CHAPTER 4: Results

4.1 Picrosides are Present in Root, Rhizome and Leaf Tissues

Leaf as well root tissues were analyzed for the presence of picrosides content. Result showed that leaves consisted of highest P-I, 1.52 ± 0.24 followed by rhizome 1.11 ± 0.31 and root 0.07 ± 0.04, respectively per 100 mg of tissue fresh weight. Whereas, P-II content was highest in rhizome 1.94 ± 0.42 followed by root 0.31 ± 0.14 and leaves 0.14 ± 0.08 in 100 mg fresh weight, respectively.

At the same time it was important to know relative distribution of biomass in these tissues to have an estimation of overall distribution of picrosides. The biomass of *P. kurrooa* was estimated on fresh weight as well as on dry weight basis. Root, rhizome and leaves had 33.3%, 29.1% and 37.6% of the biomass on dry weight basis, respectively and 43.6%, 22.2% and 34.2% on fresh weight basis, respectively on 60 DOT in the same order (Table 4.1). At 120 DOT, these values were 16.3%, 37.9% and 45.8%, respectively on dry weight basis and 58.1%, 28.0% and 15.6% on fresh weight basis, in the same order. On 60 DOT the average weights of root, rhizome and leaves were 5.8 g, 3.1 g and 7.4 g which increased to 7.8 g, 13.9 g and 28.5 g, respectively on 120 DOT (Table 4.1).

Differential accumulation of P-I and P-II was detected in different tissues of the *P. kurrooa* (Table 4.2, Fig. 4.1a). Mass spectrometric analysis confirmed the identity of P-I and P-II in the corresponding peaks (Fig. 4.1b). Picrosides were present in the leaf tissues at all the node positions, with maximum content in the youngest leaf (Table 4.2). Comparative analysis of picrosides content in different tissues revealed significantly higher P-I content in the leaf tissues as compared to the rhizomes and roots. Roots showed higher amount of P-II as compared to P-I, though the total contents of P-I and P-II was less as compared to that in the leaves. In rhizomes, both P-I and P-II were present in comparable amounts.
<table>
<thead>
<tr>
<th>Growth stage</th>
<th>Biomass (in grams)</th>
<th>Biomass (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tissue</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Root</td>
<td>Rhizome</td>
</tr>
<tr>
<td>60 DOT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fresh weight</td>
<td>5.77 ± 1.56</td>
<td>3.08 ± 0.60</td>
</tr>
<tr>
<td>Dry weight</td>
<td>1.15 ± 0.31</td>
<td>0.89 ± 0.17</td>
</tr>
<tr>
<td>120 DOT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fresh weight</td>
<td>7.78 ± 1.16</td>
<td>13.91 ± 0.93</td>
</tr>
<tr>
<td>Dry weight</td>
<td>1.33 ± 0.19</td>
<td>3.13 ± 0.22</td>
</tr>
</tbody>
</table>

Table 4.1: Biomass and per cent distribution of biomass (on fresh weight and dry weight basis) in different parts of *P. kurrooa*. Data represents the average of eleven independent plants ± SE each, at two different growth stages of plants [60 and 120 DOT]. * Indicates statistical significant difference with corresponding root biomass, and ** indicates statistical significant difference with both corresponding root as well as rhizome biomass (p< 0.05)

Further, in order to find whether or not the picrosides accumulation varied during the day, picrosides content were estimated at four different time points of the day. Picrosides content were found higher during light period (at 9:00 h and 13:00 h) as compared to the low light/dark period (at 5:00 h and 17:00 h), though statistically the differences were not significant (Table 4.2). Analysis of picrosides content during different growth cycle of the plant showed its presence in all the tissues analyzed (root rhizome and leaf) throughout, though the content declined in senescing plants (Table 4.3). Picrosides content was also estimated in the plants collected from natural population at Rohtang pass (District Kullu, Himachal Pradesh, India). The root, rhizome and leaves had 0.10 ± 0.001, 1.22 ± 0.25 and 2.32 ± 0.64% P-I, respectively. It was difficult to assign specific age to the populations and compare the data of different locations in the present study, however, our experience showed that naturally growing population always exhibited presence of picrosides in the leaf tissue irrespective of age of the plant.
Fig. 4.1: a, HPLC elution profile of P-I and P-II standards, extracts of leaf, rhizome and root tissue of *P. kurrooa*; b, mass spectrum (in positive mode of ionization) of P-I (molecular weight, 492.47) and P-II (molecular weight, 412.46).
### Table 4.2: Picroside content in different tissues/organs of *P. kurrooa*. Leaf positions are numbered starting from the top designated as 1st leaf, the youngest leaf. Values are means of three separate biological replicates ± SE. *Light intensity* was measured as photosynthetically active radiation (PAR) using PAR sensor attached to an infra red gas analyzer (LI- 6400; LI-COR Biosciences, USA). Superscripted numbers 1, 2, 3, 4, 5, 6, 7 and 8 indicate that the corresponding value is statistically different with respect to value of 1st leaf, 2nd leaf, 3rd leaf, 4th leaf, 5th leaf, 6th leaf, rhizome, and root, respectively, in the same column (p< 0.05). No significant difference was observed among corresponding values in a row.

<table>
<thead>
<tr>
<th>Tissue</th>
<th>Developmental stage</th>
<th>Picroside content (mg/100 mg fresh weight)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Young (chl: 1.08±0.10 mg/g fresh weight)</td>
<td>Mature (chl: 0.98±0.04 mg/g fresh weight)</td>
</tr>
<tr>
<td></td>
<td>P-I</td>
<td>P-II</td>
</tr>
<tr>
<td>1st Leaf</td>
<td>2.40±0.165,8</td>
<td>0.26±0.155</td>
</tr>
<tr>
<td>2nd Leaf</td>
<td>1.52±0.218</td>
<td>0.14±0.087</td>
</tr>
<tr>
<td>3rd Leaf</td>
<td>1.43±0.32</td>
<td>0.16±0.097</td>
</tr>
<tr>
<td>4th Leaf</td>
<td>1.05±0.18</td>
<td>0.06±0.037</td>
</tr>
<tr>
<td>5th Leaf</td>
<td>0.86±0.39</td>
<td>0.06±0.037</td>
</tr>
<tr>
<td>6th Leaf</td>
<td>1.05±0.49</td>
<td>0.19±0.118</td>
</tr>
<tr>
<td>Rhizome</td>
<td>1.11±0.31</td>
<td>1.94±0.421,2,3,4,5,6,8</td>
</tr>
<tr>
<td>Root</td>
<td>0.07±0.04</td>
<td>0.31±0.141</td>
</tr>
</tbody>
</table>

### Table 4.3: Picroside content during different stages of development, in leaf, rhizome and root of *P. kurrooa*. Values are means of three separate biological replicates ± SE. In case of leaf tissue, the leaf at position 2 (from the top) was selected due to its presence in the early as well as late stage of plant growth. Young, mature and senescent represents 30, 120 and 180 DOT. Numerical values of chlorophyll (chl) are mean±SE of ten independent biological samples (2nd leaf) at all three stages (young, mature and senescent). Superscripted numbers 1, 2 and 3 indicates that the value is statistically different with respect to value of 2nd leaf, rhizome and root respectively, in the same column (p< 0.05).
Chapter 4

4.2 Pathway Specific Inhibitors Suggests the Predominant Role of MEP Pathway in Picrosides Biosynthesis

Fosmidomycin decreased picrosides content in a concentration and time dependent manner. At 24 h and 72 h of the treatment, respective decrease in the P-I was 85% and 91.7% at 50 μM; 90.2% and 92.7% at 100 μM; 92.0% and 96.9% at 200 μM of fosmidomycin (Fig. 4.2 a). Though, a concentration of 50 μM was effective in reducing the P-I, P-II required either higher concentration or longer time period (72 h) of lower concentration of fosmidomycin to be effective in inhibiting picrosides accumulation (Fig. 4.2 a).

![Graph showing the effect of fosmidomycin on picrosides content at 24 h and 72 h of treatment]

**Fig. 4.2 a: Effect of different concentrations of fosmidomycin on picrosides content at 24 h and 72 h of the treatments.** Percent change in picrosides (P-I and P-II) was calculated with respect to the respective control (did not receive any inhibitor) for each time point. Each value is mean of three separate biological replicate with standard deviation shown in the standard deviation bar at both the time intervals.
Chapter 4

Results

The effect of mevinolin was an interesting one. The lower concentration (50 μM) inhibited P-I and P-II by 20% and 53%, respectively at 24 h of treatment as compared to the respective control (Fig. 4.2 b). However, incubation of shoot at higher concentrations of mevinolin (200 μM) for longer period (72 h) enhanced P-I to 1000% (Fig. 4.2 b). This enhancement for P-II was 604.2%.

Fig. 4.2 b: Effect of different concentrations of mevinolin on picrosides content at 24 h and 72 h of the treatments. Percent change in picrosides (P-I and P-II) was calculated with respect to the respective control (did not receive any inhibitor) for each time point. Each value is mean of three separate biological replicate with standard deviation shown in the standard deviation bar at both the time intervals.
4.3 [1-$^{13}$C] Incorporation Experiment Suggests the Predominant Role of MEP Pathway in the Biosynthesis of Iridoid Moiety

Since P-1 is the predominant form of picrosides present in the leaf tissue, [1-$^{13}$C] enrichment experiment was performed using P-I.

![Diagram of carbon atoms enrichment](image)

**Fig. 4.3: Degree of enrichment of various carbon atoms of the P-I.** P-I was extracted from leaf tissue after being fed with either [1-$^{13}$C] glucose (S) or unlabelled glucose (C). The value indicated at a particular carbon is the S:C ratio calculated by taking the ratio of $^{13}$C- signal integral of $^{13}$C-labelled and unlabelled samples. For these calculations C-4 of P-I was set as reference and its integral was taken as unity.

$^{13}$C-NMR spectra revealed that out of 24 carbons, 8 were found to be enriched (Fig. 4.3 and Table 4.4). $^{13}$C enrichment was observed in: (a) $\beta$-carbon in cinnamoyl moiety, (b) C-3’’ and C-5’’ (at meta-positions) in cinnamoyl moiety, (c) C-6 and C-10 positions in iridoid moiety. In addition to this, enrichment was observed at positions C-1’, C-4’ and C-6’ in glucose moiety.
<table>
<thead>
<tr>
<th>Carbon Position</th>
<th>Chemical Shift $\delta_c$ (ppm)</th>
<th>$^{13}$C Signal Intensity</th>
<th>[1-$^{13}$C] glucose($I_c$)</th>
<th>Unlabeled glucose($I_c$)</th>
<th>$I_c / I_c$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Iridoid moiety</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>94.3</td>
<td>1.008</td>
<td>0.9575</td>
<td>1.05</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>140.7</td>
<td>1.019</td>
<td>1.017</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>103.1</td>
<td>1.000</td>
<td>1.000</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>38.0</td>
<td>0.9655</td>
<td>0.9165</td>
<td>1.05</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>78.6</td>
<td>0.9705</td>
<td>0.750</td>
<td><strong>1.29</strong></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>61.4</td>
<td>1.062</td>
<td>1.028</td>
<td>1.03</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>65.1</td>
<td>0.873</td>
<td>0.9700</td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>42.3</td>
<td>0.7985</td>
<td>0.898</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>60.8</td>
<td>1.029</td>
<td>0.7925</td>
<td><strong>1.30</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Glucose moiety</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1'</td>
<td>98.7</td>
<td>1.719</td>
<td>0.9845</td>
<td><strong>1.75</strong></td>
<td></td>
</tr>
<tr>
<td>2'</td>
<td>73.7</td>
<td>0.944</td>
<td>0.897</td>
<td>1.05</td>
<td></td>
</tr>
<tr>
<td>3'</td>
<td>76.4</td>
<td>1.002</td>
<td>0.9235</td>
<td>1.09</td>
<td></td>
</tr>
<tr>
<td>4'</td>
<td>70.5</td>
<td>1.021</td>
<td>0.810</td>
<td><strong>1.26</strong></td>
<td></td>
</tr>
<tr>
<td>5'</td>
<td>74.8</td>
<td>0.8395</td>
<td>0.8965</td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td>6'</td>
<td>63.1</td>
<td>1.024</td>
<td>0.8555</td>
<td>1.20</td>
<td></td>
</tr>
<tr>
<td><strong>Cinnamoyl moiety</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1&quot;</td>
<td>134.5</td>
<td>1.0315</td>
<td>1.014</td>
<td>1.02</td>
<td></td>
</tr>
<tr>
<td>2&quot;</td>
<td>129.1</td>
<td>2.256</td>
<td>2.3845</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>3&quot;</td>
<td>128.3</td>
<td>3.4225</td>
<td>2.1305</td>
<td><strong>1.61</strong></td>
<td></td>
</tr>
<tr>
<td>4&quot;</td>
<td>130.7</td>
<td>1.1415</td>
<td>0.9685</td>
<td>1.18</td>
<td></td>
</tr>
<tr>
<td>5&quot;</td>
<td>128.3</td>
<td>3.4225</td>
<td>2.1305</td>
<td><strong>1.61</strong></td>
<td></td>
</tr>
<tr>
<td>6&quot;</td>
<td>129.1</td>
<td>2.256</td>
<td>2.3845</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>CO</td>
<td>167.3</td>
<td>1.101</td>
<td>1.080</td>
<td>1.02</td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>117.5</td>
<td>0.9945</td>
<td>0.992</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>$\beta$</td>
<td>145.7</td>
<td>1.731</td>
<td>0.9215</td>
<td><strong>1.88</strong></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.4: $^{13}$C NMR analysis of P-I in P. kurrooa leaves after being fed with [1-$^{13}$C] glucose (S) or unlabelled glucose (C) under same conditions. Assigned carbon are shown in Fig. 4.3. The intensity ratio ($I_c / I_c$) greater than threshold value (1.20) are bold and underlined, and represents the significant enrichment of respective carbon position. C-4 of P-I was set as reference and its integral was taken as unity.
4.4 Effect of External Cues on Picrosides Accumulation

ABA treatment differentially modulated the levels of P-I and P-II. ABA enhanced P-I by 15% but decreased P-II by 50% as compared to the respective control at 24 h of the treatments. These values were 45% and 23% at 48 h of the treatment (Fig. 4.4).

MeJa did not alter the level of P-I, however, P-II decreased by 50% at 48 h of the treatment as compared to the respective control (Fig. 4.4).

![Graph](image)

**Fig. 4.4:** Effect of ABA, MeJA, PEG, H₂O₂, SA and GA₃ on picrosides accumulation: P-I and P-II was estimated at 24 h and 48 h of the treatment. Values for picrosides content are shown as percent over respective control (did not receive the treatment). Each value is mean of three separate biological replicate with standard deviation shown in the standard deviation bar at both the time intervals.

PEG induced the accumulation of P-I by 40% at both the time intervals of the treatment as compared to the respective control (Fig. 4.4). These values for P-II at 24 h and 48 h were 20% and 30%, respectively.
Chapter 4

Results

P-I increased by 50% and 15% at 24 h and 48 h of the H$_2$O$_2$ treatment, respectively as compared to the respective control (Fig. 4.4). P-II, however, showed initial decrease by 54% at 24 h and thereafter an increase (30%) was recorded at 48 h of treatment as compared to respective control.

SA treatment decreased P-I by 29.5% and 3.3% at 24 h and 48 h of the treatment, respectively (Fig. 4.4). The decrease in P-II was by 55% at above time intervals of treatment as compared to the respective control.

GA$_3$ treatment did not alter the picrosides accumulation to a appreciable level. The decrease and increase in P-I was by 10% and 5% at 24 h and 48 h of the treatment, respectively as compared to the respective control. P-II was decreased by 20% and 6% at 24 h and 48 h of the treatment, as compared to the respective control (Fig. 4.4).

4.5 Development of RNA Isolation System

Using the developed RNA isolation system, RNA was successfully extracted from leaves of *P. kurrooa* as explained in section 3.8. The yield from this system was 62.66±15.5 μg of RNA from 100 mg of leaf tissue with an A$_{260}$/A$_{280}$ ratio of 1.9 and A$_{260}$/A$_{230}$ ratio of 2.1, respectively. The developed system was found superior for isolation of RNA as compared with TRIzol® (Invitrogen, USA), RNeasy® (Qiagen, Germany) and guanidinium salt based method (Lal *et al.*, 2001). These methods either failed to isolate RNA or yielded extremely low quantity of RNA.

Integrity of isolated RNA was analyzed on a 1% formaldehyde-agarose denaturing gel. The RNA was intact as evidenced by the distinct 28S and 18S rRNA bands (Fig. 4.5). As expected, the intensity of the 28S rRNA band was higher as compared to that of the 18S rRNA band. Further, the 5S rRNA band was less intense than both the 28S and the 18S rRNA bands.
Fig. 4.5: Denaturing gel electrophoresis of RNA isolated from leaf tissues of *P. kurrooa* using different methods of RNA isolation: The legends on top indicate the method used I, present system; G, guanidine-HCl method; T, Trizol and R, RNeasy (Ghawana et al., 2007, 2011). RNA yield (μg/100 mg tissue) is mentioned below each panel, each value represents mean ± standard deviation (SD) of three biological samples. Equal volume of RNA (2 μl) was loaded. The starting material and the volume of the DEPC-treated ADW used to dissolve RNA were kept same in all the four procedures.

4.6 Cloning and Characterization of Genes Related to Picrosides Biosynthesis:

4.6.1 ESTs cloned by DD

DD with arbitrary primers H-AP-21 and H-AP-26 (Appendix III; Page XIV; Fig. 4.6 1-6) yielded a total of 70 differentially expressed cDNA fragments. After removal of poor quality and non-readable sequences, 43 sequences were obtained. Sixteen of the total 43 cDNA fragments obtained were either hypothetical proteins or did not show significant homology to any protein in the database (Table 4.4). Four cDNA fragments exhibited homology to *thiamine biosynthesis gene (Pkti)*, *3-hydroxy-3-methylglutaryl-CoA reductase (Pkhmgr)*, *geranyl diphosphate synthase (Pkgdps)* and *1-deoxy-D-xylulose-5-phosphate reductoisomerase (Pkdxr)*, respectively. Although, 3’ end is expected to generate. The cDNAs. However fragment obtained in our case was not from 3’ end. These were cloned to full-length cDNAs as described in section 4.6.3.
Fig. 4.6 (1-6): Autoradiogram of DD-gel depicting ESTs generated from control and PEG-treated leaves of *P. kurrooa*. Arrows indicate the cDNA fragments, which were excised from the gels and used in the present work. C, control; P, PEG treatment.
Table 4.4: Summary of ESTs, generated using DD. Shaded area represents the ESTs selected for the cloning of full-length cDNA using RACE.

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Clone ID</th>
<th>Homology in Database</th>
<th>cDNA Insert Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pk1.1.1</td>
<td>Proton-translocating NADH-quinone oxidoreductase</td>
<td>517 bp</td>
</tr>
<tr>
<td>2</td>
<td>Pk1.2.1</td>
<td>Cell wall-associated hydrolase</td>
<td>408 bp</td>
</tr>
<tr>
<td>3</td>
<td>Pk1.2.2</td>
<td>No significant similarity</td>
<td>230 bp</td>
</tr>
<tr>
<td>4</td>
<td>Pk1.2.3</td>
<td>No significant similarity</td>
<td>400 bp</td>
</tr>
<tr>
<td>5</td>
<td>Pk1.4.3</td>
<td>Pyruvate kinase</td>
<td>400 bp</td>
</tr>
<tr>
<td>6</td>
<td>Pk1.4.4</td>
<td>Putative senescence-associated protein</td>
<td>572 bp</td>
</tr>
<tr>
<td>7</td>
<td>Pk1.4.5</td>
<td>Un-characterized protein</td>
<td>578 bp</td>
</tr>
<tr>
<td>8</td>
<td>Pk1.4.6</td>
<td>Cytochrome P450</td>
<td>428 bp</td>
</tr>
<tr>
<td>9</td>
<td>Pk1.4.7</td>
<td>Methyl-coenzyme M reductase alpha subunit</td>
<td>572 bp</td>
</tr>
<tr>
<td>10</td>
<td>Pk1.4.8</td>
<td>UDP-glucuronate 4-epimerase 1-like</td>
<td>578 bp</td>
</tr>
<tr>
<td>11</td>
<td>Pk1.4.9</td>
<td>Aminoglycoside 3'-phosphotransferase</td>
<td>389 bp</td>
</tr>
<tr>
<td>12</td>
<td>Pk1.4.10</td>
<td>Plasma membrane H⁺-ATPase</td>
<td>560 bp</td>
</tr>
<tr>
<td>13</td>
<td>Pk1.4.11</td>
<td>ABC superfamily ATP binding cassette</td>
<td>208 bp</td>
</tr>
<tr>
<td>14</td>
<td>Pk2.4.12</td>
<td>SocE</td>
<td>578 bp</td>
</tr>
<tr>
<td>15</td>
<td>Pk2.4.13</td>
<td>Thiamine biosynthetic gene</td>
<td>400 bp</td>
</tr>
<tr>
<td>16</td>
<td>Pk3.5.11</td>
<td>Cytochrome P450 like TBE</td>
<td>580 bp</td>
</tr>
<tr>
<td>17</td>
<td>Pk3.5.12</td>
<td>Unknown protein product</td>
<td>580 bp</td>
</tr>
<tr>
<td>18</td>
<td>Pk3.5.13</td>
<td>hydroxymethylglutaryl-CoA reductase</td>
<td>470 bp</td>
</tr>
<tr>
<td>19</td>
<td>Pk4.5.7</td>
<td>rRNA intron-encoded homing endonuclease</td>
<td>580 bp</td>
</tr>
<tr>
<td>20</td>
<td>Pk4.6.2</td>
<td>GK21101 Drosophila willistoni</td>
<td>300 bp</td>
</tr>
<tr>
<td>21</td>
<td>Pk4.6.3</td>
<td>Putative photosystem II component</td>
<td>300 bp</td>
</tr>
<tr>
<td>22</td>
<td>Pk5.1.2</td>
<td>Hypothetical protein</td>
<td>320 bp</td>
</tr>
<tr>
<td>23</td>
<td>Pk5.2.5</td>
<td>No significant similarity</td>
<td>308 bp</td>
</tr>
<tr>
<td>24</td>
<td>Pk5.3.7</td>
<td>ATP synthase F0, A subunit</td>
<td>307 bp</td>
</tr>
<tr>
<td>25</td>
<td>Pk5.6.11</td>
<td>Unnamed protein product</td>
<td>311 bp</td>
</tr>
<tr>
<td>26</td>
<td>Pk6.4.4</td>
<td>Probable aquaporin PIP1-4-like</td>
<td>580 bp</td>
</tr>
<tr>
<td>27</td>
<td>Pk6.4.5</td>
<td>Geranyl diphosphate synthase</td>
<td>520 bp</td>
</tr>
<tr>
<td>28</td>
<td>Pk6.2.11</td>
<td>Hypothetical protein</td>
<td>296 bp</td>
</tr>
<tr>
<td>29</td>
<td>Pk6.8.8</td>
<td>Cell wall-associated hydrolase</td>
<td>560 bp</td>
</tr>
<tr>
<td>30</td>
<td>Pk6.9.4</td>
<td>Metallocarboxypeptidase inhibitor</td>
<td>580 bp</td>
</tr>
<tr>
<td>31</td>
<td>Pk6.9.4</td>
<td>Hypothetical protein</td>
<td>560 bp</td>
</tr>
<tr>
<td>32</td>
<td>Pk6.10.4</td>
<td>Hypothetical protein found</td>
<td>540 bp</td>
</tr>
<tr>
<td>33</td>
<td>Pk7.2.4</td>
<td>Putative senescence</td>
<td>480 bp</td>
</tr>
<tr>
<td>34</td>
<td>Pk7.3.2</td>
<td>Uncharacterized protein</td>
<td>458 bp</td>
</tr>
<tr>
<td>35</td>
<td>Pk7.10.8</td>
<td>Aquaporin</td>
<td>560 bp</td>
</tr>
<tr>
<td>36</td>
<td>Pk7.16.1</td>
<td>Calvin cycle protein CP12</td>
<td>580 bp</td>
</tr>
<tr>
<td>37</td>
<td>Pk7.16.2</td>
<td>Hypothetical protein</td>
<td>560 bp</td>
</tr>
<tr>
<td>38</td>
<td>Pk8.1.2</td>
<td>Putative purowidoline b protein</td>
<td>580 bp</td>
</tr>
<tr>
<td>39</td>
<td>Pk8.3.2</td>
<td>Uncharacterized protein</td>
<td>432 bp</td>
</tr>
<tr>
<td>40</td>
<td>Pk8.20.4</td>
<td>Epsin 1</td>
<td>363 bp</td>
</tr>
<tr>
<td>41</td>
<td>Pk8.20.6</td>
<td>1-deoxy-D-xylulose-5-phosphate reductoisomerase</td>
<td>580 bp</td>
</tr>
<tr>
<td>42</td>
<td>Pk8.25.8</td>
<td>Putative polyprotein</td>
<td>400 bp</td>
</tr>
<tr>
<td>43</td>
<td>Pk9.2.4</td>
<td>Hypothetical protein</td>
<td>400 bp</td>
</tr>
</tbody>
</table>
4.6.2 ESTs cloned by degenerate primers

DD involves the cloning of each band, which was expensive and time consuming. DD generated short-sized cDNA fragments that correspond to 3’ end of the gene. Generally, 3’ ends do not contain enough of coding region to provide significant information on gene homology and is apparently labeled as unknown or novel (Debouck, 1995; Lievens et al., 2001).

Hence degenerate primer based approach was followed to clone rest of the genes associated with MEP and MVA pathways, pathways central to GDP biosynthesis and hence picrosides biosynthesis. By utilizing degenerate primers, eight partial cDNAs were cloned as follow: 1-deoxy-D-xylulose-5-phosphate synthase (Pkdxs; 400 bp), 4-(cytidine-5'-diphospho)-2-C-methylerythritol kinase (Pkmck; 410 bp), 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate reductase (Pkhdr; 450 bp), acetoacetyl-CoA thiolase (Pktacth; 400 bp), 3-hydroxy-3-methylglutaryl-CoA synthase (Pkhmsgs; 400 bp), isopentenyl pyrophosphate isomerase (Pkpipp; 400 bp), phenylalanine ammonia lyase (Pkpalf; 200 bp) and caffeoyl-CoA O-methyltransferase (Pkcalf; 250 bp).

4.6.3 Full Length cloning of cDNAs

The partial sequences allowed designing of primers for RACE to clone full-length cDNAs. Nine full-length cDNAs were cloned as follow: Pkdxs (2317 bp), Pkdxs (1767 bp), Pkmck (1674 bp), Pkhdr (1701 bp), Pktacth (1545 bp), Pkhmsgr (2241 bp), Pkpipp (987 bp), Pkgdps (1434 bp) and Pkthi (1065 bp). The cloning and important characteristics of these genes are detailed in the following sections:
4.6.3.1 1-deoxy-D-xylulose-5-phosphate synthase (Pkdxs)

The PCR with degenerate primers yielded an amplicon of 400 bp (Fig. 4.7 A). The nucleotide sequence showed 74-78 % homology with dxs of Antirrhinum majus, Tagetes erecta and Nicotiana tabacum.

The partial sequence was used to design primers for 5' and 3' RACE to obtain full length cDNA. The nested 5' and 3' RACE yielded distinct bands of 1,000 bp and 1,500 bp, respectively (Fig. 4.7 B).

![Image](image_url)

Fig. 4.7: Agarose gel electrophoresis showing amplified products of Pkdxs: (A) partial cDNA using degenerate primer; (B) amplicons of nested RACE; lane 1: 5' RACE, lane 2: 3' RACE; unlabelled arrow on the right side of the gel indicates the desired amplicons, M: DNA size marker.

Full-length sequence of putative Pkdxs was obtained by aligning 5' and 3' RACE sequences. The full-length cDNA (GenBank accession number EU561005.1) was 2,317 bp, consisting of a coding region of 2,064 bp, 5' UTR of 50 bp, 3' UTR of 203 bp including a poly A+ tail of 23 bp (Fig. 4.8). The stop codon TAA and polyadenylation signal AAGAAA was present at positions 2,052 bp and 2,155 bp, respectively.
Fig. 4.8: Nucleotide and deduced amino acid sequence of *Pkdx* (GenBank accession number EU561005.1). The start codon is shown in bold; the termination codon (TAA) is shown in bold and asterisk (*), polyadenylation (AAGAAAA) signal is shown in bold and underlined.
Analysis of full length cDNA of Pkdxs

The 2064 bp of cDNA encoded a putative Pkdxs of 687 aa with predicted molecular mass of 73.62 kDa and pI of 8.60.

Deduced amino acid sequence exhibited 78% identity with dxs of Ricinus communis and Lycopersicon esculentum (Fig. 4.9). This value was 74% for Antirrhinum majus, Tagetes erecta, Nicotiana tabacum, and Stevia rebaudiana.

Conserved domain analysis at NCBI and SMART showed the presence of domains related to Pkdxs. The alignment of translated Pkdxs protein with other plant homologue revealed putative thiamine binding site, which is represented by motif GDGX(7-8)EX(3-4)AX(11-13)N and was present at position 185-216 aa residues. Pkdxs possessed signature motif VGAL (at 234-238 a.a. residue), which is specific to plant dxs (Fig. 4.9). The conserved domain characteristics to transketolase superfamily, TPP binding sites were also present in the cloned gene (Fig. 4.10 E).
Fig. 4.9: Alignment of deduced amino acid sequence of \textit{Pdcs} with \textit{dcs} homologs of other plant species. Hyphen (-) indicates gap introduced to optimize alignment. Identical amino acids are shown by star (*) whereas colons (:) indicates the consensus sequences. The GenBank accession numbers of different \textit{dcs} used for alignment were: \textit{AY 770407} (Antirrhinum; \textit{Antirrhinum majus}), \textit{AF 251026} (Tagetes; \textit{Tagetes errecta}), \textit{EQ 974568} (Ricinus; \textit{Ricinus communis}), \textit{FJ 214107} (Stevia; \textit{Stevia reudiana}) and \textit{AY 687353} (Lycopersicon; \textit{Lycopersicon esculentum}). The consensus sequence of conserved domains are underlined and mentioned above the sequence.
Picrorhiza  
Antirrhinum  
Ricinus  
Stevia  
Lycopersicon  

Picrorhiza  
Antirrhinum  
Ricinus  
Stevia  
Lycopersicon  

Picrorhiza  
Antirrhinum  
Ricinus  
Stevia  
Lycopersicon  

Fig. 4.9: Continued

SOPMA analysis (Geourjon, C. & Deléage, 1995) showed that Pkdxs consisted of 36.10 % α-helix, 40.32 % random coils, joined by 16.74 % extended sheet and 6.84 % β-turns (Fig. 4.9 D). Hydropathy plots (Kyte and Doolite, 1982) of deduced amino acid sequence of Pkdxs suggested it to be hydrophilic protein (Fig. 4.9 B). This prediction was further supported by the DAS-TM segment prediction (Fig. 4.9 A) and PSORT (Fig. 4.9 C) analysis. These predictions mapped the location of this protein at chloroplast stroma and inner membrane chloroplast with certainty of 0.95 and 0.82, respectively.

97
Fig. 4.10: Structural analysis of *Pdxs*; (A) transmembrane domain, (B) hydropathy plot, (C) cellular locations, (D) predicted secondary structures (where alpha helices, extended sheets, beta turns and random coils are indicated by the longest, the second longest, the second shortest and the shortest vertical lines, respectively), (E) Predicted conserved domain.
4.6.3.2 1-deoxy-D-xylulose-5-phosphate reductoisomerase (Pkdxr)

An amplicon of 400 bp was amplified from DD-gel using respective anchored and arbitrary primers (Fig. 4.11 A). The nucleotide sequence obtained from this amplicon showed 82% to 80% homology (E-value= 2e-80) with dxr of Nicotiana tabacum, Salvia miliorrhiza and cathranthus roseus.

![Image of agarose gel electrophoresis]

Fig. 4.11: Agarose gel electrophoresis showing amplified products of Pkdxr: (A) partial cDNA using DD primers; (B) amplicons from RACE, lane 1: 5' RACE, lane 2: 3' RACE; (C) full length ORF; M, DNA size marker (bp) loaded on the gel; unlabelled arrow on the opposite side of marker indicates the desired amplicon.

The partial sequence was used to design primers for 5' and 3' RACE to obtain full length cDNA. Primary 5' and 3' RACE yielded multiple bands. In second round, nested PCR generated distinct bands of 1,500 bp and 800 in 5' RACE. (Fig. 4.11 B). The full length ORF was also cloned (Fig. 4.11 C).

Full length sequences of putative Pkdxr was obtained by aligning 5' RACE sequences and DD sequence. The full-length cDNA (GenBank accession number DQ347963) was 1,767 bp consisting of a coding region of 1,458 bp, 5' UTR of 115 bp, 3' UTR of 194 bp including a poly A⁺ tail of 28 bp (Fig. 4.12). The stop codon TAA and polyadenylation signal ATTAAC was present at position 1570 bp and 1706 bp, respectively.
Chapter 4

Fig. 4.12: Nucleotide and deduced amino acid sequence of *Pkdxr* (GenBank accession number DQ347963). The start codon (ATG) is shown in bold; the termination codon (TAA) is shown in bold and underlined.
Analysis of full length cDNA of \( Pkd\alpha \)

The 1458 bp cDNA encoded a putative \( Pkd\alpha \) of 486 aa with predicted molecular mass of 51.42 kDa and pI of 5.89.

Deduced amino acid sequence exhibited 89% similarity with \( dxr \) of \textit{Antirrhinum majus}, \textit{Nicotiana tabacum}, and \textit{Hevea brasiliensis}. This value was 85% for \textit{Stevia rebaudiana}, \textit{Artemisia annua} and \textit{Horeudum vulgare} (Fig. 4.13).

Fig. 4.13: Alignment of deduced amino acid sequence of \( Pkd\alpha \) with \( dxr \) homologs of other plant species. Hyphen (-) indicates gap introduced to optimize alignment. Identical amino acids are shown by star (*) whereas colons (:) indicates the consensus sequences. The GenBank accession numbers of different \( dxr \) used for alignment were: FJ 214108 (Stevia; \textit{Stevia rebaudiana}), AF 182287 (Artemisia; \textit{Artemisia annua}), AY 770406 (Antirrhinum; \textit{Antirrhinum majus}), AF 331705 (Lycopersicon; \textit{Lycopersicon esculentum}), DQ 839130 (Nicotiana; \textit{Nicotiana tabacum}), AJ 583446 (Horeudum; \textit{Horeudum vulgare}) and AY 494186 (Ginkgo; \textit{Ginkgo biloba}). The consensus sequence of conserved domains are underlined and mentioned above the sequences.
At the N-terminal side, the sequence is poorly conserved but enriched in Serine residues (Fig. 4.13). The Pkdxr possessed processing site near N terminus which was represented by conserved Cys-Ser-X motif (at position 49-51 aa), where X means any of the hydrophobic residues Ala, Val, or Met. Following this, highly conserved consensus sequence rich in Proline (Pro) residues was detected, it was represented by consensus motif P(P/Q) PAWPG(R/T)A (at positions 58-66 aa of the picrorhiza sequence). This is characteristics signature motif for plant dxr. Conserved domain analysis at NCBI and SMART revealed the presence of important domain related to Pkdxr such as to the catalytic site, substrate binding site (Fig. 4.14 E). SOPMA analysis showed that Pkdxr consisted of 40.42% α-helix, 34.32% random coils, joined by 17.47% extended sheet and 7.79% β-turns (Fig. 4.14 D).

Hydropathy plot of deduced amino acid sequence of Pkdxr suggested it to be hydrophobic protein (Fig. 4.14 B). This prediction further supported by the transmembrane (Fig. 4.14 A) and PSORT (Fig. 4.14 C) analysis of deduced amino acid sequences. These predictions map the location of this protein to the chloroplast stroma with 0.89 certainty and to mitochondrial matrix with 0.62 certainty.
Fig. 4.14: Structural analysis of *Pkdxr*; (A) transmembrane domain, (B) hydropathy plot, (C) cellular location, (D) predicted secondary structures (where alpha helices, extended sheets, beta turns and random coils are indicated by the longest, the second longest, the second shortest and the shortest vertical lines, respectively and (E) predicted conserved domains.
4.6.3.3 4-diphosphocytidyl-2C-methyl-D-erythritol kinase (Pkcmk)

The PCR with degenerate primers yielded an amplicon of 410 bp (Fig. 4.15 A). The nucleotide sequence of this amplicon exhibited about 70% homology to cmk of Catharanthus roseus with E-value=2e-173. The partial sequence was used to design primers for 5' and 3' RACE to obtain full length cDNA. The nested PCR generated bands of 800 bp and 900 bp in 5' and 3' RACE, respectively (Fig. 4.15 B and C).

![Figure 4.15](image)

Fig. 4.15: Agarose gel electrophoresis showing amplified products of Pkcmk: (A) partial cDNA using degenerate primers; (B) 5' RACE amplicon; (C) 3' RACE amplicon; unlabelled arrow on the right side of the gel indicates the desired band. M: DNA size marker (bp) loaded on the gel.

Full length sequence of putative Pkcmk was obtained by aligning 5' and 3' RACE sequence. The full-length cDNA (including 5' and 3'-untranslated region) of Pkcmk (GenBank accession number EF 199769) was 1674 bp, consisting of a coding region of 1395 bp, 5' UTR of 174 bp, and 3' UTR of 279 bp including a poly A⁺ tail of 61 bp (Fig. 4.16). The stop codon TAG and polyadenylation signal AATAAA was present at position 1,393 bp and 1,436 bp, respectively. The isolated gene from this study showed 70% homology to Salvia miltiorrhiza, Lycopersicon esculentum and Nicotiana tabacum, respectively.
Figure 4.16: Nucleotide and deduced amino acid sequence of *Pkcmk* (GenBank accession number EF199769). The start codon (ATG) is shown in bold; the termination codon (TAG) is shown in bold and underlined. (*)
**Fig. 4.17:** Alignment of deduced amino acid sequence of *Pckm* with *cmk* homologs of other plant species. Hyphen (-) indicate gap introduced to optimize alignment. Identical amino acids are shown by star (*) whereas colons (:) indicate the consensus sequences. The GenBank accession numbers of different *cmk* used for alignment are: EF 534309 (Salvia; *Salvia miltiorrhiza*), EF 474476 (Nicotiana; *Nicotiana tabacum*), DQ 269453 (Stevia; *Stevia rebaudiana*), FJ 217703 (Hevea; *Hevea brasiliensis*), NM 001155666 (Zea; *Zea mays*), AF 288615 (*Arabidopsis; Arabidopsis thaliana*). The consensus sequences of the conserved domains are underlined.

107


Analysis of full length cDNA of Pkcmk

The 1395 bp cDNA encoded a putative Pkcmk of 465 aa with predicted molecular mass of 44.46 kDa and pI of 5.87.

Deduced amino acid sequence exhibited 88% similarity with cmk of Nicotiana tabacum, 87% with Hevea brasiliensis. The value was 84% to 80% for Oryza stavia, Arabidopsis thaliana and Raphanus sativus (Fig. 4.17).

Conserved domain analysis at NCBI and SMART showed the presence of domain related to Pkcmk (Fig. 4.18 E). The three characteristics conserved motifs were found. The first region (K-L-N-L-F-L) reported to be involved in CDP-ME binding (Miallau et al., 2003; Sgraja et al., 2008) was present at position 164-168 aa. The second region G-L-G-G-G-S and a variant of the third region (L-T-GT-G,) which was represented by C-T-GR-G, was present at position 191-198 aa and at 242-247 aa (Fig. 4.17), respectively. These motifs represent ATP binding site (Sgraja et al., 2008; Miallau et al., 2003) and are characteristics of functional cmk.

SOPMA analysis showed that Pkcmk consisted of 32.27% α-helix, 44.09% random coils, joined by 17.33% extended sheet and 5.91% β-turns (Fig. 4.18 D). Hydropathy plots of deduced amino acid sequence of Pkcmk suggested it to be hydrophobic protein (Fig. 4.18 B). This prediction was further supported by the DAS-TM (Fig. 4.18 A) and PSORT (Fig. 4.18 C) analysis of deduced amino acid sequences. These predictions map the location of this protein to the chloroplast stroma with 0.89% certainty and to thylakoid with 0.56% certainty.
**Fig. 4.18: Structural analysis of Pkcmk:** (A) transmembrane domain, (B) hydropathy plot, (C) cellular locations, (D) predicted secondary structures (where alpha helices, extended sheets, beta turns and random coils are indicated by the longest, the second longest, the second shortest and the shortest vertical lines, respectively) and (E) predicted conserved domain.
4.6.3.4 1-hydroxy-2-methyl-2-(E)-butenyl-4-PP reductase (Pkhdr)

The PCR with degenerate primer yielded an amplicon of 450 bp (Fig. 4.19 A). The nucleotide sequence obtained from this amplicon showed 70% homology with hdr of Adonis palaestinas. From this nucleotide sequence information gene specific primer were designed and utilized in RACE-PCR to obtain the full length cDNA.

The first round of primary PCR was performed using 5' RACE ready cDNA population, to clone 5' end of hdr. Similarly, primary PCR using 3' RACE ready cDNA population, to clone 3' end. Both of these PCR did not give any amplification. In second round, nested PCR yielded bands of 980 bp and 800 bp in 5' and 3' RACE, respectively (Fig. 4.19 B).

![Image](image_url)

**Fig. 4.19:** Agarose gel electrophoresis showing amplified products of Pkhdr: (A) partial cDNA using degenerate primer, (B) amplicons of 5' and 3' RACE; lane 1: 5' RACE, lane 2: Nested 5' RACE lane 3: 3' RACE; unlabelled arrow on the right side of the gel indicates the desired band, M: DNA size marker.

Full length sequence of putative Pkhdr was obtained by aligning 5' and 3' RACE sequences. The full-length cDNA (GenBank accession number EF199770.1) was 1701 bp, consisting of a coding region of 1,266 bp, 5' UTR of 47 bp, 3' UTR of 277 bp including a poly A+ tail of 27 bp (Fig. 4.20). The stop codon TAG and polyadenylation signal CATAAA was present at position 1,431 bp and 1,466 bp, respectively.
Fig. 4.20: Nucleotide and deduced amino acid sequence of Phkhrd (GenBank accession number EF199770). The start codon (ATG) is shown in bold; the termination codon (TAG) is shown in bold and asterisk (*), polyadenylation signal (CATAAA) is shown in bold and underlined.

111
Fig. 4.21: Sequence alignment of the deduced \textit{Pkdhr} and other \textit{hdr} homologs from other plants.

The GenBank accession numbers of different \textit{hdr} used for alignment are: ABB55395 (StIspH; \textit{Solanum tuberosum}), AF 270978 (ApIspH; \textit{Adonis palestinae}), ABB 88836 (SrIspH; \textit{Stevia rebaudiana}), NP 0015467 (OsHDR; \textit{Oryza sativa}), AAW 82381 (AtIspH; \textit{Arabidopsis thaliana}).

The completely identical amino acids are indicated with (*) Conserved amino acids are indicated with colons (:). Four conserved cysteine residues are shown by arrows.
Analysis of full length cDNA of Pkhdr

The 1266 bp cDNA encoded a putative Pkhdr of 422 aa with predicted molecular mass of 51.90 kDa and isoelectric point (pI) of 5.96.

Deduced amino acid sequence shared 80% similarity with hdr of *Oryza stavia*, *Arabidopsis thaliana* and *Raphanus sativus*. The value was 78% for *Adonis palaestina* and *Stevia rebaudiana* (Fig. 4.21).

Conserved domain analysis at NCBI and SMART showed the presence of domain related to Pkhdr (Fig. 4.22 E). The four conserved cysteine residues which are present in all the Pkhr, were found in predicted amino acid sequences. These had specific function and play vital role in catalytic activity. SOPMA analysis showed that Pkhdr consisted of 45.02 % α-helix, 32.03% random coils, joined by 16.02% extended sheet and 6.93% β-turns (Fig. 4.22 D).

Hydropathy plots of deduced amino acid sequence of PkDXR suggested it to be hydrophobic protein (Fig. 4.22 B). This prediction was further supported by the DAS-TM (Fig. 4.22 A) and PSORT (Fig. 4.22 C) analysis of deduced amino acid sequences. These predictions map the location of this protein to the mitochondrial matrix with 0.59 certainty and to microbody with 0.39 certainty.
Fig. 4.22: Structural analysis of \textit{Pkhdr}; (A) transmembrane domain, (B) hydropathy plot, (C) cellular locations, (D) predicted secondary structures (where alpha helices, extended sheets, beta turns and random coils are indicated by the longest, the second longest, the second shortest and the shortest vertical lines, respectively) and (E) predicted conserved domains.
4.6.3.5 Acetoacetyl-CoA thiolase (Pkaclh)

The PCR with degenerate primers yielded an amplicon of 400 bp (Fig. 4.23 A). The nucleotide sequence of this amplicon showed 80% homology with acth of Nicotiana tabacum. The partial sequence was used to design primers for 5' RACE and 3' RACE to obtain full length cDNA.

The first round of primary PCR was performed using 5' RACE ready cDNA population, to clone 5' end of Pkaclh. Similarly, primary RACE using 3' RACE ready cDNA population, to clone 3' end. Both of these PCR yielded multiple bands. In second round, nested PCR yielded bands of 1,500 bp and 900 bp in 5' and 3' RACE, respectively (Fig.4.23 B).

Fig. 4.23: Agarose gel electrophoresis showing amplified products of Pkaclh: (A) partial cDNA using degenerate primers, (B) amplicons of 5' and 3' RACE of Pkaclh; Lane 1: 5' RACE, Lane 2: 3' RACE; M:DNA size marker (bp) loaded on the gel, unlabelled arrow on opposite side of the marker indicates the desired band.

Full length sequence of putative Pkaclh was obtained by aligning 5' and 3' RACE sequences. The full-length cDNA of Pkaclh (GenBank accession number DQ347964) was 1545 bp consisting of a coding region of 1245 bp, 5' UTR of 84 bp, 3' UTR of 246 bp including a poly A+ tail of 24 bp (Fig. 4.24). The stop codon TAA and polyadenylation signal AATAAAA was present at position 1297 bp and 1436 bp, respectively. The isolated gene from this study exhibited 82% homology with Nicotiana tabacum, 78% to Hevea brasiliensis and Arabidopsis thaliana, respectively.
Fig. 4.24: Nucleotide and deduced amino acid sequence of Phacth (GenBank accession number DQ347964). The start codon (ATG) is shown in bold; termination codon (TAA) is shown in bold and asterisk (*), and polyadenylation signal (AATAAA) is shown in bold and underlined.
**Chapter 4**

**Results**

| ARABIDOPSIS | MAPPVSDSLQPRDVCVVGARTP1GDGLGSLTATRGLGIAQAALKRAHVDALVE | 60 |
| HEVEA | MFP-SSDSINPRDVICVVGARTP1GDGLGSLTATRGLGIAQAALKRAHVDALVE | 58 |
| FICORRHIZA | MAP-SSDSINPRDVICVVGARTP1GDGLGSLTATRGLGIAQAALKRAHVDALVE | 59 |
| NICOTIANA | MAP-SSDSINPRDVICVVGARTP1GDGLGSLTATRGLGIAQAALKRAHVDALVE | 59 |
| RAPHANUS | MAU-SSDSINPRDVICVVGARTP1GDGLGSLTATRGLGIAQAALKRAHVDALVE | 59 |
| ORYZA | --- | 59 |
| ARNEBIA | --- | 56 |

---

**Fig. 4.25:** Alignment of deduced amino acid sequence of *P. acth* with *acth* homologs of other plant species. Hyphen (-) indicate gap introduced to optimize alignment. Identical amino acids are shown by star (*) whereas colons (:) indicate consensus sequences. The GenBank accession numbers of different *acth* used for alignment are: AF 364059 (Arabidopsis; *Arabidopsis thaliana*), AB 294687 (Hevea; *Hevea brasiliensis*), AY 748245 (Nicotiana; *Nicotiana tabacum*), X 78116 (Raphanus; *Raphanus sativus*), BAB 39872 (Oryza; *Oryza sativa*), DQ 395086 (Arnebia; *Arnebia euchroma*). The thioase signature motif and active site are underlined.
Chapter 4

Results

Analysis of full length cDNA of Pkacth

The 1245 bp cDNA encoded a putative Pkacth of 415 aa with predicted molecular mass of 41.41 kDa and pI of 6.33.

Deduced amino acid sequence shared 88% similarity with acth of Nicotiana tabacum, 87% with Hevea brasiliensis. This value was 84% to 80% for Oryza stavia, Arabidopsis thaliana and Raphanus sativus (Fig. 4.25).

Conserved domain analysis at NCBI and SMART showed the presence of important domain related to Pkacth. This enzyme had multiple domains, which consist of domain for thiolase activity, condensation activity, decarboxylation etc. (Fig. 4.26 E). The thiolase signature with consensus N-x(2)-G(2)-x[LIVM]-SA-x-G-H-P-x-[GAS]-x-[ST]-G was found near the C-terminus at position 350-366 aa. The thiolase active site represented by consensus sequence GVGGVCNGgGgAsA was present at position 385-398 aa (Fig. 4.25).

SOPMA analysis showed that Pkacth consisted of 40.48% α-helix, 32.43% random coils, joined by 15.35% extended sheet and 7.43% β-turns (Fig. 4.26 D). Hydropathy plots of deduced amino acid sequence of Pkacth suggested it to be hydrophilic protein (Fig. 4.26 B). This prediction was further supported by the transmembrane (Fig. 4.26 A) and PSORT (Fig. 4.26 C) analysis of deduced amino acid sequences. These predictions map the location of this protein to the cytoplasm with 0.45 certainty and to microbody matrix with 0.30 certainty.
Fig. 4.26: Structural analysis of PkACTH: (A) transmembrane domain, (B) hydropathy plot, (C) cellular locations, (D) predicted secondary structures (where alpha helices, extended sheets, beta turns and random coils are indicated by the longest, the second longest, the second shortest and the shortest vertical lines, respectively and (E) predicted conserved domain.
4.6.3.6 3-hydroxy-3-methylglutaryl-Coenzyme A reductase (Pkhmgr)

An amplicon of 500 bp was amplified from the DD gel (Fig. 4.27 A). The sequence of this differentially expressed amplicon exhibited 70% homology with E-value = 0.0 to *hmgr* of *Nicotiana tabacum*. This partial sequence information was utilized to design primers for 5' RACE and 3' RACE.

Primary RACE was performed with 5' RACE ready cDNA population to clone the 5' ends of the gene, did not give any visible amplification. The nested PCR yielded an amplicon of 800 bp and 1,300 bp in the 5' RACE and 3' RACE, respectively (Fig. 4.27 B).

![Fig. 4.27: Agarose gel electrophoresis showing amplified products of Pkhmgr](image)

Full length sequence of putative *Pkhmgr* was obtained by aligning the sequences from 5' RACE and DD amplicon. The full-length cDNA (GenBank accession number DQ 347962) was 2,241 bp consisting of a coding region of 1,683 bp, 5' UTR of 103 bp, 3' UTR of 452 bp including a poly A⁺ tail of 27 bp (Fig. 4.28). The stop codon TAA and polyadenylation signal AGGAAA was present at position 1788 and 2198 bp, respectively.

```
1  ACGCGGGGATTCTTCAGTCTTTCCACTATAACGAAACCACTCCCGTAGGCATCGCTCTTC
1  M D F R R
61  CAGGCCACCTAAAATTACCCCATTTCTCTCCACACCACCGAAGATATGAGCTTGCCGG
6  R P P R P S Q I K H Q L Q S P A S G A S
121  AGGCCCTCTAGACCCTTACAGATCAAACGATCAATCTACATCTCCAGCCCTCCGGCGCTCC
26  A D H H N Q S S S P K A S D A L P L P L
181  GCCGATCACCACACCGAGTCGCTCCTCCCAAGGCTCTGACTGCGGCTTACCTTTAA
```
Chapter 4

Fig. 4.28: Nucleotide and deduced amino acid sequence of *Pkhmgr* (GenBank accession number DQ 347962). The start codon (ATG) is shown in bold; termination codon (TAA) is shown in bold and marked by asterisk (*), polyadenylation signal (AGGAAA) is shown in bold and underlined.
**Motif A**

**Motif B**

**Fig. 4.29:** Sequence alignment of the deduced Pkhmgr with hmgr homologs of other plants. Hyphen (-) indicate gap introduced to optimize alignment. Identical amino acids are shown by star (*) whereas colons (:) indicate the consensus sequences. The GenBank accession numbers of different hmgr used for alignment are AAB 87727 (Nicotiana; Nicotiana tabacum), AAB 62581 (Lycopersicon; Lycopersicon esculentum), AY 254389.2 (Andrographis; Andrographis paniculata), EF 0625691 (Eucommia; Eucommia perkinis), AAR 83122 (Arabidopsis; Arabidopsis italiana), AA48610 (Raphanus; Raphanus sativus), BAA 36291 (Cucumis; Cucumis melo), CAO 47551 (Vitis; Vitis vinifera). Conserved motifs were boxed and marked as A, B, C, D, E, F and G, respectively.
Fig. 4.29 continued
Analysis of full length cDNA of Pkhmgr

The 1683 bp cDNA encoded a putative Pkhmgr of 561 aa with predicted molecular mass of 60.17 kDa and pI of 6.82.

Deduced amino acid sequence shared 84% similarity with hmgr of *Nicotiana tabacum* and *Lycopersicon esculentum*. This value was 78% for *Arabidopsis thaliana* and 74% to 70% for *Andrographic paniculata*, *Eucommia perkinsis* and *Cucumis melo*, respectively (Fig. 4.29).

Conserved domain analysis at NCBI and SMART showed the presence of characteristics domains related to HMG-CoA reductase activity. The HMG-CoA binding motif (E83 loop), represented by consensus DAMGXN was present at position 374-380 aa and NAD(P) binding motif, represented by GX2G2XT was present at position 526-534 aa, in Pkhmgr. A EX3GX4P motif (motif b), which had reported to be essential for the HMG-CoA binding site, was present at position 251-261 aa (Fig. 4.29). The domain related to the catalytic site, substrate binding site and NADP binding site has been found (Fig. 4.30 E). SOPMA analysis showed that Pkhmgr consisted of 45.63% α-helix, 34.32% random coils, joined by 13.90% extended sheet and 4.63% β-turns (Fig. 4.30 D).

Hydropathy plots of deduced amino acid sequence of Pkhmgr suggested it to be transmembrane protein (Fig. 4.30 B). This prediction further supported by the DAS-TM (Fig. 4.30 A) and PSORT (Fig. 4.30 C) analysis of deduced Pkhmgr. These predictions map location of this protein to the plasma membrane with 0.60 certainty, to golgi body with 0.40 certainty and to chloroplast thylakoid membrane with 0.38 certainty.
Fig. 4.30: Structural analysis of Pkhmgr; (A) transmembrane domain, (B) hydropathy plot, (C) cellular locations, (D) predicted secondary structures (where alpha helices, extended sheets, beta turns and random coils are indicated by the longest, the second longest, the second shortest and the shortest vertical lines, respectively and (E) predicted conserved domain.
4.6.3.7 Geranyl diphosphate synthase (Pkgdps)

An amplicon of 260 bp was amplified from DD-gel (Fig. 4.31). The nucleotide sequence of this amplicon shared 68% homology with E-value = 6e-148 to gdps of Tagetes erecta.

![Fig. 4.31: Agarose gel electrophoresis showing partial cDNA of Pkgdps amplified from DD-gel.](image)

The DNA size markers (M) are indicated along with their size. Arrow on the right side indicates the desired band.

The nested RACE PCR generated distinct bands of approximately 1 kb and 1.3 kb in 5' and 3' RACE respectively (Fig. 4.32 A).

![Fig. 4.32: Agarose gel electrophoresis showing amplicons of Pkgdps; (A) 5' RACE; (B) 3' RACE; (C) full length ORF of Pkgdps. M: DNA size marker (bp); unlabelled arrow on the opposite side of marker indicates the desired band.](image)

Full length sequence of putative Pkgdps was obtained by aligning 5' and 3' RACE sequences. The gene was also cloned to full length (Fig. 4.32 B). The full-length cDNA (GenBank accession number AY 866498) was 1434 bp consisting of a coding region of
Chapter 4

Results

1128bp, 5' UTR of 113 bp, 3' UTR of 205 bp including a poly A+ tail of 33 bp (Fig. 4.33). The stop codon TGA and polyadenylation signal TATAAA was present at position 1225 and 1267 bp, respectively.

```
1  GCGGGAAGGCAATTCTATACAAGATTCACATAATTCTCTTCTCTGCACCTGAATCAGCAA
  1  M S L
61  TCACACTACTATCTGATCTTTAACTCCACACTTGAATTTTCTTCCACCACAAATGAGCCTT
  61  V L T W S T S S I L Q S N I S
121  GAAATCTATTTACATGGTCACAAACAAGCTCCATACTCAACATTCAAAGCAACATATCA
  121  ATG AGCCTT
  4    V N S I T W S Q T S S I L N I Q S N I S
181  GTAAATTCTATTACATGGTCACAAACAAGCTCCATACTCAACATTCAAAGCAACATATCA
  181  ATG AGCCTT
  4    V N S I T W S Q T S S I L N I Q S N I S
241  GTAAATTCTATTACATGGTCACAAACAAGCTCCATACTCAACATTCAAAGCAACATATCA
  241  ATG AGCCTT
  4    V N S I T W S Q T S S I L N I Q S N I S
301  GTAAATTCTATTACATGGTCACAAACAAGCTCCATACTCAACATTCAAAGCAACATATCA
  301  ATG AGCCTT
  4    V N S I T W S Q T S S I L N I Q S N I S
361  GTAAATTCTATTACATGGTCACAAACAAGCTCCATACTCAACATTCAAAGCAACATATCA
  361  ATG AGCCTT
  4    V N S I T W S Q T S S I L N I Q S N I S
421  GTAAATTCTATTACATGGTCACAAACAAGCTCCATACTCAACATTCAAAGCAACATATCA
  421  ATG AGCCTT
  4    V N S I T W S Q T S S I L N I Q S N I S
481  GTAAATTCTATTACATGGTCACAAACAAGCTCCATACTCAACATTCAAAGCAACATATCA
  481  ATG AGCCTT
  4    V N S I T W S Q T S S I L N I Q S N I S
541  GTAAATTCTATTACATGGTCACAAACAAGCTCCATACTCAACATTCAAAGCAACATATCA
  541  ATG AGCCTT
  4    V N S I T W S Q T S S I L N I Q S N I S
601  GTAAATTCTATTACATGGTCACAAACAAGCTCCATACTCAACATTCAAAGCAACATATCA
  601  ATG AGCCTT
  4    V N S I T W S Q T S S I L N I Q S N I S
661  GTAAATTCTATTACATGGTCACAAACAAGCTCCATACTCAACATTCAAAGCAACATATCA
  661  ATG AGCCTT
  4    V N S I T W S Q T S S I L N I Q S N I S
721  GTAAATTCTATTACATGGTCACAAACAAGCTCCATACTCAACATTCAAAGCAACATATCA
  721  ATG AGCCTT
  4    V N S I T W S Q T S S I L N I Q S N I S
781  GTAAATTCTATTACATGGTCACAAACAAGCTCCATACTCAACATTCAAAGCAACATATCA
  781  ATG AGCCTT
  4    V N S I T W S Q T S S I L N I Q S N I S
841  GTAAATTCTATTACATGGTCACAAACAAGCTCCATACTCAACATTCAAAGCAACATATCA
  841  ATG AGCCTT
  4    V N S I T W S Q T S S I L N I Q S N I S
901  GTAAATTCTATTACATGGTCACAAACAAGCTCCATACTCAACATTCAAAGCAACATATCA
  901  ATG AGCCTT
  4    V N S I T W S Q T S S I L N I Q S N I S
961  GTAAATTCTATTACATGGTCACAAACAAGCTCCATACTCAACATTCAAAGCAACATATCA
  961  ATG AGCCTT
  4    V N S I T W S Q T S S I L N I Q S N I S
1021  GTAAATTCTATTACATGGTCACAAACAAGCTCCATACTCAACATTCAAAGCAACATATCA
  1021  ATG AGCCTT
  4    V N S I T W S Q T S S I L N I Q S N I S
1081  GTAAATTCTATTACATGGTCACAAACAAGCTCCATACTCAACATTCAAAGCAACATATCA
  1081  ATG AGCCTT
  4    V N S I T W S Q T S S I L N I Q S N I S
1141  GTAAATTCTATTACATGGTCACAAACAAGCTCCATACTCAACATTCAAAGCAACATATCA
  1141  ATG AGCCTT
  4    V N S I T W S Q T S S I L N I Q S N I S
1201  GTAAATTCTATTACATGGTCACAAACAAGCTCCATACTCAACATTCAAAGCAACATATCA
  1201  ATG AGCCTT
  4    V N S I T W S Q T S S I L N I Q S N I S
```

Fig. 4.33: Nucleotide and deduced amino acid sequence of *Pkgdps* (GenBank accession number AY866498). The start codon (ATG) is shown in bold, termination codon (TGA) is shown in bold and asterisk (*), and polyadenylation signal (TATAAA) is shown in bold and underlined.
Fig. 4.34: Alignment of deduced amino acid sequence of *Pkgdps* with *gdp* homologs of other plant species. Hyphen (-) indicate gap introduced to optimize alignment. Identical amino acids are shown by star (*) whereas colons (:) indicate the consensus sequences. The conserved domains are underlined and mentioned. The GenBank accession numbers of different *gdp* used for alignment are: AF 251012 (Tagetes; *Tagetes erecta*), DQ 432013 (Stevia; *Stevia reediana*), EF 382626 (Nicotiana; *Nicotiana tabacum*), AB 055496 (Hevea; *Hevea brasiliensis*) and AB027705 (Daucus; *Dacus carota*).
Chapter 4

Results

Analysis of full length cDNA of Pkgdps

The 1128 bp of Pkgdps encoded a putative Pkgdps of 375 aa with predicted molecular mass of 40.39 kDa and pI of 5.38.

Deduced amino acid sequence exhibited 73% similarity with gdps of Tagetes erecta, Stevia reebudiana and Dacus carota. The value was 70% for Nicotiana tabacum, and Hevea brasiliensis (Fig. 4.34).

The alignment of Pkgdps with its homologue from other plants revealed the presence of conserved aspartate rich motifs like DD(X)2-4,D present at 150-157 aa, the other two such motifs like DDXXD were present at 287-291 aa and DXXXD 306-310 aa, respectively (Fig. 4.34). The first 61 deduced amino acid residues showed the required characteristics of an N-terminal plastid-targeting sequence i.e. the sequence is rich in serine residues and small hydrophobic amino acids and is low in acidic residues. Conserved domain analysis at NCBI and SMART showed the presence important domain related to Pkgdps. The domain related to the catalytic site, substrate binding site and substrate-Mg\(^{2+}\) binding site were found (Fig. 4.35 E). The presence of various essential domains like substrate binding domain, active site lid, chain length determination domain which is specific to prenyl transferases were also identified. SOPMA analysis showed that Pkgdps consisted of 54.45% α-helix, 33.64% random coils, joined by 7.28% extended sheet and 4.58% β-turns (Fig. 4.35 D).

Hydropathy plots (Kyte and Doolite, 1982) of deduced amino acid sequence of Pkgdps suggested it to be hydrophilic protein (Fig. 4.35 B). This prediction further supported by the DAS-TM (Fig. 4.35 A) and PSORT (Fig. 4.35 C) analysis of deduced amino acid sequences. These predictions map the location of this protein to the chloroplast stroma with 0.86 certainty and to mitochondrial matrix with 0.63 certainty.
Fig. 4.35: Structural analysis of gene PkGDPS. (A) transmembrane domain, (B) hydropathy plot, (C) cellular locations, (D) predicted secondary structures (where alpha helices, extended sheets, beta turns and random coils are indicated by the longest, the second longest, the second shortest and the shortest vertical lines, respectively and (E) predicted conserved domain
4.6.3.8 Isopentyl pyrophosphate isomerase (Pkippi)

An amplicon of 400 bp was amplified utilizing degenerate primers (Fig. 4.36A). The nucleotide sequence of this amplicon exhibited 80% homology with ippi of *Nicotiana tabacum* (E-value=9e-124) and other reported ipp genes from plant and bacteria.

The nucleotide sequence of this amplicon was used to design primer for 5' and 3' RACE to obtain full length cDNA. The nested 5' and 3' RACE yielded distinct bands of 800 bp and 600 bp, respectively (Fig. 4.36 B).

![Fig. 4.36: Agarose gel electrophoresis showing amplified products of Pkippi: (A) partial cDNA using degenerate primers; (B) amplicons of 5' and 3' RACE. **Lane 1:** 5' RACE, **Lane 2:** 3' RACE, **Lane 3:** Nested 3' RACE; M: DNA size marker (bp) loaded on the gel; unlabelled arrow on the right hand side of the gel indicates the desired band.](image)

Full length sequence of putative *Pkippi* was obtained by aligning 5' and 3' RACE sequences. The full-length cDNA of *Pkippi* (GenBank accession number EF 421829) was 987 bp consisting of a coding region of 657 bp, 5' UTR of 138 bp, 3' UTR of 192 bp including a poly A⁺ tail of 27 bp (Fig. 4.37). The stop codon TAG and polyadenylation signal ATAAA was present at 793 bp and 853 bp, respectively. On comparing with other plant, it was found that *Pkippi* exhibited 80% to 85% homology with ippi of other plants.
Fig. 4.37: Nucleotide and deduced amino acid sequence of *Pkippi* (GenBank accession number EF199770). The start codon (ATG) is shown in bold, termination codon (TAG) is shown in bold and asterisk (*), polyadenylation signal (AATAA) is shown in bold and underlined.
<table>
<thead>
<tr>
<th>Species</th>
<th>Sequence</th>
<th>Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nicotiana</td>
<td>MGDVEADANMDAVQR1MFEDECILVDENDQVVGHDTKYUNCHLMERIESLNLHRAFSVF</td>
<td>60</td>
</tr>
<tr>
<td>Solanum</td>
<td>MVDVTDANMDAVQVR1MFDECAILVDENDQVVGHDTKYUNCHLMEKESENLNLHRAFSVF</td>
<td>60</td>
</tr>
<tr>
<td>Picrorhiza</td>
<td>MAAVTDSTMDAVQVR1MFDECAILVDENDQVVGHDTKYUNCHLMEKESENLNLHRAFSVF</td>
<td>60</td>
</tr>
<tr>
<td>Arnebia</td>
<td>---MAADAGMDAVQVR1MFDECAILVDENDQVVGHDTKYUNCHLMEKESENLNLHRAFSVF</td>
<td>57</td>
</tr>
<tr>
<td>Periploca</td>
<td>MGDVADSMADAVQVR1MFDECAILVDENDQVVGHDTKYUNCHLMEKESENLNLHRAFSVF</td>
<td>60</td>
</tr>
<tr>
<td>Stevia</td>
<td>---MGDSGMADVQVR1MFDECAILVDENDQVVGHDTKYUNCHLMEKESENLNLHRAFSVF</td>
<td>57</td>
</tr>
<tr>
<td>Hevea</td>
<td>MGDAPE-DTGMADVQVR1MFDECAILVDENDRAVGASKYSCHELWSNLNLHRAFSVF</td>
<td>59</td>
</tr>
</tbody>
</table>

Fig. 4.38: Alignment of deduced amino acid sequence of *P. nipponicum* with *ippi* homologs of other plant species. Hyphen (-) indicate gap introduced to optimize alignment. Identical amino acids are shown by star (*) whereas colons (:) indicate the consensus sequences. The conserved motifs are underlined. The GenBank accession numbers of different *ippi* used for alignment are: AB 049816 (Nicotiana; *Nicotiana tabacum*), EU 253957 (Solanum; *Solanum lycopersicum*), DQ 453138 (Arnebia; *Arnebia euchroma*), AB 091677 (Periploca; *Periplocea sepium*), DQ 989585 (Stevia; *Stevia rebaudiana*), AB 294697 (Hevea; *Hevea brasiliensis*).
Analysis of full length cDNA of Pkippi

The 657 bp cDNA encoded a putative Pkippi of 219 aa with predicted molecular mass of 27.19 kDa and pI of 4.98.

Deduced amino acid sequence shared 88% similarity with ippi of *Nicotiana tabacum*, *Stevia rebaudiana* and *Solanum lycopersicum*. This value was 80% for *Arnebia euchroma*, *Stevia rabaudiana* and *Hevea brasiliensis* (Fig. 4.38).

Conserved domain analysis at NCBI and SMART showed the presence of conserved cysteine residue in NXXCXHP motif, present at position 85-91 aa and conserved glutamate residue in EXE motif, present at position 95-98 aa (Fig. 4.38). The domain related to catalytic site, substrate binding site has been found (Fig. 4.39 E). The nudix domain, which is specific to *ippi* was also found. SOPMA analysis showed that Pkippi consisted of 52.34% α-helix, 30.64% random coils, joined by 11.06% extended sheet and 5.96% β- turns (Fig. 4.39 D).

Hydropathy plots (Kyte and Doolite, 1982) of deduced amino acid sequence of *Pkippi* suggested it to be hydrophilic protein (Fig. 4.39 B). This prediction was further supported by the DAS-TM (Fig. 4.39 A) and PSORT (Fig. 4.39 C) analysis of deduced amino acid sequences. These predictions map the location of this protein to the mitochondrial matrix and cytoplasm with nearly equal certainty (0.48 and 0.45, respectively).
Fig. 4.39: **Structural analysis of Pkippi:** (A) transmembrane domain, (B) hydropathy plot, (C) cellular location, (D) predicted secondary structures (where alpha helices, extended sheets, beta turns, and random coils are indicated by the longest, the second longest, the second shortest and the shortest vertical lines, respectively) and (E) predicted conserved domain.
4.6.3.9 Thiamine biosynthesis gene (*Pkthi*)

An amplicon of 350 bp was amplified from DD gel (Fig. 4.40). Nucleotide sequences analysis revealed that this 400 bp amplicon exhibited 78 % homology (E-value = 1e-149) with *thi* gene of *Nicotiana tabacum*.

![Agarose gel electrophoresis showing amplicon of the *Pkthi* using arbitrary and anchored primer. The DNA size markers M, loaded on the right side of the gel.](image)

The partial sequence was used to design primer for 5' and 3' RACE to obtain full length cDNA. The nested 5' RACE and 3' RACE generated distinct bands of 1kb and 400 bp, respectively (Fig.4.41 A and B).

![Agarose gel electrophoresis showing amplified product of *Pkthi* ; (A) amplicons of 5' RACE, (B) amplicons of 3'RACE and (C) full length ORF, M, DNA size marker (bp); unlabelled arrow on the right hand side of the gel indicates the desired band.](image)
Full length sequence of putative *Pkti* was obtained by aligning 5' RACE and DD sequences and was also cloned to full length using specific primers.

```
1  CGCGGGGATCCAGAGTCAGTGCTGTTGCATCTTTCGCTCTCTCAACTATCCCCTTTTTTT
  2  M      A      A      T      L      T      S      S      L      T
  30  AGTTCATCA
  48  GCAGCCATGGCCGCAACCTTAACCTCTTCCCTCACCAAATCTTCATTC
  56  L      D      T      K      S      S      I      Y      G      S
  74  L      D      N      L      T      F      A      P      I      K
  92  V      K      S      S      P      Q      N      L      S      V
  110  CTCGACACCAAGTCCTCCATCTATGGCTCTCCACTTTCCTCACGCGCCACCATCCAGCGC
  128  M      V      V      R      K      P      A      P      K      F
  146  Q      D      D      Y      V      V      I      K      H      A
  164  L      L      A      R      P      N      V      K      L      F
  182  GCGAGAGAGATCGTGCCTGGAATGATTGTTACTGGAATGGAAGTTGCTGAAATTGATGGT
  200  TCACCGAGAATGGGACCTACATTTGGAGCAATGATGATTTCAGGGCAAAAAGCAGCGCAC
  218  ATGGTCGTCCGCAAACCGGCCCCAAATTTCTCGACGAGCTGGAAATTGCTTACGACGAG
  236  CGGTACATGATGGATATGATCACCTACGCCGACACCGACGTCGTCGTCGTCGGCGCCGGA
  254  CAAGACGACTACGTTGTCATCAAGCACGCCGCTCTCTTCACCTCCACCATCATGAGCAAG
  272  GTGTTGATATGTGAAGATGGTGTACTCTGTTTTTTTGCTTGGAAATGCTAATCTTCTCTG
  290  GTTTTGTCCTTGCAAAAAAAAAAAAAAAAAAAAAAAAAAA
  308  TTGGCTTTGAAGGCTTTGGGGCAACCAAATGCGTTGGATGATTCATATGATGGATTGAAA
  326  ACGAAGACCGTAACATCTCAGACGCCGCTCTTCCCTTCACCTTCCACCATCATGAGCAAG
  344  GCTGAGGAATTTGTTTTGGCTGCGGCAGATGCTGGTGATATTGTCGATGCA
  362  AAGGGGGAATAAGAAAAAAAGTTGCAGATTGTGGTCAGAGGGGGGTGGAATTGGTG
  380  GTGTTGATATGTGAAGATGGTGTACTCTGTTTTTTTGCTTGGAAATGCTAATCTTCTCTG
  408  GTTTTGTCCTTGCAAAAAAAAAAAAAAAAAAAAAAAAAAA
  426  AGCGTGCCTGGGATGAAGGCGCTGGACATGAACACGGCGGAGGACGCCATTGTTAGGCTC
  444  GCGAGAGAGATCGTGCCTGGAATGATTGTTACTGGAATGGAAGTTGCTGAAATTGATGGT
  462  TCACCGAGAATGGGACCTACATTTGGAGCAATGATGATTTCAGGGCAAAAAGCAGCGCAC
  480  ATGGTCGTCCGCAAACCGGCCCCAAATTTCTCGACGAGCTGGAAATTGCTTACGACGAG
  508  CGGTACATGATGGATATGATCACCTACGCCGACACCGACGTCGTCGTCGTCGGCGCCGGA
  526  CAAGACGACTACGTTGTCATCAAGCACGCCGCTCTCTTCACCTCCACCATCATGAGCAAG
  544  GCTGAGGAATTTGTTTTGGCTGCGGCAGATGCTGGTGATATTGTCGATGCA
  562  AAGGGGGAATAAGAAAAAAAGTTGCAGATTGTGGTCAGAGGGGGGTGGAATTGGTG
  580  GTGTTGATATGTGAAGATGGTGTACTCTGTTTTTTTGCTTGGAAATGCTAATCTTCTCTG
  608  GTTTTGTCCTTGCAAAAAAAAAAAAAAAAAAAAAAAAAAA
  626  AGCGTGCCTGGGATGAAGGCGCTGGACATGAACACGGCGGAGGACGCCATTGTTAGGCTC
  644  GCGAGAGAGATCGTGCCTGGAATGATTGTTACTGGAATGGAAGTTGCTGAAATTGATGGT
  662  TCACCGAGAATGGGACCTACATTTGGAGCAATGATGATTTCAGGGCAAAAAGCAGCGCAC
  680  ATGGTCGTCCGCAAACCGGCCCCAAATTTCTCGACGAGCTGGAAATTGCTTACGACGAG
  708  CGGTACATGATGGATATGATCACCTACGCCGACACCGACGTCGTCGTCGTCGGCGCCGGA
  726  CAAGACGACTACGTTGTCATCAAGCACGCCGCTCTCTTCACCTCCACCATCATGAGCAAG
  744  GCTGAGGAATTTGTTTTGGCTGCGGCAGATGCTGGTGATATTGTCGATGCA
  762  AAGGGGGAATAAGAAAAAAAGTTGCAGATTGTGGTCAGAGGGGGGTGGAATTGGTG
  780  GTGTTGATATGTGAAGATGGTGTACTCTGTTTTTTTGCTTGGAAATGCTAATCTTCTCTG
  808  GTTTTGTCCTTGCAAAAAAAAAAAAAAAAAAAAAAAAAAA
  826  AGCGTGCCTGGGATGAAGGCGCTGGACATGAACACGGCGGAGGACGCCATTGTTAGGCTC
  844  GCGAGAGAGATCGTGCCTGGAATGATTGTTACTGGAATGGAAGTTGCTGAAATTGATGGT
  862  TCACCGAGAATGGGACCTACATTTGGAGCAATGATGATTTCAGGGCAAAAAGCAGCGCAC
  880  ATGGTCGTCCGCAAACCGGCCCCAAATTTCTCGACGAGCTGGAAATTGCTTACGACGAG
  908  CGGTACATGATGGATATGATCACCTACGCCGACACCGACGTCGTCGTCGTCGGCGCCGGA
  926  CAAGACGACTACGTTGTCATCAAGCACGCCGCTCTCTTCACCTCCACCATCATGAGCAAG
  944  GCTGAGGAATTTGTTTTGGCTGCGGCAGATGCTGGTGATATTGTCGATGCA
  962  AAGGGGGAATAAGAAAAAAAGTTGCAGATTGTGGTCAGAGGGGGGTGGAATTGGTG
  980  GTGTTGATATGTGAAGATGGTGTACTCTGTTTTTTTGCTTGGAAATGCTAATCTTCTCTG
 1008  GTTTTGTCCTTGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

Fig. 4.42: Nucleotide and deduced amino acid sequence of *Pkti* (GenBank accession number AY866497). The start codon (ATG) is shown in bold, whereas the termination codon (TAA) is shown in bold and marked by asterisk (*), polyadenylation signal is shown in bold and underlined.
Chapter 4

Results

The full-length cDNA (GenBank accession number AY866497) was 1361 bp, consisting of a coding region 1065 bp, 5' untranslated region (UTR) of 70 bp, 3' UTR of 226 bp including a poly A+ tail of 27 bp (Fig. 4.42). The stop codon TAA and polyadenylation signal AAATAAA were present at 1131 bp and 1198 bp position, respectively. The full length ORF was approximately 1100 bp was also cloned (Fig. 4.41 B).

Analysis of full length cDNA of Pkthi

The 1065 bp cDNA encoded a protein of 355 aa with predicted molecular mass 37.48 kDa and isoelectric point (pI) of 5.03.

Deduced amino sequence exhibited 81% similarity with Pkthi of Glycine max and Populus trichocarpa. This value was 78% to 76% for Nicotiana tabacum, Arabidopsis thaliana and 74% to 70% for Oryza stavia, Pseudotsuga menziesii and Gulliardia theta, respectively (Fig. 4.43). The alignment of translated Pkthi with other plant homologue revealed conserved secondary structure β1-α1- β2 that is known to bind FAD and the FAD binding, represented by consensus sequence GXGXGLXXA (where X denotes any residue) was present at position 94-103 aa residues. Important domains related to activity site, substrate binding site were also represented (Fig. 4.43). Conserved domain analysis at NCBI and SMART showed the presence of THI 4 domain (Fig. 4.44 E), which involved in thiazole biosynthesis in Pkthi. SOPMA analysis showed that Pkthi consisted of 44.07% α-helix, 31.36% random coils, joined by 16.67% extended sheet and 7.91% β- turns (Fig. 4.44 D).

Hydropathy plots (Kyte and Doolite, 1982) of deduced amino acid sequence of Pkthi suggested that this is hydrophilic protein (Fig. 4.44 B). This prediction further supported by the transmembrane (Fig. 4.44 A) and PSORT (Fig. 4.44 C) analysis of deduced amino acid sequences. These prediction map the location of this protein to the chloroplast stroma with certainty of 0.92.
Fig. 4.43: Alignment of deduced amino acid sequence of Pkthi with thi homologs from other plant species. Hyphen (-) indicates gap introduced to optimize alignment. Identical amino acids are shown by star (*) and colons (:) indicates the consensus sequences. The GenBank accession numbers of different thi used for alignment are: AK 244555 (Glycine; Glycine max), Y202080 (Nicotiana; Nicotiana tabacum), NM 124858 (Arabidopsis; Arabidopsis thaliana), AB110170 (Oryza; Oryza stavia), 832850 (Pseudotsuga; Pseudotsuga menziesii), CAH 25337 (Guillardia; Guillardia theta). The consensus sequence GXGXXGLXXA (where X denotes any residue) is underlined.
Fig. 4.44: Structural analysis of gene *PkThi*; (A) transmembrane domain, (B) hydropathy plot, (C) cellular, (D) predicted secondary structures predicted secondary structures (where alpha helices, extended sheets, beta turns and random coils are indicated by the longest, the second longest, the second shortest and the shortest vertical lines, respectively) and (E) predicted conserved domain.
Partial cloned genes

In spite of several attempts, full-length cDNAs of Pkcaff, PkhmgS and Pkpal could not be cloned to full length. Pkcaff and Pkpal RACE PCR could not yield the 5’ end while PkHMGs could not be extended at either length.

4.6.3.10 Caffeoyl-CoA-O-methyltransferase (Pkcaff)

The PCR with degenerate primer yielded an amplicon of 250 bp (Fig. 4.45). The sequences analysis revealed that the amplicon exhibited 78% homology (E-value = 1e-129) to caffeoyl-CoA methyltransferase sequences reported from the other plants.

![Image of gel electrophoresis](image)

**Fig.4.45:** Agarose gel electrophoresis showing amplified products of Pkcaff; (A) partial cDNA using degenerate primer, (B) amplicons of 5’ and 3’ RACE, Lane 1: 5’ RACE, Lane 2: Nested 3’ RACE Lane 3: 3’ RACE; unlabelled arrow on the right hand side of the gel indicates the desired, M:DNA size marker.

The partial sequence of Pkcaff was extended up to 800 bp by aligning 5’ (600 bp) and 3’-RACE (400 bp) sequence encompassing an ORF of 700 bp (Fig. 4.45 A and B). Homology search using BLAST showed more than 72% homology with caffeoyl-CoA O-methyltransferase of Plantago major. It had no start codon ATG (the 5’-end is incomplete), however it had the stop codon at position 597 bp (Fig. 4.46).
Chapter 4

Results

Fig. 4.46: Nucleotide and deduced amino acid sequence of Pkcaff. The termination codon (TAA) is shown in bold and marked by asterisk (*), and polyadenylation signal (AATAAAA) is marked bold and underlined.
Istatis
LREVTAK--HPWNMTTSADEQFLNMLIKLINAKNTEIGVYTGYSLLATALALPEDG-- 115
Arabidopsis
LREVTAK--HPWNMTTSADEQFLNMLIKLINAKNTEIGVYTGYSLLATALALPEDG-- 117
carthamus
LREVTAK--HPWNMTTSADEQFLNMLIKLINAKNTEIGVYTGYSLLATALALPEDG-- 110
codonopsis
LREWTAK--HPWNMTTSADEQFLNMLIKLINAKNTEIGVYTGYSLLATALAPDDG-- 105
Ammi
LREWTAK--HPWNMTTSADEQFLNMLIKLINAKNTEIGVYTGYSLLATALAPDDG-- 99
Bambusa
LREWTAK--HPWNMTTSADEQFLNMLIKLINAKNTEIGVYTGYSLLATALAPDDG-- 117
Glycine
LREWTAK--HPWNMTTSADEQFLNMLIKLINAKNTEIGVYTGYSLLATALAPDDG-- 117
Picrorhiza
-----------------SVNVFADEQGQTMMLKLMMKNTIEVGFYSGYLLATALALKDGK-- 48
Plantago
LRYATVEKVQKWLSMVPADEQDFSLMLKMNNKTIEVGFYSGYLLATALAPDDG-- 115

Istatis
-KILAMQVRENVELPGLPIIEKAGAHKIDFREGPALPVLDOVLVADKHNHGTYDFIVFDA-- 174
Arabidopsis
-KILAMQVRENVELPGLPIIEKAGAHKIDFREGPALPVLDOVLVADKHNHGTYDFIVFDA-- 176
carthamus
-KILAMQVRENVELPGLPIIEKAGAHKIDFREGPALPVLDOVLVADKHNHGTYDFIVFDA-- 169
codonopsis
-KILAMQVRENVELPGLPIIEKAGAHKIDFREGPALPVLDOVLVADKHNHGTYDFIVFDA-- 169
Ammi
-KILAMQVRENVELPGLPIIEKAGAHKIDFREGPALPVLDOVLVADKHNHGTYDFIVFDA-- 158
Bambusa
-TILAMQVRENVELPGLPIIEKAGAHKIDFREGPALPVLDOVLVADKHNHGTYDFIVFDA-- 176
Glycine
-KIILPQDREAYEIGLPEIKAGVEHIDFIESPALPVLKLEDPNSKESPFADFVFA-- 166
Picrorhiza
-VQVIAIPEAYIGLPEIKAGQVQFIDSNSALSIVKELIDNGE-EGTDFDFAFVDA-- 107
Plantago
-KIVAIDPQREAYEGLPIQKenAIHIGQFSQDAMKVMNEFLARGE-EGTDFDFAFVDA-- 173

нструments

Fig. 4.47: Alignment of deduced amino acid sequence of Pkea with caff homologs of other plant species. Hyphen (-) indicates gap introduced to optimize alignment. Identical amino acids are shown by star (*) whereas colons (:) indicates the consensus sequences. The GenBank accession numbers of different Caff used for alignment are: AAZ95245.1 (Isatis; Isatis tinctoria), NP_195131.1 (Arabidopsis; Arabidopsis thaliana), BAG71894.1 (Carthamus; Carthamus tinctorius), BAE48788.1 (Codonopsis; Codonopsis lanceolata), AAT4011.1 (Ammi majus; Ammi majus), ABO26812.1 (Bambusa; Bambusa oldhamii), ADX43927.1(Glycine; Glycine max) and CAJ43712.1 (Plantago; Plantago major).
Deduced amino acid sequence demonstrated significant similarity with *caffeoyl-CoA O-methyltransferase* proteins reported from *Plantago major* (81 %), *Populus trichocarpa* (80 %), *Isatis tinctoria* (79 %), and *Ammi majus* (77 %) (Fig. 4.47). Characteristic domain, methyl transferases was detected between 1-170 aa (Fig. 4.48 A). SOPMA analysis revealed that α-helices (41 %), β-turns (12 %), extended strands (21 %) and random coils (25.57 %) (Fig. 4.48 B). Hydropathy plot at using a window of size 19 indicated there is no TM domain (Fig. 4.48 C).

**Fig. 4.48: Structural analysis of Pkcaff;** (A) predicted conserved domain .and secondary structures, (B) predicted secondary structures (where alpha helices, extended sheets, beta turns and random coils are indicated by the longest, the second longest, the second shortest and the shortest vertical lines, respectively) and) and (C) hydropathy plot.
4.6.3.11 Phenylalanine ammonia lyase (Pkpal)

The PCR with degenerate primers yielded an amplicon of 250 bp (Fig. 4.49 A). The sequences analysis revealed that the amplicon exhibited 82 % homology (E-value = 1e-129) to pal sequence reported from Digitalis lanata the other plants.

![Image of gel electrophoresis](image)

**Fig. 4.49:** Agarose gel electrophoresis showing amplified products of *Pkcmk*; (A) partial cDNA using degenerate primer, (B) amplicons of 5'RACE, (C) amplicons of 3'RACE. M, DNA size marker (bp) loaded on the gel; arrow on the opposite side of the marker indicates the desired band.

The 862 bp long sequence was assembled by aligning 5' (600 bp) and 3'RACE (450 bp) sequence (Fig. 4.49 B and C). It encompassed an ORF of 647 bp. The start codon ATG could not be located suggesting that 5' end was incomplete; however stop codon TAG was present at position 643 bp. The polyadenylation signal AACAAA was present at position 665 bp along with a poly A+ tail of 25 bp (Fig. 4.50).
Fig. 4.50: Nucleotide and deduced amino acid sequence of Pkpal. The termination codon (TAA) is shown in bold and marked by asterisk (*), and polyadenylation signal (AACAAA) is marked bold and underlined.

Deduced amino acid sequence of Pkpal demonstrated significant similarity with pal reported from *Melissa officinalis* (92 %), *Salvia miltiorrhiza* (90%) and *Nicotiana attenuata* (86 %) (Fig. 4.51). It was found to be highly conserved protein among the all the plant species.
| Melissa  | TDLTHLKLHHPQIEAAAMEHILDSGYVKAQQKLHMDPLQPKQD4YARLRTSQW   | 353 |
| Salvia   | TDLTHLKLHHPQIEAAAMEHILDSGYVKAQQKLHMDPLQPKQD4YARLRTSQW   | 355 |
| Rehmannia| TDLTHLKLHHPQIEAAAMEHILDSGYVKAQQKLHMDPLQPKQD4YARLRTSQW   | 352 |
| Nicotiana| TDLTHLKLHHPQIEAAAMEHILDSGYVKAQQKLHMDPLQPKQD4YARLRTSQW   | 356 |
| Picrorhiza|                          |     |

| Melissa  | GPQIEVIRATMKMERIEINSVNDLPLDVARSKEAGGNGQGQTVGVMNALARIALSIG  | 413 |
| Salvia   | GPQIEVIRATMKMERIEINSVNDLPLDVARSKEAGGNGQGQTVGVMNALARIALSIG  | 415 |
| Rehmannia| GPQIEVIRATMKMERIEINSVNDLPLDVARSKEAGGNGQGQTVGVMNALARIALSIG  | 412 |
| Nicotiana| GPQIEVIRATMKMERIEINSVNDLPLDVARSKEAGGNGQGQTVGVMNALARIALSIG  | 416 |
| Picrorhiza|                          |     |

| Melissa  | KLLFLQFSELVNDVNYNGLPSNILSGRNPSLDYFGKSEIAMASYCSELQFLANPVTH    | 473 |
| Salvia   | KLLFLQFSELVNDVNYNGLPSNILSGRNPSLDYFGKSEIAMASYCSELQFLANPVTH    | 475 |
| Rehmannia| KLLFLQFSELVNDVNYNGLPSNILSGRNPSLDYFGKSEIAMASYCSELQFLANPVTH    | 472 |
| Nicotiana| KLLFLQFSELVNDVNYNGLPSNILSGRNPSLDYFGKSEIAMASYCSELQFLANPVTH    | 476 |
| Picrorhiza|                          |     |

| Melissa  | QSAEQHQDNVSLGLISSRTVEALILKLMSSTYVGLCQAIIDRLRLEENLKHAYMVNT  | 533 |
| Salvia   | QSAEQHQDNVSLGLISSRTVEALILKLMSSTYVGLCQAIIDRLRLEENLKHAYMVNT  | 535 |
| Rehmannia| QSAEQHQDNVSLGLISSRTVEALILKLMSSTYVGLCQAIIDRLRLEENLKHAYMVNT  | 532 |
| Nicotiana| QSAEQHQDNVSLGLISSRTVEALILKLMSSTYVGLCQAIIDRLRLEENLKHAYMVNT  | 536 |
| Picrorhiza|                          |     |

| Melissa  | SQVAKRLTGANGKPSRCEKDIILVVDREYVFAYDDCSTYPLIKLGKLQVLVEH  | 593 |
| Salvia   | SQVAKRLTGANGKPSRCEKDIILVVDREYVFAYDDCSTYPLIKLGKLQVLVEH  | 595 |
| Rehmannia| SQVAKRLTGANGKPSRCEKDIILVVDREYVFAYDDCSTYPLIKLGKLQVLVEH  | 596 |
| Nicotiana| SQVAKRLTGANGKPSRCEKDIILVVDREYVFAYDDCSTYPLIKLGKLQVLVEH  | 597 |
| Picrorhiza|                          |     |

| Melissa  | ALKNGGEGKAISTSFIPKIEAEEELKTLLPKEVESARTASESGNPAINANIRECSYPL  | 653 |
| Salvia   | ALKNGGEGKAISTSFIPKIEAEEELKTLLPKEVESARTASESGNPAINANIRECSYPL  | 652 |
| Rehmannia| ALKNGGEGKAISTSFIPKIEAEEELKTLLPKEVESARTASESGNPAINANIRECSYPL  | 654 |
| Nicotiana| ALKNGGEGKAISTSFIPKIEAEEELKTLLPKEVESARTASESGNPAINANIRECSYPL  | 656 |
| Picrorhiza|                          |     |

| Melissa  | YKFRÑEELGALSLTEGKVSPEECEDKVFTALSNGLIDPLLECLIQWGWAAPLIC  | 709 |
| Salvia   | YKFRÑEELGALSLTEGKVSPEECEDKVFTALSNGLIDPLLECLIQWGWAAPLIC  | 711 |
| Rehmannia| YKFRÑEELGALSLTEGKVSPEECEDKVFTALSNGLIDPLLECLIQWGWAAPLIC  | 708 |
| Nicotiana| YKFRÑEELGALSLTEGKVSPEECEDKVFTAMCNGQIIIDLELCEKEWGWAAPLIC  | 712 |
| Picrorhiza|                          |     |

**Fig. 4.51:** Alignment of deduced amino acid sequence of *PkPal* with *pal* homologs of other plant species. Hyphen (-) indicates gap introduced to optimize alignment. Identical amino acids are shown by star (*) whereas colons (:) indicates the consensus sequences. The GenBank accession numbers of different *Caff* used for alignment are: CBJ23826.1(Melissa; *Melissa officinalis*), ABD73282.1(Salvia; *Salvia miltiorrhiza*), AAK84225.1(Rehmannia; *Rehmannia glutinosa*) and  ABG75910.1(Nicotiana; *Nicotiana attenuate*).
A. 

![Diagram showing query set and non-specific hits]

B. 

![Graph showing secondary structures with peaks at 50, 100, and 150]

C. 

![Graph showing hydropathy plot with Window: 9]

**Fig. 4.52: Structural analysis of PkPal;** (A) predicted conserved domain and secondary structures, (B) predicted secondary structures (where alpha helices, extended sheets, beta turns and random coils are indicated by the longest, the second longest, the second shortest and the shortest vertical lines, respectively) and (C) hydropathy plot.

Characteristic *phenylalanine and histidine ammonia-lyases signature* was detected between 1-200 aa (Fig. 4.52 A). SOPMA analysis revealed α-helices (134 aa), β-turns (8 aa), extended strands (14 aa) and random coils (58 aa) (Fig. 4.52 B). Hydropathy plot analysis using a window of size 9 indicated that it is a hydrophilic enzyme (Fig. 4.52 C).
4.6.3.12 3-hydroxy-3-methylglutaryl CoA synthase (hmgs)

A 400 bp showed 90% homology with *Oryza sativa* and 87-84% homology with *Medicago truncatula, Hevea brasiliensis* and *Brassica juncea* (Fig. 4.53).

![Agarose gel electrophoresis](image)

**Fig. 4.53:** Agarose gel electrophoresis showing partial amplificon of Pkhmgs using degenerate primer. **M,** DNA size marker (bp) loaded on the gel; arrow on the right hand side of the gel indicates the desired band.

*Pkhmgs* could not be extended further by RACE inspite of several attempts, however, conceptual translation showed that it might be a part of ORF as it was free from any intron sequence (Fig. 4.54). Deduced amino acid sequence demonstrated significant similarity with hmg proteins reported from *Solanum lycopersicum* (81 %), *Hevea brasiliensis* (80 %), *Brassica juncea* (79 %), *Ricinus communis* (78 %) and *Camptotheca acuminata* (77 %) (Fig. 4.55).
Fig. 4.54: Nucleotide and deduced amino acid sequence of *Pkhms* of *P. kurrooa*. 
Fig. 4.55: Alignment of deduced amino acid sequence of PkHMGS with HMGS homologs of other plant species. Hyphen (-) indicates gap introduced to optimize alignment. Identical amino acids are shown by star (*) and colons (:) indicates the consensus sequences. The GenBank accession numbers of different HMGS used for alignment are: ACV65039.1 (Salvia; Salvia miltiorrhiza), AEC13715.1 (Catharanthus; Catharanthus roseus), BAF98279.1(Hevea; Hevea brasiliensis), XP_002509692.1 (Ricinus; Ricinus communis), XP_002298076.1 (Populus; Populus trichocarpa), AAG32923.1 (Brassica; Brassica juncea), XP_002468628.1(Sorghum; Sorghum bicolor) and NP_001170696.1 (Zea; Zea mays). Active site domain NtDVEGVDstNACYgG is underlined.
A.

B.

C.

**Fig. 4.56: Structural analysis of gene Phhmgs;** (A) predicted conserved, (B) predicted secondary structures (where alpha helices, extended sheets, beta turns and random coils are indicated by the longest, the second longest, the second shortest and the shortest vertical lines, respectively) and (C) hydropathy plot.

Characteristic domain, *hydroxymethylglutaryl-coenzyme A synthase* was detected between 10-120 aa and active site domain represented by motif NtDVEGVDstNACYgG was present at position 21-33 aa (Fig. 4.55 and 4.56 A). SOPMA analysis revealed α-helices (42 aa), β-turns (5 aa), extended strands (31 aa) and random coils (58 aa) (Fig. 4.56 B). Hydropathy plot indicated that it had characteristics of transmembrane domain (Fig. 4.56 C).
Chapter 4

4.7 Comparative Analysis of Cloned Genes

Intensive bioinformatics analyses showed that all the twelve genes contained the putative functional/conserved domains and predicted secondary structures, which are essential to render characteristic functionality in heterologous system such yeast, *E.coli* and complementation of function in mutants (Table 4.4). Presence of the desired functional domains and secondary structures in the genes cloned from *P.kurrooa* suggest them to be functional genes.

**Table 4.4**: Bioinformatics analyses and functional validation of genes from other plants, conserved domain present in *P. kurrooa* cDNA is represented by *.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Functional / conserved domain</th>
<th>Predicted secondary structure</th>
<th>Validation method</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>acdh</em> from <em>Raphanus sativus</em></td>
<td>*Thiolases acyl-enzyme intermediate signatures VNKvCASGMkAVmi aaqsl, and NvnGaVSIGPcGcG, <em>Thiolases active site GVGGVCNGgGgAsA</em></td>
<td>α-helices (183 aa), β-turns (34 aa), extended strands (54 aa) and random coils (135 aa)</td>
<td>Overexpression in yeast (Vollack and Bach, 1996)</td>
</tr>
<tr>
<td><em>hmgs</em> from <em>Brassica juncea</em></td>
<td><em>Hydroxymethylglutaryl-coenzyme A synthase active site: NdVEGVDstNACyG</em></td>
<td>α-helices (206 aa), β-turns (24 aa), extended strands (66 aa) and random coils (165 aa)</td>
<td>Overexpression in <em>E. coli</em> (Nagegowda et al., 2004)</td>
</tr>
<tr>
<td><em>hmgr</em> from <em>Arabidopsis thaliana</em></td>
<td><em>Hydroxymethylglutaryl-coenzyme A reductases signatures: signature 1 RfCCsTGDaMGcMnl signature 2 VGGVVGT signature 3 ALaAgqLvRSHMkY</em></td>
<td>α- helices (259 aa), β-turns (30 aa), extended strands (82 aa) and random coils (196 aa)</td>
<td>Heterologous expression in yeast (Learned and Fink, 1989)</td>
</tr>
<tr>
<td><em>ippi</em> from <em>Arabidopsis thaliana</em></td>
<td><em>Nudix hydrolase domain profile</em></td>
<td>α- helices (127 aa), β-turns (6 aa), extended strands (33 aa) and random coils (125 aa)</td>
<td>Enzyme activity assay (Campbell et al., 1997)</td>
</tr>
<tr>
<td><em>gdp</em> from <em>Abies grandis</em></td>
<td><em>Aspartate rich motif DD(X)2-4, D</em></td>
<td>α- helices (249 aa), β-turns (27 aa), extended strands (17 aa) and random coils (90 aa)</td>
<td>Heterologous expression in <em>E. coli</em> and enzyme activity assay (Burke and Croteau, 2002)</td>
</tr>
<tr>
<td>Gene</td>
<td>Source</td>
<td>Function/Structure</td>
<td>Secondary Structure/Activity</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------------------------</td>
<td>------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>dxs</td>
<td>Stevia rebaudiana</td>
<td>*TPP binding domain and conserved Histidine residue motifs</td>
<td>α-helices (263 aa), β-turns (50 aa), extended strand (103 aa) and random coil (299)</td>
</tr>
<tr>
<td>dsr</td>
<td>Stevia rebaudiana</td>
<td>*Conserved NADPH motif</td>
<td>α-helices (172 aa), β-turns (24 aa), extended strand (86 aa) and random coil (191)</td>
</tr>
<tr>
<td>cmk</td>
<td>Ginkgo biloba</td>
<td>*ATP-binding sites for the functional activity</td>
<td>α-helices (173 aa), β-turns (23 aa), extended strand (80 aa) and random coil (157)</td>
</tr>
<tr>
<td>hsr</td>
<td>Ginkgo biloba and Pinus taeda</td>
<td>*Three conserved cysteine residues</td>
<td>α-helices (263 aa), β-turns (50 aa), extended strand (103 aa) and random coil (299)</td>
</tr>
<tr>
<td>pal</td>
<td>Arabidopsis thaliana</td>
<td>*Phenylammonia lyase domain</td>
<td>α-helices (409 aa), β-turns (42 aa), extended strands (60 aa) and random coils (214 aa)</td>
</tr>
<tr>
<td>caffe</td>
<td>Vanilla planifolia</td>
<td>*Methyl transferase domain</td>
<td>α-helices (119 aa), β-turns (25 aa), extended strands (47 aa) and random coils (105 aa)</td>
</tr>
<tr>
<td>thi</td>
<td>Rhizobium eti</td>
<td>*Secondary structure β1-α1- β 2 and consensus motif GXGXXGLXXA</td>
<td>α-helices (123 aa), β-turns (24 aa), extended strands (56 aa) and random coils (120 aa)</td>
</tr>
</tbody>
</table>
4.8 Gene Expression Studies

Gene expression was studied by semi-quantitative RT-PCR using gene-specific primers and the cDNA prepared from different treatments. Since internal control gene primers were not reported for *P. kurrooa*, these were also developed as described in section 3.18 (Singh *et al.*, 2004).

4.8.1 Standardization of internal control gene primers for expression analysis

Primers designed from the conserved region of 26S *rRNA* were standardized, for use as an internal control for expression studies. RT-PCR performed with these primers amplified a 534 bp band from cDNA of *P. kurrooa*. Sequencing of the amplified fragment confirmed it to be a 26S *rRNA* fragment.

4.8.1.1 Standardization of PCR cycles and annealing temperature for 26s *rRNA* primers

Analysis of amplicon at 15, 20, 25 and 30 cycles showed that amplicon was well detectable at 25 cycles and appeared at exponential phase of amplification (Fig. 4.57). Hence, further experiments were carried out at 25 cycles for 26S *rRNA* amplification.

![Fig. 4.57: Amplification of 26S rRNA at varied numbers of PCR cycles.](image)

Although, the calculated Tm for primer was 52°C, PCR amplification of 26S *rRNA* amplicon was performed at temperatures ranging from 48-68°C. Amplification was obtained across all temperatures, though much less amplification was observed at 48°C (Fig. 4.58).

![Fig. 4.58: Amplification of 26S rRNA at different annealing temperatures.](image)
4.8.1.2 Expression of 26S rRNA in response to cDNA template quantity

Expression of 26S rRNA was studied in response to different concentrations (0.5 μl, 1.0 μl and 1.5 μl) of cDNA template. Increasing the cDNA template from 0.5 to 1.5 μl, amplification yield also increased accordingly, as evidenced by the increase in IDV from 25 to 83 and 131, respectively (Fig. 4.59).

![Amplification of 26S rRNA with varied concentrations of cDNA template.](image)

Fig. 4.59: Amplification of 26S rRNA with varied concentrations of cDNA template. Numerals above each lane represent microliters of the cDNA template used in PCR. Values inside the panel represent the IDV of the amplicon as measured using Alpha DigiDoc 1000 software.

4.8.1.3 Expression of 26S rRNA in response to developmental stages and environmental cues

Expression of 26S rRNA was studied in response to various developmental stages of leaves, different tissues and environmental cues (PEG 8000, GA₃). However, no change in gene expression was observed in all the tissues and the treatments under observation (Fig 4.60).

![Expression of 26S rRNA in different tissues and in response environmental cues.](image)

Fig. 4.60: Expression of 26S rRNA in different tissues and in response environmental cues. Expression was also studied in different tissues and in response to PEG and GA₃ as mentioned above each lane.

4.8.2 Gene expression during different time points of the day

Gene expression during different time points of the days was studied in first leaf, fourth leaf and rhizome of *P. kurrooa*. As shown in Table 4.2 and 4.3 all the leaf tissues has picrosides at various time of the day and also at various stages of developments.
4.8.2.1 Gene expression in the first leaf and fourth leaf:

In the first leaf (youngest leaf), out of the 11 gene analyzed (Fig. 4.61 A), expression of \textit{Pkcaff} and \textit{Pkpal} was maximum during late hours of the day (17:00 h) as compared to early hours (5:00 h to 9:00 h).

\textbf{Fig. 4.61: Expression of various genes during different time points of the day in the first and fourth leaf.} Light intensity (\text{\textmu}mol m^{-2} s^{-1}) at time of sampling is shown in parenthesis. Time of sample harvesting is mentioned above each panel and name of the genes are mentioned in between panels A and B. (ND: not detectable).

In mature leaf (fourth leaf), except for \textit{Pkpal}, \textit{Pkcaff} and \textit{Pkdxs} other genes did not exhibit alternation in their expression during early vs late hour (Fig. 4.61 B). As can be seen from Fig. 4.61 A and B, expression of \textit{Pkacth}, \textit{Pkhmgs}, \textit{Pkhmgr}, \textit{Pkippi}, \textit{Pkdxs}, \textit{Pkdxr},
Chapter 4

Results

*Pkcmk* and *Pkhdr* and *Pkgdps* did not exhibit any variation during the entire period of experimentation in young as well as mature leaf.

4.8.2.2 Gene expression in the rhizome

Fig. 4.62 depicts the expression of genes in rhizome tissue (underground parts). Expression of various genes of the picrosides biosynthetic pathway was similar at all the time points of the day.

![Gene expression in rhizomes](image)

**Fig. 4.62:** Expression of various genes during different time points of the day in the rhizome. Light intensity (μmol m⁻² s⁻¹) at time of sampling is shown in parenthesis. Time of sample harvesting is mentioned above each panel and name of the genes are mentioned in between panels A (ND; not detectable).
4.8.4 Expression of cloned genes in response to external cues

The leaves were treated with different types of putative modulators as described in the section 3.6. Leaf tissue showed the sign of wilting after 48 h of these treatments therefore, expression analysis was restricted up to 48 h and effect of these modulators on gene expression is described in following section:

4.8.4.1 Expression of genes in response to GA3

GA3 up-regulated the expression of Pkhrmg, Pkpal and Pkcaff at 24 h, whereas Pkcmk, Pkdxs, Pkdxr, Pkhdr and Pkgdps exhibited down-regulation at 24 h and 48 h of the treatment as compared to the respective control (Fig. 4.63).

Fig. 4.63: Effect of GA3 (200 μM) on expression of various genes in fifth leaf (from top, whereas top leaf was designated as first leaf). Leaves were treated with GA3 (200 μM) as described in section 3.6 and corresponding picrosides content for this experiment is shown in Fig. 4.4. Name of gene is shown on right side of the panel whereas duration of the treatment is mentioned above each panel. 26S rRNA was used as an internal control.
4.8.4.2 Expression of genes in response to ABA

Application of ABA down-regulated the expression of *Pkhmgr, Pkcmk, Pkdxs, Pkcaff* and *Pkgdps* at 48 h of the treatment as compared to the respective control. Expression of *Pkippi* exhibited up-regulation, whereas *Pkpal* and *Pkhdr* did not modulate at both the time intervals (Fig. 4.66).

![Image of gel electrophoresis showing expression of various genes in response to ABA](image)

**Fig. 4.64:** Effect of ABA (200 μM) on expression of various genes in fifth leaf (from top, whereas top leaf was designated as first leaf). Leaves were treated with ABA (200 μM) as described in section 3.6 and corresponding picrosides content for this experiment is shown in Fig. 4.4. Name of gene is shown on right side of the panel whereas duration of the treatment is mentioned above each panel. 26S rRNA was used as an internal control.
4.8.4.4 Expression of genes in response to MeJA

MeJA up-regulated the expression of \textit{Pkhmgr}, \textit{Pkippi} and \textit{Pkpal} at 24 h and 48 h of treatment as compared to the respective control. \textit{Pkdxs} exhibited up-regulation at 24 h and down-regulation at 48 h of treatment. The down-regulation was exhibited by (i) \textit{Pkhdr} at 24 h and 48 h, (ii) \textit{Pkaclth}, \textit{Pkdxr} at 48 h, and (iii) \textit{Pkgdps} at 24 h of the treatment (Fig. 4.67).

![Diagram of gene expression](image)

**Fig. 4.65:** Effect of MeJA (200 μM) on expression of various genes in fifth leaf (from top, whereas top leaf was designated as first leaf). Leaves were treated with MeJA (200 μM) as described in section 3.6 and corresponding picrosides content for this experiment is shown in Fig. 4.4. Name of gene is shown on right side of the panel whereas duration of the treatment is mentioned above each panel. 26S \textit{rRNA} was used as an internal control.
4.8.4.4 Expression of genes in response to H$_2$O$_2$

H$_2$O$_2$ up-regulated the expression of *Pkippi, Pkdxs, Pkhdr* and *Pkcaff* at 24 h and 48 h of the treatment as compared to the respective control. However, *Pkhmgr, Pkcmk, Pkdxr, Pkippi, Pkgdps,* and *Pkpal* did not modulate at both the time intervals as compared to the respective control (Fig. 4.68).

**Fig. 4.66:** Effect of H$_2$O$_2$ on expression of various genes of *P. kurrooa*. Leaves were treated with H$_2$O$_2$ (20 mM) as described in section 3.6 and picrosides content for these experiment is shown in Fig. 4.4. Name of gene is shown on right side of the panel; duration of the treatment is mentioned above each panel; 26S rRNA was used as an internal control.
4.8.4.5 Expression of genes in response to PEG

PEG up-regulated expression of *Pkacth*, *Pkhmgr*, *Pkcmk*, *Pkdxr*, *Pkippi*, and *Pkcaff* and down-regulated the expression of *Pkdxs* and *Pkgdps* at 24 h and 48 h of the treatment as compared to the respective control. Expression of *Pkpal* was unaffected during the experimentation period (Fig. 4.69).

Fig. 4.67: Effect of PEG (15%) on expression of various genes in fifth leaf (from top, whereas top leaf was designated as first leaf). Leaves were treated with PEG (15%) as described in section 3.6 and corresponding picrosides content for this experiment is shown in Fig. 4.4. Name of gene is shown on right side of the panel whereas duration of the treatment is mentioned above each panel. *26S rRNA* was used as an internal control.
4.8.4.6 Expression of genes in response to SA

SA up-regulated the expression of *Pkhmgr*, *Pkcmk*, *Pkdxr*, *Pkippi*, *Pkpal* and *Pkcaff* whereas down-regulated the expression of *Pkaclh*, *Pkdxs*, *Pkhdr* and *Pkgdps* as compared to respective control during the entire period of study (Fig. 4.70).

**Fig. 4.68:** Effect of SA (200 μM) on expression of various genes in fifth leaf (from top, whereas top leaf was designated as first leaf). Leaves were treated with SA (200 μM) as described in section 3.6 and corresponding picrosides content for this experiment is shown in Fig. 4.4. Name of gene is shown on right side of the panel whereas duration of the treatment is mentioned above each panel. 26S rRNA was used as an internal control.