Available online at www.sciencedirect.com
SCIENCE DIRECT®
Phytochemistry 66 (2005) 2072–2091
PHYTOCHEMISTRY
www.elsevier.com/locate/phytochem

Characterization in vitro and in vivo of the putative multigene 4-coumarate:CoA ligase network in *Arabidopsis*: syringyl lignin and sinapate/sinapyl alcohol derivative formation ☆

Michael A. Costa, Diana L. Bedgar, Syed G.A. Moinuddin, Kye-Won Kim, Claudia L. Cardenas, Fiona C. Cochrane, Jay M. Shockey [1], Gregory L. Helms, Yoshiaki Amakura, Hironobu Takahashi, Jessica K. Milhollan, Laurence B. Davin, John Browse, Norman G. Lewis *

Institute of Biological Chemistry, Washington State University, Pullman, WA 99164-6340, USA

Received 15 June 2005; accepted 20 June 2005
Available online 15 August 2005

Abstract

A recent in silico analysis revealed that the *Arabidopsis* genome has 14 genes annotated as putative 4-coumarate:CoA ligase isoforms or homologues. Of these, 11 were selected for detailed functional analysis in vitro, using all known possible phenylpropanoid pathway intermediates (*p*-coumaric, caffeic, ferulic, 5-hydroxyferulic and sinapic acids), as well as cinnamic acid. Of the 11 recombinant proteins so obtained, four were catalytically active in vitro, with fairly broad substrate specificities, confirming that the *4CL* gene family in *Arabidopsis* has only four members. This finding is in agreement with our previous phylogenetic analyses, and again illustrates the need for comprehensive characterization of all putative 4CLs, rather than piecemeal analysis of selected gene members. All 11 proteins were expressed with a C-terminal His_6-tag and functionally characterized, with one, At4CL1, expressed in native form for kinetic property comparisons.

Of the 11 putative His_6-tagged 4CLs, isoform At4CL1 best utilized *p*-coumaric, caffeic, ferulic and 5-hydroxyferulic acids as substrates, whereas At4CL2 readily transformed *p*-coumaric and caffeic acids into the corresponding CoA esters, while ferulic and 5-hydroxyferulic acids were converted quite poorly. At4CL3 also displayed broad substrate specificity efficiently converting *p*-coumaric, caffeic and ferulic acids into their CoA esters, whereas 5-hydroxyferulic acid was not as effectively utilized. By contrast, while At4CL5 is the only isoform capable of ligating sinapic acid, the two preferred substrates were 5-hydroxyferulic and caffeic acids. Indeed, both At4CL1 and At4CL5 most effectively utilized 5-hydroxyferulic acid with $k_{enz} \sim 10$-fold higher than that for At4CL2 and At4CL3.

The remaining seven 4CL-like homologues had no measurable catalytic activity (at ~100 μg protein concentrations), again bringing into sharp focus both the advantages to, and the limitations of, current database annotations, and the need to unambiguously demonstrate true enzyme function.

Abbreviations: 4CL, 4-coumarate:CoA ligase; CWR, cell wall residue; G, guaiacyl; H, *p*-hydroxyphenyl; IPTG, isopropyl-β-D-thiogalactopyranoside; RT-PCR, reverse transcription-polymerase chain reaction; S, syringyl; TAIR, The *Arabidopsis* Information Resource; TIGR, The Institute for Genomic Research; WT, wild type.

☆ *Data deposition*: The cDNA sequences obtained during the course of this research were deposited in the GenBank database with the following accession numbers: AY376729 (*At4CL1*), AY376728 (*At4CL2*), AY376730 (*At4CL3*), AY376731 (*At4CL4*), AY376732 (*At4CL5*), AY250839 (*At4CL6*), AY376733 (*At4CL7*), AY376734 (*At4CL10*), AY376735 (*At4CL11*).

* Corresponding author. Tel.: +1 509 335 2682; fax: +1 509 335 8206.
E-mail address: lewisn@wsu.edu (N.G. Lewis).
[1] Present address: Southern Regional Research Center, USDA-ARS, New Orleans, LA 70124, USA.

0031-9422/$ - see front matter
doi:10.1016/j.phytochem.2005.06.022

Lastly, although At4CL5 is able to convert both 5-hydroxyferulic and sinapic acids into the corresponding CoA esters, the physiological significance of the latter observation in vitro was in question, i.e. particularly since other 4CL isoforms can effectively convert 5-hydroxyferulic acid into 5-hydroxyferuloyl CoA. Hence, homozygous lines containing T-DNA or enhancer trap inserts (knockouts) for *4cl5* were selected by screening, with *Arabidopsis* stem sections from each mutant line subjected to detailed analyses for both lignin monomeric compositions and contents, and sinapate/sinapyl alcohol derivative formation, at different stages of growth and development until maturation. The data so obtained revealed that this "knockout" had no significant effect on either lignin content or monomeric composition, or on the accumulation of sinapate/sinapyl alcohol derivatives. The results from the present study indicate that formation of syringyl lignins and sinapate/sinapyl alcohol derivatives result primarily from methylation of 5-hydroxyferuloyl CoA or derivatives thereof rather than sinapic acid ligation. That is, no specific physiological role for At4CL5 in direct sinapic acid CoA ligation could be identified. How the putative overlapping 4CL metabolic networks are in fact organized in planta at various stages of growth and development will be the subject of future inquiry.

Keywords: *Arabidopsis thaliana*; Cruciferae; 4-Coumarate CoA ligases; 4-Coumarate CoA ligase homologues; Lignin; Knockout analysis; Metabolic networks; Sinapate/sinapyl alcohol derivatives

1. Introduction

Sequencing of the *Arabidopsis thaliana* genome (The Arabidopsis Genome Initiative, 2000) as a model plant system has provided an avenue for systematic functional analysis of its (>28,000) genes, this being a goal of the National Science Foundation entitled the "NSF *Arabidopsis* 2010 Initiative". This laboratory has recently undertaken the task of the comprehensive study of some 248 genes annotated in *A. thaliana* databases, with each having sequence homology to genes primarily involved and/or having suspected roles in phenylpropanoid pathway related metabolism and/or cell wall formation. Recent reports have described, in a comprehensive manner, the cinnamyl alcohol dehydrogenase (Kim et al., 2004), the phenylalanine ammonia lyase (Cochrane et al., 2004) and the acyl activating enzyme (Shockey et al., 2003) gene families. Of particular interest is how various metabolic networks associated with phenylpropanoid pathway metabolism are organized in planta.

The present investigation has as its objective to address further the precise physiological/enzymatic functions of the various members in the *Arabidopsis* gene family annotated as *4-coumarate CoA ligase* (*4CL*) or *4CL*-like, with initial attention being given to those with bona fide 4CL activities. In this regard, a recent in silico analysis of available databases (Costa et al., 2003) revealed that 14 genes were annotated as (putative) *4CL*s (*At4CL1–14*) in the *Arabidopsis* genome, of which At4CL1, 2, 3, 5 and 10 grouped together (Shockey et al., 2003) and were thus the most likely candidates for the *At4CL* multigene family. [4CL12–14 were eliminated from further consideration in part because of the presence of C-terminus peroxisome targeting signal sequences which are lacking in bona fide cytosolic 4CLs (Costa et al., 2003; Shockey et al., 2003).] The 11 genes (*At4CL1–11*) were thus cloned, and the corresponding recombinant proteins (each with a C-terminal His_6-tag) expressed in *Escherichia coli*. All 11 proteins were then characterized through a combination of sequence analyses and in vitro kinetic studies using six hydroxycinnamic acid derivatives as possible substrates [i.e. cinnamic (**1**), *p*-coumaric (**2**), caffeic (**3**), ferulic (**4**), 5-hydroxyferulic (**5**), and sinapic (**6**) acids].

Prior to discussing the results obtained and their significance, a brief historical summary of progress in the 4CL field is provided for needed context (see Lewis et al., 1999, for a detailed review and references therein). Thus, recognition that hydroxycinnamic acids were likely to be activated as their CoA esters, prior to subsequent metabolism, was first proposed by Brown et al. (1959) based on analogous observations in other biochemical pathways. Later, Gross and Zenk (1966) provided the first indirect proof for involvement of hydroxycinnamoyl CoA esters in monolignol formation, and subsequently in the 1970s, hydroxycinnamate CoA ligase activities were demonstrated in various plants at either the cell-free extract level or with partially purified protein preparations (Hahlbrock and Grisenbach, 1970; Walton and Butt, 1970; Mansell et al., 1972; Gross et al., 1973; Lindl et al., 1973; Rhodes and Wooltorton, 1973, 1975; Gross and Zenk, 1974; Knobloch and Hahlbrock, 1975; Wallis and Rhodes, 1977). Although such studies did not result in pure 4CL isoforms being obtained, it soon became apparent that sinapic acid (**6**) did not serve as a substrate for ligation using, for example, the partially purified CoA ligase preparations from various species, such as *Acer saccharinum*, *Forsythia suspensa*, *Taxus baccata* (Gross et al., 1975) and parsley (*Petroselinum crispum*) (Knobloch and Hahlbrock, 1977). However, in the latter case, it was shown that 5-hydroxyferulic acid (**5**) was effectively converted into 5-hydroxyferuloyl CoA (**11**) suggesting that it (or a reduced derivative thereof) served as a methylation substrate during syringyl (sinapyl alcohol **17** derived) lignin formation.

It was not, however, until the early 1980s that purification of 4CL isoforms to apparent homogeneity was

achieved using parsley (Ragg et al., 1981) and spruce (*Picea abies*) (Lüderitz et al., 1982). The first cloning of near full-length cDNAs for two 4CL parsley isoenzymes followed in 1987–1988 (Douglas et al., 1987; Lozoya et al., 1988), and preliminary characterization of both isoforms indicated that they were able to convert cinnamate (**1**), *m*-coumarate (**13**), *o*-coumarate (**14**), *p*-coumarate (**2**), caffeate (**3**) and ferulate (**4**) into the corresponding CoA esters **7**, **15**, **16**, **8**–**10**, whereas sinapate (**6**) did not undergo ligation, i.e. in agreement with previous observations. Definition of the ATP/AMP presumed binding domains then became possible by sequence homology comparisons with other much more well-studied ATP-requiring enzymes used in aromatic carboxylic acid activation (Becker-André et al., 1991). Other workers also reported homologous genes encoding 4CLs in rice (Zhao et al., 1990), soybean (*Glycine max*) (Uhlmann and Ebel, 1993); *Arabidopsis* (Lee et al., 1995, 1997; Ehlting et al., 1999; Cukovic et al., 2001), loblolly pine (*Pinus taeda*) (Voo et al., 1995; Zhang and Chiang, 1997); tobacco (Kajita et al., 1996; Lee and Douglas, 1996), aspen (*Populus tremuloides*) (Hu et al., 1998; Harding et al., 2002) and hybrid poplar (*Populus trichocarpa* × *Populus deltoides*) (Allina et al., 1998). None of these, however, had any demonstrable capacity for ligating sinapic acid (**6**).

Discovery of a gene encoding a 4CL capable of ligating sinapic acid (**6**) to afford sinapoyl CoA (**12**) was first reported in the study of different members of a soybean (*Glycine max*) *4CL* gene family (Lindermayr et al., 2002). Its gene (trivially annotated as *Gm4CL1*) encoded a protein of broad substrate versatility that was able to convert cinnamate (**1**), *p*-coumarate (**2**), caffeate (**3**), ferulate (**4**) and sinapate (**6**) into the corresponding CoA esters **7**–**10** and **12**, respectively, as gauged by spectrophotometric analyses of the assay mixtures; surprisingly, 5-hydroxyferulic acid (**5**) was not evaluated as a substrate. The overall study was quite limited, however, in terms of detailed biochemical characterization, with only K_m and "relative V_{max} (% of coumarate)" values being determined. [The remaining other three 4CLs (Gm4CL2–4) were, by contrast, unable to ligate sinapate (**6**) under the conditions employed.] Nevertheless, encouraging evidence was reported for the first 4CL isoform able to convert sinapic acid (**6**) into sinapoyl CoA (**12**).

The study of 4CL in *Arabidopsis* is quite noteworthy, particularly given the twists and turns that this endeavor has taken. Initially, it was reported that there was only one *4CL* gene in *Arabidopsis* (Lee et al., 1995), whose subsequent downregulation (under control of either a *CaMV 35S* or a parsley *4CL* promoter) gave transformants with reportedly highly variable characteristics in terms of lignin contents and/or monomeric compositions (Lee et al., 1997). These experimental data, and their limitations, have been exhaustively and critically reviewed elsewhere (Anterola and Lewis, 2002). They need not be repeated here, other than to summarize that the syringyl and guaiacyl lignin derivative recoveries were between 80% and 115% of the estimated lignin contents, rather than the much lower recoveries to be expected (30–50%). Moreover, there was no obvious correlation between presumed lignin downregulation and observed G/S ratios. These data raised legitimate questions about the choice of techniques used for plant analysis, and hence of the conclusions made.

Later studies (Ehlting et al., 1999; Cukovic et al., 2001) then extended the number of 4CLs present in *Arabidopsis* from one to three, each of which displayed the expected broad substrate versatilities with several cinnamic acid derivatives, except for sinapate (**6**) which was not a substrate. As for Lindermayr et al. (2002), enzyme assays were carried out spectrophotometrically, with each of the three isoforms being only preliminarily characterized in terms of K_m and "relative V_{max} (% of coumarate)" estimations. The *Arabidopsis 4CL* gene family was then further extended to four isoenzymes (including an isoform apparently capable of ligating sinapic acid (**6**) as determined spectrophotometrically) (Schneider et al., 2003). This same gene was yet again reported as encoding a 4CL capable of ligating sinapic acid (**6**) (Hamberger and Hahlbrock, 2004). Interestingly, in none of these studies was 5-hydroxyferulic acid (**5**) used as substrate, nor were the catalytic properties in vitro of other homologues described.

In the present study, we thus undertook a comprehensive and detailed analysis of the 11 putative members of the *Arabidopsis 4CL* gene family with no prejudice or preference being given to the degree of homology. In each case, the recombinant proteins were purified to apparent homogeneity, with definitive enzymatic data determined by HPLC analysis of assay mixtures in order to verify unambiguously product authenticity rather than by spectrophotometric assays. Compounds **1**–**6** were used as potential substrates, and the data obtained establish that there is a family of only four *4CL* genes whose encoded proteins can ligate various (hydroxy)cinnamic acids; the relative properties of each 4CL are described below. An *4cl5* knockout was also analyzed to explore the extent, if any, to which syringyl lignin or sinapate/sinapyl alcohol derivative formation occurs if this *4CL* gene family member is disrupted.

2. Results and discussion

2.1. Arabidopsis genome analysis and reclassification of annotated putative 4CL homologues

The Institute for Genomic Research (TIGR) database (Costa et al., 2003) has 14 entries (*At4CL1–14*)

annotated as putative *4CL* genes, of which *At4CL12–14* were eliminated as previously discussed. Three of the eleven genes (i.e. *At4CL1–3*) encoded bona fide 4CL proteins (Ehlting et al., 1999) (Table 1 and Fig. 1), with At4CL2 and At4CL3 having quite high similarity (87.2% and 71.3%) and identity (83.4% and 61.3%) to At4CL1. Relative to *At4CL1*, comparison of the remaining eight genes (*At4CL4–At4CL11*) indicated that they encoded proteins with similarities and identities ranging from 74.5% to 42.8% and 66.4% to 31.1%, respectively (Table 1); of these, At4CL5 displayed quite good homology to At4CL1 (74.5% similarity; 66.4% identity) suggesting it to be another likely *At4CL* gene family member. Although the remainder (At4CL6–11) had significantly lower levels of homology (55.3–42.8% similarity; 43.9–31.1% identity) to that of At4CL1, our previous phylogenetic comparison of the acyl activating enzyme superfamily in *Arabidopsis* suggested that At4CL10 should also be grouped in the At4CL family (Shockey et al., 2003). Therefore, we provisionally proceeded with the prediction that the *At4CL* family consisted of five members. The remaining genes *At4CL4*, *At4CL6–9*, and *At4CL11*, although annotated as *4CL*s were less likely to be as such based also on our previous phylogenetic analyses (see Shockey et al., 2003). Nevertheless, each needed to be evaluated experimentally in vitro on an individual basis to assess catalytic function.

2.2. Chromosomal organization and gene cloning

The putative *At4CL* genes (*At4CL1–11*) of interest are distributed in four of the five *Arabidopsis* chromosomes (Table 1): chromosome 1 (*At4CL1*, *3*, *9* and *10*), chromosome 3 (*At4CL2*, *5* and *8*), chromosome 4 (*At4CL6* and *7*), and chromosome 5 (*At4CL4* and *11*), with *At4CL2* and *5* being clustered together. Of these genes, the ultimately experimentally verifiable bona fide *At4CL*s (i.e. *At4CL1–3* and *5* discussed below) were localized on chromosomes 1 and 3. It was also noted that all 11 putative *At4CL* genes (*At4CL1–11*) were characterized by having a minimum of three introns, with the shortest being a little less than 100 bp and the longest ~1700 bp (*At4CL5*).

At4CL1–7 as well as *At4CL10* and *11* genes were obtained by RT-PCR using gene-specific primers (Table 2) designed for each annotated *At4CL* DNA sequence, and RNA isolated from 16-day old tissue as a template. Clones for *At4CL8* and *At4CL9* were obtained from the Plant Genome Expression Center.

2.3. 4CL recombinant protein expression and kinetic characterization

Following individual induction of transformed *E. coli* cells (see Section 4) with IPTG, crude protein extracts for each At4CL homologue were prepared from cell pellets (derived from 500 ml of culture). Each C-terminal His_6-tagged At4CL homologue was then individually purified to apparent homogeneity using metal chelate affinity chromatography; Fig. 2 shows a representative example (At4CL2), with purity assessed by SDS–PAGE and silver staining. The amount of pure recombinant At4CL protein obtained for each homologue was typically 30–50 mg/l of culture, except for At4CL5 which gave lower amounts (~12 mg/l). It is noteworthy that while the C-terminal His_6-tag did not adversely affect the activity of bona fide At4CLs (described below), the presence of >5 mM imidazole drastically reduced detectable activity. The imidazole was therefore removed by exchanging the eluting solution composition with 20 mM Tris–HCl (pH 7.5) during concentration. Purified proteins were stable at −20 °C in 10% (v/v) glycerol for 30 days but only for circa 4 days at 4 °C.

For kinetic property comparisons, native At4CL1 was also expressed in recombinant form, this being

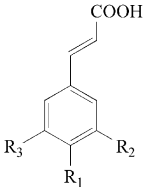
Table 1
Sequence comparison of putative *Arabidopsis thaliana 4CL* genes and corresponding proteins with respect to At4CL1

Gene	*At4CL* homologue[a]	cDNA (bp)[b]	Protein (kDa)	Comparison with At4CL1	
				Similarity (%)	Identity (%)
At1g51680	*At4CL1*	1686	61.0	100.0	100.0
At3g21240	*At4CL2*	1671	60.8	87.2	83.4
At1g65060	*At4CL3*	1686	61.3	71.3	61.3
At5g63380	*At4CL4*	1689	61.3	50.6	37.4
At3g21230	*At4CL5*	1713	62.5	74.5	66.4
At4g05160	*At4CL6*	1635	59.8	55.3	43.9
At4g19010	*At4CL7*	1701	61.9	49.6	39.8
At3g48990	*At4CL8*	1545	55.5	42.8	31.1
At1g20510	*At4CL9*	1641	59.3	50.6	40.9
At1g62940	*At4CL10*	1629	59.1	52.8	39.6
At5g38120	*At4CL11*	1653	60.4	47.2	36.2

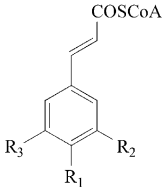
[a] *At4CL1–At4CL11* numbers are based on previous annotations (Costa et al., 2003), with *At4CL1–3*, *5* and *10* proposed as being *4CL*s proper (Shockey et al., 2003).
[b] Including stop codon.

purified to apparent homogeneity (see Section 4), in order to assess whether there were any substantial differences in the enzyme properties of either the native or recombinant forms.

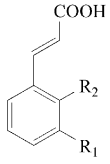
Enzyme assays and determination of kinetic properties for the recombinant His_6-tag At4CL and native At4CL1 proteins were then carried out using six possible phenylpropanoid substrates viz, cinnamic (**1**), *p*-coumaric (**2**), caffeic (**3**), ferulic (**4**), 5-hydroxyferulic (**5**) and sinapic (**6**) acids, with enzyme activities unambiguously determined by HPLC analyses using both UV and MS detection (data not shown). Synthesis of 5-hydroxyferulic acid (**5**) was as described in Kim et al. (2004), whereas the other cinnamic acid derivatives (**1**–**4** and **6**) were from commercial sources (see Section 4). The corresponding potential enzymatic products, cinnamoyl (**7**), *p*-coumaroyl (**8**), caffeoyl (**9**) and feruloyl (**10**) CoAs were generated as for Beuerle and Pichersky (2002), using recombinant tobacco 4CL as enzyme source. By contrast, both 5-hydroxyferuloyl



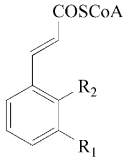
R_1, R_2, R_3 = H, cinnamic acid (**1**)
R_1 = OH, R_2, R_3 = H, *p*-coumaric acid (**2**)
R_1, R_2 = OH, R_3 = H, caffeic acid (**3**)
R_1 = OH, R_2 = OCH_3, R_3 = H, ferulic acid (**4**)
R_1, R_3 = OH, R_2 = OCH_3, 5-hydroxyferulic acid (**5**)
R_1 = OH, R_2, R_3 = OCH_3, sinapic acid (**6**)



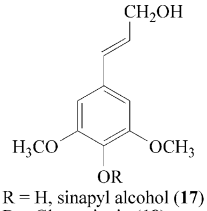
R_1, R_2, R_3 = H, cinnamoyl CoA (**7**)
R_1 = OH, R_2, R_3 = H, *p*-coumaroyl CoA (**8**)
R_1, R_2 = OH, R_3 = H, caffeoyl CoA (**9**)
R_1 = OH, R_2 = OCH_3, R_3 = H, feruloyl CoA (**10**)
R_1, R_3 = OH, R_2 = OCH_3, 5-hydroxyferuloyl CoA (**11**)
R_1 = OH, R_2, R_3 = OCH_3, sinapoyl CoA (**12**)



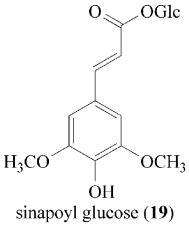
R_1 = OH, R_2 = H, *m*-coumaric acid (**13**)
R_1 = H, R_2 = OH, *o*-coumaric acid (**14**)



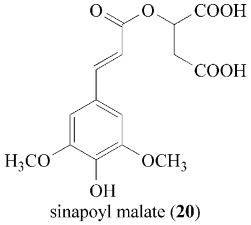
R_1 = OH, R_2 = H, *m*-coumaroyl CoA (**15**)
R_1 = H, R_2 = OH, *o*-coumaroyl CoA (**16**)



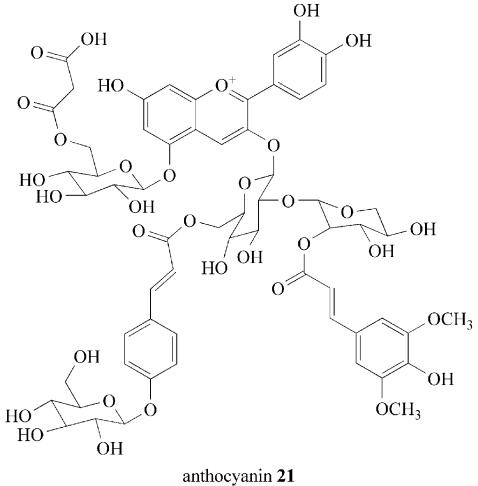
R = H, sinapyl alcohol (**17**)
R = Glc, syringin (**18**)



sinapoyl glucose (**19**)



sinapoyl malate (**20**)



anthocyanin **21**

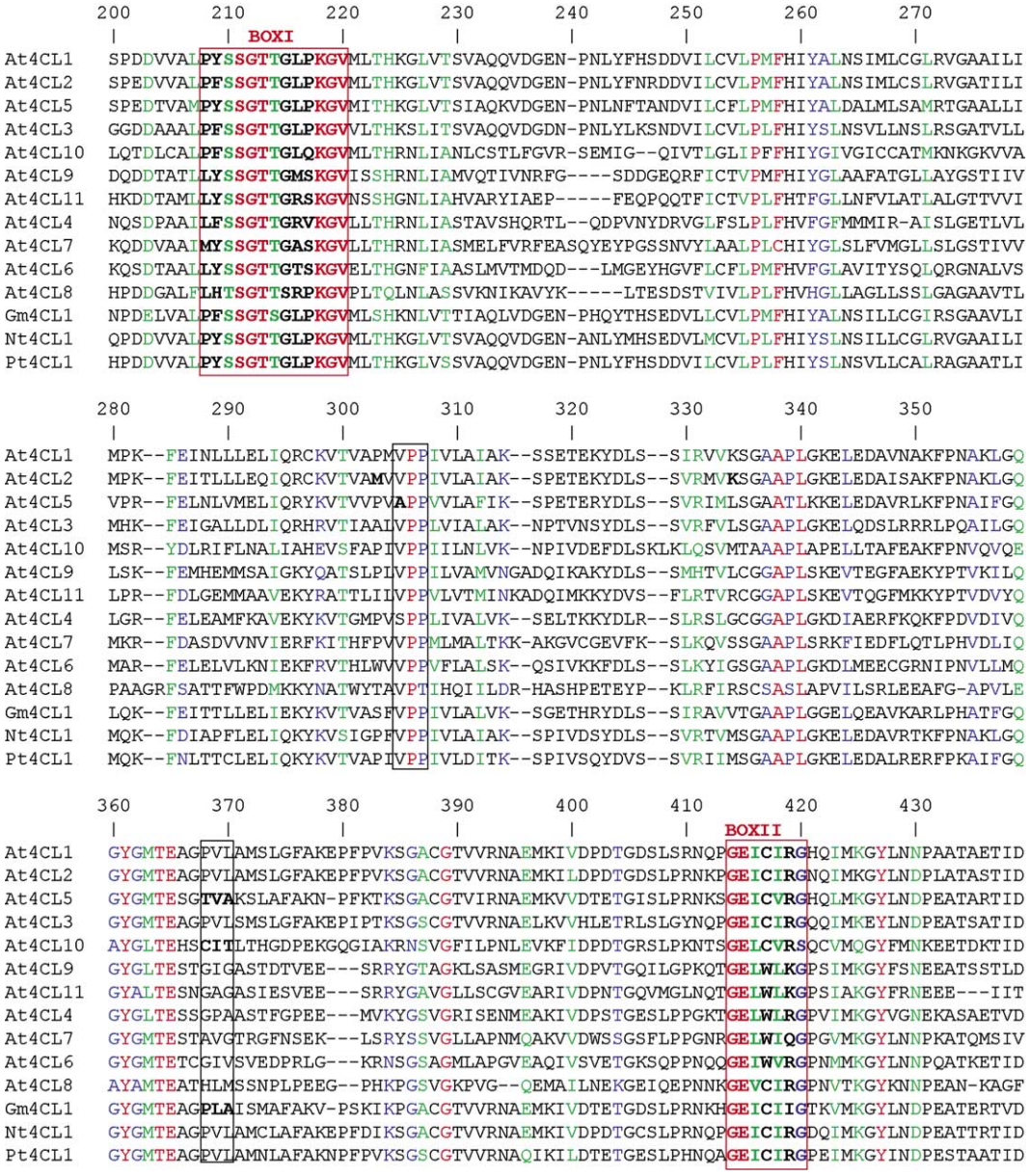


Fig. 1. Amino acid sequence comparison of putative At4CLs with 4CL1 from *Glycine max*, *Nicotiana tabacum*, and *Pinus taeda*. Alignment was performed by http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_clustalw.html using default programme settings. Box I represents the putative AMP binding domain and adjacent conserved amino acid residues, and Box II the conserved *GEICIGR* motif (Stuible et al., 2000). Conserved *VPP* and *PVL* domains with reference to Gm4CL1 (Schneider et al., 2003) are also boxed. Red amino acid residues indicate identity amongst all of the sequences; green residues indicate strong similarity, while blue residues indicate weak similarity.

(**11**) and sinapoyl (**12**) CoA were synthesized via the imidazolide intermediates **23** and **24** (see Scheme 1), a strategy based on the general procedure of Pabsch et al. (1991); purification of derivatives **11** and **12** used the protocol of Beuerle and Pichersky (2002) as before.

Of the 11 purified putative 4CL homologues (At4CL1–11), only At4CL1–3 were biochemically competent to reduce substrates **1**–**5** to afford the corresponding CoA derivatives **7**–**11**, whereas At4CL5 utilized substrates **2**–**6**: At4Cl5 is the sole member able to convert sinapic acid (**6**) into sinapoyl CoA (**12**) in agreement with earlier reports (Schneider et al., 2003; Hamberger and Hahlbrock, 2004). All of the active enzymes, however, displayed Michaelis–Menten kinetics. The remaining proteins (At4CL4, 6–10 and 11) were catalytically unable to convert any of the substrates into the corresponding CoA esters even at protein amounts of up to ~100 μg/assay; extremely weak activities were, however, noted for some at very high protein amounts (>120 μg) (discussed later). [Interestingly, while initial experiments with At4CL10 suggested bona fide catalytic activity for several of the substrates, this could not be reproduced consistently.]

Table 2
Gene specific primers used to amplify putative *Arabidopsis thaliana* *4CL1–11* genes

Genes	Primers
At4CL1	5′-ATGGCGCCACAAGAACAAGCAG-3′ 5′-CAATCCATTTGCTAGTTTTGCCCTC-3′
At4CL2	5′-ATGACGACACAAGATGTGATAG-3′ 5′-GTTCATTAATCCATTTGCTAG-3′
At4CL3	5′-ATGATCACTGCAGCTCTACAC-3′ 5′-ACAAAGCTTAGCTTTGAGGTC-3′
At4CL4	5′-ATGCTGACGAAAACCAACGAC-3′ 5′-AAGTTTTGATGCATTGCCATC-3′
At4CL5	5′-ATGGTGCTCCAACAACAAACGC-3′ 5′-TTTAGAGCACATGGTTTCCAATTTAG-3′
At4CL6	5′-ATGGAGAAATCCGGCTACGGC-3′ 5′-CATCTTGGATCTTACTTGCTGA-3′
At4CL7	5′-ATGGCGGCGACACATCTTCAC-3′ 5′-TAGTCTTGAAGAAACAGAGTTTG-3′
At4CL8	5′-ATGGATAGCGATACTCTCTCAG-3′ 5′-GGGCTTCTCAAGGAAATGTTG-3′
At4CL9	5′-ATGGCTTCAGTGAATTCTCG-3′ 5′-AAGCTTGGAGTTGGAAGTTGCTATC-3′
At4CL10	5′-ATGGAGAGTCAAAAGCAAGAAG-3′ 5′-CTTCTTGTTGATGCTGAGGATC-3′
At4CL11	5′-ATGGCGAATTCTCAAAGATCA-3′ 5′-AATTTTTGAAATGGCAAATTTG-3′

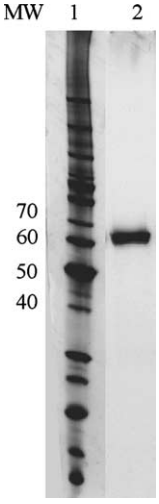


Fig. 2. SDS–PAGE of purified At4CL2. Lane 1, molecular weight markers and lane 2, At4CL2.

For the four catalytically competent family members, At4CL1–3 and 5, pH and temperature optima were determined, using as potential substrates **1**–**6** and coenzyme A (data not shown). For kinetic measurements (described below), a temperature of 30 °C at pH 7.5 was employed for all assays.

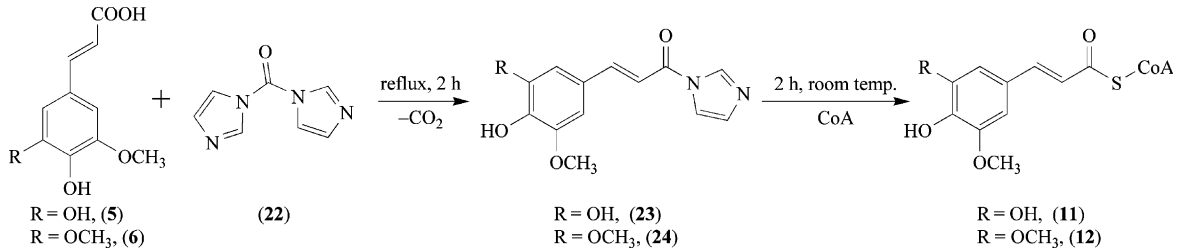
The properties of At4CL1 were first determined using both the His_6-tagged and native forms, and their relative catalytic properties are summarized (Table 3 and Fig. 3). Although the "apparent" k_{enz} value for the native protein is reduced by up to an order of magnitude, relative to the His_6-tagged form, this is presumably not the case. Instead, it reflects significant activity losses accompanied by the much lengthier purification protocol for the native At4CL1 (see Section 4). The overall substrate versatility preferences, however, remained the same with *p*-coumaric (**2**) and caffeic (**3**) acids being the best substrates, followed by ferulic (**4**) and 5-hydroxyferulic (**5**) acids. By contrast, cinnamic acid (**1**) was poorly utilized and sinapic acid (**6**) did not serve as a substrate. Interestingly, the lower k_{enz} with cinnamic acid (**1**) was in part due to a higher K_m (weaker binding), whereas the other substrates **2**–**5** were much more strongly bound.

Somewhat analogous trends in K_m values were reported for substrates **2**–**4** (although, 5-hydroxyferulic acid (**5**) was not examined) by Ehlting et al. (1999) following spectrophotometric assays using crude bacterial extracts containing recombinant At4CL1. However, that study was limited since no detailed kinetic data was reported other than "relative V_{max} values", and thus there is no basis for direct comparison with the present study which uses purified proteins.

Interestingly, 5-hydroxyferulic acid (**5**) was utilized the most effectively with both At4CL1 and At4CL5; for At4CL2 and 3, by contrast, k_{enz} values for 5-hydroxyferulic acid (**5**) were approximately an order of magnitude lower. Thus, with the exception of sinapic acid (**6**) which does not serve as a substrate, At4CL1 was overall the catalytically most efficient protein and, as expected, displayed relatively broad substrate versatility (Table 3 and Fig. 3).

At4CL2 has the second highest homology (87.2% similarity, 83.4% identity, Table 1) to that of At4CL1, but with reduced catalytic efficacy (Table 3 and Fig. 3). As for At4CL1, sinapic acid (**6**) did not serve as a substrate. In our hands, both caffeic (**3**) and *p*-coumaric (**4**) acids served by far as the best substrates (k_{enz} 492,400 and 66,600 M^{-1} s^{-1}), with much reduced catalytic activity being noted for utilization of ferulic (**4**), 5-hydroxyferulic (**5**), and cinnamic (**1**) acids.

Recombinant At4CL2, expressed as a fusion protein either lacking (Ehlting et al., 1999) or containing an N-terminal His_6-tag (Ehlting et al., 1999; Stuible et al., 2000; Stuible and Kombrink, 2001; Schneider et al., 2003) was previously preliminarily characterized, with some – but not all – characteristics in harmony with that of the C-terminal His_6-tagged At4CL2 examined in this study. That is, as for At4CL1, the study by Ehlting et al. (1999) provided only very limited kinetic data, with K_m and "relative V_{max} (as % coumarate)" values reported using bacterial extracts containing recombinant At4CL2. Their data with K_m's for cinnamate (**1**), *p*-coumarate (**2**) and caffeate (**3**) of 6630, 252 and 20 μM, respectively, showed the same general trend as



Scheme 1. Synthetic routes to 5-hydroxyferuloyl CoA (**11**) and sinapoyl CoA (**12**).

Table 3
Kinetic parameters of *Arabidopsis thaliana* 4CL isoforms against various phenylpropanoid substrates

Substrate	K_m (μM)	V_{max} (pkat/μg)	k_{cat} (s^{-1})	k_{enz} (M^{-1} s^{-1})
At4CL1				
Cinnamic (**1**)	135 ± 4.9	9.2 ± 0.10	0.59	4400
p-Coumaric (**2**)	6 ± 0.4	61.3 ± 0.95	3.96	666000
Caffeic (**3**)	2 ± 0.1	42.0 ± 0.25	2.71	1214500
Ferulic (**4**)	16 ± 1.3	39.5 ± 1.10	2.55	154600
5-OH Ferulic (**5**)	21 ± 1.8	10.2 ± 0.31	0.66	31000
Sinapic (**6**)	No conversion			
At4CL1 native				
Cinnamic (**1**)	Not determined			
p-Coumaric (**2**)	6 ± 0.5	8.7 ± 0.14	0.53	91000
Caffeic (**3**)	2 ± 0.2	3.9 ± 0.07	0.24	97400
Ferulic (**4**)	17 ± 1.0	6.9 ± 0.13	0.42	25000
5-OH Ferulic (**5**)	15 ± 2.9	1.1 ± 0.01	0.06	4260
Sinapic (**6**)	No conversion			
At4CL2				
Cinnamic (**1**)	181 ± 4.3	3.2 ± 0.03	0.21	1140
p-Coumaric (**2**)	45 ± 2.4	46.6 ± 1.14	3.00	66600
Caffeic (**3**)	4 ± 0.4	30.4 ± 0.54	1.96	492400
Ferulic (**4**)	686 ± 32.8	9.9 ± 0.22	0.64	930
5-OH Ferulic (**5**)	142 ± 7.9	7.0 ± 0.12	0.45	3200
Sinapic (**6**)	No conversion			
At4CL3				
Cinnamic (**1**)	42 ± 1.8	4.0 ± 0.06	0.26	6200
p-Coumaric (**2**)	4 ± 0.2	12.9 ± 0.16	0.84	227900
Caffeic (**3**)	13 ± 1.0	7.0 ± 0.21	0.45	33600
Ferulic (**4**)	10 ± 0.7	6.4 ± 0.11	0.41	41000
5-OH Ferulic (**5**)	37 ± 2.3	0.9 ± 0.02	0.06	1600
Sinapic (**6**)	No conversion			
At4CL5				
Cinnamic (**1**)	No conversion			
p-Coumaric (**2**)	25 ± 1.7	3.1 ± 0.07	0.20	8000
Caffeic (**3**)	7 ± 0.9	2.9 ± 0.09	0.19	26000
Ferulic (**4**)	6 ± 0.9	0.9 ± 0.03	0.06	9300
5-OH Ferulic (**5**)	6 ± 0.8	2.5 ± 0.05	0.17	26500
Sinapic (**6**)	10 ± 1.1	1.1 ± 0.04	0.08	7900

in this study (181, 45, and 4 μM). One significant difference in the present study is that ferulic acid (**4**) also served as a substrate, and 5-hydroxyferulic acid (**5**) was converted into its corresponding CoA ester (**11**) as well. By contrast, in the study by Ehlting et al. (1999), spectrophotometric assays failed to detect any binding of ferulate (**4**) or its conversion into feruloyl CoA (**10**). In the current study, however, enzymatically formed feruloyl CoA (**10**) was isolated and unambiguously identified on the basis of one and two dimensional NMR, mass and UV spectroscopic analyses (see Section 4). Surprisingly, in neither study (Ehlting et al., 1999; Schneider et al., 2003) was 5-hydroxyferulic acid (**5**) evaluated as a potential substrate.

In spite of the lower catalytic efficiency of At4CL2 (compared to At4CL1), several studies have been carried out with it, particularly as regards mutational effects on catalytic activity and substrate versatility

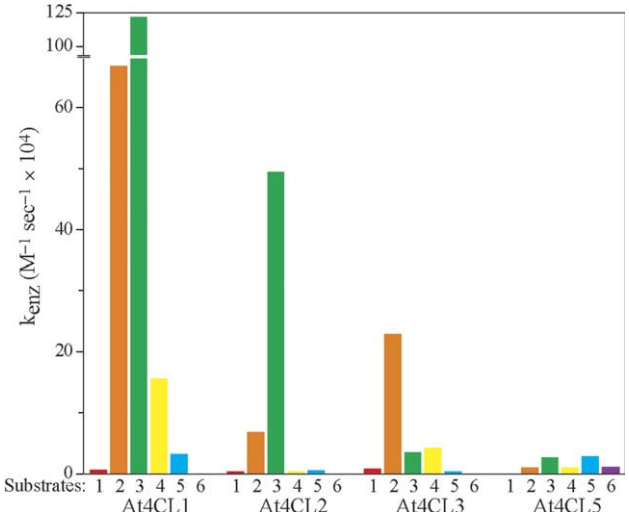


Fig. 3. Relative substrate efficacies of *Arabidopsis* 4CL homologues, At4CL1–3 and 5. Substrates **1**–**6** structures are shown.

(Stuible et al., 2000; Stuible and Kombrink, 2001; Schneider et al., 2003). In those studies, however, only a limited repertoire of potential substrates, viz, caffeic (**3**), ferulic (**4**) and *p*-coumaric (**2**) acids, was again used in the characterization of the N-terminal His_6-tagged At4CL2. Yet, several notable differences between these studies and the present investigation were noted. Thus, in the initial study by Stuible et al. (2000), a K_m for caffeic acid (**3**) of 24 μM was reported, this being in good agreement with the value reported herein (17 μM); however, no V_{max} was reported. In a subsequent study (Stuible and Kombrink, 2001), both caffeic (**3**) and ferulic (**4**) acids were now examined as test substrates, with a V_{max} of 0.158 pkat/μg being reported for caffeic acid (**3**), this differing by a factor of nearly 200 from our own experimental data (30.4 pkat/μg, Table 3). However, the same investigators next reported additional kinetic data for the N-terminal His_6-tagged At4CL2 in a subsequent paper (Schneider et al., 2003), but now with a V_{max} of 236 ± 30 pkat/μg protein for caffeic acid (**3**), rather than the 0.158 pkat/μg protein previously described (Stuible and Kombrink, 2001). No explanation for this apparent ~500-fold increase in V_{max} was given, which, if correct, would differ substantially from the measurements made in this study using the C-terminal His_6-tagged At4CL2.

More comparable data reported by Schneider et al. (2003) with those herein are the values of K_m (233 ± 15 μM) and V_{max} (475 ± 94 pkat/μg protein) for *p*-coumaric acid (**2**). The K_m (45 μM) obtained in the present study is somewhat comparable, as is our experimentally determined V_{max} (46.6 pkat/μg protein). Thus, we have proceeded on the assumption that the study by Stuible and Kombrink (2001) presumably contains the actual kinetic data, which would be in harmony with the measurements made herein. Whatever the reasons for the changing values of V_{max} in the three studies of the Kombrink group (Stuible et al., 2000; Stuible and Kombrink, 2001; Schneider et al., 2003), the impact on our own investigation should be minimal. This is because the study described herein involves the direct comparison of all C-terminal His_6-tagged At4CL isoforms, and thus effects on catalytic activity, if any, introduced by the His_6 tag should be normalized. Lastly, another distinguishing feature in the present study from that of Stuible and Kombrink (2001) and Schneider et al. (2003) involves their absence of detectable binding and conversion of ferulic acid (**4**) by At4CL2. In our hands, by contrast, a K_m of 686 μM and k_{enz} of 930 M^{-1} s^{-1} from ferulic acid (**4**) were obtained. One possible explanation is that the spectrophotometric assays being used (Stuible and Kombrink, 2001; Schneider et al., 2003) do not permit unambiguous identification of products being generated, in contrast to the HPLC studies with UV and MS detection utilized in this investigation.

For At4CL3, which also displayed broad substrate versatility (Table 3 and Fig. 3), the experimentally determined k_{enz} for *p*-coumaric (**2**), caffeic (**3**), and 5-hydroxyferulic (**5**) acids of 227,900, 33,600 and 41,000 M^{-1} s^{-1} were significantly higher than that for either cinnamic (**1**), or 5-hydroxyferulic (**5**) acids (6200 and 1600 M^{-1} s^{-1}, respectively). Although the kinetic data (K_m's) are substantially different in this study from that of Ehlting et al. (1999), who preliminarily characterized bacterial extracts containing At4CL3, the overall trend in relative catalytic activity was the same. However, since **5** was not evaluated in the Ehlting et al. (1999) study, and no detailed kinetic characterization was carried out, the earlier overall interpretations were limited by the lack of additional data as now generated in the present study.

At4CL5 has the third closest homology to At4CL1 with 74.5% similarity and 76.4% identity (Table 1). It also displayed quite broad substrate versatility with the various substrates evaluated, but differed from At4CL1–3 in an inability to ligate cinnamic acid (**1**) and an ability to ligate sinapic acid (**6**) (Table 3 and Fig. 3). However, the best substrates were caffeic (**3**) and 5-hydroxyferulic (**5**) acids, k_{enz} of 26,000 and 26,500 M^{-1} s^{-1}, with these k_{enz} values being some 3-fold higher than that for either *p*-coumaric (**2**), ferulic (**4**) or sinapic (**6**) acids. Overall, it has the broadest substrate versatility for these phenolic analogues in being able to employ sinapic acid (**6**); the possible physiological significance of the latter observation is discussed later (Sections 2.4 and 2.6).

For needed context, the first report of a gene encoding a sinapic acid (**6**) activating 4CL was from a study of a four-membered *4CL* gene family in soybean (*Glycine max* L) (Lindermayr et al., 2002). In that investigation, preliminary kinetic data for all four 4CL isoforms were determined, with Gm4CL1 being observed to bind *p*-coumaric (**2**), caffeic (**3**), ferulic (**4**) and sinapic (**6**) acids quite strongly (K_m's ranging from 8 to 33 μM). Once again,

however, 5-hydroxyferulic acid (**5**) was not used as a substrate in that study. Preliminary data (relative V_{max} as % coumarate consumed) with the above other substrates suggested that *p*-coumaric acid (**2**) was the best substrate. The soybean Gm4CL1, however, only shares some 74.0% similarity and 57.6% identity to that of At4CL5, thereby making homology comparisons as regards possible catalytic function of the putative At4CLs tenuous.

By contrast, there are interesting findings in our own study with At4CL5, which include comparable K_m's (6–25 μM) for all substrates to that of Lindermayr et al. (2002). However, in our hands, there was no indication of any preference for *p*-coumaric acid (**2**) over the potential substrates **3**, **4** and **6** as suggested from the soybean Gm4CL1 data of Lindermayr et al. (2002). It is also noteworthy that since their initial report of a 4CL able to activate sinapic acid (**6**) in soybean, there have been several descriptions of a sinapic acid (**6**) activating enzyme in *Arabidopsis*, i.e. in summary form to the *Arabidopsis* community (Lewis, Davin and Franceschi, June 2003, NSF MCB-0117260 report), and in papers by Schneider et al. (2003) and Hamberger and Hahlbrock (2004). Additionally, the limited kinetic data provided by Hamberger and Hahlbrock (2004) for At4CL5 (also linked via a C-terminal His_6-tag), as well as that of Schneider et al. (2003), further underscored the necessity for comprehensive detailed analyses of gene families, rather than (as historically has been done) piece-meal gene-by-gene analyses using a limited repertoire of potential substrates.

In an analogous situation as for At4CL2, there are substantial differences in the kinetic characterization of At4CL5 in our study, relative to both the findings of Hamberger and Hahlbrock (2004) and Schneider et al. (2003). Firstly, neither of the latter two studies used 5-hydroxyferulic acid (**5**) as substrate, whose omission missed the opportunity to establish that it is a preferred substrate for catalysis. Secondly, the kinetic properties in Hamberger and Hahlbrock (2004) give "specific activity" values rather than V_{max} proper. That is, "specific activities" of 100, 187, 153 and 105 pkat/μg were reported for *p*-coumaric (**2**), caffeic (**3**), ferulic (**4**) and sinapic (**6**) acids, thereby indirectly suggesting caffeic acid (**3**) as the best substrate for At4CL5. Our own findings, however, differ in terms of both caffeic (**3**) and 5-hydroxyferulic (**5**) acids serving as the best substrates. Furthermore, while the K_m values for sinapic acid (**6**) in this study, and that of Schneider et al. (2003), are comparable (23 vs. 10 ± 1.1 μM), their reported V_{max} of 83 pkat/μg again differs from our own by circa 75-fold. Whether this is a consequence of differences in catalytic properties of At4CL5 with either N-terminal or C-terminal His_6-tag attachments, or in the V_{max} determinations themselves, cannot be gauged at this time.

The proteins encoded by the remaining candidate *4CL* genes (*At4CL4*, *6–10* and *11*), respectively, were expressed in C-terminal His_6-tagged recombinant forms and purified as before. However, none of these displayed any detectable catalytic activity when protein amounts (up to 100 μg) were assayed. It is noteworthy though, albeit of no apparent physiological significance, that several of these proteins were marginally able to activate certain hydroxycinnamic acids, but only at protein levels (>120 μg) per assay. That is, At4CL4 was able to poorly activate caffeic (**3**), ferulic (**4**) and 5-hydroxyferulic (**5**) acids, whereas At4CL7 marginally utilized caffeic (**3**) and ferulic (**4**) acids. Additionally, At4CL9 was able to activate *p*-coumaric acid (**2**), whereas At4CL11 engendered formation of barely detectable levels of cinnamoyl (**7**) and *p*-coumaroyl (**8**) CoA when incubated in the presence of cinnamic (**1**) and *p*-coumaric (**2**) acids. [As indicated earlier, At4CL10 gave some initially positive assays, but these could not be reproduced consistently.] By contrast, the remainder (At4CL6 and 8) were inactive with any of the substrates utilized, even at such high protein levels. These data are included for comparative purposes only. It is concluded, however, that the data are of no physiological significance, and that the proteins have distinct physiological roles which await discovery, rather than (hydroxy)cinnamic acid activation. Indeed, recently, it has been proposed that At4CL4 and 6 may be involved in the biosynthetic pathway to jasmonic acid (Schneider et al., 2005).

Thus, this phase of the study demonstrated that there are four genes which encode catalytically active At4CLs, all of which show rather broad but distinct substrate versatilities. Ongoing work on the X-ray crystal structural determinations of the various At4CL isoforms is now in progress, and this should provide a definitive explanation as to the differences in substrate versatilities encountered, and thus add to our developing knowledge of key 4CL residues required for binding and catalysis (Stuible et al., 2000; Stuible and Kombrink, 2001; Lindermayr et al., 2003) (see Fig. 3). Such X-ray studies have been completed for pinoresinol–lariciresinol reductase (Min et al., 2003), phenylcoumaran benzylic ether reductase (Min et al., 2003), secoisolariciresinol dehydrogenase (Youn et al., 2005), as well as in the modeling of both isoflavone reductase (Min et al., 2003) and cinnamyl alcohol dehydrogenase (Kim et al., 2004).

2.4. Phylogenetic analysis

Two approaches were undertaken for the phylogenetic analysis of the putative *At4CL* gene family annotated as such in The Institute for Genomic Research (TIGR) database. The first approach employed was to analyze the large superfamily of acyl activating enzymes (Shockey et al., 2003), which compares protein sequence alignments using the CLUSTAL_X program (Thompson et al., 1997) with default gap creation and gap extension penalty scores. The subsequent phylogenetic tree was then drawn from the deduced alignments and

employed the TreeView program (Page, 1996). Using this approach, seven clades (I–VII) were segregated (Shockey et al., 2003), with clade IV grouping only At4CL1–3, 5 and 10 together, but of which only At4CL1–3 and 5 had 4CL activity. Clade V, by comparison, grouped together seven of the annotated putative 4CLs (At4CL6, 7, 9, 11–14, including At4CL12–14 which had been eliminated earlier) for consideration as bona fide 4CLs. Clade VII, on the other hand, contained three genes including the putative At4CL8 homologue. However, from this study, it is now confirmed that none of those annotated as putative At4CLs in Clades V and VII correspond to *4CL* gene family members.

In an analogous manner, comparison of the derived amino acid sequences of the putative *At4CL1–11* gene family with that of three known 4CLs, i.e. from *Glycine max* (Gm4CL1) (Lindermayr et al., 2002), tobacco (Nt4CL1, *Nicotiana tabacum*) (Lee and Douglas, 1996) and loblolly pine (Pt4CL1, *Pinus taeda*) (Voo et al., 1995) generated similar results. That is, using both the CLUSTAL W program (Thompson et al., 1994) with default settings, and JALVIEW (http://www.ebi.ac.uk/jalview) multiple alignment editor V1.8 (Cuff et al., 2000), it was found that At4CL1–3, 5 and 10 grouped together with Gm4CL1, Pt4CL1 and Nt4CL1. As before At4CL4, 6, 7, 9 and 11 formed a distinct second group; by contrast, At4CL8 was a completely different outgroup with maximum divergence (Fig. 4).

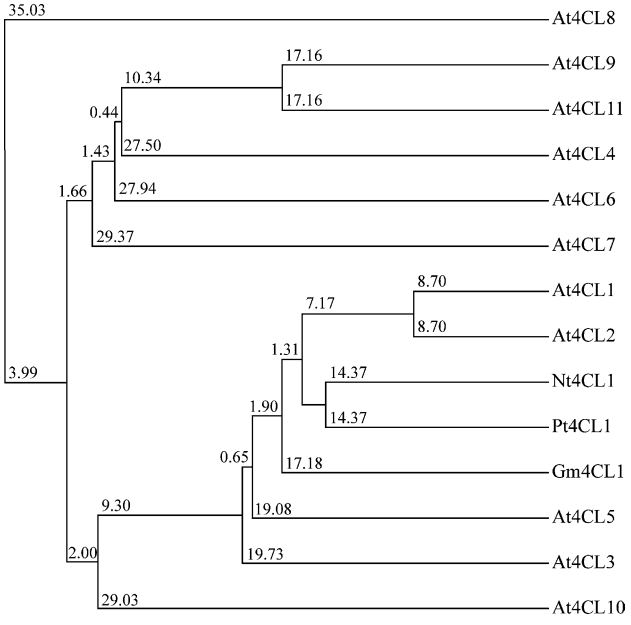


Fig. 4. Phylogenetic analysis of At4CL proteins with soybean, pine and tobacco 4CL1. Phylogenetic tree was constructed based on the multiple alignment generated from http://www.ebi.ac.uk/cgi-bin/clustalw, version 1.82 and using http://www.ebi.ac.uk/jalview/. The tree was generated by average distance using percentage identity value (PID). The numbers indicate the branch lengths that are the percentage mismatches between two nodes.

2.5. Expression patterns of At4CL genes

Quantitative real-time PCR, using gene specific primers for each *At4CL*, was conducted to study the overall relative expression profile of each of the *4CL* genes, except *At4CL3* which could not be analyzed since no primer sets specific to *At4CL3* could be designed either with the SPAD program or manually (see Section 4). RNA was isolated from the *Arabidopsis* ecotype Columbia seedlings (grown 5 and 14 days on agar plates, see Section 4), as well as from the basal stems (~1.5 cm) of 4- and 6-week old plants.

Relative to the other genes, *At4CL1*, *6*, *8*, and *9* transcript accumulation was generally highest in the 5- and 14-day old seedlings as well as in the basal stem tissues at 4 and 6 weeks. Overall, *At4CL8* had the highest relative value (RE ~ 1.5, see Section 4) at 5 and 10 days with respect to the *actin 2* (*ACT2*) gene (see Fig. 5(a)), although the function of At4CL8 is as yet unknown. These expression patterns were in good agreement with the EST database information reported in Costa et al. (2003). The *At4CL1* gene was consistently highly expressed in all samples examined, with its expression being ~2-fold higher in 4- and 6-week old stem tissues compared to young seedlings, in agreement with Raes et al. (2003). On the other hand, *At4CL10* and *At4CL11* did not show any detectable transcript accumulation under the conditions studied, even when 10 times more cDNA template were used (data not shown).

2.6. Gene knockout studies

The ability of At4CL5 to ligate sinapic acid (**6**) had potential ramifications for syringyl lignin (sinapyl alcohol **17**-derived) and sinapate/sinapyl alcohol derivative biosynthesis. In this context, an earlier preliminary report by Yamauchi et al. (2003) indicated that in *Arabidopsis* sinapic acid (**6**) is neither ligated by 4CL nor is it converted into syringyl lignin.

We, therefore, chose to investigate this question more thoroughly by examining the effects of "knocking-out" *At4CL5*. In this respect, two potential "knockout" lines were available, one from the Salk T-DNA insertion line collection (SALK_057020) and a second from the Cold Spring Harbor Laboratory collection (ET6180). The SALK_057020 mutant in the ecotype Columbia contains a T-DNA insertion 200 bp upstream of the translation start codon (ATG), while the ET6180 mutant in the ecotype Landsberg (*L*er) contains an enhancer trap insertion between the first and second exons of the *At4CL5* gene. Homozygous lines for SALK_057020 were thus selected by PCR using left and right border gene specific primers, whereas for ET6180 these were selected on naphthalene acetamide (NAM) and kanamycin, with the corresponding seeds for each harvested (see Section 4). Quantitative real time PCR analyses,

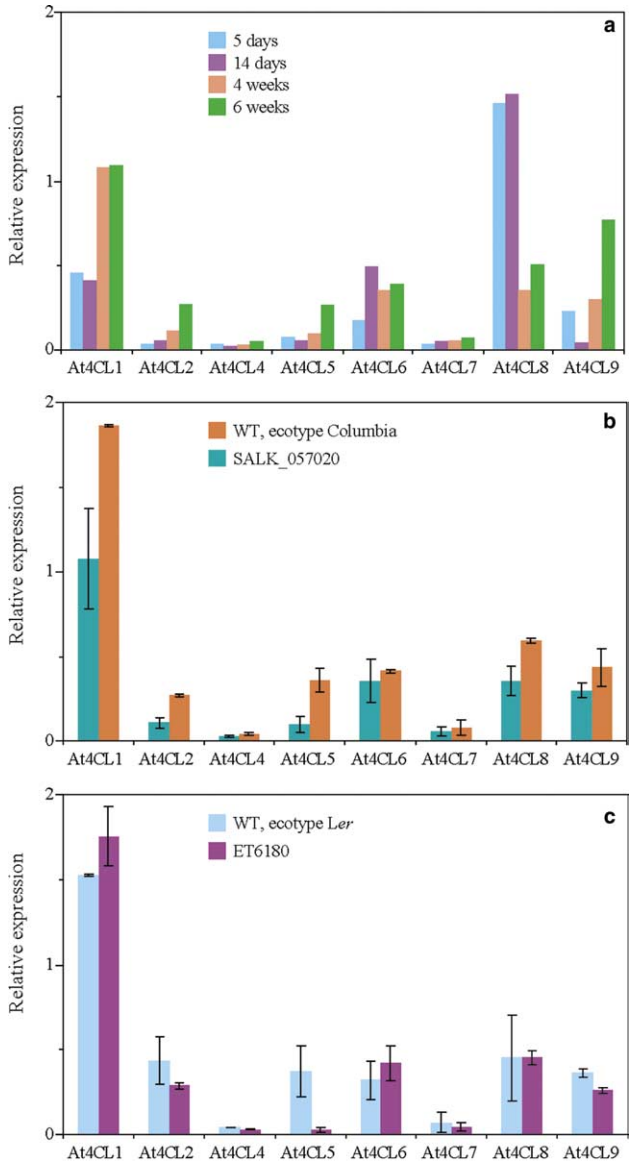


Fig. 5. Expression patterns of bona fide and putative *4CL* genes in *Arabidopsis*: (a) using total RNA isolated from 5- and 14-day old seedlings as well as 4- and 6-week old basal stems. (b,c) Transcript profile of the *At4CL* genes in 4-week old basal stems of wild type and *4cl5* knockout *Arabidopsis* mutants. (b) Ecotype Columbia and SALK_057020 mutant; (c) ecotype *L*er and Cold Spring Harbor Laboratory mutant, ET6180. Relative accumulation of the *At4CL* transcripts was determined by quantitative real-time PCR, with the results normalized with respect to the *ACT2* transcript levels (RE = 1, see Section 4).

lignin content and monomer composition estimations, as well as sinapate/sinapyl alcohol derivative contents were next conducted on these selected lines.

Quantitative analysis of expression for the *At4CL* genes using real time RT-PCR with a primer set specific to each of the *At4CL* mRNA's (see Table 4) was next performed in order to gauge to what extent the expression levels of each of the genes were affected in the *4cl5* mutants. However, even though the insertion of the T-DNA in the SALK_057020 mutant was confirmed, we were unable to detect any decrease of *At4CL5* gene transcript accumulation in 4-week old basal stems of the SALK_057020 mutant (Fig. 5(b)). In fact, the opposite effect was observed not only for *At4CL5* but also for all other *At4CL* genes detected. On the other hand, the enhancer trap insertion effectively disrupted transcript accumulation of *At4CL5* (>90%) in the ET6180 mutant compared with levels in the wild type Landsberg plants (Fig. 5(c)). By contrast, disruption of the *At4CL5* gene function in the ET6180 mutant had essentially no effect on expression of the other bona fide *At4CL* genes tested.

Homozygous lines for the two *4cl5* lines were next grown until maturity, with groups of plants (circa 20) harvested on a weekly basis from 4 to 10 weeks after seeding. For each weekly harvesting stage, the bolting stems from each plant line were individually freeze-dried, ball milled, and solvent extracted to remove any readily soluble materials. The resulting extractive-free stem material for each line was then subjected to acetyl bromide determinations for lignin contents (Iiyama and Wallis, 1990), as well as thioacidolysis (Lapierre et al., 1986; Rolando et al., 1992) and nitrobenzene oxidation (Iiyama and Lam, 1990) (data not shown) analyses, in order to estimate lignin monomer composition (1990) as described in Blee et al. (2001) and Kim et al. (2004).

The two homozygous (*4cl5*) and WT *Arabidopsis* (Columbia and *L*er) plant lines were initially analyzed at weekly intervals to measure both stem lengths and widths (data not shown); however, none of these lines differed in any significant way, showing for all intents and purposes comparable trends in stem growth and development. Note that while the plant stems reached maximum length and width by 6 weeks, all growth and development analyses were carried out weekly until week 10 in order to identify any effects occurring during maturation/senescence.

Fig. 6 plots the total acetyl bromide lignin contents (Fig. 6(a) and (b)) and lignin monomeric composition (% CWR), as represented by syringyl/guaiacyl (S/G) ratios (Fig. 6(c) and (d)), respectively. The data as plotted reveal no differences in lignin deposition from weeks 4 to 10 for the SALK_057020 and ET6180 mutants. That is, there is no notable effect on either lignin deposition or on syringyl monomer content.

Additionally, we examined whether there was any deleterious effect on either syringin (**18**) sinapoyl glucose (**19**), sinapoyl malate (**20**) and the anthocyanin (**21**) (Bloor and Abrahams, 2002) amounts. Accordingly, *Arabidopsis* WT (ecotype Columbia and *L*er) and *4cl5* knockout (SALK_057020 and ET6180) stem and leaf samples were individually examined at weekly intervals, but there was no effect on overall amounts of any of these components (data not shown).

Table 4
Primers used for quantitative real time PCR analysis of knockout mutants and wild type *Arabidopsis*

Gene	Primer	Length of amplified fragment	Position
At4CL1	5′-CATATCTACGCTTTGAACTC-3′ 5′-ACCAGATTTCACCACTCTTATC-3′	216	Exon I
At4CL2	5′-GTTGTTCGTTTTGCAGTCG-3′ 5′-GTAAGCAGTTTGAAACAGAGATGC-3′	280	3′-end
At4CL4	5′-TGGATGGCAATGCATCAAAACT-3′ 5′-GGTGACGAATGACGATGCTATAA-3′	196	3′-end
At4CL5	5′-AAGGATCTGCGAGCTAAATTGG-3′ 5′-TCTGAGGGTAGCTTAGCTTTAGGTT-3′	199	3′-end
At4CL6	5′-TTCATCAGCCTTGTTCCA-3′ 5′-CTCTTTTGTGGGTTGTGTAA-3′	183	3′-end
At4CL7	5′-GTGAGGTTTGAAGCTTCTCAGTA-3′ 5′-ATGAGGAAGAGTCTGAAGGA-3′	345	Exon I
At4CL8	5′-TTCATCACTGATAACCTCCCCAAAA-3′ 5′-CCAAGAACCAATAATAGAGAACGC-3′	185	3′-end
At4CL9	5′-CATCAAGATAGCAACTTCCA-3′ 5′-TAGTGGCCATTTATGTGTG-3′	175	3′-end
ACT2	5′-CTGGGGTTTTATGAATGGGATCAA-3′ 5′-CTCAGAGCTACAAAACAATGGGA-3′	203	3′-end

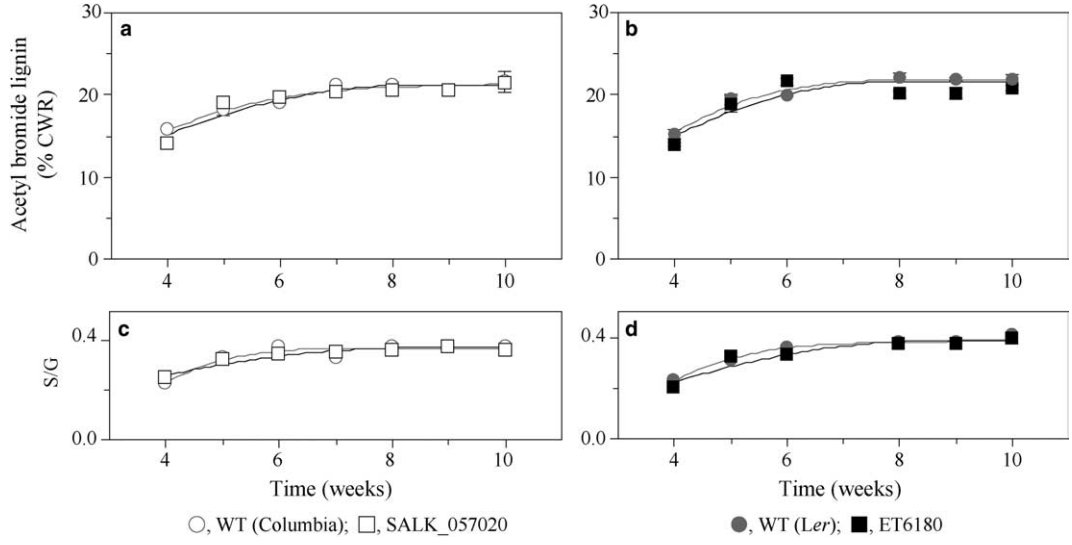


Fig. 6. Lignin deposition patterns (a,b) and syringyl (S)/guaiacyl (G) ratios (c,d) in stem sections of wild type and knockout *Arabidopsis* plant lines at different stages of growth and development (4–10 weeks).

Thus, in summary, disruption of the *At4CL5* gene (>90%) had no effect on either sinapate/sinapyl alcohol derivative accumulation and/or syringyl lignin contents.

The most likely possible interpretation to these findings is that At4CL5 has no specific role in sinapic acid (**6**) ligation, and hence is not directly involved in the formation of either sinapate/sinapyl alcohol derivatives and/or syringyl lignin deposition. This will only be unambiguously clarified when the comprehensive expression patterns of the bona fide *At4CL* genes are delineated at all stages of *Arabidopsis* growth and development. It appears to be significant, however, that all four bona fide At4CL isoforms were able to differentially ligate 5-hydroxyferulic acid (**5**); this, in turn, strongly suggests that sinapyl alcohol (**17**), syringin (**18**) and the sinapate esters **19**, **20** are mainly, if not exclusively, derived from 5-hydroxyferuloyl CoA (**11**) and/or derivatives thereof.

3. Conclusions

Comprehensive characterization of the annotated *Arabidopsis* multigene family revealed that there are

only four bona fide At4CL's. Of these, the catalytic properties suggested that syringyl lignin and sinapate/sinapyl alcohol derivative formation occurs via ligation of 5-hydroxyferulic acid (**5**) or reduced derivatives thereof rather than from sinapic acid (**6**) directly. These provisional conclusions were further supported by analysis of an *4cl5* knockout which revealed essentially no effect on (syringyl) lignin content nor on amounts of sinapate/sinapyl alcohol derivatives (**18**–**20**). Future work will investigate the overall patterns of gene expression for each of the putative *At4CL* genes in order to gauge to what extent there is overlapping expression of the bona fide *At4CL* genes.

4. Experimental

4.1. General

4.1.1. Instrumentation

PCR amplifications were carried out using a PTC-0220 DNA Engine Dyad Peltier Thermal Cycler (MJ Research, Waltham, MA, USA) whereas quantitative real time PCR analyses employed a Mx4000 (Stratagene) (Anterola et al., 2002). Clones harboring *At4CL* genes were sequenced on an ABI PRISM 377 (Applied Biosystems, Foster City, CA, USA) automated DNA sequencer using Big-Dye™ terminator technology.

Metal-chelate affinity chromatography was performed using a BioCAD 700E work station with perfusion chromatography (Applied Biosystems) employing a POROS® 20 Metal-chelate (Applied Biosystems) column (100 mm × 4.6 mm); all chromatography steps for purification of native At4CL1 were carried out on a FPLC (Amersham). Reversed-phase HPLC analyses were carried out on an Alliance 2695 HPLC system (Waters, Milford, MA, USA) using a Novapak C_{18} column (Waters, 150 × 3.9 mm). LC-MS was performed on a Finnigan MAT (San Jose, CA, USA) ion trap mass spectrometer equipped with a Finnigan electrospray.

4.1.2. NMR spectroscopy

One dimensional ^{1}H, ^{13}C and two dimensional ^{1}H–^{13}C HSQC (heteronuclear single quantum correlation) and HMBC (heteronuclear multiple bond correlation) spectra were collected on a Varian Inova 500 MHz spectrometer operating at 499.85 MHz for ^{1}H and 125.67 MHz for ^{13}C, respectively. One dimensional ^{31}P spectra (121.5 MHz) and single frequency ^{31}P decoupled ^{1}H spectra (300.14 MHz) were acquired on a Varian Mercury 300. ^{1}H proton spectra obtained in D_2O were referenced to the DSS (sodium 2,2-dimethyl-2-silapentane-5-sulfonate) proton signal at 0.0 ppm, whereas the ^{13}C dimension of the HSQC spectra was referenced indirectly to the DSS signal via the method described by Wishart et al. (1995). ^{31}P spectra were referenced to 85% phosphoric acid (neat) at 0 ppm. Gradient enhanced phase-sensitive HSQC spectra were obtained using the standard Varian pulse sequence at 22 °C, with HSQC spectra collected with sweep widths (acquisition times) of 5182 Hz (198 ms) in t2 (^{1}H) and 10,055 Hz (19.9 ms), 200 × 2 hypercomplex increments in t1. Data were processed in F2 by applying a Gaussian function with a 0.101 s time constant prior to Fourier transformation. The F1 processed data utilized a linear prediction of the original 200 real points, in the case of data sets acquired for 19.9 ms in t1, to 512 points, apodizing with a Gaussian function with a 0.022 s time constant followed by zero filling to 2K complex points and Fourier transformation. Gradient enhanced phase-sensitive HMBC spectra were obtained using the standard Varian pulse sequence at 22 °C, with sweep widths (acquisition times) of 5182 Hz (198 ms) in t2 (^{1}H) and 25,388 Hz (6.3 ms), 160 increments in t1. Data were processed in F2 by applying a pseudo-echo function with −12.1 Hz of exponential line broadening followed by a Gaussian function with a 0.067 s time constant prior to Fourier transformation. The F1 processed data utilized a linear prediction of the original 160 real points in t1, to 512 points, apodizing with a sinebell function shifted by 30 degrees followed by zero filling to 2K complex points and Fourier transformation. ^{31}P decoupled ^{1}H spectra were obtained by single frequency decoupling of the individual ^{31}P peaks using a decoupling B_1 field of approximately 55 Hz. Care was taken to avoid unintended saturation of the neighboring ^{31}P resonance and the ^{1}H spectrum was examined to see which peaks had undergone ^{31}P decoupling.

4.1.3. Materials

All solvents and chemicals used were either reagent or high performance liquid chromatography (HPLC) grade unless otherwise specified. Cinnamic (**1**), *p*-coumaric (**2**), caffeic (**3**), ferulic (**4**) and sinapic (**6**) acids, as well as coenzyme A, were purchased from Sigma (St. Louis, MO, USA). Affi-Gel Blue gel and CHT ceramic hydroxyapatite were obtained from BioRad (Hercules, CA, USA), whereas PD10 columns were from Amersham Biosciences (Piscataway, NJ, USA).

The expression vector pCRT7-CT TOPO®, *E. coli* TOP10 F′ and BL21 STAR™ (DE3) cells, SuperScript™ III First-Strand Synthesis System for RT-PCR, 1-kb DNA molecular weight and Benchmark™ protein ladders, Low Melting Point (LMP) Agarose and Zeocin™ antibiotic were purchased from Invitrogen (Carlsbad, CA, USA). The RNeasy® Plant Mini Kit, QIAquick® Gel Extraction Kit, and QIAprep® Spin Miniprep Kits were from Qiagen Inc. (Valencia, CA, USA), whereas the *Pfu Turbo*® DNA Polymerase was purchased from Stratagene (La Jolla, CA, USA). AgarACE® Agarose-Digesting Enzyme and Wizard® *Plus* SV Miniprep DNA Purification System were purchased from

Promega (Madison, WI, USA). BugBuster® Protein Extraction Reagent, Benzonase® Nuclease, rLysozyme™ Solution, and pET24d expression vector were obtained from Novagen (Madison, WI, USA). Oligonucleotide primers for sequencing, polymerase chain reactions (PCR) and Reverse Transcriptase (RT) mediated PCR were synthesized by Invitrogen. Discontinuous SDS–PAGE was performed using 12% pre-cast minigels (BioRad, Hercules, CA, USA).

4.2. Plant materials

Wild type (WT) ecotype Columbia and Landsberg (*L*er) as well as the knockout mutant *A. thaliana* plant lines were grown in Washington State University greenhouse facilities as described in Patten et al. (2005).

Transposon and T-DNA insertion lines (T_3 seeds) were obtained from the Cold Spring Harbor Laboratory (ET6180; http://genetrap.cshl.org) and the Arabidopsis Biological Resource Center (SALK_057020; ABRC; http://www.arabidopsis.org/abrc/) at Ohio State University (Columbus, Ohio), respectively. Homozygous plant lines for the SALK_057020 knockout were screened by PCR amplification using a REDExtract-N-AMP plant PCR kit (Sigma, St. Louis, MO, USA) according to the manufacturer's instructions. The primers, designed using the SIGnAL web site (http://signal.salk.edu/isects.html), were LP, 5′-TCCTACCAAACAAACAAACGAAC-3′; RP, 5′-CACCGTCGATGATACAGGTTG-3′; LBa1, 5′-TGGTTCACGTAGTGGGCCATCG-3′. PCR, performed using two different primer sets, a three-primer set (LP, RP and LBa1) and a two-primer set (LP and RP), for confirmation of the homozygous lines, was conducted by using a touch-down method as follows: initial denaturation (94 °C, 5 min), 10 cycles of denaturation (94 °C, 1 min), annealing (62 to 52 °C, 30 s) with the temperature decreasing one degree each cycle and extension (72 °C, 1 min) followed by 25 cycles of denaturation (94 °C, 1 min), annealing (52 °C, 30 s), extension (72 °C, 1 min), and final extension (72 °C, 10 min). Amplified products were sequenced and analyzed to confirm the insertion position. The ET6180 homozygous line was selected on a Murashige and Skoog medium containing naphthalene acetamide (NAM, 0.65 μg ml^{-1}) and kanamycin (50 μg ml^{-1}). Individual selected homozygous lines were reconfirmed by using the same procedures described above before use for mRNA expression and lignin analyses.

4.3. Enzymatic and chemical syntheses

E-5-Hydroxyferulic acid (**5**) was synthesized as described (Kim et al., 2004), whereas cinnamoyl (**7**), *p*-coumaroyl (**8**), caffeoyl (**9**) and feruloyl (**10**) CoAs were enzymatically generated using recombinant tobacco 4CL and further purified using the procedure of Beuerle and Pichersky (2002).

Sinapoyl CoA (**12**) was synthesized as follows: sinapic acid (**6**) (1.68 g, 7.49 mmol) and carbonyldiimidazolide (CDI, **22**; 2.43 g, 14.98 mmol) were converted into crude sinapic acid imidazolide (**24**; 1.33 g, 4.8 mmol, 65%) exactly as described (Pabsch et al., 1991) (Scheme 1) and used without further purification. Compound **24** (10 mg, 0.036 mmol) in dry THF (5 ml) and coenzyme A (15 mg, 0.019 mmol) in distilled H_2O (~10 ml) were next converted into crude sinapoyl CoA ester (**12**) (Pabsch et al., 1991), which was subsequently purified as described by Beuerle and Pichersky (2002) to give pure sinapoyl CoA (**12**) (12 mg, 0.012 mmol, 35%).

5-Hydroxyferuloyl CoA (**11**) was synthesized in an analogous manner: 5-hydroxyferulic acid (**5**; 0.021 g, 0.1 mmol) and CDI (**22**; 0.032 g, 0.19 mmol) afforded 5-hydroxyferulic acid imidazolide (**23**; 0.015 g, 0.059 mmol, 60%; Scheme 1) was obtained, a portion (10 mg) of which was then converted into 5-hydroxyferuloyl CoA (**11**) (11 mg, 0.011 mmol, 30%).

4.4. At4CL homologues

The putative *At4CL* homologues characterized in this study were designated *At4CL1–At4CL11* (Costa et al., 2003) (Table 1); corresponding cDNA sequences were obtained from The *Arabidopsis* Information Resource (TAIR; http://www.arabidopsis.org) database and used to design and synthesize gene specific primers (Table 2).

4.5. Cloning of A. thaliana 4CL homologues

cDNA clones of *At4CL8* (U11110, GenBank Accession number AY117254) and *At4CL9* (U09223, GenBank Accession number AY040047), each in a pUni51 vector, were provided by the Plant Gene Expression Center (Albany, CA), whereas *At4CL 1–7*, *10* and *11* were obtained as follows: whole tissues of 16-day old wild type *A. thaliana* plants were used for total RNA isolation using RNeasy® Plant Mini Kits, following the manufacturer's instructions. Each *At4CL* homologue, *At4CL1–7*, *10* and *11*, was individually amplified, using *Pfu Turbo*® DNA Polymerase and gene specific primers (Table 2) and a cDNA product obtained from total RNA (1 μg) with the SuperScript™ First Strand Synthesis System for RT-PCR. Each PCR mixture contained template (10 ng), 0.2 mM dNTP mixture, gene-specific primers (100 ng each of forward and reverse), 2.5 units *Pfu Turbo*® DNA Polymerase, 2 μl 10× *Pfu Turbo*® buffer in a total final volume of 20 μl. The PCR conditions for cDNA amplification were as follows: initial denaturation for 2 min at 96 °C, followed by 35 cycles of 95 °C, 30 s; 55 °C, 30 s; and 70 °C, 3 min, with a final extension at 70 °C for 10 min. After

A-overhang addition using *Taq* DNA polymerase as suggested by Invitrogen, the PCR products were individually analyzed on 0.7% low melting point agarose gels. Fragments of interest for each putative *At4CL* gene were purified by cutting out the gel fragment containing the band of interest, followed by digestion with Agar*ACE*® enzyme according to Promega's protocol. These fragments were then individually cloned into a pCRT7-CT TOPO® vector and transformed into *E. coli* TOP10 F′ cells. Plasmid DNA for each of the putative *At4CL* clones was isolated using the Wizard® Plus SV Miniprep DNA Purification System and the authenticity and accuracy of each clone was verified by sequencing.

Putative *At4CL* clones obtained from the Plant Gene Expression Center were PCR amplified from the pUni51 vector using the individual template (U11110 for *At4CL8*, U09223 for *At4CL9*) with the corresponding gene specific primers. Each PCR mixture contained template (10 ng), 0.2 mM dNTP mixture, gene-specific primer (100 ng) (Table 2), 5 units *Pfu Turbo*® DNA Polymerase (Stratagene), 2.5 μl 10× *Pfu Turbo*® buffer, and 2.5 mM $MgCl_2$ in a total volume of 25 μl, with PCR amplifications performed as follows: initial denaturation at 94 °C for 5 min, followed by 30 cycles at 94 °C, 1 min; 55 °C, 1 min; 72 °C, 2 min with a final extension at 72 °C for 7 min. The PCR products were individually analyzed on 1.2% agarose gels with fragments of interest for each *At4CL* gene isolated and purified using the QIAquick® Gel Extraction Kit, following the manufacturer's instructions. These were then individually cloned into a pCRT7-CT TOPO® vector and transformed into *E. coli* TOP10 F′ cells. Plasmid DNA for each of these putative *At4CL* clones was isolated using the QIAprep® Spin Miniprep Kit and sequenced as described above.

4.6. DNA and protein sequence analyses

cDNA sequences of the *At4CL*'s were analyzed by BLAST 2 SEQUENCES (http://www.ncbi.nlm.nih.gov/) to ensure there were no point mutations; ExPASy molecular biology server (http://us.expasy.org/) was also used to confirm translation of all putative *At4CL* open reading frames, whereas protein Mw and CLUSTAL_W (1.8) from http://www.ebi.ac.uk was employed to calculate percentages of similarity and identity as well as being used for phylogenetic analysis. CLUSTAL_W (http://npsa-pbil.ibcp.fr/cgi-bin) was used to align amino acid sequences by sequence weighting, position-specific gap penalties and weight matrix choices (Thompson et al., 1994). The following 4CL sequences from GenBank were used for CLUSTAL analyses: tobacco Nt4CL1 (U50845), soybean Gm4CL1 (AF279267) and pine Pt4CL1 (U12012).

4.7. Expression of A. thaliana 4CL homologues

cDNAs of the 11 *At4CL* homologues, (*At4CL1–11*), either with a stop codon for native protein production or minus the stop codon to allow for His_6-tagged purification, were PCR amplified and individually cloned into a pCRT7-CT® expression vector, followed by sequence confirmation. For protein expression and purification, the resulting individual expression constructs containing the *At4CL*genes were then transformed into *E. coli* BL21 STAR™ (DE3). In addition, the *At4CL1* and *At4CL10* genes were cloned into a pET24d expression vector (C-terminal His_6-tagged), transformed into competent *E. coli* BL21 STAR™ (DE3) cells and selected on LB medium containing kanamycin (50 μg ml^{-1}) for protein expression.

Transformed *E. coli* were inoculated into YEP medium (5 ml) containing 50 μg ml^{-1} zeocin (50 μg ml^{-1} kanamycin for*At4CL1* and *At4CL10* in pET24d expression vector), and grown overnight at 37 °C/250 rpm. On the next day, an aliquot of this seed culture (100 μl) was inoculated into YEP medium (500 ml) containing the respective antibiotic as above and incubated at 37 °C with shaking (250 rpm) until cell densities at A_{600} reached 0.6 at which time the temperature was then changed to 20 °C. When the cells had reached this temperature, they were induced with isopropyl-β-D-thiogalactopyranoside (IPTG, 0.5 mM final concentration) and allowed to grow for 24 h. Each cell culture was then harvested by centrifugation (15,000*g*/4 °C for 10 min), with the cell pellets stored at −80 °C. Frozen cell pellets of each At4CL homologue were thawed and resuspended in BugBuster® Protein Extraction Reagent-100 mM Tris–HCl buffer (pH 7.5) containing 2 mM dithiothreitol, rLysozyme™ (1 KU ml^{-1}) and Benzonase® Nuclease (25 units ml^{-1}) (1:1; 7 ml). Each suspension was next incubated on ice on a platform shaker for 30 min. Supernatants containing soluble At4CL proteins were individually recovered after centrifugation (16,000*g*, 4 °C, 15 min) and filtered through an 0.45 μm Acrodisc® syringe filter (Pall Life Sciences, Ann Arbor, MI, USA), with aliquots (5 μg protein) analyzed using SDS–PAGE and visualization by silver staining as before.

4.8. At4CL purification

4.8.1. His_6-tagged proteins

At4CL proteins, previously tagged in the cloning vector pCRT7-CT to His_6 at their specific C-termini, were individually purified by perfusion chromatography using the BioCAD 700E work station. Aliquots (2 ml) of each crude At4CL protein preparation (see Section 4.7) were applied to POROS® 20 Metal-chelate affinity columns previously equilibrated in binding buffer (20 mM Tris–HCl, pH 7.9; 500 mM NaCl; 5 mM

imidazole), at a flow rate of 10 ml min^{-1}. Each column was first washed with binding buffer (8 ml) to remove unbound contaminant proteins. Target proteins were then individually eluted with a linear gradient of 10–100% elution buffer (20 mM Tris–HCl, pH 7.9; 500 mM NaCl; 500 mM imidazole) in binding buffer, with 1 ml fractions collected. Aliquots (14 μl) of alternate fractions were analyzed by SDS–PAGE with silver staining for visualization as above. Fractions containing each His_6-tagged At4CL protein (MW ~65 kDa) were pooled and concentrated in a Centricon® Plus-80 (Millipore, Billerica, MA, USA) to a volume of 1 ml at 4 °C. Each resulting supernatant was diluted to 80 ml with Tris–HCl buffer (20 mM, pH 7.5) and re-concentrated to remove excess imidazole to a final volume of 1–2 ml. Enzyme preparations were stored in 20 mM Tris–HCl buffer, pH 7.5 with 10% (v/v) glycerol at −80 °C until needed.

4.8.2. Native At4CL1 purification

The supernatant containing native At4CL1 (see Section 4.7) was subjected to ammonium sulfate precipitation. Proteins precipitating between 30% and 70% ammonium sulfate were recovered by centrifugation (14,500*g*, 15 min), with the pellet reconstituted in 10 ml of Tris–HCl buffer (20 mM, pH 7.5) containing DTT (5 mM) (Buffer A) and desalted on PD10 column equilibrated with Buffer A. This enzyme preparation was next applied to an Affi-Gel Blue gel column (32 × 1.6 cm) equilibrated in Buffer A at a flow rate of 1 ml min^{-1}. After washing the column with Buffer A (100 ml), At4CL1 was eluted with a gradient of NaCl (0 to 5 M) in Buffer A (100 ml). Active fractions, eluted between 2.5 and 4 M NaCl were combined, desalted on PD10 equilibrated in KPi buffer (5 mM, pH 7.5) containing DTT (5 mM) (Buffer B). The resulting solution was next loaded onto a CHT ceramic hydroxyapatite column (8.5 × 1 cm) equilibrated in Buffer B at a flow rate of 1 ml min^{-1}. After washing the column with Buffer B (10 ml), At4CL1 was eluted with linear gradients of KPi first from 5 to 125 mM in 100 ml and then from 125 to 250 mM in 30 ml. Active fractions eluted between 35 and 60 mM KPi were pooled, concentrated (4 ml) and used for kinetic analyses.

4.9. Enzyme characterization

4.9.1. General assays

Standard assays consisted of 200 μM substrate **1**–**6**, 2.5 mM ATP, 2.5 mM $MgCl_2$, 0.2 mM Coenzyme A, 1–120 μg of purified At4CL protein (depending on activity level), Bis–Tris–propane buffer (100 mM, pH 7.5) in a total volume of 250 μl. Enzymatic reactions were initiated by addition of At4CL. Controls were carried out with boiled (96 °C, 10 min) At4CL protein. After 10 min incubation at 30 °C, assays were individually terminated by addition of HOAc (10 μl), with an aliquot (80 μl) of each assay mixture analyzed by reversed-phase HPLC, with individual separation of CoA esters from substrate acids achieved as follows: eluants A (100 mM ammonium acetate, pH 4.3) and B (CH_3CN) in a linear gradient of A:B from 100:0 to 90:10 from 0 to 5 min, followed by a concave gradient (Waters curve 7) to A:B (1:1) over 20 min and finally a linear gradient from A:B (1:1) to 100% B in 5 min, this being held for 2 min at a flow rate of 1 ml min^{-1}. Detection was carried out at 311 nm for cinnamoyl-CoA (**7**) and 335 nm for *p*-coumaroyl (**8**), caffeoyl (**9**), feruloyl (**10**), 5-hydroxyferuloyl (**11**) and sinapoyl (**12**) CoAs, using a photodiode array 994 system (Waters, Milford, MA, USA); calibration curves were established for each CoA ester product (**7–12**) prior to assays.

pH and temperature optima were determined for the most active isoforms (At4CL1–3, 5 and10) using standard assay conditions as described above. For pH optima, incubations were carried at 30 °C with Bis–Tris–propane buffer (100 mM, pH 6–9) or CAPSO buffer (100 mM, pH 9–11), whereas for temperature optima, assays were performed in Bis–Tris–propane buffer (100 mM, pH 7.5) with incubation temperatures ranging from 20 to 70 °C.

4.9.2. Feruloyl CoA formation by At4CL2

To Tris–HCl buffer (100 mM, pH 7.5, 5 ml) was added $MgCl_2$ (62.5 mM, 400 μl), CoA (5 mM, 390 μl), ATP (62.5 mM, 200 μl), ferulic acid (**4**, 3.3 mg in 500 μl MeOH) and At4CL2 (233 μg, 200 μl), with the volume brought to 10 ml with H_2O. After stirring at room temperature for 2 h, ammonium acetate (0.4 g) was added, with the enzymatically formed **9** purified exactly as described in Beuerle and Pichersky (2002) to give feruloyl CoA (**9**, 2 mg, 2.1 μmol).

UV $\lambda_{max}^{H_2O}$ nm (log ε): 347.2 (0.14), 261.4 (0.118); ^{1}H NMR (500 MHz; D_2O): δ 0.75 (*s*, 3H, 14-CH_3), 0.89 (*s*, 3H, 13-CH_3), 2.46 (*m*, 2H, 18-CH_2), 3.11 (*m*, 1H, 21-CH), 3.15 (*m*, 1H, 21-CH), 3.42 (*m*, 2H, 20-CH_2), 3.46 (*m*, 2H, 17-CH_2), 3.54 (*dd*, 1H, $J = 10.2$, $J_{HP} = 4.2$ Hz, 11-CH), 3.83 (*dd*, 1H, $J = 10.2$, $J_{HP} = 4.7$ Hz, 11-CH), 3.85 (*s*, 3H, OCH_3), 4.02 (*s*, 1H, 15-CH), 4.20 (*br m*, 2H, 10-CH_2), 4.53 (*br* m, 1H, 9-CH), 4.73 (*br m*, 1H, 8-CH), 4.80 (*m*, 1H, 7-CH), 6.02 (*d*, 1H, $J = 6.1$ Hz, 6-CH), 6.59 (*d*, 1H, $J = 15.8$ Hz, 23-CH), 6.83 (*d*, 1H, $J = 8.3$ Hz, 27-CH), 6.99 (*dd*, 1H, $J = 8.3$, 1.2 Hz, 26-CH), 7.04 (*br* s, 1H, 30-CH), 7.36 (*d*, 1H, $J = 15.8$ Hz, 24-CH), 8.05 (*s*, 1H, 2-CH), 8.43 (*s*, 1H, 5-CH); ^{13}C NMR (125 MHz; D_2O): δ 20.98 (C-14), 23.58 (C-13), 30.84 (C-21), 38.13 (C-18), 38.20 (C-17), 41.07 (*d*, $^3J_{CP} = 8.2$ Hz, C-12), 41.47 (C-20), 58.47 (OCH_3), 68.4 (*d*, $^2J_{CP} = 4.3$ Hz, C-10), 74.55 (*d*, $^2J_{CP} = 5.8$ Hz, C-11), 76.36 (*d*, $^2J_{CP} = 3.9$ Hz, C-8), 76.85 (C-15), 77.16 (*br*, C-7), 86.51 (*br*, C-9), 89.25 (C-6), 113.62 (C-30), 118.29 (C-27), 121.16 (C-4), 124.44 (C-23), 126.54 (C-26), 128.84

(C-25), 142.22 (C-5), 144.54 (C-24), 150.18 (C-29), 150.84 (C-28), 151.71 (C-3), 155.22 (C-2), 158.0 (C-1), 176.80 (C-19), 177.48 (C-16), 195.94 (C-22); ESI-MS *m/z*: 942.3 $[M–H]^-$.

4.9.3. Enzyme kinetics

Kinetic constants (K_m and V_{max}) for each functional At4CL isoform were performed under the standard conditions described above at pH 7.5 and at 30 °C, with substrate concentrations (for **1**–**6**) ranging from 2 to 400 μM and purified At4CL protein (33 ng–11 μg for At4CL1, 0.1–47 μg for At4CL2, 0.5–14 μg for At4CL3 and 1–3 μg for At4CL5). Assays for individual substrates were carried out in triplicate for 5 min for substrates **1**, **2**, **4** and **5**, and 2 min for **3** and **6**, with controls as above but without CoA. Kinetic constants (K_m and V_{max}) were determined from Lineweaver–Burk plots.

4.10. Quantitative real time PCR

For analysis of transcript accumulation of *At4CLs* in wild type (ecotype Columbia and Landsberg) and *4cl5* mutant lines, total RNA was extracted from 5- and 14-day old seedlings as well as from 4- and 6-week old inflorescence basal stems (~1.5 cm, 80–100 mg) using the RNeasy® Plant RNA Isolation Kit (Qiagen) following the manufacturer's directions. An aliquot (2 μg) of isolated RNA was treated with DNase I (Invitrogen) prior to cDNA synthesis using Superscript™ II Reverse Transcriptase (Invitrogen) and oligo(dT) primers according to the manufacturer's protocol. PCR primers specific to each of the *4CL* genes from the *Arabidopsis* genome were designed by using the SPAD program (http://genoplate-info.infobiogen.fr/spads) (see Table 4). Positions and the predicted PCR product sizes are shown in Table 4. The specificity of each primer set was confirmed by the appearance of only one peak in melting curve analysis and by a single band in agarose gel electrophoresis as well as sequencing of amplified fragments. Since the SPAD program could not generate specific primers for *At4CL3* and several primers designed manually gave rise to multiple bands, the *At4CL3* was not analyzed in this study. The *ACT2* gene (At3g18780), to which the RT-PCRs were normalized when expressed as relative expression levels, was also amplified with sense and antisense primers (Table 4). Quantitative real time PCR was performed on the MX4000 and fluorescence intensities were normalized against the reference dye, ROX. The RT-PCR mixture contained a 5-fold diluted cDNA mixture for a template, 2× Platinum SYBR® Green qPCR SuperMix-UDG (12.5 μl, Invitrogen), ROX reference dye and 10 μM of gene-specific primers. Cycling parameters were 50 °C for 2 min, 95 °C for 2 min, followed by 40 cycles of 95 °C for 30 s and 60 °C for 1 min. The relative expression (RE) level was normalized with *ACT2* gene expression level using the modified equation of Pfaffl (2001) and calculated as $RE = RE = E^{\Delta Ct_{4CLs}}/E^{\Delta Ct_{ACT2}}$, where E is the amplification efficiency. Amplification PCR efficiency (75–85% in this study) was determined by amplification of a dilution series of reverse transcribed cDNA (100, 50, 10, 5, 1, 0.5 ng per reaction).

4.11. Sample preparation for lignin analysis and estimations of lignin content and lignin monomeric compositions

For lignin analysis, stems (from 20 to 30 individual plants) from both knockout lines and wild type plants (Columbia and *L*er) were initially harvested 4 weeks after seeding and every week thereafter until maturation/senescence (10-week old). Samples were prepared as described in Kim et al. (2004). Lignin contents were estimated using the acetyl bromide method (Iiyama and Wallis, 1990), whereas nitrobenzene oxidation was carried out as described in Iiyama and Lam (1990); thioacidolysis employed the procedures of Lapierre et al. (1986) and Rolando et al. (1992), as described in Blee et al. (2001).

4.12. Sinapate/sinapyl alcohol derivative determinations

For sinapate/sinapyl alcohol derivative determinations, stems and leaves from both WT (Columbia and *L*er) and knockout plants were harvested weekly from weeks 4 to 12. Frozen leaves and stems were pulverized in a mortar by means of a pestle and extracted with MeOH–HOAc–H_2O (9:1:10, 5 ml g^{-1} fresh weight) for 24 h. After centrifugation (16,000*g*, 5 min), an aliquot (80 μl) was subjected to reversed-phase HPLC analysis using the following solvent system: isocratic eluants A (MeOH–MeCN, 1:1) and B (3% HOAc in H_2O) 10:90 from 0 to 5 min then a linear gradient of A:B to 56:44 over 30 min at a flow rate of 1 ml min^{-1} with detection at 320 nm for compounds **18**–**20** and 520 nm for the anthocyanin **21**.

Acknowledgements

This research project is supported by the National Science Foundation (Arabidopsis 2010, MCB-0117260), the National Aeronautics and Space Administration (NAG 2-1513), the United States Department of Energy (DE FG03-97ER20259) and the G. Thomas and Anita Hargrove Center for Plant Genomic Research, as well by Dow Chemical Co. (Dow AgroSciences, to J.B.) and the Agricultural Research Center at Washington State University. The authors thank Dr. Erin Pichersky for the *4CL* tobacco clone.

References

Allina, S.M., Pri-Hadash, A., Theilmann, D.A., Ellis, B.E., Douglas, C.J., 1998. 4-Coumarate:coenzyme A ligase in hybrid poplar. Properties of native enzymes, cDNA cloning, and analysis of recombinant enzymes. Plant Physiol. 116, 743–754.

Anterola, A.M., Jeon, J.-H., Davin, L.B., Lewis, N.G., 2002. Transcriptional control of monolignol biosynthesis in *Pinus taeda*: factors affecting monolignol ratios and carbon allocation in phenylpropanoid metabolism. J. Biol. Chem. 277, 18272–18280.

Anterola, A.M., Lewis, N.G., 2002. Trends in lignin modification: a comprehensive analysis of the effects of genetic manipulations/mutations on lignification and vascular integrity. Phytochemistry 61, 221–294.

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.

Beuerle, T., Pichersky, E., 2002. Enzymatic synthesis and purification of aromatic coenzyme A esters. Anal. Biochem. 302, 305–312.

Blee, K., Choi, J.W., O'Connell, A.P., Jupe, S.C., Schuch, W., Lewis, N.G., Bolwell, G.P., 2001. Antisense and sense expression of cDNA coding for CYP73A15, a class II cinnamate 4-hydroxylase, leads to a delayed and reduced production of lignin in tobacco. Phytochemistry 57, 1159–1166.

Bloor, S.J., Abrahams, S., 2002. The structure of the major anthocyanin in *Arabidopsis thaliana*. Phytochemistry 59, 343–346.

Brown, S.A., Wright, D., Neish, A.C., 1959. Studies of lignin biosynthesis using isotopic carbon. VII. The role of *p*-hydroxyphenylpyruvic acid. Can. J. Biochem. Physiol. 37, 25–34.

Cochrane, F.C., Davin, L.B., Lewis, N.G., 2004. The *Arabidopsis* phenylalanine ammonia-lyase multigene family: kinetic characterization of the four PAL isoforms. Phytochemistry 65, 1557–1564.

Costa, M.A., Collins, R.E., Anterola, A.M., Cochrane, F.C., Davin, L.B., Lewis, N.G., 2003. An in silico assessment of gene function and organization of the phenylpropanoid pathway metabolic networks in *Arabidopsis thaliana* and limitations thereof. Phytochemistry 64, 1097–1112.

Cuff, J.A., Birney, E., Clamp, M.E., Barton, G.J., 2000. ProtEST: protein multiple sequence alignments from expressed sequence tags. Bioinformatics 16, 111–116.

Cukovic, D., Ehlting, J., VanZiffle, J.A., Douglas, C.J., 2001. Structure and evolution of 4-coumarate:coenzyme A ligase (*4CL*) gene families. Biol. Chem. 382, 645–654.

Douglas, C., Hoffmann, H., Schulz, W., Hahlbrock, K., 1987. Structure and elicitor or UV-light-stimulated expression of two 4-coumarate:CoA ligase genes in parsley. EMBO J. 6, 1189–1195.

Ehlting, J., Büttner, D., Wang, Q., Douglas, C.J., Somssich, I.E., Kombrink, E., 1999. Three 4-coumarate:coenzyme A ligases in *Arabidopsis thaliana* represent two evolutionarily divergent classes in angiosperms. Plant J. 19, 9–20.

Gross, G.G., Mansell, R.L., Zenk, M.H., 1975. Hydroxycinnamate: coenzyme A ligase from lignifying tissue of higher plants. Biochem. Physiol. Pflanzen 168, S41–S51.

Gross, G.G., Stöckigt, J., Mansell, R.L., Zenk, M.H., 1973. Three novel enzymes involved in the reduction of ferulic acid to coniferyl alcohol in higher plants: ferulate:CoA ligase, feruloyl-CoA reductase and coniferyl alcohol oxidoreductase. FEBS Lett. 31, 283–286.

Gross, G.G., Zenk, M.H., 1966. Darstellung und Eigenschaften von Coenzym A-Thiolestern substituierter Zimtsäuren. Z. Naturforsch. 21b, 683–690.

Gross, G.G., Zenk, M.H., 1974. Isolation and properties of hydroxycinnamate:CoA ligase from lignifying tissue of *Forsythia*. Eur. J. Biochem. 42, 453–459.

Hahlbrock, K., Grisenbach, H., 1970. Formation of coenzyme A esters of cinnamic acids with an enzyme preparation from cell suspension cultures of parsley. FEBS Lett. 11, 62–64.

Hamberger, B., Hahlbrock, K., 2004. The *4-coumarate:CoA ligase* gene family in *Arabidopsis thaliana* comprises one rare, sinapate-activating and three commonly occurring isoenzymes. Proc. Natl. Acad. Sci. USA 101, 2209–2214.

Harding, S.A., Leshkevich, J., Chiang, V.L., Tsai, C.J., 2002. Differential substrate inhibition couples kinetically distinct 4-coumarate:coenzyme A ligases with spatially distinct metabolic roles in quaking aspen. Plant Physiol. 128, 428–438.

Hu, W.-J., Kawaoka, A., Tsai, C.-J., Lung, J., Osakabe, K., Ebinuma, H., Chiang, V.L., 1998. Compartmentalized expression of two structurally and functionally distinct 4-coumarate:CoA ligase genes in aspen (*Populus tremuloides*). Proc. Natl. Acad. Sci. USA 95, 5407–5412.

Iiyama, K., Lam, T.B.T., 1990. Lignin in wheat internodes. Part I: the reactivities of lignin units during alkaline nitrobenzene oxidation. J. Sci. Food Agric. 51, 481–491.

Iiyama, K., Wallis, A.F.A., 1990. Determination of lignin in herbaceous plants by an improved acetyl bromide procedure. J. Sci. Food Agric. 51, 145–161.

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.

Kim, S.-J., Kim, M.-R., Bedgar, D.L., Moinuddin, S.G.A., Cardenas, C.L., Davin, L.B., Kang, C.-H., Lewis, N.G., 2004. Functional reclassification of the putative cinnamyl alcohol dehydrogenase multigene family in *Arabidopsis*. Proc. Natl. Acad. Sci. USA 101, 1455–1460.

Knobloch, K.-H., Hahlbrock, K., 1975. Isoenzymes of *p*-coumarate:CoA ligase from cell suspension cultures of *Glycine max*. Eur. J. Biochem. 52, 311–320.

Knobloch, K.-H., Hahlbrock, K., 1977. 4-Coumarate:CoA ligase from cell suspension cultures of *Petroselinum hortense* Hoffm. Arch. Biochem. Biophys. 184, 237–248.

Lapierre, C., Monties, B., Rolando, C., 1986. Thioacidolysis of poplar lignins: identification of monomeric syringyl products and characterization of guaiacyl-syringyl lignin fractions. Holzforschung 40, 113–118.

Lee, D., Douglas, C.J., 1996. Two divergent members of a tobacco 4-coumarate:coenzyme A ligase (*4CL*) gene family. cDNA structure, gene inheritance and expression, and properties of recombinant proteins. Plant Physiol. 112, 193–205.

Lee, D., Ellard, M., Wanner, L.A., Davis, K.R., Douglas, C.J., 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., Meyer, K., Chapple, C., Douglas, C.J., 1997. Antisense suppression of 4-coumarate:coenzyme A ligase activity in *Arabidopsis* leads to altered lignin subunit composition. Plant Cell 9, 1985–1998.

Lewis, N.G., Davin, L.B., Sarkanen, S., 1999. The nature and function of lignins. In: Barton, Sir D.H.R., Nakanishi, K., Meth-Cohn, O. (Eds.), Comprehensive Natural Products Chemistry, vol. 3. Pergamon Press, Oxford, pp. 617–745.

Lindermayr, C., Fliegmann, J., Ebel, J., 2003. Deletion of a single amino acid residue from different 4-coumarate:CoA ligases from soybean results in the generation of new substrate specificities. J. Biol. Chem. 278, 2781–2786.

Lindermayr, C., Möllers, B., Fliegmann, J., Uhlmann, A., Lottspeich, F., Meimberg, H., Ebel, J., 2002. Divergent members of a soybean (*Glycine max* L.) 4-coumarate:coenzyme A ligase gene family. Primary structures, catalytic properties, and differential expression. Eur. J. Biochem. 269, 1304–1315.

Lindl, T., Kreuzaler, F., Hahlbrock, K., 1973. Synthesis of *p*-coumaroyl coenzyme A with a partially purified *p*-coumarate:CoA ligase from cell suspension cultures of soybean (*Glycine max*). Biochem. Biophys. Acta 302, 457–464.

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.

Mansell, R.L., Stöckigt, J., Zenk, M.H., 1972. Reduction of ferulic acid to coniferyl alcohol in a cell free system from a higher plant. Z. Pflanzenphysiol. 68, S286–S288.

Min, T., Kasahara, H., Bedgar, D.L., Youn, B., Lawrence, P.K., Gang, D.R., Halls, S.C., Park, H., Hilsenbeck, J.L., Davin, L.B., Lewis, N.G., Kang, C., 2003. Crystal structures of pinoresinol–lariciresinol and phenylcoumaran benzylic ether reductases and their relationship to isoflavone reductases. J. Biol. Chem. 278, 50714–50723.

Pabsch, K., Petersen, M., Rao, N.N., Alfermann, A.W., Wandrey, C., 1991. Chemo-enzymatic synthesis of rosmarinic acid. Recl. Trav. Chim. Pays-Bas 110, 199–205.

Page, R.D., 1996. TreeView: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. 12, 357–358.

Patten, A.M., Cardenas, C.L., Cochrane, F.C., Laskar, D.D., Bedgar, D.L., Davin, L.B., Lewis, N.G., 2005. Reassessment of effects on lignification and vascular development in the *irx4 Arabidopsis* mutant. Phytochemistry (in press).

Pfaffl, M.W., 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 29, 2002–2007.

Raes, J., Rohde, A., Christensen, J.H., Van de Peer, Y., Boerjan, W., 2003. Genome-wide characterization of the lignification toolbox in *Arabidopsis*. Plant Physiol. 133, 1051–1071.

Ragg, H., Kuhn, D.N., Hahlbrock, K., 1981. Coordinated regulation of 4-coumarate:CoA ligase and phenylalanine ammonia-lyase mRNAs in cultured plant cells. J. Biol. Chem. 256, 10061–10065.

Rhodes, M.J.C., Wooltorton, L.S.C., 1973. Formation of CoA esters of cinnamic acid derivatives by extracts of *Brassica napobrassica* root tissue. Phytochemistry 12, 2381–2387.

Rhodes, M.J.C., Wooltorton, L.S.C., 1975. The *p*-coumaryl CoA ligase of potato tubers. Phytochemistry 14, 2161–2164.

Rolando, C., Monties, B., Lapierre, C., 1992. Thioacidolysis. In: Lin, S.Y., Dence, C.W. (Eds.), Methods in Lignin Chemistry. Springer-Verlag, Berlin, pp. 334–349.

Schneider, K., Hövel, K., Witzel, K., Hamberger, B., Schomburg, D., Kombrink, E., Stuible, H.-P., 2003. The substrate specificity-determining amino acid code of 4-coumarate:CoA ligase. Proc. Natl. Acad. Sci. USA 100, 8601–8606.

Schneider, K., Kienow, L., Schmelzer, E., Colby, T., Bartsch, M., Miersch, O., Wasternack, C., Kombrink, E., Stuible, H.-P., 2005. A new type of peroxisomal acyl-coenzyme A synthetase from *Arabidopsis thaliana* has the catalytic capacity to activate biosynthetic precursors of jasmonic acid. J. Biol. Chem. 280, 13962–13972.

Shockey, J.M., Fulda, M.S., Browse, J., 2003. *Arabidopsis* contains a large superfamily of acyl-activating enzymes. Phylogenetic and biochemical analysis reveals a new class of acyl-coenzyme A synthetases. Plant Physiol. 132, 1065–1076.

Stuible, H.-P., Büttner, D., Ehlting, J., Hahlbrock, K., Kombrink, E., 2000. Mutational analysis of 4-coumarate:CoA ligase identifies functionally important amino acids and verifies its close relationship to other adenylate-forming enzymes. FEBS Lett. 467, 117–122.

Stuible, H.-P., Kombrink, E., 2001. Identification of the substrate specificity-conferring amino acid residues of 4-coumarate:coenzyme A ligase allows the rational design of mutant enzymes with new catalytic properties. J. Biol. Chem. 276, 26893–26897.

The Arabidopsis Genome Initiative, 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408, 796–815.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

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.

Voo, K.S., Whetten, R.W., O'Malley, D.M., Sederoff, R.R., 1995. 4-Coumarate:coenzyme A ligase from loblolly pine xylem. Isolation, characterization, and complementary DNA cloning. Plant Physiol. 108, 85–97.

Wallis, P.J., Rhodes, M.J.C., 1977. Multiple forms of hydroxycinnamate:CoA ligase in etiolated pea seedlings. Phytochemistry 16, 1891–1894.

Walton, E., Butt, V.S., 1970. The activation of cinnamate by an enzyme from leaves of spinach beet (*Beta vulgaris* L. ssp. *vulgaris*). J. Exp. Bot. 21, 887–891.

Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E., Markley, J.L., Sykes, B.D., 1995. ^{1}H, ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. J. Biomol. NMR 6, 135–140.

Yamauchi, K., Yasuda, S., Hamada, K., Tsutsumi, Y., Fukushima, K., 2003. Multiform biosynthetic pathway of syringyl lignin in angiosperms. Planta 216, 496–501.

Youn, B., Moinuddin, S.G.A., Davin, L.B., Lewis, N.G., Kang, C.H., 2005. Crystal structures of apo-form, and binary/ternary complexes of *Podophyllum* secoisolariciresinol dehydrogenase, an enzyme involved in formation of health-protecting and plant defense lignans. J. Biol. Chem. 280, 12917–12926.

Zhang, X.-H., Chiang, V.L., 1997. Molecular cloning of 4-coumarate:coenzyme A 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, S.D., Dube, S.K., 1990. Nucleotide sequence of rice 4-coumarate:CoA ligase gene, 4-CL1. Nucleic Acids Res. 18, 6144.