

Full Length Research Paper

Characterization and the expression profile of 4-coumarate: CoA ligase (*li4CL*) from hairy roots of *Isatis indigotica*

Peng Di¹, Yongsheng Hu^{1,3}, Hongjiao Xuan^{1,5}, Ying Xiao¹, Junfeng Chen¹, Lei Zhang² and Wansheng Chen^{1,4*}

¹Department of Pharmacy, Changzheng Hospital, Second Military Medical University, Shanghai 200003, China.

²Department of Pharmacognosy, School of Pharmacy, Second Military Medical University, Shanghai 200433, China.

³Department of Pharmacy, the 118th Hospital of PLA, Wenzhou 325000, China.

⁴Modern Research Center for Traditional Chinese Medicine, Second Military Medical University, Shanghai 200433, China.

⁵School of Pharmacy, Jiamusi University, Jiamusi 154007, China.

Accepted 15 May, 2012

The enzyme 4-coumarate: CoA ligase (4CL) activates cinnamic acid and its hydroxylated derivatives by forming the corresponding CoA thioesters. These serve as substrates for biosynthesis of phenylpropanoid-derived secondary metabolites for the medicinal herbal *Isatis indigotica*. To investigate the function of 4CL gene, we have characterized this gene named *li4CL* with GenBank Accession No. GQ872418. The full-length cDNA of *li4CL* was 1967 bp and contained a 1632 bp open reading frame (ORF) encoding a 543 amino acid protein. Phylogenetic analysis places the *li4CL* into a third group distinct from the common type I and type II 4CL. Real-time quantitative polymerase chain reaction (PCR) analysis indicated that *li4CL* was expressed in roots, stems, leaves and flowers of *I. indigotica*, with the highest expression level in roots and flowers. The elicitor treatment experiments using methyl jasmonate (MeJA), abscisic acid (ABA) and UV-B revealed that *li4CL* respond to these elicitors in different manners. The full-length of ORF was sub-cloned into bacterial expression vector pET32a(+) and transferred into *Escherichia coli* BL21(DE3). The recombinant protein had high expression level in *E. coli* BL21(DE3) with isopropyl- β -D-thiogalactoside (IPTG) induction.

Key words: Phenylpropanoid pathway, UV-B, lignan, *Isatis indigotica* Fort., 4-coumarate:CoA ligase (4CL).

INTRODUCTION

Isatis indigotica Fort. (*Isatis tinctoria*) is a species of the economically valuable flower plant in Brassicaceae family. It is a biennial herbaceous plant species distributed widely in China. Indigowoad Root (Chinese name Ban-Lan-Gen) is a traditional Chinese medicine herb that comes from the roots of woad, and it is also known as Radix isatidis. In China, it is usually used for treating seasonal febrile diseases, pestilence, mumps, eruptive diseases, inflammatory diseases with redness of skin, sore throat, etc. (Ho and Chang, 2002). "Ban-Lan-

Gen-Ke-Li" and another two Chinese patent medicines which also contain the *I. indigotica* belonged to the eight selected Chinese patent medicine, which were screened from thousands of Chinese patent medicines for anti-SARS (Ni et al., 2008). Previous *in vitro* studies in our laboratory indicated that the lignans such as Lariciresinol and Larch Lignan glycosides from *I. indigotica* could inhibit Influenza virus A (AI/JF/262/95) (Bin, 2003).

The phenylpropanoid metabolic pathway, consists of a complex of diverse branches of biochemical reactions that synthesize an array of plant secondary metabolites, such as lignans, flavonoids, isoflavonoids, anthocyanins, stilbenes etc (Alfermann et al., 2008).

In most species, the general phenylpropanoid pathway

*Corresponding author. E-mail: smmuipb@gmail.com.

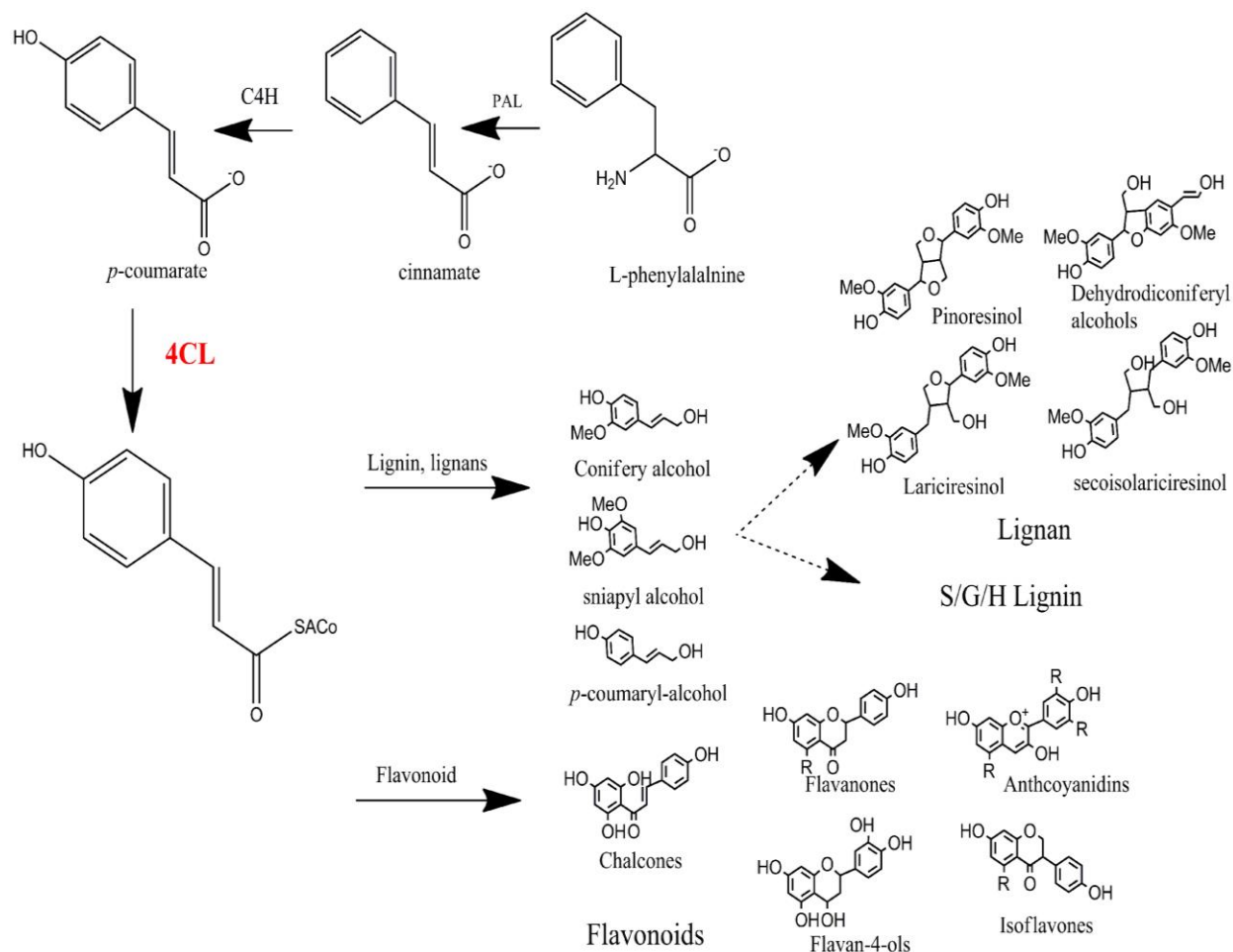


Figure 1. Phenylpropanoid biosynthetic pathways. The general phenylpropanoid pathway leading from phenylalanine to *p*-coumaroyl-CoA, the entry point to each major downstream pathway. Additional offshoot product classes are indicated. PAL, phenylalanine ammonia-lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate: CoA ligase.

transforms phenylalanine into the coenzyme A (CoA)-activated hydroxycinnamoyl (phenylpropanoid) thioesters. This structure is capable for the two major downstream pathways, monolignol and flavonoid biosynthesis. Three enzymes participate in the transformation process. Firstly, phenylalanine ammonia-lyase (PAL; EC 4.3.1.5) transforms the L-phenylalanine to cinnamic acid, the form of the phenylpropanoid skeleton, by decarboxylation. Then cinnamic acid 4-hydroxylase (C4H, EC 1.14.13.11) catalyzes the introduction of a hydroxyl group at the para position of the phenyl ring of cinnamic acid, producing *p*-coumaric acid. The last step, 4-coumarate: CoA ligase (4CL; EC 6.2.1.12) catalyzes the carboxyl group of *p*-coumaric acid by formation of a thioester bond with CoA. These activated intermediates serve as precursors for the formation of a number of natural products, such as monolignols, lignin and lignans, phenylpropenes and the flavonoids and stilbenes. 4CL catalyzes the final step in this three-step catalytic

process, it converts the C-flow to the *p*-coumaroyl-CoA which is capable for the downstream catalyze; so it is the crucial step in the common phenylpropanoid pathway. Because of the importance of phenylpropanoid-derived compounds in plants and the crucial catalyzed function in the phenylpropanoid pathway (Figure 1), 4CL has been the subject of extensive study for many years, mainly in higher plants. It has been cloned from *Petroselinum crispum* (Mill.) Hill (Douglas et al., 1987), *Solanum tuberosum* Linn (Becker-André et al., 1991), *Pinus taeda* Linn (Voo et al., 1995), *Nicotiana tabacum* (Lee and Douglas, 1996), *Glycine max* (Lindermayr et al., 2002) and *Arabidopsis thaliana* (Ehltting et al., 1999). 4CL is grouped into the super family of Adenylate-forming enzymes due to the presence of a highly conserved putative AMP-binding domain (Cukovic et al., 2001), and in many plants, 4CL is encoded by a small gene family (Cukovic et al., 2001).

By now, many researches about the function of 4CL is

mainly focused on lignins, the second most abundant group of plant biopolymers representing the most renewable carbon source on earth (Whetten et al., 1998), but few of the studies focus on other types of the secondary metabolites especially lignans, the structures with biological activities of antitumor, antiviral, protect liver, antioxidant, blood plate activating factor inhibiting (Lee and Xiao, 2003; Saleem et al., 2005). 4CL is on the first step of lignan biosynthesis pathway and it may be the initial step of the valuable lignans' accumulation.

In the present study, based on sequence homology, a 4CL gene from *I. indigotica* (*li4CL*) was isolated and characterized. The full-length of open reading frame (ORF) was sub-cloned into prokaryotic expression vector pET32a(+). The recombinant protein had high expression level in *Escherichia coli* BL21(DE3). The expression profiling analyses of *li4CL* under stress conditions and phytohormone treatment suggested that *li4CL* may be involved in environmental stress and hormone signaling, which serves as an initial step to further study about *li4CL* biosynthesis and improve the production of phenylpropanoid compounds in *I. indigotica* by metabolic engineering in the future. This study will enable us to further understand the role *li4CL* in the synthesis of phenylpropanoid compounds in *I. indigotica* at the molecular level

MATERIALS AND METHODS

Plant materials

Seeds of *I. indigotica* were purchased from the local market. The plant was grown in the gardens of Second Military Medical University, Shanghai, China, and identified by Professor ZHANG Han-Ming (School of Pharmacy, Second Military Medical University, SMMU). The specimen (Voucher specimen No. 138) was deposited in the Department of Pharmacognosy (SMMU).

Hairy roots culture

The *I. indigotica* hairy roots used in the present study were derived after the infection of plantlets with the Ri T-DNA (transfer DNA)-bearing *Agrobacterium rhizogenes* C58C1. The stock culture of hairy roots was maintained on solid, hormone-free 1/2B5 medium with 13 g/L agar and 30 g/L sucrose but without ammonium nitrate at 25°C in the dark.

Hairy roots for all the experiments in the present study were maintained in the shake-flask culture of 250 ml Erlenmeyer flasks, each filled with 100 ml of liquid 1/2B5 medium, on an orbital shaker at 120 rev./min at 25°C under dark for 30 days until further treatments, RNA and DNA isolation.

Treatments

The 30 days *I. indigotica* old hairy roots grown in shake-flask were sprayed with solution of 100 μM methyl jasmonate (MeJA), 100 μM abscisic acid (ABA) followed by expression analysis. Another set of control were similarly treated with distilled water. For UV-B treatment, the hairy-roots were exposed to 1500 J/m² UV-B light for 30 min, then take the hairy roots in dark.

RNA and DNA isolation

Root, leaves, stem and flowers of *I. indigotica*, as well as *I. indigotica* hairy roots under various treatments at selected time point were used for RNA isolation. Total RNA was extracted using TRIzol reagent (GIBCO BRL) according to the manufacturer's protocol. The genomic DNA of *I. indigotica* was isolated using the modified cetyltrimethyl ammonium bromide (CTAB) method. The quality and concentration of RNA and DNA samples were examined by ethidium bromide (EB)-stained agarose gel electrophoresis and spectrophotometric analysis.

Molecular cloning of the *li4CL* full-length cDNA

cDNA synthesis was based on the rapid amplification of cDNA ends method using oligonucleotide primer 5'-GACTCGAGTCGACATCGA(T)17-3'. RT-PCR reactions were carried out using the following primers- F: 5'-CAGGG(A/C)TA(T)/CGG(A/G)ATGAC(G/A/T)GA-3', R: 5'-CCTG(G/A)AA(G/A)CC(C/T)TT(G/A)TACTTGAT-3'. The degenerate primers F and R were designed based on highly conserved sequences in known 4CLs from other species. Two-round PCR reactions were conducted to amplify the 4CL core sequence. The first round PCR reaction used the primers F and R, and second round PCR (nested PCR) reaction used primers NF and NR. The amplified PCR product was purified and cloned into PMD18-T vector (TaKaRa, Japan) for sequencing.

According to the sequencing result, gene specific primers for 3'-RACE and 5'-RACE were designed. Molecular cloning of *li4CL* from *I. indigotica* was carried out by rapid amplification of cDNA ends (RACE) method using a SMARTer™ RACE cDNA amplification kit (Clontech, USA). For 3'-RACE of *li4CL*, about 100 ng of total RNA was reverse transcribed with 3'-CDS primer by SMARTScribe™.

Reverse transcriptase (Clontech, USA), Universal Primer A Mix, gene-specific primers 4CLF1069 (5'-CAGGGCTATGGAATGACTGAA-3') were used. The PCR was conducted accordingly to the protocol provided by the manufacturer (Clontech, USA). For 5'-RACE of *li4CL*, about 100 ng of total RNA was reverse transcribed with 5'-CDS primer and SMARTer IIA oligonucleotide (Clontech, USA). Universal Primer A Mix, gene-specific primers 4CLR1232 (5'-CAAGCATTCCCGCTGAACCAG-3') were used. The PCR was conducted accordingly to the manufacturer's protocol. Amplification conditions of 3' and 5'-RACE were performed as follows: 5 min at 94°C, 35 cycles of 30 s at 94°C, 30 s at 58°C and 3 min at 72°C, and a final extension of 10 min at 72°C. The amplified PCR product was purified and cloned into PMD18-T vector for sequencing.

The full-length *li4CL* from *I. indigotica* was obtained by aligning the products of 3', 5'-RACE and core sequence, and subsequently amplified by proof-reading PCR amplification with primers 4CLF (5'-AAAATGGAGAAATCCGGCTAC-3') and 4CLR (5'-TTCTTCACTTTGTGCGGTTTG-3'). The PCR was carried out in a total volume of 50 μl reaction solution containing 5 μl 10 × LA PCR buffer II (Mg²⁺ Plus), 8 μl dNTP Mix (2.5 mM), 2 μl of each primer (10 μM), 2 μl cDNA (3'-RACE RT product) and 2.5 units of LA Taq DNA polymerase using the following protocol: 5 min at 94°C, 35 cycles of 30 s at 94°C, 30 s at 56°C and 3 min at 72°C, and a final extension of 10 min at 72°C. The amplified PCR product was purified and cloned into PMD18-T vector for sequencing.

Isolation of cDNA and genomic DNA sequences coding for *li4CL* by PCR method

In order to detect potential introns within the *Li4CL* sequence, PCR amplification was carried out using the same conditions as that for

the cloning of the full-length cDNA except that the template was substituted by 1.5 µg of total genomic DNA and the extension time at 72°C in the amplification cycles was prolonged up to 4 min.

Gene accession numbers

The cDNA sequences obtained during the course of this research were deposited in the GenBank database with the following accession numbers: AY376729 (At4CL1), AY376728 (At4CL2), AY376730 (At4CL3), AY376731 (At4CL4), AY376732 (At4CL5), AY250839 (At4CL6), AY376733 (At4CL7), Q84P21 (At4CL9), AY376735 (At4CL11), AF279267 (Gm4CL1), AF002259 (Gm4CL2), AF002258 (Gm4CL3), X69955 (Gm4CL4), AF041049 (Pt4CL1), AF041050 (Pt4CL2), AF008184 (Pg4CL1), AF008183 (Pg4CL2), O24145 (Nt4CL1), O24146 (Nt4CL2), AY237163 (Sm4CL1), AY237164 (Sm4CL2), X13324 (Pc4CL1) and X13325 (Pc4CL2)

Bioinformatics analysis

ORF translation and BLAST were done on NCBI. Phylogenetic analysis of *li4CL* and other known 4CLs from other plant species retrieved from GenBank were aligned by MEGA software (version 5) and subsequently a phylogenetic tree was constructed by the neighbor joining (NJ) method using MEGA 5.0 software (Kumar et al., 2001). Molecular weight, isoelectric point and structural analysis of the deduced *li4CL* protein were predicted (Vector NTI Suite 11). The secondary and tertiary structures were predicted by GOR IV program (Combet et al., 2000) and SWISS-MODEL respectively (Kiefer et al., 2009).

Expression of *li4CL* in *Escherichia coli*

Using cDNA from RT-PCR as template, the *li4CL* gene was amplified with primer PF (5' - GCCgatatcATGGAGAAATCCGGCTACGG-3') and primer PR (5' - CGCggatcTCACATCTTGATCTTACTT-3'). The sequence underlined is the recognition site of the restriction enzymes EcoRV and BamHI, respectively. The amplified PCR products were ligated into the pEASY-T1 vector (TransGen Biotech, China) resulting in pEASY-T1/*li4CL*. Positive clones were confirmed by digestion with EcoRV and BamHI. The ligation of the *li4CL* fragment into the pET32a(+) vector (Novagen, USA) was followed by digestion using the same enzymes to obtain an in-frame fusion gene. The recombinant construct pET32a(+)-*li4CL* was identified by restriction digestion and the *li4CL* insert was verified by sequencing. The correct recombinant bacterial expression vector was named pET32a-*li4CL*. Transformation into *E. coli* Trans5α was carried out for amplification of the recombinant plasmid and *E. coli* BL21(DE3) competent cells (Novagen, USA) were used for T7-*li4CL* fusion protein expression. The transformant harboring plasmid pET32a(+)-*li4CL* were cultured at 37°C in Luria-Bertani (LB) medium until OD600 reached 0.6 to 0.8, isopropyl-β-D-thiogalactoside (IPTG, Merck, Germany) was then added to a final concentration of 1 mM. The cultivation was continued for another 1 to 6 h at 37°C. The cells were harvested by centrifugation and lysed by sonication. The total bacterial protein was analyzed by 12% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). The amount of the expression product was analyzed by UMAX Magic Scan software.

Expression profile of *li4CL* in different tissues and under various stresses

Total RNA was reversely transcribed by using AMV reserve transcriptase (Takara, Japan) to generate cDNA. Gene specific

primers 4CLF (5'-CTCGAATCCGAAGCTTATCG-3') and 4CLR (5'-AACACGCCGTGATACTCTCC-3') were designed according to the corresponding sequences of *li4CL*. Actin gene was amplified with primers (5'-ACCCTCACGGGGAAGACCATC-3') and (5'-ACCACGGAGACGGAGGACAAG-3') as a control. Real-time PCR was performed according to manufacturer's instruction (TaKaRa, Japan) under the following condition: 10 s pre-denaturation at 95°C, 1 cycle; 10 s denaturation at 95°C, 30 s annealing at 58.1°C, 40 cycles. The products of real-time quantitative PCR were run on 1.5% agarose gel electrophoresis and showed an equal-sized band as predicted. Quantification of the gene expression was done with comparative CT method. Experiments were performed in triplicate, and the results were represented by their mean ± standard error (SE).

RESULTS

Cloning of the full-length cDNA of *li4CL*

Using the RACE method and primers previously mentioned, cDNA ends of 889 bp and 1159 bp were amplified by 3'- and 5'-RACE, respectively. The two sequences were assembled with Vector NTI Advance 11 and the full-length *li4CL* cDNA was subsequently amplified by proof-reading PCR amplification with primers previously mentioned. The full-length cDNA of *li4CL* was 1967 bp and contained a 1632 bp ORF encoding a 543 amino acid protein (Figure 2A). BLAST search revealed that the nucleotide sequence of *li4CL* had sequence similarities of 90% with 4CL from *A. thaliana* (NM_116755.4) and 70% identity to 4CL from *P. trichocarpa* (XM_002329287.1), respectively. The analyses of genomic DNA sequences revealed that *li4CL* cDNA sequence and the genomic sequence were identical, which indicated that the *li4CL* gene contain 5 introns the size are 215, 80, 105, 94 and 102 bp.

Characterization of the deduced *li4CL* protein

By using the Vector NTI Advance 11 software, the calculated isoelectric point (pI) and molecular weight of the deduced *li4CL* were predicted to be 8.59 and 59.52 kDa, respectively. Protein-protein BLAST showed that on the amino acid level *li4CL* had high homology to 4CLs from other plant species. Multiple sequences alignments (Figure 2B) showed that two highly conservative regions were identified, the AMP binding site, a 11 amino acid residue SSGTTGISKGV motif I was found at position 197 to 207 of 4CL peptide chain; another motif 7 amino acid residue GEIWWVRG was found at position 389 to 395 and this is 4CL protein catalysis site. (Stuible et al., 2000). Conserved VPP and PVL domains with reference to *Gm4CL1* (Schneider et al., 2003) are also boxed. The secondary structure of *li4CL* protein, predicted by GOR IV (Combet et al., 2000), consists of 29.10% alpha helix, 22.88% extended strand and 44.62% random coil. Alpha helices mainly reside at the middle region of *li4CL* protein; while random coils and extended strand distribute at most

A

```

1  acatggggatcatcacactctctctctcttattcctctttagatcgggcaaa
  M E K S G I G R D G I R S L R P T L V L P K D
61  ATGSAAGAAATCCGCTACGCGCAGAGACGGCATTTACCGATCACTTCTGTCGAGCGCTCGTTCITCCCAAAGT
  P N L S L V S F L F R N S S S I P S K P A I V D
133  CCCAACCTATCTTATGCTCTCTTCTTATCCGTAATTCATCTTCTTACCGCTCGAAGCCAGCTATCGTCGAC
  S D S G D S L S F S Q L K S A V S R L A Q G F L
205  TCCGACTCCGCGGATCTTATCTCTCTCAGCTCAAATCGCGCTCTCTAGACTCGCCCAAGGATTTCTC
  R L G I R K N D V V L I F A P N S I Q F P L C F
277  CGTCTGGGCATCCGTAAGAACGCTCGTGTGTGATCTTCCGTCCTCAACTCTTACCAAGTCCCTCTCTGTTTC
  L A V S A I G G V F T T A N P L I T T N E V S K
349  CTCGCGTATCAGCAATCGCGCGGTTTTCACCCACCGCAATCTCTTTACACAGCAAGCAGGTTTCGAAG
  Q I K D S N P K L I V S V E R L F D K V K G F N
421  CAGATCAAGGACTCGAATCCGAAGCTTATCGTCTCCGTTGAGCGGTATTTCAGCAAGTCAAGGCTTTAAT
  L P V V L L G S G E S V Q I P E S D S K I L T F
493  CTCGCCGTCGCTGCTCGGTTCTGCGCAGTCCGATCAGATTCGCCAATCGGATCAAAAAATCCCTCACTTTC
  D N V M E L S D P V S D L P V V D I K Q S D T A
565  GACAAACGATGAGCTCTCCGACCCGTTTCGATCTCCCGTCTGTCGACATCAAGCAGTCGGACACAGCT
  A L L I S S G T T G I S K G V E L T H G N F V A
637  A S T L M T T M D Q D L M G E I H G V F L C F L P
  CGCTCTCGATGACGAGTGGATCAAGATCATGCGGAGATATCAGCGCTGTCTTGTGTTTCTTCCC
709  M F H V F G L A V I A I S Q L Q R G N A L I S M
  ATGTTTCATGTGTTTGGGCTTCCGCTGATTCGATTCACAGCTTCAGAGCGGAATGCTTTGATCTCCATG
781  A R F E L E A L L K N I E K I R V T Q L W V V P
  CGGAGATCGAGCTCGAGCGGCTTCTTAAGAACATTAAGAAATATAGAGTCACTCAGCTTTGGGTTGTTCCC
853  P V F L A L S K Q S I V K K I D L S S L K I I G
  CCTGTGTTCTCTTCTTCCAAAGCAGATATTGTAAGAAATATGATCTTCTTCCGTTAAGTACATTGGT
925  S G A A P L G K D L M E E C G R N I P N I V L M
  TCAGGTGCTGCTCTCTTGGGAAGATTTGATGGAAGAGTGTGGAAAAACATCCCTAATATTGTCCTTATG
997  Q G I G M T E T C G I V S V E D P R L G K R N S
  CAGGGCTATGGAATGACTGAAACGCTGTGGGATTTGCTCTGTGGAAGACCCGAGGCTCGGGAAGGAACCTC
1069  G S A G M L A P G V E A Q I V S V E T G K S Q P
  GGTTGAGCGGGAATGCTTGCACAGGTGTTGAGCTCAGATAGTGTGTGTAAGAACTGGAATACTCAACT
1141  P N Q Q G E I W V R G A N M M K G I F N N P Q A
  CCGAATCAACAGGCGAATTTGGGTTCCGAGGGCTAACATGATGAAAGGTATTTCATTAACCTCAGGCT
1213  T K E T I D K K G W V H T G D L G I F N E D G N
  ACTAAGAAACAATAGATAAAAAAGGTGGGTACATACCGAGATCTTGGTTACTCAACGAAGATGGTAAC
1285  L F V V D R I K E L I K Y K G P Q V A P A E L E
  CTTTTGTTGATGATGAAATCAAGAACTTATCAAAATCAAAAGGCTTCCAGGTTGCTCCAGCTGAACAGTAA
1357  G L L V S H P E I L D A V V I P P P D E E A G E
  GGTCTTCTGTGTGCGCAACCCGAGATACTAGATGCTGTGCTATCCCGTTCCTGATGAAGAAGCAGGCGAA
1429  V P I A F V V R S P K S S I T Q E D I Q K F I A
  GTTCCAAATGCGTTTGTGTGATGATCAACAAAAGCTCAATACACAGGAAGATATCAAAAGTTCATAGCC
1501  K Q V A P I K R L R R V S F I S S V P K S A A G
  AARACGGTCTGCTCTTAACAAAGATTCGGAAGAGTGAAGCTTATTAGCAGTGTTCAAAATCTGCTGCTGT
1573  K I L R R E L V Q Q V R S K M *
  AAAATCTTGAGAAGAGAGCTTGTTCAGCAAGTAAAGATCCAAAGTGTGA
1645  AGCATCTTACTACCGGAGCCTATCTATGTTGCTCGTCTTTTCAACATTTTTATGTAACATTCNAAT
1633  CAACAGTTACGCAACCGCACAAAGTGAAGAAAAATATCAAGAACCAAGGTCAGATATATACAAAT
1705  TGTTCTCTCTATCTTATGCTATAAACCAAGCACTTTTATTGCTCTTCAACAGAAAACATAACTCTGTATA
1777  TTATTAATCTTTTTTGTACAGAGCTTGAATGATAAAAAAARAAAAAAAAAAAAAAAAAAAA
1849  TTATTAATCTTTTTTGTACAGAGCTTGAATGATAAAAAAARAAAAAAAAAAAAAAAAAAAA
  
```

B

BOX I

```

* 240 * 260 * 280 * 300 * 320 *
Ii4CL : QSDTAALLYSSGTTGLPKGVLTTHKGLVTHGNFIA--ASLMTMDQDLMEYGHVFLCFLPMFHVFLAVIAYSQLRGNALISMARFELEALLKNIKRYVTQLWVPPVFLALS : 295
At4CL1 : PDDVVALFYSSGTTGLPKGVLTTHKGLVTSVAQQVDGENPNLYFHSDDVILCVLPMFHYALNSIMLCSLRVGAAILIMPKEIINLLELQIRCKVTVAMVPPVLAIA : 310
At4CL2 : PEDVVALFYSSGTTGLPKGVLTTHKGLVTSVAQQVDGENPNLYFNRDDVILCVLPMFHYALNSIMLCSLRVGAAILIMPKEIITLLELQIRCKVTVAMVPPVLAIA : 303
At4CL3 : GDDAAALFYSSGTTGLPKGVLTTHKGLVTSVAQQVDGENPNLYLKSNDVILCVLPMFHYALNSVLLNSRSGATVLLMHHKFEIGALLDLIQRHVTIALLPPLVLAIA : 313
At4CL5 : PEDVVALFYSSGTTGLPKGVLTTHKGLVTSVAQQVDGENPNLNFANDVILCVLPMFHYALDALMLSAMRTGAALLVPRFELNVMELIQRVTVVPPVLAIA : 318
Gm4CL1 : PDELVALFYSSGTTGLPKGVLTSHKLVTTIAQLVDGENPHQYTHSEVILCVLPMFHYALNSILCGLRVGAAILIQKFEITLLELIEKYKVTVASFPPIVLAIA : 292
Nt4CL1 : PDDVVALFYSSGTTGLPKGVLTTHKGLVTSVAQQVDGENANLMHSEVILCVLPMFHYALNSILCGLRVGAAILIMQKFDIAPFLELIQKYKVTVASFPPIVLAIA : 294
Nt4CL2 : PDDVVALFYSSGTTGLPKGVLTTHKGLVTSVAQQVDGENPNLYIHSDEVMLCVLPMFHYALNSVLLCGLRVGAAILIMQKFDVSVLELIQRVTVVPPVLAIA : 289
Sm4CL1 : ADDVVALFYSSGTTGLPKGVLTSHKLVACVSVQVDGENPAVHIDREDVILCVLPMFHYALNSVLLCGLRVGAAILIMQKFEISELMELIEKYKVTVASFPPIVLAIA : 289
Sm4CL2 : PEDAVALFYSSGTTGLPKGVLTTHKGLVTSVAQQVDGENPNLYIHSDEVMLCVLPMFHYALNSVLLCGLRVGAAILIMQKFEIVPFLDMQRKYKVTVASFPPIVLAIA : 289
  
```

BOX II

```

340 * 360 * 380 * 400 * 420 * 440
Ii4CL : KQSIKKYDLSSLYKYGSAAPLGLKDLMEECGNLPTMLMQGYMTEGCHVSVV--DPRLGKRNKSGSAGMLAPGVEAQIVSVETGKSGPPNQGEIIVRCAANMVKGY : 402
At4CL1 : KSSETEKYDLSSIRVVKSAAPLGLKELEDVNAKFPNAKLGQGYMTEAGPFLAMSLGFAKEPFPVKSGAGCTVVRNAEMKIVDPDGTGSLRNQFGEIIVRCAANMVKGY : 420
At4CL2 : KSPETERKYDLSSVRMVKSAAPLGLKELEDVNAKFPNAKLGQGYMTEAGPFLAMSLGFAKEPFPVKSGAGCTVVRNAEMKIVDPDGTGSLRNQFGEIIVRCAANMVKGY : 413
At4CL3 : KNPTVNSYDLSSVRVLSGAAPLGLKELQDLRRRLPQALIGQGYMTEAGPFLAMSLGFAKEPFPVKSGAGCTVVRNAELKVVHLETRLSLQVYFGEIIVRCAANMVKGY : 423
At4CL5 : KSPETERKYDLSSVRMVKSAAPLGLKELEDVNAKFPNAKLGQGYMTEAGPFLAMSLGFAKEPFPVKSGAGCTVVRNAEMKIVDPDGTGSLRNQFGEIIVRCAANMVKGY : 427
Gm4CL1 : KSGETHRYDLSSIRAVVTAAPLGLGELQEAVKARLPHATFGQGYMTEAGPFLAMSLGFAKEPFPVKSGAGCTVVRNAEMKIVDPDGTGSLRNQFGEIIVRCAANMVKGY : 401
Nt4CL1 : KSPVDSYDLSSVRTVMSGAAPLGLKELEDVNAKFPNAKLGQGYMTEAGPFLAMSLGFAKEPFPVKSGAGCTVVRNAEMKIVDPDGTGSLRNQFGEIIVRCAANMVKGY : 404
Nt4CL2 : KSPMVDYDLSSVRTVMSGAAPLGLKELEDVNAKFPNAKLGQGYMTEAGPFLAMSLGFAKEPFPVKSGAGCTVVRNAEMKIVDPKTNLSLRNQFGEIIVRCAANMVKGY : 399
Sm4CL1 : KSPAARKDFDSSVRVVKSAAPLGLKELEDVNAKFPNAKLGQGYMTEAGPFLAMSLGFAKEPFPVKSGAGCTVVRNAEMKIVDPDGTGSLRNQFGEIIVRCAANMVKGY : 398
Sm4CL2 : KSPLVAKYDLSSVRMVKSAAPLGLKELEDVNAKFPNAKLGQGYMTEAGPFLAMSLGFAKEPFPVKSGAGCTVVRNAEMKIVDPDGTGSLRNQFGEIIVRCAANMVKGY : 399
  
```

Figure 2. (A), The full-length cDNA sequence and the deduced amino acid sequence of Ii4CL; (B), amino acid sequence comparison of At4CLs with 4CL1 from *A. thaliana*, *G. max*, *N. tabacum*, and *Salvia miltiorrhiza*. 4CLs alignment was performed by ClustalX2(2.0) motif I SSGTTGISKGV putative AMP binding domain with 11amion acid and , and motif II the conserved GEICIGR motif (Stuible et al., 2000). Conserved VPP and PVL domains with reference to Gm4CL1 (Schneider et al., 2003) are also marked.

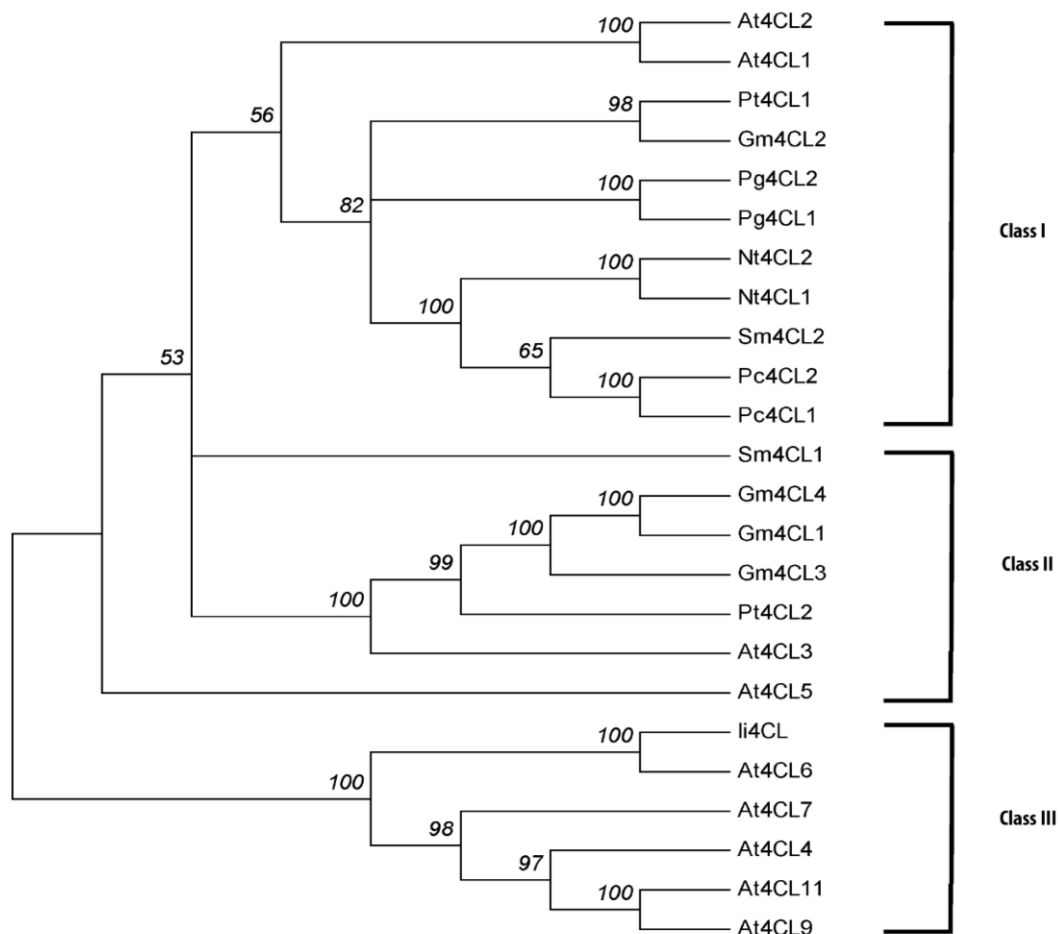


Figure 3. Phylogenetic relationships among plant 4CLs from *A. thaliana*, *G. max*, *N. tabacum*, *S. miltiorrhiza*, *Populus generosa*, *P. tremuloides* and *Petroselinum crispum*. The phylogenetic tree was constructed by the NJ method using MEGA 5.0 software (Thompson et al., 1994; Saitou and Nei, 1987; Kumar et al., 2001).

region of *li4CL* protein.

Phylogenetic analysis

In order to study the evolutionary relationships among different 4CL proteins from various plant species, a phylogenetic tree was constructed based on the deduced amino acid sequences of predicted *li4CL* and other 4CLs (Figure 3). Phylogenetic analysis revealed that the 4CL genes form separate clades. 4CLs in dicots can be divided into Class I and Class II. But *li4CL* was grouped into Class III with some *At4CLs*, which is distinct from the lignin-associated type I and type II 4CLs found in dicots.

Expression of *Li4CL* in *Escherichia coli*

The *li4CL* gene ORF was expressed as a fusion protein in *E. coli* BL21 (DE3). The pET32a(+)-*li4CL* recombinant

plasmid was constructed by sub-cloning the ORF of *li4CL* into the EcoRV and BamHI site of pET32a(+). Figure 4 (lanes 4 to 6) shows the expression of the fusion protein. A specific protein band of approximately 75 kDa (including the his-tag) was observed from IPTG induced pET32a(+)-*li4CL* cultures, which absent in cultures without IPTG induction. The results illustrate the recombinant proteins had high expression level in *E. coli*

Tissue-specific and induce expression profile of *li4CL*

To investigate the expression profile of *li4CL* in different tissues of *I. indigotica*, total RNA was isolated from root, stem, leaves and flowers tissues, respectively, and then the expression profile were determined by real-time quantitative PCR. The results showed that *li4CL* expression could be detected in all tissues but at different expression levels, with the strong expression in roots,

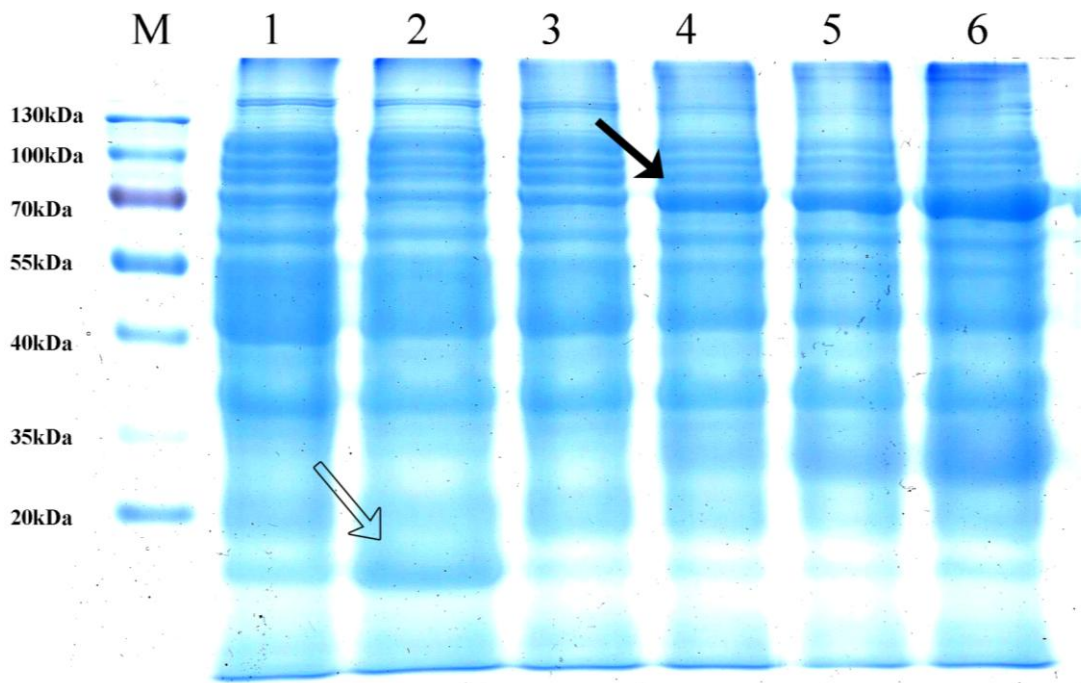


Figure 4. Expression of pET32a-*li4CL* fusion for different time. M, protein marker; lane 1, uninduced pET32a; lane 2, induced pET32a for 3 h; lane 3, uninduced pET32a-*li4CL*; lane 4, 5, and 6, pET32a-*li4CL* induced by 1 mM IPTG for 1, 3 and 5 h, respectively. The white arrow shows the his-tag protein 16 kDa, and the black arrow shows the recombinant *li4CL* protein 75 kDa.

moderate expression in flowers and weak expression in stems and leaves (Figure 5). The downstream metabolites of phenylpropanoid pathway such as lignins and flavonoids are in response to various environmental signals, many genes were reported inducing by several signal molecules with low molecular masses such as MeJA, GA₃ and ABA. However, little is known about the expression profiles of *4CL* in *I. indigotica* under elicitor treatments. Therefore, it is worthwhile studying expression patterns of *4CL* in this specie under various treatments, which will be helpful to reveal molecular induction mechanism for further improving the biosynthesis of metabolites on phenylpropanoid pathway in *I. indigotica*.

To understanding the role of *li4CL* regarding to environment signals, plants were treated with signal molecules MeJA, ABA and UV-B. For the time-course experiment, hairy roots of *I. indigotica* were treated with MeJA, ABA and UV-B, harvested for RNA isolation at different time points. Under the MeJA induction experiment, *li4CL* gene expression was responded to MeJA induction directly, the expression rises after 2 h reaches to the highest at 4 h then keeps high expression to 8 hours then decreases a little until 24 h (Figure 5). As shown in Figure 5, ABA induced *li4CL* expression in a different way, with a little response until 12 h then reach the highest level observed at 24 h (More than 6-fold higher compared to the control experiment). Under UV-B

treatment, the response is very fast, it rises to the highest expression level from 5 to 10 min and decrease at 30 min under UV-B. 30 min after turn off UV-B the expression level was decrease to low at 30 min then increase to high at 60 min then decrease a little at 120 min (Figure 5).

The differences between MeJA and ABA induction may be caused by the different mechanism of signal molecular and phytohormones. The UV-B results may be caused by the protection and repair mechanism.

DISCUSSION

The biosynthesis of several important hormones and other secondary metabolites depend on benzoyl- and/or cinnamoyl-CoA as substrates (Ribnicky et al., 1998; Graser et al., 2001) so characterization of the *4CL/4CL*-like genes may help to expand our understanding of the biosynthesis of these compounds. In dicots, 4CLs can be classified into Class I 4CLs, which are associated with lignin accumulation, and Class II 4CLs, which are involved in the metabolism of other phenolic compounds (Hu et al., 1998; Ehling et al., 1999; Harding et al., 2002). Sequence analysis revealed that *li4CL* belonged to a third phylogenetic group (Class III). Compare to Shockey and Fulda (2003) study, *li4CL* was grouped into Clades V in their study, and they found that the clade V enzymes might participate in the metabolism of very long-

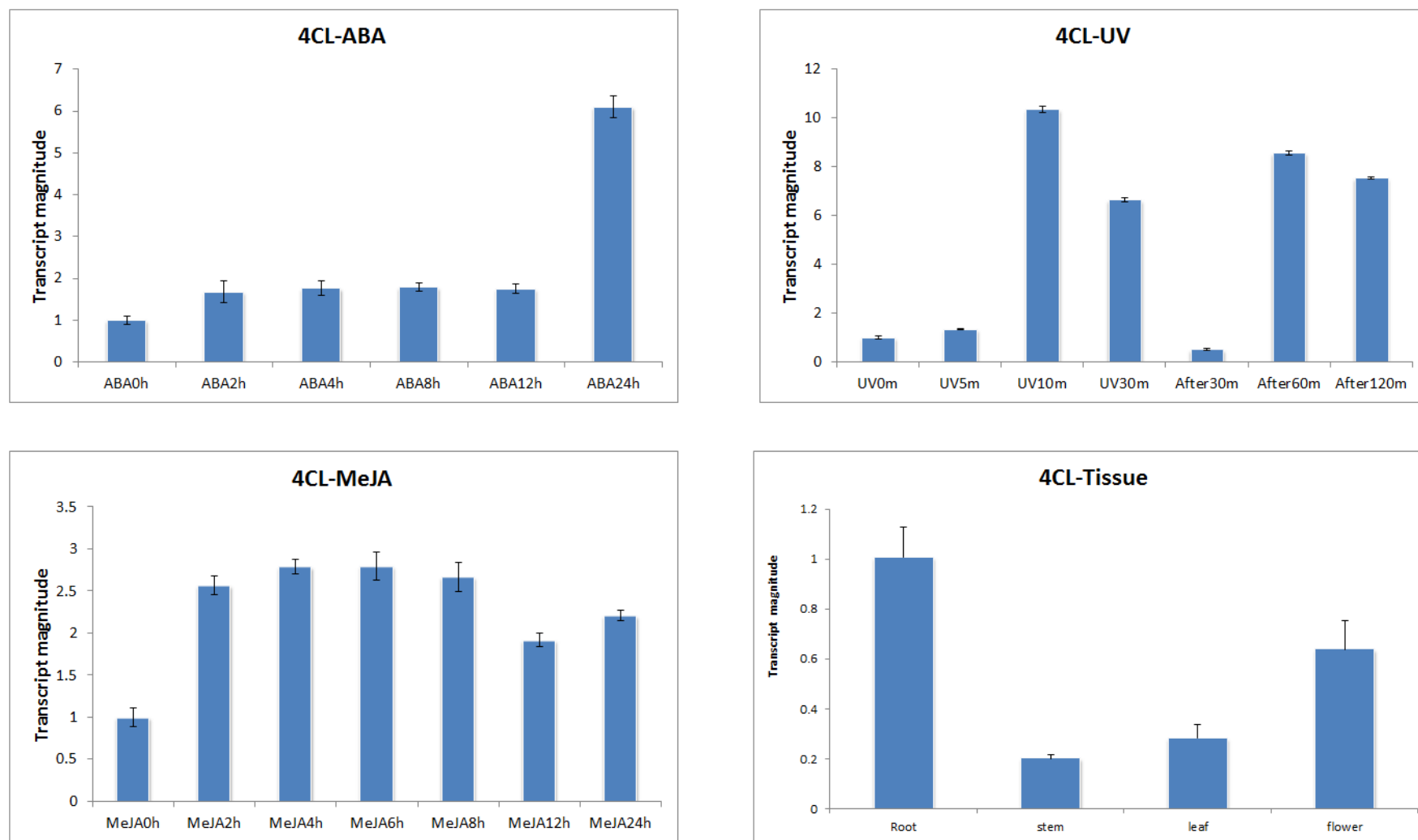


Figure 5. Expression profiling analysis of *li4CL* in different tissues of *I. indigotica*, and under various treatments (100 Mm MeJA, 100 μ M ABA, 1500 J/m² UV-B) during plant culture period, with plants similarly treated with distilled water as control. Data represents the mean values \pm SE of three replicates.

chain (>C22) fatty acids. Very long-chain fatty acids, and the esters, alcohols, ketones, and

aldehydes derived from them are major components of the surface wax layer that covers

the aerial parts of all terrestrial plants.

The general phenylpropanoid metabolism

generates an enormous array of secondary metabolites, these compounds anticipate into various plant physiology processions. As a key enzyme on the pathway, the expression level of 4CLs in different plant tissue and organ may concern about accumulation of some functional compounds which can affect and regulate plant growth and development. We found that *li4CL* was highly expressed in roots and flowers, in stem and leaves it expressed weakly. The point that 4CL is related to the lignin production has been extensively investigated, 4CL silencing in tobacco (*N. tabacum*), Arabidopsis (*A. thaliana*), and *P. tremuloides* caused lignin reductions in the range of 25 to 45% (Kajita et al., 1997; Lee et al., 1997; Hu et al., 1999; Li et al., 2003), so the highly expression of 4CL in root may cause by the lignification procession which needs plenty of lignin. This result also indicate that the synthesis of phenylpropanoid is mainly located in root, which may confirm that the root of *I. indigotica* "Ban-Lan-Gen" was used as a traditional Chinese medicine. Previous study reveal that in Raspberry anthocyanin pigments (Kumar and Ellis, 2003) whose biosynthesis is also dependent on activated 4-coumaric acid as a precursor, in *I. indigotica*, flower may need more precursor for anthocyanin pigments so it caused the highly expression of *li4CL*. In stem and leaves, they might not have the strong lignification process and the flavonoid and anthocyanin was not needed for the growth, the precursor support pathway was not stimulation so that the expression of 4CL was on a low level.

Transcriptional activation of 4CL genes was previously found after elicitor treatment of cell cultures from various plant species, including parsley, soybean and arabidopsis (Douglas et al., 1987; Trezzini et al., 1993; Lindermayr et al., 2002). Expression of 4CL genes was also induced in different plants of parsley, potato, soybean and arabidopsis following with bio or abiotic elicitor and environmental stress, such as pathogens, wounding, UV-B and phytohormone induction (Schmelzer et al., 1989; Becker-André et al., 1991; Lindermayr et al., 2002). Under these conditions, phenylpropanoid derivatives play an important defensive role (Dixon and Paiva, 1995).

UV-B boosts transcription levels for phenylalanine ammonia lyase, 4CL chalcone-flavone isomerase and dihydroflavonol-4-reductase (Greenberg et al., 1997) these enzymes are important on the phenylpropanoid pathway. It is reported that flavonoids and sinapic esters protect by specifically absorbing in the wavelength region from 280 to 340 nm. So we predict that in *I. indigotica* the 4CL should be responded to UV-B. *li4CL* responded to UV-light treatment intensely in our research is in agreement with our previous prediction, the response is very fast while the hairy-roots under UV-B. But 30 min after turn off, UV-light the expression level was decreased to low at 30 min then increases to high at 60 min then decrease a little at 120 min; this result was interesting. Since plants are exposed to continuously varying levels of UV-B, they may well be continuously adjusting their UV

defense. But in our research, the second high expression time was at 60 min after turn off UV-B, it should not cause by the protection and probable caused by the repair process. So we infer that, the protection and repair process may happen sequentially, and the compounds anticipated into protection and repair may different, so that the 4CL needed to express highly in different time, under the UV-B and after the UV-light. The details about the specific compounds anticipated into different process need to further research.

Jasmonates (JAs) are plant-specific signaling molecules that steer a diverse set of physiological and developmental processes. Pathogen attack and wounding inflicted by herbivores induce the biosynthesis of these hormones, triggering defense responses both locally and systemically. MeJA can be the induction of phenylpropanoid and its downstream metabolic metabolism such as lignins and flavonoids. (Pauwels et al., 2008). The result is in accord with our prediction, *li4CL* expression was upregulated in response to MeJA treatment after 2 h then kept at the high level until 8 h, then decreased slightly at 12 and 24 h; we found that the increase level of MeJA treatment was not significant compare to UV-light treatment, this difference may be caused by the individual induction mechanisms.

The phytohormone ABA is a major regulator of plant development and stress responses, including seed dormancy, germination, stomatal aperture regulation and drought resistance responses (Cutler et al., 2010; Raghavendra et al., 2010). The expression of *li4CL* was slightly changed after treatment until 24 h. At 24 h, it significantly raised to 6-fold higher than untreated. In Arabidopsis, 4CL expression was upregulated after 20 h treatment with 50 μ M ABA after 20 h (Böhmer and Schroeder, 2011). These results indicated that the phenylpropanoid pathway may not be fast but high degree responded to ABA.

In conclusion, we successfully isolated and characterized the full-length *li4CL* cDNA from the medicinal plant *I. indigotica*. We investigated the expression profile of *li4CL* in different tissues and in different treatments. The results will help us to understand the role of 4CL gene involved in the defense/stress and hormone response pathways at the molecular level. The hairy roots regeneration and transgenic system of *I. indigotica* has already been established in our laboratory (Hu et al., 2011), which will help to prompt the possibility to apply a target of 4CL gene for the exploitation of a metabolic-engineering approach to manipulating the lignan biosynthesis in *I. indigotica*.

ACKNOWLEDGEMENTS

This research was financially supported by National Natural Science Foundation of China (31100221, 30900786); and Domestic science and technology

cooperation projects (10495801400, 10395820200), Shanghai Science and Technology Committee.

REFERENCES

- Alfermann AW, Fuss E, Bayindir U (2008). Hinokinin biosynthesis in *Linum catharticum* Reichenb. *Plant J. Cell Mol. Boil.* 55:810-820.
- Becker-André M, Schulze-Lefert P, Hahlbrock K (1991). Structural comparison, modes of expression, and putative cis-acting elements of the two 4-coumarate: CoA ligase genes in potato. *J. Boil. Chem.* 266:8551-8559.
- Bin L (2003). Studies on Active Constituents and Quality Evaluation of Banlangen. PhD dissertation, Second Military Med. Univ., Shanghai, China. pp. 36-37.
- Böhmer M, Schroeder JI (2011). Quantitative transcriptomic analysis of abscisic acid-induced and reactive oxygen species-dependent expression changes and proteomic profiling in Arabidopsis suspension cells. *Plant J. Cell Mol. Boil.* 67:105-118.
- Combet C, Blanchet C, Geourjon C (2000). NPS@: network protein sequence analysis. *Trends Biochem. Sci.* 25:147-150.
- Cukovic D, Ehrling J, VanZiffle JA (2001). Structure and evolution of 4-coumarate:coenzyme A ligase (4CL) gene families. *Boil. Chem.* 382:645-654.
- Cutler SR, Rodriguez PL, Finkelstein RR (2010). Abscisic acid: emergence of a core signaling network. *Ann. Rev. Plant Biol.* 61:651-679.
- Dixon RA, Paiva NL (1995). Stress-induced phenylpropanoid metabolism. *Plant Cell* 7:1085.
- Douglas C, Hoffmann H, Schulz W (1987). Structure and elicitor or u.v.-light-stimulated expression of two 4-coumarate: CoA ligase genes in parsley. *EMBO J.* 6:1189-1195.
- Ehrling J, Büttner D, Wang Q (1999). Three 4-coumarate:coenzyme A ligases in Arabidopsis thaliana represent two evolutionarily divergent classes in angiosperms. *Plant J. Cell Mol. Biol.* 19:9-20.
- Graser G, Oldham NJ, Brown PD (2001). The biosynthesis of benzoic acid glucosinolate esters in Arabidopsis thaliana. *Phytochemistry* 57:23-32.
- Greenberg BM, Wilson MI, Huang XD (1997). The effects of ultraviolet-B radiation on higher plants. *Plants Environ. Stud.* 1-35.
- Harding SA, Leshkevich J, Chiang VL (2002). Differential substrate inhibition couples kinetically distinct 4-coumarate:coenzyme A ligases with spatially distinct metabolic roles in quaking aspen. *Plant Physiol.* 128:428-438.
- Ho Y, Chang Y (2002). Studies on the antinociceptive, anti-inflammatory and anti pyretic effects of *Isatis indigotica* root. *Phytomed. Int. J. Phytother. Phytopharmacol.* 9:419-424.
- Hu W, Kawaoka A, Tsa C (1998). Compartmentalized expression of two structurally and functionally distinct 4-coumarate:CoA ligase genes in aspen (*Populus tremuloides*). *Proc. Natl. Acad. Sci. USA* 95:5407-5412.
- Hu WJ, Harding SA, Lung J (1999). Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. *Nat. Biotechnol.* 17:808-812.
- Hu Y, Di P, Chen J (2011). Isolation and characterization of a gene encoding cinnamoyl-CoA reductase from *Isatis indigotica* Fort. *Mol. Biol. Rep.* 38:2075-2083.
- Kajita S, Hishiyama S, Tomimura Y (1997). Structural characterization of modified lignin in transgenic tobacco plants in which the activity of 4-coumarate: coenzyme A ligase is depressed. *Plant Physiol.* 114:871.
- Kiefer F, Arnold K, Künzli M (2009). The SWISS-MODEL Repository and associated resources. *Nucl. Acids Res.* 37:D387-392.
- Kumar A, Ellis BE (2003). 4-coumarate:CoA ligase gene family in Rubusidæus: cDNA structures, evolution, and expression. *Plant Mol. Biol.* 51:327-340.
- Kumar S, Tamura K, Jakobsen IB (2001). MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17:1244-1245.
- Lee D, Douglas CJ (1996). Two divergent members of a tobacco 4-coumarate:coenzyme A ligase (4CL) gene family. cDNA structure, gene inheritance and expression, and properties of recombinant proteins. *Plant Physiol.* 112:193-205.
- Lee D, Meyer K, Chapple C (1997). Antisense suppression of 4-coumarate: coenzyme A ligase activity in Arabidopsis leads to altered lignin subunit composition. *Plant Cell Online* 9:1985-1998.
- Lee K, Xiao Z (2003). Lignans in treatment of cancer and other diseases. *Phytochem. Rev.* 2:341-362.
- Li L, Zhou Y, Cheng X (2003). Combinatorial modification of multiple lignin traits in trees through multigenecotransformation. *Proc. Natl. Acad. Sci.* 100:4939.
- Lindermayr C, Möllers B, Fliegmann J (2002). Divergent members of a soybean (*Glycine max* L.) 4-coumarate: coenzyme A ligase gene family. Primary structures, catalytic properties, and differential expression. *Eur. J. Biochem.* 269:1304-1315.
- Ni H, Shi X, Liu H (2008). Advances of pharmacological effects of triterpenes from *Isatis indigotica*. *J. Qiqihar Univ. Med.*, 29: 1609-1611.
- Pauwels L, Morreel K, De Witte E (2008). Mapping methyl jasmonate-mediated transcriptional reprogramming of metabolism and cell cycle progression in cultured Arabidopsis cells. *Proc. Natl. Acad. Sci.* 105:1380.
- Raghavendra AS, Gonugunta VK, Christmann A (2010). ABA perception and signalling. *Trends Plant Sci.* 15:395-401.
- Ribnicky D, Shulaev V, Raskin I (1998). Intermediates of salicylic acid biosynthesis in tobacco. *Plant Physiol.*, 118: 565-572.
- Saitou N, Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4:406-425.
- Saleem M, Kim J, Ali S (2005). An update on bioactive plant lignans. *Nat. Prod. Rep.* 22:696-716.
- Schmelzer E, Kruger-Lebus S, Hahlbrock K (1989). Temporal and spatial patterns of gene expression around sites of attempted fungal infection in parsley leaves. *Plant Cell Online* 1:993-1001.
- Schneider K, Hövel K, Witzel K (2003). The substrate specificity-determining amino acid code of 4-coumarate: CoA ligase. *Proc. Natl. Acad. Sci. USA* 100:8601-8606.
- Shockey JM, Fulda MS (2003). Arabidopsis contains a large superfamily of acyl-activating enzymes. Phylogenetic and biochemical analysis reveals a new class of acyl-coenzyme A synthetases. *Plant Physiol.* 132:1065-1076.
- Stuible H, Büttner D, Ehrling J (2000). Mutational analysis of 4-coumarate: CoA ligase identifies functionally important amino acids and verifies its close relationship to other adenylate-forming enzymes. *FEBS Lett.* 467:117-122.
- Thompson JD, Higgins DG, Gibson TJ (1994). CLUSTAL W: improving the sensitivity of progressiv. [Nucleic Acids Res. 1994] - PubMed result. *Nucl. Acids Res.* 22:4673-4680.
- Trezzini GF, Horrichs A, Somssich IE (1993). Isolation of putative defense-related genes from Arabidopsis thaliana and expression in fungal elicitor-treated cells. *Plant Mol. Biol.* 21:385-389.
- Voo KS, Whetten RW, O'Malley DM (1995). 4-coumarate:coenzyme A ligase from loblolly pine xylem. Isolation, characterization, and complementary DNA cloning. *Plant Physiol.* 108:85-97.
- Whetten RW, MacKay JJ, Sederoff RR (1998). Recent Advances In Understanding Lignin Biosynthesis. *Ann. Rev. Plant Physiol. Plant Mol. Biol.* 49:585-609.