Isolation, Cloning and Characterization of Cinnamate 4-Hydroxylase (*C4H*) Gene from *Leucaena leucocephala* and its Expression Studies

A THESIS
SUBMITTED TO THE
UNIVERSITY OF PUNE

FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
BIOTECHNOLOGY

BY
SANTOSH KUMAR

UNDER THE GUIDANCE OF
Dr. B. M. KHAN

DIVISION OF PLANT TISSUE CULTURE
NATIONAL CHEMICAL LABORATORY
PUNE – 411 008, INDIA

January, 2012

Dedicated to my Grand Parents

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

First of all, I bow my head and express gratitude with all reverence and devotion to the "Almighty" for showering His blessings. He is the one who has given me a chance to acknowledge whole-heartedly all those who have provided me utmost support all through my thesis research.

I wish to express my heartfelt regards and deep sense of gratitude to my research guide, Dr B M Khan, Scientist-G and HOD, Division of Plant Tissue Culture, NCL for his gracious initiation, inspiring guidance, liberal attitude, valuable suggestions, never ending affection and unstrained encouragement during the course of investigation and preparation of the manuscript.

I would like to specially thank Dr S K Rawal (Former HOD, Division of PTC, NCL) for his critical suggestions and evaluations of my work periodically. His words of wisdom were always helpful for me to provide impetus to my work.

With stupendous ecstasy, I express a deep sense of indebtedness and gratitude to Dr David Nelson, University of Tennessee, USA for classifying and assigning names to LlC4H proteins.

I wish my sincere thanks to all the scientists of Division of PTC and Biochemical Sciences, especially Dr D C Agrawal, Dr U J Mehta and Dr S V Kendurkar and for allowing me to use their lab facility as and when required. Dr D K Kulkarni finds a special mention here for teaching me to take free hand sections. I am highly indebted to Late Dr M I Khan and his family for their support extended to me at the time of need. I also like to take this opportunity to thank all my teachers, who have taught me at various levels right from my childhood.

I am elated with delight to avail this wonderful opportunity to record my gratitude to my lab seniors Dr Sucheta, Dr Sushim, Dr Sameer, Dr Noor, Dr Arun, Dr Abhilash and Dr (Mrs) Pallavi for their cooperation and help. I am highly indebted to my lab colleagues Prashant, Rishi, Krunal, Somesh, Parth, Neha, Kannan, Poonam, Uma, Ruby, Santosh Jadhav and Shakeel for their support and cordial atmosphere in lab. My special words of thanks should also go to my trainees Shruti and Gayatri for their assistance during critical phase of my work. I would also like to thank all the trainees in the lab, who have worked or are working in the lab, for their politeness and respect.

Some persons come into one's life and again they make a bye. But, its never bye for the nice bond of friends Ramesh, Prabhash, Ravi, Nishant, Asad, Shadab, Srikant, Prasad, Sajid, Avinash, Ansary, Shabab, Dr Atul, Dr Feroz, Dr Anil, Dr Nagraj, Dr JP, Dr Shashi, Sathish, Vaibhav, Shiv bhaiya, Trupti, Urvashi, Gyan, Yashwant, Raju, Dr Prasad, Dr Bhuvan, Poonam and Rita whose encouragement and best wishes provided a homely environment during my stay at NCL. I also wish to thank all my school level, B.Sc. and M.Sc. friends, most of them although not with me in Pune, were always with me with their well wishes.

I express my appreciation to all the laboratory staff members of NCL in general and Mr Surya Prasad and Mr Jagtap in particular, who took immense pain in maintaining the electrical equipments and managing the green house facility, respectively.

The financial assistance provided by Council of Scientific and Industrial Research, Government of India is gratefully acknowledged. I also thank Dr Saurav Pal, Director, NCL and Dr S Sivaram (Former Director, NCL) for giving me a chance to carry out my work at NCL and letting me use the facilities of NCL.

It would be like a drop in the ocean of words that can never reach their mark for my lab-mate and wife, Sumita who gave me infinite love and courage to go for this achievement. I am sure we will always be each other's strength through the ups and downs of life yet to come. I am equally grateful to my mother and father-in-law for their blessings and support in several ways. I cannot forget to mention my only *saali saahiba*, Namita, who despite her business, used to make our weekends nice by visiting us.

Words are small trophies to express my deep sense of indebtedness and affection to my parents, sister and brothers, my family members and relatives for their blessings, encouragement and inspiration, which helped me in building my educational career.

Last but not the least, thanks to all beloved and respected people around me who directly and indirectly helped me during my PhD program and whose names could not be mentioned here.

December, 2011 Santosh Kumar

DECLARATION

I hereby declare that the thesis entitled **"Isolation, Cloning and Characterization of Cinnamate 4-Hydroxylase (*C4H*) Gene from *Leucaena leucocephala* and its Expression Studies"**, submitted for the Degree of **Doctor of Philosophy** to the University of Pune, has been carried out by me at the Division of Plant Tissue Culture, National Chemical Laboratory, Pune - 411 008, Maharashtra, India, under the supervision of **Dr. B. M. Khan** (Research Guide). The work is original and has not been submitted in part or full by me for any other degree or diploma to any other University.

December, 2011

(Santosh Kumar)
Research Scholar
Division of PTC, NCL
Pune – 411 008

CERTIFICATE

This is to certify that the work incorporated in the thesis entitled **"Isolation, Cloning and Characterization of Cinnamate 4-Hydroxylase (*C4H*) Gene from *Leucaena leucocephala* and its Expression Studies"** submitted by **Santosh Kumar** was carried out under my supervision at the Plant Tissue Culture Division, National Chemical Laboratory, Pune. Materials obtained from other sources have been duly acknowledged in the thesis.

Dr. B. M. Khan
Research Guide

Abstract

Leucaena leucocephala is one of the most versatile, fast growing commercially important trees for paper and pulp industry in India, contributing for about 1/4th of the total raw material supplied to the industry. *L. leucocephala* is traditionally high in lignin content which must be removed from pulp involving various chemical and energy intensive processes, which in turn release various toxic pollutants and damage the polysaccharide components of wood. There is currently intense interest to reduce, modify or alter lignin content of *L. leucocephala* to be used as model plant in paper industry. Phenylpropanoid pathway is operative in plants and responsible for the production of a variety of compounds including lignin. Cinnamate 4-Hydroxylase (C4H) catalyzes the second step of phenylpropanoid pathway and hydroxylates *trans*-cinnamic acid at the *para* position. C4H plays central role in lignification as it catalyzes slow reaction early in the pathway, which branches after the action of C4H.

To have a better understanding of the role of C4H in *L. leucocephala*, three *C4H* isoforms (*LlC4H*) were isolated from *L. leucocephala* and characterized. The isoforms share more than 98% sequence identity at amino acid level to each other. These isoforms contained all the conserved sequences found in P450 superfamily and C4H from other plants. Phylogenetic analysis grouped LlC4H with class I C4Hs, some members of which are functionally well characterized. Southern hybridization study suggested that there may be 4-6 *C4H* isoforms in *Leucaena* genome with some isoforms may be tandemely arranged.

Three isoforms of C4H were expressed in *Escherichia coli*, strain Rosetta (DE3) and one isoform was purified from inclusion body. LlC4H proteins are highly hydrophobic and hence they profusely formed inclusion body in bacteria. C4H bound very weakly to the Ni-NTA matrix and was difficult to obtain in pure form from inclusion body. Under the optimized conditions, yield of purified C4H from inclusion body varied between 50-100 µg/mL. Secondary and

tertiary structure comparison between the LlC4H isoforms suggested that all the isoforms are identical in their secondary and tertiary structure.

Tissue and age specific Quantitative Real Time (Q-RT) PCR study suggested that *C4H* transcripts were highly abundant in root tissues followed by stem and leaves. Maximum transcript level at any time in any tissue was observed in 30 day old root tissue. Among the tissues investigated, demonstrable C4H activity was found maximum in 1 year old stem tissues. *C4H* was found to be stress responsive and its transcript level increased in response to external abiotic stimuli (salt stress, methyl jasmonate stress and UV-C stress). Tissue-wise quantitative comparison of lignin from developing seedling stage to one year old tree stage indicated that while acid insoluble lignin increased with age, acid soluble lignin first decreased and then slightly increased.

Transgenic *L. leucocephala* and tobacco plants with reduced C4H expression were raised using modified pCAMBIA1301 vector harbouring partial *L. leucocephala C4H* in antisense orientation. Careful study of transcript level of different phenylpropanoid pathway genes in transgenics, down-regulated for C4H activity showed coordinated down-regulation and gave indication that C4H might be catalyzing rate limiting step early in the pathway. Transgenic plants had reduced lignin content consistent with its role in lignification.

ABBREVIATIONS

Cinnamate 4-Hydroxylase	(C4H)
Cytochrome P450	(P450)
NADPH: Cytochrome P450 Reductase	(CPR)
4-Coumarate coenzyme A ligase	(4CL)
Abridged Anchor Primer	(AAP)
6- Benzylaminopurine	(BAP)
Coumarate 3- hydroxylase	(C3H)
Cinnamyl alcohol dehydrogenase	(CAD)
Caffeoyl coenzyme A 3-O- methyltransferase	(CCoAOMT)
Cinnamoyl coenzyme A reductase	(CCR)
Caffeate O-methyltransferase	(COMT)
Ferulate 5-hydroxylase/ coniferaldehyde 5-hydroxylase	(F5H/CAld5H)
Guaiacyl	(G)
1-Napthyl acetic acid	(NAA)
Overnight	(O/N)
Phenylalanine ammonia lyase	(PAL)
Rapid amplification of cDNA ends	(RACE)
Room temperature	(RT)
Syringyl	(S)
Sinapyl alcohol dehydrogenase	(SAD)
Thidiazuron	(TDZ)
Untranslated Region	(UTR)
(2-isopentenyl)adenine	(2ip)
∆-Amino levulinic acid	(ALA)

CHAPTER - 1

General Introduction

Chapter 1: General Introduction

Indian paper industry holds a share of 2% of the world paper production and is continuously expanding (FAOSTAT 2011). About 660 manufacturing units in India are currently engaged in manufacturing paper, paperboards, newsprints and paper products like tissue paper, paper bag, filter paper, tea bag, medical grade coated paper and light weight online coated paper etc. To keep the Indian paper industry rolling, availability of superior quality raw material is must. Paper industry is primarily dependent upon forest-based raw materials with Indian paper industry currently relying on four major species viz. Bamboo, Eucalyptus, Casuarina, and *L. leucocephala*. Among these plants, *L. leucocephala* accounts for about 25% of the raw material (Srivastava et al. 2011). The wood pulp, that is processed to form paper, consists of three major components, cellulose, lignin and hemicellulose. Cellulose is the fibrous component of wood used to make paper, while lignin is responsible for sticking of the cellulose microfibrils together. As a major cell wall polymer, lignin imparts rigidity to the plant cell and renders the cell wall impermeable to water. This impermeability holds key role in water conduction through the xylem vessles to withstand the negative pressure, thought to be generated by transpiration and water conduction along the vascular bundle (Steudle 2001). In addition, lignin acts as a barrier to plant pathogens and prevents lodging, a problem in many agronomically important plants. Other than its critical role in plants' sustenance, it is seen as a major hindrance for the agro-industrial exploitation of various plant species such as source of pulp for paper industry; and fodder for ruminants. The presence of lignin adversely affects paper quality and is responsible for poor performance characteristics and brightness of paper associated with yellowing of paper with age (Rastogi and Dwivedi 2006). To produce high quality paper, lignin needs to be removed from the pulp involving various chemical and energy intensive processes which in turn release various toxic pollutants and damage the polysaccharide components of wood (Rastogi and Dwivedi 2006). There is currently intense interest to reduce, modify or alter lignin content of *L. leucocephala* to be used as model plant in paper industry or as easily digestible and palatable fodder for animals. Therefore, various biotechnological programs have been undertaken to raise transgenics down-

regulated for key lignin biosynthetic gene(s) in order to tackle this problem. In the current work, the author has undertaken one such step by isolating a lignin biosynthetic pathway gene, Cinnamate 4-Hydroxylase (*C4H*) from *L. leucocephala* and characterizing it at molecular level to understand its role in lignification in *L. leucocephala.* Downregulation of *C4H* expression has also been studied in transgenic tobacco and *L. leucocephala* plants and the transgenics have been analysed for parameters like lignin content and the effect of reduction of C4H activity on the transcript abundance of *C4H* and other phenylpropanoid pathway genes.

1.1 *Leucaena leucocephala*

We need to have high cellulosic pulp yielding tree acclimatized to grow in a range of climatic conditions to serve as a source of pulp for paper industry in India. *L. leucocephala* is one such multiple purpose and one of the most versatile and productive tree legume commonly found in India (Fig. 1.1). This tree species was once known as the 'miracle tree' because of its worldwide success of being a perennial and highly nutritious forage tree and a variety of other uses (Shelton and Brewbaker 1994). Besides forage, *L. leucocephala* has been used as a potential source of pulp, firewood, timber, mulch, organic fertilizer, gum, shade and as cover to control soil erosion (Shelton and Brewbaker 1994; Vietmeyer and Cottom 1977). *L. leucocephala* pulp is of high quality and is suitable for use in printing and writing papers owing to its high opacity, good printability and good formation. It has been estimated that the dried leaves of *L. leucocephala* contain 27-34% protein of high nutritional quality (Rastogi and Dwivedi 2006) making it very healthy source of fodder. A large volume of medium-light hardwood for fuel (specific gravity 0.5 – 0.75) with low moisture and high calorific value that produces excellent quality charcoal, low smoke and ash can be produced from this species (Shelton and Brewbaker 1994). Being a leguminous tree, it is also associated with atmospheric nitrogen fixation and has been used in alley cropping in Philippines and Indonesia (Shelton and Brewbaker 1994; Vietmeyer and Cottom 1977). The *Leucaena-Rhizobium* partnership can fix more than 500 kg of nitrogen per hectare of land (Vietmeyer and Cottom 1977). The major limitation in using *L. Leucocephala* as a source of pulp or forage is its extent

of lignification. Despite being a good quality source of pulp, *L. leucocephala* is traditionally high in lignin content (Dutt et al. 2007) and this problem needs to be addressed by raising transgenics with low lignin content.

Fig. 1.1: *Leucaena leucocephala* (A) flower; (B) Young pods; (C) Mature pods; and (D) seeds

1.1.1 Botanical description of *Leucaena leucocephala*

L. leucocephala, previously known as *L. glauca*, is a perennial tree or shrub belonging to fabaceae family, that grows up to heights of 7-18 m (Shelton and Brewbaker 1994). Other than *L. leucocephala*, there are at least 14 other species recognized in the genus. Majority of these species are diploid, but *L. leucocephala* is a self-pollinating tetraploid (2n=4x=104) species (Shelton and Brewbaker 1994). *L. leucocephala* is the most common species of *Leucaena* genus and hence is sometimes referred to as '*Leucaena'* only (Vietmeyer and

Cottom 1977). The author has followed this convention and hence in this thesis, '*Leucaena*' written anywhere in the text refers to *Leucaena leucocephala* only and not to other species of the genus *Leucaena*.

1.1.2 Botanical classification of *Leucaena leucocephala*

Kingdom: Viridiplantae

Division: Streptophyta

Class: Magnoliopsida

Order: Fabales

Family: Fabaceae

Genus: *Leucaena*

Species: *leucocephala*

1.2 Pulp and paper industry statistics

The world's paper and paperboard production decreased from 392 million tonnes in 2008 to 377 million tonnes in 2009 (FAOSTAT 2011) whereas the demand for paper is growing day by day. In India, annually, 6.8 million tonnes (mt) of wood is consumed to produce 1.9 mt of pulp, where nearly 80% of wood is procured from agro-forestry sources and rest 20% from government sources (Kulkarni 2008). To maintain the supply of raw materials to the industry, the industry relies on input from social and farm forestry. *Leucaena* accounts for approximately $1/4^{th}$ of the raw material supplied to the industry and thus plays crucial role to maintain the demand-supply chain. Quality wise, pulp generated from *Leucaena* is rich in holocellulose and low in ash, silica, hot water solubles and alcohol-benzene solubles, which is comparable to other fast growing hardwoods used in paper industry (Dutt et al. 2007). *Leucaena* wood contains about 60% holocellulose and the pulp yield varies between 48.5 – 51.5% depending upon the age of the tree (Dutt et al. 2007).

India, despite featuring among the top 15 paper producing countries of the world, the per capita consumption of paper in India is among the lowest and is poised at about 4.5 kg per year, far below the South Asian and global average of 10.89 and 53.07 kg per year, respectively. But the demand for paper is expected to increase particularly because of the upsurge of the

disposable income of the expanding middle income group. So far, the growth of paper industry in India has matched the growth of GDP and on an average stands between 6-7% over the last few years. India is the fastest growing market for paper globally and the paper production is estimated to touch 13.95 mt by 2015-16 (http://www.ipma.co.in/paper_industry_overview.asp). As per the estimates from paper industry, paper production is expected to grow at a Compounded Annual Growth Rate (CAGR) of 8.4% while paper consumption will grow at a CAGR of 9% till 2012-13. The import of pulp & paper products is hence, likely to show a growing trend.

1.3 The pulping process (Pulp production)

Pulping is the process to reduce wood to a fibrous mat by separating the wood component cellulose from lignin and hemicellulose. Pulping can be divided broadly into three steps.

Step 1: Initial processing

Initial processing can be done either through (a) chemical, (b) mechanical or (c) semi-mechanical (combination of both).

a) Chemical method involves cooking wood in a digester at elevated pressure with a solution of chemical that dissolves lignin and leave behind cellulose. This process is hazardous releasing air pollutants including formaldehyde, methanol, acetaldehyde and methyl ethyl ketones.

b) Mechanical method physically separates the fibres by pressing the wood against a grinder. This is energy demanding and produces weak and opaque product that discolours easily after exposure to light.

c) Semi-mechanical pulping involves both chemical and mechanical pulping methods. Wood chips are first partially cooked with chemicals and then is subjected to mechanical pulping.

Step 2: Washing the pulp

After pulping, dissolved lignin and chemicals are removed from pulp by passing pulp through a series of washers and screens at high temperature, again generating hazardous exhausts and air pollutants. The chemical recovery process to recover used chemicals from the effluents also causes severe air pollution.

Step 3: Bleaching the pulp

After washing, pulp is bleached to remove colour associated with remaining residual lignin in the pulp. This is achieved by using either elemental chlorine and hypochlorite; or oxygen and peroxide and washing the pulp with several volumes of water releasing polluted waste water stream and vents releasing air pollutants. Depending upon the bleaching method used, the effluents may contain chlorinated organic compounds and toxic chemicals like dioxins, furans and chlorinated organic compounds and vents discard chloroform, methanol, formaldehyde and methyl ethyl ketone. Many a times, these chemicals pass untreated, through the treatment plants and contaminate water reservoirs like river and ocean.

1.4 Lignin and its biosynthesis

Eclipsed only by cellulose, lignin is the second most abundant biopolymer on earth and represents approximately 30% of the organic carbon (Boerjan et al. 2003). It is a complex racemic aromatic heteropolymer composed of three 4-hydroxycinnamoyl alcohol units differing in their degree of methoxylation in different combinations (Vanholme et al. 2010; Boerjan et al. 2003; Vogt 2010). The hydroxycinnamoyl alcohol units of lignin are termed monolignols (coniferyl, sinapyl and *p*-coumaryl alcohols) and they constitute the main building block of lignin (Fig. 1.2). These monolignols, when incorporated into the lignin polymer are called guaiacyl (G), syringyl (S) and p-hydroxyphenyl (H) units (Boerjan et al. 2003; Vanholme et al. 2010; Goujon et al. 2003). The amount and composition of lignin vary among different taxa, cell type, individual cell wall layers and is dependent on environmental and developmental factors (Achyuthan et al. 2010; Boerjan et al. 2003). Hardwood (dicotyledonous angiosperm) lignin primarily consists of G and S units with only traces of H units whereas softwood (gymnosperm) lignin consists mostly of G with low level of H units (Achyuthan et al. 2010; Boerjan et al. 2003; Vanholme et al. 2010). G monolignol units are methylated on 3-hydroxy position of the aromatic ring and are derived from caffeic acid (and/or its related aldehye and alcohol), whereas S monolignol units are methylated on both 3-hydroxy and 5-hydroxy positions and are derived from sinapic acid (and/or its related aldehye and alcohol) (see Fig. 1.2).

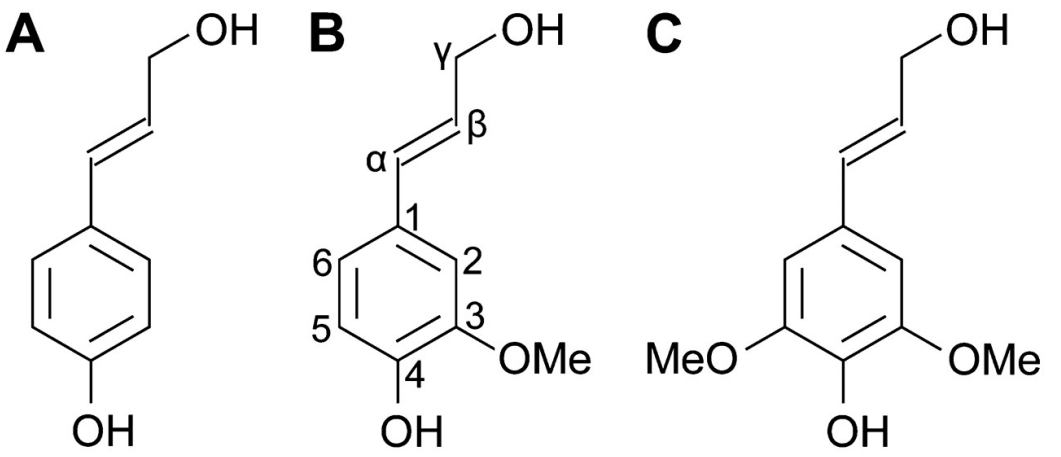


Fig. 1.2: Chemical structures of three monolignols. (A) H-monomer (p-coumaryl alcohol); (B) G-monomer (coniferyl alcohol); (C) S-monomer (sinapyl alcohol) (source: van Parijs et al. (2010) *Plant Physiology* 153, 1332-1344)

The relative abundance of S and G units in wood plays important role in paper and pulp industry with higher S units favourable than G units. S units in lignin are linked via relatively labile ether bonds which are more easily hydrolyzed than the linkage between G units. On the other hand, majority of G units are linked via more stable (and stubborn) biphenyl and other carbon-carbon linkages difficult to hydrolyze (Eckardt 2002). As a result, higher G content in wood requires more expensive and hazardous chemicals as well as vigorous treatment methods. High variation in the S/G ratio has been observed in between species, within a single species, within individual plant and cell type as a result of genetic, developmental and environmental cues (Eckardt 2002). Even a fractional increase in S/G ratio associated with overall decrease in lignin content has the potential to significantly improve pulping process and bring down the use of environmentally hazardous chemicals to a large extent.

Lignin deposition mainly takes place during secondary thickening of the cell wall in the final stage of xylem cell differentiation (Boerjan et al. 2003). The secondary cell wall consists of three layers: the outer (S1), middle (S2) and the inner (S3) (Fig. 1.3). Lignin deposition takes place in different phases in these layers each preceded by deposition of carbohydrates starting at cell corner of middle lamella and the primary wall when S1 formation has started (Boerjan et al. 2003). (For detailed phenylpropanoid pathway, see Fig. 1.4)

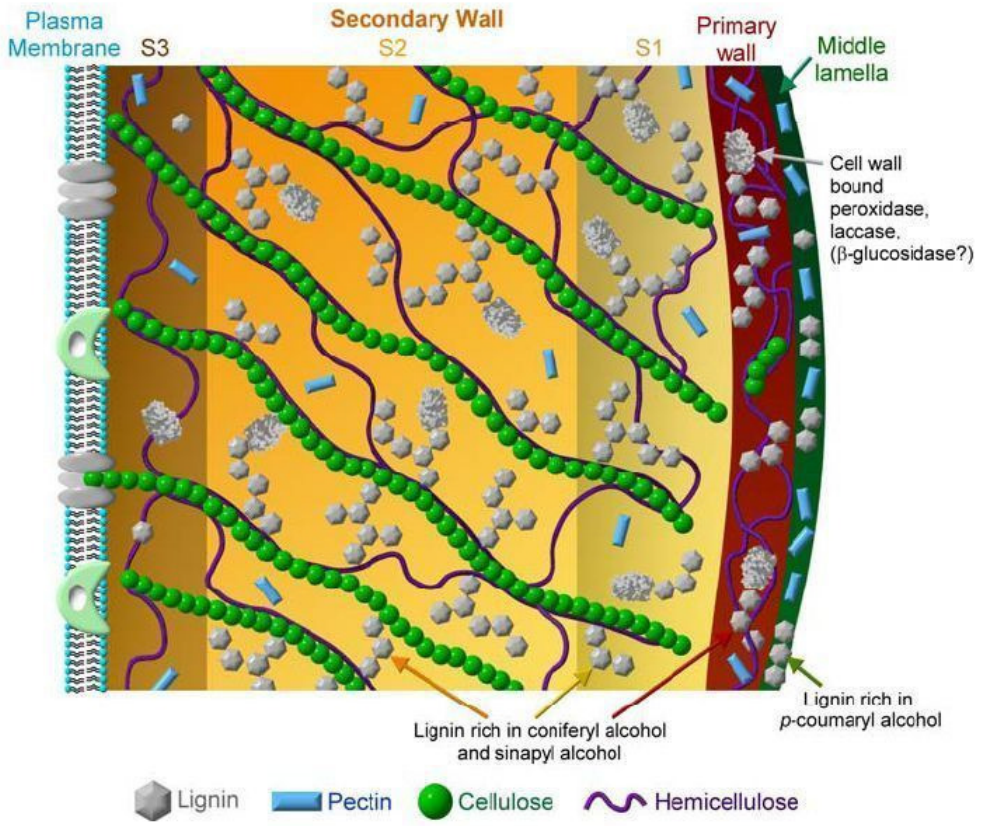


Fig. 1.3: Lignin, pectin, cellulose and hemicellulose and their occurrence within the boundary of plasma membrane and middle lamella (source: Achyuthan et al. (2010) *Molecules* 15, 8641-8688

1.4.1 General phenylpropanoid pathway enzymes

Lignin biosynthesis in plants begins with a core group of three reactions termed as general phenylpropanoid pathway catalyzed by Phenylalanine ammonia lyase (PAL), Cinnamate 4-Hydroxylase (C4H) and 4-Coumarate: CoA Ligase (4CL) respectively (Boerjan et al. 2003; Vogt 2010; Hahlbrock and Scheel 1989). The first step towards formation of monolignols commences with the deamination of phenylalanine catalyzed by PAL. Subsequent reactions involve hydroxylation of the aromatic ring followed by phenolic O-methylation and sequential reduction of the side chain carboxyl group to an alcohol group. PAL, being the first enzyme of phenylpropanoid pathway, plays the role of a connecting link between primary and secondary metabolism by directing the carbon flow from Shikimate pathway to phenylpropanoid pathway (Boerjan et al. 2003; Boudet 2007; Hahlbrock and

Scheel 1989; Vanholme et al. 2010; Vogt 2010). The pathway branches after the action of C4H (second step of the pathway) and leads to a variety of compounds including lignin, hydroxycinnamic acid conjugates and flavonoides. Many intermediates and end products of this pathway play important role in plants as phytoalexins, antioxidants, UV ray protectants, pigments, aromatic compounds and antiherbivory compounds (Boudet 2007; Naoumkina et al. 2010). The last step of general phenylpropanoid pathway is catalyzed by 4CL, which commonly exists in plants as a small gene family with distinct roles (Boudet 2007; Vogt 2010). A particular 4CL isoform may direct the flux towards lignin biosynthesis while the other isoform towards other phenolics such as flavonoides with different level of activity in different tissues (Boudet 2007).

1.4.2 Monolignol biosynthesis committed enzymes

Cinnamoyl-CoA-reductase (CCR) and cinnamyl / sinapyl alcohol dehydrogenase (CAD / SAD) belong to the group of monolignol biosynthesis committed enzymes (Boudet 2007; Boudet 2000; Grima-pettenati and Goffner 1999; Vanholme et al. 2008). CCR as the first enzyme committed to monolignol biosynthesis channels phenylpropanoid metabolites into lignin biosynthesis. There may be more than one CCR gene per haploid genome, but in that case only one CCR isoform appears to be regulating lignin and other(s) seem to be related to the production of other phenolic compounds whose accumulation may lead to resistance (Boudet 2000). Another monolignol specific multifunctional enzyme, CAD, catalyzes NADPH dependent reduction of cinnamyl aldehydes to the corresponding cinnamyl alcohol, the last step of monolignol biosynthesis before oxidative polymerization in the cell wall (Boudet 2000; Grima-pettenati and Goffner 1999; Sibout et al. 2005). A CAD homolog from aspen was reported with very high specificity towards reducing sinapaldehyde than the usual CAD substrate and hence was named SAD (Boerjan et al. 2003; Boudet 2007).

1.4.3 Monolilgnol modification specific phenylpropanoid enzymes

Enzymes of this group control lignin monomeric composition and comprise of methylating enzymes, i.e. caffeic acid-*O*-methyltransferase (COMT), caffeoyl-CoA *O*-methytransferase (CCoAOMT); hydroxylating enzymes coumarate-3-

hydroxylase (C3H) and ferulate-5-hydroxylase (F5H) and a newly discovered enzyme HCT (p-hydroxycinnamoyl-CoA:D-quinate; CQT or p-hydroxy-cinnamoyl-CoA:shikimate p-hydroxycinnamoyltransferase; CST, collectively called HCT) (Boerjan et al. 2003; Boudet 2000; Grima-pettenati and Goffner 1999). These enzymes are likely candidates in regulating the types of monolignols incorporated into polymeric lignin. C3H was long thought to hydroxylate coumaric acid at C_3 position of the aromatic ring to form caffeic acid, but later it was found that the preferred substrates of C3H are shikimate and quinate esters of coumaric acid and not the conventionally thought *p*-coumaric acid, *p*-coumaroyl-CoA, p-coumaraldehyde, p-coumaryl alcohol, nor the 1–*O*-glucose ester and the 4–O-glucoside of p-coumaric acid (Boerjan et al. 2003; Vanholme et al. 2010). F5H is thought to hydroxylate ferulic acid, coniferaldehyde, coniferyl alcohol and possibly feruloyl-CoA at C_5 of the aromatic ring to form 5-hydroxyferulic acid, 5-hydroxyconiferaldehyde, 5-hydroxyconiferyl alcohol and 5-hydroxyferuloyl-CoA respectively. Whereas, HCT converts *p*-coumaroyl-CoA to *p*-coumaroyl shikimic acid / quinic acid (both are substrates of C3H; C3H converts them to caffeoyl shikimic acid / quinic acid) and caffeoyl shikimic acid / quinic acid to caffeoyl-CoA.

The classic historical view of lignin biosynthesis suggested that methylation by COMT occurred only at free acid level; COMT methylating caffeic acid on 3-OH position of the aromatic ring to form ferulic acid (and after the action of F5H on ferulic acid to form 5-hydroxy ferulic acid) and methylating 5-hydroxy ferulic acid at 5-OH position to form sinapic acid. This view was challenged after the finding that COMT is capable of methylating at aldehyde and alcohol levels as well (Boerjan et al. 2003; Boudet 2000; Eckardt 2002). This was further strengthened after the discovery of another class of *O*-methyltransferase, CCoAOMT. It was successfully proved and is now established that CCoAOMT can methylate CoA esters of 3-hydroxycaffeoyl CoA as well as 5-hydroxyferuloyl CoA (Boudet 2000; Eckardt 2002). The current information on lignin biosynthetic pathway suggests that the pathway can best be described as a metabolic grid with modification of aromatic ring possible at any or all of these levels.

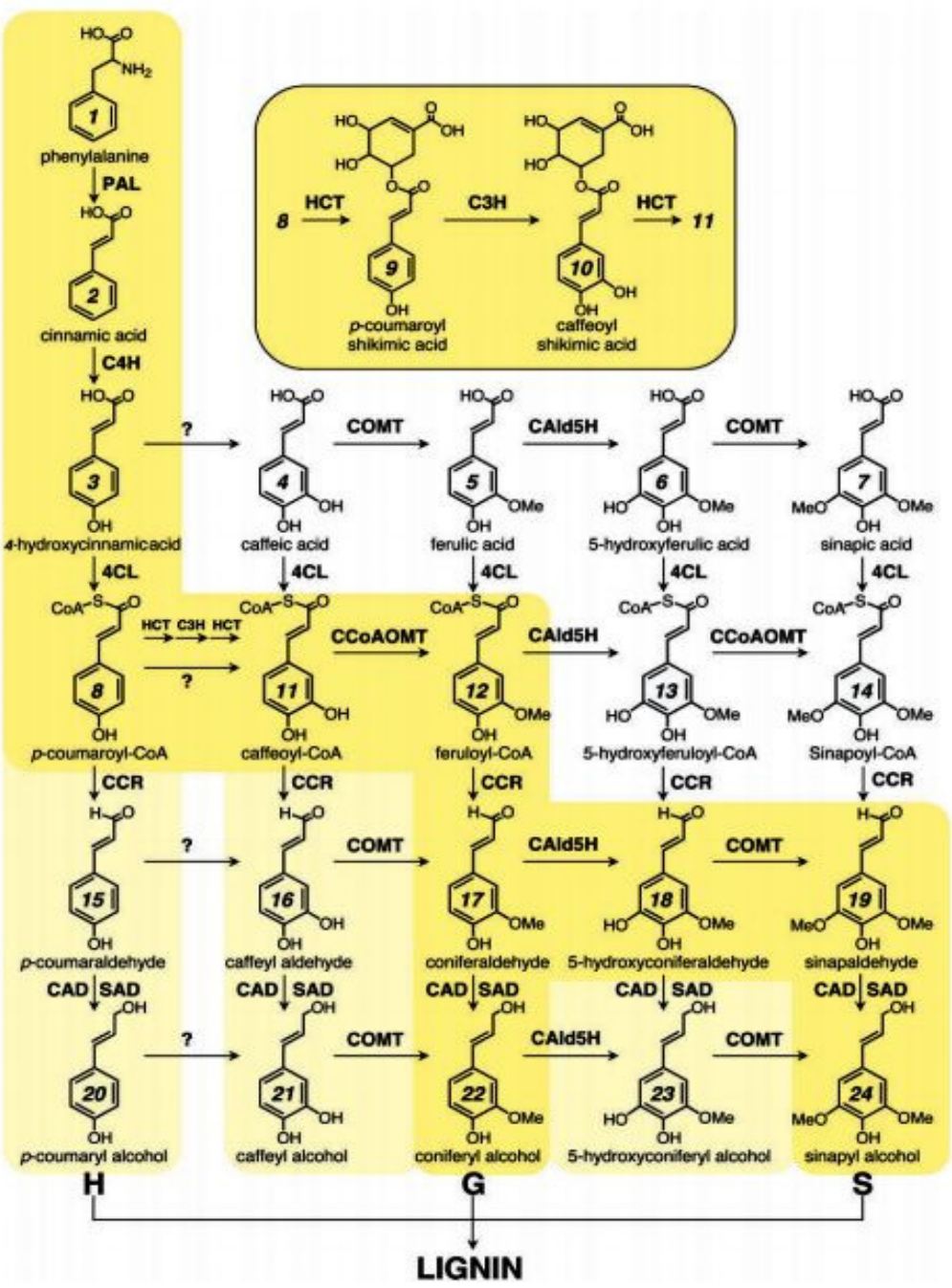


Fig. 1.4: Lignin biosynthetic pathway based on current knowledge (as adapted from Shi et al. (2010) *Plant & Cell Physiology* 51, (1) 144-163)

1.4.4 Monolignol transport, coupling and polymerization

The monolignols are relatively unstable and toxic to the cell that do not get accummulated in the cytosol but are transported to the cell wall where they are polymerized. Several hypotheses have been put forward to explain the transport mechanism of monolignols to the cell wall but none of them is widely acceptable. The involvement of phenolic glucosides, coniferin and syringin, was proposed to explain the transport of monolignol but it could not be proved (Vanholme et al. 2008). Transport through Golgi body derived vesicles or through exporter proteins to plamsa membrene also did not succeed (Vanholme et al. 2010). The picture is still not very clear about the transport of monolignols to the plasma membrane (Boerjan et al. 2003; Vanholme et al. 2010; Vanholme et al. 2008).

After transport of monolignols to the cell wall, they are polymerized through dehydrogenative (oxidative) polymerization which involves oxidative radicalization of phenols followed by combinatorial radical coupling. The dehydrogenation of monolignols has been ascribed to different classes of proteins like, peroxidases, laccases, polyphenol oxidases, and coniferyl alcohol oxidases, but the involvement of any particular protein falling in these categories is very limited because these enzymes belong to large gene families (Vanholme et al. 2010). After dehydrogenation, two monomer radicles may couple to form a dehydro dimer by forming covalent bond between both subunits at their *β* position, resulting in only *β*-*β*, *β*-*O*-4 and *β*-5 dimers (Vanholme et al. 2010). Since the coupling is in chemical combinatorial fashion, any phenol present in zone of cell wall may enter into the coupling process to the extent as allowed. Thus, polymerization of monolignols follows simple chemical rules and largely depend upon the chemical and physical environment of the site of synthesis as pH, ionic strength, temperature, supply and concentration of individual monolignols, physical proximity of the subunits, hydrogen peroxide and peroxidase concentrations, and the matrix in general (Boerjan et al. 2003; Vanholme et al. 2010; Vanholme et al. 2008). The formed dimer in this fashion again gets dehydrogenated to a phenolic radical and followed by coupling of another monomer radicle. This mode of action of growing the lignin polymer chain is called endwise coupling.

1.5 Cinnamate 4-Hydroxylase (C4H) as target gene

C4H catalyzes the second step of phenylpropanoid pathway and hydroxylates trans-cinnamic acid at the *para* position, which is the reaction product of first enzyme of phenylpropanoid pathway, PAL (Russell 1971; Hahlbrock and Scheel 1989). *C4H* is a member of the large group of cytochrome P450 monooxygenases (P450) from plants and exclusively constitute the *CYP73* family, a typical group A P450 (Chapple 1998). Russell and Conn (1967) provided first evidence of C4H activity from the crude extract of pea seedlings whereas C4H was first purified in active form from *Jerusalem artichoke* (Gabriac et al. 1991). C4H is one of the most abundant P450 in plants, nevertheless it was a difficult target to purify in active form because of its low concentration, hydrophobicity, instability and membrane bound nature (Chapple 1998; Russell 1971). It took more than 20 years after demonstrating its activity from crude extract, to bring to purity from *J. artichoke* (Gabriac et al. 1991). Following C4H purification from *J. artichoke*, with the aid of peptide sequencing and antibodies raised against the purified C4H, genes corresponding to *C4H* were isolated from alfalfa, mung bean, *J. artichoke*, *Catharathus rosues*, hybrid aspen, pea, *Arabidopsis* (Fahrendorf and Dixon 1993; Frank et al. 1996; Hotze et al. 1995; Kawai et al. 1996; Mizutani et al. 1997; Mizutani et al. 1993a; Teutsch et al. 1993) and the number is still growing. C4H isolated from these plants had high level of identity ranging from 75-80% at nucleotide level to more than 88-90% identity at amino acid level. C4H sequences from plants share many conserved regions and form a cluster of closely related sequences highly similar to each other. Other than this, C4H possess all the characteristics of a typical P450 protein including the light reversible inhibition of enzymatic activity by carbon monoxide (Chapple 1998). C4H exists as multi-family in trees like *Populus tremuloides*, *P. trichocarpa* (Lu et al. 2006), periwinkle (Hotze et al. 1995) and members of fabaceae family plants like pea (Whitbred and Schuler 2000) and alfalfa (Fahrendorf and Dixon 1993). Despite the presence of multiple isoforms of *C4H*, the precise roles of individual isoforms are not known in every case. Promoter region of C4H genes contain putative cis-acting elements known to occur in other phenylpropanoid pathway genes as well supporting the view that phenylpropanoid genes are subjected to coordinated transcriptional

regulation. Expression pattern of C4H is consistent with the role of this enzyme in lignification. Heterologous expression of C4H in yeast system efficiently converts *trans*-cinnamic acid to *para*-coumaric acid.

C4H is cotranslationally inserted to the endoplasmic reticulum (ER) in a signal recognition particle (SRP) dependent fashion through its helical N-terminal hydrophobic domain and rest of the protein including the heme binding domain faces the cytosol (Chapple 1998). The N terminal transmembrane hydrophobic helix is immediately followed by a proline rich region supposed to destabilize the α helix and produce kink to optimally orient the rest of the sequence with respect to the plasma membrane. C4H also contains heme binding motif near the C- terminus (the conserved sequence is PFGVGRRSCPG) in addition to conserved helices like I helix, K helix, and K' helix (Durst and Nelson 1995).

C4H catalyzes slow reaction with high substrate specificity (Ehlting et al. 2006) and is located at strategic important point as the pathway branches after C4H. *Pinus taeda* cell suspension culture when fed with saturating level of phenylalanine did not result in proportional increase in *C4H* and *C3H* transcript level as with other phenylpropanoid genes like PAL, 4CL, COMT, CCR and CAD indicating both the P450 genes indeed are rate limiting step in the phenylpropanoid pathway (Anterola et al. 2002). Also, the reaction catalyzed by C4H is exothermic and irreversible (i.e. represent a point of no return) (Ehlting et al. 2006); hence C4H plays a vital role at the very beginning of the phenylpropanoid pathway.

It has been proposed that C4H anchors in the ER membrane and other phenylpropanoid enzymes form multi-enzyme complex by interacting and associating with C4H, although metabolic channelling has only been successfully demonstrated between a specific form of PAL and C4H. This metabolic channelling helps in keeping the local concentration of PAL generated cinnamic acid pool high and does not let it mix with the cytosolic contents (Achnine et al. 2004; Rasmussen et al. 1999). The physical association of membrane anchored C4H, PAL and possibly later

phenylpropanoid pathway enzymes provide indication of effective regulation of phenylpropanoid pathway at the entry level.

In agreement with early position of C4H in the pathway, mutants for C4H exhibit reduced level of phenylpropanoid pathway end products, reduced lignin with altered monomer content with concomitant accumulation of a novel hydroxycinnamic ester, cinnamoymalate not found in wild type plants. Other phenotypic changes observed were dwarfism, male sterility and swelling and branch junctions (Schilmiller et al. 2009). Considering this, several workers generated transgenic plants down-regulated for *C4H* in order to reduce lignin content in plants to be used as model plants in paper industry (Sewalt et al. 1997) or to be used as forage crops with increased palatability. Sewalt et al. (1997) showed that C4H downregulation in transgenic tobacco reduced C4H activity to up to 20% of wild type and the lines expressing less than 50% C4H activity had significantly reduced lignin level. In another study, Blount et al. (2000) found that transgenic tobacco lines over-expressed for PAL but down-regulated for C4H had reduced PAL activity. This emphasizes the dominant effect of C4H downregulation over PAL over-expression. Rastogi and Dwivedi (2006) generated transgenic *L. leucocephala* with decreased *O*-methyl transferase (*OMT*) and showed that reduction of 55-60% OMT activity in transformants had 13-28% lesser lignin content. Extrapolating this to down-regulation of C4H in *L. leucocephala* may yield transgenics with significantly reduced lignin level and higher extractability of cellulose from the plants. Keeping this in mind, the present work has been undertaken.

CHAPTER - 2

Materials and Methods

Chapter 2: Materials and Methods

This chapter describes all the general consumables and non-consumable materials used during the course of this work and their source. It also deals with general laboratory protocols and techniques used during experiments. Other specific methodologies have been discussed in the respective chapters.

2.1 Materials/Equipments used in the study

2.1.1 Glasswares and their preparation

Glasswares used in all the experiments were procured from "Borosil", India. Test tubes (25 mm x 150 mm), glass bottles (70 mm x 125 mm), petridishes (85 mm x 15 mm; 55 mm x 15 mm), conical flasks (100, 250 & 500 mL; 1, 2 & 5 L capacity) and pipettes (1, 2, 5, 10 and 25 mL capacity) were used during the course of study. Glasswares were washed and dried using standard procedures. Glasswares were autoclaved when necessary at 121 ºC and 15 psi for 1 h.

2.1.2 Plastic ware

Sterile disposable filter sterilization units (0.22 µm) and petridishes (55 mm and 85 mm diameter) were procured from "Laxbro" or "Axygen", India. Microfuge tubes (1.5 mL and 2 mL capacity), microtips (10, 200 and 1000 µL capacity) and PCR tubes (0.2 mL and 0.5 mL capacity) were obtained from "Tarsons" and "Axygen", India and EIA/RIA medium size flat bottom, 96 well plates from "Costar".

2.1.3 Chemicals

Tris, IPTG, X-gal, SDS, BSA, EDTA, PMSF, PIPES, DTT, PVPP, APS, imidazole, urea, agarose, and ethidium bromide were purchased from Sigma-Aldrich (USA) and Bioworld (USA). Restriction enzymes, T4 DNA ligase, RNase A and lysozyme were obtained from NEB (USA), Promega (USA), Bioenzymes (USA) and Amersham (UK). Different kits were purchased from BD Clonetech (Japan), Invitrogen (USA), Promega (USA) and Sigma-Aldrich (USA). *Taq* DNA polymerase was obtained from Sigma-Aldrich (USA) and Bangalore Genei (India). Plasmid vectors, pGEM-T Easy Vector, pET30b(+)

was purchased from Promega (USA) and Novagen (USA), respectively. DNA Labeling kit and Hybond-N+ membrane were obtained from Amersham (UK). [α-^{32}P]-dCTP was obtained from Bhabha Atomic Research Centre (BARC), India. All other chemicals and solvents of analytical grade were purchased from HiMedia, Qualigens Fine Chemicals and E-Merck Laboratories, India. All chemicals used in the tissue culture study were of analytical grade (AR) and were obtained from "Qualigens", "S.D. Fine Chemicals" or "HiMedia", India. The oligonucleotides/ primers were synthesized by "Eurofins Genomics India Limited, India".

2.1.4 Equipments used for the study

S.No.	Equipment	Make
1	Balances	Contech/ Sartorious
2	Water bath	Fisher Scientific
3	Dry Bath	Eppendorf
4	Incubator	New Brunswick
5	Centrifuge	Sorvall/eppendorf
6	Gel Documentation system	BioRad
7	Thermo Cycler PCR machine	BioRad
8	Spectrophotometer	Perkin Elmer
9	Power pack	Bio-Rad/GeNei
10	Agarose Gel Electrophoresis Units	Bangalore GeNei/ Bio-Rad
11	Protein Gel Electrophoresis Units	GE HealthCare
12	Sonicator	Misonix
13	pH-Meter	Digital corp.
14	Water purification system	Millipore Unit (Milli RO/ MilliQ)
15	Microwave oven	Electrolux
16	Fridge/ Deep freezer	Vestfrost/Godrej
17	Magnetic rotator	REMI
18	Laminar Air Flow	Microfilt India
19	Typhoon trio+	GE Healthcare
20	Icematic	Saksham Technologies
21	ELISA Plate Reader	BioRad

22	Real Time PCR machine	Stratagene
23	Speed Vac concentrator	Eppendorf
24	Hybridisation incubator	SciGene
25	Light microscope	Axioplan2(Carl Zeiss)
26	NanoVue system (Nanodrop)	GE Healthcare

2.1.5 Bacterial strains

Strain	**Genotype**
Escherichia coli XL-10 Gold	*endA1 glnV44 recA1 thi-1 gyrA96 relA1 lac Hte Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 tet^R F'[proAB lacI^{q}ZΔM15 Tn10(Tet^R Amy Cm^R)]* (Invitrogen, USA).
E. coli XL-1 Blue	*RecA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F´ proABlacIqZΔM15 Tn10 (Tetr)* (Stratagene, USA)
TOPO 10	*F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1araD139 Δ(araleu) 7697 galU galK rpsL (StrR) endA1 nupG*
E. coli Rosseta(DE3)	F– *ompT hsdS$_B$*(r_B^- m_B^-) *galdcm lacY1* (DE3) pRARE (*argU, argW, ileX, glyT, leuW, proL*) (Cm^r)
Agrobacterium tumefaciens GV2260	*C58, Rifr, pGV2260 (pTiB6S3_T-DNA), Carbr, Octopine type*

2.1.6 Vectors (plasmid) used

Plasmid construct	**Important features (reference or source)**
pGEM-T Easy Vector	Cloning vector (Promega, USA)
pET-30b(+)	Expression vector (Novagen)
Modified pCAMBIA1301	Binary vector

2.2. Plant material and source

Leucaena leucocephala cultivar K-636 growing in the orchard of NCL campus was used for the studies. Seeds from the plant were treated as per Shaik et al. (2009) and grown either in soil or in modified MS medium (Shaik et al. 2009). Cultivar K-636 produces erect boles suitable for timber production. Tobacco seeds (*Nicotiana tabacum* var. Anand 119) were procured from the Horticultural Division, NCL.

2.3 Buffers and Solutions

2.3.1 Buffers and Solutions for DNA Electrophoresis

Name	Ingredients	Preparation and storage
50xTAE buffer	2 M Tris Acetic acid 0.05 M EDTA	pH was adjusted to 8.0 with NaOH and stored at room temperature
10xTBE buffer	90 mM Tris 90 mM Boric acid 2 mM EDTA	Stored at room temperature
DNA loading buffer	0.25 g Xylene cyanol 0.25 g Bromophenol blue 0.25 g Ficoll 400 1.46 g EDTA Make up the volume to 100 ml with MB grade H_2O	The solutions were filter sterilized using 0.22 µ filter and stored at 4 ºC.

2.3.2 Buffers and Solutions for Southern blotting and hybridization

Name	Ingredients	Preparation and storage
Depurination buffer	0.25 N HCL	Freshly prepared
Denaturation buffer	1.5 M NaCl 0.5 M NaOH	Room temperature
Neutralization buffer	1.5 M NaCl 1.0 M Tris HCl (pH 7.4)	Room temperature
20 X SSC	3 M NaCl 0.3 M Sodium citrate (pH 7.0)	Room temperature
Hybridization buffer	1% BSA 1.0 mM EDTA, pH 8.0 0.5 M Sodium phosphate (pH 7.2) 7% SDS	Room temperature
Low stringency wash Buffer	2 X SSC 0.1% SDS	Freshly prepared
High stringency wash Buffer	0.1 X SSC 0.1% SDS	Freshly prepared

2.3.3 Buffers and Solutions for Plasmid isolation (Alkaline lysis method)

Name	**Ingredients**	**Preparation and storage**
Solution I or Resuspension buffer	50 mM Glucose, 25 mM Tris-HCl (pH 8.0), 10 mM EDTA (pH 8.0)	4 ºC
Solution II or Lysis buffer	0.2 N NaOH and 1% SDS	Freshly prepared
Solution III or Neutralisation buffer	3 M Potassium acetate (pH 4.8)	4 ºC
RNase A	10 mg/mL	-20 ºC
Other solutions or Reagents	Chloroform, Absolute ethanol, 3.0 M Sodium acetate, 70% ethanol & Deionized sterile water	Room temperature

2.3.4. Different media and buffers used for bacterial studies

Name	**Ingredients**	**Preparation and storage**
Luria Bertani Broth (LB)	1% Bactotryptone 0.5% Yeast extract 1% NaCl	pH adjusted to 7.0 with NaOH, store at room temperature or at +4 ºC
SOB media	2% Bactotryptone 0.5% Yeast extract 10 mM NaCl 10 mM $MgCl_2.6H_2O$ 2 mM KCl	pH adjusted to 6.8 with NaOH, store at room temperature or at +4 ºC
TB buffer	10 mM PIPES 15 mM $CaCl_2$ 250 mM KCl	pH was adjusted to 6.8 with KOH. $MnCl_2$ was added to final concentration of 55 mM and filter sterilized
YEP	0.5% Beef extract 0.1% Yeast extract 0.5% Peptone 0.5% Sucrose 0.049% $MgSO_4.7H_2O$	Stored at room temperature

2.3.5 Antobiotic stock and working solutions for bacterial work

Antibiotic	**Vector**	**Stock concentration**	**Working concentration**
Ampicillin	pGEM-T Easy	100 mg/mL	100 µg/mL
Kanamycin	pET-30b(+) and Modified pCAMBIA1301	50 mg/mL	50 µg/mL

2.3.6 Buffers and Solutions for Protein Gel Electrophoresis (PAGE) and Coomassie-blue staining

Name	Components	Preparation and storage
Monomer solution	29.2% acrylamide 0.8% bis-acrylamide in water	Store 4 ºC (in darkness)
Stacking gel	Distilled water 3.40 mL 1 M Tris-HCl (pH 6.8) 00.63 mL Acrylamide/bis 30%, 0.83 mL 10% SDS 0.05 mL 10% (w/v) APS 0.05 mL TEMED 4 µL	Freshly prepared
Separating gel	Distilled water 3.3 mL 1.5 M Tris-HCl (pH 8.8) 2.0 mL Acrylamide/Bis 30% 4.0 mL 10% SDS (SDS-PAGE) 0.1 mL 10% (w/v) APS 0.1 mL TEMED 07 µL	Freshly prepared
5X Protein loading dye	Distilled water 2.7 mL 0.5 M Tris-HCl (pH 6.8) 1.0 mL Glycerol, 2.0 mL 10% SDS(SDS-PAGE) 3.3 mL β-Mercaptoethanol 0.5 mL 0.5% Bromophenol blue 0.5 mL	Room temperature
10X SDS electrode buffer	Tris base 15.1 g Glycine 94.0 g SDS 0.5 g Adjust pH-8.3. Make up the volume to 500 mL with distilled water.	Room temperature
Coomassie-blue Staining solution	Coomassie-blue R 250 0.1 g Methanol 40 mL Acetic acid 10 mL Make up volume to 100 mL with distilled water	Room temperature
Destaining solution (I)	Methanol 50 mL Acetic acid 10 mL Make up volume to 100 mL with distilled water	Room temperature
Destaining solution (II)	Methanol 10 mL Acetic acid 10 mL Make up volume to 100 mL with distilled water	Room temperature

2.3.7 Solutions for Protein Gel Electrophoresis (PAGE) silver staining

Silver staining Fixer solution	Methanol Acetic acid	40% 10%	Room temperature
Sensitising solution	$Na_2S_2O_3$	0.2%	Room temperature
Silver stain solution	Silver nitrate Formaldehyde	0.2% 0.01%	Freshly prepared in dark
Developing solution	Na_2CO_3 Formaldehyde	6% 0.02%	Freshly prepared

2.3.8 Buffers for protein extraction and purification under denaturing conditions from *E. coli*

Name	**Components**	**Preparation and storage**
Lysis buffer	20 mM Tris-HCl (pH 8.0) 1 mM EDTA 10 mM $MgCl_2$ 0.4 mM NaCl 10% Glycerol Lysozyme 200 µg/mL (Added freshly)	Filtered sterilized and stored at 4 ºC
Dispersion buffer/Binding buffer	10 mM Tris-HCl (pH 8.0) 100 mM NaH_2PO_4 2 M NaCl 8 M Urea 10 mM imidazole	Room temperature (pH adjusted to 8.0 every time before use)
Wash buffer	10 mM Tris (pH 8.0) 100 mM NaH_2PO_4 2 M NaCl 8 M Urea 30 mM imidazole	Room temperature (pH adjusted to 8.0 every time before use)
Elution buffer	10 mM Tris (pH 8.0) 100 mM NaH_2PO_4 2 M NaCl 8 M Urea 250 mM imidazole	Room temperature (pH adjusted to 8.0 every time before use)

2.3.9 DNA and protein markers used

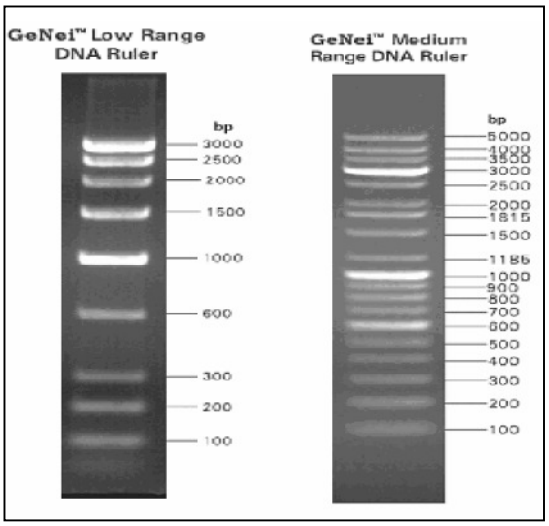


Fig. 2.1: DNA ladder (Bangalore Genei)

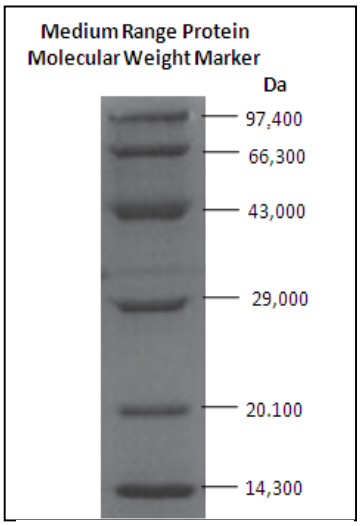


Fig. 2.2: Protein Molecular Weight Marker

2.4 Methods/Protocols for bacteriological work

2.4.1 Bacterial culture/storage conditions

E. coli was grown at 37 ºC with shaking at 200 rpm in Luria Bertani (LB) broth/SOB and maintained on LB/SOB plates with 1.5% agar. For plasmid DNA preparation recombinant *E. coli* was grown in LB medium supplemented with appropriate antibiotic(s).

2.4.2 Competent cell preparation and bacterial transformation

E. coli competent cells were prepared as per the standard molecular biology protocol, suspended in TB buffer, aliquot prepared and stored at -80 ºC. The competent *E. coli* cells were transformed as per Sambrook and Russell (2001). X-gal (40 µL from stock concentration 20 mg/mL) and IPTG (4 µL from 1 M stock) was spread on the medium with appropriated antibiotic as and when necessary.

2.4.3 Colony screening by PCR

This method relies on the selectivity of PCR template and to the potential of PCR to greatly amplify very small amount of template to detectable level. A part of colony on the plate is picked using a sterile pipette tip and suspended

in sterile water and 2-3 µL of the soup is used as template in PCR taking either the insert specific primers, vector specific primers or a combination of both to detect correctly the absence or presence of insert as well as its orientation with respect to the vector, depending upon the primer combination used. An additional cycle of 5 min at 95 °C is kept at the start of PCR to lyse the content of bacterial cells to the PCR mix.

2.4.4 Preservation of bacterial culture

Bacterial cultures harbouring plasmid with target of interest were stored in glycerol at ultra low temperature. In a microfuge tube, 800 µL of bacterial culture and 200 µL of sterile glycerol were mixed by vortexing briefly and snap frozen in liquid N_2 and stored at -80 ° C. The glycerol stocks were revived every six months and fresh stocks were prepared.

2.5 Isolation of Nucleic Acids, restriction digestion and PCR

2.5.1 Isolation of plasmid DNA from *E. coli* cells

Plasmid was isolated from *E. coli* using alkaline lysis method as reported by Sambrook and Russell (2000).

2.5.2 Isolation of plant Genomic DNA

For maxiprep, maxiprep gDNA isolation kit (Himedia, India) was used. For miniprep analysis, miniprep genomic DNA isolation kit of Bangalore Genei or HImedia (both India) was used.

2.5.3 Restriction digestion and ligation of DNA

Plasmid and genomic DNA digestion and ligation were performed as per the supplier's recommendations.

2.5.4 Extraction and purification of DNA from agarose gels

DNA was gel eluted as per the supplier's (Invitrogen, USA; Axygen, India) recommendation.

2.5.5 RNA Extraction and DNase I treatment

RNA was isolated from the plant materials using RNA isolation kit from Invitrogen (USA) in an RNase free environment (Sambrook and Russell

2000) following the manufacturer's recommendation. Trizol was used to disrupt the nucleo-protein complex and rest kit components for RNA purification. DNase I treatment was given to eliminate DNA contamination in RNA as per the supplier's recommendations. Absence of gDNA in RNA sample was confirmed by PCR, taking isolated RNA as template and C4H specific primers for amplification using *Taq* DNA polymerase which does not accept RNA as template.

2.5.6 Spectrophotometric determination of nucleic acid concentration

DNA concentration was determined by measurement of the absorption at 260 nm in NanoVeu plus. Concentration of 1 µL of undiluted nucleic acid solution was determined in triplicate and the average value considered for calculation. The ratio of absorbance at 260/280 nm should be near 1.8 for DNA and near 2.0 for RNA. The integrity of nucleic acid was checked by gel electrophoresis.

2.5.7 First strand cDNA synthesis by Reverse Transcription

cDNA synthesis kit of Invitrogen (Superscript III, MMLV Reverse Transcriptase) was used to synthesize first strand cDNA. This is an engineered Reverse Transcriptase with reduced RNase H activity and has potential to reverse transcribe full length of transcripts. First strand cDNA was synthesised taking 1 µg total RNA (DNA free) as template and oligo dT or gene specific reverse primer, as the case may be, in RNase free tube following supplier's information. After RNase H treatment, cDNA were diluted and used as template in PCR reactions.

2.5.8 Polymerae Chain Reaction (PCR)

PCR is a very powerful technique to exponentially amplify a given DNA molecule (fragment of interest) using a DNA polymerase enzyme (Mullis and Faloona 1987). This amplification is very specific in the sense only the sequence flanked by a pair of primers gets synthesized. The template may be single stranded, double stranded, pure or a mixture of DNA molecules (Arnheim and Erlich 1992; Mullis and Faloona 1987). PCR may be used to amplify a given DNA fragment, introduce restriction site to a fragment for its directional cloning to a vector, create sequence mismatch, deletion, addition or to differentiate between two alleles. In the present study, PCR has been

extensively used for various purposes. The PCR condition used to amplify a certain fragment (or any recommendation/deviation from instruction manual) has been mentioned in the respective section.

2.5.9 Quantitative Real Time PCR (Q-RT PCR)

Quantitative Real Time PCR is a powerful technique to detect and monitor the exponentially amplifying product on a real time basis without opening the reaction tube or terminating the reaction. This was first achieved by adding EtBr to the reaction mixture and capturing the fluorescence emitted on a real time basis as the concentration of double stranded DNA increased (Higuchi et al. 1992). Further advancement of the procedure included the use of a video camera to monitor multiple PCR reactions taking place at the same time. The change in fluorescence in each individual tube became possible this way (Higuchi et al. 1993). QRT-PCR has gone a long way after that by the introduction of automated machines to detect the fluorescence, introduction of probes with phlorophores and quenchers, and dyes that absorb and emit at different wavelengths making multiplex PCR possible.

In the present work, for *C4H* transcript level study from *L. leucocephala*, Brilliant® II QPCR Master Mix for *Taq*man probe with low rox (reference dye) (Stratagene, USA) has been used. The Brilliant II QPCR master mix includes SureStart® Taq DNA polymerase, a modified version of *Taq*2000™ DNA polymerase with hot start capability and a passive reference dye (rox). For transcript level of *C4H* and other phenylpropanoid pathway genes, like *PAL*, *4CL, CAD* and *CCoAOMT* from transgenic tobacco plants downregulated for *C4H* sequence derived from *L. leucocephala*, mastermix for syber green with low rox (Eurogentech, Belgium) has been used. The concentration of probe (for *Taq*man assay), and forward and reverse primers for individual genes to be used in reactions were empirically determined. Reactions were performed in independent triplicates and the Ct values were averaged. An individual reaction was of 25 µL volume that contained 12.5 µL of 2X master mix and empirically determined amount of primers (and probe for *Taq*man assay) and normalized amount of template (first strand cDNA in our case). The sequence of primers and probe has been given in respective chapters.

2.6 Southern hybridization

Southern hybridization was first reported by Southern (1975). This is a powerful technique to detect specific sequence among DNA fragments that have been separated through electrophoresis. Genomic DNA is first digested using appropriate restriction enzymes and electrophoretically separated on agarose gel and then blotted on to a nylon membrane. The membrane is then kept in hybridization buffer containing radioactive probe for hybridization to take place. The blot (membrane) is then washed and then exposed to a sensitive screen or X-ray film and the screen is either scanned or the film is developed.

2.6.1 Southern hybridization procedure

20 µg of gDNA was digested with *Eco*RI, *Sac*I and *Nde*I in separate tubes and electrophoresed on 0.8% agarose gel. DNA was blotted on to Hybond-N^+ membrane (Amersham) and approximately 400 bp fragment (3' terminus of the gene) was used as probe. Pre-hybridization was performed for 8 h at 60 ºC in hybridization buffer (0.5 M $Na_2HPO_4 \cdot 2H_2O$, pH 7.2; 1 mM EDTA, pH 8.0; 7% SDS and 1% BSA) in a hybridization chamber (SciGene). Radiolabelled probe was synthesized using α'-P^{32} labelled dCTP, unlabelled dNTPs (without dCTP), approximately 400 bp purified PCR product produced by amplifying *L. leucocephala* gDNA as template with primers C4H-RL and C4H-Stop (primer sequences are given in Table 3.1) using Klenow fragment. Hybridization was carried out for 16-20 h at 60 ºC in fresh buffer containing probe. Membrane was first rinsed and then washed once in low stringency buffer (2X SSC, 0.1% SDS, 42 ºC) for 10 min and once in high stringency buffer (0.1X SSC, 0.1% SDS, 65 ºC) for 5 min. Washed membrane was exposed to storage phosphor screen for overnight and scanned by phosphor imager system (Typhoon Trio+, GE Healthcare, USA).

2.6.2 Radioactive labelling of the probe DNA

Radioactive labeling of the probe DNA was done using the Megaprime DNA labeling kit (Amersham, UK). The manufacturer's recommendations were followed to label 20 ng of template DNA with the help of Klenow fragment.

2.7 Protocols for running/analyzing PAGE gels

2.7.1 Preparation, loading and running the gel

10 or 12% SDS-PAGE gels of appropriate dimensions were prepared as per the need of the experiment. Utmost care was taken to avoid contact of acrylamide solution to skin or inhale its fume. Protein sample was mixed with protein loading dye and heated for 3-5 min to denature the proteins for SDS-PAGE analysis. Gel was run at room temperature at constant current of 2-3 mA per well. Gel was run till the dye front passed through the bottom of the gel.

2.7.2 Coomassie Brilliant blue staining/destaining of gel

The run gel was carefully transferred to staining box and rinsed with de-ionized water. Enough staining solution was poured in to the staining box and the kept in rocking position for 6 hr to overnight to stain the gel. For destaining, gel was rinsed with de-ionized water after discarding the used staining solution. The gel was destained in destaining solution (I) for 1-2 hr and then destaining solution (II) till the gel swells and become clear.

2.7.3 Silver staining of SDS-PAGE gels

The run gel was rinsed with MB grade water and fixer solution and kept in fixer solution for 1 hr to overnight. The gel is given one or two washes in 50% ethanol of 20 min duration. The gel is rinsed at least thrice with MB grade water between every two steps. The gel is transferred to thiosulphate solution for 1 min. The gel is transferred in staining solution and with formaldehyde and incubated in dark for 20 min with intermittent shaking. Developer solution is then poured in to the staining box and the box is shaken gently in hand till bands appear. The gel is then stored in fixer solution.

2.8 Determination of protein concentration

The assay is based on the observation that the absorbance maxima for an acidic solution of Coomassie Brilliant Blue G-250 shifts from 465 nm to 595 nm when binding to protein occurs. This was first demonstrated by Bradford (1976) and also known as Bradford assay. Bradford reagent was purchased from Biorad, USA. A standard curve of OD verses known concentration of

BSA as standard was plotted and the concentration of unknown samples were determined accordingly.

2.9. Phloroglucinol Staining

2% Phloroglucinol solution was prepared in 95% ethanol and stored under dark condition. Free hand sections of tissues were taken and submerged in phloroglucinol solution for 5 minutes. After that, equal volume of concentrated HCl was added to the tissue sections submerged in phloroglucinol. Lignified tissues take red colour. Photographs were immediately taken under a light microscope. The staining is not permanent.

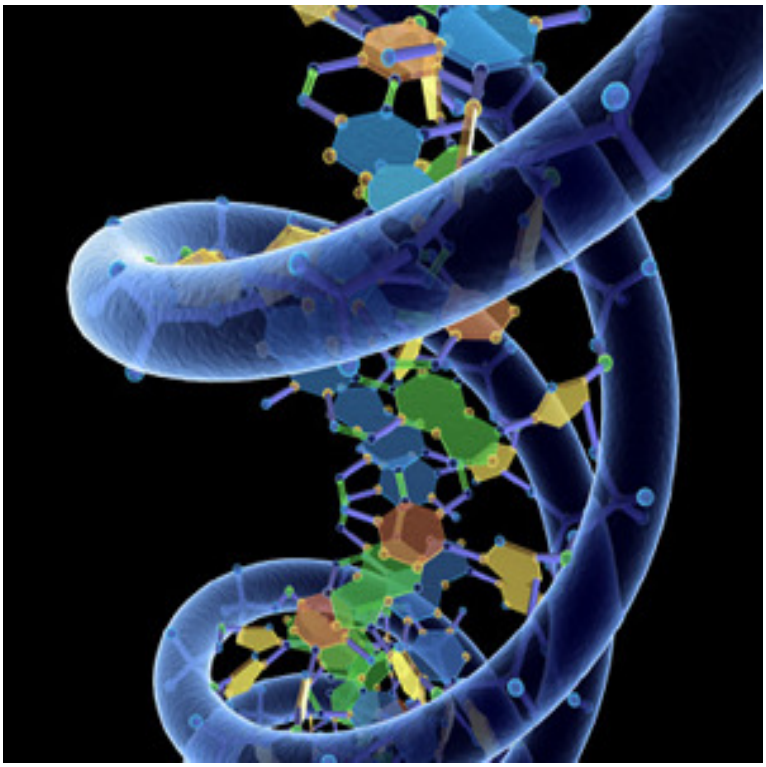
2.10 Lignin estimation

Lignin content from the plant tissues was determined using the Klason method (Kirk and Obst 1988) with slight modification. Tissues were air dried and ground to fine powder and continuously extracted in acetone: water (10:1 v/v) for 48 hours at 55 ºC. Tissue powder was then dried to constant weight at 105 ºC to get cell wall residue (CWR). About 200 mg of CWR was acid hydrolyzed (72% H_2SO_4) for 3 h at 25 ºC in shaking condition. Hydrolyzed CWR is then diluted to 3% H_2SO_4 using molecular grade water and autoclaved for one hour. After cooling down, residue was filtered and washed with hot water to remove acid and dried to constant weight at 105 ºC to get acid insoluble or Klason lignin (KL). KL was expressed as percentage of CWR. The filtrate after appropriate dilution was used for spectrophotometric determination of acid soluble lignin (ASL) at 205 nm using the formula given below. Total lignin was calculated as the sum of ASL and KL.

$$\text{Acid soluble lignin (\%)} = \frac{100 \times A205 \times \text{Volume of filtrate (mL)} \times \text{dilution factor}}{110 \times \text{Initial dry powder weight (mg)}}$$

Here, A205 denotes absorbance at 205 nm.

CHAPTER - 3



Isolation, Cloning and Characterization of Cinnamate 4-Hydroxylase (C4H) Gene(s) from Leucaena leucocephala

Chapter 3: Isolation, Cloning and Characterization of Cinnamate 4-Hydroxylase (*C4H*) Gene(s) from *Leucaena leucocephala*

This chapter describes the isolation, cloning and characterization of Cinnamate 4-Hydroxylase gene(s) from *L. leucocephala*. This chapter also deals with the specific methods used during the course of work carried out in this chapter.

3.1 Review of literature

C4H has been extensively studied post its discovery because of its crucial role in plant metabolism. *C4H* is the second gene of phenylpropanoid pathway that gives rise to a plethora of compounds including lignin, flavonoids, hydroxycinnamic acid esters, lignans, stilbenes, and a host of other secondary metabolites that act against plant pathogens (Boudet 2007; Chapple 1998; Ehlting et al. 2006; Frank et al. 1996; Hahlbrock and Scheel 1989; Lu et al. 2006; Naoumkina et al. 2010). Identification of cDNA sequence corresponding to C4H was the first plant P450 that could be associated to an enzyme function (Fahrendorf and Dixon 1993; Mizutani et al. 1993a; Teutsch et al. 1993). Isolation of C4H protein first took place from *Jerusalem artichoke* (Gabriac et al. 1991) and mung bean (Mizutani et al. 1993a). Antibodies raised against *J. artichoke* C4H protein strongly inhibited C4H activity from other plans implying close sequence identity between C4H from different plants, which was later proved correct and in fact, C4H from different plants share more than 88% sequence identity at amino acid level. Following C4H protein isolation in mung bean, *C4H* cDNA sequence isolation in mung bean became possible through partial peptide sequencing (Mizutani et al. 1993b). Polyclonal antibodies raised against purified C4H from *J. artichoke* was used to screen wound induced cDNA expression library from *J. artichoke* to isolate partial *C4H* cDNA sequence. This partial cDNA sequence was used as probe to isolated full length *C4H* cDNA sequence from *J. artichoke* (Teutsch et al. 1993). At the same time, another group using antibody against avocado P450 (CYP73A1) expressed during fruit ripening, isolated *C4H* cDNA sequence from alfalfa cDNA expression library (Fahrendorf and Dixon 1993). Fahrendorf and Dixon (1993) expressed the cDNA sequence in yeast and concluded that the cDNA sequence possess

C4H activity. Further, they showed that C4H is encoded by a small gene family in alfalfa and is strongly induced by fungal elicitor in alfalfa cell suspension culture. Since then, *C4H* cDNA and gDNA clones have been isolated and characterized from many plants (Akashi et al. 1997; Bell-lelong et al. 1997; Betz et al. 2001; Chen et al. 2007; Frank et al. 1996; Hotze et al. 1995; Hübner et al. 2003; Kawai et al. 1996; Koopmann et al. 1999; Liu et al. 2009; Lu et al. 2006; Mizutani et al. 1997; Overkamp et al. 2000; Petersen 2003; Whitbred and Schuler 2000; Yamamura et al. 2001).

3.2 Materials and methods

3.2.1 Primer designing

C4H nucleotide sequences of Fabaceae family plants and other tree species available in NCBI database were retrieved and aligned for conserved regions. Degenerate primers were designed from those regions and used in PCR for isolating partial *C4H* sequence from *Leucaena*. Tm of the primers was theoretically calculated using the software pDRAW32 or primer blast (NCBI). The list of all the primers used in the study is given in Table 3.1.

3.2.2 Isolation of partial *C4H* sequence

After RNA isolation from *L. leucocephala* debarked xylem scrapings and first strand cDNA synthesis, PCR with two sets of primers (C4H-F1 with C4H-R1) gave amplification of around 600 bp, which was cloned in to pGEM-T Easy vector and sequenced. From within this sequence, primers specific for isolating the rest of the sequence including the 5' and 3' UTR were designed.

3.2.3 Isolation of 3' UTR by 3' RACE PCR

3' UTR of *C4H* was isolated using Rapid Amplification of cDNA Ends (RACE) kit (Invitrogen) following the manufacturer's recommendations. Briefly, cDNA was synthesised using RNA isolated from *L. leucocephala* as template and GeneRacer Oligo dT as reverse primer. cDNA was diluted and used as template in primary 3' RACE PCR reaction, which selectively enriches the desired transcripts in question. The primers used in the primary reaction were

C4H-RACE-F1 and GeneRacer 3' Primer (supplied with the kit). The protocol suggested by the supplier was followed during primary 3' RACE PCR. The primary reaction product was 200 times diluted and used as template in secondary 3' RACE PCR. The primers in the reaction were C4H-RACE-F2 and GeneRacer 3' Nested Primer (supplied with the kit). PCR was set up under following condition: 95 ºC for 5 min; 30 cycles of 30s at 95 ºC, 30s at 65 ºC and 1.5 min at 72 ºC; and one last cycle of 7 min at 72 ºC. The PCR product was gel eluted, ligated with pGEM-T Easy vector and cloned in to *E. coli*. Positive clones were selected on the basis of blue-white screening and colony PCR (section 2.4.3). Plasmid isolated from the positive clones and sequenced.

Table 3.1 Sequence of primers used in the study

Serial no.	Primer name	Primer sequence (5'→3')
1	C4H-F1	ATGGATYTYCYMYTMHTVGAGAAG
2	C4H-R1	WBRAASADMGGRTCRTCBWCRCTCTC
3	GeneRacer Oligo dT Primer	GCTGTCAACGATACGCTACGTAACGGCATGACAGTG$(T)_{24}$
4	C4H-RACE-F1	CAGCAATACAGGGATGGATGGGA
5	GeneRacer 3' Primer	Position 1-25 of GeneRacer™ Oligo dT Primer
6	C4H-RACE-F2	ATCCCGAGTCGGCGACCAATGGGA
7	GeneRacer 3' Nested Primer	Position 14-36 of GeneRacer™ Oligo dT Primer
8	C4H-Stop	CTAGAAGGACCTWGGYTTGGCAAC
9	Abridged Anchor Primer	GGCCACGCGTCGACTAGTACGGGIIGGGIIGGGIIG
10	C4H-RACE-R1	ATCCTCGCTCTCGAACCTTCTATCGA
11	C4H-Start	ATGGATCTCCTACTCCTGGAGAAGAC
12	C4H-RL	CTGGTCCCCCACATGAACCTCCA

3.2.4 Isolation of 5' UTR of C4H

To isolate 5' end of *C4H*, anchored PCR approach as described by Chen et al. (2005) to isolate upstream sequence of a MYB gene from genomic DNA of *Boea crassifolia* was followed except that in our case we used cDNA as template. In short, 'C4H-Stop' primer (primer that includes stop codon of *C4H*) was used to selectively reverse transcribe *C4H* mRNA from total RNA pool in a reaction catalyzed by Moloney Murine Leukemia Virus Reverse Transcriptase (M-MLV RT, Superscript III, Invitrogen) that is capable of transcribing full length of long transcripts. After RNaseH treatment, the cDNA reaction mixture was purified using PCR purification kit (Sigma, USA) to remove unspent dNTPs, primers etc. The purified cDNA was subjected to C-tailing reaction using the enzyme terminal deoxynucleotidyl transferase and dCTP (Promega, USA) as specified by Chen et al (2005). The poly(dC) tailed cDNA was used as template in subsequent PCR amplification using Abridged Anchor Primer (AAP) and a nested primer designed for 5'RACE of *C4H* (C4H-RACE-R1) under the following conditions: 95 ºC for 3 min, 35 cycles of 30s at 95 ºC, 30s at temperature gradient between 60 ºC to 65 ºC and 1 min at 68 ºC, and one cycle of 7 min at 72 ºC. The PCR product obtained was purified and cloned into pGEM-T Easy vector (Promega, USA). After transformation into XL-10 blue cells, 10 positive clones were sequenced.

3.2.5 Isolation of full length cDNA / genomic DNA sequence of *C4H*

C4H-Start and C4H-Stop primers designed on the basis of the sequences got after 3' and 5' RACE of C4H were used to isolate full length of *C4H* from *L. leucocephala*. PCR was carried out using LA *Taq* DNA polymerase (Sigma, USA) under the following condition: 95 ºC for 3 min, 35 cycles of 30s at 95 ºC, 30s at 60 ºC and 1.3 min at 68 ºC, and one cycle of 7 min at 72 ºC. PCR product was purified and cloned in to *E. coli* and 20 positive clones were fully or partially sequenced.

Genomic DNA clone of *C4H* was isolated by amplifying genomic DNA of *L. leucocephala* using start and stop primers of *C4H* and LA *Taq* DNA polymerase under following conditions: 95 ºC for 5 min, 35 cycles of 30s at 95 ºC, 1min at 60 ºC and 2.3 min at 68 ºC, and one cycle of 7 min at 72 ºC. Three gDNA clones were fully sequenced.

3.2.6 Bioinformatic and phylogenetic analysis

Homology searches were performed on NCBI BLAST webpage (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Conceptual translation of nucleotide sequence, prediction of molecular mass and isoelectric point was carried out using ExPASy proteomics server (http://expasy.org/tools/). Signal sequence was predicted using SignalP 4.0 (http://www.cbs.dtu.dk/services/SignalP/) programme. Pairwise sequence alignment and percent identity was calculated using EMBOSS Pairwise Sequence Alignment tool (http://www.ebi.ac.uk/Tools/psa/) for both local and global alignments. Multiple sequence alignments were carried out using the program ClustalW2 (http://www.ebi.ac.uk/Tools/msa/) or by logging in to www.justbio.com. *In silico* restriction analysis of DNA fragments was performed using NEBcutter V2.0 tool (http://tools.neb.com/NEBcutter2/). Phylogenetic tree was constructed using the executable program MEGA5 (Tamura et al. 2011). The evolutionary relationship was inferred using neighbour joining method with 1000 replicates.

3.2.7 Determination of copy number of *C4H* in *L. leucocephala* genome

Copy number of *C4H* in *L. leucocephala* was determined using DNA hybridization technique. The protocol followed for the hybridization has been discussed in chapter 2, section 2.6.1. The protocol for radioactive labelling of probe DNA has been discussed in section 2.6.2. The template for probe preparation was prepared as follows: *L. leucocephala* genomic DNA was amplified using the primers C4H-RL and C4H-Stop and purified using PCR purification kit (Invitrogen, USA). 20 ng of the ~400 bp purified PCR product was used as template in the labelling reaction.

3.3 Results

3.3.1 Multiple sequence alignment of *C4H* nucleotide sequences from different plants and primer design

C4H protein sequences are more than 88% identical to each other. This level of identity has long stretches of identical sequences. Also, C4H belongs to a large group of cytochrome P450 monooxygenases, which contains members other than C4H with less than 40% sequence identity at amino acid sequence level. Every amino acid is encoded by triplets of nucleotides called codon, and sometimes one amino acid is coded by more than one codon, greater level of dissimilarity is observed in the nucleotide sequence. *C4H* sequences from different plants are about 75% identical at nucleotide level; it offered greater level of control in designing primers. Hence, *C4H* nucleotide sequences from plants/trees were retrieved from NCBI sequence database and aligned for consensus regions. Different sets of degenerate primers were designed. Fig. 3.1 shows the consensus sequence of the different sequences and the region from where degenerate primers were designed.

```
  1 ATGGATYTYC YMYTMHTVGA GAAGRCSYTM HTNRSBYYYT TCKYNGCSRY
 51 BBTHVTHGCB RYNVYBRTNT CBARRCTHCG HKSVAARMRN TTYAARYTBC
101 CBCCNGGDCC NHTHSSNKTN CCNRTBTTYG GHAAYTGGCT YCARGTBGGB
151 GAYGAYYTVA ACCAMMGNAA HYTBRCHRAY WWNKCSAARM RDTTYGGHGA
201 BVTNTWYHTN YTHMGMATGG GNCARCGYAA YYTNGYBGTB GTBTCNTCNC
251 CKGAVHTVKC BAARGARGTB YTVCAYACDC ARGGHGTBGA RTTYGGNTCN
301 MGVACNMGVA AYGTNGTNTT YGAYATHTTY ACBGGDAARG GHCARGAYAT
351 GGTKTTYACB GTNTAYRGHG ARCAYTGGMG VAARATGMGG MGRATYATGA
401 CBGTVCCDTT YTTYAYNAAY AARGTBGTVC ARCARYABMG NBDNRRVTGG
451 GARDHNGARG YBGMNVVNGT NRTBGMGGAY GTKARRAARV AYCCNSARKC
501 DGCNAMBAMY GGRRTBGTDY TRMGKMRRMG RYTRCARYTB MTGATGTAYA
551 AYAAYATGTW YMGRATYATG TTYGAYASRA GRTTYGAGAG YGWVGAYGAY
601 CCKHTSTTYV WBARRYTNAR RGSBYTGAAY GGNGARAGRA GYMGRTTGGC
651 KCAGAGYTTY GADTAYAAYT AYGGBGATTT YATYCCNRTB YTVAGRCCYT
701 TYYTBAGRGG WTAYYTSAAS MTSKKBMARS ARGTKAARGA VAVRAGRHTN
751 MARCTBTTCA ARGAYYAYTT YGTYRAHGAR AGRAAGAARY TDDNRARYAC
801 MMARRVVDBD rcggvhRRBR AWRVNYTNAA RTGYGCVATH GAYCAYRTBY
851 TKGANGCYCA RMAGAARGGR GARATCAAYG ARGAYAAYGT YCTBTACATY
901 GTYGARAAYA TYAAYGTYGC HGCWATWGAR ACVRCWYTRT GGTCVATHGA
```

```
 951 GTGGGGHATY GCNGAGYTDG TSAAYCAYCC NSAVRTYCAR AASAARYTSM
1001 GNSMYGARMT BGAYASNKTN CTHGGHSHHG RNVWBCWRRT NACNGARCCN
```

Fig. 3.1: Consensus sequence arising from different *C4H* sequences. Region from where forward primer was designed is shaded in yellow whereas region from where reverse primer is designed is shaded in grey. The accession numbers of sequences used to derive this consensus sequence are FJ968526 (*Glycine max*), NM_001250388 (*Glycine max*), EU275980 (*Acacia auriculiformis* x *Acacia mangium*), AF175275 (Pea), NM128601 (Arabidopsis thaliana), EU760387 (Populus tomentosa),

3.3.2 Isolation of partial sequence of *C4H* from *L. leucocephala*

Use of degenerate primer combination (C4H-F1 and C4H-R1) with cDNA pool of *Leucaena* xylem scrapings fetched a band of about 600 bp Fig. 3.2). The band was cloned in pGEM-T Easy vector and sequenced. Blast analysis of the sequence showed that it is more than 75% identical to the previously reported *C4H* sequences at nucleotide level and more than 85% identical at amino acid level. Blast search of the conceptually translated amino acid sequence predicted the presence of putative conserved domains specific to P450 superfamily (Fig. 3.3). This confirmed that the isolated partial sequences indeed represent *C4H* from *L. leucocephala*.

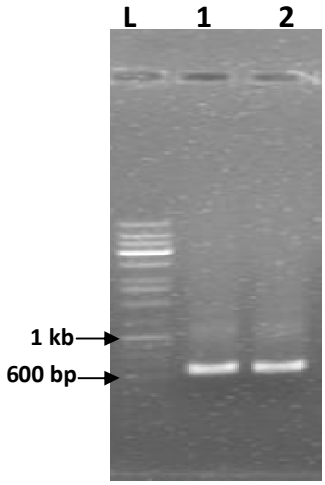


Fig. 3.2 Gel photograph showing PCR amplified partial *C4H* band of about 600 bp. (Lane L: Medium range ladder and Lanes 1 and 2: partial fragment of about 600 bp)

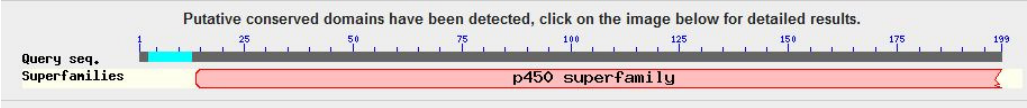


Fig. 3.3: Blast search of the conceptually translated sequence predicted putative conserved domains specific to P450 superfamily within the sequence

3.3.3 RACE reactions (3' and 5' RACE) of *C4H*

3' and 5' RACE reactions for C4H were carried out as mentioned in sections 3.2.3 and 3.2.4 respectively. 3' RACE reaction for C4H yielded a band of around 1.3 kb (Fig. 3.4 A), while 5' RACE reaction using the above protocol yielded a band of around 700 bp (Fig. 3.4 B) at all the temperature gradient between 60-65 ºC. The 3' and 5' RACE products were cloned in pGEM-T Easy vector and sequenced.

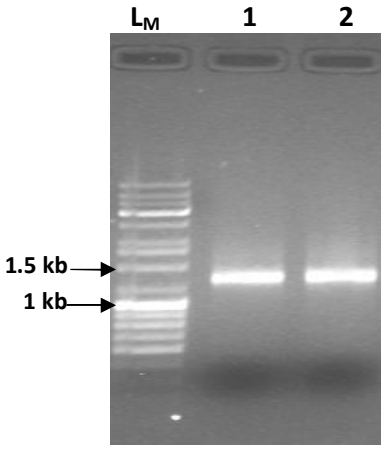


Fig. 3.4 (A): Gel photo-graph showing 3' RACE product of about 1.3 kb. Lane L_M: medium range ladder

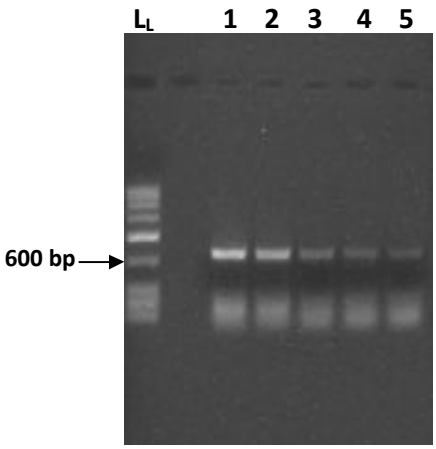


Fig. 3.4 (B): Gel photograph showing 5' RACE product of about 700 bp at all the temperatures between 60-65 º C. Lane L_L: low range ladder

3.3.4 Isolation of full length cDNA sequence of *C4H* from *L. leucocephala*

C4H-Start and C4H-Stop primers were designed from the obtained RACE sequences and PCR was carried out using LA *Taq* DNA polymerase (Sigma,

USA) under the following condition: 95 ºC for 3 min, 35 cycles of 30s at 95 ºC, 30s at 60 ºC and 1.3 min at 68 ºC, and one cycle of 7 min at 72 ºC. PCR reaction using cDNA as template resulted in a band of approx 1.5 kb (Fig. 3.5 A) which was purified and cloned in pGEM-T Easy vector. After transformation into competent *E. coli* cells, 20 clones were fully or partially sequenced.

Three cDNA clones of *C4H* were isolated this way from *L. leucocephala* which were designated as *LlC4H*1, *LlC4H*2 and *LlC4H*3 with NCBI Genbank accession numbers GU183363, HQ191221 and HQ191222 respectively. The ORF of *C4H* in *L. leucocephala* is 1518 nucleotides long coding for 505 amino acid residues. The 5' UTR of *C4H* in *Leucaena* is 74-75 nucleotides and the 3' UTR is 242-244 nucleotides long in different isoforms. Sequence alignment of the coding region of the three isoforms revealed that they are highly identical. *LlC4H*1 and *LlC4H*2 were found to be 99% identical to each other, whereas, *LlC4H*3 was found to be 97% identical to both *LlC4H*1 and *LlC4H*2. Most of the differences at nucleotide level occurred at the third nucleotide position of codons and hence very high degree of identity was observed between the three isoforms at amino acid level with LlC4H1 and LlC4H2 being 99% identical and LlC4H3 being 98% identical to both LlC4H1 and LlC4H2. Sequences showing more than 97% sequence identity at amino acid level are considered allelic variants of the gene.

3.3.5 Isolation of full length gDNA sequence of *C4H* from *L. leucocephala*

Genomic DNA clone of *C4H* was isolated by amplifying genomic DNA of *L. leucocephala* using start and stop primers of C4H and LA *Taq* DNA polymerase under following conditions: 95 ºC for 5 min, 35 cycles of 30s at 95 ºC, 1min at 60 ºC and 2.3 min at 68 ºC, and one cycle of 7 min at 72 ºC. PCR reaction using gDNA as template resulted in a band of approx 2.5 kb (Fig. 3.5 B), which was cloned into pGEM-T Easy vector following purification. Three positive clones were fully sequenced.

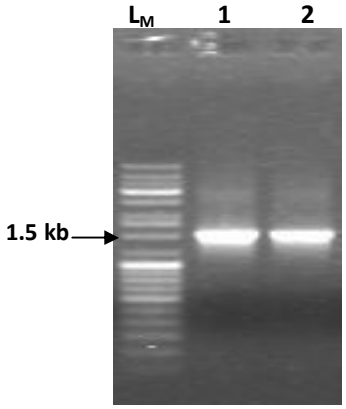


Fig. 3.5 (A): Gel picture showing band of about 1.5 kb (full length cDNA clone)
L_M= medium range ladder

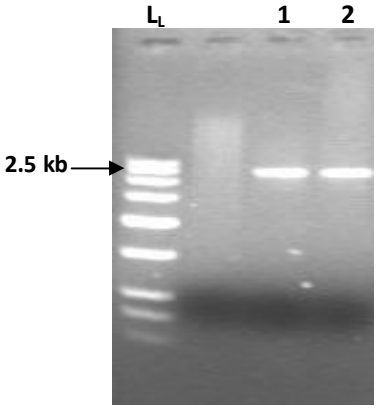


Fig. 3.5 (B): Gel picture showing band of about 2.5 kb (full length gDNA clone)
L_L= low range ladder

Three gDNA clones of *C4H* were fully sequenced and as a matter of chance, all sequences confirmed to be the gDNA clone corresponding to *LlC4H*3, which was named *gLlC4H3* (NCBI accession number JN874563). Sequence comparison between *LlC4H3* and *gLlC4H*3 indicated that *gLlC4H*3 consists of three exons of 785, 134 and 599 nucleotides and two introns of 154 and 662 nucleotides, respectively amounting to 2,334 nucleotides (Fig. 3.6).

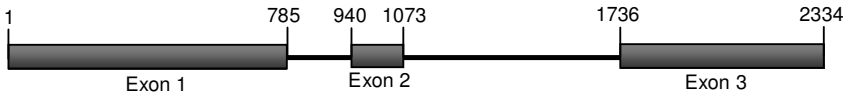


Fig. 3.6 Exon-intron pattern of *gLlC4H*3. *gLlC4H*3 contains three exons and two introns

3.3.6 Alignment of *LIC4H* isoforms' nucleotide and amino acid sequences

Deduced amino acid sequences from *LIC4H*1, *LIC4H*2 and *LIC4H*3 indicate that they all are composed of 505 amino acid residues featuring the motifs conserved in P450 family and C4H. The theoretical pI and MW as calculated using ExPASy tool came out to be 9.18, 9.24, 9.25 and 57.94 kDa, 57.93 kDa, 57.99 kDa for LIC4H1, LIC4H2 and LIC4H3, respectively. The ORF and amino acid sequences of *LIC4H*s were aligned to study the differences existing in *LIC4H* at nucleotide and amino acid level, respectively. The alignment score table and Clustal 2.1 ORF sequence alignment of *LIC4H*s at nucleotide level has been given in Table 3.2 and Fig. 3.7, respectively.

Table 3.2 Alignment score table of the ORF of the three isoforms at nucleotide level

SeqA	Name	Length	SeqB	Name	Length	Score
1	LIC4H1	1518	2	LIC4H2	1518	99.0
1	LIC4H1	1518	3	LIC4H3	1518	97.0
2	LIC4H2	1518	3	LIC4H3	1518	97.0

```
LlC4H1          ATGGATCTCCTACTCCTGGAGAAGACCCTGCTCGGCCTTTTCGTCGCCGCTGTCGTCGCC 60
LlC4H2          ATGGATCTCCTACTCCTGGAGAAGACCCTGCTCGGCCTTTTCGTCGCCGCTGTCGTCGCC 60
LlC4H3          ATGGATCTCCTACTCCTGGAGAAGACCCTGCTCGGCCTTTTCTTCGCCGCTGTCGTCGCC 60
                ************************************** ****************

LlC4H1          ATTGCTGTTTCCAAGCTCCGAGGGAAGCGCTTCAAGCTTCCGCCGGGTCCTCTTCCGGTC 120
LlC4H2          ATTGCTGTTTCCAAGCTCCGAGGGAAGCGCTTCAAGCTTCCGCCGGGTCCTCTTCCGGTC 120
LlC4H3          ATTGCCGTTTCCAAGCTCCGAGGGAAGCGCTTCAAGCTTCCGCCGGGTCCTCTTCCGGTC 120
                ***** ************************************************

LlC4H1          CCCATCTTCGGAAATTGGCTTCAGGTCGGCGACGATCTTAACCACCGGAATTTGACTGAT 180
LlC4H2          CCCATCTTCGGAAATTGGCTTCAGGTCGGCGACGATCTTAACCACCGGAATTTGACTGAT 180
LlC4H3          CCCATCTTCGGAAATTGGCTTCAGGTCGGCGATGATCTTAACCACCGGAATTTGACCGAT 180
                ******************************** ********************** ***

LlC4H1          TTAGCGAAGAAGTTTGGCGATATCTTCTTGCTCCGGATGGGCCAGAGGAACCTTCTGGTG 240
LlC4H2          TTAGCGAAGAAGTTTGGCGATATCTTCTTGCTCCGGATGGGCCAGAGGAACCTTCTGGTG 240
LlC4H3          TTAGCGAAGAAGTTTGGCGATATCTTCTTGCTTCGGATGGGCCAGAGGAACCTTCTGGTG 240
                ******************************** **************************

LlC4H1          GTTTCGTCGCCGGAGTTGGCAAAGGAGGTGCTGCACACGCAGGGGGTGGAGTTCGGATCC 300
LlC4H2          GTTTCGTCGCCGGAGTTGGCAAAGGAGGTGCTGCACACCCAGGGGGTGGAGTTCGGATCC 300
LlC4H3          GTTTCGTCGCCGGAGTTGGCAAAGGAGGTGCTGCACACGCAGGGAGTGGAGTTCGGATCC 300
                ************************************ ***** ***************
```

```
LlC4H1          AGGACTCGGAATGTGGTGTTCGACATCTTCACCGGAAAAGGTCAGGATATGGTGTTTACG 360
LlC4H2          AGGACTCGGAATGTGGTGTTCGACATCTTCACCGGAAAAGGTCAGGATATGGTGTTTACG 360
LlC4H3          AGGACTCGGAATGTGGTGTTCGACATCTTCACTGGAAAAGGTCAGGATATGGTGTTTACG 360
                ******************************** ************************** 

LlC4H1          GTGTACGGAGAGCACTGGAGGAAGATGCGGCGGATCATGACGGTGCCGTTTTTCACGAAC 420
LlC4H2          GTGTACGGAGAGCACTGGAGGAAGATGCGGCGGATCATGACGGTGCCGTTTTTCACGAAC 420
LlC4H3          GTGTACGGGGAGCACTGGAGGAAGATGCGGCGGATCATGACGGTGCCGTTTTTCACGAAC 420
                ******** ************************************************** 

LlC4H1          AAGGTGGTTCAGCAATACAGGGATGGATGGGAGAAAGAGGTGGCGAGCGTGGTGGAGGAC 480
LlC4H2          AAGGTGGTTCAGCAATACAGGGATGGATGGGAGAAAGAGGTGGCGAGCGTGGTGGAGGAC 480
LlC4H3          AAGGTGGTTCAGCAATACAGGGATGGATGGGAGAAAGAGGCGGCGAGCGTGGTGGAGGAC 480
                ************************************** ******************** 

LlC4H1          GTGAAGAAGAATCCCGAGTCGGCGACCAATGGGATCGTGCTGAGGAGGCGGTTACAGCTG 540
LlC4H2          GTGAAGAAGAATCCCGAGTCGGCGACCAATGGGATCGTGGTGAGGAGGCGGTTACAGCTG 540
LlC4H3          GTGAAGAAGAATCCCGAGTCGGCGACCAATGGGATCGTGCTGAGGAGGCGGTTACAGCTG 540
                ************************************** ******************** 

LlC4H1          ATGATGTACAACAACATGTACAGAATCATGTTCGATAGAAGGTTCGAGAGCGAGGATGAT 600
LlC4H2          ATGATGTACAACAACATGTACAGAATCATGTTCGATAGAAGGTTCGAGAGCGAGGATGAT 600
LlC4H3          ATGATGTACAACAACATGTACAGAATCATGTTCGATAGAAGGTTCGACAGCGAGGATGAT 600
                ********************************************** ************ 

LlC4H1          CCACTTTTCCAGAGGCTGAAGTCTCTGAACGGAGAGAGGAGTCGTCTGGCTCAGAGCTTT 660
LlC4H2          CCACTTTTCCAGAGGCTGAAGTCTCTGAACGGAGAGAGGAGTCGTCTGGCTCAGAGCTTT 660
LlC4H3          CCTCTTTTCCAGAGGCTGAGGTCTCTGAACGGAGAGAGGAGCCGTCTGGCTCAGAGCTTT 660
                ** ***************** ********************** ****************** 

LlC4H1          GAGTATAATTATGGCGATTTTATTCCCATTCTCAGGCCATTCTTGCGAGGATACCTGAAA 720
LlC4H2          GAGTATAATTATGGCGATTTTATTCCCATTCTCAGGCCATTCTTGCGAGGATACCTGAAA 720
LlC4H3          GAGTATAATTATGGCGATTTTATTCCCATTCTCAGGCCTTTCTTGCGAGGATACCTGAAA 720
                ************************************ ********************** 

LlC4H1          ATCTGTAAGGAAGTTAAGGAGACAAGGTTGAAACTCTTCAAAGACTACTTCGTTGATGAG 780
LlC4H2          ATCTGTAAGGAAGTTAAGGAGACAAGGTTGAAACTCTTCAAAGACTACTTCGTTGATGAG 780
LlC4H3          ATCTGTAAGGAAGTTAAGGAGACAAGGTTGAAACTCTTCAAAGACTACTTCGTTGATGAG 780
                ************************************************************ 

LlC4H1          AGGAAGAAACTGTCAAGCACTAGGACTTCAAGCAACGGGGAATTAAAATGTGCGATTGAC 840
LlC4H2          AGGAAGAAACTGTCAAGCACTAGGACTTCAAGCAACGGGGAATTAAAATGTGCGATTGAC 840
LlC4H3          AGGAAGAAACTGTCAAGCACTAGGACTTCAAGCAACGGGGAATTAAAATGTGCTATTGAC 840
                ******************************************************** ****** 

LlC4H1          CACATCTTGGACGCCCAGAAAAAGGGCGAGATCAACGAGGACAATGTACTCTACATTGTA 900
LlC4H2          CACATCTTGGACGCCCAGAAAAAGGGCGAGATCAACGAGGACAATGTACTCTACATTGTA 900
LlC4H3          CACATCTTGGACGCCCAGAAAAAGGGCGAGATCAACGAGGACAATGTACTCTACATTGTA 900
                ************************************************************ 

LlC4H1          GAGAACATCAACGTTGCCGCAATCGAGACAACCCTCTGGTCGATTGAGTGGGGTATCGCC 960
LlC4H2          GAGAACATCAACGTTGCCGCAATCGAGACAACCCTCTGGTCGATTGAGTGGGGTATCGCC 960
LlC4H3          GAGAACATTAACGTCGCCGCAATCGAGACAACCCTCTGGTCAATTGAGTGGGGTATCGCC 960
                ******** ***** ************************** ****************** 

LlC4H1          GAGCTGGTGAACCACCCAGAGATTCAAAAGAAGCTCCGCGATGAGATCGACACCGTCCTG 1020
LlC4H2          GAGCTGGTGAACCACCCAGAGATTCAAAAGAAGCTCCGCGATGAGATCGACACCGTCCTG 1020
LlC4H3          GAGCTGGTGAACCACCCAGAGATTCAAAAGAAGCTCCGCGATGAGATCGACACCGTCCTG 1020
                ************************************************************ 

LlC4H1          GGACCCGGCCACCAGGTGACCGAGTCAGACACCCACAAGCTCCCATACCTACAGGCCGTG 1080
LlC4H2          GGACCCGGCCACCAGGTGACCGAGCCAGACACCCACAAGCTCCCATACCTACAGGCCGTG 1080
LlC4H3          GGACCTGGCCACCAGGTGACCGAGCCAGACACCCACAAGCTCCCATACCTACAGGCCGTG 1080
                ***** ****************** ********************************** 

LlC4H1          GTCAAGGAGACCCTTCGTCTGCGAATGGCCATCCCGCTCCTGGTCCCCCACATGAACCTC 1140
LlC4H2          GTCAAGGAGACCCTTCGTCTGCGAATGGCCATCCCGCTCCTGGTCCCCCACATGAACCTC 1140
LlC4H3          ATCAAGGAGACACTTCGCCTGCGAATGGCCATCCCACTCCTGGTCCCCCACATGAACCTC 1140
                 ********** ***** ****************** ************************ 



LlC4H1          CACGACGCCCAACTCGGCGGTTACGACATCCCCGCCGAGAGCAAGATCCTCGTTAACGCA 1200
LlC4H2          CACGACGCCAAGCTCGGCGGTTACGACATCCCCGCCGAGAGCAAGATCCTCGTTAACGCA 1200
LlC4H3          CACGACGCCAAGCTCGGCGGCTACGACATCCCCGCCGAGAGCAAGATCCTAGTCAACGCA 1200
                ********* * ******** ***************************** ** ****** 
```

```
LlC4H1          TGGTGGCTCGCCAACAACCCCGCTCACTGGAAGAACCCGGAGCAATTCCGACCCGAGAGA 1260
LlC4H2          TGGTGGCTCGCCAACAACCCCGCTCACTGGAAGAACCCGGAGCAATTCCGACCCGAGAGA 1260
LlC4H3          TGGTGGCTCGCCAACAACCCCGCTCACTGGAAGAACCCGGAGCAATTCCGACCCGAGAGA 1260
                ************************************************************

LlC4H1          TTCCTCGAGGAGGAATCAAAGGTTGATGCCAACGGGAACGACTTCAGGTACCTTCCTTTT 1320
LlC4H2          TTCCTCGAGGAGGAATCAAAGGTTGATGCCAACGGGAACGACTTCAGGTACCTTCCTTTT 1320
LlC4H3          TTCCTCGAGGAGGAATCAAAGGTTGATGCCAACGGGAACGACTTCAGGTACCTTCCTTTT 1320
                ************************************************************

LlC4H1          GGTGTTGGGAGGAGGAGCTGCCCTGGAATTATTCTGGCACTGCCAATTCTTGGAATTACG 1380
LlC4H2          GGTGTTGGGAGGAGGAGCTGCCCTGGAATTATTCTGGCACTGCCAATTCTTGGAATTACG 1380
LlC4H3          GGTGTTGGGAGGAGGAGCTGCCCTGGAATTATTCTGGCCCTGCCAATTCTTGGAATTACG 1380
                ************************************ ********************

LlC4H1          TTGGGACGTTTGGTTCAGAACTTTGAGCTGTTGCCTCCTCCAGGACACTCCAAGCTCGAC 1440
LlC4H2          TTGGGACGTTTGGTTCAGAACTTTGAGCTGTTGCCTCCTCCAGGACACTCCAAGCTCGAC 1440
LlC4H3          TTGGGACGTTTGGTTCAGAACTTTGAGCTGTTGCCTCCTCCCGGACAGTCCAAGCTCGAC 1440
                ****************************************** ***** ************

LlC4H1          ACCACTGAGAAAGGAGGCCAGTTTAGTTTGCACATATTGAAGCATTCCACCATTGTTGCC 1500
LlC4H2          ACCACTGAGAAAGGAGGCCAGTTTAGTTTGCACATATTGAAGCATTCCACCATTGTTGCC 1500
LlC4H3          ACCACTGAGAAAGGAGGCCAGTTTAGTTTGCACATATTGAAGCATTCCACCATTGTTGCC 1500
                ************************************************************

LlC4H1          AAGCCAAGGTCCTTCTAG 1518
LlC4H2          AAGCCAAGGTCCTTCTAG 1518
LlC4H3          AAACCTAGGTCCTTCTAG 1518
                ** ** ************
```

Fig. 3.7 Alignment of *LlC4H* sequences at nucleotide level

The score table and Clustal 2.1 sequence alignment of LlC4Hs at amino acid level has been shown in Table 3.3 and Fig. 3.8, respectively.

Table 3.3 Alignment score table of the three isoforms at amino acid level

SeqA	Name	Length	SeqB	Name	Length	Score
1	LlC4H1	505	2	LlC4H2	505	99.0
1	LlC4H1	505	3	LlC4H3	505	98.0
2	LlC4H2	505	3	LlC4H3	505	98.0

```
LlC4H1          MDLLLLEKTLLGLFVAAVVAIAVSKLRGKRFKLPPGPLPVPIFGNWLQVGDDLNHRNLTD 60
LlC4H2          MDLLLLEKTLLGLFVAAVVAIAVSKLRGKRFKLPPGPLPVPIFGNWLQVGDDLNHRNLTD 60
LlC4H3          MDLLLLEKTLLGLFFAAVVAIAVSKLRGKRFKLPPGPLPVPIFGNWLQVGDDLNHRNLTD 60
                **************.*********************************************

LlC4H1          LAKKFGDIFLLRMGQRNLLVVSSPELAKEVLHTQGVEFGSRTRNVVFDIFTGKGQDMVFT 120
LlC4H2          LAKKFGDIFLLRMGQRNLLVVSSPELAKEVLHTQGVEFGSRTRNVVFDIFTGKGQDMVFT 120
LlC4H3          LAKKFGDIFLLRMGQRNLLVVSSPELAKEVLHTQGVEFGSRTRNVVFDIFTGKGQDMVFT 120
                ************************************************************
```

```
LlC4H1          VYGEHWRKMRRIMTVPFFTNKVVQQYRDGWEKEVASVVEDVKKNPESATNGIVLRRRLQL 180
LlC4H2          VYGEHWRKMRRIMTVPFFTNKVVQQYRDGWEKEVASVVEDVKKNPESATNGIVVRRRLQL 180
LlC4H3          VYGEHWRKMRRIMTVPFFTNKVVQQYRDGWEKEAASVVEDVKKNPESATNGIVLRRRLQL 180
                ********************************.******************:******

LlC4H1          MMYNNMYRIMFDRRFESEDDPLFQRLKSLNGERSRLAQSFEYNYGDFIPILRPFLRGYLK 240
LlC4H2          MMYNNMYRIMFDRRFESEDDPLFQRLKSLNGERSRLAQSFEYNYGDFIPILRPFLRGYLK 240
LlC4H3          MMYNNMYRIMFDRRFDSEDDPLFQRLRSLNGERSRLAQSFEYNYGDFIPILRPFLRGYLK 240
                ***************:**********:********************************

LlC4H1          ICKEVKETRLKLFKDYFVDERKKLSSTRTSSNGELKCAIDHILDAQKKGEINEDNVLYIV 300
LlC4H2          ICKEVKETRLKLFKDYFVDERKKLSSTRTSSNGELKCAIDHILDAQKKGEINEDNVLYIV 300
LlC4H3          ICKEVKETRLKLFKDYFVDERKKLSSTRTSSNGELKCAIDHILDAQKKGEINEDNVLYIV 300
                ************************************************************

LlC4H1          ENINVAAIETTLWSIEWGIAELVNHPEIQKKLRDEIDTVLGPGHQVTESDTHKLPYLQAV 360
LlC4H2          ENINVAAIETTLWSIEWGIAELVNHPEIQKKLRDEIDTVLGPGHQVTEPDTHKLPYLQAV 360
LlC4H3          ENINVAAIETTLWSIEWGIAELVNHPEIQKKLRDEIDTVLGPGHQVTEPDTHKLPYLQAV 360
                ************************************************.***********

LlC4H1          VKETLRLRMAIPLLVPHMNLHDAQLGGYDIPAESKILVNAWWLANNPAHWKNPEQFRPER 420
LlC4H2          VKETLRLRMAIPLLVPHMNLHDAKLGGYDIPAESKILVNAWWLANNPAHWKNPEQFRPER 420
LlC4H3          IKETLRLRMAIPLLVPHMNLHDAKLGGYDIPAESKILVNAWWLANNPAHWKNPEQFRPER 420
                :**********************:***********************************

LlC4H1          FLEEESKVDANGNDFRYLPFGVGRRSCPGIILALPILGITLGRLVQNFELLPPPGHSKLD 480
LlC4H2          FLEEESKVDANGNDFRYLPFGVGRRSCPGIILALPILGITLGRLVQNFELLPPPGHSKLD 480
LlC4H3          FLEEESKVDANGNDFRYLPFGVGRRSCPGIILALPILGITLGRLVQNFELLPPPGQSKLD 480
                ************************************************************:****

LlC4H1          TTEKGGQFSLHILKHSTIVAKPRSF 505
LlC4H2          TTEKGGQFSLHILKHSTIVAKPRSF 505
LlC4H3          TTEKGGQFSLHILKHSTIVAKPRSF 505
                *************************
```

Fig. 3.8 Alignment of LIC4H sequences at amino acid level

3.3.7 Bioinformatic characterization of LIC4Hs

3.3.7.1 BLAST search and prediction of conserved domains

Since all the isoforms are more than 98% identical at amino acid level, all the bioinformatic characterization of LIC4Hs have been done taking LIC4H1 as representative of all LIC4Hs. Blast search of LIC4H1 sequence against the known C4Hs from other plants showed that LIC4H1 share maximum identity of 91% with C4H from *Glycine max* (a Fabaceae family plant) followed by *Nicotiana tabacum* and other plants (84-90%). Blast result also showed presence of putative conserved domains in the sequence predicting the protein to be a member of p450 superfamily (Fig. 3.9).

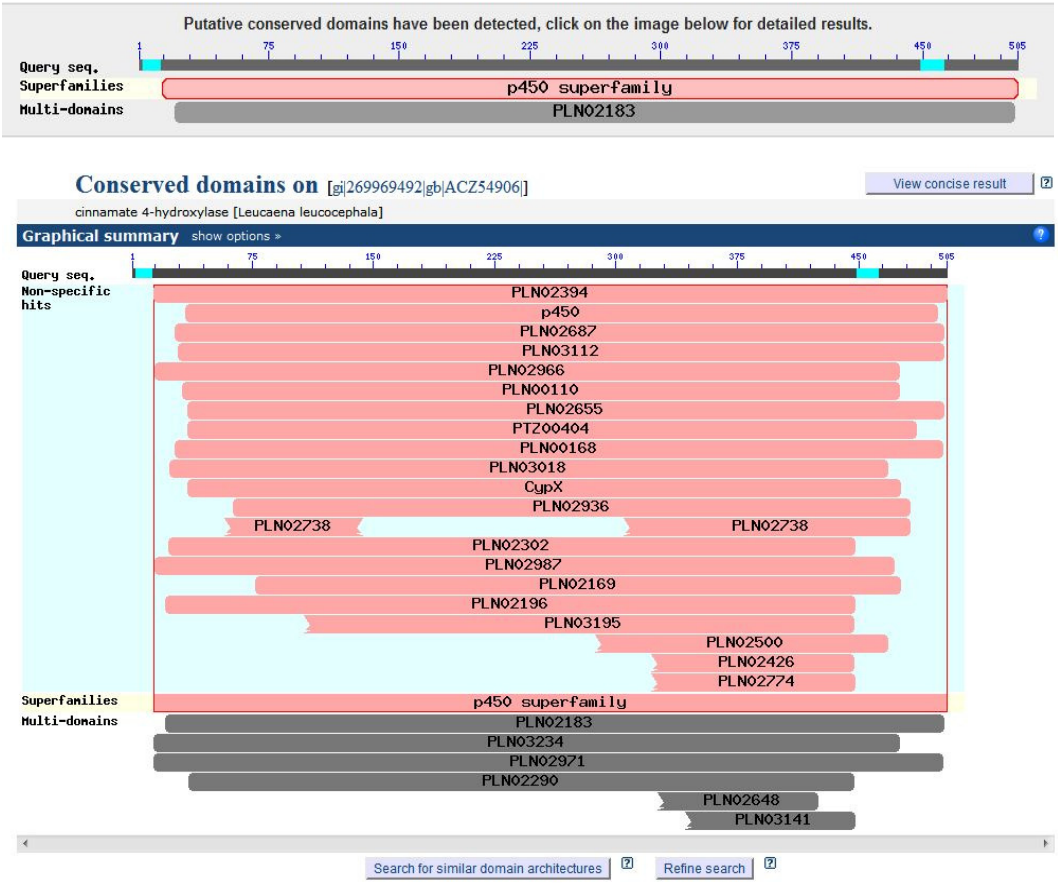


Fig. 3.9 Blast search of LIC4H1 showing the presence of putative conserved domains in the sequence

3.3.7.2 Codon usage and GC content of *LIC4H*1 sequence

Codon usage and GC content of *LIC4H*1 was determined (Fig. 3.10) using online tool available at www.justbio.com. Codon usage has been expressed as% of total codons.

Codon Usage Means in whole sequences

Expressed as % of total codons

genetic code: Standard Genetic Code

codon	mean	codon	mean	codon	mean	codon	mean
UUU ( )	1.58	UCU ( )	0.20	UAU ( )	0.40	UGU ( )	0.40
UUC ( )	3.56	UCC ( )	0.99	UAC ( )	1.98	UGC ( )	0.20
UUA ( )	0.59	UCA ( )	0.79	UAA ( )	0.00	UGA ( )	0.00
UUG ( )	2.17	UCG ( )	0.79	UAG ( )	0.20	UGG ( )	1.58
CUU ( )	1.98	CCU ( )	0.99	CAU ( )	0.20	CGU ( )	0.59
CUC ( )	3.56	CCC ( )	1.58	CAC ( )	2.37	CGC ( )	0.40
CUA ( )	0.40	CCA ( )	1.38	CAA ( )	0.79	CGA ( )	0.79
CUG ( )	3.56	CCG ( )	1.38	CAG ( )	2.57	CGG ( )	1.19
AUU ( )	2.37	ACU ( )	0.99	AAU ( )	1.38	AGU ( )	0.40
AUC ( )	3.16	ACC ( )	1.98	AAC ( )	3.75	AGC ( )	1.38
AUA ( )	0.20	ACA ( )	0.40	AAA ( )	1.78	AGA ( )	0.59
AUG ( )	2.17	ACG ( )	0.99	AAG ( )	5.14	AGG ( )	3.36
GUU ( )	2.17	GCU ( )	0.79	GAU ( )	2.57	GGU ( )	0.99
GUC ( )	1.58	GCC ( )	2.37	GAC ( )	2.57	GGC ( )	1.78
GUA ( )	0.40	GCA ( )	0.79	GAA ( )	0.59	GGA ( )	2.57
GUG ( )	3.36	GCG ( )	0.79	GAG ( )	6.32	GGG ( )	1.19

Fig. 3.10 Codon usage of *LIC4H*1 gene expressed as percentage of all codons

3.3.7.3 *In silico* restriction analysis of *LIC4H*1 and g*LIC4H*3

Nucleotide sequences of a cDNA clone (*LIC4H*1, Fig. 3.11) and a gDNA clone of *C4H* (g*LIC4H*3, Fig. 3.12) was carried out using free online software NEBcutter V2.0 (http://tools.neb.com/NEBcutter2/). The analysis was limited to type II and some type III restriction enzymes commercially available and supplied by NEB. Type I and rest type III restriction enzymes, endonucleases and nicking enzymes were not considered for the analysis. Restriction enzymes cutting the nucleotide sequence more than once were also not considered for display here. In silico restriction digestion is useful in designing strategy for cloning and Southern hybridization experiments.

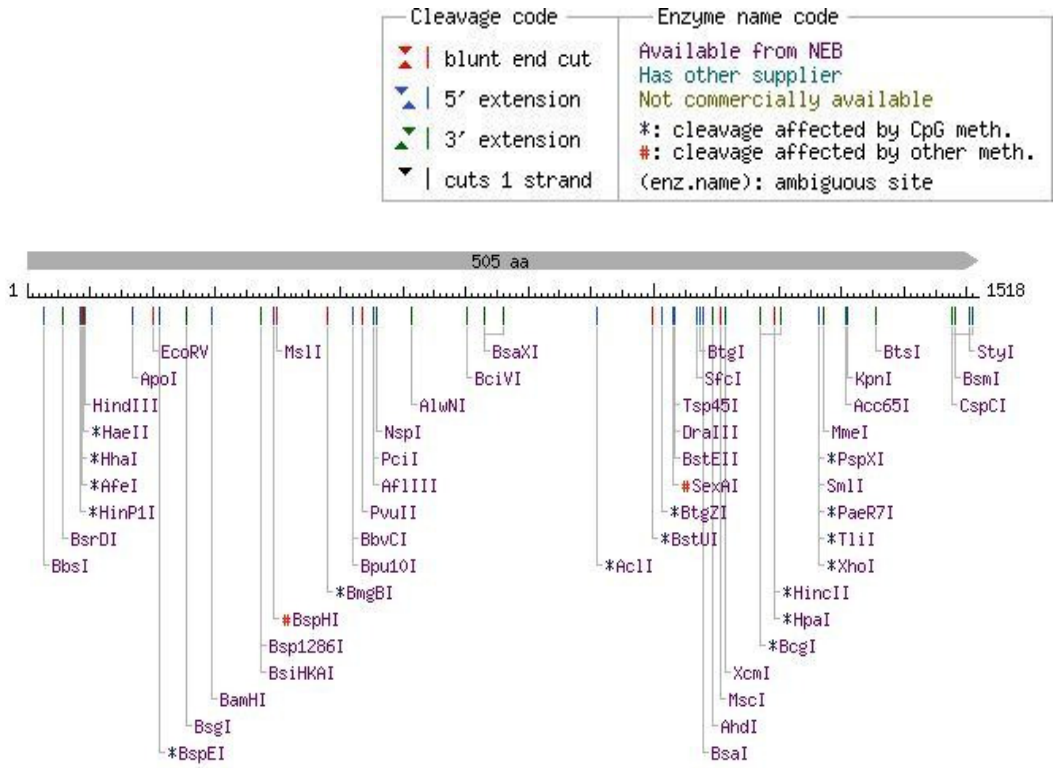


Fig. 3.11 *In silico* restriction analysis of *LIC4H*1. Only single cutter restriction enzymes supplied by NEB are on display

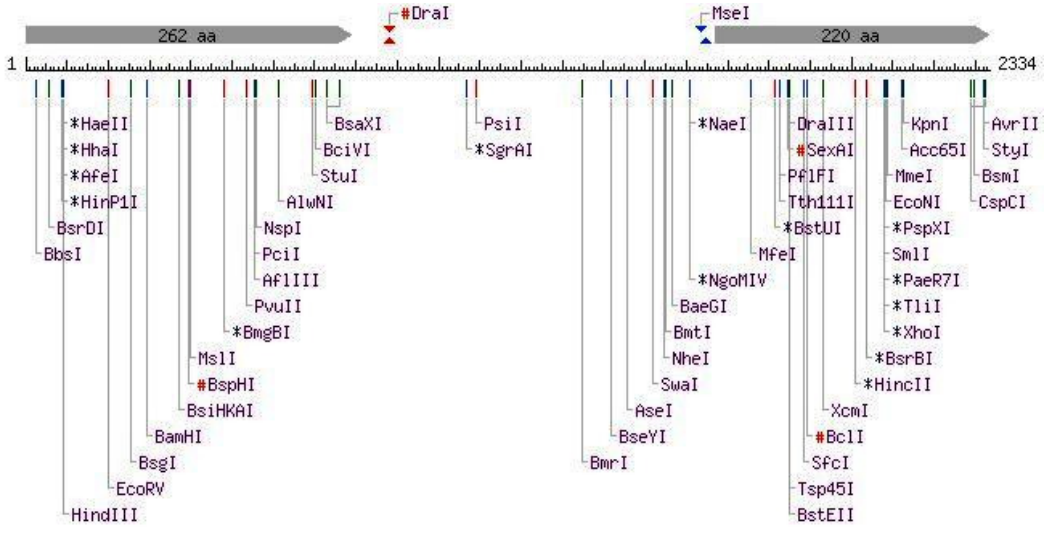


Fig. 3.12 *In silico* restriction analysis of g*LIC4H*3. Only single cutter restriction enzymes supplied by NEB are on display

3.3.7.4 Physico-chemical properties and amino acid composition of LIC4H1

Physico-chemical properties of LIC4H1 as predicted using ExPASy proteomic tools are given below and the amino acid composition of LIC4H is given in Table 3.4.

- Number of amino acids: 505
- Molecular weight: 57939.1
- Theoretical pI: 9.18
- Total number of negatively charged residues (Asp + Glu): 61
- Total number of positively charged residues (Arg + Lys): 70
- Formula: $C_{2626}H_{4156}N_{722}O_{728}S_{14}$
- Atomic composition:

Carbon	C	2626
Hydrogen	H	4156
Nitrogen	N	722
Oxygen	O	728
Sulfur	S	14

- Total number of atoms: 8246
- Instability Index (II): 42.26 (classifies the protein as unstable)
- The estimated half-life is:

  30 hours (mammalian reticulocytes, *in vitro*).

  >20 hours (yeast, *in vivo*).

  >10 hours (*Escherichia coli*, *in vivo*).

- Extinction coefficients (M^{-1} cm^{-1}, at 280 nm measured in water):

  Ext. coefficient 62005
  Abs 0.1% (=1 g/l) 1.070, assuming all pairs of Cys residues form cystines

- Ext. coefficient 61880
  Abs 0.1% (=1 g/l) 1.068, assuming all Cys residues are reduced

- Aliphatic index: 96.85
- Grand average of hydropathicity (GRAVY): -0.282

Table 3.4 Amino acid composition of LIC4H1

Amino acid	No. of residues	Percent composition	Amino acid	No. of residues	Percent composition
Ala (A)	24	4.8%	Leu (L)	62	12.3%
Arg (R)	35	6.9%	Lys (K)	35	6.9%
Asn (N)	26	5.1%	Met (M)	11	2.2%
Asp (D)	26	5.1%	Phe (F)	26	5.1%
Cys (C)	3	0.6%	Pro (P)	27	5.3%
Gln (Q)	17	3.4%	Ser (S)	23	4.6%
Glu (E)	35	6.9%	Thr (T)	22	4.4%
Gly (G)	33	6.5%	Trp (W)	8	1.6%
His (H)	13	2.6%	Tyr (Y)	12	2.4%
Ile (I)	29	5.7%	Val (V)	38	7.5%

3.3.7.5 Hydropathy plot of LIC4H1

Kyte and Doolittle hydropathy plot of a protein is drawn to predict trans-membrane and surface exposed regions of globular protein. Depending upon the hydrophobic or hydrophilic properties of its side chain, every amino acid is given a hydrophobicity score between -4.5 to +4.5. Higher score corresponds to hydrophobic amino acids and *vice versa*. A window size is then set which is actually the number of consecutive amino acids in a protein chain, whose hydropathy values are averaged and assigned to the middle amino acid in the window. The calculation advances to the next amino acid in the first window itself and the pattern continues till the end of the protein sequence. The nature of the graph depends upon the window size. Hydropathy plot of LIC4H1 has been drawn in the window sizes of both 9 and 19. When the window size is 19, peaks with scores greater than 1.8 indicate possible trans-membrane region, whereas, when the window size is 9, strong negative peaks indicate possible surface regions of globular proteins.

Fig. 3.13 shows the Kyte and Doolittle hydropathy plot of LIC4H1 in window size 9. The surface exposed region of LIC4H1 predicted using this tool may be putative antigenic epitope while positive values may represent hydrophobic regions. The grand average of hydropathicity (GRAVY),

calculated as the sum of hydropathy values of all the amino acids, divided by the number of residues in the sequence, was predicted to be -0.282 by ExPASy server tool.

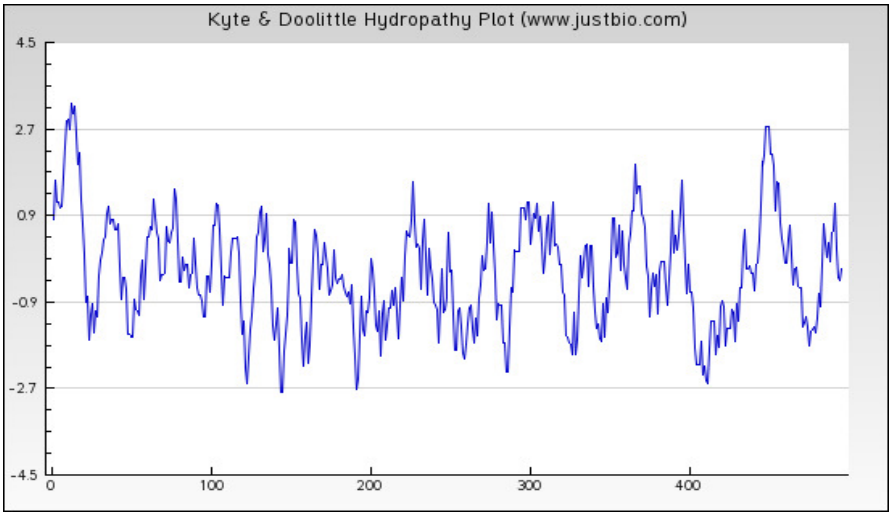


Fig. 3.13 Kyte and Doolittle hydropathy plot of LIC4H1 (window size: 9)

When the window size is 19, peaks with scores greater than 1.8 indicate possible trans-membrane regions. As evident from the plot (Fig. 3.14), LIC4H1 contains a membrane spanning region near the N-terminal of the protein. The membrane spanning region is highly hydrophobic since it has highest positive value in the plot.

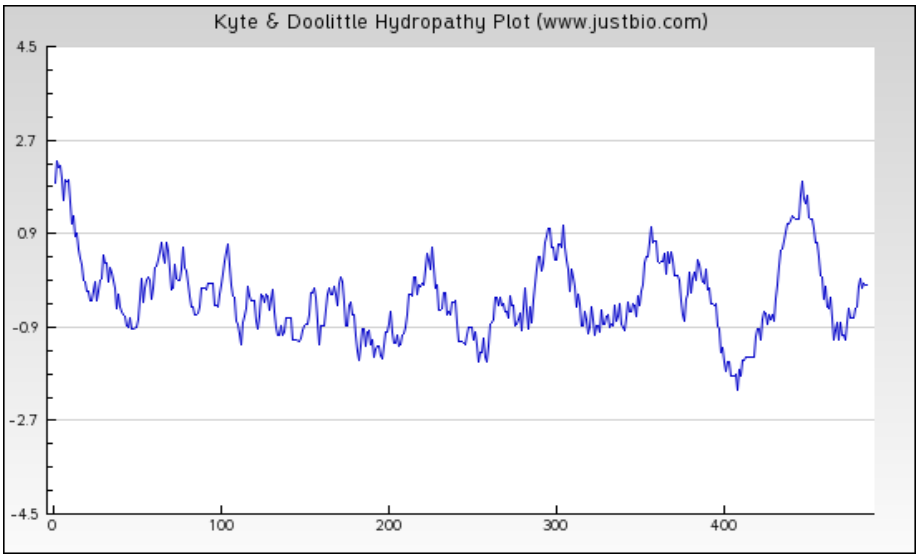


Fig. 3.14 Kyte and Doolittle hydropathy plot of LIC4H1 (window size: 19)

3.3.7.6 Phylogenetic analysis of LIC4Hs

Homology tree of LIC4Hs and other known C4Hs from other plants was generated using MEGA5 and the evolutionary distances were measured using Poisson correction method. The evolutionary relationship was inferred using neighbour joining method with 1000 replicates. The homology tree (Fig. 3.15) can be divided into three clusters: one each representing dicots and monocots' C4H and the other mixed population. LIC4Hs grouped with C4Hs from dicot plants. Sequences in this cluster share 82-99% identnty at amino acid level and belong to well characterized class I CYP73, home to majority of dicot C4Hs.

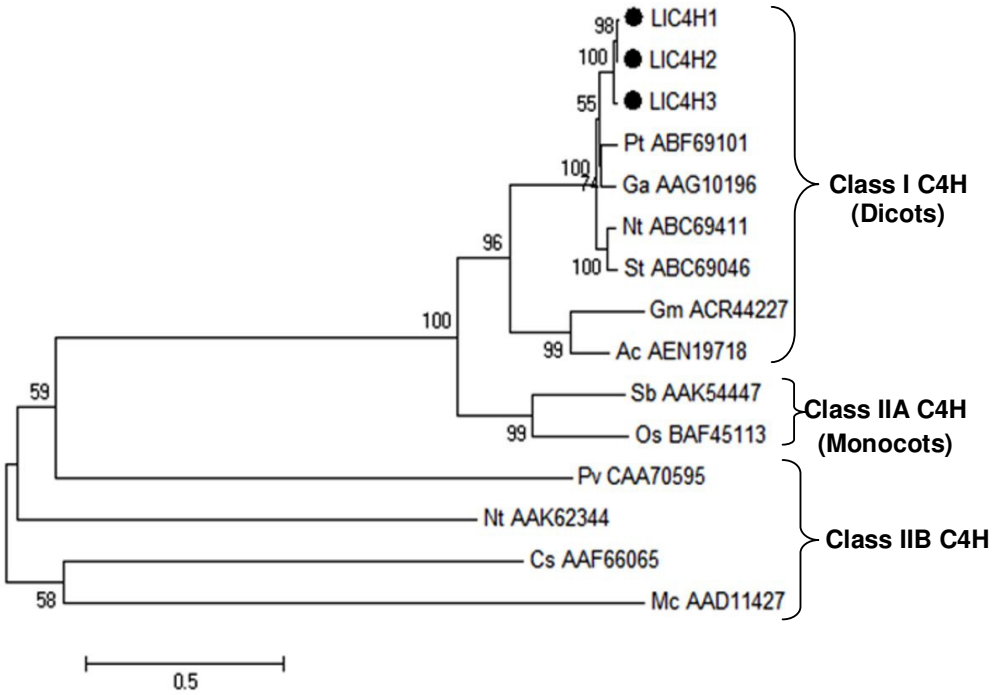


Fig. 3.15 Homology tree of LIC4Hs and other known C4Hs from various plants. Numbers at node represent the percentage of replica trees in which the associated taxa cluster together in the bootstrap test for 1000 replicates. This tree clearly classified LIC4H as Class I CYP73. Pt ABF69101 (*Populus tremuloides*), Ga AAG10196 (*Gossypium arboretum*), Nt ABC69411 (*Nicotiana tabacum*), St ABC69046 (*Solanum tuberosum*), Gm ACR44227 (*Glycine max*) and Ac AEN19718 (*Acacia auriculiformis* x *Acacia mangium*) belong to Class I CYP73. Sb AAK54447 (*Sorghum bicolour*), Os BAF45113 (*Oryza sativa*, Japonica), Pv CAA70595 (*Phaseolus vulgaris*), Nt AAK62344 (*Nicotiana tabacum*), Cs AAF66055 (*Citrus sinensis*), Mc AAD11427 (*Mesembryanthemum crystallinum*)

3.3.8 Determination of copy number of *C4H* in *L. leucocephala*

The restriction enzymes used in the hybridization experiment did not cut within the probe used for hybridization or the entire length of gene. As shown in Fig 3.16, lanes corresponding to *Eco*RI, *Sac*I and *Nde*I have 3, 3 and 2 bands respectively implying at least 3 copies of *C4H* in *Leucaena* genome. However, as the bands correspond to high molecular weight, there may be some copies of *C4H* tandemly arranged and hence detectable as single band. Copy number determination by DNA hybridization in a self-pollinating polyploid species like *Leucaena* may also under-represent the actual copy number of a gene in the genome. Considering all these things, we propose that there may be 4-6 *C4H* isoforms in *L. leucocephala*.

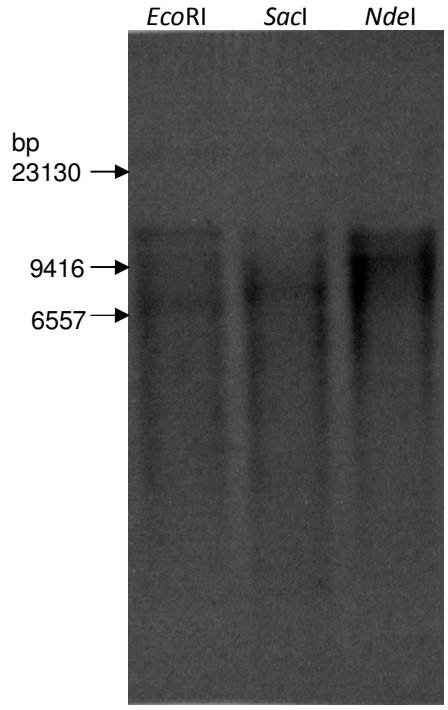


Fig. 3.16 DNA blot analysis of Ll-C4H gene in *Leucaena leucocephala*. 20 µg genomic DNA was digested with *Eco*RI, *Sac*I and *Nde*I enzymes, fractioned on 0.8% agarose gel and blotted on to Hybond-N^+ membrane. Hybridization was performed using α'-P^{32} labelled probe (~400 bp 3' terminus of the gene). *Hin*dIII digested λ-Phage DNA was used as ladder. None of the restriction enzymes cut within the probe or *C4H* gene

3.4 Discussion

PCR based approach was successfully used in our case to isolate full length of *C4H* from *L. leucocephala*. Three forms of C4H were isolated this way from *L. leucocephala* which were designated as *LlC4H*1, *LlC4H*2 and *LlC4H*3 with NCBI Genbank accession numbers GU183363, HQ191221 and HQ191222 respectively. The ORF of *C4H* in *L. leucocephala* is 1518 nucleotides long coding for 505 amino acid residues as in *J. artichoke* (Teutsch et al. 1993), mung bean (Mizutani et al. 1993), *Nicotiana tabacum* (ABC69411, ABC69413), *Populus trichocarpa* (ACC63873), *Gossypium hirsutum* (ACZ06240) etc.

C4H exists as multigene family in trees like *Populus tremuloides*, *P. trichocarpa* (Lu et al. 2006), periwinkle (Hotze et al. 1995) and members of Fabaceae family plants like pea (Whitbred and Schuler 2000) and alfalfa (Fahrendorf and Dixon 1993). Among the *C4H* isoforms in *L. leucocephala*, most of the differences at nucleotide level occurred at the third nucleotide position of codons and hence very high degree of identity was observed between the three isoforms at amino acid level with LlC4H1 and LlC4H2 being 99% identical and LlC4H3 being 98% identical to both LlC4H1 and LlC4H2. Sequences showing more than 97% sequence identity at amino acid level are considered allelic variants of the gene. Four isoforms of *C4H* isolated from *Populus tremuloides* were ≥ 97% identical to each other at amino acid level, although on the basis of nucleotide sequence comparison they could be divided into two groups (*PtreC4H1* and *PtreC4H2*) of two isoforms each (Lu et al. 2006). Three isoforms of *C4H* highly similar in their coding regions as well as promoter regions and introns varying only at some places caused by deletion or insertion of nucleotides exist in hybrid aspen (Kawai et al. 1996). Another member of Fabaceae family, pea was also found to have only eight nucleotide difference over two *C4H* partial sequences (accession no. AF175276 and AF175277) of 1 kb each (Whitbred and Schuler 2000). These results support our finding that the three isoforms reported are actually isoforms of *C4H* and not *Taq* polymerase generated errors in PCR or sequencing errors. These little variations in *C4H* isoforms in *L. leucocephala* may be due to the presence of multiple highly similar *C4H*

members in the genome or because of the self pollinating polyploid (2n=4x=104, Shelton and Brewbaker 1994) nature of this species.

Despite the presence of multiple isoforms of *C4H* in several plants, the precise roles of individual isoforms are not known in every case. In case of *Populus tremuloides*, *PtreC4H2* was strongly expressed in developing xylem, weakly in phloem and undetectable in mature leaf tissues whereas *PtreC41* was most abundant in developing xylem followed by phloem and mature leaf tissues (Lu et al. 2006). In case of *L. leucocephala*, the sequences were so much similar at nucleotide level that it did not allow us to analyze their expression individually.

Sequence comparison between *LlC4H3* and *gLlC4H*3 indicated that *gLlC4H*3 consists of three exons of 785, 134 and 599 nucleotides and two introns of 154 and 662 nucleotides, respectively amounting to 2,334 nucleotides. This pattern of intron location is conserved among the *C4H* gDNA clones isolated so far from different plants (Bell-lelong et al. 1997; Kawai et al. 1996; Mizutani et al. 1997), albeit the sizes of genes in different plants vary owing to different intron lengths. Both the intron splice junctions in *gLlC4H*3 obeyed the standard GT-AG rule.

Theoretical translation of *LlC4H* isoforms' nucleotide sequence yielded highly identical proteins of 505 amino acids which possessed all the features of a typical P450 protein and conserved sequences found in all the C4Hs isolated from different plants. The N-terminal hydrophobic transmembrane helix is followed by a proline rich region (consensus sequence (P/I)PGPx(G/P)xP) supposed to destabilize the α-helix and produce kink to optimally orient the enzyme with regard to the membrane (Yamazaki et al. 1993). Heme binding motif, PFGVGRRSCPG conserved in C4Hs near the C-terminus (the consensus sequence in P450s is PFGXGRRXCXG) and helices like I helix (AAIETT), K helix (ETLR) and K' helix (AWWLANN) (Durst and Nelson 1995) were also found in *L. leucocephala* C4Hs. However, there is a slight deviation in all the three isoforms of LlC4H from the consensus sequence PEEFRPERF found in most C4Hs in that E415 was replaced by Q415. Unusual residue in this consensus sequence has also been reported

from *Ammi majus* (E415D, AAO62904) (Hübner et al. 2003), *Ruta graveolens* (E414Q and E415V, AAN63028) (Gravot et al. 2004), ACJ37399 (E415V), BAJ17666 (P413A, E414N), ADG43134 (P413L) and AAC41660 (E415D). In addition, LlC4Hs also possess the residues N302 and I371 supposed to form hydrogen bond and hydrophobic interaction with the anionic site or the aromatic ring of cinnamic acid, and K484 that is required to spatially orient the substrate during or after the reaction (Schoch et al. 2003).

Construction of phylogenetic tree of LlC4H protein sequences with the reported C4H sequences from other plants that contained C4H representatives from dicots, monocots, trees and shrubs can be divided into three major clusters. First cluster is formed by a group of C4Hs belonging to dicot plants. Majority of dicot C4H share more than 82% sequence identity and belong to a functionally well characterized class I C4H. LlC4Hs belong to this class of C4H. The other two clusters are formed by class IIA and class IIB C4H, which are about 60% identical to class I C4Hs and 72-95% identical to each other (Ehlting et al. 2006). Class IIA C4H are formed exclusively by C4H belonging to monocots, whereas class IIB C4H are formed by rest of dicot C4H. Monocots contain only class IIA C4H and not class I C4H, whereas dicots must contain at least one class I C4H, additionally they may contain class IIB C4H (Ehlting et al. 2006). Among the sequences taken for phylogenetic tree generation, LlC4Hs showed closer evolutionary lineage to C4H from trees/shrub like *Populus tremuloides* and *Gossypium arboreum* than *Glycine max*, a leguminous plant, emphasizing its proximity to trees in the hierarchal organization.

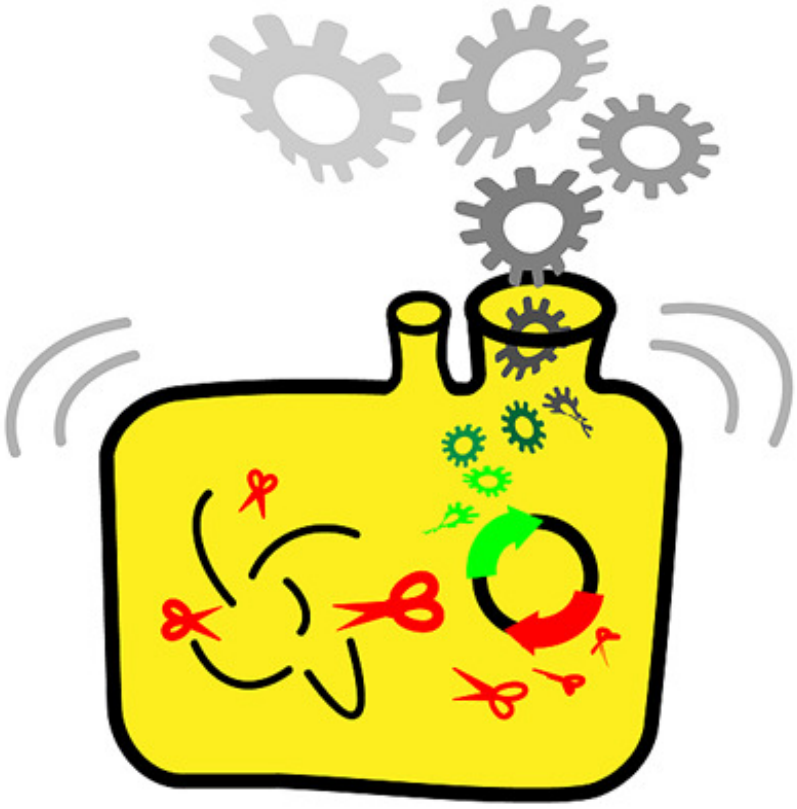
~400 bp long radioactive probe was used in the DNA hybridization experiment. Since the probe size was small, restriction enzymes not cutting within the probe were selected for digesting genomic DNA isolated from *Leucaena*. DNA hybridization study suggested that there may be 3 *C4H* isoforms present in the genome of *Leucaena*. However, as the bands correspond to high molecular weight, some copies of *C4H* may be tandemly arranged there and hence detectable as single band. Another Fabaceae member, pea also contains two highly identical *C4H* alleles tandemly linked

on a 10 kb *Eco*RI digested fragment (Frank et al. 1996; Whitbred and Schuler 2000). Copy number determination by DNA hybridization in a self-pollinating polyploid species like *Leucaena* (Shelton and Brewbaker 1994) may also inadequately represent the actual copy number of a gene in the genome. On account of this, we propose that there may be 4-6 highly identical *C4H* isoforms in *L. leucocephala*. It is not known whether all the *C4H* loci code for active proteins. Some of the isoforms may have become redundant or have evolved to become pseudogene.

3.5 Conclusion

Three highly identical *C4H* isoforms were isolated from *L. leucocephala* which bear more than 82% sequence identity to C4H enzymes from other plants. These isoforms contained all the conserved sequences found in P450 superfamily and C4H from other plants. Phylogenetic analysis grouped LlC4H with class I C4Hs, some members of which are functionally well characterized. Southern hybridization study suggested that there may be 4-6 *C4H* isoforms in *Leucaena* genome with some isoforms may be tandemely arranged.

CHAPTER - 4



Heterologous Expression, Purification and Characterization of Leucaena leucocephala C4H gene

Chapter 4: Heterologous Expression, Purification and Characterization of *Leucaena leucocephala C4H* gene

This chapter describes the heterologous expression of *Leucaena leucocephala C4H* gene in a prokaryotic host, its purification and characterization at structural level. This chapter also describes all the specific techniques, materials and methods used in the experiments performed in this chapter.

4.1 Review of literature

Following C4H isolation from *Helianthus tuberosus* (Gabriac et al. 1991) and mung bean (Mizutani et al. 1993a), attempts were made to identify its cDNA sequence from plants. cDNA sequence of *C4H* were isolated from alfalfa (Fahrendorf and Dixon 1993), mung bean (Mizutani et al. 1993b) and *Helianthus tuberosus* (Teutsch et al. 1993) independently and immediately upon protein isolation and identification possessing hydroxylase activity. In the next years to come, cDNA sequences of *C4H* were reported from other plants. In the mean time, some other sequences were described belonging to P450 gene family, based on sequence similarity with previously reported sequences. Since P450 proteins are difficult to purify from plants system, need was felt to express the protein in a heterologous system (other than plant system) and then functionally characterize the protein. Heterologous expression of a P450 protein in bacteria suffers from some limitations. The most important limitation is the lack of a suitable redox partner (NADPH:cytochrome P450 reductase, CPR) in *Escherichia coli*, the most studied bacterial system. Another limitation is the translational incompatibility of the N-terminal signal sequence of P450 proteins in *E. coli*. These problems were overcome and the first report of functional expression of *C4H* in bacterial system was provided by Hotze et al. (1995), when the authors expressed cDNA sequence of *C4H* isolated from *Catharanthus roseus* in *E. coli*. The two problems were managed by fusing the redox partner (CPR) to C4H and by deleting the signal peptide while expressing the protein in *E. coli*. Δ-Amino luvulinic acid and $FeCl_3$ (as a source of iron) needs to be supplemented in the medium for the formation of porphyrin ring that forms the backbone of heme-thiolate group of P450 protein (Nielsen and Møller 2000).

An alternate of expressing C4H protein without encountering these problems is expressing C4H in a eukaryotic system, yeast. Yeast offers the advantage that proteins with signal peptide can be expressed and fusion of redox partner is not absolutely necessary. For functional expression, yeast has been used to express C4H from a variety of plants (Hübner et al. 2003; Koopmann et al. 1999; Ro and Douglas 2004; Ro et al. 2001; Yamamura et al. 2001). Although C4H has been expressed in yeast system for activity assay, expression in yeast suffers from low yield. As I wanted to raise antibodies against C4H in rabbit; I needed purified protein in mg quantity. This was possible only when C4H was expressed in *E. coli* and purified for my purpose. As discussed earlier, problems associated with C4H expression in *E. coli* could be overcome. For activity purpose, the redox partner of C4H can be supplied in the form of purified CPR from any source including human origin, as it has been successfully proved that there is sufficient sequence similarity between a plant and mammalian CPR to replace each other (Ronald 1993), although the efficiency of a CPR to transfer electrons to a particular P450 may vary. In Chapter 3, I had mentioned isolation of three isoforms of *C4H* from *L. leucocephala*. In this chapter, I am explaining the expression of the isoforms in *E. coli* and purification of one isoform from *E. coli*.

4.2 Materials and methods

4.2.1 Stock solutions

IPTG (1 M) in MB grade water and filter sterilized

Δ-amino Levulinic acid (δ-ALA) (0.5 M) in MB grade water and filter sterilized

$FeCl_3$ (0.5 M) in MB grade water and filter sterilized

Chloramphenicol (35 mg/mL) in ethanol and filter sterilized

Kanamycin (50 mg/mL) in MB grade water and filter sterilized

4.2.2 Signal peptide prediction, primer design and PCR

Signal peptide sequence was predicted using the software signalP (http://www.cbs.dtu.dk/services/SignalP/). Restriction sites (*Nde*I and *Xho*I) in the expression vector were carefully chosen and primers for amplification of C4H sequence were designed in such a way which excluded the signal

sequence from PCR amplification and did not change the reading frame of the amplified insert. The list of primer sequences to amplify the C4H sequence for expression in *E. coli* is given in table 4.1. The restriction sites in the primer sequence are shown underlined. After primer designing, conceptual translation was performed to check that the reading frame of the insert had not changed. For PCR amplification, cloned C4H sequences of different isoforms were used as template and were amplified using LA *Taq* DNA polymerase (Sigma, USA). The PCR condition followed was as follows: 95 ºC for 5 min, 35 cycles of 95 ºC for 30 s, 62 ºC for 30 s and 68 ºC for 1.3 min with the final extension at 72 ºC for 7 min. The PCR products were purified and cloned in pGEM-T Easy vector and cloned in XL-1 blue cells. Positive colonies were picked following colony PCR and plasmids were individually isolated from the colonies positive for the three isoforms.

Table 4.1 List of primers used for *C4H* expression in *E. coli*

Primer Name	Primer sequence (5' → 3')
C4H-Exp-F-1	CATATGATTGCTGTTTCCAAGCTCCGA
C4H-Esp-R-1	CTCGAGGAAGGACCTTGGCTTGGCAAC

4.2.3 Construct preparation for *C4H* expression

The isolated plasmids for the three isoforms were individually digested by the enzymes *Nde*I and *Xho*I and the digested insert fragments were gel purified. The expression vector used in study, pET-30b(+) was also digested using the same restriction enzymes and the digested vector backbone was gel purified. The digested vector backbone and the digested inserts from pGEM-T Easy vector were ligated in individual tubes for different isoforms. Ligated products were then cloned in XL-1 blue cells and the positive clones were picked after colony PCR. Plasmids from positive clones were also sequenced to confirm its identity, direction of the insert with respect to the vector backbone and to identify the theoretical translational product. The cloning/expression region of pET30-b(+) has been given in Fig. 4.1 A and B. The entire sequence of events that was followed to express *C4H* in *E. coli* has been graphically represented in Fig. 4.2.

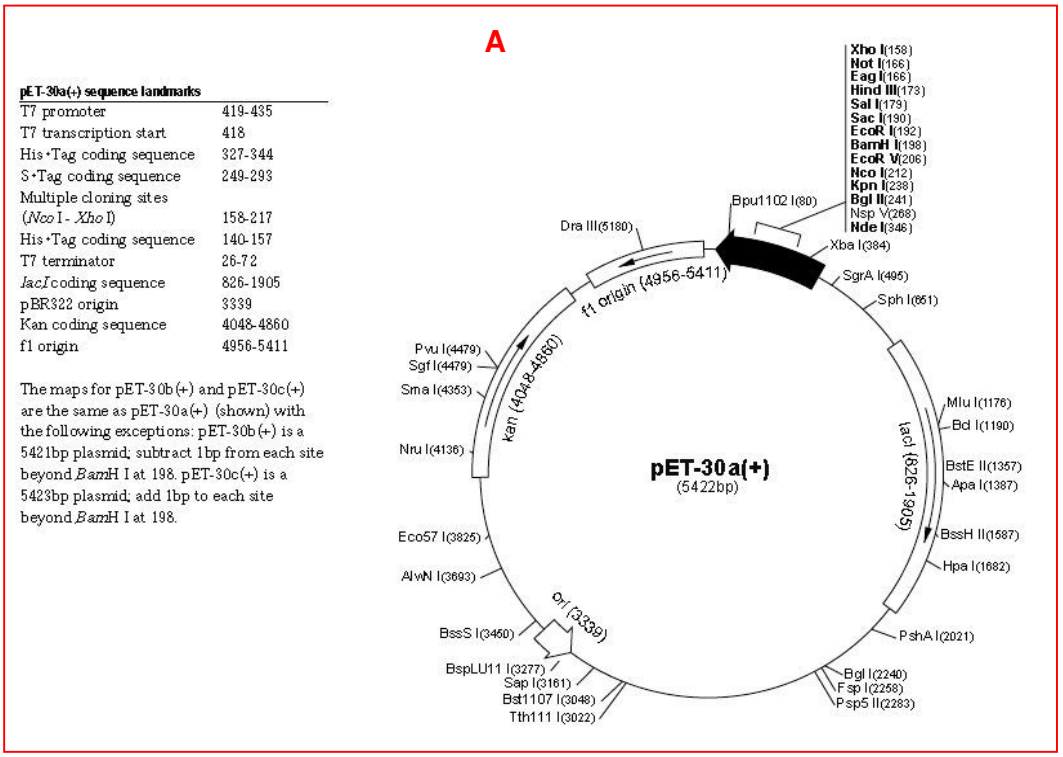


Fig. 4.1 (A) pET-30a(+) vector map and the sequence elements present on the vector. The maps for pET-30b(+) and pET-30c(+) are the same as pET-30a(+) with the following exceptions: pET-30b(+) is a 5421bp plasmid; subtract 1bp from each site beyond *Bam*HI at 198. pET-30c(+) is a 5423bp plasmid; add 1bp to each site beyond *Bam*HI at 198 (*reference*: pET-30a-c(+) vector manual, Novagen)

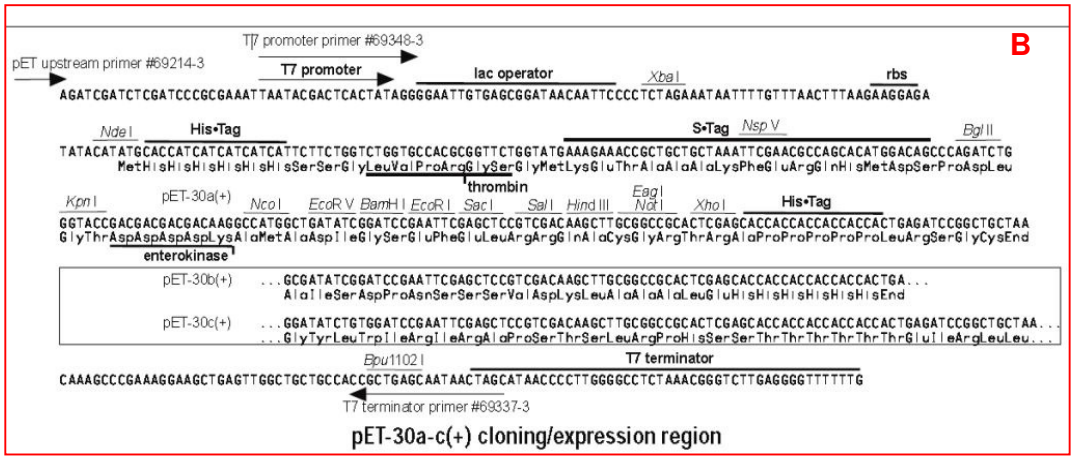


Fig. 4.1 (B) pET-30a-c(+) cloning/expression region (as adapted from pET-30a-c(+) vector manual, Novagen

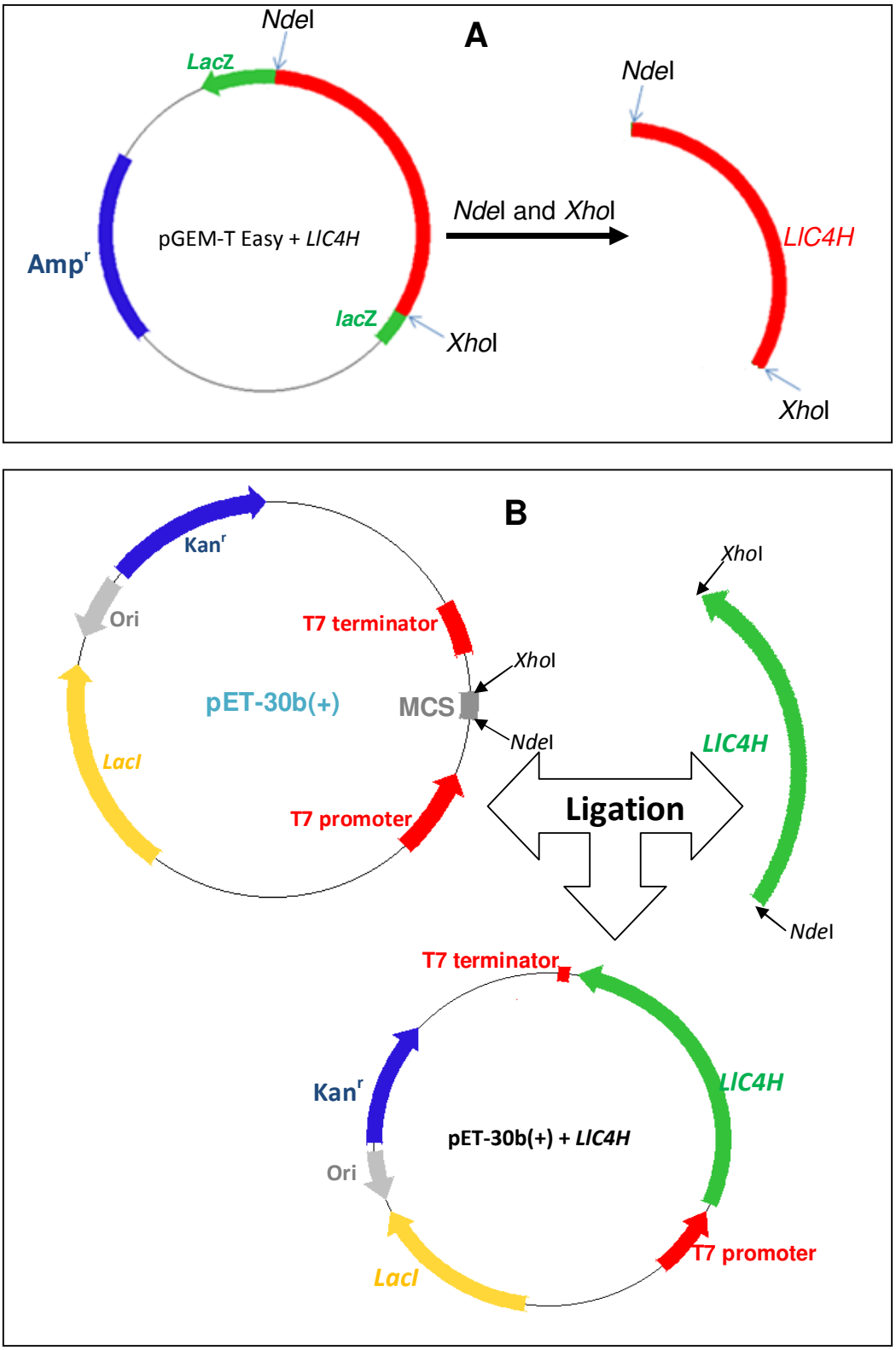


Fig. 4.2 (A) and (B) Strategy for directional cloning of *LIC4H* fragment in pET-30b(+) for expression in *E. coli*

4.2.4 Initial screening of *E. coli* strains for C4H expression

Sequenced and confirmed plasmids (expression vector containing signal sequence removed C4H isoforms in proper orientation) were then cloned in different *E. coli* strains, like BL-21(DE3), Rosetta (DE3) and Rosetta-Gami2 (DE3) for expression. Cells after transformation were picked by colony PCR and were initially screened by growing in LB broth supplemented with δ-ALA (working concentration 1 mM) and $FeCl_3$ (working concentration 0.5 mM) and induced by 1mM IPTG (working concentration). After 6 hours of induction, cells were harvested by centrifugation at 10,000Xg for 1 min. Supernatant was thrown and the pellet was washed two times in MB grade water. The cell pellet was suspended in 50-100 µL of MB grade water and denatured in the presence of loading buffer by heating at 95 ºC for 5 min and loaded on to the 10-12% SDS-PAGE gel to check for expression with protein marker. In bacteria, the over-expressed protein looks like a bulb in the gel.

4.2.5 Optimization of C4H expression in *E. coli*

Cells showing promising expression of C4H were optimized for IPTG concentration (for induction), incubation temperature (to grow the culture) and duration of incubation. After initial screening, cells were first grown in 5 mL LB medium using the colony as inoculum and grown for overnight as primary culture. Next morning, when the cells are at the same growth stage, fresh 100 mL LB medium was inoculated taking 1 mL of the primary culture as inoculum (or 1% of the primary culture as inoculum for the secondary culture). This ensures that the cells are at the same stage of growth when they are induced. Both IPTG concentration and incubation temperature were reduced for increased solubilisation of C4H and the incubation time was increased to make up for the decreased growth rate of the cells.

4.2.6 Purification protocol of C4H from inclusion bodies of *E. coli*

C4H was extracted and purified from the cell extract using the solutions mentioned in chapter 2, section 2.3.8. Cells harboring recombinant pET vector were grown overnight in 5 mL of LB Broth at 37 ºC on a shaker at 200 rpm. Secondary culture was inoculated by taking 1% primary inoculum. When the absorbance of the secondary culture was around 0.6, cells were induced

with IPTG (0.5 mM) and allowed to grow for 5 h. Cells were harvested by centrifuging at 5000Xg for 10 min at 4 ℃ and suspended in 6 mL lysis buffer. The cell suspension was sonicated six times for 30 s pulses with 10 s interval at 70% amplitude in Branson 450 sonicator. $MgCl_2$ (final concentration 10 mM) and lysozyme (final concentration 200 µg/mL) were added after sonication and kept at 4 ºC for 20 min. The suspension was centrifuged at 10,000Xg for 10 min. Pellet was resuspended in 5 mL of dispersion buffer, filtered with 0.22 µ filter to get rid of particulate matters; and stored at 4 ºC for up to one week.

The recombinant C4H protein was tagged with hexa-histidine residues at the C-terminal, which was used to purify the recombinant protein from the inclusion body (see Fig. 4.1 B). Inclusion body in bacteria is formed when identical proteins aggregate and form imperfectly ordered macroscopic entity or highly ordered fibrils. In the present study, Ni-NTA resin from QIAGEN was used to pack the column. This affinity matrix contains bound nickel ion, to which polyhistidine-tag binds with micromolar affinity (the binding of hexa-histidine tag has been found optimum for most proteins). Composition of all the buffers used in this study along with their storage conditions is given in chapter-2, section 2.3.8. The pH of all the buffers was adjusted to 8.0 every time, shortly before use. The column was packed and washed first with sterile MB grade water and then equilibrated with the binding buffer. The matrix is then loaded with the bacterial membrane fraction (solubilised, denatured and dispersed mixture of inclusion body with several other bacterial proteins) and kept in gently rocking condition for about half an hour for the proteins to bind the matrix. The column was then kept in standing position to settle the matrix and the flow-through was collected and stored later to be fractioned on gel. The column was then washed with wash buffer containing 30 mM imidazole to remove unbound or loosely bound proteins. Washing was continued till the absorbance of the washes at 280 nm became zero. The tightly bound protein was eluted in the last step using elution buffer containing 250 mM imidazole. Elutions were collected in fractions, and all the fractions and the 'flow-through' collected through the column before the washing step, were fractioned on to 10% SDS-poly-acrylamide gel and appropriately stained. The

purified fractions were quantified using Bradford reagent (Bradford 1976), pooled and stored at +4 ºC. The protocols for running/analyzing SDS-polyacrylamide gels, Coomassie brilliant blue (CBB) staining, silver staining and quantifying protein using Bradford reagent is discussed in chapter-2, sections 2.7.1, 2.7.2, 2.7.3 and 2.8 respectively.

4.2.7 Prediction of secondary and tertiary structure of C4H

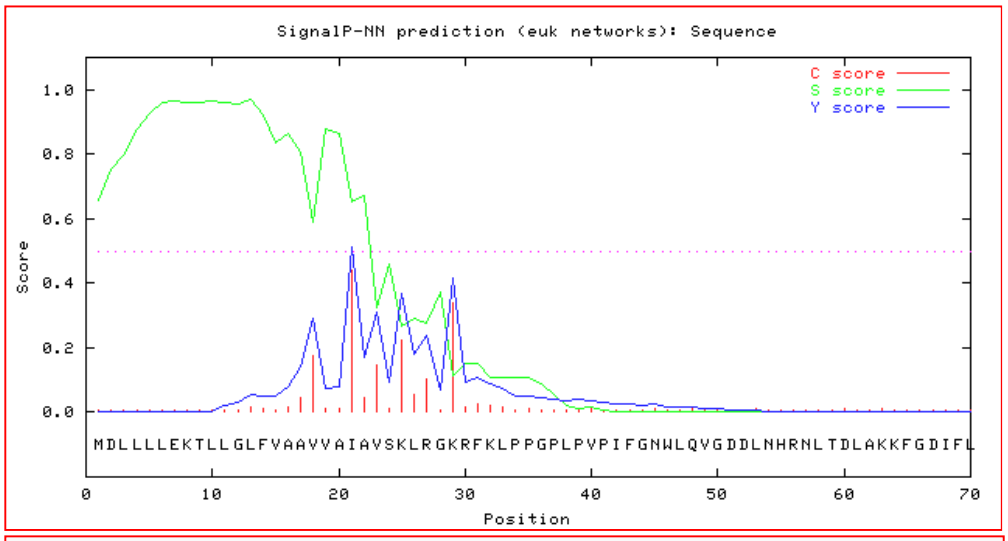
The secondary and tertiary structures were predicted by the program HNN tool and Modeller 9v9, respectively. With the aim of finding an adequate template for homology modelling of C4H, amino acid sequence of LlC4H1 was aligned against protein structures deposited in Brookhaven Protein Data Bank (PDB) using pBLAST. Crystal structures of progesterone 21-hydroxylase (PDB: 1NR6), CYPIA2 (PDB: 2HI4), mephenytoin 4-hydroxylase (PDB: 2NNJ), and CYPIIE1 (PDB: 3E6I) were used as templates (14-33% sequence identity). First 24 residues at the N-terminal of LlC4H1 were not included in model because the templates structures did not contain residues pertaining to the N-terminal membrane-binding domain. SALIGN was employed to construct multiple structure alignments of templates. The target sequence was then aligned with this multiple structure-based alignment. Based on the alignment information, the 100 structural models were built by using MODELLER 9v9 program. The coordinates of heme in the models were obtained from mephenytoin 4-hydroxylase (2NNJ) and positioned in targets as in the 2NNJ template. The model having least the discrete optimized protein energy (DOPE) score and lowest PDF energy was selected as the best model. Further correction of model in the loop regions was done using the Modeller loop refinement module. Energy minimization of final models was performed using chimera to reduce steric clashes among residues. The stereochemical quality of the structures was evaluated using PROCHECK with Ramachandran plot, the distribution of residual energy was evaluated in ProSA, and amino acid environment was assessed using Errat from the SAVES server. Additionally, a quantitative assessment of the quality of the protein structure predictions, relative to its parent structure, was done by comparing C4H and its templates using the TM-score program. The root mean square deviation (RMSD) of the C4H structure was calculated by

superimposing the template (2NNJ) by using Chimera. To view the graphical representation of the tertiary structure model of C4H, PyMol 1.3 program was used. The model was validated using Procheck program from SAVES server (http://nihserver.mbi.ucla.edu/SAVES/).

4.3 Results

4.3.1 Signal peptide prediction

Signal peptide was predicted on the basis of signalP software (Bendtsen et al. 2004). Amino acid sequences of all the three isoforms of C4H in *Leucaena leucocephala* were individually scanned for signal peptide. The result of scanning LlC4H1 has been shown in Fig. 4.3. The results of scanning the three isoforms for the signal peptide are almost similar and hence, result with only one isoform has been shown (Fig. 4.3).




```
# data


>Sequence                length = 70
# Measure   Position  Value   Cutoff   signal peptide?
  max. C     21       0.438    0.32    YES
  max. Y     21       0.514    0.33    YES
  max. S     13       0.968    0.87    YES
  mean S      1-20    0.873    0.48    YES
       D      1-20    0.693    0.43    YES
# Most likely cleavage site between pos. 20 and 21: VVA-IA
```

Fig. 4.3 Signal peptide prediction in LlC4H1 using SignalP 3.0 server

SignalP-NN prediction predicted an N-terminal signal sequence in LlC4H1 with most likely cleavage site between amino acid positions 20 and 21. By default, initially the first 70 amino acids are considered for signal peptide prediction out of a given sequence of amino acids.

4.3.2 PCR amplification of *LlC4H* genes without signal sequences

PCR amplification of the three signal sequence removed *LlC4H* genes as mentioned in section 4.2.2 gave amplification of around 1.5 kb each (Fig. 4.4), which were individually cloned in pGEM-T Easy vector, transformed in to XL-1 blue cells, positive colonies confirmed using colony PCR and plasmids isolated from them individually for the three isoforms. The plasmids were named 'pGEM-T Easy+LlC4H', followed by the clone number.

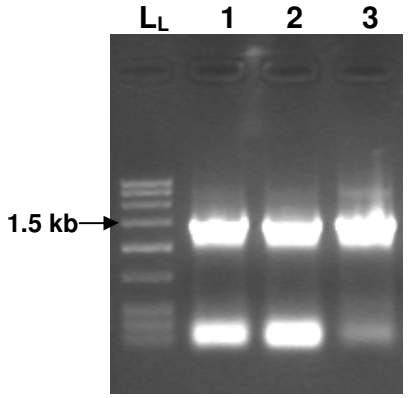


Fig. 4.4 PCR amplification of *LlC4H* genes without signal sequences. L_L= Low range ladder and lanes 1, 2 and 3 represent amplifications for LlC4H1, LlC4H2 and LlC4H3, respectively

4.3.3 Restriction digestion of 'pGEM-T Easy+LlC4Hs' plasmids and pET-30b(+) vector with *Nde*I and *Xho*I

pGEM-T Easy+LlC4H plasmids and pET-30b(+) vector were digested with *Nde*I and *Xho*I enzymes to be ligated to each other. pET-30b(+) vector was digested with these enzymes first (Fig. 4.5 A) and the vector backbone was gel eluted and stored at -20 ºC. Careful examination of *LlC4H* nucleotide sequences had revealed that all the three isoforms contained one *Xho*I restriction site near the 3' terminus of the gene (approximately 250 nucleotides upstream of stop codon). Hence, pGEM-T Easy+LlC4H plasmids

were first completely digested with *Nde*I and then partially digested with *Xho*I for 5 minutes only. Loading the digestion product on to 1.2% agarose gel and running the gel clearly showed two bands near 1.5 kb mark representing both partial C4H resulting from complete digestion; and another band comprising of complete 3' end (Fig. 4.5 B). The upper band (bigger fragment) containing complete 3' end of *LlC4H* gene was carefully cut from the gel and individually eluted. The eluted fragments were ligated to *Nde*I and *Xho*I digested pET-30b(+) and transformed in to XL-1 cells and plasmids were individually isolated from positive colonies. These plasmids after sequence confirmation through sequencing were used to transform various expression strains of *E. coli* for expression of *LlC4H*s.

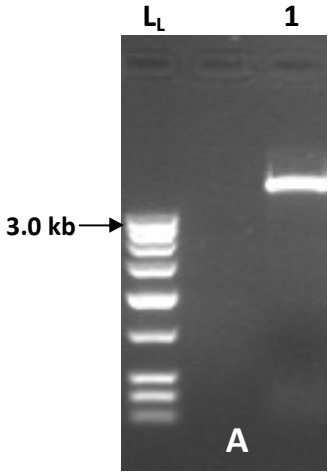


Fig. 4.5 (A) Digestion of pET-30b(+) with *Nde*I and *Xho*I. Lane L_L= Low range ladder; Lane 1= digested pET-30b(+) vector

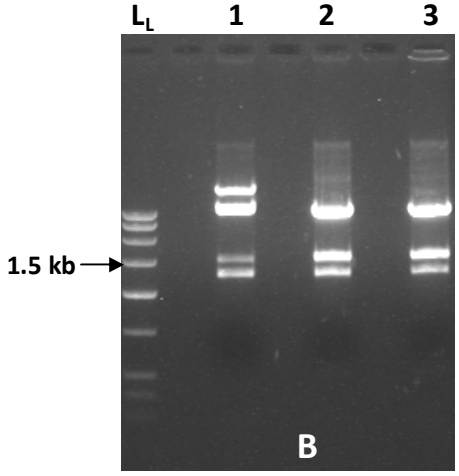


Fig. 4.5 (B) Digestion of pGEM-T Easy+*LlC4H*s with *Nde*I (complete) and then *Xho*I (partial). Lane L_L= Low range ladder; Lane 1= LlC4H1; Lane 2= LlC4H2; Lane 3= LlC4H3

4.3.4 *E. coli* strain selection for C4H expression

E. coli strains like BL-21, Rosetta (DE3) and Rosetta-gami2 (DE3) were tried to express C4H gene(s) cloned in pET-30b(+) vector. BL-21 failed to express *LlC4H* isoforms under all conditions. Only the strain Rosetta (DE3) over-expressed a protein of molecular weight near about 55 kDa, as estimated upon conceptual translation of LlC4H protein lacking the signal peptide.

Surprisingly, Rosetta-gami2 (DE3) did not express the proteins, despite carrying all the advantages of Rosetta (DE3) strain. Initially, to screen which strains expresses LIC4H, cells were lysed and the crude extract was loaded on to the gel for screening purpose as explained in section 4.2.4 of this chapter. Not all the Rosetta (DE3) colonies which were positive in colony PCR produced the recombinant protein (Fig. 4.6).

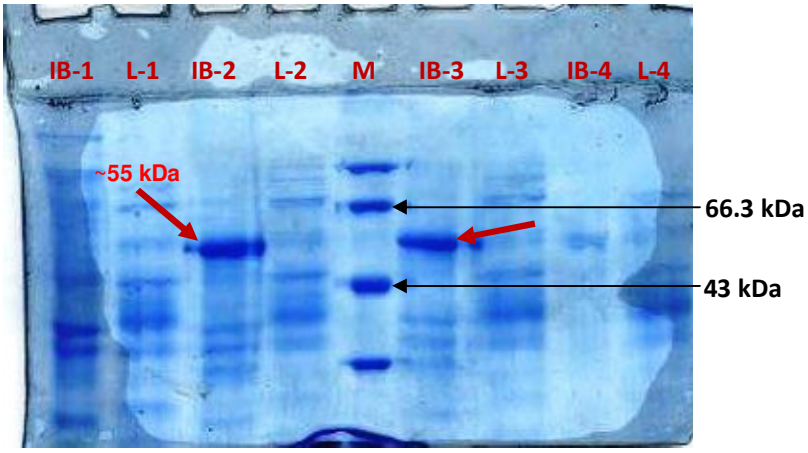
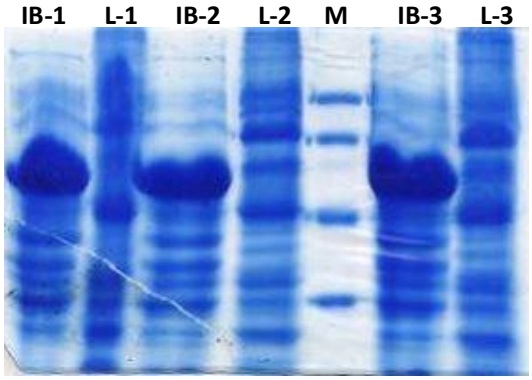


Fig. 4.6 Screening of Rosetta (DE3) strains for over-expression of *LIC4H*. Not all colonies positive for the presence of insert expresses the recombinant protein. IB= inclusion body; L= lysate; M= Protein molecular weight marker

4.3.5 Over-expression of three isoforms of LIC4H

One colony from each of the three isoforms was selected on the basis of loading equal amount of proteins and comparing the intensity of over-expressing, bulb like ~55 kDa protein, visually, after CBB staining (Fig. 4.7).



IB-1= Inclusion body of LIC4H1

L-1= Lysate of LIC4H1

IB-2= Inclusion body of LIC4H2

L-2= Lysate of LIC4H2

IB-3= Inclusion body of LIC4H3

L-3= Lysate of LIC4H3

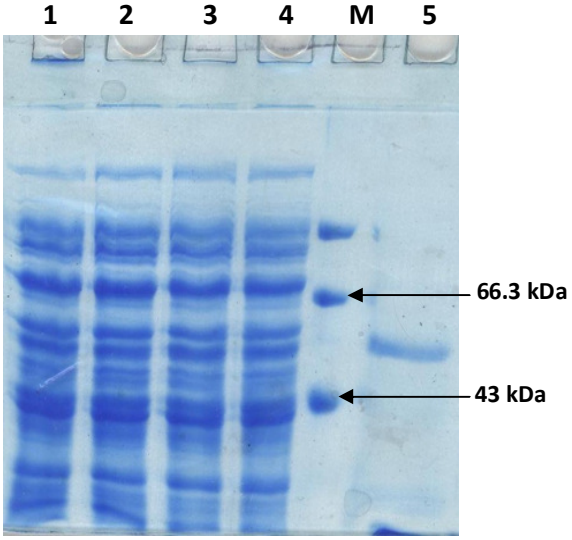
M= Protein mol. wt. marker

Fig. 4.7 Gel picture showing over-expression of three LIC4H isoforms

Fig. 4.7 shows a gel with over-expression of one colony each of the three isoforms. These colonies are the representative colonies from each isoform which showed maximum over-expression on the gel, after visual observation.

4.3.6 Attempts to maximize solubilisation of LIC4H1 and its purification from crude lysate

Since one of the reasons to heterologously express a plant protein in bacteria is its functional expression, it was decided to maximize the solubility of LIC4H1 in the bacteria's cytoplasm. A colony over-expressing LIC4H1 was selected for this purpose. Initial observation with mild decrease in IPTG concentration and temperature suggested that LIC4H proteins are highly hydrophobic and form inclusion body very rapidly. Hence it was planned to increase solubility by going for much lower temperature and IPTG concentration (Fig. 4.8 and 4.9).



Culture conditions:
0.05 M IPTG
15 ºC Temperature
Lane 1= 12 hour
Lane 2= 16 hour
Lane 3= 18 hour
Lane 4= 20 hour
Lane M= Protein mol. wt. marker
Lane 5= partially purified protein

Fig. 4.8 Gel photograph showing fractioned lysate on 10% SDS-Polyacrylamide gel. Culture conditions are mentioned above

Equal amount of secondary culture was harvested and equal amount of lysate was prepared. Protein concentration was determined in each culture condition mentioned above and equal amount of protein was loaded on to the gel. After visual observation of the gel following staining, conclusion to follow or discard a culture condition was reached (Fig. 4.8 and 4.9).

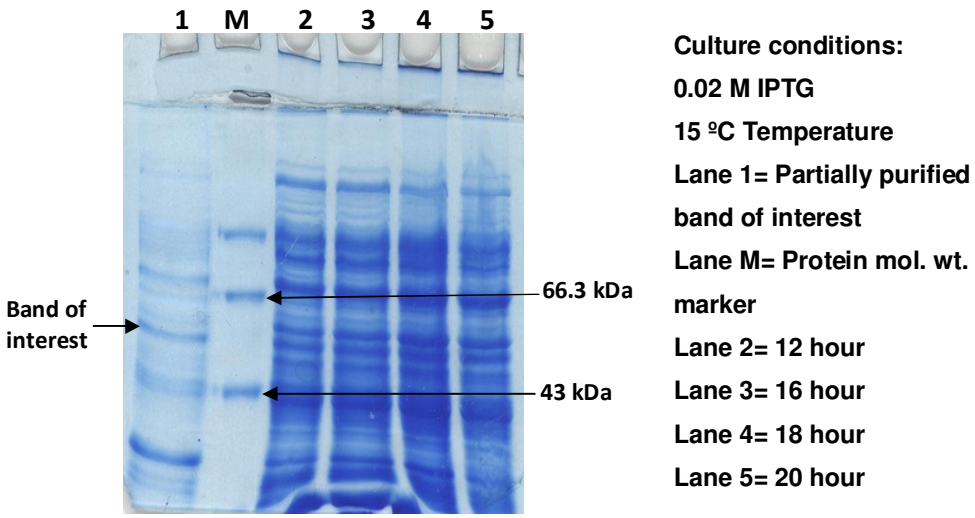


Fig. 4.9 Gel photograph showing fractioned lysate on 10% SDS-Polyacrylamide gel. Culture conditions are mentioned above

Comparing the gels in Fig. 4.8 and 4.9, it was decided to go for 15 ºC as incubation temperature, 0.02 mM of IPTG concentration for induction and 18 hour of incubation time, as maximum concentration of the expected sized protein was seen in that condition.

When it was tried to purify LIC4H1 from the crude lysate, I could not purify LIC4H1 from the crude lysate, as the binding of LIC4H1 to the Ni-NTA agarose matrix was very poor. When I tried to purify LIC4H1 using mild conditions, some of the bacterial intrinsic proteins also got bound to the Ni-NTA matrix which were co-eluted with LIC4H1 in the elution step. Increasing the stringency of purification resulted in non-binding of the protein (LIC4H1) to the matrix and elution of LIC4H1 either in the flow-through or during the washing step. It may be possible that the C-terminal hexa-histidine tag was buried inside the folded protein and hence was not available for interacting with the Ni-NTA matrix. Hence, neither purification of LIC4H1 from crude lysate nor its activity assay could be reported in this study. Protocol for purification of LIC4H1 from inclusion body is reported in the previous section of this chapter (section 4.2.6) and the results are shown in the next section 4.3.7.

4.3.7 Purification of LIC4H1 from inclusion body

LIC4H1 was purified from the inclusion body of *E. coli* secondary culture grown at 37 ºC for 5-6 hours after being induced using 0.5 mM IPTG. Fig. 4.10 shows one such gel silver stained after purifying LIC4H1 from *E. coli*.

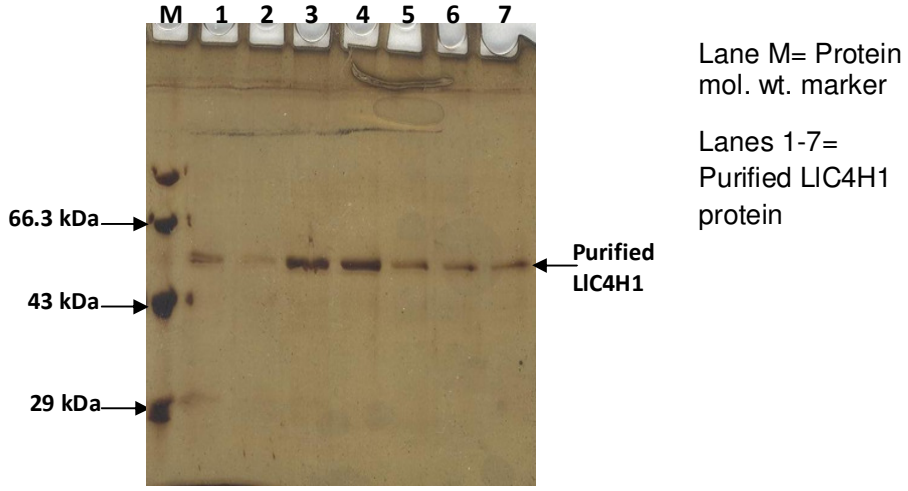


Fig. 4.10 Purified LIC4H1 from inclusion body

Purification of LIC4H1 became possible by employing highly stringent conditions. As was the case with crude lysate, LIC4H1 present in crude inclusion body also, did not bind strongly to the Ni-NTA matrix resulting in poor yield. A bacterial intrinsic protein of ~29 kDa size used to bind strongly to the matrix and used to co-elute with LIC4H1 during elution steps. The concentration of the contaminant bacterial protein looked only slightly lesser than the concentration of eluted LIC4H1 when checked on gel and stained. Due to this, binding and washing steps were made more and more stringent by increasing the concentration of imidazole and NaCl in the buffers. During the binding and washing steps, imidazole concentrations were increased to 10 mM and 30 mM respectively. Whereas NaCl concentration as high as 2 M was used to disturb the ionic interaction of the contaminant protein with the matrix in all the buffers. All these strategies although increased the purity of LIC4H1 from inclusion body, they compromised with the concentration of eluted LIC4H1. Eluted LIC4H1 protein was in the range of 50-100 µg/mL of elution buffer.

4.3.8 Prediction of secondary structure of LIC4Hs

Secondary structures of LIC4Hs were predicted using the software HNN tool. Since, the primary structures of LIC4Hs are more than 98% similar, their secondary structures are also very similar (Fig. 4.11).

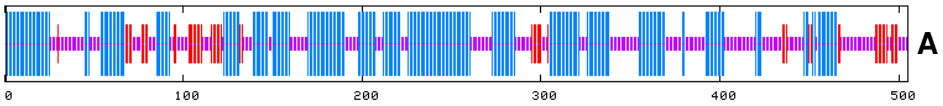


Fig. 4.11 (A) Secondary structure prediction of LIC4H1

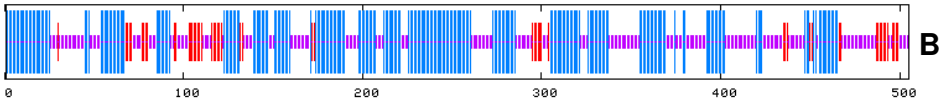


Fig. 4.11 (B) Secondary structure prediction of LIC4H2

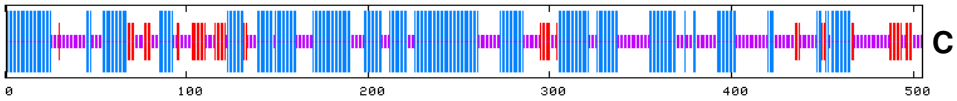


Fig. 4.11 (C) Secondary structure prediction of LIC4H3

In Fig. 4.11 (A, B and C), the longest blue coloured bar represent α-helix, the second longest red coloured bar represent extended strand and the smallest violet coloured bar represent random coil. In all the three isoforms, α-helices are mainly concentrated at the N-terminus and near the central region of the proteins. Extended strands are mainly concentrated downstream of the N-terminus and near the C-terminus, whereas random coil are uniformly distributed throughout the entire length of the protein. The percentage of the three structural elements in the three isoforms is given in Table 4.2.

Table 4.2 Comparison between secondary structures of the three isoforms

Core structure	LIC4H1	LIC4H2	LIC4H3
Alpha helix	48.32%	47.52%	48.71%
Extended strand	10.30%	10.69%	10.30%
Random coil	41.39%	41.78%	40.99%

4.3.9 Prediction of tertiary structure of LIC4H1

Tertiary structure of LIC4H1 was predicted using the programme Modeller 9v9 and viewed using the programme PyMol 1.3. The tertiary structure as predicted by Modeller 9v9 is that of a globular protein (Fig. 4.12). First 24 residues at the N-terminal of LIC4H1 were not included in modeling because the template structures did not contain residues of the N-terminal membrane-binding domain. The coordinates of heme in the model was obtained from 2NNJ and positioned in the target according to the template. The model having the least discrete optimized protein energy (DOPE) score and lowest PDF energy was selected as the best model. Further correction of model in the loop regions was done using the Modeller loop refinement module.

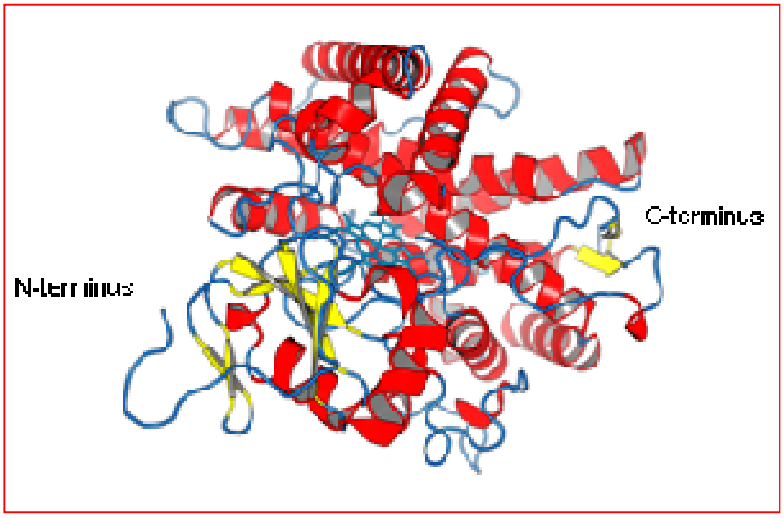


Fig. 4.12 Tertiary structure of LIC4H1 as predicted using the program Modeller 9v9

Ramachandran plot of the model was generated by the program PROCHECK, which assured very good confidence for the predicted model with 91.7% residues in most allowed region, 8.3% in additional allowed region and no residue in generously allowed region and disallowed region (Fig. 4.13). Glycine and proline residues are not considered while preparing Ramchandran plot.

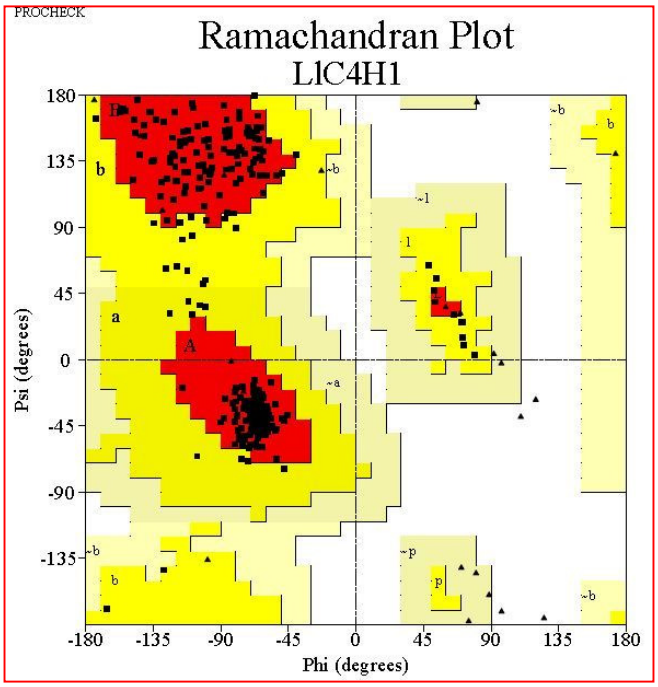


Fig. 4.13 Ramchandran plot of LlC4H1 as generated by the program PROCHECK

4.4 Discussion

LlC4H proteins were expressed in an *E. coli* strain, Rosetta (DE3). Attempts to express LlC4H in BL-21 (DE3) failed; this failure may be attributed to the translational incompatibility of this strain to translate the signal peptide of LlC4H. Living systems have biased preference for a particular codon out of the few possible codons to code for a particular amino acid, this is called codon biasness (Gustafsson et al. 2004; Hershberg and Petrov 2008). Bacteria and plants may have different codon biasness for a particular amino acid and hence, a simple system like bacteria may fail to synthesize a protein if it encounters a rarely used codon in bacterial system to code for an amino acid (Gustafsson et al. 2004; Hershberg and Petrov 2008). This problem was overcome by the introduction of genetically engineered *E. coli* strains which have the capability of coding for t-RNA specific to rarely used codons in prokaryotes. Rosetta (DE3) is one such strain and hence, LlC4H, which was otherwise not being expressed in BL-21 (DE3), was expressed in Rosetta (DE3). RosettaGami-2 (DE3) also contains t-RNA specific for codons rarely used in *E. coli*, but this strain failed to express LlC4H in it. Inability of

RosettaGami-2 (DE3) to express LlC4H revealed some unexplained difference between the two strains and is beyond simple logistics to suggest a possible reason behind this. Signal peptide from anchored proteins also needs to be removed to aid in its expression in bacteria, as signal peptides are normally hydrophobic and remain anchored to a particular eukaryotic membrane (Hotze et al. 1995). In LlC4Hs, N-terminal signal peptide was predicted in all the isoforms and hence, primers were designed to exclude the membrane spanning region from the recombinant protein.

As per my knowledge, so far there is only one report of functional expression of C4H in bacteria (Hotze et al. 1995), despite being the simplest expression system. Use of bacterial system to express C4H has been limited because of two major reasons. First is the absence of a suitable redox partner in bacteria and the other one is the translational incompatibility of the N-terminal signal sequence present in C4H. The functional expression for the first time became possible when the redox partner of C4H, NADPH:cytochrome P450 reductase (CPR) was fused to C4H in addition to deleting the signal sequence of C4H (Hotze et al. 1995). Hotze et al. (1995) did not report the use of any additional medium component, like δ-amino levulinic acid (ALA) or any iron source in to the medium, which is otherwise reported to be added to the medium for functional assay of some other P450 enzymes in bacteria (Nielsen and Møller 2000). Another way to functionally express C4H in bacteria may become possible through supplementing C4H with its redox partner (CPR), but higher efficiency may require purified proteins (both C4H and CPR) and in fact, applicability of this has never been tested for C4H. The stability of C4H as an enzyme is very poor and it loses 64% and 70% of its activity within 24 hours even at 4 ºC and -20 ºC respectively (Russell 1971). One cycle of freeze-thaw reduces C4H activity by 54% (Russell 1971). Its vulnerable activity makes it very susceptible to carry out activity assay experiments.

Although for functional analysis, C4H has been expressed in yeast (Hübner et al. 2003; Koopmann et al. 1999; Ro and Douglas 2004; Ro et al. 2001; Yamamura et al. 2001), it is associated with low yield and may not be

suitable for raising antibodies against C4H in rabbits, as it requires mg amount of purified protein to raise antibodies in rabbits. LlC4Hs were expressed in Rosetta (DE3) and as can be seen in Fig. 4.7, yield of LlC4H was optimum in my case to purify the protein in enough quantity. But, because binding of LlC4H with the matrix was so poor that it did not allow us to purify the protein in a quantity that could be used to raise antibody against the protein in rabbit. Hydropathy index of LlC4H1 (chapter 3, Fig. 3.14) also revealed that LlC4H1 is highly hydrophobic. The typical yield after purification varied between 50-100 µg/mL of eluent. This is very insignificant considering the huge amount of protein required to raise antibody in rabbit. Sebsequently upon purification, C4H was dialyzed or desalted through desalting column. But, due to recurring losses in every step, sufficient amount of protein could never be collected for raising antibody. Since the three isoforms are very much similar in amino acid sequence and composition, the rest two LlC4Hs may also be supposed to bear the same level of hydrophobicity and hence they were not considered for purification, once the purification of LlC4H1 failed to give substantial amount of LlC4H1 for raising antibodies.

Secondary structure prediction using HNN tool predicted almost similar structure for the three isoforms (Fig. 4.11). In LlC4H1, 244 amino acids (48.32%) form alpha helix, 52 amino acids (10.30%) form extended strands and 209 amino acids (41.39%) from random coil. HNN tool also predicted that LlC4H2 and LlC4H3 had 240 and 246 amino acids as alpha helix, 54 and 52 amino acids as extended strand and 211 and 207 amino acids as random coil, respectively. In another study, Liu et al. (2009) isolated *C4H* from *Parthenocissus henryana* and reported that secondary structure prediction using GOR IV (another software to predict protein secondary structure) predicted 41.78% alpha helix, 15.64% extended strand and 42.57% random coil in it. The distribution of these structures is also similar between LlC4Hs and C4H from *Parthenocissus henryana.* These data show that C4H proteins from two distantly related plants had similar secondary structure and this pattern may be present in other C4H sequences reported in the literature.

Tertiary structure prediction showed that LlC4H1 is a globular protein, consistent with earlier findings (Liu et al. 2009; Rupasinghe et al. 2003).

Rupasinghe et al. (2003) reported that four P450 proteins of phenylpropanoid pathway, despite sharing only 13% sequence identity in *Arabidopsis*, had highly conserved structural cores and several loop regions. Rupasinghe et al. (2003) also reported that the four enzymes employ common strategy to identify their substrates in that their cinnamate derived substrates align along helix I of the P450 with its aliphatic side chain facing the N-terminus and the aromatic ring facing the C-terminal of the enzyme. Sequences of LIC4Hs are so much identical that their tertiary structures are nearly the same.

4.5 Conclusion

Three isoforms of C4H were expressed in *E. coli*, strain Rosetta (DE3) and one isoform was purified from inclusion body. LIC4H proteins are highly hydrophobic and hence they profusely formed inclusion body in bacteria. Secondary and tertiary structure comparison between the LIC4H isoforms suggested that all the isoforms are identical in their secondary and tertiary structure.

CHAPTER - 5

Spatial and Temporal Expression of C4H in Leucaena leucocephala

Chapter 5: Spatial and Temporal Expression of *C4H* in *Leucaena leucocephala*

This chapter describes the spatio-temporal transcript abundance of *C4H* and the age and tissue specific activity of C4H in *L. leucocephala*. Tissue and age specific lignin content (both acid soluble and Klason lignin) in *L. leucocephala* in relation to C4H activity has also been discussed in this chapter. This chapter also features any specific materials and methods used in this chapter.

5.1 Review of literature

C4H activity was first demonstrated in pea seedlings (Russell and Conn 1967). Russell (1971) reported the requirement of molecular oxygen, NADPH and 2-mercaptoethanol for its activity from pea seedlings microsomal fraction. Reaction catalyzed by C4H was found to be inhibited by carbon monoxide (light reversible) and azide (Russell 1971). The activity of C4H in etiolated pea seedlings could be increased by exposure to light, 12 hours before extraction and assay (Russell 1971). The enzyme was expressed in young developing tissues during seedling stage and highest activity was recorded in apical bud with non-detectable activity in mature leaves (Russell 1971). This was followed by its detection in cell free extracts of potato tubers, buckwheat seedlings, parsley cell suspension culture, soybean, *Jerusalem artichoke* tubers and sorghum seedlings (Lamb and Rubery 1975). Later, it was discovered that PAL and C4H undergo concomitant changes in their enzyme activity following potato disc formation and illumination (Lamb 1977). There was a lag of 2 hours between illumination and induction of enzyme activity (Lamb 1977). C4H can be induced using manganese and this idea was used to over-express C4H in *J. artichoke* and led to first instance of its purification from *J. artichoke* (Gabriac et al. 1991). Further report of purification and characterization of C4H was from etiolated mung bean seedlings (Mizutani et al. 1993a). Fahrendorf and Dixon (1993) reported that *C4H* transcript level is stress inducible and strongly induced by fungal elicitor in alfalfa cell suspension culture at the beginning of accumulation of isoflavonoid phytoalexins. Following the elucidation of mRNA sequence of *C4H* from different plant species (Fahrendorf and Dixon 1993; Mizutani et al. 1993b;

Teutsch et al. 1993) and with the introduction of semi-quantitative RT PCR, northern hybridization and Q-RT PCR, flow of information on the inducibility and organ or tissue specificity of C4H started coming. Soon, it was reported that C4H is wound inducible in pea seedlings (Frank et al. 1996). Induction of enzyme activity in wounded or chemically treated *J. artichoke* tubers was reported to be the primary result of gene activation, although involvement of post-transcriptional modification in regulating C4H activity was not entirely excluded (Batard et al. 1997). In *Arabidopsis*, *C4H* transcript level was highest in inflorescence stems and was significantly higher in roots and siliques than in leaves and flowers (Mizutani et al. 1997), well supported by another independent study (Bell-lelong et al. 1997). *C4H* exists as multigene family in trees like *Populus tremuloides*, *P. Trichocarpa* (Lu et al. 2006), periwinkle (Hotze et al. 1995) and members of fabaceae family plants like pea (Whitbred and Schuler 2000) and alfalfa (Fahrendorf and Dixon 1993). Despite the presence of multiple isoforms of *C4H*, the precise roles of individual isoforms are not known in every case. In case of *L. leucocephala*, the sequences were so much similar at nucleotide level that it did not allow us to analyze their expression individually.

5.2 Materials and methods

5.2.1 *L. leucocephala* plant material and its source

Seeds from a healthy *L. leucocephala* cultivar K636 plant (this cultivar produces erect boles suitable for timber production) growing in our Institute (NCL) campus were collected and soaked in water. Floating seeds were discarded and the wet seeds were sundried for seven days and sterilized as reported by Shaik et al. (2009). Soaking was done to ensure that unhealthy seeds, if any, are not considered for the study. Imbibed seeds showing radical emergence (considered zero day) were inoculated in to sterile mixture of soil and farm yard manure (FYM) in plastic trays. For *C4H* transcript level study in stress induced plants, seeds were grown in modified MS medium (Shaik et al. 2009) in glass culture bottles. Experimental plants were grown in culture room with temperature of 25 ± 2 ºC, 16 hour photoperiod and 70% relative humidity (RH). Leaf, root and stem tissues were harvested from 5, 10,

15, 20, 30 and 60 days (d) plants (with the exception of 5 days old leaf tissue, as leaf had not emerged in 5 days old plants). In addition, leaf and stem branches (hereafter referred as stem) were also harvested from a one year old plant growing in the orchard of NCL.

5.2.2 RNA isolation and quantification

As explained in chapter 2, sections 2.5.5 and 2.5.6, respectively

5.2.3 First strand cDNA synthesis

As explained in chapter 2, sections 2.5.7

5.2.4 Quantitative Real Time PCR study

Quantitative RT PCR study was carried out using Mx3000P (Stratagene) Real Time PCR machine. cDNA were diluted and normalized against 18S ribosomal RNA gene. Customized 18S (Leu-18SF and Leu-18S-R) and *C4H* primers (C4H-RT-F and C4H-RT-R); and probes specific to *Leucaena* 18S ribosomal RNA gene (Leu-18S) and *C4H* gene (Leu-C4H) with 5' 6-FAM and 3' BHQ-1 modification were used in the study. Sequences of primers and probes are given in Table 5.1. In un-induced plants, spatio-temporal *C4H* transcript abundance in terms of fold increase/decrease was calculated taking *C4H* transcript level on zero day as basal value, arbitrarily set to 1.

Table 5.1 Primers and probes sequences used in this chapter (Probe sequences are marked with *)

Primer / Probe name	Primer / Probe sequence (5' → 3')
Leu-18S-F	CATTCGAACGTCTGCCCTATCA
Leu-18S-R	GATGTGGTAGCCGTTTCTCAGG
*Leu-18S	6-FAM-ATTCTCCGTCACCCGTCACCACCAT-BHQ-1
C4H-RT-F	GTCGCCAATGCCGTTTCC
C4H-RT-R	CTGAAGCCAATTTCCGAAGATG
*Leu-C4H	6-FAM-CGCTTCAAGTTTCCGCCGGGTCCTC-BHQ-1

5.2.5 NaCl, jasmonic acid and UV- stress

Seeds sterilized by the method described by Shaik et al. (2009) were grown in modified MS medium (Shaik et al. 2009) for stress related studies. 15 d old *L. leucocephala* seedlings were subjected to salt and jasmonic acid stress, by dipping the roots of the seedlings completely in modified MS liquid medium containing 100 mM NaCl and 100 µM methyl-jasmonate, respectively in standard tissue culture conditions. Control plants were dipped in the medium devoid of salt or methyl jasmonate. After 24 hour leaf, stem and root tissues were harvested for RNA isolation.

15 d old *L. leucocephala* seedlings growing in modified MS semi-solid medium was exposed to germicidal UV-C radiation (100 nm – 280 nm) produced by medium pressure lamps of the laminar air flow for 15 min, by keeping the caps of the culture bottles opened and the bottle kept on the cabinet floor of the laminar air flow. After 15 min of exposure, the seedlings were grown for 24 h under standard tissue culture conditions to allow the accumulation of UV-responsive transcripts in the plant tissues. The plant tissues were then harvested for RNA isolation. For control plants, the bottle cap was opened but the plants were not exposed to UV irradiation. For stressed tissues, fold increase/decrease in *C4H* transcript level was calculated with respect to the control tissue, arbitrarily assumed 1.

5.2.6 Stock solutions

NADPH (2 mM), dissolved in 0.1 N NaOH

trans-Cinnamic acid (2 mM), dissolved in HPLC grade methanol

PMSF (1 M), dissolved in DMSO

5.2.7 C4H enzyme assay from crude tissue extract

Spectrophotometric analysis of C4H activity from crude tissue extract was performed as described by Lamb and Rubery (1975). In the assay experiment, all the steps were performed in ice or at 4 ºC unless otherwise stated. Briefly, about 4 gm of the tissue was crushed in liquid N_2 and extracted in 12 mL of cold extraction buffer (Chen et al. 2006) with slight modification (50 mM Tris-HCl buffer, pH 7.5, 15 mM 2-mercaproethanol, 4 mM $MgCl_2$, 5 mM ascorbic acid, 10 mM leupeptin, 1 mM PMSF, 10% glycerol

and 0.15% w/v PVP). 2-mercaptoethanol and PMSF were added freshly to the extraction buffer. The homogenate was filtered through four layers of muslin cloth and centrifuged at 12,000Xg for 20 min at 4 ºC. Supernatant was transferred to a new tube, quantified using Bradford reagent (Bradford 1976) taking BSA as standard and used as a source of crude enzyme for C4H assay. The reaction was performed in uncapped 15 ml falcon tube in a volume of 3 mL which contained 50 mM Tris-HCl buffer (pH 7.5), 2 µM NADPH (Sigma, USA), 2 µM *trans*-cinnamic acid (Sigma) and 0.5 mL crude extract. All components without substrate served as control. Reaction mixture without substrate was incubated at room temperature for 5 min and the reaction was started by addition of the substrate. Reaction mixture was incubated in shaking condition (160 rpm) at 25 ºC for 1 hour and the reaction was terminated by addition of 100 µL 11.6 M HCl. Denatured protein was pelleted by centrifuging for 10 min at 12,000Xg. The pH of 2.5 mL of supernatant was adjusted to 11.0 by adding 1.0 N NaOH and absorbance was measured at 340 nm in PerkinElmer Lambda 650 UV/VIS Spectrometer. The reactions were performed in triplicates and the absorbance values were averaged. One unit of C4H activity equals to a change of 0.01 in absorbance per hour and expressed as U/mg protein/h (Chen et al. 2006).

5.2.7 Above ground fresh weight measurement of *L. leucocephala* seedlings

Five 5, 10 and 15 d old seedlings were uprooted carefully and cut at their stem-root junction and fresh weight of the above ground region (stem, cotyledon and leaves) were taken on a Contech precision balance up to four significant digits in gram.

5.2.8 Lignin estimation from tissues

Leaf, stem and root tissues were harvested from 30 d and 60 d old experimental plants. Leaf and stem tissues from a 1 year old plant were also harvested. Acid soluble lignin and Klason lignin were determined from the tissues as discussed in chapter 2, section 2.10.

5.3 Results

Owing to very high identity of *C4H* isoforms at nucleotide level, it was not possible to design primers and probe specific to a particular isoform as there always stood possibility of either non-target amplification by the primers or a probe binding to non-target amplicon. Hence, one set of primers and probe capable to detect the three isoforms simultaneously was employed in the study.

Crude extract of various tissues at different age was used for enzyme assay reactions. Since crude tissue extract contained all the proteins at that particular point of time in that tissue, it was not possible to discriminate between C4H allelic variants and attribute the reaction catalyzed to a particular variant. Thus, spatio-temporal expression pattern of *C4H* represents the collective transcript level or enzyme activity of the allelic variants.

5.3.1 Spatio-temporal transcript abundance of *C4H* in *L. leucocephala* different tissues

The investigations have been carried out in developing seedling stage (starting from 0 d to 5, 10, 15, 20, 30 and 60 d) to mature 1 year old plant. Transcript level of *C4H* in *L. leucocephala* was detectable as early as just after imbibition (0 d embryo). Transcript level in leaf, root and stem tissues of various ages with respect to the transcript level at zero day embryo is given in Table 5.2. Transcript level in zero day embryo has arbitrarily been assigned value 1, and the respective values in other tissues have been calculated.

In leaf tissue, there was progressive increase in the *C4H* transcript level with the advancing age of the plant and highest level was detected in one year old tree leaves (Fig. 5.1 A). In root (Fig. 5.1 B) and stem (Fig. 5.1 C) tissues, there was increase in *C4H* transcript level till 30 d, although a phase of slight decrease in the transcript level was observed in 15 d old seedling with respect to 10 d old root and stem tissues. On 60 d, both root and stem transcript level was found decreased with respect to 30 d old tissues, but in one year old leaf and stem tissue, the transcript level was found to be all time high. Maximum transcript abundance at any time in any tissue was observed in 30 d old root tissue.

Table 5.2 *C4H* transcript level in tissues of different age (Transcript level at zero day was arbitrarily assumed 1 and relative values calculated)

Tissue	5 d	10 d	15 d	20 d	30 d	60 d	1 Yr
Leaf	NA	1.9	9.5	18	98	404	1031
Root	2.5	12.4	10.7	214	3061	519	ND
Stem	3.2	29	16.6	31	936	133	1698

(NA= Not applicable; ND= Not determined)

Fig. 5.1 (A) Transcript levels of *C4H* in leaf tissues of different age

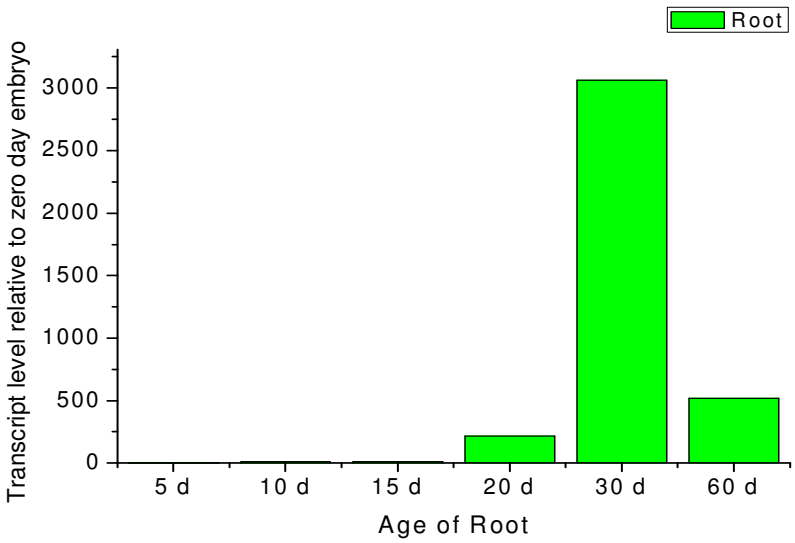


Fig. 5.1 (B) Transcript levels of *C4H* in root tissues of different age

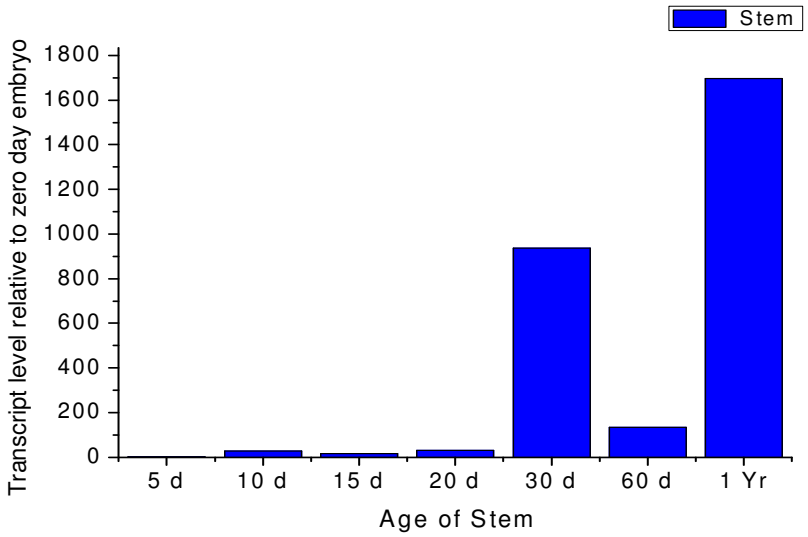


Fig. 5.1 (C) Transcript levels of *C4H* in stem tissues of different age

Since great difference in transcript levels was observed before 20 d and after 30 d, for better comparison, spatio-temporal transcript abundance has been compared into two phases, first phase of 0-20 d (Fig 5.2 A) and second phase of 20 d - 1 year (Fig 5.2 B). Transcript level of 20 d root is shown in Fig. 5.2 (B) and transcript level in 1 Yr root was not determined.

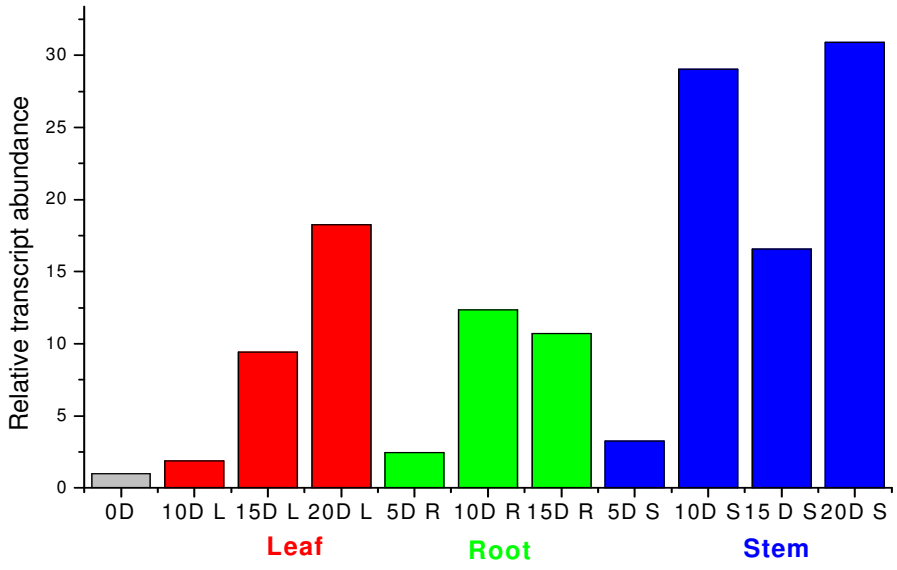


Fig. 5.2 (A) Comparison of *C4H* transcript levels in 0- 20 d old tissues. L denotes leaf (red bars), R denotes root (green bars) and S denotes stem (blue bars)

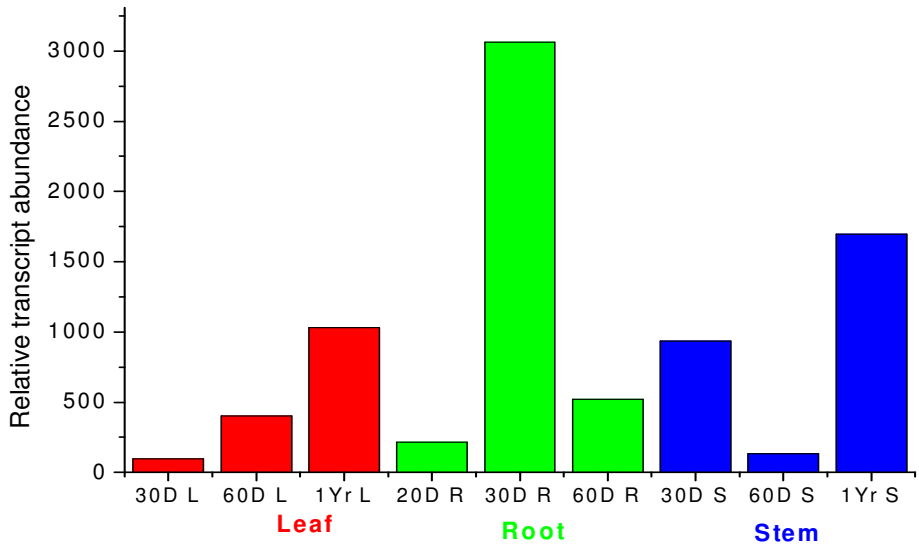


Fig. 5.2 (B) Comparison of *C4H* transcript levels in 30 d - 1 Yr old tissues. L denotes leaf (red bars), R denotes root (green bars) and S denotes stem (blue bars)

5.3.2 Effect of various stresses on the transcript level of *C4H* in different tissues of *L. leucocephala*

To the best of my knowledge, effect of salt stress on the transcript level of *C4H* has been not been described earlier in other plants. Dipping root of *Leucaena* seedlings in salt solution (100 mM NaCl) for 24 hours, marginally increased *C4H* transcript level in leaf and stem (9.6 and 3 times respectively), but substantially increased the m-RNA level in root (44 times). Dipping root of *Leucaena* seedlings in methyl jasmonate (100 µM) solution for 24 hours, substantially increased *C4H* transcript level in leaf, root and stem tissues (63, 47 and 179 times respectively of control plants). UV-C stress in seedlings increased *C4H* transcript level in leaf and stem (17.5 and 5.7 times), but surprisingly decreased its level in root (0.6 times of control root). The results are summarised in Table 5.3 and Fig. 5.3.

Table 5.3 Effects of various stimuli on *C4H* transcript level

Tissue	Salt stress	Methyl jasmonate stress	UV-C stress
Leaf	9.6	63	17.5
Root	44	47	0.6
Stem	3	179	5.7

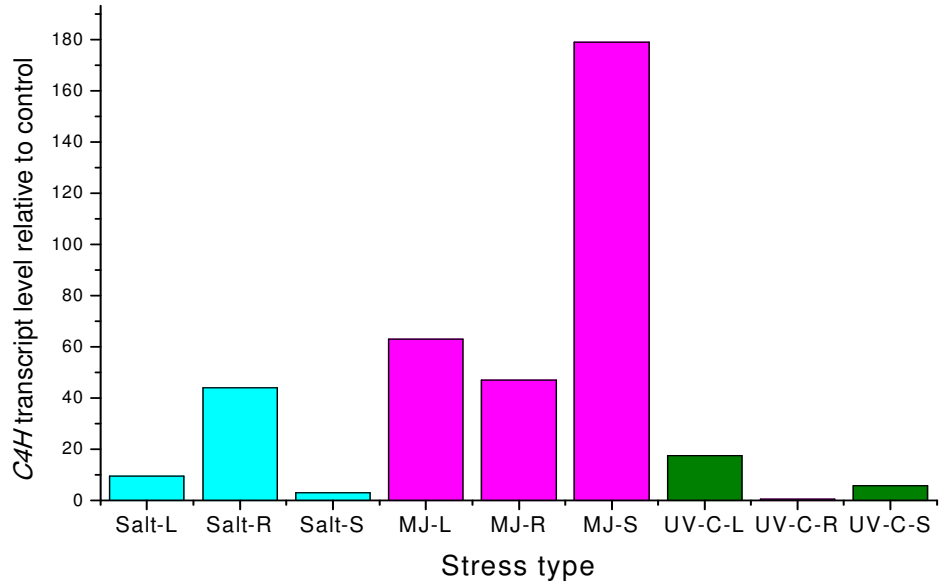


Fig. 5.3 Induction of *C4H* transcript levels in leaf, root and stem tissues of *L. leucocephala* 15 day old seedlings by various stress types. L, R and S denote leaf, root and stem respectively. MJ denotes methyl jasmonate

5.3.3 Extractable C4H activity from leaf, root and stem tissues of *L. leucocephala* of various ages

The protocol employed did not show detectable level of C4H activity from crude tissue extract until 30 d except in root tissue where C4H activity was demonstrable by 20 d. Leaf, root and stem tissues tested on 30 d after germination had detectable activity with the highest activity observed for root and least for leaf tissue (Table 5.4). In all the tissues investigated, an age dependent positive correlation between C4H activity and the age of the tissue was found with greater activity in older tissues (Fig. 5.4).

Table 5.4 Extractable C4H activity assay from leaf, root and stem tissues of *L. leucocephala* of various ages

Tissue	20 d	30 d	60 d	1 Yr
Leaf	UD	0.06 ± 0.01	0.11 ± 0.03	0.41 ± 0.1
Root	0.76 ± 0.1	1.45 ± 0.08	2.25 ± 0.32	ND
Stem	UD	0.54 ± 0.17	1.06 ± 0.15	2.85 ± 0.23

(UD= un-detectable; ND= Not determined)

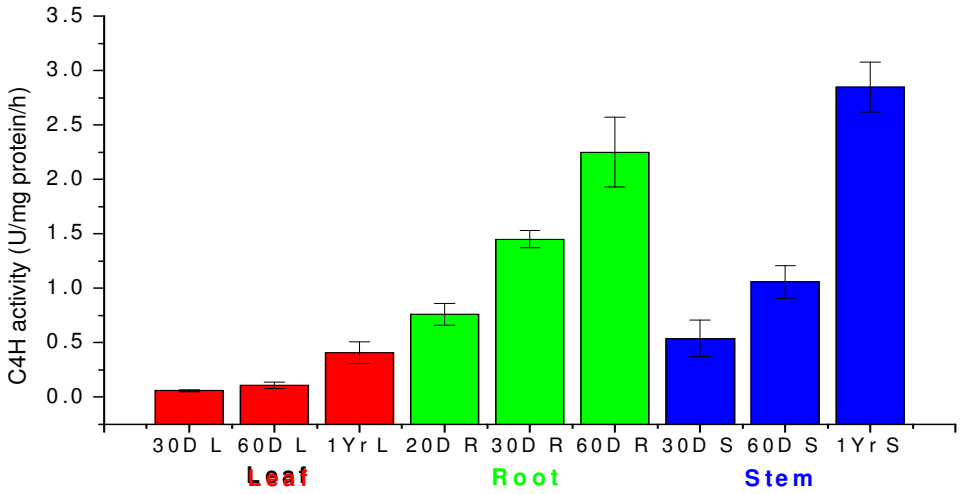


Fig. 5.4 Extractable C4H activity from *L. leucocephala* tissues of different age. L, R and S denote leaf, root and stem respectively

5.3.4 Above ground fresh weight measurement of *L. leucocephala* seedlings

Fresh weight of above ground portion of five *L. leucocephala* seedlings (stem, leaves and cotyledons) of 5 d, 10d and 15 day of age were taken. The seedlings were cut at the root-stem junction and the stem portion was weighed in a precision balance up to four significant digits in gram. In the same span of five days, there is higher difference in weight between 5 d and 10 d old seedlings than 10 d and 15 d old seedlings. Table 5.5 clearly shows that the seedlings had gained little weight in between 10 and 15 day of age, whereas the height of the 15 d old seedling is about double the height of 10 d old seedling. For a comparison between the sizes of different aged seedlings (0 day seed to 20 d old seedlings), please see Fig. 5.5.

Table 5.5 Above ground fresh weight of 5-15 d old *Leucaena* seedlings

Seedling age	Seedling fresh above ground Weight (gram)
5 day	0.1704 ± 0.025
10 day	0.2332 ± 0.032
15 day	0.2547 ± 0.014

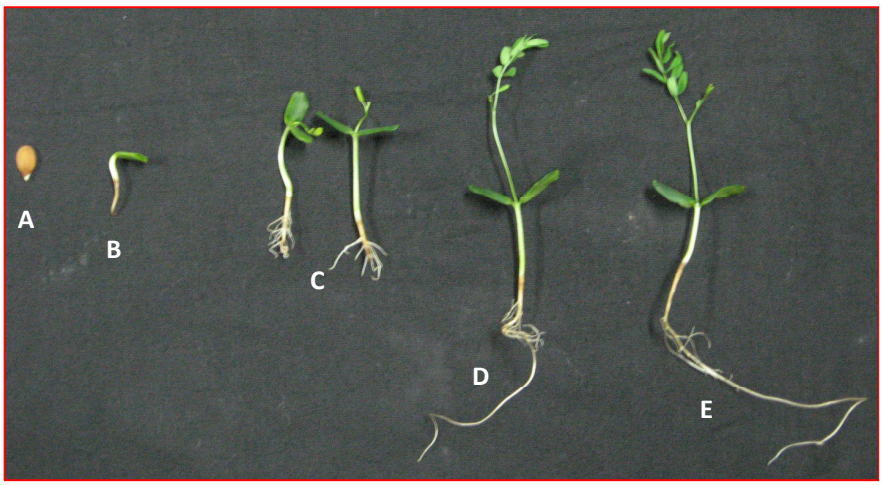


Fig. 5.5 Comparative size of *L. leucocephala* seedlings. A (0 day), B (5 day), C (10 day), D (15 day), E (20 day) old seedlings

5.3.5 Lignin estimation from *L. leucocephala* tissues of different age

Analysis of lignin content in *L. leucocephala* different tissues showed that while the acid soluble lignin first decreases and then increases marginally with age, the Klason lignin increases with age. As predictable from transcript level and enzyme activity study, root tissues are very highly lignified followed by stem and leaves. The lignin content of *L. leucocephala* different tissues as analyzed is presented in the form of Table 5.6 and Fig. 5.6.

Table 5.6 Lignin content of *L. leucocephala* tissues of different age

Tissue	Acid soluble lignin (ASL)	Klason lignin (KL)	Total lignin
30 d Leaf	4.09 ± 0.31	2.51 ± 0.33	6.6 ± 0.64
60 d Leaf	3.33 ± 0.35	6.11 ± 0.99	9.44 ± 1.34
1 Yr Leaf	3.38 ± 0.25	10.25 ± 0.71	13.63 ± 0.96
30 d Root	1.70 ± 0.18	15.49 ± 1.09	17.19 ± 1.27
60 d Root	1.33 ± 0.09	21.12 ± 0.69	22.45 ± 0.78
30 d Stem	2.47 ± 0.17	9.7 ± 0.67	12.17 ± 0.84
60 d Stem	1.41 ± 0.11	13.53 ± 0.70	14.94 ± 0.81
1 Yr Stem	1.70 ± 0.06	25.14 ± 0.33	26.84 ± 0.39

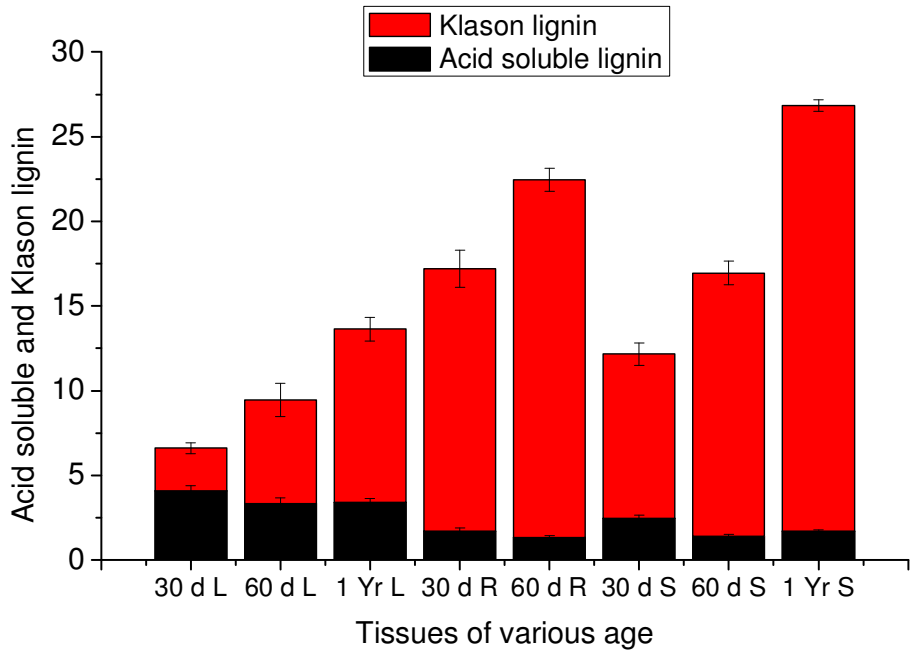


Fig. 5.6 Acid soluble and Klason lignin content from leaf, root and stem tissues of *L. leucocephala* of different age

5.3.6 Phloroglucinol staining of root and stem tissue sections

Free hand tissue sections of 10 d and 60 d old stem and root sections were taken and stained using acidic phloroglucinol as mentioned in chapter 2, section 2.9. Acidic phloroglucinol stains lignin red. As can be seen in Fig. 5.7 (A) and 5.7 (C), there is vast difference in the staining patterns of 10 d and 60 d old stem. Similarly, between 10 d and 60 d old root samples (Fig. 5.7 (B) and 5.7 (D), staining is more intense in 60 d old root section compared to 10 d old root sections. Photographs were taken immediately after staining, as phloroglucinol staining is not permanent. All the photographs have been taken in 10X magnification. Xylem tissues are highly lignified as can be seen in all the sections. Even, 10 d old stem and root sections easily get stained by phloroglucinol.

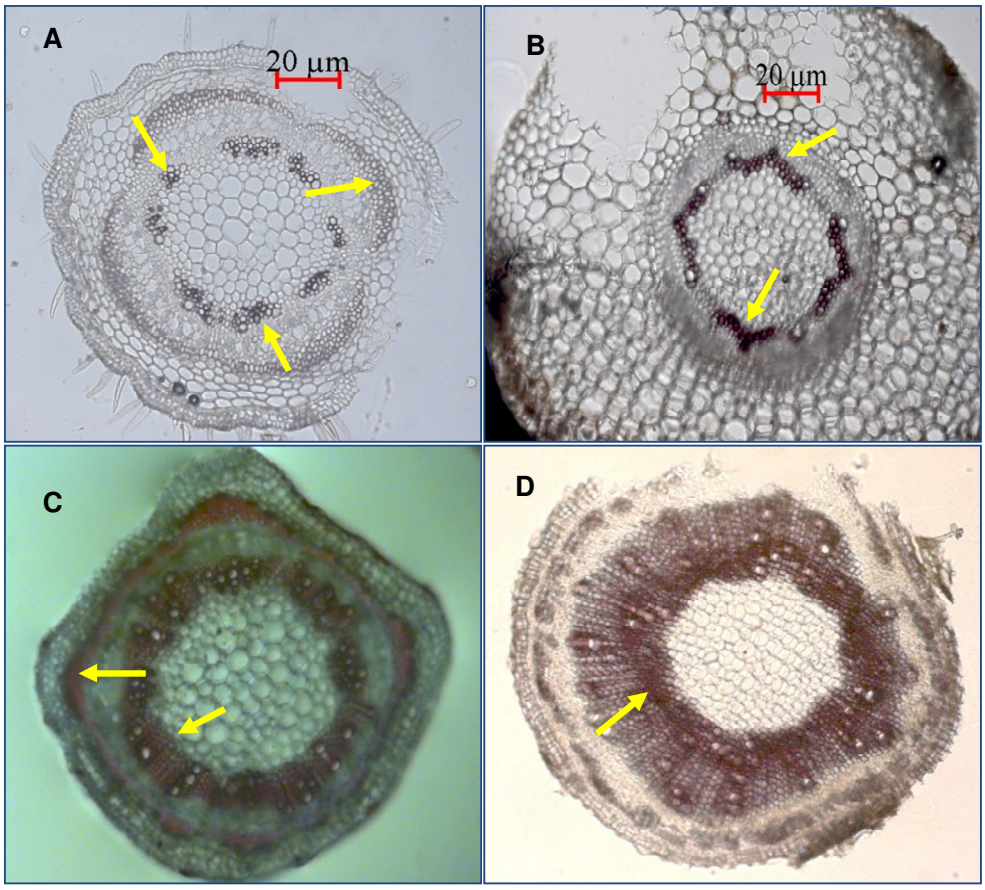


Fig. 5.7 Acidic-Phloroglucinol staining of (A) 10 d old stem section, (B) 10 d old root section, (C) 60 d old stem section, and (D) 60 d old root section. Acidic-phloroglucinol stains lignin red. Magnification is 10X for all the photographs. As can be seen, xylem tissues (indicated by arrows) are highly lignified in both stem and root sections.

5.4 Discussion

To the best of my knowledge, developmental transcript profiling and activity study of C4H in various tissues of a tree species starting from young stage to mature stage has not been reported in detail. These two studies in conjunction will further enhance our understanding of time and tissue dependent regulation of C4H transcription, translation and the developmental role that C4H plays during initial years of plant growth in relation to lignification in *L. leucocephala*. The investigations have been carried out in

developing seedling stage (starting from 0 d to 5, 10, 15, 20, 30 and 60 d) to mature 1 year old plant. Transcript level of *C4H* in *L. leucocephala* was detectable as early as just after imbibition (0 d embryo). This finding is consistent with earlier report in *Arabidopsis* (Bell-lelong et al. 1997) where the authors detected *C4H* transcripts in recently imbibed seeds using RNA blot hybridization experiment. In leaf tissue, there was progressive increase in the *C4H* transcript level with the advancing age of the plant and highest level was detected in one year old tree leaves. In root and stem tissues, there was increase in *C4H* transcript level till 30 d, although a phase of slight decrease in the transcript level was observed in 15 d old seedling with respect to 10 d old root and stem tissues. A plausible reason for the slight decrease may be the voluminous increase in size of 15 d old seedling relative to 10 d old seedling. It may be proposed that due to active cell division without accumulating much biomass (Table 5.5, Fig 5.5), and because of the cell machinery paying more attention towards primary metabolism than secondary metabolism, the transcription of *C4H* slowed down. Another reason of the decrease in *C4H* transcript level may be the maturation of vascular bundle in the stem from isolated to a continuous ring and the appearance of secondary xylem tissue in stem. This temporary halt was also observed in case of another general phenylpropanoid pathway gene, 4CL (Gupta 2008) in an independent work carried out in our lab. On 60 d, both root and stem transcript level was found decreased with respect to 30 d old tissues which may be due to attaining very high transcript level by 30 d and subsequently upon active enzyme formation, there is no longer need of maintaining high transcript level. In one year old leaf and stem tissue, the transcript level was found to be all time high. This is expected because as the plant matures and its mass increases due to its entry to the secondary growth phase, there is increased requirement of lignification to provide mechanical strength to the tree. Maximum transcript abundance at any time in any tissue was observed in 30 d old root tissue consistent with earlier reports that *C4H* is highly expressed in roots (Bell-lelong et al. 1997; Frank et al. 1996; Mizutani et al. 1997; Park et al. 2010) and cells undergoing active lignification especially xylem cells (Lu et al. 2006; Ro et al. 2001).

As per best of my knowledge, induction of *C4H* transcript level as a result of salt stress has not been reported earlier in other plants. Although, there is only marginal increase in the transcript level in leaves and stem tissues compared to control tissues, the increase in root tissue is substantial. This substantial increase may be due to the fact that roots were kept immersed in the salt solution and hence root *C4H* transcription level increased. Involvement of C4H in plant's defence is well documented (Belles et al. 2008; Fahrendorf and Dixon 1993; Naoumkina et al. 2010; Whitbred and Schuler 2000). Plants, when challenged with pathogens, produce jasmonic acid. Reversibly speaking, plants induced with jasmonic acid (or methyl jasmonate) behave as if challenged by pathogen. Hence, addition of methyl jasmonate to the growing medium induced *C4H* transcript level in tissues (Huang et al. 2008). It is interesting to note that although the roots of the seedlings were dipped in methyl jasmonate, greater induction was observed in stem and leaves than root tissues. *C4H* is also reported to be induced by UV-B rays (Huang et al. 2008; Jiang et al. 2005; Lee et al. 2008; Lu et al. 2006; Park et al. 2010; Shi et al. 2010) and *C4H* promoters are reported to contain UV responsive elements (Shi et al. 2010). That *C4H* transcription can be increased or decreased by germicidal UV-C rays is not reported. Present work reports that *C4H* transcript level in stem and leaves increased upon UV-C exposure, whereas the transcript level in roots decreased surprisingly to about 60% of control tissue. This may be due to the fact that UV-C radiation could not penetrate the semi-solid medium and remained limited to leaves and stem tissues. Hence *C4H* transcript level in leaf and stem increased to a high level shortly after the induction and within 24 hours, transcription factors responsible for repressing *C4H* transcription became active in the seedling to repress *C4H* transcription. These transcription factors reduced *C4H* transcription level in root tissues, as *C4H* was not induced in root tissues because UV-C rays could not penetrate the medium. Omer (2011) reported that after 24 hours of *Leucaena* seedlings being induced by UV-C rays, *Ll*MYB_SSM (an R2R3 type MYB transcription factor gene which is a repressor of phenylpropanoid pathway in *L. leucocephala*) transcript level increased to about 1300 times in leaves and about 35 times in stem. This

repressor may be responsible to decrease *C4H* transcription in UV-C induced roots in my experiment.

Spectrophotometric determination of C4H activity from crude tissue extract according to the protocol of Lamb and Rubery (1975) did not show detectable level of C4H activity from *L. leucocephala* crude tissue extract until 30 d except in root tissue where C4H activity was demonstrable by 20 d. This is in agreement with the transcript abundance study where *C4H* transcript level in root on 20 d had jumped abruptly to a higher value as compared to the level on 15 d (Fig. 5.1 B). This result also suggests that enhanced transcript level is the primary force affecting C4H activity (Batard et al. 1997; Frank et al. 1996; Mizutani et al. 1997). Leaf, root and stem tissues tested on 30 d after germination had detectable activity with the highest activity observed for root and least for leaf tissue (Fig 5.4). 30 d root and stem activity were 24 and 9 times higher than the same aged leaves, respectively, whereas root and stem transcript level were around 30 and 9 times higher than in leaves, respectively. The disproportionate activity and transcript level in the root on 30 d may be attributed to more than 300 times increase in the transcript level in a span of 15 days (when compared to 15 d old root tissue); and that the transcripts were either not translated or if translated, not modified post-translationally to form active enzyme at such a fast pace. Lagging of C4H activity behind *C4H* transcript or protein accumulation has been reported from wound induced *J. artichoke* (Batard et al. 1997) and pea (Frank et al. 1996) where the peak transcript level was followed by peak activity after a gap of 10-20 hours (Batard et al. 1997) in *J. artichoke* and 7 hours (Frank et al. 1996) in pea. Post translational modification of C4H includes its attachment to heme ligand, insertion to ER membrane and coupling with its redox partner (NADPH:cytochrome P450 reductase, CPR). Among 60 d old tissues, maximum activity was observed in 60 d old root tissue followed by stem and leaf tissues. Although quantitative real time data showed decreased transcript level in root and stem on 60 d than 30 d, the enzyme activity is higher on 60 d than 30 d. After attaining high transcript level in tissues by 30 d, it may be assumed that they got translated to active enzymes and once a basal level of enzyme activity is achieved, the transcript level becomes

dependent on the requirement of active enzyme level in the tissue. In the absence of data pertaining to activity of 1 year old root tissue, highest activity was observed in 1 year old stem, slightly above the activity in root at 60 d. C4H activity assay was conducted taking crude protein extract in which the fraction of C4H present varied as per the tissue/age. Nevertheless, in all the tissues investigated, an age dependent positive correlation between C4H activity and the age of the tissue was found with greater activity in older tissues (Fig 5.4); whereas *C4H* transcript level was not found to follow such an increasing pattern indicating modulation in *C4H* transcript level may be the primary force affecting C4H activity, but its level is tightly regulated as per the requirement of the tissue. Another conclusion that can be drawn from the correlation between the two studies is while the transcript level induction or modulation is a transient phenomenon; its effect at protein level is stable and lasts longer.

By looking at the vigour of 15 d old seedlings (for comparison between the sizes of seedlings, please see Fig. 5.5) we expected that the fresh weight of 15 d old seedlings should be substantially higher than 10 d old seedlings. Contrary to our assumption, although 15 d old seedlings were more vigorously grown than 10 day old seedlings, they had gained little fresh weight (Table 5.5) emphasizing the active primary metabolism and growth phase of seedlings. At this stage, cells may be actively dividing, elongating and differentiating into specific tissues. And hence, the transcript level of *C4H*, a secondary metabolic pathway gene reduced or came to a halt in 15 d old tissues than in than 10 d old tissues.

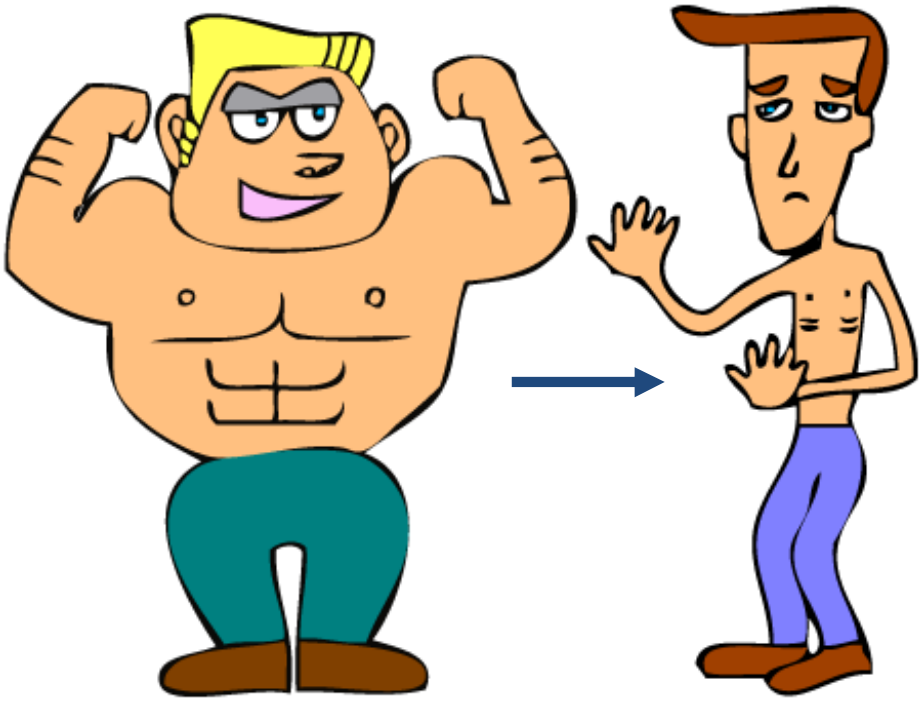
Lignin content was analysed in 30 and 60 day old leaf, root and stem and leaf and stem tissue of 1 year old plants. Total lignin was found to increase in tissues with advancing age (Table 5.6). Acid soluble lignin first decreased in tissues (60 d in comparison to 30 day) and then increased slightly (1 Yr in comparison to 60 d). On the other hand, acid insoluble lignin increased with age. As expected from the transcript level and activity study of C4H, lignification in roots is highest (Bell-lelong et al. 1997; Frank et al. 1996; Mizutani et al. 1997; Park et al. 2010) followed by stem and leaf tissues

respectively. During immature stage of plants, ASL forms a major proportion of total lignin; but as the plant matures, the proportion of KL increases and in one year old stem tissue it reaches near 94 % of total lignin. Rastogi and Dwivedi (2006) reported that KL constituted nearly 96 % of total lignin in 1.5 year old *L. leucocephala* stem. Total lignin content followed the pattern of extractable C4H activity from crude extract. Since P450 enzymes are catalytically slow (Ehlting et al. 2006), they might be supposed to be catalyzing the rate limiting step. C4H in this way plays determining role in lignification. As removal of insoluble lignin from pulp employing environmentally or economically benign methods is the major challenge in paper industry, downregulation of *C4H* in *L. leucocephala* offers very prospective alternative to decrease insoluble lignin (or for that matter total lignin) content in transgenics. Phloroglucinol staining of different aged root and stem transverse sections showed clear differences in accumulated lignin over time and gave histological details of the sections (Fig. 5.7).

5.5 Conclusion

Maximum *C4H* transcript level was found in root tissue followed by stem and leaves. Highest tissue specific activity was recorded in 1 year stem tissue. Lignin quantification from tissues of different age suggests that lignification is directly proportional to C4H activity in *L. leucocephala*. This study strengthens our hypothesis that decreasing C4H activity in transgenic *L. leucocephala* may result in reduced lignification in plants.

CHAPTER - 6



Genetic Engineering of tobacco and Leucaena leucocephala for down-regulation of C4H

Chapter 6: Genetic Engineering of tobacco and *Leucaena leucocephala* for down-regulation of C4H

This chapter describes the genetic transformation of tobacco and *Leucaena leucocephala* for downregulation of *C4H* and transgenics analysis. This chapter also deals with the specific materials and methods for raising transgenics and their analysis carried out during the course of this chapter.

6.1 Review of literature

To engineer plants with agronomically desirable lignin related traits, it is necessary to devise strategies to predictably and flexibly reduce lignin content and/or alter the monomer composition in favour of higher S/G ratio. Reduction in lignin content or alteration in lignin composition can be brought about either by tree breeding programs or by exploring the potential of genetic engineering by means of down-regulating key lignin biosynthesis gene(s). Downregulation of enzymes other than general phenylpropanoid pathway enzymes (PAL, C4H and 4CL) in transgenic plants have presented unpredictable and sometimes contradictory results. Ni et al. (1994) reported that modest reduction in COMT activity in transgenic tobacco resulted in small reduction in lignin content with no significant change in monomer composition. In contrast, 60% reduction in OMT activity in transgenic *Leucaena* resulted in 28% lesser lignin and a probable decrease in S units predicted on the basis of histo-chemical staining of tissue sections (Rastogi and Dwivedi 2006). Halpin et al. (1994) showed that downregulation of CAD, the last enzyme of phenylpropanoid pathway lead to more qualitative change in lignin than quantitative change and produced modified lignin in otherwise phenotypically normal plants. On the other hand, Saathoff et al. (2011) reported that RNA mediated silencing of CAD had resulted in decreased overall lignin with altered lignin structure associated with higher glucose recovery from tissues when treated with cellulase.

PAL is the first enzyme of general phenylpropanoid pathway and its strong suppression lead to reduced lignin content and high S/G ratio (Sewalt et al. 1997). The increase in S/G ratio was the result of decrease in G monomer while no change in the level of S monomer. *Pinus taeda* cell suspension culture, when fed with exogenously supplied phenylalanine to

saturating level in the phenylpropanoid pathway induction medium, resulted in coordinated up-regulation in the transcript level of *PAL, 4CL, CCoAOMT*, *CCR* and *CAD*. By contrast, there was only slight increase in the transcript level of the two P450 genes, *C4H* and *C3H* (Anterola et al. 2002). That was accompanied by large increase in the intracellular concentration of cinnamic acid and *p*-coumaric acid. These results suggest that PAL, 4CL, CCoAOMT, CCR and CAD cannot be truly rate limiting as their transcription level gets modulated as per the metabolic demand. On the other hand, limited expression of C4H and C3H enables them to keep control over metabolic flux into phenylpropanoid pathway. Decrease in the lignin content as a result of PAL or 4CL downregulation has been reported but it is when there is severe suppression of their activity.

C4H catalyzes slow reaction with high substrate specificity (Ehlting et al. 2006) and is located at strategic important point as the pathway branches after C4H. Also, the reaction catalyzed by C4H is exothermic and irreversible (i.e. represent a point of no return) (Ehlting et al. 2006); hence C4H plays a vital role at the very beginning of the phenylpropanoid pathway. Metabolic channelling between a specific form of PAL and C4H has been successfully demonstrated in the form of multi-enzyme complex. This complex helps in keeping the local concentration of PAL generated cinnamic acid high and does not let it to diffuse to the cytosol. Exogenously supplied cinnamic acid inhibits PAL transcription and activity with simultaneous induction of the synthesis of a proteinaceous inhibitor of PAL (Bolwell et al. 1986; Lamb 1977; Mavandad et al. 1990). *In vivo* chemical inhibition of C4H activity has also been reported to result in reduced PAL expression (Bolwell et al. 1988).

Sewalt et al. (1997) showed that C4H downregulation in transgenic tobacco reduced C4H activity up to 20% of control plants and the lines expressing less than 50% C4H activity had significantly reduced lignin level. Blount et al. (2000) reported that tobacco transgenics downregulated for C4H has approximately proportional PAL activity, whereas tobacco transgenics downregulated for PAL alone had no significant reduction in C4H activity. Blount et al. (2000) also found that transgenic tobacco lines over-expressed for PAL but down-regulated for C4H had reduced PAL activity. This

emphasizes the dominant effect of C4H downregulation over PAL over-expression. Downregulation of C4H in transgenics or mutation in C4H gene induces pleiotropic effects on phenolic end-products including reduction in lignin and flavonoides, accumulation of novel compound- cinnamoylmalate not found in wild type plants (Schilmiller et al. 2009). All these data support the view that C4H catalyzes rate limiting step at an early point in the pathway and thus the major determinant of the carbon flux in to the pathway. Hence, C4H may be a suitable candidate gene for down-regulation in *L. leucocephala* and may yield transgenics with significantly reduced lignin level and higher extractability of cellulose from the plants.

6.2 Materials and methods

6.2.1 *L. leucocephala* plant material

L. leucocephala seeds and their source have been discussed in chapter 2, section 2.2. Seeds from the plant were treated and imbibed as per Shaik et al. (2009). Embryo axes from the imbibed seeds were excised just after radicle emergence from the seeds. Excised embryos were inoculated on multiple shoot induction medium for 48 hours (for pre-culture) and then used as explants for transformation.

6.2.2 *Nicotiana tabacum* plant material

Sterile tobacco seeds (*Nicotiana tabacum* var. Anand 119) were germinated on wet sterile paper towel. Germinated seeds were transferred to ½ MS basal medium (Murashige and Skoog 1962) supplemented with 2% sucrose and 1.5% glucose for germination. One month old axenic cultured plant leaves were used as the explant for further transformation experiments.

6.2.3 Agrobacterium strain and binary vector

Agrobacterium tumefaciens strain GV2260 was used in the present study. The binary vector, modified pCAMBIA1301 harboured partial *LlC4H*1 gene in antisense orientation under the control of a constitutive CaMV35S promoter and *Nos* poly A terminator. The modified pCAMBIA1301 (hereafter referred as pCAMBIA1301) also contained kanamycin resistance gene for bacterial selection and hygromycin phosphotranferase II gene (*hpt*II) for plant selection. Besides the vector contained *gus* as reporter gene (See Fig. 6.1).

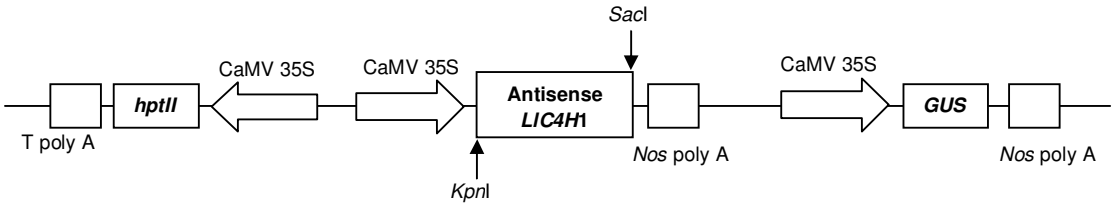


Fig 6.1 Schematic representation of the T-DNA sequence of the modified pCAMBIA 1301 binary vector harbouring partial antisense *LIC4H*1 gene. CaMV 35S: Cauliflower mosaic virus 35S RNA promoter, *hpt*II: hygromycin phosphotranferase gene, *GUS*: β-glucuronidase gene

6.2.4 Vector construct preparation

The precise protocol for vector preparation has been discussed by Gupta (2008). In short, the modified construct prepared by Gupta (2008) was digested by *Kpn*I and *Sac*I and the partial *LI4CL* gene insert was seperated from the vector backbone. Following gel electrophoresis, the vector backbone was recovered from the gel and purified. *Sac*I and *Kpn*I sites were introduced in to the forward and reverse primers (Table 6.1) respectively and 438 bp fragment of *LIC4H*1 was amplified using a plasmid harbouring *LIC4H*1 cDNA clone as template. This 438 bp partial *LIC4H*1 fragment along with its restriction site overhangs was cloned in pGEM T-easy and transformed in competent *E. coli* host strain (XL-1 blue). Plasmid isolated from positive colony was digested with *Kpn*I and *Sac*I and gel electrophoresis was done. The smaller fragment (insert) was then purified from the gel and directionally ligated with the pCAMBIA vector backbone digested earlier using the same restriction enzymes. The ligated product was then cloned in XL-1 blue and positive colonies identified. Plasmid was isolated form a positive colony and was used to transform *Agrobacterium tumefaciens* strain GV2260. Transformed bacterial colony grew on kanamycin and rifampicin containing YEP medium, as *A. tumefaciens* are resistant to rifampicin and the modified pCAMBIA 1301 has kanamycin resistance gene for bacterial selection.

Table 6.1 Primer sequences used to amplify partial *LIC4H* fragment for antisense down-regulation (restriction sites are underlined)

Primer name	Primer sequence (5'→3')
C4H-ANTS-F3	GAGCTCATGGATCTCCTACTCCTGGAG
C4H-ANTS-R3	GGTACCTGTATTGCTGAACCACCTTGT

6.2.5 *A. tumefaciens* transformation protocol and selection

A single colony of *A. tumefaciens* (GV2260) was inoculated in 5 mL LB broth containing Rifampicin 50 µg/mL and incubated at 28 ºC with shaking at 200 rpm for overnight. 1 mL of the culture was used to inoculate 50 mL of YEP broth and grown at the same condition till the OD reaches 0.6. Grown culture were centrifuged at 5000 g for 10 min at 4 ºC and washed twice with ice cold 150 mM $CaCl_2$. The cells were pelleted and resuspended in 1 mL of ice cold 20 mM $CaCl_2$. Aliquots of 100 µL were made. For transformation, 10 µg DNA of the modified pCAMBIA1301 vector was added to an aliquot of the competent *A. tumefaciens* cells and incubated on ice for 30 min. The cells were then snap frozen in liquid nitrogen and allowed to thaw at 37 ºC for 5 min. The freeze thaw process was repeated once. After thawing, 1mL LB medium was added and the tube incubated at 28 ºC for 4 h with gentle shaking. The cells were briefly spun, supernatant thrown retaining 100 µL of volume to resuspend the pellet and plated on LB agar medium with rifampicin (50 µg/mL) and kanamycin (50 µg/mL). Plate incubated at 28 ºC for 48 hours and transformed *Agrobacterium* clones were used to infect explants.

6.2.6 Media composition for *Leucaena* and tobacco tissue culture

Explants were maintained on modified ½ MS medium (Murashige and Skoog 1962) with the addition of different plant hormones for different purposes as explained below. ½ MS basal medium contained half strength of major and minor salts and full strength of vitamins, iron, *myo*-inositol supplemented with 2% sucrose and 1.5% glucose. 0.8% of agar agar was used as gelling material, when required. The pH of the medium was adjusted to 5.8 prior to autoclaving. The culture bottles were incubated at 25±2 ºC and 70% relative humidity under 16 hour photoperiod.

6.2.6.1 Multiple shoot induction medium for *L. leucocephala* embryo

½ Modified MS medium + Thidiazuron (TDZ) at the concentration of 0.5 mg/L

6.2.6.2 Shoot elongation medium for *L. leucocephala*

½ Modified MS medium + 2ip (1 mg/L)

6.2.6.3 Root initiation medium for *L. leucocephala*

½ Modified MS medium + NAA (0.2 mg/L)

6.2.6.4 Tobacco shoot induction medium

½ Modified MS medium + 4.4 µM BAP + 5.37 µM NAA

6.2.6.5 Tobacco root initiation medium

½ Modified MS medium (without hormones)

6.2.7 Coating of plasmid DNA on gold micro-particle

Approximately 60 mg of gold particles of 1 µm size were used for bombardment. The gold particles taken in microcentrifuge tubes were washed with sterile water and supernatant was decanted after centrifuging at 13,000 g for 10 min. DNA coating on the gold particles was carried out following the recommendations and protocol of Sanford et al. (1993), which ensured homogenous coating of the DNA on the particles. Particle bombardment was done using BioRad PDS 1000/He system. Helium gas pressure (1300 psi) was used to bombard the coated plasmid on to the explant. Rupture disk that can withstand up to 1100 psi was used in the process. Microcarrier travel distance was adjusted to 6 cm. The embryo axes were arranged in the centre of the petridishes and bombarded with gold particles coated with plasmid DNA.

6.2.8 Transformation protocol and Multiple shoot induction from *L. leucocephala* embryo

Multiple shoot regeneration from *L. leucocephala* embryo axis was previously optimized in our lab. In short, embryo were excised from the germinated *Leucaena* seeds and kept on multiple shoot induction medium for 48 hours for preculture. For transformation of *Leucaena* embryo, a combined method of particle bombardment followed by co-cultivation with *Agrobacterium tumefaciens* harbouring the recombinant binary vector was used. The embryo axes were first bombarded with gold particles coated with recombinant binary vector as mentioned above followed by co-cultivation with *A. tumefaciens* (GV2260) harboring pCAMBIA1301 containing partial *LIC4H*1 gene in anti-sense orientation. Cultured *Agrobacterium* cells were pelleted by centrifugation at 3,000 g for 5 min, the supernatant was thrown and the cell pellet was resuspended in ½ MS liquid medium. The selected embryo axes after bombardment were co-cultured with *Agrobacterium* for half an hour by immersing them into the bacterial suspension. The agro infected embryo

axes were then transferred onto the multiple shoot induction medium with 0.1 mM acetosyringone (3', 5'-dimethoxy-4'-hydroxy-acetophenone; Sigma, USA) and co-cultivated in the dark at 28 ºC for 2 days. After co-cultivation, the embryo axes were washed thoroughly with cefotaxime 250 mg/L in sterile ½ MS liquid medium and transferred onto fresh multiple shoot induction medium containing cefotaxime (250 mg/L). After one week, explants were again washed with cefotaxime and transferred on fresh multiple shoot induction medium containing cefotaxime and hygromycin (50 mg/L), concentration of which was empirically determined in an earlier experiment to find out the LD_{50} value of untransformed *Leucaena* embryo. The explants were shifted to fresh medium every fortnight. After emergence of multiple shoots from the embryo in nearly one month, they were plucked from the parent embryo and shifted to shoot elongation medium for further period till they reached a height of 4-5 cm. After attaining requisite height, they were shifted to root initiation medium for about 15 days. Rooted plantlets were shifted to soil and kept for hardening. After hardening, they were shifted to green house.

6.2.9 Transformation protocol of tobacco

Tobacco (*Nicotiana tabacum* var. Anand 119) was transformed by modified pCAMBIA1301 vector containing antisense partial *LlC4H*1gene using leaf disc method as described by Horsch et al. (1985). Leaf disks prepared this way were co-cultivated with *Agrobacterium* after 48 hours of pre-culture in tobacco shoot induction medium. LD_{50} of tobacco discs for hygromycin was determined in previous experiments to be 25 mg/L. Tobacco shoots that developed in about 3-4 weeks were excised from the parent tissue and allowed to grow for 12 weeks on the same composition medium with shifting every two weeks. Shoots were then shifted to tobacco root initiation medium. As soon as roots appeared, the rooted shoots were transplanted to soil and kept for hardening and later shifted to green house.

6.2.10 *GUS* assay protocol

GUS assay is a destructive assay to confirm transgenics in which *GUS* gene has been immobilized. *GUS* staining solution was prepared by taking 1 mM X-Gluc (5-bromo, 4chloro, 3-indolyl-β-D-glucuronide: Cyclohexylammonium

salt) from a 20 mM stock made in DMF, 100 mM sodium dihydrogen phosphate dihydrate and 0.5% Tween-20. The pH of the solution was adjusted to 7.0 with 1N NaOH. Control and transgenic leaf tissues or explants were preincubated at 55 °C for 1 h before incubation with the GUS staining solution to heat inactivate the endogenous GUS-like activity. After this the tissues were incubated at 37 °C in dark for overnight, dipped in the assay solution. GUS gene expression was distinctly observed in the transformed tissue as compared to the non-transformed one by bleaching in 70% alcohol for 3-4 h, later replaced by absolute ethanol and photographed by using an Axioplan 4 microscope from Carl Zeiss.

6.2.11 Selection for transformants

Putative transformants for tobacco were selected on the basis of *GUS* assay taking a small leaf/or small portion of leaf from the regenerated shoot of tobacco. *Leucaena* putative transformants were selected only at the time of being kept for hardening. Transformants were confirmed on the basis of PCR using *hpt*II specific primers and gDNA isolated from the putative transformants as template. The list of all the primers used for the selection of transformants and for the Q-RT PCR analysis performed in this chapter is given in Table 6.2.

6.2.12 Transcript abundance study of phenylpropanoid pathway genes in transgenics down-regulated for *C4H*

Petioles of transgenic tobacco and three control plants, emerging from nodes 8-10 from the first fully opened leaf at the top was crushed in liquid nitrogen and RNA was isolated from the tissue. First strand cDNA was prepared as mentioned in chapter 2, section 2.5.7 and used as template in quantitative real time PCR. Transcript level of *C4H*, *PAL*, *4CL*, *CAD* and *CCoAOMT* were averaged form the three control plants. Transcript level of *C4H*, *PAL*, *4CL*, *CAD* and *CCoAOMT* were determined from transgenics and expressed as the percentage transcript level of control plants. Primer sequences specific to tobacco gene sequences were designed by retrieving the specific gene sequences from NCBI database. Genbank accession numbers DQ350352, DQ350353 and DQ350354 were used to design tobacco specific *C4H*

primers; AB008199, AB008200, D17467 and EU883670 for tobacco specific *PAL* primers; U50845 and U50846 for tobacco specific *4CL* primers; X62343 and X62344 for tobacco specific *CAD* primers; and U38612, U62734, U62735 and U62736 for tobacco specific *CCoAOMT* primers. Primer sequences used for Q-RT PCR in this chapter are given in Table 6.2 and optimization of Q-RT PCR has been discussed in chapter 2, section 2.5.9.

Table 6.2 List and sequence of Q-RT PCR primers used in this chapter

Name of the primer	Sequence of the primer (5' → 3')
hptII-F	GCGTCGGTTTCCACTATCGG
hptII-R	TCTCGGAGGGCGAAGAATCT
Tob-*C4H*-F	GTGGCAAGTGTAATTGAGGATGTG
Tob-*C4H*-R	GCCAATCTACTCCTTTCACCATTC
Tob-*PAL*-F	CGGTGGATTTTTCGAGTTGCAGCC
Tob-*PAL*-R	TGAGCCGCCTTCACATAAGAGCT
Tob-*4CL*-F	GGTTACACACTGGCGACATTGG
Tob-*4CL*-R	GGAACTTCTCCTGCTTGCTCATC
Tob-*CAD*-F	GGGAGTGAAAATAGCAAAGG
Tob-*CAD*-R	GCCAACAGGGACAGTATC
Tob-*CCoAOMT*-F	ACACCCTATGGAATGGATCAG
Tob-*CCoAOMT*-R	GCCTTGTTGAGTTCCAATACG

6.2.13 Lignin estimation from stem tissues and phloroglucinol staining

Phloroglucinol staining protocol has been discussed in chapter 2, section 2.9. Images were captured using an Axioplan 4 microscope from Carl Zeiss. Lignin from stem tissues of transgenics and control plants were estimated using the protocol of Kirk and Obst (1988) with slight modification which has been discussed in chapter 2, section 2.10.

6.3 Results

6.3.1 Transformation of *Leucaena* embryo

Leucaena was stably transformed using the protocols mentioned above, but the growth of *Leucaena in vitro* was very slow compared to its growth *ex vitro*. Control embryo gave about 5-6 shoots per explant in the multiple shoot induction medium, which were, if not plucked from the parent tissue and frequently recultured to medium devoid of TDZ, failed to elongate. *Leucaena* transgenics were stubbornly slow growing (See Fig. 6.1).

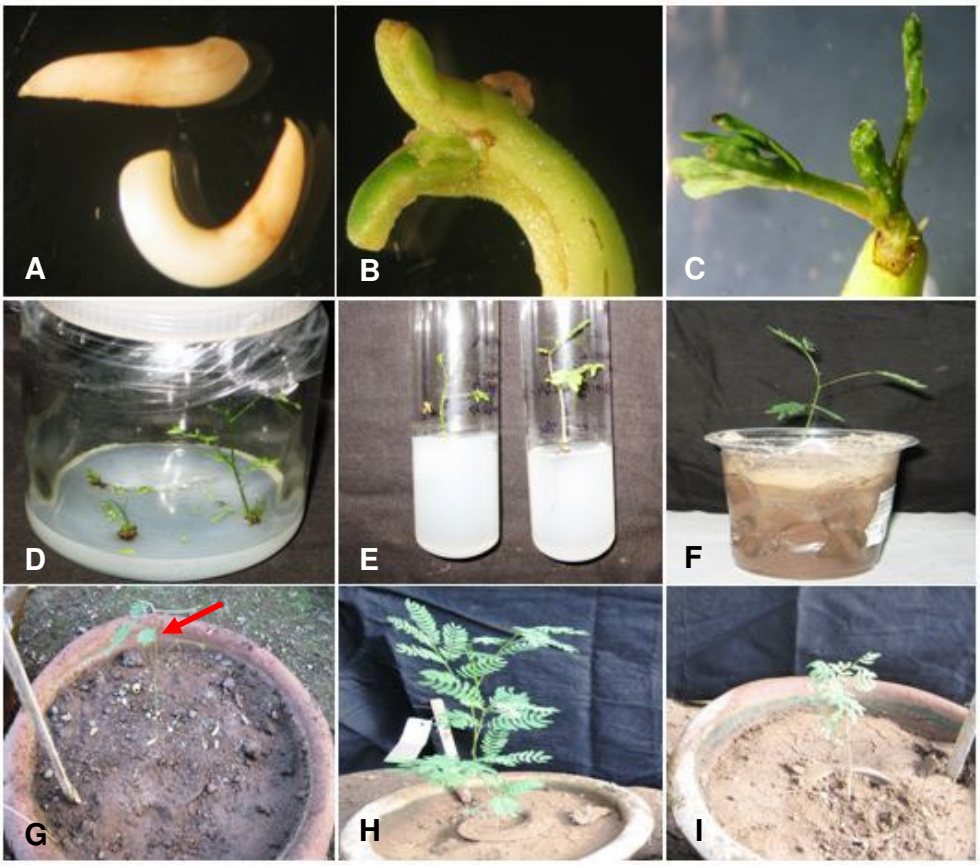


Fig. 6.1 Transformation of *Leucaena* embryo (A) *Leucaena* embryo excised from the seeds; (B) *Leucaena* embryo 3 days after bombardment and co-cultivation; (C) Emergence of multiple shoots from *Leucaena* embryo; (D) Individual shoots growing in a glass bottle; (E) Shoots in the root induction medium; (F) Rooted plantlet transplanted to soil; (G) Rooted *Leucaena* plant in green house; (H) Control *Leucaena* plant in green house; (I) Transformed *Leucaena* plant in green house

6.3.2 Transformation of tobacco leaf disc

Tobacco explants gave rise to transgenics frequently, faster and much easily than *Leucaena* explants and hence, analysis of transgenic tobacco in greater depths became possible than transgenic *Leucaena*. Fig. 6.2 shows the sequential events in the regeneration of tobacco.

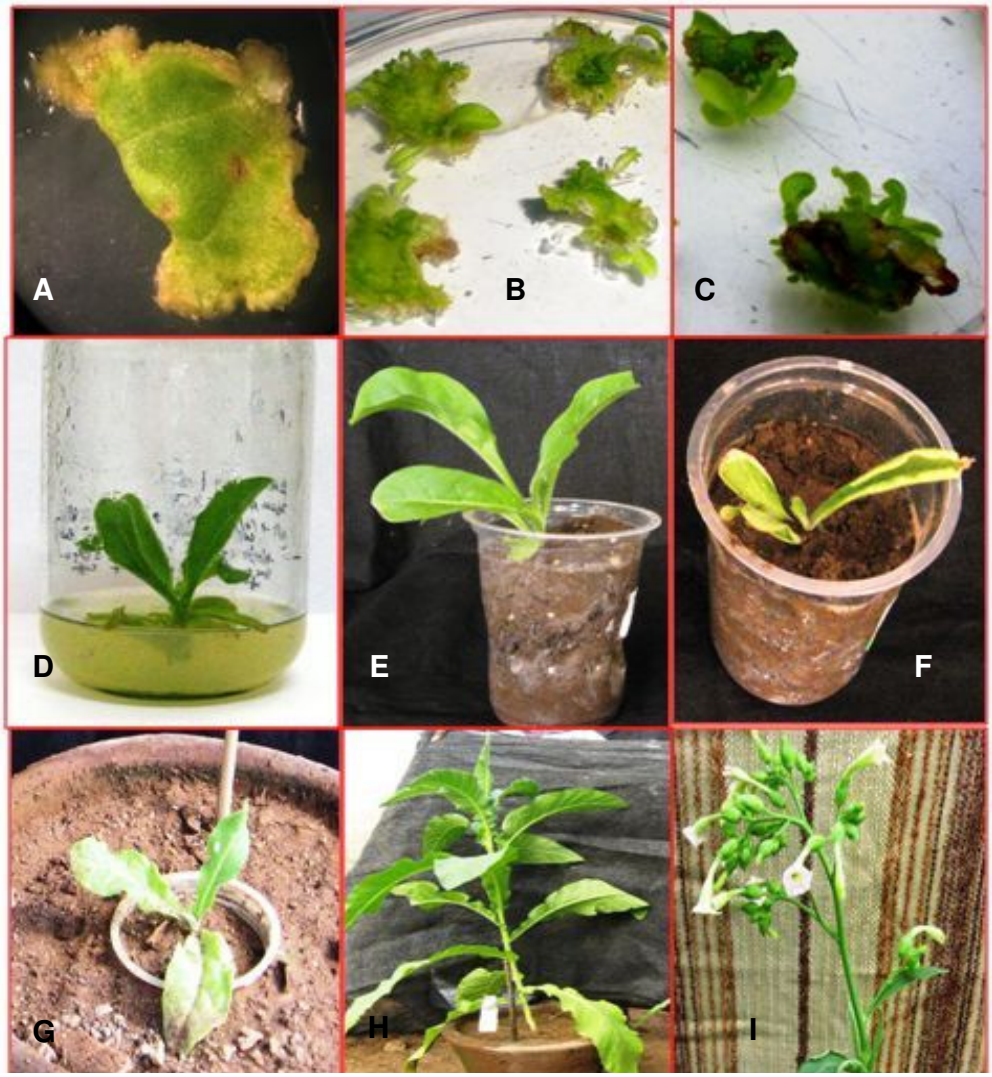


Fig. 6.2 Transformation of tobacco explants (A) Tobacco leaf disc 15 days after co-cultivation; (B) Shoot emergence from control leaf discs; (C) Shoot emergence from co-cultivated leaf discs on hygromycin containing selection medium; (D) Tobacco shoot growing in a bottle; (E) Control rooted shoot transplanted in soil; (F) A transgenic tobacco plant transplanted in soil; (G) A tobacco plant shifted to green house; (H) A mature tobacco plant; (I) Tobacco inflorescence

6.3.3 *GUS* assay

To look for transient *gus* assay in bombarded and co-cultivated *Leucaena* embryo, gus assay taking *Leucaena* embryo was performed 5 days after co-cultivation. Tobacco leaves were also subjected to *gus* assay to search for putative transformants. Some of the photographs are summarised in Fig 6.3.

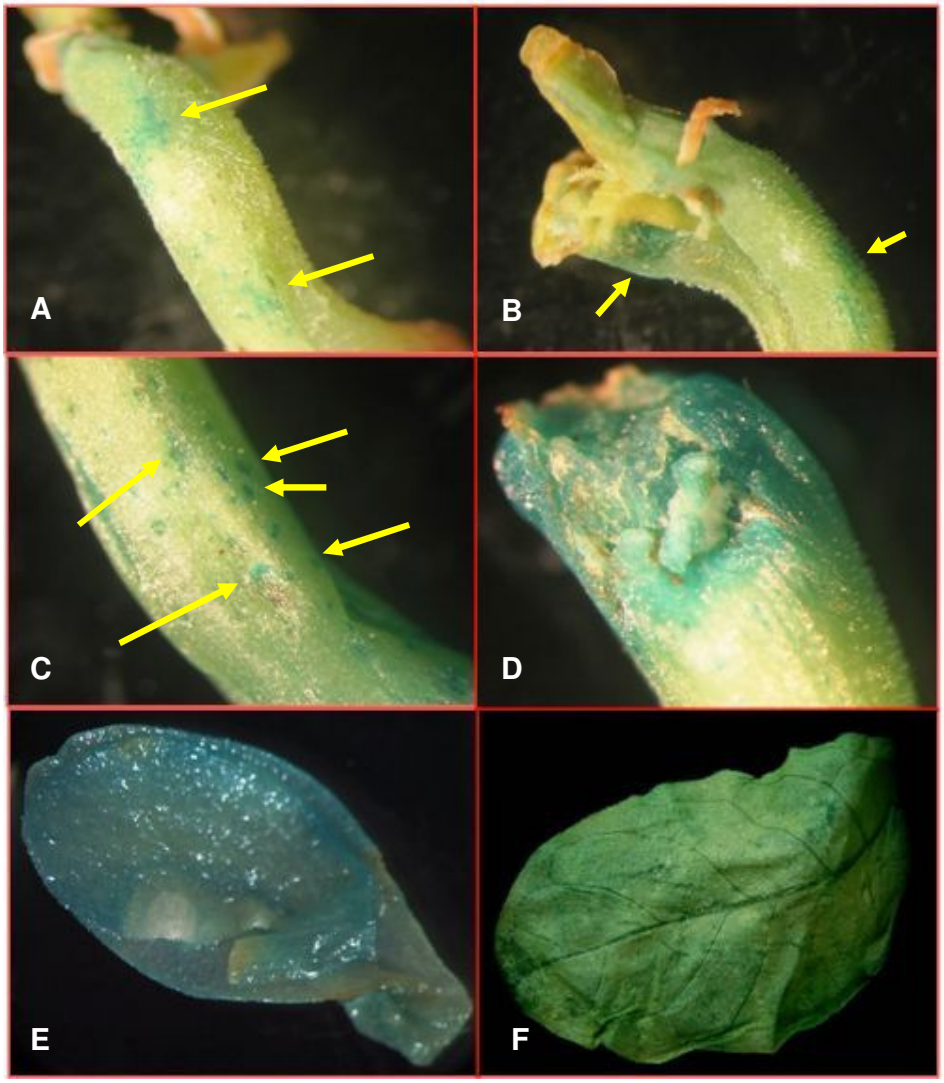


Fig. 6.3 *GUS* assay photographs of *Leucaena* embryo and tobacco leaves (A, B, C and D) *Leucaena* embryo 5 days after bombardment and co-cultivation. Blue spots characteristic of *gus* assay can be clearly seen (see arrow positions). At some places, spots are concentrated and at some places they are widespread (see arrows). (E) *GUS* assay of tobacco leaf at young stage and (F) *GUS* assay of tobacco leaf at transplanting stage

6.3.4 Screening of transgenics using PCR

Transgenics were identified on the basis of PCR done with *hpt*II specific primers and genomic DNA isolated from the putative transgenics as template. True transgenics gave expected size amplification of ~900 bp, which was gel purified, ligated to pGEM-T Easy vector, cloned and sequenced (See Fig. 6.4). Every putative transgenic that gave the amplification of ~900 bp was sequenced this way. The sequence got after amplification with the primers was matched with the *hpt*II sequence of pCAMBIA vectors available in NCBI database to reach to any conclusion.

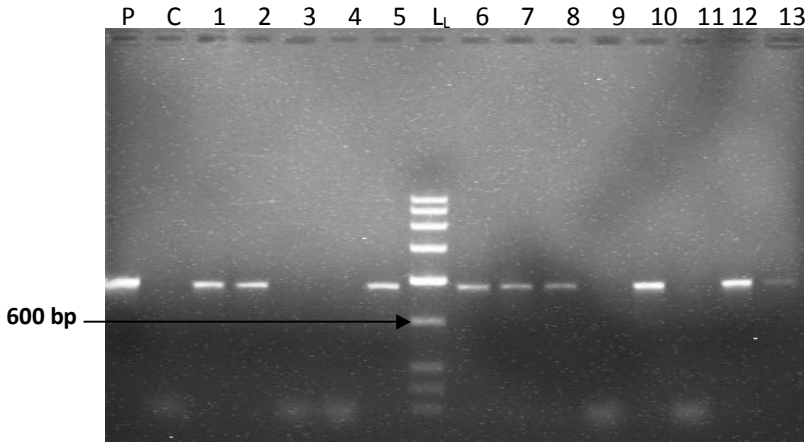


Fig. 6.4 Gel photograph showing ~900 bp fragment amplified using *hpt*II specific primers and genomic DNA isolated from putative transgenics as template. Lane P= recombinant pCAMBIA1301 plasmid as template serving as positive control and reference size; C= control (non-transformed) plant; L_L= low range ladder; Lane 1-5 putative transgenics of *Leucaena* and Lane 6-13 putative transgenics of tobacco

Alignment of *hpt*II sequence of the PCR amplicons and *hpt*II sequence of pCAMBIA1301 available on internet has been shown in Fig. 6.5. As can be seen, there is a difference of one nucleotide in the two sequences, possibly due to replication error during any of the DNA manipulation steps. This resulted in a mis-sense mutation and converted alanine to serine, but supposedly, this did not affect the catalytic property of hygromycin phosphotransferase, as there are some binary vectors in which alanine is replaced by serine in catalytically active hygromycin phosphotransferase

enzyme. These enzymes show 100% sequence identity at amino acid level with the sequence or our hptII enzyme. For comparison between the two sequences at amino acid level, see Fig. 6.6(A) and 6.6(B).

Aligned_sequences: 2
1: hpt_transgenics
2: hpt_pCAMBIA1301
Matrix: EDNAFULL
Length: 1026
Identity: 913/1026 (89.0%)
Similarity: 913/1026 (89.0%)
Gaps: 112/1026 (10.9%)
Score: 4561.0

```
hpt_transgenics      1 --------------------------------GCGTCGGTTTCCACTATCGG     20
                                                       ||||||||||||||||||||
hpt_pCAMBIA1301      1 CTATTTCTTTGCCCTCGGACGAGTGCTGGGGCGTCGGTTTCCACTATCGG     50

hpt_transgenics     21 CGAGTACTTCTACACAGCCATCGGTCCAGACGGCCGCGCTTCTGCGGGCG     70
                       ||||||||||||||||||||||||||||||||||||||||||||||||||||
hpt_pCAMBIA1301     51 CGAGTACTTCTACACAGCCATCGGTCCAGACGGCCGCGCTTCTGCGGGCG    100

hpt_transgenics     71 ATTTGTGTACGCCCGACAGTCCCGGCTCCGGATCGGACGATTGCGTCGCA    120
                       ||||||||||||||||||||||||||||||||||||||||||||||||||||
hpt_pCAMBIA1301    101 ATTTGTGTACGCCCGACAGTCCCGGCTCCGGATCGGACGATTGCGTCGCA    150

hpt_transgenics    121 TCGACCCTGCGCCCAAGCTGCATCATCGAAATTGCCGTCAACCAAGCTCT    170
                       ||||||||||||||||||||||||||||||||||||||||||||||||||||
hpt_pCAMBIA1301    151 TCGACCCTGCGCCCAAGCTGCATCATCGAAATTGCCGTCAACCAAGCTCT    200

hpt_transgenics    171 GATAGAGTTGGTCAAGACCAATGCGGAGCATATACGCCCGGAGTCGTGGC    220
                       ||||||||||||||||||||||||||||||||||||||||||||||||||||
hpt_pCAMBIA1301    201 GATAGAGTTGGTCAAGACCAATGCGGAGCATATACGCCCGGAGTCGTGGC    250

hpt_transgenics    221 GATCCTGCAAGCTCCGGATGCCTCCGCTCGAAGTAGCGCGTCTGCTGCTC    270
                       ||||||||||||||||||||||||||||||||||||||||||||||||||||
hpt_pCAMBIA1301    251 GATCCTGCAAGCTCCGGATGCCTCCGCTCGAAGTAGCGCGTCTGCTGCTC    300

hpt_transgenics    271 CATACAAGCCAACCACGGCCTCCAGAAGAAGATGTTGGCGACCTCGTATT    320
                       ||||||||||||||||||||||||||||||||||||||||||||||||||||
hpt_pCAMBIA1301    301 CATACAAGCCAACCACGGCCTCCAGAAGAAGATGTTGGCGACCTCGTATT    350

hpt_transgenics    321 GGGAATCCCCGAACATCGCCTCGCTCCAGTCAATGACCGCTGTTATGCGG    370
                       ||||||||||||||||||||||||||||||||||||||||||||||||||||
hpt_pCAMBIA1301    351 GGGAATCCCCGAACATCGCCTCGCTCCAGTCAATGACCGCTGTTATGCGG    400

hpt_transgenics    371 CCATTGTCCGTCAGGACATTGTTGGAGCCGAAATCCGCGTGCACGAGGTG    420
                       ||||||||||||||||||||||||||||||||||||||||||||||||||||
hpt_pCAMBIA1301    401 CCATTGTCCGTCAGGACATTGTTGGAGCCGAAATCCGCGTGCACGAGGTG    450

hpt_transgenics    421 CCGGACTTCGGGGCAGTCCTCGGCCCAAAGCATCAGCTCATCGAGAGCCT    470
                       ||||||||||||||||||||||||||||||||||||||||||||||||||||
hpt_pCAMBIA1301    451 CCGGACTTCGGGGCAGTCCTCGGCCCAAAGCATCAGCTCATCGAGAGCCT    500

hpt_transgenics    471 GCGCGACGGACGCACTGACGGTGTCGTCCATCACAGTTTGCCAGTGATAC    520
                       ||||||||||||||||||||||||||||||||||||||||||||||||||||
hpt_pCAMBIA1301    501 GCGCGACGGACGCACTGACGGTGTCGTCCATCACAGTTTGCCAGTGATAC    550

hpt_transgenics    521 ACATGGGGATCAGCAATCGCGCATATGAAATCACGCCATGTAGTGTATTG    570
                       ||||||||||||||||||||||||||||||||||||||||||||||||||||
hpt_pCAMBIA1301    551 ACATGGGGATCAGCAATCGCGCATATGAAATCACGCCATGTAGTGTATTG    600

hpt_transgenics    571 ACCGATTCCTTGCGGTCCGAATGGGCCGAACCCGCTCGTCTGGCTAAGAT    620
                       ||||||||||||||||||||||||||||||||||||||||||||||||||||
hpt_pCAMBIA1301    601 ACCGATTCCTTGCGGTCCGAATGGGCCGAACCCGCTCGTCTGGCTAAGAT    650
```

```
hpt_transgenics     621 CGGCCGCAGCGATCGCATCCATAGCCTCCGCGACCGGTTGTAGAACAGCG    670
                        ||||||||||||||||||||||||||||||||||||||||||||||||||||
hpt_pCAMBIA1301     651 CGGCCGCAGCGATCGCATCCATAGCCTCCGCGACCGGTTGTAGAACAGCG    700

hpt_transgenics     671 GGCAGTTCGGTTTCAGGCAGGTCTTGCAACGTGACACCCTGTGAACGGCG    720
                        ||||||||||||||||||||||||||||||||||||||||||||||.||||||
hpt_pCAMBIA1301     701 GGCAGTTCGGTTTCAGGCAGGTCTTGCAACGTGACACCCTGTGCACGGCG    750

hpt_transgenics     721 GGAGATGCAATAGGTCAGGCTCTCGCTAAACTCCCCAATGTCAAGCACTT    770
                        ||||||||||||||||||||||||||||||||||||||||||||||||||||
hpt_pCAMBIA1301     751 GGAGATGCAATAGGTCAGGCTCTCGCTAAACTCCCCAATGTCAAGCACTT    800

hpt_transgenics     771 CCGGAATCGGGAGCGCGGCCGATGCAAAGTGCCGATAAACATAACGATCT    820
                        ||||||||||||||||||||||||||||||||||||||||||||||||||||
hpt_pCAMBIA1301     801 CCGGAATCGGGAGCGCGGCCGATGCAAAGTGCCGATAAACATAACGATCT    850

hpt_transgenics     821 TTGTAGAAACCATCGGCGCAGCTATTTACCCGCAGGACATATCCACGCCC    870
                        ||||||||||||||||||||||||||||||||||||||||||||||||||||
hpt_pCAMBIA1301     851 TTGTAGAAACCATCGGCGCAGCTATTTACCCGCAGGACATATCCACGCCC    900

hpt_transgenics     871 TCCTACATCGAAGCTGAAAGCACGAGATTCTTCGCCCTCCGAGA------    914
                        ||||||||||||||||||||||||||||||||||||||||||||
hpt_pCAMBIA1301     901 TCCTACATCGAAGCTGAAAGCACGAGATTCTTCGCCCTCCGAGAGCTGCA    950

hpt_transgenics     915 --------------------------------------------------    914

hpt_pCAMBIA1301     951 TCAGGTCGGAGACGCTGTCGAACTTTTCGATCAGAAACTTCTCGACAGAC   1000

hpt_transgenics     915 ---------------------------    914

hpt_pCAMBIA1301    1001 GTCGCGGTGAGTTCAGGCTTTTTCAT   1026
```

Fig. 6.5 Result of alignment of the two sequences at nucleotide level. The nucleotide difference has been shaded green

```
>gb|AAG38025.1|AF294976_1  hygromycin phosphotransferase [Cloning vector pC1300intA]
 gb|AAG38026.1|AF294977_1  hygromycin phosphotransferase [Cloning vector pC1300intB]
 gb|AAG38027.1|AF294978_1  hygromycin phosphotransferase [Cloning vector pC1300intC]
 ▶80 more sequence titles
 Length=341

 Score =  630 bits (1625),  Expect = 0.0, Method: Compositional matrix adjust.
 Identities = 303/304 (99%), Positives = 304/304 (100%), Gaps = 0/304 (0%)

Query  1    SEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAALPIPEVLDIGEFSESLT  60
            SEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAALPIPEVLDIGEFSESLT
Sbjct  29   SEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAALPIPEVLDIGEFSESLT  88

Query  61   YCISRRSQGVTLQDLPETELPAVLQPVAEAMDAIAAADLSQTSGFGPFGPQGIGQYTTWR  120
            YCISRRSQGVTLQDLPETELPAVLQPVAEAMDAIAAADLSQTSGFGPFGPQGIGQYTTWR
Sbjct  89   YCISRRAQGVTLQDLPETELPAVLQPVAEAMDAIAAADLSQTSGFGPFGPQGIGQYTTWR  148

Query  121  DFICAIADPHVYHWQTVMDDTVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTDN  180
            DFICAIADPHVYHWQTVMDDTVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTDN
Sbjct  149  DFICAIADPHVYHWQTVMDDTVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTDN  208

Query  181  GRITAVIDWSEAMFGDSQYEVANIFFWRPWLACMEQQTRYFERRHPELAGSPRLRAYMLR  240
            GRITAVIDWSEAMFGDSQYEVANIFFWRPWLACMEQQTRYFERRHPELAGSPRLRAYMLR
Sbjct  209  GRITAVIDWSEAMFGDSQYEVANIFFWRPWLACMEQQTRYFERRHPELAGSPRLRAYMLR  268

Query  241  IGLDQLYQSLVDGNFDDAAWAQGRCDAIVRSGAGTVGRTQIARRSAAVWTDGCVEVLADS  300
            IGLDQLYQSLVDGNFDDAAWAQGRCDAIVRSGAGTVGRTQIARRSAAVWTDGCVEVLADS
Sbjct  269  IGLDQLYQSLVDGNFDDAAWAQGRCDAIVRSGAGTVGRTQIARRSAAVWTDGCVEVLADS  328

Query  301  GNRR  304
            GNRR
Sbjct  329  GNRR  332                                                    A
```

Fig. 6.6(A) NCBI blast showing one amino acid difference between conceptually translated *hpt*II sequence isolated from transgenics and hptII sequence of pCAMBIA1301 available on net. The difference is shaded green

```
>gb|ABE03626.1|  hygromycin phosphotransferase [Binary vector pGPro1]
 gb|ABV71995.1|  hygromycin phosphotransferase [Binary vector pGPro2]
 gb|AEG42343.1|  hygromycin phosphotransferase [Binary vector pCMIRT15]
 gb|AEJ22718.1|  HptII [cloning vector pSOL8DCL2]
 gb|AEJ22722.1|  HptII [cloning vector pSOL9CAP]
 gb|AEJ22727.1|  HptII [cloning vector pSOL8DC3]
Length=341

 Score =  632 bits (1629),  Expect = 0.0, Method: Compositional matrix adjust
 Identities = 304/304 (100%), Positives = 304/304 (100%), Gaps = 0/304 (0%)

Query  1    SEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAALPIPEVLDIGEFSESLT  60
            SEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAALPIPEVLDIGEFSESLT
Sbjct  29   SEGEESRAFSFDVGGRGYVLRVNSCADGFYKDRYVYRHFASAALPIPEVLDIGEFSESLT  88

Query  61   YCISRRSQGVTLQDLPETELPAVLQPVAEAMDAIAAADLSQTSGFGPFGPQGIGQYTTWR  120
            YCISRRSQGVTLQDLPETELPAVLQPVAEAMDAIAAADLSQTSGFGPFGPQGIGQYTTWR
Sbjct  89   YCISRRSQGVTLQDLPETELPAVLQPVAEAMDAIAAADLSQTSGFGPFGPQGIGQYTTWR  148

Query  121  DFICAIADPHVYHWQTVMDDTVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTDN  180
            DFICAIADPHVYHWQTVMDDTVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTDN
Sbjct  149  DFICAIADPHVYHWQTVMDDTVSASVAQALDELMLWAEDCPEVRHLVHADFGSNNVLTDN  208

Query  181  GRITAVIDWSEAMFGDSQYEVANIFFWRPWLACMEQQTRYFERRHPELAGSPRLRAYMLR  240
            GRITAVIDWSEAMFGDSQYEVANIFFWRPWLACMEQQTRYFERRHPELAGSPRLRAYMLR
Sbjct  209  GRITAVIDWSEAMFGDSQYEVANIFFWRPWLACMEQQTRYFERRHPELAGSPRLRAYMLR  268

Query  241  IGLDQLYQSLVDGNFDDAAWAQGRCDAIVRSGAGTVGRTQIARRSAAVWTDGCVEVLADS  300
            IGLDQLYQSLVDGNFDDAAWAQGRCDAIVRSGAGTVGRTQIARRSAAVWTDGCVEVLADS
Sbjct  269  IGLDQLYQSLVDGNFDDAAWAQGRCDAIVRSGAGTVGRTQIARRSAAVWTDGCVEVLADS  328

Query  301  GNRR  304
            GNRR
Sbjct  329  GNRR  332                                                    B
```

Fig. 6.6(B) NCBI blast showing 100% identity between conceptually translated *hpt*II sequence isolated from transgenics and hptII sequence of some binary vectors available on net. Identity % shown above is shaded green

6.3.5 Difference in the lignin staining pattern of tobacco control and a transgenic plant

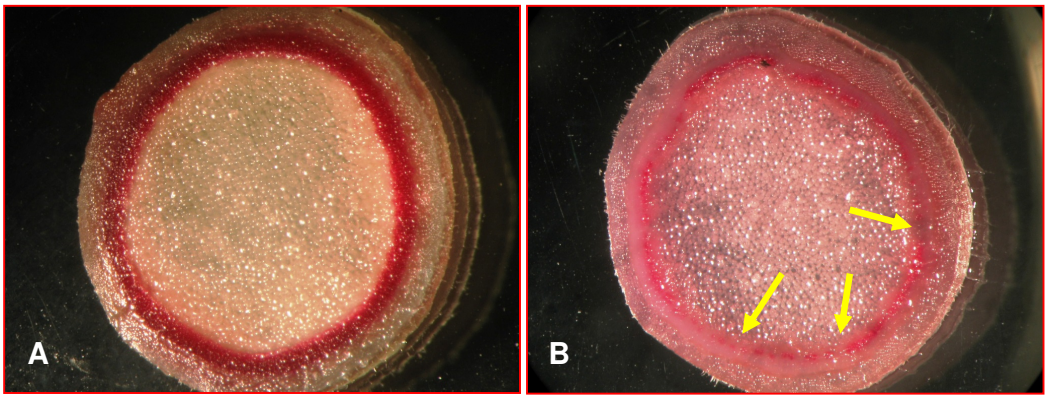


Fig. 6.7 Phloroglucinol staining of 2-3 mm thick stem sections of tobacco: (A) control plant; (B) transgenic plant down-regulated for C4H activity

The sections clearly showed differences in staining pattern of control and transgenic stems. Acidic phloroglucinol stains lignin red. While sections from control plants clearly and uniformly stained dark red (Fig. 6.7A), sections from transgenic plants down-regulated for C4H activity stained pink in colour with non-uniform staining (Fig. 6.7B) suggesting both decrease in lignin content and alteration in its composition. Gaps are visible in the vascular ring of transgenic tobacco showing zone of little lignification (see arrow direction).

Phloroglucinol staining of free hand sections taken from a control and a transgenic, severely reduced for C4H activity, also showed remarkable differences in the accumulated lignin in xylem vessels (Fig. 6.8). Lignin staining of transgenic AS1 showed that at some places, xylem vessels were devoid of lignin. Other transgenics considerably reduced for C4H activity also showed similar pattern of lignification. Transgenics with marginally reduced C4H activity had similar level of stainable lignin as that of a control plant.

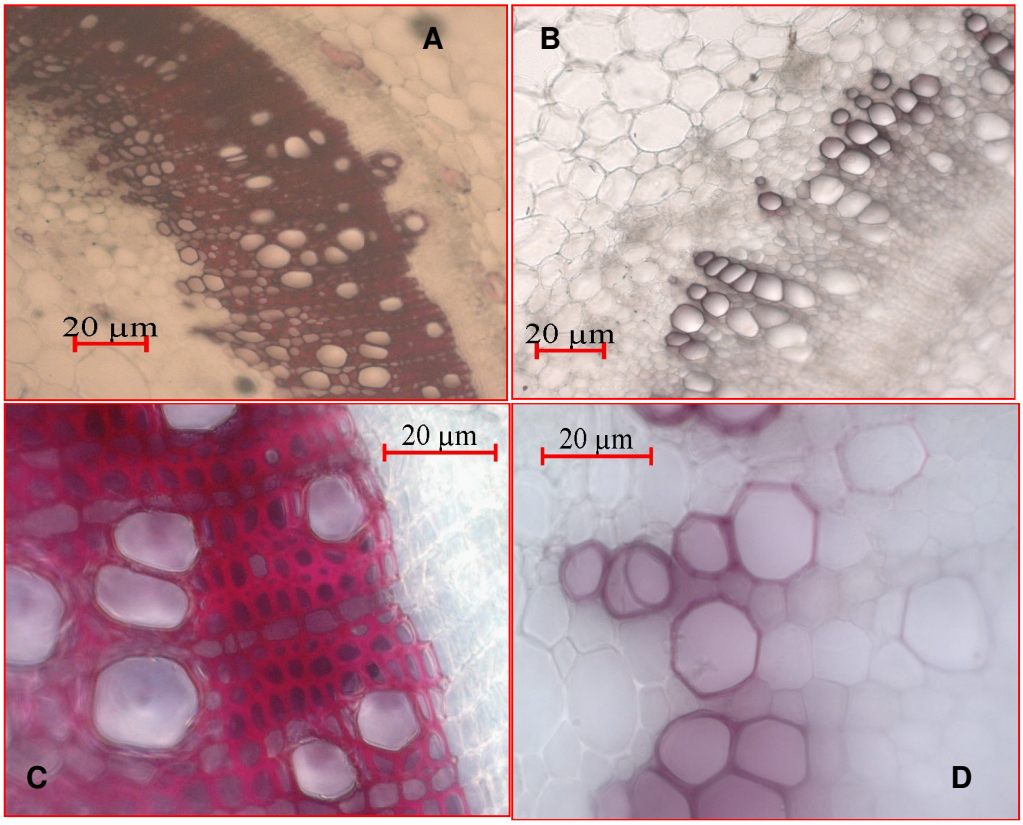


Fig. 6.8 Photographs of phloroglucinol staining of free hand sections of a control and a transgenic, severely reduced for C4H activity, at different magnifications. (A) control plant at 10X; (B) transgenic AS1 at 10X; (C) control at 40X; and (D) transgenic AS1 at 40X

6.3.6 Variation in phenotype between control and transgenic tobacco

Tobacco transgenics down-regulated for C4H were initially slow growing in nature. Some of the transgenics died in the course of shifting them to green house or after being there in green house for about 2 of weeks. Fig. 6.9 shows one such transgenic plant which died in the green house. The leaves of the transgenic plant are curled inward and their tips are brown. Root system of the transgenic was very poorly developed (not visible in the figure).

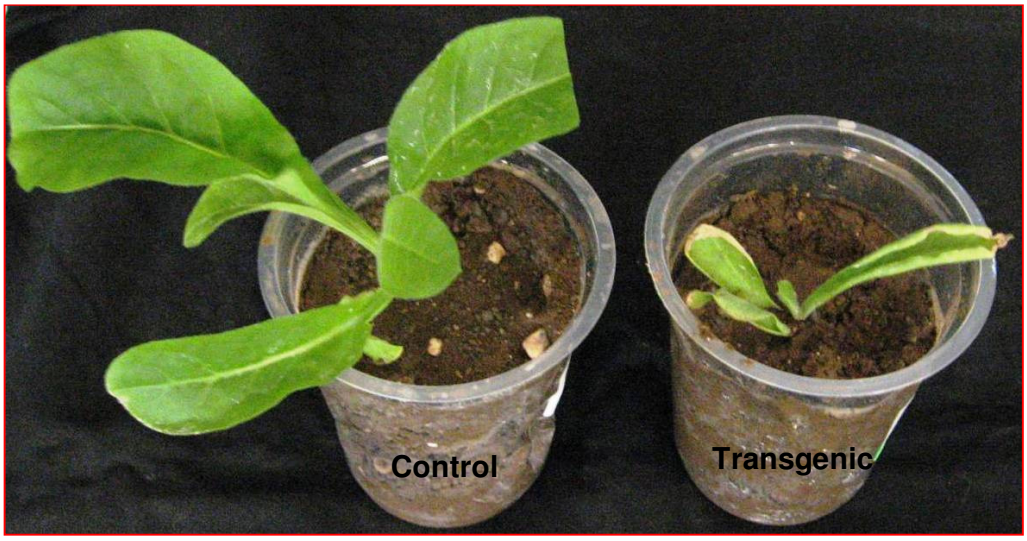


Fig. 6.9 Phenotypic differences between a control and a transgenic tobacco. This transgenic plant failed to grow and died after some time

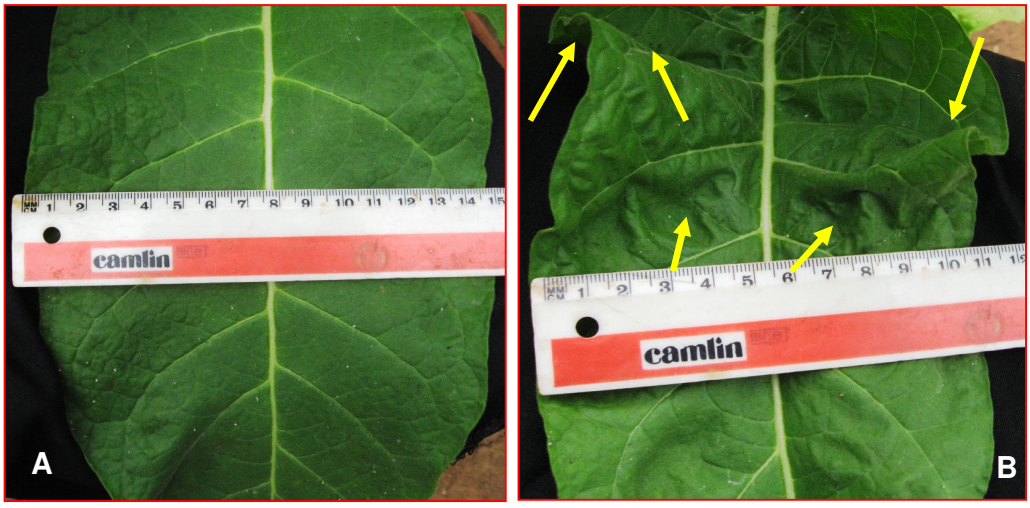


Fig. 6.10 Morphological difference between a control (A); and a mature transgenic leaf (B)

Fig. 6.10 clearly shows morphological differences between a control and a mature transgenic leaf. Control plant's leaf is flat whereas mature transgenic leaf has undulating surface (see arrow head). Leaves from only few transgenics showed this morphology.

6.3.7 Phenotypic difference between control and transgenic *Leucaena*

Unlike the observation in tobacco, where transgenics were initially slow growing which later caught the control plants in growth (unless severely reduced for C4H activity or they died prematurely presumably because of the integration of the transgene within a vital gene); *Leucaena* transgenics and control plants both showed almost same pace of growth until transplanted to soil. After 3 months of hardening, the control plant's growth was slightly better than the transgenics (Fig. 6.11A). After being shifted to green house, while control plants started gaining vigour, growth of transgenics still remained poor (Fig. 6.11B). In fact, the growth of transgenic *Leucaena* plants is so slow that it is currently not possible to carry out their analysis at this stage. *L. leucocephala* transgenics will later be analyzed and characterized for their lignin content, modified lignin composition and structure, extractable C4H activity, impact of C4H down-regulation on other phenylpropanoid pathway genes and other relevant aspects and may be utilized in planting forests for paper industry, contingent upon the approval of government of India.

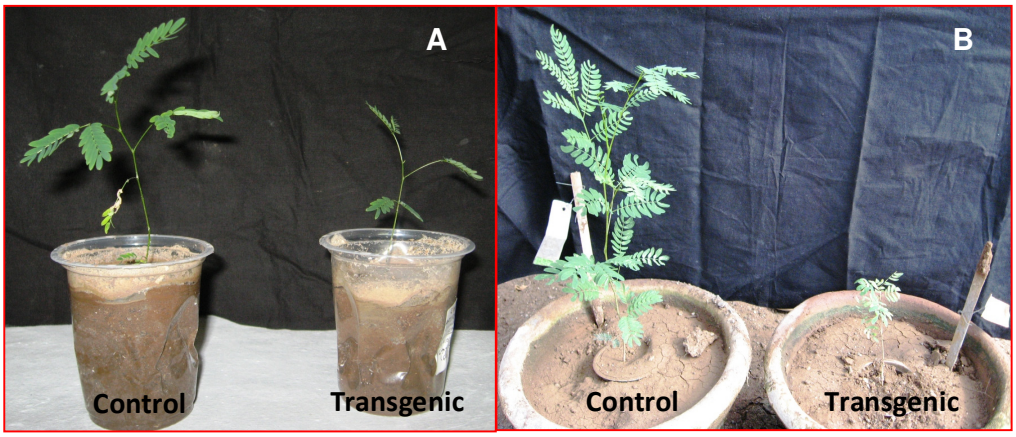
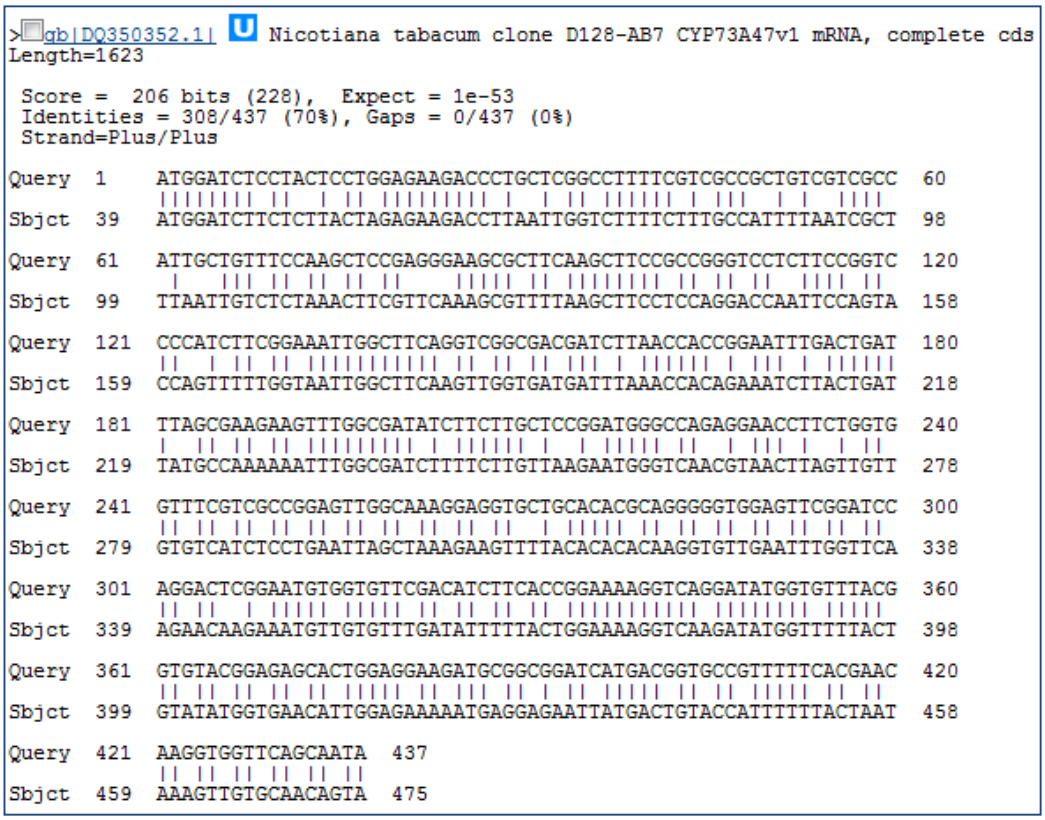


Fig. 6.11 Phenotypic differences between control and transgenic *Leucaena* at different stages. (A) At the time of shifting to green house; (B) 3 months after shifting to green house

6.3.8 Sequence identity between *C4H* from *Leucaena* and *C4H* from tobacco

Three highly similar isoforms of *C4H* are reported from tobacco (NCBI accession numbers DQ350352, DQ350354 and DQ350353 respectively). The partial *C4H* nucleotide sequence from *Leucaena* used to down-regulate C4H activity in transgenic tobacco shares 70%-71% identity with tobacco *C4H* sequences available on internet (Fig. 6.12). Apart from these three isoforms, two additional wound inducible *C4H* sequences from tobacco are also reported (NCBI accession numbers AF368379 and AF368378). These two sequences share more than 50% identity with the partial *Leucaena C4H* sequence. *Leucaena C4H* sequence does not share significant similarity with any other known gene sequence so far in tobacco. Thus, it may be expected that down-regulation of C4H in tobacco using *Leucaena C4H* will not directly influence the transcript level of any other transcript in tobacco.



```
>gb|DQ350352.1| Nicotiana tabacum clone D128-AB7 CYP73A47v1 mRNA, complete cds
Length=1623

 Score =  206 bits (228),  Expect = 1e-53
 Identities = 308/437 (70%), Gaps = 0/437 (0%)
 Strand=Plus/Plus

Query  1    ATGGATCTCCTACTCCTGGAGAAGACCCTGCTCGGCCTTTTCGTCGCCGCTGTCGTCGCC  60
            ||||||||  ||  | ||  |||||||||  |  |  ||  ||||||  |  |||  |  |  ||||
Sbjct  39   ATGGATCTTCTCTTACTAGAGAAGACCTTAATTGGTCTTTTCTTTGCCATTTTAATCGCT  98

Query  61   ATTGCTGTTTCCAAGCTCCGAGGGAAGCGCTTCAAGCTTCCGCCGGGTCCTCTTCCGGTC  120
             |   |||  ||  ||  ||  ||     |||||  ||  ||||||||  ||  ||  ||   ||||  ||
Sbjct  99   TTAATTGTCTCTAAACTTCGTTCAAAGCGTTTTAAGCTTCCTCCAGGACCAATTCCAGTA  158

Query  121  CCCATCTTCGGAAATTGGCTTCAGGTCGGCGACGATCTTAACCACCGGAATTTGACTGAT  180
            ||  |  ||  ||  ||||||||||||  ||  ||  ||  |||  |  |||||||  |  |||  |  ||||||
Sbjct  159  CCAGTTTTTGGTAATTGGCTTCAAGTTGGTGATGATTTAAACCACAGAAATCTTACTGAT  218

Query  181  TTAGCGAAGAAGTTTGGCGATATCTTCTTGCTCCGGATGGGCCAGAGGAACCTTCTGGTG  240
            |  ||  ||  ||  ||||||||||  |  ||||||  |   |  |||||  ||   |  |||  |   |  ||
Sbjct  219  TATGCCAAAAAATTTGGCGATCTTTTCTTGTTAAGAATGGGTCAACGTAACTTAGTTGTT  278

Query  241  GTTTCGTCGCCGGAGTTGGCAAAGGAGGTGCTGCACACGCAGGGGGTGGAGTTCGGATCC  300
            ||  ||  ||  ||  ||  ||  ||  ||  ||  ||   |  |||||  ||  ||  ||  ||  ||  ||
Sbjct  279  GTGTCATCTCCTGAATTAGCTAAAGAAGTTTTACACACACAAGGTGTTGAATTTGGTTCA  338

Query  301  AGGACTCGGAATGTGGTGTTCGACATCTTCACCGGAAAAGGTCAGGATATGGTGTTTACG  360
            ||  ||   |  |||||  |||||  ||  ||  ||  ||  ||||||||||||  ||||||||  |||||
Sbjct  339  AGAACAAGAAATGTTGTGTTTGATATTTTTACTGGAAAAGGTCAAGATATGGTTTTTACT  398

Query  361  GTGTACGGAGAGCACTGGAGGAAGATGCGGCGGATCATGACGGTGCCGTTTTTCACGAAC  420
            ||  ||  ||  ||  ||  |||||  ||  |||  ||  |  ||  |||||  ||  ||  |||||  ||  ||
Sbjct  399  GTATATGGTGAACATTGGAGAAAAATGAGGAGAATTATGACTGTACCATTTTTTACTAAT  458

Query  421  AAGGTGGTTCAGCAATA  437
            ||  ||  ||  ||  ||  ||
Sbjct  459  AAAGTTGTGCAACAGTA  475
```

Fig. 6.12 BLAST result of partial *Leucaena C4H* sequence used to down-regulate C4H expression in tobacco. BLAST search was limited to tobacco (txid4097) and identity with only DQ350352 has been shown

6.3.9 Transcript levels of different phenylpropanoid pathway genes in the C4H down-regulated transgenic tobacco plants

Effect of down-regulation of *C4H* on the transcript levels of various phenylpropanoid pathway genes' was elucidated using the powerful technique of real time PCR. The Ct values and transcript level of various genes in the transgenics, compared to control plants are summarised in Tables 6.3 and 6.4, respectively. AS1-AS10 are transgenic (antisense C4H) tobacco plants and C stands for control tobacco plants.

Table 6.3 Ct values of various phenylpropanoid pathway genes obtained after normalization against 5.8S rRNA gene in the petiole samples analyzed between nodes 8-10 from first fully opened leaf at top

Plant name	*C4H*	*PAL*	*4CL*	*CAD*	*CCoAOMT*
AS1	30.81±0.09	31.31±0.51	30.33±0.14	28.26±0.05	31.42±0.24
AS4	28.97±0.15	29.06±0.43	28.64±0.14	26.99±0.02	29.43±0.29
AS5	28.35±0.29	27.5±0.07	27.56±0.03	25.80±0.05	29.29±0.41
AS6	29.02±0.42	27.63±0.16	27.81±0.12	26.45±0.05	29.61±0.22
AS8	28.64±0.01	28.18±0.1	28.36±0.04	26.85±0.17	28.23±0.15
AS9	26.47±0.24	26.73±0.11	26.77±0.08	25.55±0.06	27.52±0.15
AS10	29.02±0.05	28.67±0.13	28.73±0.07	27.07±0.05	29.39±0.11
C	26.26±0.16	23.49±0.07	26.36±0.1	25.1±0.37	27.19±0.35

Table 6.4 Transcript levels of various phenylpropanoid pathway genes in the transgenic petiole samples relative to control plants (Petiole samples analyzed between nodes 8-10 from first fully opened leaf at top). Transcript level in control plants was averaged to 100%

Plant name	*C4H* (%)	*PAL* (%)	*4CL* (%)	*CAD* (%)	*CCoAOMT* (%)
AS1	4.3	0.4	6.4	11.2	5.3
AS4	15.3	2.1	20.6	27	21.2
AS5	23.5	6.2	43.5	61.5	23.3
AS6	14.8	5.7	36.6	39.2	18.7
AS8	19.2	3.9	25.0	29.7	48.5
AS9	86.4	10.6	75.3	73.2	79.4
AS10	14.8	2.8	19.3	26.4	21.8
C	100	100	100	100	100

Careful study of Table 6.4 suggests that *C4H* down-regulation has profound effect on the transcript level of *C4H* in transgenics. The minimum transcript level was observed in AS1 with only 4.3% of the transcript level observed in control. Transcript level of all the genes of phenylpropanoid pathway studied in transgenics, showed reduced transcript level compared to control. It is interesting to see that *C4H* down-regulation in tobacco resulted in higher reduction in the transcript level of *PAL*, the enzyme preceding *C4H* in the pathway, than *C4H* itself. Another general phenylpropanoid pathway enzyme, *4CL* had comparable reduction in transcript level with that of *C4H*. *CAD* and *CCoAOMT* transcript levels had also reduced in transgenics. Graphs in the form of Fig. 6.13A, B, C, D and E represent the transcript level of *C4H*, *PAL*, *4CL*, *CAD* and *CCoAOMT*, respectively from different transgenics and averaged value of control plants.

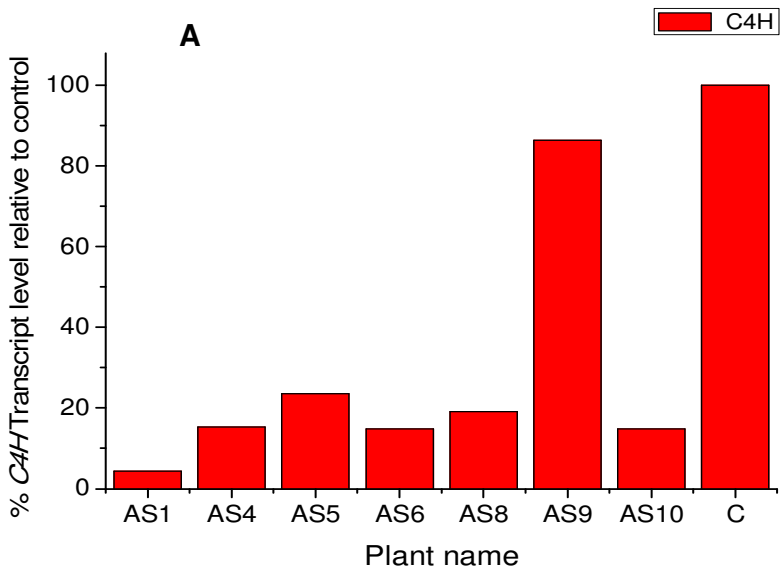


Fig. 6.13 (A) Transcript level of *C4H* from various transgenics and averaged value of control plant. The transcript level of *C4H* in control plants were arbitrarily assigned to be 100%

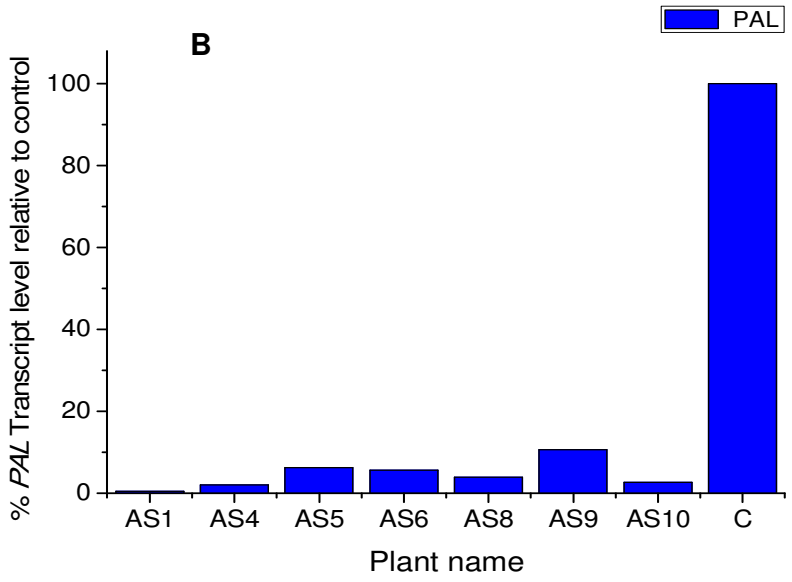


Fig. 6.13 (B) Transcript level of *PAL* from various transgenics and averaged value of control plants. The transcript level of *PAL* in control plants were arbitrarily assigned to be 100%

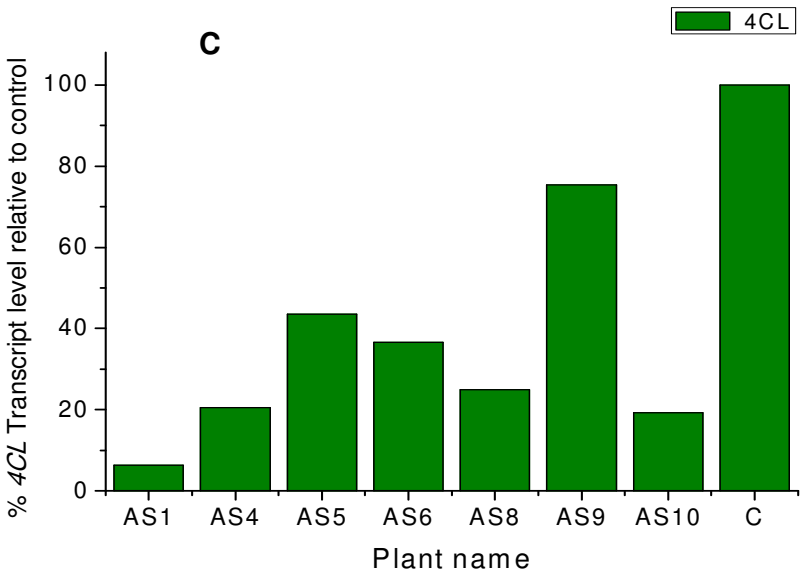


Fig. 6.13 (C) Transcript level of *4CL* from various transgenics and averaged value of control plants. The transcript level of *4CL* in control plants were arbitrarily assigned to be 100%

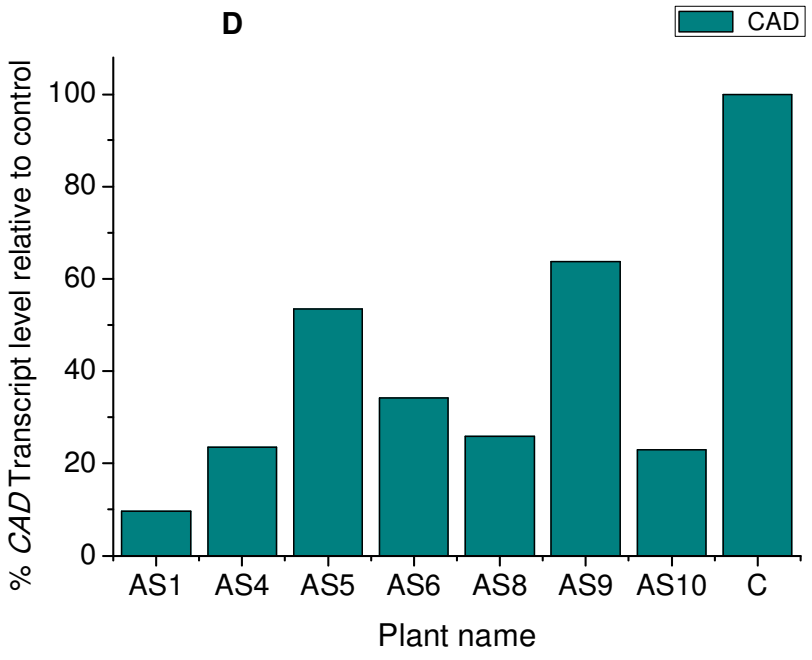


Fig. 6.13 (D) Transcript level of *CAD* from various transgenics and averaged value of control plants. The transcript level of *CAD* in control plants were arbitrarily assigned to be 100%

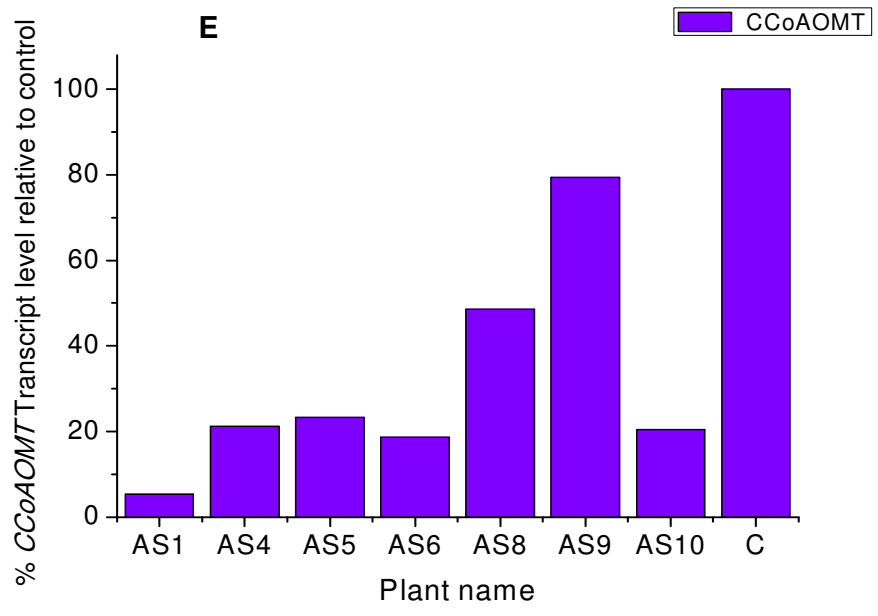


Fig. 6.13 (E) Transcript level of *CCoAOMT* from various transgenics and averaged value of control plants. The transcript level of *CCoAOMT* in control plants were arbitrarily assigned to be 100%

6.3.10 Lignin estimation from tobacco stem

Lignin from three control plants and 2 transgenic tobacco plants were estimated as mentioned in chapter 2, section 2.10. The lignin contents of the control plants were averaged. For the transgenics, lignin was estimated in duplicates and the values were averaged. Acid soluble and insoluble lignin content of the transgenics with respect to average lignin content of control plants has been presented in Table 6.5.

Table 6.5 Total lignin estimation from control and transgenic plants

↓ Plant name	Acid soluble lignin % (ASL)	Klason lignin% (KL)	Total lignin (ASL+KL)
C (averaged)	2.34 ± 0.03	20.3 ± 0.37	22.64 ± 0.4
AS9	1.98 ± 0.02	19.56 ± 0.24	21.54 ± 0.26
AS10	2.28 ± 0.04	13.62 ± 0.36	15.9 ± 0.4

6.4 Discussion

C4H downregulation in *Leucaena* and tobacco affected the normal growth of both plants. Both *Leucaena* and tobacco transgenics grew slowly initially. *Leucaena* transgenics are still very small for further analysis and hence most of the study is based on tobacco transgenics. A few tobacco transgenics displayed distortion in leaf morphology with the leaves being wrinkled and curled (Fig. 6.10B). This is consistent with the result of Prashant et al. (2011), where they used CCR gene from *Leucaena leucocephala* to down-regulate CCR activity in transgenic tobacco and observed the same phenotype in some of their transgenics. The phenotypic differences observed in transgenics or mutants appear to be directly related to the reduced and altered lignin in tissues, because lignin provides strength to the tissues (Bjurhager et al. 2010; Blount et al. 2000; Prashant et al. 2011; Ralph et al. 2006; Rastogi and Dwivedi 2006; Schilmiller et al. 2009; Sewalt et al. 1997). Differential phenotypic results in transgenics are observed depending upon the site of integration of t-DNA, which varies randomly in different transgenics and influences the transgene activity in both a qualitative and a quantitative manner (Van Der Krol et al. 1988). Mutants for C4H exhibited reduced level

of phenylpropanoid pathway end products, reduced lignin with altered monomer content with concomitant accumulation of a novel hydroxycinnamic ester, cinnamoymalate not found in wild type plants (Schilmiller et al. 2009). Other phenotypic changes observed were dwarfism, male sterility and swelling of branch junctions (Schilmiller et al. 2009). Transgenics down-regulated for C4H were raised in different plants and contrasting results were reported. Tobacco transgenics carrying an alfalfa (*Medicago sativa*) C4H transgene in the antisense orientation resulted in reduced level of Klason lignin and decreased S/G ratio as determined by pyrolysis gas chromatography/mass spectrometry (Sewalt et al. 1997), whereas *Populus tremula x tremuloides* (hybrid aspen) transgenics down-regulated for C4H using homologous partial C4H in antisense orientation resulted in reduced Klason lignin with a corresponding increase in cellulose content, but there was no change in S/G ratio determined by pyrolysis gas chromatography/mass spectrometry (Bjurhager et al. 2010). Introducing homologous sense construct of C4H in tomato resulted in delayed and reduced lignification, in some cases to an extent that is observed in plants heavily suppressed for C4H activity. This was further accompanied by an increase in S/G ratio (Millar et al. 2007) contrary to the observation in transgenic tobacco where reduced lignin level in down-regulated tobacco transgenics was associated with decreased S/G ratio (Sewalt et al. 1997). Sense or antisense expression of a truncated French bean class II *C4H* in tobacco resulted in delayed lignification and reduced lignin in most of the lines with slightly reduced S/G ratio compared to wild type (Blee et al. 2001). Over-expression of C4H had no effect on lignin content in tobacco containing C4H transgene from alfalfa (Sewalt et al. 1997).

Although down-regulation of C4H in plants has been studied and the resulting effect on lignin content and composition has been extensively studied in some cases, the relationship between C4H down-regulation and its effect on the transcript level or activity on other phenylpropanoid pathway gene(s) is largely underrepresented and is mostly limited to the study of PAL, the first enzyme of the pathway (Blount et al. 2000). In this study, I am reporting the quantitative changes in the transcript level of some

phenylpropanoid pathway genes as a result of C4H down-regulation in transgenic tobacco carrying partial *C4H* gene from *L. leucocephala* in antisense orientation. This study in conjunction with lignin estimation from stem tissues of tobacco transgenics will help us to understand whether or not C4H is the rate controlling enzyme at the entry point into the pathway.

Association of C4H and PAL is well documented. It has been successfully demonstrated that C4H was co-regulated with PAL in response to co-induction under various conditions in Parsley. In fact, the association of C4H was stronger with PAL than the redox partner of C4H, NADPH:cytochrome P450 oxidoreductase (CPR) (Koopmann et al. 1999). PAL and C4H activities concomitantly increase in potato tuber when they are illuminated with light (Lamb 1977). Transgenic tobacco down-regulated for C4H showed similar level of PAL activity as that of C4H (Blount et al. 2000). In the present study, *PAL* transcript level in transgenic tobacco decreased to a much lower fraction of the control plants. In AS1, C4H transcript level reached about 4.3% of its level in control plants, whereas PAL transcript level decreased to a much lower level, 0.4% of its level in control plants. In all the transgenics, there is correspondingly higher reduction in the transcript level of PAL than in the transcript level of C4H. The transcript level (in terms of number of mRNA transcripts) of PAL was 6.8 times higher than the transcript level of C4H in control plants under the green house condition, deduced on the basis of change in Ct values of the two genes. Down-regulation of C4H in transgenics shifted the equilibrium to between 0.7- 1.8, i.e. number of mRNA transcripts of PAL was in between 0.7 to 1.8 times the number of transcripts of C4H in transgenics. This may be one reason why Blount et al. (2000) found equal activity of C4H and PAL in C4H down-regulated transgenics, although, while stating this, I did not take into account post-transcriptional regulation, post-transcriptional modification and the relative activity of these two enzymes *in vivo*. Reduction of PAL activity in C4H down-regulated plants has previously been reported by Blount et al. (2000). Blount et al. (2000) also reported that C4H activity was not reduced in plants genetically engineered for reduced PAL activity. Also, progeny arising as a result of cross between a tobacco line over-expressing bean PAL transgene and a C4H antisense

transgenic line had reduced PAL activity than progeny of PAL over-expresser. This emphasizes the dominant effect of C4H down-regulation over PAL over-expression.

Bjurhager et al. (2010) reported that in two transgenic hybrid aspen trees down-regulated for homologous C4H transgene had C4H transcript level in the range of 23.5 ± 15% and 12.5 ± 4% respectively. In the present study, transcript level as low as 4% has been recorded from AS1 as compared to control. Given that P450 enzymes catalyze slow reactions (low K_{cat}) (Ehlting et al. 2006), a reduction of this level has the potential to abnormally bring down the overall activity of C4H from the transgenic plants leading to accumulation of its substrate, cinnamic acid. Inhibition of PAL transcript level and activity was recorded accompanied with the synthesis of a proteinaceous PAL inhibitor in potato tubers, when cinnamic acid was exogenously supplied (Lamb 1977). Blount et al. (2000) also established that cinnamic acid plays the role of a negative modulator of PAL activity, whose accumulation beyond a certain level in the cell is perhaps sensed by the cell machinery to withhold PAL transcription. Studies have also demonstrated that potato tubers when illuminated with light, lead to concomitant increase in PAL and C4H activity, but while PAL activity go on increasing, C4H activity does not increase beyond a level even after extended illumination (Lamb 1977) emphasizing that it is perhaps C4H that checks and monitors carbon flux into phenylpropanoid pathway. Physical association in the form of a multi-enzyme complex formed of a particular isoform of PAL and C4H in transgenic tobacco is also reported (Achnine et al. 2004). Formation of multi-enzyme complex does not let PAL generated cinnamic acid to diffuse to the cytosol and thus forms a sort of channelling between PAL and C4H. Because PAL and C4H are so closely associated, there is no surprise that down-regulation of C4H reduced PAL transcript to a greater level than the reduction of C4H transcripts in the present study.

Besides PAL, C4H downregulation also lead to the decrease in the transcript level of other phenylpropanoid pathway genes used in the present study. Since as per my best knowledge, there are no previous reports of

reduction of transcript levels of these enzymes, especially *CAD* and *CCoAOMT* in C4H transgenics, this study holds its own importance. Q-RT PCR analysis suggested that *C4H* and *4CL* transcripts are almost equally abundant in control plants. In all transgenics, *4CL* transcript level is slightly above *C4H* transcript level except in AS9, where, *4CL* level (75.3% of control) is below *C4H* level (86.4% of control). This data shows that even a slight downregulation of *C4H* in transgenics can reduce *4CL* transcript level. In the same plant (AS9), *PAL* transcript level is only 10.6% of the level present in control plants. In control plants, *CAD* transcript level was about 2.5 times higher than *C4H* transcript level, whereas *CCoAOMT* transcript level was 0.53 times (i.e. transcript level lesser than *C4H*) of *C4H* transcript level. Both, *CAD* and *CCoAOMT* transcript level decreased as a result of reducing C4H activity in transgenic tobacco. In AS9, a small reduction in *C4H* level reduced *CAD* and *CCoAOMT* level to 73.2% and 79.4% of their control level, respectively. In general, reduction in *CCoAOMT* was greater than the reduction in *CAD* level except in AS8. Since the transcript levels of these genes have decreased in *C4H* transgenics, it may be proposed that C4H catalyzes a rate limiting step at least in part, at the early point in the phenylpropanoid pathway. Anterola et al. (2002) showed that *Pinus taeda* cell suspension culture were when fed with saturating level of phenylalanine and kept in phenylpropanoid induction medium, induced the transcription level of *PAL*, *4CL*, *COMT*, *CCR* and *CAD*, but did not induce transcription level of *C4H* and *C3H*. This shows that transcript levels of *C4H* and *C3H* are not dependent upon the metabolic requirement of the cells and may represent controlling point in the phenylpropanoid pathway. The present study further supports the hypothesis of Anterola et al. (2002). As expected, lignin content of the transgenics analyzed were reduced compared to the control plants. In AS9, displaying only partial decrease in *C4H* transcript level (86.4% of control) had 4.8% lesser total lignin content primarily due to major decrease in the acid soluble lignin and slight decrease in Klason lignin. Contrary to it, in AS10, which had only 14.8% *C4H* transcript level compared to control, had 30% lesser total lignin, mainly due to decrease in Klason lignin than acid soluble lignin. It suggests that more transgenics need to be analyzed for their lignin content to reach to a conclusion.

6.5 Conclusion

Transgenic *Leucaena* and tobacco plants down-regulated for C4H were raised using a modified pCAMBIA1301 vector. Q-RT PCR study of transgenic tobacco showed reduced transcript level of *PAL, 4CL, CAD* and *CCoAOMT* in addition to *C4H* in all the transgenics. Lignin level was also found decreased in stem tissues of the transgenics analyzed. These findings support the hypothesis of Anterola et al. (2002) that *C4H* might be a rate limiting gene early in the phenylpropanoid pathway.

SUMMARY

&

FUTURE PROSPECTS

Summary and future prospects

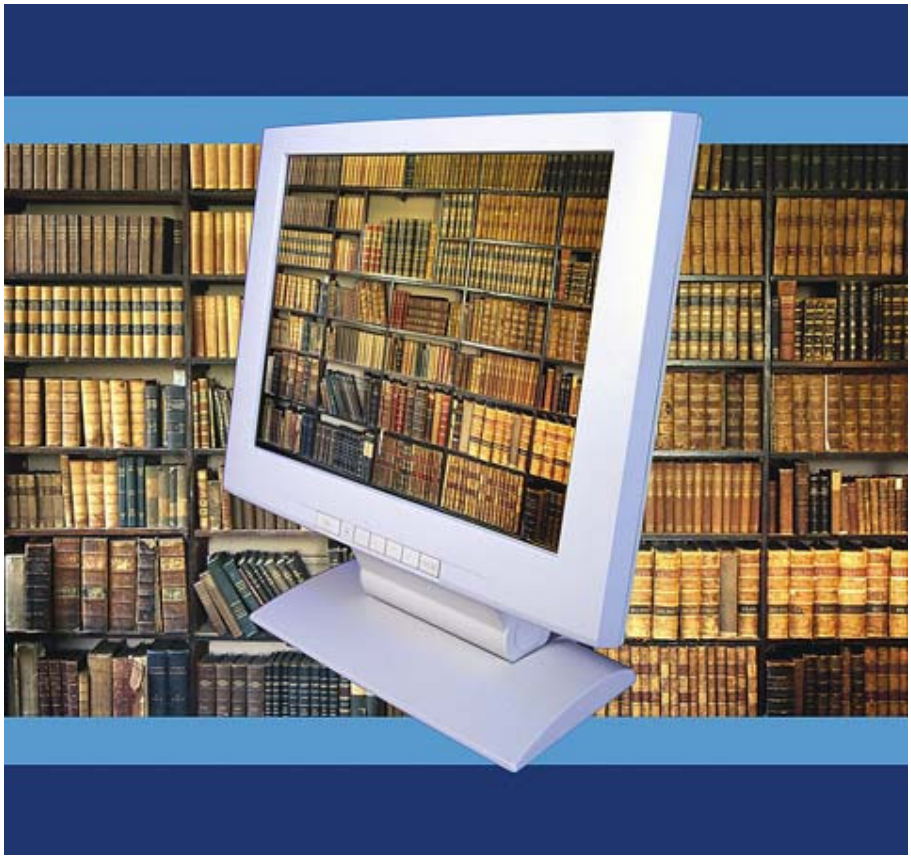
To meet the growing demand of paper, we need to have high cellulosic pulp yielding trees well adapted to grow in a range of climatic conditions. *Leucaena leucocephala*, hailed once as the 'miracle tree' because of its worldwide reputation of being a highly nutritious perennial forage tree has largely emerged as a multi-utility tree of late. *L. leucocephala* accounts for approximately 25% of the raw material supplied to the paper and pulp industry in India. Quality wise, pulp generated from *L. leucocephala* is rich in holocellulose and low in ash, silica, hot water solubles and alcohol-benzene solubles, which is at par with other fast growing hardwoods used in paper industry. The one factor that is of prime concern to the paper and pulp industry is the lignin content of *L. leucocephala*. The presence of lignin adversely affects paper quality and is responsible for poor performance characteristics and brightness of paper associated with yellowing of paper with age. To produce high quality paper, lignin needs to be removed from the pulp involving various hazardous chemicals. Every unit decrease in lignin content associated with favourable S/G ratio will be highly welcomed by paper industry. In this context, the present work was undertaken to isolate and characterize *Cinnamate 4-Hydroxylase*, a lignin biosynthesis pathway gene situated at strategically important point in the pathway. It was also planned to raise transgenic *L. leucocephala* plants and study the extent of downregulation of C4H and the resulting effect on lignin content.

Three highly identical *C4H* isoforms were isolated from *L. leucocephala* which bear more than 82% sequence identity to C4H enzymes from other plants. These isoforms contained all the conserved sequences found in P450 superfamily and C4H from other plants. All the three isoforms of *C4H* were expressed in *E. coli*, strain Rosetta (DE3) and one isoform was purified from inclusion body. Spatio-temporal transcript abundance study and activity assay form crude tissue extract was also done to understand the expression pattern of C4H in vivo. Maximum *C4H* transcript level was found in root tissue followed by stem and leaves. Lignin quantification from tissues of different age suggests that lignification is directly proportional to C4H activity in *L. leucocephala*. Thus, it provided direct evidence that C4H down-regulation may lead to substantial reduction in lignin content in transgenic

plants. Transgenic *Leucaena* and tobacco plants down-regulated for C4H were raised using a modified pCAMBIA1301 vector. Transgenic *Leucaena* plants are slow growing and could not be analyzed owing to their poor growth and small size. Q-RT PCR study of transgenic tobacco showed reduced transcript level of *PAL, 4CL, CAD* and *CCoAOMT* in addition to *C4H* in all the transgenics. Lignin level was also found decreased in stem tissues of the transgenics analyzed. This suggested that transgenic *Leucaena* plants down-regulated for C4H may also have reduced lignin content near critical level, due to which they are so much slow in growth.

Transgenic tobacco plants may be analyzed for lignin composition in future. The transgenic *Leucaena* plants that recover in near future and start gaining height and weight may be analyzed for lignin content and the activity of C4H in those plants. In future, it may be planned to reduce the expression of two lignin biosynthetic pathway genes simultaneously, one being C4H which comes early in the pathway and the other one CCR or CAD which comes late in the pathway. The plants can then be characterized for lignin related traits. This may be done in tobacco first, because tobacco is fast growing compared to *Leucaena*. Although, it may be noted that the results in the two plant systems may not be the same.

REFERENCES



Achnine L, Blancaflor EB, Rasmussen S, Dixon RA (2004) Colocalization of L -phenylalanine ammonia-lyase and cinnamate 4-hydroxylase for metabolic channeling in phenylpropanoid biosynthesis. Plant Cell 16:3098-3109. doi: 10.1105/tpc.104.024406.pathways

Achyuthan KE, Achyuthan AM, Adams PD, et al. (2010) Supramolecular self-assembled chaos: polyphenolic lignin's barrier to cost-effective lignocellulosic biofuels. Molecules 15:8641-8688. doi: 10.3390/molecules15118641

Akashi T, Aoki T, Takahashi T, et al. (1997) Cloning of cytochrome P450 cDNAs from cultured *Glycyrrhiza echinata* L. cells and their transcriptional activation by elicitor-treatment. Plant Sci 126:39-47. doi: 10.1016/S0168-9452(97)00091-5

Anterola AM, Jeon J-H, Davin LB, Lewis NG (2002) Transcriptional control of monolignol biosynthesis in *Pinus taeda*: factors affecting monolignol ratios and carbon allocation in phenylpropanoid metabolism. J Biol Chem 277:18272-18280.

Arnheim N, Erlich H (1992) Polymerase chain reaction strategy. Ann Rev Biochem 61:131-156.

Batard Y, Schalk M, Pierrel MA, et al. (1997) Regulation of the cinnamate 4-hydroxylase (CYP73A1) in *Jerusalem artichoke* tubers in response to wounding and chemical treatments. Plant Physiol 113:951-959.

Belles J, Lopezgresa M, Fayos J, et al. (2008) Induction of cinnamate 4-hydroxylase and phenylpropanoids in virus-infected cucumber and melon plants. Plant Sci 174:524-533. doi: 10.1016/j.plantsci.2008.02.008

Bell-lelong DA, Cusumano JC, Meyer K, Chapple C (1997) Cinnamate-4-hydroxylase expression in *Arabidopsis* regulation in response to development and the environment. Plant Physiol 113:729-738.

Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340:783-95. doi: 10.1016/j.jmb.2004.05.028

Betz C, Mccollum TG, Mayer RT (2001) Differential expression of two cinnamate 4-hydroxylase genes in "Valencia" orange (*Citrus sinensis* Osbeck). Plant Mol Biol 46:741-748.

Bjurhager I, Olsson A-M, Zhang B, et al. (2010) Ultrastructure and mechanical properties of populus wood with reduced lignin content caused by transgenic down-regulation of cinnamate 4-hydroxylase. Biomacromolecules 11:2359-2365.

Blee K, Choi JW, O'Connell a P, et al. (2001) Antisense and sense expression of cDNA coding for CYP73A15, a class II cinnamate 4-

hydroxylase, leads to a delayed and reduced production of lignin in tobacco. Phytochemistry 57:1159-66.

Blount JW, Korth KL, Masoud S a, et al. (2000) Altering expression of cinnamic acid 4-hydroxylase in transgenic plants provides evidence for a feedback loop at the entry point into the phenylpropanoid pathway. Plant Physiol 122:107-16.

Boerjan W, Ralph J, Baucher M (2003) Lignin biosynthesis. Annu Rev Plant Biol 54:519-46. doi: 10.1146/annurev.arplant.54.031902.134938

Bolwell GP, Cramer CL, Lamb CJ, et al. (1986) L-Phenylalanine ammonia-lyase from *Phaseolus vulgaris*: Modulation of the levels of active enzyme by trans-cinnamic acid. Planta 169:97-107.

Bolwell GP, Mavandad M, Millar DJ, et al. (1988) Inhibition of mRNA levels and activities by trans-cinnamic acid in elicitor-induced bean cells. Phytochemistry 27:2109-2117. doi: 10.1016/0031-9422(88)80106-7

Boudet A-M (2007) Evolution and current status of research in phenolic compounds. Phytochemistry 68:2722-2735. doi: 10.1016/j.phytochem.2007.06.012

Boudet A-M (2000) Lignins and lignification: Selected issues. Plant Physiol Biochem 38:81-96.

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 72:248-254.

Chapple C (1998) Molecular-genetic analysis of plant cytochrome P450-dependent monooxygenases. Annu Rev Plant Physiol Plant Mol Biol 49:311-343. doi: 10.1146/annurev.arplant.49.1.311

Chen A-H, Chai Y-R, Li J-N, Chen L (2007) Molecular cloning of two genes encoding cinnamate 4-hydroxylase (C4H) from oilseed rape (*Brassica napus*). J Biochem Mol Biol 40:247-260.

Chen BJ, Wang Y, Hu YL, et al. (2005) Cloning and characterization of a drought-inducible MYB gene from *Boea crassifolia*. Plant Sci 168:493-500. doi: 10.1016/j.plantsci.2004.09.013

Chen JY, Wen PF, Kong WF, et al. (2006) Changes and subcellular localizations of the enzymes involved in phenylpropanoid metabolism during grape berry development. J Plant Physiol 163:115-27. doi: 10.1016/j.jplph.2005.07.006

Durst F, Nelson DR (1995) Diversity and evolution of plant P450 and P450-reductases. Drug Metab Drug Interact 12:189-206.

Dutt D, Tyagi CH, Malik RS (2007) Studies on effect of growth factor on morphological, chemical and pulp and paper making characteristics and its impact on fluff generation. Indian J Chem Techn 14:626-634.

Eckardt N a (2002) Probing the mysteries of lignin biosynthesis: the crystal structure of caffeic acid/5-hydroxyferulic acid 3/5-O-methyltransferase provides new insights. Plant cell 14:1185-1189.

Ehlting J, Hamberger B, Million-Rousseau R, Werck-Reichhart D (2006) Cytochromes P450 in phenolic metabolism. Phytochem Rev 5:239-270. doi: 10.1007/s11101-006-9025-1

FAOSTAT (2011) Highlights on paper and paperboard: 1999- 2009. In: FAO Forestry Department. http://faostat.fao.org/Portals/_Faostat/documents-/pdf/Paper and paperboard.pdf. Accessed 1 Oct 2011

Fahrendorf T, Dixon RA (1993) Stress responses in alfalfa (*Medicago sativa* L.). XVIII: Molecular cloning and expression of the elicitor-inducible cinnamic acid 4-hydroxylase cytochrome P450. Arch Biochem Biophys 305:509-515.

Frank MR, Deyneka JM, Schuler MA (1996) Cloning of wound-induced cytochrome P450 monooxygenases expressed in pea. Plant Physiol 110:1035-1046.

Gabriac B, Werck-Reichhart D, Teutsch H, Durst F (1991) Purification and immunocharacterization of a plant cytochrome P450: the cinnamic acid 4-hydroxylase. Arch Biochem Biophys 288:302-309.

Goujon T, Sibout R, Eudes A, et al. (2003) Genes involved in the biosynthesis of lignin precursors in *Arabidopsis thaliana*. Plant Physiol Biochem 41:677-687. doi: 10.1016/S0981-9428(03)00095-0

Gravot A, Larbat R, Hehn A, et al. (2004) Cinnamic acid 4-hydroxylase mechanism-based inactivation by psoralen derivatives: cloning and characterization of a C4H from a psoralen producing plant—*Ruta graveolens*—exhibiting low sensitivity to psoralen inactivation. Arch Biochem Biophys 422:71-80. doi: 10.1016/j.abb.2003.12.013

Grima-pettenati J, Goffner D (1999) Lignin genetic engineering revisited. Plant Sci 145:51 - 65.

Gupta SK (2008) Isolation, cloning and characterization of lignin biosynthesis pathway gene(s) 4-Coumarate Co A ligase (4CL) from *Leucaena leucocephala*. University of Pune

Gustafsson C, Govindarajan S, Minshull J (2004) Codon bias and heterologous protein expression. Trends Biotechnol 22:346-53. doi: 10.1016/j.tibtech.2004.04.006

Hahlbrock K, Scheel D (1989) Physiology and molecular biology of phenylpropanoid metabolism. Annu Rev Plant Physiol Plant Mol Biol 40:347-369.

Halpin C, Knight ME, Foxon GA, et al. (1994) Manipulation of lignin quality by downregulation of cinnamyl alcohol dehydrogenase. Plant J 6:339-350.

Hershberg R, Petrov D a (2008) Selection on codon bias. Annu Rev Genet 42:287-99. doi: 10.1146/annurev.genet.42.110807.091442

Higuchi R, Dollinger G, Walsh PS, Griffith R (1992) Simultaneous amplification and detection of specific DNA sequences. Nat Biotechnol 10:413-417. doi: 10.1038/nbt0492-413

Higuchi R, Fockler C, Dollinger G, Watson R (1993) Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. Nat Biotechnol 11:1026-1030.

Horsch RB, Fry JE, Hoffmann NL, et al. (1985) A Simple and General Method for transferring genes into plants. Science 227:1229-1231.

Hotze M, Schröder G, Schröder J (1995) Cinnamate 4-hydroxylase from *Catharanthus roseus*, and a strategy for the functional expression of plant cytochrome P450 proteins as translational fusions with P450 reductase in *Escherichia coli*. FEBS Letters 374:345-350.

Huang B, Duan Y, Yi B, et al. (2008) Characterization and expression profiling of cinnamate 4-hydroxylase gene from *Salvia miltiorrhiza* in rosmarinic acid biosynthesis pathway. Russ J Plant Physl+ 55:390-399. doi: 10.1134/S1021443708030163

Hübner S, Hehmann M, Schreiner S, et al. (2003) Functional expression of cinnamate 4-hydroxylase from *Ammi majus* L. Phytochemistry 64:445-452. doi: 10.1016/S0031-9422(03)00265-6

Jiang H, Wood KV, Morgan JA (2005) Metabolic engineering of the phenylpropanoid pathway in *Saccharomyces cerevisiae*. Appl Environ Microb 71:2962-2969. doi: 10.1128/AEM.71.6.2962

Kawai S, Mori A, Shiokawa T, et al. (1996) Isolation and analysis of cinnamic acid 4-hydroxylase homologous genes from a hybrid aspen, *Poplus kitakamiensis*. Biosci Biotech Biochem, 60:1586-1597.

Kirk TK, Obst JR (1988) Lignin determination. Methods Enzymol 161:87-101.

Koopmann E, Logemann E, Hahlbrock K (1999) Regulation and functional expression of cinnamate 4-hydroxylase from Parsley. Plant Physiol 119:49-55.

Kulkarni HD (2008) Private farmerprivate industry partnerships for industrial wood production: a case study. Int Forest Rev 10:147-155.

Lamb CJ (1977) Phenylalanine Ammonia-Lyase and Cinnamic Acid 4-Hydroxylase: Characterization of the concomitant changes in enzyme activities in illuminated potato tuber discs. Planta 175:169-175.

Lamb CJ, Rubery PH (1975) A spectrophotometric assay for *trans*-cinnamic acid 4-hydroxylase activity. Anal Biochem 68:554-561.

Lee SS, Lee EM, An BC, et al. (2008) Molecular cloning and characterization of cinnamate-4-hydroxylase gene from *Rubus coreanus*. Open Plant Sci J 2:31-36.

Liu S, Hu Y, Wang X, et al. (2009) Isolation and characterization of a gene encoding cinnamate 4-hydroxylase from *Parthenocissus henryana*. Mol Biol Rep 36:1605-1610. doi: 10.1007/s11033-008-9357-6

Lu S, Zhou Y, Li L, Chiang VL (2006) Distinct roles of cinnamate 4-hydroxylase genes in *Populus*. Plant Cell Physiol 47:905-914. doi: 10.1093/pcp/pcj063

Mavandad M, Edwards R, Liang X, et al. (1990) Effects of trans-cinnamic acid on expression of the bean phenylalanine ammonia-lyase gene family. Plant Physiol 94:671-680.

Millar DJ, Long M, Donovan G, et al. (2007) Introduction of sense constructs of cinnamate 4-hydroxylase (CYP73A24) in transgenic tomato plants shows opposite effects on flux into stem lignin and fruit flavonoids. Phytochemistry 68:1497-509. doi: 10.1016/j.phytochem.2007.03.018

Mizutani M, Ohta D, Sato R (1997) Isolation of a cDNA and a genomic clone encoding cinnamate 4-hydroxylase from *Arabidopsis* and Its expression manner in planta. Plant Physiol 113:755-763.

Mizutani M, Otha D, Sato R (1993a) Purification and characterization of a cytochrome P450 (trans-cinnamic acid 4-hydroxylase) from etiolated mung bean seedlings. Plant Cell Physiol 34:481-488.

Mizutani M, Ward E, DiMaio J, et al. (1993b) Molecular cloning and sequencing of a cDNA encoding mung bean cytochrome P450 (P450C4H) possessing cinnamate 4-hydroxylase activity. Biochem Biophys Res Commun 190:875-880.

Mullis KB, Faloona FA (1987) Specific synthesis of DNA in vitro via a polymerase-catalysed chain reaction. Methods Enzymol 155:335-350.

Murashige T, Skoog F (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiol Plantarum 15:473-497.

Naoumkina MA, Zhao Q, Gallego-giraldo L, et al. (2010) Genome-wide analysis of phenylpropanoid defence pathways. Mol Plant Pathol 11:829-846. doi: 10.1111/J.1364-3703.2010.00648.X

Nielsen JS, Møller BL (2000) Cloning and expression of cytochrome P450 enzymes catalyzing the conversion of tyrosine to p-hydroxyphenylacetaldoxime in the biosynthesis of cyanogenic glucosides in *Triglochin maritima*. Plant Physiol 122:1311-21.

Ni W, Paiva NL, Dixon RA (1994) Reduced lignin in transgenic plants containing an engineered caffeic acid *O*-methyltransferase antisense gene. Transgenic Res 3:120-126.

Omer S (2011) Studies on regulation of lignin biosynthesis gene(s) in *Leucaena leucocephala*. University of Pune, pp-200-202.

Overkamp S, Hein F, Barz W (2000) Cloning and characterization of eight cytochrome P450 cDNAs from chickpea (*Cicer arietinum* L.) cell suspension cultures. Plant Sci 155:101-108.

Park JH, Park NI, Xu H, Park SU (2010) Cloning and characterization of phenylalanine ammonia-lyase and cinnamate 4-hydroxylase and pyranocoumarin biosynthesis in *Angelica gigas*. J Nat Prod 73:1394-1397.

Petersen M (2003) Cinnamic acid 4-hydroxylase from cell cultures of the hornwort *Anthoceros agrestis*. Planta 217:96-101. doi: 10.1007/s00425-002-0960-9

Prashant S, Srilakshmi Sunita M, Pramod S, et al. (2011) Down-regulation of *Leucaena leucocephala* cinnamoyl CoA reductase (LlCCR) gene induces significant changes in phenotype, soluble phenolic pools and lignin in transgenic tobacco. Plant Cell Rep. doi: 10.1007/s00299-011-1127-6

Ralph J, Akiyama T, Kim H, et al. (2006) Effects of coumarate 3-hydroxylase down-regulation on lignin structure. J Biol Chem 281:8843-53. doi: 10.1074/jbc.M511598200

Rasmussen S, Dixon RA, Division PB, et al. (1999) Transgene-mediated and elicitor-induced perturbation of metabolic channeling at the entry point into the phenylpropanoid pathway. Plant Cell 11:1537-1551.

Rastogi S, Dwivedi UN (2006) Down-regulation of lignin biosynthesis in transgenic *Leucaena leucocephala* harboring O-methyltransferase gene. Biotechnol Prog 22:609-16. doi: 10.1021/bp050206+

Ro D-K, Douglas CJ (2004) Reconstitution of the entry point of plant phenylpropanoid metabolism in yeast (*Saccharomyces cerevisiae*): implications for control of metabolic flux into the phenylpropanoid pathway. J Biol Chem 279:2600-2607. doi: 10.1074/jbc.M309951200

Ro DK, Mah N, Ellis BE, Douglas CJ (2001) Functional characterization and subcellular localization of poplar (*Populus trichocarpa* x *Populus deltoides*) cinnamate 4-hydroxylase. Plant Physiol 126:317-329.

Ronald W (1993) Purification, characterization, and cDNA cloning of an NADPH-cytochrome P450 reductase from mung bean. Proc Natl Acad Sci USA 90:2890-2894.

Rupasinghe S, Baudry J, Schuler M a. (2003) Common active site architecture and binding strategy of four phenylpropanoid P450s from *Arabidopsis thaliana* as revealed by molecular modeling. Protein Eng 16:721-731. doi: 10.1093/protein/gzg094

Russell DW (1971) The metabolism of aromatic compounds in higher plants X. Properties of the cinnamic acid 4-hydroxylase of pea seedlings and some aspects of its metabolic and developmental control. J Biol Chem 246:3870-3878.

Russell DW, Conn EE (1967) The cinnamic acid 4-hydroxylase of pea seedlings. Arch Biochem Biophys 122:256-258.

Saathoff AJ, Sarath G, Chow EK, et al. (2011) Downregulation of cinnamyl-alcohol dehydrogenase in Switchgrass by RNA silencing results in enhanced glucose release after cellulase treatment. PLoS ONE 6:10.

Sambrook J, Russell D (2000) Molecular cloning: A laboratory manual, 3rd ed. Cold Spring Harbor Laboratory Press, NewYork

Sanford JC, Smith FD, Russell JA (1993) Optimizing the biolistic process for different biological applications. Methods Enzymol 217:483-509.

Schilmiller AL, Stout J, Weng J-K, et al. (2009) Mutations in the cinnamate 4-hydroxylase gene impact metabolism, growth and development in *Arabidopsis*. Plant J 60:771-782.

Schoch G a., Attias R, Le Ret M, Werck-Reichhart D (2003) Key substrate recognition residues in the active site of a plant cytochrome P450, CYP73A1. Homology model guided site-directed mutagenesis. Eur J Biochem 270:3684-3695. doi: 10.1046/j.1432-1033.2003.03739.x

Sewalt VJH, Ni W, Blount JW, et al. (1997) Reduced lignin content and altered lignin composition in transgenic tobacco down-regulated in expression of L-phenylalanine ammonia-lyase or cinnamate 4-hydroxylase. Plant Physiol 115:41-50.

Shaik NM, Arha M, Nookaraju A, et al. (2009) Improved method of in vitro regeneration in *Leucaena leucocephala* — a leguminous pulpwood tree species. Physiol Mol Biol Plants 15:311-318. doi: 10.1007/s12298-009-0035-5

Shelton HM, Brewbaker JL (1994) *Leucaena leucocephala* - the most widely used forage tree legume. http://www.fao.org/ag/AGP/AGPC/doc-/Publicat/Gutt-shel/x5556e06.htm. Accessed 1 Oct 2011

Shi R, Sun Y-H, Li Q, et al. (2010) Towards a systems approach for lignin biosynthesis in *Populus trichocarpa*: transcript abundance and specificity of the monolignol biosynthetic genes. Plant Cell Physiol 51:144-163. doi: 10.1093/pcp/pcp175

Sibout R, Eudes A, Mouille G, et al. (2005) Cinnamyl alcohol dehydrogenase-C and -D are the primary genes involved in lignin biosynthesis in the floral stem of *Arabidopsis*. Plant Cell 17:2059-2076. doi: 10.1105/tpc.105.030767.1

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503-517.

Srivastava S, Gupta RK, Arha M, et al. (2011) Expression analysis of cinnamoyl-CoA reductase (CCR) gene in developing seedlings of *Leucaena leucocephala*: a pulp yielding tree species. Plant Physiol Biochem 49:138-45. doi: 10.1016/j.plaphy.2010.11.001

Steudle E (2001) The cohesion-tension mechanism and the acquisition of water by plant roots. Annu Rev Plant Physiol Plant Mol Biol 52:847-875. doi: 10.1146/annurev.arplant.52.1.847

Tamura K, Peterson D, Peterson N, et al. (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731-2739. doi: 10.1093/molbev/msr121

Teutsch HG, Hasenfratz MP, Lesot A, et al. (1993) Isolation and sequence of a cDNA encoding the *Jerusalem artichoke* cinnamate 4-hydroxylase, a major plant cytochrome P450 involved in the general phenylpropanoid pathway. Proc Natl Acad Sci USA 90:4102-4106.

Van Der Krol A, Lenting PE, Veenstra J, et al. (1988) An anti-sense chalcone synthase gene in transgenic plants inhibits flower pigmentation. Nature 333:866-869. doi: 10.1038/333866a0

Vanholme R, Demedts B, Morreel K, et al. (2010) Lignin biosynthesis and structure. Plant Physiol 153:895-905. doi: 10.1104/pp.110.155119

Vanholme R, Morreel K, Ralph J, Boerjan W (2008) Lignin engineering. Curr Opin Plant Biol 11:278-285. doi: 10.1016/j.pbi.2008.03.005

Vietmeyer N, Cottom B (1977) *Leucaena*: Promising forage and tree crop for the tropics, Nat Acad Sci USA 22-80.

Vogt T (2010) Phenylpropanoid biosynthesis. Mol Plant 3:2-20.

Whitbred JM, Schuler MA (2000) Molecular characterization of CYP73A9 and CYP82A1 P450 genes involved in plant defense in pea. Plant Physiol 124:47-58.

Yamamura Y, Ogihara Y, Mizukami H (2001) Cinnamic acid 4-hydroxylase from *Lithospermum erythrorhizon*: cDNA cloning and gene expression. Plant Cell Rep 20:655-662. doi: 10.1007/s002990100373

Yamazaki S, Sato K, Sakaguchi M, et al. (1993) Importance of the proline-rich region following signal-anchor sequence in the formation of correct conformation of microsomal cytochrome P-450s. J Biochem 114:652-657.

Book chapters / Research Papers published / under review / under preparation

- Shabab Mohammad, Vaibhav P. Mhaindarkar, **Santosh Kumar**, Mohammad I. Khan, Siddharth H. Bhosale (2011) Isolation and phylogenetic analysis of marine fungus *Penicillium sp.* sdbf1 and partial characterization of its cysteine protease inhibitor. *International Journal of Advanced Biotechnology and Research* 2: 135-142

- Bashir M. Khan, Shuban K. Rawal, Manish Arha, Sushim K. Gupta, Sameer Srivastava, Noor M. Shaik, Arun K. Yadav, Pallavi S. Kulkarni, Abhilash O. U., **Santosh Kumar**, Sumita Omer *et al*. (2011) Genetic Engineering of Phenylpropanoid Pathway in *Leucaena leucocephala* (Book chapter accepted for publication *In* genetic Engineering, Intech – Open Access Publishers (http://www.intechweb.org/booksprocess/-allchapters/)

- **Santosh Kumar**, Sumita Omer and B. M. Khan. Molecular studies on Cinnamate 4-hydroxylase from *Leucaena leucocephala*: a pulp yielding legume tree (under review in *Plant Cell Reports*)

- **Santosh Kumar**, Sumita Omer, Shruti Chitransh and B. M. Khan. (2011) Downregulation of *Cinnamate 4-Hydroxylase* in transgenic tobacco alters the transcript abundance of other phenylpropanoid pathway gene(s) (manuscript under preparation for *Plant Cell, Tissue and Organ Culture*)

- Sumita Omer, **Santosh Kumar** and B. M. Khan. (2011) Isolation and characterization of a novel R2R3 type MYB transcription factor gene from *Leucaena leucocephala* (manuscript under preparation)

Best Poster Awards

- **Santosh Kumar**, Sumita Omer and B.M. Khan. Molecular Studies on Cinnamate 4-Hydroxylase (*C4H*) and its Down-regulation studies in *Leucaena leucocephala* and tobacco. Best poster prize in the area of Biological Sciences in the poster presentation held on 24-25th February, 2011 as part of the National Science Day and International Year of Chemistry celebrations, 2011 at National Chemical Laboratory, Pune, India

- **Santosh Kumar**, Sumita Omer and B.M. Khan. Isolation, Cloning and Expression of three Isoforms of Cinnamate 4-Hydroxylase (C4H) and their Down-regulation studies in *Leucaena leucocephala*. Best poster (second prize) award in National Symposium and XXXII Annual Meet of Plant Tissue Culture Association (India) held from 4th to 6th February 2011 at Bikaner, India

***Santosh Kumar**, the author of this manuscript was born on 8th March, 1981 at Dalsingsarai, Distt-Samastipur in Bihar. He did his schooling (Secondary & Higher Secondary) from Jawahar Navodaya Vidyalaya, Begusarai, Bihar. He joined Assam Agricultural University, Jorhat for B.Sc. Ag. Degree programme in 2004 and awarded Bachelor of Science in Agriculture in 2004. After getting selected in All India JNU combined Entrance Test he joined G.B.P.U.A&T, Pantnagar, Uttanchal in July 2004. He received his post graduation (M.Sc. Ag. in Molecular Biology and Biotechnology) at G.B.PUA&T Pantnagar in the department of Molecular Biology & Genetic Engineering in 2006. He then joined National Chemical Laboratory, Pune, India for his PhD in Biotechnology in the month of September, 2006.*

He was awarded University merit scholarship in all the semesters during his under graduation, and Department of Biotechnology; Government of India merit scholarship during post graduation programme. He qualified Council of Scientific and Industrial Research-Junior Research Fellowship (CSIR-JRF) award in the month of December, 2005 and subsequently joined NCL as a research scholar.

Permanent address:
Santosh Kumar
New Bharat Drug Agency
Ganj Road
P.O.- Dalsingsarai (pin- 848114)
Dist- Samastipur

e-mail: skg.research@gmail.com
phone: +91-9970747201