Isolation of cinnamyl alcohol dehydrogenase cDNAs from two important economic species : alfalfa and poplar. Demonstration of a high homology of the gene within angiosperms

J. van Doorsselaere, Marie Baucher, Catherine Feuillet, A.M. Boudet, Marc van Montagu, Dirk Inzé

Plant Physiol. Biochem. 1995, 33 (1), 105-109

Short Communication

Isolation of cinnamyl alcohol dehydrogenase cDNAs from two important economic species: alfalfa and poplar. Demonstration of a high homology of the gene within angiosperms

Jan Van Doorsselaere[1], Marie Baucher[1], Catherine Feuillet[2], Alain M. Boudet[2], Marc Van Montagu[1]* and Dirk Inzé[3]

[1] Laboratorium voor Genetica, Universiteit Gent, K. L. Ledeganckstraat 35, B-9000 Gent, Belgium.
[2] Centre de Biologie et Physiologie Végétales, Université Paul Sabatier, (CNRS-URA, 1457), 118, route de Narbonne, F-31062 Toulouse Cedex, France.
[3] Laboratoire Associé de l'Institut National de la Recherche Agronomique, France, Universiteit Gent, K. L. Ledeganckstraat 35, B-9000 Gent, Belgium.
* Author to whom correspondence should be addressed (fax 32-92645349; E-mail mamon@gengenp.rug.ac.be)

Abstract Cinnamyl alcohol dehydrogenase (CAD), a key enzyme in lignification, was cloned from poplar (*Populus trichocarpa* x *P. deltoides*) and alfalfa (*Medicago sativa* L.). Both cDNAs contain one continuous open reading frame encoding a protein of 357 amino acid residues (M_r 39,036) for the poplar CAD and 358 amino acids (M_r 38,950) for the alfalfa CAD. Comparison of the poplar and alfalfa CAD with other reported plant CAD sequences confirms the identity of the isolated cDNAs and shows the poplar and alfalfa CAD to be closely related to CAD from tobacco, eucalyptus and *Aralia cordata* (80% amino acid identity) while 70% amino acid identity was found with the CAD from the gymnosperm spruce. Sequence comparisons allowed also to identify important conserved domains and specific functional motifs.

Key words Lignin, cinnamyl alcohol dehydrogenase, cDNA cloning, sequence comparisons, poplar, alfalfa.

Abbreviations CAD, Cinnamyl alcohol dehydrogenase (EC 1.1.1.195).

INTRODUCTION

Lignin is an aromatic biopolymer found in higher plants (ferns, gymnosperms and angiosperms), especially in secondary cell walls of vascular tissue (*e.g.* xylem) where it provides strength and rigidity to the cell walls and an impermeable barrier for nutrient conductance (Lewis and Yamamoto, 1990).

Lignin results from the dehydrogenative polymerization of three different monomers (or monolignols) sinapyl alcohol, coniferyl alcohol and para-coumaryl alcohol. These monolignols are formed by reduction of the corresponding aldehydes (sinapaldehyde, coniferaldehyde and para-coumaraldehyde), a reaction catalyzed by the cinnamyl alcohol dehydrogenase (CAD; EC 1.1.1.195).

Recently the CAD enzyme has been purified from different plants and tissues (for a review see Boudet *et al.*, 1994). In these species, the enzyme appears as a dimer with a subunit molecular mass of approximately 40 kDa. The pine and spruce CAD use mainly coniferaldehyde and *p*-coumaraldehyde as substrate and not sinapaldehyde; this is in contrast with CAD from angiosperms which use all three cinnamaldehydes. In parallel studies, *cad* cDNAs have been isolated from tobacco (Knight *et al.*, 1992), eucalyptus (Grima-Pettenati *et al.*, 1993), spruce (Galliano *et al.*, 1993), *Aralia cordata* (Hibino *et al.*,

Plant Physiol. Biochem., 0981-9428/95/01/$ 4.00/© Gauthier-Villars

1993) and loblolly pine (O'Malley *et al.*, 1992).

Here, we would like to report on the isolation and characterization of *cad* cDNAs from alfalfa and poplar, two species of economic importance, poplar for the pulp industry and alfalfa as a forage crop. The negative impact of lignin on forage digestibility (Akin and Chesson, 1989) and paper pulping (Chiang *et al.*, 1988) is well established. The cloning of these genes could open perspectives for modifying lignin content or quality in these plants (Halpin *et al.*, 1994). In addition, sequence comparisons with CAD from tobacco, eucalyptus, *Aralia cordata* and spruce were also performed in order to evaluate the conservation of the gene within angiosperms and gymnosperms.

RESULTS AND DISCUSSION

Isolation of poplar and alfalfa *cad* cDNAs

Screening of a poplar leaf cDNA library with the tobacco CAD (PTCAD19) as a probe resulted in the isolation of two clones with inserts of different sizes. Partial sequencing showed them to be representative of the same mRNA species. The nucleotide sequence of the longest cDNA clone (cad.PdxPt.1, 1,305 bp, EMBL/Genbank accession No. Z19568) coding for the cinnamyl alcohol dehydrogenase from poplar contains one open reading frame (ORF) of 1071 nucleotides ending at nucleotide 1099, a 5' untranslated end of 27 nucleotides, a 3' end of 204 nucleotides and a putative poly(A)+ signal (position 1123-1128). The ORF corresponds to a protein of 357 amino acids (calculated M_r 39,036, Pi 5.94).

In addition, a λZAPII cDNA library from elicitor treated alfalfa cells was screened using the cad.PdxPt.1 insert as a probe. One positive clone designated cad.Ms.1 (insert size 2.7 kb; EMBL/Genbank accession No. Z19573) was partially sequenced. The CAD ORF consisted of 1074 nucleotides encoding a protein of 358 amino acids (calculated M_r 38,950; Pi 5.43). An eighty nucleotide 5'-untranslated sequence was found upstream from the ATG (position 81-83). A possible poly(A)+ signal (AATAG) appears at positions 1307-1311.

Table 1. *Percentage identity (upper) and similarity (lower) at the amino acid level between the CAD from eucalyptus (EUCAD), poplar (POCAD), tobacco (TBCAD), alfalfa (ALCAD),* Aralia cordata *(ARCAD) and spruce (SPCAD).*

	EUCAD	POCAD	TBCAD	ALCAD	ARCAD	SPCAD
EUCAD	XXX					
POCAD	80.9	XXX				
	10.4					
TBCAD	78.1	79.8	XXX			
	12.1	11.2				
ALCAD	77.8	80.1	78.2	XXX		
	12.1	11.2	11.8			
ARCAD	81.5	84.0	81.2	80.2	XXX	
	10.7	8.4	9.8	11.7		
SPCAD	68.3	69.2	69.7	67.8	69.7	XXX
	16.0	16.0	15.4	16.2	16.0	

Evolutionary relationship between CAD sequences

As mentioned previously, *cad* cDNAs have been isolated from tobacco, eucalyptus, *Aralia cordata* and spruce. The identity of the eucalyptus and spruce clone was unambiguously demonstrated by production of functional CAD proteins. Table 1 shows the percentages identity and similarity (at the amino acid level) found between six plant cinnamyl alcohol dehydrogenases. Both the poplar and alfalfa CAD

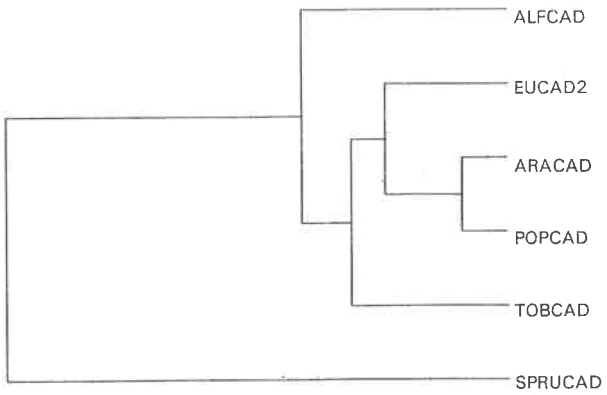

Figure 1. *Evolutionary relationship between alfalfa (ALFCAD), eucalyptus (EUCAD2),* Aralia cordata *(ARACAD), poplar (POPCAD), tobacco (TOBCAD) and spruce (SPRUCAD) CAD.*

Figure 2. *Comparison of the alfalfa (ALFCAD), eucalyptus (EUCAD2),* Aralia cordata *(ARACAD), poplar (POPCAD), tobacco (TOBCAD) and spruce (SPRUCAD) sequences.* The general Zn-containing AD motif (positions 68-82) and the Co-enzyme-binding domain (positions 188-193) are underlined. (■), active site Zn-binding cysteines (position 47 and 163) and one histidine (position 69); (▼), structural Zn ligand cysteine (position 100, 103, 106 and 114); (●●●), targeting signal for microbodies. Identical amino acids are denoted with an asterisk, similar amino acids are also indicated (.).

```
                                                                    ■
ALFCAD      MGSIEAAERTTVGLAAKDPSGILTPYTYTLRNTGPDDVYIKIHYCGVCHS     50
EUCAD2      MGSLEK-ERTTTGWAARDPSGVLSPYTYSLRNTGPEDLYIKVLSCGVCHS     49
ARACAD      MGSLEA-ERKTTGWAARDPSGVLSPYTYTLRETGPEDVFIKIIYCGICHT     49
POPCAD      MGSLET-ERKIVGWAATDSTGHLAPYTYSLRDTGPEDVFIKVISCGVCHT     49
TOBCAD      MGSLDV-EKSAIGWAARDPSGLLSPYTYTLRNTGPEDVQVKVLYCGLCHS     49
SPRUCAD     MGSLES-ERTVTGYAARDSSGHLSPYTYTLRNKGPEDVIVRVIYCGICHS     49
            ***..  *.. .* ** *..* *.****.**..**.*. ...  **.**.

                        ■
ALFCAD      DLHQIKNDLGMSNYPMVPGHEVVGEVLEVGSNVTRFKVGEIVGVGLLVGC    100
EUCAD2      DIHQIKNDLGMSHYPMVPGHEVVGEVLEVGSEVTKYRVGDRVGTGIVVGC     99
ARACAD      DIHQIKNDLGASNYPMVPGHEVVGEVVEVGSDVTKFKVGDCVGDGTIVGC     99
POPCAD      DIHQIKNDLGMSHYPMVPGHEVVGEVVEVGSDVTRFKVGDVVGVGVIVGS     99
TOBCAD      DLHQVKNDLGMSNYPLVPGHEVVGKVVEVGADVSKFKVGDTVGVGLLVGS     99
SPRUCAD     DLVQMHNEMGMSNYPMVPGHEVVGVVTEIGSEVKKFKVGEHVGVGCIVGS     99
            *. *..*..* *.**.******** * *.*..*....**. ** * .**.

            ▼  ▼  ▼           ▼
ALFCAD      CKSCRACDSEIEQYCNKKIWSYNDVYTDGKITQGGFAESTVVEQKFVVKI    150
EUCAD2      CRSCSPCNSDQEQYCNKKIWNYNDVYTDGKPTQGGFAGEIVVGERFVVKI    149
ARACAD      CKTCRPCKADVEQYCNKKIWSYNDVYTDGKPTQGGFSGHMVVDQKFVVKI    149
POPCAD      CKNCHPCKSEIEQYCNKKIWSYNDVYTDGKPTQGGFAESMVVHQKFVVRI    149
TOBCAD      CRNCGPCKREIEQYCNKKIWNCNDVYTDGKPTQGGFANSMVVDQNFVVKI    149
SPRUCAD     CRSCSNCNGSMEQYCSKRIWTYNDVNHDGTPTQGGFASSMVVDQMFVVRI    149
            *..*  *. . ****.*.**..***  **. *****..  ** . ***.*

                        ■
ALFCAD      PEGLAPEQVAPLLCAGVTVYSPLSHFGLKT-PGLRGGILGLGGVGHMGVK    199
EUCAD2      PDGLESEQAAPLMCAGVTVYSPLVRFGLKQ-SGLRGGILGLGGVGHMGVK    198
ARACAD      PDGMAPEQAAPLLCAGVTVYSPLTHFGLKEISGLRGGILGLGGVGHMGVK    199
POPCAD      PDGMSPEQAAPLLCAGLTVYSPLKHFGLKQ-SGLRGGILGLGGVGHMGVK    198
TOBCAD      PEGMAPEQAAPLLCAGITVYSPFNHFGFNQ-SGFRGGILGLGGVGHMGVK    198
SPRUCAD     PENLPLEQAAPLLCAGVTVYSPMKHFGMTE-PGKKCGILGLGGVGHMGVK    198
            *...  **.***.***.*****. .**..  .* . **************

ALFCAD      VAKALGHHVTVISSSDKKKKEALEDLGADNYLVSSDTVGMQEAADSLDYI    249
EUCAD2      IAKAMGHHVTVISSSDKKRTEALEHLGADAYLVSSDENGMKEATDSLDYI    248
ARACAD      LAKAMGHHVTVISSSDKKKEEAIDHLGADAYLVSSDATQMQEAADSLDYI    249
POPCAD      IAKAMGHHVTVISSSDKKREEAMEHLGADEYLVSSDVESMQKAADQLDYI    248
TOBCAD      IAKAMGHHVTVISSSNKKRQEALEHLGADDYLVSSDTDKMQEAADSLDYI    248
SPRUCAD     IAKAFGLHVTVISSSDKKKEEALEVLGADAYLVSKDAEKMQEAAESLDYI    248
            .***.* ********.**. **.. ****.****.*   *..*.. ****

ALFCAD      IDTVPVGHPLEPYLSLLKIDGKLILMGVINTPLQFVTPMVMLGRKSITGS    299
EUCAD2      FDTIPVVHPLEPYLALLKLDGKLILTGVINAPLQFISPMVMLGRKSITGS    298
ARACAD      IDTVPVFHPLEPYLSLLKLDGKLILMGVINTPLQFISPMVMLGRKAITGS    299
POPCAD      IDTVPVVHPLEPYLSLLKLDGKLILMGVINAPLQFVTPMVMLGRKSITGS    298
TOBCAD      IDTVPVGHPLELYLSLLKIDGKLILIGVINTPLQFISPMVMLGRKSITGS    298
SPRUCAD     MDTIPVAHPLEPYLALLKTNGKLVMLGVVPEPLHFVTPLLILGRRSIAGS    298
            .**.** **** **.*** .***.. **. .**.*..*...***..*.**

ALFCAD      FVGSVKETEEMLEFWKEKGLTSMIEIVTMDYINKAFERLEKNDVRYRFVV    349
EUCAD2      FIGSMKETEEMLEFCKEKGLTSQIEVIKMDYVNTALERLEKNDVRYRFVV    348
ARACAD      FIGSMKETEEMLDFCNEKGITSTIEVVKMDYINTAFERLEKNDVRYRFVV    349
POPCAD      FIGSMKETEEMLEFCKEKGVASMIEVIKMDYINTAFERLEKNDVRYRFVV    348
TOBCAD      FIGSMKETEEMLDFCKEKGVTSQIEIVKMDYINTAMERLEKNDVSYRFVV    348
SPRUCAD     FIGSMEETQETLDFCAEKKVSSMIEVVGLDYINTAMERLVKNDVRYRFVV    348
            *.**..**.* *.*  ** ..* **.. .**.*.*.*** ****.*****

            ●●●
ALFCAD      DVKGSKFEE--    358
EUCAD2      DVVGSKLD---    356
ARACAD      DVAGSKLDQET    360
POPCAD      DVAGSKLIH--    357
TOBCAD      DVAGSKLDQ--    357
SPRUCAD     DVAASNLDK--    357
            ** .*..
```

have approximately 80% homology (identity) with the CAD from eucalyptus confirming the true nature of the corresponding cDNAs. Seventy % amino acid identity was found with the spruce CAD. The percentages of identity found between the different CAD proteins allowed us to establish a phylogenetic tree using the CLUSTAL program (PC/GENE) (*fig.* 1). The high homology found between the angiosperm CAD sequences indicates a close relationship between these species. As expected, a lower homology is found between angiosperm CAD and spruce CAD. Hereby we have to mention that the complete CAD sequence of only one gymnosperm (spruce) is presently available. Comparison of the N-terminal amino acid sequence of loblolly pine (O'Malley *et al.*, 1992) and spruce showed 91.5% amino acid identity (while only approximately 60% was found with the tobacco CAD N-terminal sequence) (Galliano *et al.*, 1993) demonstrating a close evolutionary relationship of pine and spruce.

Comparison of the alfalfa and poplar CAD with other plant CAD sequences and identification of important motifs

Sequence analysis of the poplar and alfalfa CAD shows that they can be classified in the group of the long chain Zn containing alcohol dehydrogenases (ADH). These ADHs (from both procaryotic and eucaryotic species) share 22 amino acid residues which are strictly conserved (Jörnvall *et al.*, 1987). From these 22 residues, we were able to identify 19 in the poplar and alfalfa CAD (and in other plant CAD proteins, *fig.* 2). These residues include Gly 188, 190, 193 and a basic residue (Arg or Lys) at position 217 (important for Co-enzyme binding), Cys (47 and 163) and His 69 involved in binding the Zn atom at the active site and Cys 100, 103, 106 and 114 implicated in binding a second (structural) Zn atom and three acidic amino acids (Asp 50, Asp 89 and Glu 70) with important binding properties. Between positions 68 and 82 a consensus sequence Gly-His-Glu-X-X-Gly-X-X-X-X-X-Gly-X-X-Val, present in all long chain Zn ADHs, is seen. Like the tobacco, *Aralia cordata* and the eucalyptus CAD, the poplar CAD contains a Ser-Lys-Leu-COOH terminal targeting signal for microbodies (amino acid residues 353-355) (Keller *et al.*, 1991). The Leu is replaced by a Phe in the alfalfa CAD. It is not known whether this sequence in the CAD has a function *in vivo*.

CONCLUSION

Due to its specific role in lignin synthesis, CAD is one of the potential target enzymes for modulating the quality and quantity of lignins in plants by genetic engineering. The cloning of CAD from poplar (a commercially important hardwood for pulp and paper industry) and alfalfa (frequently used as forage crop) could be important for modifying monolignol biosynthesis in these plants. By expressing antisense constructs in these organisms, a different form of lignin could eventually be achieved (in progress). Genetically modified poplars might be produced with altered wood and better bleaching qualities. In alfalfa the digestibility might be improved. In addition our results show that to down regulate the expression of *cad* genes in other angiosperm plants (via antisense technology), one of the *cad* cDNAs described here could be used since *cad* sequences from angiosperms are highly homologous.

METHODS

The poplar *cad* cDNA (cad.PdxPt.1) was isolated by screening a poplar (*Populus trichocarpa* x *P. deltoides*) leaf cDNA library in pUC18 with the full-length tobacco *cad* cDNA (pTCAD19; Knight *et al.*, 1992) as a probe. The alfalfa *cad* cDNA (cad.Ms.1) was isolated from an alfalfa (*Medicago sativa* L.) cDNA library in λZAPII (Gowri *et al.*, 1991) by using the cad.PdxPt.1 insert as a probe. The screening of both libraries was essentially performed according to standard procedures (Maniatis *et al.*, 1982). One positive alfalfa clone was recovered as an *Eco*RI fragment in pBluescript (Stratagene) by *in vivo* excision from λZAPII following the manufacturer's protocol. DNA sequence determinations were carried out on DNA isolated by a Qiagen plasmid kit by the dideoxy chain termination method (Sanger *et al.*, 1977) and the Maxam and Gilbert procedure (1977). DNA sequences were analyzed using Ig (Intelligenetics) and PC/GENE programs.

Acknowledgements. The authors wish to thank Vera Vermaercke and Martine De Cock for preparing the figures and the manuscript; Dr. Simon Hawkins for providing the slide for the cover front of this issue; Jacqueline Grima-Pettenati and Tom Gerats for critical reading of the manuscript. This work was supported by a grant from the Commission of the European Communities (ECLAIR-OPLIGE no. AGRE-0021-C[EDB]). JVD is indebted to the "Instituut ter aanmoediging van het Wetenschappelijk Onderzoek in Nijverheid en Landbouw" for a predoctoral fellowship; DI is a research Director of the Institut National de la Recherche Agronomique (France).

(Received June 23, 1994; accepted August 3, 1994)

REFERENCES

Akin D. E. and Chesson A., 1989. Lignification as the major factor limiting forage feeding value especially in warm conditions. In *Proceedings 16th International Grassland Congress, Association française pour la production fourragère*, ed., Versailles, INRA, 1753-1760.

Boudet A. M., Lapierre C. and Grima-Pettenati J., 1994. Biochemistry and molecular biology of lignification. *New Phytol.*, **00,** 000-000.

Chiang V. L., Puumala R. J., Takeuchi H. and Eckert R. E., 1988. Comparison of softwood and hardwood kraft pulping. *Tappi J.*, **71,** 173-176.

Galliano H., Cabane M., Eckerskorn C., Lottspeich F., Sandermann H. Jr. and Ernst D., 1993. Molecular cloning, sequence analysis and elicitor-/ozone-induced accumulation of cinnamyl alcohol dehydrogenase from Norway spruce (*Picea abies* L.). *Plant Mol. Biol.*, **23,** 145-156.

Gowri G., Bugos R. C., Campbell W. H., Maxwell C. A. and Dixon R. A., 1991. Stress responses in alfalfa (*Medicago sativa* L.) X. Molecular cloning and expression of *S*-adenosyl-L-methionine:caffeic acid 3-*O*-methyltransferase, a key enzyme of lignin biosynthesis. *Plant Physiol.*, **97,** 7-14.

Grima-Pettenati J., Feuillet C., Goffner D., Borderies G. and Boudet A. M., 1993. Molecular cloning and expression of a *Eucalyptus gunnii* cDNA clone encoding cinnamyl alcohol dehydrogenase. *Plant Mol. Biol.*, **21,** 1085-1095.

Halpin C., Knight M. E., Foxon G. A., Campbell M. M., Boudet A. M., Boon J. J., Chabbert B., Tollier M. T. and Schuch W., 1994. Manipulation of lignin quality by downregulation of cinnamyl alcohol dehydrogenase. *Plant J.*, **6,** 339-350.

Hibino T., Shibata D., Chen J.-Q. and Higuchi T., 1993. Cinnamyl alcohol dehydrogenase from *Aralia cordata*: Cloning of the cDNA and expression of the gene in lignified tissues. *Plant Cell Physiol.*, **34,** 659-665.

Jörnvall H., Persson B. and Jeffery J., 1987. Characteristics of alcohol/polyol dehydrogenases. The zinc-containing long-chain alcohol dehydrogenase. *Eur. J. Biochem.*, **167,** 195-201.

Keller G.-A., Krisans S., Gould S. J., Sommer J. M., Wang C. C., Schliebs W., Kunau W., Brody S. and Subramani S., 1991. Evolutionary conservation of a microbody targeting signal that targets proteins to peroxisomes, glyoxisomes, and glycosomes. *J. Cell Biol.*, **114,** 893-904.

Knight M. E., Halpin C. and Schuch W., 1992. Identification and characterization of cDNA clones encoding cinnamyl alcohol dehydrogenase from tobacco. *Plant Mol. Biol.*, **19,** 793-801.

Lewis G. N. and Yamamoto E., 1990. Lignin: occurrence, biogenesis and biodegradation. *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **41,** 455-496.

Maniatis T., Fritsch E. F. and Sambrook J., 1982. Molecular cloning, a laboratory manual. New York, Cold Spring Harbor Laboratory.

Maxam A. M. and Gilbert W., 1977. A new method of sequencing DNA. *Proc. Natl. Acad. Sci. USA*, **74,** 560-564.

O'Malley D. M., Porter S. and Sederoff R. R., 1992. Purification, characterization, and cloning of cinnamyl alcohol dehydrogenase in loblolly pine (*Pinus taeda* L.). *Plant Physiol.*, **98,** 1364-1371.

Sanger F., Nicklen S. and Coulson A. R., 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*, **74,** 5463-5467.