTCCATCGGCCGCTATGTTACAGGCGAAAGGATCAAGAGG
CTAAAAAACTACACTGGAAGGAGCTGGAATTGCGTGAG

Bases in green= 5*gene racer primer
Bases in pink= 5* gene racer GSP reverse primer binding site
Bases in blue= Overlapping region

> 5* RACE product for S3(R)_E12

GGGGGGAAAGTTGAAATACGGAATATATCATACAAACGAGTAATAGAGCTTTGGAAGTTTCTGTATCT
CCATCTTTTCTCTATCGATTTCGCTGCGAGAGCTTTACAAATTCGATTCCGATCCGAT
TCGAGATGGCAGATTGAAAGGCTGTTGCTCGACCCGAGTTGCTGTTTCTCTCTG
GGAGCTTTCTACAAAGGCAATACACCTGAAATATCCGAATTTTGTGGGCTGTCGAGAG
AGCACTGTCCTGAGCTGACTGCTTGGGAAATCACAAGGCGACCAATACATTTGTCAACCGATCGAA
> Full length mRNA_S3(R)_E12 (760 bases without poly A)

GCGGGGAAGTGTAACGGAATAATATATATACAAACACGTAATAGAGCTTTGGAAGTTTCTGTATCT
CCATCTTTTCTCTATCCGAGATCTTCTGGGAGAGCTTTACCAATCCGATTCCGATCAGTT
TCCGAGATGGCCGATGATGAAAGGCTTGTCCTCCGACGCCAGTTCTGCTCTTCTCTCCGAG
GGTCAATTTTCTCTGATATCCTCCATCACACCCATTCATCCTCTGTTTCTGAGTTAAAAAG
GGATGTTCTACAGGAAATAGACTGCAAAAATGGAATCCGAATTTTTGTGCGCTGTCGGCGAGA
GACGGATGACGATGCTGGTGTGGAATCAAAGACACACACAAAATATACTGCTGTCTAAC
AGCTGGCAATTTGCGCTGTATGATGTTAGATGCAGCTGGCAGCTGCTGCTGGAATTTAGC
GAGTCATCTGGAACGCTAGTCTCTCATGAGGATGAGTCATGAAATTCTACTTCTCTCG
TTAGGAATTTTCTGCTGGGATACCAAATAGTTAACATCAAGTGGAAGTTGCTGCTAGCATTTG
TCATAGAGGCTTTATTTTTATACATTTGTTGTTACAAAATACGTTTTTTAAAAAAAAAAAAA
AAA

> COMPLETE ORF FOR FDH LIKE PROTEIN WITHIN S3(R)_E12 mRNA (576 BASES)

ATGGCGATGATGAAGGCTTGTGCTGCAAGCCGAGTTTTGCTCTGCTCTGGAGTTCAAAT
TTCTCTCTCATACTCCAAAGCAAGACTCCACCTGCTGTCTCCAGTGAAGAAAAGATAGTTGGAGT
GGTCTCAAGGAAATGATAGCTGCAAAATGGAATCCGAATTTTTGTGCGCTGCTAGCAGATT
GGCATCGTACGTGTTGGAATACACAAGGCAACATAATCTGTCAAGCAGTAGCAAAAGAAGG
ACCAATGTTGGAATCGCGAGAGAGATATGCTGTCTCCATGTAATACAGGTCGATGTTG
CGCCTATGTACAGCGAAGAGCTAAGAATCTGCTGCTCGCAAGTTGGAATTTAGTGGAGTTGCT
ACTGGAAGCATTGTAATCGTCAGTGCCAGATGATAGCTGCAAGAATTTCTACTTCTCTGTAGGA
ATTCTTGTCCCTGGATACCATCAAGTTATCAATGGGAGATGTTGCTGCTAGCTATTAG

> TRANSLATED PROTEIN FROM S3(R)_E12 ORF (192 amino acid)

| A A M K R V A S T A V R A F A S S G S S I S S I P T R D L H A S P G S |
| G H Q Y I V T D K E G P N C E L K H I P D L H V L L I T T P F H P A Y V T A |

558 bp RACE amplicon

100bp ladder

Fig 5.16. 5'RACE PCR amplification for *Camellia sinensis* FDH like protein
>BLAST P analysis and alignment details of the nearest hit

Gene: XP_002278444.1
Predicted: hypothetical protein [Vitis vinifera]
Emb|GBI34925.3| Unnamed protein product [Vitis vinifera]
Length = 303

Gene ID: 100261093 LOC:100261093 [Vitis vinifera] (10 or fewer PubMed links)

Score = 313 bits (802), Expect = 7e-86, Method: Compositional matrix adjust.
Identities = 158/183 (86%), Positives = 168/183 (92%), Gaps = 3/183 (2%)

Query 1

MAMMKRVASTAVRAFASSGSSSIISSSIPTRDLHASPGSKKIVGYFVYKANEYMNNNFVFVC
MAMMKRVA +AVRAFA +8  T+ LHASGKKIVGFYKANEYAMNNNFVFVC

Subject 1

MAMMKRVASTAVRAFASSGSSSIISSSIPTRDLHASPGSKKIVGFYKANEYAMNNNFVFVC
MAMMKRVA +AVRAFA +8  T+ LHASGKKIVGFYKANEYAMNNNFVFVC

Query 61

AELRGIDWLESGQGHQYTVTDKKEFLKHKHFFVLPFHPHAAYVTAERI1KAK
E ALGIDWLESGQGHQYTVTDKKEFLKHKHFFVLPFHPHAAYVTAERI1KAK

Subject 58

VEALRGIDWLESGQGHQYTVTDKKEFLKHKHFFVLPFHPHAAYVTAERI1KAK
VEALRGIDWLESGQGHQYTVTDKKEFLKHKHFFVLPFHPHAAYVTAERI1KAK

Query 121

NLQLIITAGIGSDHDLQAAAAGLTVAETGSNVVSVAEDELRILLLYLRNFLFG+HQQV
NLQLIITAGIGSDHDLQAAAAGLTVAETGSNVVSVAEDELRILLLYLRNFLFG+HQQV

Subject 118

NLQLIITAGIGSDHDLQAAAAGLTVAETGSNVVSVAEDELRILLLYLRNFLFG+HQQV
NLQLIITAGIGSDHDLQAAAAGLTVAETGSNVVSVAEDELRILLLYLRNFLFG+HQQV

Query 181

ING 183
I+G

Subject 178

ISG 180

Conserved domains on [K1582]

Graphical summary

List of domain hits

Blast search parameters

Date Source: Live blast search RID = V0703BEA411
User Options: Database: clustalw-v2.2 Low complexity filter: yes E-value threshold: 0.01 Maximum number of hits: 500
Conserved domain search revealed the presence of six individual types of domains (mentioned in the chart above). The formate dehydrogenase (FDH, 1.2.1.2.) catalyzes the NAD-linked oxidation of formate to CO$_2$, and FDH transcript accumulation has been reported after various abiotic stresses.

Heterologous expression of the protein (192 amino acids, 21.12kDa) was accomplished as a fusion protein along with N terminal tag. Maximum induction occurred when the cell culture was incubated for 10 hours at 30°C with a final concentration of 1Mm IPTG in the media (Fig. 5.17). The purified (fusion) protein migrated as a single unit when electrophoresed depicting a molecular weight of 237pprox..75900 daltons (or 21120 daltons upon enterokinase treatment and separation from the fused tag protein), when run along with a 10-250 Kda protein molecular marker (NEB protein ladder, cat no. P7703S).

![Fig. 5.17 Heterologous expression of Camellia sinensis FDH (P=Purified, M=marker, IN=induced total protein,Molecular weights in KDa)](image)

**In-silico Homology modelling:**

The modelling for the protein was done using the template d2naca2, a flavodoxin like fold, belonging to Formate/glycerate dehydrogenases, substrate-binding domain, with a prediction confidence of 100% and coverage 69% (133 residues, 69% of sequence) have been 237 approximately with 100.0% confidence by the single highest scoring template. Thirty one (31%) of the amino acids could not be predicted.
for their models, being concentrated primarily near the N terminus. Further structural studies are therefore necessary.

Fig. 5.18 Secondary structure prediction of *Camellia sinensis* FDH

Based on the above secondary structure prediction (Fig. 5.18), the following probable 3D model of the protein could be proposed (5.19)
5.3.4. Leucine Rich Repeat (LRR) like protein

5’ RACE analysis revealed full length Mrna sequence of the above protein, amplifying a 563bp region (Fig.5.20) upstream to the original subtracted sequence. The protein is known to play important role in disease resistance and stress mitigation in plants. The details are given below.
>GA3130XL_S5_P2F1_pGEMT_M13F3.2_TB_4.2.09_016.ab1(Subtracted sequence, 314 bases)
ACCTGGATCATATAATCATACATTATCTGGGAGAACATCCGCGATGCCACACTTTTCTGGAAA
GAAATGTATTGGAAGAAATGCTTTTCGCGCCAGGTGTAAGACTATTATGCGTC
AACAATGTCTCGAAGTTTTGACACAGAGTTTCTTGTGTTTATTGCTAAAGCATTAGGTGTTTG
AGTTGGAAATGGAATTGAGTTCTATAAATAGGATGTATATAATGTCACACTAGTGTTTTGGAAAATTAAATTTAAGAAATTTTGGGAAAGTAGTTTGGTGGGACAAA

>GA3130XL_pGMTEZ_M13R_TB_S5P2F1_800bp (5'RACE amplicon_563 bases)
CGACTGGAAGCAGCGAGGACACTGAAGACGATGGCGAGCAGCGAGCGAGCGACGACG
GTGAGCATCGAAGTGACGATAACTAGCAGCCAGGCGAGCGGTGTTGTTGGCCAGAGTCA
GATAACCGCCGATCCACCCTCCCGATTTGAGCTATTAGGGTTTTGCCC
CAAAAACGAAACCTGGCAACGCGACTGAGAATTATTATTGCTTTTGTGATGA
AGCAGGGTCTCGTTCCAGGACAAATTCTTCAAGCGTGACTACTCTTCCACTTCTCT
CGCTTCTTTCTTCTACATCCCCTCGTGCTCTCCGCTCCTCCAAATTCATCTCCTCT
CACTTTTCTCTCATCTTTCATCTCGGGGAAATTTCAAGAGCCTTTGCAGTTTTCTG
TATTGACTGTTGTCAAGATGGTGAAGTGGAGGATGGGGAGGAACTTTCCATTGCA
GATACTCTTTTTGCTAAGATTTTTCTTGTGCTAGTTGTTGCTATTAGTCACTAGATTA
TCACTTGACTCTCTGGTTACACGTTGGTTATTCTAGGTACCTGGATCATAATCATACATTATCTGGGAGAACATCCGCGATGCCACACT
TTACTGGGAGAACATCCGCGATGCCACACT

Bases in green= 5'gene racer primer
Bases in pink= 5' gene racer GSP reverse primer binding site
Bases in blue= Overlapping region
>GA3130XL_S5_P2F1_TB_4.2.09_016.ab1 (Complete mRNA, 812 bases)

GGACGATGGCGAGCAGCAGCAGCGAGGTCAGCTGACTGACGATAACTA
GCAGCGCGACGCGGTTGCTGCGTTCAGATACCTCTCTGCGATTCAACCGTC
CCTCGGATTAGCTATAGGTTGGGCCCCAAAACGAAAACCTGGGCAAGCGGG
ACTGAGATATTATTTGCTTTTGTTGATGAACACGCGGTGCTGCCAGGACAAATT
CTTCAGGCGGACTACGCTCTCCTCACTGCTTGTTCTTTCTCTGACTACATCCC
CGCTGCTCAGGCTCTCCAAATTCTGCTCTCCTCTACTTTCTCTCTACTTTCTTC
GGGAATTCTAACGAGCTCTTGGATTTCTGTTGACTGTTGCTCAAGAATGGTGAA
AAATGAGGGCATGGAGGAGCCCTCATGTCGAGATCTCTCTTTTTGGCTAGATTT
TTCTTGTTGCCATTGCTTTAGTCAAGTTATCAACTGCTCTGTGACAGCTTT
GGTATTTCTAGGTACCTGGAATCATAATCAATTTACTGGGAGAATCCCGGATGCA
CAGTTTTTCAGAGAAAGATGTATATTGAGGAAGAATGCTTTCCGCGGACAGGTG
ACTATTGCGCTCAACAATGTCTGAGGCTTTGGACACAGAGTTCTTTTTAG
TCAAGGCTTCCACATATTATTTTTCTCGTAGAAATTTGTTTTTTGTCTGCTAAA
ACTAGGGTGTGTGAGTTGAGTTGAAATTGGAATATGATTCTTAAATTAGATGTATATAA
TGTACACTAGTTGTTTGGGAAATTTTAAGAATTTTTGGAAGTAGGTTGTTGACAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

>S5P2F1_frame 1 CDS (276 BASES)

ATG GTG AAA GTG GAG GCA TGG GAG GCC TTC ATT GCA GAT ACT CTC TTT
TTT GCT AAG ATT TTT CTT GTT GGC ATT GCT TTA GTC ATG GAT TAT CAC
TTG ACT CTC TGG TAC ACG TTG GTA TTT CTA GGT ACC TGG ATC ATA ATC
AAT TTA CTG GGA GAA TCC CGG ATG CAC ACT TTC TGG AAA GAA ATG TAT
ATT GAA GGA AAT GCT TTC CGG CCA GGT GTG AAA CTG ATT GGC GTC AAC
ATT GTC CTC GAA GTT TTT GAC ACA GAG TTC TGG TTT

>S5P2F1_amino acid sequence in frame 1 (92 amino acids)

MVKEAVWEAFIADTLFFAKIFLVGIALVMDYHLLWETYLTLVFLGTVIINLLGESRMH
TFLKEMYIEGNARFPGVKLIGVNNVLEVFDTEFLF
Upon BLAST P analysis, no conserved domain was found and without any significant match statistics with the most closest hit protein indicating significant variation of the protein in tea.

Heterologous expression of the protein was accomplished as a fusion protein along with commercially available tag (Fig. 5.21). Maximum induction occurred when the cell culture was incubated for 8 hours at 30°C with a final concentration of 0.5Mm IPTG in the media. The purified (fusion) protein migrated as a single unit when electrophoresed depicting a molecular weight of 242pprox..64900 daltons (or 10120 daltons upon enterkinase treatment and separation from the fused tag protein).

563 base RACE amplicon

100 bp ladder

Fig.5.20  5’RACE amplicon for S5P2F1

Upon BLAST P analysis, no conserved domain was found and without any significant match statistics with the most closest hit protein indicating significant variation of the protein in tea.

Heterologous expression of the protein was accomplished as a fusion protein along with commercially available tag (Fig. 5.21). Maximum induction occurred when the cell culture was incubated for 8 hours at 30°C with a final concentration of 0.5Mm IPTG in the media. The purified (fusion) protein migrated as a single unit when electrophoresed depicting a molecular weight of 242pprox..64900 daltons (or 10120 daltons upon enterkinase treatment and separation from the fused tag protein).
In silico homology modelling:

Phyre2 based homology modelling could successfully predict 15% of 92 amino acids protein by only 15.1% confidence with a 16% coverage. This suggests that majority of the protein structure remains to be elucidated and modelled. The predicted part of the protein was modelled on the chain A of the PDB molecule kalata b3/b6 bearing the fold library id c1wn8A. Moreover, transmembrane helices could be predicted (Fig. 5.22) from the protein sequence to adapt to the following topology shown below.
Secondary structure prediction (Fig. 5.23) showed that both the N and C terminals of the protein possessed regions of prediction which necessitates further enquiry into its structure. Such investigation will also help in understanding the functional role of the same in relation with its structure (Kelley and Sternberg, 2009).

Fig. 5.23 Secondary structure prediction of *Camellia sinensis* LRR like protein.

Based on the above interpretation, three dimensional model of the predicted part of the protein was generated using the phyre2 protein homology prediction software (Fig.5.24)

Fig.5.24. Proposed 3-D image model of the predicted region of the *Camellia sinensis* LRR like protein.
References-


