

# Characterization of three cinnamyl alcohol dehydrogenases from *Carthamus tinctorius*

Safendrri Komara Ragamustari<sup>1,2</sup>, Naoko Shiraiwa<sup>1</sup>, Takefumi Hattori<sup>1,3</sup>,  
Tomoyuki Nakatsubo<sup>1</sup>, Shiro Suzuki<sup>1</sup>, Toshiaki Umezawa<sup>1,2,\*</sup>

<sup>1</sup>Research Institute for Sustainable Humanosphere, Kyoto University, Uji, Kyoto 611-0011, Japan; <sup>2</sup>Institute of Sustainability Science, Kyoto University, Uji, Kyoto 611-0011, Japan; <sup>3</sup>Institute of Socio-Arts and Sciences, The University of Tokushima, Tokushima 770-8502, Japan

\*E-mail: tumezawa@rish.kyoto-u.ac.jp Tel: +81-774-38-3625 Fax: +81-774-38-3682

Received December 26, 2012; accepted February 6, 2013 (Edited by T. Demura)

**Abstract** We report the isolation and characterization of three cDNAs encoding cinnamyl alcohol dehydrogenase (CAD) from *Carthamus tinctorius* (safflower). All three recombinant CADs were able to reduce coniferaldehyde and sinapaldehyde into coniferyl alcohol and sinapyl alcohol, respectively, and were designated as CtCAD1, CtCAD2, and CtCAD3. Phylogenetic analysis of CAD amino acid sequences and homology modeling revealed that CtCAD1 and CtCAD3 were closely related to the sinapaldehyde-specific aspen (*Populus tremuloides*) sinapyl alcohol dehydrogenase (PtresAD). CtCAD2 was in a clade containing class I plant CADs. Gas chromatography-mass spectrometry-based kinetic analysis using two different substrates, coniferaldehyde and sinapaldehyde, indicated that the CtCADs showed no strong preference for either substrate. CtCAD2 has the highest catalytic efficiency ( $k_{cat}/K_m$ ) (81.49 mM<sup>-1</sup> min<sup>-1</sup> and 95.3 mM<sup>-1</sup> min<sup>-1</sup> for coniferaldehyde and sinapaldehyde, respectively) compared with the other CtCADs. Inhibition kinetics showed that coniferaldehyde was a stronger inhibitor than sinapaldehyde for all CtCADs. Quantitative real-time PCR revealed that CtCAD2 was expressed at higher levels than CtCAD1 and CtCAD3 in all samples, except developing seeds at 3 days after flowering, where CtCAD1 had a higher expression level. In plant protein assays with coniferaldehyde and sinapaldehyde, plant protein extracted from seeds at 7 days after flowering, showed the highest specific activity. The product yields in plant protein assays were strongly correlated with gene expressions of CtCAD2 and CtCAD3 in the respective organs.

**Key words:** CAD, lignin biosynthesis, *Carthamus tinctorius*, coniferaldehyde, sinapaldehyde.

Lignin is a major component of the secondary cell wall of vascular plants, where it fills the spaces between cell wall polysaccharides. It confers mechanical strength and makes the cell wall impervious (Boerjan et al. 2003). Therefore, this biopolymer makes an important contribution to the overall development of vascular plants. Lignin is biosynthesized via oxidative coupling of *p*-hydroxycinnamyl alcohols (monolignols) and related compounds formed in the cinnamate/monolignol pathway (Umezawa 2010). Cinnamyl alcohol dehydrogenase (CAD) is an NADPH-dependent enzyme responsible for the last reductive

step in the monolignol biosynthesis pathway, in which *p*-hydroxycinnamaldehydes, such as coniferaldehyde and sinapaldehyde, are reduced into their respective monolignols (Higuchi 2006; Kutsuki et al. 1982; Umezawa 2010) (Figure 1). *p*-Hydroxycinnamyl alcohols are precursors of lignin, and are also the precursors of phenylpropanoid dimers known as lignans and neolignans, which are of interest mainly because of their medicinal values (Umezawa 2003).

Lignin formation in gymnosperms involves only one CAD isoform per species (Li et al. 2012; Ma 2010; Mackay et al. 1995; O'Malley et al. 1992). Individual

Abbreviations: CAD, cinnamyl alcohol dehydrogenase; CoAOMT, *p*-hydroxycinnamoyl CoA *O*-methyltransferase; GCMS, gas chromatography-mass spectrometry; MAFFT, multiple alignment using fast Fourier transform; OMT, *O*-methyltransferase; PCR, polymerase chain reaction; qRT-PCR, quantitative real-time PCR; SAD, sinapyl alcohol dehydrogenase.

The accession numbers of the cDNAs used in this research are as follows: *Apium graveolens mannitol dehydrogenase* (MTD), AF067082; *Arabidopsis thaliana CAD1*, BT008840; *Arabidopsis thaliana CAD5*, NM\_119587; *Arabidopsis thaliana ELI3-2*, X67815; *Camptotheca acuminata 10-hydroxygeraniol oxidoreductase* (HGO), AY342355; *Carthamus tinctorius CAD1*, AB576772; *Carthamus tinctorius CAD2*, AB576773; *Carthamus tinctorius CAD3*, AB576774; *Eucalyptus gunnii CAD2*, X65631; *Eucalyptus globulus CAD*, AF038561; *Fragaria×ananassa* cv. Chandler CAD1, AF320110; *Medicago sativa CAD*, AF083332; *Nicotiana tabacum CAD1*, X62343; *Nicotiana tabacum CAD2*, X62344; *Ocimum basilicum geraniol dehydrogenase* (GEDH), AY879284; *Petroselinum crispum ELI3*, X67817; *Pinus taeda CAD1*, Z37992; *Pinus taeda CAD2*, Z37991; *Populus tremuloides CAD*, AF217957; *Populus tremuloides SAD*, AF273256; *Oryza sativa CAD1-1*, DP000086.1; *Oryza sativa CAD1-2*, DP000086.1; *Oryza sativa CAD4*, DP000010.2; and *Vitis vinifera CAD2*, XM\_002277339.

This article can be found at <http://www.jspcmb.jp/>

Published online August 7, 2013

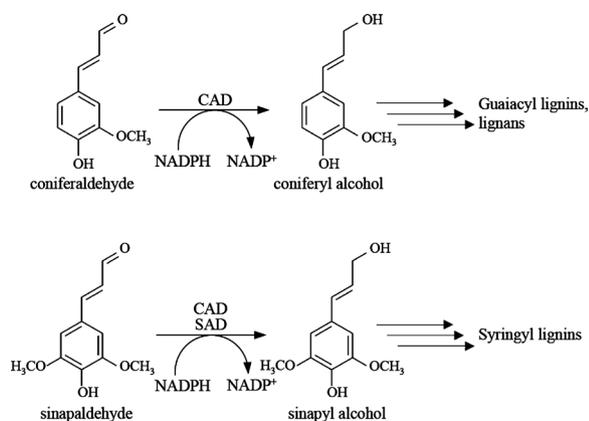


Figure 1. Last step of monolignol biosynthesis involving cinnamyl alcohol dehydrogenase.

angiosperm species, in contrast, have a number of CAD isoforms. These isoforms can show different or comparable characteristics (Goffner et al. 1992; Kim et al. 2004; Li et al. 2001; Mansell et al. 1976; Pillonel et al. 1992; Wyrambik and Grisebach 1975). Wyrambik and Grisebach (1975) undertook one of the earliest studies of different plant CAD isoforms. They isolated and characterized two CAD isoforms that showed different substrate specificities from soybean cell suspension cultures. *Eucalyptus gunnii* was also reported to have several CAD isoforms. Goffner et al. (1992) reported the isolation of two structurally distinct CAD isoforms from *E. gunnii*. One *E. gunnii* CAD (EgCAD1) was only capable of catalyzing coniferaldehyde reduction, whereas the second CAD (EgCAD2) reduced a wider range of substrates: coniferaldehyde, sinapaldehyde, and *p*-coumaraldehyde. Interestingly, a cDNA encoding an EgCAD1-type enzyme showed higher similarity to CCRs (cinnamoyl CoA reductases) than to typical plant CADs (Goffner et al. 1992).

Although intensively studied, the physiological roles of CAD isoforms and the complexity of their relationships in individual species remain unclear (Barakat et al. 2009). Elucidating the role of CADs in the monolignol pathway will provide a better understanding of the nature of lignin, lignan, and neolignan formation, including the determination of the syringyl-guaiacyl composition in lignin polymers. In 2001, Li et al. isolated a CAD-encoding cDNA from *Populus tremuloides* (aspen), and its recombinant protein selectively reduced sinapaldehyde, giving rise to sinapyl alcohol. This enzyme was re-annotated as sinapyl alcohol dehydrogenase (PtreSAD). This finding led to the hypothesis that CAD isoforms with high homology to PtreSAD should exhibit similar characteristics. However, in *Arabidopsis thaliana* lignin biosynthesis, no sinapaldehyde-specific CAD (SAD) activity has been detected from CAD enzymes homologous to PtreSAD (Kim et al. 2004; Sibout et al. 2005). Although *AtELI3-2*

(*AtCAD8* or *At4g37990*) showed a high similarity to *PtreSAD*, the recombinant protein expressed from the gene showed similar specific activities towards both coniferaldehyde and sinapaldehyde (Kim et al. 2004). This is also the case for *Nicotiana tabacum* SAD2 as demonstrated by Barakate et al. (2011). Nevertheless, research on sinapaldehyde-specific CADs are important and can lead to strategies for increasing syringyl lignin content, which can significantly improve effectiveness in kraft pulping (Bugos et al. 1991).

Developing seeds of *Carthamus tinctorius* (safflower) produce significant amounts of lignin and lignans (Sakakibara et al. 2007). In the present study, we isolated two *C. tinctorius* CAD cDNAs (*CtCAD1* and *CtCAD3*) that showed high similarity to *PtreSAD* and one *CtCAD* cDNA (*CtCAD2*) that showed high similarity to *PtreCAD*. The *CtCADs* were characterized biochemically using their recombinant proteins. None of the *CtCADs* showed a strong substrate preference for sinapaldehyde, including *CtCAD1* and *CtCAD3*, despite their high identity and structural similarity with *PtreSAD*. The three *CADs* showed distinguishable characteristics. *CtCAD2* and *CtCAD3* are most likely involved in lignin and lignan biosynthesis, with different degrees of significance.

## Materials and methods

### Isolation and cloning of *C. tinctorius* CAD cDNAs

The cloning and isolation of cDNAs encoding *C. tinctorius* CADs were performed by screening a lambda ZAPII cDNA library prepared from developing seeds of *C. tinctorius* collected at 12 DAF (days after flowering) (Nakatsubo et al. 2007). Conserved regions of the *Pinus taeda* CAD cDNA (GenBank accession no. Z37992) and *Populus tremuloides* SAD cDNA (AF273256) (Li et al. 2001) were used as probes by labeling their NADP- and Zn-binding sites with  $^{32}\text{P}$  using the DECAprime II Random Priming DNA Labeling Kit (Ambion, Austin, TX, USA). The probes were used to conduct two rounds of screening.

### Phylogenetic analysis

Sequences of the *CtCAD* clones were aligned with those of other plant CADs using the E-INS-i method in the "multiple alignment using fast Fourier transform" (MAFFT) 5.0 program (Katoh et al. 2005), and visualized by Bioedit (Hall 1999). A phylogenetic tree was created by the Neighbor-joining method using MAFFT 5.0 (Katoh et al. 2005). The phylogenetic tree was viewed and edited using Dendroscope (Huson et al. 2007).

### Quantitative real-time PCR analysis

For quantitative real-time PCR (qRT-PCR), we used SYBR green-based chemistry (Applied Biosystems LLC, Foster City, CA, USA). All primers were designed from their corresponding cDNA sequences using the Primer Express software (Applied

Biosystems). For *CtCAD1*, the forward primer 5'-TGT AAT GGC ACC AGA TGC AAA G-3' and reverse primer 5'-CAA TAA ACT CAG AGA GAA ATC AAA CTC AA-3' were used to amplify a 69-bp amplicon from the *CtCAD1* 3'-untranslated region (UTR). For *CtCAD2*, the forward primer 5'-ATG TTA ACC GGA TGC AGG AGT T-3' and reverse primer 5'-CAA CGA GAG GCT CGG TTC A-3' were used to amplify a 79-bp amplicon from the *CtCAD2* coding region. For *CtCAD3*, the forward primer 5'-GGT TGC TTC GAT GTG TGC TT-3' and reverse primer 5'-TCG TTC AAA ACC AAG ATT GTT TTA TG-3' were used to amplify an 82-bp amplicon from the *CtCAD3* 3'-UTR.

Total RNA samples were isolated from *C. tinctorius* developing seeds (harvested at 3, 4, 6, 9, 12, 15, and 18 DAF), stems, and leaves using the Plant RNeasy extraction kit (Qiagen, GmbH, Hilden, Germany) or the method of Bugos et al. (1995). Standard curves were generated using serial dilutions of *CtCAD*-plasmid solutions of known concentrations. Ribosomal RNA was chosen as the internal reference for normalization (Bustin 2000) of the expression profile, and was quantified using Taqman Ribosomal RNA Control Reagents (Applied Biosystems). Amplification and fluorescence measurements were performed using an ABI 7300 real-time PCR apparatus with default parameters.

#### **Total plant protein assays with coniferaldehyde and sinapaldehyde**

All of the substrates and internal standards used in enzyme assays were prepared previously (Sakakibara et al. 2007). *C. tinctorius* seeds at different developmental stages (6, 7, 8, 9, 10, 12, 14, 15, and 16 DAF), stems, and leaves were harvested and stored in liquid nitrogen until use. Each sample (0.5 g) was crushed in liquid nitrogen using a chilled mortar and pestle along with 100 mg each of sea sand and polyclar AT. After reaching ice temperature, 1.75 ml chilled 0.1 M potassium phosphate buffer (KPB) pH 7.5 containing 10 mM dithiothreitol (DTT) was added. The sample was ground until it became a fine slurry and then centrifuged at 10,000×g at 4°C for 15 min. The supernatant was filtered through a Millex-GV PVDF 0.22-μm diameter filter (Millipore, Carrigtwohill Co., Cork, Ireland) and submitted to a Sephadex G-25 column (Pharmacia, Uppsala, Sweden) to change the buffer to 50 mM Tris-HCl (pH 8.0). The obtained protein solution was immediately used for enzyme assays with coniferaldehyde or sinapaldehyde in an enzyme assay mixture containing 50 mM Tris-HCl buffer (pH 8.0), 250 μM β-mercaptoethanol, 50 μM NADPH, and 120 μl of cell-free plant protein solution. The reactions were carried out at 30°C for 60 min. The reactions were terminated by extraction with ethyl acetate containing internal standards (coniferyl alcohol-*d*<sub>3</sub> or sinapyl alcohol-*d*<sub>3</sub>), dried up, and stored in -20°C until use.

Analyses of all enzyme assays in this study were carried out by gas chromatography-mass spectrometry (GCMS) analysis according to Nakatsubo et al. (2007). GCMS analysis was performed using a Shimadzu QP-5050A GC-MS system

(Shimadzu Co., Kyoto, Japan) [electron impact mode (70 eV); column: Shimadzu HiCap CBP10-M25-025 (20 m × 0.22 mm); carrier gas: helium; injection temperature: 240°C; column temperature, 40°C at *t*=0 to 2 min, then to *t*=230 at 25°C min<sup>-1</sup>] or a Shimadzu QP-2010 plus GC-MS system [electron-impact mode (70 eV); column, Shimadzu HiCap CBP10-M25-025 column (10 m × 0.22 mm); carrier gas, helium; injection temperature, 250°C; column temperature, 80°C at *t*=0 to 2 min, then to 250°C at 10°C min<sup>-1</sup>]. Dried samples were dissolved in *N*, *O*-bis(trimethylsilyl)acetamide, heated at 60°C for 45 min to produce their trimethylsilyl (TMS) derivatives and then subjected to GCMS analysis.

#### **Statistical analysis**

Data obtained from qRT-PCR and plant protein assays were collected from three identical experiments. The normal distribution of the collected data was tested using the Shapiro-Wilk normality test in the "R" software suite (R development team 2011). To determine statistical significance among data sets, depending on normality of the data and number of data sets compared, one-way ANOVA, the Kruskal-Wallis test, or an independent *t*-test in Brightstat (Stricker 2008) was used. The correlations between *CtCAD* expression and specific activity of total plant protein towards coniferaldehyde and sinapaldehyde were determined by linear regression and Pearson correlation tests in Brightstat (Stricker 2008).

#### **Expression of recombinant CtCADs in Escherichia coli**

PCR was used to introduce an *Nde*I site at the 5'-end and a *Not*I site at the 3'-end of the coding sequences of the *CtCAD* clones, using sense primers (*CtCAD1*, 5'-TCA TAT GTC ACT AGA GTC AGT GCA-3'; *CtCAD2*, 5'-TCA TAT GGA AAG TTT GAA AGA AGA AAG-3'; *CtCAD3*, 5'-TCA TAT GGT GAA ATC TCC AGA AGC-3') and antisense primers (*CtCAD1*, 5'-TGC GGC CGC CAT AGA AGA CTT GAG GG-3'; *CtCAD2*, 5'-TGC GGC CGC TTC GTC TTC GAG CTT GCT-3'; *CtCAD3*, 5'-TGC GGC CGC TGG AGC TTT TAG GGA GTT G-3'). The *CtCAD* ORFs were then subcloned into pET23a expression vectors with a His-tag fused at the C-terminal of the cDNA sequences. After sequencing to confirm the accuracy of the *CtCAD*-pET23a constructs, each construct was introduced into the *E. coli* strain BL21 (DE3) (Novagen, San Diego, CA, USA). The *E. coli* BL21 (DE3) cells transformed with *CtCAD1* and *CtCAD2* were grown in LB medium enhanced with the Overnight Express Autoinduction System (Novagen) according to the manufacturer's protocol. For *CtCAD3*, expression was achieved by IPTG induction following the method of Li et al. (2001). After harvesting the bacterial cells by centrifugation at 2,000×g for 10 min at 4°C, the pellets were processed for chromatography using His-bind Resin (Novagen) according to the manufacturer's protocol. The obtained eluate was desalted with a Sephadex G-25 column. Confirmation of the purified *CtCAD* molecular mass was performed by SDS-PAGE by comparing it with standard proteins in the Low Molecular

Weight (LMW) Electrophoresis Calibration Kit (Pharmacia). The concentration of the recombinant protein was determined using the Bradford method (Bradford 1976), with bovine serum albumin as the standard.

#### **Kinetic characterization of recombinant CtCADs**

The basic enzyme assay mixture contained 50 mM KPB or Tris-HCl buffer, ranging from pH 6.5 to 7.5, 500  $\mu$ M  $\beta$ -mercaptoethanol, 50  $\mu$ M NADPH, and either 1214 ng (CtCAD1), 642 ng (CtCAD2), or 1477 ng (CtCAD3) protein in a total volume of 200  $\mu$ l. The optimum pH and temperature were determined using standard assay conditions. A concentration ranging from 2 to 500  $\mu$ M of the substrate (coniferaldehyde or sinapaldehyde) was used in the kinetic assays. Reactions were carried out for 15 min at 26°C (CtCAD1) or 30°C (CtCAD2 and CtCAD3). The reactions were terminated by extraction with ethyl acetate containing internal standards (coniferyl alcohol- $d_3$  or sinapyl alcohol- $d_3$ ), dried up, and stored in  $-20^\circ\text{C}$  until submitted to GCMS analysis. The data was then analyzed according to the Lineweaver–Burk equation.

All assays for inhibition kinetics were performed according to the basic enzyme assay mixture with 50 mM KPB (pH 6.5), and incubated for 15 min at 30°C. When coniferaldehyde was used as the substrate, sinapaldehyde was introduced as the inhibitor, and *vice versa*.  $K_i$  values were determined according to Dixon (1953) based on GCMS data.

#### **Homology modeling and molecular dynamics**

The SWISS-MODEL automated protein structure homology-modeling server (Arnold et al. 2006; Bordoli et al. 2009) used the PtreSAD crystal structure (PDB ID: 1YQD) (Bomati and Noel 2005) as the template for CtCAD1 and CtCAD3, and AtCAD5 crystal structure (PDB ID: 2CF6) (Youn et al. 2006) as the template for CtCAD2. Identity between the CtCADs and their respective templates ranged between 68 to 79%. Bordoli et al. (2009) mentioned that proteins that share more than 50% identity with their template could generally be submitted to homology modeling to produce reliable models. The models were refined by adding missing side chains using the “complete a structure” program on the WHAT IF server (China et al. 1995; Rodriguez et al. 1998; Vriend 1990).

The structures were evaluated by three independent methods. The first method was to calculate the Root Mean Square Deviation (RMSD) between the backbones of the CtCAD models and their templates by using the matchmaker program in the Chimera software suite (Meng et al. 2006; Pettersen et al. 2004) and SSM superpose program in the coot software suite (Emsley et al. 2010). The second method was to verify quality of the structures by using ProSA (Wiederstein and Sippl 2007) and PROCHECK (Laskowski et al. 1993).

The last method for assessing the quality of the constructed models was by submitting them to molecular dynamics using the GROMACS software suite (Hess et al. 2008). Molecular dynamics was performed with an OPLS-AA force field

(Kaminski et al. 2001) and the SPC/E water model (Berendsen et al. 1987). Energy minimization using the steepest descent minimization method was conducted until  $F_{\text{max}}$  reached  $\leq 1,000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ . The CtCAD models were solvated in a rectangular water box whose planes were set to be 1 nm away from the surface of the models. Sodium cations were added to neutralize the charge of the system and periodic boundary conditions were imposed. The particle-mesh Ewald method (Darden et al. 1993) was used to describe long-range electrostatic interactions. The systems were equilibrated for 100 ps in an NVT ensemble (constant system volume and temperature) to reach the target temperature (300 K) and an additional 100 ps in an NPT ensemble (constant system pressure and temperature) to reach the pressure of 1 bar. Production simulations were performed in the NPT ensemble using the velocity rescaling thermostat (Bussi et al. 2007) and Parrinello–Rahman barostat (Parrinello and Rahman 1981) with 0.1 ps and 2.0 ps time constants, respectively. The length of the simulation was 1 ns, and coordinates were saved every 2 ps for analysis.

Visualization of the CtCAD substrate binding pockets and their molecular surfaces were performed using the Chimera software suite (Pettersen et al. 2004; Sanner et al. 1996).

#### **Molecular docking**

DOCK 6 (Ewing et al. 2001) was used for molecular docking procedures of the substrates and co-substrates to the CtCAD models at default settings, unless stated otherwise. Structures were obtained from the RCSB Protein Data Bank: Coniferaldehyde and sinapaldehyde from Louie et al. (2010), and NADP<sup>+</sup> from Bomati and Noel (2005) or Youn et al. (2006). We conducted flexible ligand docking of NADP<sup>+</sup> to the CtCADs, which was followed with flexible ligand docking of substrates (coniferaldehyde and sinapaldehyde) to the CtCAD–NADP<sup>+</sup> complex, cycling through 1,000 maximum orientations for each docking procedure.

## **Results**

#### **Isolation of CtCAD cDNAs**

<sup>32</sup>P-labeled conserved regions of *PtaeCAD* cDNA and *PtreSAD* cDNA were used as probes to screen the cDNA library (ca.  $1.6 \times 10^5$  plaques) prepared from 12 DAF seeds of *C. tinctorius*. After second-round screening, 14 positive clones were obtained. Sequence analysis of the positive clones revealed that nine showed high sequence homology to previously isolated CADs. The nine clones were classified into three groups. From each group, a full-length cDNA was obtained and designated as *CtCAD1*, *CtCAD2*, and *CtCAD3*. The ORFs of *CtCAD1*, *CtCAD2*, and *CtCAD3* were 1,080, 1,077, and 1,083 bp, respectively. The predicted amino acid sequences encoded proteins with a molecular mass of approximately 38 kDa.

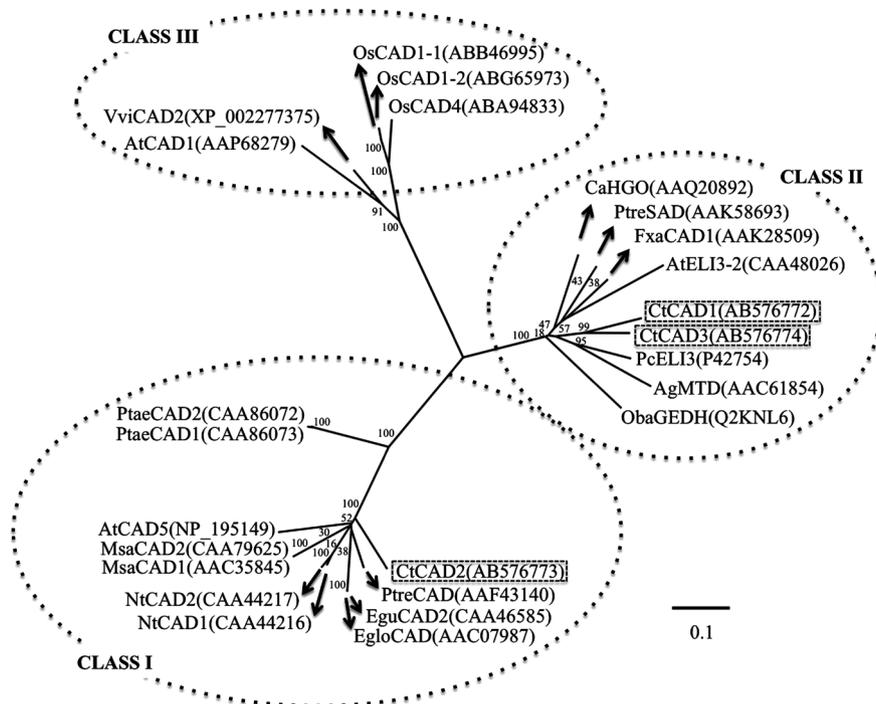


Figure 2. Phylogenetic tree from a CAD amino acid sequence alignment, constructed using the Neighbor-joining method. The tree was generated using MAFFT 5.0 (Kato et al. 2005) and edited with Dendroscope (Huson et al. 2007). The classification of class I, class II and class III are after Barakat et al. (2009). Bootstrap confidence values ( $n=100$ ) are shown at the forks.

### *CtCAD* phylogenetic analysis

All of the deduced amino acid sequences encoded by the *CtCAD* cDNAs had highly conserved zinc-binding domains and an NADP-binding domain that were identified in previously characterized CAD isoforms (Knight et al. 1992; Grima-Pettenati et al. 1993; Li et al. 2001; MacKay et al. 1995). The catalytic Zn1-binding consensus and structural Zn2 consensus were identified at amino acid residues 69 to 83 and 89 to 114 (*CtCAD1*), 68 to 82 and 88 to 113 (*CtCAD2*), and 71 to 85 and 91 to 116 (*CtCAD3*), respectively. The NADP-binding domain was located at amino acid residues 189 to 194 (*CtCAD1*), 188 to 193 (*CtCAD2*), and 191 to 196 (*CtCAD3*) (Supplemental Figure S1).

Sequence identity between CAD isoforms was deduced using the IDENTIFY matrix suite in Bioedit (Hall 1999). *CtCAD1* and *CtCAD3* were closely related to *PtreSAD* (68.5% and 73.7% amino acid identity, respectively). *CtCAD2*, in contrast, showed high similarity to *PtreCAD* (79.8% identity) (Li et al. 2001) and *AtCAD5* (74.8% identity) (Youn et al. 2006). The *CtCADs* showed low amino acid sequence identities to *EgCAD1*-type enzymes (8.6–10.5%), which are close to cinnamoyl CoA reductases (CCRs) (Goffner et al. 1992).

A phylogenetic tree (Figure 2) constructed using the neighbor-joining method showed that *CtCAD1* and *CtCAD3* were located in a clade that included *P. tremuloides* sinapaldehyde-specific SAD (*PtreSAD*) and plant defense-related CADs from *A. thaliana* (Somssich

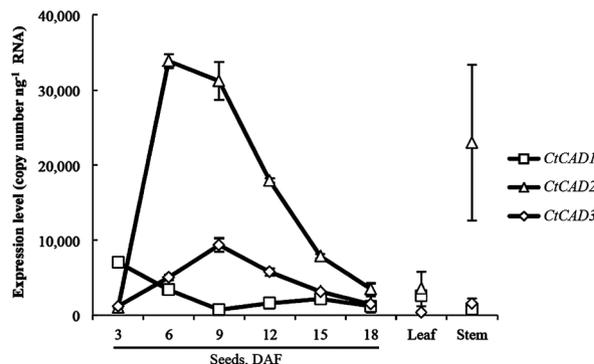


Figure 3. Quantitative real-time PCR analysis of the expressions of *CtCAD1*, *CtCAD2* and *CtCAD3*. Data are means obtained from three identical experiments ( $\pm$  standard deviation).

et al. 1996), *Apium graveolens* (Williamson et al. 1995), *Camptotheca acuminata* (Valletta et al. 2010), *Ocimum basilicum* (Iijima et al. 2006), *Petroselinum crispum* (Kiedrowski et al. 1992), and the lignin-related CAD from *Fragaria*  $\times$  *ananassa* cv. Chandler (Blanco-Portales et al. 2002). Barakat et al. (2009) classified CADs in this clade as class II CADs. *CtCAD2* was located in a different clade that contained class I CADs (Barakat et al. 2009) that are involved in monolignol biosynthesis.

### *qRT-PCR analysis of gene expression*

As shown in Figure 3, *CtCAD2* was expressed at higher levels compared to *CtCAD1* and *CtCAD3* in all organs, except for 3 DAF developing seeds, in which the *CtCAD1*

expression level was higher. *CtCAD2* and *CtCAD3* showed similar expression patterns in developing seeds, with the highest expressions at 6–9 DAF. Conversely, the *CtCAD1* expression level was highest at 3 DAF. *CtCAD2* expression in the stem was comparable to that at 6–12 DAF in the seeds, while *CtCAD3* expression in the stem was lower than that in the seeds (Figure 3). *CtCAD2* and *CtCAD3* were expressed at significantly higher levels in the stem than in the leaves ( $p < 0.05$ ), whereas there was no significant difference in expression levels of *CtCAD1* between stems and leaves ( $p > 0.05$ ).

#### Plant protein assay

As shown in Figure 4, plant protein assays showed that the highest specific activity was in seeds at 6 to 7 DAF. The specific activity gradually became lower as the seeds matured. This was the case for both coniferaldehyde and sinapaldehyde activities. Assays of proteins from stems also showed considerable specific activity (approx.  $20 \text{ nmol h}^{-1} \text{ mg}^{-1}$  protein) for both coniferaldehyde and sinapaldehyde.

Correlation analysis was conducted to analyze the relationship between gene expression and specific activity of total plant proteins towards coniferaldehyde and

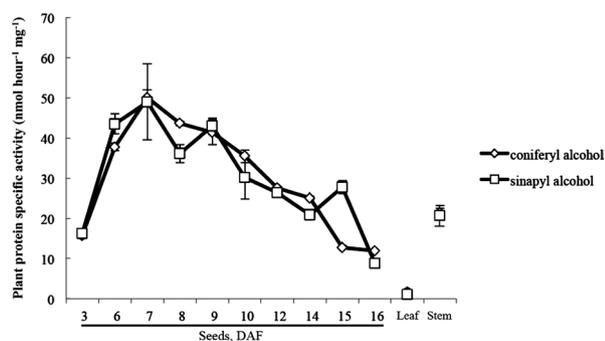


Figure 4. Specific activity of total plant protein of *C. tinctorius* seeds at indicated developmental stages (DAF=days after flowering), leaves, and stems, in assays with coniferaldehyde and sinapaldehyde as substrates. Generally, coniferyl alcohol and sinapyl alcohol yields (product) were significantly higher in assays incubated with total plant protein from 7 DAF developing seeds ( $p < 0.05$ ). The only exception was in sinapaldehyde assays incubated with total plant proteins from 6 DAF developing seeds. Data are means obtained from three identical experiments ( $\pm$  standard deviation).

sinapaldehyde. The results are shown in Supplemental Table S1. *CtCAD1* expression was not correlated with the specific activity of total plant proteins towards coniferaldehyde or sinapaldehyde. In contrast, *CtCAD2* and *CtCAD3* expressions in different organs were strongly correlated ( $p < 0.001$ ) with the specific activity of total plant proteins for the tested substrates. However, in contrast to the high *CtCAD2* expression and high specific plant protein activity in the stem, the expression of *CtCAD3* was relatively lower in the stem.

#### Preparation and kinetic analysis of recombinant CtCADs

After purification by His-bind affinity chromatography, SDS-PAGE analysis of the recombinant putative CtCADs showed apparent homogeneity (Supplemental Figure S2). Enzyme assays showed that the recombinant protein of all putative CtCADs reduced coniferaldehyde and sinapaldehyde into their corresponding alcohols, coniferyl alcohol and sinapyl alcohol, confirming their identity as CADs.

For further characterization of the CtCADs, we examined their substrate preferences by conducting kinetic analyses on the purified recombinant CtCADs. GCMS-based analysis was used to determine the optimum assay conditions (pH, temperature, and reaction time) for each CtCAD. We used Lineweaver-Burk analysis to determine the kinetic attributes of each CtCAD, including  $K_m$ ,  $V_{max}$ , and  $k_{cat}$  values towards two substrates (coniferaldehyde and sinapaldehyde). The analysis indicated that there was no preferable substrate affinity. For each CtCAD, the  $K_m$  values toward both substrates did not differ significantly (Table 1). *CtCAD2* had the highest  $k_{cat}$  values for coniferaldehyde and sinapaldehyde reduction (Table 1), in which the values towards the substrates were approximately 3- to 90-times greater than those obtained in reactions catalyzed by *CtCAD1* and *CtCAD3*.

#### Inhibition kinetics

According to Lineweaver-Burk plots from CtCAD assays, the inhibition types for all CtCADs when coniferaldehyde (substrate) was incubated with

Table 1. Kinetic properties of CtCADs with coniferaldehyde and sinapaldehyde as substrates. All assays were conducted at 30°C, except for *CtCAD1* assays (26°C), and were incubated for 15 min in 50 mM potassium phosphate buffer (pH 6.5 or 7.5) or Tris-HCl buffer (pH 7.5). Results are compiled from 2 independent experiments.

Clone	Substrate	$K_m$ (mM)	$V_{max}$ (nmol min <sup>-1</sup> μg <sup>-1</sup> )	$k_{cat}$ (min <sup>-1</sup> )	$k_{cat}/K_m$ (mM <sup>-1</sup> min <sup>-1</sup> )
CtCAD1	Coniferaldehyde	0.068	0.019	0.728	10.62
	Sinapaldehyde	0.058	0.008	0.318	5.39
CtCAD2	Coniferaldehyde	0.024	0.049	1.901	81.49
	Sinapaldehyde	0.026	0.063	2.460	95.30
CtCAD3	Coniferaldehyde	0.032	0.001	0.059	2.15
	Sinapaldehyde	0.055	0.002	0.081	1.38

sinapaldehyde (inhibitor), and *vice versa*, were competitive inhibitions. The  $K_i$  values for coniferaldehyde were lower than those for sinapaldehyde for all recombinant CtCADs, indicating that coniferaldehyde was a stronger inhibitor than sinapaldehyde. The detailed results are shown in Table 2.

#### Quality assessment of the CtCAD models

CtCAD sequences were submitted to the SWISS-MODEL automated protein structure homology-modeling server (Arnold et al. 2006; Bordoli et al. 2009) to generate their predicted structures. Superposition of the backbones of CtCAD1, CtCAD2, and CtCAD3 with their respective templates using Chimera or Coot, gave RMSD values ranging from 0.11 to 0.275 Å. This indicated good structural alignment of the CtCAD models with their respective templates. ProSa analysis showed that the Z-scores and pseudo-energy profiles based on the knowledge-based mean field for each model were similar to their crystal structure templates. Furthermore, Ramachandran plots from PROCHECK analysis revealed that the CtCAD models also showed comparable stereochemical properties to those of the templates

Table 2. Inhibition kinetics of CtCADs. All observed inhibitions were competitive-type inhibitions. Kinetic parameters were determined for the inhibition of the reduction of each substrate by inhibitors. Results are compiled from 2 independent experiments.

Clone	Substrate	Inhibitor	$K_i$ ( $\mu$ M)
CtCAD1	Coniferaldehyde	Sinapaldehyde	533.87
	Sinapaldehyde	Coniferaldehyde	41.77
CtCAD2	Coniferaldehyde	Sinapaldehyde	906.07
	Sinapaldehyde	Coniferaldehyde	105.62
CtCAD3	Coniferaldehyde	Sinapaldehyde	269.02
	Sinapaldehyde	Coniferaldehyde	87.29

Table 3. Residues involved in substrate-binding site of CtCAD1, CtCAD2, CtCAD3, PtreSAD, and AtCAD5. Important residues determining substrate specificity in PtreSAD-type enzymes are shaded.

	CtCAD1	CtCAD3	PtreSAD	CtCAD2	AtCAD5
Chain A	CYS 48	CYS 50	CYS 50	CYS 47	CYS 47
	SER 50	SER 52	SER 52	THR 49	THR 49
	—	—	—	GLN 53	GLN 53
	TRP 59	TRP 61	TRP 61	LEU 58	LEU 58
	—	—	—	MET 60	MET 60
	HIS 70	HIS 72	HIS 72	HIS 69	HIS 69
	—	—	—	GLU 70	GLU 70
	CYS 96	CYS 98	CYS 98	CYS 95	CYS 95
	GLN 112	ASN 115	ASN 115	—	—
	TYR 114	TYR 116	TYR 116	—	—
	MET 120	LEU 122	LEU 122	TRP 119	TRP 119
	CYS 164	CYS 166	CYS 166	CYS 163	CYS 163
	ILE 277	ALA 279	ALA 279	VAL 276	VAL 276
	ASN 300	ALA 302	GLY 302	PHE 299	PHE 299
ILE 301	ILE 303	ILE 303	ILE 300	ILE 300	
Chain B	PHE 287	PHE 289	PHE 289	PRO 286	PRO 286
	ILE 290	ILE 292	ILE 292	MET 289	MET 289
	MET 291	MET 293	ALA 293	LEU 290	LEU 290

(Supplemental Table S2).

During the 1 ns production molecular dynamics simulation, the models were relatively stable, with their backbone moving at an RMSD value of around 1 to 3 Å relative to their energy-minimized structure. Furthermore, the radius of gyration of the models, which is the measure of protein compactness and an indicator of folding stability, was also stable throughout the molecular dynamics simulation at approximately 3 nm. Taking the results together, we concluded that the CtCAD models have reasonable quality and can be used for subsequent analyses.

#### Substrate-binding pockets of the CtCAD models

CtCAD1 and CtCAD3 substrate-binding sites were constructed with the similar residues that are involved in the PtreSAD binding-site (Bomati and Noel 2005), while the CtCAD2 substrate-binding site resembled to that of AtCAD5 (Youn et al. 2006). In fact, all 16 residues involved in the CtCAD2 and template (AtCAD5) substrate-binding sites were identical with no substitutions (Table 3). The PtreSAD substrate-binding site consisted of 15 amino acid residues. Among them, there were five and two residue substitutions in the CtCAD1 and CtCAD3 substrate-binding pockets, respectively, compared with the template. For CtCAD1, residues ASN115, LEU122, ALA279, GLY302 (chain A), and ALA293 (chain B) in the PtreSAD substrate-binding pocket were substituted with GLN112, MET120, ILE277, ASN300, and MET291, respectively. For CtCAD3, only two substitutions occurred with regard to the PtreSAD substrate-binding site, in which GLY302 (chain A) and ALA293 (chain B) in PtreSAD were substituted by the slightly larger ALA302 and MET293 in CtCAD3. These

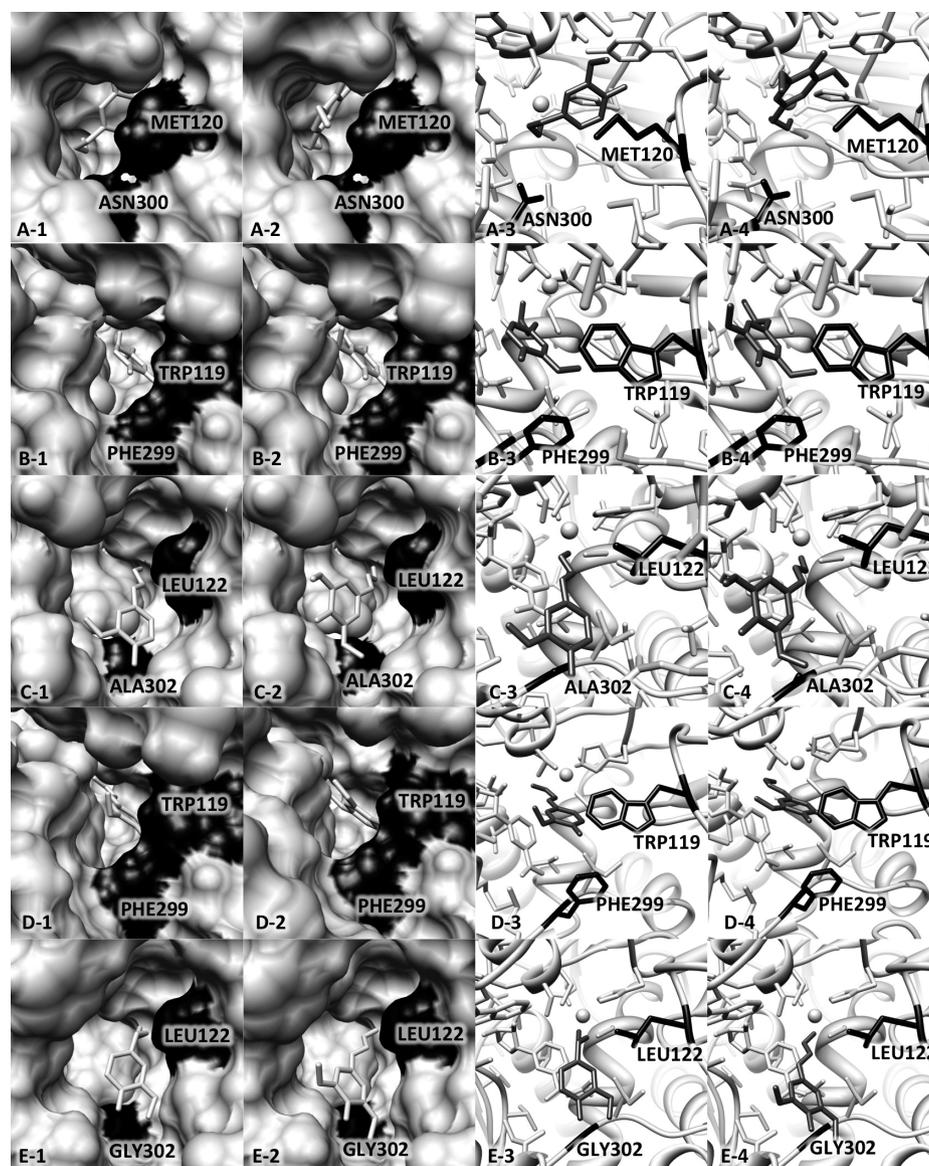


Figure 5. Molecular surface and stick representation of substrate docking in (A) CtCAD1, (B) CtCAD2, (C) CtCAD3, (D) AtCAD5, (E) and PtreSAD substrate-binding pockets (1=molecular surface representation of coniferaldehyde docking poses; 2=molecular surface representation of sinapaldehyde docking poses; 3=stick model representation of coniferaldehyde docking poses; 4=stick model representation of sinapaldehyde docking poses).

substitutions contributed to the different pocket shapes between CtCAD1, CtCAD3 and PtreSAD, even though they share high identity. The residue differences had only a subtle effect in case of CtCAD3. However, it profoundly changed the shape of the CtCAD1 substrate-binding site in regards to the template, resulting in a more constricted binding pocket.

#### Molecular docking

Molecular docking was performed to predict the orientation of the substrates (coniferaldehyde and sinapaldehyde) when they bound to the CtCAD substrate-binding pockets (Figure 5). The substrates fitted well into the pockets, with their aldehyde carbonyls directly coordinating with the  $Zn^{2+}$  catalytic ion. The

different shapes and sizes of the pockets consequently resulted in apparent differences in the orientation as to how the substrates resided in them. The bulky residues of MET120 and ASN300 in the CtCAD1 pocket made it sterically restricted relative to its template (PtreSAD) (Table 3). Furthermore, the added polarity of ASN300 may also contribute to substrate-binding orientation. The phenyl ring of both coniferaldehyde and sinapaldehyde resided above the MET120 residue (Figure 5A). The CtCAD3 substrate-binding pocket is similar to that of PtreSAD. However, because of the substitution of GLY302 to ALA302 (Table 3), some differences can be observed in the substrate binding orientation. Coniferaldehyde was packed flat to the base of the pocket just above ALA302, similar to its orientation in PtreSAD,

whereas sinapaldehyde was oriented differently, coplanar to the co-substrate NADP<sup>+</sup> (Figure 5C). The CtCAD2 binding site is identical to that of its template (AtCAD5). The docking of both coniferaldehyde and sinapaldehyde to the binding pocket of CtCAD2 resulted in similar docking orientations, coplanar to the co-substrate NADP<sup>+</sup> (Figure 5B).

## Discussion

The CtCADs showed high amino acid sequence homology and close phylogenetic relationships with previously characterized typical plant CADs and not CCR-like CADs (Figure 2). CtCAD1 and CtCAD3, which showed close phylogenetic relationships with the sinapaldehyde-specific PtreSAD, did not show an exclusive sinapaldehyde preference. In fact, the enzyme kinetics analyses indicated that the recombinant CtCAD1 and CtCAD3 showed a preference to reduce coniferaldehyde rather than sinapaldehyde. This was concluded from the slightly higher  $k_{\text{cat}}/K_m$  values for coniferaldehyde than for sinapaldehyde, and smaller  $K_i$  values for sinapaldehyde reduction to which coniferaldehyde was added as the inhibitor. The  $K_i$  values indicated that coniferaldehyde was a stronger inhibitor than sinapaldehyde for the recombinant enzymes.

CtCAD2 showed similar characteristics to previously isolated angiosperm CADs, especially in its ability to reduce both coniferaldehyde and sinapaldehyde at comparable rates. Enzyme kinetics analyses of CtCAD2 showed that, when coniferaldehyde and sinapaldehyde were supplied as substrates separately, both substrates showed similar  $k_{\text{cat}}/K_m$  values. Even though sinapaldehyde was a slightly better substrate, judging from the  $k_{\text{cat}}/K_m$  values, the difference was statistically insignificant. Inhibition kinetics results suggested that when the two substrates were supplied together, coniferaldehyde was a stronger inhibitor than sinapaldehyde, as shown by the  $K_i$  values towards the two compounds.

Quantitative real-time PCR results showed that CtCAD2 was expressed at relatively higher levels in all tested organs, except for 3 DAF developing seeds, where CtCAD1 expression was marginally higher (Figure 3). Even so, CtCAD2 and CtCAD3 showed similar gene expression patterns in developing seeds; that is, a significant increase at 6–9 DAF, as well as an increase of expression levels in stems. Furthermore, plant protein assays of developing seeds showed that specific activity towards coniferaldehyde and sinapaldehyde, giving rise to their respective products, was highest at 6 to 7 DAF (Figure 4). Sakakibara et al. (2007) reported an increase in the amount of lignins up to approximately 700 nmol coniferyl alcohol unit mg<sup>-1</sup> and nitrobenzene oxidation products up to approximately

90 pmol mg<sup>-1</sup> in *C. tinctorius* seeds from 6 to 9 DAF, indicating that this period marked the beginning of lignin accumulation. Furthermore, other monolignol biosynthetic pathway genes of *C. tinctorius*, encoding caffeoyl CoA *O*-methyltransferase (*CtCoAOMT*) and 5-hydroxyconiferaldehyde *O*-methyltransferase (*CtAldOMT*), showed the same expression pattern in the same tissues (Nakatsubo et al. 2007).

There were considerable amounts of CtCAD2 expression in stem samples, the organ where CtCAD1 and CtCAD3 were expressed at much lower levels. It has been reported that some of the important genes in monolignol biosynthesis are expressed specifically in differentiating xylem and other vascular tissues (Dixon et al. 2001), which are abundant in stems. This may also be the case for CADs.

Through statistical analysis, our results showed strong correlations between CtCAD2 and CtCAD3 gene expression levels and specific activity of plant proteins towards coniferaldehyde and sinapaldehyde. CtCAD2 gene expression levels had higher correlation values compared to CtCAD3 gene expression levels toward the specific activity of plant proteins. Gene expression can be correlated with production of its corresponding protein in organs. Although this does not necessarily reflect a cause-and-effect relationship and is not the case for all genes because of the complex mechanism of protein translation (Greenbaum et al. 2003), we conclude that the compared variables (gene expression and specific activity of total plant protein), as judged by their high correlation values, are somewhat connected. Combining the obtained data, there was a strong indication that CtCAD2, and to a lesser extent, CtCAD3, were indeed the main CADs involved in supplying monolignols in lignification in *C. tinctorius*.

In addition to lignin, large amounts of lignans (up to approximately 10 nmol mg<sup>-1</sup> for the lignan matairesinol) are accumulated in the seeds, which started at around 7 DAF. However, only negligible amounts of lignans were detected in the stems and leaves (Sakakibara et al. 2007; Umezawa et al. 2013). The lignan accumulation profile in the seeds is similar to the expression profiles of CtCAD2 and CtCAD3. Furthermore, the accumulation of lignans in the leaves and stems (Sakakibara et al. 2007; Umezawa et al. 2013) is in accordance with the expression level and profile of CtCAD3, which has apparent, yet low, gene expression levels. These results suggest the possible involvement of CtCAD3 in lignan biosynthesis. Involvement of CtCAD2 in lignan biosynthesis is also a possibility, despite it being more dominant in lignin biosynthesis.

Although CtCAD1 may not be dominant in monolignol biosynthesis, it may have other roles in plant metabolism, such as in stress-induced defenses, similar to the *ELI3* genes encoding pathogen defense-

related CADs in *Apium graveolens*, *Arabidopsis*, or *P. crispum* (parsley) (Logemann et al. 1997; Somssich et al. 1996; Williamson et al. 1995). It is also possible that CtCAD1 is involved in extremely localized processes, such as lignification during *Arabidopsis* silique or anther dehiscence (Liljegren et al. 2000; Mitsuda et al. 2005). To investigate this further, GUS promoter analysis or gene expression analysis at a higher resolution using more specific samples could be carried out; that is, analysis of gene expression in different sections of the seed instead of whole seeds.

*In silico* protein structure analysis is a useful tool for analyzing the details of substrate specificity. Structural analysis of PtreSAD and kinetic analysis of the PtreSAD mutants indicated that the residue GLY302, which contributes to the rather flat active site floor, and the residue LEU122, which builds the right side wall of the binding pocket, are important factors in determining substrate specificity of SAD-type enzymes (Bomati and Noel 2005). PtreSAD showed a 3-fold higher activity towards sinapaldehyde compared to coniferaldehyde (Li et al. 2001), while a member of Class II CADs (Figure 2), PcELI3, which has the bulkier ASN279 and GLN99 as the equivalent of PtreSAD's GLY302 and LEU122, respectively, resulted in total loss of activity towards sinapaldehyde (Bomati and Noel 2005; Logemann et al. 1997). However, CtCAD1 and CtCAD3 maintained the activity towards sinapaldehyde at an equal level to that of coniferaldehyde, although the homology modeling analysis showed that the GLY302/LEU122 residue combination in PtreSAD were substituted by ASN300/MET120 and ALA302/LEU122 in CtCAD1 and CtCAD3, respectively (Table 3). In addition, NtSAD2 with ALA301/MET122 did not show strong preference towards either coniferaldehyde or sinapaldehyde (Barakate et al. 2011). The significant sinapaldehyde reducing activity of the CtCADs and NtSAD2 can be rationalized by the multiple effects of amino acid substitution on their kinetic behavior. Indeed, the presence of the GLY302/LEU122 combination does not guarantee a clear preference for sinapaldehyde, as demonstrated by FxaCAD1 (Blanco-Portales et al. 2002). This CAD still showed significant activity towards coniferaldehyde (80% activity compared to the activity for sinapaldehyde), although it possesses the important GLY300/LEU120 residues as the equivalent of PtreSAD's GLY302/LEU122, in addition to similar residues building the binding pocket as PtreSAD (Blanco-Portales et al. 2002). Taken together, the data suggest that although the combination of residues at the position of GLY302/LEU122 is important in determining substrate specificity in PtreSAD-type enzymes, there are most likely other factors that also need to be taken into account.

In conclusion, this report describes the isolation and characterization of three genes encoding cinnamyl

alcohol dehydrogenase (CAD) in *C. tinctorius*; CtCAD1, CtCAD2, and CtCAD3. CtCAD2 is most likely to be the dominant CtCAD in monolignol biosynthesis. The CtCADs did not show strong preferences for either coniferaldehyde or sinapaldehyde, including CtCAD1 and CtCAD3, despite their close phylogenetic relationship and/or structural similarity with the sinapaldehyde-specific PtreSAD, leading to the notion that they are somewhat positioned between PtreCAD and PtreSAD in terms of structure and kinetic behavior.

## Acknowledgements

We thank Professor Bunzo Mikami, Kyoto University, for the many advices regarding homology modeling and for critical reading of the manuscript. We thank Professor Vincent L. Chiang, North Carolina State University, for kindly providing the *P. taeda* CAD and *P. tremuloides* CAD cDNAs used in this research and for critical reading of the manuscript. We thank Professor ChulHee Kang, Washington State University, for providing the partner molecule file of 2CF6 (AtCAD5), which was used in homology modeling. This research was partly supported by a grant-in-aid from the New Energy and Industrial Technology Development Organization (Development of Fundamental Technologies for Controlling the Process of Material Production of Plants) and by grants-in-aid from the Japan Society for the Promotion of Science (nos. 12660150, 16380116, and 18658069). A part of this study was conducted at the Forest Biomass Analytical System, Research Institute for Sustainable Humanosphere, Kyoto University, Japan.

## References

- Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22: 195–201
- Barakat A, Bagniewska-Zadworna A, Choi A, Plakkat U, DiLoreto DS, Yellanki P, Carlson JE (2009) The cinnamyl alcohol dehydrogenase gene family in *Populus*: phylogeny, organization, and expression. *BMC Plant Biol* 9: 26
- Barakate A, Stephens J, Goldie A, Hunter WN, Marshall D, Hancock RD, Lapierre C, Morreel K, Boerjan W, Halpin C (2011) Syringyl lignin is unaltered by severe sinapyl alcohol dehydrogenase suppression in tobacco. *Plant Cell* 23: 4492–4506
- Berendsen HJC, Grigera JR, Straatsma TP (1987) The missing term in effective pair potentials. *J Phys Chem* 91: 6269–6271
- Blanco-Portales R, Medina-Escobar N, López-Ráez JA, González-Reyes JA, Villalba JM, Moyano E, Caballero JL, Muñoz-Blanco J (2002) Cloning, expression and immunolocalization pattern of a cinnamyl alcohol dehydrogenase gene from strawberry (*Fragaria × ananassa* cv. Chandler). *J Exp Bot* 53: 1723–1734
- Boerjan W, Ralph J, Baucher M (2003) Lignin biosynthesis. *Annu Rev Plant Biol* 54: 519–546
- Bomati EK, Noel JP (2005) Structural and kinetic basis for substrate selectivity in *Populus tremuloides* sinapyl alcohol dehydrogenase. *Plant Cell* 17: 1598–1611
- Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, Schwede T (2009) Protein structure homology modeling using SWISS-MODEL workspace. *Nat Protoc* 4: 1–13
- Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72: 248–254

- Bugos RC, Chiang VL, Campbell WH (1991) cDNA cloning, sequence analysis and seasonal expression of lignin-bispecific caffeic acid/5-hydroxyferulic acid O-methyltransferase of aspen. *Plant Mol Biol* 17: 1203–1215
- Bugos RC, Chiang VL, Zhang XH, Campbell ER, Podila GK, Campbell WH (1995) RNA isolation from plant tissues recalcitrant to extraction in guanidine. *Biotechniques* 19: 734–737
- Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. *J Chem Phys* 126: 014101
- Bustin SA (2000) Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol* 25: 169–193
- China G, Padron G, Hooft RWW, Sander C, Vriend G (1995) The use of position-specific rotamers in model building by homology. *Proteins* 23: 415–421
- Darden T, York D, Pedersen L (1993) Particle Mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J Chem Phys* 98: 10089–10092
- Dixon M (1953) The determination of enzyme inhibitor constants. *Biochem J* 55: 170–171
- Dixon RA, Chen F, Guo D, Parvathi K (2001) The biosynthesis of monolignols: a “metabolic grid”, or independent pathways to guaiacyl and syringyl units? *Phytochemistry* 57: 1069–1084
- Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66: 486–501
- Ewing TJ, Makino S, Skillman AG, Kuntz ID (2001) DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J Comput Aided Mol Des* 15: 411–428
- Goffner D, Joffroy I, Grima-Pettenati J, Halpin C, Knight ME, Schuch W, Boudet AM (1992) Purification and characterization of isoforms of cinnamyl alcohol dehydrogenase (CAD) from *Eucalyptus xylem*. *Planta* 188: 48–53
- Greenbaum D, Colangelo C, Williams K, Gerstein M (2003) Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol* 4: 117
- Grima-Pettenati J, Feuillet C, Goffner D, Borderies G, Boudet AM (1993) Molecular cloning and expression of a *Eucalyptus gummi* cDNA clone encoding cinnamyl alcohol dehydrogenase. *Plant Mol Biol* 21: 1085–1095
- Hall TA (1999) Bioedit: a user friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 41: 95–98
- Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* 4: 435–447
- Higuchi T (2006) Look back over the studies of lignin biochemistry. *J Wood Sci* 52: 2–8
- Huson DH, Richter DC, Rausch C, DeZulian T, Franz M, Rupp R (2007) Dendroscope: an interactive viewer for large phylogenetic trees. *BMC Bioinformatics* 8: 460
- Iijima Y, Wang G, Fridman E, Pichersky E (2006) Analysis of the enzymatic formation of citral in the glands of sweet basil. *Arch Biochem Biophys* 448: 141–149
- Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B* 105: 6474–6487
- Katoh K, Kuma K, Toh H, Miyata T (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* 33: 511–518
- Kiedrowski S, Kawalleck P, Hahlbrock K, Somssich IE, Dangel JL (1992) Rapid activation of a novel plant defense gene is strictly dependent on the *Arabidopsis RPM1* disease resistance locus. *EMBO J* 11: 4677–4684
- Kim SJ, Kim MR, Bedgar DL, Moinuddin SG, Cardenas CL, Davin LB, Kang C, Lewis NG (2004) Functional reclassification of the putative cinnamyl alcohol dehydrogenase multigene family in *Arabidopsis*. *Proc Natl Acad Sci USA* 101: 1455–1460
- Knight ME, Halpin C, Schuch W (1992) Identification and characterization of cDNA clones encoding cinnamyl alcohol dehydrogenase from tobacco. *Plant Mol Biol* 19: 793–801
- Kutsuki H, Shimada M, Higuchi T (1982) Regulatory role of cinnamyl alcohol dehydrogenase in the formation of guaiacyl and syringyl lignins. *Phytochemistry* 21: 19–23
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* 26: 283–291
- Li L, Cheng XF, Leshkevich J, Umezawa T, Harding SA, Chiang VL (2001) The last step of syringyl monolignol biosynthesis in angiosperms is regulated by a novel gene encoding sinapyl alcohol dehydrogenase. *Plant Cell* 13: 1567–1586
- Li X, Ma D, Chen J, Pu G, Ji Y, Lei C, Du Z, Liu B, Ye H, Wang H (2012) Biochemical characterization and identification of a cinnamyl alcohol dehydrogenase from *Artemisia annua*. *Plant Sci* 193–194: 85–95
- Liljgren SJ, Ditta GS, Eshed Y, Savidge B, Bowman JL, Yanofsky MF (2000) SHATTERPROOF MADS-box genes control seed dispersal in *Arabidopsis*. *Nature (London)* 404: 766–770
- Logemann E, Reinold S, Somssich IE, Hahlbrock K (1997) A novel type of pathogen defense-related cinnamyl alcohol dehydrogenase. *Biol Chem* 378: 909–913
- Louie GV, Bowman ME, Tu Y, Mouradov A, Spangenberg G, Noel JP (2010) Structure-function analyses of a caffeic acid O-methyltransferase from perennial ryegrass reveal the molecular basis for substrate preference. *Plant Cell* 22: 4114–4127
- Ma Q-H (2010) Functional analysis of a cinnamyl alcohol dehydrogenase involved in lignin biosynthesis in wheat. *J Exp Bot* 61: 2735–2744
- MacKay JJ, Liu W, Whetten R, Sederoff RR, O'Malley DM (1995) Genetic analysis of cinnamyl alcohol dehydrogenase in loblolly pine: single gene inheritance, molecular characterization and evolution. *Mol Gen Genet* 247: 537–545
- Mansell RB, Babbel GR, Zenk MH (1976) Multiple forms and specificity of coniferyl alcohol dehydrogenase from cambial regions of higher plants. *Phytochemistry* 15: 1849–1853
- Meng EC, Pettersen EF, Couch GS, Huang CC, Ferrin TE (2006) Tools for integrated sequence-structure analysis with UCSF Chimera. *BMC Bioinformatics* 7: 339
- Mitsuda N, Seki M, Shinozaki K, Ohme-Takagi M (2005) The NAC transcription factors NST1 and NST2 of *Arabidopsis* regulate secondary wall thickenings and are required for anther dehiscence. *Plant Cell* 17: 2993–3006
- Nakatsubo T, Li L, Hattori T, Lu S, Sakakibara N, Chiang VL, Shimada M, Suzuki S, Umezawa T (2007) Roles of 5-hydroxyconiferaldehyde and caffeoyl CoA O-methyltransferases in monolignol biosynthesis in *Carthamus tinctorius*. *Cellul Chem Technol* 41: 511–520
- O'Malley DM, Porter S, Sederoff RR (1992) Purification, characterization, and cloning of cinnamyl alcohol dehydrogenase in Loblolly Pine (*Pinus taeda* L.). *Plant Physiol* 98: 1364–1371
- Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. *J Appl Phys* 52: 7182–7190

- Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25: 1605–1612
- Pillonel C, Hunziker P, Binder A (1992) Multiple forms of the constitutive wheat cinnamyl alcohol dehydrogenase. *J Exp Bot* 43: 299–305
- R Development Core Team (2011) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3–900051–07–0, URL <http://www.R-project.org/>
- Rodriguez R, Chinae G, Lopez N, Pons T, Vriend G (1998) Homology modeling, model and software evaluation: three related resources. *Bioinformatics* 14: 523–528
- Sakakibara N, Nakatsubo T, Suzuki S, Shibata D, Shimada M, Umezawa T (2007) Metabolic analysis of the cinnamate/monolignol pathway in *Carthamus tinctorius* seeds by a stable-isotope-dilution method. *Org Biomol Chem* 5: 802–815
- Sanner MF, Olson AJ, Spehner JC (1996) Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* 38: 305–320
- Sibout R, Eudes A, Mouille G, Pollet B, Lapierre C, Jouanin L, Séguin A (2005) Cinnamyl alcohol dehydrogenase-C and -D are the primary genes involved in lignin biosynthesis in the floral stem of *Arabidopsis*. *Plant Cell* 17: 2059–2076
- Somssich IE, Wernert P, Kiedrowski S, Hahlbrock K (1996) *Arabidopsis thaliana* defense-related protein ELI3 is an aromatic alcohol: NADP<sup>+</sup> oxidoreductase. *Proc Natl Acad Sci USA* 93: 14199–14203
- Stricker D (2008) BrightStat.com: free statistics online. *Comput Methods Programs Biomed* 92: 135–143
- Umezawa T (2003) Diversity in lignan biosynthesis. *Phytochem Rev* 2: 371–390
- Umezawa T (2010) The cinnamate/monolignol pathway. *Phytochem Rev* 9: 1–17
- Umezawa T, Ragamustari SK, Nakatsubo T, Wada S, Li L, Yamamura M, Sakakibara N, Hattori T, Suzuki S, Chiang VL (2013) A novel lignan O-methyltransferase catalyzing the regioselective methylation of matairesinol in *Carthamus tinctorius*. *Plant Biotechnol* 30: 97–109
- Valletta A, Trainotti L, Santamaria AR, Pasqua G (2010) Cell-specific expression of tryptophan decarboxylase and 10-hydroxygeraniol oxidoreductase, key genes involved in camptothecin biosynthesis in *Camptotheca acuminata* Decne (Nyssaceae). *BMC Plant Biol* 10: 69
- Vriend G (1990) WHAT IF: a molecular modeling and drug design program. *J Mol Graph* 8: 52–56
- Wiederstein M, Sippl MJ (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 35 (Web Server issue): W407–W410
- Williamson JD, Stoop JM, Massel MO, Conkling MA, Pharr DM (1995) Sequence analysis of a mannitol dehydrogenase cDNA from plants reveals a function for the pathogenesis-related protein ELI3. *Proc Natl Acad Sci USA* 92: 7148–7152
- Wyrambik D, Grisebach H (1975) Purification and properties of isoenzymes of cinnamyl-alcohol dehydrogenase from soybean-cell-suspension cultures. *Eur J Biochem* 59: 9–15
- Youn B, Camacho R, Moinuddin SGA, Lee C, Davin LB, Lewis NG, Kang C (2006) Crystal structures and catalytic mechanism of the *Arabidopsis* cinnamyl alcohol dehydrogenases AtCAD5 and AtCAD4. *Org Biomol Chem* 4: 1687–1697

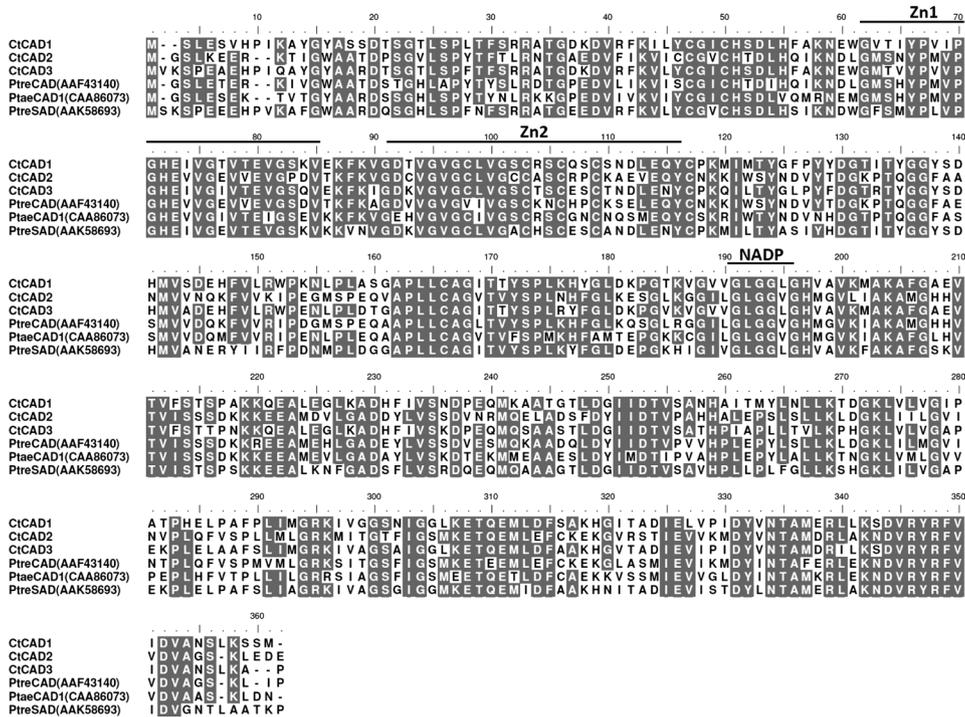


Figure S1. Alignment of CtCAD1, CtCAD2, CtCAD3 and selected CAD amino acid sequences. The alignment was created using MAFFT 5.0 and edited using Bioedit sequence alignment editor. Lines above the alignment indicate Zn1, Zn2, and NADP conserved binding domains.

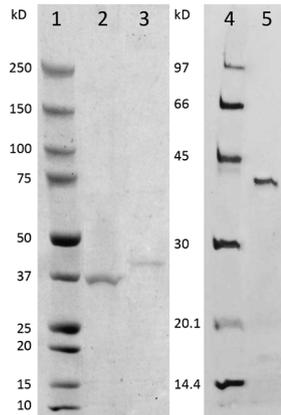


Figure S2. SDS-PAGE analysis of purified CtCADs. Proteins were stained with Coomassie Blue staining after His-Bind chromatography and desalting. Lane 1: Precision Plus Protein Dual Color Standards (BioRad, Hercules, CA, USA); lane 2: Purified CtCAD1; lane 3: Purified CtCAD2; lane 4: Molecular weight protein markers (Pharmacia, Uppsala, Sweden); lane 5: Purified CtCAD3. Recombinant CtCADs were purified to homogeneity using His-bind chromatography.

Table S1. Correlation analysis of *CtCAD* gene expression with specific activity of total plant protein towards coniferaldehyde or sinapaldehyde. Statistical analyses provided Pearson correlation and  $R^2$  values, which determined the strength of the correlation. Values closer to 1 indicate stronger correlation between the two data.

Correlation (gene expression-plant protein assay)	Pearson correlation	$R^2$ value
<i>CtCAD1</i> -coniferaldehyde assay	-0.06	0.017
<i>CtCAD1</i> -sinapaldehyde assay	-0.05	0.022
<i>CtCAD2</i> -coniferaldehyde assay	0.88	0.803
<i>CtCAD2</i> -sinapaldehyde assay	0.83	0.683
<i>CtCAD3</i> -coniferaldehyde assay	0.84	0.701
<i>CtCAD3</i> -sinapaldehyde assay	0.79	0.620

Table S2. Distribution of the CtCAD residues in the Ramachandran plot calculated by PROCHECK (Laskowski et al. 1993).

Protein model	Core	Allowed	Generously allowed	Disallowed
CtCAD1	89.70%	10.00%	0.00%	0.30%
CtCAD2	74.50%	21.80%	2.00%	1.70%
CtCAD3	90.30%	9.40%	0.00%	0.30%
PtreSAD	89.10%	10.60%	0.30%	0.00%
AtCAD5	74.70%	20.60%	3.40%	1.40%