CLONING AND EXPRESSION OF GENES ENCODING DIVERGENT 4-COUMARATE : CoA LIGASE IN POPLAR

By

DANIELA CUKOVIC

B.Sc., University of Sarajevo, 1991

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

In

THE FACULTY OF GRADUATE STUDIES

(Department of Botany)

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

August 1999

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Botany

The University of British Columbia
Vancouver, Canada

Date August 19th 1993

DE-6 (2/88)

Abstract

The enzyme 4-coumarate:Coenzyme A ligase catalyzes the third and final step in the general phenylpropanoid metabolic pathway, providing activated thioester substrates for different phenylpropanoid products. To clone putative divergent members of the poplar *4CL* gene family, degenerate primers directed to a part of the first exon, containing a putative AMP-binding motif, were used. Three new 4CL-like classes were distinguished among 72 cloned PCR amplification products. The new 4CL-like sequence classes, arbitrarily named *4CL6*, *4CL10* and *4CL14*, shared 61% to 96% deduced amino acid sequence identity over the 600 bp region compared. A full-length cDNA clone of *4CL6* was isolated by screening a xylem cDNA library with the PCR fragment. Isolation of a full-length cDNA clone of *4CL10* involved both screening a young leaf cDNA library, and 5' and 3' RACE. Sequence analysis showed that the most divergent member of the poplar *4CL* gene family is *4CL10* (61% -65% amino acid identity with others), while *4CL6* is moderately divergent to previously isolated members. RNA blot analysis showed that the *4CL6* gene is expressed exclusively in xylem, green stem and root, while *4CL10* expression is restricted to young leaf, old leaf and root. Recombinant enzymes corresponding to the *4CL6* and *4CL10* genes were expressed using a baculovirus system. The recombinant 4CL6 was partially purified by FPLC, and showed a substrate utilization profile (4-coumaric acid > ferulic acid > caffeic acid > cinnamic acid, and no activity toward sinapic acid) similar to the profiles of the previously studied native and recombinant 4CL proteins.

Table of Contents

List of Figures and Tables

List of Abbreviations

AcM-NPV *Autographa californica* nuclear polyhedrosis virus

AMP adenosine monophosphate

AP1 adaptor primer

ATP adenosine triphosphate

BSA bovine serum albumin

CAD cinnamyl alcohol dehydrogenase

CCo3H 4-coumaryl:CoA 3-hydroxylase

CCoOMT caffeoyl:CoA *O*-methyltransferase

CCR cinnamoyl CoA reductase

C3H 4-coumarate 3-hydroxylase

C4H cinnamate-4-hydroxylase

4CL 4-coumarate:CoA ligase

CHS chalcone synthase

CoASH coenzyme A

COMT catechol *O*-methyltransferase

cDNA complementary DNA

ds cDNA double stranded cDNA

F5H ferulate 5-hydroxylase

FPLC fast protein liquid chromatography

GSP gene-specific primer

H11 poplar hybrid (*Populus trichocarpa* X *Populus deltoides*)

MOI multiplicity of infection

NGSP nested gene-specific primer

dNTP deoxynucleotide triphosphate

PAL phenylalanine ammonia lyase

PCR polymerase chain reaction

pfu plaque forming unit

PPi pyrophosphate

RACE rapid amplification of cDNA ends

RAPD randomly amplified polymorphic DNA

RFLP restriction fragment length polymorphism

SDS-PAGE sodium dodecyl sulfate polyacrylamide gel electrophoresis

Sf9 cells insect cells of *Spodoptera furgipedra*, line 9

Acknowledgement

I would like to thank Dr. Carl Douglas for giving me the opportunity of doing a master of science degree under his supervision, for his guidance in developing this thesis research, and for his financial assistance that he provided over the course of three summers. I am greatly indebted to Dr. Brian Ellis and Dr. Beverley Green for their suggestions and guidance in developing this thesis research.

I would like to thank Dr. Elizabeth Molitor for introducing me to the lab, and thanks to the Douglas, Ellis, Haughn and Kunst lab members for a co-operative and fun environment.

My special gratitude goes to Amrita Singh for providing degenerate primers and sharing PCR experience, to Dr. Sandra Allina and Dr. Cindy Shippam-Brett for their assistance with the baculovirus vector expression system, and to Janis Pereira for her assistance with the FPLC work.

CHAPTER I

Introduction

In plants, the phenylpropanoid pathway is responsible for the formation of several classes of chemical compounds including coumarins, flavonoids, stilbenes, suberin, lignin and other cell-wall associated phenolics. The functions of phenylpropanoid derivatives are as diverse as their structural variations. Phenylpropanoids serve as flower pigments (anthocyanins), antimicrobial compounds (isoflavans, isoflavonoids, stilbenes, psoralens, coumarins and flavonols), UV protectants (anthocyanins, flavones, psoralens and isoflavonoids) and signal molecules in symbiotic interactions (flavonoids). Phenylpropanoids with a complex heteropolymer structure, lignin and suberin, function as water-impermeable diffusion barriers in dermal tissues and vascular tissue; they are also induced in response to wounding (Hahlbrock and Scheel 1989; Dixon and Paiva 1995).

Phenylpropanoid metabolism can be divided into a general pathway, required for the synthesis of all phenylpropanoid metabolites, and specific branch pathways, which lead to synthesis of specific phenolic end products (Douglas *et al.*, 1992). The biosynthesis of phenylpropanoid compounds is developmentally activated in specific tissues and cell types, but can be induced by various biotic and abiotic stresses.

The evolutionary development of phenylpropanoid metabolism must have been a critical juncture in the transition of plants from an aquatic to a terrestrial environment, since it provided vascular plants with ability to transport water and nutrients and to cope with new abiotic and biotic stresses (Davin and Lewis, 1992).

1.1 General Phenylpropanoid Metabolism

The central pathway of phenylpropanoid metabolism consists of three core reactions (Figure 1). The aromatic amino acid phenylalanine is deaminated by the enzyme phenylalanine ammonia lyase (PAL) to produce cinnamic acid. The next enzyme in the pathway, cinnamate-4-hydroxylase (C4H), uses O_2 and NADPH to hydroxylate the 6-membered aromatic ring of cinnamic acid at the *para*-position, making 4-coumarate. The enzyme 4-coumarate:CoA ligase (4CL) catalyzes the last step in the general phenylpropanoid pathway, formation of Coenzyme A esters of *p*-coumaric acid and other hydroxy- or methoxy- derivatives of cinnamic acid, such as caffeic acid, ferulic acid, and sinapic acid (Figure 2). CoA esters of hydroxycinnamic acids serve as substrates for specific branch pathways. Thus, coumaryl-CoA is a substrate for the enzyme chalcone synthase (CHS), which catalyzes the branch point step in the biosynthesis of flavonoid compounds and coumaryl-CoA, feruloyl CoA and sinapyl CoA are reduced by cinnamoyl CoA reductase (CCR), and directed to the biosynthesis of lignin monomers.

Hydroxylation and methylation of the phenylpropanoid aromatic ring, important in the generation of different monolignols, can occur at the level of free acids, or at the level of the corresponding CoA-esters (Figure 2). The traditional view of the pathway has been that 4-coumaric acid is hydroxylated at the 3-position by 4-coumarate 3-hydroxylase (C3H). Enzymes that carry out this hydroxylation reaction in vitro have been detected, but little is known about their properties or physiological role (reviewed by Whetten and Sederoff, 1995).

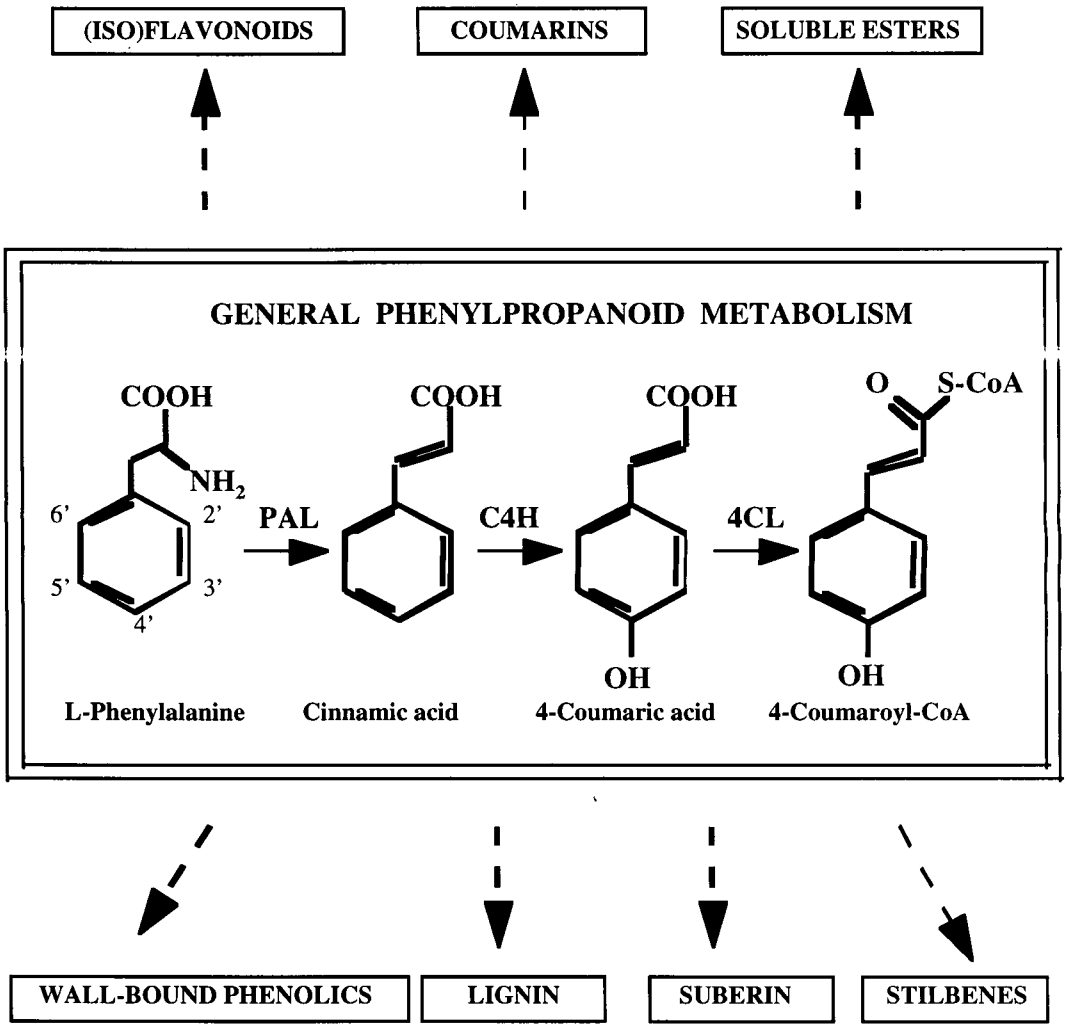


Figure 1. The reactions of general phenylpropanoid matabolism. Dashed arrows indicate branch pathways emanating from the general pathway. PAL, phenylalanine ammonia-lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate:CoA ligase (Hahlbrock and Scheel, 1989).

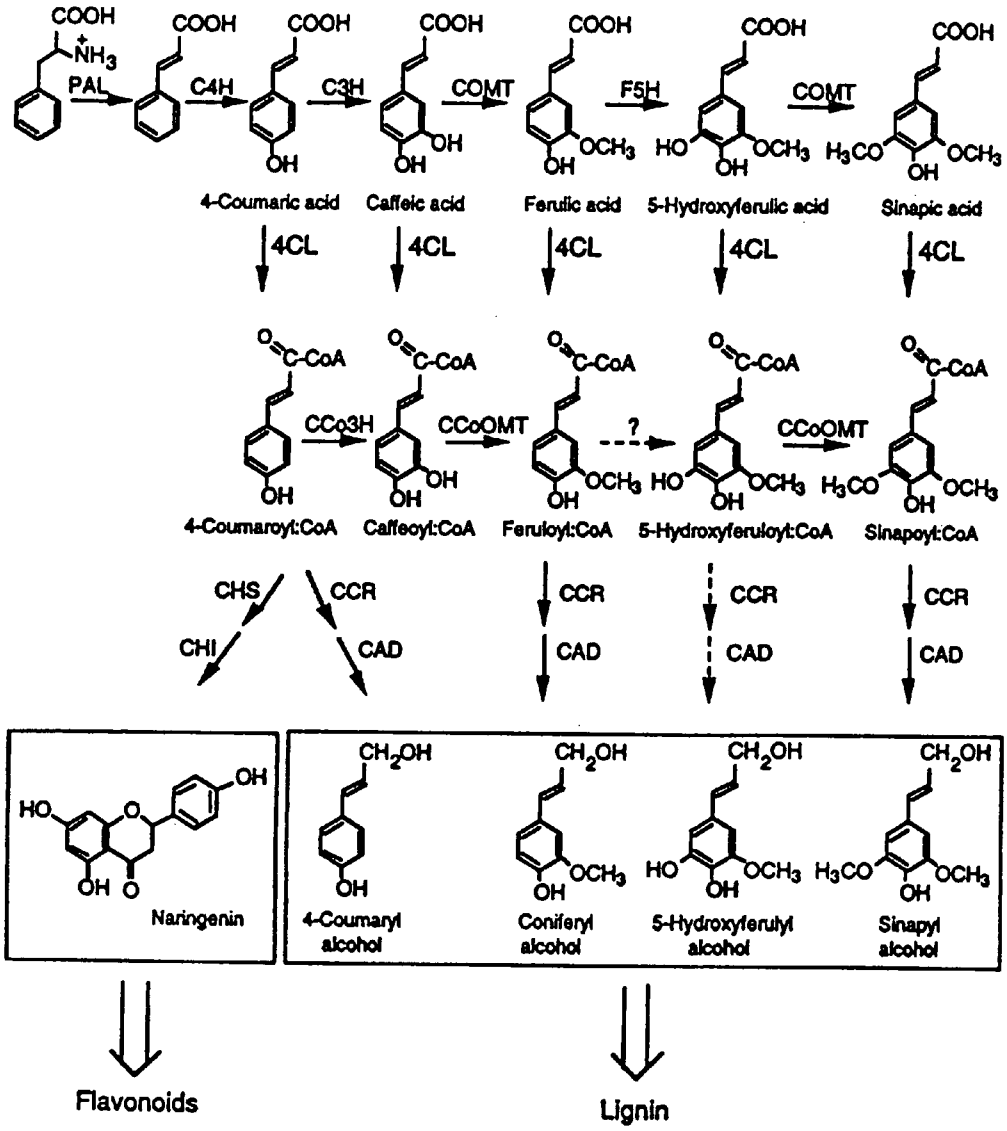


Figure 2. Schematic diagram of the biosynthesis of flavonoid- and lignin- precursors. The identification of novel enzymes suggests that the biosynthesis of CoA esters from 4-coumarate may proceed through a number of enzymatic steps potentially resulting in a metabolic grid rather than a linear pathway. C3H, 4-coumarate 3-hydroxylase; COMT, caffeic acid/5-hydroxyferulic acid O-methyltransferase; F5H, ferulate 5-hydroxylase; CCo3H, 4-coumaroyl:CoA 3-hydroxylase; CCoOMT, caffeoyl:CoA O-methyltransferase; ?, uncharacterized metabolic step; CHS, chalcone synthase; CHI, chalcone isomerase; CCR, cinnamoyl-CoA reductase; CAD, cinnamyl alcohol dehydrogenase. Dashed arrows represent enzymatic steps, which have not been clearly demonstrated (Whetten and Sederoff, 1995).

An alternative to hydroxylation of free 4-coumarate has been supported by discovery of a 4-coumaroyl/caffeoyl-CoA hydroxylase (CCo3H) in an anthocyanin mutant of *Silene dioica*, but the potential involvement of CCo3H in monolignol biosynthesis has not yet been adequately tested (reviewed by Whetten and Sederoff, 1995).

The 3- and 5-methylation of the aromatic ring of free acids is catalyzed by catechol *O*-methyltransferase (COMT), which in angiosperms utilize both caffeic acid and 5-hydroxyferulic acid (reviewed by Whetten and Sederoff, 1995). Ye *et al.* (1994) showed that a distinct OMT enzyme, caffeoyl CoA 3-*O*-methyltransferase (CCoOMT), acts at the level of the caffeoyl-CoA ester and 5-hydroxyferuloyl-CoA during monolignol biosynthesis (Figure 2). This portion of the monolignol biosynthesis pathway appears to be more of a network or grid than a linear pathway (Whetten *et al.*, 1998).

Hydroxylation of ferulate to 5-hydroxyferulate, thought to be catalyzed by ferulate 5-hydroxylase (F5H), appears to be a key step in sinapyl alcohol and syringyl lignin biosynthesis. A mutant *(fah-1)* in *Arabidopsis* shows a lack of sinapate-derived residues in lignin (Chapple *et al.*, 1992). Meyer *et al.* (1998) showed that overexpression of the F5H gene from the CaMV35S, or *AtC4H* promoter leads to biosynthesis of a large amount of syringyl lignin. Characterization of the catalytic properties of recombinant F5H protein has recently revealed that F5H has the best activity toward coniferyl aldehyde rather than free ferulic acid or feruloyl:CoA (Chapple, C.C., unpublished data).

1.2 4-Coumarate:CoA Ligase

The last step of general phenylpropanoid metabolism is the activation of cinnamic acids to form CoA thioesters. This reaction is catalyzed by 4-coumarate:CoA ligase

(E.C. 6.2.1.12 the group of acid-thiol ligases) and proceeds via a two-step process involving an acyl adenylate intermediates, in analogy to the activation of fatty acids with ATP and CoASH according to the following equations (Gross, 1985):

$$\text{4CL + hydoxycinnamic acid + ATP} \xrightarrow{Mg^{2+}} \text{4CL • hydroxycinnamyl-AMP + PPi}$$

$$\text{4CL • hydroxycinnamyl-AMP + CoASH} \rightarrow \text{4CL + cinnamoyl-CoA + ATP}$$

4CL activity requires the presence of ATP as a cosubstrate Mg^{2+} as a cofactor, and CoASH (Gross, 1985). Knobloch and Hahlbrock (1977) showed a sigmoidal dependence of parsley 4CL activity on ATP and CoASH, indicating an allosteric character for this enzyme.

During the 1970's and the early 1980's classical biochemical techniques were used to isolate and purify different CoA ligase isoforms from various plant species, such as *Forsythia*, soybean, *Petunia*, parsley, pea, carrot, spruce, *Erythrina crista-galli*, poplar and maize (Gross and Zenk, 1974: Knobloch and Hahlbrock, 1975; Ranjeva *et al.*, 1976; Knobloch and Hahlbrock, 1977; Heinzmann *et al.*, 1977; Wallis and Rhodes, 1977; Lüderitz *et al.*, 1982; Kutsuki *et al.*, 1982; Grand *et al.*, 1983; Vincent and Nicholson, 1987). The enzyme is monomeric and molecular weights of 40 kDa (*Erythrina crista-galli*), 55 kDa (soybean, *Forsythia*), 67 kDa (parsley) and 75 kDa (pea) were estimated.

Physically distinct 4CL isoforms have been reported from soybean, *Petunia*, pea, poplar, carrot and mesocotyl of maize, (Knobloch and Hahlbrock, 1975; Ranjeva *et al.*, 1976; Wallis and Rhodes, 1977; Grand *et al.*, 1983; Heinzmann *et al.*, 1977; Vincent and Nicholson, 1987). All 4CL isoforms examined to date have the highest activity toward 4-coumaric acid. However, in soybean, *Petunia*, and poplar, partially purified 4CL isoforms exhibit different substrate specificity toward substituted cinnamic acids (Knobloch and Hahlbrock, 1975; Ranjeva *et al.*, 1976; Grand *et al.*, 1983). It has been hypothesized that

4CL, with its ability to utilize a number of related substrates, could control the partitioning of carbon into different branch-pathways through the activity of distinct 4CL-isoforms (Knobloch and Hahlbrock, 1975; Grand *et al.*, 1983). For example, in *Petunia* leaves one isoenzyme preferentially utilizes caffeic acid (caffeate: CoA ligase), leading to the suggestion that it was involved in caffeic acid ester formation. The second isoenzyme (ferulate: CoA ligase) could be involved in guaiacyl lignin biosynthesis, and the third enzyme, which utilized sinapic acid, could take part in the formation of syringyl lignin (Ranjeva *et al.*, 1976).

In contrast, a single 4CL form was purified from *Forsythia, Erythrina crista-galli*, maize (leaves) and loblolly pine (Gross and Zenk, 1974; Lüderitz *et al.*, 1982; Hipskind *et al.*, 1993; Voo *et al.*, 1995), suggesting that 4CL does not participate in the metabolic channeling of phenylpropanoid derivatives in these species.

1.3 *4CL* Gene Families

More detailed knowledge about 4CL isoforms was provided by cloning of 4CL genes during the 1980's and 1990's. 4CL is encoded by multiple divergent genes in some plants like rice, soybean, poplar, *Lithospermum erythrorhizon*, tobacco and aspen (Zhao *et al.*, 1990; Uhlmann and Ebel, 1993; Allina and Douglas, 1994; Yazaki *et al.*, 1995; Lee and Douglas, 1996; Hu *et al.* 1998); by very similar duplicated genes as in the case of parsley, potato and loblolly pine (Lozoya *et al.*, 1988; Becker-André *et al.*, 1991; Zhang and Chiang 1997); and apparently by a single-gene in *Vanilla planifolia* (Brodelius and Xue, 1997). Interestingly, *Arabidopsis* 4CL was previously assumed to be encoded by a single-copy gene

(Lee *et al.*, 1995), but two divergent *4CL* classes have since been cloned in *Arabidopsis* by Ehlting *et al.* (1999).

Sequence comparisons between deduced amino acid sequences of *4CL* cDNAs of potato, parsley, soybean, rice and tobacco, and other cloned *4CL* genes (Becker-André *et al.*, 1991; Uhlman and Ebel, 1993; Lee and Douglas, 1996) indicate that the 4CL proteins contain a highly conserved seven amino acid motif "GEICIRG " clustered around a conserved cysteine residue. This motif is conserved in several apparently unrelated enzymes dependent on ATP, such as luciferase from firefly (deWet *et al.*, 1987; Schröder, 1989); tyrocidin synthetase A and tyrocidin synthetase from *Bacillus brevis*, and may be associated with catalytic activity (Backer-André *et al.*, 1991). A second conserved motif has been proposed (Bairoch, 1991) as a signature for a putative AMP-binding domain that is common to a number of prokaryotic and eukaryotic ATP-dependent enzymes. In addition, the predicted amino acid sequences of 4CL proteins from the plants listed above each contain a total of six conserved cysteine residues.

Genomic clones of *4CL* genes from rice, parsley and potato show the presence of five exons (the first exon is longest) and four introns (Zhao *et al.*, 1990; Lozoya *et al.*, 1988; Becker-André *et al.*, 1991). Zhang and Chiang (1997) reported that two loblolly pine *4CL* genes have three introns, whereas *Arabidopsis 4CL* genomic sequences have four introns at the conserved positions, and one of them contains three additional introns (Ehlting *et al.* 1999).

4CL gene expression, like that of many of the phenylpropanoid genes, is regulated developmentally and is also activated by external stimuli such as pathogen infection, elicitor treatment, wounding, and UV-light irradiation (Douglas *et al.*, 1987; Schmelzer *et al.*, 1989;

Wu and Hahlbrock, 1992). In tobacco flowers, *in situ* hybridization shows that endogenous tobacco *4CL* transcripts and those of an introduced parsley *4CL1* gene accumulate in a cell-type specific manner, and that the patterns of accumulation are generally consistent with the sites of phenylpropanoid natural-product accumulation (Reinold *et al.*, 1993). As well, *4CL* expression in tobacco is activated by wounding, light, and methyl jasmonate treatment (Douglas *et al.*, 1991; Ellard-Ivey and Douglas, 1996). The *4CL* genes from parsley and soybean are differentially regulated. *4CL2* from parsley is preferentially expressed in the flowering stem and is light inducible, whereas *4CL1* is wound inducible in roots (Lois and Hahlbrock, 1992). In soybean, *4CL16* is inducible by fungal infection whereas *4CL14* is not (Uhlmann and Ebel, 1993).

1.4 Phenylpropanoid Metabolism in Poplar

Phenylpropanoid metabolism plays an important role in the growth and development of woody plants, since an important component of wood is lignin. For example, between 21% to 23 % of *P. trichocarpa* wood dry weight is lignin (Swan and Kellogg, 1986).
Many derivatives of phenolic compounds have been isolate from the bark of *P. trichocarpa* such as salicin, trichocarpin, salireposide, salicyl alcohol, cinnamic acid, 4-coumaric acid and others (Pearl and Darling, 1968). Bud exudate of *P. deltoides* is predominantly composed of flavanones, chalcones and ester of flavanones together with the flavone galangin, the flavanon pinocembrin and the flavanonol pinobanksin (Greenaway *et al.*, 1990). In contrast, the major phenolic compounds in bud exudate of *P. trichocarpa* are dihydrochalcone, benzyl salicylate, cinnamic acid, and minor amount of flavanones, chalcones and flavones (English *et al.*, 1991).

Shain and Miller (1982) reported that pinocembrin (5,7-dihydroxyflavanone), as a major component of poplar bud exudate is active against a fungal pathogen *Melampsora medusae.* Thus, a sufficient amount of pinocembrin is present on the surface of young, expanding leaves of *P. deltoides* contributing to their resistance to *M. medusae.* As leaves age, however, the concentration of pinocembrin is depleted as a result of weathering, leaves expansion and insufficient replenishment, and leaves become more susceptible to the fungal infection.

Induction of defense response in poplar is associated with induction of phenylpropanoid metabolism. Hybrid poplar suspension-cultured cells treated with elicitor showed a coordinated and transient increases in extractable PAL and 4CL enzyme activity at about 7 h after elicitation (Moniz de Sà *et al.*, 1992). These increases were proceeded by an accumulation in extractable cell wall-bound phenolic compounds.

Despite of the economic importance of wood, little is known about the genetic control of wood formation. Poplar is an useful organism as a pulp source for paper and as a model organism for the study of physiology, genetics and molecular biology of a woody perennial. Poplar is easily propagated vegetatively, transformed by *Agrobacterium tumefaciens* and regenerated as a transgenic tree (Laple *et al.,* 1992; Kajita *et al.*, 1994). Furthermore, a three generation pedigree is available for genetic analysis, derived from a cross between *Populus deltoides* and *P. trichocarpa.* A *Populus* genome linkage map was constructed using RAPD and RFLP markers in this pedigree (Bradshaw *et al,* 1994).

Several phenylpropanoid genes have been characterized in the genus *Populus* (poplars and aspens). PAL is encoded by a small gene family, two members of which (*PAL1* and *PAL2*) are highly expressed in young leaves and stems, where large amounts of soluble

phenylpropanoid products accumulate in addition to lignin (Subramaniam *et al.*, 1993). Two more divergent members of the *Populus* gene family have recently been identified, which, in contrast to *PAL1* and *PAL2*, are strongly expressed in older stems in which secondary xylem is undergoing differentiation (Osakabe *et al.*, 1995). This differential expression indicates that specific *PAL* gene family members could be used to direct phenylalanine into secondary metabolism in different tissues.

The studies by Grand *et al.* (1983) strongly suggest the possible existence of multiple 4CL isoenzymes in poplar. Three partially purified 4CL isoenzymes were obtained from a poplar "*euramericana*" hybrid stems by chromatofocusing, using 4-coumaric acid as a substrate to measure enzyme activity. The molecular mass of 41 kDa is similar to the value for 4CL from *Erythrina crista-galli* (40 kDa), but lower then values determined for other plants (Grand *et al.*, 1983). The partially purified poplar 4CL isoenzymes described in that study showed differences in substrate specificity. 4CL1 reacts with 4-coumarate, ferulate, and sinapate as substrates; 4CL2 can use 4-coumarate and ferulate as substrates; and 4CL3 can use 4-coumarate and caffeate as substrates. The 4CL activity was mainly localized in lignified tissues, such as xylem and sclerenchyma (phloem fiber cells) (Grand *et al.*, 1983).

1.5 Poplar *4CL* Gene Family

Recent characterization of poplar 4CL isoenzymes and members of the *4CL* gene family has been done in a poplar hybrid, clone H11-11, derived from a cross between *Populus trichocarpa* and *Populus deltoides*. Two classes of *4CL* cDNA clones (*4CL9* and *4CL216*), that do not cross hybridize at high stringency, were isolated by screening an H11 young leaf cDNA library with a parsley *4CL* cDNA probe (Douglas *et al.*, 1992; Allina and

Douglas, 1994; Allina *et al.*, 1998). The clones showed about 86% identity at the nucleotide and amino acid levels. Genomic Southern blots revealed restriction fragments that strongly cross-hybridize under high stringency conditions specifically either with *4CL216* and *4CL9*. However, several weakly hybridizing bands were visible, especially after low stringency hybridization, suggesting that the *4CL* gene family might include a divergent gene(s) that weakly hybridize with *4CL9* and *4CL216*. The results of northern analysis indicated that the *4CL9* gene is expressed strongly in young leaf and weakly in old leaf, green stem and xylem, while the *4CL216* gene shows strong expression in all those tissues. Neither the *4CL9* or the *4CL216* genes is expressed in elicitor-treated suspension tissue culture cells.

Expression of recombinant 4CL proteins from these cDNA clones showed that they have identical substrate utilization profiles (Allina *et al.*, 1998). Furthermore, 4CL forms from poplar hybrid H11-11, xylem, elicited cell culture, and young leaf were separated by FPLC ion column exchange, indicating the presence of 3-4 putative 4CL isoforms. However, the different forms appear to have identical substrate utilization profiles (Allina *et al.*, 1998). These results do not support previous evidence that three catalytically distinct 4CL forms exist in poplar.

1.6 Lignin

Lignin (Latin: lignum – wood) is a three-dimensional heteropolymer, resulting from the dehydrogenative polymerization of three different hydroxycinnamyl alcohols (monolignols): *p*-coumaryl, coniferyl and sinapyl alcohol, which differ in the extent of methylation. These monolignols become cross-linked by about 20 types of intersubunit linkages (Chen, 1991).

Lignin accumulates in the secondary wall and the middle lamellae of specialized cells such as tracheids, vessel elements, xylem and phloem fibers, and sclereids. When the lignin deposition is complete, the protoplast degenerates leaving a dead cell protected against collapse by its strengthened walls (Boudet *et al.*, 1995). Lignin provides rigidity and hydrophobicity to the cell wall, which is particularly important for cells which conduct water. Also, stress-induced lignin deposition provides a mechanism for sealing off sites of pathogen infection and wounding (Dixon and Paiva, 1995).

The monomeric composition of lignins is one of the most important characteristics of lignins from both a taxonomic and an industrial point of view (Boudet, 1998). The chemical complexity of lignin has increased from pteridophytes and gymnosperm lignins to the angiosperms. Grasses have lignin with the most complex composition (Boudet *et al.*, 1995). In conifers, lignin is typically composed only of coniferyl alcohol (guaiacyl or G units), although, in compression wood, lignin is polymerized from coniferyl and a small portion of p-coumaryl alcohol (H units). This composition makes compression wood lignin less methylated and more difficult to hydrolyze (Higuchi, 1985). The typical lignin of angiosperms is composed of a mixture of guaiacyl and syringyl units. The presence of the more highly methylated syringyl units increases the ability of hardwood lignin to be hydrolyzed (Sederoff *et al.*, 1994). In monocotyledon plants, lignins contain all three monolignols, but they also have ester- and ether- linked hydroxycinnamic acids, which could be derived from L-tyrosine (Higuchi, 1985; Lewis and Yamamoto, 1990).

Lignin composition and intermolecular linkages vary according to cell type, stage of tissue development and individual cell wall layers (Davin and Lewis, 1992). For example, guaiacyl subunits predominate in the xylem of vascular bundles of *Arabidopsis* stems, while

adjacent, heavily lignified sclerenchyma cells contain syringyl units (Chapple, *et al.* 1992; Meyer, *et al.* 1998).

Monolignols are derived from CoA esters of hydroxycinnamic acids via a two-step reduction process catalyzed by cinnamyl-CoA reductase (CCR) and cinnamyl alcohol dehydrogenase (CAD) (Griesbach, 1981). Once made, the lignin monomers may be glycosylated at the 4-hydroxy position and then transported to the site of lignin deposition. The mechanism of monolignol transport from cytoplasm to the cell wall is not well understood. In conifers it has been hypothesized that monolignol glucosides are stored in the vacuole, transported from symplast to apoplast, and then de-glycosylated (Whetten and Sederoff, 1995). The polymerization of monolignols to make lignin is believed to be catalyzed by some combination of peroxidase and laccase (reviewed by Campbell and Sedoroff, 1996; Whetten *et al.*, 1998).

Proposed mechanisms that control lignin content and composition are (i) substrate specificity of different isoforms of monolignol biosynthetic enzymes (e.g. COMT, 4CL, CAD) and (ii) transcriptional regulation of genes encoding isoforms of monomer specific pathways (e.g. PAL, 4CL, CAD) (Higuchi, 1985; Sederoff *et al*, 1994; Campbell and Sederoff, 1996). The existence of 4CL isoforms in some plants with different substrate utilization profiles has led to the proposal that such isoforms may help control carbon flow into different lignin monomer (Knobloch and Hahlbrock, 1975; Ranjeva *et al*, 1976; Grand *et al.*, 1983).

Characterization of lignin content and composition in mutants of maize, sorghum, and *Arabidopsis,* and in transgenic tobacco, poplar and *Arabidopsis* with suppressed enzyme activities, has led to increased understanding of the monolignol biosynthesis pathway, and of

the mechanisms by which lignin composition is regulated in different plants. For example, lignin content remained mostly constant in transgenic plants with suppressed OMT activity in tobacco and poplar (Atanassova *et al.*, 1995, van Doorsselaere *et al.*, 1995), suppressed CAD activity in tobacco and poplar (Halpin *et al.*, 1994; Hibino *et al.* 1995; Baucher *et al.*, 1996), suppressed 4CL activity in *Arabidopsis* (Lee *et al.*, 1997), and ectopic F5H activity in *Arabidopsis* (Meyer *et al.*, 1998). In contrast, changes in lignin composition (increases or decreases in the S:G ratio) have been observed in many transgenic plants with altered enzyme activities generated by suppression of PAL, 4CL, COMT and CCR expression or ectopic F5H expression (reviewed by Whetten *et al.*, 1998). Therefore, the amount of lignin in a plant and a given tissue can be maintained at a relatively constant level even when an enzyme normally involved in the synthesis of lignin precursors is suppressed or overexpressed. This response reflects the importance of lignin to vascular function and mechanical support in land plants, and indicates a high degree of plasticity in lignin biosynthesis.

When this project was initiated, the nature of the *4CL* gene family in poplar and other plants was poorly understood. While different isoforms had been described in many plants, the extent to which the enzymatic properties of these isoforms differed was unknown. Furthermore, it was possible that genes encoding highly divergent isoforms might not have been detected due to lack of cross-hybridization to cloned genes.

The purpose of my project was to test the hypothesis that the poplar *4CL* gene family contains, besides the already isolated genes, other genes that encode divergent 4CL enzymes. It was further hypothesized that such divergent genes have different expression patterns, and that the corresponding isoenzymes have distinct substrate specificities.

CHAPTER II

Materials and Methods

2.1 Plant Material

Clonally propagated individuals of *Populus trichocarpa* 93-968 and poplar hybrid (*P. trichocarpa* Torr. & Gray X *P. deltoides* Marsh) H11 were used for organ and tissue isolation. Young leaves (0.5-2 cm in length), old (fully expanded) leaves and green (nonwoody) stems were harvested from H11 plants maintained in growth chambers at 23°C in a 16-h light/8-h dark regime. Harvested material was immediately frozen in liquid N_2 and stored at –80°C until use. Secondary xylem was isolated from field-grown trees as described previously (Allina *et al.*, 1998). The suspension cultures of H11 cells and elicitor treatments were prepared as described by Moniz de Sá *et al.* (1992).

2.2. Preparation of Genomic DNA for PCR Analysis

Genomic DNA, used in PCR reactions, was isolated from fresh leaves of chamber-grown *P. trichocarpa* cuttings by a CTAB method as described by Roger and Bendich (1988).

2.3. PCR Search of Poplar Genomic DNA

Poplar (*P. trichocarpa*) genomic DNA (~100 ng) was used as a template for PCR amplification with 20 pmol of each degenerate primer, 200 μM dNTPs, 1.5 mM $MgCl_2$ and 2.5.units of Taq DNA polymerase (Gibco-BRL) in 50 μl reaction. Conditions for PCR runs with primer combinations 3 & 6 and 3 & 8 were 94°C/ 10 min, 35 cycles of 94°C/ 50 sec, 55°C/50 sec, 72°C/70 sec, and 72°C/10 min. Primer combinations 3 & 2 and 3 & 4 were

used under conditions: 94°C/ 10 min, 3 cycles of 94°C/ 50 sec, 50°C/50 sec, 72°C/70 sec, 30 cycles of the same conditions as previously described. PCR reactions without the template, or reactions with one primer and the template were negative controls for each PCR run. PCR fragments were digested with *Bam* HI and *Xba*I, and subcloned into a pBluScript plasmid.

2.4 Screening cDNA Library

About 4 x 10^5 recombinant phage from a poplar xylem cDNA library in λZAPII (prepared by Yuji Tsutsumi) were screened using the 600 bp PCR fragments of *4CL6*, *4CL10* and *4CL14* as described by Sambrook *et al.* (1989). Probes were radioactively labelled using the Random Primer DNA Labelling System (Gibco-BRL) according to the manufacturer's specifications. Hybridized filters were washed at high stringency (0.2 X SSC, 0.1% SDS, 65°C) for 1 h. After purification of positive plaques, a single positive clone obtained was subjected to in vivo excision to yield pBluescript phagemid with helper phage strain ExAssist (Stratagene). The clone was double digested with *Xba*I and *Xho*I, and separated on 1% agarose gel, followed by partial sequencing of the 5' and 3' ends of the clone. It was shown that the clone was a full-length (~ 2 kb) cDNA of *4CL6*. The same approach was applied in screening a λZAPII (Stratagene) H11 young-leaf cDNA library (Subramaniam *et al.*, 1993) with the 600 bp *4CL10* PCR fragment. Eight positive plaques were purified and further characterized. All clones contained an incomplete (~ 1000 bp) cDNA fragment of *4CL10*.

2.5 Rapid Amplification of 5'-cDNA and 3'-cDNA Ends

Total RNA was isolated from young leaves of hybrid poplar H11 by the method of Hughes and Galau (1988) from young leaves. Purification of poly A+ RNA from total RNA was based on a protocol using dynabeads (Dynal). A library of uncloned, adaptor ligated double strand cDNA was generated using Marathon cDNA amplification kit (Clontech Laboratories, Inc.) and polyA+ RNA from young leaves. Taq polymerase with high fidelity (Boehringer) and the adaptor ligated cDNA, as a template, were used in PCR of 5'RACE and 3' RACE. The first PCR of 3'RACE was performed with the adaptor primer AP1 (5' –CCATCCTAATACGACTCACTATAGGGC– 3') and a gene specific primer 4CL10.2 at the position 950 (5' -AGAACCCAATGGTGGCGAACTTCGAC- 3'). To confirm that the amplified region was derived from the desired target, the second PCR of 3'RACE was performed with AP1 and a nested gene-specific primer, 4CL10.4 at the position 1067 (5' –GGGACAGGGTTATGGGATGACAGAG- 3'). Conditions for both PCR runs were 94°C/1 min, 30 cycles of 94-°C/ 30 sec, 65°C/30 sec, 68°C/2 min. An antisense gene-specific primer (GSP), 4CL10.7 at the position 1125 (5' –AGGCTAAGCA CATTGATAGCACTGGC- 3'), and AP1 were used in PCR of 5'RACE under these conditions: 94°C/1 min, 30 cycles of 94 °C /30 sec, 62°C/30 sec, 68°C/3 min. Primers 4CL10.0 at the position 1 (5' –ATGATGTCCGTGGCCACGGTTGAG- 3') and 4CL10.9 at the position (5' –GGGCACAATGAGTGAAGACAACACA- 3') were used to amplify full-length cDNA of *4CL10* (conditions: 94°C/1 min, 30 cycles of 94 °C /30 sec, 60°C/30 sec, 68°C/3 min). The amplified DNA fragments were cloned into the pCR2.1 plasmid vector using the TA-cloning kit (Invitrogen).

2.6. DNA Sequencing and Sequence Analysis

PCR fragments and cDNA clones were sequenced by the University of British Columbia Nucleic Acid-Protein Service Unit using the PRISM Ready Reaction DyeDeoxy Terminator Cycle Sequencing kit (Applied Biosystems) and DNA sequencer 373 (Applied Biosystems). Sequencing was carried out on DNA isolated either by a modified mini alkaline/lysis/PEG precipitation procedure (suggested by the NAPS) or by a Qiagen plasmid kit. DNA and predicted amino acid sequences were analyzed using University of Wisconsin Genetics Computer Group (Medison) software. Sequenced were searched against the GenBank non-redundant protein and EST database (National Center of Biotechnology Information) using BLASTX. Multiple sequence alignment was obtained using Clustal-W program (Baylor College of Medicine) and BOXSHADE. Predicted molecular weight of deduced amino acid sequences was calculated by software Compute pI/Mw Tool (http://www.expasy.ch/tools/pi_tool.html).

2.7 RNA Extraction and Northern Blot Analysis

Total RNA was isolated by the method of Hughes and Galau (1988) from young leaves, old leaves, green stem and root of poplar hybrid clone H11. A trizol reagent (Gibco-BRL) was used to extract total RNA from secondary xylem and the suspension cultures of H11 cells, according to the manufacturer's recommendations. RNA (10 μg per lane) was denatured by 50% formamide and 2.2 M formaldehyde at 65°C for 15 min.; separated by 2.2 M formaldehyde and 1.2% agarose gel electrophoresis; rinsed with water for 45 min; stained with 0.5 μg/ml and blotted onto Hybond–N nylon membrane (Amersham). Probes generated from *4CL6* and *4CL10* cDNA clones (1.9 kb), were radioactively labelled using

the Random Primer DNA Labelling System (Gibco-BRL) according to the manufacturer's specifications. Prehybridization and hybridization were performed at 65°C for 16 h in 7% SDS sodium phosphate buffer (Church and Gilbert, 1984). The membranes were washed at high stringency (0.2 X SSC, 0.1% SDS, 65°C) for 1 h, and autoradiographed for 2 days (the blot probed with *4CL6*) or 5 days (the blot probed with *4CL10*). To demonstrate evenness of loading between lanes, the blots were stripped and rehybridized with a probe for a pea rRNA (Jorgensen *et al.*, 1982).

2.8 Generating Recombinant 4CL6 Baculovirus Particles

To subclone the full-length cDNA clones of *4CL6* and *4CL10* into pVL1392, a Baculovirus transfer vector, pVL1392, was digested by *Sma* I and *Not* I.
The *4CL6* cDNA was cut out from a BlueScript plasmid using *Xho*I and *Not* I, and a blunt end was created at the *Xho*I site by filling in with Klenow polymerase, while the *4CL10* cDNA was cut out from pCR 2.1 plasmid (TA plasmid, Invitrogene) using *Kpn* I and *Not* I, and a blunt end was created at the *Kpn* I site. Recombinant plasmids: pVL1392::*4CL6* and pVL1392::*4CL10*, created after ligation, contained flanking sequences, which were homologues to the Baculovirus genome.

To transfer the genes into genome of *Autographa californica* nuclear polyhedrosis virus (AcM-NPV), 2 μg of highly purified DNA of the pVL1392::*4CL6* construct and 0.2 μg of AcM-NPV (BaculoGold DNA, Pharmingen) were co-transfected into *Spodoptera furgipedra,* Sf9 insect cells (American type Culture collection, Accession Number CRL-1711). Recombination took place within insect cells between the homologous regions in the transfer vector and the BaculoGold DNA. Recombinant baculovirus particles were plaque-

purified (Summers and Smith, 1987). Based on an immunoblot and 4CL-enzyme essay, a single recombinant 4CL6 baculovirus plaque was chosen for production of high-titre virus stock. The viral titre (2.45 $X10^8$ pfu/ml) was calculated by end point dilution (Summers and Smith, 1987). Expression and activity of the recombinant 4CL6 protein was tested over time (every 12 h/3 days). A multiplicity of infection (MOI) of 1.5 and 36-h infection time was selected based on 4CL enzyme assay as the best condition for production of the recombinant 4CL6 protein.

2.9 Expression and Purification of Recombinant 4CL

50 x 10^6 Sf9 cells were infected with the recombinant 4CL6 baculovirus at MOI 1.5 and harvested 36 hours after infection. Cells were centrifuged at 1000 xg, for 5 minutes at 4°C. The cell pellet was washed twice with Dulbecco's Phosphate Buffered Saline (Sigma), resuspended in 4 ml 50mM Tris, pH 7.8, and lysed in a 15 ml Weaton Homogenizer. Cellular debris was removed through centrifugation at 15,000 xg, for 10 minutes, then the supernatant was filtered through a 0.22 μm filter. The supernatant was loaded onto a High Q column (Bio-Rad) and subjected to anion-exchange FPLC as described by Allina *et al.* (1998). Fractions were collected (1 ml), glycerol was added to 30% (v/v) and they were stored at –20°C. Protein content in crude extract and FPLC fractions was quantified by the Bradford (1976) method using BSA as a standard and the Bio-Rad Protein Assay Kit.

2.10 4CL Enzyme Assays

4CL activity was measured at room temperature spectrophotometrically as described by Knobloch and Hahlbrock (1977) using 5 mM ATP, 5 mM $MgCl_2$, 470 mM Tris, pH 7.8,

0.33 mM CoA, and 0.2 mM cinnamic acid derivatives as substrates. The pH of the reaction mixture was adjusted to pH 7.8 with 4 M KOH (0.5 ml per 10 ml). The change in absorbance of the reaction mixtures was monitored at wavelengths of 311, 333, 346, 345, or 352 nm, according to the absorption maxima for cinnamoyl:CoA, 4-coumaroyl:CoA, caffeoyl:CoA, feruloyl:CoA, and sinapoyl:CoA, respectively (Stöckigt and Zenk, 1975).

2.11 SDS-PAGE and Immunoblots

Protein samples were electrophoresed in 10% SDS-polyacrylamide separating gels as described by Laemmli (1970). The proteins were blotted onto nitrocellulose Hybond-C (Amersham), blocked with 5% (w/v) nonfat powdered milk, probed with antisera raised against the recombinant 4CL9 at a 1:5000 dilution (Allina *et al.* 1998), and reacted with goat anti-rabbit IgG conjugated to alkaline phosphatase (Pharmacia) at 1:5000 dilution. Alkaline phosphatase activity was visualized using nitroblue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphatase as the substrates (Gibco-BRL).

CHAPTER III

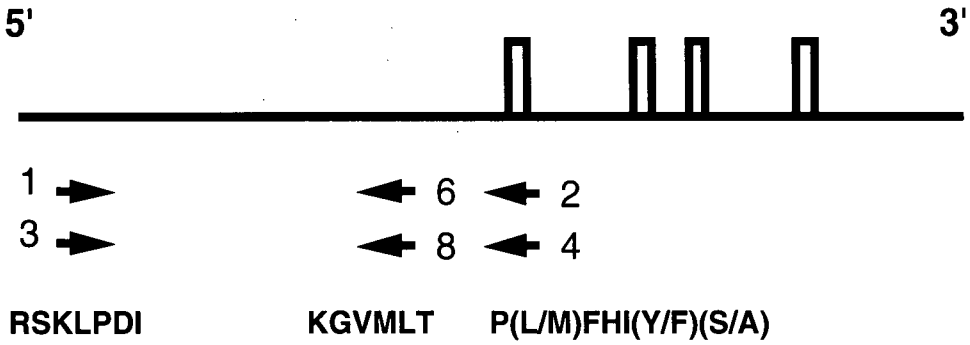
Results

3.1 Cloning Divergent *4CL* Genes

3.1.1 PCR-Based Search

Two *4CL* cDNA clones, *4CL216* and *4CL9,* were previously isolated from a H11 young leaf cDNA library, and shown to be derived from two poplar *4CL* gene family members, *4CL1* and *4CL2* (Allina *et al.*, 1998). The identity between the predicted amino acid sequences of these two clones is 86%. To clone other members from the *4CL* gene family in poplar, a PCR strategy was employed. Genomic DNA of *Populus trichocarpa*, clone 93-968, was used as a PCR template for amplification of *4CL* genes using sets of degenerate *4CL*-gene-specific primers designed by Amrita Singh. These primers had been designed based on amino acid sequences conserved between all predicted 4CL proteins from several plants. The primers were directed to a part of the first exon containing a conserved putative AMP-binding motif (Figure 3).

Forward primers 1 and 3 correspond to the nucleotide sequence of seven amino acids conserved near the N-terminus of 4CL proteins. This site contains consensus arginine, serine, and leucine residues, which have high codon degeneracy (six codons each). In order to reduce the degeneracy of primer 3, two codons for arginine (CGG, CGA) and two codons for leucine (UUG, UUA) were selected as the codons of preference in known *4CL* genes. Also, a deoxyinosine was incorporated at the third position of the serine codon, further decreasing the complexity of primer 3. Primer 1 differs from the primer 3 only in the presence of deoxyinosine at the third position of the leucine codon (Figure 3).



primer 1 5' CG GGA TCC G(A/G) CIA A(A/G)T TIC CIG A(CT)A T 3'
Bam HI

primer 3 5' CG GGA TCC G(A/G) CIA A(A/G)T T(A/G)C CIG A(C/T)A T 3'
Bam HI

primer 2 5' GCTCTAGA GA(G/A) TA(AGT) AT(A/G)TG(A/G) AAI AG(A/G)GG 3'
Xba I

primer 4 5' GC TCTAGA GA(GA)TA(AGT) AT(G/A) TG(A/G) AA(T/C) AA(A/G) GG 3'
Xba I

primer 6 5' GCT CTA GAG TIA GCA T(C/A)A CIC C(C/T)T TT 3'
Xba I

primer 8 5' GCT CTA GAG T(T/A)A ACA T(C/A)A CIC C(C/T)T T 3'
Xba I

Figure 3. Generic *4CL* gene and degenerate primers. A generic *4CL* gene has five exons and four introns. Primers 1 and 3 are forward degenerate primers corresponding to the sequence which encodes seven conserved amino acids (RSKLPDI) at the 5' end of the first exon. Primers 2, 4, 6 and 8 are reverse oligodeoxynucleotides directed to the conserved sequences encoding amino acids KGVMLT and P(L/M)FH(Y/F)(S/A), respectively. Each primer incorporates a restriction enzyme recognition site to facilitate PCR product cloning into a pBlueScript vector.

Reverse primers 6 and 8 were specific to the codons of six conserved amino acids located next to the putative AMP-binding motif. Another set of reverse primers, 2 and 4, were directed to a region downstream from the putative AMP-binding motif (Figure 3). Alternative conserved amino acids of this region are found in different 4CL proteins, such as leucine/methionine, tyrosine/phenylalanine and serine/alanine. Thus, primers 2 and 4 recognized the same region, but they accounted for codons of the different amino acids. All primers incorporated restriction enzyme recognition sites at their 5' ends, to facilitate PCR product cloning into the pBlueScript KS+ plasmid. PCR reactions using combinations of forward (1,3) and reverse (2,4,6,8) primers were expected to generate amplified products about 600-750 bp in length.

PCR reactions using the combination of primer 1 with any reverse primer, and genomic DNA as a target, yielded either attenuated amplification or no amplification product (Figure 4A: lane 3, and 4B: lane3). Therefore, all PCR reactions were carried out using primer 3 as the forward primer. Figures 4A and 4B show that primer sets 3+6 and 3+8 amplified a fragment of ~ 600 bp, as predicted, without background bands. In contrast, primer sets 3+2 and 3+4 amplified fragments of the predicted size (~750 bp), plus several fragments of different size (Figures 4C and 4D). Also, PCR negative controls with the template and single primers indicated that primer 4 alone generated non-specific products. Potential contamination of primers with DNA from other sources was excluded, because neither of the primer sets in negative controls without the template amplified any fragment.
After subcloning the amplification products into *Bam*HI / *Xba*I-digested pBlueScript KS+ plasmid, 72 clones were obtained. Each clone was fingerprinted by restriction enzyme analysis using *Eco* RI, *Eco* RV, *Hind* III, *Sac*I and *Stu*I (Figure 5).

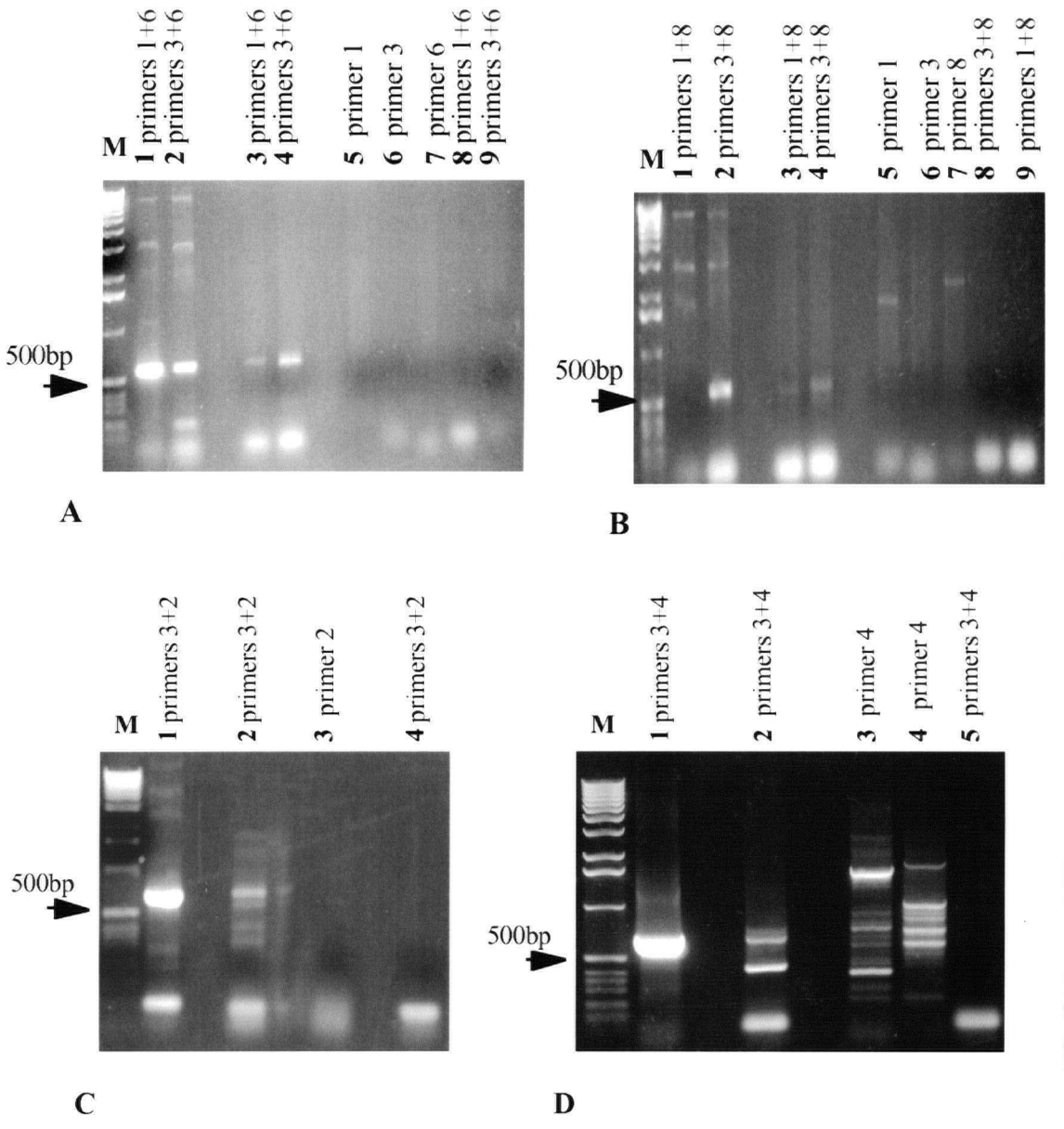


Figure 4. PCR amplification of 4CL gene fragments from *P. tichocarpa* using different degenerate primers. **A**, amplification with primer sets 1-6 and 3-6; lanes 1 and 2 with cDNA of 4CL9 as a template; lanes 3-6 genomic DNA as template; line 7- 8 with no template. **B**, amplification with primers 1-8 and 3-8; lanes 1 and 2 with cDNA; lanes 3-7 with genomic DNA; lanes 8 and 9 without template. **C**, amplification with primer set 3-2; lane 1 with cDNA, lane 2and 3 with genomic DNA; lane 4 primers 3 and 2 with no template. **D**, amplification with primer set 3-4; lane 1 with cDNA; lane 2 and 3 with genomic DNA; lane 4 genomic DNA and primer 4 (the second stock); lane 5 with no template.

Clones were placed into classes based on their restriction fragment length polymorphism (RFLP). These fell into *4CL9*-like, *4CL216*-like and classes containing potentially new 4CL sequences.

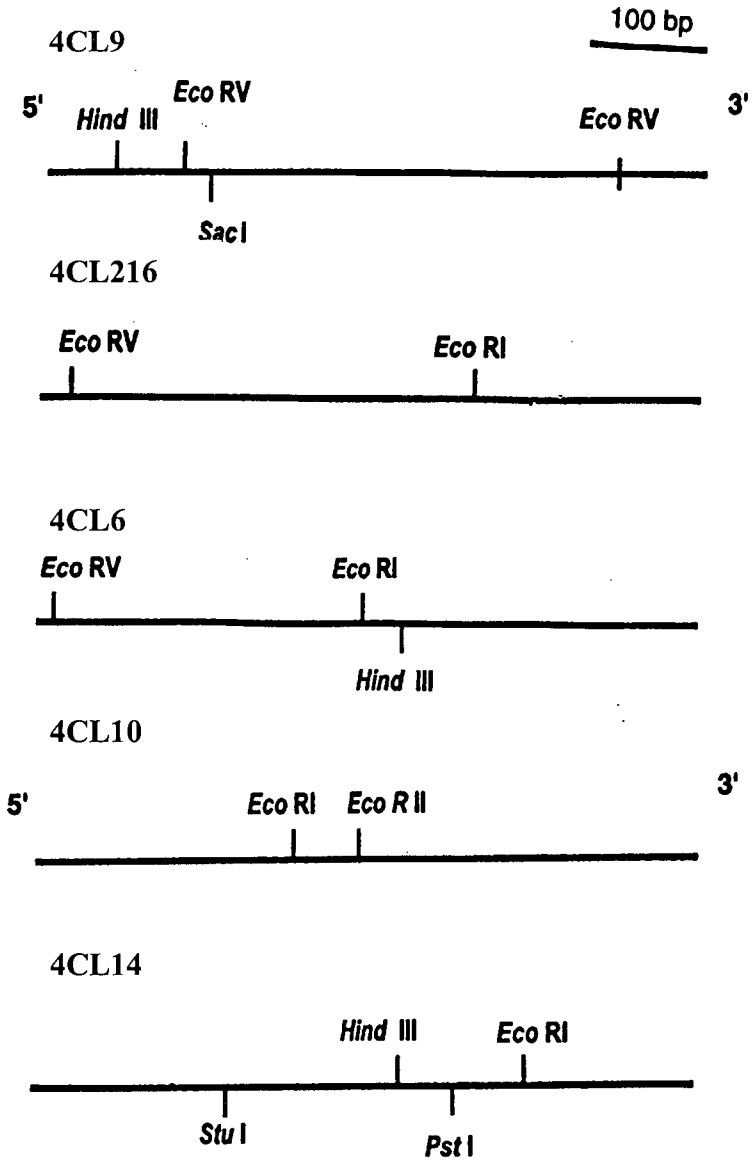


Figure 5. Restriction map of the PCR amplified DNA fragments (~600 bp) of poplar *4CL* genes.

Sequence analysis showed that all clones from the *4CL9* RFLP class contained the predicted amplified region of the *4CL9* gene, but neither sequence from the putative *4CL216* RFLP class matched the corresponding amplified region of the *4CL216* gene. Furthermore, sequence analysis showed that three of the eight RFLP classes represented new *4CL*-like sequences: *4CL6*, *4CL10* and *4CL14*. The remaining five classes contained non-*4CL* sequences. BESTFIT and GAP sequence analysis revealed that the *4CL6*, *4CL10* and *4CL14* classes share between 65% and 73% nucleotide identity with the previous two poplar *4CL* cDNAs (*4CL216* & *4CL9*). *4CL6* and *4CL14* share 89% identical nucleotide sequence, and 90% identical amino acid sequence, respectively (Figure 6). Since *4CL6* and *4CL10* showed relatively high divergence from each other (61% identity at the predicted amino acid level), PCR clones from these classes were chosen for cloning full length cDNA clones.

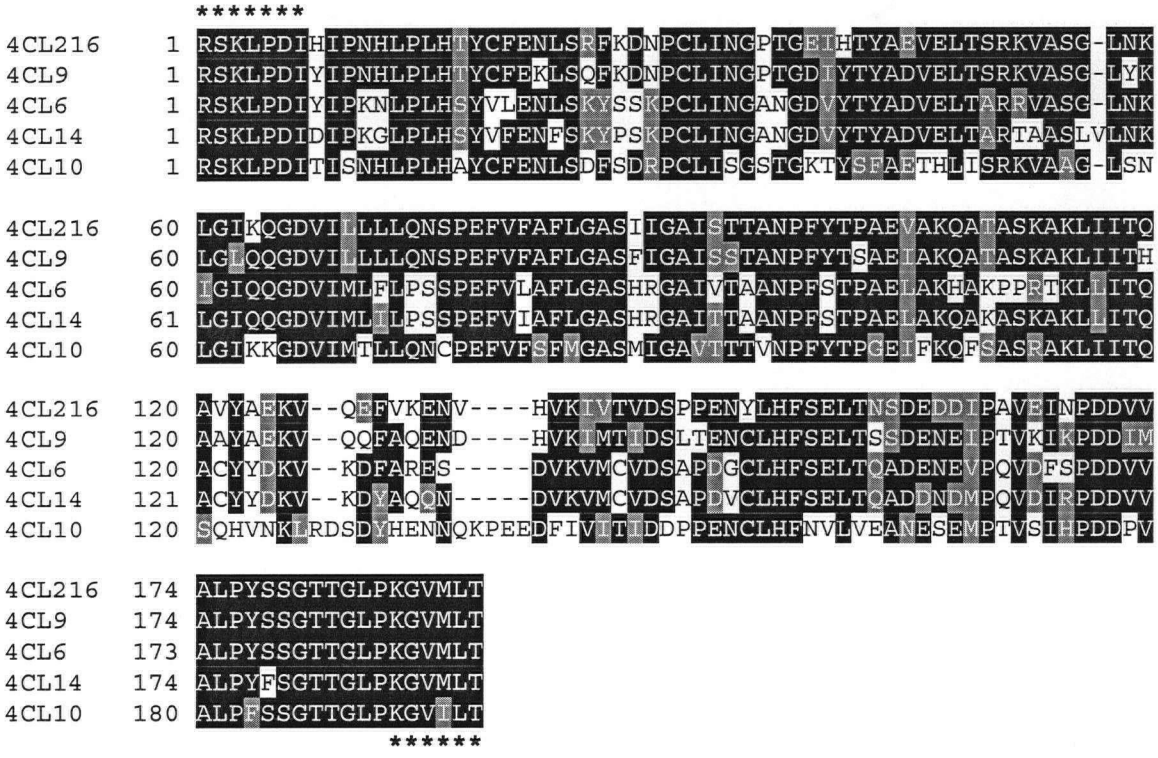


Figure 6. Alignment of deduced amino acid sequences of PCR fragments of poplar *4CL* genes. Asterisks indicate conserved amino acids used for designing degenerate primers.

Table I. Putative 4CL gene fragments amplified by PCR.

Primer set	4CL6	4CL10	4CL14	4CL9	Non specific	Total	% Non-specific
3 x 6	11	1	-	-	7	19	37
3 x 8	-	6	-	10	5	21	24
3 x 2	-	-	8	-	6	14	42
3 x 4	-	-	-	-	18	18	100

The efficiency of different primer sets in amplification of 4CL fragments is summarized in Table I. Thus, primer sets 3+6 and 3+8 successfully amplified two new classes, *4CL6* and *4CL10*, as well as the known *4CL9* class. Primer set 3+2 was biased toward amplification of the new class *4CL14*, while primer set 3+4 amplified only non-specific fragments.

3.1.2 Cloning Full-Length cDNA Clones

A cDNA library from xylem RNA of hybrid poplar clone H11-11was screened using a mixture of the PCR fragments *4CL6* and *4CL10* as a probe. Three positive clones out of approximately 400,000 pfu screened were detected in the first plaque hybridization experiment. After four rounds of plaque purification, the screening resulted in purification of a single 1.88 kb cDNA clone corresponding to the *4CL6* PCR fragment. Neither *4CL10* nor *4CL14* were found in this library. The resulting *4CL6* cDNA was sequenced and found to contain an open reading frame of 1647 bp, which is predicted to encode a protein of 548 amino acids with a molecular weight of 59.2 kDa. There was a 6- bp 5'-untranslated region, a 227 bp 3'-noncoding region, and a poly (A) tail (Figure 7).

```
   1  CCCGCAATGG ACGCCATAAT GAATTCACAA GAAGAATTCA TCTTTCGCTC
  51  AAAATTACCA GACATCTACA TCCCGAAAAA CCTTCCTCTG CATTCATACG
 101  TTCTTGAAAA CTTGTCTAAA TATTCATCAA AACCTTGCCT GATAAATGGC
 151  GCAAACGGAG ATGTCTACAC CTATGCTGAC GTTGAGCTCA CAGCAAGAAG
 201  AGTTGCTTCT GGTCTTAACA AGATTGGTAT TCAACAAGGT GACGTGATCA
 251  TGCTCTTCCT ACCAAGTTCA CCTGAATTCG TGCTTGCTTT CCTAGGCGCT
 301  TCACACAGAG GTGCCATTGT CACCGCTGCC AATCCTTTCT CCACCCCTGC
 351  AGAGCTAGCA AAACATGCCA AGCCTCCAAG AACAAAGCTT TTGATAACAC
 401  AGGCTTGTTA CTACGACAAG GTTAAAGATT TTGCACGAGA AAGTGATGTT
 451  AAGGTCATGT GCGTAGACTC TGCCCCAGAT GGGTGCTTGC ACTTTTCAGA
 501  GCTAACACAG GCTGACGAAA ATGAAGTGCC CCAGGTCGAC TTTAGTCCTG
 551  ATGATGTTGT AGCATTGCCT TATTCATCAG GGACTACAGG GTTACCAAAA
 601  GGGGTCATGC TAACGCACAA AGGGCTAATA ACCAGTGTGG CTCAACAAGT
 651  AGATGGAGAC AATCCTAACC TGTATTTTCA CAGTGAAGAT GTGATTTTGT
 701  GTGTGTTGCC TATGTTCCAT ATCTATGCTC TGAATTCAAT AATGCTTTGT
 751  GGGCTGAGAG TTGGTGCCTC GATTTTGATA ATGCCAAAGT TTGATATTGG
 801  TACTCTGCTG GGATTGATTG AGAAGTACAA GGTATCTATA GCACCAGTTG
 851  TTCCACCTGT GATGTTGGCA ATTGCTAAGT CACCTGATTT TGACAAGCAC
 901  GACTTGTCTT CTTTGAGGAT GATAAAATCT GGAGGGGCTC CATTGGGCAA
 951  GGAACTTGAA GATACTGTCA GAGCTAAGTT TCCTCAGGCC AGACTTGGTC
1001  AGGGATATGG AATGACCGAG GCAGGACCTG TTCTAGCAAT GTGCTTGGCA
1051  TTTGCCAAGG AACCATTTGA CATAAAACCA GGTGCATGTG GGACTGTCGT
1101  CAGGAATGCA GAAATGAAGA TTGTTGACCC AGAAACAGGG GCCTCTCTAC
1151  GGAGGAACCA GCCTGGTGAG ATCTGCATCC GGGGTGATCA GATCATGAAA
1201  GGATATCTTA ATGACCCTGA GGCAACCTCA AGAACAATAG ACAAAGAAGG
1251  ATGGTTGCAC ACAGGCGATA TCGGCTACAT TGATGACGAT GATGAGCTTT
1301  TCATCGTTGA CAGATTGAAG GAATTGATCA AATATAAAGG GTTTCAGGTT
1351  GCTCCTGCTG AACTCGAAGC TTTGTTACTA GCCCATCCAC AGATATCCGA
1401  TGCTGCTGTA GTAGGAATGA AAGATGAGGA TGCAGGAGAA GTTCCTGTTG
1451  CATTTGTAGT GAAATCAGAA AAGTCTCAGG CCACCGAAGA TGAAATTAAG
1501  CAGTATATTT CAAAACAGGT GATATTCTAC AAGAGAATAA AACGAGTTTT
1551  CTTCATTGAA GCAATTCCCA AGGCGCCATC AGGCAAAATC CTTAGGAAGA
1601  ATCTGAGAGA AACGTTGCCA GGCATATAAC TGAAGACGTT ACTGAACATT
1651  TAACCCTCTG TCTTATTTCT TTAATACTTG CGAAAATGCC AATGAATCAT
1701  TGTAGTGTTG AATCAAGCGT GCTTGGAAAA GACACGTTAC CAAACGTTAA
1751  GAACATTACT GTTCTTGTTA TACAAGCTCT TTAATGTTGC TTTTGTACTT
1801  GGGAAAACAT AAGTTCTCCT GTCGCCATAT GGAGTAATTC AATTGAATAT
1851  TTTGGTTTTT TTAAAAAAAA AAAAAAAAAA
```

Figure 7. Nucleotide sequence of the *4CL6* cDNA clone. The sequence includes an open reading frame of 1647 nucleotides, plus untranslated 5' and 3' end sequences. The putative initiation and termination codons are underlined.

Since no *4CL10* clones were identified in the xylem library, a young leaf cDNA library from H11-11 (Subramaniam *et al.*, 1993) was screened with the *4CL10* PCR fragment. Several positive clones were plaque purified, but all were only ~1000-bp in size. Sequencing analysis showed that these clones contained a 4CL-like coding region corresponding to the *4CL10* sequence, but that they lacked the 5' and 3' ends of the cDNA. In order to obtain a cDNA with the full 4CL10 coding region, RACE technology was employed. Gene specific primers (GSPs) were designed based on the sequence of the partial cDNA *4CL10* clones. A pool of adaptor-ligated cDNAs was made from H11 young leaf RNA, and *4CL10* GSPs and adaptor primers were used in 5'RACE and 3'RACE PCR to amplify the missing ends of *4CL10* cDNA (Figure 8). The RACE products were cloned into TA vector pCR2.1 and sequenced. The 3'RACE PCR amplified a 850-bp region including the 3'end of the predicted *4CL10* coding, a 3' noncoding region, and a poly (A) tail. The 5' RACE PCR amplified a 1050-bp region corresponding to the predicted 5' end of the *4CL10* cDNA, but the fragment did not contain a predicted ATG start codon. A full length cDNA clone of the *4CL10* gene was obtained by PCR amplification using the primer-ligated leaf cDNA pool and another set of *4CL10*-specific primers. A 5' primer was designed based on the published sequence of the 5' end of the aspen *Pt4CL2* gene, since that gene is very similar to my 4CL10 sequence. A 3' primer was designed based on the 3' untranslated portion of the 3'RACE fragment of *4CL10* cDNA. The amplification product was cloned and sequenced. This *4CL10* cDNA was 1917 bp long, and contained an open reading frame of 1740 bp, predicted to encode a protein of 579 amino acids (63.1 kDa, predicted molecular weight), including the putative initiation and termination codons (Figure 9).

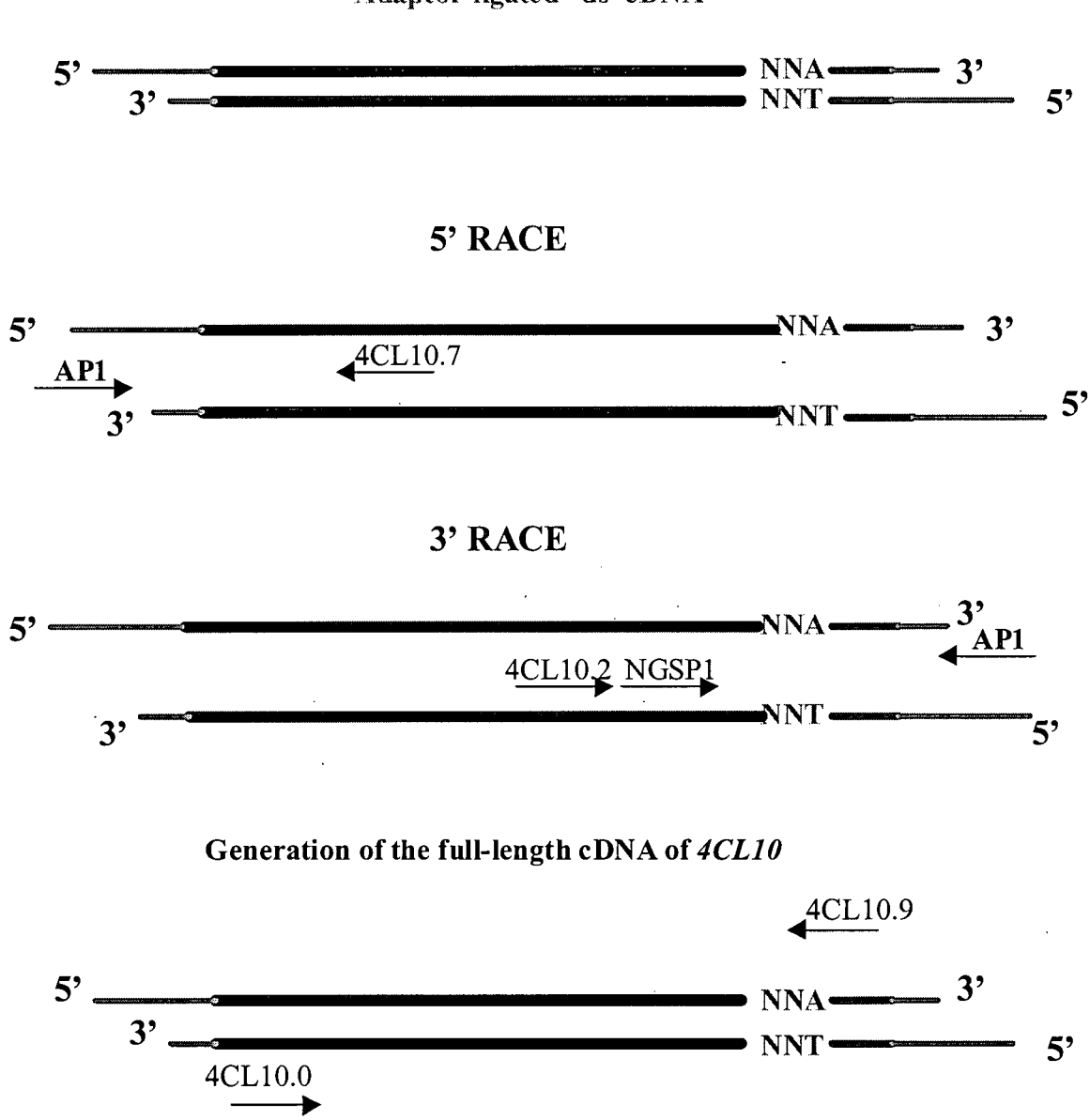


Figure 8. The template and primers used in RACE reactions and amplification of the full-length cDNA. Double-stranded cDNA (ds cDNA) was generated from young leaf poly A+ RNA by reverse transcription and second-strand synthesis, and it was ligated to the adaptor. AP1= adaptor primer 1; gene-specific primers = 4CL10.7, 4CL10.2, 4CL10.0, and 4CL10.9. NGSP1, nested gene- specific primer. The specificity of RACE reactions is greatly enhanced by absence of an AP1 binding site on the adaptor-ligated cDNAs. This site is created on the cDNA of interest by extension from the inner, gene-specific primer during the first RACE cycle. The amine group on the cDNA adaptor blocks extension of the 3'end of the adaptor-ligated ds cDNA, and thus prevents formation of an AP1 binding site on the general population of cDNAs.

```
   1  ATGATGTCCG TGGCCACGGT TGAGCCCCCG AAACCGGAAC TCTCCCCCCC
  51  ACAAAACCAA AACGCACCAT CCTCTCATGA AACTGATCAT ATTTTCAGAT
 101  CAAAACTACC AGATATAACC ATCTCGAACC ACCTCCCTCT GCACGCATAC
 151  TGCTTTGAAA ACCTCTCTGA TTTCTCAGAT AGGCCATGCT TGATTTCAGG
 201  TTCCACGGGA AAAACCTACT CTTTTGCCGA AACTCACCTA ATATCTCGAA
 251  AGGTCGCTGC TGGGTTATCC AATTTGGGCA TCAAGAAAGG CGATGTAATC
 301  ATGACCCTGC TCCAAAACTG CCCAGAATTC GTCTTCTCCT TCATGGGTGC
 351  TTCCATGATT GGTGCAGTCA CCACCACTGT GAACCCTTTC TACACTCCAG
 401  GTGAAATATT CAAGCAATTC TCTGCTTCTC GTGCGAAACT GATTATCACC
 451  CAGTCTCAAC ATGTGAACAA GCTAAGAGAT AGTGATTACC ATGAAAACAA
 501  CCAAAAACCG GAGGAAGATT TCATAGTAAT CACCATTGAT GACCCACCAG
 551  AGAACTGTCT ACATTTCAAT GTGCTTGTTG AGGCTAACGA GAGTGAAATG
 601  CCAACAGTTT CAATCCATCC GGATGATCCT GTGGCATTAC CATTCTCTTC
 651  AGGGACAACA GGGCTCCCAA AAGGAGTGAT ACTGACTCAC AAGAGCTTGA
 701  TAACAAGTGT GGCTCAACAA GTTGATGGAG AGATCCCAAA TTTATACTTG
 751  AAACAAGATG ATGTCGTTTT ATGCGTTTTA CCTTTGTTTC ACATCTTTTC
 801  ATTGAACAGC GTGTTGTTAT GCTCGTTGAG AGCCGGTTCT GCTGTACTTT
 851  TAATGCAAAA GTTTGAGATC GGATCACTGC TAGAGCTCAT TCAGAAACAC
 901  AATGTTTCGG TTGCGGCTGT GGTGCCACCA CTGGTGCTGG CGTTGGCCAA
 951  GAACCCAATG GTGGCGAACT TCGACTTGAG TTCGATCAGG GTAGTCCTCT
1001  CAGGGGCTGC GCCACTGGGG AAGGAGCTCG AGGAGGCCCT CAGGAGCAGG
1051  GTTCCACAGG CCATCCTGGG ACAGGGTTAT GGGATGACAG AGGCGGGGCC
1101  AGTGCTATCA ATGTGCTTAG CCTTCTCAAA GCAACCTTTA CCCACCAAGT
1151  CTGGATCATG TGGAACAGTG GTTAGAAACG CAGAGCTCAA GGTCATTGAC
1201  CCTGAGACCG GTAGCTCTCT TGGTCGCAAC CAACCTGGTG AAATCTGCAT
1251  CCGGGGATCC CAAATCATGA AAGGATATTT GAATGACGCG GAAGCCACGG
1301  CAAACATCAT AGACGTTGAG GGTTGGCTCC ACACTGGAGA TATAGGTTAT
1351  GTCGACGACG ACGACGAGAT TTTCATTGTT GATAGAGTGA AGGAAATCAT
1401  AAAATTCAAA GGCTTCCAGG TGCCGCCAGC GGAGCTTGAG GCTCTCCTTG
1451  TAAACCACCC TTCAATTGCG GATGCGGCTG TTGTTCCCCG AGATAACTTG
1501  TATGGAAACA ACAGGCAAAA AGACGAGGTT GCTGGTGAAG TTCCTGTCGC
1551  GTTTGTGGTC CGCTCAAATG ATCTTGACCT TAATGAAGAG GCTGTAAAAG
1601  ACTACATTGC AAAGCAGGTG GTGTTCTACA AGAAACTGCA CAAGGTGTTC
1651  TTCGTTCATT CTATTCCCAA ATCGGCTTCT GGAAAGATTC TAAGAAAAGA
1701  CCTCAGAGCC AAGCTTGCCA CAGCCACCAC CATGTCCTAG ATTTCATTAC
1751  GTTAAATCTG CATTTATTAT TTTGTGTTGT CTTTCACTCG CTGTGGAAAG
1801  ATTCTAAGAA AAGACCTCAG AGCCAAGCTT GCCACAGCCA CCACCATGCA
1851  TGTCCTAGAA TTCATTCCGT TAAATCTGCA TTTATATTAT TTTGTGTTGT
1901  CTTCACTCAT TGTGCCC
```

Figure 9. Nucleotide sequence of the *4CL10* cDNA clone. The sequence includes an open reading frame of 1740 nucleotides, and an untranslated 3'end sequence. The putative initiation and termination codons are underlined.

3.1.3 Sequence Analysis

The deduced amino acid sequences of the new *4CL* cDNAs, designated *4CL6* and *4CL10*, were compared to each other using the GAP and BESTFIT functions of the GCG sequence analysis software (Figure 10).

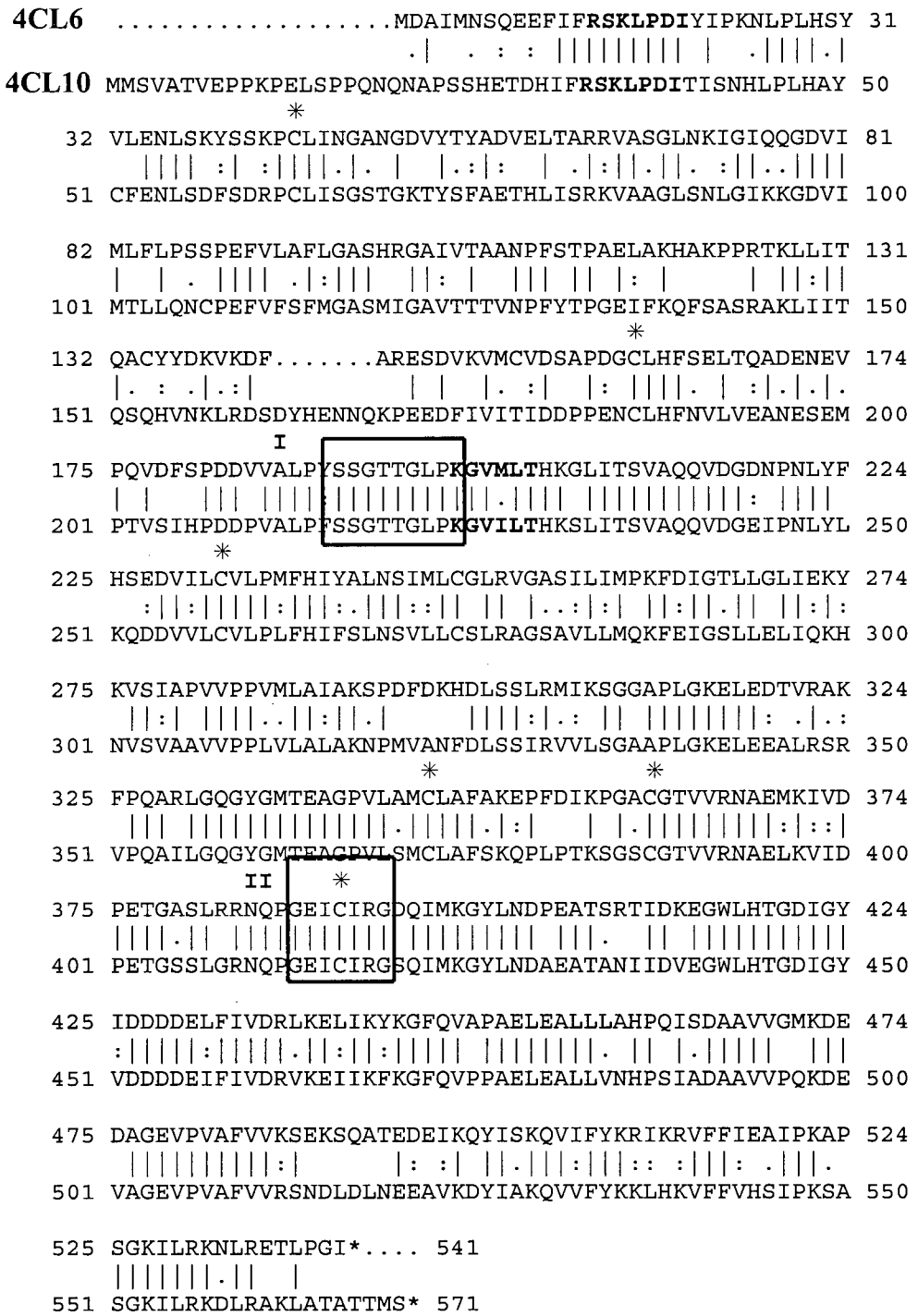


Figure 10. Comparison of the deduced amino acid sequences of the *4CL6* and *4CL10* cDNA clones. Identical residues are denoted by | and similar residues by : Conserved regions **I** (AMP-binding motif) and **II** (putative catalytic site) are boxed. Asterisks indicate conserved cysteine residues. Bold letters indicate conserved amino acids used for designing degenerate primers.

Although the predicted amino acid sequences of 4CL6 and 4CL10 showed only 61 % identity, they have significant similarity throughout their complete lengths. Both proteins contain conserved amino acid motifs at the positions observed in all other 4CL proteins. Motif I (Box I, Figure 10) has been suggested to form part of the AMP-binding domain (Schroder, 1989; Bairoch, 1991), while the Box II amino acid motif GEICIRG (Figure 10) has been proposed to be associated with stability and catalytic activity of 4CL and related enzymes (Becker-André *et al.*, 1991).

Table II Comparison of 4CL6 and 4CL10 predicted amino acid sequences to each other and to other poplar 4CL cDNA sequences.

cDNA	4CL9 % identity	4CL216 % identity	4CL6 % identity	4CL10 % identity
4CL9	*	86	74	64
4CL6	74	74	*	61
4CL10	64	65	61	*

Table II shows a comparison of the predicted amino acid sequences of the four known poplar *4CL* genes (*4CL9* and *4CL216* previously cloned, and *4CL6* and *4CL10* from this study). The most divergent *4CL* gene within the poplar *4CL* gene family is *4CL10*. This gene shows less identity (61% to 65%) to the three other poplar *4CL* genes than to certain *4CL* genes from other plants such as aspen *Pt4CL2* (96%), *Lithospermum 4CL2* (76%) and soybean 4CL2 (78%) (Table III). The *4CL6* gene shows moderate divergence from other poplar 4CL genes (74%), and higher identity to aspen *Pt4CL1* (94%), soybean 4CL1 (80%) and tobacco *Nt4CL1* (78%).

Table III Comparison of 4CL6 and 4CL10 predicted amino acid sequences to each other and to other 4CL sequences.

cDNA	**4CL6** % identity	**4CL10** % identity	**Ref.**
4CL6	*	61	
4CL10	61	*	
poplar 4CL1	74	65	Allina & Douglas (1997)
poplar 4CL2	74	64	Allina & Douglas (1997)
aspen 4CL1	94	60	Hu *et al.* (1998)
aspen 4CL2	60	96	Hu *et al.* (1998)
tobacco Nt4CL19	77	67	Lee & Douglas (1996)
tobacco Nt4CL1	78	65	Lee & Douglas (1996)
Lithospermum 4CL1	73	60	Yazaki *et al.* (1995)
Lithospermum 4CL2	60	76	Yazaki *et al.* (1995)
Arabidopsis 4CL1	70	60	Lee *et al.* (1995)
soybean 4CL1	80	64	Uhlmann & Ebel (1993)
soybean 4CL2	66	78	Uhlmann & Ebel (1993)

Alignment of deduced amino acid sequences of the cDNA clones for the four poplar 4CL gene (Figure 11) revealed that the most pronounced similarity between all four predicted amino acid sequences occurs within the central and C-terminal part of the proteins. Based on the sequence alignments, unique sequence features of the predicted 4CL10 relative to the other 4CL proteins are a 19 amino-acid N-terminal extension and two insertions of 4-5 amino acids each near the N-terminus. Similar N-terminal extensions and insertions are present in *Lithospermum 4CL2*, aspen *Pt4CL2* and rice *4CL2*. The BLAST search for proteins with significant similarity to 4CL at the amino acid sequence level, identified several non-4CL proteins, including luciferin-4-monooxygenase and long-chain-fatty-acid-CoA ligase. These proteins share many conserved amino acids throughout their entire length with 4CL6 and 4CL10, particularly in the AMP-binding and the GEICIRG motifs (Figure 12).

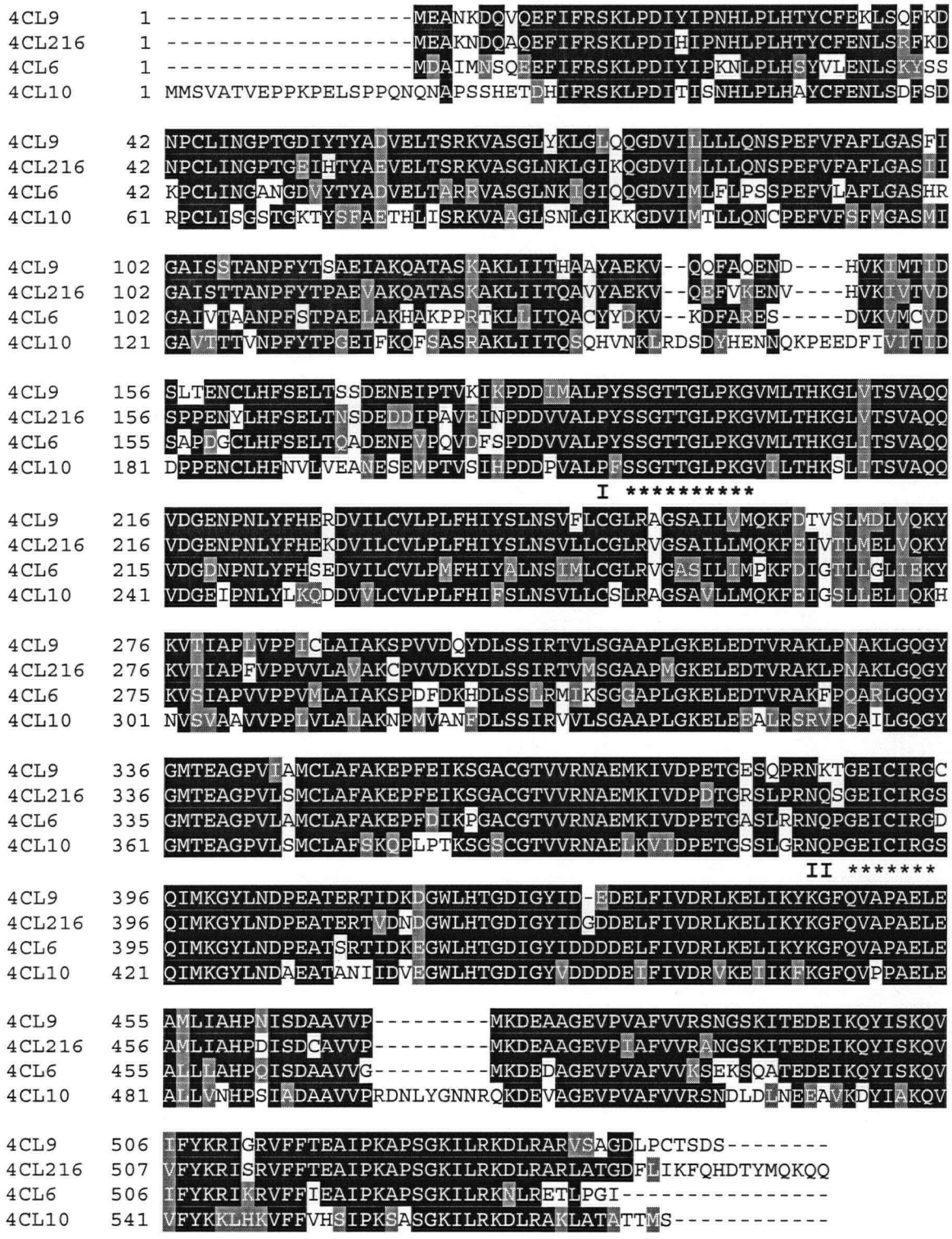


Figure 11. Alignment of deduced amino acid sequences of poplar *4CL* cDNA clones. Black shade indicates identical amino acids; grey shade indicates similar amino acids. Conserved regions I (AMP-binding motif) and II (putative catalytic site) are indicates by asterisks.

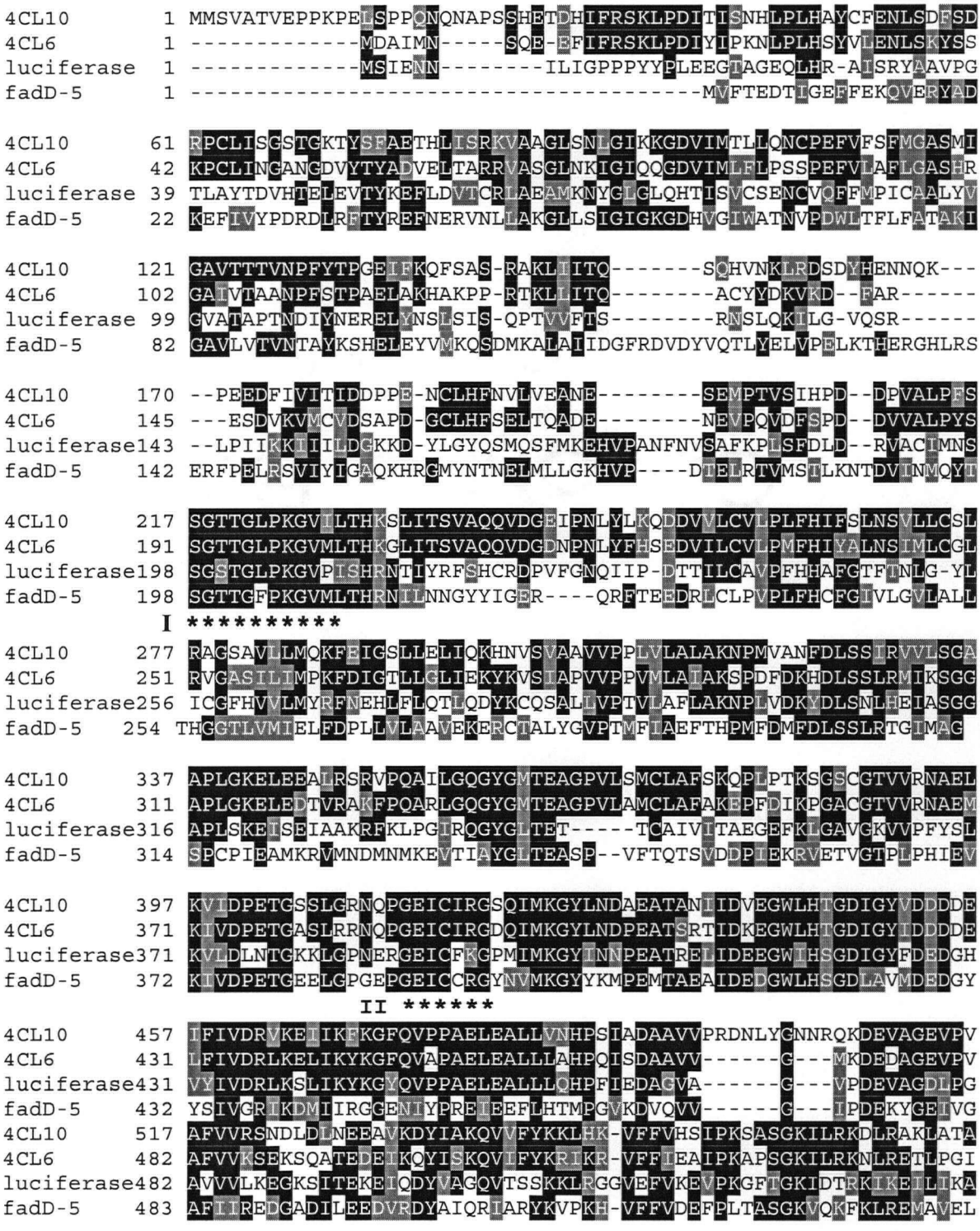


Figure 12. Alignment of deduced amino acid sequences of poplar *4CL6* and *4CL10* cDNA clones and other ATP-dependant enzymes. Luciferase= luciferin-4-monooxygenase from *Photuris pennsylvanica* (D25416.1); fadD-5= long-chain-fatty-acid-CoA ligase (AE001034). Conserved regions I (AMP-binding motif) and II (putative catalytic site) are indicated by asterisks. Black shade indicates identical amino acids; grey shade indicates similar amino acids.

3.2 Expression of Divergent *4CL* Genes

To study the regulation of *4CL6* and *4CL10* gene expression, total RNA from different tissues of poplar clone H11-11 (old leaf, young leaf, green stem, xylem, elicitor-treated and untreated poplar suspension–culture cells) was blotted and hybridized with probes prepared from the *4CL6* or *4CL10* cDNA clones. The level of *4CL6* mRNA accumulation was high in xylem, green stem and root, and quite low in old leaf (Figure 13A). No expression of *4CL6* was detected in young leaf, or in control or elicitor-treated suspension cultures.

A very different expression pattern was observed for *4CL10* (Figure 13B). *4CL10* transcripts were detected in old and young leaves as well as in root tissue, but not in green stem and xylem. The blot hybridized with the *4CL10* probe was exposed twice as long as the blot hybridized with the *4CL6* probe. Thus, the level of *4CL10* mRNA accumulation in old leaf and root appears to be lower than that of *4CL6,* and the overall expression level of *4CL10* is quite low.

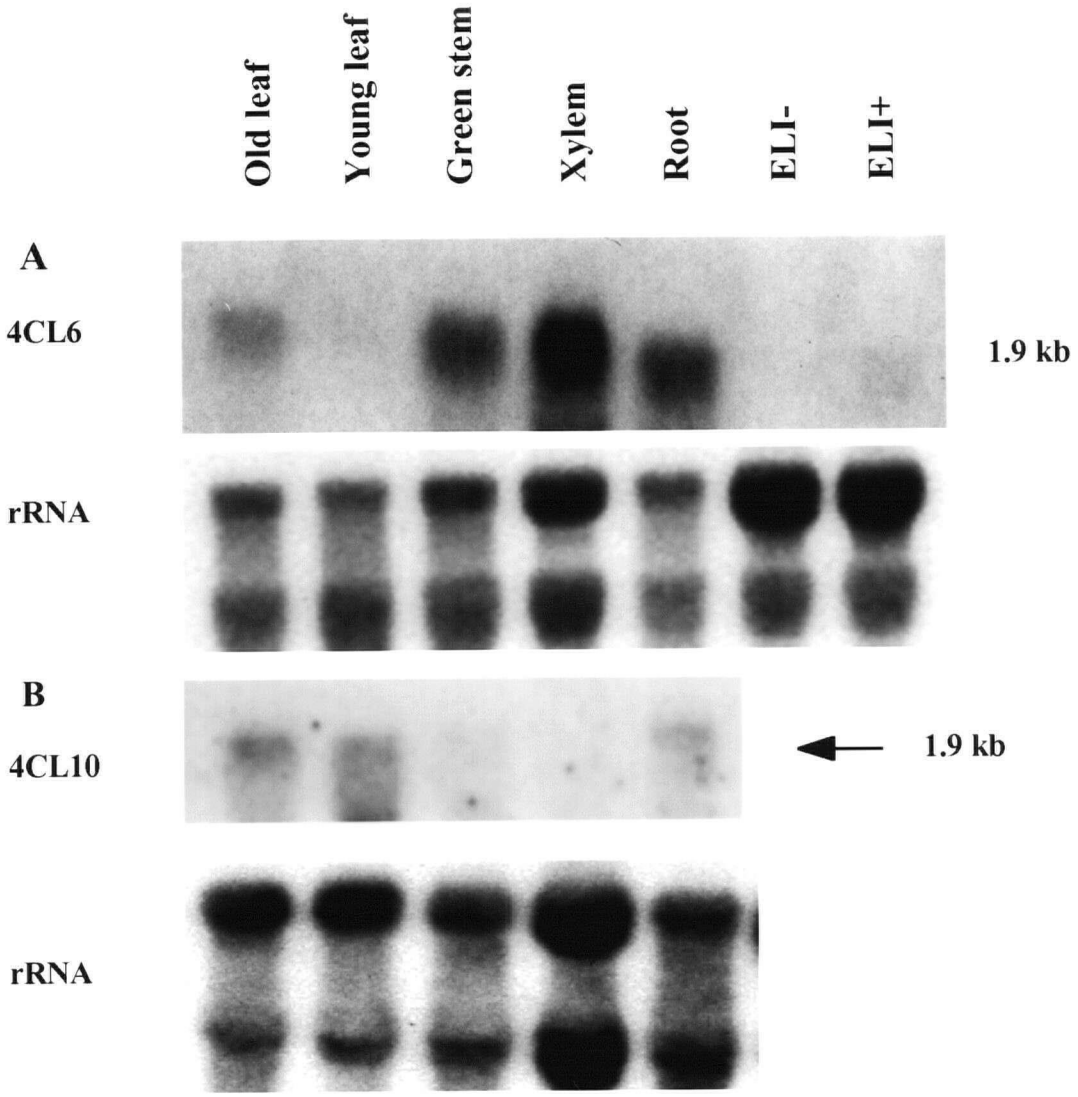


Figure 13. Northern-blot analysis of *4CL6* and *4CL10* mRNA levels. Total RNA (10 μ g) From different tissues and organs was on duplicated formaldehyde agarose gels, blotted to nylon membranes, hybridized to either *4CL6* or *4CL10* cDNA probes, and washed at high stringency (0.2XSSC, 0.1%SDS, 65°C). Even loading was shown by stripping blots and rehybridizing with a pea rRNA probe. EL+, elicitor-treated tissue culture cells: ELI-, untreated control cells.

3.3 Characterization of Recombinant 4CL Proteins

3.3.1 Expression and Purification of Recombinant 4CL6 and 4CL10 Proteins

To test whether the novel *4CL* genes from poplar actually encode 4CL enzymes and whether the enzymes have catalytic properties distinct from previously characterized poplar 4CL enzymes (Allina *et al.*, 1998), the *4CL6* and *4CL10* cDNAs were expressed as recombinant proteins using a baculovirus vector expression system. Recombinant baculovirus particles expressing 4CL6 were generated as a result of homologous recombination between pVL1392::*4CL6* and AcNPV viral DNA in Sf9 insect cells. Identification and purification of the recombinant 4CL6 baculovirus particles were done as described in Materials and Methods, while a similar procedure was used to transfer *4CL10* from pVL1392::*4CL10* into AcNPV viral DNA in the lab of David Theilman, Pacific Agri-Food Research Center, Summerland, BC.

To generate the recombinant 4CL6 protein, 50 x 10^6 Sf9 cells were infected with recombinant 4CL6 baculovirus particles from a high titer stock for 36 hours. The 4CL6 protein was partially purified from the crude protein extract by fast-protein liquid chromatography using an ion-exchange High-Q column. The recombinant 4CL6 protein eluted as a single peak of 4CL activity, assayed using 4-coumaric acid as a substrate (Figure 14). The most active fractions of 4CL6 had specific activities between 498 and 597 pkat/μg protein.

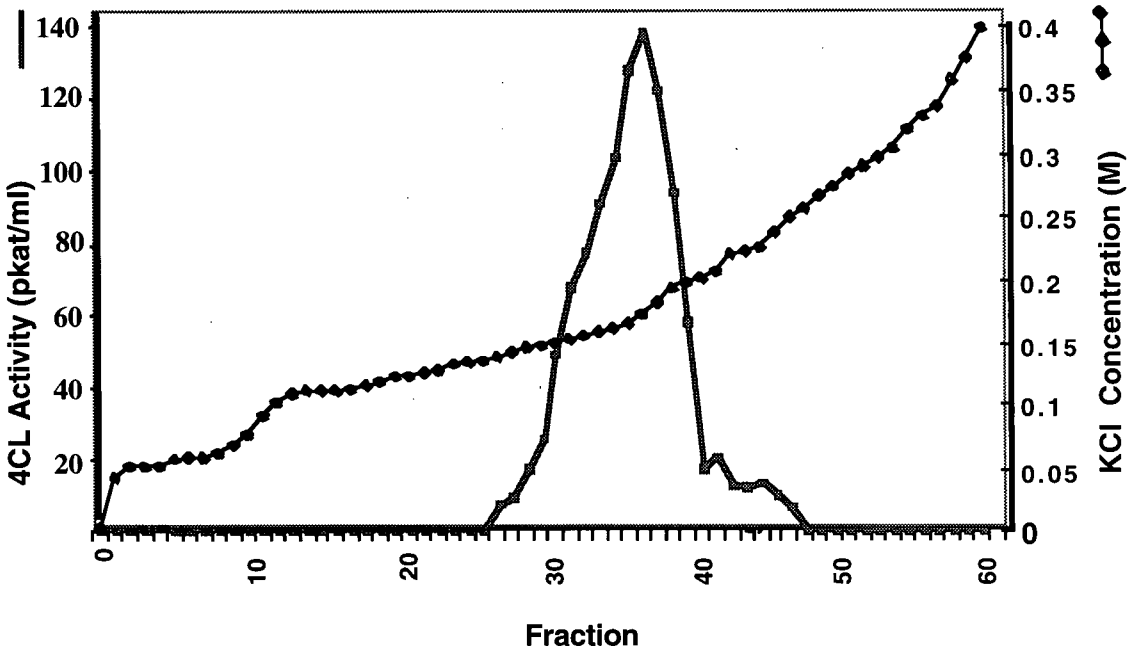


Figure 14. Partial purification of recombinant protein 4CL6 by FPLC on a High-Q anion exchange column using KCl gradient. 4CL activity was assayed using 4-coumarate as a substrate.

To assess the purity of FPLC-purified 4CL, two peak FPLC fractions were separated by SDS-PAGE, in parallel with crude extracts from uninfected Sf9 and infected Sf9 cells. This analysis showed that FPLC resulted in an efficient enrichment of the presumed recombinant 4CL6 protein, whose migration corresponded to the predicted molecular mass of 60 kDa (Figure 15A). A parallel immunoblot showed that poplar 4CL-specific polyclonal antibodies (arised against recombinant 4CL9, Allina *et al.*, 1998) cross-reacted with the recombinant 4CL6 (Figure 15B), confirming the identity of the recombinant protein as 4CL.

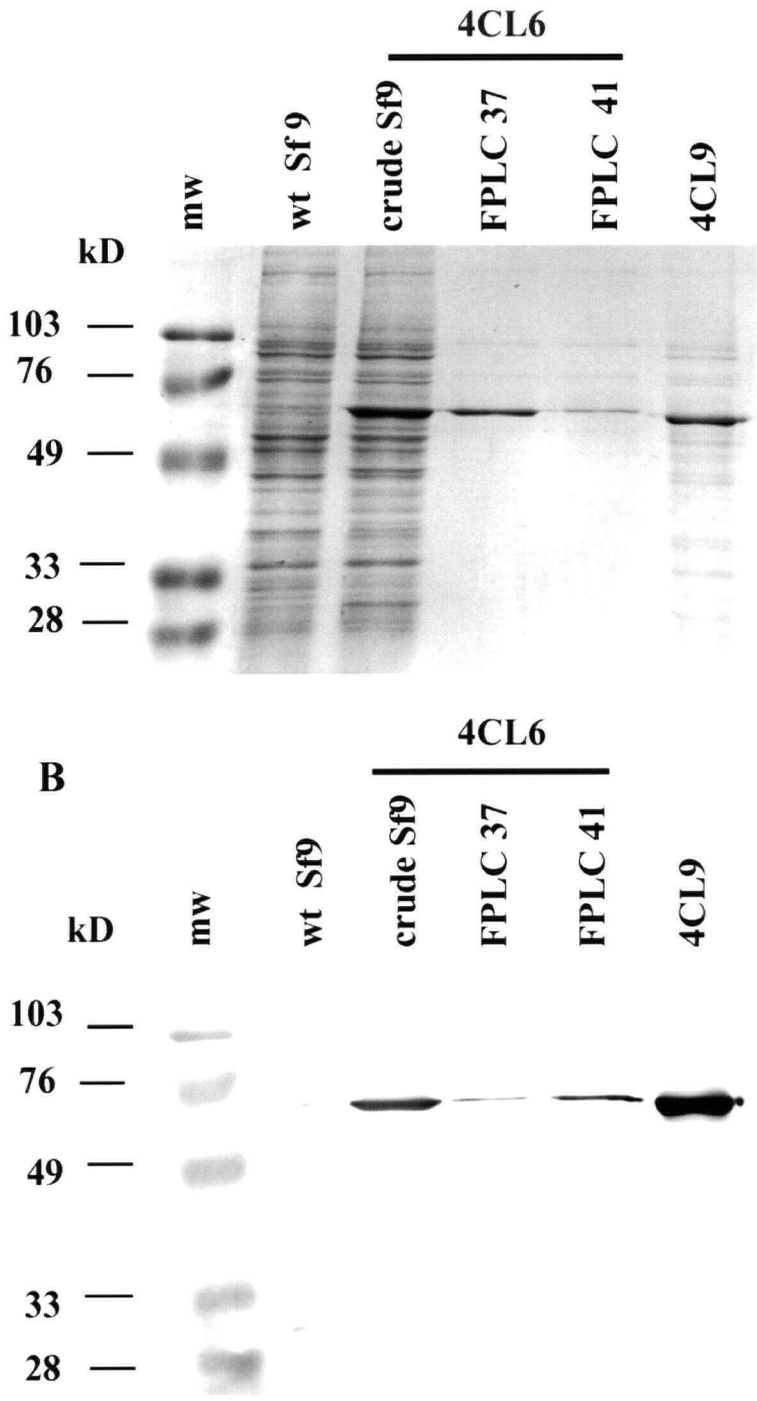


Figure 15. SDS-PAGE and immunoblot analysis of recombinant 4CL6 protein. SDS-PAGE gel (A) and an immunoblot of a parallel gel (B) reacted with antiserum specific to recombinant 4CL9 protein. mw, molecular weight standard, wt Sf9, uninfected insect cells, crude extracts of Sf9 cells infected with 4CL6 and 4CL9 baculovirus constructs, 4CL6 FPLC-purified from Sf9 infected cells.

As shown on Figure 16, eight 4CL10-expressing baculovirus independent stocks, express recombinant proteins of the size predicted for 4CL10 protein. The protein has a higher molecular weight than the recombinant 4CL6 and 4CL9 proteins, as expected from the predicted amino acid sequence. Crude extracts of Sf9 cells infected with the 4CL10 baculovirus stocks showed 5-10 fold less 4CL activity toward 4-coumaric acid than crude extract of Sf9 infected with 4CL6 baculovirus particles.

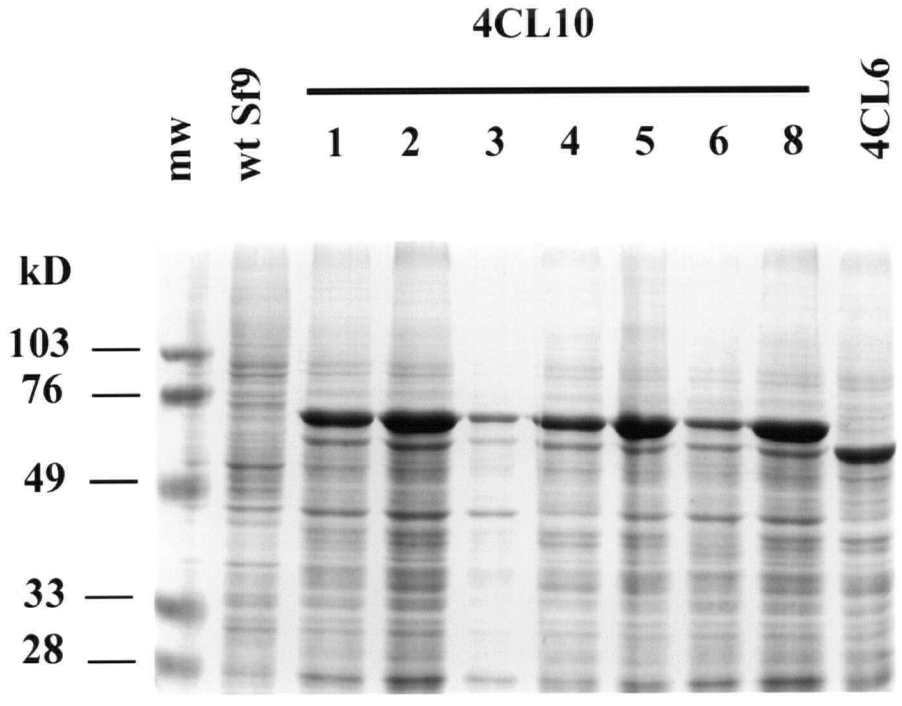


Figure 16. SDS-PAGE analysis of recombinant 4CL10 protein. SDS-PAGE gel stained in Coomassie brilliant blue. MW, molecular weight standard, wt Sf9, uninfected insect cells, crude extracts of Sf9 cells infected with 4CL10 (1-6 & 8 plaques) and 4CL6 baculovirus constructs.

3.3.2 Substrate Utilization Profile of Recombinant 4CL6 Protein

The ability to use different hydroxycinnamic acids as substrates was tested for the 4CL6 protein. Enzyme assays were carried out using 0.2 mM concentrations of each substrate. Figure 17 shows that the partially purified recombinant 4CL6 protein showed a strong preference for 4-coumaric acid and decreasing activities toward ferulic acid, caffeic acid, and cinnamic acid. No activity toward sinapic acid was detected. Activity toward 5-hydroxyferulic acid was not tested because 5-hydroxyferulic acid was available in insufficient amount.

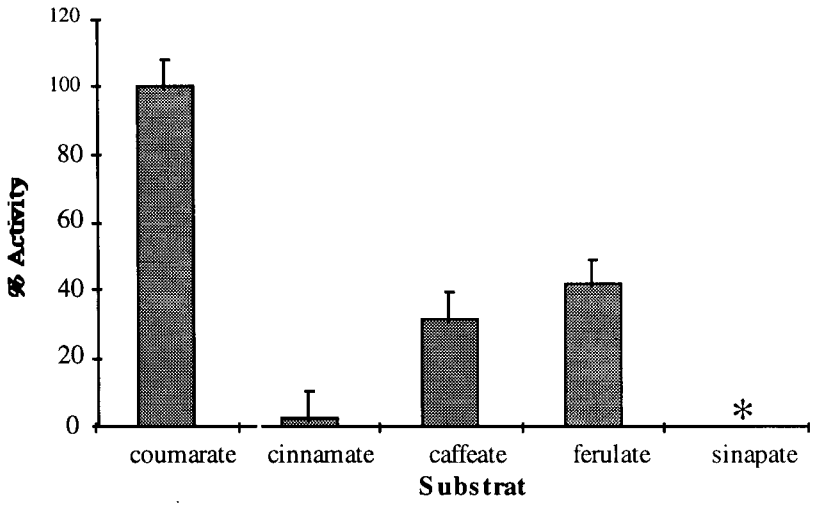


Figure 17. Substrate-utilization profile of recombinant 4CL6. Recombinant 4CL6 enzyme activity was measured using FPLC-purified 4CL6 and 0.2 mM concentrations of hydroxycinnamic acids. Results are averages of three trials: error bars represent SD values. Asterisks indicate the absence of detectable activity. Results are reported as a precentage of the activity against 4-coumaric acid, which was 480.7 pkat/mg.

CHAPTER IV

Discussion

The results of this work confirm to the existence of at least four 4CL isoenzymes in poplar. This conclusion is based on isolation of two full-size *4CL* cDNA clones in addition to those previously described (Allina *et al.*, 1998), and is consistent with multiple 4CL proteins detected by FPLC chromatography (Allina *et al.*, 1998).

The use of degenerate PCR primers that target conserved sequences in *4CL* genes allowed the efficient identification of divergent *4CL* genes. Thus, while previous screening of a poplar young leaf cDNA library resulted in identification of two *4CL* cDNA clones with 86% identity at the predicted amino acid level (Allina *et al.*, 1998), my work showed that the poplar *4CL* gene family in fact contains three new *4CL* classes (*4CL6, 4CL10* and *4CL14*) in addition to the previously studied genes. These were distinguished among 72 clones characterized from the products of PCR amplification of poplar genomic DNA. PCR techniques have been successfully applied to isolate a divergent *Pt4CL2* gene from aspen (Hu *et al.*, 1998). Degenerate PCR primers that target evolutionarily conserved sequences in *PAL* genes also revealed the presence of at least eight *pal* loci in the jack pine genome (Butland *et al.*, 1998). In contrast, the previous study on the *PAL* gene family in loblolly pine had reported that this pine genome contains a single *PAL* gene (Whetten and Sederoff, 1992). Thus, it is clear that this PCR strategy can potentially identify phenylpropanoid gene family members that might be difficult to detect by hybridization.

The degenerate primers used in this work performed with different efficiency in amplifying *4CL* gene fragments from *P. trichocarpa*. Thus, primer sets 3+6 and 3+8

amplified fragments of three poplar *4CL* genes, *4CL6*, *4CL10* and *4CL9* (Table I). Relatively low frequency of nonspecific sequences amplified by these primer sets (24% and 37%) was consistent with results obtained in other studies with PCR-based cloning of different genes. For example, Sells and Chernoff (1995) found that each set of degenerate primers specific for tyrosine phosphatase genes, amplified about10 nonspecific sequences out of 30 characterized sequences. In contrast, the 4CL primer set 3+2 amplified only the *4CL14* class. Similar results with degenerate primers were found by Butland *et al.* (1998), where PAL primers P7-P6 amplified only *pal* 5gene, while P5-P2 primers amplified four *pal* sequence classes. 4CL primer set 3+ 4 amplified only nonspecific sequences, suggesting that reverse primer 4 annealed inappropriately due to self-annealing or formation of hairpin structures within the oligonucleotide pool. Moreover, the low efficiency of reverse primers 2 and 4 could be related to the high degeneracy of their sequences, necessary because of the alternative conserved amino acids found between different *4CL* genes in this target region. Reverse primers 2 and 4 are also longer (20+8 nt) than the reverse primers 6 and 8 (17+8 nt). Wilks *et al.* (1989) reported that long degenerate primers (20-30 nt) are less efficient than short (14-20 nt) degenerate primers. Therefore, this work showed that the consensus amino acids RSKLPDI and KGVMLT are a better choice for designing degenerate primers specific for *4CL* than the region with the alternating amino acids P(L/M)FHI(Y/F)(S/A).

The possibility that additional poplar *4CL* genes remain to be isolated appears small due to the exhaustive PCR-based search of poplar genomic DNA and the screening of two cDNA libraries. Furthermore, a poplar database of expressed sequence tags (ESTs) from cambium and xylem (http://www.biochem.kth.se/PopulusDB, Sterky *et al.*, 1998) revealed three ESTs from cambium, which show 90-96% identity to the *4CL6* cDNA clone.

The three ESTs represent different parts of the same gene, which supports the result of this work that 4CL6 is the major 4CL isoform expressed in the wood-forming tissues.

Neither of the degenerate primer sets amplified any sequence that corresponded to the *4CL216* gene. A comparison of the PCR primer sequences with the actual nucleotide sequences of the four poplar *4CL* genes indicates that mispairing between the primer 3 and *4CL216* could explain the absence of *4CL216* fragments after PCR amplification (Figure 18). This indicates that the PCR primers used in this work were biased toward certain *4CL* genes, making it possible that some *4CL* genes remain to be isolated.

A

		# of mismatches
4CL9	3'**T**CC AGG TTT **GAG** GGA CTG TAG 5'	3
Primer 3	5'**C**GG TCI AAA **TTG** CCI GAC AT 3'	
4CL216	3'**T**CC AGG TTT **GAG** GGA CT**A** TAG 5'	4
4CL6	3'GCG AGT TTT AAT GGT CTG TAG 5'	0
Primer 3	5'CGG TCI AAA TTA CCI GAC AT 3'	
4CL10	3'**T**CT AGT TTT **GAT** GGT CTA TAT 5'	2
Conserved amino acids	R S K L P D I	

B

4CL9	5'AAA GGT GT**C** ATG TTG ACT C 3'	2
Primer 8	3'TTT CCI CA**C**(A)TAC AAA TGA G 5'	
4CL216	5'AAA GGT GT**C** ATG TTG ACT C 3'	2
4CL6	5'AAA GGG GT**C** ATG CTA AC**G** C 3'	2
Primer 6	3'TTT CCI CAC(A)TAC GAI TGA G 5'	
4CL10	5'AAA GGT GTG ATA CTG ACT C 3'	1
Conserved amino acids	K G V I L T	

Figure 18. Degenerate primers and corresponding sequences in poplar 4CL genes. A, comparison of the forward primer 3 with the actual nucleotide sequences; B, Comparison of the reverse primers 6 and 8 with the actual nucleotide sequences of poplar *4CL* genes; bold letters indicate mismatches.

All four poplar 4CL cDNAs encode predicted polypeptides with a highly conserved region designated as a putative AMP-binding motif (region I in Figure 11), a seven-amino-

acid motif containing one cysteine residue (region II in Figure 11) representing a putative catalytic site, and six cysteine residues, suggesting a conserved tertiary structure. These 4CL landmarks have been found in deduced amino acid sequences of all other 4CL predicted products such as those from parsley, potato, soybean, *Lithospermum,* tobacco and aspen (Lozoya *et al.*, 1988; Backer-André *et al.*, 1991; Uhlmann and Ebel, 1993; Yazaki *et al.*, 1995; Lee and Douglas, 1996; Hu *et al.* 1998). Similarity between the predicted amino acid sequences of poplar 4CL proteins is especially high in the central and C-terminal parts of the proteins. In these regions there are long stretches of essentially identical amino acid residues. These amino acid residues are likely important for enzyme structure and function. However, to date there is no direct experimental evidence to support the hypothetical function assignments of motifs I and II. Site-directed mutagenesis of the cysteine residue in the region II (potato 4CL), carried out in order to test the hypothesis that this cysteine residue is associated with catalytic activity, failed because of instability of the mutated protein in *E. coli* cells (Becker-Anderé *et al.*, 1991). In addition, the lowest homology among poplar 4CLs is detected in the N-terminal sequences. Therefore the N-terminal region might be involved in phenolic substrate binding specificity (Hu *et al.*, 1998).

Different luciferin-4-monooxygenase (firefly luciferase) and long-chain-fatty-acid-CoA ligases genes showed similarity to *4CL6*, *4CL10* and other 4CLs (Figure 12). These proteins share many conserved amino acids throughout their entire lengths, particularly in the AMP-binding (I) and the GEICIRG (II) motifs (overall 34-36% identity). Although luciferase, long-chain-fatty-acid-CoA ligase, and 4CL catalyse different metabolic reactions, all three enzymes are ATP-dependent, catalyze a two-step reaction which includes an adenylate intermediate compound, and have a hydrophobic substrate (luciferin, a fatty acid,

a hydroxycinnamic acid) and a polar cosubstrate (O_2 for luciferase, CoASH for both ligases) (de Wet, 1986; Gross, 1985). The low similarity between these enzymes in their N-terminal regions could be due to their different substrate specificities, while more conserved amino acids in the central and C-terminal parts are probably involved in the conserved function of ATP binding and hydrolysis. It is note worthy that Gross and Zenk (1966) reported that a fairly non-specific acyl-CoA ligase from beef liver mitochondria was able to activate cinnamic acids at low efficiency.

Comparison of deduced amino acid sequences of 4CL6 and 4CL10 with aspen 4CLs showed that aspen Pt4CL1 and poplar 4CL6 proteins are likely the products of orthologues genes (94% amino acid identity), as are the aspen Pt4CL2 and poplar 4CL10 proteins (96% identity). These two pairs of orthologues genes are very distantly related based on an analysis of phylogenetic relationships of *4CL* genes (Ehlting *et al.*, 1999). Thus, *4CL6* and aspen *Pt4CL1* belong to the class I, while the *4CL10* and *Pt4CL2* orthologous belong to a distinct class II. 4CL isoforms encoded by genes isolated from parsley, tobacco, potato and poplar (*4CL9* and *4CL216*) also belong to class I, and they are more closely related to each other than to 4CL isoforms in class II.

The deduced amino acid sequence of *4CL10* is 61-65% identical to the other three poplar 4CLs. This sequence shows more identity to *Lithospermum 4CL2*, soybean *4CL2*, *Arabidopsis 4CL3* and aspen *Pt4CL2* (78-96%, Table II), all of class II, then to the other poplar *4CL* genes. The N-terminal part of poplar *4CL10* and other class II 4CL proteins differ from other dicot 4CLs in having a terminal extension of 19-24 amino acids and in a two short insertions. Furthermore, genomic clones showed that soybean *4CL2* and *Arabidopsis 4CL3* genes (class II genes) have 6 introns, in contrast to genomic clones of

4CLs from class I (Lozoya *et al.*, 1988; Beecker-André *et al.*, 1991). Hu *et al.* (1998) reported that the promoter region of aspen *Pt4CL2* has none of the boxes consistently found in all known plant PAL and 4CL gene promoters.

Future work on *4CL* gene families from different plants will clarify whether other plants like parsley, potato and tobacco also have 4CL isoforms from the class II. One likely possibility is that divergent class II isoforms exist in these plants, but that they remain to be isolated, perhaps due to their divergent sequences (Ehlting *et al.*, 1999).

Clear differences in expression patterns have been described that distinguish class I and class II 4CL isoforms in aspen, *Arabidopsis*, and soybean (Hu *et al.*, 1998; Ehlting *et al.*, 1999; Uhlmann and Ebel, 1993). My work showed that *4CL6* and *4CL10* also have distinct expression patterns, and that these patterns differ from those of the previously isolated *4CL9* and *4CL216* genes (Allina *et al.*, 1998). *4CL6* is strongly expressed in xylem, green stem and root, but the expression was not detected in young leaves, while *4CL10* is expressed only in leaves and root. In contrast, *4CL9* mRNA is most abundant in young leaf, and *4CL216* is preferentially expressed in old leaf (Allina *et al.*, 1998).

Based on these results, the function of the 4CL6 isoform (class I) appears to be related to lignin biosynthesis, and the 4CL10 isoform (class II) appears to play role in the biosynthesis of nonlignin-related phenylpropanoids (e.g. flavonoids and phenylpropanoid esters), that accumulate in poplar leaves. A similar difference in the expression patterns of aspen *Pt4CL1* and *Pt4CL2* has been observed (Hu *et al*, 1998.). Also, heterologous *Pt 4CL1* and *Pt4CL2* promoter-GUS fusion analyze indicated that *Pt 4CL1* (the orthologues to *4CL6*) gene expression is xylem-specific and *Pt4CL2* (the orthologues to *4CL10)* is epidermis-specific (Hu *et al.*, 1998). In *Arabidopsis,* the most divergent (class II) gene, *4CL3,* was

expressed to relatively high levels in flowers, but not in lignified tissue such as bolting stem, while the *4CL1* and *4CL2* (class I) are specific for bolting stem and seedling roots. Similarly, the class II *4CL* isoform from soybean (*4CL16*) is specifically induced by pathogen infection, after which soybean isoflavanoid phytoalexins accumulate (Uhlmann and Ebel, 1993). This led Ehlting *et al.*, (1999) to propose that a primary function of the class II 4CL isoforms is to channel activated 4-coumarate to chalcone synthase and subsequently to different branch pathways of flavonoid secondary metabolism. On the other hand, a role in lignin biosynthesis has been postulated for the class I 4CL isoforms, based on the high expression level in lignifying tissues of 4CL genes from parsley, tobacco, aspen *Pt4CL1* and poplar *4CL6,* described in this work (Hauffe *at al.*, 1991; Lee and Douglas, 1996; Hu *et al.*, 1998).

Neither the previously studied genes nor *4CL6* was activated in response to elicitor treatment, although the same RNA samples hybridized to a poplar *PAL* probe showed that expression of this gene was strongly stimulated by the elicitor treatment (data not shown). Earlier studies using a heterologous *4CL* probe indicated that *4CL* expression is elicitor-activated in hybrid poplar suspension-cultured cells, followed by a 10- to 20- fold increase in extractable PAL and 4CL enzyme activities (Moniz de Sá *et al.*, 1992). Furthermore, accumulation of a single 4CL isoform in the elicited suspension culture was detected by partial purification with FPLC (Allina *et al.*, 1998). Therefore, a good candidate for an elicitor-induced poplar *4CL* gene is *4CL10*, and its expression in elicitor-treated suspension-cultured cells should be tested.

To further investigate the regulation of tissue-specific and cell-specific expression of poplar *4CL* genes, *in situ* hybridization could be done using leaf and root sections. That

approach could show cell-type specific differences in the expression patterns. For example, *4CL6* and *4CL10* are expressed in root, but their mRNAs might be localised in different tissues of root (epidermis vs. vascular tissue). Also, RFLP analysis of RT-PCR products from different tissues would give better insight into tissue-specific expression of the poplar 4CL genes, and clarify differences between *4CL216* and *4CL9* expression patterns, since they have the highest sequence identity and some cross-hybridization between them was reported (Allina *et al.*, 1998).

The baculovirus-expressed protein encoded by *4CL6* shows a strong preference for 4-coumaric acid as a substrate and decreasing activities toward ferulic acid, caffeic acid and cinnamic acid. This is consistent with previous results (Allina *et al.* 1998) indicating that all native isoforms of hybrid poplar had similar substrate utilization profiles, similar to those of recombinant proteins 4CL9 and 4CL216 proteins. Thus, the increase in divergence between the poplar 4CL deduced amino acid sequences (75% identity between 4CL6 and 4CL9; 86% identity between 4CL9 and 4CL216) does not result in a change in the substrate specificities of these isoforms. Since analysis of the substrate specificity of recombinant 4CL10 protein and the kinetic analysis of both recombinant proteins (4CL6 and 4CL10) have not yet been completed, it remains a possibility that 4CL10 could have a distinct substrate-utilization profile. In addition, Hu *et al.* (1998) reported that the major differences between aspen recombinant Pt4CL1 (orthologues to 4CL6) and Pt4CL2 (orthologues to 4CL10) were that Pt4CL2 is not active toward 5-hydroxyferulic acid, and that the amount of total specific activity of Pt4CL2 is strikingly lower than that of Pt4CL1.

To date, no 4CL proteins have been found to accept substrates other than cinnamic acids. This fact, together with their pronounced preference for hydroxylated cinnamic acids,

has led to proposing the systematic name "hydroxycinnamate:CoA ligase (AMP)" (Gross and Zenk, 1974). Alternatively, the name for this enzyme can refer to the best substrate. Since, 4-coumaric acid is the best substrate for the recombinant 4CL9, 4CL216 and 4CL6, the name "4-coumarate:CoA ligase" is appropriate for these isoforms.

The lack of any detectable conversion of sinapic acid by recombinant 4CL6 is consistent with results obtained both with poplar native 4CLs and recombinant 4CL9 and 4CL216 (Allina *et al.*, 1998). Moreover, none of the recombinant 4CL enzymes from tobacco, soybean or *Arabidopsis* showed activity toward sinapic acid. The occurrence of sinapic acid:CoA ligase has been reported only for some native 4CL isoforms from soybean, *Petunia* and poplar (Knobloch and Hahlbrock, 1975; Ranjeva *et al.*, 1976; Grand *et al.*, 1983). Therefore, a number of studies indicate that sinapic acid cannot be efficiently activated to the CoA thioester by most 4CLs, and that the biosynthesis of sinapyl alcohol and syringyl lignin may occur via an alternative pathway. Thus, Higuchi (1985) postulated a biosynthetic pathway of sinapyl alcohol through 5-hydroxy-coniferaldehyde, followed by methylation by an OMT (*O*-methyltransferse) and reduction by a CAD. Also, Ye and Varner (1995) suggested that 5-hydroxyferuloyl-CoA might be methylated by CCoAOMT (caffeoyl-coenzyme A-3-*O*-methyltransferase) to produce sinapyl-CoA. The genes encoding COMT and CCoAOMT in aspen have been cloned (Bugos *et al.*, 1991; Meng and Campbell, 1995). Meng and Campbell (1998) found that aspen recombinant COMT utilizes free hydroxycinnamic acid, while aspen CCoOMT has activity toward CoA ester substrates with a preference for caffeoyl-CoA. These authors hypothesised that CCoAOMT is likely to be responsible for biosynthesis of lignin precursors in the guaiacyl pathway, while COMT is more likely to be involved in the syringyl pathway. In support of this, COMT-down -

regulated poplar trees displayed unchanged lignin content, but the proportion of G units and resistant biphenyl structures was dramatically enhanced (Lapierre *et al.* 1999). Because down-regulation of 4CL activity in *Arabidopsis* leads to a decrease in G residues, but no change in S residues, Lee *et al.* (1997) proposed that S-lignin might be synthesized by a 4CL-independent pathway.

The absence of activity toward sinapic acid and absence of catalytically distinct 4CL isoforms in poplar and other plants (Lozoya *et al.*, 1988; Voo *et al.*, 1995; Lee and Douglas, 1996), makes it unlikely that 4CL controls partitioning of carbon into guiacyl and syringyl lignin, or into other phenylpropanoid end products. However, 4CL is an end point enzyme between general phenylpropanoid metabolism and branching pathways, and could have an impact on carbon flow from phenylalanine amino acid into monolignols. Thus, the lignin content was decreased to 50% in 4CL down-regulated transgenic tobacco (Kajita *et al.*, 1996), and somewhat reduced in 4CL down-regulated *Arabidopsis* (Lee *et al.*, 1997). Antisense suppression of *Pt4CL1* in transgenic aspen showed up to a 45% reduction of lignin content and increased in cellulose content (Hu *et al*, 1999). To further analyse the role of the 4CL6 isoform in lignifying tissues of poplar, an antisense construct of *4CL6* gene could be transferred into poplar, and the transgenic trees tested for changes in lignin structure and content. A similar approach could be used to test the proposed role of the 4CL10 isoform in other phenylpropanoid pathways (e.g. a role in pathogen resistance).

Bibliography

Allina SM, Douglas CJ (1994) Isolation and characterization of the 4-coumarate:CoA ligase gene family in poplar hybrid (abstract No. 852). Plant Physiolo **105:** S-154.

Allina SM, Pri-Hadash A, Theilmann DA, Ellis BE, Douglas CJ (1998) 4-coumarate:Coenzyme A ligase in hybrid poplar. Plant Physiol **116:** 743-754.

Atanassova R, Favet N, Martz F, Chabbert B, Tollier MT, Monties B, Fritig B, Legrand M (1995) Altered lignin composition in transgenic tobacco expressing O-methyltransferase sequences in sense and antisense orientation. Plant J **8:** 465-477.

Bairoch A (1991) Putative AMP-binding domain signature. PROSITE: a dictionary of sites and patterns in proteins. Release 8.00, Medical Biochemistry Department, University of Geneva, Switzerland.

Baucher M, Chabbert B, Pilate G, van Doorsselaere J, Tollier M-T, Petit-Conil M, Cornu D, Monties B, van Montagu M, Inzé D, Jouanin L, Boerjan W (1996) Red xylem and higher lignin extractability by down-regulating cinnamyl alcohol dehydrogenase in poplar (*Populus tremula* and *Populus alba*). Plant Physiol **112:** 1479-1490

Becker-André M, Schulze-Lefert P, Hahlbrock K (1991) Structural comparison, modes of expression, and putative cis-acting elements of the two 4-coumarate:CoA ligase genes in potato. J Biol Chem **266:** 8551-8559.

Boudet AM, Lapierre C, Grima-Pettenati J (1995) Biochemistry and molecular biology of lignification. New Phytol **129:** 203-236.

Boudet AM (1998) A new view of lignification. Trends Plant Sci **3:** 67-71.

Bradshaw HD, Villar M, Watson BD, Otto KG, Stewart S, Stettler RF (1994) Molecular genetics of growth and development in *Populus*. III. A genetic linkage map of a hybrid poplar composed of RFLP, STS and RAPD markers. Theor Appl Genet **89:** 167-178.

Brodelius PE, Xue ZT (1997) Isolation and characterization of cDNA from cell suspension cultures of *Vanilla planifolia* encoding 4-coumarate:Coenzyme A ligase. Plant Physiol Biochem **35:** 497-506.

Bugos RC, Chiang VLC, Campbell WH (1991) cDNA cloning, sequencing analysis and seasonal expression of lignin-bispecific caffeic acid/5-hydroxyferulic acid *O*-methyltransferase of aspen. Plant Mol Biol **17:** 1203-1215.

Butland SL, Chow ML, Ellis BE (1998) A diverse family of phenylalanine ammonia-lyase genes in pine trees and cell cultures. Plant Mol Biol **37:** 15-24.

Campbell MM, Sederoff RR (1996) Variation in lignin content and composition. Plant Physiol **110:** 3-13.

Chapple CC, Vogt T, Ellis BE, Somerville CR (1992) An *Arabidopsis* mutant defective in the general phenylpropanoid pathway. Plant Cell **4:** 1413-24.

Church GM, Gilbert W (1984) Genomic sequencing. Proc Natl Acad Sci USA **81:** 1991-1995

Chen C-L (1991) Lignin: occurrence in woody tissues, isolation, reactions and structure. *In* M Lewin, IS Goldstein, eds, Wood structure and composition, Dekker, New York, pp 183-261

Davin LB , Lewis NG (1992) Phenylpropanoid metabolism: biosynthesis of monolignols, lignans and neolignals, lignin and suberins. *In* HA Stafford, RK Ibrahim, eds, Phenolic Metabolism in Plants, Plenum Press, New York, pp 325-375.

Dixon RA , Paiva NL (1995) Stress-induced phenylpropanoid metabolism. Plant Cell **7:** 1085-1097.

Douglas CJ, Hoffmann H, Schulz W, Hahlbrock K (1987) Structure and elicitor or u.v.-light stimulated expression of two 4-coumarate:CoA ligase in parsley. EMBO J **10:** 1189-1195.

Douglas CJ, Hauffe KD, Ites-Moralse ME, Ellard M, Paszkowski U, Hahlbrock K, Dangl JL (1991) Exonic sequences are required for elicitor and light activator of a plant defense gene, but promoter sequences are sufficient for tissue specific expression. EMBO J **10:** 1767-1775.

Douglas CJ, Ellard M, Hauffe KD, Molitor E, Moniz de Sá M, Reinold R, Subramaniam R, Wiliams F (1992) General phenylpropanoid metabolism: regulation by environmental and developmental signals. *In* HA Stafford, RK Ibrahim eds, Recent Advances in Phytochemistry, Vol 26, Plenum Press, New York, pp 63-89.

deWet JR, Wood KV, deLuca M, Helinski DR, Subramani S (1987) Firefly luciferase gene: structure and expression in mammalian cells. Mol Cell Biol **7:** 725-737.

Ehlting J, Büttner D, Wang Q, Douglas CJ, Somssich IE, Kombrink E (1999) Three 4-coumarate:Coenzyme A ligases (4CL) in *Arabidopsis thaliana* represent two evolutionary divergent 4CL classes in angiosperm plants. Plant J **19:** 9-20.

Ellard-Ivey M, Douglas CJ (1996). Role of jasmonates in the elicitor- and wound-inducible expression of defense genes in parsley and transgenic tobacco. Plant Physiol **112:** 183-192.

English S, Greenway W, Whatley FR (1991) Analysis of phenolics of *Populus trichocarpa* bud exudates by GC-MS. Phytochemistry **30:** 531-533.

Grand C, Boudet A, Boudet AM (1983) Isoenzyme of hydroxycinnamate:CoA ligase from poplar stems properties and tissue distribution. Planta **158:** 225-229.

Greenway W, English S, Whatley FR (1990) Phenolic composition of bud exudates of *Populus deltoides*. Z Naturforsch **45c:** 587-593.

Griesbach H (1981) Lignins. *In* PK Stumff, EE Conn, eds, The Biochemistry of Plants. A Comprehensive Treatise of Secondary Plant Products, Vol 7, Academic Press, New York, pp. 457-478.

Gross GG, Zenk MH (1966) Darstellung und Eigenschaften von Coenzyme A Thioestern substituiter Zinsturen. Z Naturforsch **21b:** 683-690.

Gross GG, Zenk MH (1974) Isolation and properties of hydroxycinnamate:CoA ligase from lignifying tissue of *Forsythia*. Eur J Biochem **42:** 453-459.

Gross GG (1985) Biosynthesis and metabolism of phenolic acids and monolignols. *In* T Higuchi, ed, Biosynthesis and biodegradation of wood components, Academic Press Inc, New York, pp. 229-271.

Halpin C, Knight ME, Foxon GA, Campbell MM, Boudet AM, Boon JJ, Chabbert B, Tollier M-T, Schuch W (1994) Manipulation of lignin quality by downregulation of cinnamyl alcohol dehydrogenase. Plant J **6:** 339-350.

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

Hauffe KD, Paszkowski U, Schulze-Lefert P, Hahlbrock K, Dangl JL, Douglas CJ (1991) A parsley 4CL-1 promoter fragment specifies complex expression patterns in transgenic tobacco. Plant Cell **3:** 435-443.

Heinzmann U, Seitz U, Seitz U (1977) Purification and substrate specificities of hydroxycinnamate:CoA ligase from anthocyanin-containing and anthocyanin-free carrot cells. Planta **135:** 313-318.

Hibino T, Takabe K, Kawazu T, Shibata D, Higuchi T (1995) Increase of cinnamaldehyde groups in lignin of transgenic tobacco plants carrying an antisense gene for cinnamyl alcohol dehydrogenase. Biosci Biotechnol Biochem **59:** 929-931.

Higuchi T (1985) Biosynthesis of lignin. *In* T Higuchi, ed, Biosynthesis and Biodegradation of Wood Components, Academic Press Inc, New York, pp 141-160.

Hipskind JD, Vincent JR, Butler LG, Nicholson RL (1993) Partial purification and characterization of 4-hydroxycinnamic acid:CoA ligase from maize leaves infected with *Bipolaris maydis*. Pysiol Mol Plant Pathology **43:** 365-377.

Hu WJ, Kawaoka A, Tsai CJ, Lung J, Osakabe K (1998) Compartmentalized expression of two structurally and functionally distinct 4-coumarate:CoA ligase genes in aspen (*Populus tremuloides*). Proc Nat Acad Sci USA **95:** 5407-5412.

Hu WJ, Harding SA, Lung J, Popko JL, Ralph J, Stokke DD, Tsai CJ, Chiang VL (1999) Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees. Nat Biotechnol **17**: 808-812.

Hughes WD, Galau G (1988) Preparation of RNA from cotton leaves and pollen. Plant Mol Biol Rep **6:** 253-257.

Jorgensen A, Cuellar E, Thomson F (1982) Modes and tempos in the evolution of nuclear-encoded ribosomal RNA genes in legumes. Carnegie Inst Wash Year Book **81:** 98-101.

Kajita S, Osakabe K, Katayama Y, Kawai S, Matsumoto Y, Hata K, Moroshoshi N (1994) *Agrobacterium*-mediated transformation of poplar using a disarmed binary vector and the overexpression of specific member of poplar peroxidase genes in transgenic poplar cell. Plant Sci **103:** 231-239.

Kajita S, Katayama Y, Omori S (1996) Alterations in the biosynthesis of lignin in transgenic plants with chimeric genes for 4-coumarate:Coenzyme A ligase. Plant Cell Physiol **37:** 957-965.

Knobloch KH , Hahlbrock K (1975) Isoenzyme of *p*-coumarate:CoA ligase from cell suspension cultures of *Glycine max*. Eur J Biochem **52:** 331-320.

Knobloch KH, Hahlbrock K (1977) 4-coumarate:CoA ligase from cell suspension cultures of *Petroselinum hortense* Hoffm. Arch Biochem Biophys **184:** 237-248.

Kutsuki H, Shimada M, Higuchi T (1982) Distribution and roles of *p*-hydroxycinnamate:CoA ligase in lignin biosynthesis. Phytochemistry **21:** 267-271.

Laemmli UK (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227:** 680-685

Lapierre C, Pollet B, Petit-Conil M, Toval G, Romero J, Pilate G, Leplé J-C, Boerjan W, Ferret V, de Nadai V, Jouanin L (1999) Structural alternations of lignins in transgenic poplars with depressed cinnamyl alcohol dehydrogenase or caffeic acid *O*-methyltransferase activity have opposite impact on the efficiency of industrial kraft pulping. Plant Physiol **119:** 153-163.

Laplé JC, Brasileiro ACM, Michel MF, Delmo NSF, Jouanin L (1992) Transgenic poplars: expression of chimeric genes using four different constructs. Plant Cell Reports **11:** 137-141.

Lee D, Ellard M, Wanner LA, Davis KR, Douglas CJ (1995) The *Arabidopsis thaliana* 4-coumarate:CoA ligase (*4CL*) gene: stress and developmentally regulated expression and nucleotide sequence of its cDNA. Plant Mol Biol **28:** 871-884.

Lee D, Douglas CJ (1996) Two divergent members of a tobacco 4-coumarate:coenzyme A ligase (*4CL*) gene family. Plant Physiol **112:** 193-205

Lee D, Meyer K, Chapple C, Douglas CJ (1997) Antisense suppression of 4-coumarate: Coenzyme A ligase activity in *Arabidopsis* leads to altered lignin subunit composition. Plant Cell **9:** 1985-1998.

Lewis NG, Yamamoto E (1990) Lignin: occurrence, biogenesis and biodegradation. Ann Rev Plant Physiol Plant Mol Biol **41:** 455-496.

Lois R, Hahlbrock K (1992) Differential wound activation of members of the phenylanine ammonia-lyase and 4-coumarate:CoA ligase gene families in various organs of parsley plants. Z Natureforsch **47c:** 90-94.

Lozoya E, Hoffmann H, Douglas C, Schulz W, Scheel D, Hahlbrock K (1988) Primary structures and catalytic properties of isoenzymes encoded by the two 4-coumarate:CoA ligase genes in parsley. Eur J Biochem **176:** 661-667.

Lüderitz T, Schatz G, Grisebach H (1982) Enzymic synthesis of lignin precursors: purification and properties of 4-coumarate:CoA ligase from cambial sap of spruce (*Picea abies* L.). Eur J Biochem **123:** 583-586.

Meng H, Campbell WH (1995) Cloning of aspen (*Populus tremuloides*) xylem caffeoyl-CoA 3-*O*-methyltransferase (GenBank U27116). Plant Physiol **108:** 1749; Plant Gene Register PGR95-040.

Meng H, Campbell WH (1998) Substrate profiles and expression of caffeoyl coenzyme A and caffeic acid *O*-methyltransferase in secondary xylem of aspen during seasonal development. Plant Mol Biol **38:** 513-520.

Meyer K, Shirley AM, Cusumano JC, Bell-Lelong DA, Chapple C (1998) Lignin monomer composition is determined by the expression of a cytochrome P450-dependent monooxygenase in *Arabidopsis*. Proc Natl Acad Sci USA **95:** 6619-6623.

Moniz de Sá M, Subramaniam R, Williams FE , Douglas CJ (1992) Rapid activation of phenylpropanoid metabolism in elicitor-treated hybrid poplar (*Populus trichocarpa* Torr. & Gray x *Populus deltoides* Marsh) suspension-cultured cells. Plant Physiol **98:** 728-737.

Osakabe Y, Osakabe K, Kawai S, Katayama Y, Morohoshi N (1995) Characterization of the structure and determination of mRNA level of the phenylalanine ammonia-lyase gene family from *Populus kitakamiensis*. Plant Mol Biol **28:** 1133-1141.

Pearl IA, Darling SF (1968) Studies on the barks of the family Salicaceae-XIX. Phytochemistry **7:** 1851-1853.

Ranjeva R, Boudet AM , Faggion R (1976) Phenolic metabolism in petunia tissues. Biochemie **58:** 1255-1262.

Reinold S, Hauffe KD, Douglas CJ (1993) Tobacco and parsley 4-coumarate:CoA ligase genes are temporally and spatially regulated in a cell type-specific manner during tobacco flower development. Plant Physiol **101:** 373-383.

Rogers SO, Bendich AJ (1988) Extraction DNA from plant tissues. *In* SB Gelvin, RA Schilperoort, D-PS Verma eds, Plant Molecular Biology Manual, Kluwer Academic Press, Dordrecht Boston London, pp A2: 1-3.

Sambrook L, Fritsch EF, Maniatis T (1989) Molecular Cloning: A laboratory Manual, 2nd edn. Cold Spring Harbor Laboratory Press, New York.

Schmelzer E, Kruger-Lebus S, Hahlbrock K (1989) Temporal and spatial patterns of gene expression around sites of attempted fungal infection in parsley leaves. Plant Cell **1:** 993-1001

Schröder J (1989) Protein sequence homology between plant 4-coumarate: CoA ligase and firefly luciferase. Nucl Acids Res **17:** 460.

Sederoff R, Campbell M, O'Malley D, Whetten R (1994) Genetic regulation of lignin biosynthesis and the potential modification of wood by genetic engineering in loblolly pine. *In* BE Ellis, GW Kuroki, HA Stafford, eds, Genetic engineering of plant secondary metabolism, Plenum Press, New York, pp 313-355.

Sells MA, Chernoff J (1995) Polymerase chain reaction cloning of related genes. Methods Enzym **254:** 184-195.

Shain L, Miller JB (1982) Pinocembrin: An antifugal compound secreted by leaf glands of eastern cottonwood. Phytopathology **72:** 877-880.

Sterky F, Regan S, Karlsson J, Hertzberg M, Rohde A, Holmberg A, Amini B, Bhalerao R, Larsson M, Villarroel R, van Montagu M, Sandberg G, Olsson O, TT., Boerjan W, Gustafsson P, Uhlen M, Sundberg B , Lundeberg J (1998) Gene discovery in the wood-forming tissues of poplar: analysis of 5,692 expressed sequence tags. Proc Natl Acad Sci USA **95:** 1330-1335.

Stöckigt J, Zenk MH (1975) Chemical synthesis and properties of hydroxycinnamoyl-coenzyme A derivatives. Z Naturforsch **30c:** 352-358.

Swan EP, Kellog RM (1986) Chemical properties of black cottonwood and balsam poplar. Can J For Res **16:** 497-501.

Subramaniam R, Reinold S, Molitor EK, Douglas CJ (1993) Structure, inheritance, and expression of hybrid poplar (*Populus trichocarpa* X *Populus deltoides*) phenylalanine ammonya-lyase genes. Plant Physiol **102:** 71-83.

Summers MD, Smith GE (1987) A manual of methods for baculovirus vectors and insect cell culture procedures. Tex Agric Exp Stn Bull No1555.

Uhlmann A, Ebel J (1993) Molecular cloning and expression of 4-coumarate:Coenzyme A ligase, an enzyme involved in the resistance response of soybean (*Glycine max* L.) against pathogen attack. Plant Physiol **102:** 1147-1156.

van Doorsselaere J, Baucher M, Chognot E, Chabbert B, Tollier M-T, Petit-Conil M, Leplé J-C, Pilate G, Cornu D, Monties B, van Montagu M, Inzé D, Jouanin L (1995) A novel lignin in poplar trees with reduced caffeic acid/5-hydroxyferulic acid *O*-methyltransferase activity. Plant J **8:** 855-864.

Vincent JR, Nicholson RL (1987) Evidence for isoenzymes of 4-hydroxycinnamic acid: CoA ligase in maize mesocotyls and their response to infection by *Helminthosporium maydis* race 0. Physiol Mol Plant Pathology **30:** 121-129.

Voo KS, Whetten RW, O'Malley DM, Sederoff RR (1995). 4-coumarate:Coenzyme A ligase from loblolly pine xylem. Plant Physiol **108:** 85-97.

Wallis PJ, Rhodes MJC (1977) Multiple forms of hydroxycinnamate:CoA ligase in etiolated pea seedlings. Phytochemistry **16:** 1891-1894.

Whetten R , Sederoff R (1995). Lignin biosynthesis. Plant Cell **7:** 1001-1013.

Whetten RW, MacKay JJ, Sederoff RR (1998). Recent advances in understanding lignin biosynthesis. Annu Rev Plant Physiol Plant Mol Biol **49:** 585-609.

Wilks AF, Kurban RR, Hovens CM, Ralph SJ (1989) The application of the polymerase chain reaction to cloning members of the protein tyrosine kinase. Gene **85:** 67-74.

Wu S-C, Hahlbrock K (1992) *In situ* localization of phenylpropanoid-related gene expression in different tissues of light- and dark-grown parsley seedlings. Z Naturforsch. **47c:** 591-600.

Yazaki K, Ogawa A , Tabata M (1995) Isolation and characterization of two cDNAs encoding 4-coumarate:CoA ligase in *Lithospermum* cell cultures. Plant Cell Physiol **36:** 1319-1329.

Ye ZH, Kneusel RE, Matern U, Varner JE (1994) An alternative methylation pathway in lignin biosynthesis in *Zinnia*. Plant Cell **6:** 1427-1439.

Ye ZH,Varner JE (1995) Differential expression of two *O*-methyltransferases in lignin biosynthesis in *Zinnia elegans*. Plant Physiol **108:** 459-467.

Zhang XH, Chiang VL (1997) Molecular cloning of 4-coumarate:CoA ligase in loblolly pine and the roles of this enzyme in the biosynthesis of lignin in compression wood. Plant Physiol **113:** 65-74.

Zhao Y, Kung SD, Dube SK (1990) Nucleotide sequence of rice 4-coumarate:CoA ligase gene, 4-CL.1. Nucl Acids Res **18:** 6144.