Crystal structures and catalytic mechanism of the *Arabidopsis* cinnamyl alcohol dehydrogenases AtCAD5 and AtCAD4†

Buhyun Youn,[a] Roy Camacho,[a] Syed G. A. Moinuddin,[b] Choonseok Lee,[b] Laurence B. Davin,[b] Norman G. Lewis[b] and ChulHee Kang*[a]

Received 3rd February 2006, Accepted 24th February 2006
First published as an Advance Article on the web 4th April 2006
DOI: 10.1039/b601672c

The cinnamyl alcohol dehydrogenase (CAD) multigene family *in planta* encodes proteins catalyzing the reductions of various phenylpropenyl aldehyde derivatives in a substrate versatile manner, and whose metabolic products are the precursors of structural lignins, health-related lignans, and various other metabolites. In *Arabidopsis thaliana*, the two isoforms, AtCAD5 and AtCAD4, are the catalytically most active being viewed as mainly involved in the formation of guaiacyl/syringyl lignins. In this study, we determined the crystal structures of AtCAD5 in the apo-form and as a binary complex with $NADP^+$, respectively, and modeled that of AtCAD4. Both AtCAD5 and AtCAD4 are dimers with two zinc ions per subunit and belong to the Zn-dependent medium chain dehydrogenase/reductase (MDR) superfamily, on the basis of their overall 2-domain structures and distribution of secondary structural elements. The catalytic Zn^{2+} ions in both enzymes are tetrahedrally coordinated, but differ from those in horse liver alcohol dehydrogenase since the carboxyl side-chain of Glu70 is ligated to Zn^{2+} instead of water. Using AtCAD5, site-directed mutagenesis of Glu70 to alanine resulted in loss of catalytic activity, thereby indicating that perturbation of the Zn^{2+} coordination was sufficient to abolish catalytic activity. The substrate-binding pockets of both AtCAD5 and AtCAD4 were also examined, and found to be significantly different and smaller compared to that of a putative aspen sinapyl alcohol dehydrogenase (SAD) and a putative yeast CAD. While the physiological roles of the aspen SAD and the yeast CAD are uncertain, they nevertheless have a high similarity in the overall 3D structures to AtCAD5 and 4. With the *bona fide* CAD's from various species, nine out of the twelve residues which constitute the proposed substrate-binding pocket were, however, conserved. This is provisionally considered as indicative of a characteristic fingerprint for the CAD family.

Introduction

Cinnamyl alcohol dehydrogenases (CAD's, EC 1.1.1.195) catalyze the substrate versatile, NADPH-dependent, conversion of *p*-hydroxycinnamyl aldehydes **1–5** *in vitro* into the corresponding alcohols **7–11**,[1,2] (Fig. 1). The latter metabolites, in turn, are obligatory precursors of *e.g.* structural cell-wall lignins[3,4] and health-related lignans,[5–10], including the antiviral agent podophyllotoxin (**13**) from *Podophyllum* species.[8] Podophyllotoxin (**13**) finds application in the semi-synthesis of the widely employed anticancer compounds teniposide (**14**), etoposide (**15**) and Etopophos® (etoposide phosphate, **16**).[11] CAD is also an essential step on the pathway to other lignans, such as matairesinol (**17**)[8] and secoisolariciresinol (**18**),[5,6] which can serve as plant precursors of the "mammalian" lignans, enterolactone (**19**) and enterodiol (**20**).[12,13]

In *Arabidopsis*, there are nine Zn^{2+}-dependent CAD's or homologues thereof[14] belonging to the medium chain dehydrogenase/reductase (MDR) superfamily, with two Zn^{2+} ions per subunit.[15] This MDR superfamily is found in a variety of organisms, such as bacteria, fungi, plants, cephalopods and vertebrates, of which the best known example is horse liver alcohol dehydrogenase.[16] In *Arabidopsis*, two isoforms of the CAD family [AtCAD4 (At3g19450) and AtCAD5 (At4g34230)] are catalytically the most active *in vitro* with *p*-hydroxycinnamyl aldehydes **1–5**, albeit differing in their relative abilities to process sinapyl aldehyde (**5**); AtCAD5 is evidently ~270-fold more efficient than AtCAD4.[14] Overall, however, *p*-hydroxycinnamyl aldehyde (**1**) is the preferred substrate for both isoforms, as well as for the catalytically less active AtCAD2, 3, 7 and 8. The remaining 3 putative homologues (AtCAD1, 6 and 9) displayed no activities when incubated with substrates **1–5**.

In terms of the precise physiological function, a double mutant of both AtCAD4 and 5 in *Arabidopsis*[17] markedly reduced the ability to synthesize either coniferyl (**9**) or sinapyl (**11**) alcohols, and hence the formation of the corresponding guaiacyl (G) and syringyl (S) lignins. This suggests therefore that in *Arabidopsis* AtCAD4 and 5 are largely responsible for the formation of **9** and **11** in lignifying tissues.

Much of our current understanding of alcohol dehydrogenases (ADH's) stems from comprehensive studies on the aliphatic horse liver ADH crystal structure,[18] and this knowledge has been applied to many other aliphatic ADH's. By contrast,

[a]*School of Molecular Biosciences, Washington State University, Pullman, WA 99164-4660, USA. E-mail: chkang@wsunix.wsu.edu; Fax: 1 509 335 9688; Tel: 1 509 335 1409*
[b]*Institute of Biological Chemistry, Washington State University, Pullman, WA 99164-6340, USA*

† Abbreviations: ADH, alcohol dehydrogenase; CAD, cinnamyl alcohol dehydrogenase; MDR, medium chain dehydrogenase/reductase; PDB, protein data bank; SAD, sinapyl alcohol dehydrogenase.

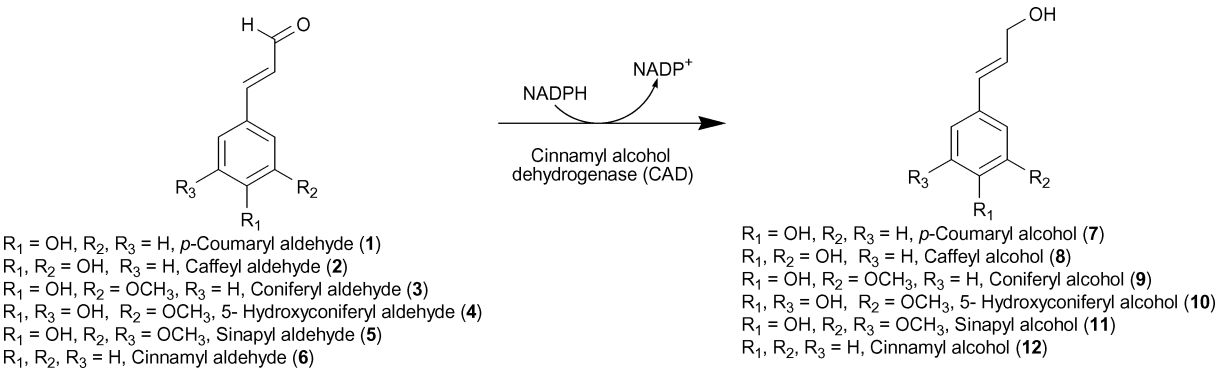
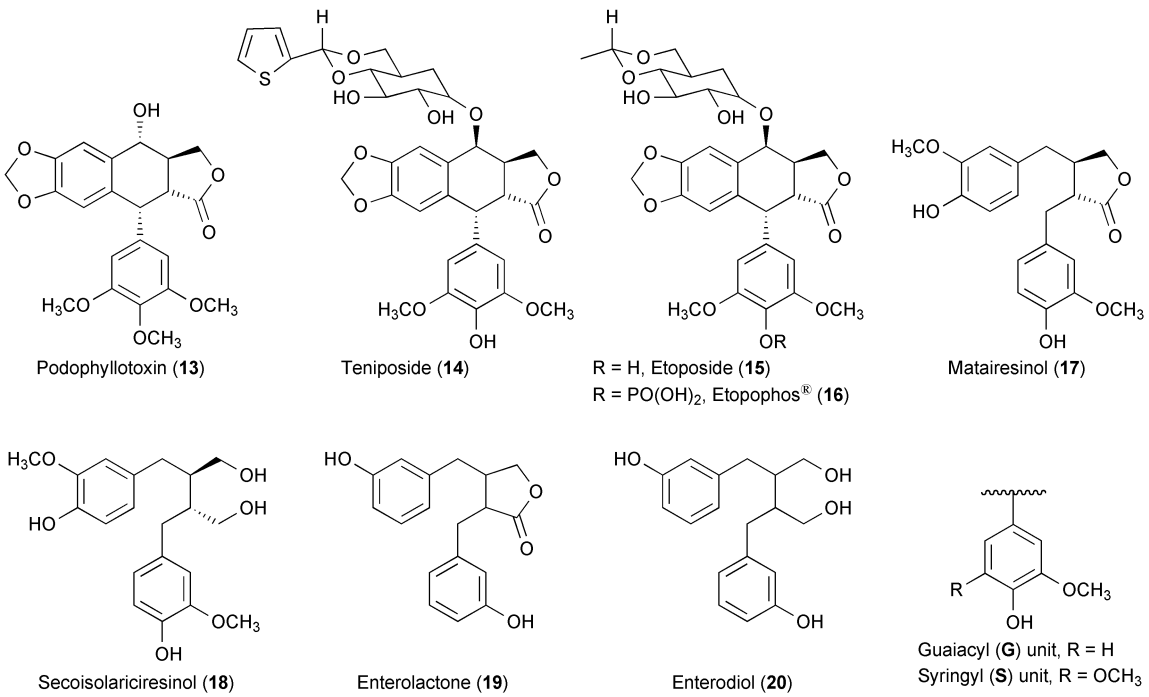


Fig. 1 Cinnamyl alcohol dehydrogenase substrates and products.



little is known about aromatic alcohol dehydrogenases, such as CAD's, in spite of their enormous importance in vascular plant formation, in health-related areas, in biotechnology and so forth. There are useful reports, however, of crystal structures of a putative CAD in yeast (*Saccharomyces cerevisiae*)[19] and an aspen sinapyl alcohol dehydrogenase (SAD).[20] These enzymes, like horse liver ADH, show some similarity to AtCAD5 and AtCAD4 in terms of their amino acid sequences, even though their precise biochemical/physiological significance is uncertain. *S. cerevisiae* lacks a biochemical pathway to either monolignols **7–11** or cinnamyl alcohol **12**, and the biochemical/physiological role of the putative SAD is also uncertain. Furthermore, current evidence of the biochemical/physiological properties of the *Arabidopsis* CAD isoforms *in vitro*[14] and *in vivo*[17] has given no indication that there is a SAD-specific enzyme as previously reported.[21] Indeed, from an enzymological perspective, the aspen SAD displays considerable substrate versatility for aldehydes **1–5** *in vitro*, suggesting that there is no role specific to a particular substrate. The analysis of the *Arabidopsis* AtCAD4 and AtCAD5 double mutant also indicated that there was not a specific SAD responsible solely for sinapyl alcohol (**11**) formation in lignification.[17]

In this study, we examined the crystal structures of the two *bona fide* CAD's, AtCAD4 and AtCAD5, both of which have established roles in monolignol, lignan and lignin formation *in vivo*. In addition, we considered it instructive to conduct a detailed comparison of AtCAD4 and AtCAD5 with the putative aspen SAD and yeast CAD, particularly with respect to their binding site geometries and amino acid compositions.

Results and discussion

Overall structure

Recombinant AtCAD5 was first crystallized in both its apo-form and as a binary complex, with crystals of the latter obtained by mixing with $NADP^+$. The structure of the apo-form was determined at 2.0 Å resolution by molecular replacement using coordinates of a yahK, a zinc-type alcohol dehydrogenase-like protein (1UUF) from *Escherichia coli*, which at the initiation of

our studies had the highest sequence similarity to AtCAD5 in the Protein Data Bank (PDB). The binary complex structure of AtCAD5 (with $NADP^+$) was determined at 2.6 Å resolution, using the coordinates of the deduced structure of the apo-form.

The asymmetric unit of the AtCAD5 crystal is composed of one molecule and the crystallographic 2-fold axis produces a dimer tightly associated through two 2-fold related β strands (βF). Consequently, its dimer forms an extended 12 stranded β-sheet with 6 strands (βA-βF) from each subunit (Fig. 2). Additionally, with the differences in substrate preferences noted for AtCAD5 and AtCAD4 regarding sinapyl aldehyde (**5**),[14] the structure of AtCAD4 was also modeled using the refined coordinates of AtCAD5 followed by the process of energy minimization. As expected, since AtCAD4 has 77.5% identity with AtCAD5, the overall backbone structure of the modeled AtCAD4 had no significant changes (Fig. 2, inset).

Fig. 2 Crystal structure of AtCAD5 homodimer and energy minimized model of AtCAD4 (inset). The catalytic and nucleotide-binding domains are colored in light blue and violet for the bottom subunit, and green and dark orange for the upper subunit, respectively. Catalytic and structural Zn^{2+} are depicted as red dots. Secondary structural elements have been numbered sequentially as α1–α6/αA–αE and β1–β12/βA–βF for the α-helices and β-strands, respectively.

AtCAD5 and AtCAD4 were studied by light scattering to determine their tendencies to form oligomers. Both static and dynamic light-scattering experiments further confirmed that most of the population of AtCAD5 and AtCAD4 were in the dimer form in both phosphate buffered saline (pH 7.0) and 20 mM Tris–HCl buffer (pH 8.0) (Fig. 3). This quaternary structure of AtCAD5 and AtCAD4 is similar to that of other Zn-containing ADH's from higher plants and mammals. In addition, the elution volumes of both AtCAD4 and AtCAD5 reflected their extended hydrodynamic volumes in agreement with the asymmetric shape of the dimers shown in Fig. 2.

Each AtCAD5 subunit is composed of two distinct domains, namely a Rossmann fold forming the nucleotide-binding domain

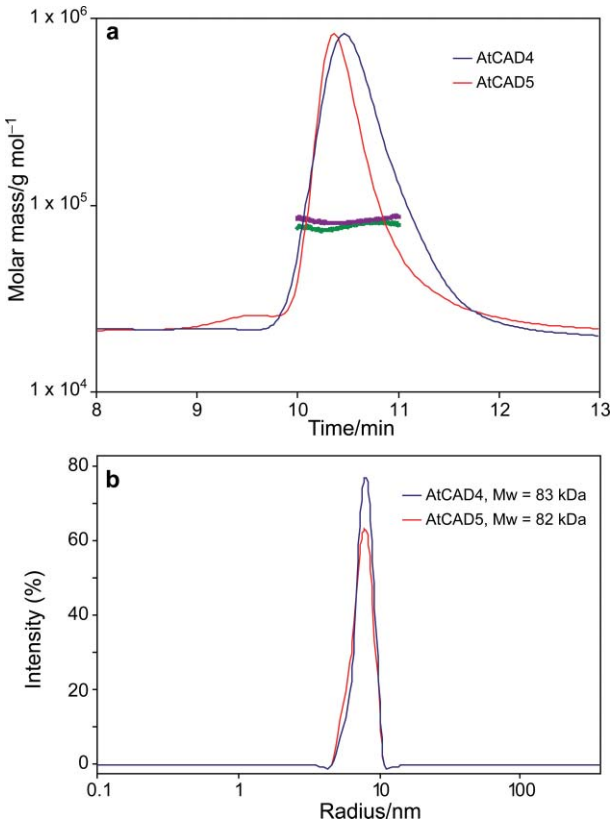


Fig. 3 Molecular mass determination of AtCAD5 and AtCAD4. (**a**) Multiangle laser light-scattering elution profile of AtCAD5 (red) and AtCAD4 (blue) (~2 mg cm^{-3} each). Elution profile is shown as molecular weight *versus* elution time. The *thin* solid lines represent changes in refractive index on an arbitrary scale that is proportional to protein concentration. The *thick* solid lines indicate calculated molecular masses (thick green line: AtCAD5, thick purple line: AtCAD4). (**b**) Dynamic light scattering data of AtCAD5 (red) and AtCAD4 (blue) (~2 mg cm^{-3} each). The calculated molecular radius and molecular weight are 3.86 nm and 82 kDa for AtCAD5, and 3.94 nm and 83 kDa for AtCAD4, respectively.

(residues 163–301) and a catalytic domain (residues 1–162 and 302–357) (Fig. 4). Interestingly, AtCAD5 has a similar overall fold to other ADH's that belong to the same MDR superfamily in spite of its relatively low level of sequence similarity (20–23%). Hence, in Fig. 2 and 4, the secondary structural elements of AtCAD5 can be conveniently described by comparison with horse liver ADH (8ADH) nomenclature. The nucleotide-binding domain is thus composed of a β-pleated sheet of six parallel strands (βA to βF) flanked by five helices (αA to αE), whereas the catalytic domain consists mainly of a core of antiparallel β-strands (β1 to β9) with six helical segments (α1 to α6) at the surface of the molecule. As with mammalian ADH's, there are two Zn^{2+} ions in AtCAD5, one catalytic and the other structural, with both being coordinated by residues solely from the catalytic domain (Fig. 5b–d). The catalytic Zn^{2+} is located inside a cleft formed between the two domains and is positioned at the bottom of the hydrophobic substrate-binding pocket, coordinated by four residues, Cys47, His69, Glu70 and Cys163, in an approximate tetrahedral symmetry (Fig. 5b). Notably, the $O^{\varepsilon 1}$ atom of the carboxyl side-chain of Glu70 coordinates the catalytic Zn^{2+} instead of water, with the latter being more typical of other Zn-dependent MDR's including

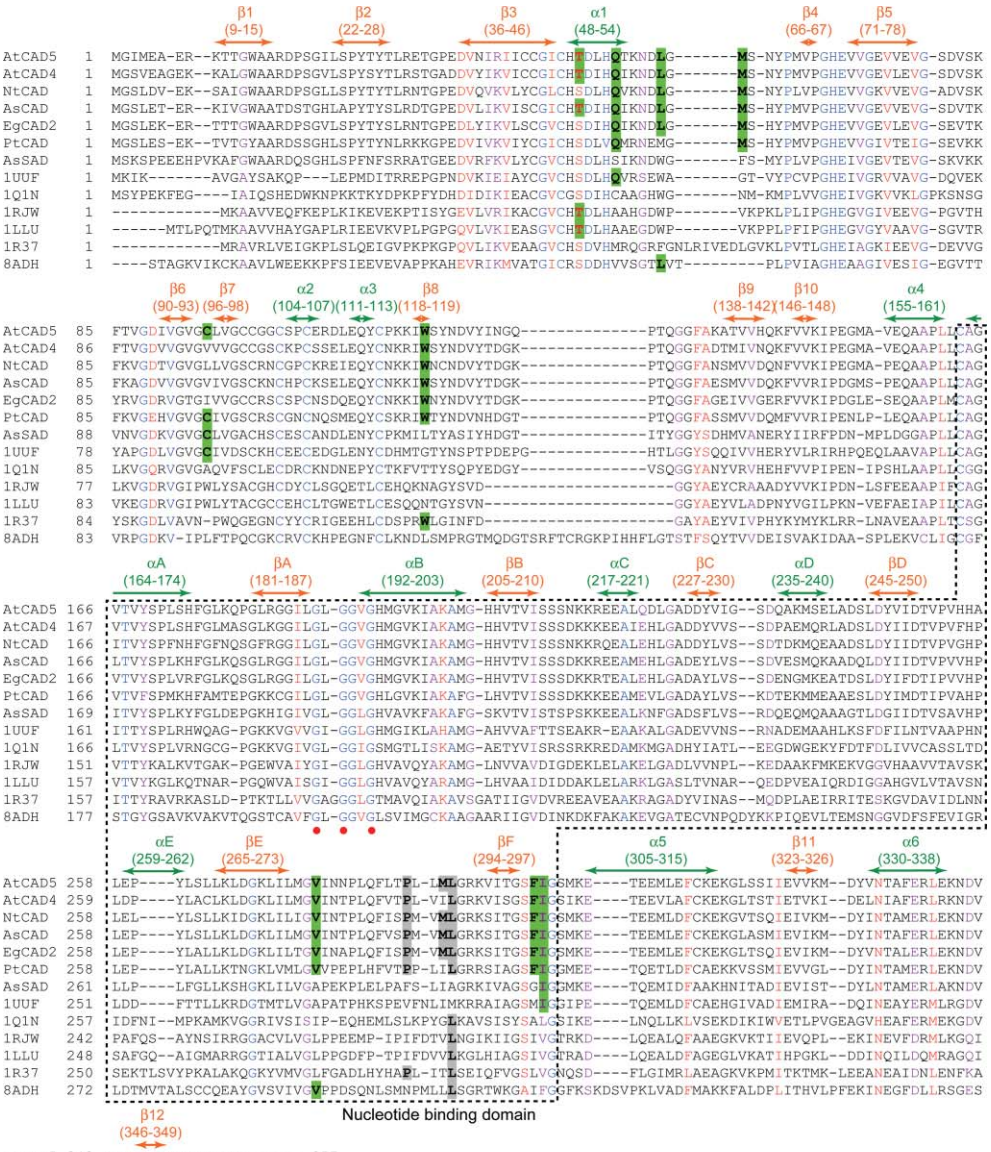


Fig. 4 Amino acid sequence comparisons of AtCAD5, AtCAD4, *Nicotiana tabacum* CAD (NtCAD, Genbank accession number X62344), aspen (*Populus tremuloides*) CAD (AsCAD, Genbank accession number AF217957), *Eucalyptus gunnii* CAD (EgCAD2, Genbank accession number X65631), *Pinus taeda* CAD (PtCAD, Genbank accession number Z37992), putative sinapyl alcohol dehydrogenase (SAD) from aspen (AsSAD, 1YQX, Genbank accession number AF273256), yahK from *Escherichia coli* (1UUF), a putative CAD from *Saccharomyces cerevisiae* (1Q1N), ADH from *Bacillus stearothermophilus* (1RJW), ADH from *Pseudomonas aeruginosa* (1LLU), ADH from *Sulfolobus solfataricus* (1R37) and horse liver ADH from *Equus caballus* (8ADH). Secondary structural elements of AtCAD5 are highlighted in colored bars on top of the corresponding sequence and the nucleotide-binding domain is boxed by a thin dotted line. The conserved glycine residues in GX(X)GXXG motif are marked with red dots. The conserved residues constituting a substrate-binding site are highlighted by green and gray colors indicating their belonging to different subunits. Eukaryotic 1YQX, 1Q1N and 8ADH are dimeric, whereas prokaryotic 1RJW, 1LLU and 1R37 are tetrameric Zn-dependent MDR's. (The corresponding data for 1UUF has not been published.)

horse liver ADH (8ADH).[22,23] By contrast, the structural Zn^{2+} is located in a short α-helix-containing loop (residues 98–116), which protrudes somewhat from the main catalytic domain. It is also in a tetrahedral coordination with cysteine residues, Cys100, Cys103, Cys106 and Cys114, respectively. Thus, taken together the observed structural features establish that the crystal structure of AtCAD5 belongs to the Zn^{2+}-dependent MDR family, as previously predicted.[14]

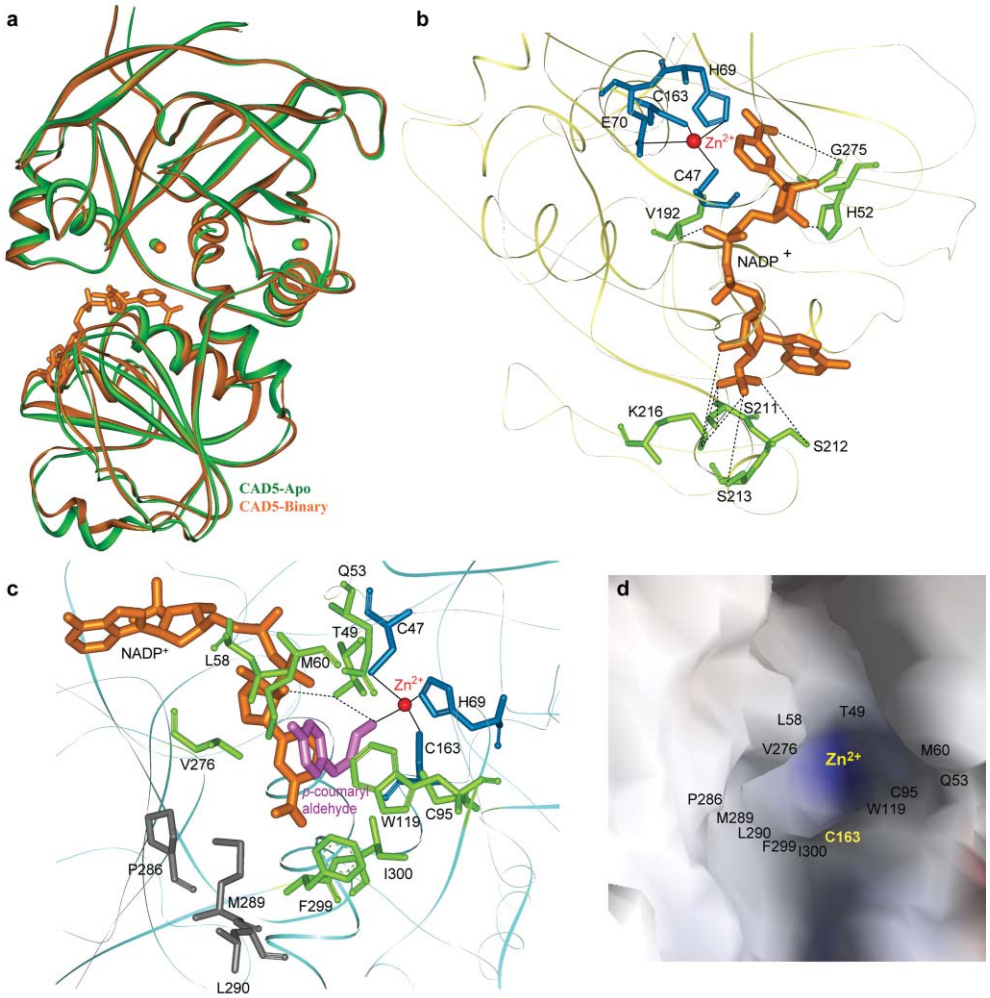


Fig. 5 (**a**) Superimposed views of AtCAD5 in its apo- and binary complex forms. (**b**) Structure of the substrate-binding pocket of the $NADP^+$ binary form of AtCAD5: The catalytic Zn^{2+} ion (red sphere) is tetrahedrally coordinated by Cys47, His69, Cys163 and Glu70 (blue) and the $NADP^+$ molecule (orange) is held by Val192, Ser211, Ser212, Ser213, Lys216 and Gly275 (green). (**c**) Ternary complex model of AtCAD5 with *p*-coumaryl aldehyde (**1**) showing the structure of the substrate-binding pocket. Participating residues from one subunit are marked in green and from the other subunit in gray. The catalytic Zn^{2+} (red sphere) is tetrahedrally coordinated by Cys47, His69, Cys163 (blue) and the aldehyde oxygen of *p*-coumaryl aldehyde (**1**) (purple). (**d**) Surface representation of AtCAD5 active site cavity. Twelve residues (black) that constitute the substrate-binding pocket are indicated together with the one Zn-coordinating residue, Cys163, (yellow). In figures **b** and **c**, possible hydrogen bonds are shown as black dotted lines.

In the binary complex, $NADP^+$ is located at the active site clefts between the catalytic and nucleotide-binding domains centered on the catalytic Zn^{2+} as described in detail below. Upon cofactor binding, however, the backbone conformation of AtCAD5 did not change significantly (Fig. 5a).

Comparison to other MDR's

In terms of quaternary structure, the homodimeric MDR's tend to occur in higher eukaryotes, whereas tetrameric MDR's are found in prokaryotes and lower eukaryotes. Thus, in order to gain a preliminary insight into the catalytic mechanism and other functionally significant issues operative, structural comparisons with other MDR's were next performed, through comparison of amino acid sequences (BLASTP), as well as 3D structures (Dali search), respectively.

First, amino acid sequence comparisons through a BLASTP search in the NCBI database revealed that AtCAD5 and AtCAD4 have the highest similarity (82.9 and 81.5%) and identity (76.5 and 75.1%) to a *bona fide* tobacco (*Nicotiana tabacum*) CAD.[14,24] It also indicated that the similarity and identity of aspen SAD to AtCAD5 and 4 were rather low (*i.e.*, 62.6/62.5% and 53.1/53.3%,[14] respectively), as were those for the putative yeast CAD (69.2/67.9% and 35.7/35.4%, respectively).

On the other hand, a Dali search[25] indicated that the most similar 3D structure to AtCAD5/4 currently in the PDB was the aspen SAD (AsSAD, 1YQX)[20] with a Z-score of 49.3 followed by the putative yeast CAD (1Q1N)[19] with a Z-score of 45.6. Additional PDB entries included an ADH-like protein yahK from *E. coli* (1UUF), an ADH from *Bacillus stearothermophilus* (1RJW)[26] with Z-scores of 45.4 and 45.3, respectively, as well as several other MDR's, including 1R37 from *Sulfolobus solfataricus*[27] and 1LLU from *Pseudomonas aeruginosa*[28] (Fig. 4).

The amino acid sequence identities of the latter to AtCAD5/4 were much lower, *i.e.* 1UUF (42.9/41.2%), 1RJW (33.2/30.9%), 1R37 (26.5/27.8%) and 1LLU (29.2/29.4%), respectively.

Detailed sequence comparisons of the above-mentioned eukaryotic dimeric (8ADH, 1YQX, 1Q1N) and prokaryotic tetrameric (1RJW, 1LLU, 1R37) Zn-dependent MDR's were also carried out. These enzymes have high sequential and structural similarities to AtCAD5, with the longest region of similarity located between Cys163 and Asp250, which in turn covers most of the nucleotide-binding domain (Fig. 4). Additionally, there is a substantial amount of sequential heterogeneity among the MDR's in both the N and C-terminus regions, which are structurally disordered in most cases. There are also several areas of deletions and insertions in the amino acid sequences of the MDR's compared. In particular, all the sequences shown in Fig. 4 have a significant deletion in the loop area between β8 and β9 relative to horse liver ADH (8ADH), even though the deletion size is less severe than for tetrameric MDR's. In horse liver ADH (8ADH), the flexible loop between β8 and β9 is located at the entrance to the active site contributing to substrate-binding, while at the same time restricting entry. The same deletion occurs in a putative yeast CAD, this being considered to explain its broad range of substrate specificity.[19]

Cofactor binding site

The initial *Fo–Fc* map using diffraction data from the binary complex crystal, and the coordinates of the apo AtCAD5, clearly shows the electron density corresponding to $NADP^+$ (Fig. 6). The nicotinamide ring is close to the catalytic Zn^{2+} at the bottom of the substrate-binding pocket (see Fig. 5b) and both the adenine and nicotinamide rings of $NADP^+$ are in the *syn*-conformation. In the apo-form, the cofactor-binding pocket contains several water molecules, thereby forming a hydrogen-bonding network with the side-chains of the lined residues, especially in areas where phosphate groups reside (figure not shown). As predicted earlier, the nicotinamide ring is in the proper orientation for the A-face specific hydride transfer from C4 to the corresponding substrate.[14] The binary complex with $NADP^+$ maintains the same tetrahedral geometry of Zn^{2+} being coordinated by Cys47, His69, Glu70 and Cys163, as for the apo-form, which is different from two reported $NADP^+$-binary complexes: ADH's from *Thermoanaerobacter brockii* (TbADH) and *Clostridium beijerinckii* (CbADH).[29] In the binary complex of CbADH and TbADH with $NADP^+$, however, the carboxyl side-chain of Glu is no longer able to coordinate with Zn^{2+} indicative of a perturbation of the Zn^{2+} coordination upon cofactor-binding.[29,30]

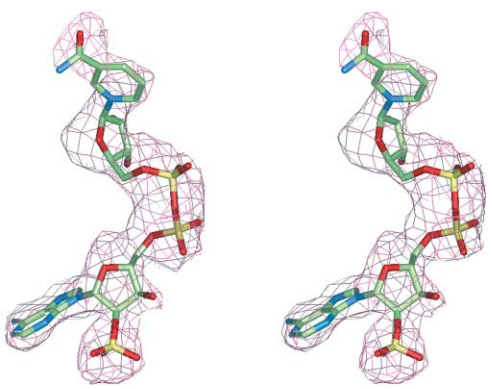


Fig. 6 Difference Fourier maps for $NADP^+$ binary complexes of AtCAD5. A difference (*Fo* − *Fc*) electron density map of the active site area is contoured at 2.0 σ.

Cofactor-binding by AtCAD5 occurs through three flexible loops. As for other typical NAD(P)(H) dependent enzymes,[9,10] AtCAD5 has a glycine-rich motif at the first β-α-β unit of the cofactor-binding domain (^{188}GLGGVG^{193}), which participates in binding the pyrophosphate group of $NADP^+$ through a helical dipole of αA. This typical GxGxxG spacing is shared by all Zn-dependent MDR's of high similarity in Fig. 4, except for ADH from *S. solfataricus* (1R37). In particular, the pyrophosphate group of the $NADP^+$ is within hydrogen-bonding distance to the backbone amide nitrogen of residue Val192 (Fig. 5b), thereby N-capping and compensating for the helix macro-dipole.[31] All the amino acids in this tight turn located between βA and αB thus show relatively high temperature factors which probably facilitates their interaction with $NADP^+$ *via* conformational flexibility.

Another loop region composed of residues 211–216 contains many conserved residues that also interact with $NADP^+$. The side-chain of the highly conserved Ser213 in this area enables a preference for NADP(H) over NAD(H). The Ser213 is located at the carboxy end of the βB strand forming a hydrogen bond with the 2′-phosphate group of NADP(H), and this position is normally occupied by aspartic acid in enzymes that preferentially bind NAD(H), *i.e.* to form a hydrogen bond to both hydroxyl groups of the adenine ribose of NAD(H).[31,32] Moreover, site-directed mutagenesis of the Ser212 amino acid in *E. gunnii* CAD2 to Asp212, followed by kinetic studies of both WT and mutant CAD2, had previously demonstrated the involvement of this residue in determining cofactor specificity.[33]

In addition to Ser213 in the AtCAD5, the side-chains of neighboring residues, Ser211, Ser212 and Lys216 are located in the loop between βB and αC and are highly conserved among *bona fide* CADs (Fig. 4). They form hydrogen bonds and electrostatic interactions with the 2′-phosphate of the adenine ribose (Fig. 5b), and the side-chain of Lys216 is also within hydrogen-bonding distance of the O3′ of the same ribose ring. The third region involves cofactor binding from 275–286 (Fig. 4), where the amide group of $NADP^+$ interacts with the backbone of Gly275 located in the loop between βE and βF connecting two domains of the enzyme (Fig. 5b).

Interestingly, the direct interaction of the above-mentioned three flexible loops, 188–191, 211–216 and 275–286, with the cofactor has been suggested to facilitate the dissociation of the products, a frequent rate-limiting step of many MDR enzymes.[19]

Substrate binding

In spite of our extensive efforts to crystallize the ternary complex for AtCAD5, the initial *Fo–Fc* maps for all complex data gave no significant electron density suitable for the positional refinement of the corresponding substrate *p*-coumaryl aldehyde (**1**). This was also the case for both the putative yeast CAD[19] and the putative aspen SAD.[20] On the other hand, the unique and properly sized pocket for substrate-binding was easily deduced to be near the catalytic Zn^{2+}. Indeed, after constraining the relevant aldehydic oxygen atom of the substrate *via* coordination to the catalytic Zn^{2+} (with the concomitant displacement of Glu70 from the Zn^{2+}), the

resulting A-face of the nicotinamide ring was thus positioned for *pro-R* hydride transfer to and from its C4 atom. For example, Fig. 5c shows the *pro-S* hydrogen of the C9 atom of the modeled *p*-coumaryl aldehyde (**1**) facing the C4 atom of the nicotinamide ring at a proper distance (~2.0 Å).

The putative substrate-binding site of AtCAD5 is lined with twelve residues, mostly hydrophobic amino acids. Nine residues from one subunit, Thr49, Gln53, Leu58, Met60, Cys95, Trp119, Val276, Phe299, Ile300 and three residues from the other subunit, Pro286, Met289 and Leu290, constitute the substrate-binding pocket (Fig. 5c, 5d). The Cα carbon position of those residues thus superimposes well with those of the horse liver ADH[23] and the *Eucalyptus gunnii* CAD2 model.[34] Notably, out of those 12 residues, nine amino acids are completely conserved among all *bona fide* CAD's (Fig. 4, residues in green). Two out of the remaining three also show a conservative heterogeneity with either serine or threonine for residue 49, and methionine or isoleucine for residue 289. Only the residue at 95 shows a higher polarity heterogeneity among *bona fide* CAD's, that is, Leu, Val, Ile and Cys (Fig. 4).

Interestingly, the entry region of the binding pocket of AtCAD5 is larger than that of horse liver ADH due to the major deletion between β8 and β9, although the precise physiological significance of this deletion cannot be gauged fully at this time.

As observed in some other Zn^{2+}-dependent MDR's, the modeled substrate through its aldehydic group ligation to the catalytic Zn^{2+} results in an extensive hydrogen-bonded network, thereby allowing coupling through hydride abstraction, transfer and protonation (Fig. 5c). Specifically, the aldehydic oxygen of the substrate is within hydrogen-bonding distance to the hydroxyl group of Thr49. In turn, the hydroxyl group of Thr49, which is conserved as either Thr or Ser in all ADH's compared, is within hydrogen-bonding distance of the O2′ of the nicotinamide ribose (not shown). As shown in Fig. 5c and Fig. 7, this residue is also in close proximity to the catalytic Zn^{2+}. Another highly conserved residue, His52, is also within hydrogen bonding distance of the O3′ of the nicotinamide ribose (Fig. 5b and 7). Therefore, both Thr49 and His52 not only fix the position of the nicotinamide ring during catalysis, but also permit hydride transfer from cofactor to substrate. As for several other Zn^{2+}-dependent ADH's, the Thr49 (or Ser) and His52 in AtCAD5, together with the O2′ and O3′ hydroxyl groups of the ribose ring, enable the proton relay mechanism through the hydrogen bond network among those functional groups, *i.e.* by

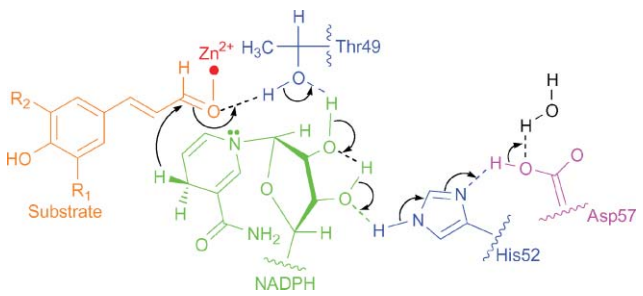


Fig. 7 Proposed proton shuttling mechanism during the reduction process in the active site of the AtCAD5. Solid arrows indicate the movement of two electrons among the functional groups during substrate reduction. The possible hydrogen bonds involved are shown with dotted lines.

shuttling a proton from the bulk solvent to the carbonyl oxygen of the aldehyde ligated to the catalytic Zn^{2+} (Fig. 7).

Significantly, there are no notable differences in the local conformations between AtCAD4 and AtCAD5. Indeed, most of the residues, within either contact or potential interacting distances to the substrates, were conserved between both. Only two of the 12 residues in the substrate-binding pocket differ; *i.e.* Cys95 and Met289 in AtCAD5 were replaced by Val96 and Ile290, respectively, in AtCAD4. This difference makes the binding pocket of the latter slightly more hydrophobic. Importantly, Cys95 that resides underneath Trp119 (Fig. 5c, 5d), and which corresponds to Phe93 in horse liver ADH (8ADH), has also been proposed to affect the substrate specificity in mammalian ADH's and *S. solfataricus* ADH (1R37) by steric hindrance.[27,35]

In an analogous manner to AtCAD5/4, provisional substrates sinapyl (**5**), coniferyl (**3**) and cinnamyl (**6**) aldehydes were also modeled into the coordinates of the binary $NADP^+$ complexes with the putative aspen SAD (AsSAD, 1YQX)[20] and yeast CAD (1Q1N),[19] respectively.

In this regard, the presumed active site of aspen SAD substantially differed from that of AtCAD5/4. Among the above-mentioned 12 residues constituting the proposed substrate-binding site, only two residues were identical to AtCAD5. Most of these substitutions were non-conservative, although there was no significant polarity change among them. As a result, the proposed tethering mechanism for the phenolic hydroxyl group of the corresponding substrates through the backbone carbonyl oxygen of Ala293 in the putative aspen SAD[20] cannot occur in the binding pocket of AtCAD5, due to both the changed conformation and the resulting steric hindrance. Specifically, there were several other somewhat symmetrical substitutions, including the two pairs of amino acid residues previously noticed, namely Leu58/Pro286 and Trp119/Phe299 in AtCAD5 to Trp61/Phe289 and Leu122/Gly302 in SAD, respectively[20] (Fig. 4). The bulky side-chains of the hydrophobic residues, Trp119/Phe299, substantially shrink the size of the binding pocket and its entrance in AtCAD4/5 compared to that in the putative aspen SAD. Moreover, the Val276 residue in AtCAD4/5, that is in close contact with the modeled substrate, is substituted to Ala in aspen SAD, thus increasing the size of the pocket further. Taken together, these differences in binding pocket (size and shape) would appear to enable the putative aspen SAD to display even higher substrate versatility than either the AtCAD5 or AtCAD4.

The substrate-binding site of the putative yeast CAD (1Q1N) also substantially differs from that of AtCAD5. This is reflected in the observation that there is only one amino acid conserved among the 12 residues, with two of the others having substitutions of different polarity. Consequently, the funnel-shaped cavity of the putative yeast CAD is also relatively larger than that of AtCAD5. This perhaps helps to explain further its very broad substrate specificity range,[19] as well as its being more polar than that of either AtCAD4/5 or the putative aspen SAD.

One of the two residues participating in the proton shuttle system, His52, is exposed to the bulk solvent, and is completely conserved among all of the compared MDR's. This occurs even though Thr49 is substituted to Ser in some MDR's, probably indicating that all of these reductases adopt an identical proton transfer system. By contrast, the alternative proton pathway that was proposed in horse liver ADH,[27,36] from the 3′OH of

a nicotinamide ring to the backbone carbonyl oxygen of Ile269 (Pro253 in AtCAD5), is not possible due to the relatively long distance between the two acceptor and donor groups. However, while this might be possible in the ternary complex (following conformational changes) this was not investigated.

Site-directed mutagenesis of the fourth Zn^{2+} coordination ligand (Glu70)

In most ADH's, including horse liver ADH, the fourth ligand of the catalytic Zn^{2+} is a water molecule hydrogen-bonded to the hydroxyl group of Thr49 (or Ser) and which can be replaced by substrate.[37] However, in some prokaryotic ADH's, the catalytic Zn^{2+} is instead coordinated to a strictly conserved Glu residue located opposite to the substrate binding pocket *e.g.* as observed for *S. solfataricus* (1R37), *C. beijerinckii* (CbADH) and *T. brockii* (TbADH).[27,29] Additionally, among eukaryotic ADH's, Zn^{2+} coordination to a Glu residue is very rare. Moreover, since AtCAD5 (coordinated to Glu70) is only the second example of such a coordination in the eukaryotes, following that of rat sorbitol dehydrogenase,[38] this potentially suggested a role in catalysis. Indeed, upon substrate-binding, the aldehydic oxygen of the substrate is envisaged to displace Glu70 *via* direct coordination to Zn^{2+}, so that it remains tetra-coordinated.[27,39,40] The resulting free side-chain of the Glu70 might then form a salt bridge to the nearby Arg345 as observed in apo-horse liver ADH.[23] Interestingly, this Glu residue is also one of the most highly conserved in class I ADH's, whose replacement in yeast ADH resulted in a considerable reduction in overall catalytic efficiency, even though the residue was not directly bound to the catalytic Zn^{2+} in that enzyme.[41] It has thus been proposed that the intermittent role of this dynamic Glu residue is both in facilitating the binding of the aldehydic substrate and in the subsequent product release.

The Zn^{2+}-coordination of the Glu70 is, however, observed in the apo-form of AtCAD5 and is maintained even after $NADP^+$ binding, even though the opposite occurred in the case of the $NADP^+$ complexes of both CbADH and TbADH.[27,29] We, therefore, considered it instructive to replace Glu70 with Ala70 *via* site-directed mutagenesis (see Experimental) to ascertain what effect, if any, this had on the overall catalytic efficacy. The resulting mutagenized AtCAD5 (Glu70Ala) was thus next heterologously expressed, purified to apparent homogeneity, and assayed for its capacity to convert the corresponding aldehydes **1–5**. The resulting mutagenized protein was catalytically inactive, in agreement with our hypothesis that perturbation of the Zn^{2+} catalytic center coordination would adversely affect catalysis.

Conclusions

Herein we report the crystal structures of AtCAD5 and the modeled structure of AtCAD4, the two catalytically most active CAD's from *A. thaliana.* The observed structures underscore previous observations reported for various MDR-related ADH's and provide us with a useful tool for a comparative study of other enzymes of putative CAD and SAD-like character. In particular, we identified 12 residues apparently constituting the substrate-binding site, which are well conserved among the *bona fide* CAD's from various species known thus far. By contrast, only 2 or 1 of these residues are conserved in the putative aspen SAD and in the putative yeast CAD, which in both cases results in significantly larger binding pockets. The true significance of such binding pocket differences will, however, only be objectively determined when the actual physiological roles of each are determined *in vivo* as previously noted by Bomati and Noel.[20] In terms of the CAD's whose roles are unambiguously established, it is of interest that their overall structures, including the location and number of the secondary structural elements, are somewhat different to horse liver ADH, *i.e.* only four of the 12 amino acids are conserved. This is also a reflection of the similarities and differences between CAD and horse liver ADH with respect to their substrate-binding pocket, *i.e.* in terms of preferentially utilizing aromatic and aliphatic substrates, respectively. Nevertheless, the horse liver ADH coordinates were such as to enable provisional modelling of the tobacco CAD[34] (discussed in Lewis *et al.*[3]) as were the coordinates of the *S. solfataricus* ADH[30] for AtCAD5 and AtCAD4.[14] Finally, the relatively tight substrate-binding pocket observed for the various hydroxycinnamyl aldehydes **1–5** examined using AtCAD5 and AtCAD4 now offers the opportunity to biotechnologically engineer the pocket to more readily utilize one substrate over another, *e.g.* to more efficiently channel metabolic flux, for example, for the formation of health-related lignans.

Experimental

Materials

Coniferyl (**3**), sinapyl (**5**) and cinnamyl (**6**) aldehydes were purchased from Aldrich whereas *p*-coumaryl (**1**), caffeyl (**2**), 5-hydroxyconiferyl (**4**) aldehydes and *p*-coumaryl (**7**), caffeyl (**8**), coniferyl (**9**), 5-hydroxyconiferyl (**10**) and sinapyl (**11**) alcohols were synthesized as described in Kim *et al.*[14]

Expression and purification of AtCAD5 and AtCAD4

Plasmids (pET151D-TOPO® TA vector, Invitrogen) harboring *AtCAD5* (GenBank accession number AY302082) and *AtCAD4* (GenBank accession number AY302081) were used to individually transform ER2566 *Escherichia coli* cells (New England BioLabs Inc.), with the cells grown at 37 °C with shaking (250 rpm) in Luria-Bertani (LB) medium supplemented with ampicillin (100 μg cm^{-3}). Over-expression of AtCAD5 (or AtCAD4) was induced by the addition of isopropyl β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM at mid-log phase (A_{600} = 0.5). After agitation (250 rpm) for 20 h at 20 °C, the cells were individually harvested by centrifugation (3000 × *g* for 20 min). The cell pellets were individually resuspended in lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0, 10% glycerol, 2 mM β-mercaptoethanol), and lyzed by sonication (5 × 10 s, model 450 sonifier®, Branson Ultrasonics Co.), with the lysates cleared by centrifugation (20 000 × *g* for 40 min). To each supernatant, Ni-nitrilotriacetic acid agarose (Qiagen, Hilden, Germany) was then added, with the slurry placed on a rocking shaker at 4 °C for 1 h. After eluting the unbound proteins and washing the resin ten times with washing buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, pH 8.0), the AtCAD5 (or AtCAD4) was then eluted stepwise with elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 100–300 mM imidazole pH 8.0) at a 50 mM increment of imidazole concentration.

Thereafter, AtCAD5 (or AtCAD4)-enriched fractions (150–250 mM imidazole) were pooled, concentrated and the buffer exchanged to 5 mM Na phosphate (pH 6.8) using Amicon YM10 membrane (Millipore). This concentrated sample was passed over a CHT-10 hydroxyapatite column (BioRad, 1 × 10 cm), which was pre-equilibrated in Na phosphate buffer (5 mM, pH 6.8), at a flow rate of 3.0 cm^3 min^{-1}. Each column was eluted with a linear gradient of Na phosphate (from 5 to 500 mM in 200 cm^3). AtCAD5 and AtCAD4 were individually eluted at 50 and 100 mM Na phosphate, respectively. Fractions containing AtCAD5 (or AtCAD4) were concentrated as described previously, with the buffer exchanged to 50 mM Tris–HCl (pH 8.0) containing EDTA (1 mM) and dithiothreitol (1 mM) (Buffer A). The resulting AtCAD5 (or AtCAD4)-enriched protein fraction was loaded onto a MonoQ™ GL10/100 anion exchange column (Amersham Biosciences) equilibrated in Buffer A at a flow rate of 2 cm^3 min^{-1}, with proteins eluted with a NaCl step gradient (0.05, 0.1, 0.2, 0.4, and 2 M; 20 cm^3 for each step); the catalytically active AtCAD5 (or AtCAD4) fractions were eluted at 0.1 M NaCl. Confirmation of the presence and purity of the AtCAD5 (or AtCAD4) was made by SDS-PAGE.

Size exclusion chromatography and multiangle laser light scattering

Determination of the molecular mass of AtCAD5 (and AtCAD4) in solution, using a static light scattering device, was performed as described previously in Youn *et al.*[10] by loading a solution of AtCAD5 or AtCAD4 (100 µl, 2 mg cm^{-3} in PBS) onto a KW-803 column (8 × 300 mm, Shodex, Japan) pre-equilibrated in PBS buffer.

Dynamic light scattering

The radius and molecular weight of AtCAD5 (and AtCAD4) were estimated using a DynaPro-Titan (Wyatt Technology Corp.) instrument at 22 °C. Purified AtCAD5 or AtCAD4 (2 mg cm^{-3}) in a freshly prepared Tris–HCl buffer (20 mM, pH 8.0), containing 1 mM EDTA and 1 mM dithiothreitol, were filtered through a polyvinylidene difluoride filter (0.1 µm, Millipore). Scattering data were acquired through accumulation (5 times) of 10 scans with 10 s per scan, with the laser intensity set to a range of 50–60% (30–36 mW). The corresponding molecular weight and radius were calculated using the software package 'DYNAMICS V6' supplied with the instrument.

Crystallization of AtCAD5

AtCAD5 was crystallized in a range of enzyme concentrations from 4 to 10 mg cm^{-3} in 20 mM Tris–HCl buffer (pH 8.0) containing 1 mM EDTA and 1 mM dithiothreitol using the hanging drop vapor diffusion method at two different temperatures of 277 K and 293 K. For crystallization, apo-AtCAD5 was diluted two-fold with the reservoir solution [20% (w/v) PEG 3350 and 0.2 M trilithium citrate tetrahydrate, pH 8.1] for a final hanging drop volume of 3 mm^3. Diamond-shaped crystals usually appeared after 3 days, and the best diffracting crystals with dimensions of ~0.2 × 0.2 × 0.8 mm were obtained in about 2 weeks. Apo-AtCAD5 crystallized in a tetragonal space group, $P4_12_12$ ($a = b =$ 54.22, c = 312.29 Å), with one molecule in an asymmetric unit. For binary complex crystals, purified AtCAD5 (10 mg cm^{-3}) in 30 mM Tris–HCl buffer (pH 8.5) containing 100 mM NaCl, 2 mM dithiothreitol and 1.5 mM $NADP^+$ was used. The binary AtCAD5 crystals were obtained by mixing the above solution of AtCAD5 (1.5 mm^3) with an equal volume of a reservoir solution containing 20% (w/v) PEG 3350 and 0.2 M diammonium tartrate (pH 6.6). The binary complex was also crystallized in the same tetragonal space group, $P4_12_12$, with corresponding unit cells of $a = b =$ 54.71, c = 303.93 Å. Diffraction data for the apo-form at 2.0 Å resolution, and the binary complex at 2.6 Å resolution, were collected using an ADSC Q210 CCD detector in the beam line 8.2.1 at Berkeley Advanced Light Source (ALS) and a Rigaku Saturn 92 CCD detctor/MicroMax-007 X-ray generator, respectively. All data were collected at a temperature of 100 K. Before crystal freezing, the corresponding crystals were soaked for 5 min in cryoprotectant (25% glycerol in the corresponding reservoir solution).

Structural solution and refinement

In order to obtain initial phase information of apo-AtCAD5 intensity data, the coordinates of the *E. coli* alcohol dehydrogenase-like protein yahK (1UUF) and the software package, AmoRe,[42] were used for molecular replacement. Rigid body refinement was carried out with X-PLOR,[43] beginning with the best solution of molecular replacement. The initial crystallographic R-value of a solution was 40.0%, with the data resolution ranging from 15.0 to 3.0 Å. After several cycles of positional and temperature factor refinements, and a series of simulated annealing omit maps, a reasonable quality electron density map was achieved and all residues were fitted. The structure of the binary complex of AtCAD5 was solved by the molecular replacement method using the refined coordinates of apo-AtCAD5. The final R-factors for the apo-form and the binary complex were 19.5% (R_{free} = 23.5% for the random 5% data) and 20.1% (R_{free} = 23.0% for the random 5% data), respectively (Table 1). The number of reflections above 2σ for the apo-form was 25 958 (98.2% completeness) between 10.0 and 2.0 Å resolution. The crystals of the binary complex did not diffract as well as the apo-form and gave reflection numbers of 9631 (above 2σ, 93.2% completeness) between 10.0 and 2.6 Å resolution. The root mean square deviations (rmsd) (from ideal geometry) of the final coordinates corresponding to the apo-form and the binary complex are 0.01 and 0.02 Å for bonds and 2.6 and 2.9° for angles, respectively. All AtCAD5 coordinates have been deposited in the Protein Data Bank (apo-form: 2CF5, binary complex: 2CF6).

In turn, amino acid substitutions, insertions, and deletions for the AtCAD4 were next performed by using the graphics program O,[44] starting from the refined coordinates of AtCAD5, followed by quick-energy minimization by using X-PLOR[43] with potential function parameters of CHARMM19 as described previously.[14] The initial position of the substrate was obtained through the solid docking module on QUANTA (BioSYM/Micron Separations), which is based on conformational space, followed by a quick-energy minimization by X-PLOR.[43]

Site-directed mutagenesis

Forward (5′-CATGGTTCCTGGGCATGC*G*GTGGTAGGGG AAGTAG-3′) and reverse (5′-CTACTTCCCCTACCACC*G*C ATGCCCAGGAACCATG-3′) primers were designed and

Table 1 Crystallographic data for AtCAD5

Data	Apo-form	Binary complex[a]
Beam line	ALS 8.2.1	WSU (MM007)
Wavelength/Å	1.07812	1.5418
Resolution/Å	50 to 2.0	50 to 2.6
Space group	$P4_12_12$	$P4_12_12$
Cell dimensions/Å	$a = 54.22$	$a = 54.71$
	$b = 54.22$	$b = 54.71$
	$c = 312.99$	$c = 303.93$
Asymmetric unit	1 molecule	1 molecule
Total observations	129 883	59 808
Unique reflections	32 473	14 177
Completeness (%)	98.2	93.2
R_{sym} (%)[b,c]	5.8 (13.3)	6.5 (13.5)
Refinement		
Resolution/Å	10 to 2.0	10 to 2.6
Number of reflections	25 958	9631
R_{cryst} (%)[d]	19.5	20.1
R_{free} (%)[e]	23.5	23.0
rmsd[f] bonds/Å	0.011	0.012
rmsd angles/°	3.019	3.463
Number of atoms		
Protein + ion + ligand	2667	2715
Water	199	140

[a] AtCAD5 + $NADP^+$. [b] Numbers in parentheses refer to the highest shell. [c] $R_{sym} = \Sigma\,|I_h - \langle I_h\rangle|\,/\Sigma I_h$, where $\langle I_h\rangle$ is the average intensity over symmetry equivalent reflections. [d] $R_{cryst} = \Sigma\,|F_{obs}-F_{calc}|\,/\Sigma F_{obs}$, where summation is over the data used for refinement. [e] R_{free} was calculated as for R_{cryst} using 5% of the data that was excluded from refinement. [f] Root mean square deviations.

synthesized (Invitrogen, Carlsbad, USA) to convert AtCAD5 Glu70 into alanine (asterisks indicate mismatch for the Glu→Ala substitution). Site specific mutagenesis was carried out with a QuikChange XL site-directed mutagenesis kit (Stratagene, La Jolla, USA) following the manufacturer's instructions with PCR conditions as follows: initial denaturation at 95 °C for 1 min, followed by 18 cycles of 95 °C for 50 s, 60 °C for 50 s and 68 °C for 6 min, with 7 min at 68 °C and an indefinite hold at 4 °C. After PCR completion, the PCR product was treated with *Dpn* I for 1 h at 30 °C to digest the non-mutated parental DNA template, and was transformed into *E. coli* TOP10 cells. After selection on LB plates containing 100 μg cm^{-3} carbenicillin, a positive clone, containing the mutation Glu70→Ala70, was confirmed by sequencing both strands (using pTrcHis2 forward and reverse primers (Invitrogen)) to ensure that there were no other mutation(s) in the open reading frame as a result of the PCR. Heterologous expression of the Glu70→Ala70 mutant as well as of recombinant AtCAD5 was carried out as described above. Both were purified and assayed using aldehydes **1–5** as described in Kim *et al.*[14]

Acknowledgements

This research was supported in part by the National Institutes of Health (GM66173), the National Science Foundation (MCB-0417291), the United States Department of Energy (DE-FG-0397ER20259), the G. Thomas and Anita Hargrove Center for Plant Genomic Research, McIntire-Stennis, and the Murdock Charitable Trust. We thank D. J. Pouchnik for DNA sequencing.

References

1 G. G. Gross, J. Stöckigt, R. L. Mansell and M. H. Zenk, *FEBS Lett.*, 1973, **31**, 283–286.
2 R. L. Mansell, G. G. Gross, J. Stöckigt, H. Franke and M. H. Zenk, *Phytochemistry*, 1974, **13**, 2427–2435.
3 N. G. Lewis, L. B. Davin and S. Sarkanen, in *Comprehensive Natural Products Chemistry*, ed. Sir D. H. R. Barton, K. Nakanishi and O. Meth-Cohn, Elsevier, Oxford, 1999, vol. 3, pp. 617–745.
4 A. M. Anterola, J.-H. Jeon, L. B. Davin and N. G. Lewis, *J. Biol. Chem.*, 2002, **277**, 18272–18280.
5 A. T. Dinkova-Kostova, D. R. Gang, L. B. Davin, D. L. Bedgar, A. Chu and N. G. Lewis, *J. Biol. Chem.*, 1996, **271**, 29473–29482.
6 M. Fujita, D. R. Gang, L. B. Davin and N. G. Lewis, *J. Biol. Chem.*, 1999, **274**, 618–627.
7 N. G. Lewis and L. B. Davin, in *Comprehensive Natural Products Chemistry*, ed. Sir D. H. R. Barton, K. Nakanishi and O. Meth-Cohn, Elsevier, Oxford, 1999, vol. 1, pp. 639–712.
8 Z.-Q. Xia, M. A. Costa, H. C. Pélissier, L. B. Davin and N. G. Lewis, *J. Biol. Chem.*, 2001, **276**, 12614–12623.
9 T. Min, H. Kasahara, D. L. Bedgar, B. Youn, P. K. Lawrence, D. R. Gang, S. C. Halls, H. Park, J. L. Hilsenbeck, L. B. Davin, N. G. Lewis and C. Kang, *J. Biol. Chem.*, 2003, **278**, 50714–50723.
10 B. Youn, S. G. A. Moinuddin, L. B. Davin, N. G. Lewis and C. Kang, *J. Biol. Chem.*, 2005, **280**, 12917–12926.
11 C. Canel, R. M. Moraes, F. E. Dayan and D. Ferreira, *Phytochemistry*, 2000, **54**, 115–120.
12 S. P. Borriello, K. D. R. Setchell, M. Axelson and A. M. Lawson, *J. Appl. Bacteriol.*, 1985, **58**, 37–43.
13 K. H. Teoh, J. D. Ford, M.-R. Kim, L. B. Davin and N. G. Lewis, in *Flaxseed in Human Nutrition*, ed. L. U. Thompson and S. C. Cunnane, AOCS Press, Champaign, IL, 2nd edn, 2003, pp. 41–62.
14 S.-J. Kim, M.-R. Kim, D. L. Bedgar, S. G. A. Moinuddin, C. L. Cardenas, L. B. Davin, C. Kang and N. G. Lewis, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 1455–1460.
15 H. Jörnvall and J.-O. Höög, *Alcohol Alcohol.*, 1995, **30**, 153–161.
16 E. Nordling, B. Persson and H. Jörnvall, *Cell. Mol. Life Sci.*, 2002, **59**, 1070–1075.
17 R. Sibout, A. Eudes, G. Mouille, B. Pollet, C. Lapierre, L. Jouanin and A. Séguin, *Plant Cell*, 2005, **17**, 2059–2076.
18 C.-I. Brändén and H. Eklund, *Experientia*, 1980, **S36**, 40–84.
19 E. Valencia, C. Larroy, W. F. Ochoa, X. Parés, I. Fita and J. A. Biosca, *J. Mol. Biol.*, 2004, **341**, 1049–1062.
20 E. K. Bomati and J. P. Noel, *Plant Cell*, 2005, **17**, 1598–1611.
21 L. Li, X. F. Cheng, J. Leshkevich, T. Umezawa, S. A. Harding and V. L. Chiang, *Plant Cell*, 2001, **13**, 1567–1585.
22 F. Colonna-Cesari, D. Perahia, M. Karplus, H. Eklund, C. I. Brändén and O. Tapia, *J. Biol. Chem.*, 1986, **261**, 15273–15280.
23 H. Eklund and C. I. Brändén, in *Biological Macromolecules and Assemblies*, ed. F. A. Jurnak and A. McPherson, John Wiley & Sons, Inc., London, 1987, vol. 3, pp. 75–142.
24 M. E. Knight, C. Halpin and W. Schuch, *Plant Mol. Biol.*, 1992, **19**, 793–801.
25 L. Holm and C. Sander, *J. Mol. Biol.*, 1993, **233**, 123–138.
26 C. Ceccarelli, Z.-X. Liang, M. Strickler, G. Prehna, B. M. Goldstein, J. P. Klinman and B. J. Bahnson, *Biochemistry*, 2004, **43**, 5266–5277.
27 L. Esposito, I. Bruno, F. Sica, C. A. Raia, A. Giordano, M. Rossi, L. Mazzarella and A. Zagari, *Biochemistry*, 2003, **42**, 14397–14407.
28 I. Levin, G. Meiri, M. Peretz, Y. Burstein and F. Frolow, *Protein Sci.*, 2004, **13**, 1547–1556.
29 Y. Korkhin, A. J. Kalb, M. Peretz, O. Bogin, Y. Burstein and F. Frolow, *J. Mol. Biol.*, 1998, **278**, 967–981.
30 L. Esposito, F. Sica, C. A. Raia, A. Giordano, M. Rossi, L. Mazzarella and A. Zagari, *J. Mol. Biol.*, 2002, **318**, 463–477.
31 P. J. Barbosa Pereira, S. Macedo-Ribeiro, A. Párraga, R. Pérez-Luque, O. Cunningham, K. Darcy, T. J. Mantle and M. Coll, *Nat. Struct. Biol.*, 2001, **8**, 215–220.
32 A. J. Powell, J. A. Read, M. J. Banfield, F. Gunn-Moore, S. D. Yan, J. Lustbader, A. R. Stern, D. M. Stern and R. L. Brady, *J. Mol. Biol.*, 2000, **303**, 311–327.

33 V. Lauvergeat, K. Kennedy, C. Feuillet, J. H. McKie, L. Gorrichon, M. Baltas, A. M. Boudet, J. Grima-Pettenati and K. T. Douglas, *Biochemistry*, 1995, **34**, 12426–12434.
34 J. H. McKie, R. Jaouhari, K. T. Douglas, D. Goffner, C. Feuillet, J. Grima-Pettenati, A. M. Boudet, M. Baltas and L. Gorrichon, *Biochim. Biophys. Acta*, 1993, **1202**, 61–69.
35 C. L. Stone, T.-K. Li and W. F. Bosron, *J. Biol. Chem.*, 1989, **264**, 11112–11116.
36 J. Luo and T. C. Bruice, *J. Am. Chem. Soc.*, 2001, **123**, 11952–11959.
37 B. L. Vallee and D. S. Auld, *Proc. Natl. Acad. Sci. U. S. A.*, 1990, **87**, 220–224.
38 K. Johansson, M. El-Ahmad, C. Kaiser, H. Jornvall, H. Eklund, J. Hoog and S. Ramaswamy, *Chem.–Biol. Interact.*, 2001, **130–132**, 351–358.
39 T. D. Colby, B. J. Bahnson, J.K. Chin, J. P. Klinman and B. M. Goldstein, *Biochemistry*, 1998, **37**, 9295–9304.
40 I. Levin, R. Schwarzenbacher, D. McMullan, P. Abdubek, E. Ambing, T. Biorac, J. Cambell, J. M. Canaves, H. J. Chiu, X. Dai, A. M. Deacon, M. DiDonato, M. A. Elsliger, A. Godzik, C. Grittini, S. K. Grzechnik, E. Hampton, L. Jaroszewski, C. Karlak, H. E. Klock, E. Koesema, A. Kreusch, P. Kuhn, S. A. Lesley, T. M. McPhillips, M. D. Miller, A. Morse, K. Moy, J. Ouyang, R. Page, K. Quijano, R. Reyes, A. Robb, E. Sims, G. Spraggon, R. C. Stevens, H. van den Bedem, J. Velasquez, J. Vincent, F. von Delft, X. Wang, B. West, G. Wolf, Q. Xu, K. O. Hodgson, J. Wooley and I. A. Wilson, *Proteins*, 2004, **56**, 629–633.
41 O. Kleifeld, S. P. Shi, R. Zarivach, M. Eisenstein and I. Sagi, *Protein Sci.*, 2003, **12**, 468–479.
42 J. Navaza, *Acta Crystallogr., Sect. A: Fundam. Crystallogr.*, 1994, **A50**, 157–163.
43 A. T. Brünger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson and G. L. Warren, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 1998, **D54**, 905–921.
44 T. A. Jones, J.-Y. Zou, S. W. Cowan and M. Kjeldgaard, *Acta Crystallogr., Sect. A: Found. Crystallogr.*, 1991, **A47**, 110–119.