2012

Switchgrass PviCAD1: Understanding Residues Important for Substrate Preferences and Activity

Aaron J. Saathoff
USDA-ARS Grain, Forage, asaathoff2@unl.edu

Mark S. Hargrove
Iowa State University, msh@iastate.edu

Eric J. Haas
Creighton University, EricHaas@creighton.edu

Christian M. Tobias
USDA-ARS Genomics and Gene Discovery Unit, christian.tobias@ars.usda.gov

Paul Twigg
University of Nebraska at Kearney, twiggp@unk.edu

See next page for additional authors

Appl Biochem Biotechnol (2012) 168:1086–1100
DOI 10.1007/s12010-012-9843-0

Switchgrass PviCAD1: Understanding Residues Important for Substrate Preferences and Activity

Aaron J. Saathoff · Mark S. Hargrove · Eric J. Haas · Christian M. Tobias · Paul Twigg · Scott Sattler · Gautam Sarath

Received: 20 January 2012 / Accepted: 9 August 2012 /
Published online: 23 August 2012

Abstract Cinnamyl alcohol dehydrogenase (CAD) catalyzes the final step in monolignol biosynthesis. Although plants contain numerous genes coding for CADs, only one or two CADs appear to have a primary physiological role in lignin biosynthesis. Much of this distinction appears to reside in a few key residues that permit reasonable catalytic rates on monolignal substrates. Here, several mutant proteins were generated using switchgrass wild type (WT) PviCAD1 as a template to understand the role of some of these key residues, including a proton shuttling HL duo in the active site. Mutated proteins displayed lowered or limited activity on cinnamylaldehydes and exhibited altered kinetic properties compared to the WT enzyme, suggesting that key residues important for efficient catalysis had been identified. We have also shown that a sorghum ortholog containing EW, instead of HL in its active site, displayed negligible activity against monolignals. These results indicate that lignifying CADs require a specific set of key residues for efficient activity against monolignals.

Keywords Switchgrass · Cinnamyl alcohol dehydrogenase · Lignin · Protein mutagenesis

A. J. Saathoff · S. Sattler · G. Sarath (✉)
USDA-ARS Grain, Forage, and Bioenergy Research Unit, University of Nebraska, 137 Keim Hall, Lincoln, NE 68583-0937, USA
e-mail: Gautam.Sarath@ars.usda.gov

M. S. Hargrove
Department of Biochemistry, Biophysics & Molecular Biology, Iowa State University, 4114 Molecular Biology Building, Ames, IA 50011, USA

E. J. Haas
Department of Chemistry, Creighton University, 2500 California Plaza, Omaha, NE 68178, USA

C. M. Tobias
USDA-ARS Genomics and Gene Discovery Unit, 800 Buchanan St., Albany, CA 94710, USA

P. Twigg
Department of Biology, University of Nebraska-Kearney, Kearney, NE 68849, USA

Introduction

Terrestrial land plants have been estimated to assimilate about 56 Gt of carbon annually [1]. Much of this carbon is used in the synthesis of cellulose [2, 3], hemicellulose [4], and lignin [5], which make up the three main components of plant cell walls and are the most abundant polymers on earth. Because of their abundance, these three polymers offer attractive targets for producing renewable fuels, power, and chemicals. However, the complex nature in which these polymers are interconnected in plant cell walls has presented formidable technical challenges in developing cost-effective and environmentally benign conversion technologies based on a biochemical platform. Therefore, understanding how plants control and synthesize their cell wall components, especially lignins, is important for targeting efforts towards developing perennial plants, such as switchgrass (*Panicum virgatum* L.), which are more amenable to conversion processes [6, 7].

Lignin is a complex heteropolymer consisting primarily of three main phenylpropanoid subunits: *p*-hydroxyphenyl (H-lignin), guaiacyl (G-lignin), and syringyl (S-lignin) units which are derived from their respective alcohols [8, 9]. Lignin has also been found to contain additional constituents such as acylated and aldehyde units and, in grasses, ferulate [5]. Compared to dicots, grasses have substantially more lignin in their secondary cell walls [10]. The various types of chemical bonds found in lignin have made compositional and structural analysis difficult [11].

Current research has implicated approximately ten enzymes in monolignol biosynthesis [5, 8, 12]. A key enzyme involved in the later steps of lignin biosynthesis is cinnamyl alcohol dehydrogenase (CAD), which catalyzes the reduction of *p*-hydroxycinnamaldehydes to their corresponding alcohols (Fig. 1). Cinnamylalcohols are the primary monomers incorporated into the lignin polymer via enzymes such as laccases, peroxidases, or polyphenol oxidases [5]. CAD has been isolated and characterized from a variety of plant species [13–16], and its crystal structure has been published [17]. The crystal structure indicated that AtCAD4 and AtCAD5 share many structural features with the Zn-dependent medium chain reductase superfamily, although the catalytic Zn^{2+} of both CADs was coordinated to Glu, a

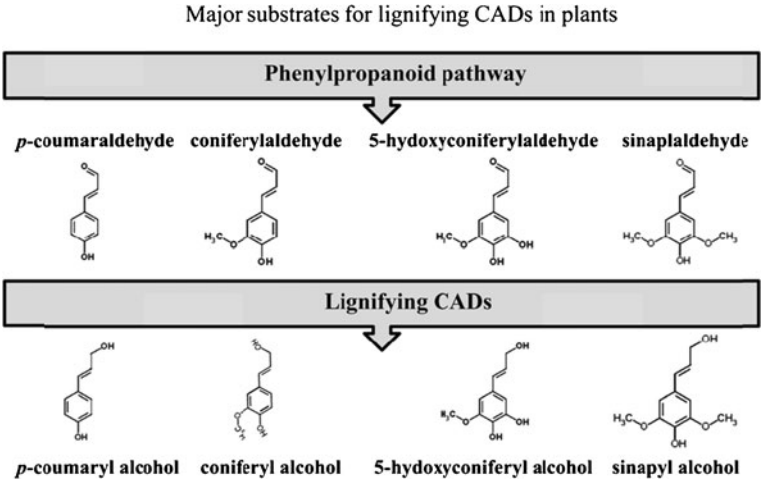

Fig. 1 Major substrates for lignifying CADs in plants. A generalized schematic showing the major aldehyde substrates available for conversion into alcohols (monolignols) by CADs involved in lignification. The ratios and amounts of each substrate and product are plant specific (adapted from Jung et al. [46])

motif rarely observed in other eukaryotes. Additional analysis and substrate modeling revealed a binding pocket that contained 12 mostly hydrophobic residues involved in substrate recognition and binding [17]. However, genetic analysis has shown plants contain relatively large CAD families whose functions are largely unknown but are unlikely to be involved in lignification [18–22]. In other studies [23, 24], domains and specific residues that differentiated between monocot and dicot CADs and bona fide CAD enzymes involved in lignification were identified. The goals of this research were to: (1) identify key protein residues that appear to differentiate lignifying CADs from ones that may have other primary biological roles and (2) establish an initial set of targets for protein engineering efforts towards altering CADs to accept novel substrates that could alter downstream lignin composition. Here, we used the wild type (WT) PviCAD1 enzyme, which is involved in lignification in switchgrass [23], as a template for generating several mutated proteins to understand the roles of specific residues in controlling catalytic rates and substrate preference.

Materials and Methods

PviCAD1

PviCAD1 was previously cloned into the pET28a vector (EMD Chemicals, Inc.) using publicly available EST resources as described previously [23]. A list of residues that were selected for site directed mutagenesis is presented in Table 1, along with a brief rationale of why a specific mutation was made. Primers were obtained from Bioneer and MWG Operon, and sequences are provided in Table 2. For site-directed mutagenesis experiments, the reverse primers were the exact complement of the forward primers listed in the table.

Generation of PviCAD1 Mutants

PviCAD1 mutants were generated using site-directed mutagenesis by generally following the Stratagene's Quick Change site-directed mutagenesis protocol. PviCAD1 in pET28a (EMD Chemicals, Inc.) was used as a template for all reactions except for the double mutant which used H57D as a template. Each PCR reaction consisted of 1 μL of each primer, 5 μL 10× reaction buffer, 39 μL diethyl pyrocarbonate (DEPC) water, 1 μL dNTPs, and 1 μL *Pfu* turbo polymerase (Invitrogen Corp.). Reactions were placed in a thermocycler (Biometra T-Gradient, Biomedizinishce Analytik, GmbH) and used the following cycling parameters:

Table 1 List of mutations and rationale

Mutation	Rationale
H57D	Mimics D58/57 seen in AtCAD4/5, part of proton shuttling motif
L58W	58 L expected to be involved in substrate binding; many CADs of unknown function have W at equivalent position
H57D/L58W	Double mutant; DW motif observed in some CADs
W119F	Altered residue thought to be important for substrate stabilization; some CADs contain equivalent F at this position
S120T	Many CADs contain an equivalent T at this position
D123S	All lignifying CADs have an invariant D; many other CADs contain an equivalent S

Table 2 List of primers used for PCR amplifications

Mutation	Primer sequence
H57D	gacatccaccaggccaagaacgacctcggcgcttccaagtacccc
L58W	gacatccaccaggccaagaaccactggggcgcttccaagtacccc
H57D/L58W	gacatccaccaggccaagaacgactggggcgcttccaagtacccc
W119F	gagcagtactgcaacaagaggatcttctcctacaacgacgtcta
S120T	cagtactgcaacaagaggatctggacctacaacgacgtctacactg
D123S	aggatctggtcctacaactccgtctacactgacggccgg
SbCAD6	
Forward	cgctataaagagagaggcaa
Reverse	tgatcatttgttgatccaag

95 °C (5 min), 55 °C (3 min.), and 72 °C (6 min) for 18 cycles. PCR reactions were then digested with 1 μL Dpn1 (20 U/μL, New England Biolabs) for a minimum of 4 h before ethanol precipitation to concentrate PCR product. The resulting DNA pellet was taken up in 10 μL DEPC water, and 1–7 μL were used in subsequent transformation into NovaBlue (EMD Chemicals, Inc.) or Top10F′ (Invitrogen Corp.) cells. For each mutant, several of the resultant colonies were grown up in Luria–Bertani (LB) media, and purified plasmids were sent off for sequence verification. After sequence verification, an appropriate plasmid stock was selected and transformed into Rosetta 2 cells (EMD Chemicals, Inc.) for subsequent protein expression.

SbCAD6 Cloning

Primers for amplification of SbCAD6 cDNA were designed using MacVector, and the sequences are provided in Table 2. PCR amplification used a cDNA library derived from *Sorghum bicolor* cv. Atlas. PCR product was verified by restriction digest, gel purified, and cloned into the pET30a vector (EMD Chemicals, Inc.) with another set of primers containing *EcoRI* and *XhoI* restriction sites.

Protein Expression and Purification

All proteins were expressed by growing Rosetta 2 cells in 250 or 400 mL of LB media to an OD_{600} of 0.4–0.6 and adding 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). After addition of IPTG, cultures were allowed to grow overnight in an incubator at 18 °C and shaken at 225 rpm. Cells were harvested by centrifugation at 2,800 rpm for 15 min, and pellets were washed twice in 20 mL 100 mM Tris-Cl, pH 7.5, centrifuged as indicated previously, and then taken up in 5–10 mL 100 mM Tris-Cl, pH 7.5, 5 mM dithiothreitol (DTT), and 5 % ethylene glycol. To release soluble protein, 100 mg lysozyme was added to the buffer, and the suspension was sonicated (Branson Digital Sonifier 450) on ice using 10 s bursts at 30 % amplitude with 30 s between bursts for 10 min. The suspension was then centrifuged at 2,800 rpm for 15 min at 4 °C, and supernatant was collected into new 50-mL plastic centrifuge tubes (VWR, Inc.). Protein purification used an AKTA Purifier 900 with a nickel column (GE Healthcare HisTrap FF) that bound the 6× His-tag on the N-terminus of the protein. The appropriate column fractions were pooled and dialyzed overnight at 4 °C into 1 L of 100 mM Tris-HCl, pH 7.5, 5 mM DTT, and 5 % ethylene glycol. Small aliquots (~150 μL) of dialyzed protein was placed into 1.7-mL tubes the next morning, flash frozen

using liquid N_2, and stored at −80 °C until needed. Protein purity was assessed using SDS-PAGE (12 % Bis-Tris Criterion XT, Bio-Rad Laboratories, Inc.) stained with Coomassie Brilliant Blue [25].

Enzyme Assays

Activity was determined by generally following previously published procedures [13, 23, 24, 26]. Protein activity was tested by monitoring absorbance changes on a BioTek Synergy HT plate reader at A_{325} (coumaryl aldehyde), A_{340} (sinapaldehyde and coniferaldehyde), A_{290} (*trans*-cinnamaldehyde), or A_{400} (coniferyl and coumaryl alcohols). All reactions were carried out in a volume of 200 μL and consisted of 100 mM buffer, 200 μM cofactor, substrate (1–200 μM depending on the experiment), and 10 μL of diluted enzyme. For aldehyde substrates, the buffer was 100 mM MES, pH 6.5, and nicotinamide adenosine dinucleotide phosphate (NADPH) was used as a cofactor; for alcohol substrates, the buffer was 100 mM Tris-HCl, pH 8.8, and $NADP^+$ was used as a cofactor. For each mutant, the appropriate amount of enzyme to use for activity tests was determined by finding a dilution that resulted in a linear change in absorbance over 3 min on 5 μM sinapaldehyde. For some of the mutants, higher enzyme amounts had to be used on alcohol substrates due to activity levels that were too low for reliable detection at the initial dilution factor that was determined using sinapaldehyde. During activity tests, enzymes were kept on ice, and new working dilutions were prepared every 10–15 min in order to minimize the effects of activity loss. Protein concentrations were determined using the Pierce 660 nm protein assay, and lysozyme was used to generate standard curves.

Computer Modeling

Protein modeling and visualization was done using PyMOL v. 1.4.1 to visualize the PviCAD1 pdb file. When examining the potential effects of site-directed mutagenesis, residues were mutated in PyMOL, and the most stable side-chain rotamer was used as the basis for structural analysis.

Sequence Alignments and Phylogenetic Trees

Sequence alignments were made using ClustalW 2.1 [27]. Phylogenetic analysis and the resulting trees were made using tools available at Phylogeny.fr [28]. In both cases, the default program settings were used.

Data Analysis

Statistical analyses were conducted using PROC GLM in SAS for Windows 9.2 (SAS Institute, Inc., Cary, NC). Nonlinear parameter estimation used routines available in SigmaPlot 11 (Systat Software, Inc., San Jose, CA). Where applicable, standard error propagation techniques were used to propagate error in calculations where two or more variables each had some associated uncertainty that contributed to the final result.

Results

Velocities for WT and mutant PviCAD1 proteins when tested on a variety of substrates are shown in Fig. 2. WT PviCAD1 displayed the highest velocity when coumaryl and sinapyl

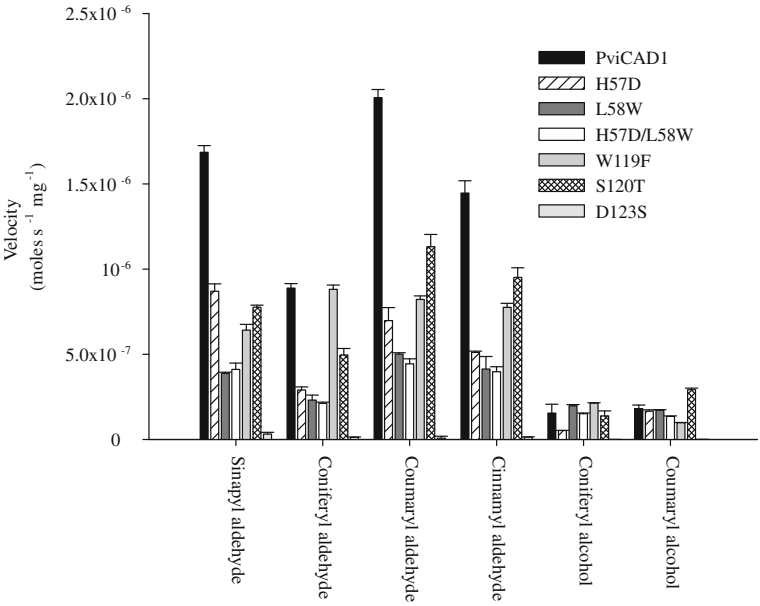

Fig. 2 WT CAD and mutant velocities on different substrates. Activity for each purified protein was monitored on five substrates that had an initial concentration of 100 μM. Reactions were allowed to proceed for 3 min and were conducted in quadruplicate. The mean velocity is shown in the *graph*, and *error bars* represent the standard deviation

aldehydes were used as substrates (2,000 and 1,690 nkat mg^{-1} protein, respectively) and exhibited approximately ~50 % of this velocity on coniferylaldehyde. Both coniferyl and coumaryl alcohols were poor substrates. In single concentration comparative tests and relative to WT catalytic rates, mutant H57D velocity on sinapaldehyde (52 %), coniferaldehyde (33 %), coumaryl aldehyde (35 %), *trans*-cinnamaldehyde (35 %), coumaryl alcohol (92 %), and coniferyl alcohol (34 %) was lower. Mutant L58W velocity was lower on sinapaldehyde (23 %), coniferaldehyde (26 %), coumaryl aldehyde (25 %), *trans*-cinnamaldehyde (29 %), and coumaryl alcohol (95 %) and higher on coniferyl alcohol (128 %). The H57D/L58W double mutant velocity showed a similar trend; sinapaldehyde (24 %), coniferaldehyde (24 %), coumaryl aldehyde (22 %), *trans*-cinnamaldehyde (27 %), and coumaryl alcohol (75 %) velocities were lower, while coniferyl alcohol velocity (98 %) was nearly identical to WT. The W119F mutation resulted in higher velocity on coniferyl alcohol (138 %) and lower velocities on sinapaldehyde (55 %), coniferaldehyde (87 %), coumaryl aldehyde (41 %), *trans*-cinnamaldehyde (54 %), and coumaryl alcohol (55 %). The S120T mutant displayed higher activity on coumaryl alcohol (162 %) and lower activities on sinapaldehyde (46 %), coniferaldehyde (56 %), coumaryl aldehyde (56 %), and *trans*-cinnamaldehyde (66 %). Mutant D123S displayed sharply reduced activity on all of the tested substrates.

The ratio of enzyme activity against different substrates normalized to its activity on sinapyl aldehyde is shown in Fig. 3. One-way ANOVA and Tukey's pairwise comparison procedure were used to compare the substrate velocities within each CAD. Therefore, the statistical model was designed to analyze velocities of different substrates of each CAD individually rather than across different CADs. Substrate velocity rankings for each protein are provided in Table 3. Overall, WT enzyme displayed an apparent preference for coumaryl

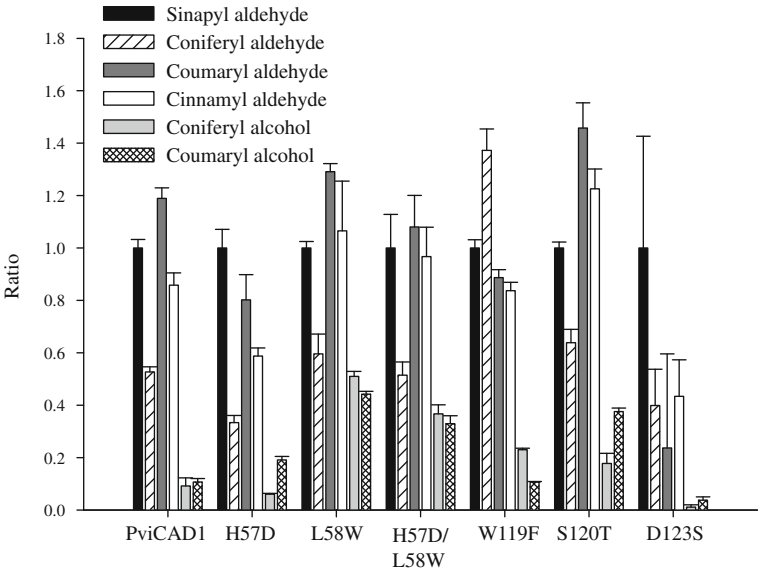

Fig. 3 Velocity results normalized to the sinapaldehyde rate for each enzyme. Velocities for each CAD, when normalized to their respective sinapaldehyde velocity, allowed for better representation of the substrate preferences of each enzyme, although absolute velocity differences between enzymes are subsumed. The *error bars* represent the standard deviation associated with the calculated ratio and was derived using standard error propagation techniques

aldehyde, while sinapaldehyde and *trans*-cinnamaldehyde velocities were somewhat lower; all of these aldehydes were strongly preferred over coniferaldehyde. Compared to the WT, velocity rankings for mutant H57D were slightly altered because activity on sinapaldehyde was higher than on all other aldehydes. Mutants L58W and H57D/L58W both showed a preference for coumaryl aldehyde followed by relatively similar velocities on *trans*-cinnamaldehyde and sinapaldehyde; alcohol velocities were also relatively higher to the aldehyde velocities of these mutants. The W119F substrate profile showed a preference for coniferaldehyde over all other aldehydes, and *trans*-cinnamaldehyde was the least preferred aldehyde which was unique among the tested mutations. The S120T mutation resulted in a protein with a similar substrate profile to WT CAD, although *trans*-cinnamaldehyde was

Table 3 Ranking of substrate velocities for each mutation

Substrate	PviCAD1	H57D	L58W	H57D/L58W	W119F	S120T	D123S
Sinapaldehyde	2 b	1 a	3 b	2 a	4 c	3 c	1 a
Coniferaldehyde	4 d	4 d	4 c	4 b	1 a	4 d	4 b
Coumaryl aldehyde	1 a	2 b	1 a	1 a	2 b	1 a	3 b
trans-Cinnamaldehyde	3 c	3 c	2 b	3 a	3 b	2 b	2 b
Coniferyl alcohol	6 e	6 f	5 c	5 c	5 d	6 f	ND
Coumaryl alcohol	5 e	5 e	6 c	6 c	6 e	5 e	5 c

Values with the same letter reflect velocities that were not statistically different based on Tukey's HSD test (α=0.05). Comparisons are valid for values compared within each CAD, not across different CADs

ND not detected

more preferred than sinapaldehyde in the S120T mutant. Large error values for the calculated D123S ratios made determination of preferences among the tested aldehyde substrates difficult to accurately measure, although activity on hydroxycinnamylaldehydes was still higher than on hydroxycinnamyl alcohols. In all cases, CAD velocities for WT or any mutant protein on alcohol substrates were lower than their respective aldehyde velocities; however, for PviCAD1, L58W, and H57D/L58W, the coniferyl alcohol and coumaryl alcohol rates were statistically indistinguishable within each CAD, respectively.

In order to better understand the apparent velocity differences among the engineered CADs, enzyme kinetics were investigated for each CAD on two substrates: sinapaldehyde and coumaryl alcohol. These substrates were chosen because previous work demonstrated that they were suitable reference substrates for switchgrass CAD [23]. In all cases, a Michaelis–Menten kinetic model was fit to the data, and parameter estimates of the kinetic constants K_m, V_{max}, k_{cat}, and k_{cat}/K_m are shown in Table 4. Relative to the WT protein, for sinapaldehyde, all of the CAD mutants were found to have lower estimated V_{max} values by at least a factor of 4.8; however, the sinapaldehyde K_m values for the H57D and S120T mutants were relatively similar to WT, while the L58W, H57D/L58W, and W119F mutations resulted in much higher K_m values for this substrate. Relative to the WT, most of the CAD mutants also had lower V_{max} values on coumaryl alcohol, with the exception being S120T which had a similar value. Estimated K_m values for coumaryl alcohol were low (below 10 μM) for WT PviCAD1 and all of the mutants except for the W119F mutant which had the highest estimated K_m of 13.8 μM on this substrate.

The D123S mutant exhibited surprisingly low activity on monolignal substrates. Therefore, the PviCAD1 structure was visualized in PyMOL, and Asp123 was changed to a Ser in order to find a potential explanation for this result (Fig. 4). The electrostatic interaction between the Asp123 side chain (partially transparent red bonds) and the Trp119 is characterized by contact distances of 2.9 Å (between Asp123 Od2 and the Trp119 amide N) and 4.3 Å (between Asp123 Od2 and the Trp119 Ne). When the Asp123 side chain is replaced by that of Ser by mutation in PyMOL (opaque red bonds), the rotamer that best accommodates the comparable electrostatic interactions with Trp119 has bond distances of 3.3 and 4.7 Å, respectively. Thus, Ser at this position has a diminished capacity for positive electrostatic interactions with Trp119.

Table 4 Parameter estimates for kinetic constants

Protein	Substrate	V_{max} ($mol\,s^{-1}\,mg^{-1}$)	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($s^{-1}\,\mu M^{-1}$)
PviCAD1	Sinapaldehyde	$2.53\times10^{-6}\pm9.89\times10^{-8}$	14.5±1.85	207.3	14.30
	Coumaryl alcohol	$3.18\times10^{-7}\pm1.82\times10^{-8}$	4.2±1.15	26.03	6.20
H57D	Sinapaldehyde	$9.77\times10^{-7}\pm1.56\times10^{-8}$	12.4±0.82	80.02	6.47
	Coumaryl alcohol	$1.97\times10^{-7}\pm8.04\times10^{-9}$	7.35±1.27	16.18	2.20
L58W	Sinapaldehyde	$5.23\times10^{-7}\pm3.13\times10^{-8}$	46.6±7.7	42.82	0.92
	Coumaryl alcohol	$1.88\times10^{-7}\pm5.39\times10^{-9}$	7.39±0.90	15.36	2.08
H57D/L58W	Sinapaldehyde	$5.56\times10^{-7}\pm3.71\times10^{-8}$	41.7±8.0	45.56	1.09
	Coumaryl alcohol	$1.53\times10^{-7}\pm4.46\times10^{-9}$	8.87±1.04	12.53	1.41
W119F	Sinapaldehyde	$1.23\times10^{-6}\pm3.27\times10^{-8}$	42.3±3.2	100.5	2.38
	Coumaryl alcohol	$1.16\times10^{-7}\pm5.26\times10^{-9}$	13.8±2.11	9.52	0.69
S120T	Sinapaldehyde	$8.05\times10^{-7}\pm4.37\times10^{-8}$	14.7±2.9	65.9	4.48
	Coumaryl alcohol	$3.22\times10^{-7}\pm1.41\times10^{-8}$	7.30±1.36	26.38	3.62

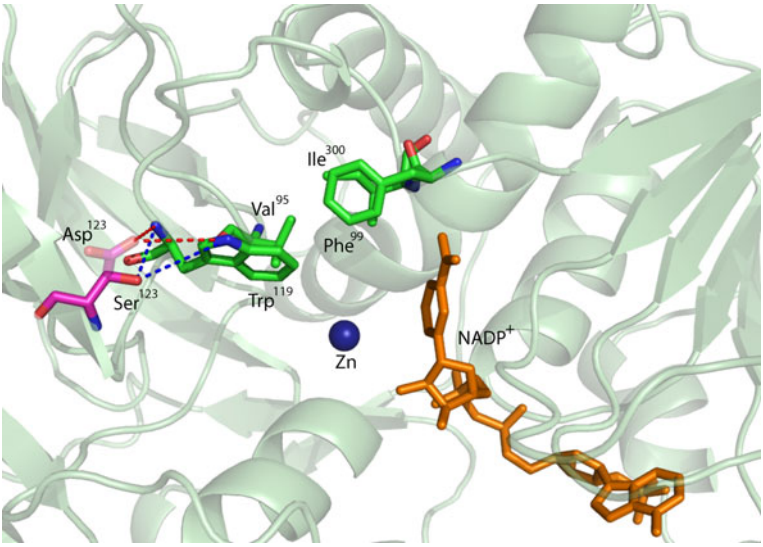

Fig. 4 Molecular model of the CAD-binding pocket. This model shows the potential changes that occurred in the D123S mutant, which resulted in the substantial velocity reduction observed for all tested substrates. Critically, in the D123S mutant, local hydrogen-bonding patterns may have been altered that resulted in a different orientation of Trp119, which appears to be involved in substrate recognition and stabilization

An alignment of selected CAD residues, which was generated using CLUSTAL 2.1, is shown in Fig. 5. The first 13 sequences represent proteins that have been shown to have high activity on monolignol substrates and appear to have roles in lignification. One ancestral CAD from *Selaginella* also grouped with the lignifying CADs and appeared to share many of their important features. The next 13 sequences consist of CAD-like proteins with generally unknown functions and substrate preferences, which were included to highlight differences between them and lignifying CADs. The residues highlighted in black include most of the AtCAD5 substrate-binding residues that were previously identified by Youn et al. [17]. The sequence alignment indicated that there were clear differences in binding site residues at nearly all positions between the lignifying and nonlignifying CADs.

The conservation of key residues in all bona fide CADs and the enzymatic data obtained from mutational studies described above, which suggested detailed phylogenetic analyses of the CADs and CAD-like proteins as shown in Fig. 5, could yield more insights into the relationships of these proteins and the potential of CAD-like enzymes to participate in lignification reactions. A phylogenetic tree of these proteins is shown in Fig. 6. The CAD and CAD-like sequences analyzed yielded a tree that could easily be distinguished into three clades, namely, lignifying CADs, nonlignifying CADs (almost all the CAD-like sequences analyzed), and an ancestral group containing CAD-like sequences from *Selaginella* and *Physcomitrella*. One of the *Selaginella* sequences (National Center for Biotechnology Information (NCBI) accession number XM_002993954.1) clustered closer to the lignifying CADs and was included in that clade. The lignifying CADs could be further differentiated into the monocot and dicot clades (Fig. 6) that were consistent with earlier published work [15, 18, 23, 24].

We had shown that switchgrass [23] and sorghum [24] contained two and one CADs, respectively, that appeared to be involved in lignification in stems. These species contained additional CAD-like proteins, namely, SbCAD4 from sorghum and PviAroADH from

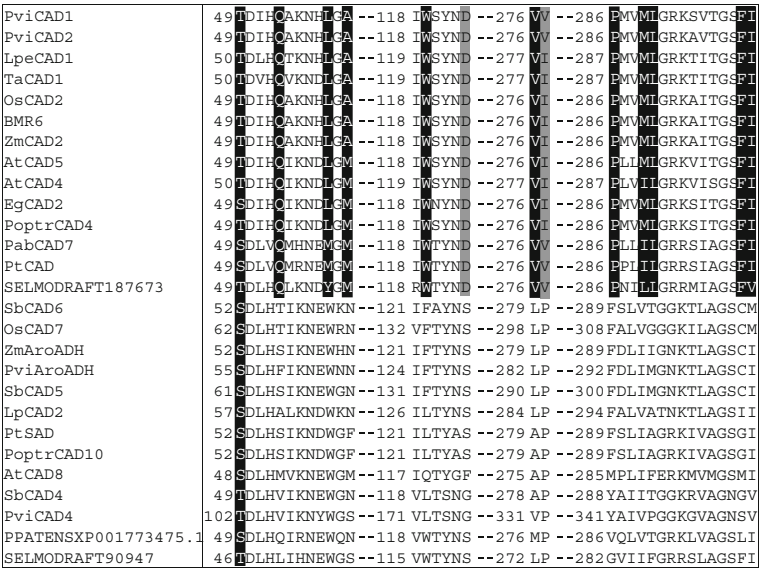

Fig. 5 Alignment of selected CAD protein residues. The first 12 sequences are from proteins likely to be involved in monolignol biosynthesis, and many have been shown to have high activity on monolignol substrates. The next 12 sequences are from other CADs that appear less likely to be involved in lignification. The residues *highlighted in black* are residues that were predicted to be important in AtCAD5 substrate binding by Youn et al. [17]. Other potentially important or distinguishing residues are *highlighted in gray*. NCBI nucleotide accession numbers for each gene are as follows: PviCAD1 (GU045611.1), PviCAD2 (GU045612.1), SbCAD6 (XM_002447037.1), TaCAD1 (GU563724.1), OsCAD2 (NM_001052667.1), BMR6 (FJ554574.1), ZmCAD2 (NM_001112184.1), AtCAD5 (NM_119587.3), AtCAD4 (NM_112832.3), EgCAD2 (P31655), PoptrCAD4 (XM_002313839.1), PabCAD7 (Q08350), PtaCAD1 (P41637), SELMODRAFT187673 (XM_002993954.1), SbCAD6 (XM_002447037.1), OsCAD7 (NM_001060383.1), ZmAroADH (NM_001154254.1), SbCAD5 (XM_002445169.1), LpCAD2 (AF472592.1), PtSAD (AF273256.1), PoptrCAD10 (XM_002322786.1), AtCAD8 (NM_119960.2), SbCAD4 (XM_002462303.1), PPATENSXP001773475.1 (XM_001773423.1), and SELMODRAFT90947 (XM_002968611.1)

switchgrass, which grouped to the nonlignifying clade (Fig. 6) and, in the case of SbCAD4, exhibited low or negligible activity against monolignol substrates. Here, we tested another sorghum CAD-like protein, SbCAD6, which appears to be the sorghum ortholog of rice flexible culm (FC1) protein. The rice flexible culm mutation was mapped to a CAD-like gene within the rice genome [29]. Plants carrying this mutation showed increased culm flexibility, reduced cell wall thickness, lowered lignin content, and lowered total CAD activity in internode extracts. However, the sequence for this protein did not have the signature residues at equivalent positions, including Gln53, His57 or Asp57, Trp119, Asp123, Val276, Pro287, Leu290, and Phe299, which appear to be invariant in biochemically characterized lignifying CADs (Fig. 5). The recombinant SbCAD6 protein was analyzed for enzymatic activity. On all of the tested substrates, SbCAD6 velocity was markedly lower when compared to the velocity of the switchgrass WT protein (Fig. 7).

Discussion

For biochemically based conversion platforms, development of feedstocks that have high conversion efficiency into renewable fuels and chemicals is an important goal. Lignin

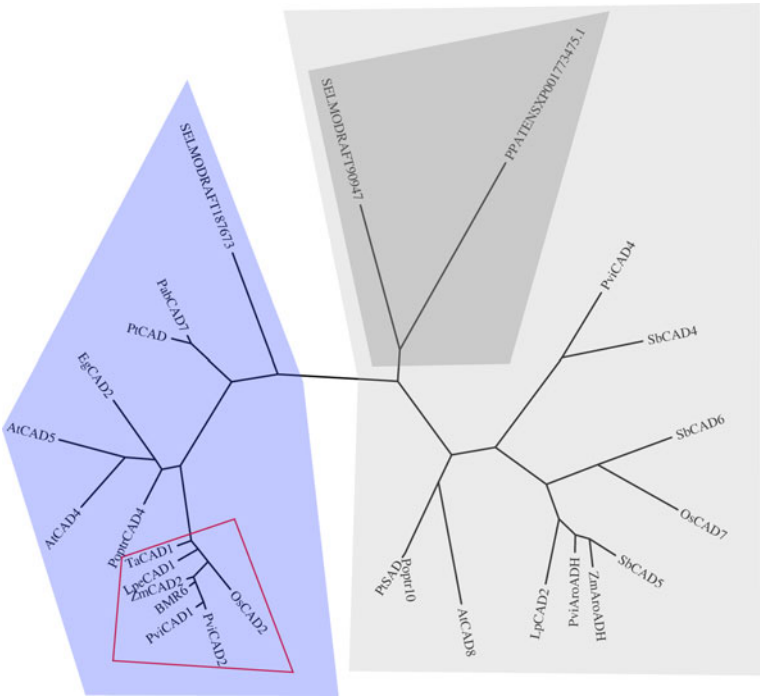

Fig. 6 Phylogenetic tree of CAD protein sequences. The sequences used here are the full sequences of the CADs used in the partial sequence alignment shown in Fig. 4. Sequence grouping: *blue polygon*=lignifying CADs, *light gray polygon*=nonlignifying CADs, *dark gray polygon*=ancestral CADs, and *red polygon*=monocot CADs

content remains an important bottleneck for biochemical conversion into liquid fuels [30], and lignin reduction or alteration has been shown to improve liquid fuel yields [31, 32]. Considerable efforts are being devoted towards understanding how plants construct and maintain their cell walls for selecting targets for future manipulation using both conventional breeding and genetic engineering strategies that have a goal of producing plants with cell walls that are easier to convert into liquid fuels, such as ethanol. Lignin biosynthesis, including the relevant CADs, represents one such target, although the role(s) of the many other plant CAD-like proteins is unclear. Previous work on switchgrass indicated that it contained two bona fide CADs involved in monolignol biosynthesis [23, 33]. This finding was in agreement with other work that has demonstrated that one or a very limited number of CAD genes have a primary physiological role in cell wall lignification. For instance, in sorghum, a mutation in SbCAD2 was found to be responsible for a brown midrib phenotype [21, 24], while in *Arabidopsis*, only two CADs had meaningful activity against 4-hydroxycinnamylaldehyde substrates, and the same two CADs were critical for maintaining suitable stem lignin levels [20, 34]. In other plants, CAD disruption has led to one or more of the following: altered phenotypes, altered levels of cell wall solubles, altered lignin content or composition, and/or higher digestibility or saccharification efficiency [35–38]. Transgenic work in monocots has demonstrated that CAD downregulation resulted in lower (Klason) lignin levels and higher IVDMD in tall fescue [39] and lowered total lignin levels and improvements in saccharification efficiency in switchgrass, although no red coloration of plant tissue was noted [33, 40]. Understanding the biological function(s) and specific

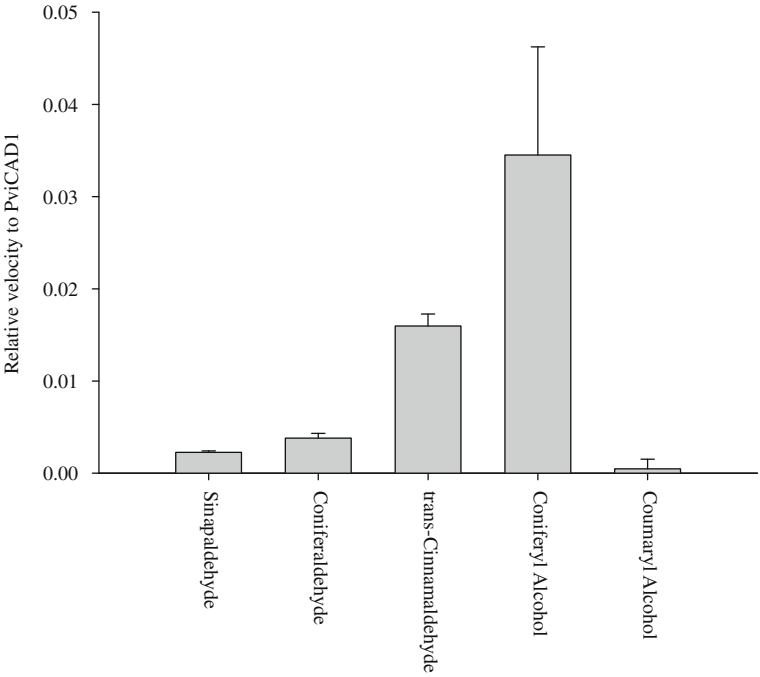

Fig. 7 Activity of SbCAD6 on several different substrates. The *displayed values* represent the ratio of SbCAD6 velocity to WT PviCAD1 velocity on the same substrate. Substrate concentration was 100 μM. *Error bars* represent the standard deviation

residues that govern substrate specificity, particularly of the CADs that are involved in cell wall lignification, is an important part of overall efforts towards understanding cell wall synthesis and potential manipulation for easier deconstruction of plants for renewable fuels and chemicals [41, 42].

The crystal structure of a dicot CAD, AtCAD5 [17], revealed a binding site that consisted of 12 residues: Thr49, Gln53, Leu58, Met60, Cys95, Trp119, Val276, Pro286, Met289, Leu 290, Phe299, and Ile300. Additionally, Youn et al. [17] proposed a proton shuttling mechanism that included Thr49, His52, and Asp57. In at least some monocots, this shuttling mechanism has been shown to apparently use His57 rather than Asp57 [23, 24]. An earlier model of *Eucalyptus gunnii* CAD2, based on the structure of horse liver alcohol dehydrogenase, suggested that Phe299 and Trp119 were important for stabilizing the phenolic ring of hydroxycinnamylaldehyde substrates [43]. Additionally, Cys95 in AtCAD5, which corresponds to Val96 in AtCAD4, Ile95 in EgCAD2, and Val95 in PviCAD1, was suggested to also be involved in determining substrate specificity [17].

Earlier site-directed mutagenesis of CADs showed the importance of Ser212 in cofactor binding [44] and Glu70 as a critical residue for coordination of the catalytic zinc residue [17]. In this study, residues were specifically chosen that either mimicked the corresponding residues in apparent nonlignifying CADs or changed the characteristics of the substrate-binding pocket (see Table 1). Two of the most conservative mutations H57D and S120T had relatively modest effects on activity. Changing His57 to Asp resulted in lower velocities on all tested substrates except coumaryl alcohol, although preference for coumaryl aldehyde appeared to be reduced in the H57D mutant and enhanced in the S120T mutant. Dicots

generally appear to use an Asp for proton shuttling at position 57, while many monocots seem to use His for this purpose. The H57D results may thus be a reflection of structural optimization in PviCAD1and other monocot lignifying CADs for use of His rather than Asp. The other highly conservative mutation, S120T, was chosen because our analysis revealed that all known bona fide CADs with a primary physiological role in lignification contain Ser at this position, while many other CADs have an equivalent Thr. This division was also apparent in previous analysis of plant-specific CAD families in *Arabidopsis* and sorghum [20, 21]. Changing Ser120 to Thr resulted in lower velocities against cinnamylaldehydes when compared to WT, although the reductions for coniferyl, coumaryl, and *trans*-cinnamylaldehydes were less pronounced than in the H57D mutant. However, S120T apparent efficiency (k_{cat}/K_m) was still lower than WT on these substrates, which again indicated structural optimization for Ser over Thr.

Three other mutations in the substrate-binding pocket had stronger impacts on velocity and apparent substrate preferences. The first of these, L58W, was chosen because many CADs contain a Trp at this equivalent position, while CADs with confirmed high catalytic rates against monolignal substrates have a nearly invariant Leu (see Fig. 5). This mutation, while not changing polarity within the active site, reduced velocity on all of the aldehyde substrates and sharply increased the sinapaldehyde estimated K_m, which suggested that the bulky Trp side chain was hindering substrate movement into the binding pocket. Molecular modeling confirmed that this was likely the case as the Trp side chain appeared to protrude into and shrink the overall size of the binding pocket. The reduced binding pocket size also altered apparent substrate preferences since activity on coumaryl aldehyde and *trans*-cinnamaldehyde was higher than both sinapyl and coniferyl aldehydes. The double mutant, H57D/L58W, was created based on the fact that many CADs have an equivalent DW or EW sequence motif at the H57/L58 positions when compared to WT. The double mutant displayed a nearly identical activity and kinetic profile to L58W, which suggested that the Trp was clearly the dominant factor in determining substrate velocities and preferences. Notably, two conifer CAD sequences, PtaCAD [45] and PabCAD7 [46], have a 57EM58 sequence at this position. Because gymnosperm CADs have displayed very low catalytic activity on sinapaldehyde [14, 47], it is possible that the HL or DL motif present in monocot and eudicot CADs may have evolved to increase activity on a range of monolignals, including sinapaldehyde and *p*-coumaryl aldehyde.

The W119F mutation was selected because all known lignifying CADs have a Trp at this position, while the equivalent position appears to be more variable in other CADs. As previously noted, W119 is expected to serve an important role in stabilizing the six-carbon ring structure of hydroxycinnamylaldehyde substrates, and thus, a residue change at this position could alter the overall enzyme activity and substrate preferences. This was indeed the case: the W119F mutation resulted in a protein with generally higher catalytic rates than the other single (H57D) or double (H57D/L58W) active site mutants. Additionally, the W119F mutation had a strong influence on substrate preferences since this change resulted in a protein with a higher velocity on coniferaldehyde than on sinapaldehyde and a clear preference for coniferyl alcohol over coumaryl alcohol, which was unique among the tested CAD mutants. Kinetically, the higher estimated K_m values, coupled with lower turnover numbers and lower k_{cat}/K_m values, argue for Trp119, serving an important role in helping to stabilize substrate interactions, which was suggested by earlier CAD structural studies [17, 43].

One mutation, D123S, resulted in a protein that was essentially inactive. This was an unexpected finding since Asp123 was not predicted to be involved in substrate binding or catalysis [17]. Therefore, it is apparent that Asp123 is critical to WT activity and, likely, all

bona fide lignifying CADs since this residue appears to be absolutely conserved among them. Many other CADs have a Ser at this equivalent position, which may be equally important for maintaining native activity, although other residues, including Gly and Phe, were also present in some CADs at this site. In an analysis of sorghum CADs, SbCAD7 and SbCAD5 contained Ser at this equivalent site, while other CAD family members contained Gly [21], which was in agreement with the results obtained here. Molecular modeling suggested that localized hydrogen-bonding patterns were altered in the D123S mutant. This change in hydrogen bonding may have resulted in a localized conformational change that prevented Trp119 from serving its important substrate stabilizing function through aromatic stacking interactions, resulting in the observed dramatically lower activity rates. A crystal structure of this mutant is needed, though, to confirm this hypothesis.

The biochemical results with the sorghum ortholog to rice FC1 confirmed the phylogenetic analysis of lignifying CADs and the mutation work reported here (Fig. 6 and [23, 24]). Our results would suggest that CADs with sequence similarities to rice FC1 or sorghum SbCAD6 are unlikely to utilize monolignals as effective substrates. The significant divergence in amino acid sequences between the lignifying CADs relative to rice FC1 would suggest that key amino acid substitutions, along the entire length of protein, would be needed to convert such CAD orthologs into enzymes that could efficiently convert monolignals into monolignols. Since Li et al. [29] did not biochemically characterize the rice protein, it is unclear if this rice CAD-like protein might exhibit different substrate specificities than its sorghum ortholog. Overall, the data presented here underscore the changes in amino acid sequence that has occurred in the bona fide lignifying CADs as compared to the other CADs and CAD-like genes present in most plant genomes. It also suggests that it should be possible to rapidly identify the potential lignifying CADs in as yet unexplored plant genomes. Future engineering of CADs to potentially accept novel lignin monomers would need to consider changes to the binding pocket that maintain key residues, such as W119 and D123, in an optimal position for substrate binding, while assuring efficient transfer of electrons through proton shuttling from NADPH.

Acknowledgments We thank Nathan Palmer for his excellent technical assistance. This work was supported by the USDA-ARS CRIS project 5440-21000-028-00D and in part by the Office of Science (BER), US Department of Energy grant number (DE-AI02-09ER64829). The US Department of Agriculture, Agricultural Research Service, is an equal opportunity/affirmative action employer, and all agency services are available without discrimination. Mention of commercial products and organizations in this manuscript is solely to provide specific information. It does not constitute endorsement by USDA-ARS over other products and organizations not mentioned.

References

1. Field, C. B., Behrenfeld, M. J., Randerson, J. T., & Falkowski, P. (1998). *Science, 281*, 237–240.
2. Delmer, D. P., & Amor, Y. (1995). *Plant Cell, 7*, 987–1000.
3. Somerville, C. (2006). *Annual Review of Cell and Developmental Biology, 22*, 53–78.
4. Pauly, M., & Keegstra, K. (2008). *Plant Journal, 54*, 559–568.
5. Boerjan, W., Ralph, J., & Baucher, M. (2003). *Annual Review of Plant Biology, 54*, 519–546.
6. Chapple, C., Ladisch, M., & Meilan, R. (2007). *Nature Biotechnology, 25*, 746–748.
7. Sarath, G., Mitchell, R. B., Sattler, S. E., Funnell, D., Pedersen, J. F., Graybosch, R. A., & Vogel, K. P. (2008). *Journal of Industrial Microbiology & Biotechnology, 35*, 343–354.
8. Li, X., Weng, J. K., & Chapple, C. (2008). *Plant Journal, 54*, 569–581.
9. Dixon, R. A., Chen, F., Guo, D. J., & Parvathi, K. (2001). *Phytochemistry, 57*, 1069–1084.
10. Vogel, J. (2008). *Current Opinion in Plant Biology, 11*, 301–307.
11. Boudet, A. M., Lapierre, C., & Grimapettenati, J. (1995). *New Phytologist, 129*, 203–236.

12. Hisano, H., Nandakumar, R., & Wang, Z. Y. (2009). *In Vitro Cellular & Developmental Biology - Plant, 45*, 306–313.
13. Hawkins, S. W., & Boudet, A. M. (1994). *Plant Physiology, 104*, 75–84.
14. Luderitz, T., & Grisebach, H. (1981). *European Journal of Biochemistry, 119*, 115–124.
15. Ma, Q. H. (2010). *Journal of Experimental Botany, 61*, 2735–2744.
16. McAlister, F. M., Lewis-Henderson, W. R., Jenkins, C. L. D., & Watson, J. M. (2001). *Australian Journal of Plant Physiology, 28*, 1085–1094.
17. Youn, B., Camacho, R., Moinuddin, S. G. A., Lee, C., Davin, L. B., Lewis, N. G., & Kang, C. H. (2006). *Organic & Biomolecular Chemistry, 4*, 1687–1697.
18. Barakat, A., Bagniewska-Zadworna, A., Choi, A., Plakkat, U., DiLoreto, D. S., Yellanki, P., & Carlson, J. E. (2009). *BMC Plant Biology, 9*, 26.
19. Costa, M. A., Collins, R. E., Anterola, A. M., Cochrane, F. C., Davin, L. B., & Lewis, N. G. (2003). *Phytochemistry, 64*, 1097–1112.
20. Kim, S.-J., Kim, M.-R., Bedgar, D. L., Moinuddin, S. G. A., Cardenas, C. L., Davin, L. B., Kang, C., & Lewis, N. G. (2004). *Proceedings of the National Academy of Sciences of the United States of America, 101*, 1455–1460.
21. Saballos, A., Ejeta, G., Sanchez, E., Kang, C., & Vermerris, W. (2009). *Genetics, 181*, 783–795.
22. Tobias, C. M., & Chow, E. K. (2005). *Planta, 220*, 678–688.
23. Saathoff, A. J., Tobias, C. M., Sattler, S. E., Haas, E. J., Twigg, P., & Sarath, G. (2011). *BioEnergy Research, 4*, 120–133.
24. Sattler, S. E., Saathoff, A. J., Haas, E. J., Palmer, N. A., Funnell-Harris, D. L., Sarath, G., & Pedersen, J. F. (2009). *Plant Physiology, 150*, 584–595.
25. Laemmli, U. K. (1970). *Nature, 227*, 680–685.
26. Mansell, R. L., Gross, G. G., Stockigt, J., Franke, H., & Zenk, M. H. (1974). *Phytochemistry, 13*, 2427–2435.
27. Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., & Thompson, J. D. (2003). *Nucleic Acids Research, 31*, 3497–3500.
28. Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J. F., Guindon, S., Lefort, V., Lescot, M., Claverie, J. M., & Gascuel, O. (2008). *Nucleic Acids Research, 36*, W465–W469.
29. Li, X. J., Yang, Y., Yao, J. L., Chen, G. X., Li, X. H., Zhang, Q. F., & Wu, C. Y. (2009). *Plant Molecular Biology, 69*, 685–697.
30. Carroll, A., & Somerville, C. (2009). *Annual Review of Plant Biology, 60*, 165–182.
31. Dien, B., Sarath, G., Pedersen, J., Sattler, S., Chen, H., Funnell-Harris, D., Nichols, N., & Cotta, M. (2009). *BioEnergy Research, 2*, 153–164.
32. Chen, F., & Dixon, R. A. (2007). *Nature Biotechnology, 25*, 759–761.
33. Saathoff, A. J., Sarath, G., Chow, E. K., Dien, B. S. and Tobias, C. M. (2011). *PLoS One*, *6*.
34. Sibout, R., Eudes, A., Mouille, G., Pollet, B., Lapierre, C., Jouanin, L., & Seguin, A. (2005). *Plant Cell, 17*, 2059–2076.
35. Baucher, M., Bernard-Vailhe, M. A., Chabbert, B., Besle, J. M., Opsomer, C., Van Montagu, M., & Botterman, J. (1999). *Plant Molecular Biology, 39*, 437–447.
36. Baucher, M., Chabbert, B., Pilate, G., VanDoorsselaere, J., Tollier, M. T., PetitConil, M., Cornu, D., Monties, B., VanMontagu, M., Inze, D., Jouanin, L., & Boerjan, W. (1996). *Plant Physiology, 112*, 1479–1490.
37. Halpin, C., Knight, M. E., Foxon, G. A., Campbell, M. M., Boudet, A. M., Boon, J. J., Chabbert, B., Tollier, M. T., & Schuch, W. (1994). *Plant Journal, 6*, 339–350.
38. Jackson, L. A., Shadle, G. L., Zhou, R., Nakashima, J., Chen, F., & Dixon, R. A. (2008). *BioEnergy Research, 1*, 180–192.
39. Chen, L., Auh, C.-K., Dowling, P., Bell, J., Chen, F., Hopkins, A., Dixon, R. A., & Wang, Z.-Y. (2003). *Plant Biotechnology Journal, 1*, 437–449.
40. Fu, C. X., Xiao, X. R., Xi, Y. J., Ge, Y. X., Chen, F., Bouton, J., Dixon, R. A., & Wang, Z. Y. (2011). *BioEnergy Research, 4*, 153–164.
41. Grabber, J. H., Schatz, P. F., Kim, H., Lu, F. C., & Ralph, J. (2010). *BMC Plant Biology, 10*, 13.
42. Vanholme, R., Morreel, K., Ralph, J., & Boerjan, W. (2008). *Current Opinion in Plant Biology, 11*, 278–285.
43. McKie, J. H., Jaouhari, R., Douglas, K. T., Goffner, D., Feuillet, C., Grimapettenati, J., Boudet, A. M., Baltas, M., & Gorrichon, L. (1993). *Biochimica Et Biophysica Acta, 1202*, 61–69.
44. Lauvergeat, V., Kennedy, K., Feuillet, C., Mckie, J. H., Gorrichon, L., Baltas, M., Boudet, A. M., Grimapettenati, J., & Douglas, K. T. (1995). *Biochemistry, 34*, 12426–12434.
45. MacKay, J. J., Liu, W. W., Whetten, R., Sederoff, R. R., & Omalley, D. M. (1995). *Molecular & General Genetics, 247*, 537–545.
46. Schubert, R., Sperisen, C., Muller-Starck, G., La Scala, S., Ernst, D., Sandermann, H., & Hager, K. P. (1998). *Trees-Structure and Function, 12*, 453–463.
47. Kutsuki, H., Shimada, M., & Higuchi, T. (1982). *Phytochemistry, 21*, 19–23.